LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Essays in Empirical Asset Pricing

Jiantao Huang

Thesis submitted to the Department of Finance of the London School of

Economics and Political Science for the degree of

Doctor of Philosophy

April 2022

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 68,640 words.

Statement of conjoint work

I confirm that Chapter 2 is jointly co-authored with Svetlana Bryzgalova and Christian Julliard, and I contributed 33% of this work.

I confirm that Chapter 3 is jointly co-authored with Ran Shi, and I contributed 50% of this work.

Acknowledgements

Abstract

In this thesis, I develop new econometric techniques to measure and understand the sources of economic risks in equity markets.

The first chapter studies frequency-dependent risks in the factor zoo. My approach generalizes canonical principal component analysis (PCA) by exploiting frequency-dependent information in asset returns. Empirically, the linear stochastic discount factor (SDF) composed of the first few low-frequency principal components (PCs) capture all the risk premia in asset returns. It also explains well the cross-section of characteristic-sorted portfolios. In contrast, high-frequency and canonical PCA have inferior performance since they fail to identify slow-moving information in asset returns. Moreover, I decompose the low-frequency SDF into two orthogonal priced components. The first component is constructed by high-frequency or traditional PCA. It is almost serially uncorrelated and relates to discount-rate news, intermediary factors, jump risk, and investor sentiment. The second component is slow-moving and captures business-cycle risks related to consumption and GDP growth. Hence, only low-frequency PCA identifies the second persistent component emphasized by many macro-finance models.

The second chapter (with Svetlana Bryzgalova and Christian Julliard) proposes a novel framework for linear asset pricing models: simple, robust, and applicable to high-dimensional problems. For (potentially misspecified) standalone models, it provides reliable estimates of risk prices for both tradable and non-tradable factors and detects those weakly identified. For competing factors and (possibly non-nested) models, the method automatically selects the best specification – if a dominant one exists – or provides a Bayesian model averaging (BMA-SDF) if there is no clear winner. We analyse 2.25 quadrillion models generated by a large set of factors and find that the BMA-SDF outperforms existing models in- and out-of-sample.

The third chapter (with Ran Shi) develops a Bayesian approach to quantify model uncertainty about linear SDFs, defined as the entropy of posterior model probabilities. We show that model uncertainty displays massive fluctuations over time, and high model uncertainty coincides with major market events. These observations hold not only in US markets but also in European and Asian Pacific equity markets. Moreover, positive model uncertainty shocks relate to sharp outflows from US equity mutual funds but significant inflows to government bond funds, with effects persisting for three years. In survey data, investors tend to be more pessimistic about equity performance during periods of higher model uncertainty.

Contents

List of Figures

List of Tables

Chapter 1

Frequency Dependent Risks in the Factor Zoo

Jiantao Huang[1]

1.1 Introduction

Explaining the cross-section of expected returns has been an important challenge in asset pricing literature. Researchers have acknowledged that the consumption-based capital asset pricing model (CCAPM)[2] provides little explanatory power, which has inspired a wide variety of new models. Some models introduce slow-moving components into the stochastic discount factor (SDF), such as the surplus consumption ratio in Campbell and Cochrane (1999) and the stochastic mean and variance of consumption growth in Bansal and Yaron (2004). In other models, the SDF consists only of fast-moving components, e.g., output jumps in Barro (2006), the intermediary's consumption growth in He and Krishnamurthy (2013), and sentiment-driven demand shocks in Kozak, Nagel, and Santosh (2018). Identifying the key determinants of the SDF, particularly the slow-moving components that are notoriously difficult to measure (see Alvarez and Jermann (2005)), remains an open question. This paper addresses this question through the lens of frequency-dependent risks. In addition, I seek to understand the frequency-specific drivers of expected returns and explore the role of distinct asset pricing models at different frequencies.

_____

[1]Any errors or omissions are my responsibility. I thank Thummim Cho, Ian Martin, Cameron Peng, Ran Shi, Dimitri Vayanos, and seminar participants at the London School of Economics, Tsinghua University PBC School, Chinese University of Hong Kong, University of British Columbia, Peking University HSBC Business School, and University of Hong Kong for helpful comments, discussions, and suggestions. I am particularly grateful to Christian Julliard, Dong Lou, and Svetlana Bryzgalova for their invaluable guidance and support.

[2]I refer to earlier versions of CCAPM developed by Rubinstein (1976), Lucas (1978), and Breeden (1979).

This paper generalizes canonical principal component analysis (PCA) to construct latent factors that explain the cross-section of monthly expected returns. The key novelty of my approach is that I exploit frequency-dependent information of asset returns to estimate latent factors. Using standard Fourier transform, I decompose the covariance matrix of monthly returns into high- and low-frequency components and estimate systematic factors in each frequency interval. I denote them as high- and low-frequency principal components (PCs) and use them as monthly tradable proxies for short- and long-term systematic risks.

When do frequency-dependent risks matter? I show that when asset returns are independent, high- and low-frequency latent factors are precisely identical to the canonical PCs. In other words, only when asset returns deviate from the independence assumption we need to study frequency-dependent risks. Empirically, low-frequency PCs contain a persistent element missed by high-frequency and conventional PCs. Moreover, this persistent missing part is essential in explaining expected returns and reflects business-cycle risks.

Asset pricing models often make parametric assumptions enforcing whether fast- or slow-moving economic shocks drive the SDF. Rather than assuming the existence of fast- or slow-moving elements, this paper lets the data speak and suggests that both two components are priced but reflect different economic fundamentals. The key for detecting the slow-moving component is the rich persistent information in the factor zoo. For example, Gupta and Kelly (2019) find that 48 of 65 investment anomalies have significantly positive AR(1) coefficients.[3] The low-frequency PCA boosts the signal of persistent information in the factor zoo and combines the factors' persistence into a few low-frequency PCs. Instead, the high-frequency or conventional PCA fails to detect them.

My empirical results are based on a large cross-section of 78 portfolios.[4] I divide the whole sample equally into two subsamples. I estimate the factor compositions and risk prices of frequency-specific PCs in the first subsample and examine their *out-of-sample (OOS)* performance in the second subsample. In the main analysis, the LF interval is between three and ten years, and I interpret it as the business-cycle frequency interval. In contrast, the HF interval is between zero and three years. The empirical findings are fourfold.

First, the SDF is sparse only in the space of low-frequency PCs. The low-frequency SDF comprising the seven largest low-frequency PCs is the "proper" benchmark: It yields an OOS Sharpe ratio of around 0.37 per month. Additional low-frequency PCs are redundant. In contrast, I need more than 20 high-frequency or canonical PCs to gain a comparable Sharpe ratio. Since high-frequency components account for 94% of time-series variations in asset returns, the large canonical PCs are virtually equivalent to the high-frequency latent

---

[3]The other 11 have positive yet insignificant coefficients. No factor has significantly negative coefficients.

[4]Test assets are long and short legs sorted by 39 firm features in Kozak, Nagel, and Santosh (2020).

factors. I also split the whole high-frequency interval into a few subintervals, but the SDF is dense even in the space of highly fast-moving factors (with a cycle length shorter than three months). Past research (e.g., Kozak, Nagel, and Santosh (2018, 2020)) often uses the first few PCs of single-period returns (identical to high-frequency PCs in the data) to construct the SDF. My paper shows that this standard practise can be improved by exploiting frequency-dependent information in asset returns.

Second, the low-frequency SDF cannot be explained by the high-frequency SDF or celebrated factor models in Fama and French (1993, 2015), Carhart (1997), and Hou, Xue, and Zhang (2015). Monthly alphas of the low-frequency SDF are significantly greater than 0.6%. In contrast, the low-frequency SDF can entirely span the high-frequency one. This evidence provides further justification for using the low-frequency SDF as the benchmark.

Third, I decompose the low-frequency SDF into fast- and slow-moving components. The first component is the optimal portfolio composed of high-frequency PCs. This SDF component is nearly identical to the SDF constructed by Kozak, Nagel, and Santosh (2018, 2020). I observe that the high-frequency SDF is almost serially uncorrelated and yields a monthly Sharpe ratio of 0.29, so I denote it as the fast-moving component. However, it still misses an essential slow-moving element. I project the low-frequency SDF into the space of the high-frequency SDF and extract an orthogonal part, denoted as the *missing-SDF*. This missing part, displaying a persistent dynamic according to the variance ratio test, explains 30% of the time-series variation of the low-frequency SDF and earns a monthly Sharpe ratio of 0.24.

Fourth, fast- and slow-moving components of the low-frequency SDF embody entirely different sources of economic risks. Precisely, the high-frequency SDF is correlated with market discount-rate news in Campbell and Vuolteenaho (2004), intermediary factors in He, Kelly, and Manela (2017), market jump risk proxied by the VXO index, and the sentiment-driven demand shocks from Baker and Wurgler (2006) investor sentiment. Instead, the slow-moving part of the SDF is related to consumption and GDP growth. It also predicts the next quarter economic growth. Hence, the missing-SDF reflects slow-moving business-cycle risks.

My empirical findings have implications for asset pricing models, which link the SDFs to different economic fundamentals. Macro-finance models often use persistent shocks to macro variables, such as the stochastic mean of consumption growth, to magnify their prominence in the SDF. My paper confirms that asset returns carry useful persistent information related to macro fundamentals, but I can identify them only at low frequencies. My paper also reconciles the disconnection between asset returns and some macro fundamentals. For example, asset returns and consumption growth are almost uncorrelated at the quarterly frequency, so asset pricing seems to disconnect with the macroeconomy in short horizons.

My paper confirms that the large PCs of short-horizon returns are unrelated to consumption growth. After removing high-frequency variations from asset returns, the remaining slow-moving component strongly correlates with macro fundamentals. Therefore, identifying the slow-moving component is salient for understanding and testing macro-finance models.

Furthermore, macro risks are insufficient to explain the cross-section. The fast-moving component of the benchmark SDF commands a significant price of risk but is orthogonal to macro risks. Instead, the demand shocks from sentiment investors, the shocks to the intermediary sector, and market discount-rate news are essential in understanding the fast-moving component of the SDF (high-frequency SDF). Hence, different asset pricing models explain either fast- or slow-moving components of the SDF but not both.

There are two appealing benefits to studying asset returns at different frequencies. *First*, it helps to explore the dynamics of state variables in the SDF. I decompose the variance of an SDF (equivalently, the maximal achievable Sharpe ratio) into frequency-specific components. Also, I prove that if the SDF has a larger variance at high (low) frequencies, state variables entering the SDF are, on average, more fast-moving (slow-moving). Since a sparse low-frequency SDF embodies a significantly higher Sharpe ratio than a high-frequency one, slow-moving state variables are empirically more prominent than fast-moving ones.[5]

Second, frequency-dependent PCA strengthens the signal of some systematic factors. Generally, a slow-moving (fast-moving) latent factor has a stronger signal at low (high) frequencies. Suppose a weak latent factor explains a tiny proportion of single-period returns.[6] In that case, the canonical PCA fails to identify it. However, frequency-specific PCA can recover this weak factor if its variance is large enough in a specific frequency interval. This paper shows that the low-frequency PCA recovers some essential priced weak factors with strong enough signals only at low frequencies. Instead, the high-frequency and canonical PCA identify them as idiosyncratic noises, so many small high-frequency and canonical PCs are needed to attain the same Sharpe ratio as a sparse low-frequency SDF.

It is worth noting that economic theory predicts the sparsity of latent factor models. The absence of near-arbitrage opportunities in Kozak, Nagel, and Santosh (2018) argues that only the largest PCs enter the SDF. However, this paper observes some small high-frequency (also canonical) PCs bringing nontrivial risk premia, so the absence of near-arbitrage opportunities fails. One explanation is that some economic shocks, such as the stochastic mean of consumption growth, are slow-moving and explain only a tiny fraction of single period returns. Hence, traditional PCA fails to detect these small but persistent shocks. Suppose

---

[5]The importance of a state variable X_t comes from the variance of X_t and its risk price squared (b_X^2). In latent factor models, I can identify only $\text{Var}(X_t)b_X^2$ rather than $\text{Var}(X_t)$ and b_X^2 individually.

[6]Onatski (2012) and Lettau and Pelger (2020a) assume that the variance of a weak factor does not grow as the number of test assets converges to infinity.

market participants have Epstein-Zin preferences as in the long-run risk model. In that case, persistent shocks to economic fundamentals command sizable risk premia and constitute a considerable part of the SDF. Since the low-frequency PCA successfully captures these slow-moving elements, we observe the sparsity of the low-frequency SDF.

1.1.1 Related Literature

This paper mainly contributes to two strands of literature. The first closely related branch of literature is the study of asset pricing models at different frequencies. We have known for a long time that both CAPM and CCAPM have better performance in the long horizon. For example, Handa, Kothari, and Wasley (1989) show that the size effect becomes statistically insignificant when the market beta is estimated using annual returns. Parker and Julliard (2005) measure ultimate consumption risk at a horizon of three years and document that it explains a large proportion of expected returns. Brennan and Zhang (2020) derive the CAPM with a stochastic investor horizon, and their estimates show that the probability distribution of investor horizons puts a massive weight on the interval between 8 and 20 months. Chernov, Lochstoer, and Lundeby (2022) test asset pricing models using multi-horizon returns and report that single-period estimates of those models typically do a poor job of explaining long-term returns.

However, all the above papers study factor models at a specific frequency instead of in a frequency interval. A few recent papers adopt spectral analysis to study frequency-dependent risks. First, Dew-Becker and Giglio (2016) study frequency-dependent risk prices in consumption-based models and show that only the long-run risk model can explain asset returns. Instead, my paper does not make a parametric assumption of the SDF. I construct the SDF using latent factors of asset returns and find that the SDF contains a huge fast-moving component that the consumption risk cannot explain.

Second, Bandi, Chaudhuri, Lo, and Tamoni (2021) use a Wold representation of the CAPM beta. Only the business cycle components within the frequency interval between 32 and 64 months can price the cross-sections. One key feature of their approach is assuming a vector autoregressive (VAR) process for state variables. In contrast, my paper takes a nonparametric point of view and is more robust to the model misspecification of the state vector dynamics. In addition, we have different economic interpretations. Their paper claims that the business-cycle component of the market beta captures delayed price adjustments to new information in the market portfolio. Instead, my paper finds low-frequency systematic factors capture business-cycle risks, but short-term factors miss them.

Last but not least, Neuhierl and Varneskov (2021) decompose the covariance between asset returns and pricing factors via the Fourier transform and study the frequency-dependent

risks. My paper improves their framework in a few aspects. Their paper studies factors individually, and their framework cannot handle the factor zoo. Instead, the framework in my paper is more suitable for the high-dimensional case. Also, they do not explore whether high- or low-frequency factors can explain the cross-section of average returns. Unlike their paper, I show that low-frequency latent factors are salient for cross-sectional asset pricing. The decomposition of the SDF into fast- and slow-moving components also improve our understanding of economic risks in the factor zoo.

The second branch of related literature is the abundant study of latent factor models after Ross (1976). Early empirical applications include Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), Connor and Korajczyk (1988). Kozak, Nagel, and Santosh (2020) use PCA to estimate latent factors of a large cross-section of characteristic-managed portfolios and then estimate their risk prices via an elastic-net algorithm. Kelly, Pruitt, and Su (2019) propose the instrumented PCA to model both pricing errors and factor loadings as functions of firm characteristics, and they find that four IPCA factors explain the cross-section of individual stock returns. Lettau and Pelger (2020a) and Lettau and Pelger (2020b) generalize PCA by including a penalty term on the pricing errors in expected returns. Their method can identify weak factors with high Sharpe ratios, even when the canonical PCA omits them. This paper differs from previous literature in that I estimate latent factors using frequency-dependent information in asset returns. As I show in Sections 1.2 and 1.3, the importance of latent factors can change across frequencies, and the frequency-dependent PCA can also strengthen a factor's signal if it is not independent. Giglio and Xiu (2021) and Giglio, Xiu, and Zhang (2021) show that we can project a nontradable factor into the space of the largest several PCs of single-period returns. The risk premium of a nontradable factor is the expected return of its mimicking portfolio composed of the largest several PCs of single-period returns.

Nevertheless, I do not intend to develop a method that can outperform all previous forms of PCA. Instead, I aim to provide a novel framework that is suitable for analyzing frequency-dependent risks in the factor zoo. Moreover, my frequency-dependent PCA can also be integrated with other PCA methods. For example, we can construct the factor-mimicking portfolio composed of frequency-specific PCs and use the three-pass procedure in Giglio and Xiu (2021) to estimate the risk premium of nontradable factors.

1.2   Methodology

Notation. $\mathbb{E}[\cdot]$, $\text{Var}[\cdot]$, and $\text{Cov}[\cdot]$ are the expectation, variance, and covariance operators. Suppose that $\boldsymbol{X_t}$ is an arbitrary $N \times 1$ vector of covariance-stationary random variables. $\boldsymbol{\mu_X}$,

$\mathbb{E}_{t-1}[\boldsymbol{X_t}]$, and $\bar{\boldsymbol{X}}_{\boldsymbol{t}}$ denote the unconditional, conditional, and sample mean of $\boldsymbol{X_t}$. $\boldsymbol{\Sigma_X}(h)$ is the autocovariance matrix with lag h: $\boldsymbol{\Sigma_X}(h) = \mathbb{E}\big[\boldsymbol{X_{t+h}X_t}\big] - \mathbb{E}[\boldsymbol{X_{t+h}}]\mathbb{E}[\boldsymbol{X_t}]^{\top}$. Particularly, $\boldsymbol{\Sigma_X}(0)$ is the unconditional covariance matrix, simply denoted by $\boldsymbol{\Sigma_X}$. $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}(h)$ is the sample estimate of $\boldsymbol{\Sigma_X}(h)$. $\text{Tr}\big[\boldsymbol{A}\big]$ is the trace of a matrix $\boldsymbol{A}$.

1.2.1   Asset Pricing Models

Suppose that there are N test assets, denoted by $\boldsymbol{R_t} = (R_{1t}, \ldots, R_{Nt})^{\top}$, and the sample size is T. This paper considers empirical applications in which both N and T are reasonably large, in particular, $\frac{N}{T} \to c < 1$. Motivated by the arbitrage pricing theory (APT) developed by Ross (1976), this paper studies an approximate factor pricing model, where the excess return on asset n, R_{nt}, is driven by a systematic component captured by K ($K < N$) latent factors and an idiosyncratic shock,

$$\underbrace{\boldsymbol{R_{t+1}}}_{N \times 1} = \underbrace{\boldsymbol{\alpha}}_{N \times 1} + \underbrace{\boldsymbol{\beta}}_{N \times K}\underbrace{\boldsymbol{F_{t+1}}}_{K \times 1} + \underbrace{\boldsymbol{e_{t+1}}}_{N \times 1}, \tag{1.1}$$

where $\boldsymbol{\alpha}$ denotes a vector of potential mispricings, $\boldsymbol{\beta F_{t+1}}$ is a vector of common components that are the product of factor loadings $\boldsymbol{\beta}$ and latent factors $\boldsymbol{F_{t+1}}$, and $\boldsymbol{e_{t+1}}$ is a vector of idiosyncratic shocks. I further require $\boldsymbol{\beta F_{t+1}}$ and $\boldsymbol{e_{t+1}}$ to be orthogonal. Empirically, I need to estimate the common component and cannot identify $\boldsymbol{\beta}$ and $\boldsymbol{F_{t+1}}$ separately.

Moreover, I require only systematic risks, proxied by $\boldsymbol{F_{t+1}}$, to enter the SDF. In other words, this paper assumes a strong form of APT, whereas unsystematic risks $\boldsymbol{e_{t+1}}$ earn zero risk premia. Specifically, $\mathcal{M}_{t+1}$ is linear in factors $\boldsymbol{F_{t+1}}$,[7]

$$\mathcal{M}_{t+1} = 1 - \boldsymbol{b}^{\top}(\boldsymbol{F_{t+1}} - \boldsymbol{\mu_F}), \tag{1.2}$$

where $\boldsymbol{\mu_F}$ is the unconditional expectation of latent factors, and $\boldsymbol{b}$ is the vector of risk prices for systematic factors, capturing the compensation for bearing systematic risks. According to the Hansen and Jagannathan (1991) (**HJ**) bound, if $\boldsymbol{F_{t+1}}$ are tradable factors, $\boldsymbol{b}^{\top}\boldsymbol{F_{t+1}}$ is the mean-variance efficient (MVE) portfolio. Therefore, constructing the linear SDF is the equivalent of finding the MVE portfolio in the cross-section of test assets.

According to the fundamental asset pricing equation,

$$\mathbb{E}[\mathcal{M}_{t+1}\boldsymbol{R_{t+1}}] = \mathbb{E}\big\{\boldsymbol{R_{t+1}}[1 - \boldsymbol{b}^{\top}(\boldsymbol{F_{t+1}} - \boldsymbol{\mu_F})]\big\} = \boldsymbol{0_N} \tag{1.3}$$

---

[7]I consider only excess returns in this paper, so the unconditional mean of $\mathcal{M}_{t+1}$ is unidentified. Without loss of generality, I normalize its mean to be one.

$$\implies \mathbb{E}[\boldsymbol{R_{t+1}}] = \text{Cov}(\boldsymbol{R_{t+1}}, \boldsymbol{F_{t+1}})\boldsymbol{b}, \tag{1.4}$$

so systematic risks, quantified by the covariance matrix, fully explain the cross-section of expected returns.

Past research documents the deviation of the independently and identically distributed (IID) assumption for asset returns. For instance, Chernov, Lochstoer, and Lundeby (2022) calculate the variance ratio of the mean-variance efficient portfolios in notable factor models. According to their results, factor returns are far from IID. In addition, Haddad, Kozak, and Santosh (2020) show that the first few principal components of asset returns are predictable by their own portfolio-level log book-to-market ratio. Motivated by their findings, this paper deviates from the IID assumption of $\boldsymbol{R_{t+1}}$ by assuming that latent factors subsume all the time-series dependency. Specifically, I assume that a $p \times 1$ vector of mean-zero "latent" state variables $\boldsymbol{X_t}$ can predict factors $\boldsymbol{F_{t+1}}$ as follows:

$$\underbrace{\boldsymbol{F_{t+1}}}_{K \times 1} = \underbrace{\boldsymbol{\mu_F}}_{K \times 1} + \underbrace{\boldsymbol{\Phi_X}}_{K \times p} \underbrace{\boldsymbol{X_t}}_{p \times 1} + \underbrace{\boldsymbol{f_{t+1}}}_{K \times 1}, \tag{1.5}$$

where $\boldsymbol{\mu_F}$ is the unconditional mean of latent factors, $\boldsymbol{\Phi_X X_t}$ captures the time-varying conditional mean of latent factors, $\boldsymbol{f_{t+1}}$ is conditionally uncorrelated: $\mathbb{E}[\boldsymbol{f_{t+1}}] = \mathbb{E}_t[\boldsymbol{f_{t+1}}] = \boldsymbol{0_K}$. Similarly, idiosyncratic shocks $\boldsymbol{e_{t+1}}$ are conditionally uncorrelated. Chamberlain and Rothschild (1983) also model idiosyncratic components as being cross-sectionally but not serially correlated. Since pricing errors are poorly predictable, such an assumption can be viewed as a good first-order approximation. I further plug equation (1.5) into the SDF,

$$\mathcal{M}_{t+1} = 1 - \boldsymbol{b}^\top \boldsymbol{f_{t+1}} - \boldsymbol{b_X}^\top \boldsymbol{X_t}, \tag{1.6}$$

where $\boldsymbol{b_X}^\top = \boldsymbol{b}^\top \boldsymbol{\Phi_X}$, and $\boldsymbol{b_X}^\top \boldsymbol{X_t}$ drives the conditional mean of the SDF, capturing its full conditional dynamics. The formula for $\mathcal{M}_{t+1}$ in equation (1.6) relates to previous studies that decompose the SDF into permanent and transitory components (see Alvarez and Jermann (2005) and Hansen and Scheinkman (2009)).

In addition, Hansen, Heaton, and Li (2008) study parametric models of state variables, modelling them using a stationary vector autoregressive (VAR) model. Instead, this paper is agnostic about the state vector $\boldsymbol{X_t}$. $\boldsymbol{X_t}$ can be firm characteristics and macro indicators, such as book-to-market ratio and cay (see Lettau and Ludvigson (2001a)). By decomposing asset returns into frequency-dependent components, this paper can infer whether state variables critical in pricing the cross-section, on average, are more important at high or low frequencies. The next subsection introduces the Fourier transform as the non-parametric solution.

1.2.2 Frequency Domain Analysis

This paper uses the techniques in frequency domain analysis to model the time-series dependence of asset returns and decompose an empirical series into its repetitive or regular components. I start by motivating why the Fourier transform is a natural approach to study long-horizon asset returns. Suppose that an excess return process x_t follows an AR(1) process, $x_{t+1} = \rho_x x_t + \sqrt{1 - \rho_x^2}\sigma_x\eta_{x,t+1}$, where ρ_x is the AR(1) coefficient, σ_x^2 is the unconditional variance, and $\eta_{x,t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. When ρ_x is zero (negative, positive), the asset return follows an IID (fast-moving, slow-moving) process. When ρ_x is more positive, the asset return tends to be more persistent.

Figure 1.A.1 plots the cumulative returns in a 24-month rolling window for three AR(1) processes: $\rho_x \in \{-0.5, 0, 0.5\}$. No matter how persistent the time series is, its long-horizon return always exhibits a cyclical pattern. Hence, it is natural to project the long-horizon return on the sine and cosine functions: $x_{t,t+24} = a_0 + a_1\sin(\frac{2\pi t}{48}) + a_2\cos(\frac{2\pi t}{48}) + e_{t,t+24}$. Note that the deterministic processes $\sin(\frac{2\pi t}{48})$ and $\cos(\frac{2\pi t}{48})$ complete a cycle in 48 months, or equivalently, it has a cycle length of 48 months. Motivated by this observation, I can study the cyclical pattern of long-horizon asset returns by projecting them on the space of sine and cosine functions, and an M-month cumulative return corresponds to a cycle length of $2M$. The frequency-domain analysis is the natural solution. Technically speaking, the Fourier transform decomposes a time series into orthogonal components at different frequencies. In the language of regression, it regresses the original time series into a sequence of sines and cosines functions.[8]

This paper uses ω to denote the frequency of a time-series process, which quantifies the number of cycles that this process completes per unit of time. Of equal interest is the period (or cycle length) of a time series, defined as the number of time points in a cycle: $\tau = \frac{1}{\omega}$. For instance, if ω is 0.1 in monthly data, the time series will finish 0.1 cycles in a month. Equivalently, it will take this process 10 months to complete one cycle.[9]

The spectral density matrix of $\boldsymbol{R_t}$ is defined as the Fourier transform of its auto-covariance matrices,

$$\boldsymbol{f_R}(\omega) = \sum_{h=-\infty}^{\infty} \boldsymbol{\Sigma_R}(h)\exp\{-2\pi ih\omega\}.$$

---

[8]According to the spectral representation theorem in Hannan (2009), a covariance-stationary time series can be approximated by a sum of sine and cosines random variables with different variances across frequencies. (see appendix 1.A.1.1).

[9]In addition, the absolute value of ω is no larger than 0.5 since any time series spends at least two months completing a cycle.

Through inverse Fourier transform, I can reverse engineer the auto-covariance matrix,

$$\boldsymbol{\Sigma_R}(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{2\pi i h \omega\} \boldsymbol{f_R}(\omega) d\omega.$$

In the study of asset pricing models, such as finding the tangency portfolio, investors focus on the covariance matrix of asset returns, that is, $h = 0$,

$$\mathrm{Cov}(\boldsymbol{R_t}) := \boldsymbol{\Sigma_R} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{R}(\boldsymbol{f_R}(\omega)) d\omega. \tag{1.7}$$

The Fourier transform of $\boldsymbol{R_t}$ decomposes asset returns as an equally weighted average of orthogonal components at different frequencies, so the covariance matrix of $\boldsymbol{R_t}$ equals the integral of the covariance matrix of its frequency-ω component, $\boldsymbol{f_R}(\omega)$, as in equation (1.7). In Appendix A1, I further show that only the real part of the spectral density matrix plays a role in estimating PCs. Hence, I focus on $\mathcal{R}(\boldsymbol{f_R}(\omega))$, the real part of the spectral density. Equation (1.7) also implies that $\mathcal{R}(\boldsymbol{f_R}(\omega))$ is the contribution to the covariance matrix from the frequency-ω component. If test asset returns are IID, the spectral density matrix of asset returns is constant across frequencies; that is $\boldsymbol{f_R}(\omega) = \boldsymbol{\Sigma_R}$ for every ω.

Why should we study the frequency-specific covariance matrices of asset returns? One reason is that the single-period covariance matrix often fails to capture systematic risks critical in explaining risk premia. For example, Brennan and Zhang (2020) show that yearly CAPM beta, which equals the covariance between annualized asset returns and the market portfolio, can explain the cross-section of 25 Fama-French size-B/M monthly portfolios. In contrast, the monthly CAPM beta entirely fails. Therefore, the single-period covariance, like the single-period CAPM beta, possibly misspecifies actual systematic risks, rendering the estimation of risk prices difficult or even impossible. This observation calls for the study of frequency-dependent systematic risks.

I estimate the spectral density matrix via discrete Fourier transform (DFT).[10] A simple example is in Figure 1.A.3, where a deterministic time series x_t in panel (c) consists of two components. The first component in panel (a) is slow-moving, with a frequency equal to 0.05, which completes a cycle every 20 periods. Another component of x_t in panel (b) is fast-moving, spending only two periods repeating a cycle. As in panel (d), DFT decomposes the variance of x_t into two parts contributed by low-frequency and high-frequency fluctuations.

Like other non-parametric estimation methods, the DFT estimate of the spectral density matrix at a particular frequency is susceptible to significant uncertainties. To reduce the variance, I divide the frequency intervals into three groups and estimate the spectral density

---

[10]Details about DFT can be found in Appendix 1.A.1.2.

matrix in each frequency interval.

What are the ideal cutoff points of the entire frequency interval? Past research can give some hints on this question. Dew-Becker and Giglio (2016) derive a closed-form solution to frequency-specific risk prices of parametric CCAPMs. They observe that only the long-run risk with cycles more prolonged than the business cycle is priced in the cross-section. In addition, Bandi, Chaudhuri, Lo, and Tamoni (2021) use the orthogonal Wold decomposition of CAPM's beta, and they find that only the business-cycle component of CAPM beta in the frequency interval between 32 and 64 months is priced. These observations motivate the following division of frequency intervals.

More specifically, I consider the following divisions of frequency intervals: (1) $\tau = \frac{1}{\omega} < 36$ months (high-frequency, denoted as HF), (2) $\tau = \frac{1}{\omega} \in [36, 120]$ months (low-frequency, or business cycle frequency, denoted as LF), and (3) $\tau = \frac{1}{\omega} > 120$ months (Above-LF, or A-LF). This paper considers the second frequency interval as being closely related to business cycles. Generally, the covariance matrix in the third group, with a cycle length greater than 120 months, is difficult to estimate non-parametrically. In later analysis, I also divide the HF interval into several sub-intervals. In addition, I consider alternative LF intervals in the robustness check. With the above division, I decompose the covariance matrix of $\boldsymbol{R_t}$,

$$\boldsymbol{\Sigma_R} = \int_{\omega \in \Omega_{HF}} \mathcal{R}(\boldsymbol{f_R}(\omega))d\omega + \int_{\omega \in \Omega_{LF}} \mathcal{R}(\boldsymbol{f_R}(\omega))d\omega + \int_{\omega \in \Omega_{A-LF}} \mathcal{R}(\boldsymbol{f_R}(\omega))d\omega \qquad (1.8)$$

$$= |\Omega_{HF}|\boldsymbol{\Sigma_R^{HF}} + |\Omega_{LF}|\boldsymbol{\Sigma_R^{LF}} + |\Omega_{A-LF}|\boldsymbol{\Sigma_R^{A-LF}}, \qquad (1.9)$$

where Ω_{HF}, Ω_{LF}, and Ω_{A-LF} denote the set of HF, LF, and Above-LF, with lengths $|\Omega_{HF}|$, $|\Omega_{LF}|$ and $|\Omega_{A-LF}|$ ($|\Omega_{HF}| + |\Omega_{LF}| + |\Omega_{A-LF}| = 1$).

Proposition 1.1 (Decomposition of asset returns' spectral density matrix) *I assume that $\boldsymbol{e_{t+1}}$ and $\boldsymbol{f_{t+1}}$ are conditional uncorrelated, and they are orthogonal. Then the spectral density matrices of $\boldsymbol{e_{t+1}}$ and $\boldsymbol{f_{t+1}}$ are constant across frequencies and equal to their unconditional covariance matrices $\boldsymbol{\Sigma_e}$ and $\boldsymbol{\Sigma_f}$, respectively. Moreover, I can decompose the spectral density matrix of $\boldsymbol{R_t}$ as,*

$$\boldsymbol{f_R}(\omega) = \boldsymbol{\beta} \boldsymbol{f_F}(\omega)\boldsymbol{\beta}^\top + \boldsymbol{\Sigma_e} = \boldsymbol{\beta}\boldsymbol{\Sigma_f}\boldsymbol{\beta}^\top + \boldsymbol{\Sigma_e} + \boldsymbol{\beta_X}\boldsymbol{f_X}(\omega)\boldsymbol{\beta_X^\top}, \qquad (1.10)$$

where $\boldsymbol{f_F}(\omega)$ and $\boldsymbol{f_X}(\omega)$ are the spectral density matrices of latent systematic factors and state variables, and $\boldsymbol{\beta_X} = \boldsymbol{\beta}\boldsymbol{\Phi_X}$.

A simple derivation of proposition 1.1 is in Appendix 1.A.2.1. A key observation in equation (1.10) is that only the last component related to state variables is frequency-dependent,

as e_{t+1} and f_{t+1} are conditionally uncorrelated. Furthermore, if I estimate latent factors of asset returns at different frequencies, I can study the difference between HF and LF systematic risks. More precisely, equation (1.10) indicates that the dynamics of state variables entirely drive the difference between HF and LF systematic risks. Similarly, I also decompose the SDF into frequency-dependent components and illustrate how the maximal Sharpe ratio implied by the SDF varies across frequencies.

Proposition 1.2 (Spectral density function of the SDF) *I normalize latent state variables $\boldsymbol{X_t}$ such that they are uncorrelated. Define risk prices of $\boldsymbol{X_t}$ as $\boldsymbol{b_X} = \boldsymbol{\Phi_X^\top b}$: $\boldsymbol{b_X} = (b_{X,1}, \ldots, b_{X,p})^\top$. Then the unconditional variance of the SDF is*

$$\mathrm{Var}(\mathcal{M}_{t+1}) = \boldsymbol{b}^\top \boldsymbol{\Sigma_f} \boldsymbol{b} + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{j=1}^{p} b_{X,i}^2 f_{X_i}(\omega) d\omega, \tag{1.11}$$

and the spectral density function of $\mathcal{M}_{t+1}$ is

$$f_{\mathcal{M}}(\omega) = \boldsymbol{b}^\top \boldsymbol{\Sigma_f} \boldsymbol{b} + \sum_{j=1}^{p} b_{X,i}^2 f_{X_i}(\omega), \tag{1.12}$$

where $f_{X_i}(\omega)$ is the spectral density function of the i-th state variable X_i.

I derive proposition 1.2 in Appendix 1.A.2.2. Since state variables are latent, I can always normalize them such that they are uncorrelated. An alternative interpretation of the normalization of $\boldsymbol{X_t}$ is that latent state variables are PCs of conditional expectations of factors. The above derivation shows that the maximal Sharpe ratio of the economy is frequency dependent. Moreover, the spectral density function of $\mathcal{M}_{t+1}$, denoted by $f_{\mathcal{M}}(\omega)$, varies across frequencies only due to the second term $\sum_{j=1}^{p} b_{X,i}^2 \boldsymbol{f_{X_i}}(\omega)$. I interpret this quantity as the weighted-average spectral density function of latent state variables, with weights proportional to the squared risk prices of state variables. If, on average, $\sum_{j=1}^{p} b_{X,i}^2 \boldsymbol{f_{X_i}}(\omega)$ is larger at high (low) frequencies, it implies that high (low) frequency information is more prominent in this cross-section of asset returns.[11] In addition, the spectral decomposition of $\mathcal{M}_{t+1}$ can only identify the state variable with a non-zero price of risk $b_{X,i}$.

Similar to this paper, Neuhierl and Varneskov (2021) use the Fourier transform to study the dynamics of state variables driving asset returns. They also show how to map their SDF into some canonical consumption-based asset pricing models. While the spectral density function of the SDF in IID CCAPM is constant across frequencies, other candidate models

---

[11]The prominence of a state variable X_t comes from two sources: the variance of X_t and its risk price squared (b_X^2). Since I use PCs as factors and state variables are latent, I can identify only $\mathrm{Var}(X_t)b_X^2$ rather than $\mathrm{Var}(X_t)$ and b_X^2 individually.

such as the long-run risk model in Bansal and Yaron (2004) have persistent SDFs. From the theoretical point of view, the LF component is more critical than the HF one in the SDF. A limitation of Neuhierl and Varneskov (2021) is that they consider only a single factor and explore how its risk premium varies across frequency. Differently, this paper aims to handle a factor zoo and extract frequency-dependent systematic risks in the large cross-section using the techniques introduced in the following subsections.

Remark 1.3 (Interpretation of the Frequency) *The frequency is different from the turnover of a factor strategy. Let us consider two value strategies: the first is the monthly rebalanced HML, or HML devil, from the AQR library. The second strategy is the yearly rebalanced HML from Ken French library. Even though these two strategies have different turnovers, their correlation is as high as 0.9. Furthermore, I compute their autocorrelations and variance ratios — these two HML strategies have pretty similar patterns.*

1.2.3 Estimation of Risk Factors

Under the assumption that $\boldsymbol{F_{t+1}}$ and $\boldsymbol{e_{t+1}}$ are orthogonal, I can represent the covariance matrix of asset returns as following:

$$\boldsymbol{\Sigma_R} = \boldsymbol{\beta}\,\boldsymbol{\Sigma_F}\,\boldsymbol{\beta}^\top + \boldsymbol{\Sigma_e}. \tag{1.13}$$

Systematic factors $\boldsymbol{F_{t+1}}$ are not directly observable, so I aim to estimate tradable proxies for them. A common way to estimate model (1.1) is the Principal Component Analysis (PCA), which relies on the eigendecomposition of $\boldsymbol{\Sigma_R}$,

$$\boldsymbol{\Sigma_R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top, \text{ with } \boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_N\}, \tag{1.14}$$

where $\boldsymbol{Q}$ is the matrix of eigenvectors ($\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I_N}$), and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues in a descending order. A common practice of PCA is to estimate $\boldsymbol{\beta}$ as the first K columns of $\boldsymbol{Q}$, denoted by $\boldsymbol{Q_K}$. Moreover, the estimates of K principal components (PC) are $\hat{\boldsymbol{F}}_t = \boldsymbol{Q_K}^\top\boldsymbol{R_t}$, uncorrelated by construction.

This paper uses the normalization $\boldsymbol{\beta}^\top\boldsymbol{\beta} = \boldsymbol{I_K}$ and $\boldsymbol{\Sigma_F} = \text{diag}\{\sigma_{F,1}^2, \ldots, \sigma_{F,K}^2\}$ in all following analyses, with exceptions unless stated. Furthermore, I allow for a mixture of strong and weak factors. In fact, most of asset pricing studies, as in this paper, use diversified portfolios as test assets, and the number of test assets is not truly infinite. In this paper, I differentiate strong and weak factors based on their variances.

Bai (2003) proves the asymptotic consistency of PCA when all factors are strong factors, which affect an increasing number of test asset returns as N goes to infinity. Mathematically,

$\frac{\sum_{t=1}^{T} \boldsymbol{F_t F_t^\top}}{T} \to \boldsymbol{\Sigma_F}$ and $\frac{\boldsymbol{\beta^\top \beta}}{N} \to \boldsymbol{\Sigma_N}$, where both $\boldsymbol{\Sigma_F}$ and $\boldsymbol{\Sigma_N}$ are positive definite matrices. Bai and Ng (2002) make the same assumption and propose an asymptotically consistent algorithm to determine the number of latent factors in model (1.1). Under the normalization chosen by this paper, the above assumption is the equivalent of explosive eigenvalues of $\boldsymbol{\Sigma_F}$. That is to say, the largest K eigenvalues of $\boldsymbol{\Sigma_R}$ and $\boldsymbol{\Sigma_F}$ will go to infinity as $N \to \infty$.

In addition, Bai (2003), like other papers, allows the idiosyncratic terms $\boldsymbol{e_t}$ to be only weakly correlated both cross-sectionally and over time. Furthermore, idiosyncratic shocks explain a finite amount of time-series variations in asset returns; that is

$$\lim_{N,T\to\infty} \sup \max \mathrm{eval}\big(\boldsymbol{\Sigma_e}\big) < \infty,$$

where $\max \mathrm{eval}(\boldsymbol{A})$ denotes the maximal eigenvalue of matrix $\boldsymbol{A}$.

In empirical applications, however, it is uncommon to come across the ideal case in which we can clearly separate large eigenvalues related to latent factors from small eigenvalues representing idiosyncratic shocks. A few papers have documented that PCA cannot consistently estimate model (1.1) when some latent factors are weak (see Onatski (2012), Lettau and Pelger (2020a)). Contrary to a strong factor, a weak one explains a relatively smaller fraction of time-series variations in asset returns. Alternatively, we can interpret a weak factor as having a finite variance or a relatively small variance in a finite sample.

Some weak factors are necessary to explain the cross-section of asset returns. For example, Lettau and Pelger (2020b) show that the omission of weak factors with a high Sharpe ratio can deteriorate the performance of latent factor models. However, the question is, when will PCA ignore the weak factors?

Benaych-Georges and Nadakuditi (2011) shed light on this question. Suppose that the covariance matrix of asset returns can be decomposed into the sum of two matrices as in equation (1.13), and one of them, such as $\boldsymbol{\Sigma_e}$, has bounded eigenvalues. Under this setup, the k-th ($k \leq K$) eigenvalue of $\boldsymbol{\Sigma_R}$, representing the k-th systematic factor F_{kt}, is identified if the k-th eigenvalue of $\boldsymbol{\beta} \, \boldsymbol{\Sigma_F} \, \boldsymbol{\beta^\top}$, equal to $\sigma_{F,k}^2$ under the normalization, is greater than a certain threshold; that is

$$\lambda_k\big(\boldsymbol{\beta} \, \boldsymbol{\Sigma_F} \, \boldsymbol{\beta^\top}\big) = \sigma_{F,k}^2 > \lambda_{crit},$$

where $\lambda_k(\boldsymbol{A})$ denotes the k-th largest eigenvalue of matrix $\boldsymbol{A}$, and λ_{crit} is related to the limit of $\frac{N}{T}$ and the upper bound of eigenvalues for $\boldsymbol{\Sigma_e}$. Otherwise, a phenomenon called eigenvalue phase transition occurs, and the factor k is no longer identified. Now let us look at a simple example.

Example 1.4 (Single-factor model) *Suppose that there is only one systematic factor in model (1.1), and the idiosyncratic vector $\boldsymbol{e_t}$ has a covariance matrix $\sigma^2 \boldsymbol{I_N}$ ($\sigma^2 < \infty$). I*

further normalize the factor loadings such that $\boldsymbol{\beta}^\top \boldsymbol{\beta} = 1$, and the variance of $\boldsymbol{F_t}$ is σ_F^2 ($\sigma_F^2 < \infty$). As $\frac{N}{T} \to c < 1$, the distribution of eigenvalues for $\mathrm{Var}(\boldsymbol{e_t})$ converges to the Marchenko–Pastur distribution, with lower and upper bounds $\sigma^2(1 - \sqrt{c})^2$ and $\sigma^2(1 + \sqrt{c})^2$.

According to corollary 2 in Lettau and Pelger (2020a), when $\sigma_F^2 \leq \sqrt{c}\sigma^2$, the top eigenvalue converges to $\sigma^2(1 + \sqrt{c})^2$. Consequently, PCA can no longer identify $\boldsymbol{F_t}$, and the correlation between true factor $\boldsymbol{F_t}$ and the PCA estimate $\hat{\boldsymbol{F}}_t$ converges to zero.

A strong factor has a variance that is much more considerable than the critical point at all frequencies, so it is always identifiable. However, there are some "marginal factors" whose signals are strong enough only at high or low frequencies. It depends on the dynamics of state variables driving this factor. Suppose that a weak factor in example 1.4 has a variance slightly less than the critical value $\sqrt{c}\sigma^2$, but it follows an AR(1) process: $F_t = \rho_F F_{t-1} + \sqrt{1 - \rho_F^2}e_{F,t}$, $e_{F,t} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_F^2)$. If ρ_F is more positive (negative), F_t is more slow-moving (fast-moving). The spectral density of F_t is in Figure 1.A.2. For instance, when ρ_F is 0.5, the variance of F_t at low frequencies is roughly four times the unconditional variance. It is possible that a weak factor is unidentified by canonical PCA but stands out at low frequencies if its signal is persistent and strong enough in the long horizon. This observation also motivates the frequency-dependent PCA.

Definition 1.1 (Frequency-dependent PCA) *Suppose that $\boldsymbol{\Sigma_R^{HF}}$, $\boldsymbol{\Sigma_R^{LF}}$, and $\boldsymbol{\Sigma_R^{A\text{-}LF}}$ are high-, low-, and above-low-frequency covariance matrices of the N-dimensional random vector $\boldsymbol{R_t}$. The eigendecomposition of $\boldsymbol{\Sigma_R^Z}$, $Z \in \{HF, LF, A\text{-}LF\}$, is*

$$\boldsymbol{\Sigma_R^Z} = \boldsymbol{Q^Z}\boldsymbol{\Lambda^Z}(\boldsymbol{Q^Z})^\top, \ \text{with } \boldsymbol{\Lambda^Z} = diag\{\lambda_1^Z, \ldots, \lambda_N^Z\},$$

where $\boldsymbol{Q^Z}$ are eigenvectors of $\boldsymbol{\Sigma_R^Z}$, that is, $(\boldsymbol{Q^Z})^\top \boldsymbol{Q^Z} = \boldsymbol{I_N}$, and $\boldsymbol{\Lambda^Z}$ is the diagonal matrix of eigenvalues in descending order. Define the latent factors in the frequency $Z \in \{HF, LF, A\text{-}LF\}$ as $\boldsymbol{F_t^Z} = (\boldsymbol{Q^Z})^\top \boldsymbol{R_t}$.

Intuitively, I rotate the space of canonical PCs to target the short-term and long-term common variations in asset returns. In other words, frequency-dependent PCA aims to select monthly proxies for short-term and long-term systematic risks. A special case is when asset returns are IID. Since $\boldsymbol{\Sigma_R^Z}$ are identical in this case, PCA, HF-PCA, and LF-PCA will deliver the exact estimates across frequencies.

1.2.4 Estimation of Risk Prices

This paper always uses principal components of asset returns as systematic factors, $\boldsymbol{F_t} = (\boldsymbol{Q^Z})^\top \boldsymbol{R_t}$, $Z \in \{\text{HF}, \text{LF}, \text{A-LF}\}$. Since $\boldsymbol{F_t}$ are always tradable, estimating the linear SDF in

equation (1.2) is the same as finding the optimal portfolio weights, $\boldsymbol{b}$, such that $\boldsymbol{b}^\top \boldsymbol{F_t}$ is the MVE portfolio. If the SDF prices the cross-section of asset returns, it also prices all tradable factors, so I can rewrite equation (1.4) as follows:

$$\boldsymbol{\mu_F} = \boldsymbol{\Sigma_F} \boldsymbol{b}. \tag{1.15}$$

I can solve the risk prices from equation (1.15), $\boldsymbol{b} = \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\mu_F}$. Therefore, risk prices $\boldsymbol{b}$ are proportional to the MVE portfolio weights. In a finite sample, I need to estimate both $\boldsymbol{\Sigma_F}$ and $\boldsymbol{\mu_F}$. Past research have shown that a naive estimator such as $\hat{\boldsymbol{b}} = \hat{\boldsymbol{\Sigma}}_F^{-1} \bar{\boldsymbol{F}}_t$ does not perform well in real datasets. For example, Tu and Zhou (2011) show that the estimated Markowitz (1952) portfolio not only underperforms the naive 1/N rule, in which investors invest equally across N assets, but also earns negative risk-adjusted returns. Kozak, Nagel, and Santosh (2020) argue that the majority of uncertainty comes from the estimation of factor means $\boldsymbol{\mu_F}$ and propose a simple Bayesian procedure to estimate $\boldsymbol{b}$.

To compare with Kozak, Nagel, and Santosh (2020), this paper adopts a similar strategy, which assumes that the covariance matrix of factor returns, $\boldsymbol{\Sigma_F}$, is known and focuses on the modelling of mean factor returns. Furthermore, equation (1.15) does not hold exactly in finite sample, so I include pricing errors $\boldsymbol{\alpha}$ on the right-hand side of equation (1.15): $\boldsymbol{\mu_F} = \boldsymbol{\Sigma_F} \boldsymbol{b} + \boldsymbol{\alpha}, \quad \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0_N}, \frac{1}{T}\boldsymbol{\Sigma_F})$. Finally, I assign a normal prior for risk prices: $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0_K}, \frac{\kappa^2}{\tau}\boldsymbol{I_K})$, $\tau = \mathrm{Tr}\left[\boldsymbol{\Sigma_F}\right]$. Under such a prior distribution, the prior expectation on the squared Sharpe ratio of factor returns is equal to

$$\mathbb{E}_{prior}[SR_F^2] = \mathbb{E}_{prior}[\boldsymbol{b}^\top \boldsymbol{\Sigma_F} \boldsymbol{b}] = \frac{\kappa^2}{\tau} \mathrm{Tr}\left[\boldsymbol{\Sigma_F}\right] = \kappa^2.$$

Also, the posterior distribution of $\boldsymbol{b}$, conditional on $(\boldsymbol{\mu_F}, \boldsymbol{\Sigma_F})$, is

$$p(\boldsymbol{b} \mid \boldsymbol{\mu_F}, \boldsymbol{\Sigma_F}) \propto \exp\left\{ -\frac{T}{2}\left[(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F}\boldsymbol{b})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F}\boldsymbol{b}) + \frac{\tau}{\kappa^2 T}\boldsymbol{b}^\top \boldsymbol{b} \right] \right\}.$$

Therefore, the posterior mode of $\boldsymbol{b}$ is the solution to the below objective function I,

$$\min_{\boldsymbol{b}} \left\{ (\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F}\boldsymbol{b})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F}\boldsymbol{b}) + v_2 \boldsymbol{b}^\top \boldsymbol{b} \right\}, \tag{1.16}$$

where $v_2 = \frac{\mathrm{Tr}\left[\boldsymbol{\Sigma_F}\right]}{\mathbb{E}_{prior}[SR_F^2] \times T}$. A detailed derivation of equation (1.16) is in Appendix 1.A.2.3. For simplicity, I will denote $\sqrt{\mathbb{E}_{prior}[SR_F^2]}$ as SR_{prior}, or simply call it the prior Sharpe ratio. However, due to Jensen's inequality, $\mathbb{E}_{prior}[SR_F] \leq \sqrt{\mathbb{E}_{prior}[SR_F^2]}$, so SR_{prior} is an upper bound on the expected Sharpe ratio under prior distribution. Objective function I is to

minimize squared Sharpe ratio of pricing errors (or equivalently maximize R^2_{GLS}), subject to L_2-penalty. In addition, I include factors into the model based on their ability of explaining time-series variations. That is to say, I include the first K largest PCs into analysis when I consider a $K-$factor model.

Kozak, Nagel, and Santosh (2020) extends equation (1.16) by including the L_1-penalty,

$$\min_{\boldsymbol{b}}\left\{(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b}) + 2v_1|\boldsymbol{b}|_1 + v_2\boldsymbol{b}^\top\boldsymbol{b}\right\}. \tag{1.17}$$

Since the principal components are uncorrelated by constructions, its covariance matrix $\boldsymbol{\Sigma_F}$ is diagonal with elements equal to eigenvalues of the covariance matrix of test assets. The closed-form solution to optimization problem (1.17) is

$$\hat{\lambda}_{i,KNS} = \begin{cases} \frac{\mu_{F,i}-v_1}{\sigma^2_{F,i}+v_2}, & \text{if } \mu_{F,i} \geq v_1, \\ 0, & \text{if } \mu_{F,i} < v_1, \end{cases} \tag{1.18}$$

so the above algorithm selects a certain factor j whenever it has a mean greater than v_1. In other words, v_1 controls the sparsity of factor models. Moreover, factors with small variances are shrunk more heavily by the L_2-penalty. This makes sense as those factors are more likely to be idiosyncratic risks that should not command sizeable risk premia.

This paper shows the empirical results using both objective functions (1.16) and (1.17). Suppose systematic factors that explain a large amount of time-series variations can capture most of the risk premium. In that case, estimates by (1.16) or (1.17) should be largely similar. It also implies that we can find a sparse factor model consisting of large PCs of asset returns. Finally, I emphasize that I also assume there is only one true SDF. One of this paper's main objectives is to determine whether the SDF comprised of canonical, HF, or LF PCs is a better approximation to the tangency portfolio.

1.3 Empirics

I now proceed to the empirical studies. The first step is to estimate frequency-specific risk factors using the techniques introduced in Section 1.2 and to investigate whether the SDF composed of HF or LF factors is a better approximation to the tangency portfolio. Next, I explore whether some celebrated factor models proposed in the literature, such as Fama-French three factors, can explain these SDFs. Finally, I attempt to understand the economic fundamentals behinds SDFs. I begin this section with the analysis of 25 Fama-French portfolios to show how the factor structure of asset returns varies across frequencies.

I then carry out the main analysis in a large cross-section of portfolios, studying which frequency is salient for the cross-sectional asset pricing.

1.3.1 Sample and Data

My primary data source comes from the characteristic-managed portfolios in Kozak, Nagel, and Santosh (2020). The definition of firm characteristics and the data are on Serhiy Kozak's website. There are 51 firm characteristics in Kozak, Nagel, and Santosh (2020),[12] but I select 39 of them to ensure the sample size large enough to estimate the low-frequency covariance matrix of asset returns.

In the benchmark analysis, I split the sample from August 1963 to December 2019 into two halves and focus on the out-of-sample performance, which imposes additional challenges on the estimation due to the smaller subsamples. Firm characteristics can be further categorized into eight groups, as in Table 1.A.2. Kozak, Nagel, and Santosh (2020) also exclude stocks with market equity below 0.01% of the aggregate US market cap, alleviating the impact of microcaps. Each month, individual stocks are sorted into 10 portfolios based on each of the 39 firm characteristics. They construct portfolios with weights equal to cross-sectional ranks of a given stock's characteristic, which is centered and normalized by the sum of absolute values of all ranks in the cross-section.

I also use the Fama-French 25 size and book-to-market (total variance) portfolios in Section 1.3.2. I download the data from Ken French's library. In succeeding analysis, I supplement the main dataset with additional economic variables. Detailed variable definitions and data sources are provided in Table 1.A.1.

1.3.2 Starting Examples: 25 Fama-French Portfolios

To illustrate how the factor structures vary across frequencies, I start with the 25 Fama-French size and book-to-market portfolios. The numbers in Figure 1.1 are 25 portfolios' factor loadings, equivalently their portfolio weights. The sample spans from August 1963 to December 2019. In each graph, the x-axis shows five buckets of book-to-market ratio, whereas the y-axis plots five levels of firm size. For instance, ME1 represents small firms, and BM5 means high book-to-market portfolios. Since PC1 is always the level (identically market) factor, I will display only the second and third PCs. In addition, I note that PCs and HF-PCs are almost identical; therefore, I will focus on explaining the difference between HF-PCs and LF-PCs.

---

[12]I thank the authors for sharing the data on their website. A more specific description of how to construct this dataset can be found on Serhiy Kozak's website: https://sites.google.com/site/serhiykozak/data.
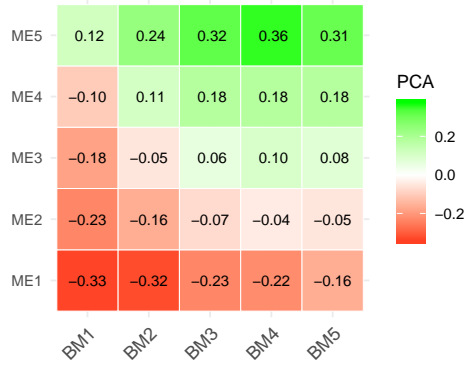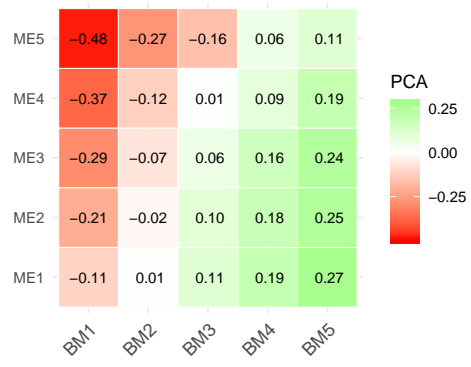
(a) 2nd PC of PCA

(b) 3rd PC of PCA
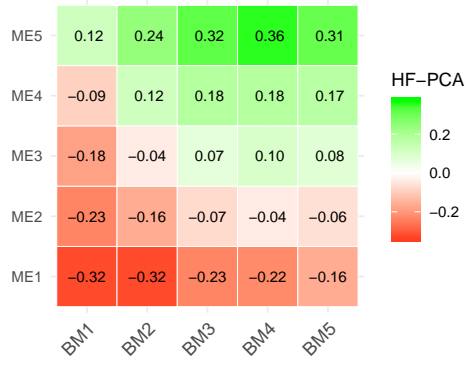
(c) 2nd PC of HF-PCA
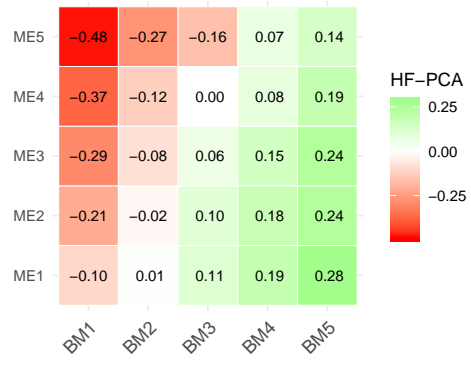
(d) 3rd PC of HF-PCA
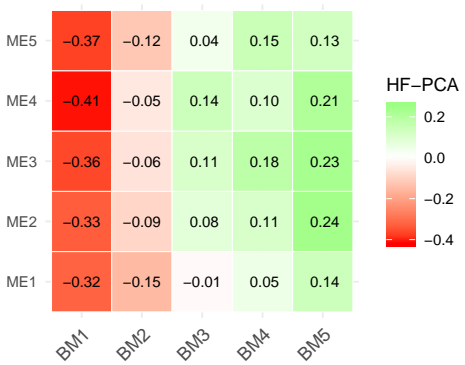
(e) 2nd PC of LF-PCA
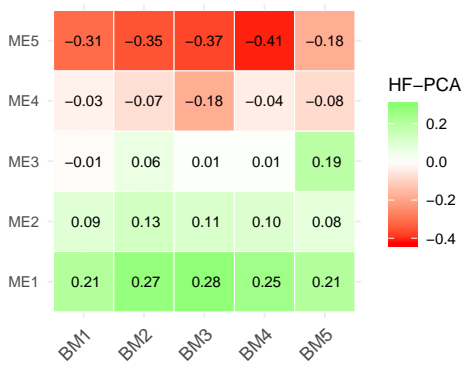
(f) 3rd PC of LF-PCA

Figure 1.1: 25 Fama-French Size-Value Portfolios: 2nd and 3rd PCs

This figure shows the factor loadings of the second and third principal components in the cross-section of 25 Fama-French size-value portfolios. I estimate the factor loadings using the canonical PCA, HF-PCA, and LF-PCA. See Definition 1.1 for the algorithm. The sample is from August 1963 to December 2019.

First of all, let us look at Panel (c): In each column, the second HF-PC positively loads on all large portfolios in ME5 but negatively on small portfolios in ME1. Therefore, HF-PC2 is a size factor. Next, in Panel (d), the portfolio weights of all five portfolios in BM5 (BM1) are always positive (negative), so HF-PC3 is a value factor. Overall, the size factor is more important than the value factor at high frequencies.

On the contrary, I observe the opposite at low frequencies. The heat-map in Panel (e) reveals that the second most crucial LF-PC is the value factor, while the size factor becomes the third-largest LF-PC, as is evident in Panel (f). This observation is largely compatible with the economic theory because the value factor often captures the business-cycle risk at low frequency. For example, Lettau and Ludvigson (2001b) point out that value stocks are more highly correlated with consumption growth in bad economic states than growth stocks, so they earn higher average returns. In short, the example in Figure 1.1 shows that the relative importance of latent factors can vary across frequencies.

1.3.3 Simple Simulation

In this section, I design a simple simulation to illustrate how the frequency-dependent PCA can recover the conditional information in asset returns. I assume that each asset return is driven by an IID systematic component (strong factor), a persistent state variable (weak factor), and an idiosyncratic element,

$$\boldsymbol{R_{t+1}} = \boldsymbol{\mu_R} + \boldsymbol{\beta_F} F_{t+1} + \boldsymbol{\beta_x} x_t + \boldsymbol{e_{t+1}}, \quad \boldsymbol{\beta_F^\top \beta_F} = 1, \quad \boldsymbol{\beta_x^\top \beta_x} = 1, \quad \boldsymbol{\beta_F^\top \beta_x} = 0. \quad (1.19)$$

$$F_{t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_F^2), \quad \boldsymbol{e_{t+1}} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0_N}, \sigma^2 \boldsymbol{I_N}), \quad F_{t+1} \perp x_t \perp \boldsymbol{e_{t+1}},$$

$$x_{t+1} = \rho_x \cdot x_t + \sqrt{1 - \rho_x^2} \, \sigma_x \, \eta_{x,t+1}, \quad \eta_{x,t+1} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1). \quad (1.20)$$

In the above model, only the state variable x_t can predict asset returns, and it follows an AR(1) process. Examples of x_t include (1) the time-varying mean and variance of consumption growth in the long-run risk model (Bansal and Yaron (2004)), (2) the systemic- and stock-specific resilience in the recovery following a disaster in the disaster model (Gabaix (2012)), (3) the surplus consumption ratio in Campbell and Cochrane (1999), and (4) portfolio-level book-to-market ratio (Haddad, Kozak, and Santosh (2020)). Idiosyncratic shocks are homogeneous and have an identical variance σ^2.

According to example 4, the conditional information $\boldsymbol{\beta_x} x_t$ is asymptotically unidentified if $\sigma_x^2 < \sqrt{\lim \frac{N}{T}} \sigma^2$. Factor returns are weakly predicted, so the assumption for a relatively small σ_x^2 is reasonable. Furthermore, F_{t+1} and x_t are priced in the cross-section, so they

enter the linear SDF,

$$\mathcal{M}_{t+1} = 1 - b_F \cdot F_{t+1} - b_x \cdot x_t, \tag{1.21}$$

where b_F and b_x are risk prices of F_{t+1} and x_t. The expected returns are determined by the fundamental asset pricing equation $\mathbf{E}[\mathcal{M}_{t+1}\boldsymbol{R_{t+1}}] = \mathbf{0_N}$, which implies

$$\boldsymbol{\mu_R} = -\text{Cov}(\mathcal{M}_{t+1}, \boldsymbol{R_{t+1}}) = b_F \boldsymbol{\beta_F}\sigma_F^2 + b_x \boldsymbol{\beta_x}\sigma_x^2. \tag{1.22}$$

F_{t+1} and x_t are latent, so I extract their tradable proxies. Specifically, I project them into the space of asset returns and find factor-mimicking portfolios with the highest Sharpe ratio:

$$F_{t+1} : \tilde{F}_{t+1} = \boldsymbol{\beta_F}^\top \boldsymbol{R_{t+1}}, \quad \mathbf{E}[\tilde{F}_{t+1}] = b_F\sigma_F^2, \quad \text{Var}(\tilde{F}_{t+1}) = \sigma_F^2 + \sigma^2, SR_F^2 = \frac{b_F^2\sigma_F^2}{1 + \frac{\sigma^2}{\sigma_F^2}},$$

$$x_t : \tilde{X}_{t+1} = \boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, \quad \mathbf{E}[\tilde{X}_{t+1}] = b_x\sigma_x^2, \quad \text{Var}(\tilde{X}_{t+1}) = \sigma_x^2 + \sigma^2, SR_x^2 = \frac{b_x^2\sigma_x^2}{1 + \frac{\sigma^2}{\sigma_x^2}},$$

where $\tilde{F}_{t+1}$ and $\tilde{X}_{t+1}$ are tradable proxies for F_{t+1} and x_t and are orthogonal by construction.

Regarding the simulation setup, I estimate the first PC in the cross-section of 78 test assets and assume that $\boldsymbol{\beta_F}$ is equal to its factor loadings and the volatility of F_{t+1} (σ_F) is 8.[13] Also, the idiosyncratic shocks have a unit variance ($\sigma^2 = 1$). In addition, the weak factor x_t has an identical factor loading ($\boldsymbol{\beta_x}$) as the second PC. However, its variance is small, $\sigma_x^2 = 0.5$; in other words, the weak factor explains a tiny fraction of time-series variations in single-period returns. In the cross-section of 78 test assets and 338 monthly observations, the canonical PCA has difficulty in identifying this factor: The critical value, $\sqrt{\lim \frac{N}{T}}\sigma^2 \approx 0.48$, is close to the variance of weak factor. According to past literature, state variables that can predict asset returns tend to be extremely persistent, such as those in the long-run risk model, so I set ρ to be 0.9. Finally, I choose the Sharpe ratio of $\tilde{F}_{t+1}$ and $\tilde{X}_{t+1}$ to be 0.25 and 0.30 per month, and their risk prices can be reverse-engineered: $b_F = 0.032$ and $b_x = 0.735$.

Suppose I simulate factors and asset returns using the above model setup. I estimate the latent factors using canonical PCA, HF-PCA and LF-PCA. Figure 1.2 is one such example. The blue solid lines are "true" tradable factors $\boldsymbol{\beta_F}^\top \boldsymbol{R_{t+1}}$ and $\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}$, and the red dotted lines show the estimates of factors. As is evident in Panels (a), (c), and (e), I can always identify the first latent factor. Equivalently, the first latent factor explains the largest fraction of time-series variations in both short-horizon and long-horizon asset returns.

---

[13]I assume a relatively large σ_F to make sure that it is a strong factor that is always identified by (HF- or LF- or canonical) PCA. The simulation results are robust to other $\sigma_F \in \{3, 5, 10\}$. In the data, the first three - five latent factors often have sizable eigenvalues (volatility) compared to the idiosyncratic volatility, so this assumption for the strong factor is reasonable.
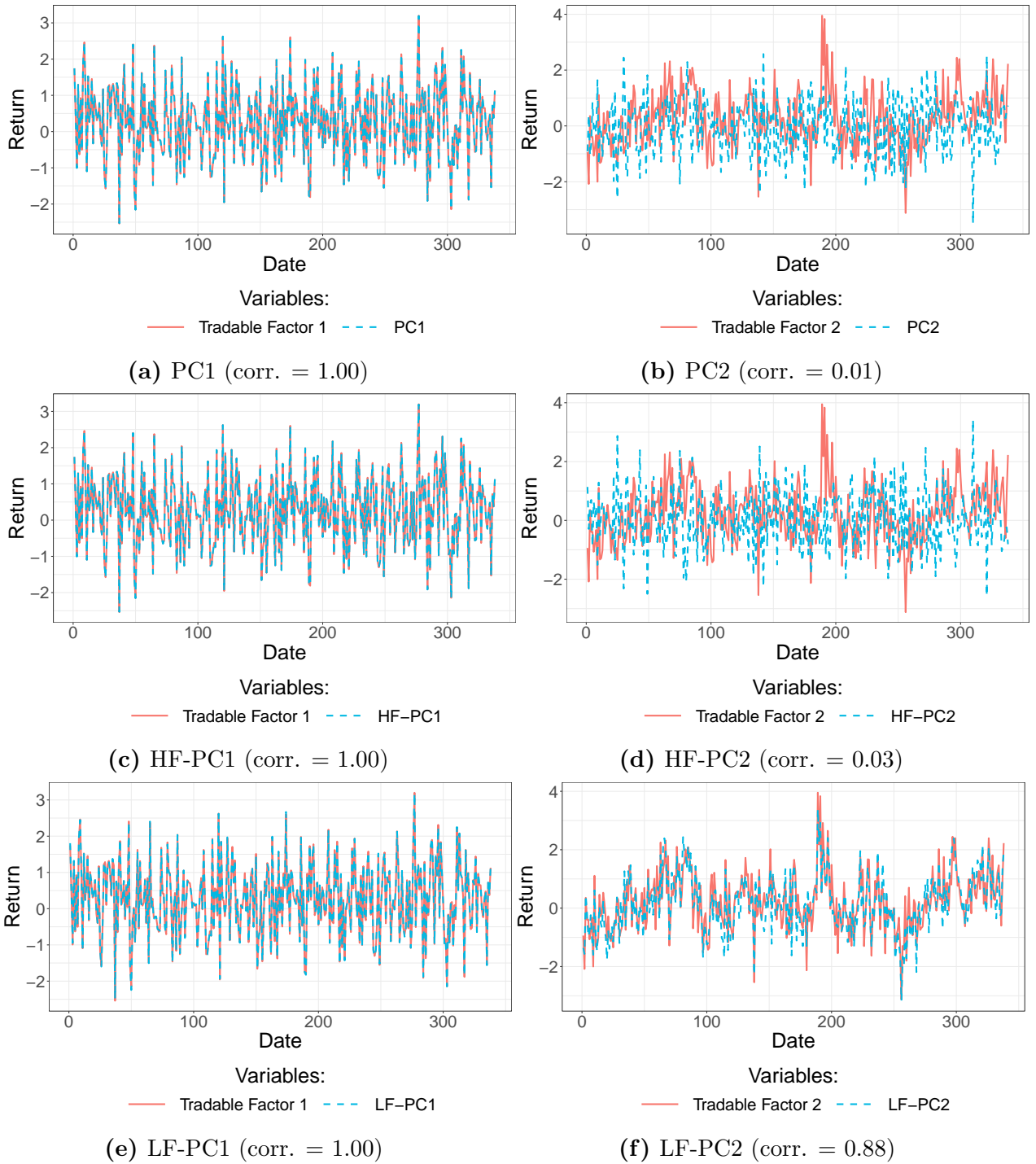
Figure 1.2: Starting example: first two latent factors from canonical, HF-, and LF-PCA

I simulate one sample path of systematic factors and asset returns using the model setup described in the main text. This graph plots the time series of the first two "true" tradable factors (blue solid lines) and their estimates (red dotted lines) from canonical PCA (Panels (a) and (b)), HF-PCA (Panels (c) and (d)), and LF-PCA (Panels (e) and (f)). In addition, corr. refers to the correlation between the true tradable factor and its estimate. I standardize all time series to have unit variance.

On the contrary, the weak factor is difficult to identify. Panels (b) and (d) show that neither canonical PCA and HF-PCA can recover the weak factor related to x_t, and the second PC or HF-PC has an almost zero correlation with the true factor-mimicking portfolio of x_t. The maximal Sharpe ratio implied by the first two PCs or HF-PCs is 0.276. However, the persistence of the state variables x_t magnifies its signal at low frequencies and allows me to detect it empirically. Panel (f) plots the second LF-PC, which closely tracks the "true" factor and has a correlation of 0.88. Moreover, the maximal Sharpe ratio implied by the first two LF-PCs is 0.321. Intuitively, as Figure 1.A.2 displays, the low-frequency variance of the persistent factor is much more considerable than its unconditional variance, so the signal of this factor passes the critical value $\sqrt{\lim \frac{N}{T} \sigma^2}$ at low frequencies and becomes identified.

Table 1.1 reports the simulation results of estimation using canonical PCA, HF-PCA, and LF-PCA in 1,000 simulations. The time-series sample size is 338. For each statistic, I show its 5th, 25th, 50th, 75th, 95th, mean, and mode. Panel (A) displays the correlation between the second true factor and estimated PC2 from canonical PCA, HF-PCA, and LF-PCA. Ideally, the correlation is 1. I focus on the second PC since the first PC is always identified, so there is no difference among different types of PCAs. The most important observation is that the LF-PC2 has a much more significant correlation with the true second factor. Specifically, the average correlation between LF-PC2 and the true factor 2 is 0.754, whereas they are only 0.176 and 0.434 for HF-PC2 and canonical PC2, respectively. Hence, studying the LF components asset returns recovers a huge part of the persistent state variable.

In panel (B), I construct the MVE portfolios consisting of the first two latent factors: $\mathbf{MVE}_t = \hat{\boldsymbol{\mu}}_{\boldsymbol{F}}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{-1} \boldsymbol{F_t}$, where $\boldsymbol{F_t}$ is either the first two PCs, HF-PCs, or LF-PCs. Since the LF-PCA recovers the persistent priced state variable, the LF-MVE portfolio has a greater Sharpe ratio than the other two portfolios. Panel (C) further reports the correlation among the HF-MVE, canonical-MVE, and LF-MVE portfolios. The MVE portfolios composed of the HF- and canonical PCs are highly correlated, with an average correlation of 0.937. However, the PCA can identify the state variable x_t in some simulations when the HF-PCA fails, so their MVE portfolios have correlations less than 0.746 in 5% of these simulations.

Finally, I decompose $\mathbf{MVE}_t^{HF}$ and $\mathbf{MVE}_t^{LF}$ in Panel (D),

$$\mathbf{MVE}_t^{LF} = \gamma^{HF} \mathbf{MVE}_t^{HF} + \mathbf{MVE}_t^{missing}, \quad \mathbf{MVE}_t^{HF} = \gamma^{HF} \mathbf{MVE}_t^{LF} + \mathbf{MVE}_t^{unpriced}.$$

I report the Sharpe ratio of $\mathbf{MVE}_t^{missing}$ and $\mathbf{MVE}_t^{unpriced}$ and also their correlation coefficients with the second factor. On average, the missing-MVE portfolio ($\mathbf{MVE}_t^{missing}$) has a correlation of 0.666 with the second true factor and yields a Sharpe ratio of 0.193. It implies that the HF-PCA misses important conditional information x_t. On the contrary, the unpriced MVE portfolio ($\mathbf{MVE}_t^{unpriced}$) has a negative correlation with the second true factor,

but its Sharpe ratio is almost zero. Hence, the LF-MVE portfolio can be decomposed into two components: The first component, which is linear in the HF-MVE portfolio, identifies the IID shock driving a large proportion of common variations in asset returns, and the second component is the missing-part, mainly reflecting the slow-moving state variable.

In short, the simulation results confirm that both canonical PCA and HF-PCA often fail to identify the weak factor. However, if the weak factor is slow-moving, its signal can soar at low frequencies so that the LF-PCA can identify it.

Table 1.1: Simulation Results

	5th	25th	50th	75th	95th	Mean	Mode
Panel (A). Correlation between 2nd true factor and its estimate							
$corr(\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, (\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^{\boldsymbol{PC}})^\top \boldsymbol{R_{t+1}})$	0.041	0.228	0.449	0.640	0.798	0.434	0.556
$corr(\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, (\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^{\boldsymbol{HF}})^\top \boldsymbol{R_{t+1}})$	0.012	0.069	0.154	0.257	0.431	0.176	0.053
$corr(\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, (\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^{\boldsymbol{LF}})^\top \boldsymbol{R_{t+1}})$	0.451	0.706	0.791	0.847	0.900	0.754	0.831
Panel (B). Sharpe ratio of MVE portfolios							
Sharpe ratio of $\mathbf{MVE}_t^{PC}$	0.179	0.232	0.272	0.313	0.374	0.274	0.264
Sharpe ratio of $\mathbf{MVE}_t^{HF}$	0.163	0.218	0.254	0.292	0.347	0.255	0.249
Sharpe ratio of $\mathbf{MVE}_t^{LF}$	0.210	0.279	0.325	0.378	0.473	0.330	0.332
Panel (C). Correlation between MVE portfolios							
$corr(\mathbf{MVE}_t^{PC}, \mathbf{MVE}_t^{HF})$	0.746	0.919	0.971	0.993	0.999	0.937	0.992
$corr(\mathbf{MVE}_t^{HF}, \mathbf{MVE}_t^{LF})$	0.497	0.665	0.789	0.899	0.979	0.770	0.917
$corr(\mathbf{MVE}_t^{PC}, \mathbf{MVE}_t^{LF})$	0.599	0.771	0.868	0.937	0.985	0.840	0.943
Panel (D). Difference between $\mathbf{MVE}_t^{HF}$ and $\mathbf{MVE}_t^{LF}$							
Sharpe ratio of $\mathbf{MVE}_t^{missing}$	0.035	0.114	0.189	0.260	0.363	0.193	0.199
Sharpe ratio of $\mathbf{MVE}_t^{unpriced}$	0.001	0.003	0.008	0.019	0.055	0.015	0.004
$corr(\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, \mathbf{MVE}_t^{missing})$	0.197	0.623	0.740	0.812	0.881	0.666	0.788
$corr(\boldsymbol{\beta_x}^\top \boldsymbol{R_{t+1}}, \mathbf{MVE}_t^{unpriced})$	-0.741	-0.627	-0.514	-0.392	-0.087	-0.475	-0.520

This table reports the simulation results in 1,000 simulations. I estimate the systematic factors using canonical PCA, HF-PCA, and LF-PCA. For each statistic, I show its 5th, 25th, 50th, 75th, 95th, mean, and mode. In Panel (A), I consider the correlation between the second true factor and estimated PC2 from canonical PCA, HF-PCA, and LF-PCA. Ideally, the correlation should be 1. In Panel (B), I construct the mean-variance efficient (MVE) portfolios consisting of the first two latent factors: $\mathbf{MVE}_t = \hat{\boldsymbol{\mu}}_{\boldsymbol{F}}^\top \hat{\boldsymbol{\Sigma}}_{\boldsymbol{F}}^{-1} \boldsymbol{F_t}$, where $\boldsymbol{F_t}$ is either the first two canonical PCs, HF-PCs, or LF-PCs. Panel (c) reports the correlation between $\mathbf{MVE}_t^{PC}$, $\mathbf{MVE}_t^{HF}$, and $\mathbf{MVE}_t^{LF}$. Next, I decompose $\mathbf{MVE}_t^{HF}$ and $\mathbf{MVE}_t^{LF}$ in Panel (D) as follows:

$$\mathbf{MVE}_t^{LF} = \gamma^{HF}\mathbf{MVE}_t^{HF} + \mathbf{MVE}_t^{missing},$$

$$\mathbf{MVE}_t^{HF} = \gamma^{HF}\mathbf{MVE}_t^{LF} + \mathbf{MVE}_t^{unpriced}.$$

Finally, I report the Sharpe ratio of $\mathbf{MVE}_t^{missing}$ and $\mathbf{MVE}_t^{unpriced}$ and their correlation coefficients with the second factor.

1.3.4 Out-of-Sample Performance: 78 Test Assets

In this section, I examine a large cross-section of 39 firm characteristics from Kozak, Nagel, and Santosh (2020). Following the past literature such as Lettau and Pelger (2020b), I include both the short and long legs into my analysis, so there are 78 test assets. I focus only on two extreme portfolios for two reasons. First, if I consider all 10 sorted portfolios for each characteristic, there are 390 portfolios. This large cross-section is particularly challenging for the LF-PCA, which uses only long-run components of asset returns in estimation. It implies a trade-off between signaling extraction and estimation noise, so I include only two extreme portfolios to control estimation errors. Also, when I include all 10 sorted portfolios, the portfolio weights are often the most enormous for portfolios in deciles 1 and 10. In other words, most of the relevant information comes from two extreme portfolios.

The entire sample is from August 1963 to December 2019. I further split the whole sample into two equal subsamples. Subsample 1 has 339 monthly observations, spanning from August 1963 to October 1991, and I treat it as the in-sample. Subsample 2 is the out-of-sample (OOS), which is from November 1991 to the end of the sample. As I show in Section 1.2, estimating a linear SDF composed of asset returns is identical to finding the MVE portfolio with the highest achievable Sharpe ratio. It requires me to focus on the OOS performance of asset pricing models, as the in-sample estimate often exaggerates the attainable Sharpe ratio in the real world. For instance, the annualized Sharpe ratio of 78 test assets is higher than 3 in the full sample, which is unreasonably large, according to the good deal bounds in Cochrane and Saa-Requejo (2000). In addition, McLean and Pontiff (2016) document the declining performance of many anomalies post-publication, and Kozak, Nagel, and Santosh (2018) also show that the average returns of 15 factors decrease considerably in the second subsample. Motivated by the previous papers, this paper estimates the PCs and their risk prices using data in the first subsample and evaluates the OOS performance in subsample 2.

HF vs. LF Time-Series Variations

First, I look at the time-series variations (TSVs) explained by different frequency-specific components in the in-sample. The results in the second subsample are largely similar (see Figure 1.A.4). I estimate the spectral density matrix $\hat{\boldsymbol{f}}_{\boldsymbol{R}}(\omega)$ via the DFT described in Appendix 1.A.1.2, where the algorithm estimates only at frequencies $\frac{h}{360}$, $h \in \{1, \ldots, 180\}$. Specially, the HF component corresponds to the cycle length shorter than 36 months, that is, $\frac{360}{h} < 36$, whereas the LF part has a cycle period between 36 and 120 months, that is,

$36 \leq \frac{360}{h} \leq 120$. Therefore, the sample estimates of $\mathbf{\Sigma}_{\boldsymbol{R}}^{HF}$ and $\mathbf{\Sigma}_{\boldsymbol{R}}^{LF}$ are as follows:

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{R}}^{HF} = \frac{1}{170} \sum_{h=11}^{180} \mathcal{R}(\hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})), \quad \hat{\mathbf{\Sigma}}_{\boldsymbol{R}}^{LF} = \frac{1}{8} \sum_{h=3}^{10} \mathcal{R}(\hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})).$$

I further define the fraction of time-series variations explained by HF and LF components as follows:

$$TSV^{HF} = \frac{tr\left[\sum_{h=11}^{180} \hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})\right]}{tr\left[\sum_{h=1}^{180} \hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})\right]} \quad \text{and} \quad TSV^{LF} = \frac{tr\left[\sum_{h=3}^{10} \hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})\right]}{tr\left[\sum_{h=1}^{180} \hat{\boldsymbol{f}}_{\boldsymbol{R}}(\frac{h}{360})\right]}.$$

If returns are uncorrelated, the spectral density matrix is approximately constant across frequencies, so the HF (LF, or above-LF) component accounts for $\frac{170}{180} = 94.5\%$ ($\frac{8}{180} = 4.4\%$, or $\frac{2}{180} = 1.1\%$) of time-series variations. Empirically, however, this LF part explains 5.1% of time-series variations, so this slow-moving component is slightly more important than that predicted by the uncorrelated assumption (see Figure 1.3(a)).

