Learning the Stochastic Discount Factor*

Zhanhui Chen† Yi Ding‡ Yingying Li§ Xinghua Zheng¶

March 6, 2024

Abstract

We develop a statistical framework to learn the high-dimensional stochastic discount factor (SDF) from a large set of characteristic-based portfolios. Specifically, we build on the maximum-Sharpe ratio estimated and sparse regression method proposed in Ao et al. (2019) to construct the SDF portfolio, and develop a statistical inference theory to test the SDF loadings. Applying our approach to 194 characteristic-based portfolios, we find that the SDF constructed by about 20 of them performs well in explaining stock returns.

JEL classification: C55, C58, G11, G12

Keywords: Stochastic discount factor; Factor models; High dimensions; Sparse regressions; Maximum Sharpe ratio regression

*Zhanhui Chen acknowledges the financial support from the Hong Kong Research Grants Council (GRF 16504522, GRF 16503423, and T31-603/21-N). Yi Ding acknowledges the financial support from NSFC of China (72101226), SRG2023-00001-FBA and MYRG-GRG2023-00119-FBA of the University of Macau. Yingying Li acknowledges the financial support from RGC grants GRF 16504223 and SRFS2324-6H02. Xinghua Zheng acknowledges the financial support from RGC grants GRF-16304521 and GRF 16304019.

†Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: chenzhanhui@ust.hk.

‡Faculty of Business Administration, University of Macau, Taipa, Macau. E-mail: yiding@um.edu.mo.

§Department of ISOM and Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: yyli@ust.hk.

¶Department of ISOM, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. E-mail: xhzheng@ust.hk.

1 Introduction

Constructing the stochastic discount factor (SDF) or factor models is critical in asset pricing and factor investing. Decades of empirical studies have identified a fruitful set of factors and factor models. This large "zoo" of factors poses challenges as well as opportunities to better understand the compositions of the SDF. The central question of the SDF learning is, as Cochrane (2011) points out, which characteristics provide truly independent useful information explaining the returns, and hence contribute to the SDF? In this paper, we attempt to address this issue by developing a statistical learning framework to construct the SDF and explore its properties.

Learning the SDF is challenging due to the high-dimensional nature of the problem. There are over hundreds of potential factors, while the number of monthly historical observations are only hundreds. When the number of assets is about the same order of magnitude or even larger than the number of periods, estimating the high-dimensional covariance matrix is non-trivial and creates substantial statistical challenges. In addition, getting an accurate estimation of the expected returns is difficult because we need a long period to estimate the means but factors might have structural changes over time.

Traditionally, factors and test assets are often constructed by portfolio sorts (e.g. based on some firm characteristics), and then factor models are tested against a set of test assets. This approach suffers from two limitations. First, it is limited by the dimensionality of characteristics and often can't fully address the interaction among different characteristics.[1] Second, the choice of test assets is less explored, which influences the empirical evaluation. Especially, when weak factors or even irrelevant factors which test assets do not have strong exposure to are present, the empirical tests may be biased (Kan and Zhang, 1999; Kleibergen, 2009; Bryzgalova, 2017; Gospodinov et al., 2017; Anatolyev and Mikusheva, 2022; Giglio et al., 2022).[2]

To avoid the limitations of the above approach, in this paper, we follow the right-hand-side approach suggested in Barillas and Shanken (2017) and Fama and French (2018). Barillas and Shanken (2017) and Fama and French (2018) show that it is sufficient to examine the squared Sharpe ratio (which is equivalent to the Hansen and Jagannathan (1997) distance) for models with tradable factors, while the choice of test assets is irrelevant. That is, the model with "less mispricing" should have a higher squared Sharpe ratio.[3] Therefore, we can identify the SDF portfolio to be the one with the maximum

[1] Recently, Bryzgalova et al. (2023b) apply decision trees to build portfolios.

[2] For example, many macroeconomic factors appear to be weak factors and test assets may be unable to identify risk premia of these weak factors.

[3] Barillas and Shanken (2018) apply this observation and suggest a Bayesian procedure which computes the probabilities for all models created by a given set of factors. Barillas et al. (2020) apply this metric to empirically compare several prevailing factor models.

squared Sharpe ratio on the mean-variance efficient frontier. Such a SDF portfolio is mean-variance efficient, which explains asset returns. This is also consistent with the no arbitrage condition that if there exists a factor model that prices all assets, the factor portfolios are mean-variance efficient.

Specifically, we build on a large set of potential factor portfolios in Hou et al. (2020a) to construct the SDF portfolio. To tackle the high-dimensional SDF loading estimation problem, we apply the *maximum - Sharpe ratio estimated & sparse regression* (MAXSER) approach proposed by Ao et al. (2019) to select significant inputs for the SDF. MAXSER approach can achieve the mean-variance efficiency asymptotically under the high-dimensional setting. We show that the MAXSER estimator can consistently select the factors with nonzero weights in constructing the SDF. Our setting allows for common factors in the characteristic-based portfolios. With the number of factors significantly reduced by the MAXSER selection, we apply the plug-in method to estimate the optimal weights on the selected portfolios. We show both consistency and asymptotic normality of such estimated weights, which enable us to conduct post inference of the SDF loadings.

We apply our methodology to learning the SDF loadings on the monthly returns of 188 anomaly portfolios from Hou et al. (2020a).[4] We also include the widely used Fama-French six factors (Fama and French, 2018), so there are totally 194 factor portfolios. We find that our proposed SDF performs well in achieving a high Sharpe ratio and explaining the cross-section of expected returns of various portfolios. After applying the post selection of the SDF loadings based on our inference theory, we find that the SDF can be constructed by about 20 characteristic-based portfolios. This complements the finding in Bryzgalova et al. (2023a) that a large set of factors (23 to 25 factors) is necessary to fully capture the SDF. We find that important variables are mostly related to momentum and earnings, including the customer momentum, the cumulative abnormal stock return, the earnings predictability, the market portfolio, and the quarterly earnings-to-price. Except the market factor, the Fama-French factors are not significant given the set of selected factors in our estimated SDF.

Next, we test our estimated SDF against benchmark factor models such as CAPM, the Fama-French three-factor model (Fama and French, 1992, hereafter, FF3), the Fama-French five-factor model (Fama and French, 2015, hereafter, FF5), the Fama-French six-factor model (Fama and French, 2018, hereafter, FF6), Q4/Q5 models (Hou et al., 2015, 2021), the six-factor model in Barillas and Shanken (2018, hereafter, BS6), and behavioral models proposed in Stambaugh and Yuan (2017, hereafter, SY4) and Daniel et al. (2020,

[4]One might start with factors constructed from individual stocks by using firm characteristics. For example, Kozak and Nagel (2023) show that these could span SDF if the covariance of stock returns satisfies a specific structure. They argue that such requirement is more likely satisfied if a large number of firm characteristics are used.

hereafter, DHS3). We also include the SDF estimator of Kozak et al. (2020, hereafter, KNS) as a benchmark case. The test results show that our SDF significantly outperforms the benchmark factor models both statistically and economically. For example, our SDF captures the 188 anomaly portfolios and other prevailing factors. Our results are robust after taking into account the transaction costs.

Although we use anomaly portfolios to construct the SDF, it is important to note that this is not equivalent to using the most significant anomalies. It is possible that a particular factor has a zero alpha relative to some factor models but has a significant loading in our SDF because it helps to reduce the overall risk and achieve a higher overall Sharpe ratio due to the return dependence in characteristic-based portfolios. Therefore, it is important to consider the interactions among characteristics, as emphasized in Bryzgalova et al. (2023b). That means, to construct the SDF, purely chasing factors that have significant alphas, as some studies do, is unnecessary.

In addition to directly using the characteristic-based portfolios, we also construct the SDF based on the principal component (PC) space. Kozak et al. (2020) document that the SDF is sparse in the sense that it can be spanned by a small number of principal components in the characteristic-based portfolios. That is, a factor model with a few statistical factors using the principal components performs similarly to reduced-form models. We estimate the top 20 PCs as in Kozak et al. (2020). First, we find that none of the 20 PCs has a statistically significant alpha relative to our SDF. Second, we further add these 20 PCs to the input set and re-estimate the SDF from totally 214 portfolios. We find that the 20 PCs do not help to improve the SDF. In the SDF estimated with the 214 portfolios, the 20 PCs are rarely selected and have very low weights even when selected. In addition, when we project all characteristic-based portfolios onto the PC space and re-estimate the SDF using the projected portfolio returns, we find that this also doesn't improve our SDF. These results suggest that, different from Kozak et al. (2020), the cross-section of returns can not be adequately explained by a small number of PCs. Our results echo Bryzgalova et al. (2023a) who find that the SDF is dense in the PC space and a characteristic-based factor model using a few PCs is unlikely to perform well.

Our paper is closely related to the emerging literature on applying machine learning tools to identify factors and construct factor models.[5] Some papers apply machine learning technique to estimate factor loadings. For example, Kelly et al. (2019) develop an instrument principal component analysis (PCA) method to estimate loadings of latent factor models as linear functions of characteristics. Lettau and Pelger (2020b,a) propose

[5]Various machine learning approaches have been employed to predict asset returns. For example, Freyberger et al. (2020) use the adaptive group Lasso to select characteristics that provide incremental information for the cross-section of expected returns. Gu et al. (2020) use a variety of machine learning techniques to predict stock returns using characteristics.

a risk-premium PCA to estimate the SDF at the presence of weak factors. Kozak et al. (2020) impose economically motivated prior about SDF coefficients and apply shrinkage and selection methods to construct SDF. Gu et al. (2021) propose a nonlinear latent conditional factor model based on asset characteristics learned with autoencoder neural networks. Giglio et al. (2022) suggest using a supervised PCA to select test assets and estimate risk premia for potentially weak factors. Meanwhile, some papers develop empirical tests. For example, Feng et al. (2020) propose a test for the significance of new factors under the existence of a factor model. Giglio et al. (2021) study multiple hypothesis testing problems in linear asset pricing models. We contribute to the literature in two ways. First, earlier studies mostly focus on individual factors while we consider the mean-variance efficient SDF portfolio. Second, we establish the statistical properties of the SDF loading estimators in the high-dimensional setting, show the consistency of our estimators and develop a statistical inference theory of these estimators, which is lack in the earlier studies.

Our study also relates to the empirical literature on identifying and interpreting characteristic-based portfolios. Numerous studies examine whether these portfolios are significant, robust, and replicable. For example, Hou et al. (2020a) investigate over hundreds of anomalies and find that half of them fail to replicate with more recent data or a different portfolio construction scheme than the original paper. The factor zoo raises concerns about data mining, transaction costs, and replication (McLean and Pontiff, 2016; Harvey et al., 2016; Andrews and Kasy, 2019; Jacobs and Müller, 2020; Chen, 2021; Harvey and Liu, 2021a,b; Jensen et al., 2023). Different from these papers, our paper uses these portfolios to construct the pricing kernel, which may not be anomaly portfolios. In fact, we find the most important characteristic-based portfolios in our SDF do not have significant alphas relative to other factor models.

Last, our paper contributes to the literature of high-dimensional mean-variance efficient portfolio optimization. It is well known that using sample plug-in portfolio performs poorly in constructing mean-variance optimal portfolios when the number of assets is large due to the sampling error in high-dimensions (see, e.g., Michaud, 1989; Britten-Jones, 1999). Various approaches have been proposed to improve the performance of large portfolios. One way is to adopt better mean and covariance matrix estimators, e.g., the Black-Litterman's approach of expected return estimation (Black and Litterman, 1991), factor-model-based covariance matrix estimators (Fan et al., 2008, 2011, 2013), and shrinkage estimators (Ledoit and Wolf, 2017). The MAXSER proposed in Ao et al. (2019) can simultaneously achieve the optimal Sharpe ratio and the target risk constraints. This paper builds on Ao et al. (2019) and establishes statistical inference theory for the SDF weights. The inference theory can be used to further reduce the number of assets in the

final SDF portfolio.

The rest of this paper is organized as follows. We present our methodology supported with statistical theory in Section 2. Section 3 presents main empirical results. Section 4 considers the robustness of our approach, while Section 5 further explores the economic interpretations of our estimated SDF. Finally, we conclude in Section 6. The proofs are collected in Appendix A and additional empirical results are presented in Appendix B.

2 Model

2.1 SDF and the mean-variance efficient portfolio

Under the beta-pricing models, factors ($\mathbf{R}_t$) can be constructed using characteristic-based stock portfolios, i.e., $\mathbf{R}_t = \mathbf{Z}_{t-1}^T \mathbf{R}_{s,t}$, where $\mathbf{R}_t$ is an $N \times 1$ vector of characteristic sorted factors, $\mathbf{Z}_{t-1}$ are $N_s \times N$ asset characteristics, $\mathbf{R}_{s,t}$ are $N_s \times 1$ excess stock returns. The stochastic discount factor (SDF) in this beta-pricing model can be written as

$$\begin{aligned} M_t &= \frac{1 - \mathbf{b}^T(\mathbf{R}_t - E(\mathbf{R}_t))}{R_{f,t}}, \\ E(M_t \mathbf{R}_t) &= 0, \end{aligned} \tag{2.1}$$

where $\mathbf{b}$ is the loading of the SDF on the characteristic sorted portfolio, and $R_{f,t}$ is the risk-free rate. Solving (2.1), we get

$$\mathbf{b} = \mathbf{\Sigma}^{-1} \boldsymbol{\mu}, \tag{2.2}$$

where $\mathbf{\Sigma}$ is the covariance matrix and $\boldsymbol{\mu}$ is the mean return of factor returns $\mathbf{R}_t$.

Note that finding the SDF loadings $\mathbf{b}$ is equivalent to solving the weights of the optimal mean-variance portfolio using the characteristics based factors. Specifically, the mean-variance optimal portfolio solves

$$\max \boldsymbol{\mu}^T \boldsymbol{\omega}, \text{ subject to } \boldsymbol{\omega}^T \mathbf{\Sigma} \boldsymbol{\omega} \leq \sigma^2,$$

where σ is the target risk constraint.[6] The optimal weight is

$$\boldsymbol{\omega}^* = \frac{\sigma}{\sqrt{\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1} \boldsymbol{\mu}}} \mathbf{\Sigma}^{-1} \boldsymbol{\mu}. \tag{2.3}$$

Comparing (2.3) with (2.2), we see that the SDF loading $\mathbf{b}$ differs from the optimal portfolio weights $\boldsymbol{\omega}^*$ up to a scalar, which is invariant to the composition of the SDF.

[6] Note that the optimization problem considers the investment on risky assets so the weights do not sum up to one. The investment weight on the risk-free asset is $1 - \sum_i w_i$.

Motivated by such an observation, we will consider the SDF as a mean-variance portfolio optimization problem.

Given the high dimension of risky assets, it is challenging to estimate SDF when the number of characteristic-sorted portfolios, N, is large. In the next subsections, we develop a selection approach to estimate SDF. We also explore a central limit theorem of the estimated SDF loadings to establish the statistical inference theory.

