Three Essays on Machine Learning in Empirical Finance

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

vorgelegt

von

M.Sc. Sebastian Weibels

aus

Oberhausen

Referent: Prof. Dr. Tom Zimmermann

Korreferent: Prof. Dr. Dieter Hess

Tag der Promotion: 27.04.2026

Acknowledgements

This thesis marks the completion of my doctoral studies in the Department of Econometrics and Statistics and the Department of Finance at the University of Cologne.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Tom Zimmermann, for his continuous support, guidance, and advice throughout my doctoral journey. His expertise has been instrumental not only in shaping my research but also in influencing my personal development. I am also indebted to my co-supervisor, Prof. Dr. Dieter Hess, for his mentorship and encouragement. In addition, I am sincerely grateful to Prof. Dr. Hartmann-Wendels for his supervision during the early stages of my doctoral studies.

I am also deeply grateful to my co-authors, Dr. Frederik Simon, Prof. Dr. Dieter Hess, and Prof. Dr. Tom Zimmermann, for their collaboration, dedication, and invaluable contributions to our research. I have greatly benefited from their input and expertise.

Moreover, I am thankful to my colleagues in the Department of Econometrics and Statistics, the Department of Finance, and the Centre for Financial Research at the University of Cologne for their support and for the many stimulating discussions we shared. It was a wonderful time that will always remain a special memory for me.

I would particularly like to thank my friends, who supported me throughout this journey and whose friendship and sense of humor brought me joy and balance beyond academia, providing welcome distractions from the demands of my research.

Lastly, I would like to thank my family and loved ones for their unwavering encouragement and the opportunities they have given me. Your support throughout all stages of my life has been a constant source of strength. Without a doubt, I would not have achieved this without you. For this, and for so much more, I am endlessly grateful.

Sebastian Weibels

Contents

Acknowledgements **I**

List of Tables **VIII**

List of Figures **X**

1 Introduction **1**

2 Deep Parametric Portfolio Policies **7**

2.1 Introduction 7

2.2 Theory 12

2.2.1 Expected utility framework and parametric portfolio policies . . . 12

2.2.2 Risk aversion as economic regularization 14

2.3 Estimation and results 19

2.3.1 Network architecture 19

2.3.2 Data 23

2.3.3 Performance results for CRRA investors 24

2.3.4 Supporting results 28

2.3.5 Robustness 31

2.3.6 Market frictions 33

2.4 Alternative investor utility functions 36

2.4.1 Mean-variance and loss aversion 36

2.4.2 Portfolio moments across utility functions and models 39

2.5 Conclusion 41

3 Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data **43**

3.1 Introduction . . . 43
3.2 Empirical approach . . . 51
3.2.1 General setup . . . 51
3.2.2 Data . . . 52
3.2.3 Models . . . 53
3.2.4 Out-of-sample approach . . . 54
3.2.5 Evaluation . . . 55
3.2.6 Interpretation . . . 56
3.3 Evaluation . . . 58
3.3.1 Accuracy and bias . . . 58
3.3.2 Out-of-sample R^2 . . . 61
3.3.3 Implied cost of capital . . . 63
3.4 Interpretation . . . 65
3.4.1 Variable importance . . . 65
3.4.2 Group importance . . . 67
3.4.3 Nonlinearity . . . 72
3.4.4 Subsamples . . . 74
3.5 Conclusion . . . 83

4 Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns 85

4.1 Introduction . . . 85
4.2 Constructing a measure of firm atypicality . . . 93
4.3 Data and model . . . 97
4.3.1 Data . . . 97
4.3.2 Control variables . . . 98
4.3.3 Autoencoder model . . . 99
4.3.4 Descriptive statistics of ATYP . . . 101
4.4 ATYP as a measure of information-processing difficulty . . . 104
4.4.1 Relation to other proxies of information-processing difficulty . . . 104
4.4.2 Incremental explanatory power . . . 106
4.4.3 Post-earnings-announcement drift . . . 108
4.4.4 Industry information diffusion . . . 108