I further compare $\text{Tr}\left[\hat{\mathbf{\Sigma}}_{\boldsymbol{R}}^{LF}\right]$ to $\text{Tr}\left[\hat{\mathbf{\Sigma}}_{\boldsymbol{R}}^{HF}\right]$ and find that the former is around 1.2 times as the latter, which means the LF risk is slightly higher than the HF risk. Next, Figure 1.3(b) computes the ratio of LF-eigenvalues over HF-eigenvalues. An interesting observation is that the top 10 eigenvalues of the LF covariance matrix are 1.5 to 2.5 times as those of the HF one, except for the PC1. Therefore, the LF component has a clearer factor structure.
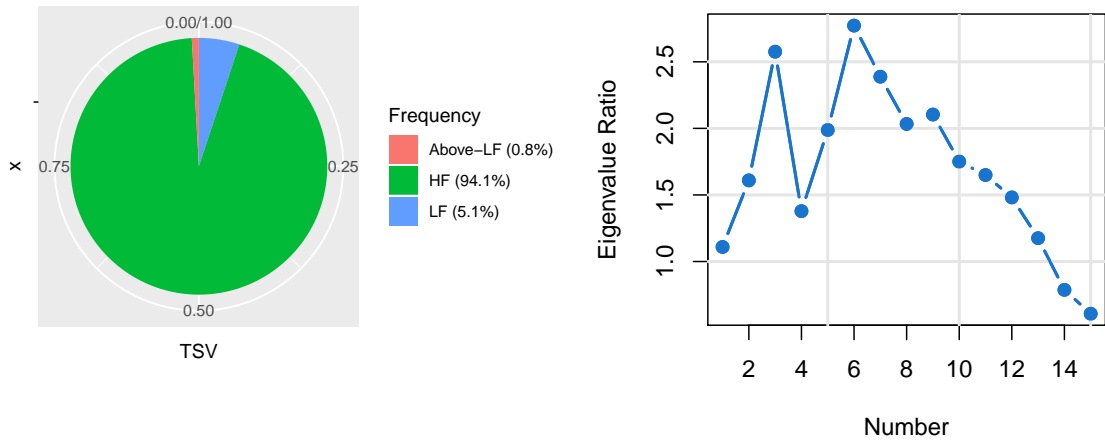
(a) Time-series variations explained by frequency-specific components

(b) LF-eigenvalues / HF-eigenvalues

Figure 1.3: Time-series variations in 78 assets, subsample 1

Panel (a) plots the fraction of time-series variations in 78 asset returns explained by the HF, LF, and above-LF components. Panel (b) plots the ratio of the first 15 low-frequency eigenvalues over high-frequency eigenvalues. The sample is from August 1963 to October 1991.

Out-of-Sample Sharpe Ratio

With the eigendecomposition of the frequency-dependent covariance matrix of asset returns, I construct OOS latent factors following definition 1.1: $\boldsymbol{F}_t^{OOS} = (\boldsymbol{Q}^{IN})^\top \boldsymbol{R}_t^{OOS}$, where $\boldsymbol{Q}^{IN}$ is the eigenvector of the frequency-dependent covariance matrix estimated in the first subsample, and $\boldsymbol{R}_t^{OOS}$ denotes asset returns in the out-of-sample. In addition, I estimate risk prices $\boldsymbol{b}$ for different prior Sharpe ratios in the in-sample, with the estimate denoted as $\hat{\boldsymbol{b}}^{IN}$. In the benchmark case, I use the objective function in equation (1.16) and include latent factors into the regression based on their eigenvalues. In other words, latent factors that drive more common movements among asset returns enter the SDF first. Next, I construct the OOS MVE portfolio, $\mathbf{MVE}_t^{OOS} = (\hat{\boldsymbol{b}}^{IN})^\top \boldsymbol{F}_t^{OOS}$.

Figure 1.4 is the heat-map of the OOS Sharpe ratio. I present only the HF- and LF-PCA results to save space, while the graphs of above-LF-PCA and PCA are in the appendix. In each panel, the x-axis denotes the prior Sharpe ratio of factor models, corresponding to different levels of L_2-shrinkage v_2 in equation (1.16). If I choose a larger prior Sharpe ratio, I impose a gentler shrinkage to risk prices $\boldsymbol{b}$. The y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. For example, the red color represents the "nearly" maximal monthly Sharpe ratio that these factor models can achieve in the out-of-sample, around 0.35 - 0.38 in the data.

Panel (a) in Figure 1.4 and Panel (b) in Figure 1.A.5 show the results of HF-PCA and canonical PCA respectively — they have almost identical heat-maps. Generally, the first six or seven canonical or high-frequency PCs deliver an OOS Sharpe ratio of 0.28-0.29 across a wide range of prior Sharpe ratios. However, this low-dimensional (HF-)PC model still ignores an important priced component in the SDF. For example, when the prior Sharpe ratio is 0.4 or 0.5, in Figure 1.5, the OOS Sharpe ratio increases gradually from 0.28 to 0.37 as more (HF-)PCs enter the SDF. Besides, a substantial L_2-shrinkage helps reduce the required number of latent factors. Especially when $SR_{prior} = 0.2$, I need 20-25 (HF-)PCs to reach the nearly optimal OOS Sharpe ratio. The factor model composed of the extremely low-frequency PCs has a similar observation, as is indicated in Panel (a) in Figure 1.A.5. Since the above-LF-component is moving considerably slowly, estimating the covariance matrix is challenging, so I compare the HF and LF systematic factors and the SDFs composed of them in the following analysis.

Panel (b) plots the OOS Sharpe ratio of LF-PCs. A distinguishing feature is the sparsity in the space of LF-PCs. In Figure 1.5, the first *seven* LF-PCs are almost sufficient to achieve the optimal OOS Sharpe ratio, at around 0.37 per month. In other words, seven systematic factors that explain the most LF common variations in asset returns can span the whole asset space in the out-of-sample. Moreover, this observation is not sensitive to the choice of prior

Sharpe ratio. With a wide range of reasonable prior, such as $SR_{prior} \in [0.3, 0.8]$, the SDF constructed by the first seven LF-PCs is always nearly optimal in the out-of-sample. Last but not least, since PCs are no more than linear transformations of original test asset returns $\boldsymbol{R_t}$, they contain almost identical information. Therefore, the MVE portfolios consisting of HF-PCs, LF-PCs, and original PCs earn just about the same OOS Sharpe ratio as the number of factors entering the SDF approximates N.
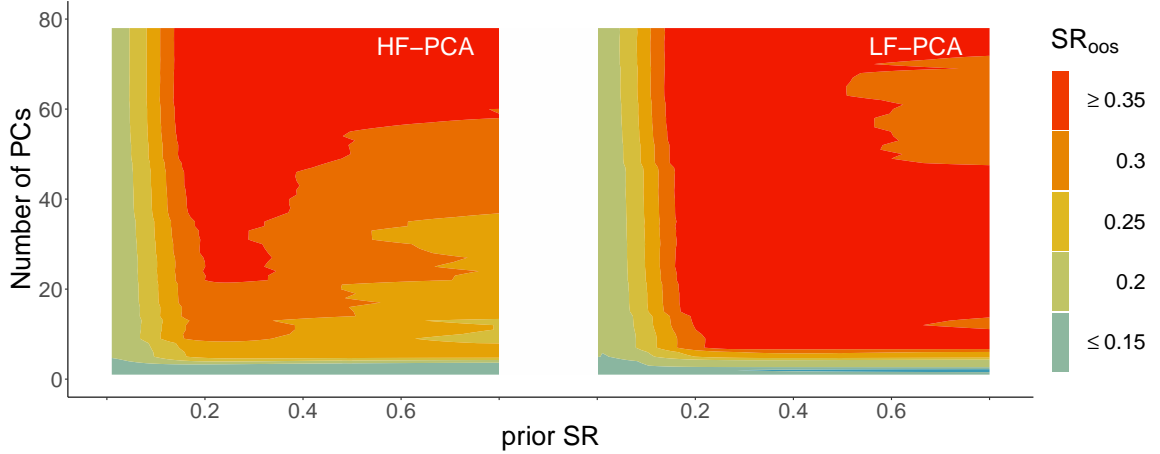
Figure 1.4: OOS Sharpe ratio of HF- vs. LF-PCA, 78 test assets

This graph plots the heat-maps of the OOS Sharpe ratio of HF- vs. LF-PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different levels of OOS Sharpe ratios. I include the PCs into the SDF based on their ability to explain time-series variations.

Out-of-Sample R_{gls}^2

In addition to the OOS Sharpe ratio, I also investigate the GLS R-squared of factor models, denoted by R_{gls}^2. With the in-sample estimate of risk prices, I compute the OOS pricing errors predicted by a factor model, $\boldsymbol{\alpha_R^{OOS}} = \bar{\boldsymbol{R}}_t^{OOS} - \text{Cov}(\boldsymbol{R_t^{OOS}}, \boldsymbol{F_t^{OOS}})\hat{\boldsymbol{b}}^{IN}$, where $\text{Cov}(\boldsymbol{R_t^{OOS}}, \boldsymbol{F_t^{OOS}})$ is the sample covariance matrix between OOS asset returns $\boldsymbol{R_t^{OOS}}$ and OOS factors $\boldsymbol{f_t^{OOS}}$, $\bar{\boldsymbol{R}}_t^{OOS}$ is the sample average of OOS asset returns. R_{gls}^2 is defined as

$$R_{gls}^2 = 1 - \frac{(\boldsymbol{\alpha_R^{OOS}})^\top (\hat{\boldsymbol{\Sigma}}_R^{OOS})^{-1} \boldsymbol{\alpha_R^{OOS}}}{(\bar{\boldsymbol{R}}^{OOS})^\top (\hat{\boldsymbol{\Sigma}}_R^{OOS})^{-1} \bar{\boldsymbol{R}}^{OOS}}. \tag{1.23}$$

R_{gls}^2 has a few satisfying properties. First, R_{gls}^2 has a straightforward economic interpretation: It quantifies the proportion of the squared Sharpe ratio of test assets explained by a factor model. Also, the objective function in equation (1.16) is to maximize R_{gls}^2. Therefore, it is natural to compare R_{gls}^2 in the out-of-sample, consistent with my objective function. Last but not least, R_{gls}^2 is invariant to any non-singular linear transformation of

(a) $SR_{prior} = 0.4$ **(b)** $SR_{prior} = 0.5$
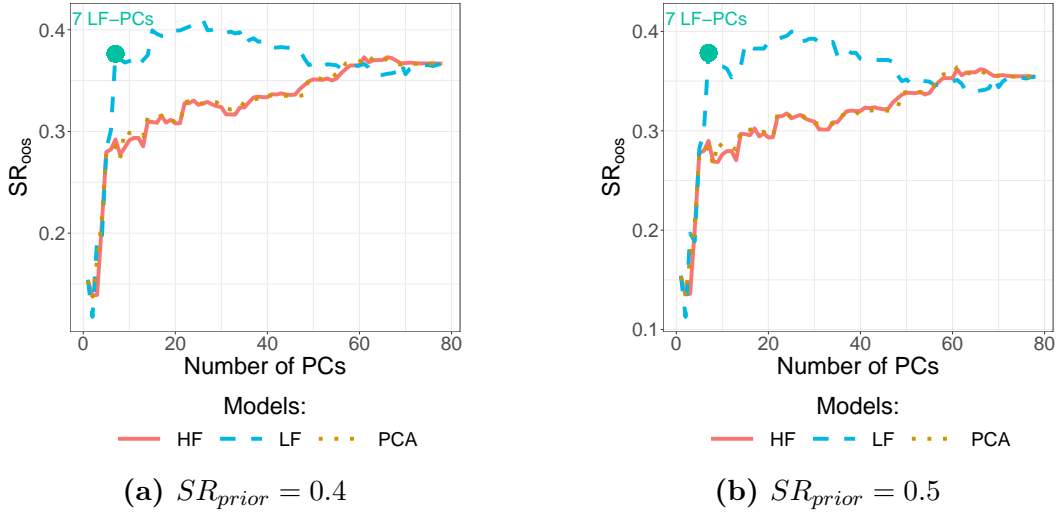
Figure 1.5: Zoom in OOS Sharpe ratio, $SR_{prior} \in \{0.4, 0.5\}$

This graph zooms in the OOS Sharpe ratio of PCA, HF-PCA, and LF-PCA. Different from figure 1.4, this figure shows the estimates using two prior Sharpe ratios, $SR_{prior} \in \{0.4, 0.5\}$.

the original asset space. Specifically, for an arbitrary transformation of asset returns, such as $\boldsymbol{Y_t^{OOS}} = \boldsymbol{P}^\top \boldsymbol{R_t^{OOS}}$, where $\boldsymbol{P}$ is nonsingular, R_{gls}^2 of pricing $\boldsymbol{Y_t^{OOS}}$ is exactly identical to that of $\boldsymbol{R_t^{OOS}}$. By focusing on R_{gls}^2, there is no need to choose whether the SDF should price original asset returns or their transformation, such as PCs.
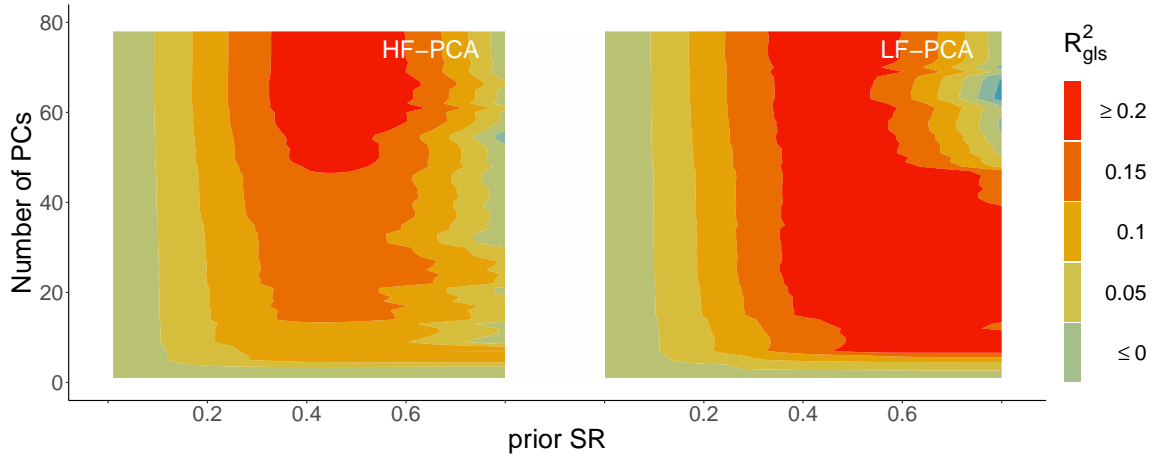
Figure 1.6: OOS R_{gls}^2 of HF- vs. LF-PCA, 78 test assets

This graph plots the heat-maps of the OOS R_{gls}^2 of HF- vs. LF-PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different levels of OOS R_{gls}^2. I include the PCs into the SDF based on their ability to explain time-series variations.

Figure 1.6 plots the heat-maps of OOS R_{gls}^2 of HF- and LF-PCA. Related plots of above-LF-PCA and original PCA can be found in Figure 1.A.6. Similar to the OOS Sharpe ratio,

the PCA and HF-PCA share similar patterns — I need many latent factors to obtain the optimal OOS R^2_{gls}. On the contrary, I can choose a relatively parsimonious SDF consisting of LF-PCs. For instance, when SR_{prior} is between 0.5 and 0.8, the OOS R^2_{gls} of a 7 LF-factor-model is around 18% – 20%. Overall, the exploration of R^2_{gls} provides further evidence on the sparsity of LF-SDF. At the same time, the HF-SDF always needs more than 30 latent factors to achieve a nearly optimal OOS R^2_{gls}.

Zoom in High-Frequency Intervals

In the previous analysis, I define the HF interval as $\tau^{HF} \in [2, 36)$ and find the sparsity of latent factor models only at low frequencies. However, the definition of the HF interval is probably too wide to capture certain pricing information at a specific high frequency. For instance, the short-term reversal in Jegadeesh (1990) manipulates extremely fast-moving information in predicting future stock returns. To explore whether the performance of latent factor models varies significantly under alternative definitions of HF intervals, I consider a further division of $\tau^{HF} \in [2, 36)$: (1) $[2, 3)$, (2) $[3, 6)$, (3) $[6, 12)$, and (4) $[12, 36)$. Next, I will examine the OOS Sharpe ratio of latent factor models in these four HF intervals.

Figure 1.7 plots the heat-maps of the OOS Sharpe ratio of latent factor models composed of PCs in these four HF intervals. Clearly, I always need more than 20 latent factors to achieve the optimal OOS Sharpe ratio. In addition, the performance of factor models is sensitive to the choice of the L_2-penalty — a significant penalty or a small prior Sharpe ratio is necessary to ensure a decent OOS performance. Hence, the sparsity of latent factor models only exists in the LF frequency interval $[36, 120)$.

The previous empirical results shed light on the dynamics of priced state variables in the cross-section of 78 test assets. According to Proposition 1.2, the maximal Sharpe ratio implied by the SDF is frequency-dependent only due to state variables $\boldsymbol{X_t}$. More importantly, it implies that $\sum_{j=1}^{p} b_{X,i}^2 \boldsymbol{f}_{\boldsymbol{X_i}}(\omega)$, the second term in the spectral density function of the SDF, is on average larger at low frequencies. While either the canonical or HF-PCA fails to identify this persistent state variable, the LF-PCA recovers it as one of the largest factors explaining time-series variations in long-horizon asset returns. This conditional information is also priced in the cross-section. In the language of ICAPM (Merton (1973)), stock market participants have the incentive to hedge some slow-moving state variables, as those state variables can predict the future stock returns and economic environments and therefore affect investors' future investment opportunity set. Because of the hedging demand, the state variables command non-zero risk prices, so a valid SDF should not omit them. Finally, the fast-moving state variables are not essential in the monthly data. As Figure 1.7 indicates, the SDF is similarly dense in the space of extremely HF systematic factors.
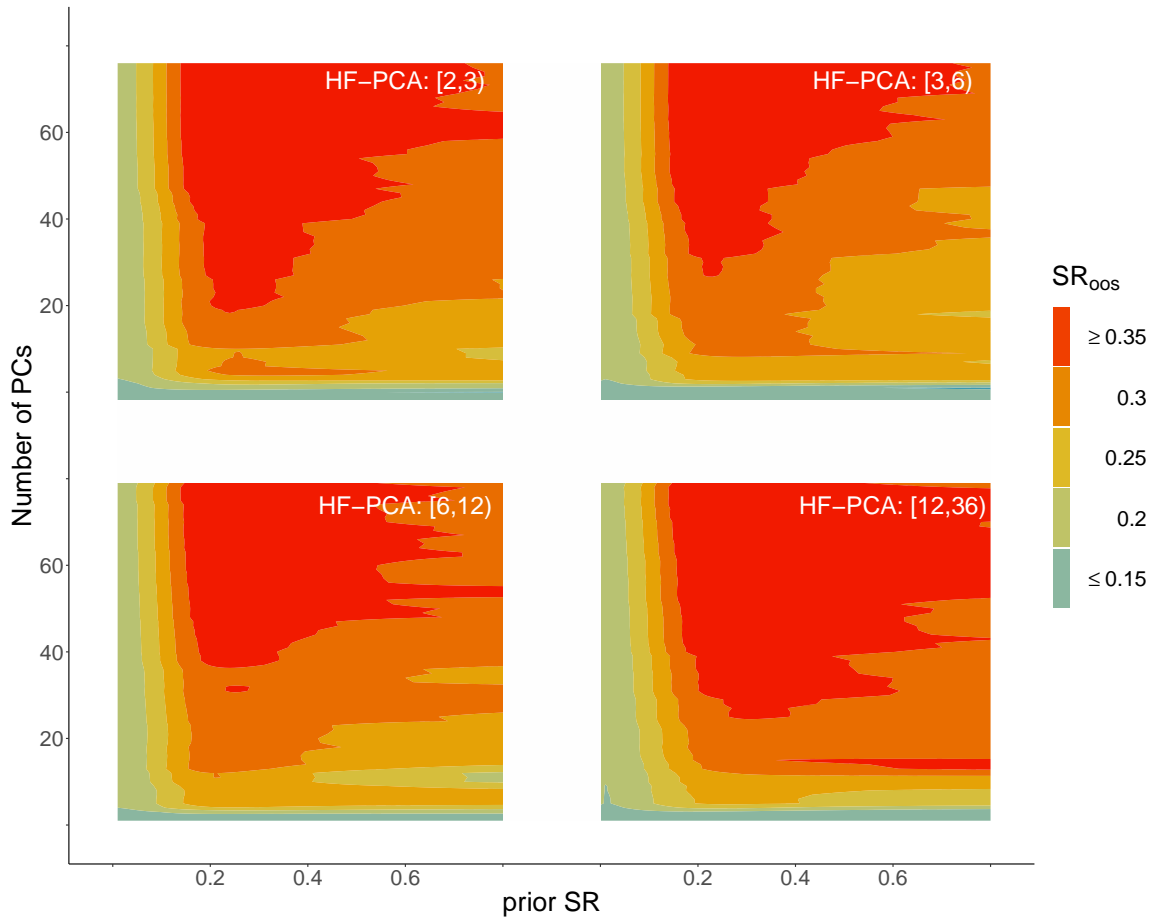
Figure 1.7: Robustness Check: OOS Sharpe ratio of 78 test assets, different HF intervals

This graph plots the heat-maps of OOS Sharpe ratios of HF-PCA, where I divide the HF intervals into four sub-intervals: $[2, 3)$, $[3, 6)$, $[6, 12)$ and $[12, 36)$. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF.

Kozak, Nagel, and Santosh (2020) Estimation: Imposing Model Sparsity

As previous empirical results indicate, an SDF composed of (HF-)PCs cannot be parsimonious in terms of either the OOS Sharpe ratio or R_{gls}^2. What if I impose the sparsity of factor models? In this part, I follow the Kozak, Nagel, and Santosh (2020) procedure, described in equation (1.17), that includes L_1 shrinkage. According to the closed-form solution in equation (1.18), this procedure selects PCs with higher in-sample average returns first. Also, a larger v_1 renders more factors to have zero risk prices, so this algorithm enforces the sparsity of factor models.

I show the OOS Sharpe ratio of the MVE portfolios from the Kozak, Nagel, and Santosh (2020) estimation in Figure 1.8. First, 15 PCs or 20 HF-PCs and canonical PCs can deliver the nearly optimal Sharpe ratio, so the SDF does become sparser. However, it must be the case that the objective function chooses some small PCs that have essential pricing

information. At the same time, the Kozak, Nagel, and Santosh (2020) procedure can still discover a sparse LF-SDF, with the first seven LF-PCs commanding a 0.4 monthly Sharpe ratio. According to the heat-maps of OOS Sharpe ratio in Figure 1.A.7, the LF-SDF is always sparse when the prior Sharpe ratio that I use to estimate the risk prices of LF latent factors is greater than 0.3.

(a) $SR_{prior} = 0.4$ **(b)** $SR_{prior} = 0.5$
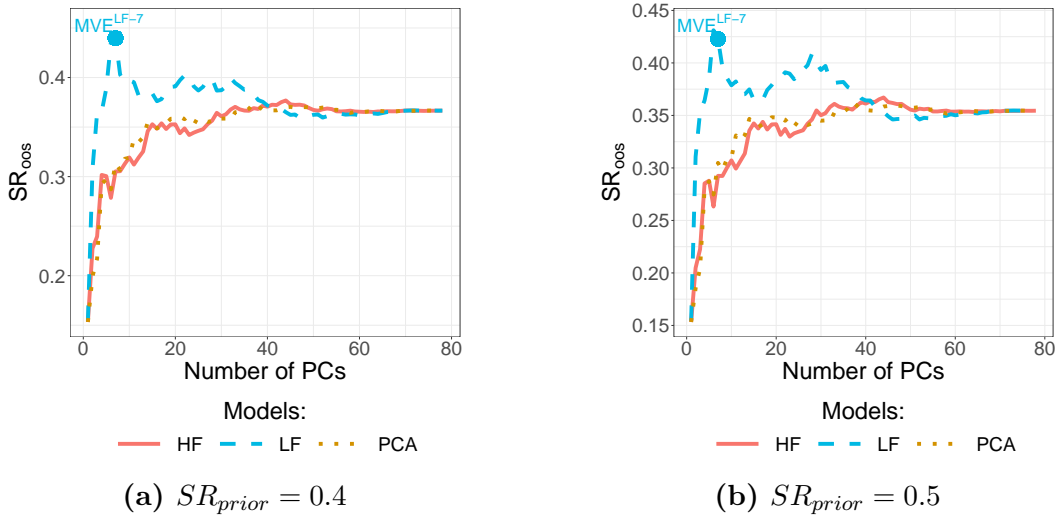
Figure 1.8: OOS Sharpe ratio, Kozak, Nagel, and Santosh (2020) estimation

This graph zooms in the OOS Sharpe ratio of the canonical PCA, HF-PCA, and LF-PCA. The prior monthly Sharpe ratio is set to be 0.4 in Panel (a) and 0.5 in Panel (b). I estimates risk prices using the Kozak, Nagel, and Santosh (2020) algorithm described in equation (1.17). Latent factors are PCs of 78 test asset returns.

Factor models that have been proposed in past literature are mostly sparse, such as the Fama-French three-factor model. However, there is no particular reason why a factor model must be sparse, even though people often pursue parsimonious models. Giannone, Lenza, and Primiceri (2021a) call this "the illusion of sparsity." Moreover, it is almost unlikely to select a few firm characteristics, such as size and book-to-market ratio, and use them to span the whole asset space. For instance, Kozak, Nagel, and Santosh (2020) show that characteristics-sparse SDFs formed from a few factors cannot appropriately explain the cross-section of expected stock returns in the out-of-sample. In addition, Bryzgalova, Huang, and Julliard (2021) use a continuous spike-and-slab Bayesian prior to study 51 observable tradable and nontradable factors, and they find that within a wide range of reasonable prior Sharpe ratios of the SDF, 95% posterior credible intervals of the number of factors in the true model are between 16 and 32.

Why do we desire the sparsity of latent factor models? According to Kozak, Nagel, and Santosh (2018), the absence of near arbitrage opportunities implies that factors capturing the most systematic common variations in asset returns are non-diversifiable, so market participants earn non-zero risk premia for taking these risks. However, I observe some small

(HF-)PCs bringing nontrivial risk premia. These factors explain less than 0.1% of the time-series variation, which implies that they are idiosyncratic shocks. Accordingly, they should not command sizable risk premia; otherwise, arbitrageurs can include those small PCs into their portfolio without increasing their investment risk significantly — instead, the sparsity of LF-SDF solves this puzzle to some extent.

How should we interpret the sparsity of the LF-SDF? On the one hand, it makes economic sense to observe a sparse LF-SDF. Suppose investors are buy-and-hold investors who pay more attention to the long-term trade-off between risk and returns, or investors have Epstein-Zin preference and are particularly risk-averse to the long-run uncertainty. Under these scenarios, the LF-SDF should imply a higher Sharpe ratio than the HF-SDF, because those LF systematic factors are the most risky in the long horizon. On the other hand, some persistent state variables explain a small fraction of common variations in single-period returns, but they are much more prominent in the long horizon. Hence, the LF-PCA boosts the signal of this persistent conditional information and recovers them partially or wholly.

1.3.5   Do Celebrated Models Explain HF and LF Risks?

This section further compares the OOS MVE portfolios of latent factors with the following benchmark models: (1) CAPM, (2) Fama and French (1993) three factors (FF3), (3) Fama and French (2015) five factors (FF5), (4) Carhart (1997) four factors (Carhart4), and (5) Hou, Xue, and Zhang (2015) four factors (Q4). First of all, I examine whether these five sparse factor models can explain HF- and LF-MVE portfolios by running time-series regressions as follows:[14]

$$\mathbf{MVE}_t^Z = \alpha + \boldsymbol{\beta}^\top \boldsymbol{B_t} + \eta_t, \quad Z \in \{\mathrm{HF}, \mathrm{LF}\},$$

where $\boldsymbol{B_t}$ is one of the five benchmark models mentioned before. I report three test-statistics in Table 1.2: (1) α, (2) t-statistics of α, and (3) the adjusted R-squared, denoted as R_{adj}^2. To control for the serial dependence of pricing errors, I use Newey and West (1987) standard errors with both 36 lags (t-stat I) and 12 lags (t-stat II). In Table 1.2, I estimate risk prices of PCs under the prior Sharpe ratio equal to 0.4. To enhance interpretability, I normalize all MVE portfolios to have the same volatility as the market factor.

The first panel in Table 1.2 examines CAPM. Not surprisingly, the market factor alone entirely fails to explain neither HF- nor LF-MVE portfolios. The pricing errors are enormous, always greater than 1% per month. Moreover, LF-MVE portfolios always have higher alphas and t-statistics than HF ones; hence, they are more difficult ro explain. Interestingly, I

---

[14]Empirically, the MVE portfolios composed of the first several HF or canonical PCs are almost identical. Specifically, their correlation coefficients are around 98 – 99%. Therefore, I focus on comparing HF- and LF-MVE portfolios.

observe relatively low R^2_{adj}, less than 10% in all columns. Since the first PC in the cross-section is always a level factor that is highly correlated with the market factor, the low R^2_{adj} in CAPM implies that the SDF loads heavily on other lower-order latent factors.

FF3 extends CAPM by including the size and value factors. Compared to CAPM, FF3 slightly reduces the pricing errors and significantly increases R^2_{adj}, particularly in the regression of HF-MVE portfolios. However, FF3 still fails to explain the OOS MVE portfolios of latent factors.

Carhart4 includes the momentum factor into FF3. Intriguingly, the alphas of MVE portfolios reduce by more than 40% compared to the previous two regressions, although all remain significantly positive. Moreover, the inclusion of the momentum factor improves the time-series fit dramatically. For example, Carhart4 explains 52% of time-series variation in the MVE portfolio of seven LF-PCs, while R^2_{gls} in FF3 is just 16%.

Table 1.2: Do celebrated models explain HF and LF risks?

		Panel (A). $\mathbf{MVE}_t^{HF}$				Panel (B). $\mathbf{MVE}_t^{LF}$			
		7 PCs	8 PCs	9 PCs	10 PCs	7 PCs	8 PCs	9 PCs	10 PCs
CAPM	α	1.03%	1.00%	1.16%	1.18%	1.39%	1.38%	1.38%	1.41%
	t-stat I	(2.89)	(2.47)	(2.88)	(2.99)	(3.57)	(3.72)	(3.72)	(3.76)
	t-stat II	(3.05)	(2.72)	(3.16)	(3.28)	(4.48)	(4.60)	(4.60)	(4.60)
	R^2_{adj}	8.02%	5.55%	0.30%	0.44%	7.41%	6.38%	5.76%	4.36%
FF3	α	0.80%	0.77%	0.94%	0.98%	1.27%	1.27%	1.27%	1.29%
	t-stat I	(4.87)	(3.91)	(4.29)	(4.44)	(4.45)	(4.65)	(4.65)	(4.75)
	t-stat II	(4.76)	(4.13)	(4.60)	(4.73)	(5.82)	(5.95)	(5.97)	(6.05)
	R^2_{adj}	40.97%	38.39%	31.92%	26.44%	15.84%	14.24%	13.06%	12.05%
Carhart4	α	0.44%	0.39%	0.57%	0.59%	0.83%	0.82%	0.81%	0.84%
	t-stat I	(3.18)	(2.61)	(3.09)	(3.18)	(3.75)	(3.88)	(3.87)	(3.91)
	t-stat II	(2.81)	(2.39)	(3.08)	(3.14)	(4.40)	(4.46)	(4.45)	(4.44)
	R^2_{adj}	65.13%	65.62%	57.00%	53.91%	52.36%	51.2%	51.81%	50.21%
FF5	α	0.43%	0.43%	0.48%	0.53%	0.87%	0.84%	0.83%	0.84%
	t-stat I	(2.55)	(2.36)	(2.87)	(3.11)	(3.15)	(3.17)	(3.12)	(3.13)
	t-stat II	(2.43)	(2.42)	(3.24)	(3.43)	(3.94)	(3.88)	(3.81)	(3.87)
	R^2_{adj}	49.98%	48.09%	48.77%	41.91%	27.95%	28.02%	27.52%	27.26%
Q4	α	0.32%	0.23%	0.26%	0.33%	0.76%	0.73%	0.71%	0.72%
	t-stat I	(2.07)	(1.30)	(1.47)	(1.81)	(2.77)	(2.82)	(2.75)	(2.81)
	t-stat II	(1.60)	(1.10)	(1.30)	(1.61)	(3.25)	(3.22)	(3.14)	(3.24)
	R^2_{adj}	42.81%	39.87%	44.7%	40.47%	31.31%	31.35%	31.71%	31.16%

This table tests whether five sparse factor models proposed in past literature can explain the MVE portfolios composed of latent factors. I construct the MVE portfolios using the first seven to 10 latent factors following the same steps as in the section 1.3.4. I estimate the factors' risk prices under the prior Sharpe ratio of 0.4. The five benchmark models include (1) CAPM, (2) Fama and French (1993) three factors (FF3), (3) Fama and French (2015) five factors (FF5), (4) Carhart (1997) four factors (Carhart4), and (5) Hou, Xue, and Zhang (2015) four factors (Q4). I report three test-statistic in table 1.2: (1) α, (2) t-statistics of α, and (3) adjusted R-squared, denoted as R^2_{adj}. To control for the serial dependence of pricing errors, I use Newey and West (1987) standard errors with both 36 lags (t-stat I) and 12 lags (t-stat II).

In the last two panels, I consider two models with both investment and profitability

factors in them. Simply speaking, FF5 differs from Q4 in the additional value factor in FF5, and they adopt a slightly distinct approach to construct factors. In addition, Q4 is better at explaining the MVE portfolios than FF5. Notably, pricing errors of HF-MVE portfolios are remarkably smaller, declining to around 0.3% per month, and are no longer significant, except for t-statistic I in the column of seven HF-PCs . On the other hand, LF-MVE portfolios still have sizable and statistically significant pricing errors, at around 0.7% per month. I have similar empirical findings under another prior Sharpe ratio equal to 0.5 (see Table 1.A.3 in the appendix).

In short, none of the five benchmark models can explain LF-MVE portfolios, while the Q4 model in Hou, Xue, and Zhang (2015) is capable of rationalizing the abnormal returns of the HF-MVE portfolios.

Next, I test whether LF-MVE portfolios can explain HF-MVE ones or whether the opposite is valid. Similarly, I run time-series regressions, but the benchmark model $\boldsymbol{B_t}$ becomes either $\mathbf{MVE}_t^{HF}$ or $\mathbf{MVE}_t^{LF}$. Table 1.3 reports the results under the prior Sharpe ratio 0.4.

In Panel (a), I regress $\mathbf{MVE}_t^{HF}$ on $\mathbf{MVE}_t^{LF}$, both of which are constructed by the first seven, eight, nine, and 10 PCs. First, pricing errors are almost zeros in the statistical sense and less than 0.1% per month. In other words, the LF-MVE portfolios can span the HF-MVE portfolios. On the other hand, I regress $\mathbf{MVE}_t^{LF}$ on $\mathbf{MVE}_t^{HF}$ in Panel (b). Unlike Panel (a), pricing errors are always significantly positive, implying that the HF-MVE ignores an essential priced component of LF-MVE.

To sum up, the evidence in Tables 1.2 and 1.3 indicates that MVE portfolios, or SDFs, consisting of LF-PCs, should be the right benchmark. The first few LF-PCs can construct an LF-SDF that yields nearly optimal OOS Sharpe ratio, and none of the five notable factor models proposed in the past literature or HF-MVE portfolios can explain them. At the same time, they can fully explain HF-MVE portfolios in the out-of-sample.

Table 1.3: Which benchmark? HF vs. LF Tangency Portfolios

	Panel (A): $\mathbf{MVE}_t^{HF} = \alpha + \beta\mathbf{MVE}_t^{LF} + e_t$				Panel (B): $\mathbf{MVE}_t^{LF} = \alpha + \beta\mathbf{MVE}_t^{HF} + e_t$			
	7 PCs	8 PCs	9 PCs	10 PCs	7 PCs	8 PCs	9 PCs	10 PCs
α	-0.10%	-0.10%	0.10%	0.00%	0.60%	0.60%	0.70%	0.60%
t-stat I	(-0.74)	(-0.62)	(0.51)	(-0.08)	(3.29)	(4.19)	(2.79)	(2.70)
t-stat II	(-0.63)	(-0.52)	(0.44)	(-0.07)	(4.01)	(4.50)	(3.04)	(3.06)
R^2_{adj}	68.89%	62.68%	53.86%	63.23%	68.89%	62.68%	53.86%	63.23%

This table tests whether the LF-MVE portfolio can explain the HF-MVE or whether the opposite is valid. I construct the MVE portfolios using the first 7 – 10 latent factors following the same steps as in Section 1.3.4. I estimate the factors' risk prices under the prior Sharpe ratio of 0.4. I report three test-statistic in Table 1.2: (1) α, (2) t-statistics of α, and (3) adjusted R-squared, denoted as R^2_{adj}. To control for the serial dependence of pricing errors, I use Newey and West (1987) standard errors with both 36 lags (t-stat I) and 12 lags (t-stat II).

1.3.6 Origins of Economic Risks in SDFs

Why do I observe the sparsity of latent factor models only in the space of LF-PCs? Why do sparse LF-MVE portfolios earn higher Sharpe ratios than those composed of the first few HF or canonical PCs? Do they represent different sources of economic fundamentals? This section attempts to answer these questions by studying the economic drivers behind the linear SDFs consisting of HF and LF systematic factors.

I consider the SDFs composed of the first *seven* HF-PCs or LF-PCs. I denote them as the HF-SDF and LF-SDF, respectively. From the heat-maps in Figure 1.4, the first seven LF-PCs can generate nearly optimal OOS Sharpe ratios under a wide range of prior distributions. In addition, the inclusion of extra PCs into the SDF adds enormous unpriced noises but minimal additional pricing information. Last but not least, the space of the first seven HF-PCs is almost identical to that of the first seven canonical PCs, so I focus on comparing the HF-SDF to LF-SDF.

Past literature often uses the first several largest PCs of single-period returns, which are empirically identical to HF-PCs, to construct the linear SDF. However, my previous empirical findings indicate that such SDFs can neglect a vast priced component. Hence, I decompose the LF-SDF[15] ($\mathcal{M}_t^{LF}$) into two components, the first of which is perfectly correlated with the HF-SDF ($\mathcal{M}_t^{HF}$) and another of which is the orthogonal part as follows:

$$\mathcal{M}_t^{LF} = \beta_{HF}\mathcal{M}_t^{HF} + \mathcal{M}_t^{missing}, \quad \mathcal{M}_t^{HF} \perp \mathcal{M}_t^{missing}. \tag{1.24}$$

Similarly, I project the HF-SDF into the linear space of the LF-SDF and extract an uncorrelated component, denoted by $\mathcal{M}_t^{unpriced}$,

$$\mathcal{M}_t^{HF} = \beta_{LF}\mathcal{M}_t^{LF} + \mathcal{M}_t^{unpriced}, \quad \mathcal{M}_t^{LF} \perp \mathcal{M}_t^{unpriced}. \tag{1.25}$$

Table 1.4 reports the correlation matrix and Sharpe ratios implied by $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$. As mentioned before, $\mathcal{M}_t^{LF}$ implies a higher Sharpe ratio than $\mathcal{M}_t^{HF}$, and both SDFs imply statistically significant Sharpe ratios with t-statistics greater than 4. Moreover, $\mathcal{M}_t^{HF}$ accounts for only 69% of the time-series variation of $\mathcal{M}_t^{LF}$ but misses a considerable component $\mathcal{M}_t^{missing}$ that earns a monthly Sharpe ratio of around 0.2 and has a t-statistic equal to 4.6. In the following tables, I call $\mathcal{M}_t^{missing}$ the *missing-SDF*, which means that the traditional PCA or HF-PCA misses a huge priced component of the LF-SDF. Not surprisingly, the part of HF-SDF orthogonal to LF-SDF has almost zero Sharpe ratio, so this is an unpriced component. Hence, I will call $\mathcal{M}_t^{unpriced}$ the *unpriced-SDF*.

---

[15]In this paper, the SDF is equal to one minus the MVE portfolio: $\mathcal{M}_t = 1 - \mathbf{MVE}_t$. Hence, it is equivalent to studying the MVE portfolios.

Table 1.4: Correlation among $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$

Corr.	$\mathcal{M}_t^{LF}$	$\mathcal{M}_t^{unpriced}$	$\mathcal{M}_t^{HF}$	$\mathcal{M}_t^{missing}$	SR	t-stat (36 lags)
$\mathcal{M}_t^{LF}$	1.00				0.376	6.22
$\mathcal{M}_t^{unpriced}$	0.00	1.00			0.037	0.67
$\mathcal{M}_t^{HF}$	0.83	0.56	1.00		0.292	4.71
$\mathcal{M}_t^{missing}$	0.56	-0.83	0.00	1.00	0.240	4.56

This table plots the correlation matrix and Sharpe ratio (SR) of the following four variables: $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$. $\mathcal{M}_t^{LF}$ and $\mathcal{M}_t^{HF}$ are OOS MVE portfolios composed of the first seven HF- or LF-PCs, and they are constructed by the procedures in Section 1.3.4 under the prior Sharpe ratio 0.4. The last column reports the t-statistics of Sharpe ratio using the Newey and West (1987) standard errors with 36 lags. The out-of-sample period spans from November 1991 to December 2019.

The findings in Table 1.4 also relate to literature that attempts to denoise the tradable factor. For instance, Golubov and Konstantinidi (2019) decompose the market-to-book ratio into market-to-value and value-to-book components — the market-to-value component drives nearly all the risk premium of the value strategy. In addition, Daniel, Mota, Rottke, and Santos (2020) document that unpriced components explain a reasonably large amount of Fama-French five factors, and they propose a novel way to hedge the unpriced components. By focusing on the long-term comovement of asset returns, the LF-SDF significantly reduces the unpriced component.

Dynamics of SDFs: Variance Ratio Test

Theoretically, the LF-PCA has a better finite sample performance than the HF-PCA and canonical PCA because studying the long-horizon returns boosts the signal of some persistent conditional information driving the asset returns and detects them empirically. If the previous argument is valid, the LF-SDF must capture some conditional information ignored by the HF-SDF, so LF-SDF should have a different dynamic across multiple horizons. To explore the dynamics of SDFs, I resort to the variance ratio test, which is calculated as

$$VR(h) = \frac{\text{Var}(\mathcal{M}_{t,t+1} + \cdots + \mathcal{M}_{t+h-1,t+h})}{h \times \text{Var}(\mathcal{M}_{t,t+1})}, \tag{1.26}$$

where $\mathcal{M}_{t,t+1}$ is the single-period SDF. I can also rewrite the variance ratio as a weighted average of the autocorrelations of $\mathcal{M}_{t,t+1}$. A useful benchmark is the IID case, where the variance ratio test equals 1 at any horizon.

Figure 1.9 displays the variance ratios for $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$. The blue dotted lines are 95% confidence intervals of the variance ratios. If the solid red line crosses the dotted blue lines, I can reject the null hypothesis of the IID assumption. Panels (a) and (b) show the variance ratios for the HF-SDF and LF-SDF, respectively. While

(a) HF-SDF

(b) LF-SDF

(c) Missing-SDF
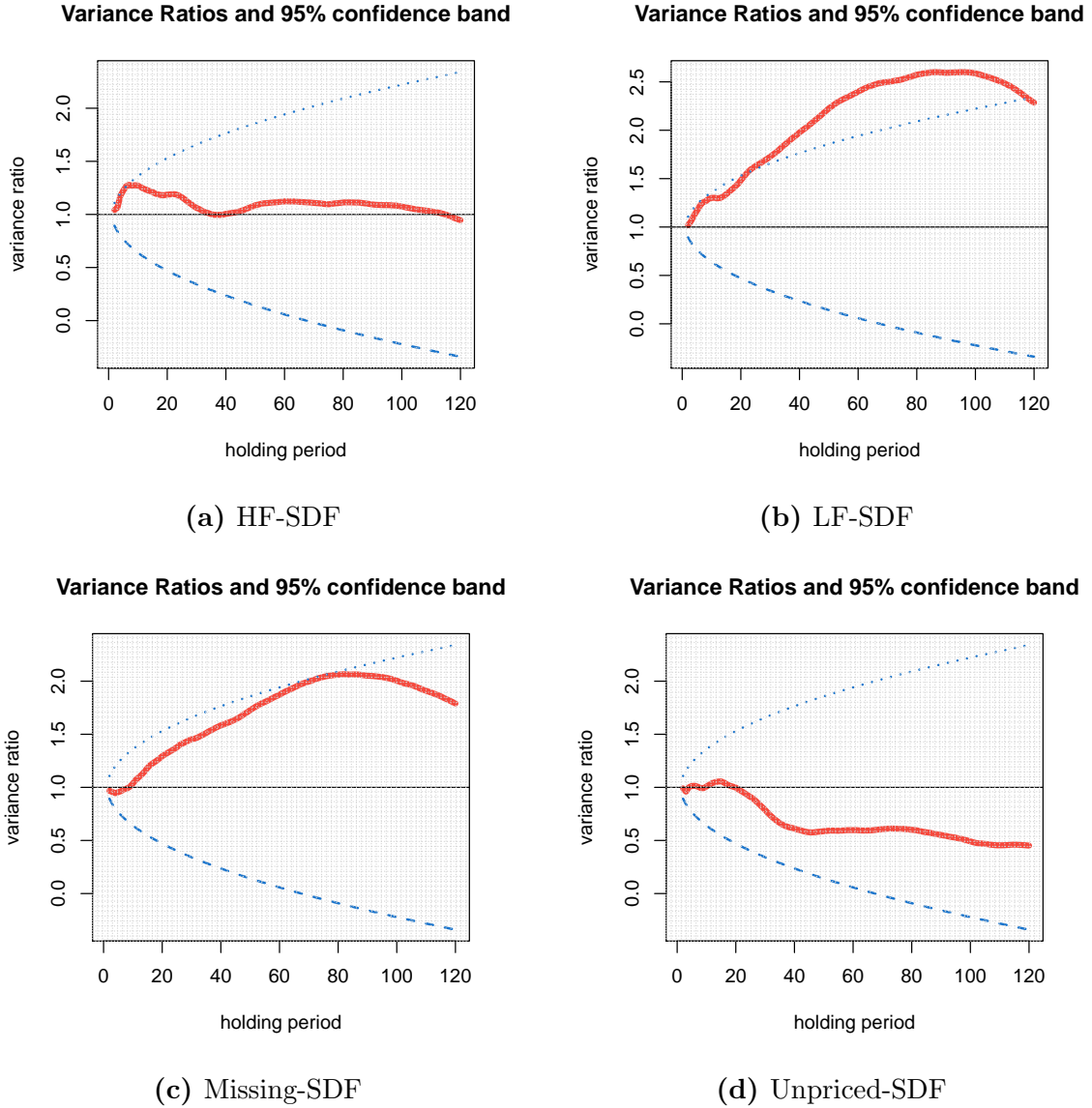
(d) Unpriced-SDF

Figure 1.9: Variance Ratio of the SDF Components

This graph plots the variance ratios of $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$, calculated as

$$VR(h) = \frac{\text{Var}(\mathcal{M}_{t,t+1} + \cdots + \mathcal{M}_{t+h-1,t+h})}{h \times \text{Var}(\mathcal{M}_{t,t+1})},$$

where $\mathcal{M}_{t,t+1}$ is the single-period SDF. The HF-SDF and LF-SDF consist of the largest seven HF- and LF-PCs. The prior (monthly) Sharpe ratio used to estimated the risk prices is set to be 0.4. The blue dotted lines are 95% confidence intervals of the variance ratios. If the red solid line crosses the blue dotted lines, I can reject the null hypothesis of the IID assumption for the linear SDFs.

the HF-SDF exhibits limited autocorrelation over time, the LF-SDF displays a remarkable deviation from the IID assumption. For example, a five-year investor holding the LF-MVE portfolio is subject to double the variance of an investor with a monthly holding period. In addition, the variance ratio of the LF-SDF peaks between the six- and seven-year horizon, but it starts to decrease slowly after the seven-year horizon. Intuitively, the LF-SDF is riskier than HF-SDF from the perspective of long-term investors, so it should command a higher Sharpe ratio to compensate for bearing additional low-frequency risks.

Clearly, there are essential persistent components in the dynamic of the LF-SDF. Panel (c) further plots the variance ratio for the missing-SDF, which manifests a similar dynamic as the LF-SDF. Combining the evidence in Panels (a), (b), and (c), I conjecture that the LF-SDF, which is the suitable benchmark and captures the highest attainable Sharpe ratio, contains two components: (1) the first component is spanned by the HF-SDF, which mainly captures the short-term information in asset returns and is roughly conditionally uncorrelated over time, and (2) the second component is ignored by canonical PCA, which identifies some persistent information that commands sizable risk premium. Panel (d) also presents the variance ratio for the unpriced-SDF. I confirm that this component mainly reflects short-horizon information, with a decreasing variance ratio after two years.

Cumulative Returns of MVE Portfolios

Next, I examine the cumulative performance from the perspective of an investor of the LF-MVE, HF-MVE, missing-MVE, and the market portfolio. To increase interpretability, I normalize all portfolio returns to have the same volatility as the market portfolio, about 4.2% per month. The portfolio weights of the MVE portfolios are determined by the data in the first subsample, so there is no looking-forward bias. Figure 1.10 plots the log of cumulative excess returns from November 1991 to December 2019 (OOS). The solid red line indicates that the LF-MVE portfolio has the best long-horizon performance, with a log cumulative excess return of around 5. As a comparison, the investor earns cumulative log-returns of 3.83 and 2.03 in the HF-MVE portfolio (solid blue line) and the market portfolio (solid green line), respectively. Another surprising fact is that investors of the MVE portfolios do not lose money during the dot-com bubble, while the market portfolio experiences a -40% return. However, the market values of all portfolios plummet during the 2008 global financial crisis.

The missing-MVE portfolio (solid orange line) is the component of the LF-MVE that is uncorrelated with the HF-MVE portfolio, and its behaviors are different from the HF-MVE portfolio. For instance, the missing-MVE portfolio has an extraordinary performance in the late 1990s, while the HF-MVE portfolio has an almost zero excess return during the same period.

49
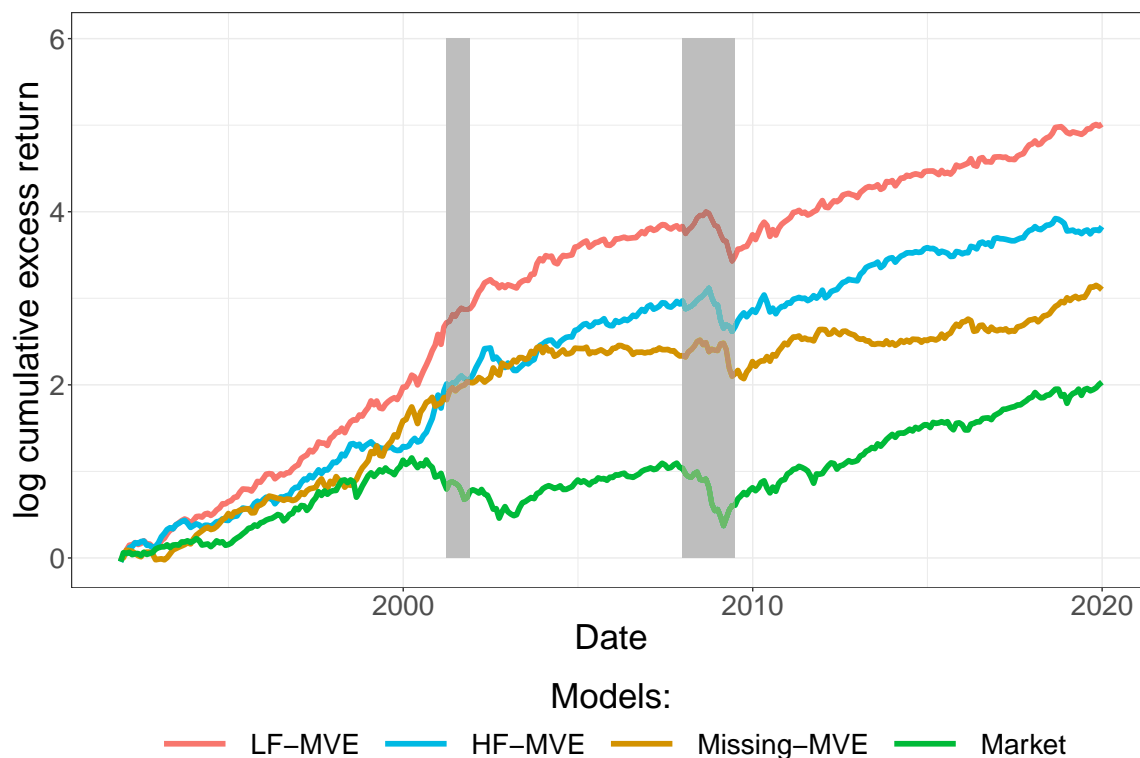
Figure 1.10: Log Cumulative Excess Returns of the MVE Portfolios

This graph plots the log of cumulative excess returns of the LF-MVE, HF-MVE, Missing-MVE (from table 1.4), and market portfolios. I normalize the LF-MVE, HF-MVE, Missing-MVE portfolios to have the same monthly volatility as the market portfolio. The sample spans from November 1991 to December 2019. Shaded areas denote the NBER recession periods: (1) 2001/03 – 2001/11 and (2) 2007/12 – 2009/06.

With the decompositions of SDFs in equations (1.24) and (1.25), I examine how each component in Table 1.4 relates to economic risks. Specifically, I regress each economic variable on different SDF components and conduct statistical tests on the correlation coefficients between each economic variable and SDFs.

There are two primary objectives for these regressions. First, I attempt to understand the economics behind SDFs. For example, the LF-SDF implies a higher Sharpe ratio than the HF-SDF. However, the extra Sharpe ratio earned by the missing-SDF ($\mathcal{M}_t^{missing}$) is probably the compensation for bearing economic risks that the HF-SDF does not load on. Moreover, given the LF-SDF as the proper benchmark, I desire to learn whether different economic risks drive the HF component ($\beta_{HF}\mathcal{M}_t^{HF}$) and the persistent component ($\mathcal{M}_t^{missing}$).

Second, it helps study the risk premium of a nontradable economic factor. As Cochrane (2009) indicates, we can define its risk premium as $-\text{Cov}(Y_t, \mathcal{M}_t)$, where Y_t is the nontradable factor. Similarly, Giglio and Xiu (2021) show that we should project a nontradable factor into the space of the largest principal components of a huge cross-section of test assets. However, Section 1.3.4 points out that the SDF constructed by either PCs or HF-PCs

50

potentially ignore an important priced component of the true SDF. According to equation
(1.24), $\text{Cov}(Y_t, \mathcal{M}_t^{LF}) = \beta_{HF}\text{Cov}(Y_t, \mathcal{M}_t^{HF}) + \text{Cov}(Y_t, \mathcal{M}_t^{missing})$. An economic variable can
be uncorrelated with $\mathcal{M}_t^{HF}$ but significantly correlated with the missing part, $\mathcal{M}_t^{missing}$.
Hence, the study of LF latent factors provides additional insights into nontradable economic
risks.

Table 1.5 reports the results. I consider eight economic variables, whose definitions are
in Table 1.A.1. I standardize both dependent and independent variables so that readers can
interpret all coefficient estimates as correlations. Similar to previous tables, I report two
t-statistics using Newey and West (1987) standard errors with (1) 36 lags (t-stat I) and (2)
12 lags (t-stat II). Since macro variables are sometimes extremely persistent, I also report
dependent variables' first-order autoregressive (AR(1)) coefficients (ρ). For example, if ρ is
close to 1, the economic variable is virtually a random walk process, making all statistical
inference based on asymptotic normality invalid.

In Panel (A), I regress each nontradable economic variable Y_t on $\mathcal{M}_t^{HF}$ and $\mathcal{M}_t^{missing}$,
while in Panel (B), I regress Y_t on $\mathcal{M}_t^{LF}$ and $\mathcal{M}_t^{unpriced}$. Since past literature often uses
canonical PCs, which are almost identical to HF-PCs, it is intriguing to compare the co-
efficient estimates of $\mathcal{M}_t^{HF}$ and $\mathcal{M}_t^{LF}$. Also, if their coefficients are hugely different, the
missing-SDF $\mathcal{M}_t^{missing}$ or the unpriced-SDF $\mathcal{M}_t^{unpriced}$ should explain the difference. In Ta-
ble 1.5, I estimate the risk prices of factors under the prior Sharpe ratio 0.4. I consider a
robustness check by adopting another prior Sharpe ratio 0.5, and Table 1.A.6 presents the
related results. Overall, the results in Table 1.5 are not considerably different from those in
1.A.6.

Quarterly Real Consumption Growth

First, I consider the textbook CCAPM, which predicts a negative correlation between
consumption growth and SDFs. However, past research (e.g., Kan and Zhang (1999a))
find that the quarterly real nondurable consumption growth, commonly used in the past
literature, is not strongly correlated with test assets. In other words, the risk premium of
consumption risk is zero, contradicting the standard textbook prediction.

Column (1) in Table 1.5 presents the coefficient estimate of quarterly real consumption
growth. Theoretically, when the consumption growth is low in bad economic states, marginal
utility of investors, proxied by the SDF, should be higher, so economic theory predicts
negative correlations. However, the correlation between consumption growth and the HF-
SDF is only marginally negative, with a t-statistic of -0.4. However, Panel (B) shows
that the consumption growth is closely associated with the LF-SDF, with a much higher
correlation coefficient -0.15. The t-statistic (optimal lags) equals -1.2, so I cannot reject

Table 1.5: Economic Fundamentals related to HF- vs. LF-SDFs

$Y_t:$	C_t^{nd}	C_{t+1}^{nd}	GDP_t	GDP_{t+1}	N_t^{CF}	N_t^{DR}	HKM_t^{ntr}	HKM_t^{tr}	VXO_t^{ar1}	BW_t^{ar1}
	Panel (A): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{HF} + \beta_2 \mathcal{M}_t^{missing} + \epsilon_t$									
$\mathcal{M}_t^{HF}$	-0.037	0.148	-0.175	0.037	-0.123	-0.299	-0.238	-0.293	0.238	-0.147
t-stat I	(-0.410)	(1.396)	(-0.761)	(0.347)	(-0.909)	(-2.926)	(-2.423)	(-2.484)	(2.525)	(-2.540)
t-stat II	(-0.412)	(1.396)	(-0.815)	(0.387)	(-0.805)	(-2.557)	(-2.315)	(-2.410)	(2.557)	(-2.716)
$\mathcal{M}_t^{missing}$	-0.218	-0.229	-0.180	-0.223	-0.112	-0.043	0.136	0.168	-0.040	-0.013
t-stat I	(-1.846)	(-3.953)	(-1.583)	(-2.203)	(-1.425)	(-0.610)	(1.467)	(1.581)	(-0.826)	(-0.171)
t-stat II	(-1.940)	(-3.953)	(-1.759)	(-2.484)	(-1.431)	(-0.591)	(1.460)	(1.572)	(-0.800)	(-0.171)
	Panel (B): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{LF} + \beta_2 \mathcal{M}_t^{unpriced} + \epsilon_t$									
$\mathcal{M}_t^{LF}$	-0.147	0.004	-0.244	-0.088	-0.164	-0.272	-0.122	-0.149	0.175	-0.130
t-stat I	(-1.156)	(0.044)	(-1.003)	(-0.780)	(-1.425)	(-2.925)	(-1.420)	(-1.550)	(2.125)	(-1.922)
t-stat II	(-1.139)	(0.044)	(-1.063)	(-0.833)	(-1.426)	(-3.119)	(-1.564)	(-1.631)	(2.249)	(-1.921)
$\mathcal{M}_t^{unpriced}$	0.165	0.272	0.059	0.209	0.025	-0.131	-0.246	-0.303	0.166	-0.071
t-stat I	(3.023)	(3.513)	(0.733)	(2.300)	(0.201)	(-1.399)	(-2.328)	(-2.368)	(2.556)	(-1.074)
t-stat II	(2.634)	(3.410)	(0.939)	(2.683)	(0.195)	(-1.223)	(-2.140)	(-2.270)	(2.325)	(-1.081)
ρ	0.153	0.153	0.352	0.352	-0.189	-0.108	0.061	0.104	0.116	0.105
R_{adj}^2	4.91%	7.43%	6.30%	5.11%	2.76%	9.13%	7.51%	11.38%	5.80%	2.19%
Sample size	112	111	112	111	338	338	338	338	338	326

This table reports the results of the regressions in which I regress eight economic variables on different components of SDFs. The dependent variables include (1) and (2) current and one-period ahead quarterly real nondurable consumption growth, (3) and (4) current and one-period ahead quarterly real GDP growth, (5) cash-flow news, (6) discount-rate news, (7) nontradable intermediary factor, (8) tradable intermediary factor, (9) the AR(1) shock in VXO index, and (10) the AR(1) shock in investor (Baker and Wurgler (2006)) sentiments. The SDFs are composed of the first seven principal components of asset returns, and their risk prices are estimated under the prior Sharpe ratio equal to 0.4. I standardize both dependent and independent variables so that readers can interpret all coefficient estimates as correlations. I report two t-statistics using Newey and West (1987) standard errors with (1) 36 lags (t-stat I) and (2) 12 lags (t-stat II). In addition, I report dependent variables' first-order autocorrelation coefficients (ρ). The monthly (quarterly) out-of-sample runs from November 1991 to December 2019 (Q1 1992 – Q4 2019).

the null hypothesis of zero correlation.

What can explain this huge difference? The missing-SDF is the key, and its correlation with consumption growth is -0.22 and statistically significant. Suppose that a state variable predicts future consumption growth and portfolio returns, but it is relatively persistent. The standard PCA, which is virtually equivalent to HF-PCA, fails to identify the components related to this state variable; hence, the HF-SDF is not correlated with the consumption growth. However, the focus on the long-horizon asset returns recovers the identification of consumption risk. In short, the $\mathcal{M}_t^{missing}$ is significantly and negatively correlated with consumption growth, which implies a positive risk premium of consumption risk.

Campbell (1999) suggest an alternative timing convention to calculate the correlation between consumption growth and asset returns. Specifically, the consumption during a quarter is a flow. If we think of the consumption observed at quarter t as the consumption level at the beginning of this quarter, we should use the next-period consumption to compute

consumption growth at quarter t. In other words, I should estimate the correlation between C_{t+1}^{nd} and $\mathcal{M}_t$. Column (2) in Table 1.5 displays the results. The missing-SDF is still significantly correlated with the next-period consumption growth, and its t-statistic is around -4.

Quarterly Real GDP Growth

Liew and Vassalou (2000) show that HML and SMB positively predict future real GDP growth. Motivated by this finding, I study whether the quarterly real GDP growth correlates with SDFs. Intriguingly, only the coefficient estimate of $\mathcal{M}_t^{missing}$ is significantly negative, with a t-statistic around -1.8 (see t-statistic II) in Column (3) of Table 1.5. Although the correlation coefficients of both $\mathcal{M}_t^{HF}$ and $\mathcal{M}_t^{LF}$ are not trivial, around -0.2, their standard errors are so enormous that I cannot reject the null hypothesis of zero correlation. The high autocorrelation coefficient, equal to 0.35, and small sample size, potentially contribute to the notable estimation uncertainty.

In Column (4), I explore whether the SDFs can predict GDP growth in the next quarter. While both the HF-SDF and LF-SDF have almost zero prediction power, the missing-SDF negatively predicts the GDP growth. The coefficient estimate is -0.22 and has a t-statistic around -2. In other words, if the MVE portfolio implied by the missing-SDF experiences a negative return (or the missing-SDF increases) at quarter t, it predicts that the future GDP growth will decrease over the next quarter. This finding indicates that persistent state variables contained in asset returns can predict GDP growth. The missing-SDF captures this persistent predictor, so it is closely related to GDP growth.

Cash-Flow vs. Discount-Rate News

Campbell and Vuolteenaho (2004) decompose the shocks in the market portfolio into cash-flow news and discount-rate news.[16] In their language, cash-flow news is bad, for investors' wealth decreases and the future investment opportunity set is unchanged. On the contrary, discount-rate news is good since future investment opportunities, quantified by expected returns, improve.

Campbell and Vuolteenaho (2004) include four state variables: (1) the excess log return on the market, (2) the term yield spread that is the yield difference between 10-year and short-term constant-maturity taxable bonds, (3) the pricing-earnings ratio (PE) from Shiller (2000), and (4) the small-stock value spread that is the difference between $\log(\frac{BE}{ME})$ of the

---

[16]Campbell and Vuolteenaho (2004) estimate cash-flow and discount-rate news using a first-order VAR model: $\boldsymbol{Z_{t+1}} = \boldsymbol{a} + \boldsymbol{\Gamma Z_t} + \boldsymbol{u_{t+1}}$, where $\boldsymbol{Z_{t+1}}$ is a m-by-1 state vector with the log market excess return as its first entry. After estimating the VAR(1) model via OLS, they define cash-flow and discount-rate news as follows: $N_{t+1}^{CF} = \left[\boldsymbol{e_1}^\top + \boldsymbol{e_1}^\top \rho\boldsymbol{\Gamma}(\boldsymbol{I_m} - \rho\boldsymbol{\Gamma})^{-1}\right]\boldsymbol{u_{t+1}}$ and $N_{t+1}^{DR} = \boldsymbol{e_1}^\top \rho\boldsymbol{\Gamma}(\boldsymbol{I_m} - \rho\boldsymbol{\Gamma})^{-1}\boldsymbol{u_{t+1}}$, where $\boldsymbol{e_1}^\top = (1, 0, \ldots, 0)^\top$.

small high-book-to-market portfolio and $\log(\frac{BE}{ME})$ of the small low-book-to-market portfolio. However, the term yield spread that they used originally is no longer updated, so I replace it with the difference between the log yield on the 10-year U.S. Constant Maturity Bond and the log yield on the three-month U.S. Treasury bill, as in Campbell, Giglio, and Polk (2013). Campbell, Giglio, and Polk (2013) additionally include as a state variable the default spread (DEF), defined as the difference between the log yield on Moody's BAA and AAA bonds.

In the monthly data, I find that the default spread does not predict the market portfolio, so I stick to the four-state-variable VAR regression in Campbell and Vuolteenaho (2004). Moreover, I estimate the VAR model using monthly data from December 1928 to December 2019 and extract cash-flow and discount-rate news from November 1991 to December 2019. In Columns (5) and (6) of Table 1.5, I report the correlation coefficients between SDFs and two sources of shocks in the market portfolio.

Cash-flow news is negatively correlated with $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{LF}$, but none of their coefficients is statistically significant. The LF-SDF is slightly more relevant to cash-flow news than the HF-SDF. Overall, the statistical power of these tests is not strong enough to make decisive conclusions.

Differently, discount-rate news is strongly and negatively correlated with both $\mathcal{M}_t^{HF}$ and $\mathcal{M}_t^{LF}$, with correlation coefficients around -0.3 and t-statistics around -3. In other words, discount-rate news earns a significantly positive risk premium. The time-series R^2, equal to 9% in column (6), is also considerably higher than in the regression of cash-flow news.

As a robustness check, I also estimate cash-flow and discount rate news by including the default spread as the fifth state variable in the VAR(1) regression. Columns (1) and (2) of Table 1.A.5 present similar results. The coefficient estimates in the regression are almost unchanged. In short, not only does discount-rate news explain most of the time-series variation in return news (see Campbell (1990)), but it is also more critical than cash-flow news as a source of economic risk for which investors in stock markets require risk compensation.

Intermediary Factor

He, Kelly, and Manela (2017) show that their intermediary factor can price many asset classes and conclude that financial intermediaries are important marginal investors and key to understanding asset prices. Their paper defines the intermediary capital ratio as the aggregate value of market equity divided by aggregate market equity plus aggregate book debt of primary dealers active. The intermediary capital risk factor, HKM_t^{ntr}, is the AR(1) innovation to the market-based capital ratio of primary dealers. He, Kelly, and Manela (2017) also define a tradable intermediary factor, denoted as HKM_t^{tr}. As predicted by intermediary asset pricing theory, such as He and Krishnamurthy (2013), the SDF of financial

intermediaries is higher when a negative shock hits them, so the correlation between their SDF and the intermediary factor is expected to be negative.

Column (7) in Table 1.5 studies the nontradable intermediary factor. Panel (A) shows that the HF-SDF has a significant negative correlation with HKM_t^{ntr}, equal to around -0.24 with a t-statistic of -2.4. However, the LF-SDF has a smaller correlation (-0.12) in absolute terms, and its t-statistic is only -1.4. Hence, I am on the edge of rejecting the null hypothesis of zero correlation between HKM_t^{ntr} and $\mathcal{M}_t^{LF}$. More surprisingly, $\mathcal{M}_t^{missing}$ is positively associated with the intermediary factor, which implies that it hedges the intermediary risk in the HF-SDF. In other words, the intermediary factor cannot explain the high Sharpe ratio of $\mathcal{M}_t^{missing}$, or makes it even more puzzling.

Column (8) in Table 1.5 runs similar regressions but uses the tradable intermediary factor. The observations are largely compatible with those in Column (7). I also report the correlation between SDFs and quarterly intermediary factors in Columns (3) and (4) of Table 1.A.5, and the empirical patterns are virtually identical.

Even though I do not discover a significant correlation coefficient between the LF-SDF and the intermediary factor, it does not imply that financial intermediaries do not play an important role in understanding asset prices. On the one hand, the intermediary factor is significantly correlated with the HF-SDF, especially its unpriced component, so the intermediary factor, at the very least, drives the common variations in asset returns. On the other hand, the risk premium of the nontradable factor in He, Kelly, and Manela (2017) is not statistically different from zero in the monthly regression of stock portfolios, which is consistent with the insignificant correlation between the LF-SDF and the intermediary factor. Also, financial intermediaries should be more important in other asset markets, such as CDS and derivative markets, in which they get involved actively.

Jump Risk

Investors require compensation for bearing downside risk (e.g., Ang, Chen, and Xing (2006)). While there is no consensus on which variable represents downside risk, I use the VXO index as the proxy, since it is commonly accepted as the fear index in the industry. Specifically, the VXO index is the risk-neutral entropy of the market excess return and is particularly sensitive to the left tail of the return distribution.

It is problematic to regress the VXO index on SDFs. The VXO index is highly persistent, with an AR(1) coefficient of around 0.9, so standard errors of coefficient estimates are enormous. In other words, the high persistence makes the statistical inference almost impossible in small samples. Hence, I extract the shock in the VXO index via an AR(1) regression, $Y_t = a + \rho Y_{t-1} + n_t$, and jump risk is defined as the AR(1) innovation $Y_t - a - \rho Y_{t-1}$.

Column (9) of Table 1.5 reports the correlation between SDFs and jump risk. Both the HF-SDF and LF-SDF have significantly positive correlations $(0.18 - 0.24)$ with jump risk. Interestingly, coefficient estimates of the HF-SDF and LF-SDF in Table 1.5 are similar to those in Table 1.A.5, in which I regress the original VXO index on SDFs. Hence, it can increase the power of statistical tests to focus on the much less persistent AR(1) innovation.

Intuitively, when investors are particularly fearful, the SDFs, proxying for their marginal utility functions, are likewise high. In other words, investors are willing to pay a positive risk premium to hedge jump risk. Nevertheless, the missing-SDF is almost unrelated to jump risk, so jump risk does not explain the risk premium of $\mathcal{M}_t^{missing}$.

Investor Sentiment

Rational economic models cannot always explain economic phenomena that we observe in the real world, such as the tech stock bubble in the late 1990s and the housing bubble in 2008. Instead, investor sentiments are also essential in understanding asset prices. For instance, De Long, Shleifer, Summers, and Waldmann (1990) build a theoretical model in which the presence of noise traders with stochastic beliefs can create a source of risk that requires a positive risk premium. Kozak, Nagel, and Santosh (2018) show that if the demand from sentiment investors drives a large proportion of asset returns' common variations, their demand shocks, or investor sentiments, should enter the SDF as well.

Motivated by these papers, I go on to explore how SDFs extracted purely from asset returns correlate with the proxy for investor sentiments. First of all, I use the BW sentiment index in Baker and Wurgler (2006), which estimate the first principal component of six variables: the closed-end fund discount, the NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium.

The AR(1) coefficient of the BW index is close to 1, so I extract its AR(1) shock following the same steps as for the VXO index. The last column of Table 1.5 demonstrates that the HF-SDF is negatively correlated with the investor sentiment, with a t-statistic of around -2.6. The LF-SDF has a similar coefficient estimate (-0.13), and its t-statistic is about -1.9. Column (6) of Table 1.A.5 reports the correlation between SDFs and the original BW sentiment. Even though the magnitudes of coefficient estimates are extremely similar, their t-statistics are much lower due to the persistence of the BW sentiment index. Last but not least, the missing-SDF is virtually unrelated to the BW sentiment index. Overall, Table 1.5 indicates that only macro risk, such as consumption and GDP growth, can potentially explain the risk premium of $\mathcal{M}_t^{missing}$.

Huang, Jiang, Tu, and Zhou (2015) modify the BW sentiment index using the partial least squares method. Precisely, they extract the most important component that can simul-

taneously predict the future market return and explain time-series variations of the original six proxies. I call their sentiment index HJTZ sentiment. Columns (7) and (8) of Table 1.A.5 show that only the HF-SDF is weakly correlated with the AR(1) shock of HJTZ sentiment. On the contrary, its correlation with the LF-SDF is only -0.07, compared to -0.14 for BW sentiment. Overall, the HJTZ sentiment is less correlated with SDFs than the BW sentiment.

Summary

The findings in Table 1.5 deepen our understanding of the economics behind the factor zoo. One potential reason for the existence of the factor zoo is that factors are noisy proxies for economic fundamentals and therefore do not span each other. For example, Liew and Vassalou (2000) report that both the value and size factors can predict GDP growth, but they are never comprehensive predictors and cannot replace each other. This paper further shows that the importance of economic risks varies at different frequencies. In Table 1.5, I confirm that a sparse LF-SDF, earning a nearly optimal Sharpe ratio, captures two elements: (1) the first one is perfectly linear in the HF-SDF and almost uncorrelated over time, which is statistically associated with discount-rate news of the market excess return, the intermediary factor, jump risk, and investor sentiment, whereas (2) the second one is neglected by the HF-SDF and captures some persistent state variables, reflecting business-cycle risks related to consumption and GDP growth.

1.4 Additional Robustness Checks

In this part, I present a few of the robustness checks of Section 1.3. Specifically, I investigate whether the sparsity of the LF-SDF is robust (1) when I consider only long-short portfolios, (2) if I impose the CAPM, or (3) if I slightly modify the definition of the LF interval.

1.4.1 39 Long-Short Portfolios

Until now, I have included long and short portfolios separately for each firm characteristic in Section 1.3. However, many papers in cross-sectional asset pricing literature handle long-short portfolios, such as the size factor in FF3. To confirm the robustness of the main results, I further analyze long-short portfolios of 39 firm characteristics.

Figure 1.11 plots the OOS Sharpe ratio of PCA, HF-PCA and LF-PCA under prior Sharpe ratios $\in \{0.4, 0.5\}$. A more comprehensive heat-map is in Figure 1.A.8. First, the maximal Sharpe ratio is around 0.35, slightly less than that in the cross-section of 78 portfolios. Second, I can still discern a parsimonious factor model composed of low-frequency PCs. Particularly, a six-factor LF-PC model delivers an optimal OOS Sharpe ratio, and this

finding is robust across a wide range of L_2-penalty, as I observe in Figure 1.A.8. On the contrary, latent-factor models constructed by canonical and high-frequency PCs are dense, consistent with my observations in Section 1.3.4. Overall, the sparsity of LF-PC models is robust in the cross-section of 39 long-short portfolios. In the following robustness analyses, I stick to the original cross-section of 78 test assets.

(a) $SR_{prior} = 0.4$ **(b)** $SR_{prior} = 0.5$
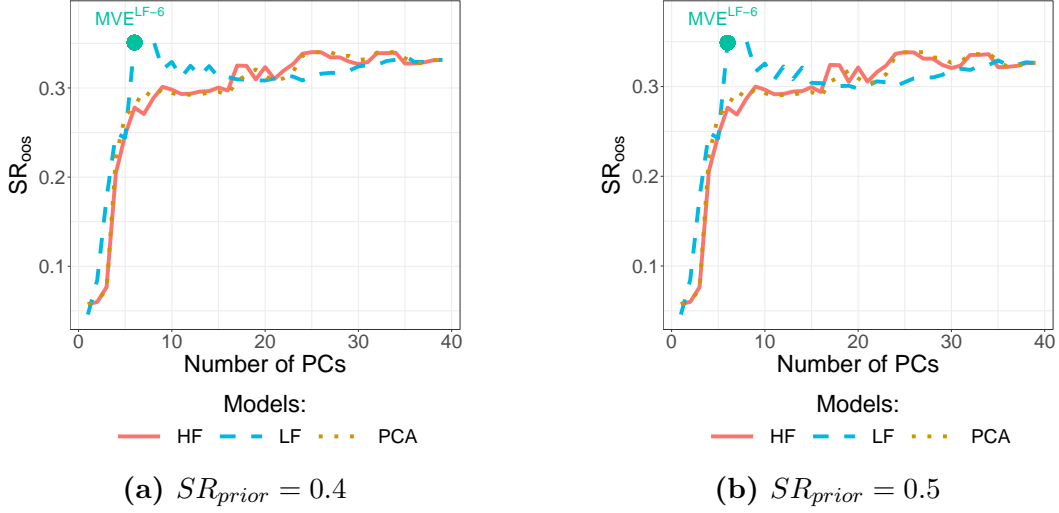
Figure 1.11: Zoom in OOS Sharpe ratio of 39 long-short portfolios, $SR_{prior} \in \{0.4, 0.5\}$

This graph zooms in the OOS Sharpe ratio of PCA, HF-PCA, and LF-PCA. The cross-section of test assets is 39 long-short portfolios. Different from figure 1.A.8, this figure shows the estimates using two prior Sharpe ratios, $SR_{prior} \in \{0.4, 0.5\}$.

1.4.2 Imposing the CAPM

Since the introduction of the CAPM by Sharpe (1964) and Lintner (1965), the market factor has become the most influential factor in cross-sectional asset pricing. For example, Barillas and Shanken (2018a) use the market factor as the anchor to compare a few famous factor models via Bayes factors. Also, Kozak, Nagel, and Santosh (2020) extract the CAPM α of 50 long-short anomalies and estimate other systematic factors via the eigendecomposition of the CAPM $\boldsymbol{\alpha}$. Following the past literature, I turn to study the CAPM $\boldsymbol{\alpha}$.

Here is my empirical strategy. First, I regress $\boldsymbol{R_t}$ on the market factor using the in-sample observations: $\boldsymbol{R_t^{IN}} = \boldsymbol{\beta_0^{IN}} + \boldsymbol{\beta_m^{IN}} R_{mt}^{IN} + \boldsymbol{e_t^{IN}}$, and the CAPM $\boldsymbol{\alpha}$ is defined as: $\boldsymbol{\alpha_t^{IN}} = \boldsymbol{R_t^{IN}} - \boldsymbol{\beta_m^{IN}} R_{mt}^{IN}$. Next, I decompose the covariance matrix of $\boldsymbol{\alpha_t^{IN}}$ into frequency-dependent components and estimate frequency-dependent PCs as in definition 1.1. When I mention a K-factor model, the SDF consists of the market factor and the first K PCs of $\boldsymbol{\alpha_t^{IN}}$: $\mathcal{M}_t = 1 - b_m(R_{mt}^{IN} - \mu_m) - \boldsymbol{b_F^\top}(\boldsymbol{F_t} - \boldsymbol{\mu_F})$. Finally, I estimate risk prices $(b_m, \boldsymbol{b_F^\top})^\top$ using the objective function in equation (1.16). To evaluate the OOS performance, I use the in-sample estimate

of market loadings $\boldsymbol{\beta^{IN}}$ to construct the OOS CAPM $\boldsymbol{\alpha}$: $\boldsymbol{\alpha_t^{OOS}} = \boldsymbol{R_t^{OOS}} - \boldsymbol{\beta^{IN}} R_{mt}^{OOS}$. Then I construct the OOS latent factors and MVE portfolio as before.

Figure 1.12 plots the Sharpe ratio of the OOS MVE portfolio following the procedures described in the previous paragraph. The prior monthly Sharpe ratio is set to be 0.4 in Panel (a) and 0.5 in Panel (b). Like the benchmark case, the MVE portfolio consisting of the market factor and another 6 LF-PCs can earn a virtually optimal OOS Sharpe ratio, around 0.37 monthly. HF-PCA, however, needs much more PCs, literally more than 60, to reach the highest point, and PCA has almost an identical pattern. In short, a seven LF factor model can nearly span the whole asset space in the out-of-sample.

(a) $SR_{prior} = 0.4$ (b) $SR_{prior} = 0.5$
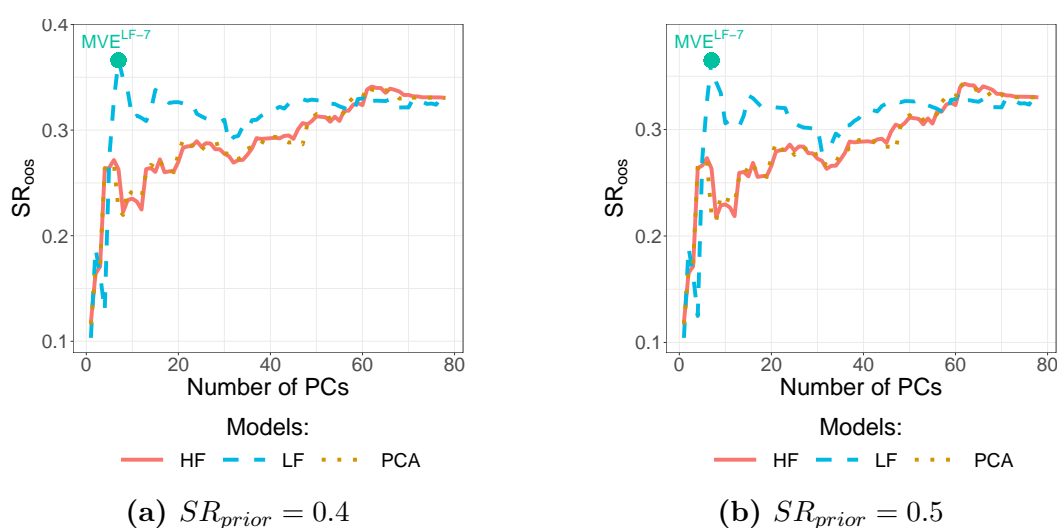
Figure 1.12: Imposing CAPM: OOS Sharpe ratio of 78 portfolios, $SR_{prior} \in \{0.4, 0.5\}$

This graph zooms in the OOS Sharpe ratio of PCA, HF-PCA, and LF-PCA after imposing the CAPM. This figure shows the estimates using two prior Sharpe ratios, $SR_{prior} \in \{0.4, 0.5\}$.

1.4.3 Alternative Cutoffs of the LF Interval

Earlier, I define the low-frequency component of an asset return as the part with a cycle length between 36 and 120 months. This section investigates whether the sparsity of LF factor models is particularly sensitive to alternative cutoffs of the LF intervals.

Panel (a) in Figure 1.13 defines the period of the LF component between 24 and 120 months. The OOS MVE portfolios of HF-PCA and PCA are identical to previous ones. The LF-MVE portfolios, however, are still more parsimonious. For example, The LF-MVE portfolio composed of the first seven LF-PCs earns a monthly Sharpe ratio of 0.35 in the out-of-sample. In addition, Figure 1.A.9 plots the heat-map for LF-PCA, whose pattern is virtually identical to that of Figure 1.4.