2.2 Estimating SDF with MAXSER

Given a large number of characteristic-based portfolios, we allow some factors to be portfolios of other characteristic-based factors, e.g., the market factor. That is, the characteristic-based portfolios may not be orthogonal as they might share some common factor components. Suppose that the return generating process of the characteristic-based portfolios is as follows:

$$\mathbf{R}_t = \boldsymbol{\beta}\mathbf{f}_t + \mathbf{U}_t, \tag{2.4}$$

where $\mathbf{f}_t$ is a $K \times 1$ vector of common factors, $\boldsymbol{\beta}$ is an $N \times K$ factor loadings, and $\mathbf{U}_t$ is the idiosyncratic return. We will use $(\mathbf{f}_t, \mathbf{R}_t)$ to construct the SDF. When the common factor components are reasonably presented, they explain much of the variations in $\mathbf{R}_t$. Hence, it is reasonable to assume that the factor $\mathbf{f}_t$ is likely to be included in the SDF and the SDF loadings on the idiosyncratic component $\mathbf{U}_t$ is likely to be sparse after excluding the common factor components.

Next, we follow Ao et al. (2019) and adopt the maximum-Sharpe-ratio estimated and sparse regression (MAXSER) approach to estimate the SDF loadings. The MAXSER approach combines mean-variance portfolio optimization with Lasso regression to find the optimal sparse portfolio. MAXSER finds the mean-variance efficient portfolios by maximizing the Sharpe ratio directly, instead of relying on the estimates of mean and covariance matrix separately. Ao et al. (2019) argue that MAXSER enjoys desirable theoretical properties in terms of portfolio risk control and expected return maximization under the high-dimensional setting when both the number of assets and the observations are large.

We summarize the MAXSER approach as follows. Using both $\mathbf{f}_t$ and $\mathbf{R}_t$ under the model (2.4) to construct the optimal mean-variance portfolio, we denote the optimal weights by $\boldsymbol{\omega}^* = ((\boldsymbol{\omega}_f^*)^T, (\boldsymbol{\omega}_R^*)^T)^T$. Proposition 3 of Ao et al. (2019) shows that

$$\begin{aligned}
\boldsymbol{\omega}_f^* &= \frac{\sigma}{\sqrt{\theta_{all}}}\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}_f - \frac{\sigma}{\sqrt{\theta_{all}}}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}, \\
\boldsymbol{\omega}_R^* &= \frac{\sigma}{\sqrt{\theta_{all}}}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha},
\end{aligned}$$

where $\mathbf{\Sigma}_f$ and $\boldsymbol{\mu}_f$ are the covariance matrix and the mean of $\mathbf{f}_t$; $\mathbf{\Sigma}_u =: (\sigma_{u,ij})_{1\le i,j\le N}$ and $\boldsymbol{\alpha}$ are the covariance matrix and the mean of $\mathbf{U}_t$. Also, $\theta_{all} = \boldsymbol{\mu}_f^T \mathbf{\Sigma}_f^{-1} \boldsymbol{\mu}_f + \boldsymbol{\alpha}^T \mathbf{\Sigma}_u^{-1} \boldsymbol{\alpha}$ is the squared Sharpe ratio of the optimal portfolio investing in factors and all assets.

Similarly, the optimal portfolio invested in the idiosyncratic components $\mathbf{U}$ with a risk constraint being one is

$$\boldsymbol{\omega}_u^* = \frac{1}{\sqrt{\theta_u}} \mathbf{\Sigma}_u^{-1} \boldsymbol{\alpha},$$

where $\theta_u = \boldsymbol{\alpha}^T \mathbf{\Sigma}_u^{-1} \boldsymbol{\alpha}$ is the squared Sharpe ratio of the optimal portfolio investing in idiosyncratic components. We see that $\boldsymbol{\omega}_R^*$ is proportional to $\boldsymbol{\omega}_u^*$. Therefore, we first obtain the optimal weights on idiosyncratic components, $\widehat{\boldsymbol{\omega}}_u^*$, using the following Lasso-type regression

$$\widehat{\boldsymbol{\omega}}_u^* = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^{T} (\widehat{r}_{u,c} - \boldsymbol{\omega}^T \widehat{\mathbf{U}}_t)^2 + \lambda_T \|\boldsymbol{\omega}\|_1, \tag{2.5}$$

where $\|\mathbf{x}\|_1 := \sum |x_i|$ for any vector $\mathbf{x} = (x_i)$. In the above regression, $\widehat{\mathbf{U}}_t = \mathbf{R}_t - \widehat{\boldsymbol{\beta}} \mathbf{f}_t$ are the estimated idiosyncratic returns, and $\widehat{\boldsymbol{\beta}}$ are the coefficients from regressing $(\mathbf{R}_t)$ over $(\mathbf{f}_t)$:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T} (\mathbf{R}_t - \bar{\mathbf{R}})(\mathbf{f}_t - \bar{\mathbf{f}})^T \right) \left(\sum_{t=1}^{T} (\mathbf{f}_t - \bar{\mathbf{f}})(\mathbf{f}_t - \bar{\mathbf{f}})^T \right)^{-1},$$

where $\bar{\mathbf{f}} = \sum_{t=1}^{T} \mathbf{f}_t / T$ and $\bar{\mathbf{R}} = \sum_{t=1}^{T} \mathbf{R}_t / T$, assuming a sample size of T. The tuning parameter λ_T is chosen by cross-validation with the criteria of minimizing the difference between the risk computed using the validation set and the given risk constraint. The $\widehat{r}_{u,c}$ is a constant defined as follows. Denote $\widehat{\boldsymbol{\mu}}_f$, $\widehat{\mathbf{\Sigma}}_f$ as the sample mean and sample covariance matrix of $\mathbf{f}_t$, while $\widehat{\boldsymbol{\mu}}_{full}$ and $\widehat{\mathbf{\Sigma}}_{full}$ are the sample mean and sample covariance matrix of $(\mathbf{f}_t, \mathbf{R}_t^T)^T$, respectively. We define

$$\widehat{r}_{u,c} = \sqrt{\frac{1}{\widehat{\theta}_u}} + \sqrt{\widehat{\theta}_u},$$

where

$$\widehat{\theta}_u = \widehat{\theta}_{all} - \widehat{\boldsymbol{\mu}}_f^T \widehat{\mathbf{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f, \qquad \text{and}$$
$$\widehat{\theta}_{all} = \frac{(T - N - K - 2) \widehat{\boldsymbol{\mu}}_{full}^T \widehat{\mathbf{\Sigma}}_{full}^{-1} \widehat{\boldsymbol{\mu}}_{full} - N - K}{T}.$$

Finally, the estimated optimal weights on all portfolios are $\widehat{\boldsymbol{\omega}}^* = ((\widehat{\boldsymbol{\omega}}_f^*)^T, (\widehat{\boldsymbol{\omega}}_R^*)^T)^T$ with

$$\widehat{\boldsymbol{\omega}}_f^* = \frac{\sigma}{\sqrt{\widehat{\theta}_{all}}} \widehat{\mathbf{\Sigma}}_f^{-1} \widehat{\boldsymbol{\mu}}_f - \frac{\sigma}{\sqrt{\widehat{\theta}_{all}}} \widehat{\boldsymbol{\beta}}^T \widehat{\boldsymbol{\omega}}_u^*,$$
$$\widehat{\boldsymbol{\omega}}_R^* = \sigma \sqrt{\frac{\widehat{\theta}_u}{\widehat{\theta}_{all}}} \widehat{\boldsymbol{\omega}}_u^*.$$

Ao et al. (2019) focus on the MAXSER portfolio's performance. In particular, they show that the estimated portfolio achieves asymptotically the optimal Sharpe ratio under a risk constraint (see Theorem 2 therein for more details). In this paper, we apply and further develop their approach to estimate the SDF via finding the mean-variance efficient portfolio. Our goal is to find the SDF loading estimators that are economically meaningful, easy to estimate, and enjoy desirable statistical properties. In the next subsections, we validate the consistency of variable selection and develop a statistical inference theory about the SDF estimates.

2.3 Sign consistency of MAXSER weights

We make the following assumptions.

Assumption 1 *The number of assets N and the sample size T satisfy that $N, T \to \infty$ and $N/T \to \rho \in (0, 1)$.*

Assumption 2 $\mathbf{f}_t \overset{i.i.d.}{\sim} N(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$. $\mathbf{U}_t$ *is independent of* $\mathbf{f}_t$ *with* $\mathbf{U}_t \overset{i.i.d.}{\sim} N(\boldsymbol{\alpha}, \boldsymbol{\Sigma}_u)$.

Assumption 3 *There exists a constant M such that*

$$\max\left\{\boldsymbol{\alpha}\boldsymbol{\Sigma}_u^{-1}\boldsymbol{\alpha}, \max_{1\le i\le N, 1\le k\le K} |\alpha_i|, |\beta_{ik}|, \sigma_{u,ii}\right\} < M.$$

Denote $S_1 = \{i : \omega_{u,i}^* \neq 0, 1 \le i \le N\}$, $q_u = |S_1|$, $\boldsymbol{\Sigma}_{u,11} = (\boldsymbol{\Sigma}_u)_{S_1,S_1}$, $\boldsymbol{\Sigma}_{u,21} = (\boldsymbol{\Sigma}_u)_{S_1^c,S_1}$, $\boldsymbol{\alpha}_1 = (\alpha_i)_{i\in S_1}$, and $\boldsymbol{\alpha}_2 = (\alpha_i)_{i\in S^c}$. The following notation is used thoughout the paper. For any matrix $\mathbf{A} = (a_{ij})$, its spectral norm is defined as $\|\mathbf{A}\| = \max_{\|\mathbf{x}\|\le 1}\sqrt{\mathbf{x}^T\mathbf{A}\mathbf{x}}$, where $\|\mathbf{x}\| = \sqrt{\sum x_i}$ for any vector $\mathbf{x} = (x_i)$; the max norm is defined as $\|\mathbf{A}\|_{\max} = \max_{i,j}|a_{ij}|$; and the ℓ_1 norm is defined as $\|\mathbf{A}\|_1 = \max_i \sum_j |a_{ij}|$.

Assumption 4 $q_u\sqrt{\log N/N} = o(1)$, $\min_{i\in S_1}|\omega_i^*| \gg \lambda_T\sqrt{q_u}/N$ *with* $\lambda_T \gg q_u\sqrt{\log N}\sqrt{N}$.

Assumption 5 $\|\boldsymbol{\Sigma}_{u,21} + \boldsymbol{\alpha}_2\boldsymbol{\alpha}_1^T\|_{\max} < 1 - \eta$ *for some constant* $\eta > 0$, *and* $\|\boldsymbol{\Sigma}_{u,11}^{-1}\| < M$ *for constant* $M > 0$.

Assumptions 1–3 are from Ao et al. (2019). The sparsity assumption, Assumption 4, is slightly different from Ao et al. (2019). We require the number of nonzero weights to be not too big. Moreover, the minimum nonzero weight is not too small. Assumption 5 includes the irrepresentable condition and the regularity condition on the precision matrix that are widely imposed in the Lasso literature.

Zhao and Yu (2006) establish the sign consistency of the estimated coefficients from the Lasso regressions. It is important to note that our setting is essentially different from the standard Lasso regression. The difference lies in the fact that in the standard Lasso regression, the inference theory is obtained by conditioning on the x-variable and utilizing the i.i.d. assumption imposed on the noise term. In our setting, our inference theory is built upon the randomness of the variable $\widehat{\mathbf{U}}_t$ which is the x-variable. The next theorem establishes the sign consistency of the MAXSER weights in our setting.

Theorem 1 *Under Assumptions 1–5,*

$$P\Big(sign(\widehat{\boldsymbol{\omega}}_R^*) = sign(\boldsymbol{\omega}_R^*)\Big) \to 1. \tag{2.6}$$

Proof: See Appendix A.1.

2.4 Statistical inferences of SDF loadings

Inference of sparse regressions has been explored in the literature. For example, a widely used approach is the debiased low-dimensional linear projection method developed by Zhang and Zhang (2014) for post inference of sparse regressions. However, for the same reason mentioned in the previous subsection, the approach is not applicable to our setting. Another widely used statistical inference approach is to perform post inference based on variable selection from sparse regressions. For example, in the ordinary regression setting, Belloni et al. (2014) develop testing methods for variables estimated from ordinary least squares (OLS) based on post-sparse regressions. In financial applications, Feng et al. (2020) develop factor testing methods that applies OLS based on the post-LASSO model selection. The framework of Feng et al. (2020) follows the ordinary regression setting as they use test asset returns as responses in the regression. The inference results established for the ordinary regression do not apply to our setting either.

In this paper, we develop an alternative approach for statistical inference of the SDF loadings. We will use the variable selection results from the MAXSER estimator for optimal portfolios and conduct a post plug-in estimation of the SDF loadings using Eq. (2.2). We will then make statistical inference based on the plug-in estimator.

Define $S_{full,1} = \{1, ..., K\} \cup \{i + K : i \in S_1\}$. The set $S_{full,1}$ includes all factors and the variables with non-zero SDF loadings. It is straightforward to verify that

$$\boldsymbol{\omega}^* = \Big((\boldsymbol{\omega}_{full,1}^*)^T, \mathbf{0}^T\Big)^T,$$

where $\boldsymbol{\omega}_{full,1}^*$ is the optimal weights of length $K + |S_1|$, which use the factors and the

assets with nonzero weights and the same risk target, given by

$$\boldsymbol{\omega}^*_{full,1} = \frac{\sigma}{\sqrt{\boldsymbol{\mu}^T_{full,1}\boldsymbol{\Sigma}^{-1}_{full,1}\boldsymbol{\mu}_{full,1}}}\boldsymbol{\Sigma}^{-1}_{full,1}\boldsymbol{\mu}_{full,1},$$

where $\boldsymbol{\mu}_{full,1}$, and $\boldsymbol{\Sigma}_{full,1}$ are the mean and the covariance matrix matrix of $(\mathbf{f}_t^T, \mathbf{R}_{S_1}^T)^T$. Note that the SDF weight $\mathbf{b} = \boldsymbol{\Sigma}^{-1}_{full}\boldsymbol{\mu}_{full}$ is proportional to the optimal portfolio weight $\boldsymbol{\omega}^*_{full}$. We have $\mathbf{b} = (\mathbf{b}^T_{full,1}, \mathbf{0}^T)^T$ with $\mathbf{b}_{full,1} = \boldsymbol{\Sigma}^{-1}_{full,1}\boldsymbol{\mu}_{full,1}$.

Denote $\widehat{S}_{full,1} = \{1, ..., K\} \cup \{i + K : (\widehat{\boldsymbol{\omega}}^*_R)_i \neq 0, 1 \leq i \leq N\}$. The sign consistency established in Theorem 1 implies that

$$P(\widehat{S}_{full,1} = S_{full,1}) \to 1.$$

Therefore, we focus on $S_{full,1}$ and perform post inference about $\mathbf{b}_{full,1}$. Specifically, we compute a plug-in estimator $\widehat{\mathbf{b}} = (\widehat{\mathbf{b}}^T_{full,1}, \mathbf{0}^T)^T$ using the assets in $\widehat{S}_{full,1}$:

$$\widehat{\mathbf{b}}_{full,1} = \widehat{\boldsymbol{\Sigma}}^{-1}_{full,1}\widehat{\boldsymbol{\mu}}_{full,1}, \tag{2.7}$$

where $\widehat{\boldsymbol{\Sigma}}_{full,1}$ and $\widehat{\boldsymbol{\mu}}_{full,1}$ are the sample covariance and the sample mean of the asset returns in $\widehat{S}_{full,1}$ that includes the K factors and assets with nonzero weight estimates.

We make the following assumption.

Assumption 6 $\|\boldsymbol{\Sigma}^{-1}_{full,1}\| \leq C$ *for some constant* $C > 0$.

Theorem 2 gives the central limit theorem of $\widehat{\mathbf{b}}_{full,1}$ defined in Eq. (2.7).

