Sparsity-driven factor selection: A time-varying framework for factor zoo screening*

Hugo Finizola Stellet†

This version: July 2023

Abstract

This paper proposes a framework for reducing dimensionality in Stochastic Discount Factor (SDF) models by selecting only the most relevant factors, utilizing sparsity-inducing regression techniques, and separating the time series considered during factor selection from that used for predicting returns. We argue that traditional factor-based asset pricing models assume high sparsity levels arbitrarily and propose ensuring a similar scarcity of factors. To this end, we suggest an alternative criterion for setting the penalization parameter of a shrinkage regression, ensuring the selection of a predefined number of factors. The paper applies the proposed framework to a large set of factors, widely used shrinkage techniques, and various candidate time periods, and demonstrates that even simple regressions can achieve positive results when combined with the proposed methodology. Our work provides a suitable framework for researchers to screen useful factors while using the SDF approach and offers an additional criterion of choice for the penalization parameter, especially useful for ensuring sparsity.

Keywords— Factor investing, SDF, Time-varying asset pricing, Shrinkage penalization

May God forgive those bad people.

Adriano "Imperador" Ribeiro

*I am indebted to Marcelo Medeiros for advising this project. I must also thanks Chuanping Sun for his help. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

†PhD candidate at PUC-Rio.

Contents

1 Introduction

Ever since the proposal of the Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965), factor-like models have been the subject of numerous studies. While the original model only considered the market factor, several refinements have been proposed over time, with the Fama-French 3/5 factors model (see Fama and French, 1993, 2015) being the most notable benchmark for market efficiency.

The search for factors that potentially explain some part of the cross-section of expected returns has led to hundreds of factors being claimed as relevant in the literature. However, Cochrane (2011) argues that researchers may have gone too far in publishing an overwhelming number of factors, rendering it impractical and conceptually inadvisable to consider them all jointly. He refers to this phenomenon as a "factor zoo" and advocates against using too many factors to explain the cross-sectional average of returns.

This criticism has raised an important question: which factors truly matter? With an overwhelming number of potentially relevant factors, this situation presents itself as a high-dimensional problem. One straightforward solution is to use shrinkage techniques to impose sparsity, such as the LASSO (Tibshirani, 1996), the Elastic Net (Zou and Hastie, 2005), the Adaptive LASSO (Zou, 2006) - or any of the many other available techniques.

Feng et al. (2020) tackled the issue of the "factor zoo" by using a double-LASSO selection procedure (see Belloni et al., 2014) favoring a more parsimonious asset pricing model and setting a new benchmark to which newly proposed factors should be compared. However, when factors are highly correlated, the LASSO estimator can be unstable, as pointed out by Kozak et al. (2020) and Figueiredo and Nowak (2016), and other regression techniques may be better suited to the task.

Sun (2023) applied the Ordered Weighted LASSO (OWL, Figueiredo and Nowak, 2016) to "dissect the factor zoo" and identify factors that are jointly able to explain cross-sectional returns. Similarly, Freyberger et al. (2020) used the Adaptive Group LASSO (see Huang et al., 2010) to "dissect non-parametrically the factor zoo". Unlike Feng et al. (2020), both of these studies placed a greater focus on out-of-sample predictability rather than determining a parsimonious model.

The degree to which shrinkage is applied is crucial when utilizing any of the regression techniques mentioned, and researchers must adjust the penalization parameter(s) appropriately. Numerous methodologies have been proposed, with some prioritizing predictability, such as K-fold Cross-Validation (CV, Stone, 1974), and others emphasizing in-sample fits, such as the Bayesian Information Criterion (BIC, Schwarz, 1978) and the Akaike Information Criterion (AIC, Akaike, 1974). However, it is worth noting that none of these commonly used criteria focus on ensuring a minimum level of sparsity.

Our contributions encompass three main fronts. Firstly, we introduce a framework for researchers seeking to address the issue of the factor zoo through the SDF methodology. This framework enables the distinction between the time series used for estimating covariances between factors and test assets to find the relevant factors, and the time series used for returns predictions. We show that if a researcher aims to validate her hypothesis through out-of-sample returns of hedge portfolios, factor selection must be done in a rolling-window scheme. Secondly, we propose a new benchmark for evaluating factor selection models' out-of-sample predictability, inspired by the acclaimed Fama-French 3 factors model (Fama and French, 1993). This benchmark is more stringent than simply verifying the hedge portfolios' resulting Sharpe ratios.

In addition to proposing a framework for selecting factors in the SDF methodology, we also suggest methods for guaranteeing a certain level of sparsity in shrinkage regressions. In our study, we applied basic shrinkage regression techniques (Elastic Net, LASSO, and Adaptive LASSO) to the SDF problem, but found that simply using traditional Bayesian Information Criteria to set the penalization parameter did not result in impressive predictive capability. However, when we enforced a pre-defined level of sparsity - by only including factors with the largest absolute coefficient values - or ensured the selection of the exact number of desired factors - by adjusting the penalization parameter accordingly - we observed significant out-of-sample performance improvements. We were able to generate long-short portfolios with annualized Sharpe ratios as high as 2.00.[1]

After applying our proposed framework and sparsity-guaranteeing mechanisms to the SDF problem, we were able to observe how dissociating the periods used for factor selection and returns prediction impacted the SDF dimensionality problem. Traditionally, the literature has used the same period, usually 120 months, for both selecting the factors and predicting the returns. However, our results suggest that a shorter window of 60 months is more suitable for factor selection, while a longer time series of 180 to 240 months should be considered for predicting assets' returns. We found that this dissociation led to more accurate out-of-sample predictions, improving the overall performance.

This paper is structured into five sections. The current section served as an introduction to the related literature and the achievements of this study. In the second section, we present the framework proposed, which includes the Stochastic Discount Factor model, exploration of the factor zoo high-dimensional environment, traditional and new ways of setting the level of sparsity, and time-variant windows for screening relevant factors - and predicting assets' returns. The third section describes the data used in this paper, as well as the methodology for constructing the anomaly factors and test portfolios. In the fourth section, we present obtained results and propose a stricter benchmark for out-of-sample predictability. Finally, in the conclusion, we summarize the contributions and results.

[1]Net of trading costs.

2 Methodology

We adopt the Stochastic Discount Factor (SDF) model (see Cochrane, 2009), as in recent works by Feng et al. (2020), Freyberger et al. (2020), and Sun (2023), to investigate the joint explanation of cross-sectional returns. We present the SDF model in Subsection 2.1, before examining the similarities and differences between risk price and risk premium. In Subsection 2.2, we introduce the zoo of factors, a high-dimensional environment that poses a clear challenge for factor selection. We explore how sparsity can be imposed in high-dimensional problems in Subsection 2.3 and propose ways of guaranteeing a desired degree of it. Finally, in Subsection 2.4, we introduce a rolling-window approach to the factor selection problem.

