Pontifícia Universidade Católica
do Rio de Janeiro

Hugo Finizola Stellet

Essays concerning high-dimension asset pricing: time-variability, the SDF model error and factors returns' alphas

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor : Prof. Nathalie Christine Gimenes
Co-advisor: Prof. Marcelo Cunha Medeiros

Rio de Janeiro
September 2024

Pontifícia Universidade Católica do Rio de Janeiro

Hugo Finizola Stellet

Essays concerning high-dimension asset pricing: time-variability, the SDF model error and factors returns' alphas

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the Examination Committee.

Prof. Nathalie Christine Gimenes
Advisor
Department of Economics – PUC-Rio

Prof. Marcelo Cunha Medeiros
Co-advisor
Department of Economics – UIUC

Prof. Marcelo Fernandes
São Paulo School of Economics – FGV EESP

Prof. Márcio Gomes Pinto Garcia
Department of Economics – PUC-Rio

Prof. Daniela Kubudi Glasman
Brazilian School of Economics and Finance – FGV EPGE

Prof. Rodrigo dos Santos Targino
School of Applied Mathematics – FGV EMAp

Rio de Janeiro, September 30th, 2024

Hugo Finizola Stellet

BA in Production Engineering from the Federal University of Rio de Janeiro (UFRJ) in 2016.

MSc in Economics from the Brazilian School of Economics and Finance (EPGE) in 2019.

Bibliographic data

Stellet, Hugo Finizola

Essays concerning high-dimension asset pricing: time-variability, the SDF model error and factors returns' alphas / Hugo Finizola Stellet; advisor: Nathalie Christine Gimenes; co-advisor: Marcelo Cunha Medeiros. – Rio de janeiro: PUC-Rio, Departamento de Economia, 2024.

v., 86 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia – Teses. 2. Econometria – Teses. 3. Finanças – Teses. 4. Apreçamento de ativos variável no tempo. 5. Investimento em fatores. 6. Penalização por ecolhimento. 7. Erro do Fator Estocástico de Desconto. 8. Regressões de *span*. I. Gimenes, Nathalie Christine. II. Medeiros, Marcelo Cunha. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 330

To Cleber, his wife, sons, and,
and above all, to his caring mother.

Acknowledgments

To my parents Marcelo and Raquel, and my brothers Pedro and Mateus, for being my cornerstone, the safe harbor in times of need, always giving the purest, unconditional, love - that wouldn't fit even in the largest room in the world;

To my grandparents, Nilse, Beth, and Rodney, for all the lessons, kindness, and Flamengo games watched through all my life. To Marcelina, for all the effort spent on providing me with comfort, always with the warmest smile in the world;

To my co-advisor, Marcelo Medeiros, whom I had the privilege of getting to know during my PhD years, and learning a lot about being an incredible father, caring husband, and wine appreciator (and also a fair bit about econometrics);

To my advisor, Nathalie Gimenes, for the support whenever needed;

To Professor Daniela Kubudi, for being the first person to advise me to pursue my PhD title, and for coming back to my examination committee after being my MSc's advisor - and accepting me as an intern back in 2015;

To professors Marcelo Fernandes, Márcio Garcia, and Rodrigo dos Santos Targino for their participation in the examination committee, as well as for their comments, suggestions, and critiques;

To Fernando Tassinari, for working with me on one of the articles;

To all the professors who have contributed to my academic development and personal growth. Especially for the ones from the Department of Economics at PUC-Rio, who nudged me into the path that led to this dissertation;

To the administrative staff of the Department of Economics at PUC-Rio, I am grateful for their support and assistance;

To all new and old colleagues, whom I have been collecting since graduation days at UFRJ. Each one of you has left a mark on my personal, professional, and academic development, and I will always be grateful for crossing your paths;

Especially to you, Isabelle. "*E eu tô com uma saudade apertada de ir dormir bem cansado, e de acordar do teu lado pra te dizer que eu te amo. Que eu te amo demais.*"

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001

Abstract

Stellet, Hugo Finizola; Gimenes, Nathalie Christine (Advisor); Medeiros, Marcelo Cunha (Co-Advisor). **Essays concerning high-dimension asset pricing: time-variability, the SDF model error and factors returns' alphas**. Rio de Janeiro, 2024. 86p. Tese de doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation comprises two essays on high-dimensional factor-based asset pricing, with an application to the United States equity market. We begin by introducing a time-varying framework to select relevant pricing factors, while actively avoiding biases, and evaluate the effectiveness of applied dimension-reducing techniques. We present the main findings of the articles that constitute this thesis. The first article builds on the Stochastic Discount Factor (SDF) model error to identify the most relevant pricing factors, employing sparsity-inducing regression techniques. Contending that traditional factor-based asset pricing models arbitrarily assume high levels of sparsity, we propose an alternative criterion for determining the penalization parameter in shrinkage regression to ensure similar factor scarcity. Our results demonstrate that even simple regressions can achieve good predictability - and factor sparsity is desirable. In the second paper, we leverage shrinkage techniques on regressions spanning pricing anomalies returns' to identify statistically significant factors. This framework proposes methods for selecting a limited number of impactful factors under sparsity. This approach outperforms the benchmarks, strengthening the sparse pricing anomalies idea. Finally, we compare the results of the two articles' methodologies, highlighting the qualitative aspects of factor selection.

Keywords

Time-varying asset pricing   Factor investing   Shrinkage penalization
Stochastic Discount Factor error   Spanning regressions

Resumo

Stellet, Hugo Finizola; Gimenes, Nathalie Christine; Medeiros, Marcelo Cunha. **Ensaios sobre precificação de ativos em alta dimensão: variações temporais, erro do SDF e alfas dos retornos de fatores**. Rio de Janeiro, 2024. 86p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese é composta por dois ensaios sobre precificação de ativos baseada em fatores em alta dimensão, trazendo uma aplicação ao mercado de ações dos Estados Unidos. Iniciamos introduzindo uma estrutura que comporta viariabilidade temporal para selecionar fatores de precificação relevantes, ativamente evitando viéses, e avalia a eficácia das técnicas aplicadas para redução de dimensionalidade. Em seguida, apresentamos as principais descobertas dos artigos que constituem esta tese. O primeiro artigo parte do erro do modelo de Fator de Desconto Estocástico para identificar os fatores de precificação mais relevantes, empregando técnicas de regressão específicas para induzir esparsidade. Como os modelos tradicionais de precificação de ativos baseados em fatores assumem arbitrariamente altos níveis de esparsidade, propomos um critério alternativo para determinar o parâmetro de penalização na regressão, a fim de garantir uma escassez de fatores semelhante para um modelo de alta dimensão. Nossos resultados demonstram que até mesmo regressões simples podem proporcionar boa previsibilidade - e a esparsidade na seleção de fatores é desejável. No segundo artigo, utilizamos técnicas de encolhimento em regressões que "spanam" retornos de anomalias de precificação para identificar fatores estatisticamente significativos. Esta estrutura enfatiza a esparsidade, propondo métodos para selecionar um número limitado de fatores relevantes. Esta abordagem supera os benchmarks e corroborou com a ideia de esparcidade na seleção de anomalias de precificação. No final, comparamos os resultados das metodologias dos dois artigos, destacando os aspectos qualitativos da seleção de fatores.

Palavras-chave

Apreçamento de ativos variável no tempo Investimento em fatores Penalização por ecolhimento Erro do Fator Estocástico de Desconto Regressões de *span*

Table of contents

Introduction **12**

1 Time-varying asset pricing: A framework **13**

1.1 Introduction 14

1.2 The factor zoo 15

1.3 Factor models' sparsity 18

1.4 Assessing models' results 21

1.5 The time-varying framework 23

1.6 Final considerations 26

2 Applying the time-varying framework on a Stochastic Discount Factor methodology **27**

2.1 Introduction 28

2.2 Methodology 29

2.3 Empirical analysis 35

2.4 Final considerations 40

2.A Appendix 41

3 Modifying traditional low-dimensional methodology to a high-dimensional environment **44**

3.1 Introduction 45

3.2 Methodology 47

3.3 Empirical analysis 53

3.4 Final considerations 62

3.A Appendix 63

4 Which factors matter? **66**

4.1 Introduction 67

4.2 Methodology 68

4.3 Factor hall of fame 72

4.4 Final considerations 78

4.A Appendix 79

Conclusion **81**

List of figures

Figure 1.1 Variable section rolling-window framework scheme 24
Figure 1.2 Full panel variable section framework scheme 25

Figure 3.1 Pricing anomalies accepted by the Variance Inflation Factor control through time 54
Figure 3.A.1 Factors with intercepts' p-values bellow 0.10 - BIC penalty 63
Figure 3.A.2 Factors with intercepts' p-values bellow 0.10 - FRC5 penalty 63

Figure 4.1 Relevant factor categories over time - SDF-error model 76
Figure 4.2 Relevant factor categories over time - factor spanning methodology (FRC5) 77
Figure 4.A.1 Relevant factor categories over time - factor spanning methodology (BIC) 80

List of tables

Table 1.1 Anomaly factors 17
Table 1.2 Out-of-sample returns prediction results - Fama-French benchmark 22
Table 1.3 Out-of-sample returns prediction results - Monte Carlo benchmarks 23

Table 2.1 Out-of-sample results - BIC penalty 35
Table 2.2 Out-of-sample results - FRC penalty considering 5 and 10 factors 37
Table 2.3 Out-of-sample results - FRC penalty considering 1 and 3 factors 38
Table 2.4 Outperforming hedge portfolios 39
Table 2.A.1 Out-of-sample results - BIC penalty 41
Table 2.A.2 Out-of-sample results - FRC penalty considering 5 and 10 factors 42
Table 2.A.3 Out-of-sample results - FRC penalty considering 1 and 3 factors 43

Table 3.1 Summary of pricing anomalies rejected by the Variance Inflation Factor control 55
Table 3.2 Spanning factors returns shrinkage outcome 55
Table 3.3 Out-of-sample results - BIC penalty and p-values under 0.10 57
Table 3.4 Out-of-sample results - BIC penalty and 1, 3, and 5 lowest p-values 58
Table 3.5 Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and p-values under 0.10 59
Table 3.6 Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and 1, 3, and 5 lowest p-values 60
Table 3.7 Outperforming hedge portfolios 61
Table 3.A.1 Out-of-sample results - BIC penalty and p-values under 0.10 64
Table 3.A.2 Out-of-sample results - BIC penalty and 1, 3, and 5 lowest p-values 64
Table 3.A.3 Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and p-values under 0.10 65
Table 3.A.4 Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and 1, 3, and 5 lowest p-values 65

Table 4.1 Anomaly factors - Category classification 71
Table 4.2 Most relevant factors - SDF-error model 72
Table 4.3 Most relevant factors - factor spanning methodology (FRC5) 73
Table 4.4 Most relevant factor categories - SDF-error model 74
Table 4.5 Most relevant factor categories - factor spanning methodology (FRC5) 75
Table 4.A.1 Most relevant factors - factor spanning methodology (BIC) 79
Table 4.A.2 Most relevant factor categories - factor spanning methodology (BIC) 79

“May God forgive those bad people.”

Adriano "Imperador" Ribeiro, *Flamengo 1-0 Vasco on March 14, 2010*.

Introduction

The pursuit of factors that might explain variations in the cross-section of expected returns has yielded an extensive array of proposed factors in the literature. However, Cochrane (2011) critiques this trend, suggesting that the proliferation of factors has become excessive, making it both impractical and conceptually unsound to evaluate them 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.

Cochrane's critique raises an intriguing question: which factors hold genuine significance? The sheer volume of potential factors transforms the challenge of factor selection into a high-dimensional puzzle

This dissertation addresses high-dimensional factor selection in asset pricing through two distinct essays, each investigating the role of sparsity in factor selection using fundamentally different methodologies. Chapter 2 builds upon a specification of the factor model based on a projection of the Stochastic Discount Factor (SDF) onto the subspace of the US equity market (see Feng *et al.*, 2020) while leveraging shrinkage techniques to identify relevant factors. In chapter 3 we adapt traditional low-dimensional methods to a high-dimensional environment to evaluate the statistical significance of factors by spanning each factor's returns against all other available factors and assessing the LASSO's intercept statistical significance (see Jensen *et al.*, 2023).

The remaining chapters examine the convergences and divergences between these methodologies, enhancing our understanding of their distinct implications for factor selection in asset pricing. Chapter 1 introduces the time-varying framework applied in both studies while presenting intuition about factor models' sparsity, factor multicollinearity, and benchmarks to confront obtained results. Finally, Chapter 4 expands the analysis qualitatively by examining the nature of the most significant pricing factors and comparing the outcomes of both methodologies.

1
Time-varying asset pricing: A framework

Abstract. This chapter introduces a flexible, time-varying framework to address the curse of dimensionality in the factor zoo, targeting issues in selecting relevant asset-pricing factors. By separating rolling windows for multicollinearity treatment, factor selection, and returns prediction, the framework avoids look-ahead bias and ensures robustness. It is adaptable to diverse shrinkage techniques, encouraging tailoring to specific research needs. Additionally, two rigorous benchmarks — one based on a classic low-dimensional model and the other on Monte Carlo simulations — are incorporated to assess out-of-sample predictability. We also explore the concept of implicit sparsity in traditional models, proposing methods to achieve similar sparsity levels in high-dimensional contexts. This framework lays the groundwork for two novel dimensionality-reduction approaches evaluated in subsequent chapters.

Keywords: factor asset pricing; time-varying asset pricing; high-dimensionality asset pricing; Sharpe ratio benchmarks.

1.1 Introduction

Cochrane (2011) noted that it is unlikely for all pricing factors proposed in the literature to be jointly relevant for the cross-section of average returns. We will apply two distinct dimension-reducing techniques, to two different interpretations of the factor zoo, to select relevant pricing factors. The factor selections will be evaluated according to their returns forecast capacities.

However, before presenting our shrinkage methodologies (see Chapters 2 and 3) we dedicate this first chapter to properly setting up a *time-varying* framework for managing the dimensionality issue, without incurring any biases.

In contrast to Feng *et al.* (2020), which proposed a parsimonious, *time-invariant* approach to assess the relevance of newly proposed pricing anomalies, Freyberger *et al.* (2020) used the Adaptive Group LASSO (see Huang *et al.*, 2010) in a rolling-window scheme in a study focused on *predictability*, evaluating shrinkage exercises performances according to the observed out-of-sample returns of hedge portfolios built concerning predicted assets returns.

We develop a time-variant framework, enabling the distinction between the time series used for finding the relevant pricing factors and for testing asset returns' predictions. Moreover, we added the possibility of a distinct rolling window for addressing multicollinearity issues. We show that, if the researcher aims to validate her hypothesis through out-of-sample returns of hedge portfolios, factor selection should be conducted within a rolling window scheme to avoid biases - especially *look-ahead bias*.

Furthermore, we propose new benchmarks for evaluating factor selection models' out-of-sample predictability: one inspired by the acclaimed Fama-French 3 factors model (Fama & French, 1993) and another grounded on Monte Carlo simulations.

We also describe two usual methodologies for shrinking the zoo (LASSO and Elastic Net - see Tibshirani (1996) and Zou & Hastie (2005)) and the classic criteria to determine the severity of the shrinkage (CV, BIC, and AIC - see Stone (1974), Schwarz (1978), and Akaike (1974)). However, we believe that there is a *silent* sparsity assumption in classic low-dimensional asset-pricing literature, arguing that it may be beneficial to high-dimensional models to impose similar sparsity degrees.

Contributions for the literature. We can summarize the main contributions of this essay on two fronts. First, this essay advances the literature on time-varying asset pricing by introducing a robust, flexible framework for working within the factor zoo environment. Second, the proposed stricter benchmarks for accessing out-of-sample results free researchers from being subject to a raw Sharpe ratio analysis, offering benchmarks that combine intuitive metrics with quantitative precision.

We also point out minor contributions, as we elaborate an argument for imposing quantitative/high-dimensional models to sparsity levels comparable to classic low-dimension asset pricing, and shed light on a relevant bias both researchers and practitioners might be introducing in their applications.

Outline. This chapter has five more sections in addition to this Introduction. Section 1.2 presents the so-called factor zoo, while Section 1.3 explores its high-dimensionality, comparing it to low-dimensional settings, and introducing candidate frameworks to tackle the curse of dimensionality. Section 1.4 encompasses discussions about how to access models' results, including the ideal set of test assets, metrics to verify predictability, and stricter benchmarks. Finally, Section 1.5 wraps up all relevant aspects into a time-varying setup, and Section 1.6 concludes this introductory chapter.

1.2
The factor zoo

Factor-based asset pricing models have had a groundbreaking impact on the finance field since the introduction of the Capital Asset Pricing Model (CAPM) over half a century ago by Sharpe (1964) and Lintner (1965). Originally solely composed of the market factor, multiple researchers refined the CAPM over the last decades. This core intuition continues in modern models, such as the Fama-French, Carhart, and q-4 factor models (see Fama & French, 1993, 2015; Carhart, 1997; Hou *et al.*, 2015), where asset returns are evaluated based on *relevant* pricing factors.

However, the literature on factors that allegedly 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 the *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. In simpler terms, the abundance of factors calls for researchers to focus on two primary objectives: establishing a parsimonious benchmark for assessing new factors and developing models that accurately predict market returns.

In addition, the traditional Fama & 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, pricing anomalies tend to be highly correlated, which can lead to the selection of redundant factors - especially in high-dimensional settings. Additionally, when factors are correlated, the Fama-MacBeth approach may suffer from weak factor identification - see Kleibergen (2009). Therefore, alternative methods are required to effectively estimate the factor loadings and test for the significance of the estimated coefficients.

1.2.1
Anomaly factors

Factors are constructed based on published asset pricing anomalies, which are defined by Brennan & 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"*.

In addition to the market factor, we examine 80 additional characteristics-based factors as possible regressors - see Table 1.1. We exclude micro stocks with a market capitalization smaller than the 20th percentile of NYSE-listed stocks.[1]

We compute the factors as the spread returns between top and bottom decile portfolios, controlling for size. This approach is akin to a more tail-oriented version of Fama & French (1993)'s methodology, which uses deciles instead of 30% percentiles. As the key consideration is whether the factor survives the dimensionality-lowering procedure, all factors are calculated on a high-minus-low basis - regardless of whether they are characterized as low-minus-high in the literature. We demean and adjust all factors to share the same standard deviation as the market factor. This facilitates interpretation and turns the estimated coefficients' magnitudes comparable.

[1] Micro stocks are classified monthly.

1.2.2
Factor multicollinearity

One of the primary concerns highlighted by Cochrane in his presidential address (Cochrane, 2011) is that is highly unlikely that all proposed pricing factors are jointly relevant in pricing the cross-section of asset returns. This skepticism arises because some factors capture similar qualitative information (e.g., illiquidity and zero trading days), some are combinations of others (e.g., current ratio and percentage change in current ratio), some vary only by a time frame of interest (e.g., 1, 6, 12, 36-month momentum), and some are even squared versions of others (e.g., beta and beta squared).

Table 1.1: Anomaly factors

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

Notes: This table lists all used factors. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024). Detailed information is available at Green *et al.* (2017).

Although multicollinearity does not bias estimated slope coefficients, it inflates their standard errors. For instance, severe multicollinearity poses a challenge for shrinkage methods like the LASSO regression (see Tibshirani, 1996), which may lead to arbitrary exclusion of certain covariates from the model if they are sufficiently correlated. As correlated factors capture similar economic reasoning, it may be interesting to treat possible multicollinearity issues before imposing shrinkage.