Theorem 2 *Under Assumptions 1–6, for any fixed* k *and any deterministic* $k \times (q_u + K)$ *matrix* $\mathbf{A}$, $\|\mathbf{A}\| = O(1)$,*we have*

$$\sqrt{n}\mathbf{A}(\widehat{\mathbf{b}}_{full,1} - \mathbf{b}_{full,1}) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}_A),$$

where $\boldsymbol{\Sigma}_A = \lim_{N\to\infty} \mathbf{A}((\boldsymbol{\mu}^T_{full,1}\boldsymbol{\Sigma}^{-1}_{full,1}\boldsymbol{\mu}_{full,1} + 1)\boldsymbol{\Sigma}^{-1}_{full,1} + \boldsymbol{\Sigma}^{-1}_{full,1}\boldsymbol{\mu}^T_{full,1}\boldsymbol{\mu}_{full,1}\boldsymbol{\Sigma}^{-1}_{full,1})\mathbf{A}^T$.

Proof: See Appendix A.2.

Theorem 2 states that the SDF estimator $\widehat{\mathbf{b}}_{full,1}$ enjoys the asymptotic normality, which allows us to perform post inference of the SDF loadings.

3 Empirical tests

3.1 Data

We build our work on the fruitful literature on return predictability and anomalies, i.e., using the characteristic-based portfolio data from Hou et al. (2020a). The data set includes monthly returns of 188 anomaly portfolios from January 1980 to December 2021. Hou et al. (2020a) categorize these anomalies into 6 groups, namely, frictions, intangibles, investment, momentum, profitability, and value-growth. We also include the Fama-French six factors (Fama and French (2018)), i.e., the market portfolio, SMB, HML, CMA, RMW and MOM portfolios. Monthly data of these six factors are obtained from French's data library.[7] Totally there are 194 test portfolios with 504 monthly observations.

3.2 SDF learning: Various approaches

We consider various SDF learning approaches, including MAXSER-based approaches and others proposed in the literature.

First, we use the 194 test portfolios to learn the SDF based on the MAXSER method. We note that there exits a factor structure among testing portfolios, as is clearly shown in Figure 1. The market portfolio is correlated with other test portfolios with a median magnitude of 0.15 and interquarter range of 0.07 and 0.26. For our proposed MAXSER method, we choose the market portfolio as the base factor for the remaining factors. Because the SDF learning is equivalent to computing the mean-variance optimal portfolio, the proposed MAXSER approach developed in Section 2.2 can consistently select the variables from a large number of candidate factors to construct the SDF, as shown in Theorem 1. The portfolio is estimated every year using a rolling window of the past 20 years, that is $T = 240$. The target risk σ is set to be the market risk during the training period. The out-of-sample period is between 2000 and 2021. We denote the portfolio by MAXSER.

[7] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

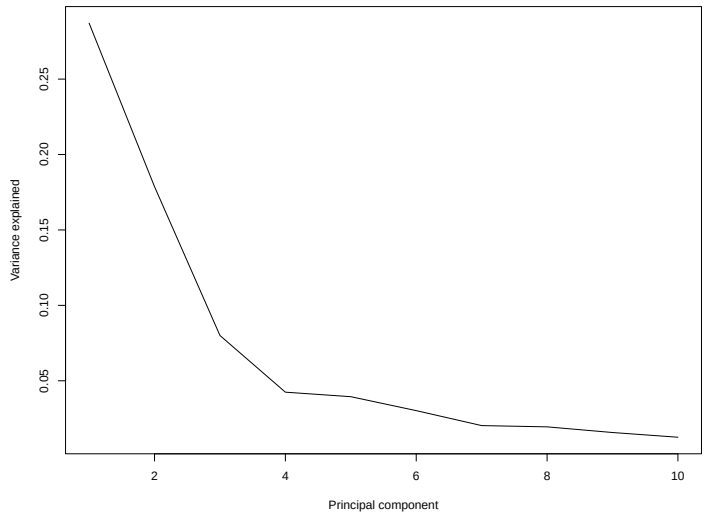

Figure 1: **Scree plot of PCA over 194 test portfolios.**

Because many anomaly portfolios are highly co-linear, to facilitate the learning, in each training period, we pre-screen the portfolios and exclude the ones that have high multi-colinearity with other portfolios. Specifically, we sequentially remove portfolios, one at a time, that has a variance inflation factor (VIF) higher than 10 and is the highest among all remaining portfolios. After the pre-screening, the VIFs of all portfolios in the training set are lower than 10.

Second, the sparsity of the estimated parameters from the Lasso regression depends on the tuning parameter λ_T. In a finite sample, it is well known that Lasso estimation tends to overselect non-signals with a tuning parameter λ chosen from cross-validation using criteria to achieve prediction accuracy (see, e.g., Leng et al. (2006)). In addition, for the model selection purpose, Leng et al. (2006) suggest using statistical tests to perform post variable selection. This motivates us to do a post screening based on the SDF weight inference theory established in Section 2.4. We perform a backward variable selection procedure as follows. Set M_0 to be the set of factors selected by MAXSER. For $j = 0, ...,$ within the active set M_j, if $|M_j| > N_s$ for a pre-specified size N_s, we perform statistical inference on each variable in M_j and get the p-values as

$$p_i^{(M_j)} = 2\left(1 - F\left(\frac{|\widehat{b}_i^{(M_j)}|}{\widehat{sd}(\widehat{b}_i^{(M_j)})}\right)\right),$$

where $F(\cdot)$ is the cumulative distribution function of a standard normal distribution, $\widehat{b}_i^{(M_j)}$ is the plugin loading estimator of the ith variable from (2.7) for $i \in M_j$, and $\widehat{sd}(\widehat{b}_i^{(M_j)})$ is the standard deviation estimator from Theorem 2. We then remove one variable from

M_j with the highest p-value and get M_{j+1}. This inference process is valid because we show that the post-MAXSER plugin SDF estimator is asymptotically jointly normal in Theorem 2. We continue this process until $|M_k| \leq N_s$ for some k. Finally, the resulting SDF portfolio is estimated using the plug-in method based on the resulting post-selected set with at most N_s selected variables for a pre-specified N_s. We denote this post-screened portfolio as MAXSER-S(N_s). We consider $N_s = 5, 10, 20, 30$, and 40.

Third, we compare our SDF portfolio with the shrinkage SDF estimator in Kozak et al. (2020), denoted by KNS. Specifically, KNS considers the SDF loading estimates that solve the following problem:

$$\widehat{\mathbf{b}}^{KNS} = \underset{\mathbf{b}}{\text{argmin}}(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\Sigma}}\mathbf{b})^T\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\Sigma}}\mathbf{b}) + \gamma_1 \sum_{i=1}^{N} |\mathbf{b}_i| + \gamma_2 \mathbf{b}^T\mathbf{b},$$

where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are sample mean and sample covariance matrix of the training data, γ_1 and γ_2 are tuning parameters. Following Kozak et al. (2020), the tuning parameters are chosen by cross-validation with the criteria of maximizing the out-of-sample (oos) R^2:

$$R^2_{oos} = 1 - \frac{(\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\Sigma}}_2\mathbf{b})^T(\widehat{\boldsymbol{\mu}}_2 - \widehat{\boldsymbol{\Sigma}}_2\mathbf{b})^T}{\widehat{\boldsymbol{\mu}}_2^T\widehat{\boldsymbol{\mu}}_2},$$

where $\widehat{\boldsymbol{\mu}}_2$, and $\widehat{\boldsymbol{\Sigma}}_2$ are the withheld sample in cross-validation. The weight $\widehat{\mathbf{b}}^{KNS}$ is then normalized to make the portfolio risk the same as the target risk.

Fourth, Kozak et al. (2020) find that the SDF can be formed by a few principal components (PC) in characteristic-based portfolios. In order to verify whether the PCs are useful in constructing the SDF, we enlarge the pool of characteristic-based portfolios with their top 20 PCs and estimate the SDF using our proposed approach. Specifically, we use the first 20 years of training period to perform PCA on the 193 portfolios (except the market portfolio) and obtain the principal eigenvectors. We then use the eigenvectors to perform orthogonal transformation and project all returns onto the PC space. Based on the enlarged pool of variables with top 20 PCs, we use the proposed MAXSER approach to estimate SDF loadings. We denote this estimated portfolio as MAXSER(+20PCs). We also consider investing solely on the 20 PCs and construct a plugin estimator, and a plugin estimator that invests in the 20 PCs and the market portfolio. Denote these portfolios by 20PCs and Mkt+20PCs, respectively.

Last, we also include the following benchmark portfolios based on various factor models for comparison.

- CAPM: using the market portfolio.
- FF3/FF5/FF6: constructing a plug-in optimal portfolio using Mkt-Rf, SMB and

HML factors from the Fama-French three-factor model (Fama and French (1992)), the Fama-French five-factor model with CMA and RMW portfolios (Fama and French (2015)), and further including MOM for the Fama-French six-factor model (Fama and French (2018)).

- Q4/Q5: Constructing a plug-in optimal portfolio using Mkt-Rf, R_ME, R_IA and R_ROE factors from the Q4 factor model (Hou et al. (2015)) and additional R_EG factor for the Q5 factor model (Hou et al. (2021))[8].

- BS6: Constructing a plug-in optimal portfolio using Mkt-Rf, SMB, IA, ROE, MOM, HML(m) factors from Barillas and Shanken (2018)[9].

- SY4: Constructing a plug-in optimal using Mkt-Rf, SMB, PERF and MGMT factors from Stambaugh and Yuan (2017)[10].

- DHS3: Constructing a plug-in optimal using Mkt-Rf, PEAD and FIN factors from Daniel et al. (2020)[11].

3.3 Performance of SDF portfolios

We summarize the out-of-sample performance of the SDF portfolios computed from our proposed approach and other benchmark models in Table 1. We report the monthly mean, standard deviation, Sharp ratio, maximum drawdown, together with different monthly distribution percentiles. In addition, we follow Memmel (2003) and report the p-value of the test for the difference in Sharpe ratios:

$$H_0 : SR_{MAXSER-S(20)} \leq SR_0 \qquad \text{vs.} \qquad H_1 : SR_{MAXSER-S(20)} > SR_0, \tag{3.1}$$

where $SR_{MAXSER-S(20)}$ is the Sharpe ratio of MAXSER-S(20) and SR_0 denotes the Sharpe ratio of another approach. The cumulative returns of the mean-variance optimal portfolios are plotted in Figure 2.

Table 1 summarizes the out-of-sample performances of the SDF portfolios (the mean-variance optimal portfolios). We see from Panel A of Table 1 that the MAXSER portfolio without post-screening achieves a monthly Sharpe ratio of 0.46, which is the highest among all approaches considered. Among the MAXSER portfolios with post screening, we see that as the number of assets included in the SDF portfolio increases from 5 to 20, the post-screened SDF portfolios' monthly Sharpe ratio increases from 0.29 to 0.42. When

[8]Monthly data of the q-factors are from https://global-q.org/factors.html.
[9]HML(m) data are from https://pages.stern.nyu.edu/~afrazzin/data_library.htm.
[10]Monthly data of SY4 factors are from https://finance.wharton.upenn.edu/~stambaug/.
[11]Monthly data of DHS factors are from https://sites.google.com/view/linsunhome.

the number of assets included into the SDF portfolio grows to be greater than 20, its performance deteriorates. Compared with the MAXSER portfolio without post-screening, the MAXSER-S(20) successfully reduce the number of assets used from about 50 to 20 while achieving a good performance, i.e., its Sharpe ratio is statistically indifferent to the case without post screening (MAXSER). The results demonstrate the advantage of utilizing the statistical inference theory for the SDF loadings established in Theorem 2. The formal statistical tests show that the difference in Sharpe ratios from various MAXSER-based portfolios is mostly statistically insignificant, except MAXSER-S(5) which has a low Sharpe ratio. For example, MAXSER and MAXSER-S(20) is not statistically significant. Therefore, balancing between the performance and the number of variables used, we will focus on the MAXSER-S(20), which is our main SDF later.

Panel B of Table 1 shows that the MAXSER-S(20) significantly outperforms most of the benchmark cases, except 20 PCs, Mkt+20PCs, Q5, and DHS3. For example, although KNS includes about 50 variables on average in the optimal portfolios, it has a lower Sharpe ratio of 0.27. Among all the benchmark portfolios in Panel B, the Q5 portfolio has the highest monthly Sharpe Ratio of 0.36, which is still smaller than that of MAXSER-S(20).

Turning to the performance of portfolios using PCs in characteristic-based portfolios, we see that including PCs does not improve the performance of the SDF portfolio. The Sharpe ratios of the portfolios that invest solely on 20 PCs and the market with 20 PCs (Mkt+20PCs), and the MAXSER portfolio with an enlarged pool includings PCs (MAXSER(+20PCs)) are smaller than that of MAXSER-S(20), which invests in characteristic-based portfolios directly.

Table 1: **Out-of-sample performances of SDF.** This table reports the out-of-sample performances of the optimal portfolios constructed from the proposed approach and other benchmark models. The summary statistics include mean return (MEAN), standard deviation (SD), Sharpe ratio (SR), the maximum drawdown (MDD), and the distribution percentiles. All are reported in percentage except the Sharpe ratio. The column "SR-pvalue" reports the p-value of testing the difference in the Sharpe ratios between the proposed MAXSER-S(20) and other portfolios. *, ** and *** indicate statistical significance for testing (3.1) at the level of 5%, 1% and 0.1%, respectively. The optimal portfolios are estimated every year, using a rolling window of the past 20 years, and portfolio weights are rebalanced every month. The out-of-sample period is between 2000 and 2021. We compare the MAXSER without post screening, MAXSER with various post screening levels (5,..., 40), MAXSER based on the portfolio pool enlarged with 20 PCs (MAXSER(+20PCs)), the shrinkage SDF portfolio (KNS) by Kozak et al. (2020), the plugin portfolio that invest in 20 PCs or Market+20 PCs, and the SDF portfolio constructed from various factor models.