2.1 Model setup

Denoting the SDF as m, we start from a linear specification of it,

$$m := r_0^{-1}[1 - b'(f - \mathbb{E}[f])], \tag{1}$$

where r_0^{-1} is a constant zero-beta rate, f is a $K \times 1$ vector of K factor returns and b is the $K \times 1$ vector of SDFs coefficients - interpreted as *risk prices*.

Our goal is to identify relevant factors from the vast array of factors available in the literature by examining their impact on the SDF. Specifically, we seek factors that are responsible for movements in the SDF, as evidenced by their non-zero risk prices, i.e., these risk prices should be reflecting the marginal utility of the factors in explaining the cross-section of average returns.

The existence of useless factors and redundant factors in the high-dimensional environment of factor selection and building is a well-known problem. Useless factors simply do not contain any relevant information for the cross-section of asset returns, are not correlated with useful factors, and hence their prices are zero. Redundant factors, on the other hand, have their effects explained by other, relevant, factors and can be rewritten as a combination of them. In other words, redundant factors have zero risk prices but are correlated with the relevant factors.

The literature also distinguishes what is known as the risk premium, which is measured by the second pass free parameter in Fama and MacBeth (1973)'s regression. The covariance matrix of factor returns determines the relationship between risk price and risk premium, which can be expressed as $\zeta = \mathbb{E}[ff']b$, see Cochrane (2009), where ζ denotes the $K \times 1$ vector of risk premiums.

Despite such an embryonic relationship, price and risk premium have substantially distinct interpretations. The premium is related to an investor's willingness to hedge a certain risk factor, regardless of whether or not this factor helps to price the average cross-section of returns. It is therefore possible for a factor not included in the SDF model to exhibit a non-zero premium if it is correlated with some useful factor(s). However, since our goal is to identify which factors are relevant for pricing the average cross-section of asset returns, our attention is focused on SDF loadings, i.e. risk prices, rather than risk premiums. The challenge lies in selecting only the relevant factors that contain unique information that is not captured by any other factor, and therefore have non-zero risk prices.

Back to the theoretical model, it requires that any admissible value for the SDF factor m satisfies the fundamental asset pricing equation, i.e., $\mathbb{E}[Rm] = 0$, as thoroughly explored in Cochrane (2009). However, it is also noted that this equality may not hold when the SDF factor m is unknown and must be estimated from some model. To address this issue, we define the pricing error $(e(b))$ as the deviation from zero of the fundamental asset pricing equation and denote the SDF as $m(b)$, unknown due to its dependence on the also unknown risk price b. The pricing error can then be written as:

$$\begin{aligned} e(b) &= \mathbb{E}[Rm(b)] \\ &= \mathbb{E}[R]\,\mathbb{E}[m(b)] + \mathbb{C}\mathrm{ov}(R, m(b)) \\ &= r_0^{-1}\,\mathbb{E}[R]\,\mathbb{E}[1 - b'(f - \mathbb{E}[f])] + r_0^{-1}\,\mathbb{C}\mathrm{ov}(R, 1 - b'(f - \mathbb{E}[f])) \\ &= r_0^{-1}(\mathbb{E}[R] - \mathbb{C}\mathrm{ov}(R, f)b) \\ &= r_0^{-1}(\mu_R - Cb), \end{aligned} \tag{2}$$

where R is a $N \times 1$ vector of excess return of N assets, $\mu_R := \mathbb{E}[R]$ is the $N \times 1$ vector of assets' excess return expectation and $C := \mathbb{C}\mathrm{ov}(R, f)$.

As the pricing error quadratic form is defined as $\mathrm{Q}(b) = e(b)'We(b)$, where W is some appropriate $N \times N$ weighting matrix, it is possible to estimate the risk prices b by minimizing their quadratic error $\mathrm{Q}(b)$, as follows:

$$\begin{aligned} \hat{b} &= \arg\min_b \mathrm{Q}(b) \\ &= \arg\min_b [(\mu_R - Cb)'W(\mu_R - Cb)], \end{aligned} \tag{3}$$

leading to

$$\hat{\hat{b}} = (\hat{C}'\hat{W}\hat{C})^{-1}\hat{C}'\hat{W}\hat{\mu}_R, \tag{4}$$

where $\hat{C} = \hat{\mathbb{C}}\mathrm{ov}(R, f) = (1/T)\sum_{t=1}^{T}(R_t - \hat{\mu}_R)(f_t - \hat{\mu}_f)'$, $\hat{\mu}_f = (1/T)\sum_{t=1}^{T} f_t$ and $\hat{\mu}_R = (1/T)\sum_{t=1}^{T} R_t$ - notice that, as a constant, r_0 could be disregarded at the optimization problem. In this specification $\hat{\hat{b}}$ is an empirical estimate of $\hat{b}$, which makes use of sample estimates of C and μ_R.

In choosing a functional form for the weighting matrix W, Ludvigson (2013) proposes two options. When there are plenty of test assets, she recommends a surprisingly simple choice: the identity matrix. This matrix ensures that the weights are not tilted towards any particular subset of test assets, which can be useful in situations where these assets have economic interpretation. Another option is to set $W := \mathbb{E}(RR')^{-1}$, which connects $\mathrm{Q}(b)$ to the Hansen-Jagannathan (H-J) distance. According to Ludvigson (2013), in settings where test assets are somewhat limited (i.e., when K is large compared to N), using the H-J distance leads to more stable estimators.

In addition, it is also noteworthy that our model reflects a projection of the SDF in a specific subspace, as both anomaly factors and test assets are constructed solely from the information on stocks' returns (as discussed in Subsections 3.1 and 3.2). This approximation is widely accepted in the literature that explores the high-dimensionality of the zoo of factors - see Feng et al. (2020); Freyberger et al. (2020); Sun (2023).

In subsection 3.2, we showed that our test portfolios are abundant and have clear economic interpretations. Therefore, the identity matrix is the obvious choice for the weighting matrix W, and we can write the optimization problem as:

$$\hat{\hat{b}} = \arg\min_{b}[(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] \tag{5}$$

2.2 The zoo of factors

The factors considered in this study are constructed based on published asset pricing anomalies, which are defined by Brennan and Xia (2001) as *"statistically significant differences between the realized average returns associated with certain characteristics of securities, or on portfolios of securities formed based on those characteristics, and the returns that are predicted by a particular asset pricing model"*.