In a tour through the zoo of factors, Sun (2024) briefly evaluates factor correlation using two distinct methods: simply using time series of the factors' returns or using factor loadings (coefficients of explanatory variables in the second stage of Fama-MacBeth regression - see Fama & MacBeth (1973)). Sun (2024) shows that overall correlation levels between factors are *substantially higher* when measured using factor loadings, evidencing that Fama-MacBeth's regression may encounter complications for this exercise.

1.3
Factor models' sparsity

Despite recent advances in high-dimensional asset pricing models, such as those by Bryzgalova *et al.* (2023) and Jensen *et al.* (2023), most widely known models are low-dimensional due to either economic fundamentals or technical limitations, such as computational constraints.

In this section, we present what we name a "silent" sparsity assumption taken by traditional models, before providing insights on addressing the dimensionality problem imposed by the vast universe of candidate factors.

1.3.1
"Silent" sparsity assumption

Classic, low-dimensional, asset pricing models - see Fama & French (1993, 2015), Carhart (1997), or Hou *et al.* (2015) - are broadly accepted by both academics and practitioners. Those models are widely used as benchmarks of markets' efficiency, and new pricing factors often have their statistical significance assessed against one of those models.

Given the vast pool of candidate factors introduced in Section 1.2, we argue that low-dimensional models implicitly assume high sparsity by focusing on only a handful of pricing factors - even if not acknowledging it. We believe that this *silent sparsity assumption*, where the focus is only on a handful of pricing factors, might be a relevant source of strength for those models.

According to Feng *et al.* (2020), the asset pricing literature has implicitly relied on the concept of sparsity for a long time. 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, possibly carrying researchers´ personal biases - possibly even unintentionally.

This *sparsity* differs from a machine learning-based approach, which imposes no bias toward any specific explanatory factor.

As the idea of sparsity is widely adopted in asset pricing models, we test two distinct methodologies for imposing sparsity into two distinct higher-dimension settings, turning the silent assumption into an explicit one. The intuition behind this is that instead of qualitatively choosing pricing factors according to some economic/financial intuitions, we rely on past data to select factors relevant to pricing the cross-section of returns.

1.3.2
Tackling zoo's high-dimensionality

Regularization and dimension-reduction techniques emerge as natural solutions to the factor zoo problem, given that traditional methodologies do not survive the *curse of dimensionality*. Machine-learning literature has produced several methods that help tackle the curse of dimensionality. While acknowledging the existence of more elaborated techniques,[2] in this dissertation we focus on two of the most widely used shrinkage techniques: Elastic Net (eNet) and LASSO (see Zou & Hastie (2005) and Tibshirani (1996)). Focusing on less sophisticated regressions will hopefully draw attention to the established time-varying framework.

These methods work by adding a penalty term $\Omega(b)$ to the usual Ordinary Least Squares (OLS) regression, such that:

$$\hat{\beta} = \arg\min_{b}[(Y - Xb)'(Y - Xb)] + \Omega(b) \tag{1.1}$$

The penalty term $\Omega(b)$ differs across regression techniques, defined by the form of shrinkage applied to the model. The LASSO estimator (Tibshirani, 1996) includes a $\mathbb{L}_1$ norm penalty function for parameters. On the other hand, the Elastic Net regression (Zou & Hastie, 2005) combines LASSO's $\mathbb{L}_1$ norm penalization with Ridge's (Hoerl & Kennard, 1970) $\mathbb{L}_2$ norm. The penalty term for each regression can be expressed as:

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

[2]See Figueiredo & Nowak (2016) and Huang *et al.* (2010) for the Ordered-Weighted and Adaptive Group LASSO, respectively.

The shrinkage penalty parameter merits special attention as a crucial component of the objective function, set by the researcher to promote sparsity and prevent overfitting. It dictates how severe the penalization will be, thus regulating the extent of shrinkage. Various techniques exist to aid in setting penalization parameters, including K-fold Cross-Validation (CV) and Information Criteria, such as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC).

Cross-Validation (CV), proposed by Stone (1974), partitions the data into training and validation sets to evaluate model performance and prevent overfitting, making no strong assumptions about data distribution. Conversely, Bayesian Information Criterion (BIC) (see Schwarz, 1978) and Akaike Information Criterion (AIC) (see Akaike, 1974) weigh the trade-off between model fit and complexity, with BIC often favoring more parsimonious models than AIC.

One advantage of employing Information Criteria over Cross-Validation lies in their computational efficiency and ease of application to large datasets. However, Information Criteria hinge on stronger data distribution assumptions and may be sensitive to violations. In contrast, Cross-Validation offers robustness against such violations but demands computational resources and a substantial sample size.

Usually, both methodologies would be well-suited for our applications. However, since the SDF error regression presented in Chapter 2 does not pair well with CV,[3] following Freyberger *et al.* (2020), we adopt BIC as the more suitable method given its computational efficiency and suitability for our applications.

1.3.3
Methodologies for imposing sparsity

Chapter 2. The next chapter exposes a methodology built upon the Stochastic Discount Factor (SDF), approximating the US equity market as a projection, *a subspace*, of the SDF. Factor selection is done directly through shrinkage regressions, where the objective function is obtained algebraically after some manipulation of the SDF pricing error expression.

Chapter 3. The following chapter adapts the usual low-dimensional methodology employed to verify a pricing anomaly relevance, of facing the new factor over a given set of benchmark factors, to higher-dimensional environments. Fac-

[3]As the independent variable is a covariance matrix of factors´ and assets´ returns, properly partitioning the data would require re-estimating the matrix, imposing unnecessary computational costs into the model.

tor selection is accessed by observing the p-values of the intercepts of a series of factors' returns spanning regressions - estimated through LASSO.

1.4
Assessing models' results

We assess the predictability of the chosen factors through an out-of-sample analysis, following Freyberger *et al.* (2020), by examining returns on hedge portfolios constructed from test assets' return predictions based on these selected factors.

There is a divide in the literature on the optimal set of test assets for asset pricing models. While some researchers argue for individual stocks - see Harvey & Liu (2021) and Lewellen (2015) - others favor characteristic-sorted portfolios due to their more stable betas and better signal-to-noise ratios, which help mitigate missing data issues (see Feng *et al.* (2020)).[4] In line with these arguments, we generate our test assets by sorting the stocks into portfolios based on their factors' characteristics.

Using the methodology from Sun (2024), we construct bivariate sorted portfolios by intersecting stock size with each of the 80 characteristics from the prior subsection - in line with Feng *et al.* (2020) and Freyberger *et al.* (2020). These portfolios, formed in a 5×5 grid, are long-only and less extreme (in comparison to anomaly factor portfolios - see Subsection 1.2.1), as thresholds are set two deciles apart instead of one. At the end of the process, any bivariate portfolio that fails to generate diversified portfolios, i.e., portfolios with less than 20 assets, 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.

Freyberger *et al.* (2020) employ the returns of the selected factors, delayed by one period, as predictors for test asset returns in simple OLS regressions conducted over rolling windows of 120 months. Subsequently, they utilize the OLS coefficients to forecast the returns of test assets one period ahead. A trading strategy is then formulated, involving hedge portfolios: assets in the top decile of predicted returns are bought, while those in the bottom decile are sold. If the strategy yields significant alpha, the variable selection is deemed successful.

The one-period delay approach is widely used in the literature, implicitly assuming factor momentum: the best estimate for a factor's return at time t is its return at $t-1$. Given the broad acceptance of factor momentum (see Houweling & Van Zundert, 2017; Gupta & Kelly, 2019), we find this assumption reasonable,

[4] Fama & French (2008) and Hou *et al.* (2015) have also advocated using sorted portfolios.

particularly considering the interpretability of the results.

We adapt the described methodology, allowing for variations in the rolling window sizes. The choice to vary window lengths may reflect different economic implications: shorter windows should capture more immediate market dynamics, while longer windows offer broader datasets, providing more stable estimates.

This methodology provides us with several series of cash-neutral portfolio returns, comparable over many possible performance metrics. Moreover, those test portfolios are both abundant and - as based on securities characteristics - have economic interpretations. Consistent with common practice in empirical asset pricing, we use the Sharpe ratio (see Sharpe, 1998) as the main performance metric, supplemented by the average monthly turnover to approximate for implementation costs.

1.4.1 Benchmarks

Practitioners often use an annualized Sharpe ratio, gross of trading costs, above one as a "rule of thumb" for considering hedge portfolio returns "interesting". Moreover, as all factor anomalies considered were reported as statistically significant return predictors in other studies, abnormal results may be a byproduct of that documented relevance.

Simply evaluating the hedge portfolios' Sharpe ratios could raise doubts about the full benefits of enforcing sparsity, so we introduce more stringent benchmarks to allow for more meaningful comparisons.

Table 1.2: Out-of-sample returns prediction results - Fama-French benchmark

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

Notes: 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 & French (1993), for different RW_{pred}.

Fama-French One way to establish a fair benchmark is to replicate standard procedures for testing pricing anomalies while adjusting them to our framework. In that spirit, we generated out-of-sample results by fixing the factors from the classic 3-factor model (Fama & French, 1993) as predictors - the aim is to incorporate the classic benchmark into our setup. The performance of this benchmark, measured by its out-of-sample Sharpe ratio, for all prediction windows, is presented in Table 1.2.

Monte Carlo simulation Abnormal returns might also be a byproduct of the considered factors returns, rather than from proper factor selection. To further examine this possibility, we conduct eight Monte Carlo simulations, selecting 3 or 5 factors at random in each to assess out-of-sample performance.

Table 1.3 presents the average Sharpe ratios obtained in our Monte Carlo simulations (of 100 repetitions), where we randomly select distinct n factors for each forecasting period.

Table 1.3: Out-of-sample returns prediction results - Monte Carlo benchmarks

RW_{pred}	60	120	180	240
3 factors	1.14	1.30	1.45	1.49
5 factors	0.97	1.13	1.26	1.31

Notes: 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, selecting factors randomly (following usual Monte Carlo approach), for different RW_{pred}.

Both proposed benchmarks could generate higher-than-expected out-of-sample Sharpe ratios, and results slightly favor the Fama-French-based portfolios. Although those outstanding out-of-sample Sharpe ratios are probably a byproduct of the nature of test assets, we believe they provide viable metrics for comparing the results of distinct asset pricing methodologies.

Our factor selection methodologies can only be deemed *significant* if they outperform these stricter benchmarks, ideally across various combinations of selection and forecasting windows (RW_{shrk} and RW_{pred}), addressing concerns about potential data mining biases.

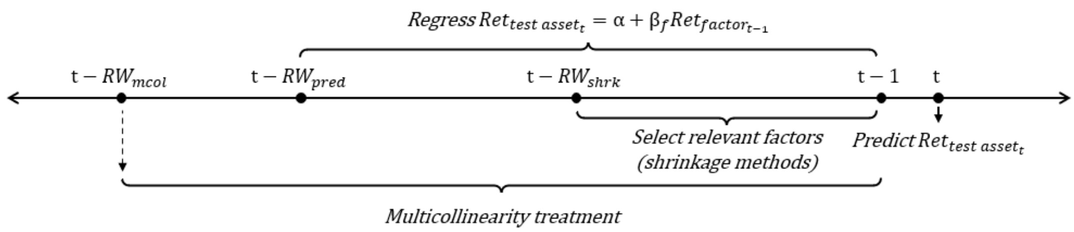
1.5
The time-varying framework

Data. Our dataset, drawn from CRSP and Compustat, spans January 1980 to December 2021, covering all NYSE, AMEX, and NASDAQ-listed common stocks and includes the 80 characteristics used by Sun (2024). Risk-free rate and market excess returns are obtained from Kenneth French's online data library.

Setup. This time-varying framework allows flexible use of multiple time windows, supporting different financial perspectives while avoiding look-ahead bias.

Following Freyberger *et al.* (2020), we separate the rolling windows for each step of our analysis:[5] one for addressing multicollinearity, another for estimating relevant factors, and a final one for conducting out-of-sample prediction regressions - denoted as RW_{mcol}, RW_{shrk}, and RW_{pred}, respectively, as illustrated in Figure 1.1.

Figure 1.1: Variable section rolling-window framework scheme



Notes: Illustration of the proposed time-variant factor selection framework. In this scenario, for forecasting test assets' returns for time t, VIFs are computed with data from $t-1$ to $t-RW_{mcol}$, factor selection considers data from $t-1$ to $t-RW_{shrk}$, and assets' returns at time t are predicted using a time series from $t-1$ to $t-RW_{pred}$.

This segmentation enables a thorough exploration of factor selection by allowing each step — multicollinearity treatment, factor shrinkage, and returns forecasting — to be conducted over different time frames, supporting a more adaptable approach to factor selection.

The schematic representation of the proposed variable selection framework is presented in Figure 1.1. To forecast test assets' returns at time τ:

1. ***Multicollinearity treatment.*** If the chosen methodology involves a multicollinearity filter, run the treatment using data from $t-1$ to $t-RW_{mcol}$;
2. ***Shrinkage.*** For every factor accepted by the multicollinearity treatment, f_i, select relevant factors using data from $t-1$ to $t-RW_{shrk}$;
3. ***Returns forecast.*** Over a different time series, from $t-1$ to $t-RW_{pred}$, regress the returns of selected factors against every test asset's returns, delayed by one period - as described in Section 1.4;
4. ***Hedge portfolios construction.*** Finally, use the OLS coefficients to project test assets' returns for time t: build neutral long-short portfolios based on predicted returns, buying (selling) the top (bottom) decile.

Repeat this process across all periods, storing long-short portfolio returns for performance evaluation metrics (see Section 1.4).

[5] In their article, Freyberger *et al.* (2020) keep the rolling window fixed for all analysis.

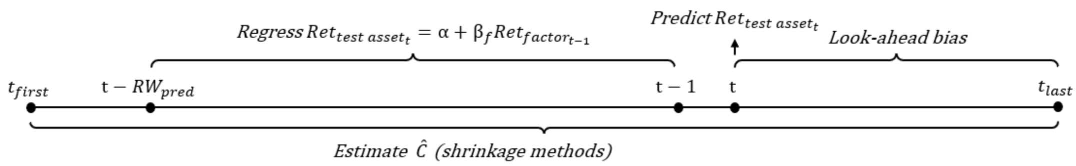
This methodology facilitates examining the factor zoo problem by ensuring a clear separation between the time series used for avoiding major multicollinearity issues, the one to select relevant factors, and the time series utilized to forecast returns based on that selection. It offers versatility, accommodating studies focused on stable factor models spanning decades of data, and those looking at shorter periods such as intraday estimations.

1.5.1 Avoiding look-ahead biases

While verifying our models' forecasting abilities, we should always avoid any source of biases. Time-variant exercises like those in this dissertation must carefully define data samples to ensure valid predictability assessments.

Also focused on good out-of-sample returns prediction, Sun (2024) selects variables using all available data[6] before conducting out-of-sample exercises to evaluate the performance of hedged portfolios built upon selected variables. When focused on predictability, using the full panel for selecting relevant factors is not recommended as it introduces a look-ahead bias, using "future data" to predict "past" asset returns.

Figure 1.2: Full panel variable section framework scheme



Notes: Illustration of a full panel factor selection framework. In this scenario, factor selection considers all data available, and assets' returns at time t are predicted using a time series from $t-1$ to $t-RW_{pred}$. The look-ahead bias is characterized by the utilization of data from periods later than t.

In summary, using the complete data panel when estimating covariance matrices and factor models introduces look-ahead bias and can compromise the accuracy of empirical findings. As illustrated in Figure 1.2, this bias arises from using data that would not have been available at the prediction time.

[6] Sun (2024) also breaks the time series into two disjoint parts when observing the time-varying nature of the selected factors. However, the bias persists within each disjoint sample.

1.6 Final considerations

This essay establishes a framework to address the *curse of dimensionality* in the factor zoo within a time-varying context. We carefully separate the rolling windows for factor selection, forecasting regressions, and addressing issues such as multicollinearity. The framework is adaptable to various shrinkage methods, encouraging researchers to adjust it according to specific applications.

The methodology includes two rigorous benchmarks for out-of-sample predictability: one based on classical low-dimensional models and another using Monte Carlo simulations. We also introduce the concept of an implicit sparsity assumption in classic low-dimensional asset-pricing models, positing that selecting only a few factors from the extensive factor zoo inherently imposes a strong sparsity constraint. High-dimensional models could benefit from this perspective, and we propose two methods for navigating the factor zoo while maintaining comparable sparsity levels.

The framework introduced here will be applied in the next two chapters, where we evaluate the predictive power of two distinct dimensionality-inducing approaches.

2
Applying the time-varying framework on a Stochastic Discount Factor methodology

Abstract. We apply the time-varying framework proposed in Chapter 1 for reducing dimensionality in Stochastic Discount Factor (SDF) models by selecting only the most relevant factors and utilizing sparsity-inducing regression methods. We build upon the argument that traditional factor-based asset pricing models assume high sparsity levels without empirical justification and propose a method to ensure a similar scarcity of factors. To this end, we suggest an alternative criterion for selecting the penalization parameter in shrinkage regression, ensuring the selection of a pre-defined number of factors. The paper applies the proposed framework to a large set of factors, widely used shrinkage techniques, and various candidate time periods. It demonstrates that even straightforward regression methods can yield significant results when applied within the proposed methodology. Our work provides a robust framework for researchers to identify and select relevant factors using the SDF approach while offering an additional criterion of choice for the penalization parameter - which is especially useful for ensuring sparsity. This chapter encompasses key contributions from the article *Sparsity-driven factor selection: A time-varying framework for factor zoo screening*.

Keywords: factor investing; SDF; time-varying asset pricing; shrinkage penalization; high-dimensionality asset pricing.

2.1 Introduction

As discussed in Chapter 1, the criticism of Cochrane's presidential address (see Cochrane, 2011) raised an important question: *which factors truly matter?* Given the overwhelming number of potentially relevant factors, this situation presents itself as a high-dimensional problem.

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 that sets a new benchmark for evaluating newly proposed factors.

However, a limitation of LASSO is that when factors are highly correlated, the estimator can become unstable, as highlighted by Kozak *et al.* (2020) and Figueiredo & Nowak (2016).

Sun (2024) applied the Ordered Weighted LASSO (OWL, Figueiredo & 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 level of shrinkage applied plays a crucial role when using these regression techniques, requiring researchers to adjust penalization parameters carefully. 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.

Contributions for the literature. This essay's main contributions can be summarized along two main fronts. Firstly, we apply the time-varying framework exposed in Chapter 1, which enables the distinction between the time series used for different purposes.

In addition to applying the proposed framework for selecting factors in the SDF methodology, we also suggest methods for guaranteeing a certain *degree of sparsity* in shrinkage regressions - yielding sparsity levels comparable to usual low-dimensional models. In our study, we applied basic shrinkage regression techniques (LASSO and Elastic Net) to the SDF problem. We discovered that using traditional Bayesian Information Criteria to set the penalization parameter does not yield strong predictive performance. However, we observed marked improvements in out-of-sample performance when we enforced the selection of a specified number of factors by adjusting the penalization parameter.

Findings. Among the main results of this essay are:

(i) The proposed methodology consistently outperformed the stricter benchmarks introduced;

(ii) Ensuring higher levels of sparsity, comparable to (or even *stricter* than) low-dimensional models produces higher out-of-sample Sharpe ratios;

(iii) Dissociating the periods used for factor selection and returns prediction positively impacts the SDF dimensionality problem;

(iv) Windows of 60 months are generally more suitable for factor selection;

(v) Longer time series of 180 to 240 months should be considered for forecasting test assets' returns.

Outline. In addition to this Introduction, the current chapter is structured into three more sections. In Section 2.2, we present the chosen methodology, built upon the Stochastic Discount Factor model, how we explore 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. Section 2.3 presents obtained results and faces them with stricter benchmarks for out-of-sample predictability (see Subsection 1.4.1). Finally, in the 2.4, we summarize contributions and results. The Appendix - Section 2.A - offers complementary results for the Elastic Net regression.