Model	Mean	SD	SR	SR-pvalue	MDD	P1	P5	P10	P25	P50	P75	P90	P95	P99
Panel A: MAXSER-based SDF														
MAXSER	2.92	4.95	0.46	0.94	12.76	-8.28	-5.14	-2.95	-0.53	2.19	4.41	7.56	10.81	17.47
MAXSER-S(5)	1.61	5.59	0.29	0.02*	39.73	-12.84	-7.61	-5.01	-1.22	1.36	4.73	8.03	9.90	16.91
MAXSER-S(10)	2.20	5.55	0.40	0.29	18.32	-9.98	-5.96	-4.46	-1.32	1.93	5.03	9.38	10.52	19.17
MAXSER-S(20)	2.43	5.79	0.42	–	23.33	-10.38	-6.77	-4.22	-0.83	2.16	5.48	9.24	13.02	18.55
MAXSER-S(30)	2.52	6.16	0.41	0.33	18.80	-11.76	-6.43	-4.57	-1.44	2.38	5.54	9.54	14.08	19.75
MAXSER-S(40)	2.59	6.49	0.40	0.24	19.66	-12.44	-6.70	-4.34	-1.23	2.36	5.49	10.05	14.07	21.86
MAXSER(+20PCs)	2.14	5.20	0.41	0.42	20.26	-9.82	-5.43	-3.36	-0.82	1.86	4.61	7.53	11.08	17.75
Panel B: SDF from other benchmark models														
20PCs	1.72	5.07	0.34	0.11	27.32	-8.03	-5.40	-3.50	-1.85	1.41	3.85	8.09	11.87	17.85
Mkt+20PCs	2.07	5.27	0.39	0.33	18.59	-9.95	-5.12	-3.88	-1.08	1.73	4.56	9.03	11.53	17.52
KNS	1.34	4.91	0.27	0.00**	19.63	-10.82	-6.46	-3.62	-1.13	1.27	3.62	6.07	9.51	16.50
CAPM	0.61	4.51	0.13	0.00***	49.39	-10.49	-7.89	-5.77	-1.96	1.19	3.23	5.63	7.46	10.61
FF3	0.66	5.86	0.11	0.00***	64.68	-18.79	-8.72	-6.66	-1.85	1.02	3.78	6.66	9.42	13.83
FF5	1.87	6.53	0.29	0.03*	33.60	-11.98	-6.95	-4.99	-1.77	1.87	5.05	8.53	13.15	20.22
FF6	1.81	6.45	0.28	0.02*	36.74	-11.67	-6.96	-5.41	-1.60	1.52	4.73	8.67	13.59	18.51
Q4	1.48	5.67	0.26	0.01*	34.78	-10.34	-7.65	-5.13	-1.56	1.54	4.27	7.28	11.19	17.23
Q5	1.98	5.48	0.36	0.19	24.62	-11.67	-6.34	-3.37	-1.15	1.54	4.76	8.02	10.64	16.80
BS6	1.59	6.31	0.25	0.01**	41.48	-12.89	-8.63	-5.35	-1.32	1.76	4.00	8.21	12.10	21.67
SY4	1.76	6.03	0.29	0.03*	28.22	-13.85	-7.48	-4.33	-1.21	1.79	4.41	7.00	10.41	19.61
DHS3	1.65	5.09	0.32	0.07	38.81	-11.11	-6.51	-4.25	-1.49	1.42	4.56	8.99	10.29	13.26

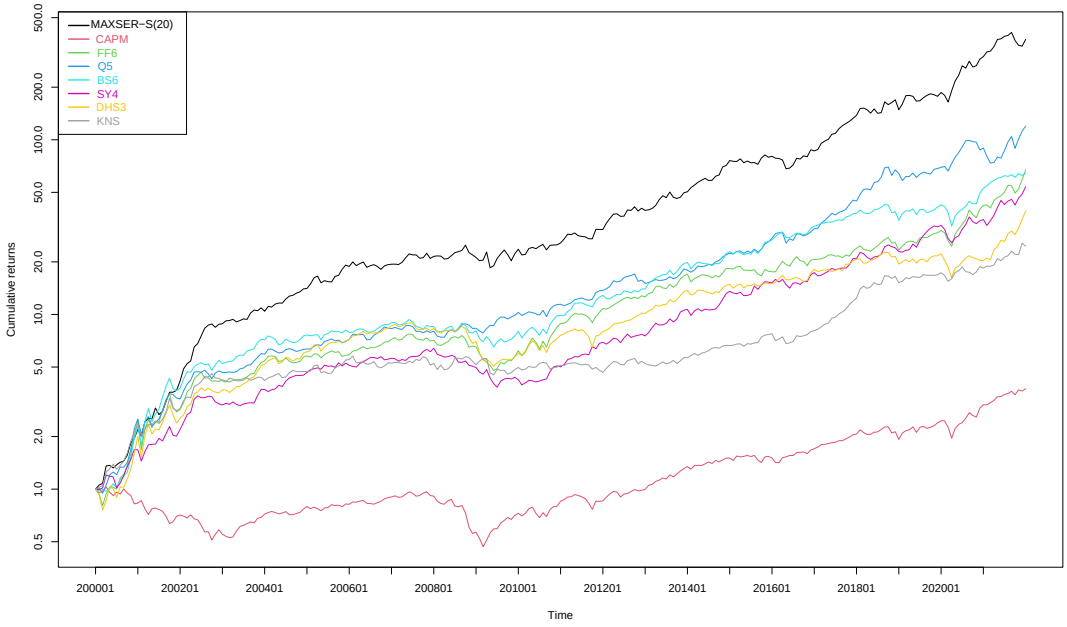

Figure 2: **Cumulative dollar returns of the SDF portfolios.** This figure plots the cumulative returns of various SDF portfolios from 2000 to 2021, starting with \$1. We include the SDF portfolios constructed from the MAXSER-S(20), CAPM, FF6, Q5, BS6, SY4, DHS3, and KNS. The portfolio risk is set to be the market risk.

Table 2: **The number of significant alphas from testing 188 anomalies against various SDFs.** We examine the explanatory power of various SDFs over 188 anomalies. This table summarizes the total number of significant alphas and the number of significant alphas in each characteristic group. The number in parenthesis is the total number of portfolios in each anomaly group. The number of significant alphas under the threshold $t > 1.96$ or $t > 3$ is reported. The evaluation period is between 2000 and 2021.

	All (188)		frictions (10)		intangibles (30)		investment (29)	
threshold of t	1.96	3	1.96	3	1.96	3	1.96	3
MAXSER-S(20)	7	1	0	0	1	1	2	0
20PCs	9	1	0	0	4	0	1	0
Mkt+20PCs	10	2	0	0	5	1	1	0
KNS	13	1	0	0	6	1	1	0
CAPM	53	5	0	0	15	1	5	1
FF3	75	49	4	3	11	9	6	3
FF5	38	16	0	0	11	4	5	1
FF6	37	17	0	0	11	7	5	1
Q4	41	15	1	0	13	4	6	0
Q5	15	2	1	0	7	2	0	0
BS6	44	21	0	0	13	6	6	1
SY4	31	3	0	0	4	1	0	0
DHS3	34	7	1	0	15	3	3	1

	momentum (41)		profitability (46)		value-growth (32)	
threshold of t	1.96	3	1.96	3	1.96	3
MAXSER-S(20)	0	0	4	0	0	0
20PCs	0	0	4	1	0	0
Mkt+20PCs	0	0	4	1	0	0
KNS	0	0	3	0	3	0
CAPM	5	0	24	4	4	1
FF3	13	4	37	30	4	0
FF5	4	0	17	11	1	0
FF6	4	0	16	9	1	0
Q4	3	0	17	11	1	0
Q5	1	0	5	0	1	0
BS6	7	0	16	14	2	0
SY4	10	0	8	1	9	1
DHS3	0	0	14	3	1	0

Next, we performs asset pricing tests in Table 2 to examine the pricing power of various SDFs. Specifically, we regress the 188 anomaly portfolios against various SDFs to examine whether the alpha is significant. Table 2 presents the number of significant alphas. We see

that the number of significant alphas is the lowest under the proposed MAXSER-S(20). There are only 7 anomalies with significant alphas under MAXSER-S(20). On the other hand, all other benchmark portfolios have more rejections. The portfolio that has the second lowest number of rejections is the plugin portfolio on the 20 PCs and the portfolio investing in market and the 20 PCs, which have 9 and 10 rejections, respectively. The KNS portfolio has 13 rejections, while Q5 has 15 rejections. In particular, KNS and Q5 do not perform well in the intangibles group where there are 6 and 7 rejections out of 30 portfolios, respectively, for a threshold $t = 1.96$.

Furthermore, we directly examine the explanatory power of the benchmark models over the SDF portfolio constructed from MAXSER-S(20). Specifically, we regress the out-of-sample return of SDF portfolio from MAXSER-S(20) against the benchmark models. Table 3 presents the alpha and its t-statistic from the regressions.

Table 3: **Testing the MAXSER-S(20) SDF against other benchmark models.** This table reports the regression results of the SDF portfolio from MAXSER-S(20) against other benchmark models, using monthly returns from 2000 to 2021. Alpha in percentage and its corresponding t-statistic are reported.

	Alpha (%)	t-statistic
CAPM	2.33	6.52
FF3	2.28	6.46
FF5	1.67	4.74
FF6	1.60	4.78
Q4	1.62	4.97
Q5	1.26	3.82
BS6	1.69	5.30
SY4	1.33	3.57
DHS3	1.54	4.60
KNS	1.35	5.01

We see from Table 3 that all benchmark models fail to explain our estimated MAXSER-S(20) SDF portfolio; the alpha of the estimated MAXSER-S(20) SDF is both economically large and statistically significant under all benchmark models. Table 3 shows that the monthly alphas are all greater than 1% with a t-statistic higher than 3.5 across all models. For example, the CAPM alpha is 2.33% per month and statistically significant (t-statistic = 6.52). These results suggest that the benchmark models can not capture the expected returns of our MAXSER-S(20) portfolio.

Finally, we switch the roles between benchmark models and our estimated SDF and

examine whether our MAXSER-S(20) SDF is able to explain the prevailing pricing factors. Specifically, we regress various pricing factors against the MAXSER-S(20) SDF. Table 4 reports the alpha and its t-statistic from regressions.

Table 4: **Testing prevailing pricing factors with the MAXSER-S(20) SDF.** This table reports the regression results of various pricing factors against the MAXSER-S(20) SDF, using monthly returns from 2000 to 2021. Alpha in percentage and its corresponding t-statistic are reported.

Factor	Alpha (%)	t-statistic
Mkt-RF	0.35	1.17
SMB	0.22	1.07
HML	-0.10	-0.48
RMW	0.21	1.06
CMA	0.08	0.63
MOM	-0.32	-0.93
R_ME	0.22	1.01
R_IA	-0.04	-0.31
R_ROE	0.01	0.03
R_EG	0.30	1.87
HmLm	0.31	1.11
MGMT	0.13	0.58
PERF	0.20	0.54
PEAD	0.18	1.23
FIN	0.07	0.25
KNS	-0.08	-0.34

We see from Table 4 that all alphas are insignificant at the t-statistic threshold level of 1.96. The highest t-statistic from the prevailing factors is the expected investment growth factor (R_EG) from the Q5 model, which is 1.87. These results suggest that our SDF well captures the prevailing pricing factors.

4 Robustness

4.1 Different numbers of assets: MAXSER-S(20) vs. MAXSER-S(10)

Table 1 shows that MAXSER-S(20) outperforms MAXSER-S(10) in terms of out-of-sample Sharpe ratio. However, the difference is not statistically significant. Therefore, one might wonder if we could reduce the number of inputs to 10. In this subsection, we further

compare the performances of the SDF estimated with 20 post-selected assets (MAXSER-S(20)) and 10 post-selected assets (MAXSER-S(10)). We perform empirical tests with the MAXSER-S(10), which are similar to those with the MAXSER-S(20) SDF. The results are presented in Appendix B. To summarize, we find that although MAXSER-S(10) exhibits good explanatory power for anomaly portfolios and prevailing factors, its empirical performances are worse than that of the MAXSER-S(20) SDF. When testing with the 188 anomaly portfolios, MAXSER-S(10) fails 13 cases using a t-statistic threshold of 1.96, while MAXSER-S(20) only fails 7 cases. When testing with the prevailing pricing factors, MAXSER-S(10) is less powerful. For example, the expected investment growth factor is significant at a t-statistic threshold of 1.96. Therefore, we choose MAXSER-S(20) over MAXSER-S(10) as our estimator of SDF.

4.2 Impacts of transaction costs

Transaction costs affect portfolio performances. In particular, MAXSER adopts Lasso, which helps screen out useless assets by imposing zeros weights, but this could lead to higher turnover and decrease the Sharpe ratio after transaction costs. In this section, we evaluate the performance of our SDF portfolio under transaction costs. We find that the portfolio has a monthly turnover of 80.7% and the absolute monthly weight exposure is 7.9. Hence our SDF portfolio turnover per dollar exposure is 10.22%. Kan et al. (2022) suggests using 20 bps per dollar of transaction when evaluating the post-transaction cost performance of investment in anomaly portfolios. Following Kan et al. (2022), we check the performance of the SDF portfolio after deducting the transaction costs, assuming a cost of 20 bps per dollar of transaction. In addition, we also check the results assuming a higher cost of 40 bps per dollar of transaction. The performance of our SDF estimator after adjusting for the transaction costs is summarized in Table 5. The cumulative returns of the SDF portfolio after transaction costs are plotted in Figure 3.

Table 5: **Performance of the MAXSER-S(20) SDF after adjusting for transaction costs.** The transaction cost is assumed to be 20 bps (40 bps) per dollar of transaction. The portfolios are rebalanced every month from 2000 to 2021. The summary statistics include the monthly mean return (Mean), standard deviation (SD), Sharpe ratio (SR), and the maximum drawdown (MDD) of the SDF portfolio. All are reported in percentage except the Sharpe ratio.

transaction cost	Mean	SD	SR	MDD
20 bps	2.27	5.77	0.39	24.96
40 bps	2.10	5.77	0.36	26.56

We see from Table 5 that, because the turnover is low relative to the weights exposure, the portfolio maintains a high Sharpe ratio after adjusting for the transaction costs. For example, the Sharpe ratio is 0.39 after removing 20 bps transaction cost per dollar, and the Sharpe ratio is 0.36 when the cost is 40 bps/dollar. Similarly, Figure 3 shows that for \$1 initial investment in 2000, at the end of 2021, the cumulative return is \$375.13 when there is no transaction costs, while it is \$244.58 (\$159.00) after adjusting for a transaction cost of 20 bps (40 bps) per dollar.

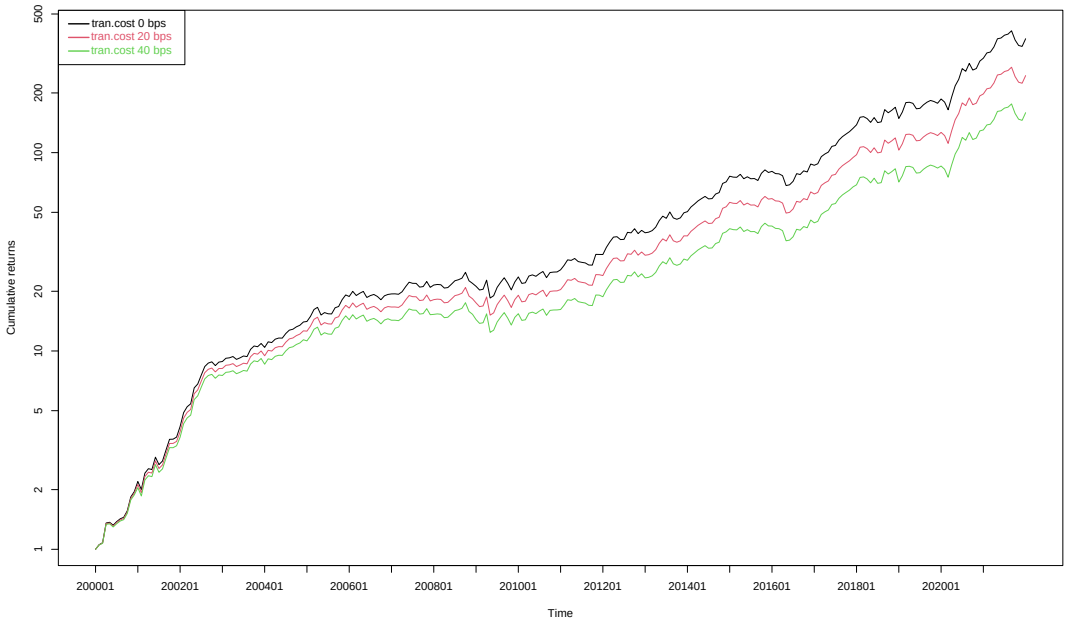

Figure 3: **Cumulative dollar returns of the MAXSER-S(20) SDF portfolio after adjusting for transaction costs.** The initial investment is \$1 in 2000. Transaction costs are assumed to be 20 bps or 40 bps between 2000 and 2021. The portfolios are rebalanced every month.