However, the literature on factors that supposedly explain the cross-section of expected returns has rapidly expanded, producing hundreds of articles, as shown by Hou et al. (2020), who even compiled a data library of 447 published anomaly variables. Cochrane (2011) labeled this situation as a "zoo of factors", likening the vast variety of animals in a zoo to the myriad of factors - each emitting a particular noise - in the literature. The criticism is that there are now too many factors, and researchers should focus on two aspects. Firstly, they should propose a parsimonious benchmark that new factors should be submitted and, secondly, they should pursue models that more accurately predict the market returns.

In addition, the traditional Fama and MacBeth (1973) regression faces theoretical issues, as in a high-dimensional world, the number of factors (K) is likely to be greater than the number of test assets (N), making the standard Fama-MacBeth approach infeasible. Furthermore, in high-dimensional settings, variables are likely to be correlated, leading to the selection of redundant factors. Additionally, when factors are correlated, the Fama-MacBeth approach may suffer from weak factor identification, as pointed out by Kleibergen (2009). Therefore, alternative methods are required to effectively estimate the factor loadings and test for the significance of the estimated coefficients.

2.3 Achieving sparsity in a high-dimension environment

Regularization and dimension-reducing techniques have emerged as possible solutions to the SDF problem, given that traditional methodologies are unable to survive the curse of dimensionality. However, before we delve into these techniques, it is important to ask ourselves a key design question: is it reasonable to assume sparsity in asset pricing?

According to Feng et al. (2020), the asset pricing literature has implicitly relied on the concept of sparsity for a long time. Researchers typically compare new factors against a model with a small set of control variables, usually the Fama-French three/five-factor models (see Fama and French, 1993, 2015) - sometimes plus momentum (see Jegadeesh and Titman, 1993). These models represent a selection of a few factors from the vast zoo of factors that could be relevant for explaining cross-sectional expected returns. This approach results in a parsimonious representation of the universe of factors. What distinguishes this type of sparsity from a machine learning-based one? We believe that, distinctively from traditional asset pricing models, the former does not have biases toward any explanatory factor.

The machine-learning literature has produced several methodologies that can help tackle the curse of dimensionality. In this paper, we focus on some of the most widely used techniques, such as Elastic Net (eNet), LASSO, and Adaptive LASSO (A-LASSO). These methods work by adding a penalty term $\Omega(b)$ to Equation 5, such that:

$$\hat{\hat{b}} = \arg\min_{b}[(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] + \Omega(b) \tag{6}$$

The definition of $\Omega(b)$ varies for each regression technique. The LASSO estimator (Tibshirani, 1996) includes a $\mathbb{L}_1$ norm penalty function for parameters. On the other hand, the Elastic Net regression (Zou and Hastie, 2005) combines LASSO's $\mathbb{L}_1$ norm penalization with Ridge's (Hoerl and Kennard, 1970) $\mathbb{L}_2$ norm. Lastly, the Adaptive LASSO (Zou, 2006) introduces a weighting vector ($\hat{w}$) to the LASSO penalty. The penalty term for each regression can be expressed as:

- LASSO: $\Omega(b)_{LASSO} = \lambda \sum_{j=1}^{K} |b_j|$
- Elastic Net: $\Omega(b)_{eNet} = \lambda \sum_{j=1}^{K} [(1-\alpha)|b_j|^2 + \alpha |b_j|]$
- Adaptive LASSO: $\Omega(b)_{A-LASSO} = \lambda \sum_{j=1}^{K} \hat{w}_j |b_j|$

Special attention should be given to λ - the penalty parameter. It is a term in the objective function, set by the researcher, that encourages sparsity and prevents overfitting. It dictates how severe the penalization will be, i.e., how much shrinkage will occur. There are several techniques available to help us set penalization parameters, such as K-fold Cross-Validation (CV) and Information Criteria, both Bayesian (BIC) and Akaike (AIC).

Cross-Validation, proposed by Stone (1974), partitions the data into training and validation sets, fits the model on the training set, and evaluates its performance on the validation set. It is a robust technique to avoid overfitting and does not rely on any assumptions about the distribution of the data. On the other hand, Bayesian Information Criterion (BIC) (Schwarz, 1978) and Akaike Information Criterion (AIC) (Akaike, 1974) evaluate the trade-off between model fit and complexity, with BIC tending to choose more parsimonious models than AIC.

One advantage of using Information Criteria over Cross-Validation is that they are computationally efficient and can be applied easily to large datasets. However, Information Criteria rely on strong assumptions about the data distribution and can be sensitive to violations of these assumptions. In contrast, CV is more robust to such violations, at the cost of being computationally expensive and requiring a large sample size to obtain accurate estimates of the prediction error. As suggested in Freyberger et al. (2020), we believe the BIC is more adequate for our application.

We could have employed more complex regression techniques to reduce the dimensionality of the SDF problem. For instance, Sun (2023) used the Ordered-Weighted LASSO (OWL, see Figueiredo and Nowak, 2016), as it should allow for the selection of more correlated regressors, while Freyberger et al. (2020) used the adaptive group LASSO (Huang et al., 2010) to propose a nonparametric method for studying which characteristics provide incremental information for the cross-section of expected returns. Despite the potential benefits of more complex regression techniques, we have chosen to stick with simpler methodologies. Our rationale for this decision is that by using less sophisticated techniques, we can focus on properly setting up our methodology and ensuring that any positive results we obtain are not simply a byproduct of the superior technique applied.

While all the regression techniques discussed above impose some degree of sparsity, it is possible that the selected factors still outnumber what the researcher considers reasonable. In such cases, it may be useful to have a methodology that guarantees a pre-defined degree of sparsity. In the next topics, we will present different approaches to achieving this goal, which we refer to as "forcing" and "ensuring" sparsity.

2.3.1 Forcing sparsity

One way to force sparsity is by considering only the regressors with the larger coefficient magnitudes, as done by Sun (2023) in his out-of-sample exercise. The approach involves determining the maximum number of regressors (n) to be considered in the analysis, and if the shrinkage process selects more than n variables, only the top n variables, ranked accordingly, are retained. To ensure that this approach works properly, it is crucial to ensure that the absolute values of the coefficients are comparable. This can be simply achieved by demeaning the regressors and scaling them to have the same standard deviation.

Forcing sparsity in this manner has its advantages, such as a lesser degree of sparsity, allowing for more drastic variable selection, and its fairly easy implementation. However, this methodology may jeopardize theoretical properties, as arbitrarily disregarding selected regressors gives up the certainty of being backed by the chosen regression's properties.

2.3.2 Ensuring sparsity

Is it possible to ensure a pre-defined degree of sparsity without compromising the theoretical properties of the chosen estimator? In the following paragraphs, we will develop a methodology that addresses this question.