2.2 Methodology

We adopt the Stochastic Discount Factor (SDF) model (see Cochrane, 2009), as in recent works by Feng *et al.* (2020) and Sun (2024), to investigate the joint explanation of cross-sectional returns. We present the SDF model in Subsection 2.2.1, before exploring how sparsity can be imposed in high-dimensional problems in Subsection 2.2.2, and proposing a methodology for guaranteeing a certain, *ad hoc* defined, degree of it - see Subsection 2.2.3.

2.2.1 Model setup

Denoting the Stochastic Discount Factor (SDF) as m, we start from a usual specification:

$$m := r_0^{-1}[1 - b'(f - \mathbb{E}[f])], \tag{2.1}$$

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*. Any admissible value for the SDF factor m must satisfy the fundamental asset pricing equation, i.e., $\mathbb{E}[Rm] = 0$.

We aim 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 responsible for movements in the SDF, as evidenced by their non-zero risk prices, i.e., these risk prices should reflect the marginal utility of the factors in explaining the cross-section of average returns.

The existence of useless 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 & 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, the fundamental asset pricing equation may not hold in the real world, where 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:

$$\begin{aligned} e(b) &= \mathbb{E}[Rm(b)] \\ &= \mathbb{E}[R]\,\mathbb{E}[m(b)] + \mathbb{C}\text{ov}(R, m(b)) \\ &= r_0^{-1}\,\mathbb{E}[R]\,\mathbb{E}[1 - b'(f - \mathbb{E}[f])] + r_0^{-1}\,\mathbb{C}\text{ov}(R, 1 - b'(f - \mathbb{E}[f])) \\ &= r_0^{-1}(\mathbb{E}[R] - \mathbb{C}\text{ov}(R, f)b) \\ &= r_0^{-1}(\mu_R - Cb), \end{aligned} \tag{2.2}$$

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 $C := \mathbb{C}\text{ov}(R, f)$.

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

$$\begin{aligned} \hat{b} &= \arg\min_b \mathsf{Q}(b) \\ &= \arg\min_b [(\mu_R - Cb)'W(\mu_R - Cb)], \end{aligned} \tag{2.3}$$

leading to

$$\hat{b} = (\hat{C}'\hat{W}\hat{C})^{-1}\hat{C}'\hat{W}\hat{\mu}_R, \tag{2.4}$$

where $\hat{C} = \hat{\mathbb{C}}\text{ov}(R, f) = (1/T)\sum_{t=1}^{T}(R_t - \hat{\mu}_R)(f_t - \hat{\mu}_f)'$, $\hat{\mu}_f = (1/T)\sum_{t=1}^{T} f_t$ 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, $\hat{b}$ 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 $\mathsf{Q}(b)$ 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 noteworthy that our model approximates the SDF through a projection onto a specific subspace, as both anomaly factors and test assets are constructed solely from the information on stocks' returns. 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 (2024).

In Section 1.4, we showed that our test portfolios are abundant and have clear economic interpretations. Therefore, the identity matrix is the clear choice for the weighting matrix W, and we can write the optimization problem as:

$$\hat{b} = \arg\min_{b}[(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] \tag{2.5}$$

2.2.2 Achieving sparsity in a high-dimension environment

As discussed in Section 1.3, estimating Equation 2.5's $\hat{b}$ coefficients using simple low-dimension techniques could be problematic, introducing possible biases and overfitting issues. Now we lean into machine-learning literature techniques to estimate them, using 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 2.5, such that:

$$\hat{b} = \arg\min_{b}[(\hat{\mu}_R - \hat{C}b)'(\hat{\mu}_R - \hat{C}b)] + \Omega(b), \tag{2.6}$$

where $\Omega(b)$ varies depending on the regression technique - as discussed in Subsection 1.3.2.

The penalty term $\Omega(b)$ is accompanied by a penalty parameter, λ, that must be set by the researcher. In Subsection 1.3.2, we show the $\Omega(b)$ forms for the most usual shrinkage regressions and classic methodologies for setting the λ value, concluding that the Bayesian Information Criteria (BIC) emerges, for our application, as the most suitable methodology for setting up our penalty parameter.

We could have employed more complex regression techniques to reduce the dimensionality of the SDF problem. For instance, Sun (2024) used the Ordered-Weighted LASSO (OWL, see Figueiredo & 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 techniques, we chose to use simpler regression techniques to focus on establishing the time-varying methodology (and the upcoming criterion to ensure sparsity), ensuring that the results obtained are not a product of the possibly 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.

One way to force sparsity is by considering only the regressors with the larger coefficient magnitudes, as done by Sun (2024) 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.

Forcing sparsity in this manner has advantages, allowing for easily implementing more drastic variable selection. However, this methodology may compromise the shrinkage theoretical foundation, as arbitrarily disregarding selected regressors gives up the certainty of being backed by the chosen regression's properties.

2.2.3 Ensuring sparsity: The Fixed Regressors Criterion

Can a pre-defined level of sparsity be ensured without compromising the theoretical properties of the estimator?

To address this challenge, we propose a distinct criterion for setting the penalization parameters in shrinkage regressions. The objective is to guarantee that a specified number of regressors (as determined by the researcher) will have non-zero coefficients. This criterion proves especially advantageous in the realm of factor-based asset-pricing models, offering *ad-hoc* means to ensure the implicit sparsity assumption inherent in models like the Fama-French 3/5, Carhart, and q-4 factors (Fama & French, 1993, 2015; Carhart, 1997; Hou *et al.*, 2015), by selecting a penalty parameter that precisely returns some desired number of non-zero coefficient regressors.

The Fixed Regressors Criterion (FRC) determines the penalty parameter (λ) through the following algorithm:

- Initialize an array of candidate values and perform the shrinkage regression on them;
- Ensure that the desired number of final regressors is included in the range of candidate λ's:
  - If the number of selected factors for the highest (lowest) candidate λ is too low (high), adjust for lower (higher) values. Re-run the regression for this new set of candidate λ's;
- If no value precisely returns the desired number of regressors, narrow the range:
  - Set the new highest (lowest) possible value as the higher (lower) penalization that yields fewer (more) than the desired number of factors.
  - Re-run the regression for this new array of candidate penalization parameters;
- Repeat the above steps until at least one candidate value returns exactly the desired number of factors with non-zero coefficients:
  - As multiple penalization values may return the set number of relevant regressors, choose the median of these values as the penalization parameter.

While implementing the FRC, the researcher should consider the specifics of the chosen estimation technique, as some do not guarantee a monotonically increasing number of selected factors as the penalization parameter decreases. Additionally, although unlikely, we cannot guarantee that the exact desired number of regressors will be selected for any given penalty value. We recommend selecting the median value from the highest penalization sequence of candidate values to address the non-monotonicity issue and ceasing the search for the perfect lambda after several iterations, opting for the highest penalization that yields one more relevant factor.[1]

[1]Concerning any additional factors, the researcher may choose what to do. If the data is scaled and the absolute values of the coefficients are comparable, we suggest disregarding the factor with the lower $|\beta|$, as suggested by Sun (2024). However, disregarding selected regressors means not necessarily being supported by the chosen regression properties.

2.3
Empirical analysis

We apply the factor selection methodology from Section 2.2 to the time-varying framework and data described in Section 1.5, to scan the "zoo of factors" - and reduce its dimension - using the LASSO shrinkage regression technique.[2] Using BIC, we determine the penalization parameter by testing 100 values ranging from 10^{-7} to 10^{-1}.[3] We test four periods - 60, 120, 180, and 240 months - for both RW_{shrk} and RW_{pred}. Beware that our results do not account for trading and slippage costs.

2.3.1
Shrinking the Zoo

We explore the efficacy of applying shrinking regressions to reduce the complexity of the "zoo of factors". We begin setting the penalty parameter using BIC. Table 2.1 displays the out-of-sample results of the pure alpha, cash-neutral, trading strategies.

Table 2.1: Out-of-sample results - BIC penalty

RW_{shrk} 60					RW_{shrk} 180				
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0091	0.0200	1.57	2.51	240	0.0078	0.0283	0.95	2.41
180	0.0091	0.0205	1.53	2.52	180	0.0079	0.0287	0.95	2.44
120	0.0080	0.0200	1.39	2.57	120	0.0070	0.0282	0.85	2.49
60	0.0076	0.0226	1.16	2.62	60	0.0044	0.0283	0.54	2.56

RW_{shrk} 120					RW_{shrk} 240				
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0069	0.0281	0.85	2.24	240	0.0069	0.0281	0.85	2.65
180	0.0067	0.0289	0.81	2.24	180	0.0066	0.0280	0.82	2.72
120	0.0055	0.0291	0.66	2.35	120	0.0056	0.0278	0.70	2.81
60	0.0040	0.0319	0.43	2.40	60	0.0051	0.0315	0.56	2.94

Notes: 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 LASSO, a penalization parameter set by BIC, for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

At first glance, the shrinkage regression results may not appear particularly promising. The best Sharpe Ratio obtained (1.57, for $RW_{shrk} = 60$, $RW_{pred} = 240$) is not able to surpass the benchmarks proposed - see 1.4.1. However, it is possible

[2]See Appendix 2.A.1 for Elastic Net (eNet) regression results.

[3]We also guarantee that the highest penalty returns at least one regressor and that the final ratio between the highest and lowest tested values is 10^3.

to observe a pattern, as Sharpe ratios appear to improve for *shorter* shrinkage, and *longer* prediction, rolling windows.

2.3.2 Guaranteeing sparsity

As Section 1.3.1 states, a sparsity assumption that suggests that only a few factors are *truly relevant* in explaining asset returns, while the rest can be 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 - a parsimonious representation of the vast universe of potential factors. However, while shrinkage regression, with hyper-parameters defined through BIC, can help achieve sparsity, it does not guarantee it. Therefore, explicitly enforcing a certain sparsity degree in asset pricing models may be beneficial.

Ensuring sparsity We now report results obtained ensuring sparsity utilizing the Fixed Regressors Criterion (FRC), as described in Section 2.2.3 - see Tables 2.2 and 2.3.

Table 2.2 reports results obtained when applying the FRC for less severe dimensional reduction. Notably, results generally improve when using the proposed methodology, even *surpassing the Fama-French benchmark* in some cases - FRC5, $RW_{shrk} = 60$ and $RW_{pred} \in \{180, 240\}$. Moreover, predictability improves when selecting factors over a shorter window, forecasting returns considering a longer window and considering a *lower number of factors*.

The role of sparsity is even more highlighted in Table 2.3, where we report out-of-sample results obtained when imposing factor selection while fixing only one or three pricing anomalies as relevant: FRC3 surpasses the benchmark's Sharpe ratio of 1.64 three times ($RW_{shrk} = 60$ and $RW_{pred} \in \{120, 180, 240\}$), while FRC1 poses an incredible performance, having *fourteen out of its sixteen portfolios beating all proposed benchmarks*.

On the one hand, the pattern of parameter sensibility observed in Tables 2.1 and 2.2 generally holds for Table 2.3, with FRC1 producing the best results, accompanied by shorter shrinkage and wider forecasting periods. On the other hand, if that pattern should hold, the best Sharpe ratio should have been observed for FRC1, $RW_{shrk} = 60$ and $RW_{pred} = 240$, which is not true: the single best performance gotten setting FRC1 and $RW_{pred} = 240$, as expected, but associated with $RW_{shrk} = 180$.

Table 2.2: Out-of-sample results - FRC penalty considering 5 and 10 factors

	FRC 5					FRC 10			
RW_{shrk}	60				RW_{shrk}	60			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0122	0.0226	1.87	2.36	240	0.0099	0.0240	1.43	2.78
180	0.0122	0.0235	1.79	2.43	180	0.0091	0.0248	1.27	2.82
120	0.0107	0.0235	1.58	2.54	120	0.0081	0.0244	1.15	2.89
60	0.0090	0.0243	1.29	2.67	60	0.0063	0.0248	0.88	3.04
RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0103	0.0224	1.59	2.44	240	0.0072	0.0255	0.98	2.85
180	0.0098	0.0243	1.40	2.49	180	0.0061	0.0275	0.76	2.87
120	0.0089	0.0244	1.27	2.53	120	0.0051	0.0291	0.61	2.88
60	0.0078	0.0262	1.03	2.68	60	0.0039	0.0285	0.48	3.01
RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0083	0.0228	1.26	2.43	240	0.0074	0.0251	1.02	2.81
180	0.0078	0.0253	1.07	2.42	180	0.0070	0.0257	0.95	2.84
120	0.0072	0.0265	0.94	2.47	120	0.0056	0.0266	0.73	2.92
60	0.0064	0.0270	0.83	2.60	60	0.0061	0.0276	0.76	3.07
RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0084	0.0246	1.18	2.45	240	0.0066	0.0256	0.89	2.78
180	0.0092	0.0259	1.23	2.44	180	0.0070	0.0262	0.92	2.82
120	0.0080	0.0259	1.08	2.53	120	0.0063	0.0264	0.83	2.89
60	0.0087	0.0270	1.11	2.69	60	0.0042	0.0278	0.53	3.04

Notes: 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 LASSO, a penalization parameter set by FRC (fixing 5 and 10 regressors), for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

This dissonant result is interesting, as it indicates that it is possible to obtain strong predictability over distinct time frames utilizing the proposed methodology.

Since predictability tends to improve with fewer factors, shorter shrinkage periods, and longer prediction windows, we advise practitioners to focus on combinations in this direction, instead of pursuing the best-performing combination of parameters. We believe that the combination of FRC1, $RW_{shrk} = 60$ and $RW_{pred} = 240$ has more evidence backing its results.

Finally, for those interested in trading costs, the turnover is also impacted by parameters' variability, decreasing with the number of considered factors and shrinkage window length, and increasing with the OLS window - unveiling the same pattern observed for out-of-sample performance. A smaller turnover associated with a lower number of relevant factors is interesting and could be a

Table 2.3: Out-of-sample results - FRC penalty considering 1 and 3 factors

FRC 1					**FRC 3**				
RW_{shrk}	60				RW_{shrk}	60			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0122	0.0201	2.09	1.24	240	0.0127	0.0214	2.06	2.02
180	0.0118	0.0209	1.95	1.31	180	0.0125	0.0217	2.00	2.12
120	0.0109	0.0222	1.71	1.35	120	0.0110	0.0219	1.74	2.16
60	0.0121	0.0231	1.81	1.54	60	0.0106	0.0232	1.58	2.38
RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0115	0.0196	2.04	1.28	240	0.0095	0.0227	1.44	2.07
180	0.0112	0.0211	1.84	1.30	180	0.0093	0.0244	1.32	2.13
120	0.0106	0.0227	1.62	1.40	120	0.0089	0.0252	1.23	2.19
60	0.0107	0.0251	1.48	1.54	60	0.0091	0.0263	1.19	2.33
RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0126	0.0199	2.19	1.33	240	0.0096	0.0207	1.61	2.13
180	0.0126	0.0215	2.04	1.35	180	0.0091	0.0229	1.38	2.14
120	0.0123	0.0231	1.85	1.38	120	0.0082	0.0242	1.17	2.15
60	0.0129	0.0250	1.79	1.49	60	0.0087	0.0247	1.22	2.29
RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0112	0.0202	1.93	1.34	240	0.0106	0.0226	1.63	2.08
180	0.0118	0.0226	1.81	1.40	180	0.0108	0.0248	1.51	2.11
120	0.0114	0.0229	1.73	1.50	120	0.0097	0.0248	1.36	2.20
60	0.0128	0.0249	1.77	1.64	60	0.0100	0.0275	1.26	2.37

Notes: 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 LASSO, a penalization parameter set by FRC (fixing 1 and 3 regressor(s)), for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

byproduct of the persistence of relevant factors.

Our empirical investigation reveals some intriguing patterns. First, our results suggest that factor sparsity is important, with Sharpe ratios generally improving when considering fewer relevant factors. Additionally, shorter periods for selecting factors tend to perform better, indicating that their relevance should be observed close to the prediction time, i.e., shrinkage should be done in smaller windows. 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.

2.3.3 Facing benchmarks

After cross-examining the results presented in Tables 2.1, 2.2, and 2.3 with the new benchmark performance on Table 1.2, we can see that outperforming the proposed benchmark is not a simple task. 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 $RW_{pred} = 180$. For instance, no combination of parameters using less severe shrinkage penalties - BIC and FCR10 - surpassed this benchmark.

Table 2.4: Outperforming hedge portfolios

Penalty criterion	RW_{shrk}	RW_{pred}	Mean	SD	Sharpe	Turnover
FRC1	180	240	0.0126	0.0199	2.19	1.33
FRC1	60	240	0.0122	0.0201	2.09	1.24
FRC3	60	240	0.0127	0.0214	2.06	2.02
FRC1	120	240	0.0115	0.0196	2.04	1.28
FRC1	180	180	0.0126	0.0215	2.04	1.35
FRC3	60	180	0.0125	0.0217	2.00	2.12
FRC1	60	180	0.0118	0.0209	1.95	1.31
FRC1	240	240	0.0112	0.0202	1.93	1.34
FRC5	60	240	0.0122	0.0226	1.87	2.36
FRC1	180	120	0.0123	0.0231	1.85	1.38
FRC1	120	180	0.0112	0.0211	1.84	1.30
FRC1	60	60	0.0121	0.0231	1.81	1.54
FRC1	240	180	0.0118	0.0226	1.81	1.40
FRC5	60	180	0.0122	0.0235	1.79	2.43
FRC1	180	60	0.0129	0.0250	1.79	1.49
FRC1	240	60	0.0128	0.0249	1.77	1.64
FRC3	60	120	0.0110	0.0219	1.74	2.16
FRC1	240	120	0.0114	0.0229	1.73	1.50
FRC1	60	120	0.0109	0.0222	1.71	1.35

Notes: This table reports the portfolios presented on Tables 2.1, 2.2, and 2.3 that yielded better annualized Sharpe ratios than the best-performing benchmark portfolio - presented on Table 1.2.

However, several hedge portfolios were able to outperform this stricter proposed benchmark. As shown in Table 2.4, our findings suggest that the zoo of factors can be reduced satisfactorily by utilizing even the most basic dimension-reducing techniques - even after confronting them with a strict benchmark. Specifically, our results show that when using monthly data in an SDF framework shrunk by the LASSO, the most relevant factors should be selected by analyzing a shorter period of approximately 60 months, predicting returns based on a longer time series exceeding 180 months, and assuming significant pre-defined sparsity by selecting fewer than 5 factors.

2.4 Final considerations

In this chapter, we applied the time-varying framework introduced in Chapter 1 to address the high dimensionality challenge in Stochastic Discount Factor (SDF) models through penalization regressions.

We explored the assumption in traditional factor-based asset pricing models, where high sparsity among factors is often presumed. This assumption led us to suggest a more flexible criterion for setting the penalization parameter in shrinkage regressions: dynamically adjusting it to select a predetermined number of jointly relevant factors. This approach can benefit researchers with specific prior beliefs about the number of factors to retain.

Our methodology was tested on a dataset comprising 80 established factors - plus the market factor. Using the LASSO, a widely recognized shrinkage technique, we evaluated periods of 60, 120, 180, and 240 months for both factor selection and return prediction. When applying the traditional Bayesian Information Criterion (BIC) to set the penalization parameter, the results were underwhelming. However, by implementing our Fixed Regressors Criterion (FRC) to dynamically select a specified number of factors, we achieved superior results, even outperforming all proposed stricter benchmarks.

Our findings indicate that shorter windows for factor selection and longer windows for return prediction yield more favorable outcomes. This result aligns with the hypothesis that factor relevance is best identified in the near term, while the relationship between selected factors' returns and test assets' returns benefits from estimation over a longer historical horizon.