Next, we repeat Table 3, adjusting for the transaction cost. That is, we regress the MAXSER-S(20) SDF portfolio against various factor models in Table 6. We see that again, our MAXSER-S(20) SDF portfolio has a statistically significant alpha against other models.

Table 6: **Testing the MAXSER-S(20) SDF portfolio against various factor models.** This table presents the regression results of the MAXSER-S(20) SDF portfolio against various factor models, using the monthly returns from 2000 to 2021. The transaction costs are assumed to be 20 bps or 40 bps per dollar of transaction. The portfolios are rebalanced every month. We report the alpha in percentage and t-statistic.

Transaction cost	20 bps		40 bps	
Model	Alpha (%)	t-statistic	Alpha (%)	t-statistic
CAPM	2.16	6.08	2.00	5.61
FF3	2.12	6.01	1.96	5.55
FF5	1.52	4.29	1.35	3.82
FF6	1.44	4.30	1.28	3.81
Q5	1.09	3.33	0.93	2.83
Q4	1.46	4.49	1.30	3.99
BS6	1.53	4.81	1.38	4.31
SY4	1.18	3.15	1.02	2.73
DHS3	1.38	4.12	1.21	3.63

Last, we repeat Table 4 to take into account of transaction costs. We regress various pricing factors against our SDF after adjusting for the transaction costs. For the variables from the benchmark models, we use the original data without removing the transaction cost. The results are summarized in Table 7. Again, we see that most pricing factors from benchmark models still do not have statistically significant alpha when regressing against our SDF after accounting for moderate level of transaction costs. The only exception is the expected investment growth factor (R_{EG}), which has a statistically significant alpha when the transaction cost is 40 bps per dollar.

Table 7: **Testing the prevailing pricing factors against the MAXSER-S(20) SDF, adjusting for transaction costs.** This table reports the regression results of various pricing factors against the MAXSER-S(20) SDF after adjusting for the transaction costs, using the monthly returns from 2000 to 2021. The transaction cost is assumed to be 20 bps or 40 bps per dollar of transaction. The portfolio is rebalanced every month. The pricing factors from various benchmark models are the original data without removing the transaction costs. Alpha in percentage and its t-statistic are reported.

Transaction cost	20 bps		40 bps	
Factor	Alpha (%)	t-statistic	Alpha (%)	t-statistic
Mkt-RF	0.37	1.24	0.39	1.31
SMB	0.23	1.12	0.24	1.17
HML	-0.08	-0.37	-0.05	-0.25
RMW	0.22	1.15	0.24	1.25
CMA	0.10	0.76	0.12	0.89
MOM	-0.29	-0.85	-0.26	-0.76
R_ME	0.23	1.08	0.24	1.16
R_IA	-0.02	-0.17	-0.00	-0.02
R_ROE	0.03	0.13	0.05	0.24
R_EG	0.31	1.95	0.32	2.04
HmLm	0.32	1.14	0.32	1.17
MGMT	0.16	0.71	0.19	0.84
PERF	0.23	0.62	0.26	0.72
PEAD	0.18	1.29	0.19	1.37
FIN	0.11	0.36	0.14	0.48

5 Investigating the MAXSER-S(20) SDF

5.1 Sources of the MAXSER-S(20) SDF

Previous results demonstrate the superior performance of MAXSER-S(20) SDF. As we selected 20 out of 194 input portfolios, one might wonder what input portfolios are important. Note that we estimate the SDF portfolio every year between 2000 and 2021 with a rolling window of 20 years. Therefore, there are 22 different trained portfolios. We collect the variables that have ever included in the estimated portfolios for at least 20% of the times, i.e., at least 5 out of 22 years. We plot the bar chart of the selected variable

frequencies in Figure 4.

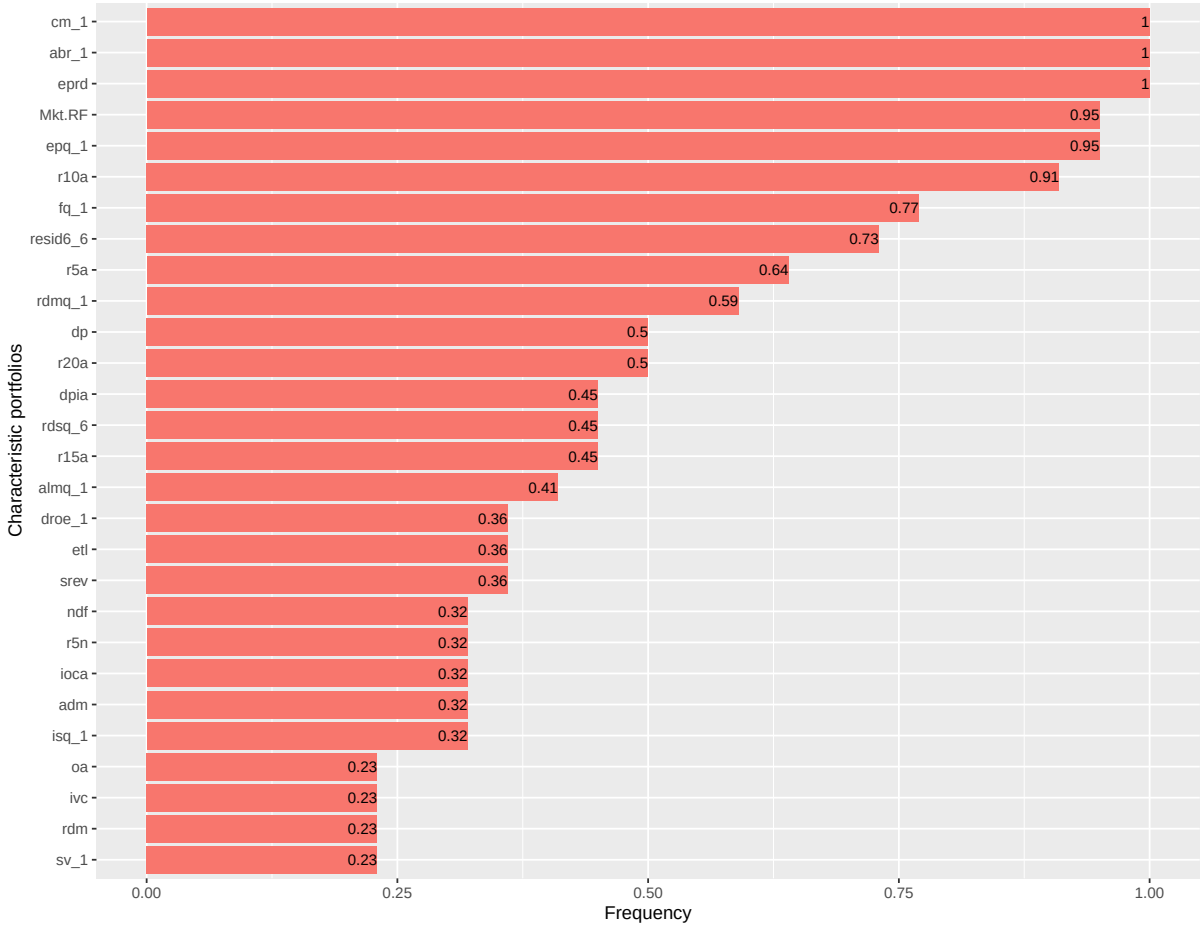

Figure 4: **Frequencies of variables included in the MAXSER-S(20) SDF.** We report the inclusion frequency of variables in the MAXSER-S(20) SDF from all training periods with an inclusion frequency higher than 20%.

We see from Figure 4 that 28 variables are frequently selected in the MAXSER-S(20) SDF, with an inclusion frequency higher than 20%. Important variables are usually related to momentum and earnings. For example, the most frequently selected variables (being selected for over 90% of the times) are the customer momentum (cm_1), the cumulative abnormal stock return (abr_1), the earnings predictability (eprd), the market portfolio (Mkt-RF), the quarterly earnings-to-price (epq_1), and the average returns across 120 months (r10a). The variables which are included for 50%–90% of the times are the quarterly fundamental score (fq_1), the six-month residual momentum (resd6_6), the average returns across 60 months (r5a), the quarterly R&D expense-to-market (rdmq_1), the dividend yield (dp), and the average returns across 240 months (r20a). See more details about the construction of these characteristic-based portfolios in Hou et al. (2020b).

It is interesting to note that the Fama-French factors are not in the list of frequently selected variables in the MAXSER-S(20) SDF, except the market portfolio.

In Table 8, we further provide some summary statistics of the most frequently selected variables in the MAXSER-S(20) SDF. We report the t-statistics of alphas of these variables when regressing against various factor models. We see from Table 8 that most of these frequently selected variables do not have statistically significant alphas under various factor models. Taking the CAPM as an example, only 7 out of 28 variables have a significant alpha. These variables are instrumental in our estimated MAXSER-S(20) SDF portfolio as they help to reduce the portfolio risk and hence improve the overall performance of the portfolio. The results demonstrate the advantage of our approach to constructing the SDF. Instead of merely selecting variables which have significant alphas under various models (e.g., strong anomalies), our approach accounts for the cross-sectional dependence in asset returns and is able to select important variables which contribute to the Sharpe ratio of the SDF portfolio. Our findings are in line with Bryzgalova et al. (2023b), who also emphasize the importance of considering the characteristic interactions.

Finally, to capture the economics meaning, we summarize the inclusion frequencies of the variables at the group-level. Specifically, in each training period, we compute the total number of variables from each anomaly group which are used in constructing the MAXSER-S(20) SDF and then report the average number from the total 22 years in Figure 5.

We see from Figure 5 that the intangible group occurs most frequently in the MAXSER-S(20) SDF. On average, 7 out of 30 variables from the intangible group are selected. The second important group is momentum. On average, 4 out of 41 momentum variables are significant in the SDF. The investment group, value-growth group and the profitability group each has on average 2 significant variables in the SDF, and the frictions group has one variable on average in the SDF. The market portfolio is selected for 95% of the time, while the SMB, RMW and HML portfolio occurs only 14%, 5% and 5% of the times, respectively. Also, the CMA and MOM in the Fama-French six-factor model are not included in the MAXSER-S(20) SDF.

Table 8: **Summary statistics of the most frequently selected variables in the MAXSER-S(20) SDF.** This table reports the summary statistics of the most frequently selected variables in constructing the MAXSER-S(20) SDF, including the frequency, group name, and the t-statistics of alphas when regressing this variable against some benchmark factor models.

			t-statistic of alphas from various factor models								
Variable	Frequency	Group	CAPM	FF3	FF5	FF6	Q4	Q5	SY4	DHS3	BS6
eprd	1.000	intangibles	2.52	4.53	3.90	3.84	3.66	3.07	1.96	3.17	3.84
abr_1	1.000	momentum	1.68	2.44	2.21	2.12	1.87	1.58	0.08	-0.17	2.12
cm_1	1.000	momentum	1.41	1.71	1.90	1.86	1.80	1.35	1.04	1.36	1.86
epq_1	0.955	value-growth	1.81	1.66	0.60	0.75	0.39	1.27	2.42	1.50	0.75
Mkt-RF	0.955	Mkt.RF	0.54	1.55	0.65	0.51	0.40	0.64	2.72	0.04	0.511
r10a	0.909	intangibles	3.76	3.02	3.43	3.70	4.02	3.46	3.38	3.72	3.70
fq_1	0.773	profitability	0.42	1.88	0.01	0.17	0.22	0.16	0.77	0.67	0.17
resid6_6	0.727	momentum	0.53	0.99	0.40	0.14	0.36	0.73	0.17	0.12	0.14
r5a	0.636	intangibles	1.48	1.33	1.22	1.20	1.61	1.63	1.59	0.83	1.20
rdmq_1	0.591	intangibles	3.00	1.81	2.30	3.07	2.99	2.27	2.82	2.67	3.07
r20a	0.500	intangibles	2.39	2.58	2.49	2.50	2.66	2.59	1.86	2.08	2.50
dp	0.500	value-growth	1.66	2.77	1.90	1.92	2.17	2.32	2.55	1.69	1.92
r15a	0.455	intangibles	0.61	0.83	0.94	0.94	0.55	0.40	0.45	0.36	0.94
rdsq_6	0.455	intangibles	1.02	1.61	3.88	3.83	2.72	1.49	0.95	2.97	3.83
dpia	0.455	investment	-0.59	0.74	0.58	0.62	0.32	0.90	0.38	0.72	0.62
almq_1	0.409	intangibles	2.23	0.77	0.35	0.54	0.77	1.68	0.82	1.26	0.54
srev	0.364	frictions	0.64	0.42	0.05	0.12	0.13	0.71	0.77	0.94	0.12
etl	0.364	intangibles	1.10	0.48	0.21	0.04	0.03	0.04	0.03	0.62	0.04
droe_1	0.364	profitability	1.73	2.51	1.08	0.89	0.45	0.14	0.75	1.06	0.89
isq_1	0.318	frictions	1.12	0.88	1.40	1.33	1.56	1.34	0.27	0.97	1.33
adm	0.318	intangibles	1.21	0.26	0.82	0.69	0.16	0.13	0.40	0.45	0.69
ioca	0.318	intangibles	2.78	3.60	1.85	1.73	2.16	1.29	0.97	2.70	1.73
r5n	0.318	intangibles	0.28	0.78	1.57	1.51	0.97	0.04	0.04	0.79	1.51
ndf	0.318	investment	1.35	1.76	0.25	0.26	0.31	0.25	0.47	1.20	0.26
sv_1	0.227	frictions	0.21	0.88	0.41	0.37	0.07	0.23	0.12	0.09	0.37
rdm	0.227	intangibles	2.20	1.36	2.16	2.24	2.27	0.95	0.09	2.49	2.24
ivc	0.227	investment	0.91	0.95	0.94	0.87	0.64	0.30	0.72	0.83	0.87
oa	0.227	investment	0.48	0.14	0.88	0.90	1.05	0.30	1.03	0.99	0.90

5.2 Nonsparsity in PCs of characteristic-based portfolios

Kozak et al. (2020) document that the SDF is sparse in the sense that it can be spanned by a small number of principal components in the characteristic portfolios. To examine the sparsity of the SDF, we first follow Kozak et al. (2020) to construct 20 PCs from the 193 characteristic portfolios (the market portfolio is excluded). Then we combine these 20 PCs with the existing 193 portfolio as the new input set. Last, we perform MAXSER to construct the SDF, denoted as MAXSER(+20PCs). From Table 1, we first see that the plugin portfolio using 20 PCs underperforms MAXSER-S(20) SDF portfolio. Second, we see that the SDF portfolio constructed from the enlarged input set does not improve over the SDF constructed by 194 characteristic-based portfolios. The MAXSER(+20PCs) SDF portfolio has a monthly Sharpe ratio of 0.41, which is lower than that of the MAXSER-S(20) portfolio constructed by the 194 portfolios (0.42). We find that the number of variables included in the MAXSER(+20PCs) SDF is 40–50, similar to the case when constructing the SDF without PCs. In addition, the 20 PCs are rarely included in the MAXSER(+20PCs) SDF portfolio. Only two PCs have ever appear in the MAXSER(+20PCs) SDF while none of them has an inclusion frequency higher than 10%. This suggests that the SDF is dense in the sense that it still largely uses the characteristic-based portfolios instead of PCs. Moreover, when evaluating the significance of the 20 PCs against our estimated MAXSER-S(20) SDF, we find that none of the PCs has a statistically significant alpha. This suggests that the 20 PCs are less important as our MAXSER-S(20) SDF fully captures the returns of the 20 PCs.