The severity of the selection placed upon candidate regressors is controlled by the penalty parameter(s) (λ for our purposes). While the literature has provided criteria to assist with setting the parameter, the most widely used methods (CV, BIC, and AIC) do not guarantee a specific level of sparsity. Cross-validation focuses on out-of-sample prediction, while Information Criteria, although penalizing the number of selected variables, do not guarantee a significant degree of variable selection.

In contrast to the traditional approach of using existing criteria, we suggest selecting the penalization parameter to achieve a predetermined number of factors - denoted as n. By fixing the number of selected factors, we can maintain the regression properties while ensuring the desired level of sparsity. The researcher can then adjust the penalization parameter dynamically to ensure that exactly n factors remain after the screening process. The penalization parameter (λ) can be set as follows:

- Starting at some reasonable λ, run the shrinkage regression of choice and count how many factors are selected:
  - If the number of selected factors is too small, decrease λ;
  - If the number of selected factors is too large, increase λ.
- Repeat the process until the regression returns the desired number of factors:
  - Notice that the pace at which λ is changing should be adjusted as the number of selected factors is approaching n.

2.4 Time-varying factor selection framework

As researchers, we should always be aware that our study objectives should guide the time frame we consider when estimating the factors-assets covariance matrix ($\hat{C}$): choosing the appropriate time frame is crucial to avoid potential biases and ensure the validity of our findings. For instance, when searching for a time-invariant asset pricing model, we should consider the longest possible time series of returns, as was done by Feng et al. (2020) when examining factor loadings to establish a benchmark for newly proposed factors.

Shifting focus to good out-of-sample returns prediction, Sun (2023) selects variables by estimating $\hat{C}$ using all available data[2] before conducting out-of-sample exercises to evaluate the performance of hedged portfolios constructed based on the selected variables. However, we believe that when the focus is on out-of-sample predictability, using the full panel of data is not recommended as it introduces a look-ahead bias, using "future data" to predict "past" asset returns. We will address this issue later, but in order to do so, we need to properly define how out-of-sample performance should be assessed.

2.4.1 Out-of-sample analysis

We propose an out-of-sample analysis inspired by the approach taken in Freyberger et al. (2020), with an additional rolling window parameter. In their study, Freyberger et al. (2020) estimate the covariance matrix $\hat{C}$ and perform variable selection regressions in rolling windows of 120 months. The selected factor returns are then used as test asset returns regressors in a simple OLS regression, which uses the same 120-month rolling window of data. The OLS coefficients are then used to predict test assets' returns one period ahead, and a trading strategy consisting of hedge portfolios is constructed: if the strategy generates sufficient alpha, the variable selection is considered successful.

To improve on the methodology proposed by Freyberger et al. (2020), we suggest disjointing the rolling windows for estimating $\hat{C}$ and running the shrinkage regressions ($RW_{\hat{C}}$) from the one for fitting the OLS prediction regression (RW_{OLS}). This separation allows for a more accurate evaluation of the Stochastic Discount Factor problem in a time-varying manner. Specifically, it enables exploration of distinct time frames for computing correlations between factors and assets (and selecting relevant factors), and for capturing the relationship between the returns of selected factors and test assets.

The schematic representation of the proposed variable selection framework is presented in Figure 1, and goes as follows:

- To predict asset returns at time T, calculate $\hat{C}$ using data from $T-1$ to $T-RW_{\hat{C}}$;
- Estimate $\hat{\mu}R$ using the same time frame, and fit the chosen shrinkage method - as shown in Equation 6;
- Over a different time frame (from $T-1$ to $T-RWOLS$), regress the returns of selected factors ($Ret_{factor_f,t-1}$) against each test asset's returns ($Ret_{test:asset_i,t}$), delayed by one period, as in:

[2] Sun (2023) also breaks the time series into two disjoint parts when observing the time-varying nature of the selected factors. However, his approach will suffer from the look-ahead bias described in Subsection 2.4.2.

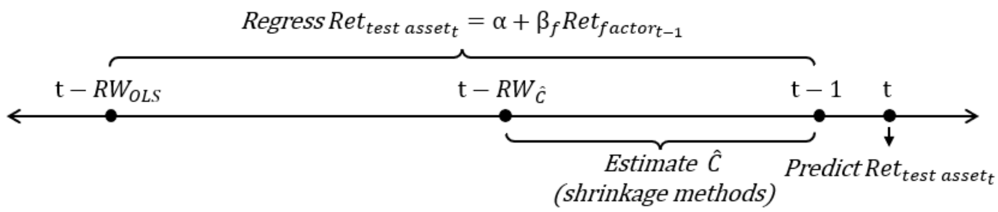


Figure 1: Variable section rolling-window framework scheme

Illustration of the proposed time-variant factor selection framework. In this scenario, factor selection considers data from $t-1$ to $t-RW_{\hat{C}}$ and assets' returns at time t are predicted using a time series from $t-1$ to $t-RW_{OLS}$.

$$Ret_{test:asset_i,t} = \alpha_i + \beta_{i,f} Ret_{factor_f,t-1};$$

- Use the OLS coefficients to project all test assets' returns for time T;
- Build neutral long-short portfolios based on predicted returns, buying (selling) the top (bottom) decile;
- Repeat the process for all possible time frames, storing the long-short portfolio returns;
- Calculate a (some) metric(s) for validating the strategy's performance - in our application, the classic Sharpe Ratio(see Sharpe, 1998).

This methodology allows for the examination of the SDF problem in a variety of applications, providing complete disassociation between the time frame used to compute covariances and select factors, and the time series used to forecast returns based on that selection. This methodology could be used in studies that focus on more stable factor models, using decades of data to estimate covariances, or in studies that use shorter time spans, such as intraday covariance estimations.

2.4.2 Complete panel look-ahead bias

As discussed in Section 2.4, the look-ahead bias arises from the unintended use of data that captures events yet to occur for predicting asset returns.

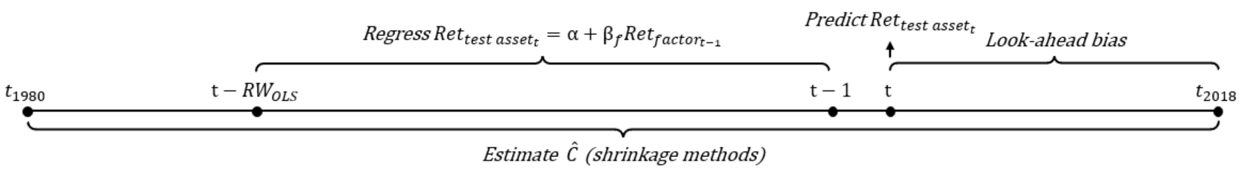


Figure 2: Full panel variable section framework scheme

Illustration of a full panel factor selection framework. In this scenario, factor selection considers all data available (in our application from January 1980, to December 2018) and assets' returns at time t are predicted using a time series from $t-1$ to $t-RW_{OLS}$. The look-ahead bias is characterized by the utilization of data from periods later than t.