2.A
Appendix

2.A.1
Elastic Net results

Table 2.A.1: Out-of-sample results - BIC penalty

RW_{shrk}	60			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0091	0.0215	1.47	2.56
180	0.0093	0.0233	1.38	2.56
120	0.0091	0.0224	1.40	2.58
60	0.0090	0.0233	1.33	2.61

RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0082	0.0292	0.98	2.18
180	0.0082	0.0295	0.96	2.21
120	0.0073	0.0291	0.87	2.24
60	0.0070	0.0294	0.82	2.30

RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0087	0.0256	1.17	2.04
180	0.0081	0.0275	1.02	2.04
120	0.0072	0.0271	0.92	2.14
60	0.0073	0.0286	0.89	2.28

RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0063	0.0273	0.80	2.54
180	0.0061	0.0276	0.77	2.62
120	0.0049	0.0288	0.59	2.70
60	0.0042	0.0312	0.47	2.89

Notes: 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 Elastic Net, a penalization parameter set by BIC, for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

Table 2.A.2: Out-of-sample results - FRC penalty considering 5 and 10 factors

FRC 5					**FRC 10**				
RW_{shrk}	60				RW_{shrk}	60			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0109	0.0229	1.65	2.36	240	0.0101	0.0253	1.38	2.75
180	0.0108	0.0235	1.59	2.45	180	0.0092	0.0257	1.25	2.82
120	0.0095	0.0240	1.37	2.53	120	0.0075	0.0251	1.04	2.87
60	0.0078	0.0235	1.15	2.73	60	0.0067	0.0233	1.00	3.03
RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0081	0.0242	1.16	2.40	240	0.0073	0.0228	1.11	2.75
180	0.0080	0.0260	1.07	2.46	180	0.0066	0.0250	0.92	2.82
120	0.0066	0.0262	0.87	2.53	120	0.0059	0.0265	0.77	2.87
60	0.0062	0.0265	0.81	2.67	60	0.0051	0.0250	0.71	3.03
RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0080	0.0239	1.15	2.41	240	0.0071	0.0260	0.95	2.75
180	0.0078	0.0265	1.02	2.43	180	0.0068	0.0281	0.83	2.82
120	0.0069	0.0276	0.86	2.47	120	0.0063	0.0293	0.74	2.87
60	0.0065	0.0271	0.83	2.60	60	0.0069	0.0277	0.86	3.03
RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0072	0.0236	1.05	2.40	240	0.0066	0.0263	0.86	2.80
180	0.0081	0.0256	1.10	2.42	180	0.0075	0.0271	0.96	2.83
120	0.0073	0.0259	0.98	2.50	120	0.0074	0.0270	0.95	2.87
60	0.0080	0.0264	1.05	2.66	60	0.0054	0.0288	0.65	3.04

Notes: 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 Elastic Net, a penalization parameter set by FRC (fixing 5 and 10 regressors), for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

Table 2.A.3: Out-of-sample results - FRC penalty considering 1 and 3 factors

FRC 1					**FRC 3**				
RW_{shrk}	60				RW_{shrk}	60			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0122	0.0201	2.09	1.24	240	0.0112	0.0210	1.85	2.00
180	0.0118	0.0209	1.95	1.31	180	0.0108	0.0217	1.73	2.10
120	0.0109	0.0222	1.71	1.35	120	0.0095	0.0211	1.57	2.14
60	0.0121	0.0231	1.81	1.54	60	0.0087	0.0223	1.35	2.38
RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0115	0.0196	2.04	1.28	240	0.0095	0.0212	1.56	2.01
180	0.0112	0.0211	1.84	1.30	180	0.0096	0.0230	1.45	2.09
120	0.0106	0.0227	1.62	1.40	120	0.0089	0.0234	1.31	2.17
60	0.0107	0.0251	1.48	1.54	60	0.0097	0.0257	1.31	2.32
RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0126	0.0199	2.19	1.33	240	0.0095	0.0208	1.57	2.16
180	0.0126	0.0215	2.04	1.35	180	0.0094	0.0229	1.43	2.18
120	0.0123	0.0231	1.85	1.38	120	0.0085	0.0242	1.21	2.20
60	0.0129	0.0250	1.79	1.49	60	0.0085	0.0241	1.22	2.30
RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0112	0.0202	1.93	1.34	240	0.0103	0.0217	1.64	2.08
180	0.0118	0.0226	1.81	1.40	180	0.0106	0.0237	1.55	2.12
120	0.0114	0.0229	1.73	1.50	120	0.0096	0.0236	1.42	2.20
60	0.0128	0.0249	1.77	1.64	60	0.0104	0.0269	1.34	2.35

Notes: 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 Elastic Net, a penalization parameter set by FRC (fixing 1 and 3 regressor(s)), for $RW_{shrk} \in [60, 120, 180, 240]$, and $RW_{pred} \in [60, 120, 180, 240]$.

3
Modifying traditional low-dimensional methodology to a high-dimensional environment

Abstract. This chapter presents a novel time-varying framework designed for high-dimensional, factor-based asset pricing models. The framework employs shrinkage techniques in regressions across pricing anomalies to identify statistically significant factors from a broad pool. Emphasizing sparsity, the framework introduces methods for selecting a limited number of impactful factors. Recognizing the implicit sparsity assumption in traditional models, this framework explicitly incorporates a similar scarcity principle during factor selection. We apply the proposed framework to an extensive set of factors across multiple time periods, demonstrating that simple, properly implemented techniques can produce robust results. Overall, this chapter offers valuable tools for researchers and practitioners, providing clear guidance on pricing factor selection while advocating for sparsity in high-dimensional models. This chapter exposes the major contributions of the article *The Factor Games: May the p-values be ever in your favor*.

Keywords: factor investing; shrinkage penalization; time-varying asset pricing; statistical significance.

3.1 Introduction

As exposed previously in Section 1.2, the quest for factors that explain the cross-section of expected returns has led to numerous contenders in the literature. Nevertheless, Cochrane (2011) contends that researchers may have overstepped by introducing an overwhelming multitude of factors, making it impractical and conceptually unwise to consider them collectively. Naming this phenomenon the *factor zoo*, he cautions against the indiscriminate use of numerous factors to explain average cross-sectional returns.

Feng *et al.* (2020) addressed the "factor zoo" challenge by employing a double-LASSO selection procedure (see Belloni *et al.*, 2014), favoring more parsimonious asset pricing models. Emphasizing out-of-sample predictability, **?** utilized the Ordered Weighted LASSO (OWL - Figueiredo & Nowak, 2016) to pinpoint factors capable of jointly explaining cross-sectional returns. Likewise, Freyberger *et al.* (2020) employed the Adaptive Group LASSO (see Huang *et al.*, 2010) to "non-parametrically dissect the factor zoo".

These studies, similarly to the exercise in Chapter 2, tackle the dimensionality issue of the factor zoo within a framework based on the Stochastic Discount Factor (SDF) model, deriving the shrinkage regression from the SDF error expression. Despite their ingenuity, this approach diverges from conventional methodologies for assessing the statistical significance of pricing factors in lower dimensions. Usually, classic methodologies involve spanning the proposed factor's returns against a given set of benchmark factors and scrutinizing the statistical significance of the intercept, as seen in studies by Jensen *et al.* (2023), Frazzini & Pedersen (2014), and Loh & Warachka (2012).

Similarly to approaches applied to the Stochastic Discount Factor (SDF) challenge, shrinkage regressions can be integrated into the spanning factors framework to address the dimensionality issue. One could use a shrinkage technique to factor spanning regressions and estimate the p-value of the intercept. Estimating a confidence interval for shrinkage regression coefficients isn't straightforward, given the bi-modal distributions of regressors' coefficients (Meinshausen & Bühlmann, 2006). However, refined techniques proposed by Meinshausen *et al.* (2009), building on the foundation laid by Wasserman & Roeder (2009), offer promising avenues for estimating the statistical significance of the LASSO's intercept - with tailored adjustments to accommodate time-series data, one may emulate lower-dimensional solutions in high-dimensional environments.

The LASSO, known for shrinking coefficients toward zero, faces challenges due to multicollinearity, which is prevalent in the returns of pricing anomalies: distinguishing the genuine effects of individual regressors becomes difficult in such scenarios. Coefficients associated with correlated features may undergo shrinkage toward zero, even if one or both harbor true effects. Efforts by Freyberger *et al.* (2020) and Sun (2024) have deployed adaptations of the raw LASSO within the SDF framework to address this challenge. Another approach to mitigate major multicollinearity issues involves conducting a Variance Inflation Factor analysis before shrinkage regressions, similarly to Green *et al.* (2017).

Again, the careful application of shrinkage is crucial when employing any of the aforementioned techniques, requiring researchers to calibrate the penalization parameter(s) carefully. Interestingly, any of the classic methodologies, like K-fold Cross-Validation (CV) (Stone, 1974), Bayesian Information Criterion (BIC) (Schwarz, 1978), and Akaike Information Criterion (AIC) (Akaike, 1974) focus on ensuring a predefined level of sparsity.

Contributions for the literature. We extend the classical factor returns spanning methodology to adapt to the *dimensionality curse* imposed by the factor zoo. This framework (see Chapter 1) enables the distinction between the time series used to handle high multicollinearity between regressors, the one for estimating spanning regressions' intercepts to identify relevant factors, and the final one to predict out-of-sample returns.

Moreover, we suggest methods to ensure a certain level of sparsity, both in shrinkage regressions (FRC - see Subsection 2.2.3) and in assessing the relevance of pricing anomalies. In our study, we used LASSO regression for spanning regressions, employing the traditional Bayesian Information Criterion and the proposed Fixed Regressors Criterion.

Findings. Among the main results of this essay are:

(i) The factor spanning regressions technique surpassed the proposed, stricter, benchmarks using a broad combination of parameters;

(ii) Using a common significance level of 0.10 for factor p-values often results in underwhelming out-of-sample performance, as the average p-value is higher than ideal and few factors are deemed relevant;

(iii) Imposing a predefined level of sparsity, using a fixed number of factors ranked by their p-values, substantially enhances predictability;

(iv) Dissociating time periods of interest positively impacts the factor zoo dimension reduction problem;

(v) Shorter factor selection and longer returns prediction windows are generally preferred;

(vi) Results from setting the LASSO's penalty parameter using the FRC5 criterion for the factor-spanning regressions are slightly superior.

Outline. The current chapter is organized into four sections. The Introduction presented related literature and achievements obtained. We expose the chosen methodology in Section 3.2, built upon the classic, low-dimensional, spanning factors returns idea, and how we adapt it for the factor zoo high-dimensional environment. Section 3.3 presents obtained results, comparing them with the benchmarks presented in Subsection 1.4.1. Finally, contributions and results are summarized in Section 3.4. The Appendix - Section 3.A - offers plots for time-variant factor statistical significance and complementary results for the Elastic Net regression.

3.2 Methodology

We evaluate the statistical significance of each factor individually through a series of spanning regressions, wherein we assess whether a combination of the other pricing anomalies can account for each factor's abnormal returns. This approach draws inspiration from low-dimension studies by Jensen *et al.* (2023), Frazzini & Pedersen (2014), and Loh & Warachka (2012).

We delineate the classical approach for assessing the significance of pricing anomalies in Subsection 3.2.1. Subsection 3.2.2 follows, being dedicated to adapting that classical approach to the high-dimensional environment of the factor zoo. This includes addressing high multicollinearity, developing a methodology for estimating statistical significance, introducing distinct criteria for setting the penalization parameter in shrinkage regressions, and proposing two criteria for assessing pricing factors' relevance within our framework.

3.2.1 Assessing factors' significance in a low-dimensional environment

Denoting f as a proposed pricing anomaly, it is possible to assess f's statistical significance by running a regression of its returns (ret_f) over some benchmark factor model, as shown in Equation 3.1 below.

$$ret_{f,t} = \alpha + \sum_{j \in F_{bench}} \beta_j ret_{f_{j,t}} + \epsilon_f, \tag{3.1}$$

where $ret_{f_{j,t}}$ represents the returns of a pricing factor that belongs to the set of relevant factors of the benchmark model (F_{bench}) at time t, and ϵ_f is the error term.

Take the classical Fama-French 3-factors model as a benchmark model example: ret_{f_j} would represent the returns of market, size, and value factors - see Fama & French (1993). Factor f will be considered relevant in pricing returns if its alpha is significant, i.e., if Equation 3.1's intercept presents a low enough p-value. This idea is broadly used in the pricing anomalies literature, as in articles like Jensen *et al.* (2023), Frazzini & Pedersen (2014), and Loh & Warachka (2012).

3.2.2 Spanning factors' returns in a high-dimension environment

We propose an approach that equalizes consideration for each potentially relevant pricing anomaly within the high-dimensional zoo of factors while allowing for time-varying importance (see Section 1.5), seeking to identify which factors have been the most statistically relevant. This is achieved by slightly modifying Equation 3.1. The model is as follows:

$$ret_{f_{i,t}} = \alpha_{f_i} + \sum_{j \neq i, j \in F} \beta_j ret_{f_{j,t}} + \epsilon_{f_i}, \tag{3.2}$$

where $ret_{f_{i,t}}$ is the return of factor f_i at time t, F is the set of all available pricing factors, α_i is the intercept of the regression, β_j is the linear coefficient for factor f_j, and ϵ_{f_i} is the error term.

The idea is to estimate each factor's alpha (α_{f_i}), calculate its p-value, and finally consider the most statistically relevant ones. If done multiple times, using rolling windows of time series, this framework should capture how factor relevance has changed over time.

The factor zoo is a well-known high-dimensional environment, therefore it is not advisable to estimate Equation 3.2 using a simple OLS regression. This is especially true as the object of interest will be the p-value of the intercept, and the more regressors we consider on the right-hand side of the model, the less likely it is to find a statistically significant intercept, as *overfitting* becomes a latent issue.

A viable approach to address high dimensionality is applying a shrinkage technique like LASSO (Tibshirani, 1996) to estimate the model in Equation 3.2.[1] However, four concerns must be addressed when implementing this type of regression:

[1] Results for the Elastic Net (Zou & Hastie, 2005) are available in the Appendix 3.A.2.

1. Regressors' multicollinearity can be detrimental when estimating associated coefficients, as LASSO's coefficients present higher standard deviations when regressors are highly correlated. We eliminate some candidate factors to bypass any possible problem caused by regressors' correlation;
2. Ordinarily, estimating the p-values for the LASSO's coefficients is a rather tricky exercise. We apply a regression technique that will calculate an unbiased estimation of the LASSO intercept's p-value, accounting for situations where the data is presented as a time series;
3. The LASSO regression is a technique sensitive to its penalization parameters. The literature provides some methodologies for setting it, like the information criteria and cross-validation, but none guarantee a certain sparsity level. We apply the proposed FRC (see Subsection 2.2.3) to ensure some desired sparsity degree;
4. Returns' predictions are notably known for being quite challenging, usually accompanied by low statistical significance. However, there may be a compelling argument to not only consider relevant factors with low enough p-values: we defend *ranking* factors according to their statistical significance and using a researcher-defined number of relevant factors.

Avoiding extreme multicollinearity Subsection 1.2.2 introduces factor models' multicollinearity issue and clarifies that it does not bias estimate slope coefficients, despite bumping up their standard errors. However, it imposes a challenge for applying the LASSO methodology, as it may drop correlated covariates that could be relevant.

Green *et al.* (2017) proposes a classic solution to mitigate the effects of multicollinearity: Variance Inflation Factors (VIFs). VIFs are defined as:

$$VIF_i = \frac{1}{1 - R_i^2}, \tag{3.3}$$

where R_i^2 denotes the unadjusted coefficient of determination when regressing the i-th independent variable on all others.

The closer R_i^2 is to zero (one), the less (more) correlated the ith independent variable is to the others, implying that multicollinearity is less (more) likely to exist. The researcher then chooses a threshold for the higher VIF she will accept, disregarding all independent variables that exceed this set value - Green *et al.*

(2017) removes factors for which $VIF_i > 7$. Used properly, this metric should capture how well the other factors' returns explain a given factor's returns.

We employ a procedure similar to Green *et al.* (2017) to address extreme factors' multicollinearity. However, we propose a few simple modifications to ensure that there is no look-ahead bias in the analysis. In their analysis, Green *et al.* (2017) calculate candidate factors' VIFs considering all available data. While this approach posed no major harm to their study focused on the cross-section of returns, our study allows for *time-variant* factors significance. Removing factors based on full-sample multicollinearity introduces a clear source of look-ahead bias. Therefore, we conduct the VIF analysis in rolling windows, considering data available only before the time of interest.[2]

Furthermore, we remove high VIF factors in decreasing order, one at a time. Traditionally, VIF values are calculated for all independent variables simultaneously, with those above a specified threshold being excluded. However, two variables could present a VIF higher than the threshold, and after removing one of them from the pool of independent variables, the other variable's VIF could potentially decrease to fit the acceptance level. To account for this behavior, we compute the VIF value for all candidate factors, remove only the one with the highest VIF, and then recalculate the VIF for all remaining factors - repeating the process until only factors with low enough Variance Inflation Factors survive.

Estimating intercept's p-value In lower dimensions, where simple OLS regression can be used without major overfitting concerns, verifying the statistical significance of the intercept in Equation 3.2 is straightforward. However, in higher dimensions, especially when applying shrinkage regressions, it becomes way more challenging.

Meinshausen *et al.* (2009) proposed a technique for calculating the intercept p-value for high-dimensional regressions, extending the concept of splitting the data into two parts, one for reducing the problem's dimensions and the other for applying classical variable selection techniques, originally proposed by Wasserman & Roeder (2009).[3]

However, their methodology does not account for time-series data - therefore, our bootstrap procedure must consider the specificities of such data. To achieve this, we modify the simple bootstrap procedure used by Meinshausen *et al.* (2009), employing a block bootstrap. For simplicity, we employ a non-

[2]Reefer to Section 1.5 for details on the time-variant framework.

[3]Wasserman & Roeder (2009) results are based on a single-split, whereas Meinshausen *et al.* (2009) proposed a multisplit method to avoid the randomness caused by data dependence.

overlapping blocks procedure (see Hall, 1985; Carlstein, 1986).

Let B represent the total number of bootstrap repetitions. For each $b = 1, ..., B$:

1. Randomly split the dataset into two disjoint groups, D^b_{shrk} and D^b_{p-val}:
   - This split must be done in blocks, as our data is of time-series-nature;
2. Run the shrinkage regression, estimating the set of active predictors ($\tilde{F}^b$), using data from D^b_{shrk}:
   - $\tilde{F}^b = \{j, \beta j \neq 0\}$, after running the shrinkage in Equation 3.2;
3. Using only D^b_{p-val}, fit the selected factors in $\tilde{F}^b$ with Ordinary Least Squares and calculate the p-value for the intercept, $P^b_{\alpha_i}$.

This procedure leads to a total of B p-values $P^b_{\alpha_i}$ - and their suitable summary statistics are quantiles (see Meinshausen *et al.*, 2009). For $\gamma \in (0, 1)$, define

$$\tilde{P}_{\alpha_i}(\gamma) = min\{1, q_\gamma([P^b_{\alpha_i}/\gamma; b = 1, ..., B])\}, \tag{3.4}$$

where $q_\gamma(\cdot)$ is the empirical γ-quantile function.