Bandi, Chaudhuri, Lo, and Tamoni (2021) decompose the CAPM β into frequency-dependent components, and they discover that only the component in the LF frequency with a period between 32 and 64 months can price conventional Fama-French portfolios. Motivated by their results, I define the period of the LF (HF) component as $\tau^{LF} \in [32, 64]$ months ($\tau^{HF} \in [2, 32)$). Panel (b) in Figure 1.13 shows that the monthly Sharpe ratio of a seven LF factor model is slightly higher than 0.37, whereas I still demand many HF-PCs to span the asset space of 78 test assets in the out-of-sample. In short, the sparsity of the LF-SDF is not sensitive to alternative definitions of the LF interval.

(a) $\tau^{LF} \in [24, 120]$ **(b)** $\tau^{LF} \in [32, 64]$
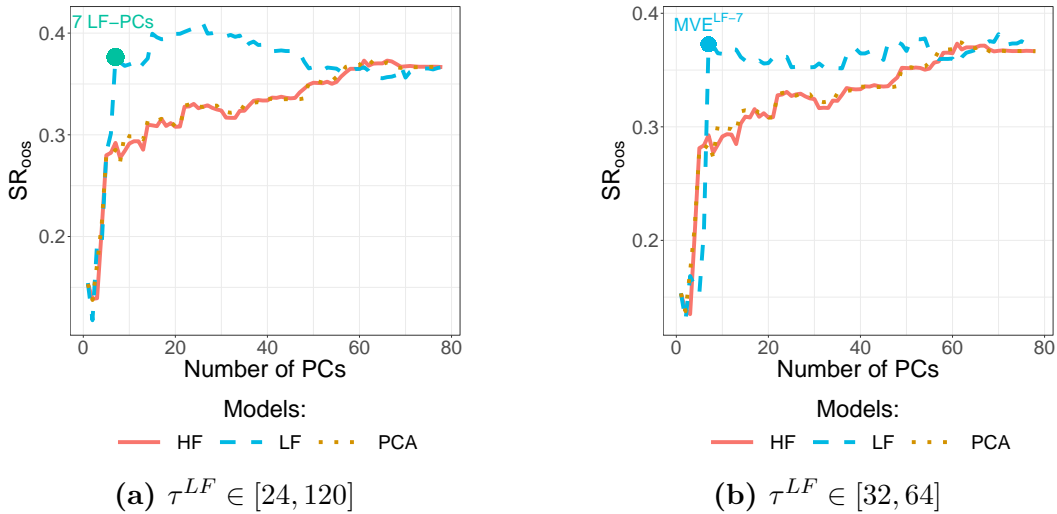
Figure 1.13: Robustness Check: OOS Sharpe ratio of 78 portfolios, $SR_{prior} = 0.4$

This graph shows additional robustness checks of the OOS Sharpe ratio of PCA, HF-PCA, and LF-PCA. I estimate risk prices under the prior Sharpe ratio 0.4. The low-frequency (LF) interval is equal to (1) panel (a): $\tau^{HF} \in [2, 24)$ and $\tau^{LF} \in [24, 120]$, and (2) panel (b): $\tau^{HF} \in [2, 32)$ and $\tau^{LF} \in [32, 64]$.

1.5   Conclusions

I use frequency-dependent risks to dissect the factor zoo and answer fundamental questions about what is salient for cross-sectional asset pricing. As a first step, I propose a new approach to quantify frequency-dependent risks and deliver monthly proxies for short-term and long-term systematic factors. Empirically, the SDF is sparse only in the space of low-frequency systematic factors. An economic interpretation of this finding is that investors are more risk-averse to low-frequency persistent systematic factors that drive a vast majority of long-run movements of asset returns, probably because they have long investment horizons or Epstein-Zin preference that imposes a considerable risk aversion to the long-run uncertainty. Hence, the first few largest LF latent factors capture almost the entire Sharpe ratio of the true SDF.

In addition, I confirm that none of the celebrated sparse factor models, such as the Fama-French three-factor model, or HF-SDF can explain the LF-SDF. At the same time, the LF-SDF can span the HF-SDF. Therefore, I conjecture that the SDF composed of the first several low-frequency factors is the proper benchmark SDF.

Furthermore, my paper deepens our understanding of the economics behind the factor zoo. It is common to use the largest several canonical PCs to construct the SDF. This SDF, virtually identical to the HF-SDF, is almost uncorrelated over time and captures economic risks related to discount-rate news of the market excess return, intermediary factors, jump risk, and investor sentiment. However, the HF-SDF still ignores an economically important component of the LF-SDF. This missing component commands a sizable monthly Sharpe ratio of about 0.2 and displays a persistent conditional dynamic, as the variance ratio test shows. More importantly, it reflects only business-cycle risks related strongly to consumption and GDP growth, and it can also predict consumption and GDP growth over the next quarter.

Traditional macro-finance models emphasize persistent conditional information and use them to rationalize the asset pricing puzzles. What I observe in this paper confirms that asset returns indeed contain useful conditional information related to macro variables, but they can be identified only at low frequencies. At the same time, the tail risk, behavioral finance, and intermediary asset pricing models are also essential in understanding asset returns, but they are more relevant in short horizons.

Appendices

1.A.1 Additional Details on Frequency Domain Analysis

1.A.1.1 Spectral Representation Theorem

Theorem 1.A.1 (Spectral Representation Theorem, Hannan (2009)) *Suppose x_t is a mean-zero covariance stationary process, with the spectral distribution function $F(\omega)$ such that its auto-covariance function $\Sigma_x(h)$ can be expressed as*

$$\Sigma_x(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{2\pi i \omega h\} dF(\omega),$$

where $F(\omega)$ is non-decreasing, $F(-\frac{1}{2}) = 0$ and $F(\frac{1}{2}) = \Sigma_x(0)$. Then there exists a complex-value stochastic process $z(\omega)$, $\omega \in [-\frac{1}{2}, \frac{1}{2}]$, having stationary uncorrelated increments, such that x_t can be written as the stochastic integral

$$x_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{-2\pi i \omega t\} dz(\omega),$$

where $\mathrm{Var}[z(\omega_2) - z(\omega_1)] = F(\omega_2) - F(\omega_1)$. Furthermore, the Spectral Representation Theorem can be extended to multivariate case.

Suppose that $x_t(\omega)$ satisfies the differential equation: $x_t(\omega)d\omega = \exp\{-2\pi i \omega t\} dz(\omega)$. The Spectral Representation Theorem implies that $x_t(\omega)$ is uncorrelated at different frequencies, and x_t is decomposed as an equally weighted average of $x_t(\omega)$, i.e., $x_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} x_t(\omega)d\omega$. Therefore, I can represent the variance of x_t as $\mathrm{Var}(x_t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{Var}[x_t(\omega)]d\omega$, where $\mathrm{Var}[x_t(\omega)]$ is the contribution from the frequency-ω component.

Suppose that $\boldsymbol{X_t}$ is a two-dimensional time series, for example, $\boldsymbol{X_t} = (x_{1t}, x_{2t})^\top$, with the auto-covariance matrix $\boldsymbol{\Sigma_X}(h)$. According to the Spectral Representation Theorem, the cross-spectrum $f_{x_1,x_2}(\omega)$ that satisfies $dF_{12}(\omega) = f_{x_1,x_2}(\omega)d\omega$ can be interpreted as the covariance between the frequency-ω components of $x_{1,t}$ and $x_{2,t}$. Next, I will consider a linear transformation of $\boldsymbol{X_t}$. Let a and b be arbitrary real numbers, and define y_t as $y_t = ax_{1,t} + bx_{2,t}$.

The spectral density function of y_t is

$$
\begin{aligned}
f_y(\omega) &= \sum_{h=-\infty}^{\infty} \mathrm{Cov}(y_{t+h}, y_t) \exp\{-2\pi ih\omega\} \\
&= \sum_{h=-\infty}^{\infty} \mathrm{Cov}(ax_{1,t+h} + bx_{2,t+h}, ax_{1,t} + bx_{2,t}) \exp\{-2\pi ih\omega\} \\
&= \sum_{h=-\infty}^{\infty} \left[a^2 \mathrm{Var}(x_{1,t}) + b^2 \mathrm{Var}(x_{2,t}) + ab\mathrm{Cov}(x_{1,t+h}, x_{2,t}) + ab\mathrm{Cov}(x_{1,t}, x_{2,t+h}) \right] \exp\{-2\pi ih\omega\} \\
&= a^2 f_{x_1}(\omega) + b^2 f_{x_2}(\omega) + ab f_{x_1,x_2}(\omega) + ab f_{x_2,x_1}(\omega) \\
&= a^2 f_{x_1}(\omega) + b^2 f_{x_2}(\omega) + 2ab\mathcal{R}[f_{x_1,x_2}(\omega)],
\end{aligned}
$$

where the last equality makes use of the fact that $f_{x_1,x_2}(\omega) = f_{x_2,x_1}(-\omega)$ and $f_{x_2,x_1}(\omega) + f_{x_2,x_1}(-\omega) = 2\mathcal{R}[f_{x_1,x_2}(\omega)]$. There are two implications. First, I can interpret the real part of the cross-spectrum as the covariance between the frequency-ω components of $x_{1,t}$ and $x_{2,t}$. Second, I need to focus only on the real part of the cross-spectrum. This paper aims to extract PCs at different frequencies. For example, the largest PC chooses a unitary linear transformation of $\boldsymbol{X_t}$ such that its variance is maximized.

1.A.1.2   Discrete Fourier Transform (DFT)

Given data $\boldsymbol{R_1}, \ldots, \boldsymbol{R_T}$, DFT and its inverse (**IDFT**) are defined as

$$
\boldsymbol{d}(\boldsymbol{\omega_j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{R_t} \exp\{-2\pi i\omega_j t\}, \ \omega_j = \frac{j}{T}, \ j = 0, 1, \ldots, T-1, \tag{27}
$$

$$
\boldsymbol{R_t} = \frac{1}{\sqrt{T}} \sum_{j=0}^{T-1} \boldsymbol{d}(\boldsymbol{\omega_j}) \exp\{2\pi i\omega_j t\}, \ t = 1, \ldots, T. \tag{28}
$$

Let's define the frequency-ω_j component of asset returns: $\boldsymbol{R_t}(\omega_j) = \frac{1}{\sqrt{T}} \boldsymbol{d}(\boldsymbol{\omega_j}) \exp\{2\pi i\omega_j t\}$. A distinguishing feature of the aforementioned decomposition is that two components from distinct frequencies are uncorrelated by construction; that is, $\mathbf{Cov}_T \left(\boldsymbol{R_t}(\omega_j) \boldsymbol{R_t}(\omega_k)^\top \right) = \boldsymbol{0}_{N \times N}$ if $j \neq k$, or $\hat{\boldsymbol{f}}_{\boldsymbol{R}}(\omega_j)$ if $j = k$. The intuition is that DFT decomposes $\boldsymbol{R_t}$ into orthogonal frequency-dependent parts.

Moreover, $\boldsymbol{d}(\omega_j)\boldsymbol{d}(\omega_j)^\star = \sum_{h=-(n-1)}^{n-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}}(h) \exp\{-2\pi i\omega_j h\} = \hat{\boldsymbol{f}}_{\boldsymbol{R}}(\omega_j)$, where $\boldsymbol{d}(\omega_j)^\star$ is the conjugate transpose operation of $\boldsymbol{d}(\omega_j)$. Therefore, we can estimate the frequency density matrix of asset returns via DFT. In practice, researchers often use a fast Fourier transform (**FFT**) algorithm to compute the transformations in replace of DFT rapidly. Figure 1.A.3

is a simple example of DFT.

1.A.2   Proofs

1.A.2.1   Proof of Proposition 1.1

Since e_{t+1} is conditionally independent, its conditional expectation is always zero: For $h > 0$, $\mathbb{E}[e_{t+h} \mid \mathbb{I}_t] = 0$, where $\mathbb{I}_t$ denotes the conditional information at time t. If $h > 0$, the auto-covariance matrix of e_{t+1} is

$$\Sigma_e(h) = \mathbb{E}(e_{t+h} e_t^\top) = \mathbb{E}[\mathbb{E}(e_{t+h} e_t^\top \mid \mathbb{I}_t)] = \mathbb{E}[\mathbb{E}(e_{t+h} \mid \mathbb{I}_t) e_t^\top] = 0_{N \times N},$$

which implies that the spectral density matrix of e_{t+1} is

$$f_e(\omega) = \sum_{h=-\infty}^{\infty} \Sigma_e(h) \exp\{-2\pi i h \omega\} = \Sigma_e(0) = \Sigma_e.$$

Therefore, even though e_{t+1} and f_{t+1} can follow stochastic volatility processes, their spectral density matrices are constant across frequencies.

In addition, F_t and e_t are orthogonal, so I can represent the covariance matrix of R_t as $\Sigma_R = \beta \Sigma_F \beta^\top + \Sigma_e$. Suppose that $f_F(\omega)$ is the spectral density matrix of F_t: $\Sigma_F = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_F(\omega) d\omega$. This implies the following spectral decomposition of Σ_R:

$$\Sigma_R = \int_{-\frac{1}{2}}^{\frac{1}{2}} \beta f_F(\omega) \beta^\top d\omega + \Sigma_e = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\beta f_F(\omega) \beta^\top + \Sigma_e \right] d\omega.$$

Due to the uniqueness of the spectral density matrix, the spectral density matrix of R_t is $f_R(\omega) = \beta f_F(\omega) \beta^\top + \Sigma_e$. Similarly, I show that $f_F(\omega) = \Sigma_f + \Phi_X f_X(\omega) \Phi_X^\top$. Therefore, I rewrite the spectral density matrix of R_t as follows:

$$f_R(\omega) = \beta \Sigma_f \beta^\top + \Sigma_e + \beta_X f_X(\omega) \beta_X^\top, \quad \beta_X = \beta \Phi_X.$$

1.A.2.2 Proof of Proposition 1.2

I can derive the unconditional variance of the linear SDF as follows:

$$
\begin{aligned}
\operatorname{Var}(\mathcal{M}_{t+1}) &= \boldsymbol{b}^\top \operatorname{Var}(\boldsymbol{f_{t+1}})\boldsymbol{b} + \boldsymbol{b}^\top \boldsymbol{\Phi_X} \operatorname{Var}(\boldsymbol{X_t})\boldsymbol{\Phi_X^\top}\boldsymbol{b} \\
&= \boldsymbol{b}^\top \operatorname{Var}(\boldsymbol{f_{t+1}})\boldsymbol{b} + \boldsymbol{b_X^\top}\operatorname{Var}(\boldsymbol{X_t})\boldsymbol{b_X} \\
&= \boldsymbol{b}^\top \operatorname{Var}(\boldsymbol{f_{t+1}})\boldsymbol{b} + \sum_{j=1}^{p} b_{X,i}^2 \operatorname{Var}(X_{jt}) \\
&= \boldsymbol{b}^\top \operatorname{Var}(\boldsymbol{f_{t+1}})\boldsymbol{b} + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{j=1}^{p} b_{X,i}^2 f_{X_i}(\omega) d\omega,
\end{aligned}
$$

where the third equality uses the fact that state variables are assumed to be uncorrelated, and the last step uses the spectral decomposition of each state variable X_i. Since the spectral density function is unique, $f_{\mathcal{M}}(\omega) = \boldsymbol{b}^\top \operatorname{Var}(\boldsymbol{f_{t+1}})\boldsymbol{b} + \sum_{j=1}^{p} b_{X,i}^2 f_{X_i}(\omega)$.

1.A.2.3 Derivation of Equation (1.16)

This section derives the objective function in equation (1.16) under a more general distributional assumption for pricing errors and risk prices. I consider only the cross-sectional regression, conditional on the observed expectation and covariance of $\boldsymbol{F_t}$ as follows:

$$
\boldsymbol{\mu_F} = \boldsymbol{\Sigma_F}\boldsymbol{b} + \boldsymbol{\alpha}, \quad \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0_N}, \sigma^2 \boldsymbol{\Sigma_F}).
$$

Therefore, the only unknowns are $\boldsymbol{b}$ and σ^2. Pástor and Stambaugh (2000) and Barillas and Shanken (2018a) also make a similar distributional assumption for $\boldsymbol{\alpha}$. Intuitively, σ^2 reflects investors' uncertainty about mispricing: When σ^2 is close to zero, the asset pricing model is almost correct. In contrast, if σ^2 is infinity, the factor model is useless, as it entirely fails to explain risk premia.

Furthermore, I assign a normal prior for risk prices: $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0_K}, \frac{\psi\sigma^2}{\tau}\boldsymbol{I_K})$, $\tau = \operatorname{Tr}\left[\boldsymbol{\Sigma_F}\right]$, and $\boldsymbol{b}$ is uncorrelated with $\boldsymbol{\alpha}$. Under such a prior distribution, the prior expectation on the squared Sharpe ratio of factor returns implied by the asset pricing model is equal to

$$
\mathbb{E}_{prior}[SR_F^2] = \mathbb{E}_{prior}[\boldsymbol{b}^\top \boldsymbol{\Sigma_F}\boldsymbol{b}] = \sum_{k=1}^{K} \sigma_{F,k}^2 \mathbb{E}_{prior}[b_k^2] = \frac{\psi\sigma^2}{\tau}\operatorname{Tr}\left[\boldsymbol{\Sigma_F}\right] = \psi\sigma^2.
$$

Next, I decompose the expected squared Sharpe ratio of factor returns as follows:

$$\mathbb{E}_{prior}[\boldsymbol{\mu_F}^\top \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\mu_F}] = \mathbb{E}_{prior}[(\boldsymbol{\Sigma_F b} + \boldsymbol{\alpha})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\Sigma_F b} + \boldsymbol{\alpha})]$$
$$= \mathbb{E}_{prior}[\boldsymbol{b}^\top \boldsymbol{\Sigma_F b}] + \mathbb{E}_{prior}[\boldsymbol{\alpha}^\top \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\alpha}]$$
$$= \psi\sigma^2 + N\sigma^2 = (\psi + N)\sigma^2;$$

therefore, $\mathbb{E}_{prior}[\boldsymbol{\mu_F}^\top \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\mu_F}]$ is the sum of $\mathbb{E}_{prior}[\boldsymbol{b}^\top \boldsymbol{\Sigma_F b}]$ and $\mathbb{E}_{prior}[\boldsymbol{\alpha}^\top \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\alpha}]$, where the former is the contribution from the SDF. Also, I derive the expected squared Sharpe ratio of the SDF as follows:

$$\mathbb{E}_{prior}[SR_F^2] = \frac{\psi}{\psi + N} \mathbb{E}_{prior}[\boldsymbol{\mu_F}^\top \boldsymbol{\Sigma_F}^{-1} \boldsymbol{\mu_F}],$$

so a larger ψ implies higher prior Sharpe ratio of the SDF. Under the above assumptions, the posterior distribution of $\boldsymbol{b}$, conditional on $(\boldsymbol{\mu_F}, \boldsymbol{\Sigma_F})$, is

$$p(\boldsymbol{b} \mid \boldsymbol{\mu_F}, \boldsymbol{\Sigma_F}) \propto \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b}) \right\} \exp\left\{ -\frac{\tau}{2\psi\sigma^2}\boldsymbol{b}^\top \boldsymbol{b} \right\}$$
$$\propto \exp\left\{ -\frac{1}{2\sigma^2}\left[(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b})^\top \boldsymbol{\Sigma_F}^{-1}(\boldsymbol{\mu_F} - \boldsymbol{\Sigma_F b}) + \frac{\tau}{\psi}\boldsymbol{b}^\top \boldsymbol{b} \right] \right\}.$$

Now let $v_2 = \frac{\tau}{\psi} = \frac{\tau\sigma^2}{\mathbb{E}_{prior}[SR_F^2]}$. Therefore, the posterior mode of $\boldsymbol{b}$ is the solution to the objective function in equation (1.16).

To compare with Kozak, Nagel, and Santosh (2020), this paper adopts a similar strategy, which assumes $\sigma^2 = \frac{1}{T}$, so $v_2 = \frac{\tau}{\psi} = \frac{\tau}{T \times \mathbb{E}_{prior}[SR_F^2]}$. Last but not least, the assumption of $\sigma^2 = \frac{1}{T}$ changes only the prior Sharpe ratio implied by the SDF. In this paper, I show the empirical results across a wide range of prior Sharpe ratios. More importantly, empirical results are robust when I estimate the model with reasonable prior *monthly* Sharpe ratios, for example, between 0.3 and 0.6.

1.A.3 Additional Tables

Table 1.A.1: Definition of Variables

Variable	Definition	Data Source
R_{mt}	Monthly market excess return	CRSP database
TY_t	Monthly term yield spread, the difference between the log yield on the 10-year U.S. Constant Maturity Bond and the log yield on the three-month U.S. treasury bills	FRED
PE_t	Monthly pricing-earnings ratio (PE) from Shiller (2000)	Robert Shiller's website
VS_t	Small-stock value spread that is the difference in $\log(\frac{BE}{ME})$ between the small high-book-to-market portfolio and the small low-book-to-market portfolio	Ken French's website
DEF_t	Monthly default spread, the difference between the log yield on Moody's BAA and AAA bonds	FRED
C_t^{nd}	Quarterly real nondurable consumption growth per capita	Table 7.1 in BEA
GDP_t	Quarterly real GDP growth per capita	Table 7.1 in BEA
N_t^{CF}	Monthly cash-flow news in Campbell and Vuolteenaho (2004) Four state variables in VAR(1): $(\log(R_{mt}), TY_t, PE_t, VS_t)$	Estimated by this paper
N_t^{DR}	Monthly discount-rate news in Campbell and Vuolteenaho (2004) Four state variables in VAR(1): $(\log(R_{mt}), TY_t, PE_t, VS_t)$	Estimated by this paper
$N_t^{CF,2}$	Monthly cash-flow news in Campbell, Giglio, and Polk (2013) Five state variables in VAR(1): $(\log(R_{mt}), TY_t, PE_t, VS_t, DEF_t)$	Estimated by this paper
$N_t^{DR,2}$	Monthly discount-rate news in Campbell, Giglio, and Polk (2013) Five state variables in VAR(1): $(\log(R_{mt}), TY_t, PE_t, VS_t, DEF_t)$	Estimated by this paper
HKM_t^I	Monthly nontradable intermediary factor in He, Kelly, and Manela (2017)	Author's website
HKM_t^{II}	Monthly tradable intermediary factor in He, Kelly, and Manela (2017)	Author's website
HKM_{qt}^I	Quarterly nontradable intermediary factor in He, Kelly, and Manela (2017)	Author's website
HKM_{qt}^{II}	Quarterly tradable intermediary factor in He, Kelly, and Manela (2017)	Author's website
VXO_t	the VXO index	WRDS database
BW_t	Sentiment index in Baker and Wurgler (2006)	Dashan Huang's website
$HJTZ_t$	Sentiment index in Huang, Jiang, Tu, and Zhou (2015)	Dashan Huang's website
VXO_t^{ar1}	AR(1) shock in VXO_t: $VXO_t - \rho \times VXO_{t-1}$	Estimated by this paper
BW_t^{ar1}	AR(1) shock in BW_t: $BW_t - \rho \times BW_{t-1}$	Estimated by this paper
$HJTZ_t^{ar1}$	AR(1) shock in $HJTZ_t$: $HJTZ_t - \rho \times HJTZ_{t-1}$	Estimated by this paper

Table 1.A.2: 39 Firm Characteristics in Kozak, Nagel, and Santosh (2020)

Category	Characteristics
Reversal	lrrev, strev, indmomrev, indrrev, indrrevlv
Momentum	mom, mom12, indmom, momrev
Value	value, valuem, divp, ep, cfp, sp
Investment	inv, invcap, igrowth, growth, noa
Profitability	prof, roaa, roea, gmargins
Value interaction	valmom, valmomprof, valprof
Trading frictions	ivol, shvol, aturnover
Others	size, price, accruals, ciss, lev, season, sgrowth, nissa, dur

Table 1.A.3: Do celebrated models explain HF and LF risks? ($SR_{prior} = 0.5$)

		Panel (A). $\mathbf{MVE}_t^{HF}$				Panel (B). $\mathbf{MVE}_t^{LF}$			
		7 PCs	8 PCs	9 PCs	10 PCs	7 PCs	8 PCs	9 PCs	10 PCs
CAPM	α	1.08%	1.03%	1.15%	1.18%	1.47%	1.46%	1.45%	1.47%
	t-stat I	(2.98)	(2.49)	(2.91)	(3.03)	(3.66)	(3.84)	(3.83)	(3.88)
	t-stat II	(3.15)	(2.75)	(3.20)	(3.33)	(4.69)	(4.84)	(4.84)	(4.83)
	R_{adj}^2	3.89%	2.29%	0.13%	0.07%	2.79%	2.05%	1.63%	0.83%
FF3	α	0.85%	0.80%	0.95%	0.99%	1.37%	1.37%	1.36%	1.38%
	t-stat I	(4.87)	(3.79)	(4.12)	(4.27)	(4.31)	(4.52)	(4.51)	(4.62)
	t-stat II	(4.79)	(4.05)	(4.44)	(4.58)	(5.75)	(5.90)	(5.92)	(6.00)
	R_{adj}^2	36.19%	33.64%	29.11%	23.11%	8.85%	7.56%	6.53%	6.19%
Carhart4	α	0.47%	0.41%	0.59%	0.61%	0.92%	0.91%	0.90%	0.92%
	t-stat I	(3.28)	(2.57)	(3.01)	(3.10)	(3.66)	(3.80)	(3.78)	(3.83)
	t-stat II	(2.89)	(2.37)	(3.02)	(3.09)	(4.43)	(4.50)	(4.49)	(4.48)
	R_{adj}^2	62.35%	62.3%	53.52%	50.1%	46.84%	45.86%	46.91%	45.52%
FF5	α	0.47%	0.45%	0.49%	0.55%	0.95%	0.92%	0.90%	0.90%
	t-stat I	(2.62)	(2.37)	(2.80)	(3.04)	(3.12)	(3.15)	(3.09)	(3.10)
	t-stat II	(2.53)	(2.48)	(3.23)	(3.42)	(4.02)	(3.98)	(3.90)	(3.95)
	R_{adj}^2	45.98%	44.58%	46.93%	39.31%	21.81%	22.4%	22.17%	22.69%
Q4	α	0.35%	0.24%	0.27%	0.34%	0.84%	0.81%	0.78%	0.79%
	t-stat I	(2.17)	(1.29)	(1.42)	(1.75)	(2.73)	(2.79)	(2.71)	(2.77)
	t-stat II	(1.69)	(1.11)	(1.29)	(1.61)	(3.34)	(3.33)	(3.24)	(3.33)
	R_{adj}^2	39.04%	36.74%	44.05%	39.22%	25.73%	26.12%	26.91%	27.12%

This table tests whether five sparse factor models proposed in past literature can explain the MVE portfolios composed of latent factors. I construct the MVE portfolios using the first seven to 10 latent factors following the same steps as in the section 1.3.4. I estimate the factors' risk prices under the prior Sharpe ratio of 0.5. The five benchmark models include (1) CAPM, (2) Fama and French (1993) three factors (FF3), (3) Fama and French (2015) five factors (FF5), (4) Carhart (1997) four factors (Carhart4), and (5) Hou, Xue, and Zhang (2015) four factors (Q4). I report three test-statistic in table 1.2: (1) α, (2) t-statistics of α, and (3) adjusted R-squared, denoted as R_{adj}^2. To control for the serial dependence of pricing errors, I use Newey and West (1987) standard errors with both 36 lags (t-stat I) and 12 lags (t-stat II).

Table 1.A.4: Correlation among $\mathcal{M}_t^{LF}$, $\mathcal{M}_t^{HF}$, $\mathcal{M}_t^{missing}$, and $\mathcal{M}_t^{unpriced}$, $SR_{prior} = 0.5$

Corr.	$\mathcal{M}_t^{LF}$	$\mathcal{M}_t^{unpriced}$	$\mathcal{M}_t^{HF}$	$\mathcal{M}_t^{missing}$	SR	t-stat (optimal lags)
$\mathcal{M}_t^{LF}$	1.00				0.378	6.24
$\mathcal{M}_t^{unpriced}$	0.00	1.00			0.014	0.25
$\mathcal{M}_t^{HF}$	0.79	0.61	1.00		0.290	4.65
$\mathcal{M}_t^{missing}$	0.61	-0.79	0.00	1.00	0.244	4.62

This table tests whether the LF-MVE portfolio can explain the HF-MVE or whether the opposite is valid. I construct the MVE portfolios using the first 7 – 10 latent factors following the same steps as in Section 1.3.4. I estimate the factors' risk prices under the prior Sharpe ratio of 0.5. I report three test-statistic in Table 1.2: (1) α, (2) t-statistics of α, and (3) adjusted R-squared, denoted as R_{adj}^2. To control for the serial dependence of pricing errors, I use Newey and West (1987) standard errors with both 36 lags (t-stat I) and 12 lags (t-stat II).

Table 1.A.5: Economic Properties of HF- vs. LF-SDFs **II**

Y_t:	$N_t^{CF,2}$	$N_t^{DR,2}$	HKM_{qt}^{ntr}	HKM_{qt}^{tr}	VXO_t	BW_t	HJTZ$_t$	HJTZ$_t^{ar1}$
	Panel (A): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{HF} + \beta_2 \mathcal{M}_t^{missing} + \epsilon_t$							
$\mathcal{M}_t^{HF}$	-0.134	-0.278	-0.301	-0.344	0.215	-0.193	-0.162	-0.125
t-stat I	(-1.125)	(-2.931)	(-2.642)	(-2.843)	(1.309)	(-1.641)	(-1.030)	(-1.658)
t-stat II	(-0.952)	(-2.426)	(-2.466)	(-2.695)	(1.383)	(-1.523)	(-0.912)	(-1.695)
$\mathcal{M}_t^{missing}$	-0.099	-0.048	0.206	0.193	-0.053	-0.085	-0.076	0.054
t-stat I	(-1.427)	(-0.695)	(1.480)	(1.398)	(-0.724)	(-1.743)	(-1.272)	(0.665)
t-stat II	(-1.325)	(-0.696)	(1.562)	(1.572)	(-0.669)	(-1.791)	(-1.164)	(0.787)
	Panel (B): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{LF} + \beta_2 \mathcal{M}_t^{unpriced} + \epsilon_t$							
$\mathcal{M}_t^{LF}$	-0.167	-0.257	-0.146	-0.189	0.148	-0.209	-0.177	-0.074
t-stat I	(-1.560)	(-2.861)	(-1.522)	(-1.726)	(0.939)	(-1.861)	(-1.116)	(-0.965)
t-stat II	(-1.570)	(-2.929)	(-2.177)	(-2.547)	(0.947)	(-1.740)	(-0.991)	(-1.000)
$\mathcal{M}_t^{unpriced}$	0.007	-0.115	-0.334	-0.346	0.164	-0.037	-0.027	-0.114
t-stat I	(0.062)	(-1.244)	(-2.258)	(-2.271)	(1.956)	(-0.557)	(-0.426)	(-1.393)
t-stat II	(0.059)	(-1.144)	(-2.006)	(-2.153)	(2.125)	(-0.578)	(-0.427)	(-1.671)
ρ	-0.173	-0.114	-0.023	0.024	0.888	0.951	0.985	0.408
R_{adj}^2	2.78%	7.96%	13.30%	15.54%	4.90%	4.50%	3.21%	1.85%
Sample size	338	338	112	112	338	326	326	326

This table reports the results of regressing economic variables on different components of SDFs. It differs from Table 1.5 in following aspects: (1) I estimate cash-flow and discount-rate news including five state variables into VAR(1) regression, as in Campbell, Giglio, and Polk (2013); (2) I use quarterly intermediary factors rather than monthly ones; (3) I use the original time-series of VXO index and Baker and Wurgler (2006) sentiment index, rather than AR(1) shocks in these variables. In addition, I consider the sentiment index in Huang, Jiang, Tu, and Zhou (2015) in the last two columns.

Specifically, the dependent variables include (1) cash-flow news, (2) discount-rate news, (3) the quarterly nontradable intermediary factor, (4) the quarterly tradable intermediary factor, (5) the VXO index, (6) the Baker and Wurgler (2006) sentiment index, (7) the Huang, Jiang, Tu, and Zhou (2015) sentiment index, and (8) the AR(1) shock in the Huang, Jiang, Tu, and Zhou (2015) sentiment.

The SDFs are composed of the first seven principal components of asset returns and their risk prices are estimated under the prior Sharpe ratio equal to 0.4. I standardize both dependent and independent variables so that readers can interpret all coefficient estimates as correlations. I report two t-statistics using Newey and West (1987) standard errors with (1) 36 lags (t-stat I) and (2) 12 lags (t-stat II). In addition, I also report dependent variables' first-order autocorrelation coefficients (ρ). The monthly (quarterly) out-of-sample is from November 1991 to December 2019 (Q1 1992 – Q4 2019).

Table 1.A.6: Economic Fundamentals related to HF- vs. LF-SDFs, $SR_{prior} = 0.5$

Y_t :	C_t^{nd}	C_{t+1}^{nd}	GDP_t	GDP_{t+1}	N_t^{CF}	N_t^{DR}	HKM_t^{ntr}	HKM_t^{tr}	VXO_t^{ar1}	BW_t^{ar1}
				Panel (A): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{HF} + \beta_2 \mathcal{M}_t^{missing} + \epsilon_t$						
$\mathcal{M}_t^{HF}$	-0.011	0.169	-0.147	0.062	-0.062	-0.227	-0.173	-0.221	0.182	-0.152
t-stat I	(-0.113)	(1.707)	(-0.650)	(0.635)	(-0.454)	(-2.299)	(-1.879)	(-1.931)	(2.154)	(-2.618)
t-stat II	(-0.128)	(1.707)	(-0.693)	(0.694)	(-0.412)	(-2.004)	(-1.709)	(-1.863)	(2.128)	(-2.772)
$\mathcal{M}_t^{missing}$	-0.192	-0.209	-0.158	-0.192	-0.081	-0.004	0.162	0.200	-0.080	-0.014
t-stat I	(-1.614)	(-3.304)	(-1.511)	(-1.985)	(-1.057)	(-0.054)	(1.769)	(1.893)	(-1.629)	(-0.188)
t-stat II	(-1.864)	(-3.304)	(-1.594)	(-2.209)	(-1.056)	(-0.051)	(1.721)	(1.857)	(-1.635)	(-0.188)
				Panel (B): $Y_t = \beta_0 + \beta_1 \mathcal{M}_t^{LF} + \beta_2 \mathcal{M}_t^{unpriced} + \epsilon_t$						
$\mathcal{M}_t^{LF}$	-0.123	0.013	-0.212	-0.064	-0.099	-0.181	-0.037	-0.052	0.094	-0.129
t-stat I	(-1.009)	(0.147)	(-0.925)	(-0.613)	(-0.895)	(-2.077)	(-0.429)	(-0.538)	(1.351)	(-1.876)
t-stat II	(-0.995)	(0.149)	(-0.965)	(-0.664)	(-0.896)	(-2.207)	(-0.475)	(-0.567)	(1.409)	(-1.871)
$\mathcal{M}_t^{unpriced}$	0.149	0.269	0.040	0.192	0.025	-0.137	-0.234	-0.294	0.175	-0.082
t-stat I	(3.217)	(3.264)	(0.415)	(2.138)	(0.201)	(-1.443)	(-2.252)	(-2.332)	(2.592)	(-1.296)
t-stat II	(2.856)	(3.281)	(0.497)	(2.418)	(0.195)	(-1.259)	(-2.050)	(-2.228)	(2.421)	(-1.304)
ρ	0.153	0.153	0.352	0.352	-0.189	-0.108	0.061	0.104	0.116	0.105
R_{adj}^2	4.91%	7.43%	6.30%	5.11%	2.76%	9.13%	7.51%	11.38%	5.80%	2.19%
Sample size	112	111	112	111	338	338	338	338	338	326

This table differs from Table 1.5 only in the prior Sharpe ratio that I use to estimate risk prices of latent factors. Specifically, this table sets SR_{prior} to be 0.5. See the footnote in Table 1.5 for details.

1.A.4 Additional Figures
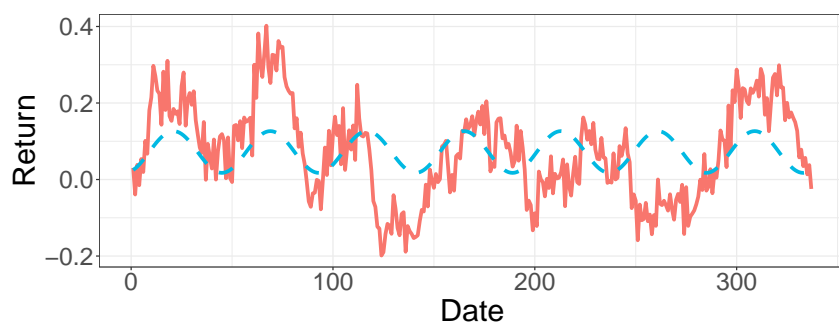
(a) Fast-Moving: AR(1) coefficient = -0.5
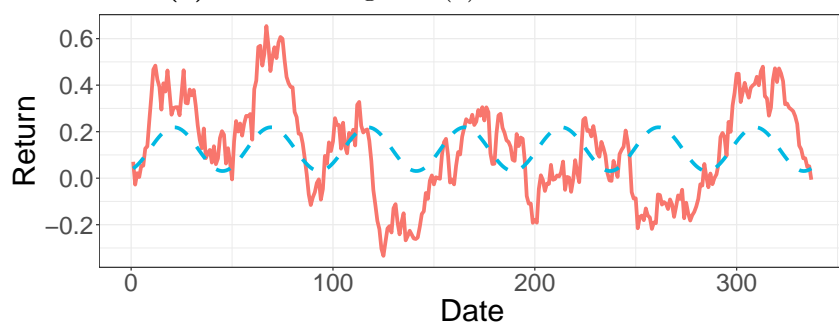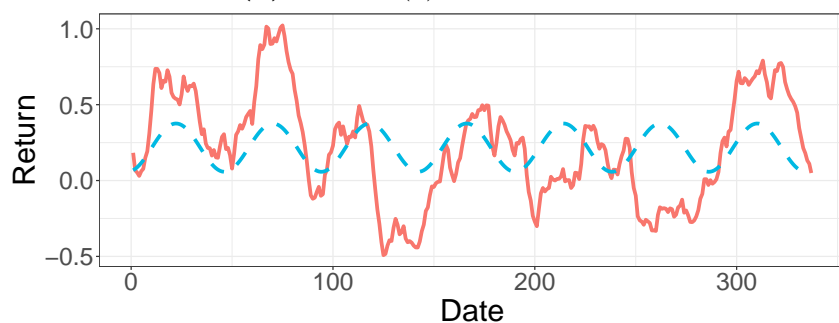
(b) IID: AR(1) coefficient = 0

(c) Slow-Moving: AR(1) coefficient = 0.5

Figure 1.A.1: Cumulative returns in a 24-month rolling window

This graph plots the cumulative returns in a 24-month rolling window. I consider three AR(1) processes for monthly (demeaned) asset returns: $x_{t+1} = \rho_x x_t + \sqrt{1 - \rho_x^2} \sigma_x \eta_{x,t+1}$, where $\sigma_x^2 = 4.5\%$, $\rho_x \in \{-0.5, 0, 0.5\}$.
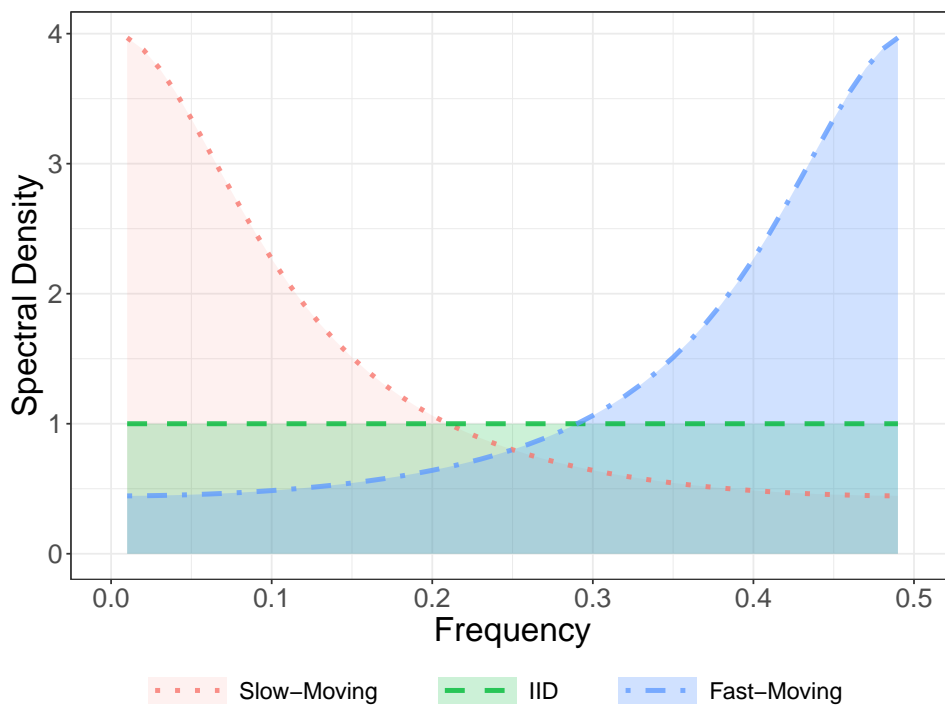
(a) $\rho_x \in \{-0.5, \ 0, \ 0.5\}$.
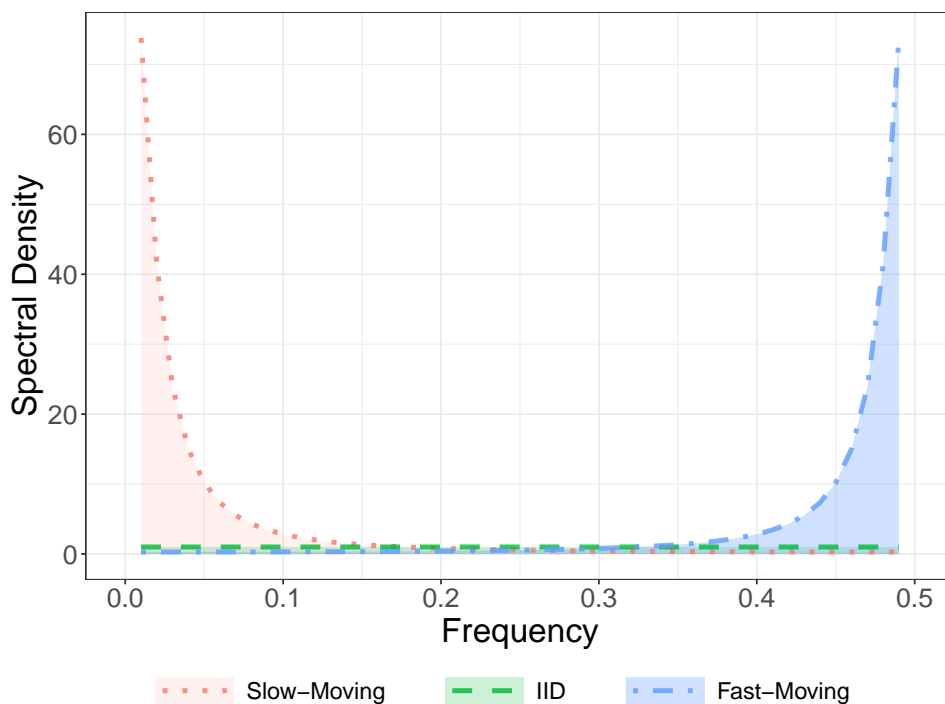
(b) $\rho_x \in \{-0.9, \ 0, \ 0.9\}$.

Figure 1.A.2: Spectral density function of AR(1) processes

This graph plots the spectral density functions of three AR(1) processes: $x_{t+1} = \rho_x x_t + \sqrt{1 - \rho_x^2} \eta_{x,t+1}$, where $\eta_{x,t+1} \overset{\text{iid}}{\sim} WN(0,1)$. When ρ_x is positive (negative), this time series is slow-moving (fast-moving).

(a) $x_t^{slow} = 3\cos(2\pi t\frac{5}{100}) + 3\sin(2\pi t\frac{5}{100})$

(b) $x_t^{fast} = 2\cos(2\pi t\frac{50}{100}) + 2\sin(2\pi t\frac{50}{100})$

(c) $x_t = x_t^{slow} + x_t^{fast}$

(d) Spectral density function $f_x(\omega)$, DFT
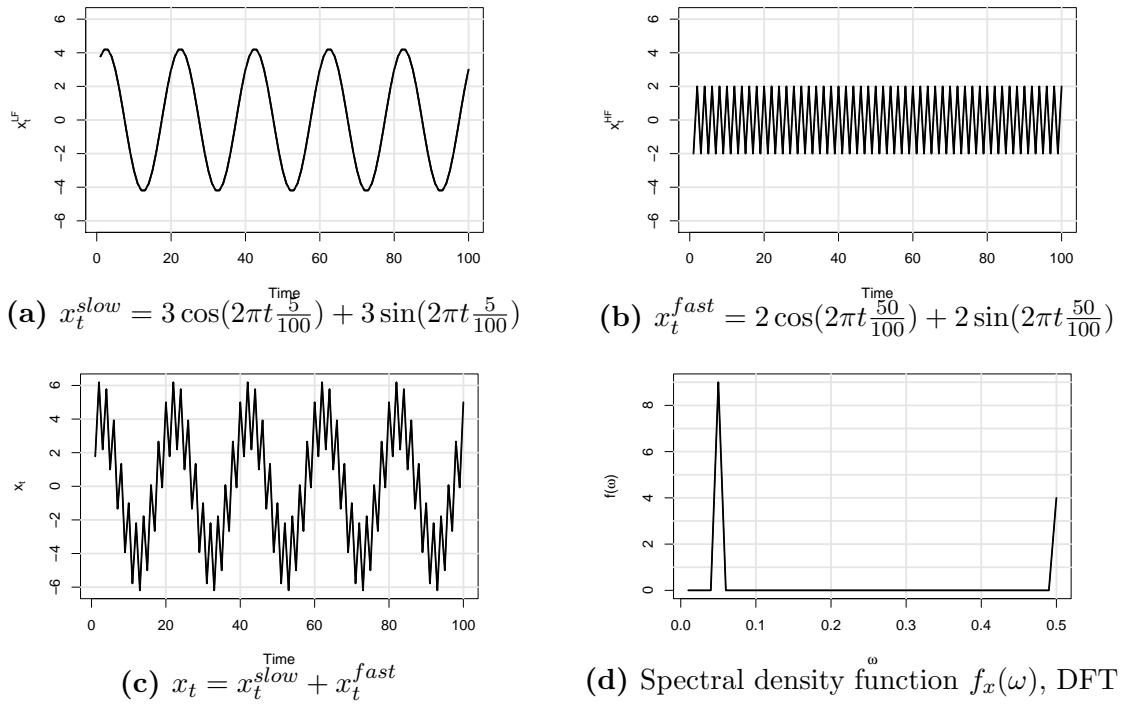
Figure 1.A.3: Example: decompose a deterministic time series via DFT

(a) TSVs explained by different components

(b) LF-eigenvalues / HF-eigenvalues
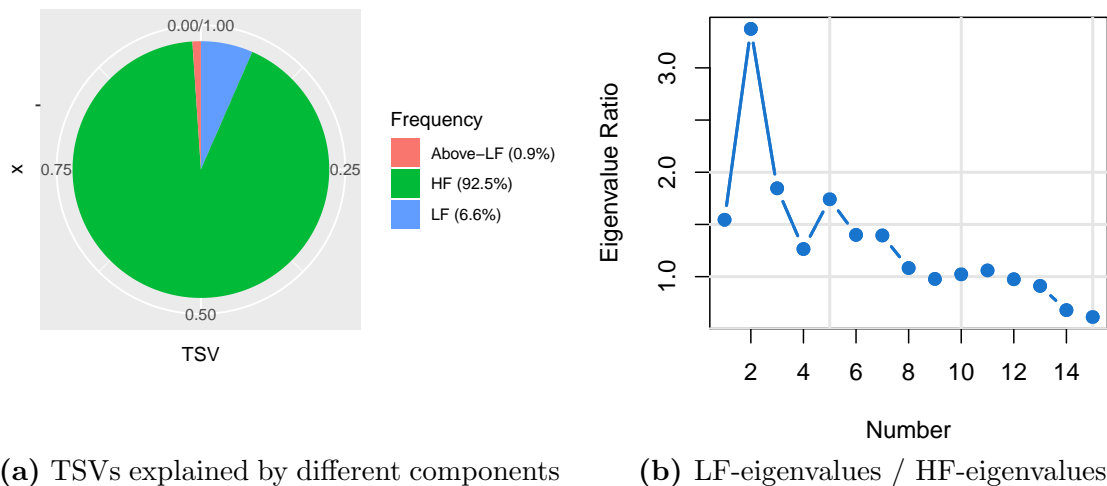
Figure 1.A.4: Time-series variations in 78 assets, subsample 2

Panel (a) plots the fraction of time-series variations in 78 asset returns explained by the HF, LF, and above-LF components. Panel (b) plots the ratio of the first 15 LF-eigenvalues over HF-eigenvalues. The sample starts from November 1991 to December 2019.
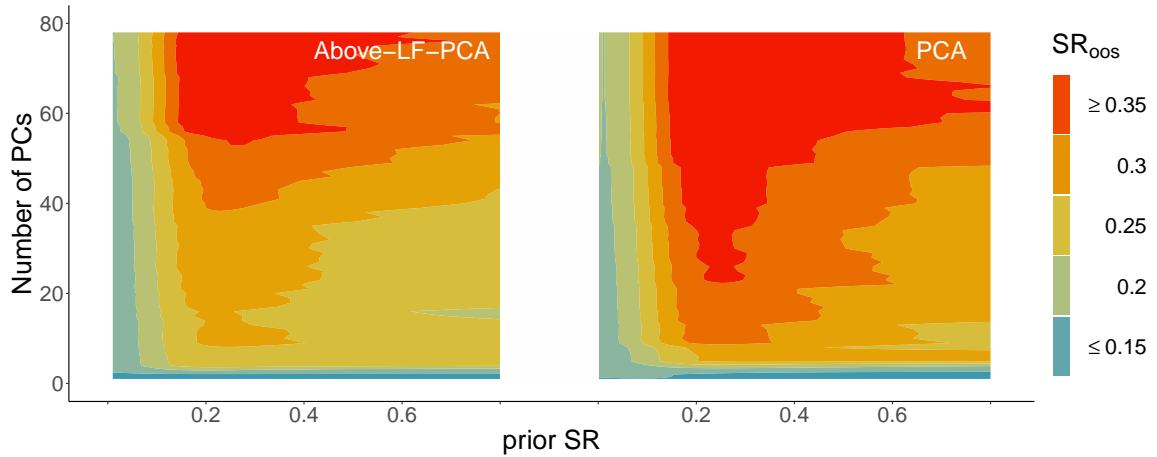
Figure 1.A.5: OOS Sharpe ratio of Above-LF-PCA and PCA, 78 test assets

This graph plots the heat-maps of the OOS Sharpe ratio of Above-LF-PCA and PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. I include the PCs into the SDF based on their ability to explain time-series variations.
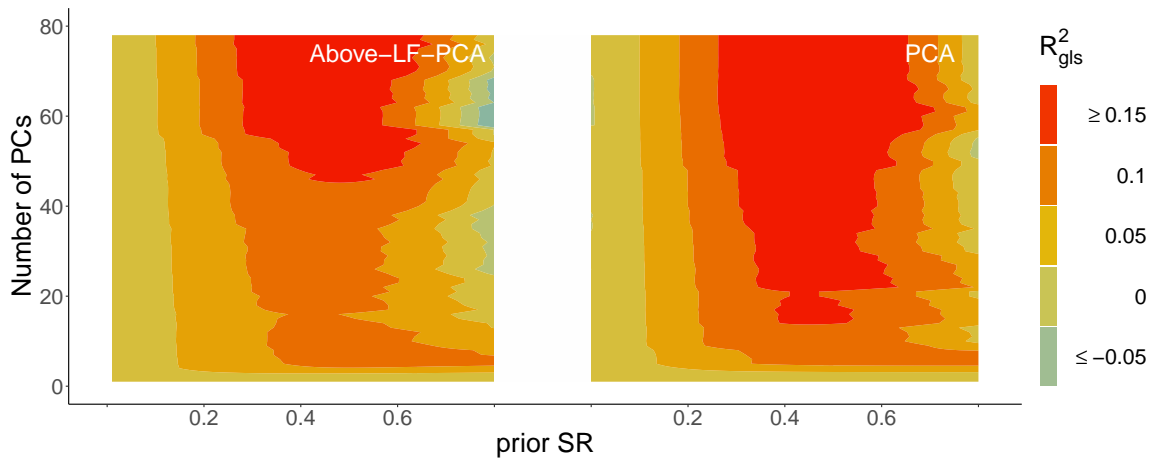
Figure 1.A.6: OOS R^2_{gls} of Above-LF-PCA and PCA, 78 test assets

This graph plots the heat-maps of the OOS R^2_{gls} of Above-LF-PCA and PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS R^2_{gls}. I include the PCs into the SDF based on their ability to explain time-series variations.
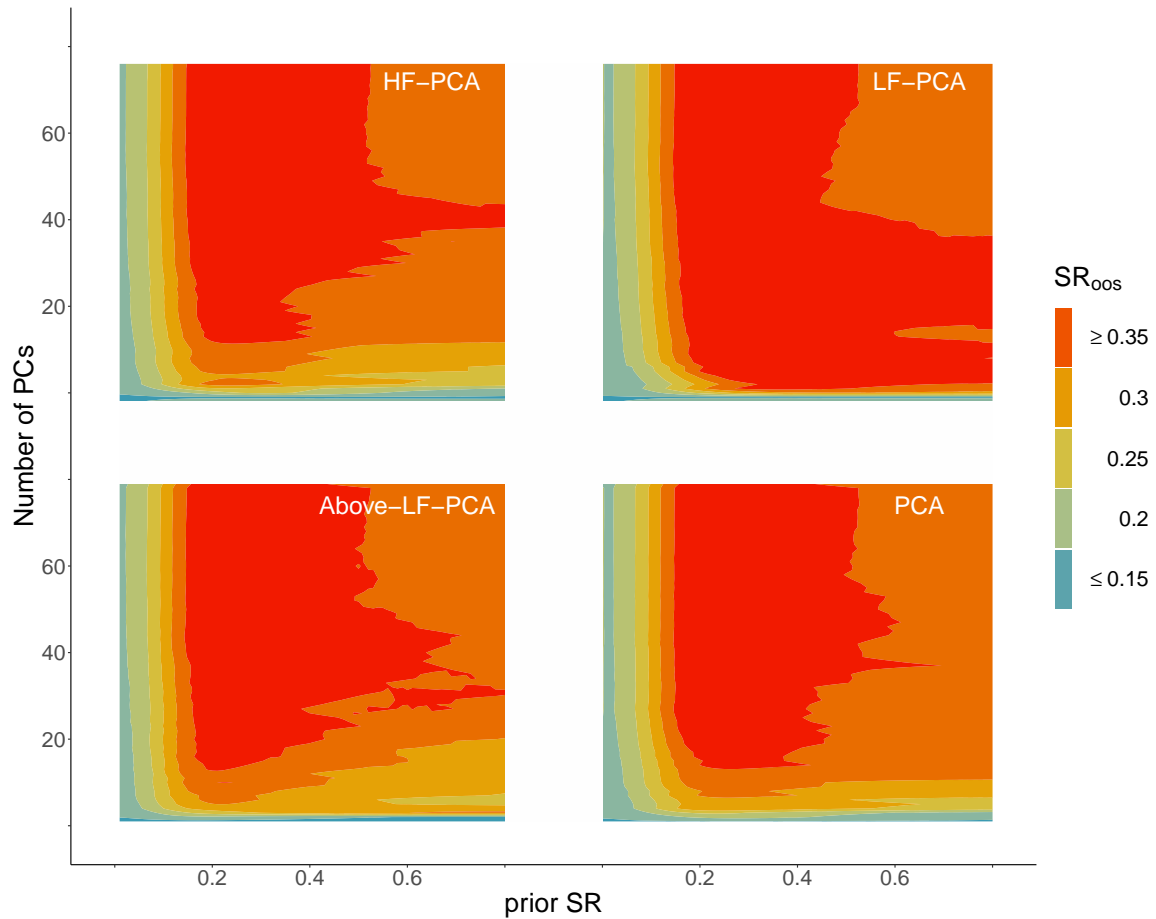
Figure 1.A.7: OOS Sharpe ratio using Kozak, Nagel, and Santosh (2020), 78 test assets

This graph plots the heat-maps of OOS Sharpe ratios of HF-PCA, LF-PCA, Above-LF-PCA and PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. The risk prices and the number of PCs entering the SDFs are determined by the Kozak, Nagel, and Santosh (2020) objective function in equation (1.17).
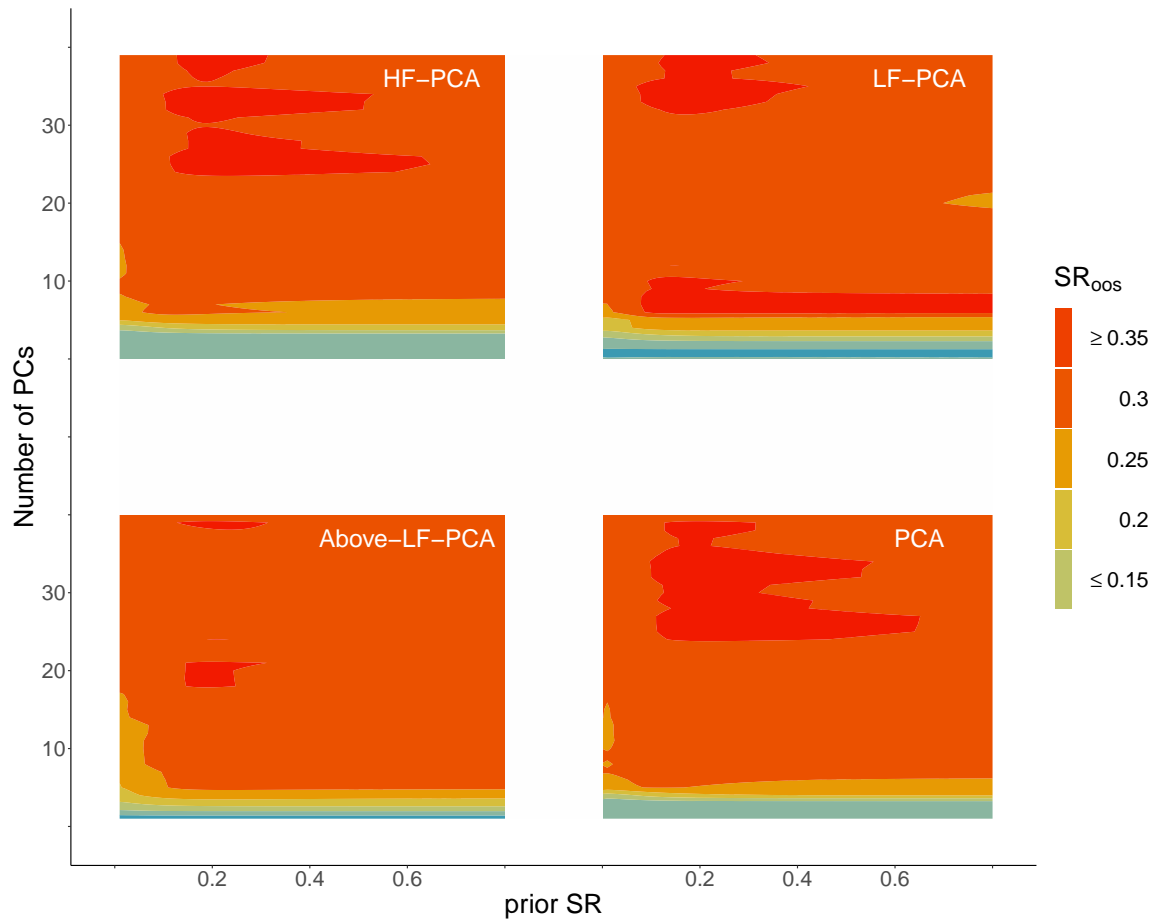
Figure 1.A.8: OOS Sharpe ratio of 39 test assets

This graph plots the heat-maps of OOS Sharpe ratios of HF-PCA, LF-PCA, Above-LF-PCA and PCA in the cross-section of 39 long-short portfolios. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. I include the PCs into the SDF based on their ability to explain time-series variations of 39 long-short portfolios.
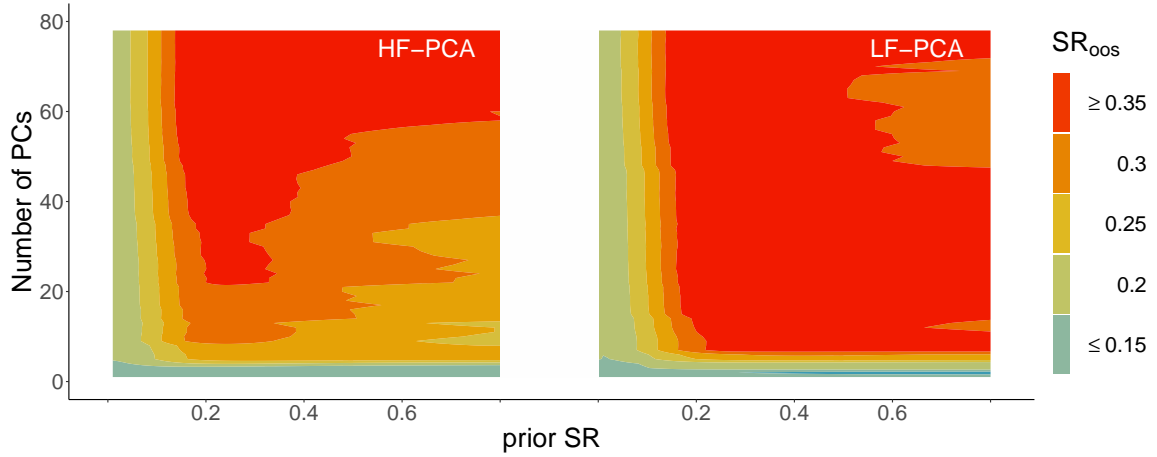
Figure 1.A.9: Robustness Check: OOS Sharpe ratio of 78 test assets, $\tau^{LF} \in [24, 120]$

This graph plots the heat-maps of OOS Sharpe ratios of HF-PCA and LF-PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. I include the PCs into the SDF based on their ability to explain time-series variations. The LF frequency interval is defined as $\tau^{LF} \in [24, 120]$.
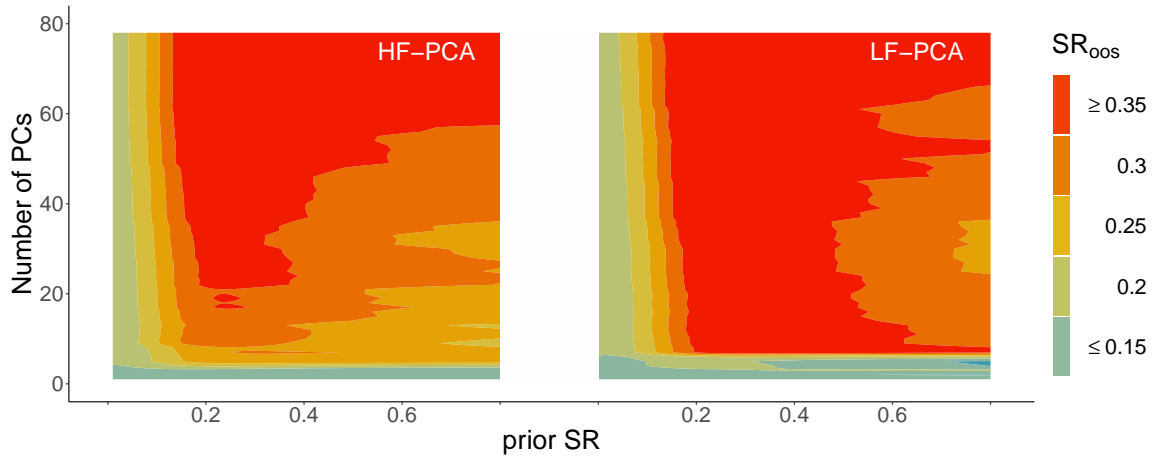
Figure 1.A.10: Robustness Check: OOS Sharpe ratio of 78 test assets, $\tau^{LF} \in [32, 64]$

This graph plots the heat-maps of OOS Sharpe ratios of HF-PCA and LF-PCA in the cross-section of 78 test assets. In each panel, the x-axis denotes the prior Sharpe ratio of the factor model, while the y-axis is the number of PCs included in the SDF. In addition, different colors represent different OOS Sharpe ratios. I include the PCs into the SDF based on their ability to explain time-series variations. The LF frequency interval is defined as $\tau^{LF} \in [32, 64]$.

Chapter 2

Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models

Svetlana Bryzgalova, Jiantao Huang, and Christian Julliard[1]

2.1 Introduction

In the last decade or so, two observations have come to the forefront of the empirical asset pricing literature. First, thanks to the factor zoo phenomenon, in the near future we might have as many empirically "priced" sources of risk as stock returns. Second, the so-called weak factors (i.e., factors whose true covariance with asset returns is asymptotically zero) are likely to both appear empirically relevant and invalidate inference on the true sources of risk (see, e.g., Gospodinov, Kan, and Robotti (2019), and Kleibergen and Zhan (2020)). Nevertheless, to the best of our knowledge, no general method has been suggested to date that: i) is applicable to both tradable and non-tradable factors, ii) can handle the entire factor zoo, iii) remains valid under misspecification, iv) is robust to the weak inference problem, and, importantly, v) delivers an empirical pricing kernel that outperforms (in- and out-of-sample) popular models (with either observable or latent factors). And that is exactly

what we provide.

We develop a *unified framework* for tackling linear asset pricing models. In the case of stand-alone model estimation, our method provides reliable price of risk estimates, hypothesis testing, and confidence intervals for these parameters, as well as all other objects of interest – alphas, R^2's, Sharpe ratios, etc. Furthermore, even when all the pricing kernels are misspecified and non-nested, our approach delivers factor selection – *if* a dominant model exists – or model averaging, if there is no clear winner given the data. The method is numerically simple, fast, easy to use, and can be feasibly applied to literally quadrillions of candidate factor models.

Empirically, we find that the Stochastic Discount Factor (SDF) constructed as the Bayesian Model Averaging (BMA) over the space of 2.25 quadrillion models, prices a wide cross-section of anomalies better than both celebrated (observable) factor models and the latent factor approach of Kozak, Nagel, and Santosh (2020). This outperformance arises not only in sample but also out-of-sample in *both* time series and cross-sectional dimensions.[2] There are three key drivers of this performance. First, our method reliably identifies a small subset of observable factors that should be included in any SDF with high probability. Second, although these factors alone are already sufficient to outperform notable (observable) factor models, they do not fully characterize the SDF. The latter, as we show, is *dense* in the space of observable factors. As a result, the BMA optimally (in the predictive density sense) aggregates multiple imperfect measures of the same sources of risk. Third, our method relies on a novel prior that is fully driven by the researcher's belief about the Sharpe ratio in the economy, and that effectively controls potential overfitting. The BMA-SDF neither requires arbitrary tuning parameters nor separates factor extraction and aggregation. Instead, unlike most of the existing literature, it delivers an SDF in one step, driven by transparent and economically motivated priors.

As stressed by Harvey (2017) in his AFA presidential address, the factor zoo naturally calls for a Bayesian solution – and we develop one. Furthermore, we show that factor proliferation and spurious inference are tightly connected problems, and a naïve Bayesian model selection fails in the presence of weak factors. We develop a reliable solution focused on the SDF representation, since the key question posed by the factor zoo lies in whether candidate risk factors have non-zero price of risk. Our Bayesian SDF formulation (B-SDF) is intuitively similar to the standard frequentist OLS/GLS estimation that imposes the self-pricing of tradable factors when they are part of the test assets. However, it is robust to identification failure, allows us to easily compare and aggregate non-nested models, and

---

[2]In cross-sectional out-of-sample exercises, we first estimate the BMA-SDF in a baseline cross-section, and then use it to price several other cross-sections *without* any further parameter estimation.

provides robust inference for all the quantities of interest within stand-alone models and across the whole model space. Remarkably, unlike the frequentist alternatives, the B-SDF estimator performs well in both small and large samples, even with fairly large cross-sections.

Our empirical results are based on what is arguably a representative cross-section of test assets: 60 portfolios based on a large number of firm-specific characteristics. We examine 51 factors proposed in the previous literature, yielding a total of 2.25 quadrillion possible models to analyze. We find that only a handful of factors proposed in the literature are robust explanators of the cross-section of returns, and a three (at most six) *most likely* factor model easily outperforms canonical reduced-form benchmarks. Nevertheless, there is no clear "winner" across the whole space of potential models: Hundreds of possible specifications that combine tradable and non-tradable factors, none of which has been examined in the previous literature, are virtually equally likely to price the cross-section of returns.

Furthermore, we find that the "true" latent SDF is dense in the space of observable factors; that is, a large subset of variables is needed to fully capture its pricing implications.[3] Nonetheless, the SDF-implied maximum Sharpe ratio in the economy is not unrealistically high, suggesting substantial commonality among the risks spanned by the factors in the zoo. BMA, therefore, emerges naturally as an optimal way of aggregating models that load on the same set of underlying risks: It aggregates all the possible factors and models based on their likelihood to have generated the data. Crucially, this approach allows for *both* selection and aggregation based on the posterior probabilities of the factors being part of the pricing kernel, and allows the data to decide on the optimal structure of the SDF. Empirically, we find that the BMA-SDF performs well both in- and out-of-sample (OOS). Its OOS performance is stable across subsamples (going both into the future and into the past), and, most importantly, it prices well cross-sections not used for its construction, including the notoriously challenging 49 industry portfolios.

Our contribution is fourfold. First, we develop a very simple Bayesian estimator for linear SDFs with both traded and non-traded factors. This approach makes weak factors easily detectable in finite sample, while providing valid inference on the strong factors' price of risk, measures of cross-sectional fit, and other objects of interest. Our robust approach is very simple to implement and use, and it does not require pre-testing or pre-estimation.

Second, we provide a method for inference on the *entire* factor zoo with model (and factor) posterior probabilities. However, as we show, model and factor selection based on marginal likelihoods (i.e., on posterior probabilities or Bayes factors) is unreliable under a flat prior for the price of risk: Asymptotically, weakly identified factors are selected with probability

---

[3]Interestingly, the SDF remains dense even when we include either the five principal components or the five RP-PCs of Lettau and Pelger (2020b).

one even if they have zero price of risk.[4] This observation, however, not only illustrates the nature of the problem; it also suggests how to restore inference: use suitable, non-informative – but yet non-flat – priors. Building upon the literature on predictor selection (see, e.g., Ishwaran, Rao, et al. (2005) and Giannone, Lenza, and Primiceri (2021b)), we provide a novel (continuous) "spike-and-slab" prior that restores the validity of model and factor selection based on posterior model probabilities and Bayes factors. It is uninformative (the "slab") for strong factors but shrinks away (the "spike") the weak ones. This prior also: *i*) makes it computationally feasible to analyze quadrillions of alternative factor models, *ii*) allows the researcher to encode prior beliefs (or lack thereof) about the sparsity of the true SDF without imposing hard thresholds, *iii*) restores the validity of hypothesis testing, and *iv*) performs well in numerous simulation settings. The prior is entirely pinned down by economic quantities: It maps into beliefs about the Sharpe ratio of the risk factors. We regard this approach as a solution for the high-dimensional inference problem generated by the factor zoo.[5]

Third, we provide a new way of selecting robust observable factors. Indeed, we find a new 3–6 observable factor model, combining variables from different papers, that dominates all the popular reduced-form benchmarks. However, even that model would be strongly rejected by the data: No sparse factor model is among the most likely 2000 data-generating processes that we consider. Furthermore, a *unique* best performing combination of the factors (sparse or dense in observables) does not seem to exist: Hundreds of possible models, never proposed in the previous literature, deliver almost equivalent performance, which indicates fragility of conventional model selection and horse races, popular among reduced-form sparse factor models.

Fourth, our results do not rely on ex ante unverifiable assumptions of existence, uniqueness, and sparsity of the true SDF representation among the candidate models (unlike LASSO and other popular frequentist methods). When a dominant model for the SDF does not arise in the data (as in our analysis), our method does not stop at selection. Instead, it efficiently aggregates pricing information from (potentially) the entire factor zoo. Interestingly, we show that solely extracting leading standard latent factors from a wide range of predictors using PCA or RP-PCA, is not sufficient to characterize the SDF. In fact, we find that observable and (some) leading latent factors are complementary for such a characterization. Therefore,

---

[4]This is similar to the effect of "weak instruments" in IV estimations, as discussed in Sims (2007).

[5]Despite a seemingly prohibitive dimension of the model space, the estimation is numerically simple and computationally feasible. Our Markov Chain, used to evaluate the whole space of 2.25 quadrillions of models and deliver all the baseline results from the paper, takes about four hours on a 3.0GHz 10-core Intel Xeon W processor and 128 GB of RAM. Furthermore, we formally test its convergence and establish that the posterior distributions converge already after *less than one fifth* of the Markov Chain draws, making our method easily applicable for most researchers.

our results indicate that there is scope for both more efficient latent factor extraction and better aggregation informed by economic fundamentals.

The remainder of the paper is organized as follows. In the next subsection we review the most closely related literature and our contribution to it. Section 2.2 provides a brief overview of the benchmark frequentist approach, while Section 2.3 outlines the Bayesian SDF estimation and its properties for inference, selection, and model aggregation. Section 2.4 provides simulation evidence on both small- and large-sample behavior of our method. Section 2.5 presents our empirical results. Finally, Section 2.6 discusses potential extensions of our procedure and concludes.

2.1.1   Related Literature

There are numerous strands of literature relying on Bayesian tools, especially for asset allocation (for an excellent overview, see Avramov and Zhou (2010)), model selection (e.g., Chib, Zeng, and Zhao (2020)), and performance evaluation (Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002), and Harvey and Liu (2019)). Therefore, we aim to provide only an overview of the literature that is most closely related to our paper.

Shanken (1987) and Harvey and Zhou (1990) are probably the first to use the Bayesian framework in portfolio choice and develop GRS-type tests (cf. Gibbons, Ross, and Shanken (1989)) for mean-variance efficiency. While Shanken (1987) is the first to examine the posterior odds ratio for portfolio alphas in the linear factor model, Harvey and Zhou (1990) set the benchmark by imposing priors on the deep model parameters. Interestingly, we show that there is a tight link between using the most popular, diffuse, priors for the price of risk and the failure of the standard estimation techniques in the presence of weak factors.

Pástor and Stambaugh (2000) and Pástor (2000) assign a prior distribution to the vector of pricing errors $\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \sim \mathcal{N}(0, \kappa \boldsymbol{\Sigma_R})$, where $\boldsymbol{\Sigma_R}$ is the variance-covariance matrix of returns and $\kappa \in \mathbb{R}_+$, and apply it to portfolio choice. This prior imposes a degree of shrinkage on the alphas: When factor models are misspecified, pricing errors cannot be too large a priori. This prior effectively places a bound on the Sharpe ratio achievable in this economy.

Barillas and Shanken (2018a) extend the aforementioned prior and derive a closed-form solution for the Bayes factor when all the risk factors are tradable and use it to compare different linear factor models exploiting the time series dimension of the data. Chib, Zeng, and Zhao (2020) show that the *improper* prior specification of Barillas and Shanken (2018a) is problematic and propose a new class of priors that leads to valid comparison for traded factor models.

There is a general close connection between the Bayesian approach to model selection and parameter estimation and the shrinkage-based one. Garlappi, Uppal, and Wang (2007)

impose a set of different priors on expected returns and the variance-covariance matrix and find that the shrinkage-based analogue leads to superior empirical performance. The ridge-based approach to recovering the SDF of Kozak, Nagel, and Santosh (2020) can also be interpreted from a Bayesian perspective with priors on the expected returns distribution.

To the best of our knowledge, our paper is the first attempt to develop a general Bayesian approach for both tradable and non-tradable factors, capable of imposing tradable restriction on the price of risk when needed. Flat priors for the price of risk, we show, lead to erroneous model selection in the presence of weak factors. Hence, we develop a novel one that depends on the degree of parameter identification. This prior is heterogenous among factors, depending on the correlation between test assets and the factor itself. In the spirit of Pástor and Stambaugh (2000), our prior directly maps into beliefs about the Sharpe ratio achievable in the economy, yet without imposing a hard threshold on it. Not only does it restore the validity of model selection, but it also allows for sharp inference in small sample on all the economic quantities of interest.

Our paper naturally contributes to the literature on weak identification in asset pricing. Starting from the seminal papers of Kan and Zhang (1999a,b), identification of risk premia has been shown to be challenging for traditional estimation procedures. Kleibergen (2009) demonstrates that the two-pass regression of Fama-MacBeth lead to biased estimates of the risk premia and spuriously high significance levels. Moreover, useless factors often crowd out the impact of the true sources of risk in the model and lead to seemingly high levels of cross-sectional fit (Kleibergen and Zhan (2015)). Gospodinov, Kan, and Robotti (2014, 2019) demonstrate that most of the estimators used to recover risk premia in the cross-section are invalidated by the presence of useless factors, and they propose alternative procedures that effectively eliminate the impact of these factors. We build upon the intuition developed in these papers and formulate the Bayesian solution to the problem by providing a prior such that when the vector of correlation coefficients between asset returns and a factor is close to zero, the prior variance for the price of risk also goes to zero, effectively shrinking the posterior toward zero.

Our method does not require any pretesting, works well in small and large time-series and cross-sectional dimensions. Furthermore, due to its hierarchical structure, it can be feasibly extended to handle time variation in the factor exposure and asset risk premia, and it accommodates both observable and latent factors. Most importantly, our approach provides a robust unified framework for evaluation of stand-alone models, factor and model selection, as well as aggregation, even when all the potential models are misspecified.

Naturally, our paper also contributes to the active (and growing) body of work that critically re-evaluates existing findings in the empirical asset pricing literature and develop robust

inference methods. There is ample empirical evidence that most linear factor models are misspecified (e.g., Chernov, Lochstoer, and Lundeby (2022), and He, Huang, and Zhou (2018)). Following Harvey, Liu, and Zhu (2016), a large body of literature has tried to understand which of the existing factors (or their combinations) drive the cross-section of asset returns. Gospodinov, Kan, and Robotti (2014) develop a general approach for misspecification-robust inference, while Giglio and Xiu (2021) exploit the invariance principle of the PCA and recover the price of risk of a given factor from the projection on the span of latent factors driving a cross-section of returns. Similarly, Uppal, Zaffaroni, and Zviadadze (2018) recover latent factors from the residuals of an asset pricing model, effectively completing the span of the SDF. Feng, Giglio, and Xiu (2020) combine cross-sectional asset pricing regressions with the double-selection LASSO of Belloni, Chernozhukov, and Hansen (2014) to provide valid uniform inference on the selected sources of risk when the true SDF is sparse. Huang, Li, and Zhou (2018) use a reduced rank approach to select from not only the observable factors but their total span, effectively allowing for sparsity in both factors and their combinations.

We do not take a stand on the origin of the factors, the "unique" true model being among the candidate specifications, and a priori SDF sparsity. Instead, we consider the whole universe of potential models that can be created from a wide set of factors proposed in the empirical literature (observable and latent) and let the data speak. We find that the cross-sectional likelihood across many best-performing (dense) models is flat. Hence, the data seem to call for aggregation, rather than selection.

Avramov (2002, 2004) are the first formal studies that bring model uncertainty to the forefront of asset pricing. Building on these seminal papers, Anderson and Cheng (2016) develop a BMA approach to portfolio choice that, with formal recognition of model uncertainty, delivers robust asset allocation and superior out-of-sample performance. Similarly, we find that there is a large degree of model uncertainty in cross-sectional asset pricing, suggesting a large degree of model misspecification and rendering canonical selection unreliable. We therefore develop a BMA method that explicitly targets cross-sectional pricing of asset returns. The resulting averaging over the space of SDFs delivers superior pricing in- and out-of-sample.

In reality, the BMA-SDF has – endogenously – elements of both selection and aggregation: While a small subset of factors delivers large individual contributions to the SDF, other factors are efficiently bundled together to deliver the best predictive density of the cross-sectional pricing kernel. In the recent literature, model *selection* (see, e.g., Feng, Giglio, and Xiu (2020)) or *aggregation* (see, e.g., Kozak, Nagel, and Santosh (2020)) of pricing factors, have been largely mutually exclusive alternatives. Our framework, instead, successfully combines both.

2.2 Frequentist Estimation of Linear SDFs

This section introduces the notation and reviews the basics of linear SDF models as well as related (frequentist) Generalized Method of Moments (GMM) estimation. Suppose that there are K factors, $\boldsymbol{f_t} = (f_{1t} \ldots f_{Kt})^\top$, $t = 1, \ldots T$, which could be either tradable or non-tradable. The returns of N test assets, which are long-short portfolios, are denoted by $\boldsymbol{R_t} = (R_{1t} \ldots R_{Nt})^\top$. Throughout the paper, $\mathbb{E}[X]$ or μ_X denotes the unconditional expectation of arbitrary random variable X and $\bar{X}$ denotes the sample mean operator.

Consider linear stochastic discount factors (M), that is models of the form $M_t = 1 - (\boldsymbol{f_t} - \mathbb{E}[\boldsymbol{f_t}])^\top \boldsymbol{\lambda_f}$. In the absence of arbitrage opportunities $\mathbb{E}[M_t \boldsymbol{R_t}] = \boldsymbol{0_N}$, which implies that expected returns are given by $\boldsymbol{\mu_R} = \mathbb{E}[\boldsymbol{R_t}] = \boldsymbol{C_f} \boldsymbol{\lambda_f}$, where $\boldsymbol{C_f}$ is the covariance matrix between $\boldsymbol{R_t}$ and $\boldsymbol{f_t}$ and $\boldsymbol{\lambda_f} \in \mathbb{R}^K$ denotes the vector of prices of risk associated with the factors. The latter can therefore be estimated via the cross-sectional regression:

$$\boldsymbol{\mu_R} = \lambda_c \boldsymbol{1_N} + \boldsymbol{C_f} \boldsymbol{\lambda_f} + \boldsymbol{\alpha} = \boldsymbol{C} \boldsymbol{\lambda} + \boldsymbol{\alpha}, \tag{2.1}$$

where $\boldsymbol{C} = (\boldsymbol{1_N}, \boldsymbol{C_f})$, $\boldsymbol{\lambda}^\top = (\lambda_c, \boldsymbol{\lambda_f^\top})$, λ_c is a scalar average mispricing (equal to zero under the null of the model being correctly specified), $\boldsymbol{1_N}$ denotes an N-dimensional vector of ones, and $\boldsymbol{\alpha} \in \mathbb{R}^N$ is the vector of pricing errors in excess of λ_c (also equal to zero under the null of the model).