Next, we perform another test to directly distinguish PCs from our characteristic-based SDF. We project all characteristic-based portfolios onto the PC space (e.g., 20 PCs and the market portfolio) and then construct the SDF using PC-based portfolio returns. We find that, again, the estimated SDF is not sparse. The number of variables estimated with MAXSER is about 20–30. In addition, the monthly Sharpe ratio of the estimated portfolio is 0.35, which is lower than the MAXSER-S(20) SDF portfolio constructed directly from the characteristic-based portfolios. This further suggests that the SDF is not sparse in the PC space of characteristic-based portfolios.

Overall, these results above suggest that, using PCs of characteristic-based portfolios do not have incremental contribution to constructing the SDF. Different from Kozak et al. (2020), we find that the cross-section of characteristic-based portfolio returns is unlikely to be sparse in the sense that they can not be adequately explained by a small number of PCs. Instead, we find that using about 20 characteristic-based portfolios performs well in constructing the SDF.

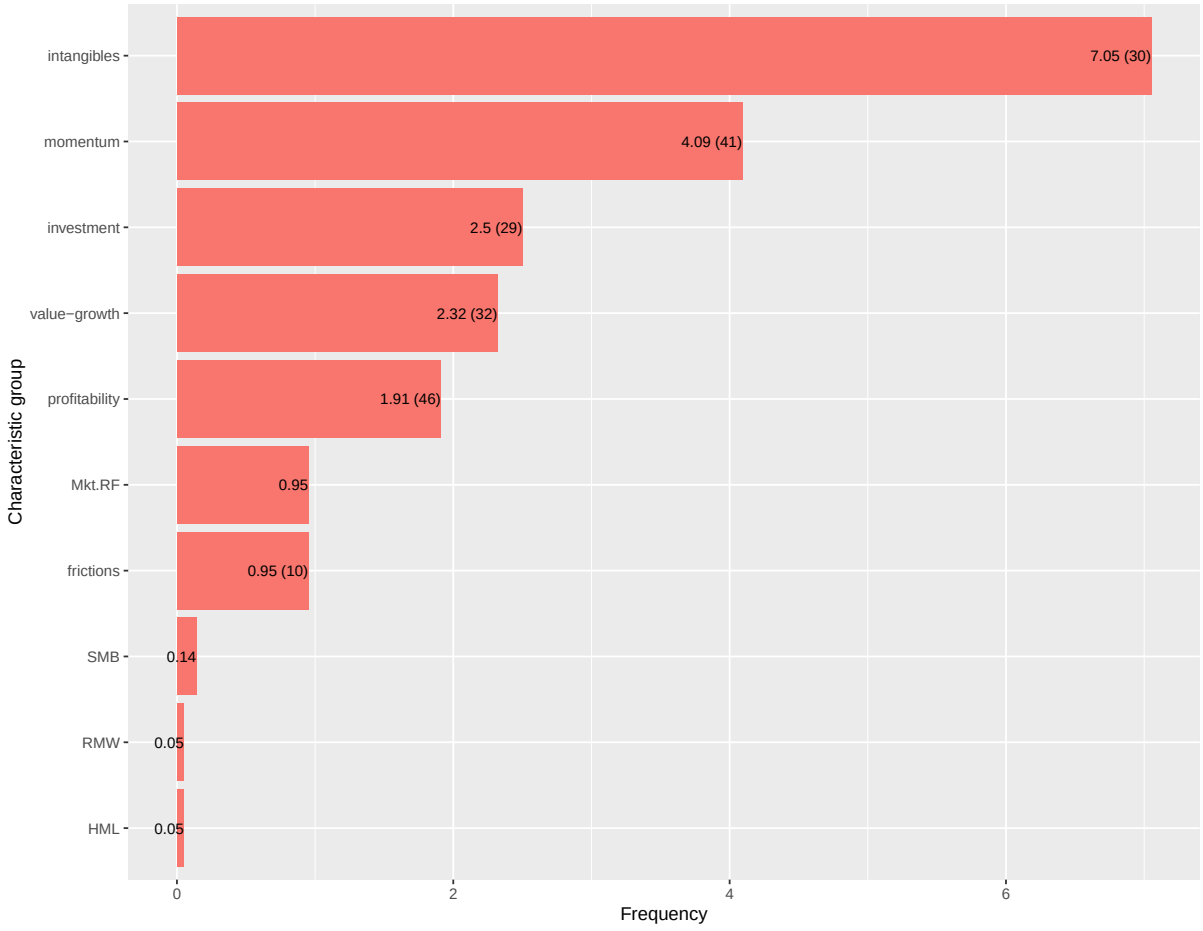

Figure 5: **Average number of variables from each characteristic group included in the MAXSER-S(20) SDF.** This barchart plots the average number of variables from each characteristic group which are included in the MAXSER-S(20) SDF. The number in parenthesis is the total number of variables in a characteristic group.

6 Conclusions

We develop a statistical framework to estimate the high-dimensional SDF loadings. We propose to estimate the loadings with the MAXSER estimator, which use a Lasso-type regression with the estimated maximized Sharpe ratio as the response. We show that our proposed approach can consistently select the factors with non-zero weights in constructing the SDF portfolio. We further develop a statistical inference theory for the SDF loading based on the asymptotic normality of the plug-in estimator from consistently post-selected variables. Our empirical tests show that the SDF can be formed by approximately 20 characteristic-based portfolios, which delivers a high monthly Sharpe ratio of 0.42. In addition, the SDF is not sparse in the PC space because it can not be spanned by a small number of principal components.

References

Anatolyev, S. and Mikusheva, A. (2022). Factor models with many assets: Strong factors, weak factors, and the two-pass procedure. *Journal of Econometrics*, 229(1):103–126.

Andrews, I. and Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, 109(8):2766–94.

Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *Review of Financial Studies*, 32(7):2890–2919.

Bai, Z.-D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Annals of Probability*, 26(1):316–345.

Barillas, F., Kan, R., Robotti, C., and Shanken, J. (2020). Model comparison with Sharpe ratios. *Journal of Financial and Quantitative Analysis*, 55(6):1840–1874.

Barillas, F. and Shanken, J. (2017). Which alpha? *Review of Financial Studies*, 30(4):1316–1338.

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

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Black, F. and Litterman, R. B. (1991). Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income*, 1(2):7–18.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, 54(2):655–671.

Bryzgalova, S. (2017). Spurious factors in linear asset-pricing models. Working Paper, London Business School.

Bryzgalova, S., Huang, J., and Julliard, C. (2023a). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance*, 78(1):487–557.

Bryzgalova, S., Pelger, M., and Zhu, J. (2023b). Forest through the trees: Building cross-sections of asset returns. Working Paper, London Business School.

Chen, A. Y. (2021). The limits of p-hacking: Some thought experiments. *Journal of Finance*, 76(5):2447–2480.

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

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

El Karoui, N. (2003). On the largest eigenvalue of wishart matrices with identity covariance when n, p and p/n tend to infinity. *arXiv preprint math/0309355*.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, 128(2):234–252.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of Statistics*, 39(6):3320.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, 75(3):1327–1370.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5):2326–2377.

Giglio, S., Liao, Y., and Xiu, D. (2021). Thousands of alpha tests. *Review of Financial Studies*, 34(7):3456–3496.

Giglio, S., Xiu, D., and Zhang, D. (2022). Test assets and weak factors. Working Paper, Yale University.

Gospodinov, N., Kan, R., and Robotti, C. (2017). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, 85(5):1613–1628.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5):2223–2273.

Gu, S., Kelly, B., and Xiu, D. (2021). Autoencoder asset pricing models. *Journal of Econometrics*, 222(1):429–450.

Hansen, L. P. and Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance*, 52(2):557–590.

Harvey, C. R. and Liu, Y. (2021a). Lucky factors. *Journal of Financial Economics*, 141(2):413–435.

Harvey, C. R. and Liu, Y. (2021b). Uncovering the iceberg from its tip: A model of publication bias and p-hacking. Working Paper, Duke University.

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

Hou, K., Mo, H., Xue, C., and Zhang, L. (2021). An augmented q-factor model with expected growth. *Review of Finance*, 25(1):1–41.

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

Hou, K., Xue, C., and Zhang, L. (2020a). Replicating anomalies. *Review of Financial Studies*, 33(5):2019–2133.

Hou, K., Xue, C., and Zhang, L. (2020b). Technical document: Testing portfolios.

Jacobs, H. and Müller, S. (2020). Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1):213–230.

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

Kan, R., Wang, X., and Zhou, G. (2022). Optimal portfolio choice with estimation risk: No risk-free asset case. *Management Science*, 68(3):2047–2068.

Kan, R. and Zhang, C. (1999). Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, 54(1):203–235.

Kelly, B. T., Pruitt, S., and Su, Y. (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.

Kozak, S. and Nagel, S. (2023). When do cross-sectional asset pricing factors span the stochastic discount factor? NBER Working Paper 31275.

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

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies*, 30(12):4349–4388.

Leng, C., Lin, Y., and Wahba, G. (2006). A note on the lasso and related procedures in model selection. *Statistica Sinica*, pages 1273–1284.

Lettau, M. and Pelger, M. (2020a). Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31.

Lettau, M. and Pelger, M. (2020b). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies*, 33(5):2274–2325.

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

Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Finance Letters*, (1):21–23.

Michaud, R. O. (1989). The Markowitz optimization enigma: is 'optimized'optimal? *Financial Analysts Journal*, 45(1):31–42.

Stambaugh, R. F. and Yuan, Y. (2017). Mispricing factors. *Review of Financial Studies*, 30(4):1270–1315.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 76(1):217–242.

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563.

Appendix A Proofs

A.1 Proof of Theorem 1

Suppose that without lost of generality, $S_1 = \{1, ..., q_u\}$. Denote by $\mathbf{U}_{S_1}$, the first q_u rows, and $\mathbf{U}_{S_1^c}$ the remaining rows of $\mathbf{U} = (\mathbf{U}_1, ..., \mathbf{U}_T) =: (U_{it})$. Denote by $\widehat{\mathbf{U}}_{S_1}$, the first q_u rows, and $\widehat{\mathbf{U}}_{S_1^c}$ the remaining rows of $\widehat{\mathbf{U}} = (\widehat{\mathbf{U}}_1, ..., \widehat{\mathbf{U}}_T) =: (\widehat{U}_{it})$. Because $\widehat{\boldsymbol{\omega}}_R^*$ and $\boldsymbol{\omega}_R^*$ are proportional to $\widehat{\boldsymbol{\omega}}_u^*$ and $\boldsymbol{\omega}_u^*$, respectively. To show (2.6), it is equivalent to show

$$P\Big(\text{sign}(\widehat{\boldsymbol{\omega}}_u^*) = \text{sign}(\boldsymbol{\omega}_u^*)\Big) \to 1.$$

By Proposition 1 of Zhao and Yu (2006),

$$P\Big(\text{sign}(\widehat{\boldsymbol{\omega}}_u^*) = \text{sign}(\boldsymbol{\omega}_u^*)\Big) \geq P\Big(A_n \cap B_n\Big), \tag{A.1}$$

where

$$A_n = \left\{ \left|(C_{11}^{-1}W_1)_i\right| < \sqrt{T}\Big(\left|(\boldsymbol{\omega}_u^*)_i\right| - \frac{\lambda_T}{2T}\left|\left(C_{11}^{-1}\text{sign}(\boldsymbol{\omega}_u^*)\right)_i\right|\Big), \text{ for all } i = 1, ..., q_u \right\},$$
$$B_n = \left\{ \left|(C_{21}C_{11}^{-1}W_1 - W_2)_i\right| < \frac{\lambda_T}{2\sqrt{T}}\eta, \text{ for all } i = 1, ..., N - q_u + 1 \right\},$$

and recall that $\widehat{r}_{u,c} = 1/\sqrt{\widehat{\theta}_u} + \sqrt{\widehat{\theta}_u}$, η is defined in Assumption 5, and

$$C_{11} = \frac{1}{T}\widehat{\mathbf{U}}_{S_1}\widehat{\mathbf{U}}_{S_1}^T, \quad C_{21} = \frac{1}{T}\widehat{\mathbf{U}}_{S_1^c}\widehat{\mathbf{U}}_{S_1}^T,$$
$$W_1 = \sqrt{\frac{1}{T}}\widehat{\mathbf{U}}_{S_1}\Big(\widehat{r}_{u,c}\mathbf{1} - \widehat{\mathbf{U}}^T\boldsymbol{\omega}_u^*\Big), \quad W_2 = \sqrt{\frac{1}{T}}\widehat{\mathbf{U}}_{S_1^c}\Big(\widehat{r}_{u,c}\mathbf{1} - \widehat{\mathbf{U}}^T\boldsymbol{\omega}_u^*\Big).$$

Define an event G, for some $C > 0$,

$$G = \Big\{ \|(W_1^T, W_2^T)\|_{\max} < C\sqrt{\log N} \Big\}$$
$$\bigcap \left\{ \|C_{11} - \boldsymbol{\Sigma}_{u,11} - \boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T\|_{\max} < C\sqrt{\frac{\log N}{T}} \right\}$$
$$\bigcap \left\{ \|C_{21} - \boldsymbol{\Sigma}_{u,21} - \boldsymbol{\alpha}_2\boldsymbol{\alpha}_1^T\|_{\max} < C\sqrt{\frac{\log N}{T}} \right\}.$$

We will show that as $N, T \to \infty$,

$$P(G) \to 1. \tag{A.2}$$

About A_n, under Assumptions 1 and 4, we have $q_u^2 \log N/T \to 0$, and under event G,

$\|C_{11} - \mathbf{\Sigma}_{u,11} - \boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T\| = O(q_u\sqrt{\log N/T}) = o(1)$. By Assumption 5 that $\|\mathbf{\Sigma}_{u,11}^{-1}\| = O(1)$, and Weyl's theorem, under event G, for all large T,

$$\|C_{11}^{-1}\| < C. \tag{A.3}$$

Therefore,

$$\|C_{11}^{-1}W_1\|_{\max} \le C\sqrt{q_u(\log N)}. \tag{A.4}$$

Moreover, by (A.3),

$$\left\|\frac{\lambda_T}{2T}C_{11}^{-1}\mathrm{sign}\Big((\boldsymbol{\omega}_u^*)_{S_1}\Big)\right\|_{\max} \le \left\|\frac{\lambda_T}{2T}C_{11}^{-1}\right\|_1 \le C\frac{\lambda_T\sqrt{q_u}}{T}. \tag{A.5}$$

By Assumption 4 that $\min_{1\le i\le q_u}|\omega_i| \gg \lambda_T\sqrt{q_u}/N$ and $\lambda_T \gg q_u\sqrt{(\log N)N}$, (A.2), (A.4) and (A.5), we get that

$$P(A_n^c) = o(1). \tag{A.6}$$

About B_n, by Assumption 5, under event G, for all large N and T, we have

$$\|C_{21}\|_{\max} < C.$$

Therefore,

$$\begin{aligned}\|C_{21}C_{11}^{-1}W_1\|_{\max} \le& \|C_{21}\|_{\max}\sum_{i\le q_u}|(C_{11}^{-1}W_1)_i| \\ \le& \sqrt{q_u}\|C_{21}\|_{\max}\cdot\|C_{11}^{-1}W_1\| \\ \le& q_u\|C_{21}\|_{\max}\cdot\|C_{11}^{-1}\|\cdot\|W_1\|_{\max} \\ =& O(q_u\sqrt{\log N}),\end{aligned}$$

where the second line holds by the Cauchy-Schwarz inequality. Therefore, by Assumption 4 that $\lambda_T \gg q_u\sqrt{(\log N)N}$, we have

$$P(B_n^c) = o(1). \tag{A.7}$$

Combining (A.1), (A.2), (A.6) and (A.7) yields

$$P\Big(\mathrm{sign}(\widehat{\boldsymbol{\omega}}_u) = \mathrm{sign}(\boldsymbol{\omega}_u^*)\Big) \ge 1 - P(A_n^c) - P(B_n^c) = 1 - o(1).$$

The desired bound (2.6) follows.