In summary, the use of the complete data panel when estimating covariance matrices and factor models introduces a look-ahead bias that can compromise the accuracy of empirical findings. As shown in Figure 2, this bias results from the use of data that is not yet available at the time of predicting asset returns. To avoid this problem, we suggest the use of the methodology described in the previous subsection.

3 Data

We utilize data from the Center for Research in Security Prices (CRSP) and Compustat databases, covering the period from January 1980 to December 2018. The data set consists of 468 months of data on all common stocks listed on NYSE, AMEX, and NASDAQ, covering the same 80 characteristics used by Sun (2023). Risk-free rate and market excess returns are obtained from Kenneth French's online data library. In the first subsection, we detail how the anomaly factors were calculated and provide an overview of the zoo of factors. In the last subsection, we explain how we constructed the test assets.

3.1 Anomaly factors

The market factor aside, we examine 80 additional characteristics as possible regressors. We exclude micro stocks that have a market capitalization less than the 20th percentile of NYSE-listed stocks.[3]

We compute the factors as the spread returns between top and bottom decile portfolios, controlling for size. This approach is like a more tail-oriented version of Fama and French (1993)'s methodology, which uses deciles instead of 30% percentiles. For our estimation, the key consideration is whether the factor survives the dimensionality-lowering procedure, and all factors are calculated on a high-minus-low basis - regardless of whether they are characterized on low-minus-high manners.[4] We demean and adjust all factors to share the same standard deviation as the market factor. This facilitates interpretation and makes the magnitudes of estimated coefficients comparable. Finally, we remove characteristics that cannot produce factors for all available dates.

3.1.1 A tour through the zoo

Now, let's take a tour through the zoo of factors with the help of Sun (2023)'s Figure 3, which displays factor correlation using two distinct measures.

In Figure 3(a), we see the correlations between the β's calculated using the time series methodology presented in Section 2. The figure shows that around 16% of the anomaly factors have absolute

[3]Micro stocks are classified on a monthly basis.

[4]Only the signal of estimated $\hat{b}$ are impacted, and the coefficients' magnitudes remain intact.

Table 1: Anomaly factors

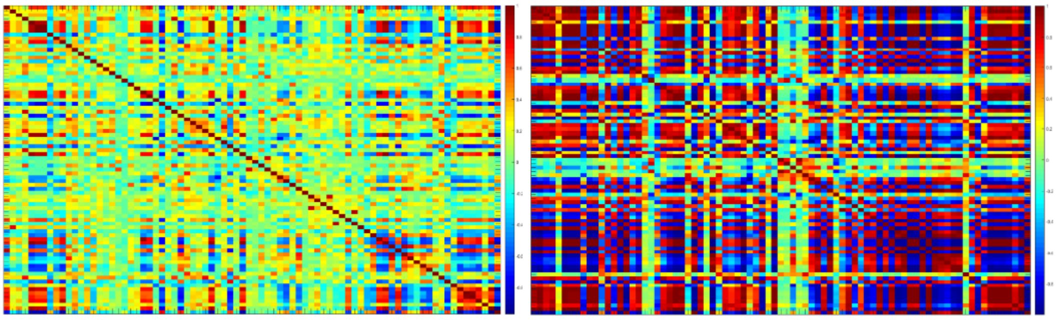
This table lists all used factors. The abbreviation is consistent with Green et al. (2017) and Sun (2023). Detailed information is available at Green et al. (2017).

Abbreviation	Description	Abbreviation	Description
absacc	Absolute accruals	mom1m	1-month momentum
acc	Working capital accruals	mom36m	36-month momentum
aeavol	Abnormal earnings announcement volume	mom6m	6-month momentum
agr	Asset growth	ms	Financial statement score
baspread	Bid-ask spread	mve	Size
beta	Beta	mve_ia	Industry adjusted size
betasq	Beta squared	nincr	Number of earnings increases
bm	Book-to-market	operprof	Operating profitability
bm_ia	Industry adjusted book-to-market	pchcapx_ia	Industry adjusted % change in capital expenditures
cash	Cash holding	pchcurrat	% change in current ratio
cashdebt	Cash flow to debt	pchdepr	% change in depreciation
cashpr	Cash productivity	pchgm_pchsale	% change in gross margin - % change in sales
cfp	Cash flow to price ratio	pchquick	% change in quick ratio
cfp_ia	Industry adjusted cfp	pchsale_pchinvt	% change in sale - % change in inventory
chatoia	Industry adjusted change in asset turnover	pchsale_pchrect	% change in sale - % change in A/R
chcsho	Change in share outstanding	pchsale_pchxsga	% change in sale - % change in SG&A
chempia	Industry adjusted change in employees	pchsaleinv	% change in sales-to-inventory
chinv	Change in inventory	pctacc	Percent accruals
chmom	Change in 6-month momentum	pricedelay	Price delay
chpmia	Industry adjusted change in profit margin	ps	Financial statement score
chtx	Change in tax expense	quick	Quick ratio
cinvest	Corporate investment	retvol	Return volatility
currat	Current ratio	roaq	Return on assets
depr	Depreciation	roavol	Earning volatility
dolvol	Dollar trading volume	roeq	Return on equity
dy	Dividend-to-price	roic	Return on invested capital
ear	Earnings announcement return	rsup	Revenue surprise
egr	Growth in common shareholder equity	salecash	Sales to cash
ep	Earnings-to-price	saleinv	Sales to inventory
gma	Gross proditability	salerec	Sales to receivables
grcapx	Growth in capital expenditure	sgr	Sales growth
grltnoa	Growth in long term net operating assets	sp	Sales-to-price
hire	Employee growth rate	std_dolvol	Volatility of liquidity (dollar trading volume)
idiovol	Idiosyncratic return volatility	std_turn	Volatility of liquidity (share turnover)
ill	Illiquidity	stdacc	Accrual volatility
invest	Capital expenditure and inventory	stdcf	Cash flow volatility
lev	Leverage	tang	Debt capacity/firm tangibility
lgr	Growth in long term debt	tb	Tax income to book income
maxret	Max daily return	turn	Share turnover
mom12m	12-month momentum	zerotrade	Zero trading days

correlation coefficients greater than 0.5: while some factors exhibit significant correlation, there is a notable degree of independence among them. However, if we measure the factors' β's using factor loadings (coefficients of explanatory variables in the second stage Fama-MacBeth regression), Figure 3(b) shows that more than 60% of the correlation coefficients (absolute value) are greater than 0.5, indicating (again) that the Fama-MacBeth regression may encounter severe complications in this application.