Such a model is usually estimated via GMM (see Hansen (1982)) with the following moment conditions:

$$\mathbb{E}[\boldsymbol{g_t}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})] = \mathbb{E}\begin{pmatrix} \boldsymbol{R_t} - \lambda_c \boldsymbol{1_N} - \boldsymbol{R_t}(\boldsymbol{f_t} - \boldsymbol{\mu_f})^\top \boldsymbol{\lambda_f} \\ \boldsymbol{f_t} - \boldsymbol{\mu_f} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0_N} \\ \boldsymbol{0_K} \end{pmatrix}, \tag{2.2}$$

with corresponding sample analogue function $\boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}) \equiv \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{g_t}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})$. Combining the latter with a weighting matrix $\boldsymbol{W}$ yields the GMM estimates as the minimizer of the following objective function:

$$\{\widehat{\lambda}_c, \widehat{\boldsymbol{\lambda}}_f, \widehat{\boldsymbol{\mu}}_f\} \equiv \underset{\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}}{\arg \min} \; \boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f})^\top \boldsymbol{W} \boldsymbol{g_T}(\lambda_c, \boldsymbol{\lambda_f}, \boldsymbol{\mu_f}).$$

Different weighting matrices deliver different point estimates. Following (Cochrane, 2009, pp. 256-258), two popular choices are

$$\boldsymbol{W_{ols}} = \begin{pmatrix} \boldsymbol{I_N} & \boldsymbol{0_{N \times K}} \\ \boldsymbol{0_{K \times N}} & \kappa \boldsymbol{I_K,} \end{pmatrix}, \text{ and } \boldsymbol{W_{gls}} = \begin{pmatrix} \boldsymbol{\Sigma_R^{-1}} & \boldsymbol{0_{N \times K}} \\ \boldsymbol{0_{K \times N}} & \kappa \boldsymbol{I_K} \end{pmatrix},$$

where $\boldsymbol{\Sigma_R}$ is the covarince matrix of returns, and $\kappa > 0$ is a large constant so that $\widehat{\boldsymbol{\mu}}_{\boldsymbol{f}} \equiv \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f_t}$. These weighting matrices yield, respectively, the following prices of risk estimates:

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{ols}} = (\widehat{\boldsymbol{C}}^{\top}\widehat{\boldsymbol{C}})^{-1}\widehat{\boldsymbol{C}}^{\top}\bar{\boldsymbol{R}}, \text{ and} \tag{2.3}$$

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{gls}} = (\widehat{\boldsymbol{C}}^{\top}\boldsymbol{\Sigma_R}^{-1}\widehat{\boldsymbol{C}})^{-1}\widehat{\boldsymbol{C}}^{\top}\boldsymbol{\Sigma_R}^{-1}\bar{\boldsymbol{R}}, \tag{2.4}$$

where $\widehat{\boldsymbol{C}} = (\boldsymbol{1_N}, \widehat{\boldsymbol{C}}_{\boldsymbol{f}})$ and $\widehat{\boldsymbol{C}}_{\boldsymbol{f}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R_t}(\boldsymbol{f_t} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{f}})^{\top}$.

GMM provides valid inference on the price of risk under a set of well-known assumptions (Newey and McFadden (1994)). In particular, equations (2.3) and (2.4) make it clear that OLS and GLS (but also GMM more generally) require the matrix of factor exposures $\boldsymbol{C}$ to have full rank – that is, the price of risk to be identified. However, there is a growing body of literature that finds this assumption to be often empirically violated.[6] Most famously, this problem arises in the case of a *weak* factor f_j that does not have enough comovement with any of the assets but is nonetheless considered to be a part of the SDF, that is $C_{i,j} \sim O(T^{-1/2})$, $i \in 1\ldots N$. In such a model, the price of risks are no longer identified and their estimates diverge with the sample size, leading to wrong inference for *both* strong and weak factors (Kan and Zhang (1999a)). Another widespread example of weak identification arises with the inclusion of a *level* factor, f_j, characterized by a lack of cross-sectional spread in factor exposures, that is, $\sum_{i=1}^{N}(C_{i,j} - \bar{C}_j)^2 \sim O(T^{-1})$, where $\bar{C}_j \equiv \frac{1}{N}\sum_{i=1}^{N} C_{i,j}$.

Identification problems arise not only when using the GMM in estimating linear SDF models but equally so in Fama-MacBeth regressions (Kan and Zhang (1999b), Kleibergen (2009)) and Maximum Likelihood Estimation (Gospodinov, Kan, and Robotti (2019)). In addition to creating inference problems for model parameters, weak identification also tends to inflate the standard measures of cross-sectional fit (Kleibergen and Zhan (2015)). Consequently, several papers have attempted to develop alternative statistical procedures that are robust to the presence of weak factors and general cases of rank deficiency of the matrix $\boldsymbol{C}$. In particular, Kleibergen (2009) proposes several novel statistics whose large sample distributions are unaffected by the failure of the identification condition. Gospodinov, Kan, and Robotti (2014) derive robust standard errors for GMM estimates of factor risk prices in the linear stochastic discount factor framework and prove that t-statistics calculated using their standard errors are robust even when the model is misspecified and a weak factor is included. Bryzgalova (2015) introduces a LASSO-like penalty term that identifies weak factors and eliminates their impact on the model. Finally, since factor strength depends on the choice of returns used in the estimation, Giglio, Xiu, and Zhang (2021) recently developed an iterative procedure for constructing a cross-section of model-specific test assets that

---

[6] For recent applications, see Kleibergen and Zhan (2020) and Gospodinov and Robotti (2021a,b).

specifically addresses the problem of weak factors.

In this paper, we provide a Bayesian inference and model selection framework that i) can be easily used for robust inference in the presence, and detection, of weak and level factors (section 2.3) and ii) can be used for both model selection and model averaging, even in the presence of a very large number of candidates (traded or non-traded, and possibly weak) risk factors – that is, the entire factor zoo.

Although we focus on the estimation of linear SDF representations, our approach can be adapted (with minimal adjustments) to deliver a robust Bayesian version of the canonical Fama-MacBeth estimation approach (see Fama and MacBeth (1973) and Fama and French (1993)).

2.3 Bayesian Analysis of Linear SDFs

This section introduces our hierarchical Bayesian estimation of linear SDF models, B-SDF. A more detailed derivation is presented in Appendix 2.A.1.1.

Consider first the time-series dimension of the estimation problem. Let $\boldsymbol{f}_t \equiv (f_{1t} \ldots f_{Kt})^\top$, $t = 1, \ldots T$ denote a vector of factors. Without loss of generality, we order the K_1 tradable factors first ($\boldsymbol{f}_t^{(1)}$), followed by K_2 non-tradable factors ($\boldsymbol{f}_t^{(2)}$), hence, $\boldsymbol{f} \equiv (\boldsymbol{f}_t^{(1),\top}, \boldsymbol{f}_t^{(2),\top})^\top$ and $k_1 + k_2 = K$.

Let $\boldsymbol{Y_t}$ denote the union of factors and returns, that is, $\boldsymbol{Y_t} \equiv \boldsymbol{f_t} \cup \boldsymbol{R_t}$, where $\boldsymbol{Y_t}$ is a p-dimensional vector. If one requires the tradable factors to price themselves (as we do in our empirical applications), then $\boldsymbol{Y_t}^\top \equiv (\boldsymbol{R_t}^\top, \boldsymbol{f}_t^{(2),\top})^\top$ and $p = N + K_2$.

We assume that $\{\boldsymbol{Y_t}\}_{t=1}^T$ follows an iid multivariate Gaussian distribution, that is, $\boldsymbol{Y_t} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, where $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ denote, respectively, the unconditional means vector and the unconditional covariance matrix. This modeling choice can easily be modified to accommodate different distributional assumptions, predictability, and time-varying volatility, albeit at the cost of losing analytical solutions in most cases. In particular, as discussed in Section 2.6, we could accommodate time-varying means and variances, as well as autocorrelations. The resulting likelihood function for the time-series layer of our hierarchical modeling is

$$p(\mathbf{Y}|\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2} tr\left[\boldsymbol{\Sigma_Y}^{-1} \sum_{t=1}^T (\boldsymbol{Y_t} - \boldsymbol{\mu_Y})(\boldsymbol{Y_t} - \boldsymbol{\mu_Y})^\top \right] \right\}, \qquad (2.5)$$

where $\mathbf{Y} \equiv \{\boldsymbol{Y_t}\}_{t=1}^T$. For simplicity, we use the diffuse prior: $\pi(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{p+1}{2}}$. This implies the following posterior distribution of $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$:

$$\boldsymbol{\mu_Y}|\boldsymbol{\Sigma_Y}, \mathbf{Y} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{Y}}, \ \boldsymbol{\Sigma_Y}/T\right), \qquad (2.6)$$

$$\Sigma_Y | \mathbf{Y} \sim \mathcal{W}^{-1}\left(T - 1, \ \sum_{t=1}^{T} \left(\mathbf{Y_t} - \widehat{\boldsymbol{\mu}}_Y\right)\left(\mathbf{Y_t} - \widehat{\boldsymbol{\mu}}_Y\right)^{\top}\right), \qquad (2.7)$$

where $\hat{\boldsymbol{\mu}}_Y \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{Y_t}$ and $\mathcal{W}^{-1}$ is the inverse-Wishart distribution (a multivariate general-ization of the inverse-gamma distribution). Note that the above posterior distribution is well defined even in the presence of weak factors, since the time-series layer does not depend on the strength of the factors or their tradability. Furthermore, the above posterior is analogous to the canonical t-distribution result for the parameters of a linear regression model.

The Normal-inverse-Wishart posterior in equations (2.6)–(2.7) implies that we can sample the distribution of the parameters $(\boldsymbol{\mu}_Y, \Sigma_Y)$ by first drawing the covariance matrix Σ_Y from the inverse-Wishart distribution conditional on the data, and then by drawing $\boldsymbol{\mu}_Y$ from a multivariate normal distribution conditional on the data and the draw of Σ_Y.

If the SDF is correctly specified, in the sense that all true factors are included, expected asset returns should be fully explained by their risk exposure, $\boldsymbol{C}$, and the prices of risk $\boldsymbol{\lambda}$, that is, $\boldsymbol{\mu_R} = \boldsymbol{C\lambda}$, where $\boldsymbol{\mu_R}$ is the sub-vector of $\boldsymbol{\mu}_Y$ corresponding to asset returns and $\boldsymbol{C}$ is the corresponding covariance sub-matrix of Σ_Y. Therefore, we can define our first estimator.[7] In Appendix 2.A.1.1 we show formally that it arises, under the assumption of correct specification, as a particular case of our general posterior presented in equations (2.11)-(2.12) below.

Definition 2.2 (Bayesian SDF (B-SDF) Estimates) *Conditional on $\boldsymbol{\mu}_Y$, Σ_Y and the data $\mathbf{Y} = \{\mathbf{Y_t}\}_{t=1}^{T}$, under the null of unique correct SDF specification[8] and any diffuse prior, the posterior distribution of $\boldsymbol{\lambda}$ is a Dirac distribution (that is, a constant) at $(\boldsymbol{C}^{\top}\boldsymbol{C})^{-1}\boldsymbol{C}^{\top}\boldsymbol{\mu_R}$. Therefore, conditional on only the data $\mathbf{Y} = \{\mathbf{Y_t}\}_{t=1}^{T}$ and the null, the posterior distribution of $\boldsymbol{\lambda}$ can be sampled by drawing $\boldsymbol{\mu}_{Y,(j)}$ and $\Sigma_{Y,(j)}$ from the Normal-inverse-Wishart (2.6)–(2.7) and computing the draw $\boldsymbol{\lambda}_{(j)} \equiv \left(\boldsymbol{C}_{(j)}^{\top}\boldsymbol{C}_{(j)}\right)^{-1}\boldsymbol{C}_{(j)}^{\top}\boldsymbol{\mu}_{R,(j)}$.*

The posterior distribution of $\boldsymbol{\lambda}$, defined above, accounts for both the uncertainty about expected returns – via the sampling of $\boldsymbol{\mu_R}$ – and the uncertainty about the factor loadings – via the sampling of $\boldsymbol{C_f}$. Note that for completeness in the above we have allowed for a common cross-sectional intercept, λ_c. However, this can be readily constrained to be equal to zero, and we consider this case in our empirical analysis.

---

[7]The B-SDF estimator, and its GLS version, as shown in Appendix 2.A.1.1, are particular cases of the more general posterior characterizations in equations (2.11)–(2.12) and (2.13)–(2.14). For expositional purposes we focus on the particular OLS- and GLS-like Bayesian estimators. Nevertheless, for any conformable matrix $\boldsymbol{A}$ such that $\boldsymbol{AC}$ is invertible, we have that under the null of unique correct specification, $\boldsymbol{\lambda}$ has (under any non-dogmatic prior) a degenerated posterior at $(\boldsymbol{AC})^{-1}\boldsymbol{A\mu_R}$ conditional on $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{\mu_R}$.

[8]That is, $\boldsymbol{\mu_R} = \boldsymbol{C\lambda}$ holds for a unique value of $\boldsymbol{\lambda}$ as assumed in standard frequentist estimation.

From the B-SDF definition, it is intuitive why we expect posterior inference to detect weak factors in finite sample. For such factors, the near singularity of $(C_{(j)}^\top C_{(j)})^{-1}$ will cause the draws for $\boldsymbol{\lambda}_{(j)}$ to diverge from zero (as in the frequentist point estimate). Nevertheless, the posterior uncertainty about factor loadings and asset risk premia will cause $C_{(j)}^\top \boldsymbol{\mu}_{R,(j)}$ to switch sign across draws, causing the posterior distribution of $\boldsymbol{\lambda}$ to put substantial probability mass on both values above and below zero. Hence, centered posterior credible intervals will tend to include zero with high probability.

In addition to risk prices $\boldsymbol{\lambda}$, we are also interested in estimating the cross-sectional fit of the model, that is, the cross-sectional R^2. Once we obtain the posterior draws of the parameters, we can easily obtain the posterior distribution of the cross-sectional R^2, defined as

$$R_{ols}^2 = 1 - \frac{(\boldsymbol{\mu_R} - C\boldsymbol{\lambda})^\top (\boldsymbol{\mu_R} - C\boldsymbol{\lambda})}{(\boldsymbol{\mu_R} - \bar{\mu}_R \mathbf{1_N})^\top (\boldsymbol{\mu_R} - \bar{\mu}_R \mathbf{1_N})}, \tag{2.8}$$

where $\bar{\mu}_R = \frac{1}{N} \sum_i^N \mu_{R,i}$. That is, for each posterior draw of $(\boldsymbol{\mu_R}, \; C, \; \boldsymbol{\lambda})$, we can construct the corresponding draw for the R^2 from equation (2.8), hence, tracing out its posterior distribution. Equation (2.8) can be thought of as the population R^2, where $\boldsymbol{\mu_R}$, C, and $\boldsymbol{\lambda}$ are unknown. After observing the data, we infer the posterior distribution of $\boldsymbol{\mu_R}$, C, and $\boldsymbol{\lambda}$, and from these we can recover the distribution of the R^2.

Often the cross-sectional step of the frequentist estimation is performed via GLS rather than least squares. In our setting, under the null of the model, this leads to the following GLS estimator (see Appendix 2.A.1.1 for a formal derivation).

Definition 2.3 (Bayesian SDF GLS (B-SDF-GLS)) *Conditional on $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma_Y}$ and the data $\mathbf{Y} = \{Y_t\}_{t=1}^T$, under the null of unique correct SDF specification and any diffuse prior, the posterior distribution of $\boldsymbol{\lambda}$ is a Dirac distribution (that is, a constant) at $(C^\top \boldsymbol{\Sigma_R}^{-1} C)^{-1} C^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{\mu_R}$. Therefore, conditional on <u>only</u> the data $\mathbf{Y} = \{Y_t\}_{t=1}^T$ and the null, the posterior distribution of $\boldsymbol{\lambda}$ can be sampled by drawing $\boldsymbol{\mu}_{Y,(j)}$ and $\boldsymbol{\Sigma}_{Y,(j)}$ from the Normal-inverse-Wishart (2.6)– (2.7) and computing $\boldsymbol{\lambda}_{(j)} \equiv (C_{(j)}^\top \boldsymbol{\Sigma}_{R,(j)}^{-1} C_{(j)})^{-1} C_{(j)}^\top \boldsymbol{\Sigma}_{R,(j)}^{-1} \boldsymbol{\mu}_{R,(j)}$.*

From the posterior sampling of the parameters in the definition above, we can also obtain the posterior distribution of the cross-sectional GLS R^2, defined as

$$R_{gls}^2 = 1 - \frac{(\boldsymbol{\mu_R} - C\boldsymbol{\lambda})^\top \boldsymbol{\Sigma_R}^{-1} (\boldsymbol{\mu_R} - C\boldsymbol{\lambda})}{(\boldsymbol{\mu_R} - \bar{\mu}_R \mathbf{1_N})^\top \boldsymbol{\Sigma_R}^{-1} (\boldsymbol{\mu_R} - \bar{\mu}_R \mathbf{1_N})}. \tag{2.9}$$

Once again, we can think of equation (2.9) as the unknown population GLS R^2, which is a function of the unknown quantities $\boldsymbol{\mu_R}$, C, and $\boldsymbol{\lambda}$. Since after observing the data we infer the posterior distribution of the parameters, we obtain the posterior distribution of the R_{gls}^2 as well.

Realistically, models are rarely true. Therefore, we now allow for the presence of model-implied *average* pricing errors, $\boldsymbol{\alpha}$.[9] This can be easily accommodated within our Bayesian framework since in this case the data-generating process in the cross-section becomes $\boldsymbol{\mu_R} = \boldsymbol{C\lambda} + \boldsymbol{\alpha}$. Adding an assumption on the cross-sectional distribution of the pricing errors yields a Bayesian hierarchical structure to the estimation that naturally separates the time series and cross-sectional dimensions of the inference problem. To continue the analogy with OLS and GLS estimators, we consider two distributional assumptions for the average pricing errors $\boldsymbol{\alpha}$.

First, we consider the case of spherical cross-sectional errors, that is, $\alpha_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, in the spirit of OLS. Under this assumption, the cross-sectional likelihood function (i.e., conditional on the time-series parameters $\boldsymbol{\mu_R}$ and $\boldsymbol{C}$) is

$$p(data|\boldsymbol{\lambda}, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{ -\frac{1}{2\sigma^2} (\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^\top (\boldsymbol{\mu_R} - \boldsymbol{C\lambda}) \right\}. \tag{2.10}$$

In the cross-sectional regression, the "data" are the expected risk premia, $\boldsymbol{\mu_R}$, and the factor loadings, $\boldsymbol{C}$. These quantities are not directly observable to the researcher but can be sampled from the Normal-inverse-Wishart posterior distribution in equations (2.6)–(2.7). Conceptually, this is not very different from the Bayesian modeling of latent variables. In the benchmark case, we assume a diffuse prior[10] for $(\boldsymbol{\lambda}, \sigma^2)$: $\pi(\boldsymbol{\lambda}, \sigma^2) \propto \sigma^{-2}$. In Appendix 2.A.1.1, we show that the posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ is then

$$\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\mu_R}, \boldsymbol{C} \sim \mathcal{N}\left(\underbrace{(\boldsymbol{C}^\top \boldsymbol{C})^{-1} \boldsymbol{C}^\top \boldsymbol{\mu_R}}_{\hat{\boldsymbol{\lambda}}}, \underbrace{\sigma^2 (\boldsymbol{C}^\top \boldsymbol{C})^{-1}}_{\Sigma_\lambda} \right) \text{ and} \tag{2.11}$$

$$\sigma^2|\boldsymbol{\mu_R}, \boldsymbol{C} \sim \mathcal{IG}\left(\frac{N-K-1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})^\top (\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})}{2} \right), \tag{2.12}$$

where $\mathcal{IG}$ denotes the inverse-Gamma distribution. The conditional distribution in equation (2.11) makes it clear that the posterior takes into account both the uncertainty about prices of risk stemming from the time series parameters $\boldsymbol{C}$ and $\boldsymbol{\mu_R}$ (that are drawn from the Normal-inverse-Wishart posterior in equations (2.6)–(2.7)) and the random pricing errors $\boldsymbol{\alpha}$ that have the conditional posterior variance distribution given in equation (2.12). If test assets' expected excess returns are fully explained by $\boldsymbol{C}$, there are no pricing errors and $\sigma^2 (\boldsymbol{C}^\top \boldsymbol{C})^{-1}$ converges to zero; otherwise, this layer of uncertainty always exists. Similarly,

---

[9]As we show in the next section, this natural assumption is essential for model selection.

[10]As shown in the next subsection, in the presence of weak factors, such a prior is not appropriate for model selection based on Bayes factors and posterior probabilities, since it does not lead to proper marginal likelihoods. Therefore, we introduce therein a novel prior for model selection.

if one assumes that the cross-sectional model is correctly specified, that is, $\sigma^2 \to 0$, we are back to the B-SDF estimator in Definition 2.2.[11]

The OLS assumption ignores the fact that average pricing errors could be cross-sectionally correlated, which motivates our second, non-spherical, cross-sectional distributional assumption for $\boldsymbol{\alpha}$. Suppose that the model is correctly specified, that is, $\boldsymbol{R_t} = \lambda_c \mathbf{1}_N + \boldsymbol{C_f}\boldsymbol{\lambda_f} + \boldsymbol{\epsilon_t}$, where $\boldsymbol{\epsilon_t} \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Sigma_R})$. Since $\mathbb{E}_T[\boldsymbol{R_t}] = \lambda_c \mathbf{1}_N + \boldsymbol{C_f}\boldsymbol{\lambda_f} + \mathbb{E}_T[\boldsymbol{\epsilon_t}]$, the pricing error $\boldsymbol{\alpha}$ should be equal to $\mathbb{E}_T[\boldsymbol{\epsilon_t}]$.[12] Hence, in the spirit of the central limit theorem, a natural distributional assumption for the pricing errors is $\boldsymbol{\alpha} \mid \boldsymbol{\Sigma_R} \sim \mathcal{N}(\mathbf{0}_N, \frac{1}{T}\boldsymbol{\Sigma_R})$. However, since we allow for mispricing, and its degree is endogeously determined by the observed data, a scaling of the covariance matrix is desirable. Therefore, we assign the following distributional assumption for $\boldsymbol{\alpha}$: $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}_N, \sigma^2\boldsymbol{\Sigma_R})$. We call this the GLS assumption. Recall that $\boldsymbol{\Sigma_R}$ is the covariance matrix of returns $\boldsymbol{R_t}$. Hence, the difference between the OLS and GLS assumption is that non-diagonal elements are non-zeros in the latter case. Since all models are misspecified to a certain degree, we would expect that the estimated σ^2 to be larger than $1/T$.

The posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ under the GLS distributional assumption, and conditional on $\boldsymbol{\mu_R}$, $\boldsymbol{\Sigma_R}$ and $\boldsymbol{C}$, is then (see derivation in Appendix 2.A.1.1)

$$\boldsymbol{\lambda}|\sigma^2, data \sim \mathcal{N}\left(\underbrace{(\boldsymbol{C}^\top\boldsymbol{\Sigma_R}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^\top\boldsymbol{\Sigma_R}^{-1}\boldsymbol{\mu_R}}_{\widehat{\boldsymbol{\lambda}}}, \underbrace{\sigma^2(\boldsymbol{C}^\top\boldsymbol{\Sigma_R}^{-1}\boldsymbol{C})^{-1}}_{\Sigma_\lambda}\right) \text{ and} \qquad (2.13)$$

$$\sigma^2|data \sim \mathcal{IG}\left(\frac{N-K-1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}\widehat{\boldsymbol{\lambda}})^\top\boldsymbol{\Sigma_R}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C}\widehat{\boldsymbol{\lambda}})}{2}\right). \qquad (2.14)$$

And once again $\boldsymbol{\mu_R}$, $\boldsymbol{\Sigma_R}$, and $\boldsymbol{C}$ can be sampled from the the Normal-inverse-Wishart posterior in equations (2.6)–(2.7). Furthermore, as before, by setting $\sigma^2 \to 0$ we recover the B-SDF-GLS in Definition 2.3.

Remark 2.2 (Generated factors) *Often factors are estimated, as, for example, in the case of principal components (PCs) and factor-mimicking portfolios (albeit the latter are not needed in our setting). This generates an additional layer of uncertainty normally ignored in empirical analysis due to the associated asymptotic complexities. Nevertheless, thanks their hierarchical structure, it is relatively easy to adjust the above-defined Bayesian estimators to account for this uncertainty. In the case of a mimicking portfolio, under a diffuse prior and Normal errors, the posterior distribution of the portfolio weights follow the standard Normal-inverse-Gamma of Gaussian linear regression models (see, e.g., Lancaster (2004)).*

---

[11] When pricing errors $\boldsymbol{\alpha}$ are assumed to be exactly zero under the null, the posterior distribution of $\boldsymbol{\lambda}$ in equation (2.11) collapses to a degenerate distribution, where $\boldsymbol{\lambda}$ equals $(\boldsymbol{C}^\top\boldsymbol{C})^{-1}\boldsymbol{C}^\top\boldsymbol{\mu_R}$ with probability one.

[12] Where $\mathbb{E}_T$ is the sample analog of the unconditional expectation operator.

Similarly, in the case of principal components as factors, under a diffuse prior, the covariance matrix from which the PCs are constructed follows an inverse-Wishart distribution.[13] Hence, the posterior distributions in Definitions 2.2 and 2.3 can account for the generated factors uncertainty by first drawing from an inverse-Wishart the covariance matrix from which PCs are constructed, or from the Normal-inverse-Gamma posterior of the mimicking portfolios coefficients, and then sampling the remaining parameters as explained above.

Note that while we focus on the case of linear SDF models, our method can be easily extended to the estimation of beta representations of the fundamental pricing equation used in the two-pass procedure, such as Fama-MacBeth regressions.

2.3.1 Model Selection and Aggregation

In the previous subsection we have derived simple Bayesian estimators that deliver, in a finite sample, credible intervals robust to the presence of weak factors and avoid over-rejecting the null hypothesis of zero prices of risk for such factors.

However, given the plethora of risk factors that have been proposed in the literature, a robust approach for model selection, across not necessarily nested models, that can handle a very large universe of possible models, as well as both traded and non-traded factors, is of paramount importance for empirical asset pricing. The canonical way of selecting models and testing hypotheses within the Bayesian framework is through Bayes factors and posterior probabilities, which is the approach we present in this section. This is, for instance, the approach suggested by Barillas and Shanken (2018a) for tradable factors. The key elements of novelty of the proposed method are that: i) our procedure is robust to the presence of weak factors, ii) it is directly applicable to both traded and non-traded factors, and iii) it selects models based on their cross-sectional performance (rather than on the time series), that is, on the basis of the risk prices that the factors command.

Our approach hinges upon the introduction of suitable and economically driven priors that deliver valid marginal likelihoods and posterior model probabilities. With valid posterior probabilities, our framework allows to also *aggregate* multiple candidate factors and specifications into the most likely, given the data, representation of the true unknown SDF (via BMA).[14] Hence, our method endogenously selects a dominant subset of factors – if such a set exists uniquely – and instead aggregates factors optimally, if no dominant low-dimensional representation arises. But, unlike the canonical dichotomy of observable factors

---

[13]Based on these two observations, Allena (2019) proposes a generalization of the Barillas and Shanken (2018a) model comparison approach for these type of factors.

[14]See, e.g., Raftery, Madigan, and Hoeting (1997), and Hoeting, Madigan, Raftery, and Volinsky (1999).

selection versus pure aggregation (e.g., principal component and entropy methods), our approach combines both. In a sense, it jointly delivers model selection and "smart" latent factor extraction.

In this subsection, we show first that flat priors for risk prices are not suitable for model selection in the presence of weak factors. Given the close analogy between frequentist testing and Bayesian inference with flat priors, this is not too surprising. But the novel insight is that the problem arises exactly because of the use of flat priors and can therefore be fixed by using non-flat, yet non-informative, priors. Second, we introduce "spike-and-slab" priors that are robust to the presence of weak factors. These priors allow us to test hypotheses using *valid* Bayes factors and model probabilities. Furthermore, they are particularly powerful in high-dimensional model selection, that is, when one wants, as in our empirical application, to consider all the factors in the zoo. Finally, we show how, as a by-product of the estimation and selection method, factors and models can be optimally aggregated.

2.3.1.1 Pitfalls of Flat Priors for Risk Prices

We start this section by discussing why flat priors for prices of risk are not suitable for model selection. Since we want to focus on and select models based on the cross-sectional asset pricing properties of the factors, for simplicity we retain flat (in the sense of Jeffreys) priors for the time-series parameters $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$.

In order to perform model selection, we relax the (null) hypothesis that models are correctly specified and allow instead for the presence of cross-sectional pricing errors. That is, we consider the cross-sectional representation $\boldsymbol{\mu_R} = \boldsymbol{C\lambda} + \boldsymbol{\alpha}$. For illustrative purposes, we focus on spherical cross-sectional errors (i.e., the case analogous to the GMM-OLS). Nevertheless, all the results in this and following subsections are also generalized to the non-spherical error setting (i.e., the case analogous to the GMM-GLS).

To model variable selection, we introduce a vector of binary latent variables $\boldsymbol{\gamma}^\top = (\gamma_0, \gamma_1, \ldots, \gamma_K)$, where $\gamma_j \in \{0, 1\}$. When $\gamma_j = 1$, factor j (with associated loadings $\boldsymbol{C_j}$) should be included into the model and vice versa. Therefore, the number of included factors is $p_\gamma \equiv \sum_{j=0}^{K} \gamma_j$. Note that we always include the intercept, that is, $\gamma_0 = 1$ always. The notation $\boldsymbol{C_\gamma} = [\boldsymbol{C_j}]_{\gamma_j = 1}$ represents a p_γ-columns sub-matrix of $\boldsymbol{C}$.

When testing whether the risk price of factor j is zero, the null hypothesis is $H_0 : \lambda_j = 0$. In our notation, this null hypothesis can be expressed as $H_0 : \gamma_j = 0$, while the alternative is $H_1 : \gamma_j = 1$. This is a small but important difference relative to the canonical frequentist testing approach: For weak factors, risk prices are not identified; hence, testing whether they are equal to any given value is problematic per se. Nevertheless, as we show in the next section, with appropriate priors, whether a factor should be included or not is a well-defined

question even in the presence of weak factors.

In the Bayesian framework, the prior distribution of parameters under the alternative hypothesis should be carefully specified. Generally speaking, the priors for nuisance parameters, such as $\boldsymbol{\mu_Y}$, $\boldsymbol{\Sigma_Y}$ and σ^2, do not greatly influence the cross-sectional inference. But, as we are about to show, this is not the case for the priors about risk prices.

Recall that when considering multiple models, say, without loss of generality model $\boldsymbol{\gamma}$ and model $\boldsymbol{\gamma}'$, by Bayes theorem we have that the posterior probability of model $\boldsymbol{\gamma}$ is

$$\Pr(\boldsymbol{\gamma}|data) = \frac{p(data|\boldsymbol{\gamma})}{p(data|\boldsymbol{\gamma}) + p(data|\boldsymbol{\gamma}')},$$

where we have given equal prior probability to each model and $p(data|\boldsymbol{\gamma})$ denotes the marginal likelihood of the model indexed by $\boldsymbol{\gamma}$. In Appendix 2.A.1.2 we show that, when using a flat prior for $\boldsymbol{\lambda}$, the marginal likelihood is

$$p(data|\boldsymbol{\gamma}) \propto (2\pi)^{\frac{p_\gamma}{2}} |\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{N-p_\gamma}{2}\right)}{\left(\frac{N\hat{\sigma}_\gamma^2}{2}\right)^{\frac{N-p_\gamma}{2}}}, \tag{2.15}$$

where $\hat{\sigma}_\gamma^2 = \frac{(\boldsymbol{\mu_R} - \boldsymbol{C}_\gamma \hat{\boldsymbol{\lambda}}_\gamma)^\top (\boldsymbol{\mu_R} - \boldsymbol{C}_\gamma \hat{\boldsymbol{\lambda}}_\gamma)}{N}$, $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma)^{-1} \boldsymbol{C}_\gamma^\top \boldsymbol{\mu_R}$, and Γ denotes the Gamma function.

Therefore, if model $\boldsymbol{\gamma}$ includes a weak factor (whose $\boldsymbol{C_j}$ asymptotically converges to zero), the matrix $\boldsymbol{C}_\gamma^\top \boldsymbol{C}_\gamma$ is nearly singular and its determinant goes to zero, sending the marginal likelihood in (2.15) to infinity. As a result, the posterior probability of the model containing the weak factor goes to one.[15] Consequently, under a flat prior for risk prices, the model containing a weak factor will always be selected asymptotically. However, the posterior distribution of $\boldsymbol{\lambda}$ for the weak factor is robust, and particularly disperse, in any finite sample.

Moreover, it is highly likely that conclusions based on the posterior coverage of $\boldsymbol{\lambda}$ contradict those arising from Bayes factors. When the prior distribution of λ_j is too diffuse under the alternative hypothesis H_1, the Bayes factor tends to favor the null H_0, even though the estimate of λ_j is far from 0. The reason is that even though H_0 seems quite unlikely based on posterior coverages, the data are even more unlikely under H_1. Therefore, a disperse prior for λ_j may push the posterior probabilities to favor H_0 and make it fail to identify true factors.[16]

---

[15] Note that a similar problem also arises when using mimicking portfolios of weak factors. In this case the singularity in the determinant in equation (2.15) would be generated by the projection of the non-tradable factors on the space of returns.

[16] This phenomenon is known as the Bartlett Paradox (see Bartlett (1957)).

Note also that flat (hence improper) priors for risk prices are not appropriate, since they render the posterior model probabilities arbitrary. Suppose that we are testing the null $H_0 : \lambda_j = 0$. Under the null hypothesis, the prior for (λ, σ^2) is $\lambda_j = 0$ and $\pi(\lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. However, the prior under the alternative hypothesis is $\pi(\lambda_j, \lambda_{-j}, \sigma^2) \propto \frac{1}{\sigma^2}$. Since the marginal likelihoods of the data, $p(data|H_0)$ and $p(data|H_1)$, are both undetermined, we cannot define the Bayes' factor $\frac{p(data|H_1)}{p(data|H_0)}$ (as stressed in, e.g., Chib, Zeng, and Zhao (2020)). In contrast, for nuisance parameters such as σ^2, we can continue to assign improper priors. Since both hypotheses H_0 and H_1 include σ^2, the prior for it will be offset in the Bayes factor and in the posterior probabilities. Therefore, we can only assign improper priors for common parameters.[17] Similarly, we can still assign improper priors for $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ in the first time-series step.

The final reason why it might be undesirable to use a flat prior for risk prices is that it does not impose any shrinkage on the parameters. This is problematic, given the large number of members of the factor zoo, while we have only limited time-series observations of both factors and test asset returns.

In the next subsection, we propose an appropriate prior for risk prices that is both robust to weak factors and can be used for model selection, even when dealing with a very large number of potential models.

2.3.1.2 Spike-and-Slab Prior for Risk Prices

To ensure that the integration of the marginal likelihood is well-behaved, we propose a novel prior specification for the factors' risk prices $\boldsymbol{\lambda_f}^\top = (\lambda_1, ..., \lambda_K)$. Since the inference in time-series regression is always valid, we only modify the priors of the cross-sectional regression parameters.

This prior belongs to the so-called *spike-and-slab* family. For illustrative purposes, in this section we consider a Dirac spike and show analytically its implications for model selection. In the next subsection we generalize the method to a "continuous spike" prior and study its finite sample performance in our simulation setup.

In particular, we model the uncertainty underlying the model selection problem with a mixture prior, $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}) \propto \pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma})$. When $\gamma_j = 1$, and, hence, the factor should be included in the model, the prior (the "slab") follows a normal distribution, given by $\lambda_j|\sigma^2, \gamma_j = 1 \sim \mathcal{N}(0, \sigma^2\psi_j)$, where ψ_j is a (crucial) quantity that we define below. When instead $\gamma_j = 0$, and the corresponding risk factor should not be included in the model, the prior (the "spike") is a Dirac distribution at zero. For the cross-sectional variance of the

---

[17]See Kass and Raftery (1995) (and also Cremers (2002)) for a more detailed discussion.

pricing errors we keep the canonical diffuse prior:[18] $\pi(\sigma^2) \propto \sigma^{-2}$.

Let $\boldsymbol{D}$ denote a diagonal matrix with elements $c, \psi_1^{-1}, \cdots \psi_K^{-1}$, and $\boldsymbol{D_\gamma}$ the sub-matrix of $\boldsymbol{D}$ corresponding to model $\boldsymbol{\gamma}$, where c is a small positive number corresponding to the common cross-sectional intercept (λ_c). The prior for the prices of risk ($\boldsymbol{\lambda_\gamma}$) of model $\boldsymbol{\gamma}$ is then

$$\boldsymbol{\lambda_\gamma}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D_\gamma}^{-1}).$$

Given this prior, we sample the posterior distribution by sequentially drawing from the conditional distributions of the parameters (i.e., we use a Gibbs sampling approach)[19] presented in the following proposition.

Proposition 2.3 (B-SDF OLS Posterior with Dirac Spike-and-Slab) *The posterior distribution of $(\lambda_\gamma, \sigma^2, \boldsymbol{\gamma})$ under the assumption of Dirac spike-and-slab prior and spherical $\boldsymbol{\alpha}$ (OLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (2.6)–(2.7), is given by the following conditional distributions:*

$$\boldsymbol{\lambda_\gamma}|data, \sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}_\gamma, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma)\right), \tag{2.16}$$

$$\sigma^2|data, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N}{2}, \frac{SSR_\gamma}{2}\right), \text{ and} \tag{2.17}$$

$$p(\boldsymbol{\gamma} \mid data) \propto \frac{|\boldsymbol{D_\gamma}|^{\frac{1}{2}}}{|\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma}|^{\frac{1}{2}}} \frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}}, \tag{2.18}$$

where $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top \mu_R}$, $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2(\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma})^{-1}$, and $SSR_\gamma = \boldsymbol{\mu_R^\top \mu_R} - \boldsymbol{\mu_R^\top C_\gamma}(\boldsymbol{C_\gamma^\top C_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top \mu_R} = \min_{\lambda_\gamma}\{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma \lambda_\gamma})^\top(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma \lambda_\gamma}) + \boldsymbol{\lambda_\gamma^\top D_\gamma \lambda_\gamma}\}$ and $\mathcal{IG}$ denotes the inverse-Gamma distribution.

Proposition 2.4 (B-SDF GLS Posterior with Dirac Spike-and-Slab) *The posterior distribution of $(\lambda_\gamma, \sigma^2, \boldsymbol{\gamma})$ under the assumption of Dirac spike-and-slab prior and and non-spherical $\boldsymbol{\alpha}$ (GLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (2.6)–(2.7), is given by the following conditional distributions:*

$$\boldsymbol{\lambda_\gamma}|data, \sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}_\gamma, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma)\right), \tag{2.19}$$

---

[18]Note that since the parameter σ is common across models and has the same support in each model, the marginal likelihoods obtained under this improper prior are valid and comparable (see Proposition 1 of Chib, Zeng, and Zhao (2020)).

[19]We do not standardize $\boldsymbol{Y_t}$ in the time-series regression. In the empirical implementation, after obtaining posterior draws for $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$, we calculate $\boldsymbol{\mu_R}$ and $\boldsymbol{C_f}$ as the standardized expected returns of test assets and correlation between test assets and factors. Then $\boldsymbol{C}$ is a matrix containing a vector of ones and $\boldsymbol{C_f}$.

$$\sigma^2 | data, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N}{2}, \frac{SSR_{\gamma}}{2}\right), \text{ and} \tag{2.20}$$

$$p(\boldsymbol{\gamma} \mid data) \propto \frac{|\boldsymbol{D}_{\gamma}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma}/2)^{\frac{N}{2}}}, \tag{2.21}$$

where $\hat{\boldsymbol{\lambda}}_{\gamma} = (\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})^{-1}\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}}$, $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_{\gamma}) = \sigma^2(\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})^{-1}$, and $SSR_{\gamma} =$ $\boldsymbol{\mu}_{\boldsymbol{R}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{\mu}_{\boldsymbol{R}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma}(\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})^{-1}\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{\mu}_{\boldsymbol{R}} = \min_{\lambda_{\gamma}}\{(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma})^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma}) + \boldsymbol{\lambda}_{\gamma}^{\top}\boldsymbol{D}_{\gamma}\boldsymbol{\lambda}_{\gamma}\}$ and $\mathcal{IG}$ denotes the inverse-Gamma distribution.

The above propositions are proved, respectively, in Appendices 2.A.1.3 and 2.A.1.4.

Note that SSR_{γ} is the minimized sum of squared errors under the spherical pricing errors assumption, and is instead the minimized squared Sharpe ratio of pricing errors in the non-spherical case, where the term $\boldsymbol{\lambda}_{\gamma}^{\top}\boldsymbol{D}_{\gamma}\boldsymbol{\lambda}_{\gamma}$ is akin to a generalized ridge regression penalty.

Our prior modeling is analogous to introducing a Tikhonov-Phillips regularization (see Tikhonov, Goncharsky, Stepanov, and Yagola (1995) and Phillips (1962)) in the cross-sectional regression step, and has the same rationale: delivering a well-defined marginal likelihood in the presence of rank deficiency (which, in our setting, arises in the presence of weak factors).

The key element and novelty of our method is that the "shrinkage" applied to the factors is endogenously heterogeneous and designed to target weak factors: It leverages the correlation between factors and returns by setting ψ_j as

$$\psi_j = \psi \times \boldsymbol{\rho_j}^{\top}\boldsymbol{\rho_j}, \tag{2.22}$$

where $\boldsymbol{\rho_j}$ is an $N \times 1$ vector of correlation coefficients between factor j and the test assets, and $\psi \in \mathbb{R}_+$ is a tuning parameter that controls the degree of shrinkage over all factors.[20] But, unlike tuning parameters in frequentist inference, as we show below, ψ is uniquely pinned down by the researcher's beliefs about Sharpe ratios being achievable in the economy.

When the correlation between f_{jt} and $\boldsymbol{R_t}$ is very low, as in the case of a weak factor, the penalty for λ_j, which is the reciprocal of $\psi\boldsymbol{\rho_j}^{\top}\boldsymbol{\rho_j} \equiv (\{\boldsymbol{D}_{\gamma}\}_{jj})^{-1}$, is very large and dominates the sum of squared errors.

Equation (2.16) (and, similarly, equation (2.19)) makes clear why this Bayesian formulation is robust to weak factors. When $\boldsymbol{C}$ converges to zero, $(\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})$ is dominated by $\boldsymbol{D}_{\gamma}$, so the identification condition for the prices of risk no longer fails. When a factor

---

[20]Alternatively, we could have set $\psi_j = \psi \times \boldsymbol{C_j}^{\top}\boldsymbol{C_j}$, where $\boldsymbol{C_j}$ is a $N \times 1$ vector of covariances of the test assets with factor j. However, $\boldsymbol{\rho_j}$ has the advantage of being invariant to the units in which factors are measured. Furthermore, in the empirical analysis the cross-sectional step is implemented using returns and factors scaled by their standard deviations, making the distinction immaterial.

is weak, its correlation with test assets converges to zero; hence, the penalty for this factor, ψ_j^{-1}, goes to infinity. As a result, the posterior mean of $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\gamma}} = (\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}})^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\mu}_{\boldsymbol{R}}$, is shrunk toward zero, and the posterior variance term $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}})$ approaches $\sigma^2 \boldsymbol{D}_{\boldsymbol{\gamma}}^{-1}$. Consequently, the posterior distribution of $\boldsymbol{\lambda}$ for a weak factor is nearly the same as its prior. In contrast, for a normal factor that has non-zero covariance with test assets, the information contained in $\boldsymbol{C}$ dominates the prior information, since in this case the absolute size of $\boldsymbol{D}_{\boldsymbol{\gamma}}$ is small relative to $\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}}$.

Remark 2.5 (Level Factors) *Identification failure of factors' risk prices can arise in the presence of "level factors," that is factors to which asset returns have non-zero exposure but lack cross-sectional spread. These factors help explain the average level of returns but not their cross-sectional dispersion, and, hence, are collinear with the common cross-sectional intercept. Our approach can handle this case by using variance standardized variables in the cross-sectional part of the estimation and replacing the penalty in (2.22) with*

$$\psi_j = \psi \times \widetilde{\boldsymbol{\rho}}_j^{\top}\widetilde{\boldsymbol{\rho}}_j, \tag{2.23}$$

where $\widetilde{\boldsymbol{\rho}}_j \equiv \boldsymbol{\rho}_j - \left(\frac{1}{N}\sum_{i=1}^{N}\rho_{j,i}\right) \times \boldsymbol{1_N}$ is the cross-sectionally demeaned vector of factor j correlations with asset returns.

When comparing two models, using posterior model probabilities for specification selection is equivalent to simply using the ratio of the marginal likelihoods, that is, the Bayes factor, which is defined as

$$BF_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = p(data|\boldsymbol{\gamma})/p(data|\boldsymbol{\gamma}'),$$

where we have given equal prior probability[21] to model $\boldsymbol{\gamma}$ and model $\boldsymbol{\gamma}'$.

Corollary 2.1 shows that, unlike in the flat prior case discussed earlier, under the Dirac spike, the Bayes factors (and posterior probabilities) are well-defined even in the presence of weak factors. Therefore, they can be used for model selection and hypotheses testing.

Corollary 2.1 (Model Selection via the Bayes Factor) *Consider two nested linear factor models, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. The only difference between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ is γ_p: γ_p equals 1 in model $\boldsymbol{\gamma}$ but 0 in model $\boldsymbol{\gamma}'$. Let $\boldsymbol{\gamma}_{-p}$ denote a $K \times 1$ vector of model index excluding γ_p: $\boldsymbol{\gamma}^{\top} = (\boldsymbol{\gamma}_{-p}^{\top}, 1)$ and $\boldsymbol{\gamma}'^{\top} = (\boldsymbol{\gamma}_{-p}^{\top}, 0)$ where, without loss of generality, we have assumed that the factor p is ordered last.*

---

[21]The corollary can be trivially extended to the case of different prior probabilities for the two models, since in this case the Bayes factor is simply the ratio of marginal likelihoods multiplied by the prior odds.

Under the spherical assumption for $\boldsymbol{\alpha}$ (OLS), the Bayes factor is

$$BF_{\gamma,\gamma'} = \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left(1 + \psi_p \boldsymbol{C_p}^\top \left[\boldsymbol{I_N} - \boldsymbol{C_{\gamma'}}(\boldsymbol{C_{\gamma'}}^\top \boldsymbol{C_{\gamma'}} + \boldsymbol{D_{\gamma'}})^{-1} \boldsymbol{C_{\gamma'}}^\top\right] \boldsymbol{C_p}\right)^{-\frac{1}{2}}, \qquad (2.24)$$

where $SSR_\gamma = \boldsymbol{\mu_R}^\top \boldsymbol{\mu_R} - \boldsymbol{\mu_R}^\top \boldsymbol{C_\gamma}(\boldsymbol{C_\gamma}^\top \boldsymbol{C_\gamma} + \boldsymbol{D_\gamma})^{-1} \boldsymbol{C_\gamma}^\top \boldsymbol{\mu_R} = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma}) + \boldsymbol{\lambda_\gamma}^\top \boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$. *Under the non-spherical assumption for* $\boldsymbol{\alpha}$ *(GLS), the Bayes factor is*

$$BF_{\gamma,\gamma'} = \left(\frac{SSR_{\gamma'}}{SSR_\gamma}\right)^{\frac{N}{2}} \left|1 + \psi_p\left[\boldsymbol{C_p}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_p} - \boldsymbol{C_p}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_{\gamma'}} \left(\boldsymbol{C_{\gamma'}}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_{\gamma'}} + \boldsymbol{D_{\gamma'}}\right)^{-1} \boldsymbol{C_{\gamma'}}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_p}\right]\right|^{-\frac{1}{2}}.$$
$$(2.25)$$

where $SSR_\gamma = \boldsymbol{\mu_R}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{\mu_R} - \boldsymbol{\mu_R}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_\gamma}(\boldsymbol{C_\gamma}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C_\gamma} + \boldsymbol{D_\gamma})^{-1} \boldsymbol{C_\gamma}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{\mu_R} = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top$
$\boldsymbol{\Sigma_R}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma}) + \boldsymbol{\lambda_\gamma}^\top \boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$.

The proof can be found in Appendix 2.A.1.5.

Since $\boldsymbol{C_p}^\top[\boldsymbol{I_N} - \boldsymbol{C_{\gamma'}}(\boldsymbol{C_{\gamma'}}^\top \boldsymbol{C_{\gamma'}} + \boldsymbol{D_{\gamma'}})^{-1}\boldsymbol{C_{\gamma'}}^\top]\boldsymbol{C_p}$ is always positive, ψ_p plays an important role in variable selection. For a strong and useful factor that can substantially reduce pricing errors, the first term in equation (2.24) dominates, and the Bayes factor will be much greater than 1, hence, providing evidence in favor of model $\boldsymbol{\gamma}$.

Recall that $SSR_\gamma = \min_{\boldsymbol{\lambda_\gamma}}\{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma}) + \boldsymbol{\lambda_\gamma}^\top \boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$, hence, we always have $SSR_\gamma \leq SSR_{\gamma'}$ in sample. There are two effects of increasing ψ_p: i) when ψ_p is large, the penalty for λ_p is small, hence, it is easier to minimize SSR_γ, and $SSR_{\gamma'}/SSR_\gamma$ becomes much larger than 1; ii) large ψ_p decreases the second term in equation (2.24), lowering the Bayes factor, and acting as a penalty for dimensionality.

A particularly interesting case is when the factor added by model $\boldsymbol{\gamma}$ is weak: $\boldsymbol{C_p}$ converges to zero, but the penalty term $1/\psi_p \propto 1/\boldsymbol{\rho_p}^\top \boldsymbol{\rho_p}$ goes to infinity. On the one hand, the first term in equation (2.24) will converge to 1; on the other hand, since $\psi_p \approx 0$ in large sample, the second term in equation (2.24) will also be around 1. Therefore, the Bayes factor for a weak factor will go to 1 asymptotically.[22] In contrast, a useful factor should be able to greatly reduce the sum of squared errors SSR_γ, so the Bayes factor will be dominated by SSR_γ, yielding a value substantially above 1.

Note that since our prior restores the validity of the marginal likelihood, *any* hypothesis on the parameters (e.g., whether the pricing errors are jointly zero) can be tested via posterior probabilities or, equivalently, Bayesian p-values. In particular, we obtain closed-form solutions for testing hypothesis about prices of risk by centering the Dirac spike at the null

---

[22]But in finite sample it may deviate from its asymptotic value, so we should not use 1 as a threshold when testing the null hypothesis $H_0 : \gamma_p = 0$.

value rather than at zero.

Corollary 2.2 (Hypothesis Testing for Risk Prices (Bayesian p-values)) *Suppose that we want to test the point hypothesis $\boldsymbol{\lambda}_{-\gamma} = \tilde{\boldsymbol{\lambda}}_{-\gamma}$ and as before we have the prior $\boldsymbol{\lambda}_{\gamma}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D}_{\gamma}^{-1})$ in model $\boldsymbol{\gamma}$. In this case, the posterior distributions in Propositions 2.3 and 2.4 still hold with SSR_{γ}, therein replaced by $\widetilde{SSR}_{\gamma}$, defined below.*

Under the spherical assumption for $\boldsymbol{\alpha}$ (OLS),

$$\begin{aligned}
\widetilde{SSR}_{\gamma} &= (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) - \\
&\quad (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top\boldsymbol{C}_{\gamma}(\boldsymbol{C}_{\gamma}^\top\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})^{-1}\boldsymbol{C}_{\gamma}^\top(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) \\
&= \min_{\boldsymbol{\lambda}_{\gamma}}\{(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma})^\top(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma}) + \boldsymbol{\lambda}_{\gamma}^\top\boldsymbol{D}_{\gamma}\boldsymbol{\lambda}_{\gamma}\},
\end{aligned}$$

where $\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} \equiv \boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}$ denotes the vector of cross-sectional residual expected returns that are unexplained by factors $f_{-\gamma}$ with prices of risk $\tilde{\boldsymbol{\lambda}}_{-\gamma}$.

Under the non-spherical assumption for $\boldsymbol{\alpha}$ (GLS),

$$\begin{aligned}
\widetilde{SSR}_{\gamma} &= (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) - \\
&\quad (\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma})^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma}(\boldsymbol{C}_{\gamma}^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\gamma} + \boldsymbol{D}_{\gamma})^{-1}\boldsymbol{C}_{\gamma}^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C}_{-\gamma}\tilde{\boldsymbol{\lambda}}_{-\gamma}) \\
&= \min_{\boldsymbol{\lambda}_{\gamma}}\{(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma})^\top\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}(\tilde{\boldsymbol{\mu}}_{\boldsymbol{R}} - \boldsymbol{C}_{\gamma}\boldsymbol{\lambda}_{\gamma}) + \boldsymbol{\lambda}_{\gamma}^\top\boldsymbol{D}_{\gamma}\boldsymbol{\lambda}_{\gamma}\},
\end{aligned}$$

A Bayesian p-value for the null hypothesis is then constructed by integrating $1-p(\boldsymbol{\gamma} \mid data)$ in equation (2.18) (equation (2.21) in the case of spherical (non-spherical) pricing errors), with respect to the Normal-inverse-Wishart in equations (2.6)–(2.7).

The proof of the corollary follows the same steps as the proofs of Propositions 2.3 and 2.4 in Appendices 2.A.1.3 and 2.A.1.4.

Corollary 2.2 can be used for joint hypothesis testing within the Bayesian framework (e.g., building confidence intervals), and it is very similar in spirit to the standard frequentist identification-robust inference.

2.3.1.3 Continuous Spike

We extend the Dirac spike-and-slab prior by encoding a continuous spike for λ_j, when γ_j equals 0. While the closed-form solutions obtained with a Dirac spike allow to feasibly evaluate *millions* of models, this extension allows to efficiently sample *quadrillions* of alternative specifications.

Following the literature on Bayesian variable selection (see, e.g., George and McCulloch (1993, 1997) and Ishwaran, Rao, et al. (2005)), we model the uncertainty underlying model

selection with a mixture prior $\pi(\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega}) = \pi(\boldsymbol{\lambda} \mid \sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)\pi(\boldsymbol{\gamma} \mid \boldsymbol{\omega})\pi(\boldsymbol{\omega})$, where

$$\lambda_j \mid \gamma_j, \sigma^2 \sim \mathcal{N}\left(0, r(\gamma_j)\psi_j\sigma^2\right). \tag{2.26}$$

Note the introduction of two new elements, $r(\gamma_j)$ and $\pi(\boldsymbol{\omega})$, in the prior. When the factor should be included, $r(\gamma_j = 1) = 1$, hence we have the same "slab" as before. When the factor should not be in the model $r(\gamma_j = 0) = r \ll 1$. Hence the Dirac "spike" is replaced by a Gaussian spike, which is extremely concentrated at zero (we set $r = 0.001$ in our empirical analysis). Note that in this case ψ_j has an effect on the spike, but given a small value for r this effect is virtually immaterial. As we explain below, the additional prior $\pi(\boldsymbol{\omega})$ encodes our ex ante beliefs about the sparsity of the true model in terms of observable factors.

We now redefine $\boldsymbol{D}$ as a diagonal matrix with elements $c, \ (r(\gamma_1)\psi_1)^{-1}, \ldots, (r(\gamma_K)\psi_K)^{-1}$, where ψ_j is given as before by equation (2.22). In matrix notation, the prior for $\boldsymbol{\lambda}$ is therefore: $\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}^{-1})$. The term $r(\gamma_j)\psi_j$ in $\boldsymbol{D}^{-1}$ is set to be small or large, depending on whether $\gamma_j = 0$ or $\gamma_j = 1$. In the empirical implementation, we set r to a value much smaller than 1 since we intend to shrink λ_j toward zero when γ_j is 0. Hence, the spike component concentrates the posterior mass of $\boldsymbol{\lambda}$ around zero, whereas the slab component allows $\boldsymbol{\lambda}$ to take values over a much wider range. Therefore, the posterior distribution of $\boldsymbol{\lambda}$ is very similar to the case of a Dirac spike in section 2.3.1.2.

Furthermore, this prior encodes beliefs about the fraction of the total Sharpe ratio of the test assets ascribable to the factors and to the pricing errors. To see this, consider the case in which (as in our empirical applications) both factors and returns are standardized. It then follows that

$$\frac{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \boldsymbol{\gamma}, \sigma^2]}{\mathbb{E}_\pi[SR_{\boldsymbol{\alpha}}^2 \mid \sigma^2]} = \frac{\sum_{k=1}^K r(\gamma_k)\psi_k}{N} = \frac{\psi \sum_{k=1}^K r(\gamma_k)\tilde{\boldsymbol{\rho}}_k^\top\tilde{\boldsymbol{\rho}}_k}{N}, \tag{2.27}$$

where $SR_{\boldsymbol{f}}$ and $SR_{\boldsymbol{\alpha}}$ denote, respectively, the Sharpe ratios of all factors[23] ($\boldsymbol{f_t}$) and of the pricing errors of all assets ($\boldsymbol{\alpha}$), and $\mathbb{E}_\pi$ denotes prior expectations. In the baseline sample of our empirical applications, $\sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top\tilde{\boldsymbol{\rho}}_k/N \simeq 3.22$.[24] Hence, for ψ in the 1–5 range, if, say, 50% of the factors are selected, our prior expectation is that the factors should explain about 62%–89% of the squared Sharpe ratio of test assets.

The prior $\pi(\boldsymbol{\omega})$ not only gives us a way of sampling from the space of potential models,

---

[23]The squared Sharpe ratio implied by the SDF is $\boldsymbol{\lambda}_{\boldsymbol{f}}^\top\boldsymbol{\Sigma}_{\boldsymbol{f}}\boldsymbol{\lambda}_{\boldsymbol{f}}$. Since $\boldsymbol{\lambda}_{\boldsymbol{f}}$ are assumed to be independently distributed in the prior level, $\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \boldsymbol{\gamma}, \sigma^2]$ is equal to $\sum_{k=1}^K \mathbb{E}_\pi[\lambda_k^2 \mid \gamma_k, \sigma^2]$.

[24]Note that in our previous study (where the cross-section was 25 Fama-French size and B/M portfolios plus 30 industry portfolios) $\sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top\tilde{\boldsymbol{\rho}}_k/N \simeq 0.51$. In that case, for ψ in the 10–20 range, if, say, 50% of the factors are selected, our prior expectation is that the factors should explain about 71%–83% of the squared Sharpe ratio of test assets.

but also encodes belief about the sparsity of the true model using the prior distribution $\pi(\gamma_j = 1|\omega_j) = \omega_j$. Following the literature on predictors selection, we set:

$$\pi(\gamma_j = 1|\omega_j) = \omega_j, \quad \omega_j \sim Beta\,(a_\omega, b_\omega).$$

Different hyper-parameters a_ω and b_ω determine whether one a priori favors more parsimonious models or not.[25] Furthermore, a_ω and b_ω can be chosen to encode prior beliefs about the Sharpe ratio achievable in the economy since $\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2] = \frac{a_\omega}{a_\omega + b_\omega}\psi\sigma^2\sum_{k=1}^K \tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$ as $r \to 0$.

The considerations above imply that an agent's expectations about the Sharpe ratio achievable i) with only *one* factor, ii) with *all* the factors jointly, and iii) the sparsity of the "true" model, *uniquely* determine the parameters ψ, a_ω, b_ω.[26]

When ω_j is constant and equal to 0.5 and r converges to 0, the continuous spike-and-slab prior is equivalent to the one with Dirac spike in Section 2.3.1.2. Instead, treating ω_j (hence, γ_j), as a parameter to be sampled is particularly useful in high-dimensional cases. For instance, suppose that there are 30 candidate factors. With the Dirac spike-and-slab prior we have to calculate the posterior model probabilities for 2^{30} different models. Given that we update $(\boldsymbol{\mu_R}, \boldsymbol{C_f})$ at each sampling round, posterior probabilities for all models are re-computed for every new draw of these quantities, rendering the computational cost very large. In contrast, with the continuous spike-and-slab approach one can simply use the posterior mean of γ_j to estimate the posterior marginal probability of the j-th factor, since they are the same quantity.

Similar to the Dirac spike-and-slab case, we use sequential sampling from the conditional distributions of the parameters $(\boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2)$ and, most importantly, $\boldsymbol{\gamma}$, as presented in the following propositions.

Proposition 2.6 (B-SDF OLS Posterior with Continuous Spike-and-Slab) *The posterior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma^2)$ under the assumption of continuous spike-and-slab prior and spherical $\boldsymbol{\alpha}$ (OLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (2.6)–(2.7), is given by the following conditional distributions:*

$$\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})\right), \tag{2.28}$$

---

[25]The prior expected probability of selecting a factor is simply $\frac{a_\omega}{a_\omega + b_\omega}$. We set $a_\omega = b_\omega = 1$ in the benchmark case; that is, each factor has an ex ante expected probability of being selected equal to 50%. However, we could for instance, set $a_\omega = 1$ and $b_\omega \gg 1$ in order to favor a sparser model.

[26]For a discussion on the importance of using priors on observables and economic quantities, rather than deep model parameters, see Jarociński and Marcet (2019).

$$\frac{p(\gamma_j = 1 | data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})}{p(\gamma_j = 0 | data, \boldsymbol{\lambda}, \boldsymbol{\omega}, \sigma^2, \boldsymbol{\gamma_{-j}})} = \frac{\omega_j}{1 - \omega_j} \frac{p(\lambda_j | \gamma_j = 1, \sigma^2)}{p(\lambda_j | \gamma_j = 0, \sigma^2)}, \tag{2.29}$$

$$\omega_j | data, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \sigma^2 \sim Beta\left(\gamma_j + a_\omega, 1 - \gamma_j + b_\omega\right), \ and \tag{2.30}$$

$$\sigma^2 | data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N + K + 1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^\top(\boldsymbol{\mu_R} - \boldsymbol{C\lambda}) + \boldsymbol{\lambda}^\top \boldsymbol{D\lambda}}{2}\right), \tag{2.31}$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^\top \boldsymbol{C} + \boldsymbol{D})^{-1} \boldsymbol{C}^\top \boldsymbol{\mu_R}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{C}^\top \boldsymbol{C} + \boldsymbol{D})^{-1}$.

Proposition 2.7 (B-SDF GLS Posterior with Continuous Spike-and-Slab) *The posterior distribution of $(\boldsymbol{\lambda}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma^2)$ under the assumption of continuous spike-and-slab prior and non-spherical $\boldsymbol{\alpha}$ (GLS), conditional on the draws of $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ from equations (2.6)–(2.7), differs from ones in Proposition 2.6 only for the posterior distributions of $(\boldsymbol{\lambda}, \sigma^2)$:*

$$\boldsymbol{\lambda} | data, \sigma^2, \boldsymbol{\gamma}, \boldsymbol{\omega} \sim \mathcal{N}\left(\hat{\boldsymbol{\lambda}}, \hat{\sigma}^2(\hat{\boldsymbol{\lambda}})\right), \ and \tag{2.32}$$

$$\sigma^2 | data, \boldsymbol{\omega}, \boldsymbol{\lambda}, \boldsymbol{\gamma} \sim \mathcal{IG}\left(\frac{N + K + 1}{2}, \frac{(\boldsymbol{\mu_R} - \boldsymbol{C\lambda})^\top \boldsymbol{\Sigma_R}^{-1}(\boldsymbol{\mu_R} - \boldsymbol{C\lambda}) + \boldsymbol{\lambda}^\top \boldsymbol{D\lambda}}{2}\right), \tag{2.33}$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C} + \boldsymbol{D})^{-1} \boldsymbol{C}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{\mu_R}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}) = \sigma^2(\boldsymbol{C}^\top \boldsymbol{\Sigma_R}^{-1} \boldsymbol{C} + \boldsymbol{D})^{-1}$.

The proofs of the above propositions are reported in Appendix 2.A.1.6.

2.3.1.4 Selection vs. Aggregation

The posterior probabilities of models and factor obtained above with spike-and-slab priors, can be used not only for model selection but also efficient aggregation using *all* possible specification.

If we are interested in some quantity Δ that is well-defined for every model $m = 1, ..., \bar{m}$ (e.g., price of risk, risk premia, and maximum Sharpe ratio), from the Bayes theorem we have

$$\mathbb{E}\left[\Delta | data\right] = \sum_{m=0}^{\bar{m}} \mathbb{E}\left[\Delta | data, model = m\right] \Pr\left(model = m | data\right), \tag{2.34}$$

where $\mathbb{E}\left[\Delta | data, model = m\right] = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} \Delta(\theta_l^{(m)})$ and $\left\{\theta_l^{(m)}\right\}_{l=1}^{L}$ denote L draws from the posterior distribution of the parameters of model m. That is, the BMA expectation of Δ, conditional on only the data is simply the weighted average of the expectation in every model, with weights equal to the models' posterior probabilities (see, e.g., Raftery, Madigan, and Hoeting (1997), and Hoeting, Madigan, Raftery, and Volinsky (1999)).

The BMA efficiently aggregates information about Δ over the space of all models, rather than conditioning on a particular model. At the same time, *if* a dominant model exists – hence it has posterior probability approaching one – the BMA will use that model alone.

For each model $\boldsymbol{\gamma}$ that one could construct with the universe of factors, we have the corresponding SDF: $M_{\boldsymbol{\gamma},t} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma},t} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma},t}])^{\top} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}$. Therefore, one can construct a BMA of the SDF using the model posterior probabilities derived in the previous sections. Note that these probabilities are based upon the ability of the factors and models to explain the cross-section of asset return; that is, they explicitly target the key property of a valid SDF. Aggregation is particularly appealing when multiple candidate factors load on the same underlying sources of risk (plus factor-specific noise). Crucially, BMA creates a weighted average that endogenously maximizes the SDF signal-to-noise ratio for cross-sectional pricing.

The BMA *is* the optimal aggregation procedure for a very wide spectrum of optimality criteria, and, in particular, it is optimal under the quadratic loss function and is "optimal on average", that is, no alternative estimator can beat the BMA for all values of the true unknown parameters (see, e.g., Raftery and Zheng (2003), and Schervish (1995)). Furthermore, the BMA predictive distribution minimizes the Kullback-Leibler information divergence relative to the true unknown data generating process. Hence, it delivers the *most likely* SDF given the data, and the estimated density is as close as possible to the true unknown one, even if all the models considered are misspecified.

A powerful feature of the BMA method is that equation (2.34) can be evaluated by generating a Markov Chain over the space of possible models. This is exactly what the continuous spike-and-slab method allows us to do: We sample models in the unrestricted space of 2.25 quadrillion specifications, computing all the desired quantities of interest for each specification sampled, and then aggregate the results. The Markov Chain endogenously over-samples the more likely specifications and under-samples the ones that are less likely to have generated the observed data. The Markov Chain can then be stopped when the posterior means of interest have converged according to the standard tests. We use as a convergence criterion the Separate Partial Mean test (see, e.g., Geweke (2005)) for *each* factor specific parameter (i.e., posterior probability and price of risk).

Recent literature has usually pursued either *selection* (see, e.g., Feng, Giglio, and Xiu (2020)) or *aggregation* (see, e.g., Kozak, Nagel, and Santosh (2020)) of pricing factors. Our approach, instead, combines *both*. The BMA-SDF includes both factors that are clear drivers of asset returns, that is, factors with posterior probability of inclusion ($\Pr[\gamma_j = 1|\text{data}]$) approaching 1, and also an optimal combination of factors that are, given the data, individually less salient.

2.4   Simulation

We build a simple setting for a linear factor model that includes both strong and weak factors and allows for potential model misspecification.

The cross-section of asset returns mimics the empirical properties of 25 Fama-French portfolios sorted by size and value. We generate both factors and test asset returns from normal distributions, assuming that HML is the only useful factor. A misspecified model also includes pricing errors from the GMM-OLS estimation, which makes the vector of simulated expected returns equal to their sample mean estimates of 25 Fama-French portfolios. Finally, a useless factor is simulated from an independent normal distribution with mean zero and standard deviation 1%. In summary,

$$f_{t,useless} \stackrel{iid}{\sim} \mathcal{N}(0, (1\%)^2), \qquad \begin{pmatrix} \boldsymbol{R_t} \\ f_{t,hml} \end{pmatrix} \stackrel{iid}{\sim} \mathcal{N} \left(\begin{bmatrix} \bar{\boldsymbol{\mu}}_{\boldsymbol{R}} \\ \bar{f}_{hml} \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}} & \hat{\boldsymbol{C}}_{\boldsymbol{hml}} \\ \hat{\boldsymbol{C}}_{\boldsymbol{hml}}^{\top} & \hat{\sigma}_{hml}^2 \end{bmatrix} \right), \text{ and}$$

$$\boldsymbol{\mu_R} = \begin{cases} \hat{\lambda}_c \mathbf{1}_{\boldsymbol{N}} + \hat{\boldsymbol{C}}_{\boldsymbol{f}} \hat{\lambda}_{HML}, & \text{if the model is correct, and} \\ \bar{\boldsymbol{R}}, & \text{if the model is misspecified,} \end{cases}$$

where factor loadings, risk prices, and variance-covariance matrix of returns and factors are equal to their sample estimates from the time series and cross-sectional regressions of the GMM-OLS procedure, applied to 25 size-and-value portfolios and HML as a factor. All the model parameters are estimated on monthly data from July 1963 to December 2017.

To illustrate the properties of the frequentist and Bayesian approaches, we consider three estimation setups: (a) the model includes only a strong factor (HML), (b) the model includes only a useless factor as a stylized example for a weak factor, and (c) the model includes both strong and useless factors. Each setting can be correctly or incorrectly specified, with the following sample sizes: T = 100, 200, 600, 1,000, and 20,000. We compare the performance of the OLS/GLS standard frequentist and Bayesian SDF estimators (GMM and B-SDF, respectively) with the focus on risk prices recovery, testing, and identification of strong and useless factors for model comparison.

2.4.1   B-SDF Estimation of Risk Prices

In this section we focus on the most realistic (and challenging) model setup, which includes both useless and strong factors and allows for model misspecification. We found similar performance of the B-SDF approach in a wide range of alternative simulation settings (e.g., considering correctly specified models and cross-sections of different dimensions).[27]

---

[27]These additional results are reported in Appendix 2.A.3.

Table 2.1 compares the performance of frequentist and Bayesian estimators of the price of risk and reports their empirical test size and confidence intervals for cross-sectional R^2. In the case of the Bayesian estimation we report results for both the flat and normal priors for the price of risk (the latter, in a single stand-alone model case, corresponds to the spike-and-slab approach). Since the model is misspecified, true cross-sectional R^2 has the population value of 43.87% (6.69%) for OLS (GLS)). In the case of the standard GMM approach, tests are constructed using standard t-statistics, and in the case of the B-SDF we rely on the quantiles of the posterior distribution to form the credible confidence intervals. The last two columns also report the quantiles of the posterior distribution of the R^2 mode across the simulations, corresponding to the peak of the cross-sectional likelihood.

As expected, in the conventional frequentist estimation, the useless factor is often found to be a significant predictor of the asset returns: Its OLS (GLS) t-statistic would be above a 5%-critical value in more than 60% (87%) of the simulations in the asymptotic case of $T = 20,000$. On the contrary, the Bayesian confidence intervals detect the useless factor and reject the null of zero price of risk attached to the useless factor with frequency asymptotically approaching the size of the tests independently from the prior.

The presence of useless factors can also bias parameter estimates for the strong ones and often leads to their *crowding out* from the model. Panel A in Table 2.1 serves as a good illustration of this possibility, with the GMM price of risk estimates for the strong factor clearly biased due to the weak identification problem. In this case B-SDF provides reliable, albeit conservative in the case of the flat prior, confidence bounds for model parameters effectively restore statistical inference. Note that the empirical size of the B-SDF (normal prior) credible confidence intervals is very close to the nominal one even for relatively small sample sizes.

Why does the Bayesian approach work while the frequentist one fails? The argument is probably best illustrated by Figure 2.1, which plots posterior distributions of B-SDF $\hat{\lambda}$ for both strong and useless factors from one of the simulations, along with their pseudo-true values of the price of risk (defined as 0 for the useless factor).

In this particular simulation, GMM estimates of $\lambda_{useless}$ imply significant price of risk for both OLS and GLS versions of the weight matrix, with traditional hypothesis testing rejecting the null of $\lambda_{useless} = 0$, even at 1% significance level. Instead, the B-SDF posteriors (blue lines in Figure 2.1) of the useless factor price of risk are diffuse and centered around 0. Intuitively, the main driving force behind it is the fact that in B-SDF, $\boldsymbol{C}$ (the covariance of factors with returns) is updated continuously: When $\hat{\boldsymbol{C}}$ is close to zero, the posterior draws of $\boldsymbol{C}$ will be randomly positive or negative, which implies that the conditional expectation of $\boldsymbol{\lambda}$ in equation 2.11 will also switch sign from draw to draw. As a result, the posterior

Table 2.1: Price of risk tests in a misspecified model with useless and strong factors

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
		Panel A: OLS										
GMM–$\boldsymbol{W_{ols}}$	100	0.083	0.033	0.007	0.065	0.03	0.005	0.082	0.029	0.004	-4.35%	70.21%
	200	0.084	0.039	0.006	0.058	0.025	0.003	0.119	0.047	0.006	-2.38%	69.17%
	600	0.075	0.034	0.009	0.074	0.032	0.005	0.255	0.140	0.024	8.42%	67.27%
	1000	0.078	0.03	0.004	0.070	0.031	0.001	0.311	0.181	0.048	16.85%	65.40%
	20000	0.066	0.019	0.001	0.052	0.022	0.001	0.752	0.585	0.288	36.92%	58.64%
B-SDF, flat prior	100	0.037	0.015	0.001	0.032	0.007	0.001	0.003	0.001	0.000	16.62%	49.24%
	200	0.054	0.021	0.002	0.036	0.013	0.001	0.006	0.001	0.000	13.54%	54.05%
	600	0.053	0.027	0.005	0.047	0.015	0.002	0.019	0.006	0.001	14.72%	58.72%
	1000	0.059	0.027	0.004	0.050	0.018	0.000	0.040	0.013	0.002	19.57%	58.85%
	20000	0.015	0.005	0.000	0.010	0.003	0.000	0.089	0.043	0.009	39.19%	52.86%
B-SDF, normal prior	100	0.062	0.029	0.005	0.047	0.019	0.002	0.003	0.001	0.000	7.47%	43.43%
	200	0.084	0.04	0.008	0.067	0.031	0.005	0.006	0.002	0.000	3.66%	48.19%
	600	0.087	0.048	0.018	0.093	0.044	0.010	0.019	0.006	0.001	4.87%	54.33%
	1000	0.094	0.052	0.011	0.106	0.051	0.010	0.040	0.013	0.002	9.64%	54.13%
	20000	0.100	0.050	0.011	0.102	0.052	0.009	0.088	0.043	0.009	34.47%	46.84%
		Panel B: GLS										
GMM–$\boldsymbol{W_{gls}}$	100	0.095	0.048	0.007	0.076	0.035	0.004	0.146	0.070	0.012	-7.66%	20.08%
	200	0.104	0.051	0.008	0.086	0.045	0.007	0.235	0.142	0.031	-6.97%	19.19%
	600	0.090	0.045	0.009	0.105	0.047	0.008	0.433	0.326	0.163	-4.81%	20.93%
	1000	0.096	0.044	0.010	0.106	0.054	0.008	0.535	0.444	0.273	-3.38%	19.52%
	20000	0.084	0.034	0.006	0.091	0.037	0.009	0.889	0.865	0.807	1.42%	19.32%
B-SDF, flat prior	100	0.114	0.061	0.011	0.046	0.020	0.001	0.029	0.009	0.000	-1.99%	9.64%
	200	0.094	0.050	0.012	0.056	0.023	0.003	0.034	0.012	0.001	-3.04%	10.27%
	600	0.090	0.045	0.008	0.066	0.028	0.004	0.068	0.029	0.004	-2.31%	12.68%
	1000	0.080	0.036	0.007	0.071	0.026	0.002	0.075	0.035	0.007	-1.10%	12.98%
	20000	0.017	0.002	0.000	0.013	0.004	0.002	0.105	0.050	0.011	3.43%	12.65%
B-SDF, normal prior	100	0.133	0.070	0.014	0.054	0.023	0.002	0.029	0.008	0.000	-3.50%	7.72%
	200	0.111	0.057	0.018	0.075	0.033	0.006	0.034	0.012	0.001	-5.08%	7.24%
	600	0.105	0.061	0.013	0.093	0.047	0.008	0.068	0.029	0.004	-5.30%	7.85%
	1000	0.108	0.055	0.014	0.099	0.049	0.010	0.075	0.035	0.007	-4.42%	7.86%
	20000	0.090	0.046	0.010	0.113	0.057	0.009	0.105	0.050	0.011	0.62%	4.10%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda_{useless}^* \equiv 0$ in a misspecified model with an intercept, a strong and a useless factor. The true value of the cross-sectional R^2_{adj} is 43.87% (6.69%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a (annualized) prior SR of the factor model equal to 1.239, 1.305, 1.386, 1.413, and 1.497 for $T \in \{100, 200, 600, 1,000,$ and $20,000\}$.

distribution of $\lambda_{useless}$ is centered around 0, and so is its confidence interval. The same logic applies to both OLS and GLS B-SDF formulations. Note that the Bayesian prior does not have any significant impact on the price of risk estimation of strong factors: In the case of well-identified sources of risk (Figure 2.1, panels (b) and (d)), the Bayesian and frequentist approach give very similar results.

Our setting also allows us to perform formal hypothesis testing via posterior probabilities and Bayes factors, following Corollary 2.2, even as $T \to \infty$, using the spike-and-slab prior of Section 2.3.1.2. We report corresponding simulation results for the Bayesian p-value in Appendix 2.A.3.2. Figure 2.A.1 shows that useless factors are easily detected (their
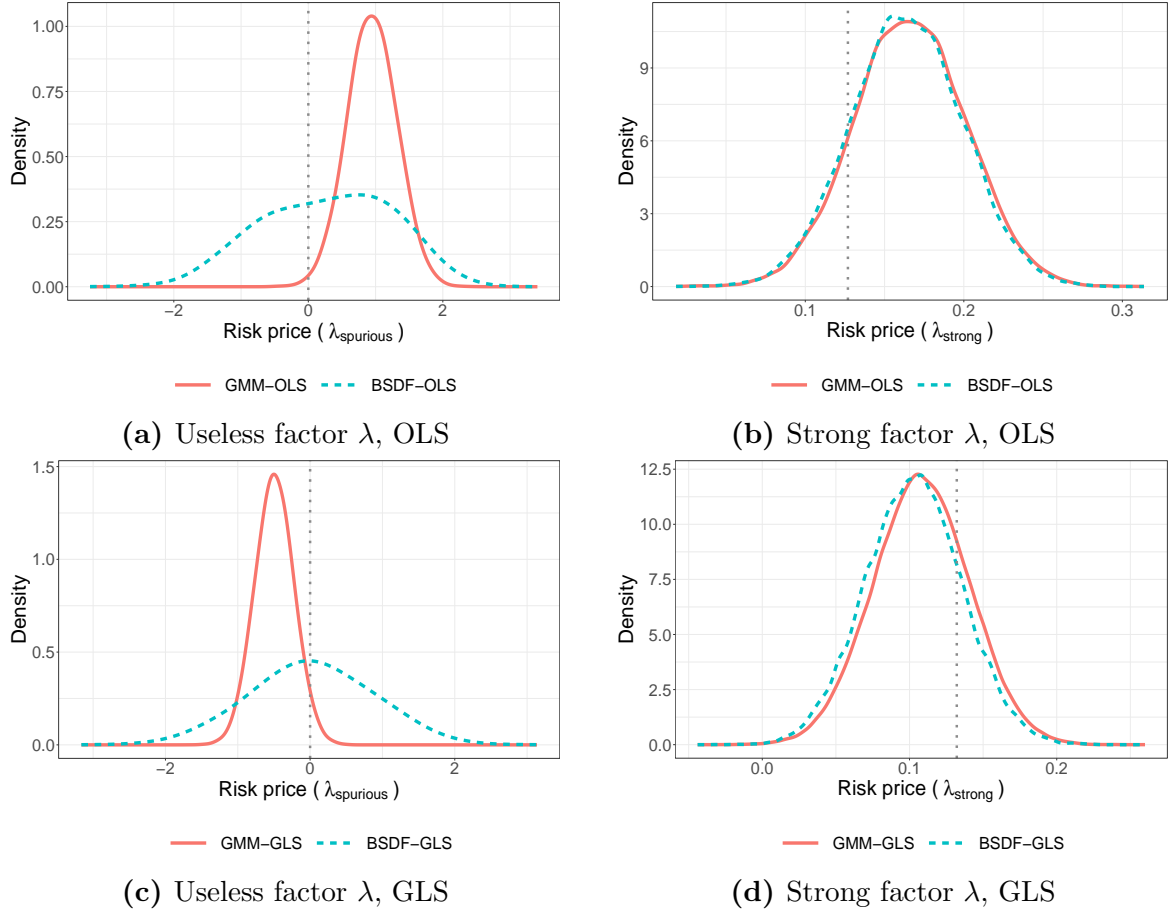
Figure 2.1: Distribution of the price of risk estimates.

Posterior distribution of the price of risk (blue dashed line) from B-SDF estimation of a misspecified one-factor model based on a single simulation with $T = 1000$ and asymptotic distribution of the frequentist GMM estimate (red solid line). The dotted line corresponds to the pseudo-true value of the parameter (defined to be 0 for a useless factor). Panels (a) and (c) correspond to the estimation of a model including a single useless factor. Panels (b) and (d) correspond to the case of including a single strong, well-identified factor.

$p-$values, as expected, are sharply concentrated around the prior inclusion probability of 50% for any sample size), while true sources of risk are successfully selected with probability fast approaching 1.

2.4.1.1 Evaluating Cross-Sectional Fit

Weak identification notoriously affects not only parameter estimates but also conventional measures of fit, such as cross-sectional R^2 (Kleibergen and Zhan (2015)). We now show that the B-SDF approach restores not only inference on the price of risk but also the validity of the measures of cross-sectional fit.

Figure 2.2 shows the distribution of cross-sectional R^2 across a large number of simulations for the asymptotic case of $T = 20,000$ and a misspecified process for returns. For

(a) OLS-type estimation   **(b)** GLS-type estimation
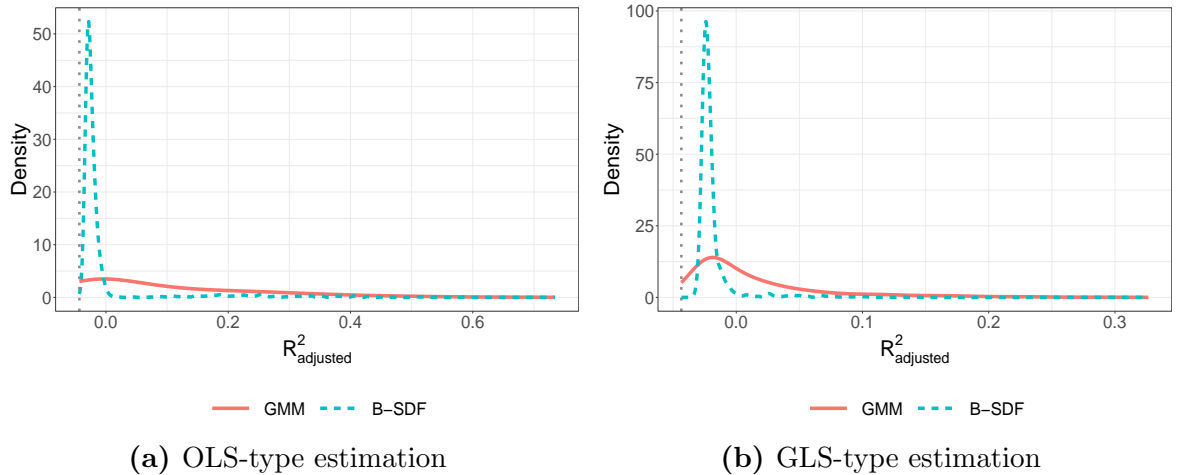
Figure 2.2: Cross-sectional distribution of OLS R^2_{adj} in a model with a useless factor.