It remains to show (A.2). Define $r_{u,c} = 1/\sqrt{\theta_u} + \sqrt{\theta_u}$, $\overline{\widehat{\mathbf{U}}} = \widehat{\mathbf{U}}\mathbf{1}/T$, $\overline{\mathbf{U}} = \mathbf{U}\mathbf{1}/T$, and

$W = (W_1^T, W_2^T)^T$. We have

$$\begin{aligned} W =& (\widehat{r}_{u,c} - r_{u,c})\sqrt{T}\overline{\widehat{\mathbf{U}}} \\ &+ r_{u,c}\sqrt{T}(\overline{\widehat{\mathbf{U}}} - \overline{\mathbf{U}}) \\ &- \sqrt{\frac{1}{T}}(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}^T\boldsymbol{\omega}_u^* \\ &- \sqrt{\frac{1}{T}}\mathbf{U}(\widehat{\mathbf{U}} - \mathbf{U})^T\boldsymbol{\omega}_u^* \\ &- \sqrt{\frac{1}{T}}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})^T\boldsymbol{\omega}_u^* \\ &+ \sqrt{\frac{1}{T}}\mathbf{U}(r_{u,c}\mathbf{1} - \mathbf{U}^T\boldsymbol{\omega}_u^*) \\ =:& I + II + III + IV + V + VI. \end{aligned}$$

About terms I and II, under Assumptions 1–3, by Bernstein's inequality,

$$P\left(\|\overline{\mathbf{U}} - \boldsymbol{\alpha}\|_{\max} > \sqrt{\frac{\log N}{T}}\right) = O\left(\frac{1}{T^2}\right),$$

and

$$P\left(|\boldsymbol{\mu}_f^T\boldsymbol{\Sigma}_f^{-1}\boldsymbol{\mu}_f - \widehat{\boldsymbol{\mu}}_f^T\widehat{\boldsymbol{\Sigma}}_f^{-1}\widehat{\boldsymbol{\mu}}_f| \geq \sqrt{\frac{\log N}{T}}\right) = O\left(\frac{1}{T^2}\right).$$

By the proof of Proposition 2 of Ao et al. (2019),

$$P\left(|\theta_{all} - \widehat{\theta}_{all}| \geq \sqrt{\frac{\log N}{N}}\right) = O\left(\frac{1}{\log N}\right).$$

Therefore

$$P\left(|\widehat{r}_{u,c} - r_{u,c}| > c\sqrt{\frac{\log N}{T}}\right) = O\left(\frac{1}{\log N}\right).$$

By the Cauchy-Schwarz inequality,

$$\|\overline{\widehat{\mathbf{U}}} - \overline{\mathbf{U}}\|_{\max} \leq \max_{1\leq i\leq N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{U}_{it} - U_{it}\right)^2}. \tag{A.8}$$

We have

$$\sum_{t=1}^{T}(\widehat{U}_{it} - U_{it})^2 = \sum_{t=1}^{T}\left((\boldsymbol{\beta}_{i\cdot} - \widehat{\boldsymbol{\beta}}_{i\cdot})\mathbf{f}_t\right)^2 \leq K\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|_{\max}^2\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbf{f}_{t,k}^2,$$

where $\mathbf{f}_{t,k} = (\mathbf{f}_t)_k$, $\widehat{\boldsymbol{\beta}}_{i\cdot}$ and $\boldsymbol{\beta}_{i\cdot}$ are the ith row of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$, respectively. Note that

$\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}} = \sum_{t=1}^{T} \mathbf{U}_t \check{\mathbf{f}}_t^T (\check{\mathbf{F}}^T \check{\mathbf{F}})^{-1}$, where $\check{\mathbf{F}} = (\check{\mathbf{f}}_{t,k})_{1\le t\le T, 1\le k\le K}$, $\check{\mathbf{f}}_t = (\mathbf{f}_{t,k} - \bar{f}_k)_{1\le k\le K}$, and $\bar{f}_k = \sum_{t=1}^{T} \mathbf{f}_{t,k}/T$. By Assumption 2, for some constant $C > 0$,

$$P\left(\sum_{k=1}^{K} \sum_{t=1}^{T} \mathbf{f}_{t,k}^2 < CT \right) > 1 - \frac{1}{T}.$$

By the independence between $\mathbf{U}$ and $\check{\mathbf{F}}$ and Bernstein's inequality, one can show that

$$P\left(\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|_{\max} > \sqrt{\frac{\log N}{T}} \right) < \frac{1}{T^2}.$$

Therefore,

$$P\left(\max_{1\le i\le N} \frac{1}{T} \sum_{t=1}^{T} (\widehat{U}_{it} - U_{it})^2 > \frac{\log N}{T} \right) = O\left(\frac{1}{T^2} \right). \tag{A.9}$$

By (A.8) and (A.9), we get

$$P\left(\|\overline{\widehat{\mathbf{U}}} - \overline{\mathbf{U}}\|_{\max} > \sqrt{\frac{\log N}{T}} \right) = O\left(\frac{1}{T^2} \right).$$

Combining the results above yields

$$P\Big(\|I\|_{\max} > \sqrt{\log N} \Big) = O\left(\frac{1}{\log N} \right),$$

and

$$P\Big(\|II\|_{\max} > \sqrt{\log N} \Big) = O\left(\frac{1}{T^2} \right).$$

About term III, by the Cauchy-Schwarz inequality,

$$\|(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}^T \boldsymbol{\omega}^*\|_{\max} \le \left(\max_{1\le i\le N} \sqrt{\sum_{t=1}^{T} \left(\widehat{U}_{it} - U_{it} \right)^2} \right) \sqrt{\sum_{t=1}^{T} \left((\boldsymbol{\omega}_u^*)^T \mathbf{U}_t \right)^2}.$$

Note that $(\boldsymbol{\omega}_u^*)^T \mathbf{U}_t \overset{\text{i.i.d.}}{\sim} N(\sqrt{\theta_u}, 1)$, we get that for some $C > 0$,

$$P\left(\left(\sum_{t=1}^{T} \left((\boldsymbol{\omega}_u^*)^T \mathbf{U}_t \right)^2 \right) < CT \right) > 1 - \frac{1}{T^2}. \tag{A.10}$$

Combining (A.9) and (A.10) yields

$$P\Big(\|III\|_{\max} > \sqrt{\log N} \Big) = O\left(\frac{1}{T^2} \right).$$

About term IV, by the Cauchy-Schwarz inequality, we get that

$$\|IV\|_{\max} \leq \sqrt{\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}U_{it}^2}\sqrt{\sum_{t=1}^{T}\Big((\boldsymbol{\omega}_u^*)^T(\widehat{\mathbf{U}}_t-\mathbf{U}_t)\Big)^2}. \tag{A.11}$$

Under Assumptions 2 and 3, there exists $C>0$,

$$P\left(\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}(U_{it})^2<C\right)\geq 1-\frac{1}{T^2}. \tag{A.12}$$

Note that $\widehat{\mathbf{U}}_t-\mathbf{U}_t=(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\mathbf{f}_t=\big(\sum_{t=1}^{T}\mathbf{U}_t\check{\mathbf{f}}_t^T(\check{\mathbf{F}}^T\check{\mathbf{F}})^{-1}\big)\mathbf{f}_t$. We have that

$$(\boldsymbol{\omega}_u^*)^T(\widehat{\mathbf{U}}_t-\mathbf{U}_t)=\left(\sum_{t=1}^{T}(\boldsymbol{\omega}_u^*)^T\mathbf{U}_t\check{\mathbf{f}}_t^T(\check{\mathbf{F}}^T\check{\mathbf{F}})^{-1}\right)\mathbf{f}_t.$$

Therefore,

$$\begin{aligned}\sum_{t=1}^{T}\Big((\boldsymbol{\omega}_u^*)^T(\widehat{\mathbf{U}}_t-\mathbf{U}_t)\Big)^2 \leq& \left\|\sum_{t=1}^{T}(\boldsymbol{\omega}_u^*)^T\mathbf{U}_t\check{\mathbf{f}}_t^T(\check{\mathbf{F}}^T\check{\mathbf{F}})^{-1}\right\|^2\left(\sum_{t=1}^{T}\|\mathbf{f}_t\|^2\right)\\ \leq& \left\|\sum_{t=1}^{T}(\boldsymbol{\omega}_u^*)^T\mathbf{U}_t\check{\mathbf{f}}_t^T\right\|^2\cdot\left\|(\check{\mathbf{F}}^T\check{\mathbf{F}})^{-1}\right\|^2\left(\sum_{t=1}^{T}\|\mathbf{f}_t\|^2\right).\end{aligned}$$

Because $(\boldsymbol{\omega}_u^*)^T\mathbf{U}_t\sim N(\sqrt{\theta_u},1)$, and independent with $\mathbf{f}_t$, we have

$$P\left(\left\|\sum_{t=1}^{T}(\boldsymbol{\omega}_u^*)^T\mathbf{U}_t\check{\mathbf{f}}_t^T\right\|>\sqrt{(\log N)T}\right)=O\left(\frac{1}{T^2}\right).$$

Moreover, because $\mathbf{f}_t\overset{\text{i.i.d.}}{\sim}N(\boldsymbol{\mu}_f,\boldsymbol{\Sigma}_f)$, we have

$$P\left(\sum_{t=1}^{T}\|\mathbf{f}_t\|^2<cT\right)>1-\frac{1}{T^2}.$$

and

$$P\left(\left\|(\check{\mathbf{F}}^T\check{\mathbf{F}})^{-1}\right\|<\frac{c}{T}\right)>1-\frac{1}{T^2}.$$

It follows that

$$P\left(\sum_{t=1}^{T}\Big((\boldsymbol{\omega}_u^*)^T(\widehat{\mathbf{U}}_t-\mathbf{U}_t)\Big)^2>C\log N\right)=O\left(\frac{1}{T^2}\right). \tag{A.13}$$

Combining (A.11), (A.12) and (A.13) yields

$$P\Big(IV > \sqrt{\log N}\Big) = O\bigg(\frac{1}{T^2}\bigg).$$

About term V, we have

$$\begin{aligned}
&\|(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})^T \boldsymbol{\omega}_u^*\|_{\max} \\
\leq &\sqrt{\max_{1\leq i\leq N}\bigg(\sum_{t=1}^{T}(\widehat{U}_{it} - U_{it})^2\bigg)}\sqrt{\sum_{t=1}^{T}\Big((\boldsymbol{\omega}_u^*)^T(\widehat{\mathbf{U}}_t - \mathbf{U}_t)\Big)^2}.
\end{aligned}$$

By (A.9) and (A.13), we then get

$$P\bigg(\|V\|_{\max} > \frac{\log N}{\sqrt{T}}\bigg) = O\bigg(\frac{1}{T^2}\bigg).$$

Finally, about term VI, we have

$$\|VI\|_{\max} = \sqrt{\frac{1}{T}}\bigg(\max_{1\leq i\leq N}\bigg|\sum_{t=1}^{T} U_{it}(r_{u,c} - \mathbf{U}_t^T\boldsymbol{\omega}_u^*)\bigg|\bigg).$$

Note that

$$\begin{aligned}
E\Big(\mathbf{U}_t(r_{u,c} - \mathbf{U}_t^T\boldsymbol{\omega}_u^*)\Big) &= \frac{1}{\sqrt{\theta_u}}\boldsymbol{\alpha} + \boldsymbol{\alpha}\sqrt{\theta_u} - \Big(\boldsymbol{\alpha}\boldsymbol{\alpha}^T + \boldsymbol{\Sigma}_{\boldsymbol{u}}\Big)\frac{1}{\sqrt{\theta_u}}\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\boldsymbol{\alpha} \\
&= \frac{1}{\sqrt{\theta_u}}\boldsymbol{\alpha} + \boldsymbol{\alpha}\sqrt{\theta_u} - \frac{1}{\sqrt{\theta_u}}\boldsymbol{\alpha} - \boldsymbol{\alpha}\sqrt{\theta_u} \\
&= \mathbf{0}.
\end{aligned}$$

In addition, under Assumption 3, for some constant $C > 0$,

$$\max_{1\leq i\leq N} E\bigg(\Big(U_{it}(r_{u,c} - \mathbf{U}_t^T\boldsymbol{\omega}_u^*)\Big)^2\bigg) \leq \max_{1\leq i\leq N}\sqrt{E\Big(U_{it}^4\Big)}\sqrt{E\Big((r_{u,c} - \mathbf{U}_t^T\boldsymbol{\omega}_u^*)^4\Big)} < C.$$

Hence, under Assumption 2, by Bernstein's inequality, we get

$$P\Big(\|VI\|_{\max} \geq \sqrt{\log N}\Big) = O\bigg(\frac{1}{T^2}\bigg).$$

Combining the results above yields

$$P\Big(\|(W_1^T, W_2^T)\|_{\max} > C\sqrt{\log N}\Big) = O\Big(\frac{1}{\log N}\Big) \to 0.$$

Next, we have

$$\begin{aligned}C_{11} =&\frac{1}{T}(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})^T \\ &+ \frac{1}{T}(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})\mathbf{U}_{S_1}^T \\ &+ \frac{1}{T}\mathbf{U}_{S_1}(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})^T \\ &+ \frac{1}{T}\mathbf{U}_{S_1}\mathbf{U}_{S_1}^T,\end{aligned}$$

and

$$\begin{aligned}C_{21} =&\frac{1}{T}(\widehat{\mathbf{U}}_{S_1^c} - \mathbf{U}_{S_1^c})(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})^T \\ &+ \frac{1}{T}(\widehat{\mathbf{U}}_{S_1^c} - \mathbf{U}_{S_1^c})\mathbf{U}_{S_1}^T \\ &+ \frac{1}{T}\mathbf{U}_{S_1^c}(\widehat{\mathbf{U}}_{S_1} - \mathbf{U}_{S_1})^T \\ &+ \frac{1}{T}\mathbf{U}_{S_1^c}\mathbf{U}_{S_1}^T.\end{aligned}$$

By Assumptions 1–3 and (A.9), with probability tending one,

$$\left\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})^T\right\|_{\max} < C\frac{\log N}{T}.$$

By Assumptions 2 and Bernstein's inequality,

$$\left\|\frac{1}{T}\mathbf{U}\mathbf{U}^T - \boldsymbol{\Sigma}_u - \boldsymbol{\alpha}\boldsymbol{\alpha}^T\right\|_{\max} < C\sqrt{\frac{\log N}{T}},$$

and

$$\begin{aligned}&\left\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})\mathbf{U}^T\right\|_{\max} \\ <&\sqrt{\left\|\frac{1}{T}(\widehat{\mathbf{U}} - \mathbf{U})(\widehat{\mathbf{U}} - \mathbf{U})^T\right\|_{\max}}\sqrt{\left\|\frac{1}{T}\mathbf{U}\mathbf{U}^T\right\|_{\max}} \\ <&C\sqrt{\frac{\log N}{T}}.\end{aligned}$$

Similarly,

$$\left\|\frac{1}{T}\mathbf{U}(\widehat{\mathbf{U}} - \mathbf{U})^T\right\|_{\max} < C\sqrt{\frac{\log N}{T}}.$$

We then get

$$\|C_{11} - \boldsymbol{\Sigma}_{u,11} - \boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T\|_{\max} < C\sqrt{\frac{\log N}{T}},$$

and

$$\|C_{21} - \boldsymbol{\Sigma}_{u,21} - \boldsymbol{\alpha}_2\boldsymbol{\alpha}_1^T\|_{\max} \leq C\sqrt{\frac{\log N}{T}}.$$

The desired result (A.2) follows.