3.2 Test assets

The literature offers different views on the ideal set of test assets for SDF models, with some scholars advocating the use of individual stocks and others promoting the utilization of sorted portfo-



(a) Factor correlation measured by time series

(b) Factor correlation measured by factor loadings

Figure 3: Factor correlation coefficients

The heat maps display matrices of correlation coefficients of considered anomaly factors. Measurement "by time series" means that correlation coefficients are computed through factors' time series data, while "by factor loadings" means they are computed through factor loadings obtained after Fama-MacBeth's first-stage regression. Retrieved from Sun (2023)'s Figure 1.

lios. While some studies have used individual stocks (see Harvey and Liu (2021) and Lewellen (2015)), Feng et al. (2020) have argued that characteristic-sorted portfolios have more stable betas, generally present better signal-to-noise ratios, and are more protected from missing data issues.[5] In line with these arguments, we generate our test assets by sorting the stocks into portfolios based on their factors' characteristics.

Following the methodology of Sun (2023), we construct a comprehensive set of bi-variate sorted portfolios as our test assets, in line with the approach proposed by Feng et al. (2020) and Freyberger et al. (2020). The approach involves creating all 5×5 bivariate-sorted portfolios, formed by intersecting stocks' size with each of the 80 characteristics considered in the previous subsection. The construction process is similar to that described in Subsection 3.1 for anomaly factors, but resulting portfolios are both long-only and less extreme.

At the end of the process, any bi-variate portfolio that fails to generate diversified portfolios for all dates of interest will be excluded from the set of test assets. Therefore we end up using a total of 1896 diversified portfolios as the full set of test assets.

4 Empirical analysis

In this section, we apply the methodology presented in Section 2 to scan the "zoo of factors" - and reduce its dimension - using common shrinkage regression techniques such as Elastic Net (eNet), LASSO, and Adaptive LASSO (A-LASSO). We test four different periods - 60, 120, 180, and 240 months - for both $RW_{\hat{C}}$ and RW_{OLS}. Note that our results do not account for trading and slippage costs.

[5]Fama and French (2008) and Hou et al. (2015) have also advocated for the usage of sorted portfolios as test assets.

We begin by presenting the results obtained when all selected factors are considered, before moving on to the approaches for achieving sparsity proposed in Subsection 2.3. The section concludes with a discussion on a new benchmark for assessing the out-of-sample performance of the resulting portfolios.

4.1 Shrinking the Zoo

In this subsection, we explore the efficacy of applying shrinking regressions to reduce the "zoo of factors". We set the penalty parameters using the BIC and include all selected factors to predict returns. Table 2 displays the Sharpe Ratios of the resulting pure alpha trading strategies.

Table 2: Out-of-sample returns prediction results - all factors

This table reports out-of-sample annualized Sharpe ratios of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering all selected factors through a penalization parameter set by BIC, for different combinations of shrinkage estimators, $RW_{\hat{C}}$, and RW_{OLS}.

$RW_{\hat{C}}$	60			120			180			240		
RW_{OLS}	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
60	0.60	0.71	0.24	0.51	0.82	0.74	0.72	0.74	0.41	0.74	0.91	0.43
120	0.82	0.88	0.59	0.69	0.86	0.94	0.91	0.94	0.71	0.82	0.90	0.54
180	0.94	1.02	0.54	0.76	1.06	0.90	1.03	1.05	0.87	0.81	0.81	0.66
240	0.89	1.14	0.43	0.69	0.95	0.86	0.95	0.91	0.76	0.82	0.92	0.60

The results of the less-labored application of the shrinking regressions are not very promising at first glance. The LASSO methodology shows the best Sharpe ratio (1.14, for $RW_{\hat{C}} = 60$, $RW_{OLS} = 240$) as well as the best average Sharpe ratio (0.91, compared to 0.79 and 0.64 for the eNet and the A-LASSO, respectively). However, overall, the resulting portfolios face difficulties in achieving significant Sharpe Ratios - especially the ones generated by the Adaptive LASSO methodology.

The average number of selected factors in each regression/covariance estimation period can shed some light on the impact of sparsity - see Table 3. Comparing Tables 2 and 3, we can see that out-of-sample predictability tends to increase as the number of selected factors decreases. Therefore, imposing a stricter degree of sparsity could potentially lead to better predictability. In the next subsection, we will explore two methods for guaranteeing sparsity, and evaluate their effectiveness in terms of out-of-sample predictability and sparsity.

Table 3: Average number of selected factors

This table reports the average number of selected factors when setting the penalization parameter by BIC, for different combinations of shrinkage estimators and $RW_{\hat{C}}$.

$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Average # of factors	15.6	9.4	22.4	17.3	9.2	20.2	17.6	10.7	21.5	16.6	10.9	20.1

4.2 Guaranteeing sparsity

The sparsity assumption, which posits that only a few factors are truly relevant in explaining asset returns while the rest can be safely disregarded, has been implicitly present in asset pricing models since their inception. For instance, when a new factor is proposed, it is typically tested against a benchmark that only includes a few key factors, such as Fama-French's factor models (Fama and French, 1993, 2015) plus momentum (Jegadeesh and Titman, 1993) - a clearly parsimonious representation of the vast universe of potential factors. However, while shrinkage regression can help achieve sparsity, it does not guarantee it. Therefore, explicitly enforcing a certain sparsity degree in asset pricing models may be beneficial.

Table 4: Out-of-sample returns prediction results - top n factors

This table reports out-of-sample annualized Sharpe ratios of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering only the top n selected factors through a penalization parameter set by BIC, for different combinations of shrinkage estimators, n, $RW_{\hat{C}}$, and RW_{OLS}.