Empirical distribution of cross-sectional R^2 achieved by a misspecified model with a useless factor across 2,000 simulations of sample size $T = 20,000$. Blue dashed lines correspond to the distribution of the posterior mode for R^2_{adj}, while red solid lines depict the pointwise sample distribution of R^2_{adj} evaluated at the frequentist GMM estimates. The grey dotted line stands for the true value of R^2_{adj}.

brevity, we focus on the most illustrative case of a single useless factor in the model. In this case frequentist estimation yields an extremely spreadout distribution of R^2 across simulations, which makes the researcher likely to conclude that the useless factor actually has significant explanatory power in the cross-section of returns.[28] This unfortunate property of the frequentist approach is not shared by our hierarchical Bayesian approach: The mode of the posterior distribution is tightly concentrated (across simulations) in the proximity of the true R^2 value.

However, the pointwise distribution of cross-sectional R^2 across the simulations is only part of the story, as it does not reveal the in-sample estimation uncertainty and whether the confidence intervals are credible in reflecting it. While B-SDF incorporates this uncertainty directly into the shape of its posterior distribution, one needs to rely on bootstrap-like algorithms to build a similar analogue in the frequentist case. As a frequentist benchmark, we use the approach of Lewellen, Nagel, and Shanken (2010) to construct the confidence interval.

Figure 2.3 presents the posterior distribution of cross-sectional R^2 for a model that contains a useless factor and contrasts it with the frequentist values and their confidence intervals. The true adjusted R^2 is marginally negative, yet not only are its frequentist estimates economically large (29% and 19% for the OLS and GLS estimation types, respectively), but also the standard approach of Lewellen, Nagel, and Shanken (2010) yields extremely wide

---

[28]Gospodinov, Kan, and Robotti (2019) show examples of perfect fit obtainable with artificially generated useless factors and a family of one-step estimators.

(a) OLS-type estimation   **(b)** GLS-type estimation
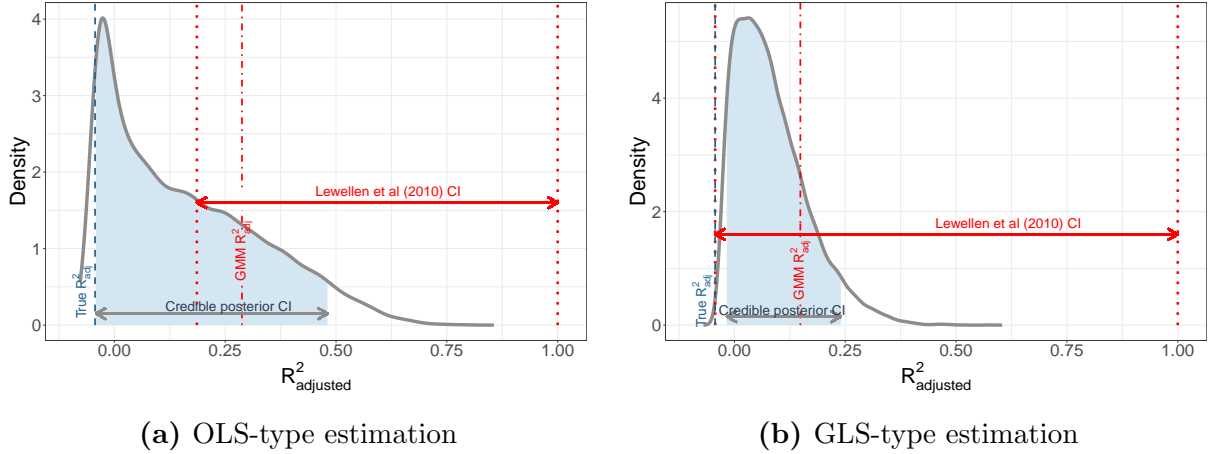
Figure 2.3: The estimation uncertainty of cross-sectional R^2.

Posterior densities of cross-sectional R^2_{adj} in one representative simulation with centered 90% confidence interval (shaded area). The blue dashed line denotes the true R^2_{adj}. The red dashed-dotted line depicts FM R^2_{adj} estimate with 90% Lewellen, Nagel, and Shanken (2010) confidence intervals (red dotted lines).

confidence intervals. Interestingly, they include a level of fit up to 100%, but not the true value. In contrast, while there is still considerable estimation uncertainty, the posterior distribution of the adjusted R^2 peaks in the proximity of 0 and is concentrated on much lower values. As shown in the last two columns of Table 2.1, this is a general property of the B-SDF estimation across simulation designs, sample sizes, and types of prior.

The B-SDF estimator performed well in a wide range of additional simulations that we have conducted. In particular, in Section 2.A.3.1 of the Appendix we show that the B-SDF-based inference stays reliable even in the presence of what is typically considered a large cross-section (100 portfolios). This is reassuring, as it implies that our estimator does not require any specific adjustments for applications with either small time-series dimension or a large cross-sectional one (unlike popular frequentist alternatives).

2.4.2 Selection via Bayes Factors

How well do flat and spike-and-slab priors work empirically in selecting relevant and detecting useless factors in the cross-section of asset returns? We revisit the theoretical results from Section 2.3 using the same simulation design therein.

We consider a misspecified model with both strong and useless factors and compute Bayes factors, corresponding to each of the potential sources of risk. Table 2.2 reports the empirical frequency of variable retention in the model across 2,000 simulations of different sample sizes ($T = 200$, 600, and 1,000). We first report the probability of retaining a factor under a flat prior, which is standard in the literature. Second, we use the continuous spike-and-slab prior

Table 2.2: The probability of retaining risk factors using Bayes factors

T		55%	57%	59%	61%	63%	65%		55%	57%	59%	61%	63%	65%
					Panel A: Flat prior									
200	f_{strong}:	0.636	0.602	0.570	0.538	0.509	0.470	$f_{useless}$:	0.980	0.950	0.856	0.724	0.581	0.437
600		0.821	0.802	0.784	0.764	0.733	0.710		0.996	0.983	0.970	0.932	0.878	0.791
1,000		0.880	0.850	0.840	0.840	0.800	0.800		1.000	1.000	0.990	0.980	0.940	0.910
				Panel B: Spike-and-Slab, prior of $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]} = 0.295$										
200	f_{strong}:	0.815	0.761	0.721	0.675	0.630	0.581	$f_{useless}$:	0.004	0.000	0.000	0.000	0.000	0.000
600		0.974	0.961	0.954	0.943	0.926	0.899		0.000	0.000	0.000	0.000	0.000	0.000
1,000		0.980	0.970	0.970	0.960	0.960	0.940		0.000	0.000	0.000	0.000	0.000	0.000
				Panel C: Spike-and-Slab, prior of $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]} = 0.807$										
200	f_{strong}:	0.527	0.489	0.449	0.412	0.381	0.349	$f_{useless}$:	0.041	0.007	0.004	0.000	0.000	0.000
600		0.859	0.832	0.811	0.774	0.734	0.690		0.001	0.000	0.000	0.000	0.000	0.000
1,000		0.910	0.910	0.870	0.850	0.830	0.820		0 .000	0.000	0.000	0.000	0.000	0.000

Frequency of retaining risk factors using BF for different samples size (T=200, 600, and 1,000) across 2,000 simulations of a misspecified model with strong and useless factors. A factor is retained if its posterior probability, $\Pr(\gamma_i = 1|data)$, is greater than a given threshold: 55%, 57%, 59%, 61%, 63%, and 65%. Returns and factors are standardized. Panel A reports results for the flat prior. Panels B and C use the spike-and-slab approach of Section 2.3.1.3 with demeaned correlations, $r = 0.001$ and $\psi = 1$ or 10, mapping into the corresponding monthly Sharpe ratios, $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]}$, listed in the table. The prior for each factor inclusion in Panels B and C is a $Beta(1,1)$, yielding a prior expectation for factor inclusion of 50%.

for the price of risk and compute the marginal probability of each factor as the posterior mean of γ_j. The decision rule is based on a range of critical values, 55%–65%, such that when the posterior factor probability ($\Pr[\gamma_j = 1|data]$) is above a particular threshold, we retain the factor.

The difference generated by the two priors is drastic in the presence of useless factors. As discussed in Section 2.3.1.1, under a flat prior for the price of risk, the posterior probability of including a useless factor in the model converges to 1 asymptotically. Table 2.2 makes it clear that the same holds even for a very short sample, making the overall process of model selection completely invalid. In turn, factor selection via spike-and-slab prior approach of Section 2.3.1.3 is reliable in both retaining strong factors and excluding useless ones (even with a very small sample size). As Panels B and C indicate, our results also remain robust to different prior values for the factor Sharpe ratio.

Overall, we find the behavior of the spike-and-slab prior very encouraging for variable and model selection: It successfully eliminates the impact of the useless factors from the model and identifies the true sources of risk.

2.5 Empirical Analysis

In this section we apply our hierarchical Bayesian method to a large set of factors proposed in the previous literature. First, we consider 51 tradable and non-tradable factors, yielding more than two quadrillion possible models, and employ our spike-and-slab priors to compute

factors' posterior probabilities and implied prices of risk (Section 2.5.1). Second, based on the results of this estimation, in Section 2.5.2 we construct an SDF via Bayesian Model Averaging and show its superior asset pricing properties. Following Martin and Nagel (2021), we consider not only in-sample but also out-of-sample performance (both in the time-series and cross-sectional dimension) and compare the BMA-SDF with both notable reduced-form models and the shrinkage-based approach to factor aggregation (Kozak, Nagel, and Santosh (2020)). Finally, in Sections 2.5.3 and 2.5.4 we study whether one can achieve an accurate representation of the SDF with low-dimensional (observable) factor models, and show that such conjecture is not supported by the data. Strikingly, our results indicate that there is scope for both selection and aggregation in linear factor models.

2.5.1   Sampling Two Quadrillion Models

We now turn our attention to a large cross-section of candidate asset pricing factors. In particular, we focus on 51 (both tradable and non-tradable) monthly factors available from October 1973 to December 2016 (i.e. $T \simeq 600$). Factors are described in Table 2.A.1 in the Appendix, with additional details available in Table OA13 of the Online Appendix.

As test assets we consider a cross-section of 60 asset returns that are meant to capture well-documented cross-sectional anomalies. These include all the (34) tradable (long-short) factors in Table 2.A.1 and additional 26 long-short portfolios based on the univariate sorting of the characteristics listed in Table 2.A.2 of Appendix 2.A.2. The inclusion of the tradable factors among the test assets and the usage of the non-spherical pricing error formulation (i.e., GLS) also imposes (asymptotically) the restriction of factors pricing themselves.[29]

Since we do not restrict the maximum number of factors to include, all the possible combinations of factors give us a total of 2^{51} possible specifications, that is 2.25 quadrillion models. We use the continuous spike-and-slab approach of Section 2.3.1.3 with non-spherical errors, since it easily handles a very large number of possible models while remaining valid in the presence of the most common identification failures. We report both posterior probabilities (given the data) of each factor (i.e., $\mathbb{E}[\gamma_j|\text{data}], \forall j$) as well as the posterior means of the factors' price of risk (i.e., $\mathbb{E}[\lambda_j|\text{data}], \forall j$) computed as the Bayesian Model Average (BMA) across the universe of models. We use the formulation of the penalty term ψ_j in equation (2.23) in order to also handle identification failures of factors' price of risk caused by level factors (see Remark 2.5).[30]

---

[29]Note that we could also have enforced this pricing restriction in finite sample using an ad hoc prior for these factors – which is analogous to estimating the model via the GLS version of the beta representation of expected returns, and then inverting the estimates to obtain the price of risk of the SDF formulation.

[30]In Appendix 2.A.4 we report results based on the formulation in equation (2.22)). The findings therein are very similar to the ones discussed below. Table 2.A.15 reports the values of the squared correlations,
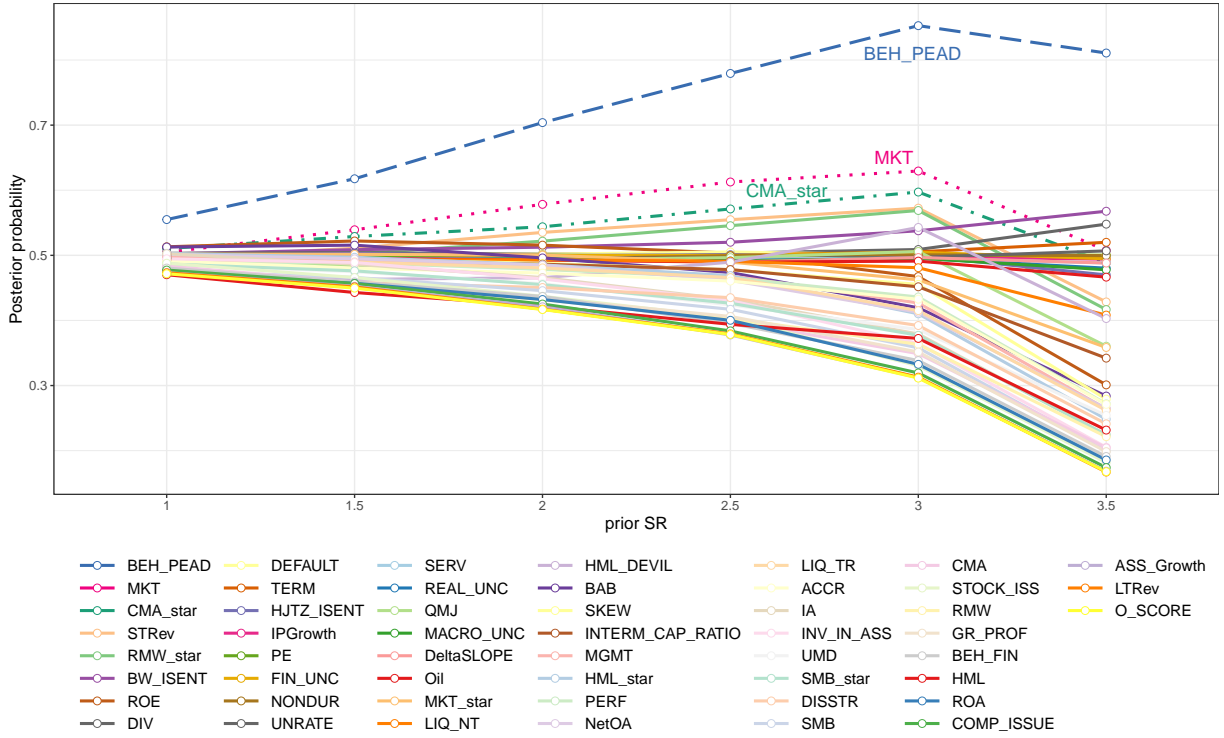
Figure 2.4: Posterior factor probabilities.

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|data]$, computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors described in Table 2.A.1 of the Appendix. Sample: 1973:10–2016:12. Test assets: 60 anomaly portfolios. Prior distribution for the j-th factor inclusion is a $Beta(1,1)$, yielding a 0.5 prior expectation for γ_j. Posterior probabilities for different values of the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]}$, annualized.

The posterior evaluation is performed and reported over a wide range for the parameter ψ (in equation (2.23)) that regulates the degree of shrinkage of potentially useless factors. This parameter controls the prior belief about the Sharpe ratio achievable with the pricing factors. We tabulate the results in units of *Sharpe ratio prior* defined as $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]}$, since this is a natural metric of beliefs. The lower value that we consider, a prior SR of 1, generates a strong shrinkage (small ψ), while the highest value reported, a prior SR of 3.5, makes the shrinkage virtually irrelevant. Since our prior gives non-zero probability to any SR value, these are *not* hard constraints.

The prior probability for each factor inclusion is drawn from a $Beta(1,1)$ (i.e., a uniform on $[0,1]$), yielding a prior expectation for γ_j equal to 50%. That is, a priori we have maximum uncertainty about whether a factor should be included or not.[31]

---

and their cross-sectionally demeaned version, of factors and test assets.

[31]We obtain virtually identical results using a $Beta(2,2)$, which still implies a prior probability of factor inclusion of 50% but lower probabilities for very dense and very sparse models. Furthermore, using a prior in favor of more sparse factor models (a $Beta(1,9)$), the empirical findings are very similar to the ones reported. These additional results are reported in Section 2.A.4 of the Appendix.

Figure 2.4 plots the posterior probabilities of the 51 factors as a function of the maximum SR in the 95% prior support. The corresponding values are reported in Table 2.3.

First, there is particularly strong evidence for including the BEH_PEAD factor of Daniel, Hirshleifer, and Sun (2020), or (behavioral) post-earnings announcement drift anomaly, as the source of priced risk in the SDF. This factor is meant to capture investors' limited attention. The posterior probability of this factor being part of the SDF is over 70% for most prior values. This might not be too surprising, given that many anomaly portfolios seem to be associated with short-term market inefficiencies.

Second, the excess return on the market (MKT) appears as a likely source of priced risk posterior probability in excess of 60% across prior specifications. This is both surprising and reassuring. Surprising, since the market return is rarely found to be significant for cross-sectional asset pricing. It is reassuring because Giglio and Xiu (2021) show that once inference is corrected for potential misspecification, the market factor appears to be priced. In our setting, estimation across all the universe of possible models is meant exactly to address the misspecification problem, and it seems to do so successfully.

Third, CMA* factor of Daniel, Mota, Rottke, and Santos (2020) shows a non-trivial increase in the posterior probability of being part of the SDF. This is the investment factor of Fama and French (2015) without its unpriced component.

Fourth, there are three more factors (RMW*, STRev, and RMW*, described in Table 2.A.1) for which the posterior probability estimate provide some (albeit not strong) support.

Fifth, there is a substantial set of factors for which the posterior probability stays roughly equal to the prior one. That is, these factors are likely to be weakly identified at best. Finally, there is a large set of factors that is unlikely to be part of the SDF pricing our data (e.g., long-short portfolios sorted by the Ohlson O-score, long-term reversal, and asset growth).

Interestingly, the results are not very sensitive to the choice of prior maximum Sharpe ratio unless there is almost no shrinkage, that is, there is no protection against weakly identified factors. In this latter case, weakly identified factors seem to drive out the statistical support for likely components of the true SDF, which is consistent with the findings of Gospodinov, Kan, and Robotti (2014) for the frequentist estimation of linear factor models.

In addition to the posterior probabilities of the factors, Table 2.3 reports the posterior means of the price of risk computed as Bayesian Model Average (BMA), that is, the weighted average of the posterior means in each possible factor model specification, with weights equal to the posterior probability of each specification being the true data-generating process (see, e.g., Roberts (1965), Geweke (1999), and Madigan and Raftery (1994)).

Several observations are in order. First, the price risk estimates for factors that are more likely to be part of the SDF (top three to six factors), the estimates are relatively stable

Table 2.3: Posterior factor probabilities, $\mathbb{E}\left[\gamma_j | \text{data}\right]$, and risk prices: 2.25 quadrillion models

| | Factor inclusion prob., $\mathbb{E}[\gamma_j|\text{data}]$ | | | | | | Price of risk, $\mathbb{E}[\lambda_j|\text{data}]$ | | | | | |
| | Total prior SR | | | | | | Total prior SR | | | | | |
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
BEH_PEAD	0.555	0.618	0.704	0.779	0.853	0.811	0.018	0.043	0.085	0.146	0.231	0.278
MKT	0.505	0.539	0.578	0.613	0.630	0.508	0.017	0.040	0.073	0.114	0.170	0.186
CMA*	0.510	0.529	0.544	0.571	0.597	0.488	0.011	0.023	0.041	0.067	0.106	0.117
STRev	0.496	0.511	0.535	0.555	0.572	0.428	0.007	0.018	0.036	0.060	0.093	0.090
RMW*	0.499	0.502	0.522	0.546	0.569	0.417	0.009	0.020	0.038	0.065	0.105	0.099
BW_ISENT	0.502	0.509	0.512	0.520	0.538	0.568	0.002	0.005	0.009	0.016	0.035	0.122
ROE	0.513	0.522	0.516	0.503	0.467	0.301	0.021	0.039	0.056	0.075	0.093	0.077
DIV	0.503	0.504	0.502	0.503	0.509	0.548	0.000	0.001	0.002	0.004	0.009	0.042
DEFAULT	0.501	0.501	0.502	0.505	0.501	0.500	0.000	0.001	0.001	0.003	0.006	0.022
TERM	0.501	0.498	0.498	0.500	0.505	0.520	0.000	-0.001	-0.002	-0.004	-0.008	-0.037
HJTZ_ISENT	0.499	0.503	0.500	0.501	0.499	0.470	0.001	0.002	0.003	0.005	0.009	0.029
IPGrowth	0.501	0.501	0.500	0.496	0.498	0.494	0.000	0.000	-0.001	-0.002	-0.004	-0.014
PE	0.497	0.497	0.500	0.498	0.500	0.500	0.000	-0.001	-0.002	-0.003	-0.007	-0.029
FIN_UNC	0.494	0.491	0.500	0.500	0.505	0.495	0.001	0.002	0.003	0.007	0.016	0.050
NONDUR	0.494	0.493	0.495	0.499	0.501	0.500	0.001	0.001	0.003	0.005	0.012	0.051
UNRATE	0.496	0.494	0.496	0.495	0.497	0.507	0.000	0.001	0.002	0.003	0.008	0.038
SERV	0.493	0.495	0.494	0.495	0.495	0.488	0.000	0.000	0.001	0.001	0.003	0.018
REAL_UNC	0.496	0.495	0.493	0.492	0.495	0.480	0.000	0.000	0.001	0.002	0.005	0.010
QMJ	0.492	0.484	0.493	0.496	0.506	0.360	0.016	0.030	0.050	0.081	0.132	0.128
MACRO_UNC	0.496	0.493	0.495	0.491	0.496	0.478	0.000	0.000	0.001	0.001	0.003	0.001
DeltaSLOPE	0.494	0.495	0.493	0.490	0.497	0.488	0.000	0.001	0.001	0.002	0.004	0.016
Oil	0.498	0.495	0.493	0.490	0.491	0.467	0.000	0.000	0.001	0.002	0.005	0.021
MKT*	0.502	0.502	0.500	0.490	0.462	0.358	0.007	0.015	0.024	0.034	0.043	0.057
LIQ_NT	0.492	0.493	0.493	0.491	0.481	0.408	0.000	0.001	0.000	-0.002	-0.010	-0.026
HML_DEVIL	0.471	0.463	0.466	0.490	0.543	0.403	0.008	0.017	0.036	0.073	0.152	0.163
BAB	0.513	0.516	0.496	0.474	0.419	0.284	0.015	0.027	0.037	0.046	0.052	0.049
SKEW	0.493	0.494	0.488	0.478	0.455	0.279	0.013	0.027	0.043	0.061	0.082	0.061
INTERM_CAP_RATIO	0.496	0.491	0.486	0.478	0.452	0.342	0.006	0.013	0.021	0.027	0.028	0.016
MGMT	0.498	0.494	0.479	0.469	0.427	0.264	0.020	0.032	0.044	0.061	0.077	0.062
HML*	0.503	0.497	0.485	0.469	0.410	0.248	0.010	0.020	0.031	0.041	0.045	0.033
PERF	0.489	0.489	0.478	0.466	0.436	0.272	0.012	0.022	0.034	0.047	0.065	0.053
NetOA	0.502	0.495	0.485	0.462	0.413	0.265	0.006	0.013	0.019	0.026	0.030	0.027
LIQ_TR	0.494	0.490	0.481	0.466	0.415	0.262	0.003	0.007	0.012	0.018	0.023	0.019
ACCR	0.491	0.480	0.473	0.460	0.433	0.271	0.004	0.008	0.016	0.028	0.041	0.034
IA	0.503	0.486	0.466	0.432	0.379	0.224	0.018	0.028	0.037	0.044	0.051	0.041
INV_IN_ASS	0.495	0.489	0.464	0.431	0.365	0.205	0.009	0.015	0.021	0.025	0.026	0.018
UMD	0.486	0.475	0.456	0.424	0.386	0.254	0.007	0.010	0.011	0.011	0.015	0.023
SMB*	0.487	0.476	0.455	0.426	0.377	0.224	0.005	0.009	0.014	0.019	0.025	0.020
DISSTR	0.474	0.459	0.451	0.435	0.392	0.241	-0.002	-0.009	-0.020	-0.034	-0.047	-0.040
SMB	0.476	0.466	0.446	0.417	0.358	0.199	0.010	0.019	0.029	0.036	0.037	0.025
CMA	0.484	0.459	0.435	0.400	0.349	0.204	0.011	0.012	0.009	0.000	-0.015	-0.015
STOCK_ISS	0.488	0.466	0.437	0.404	0.330	0.182	0.011	0.017	0.021	0.024	0.021	0.015
RMW	0.471	0.455	0.432	0.403	0.363	0.221	0.005	0.005	0.002	-0.006	-0.023	-0.019
GR_PROF	0.475	0.454	0.434	0.406	0.352	0.198	0.001	0.002	0.004	0.006	0.007	0.001
BEH_FIN	0.480	0.459	0.437	0.396	0.338	0.191	0.014	0.018	0.020	0.018	0.012	0.012
HML	0.470	0.443	0.422	0.394	0.372	0.232	0.005	0.001	-0.006	-0.019	-0.044	-0.042
ROA	0.472	0.457	0.432	0.400	0.333	0.186	0.009	0.013	0.015	0.014	0.009	0.003
COMP_ISSUE	0.477	0.457	0.425	0.384	0.319	0.174	0.006	0.007	0.007	0.005	0.002	0.004
A_Growth	0.474	0.452	0.421	0.378	0.312	0.168	0.007	0.008	0.006	0.002	-0.002	-0.003
LTRev	0.473	0.451	0.417	0.379	0.313	0.167	0.004	0.005	0.005	0.004	0.001	0.001
O_SCORE	0.472	0.450	0.417	0.378	0.311	0.168	-0.004	-0.006	-0.006	-0.005	-0.007	-0.005

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table 2.A.1 of the Appendix. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results are tabulated for different values of the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_f^2 \mid \sigma^2]}$.

for non-extreme values of the prior maximum SR. Second, for factors that are likely to be at best weakly identified the estimated price of risk is very close to zero but becomes large when the prior SR is very high, and therefore the estimation is no more robust to the weak factors. This is to be expected given the frequentist results on this issue. Third, for factors for which there is clear evidence that they should not be part of the SDF, the estimates of the price of risk are stably around zero. Furthermore, for these factors they are very close to zero even *conditional* on the factors being included in the SDF. This quantity can be easily computed by dividing the posterior mean of the price of risk by the factor posterior probability – both reported in Table 2.3.

As a reality check on the results in Table 2.3, in Table 2.5 of Section 2.5.3 below, we expand our set of candidate priced factors to include artificially generated weak factors and show that our procedure successfully singles them out. Furthermore, in the above estimation we have allowed for a common cross-sectional intercept due to allowing for an average level of mispricing. In Tables 2.A.17–2.A.18 of the Appendix we repeat the estimation imposing a zero common intercept and obtain virtually identical results.[32]

Finally, since we sample the space of 2 quadrillion models instead of estimating them one-by-one, one might wonder whether the estimation is accurate. We address this formally with the standard Separated Partial Means test (see, e.g., Geweke (2005)) for both posterior probabilities and prices of risk, which clearly indicates fast and accurate convergence of the Markov Chain-based estimates.[33]

A natural question is whether the posterior probabilities and prices of risk estimates, summarized in Table 2.3, deliver a good representation of the true latent SDF.

2.5.2 Cross-Sectional Performance

We now focus on the cross-sectional asset pricing performance of our BMA estimates of the Stochastic Discount Factor (BMA-SDF), both in- and out-of-sample, and compare it with traditional popular reduced-form factor models. Table 2.4 reports root mean squared pricing error (RMSE), mean absolute pricing errors (MAPE), and OLS and GLS cross-sectional R^2

---

[32]The fact that imposing the zero intercept restriction leaves the results virtually unchanged is not too surprising since, across all our estimates, the posterior mean of the common intercept is about 0.02–0.03 in monthly SR unit. Hence, since the average monthly variance of the baseline test assets is about 4.5%, the posterior mean of the common intercept is about 0.09%–0.135% in monthly returns units i.e. it is quite small.

[33]To implement the test we drop the first 50,000 draws and split our Markov Chain in five subsets. We compute the average frequency of rejection of posterior probability of factor inclusion and price of risk being the same for all the subsets for different values of the test size (i.e., 95%, 90%, and 80%). The corresponding empirical frequencies of rejection are 6.0%, 9.9%, and 20.2% for the posterior probability of factor inclusion and 4.1%, 9.1%, and 20.4% for the price of risk. In addition, we have estimated the model, increasing the number of draws by a factor of 10, and found virtually identical parameter estimates.

for a variety of models and test assets. For a benchmark comparison, we consider the CAPM, Fama-French five-factor model, Carhart four-factor model, and the q4 model of Hou, Xue, and Zhang (2015). Finally, we also present results for the 51 factor model that includes all the candidate risk factors considered in our analysis, as well as the shrinkage-based approach of Kozak, Nagel, and Santosh (2020) (KNS) with the optimal shrinkage level and number of factors chosen by three-fold cross-validation.[34] Results for the Bayesian (GLS) SDFs are reported for a wide range of SR priors. All the frequentist SDFs are estimated via a GLS version of the GMM (i.e., imposing the tradability restriction on the model-implied price of risk, whenever factors are tradable).[35]

Table 2.4: Cross-sectional asset pricing

Model		RMSE	MAPE	R^2_{ols}	R^2_{gls}	Model	RMSE	MAPE	R^2_{ols}	R^2_{gls}
Panel A: In-sample pricing, test assets: 60 anomalies										
BMA-SDF:	$SR_{pr} = 1$	0.287	0.227	39.2%	24.2%	51 factors	0.041	0.022	98.1%	97.7%
	$SR_{pr} = 1.5$	0.253	0.197	49.8%	30.3%	CAPM	0.418	0.338	-29.4%	16.8%
	$SR_{pr} = 2$	0.223	0.170	59.1%	37.4%	FF5	0.301	0.223	24.5%	23.2%
	$SR_{pr} = 2.5$	0.193	0.148	68.2%	45.5%	Carhart	0.317	0.244	21.5%	21.2%
	$SR_{pr} = 3$	0.162	0.128	76.6%	54.7%	q4	0.267	0.189	37.5%	28.1%
	$SR_{pr} = 3.5$	0.157	0.128	78.4%	58.8%	KNS_{CV_3}	0.296	0.237	53.7%	19.6%
Panel B: Cross-sectional out-of-sample pricing, test assets: 25 size-value portfolios										
BMA-SDF:	$SR_{pr} = 1$	0.108	0.082	42.1%	17.5%	51 factors	0.200	0.163	-98.5%	-1653%
	$SR_{pr} = 1.5$	0.094	0.070	55.7%	24.5%	CAPM	0.145	0.112	-4.6%	5.2%
	$SR_{pr} = 2$	0.085	0.063	64.5%	30.2%	FF5	0.079	0.059	69.2%	28.0%
	$SR_{pr} = 2.5$	0.077	0.058	70.5%	34.9%	Carhart	0.086	0.063	63.2%	27.1%
	$SR_{pr} = 3$	0.073	0.054	73.9%	38.4%	q4	0.083	0.065	66.1%	28.2%
	$SR_{pr} = 3.5$	0.075	0.055	72.3%	36.8%	KNS_{CV_3}	0.096	0.074	54.4%	28.0%
Panel C: Cross-sectional out-of-sample pricing, test assets: 49 industry portfolios										
BMA-SDF:	$SR_{pr} = 1$	0.097	0.080	15.6%	11.8%	51 factors	0.420	0.310	-1474.3%	-1694%
	$SR_{pr} = 1.5$	0.097	0.082	15.3%	12.8%	CAPM	0.111	0.082	-10.6%	20.9%
	$SR_{pr} = 2$	0.097	0.082	15.7%	15.8%	FF5	0.123	0.103	-35.8%	3.6%
	$SR_{pr} = 2.5$	0.098	0.081	14.9%	18.5%	Carhart	0.117	0.089	-22.1%	13.7%
	$SR_{pr} = 3$	0.100	0.083	10.9%	19.7%	q4	0.134	0.105	-60.5%	-10.9%
	$SR_{pr} = 3.5$	0.100	0.083	11.5%	20.9%	KNS_{CV_3}	0.100	0.082	10.9%	14.0%

This table compares in-sample and cross-sectional out-of-sample asset pricing performance of the B-SDF and notable frequentist factor models. We use GMM-GLS to estimate factor prices of risk for the CAPM, FF5 model of Fama and French (2015), Carhart (1997) model, q4 model of Hou, Xue, and Zhang (2015), and the model including all 51 factors. KNS stands for the SDF estimation of Kozak, Nagel, and Santosh (2020), with tuning parameter and number of factors chosen by three-fold cross-validation. For the B-SDF, we report results with risk prices under a range of prior $\sqrt{\mathbb{E}_\pi[SR^2_f \mid \sigma^2]} \in \{1\text{–}3.5\}$. In the cross-sectional OOS the models are first estimated using the baseline test assets of Panel A and then used to price (without additional parameters estimation), the test assets listed in Panels B and C. All the data is standardized, that is, pricing errors are in SR units. We show the annualized RMSE and MAPE. The out-of-sample performance relies on the SDF estimates obtained from 60 anomaly portfolios as test assets and is then used to price other cross-sectionally demeaned test assets without re-estimating the SDF.

---

[34]When applied to our sample of 60 portfolios, three-fold CV selects a model with 11 factors and the root expected SR^2 of 1.2.

[35]We have also obtained virtually identical results using time-series regressions (with tradable factors) instead of GMM, as well as other cross-sections not reported in Table 2.4. Results are available upon request.

Panel A reports in-sample asset pricing statistics for the baseline set of assets used in our estimation (60 anomaly portfolios).[36] It is striking that the Bayesian SDF tends to outperform conventional models across a wide range of metrics, and this result is stable across the whole set of SR priors. Furthermore, unlike the benchmark models, the BMA-SDF delivers cross-sectional OLS and GLS R^2 measures that are consistent with each other – without explicitly targeting any of them at the SDF estimation stage. The only model that seems to perform better than the BMA-SDF is the one using 51 factors to price 60 assets and is very likely to be overfitting the cross-section (as we show below). One might wonder whether part of the Bayesian SDF success could also be due to overfitting. We address this issue by analyzing its OOS performance, in both cross-sectional and time-series dimensions.

Panels B and C summarize the performance of SDFs estimated on a set of 60 anomaly portfolios ($\widehat{M_t}$) but then used to price a different cross-section – 25 portfolios sorted by size and value (Panel B), and 49 industry portfolios (Panel C). Since we shrink away level factors in the BMA-SDF, to put different models on equal footing we focus on cross-sectionally demeaned pricing errors. Our findings make it clear that the superior performance of the BMA-SDF observed in-sample is not due to overfitting. While the 51-factor model has a disastrous cross-sectional OOS performance, this is not the case for the BMA-SDF. Consistent with our in-sample results, the performance of the Bayesian SDF is stable across priors and metrics. Furthermore, it is either on par with that of the best reduced-form benchmark model (the FF5 model when focusing on size-value portfolios) or better. The BMA-SDF pricing ability is particularly striking in the case of industry portfolios that have long been considered a challenge for asset pricing and often advocated as an appropriate testing ground for models (e.g., Lewellen, Nagel, and Shanken (2010), and Daniel and Titman (2012)).

Figure 2.5 further illustrates the performance of different SDFs estimated on the baseline cross-section and then used to price the 49 industry portfolios. The BMA-SDF is the only model that generates predicted Sharpe ratios close to the observed ones and has positive (OLS and GLS) cross-sectional R^2s. Note that while some of the models yield predictions that have positive correlation with the actual return realizations, they are still characterized by a substantially negative R^2 since we impose the theoretical pricing restriction of $\mathbb{E}[\boldsymbol{R_t}] = -Cov(M_t, \boldsymbol{R_t})$ (using the innocuous normalization $\mathbb{E}[M_t] = 1$).

We now turn to the time-series out-of-sample performance of the BMA-SDF.[37] According

---

[36]The table reports the following measures: $RMSE \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2}$, $MAPE \equiv \frac{1}{N}\sum_{i=1}^{N}|\alpha_i|$, $R_{ols}^2 \equiv 1 - \frac{(\boldsymbol{\alpha} - \frac{1}{N}\boldsymbol{\alpha}^\top\mathbf{1}_N)^\top(\boldsymbol{\alpha} - \frac{1}{N}\boldsymbol{\alpha}^\top\mathbf{1}_N)}{(\boldsymbol{\mu_R} - \frac{1}{N}\boldsymbol{\mu_R^\top}\mathbf{1}_N)^\top(\boldsymbol{\mu_R} - \frac{1}{N}\boldsymbol{\mu_R^\top}\mathbf{1}_N)}$, and $R_{gls}^2 \equiv 1 - \frac{\boldsymbol{\alpha}^\top\boldsymbol{\Sigma_R^{-1}}\boldsymbol{\alpha}}{\boldsymbol{\mu_R^\top}\boldsymbol{\Sigma_R^{-1}}\boldsymbol{\mu_R}}$.

[37]We follow the approach canonical in the literature of performing time series OOS via a split-sample approach (see, e.g., Linnainmaa and Roberts (2018), Chen, Pelger, and Zhu (2019), Gu, Kelly, and Xiu (2020)). Nevertheless, ideally, one might want to focus on the post publication sample of the factors. This is unfortunately unfeasible in our empirical setting since a large share of the factors that we analyze have
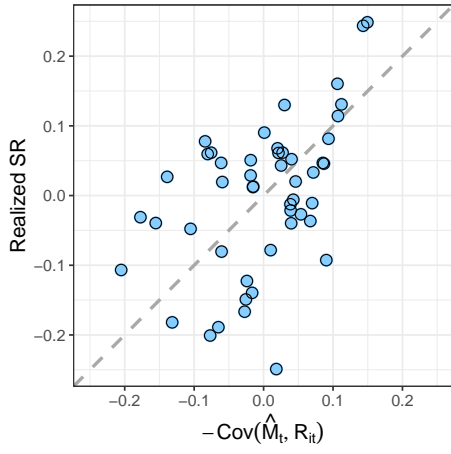
(a) B-SDF, $SR_{pr} = 2$ ($R^2_{ols} = 16\%$)
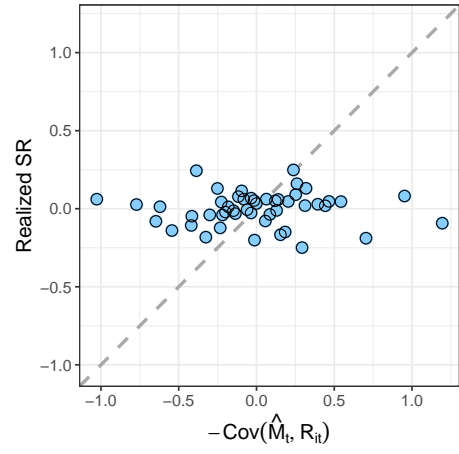
(b) 51-factor model ($R^2_{ols} = -1474\%$)

(c) FF5 ($R^2_{ols} = -36\%$)

(d) Carhart ($R^2_{ols} = -22\%$)

(e) q4 ($R^2_{ols} = -61\%$)
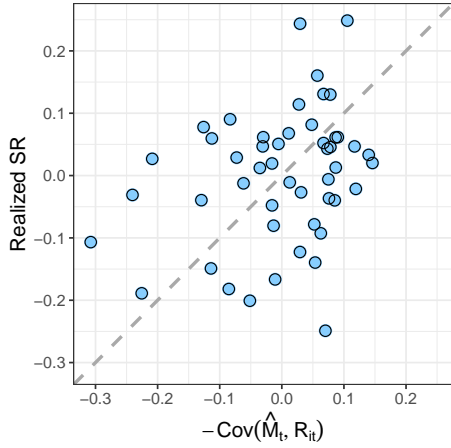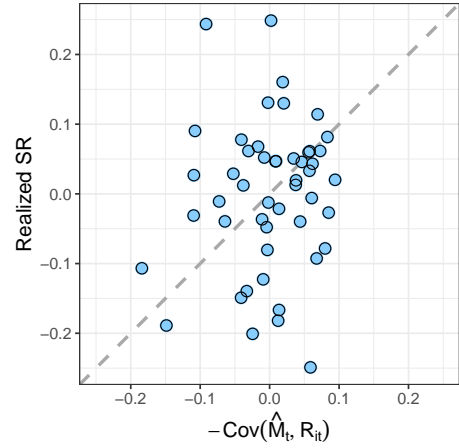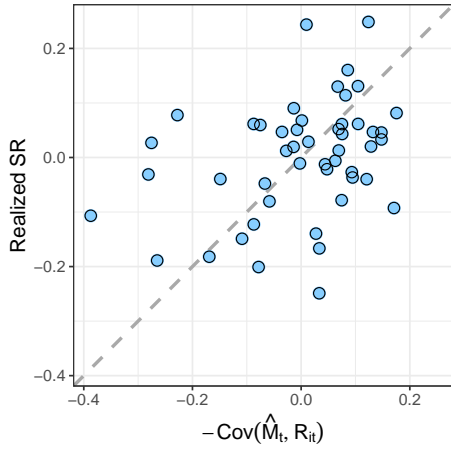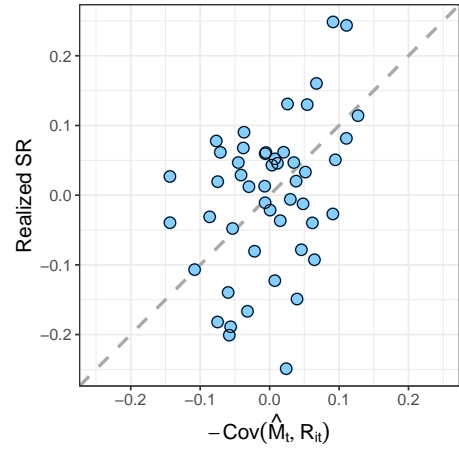
(f) KNS$_{CV_3}$ ($R^2_{ols} = 10\%$)

Figure 2.5: Out-of-sample cross-sectional pricing of 49 industry portfolios.

For each model, the figure depicts the out-of-sample performance of the SDF, obtained by using 60 anomaly portfolios as test assets, and applied to pricing 49 industry portfolios without re-estimation. All the data are standardized; that is, pricing errors are in SR units. The 45-degree line corresponds to the theoretical relationship of $\mathbb{E}[\boldsymbol{R_t}] = -Cov(M_t, \boldsymbol{R_t})$, where SDFs are normalized to have unit mean.

---

been only very recently documented.

to Table 2.4, only the shrinkage-based approach of Kozak, Nagel, and Santosh (2020) comes close to matching the performance of our Bayesian approach overall. Hence, we use it as a benchmark model for the time-series out-of-sample performance. Figure 2.6 reports out-of-sample model performance, based on the time-series difference between estimation and prediction periods. Following Martin and Nagel (2021), we use half of the time-series sample for the model estimation and SDF recovery and evaluate its cross-sectional pricing ability on the other subsample. Thus, we consider out-of-sample performance of the model, going into both future and past, without re-estimating any of the parameters. For the same value of the prior SR, BMA-SDF tends to outperform the cross-validated estimates (CV_3) of KNS, despite the fact that cross-validation was carried out on the full data sample. Furthermore, for a wide range of prior SR, our Bayesian approach performs either as well as the ex-post best combination of tuning parameters in KNS or better. This is particularly evident when recent data is used as the evaluation subsample.

2.5.3   Model Uncertainty: Selection or Aggregation?

In the previous section we have shown that averaging across the space of possible models yields an accurate representation of the SDF. A natural question is whether in the universe of models there is a *single* best model.

For consistency, frequentist model selection *demands* the existence of a unique first-best model that can be reliably distinguished from the alternatives. This is a key assumption underlying reliable factor selection via t- and χ^2- tests, LASSO, and many other approaches.

In contrast, the existence of such a dominant model can be formally assessed within the Bayesian paradigm. For instance, Giannone, Lenza, and Primiceri (2021b) study the sparsity assumption in popular empirical economic applications (using, like us, a spike-and-slab prior approach for model and variable selection). They find that the posterior distribution does not typically concentrate on a single sparse model but rather supports a wide set of models that often include a large number of predictors.

Our framework is ideally suited for evaluating the assumption of sparsity (in observable factors) in cross-sectional asset pricing, since we can compute posterior model probabilities for each possible specification generated by our 51 factors (i.e., about 2.25 quadrillion models).

Figure 2.7 presents the model posterior probabilities of the 2,000 most likely specifications (with a prior SR of 2). The first thing to notice is that even the most likely specification(s) is not a clear winner within the set of all possible models – its posterior probability is only about 0.011%. This is a remarkable improvement relative to the prior model probability that is of the order of 10^{-16}, but it clearly does not represent a substantial resolution of
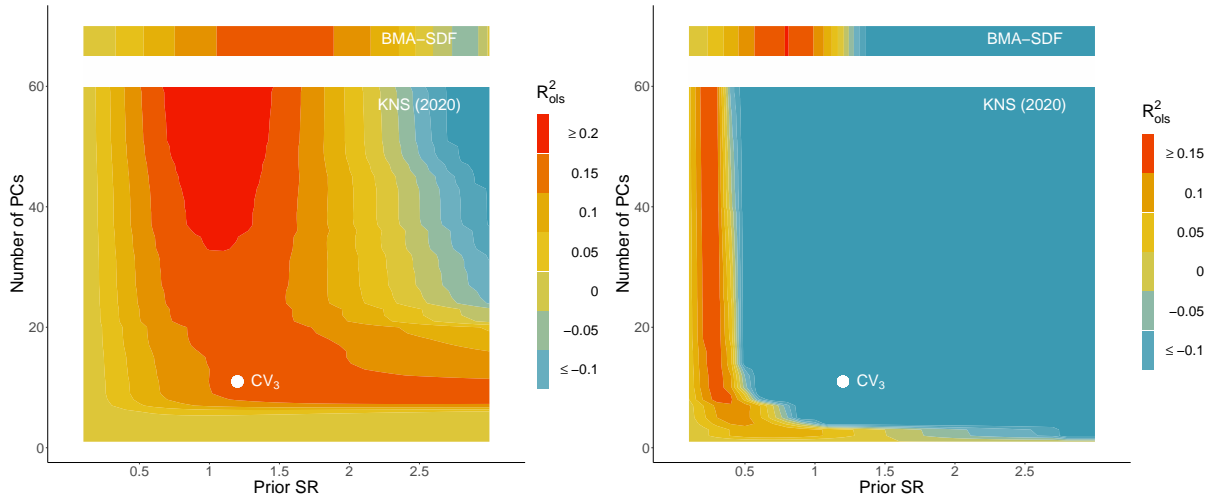
(a) R^2_{ols}, estimation period: 1995/06-2016/12,
evaluation period: 1973/10-1995/05

(b) R^2_{ols}, estimation period: 1973/10-1995/05,
evaluation period: 1995/06-2016/12
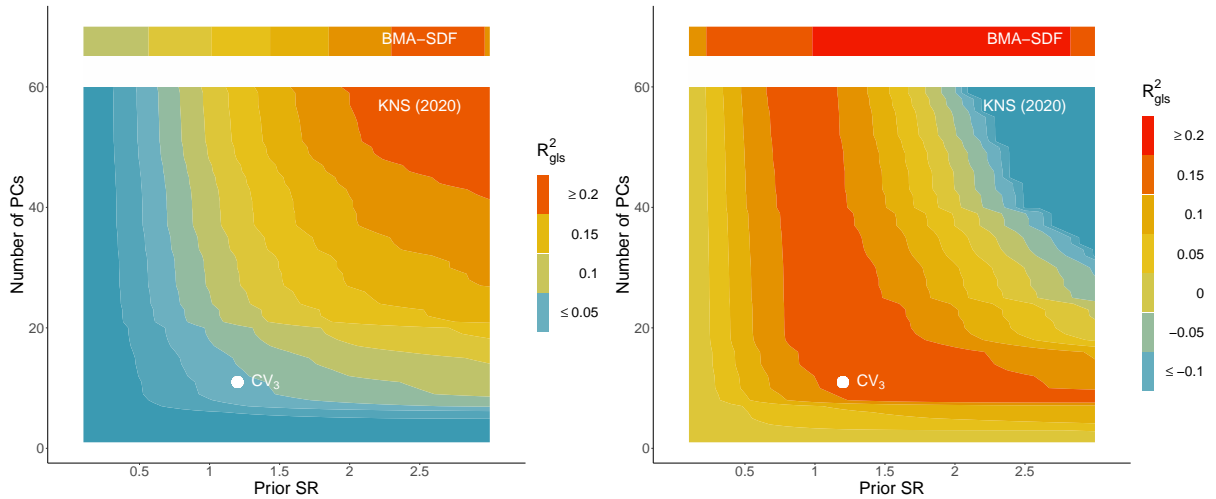
(c) R^2_{gls}, estimation period: 1995/06-2016/12,
evaluation period: 1973/10-1995/05

(d) R^2_{gls}, estimation period: 1973/10-1995/05,
evaluation period: 1995/06-2016/12

Figure 2.6: Out-of-sample cross-sectional pricing (different time samples).

The figure depicts out-of-sample performance of the SDF (R^2_{ols} and R^2_{gls}), obtained by using both BMA and Kozak, Nagel, and Santosh (2020) approaches using a time series subsample of 60 anomaly portfolios. We use half of the time-series sample for the model estimation and SDF recovery and evaluate its cross-sectional pricing ability on the other subsample. Results are reported for a range of SR prior, and in the case of KNS (2020) for different number of PCs used, as well as the optimal combination of tuning parameters chosen by a three-fold cross-validation applied to the estimation period.

model uncertainty. Furthermore, we have 10 specifications with basically the same posterior probability, and the posterior model probability decays very slowly as we move down the list of most likely models: Moving from the best model, it takes more than a thousand models
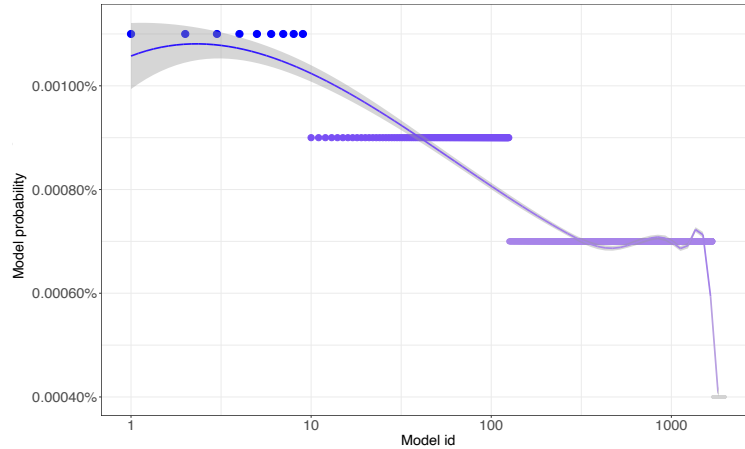
Figure 2.7: Posterior model probabilities, prior SR = 2.

Posterior model probabilities of the 2,000 most likely models computed using the continuous spike-and-slab of Section 2.3.1.3 and 51 factors. The horizontal axis uses a log scale. Sample: 1973:10–2016:12. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12.

to reach the relative odds of 2:1 (i.e., to reduce the posterior probability by 50%). That is, to a first-order approximation, the frequentist likelihood ratio test of the best performing model versus the 1000th one would yield a p-value of 30% at best (and a p-value of 15% after 2,000 models).

But how many of the factors proposed in the literature does it really take to price the cross-section? Thanks to our Bayesian method, even this question can be easily answered. In particular, by using our estimations of about 2.25 quadrillion models and their posterior probabilities, we can compute the posterior distribution of the dimensionality of the "true" model. That is, for any integer number between one and 51, we can compute the posterior probability of the linear factor model being a function of that number of factors.

Figure 2.8a reports the posterior distributions of the model dimensionality for various values of prior SR. These distributions are also summarized in Table 2.A.19 of Appendix 2.A.4.

For the most salient values of the prior SR (1–3), the posterior mean of the number of factors in the true model is in the 23–25 range, and the 95% posterior credible intervals are contained in the 16–32 factors range. That is, there is substantial evidence that the SDF is *dense* in the space of observable factors: Given the factors at hand, a relative large number of them is needed to provide an accurate representation of the "true" model. Since most of the literature has focused on very low-dimensional linear factor models, this finding suggests that most empirical results therein have been affected by a very large degree of
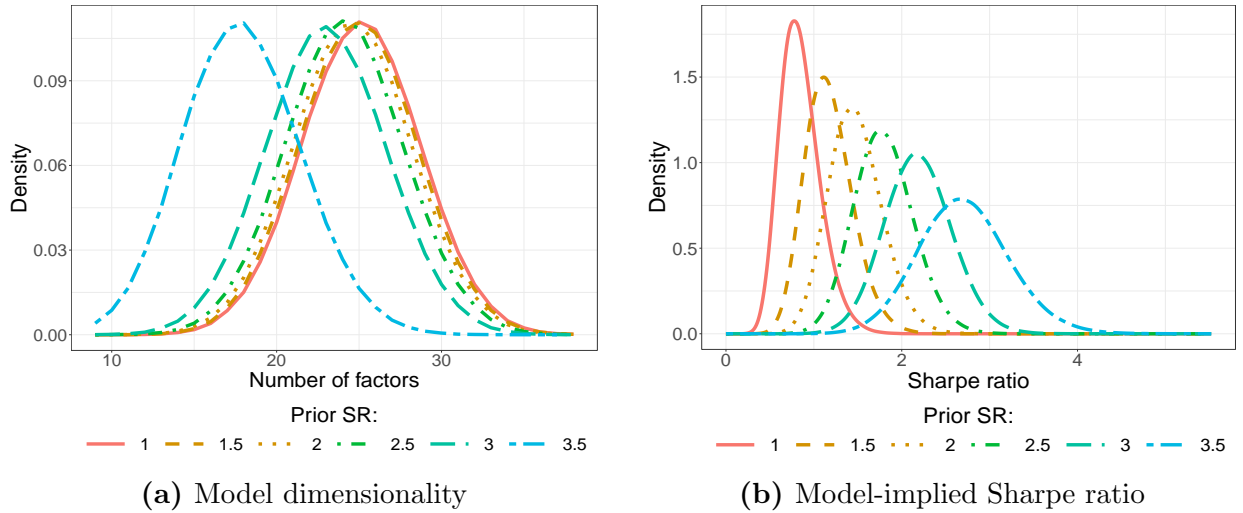
Figure 2.8: Posterior densities of model dimensionality and its implied Sharpe ratio.

Left panel: posterior density of the true model having the number of factors listed on the horizontal axis. Right panel: posterior density of the (annualized) Sharpe ratio implied by the linear factor model for various values of $SR_{prior} \in [1, 3.5]$. Sample: 1973:10–2016:12. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. The prior for each factor inclusion is a $Beta(1, 1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table 2.A.1 of the Appendix.

misspecification.

It is worth noticing that, as Figure 2.8a shows, for very large prior SR, that is, with basically a flat prior for factors' price of risk, the posterior dimensionality is reduced. This is due to two phenomena we have already outlined. First, if some of the factors are useless (and our analysis points in this direction), under a flat prior they tend to have a higher posterior probability and drive out the true sources of priced risk. Second, a flat prior for the price of risk can generate a "Bartlett Paradox" (see the discussion in Section 2.3.1.1).

Note that if the factors proposed in the literature were to capture different and uncorrelated sources of risk, one might worry that a dense model in the space of factors could imply unrealistically high Sharpe ratios (see, e.g., the discussion in Kozak, Nagel, and Santosh (2020)). Since, given a model, the SDF-implied maximum Sharpe ratio is merely a function of the factors' price of risk and covariance matrix, our Bayesian method allows us to construct the posterior distribution of the maximum Sharpe ratio for each of the 2.25 quadrillion models considered. Therefore, using the posterior probabilities of each possible model specification, we can actually construct the (BMA) posterior distribution of the SDF-implied maximum Sharpe ratio (conditional on the data only).

Figure 2.8b (and Table 2.A.19b in Appendix 2.A.4) reports the posterior distribution of the SDF-implied maximum Sharpe ratio (annualized) for several values of the prior SR. Except when a very strong shrinkage (small prior SR) is imposed (hence, Sharpe ratios

are shrunk) the posterior distributions of the Sharpe ratio are quite similar for all prior values. Furthermore, despite the model being dense in the space of factors, the posterior maximum Sharpe ratio does not appear to be unrealistically high: For example, for a prior SR $\in [1.5,\ 3]$ its posterior mean is about 1.17–2.19, and the 95% posterior credible intervals are in the 0.70–2.96 range.

Note that a model that is dense in the space of observable factors might be in principle sparse in the space of latent factors, for example, principle components. We address this issue by directly including principal components in the set of candidate factors. In particular, we consider the first five principal components of our cross-section of test assets, followed by a set of five "Risk Premia" principal components (RP-PC) of Lettau and Pelger (2020b). In addition, to confirm that our method successfully handles weak identification, we add two artificially generated useless factors (independent of returns and i.i.d. distributed). Table 2.5 reports our findings.

Panel A of Table 2.5 shows that the first five principal components do not seem to capture priced risk: Their posterior probability is substantially lower than their prior probability, and their estimated market price of risk is zero (despite them explaining 61% of the time-series variation of returns). This is quite expected since standard principal components are not designed to capture cross-sectional pricing information.

Clearly, the artificially generated useless factors are successfully handled by the estimation procedure: As expected, their posterior probability remains at the prior level (50%), and their estimated price of risk is basically zero.

In Panel B we replace the canonical PCs with RP-PCs. We find strong support for two of them (first and third) capturing priced risk, while the other three have posterior probability below the prior value and prices of risk close to zero. Interestingly, even though some of the RP-PCs seem to successfully aggregate pricing information from the cross-section of returns (and factors, since the tradable ones are part of the test assets), they do not drive out the relevance of the robust stand-alone factors we identified earlier: BEAH_PEAD, CMA*, RMW*, among others. Consequently, the underlying SDF would be best described by a combination of both observable factors and (some) latent variables. Hence, the results in Panel B highlight that, in the quest of describing the sources of priced risk, there is scope for both selection and aggregation. This is confirmed by Figure 2.A.2 in Appendix 2.A.4, which shows that the most likely SDF is dense in the combined space of observable factors and principal components.

Table 2.5: Observable factors versus Principal Components

	Factor inclusion prob., $\mathbb{E}[\gamma_j|\text{data}]$						Price of risk, $\mathbb{E}[\lambda_j|\text{data}]$					
	Total prior SR						Total prior SR					
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
Panel A: Principal Components as Factors												
BEH_PEAD	0.547	0.602	0.678	0.766	0.840	0.814	0.015	0.036	0.073	0.132	0.220	0.287
MKT	0.508	0.542	0.573	0.598	0.607	0.504	0.015	0.035	0.064	0.100	0.149	0.182
CMA*	0.509	0.523	0.539	0.564	0.597	0.516	0.009	0.020	0.037	0.061	0.101	0.124
BW_ISENT	0.499	0.502	0.509	0.514	0.528	0.555	0.002	0.004	0.008	0.014	0.030	0.105
RMW*	0.500	0.499	0.514	0.537	0.568	0.450	0.007	0.017	0.032	0.057	0.097	0.107
STRev	0.495	0.503	0.522	0.546	0.555	0.435	0.006	0.016	0.030	0.052	0.083	0.089
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Useless I	0.499	0.499	0.501	0.498	0.498	0.497	0.000	0.000	0.000	0.000	0.001	0.006
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Useless II	0.496	0.495	0.495	0.494	0.498	0.500	0.000	0.000	0.001	0.001	0.002	0.010
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PC5	0.489	0.490	0.488	0.482	0.459	0.336	0.000	0.000	0.000	0.000	0.000	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PC4	0.497	0.487	0.480	0.471	0.451	0.322	0.000	0.000	0.000	0.000	0.000	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PC3	0.483	0.477	0.467	0.449	0.420	0.280	0.000	0.000	0.000	0.000	0.000	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PC1	0.478	0.467	0.457	0.437	0.399	0.248	0.000	0.000	0.000	0.000	-0.001	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
PC2	0.473	0.455	0.444	0.429	0.397	0.249	0.000	0.000	0.000	0.000	0.000	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
LTRev	0.477	0.464	0.437	0.402	0.347	0.204	0.003	0.005	0.004	0.002	-0.002	-0.003
COMP_ISSUE	0.485	0.462	0.438	0.399	0.338	0.191	0.006	0.007	0.008	0.007	0.003	0.002
A_Growth	0.481	0.462	0.436	0.399	0.337	0.189	0.007	0.008	0.006	0.003	-0.002	-0.003
O_SCORE	0.473	0.450	0.425	0.385	0.323	0.186	-0.003	-0.005	-0.004	-0.002	-0.002	-0.003
Panel B: RP-Principal Components (Lettau and Pelger (2020b)) as Factors												
RP-PC1	0.600	0.631	0.640	0.634	0.592	0.448	-0.016	-0.030	-0.043	-0.056	-0.066	-0.067
RP-PC3	0.548	0.597	0.645	0.661	0.651	0.529	-0.004	-0.009	-0.017	-0.024	-0.032	-0.035
BEH_PEAD	0.540	0.585	0.628	0.681	0.709	0.630	0.014	0.032	0.058	0.097	0.149	0.185
CMA*	0.510	0.523	0.542	0.571	0.616	0.531	0.009	0.020	0.037	0.062	0.104	0.129
RMW*	0.500	0.504	0.517	0.547	0.583	0.466	0.007	0.017	0.033	0.059	0.101	0.112
MKT	0.507	0.518	0.525	0.516	0.493	0.391	0.013	0.028	0.044	0.061	0.081	0.103
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Useless I	0.499	0.499	0.500	0.500	0.499	0.497	0.000	0.000	0.000	0.000	0.001	0.007
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Useless II	0.495	0.495	0.495	0.498	0.496	0.499	0.000	0.000	0.000	0.001	0.002	0.010
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
RP-PC5	0.481	0.487	0.488	0.484	0.459	0.338	0.001	0.003	0.005	0.008	0.011	0.012
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
RP-PC4	0.494	0.487	0.479	0.459	0.433	0.303	0.002	0.003	0.004	0.005	0.005	0.005
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
RP-PC2	0.479	0.464	0.458	0.439	0.403	0.267	0.000	-0.001	-0.001	-0.001	-0.001	0.000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
COMP_ISSUE	0.483	0.464	0.438	0.406	0.338	0.193	0.006	0.008	0.010	0.009	0.004	0.002
A_Growth	0.483	0.466	0.443	0.396	0.337	0.196	0.007	0.007	0.005	0.000	-0.005	-0.007
LTRev	0.483	0.461	0.438	0.404	0.358	0.222	0.003	0.003	0.000	-0.006	-0.014	-0.015
O_SCORE	0.472	0.456	0.426	0.386	0.331	0.189	-0.003	-0.002	0.001	0.005	0.005	0.001

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 59 factors yielding 2^{59} models. The factors included are the 51 factors described in Table 2.A.1 of the Appendix plus two artificial i.i.d. useless factors, and five principal components. Panel A uses simple time-series principal components while Panel B uses the RP-PCs of Lettau and Pelger (2020b). Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results tabulated for different values of the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]}$.

2.5.4 A Quest for Sparsity

The previous subsections suggest that only a small number of observable factors – BEH_PEAD, MKT, CMA*, and, to a lesser extent, STRev, RMW*, and BW_ISENT – are likely stand-alone explanators of the cross-section of asset returns. A natural question is whether the Bayesian factor posterior probabilities of Table 2.3 can help identify a low-dimensional benchmark model for pricing asset returns. We answer this question by comparing the performance of a three-factor model with HML, MKT*, and SMB* as factors, to the one of several notable factor models.

Table 2.6 reports the model posterior probabilities, that is, the probability of any of these models being the true data-generating process, for the SDFs built with the most likely factors and notable linear factor models. Posterior model probabilities (for all models) are computed using the closed-form solutions for the Dirac spike-and-slab prior method of Section 2.3.1.2, giving us very precise estimates.

Strikingly, for *any* value of SR_{pr}, the best performing model is the one based on the most likely factors: Just three most likely factors (see Panel A), BEH_PEAD, MKT, and CMA*, are enough to outperform the most widely used empirical SDFs. This outperformance becomes even more pronounced when we consider the six most likely factors (see Panel B). Note that this drastic difference in performance understates the true power of our Bayesian approach to factor and model selection. Indeed, a subset of the most (individually) likely factors does not necessarily create the *most likely model*. Luckily, our approach can also be used to select the most likely model of any dimension. In particular, in Panel C we run the horse race between the most likely five-factor model that emerges using the Dirac Spike-and-Slab approach of Section 2.3.1.2. Clearly, for all the values of prior SR, the best five-factor model outperforms not only all the notable models but also the combination of six overall most likely factors (from Panel B). While different prior SR may lead to different most likely low-dimensional models, the subset of selected factors is quite stable: All the specifications include BEH_PEAD and CMA*, while RMW* and BAB are selected four times out of five, and STRev is part of the most likely model three times out of five.

Our approach can also be used to formally evaluate the space of sparse factor models. In particular, in Table 2.7 we consider the universe of all the possible models that include no more than five factors, that is, 2.6 mln models. We evaluate *all* of those models individually, computing each of their marginal likelihoods following the Dirac spike-and-slab approach of Section 2.3.1.2 (instead of sampling models, as in Section 2.5.1). The table reports both posterior probabilities of the factor inclusion and their posterior price of risk. For simplicity, we consider the prior probability of a factor being included into the model being equal to 9.58% (since we have 51 factors total and each model with up to five factors is given equal

Table 2.6: Posterior probabilities of notable models versus most likely factors

model:	SR_{pr}:	*Panel A*: 3 most likely factors					*Panel B*: 6 most likely factors					*Panel C*: Most likely 5-factor model				
		1	1.5	2	2.5	3	1	1.5	2	2.5	3	1	1.5	2	2.5	3
Most likely factors		17.5%	24.9%	36.0%	48.8%	59.1%	17.8%	27.0%	44.0%	66.5%	83.7%	23.0%	35.3%	57.0%	77.6%	88.1%
CAPM		12.7%	12.5%	11.8%	11.3%	13.1%	12.7%	12.1%	10.3%	7.3%	5.2%	11.9%	10.8%	8.0%	5.0%	3.9%
FF3		10.3%	7.9%	5.3%	3.2%	1.7%	10.3%	7.7%	4.7%	2.1%	0.7%	9.6%	6.8%	3.6%	1.4%	0.5%
FF5		9.9%	7.0%	4.2%	2.1%	0.7%	9.8%	6.8%	3.7%	1.3%	0.3%	9.2%	6.0%	2.8%	0.9%	0.2%
Carhart		10.2%	7.8%	5.2%	2.9%	1.3%	10.2%	7.6%	4.6%	1.9%	0.5%	9.6%	6.7%	3.5%	1.3%	0.4%
q4		15.7%	17.8%	17.9%	14.9%	9.6%	15.6%	17.3%	15.7%	9.9%	3.9%	14.6%	15.3%	11.9%	6.4%	2.7%
Liq-CAPM		12.5%	12.0%	10.9%	9.6%	9.0%	12.5%	11.7%	9.5%	6.2%	3.6%	11.7%	10.4%	7.4%	4.3%	2.7%
FF3-QMJ		11.2%	10.1%	8.8%	7.4%	5.5%	11.1%	9.8%	7.7%	4.8%	2.1%	10.4%	8.6%	5.8%	3.1%	1.5%

Posterior model probabilities for the specifications in the first column, for different prior Sharpe ratio values, computed using the Dirac spike-and-slab prior method of Section 2.3.1.2. Panel A includes the factors BEH_PEAD, MKT, CMA*, while Panel B considers in addition STRev, RMW*, and BW_ISENT. Panel C uses the most likely 5-factor model according to the posterior probability. Factors are: MKT, MGMT, BAB, BEH_PEAD, CMA* for $SR_{pr} = 1$; STRev, BAB, BEH_PEAD, RMW*, CMA* for $SR_{pr} = 1.5$ to 2.5; BW_ISENT, BEH_PEAD, MKT*, RMW*, CMA* for $SR_{pr} = 3$. Factors are described in Table 2.A.1 of the Appendix. Liq-CAPM stands for the liquidity-adjusted model of Pástor and Stambaugh (2003) and FF3-QMJ corresponds to the 4-factor model of Asness, Frazzini, and Pedersen (2019). Sample: 1973:10 to 2016:12. Test assets: 60 anomaly portfolios.

ex ante probability).

First, three factors clearly stand out in Table 2.7: BEH_PEAD, BW_SENT, CMA*, all of which were also among the most likely factors in the SDF identified in the whole model space of Table 2.3. Second, and strikingly, there is a large set of factors that have posterior probability of inclusion above the prior, providing support for them being included in a low-dimensional model. This group includes not only the other robust factors identified in Section 2.5.1 but also 40% of both tradable and nontradable macro-factors, such as nondurable consumption, unemployment, and industrial production growth. This second finding is consistent with our results in Section 2.5.3, where we showed that many factors seem to load on the same underlying sources of economic risks: Sparse models, therefore, tend to rely on them almost interchangeably. This is further illustrated in Figure 2.A.3 of the Appendix, which depicts posterior probabilities for the top 2,000 sparse models under the SR prior of 2. Similar to our findings in Section 2.5.3, the space of best performing models is quite flat, with their corresponding posterior probability decaying slowly. In fact, up to a first-order approximation, the frequentist likelihood ratio test of the best performing model versus the 100th (1000th) specification would yield a p-value of 19.0% (9.2%) at best.

Our findings indicate that all the low-dimensional models with observable factors are likely to be severely misspecified, and in many cases reflect noisy measures of the same underlying economic risks. While some of the factors still stand alone as significant drivers of the cross-section of asset returns, the true latent SDF is still best approximated by an efficient aggregation of many underlying variables, provided by the BMA. To further validate this point, we have performed an OOS analysis (in both time-series and cross-sectional dimension) of the BMA versus the best low-dimensional models and found that the former

Table 2.7: Posterior factor probabilities, $\mathbb{E}\left[\gamma_j|\text{data}\right]$, and risk prices: 2.6 million models

| | Factor inclusion prob., $\mathbb{E}\left[\gamma_j|\text{data}\right]$ | | | | | | Price of risk, $\mathbb{E}\left[\lambda_j|\text{data}\right]$ | | | | | |
| | Total prior SR: | | | | | | Total prior SR: | | | | | |
Factors:	0.5	1	1.5	2	2.5	3	0.5	1	1.5	2	2.5	3
BEH_PEAD	0.124	0.206	0.309	0.389	0.430	0.421	0.005	0.024	0.059	0.095	0.122	0.130
BW_ISENT	0.099	0.109	0.128	0.161	0.225	0.343	0.001	0.002	0.007	0.016	0.041	0.111
CMA*	0.104	0.122	0.136	0.141	0.137	0.120	0.003	0.010	0.017	0.023	0.026	0.025
BAB	0.112	0.133	0.140	0.136	0.125	0.105	0.005	0.014	0.021	0.024	0.025	0.022
DIV	0.097	0.102	0.109	0.121	0.141	0.183	0.000	0.000	0.001	0.002	0.005	0.016
HJTZ_ISENT	0.097	0.102	0.109	0.119	0.134	0.156	0.000	0.001	0.002	0.005	0.010	0.021
NONDUR	0.097	0.101	0.108	0.118	0.133	0.161	0.000	0.001	0.002	0.004	0.008	0.020
TERM	0.097	0.101	0.108	0.118	0.133	0.161	0.000	0.000	-0.001	-0.002	-0.005	-0.012
PE	0.097	0.101	0.108	0.117	0.132	0.160	0.000	0.000	-0.001	-0.002	-0.004	-0.012
FIN_UNC	0.097	0.101	0.108	0.117	0.131	0.149	0.000	0.001	0.002	0.004	0.008	0.017
UNRATE	0.097	0.101	0.107	0.116	0.130	0.154	0.000	0.000	0.001	0.002	0.005	0.013
DeltaSLOPE	0.097	0.101	0.107	0.116	0.129	0.154	0.000	0.000	0.001	0.002	0.003	0.010
IPGrowth	0.097	0.101	0.107	0.115	0.127	0.148	0.000	0.000	0.000	-0.001	-0.002	-0.006
DEFAULT	0.097	0.101	0.107	0.115	0.127	0.146	0.000	0.000	0.001	0.001	0.003	0.007
SERV	0.096	0.101	0.106	0.114	0.126	0.146	0.000	0.000	0.000	0.001	0.002	0.006
REAL_UNC	0.096	0.100	0.106	0.114	0.125	0.141	0.000	0.000	0.000	0.001	0.002	0.004
STRev	0.095	0.098	0.105	0.116	0.123	0.109	0.001	0.005	0.010	0.016	0.022	0.022
MACRO_UNC	0.096	0.100	0.106	0.113	0.122	0.136	0.000	0.000	0.000	0.000	0.001	0.000
Oil	0.096	0.100	0.105	0.111	0.119	0.129	0.000	0.000	0.000	0.000	0.001	0.002
MKT*	0.097	0.101	0.104	0.105	0.103	0.105	0.002	0.005	0.009	0.013	0.015	0.020
RMW*	0.096	0.098	0.102	0.106	0.103	0.083	0.002	0.006	0.011	0.015	0.018	0.016
LIQ_NT	0.096	0.098	0.100	0.102	0.101	0.096	0.000	0.000	0.001	0.001	0.002	0.003
MKT	0.094	0.099	0.103	0.103	0.095	0.080	0.003	0.009	0.014	0.018	0.020	0.018
ROE	0.107	0.113	0.106	0.093	0.078	0.060	0.006	0.013	0.016	0.017	0.015	0.012
MGMT	0.109	0.109	0.101	0.092	0.080	0.061	0.007	0.014	0.017	0.018	0.017	0.013
NetOA	0.098	0.102	0.101	0.094	0.084	0.068	0.002	0.005	0.008	0.010	0.010	0.009
IA	0.108	0.108	0.099	0.089	0.077	0.060	0.006	0.013	0.015	0.016	0.015	0.012
HML*	0.099	0.101	0.096	0.087	0.075	0.058	0.003	0.007	0.010	0.011	0.011	0.009
LIQ_TR	0.095	0.095	0.093	0.087	0.078	0.063	0.001	0.002	0.004	0.006	0.006	0.005
INTERM_CAP_RATIO	0.093	0.090	0.087	0.083	0.075	0.062	0.001	0.004	0.006	0.008	0.009	0.008
INV_IN_ASS	0.098	0.097	0.090	0.079	0.067	0.051	0.003	0.006	0.009	0.009	0.008	0.007
PERF	0.096	0.091	0.082	0.071	0.059	0.044	0.003	0.007	0.009	0.009	0.008	0.006
STOCK_ISS	0.098	0.092	0.081	0.070	0.058	0.043	0.004	0.008	0.009	0.009	0.008	0.006
ACCR	0.093	0.087	0.079	0.070	0.060	0.048	0.001	0.002	0.004	0.004	0.004	0.004
BEH_FIN	0.099	0.089	0.077	0.067	0.057	0.043	0.005	0.009	0.010	0.010	0.009	0.007
QMJ	0.095	0.086	0.076	0.066	0.055	0.040	0.004	0.008	0.010	0.010	0.009	0.007
UMD	0.094	0.087	0.076	0.065	0.055	0.043	0.002	0.004	0.005	0.005	0.004	0.003
SMB*	0.092	0.083	0.073	0.063	0.052	0.039	0.001	0.003	0.004	0.004	0.004	0.003
HML_DEVIL	0.085	0.073	0.067	0.066	0.062	0.050	0.002	0.004	0.006	0.009	0.011	0.010
CMA	0.095	0.084	0.071	0.059	0.049	0.037	0.004	0.007	0.007	0.006	0.005	0.004
SKEW	0.089	0.081	0.071	0.060	0.048	0.036	0.002	0.005	0.006	0.006	0.005	0.004
ASS_Growth	0.093	0.081	0.068	0.057	0.047	0.035	0.003	0.005	0.005	0.005	0.004	0.003
COMP_ISSUE	0.091	0.077	0.065	0.055	0.046	0.034	0.003	0.004	0.005	0.004	0.004	0.003
LTRev	0.089	0.076	0.064	0.053	0.043	0.032	0.001	0.003	0.003	0.003	0.003	0.002
RMW	0.088	0.074	0.062	0.052	0.043	0.032	0.002	0.003	0.004	0.004	0.003	0.003
ROA	0.089	0.075	0.062	0.051	0.041	0.030	0.002	0.004	0.004	0.004	0.003	0.002
GR_PROF	0.087	0.073	0.061	0.051	0.042	0.031	0.000	0.001	0.000	0.000	0.000	0.000
SMB	0.086	0.073	0.061	0.051	0.040	0.029	0.002	0.004	0.004	0.004	0.004	0.002
DISSTR	0.084	0.069	0.059	0.051	0.043	0.033	0.001	0.000	-0.001	-0.002	-0.003	-0.002
HML	0.086	0.070	0.057	0.048	0.039	0.030	0.002	0.003	0.003	0.003	0.003	0.002
O_SCORE	0.084	0.069	0.056	0.045	0.036	0.027	-0.001	-0.002	-0.002	-0.002	-0.001	-0.001

Posterior probabilities of factors, $\mathbb{E}\left[\gamma_j|\text{data}\right]$, and posterior mean of factors' risk prices, $\mathbb{E}\left[\lambda_j|\text{data}\right]$, are computed using the the Dirac spike-and-slab approach of Section 2.3.1.2 and 51 factors described in Table 2.A.1 of Appendix. Sample: 1973:10-2016:12. Test assets: 34 tradable factors and 26 investment anomalies. Prior probability of a factor being included is about 9.58% since we give each possible model equal prior probability and a factor could be included in a model with up to four other variables. Posterior probabilities are plotted for prior Sharpe ratio $\in [0.5, 3]$.

strongly outperforms the latter.[38]

2.6 Conclusions and Extensions

We develop a novel (Bayesian) method for the analysis of linear factor models in asset pricing. This approach can handle quadrillions of models generated by the zoo of traded and non-traded factors and delivers inference that is robust to the common identification failures caused by weak and level factors.

We apply our approach to the study of more than 2 quadrillion factor model specifications and find that: 1) only a handful of factors seem to be robust explanators of the cross-sections of asset returns; 2) jointly, the three to six robust factors provide a model that substantially outperforms notable benchmarks; 3) nevertheless, with very high probability the "true" latent SDF is dense in the space of factors proposed in the previous literature, likely containing 23–25 observable factors; and 4) a BMA over the universe of possible models delivers an SDF that presents a novel benchmark for in- and out-of-sample empirical asset pricing.

Our method can be feasibly extended to accommodate several salient extensions. First, one might want to bound the maximum price of risk (or the maximum Sharpe ratios) associated with the factors. This can be achieved by replacing the Gaussian distributions in our spike-and-slab priors with (rescaled and centred) Beta distributions, since the latter have bounded support. Furthermore, for the sake of expositional simplicity and closed-form solutions, we have focused on regularizing spike-and-slab priors with exponential tails. Nevertheless, our approach, which shrinks weak (and level) factors based on their correlation with asset returns, could also be implemented using polynomial tailed (i.e., heavy-tailed) mixing priors (see Polson and Scott (2011) for a general discussion of priors for regularization and shrinkage).[39] The rationale for heavy-tailed priors is that when the likelihood has thick tails while the prior has a thin tail, if the likelihood peak moves too far from the prior, the posterior eventually reverts toward the prior. Nevertheless, note that this mechanism (first pointed out in Jeffreys (1961)) is actually *desirable* in our settings in order to shrink the price of risk of useless factors toward zero.[40]

Second, thanks to its hierarchical structure, our approach can formally handle the statisti-

---

[38] These additional results are available upon request.

[39] For example, albeit alternative distributions with desirable properties exist, our spike-and-slab could be implemented using a Cauchy prior with location parameter set to zero and scale parameter proportional to ψ_j, as defined in equations (2.22) and (2.23).

[40] Since useless factors tend to generate heavy-tailed likelihoods (in the limit, the likelihood is an improper "uniform" on $\mathbb{R}$), with peaks for price of risk that deviate toward infinity, the posterior price of risk of such factors is shrunk toward the prior (zero) mean if the prior has thin tails.

cal uncertainty caused by generated factors, for example, mimicking portfolios, and provides valid inference in their presence. Furthermore, it can accommodate a wide range of both priced and unpriced latent factors.

Third, thanks to the hierarchical structure of our method, time-varying expected returns and SDF factor loadings could be accommodated by adopting the time-varying parameter approach of Primiceri (2005). Furthermore, although this would significantly increase the numerical complexity of the cross-sectional inference step, the time-varying parameters formulation could also be used for the modelling time-varying factor price of risk.

Appendices

2.A.1 Additional Derivations and Proofs

2.A.1.1 Derivation of the Posterior Distributions in Section 2.3

Let's consider first the time series layer of our hierarchical model. We assume that $\boldsymbol{Y_t} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$. The likelihood function of the observed time-series data $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^{T}$ is

$$p(\mathbf{Y}|\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^{T}(\boldsymbol{Y_t}-\boldsymbol{\mu_Y})(\boldsymbol{Y_t}-\boldsymbol{\mu_Y})^{\top}]}$$
$$\propto |\boldsymbol{\Sigma_Y}|^{-\frac{T}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^{T}(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)^{\top}+T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)^{\top}]},$$

where $\hat{\boldsymbol{\mu}}_Y = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{Y_t}$. After assigning a diffuse prior for $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$, $\pi(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{p+1}{2}}$, the posterior distribution function of $(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y})$ is

$$p(\boldsymbol{\mu_Y}, \boldsymbol{\Sigma_Y}|\mathbf{Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T+p+1}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^{T}(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)^{\top}+T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)^{\top}]},$$

Hence, the posterior distribution of $\boldsymbol{\mu_Y}$ conditional on $\boldsymbol{Y}$ and $\boldsymbol{\Sigma_Y}$ is

$$p(\boldsymbol{\mu_Y}|\mathbf{Y}, \boldsymbol{\Sigma_Y}) \propto e^{-\frac{1}{2}tr[T\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{\mu_Y}-\hat{\boldsymbol{\mu}}_Y)^{\top}]}$$

and the above is the kernel of the multivariate normal in equation (2.6). If we further integrate out $\boldsymbol{\mu_Y}$, it is easy to show that $p(\boldsymbol{\Sigma_Y}|\mathbf{Y}) \propto |\boldsymbol{\Sigma_Y}|^{-\frac{T+p}{2}} e^{-\frac{1}{2}tr[\boldsymbol{\Sigma_Y}^{-1}\sum_{t=1}^{T}(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)(\boldsymbol{Y_t}-\hat{\boldsymbol{\mu}}_Y)^{\top}]}$. Therefore, the posterior distribution of $\boldsymbol{\Sigma}$ is the inverse-Wishart in equation (2.7).