In Equation 3.4, we provide a p-value for any fixed $0 < \gamma < 1$. However, selecting γ appropriately is not straightforward, and searching for its optimal value does not guarantee error control. To determine the final p-value, we can adopt an adaptive approach that selects a suitable quantile value using a data-driven methodology. Let $\gamma_{min} \in (0, 1)$, typically set to 0.05, be a lower bound for γ. We then define the final p-value as:

$$P_{\alpha_i} = \min\{1, (1 - \log \gamma_{min}) \inf_{\gamma \in (0, \gamma_{min})} \tilde{P}_{\alpha_i}(\gamma)\} \tag{3.5}$$

We acknowledge that both the Bonferroni correction (Bonferroni, 1936) - applied after the bootstrap's final step — and the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) — employed after calculating the final p-values for all factors' intercepts — are alternatives for addressing possible data mining concerns. However, the results obtained in the empirical analysis, exposed in Section 3.3, led us to conclude that these concerns were unnecessary. If anything, higher intercepts' p-values would have led to a potentially harmful decrease in out-of-sample portfolio performance.

Setting up the penalization parameter The severity of dimensional reduction in shrinkage techniques like LASSO is closely tied to penalty parameter values, and setting their values using classic techniques - such as CV, BIC, or AIC -

won't guarantee either a stable or desirable number of factors survive the factor shrinkage process over.

This is particularly significant given that our methodology relies on assessing intercepts' p-values: the greater the number of factors with nonzero β_j coefficients in Equation 3.2, the less likely it is for α_{f_i} to differ significantly from zero.

To address this challenge, we proposed a different criterion for setting the penalization parameters in shrinkage regressions in Subsection 2.2.3. The objective is to ensure that a predetermined number of regressors (as determined by the researcher) will possess nonzero coefficients. This criterion could be especially advantageous for the current application, offering *ad-hoc* means to ensure the implicit sparsity assumption inherent in models like the Fama-French 3/5, Carhart, and q-4 factors (Fama & French, 1993, 2015; Carhart, 1997; Hou *et al.*, 2015), by selecting a penalty parameter that precisely returns the desired number of nonzero coefficient regressors.

In our empirical analysis (see Section 3.3), we use both classic (BIC) and proposed (FRC) methodologies to impose penalization in the LASSO setting. This approach enables us to access the impact of the number of factors considered in the factor model, as using BIC will permit a varying number of relevant factors in Equation 3.2, while FRC fixes nonzero beta factors to a desirable number, possibly comparable to low-dimensional models set sparsity.

Assessing factor relevance The estimation of factors' intercepts p-values emphasizes the relevance of factors with smaller p-values. Statistically correct approaches for selecting relevant factors usually involve stipulating a significance level and considering regressors with $P_{\alpha_{f_i}}$ lower than that level.

However, this idea may not be optimal in environments where predictability is, despite desirable, *challenging*, resulting in relatively large p-values. This is particularly true in fields such as return predictability, where regressions often have less-than-ideal explanatory power. In some cases, none of the candidate factors may generate p-values small enough to be considered significant. Conversely, there may even be instances where too many factors survive the shrinkage process, leading to an unwieldy number of factors to account for.

Given these concerns, instead of a fixed significance level, we suggest fixing the number of relevant factors, allowing researchers to specify the desired number of factors in their model. We believe that this could improve out-of-sample predictability.

Given this approach, we can impose the same silent sparsity assumption of classic asset pricing models, where only a handful of factors are considered. However, instead of selecting factors solely based on financial or economic intuition, we rely on the data, using statistical significance to identify relevant factors. This approach can also support extreme sparsity levels, such as models using only the factor with the single lowest p-value.

3.3 Empirical analysis

This analysis applies the methodology from Section 3.2 to reduce the dimensionality of the factor zoo by investigating the interactions between factor returns. Each factor's returns are compared against all others using the LASSO methodology to identify factors capable of generating statistically significant intercepts in the spanning regression (refer to Equation 3.2). The VIF methodology[4] uses a rolling window (RW_{VIF}) of 240 months, while three window lengths are tested for both RW_{shrk} and RW_{pred} (120, 180, and 240 months).

To determine the shrinkage penalization parameter, we consider 100 possible values, with initial values set at 10^{-4} and 10^{-1} for the lowest and highest possible penalization parameters, respectively.[5] The block bootstrap procedure is employed with eighty repetitions of blocks of five observations.

Results are presented for two approaches to select the shrinkage penalization parameter: the Bayesian Information Criterion (BIC) and the proposed Fixed Regressors Criterion (FRC), which fixes five regressors within the spanning regression.

3.3.1 Accounting for multicollinearity

The empirical analysis begins tackling potential issues arising from extreme multicollinearity among pricing anomalies. This concern is addressed using the Variance Inflation Factor (VIF) criterion, assessing each factor's returns based on the R^2 of regressions using all other factors as predictors.

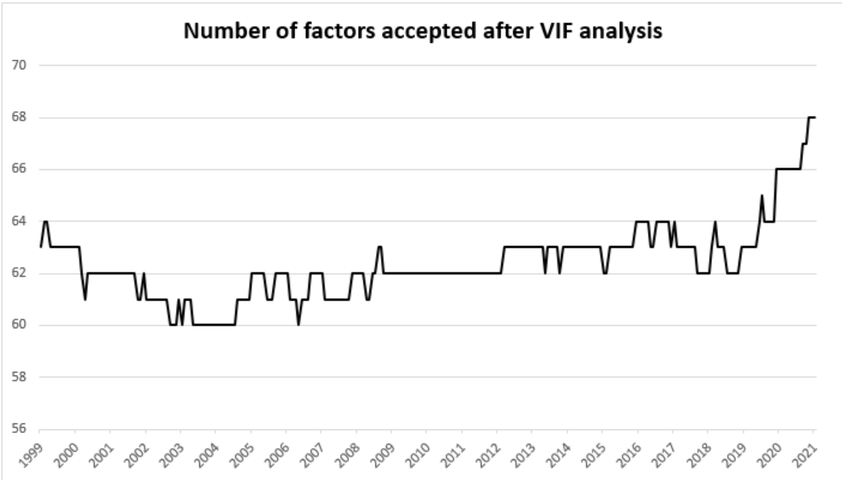
The results of this procedure, conducted with a rolling window of 240 periods, are presented in this section. Figure 3.1 illustrates the number of factors that survive this process over time, indicating a relatively constant count of accepted factors throughout the sample period. Most periods show an accepted factor count between 60 and 64, with an average of approximately 18.6 factors re-

[4] Threshold for accepted VIF value is 10 - a little more permissive than Green *et al.* (2017).

[5] When using the BIC, we guarantee that at least one factor survive the shrinkage and that the ratio between the highest and lowest penalty values tested is 10^3.

moved. Notably, there is a discernible increase in the number of factors accepted by the VIF analysis after the second half of 2019, suggesting a potential decrease in overall multicollinearity among pricing anomalies.

Figure 3.1: Pricing anomalies accepted by the Variance Inflation Factor control through time



Notes: The plot reports the number of factors that survived the VIF analysis throughout all the available periods.

Complementing the analysis, Table 3.1 lists all the anomalies rejected after the VIF procedure for at least one period. It reveals that seven candidate pricing factors were consistently disregarded due to multicollinearity, while seventeen factors were discarded in at least 80% of the time series.

Furthermore, our results demonstrate notable stability in identifying factors prone to multicollinearity issues: nineteen of the thirty-one pricing anomalies rejected at least once failed the VIF criterion in over 60% of periods.

3.3.2
Reducing dimensionality

After addressing high multicollinearity, our next step was to select relevant factors and reduce the dimensionality of the factor zoo: we proceeded to estimate the p-value of the intercept in spanning regressions - see 3.2 - to guide the selection process.

The empirical analysis uses LASSO-based shrinkage and evaluates BIC and FRC5 methodologies for penalty parameter selection. P-values were estimated using three distinct rolling windows (RW_{shrk}) of 120, 180, and 240 months.[6]

[6] As the p-value estimation methodology requires splitting the data into two disjoint parts, a shorter rolling window of 60 periods, used in Chapter 2, was not feasible.

Table 3.1: Summary of pricing anomalies rejected by the Variance Inflation Factor control

Factor	Rejections	Percentage	Factor	Rejections	Percentage
baspread	265	100.0%	quick	219	82.6%
beta	265	100.0%	roeq	176	66.4%
betasq	265	100.0%	agr	160	60.4%
cash	265	100.0%	stdacc	125	47.2%
lev	265	100.0%	pchsaleinv	95	35.8%
retvol	265	100.0%	roaq	91	34.3%
zerotrade	265	100.0%	ep	72	27.2%
idiovol	262	98.9%	currat	46	17.4%
mom6m	262	98.9%	bm	30	11.3%
dy	260	98.1%	gma	24	9.1%
stdcf	258	97.4%	roavol	21	7.9%
std_turn	253	95.5%	absacc	10	3.8%
turn	251	94.7%	pchsale_pchinvt	6	2.3%
salecash	245	92.5%	maxret	4	1.5%
ill	242	91.3%	invest	1	0.4%
sp	232	87.5%			

Notes: This table lists all factors rejected by the VIF analysis at least once, and reports the number of times the factors presented multicollinearity issues - and the percentage of available periods they were rejected. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024).

Table 3.2 summarizes the results obtained, shedding light on several interesting aspects. Firstly, our methodology is conservative in estimating p-values, with a relatively small number of factors yielding valid p-values, reaching approximately 13.4% of candidates.[7] Moreover, the number of factors with valid p-values decreases along the time series length used for regression.

Table 3.2: Spanning factors returns shrinkage outcome

Penalty style	RW_{shrk}	Average # of valid p-values	Average # of p-values < 0.10	% of periods with zero p-values < 0.10
BIC	120	10.00	1.09	34%
BIC	180	9.37	1.18	32%
BIC	240	7.15	1.24	34%
f5	120	10.69	1.24	30%
f5	180	9.99	1.20	35%
f5	240	7.15	1.12	36%

Notes: This table summarizes the outcomes of applying the LASSO regressions over the factors spanning regressions, reporting the average number of valid p-values, i.e., p-values under 1.0, the average number of p-values under the threshold of 0.10, and the percentage of time that any factor presents associated p-value smaller than 0.10, setting shrinkage penalization parameter through BIC of FRC (fixing 5 regressors - f5), and different values for RW_{shrk}.

Considering all aforementioned aspects, employing the FRC5 for setting the shrinkage penalization and conducting regressions using 120-period windows produced the most favorable results. This approach yielded the highest average

[7] Meinshausen *et al.* (2009) methodology provides a conservative approach to family-wise error rate (FWER) control, similar in spirit to Holm (1979). If many null-hypotheses rejections were to happen, the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) could be considered.

number of valid p-values, the highest average of factors with low p-values, and the lowest percentage of periods with no factor exhibiting a p-value below 0.10. However, there is no clear indication that a particular penalty style or RW_{shrk} is optimal, as there is no evidence of a relationship between RW_{shrk} and the overall level of statistical significance of the factors' p-values. Additionally, there is no discernible time-variant behavior, as illustrated by plots in the Appendix 3.A.1.

An elevated number of periods without statistically significant pricing factors could significantly undermine the performance of a hedge portfolio based on the forecasted returns of test assets — a common approach in the literature applied in this study — as it would frequently lead to deallocation. Moreover, the relatively low statistical significance is not entirely unexpected, given that time-varying asset pricing operates within a particularly noisy environment. Therefore, a method that rejects all pricing factors more than 30% of the time may be overly selective.

Sailing through the turbulent sea of pricing factors, focusing on the most promising options, regardless of their individual p-values, might be interesting. Rather than setting a threshold value for p-value, we can instead designate a desired number of factors to be considered relevant and select those with the highest statistical significance, i.e., the lowest p-values — even if they exceed some typical acceptance threshold. This proposal, involving an *ad-hoc* determination of the number of relevant factors, also aligns with the silent sparsity assumption of classic asset pricing models. Just as the literature accepts classic models like Fama-French (see Fama & French, 1993, 2015), or Carhart and q-4 factors (see Carhart, 1997; Hou *et al.*, 2015), with only a handful of pricing factors, why not let a statistical model dictate comparable sparsity levels without relying on prior qualitative information?

As p-values are comparable in magnitude and a matter of *statistical faith*, we believe that bending traditional beliefs to navigate turbulent waters with *any compass* can potentially support the navigator's decision-making.

3.3.3 Out-of-sample results

We will now present the results obtained by applying the proposed methodology - see Section 3.2. We begin with the results of using the BIC criterion in the spanning regression, subsequently discussing the results obtained by fixing five relevant regressors (FRC5). In each exposition, we initially present results obtained using only factors whose p-values are lower than 0.10, followed by the presenting results using one, three, or five factors with the lowest p-value(s).

BIC penalization Results obtained using the BIC for penalization in the spanning regressions, with a significance threshold set at p-values under 0.10, are presented in Table 3.3, showcasing not that impressive performances.

Table 3.3: Out-of-sample results - BIC penalty and p-values under 0.10

RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0088	0.0205	1.49	1.38
180	0.0085	0.0223	1.32	1.38
120	0.0088	0.0237	1.29	1.41

RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0086	0.0246	1.21	1.39
180	0.0092	0.0258	1.23	1.40
120	0.0096	0.0261	1.27	1.40

RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0100	0.0247	1.40	1.35
180	0.0104	0.0254	1.41	1.36
120	0.0102	0.0262	1.35	1.34

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with associated p-values lower than 0.10, setting shrinkage penalization parameter through BIC, for different combinations of RW_{shrk} and RW_{pred}.

At first glance, an uninformed reader might find the overall gross Sharpe Ratios above unity somewhat impressive. However, none of the rolling window combinations surpassed the benchmark (see Section 1.4.1), indicating, at most, a promising direction rather than significant performance gains.

The relatively underwhelming performance can be attributed to the method for selecting relevant factors, a p-value threshold, which results in numerous periods where constructing long-short portfolios becomes infeasible due to the absence of relevant factors for predicting test asset returns.[8] Consequently, there are many instances where the returns of the hedge portfolios are zero, thus diluting the mean returns and undermining the obtained Sharpe ratios.

Table 3.4 presents the results obtained when selecting a predetermined number of factors as relevant, using the lowest p-values obtained for each period for ranking, revealing *more promising* out-of-sample performances.

[8] As demonstrated in Table 3.2, when employing the BIC penalization, there are instances where no factor exhibits a p-value under 0.10, accounting for at least 32% of the time.

Table 3.4: Out-of-sample results - BIC penalty and 1, 3, and 5 lowest p-values

Lowest p-value					**3 lowest p-values**					**5 lowest p-values**				
RW_{shrk}	120				RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0130	0.0231	1.95	1.40	240	0.0111	0.0231	1.67	2.15	240	0.0108	0.0271	1.38	2.52
180	0.0135	0.0247	1.89	1.41	180	0.0113	0.0245	1.61	2.20	180	0.0108	0.0274	1.37	2.54
120	0.0126	0.0254	1.72	1.39	120	0.0116	0.0260	1.54	2.30	120	0.0111	0.0281	1.37	2.64
RW_{shrk}	180				RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0121	0.0235	1.79	1.38	240	0.0122	0.0261	1.63	2.10	240	0.0115	0.0283	1.41	2.45
180	0.0129	0.0252	1.78	1.36	180	0.0129	0.0274	1.62	2.06	180	0.0118	0.0292	1.40	2.45
120	0.0130	0.0258	1.74	1.37	120	0.0132	0.0279	1.64	2.12	120	0.0119	0.0302	1.36	2.54
RW_{shrk}	240				RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0119	0.0252	1.64	1.37	240	0.0140	0.0265	1.83	2.17	240	0.0115	0.0297	1.34	2.49
180	0.0121	0.0268	1.56	1.41	180	0.0144	0.0271	1.85	2.17	180	0.0111	0.0299	1.29	2.50
120	0.0120	0.0276	1.51	1.48	120	0.0144	0.0277	1.80	2.27	120	0.0114	0.0301	1.31	2.55

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with the lowest 1, 3, and 5 associated p-values, setting shrinkage penalization parameter through BIC, for different combinations of RW_{shrk} and RW_{pred}.

Hedge portfolios' mean returns reported in Table 3.4 exhibit improved Sharpe ratios, particularly when considering a smaller number of regressors, as shown in the first two blocks. Generally, smaller values for RW_{shrk} and higher values for RW_{pred} tend to perform better. However, an intriguing aspect derived from the presented results is that there is no clear optimal value for RW_{shrk} and RW_{pred} to achieve the highest Sharpe ratios. The best value (1.95) is found in the pair ($RW_{shrk} = 120$; $RW_{pred} = 240$), while the third-best result (1.85) is observed in the combination of ($RW_{shrk} = 240$; $RW_{pred} = 180$).

Out-of-sample results also indicate that researchers should aim for a *high level of sparsity*, selecting at most three factors when pricing asset returns. Most remarkably, the best outcomes are achieved when the maximum degree of sparsity is enforced. This finding suggests that, within our framework, the noise in the factor zoo is substantial enough that selecting only the most relevant pricing factor during each shrinkage period yields better results.

Finally, it is noteworthy that turnover also increases with the number of factors considered. This is a consequence of the persistence of the selected factor(s). Despite allowing for monthly changes, this framework for factor selection does not impose frequent and significant alterations in the relevant factors.[9] This observation supports the notion of opting for a more sparse factor asset-pricing

[9]Qualitative lenses are put upon the selected factors in Chapter 4.

model, as lower turnovers translate to more favorable trading conditions and reduced trading and slippage costs.

FRC5 penalization Similar to its BIC counterpart, fixing five regressors in the spanning regressions before selecting factors with p-values under 0.10 also yields unimpressive results, as shown in Table 3.5. A comparison with Table 3.3

Facing Tables 3.5 and 3.3, overall results are very similar, and none of the tested portfolios was able to yield high enough Sharpe ratios to surpass the Fama-French-based benchmark.

Table 3.5: Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and p-values under 0.10

RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0088	0.0205	1.49	1.38
180	0.0085	0.0223	1.32	1.38
120	0.0088	0.0237	1.29	1.41

RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0086	0.0246	1.21	1.39
180	0.0092	0.0258	1.23	1.40
120	0.0096	0.0261	1.27	1.40

RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0100	0.0247	1.40	1.35
180	0.0104	0.0254	1.41	1.36
120	0.0102	0.0262	1.35	1.34

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with associated p-values lower than 0.10, setting shrinkage penalization parameter through FRC, fixing 5 regressors, for different combinations of RW_{shrk} and RW_{pred}.

Results obtained with the fixed five regressors criterion for the spanning regressions and selecting the lowest one, three, and five p-values for relevant factors are displayed in Table 3.6. Once again, there is a clear parallel with results from Table 3.4. However, using the FRC5 criterion, some portfolios' *Sharpe ratios* were *greater than 2.0*, surpassing the best out-of-sample Sharpe ratio obtained when using the classic BIC: 1.95.

Again, as in Chapter 2, results advocate for a high sparsity level, and choosing only one factor at a time appears to be the superior approach — especially considering the lower turnover imposed. Overall, results tend to improve when considering smaller shrinkage windows ($RW_{shrk} \in \{120, 180\}$), longer prediction time series ($RW_{pred} \in \{240, 180\}$), and a lower number of factors considered (only the lowest p-value). Nevertheless, this is not always the case: good results were

Table 3.6: Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and 1, 3, and 5 lowest p-values

Lowest p-value					**3 lowest p-values**					**5 lowest p-values**				
RW_{shrk}	120				RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0131	0.0226	2.01	1.38	240	0.0117	0.0232	1.75	2.15	240	0.0106	0.0256	1.44	2.51
180	0.0131	0.0241	1.88	1.38	180	0.0113	0.0244	1.60	2.18	180	0.0108	0.0260	1.45	2.54
120	0.0123	0.0248	1.72	1.35	120	0.0117	0.0257	1.57	2.30	120	0.0108	0.0273	1.37	2.63
RW_{shrk}	180				RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0135	0.0229	2.04	1.36	240	0.0120	0.0251	1.65	2.23	240	0.0104	0.0271	1.33	2.48
180	0.0142	0.0243	2.03	1.34	180	0.0122	0.0265	1.60	2.20	180	0.0107	0.0286	1.29	2.50
120	0.0144	0.0250	2.00	1.36	120	0.0125	0.0277	1.56	2.28	120	0.0108	0.0301	1.25	2.60
RW_{shrk}	240				RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0121	0.0246	1.71	1.40	240	0.0143	0.0259	1.91	2.18	240	0.0117	0.0288	1.41	2.48
180	0.0123	0.0262	1.62	1.43	180	0.0143	0.0266	1.86	2.17	180	0.0110	0.0297	1.28	2.46
120	0.0120	0.0270	1.54	1.50	120	0.0141	0.0273	1.79	2.25	120	0.0112	0.0305	1.27	2.52

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with the lowest 1, 3, and 5 associated p-values, setting shrinkage penalization parameter through FRC, fixing 5 regressors, for different combinations of RW_{shrk} and RW_{pred}.

obtained when using $RW_{shrk} = 240$ with the three lowest p-values — despite imposing a significant increase in average turnover.