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.31	1.25	1.30	1.25	1.23	0.98	1.00	1.12	1.22	0.98	1.15	1.11
Top 5	1.18	1.06	1.20	1.09	1.14	1.22	0.81	0.93	1.01	1.03	1.11	1.06
Top 10	0.59	0.64	0.65	0.63	0.97	0.87	0.86	0.82	0.75	0.83	1.00	0.64

RW_{OLS}	120											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.27	1.25	1.16	1.36	1.37	1.12	1.09	1.24	1.30	0.95	1.05	1.02
Top 5	1.22	1.12	1.23	1.26	1.36	1.32	0.86	1.05	1.06	0.84	0.86	1.08
Top 10	0.93	0.86	0.80	0.79	1.10	1.15	0.94	1.00	0.98	0.87	0.91	0.68

RW_{OLS}	180											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.50	1.38	1.22	1.46	1.56	1.22	1.27	1.39	1.45	1.13	1.20	1.12
Top 5	1.37	1.26	1.13	1.37	1.50	1.46	1.02	1.22	1.22	0.94	0.91	1.23
Top 10	1.16	1.01	0.71	0.83	1.27	1.15	1.09	1.15	1.12	0.92	0.95	0.79

RW_{OLS}	240											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Top 3	1.54	1.28	1.26	1.39	1.46	1.13	1.30	1.42	1.33	1.20	1.29	1.12
Top 5	1.34	1.20	1.24	1.32	1.45	1.17	1.08	1.16	1.04	1.07	0.98	1.26
Top 10	1.08	1.10	0.85	0.77	1.13	1.01	0.98	1.12	1.00	0.88	1.04	0.85

The notion of sparsity that is widely adopted in asset pricing models is based on traditional and well-tested pricing models. However, it should be possible to achieve similar levels of sparsity through shrinkage regressions, without having to input any prior expectations.

In Subsection 2.3 we defined two methods for explicitly guaranteeing sparsity in asset pricing models. To evaluate their effectiveness, we report Sharpe ratios obtained from portfolios that force sparsity using only the top n factors, sorted by the magnitude of their coefficients, in Table 4.

The results from Table 4 suggest that utilizing only the top n factors significantly improves the observed Sharpe ratios. Additionally, the analysis reveals some interesting patterns. For instance, reducing the number of factors, shortening the period for covariance estimations, and lengthening the period for predicting returns generally enhance out-of-sample performance.

When the top n factors were used for return prediction, the LASSO regression achieved the best performance among the tested methods, with an average Sharpe ratio of 1.15 and a maximum Sharpe ratio of 1.56 (Top 3, $RW_{\hat{C}} = 120$, $RW_{OLS} = 180$). In comparison, the average Sharpe ratios for the eNet and A-LASSO were both 1.08, while the best observed Sharpe ratio for eNet was 1.54 (Top 3, $RW_{\hat{C}} = 160$, $RW_{OLS} = 240$) and 1.46 for A-LASSO (Top 5, $RW_{\hat{C}} = 120$, $RW_{OLS} = 180$).

Although forcing sparsity considering only the top n factors improve results, this method could potentially compromise the regression's properties. To address this issue, we propose ensuring a pre-defined degree of sparsity by adjusting the penalization parameter, as described in Subsection 2.3.2. The results of this approach are presented in Table 5.

Table 5: Out-of-sample returns prediction results - fixed n factors

This table reports out-of-sample annualized Sharpe ratios of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, fixing the number of selected factors as n dynamically setting the penalization parameter, for different combinations of shrinkage estimators, n, $RW_{\hat{C}}$, and RW_{OLS}.

RW_{OLS}	60											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.29	1.38	1.22	1.17	1.12	1.20	1.30	1.24	1.33	1.34	1.23	0.97
Fixed 5	1.11	1.06	1.11	0.91	1.13	1.10	0.79	0.85	1.00	1.00	1.09	0.97
Fixed 10	0.81	0.62	0.74	1.00	0.86	0.78	0.95	0.85	0.92	0.79	0.54	0.54

RW_{OLS}	120											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.53	1.56	1.47	1.21	1.25	1.38	1.39	1.26	1.38	1.34	1.26	0.89
Fixed 5	1.37	1.35	1.41	0.97	1.35	1.19	0.91	0.98	1.11	0.87	1.04	1.13
Fixed 10	0.88	0.91	1.08	1.08	0.97	1.02	0.87	0.87	1.08	0.90	0.73	0.70

RW_{OLS}	180											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.70	1.91	1.67	1.30	1.31	1.56	1.56	1.41	1.50	1.51	1.45	0.93
Fixed 5	1.57	1.72	1.43	1.04	1.46	1.38	1.05	1.10	1.25	1.01	1.23	1.19
Fixed 10	1.13	1.04	1.01	1.12	1.00	1.05	0.95	1.07	1.19	0.91	0.82	0.91

RW_{OLS}	240											
$RW_{\hat{C}}$	60			120			180			240		
	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO	eNet	LASSO	A-LASSO
Fixed 3	1.74	2.00	1.57	1.48	1.33	1.39	1.52	1.49	1.44	1.61	1.58	0.99
Fixed 5	1.56	1.74	1.55	1.05	1.45	1.15	1.11	1.19	1.19	0.96	1.24	1.13
Fixed 10	1.24	1.29	1.05	0.97	0.91	0.99	0.89	0.98	1.10	0.82	0.83	0.88

Our results suggest that adjusting the penalty factors is a much more effective method of ensuring sparsity than simply selecting factors based on their coefficient magnitudes. Fixing the number of factors improves the average Sharpe ratios for all regression methodologies, with the eNet, LASSO, and A-LASSO achieving average Sharpe ratios of 1.16, 1.19, and 1.15, respectively. Moreover, the best results also improve significantly, with the LASSO yielding a top Sharpe ratio of 2.00 (Fixed 3, $RW_{\hat{C}} = 60$, $RW_{OLS} = 240$), the Elastic Net got up to 1.74 (Fixed 3, $RW_{\hat{C}} = 60$, $RW_{OLS} = 240$), and the Adaptive LASSO achieving a ceiling of 1.67 (Fixed 3, $RW_{\hat{C}} = 60$, $RW_{OLS} = 180$).

Our empirical investigation reveals some intriguing patterns. First, our results suggest that factor sparsity is important, as both methods applied show improved out-of-sample results, with Sharpe ratios generally better when considering only three factors. Additionally, shorter periods for evaluating covariances between factors and assets tend to perform better, indicating that these covariances should be observed close to the time of prediction. Finally, it appears that longer periods for regressing selected factors' returns (delayed by one period) over test assets' returns are desirable, indicating that return predictions usually perform better when made over longer time series.

4.3 Fair comparison

Practitioners often use an annualized Sharpe ratio, gross of trading costs, above one as a "rule of thumb" for considering hedge portfolios' returns interesting. However:

- No "rule of thumb" should be accepted in relevant scientific research and;
- As all the factor anomalies considered in our study were found significant returns predictors in other studies, abnormal results may be a byproduct of their documented relevance.

Solely observing the resulting Sharpe ratios of hedge portfolios could raise doubts about the true benefits of guaranteeing sparsity in an SDF model framework. Therefore, in order to compare bananas to bananas, we need to use a more stringent benchmark that is adherent to relevant scientific research.

Table 6: Out-of-sample returns prediction results - benchmark

This table reports out-of-sample annualized Sharpe ratios of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, setting the factors as Fama and French (1993), for different RW_{OLS}.