Recall that $\boldsymbol{C} = (\boldsymbol{1_N}, \boldsymbol{C_f})$, $\boldsymbol{\lambda}^{\top} = (\lambda_c, \boldsymbol{\lambda_f}^{\top})$. Assuming $\alpha_i \sim$ iid $\mathcal{N}(0, \sigma^2)$, the cross-sectional likelihood function conditional on the time-series parameters ($\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$), $p(data|\boldsymbol{\lambda}, \sigma^2)$, is given in equation (2.10), where "data" in this cross-sectional (second) step include the observed time-series $\boldsymbol{Y} = \{\boldsymbol{Y_t}\}_{t=1}^{T}$, as well as $\boldsymbol{\mu_Y}$ and $\boldsymbol{\Sigma_Y}$ drawn from the time-series step.

Assuming the diffuse prior $\pi(\boldsymbol{\lambda}, \sigma^2) \propto \sigma^{-2}$, the posterior distribution of $(\boldsymbol{\lambda}, \sigma^2)$ is

$$p(\boldsymbol{\lambda}, \sigma^2|data) \propto (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})^{\top}(\boldsymbol{\mu_R}-\boldsymbol{C\lambda})} = (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}}+\boldsymbol{C}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))^{\top}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}}+\boldsymbol{C}(\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}))}$$
$$= (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^{\top}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})-\frac{1}{2\sigma^2}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\top}\boldsymbol{C}^{\top}\boldsymbol{C}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})}, \text{ and}$$
$$\therefore \ p(\boldsymbol{\lambda}|\sigma^2, data) \propto e^{-\frac{(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\top}\boldsymbol{C}^{\top}\boldsymbol{C}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})}{2\sigma^2}},$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^{\top}\boldsymbol{C})^{-1}\boldsymbol{C}^{\top}\boldsymbol{\mu_R}$, $\hat{\sigma}^2 = \frac{(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})^{\top}(\boldsymbol{\mu_R}-\boldsymbol{C\hat{\lambda}})}{N}$. Note that sending $\sigma^2 \to 0$ the posterior $p(\boldsymbol{\lambda}|\sigma^2, data)$ is proportional to a Dirac at $\hat{\boldsymbol{\lambda}}$ as per Definition 2.2. For non-degenerate values of σ^2, the conditional posterior of $\boldsymbol{\lambda}$ is instead the one in equation (2.11). We derive the posterior of σ^2 by integrating out $\boldsymbol{\lambda}$ in the joint posterior, $p(\sigma^2|data) = \int p(\boldsymbol{\lambda}, \sigma^2|data)d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N-K+1}{2}} e^{-\frac{N\hat{\sigma}^2}{2\sigma^2}}$, hence, obtaining equation (2.12).

Under the GLS distributional assumption, $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0_N}, \sigma^2\boldsymbol{\Sigma_R})$, where $\boldsymbol{\Sigma_R}$ is the covari-

ance matrix of returns $\boldsymbol{R_t}$, the posterior of $(\boldsymbol{\lambda}, \sigma^2)$ is then

$$p(\boldsymbol{\lambda}, \sigma^2 | data) \propto (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda})^\top \boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R} - \boldsymbol{C}\boldsymbol{\lambda})}$$

$$= (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})^\top \boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}}) - \frac{1}{2\sigma^2}(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})^\top \boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{C}(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})}, \text{ and}$$

$$\therefore \ p(\boldsymbol{\lambda} | \sigma^2, data) \propto e^{-\frac{(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})^\top \boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{C}(\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})}{2\sigma^2}},$$

where $\hat{\boldsymbol{\lambda}} = (\boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{C})^{-1} \boldsymbol{C}^\top \boldsymbol{\Sigma_R^{-1}} \boldsymbol{\mu_R}$ and the above is the kernel of a Gaussian distribution. Note that sending $\sigma^2 \to 0$, the posterior $p(\boldsymbol{\lambda} | \sigma^2, data)$ is proportional to a Dirac at $\hat{\boldsymbol{\lambda}}$ as per Definition 2.3. For non-degenerate values of σ^2 the conditional posterior of $\boldsymbol{\lambda}$ is instead the one in equation (2.13). Further integrating out $\boldsymbol{\lambda}$, we obtain $p(\sigma^2 | data) = \int p(\boldsymbol{\lambda}, \sigma^2 | data) d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N-K+1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})^\top \boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R} - \boldsymbol{C}\hat{\boldsymbol{\lambda}})}$. Hence, the posterior of σ^2 is as in equation (2.14).

2.A.1.2 Formal Derivation of the Flat Prior Pitfall for Risk Prices

Following the derivation in Section 2.A.1.1, the cross-sectional likelihood is given by equation (2.10). Assigning a flat prior to the parameters[41] $(\boldsymbol{\lambda}, \sigma^2)$, the marginal cross-sectional likelihood function conditional on model index $\boldsymbol{\gamma}$ is

$$p(data | \boldsymbol{\gamma}) = \iint p(data | \boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma^2) \pi(\boldsymbol{\lambda}, \sigma^2 | \boldsymbol{\gamma}) d\boldsymbol{\lambda} d\sigma^2 \propto \iint (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top (\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})} d\boldsymbol{\lambda} d\sigma^2$$

$$= \iint (\sigma^2)^{-\frac{N+2}{2}} e^{-\frac{N\hat{\sigma}_\gamma^2}{2\sigma^2}} e^{-\frac{(\boldsymbol{\lambda_\gamma} - \hat{\boldsymbol{\lambda}}_\gamma)^\top \boldsymbol{C_\gamma^\top} \boldsymbol{C_\gamma}(\boldsymbol{\lambda_\gamma} - \hat{\boldsymbol{\lambda}}_\gamma)}{2\sigma^2}} d\boldsymbol{\lambda} d\sigma^2$$

$$= (2\pi)^{\frac{p_\gamma}{2}} |\boldsymbol{C_\gamma^\top} \boldsymbol{C_\gamma}|^{-\frac{1}{2}} \int (\sigma^2)^{-\frac{N-p_\gamma+2}{2}} e^{-\frac{N\hat{\sigma}_\gamma^2}{2\sigma^2}} d\sigma^2 = (2\pi)^{\frac{p_\gamma}{2}} |\boldsymbol{C_\gamma^\top} \boldsymbol{C_\gamma}|^{-\frac{1}{2}} \frac{\Gamma(\frac{N-p_\gamma}{2})}{(\frac{N\hat{\sigma}_\gamma^2}{2})^{\frac{N-p_\gamma}{2}}},$$

where $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C_\gamma^\top} \boldsymbol{C_\gamma})^{-1} \boldsymbol{C_\gamma^\top} \boldsymbol{\mu_R}$, $\hat{\sigma}_\gamma^2 = \frac{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\hat{\boldsymbol{\lambda}}_\gamma)^\top (\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\hat{\boldsymbol{\lambda}}_\gamma)}{N}$ and Γ denotes the Gamma function.

---

[41]More precisely, the priors for $(\boldsymbol{\lambda}, \sigma^2)$ are $\pi(\boldsymbol{\lambda_\gamma}, \sigma^2) \propto \frac{1}{\sigma^2}$ and $\boldsymbol{\lambda_{-\gamma}} = 0$.

2.A.1.3 Proof of Proposition 2.3

Sampling $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$. From Bayes' theorem we have that

$$p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) \propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})$$
$$\propto (2\pi)^{-\frac{p_{\gamma}}{2}}|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}(\sigma^2)^{-\frac{N+p_{\gamma}}{2}}e^{-\frac{1}{2\sigma^2}[(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})+\boldsymbol{\lambda_\gamma}^\top\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}]}$$
$$= (2\pi)^{-\frac{p_{\gamma}}{2}}|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}(\sigma^2)^{-\frac{N+p_{\gamma}}{2}}e^{-\frac{(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_\gamma)^\top(\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma})(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_\gamma)}{2\sigma^2}}e^{-\frac{SSR_\gamma}{2\sigma^2}},$$

where $SSR_\gamma = \boldsymbol{\mu_R^\top}\boldsymbol{\mu_R} - \boldsymbol{\mu_R^\top}C_\gamma(\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top}\boldsymbol{\mu_R} = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})+$ $\boldsymbol{\lambda_\gamma^\top}\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$. Hence, defining $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top}\boldsymbol{\mu_R}$ and $\hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2(\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma})^{-1}$, we obtain the posterior distribution in (2.16).

Using our priors and integrating out $\boldsymbol{\lambda}$ yields

$$p(data|\sigma^2, \boldsymbol{\gamma}) = \int p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N}{2}}\frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma}|^{\frac{1}{2}}}e^{-\frac{SSR_\gamma}{2\sigma^2}}.$$

Sampling σ^2. From Bayes theorem, the posterior of σ^2 is $p(\sigma^2|data, \boldsymbol{\gamma}) \propto p(data|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2) \propto$ $(\sigma^2)^{-\frac{N}{2}-1}e^{-\frac{SSR_\gamma}{2\sigma^2}}$. Hence, the posterior distribution of σ^2 is the inverse-Gamma in (2.17).

Finally, we obtain the marginal likelihood of the data in (2.18) by integrating out σ^2 as follows:

$$p(data|\boldsymbol{\gamma}) = \int p(data|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)d\sigma^2 \propto \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma}|^{\frac{1}{2}}}\frac{1}{(SSR_\gamma/2)^{\frac{N}{2}}},$$

where $SSR_\gamma = \boldsymbol{\mu_R^\top}\boldsymbol{\mu_R} - \boldsymbol{\mu_R^\top}C_\gamma(\boldsymbol{C_\gamma^\top C_\gamma}+\boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top}\boldsymbol{\mu_R}$.

2.A.1.4 Proof of Proposition 2.4

Sampling $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$. From Bayes' theorem we have that

$$p(\boldsymbol{\lambda}|data, \sigma^2, \boldsymbol{\gamma}) \propto p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})$$
$$\propto (2\pi)^{-\frac{p_{\gamma}}{2}}|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}(\sigma^2)^{-\frac{N+p_{\gamma}}{2}}e^{-\frac{1}{2\sigma^2}[(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R}-\boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})+\boldsymbol{\lambda_\gamma}^\top\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}]}$$
$$= (2\pi)^{-\frac{p_{\gamma}}{2}}|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}(\sigma^2)^{-\frac{N+p_{\gamma}}{2}}e^{-\frac{(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_\gamma)^\top(\boldsymbol{C_\gamma^\top}\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C_\gamma}+\boldsymbol{D_\gamma})(\boldsymbol{\lambda_\gamma}-\hat{\boldsymbol{\lambda}}_\gamma)}{2\sigma^2}}e^{-\frac{SSR_\gamma}{2\sigma^2}},$$

where $SSR_\gamma = \min_{\boldsymbol{\lambda}_\gamma}\{(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma})^\top\boldsymbol{\Sigma_R^{-1}}(\boldsymbol{\mu_R} - \boldsymbol{C_\gamma}\boldsymbol{\lambda_\gamma}) + \boldsymbol{\lambda_\gamma^\top}\boldsymbol{D_\gamma}\boldsymbol{\lambda_\gamma}\}$. Hence, defining $\hat{\boldsymbol{\lambda}}_\gamma = (\boldsymbol{C_\gamma^\top}\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C_\gamma} + \boldsymbol{D_\gamma})^{-1}\boldsymbol{C_\gamma^\top}\boldsymbol{\Sigma_R^{-1}}\boldsymbol{\mu_R}, \; \hat{\sigma}^2(\hat{\boldsymbol{\lambda}}_\gamma) = \sigma^2(\boldsymbol{C_\gamma^\top}\boldsymbol{\Sigma_R^{-1}}\boldsymbol{C_\gamma} + \boldsymbol{D_\gamma})^{-1}$, we obtain the posterior distribution in (2.19).

Using our priors and integrating out $\boldsymbol{\lambda}$ yields

$$p(data|\sigma^2, \boldsymbol{\gamma}) = \int p(data|\boldsymbol{\lambda}, \sigma^2, \boldsymbol{\gamma})\pi(\boldsymbol{\lambda}|\sigma^2, \boldsymbol{\gamma})d\boldsymbol{\lambda} \propto (\sigma^2)^{-\frac{N}{2}} \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}} e^{-\frac{SSR_{\gamma}}{2\sigma^2}}.$$

Obviously, the posterior distribution of σ^2 is identical to that in equation (2.20).

Finally, we obtain the marginal likelihood of the data in (2.21) by integrating out σ^2 as follows:

$$p(data|\boldsymbol{\gamma}) = \int p(data|\sigma^2, \boldsymbol{\gamma})\pi(\sigma^2)d\sigma^2 \propto \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{R}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma}/2)^{\frac{N}{2}}}.$$

2.A.1.5 Proof of Corollary 2.1

To begin with, we introduce the following matrix notations:

$$\boldsymbol{C}_{\boldsymbol{\gamma}} = (\boldsymbol{C}_{\boldsymbol{\gamma}'}, \boldsymbol{C}_{\boldsymbol{p}}), \quad \boldsymbol{D}_{\boldsymbol{\gamma}} = \begin{pmatrix} \boldsymbol{D}_{\boldsymbol{\gamma}'} & \boldsymbol{0} \\ \boldsymbol{0} & \psi_p^{-1} \end{pmatrix},$$

where $\boldsymbol{0}$ denotes conformable matrices of zeros.

Under the spherical (OLS) distributional assumption for pricing errors $\boldsymbol{\alpha}$,

$$\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}} = \begin{pmatrix} \boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'} & \boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{p}} \\ \boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} & \boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{p}} + \psi_p^{-1} \end{pmatrix},$$

$|\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}}| = |\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'}| \times |\boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{p}} + \psi_p^{-1} - \boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'}(\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{p}}|$, and
$|\boldsymbol{D}_{\boldsymbol{\gamma}}| = |\boldsymbol{D}_{\boldsymbol{\gamma}'}| \times \psi_p^{-1}$. Equipped with the above, we have by direct calculation

$$\frac{p(data|\gamma_j = 1, \boldsymbol{\gamma}_{-j})}{p(data|\gamma_j = 0, \boldsymbol{\gamma}_{-j})} = \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}} + \boldsymbol{D}_{\boldsymbol{\gamma}}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma}/2)^{\frac{N}{2}}} \Bigg/ \frac{|\boldsymbol{D}_{\boldsymbol{\gamma}'}|^{\frac{1}{2}}}{|\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma'}/2)^{\frac{N}{2}}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \psi_p^{-\frac{1}{2}} \left|\boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{p}} + \psi_p^{-1} - \boldsymbol{C}_{\boldsymbol{p}}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'}\left(\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'}\right)^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{p}}\right|^{-\frac{1}{2}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \left(1 + \psi_p \boldsymbol{C}_{\boldsymbol{p}}^{\top}\left[\boldsymbol{I_N} - \boldsymbol{C}_{\boldsymbol{\gamma}'}\left(\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'}\right)^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\right]\boldsymbol{C}_{\boldsymbol{p}}\right)^{-\frac{1}{2}},$$

where $\boldsymbol{C}_{\boldsymbol{p}}^{\top}\left[\boldsymbol{I_N} - \boldsymbol{C}_{\boldsymbol{\gamma}'}(\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\boldsymbol{C}_{\boldsymbol{\gamma}'} + \boldsymbol{D}_{\boldsymbol{\gamma}'})^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}'}^{\top}\right]\boldsymbol{C}_{\boldsymbol{p}} = \min_{\boldsymbol{b}}\{(\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{\gamma}'}\boldsymbol{b})^{\top}(\boldsymbol{C}_{\boldsymbol{p}} - \boldsymbol{C}_{\boldsymbol{\gamma}'}\boldsymbol{b}) + \boldsymbol{b}^{\top}\boldsymbol{D}_{\boldsymbol{\gamma}'}\boldsymbol{b}\}$
is the minimal value of the penalized sum of squared errors when we use $\boldsymbol{C}_{\boldsymbol{\gamma}'}$ to predict $\boldsymbol{C}_{\boldsymbol{p}}$.

Similar to the above, in the non-spherical (GLS) pricing errors case we have

$$C_{\gamma}^{\top}\Sigma_R^{-1}C_{\gamma} + D_{\gamma} = \begin{pmatrix} C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'} & C_{\gamma'}^{\top}\Sigma_R^{-1}C_p \\ C_p^{\top}\Sigma_R^{-1}C_{\gamma'} & C_p^{\top}\Sigma_R^{-1}C_p + \psi_p^{-1} \end{pmatrix},$$

$|C_{\gamma}^{\top}\Sigma_R^{-1}C_{\gamma} + D_{\gamma}| = |C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'}| \times |C_p^{\top}\Sigma_R^{-1}C_p + \frac{1}{\psi_p} - C_p^{\top}\Sigma_R^{-1}C_{\gamma'}(C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'})^{-1}C_{\gamma'}^{\top}\Sigma_R^{-1}C_p|,$ and
$|D_{\gamma}| = |D_{\gamma'}| \times \psi_p^{-1}$. Equipped with the above, we have by direct calculation

$$\frac{p(data|\gamma_j = 1, \gamma_{-j})}{p(data|\gamma_j = 0, \gamma_{-j})} = \frac{|D_{\gamma}|^{\frac{1}{2}}}{|C_{\gamma}^{\top}\Sigma_R^{-1}C_{\gamma} + D_{\gamma}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma}/2)^{\frac{N}{2}}} \bigg/ \frac{|D_{\gamma'}|^{\frac{1}{2}}}{|C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'}|^{\frac{1}{2}}} \frac{1}{(SSR_{\gamma'}/2)^{\frac{N}{2}}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \psi_p^{-\frac{1}{2}} \left| C_p^{\top}\Sigma_R^{-1}C_p + \frac{1}{\psi_p} - C_p^{\top}\Sigma_R^{-1}C_{\gamma'}\left(C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'}\right)^{-1}C_{\gamma'}^{\top}\Sigma_R^{-1}C_p \right|^{-\frac{1}{2}}$$

$$= \left(\frac{SSR_{\gamma'}}{SSR_{\gamma}}\right)^{\frac{N}{2}} \left| 1 + \psi_p\left[C_p^{\top}\Sigma_R^{-1}C_p - C_p^{\top}\Sigma_R^{-1}C_{\gamma'}\left(C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'}\right)^{-1}C_{\gamma'}^{\top}\Sigma_R^{-1}C_p\right] \right|^{-\frac{1}{2}},$$

where $C_p^{\top}\Sigma_R^{-1}C_p - C_p^{\top}\Sigma_R^{-1}C_{\gamma'}\left(C_{\gamma'}^{\top}\Sigma_R^{-1}C_{\gamma'} + D_{\gamma'}\right)^{-1}C_{\gamma'}^{\top}\Sigma_R^{-1}C_p = \min_b\{(C_p - C_{\gamma'}b)^{\top}\Sigma_R^{-1}(C_p - C_{\gamma'}b) + b^{\top}D_{\gamma'}b\}$, which is the minimal value of the penalized sum of squared errors when we use $C_{\gamma'}$ to predict C_p, but the prediction errors are weighted by Σ_R^{-1}.

2.A.1.6   Proof of Propositions 2.6 and 2.7

Sampling λ_{γ}. Combining the likelihood and the prior for λ we have the following:

$$p(\lambda|data, \sigma^2, \gamma) \propto p(data|\lambda, \sigma^2, \gamma)p(\lambda|\sigma^2, \gamma) \propto e^{-\frac{1}{2\sigma^2}\left[\lambda^{\top}(C^{\top}C + D)\lambda - 2\lambda^{\top}C^{\top}\mu_R\right]}.$$

Therefore, defining $\hat{\lambda} = (C^{\top}C + D)^{-1}C^{\top}\mu_R$ and $\hat{\sigma}^2(\hat{\lambda}) = \sigma^2(C^{\top}C + D)^{-1}$, we have the posterior in equation (2.28).

Sampling $\{\gamma_j\}_{j=1}^{K}$. Given a ω_j, the conditional Bayes factor for the j-th risk factor is[42]

$$\frac{p(\gamma_j = 1|data, \lambda, \omega, \sigma^2, \gamma_{-j})}{p(\gamma_j = 0|data, \lambda, \omega, \sigma^2, \gamma_{-j})} = \frac{\omega_j}{1 - \omega_j} \frac{p(\lambda_j|\gamma_j = 1, \sigma^2)}{p(\lambda_j|\gamma_j = 0, \sigma^2)}$$

Sampling ω. From Bayes' theorem we have $p(\omega_j|data, \lambda, \gamma, \sigma^2) \propto \pi(\omega_j)\pi(\gamma_j|\omega_j) \propto \omega_j^{\gamma_j}(1 - \omega_j)^{1-\gamma_j}\omega_j^{a_\omega - 1}(1 - \omega_j)^{b_\omega - 1} \propto \omega_j^{\gamma_j + a_\omega - 1}(1 - \omega_j)^{1 - \gamma_j + b_\omega - 1}$. Therefore, the posterior distribution of ω_j is the Beta in equation (2.30).

Sampling σ^2. Finally, $p(\sigma^2|data, \omega, \lambda, \gamma) \propto (\sigma^2)^{-\frac{N+K+1}{2}-1}e^{-\frac{1}{2\sigma^2}[(\mu_R - C\lambda)^{\top}(\mu_R - C\lambda) + \lambda^{\top}D\lambda]}$. Hence, the posterior distribution of σ^2 is the inverse-Gamma in equation (2.31). The proof

---

[42]If we had instead imposed $\omega_j = 0.5$, as in Section 2.3.1.2, the Bayes factor would simply be $\frac{p(\lambda_j|\gamma_j = 1, \sigma^2)}{p(\lambda_j|\gamma_j = 0, \sigma^2)}$.

of Proposition 2.7 follows the same identical steps, and is therefore omitted for brevity.

2.A.2 Data

Table 2.A.1: List of factors for cross-sectional asset pricing models

Factor ID	Reference	Factor ID	Reference
MKT	Sharpe (1964, Journal of Finance), Lintner (1965, Journal of Finance)	HML_DEVIL	Asness and Frazzini (2013, Journal of Portfolio Management)
SMB	Fama and French (1992, Journal of Finance)	QMJ	Asness, Frazzini, and Pedersen (2019, Review of Accounting Studies)
HML	Fama and French (1992, Journal of Finance)	FIN_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019, AEJ: Macroeconomics)
RMW	Fama and French (2015, Journal of Financial Economics)	REAL_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019, AEJ: Macroeconomics)
CMA	Fama and French (2015, Journal of Financial Economics)	MACRO_UNC	Jurado, Ludvigson, and Ng (2015, American Economy Review), Ludvigson, Ma, and Ng (2019, AEJ: Macroeconomics)
UMD	Carhart (1997, Journal of Finance), Jegadeesh and Titman (1993, Journal of Finance)	TERM	Chen, Ross and Roll (1986, Journal of Business), Fama and French (1993, Journal of Financial Economics)
STRev	Jegadeesh and Titman (1993, Journal of Finance)	DELTA_SLOPE	Ferson and Harvey (1991, Journal of Political Economy)
LTRev	Jegadeesh and Titman (2001, Journal of Finance)	CREDIT	Chen, Ross and Roll (1986, Journal of Business), Fama and French (1993, Journal of Financial Economics)
q_IA	Hou, Xue, Zhang (2015, Review of Financial Studies)	DIV	Campbell (1996, Journal of Political Economy)
q_ROE	Hou, Xue, Zhang (2015, Review of Financial Studies)	PE	Basu (1977, Journal of Finance), Ball (1978, Journal of Financial Economics)
LIQ_NT	Pastor and Stambaugh (2003, Journal of Political Economy)	BW_INV_SENT	Baker and Wurgler (2006, Journal of Finance)
LIQ_TR	Pastor and Stambaugh (2003, Journal of Political Economy)	HJTZ_INV_SENT	Huang, Jiang, Tu, and Zhou (2015, Review of Financial Studies)
MGMT	Stambaugh and Yuan (2016, Review of Financial Studies)	BEH_PEAD	Daniel, Hirshleifer, and Sun (2019, Review of Financial Studies)
PERF	Stambaugh and Yuan (2016, Review of Financial Studies)	BEH_FIN	Daniel, Hirshleifer, and Sun (2019, Review of Financial Studies)
ACCR	Sloan (1996, Accounting Review)	MKT*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
DISSTR	Campbell, Hilscher, and Szilagyi (2008, Journal of Finance)	SMB*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
A_Growth	Cooper, Gulen, and Schill (2008, Journal of Finance)	HML*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
COMP_ISSUE	Daniel and Titman (2006, Journal of Finance)	RMW*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
GR_PROF	Novy-Marx (2013, Journal of Financial Economics)	CMA*	Daniel, Mota, Rottke, and Santos (2020, Review of Financial Studies)
INV_IN_ASSETS	Titman, Wei, and Xie (2004, Journal of Financial and Quantitative Analysis)	SKEW	Langlois (2019, Journal of Financial Economics)
NetOA	Hirshleifer, Kewei, Teoh, and Zhang (2004, Journal of Accounting and Economics)	NONDUR	Chen, Ross and Roll (1986, Journal of Business), Breeden, Gibbons, and Litzenberger (1989, Journal of Finance)
OSCORE	Ohlson (1980, Journal of Accounting Research)	SERV	Breeden, Gibbons, and Litzenberger (1989, Journal of Finance), Hall (1978, Journal of Political Economy)
ROA	Chen, Novy-Marx, and Zhang (2010, working paper)	UNRATE	Gertler and Grinols (1982, Journal of Money, Credit, and Banking)
STOCK_ISS	Ritter (1991, Journal of Finance), Fama and French (2008, Journal of Finance)	IND_PROD	Chan, Chen, and Hsieh (1985, Journal of Financial Economics), Chen, Ross and Roll (1986, Journal of Business)
INTERM_CR	He, Kelly, and Manela (2017, Journal of Financial Economics)	OIL	Chen, Ross and Roll (1986, Journal of Business)
BAB	Frazzini and Pedersen (2014, Journal of Financial Economics)		

The table presents the list of factors used in Section 2.5.1. For each of the variables, we present their identification index, the nature of the factor, and the source of data for downloading and/or constructing the time series. Full description of the factors, sources, and references can be found in Table OA13 of the Online Appendix.

Table 2.A.2: List of additional anomalies used for the construction of test assets

Anomaly ID	Reference	Anomaly ID	Reference
CashAssets	Palazzo (2012, Journal of Financial Economics)	Volume	Garfinkel (2009, Review of Accounting Studies)
FCFBook	Hou, Karolyi, and Kho (2011, Review of Financial Studies)	SGASales	Freyberger, Neuhierl, and Weber (2020, Review of Financial Studies)
CFPrice	Desai, Rajgopal, and Venkatachalam (2004, Accounting Review)	Q	Kaldor (1996, Review of Economic Studies)
CapTurnover	Haugen and Baker (1996, Journal of Financial Economics)	IVolCAPM	Ang, Hodrick, Xing, and Zhang (2006, Journal of Finance)
CapIntens	Gorodnichenko and Weber (2016, American Economic Review)	IVolFF3	Ang, Hodrick, Xing, and Zhang (2006, Journal of Finance)
DP_tr	Litzenberger and Ramaswamy (1979, Journal of Financial Economics)	DayVariance	Ang, Hodrick, Xing, and Zhang (2006, Journal of Finance)
PPE_delta	Lyandres, Sun, and Zhang (2008, Review of Financial Studies)	ProfMargin	Soliman (2008, Accounting Review)
Lev	Lewellen (2015, Critical Finance Review)	PriceCostMargin	Bustamante and Donangelo (2017, Review of Financial Studies)
SalesPrice	Lewellen (2015, Critical Finance Review)	OperLev	Novy-Marx (2011, Review of Finance)
IntermMom	Novy-Marx (2012, Journal of Financial Economics)	FixedCostSale	D'Acunto, Liu, Pflueger, and Weber (2018, Journal of Financial Economics)
YearHigh	George and Hwang (2004, Journal of Finance)	LTMom	Bondt and Thaler (1985, Journal of Finance)
PE_tr	Basu (1983, Journal of Financial Economics)	NetSalesNetOA	Soliman (2008, Accounting Review)
BidAsk	Chung and Zhang (2014, Journal of Financial Markets)	AssetsMarket	Bhandari (1988, Journal of Finance)

The table presents the list of anomalies, which, together with the tradable factors from Table 2.A.1, form a cross-section of test assets used in Section 2.5.1. For each of the variables, we present their identification index, the nature of the factor, and the source of data for downloading and/or constructing the time series. Full description of the factors, sources, and references can be found in the Table OA14 of the Online Appendix.

2.A.3 Additional Simulation Results: $N = 25$

Table 2.A.3: Price of risk tests in a misspecified model with a strong factor

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	R^2_{adj} 5th	95th
Panel A: OLS									
GMM	100	0.109	0.058	0.014	0.082	0.039	0.008	-3.92%	65.58%
	200	0.107	0.053	0.011	0.084	0.041	0.006	-3.33%	65.60%
	600	0.099	0.055	0.017	0.098	0.048	0.009	5.87%	64.16%
	1000	0.098	0.054	0.011	0.108	0.052	0.010	13.23%	61.49%
	20000	0.102	0.051	0.011	0.103	0.048	0.007	37.62%	49.45%
	100	0.064	0.029	0.006	0.047	0.019	0.002	8.14%	43.29%
	200	0.086	0.041	0.007	0.067	0.033	0.004	6.54%	49.57%
B-SDF, flat prior	600	0.087	0.050	0.018	0.097	0.046	0.009	8.72%	56.19%
	1000	0.092	0.052	0.011	0.104	0.053	0.010	13.65%	56.14%
	20000	0.099	0.052	0.010	0.104	0.051	0.009	37.36%	49.18%
	100	0.064	0.029	0.006	0.049	0.020	0.002	8.13%	43.27%
	200	0.086	0.040	0.008	0.069	0.033	0.004	6.53%	49.56%
B-SDF, normal prior	600	0.088	0.049	0.018	0.098	0.047	0.009	8.72%	56.18%
	1000	0.092	0.052	0.011	0.106	0.056	0.011	13.65%	56.14%
	20000	0.099	0.052	0.010	0.102	0.051	0.009	37.36%	49.18%
Panel B: GLS									
GMM	100	0.105	0.051	0.009	0.082	0.041	0.007	-3.96%	16.32%
	200	0.109	0.055	0.010	0.097	0.048	0.009	-3.43%	16.10%
	600	0.104	0.057	0.014	0.121	0.067	0.015	-1.65%	14.80%
	1000	0.109	0.058	0.012	0.124	0.067	0.015	-0.37%	13.53%
	20000	0.096	0.050	0.012	0.140	0.082	0.016	4.88%	8.44%
	100	0.140	0.078	0.017	0.056	0.024	0.003	-0.64%	9.35%
	200	0.116	0.059	0.017	0.074	0.036	0.005	-1.70%	9.80%
B-SDF, flat prior	600	0.104	0.061	0.014	0.096	0.048	0.008	-1.08%	11.34%
	1000	0.107	0.057	0.014	0.099	0.048	0.011	-0.09%	11.55%
	20000	0.092	0.049	0.011	0.112	0.059	0.008	4.96%	8.28%
	100	0.139	0.078	0.017	0.059	0.023	0.004	-0.65%	9.34%
	200	0.116	0.060	0.017	0.075	0.037	0.005	-1.70%	9.79%
B-SDF, normal prior	600	0.104	0.061	0.014	0.093	0.048	0.007	-1.08%	11.34%
	1000	0.106	0.057	0.014	0.101	0.047	0.010	-0.09%	11.54%
	20000	0.092	0.049	0.011	0.115	0.060	0.009	4.96%	8.28%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of the cross-sectional R^2_{adj} is 43.87% (6.69%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.239, 1.305, 1.386, 1.413, and 1.497, for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.4: Price of risk tests in a misspecified model with a useless factor

	T	$\lambda_{intercept}$ 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
				Panel A. OLS					
GMM	100	0.079	0.039	0.009	0.095	0.033	0.001	-4.28%	39.58%
	200	0.088	0.041	0.006	0.151	0.054	0.004	-4.27%	43.35%
	600	0.092	0.041	0.007	0.332	0.176	0.023	-4.26%	40.40%
	1000	0.097	0.046	0.007	0.429	0.254	0.043	-4.28%	38.52%
	20000	0.193	0.114	0.040	0.832	0.674	0.279	-4.28%	39.93%
B-SDF, flat prior	100	0.045	0.018	0.003	0.003	0.001	0.000	6.45%	18.00%
	200	0.064	0.023	0.004	0.016	0.003	0.000	4.91%	20.19%
	600	0.072	0.033	0.005	0.041	0.019	0.003	3.56%	20.78%
	1000	0.080	0.031	0.006	0.059	0.027	0.003	3.23%	21.46%
	20000	0.068	0.019	0.002	0.103	0.052	0.010	2.74%	22.95%
B-SDF, normal prior	100	0.073	0.034	0.006	0.003	0.001	0.000	-2.92%	8.17%
	200	0.101	0.051	0.011	0.016	0.003	0.000	-3.91%	2.43%
	600	0.125	0.064	0.022	0.041	0.019	0.003	-4.27%	-2.74%
	1000	0.139	0.083	0.025	0.059	0.027	0.003	-4.31%	-3.59%
	20000	0.693	0.577	0.340	0.103	0.052	0.010	-4.35%	-4.34%
				Panel B. GLS					
GMM	100	0.096	0.047	0.007	0.144	0.073	0.008	-4.06%	12.88%
	200	0.105	0.056	0.008	0.239	0.154	0.039	-3.95%	12.69%
	600	0.115	0.060	0.012	0.444	0.353	0.190	-3.53%	14.05%
	1000	0.127	0.068	0.015	0.546	0.456	0.281	-3.41%	12.88%
	20000	0.367	0.270	0.111	0.886	0.862	0.801	-2.85%	13.80%
B-SDF, flat prior	100	0.117	0.058	0.014	0.026	0.007	0.000	-0.16%	6.83%
	200	0.096	0.044	0.013	0.040	0.016	0.002	-0.89%	5.65%
	600	0.093	0.047	0.007	0.067	0.025	0.005	-1.36%	6.31%
	1000	0.094	0.049	0.007	0.070	0.030	0.008	-1.38%	6.21%
	20000	0.206	0.084	0.006	0.101	0.052	0.012	-0.88%	7.40%
B-SDF, normal prior	100	0.127	0.071	0.016	0.026	0.007	0.000	-1.50%	4.62%
	200	0.106	0.054	0.016	0.040	0.016	0.002	-2.76%	1.59%
	600	0.117	0.061	0.015	0.067	0.025	0.005	-3.74%	-0.54%
	1000	0.126	0.068	0.016	0.070	0.030	0.008	-3.86%	-1.19%
	20000	0.661	0.527	0.307	0.101	0.052	0.012	-3.20%	-2.40%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda_{useless}^* = 0$ in a misspecified model with an intercept and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of R^2 is 0%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5.

Table 2.A.5: Price of risk tests in a correctly specified model with a strong factor

	T	$\lambda_{intercept}$ 10%	5%	1%	λ_{HML} 10%	5%	1%	R^2_{adj} 5th	95th
				Panel A. OLS					
GMM	100	0.110	0.057	0.014	0.081	0.039	0.009	-3.69%	77.22%
	200	0.107	0.054	0.011	0.084	0.043	0.007	-2.37%	82.54%
	600	0.101	0.055	0.017	0.098	0.045	0.010	27.73%	91.62%
	1000	0.099	0.054	0.012	0.109	0.053	0.009	51.47%	93.90%
	20000	0.101	0.049	0.011	0.102	0.049	0.007	97.50%	99.61%
B-SDF, flat prior	100	0.064	0.028	0.006	0.048	0.018	0.002	10.87%	48.55%
	200	0.087	0.042	0.007	0.072	0.031	0.005	11.45%	59.80%
	600	0.087	0.048	0.018	0.096	0.043	0.009	22.39%	76.64%
	1000	0.092	0.051	0.011	0.103	0.054	0.009	38.42%	83.06%
	20000	0.098	0.051	0.011	0.103	0.050	0.008	96.39%	98.57%
B-SDF, normal prior	100	0.064	0.029	0.006	0.050	0.018	0.002	10.86%	48.53%
	200	0.088	0.042	0.007	0.074	0.032	0.005	11.44%	59.79%
	600	0.086	0.049	0.018	0.095	0.045	0.009	22.39%	76.63%
	1000	0.092	0.051	0.011	0.103	0.055	0.010	38.42%	83.06%
	20000	0.099	0.050	0.011	0.104	0.051	0.009	96.39%	98.57%
				Panel B. GLS					
GMM	100	0.107	0.052	0.009	0.059	0.029	0.004	-3.87%	23.28%
	200	0.113	0.056	0.009	0.072	0.032	0.006	-3.44%	32.06%
	600	0.106	0.053	0.014	0.088	0.044	0.008	3.14%	50.57%
	1000	0.111	0.058	0.012	0.097	0.047	0.008	13.68%	60.17%
	20000	0.099	0.052	0.012	0.111	0.051	0.007	88.68%	95.75%
B-SDF, flat prior	100	0.136	0.071	0.014	0.055	0.026	0.003	0.00%	12.64%
	200	0.125	0.065	0.015	0.074	0.037	0.007	-0.51%	16.42%
	600	0.109	0.057	0.016	0.095	0.045	0.008	2.56%	29.66%
	1000	0.110	0.056	0.012	0.104	0.050	0.009	7.90%	38.25%
	20000	0.098	0.050	0.013	0.117	0.053	0.008	82.06%	89.50%
B-SDF, normal prior	100	0.136	0.071	0.014	0.056	0.027	0.003	0.00%	12.63%
	200	0.125	0.064	0.015	0.075	0.040	0.007	-0.51%	16.41%
	600	0.109	0.058	0.016	0.095	0.044	0.008	2.56%	29.66%
	1000	0.110	0.056	0.012	0.105	0.051	0.008	7.90%	38.25%
	20000	0.099	0.050	0.013	0.117	0.052	0.009	82.06%	89.49%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a correctly specified model with an intercept and a strong factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The hypothetical true value of R^2 is 100%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.239, 1.305, 1.386, 1.413, and 1.497 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.6: Price of risk tests in a correctly specified model with a useless factor

	T	$\lambda_{intercept}$			$\lambda_{useless}$			R^2_{adj}	
		10%	5%	1%	10%	5%	1%	5th	95th
Panel A. OLS									
	100	0.076	0.038	0.010	0.069	0.019	0.002	-4.25%	45.71%
	200	0.081	0.036	0.006	0.105	0.033	0.000	-4.29%	52.03%
GMM	600	0.086	0.038	0.008	0.269	0.112	0.008	-4.20%	55.98%
	1000	0.091	0.041	0.006	0.365	0.184	0.013	-4.22%	57.63%
	20000	0.174	0.104	0.036	0.842	0.683	0.266	-4.10%	62.61%
	100	0.041	0.017	0.003	0.002	0.000	0.000	7.76%	19.56%
	200	0.060	0.021	0.004	0.013	0.002	0.000	6.88%	22.42%
B-SDF, flat prior	600	0.067	0.033	0.003	0.038	0.015	0.003	7.28%	27.84%
	1000	0.076	0.024	0.004	0.058	0.026	0.004	7.49%	31.19%
	20000	0.054	0.011	0.000	0.093	0.049	0.009	8.01%	38.97%
	100	0.072	0.033	0.006	0.002	0.000	0.000	-2.77%	9.52%
	200	0.101	0.051	0.010	0.013	0.002	0.000	-3.84%	3.54%
B-SDF, normal prior	600	0.125	0.064	0.021	0.038	0.015	0.003	-4.24%	-2.04%
	1000	0.139	0.083	0.025	0.058	0.026	0.004	-4.29%	-3.13%
	20000	0.692	0.574	0.340	0.093	0.049	0.009	-4.35%	-4.33%
Panel B. GLS									
	100	0.094	0.045	0.008	0.069	0.025	0.001	-4.14%	13.29%
	200	0.109	0.056	0.009	0.091	0.038	0.004	-4.19%	12.99%
GMM	600	0.120	0.067	0.015	0.139	0.072	0.012	-4.13%	13.82%
	1000	0.140	0.075	0.017	0.157	0.086	0.014	-4.19%	11.77%
	20000	0.569	0.442	0.228	0.658	0.588	0.427	-4.28%	11.85%
	100	0.119	0.061	0.010	0.011	0.002	0.000	0.35%	7.84%
	200	0.111	0.053	0.012	0.012	0.005	0.000	-0.25%	5.56%
B-SDF, flat prior	600	0.108	0.057	0.010	0.017	0.003	0.000	-0.70%	4.65%
	1000	0.118	0.070	0.016	0.013	0.005	0.000	-0.95%	4.06%
	20000	0.567	0.415	0.178	0.080	0.037	0.008	-2.01%	5.38%
	100	0.132	0.068	0.013	0.011	0.002	0.000	-1.17%	6.15%
	200	0.119	0.061	0.012	0.012	0.005	0.000	-2.52%	2.30%
B-SDF, normal prior	600	0.120	0.065	0.015	0.017	0.003	0.000	-3.56%	-1.09%
	1000	0.136	0.077	0.020	0.013	0.005	0.000	-3.81%	-2.00%
	20000	0.692	0.569	0.338	0.080	0.037	0.008	-4.29%	-4.03%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda_{useless}^* = 0$ in a correctly specified model with an intercept and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of R^2 is 0%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5.

Table 2.A.7: Price of risk tests in a correctly specified model with useless and strong factors

	T	$\lambda_{intercept}$ 10%	5%	1%	λ_{HML} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
						Panel A. OLS						
	100	0.085	0.038	0.007	0.065	0.031	0.003	0.036	0.008	0.000	-2.65%	79.55%
	200	0.090	0.037	0.004	0.062	0.023	0.003	0.031	0.008	0.000	3.19%	84.45%
GMM	600	0.075	0.035	0.010	0.069	0.033	0.005	0.041	0.012	0.000	35.99%	92.46%
	1000	0.083	0.040	0.006	0.077	0.035	0.004	0.039	0.005	0.000	57.55%	94.54%
	20000	0.073	0.038	0.006	0.072	0.032	0.003	0.035	0.007	0.001	97.95%	99.64%
	100	0.035	0.013	0.001	0.033	0.008	0.000	0.002	0.000	0.000	20.39%	54.29%
	200	0.057	0.021	0.002	0.044	0.016	0.001	0.002	0.001	0.000	20.05%	63.99%
B-SDF, flat prior	600	0.059	0.025	0.007	0.055	0.025	0.004	0.005	0.002	0.000	30.94%	79.15%
	1000	0.069	0.033	0.005	0.065	0.029	0.002	0.004	0.000	0.000	45.76%	85.03%
	20000	0.066	0.031	0.005	0.070	0.026	0.002	0.007	0.002	0.000	96.99%	98.73%
	100	0.062	0.028	0.005	0.049	0.020	0.002	0.002	0.000	0.000	10.50%	48.47%
	200	0.084	0.039	0.007	0.070	0.030	0.005	0.002	0.001	0.000	9.12%	58.83%
B-SDF, normal prior	600	0.087	0.047	0.018	0.090	0.044	0.009	0.005	0.002	0.000	19.32%	75.71%
	1000	0.093	0.052	0.011	0.105	0.052	0.009	0.004	0.000	0.000	35.52%	82.38%
	20000	0.099	0.050	0.011	0.103	0.050	0.009	0.006	0.002	0.000	96.23%	98.51%
						Panel B. GLS						
	100	0.099	0.046	0.007	0.059	0.025	0.003	0.067	0.027	0.001	-7.36%	25.68%
	200	0.105	0.050	0.009	0.068	0.030	0.004	0.069	0.029	0.002	-6.57%	34.26%
GMM	600	0.098	0.049	0.012	0.077	0.035	0.006	0.077	0.036	0.003	1.52%	52.00%
	1000	0.106	0.052	0.010	0.085	0.040	0.007	0.075	0.030	0.003	12.51%	60.51%
	20000	0.089	0.045	0.010	0.095	0.051	0.007	0.092	0.033	0.001	88.59%	95.77%
	100	0.118	0.056	0.010	0.045	0.019	0.003	0.012	0.002	0.000	-0.92%	12.91%
	200	0.109	0.056	0.012	0.059	0.029	0.004	0.010	0.004	0.000	-1.28%	16.54%
B-SDF, flat prior	600	0.101	0.049	0.012	0.076	0.034	0.005	0.009	0.001	0.000	2.31%	29.64%
	1000	0.101	0.052	0.010	0.088	0.041	0.006	0.005	0.001	0.000	7.30%	38.35%
	20000	0.090	0.041	0.007	0.096	0.045	0.008	0.005	0.000	0.000	82.07%	89.53%
	100	0.131	0.064	0.013	0.054	0.023	0.003	0.012	0.002	0.000	-2.53%	10.97%
	200	0.122	0.066	0.015	0.073	0.038	0.006	0.010	0.004	0.000	-3.64%	14.15%
B-SDF, normal prior	600	0.109	0.056	0.015	0.090	0.046	0.009	0.009	0.001	0.000	-1.37%	26.96%
	1000	0.107	0.058	0.013	0.102	0.050	0.010	0.004	0.001	0.000	3.92%	35.71%
	20000	0.100	0.048	0.012	0.113	0.054	0.009	0.005	0.000	0.000	81.20%	89.00%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda_{useless}^* \equiv 0$ in a misspecified model with an intercept, a strong, and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of the cross-sectional R^2 is 100%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.239, 1.305, 1.386, 1.413, and 1.497 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.8: Price of risk tests in a misspecified model with useless and strong factors, robustness check: $\psi \in \{2, 10\}$

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
					Panel A: OLS							
	100	0.062	0.029	0.005	0.050	0.020	0.002	0.002	0.001	0.000	5.98%	42.02%
B-SDF, normal prior	200	0.087	0.040	0.008	0.073	0.031	0.006	0.006	0.002	0.000	3.01%	47.81%
$\psi = 2$	600	0.087	0.048	0.018	0.096	0.044	0.009	0.020	0.006	0.001	4.76%	54.20%
	1000	0.095	0.052	0.011	0.106	0.052	0.009	0.040	0.011	0.002	9.55%	54.10%
	20000	0.100	0.050	0.010	0.105	0.053	0.009	0.089	0.043	0.009	34.47%	46.84%
	100	0.060	0.027	0.005	0.047	0.019	0.002	0.003	0.001	0.000	8.98%	44.71%
B-SDF, normal prior	200	0.084	0.039	0.008	0.066	0.031	0.005	0.006	0.001	0.000	4.48%	48.88%
$\psi = 10$	600	0.085	0.048	0.017	0.093	0.043	0.010	0.019	0.006	0.001	4.99%	54.37%
	1000	0.095	0.052	0.011	0.105	0.051	0.010	0.040	0.013	0.002	9.71%	54.16%
	20000	0.100	0.050	0.011	0.102	0.050	0.009	0.089	0.043	0.009	34.47%	46.84%
					Panel B: GLS							
	100	0.133	0.071	0.014	0.059	0.026	0.003	0.029	0.008	0.000	-4.19%	6.71%
B-SDF, normal prior	200	0.113	0.058	0.019	0.076	0.036	0.006	0.035	0.013	0.001	-5.64%	6.48%
$\psi = 2$	600	0.106	0.061	0.013	0.096	0.049	0.011	0.068	0.029	0.004	-5.53%	7.62%
	1000	0.107	0.055	0.013	0.101	0.049	0.011	0.075	0.036	0.007	-4.55%	7.69%
	20000	0.090	0.045	0.010	0.114	0.057	0.008	0.105	0.050	0.011	0.62%	4.10%
	100	0.129	0.068	0.013	0.051	0.022	0.002	0.029	0.009	0.000	-2.88%	8.49%
B-SDF, normal prior	200	0.108	0.057	0.018	0.073	0.032	0.006	0.034	0.012	0.001	-4.48%	8.14%
$\psi = 10$	600	0.106	0.061	0.013	0.089	0.045	0.008	0.068	0.029	0.004	-5.00%	8.18%
	1000	0.107	0.056	0.014	0.101	0.045	0.009	0.075	0.035	0.007	-4.19%	8.02%
	20000	0.091	0.046	0.010	0.111	0.058	0.008	0.105	0.050	0.011	0.63%	4.10%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda_{useless}^* \equiv 0$ in a misspecified model with an intercept, a strong and a useless factor. The true value of the cross-sectional R^2_{adj} is 43.87% (6.69%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 2 or 10. When $\psi = 2$, the normal prior implies a prior SR of the factor model equal to 1.009, 1.104, 1.235, 1.285, and 1.459 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$. Similarly, if $\psi = 10$, the normal prior implies a prior SR of the factor model equal to 1.359, 1.402, 1.450, 1.465, and 1.510 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$

2.A.3.1 Large N behavior

In this section we investigate the properties of the B-SDF procedure in estimating price of risk and measure of fit, as well as successfully identifying irrelevant factors in the model, when applied to a large cross-section.

We consider the same simulation design as described at the beginning of Section 2.4, except for the choice of the cross-section of test assets, which time series and cross-sectional features we mimic. In our baseline case in the previous subsections, we built a cross-section to emulate the 25 Fama-French portfolios, sorted by size and value. Now instead we consider the properties of the following composite cross-sections to simulate returns:

(a) $N = 55$: 25 Fama-French portfolios, sorted by size and value and 30 industry portfolios;

(b) $N = 100$: 25 Fama-French portfolios, sorted by size and value, 30 industry portfolios, 25 profitability and investment portfolios, 10 momentum portfolios, and 10 long-term reversal portfolios.

The rest of the simulation design stays unchanged; that is, the strong factor mimics the behavior of HML, with its betas and risk premia corresponding to their in-sample values, cross-sectional R^2_{adj}, portfolio average returns, and variance of the residuals.

Table 2.A.9: Price of risk tests in a misspecified model with a strong factor ($N = 55$)

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	R^2_{adj} 5th	95th
		Panel A: OLS							
GMM	100	0.113	0.059	0.010	0.072	0.032	0.004	-1.82%	30.72%
	200	0.103	0.051	0.012	0.085	0.037	0.004	-1.83%	30.92%
	600	0.092	0.045	0.008	0.090	0.040	0.006	-1.70%	27.37%
	1000	0.106	0.050	0.009	0.099	0.047	0.009	-1.29%	25.41%
	20000	0.108	0.053	0.013	0.106	0.057	0.014	6.44%	14.05%
B-SDF, flat prior	100	0.053	0.016	0.001	0.002	0.001	0.000	-0.85%	0.81%
	200	0.069	0.033	0.006	0.029	0.012	0.001	-1.27%	9.42%
	600	0.084	0.037	0.007	0.071	0.032	0.005	-1.52%	19.54%
	1000	0.100	0.047	0.009	0.088	0.038	0.006	-1.55%	20.65%
	20000	0.105	0.056	0.014	0.104	0.060	0.014	6.20%	13.73%
B-SDF, normal prior	100	0.053	0.016	0.001	0.002	0.001	0.000	-0.85%	0.81%
	200	0.069	0.032	0.006	0.030	0.012	0.001	-1.27%	9.42%
	600	0.084	0.037	0.007	0.072	0.032	0.005	-1.52%	19.54%
	1000	0.100	0.047	0.009	0.087	0.038	0.006	-1.55%	20.65%
	20000	0.105	0.056	0.014	0.104	0.058	0.014	6.20%	13.73%
		Panel B: GLS							
GMM	100	0.127	0.073	0.019	0.073	0.032	0.003	-1.07%	21.51%
	200	0.112	0.061	0.016	0.102	0.054	0.005	-0.44%	22.63%
	600	0.123	0.055	0.014	0.123	0.067	0.014	3.01%	21.13%
	1000	0.124	0.065	0.015	0.141	0.074	0.014	4.43%	20.17%
	20000	0.111	0.049	0.014	0.148	0.081	0.025	11.28%	15.27%
B-SDF, flat prior	100	0.283	0.201	0.085	0.035	0.012	0.001	-0.48%	24.04%
	200	0.174	0.106	0.036	0.068	0.029	0.006	-0.79%	19.81%
	600	0.119	0.065	0.018	0.083	0.039	0.005	1.27%	18.44%
	1000	0.115	0.065	0.020	0.095	0.043	0.006	3.00%	18.57%
	20000	0.103	0.047	0.009	0.111	0.059	0.012	11.04%	15.23%
B-SDF, normal prior	100	0.283	0.201	0.085	0.037	0.013	0.001	-0.48%	24.04%
	200	0.174	0.106	0.036	0.069	0.031	0.006	-0.79%	19.81%
	600	0.119	0.065	0.018	0.085	0.039	0.006	1.27%	18.44%
	1000	0.115	0.065	0.020	0.096	0.045	0.006	3.00%	18.57%
	20000	0.103	0.047	0.009	0.111	0.058	0.012	11.04%	15.23%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of the cross-sectional R^2_{adj} is 10.18% (13.34%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.528, 1.636, 1.773, 1.822, and 1.978 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.10: Price of risk tests in a misspecified model with a useless factor ($N = 55$)

	T	$\lambda_{intercept}$ 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
		Panel A. OLS							
	100	0.096	0.045	0.010	0.091	0.033	0.004	-1.87%	18.05%
	200	0.102	0.052	0.011	0.130	0.060	0.011	-1.87%	15.10%
GMM	600	0.101	0.051	0.011	0.258	0.158	0.030	-1.86%	13.29%
	1000	0.128	0.064	0.014	0.342	0.239	0.084	-1.88%	14.23%
	20000	0.288	0.200	0.083	0.821	0.775	0.662	-1.87%	12.69%
	100	0.055	0.014	0.002	0.000	0.000	0.000	-1.02%	-0.48%
	200	0.084	0.034	0.007	0.002	0.000	0.000	-1.33%	-0.54%
B-SDF, flat prior	600	0.102	0.050	0.011	0.012	0.003	0.001	-1.50%	-0.45%
	1000	0.121	0.064	0.014	0.030	0.010	0.000	-1.55%	-0.28%
	20000	0.264	0.154	0.029	0.090	0.037	0.008	-1.64%	0.10%
	100	0.058	0.019	0.002	0.000	0.000	0.000	-1.09%	-0.52%
	200	0.087	0.043	0.007	0.002	0.000	0.000	-1.53%	-0.81%
B-SDF, normal prior	600	0.108	0.054	0.014	0.012	0.003	0.001	-1.82%	-1.35%
	1000	0.130	0.075	0.016	0.030	0.009	0.000	-1.85%	-1.56%
	20000	0.443	0.316	0.143	0.090	0.037	0.008	-1.89%	-1.88%
		Panel B. GLS							
	100	0.125	0.071	0.019	0.104	0.056	0.005	-1.33%	21.50%
	200	0.104	0.052	0.014	0.202	0.120	0.032	-1.01%	22.14%
GMM	600	0.118	0.062	0.013	0.410	0.320	0.186	0.99%	19.94%
	1000	0.129	0.070	0.016	0.510	0.437	0.294	2.33%	19.23%
	20000	0.253	0.162	0.053	0.892	0.863	0.816	7.70%	15.65%
	100	0.283	0.199	0.084	0.057	0.026	0.002	-0.42%	25.52%
	200	0.178	0.109	0.038	0.041	0.009	0.001	-0.83%	19.54%
B-SDF, flat prior	600	0.128	0.077	0.021	0.059	0.027	0.004	-0.76%	17.53%
	1000	0.129	0.080	0.025	0.068	0.030	0.005	0.43%	17.23%
	20000	0.175	0.093	0.019	0.094	0.046	0.009	7.61%	12.41%
	100	0.285	0.200	0.084	0.057	0.026	0.002	-0.43%	25.51%
	200	0.180	0.110	0.038	0.041	0.009	0.001	-0.85%	19.47%
B-SDF, normal prior	600	0.139	0.079	0.024	0.059	0.027	0.004	-0.98%	16.90%
	1000	0.145	0.089	0.027	0.068	0.030	0.005	-0.43%	16.27%
	20000	0.342	0.234	0.095	0.094	0.046	0.009	6.86%	11.26%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda_{useless}^* = 0$ in a misspecified model with an intercept and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of R^2 is 0%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. For the uncorrelated useless factor, the normal prior implies a prior SR of 0 as the sample size T goes to infinity.

Table 2.A.11: Price of risk tests in a misspecified model with useless and strong factors ($N = 55$)

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
						Panel A: OLS						
	100	0.102	0.047	0.010	0.058	0.027	0.004	0.090	0.035	0.003	-2.92%	34.26%
	200	0.096	0.050	0.010	0.074	0.032	0.004	0.138	0.056	0.009	-2.99%	34.31%
GMM	600	0.088	0.039	0.006	0.084	0.038	0.005	0.257	0.167	0.040	-2.19%	30.22%
	1000	0.101	0.050	0.010	0.089	0.040	0.005	0.344	0.243	0.084	-1.72%	28.88%
	20000	0.113	0.059	0.015	0.080	0.033	0.006	0.816	0.773	0.654	5.92%	21.54%
	100	0.049	0.013	0.002	0.002	0.001	0.000	0.000	0.000	0.000	0.23%	4.87%
	200	0.065	0.025	0.004	0.024	0.006	0.000	0.002	0.000	0.000	-1.13%	18.26%
B-SDF, flat prior	600	0.072	0.031	0.006	0.059	0.020	0.003	0.016	0.003	0.000	-1.95%	22.44%
	1000	0.080	0.036	0.008	0.064	0.026	0.003	0.030	0.008	0.001	-2.13%	23.23%
	20000	0.041	0.017	0.002	0.023	0.010	0.001	0.086	0.040	0.006	5.78%	14.89%
	100	0.051	0.014	0.002	0.002	0.001	0.000	0.000	0.000	0.000	0.08%	4.36%
	200	0.071	0.030	0.004	0.030	0.008	0.001	0.002	0.000	0.000	-1.63%	16.04%
B-SDF, normal prior	600	0.083	0.034	0.007	0.070	0.031	0.004	0.017	0.003	0.000	-2.84%	19.26%
	1000	0.094	0.049	0.010	0.085	0.039	0.007	0.030	0.008	0.001	-3.13%	19.08%
	20000	0.107	0.054	0.015	0.104	0.061	0.012	0.086	0.040	0.006	4.36%	12.19%
						Panel B: GLS						
	100	0.125	0.070	0.018	0.069	0.034	0.003	0.099	0.050	0.007	-2.13%	22.64%
	200	0.104	0.055	0.014	0.101	0.047	0.006	0.196	0.114	0.027	-1.28%	22.85%
GMM	600	0.119	0.056	0.012	0.112	0.059	0.012	0.404	0.312	0.184	2.27%	21.91%
	1000	0.113	0.059	0.013	0.134	0.068	0.014	0.512	0.431	0.292	3.65%	21.31%
	20000	0.096	0.046	0.008	0.112	0.059	0.012	0.877	0.854	0.801	10.13%	17.87%
	100	0.278	0.197	0.084	0.028	0.008	0.001	0.053	0.027	0.003	-1.63%	24.93%
	200	0.167	0.102	0.034	0.064	0.027	0.003	0.038	0.009	0.001	-1.61%	19.05%
B-SDF, flat prior	600	0.113	0.061	0.015	0.077	0.033	0.005	0.059	0.028	0.003	0.81%	18.53%
	1000	0.111	0.060	0.016	0.082	0.038	0.005	0.068	0.031	0.004	2.47%	18.70%
	20000	0.049	0.018	0.000	0.036	0.015	0.001	0.092	0.042	0.009	10.02%	14.64%
	100	0.280	0.197	0.084	0.031	0.009	0.001	0.052	0.027	0.003	-1.64%	24.99%
	200	0.169	0.105	0.034	0.067	0.029	0.003	0.039	0.009	0.001	-1.67%	19.50%
B-SDF, normal prior	600	0.120	0.067	0.019	0.082	0.041	0.006	0.058	0.028	0.003	0.34%	17.73%
	1000	0.119	0.068	0.019	0.095	0.043	0.006	0.068	0.031	0.004	1.68%	17.45%
	20000	0.100	0.050	0.012	0.109	0.061	0.013	0.092	0.042	0.009	9.29%	13.60%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda_{useless}^* \equiv 0$ in a misspecified model with an intercept, a strong, and a useless factor. The true value of the cross-sectional R^2_{adj} is 10.18% (13.34%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.528, 1.636, 1.773, 1.822, and 1.978 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.12: Price of risk tests in a misspecified model with a strong factor ($N = 100$)

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	R^2_{adj} 5th	95th
		Panel A: OLS							
GMM	200	0.106	0.062	0.014	0.085	0.040	0.007	-0.98%	24.46%
	600	0.101	0.048	0.013	0.097	0.045	0.010	-0.70%	23.52%
	1000	0.101	0.058	0.012	0.097	0.048	0.014	-0.21%	20.38%
	20000	0.091	0.043	0.009	0.103	0.051	0.010	6.23%	11.74%
B-SDF, flat prior	200	0.082	0.039	0.006	0.022	0.006	0.000	-0.36%	1.74%
	600	0.089	0.043	0.008	0.065	0.032	0.004	-0.71%	17.12%
	1000	0.094	0.048	0.013	0.079	0.041	0.011	-0.74%	16.60%
	20000	0.092	0.042	0.010	0.098	0.047	0.011	6.02%	11.56%
B-SDF, normal prior	200	0.082	0.039	0.006	0.022	0.006	0.000	-0.36%	1.74%
	600	0.090	0.043	0.008	0.064	0.033	0.004	-0.71%	17.12%
	1000	0.094	0.048	0.013	0.079	0.041	0.011	-0.74%	16.60%
	20000	0.092	0.042	0.010	0.098	0.047	0.011	6.02%	11.56%
		Panel B: GLS							
GMM	200	0.097	0.055	0.014	0.110	0.056	0.010	-0.19%	18.33%
	600	0.121	0.071	0.015	0.163	0.101	0.025	1.68%	17.64%
	1000	0.115	0.066	0.013	0.170	0.098	0.028	2.95%	16.54%
	20000	0.104	0.054	0.013	0.177	0.105	0.036	8.37%	11.75%
B-SDF, flat prior	200	0.194	0.123	0.039	0.089	0.039	0.002	0.04%	10.53%
	600	0.123	0.067	0.018	0.122	0.065	0.015	0.16%	13.30%
	1000	0.109	0.062	0.012	0.119	0.066	0.016	1.26%	13.48%
	20000	0.088	0.040	0.010	0.095	0.046	0.010	8.07%	11.73%
B-SDF, normal prior	200	0.194	0.123	0.039	0.089	0.040	0.002	0.04%	10.53%
	600	0.123	0.067	0.018	0.126	0.067	0.016	0.16%	13.30%
	1000	0.109	0.062	0.012	0.121	0.066	0.017	1.26%	13.48%
	20000	0.088	0.040	0.010	0.094	0.045	0.010	8.07%	11.73%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_i^* in a misspecified model with an intercept and a strong factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of the cross-sectional R^2_{adj} is 8.98% (10.11%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.858, 2.010, 2.210, 2.285, and 2.529 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

Table 2.A.13: Price of risk tests in a misspecified model with a useless factor ($N = 100$)

	T	$\lambda_{intercept}$ 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
				Panel A. OLS					
GMM	200	0.108	0.060	0.015	0.206	0.117	0.031	-1.01%	13.30%
	600	0.116	0.064	0.014	0.406	0.296	0.130	-1.01%	13.39%
	1000	0.130	0.070	0.020	0.477	0.383	0.219	-1.00%	12.64%
	20000	0.357	0.251	0.110	0.875	0.841	0.764	-1.00%	12.49%
B-SDF, flat prior	200	0.105	0.051	0.007	0.000	0.000	0.000	-0.54%	-0.07%
	600	0.129	0.069	0.013	0.026	0.009	0.001	-0.72%	0.33%
	1000	0.135	0.076	0.017	0.047	0.017	0.003	-0.75%	0.45%
	20000	0.312	0.176	0.023	0.093	0.051	0.009	-0.79%	1.78%
B-SDF, normal prior	200	0.105	0.054	0.008	0.000	0.000	0.000	-0.58%	-0.11%
	600	0.142	0.077	0.019	0.026	0.009	0.001	-0.88%	-0.20%
	1000	0.153	0.087	0.025	0.047	0.017	0.003	-0.95%	-0.39%
	20000	0.623	0.500	0.274	0.093	0.051	0.009	-1.02%	-1.00%
				Panel B. GLS					
GMM	200	0.101	0.055	0.014	0.158	0.095	0.030	-0.59%	17.85%
	600	0.119	0.070	0.015	0.399	0.312	0.166	0.42%	16.67%
	1000	0.118	0.068	0.013	0.506	0.416	0.282	1.20%	15.96%
	20000	0.261	0.181	0.064	0.874	0.846	0.807	6.01%	11.11%
B-SDF, flat prior	200	0.200	0.129	0.043	0.020	0.007	0.001	-0.09%	10.29%
	600	0.122	0.069	0.016	0.063	0.026	0.003	-0.43%	12.94%
	1000	0.113	0.065	0.014	0.077	0.034	0.006	-0.25%	12.98%
	20000	0.178	0.098	0.023	0.100	0.057	0.011	5.89%	9.84%
B-SDF, normal prior	200	0.202	0.129	0.043	0.020	0.007	0.001	-0.10%	10.28%
	600	0.122	0.069	0.017	0.063	0.026	0.003	-0.48%	12.81%
	1000	0.119	0.064	0.016	0.077	0.034	0.006	-0.39%	12.70%
	20000	0.252	0.165	0.048	0.100	0.057	0.011	5.33%	9.07%

Frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true value of λ_c and $\lambda_{useless}^* = 0$ in a misspecified model with an intercept and a useless factor. Last two columns: 5th and 95th percentiles of cross-sectional R^2_{adj} across 2,000 simulations, evaluated at the point estimates for GMM and at the posterior mean for B-SDF. The true value of R^2 is 0%. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. For the uncorrelated useless factor, the normal prior implies a prior SR of 0 as the sample size T goes to infinity.

Table 2.A.14: Price of risk tests in a misspecified model with useless and strong factors ($N = 100$)

	T	λ_c 10%	5%	1%	λ_{strong} 10%	5%	1%	$\lambda_{useless}$ 10%	5%	1%	R^2_{adj} 5th	95th
					Panel A: OLS							
GMM	200	0.104	0.054	0.012	0.074	0.034	0.006	0.216	0.131	0.035	-1.26%	27.78%
	600	0.101	0.051	0.012	0.099	0.041	0.004	0.403	0.302	0.146	-0.73%	26.62%
	1000	0.095	0.056	0.008	0.094	0.040	0.009	0.488	0.398	0.231	0.22%	24.92%
	20000	0.104	0.056	0.010	0.098	0.048	0.006	0.866	0.841	0.778	6.33%	19.62%
B-SDF, flat prior	200	0.080	0.036	0.005	0.016	0.005	0.000	0.000	0.000	0.000	0.40%	9.23%
	600	0.081	0.040	0.007	0.060	0.024	0.002	0.023	0.005	0.001	-0.58%	19.69%
	1000	0.085	0.041	0.008	0.062	0.030	0.005	0.042	0.017	0.001	-0.59%	19.47%
	20000	0.024	0.004	0.001	0.020	0.003	0.001	0.087	0.042	0.009	6.16%	13.07%
B-SDF, normal prior	200	0.084	0.037	0.005	0.018	0.005	0.000	0.000	0.000	0.000	0.31%	8.57%
	600	0.092	0.043	0.009	0.070	0.033	0.003	0.023	0.005	0.001	-1.02%	17.83%
	1000	0.094	0.049	0.011	0.077	0.040	0.010	0.043	0.017	0.001	-1.23%	16.73%
	20000	0.091	0.042	0.009	0.099	0.049	0.012	0.087	0.042	0.009	5.04%	10.68%
					Panel B: GLS							
GMM	200	0.101	0.056	0.014	0.108	0.051	0.009	0.161	0.099	0.033	-0.73%	18.64%
	600	0.117	0.072	0.014	0.158	0.095	0.023	0.391	0.306	0.161	1.37%	17.53%
	1000	0.120	0.064	0.014	0.158	0.089	0.023	0.502	0.410	0.273	2.76%	16.73%
	20000	0.103	0.053	0.012	0.152	0.089	0.021	0.875	0.849	0.801	7.77%	12.81%
B-SDF, flat prior	200	0.191	0.124	0.039	0.084	0.034	0.003	0.024	0.009	0.001	-0.33%	10.58%
	600	0.122	0.064	0.016	0.122	0.061	0.017	0.059	0.024	0.002	-0.10%	12.80%
	1000	0.108	0.061	0.014	0.110	0.058	0.015	0.076	0.033	0.004	1.10%	13.36%
	20000	0.056	0.025	0.004	0.051	0.017	0.002	0.100	0.052	0.010	7.72%	11.56%
B-SDF, normal prior	200	0.193	0.123	0.040	0.087	0.036	0.003	0.024	0.009	0.001	-0.37%	10.57%
	600	0.121	0.064	0.017	0.127	0.064	0.019	0.059	0.024	0.002	-0.21%	12.75%
	1000	0.111	0.064	0.015	0.117	0.063	0.015	0.076	0.033	0.004	0.82%	13.12%
	20000	0.085	0.039	0.011	0.096	0.046	0.011	0.100	0.052	0.010	7.15%	10.81%

The table shows the frequency of rejecting the null hypothesis $H_0 : \lambda_i = \lambda_i^*$ for pseudo-true values of λ_c and λ_{strong}, $\lambda_{useless}^* \equiv 0$ in a misspecified model with an intercept, a strong, and a useless factor. The true value of the cross-sectional R^2_{adj} is 8.98% (10.11%) for the OLS (GLS) estimation. B-SDF estimates credible intervals of risk prices under (1) a flat prior or (2) a normal prior $b_j \sim \mathcal{N}(0, \sigma^2 \psi \tilde{\rho}_j^\top \tilde{\rho}_j T^d)$, where d is chosen to be 0.5, while ψ is equal to 5. The normal prior corresponds to a prior SR of the factor model equal to 1.858, 2.010, 2.210, 2.285, and 2.529 for $T \in \{100, 200, 600, 1,000, \text{ and } 20,000\}$.

2.A.3.2 Bayesian p-values

(a) Useless factor
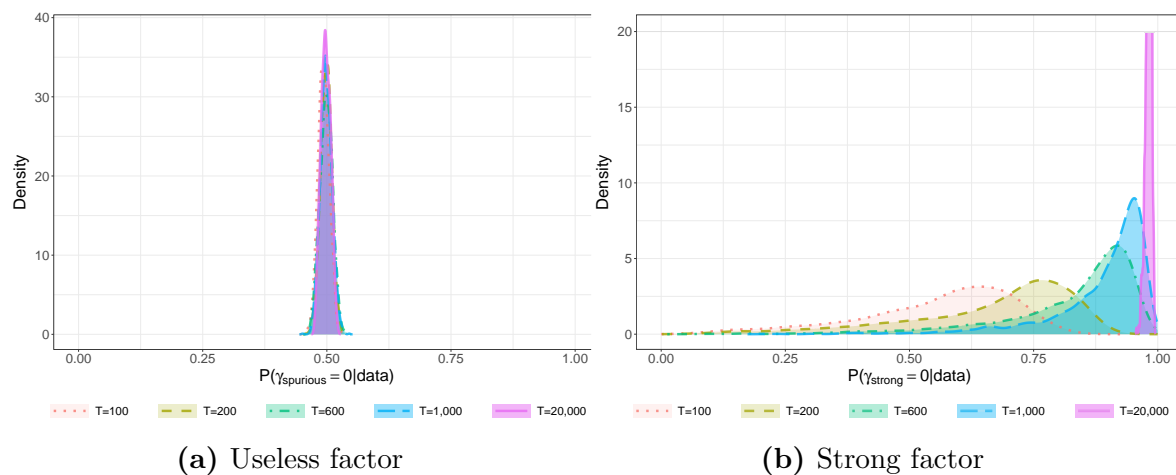
(b) Strong factor

Figure 2.A.1: Distribution of the Bayesian p-values for testing factor risk prices

Bayesian p-value, $1 - \Pr[\gamma = 1|data]$, of $H_0 : \lambda = \lambda_{pseudo-true}$, in misspecified models with both useless and strong factors, computed with the spike-and-slab prior of Section 2.3.1.2, as per Corollary 2.2, for different sample sizes. We set $\psi = 1$ in the estimation (which corresponds to a prior of $\sqrt{\mathbb{E}_\pi[SR_f^2 \mid \sigma^2]} = 0.295$) and $r = 0.001$.

2.A.4 Additional Results on the Main BMA Application

Table 2.A.15: Values of $\boldsymbol{\rho}_k^\top \boldsymbol{\rho}_k$ and $\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$ for each factor

Factor	$\boldsymbol{\rho}_k^\top \boldsymbol{\rho}_k$	$\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$	Factor	$\boldsymbol{\rho}_k^\top \boldsymbol{\rho}_k$	$\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$	Factor	$\boldsymbol{\rho}_k^\top \boldsymbol{\rho}_k$	$\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$
LIQ_NT	0.754	0.744	MKT	5.719	5.419	GR_PROF	5.01	4.84
INTERM_CAP_RATIO	3.407	3.233	SMB	7.087	7.087	INV_IN_ASS	4.007	3.683
FIN_UNC	0.353	0.339	HML	9.216	8.814	NetOA	2.252	2.103
REAL_UNC	0.137	0.137	RMW	7.575	7.031	O_SCORE	6.735	6.672
MACRO_UNC	0.162	0.162	CMA	8.757	8.157	ROA	8.002	7.533
TERM	0.155	0.155	UMD	4.133	3.8	STOCK_ISS	6.785	6.039
DEFAULT	0.145	0.144	STRev	2.573	2.51	BAB	4.439	3.981
DIV	0.108	0.108	LTRev	4.992	4.908	HML_DEVIL	8.024	8.014
UNRATE	0.226	0.224	IA	8.36	7.779	QMJ	9.555	8.879
PE	0.158	0.158	ROE	7.408	6.97	BEH_PEAD	2.604	2.466
BW_ISENT	0.736	0.575	LIQ_TR	1.3	1.28	BEH_FIN	10.223	9.292
HJTZ_ISENT	0.516	0.42	MGMT	9.603	8.714	MKT*	2.267	2.25
NONDUR	0.337	0.336	PERF	6.444	5.942	SMB*	3.426	3.321
SERV	0.152	0.152	ACCR	2.284	2.281	HML*	3.892	3.756
IPGrowth	0.116	0.115	DISSTR	7.553	6.954	RMW*	3.235	3.194
Oil	0.325	0.307	ASS_Growth	6.904	6.398	CMA*	2.869	2.673
DeltaSLOPE	0.167	0.164	COMP_ISSUE	7.455	6.72	SKEW	6.268	6.264

Values of $\boldsymbol{\rho}_k^\top \boldsymbol{\rho}_k$ and $\tilde{\boldsymbol{\rho}}_k^\top \tilde{\boldsymbol{\rho}}_k$ for each factor k, where $\boldsymbol{\rho}_k$ ($\tilde{\boldsymbol{\rho}}_k$) is an $N \times 1$ vector of (demeaned) correlation between factor k and test assets. Sample: 1973:10 to 2016:12. Test assets: cross-section of 34 tradable factors and 26 other investment anomalies. The 51 factors considered are described in Table 2.A.1.

Table 2.A.16: Posterior factor probabilities, $\mathbb{E}\left[\gamma_j|\text{data}\right]$, and risk prices: 2.25 quadrillion models, robustness check: $\omega_j \sim Beta(1,9)$

| | Factor inclusion prob., $\mathbb{E}\left[\gamma_j|\text{data}\right]$ | | | | | | Price of risk, $\mathbb{E}\left[\lambda_j|\text{data}\right]$ | | | | | |
| | Total prior SR: | | | | | | Total prior SR: | | | | | |
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
BEH_PEAD	0.179	0.231	0.252	0.244	0.196	0.121	0.020	0.043	0.059	0.066	0.061	0.067
BW_ISENT	0.108	0.120	0.129	0.162	0.189	0.141	0.002	0.006	0.012	0.027	0.055	0.063
CMA⋆	0.124	0.122	0.112	0.092	0.064	0.030	0.009	0.014	0.017	0.017	0.015	0.026
FIN_UNC	0.113	0.113	0.104	0.102	0.095	0.067	0.001	0.002	0.003	0.006	0.009	0.013
DIV	0.104	0.101	0.113	0.111	0.130	0.145	0.000	0.001	0.002	0.004	0.011	0.038
NONDUR	0.106	0.108	0.102	0.108	0.097	0.072	0.001	0.001	0.003	0.006	0.010	0.016
PE	0.107	0.106	0.102	0.107	0.107	0.097	0.000	-0.001	-0.002	-0.003	-0.007	-0.018
BAB	0.119	0.118	0.103	0.080	0.048	0.018	0.012	0.017	0.018	0.016	0.014	0.027
UNRATE	0.097	0.105	0.106	0.103	0.104	0.081	0.000	0.001	0.002	0.004	0.008	0.015
MACRO_UNC	0.103	0.098	0.108	0.100	0.092	0.066	0.000	0.000	0.001	0.001	0.001	-0.001
HJTZ_ISENT	0.099	0.099	0.102	0.108	0.105	0.069	0.001	0.002	0.004	0.008	0.013	0.014
TERM	0.100	0.100	0.098	0.102	0.098	0.098	0.000	-0.001	-0.002	-0.004	-0.007	-0.021
Oil	0.103	0.101	0.098	0.091	0.084	0.057	0.000	0.000	0.000	0.001	0.001	0.004
REAL_UNC	0.100	0.099	0.096	0.097	0.092	0.076	0.000	0.000	0.001	0.001	0.002	0.004
SERV	0.100	0.101	0.092	0.099	0.092	0.078	0.000	0.000	0.001	0.002	0.004	0.009
DeltaSLOPE	0.089	0.098	0.096	0.103	0.104	0.095	0.000	0.001	0.001	0.003	0.007	0.017
IPGrowth	0.091	0.094	0.096	0.095	0.101	0.086	0.000	0.000	-0.001	-0.002	-0.004	-0.011
DEFAULT	0.093	0.090	0.092	0.098	0.090	0.074	0.000	0.000	0.001	0.002	0.004	0.009
LIQ_NT	0.093	0.093	0.087	0.078	0.062	0.032	0.000	0.000	0.001	0.001	0.002	0.001
STRev	0.097	0.088	0.090	0.073	0.049	0.022	0.005	0.008	0.011	0.012	0.010	0.020
RMW⋆	0.100	0.093	0.079	0.064	0.041	0.019	0.006	0.009	0.010	0.010	0.009	0.021
ROE	0.109	0.092	0.072	0.054	0.036	0.018	0.012	0.014	0.013	0.012	0.014	0.039
MKT⋆	0.095	0.087	0.080	0.063	0.049	0.025	0.005	0.007	0.009	0.009	0.010	0.019
NetOA	0.101	0.089	0.075	0.057	0.039	0.018	0.005	0.007	0.008	0.007	0.007	0.013
MKT	0.096	0.086	0.075	0.055	0.041	0.030	0.008	0.011	0.013	0.011	0.013	0.045
IA	0.102	0.084	0.069	0.051	0.032	0.013	0.011	0.012	0.012	0.011	0.013	0.028
INV_IN_ASS	0.100	0.080	0.068	0.048	0.032	0.013	0.006	0.007	0.008	0.007	0.007	0.014
MGMT	0.093	0.080	0.062	0.049	0.032	0.014	0.011	0.013	0.012	0.012	0.014	0.031
ACCR	0.093	0.077	0.063	0.049	0.031	0.015	0.003	0.003	0.004	0.004	0.003	0.010
LIQ_TR	0.087	0.076	0.064	0.052	0.036	0.017	0.002	0.003	0.004	0.004	0.004	0.007
INTERM_CAP_RATIO	0.087	0.079	0.061	0.050	0.035	0.018	0.003	0.005	0.005	0.005	0.005	0.015
PERF	0.096	0.076	0.061	0.045	0.029	0.014	0.007	0.008	0.008	0.007	0.008	0.022
HML⋆	0.086	0.073	0.065	0.050	0.032	0.015	0.006	0.007	0.008	0.008	0.008	0.020
STOCK_ISS	0.088	0.070	0.057	0.041	0.027	0.011	0.007	0.007	0.007	0.007	0.007	0.017
UMD	0.087	0.071	0.054	0.043	0.030	0.012	0.004	0.005	0.005	0.004	0.005	0.010
BEH_FIN	0.086	0.071	0.053	0.039	0.024	0.010	0.008	0.009	0.008	0.008	0.010	0.019
QMJ	0.086	0.068	0.053	0.038	0.025	0.013	0.008	0.009	0.009	0.008	0.010	0.032
SMB⋆	0.082	0.064	0.052	0.040	0.024	0.011	0.003	0.003	0.004	0.003	0.003	0.008
SKEW	0.081	0.064	0.050	0.037	0.024	0.014	0.005	0.006	0.005	0.005	0.006	0.029
CMA	0.085	0.064	0.049	0.034	0.023	0.010	0.006	0.006	0.006	0.005	0.007	0.011
LTRev	0.077	0.066	0.049	0.034	0.022	0.010	0.002	0.003	0.003	0.003	0.003	0.005
HML_DEVIL	0.073	0.063	0.052	0.038	0.026	0.012	0.003	0.005	0.006	0.006	0.007	0.020
ASS_Growth	0.078	0.058	0.047	0.034	0.020	0.009	0.004	0.004	0.004	0.004	0.005	0.005
COMP_ISSUE	0.075	0.060	0.046	0.034	0.020	0.010	0.004	0.004	0.004	0.004	0.004	0.007
RMW	0.073	0.061	0.046	0.033	0.021	0.010	0.003	0.004	0.003	0.004	0.005	0.007
GR_PROF	0.073	0.058	0.042	0.032	0.020	0.010	0.001	0.001	0.000	0.000	0.000	0.002
ROA	0.075	0.054	0.041	0.031	0.020	0.010	0.004	0.003	0.004	0.003	0.005	0.013
HML	0.073	0.055	0.042	0.030	0.018	0.008	0.003	0.003	0.003	0.003	0.004	0.000
DISSTR	0.068	0.052	0.042	0.032	0.020	0.011	0.000	-0.001	-0.001	-0.001	-0.002	-0.012
O_SCORE	0.070	0.053	0.038	0.028	0.017	0.009	-0.002	-0.002	-0.002	-0.002	-0.002	-0.007
SMB	0.063	0.055	0.041	0.030	0.019	0.010	0.003	0.004	0.003	0.004	0.005	0.021

Posterior probabilities of factors, $\mathbb{E}\left[\gamma_j|\text{data}\right]$, and posterior mean of factors' risk prices, $\mathbb{E}\left[\lambda_j|\text{data}\right]$, are computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,9)$, yielding a prior expectation for γ_j equal to 10%. The data is monthly, 1973:10 to 2016:12. The 51 factors considered are described in Table 2.A.1 of the Appendix. Test assets: 34 tradable factors plus 26 investment anomalies.