A last notable observation is that the results appear to be relatively insensitive to the choice of the rolling window parameters: the sparsity level imposed by the researcher for factor selection seems to have greater importance. Overall, our findings indicate that *sparsity is crucial* in the noisy environment of the factor zoo.

3.3.4
Facing benchmarks

Upon cross-examining the results presented in Tables 3.3, 3.4, 3.5, and 3.6 alongside the new benchmark in Table 1.2, we observe (again) that surpassing the proposed benchmark Sharpe ratio of 1.64 is no simple task. For instance, even the best results obtained with the p-value threshold method (see Tables 3.3 and 3.5), respectively 1.49 ($RW_{shrk} = 120$, $RW_{pred} = 240$) and 1.44 ($RW_{shrk} = 240$, $RW_{pred} = 240$), failed to outperform the Fama-French based predictions.

However, several hedge portfolios managed to outperform this stricter benchmark. As shown in Table 3.7, all these portfolios were constructed under the fixed number of relevant factors criterion, and two key insights can be drawn:

1. Even if the most statistically significant factor has a p-value higher than 0.10, using it provided better results than not designating any factor as relevant;
2. A higher sparsity level resulted in enhanced out-of-sample predictability.

Table 3.7: Outperforming hedge portfolios

Spanning criterion	Factor selection criterion	RW_{shrk}	RW_{pred}	Mean	SD	Sharpe	Turnover
FRC5	lwst1	180	240	0.0135	0.0229	2.04	1.36
FRC5	lwst1	180	180	0.0142	0.0243	2.03	1.34
FRC5	lwst1	120	240	0.0131	0.0226	2.01	1.38
FRC5	lwst1	180	120	0.0144	0.0250	2.00	1.36
BIC	lwst1	120	240	0.0130	0.0231	1.95	1.40
FRC5	lwst3	240	240	0.0143	0.0259	1.91	2.18
BIC	lwst1	120	180	0.0135	0.0247	1.89	1.41
FRC5	lwst1	120	180	0.0131	0.0241	1.88	1.38
FRC5	lwst3	240	180	0.0143	0.0266	1.86	2.17
BIC	lwst3	240	180	0.0144	0.0271	1.85	2.17
BIC	lwst3	240	240	0.0140	0.0265	1.83	2.17
BIC	lwst3	240	120	0.0144	0.0277	1.80	2.27
FRC5	lwst3	240	120	0.0141	0.0273	1.79	2.25
BIC	lwst1	180	240	0.0121	0.0235	1.79	1.38
BIC	lwst1	180	180	0.0129	0.0252	1.78	1.36
FRC5	lwst3	120	240	0.0117	0.0232	1.75	2.15
BIC	lwst1	180	120	0.0130	0.0258	1.74	1.37
BIC	lwst1	120	120	0.0126	0.0254	1.72	1.39
FRC5	lwst1	120	120	0.0123	0.0248	1.72	1.35
FRC5	lwst1	240	240	0.0121	0.0246	1.71	1.40
BIC	lwst3	120	240	0.0111	0.0231	1.67	2.15
FRC5	lwst3	180	240	0.0120	0.0251	1.65	2.23
BIC	lwst3	180	120	0.0132	0.0279	1.64	2.12

Notes: This table reports the portfolios presented on Tables 3.3, 3.4, 3.5, and 3.6 that yielded better annualized Sharpe ratios than the best-performing benchmark portfolio - presented on Table 1.2.

Our findings suggest that the dimensionality issue of the zoo of factors can be satisfactorily addressed by estimating the statistical significance of factor-spanning shrinkage regressions' intercepts. Specifically, our results show that when using monthly data in the framework proposed in Section 3.2, the most relevant factors should be selected assuming significant pre-defined sparsity, i.e., classifying fewer than 3 factors as relevant.

More interestingly, although the best out-of-sample results are generally obtained considering a combination of shorter factor selection and longer returns prediction windows, i.e., $RW_{shrk} \in \{120, 180\}$ and $RW_{pred} \in \{180, 240\}$, it was possible to surpass the proposed benchmark with $RW_{shrk} = 240$ or $RW_{pred} = 120$. This indicates that our findings are likely not very susceptible to data mining, as there is no need for a specific combination of window sizes to obtain interesting out-of-sample results.

3.4
Final considerations

In this chapter, we assess the statistical significance of factor relevance by analyzing the intercepts of spanning regressions, building upon the framework developed in Chapter 1. Lower p-values for these intercepts indicate that certain factor returns cannot be sufficiently explained by the returns of other factors, suggesting unique information that others cannot replicate.

Furthermore, we keep underscoring the power of lower dimensional models' silent sparsity assumption, introducing it within our framework through two key steps: first, in the spanning regressions, and second, when assessing factor relevance.

We recommend the FRC criterion (see Subsection 2.2.3) for setting the penalization parameter in shrinkage regressions, dynamically adjusting it to ensure the selection of a predefined number of jointly relevant factors. One more time, this criterion proves advantageous when researchers possess prior beliefs regarding the number of factors to be considered. Additionally, we recommend enforcing a set sparsity level when selecting relevant factors, focusing on a predetermined number of factors ranked by their intercept p-values, independent of any specific statistical threshold. The outcomes presented in our study suggest that this sparsity assumption holds merit in both instances.

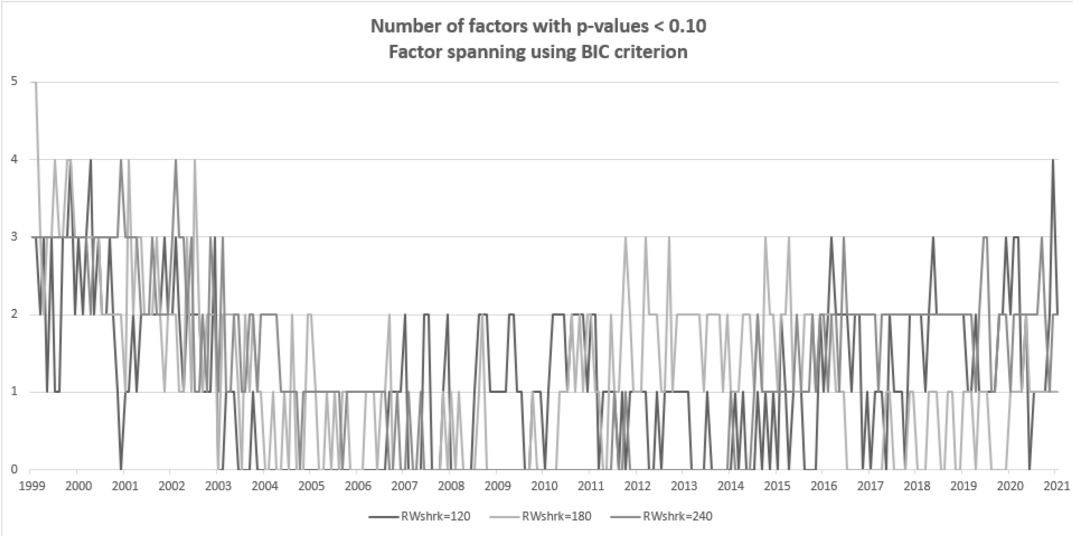
We applied our methodology to a comprehensive set of 80 anomaly factors, alongside the market factor, using the widely adopted LASSO technique for spanning regressions. We fixed the rolling window for multicollinearity treatment at 240 periods and considered candidate periods of 120, 180, and 240 months for both factor selection and returns forecasting. Although results were less favorable when only factors with p-values below 0.10 were considered, ranking factors by statistical significance and setting a predefined number allowed us to surpass the stricter benchmark across several parameter combinations.

Our results indicate that shorter windows for factor selection, combined with longer windows for return prediction, tend to yield better outcomes. This finding is grounded in the notion that the relevance of factors should closely correspond to the present moment, while the estimation of the relationship between selected factors' returns and test assets' returns benefits from a longer horizon. Nevertheless, certain combinations with extended factor selection windows *and* forecasting periods outperformed the proposed benchmark, indicating robustness against potential data-mining concerns and supporting the credibility of our findings.

3.A
Appendix

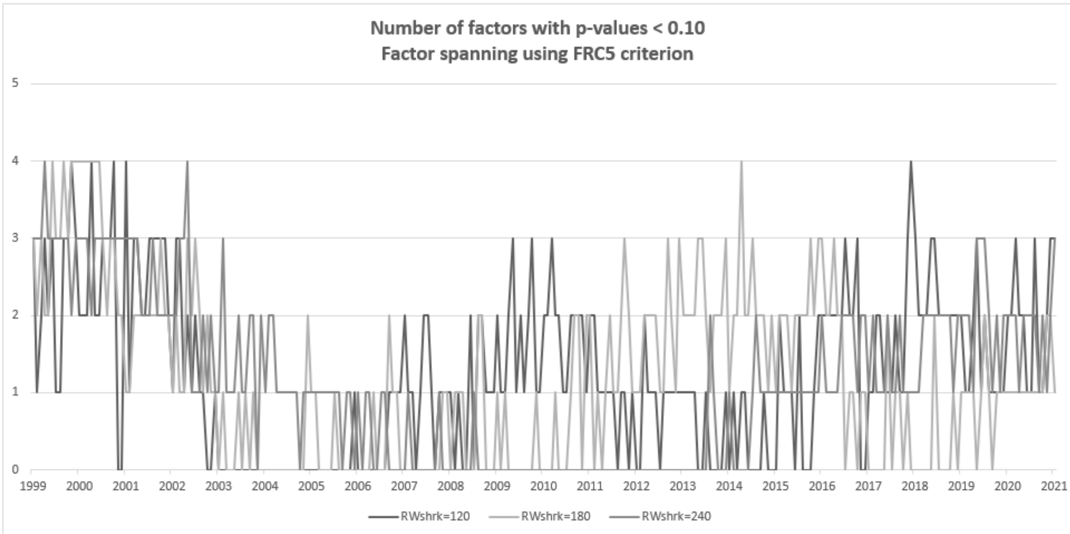
3.A.1
Time variability of statistically significant factors

Figure 3.A.1: Factors with intercepts' p-values bellow 0.10 - BIC penalty



Notes: The plot reports the number of factors with spanning regressions intercepts lower than 0.10 when setting the penalization parameter through BIC, for $RW_{shrk} \in [120, 180, 240]$.

Figure 3.A.2: Factors with intercepts' p-values bellow 0.10 - FRC5 penalty



Notes: The plot reports the number of factors with spanning regressions intercepts lower than 0.10 when setting the penalization parameter through FRC5, for $RW_{shrk} \in [120, 180, 240]$.

3.A.2 Elastic Net results

Table 3.A.1: Out-of-sample results - BIC penalty and p-values under 0.10

RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0090	0.0205	1.52	1.35
180	0.0090	0.0225	1.39	1.37
120	0.0086	0.0240	1.24	1.40

RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0099	0.0243	1.41	1.42
180	0.0104	0.0251	1.44	1.42
120	0.0108	0.0258	1.44	1.41

RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0087	0.0245	1.23	1.36
180	0.0093	0.0252	1.28	1.35
120	0.0092	0.0260	1.22	1.34

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with associated p-values lower than 0.10, setting shrinkage penalization parameter through BIC, for different combinations of RW_{shrk} and RW_{pred}.

Table 3.A.2: Out-of-sample results - BIC penalty and 1, 3, and 5 lowest p-values

Lowest p-value					**3 lowest p-values**					**5 lowest p-values**				
RW_{shrk}	120				RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0139	0.0229	2.11	1.45	240	0.0107	0.0255	1.46	2.19	240	0.0107	0.0282	1.32	2.52
180	0.0141	0.0244	2.00	1.46	180	0.0109	0.0263	1.43	2.20	180	0.0108	0.0285	1.31	2.54
120	0.0134	0.0252	1.84	1.44	120	0.0108	0.0274	1.37	2.29	120	0.0108	0.0291	1.28	2.66
RW_{shrk}	180				RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0128	0.0234	1.90	1.37	240	0.0124	0.0254	1.69	2.19	240	0.0107	0.0279	1.32	2.49
180	0.0137	0.0250	1.89	1.34	180	0.0125	0.0270	1.60	2.15	180	0.0108	0.0288	1.30	2.46
120	0.0137	0.0258	1.84	1.35	120	0.0126	0.0282	1.55	2.20	120	0.0114	0.0298	1.32	2.56
RW_{shrk}	240				RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0117	0.0255	1.58	1.40	240	0.0136	0.0264	1.78	2.15	240	0.0111	0.0292	1.32	2.49
180	0.0120	0.0271	1.54	1.43	180	0.0142	0.0269	1.83	2.15	180	0.0110	0.0298	1.28	2.49
120	0.0118	0.0279	1.46	1.52	120	0.0140	0.0274	1.77	2.25	120	0.0111	0.0311	1.24	2.56

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with the lowest 1, 3, and 5 associated p-values, setting shrinkage penalization parameter through BIC, for different combinations of RW_{shrk} and RW_{pred}.

Table 3.A.3: Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and p-values under 0.10

RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0092	0.0216	1.48	1.49
180	0.0098	0.0233	1.46	1.48
120	0.0096	0.0242	1.37	1.53

RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0076	0.0241	1.09	1.51
180	0.0077	0.0258	1.03	1.52
120	0.0082	0.0265	1.07	1.53

RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0100	0.0247	1.41	1.29
180	0.0104	0.0255	1.41	1.31
120	0.0104	0.0263	1.38	1.31

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with associated p-values lower than 0.10, setting shrinkage penalization parameter through FRC, fixing 5 regressors, for different combinations of RW_{shrk} and RW_{pred}.

Table 3.A.4: Out-of-sample results - FRC, fixing 5 regressors, (f5) penalty and 1, 3, and 5 lowest p-values

Lowest p-value					**3 lowest p-values**					**5 lowest p-values**				
RW_{shrk}	120				RW_{shrk}	120				RW_{shrk}	120			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0129	0.0229	1.95	1.40	240	0.0104	0.0246	1.47	2.16	240	0.0096	0.0255	1.30	2.52
180	0.0131	0.0243	1.87	1.39	180	0.0107	0.0260	1.42	2.19	180	0.0099	0.0258	1.33	2.54
120	0.0124	0.0251	1.71	1.36	120	0.0110	0.0269	1.42	2.28	120	0.0101	0.0268	1.31	2.65
RW_{shrk}	180				RW_{shrk}	180				RW_{shrk}	180			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0136	0.0242	1.86	1.32	240	0.0118	0.0251	1.63	2.25	240	0.0116	0.0269	1.49	2.51
180	0.0136	0.0238	1.97	1.37	180	0.0117	0.0266	1.52	2.20	180	0.0120	0.0288	1.44	2.50
120	0.0124	0.0278	1.58	1.42	120	0.0116	0.0274	1.47	2.27	120	0.0121	0.0296	1.41	2.58
RW_{shrk}	240				RW_{shrk}	240				RW_{shrk}	240			
RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover	RW_{pred}	Mean	SD	Sharpe	Turnover
240	0.0118	0.0250	1.64	1.39	240	0.0141	0.0259	1.88	2.16	240	0.0116	0.0273	1.47	2.48
180	0.0122	0.0266	1.59	1.43	180	0.0143	0.0267	1.85	2.18	180	0.0117	0.0280	1.44	2.50
120	0.0118	0.0274	1.49	1.53	120	0.0142	0.0275	1.79	2.25	120	0.0118	0.0289	1.42	2.58

Notes: This table reports out-of-sample monthly mean returns and associated standard deviations, annualized Sharpe ratios, and monthly average turnover of hedge portfolios going long-short the 10% of stocks with the highest-lowest predicted returns, considering factors with the lowest 1, 3, and 5 associated p-values, setting shrinkage penalization parameter through FRC, fixing 5 regressors, for different combinations of RW_{shrk} and RW_{pred}.

4
Which factors matter?

Abstract. The thesis' final chapter provides a qualitative analysis of the pricing factors deemed relevant by the methodologies developed in previous chapters. It begins by identifying the ten most frequently selected factors across key exercises, observing that while factor relevance shows stability within each methodology at different sparsity levels, significant variations emerge between methodologies. A further analysis groups factors by their economic fundamentals, revealing that profitability-related anomalies are prominent in both methodologies, but with distinct emphases: investment anomalies dominate within the SDF-error model, whereas trading frictions are emphasized in the factor-spanning approach. The examination also shows a consistent factor selection pattern over time, suggesting that each methodology captures unique, stable dimensions of pricing relevance. The chapter concludes with an insight into the complementary nature of these methodologies, proposing that their distinct but reliable factor selections indicate each may capture a separate facet of the asset pricing problem, which presents an opportunity for future research on the integration of these approaches.

Keywords: qualitative factor analysis; time-varying asset pricing; cross-method factor divergence; factors time-variability.

4.1 Introduction

Up to this point, we have structured a comprehensive framework to navigate the extensive landscape of pricing anomalies (see Chapter 1) and applied it to two completely different methodologies for factor selection - see Chapters 2 and 3. Analyzing the returns of hedge portfolios across both methodologies, we found each achieved impressive out-of-sample results, even *surpassing* the stricter benchmarks proposed in Subsection 1.4.1.

Up to his point, we have mainly reported returns´ characteristics, evaluating factor selection strategies by primarily comparing their Sharpe ratios. However, understanding *which* factors/factor categories are responsible for generating such impressive predictability is relevant - and one of the main advantages of using a methodology like the LASSO instead of Principal Components Regression (PCR) is that unveiling factor relevance is straightforward.

This final essay is dedicated to shedding light on the factors that survived the shrinking process of both previous chapters and looking for differences and similarities between them. We start at the individual level (looking at the pricing anomalies that were most set as relevant), then aggregate the factors in qualitative families (verifying the relevance of distinct types of factors), and finally visualize time-series plots of chosen factor categories (searching for cyclical patterns).

Contributions for the literature. As this essay focuses on comparing two methodologies for shrinking the factor zoo, it is reasonable to state that the major contributions were developed in Chapters 2 and 3 while applying the flexible framework developed in Chapter 1. However, as we are not aware of other studies that qualitatively compare two high-dimensional factor models, other essays could take advantage of our approach if in need of a comparison methodology.

Outline. This closing essay is organized into three more sections. Section 4.2 briefly exposes the comparison approaches and defines six categories for classifying the factors concerning their fundamentals. We present the factor hall of fame in Section 4.3, reporting the most relevant factors - and factor categories - for each applied methodology, before summarizing findings in Section 4.4.

4.2 Methodology

To open the *black box* of factors identified in Chapters 2 and 3, we will proceed to report which factors were most selected as relevant by the applied methodologies. The analysis begins with an evaluation of the full cross-section of data, reporting how many times each factor was found relevant by each methodology (see Subsection 4.3.1), before introducing time awareness, as we observe the persistence of factor importance and how factor selection changes over time - see Subsection 4.3.3.