RW_{OLS}	60	120	180	240
3-factors Fama-French	1.38	1.57	1.64	1.49

One way to achieve a fair benchmark is to mimic usual procedures for testing pricing anomalies while adjusting to our framework. In that spirit, we generated out-of-sample results by fixing the classic 3-factor model (Fama and French, 1993) factors as relevant - the idea is to bring the classic benchmark to our set-up. The performance of this benchmark for all OLS windows considered is presented in Table 6.

After cross-examining the results presented in Tables 2, 4, and 5 with the new benchmark performance on Table 6, we can see that outperforming the proposed benchmark is not a simple task. Simply running the out-of-sample exercise with the 3 Fama-French factors yields Sharpe ratios (gross of trading costs) as high as 1.64 for $RW_{OLS} = 180$. For instance, even the best results for Tables 2 and 4 - 1.14 (LASSO, $RW_{\hat{C}} = 60$, $RW_{OLS} = 240$) and 1.54 (LASSO, Top 3, $RW_{\hat{C}} = 120$, $RW_{OLS} = 180$) respectively - were unable to outperform the Fama-French based selection.

However, several hedge portfolios were able to outperform this stricter proposed benchmark. In fact, as shown in Table 7, these portfolios share some common characteristics:

- Fixed n methodology, with $n \in \{3, 5\}$;
- $RW_{\hat{C}} = 60$ and;
- $RW_{OLS} \in \{180, 240\}$.

Our findings suggest that the zoo of factors can be reduced satisfactorily by utilizing even the most basic dimension-reducing techniques and confronting them with a stricter benchmark. Specifically, our results show that when using monthly data in an SDF framework and considering only eNet, LASSO, and A-LASSO as candidate regression techniques, the most relevant factors should be selected by analyzing a shorter time span of approximately 60 months, predicting returns based on a longer time series exceeding 180 months, assuming significant pre-defined sparsity by selecting fewer than 5 factors, and ensuring set sparsity by adjusting the penalty parameter appropriately.

Table 7: Outperforming hedge portfolios

This table reports the portfolios presented on Tables 2, 4 and 5 that yielded better annualized Sharpe ratios than the best-performing benchmark portfolio - presented on Table 6.

Regression	Methodology	$RW_{\hat{C}}$	RW_{OLS}	SR
LASSO	Fixed 3	60	240	2.00
LASSO	Fixed 3	60	180	1.91
eNet	Fixed 3	60	240	1.74
LASSO	Fixed 5	60	240	1.74
LASSO	Fixed 5	60	180	1.72
eNet	Fixed 3	60	180	1.70
A-LASSO	Fixed 3	60	180	1.67

It is important to acknowledge that utilizing more robust regression methodologies, such as the OWL (see Sun, 2023) or the Adaptive Group LASSO (see Freyberger et al., 2020), in the proposed framework has the potential to yield even more expressive results. However, since this article does not focus on the analysis of those more complex regression models, future research should be undertaken to validate whether our findings hold under those methods.

5 Conclusion

In this paper, we have proposed a framework for addressing the issue of high dimensionality in Stochastic Discount Factor (SDF) models by utilizing penalization regressions, which effectively avoid biases. Our framework can be applied to any horizon of interest and involves separating the time series considered during factor selection from that used for predicting one-period-ahead returns. While we have suggested the use of shrinkage techniques and Ordinary Least Squares (OLS) regression for factor selection and return projection, respectively, our framework can accommodate any other factor selection and prediction methodologies.

Furthermore, we have argued that traditional factor-based asset pricing models assume high sparsity levels arbitrarily, making it possibly interesting to ensure a similar scarcity of factors. To this end, we have shown a previously used methodology and suggested an alternative criterion for setting the penalization parameter of a shrinkage regression: set it dynamically, ensuring that a predefined number of jointly relevant factors are selected. This criterion can be useful when researchers have prior beliefs about the number of factors to be considered.

Additionally, we have proposed a stricter benchmark for evaluating our portfolios' results: test assets are based on anomaly factors that are supposedly relevant, therefore a positive alpha could be a simple byproduct of their construction. This benchmark is a mixture of our out-of-sample framework with the classic Fama-French 3 factors model.

We have applied our methodology to a large set of 80 factors proposed in the literature, plus the market factor, using widely used shrinkage techniques (Elastic Net, LASSO, and Adaptive LASSO), and considering candidate time periods of 60, 90, 120, and 180 months - for both factor selection and returns prediction. When considering all selected factors and setting the penalization parameter through the Bayesian Information Criterion (BIC), the results were not inspiring. However, as we forced sparsity and considered only factors with the highest coefficients, the results improved sensibly. Moreover, when utilizing our methodology to select only a pre-defined number of factors by dynamically setting the penalization parameter, we were able even to surpass the stricter benchmark. We found that shorter windows for factor selection and longer windows for return prediction yield better results - the rationale is that the relevance of factors should be closely related to the present moment, and the relationship between selected factors' returns and test assets' returns should be estimated looking at a longer horizon.

Our work should provide a suitable framework for researchers to screen useful factors while using the SDF approach, and offer them an additional criterion of choice for the penalization parameter - especially useful for ensuring sparsity. Our results demonstrate that even simple regressions, in combination with our proposed framework and penalization criterion, can achieve positive results.

References

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

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

M. J. Brennan and Y. Xia. Assessing asset pricing anomalies. *The Review of Financial Studies*, 14 (4):905–942, 2001.

J. H. Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

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

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

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

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

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

M. Figueiredo and R. Nowak. Ordered weighted l1 regularized regression with strongly correlated covariates: Theoretical aspects. In *Artificial Intelligence and Statistics*, pages 930–938. PMLR, 2016.

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

J. Green, J. R. Hand, and X. F. Zhang. The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, 30(12):4389–4436, 2017.

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

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

K. Hou, C. Xue, and L. Zhang. Replicating anomalies. *The Review of financial studies*, 33(5):2019–2133, 2020.

J. Huang, J. L. Horowitz, and F. Wei. Variable selection in nonparametric additive models. *Annals of Statistics*, 38:2282–313, 2010.

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

F. Kleibergen. Tests of risk premia in linear factor models. *Journal of econometrics*, 149(2):149–173, 2009.

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

J. Lewellen. The cross-section of expected stock returns. critical finance review, 4 (1), 1–44, 2015.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/1924119.

S. C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. *Handbook of the Economics of Finance*, 2:799–906, 2013.

G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

W. F. Sharpe. The sharpe ratio. *Streetwise–the Best of the Journal of Portfolio Management*, 3: 169–185, 1998.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133, 1974.

C. Sun. Factor correlation and the cross section of asset returns: a correlation-robust approach. *Working paper*, 2023.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.