Table 2.A.17: Posterior factor probabilities, $\mathbb{E}\left[\gamma_j | \text{data}\right]$, and risk prices: 2.25 quadrillion models with zero common intercept.

| | Factor inclusion prob., $\mathbb{E}\left[\gamma_j | \text{data}\right]$ | | | | | | Price of risk, $\mathbb{E}\left[\lambda_j | \text{data}\right]$ | | | | | |
| | Total prior SR: | | | | | | Total prior SR: | | | | | |
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
BEH_PEAD	0.552	0.625	0.705	0.790	0.866	0.840	0.020	0.047	0.092	0.159	0.252	0.311
MKT	0.531	0.579	0.634	0.676	0.714	0.626	0.023	0.055	0.099	0.154	0.231	0.278
CMA⋆	0.518	0.534	0.560	0.595	0.632	0.553	0.012	0.027	0.048	0.080	0.126	0.151
STRev	0.505	0.519	0.552	0.589	0.615	0.480	0.010	0.023	0.045	0.076	0.115	0.113
RMW⋆	0.502	0.516	0.533	0.574	0.610	0.492	0.011	0.024	0.045	0.079	0.128	0.135
HML_DEVIL	0.477	0.476	0.496	0.553	0.632	0.548	0.012	0.027	0.055	0.112	0.227	0.291
SKEW	0.504	0.517	0.529	0.540	0.533	0.376	0.018	0.038	0.062	0.091	0.124	0.110
QMJ	0.500	0.496	0.511	0.533	0.569	0.467	0.020	0.037	0.063	0.105	0.181	0.213
ROE	0.523	0.527	0.528	0.520	0.490	0.344	0.024	0.044	0.065	0.087	0.110	0.101
BW_ISENT	0.505	0.507	0.512	0.522	0.539	0.571	0.002	0.005	0.010	0.018	0.037	0.126
FIN_UNC	0.501	0.500	0.504	0.509	0.519	0.516	0.001	0.002	0.005	0.009	0.020	0.063
UNRATE	0.503	0.503	0.502	0.504	0.508	0.514	0.000	0.001	0.002	0.004	0.009	0.038
DIV	0.501	0.501	0.499	0.502	0.508	0.541	0.000	0.001	0.002	0.004	0.009	0.042
DEFAULT	0.498	0.498	0.500	0.503	0.506	0.505	0.000	0.001	0.001	0.003	0.007	0.023
MKT⋆	0.504	0.510	0.514	0.508	0.466	0.373	0.009	0.019	0.030	0.041	0.049	0.064
PE	0.498	0.501	0.500	0.497	0.503	0.498	0.000	-0.001	-0.001	-0.002	-0.006	-0.022
TERM	0.496	0.498	0.498	0.501	0.504	0.507	0.000	-0.001	-0.002	-0.003	-0.008	-0.031
HJTZ_ISENT	0.501	0.500	0.496	0.500	0.491	0.464	0.001	0.002	0.003	0.004	0.006	0.019
Oil	0.503	0.501	0.498	0.494	0.491	0.466	0.000	0.000	0.000	0.001	0.003	0.017
REAL_UNC	0.496	0.495	0.498	0.497	0.499	0.492	0.000	0.000	0.001	0.002	0.005	0.010
DeltaSLOPE	0.498	0.496	0.498	0.494	0.497	0.489	0.000	0.001	0.001	0.002	0.003	0.012
SERV	0.495	0.497	0.496	0.497	0.494	0.500	0.000	0.001	0.001	0.002	0.004	0.023
MACRO_UNC	0.497	0.498	0.495	0.493	0.494	0.480	0.000	0.000	0.001	0.002	0.005	0.006
NONDUR	0.492	0.493	0.496	0.495	0.499	0.491	0.001	0.002	0.003	0.006	0.012	0.048
IPGrowth	0.490	0.489	0.490	0.490	0.494	0.485	0.000	0.000	0.000	-0.001	-0.002	-0.003
INTERM_CAP_RATIO	0.497	0.504	0.499	0.485	0.462	0.358	0.008	0.017	0.027	0.034	0.035	0.019
LIQ_NT	0.492	0.490	0.488	0.482	0.480	0.424	0.000	0.000	0.000	-0.004	-0.013	-0.038
HML⋆	0.506	0.506	0.501	0.483	0.427	0.272	0.013	0.024	0.038	0.049	0.054	0.043
ACCR	0.489	0.485	0.486	0.479	0.475	0.329	0.005	0.012	0.023	0.038	0.058	0.053
PERF	0.504	0.495	0.490	0.474	0.439	0.289	0.015	0.027	0.039	0.053	0.068	0.058
BAB	0.511	0.511	0.498	0.463	0.413	0.285	0.017	0.029	0.039	0.045	0.051	0.051
MGMT	0.511	0.494	0.488	0.468	0.431	0.280	0.022	0.035	0.049	0.064	0.081	0.068
LIQ_TR	0.499	0.494	0.492	0.479	0.428	0.274	0.004	0.009	0.015	0.023	0.028	0.022
NetOA	0.499	0.496	0.482	0.467	0.411	0.269	0.008	0.015	0.023	0.030	0.033	0.030
IA	0.506	0.489	0.473	0.437	0.384	0.227	0.021	0.031	0.040	0.048	0.053	0.040
SMB	0.488	0.486	0.472	0.448	0.386	0.226	0.015	0.030	0.043	0.052	0.052	0.036
UMD	0.492	0.482	0.458	0.437	0.405	0.308	0.009	0.014	0.016	0.019	0.031	0.055
INV_IN_ASS	0.499	0.490	0.467	0.435	0.372	0.214	0.010	0.018	0.025	0.030	0.031	0.021
SMB⋆	0.486	0.473	0.456	0.435	0.390	0.246	0.007	0.012	0.018	0.025	0.032	0.029
STOCK_ISS	0.490	0.478	0.449	0.411	0.340	0.186	0.013	0.020	0.025	0.027	0.023	0.015
DISSTR	0.470	0.454	0.438	0.418	0.372	0.227	0.001	-0.004	-0.012	-0.022	-0.029	-0.025
GR_PROF	0.477	0.459	0.443	0.410	0.357	0.212	0.004	0.007	0.012	0.016	0.018	0.010
ROA	0.482	0.467	0.441	0.406	0.342	0.197	0.012	0.018	0.021	0.021	0.015	0.005
RMW	0.472	0.458	0.434	0.402	0.371	0.252	0.008	0.009	0.006	-0.004	-0.028	-0.038
HML	0.471	0.450	0.424	0.399	0.388	0.281	0.008	0.006	-0.002	-0.018	-0.053	-0.071
BEH_FIN	0.487	0.462	0.439	0.402	0.341	0.200	0.017	0.021	0.022	0.019	0.006	0.006
CMA	0.483	0.457	0.431	0.397	0.348	0.206	0.013	0.014	0.011	0.003	-0.014	-0.017
COMP_ISSUE	0.483	0.459	0.433	0.392	0.324	0.182	0.008	0.010	0.010	0.008	0.003	0.002
LTRev	0.479	0.458	0.427	0.380	0.317	0.173	0.007	0.009	0.009	0.006	0.001	0.001
ASS_Growth	0.477	0.452	0.424	0.382	0.317	0.173	0.009	0.010	0.008	0.004	-0.001	-0.003
O_SCORE	0.467	0.447	0.421	0.377	0.316	0.176	0.000	0.001	0.005	0.008	0.006	0.002

Posterior probabilities of factors, $\mathbb{E}\left[\gamma_j | \text{data}\right]$, and posterior mean of factors' risk prices, $\mathbb{E}\left[\lambda_j | \text{data}\right]$, are computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The data is monthly, 1973:10 to 2016:12. The 51 factors considered are described in Table 2.A.1 of the Appendix. Test assets: 34 tradable factors plus 26 investment anomalies.

Table 2.A.18: Posterior factor probabilities, $\mathbb{E}[\gamma_j|\text{data}]$, and risk prices: 2.25 quadrillion models with zero common intercept and non-demeaned correlations

| | Factor inclusion prob., $\mathbb{E}[\gamma_j|\text{data}]$ | | | | | | Price of risk, $\mathbb{E}[\lambda_j|\text{data}]$ | | | | | |
| | Total prior SR: | | | | | | Total prior SR: | | | | | |
Factors:	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
BEH_PEAD	0.556	0.621	0.701	0.793	0.863	0.838	0.020	0.047	0.091	0.158	0.250	0.310
MKT	0.529	0.579	0.631	0.682	0.718	0.630	0.023	0.055	0.099	0.156	0.232	0.279
CMA⋆	0.517	0.532	0.561	0.595	0.636	0.554	0.013	0.027	0.049	0.080	0.127	0.151
STRev	0.500	0.518	0.549	0.587	0.617	0.484	0.009	0.023	0.044	0.075	0.113	0.113
RMW⋆	0.507	0.508	0.532	0.571	0.618	0.492	0.010	0.023	0.044	0.077	0.127	0.134
HML_DEVIL	0.481	0.477	0.493	0.545	0.638	0.553	0.011	0.026	0.053	0.107	0.225	0.293
SKEW	0.505	0.516	0.529	0.537	0.532	0.389	0.017	0.037	0.060	0.088	0.122	0.114
BW_ISENT	0.506	0.508	0.515	0.525	0.550	0.583	0.003	0.006	0.012	0.022	0.045	0.142
QMJ	0.499	0.498	0.511	0.527	0.568	0.472	0.020	0.037	0.062	0.103	0.181	0.216
ROE	0.520	0.528	0.526	0.518	0.495	0.347	0.024	0.044	0.064	0.086	0.111	0.103
FIN_UNC	0.500	0.501	0.506	0.504	0.511	0.512	0.001	0.002	0.005	0.009	0.019	0.061
UNRATE	0.504	0.504	0.502	0.503	0.504	0.513	0.000	0.001	0.002	0.004	0.009	0.037
DIV	0.501	0.501	0.500	0.501	0.509	0.540	0.000	0.001	0.002	0.004	0.009	0.040
DEFAULT	0.499	0.499	0.502	0.502	0.505	0.501	0.000	0.001	0.001	0.003	0.007	0.022
MKT⋆	0.503	0.508	0.517	0.507	0.470	0.381	0.008	0.018	0.029	0.040	0.049	0.065
HJTZ_ISENT	0.501	0.502	0.503	0.500	0.495	0.462	0.001	0.002	0.003	0.005	0.007	0.018
TERM	0.495	0.498	0.500	0.500	0.506	0.510	0.000	-0.001	-0.001	-0.003	-0.007	-0.030
PE	0.498	0.498	0.501	0.498	0.501	0.503	0.000	-0.001	-0.001	-0.002	-0.006	-0.021
Oil	0.501	0.501	0.503	0.498	0.491	0.468	0.000	0.000	0.000	0.001	0.003	0.017
DeltaSLOPE	0.498	0.497	0.497	0.501	0.495	0.493	0.000	0.001	0.001	0.002	0.003	0.012
REAL_UNC	0.496	0.495	0.498	0.497	0.501	0.491	0.000	0.000	0.001	0.002	0.004	0.010
MACRO_UNC	0.497	0.496	0.493	0.494	0.497	0.478	0.000	0.000	0.001	0.002	0.004	0.005
SERV	0.495	0.496	0.494	0.496	0.492	0.496	0.000	0.001	0.001	0.002	0.004	0.022
NONDUR	0.491	0.494	0.497	0.495	0.495	0.490	0.001	0.001	0.003	0.005	0.012	0.046
INTERM_CAP_RATIO	0.497	0.501	0.502	0.491	0.466	0.357	0.008	0.017	0.027	0.035	0.035	0.019
IPGrowth	0.489	0.489	0.488	0.490	0.491	0.486	0.000	0.000	0.000	-0.001	-0.002	-0.002
LIQ_NT	0.492	0.491	0.489	0.482	0.479	0.421	0.000	0.000	0.000	-0.003	-0.013	-0.037
ACCR	0.489	0.488	0.485	0.489	0.481	0.335	0.005	0.011	0.022	0.038	0.058	0.054
HML⋆	0.505	0.502	0.498	0.478	0.426	0.275	0.012	0.024	0.037	0.048	0.053	0.043
LIQ_TR	0.500	0.499	0.494	0.480	0.431	0.276	0.004	0.009	0.015	0.022	0.027	0.022
MGMT	0.511	0.501	0.488	0.468	0.430	0.274	0.023	0.036	0.050	0.065	0.081	0.067
PERF	0.500	0.494	0.484	0.468	0.440	0.286	0.015	0.027	0.039	0.052	0.068	0.056
BAB	0.514	0.513	0.491	0.461	0.407	0.278	0.017	0.030	0.039	0.046	0.051	0.049
NetOA	0.498	0.494	0.487	0.466	0.405	0.267	0.008	0.015	0.023	0.029	0.032	0.029
SMB	0.484	0.490	0.481	0.454	0.394	0.230	0.015	0.029	0.042	0.052	0.053	0.035
IA	0.508	0.488	0.470	0.442	0.378	0.225	0.021	0.031	0.040	0.048	0.052	0.041
UMD	0.493	0.479	0.458	0.432	0.402	0.306	0.009	0.014	0.016	0.019	0.030	0.055
INV_IN_ASS	0.498	0.489	0.464	0.433	0.370	0.212	0.011	0.018	0.025	0.030	0.031	0.021
SMB⋆	0.489	0.475	0.459	0.437	0.391	0.248	0.007	0.012	0.018	0.025	0.032	0.028
DISSTR	0.471	0.454	0.439	0.418	0.373	0.224	0.001	-0.005	-0.013	-0.022	-0.029	-0.025
STOCK_ISS	0.490	0.471	0.448	0.406	0.332	0.182	0.014	0.020	0.025	0.027	0.023	0.015
RMW	0.479	0.458	0.430	0.404	0.373	0.254	0.008	0.009	0.006	-0.003	-0.029	-0.039
GR_PROF	0.478	0.458	0.441	0.411	0.355	0.210	0.004	0.007	0.011	0.016	0.018	0.010
ROA	0.482	0.465	0.440	0.403	0.346	0.195	0.012	0.018	0.021	0.020	0.014	0.006
HML	0.472	0.443	0.423	0.403	0.387	0.284	0.008	0.005	-0.002	-0.017	-0.052	-0.072
BEH_FIN	0.485	0.463	0.435	0.398	0.337	0.196	0.017	0.021	0.022	0.019	0.009	0.006
CMA	0.479	0.461	0.431	0.399	0.348	0.207	0.013	0.014	0.011	0.002	-0.014	-0.017
COMP_ISSUE	0.486	0.459	0.429	0.384	0.322	0.177	0.009	0.010	0.010	0.008	0.003	0.003
LTRev	0.478	0.455	0.430	0.384	0.319	0.177	0.006	0.009	0.009	0.006	0.001	0.000
ASS_Growth	0.478	0.458	0.426	0.382	0.317	0.172	0.009	0.010	0.008	0.004	-0.001	-0.003
O_SCORE	0.468	0.444	0.420	0.383	0.324	0.179	-0.001	0.001	0.005	0.008	0.006	0.002

Posterior probabilities of factors, $\mathbb{E}[\gamma_j|\text{data}]$, and posterior mean of factors' risk prices, $\mathbb{E}[\lambda_j|\text{data}]$, are computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The data is monthly, 1973:10 to 2016:12. The 51 factors considered are described in Table 2.A.1 of the Appendix. Test assets: 34 tradable factors plus 26 investment anomalies.

Table 2.A.19: Posterior model dimensionality and its implied Sharpe ratio

| | (a) Number of factors | | | | | | (b) Model-implied Sharpe ratio | | | | | |
| | Total prior SR: | | | | | | Total prior SR: | | | | | |
	1	1.5	2	2.5	3	3.5	1	1.5	2	2.5	3	3.5
mean	25.14	24.93	24.62	24.12	23.02	17.91	0.85	1.17	1.47	1.80	2.19	2.70
median	25	25	25	24	23	18	0.82	1.15	1.46	1.78	2.18	2.69
2.5%	18	18	18	17	16	11	0.47	0.70	0.93	1.17	1.46	1.75
5%	19	19	19	18	17	12	0.52	0.76	1.00	1.27	1.58	1.89
95th	31	31	31	30	29	24	1.26	1.65	2.00	2.37	2.83	3.57
97.5th	32	32	32	31	30	25	1.36	1.76	2.12	2.49	2.96	3.75

Summary statistics for posterior number of the factors included in the model and the model-implied Sharpe ratio. Both are summarized for values of total prior Sharpe ratio $\in [1, 3.5]$. All the parameters are estimated over the 1973:10-2016:12 sample using a cross-section of 34 tradable factors plus 26 investment anomalies, computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 51 factors yielding $2^{51} \approx 2.25$ quadrillion models. The prior for each factor inclusion is a $Beta(1,1)$, yielding a prior expectation for γ_j equal to 50%. The 51 factors considered are described in Table 2.A.1 of the Appendix.

(a) Model dimensionality with PCs

(b) Model dimensionality with RP-PCs
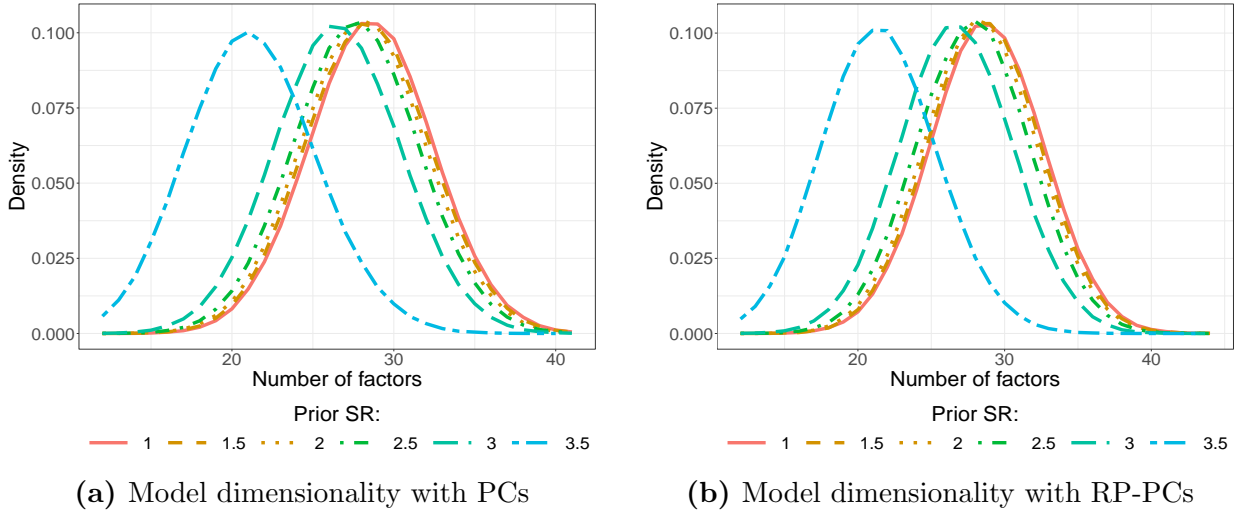
Figure 2.A.2: Model dimensionality with principal components added to the space of factors

Posterior density of the true SDF having the number of factors listed on the horizontal axis computed using the continuous spike-and-slab approach of Section 2.3.1.3 and 59 factors yielding 2^{59} models. The factors included are the 51 factors described in Table 2.A.1 of the Appendix, plus two i.i.d. useless factors, and five principal components. Panel A uses simple time series principal components, while Panel B uses the RP-PCs of Lettau and Pelger (2020b). Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results are tabulated for different values of the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_f^2 \mid \sigma^2]}$.

2.A.5   Additional Results on Sparse Models
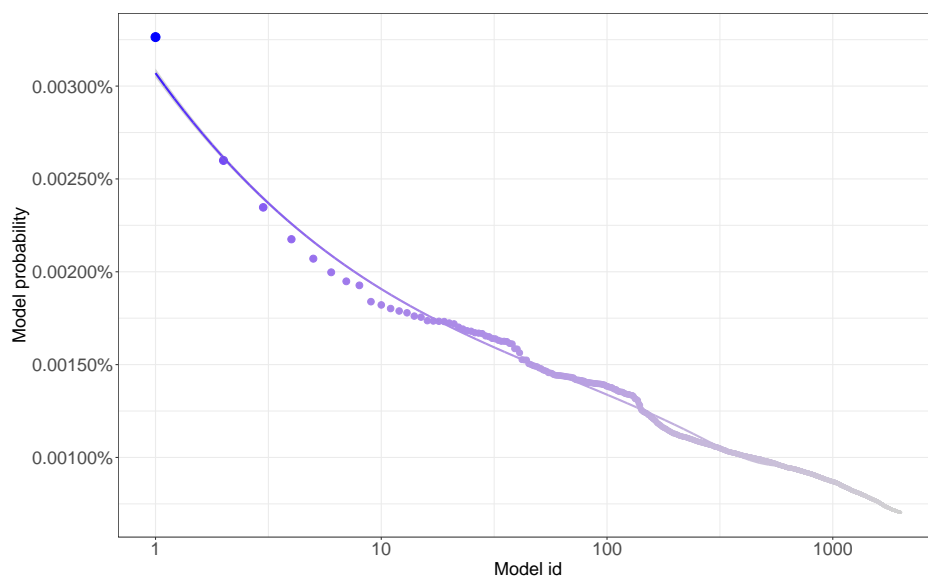
Figure 2.A.3: Posterior model probabilities, 2.6 mln sparse models

Posterior model probabilities of the 2,000 most likely models computed using the Dirac spike-and-slab of Section 2.3.1.2 and 51 factors. The horizontal axis uses a log scale. Sample: 1973:10–2016:12. Test assets: 34 tradable factors plus 26 investment anomalies, sampled monthly, 1973:10 to 2016:12. Results are reported for the prior Sharpe ratio, $\sqrt{\mathbb{E}_\pi[SR_{\boldsymbol{f}}^2 \mid \sigma^2]} = 2$.

Chapter 3

Model Uncertainty in the Cross Section

Jiantao Huang and Ran Shi [1]

3.1 Introduction

Recent literature has provided a wide spectrum of real and financial uncertainty measures.[2] They display pronounced time-series variations, and their innovations appear to be associated with business cycle fluctuations and investment decisions.

Uncertainty has ambiguous implications for investors' asset allocation decisions in equity markets. The conventional wisdom of flight-to-safety and flight-to-liquidity[3] claims that investors respond to uncertainty by curtailing risk exposures. However, uncertainty may arise in periods of "Schumpeterian growth," during which investors chase glamour stocks (which tend to be riskier) in search of the new El Dorado.[4]

Existing equity market uncertainty measures focus on second-moment uncertainty, such as realized/implied volatilities of major index returns and prediction uncertainty of economic indicators (e.g., financial uncertainty in Ludvigson, Ma, and Ng (2021)). These uncertainty

---

[1] We thank Svetlana Bryzgalova, Thummim Cho, Vicente Cuñat, Christian Julliard, Péter Kondor, Dong Lou, Ian Martin, and Cameron Peng for their comments.

[2] Bloom (2009) measures macroeconomic uncertainty using jumps in the VIX index and investigates their real impacts. Ludvigson, Ma, and Ng (2021) 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.

[3] Many theoretical papers study such phenomena, including Vayanos (2004), Caballero and Krishnamurthy (2008), Brunnermeier and Pedersen (2009), etc.

[4] This argument relates to growth options theories of uncertainty. Examples include Abel (1983), Segal, Shaliastovich, and Yaron (2015), Kraft, Schwartz, and Weiss (2018), etc.

measures do not take into account a crucial challenge equity investors face, a phenomenon dubbed the "factor zoo." If we interpret existing uncertainty measures as time-series uncertainty, a vital dimension they neglect is the cross-section.

We attempt to bridge this gap by creating a cross-sectional uncertainty measure and exploring its implications for 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 model-implied 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, i.e., 99%. Under this scenario, entropy is close to its lower bound zero (and we are clearly facing low uncertainty). On the contrary, if the two models' posterior probabilities are 50-50, the entropy reaches its maximum (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 displays considerable time-series variations and exhibits counter-cyclical behaviours, as in Figure 3.1. 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 difference). Under extreme market conditions, investors do not only face higher second-moment (volatility) and third-moment (skewness) risk but they are also confronted with higher (if not the highest) model uncertainty, i.e., they are
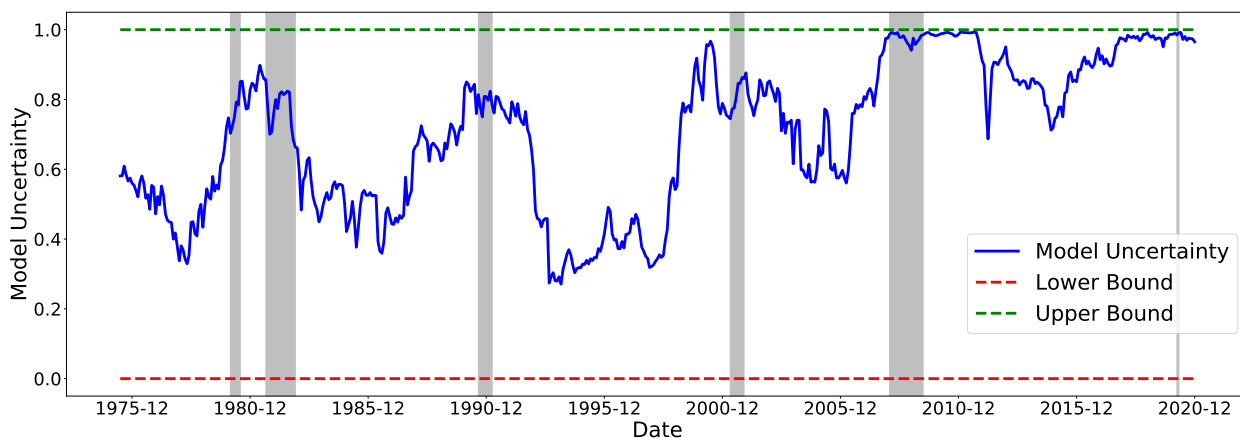
Figure 3.1: Time-Series of Model Uncertainty (3-Year Rolling Window)

The figure plots the time-series of model uncertainty about the linear stochastic discount factor (SDF). We consider 14 prominent factors from the past literature (see Section 3.3 for details). At the end of each month, we compute the posterior model probabilities using the daily factor returns in the past three years. We use the entropy of model probabilities to quantify model uncertainty in the cross-section. The sample ranges from July 1972 to December 2020. Since we use a three-year rolling window, the model uncertainty index starts from June 1975. The red line and green lines show the lower (0) and upper bounds (1) of model uncertainty. Shaded areas are NBER-based recession periods for the US.

incredibly uncertain about which model can help navigate them out of the storm.

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

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

Third, model uncertainty is a crucial determinant of mutual fund flows, regardless of being an exogenous cause or a merely propagating mechanism. We adopt the canonical Vector Autoregression (VAR) model to study the dynamic responses of fund flows to uncertainty

163

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

It is also worth noting that similar fund flows patterns do not emerge when using volatility-driven uncertainty indices such as VXO and financial uncertainty. We document some evidence that VXO and financial uncertainty innovations relate to future inflows to money market funds, consistent with "flight-to-liquidity." However, dynamic responses of fund flows to these two uncertainty measures tend to be transitory and sensitive to identification assumptions, while those to model uncertainty shocks are very persistent and robust.

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

3.1.1   Related Literature

This article mainly relates to two strands of literature. First, 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 (VIX/VXO index) and uses them as proxies for uncertainty shocks. He further shows in a VAR analysis that a positive uncertainty shock predicts declining industrial production, productivity, and employment over the next several years. Jurado, Ludvigson, and Ng (2015) measure macroeconomic uncertainty and 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. Ludvigson, Ma, and Ng (2021) further propose real and financial uncertainty indices. These two papers use the conditional volatility of prediction errors as proxies for uncertainty, so they belong to volatility-based measures. 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. Our regression analysis shows that model uncertainty is positively correlated with financial uncertainty and the VXO index but almost orthogonal to real, macro, and economic policy uncertainty mentioned above. Moreover, we detect persistent dynamic responses of mutual fund flows following model uncertainty shocks. In contrast, traditional volatility-based measures such as VXO and financial uncertainty do not have similar implications for the portfolio choice decisions of mutual fund investors. Therefore our measure is conceptually and empirically novel. Our entropy-based measure also 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).

Second, our paper 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 the size related to g. Similar ideas of specifying priors are adopted in past finance and economics 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 paper to adopt this prior in finance literature.

Some other papers, such as Barillas and Shanken (2018a), Chib, Zeng, and Zhao (2020), and Bryzgalova, Huang, and Julliard (2021), also develop Bayesian methods to estimate factor models. Unlike their papers, we aim to propose a direct measure of model uncertainty and investigate its implications for portfolio choice. Although past literature has introduced model uncertainty under the portfolio choice specification (e.g., Avramov (2002), Barillas and Shanken (2018b)), we have different motivations in the first place.

3.2 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[5]. 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

---

[5]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.

denote by $\boldsymbol{f}$, a subset of $\boldsymbol{R}$ with dimension p, the factors under consideration[6]. A linear factor model for these excess returns in the discount factor form can be written as (see Chapter 13 of Cochrane (2009) for a detailed exposition):

$$m = 1 - (\boldsymbol{f} - \mathbb{E}[\boldsymbol{f}])^\top \boldsymbol{b}, \tag{3.1}$$

$$\mathbb{E}[\boldsymbol{R} \times m] = \boldsymbol{0}, \tag{3.2}$$

or equivalently

$$\mathbb{E}[\boldsymbol{R}] = \mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}]\boldsymbol{b}, \tag{3.3}$$

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 (3.2) 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.[7] 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 (3.3).

3.2.1 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

---

[6]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.

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

model for the SDF[8], denoted by $\mathcal{M}_\gamma$, as follows: Under $\mathcal{M}_\gamma$, the linear SDF is

$$m_\gamma = 1 - (\boldsymbol{f}_\gamma - \mathbb{E}[\boldsymbol{f}_\gamma])^\top \boldsymbol{b}_\gamma, \tag{3.4}$$

and the expected excess returns are such that

$$\mathbb{E}[\boldsymbol{R}] = \text{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma] \boldsymbol{b}_\gamma, \tag{3.5}$$

where $\boldsymbol{f}_\gamma$ is a p_γ-dimensional vector that contains all the factors included under the current model;[9] $\boldsymbol{b}_\gamma$ is a p_γ-dimensional vector of nonzero factor loading; $\text{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma]$ is now an $N \times p_\gamma$ covariance matrix. The two equations above are counterparts of (3.1) and (3.3) 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 $\mathcal{M}_\gamma$ does so for the distribution of all excess returns $\boldsymbol{R}$ (conditional on vector $\boldsymbol{\gamma}$), according to equation (3.5). The expectations of this random vector $\boldsymbol{R}$ are linked to a block of its variance-covariance matrix, namely $\text{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma]$[10], through a vector of coefficients $\boldsymbol{b}_\gamma$.

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 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 (3.5). What we would like to explore is 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

---

[8]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 $\boldsymbol{f}_\gamma$ and $\boldsymbol{f}_{-\gamma}$, and the intercept of elements in $\boldsymbol{f}_\gamma$ and $\boldsymbol{f}_{-\gamma}$ is empty.

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

[10]Recall that under our setting, factors are a predetermined subset of excess returns, that is, $\boldsymbol{f}_\gamma \subseteq \boldsymbol{f} \subseteq \boldsymbol{R}$. Thus $\text{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma]$ is a sub-block of the full $N \times N$ variance-covariance matrix $\text{Var}[\boldsymbol{R}]$. It is in fact an $N \times p_\gamma$ matrix consisting of p_γ columns of $\text{Var}[\boldsymbol{R}]$. 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 $\boldsymbol{b}$.

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 $\boldsymbol{b}_\gamma$ is related to market prices of risk, because from equation (3.5)

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

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

Noticing the link between $\boldsymbol{b}_\gamma$ and $\boldsymbol{\lambda}_\gamma$, one may consider introducing the latent variable $\boldsymbol{\gamma}$ for the risk premia 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 $\boldsymbol{f}^\top = [\boldsymbol{f}_\gamma^\top, \boldsymbol{f}_{-\gamma}^\top]$, then from equation (3.5),

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

where $\boldsymbol{B} = \{\text{Var}[\boldsymbol{f}]\}^{-1}\text{Cov}[\boldsymbol{f}, \boldsymbol{R}]$. As a result, if we let $\boldsymbol{\lambda}^\top = \left[\boldsymbol{\lambda}_\gamma^\top, \boldsymbol{b}_\gamma^\top\text{Cov}[\boldsymbol{f}_\gamma, \boldsymbol{f}_{-\gamma}]\right]$, it is always the case that $\mathbb{E}[\boldsymbol{R}] = \boldsymbol{B}^\top \boldsymbol{\lambda}$ regardless of which model $\mathcal{M}_\gamma$ 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.

3.2.2 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 $\mathcal{M}_\gamma$ given the observed data $p[\mathcal{M}_\gamma \mid \mathcal{D}]$. 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:

$$\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \stackrel{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{3.6}$$

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

$$\boldsymbol{\mu} = \boldsymbol{C}_\gamma \boldsymbol{b}_\gamma, \tag{3.7}$$

where $\boldsymbol{C}_\gamma = \text{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma]$ 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 $p[\mathcal{M}_\gamma \mid \mathcal{D}]$, and then substituting this matrix with a moment estimator.[11]

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 $\mathcal{M}_\gamma$,

$$\boldsymbol{b}_\gamma \mid \mathcal{M}_\gamma \sim \mathcal{N}\left(\boldsymbol{0}, \frac{g}{T}\left(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma\right)^{-1}\right), \quad g > 0 \tag{3.8}$$

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

---

[11]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 $\boldsymbol{b}_\gamma$, as will become clear next when we introduce our prior specification. Thus we are still using data to pin down (hyper)parameters in the priors. The use of moment estimators to replace parameters in the prior distributions dates all the way back to the seminal James-Stein estimator (James and Stein (1961)). For a monograph on modern empirical Bayes methods, see Efron (2012).

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.[12] 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,[13] 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" in 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 $\boldsymbol{b}_\gamma^a$, in addition to the imaginary sample $\mathcal{D}'$ (as well as the initial diffuse prior for $\boldsymbol{b}_\gamma$). The rational expectation hypothesis says that $\boldsymbol{b}_\gamma^a = \mathbb{E}[\boldsymbol{b}_\gamma \mid \mathcal{M}_\gamma, \mathcal{D}'] = \boldsymbol{b}_\gamma'$. Now we have a reference informative prior distribution that does not depend on the hypothetical sample, which is

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

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

---

[12]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).

[13]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.

gives us the prior specification in (3.8).

   Remark. One might attempt to assign an objective prior, such as the Jeffreys prior, to $\boldsymbol{b}_\gamma$. 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 $\boldsymbol{b}_\gamma$. 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 (3.8) leads to a surprisingly simple expression for the variance of the SDF, which is summarized in Proposition 3.1.

Proposition 3.1 *Under model* $\mathcal{M}_\gamma$*, in which* $m_\gamma = 1 - (\boldsymbol{f}_\gamma - \mathbb{E}[\boldsymbol{f}_\gamma])^\top \boldsymbol{b}_\gamma$*, the g-prior specification for* $\boldsymbol{b}_\gamma$ *implies that*

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

According to Proposition 3.1, volatility of the SDF ($= \sqrt{g p_\gamma / T}$) 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 $\boldsymbol{b}_\gamma$ 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 3.2.

Proposition 3.2 *The marginal likelihood of observing excess return data* $\mathcal{D}$ *under model* $\mathcal{M}_\gamma$ *is*

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

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; $\mathrm{SR}_{\max}^2$ *is the maximal squared Sharpe ratio achievable from forming portfolios using all excess returns under consideration;* SR_γ^2 *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 3.2 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 3.2, 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 that compares model $\mathcal{M}_{\boldsymbol{\gamma}}$ with the null model $\mathcal{M}_{\mathbf{0}}$ is defined as the ratio between marginal likelihoods under two different models; that is,

$$
\begin{aligned}
\mathrm{BF}_{\boldsymbol{\gamma}}(g) &= \frac{p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]}{p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]} \\
&= \exp\left\{\frac{Tg}{2(1+g)}\mathrm{SR}^2_{\boldsymbol{\gamma}} - \frac{p_{\boldsymbol{\gamma}}}{2}\log(1+g)\right\}.
\end{aligned}
\tag{3.9}
$$

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

$$
\begin{aligned}
\mathrm{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(g) &= \frac{\mathrm{BF}_{\boldsymbol{\gamma}}(g)}{\mathrm{BF}_{\boldsymbol{\gamma}'}(g)} \\
&= \exp\left\{\frac{Tg}{2(1+g)}\left(\mathrm{SR}^2_{\boldsymbol{\gamma}} - \mathrm{SR}^2_{\boldsymbol{\gamma}'}\right) - \frac{p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}'}}{2}\log(1+g)\right\},
\end{aligned}
\tag{3.10}
$$

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 (3.10) is that although the marginal likelihood in Proposition 3.2 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 model-implied Sharpe ratios $(SR_{\boldsymbol{\gamma}})$ 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, 2009, 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 SR_γ increases as additional assets are added into the factor portfolio. Thus the penalty term on model dimensionality p_γ imposed by the g-prior plays an key role in preventing the Bayes factor to favor large models blindly. In order to properly penalize large models, g cannot be too small, as SR_γ always increases after one augments the linear SDF.

Perhaps the most desirable feature of our Bayes factor calculation in equation (3.10) is that it helps us understand the aforementioned trade-off quantitatively. When model dimension is increased by one ($p_\gamma - p_{\gamma'} = 1$), the maximal squared Sharpe ratio (times the sample size T) of the factor portfolio has to increase by at least $(1 + g)/g \times \log(1 + g)$ to lend support to the augmented model, that is,

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

However, it is always the case that $T\left(\text{SR}_\gamma^2 - \text{SR}_{\gamma'}^2\right) \leq T\text{SR}_{\max}^2$. 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 (3.9), $\text{BF}_\gamma(g) \to 0$. Paradoxically, the most favorable model will always be the null model. The case under which $g \to \infty$ corresponds to the conventional diffuse priors; and the fact that, with model uncertainty, diffuse priors always support the null model is sometimes called the Bartlett's paradox (Bartlett (1957)). Of note, this paradox poses another refutation to the use of improper diffuse priors under model uncertainty, in addition to posterior indeterminacy that has been pointed out earlier.

3.2.3 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 $\boldsymbol{b}_\gamma$. 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

$$\pi[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}] \propto \int_0^\infty \mathcal{N}\left(\boldsymbol{b}_{\boldsymbol{\gamma}} \;\middle|\; \boldsymbol{0},\, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right) \pi[g] \, \mathrm{d}g, \tag{3.11}$$

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 according to Liang, Paulo, Molina, Clyde, and Berger (2008). Because the marginal likelihood of the null model does not depend on g (recall that $p[\mathcal{D} \mid \mathcal{M_0}, g] \equiv p[\mathcal{D} \mid \mathcal{M_0}]$), improper priors will lead to indeterminacy in the ratio

$$\begin{aligned} \mathrm{BF}_{\boldsymbol{\gamma}} &= \frac{\int_0^\infty p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]\pi[g]\,\mathrm{d}g}{\int_0^\infty p[\mathcal{D} \mid \mathcal{M_0}]\pi[g]\,\mathrm{d}g} \\ &= \int_0^\infty \frac{p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]}{p[\mathcal{D} \mid \mathcal{M_0}]}\pi[g]\,\mathrm{d}g \end{aligned} \tag{3.12}$$

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 3.1, 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[14], denoted by SR_∞, then

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

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

---

[14]Note that this must be larger than the maximal in-sample Sharpe ratio of portfolios formed using excess returns under our consideration, denoted by $\mathrm{SR}_{\mathrm{max}}$ in Proposition 3.2.

world). As a result, we will require that

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

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

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

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 $\boldsymbol{\mu} = \boldsymbol{C}_\gamma \boldsymbol{b}_\gamma$, which is given as follows

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

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

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

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 $a = 4 + 2/(TR_f^2 \mathrm{SR}_\infty^2)$. Under this choice, the prior expectation for the shrinkage factor is still strictly smaller than one half, but can be very close (the ratio $2/(TR_f^2 \mathrm{SR}_\infty^2)$ is usually very small).

3.2.4   Posterior Probability of Models

We next integrate out the parameter g according to equation (3.12) to find the Bayes factors under the mixture of g-priors. Proposition 3.3 presents the results.

Proposition 3.3 *The Bayes factor for comparing model $\mathcal{M}_\gamma$ with the null model $\mathcal{M}_0$ is*

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

where

$$\underline{\Gamma}(s, x) = \int_0^x t^{s-1} e^{-t} \, \mathrm{d}t$$

is the lower incomplete Gamma function (Abramowitz and Stegun, 1965, Page 263); the scalar s_γ is defined as

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

This Bayes factor is always increasing in SR_γ^2 always decreasing in p_γ.

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

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

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

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

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

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

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:

$$\mathcal{E}[\mathcal{M}_\gamma \mid \mathcal{D}] = \sum_\gamma \log(p[\mathcal{M}_\gamma \mid \mathcal{D}]) p[\mathcal{M}_\gamma \mid \mathcal{D}]. \tag{3.13}$$

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, model uncertainty in equation (3.13) is maximized.

3.3 Data Description

In our primary empirical implementation, we combine 14 prominent factors from the past literature and measure model uncertainty in this small zoo of factors. First, we include notable Fama-French five factors (Fama and French (2015)) plus the momentum factor (Jegadeesh and Titman (1993)). In addition, we consider the q-factor model from Hou, Xue, and Zhang (2015) and include their size, investment, and profitability factors. The factor models mentioned earlier are based on rational asset pricing theory. Taking the insights from behavioural models, Daniel, Hirshleifer, and Sun (2020) propose a three-factor model consisting of the market factor, the short-term behavioural factor (PEAD), and the long-term behavioural factor (FIN). Finally, we include the HML devil, the quality-minus-junk factor, and the betting-against-beta factor from the AQR library. Appendix 3.A.1 presents the detailed description of these factors.

Table 3.A.1 reports the annualised mean returns and Sharpe ratios of 14 factors. First, most of them (except for two size factors) have enormous Sharpe ratios in the full sample from July 1972 to December 2020. In particular, the short-term behavioural factor (PEAD) seems to be the most profitable historically. Furthermore, I split the entire sample into two equal subsamples. Consistent with past literature (e.g., McLean and Pontiff (2016)), the performance of many factor strategies decline significantly from subsample one to two. Most strikingly, the annualised Sharpe ratio of the value factors has plunged from above 0.9 to nearly zero in the second subsample. This observation suggests that we should focus on the out-of-sample instead of the in-sample Sharpe ratio in evaluating factor models.

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

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

monthly mutual fund flows, so we download the monthly total net assets, monthly fund returns, and the codes of fund investment objectives. To normalise the aggregate fund flows, we divide the equity (fixed-income) fund flows across all funds within a particular investment objective by the total market capitalisation of all listed companies in CRSP (2021) (US GDP). In addition, we download the total market value of all US-listed stocks from CRSP.

Finally, we study the relationship between our model uncertainty measure and investors' expectations about future stock market performance. 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.

3.4 Measuring Model Uncertainty

We now adopt the perspective of Bayesian investors and construct the time series of model uncertainty. At the end of each month, we use all daily factor returns in the past three years to estimate the posterior model probabilities, $p[\mathcal{M}_\gamma \mid \mathcal{D}]$, and compute the entropy as in equation (3.13). We choose the hyper-parameter a to be four in the benchmark case. We also present the results obtained from alternative rolling windows and other choices of a in robustness checks (see Section 3.8).

The behavioural factors in Daniel, Hirshleifer, and Sun (2020) are available only from July 1972, and we use 36-month data in the estimation, so the model uncertainty measure starts from June 1975. Since some factors are highly correlated, we 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 ten effective factors, including market, size, profitability, value, investment, short-term and long-term behavioural factors, momentum, QMJ, and BAB.

The blue line in Figure 3.1 plots the time series of model uncertainty of linear SDFs, and the sample period spans from June 1975 to December 2020. The red and green dotted lines show the lower and upper bounds of model uncertainty, respectively. The lower entropy bound is always zero, i.e., when there is one dominant model with the posterior model probability of 100%. On the contrary, uncertainty is maximized when the posterior model probabilities are equalized across all models. Because we have 14 factors, and only one of the categorical factors could be selected into the true model, there are 5,184 different candidate

models.[15] The upper bound of model uncertainty is around 8.55.[16] To normalize the model uncertainty index, we divide it by 8.55. Hence, the upper bound is one in Figure 3.1.

The model uncertainty index has several interesting features that could shed light on the nature of uncertainty about the linear SDF. First, we observe a surprisingly high level of model uncertainty. Specifically, the average (median) model uncertainty is around 0.70 (0.75), with the first and third quartiles equal to 0.53 and 0.87, respectively. Hence, most of the time, Bayesian investors are not confident about the true SDF model. Second, model uncertainty fluctuates significantly over time. In particular, the index varies from the lowest value of 0.27 to the highest 0.99, representing economic states in which Bayesian investors find it almost unlikely to determine the true SDF model. The standard deviation of the index is 0.21. Overall, model uncertainty is a dynamic phenomenon. Finally, model uncertainty is persistent by construction since we use a rolling window of 36 months in the estimation. The first-order autocorrelation is 0.98, and the autocorrelation coefficients strictly decrease in time lags, with insignificant autocorrelations after 30 lags.

Figure 3.1 also suggests the countercyclical nature of model uncertainty. In particular, the 1990s was a remarkable period: it was remembered as a period of strong economic growth, low inflation and unemployment rate, and high stock returns. During the 1990s, model uncertainty is the lowest across our sample. As the orange dots in Figure 3.2 suggest, posterior probabilities of the top two models are significantly larger than others. Hence, investors are relatively confident about the true SDF model.

In addition, peaks in model uncertainty tend to coincide with major events in the US stock markets and economy. Important examples include the dot-com crash in 2000 and the global financial crisis in 2008 when model uncertainty almost touches its upper bound. Specifically, the blue dots in Figure 3.2 show that posterior probabilities of the top 50 models, in December 2007, are almost equalized. In other words, it is virtually infeasible to distinguish models based on the observed data. The 2008 crisis is noteworthy because model uncertainty stays at a high level for a prolonged period. In contrast, it declines shortly after other crises/recessions. In the recent five years, model uncertainty has slowly increased from 0.7 to 1 at the end of 2020.

Interestingly, we do not observe a spike in model uncertainty during the 1987 flash crash. 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

---

[15]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 $2^6 \times 3^4 = 5184$.

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

recession, and the market recovered rapidly. Instead, the leading cause was synchronous program trading, illiquidity in the market, and the subsequent market panic. Since our uncertainty measure is based on past-three-year daily data, the impact of short-term market chaos is averaged out.

In conclusion, our model uncertainty measure displays considerable time-series variations: it is particularly sizable in bad economic states. The stock market crash that lasts only for a short period, such as the 1987 flash crash, is not captured by our model uncertainty measure. Furthermore, the cyclical behaviours of model uncertainty imply another layer of investment risk: when investors experience bear stock markets, they are also the most uncertain about the true model in the cross-section, or equivalently, which portfolio of factor strategies they should hold. This further motivates us to study how model uncertainty relates to investors' portfolio choices and expectations. We investigate these topics in section 3.5 and 3.6.
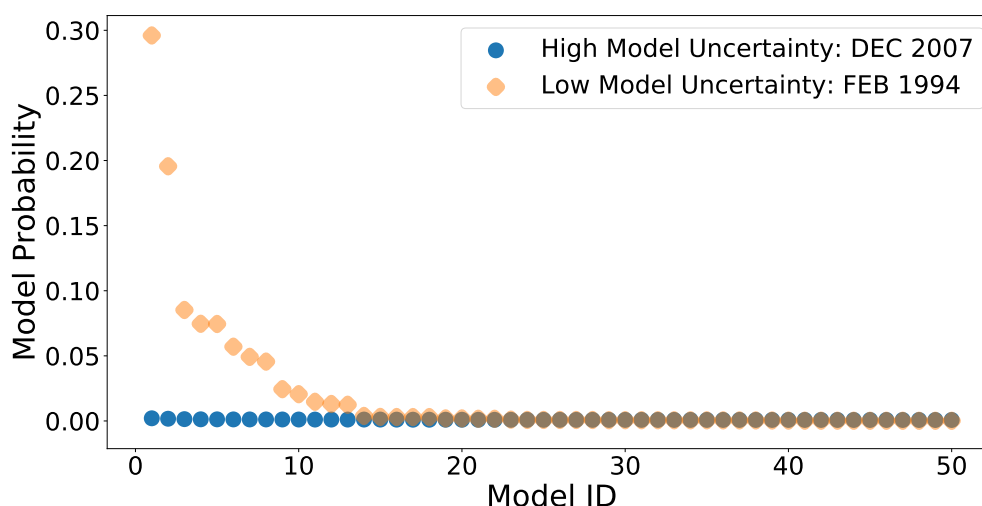
Figure 3.2: Posterior Probabilities of Top 50 models: High vs. Low Model Uncertainty

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

3.4.1 Does Model Uncertainty Matter?

Should investors take into account model uncertainty in the cross-section? A natural hypothesis is that model uncertainty plays a more critical role when it is more sizable. The logic is as follow. When model uncertainty is relatively low, the factor model with the highest model

probability dominates others, such as the orange diamonds in Figure 3.2. Hence, investors are more willing to trust the top model ranked by the Bayesian posterior probabilities. In contrast, the top model is not informative if model uncertainty is relatively high, such as during market crashes. In this case, they may prefer to aggregate the information over the space of all models.

The Bayesian model averaging (BMA) is one common approach to aggregating models. It enables us to flexibly model investors' uncertainty about potentially relevant factors. In the SDF model, we are interested in the risk prices, $\boldsymbol{b}$. The BMA of $\boldsymbol{b}$ is defined as

$$\boldsymbol{b}_{bma} := \mathbb{E}[\boldsymbol{b} \mid \mathcal{D}] = \sum_{\boldsymbol{\gamma}} \mathbb{E}[\boldsymbol{b} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}] \times P(\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}). \tag{3.14}$$

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

BMA deviates sharply from the traditional model selection, in which researchers always use a particular criterion (e.g., adjusted R2, model probabilities, etc.) to select a single model and presume that the selected model is correct. Past literature also shows the importance of model averaging in asset pricing (e.g., Avramov (2002), Bryzgalova, Huang, and Julliard (2021), Avramov, Cheng, Metzker, and Voigt (2021)).

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

For each factor model $\boldsymbol{\gamma}$ in month t, we estimate the risk prices of $\boldsymbol{f}_{\boldsymbol{\gamma}}$ via the standard GMM estimation: $\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}} = (\text{Var}[\boldsymbol{f}_{\boldsymbol{\gamma}}])^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{\gamma}t})$, where the covariance matrix and mean returns of $\boldsymbol{f}_{\boldsymbol{\gamma}}$ are estimated using the data from month $t - 35$ to month t, consistent with Figure 3.1. The tangency portfolio conditional on model $\boldsymbol{\gamma}$ is $\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{f}_{\boldsymbol{\gamma},t+1}$, and the BMA tangency portfolio is $\boldsymbol{b}_{bma}^{\top}\boldsymbol{f}_{t+1}$.[17] We update the tangency portfolio each month.[18]

We also test the null hypothesis that BMA and the model $\boldsymbol{\gamma}$ have an identical Sharpe ratio, i.e., $H_0 : SR_{bma}^2 = SR_{\boldsymbol{\gamma}}^2$, using the non-parametric Bootstrap. Under H_0, the expected return of the tangency portfolio implied by the model $\boldsymbol{\gamma}$ is linear in that of BMA: $\mathbb{E}[R_t^{\boldsymbol{\gamma}}] = \mathbb{E}[R_t^{bma}]\sigma(R_t^{\boldsymbol{\gamma}})/\sigma(R_t^{bma})$. We adjust the average return of $R_t^{\boldsymbol{\gamma}}$ using the previous equality and

---

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

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

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

Table 3.1: Out-of-Sample Model Performance

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

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

We start with describing the full-sample performance, as shown in the first row of Table 3.1. First, our Bayesian procedure successfully selects the model that outperforms traditional factor models in the out-of-sample. The top Bayesian model (see column (2)) has an OOS Sharpe ratio of 1.75, which is virtually comparable to the model composed of all 14 factors (see column (3)). Second, BMA beats the top Bayesian model. The outperformance is statistically significant, but its economic magnitude is not substantial.

One may be concerned that these 14 factors are data-mined, so choosing the top model only reflects data snooping rather than the outperformance of our Bayesian procedure. We

---

[19]In other words, we calculate the approximate achieved significance level, ASL_{boot}, by

$$ASL_{boot} = \frac{\sum_{n=1}^{B} \mathbf{1}_{\{SR(R^{bma}_{t^{\star}}) - SR(R^{\gamma}_{t^{\star}}) \geq SR(R^{bma}_{t}) - SR(R^{\gamma}_{t})\}}}{B}$$

where B is the number of Bootstraps ($B = 100,000$). If ASL_{boot} is smaller than 10% (5%, 1%), we claim that H_0 is rejected by the data at 10% (5%, 1%) significance level.

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

Whether the performance of factor models is related to model uncertainty? The short answer is yes. On average, the performance of factor models declines as model uncertainty increases. Specifically, when model uncertainty is low, both the top model and BMA have similar Sharpe ratios of around 2.57, which are exceptionally high. In other words, investors should be confident about the top model chosen by our Bayesian procedure in low uncertainty states. On the contrary, it is particularly beneficial to incorporate model uncertainty into portfolio choice when model uncertainty is high. As the last row suggests, BMA has an OOS Sharpe ratio of 1.25, significantly larger than any other specifications.

In summary, there are two takeaways from Table 3.1. First, our Bayesian procedure is competent to pick the model that has satisfactory OOS performance. Second, model uncertainty matters and is particularly noteworthy when it is relatively high. In this scenario, BMA, which aggregates the information across all models, is salient for real-time portfolio choice.

3.4.2 Decomposing Model Uncertainty

The posterior model probabilities (see Proposition 3.3) are closely related to the model-implied squared Sharpe ratio, SR_γ^2. As we include more factors, the in-sample SR_γ^2 always rises. Only when a few factor models dominate others can we be confident about the true model. In other words, when the distances in SR_γ^2 are sizable across different factor models, we can easily differentiate them and observe low model uncertainty. In contrast, when factor models have similar SR_γ^2, model uncertainty tends to be high.

Figure 3.3 plots the time-series of distances in SR_γ^2. More precisely, we show the difference between the maximal SR_γ^2 and the $90th$-quantile of SR_γ^2, as well as the difference between the maximal SR_γ^2 and medium SR_γ^2. Strikingly, the difference in SR_γ^2 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_γ^2 is close to 0.2 (daily) between 1997 and 1998, but it plunges to almost 0 from 1998 to 2000. After the tech bubble, factor models have been becoming more similar in terms of in-sample SR_γ^2.

Theoretically, SR_γ^2 is determined by mean returns of factors and their covariance matrix. We further analyze SR_γ^2 by dipping into three parts: (a) average daily factors returns in
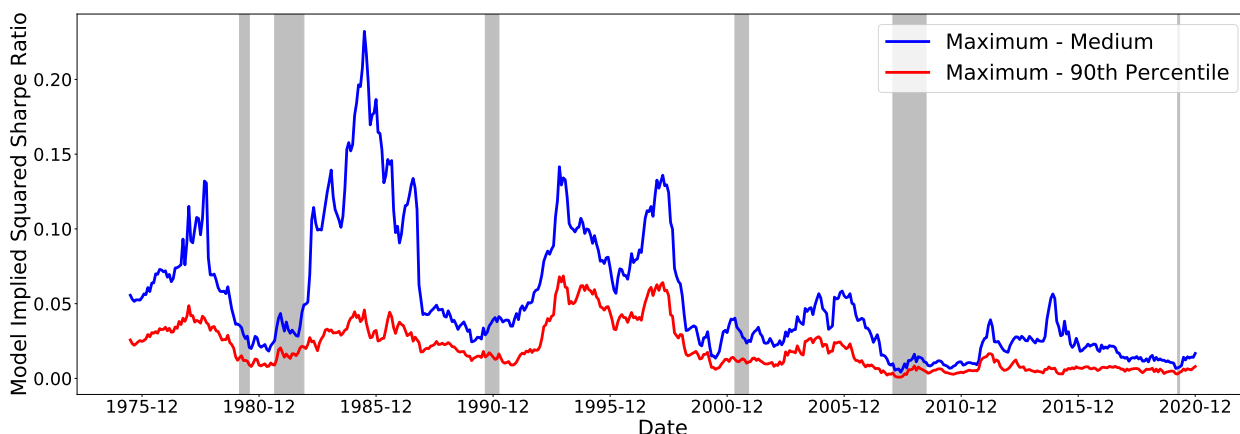
Figure 3.3: Time-Series of Model-Implied Squared Sharpe Ratio (3-Year Rolling Window)

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

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 3.4 plots these time series.

In Figure 3.4a, we show that the average daily return of all 14 factors is incredibly volatile. The average daily return also exhibits cyclical patterns. Specifically, it declines during the run-ups of stock markets. However, it plummets to the bottom during the market crash and recovers gradually after the bear markets. In the recent three most influential market crashes (dot-com bubble, 2008 global financial crisis, and the Covid-19), the average factor returns decline to near zeros. In the past decade, the profitability of these 14 factors is no longer comparable to their historical performance. One potential reason is that more investors implement the same investment strategies after the publication of these factors (see McLean and Pontiff (2016)).

Figure 3.4b plots the average volatility of 14 factors. Even though the average factor volatility increases in the bear markets, the factor returns before the dot-com bubble are not as volatile as after 2000. Typically, the average standard deviation of 14 factors is between 0.2% and 0.4%. During the dot-com bubble and recent global financial crisis, it surges to higher than 1% daily. However, it is evident from figure 3.4b that model uncertainty does not have the same time-series pattern as the average factor volatility.

During market crashes, it is highly likely that arbitrageurs who invest in these factor strategies will exit the market simultaneously, thus driving up comovements among factors.
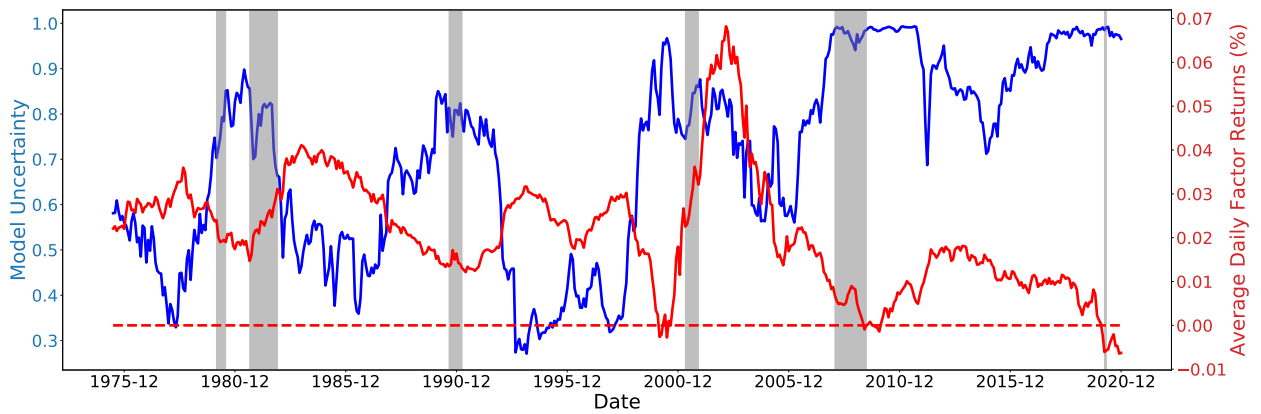
(a) Time-Series of Average (Daily) Return of 14 Factors
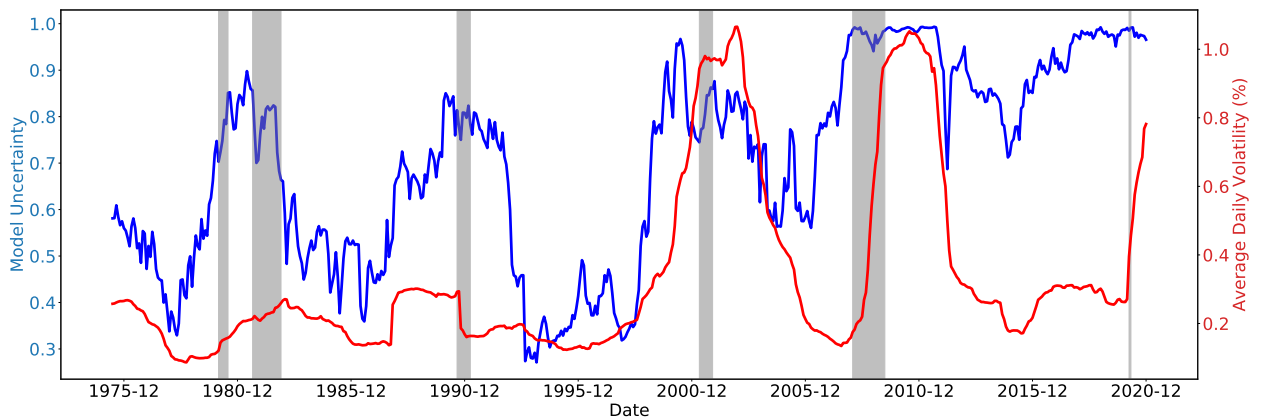
(b) Time-Series of Average (Daily) Volatility of 14 Factors
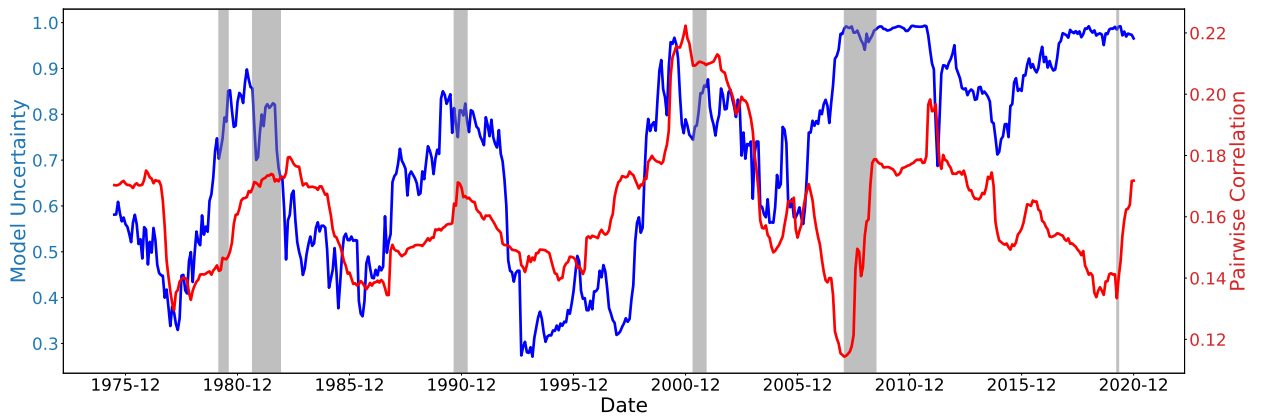
(c) Time-Series of Average Pairwise Correlation of 14 Factors

Figure 3.4: Decomposing the Model Uncertainty

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

Since the correlation matrix of factors determines the extent to which investors can diversify their investment, it could potentially influence the distances in SR_γ^2. To illustrate this point, we plot the time series of the average pairwise correlation of 14 factors.[20] The average correlation exhibits a similar cyclical pattern as model uncertainty. However, there are two key differences: (a) the average correlation decreases before the 2008 crisis while our model uncertainty starts to climb up from 2006, and (b) model uncertainty increases from 2015 to 2019, while the average correlation among factors declines during the same period.

To sum up, model uncertainty is high when the distances in SR_γ^2 among different factor models are low. Since the in-sample SR_γ^2 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, model uncertainty about linear SDFs increases dramatically during the run-ups and stands at the peak during bear markets because different factor models are highly analogous.

3.4.3 Correlation with Other Economic Variables

Figure 3.1 indicates that model uncertainty increases during times of extreme uncertainty in the financial markets and economy. A natural question is how our model uncertainty index correlates with a number of key financial and macroeconomic variables known as capturing critical financial and economic fluctuations.

There are several notable uncertainty measures in the literature. The first measure is VXO/VIX index[21] (used in Bloom (2009)), which quantifies forward-looking market volatility. Subsequent to Bloom (2009), Ludvigson, Ma, and Ng (2021) and Jurado, Ludvigson, and Ng (2015) develop the real, macro and financial uncertainty measures by exploiting a large set of macro and financial variables.[22] Baker, Bloom, and Davis (2016) use the coverage of economic or policy-related keywords in the media as proxies for economic policy uncertainty.

In addition to uncertainty measures, we compare model uncertainty with the intermediary factor from He, Kelly, and Manela (2017), the term yield spread (the yield on ten-year

---

[20]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 $\frac{1}{N \times (N-1)} \sum_{i \neq j} |\rho_{ij}|$.

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

[22]They quantify the h-period ahead uncertainty by the extent to which a particular set of economic variables (either real, macro, or financial) become more or less predictable from the perspective of economic agents. Suppose there is a set of economic indicators, $\boldsymbol{Y_t} = (y_{1t}, ..., y_{Lt})^\mathsf{T}$. 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. Our paper considers their one-period ahead uncertainty measures.

government bonds minus the yield on three-month treasury bills), and the credit spread (the yield on BAA corporate bonds minus the yield on AAA corporate bonds).

We report in Table 3.2 the results from the regression of model uncertainty on its one-period lag and some contemporaneous economic variables. By running these regressions, we do not intend to study the causal relationship between model uncertainty and other economic variables. Instead, our objective is to describe the contemporaneous relation between them. We also want to point out that model uncertainty is persistent[23] since it is constructed in a rolling window of 36 months. Therefore, we need to be careful in statistical inference. In all following tables, we use Newey-West standard errors (see Newey and West (1987)) with 36 lags in the regressions involving model uncertainty.

As Table 3.2 shows, a number of economic variables are significantly related to model uncertainty, even after we control one-period lagged entropy in the regressions. For example, model uncertainty is positively correlated with financial uncertainty and the VXO index but almost orthogonal to real, macro, and two economic policy uncertainty measures. This finding is intuitive since model uncertainty mainly refines information in financial markets. In addition, the intermediary factor and term yield spread negatively relate to model uncertainty. In column (10), we run horse racing among the VXO index, the intermediary factor, and term yield spread: While the coefficient estimates of the VXO index and term yield spread still remain significant, the intermediary factor becomes inconsequential.

Comments. Conceptually, our model uncertainty index quantifies a different layer of uncertainty from other measures. The stock market volatility, proxied by the VXO index, measures the second-moment investment risk. Three uncertainty measures in Ludvigson, Ma, and Ng (2021) and Jurado, Ludvigson, and Ng (2015) are essentially volatilities of prediction errors. In other words, they measure the dispersion of unexpected changes in economic indicators. Two economic policy uncertainty indices in Baker, Bloom, and Davis (2016) are to quantify public attention to economic policy. In contrast, our paper quantifies model uncertainty about linear SDFs. Since we know the lower and upper bounds of entropy, we can easily detect the degree of model uncertainty in the cross-section. For example, model uncertainty reaches its upper bound in some periods, implying that different models' posterior probabilities are almost identical. In short, our model uncertainty index is complementary to other uncertainty measures developed in the past literature. More importantly, ours provides a new angle of analyzing and understanding investment uncertainty.

---

[23]Strong persistence of the time-series process is ubiquitous in other uncertainty measures. Table 3.A.2 shows the AR(1) coefficients of the other six uncertainty sequences, and we find that the real, macro and financial uncertainty measures also have AR(1) coefficients less than but close to 1. It is well-known that the volatility of asset returns tends to cluster. When we run the AR(1) for the VXO index, the coefficient estimate of ρ is 0.812. Only the second economic policy uncertainty measure (EPU_2) suffers less from massive autocorrelations.

Table 3.2: Regressions of Model Uncertainty on Contemporaneous Variables

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

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

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

X_{t+1} include a) financial, macro, and real uncertainty measures from Ludvigson, Ma, and Ng (2021) and Jurado, Ludvigson, and Ng (2015) in columns (1) - (3), b) two economic policy uncertainty (EPU) indices from Baker, Bloom, and Davis (2016) in columns (4) and (5), c) VXO index in column (6), d) the intermediary factor from He, Kelly, and Manela (2017) in column (7), e) term spread in column (8), f) default spread in column (9), and g) VXO index, the intermediary factor, and the term spread in column (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.

3.5   Mutual Fund Flows

If investors consider model uncertainty a crucial source of investment risk, a natural prediction is that their portfolio choice decisions are related to our model uncertainty measure. The difficulty in empirical tests arises due to the lack of observations in their complete portfolio choice. To tackle this issue, we rely on mutual fund flows, which have been studied extensively by the past literature due to their availability. Also, the mutual fund sector is one of the largest financial intermediaries through which individual investors participate in the US stock markets. Hence, we use mutual fund flows as proxies for investors' portfolio rebalancing and study how mutual fund investors react to model uncertainty shocks.

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.[24] The CRSP style code consists of up to four letters. For example, a fund with the style "EDYG" means that i) this fund mainly invests in domestic equity markets (E = Equity, D = Domestic), and ii) it has a specific investment style "Growth" (Y = Style, G = Growth).[25] The quality of data before 1991 is low because the CRSP investment objective code is incomplete. For example, only domestic equity "style" funds and mixed fixed income and equity funds are recorded before 1991. Also, the market values of institutional holdings proportional to the total market value of all stocks (in CRSP) were tiny. Therefore, we focus on the sample from January 1991 to December 2020.

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

$$Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + RET_{i,t}) \tag{3.15}$$

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

$$Flows_t^Y = \frac{\sum_{i \in Y} Flow_{i,t}}{\text{CRSP-Market-Cap}_{t-1}}, \tag{3.16}$$

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

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

$$\boldsymbol{Y_t} = \boldsymbol{B_0} + \boldsymbol{B_1}\boldsymbol{Y_{t-1}} + \cdots + \boldsymbol{B_l}\boldsymbol{Y_{t-l}} + \boldsymbol{u_t}, \tag{3.17}$$

where l denotes the lag order, $\boldsymbol{Y_t}$ is a $k \times 1$ vector of economic variables, $\boldsymbol{u_t}$ is a $k \times 1$ vector of reduced-form innovations with the covariance matrix $\boldsymbol{\Sigma_u}$, and $(\boldsymbol{B_0}, \boldsymbol{B_1}, \ldots, \boldsymbol{B_l})$ are the coefficient matrices.

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

---

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

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

3.5.1 Aggregate Equity vs Fixed-Income Funds

Since our model uncertainty measure is based on factors in the US, we delete all foreign mutual funds. In the baseline analysis, we consider the aggregate mutual fund flows to the entire equity and fixed-income markets. That is, we study the VAR regression in equation (3.17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows_t^{FI}, Flows_t^{Equity})$. We next use impulse response functions (IRFs) to better understand the dynamic effects and propagating mechanisms of uncertainty shocks.

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

In the benchmark case, we place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. We consider a different identification assumption in robustness checks in Section 3.8; that is, we put model uncertainty as the last element in $\boldsymbol{Y_t}$. As shown below, the IRFs to model uncertainty shocks are essentially robust to the alternative identification strategy, whereas the IRFs to other uncertainty measures are not.

Table 3.3 reports the results from the VAR estimation. The sample ranges from January 1991 to December 2020. The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also include the lagged market return and VXO index as control variables in each regression. The reported t-statistics are based on the Newey-West estimate of the covariance matrix with 36 lags. First, model uncertainty only relates to its lag. Second, the VXO index positively predicts the aggregate flows to fixed-income funds: one standard deviation increase in VXO predicts 0.17 standard deviation inflows to fixed-income funds. Third, model uncertainty negatively forecasts equity fund flows, and the coefficient estimate is sizable in both economic and statistical senses. In particular, one standard deviation increase in model uncertainty implies 0.34 standard deviation equity fund outflows. Although we cannot interpret the regression results as causal, we still find that investors in domestic equity mutual funds tend to decrease their exposures when model uncertainty increases.

190

Table 3.3: VAR Estimation of Monthly Entropy, Flows to Domestic Equity Funds, and Flows to Domestic Fixed-Income Funds

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

This table reports the results from the VAR estimation in equation (3.17), where $Y_t^{\top} = (Entropy_t, Flows_t^{FI}, Flows_t^{Equity})$. $Entropy_t$ is the model uncertainty measure, and $Flows_t^{FI}$ ($Flows_t^{Equity}$) is the aggregate flows to the domestic fixed-income (equity) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (3.16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged market return (MKT_t) and VXO index (VXO_t) in each regression. The sample spans from January 1991 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.
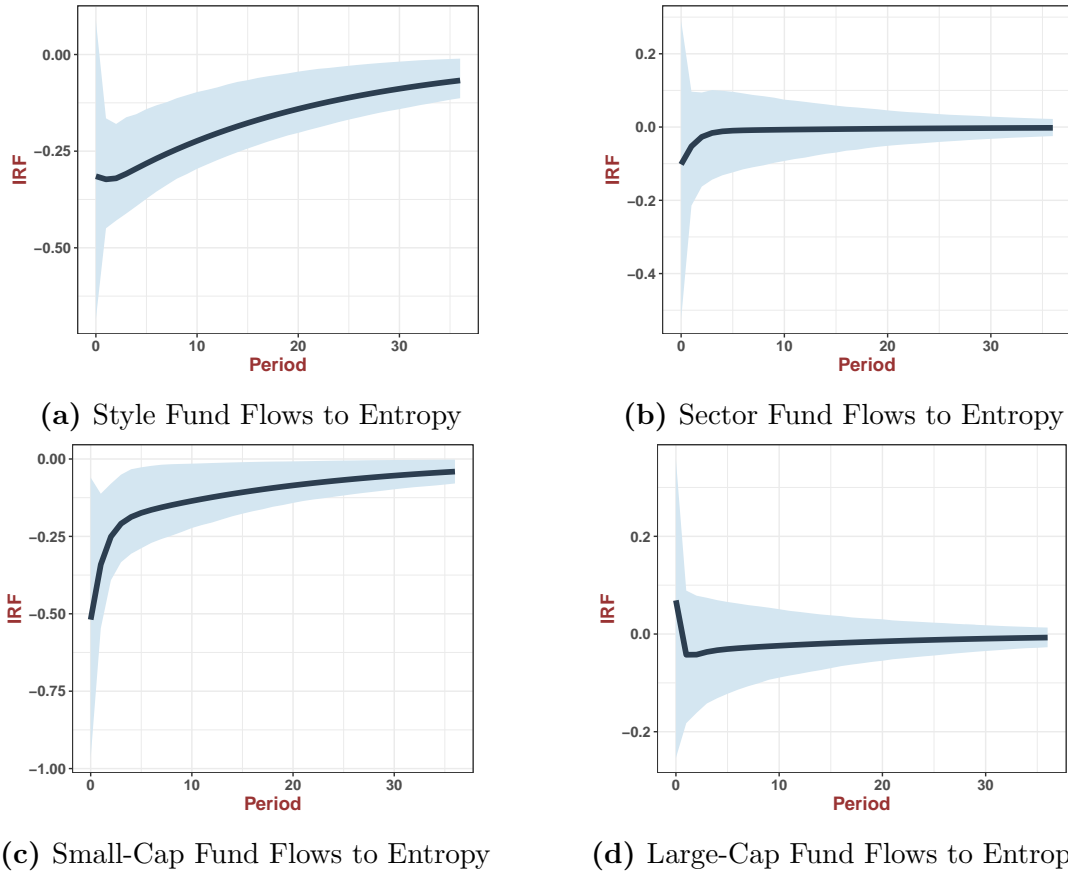
Figure 3.5 shows the dynamic responses of fund flows to model uncertainty shocks in VAR-1. Most strikingly, model uncertainty innovations sharply induce fund outflows from the US equity market, with the effects persisting even after 36 months, as depicted in Panel (a). The impulse response functions (IRFs) start from around -0.6 in period zero and slowly decline to -0.35 in period 36, significantly negative based on the 90% standard error bands. In contrast, model uncertainty has negligible effects on fixed-income fund flows (see Panel (b)).

3.5.2 Different Equity Mutual Funds

We further study the heterogeneous responses of different equity mutual funds to model uncertainty shocks. In particular, we split equity mutual funds into four categories: (a) style funds that specialize in factor investing, (b) sector funds that invest in specific industries (e.g., gold, oil, etc.), (c) small-cap funds that invest in relatively small stocks,[26] and (d) large-cap funds that invest in large stocks.

Table 3.4 reports the results from the VAR estimation in equation (3.17), where $Y_t^{\top} = (Entropy_t, Flows_t^{style}, Flows_t^{sector}, Flows_t^{small}, Flows_t^{large})$. The lag of VAR is chosen by BIC and equals one. Since the cap-based investment objective code is available after 1997, the sample begins in January 1998. First, after controlling its lag, model uncertainty is negatively predicted by large-cap fund flows and small-cap fund returns. Second, model uncertainty

---

[26]When we mention small funds, we refer to the funds with the CRSP investment objective codes equal "EDCM", "EDCS", and "EDCI".
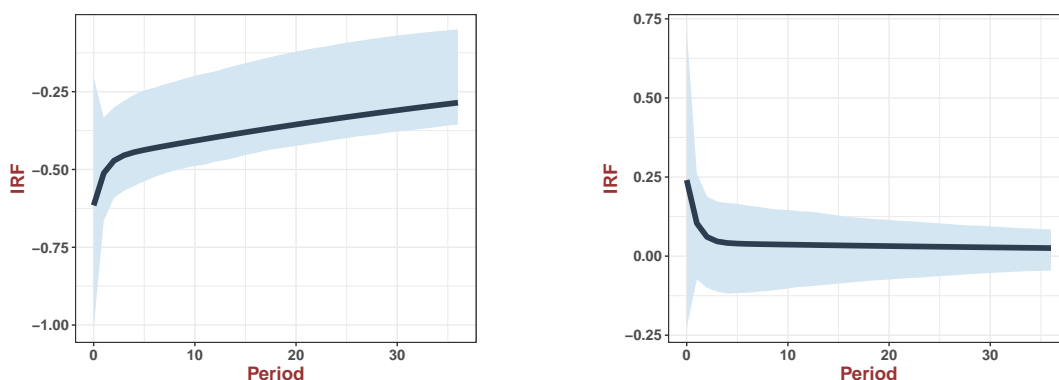
(a) Equity Fund Flows to Entropy   **(b)** Fixed-Income Fund Flows to Entropy

Figure 3.5: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows using Entropy as Uncertainty

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR-1. The shaded area denotes the 90 percent standard error bands. We consider mutual fund flows to aggregate equity and fixed-income markets in the US. We normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1991:01 - 2020:12.

negatively forecasts style and small-cap fund flows, and the coefficients are sizable. Specifically, if model uncertainty rises by one standard deviation, style (small-cap) fund flows tend to drop by 0.26 (0.12) standard deviation over the next period. On the contrary, we do not discover a significant relationship between model uncertainty and sector (large-cap) fund flows.

Different from model uncertainty, the traditional volatility-based uncertainty measure (VXO) plays a limited role in the VAR regression. It can marginally predict small-cap fund flows, but the sign of coefficient estimate is counter-intuitive: when uncertainty goes up, investors tend to invest more in small-cap funds. Instead, we observe a negative response of small-cap funds when using entropy as the uncertainty measure. Therefore, we argue that our model uncertainty index captures an essential source of investment risk for equity investors, which is omitted by the traditional VXO index.

Figure 3.6 shows the dynamic responses of four different types of equity fund flows to model uncertainty shocks in VAR-1. Consistent with Table 3.4, model uncertainty shocks reduce future style fund flows, and the effects are long-lasting (see Panel (a)). This observation is intuitive. Style funds refer to the growth, income, growth & income and "hedged" funds, so they are more likely to rely on the factor strategies used in constructing model uncertainty. Therefore, the outflows from style equity funds are remarkably enormous when the model uncertainty is high.

Table 3.4: VAR Estimation of Monthly Entropy and Flows to Domestic Equity Funds with Different Investment Objectives

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

This table reports the results from the VAR estimation in equation (3.17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows_t^{style}, Flows_t^{sector}, Flows_t^{small}, Flows_t^{large})$. $Entropy_t$ is the model uncertainty measure, and $Flows_t^{style}$ ($Flows_t^{sector}$, $Flows_t^{small}$, $Flows_t^{large}$) is the aggregate flows to the domestic style (sector, small-cap, large-cap) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (3.16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VXO index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

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

3.5.3 Different Fixed-Income Mutual Funds

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

Table 3.5 shows the results from the VAR-1 regression. According to columns (3) and (4), model uncertainty positively predicts the aggregate fund flows in US government bonds. US government bonds are notable for their superior safety over other asset classes. Hence, investors tend to allocate more wealth to safe assets when model uncertainty is more substantial. In contrast, model uncertainty negatively forecasts corporate fund flows, so mutual fund investors reduce their exposure to corporate bonds following high model uncertainty.

Next, we report the IRFs of different fixed-income funds to entropy shocks in Figure

(a) Style Fund Flows to Entropy

(b) Sector Fund Flows to Entropy

(c) Small-Cap Fund Flows to Entropy

(d) Large-Cap Fund Flows to Entropy

Figure 3.6: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes using Entropy as Uncertainty

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR-1. The shaded area denotes the 90 percent standard error bands. We consider equity fund flows with different investment objective codes (style, sector, small-cap, and large-cap). We normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1998:01 - 2020:12.

Table 3.5: VAR Estimation of Monthly Entropy and Flows to Domestic Fixed-Income Funds with Different Investment Objectives

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

This table reports the results from the VAR estimation in equation (3.17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows_t^{gov}, Flows_t^{money}, Flows_t^{corp}, Flows_t^{muni})$. $Entropy_t$ is the model uncertainty measure, and $Flows_t^{gov}$ ($Flows_t^{money}$, $Flows_t^{corp}$, $Flows_t^{muni}$) is the aggregate flows to the domestic government bond (money market, corporate bond, and municipal bond) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (3.16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the VXO index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

3.7. Not surprisingly, we document sharp dynamic inflows to government bond funds. As Panel (a) suggests, one standard deviation increase in model uncertainty corresponds to more than 0.7 standard deviation increase in government bond fund inflows at time zero, and the dynamic response persists for more than 36 periods. On the contrary, the IRFs of other fixed-income fund flows are not significant.

In addition, it is worth noting that we do not observe a significant relationship between model uncertainty and money market funds. The difference between money market and government bond funds is that the first type has a smaller duration and more liquid, while the latter consists of government bonds of different maturities. Unlike model uncertainty, the VXO index significantly predicts positive inflows to money market funds. We interpret these facts as evidence that high model uncertainty induces "flight to safety", whereas high VXO implies "flight to liquidity". Combined with the previous analyses, we conclude that mutual fund investors transfer their wealth from style and small-cap equity funds to government bonds, which are famous for their superior safety.