A.2 Proof of Theorem 2

To show Theorem 2, by Theorem 1 and Slusky's Theorem, it suffices to work under the event $\{\widehat{S}_{full,1} = S_{full,1}\}$ below. We have

$$\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1}\widehat{\boldsymbol{\mu}}_{full,1} - \boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1} \\
=&(\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1} - \boldsymbol{\Sigma}_{full,1}^{-1})(\widehat{\boldsymbol{\mu}}_{full,1} - \boldsymbol{\mu}_{full,1}) \\
&+ \boldsymbol{\Sigma}_{full,1}^{-1}(\widehat{\boldsymbol{\mu}}_{full,1} - \boldsymbol{\mu}_{full,1}) \\
&+ \boldsymbol{\Sigma}_{full,1}^{-1/2}(\boldsymbol{\Sigma}_{full,1}^{1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1}\boldsymbol{\Sigma}_{full,1}^{1/2} - \mathbf{I})(\mathbf{I} - \boldsymbol{\Sigma}_{full,1}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}\boldsymbol{\Sigma}_{full,1}^{-1/2})\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1} \\
&+ (\boldsymbol{\Sigma}_{full,1}^{-1} - \boldsymbol{\Sigma}_{full,1}^{-1}\widehat{\boldsymbol{\Sigma}}_{full,1}\boldsymbol{\Sigma}_{full,1}^{-1})\boldsymbol{\mu}_{full,1} \\
=:&I + II + III + IV.
\end{aligned} \tag{A.14}$$

Assumption 3 implies that $\|\boldsymbol{\mu}_{full,1}\|_{\max} = O(1)$, and $\|\operatorname{diag}(\boldsymbol{\Sigma}_{full,1})\|_{\max} = O(1)$. By Bernstein's inequality, we have

$$\|\widehat{\boldsymbol{\mu}}_{full,1} - \boldsymbol{\mu}_{full,1}\| = O_p\left(\sqrt{\frac{q_u(\log q_u)}{T}}\right). \tag{A.15}$$

Note that under Assumption 2, $T\boldsymbol{\Sigma}_{full,1}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}\boldsymbol{\Sigma}_{full,1}^{-1/2}$ follows Wishart distribution with df $T-1$ and covariance matrix $\mathbf{I}$, and can be written as

$$T\boldsymbol{\Sigma}_{full,1}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}\boldsymbol{\Sigma}_{full,1}^{-1/2} = \sum_{t=1}^{T-1}\mathbf{z}_t\mathbf{z}_t^T, \tag{A.16}$$

where $\mathbf{z}_t$ are i.i.d. multivariate standard normal. By Theorem 2 of El Karoui (2003), we have

$$\|\boldsymbol{\Sigma}_{full,1}^{-1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}\boldsymbol{\Sigma}_{full,1}^{-1/2} - \mathbf{I}\| = O_p\left(\sqrt{\frac{q_u}{T}}\right). \tag{A.17}$$

By Assumption 4 that $q_u^2(\log N) = o(N)$, we get that, with probability tending one, for all large N and T, $\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1}$ exists, $\|\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1}\| = O_p(1)$, and

$$\|\boldsymbol{\Sigma}_{full,1}^{1/2}\widehat{\boldsymbol{\Sigma}}_{full,1}^{-1}\boldsymbol{\Sigma}_{full,1}^{1/2} - \mathbf{I}\| = O_p\left(\sqrt{\frac{q_u}{T}}\right) = o_p(1). \tag{A.18}$$

In addition, because $\widehat{\mathbf{\Sigma}}_{full,1}^{-1} - \mathbf{\Sigma}_{full,1}^{-1} = \mathbf{\Sigma}_{full,1}^{-1/2}(\mathbf{\Sigma}_{full,1}^{1/2}\widehat{\mathbf{\Sigma}}_{full,1}^{-1}\mathbf{\Sigma}_{full,1}^{1/2} - \mathbf{I})\mathbf{\Sigma}_{full,1}^{-1/2}$, we get that

$$\|\widehat{\mathbf{\Sigma}}_{full,1}^{-1} - \mathbf{\Sigma}_{full,1}^{-1}\| = O_p\left(\sqrt{\frac{q_u}{T}}\right). \tag{A.19}$$

It follows that

$$\|I\| = O_p\left(\frac{q_u\sqrt{(\log q_u)}}{T}\right) = o_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.20}$$

where the last equality holds by the assumptions that $q_u^2(\log N) = o(N)$ and $N \asymp T$, which imply that $q_u^2(\log q_u) = o(T)$.

About term II, by Assumption 2, we have

$$II \sim N\left(0, \frac{1}{T}\mathbf{\Sigma}_{full,1}^{-1}\right). \tag{A.21}$$

About term III, we have

$$\begin{aligned}\|III\| =& \|\mathbf{\Sigma}_{full,1}^{-1/2}(\mathbf{\Sigma}_{full,1}^{1/2}\widehat{\mathbf{\Sigma}}_{full,1}^{-1}\mathbf{\Sigma}_{full,1}^{1/2} - \mathbf{I})(\mathbf{I} - \mathbf{\Sigma}_{full,1}^{-1/2}\widehat{\mathbf{\Sigma}}_{full,1}\mathbf{\Sigma}_{full,1}^{-1/2})\mathbf{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\| \\ \leq& \|\mathbf{\Sigma}_{full,1}^{-1/2}\|\|\mathbf{\Sigma}_{full,1}^{1/2}\widehat{\mathbf{\Sigma}}_{full,1}^{-1}\mathbf{\Sigma}_{full,1}^{1/2} - \mathbf{I}\|\|\mathbf{I} - \mathbf{\Sigma}_{full,1}^{-1/2}\widehat{\mathbf{\Sigma}}_{full,1}\mathbf{\Sigma}_{full,1}^{-1/2}\|\|\mathbf{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\|.\end{aligned}$$

Note that

$$\boldsymbol{\mu}_{full,1}^T\mathbf{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1} = \theta_{all} = \boldsymbol{\mu}_f^T\mathbf{\Sigma}_f^{-1}\boldsymbol{\mu}_f + \theta_u.$$

Hence by Assumption 3,

$$\|\mathbf{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\| = \sqrt{\boldsymbol{\mu}_{full,1}^T\mathbf{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}} = O(1). \tag{A.22}$$

By Assumption 6, (A.17) and (A.18), we then get

$$\|III\| = O_p\left(\frac{q_u(\log q_u)}{T}\right) = o_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.23}$$

About term IV, by (A.16), we have

$$IV = \mathbf{\Sigma}_{full,1}^{-1/2}\frac{1}{T}\sum_{t=1}^{T-1}(\mathbf{I} - \mathbf{z}_t\mathbf{z}_t^T)\mathbf{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1} + \frac{1}{T}\mathbf{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}. \tag{A.24}$$

By Assumptions 3 and 6, we have $\|\mathbf{\Sigma}_{full,1}^{-1}\| = O(1)$, and $\|\boldsymbol{\mu}_{full,1}\| = O(\sqrt{q_u})$. Therefore,

$$\frac{1}{T}\|\mathbf{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}\| = O\left(\frac{1}{T}\|\mathbf{\Sigma}_{full,1}^{-1}\|\|\boldsymbol{\mu}_{full,1}\|\right) = O\left(\frac{\sqrt{q_u}}{T}\right) = o\left(\frac{1}{\sqrt{T}}\right). \tag{A.25}$$

For any non-random matrix $\mathbf{A}$,

$$E\Big(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}(\mathbf{I}-\mathbf{z}_t\mathbf{z}_t^T)\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\Big)=\mathbf{0}. \tag{A.26}$$

In addition, for any non-random $(q_u+K)\times(q_u+K)$ matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ (that are no need to be symmetric or positive definite),

$$E(\mathbf{z}_t^T\mathbf{B}_1\mathbf{z}_t\mathbf{z}_t^T\mathbf{B}_2\mathbf{z}_t)=\operatorname{tr}(\mathbf{B}_1)\operatorname{tr}(\mathbf{B}_2)+\operatorname{tr}(\mathbf{B}_1\mathbf{B}_2)+\operatorname{tr}(\mathbf{B}_1\mathbf{B}_2^T). \tag{A.27}$$

By (A.27), we get that

$$\begin{aligned}&E\Big(e_i^T(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1})(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1})^Te_j\Big)\\ =&e_i^T(\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T+2\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T)e_j,\end{aligned} \tag{A.28}$$

where e_i is a length-k vector with the ith element being one and zero elsewhere. By (A.26) and (A.28), we have

$$\begin{aligned}&\operatorname{Var}(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1})\\ =&(\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1})\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T\\ &+\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T.\end{aligned} \tag{A.29}$$

In addition, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&E\Big(\|\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\|^4\Big)\\ =&E\Big((\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t)^2(\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t^T)^2\Big)\\ \le&\sqrt{E\Big((\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t)^4\Big)}\sqrt{E\Big((\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t^T)^4\Big)}.\end{aligned}$$

By Lemma 2.9 of Bai and Silverstein (1998),

$$\begin{aligned}&E\Big((\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t)^4\Big)\\ \le&16E\Big(\big(\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{A}^T\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t-\operatorname{tr}(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A})\big)^4\Big)\\ &+16\Big(\operatorname{tr}(\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A})\Big)^4\\ \le&Ck^4\|\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}\|^4=O(1),\end{aligned}$$

where the last equality holds by Assumption 6 and that $\|\mathbf{A}\|=O(1)$. Similarly,

$$E\Big((\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t^T)^4\Big)=O\Big((\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1})^4\Big)=O(1),$$

where the last equality holds by (A.22). It follows that

$$E\Big(\|\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1/2}\mathbf{z}_t\mathbf{z}_t^T\boldsymbol{\Sigma}_{full,1}^{-1/2}\boldsymbol{\mu}_{full,1}\|^4\Big) = O(1). \tag{A.30}$$

By the Lyapunov Central Limit Theorem,

$$\mathbf{A}(IV) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}_{A,IV}),$$

where $\boldsymbol{\Sigma}_{A,IV} = \lim_{N\to\infty}(\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1})\mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T + \mathbf{A}\boldsymbol{\Sigma}_{full,1}^{-1}\boldsymbol{\mu}_{full,1}\boldsymbol{\mu}_{full,1}^T\boldsymbol{\Sigma}_{full,1}^{-1}\mathbf{A}^T$. Using further the independence between $\widehat{\boldsymbol{\mu}}_{full,1}$ and $\widehat{\boldsymbol{\Sigma}}_{full,1}$, and (A.21), we obtain

$$\mathbf{A}(II + IV) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}_A). \tag{A.31}$$

The desired result follows from (A.14), (A.20), (A.23) and (A.31). □

Appendix B Asset pricing tests of MAXSER-S(10)

In this section, we report the asset pricing tests of the MAXSER-S(10) SDF. First, similar to the test results reported in Table 2, we examine the pricing power of the MAXSER-S(10) SDF in explaining the returns of 188 anomaly portfolios. The results are reported in Table B.1. We see that there are totally 13 (2) anomalies with significant alphas for a threshold $t = 1.96$ ($t = 3$). Comparing with the performance of the MAXSER-S(20) SDF reported in Table 2, we see that MAXSER-S(20) has fewer rejections.

Table B.1: **The number of significant alphas from testing 188 anomalies against MAXSER-S(10).** We examine the explanatory power of the MAXSER-S(10) SDF over 188 anomalies. This table summarizes the total number of significant alphas and the number of significant alphas in each characteristic group. The number in parenthesis is the total number of portfolios in each anomaly group. The number of significant alphas under the threshold $t > 1.96$ or $t > 3$ is reported. The evaluation period is between 2000 and 2021.

	All (188)		frictions (10)		intangibles (30)		investment (29)	
threshold of t	1.96	3	1.96	3	1.96	3	1.96	3
MAXSER-S(10)	13	2	0	0	4	1	3	1
	momentum (41)		profitability (46)		value-growth (32)			
threshold of t	1.96	3	1.96	3	1.96	3		
MAXSER-S(10)	0	0	6	0	0	0		

Next, similar to the test results reported in Table 3 based on the MAXSER-S(20) SDF, we evaluate the explanatory power of the benchmark models over the SDF portfolio constructed from MAXSER-S(10). Table B.2 reports the alpha and its t-statistic from the regressions.

Table B.2: **Testing the MAXSER-S(10) SDF against benchmark models.** This table reports the regression results of the SDF portfolio from MAXSER-S(10) against several benchmark models, using monthly returns from 2000 to 2021. Alpha in percentage and its corresponding t-statistic are reported.

	Alpha (%)	t-statistic
CAPM	2.17	6.27
FF3	2.07	6.31
FF5	1.49	4.57
FF6	1.45	4.52
Q4	1.44	4.67
Q5	1.18	3.71
BS6	1.43	4.66
SY4	1.33	3.56
DHS3	1.40	4.32
KNS	1.28	4.53

Table B.2 shows that the alpha of MAXSER-S(10) is both economically large and statistically significant under all benchmark models. However, comparing with the results in Table 3, we see that the alphas of MAXSER-S(10) are all smaller than those of MAXSER-S(20).

Finally, similar to the results reported in Table 4 for the MAXSER-S(20) SDF, we switch the roles between benchmark models and MAXSER-S(10), and check whether MAXSER-S(10) is able to explain the prevailing pricing factors. Table B.3 reports the alpha and its t-statistic from regressions.

Table B.3: **Testing prevailing pricing factors with the MAXSER-S(10) SDF.** This table reports the regression results of various pricing factors against the MAXSER-S(10) SDF, using monthly returns from 2000 to 2021. Alpha in percentage and its corresponding t-statistic are reported.

Factor	Alpha (%)	t-statistic
Mkt-RF	0.53	1.77
SMB	0.17	0.84
HML	-0.28	-1.38
RMW	0.10	0.53
CMA	0.06	0.47
MOM	-0.06	-0.19
R_ME	0.17	0.81
R_IA	-0.09	-0.66
R_ROE	-0.01	-0.04
R_EG	0.32	1.98
HmLm	0.05	0.16
MGMT	0.06	0.29
PERF	0.32	0.86
PEAD	0.25	1.73
FIN	-0.15	0.53
KNS	0.17	0.64

We see from Table B.3 that most alphas are insignificant at the t-statistic of 1.96. The t-statistic of the expected investment growth factor (R_EG) from the Q5 model is 1.98. Comparing with the results from Table 4, we see that MAXSER-S(20) can better capture the prevailing pricing factors than MAXSER-S(10).