While unveiling which factors were used in our previous exercises, we will maintain a direct parallel between Chapters 2 and 3 qualitative outcomes, examining if there are qualitative similarities between their selection - see Subsection 4.3.2. To further enrich the analysis, we will group the factors into qualitative families based on their fundamentals. To do so, we begin defining six distinct categories for factors, and grouping all of the 80 pricing anomalies exposed in Subsection 1.2.1 into them.

4.2.1 Factor (zoo) categories

Zoos are diverse, displaying a huge variety of wildlife for those willing to explore them. Now imagine all those vastly distinct animals are all in one vivarium, without any kind of separation between them. It would be a pandemonium. No habit would fit the needs of every single animal, behavioral conflicts between species could arise, animals could be exposed to unfamiliar diseases, predators would be side to side with their pray, and there would be *noise everywhere*

These are just some of the reasons zoos are organized into distinct habitats, usually segmented: mammal, reptile, and nocturnal houses, bird aviaries, insectaria, amphibian centers, aquariums, butterfly gardens, etc. Zoos are *organized* focusing on facilitating the visitor's experience, mitigating possible sources of confusion.

We will group the anomalies into families in order to comprehend what *kind of factor* were considered relevant by our proposed methodologies. This enables us to observe if there is a dominance of certain kinds of factors over their counterparts, or if there is a time-variant variation of relevance distributed between families.

We will follow the categories proposed at Hou *et al.* (2020) and group the studied pricing factors into six categories respecting the theoretical reasoning behind their alleged effectiveness. Those categories are:

Momentum Momentum factors are based on the observation that stocks that have performed well tend to continue doing so, while underperforming stocks are likely to keep underperforming. This phenomenon can be attributed to a variety of behavioral and market mechanisms. For instance, investors may underreact to news, leading to price adjustments that drag over time. Alternatively, herding behavior, where investors follow the actions of others, can also drive momentum.

Since its first appearance at Jegadeesh & Titman (1993), empirical evidence supporting momentum strategies has been strong, with numerous studies showing that past winners tend to outperform past losers over short to medium-term horizons. Despite its robustness, momentum can also be risky, as it is prone to sharp reversals during market downturns or periods of high volatility.

Value-versus-growth The value versus growth dichotomy is one of the most well-established concepts in financial markets. Value stocks are characterized by low prices relative to their fundamental metrics, such as earnings, book value, or cash flow - see Basu (1977), Rosenberg *et al.* (1985), and Ou & Penman (1989) respectively. These stocks are often seen as undervalued by the market, providing opportunities for higher returns as their prices adjust upward over time. In contrast, growth stocks are priced high relative to their fundamentals, reflecting expectations of strong future growth.

The value premium, i.e., the superior performance of value stocks compared to growth stocks, has been documented extensively. Theories explaining this phenomenon include the risk-based view, suggesting that value stocks are inherently riskier and thus command higher returns, and the behavioral perspective, which posits that investors systematically overpay for growth stocks due to overoptimism.

Investment Investment factors examine the relationship between a firm's investment activities and its stock returns. Firms' investment decisions, such as capital expenditures, asset growth, and changes in total assets,[1] can provide insights into their future performance. Generally, firms that invest heavily may exhibit lower subsequent returns, a phenomenon known as the *asset growth effect*.

The rationale behind the investment effect lies in the efficient allocation of capital. Firms that invest heavily might be chasing growth opportunities that do not always yield the expected returns, leading to overinvestment. Conversely, firms that are more conservative in their investment strategies might be more

[1] See Chen & Chen (2012) and Cooper *et al.* (2008) for examples.

efficient in capital allocation, resulting in higher profitability and stock performance. Investment-based anomalies are often analyzed within the broader context of firm behavior and capital market dynamics.

Profitability Profitability factors focus on the relationship between a firm's earnings and its stock returns. Measures such as return on equity (ROE) (Hou *et al.*, 2015), gross profit margin (Novy-Marx, 2013), and operating profitability Fama & French (2015) are used to gauge a firm's efficiency in generating profits from its operations. High profitability is generally associated with higher stock returns, as profitable firms are better positioned to deliver strong performance and shareholder value.

The quality dimension of investing, encompassing profitability, indicates that high-quality firms with robust earnings can yield superior returns. This approach contrasts with value investing, as it emphasizes financial health and operational efficiency rather than just low prices.

Intangibles Intangible factors consider the role of non-physical and non-financial assets in influencing stock prices. These include investments in research and development (R&D), brand equity, patents, human capital, and size.[2] Intangible assets are often key drivers of innovation and competitive advantage, particularly in industries such as technology and pharmaceuticals.

The valuation of intangible assets poses a unique challenge because they are not always reflected on the balance sheet or are easily quantifiable. However, firms with substantial intangible assets often exhibit strong future growth potential, making them attractive investment targets. For example, high R&D spending can signal a commitment to innovation, which may translate into higher future earnings and stock performance. Intangible factors highlight the importance of looking beyond traditional financial metrics to capture the full value of a firm's assets.

Trading frictions Trading frictions encompass various impediments to market liquidity and efficiency. Bid-ask spreads, trading volume, transaction costs, and market microstructure issues represent some of the factors in this group - see (Amihud & Mendelson, 1989), Chordia *et al.* (2001), and Amihud (2002) for practical examples.

Understanding trading frictions is crucial for investors, as these factors can significantly impact the execution of investment strategies and the realized returns. For instance, small-cap stocks often exhibit higher trading frictions due

[2]See Banz (1981), Asness *et al.* (2000), and Belo *et al.* (2014) for some examples.

to lower liquidity, possibly leading to higher volatility and transaction costs. On the other hand, investors may *charge* a premium to bear risks associated with frictions, and bearing them may turn out to be profitable.

Table 4.1: Anomaly factors - Category classification

Abbreviation	Description	Category	Abbreviation	Description	Category
absacc	Absolute accruals	Investment	mom1m	1-month momentum	Momentum
acc	Working capital accruals	Investment	mom36m	36-13 months momentum	Momentum
aeavol	Abnormal earnings announcement volume	Momentum	mom6m	6-1 months momentum	Momentum
agr	Asset growth	Investment	ms	Financial statement score	Profitability
baspread	Bid-ask spread	Trading Friction	mve	Size	Intangibles
beta	Beta	Trading Friction	mve_ia	Industry adjusted size	Intangibles
betasq	Beta squared	Trading Friction	nincr	Number of earnings increases	Profitability
bm	Book-to-market	Value-versus-growth	operprof	Operating profitability	Profitability
bm_ia	Industry adjusted book-to-market	Value-versus-growth	pchcapx_ia	Industry adjusted % change in capital expenditures	Investment
cash	Cash holding	Value-versus-growth	pchcurrat	% change in current ratio	Value-versus-growth
cashdebt	Cash flow to debt	Value-versus-growth	pchdepr	% change in depreciation	Investment
cashpr	Cash productivity	Profitability	pchgm_pchsale	% change in gross margin - % change in sales	Profitability
cfp	Cash flow to price ratio	Value-versus-growth	pchquick	% change in quick ratio	Value-versus-growth
cfp_ia	Industry adjusted cfp	Value-versus-growth	pchsale_pchinvt	% change in sale - % change in inventory	Profitability
chatoia	Industry adjusted change in asset turnover	Profitability	pchsale_pchrect	% change in sale - % change in A/R	Profitability
chcsho	Change in share outstanding	Investment	pchsale_pchxsga	% change in sale - % change in SG&A	Profitability
chempia	Industry adjusted change in employees	Intangibles	pchsaleinv	% change in sales-to-inventory	Profitability
chinv	Change in inventory	Investment	pctacc	Percent accruals	Investment
chmom	Change in 6-month momentum	Momentum	pricedelay	Price delay	Trading Friction
chpmia	Industry adjusted change in profit margin	Profitability	ps	Financial statement score	Profitability
chtx	Change in tax expense	Profitability	quick	Quick ratio	Value-versus-growth
cinvest	Corporate investment	Investment	retvol	Return volatility	Trading Friction
currat	Current ratio	Value-versus-growth	roaq	Return on assets	Profitability
depr	Depreciation	Investment	roavol	Earning volatility	Profitability
dolvol	Dollar trading volume	Trading Friction	roeq	Return on equity	Profitability
dy	Dividend-to-price	Value-versus-growth	roic	Return on invested capital	Profitability
ear	Earnings announcement return	Momentum	rsup	Revenue surprise	Profitability
egr	Growth in common shareholder equity	Investment	salecash	Sales to cash	Profitability
ep	Earnings-to-price	Value-versus-growth	saleinv	Sales to inventory	Profitability
gma	Gross profitability	Profitability	salerec	Sales to receivables	Profitability
grcapx	Growth in capital expenditure	Investment	sgr	Sales growth	Value-versus-growth
grltnoa	Growth in long term net operating assets	Investment	sp	Sales-to-price	Value-versus-growth
hire	Employee growth rate	Intangibles	std_dolvol	Volatility of liquidity (dollar trading volume)	Trading Friction
idiovol	Idiosyncratic return volatility	Trading Friction	std_turn	Volatility of liquidity (share turnover)	Trading Friction
ill	Illiquidity	Trading Friction	stdacc	Accrual volatility	Investment
invest	Capital expenditure and inventory	Investment	stdcf	Cash flow volatility	Value-versus-growth
lev	Leverage	Value-versus-growth	tang	Debt capacity/firm tangibility	Profitability
lgr	Growth in long term debt	Investment	tb	Tax income to book income	Intangibles
maxret	Max daily return	Trading Friction	turn	Share turnover	Trading Friction
mom12m	12-1 months momentum	Momentum	zerotrade	Zero trading days	Trading Friction

Notes: This table lists all used factors. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024), while categories are in line with Hou *et al.* (2020).

Table 4.1 brings the anomaly factors from Table 1.1 classified by their categories. It is worth noting that the distribution of factors across categories is uneven, with profitability accounting for 23 of the 80 available factors - about 29% of the pool - while intangibles are represented by only 5 characteristics (around 6% of our sample). Value-versus-growth, investment, trading friction, and momentum are represented by 16, 16, 13, and 7 pricing factors respectively.

4.3 Factor hall of fame

This section highlights key qualitative aspects of the most relevant factors from the previous chapters and examines the proposed methodologies in parallel.

We begin in Subsection 4.3.1 by presenting the top 10 most frequent factors, followed by an exploration of factor categories in Subsection 4.3.2, where we compile each family's relative importance in each methodology. Finally, Subsection 4.3.3 analyzes the time persistence of relevant factor categories.

4.3.1 Most selected factors

Table 4.2 summarizes factors selected using the shrinkage model from Chapter 2, stating how often the pricing anomalies were included in the set of relevant factors.

Table 4.2: Most relevant factors - SDF-error model

BIC	FRC1	FRC3	FRC5	FRC10
pchcapx_ia (inv) - 54.75%	grltnoa (inv) - 25.10%	mve (int) - 32.32%	mve (int) - 46.77%	ear (mom) - 66.16%
grltnoa (inv) - 50.19%	pchcapx_ia (inv) - 19.39%	pchcapx_ia (inv) - 30.04%	chpmia (prof) - 41.06%	chpmia (prof) - 65.78%
chempia (int) - 43.73%	acc (inv) - 13.31%	grltnoa (inv) - 28.52%	cfp (value) - 41.06%	pchcurrat (value) - 59.32%
mve (int) - 42.59%	mve (int) - 6.84%	lgr (inv) - 23.57%	pchcapx_ia (inv) - 34.98%	cfp (value) - 56.65%
chpmia (prof) - 42.59%	pchgm_pchsale (prof) - 5.70%	pchsale_pchrect (prof) - 21.29%	pchsale_pchrect (prof) - 33.84%	mve (int) - 56.65%
ear (mom) - 42.59%	lgr (inv) - 4.94%	acc (inv) - 20.15%	pchcurrat (value) - 31.56%	pchgm_pchsale (prof) - 55.89%
pchgm_pchsale (prof) - 41.83%	pchcurrat (value) - 4.56%	pchcurrat (value) - 18.63%	grltnoa (inv) - 28.52%	pchsale_pchrect (prof) - 45.63%
pchcurrat (value) - 41.83%	mom12m (mom) - 4.56%	chpmia (prof) - 16.73%	pchgm_pchsale (prof) - 28.52%	pricedelay (trad) - 45.63%
chatoia (prof) - 39.16%	mom36m (mom) - 3.42%	cfp (value) - 14.07%	lgr (inv) - 23.19%	chempia (int) - 43.73%
cinvest (inv) - 38.78%	rsup (prof) - 2.66%	egr (inv) - 11.79%	acc (inv) - 21.29%	mom1m (mom) - 39.92%

Notes: This table lists the top 10 factors for each exercise presented in Chapter 2, alongside their qualitative group and the percentage of periods where they were considered relevant. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024), while categories are in line with Hou *et al.* (2020).

We must recall that setting up the LASSO's penalization parameter through BIC *does not limit* the total number of relevant factors - where it's easily verified that just the top two most frequently selected factors together exceed 100% - and that BIC *did not yield* the most prominent results, indicating that more weight should be given to the results from the sparser FRC models.

It is interesting that, when the penalty parameter is set to guarantee only one relevant factor (FRC1), all top 3 factors are *investment-related* - growth in long-term net operating assets (grltnoa), industry adjusted % change in capital expenditures (pchcapx_ia) and working capital accruals (acc). In more intermediate sparsity levels the size (mve) factor dominates, while the previous most important factors, despite appearing more times, lose relative importance. Finally, earnings announcement return (ear) stands out in FRC10: a factor that did not appear in the top 10 most selected factors for any of the more sparse FRC penalties.

Allowing for more factors to be set as relevant generally jeopardized out-of-sample returns - especially when we incorporate as many as 10 factors into the model (see Subsection 2.3.3). Examining the used factors, it is possible to infer that using less relevant factors may have forced the model to focus on stronger, more *tactical*, signals, that are not relevant all the time: only three of the top 10 factors for the FRC1 hold relevant at the FRC10's top 10, and *none* of the top 3 FRC1 factors remained at the top 10 for the FRC10.

As shown in Table 4.3, Chapter 3 methodology produced a completely *distinct* set of relevant pricing anomalies, resulting in a clear most relevant factor.

Table 4.3: Most relevant factors - factor spanning methodology (FRC5)

p-val0.10	lwst1	lwst3	lwst5
dolvol (trad) - 30.80%	dolvol (trad) - 33.46%	dolvol (trad) - 45.63%	dolvol (trad) - 49.81%
bm (value) - 17.87%	bm (value) - 16.73%	roic (prof) - 30.04%	mve (int) - 38.40%
roic (prof) - 12.17%	mve (int) - 8.37%	mve (int) - 26.62%	roic (prof) - 36.50%
mve (int) - 11.41%	operprof (prof) - 7.98%	bm (value) - 24.71%	ms (prof) - 30.80%
ms (prof) - 7.98%	gma (prof) - 5.32%	ms (prof) - 23.57%	bm (value) - 30.42%
gma (prof) - 5.70%	depr (inv) - 4.94%	operprof (prof) - 17.49%	pricedelay (trad) - 24.71%
depr (inv) - 5.32%	ms (prof) - 3.42%	cinvest (inv) - 11.79%	operprof (prof) - 23.19%
operprof (prof) - 4.18%	cinvest (inv) - 3.42%	cashpr (prof) - 10.65%	cinvest (inv) - 19.39%
cashpr (prof) - 3.80%	roic (prof) - 3.04%	pricedelay (trad) - 8.75%	cashpr (prof) - 18.63%
pricedelay (trad) - 3.42%	acc (inv) - 2.66%	depr (inv) - 8.37%	chatoia (prof) - 17.11%

Notes: This table lists the top 10 factors for each FRC5 exercise presented in Chapter 3, alongside their qualitative group and the percentage of periods where they were considered relevant. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024), while categories are in line with Hou *et al.* (2020).

Dollar trading volume (dolvol), a relatively simple liquidity proxy, was consistently the most relevant pricing factor for all the exercises proposed in Chapter 3 - even when setting up the penalty parameter through BIC.[3]. It is also interesting to highlight that the other two anomalies, book-to-market (bm) and size (mve), were among the top 5 most relevant ones for all exercises.

Moreover, three more factors were in the top 10 in all exercises, return on invested capital (roic), financial statement score (ms), and operating profitability (operprof), while corporate investment (cinvest) was in the top 10 whenever we set a pre-defined number of relevant factor. Table 4.3 paints the factor spanning regressions of Chapter 3 as a *stable* methodology for factor selection, as the selected anomalies *do not vary* abruptly with desired sparsity levels.

The difference in relevant factors' stability observed between the proposed methodologies could originate from the approach used to achieve sparsity. In Chapter 3, selected factors sparsity is mainly ensured by the intercepts' p-values, as we order them and choose the n lowest ones, while in Chapter 2 the sparsity is directly guaranteed by the shrinkage's penalty parameter. Changes in the penalty

[3]See Appendix 4.A.1.

parameter can modify the region for the optimal combination of $\beta's$, incurring more drastic variations of non-zero coefficients.

A direct comparison of Tables 4.2 and 4.3 also evidences that the *de facto* used factor differs immensely between methodologies. Subsection 4.3.2 will further investigate those differences, focusing on the *type* of relevant factors, but it should now be crystal clear that there the proposed methodologies capture distinct phenomena.

4.3.2
Most relevant factor category

The individual factor significance analysis introduced, despite being relevant, must be extended for a clearer understating of the most important factors, especially as some practitioners will find of the utmost importance the comprehension of the *type* of factor that is mostly relevant. Tables 4.4 and 4.5 bring the relative importance[4] of each factor category, defined at Subsection 4.2.1.

Table 4.4: Most relevant factor categories - SDF-error model

Group	BIC	FRC1	FRC3	FRC5	FRC10	Average
Investments	23.9%	65.4%	40.2%	26.3%	16.3%	34.4%
Profitability	28.0%	14.4%	22.3%	26.5%	27.6%	23.8%
Value-versus-growth	17.4%	4.9%	13.4%	20.6%	24.2%	16.1%
Intangibles	7.6%	6.8%	13.3%	13.2%	10.2%	10.2%
Momentum	10.6%	8.4%	10.5%	11.8%	15.5%	11.3%
Trading frictions	11.2%	0.0%	0.3%	1.7%	6.2%	3.9%
Market	1.3%	0.0%	0.0%	0.0%	0.0%	0.3%

Notes: This table quantifies the relative importance of each qualitative category of factors for each exercise presented in Chapter 2.

The SDF error-based regression of Chapter 2 yielded a set of factors highly concentrated in investments anomalies, especially when imposing higher degrees of sparsity: more than 65% of the relevant factors for the FRC1 regressions were from the investments category - see Table 4.4. As the imposed sparsity levels get less strict, the distribution of categories' relevance gets less concentrated, and investment-related anomalies lose some of their relative importance: for the FRC10 penalty, the most relevant categories are profitability (27.6%) and value-versus-growth (24.6%), only then followed by investments - at 16.3%.

The average relevance concentration between factor categories for the methodology from Chapter 3 (Table 4.5) shows a similar concentration pattern across factor categories as Chapter 2's. However, profitability becomes the most represented category, particularly in lower-sparsity models. Another enormous

[4]Notice that the numbers of each column will sum 100%, and represent the percentage of relevant factors for a given exercise that was classified as being of that category.