3.5.4 Comparison with Other Uncertainty Measures

One major concern about the previous analyses is that model uncertainty is correlated with other uncertainty indicators, so the dynamic responses of mutual fund flows to model uncertainty shocks are confounded by them. Hence, we study how other uncertainty measures affect mutual fund flows in this section and compare their dynamic responses with the previous results. We consider the VXO index and financial uncertainty in Jurado, Ludvigson, and
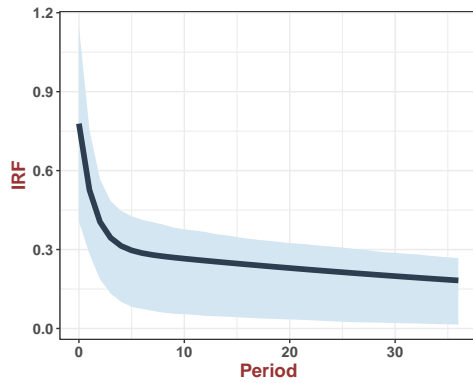
(a) Government Bonds to Entropy
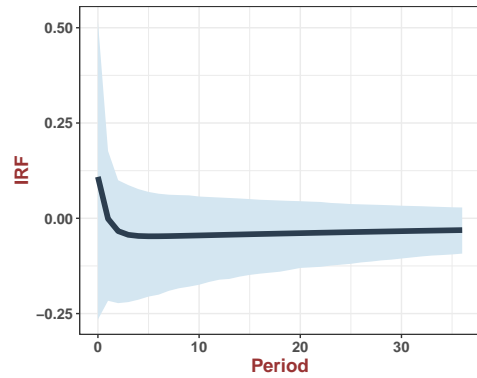
(b) Money Markets to Entropy
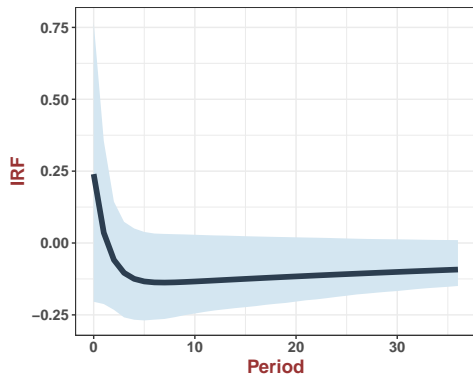
(c) Corporate Bonds to Entropy
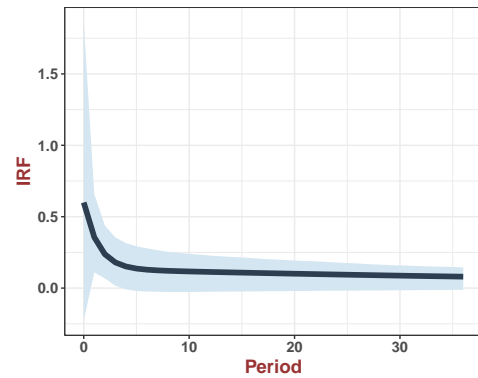
(d) Municipal Bonds to Entropy

Figure 3.7: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes using Entropy as Uncertainty

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR-1. The shaded area denotes the 90 percent standard error bands. We consider fixed-income fund flows with different investment objective codes (government bonds, money market, corporate bonds, and municipal bonds). We normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1991:01 - 2020:12.

Ng (2015) since these two measures are significantly associated with our model uncertainty measure, as we show in Table 3.2.

Figure 3.8 plots the dynamic responses of four different types of equity fund flows to VXO or financial uncertainty shocks in VAR-1. Consistent with the previous identification assumption, We place VXO or financial uncertainty first in the VAR. We also control the lagged model uncertainty in each regression. First, as Panels (a) and (b) indicate, style funds experience massive outflows when VXO or financial uncertainty increases. However, these effects are temporary; that is, the IRFs of fund flows reverse back to zeros immediately after time zero. On the contrary, model uncertainty shocks are followed by persistent outflows from style funds even beyond 36 periods. Similarly, the dynamic responses of fund flows to sector/small-cap/large-cap funds are also transitory and not significant (except for Panel (c) at period zero).

We further consider the dynamic responses of fixed-income funds in Figure 3.9. When VXO or financial uncertainty goes up, government bond funds tend to experience massive inflows, although these effects are less than 50% of those following model uncertainty shocks (see Figure 3.7(a)). Most strikingly, we document massive inflows to money market funds after positive VXO and financial uncertainty shocks. In contrast, model uncertainty does not play a part in money market funds. In other words, model uncertainty shocks primarily induce "flight to safety", while other volatility-based uncertainty measures are mainly related to "flight to liquidity".

In summary, our model uncertainty measure captures some unique dynamic responses of fund flows, and notably, they are different from traditional volatility-based measures, such as VXO and financial uncertainty. In particular, we observe significant fund inflows to government bond funds and outflows from style and small-cap equity funds. In contrast, VXO and financial uncertainty shocks fail to generate similar dynamic responses. Finally, as we will show in Section 3.8, the IRFs of fund flows to model uncertainty shocks are virtually robust to an alternative identification assumption, whereas the effects of VXO or financial uncertainty shocks tend to be fairly sensitive.

3.6   Investors' Expectations

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

(a) Style Fund Flows to VXO

(b) Style Fund Flows to Financial Unc

(c) Sector Fund Flows to VXO

(d) Sector Fund Flows to Financial Unc

(e) Small-Cap Fund Flows to VXO

(f) Small-Cap Fund Flows to Financial Unc

(g) Large-Cap Fund Flows to VXO

(h) Large-Cap Fund Flows to Financial Unc
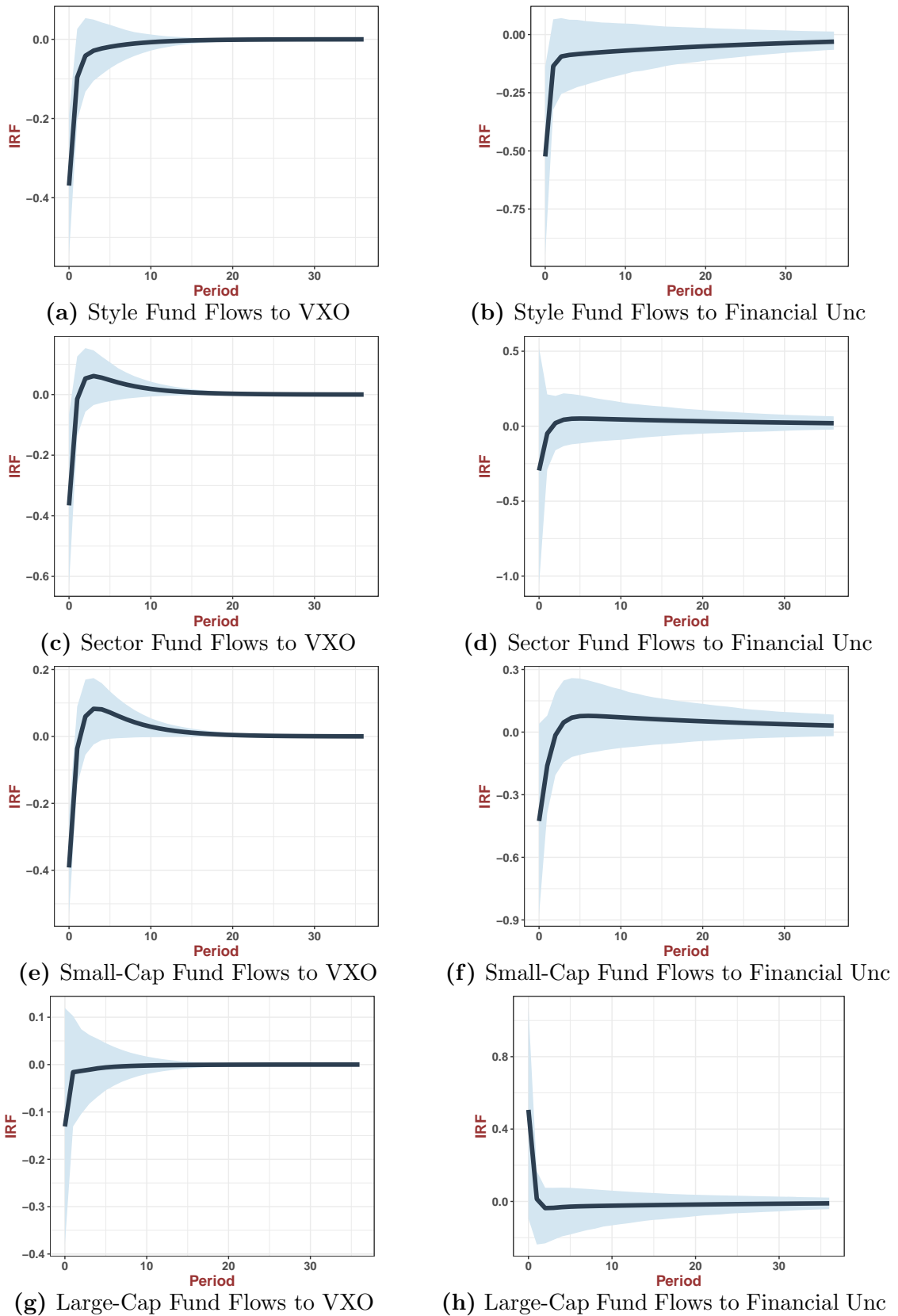
Figure 3.8: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes using VXO and Financial Uncertainty as Uncertainty Measures

This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to VXO and financial uncertainty shocks in VAR-1. Other details can be found in the footnote of Figure 3.6.

(a) Government Bonds to VXO

(b) Government Bonds to Financial Unc

(c) Money Markets to VXO

(d) Money Markets to Financial Unc

(e) Corporate Bonds to VXO

(f) Corporate Bonds to Financial Unc

(g) Municipal Bonds to VXO

(h) Municipal Bonds to Financial Unc
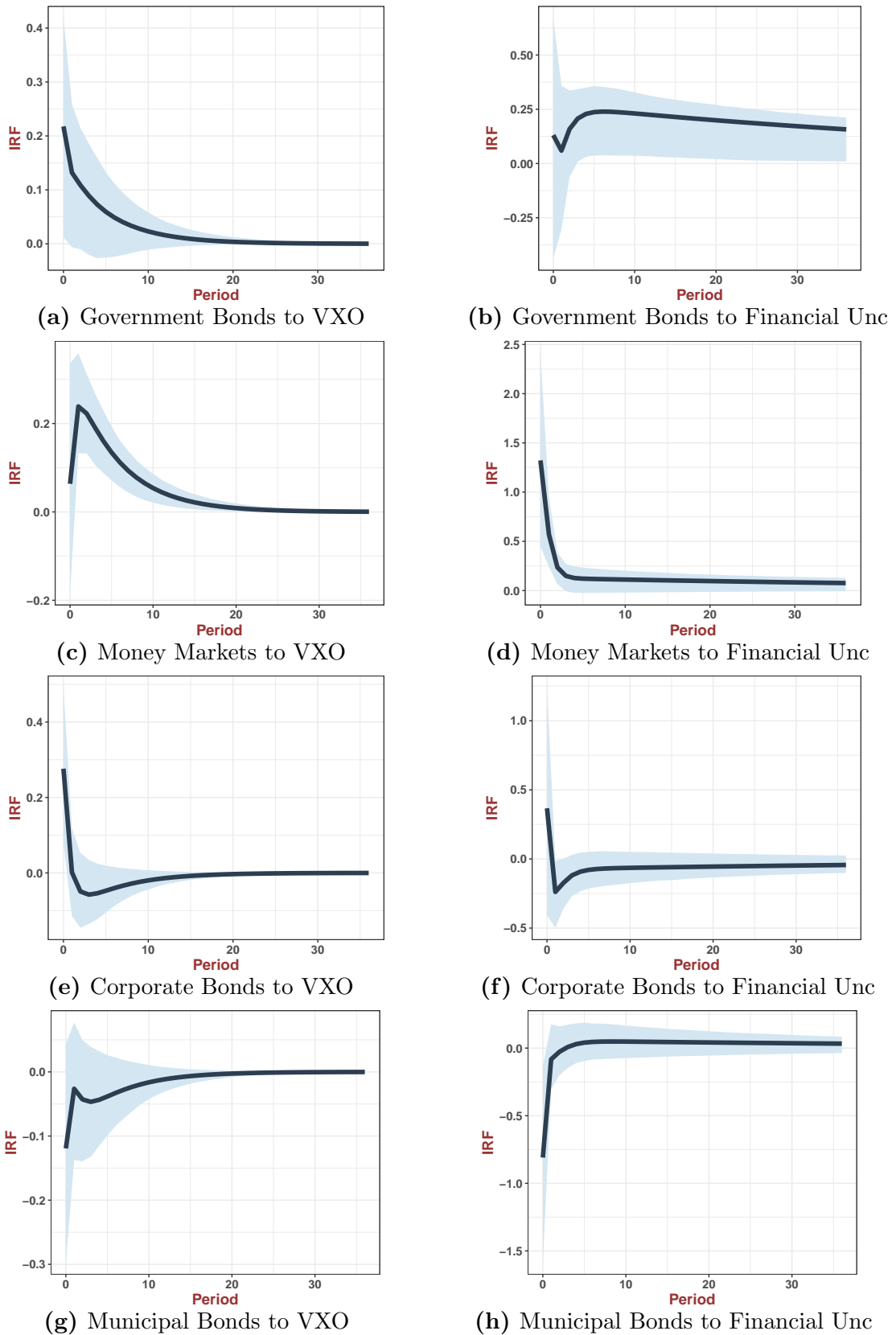
Figure 3.9: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes using VXO and Financial Uncertainty as Uncertainty Measures

This figure shows the dynamic impulse response functions (IRFs) of fixed-income fund flows to VXO and financial uncertainty shocks in VAR-1. Other details can be found in the footnote of Figure 3.7.

199

use the expectation measures in the last week of each month.

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

We consider the following time-series regression:

$$Exp_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1} \tag{3.18}$$

where Exp_{t+1} is the one-period ahead expectation measure, $Entropy_t$ is the model uncertainty measure in period t, and X_t includes other control variables up to time t, such as lagged expectation indices, VXO and etc. Since all expectation indices are autocorrelated, we control their one and two-period lags in all regressions.[27] 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 3.6(a), we regress AAII sentiment indices on model uncertainty to explore how individual investors change their attitudes towards the stock market in response to variations in model uncertainty. To increase the interpretability of our results, we standard model uncertainty to have unit variance, so coefficient estimates of $Entropy_t$ are interpreted as the increases in the percentages of bullish/neutral/bearish investors when model uncertainty grows by one standard deviation.

In columns (1) and (2), $Entropy_t$ cannot predict the next-period percentage of bullish investors. Specifically, the average investors become less bullish if model uncertainty in the cross-section goes up, but this prediction is not sharp. Columns (3) and (4) regress the percentage of neutral investors on lagged model uncertainty: If model uncertainty increases by one standard deviation, the fraction of neutral investors declines by 0.605% or 0.434%, depending on the regression setup.

The next question is, in which direction do bullish investors change their attitudes? Columns (5) and (6) indicate that investors are more likely to be bearish following an increase in model uncertainty. Our interpretation is that some neutral investors become bearish after observing a higher level of model uncertainty. Finally, we regress the difference between

---

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

fractions of bullish and bearish investors on entropy. The coefficient estimate of entropy is negative and significant at the 10% level. Overall, when model uncertainty goes up, market participants tend to be more pessimistic about the future stock market performance.

Table 3.6: Investors' Expectations, Confidence Indices, and Model Uncertainty

$Exp_{t+1} =$	Panel (a). AAII Sentiment Index							
	Bullish		Neutral		Bearish		Bullish - Bearish	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Entropy_t$	-0.280	-0.374	-0.605**	-0.434**	1.043**	1.036***	-1.511*	-1.574**
	(-0.683)	(-1.122)	(-2.121)	(-2.102)	(2.499)	(2.656)	(-1.826)	(-2.127)
VXO_t	0.022	0.079	0.016	-0.009	-0.008	-0.034	0.016	0.118
	(0.311)	(1.500)	(0.211)	(-0.161)	(-0.169)	(-0.500)	(0.157)	(1.060)
Exp_t	0.418***	0.373***	0.487***	0.452***	0.367***	0.335***	0.373***	0.331***
	(8.593)	(6.954)	(9.709)	(10.249)	(9.238)	(7.155)	(7.325)	(5.983)
Exp_{t-1}	0.098**	0.158***	0.213***	0.253***	0.182***	0.208***	0.118***	0.160***
	(2.434)	(3.531)	(6.103)	(6.209)	(5.676)	(5.850)	(3.151)	(3.623)
Lagged Market Returns	NO	YES	NO	YES	NO	YES	NO	YES
Sample Size	400	396	400	396	400	396	400	396
R^2_{adj}	21.76%	22.53%	43.11%	44.98%	27.24%	26.79%	20.92%	21.01%

$Exp_{t+1} =$	Panel (b). Shiller's Confidence Indices							
	1-Year Confidence Index - Institution		1-Year Confidence Index - Individual		Crash Confidence Index - Institution		Crash Confidence Index - Individual	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Entropy_t$	-0.365***	-0.379***	-0.546***	-0.682***	-0.562***	-0.635***	-0.754***	-0.754***
	(-2.727)	(-2.952)	(-2.733)	(-5.405)	(-3.265)	(-3.335)	(-5.790)	(-6.048)
VXO_t	0.025*	0.030	0.044*	0.080***	-0.058**	-0.034	-0.046***	-0.001
	(1.767)	(0.829)	(1.705)	(4.204)	(-2.153)	(-1.066)	(-2.712)	(-0.047)
Exp_t	1.133***	1.165***	0.931***	0.949***	1.068***	1.065***	1.086***	1.071***
	(16.820)	(18.984)	(11.730)	(15.898)	(19.165)	(21.126)	(16.459)	(13.272)
Exp_{t-1}	-0.270***	-0.304***	-0.015	-0.045	-0.217***	-0.219***	-0.241***	-0.208***
	(-3.603)	(-4.449)	(-0.212)	(-0.823)	(-3.540)	(-4.000)	(-4.268)	(-3.078)
Lagged Market Returns	NO	YES	NO	YES	NO	YES	NO	YES
Sample Size	232	228	232	228	232	228	232	228
R^2_{adj}	82.70%	83.38%	93.17%	93.25%	87.44%	87.01%	92.24%	92.82%

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

Table 3.6(b) regresses Shiller's confidence indices on entropy. Unlike the AAII sentiment index, we also observe the expectations of institutional investors. The results are generally similar to table 3.6: Investors tend to be more pessimistic about the stock market when model

uncertainty increases. They also believe that a market crash is more likely to occur following higher model uncertainty. One interesting empirical fact is that the coefficient estimates of $Entropy_t$ in the regressions of individual investors' confidence indices are always more negative than institutional investors. Hence, individual investors react more dramatically to the changes in model uncertainty than institutional ones.

In short, we conclude 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 performance.

3.7 Evidence in European and Asian Pacific Markets

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

Figure 3.10a plots the time series of model uncertainty in the European stock market from June 1996 to December 2020. Several results stand out. The time-series patterns in European markets[28] are remarkably similar to the US stock market. In particular, model uncertainty increases from 1999 and reaches its first peak between 2000 and 2001 because of the dot-com bubble burst. During these periods, model uncertainty almost touches its upper bound. After 2002, model uncertainty declines gradually and remains relatively low until the start of the 2008 global financial crisis. During this long-lasting economic and stock market crisis, model uncertainty stays close to the upper bound from 2008 to 2012 and only declines gradually after 2012. Finally, the uncertainty index shoots up again after 2015, similar to what we observe in the US market.

We next turn to discuss the findings in Asian Pacific markets.[29] It is worth noting that we observe some unique time-series variations in Asian stock markets. According to figure 3.10b,

---

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

[29]By saying the Asian Pacific market, we refer to the stock markets in Australia, Hong Kong, New Zealand, and Singapore.
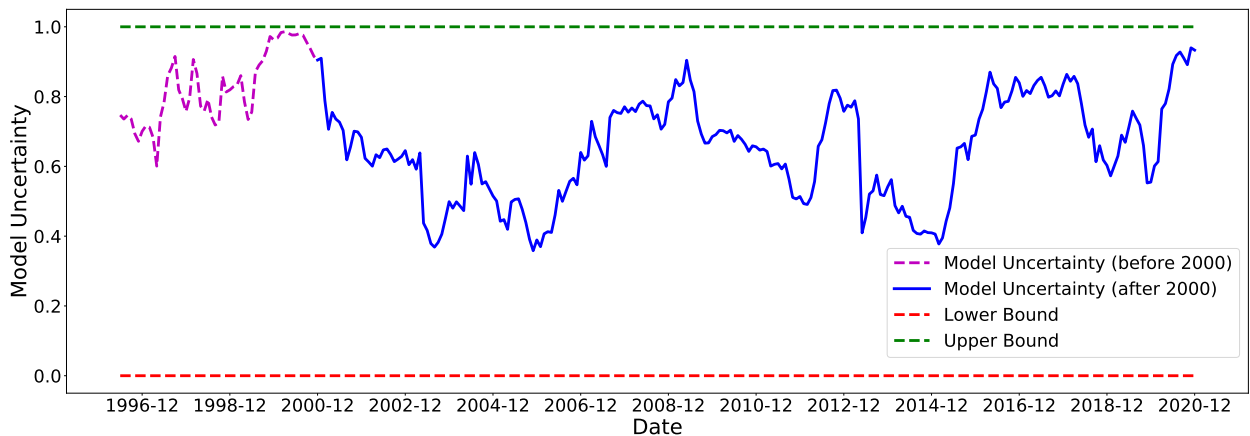
Figure 3.10: Model Uncertainty in European and Asian Pacific Markets

The figure plots the time series of model uncertainty about the linear stochastic discount factor (SDF) in European and Asian Stock Markets. The construction of model uncertainty is the same as in figure 3.1 except that we use only nine factors to calculate the posterior model probabilities. Details about used factors could be found in section 3.7. The sample ranges from July 1993 to December 2020. Since we use 3-year rolling window, the model uncertainty index starts from June 1996. The red line and green lines in the figure show the lower (0) and upper bounds (1) of model uncertainty.

model uncertainty 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 afterwards. Another steady increase in model uncertainty appears before and during the 2008 crisis, but the entropy is not as high as in the late 1990s and drops immediately from 2009. This particular pattern is unlike the US and European

markets, in which we observe higher model uncertainty of the 2008 crisis than the dot-com bubble.

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

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

3.8   Robustness Checks

This section considers several robust checks, including alternative hyper-parameter a in estimating factor models, alternative rolling windows in constructing the time series of model uncertainty, and a different identification assumption under which we re-estimate the dynamic responses of fund flows to uncertainty shocks.

3.8.1   Alternative Hyper-Parameter a

One important choice in our Bayesian inference is the value of hyper-parameter a. In the benchmark case, we assign a to be 4. Just as Section 3.2 shows, a higher a implies a stronger shrinkage for factors' risk prices, **b**.

Figure 3.A.2 plots the time series of model uncertainty using different values of a, including 3, 8, 16. Several findings stand out. First, we find that the time-series patterns in model uncertainty are not sensitive to the choice of a. In fact, the sequences under different values of a are virtually identical. Second, model uncertainty is increasing in a. This observation is not surprising since a larger a mechanically shrinks all candidate models to the null model, rendering factor models to become more similar and driving up model uncertainty.

3.8.2 Alternative Rolling Windows

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

Motivated by the above discussion, we consider four-year and five-year rolling windows in Figure 3.A.3. There is one tiny difference: Model uncertainty tends to be smoother in longer rolling windows, especially the five-year window. Beyond that, the time-series properties are similar to those found in a three-year rolling window.

3.8.3 Alternative Identification Assumption in VAR

Another robustness check concerns the identification assumption in our VAR analysis. In Section 3.5, we put uncertainty measures first in Y_t. We now consider an alternative setup, in which uncertainty measures are the last variables in Y_t. In other words, we allow uncertainty measures to correlate with contemporaneous shocks to mutual fund flows, but uncertainty shocks do not affect mutual fund flows simultaneously. Although model uncertainty is an endogenous response to innovations in fund flows under this assumption, it is still worth investigating whether model uncertainty is a key player to propagate those exogenous shocks over a long-lasting period.

Figures 3.A.5 and 3.A.6 plot the IRFs of fund flows to three uncertainty measures. Under the current assumption, the IRFs are zeros at period zero by construction. The first column shows the dynamic responses to model uncertainty shocks. Similar to the observations in Figures 3.6 and 3.7, an increase in model uncertainty relates to persistent outflows from style and small-cap funds but sharp inflows to government bond funds. The dynamic effects are bounded well away from zero even beyond 36 months, although they decline slowly over time. Hence, the main results in Figures 3.6 and 3.7 are largely robust.

The second and third columns show the IRFs of fund flows using VXO and financial uncertainty. Surprisingly, VXO shocks imply positive inflows to small-cap funds. On average, one standard deviation increase in the VXO index corresponds to more than 0.1 standard deviation fund inflows, and these positive dynamic responses last for around 20 months. However, the 90% confidence interval of IRFs covers zero effects, so they are on the edge of

being consequential. Beyond that, the IRFs in other panels are virtually zeros, so there is little evidence that mutual fund investors react to VXO or financial uncertainty shocks.

Finally, we observe significant inflows to money market funds following positive VXO shocks, and the dynamic responses have similar economic sizes to those in Figure 3.9. The key difference under the new identification assumption is that the IRFs of money market funds to financial uncertainty shocks are no longer significant. In other words, the dynamic responses to financial uncertainty shocks in Figure 3.9 are driven mainly by the identification assumption.

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

3.9 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 critical observation is that model uncertainty is countercyclical: it begins to climb up right before the stock market crashes and remains at its peaks during bear markets. Since we can calculate the lower and upper bound of entropy, we can easily discern when model uncertainty is abnormally high or low. In contrast, other uncertainty measures in past 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 the 2008 financial crisis.

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

similar countercyclical behaviours 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, at least according to our knowledge. Future research could attempt to endogenize model uncertainty in the general equilibrium model and explain its countercyclical behaviours.

Appendices

3.A.1 Description of Factors

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

Fama-French Five-factor model. Fama and French (1993) extend CAPM by introducing SMB and HML, where SMB is the return difference between portfolios of small and large stocks, and HML is the return difference between portfolios of stocks with high and low book-to-market ratios. Fama and French (2015) further include a profitability factor (RMW) and one investment factor (CMA). Again, the data comes from Ken French's website.

Momentumn. Jegadeesh and Titman (1993) find that stocks that perform well or poorly in the past three to 12 months continue their performance in the next three to 12 months. Therefore, investors can outperform the market by buying past winners and selling past losers. We download the momentum (MOM) factor from Ken French's data library.

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

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

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

Betting-against-beta. One of the most prominent failures of CAPM is that the security market line is too flat, so the risk premia of high-beta stocks are not as substantial as CAPM suggests. Frazzini and Pedersen (2014) constructs market-neutral betting-against-beta (BAB) factor that longs the low-beta stocks and shorts high-beta assets. We download the BAB factor from the AQR data library.

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

3.A.1.1 Additional Tables

Table 3.A.1: Summary Statistics of 14 Factors

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

This table reports the annualised mean returns and annualised Sharpe ratios of 14 factors listed in Appendix 3.A.1. The full sample starts from July 1972 to December 2020. We further split the entire sample into two equal subsamples.

Table 3.A.2: Summary of First-Order Autoregression

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

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 (2021) and Jurado, Ludvigson, and Ng (2015). EPU_1 and EPU_2 are two economic policy uncertainty sequences from Baker, Bloom, and Davis (2016). VXO is the forward-looking market volatility traded in CME. The t-statistics are computed using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

3.A.1.2 Additional Figures
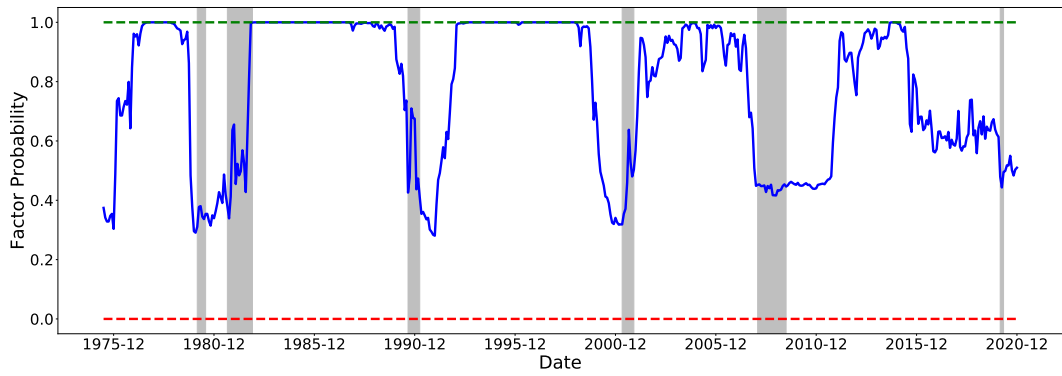
(a) MKT
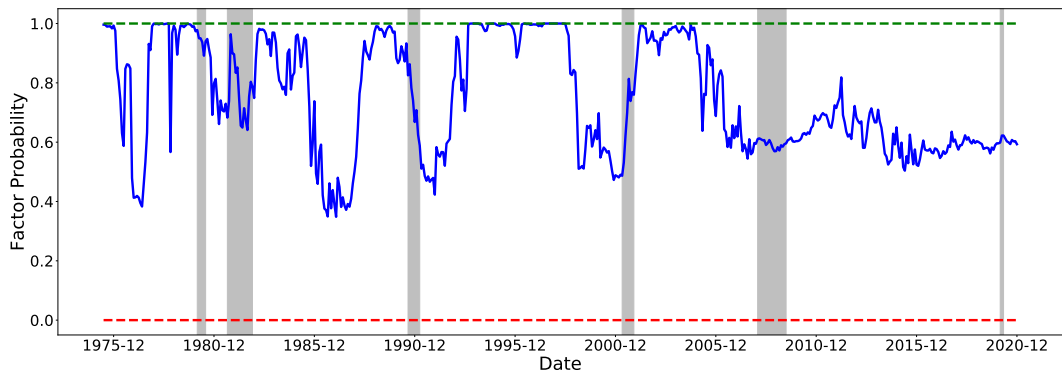
(b) Size (SMB or ME)
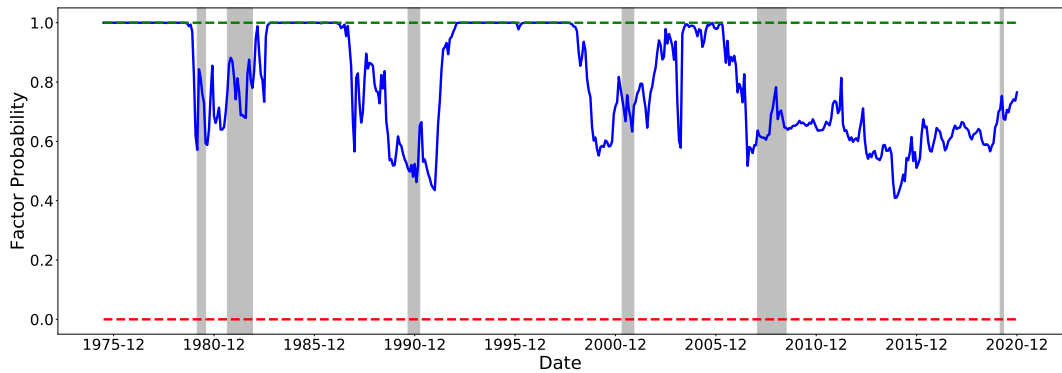
(c) Value (HML or HML_devil)
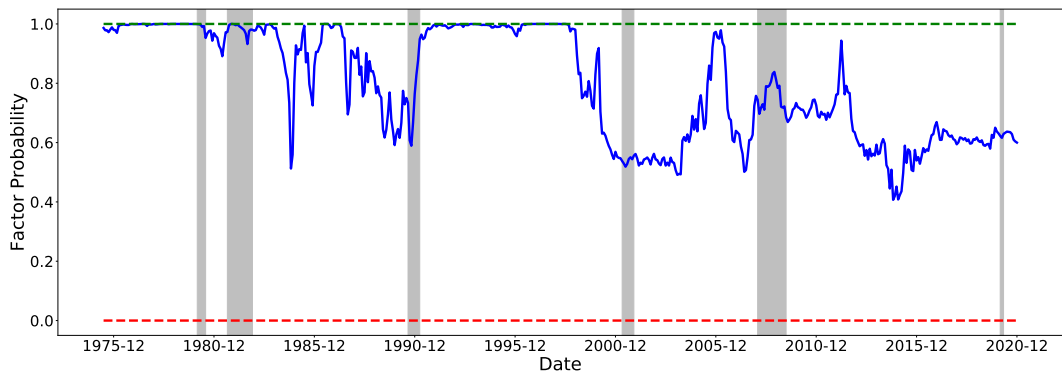
(d) Profitability (ROE or RMW)

Figure 3.A.1: Time Series of Posterior Factor Probabilities
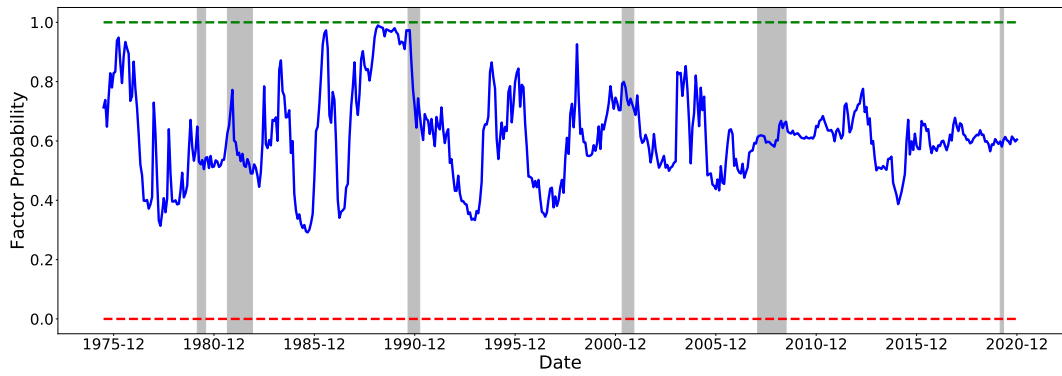
(e) Investment (IA or CMA)
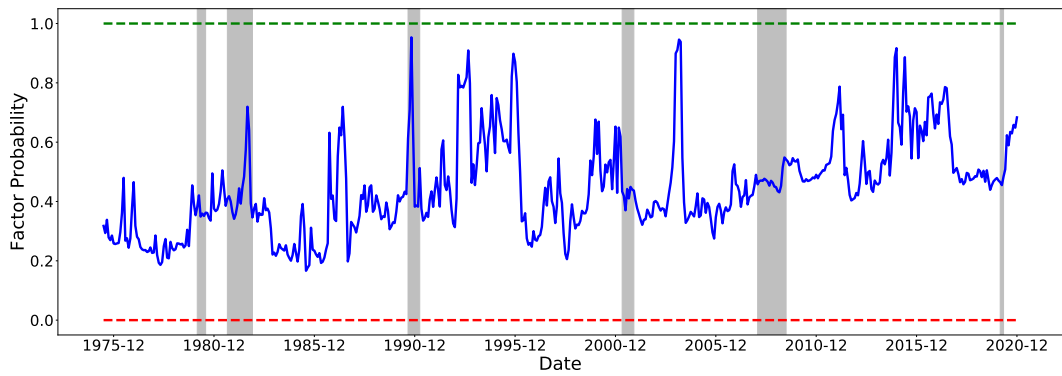
(f) Momentum
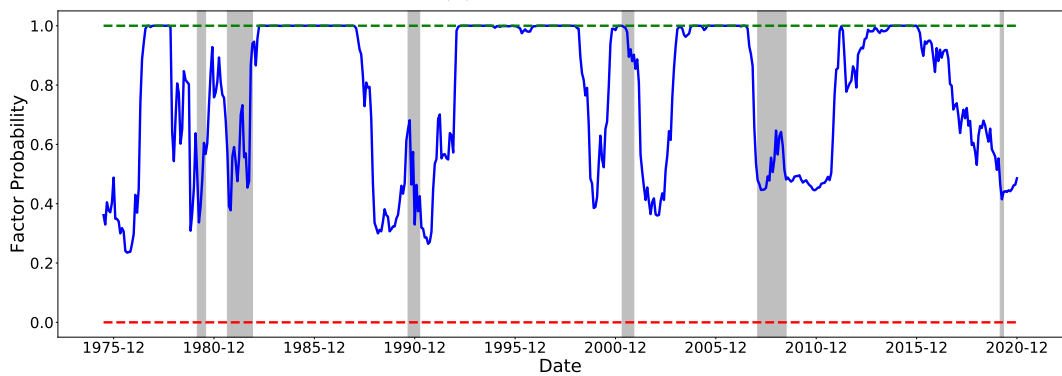
(g) BAB
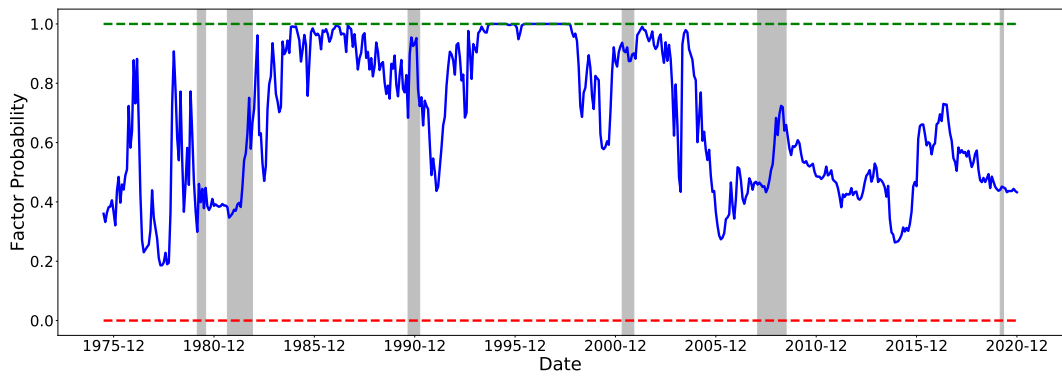
(h) QMJ

Figure 3.A.1: Time Series of Posterior Factor Probabilities (Continued)
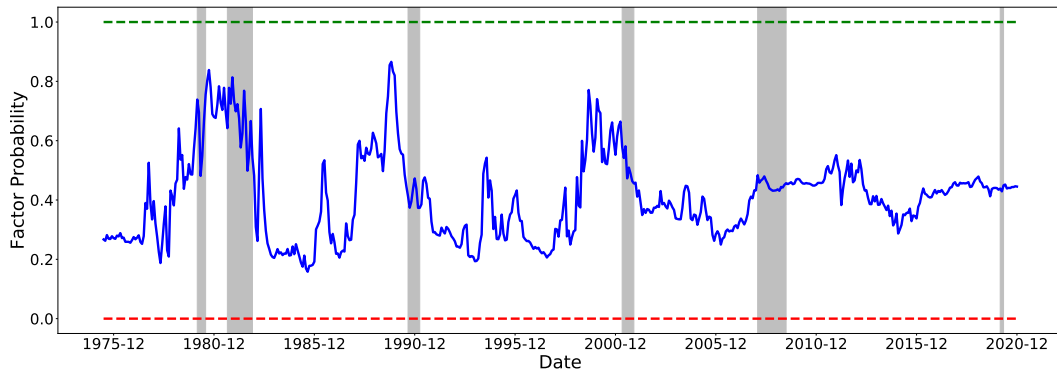
(i) FIN
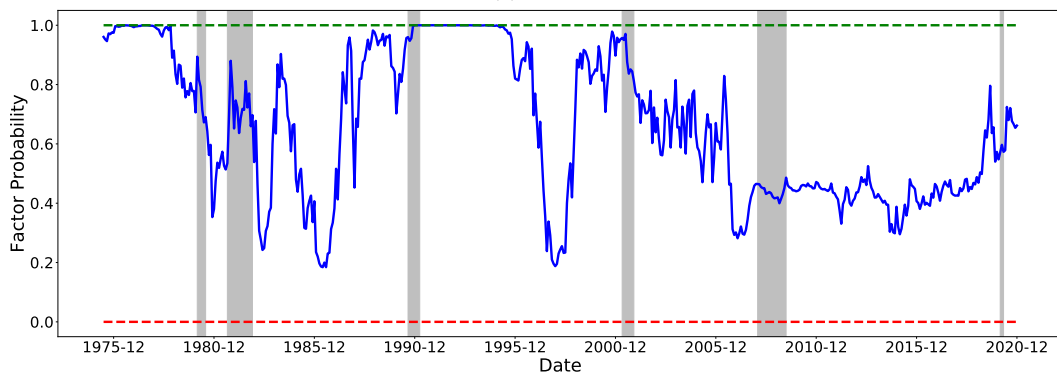
(j) PEAD

Figure 3.A.1: Time Series of Posterior Factor Probabilities (Continued)

The figures plot the time series of posterior marginal probabilities of 14 factors. At the end of each month, we estimate models using the daily factor returns in the past three years. The sample ranges from July 1972 to December 2020. Since we use a three-year rolling window, the time series of factor probabilities start from June 1975. Shaded areas are NBER-based recession periods for the US.
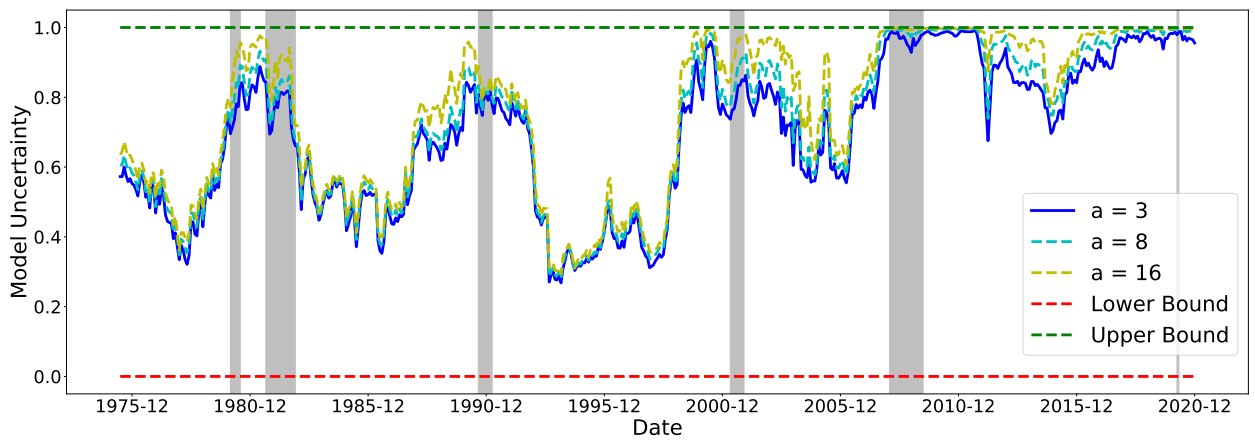
Figure 3.A.2: Time-Series of Model Uncertainty (3-Year Rolling Window) using different values of the hyper-parameter, $a \in \{3, 8, 16\}$
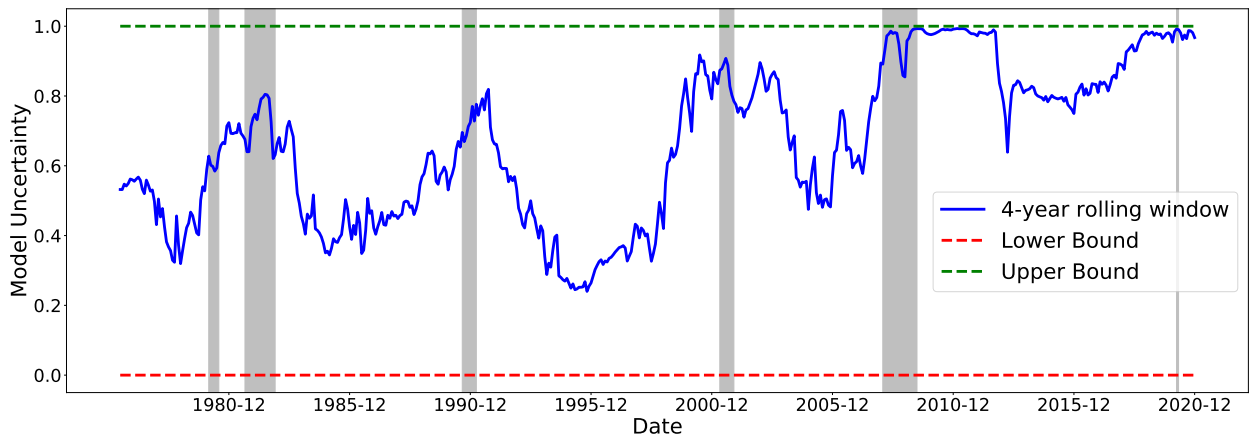
(a) Model Uncertainty in 4-Year Rolling Window
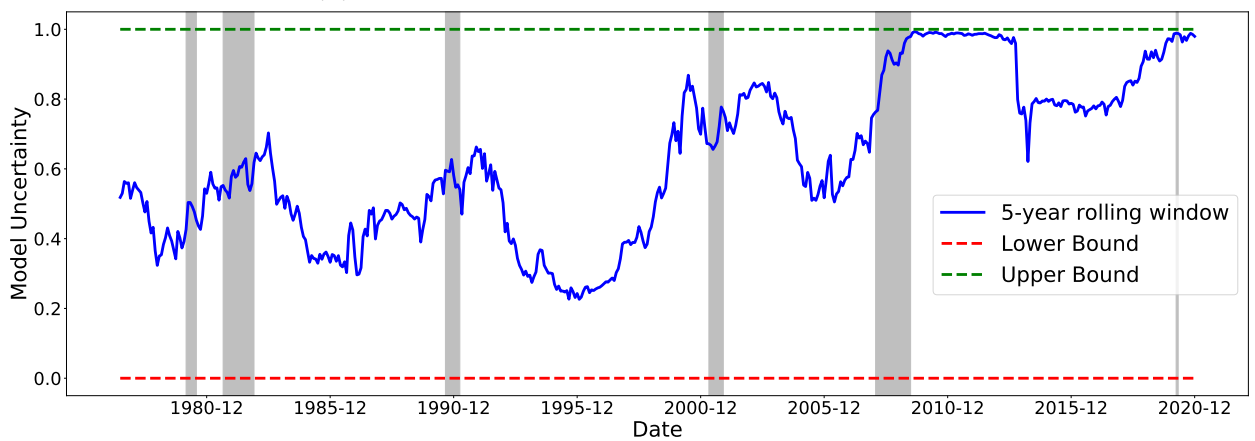
(b) Model Uncertainty in 5-Year Rolling Window

Figure 3.A.3: Alternative Rolling Windows

(a) Equity Fund Flows to Entropy

(b) Equity Fund Flows to VXO

(c) Equity Fund Flows to Financial Uncertainty

(d) Fixed-Income Fund Flows to Entropy

(e) Fixed-Income Fund Flows to VXO

(f) Fixed-Income Fund Flows to Financial Uncertainty
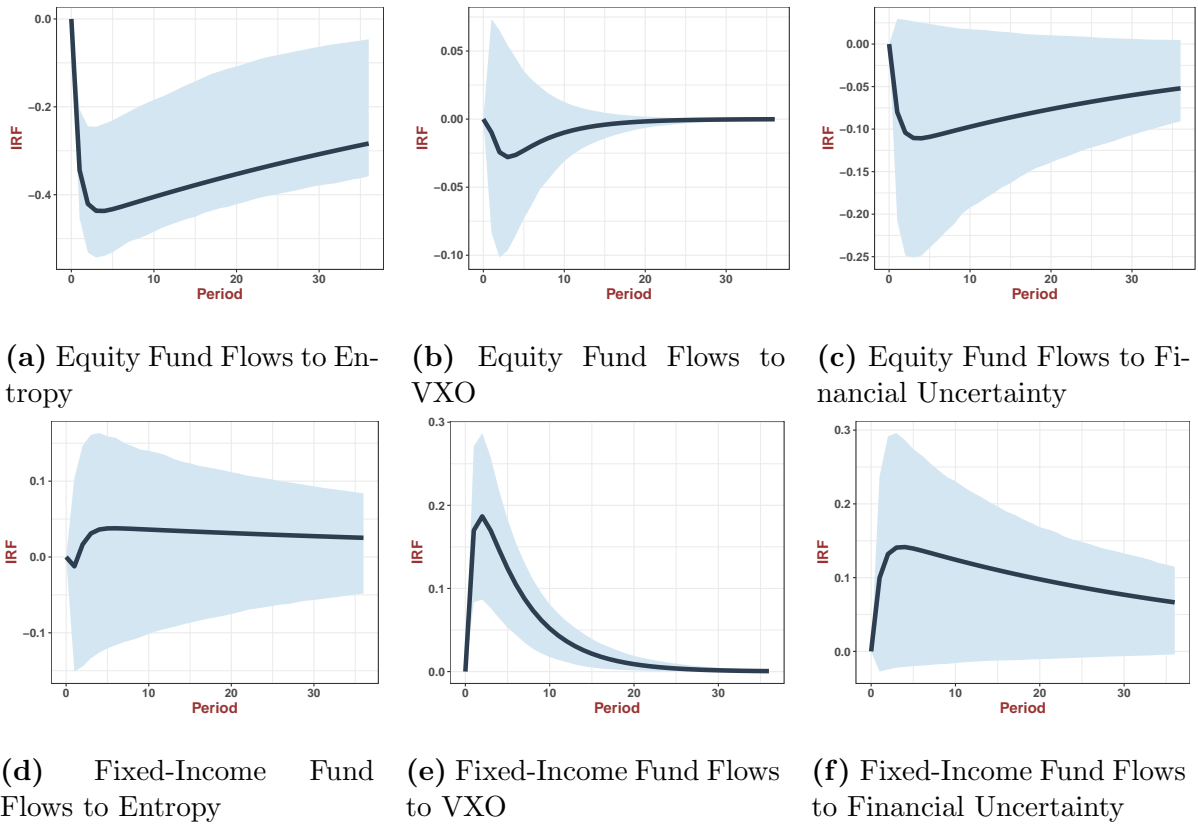
Figure 3.A.4: Robustness Check: Impulse Responses of Equity and Fixed-Income Fund Flows under Alternative Identification Assumption

This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to uncertainty shocks in VAR-1. We identity the IRFs by putting uncertainty last in VAR.
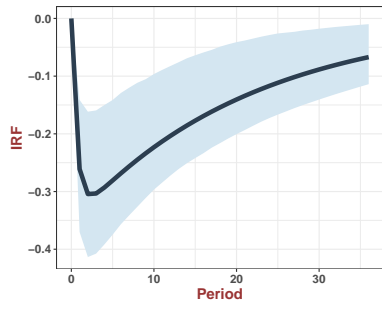
(a) Style Fund Flows to Entropy
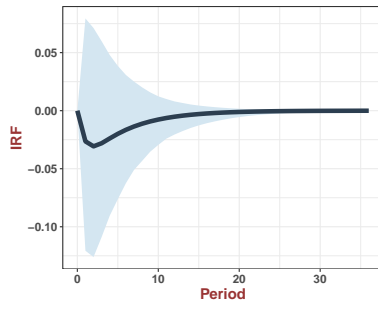
(b) Style Fund Flows to VXO
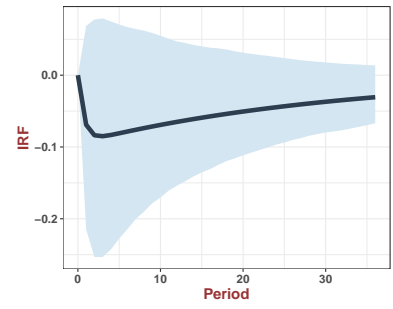
(c) Style Fund Flows to Financial Uncertainty
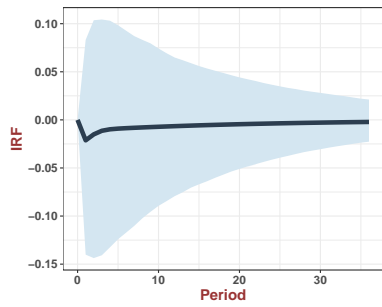
(d) Sector Fund Flows to Entropy
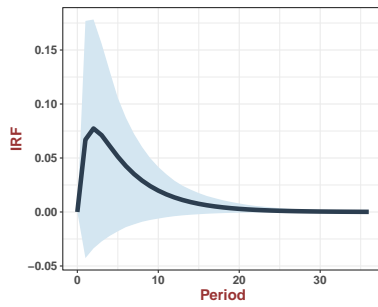
(e) Sector Fund Flows to VXO
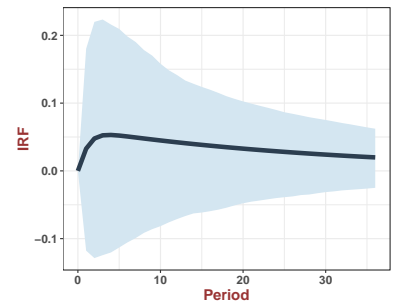
(f) Sector Fund Flows to Financial Uncertainty
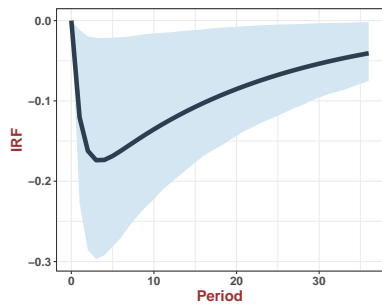
(g) Small-Cap Fund Flows to Entropy
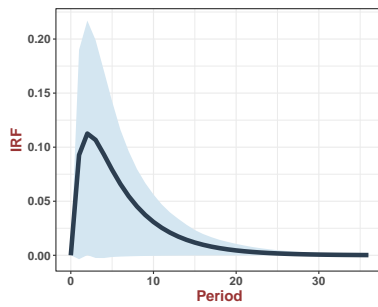
(h) Small-Cap Fund Flows to VXO
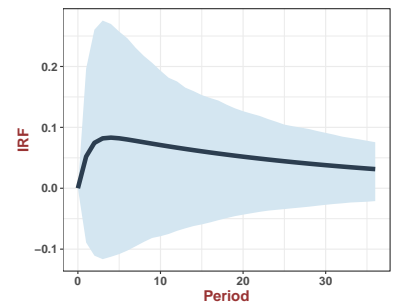
(i) Small-Cap Fund Flows to Financial Uncertainty
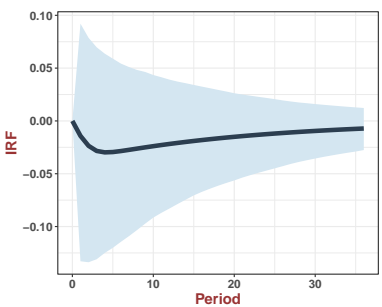
(j) Large-Cap Fund Flows to Entropy
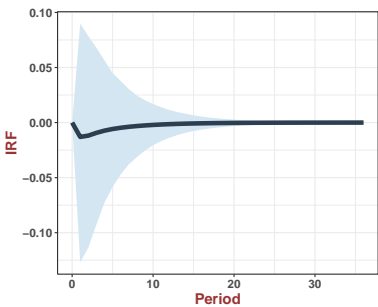
(k) Large-Cap Fund Flows to VXO
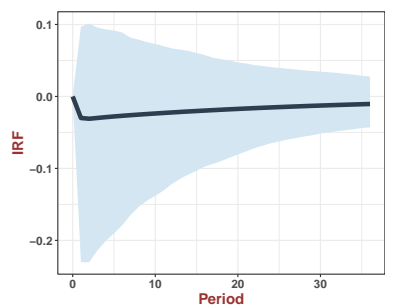
(l) Large-Cap Fund Flows to Financial Uncertainty

Figure 3.A.5: Robustness Check: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes under Alternative Identification Assumption

This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to uncertainty shocks in VAR-1. We identity the IRFs by putting uncertainty last in VAR.
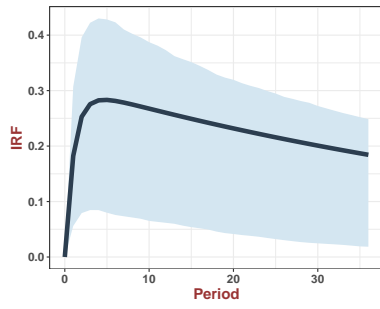
(a) Government Bonds to Entropy
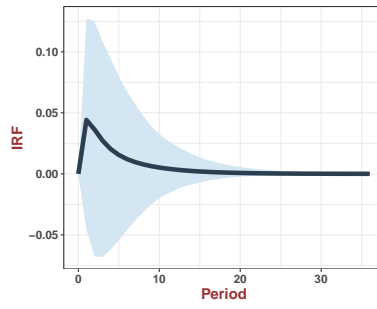
(b) Government Bonds to VXO
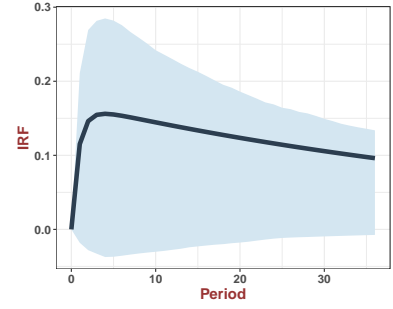
(c) Government Bonds to Financial Uncertainty
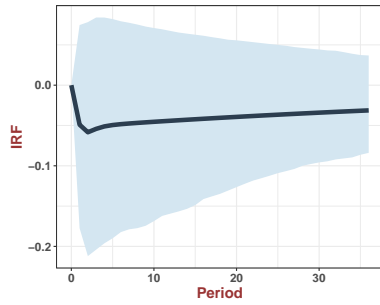
(d) Money Markets to Entropy
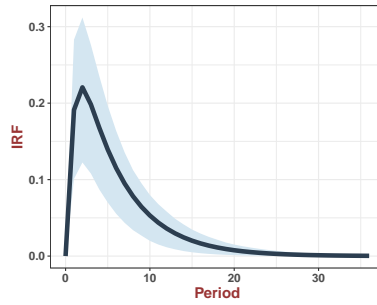
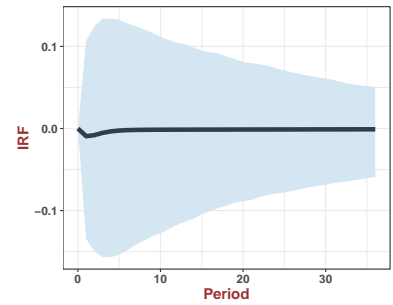**(e)** Money Markets to VXO
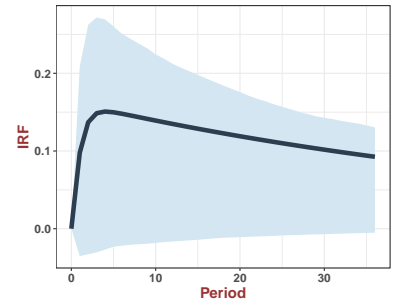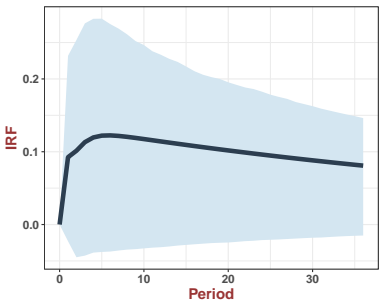
(f) Money Markets to Financial Uncertainty
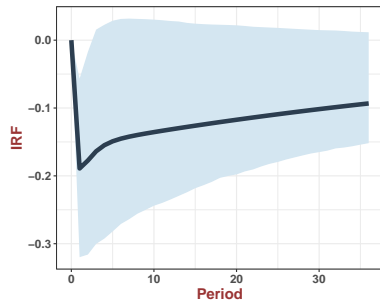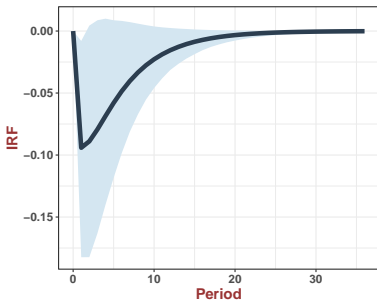
(g) Corporate Bonds to Entropy
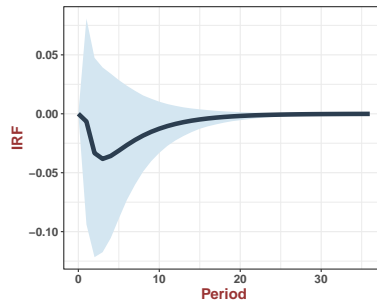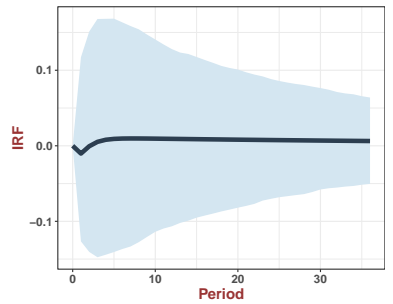
(h) Corporate Bonds to VXO

(i) Corporate Bonds to Financial Uncertainty

(j) Municipal Bonds to Entropy

(k) Municipal Bonds to VXO

(l) Municipal Bonds to Financial Uncertainty

Figure 3.A.6: Robustness Check: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes under Alternative Identification Assumption

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

3.A.2 Proofs

Lemma 3.A.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:*

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_\gamma = \begin{pmatrix} \boldsymbol{I}_{p_\gamma} \\ \boldsymbol{0}_{(N-p_\gamma)} \end{pmatrix}, \tag{19}$$

$$\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma = \boldsymbol{V}_\gamma. \tag{20}$$

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

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

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

$$= \begin{pmatrix} \boldsymbol{I}_{p_\gamma} \\ \boldsymbol{0}_{(N-p_\gamma)} \end{pmatrix},$$

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

3.A.2.1   Proof of Proposition 3.1

Proof. As in section 3.2.2, we assign g-prior for $\boldsymbol{b}_\gamma$: $\boldsymbol{b}_\gamma \mid \mathcal{M}_\gamma, g \sim \mathcal{N}\left(\boldsymbol{0}, \frac{g}{T}\left(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma\right)^{-1}\right)$. From lemma 3.A.1, $\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma = \boldsymbol{V}_\gamma$, so the prior distribution for $\boldsymbol{b}_\gamma$ is simplified as $\mathcal{N}\left(\boldsymbol{0}, \frac{g}{T}\boldsymbol{V}_\gamma^{-1}\right)$. Thus, the variance of linear SDF m_γ, conditioned that g and $\boldsymbol{V}_\gamma$ are known, is

$$\mathrm{Var}[m_\gamma] = \mathbb{E}\left[\mathrm{Var}\left[(\boldsymbol{f}_\gamma - \mathbb{E}[\boldsymbol{f}_\gamma])^\top \boldsymbol{b}_\gamma \mid \boldsymbol{b}_\gamma\right]\right] + \mathrm{Var}\left[\mathbb{E}\left[1 - (\boldsymbol{f}_\gamma - \mathbb{E}[\boldsymbol{f}_\gamma])^\top \boldsymbol{b}_\gamma \mid \boldsymbol{b}_\gamma\right]\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{b}_\gamma^\top \boldsymbol{V}_\gamma \boldsymbol{b}_\gamma\right)\right] + \mathrm{Var}\left[1 - \boldsymbol{0}^\top \boldsymbol{b}_\gamma\right]$$

$$= \mathrm{tr}\left(\boldsymbol{V}_\gamma \mathbb{E}\left[\boldsymbol{b}_\gamma \boldsymbol{b}_\gamma^\top\right]\right) + 0$$

$$= \mathrm{tr}\left(\boldsymbol{V}_\gamma \frac{g}{T}\boldsymbol{V}_\gamma^{-1}\right)$$

$$= \frac{g p_\gamma}{T}$$

This completes the proof of Proposition 3.1. ■

3.A.2.2   Proof of Proposition 3.2 and 3.3

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

$$\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \mid \mathcal{M}_\gamma, b, g \overset{\mathrm{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_\gamma \boldsymbol{b}_\gamma, \boldsymbol{\Sigma}). \tag{21}$$

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

$$p[\mathcal{D} \mid \mathcal{M}_\gamma, b, g] = (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\{-\frac{1}{2}\sum_{t=1}^T (\boldsymbol{R}_t - \boldsymbol{C}_\gamma \boldsymbol{b}_\gamma)^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\boldsymbol{R}_t - \boldsymbol{C}_\gamma \boldsymbol{b}_\gamma)\}. \tag{22}$$

In order to find the posterior model probabilities, we need to derive the marginal likelihood of data $\mathcal{D}$ conditional on model $\mathcal{M}_\gamma$. First of all, we find $p[\mathcal{D} \mid \mathcal{M}_\gamma, g]$ by integrating out $\boldsymbol{b}_\gamma$. We assign g-prior for $\boldsymbol{b}_\gamma$: $\boldsymbol{b}_\gamma \mid \mathcal{M}_\gamma \sim \mathcal{N}\left(\boldsymbol{0}, \frac{g}{T}\left(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma\right)^{-1}\right)$, thus

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

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

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

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

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

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

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

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

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

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

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

To make $p[\mathcal{D}\mid\mathcal{M}_\gamma,\,g]$ 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 3.2, that is,

$$
\begin{aligned}
p[\mathcal{D}\mid\mathcal{M}_\gamma,\,g] &= \exp\Big\{-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\boldsymbol{R}_t-\overline{\boldsymbol{R}})(\boldsymbol{R}_t-\overline{\boldsymbol{R}})^\mathsf{T})-\frac{T}{2}\Big(\mathrm{SR}_{max}^2-\frac{g}{1+g}\mathrm{SR}_\gamma^2\Big)\Big\}\frac{(1+g)^{-\frac{p_\gamma}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \\
&= \exp\Big\{-\frac{T-1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)-\frac{T}{2}\Big(\mathrm{SR}_{max}^2-\frac{g}{1+g}\mathrm{SR}_\gamma^2\Big)\Big\}\frac{(1+g)^{-\frac{p_\gamma}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}
\end{aligned}
\tag{23}
$$

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

$$
p[\mathcal{D}\mid\mathcal{M}_\gamma,\,g] \propto (1+g)^{-\frac{p_\gamma}{2}}\exp\Big\{\frac{gT}{2(1+g)}\mathrm{SR}_\gamma^2\Big\}
\tag{24}
$$

An equivalent way to think about equation (24) is to treat it as the Bayes factor of model $\mathcal{M}_\gamma$ relative to $\mathcal{M}_0$. One amazing fact is that $p[\mathcal{D}\mid\mathcal{M}_0,\,g]$ does not depend on g[30].

---

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

Therefore, the Bayes factor could be defined as

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

The prior for g is such that $\pi[g] = \frac{a-2}{2}(1 + g)^{-\frac{a}{2}}$. We calculate the marginal likelihood of data only conditional on model $\mathcal{M}_\gamma$ by integrating out g in equation (24).

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

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_γ is defined as $s_\gamma = \frac{p_\gamma + a}{2} - 1$. We have proved the formula of Bayes factor BF_γ in Proposition 3. To prove that the Bayes factor is always increasing in SR_γ^2 always decreasing in p_γ, we use the original representation of Bayes Factor, that is,

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

Take the first-order derivative with respect to SR_γ^2 and p_γ:

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

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

This completes the proof of Proposition 3. ∎

Bibliography

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

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

ALLENA, R. (2019): "Comparing Asset Pricing Models with Traded and Non-Traded Factors," working paper.

ALVAREZ, F., AND U. J. JERMANN (2005): "Using asset prices to measure the persistence of the marginal utility of wealth," *Econometrica*, 73(6), 1977–2016.

ANDERSON, E., AND A.-R. CHENG (2016): "Robust Bayesian Portfolio Choices," *Review of Financial Studies*, 29, 1330–1375.

ANG, A., J. CHEN, AND Y. XING (2006): "Downside risk," *Review of Financial Studies*, 19(4), 1191–1239.

ASNESS, C., AND A. FRAZZINI (2013): "The devil in HML's details," *Journal of Portfolio Management*, 39(4), 49–68.

ASNESS, C. S., A. FRAZZINI, AND L. H. PEDERSEN (2019): "Quality minus Junk," *Review of Accounting Studies*, 24(1), 34–112.

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

——— (2004): "Stock Return Predictability and Asset Pricing Models," *Review of Financial Studies*, 17, 699–738.

AVRAMOV, D., S. CHENG, L. METZKER, AND S. VOIGT (2021): "Integrating Factor Models," *Available at SSRN 3924337*.

AVRAMOV, D., AND G. ZHOU (2010): "Bayesian Portfolio Analysis," *Annual Review of Financial Economics*, 2, 25–47.

BAI, J. (2003): "Inferential theory for factor models of large dimensions," *Econometrica*, 71(1), 135–171.

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

BAKER, M., AND J. WURGLER (2006): "Investor sentiment and the cross-section of stock returns," *Journal of Finance*, 61(4), 1645–1680.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring economic policy uncertainty," *Quarterly Journal of Economics*, 131(4), 1593–1636.

BAKS, K. P., A. METRICK, AND J. WACHTER (2001): "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation," *Journal of Finance*, 56, 45–85.

BANDI, F. M., S. E. CHAUDHURI, A. W. LO, AND A. TAMONI (2021): "Spectral factor models," *Journal of Financial Economics*.

BANSAL, R., AND A. YARON (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance*, 59(4), 1481–1509.

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," working paper.

BARRO, R. J. (2006): "Rare disasters and asset markets in the twentieth century," *Quarterly Journal of Economics*, 121(3), 823–866.

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

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 81, 608–650.

BENAYCH-GEORGES, F., AND R. R. NADAKUDITI (2011): "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, 227(1), 494–521.

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.

BREEDEN, D. T. (1979): "An intertemporal asset pricing model with stochastic consumption and investment opportunities," *Journal of Financial Economics*, 7(3).

BRENNAN, M. J., AND Y. ZHANG (2020): "Capital asset pricing with a stochastic horizon," *Journal of Financial and Quantitative Analysis*, 55(3), 783–827.

BRUNNERMEIER, M. K., AND L. H. PEDERSEN (2009): "Market liquidity and funding liquidity," *Review of Financial Studies*, 22(6), 2201–2238.

BRYZGALOVA, S. (2015): "Spurious Factors in Linear Asset Pricing Models," working paper.

BRYZGALOVA, S., J. HUANG, AND C. JULLIARD (2021): "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *available at SSRN 3481736*.

CABALLERO, R. J., AND A. KRISHNAMURTHY (2008): "Collective risk management in a flight to quality episode," *Journal of Finance*, 63(5), 2195–2230.

CAMPBELL, J. Y. (1990): "A variance decomposition for stock returns," *Economic Journal*, 101(2), 157–179.

———— (1999): "Asset prices, consumption, and the business cycle," *Handbook of Macroeconomics*, 1, 1231–1303.

CAMPBELL, J. Y., AND J. H. COCHRANE (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107(2), 205–251.

CAMPBELL, J. Y., S. GIGLIO, AND C. POLK (2013): "Hard times," *Review of Asset Pricing Studies*, 3(1), 95–132.

CAMPBELL, J. Y., AND T. VUOLTEENAHO (2004): "Bad beta, good beta," *American Economic Review*, 94(5), 1249–1275.

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

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica*, 51(5), 1281–1304.

CHEN, L., M. PELGER, AND J. ZHU (2019): "Deep Learning in Asset Pricing," working paper.

CHERNOV, M., L. A. LOCHSTOER, AND S. R. LUNDEBY (2022): "Conditional dynamics and the multi-horizon risk-return trade-off," *Review of Financial Studies*, 35(3), 1310–1347.

CHIB, S., X. ZENG, AND L. ZHAO (2020): "On Comparing Asset Pricing Models," *Journal of Finance*, 75(1), 551–577.

COCHRANE, J. H. (2009): *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.

CONNOR, G., AND R. A. KORAJCZYK (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of Financial Economics*, 15(3), 373–394.

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

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

CRSP (2021): Calculated (or Derived) based on data from CRSP ©2021 Center for Research in Security Prices (CRSP®), The University of Chicago Booth School of Business.

DANIEL, K., D. HIRSHLEIFER, AND L. SUN (2020): "Short- and Long-Horizon Behavioral Factors," *Review of Financial Studies*, 33, 1673–1736.

DANIEL, K., L. MOTA, S. ROTTKE, AND T. SANTOS (2020): "The cross-section of risk and returns," *Review of Financial Studies*, 33(5), 1927–1979.

DANIEL, K., AND S. TITMAN (2012): "Testing Factor-Model Explanations of Market Anomalies," *Critical Finance Review*, 1(1), 103–139.

DE LONG, J. B., A. SHLEIFER, L. H. SUMMERS, AND R. J. WALDMANN (1990): "Noise trader risk in financial markets," *Journal of Political Economy*, 98(4), 703–738.

DEW-BECKER, I., AND S. GIGLIO (2016): "Asset pricing in the frequency domain: theory and empirics," *Review of Financial Studies*, 29, 2029–68.

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 (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116(1), 1–22.

FAMA, E. F., AND J. D. MACBETH (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81(3), 607–636.

FENG, G., S. GIGLIO, AND D. XIU (2020): "Taming the factor zoo: A test of new factors," *Journal of Finance*, 75(3), 1327–1370.

FRAZZINI, A., AND L. H. PEDERSEN (2014): "Betting against beta," *Journal of Financial Economics*, 111(1), 1–25.

GABAIX, X. (2012): "Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance," *Quarterly Journal of Economics*, 127(2), 645–700.

GARLAPPI, L., R. UPPAL, AND T. WANG (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *Review of Financial Studies*, 20, 41–81.

GEORGE, E. I., AND R. E. MCCULLOCH (1993): "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

——— (1997): "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–373.

GEWEKE, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2021a): "Economic predictions with big data: The illusion of sparsity," *ECB Working Paper*.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2021b): "Economic Predictions with Big Data: The Illusion of Sparsity," *Econometrica*, forthcoming.

GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57(5), 1121–1152.

GIGLIO, S., AND D. XIU (2021): "Asset pricing with omitted factors," *Journal of Political Economy*, 129(7), 1947–1990.

GIGLIO, S., D. XIU, AND D. ZHANG (2021): "Test assets and weak factors," Discussion paper, National Bureau of Economic Research.

GOLUBOV, A., AND T. KONSTANTINIDI (2019): "Where is the risk in value? Evidence from a market-to-book decomposition," *Journal of Finance*, 74(6), 3135–3186.

GOSPODINOV, N., R. KAN, AND C. ROBOTTI (2014): "Misspecification-Robust Inference in Linear Asset-Pricing Models with Irrelevant Risk Factors," *Review of Financial Studies*, 27(7), 2139–2170.

——— (2019): "Too Good to Be True? Fallacies in Evaluating Risk Factor Models," *Journal of Financial Economics*, 132(2), 451–471.

GOSPODINOV, N., AND C. ROBOTTI (2021a): "Common Pricing across Asset Classes: Empirical Evidence Revisited," *Journal of Financial Economics*, 140, 292–324.

——— (2021b): "Common Pricing across Asset Classes: Empirical Evidence Revisited," working paper.

GREENWOOD, R., AND A. SHLEIFER (2014): "Expectations of returns and expected returns," *Review of Financial Studies*, 27(3), 714–746.

GU, S., B. KELLY, AND D. XIU (2020): "Empirical Asset Pricing via Machine Learning," *Review of Financial Studies*, 33(5), 2223–2273.

Gupta, T., and B. Kelly (2019): "Factor momentum everywhere," *Journal of Portfolio Management*, 45(3), 13–36.

Haddad, V., S. Kozak, and S. Santosh (2020): "Factor timing," *Review of Financial Studies*, 33(5), 1980–2018.

Handa, P., S. P. Kothari, and C. Wasley (1989): "The relation between the return interval and betas: Implications for the size effect," *Journal of Financial Economics*, 23(1), 79–100.

Hannan, E. J. (2009): *Multiple Time Series*, vol. 38. John Wiley & Sons.

Hansen, L. P. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption strikes back? Measuring long-run risk," *Journal of Political Economy*, 116(2), 260–302.

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.

Hansen, L. P., and J. A. Scheinkman (2009): "Long-term risk: An operator approach," *Econometrica*, 77(1), 177–234.

Harvey, C., and Y. Liu (2019): "Cross-Sectional Alpha Dispersion and Performance Evaluation," *Journal of Financial Economics*, forthcoming.

Harvey, C., and G. Zhou (1990): "Bayesian Inference in Asset Pricing Tests," *Journal of Financial Economics*, 26, 221–254.

Harvey, C. R. (2017): "Presidential Address: The Scientific Outlook in Financial Economics," *Journal of Finance*, 72(4), 1399–1440.

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

He, A., D. Huang, and G. Zhou (2018): "New Factors Wanted: Evidence from a Simple Specification Test," working paper, available at SSRN: https://ssrn.com/abstract=3143752.

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

He, Z., and A. Krishnamurthy (2013): "Intermediary asset pricing," *American Economic Review*, 103(2), 732–70.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–401.

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

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015): "Investor sentiment aligned: A powerful predictor of stock returns," *Review of Financial Studies*, 28(3), 791–837.

Huang, D., J. Li, and G. Zhou (2018): "Shrinking Factor Dimension: A Reduced-Rank Approach," working paper, available at SSRN: https://ssrn.com/abstract=3205697.

Ishwaran, H., J. S. Rao, et al. (2005): "Spike-and-Slab Variable Selection: Frequentist and Bayesian Strategies," *Annals of Statistics*, 33(2), 730–773.

JAMES, W., AND C. STEIN (1961): "Estimation with Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 361–379, Berkeley, Calif. University of California Press.

JAROCIŃSKI, M., AND A. MARCET (2019): "Priors about Observables in Vector Autoregressions," *Journal of Econometrics*, 209(2), 238–255.

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.

JEFFREYS, H. (1961): *Theory of Probability.* Oxford: Calendron Press.

JEGADEESH, N. (1990): "Evidence of predictable behavior of security returns," *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," *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.

KAN, R., AND C. ZHANG (1999a): "GMM tests of stochastic discount factor models with useless factors," *Journal of Financial Economics*, 54(1), 103–127.

——— (1999b): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54(1), 203–235.

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

KASS, R. E., AND A. E. RAFTERY (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773–795.

KELLY, B. T., S. PRUITT, AND Y. SU (2019): "Characteristics are covariances: A unified model of risk and return," *Journal of Financial Economics*, 134(3), 501–524.

KLEIBERGEN, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149(2), 149–173.

KLEIBERGEN, F., AND Z. ZHAN (2015): "Unexplained Factors and Their Effects on Second Pass R-squared's," *Journal of Econometrics*, 189(1), 101–116.

——— (2020): "Robust Inference for Consumption-Based Asset Pricing," *Journal of Finance*, 75(1), 507–550.

KOZAK, S., S. NAGEL, AND S. SANTOSH (2018): "Interpreting factor models," *Journal of Finance*, 73(3), 1183–1223.

——— (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135(2), 271–292.

KRAFT, H., E. SCHWARTZ, AND F. WEISS (2018): "Growth options and firm valuation," *European Financial Management*, 24(2), 209–238.

LANCASTER, T. (2004): *Introduction to Modern Bayesian Econometrics.* Wiley and Sons.

LETTAU, M., AND S. LUDVIGSON (2001a): "Consumption, aggregate wealth, and expected stock returns," *Journal of Finance*, 56(3), 815–849.

———— (2001b): "Resurrecting the (C) CAPM: A cross-sectional test when risk premia are time-varying," *Journal of Political Economy*, 109(6), 1238–1287.

Lettau, M., and M. Pelger (2020a): "Estimating latent asset-pricing factors," *Journal of Econometrics*, 218(1), 1–31.

———— (2020b): "Factors that fit the time series and cross-section of stock returns," *Review of Financial Studies*, 33(5), 2274–2325.

Lewellen, J., S. Nagel, and J. Shanken (2010): "A skeptical appraisal of asset pricing tests," *Journal of Financial economics*, 96(2), 175–194.

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.

Liew, J., and M. Vassalou (2000): "Can book-to-market, size and momentum be risk factors that predict economic growth?," *Journal of Financial Economics*, 57(2), 221–245.

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

Linnainmaa, J. T., and M. R. Roberts (2018): "The History of the Cross-Section of Stock Returns," *Review of Financial Studies*, 31(7), 2606–2649.

Lintner, J. (1965): "Security prices, risk, and maximal gains from diversification," *Journal of Finance*, 20(4), 587–615.

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

Lucas, R. E. (1978): "Asset prices in an exchange economy," *Econometrica*, pp. 1429–1445.

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

Madigan, D., and A. E. Raftery (1994): "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89(428), 1535–1546.

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

Markowitz, H. (1952): "Portfolio Selection," *Journal of Finance*, 7(1), 77–91.

Martin, I. W., and S. Nagel (2021): "Market efficiency in the age of big data," *Journal of Financial Economics*.

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

Merton, R. C. (1973): "An intertemporal capital asset pricing model," *Econometrica*, pp. 867–887.

Neuhierl, A., and R. T. Varneskov (2021): "Frequency dependent risk," *Journal of Financial Economics*, 140(2), 644–675.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F.Engle, and D. McFadden, vol. 4. Elsevier Press.

NEWEY, W. K., AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix," *Econometrica*, 55(3), 703–708.

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

ONATSKI, A. (2012): "Asymptotics of the principal components estimator of large factor models with weakly influential factors," *Journal of Econometrics*, 168(2), 244–258.

PARKER, J. A., AND C. JULLIARD (2005): "Consumption risk and the cross section of expected returns," *Journal of Political Economy*, 113(1), 185–222.

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

PÁSTOR, L., AND R. F. STAMBAUGH (2000): "Comparing asset pricing models: an investment perspective," *Journal of Financial Economics*, 56(3), 335–381.

PÁSTOR, L., AND R. F. STAMBAUGH (2002): "Investing in Equity Mutual Funds," *Journal of Financial Economics*, 63, 351–380.

PÁSTOR, L., AND R. F. STAMBAUGH (2003): "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy*, 111(3), 642–685.

PHILLIPS, D. L. (1962): "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind," *J. ACM*, 9(1), 84–97.

POLSON, N. G., AND J. G. SCOTT (2011): "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction," in *Bayesian Statistics 9*, ed. by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West. Oxford University Press.

PRIMICERI, G. E. (2005): "Time-Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

RAFTERY, A. E., D. MADIGAN, AND J. A. HOETING (1997): "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92(437), 179–191.

RAFTERY, A. E., AND Y. ZHENG (2003): "Discussion: Performance of Bayesian Model Averaging," *Journal of the American Statistical Association*, 98, 931–938'.

ROBERTS, H. V. (1965): "Probabilistic Prediction," *Journal of the American Statistical Association*, 60(309), 50–62.

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

RUBINSTEIN, M. (1976): "The valuation of uncertain income streams and the pricing of options," *Bell Journal of Economics*, pp. 407–425.

SCHERVISH, M. J. (1995): *Theory of Statistics*, Springer Series in Statistics. Springer-Verlag.

SEGAL, G., I. SHALIASTOVICH, AND A. YARON (2015): "Good and bad uncertainty: Macroeconomic and financial market implications," *Journal of Financial Economics*, 117(2), 369–397.

SHANKEN, J. (1987): "A Bayesian Approach to Testing Portfolio Efficiency," *Journal of Financial Economics*, 19, 195–215.

SHARPE, W. F. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 19(3), 425–442.

SHILLER, R. C. (2000): "Irrational exuberance," *Philosophy and Public Policy Quarterly*, 20(1), 18–23.

SIMS, C. A. (2007): "Thinking about Instrumental Variables," mimeo.

TIKHONOV, A., A. GONCHARSKY, V. STEPANOV, AND A. G. YAGOLA (1995): *Numerical Methods for the Solution of Ill-Posed Problems*, Mathematics and Its Applications. Springer Netherlands.

TU, J., AND G. ZHOU (2011): "Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies," *Journal of Financial Economics*, 99(1), 204–215.

UPPAL, R., P. ZAFFARONI, AND I. ZVIADADZE (2018): "Correcting Misspecified Stochastic Discount Factors," working paper.

VAYANOS, D. (2004): "Flight to quality, flight to liquidity, and the pricing of risk," Discussion paper, National Bureau of Economic Research.

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.