Table 4.5: Most relevant factor categories - factor spanning methodology (FRC5)

Group	p-val0.10	lwst1	lwst3	lwst5	Average
Profitability	32.8%	24.3%	40.4%	39.8%	34.4%
Trading frictions	30.7%	35.7%	20.8%	18.0%	26.3%
Value-versus-growth	19.5%	20.2%	16.6%	18.3%	18.6%
Investments	6.5%	11.4%	10.8%	11.6%	10.1%
Intangibles	9.6%	8.4%	10.5%	10.6%	9.8%
Market	0.3%	0.0%	0.5%	1.3%	0.5%
Momentum	0.6%	0.0%	0.4%	0.5%	0.4%

Notes: This table quantifies the relative importance of each qualitative category of factors for each FRC5 exercise presented in Chapter 3.

difference is the relevance attributed to trading friction factors, which skyrocketed from 3.9% to 26.3% of average presence, topping a whopping 35.7% presence when only the factor with the lowest p-value is considered.

On the other hand, the momentum category lost almost all of its relevance observed in the first model, plummeting from 11.3% average relative importance to an almost negligible 0.4%. Finally, the market value had an underwhelming appearance in both methodologies, with less than 1% average relevance, and value-versus-growth and intangible factors displayed stable relevance in both Chapters' factor selection approaches, of about 17% and 10%, respectively.

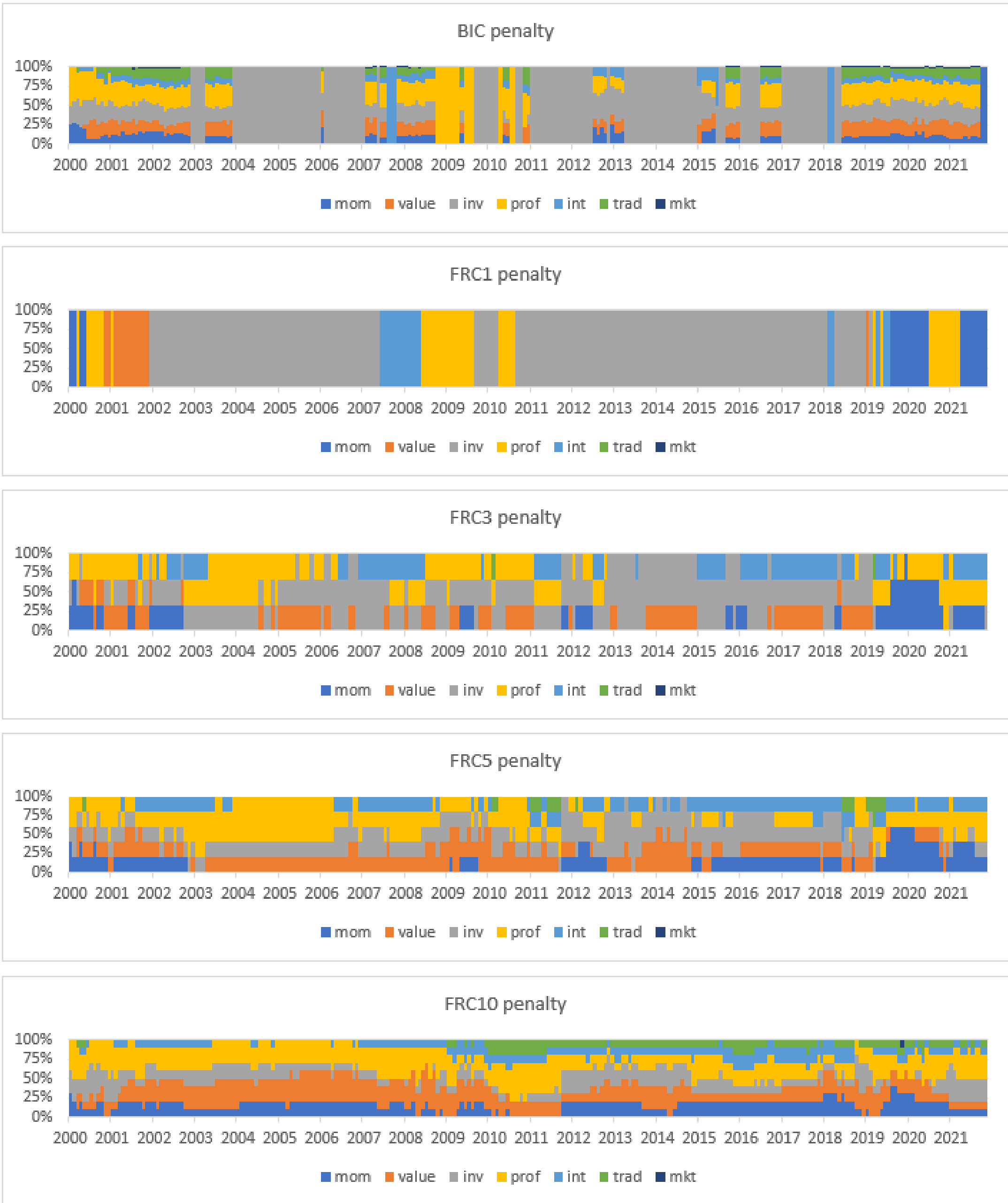
It is noteworthy to add that, as stated in Subsection 4.2.1, the distribution of the anomalies present in our database is not even between categories: profitability was as much as 23 factors, while value-versus-growth, investments, trading friction, momentum, and intangibles have 16, 16, 13, 7, and 5 factors, respectively. Results from Chapter 2, where investment factors were chosen frequently as relevant, represent a significant overweight on the category. Chapter 3 results also state a deviation from the equal-weight selection of factors, but less extreme - as investment factors already account for about 29% of the available pricing anomalies.

4.3.3 Timing of factor category relevance

The previous Subsections 4.3.1 and 4.3.2 set the foundations for comparing the *type* of factor set as relevant for each one of the exercises we have done. Now we will introduce a time-varying aspect to the analysis, as we aim to understand how the relevance of different pricing factor categories varies over time.

We explore how the category of used pricing factors vary with the time in Figures 4.1 and 4.2, where the relative factor relevance (same data from Tables 4.4 and 4.5 is plotted is respect to its time-series component, building upon the mean analysis presented on Subsection 4.3.2. Especial effort should be expended to examine potential *cyclicality* of relevant factors' categories, as those plots could unveil some macroeconomic/financial cycle-dependent patterns to be further explored.

Figure 4.1: Relevant factor categories over time - SDF-error model



Notes: This figure illustrates the time-series behavior of selected factors' categories for each exercise presented in Chapter 2.

We begin with relevant factors from Chapter 2 - see Figure 4.1. Set aside the first plot, where BIC is used to set the penalty parameter for the LASSO regression,[5] factor selection is notably stable. Most of the time, when a category

[5]As argued in Subsection 2.2.2, asset pricing is a *extremely noisy* environment, and usual

is set relevant in a period, it will stay relevant at the following selection, and the relevant categories remain unchanged for almost all the observed time.

Figure 4.2: Relevant factor categories over time - factor spanning methodology (FRC5)

Notes: This figure illustrates the time-series behavior of selected factors' categories for each FRC5 exercise presented in Chapter 3.

Advancing to results from Chapter 3, Figure 4.2 exhibits similar levels of relevant factors' categories stability, as relevant families seem to stay relevant throughout all the sample, with no major fluctuations on their representatives - especially when we focus on exercises that select factors with the lowest 3 and 5 p-values. If we focus on *the most significant* factor, we have a distinct category that is more relevant for each third of the time series: at the beginning the profitability family is dominant, in the second third the trading friction category appears more

Bayesian criterion may have trouble finding a minimum value when searching for the penalty parameter, resulting in some abrupt jumps for the used λ, leading to huge variations in the number of factors that survive the shrinkage process.

relevant, while on the final third value-versus-growth group takes over as the prevalent one.

The visualizations also highlight the distinctiveness of relevant categories across the proposed methodologies: Figure 4.1 illustrate *grayer* plots, representing a greater relevance of investment-related factors, while Figure 4.2 is *brighter* due to a dominance of the yellow and green colors that represent profitability and trading frictions anomalies. Those aspects were already enlightened in Subsection 4.3.2, but it is still interesting to visualize the phenomena.

Finally, possible macroeconomic cycle-dependent aspects of factors used in our exercises do not find support in either Figure 4.1 or 4.2, as there are any signs of cyclicality. Overall, the plots support the conclusion that relevant factor categories are *stable* across time for both methodologies, with no evident cyclicality: despite clearly selecting qualitative *distinct* pricing anomalies.

4.4 Final considerations

In this chapter, we qualitatively examined the pricing factors identified as relevant by the methodologies in Chapters 2 and 3.

We began by listing the top 10 most frequently selected factors for each of the nine main exercises in this dissertation, finding that while factor stability generally holds *within* each methodology across sparsity levels, it can shift *significantly* between methodologies.

Then we categorized factors according to their fundamentals. Profitability-related anomalies emerged as relevant in both methodologies, though they were surpassed by investment factors in the SDF-error model and closely followed by trading frictions in the factor-spanning approach. The differences in the *types* of pricing anomalies selected by each approach are notable, suggesting that each methodology captures a unique aspect of the asset pricing problem.

Finally, in observing the time variability of the selected factor types, we found considerable stability over the test period. This suggests that each proposed methodology yields a distinct yet persistently stable set of factor selections.

The existence of two distinct factor selection methodologies, both developed from the same dataset and yielding relevant out-of-sample results yet with markedly different qualitative aspects, is intriguing. We believe that each methodology may be capturing a unique aspect of price action, suggesting potential *complementarity*. Exploring this complementarity could be an interesting direction for future research.

4.A
Appendix

4.A.1
Most selected factors - Spanning regressions with BIC penalty

Table 4.A.1: Most relevant factors - factor spanning methodology (BIC)

p-val0.10	lwst1	lwst3	lwst5
dolvol (trad) - 27.38%	dolvol (trad) - 33.08%	dolvol (trad) - 45.25%	dolvol (trad) - 49.43%
ms (prof) - 16.73%	bm (value) - 12.17%	ms (prof) - 34.22%	ms (prof) - 44.49%
bm (value) - 12.93%	ms (prof) - 9.13%	roic (prof) - 31.56%	roic (prof) - 40.30%
roic (prof) - 9.89%	mve (int) - 7.22%	mve (int) - 23.19%	mve (int) - 33.84%
mve (int) - 9.13%	depr (inv) - 7.22%	bm (value) - 15.97%	pricedelay (trad) - 28.90%
depr (inv) - 8.75%	roic (prof) - 5.70%	depr (inv) - 12.93%	chatoia (prof) - 20.15%
gma (prof) - 5.32%	cinvest (inv) - 4.94%	pricedelay (trad) - 12.55%	bm (value) - 18.63%
cinvest (inv) - 1.90%	operprof (prof) - 3.04%	cinvest (inv) - 11.79%	cinvest (inv) - 17.49%
pricedelay (trad) - 1.90%	gma (prof) - 3.04%	cashpr (prof) - 10.65%	cashpr (prof) - 17.49%
currat (value) - 1.52%	chatoia (prof) - 2.66%	chatoia (prof) - 10.27%	pchcurrat (value) - 15.97%

Notes: This table lists the top 10 factors for each BIC exercise presented in Chapter 3, alongside their qualitative group and the percentage of periods where they were considered relevant. The abbreviation is consistent with Green *et al.* (2017) and Sun (2024), while categories are in line with Hou *et al.* (2020).

4.A.2
Most relevant factor category - Spanning regressions with BIC penalty

Table 4.A.2: Most relevant factor categories - factor spanning methodology (BIC)

Group	p-val0.10	lwst1	lwst3	lwst5	Average
Profitability	36.5%	27.8%	41.8%	40.3%	36.6%
Trading frictions	28.8%	35.0%	20.5%	17.6%	25.5%
Value-vesrus-growth	15.4%	15.6%	15.5%	16.4%	15.7%
Investments	9.8%	13.3%	11.3%	12.1%	11.6%
Intangibles	8.4%	7.2%	9.5%	11.2%	9.1%
Market	0.7%	0.8%	1.0%	1.5%	1.0%
Momentum	0.4%	0.4%	0.4%	0.9%	0.5%

Notes: This table quantifies the relative importance of each qualitative category of factors for each BIC exercise presented in Chapter 3.

4.A.3
Timing of factor category relevance - Spanning regressions with BIC penalty

Figure 4.A.1: Relevant factor categories over time - factor spanning methodology (BIC)

Notes: This figure illustrates the time-series behavior of selected factors' categories for each BIC exercise presented in Chapter 3.

Conclusion

This dissertation addresses the curse of dimensionality within the extensive zoo of factors in asset pricing, aiming to enhance factor selection methodologies in high-dimensional contexts. We begin by establishing a time-varying framework that minimizes potential biases and propose two distinct methodologies to manage factor dimensionality: (i) an approach leveraging the Stochastic Discount Factor (SDF) model error, and (ii) a method utilizing p-value calculations for intercepts in factor return spanning regressions.

The findings underscore the benefits of differentiating time series lengths for separate analytic purposes and imposing elevated sparsity levels in high-dimensional factor selection models, which yield comparable efficacy to classical, low-dimensional models. Our results reveal that while both methodologies produce out-of-sample Sharpe ratios of similar magnitudes, they diverge significantly in the types of factors deemed relevant, highlighting the methodological nuances in factor selection.

These findings collectively emphasize the critical role of carefully structured methodologies, indicating that even straightforward shrinkage regressions, when employed within a rigorously designed framework, can yield portfolios with substantial out-of-sample returns

Future research could extend these findings along several promising avenues. Our time-varying framework for factor selection accommodates the potential integration of more sophisticated shrinkage techniques, and further exploration in this direction is highly encouraged. It should also be interesting to replicate the proposed methodology to data sets that capture other periodicities - especially intraday data. Lastly, replicating the procedures for other economies could strengthen our findings, as, for the moment, we only have results for the United States equity market.

Bibliography

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

Amihud, Yakov. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, **5**(1), 31–56.

Amihud, Yakov, & Mendelson, Haim. 1989. The effects of beta, bid-ask spread, residual risk, and size on stock returns. *The Journal of Finance*, **44**(2), 479–486.

Asness, Clifford S, Porter, R Burt, & Stevens, Ross L. 2000. Predicting stock returns using industry-relative firm characteristics. *Available at SSRN 213872*.

Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of financial economics*, **9**(1), 3–18.

Basu, Sanjoy. 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The journal of Finance*, **32**(3), 663–682.

Belloni, Alexandre, Chernozhukov, Victor, & Hansen, Christian. 2014. Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, **81**(2), 608–650.

Belo, Frederico, Lin, Xiaoji, & Bazdresch, Santiago. 2014. Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, **122**(1), 129–177.

Benjamini, Yoav, & Hochberg, Yosef. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, **57**(1), 289–300.

Bonferroni, Carlo. 1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, **8**, 3–62.

Brennan, Michael J, & Xia, Yihong. 2001. Assessing asset pricing anomalies. *The Review of Financial Studies*, **14**(4), 905–942.

Bryzgalova, Svetlana, Huang, Jiantao, & Julliard, Christian. 2023. Bayesian solutions for the factor zoo: We just ran two quadrillion models. *The Journal of Finance*, **78**(1), 487–557.

Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of finance*, **52**(1), 57–82.

Carlstein, Edward. 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, 1171–1179.

Chen, Huafeng Jason, & Chen, Shaojun Jenny. 2012. Investment-cash flow sensitivity cannot be a good measure of financial constraints: Evidence from the time series. *Journal of Financial Economics*, **103**(2), 393–410.

Chordia, Tarun, Subrahmanyam, Avanidhar, & Anshuman, V Ravi. 2001. Trading activity and expected stock returns. *Journal of financial Economics*, **59**(1), 3–32.

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

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

Cooper, Michael J, Gulen, Huseyin, & Schill, Michael J. 2008. Asset growth and the cross-section of stock returns. *the Journal of Finance*, **63**(4), 1609–1651.

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

Fama, Eugene F, & French, Kenneth R. 2008. Dissecting anomalies. *The Journal of Finance*, **63**(4), 1653–1678.

Fama, Eugene F, & French, Kenneth R. 2015. A five-factor asset pricing model. *Journal of financial economics*, **116**(1), 1–22.

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

Feng, Guanhao, Giglio, Stefano, & Xiu, Dacheng. 2020. Taming the factor zoo: A test of new factors. *The Journal of Finance*, **75**(3), 1327–1370.

Figueiredo, Mario, & Nowak, Robert. 2016. Ordered weighted l1 regularized regression with strongly correlated covariates: Theoretical aspects. *Pages 930–938 of: Artificial Intelligence and Statistics*. PMLR.

Frazzini, Andrea, & Pedersen, Lasse Heje. 2014. Betting against beta. *Journal of financial economics*, **111**(1), 1–25.

Freyberger, Joachim, Neuhierl, Andreas, & Weber, Michael. 2020. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, **33**(5), 2326–2377.

Green, Jeremiah, Hand, John RM, & Zhang, X Frank. 2017. The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies*, **30**(12), 4389–4436.

Gupta, Tarun, & Kelly, Bryan. 2019. Factor momentum everywhere. *The Journal of Portfolio Management*, **45**(3), 13–36.

Hall, Peter. 1985. Resampling a coverage pattern. *Stochastic processes and their applications*, **20**(2), 231–246.

Harvey, Campbell R, & Liu, Yan. 2021. Lucky factors. *Journal of Financial Economics*, **141**(2), 413–435.

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

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Hou, Kewei, Xue, Chen, & Zhang, Lu. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies*, **28**(3), 650–705.

Hou, Kewei, Xue, Chen, & Zhang, Lu. 2020. Replicating anomalies. *The Review of financial studies*, **33**(5), 2019–2133.

Houweling, Patrick, & Van Zundert, Jeroen. 2017. Factor investing in the corporate bond market. *Financial Analysts Journal*, **73**(2), 100–115.

Huang, Jian, Horowitz, Joel L, & Wei, Fengrong. 2010. Variable selection in nonparametric additive models. *Annals of Statistics*, **38**, 2282–313.

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

Jensen, Theis Ingerslev, Kelly, Bryan, & Pedersen, Lasse Heje. 2023. Is there a replication crisis in finance? *The Journal of Finance*, **78**(5), 2465–2518.

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

Kozak, Serhiy, Nagel, Stefan, & Santosh, Shrihari. 2020. Shrinking the cross-section. *Journal of Financial Economics*, **135**(2), 271–292.

Lewellen, J. 2015. *The Cross-section of Expected Stock Returns. Critical Finance Review, 4 (1), 1–44.*

Lintner, John. 1965. 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.

Loh, Roger K, & Warachka, Mitch. 2012. Streaks in earnings surprises and the cross-section of stock returns. *Management Science*, **58**(7), 1305–1321.

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

Meinshausen, Nicolai, & Bühlmann, Peter. 2006. Variable selection and high-dimensional graphs with the lasso. *Annals of Statistics*, **34**, 1436–1462.

Meinshausen, Nicolai, Meier, Lukas, & Bühlmann, Peter. 2009. P-values for high-dimensional regression. *Journal of the American Statistical Association*, **104**(488), 1671–1681.

Novy-Marx, Robert. 2013. The other side of value: The gross profitability premium. *Journal of financial economics*, **108**(1), 1–28.

Ou, Jane A, & Penman, Stephen H. 1989. Financial statement analysis and the prediction of stock returns. *Journal of accounting and economics*, **11**(4), 295–329.

Rosenberg, Barr, Reid, Kenneth, & Lanstein, Ronald. 1985. Persuasive evidence of market inefficiency. *Journal of portfolio management*, **11**(3), 9–16.

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

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

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

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

Sun, Chuanping. 2024. Factor correlation and the cross section of asset returns: A correlation-robust machine learning approach. *Journal of Empirical Finance*, **77**, 101497.

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

Wasserman, Larry, & Roeder, Kathryn. 2009. High dimensional variable selection. *Annals of statistics*, **37**(5A), 2178.

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