4.5 Atypical firms and the cross-section of expected returns . . . 110
4.5.1 Univariate portfolio sorts . . . 110
4.5.2 Bivariate portfolio sorts . . . 114
4.5.3 Fama–MacBeth regressions . . . 114
4.5.4 Return persistence . . . 116
4.6 Channels of return predictability . . . 118
4.6.1 Mispricing vs. risk . . . 119
4.6.2 Limits to arbitrage . . . 120
4.6.3 Limited attention . . . 122
4.7 Robustness checks . . . 124
4.7.1 Firm characteristic clusters . . . 124
4.7.2 Within-industry construction . . . 124
4.7.3 Within-size group construction . . . 125
4.7.4 Alternative factor models . . . 127
4.7.5 Alternative data sources . . . 128
4.7.6 Different model specifications . . . 129
4.7.7 Autoencoder vs. PCA . . . 129
4.7.8 Different measurement specifications . . . 130
4.7.9 Missing data and imputation . . . 131
4.7.10 Impact of extreme observations . . . 132
4.7.11 International robustness . . . 133
4.8 Conclusion . . . 135

A Appendix to Chapter 2 **137**
A.1 Proofs . . . 137
A.1.1 Proof of Proposition 1 . . . 137
A.1.2 Proof of Proposition 3 . . . 138
A.2 Neural network configuration . . . 138
A.3 Loss aversion as economic regularization . . . 143
A.4 Additional results and extensions . . . 145
A.4.1 Partial dependence and surrogate models . . . 145
A.4.2 Rolling window estimation . . . 150
A.4.3 Adding macroeconomic variables . . . 152

A.5 Comparison of portfolio weights and variable importance 156
A.6 Supplementary figures 158
A.7 Supplementary tables 160

B Appendix to Chapter 3 **174**
B.1 Traditional earnings prediction models 174
B.2 Machine learning earnings prediction models 176
B.3 Implied cost of capital models 187
B.4 Subsample analysis 189

C Appendix to Chapter 4 **193**
C.1 Autoencoder fit 193
C.2 Supplementary figures 197
C.3 Supplementary tables 202

Bibliography **224**

Lebenslauf **236**

Eidesstattliche Erklärung **237**

List of Tables

2.1 Deep portfolio policy for CRRA investors with different degrees of risk aversion . . . 25
2.2 Comparing deep portfolio policies to benchmark strategies . . . 32
2.3 Long-only & transaction costs constrained deep portfolio policy for CRRA investors with different degrees of risk aversion . . . 35
2.4 Deep portfolio policy for mean-variance and loss-averse investors with different degrees of risk aversion (γ) and loss aversion (l) . . . 38

3.1 Median PFE . . . 60
3.2 Median PAFE . . . 61
3.3 Average out-of-sample R^2 . . . 63
3.4 Variable importance by groups . . . 69
3.5 Variable importance: financial statements . . . 71
3.6 Variable importance by financial statement type per life cycle stage . . . 77
3.7 Surrogate R^2 per life cycle stage . . . 77
3.8 Variable importance by financial statement type per size tercile . . . 79
3.9 Surrogate R^2 per size tercile . . . 80
3.10 Variable importance by financial statement type per industry . . . 82
3.11 Surrogate R^2 per industry . . . 82

4.1 Descriptive statistics . . . 99
4.2 Cross-sectional correlations to ATYP . . . 104
4.3 Explaining forecast dispersion, forecast accuracy, return volatility, and abs. earnings surprise . . . 107
4.4 ATYP and post-earnings-announcement drift . . . 109

4.5 Univariate portfolio sorts of high-ATYP firms on industry information shocks . . . 111
4.6 Univariate portfolio sorts on ATYP . . . 113
4.7 Bivariate portfolio sorts . . . 115
4.8 Fama-MacBeth cross-sectional regressions . . . 117
4.9 Mispricing and ATYP . . . 120
4.10 Limits to arbitrage and ATYP . . . 121
4.11 Attention constraints and ATYP . . . 123

A.1 Hyperparameters . . . 141
A.2 Deep portfolio policy for CRRA investors with different degrees of risk aversion with expanding vs. rolling window estimation . . . 153
A.3 Deep portfolio policy for CRRA investors with different degrees of risk aversion including macro variables . . . 154
A.4 Cross-evaluation of portfolio strategies against different utility preferences 160
A.5 Predictor variables for the (D)PPP . . . 161
A.6 Transaction cost constrained deep portfolio policy for CRRA investors with different degrees of risk aversion . . . 168
A.7 Long-only deep portfolio policy for CRRA investors with different degrees of risk aversion . . . 169
A.8 Long-only & transaction cost constrained deep portfolio policy for CRRA investors with different degrees of risk aversion . . . 170
A.9 Deep portfolio policy for mean-variance investors with different degrees of risk aversion . . . 171
A.10 Deep portfolio policy for loss-averse investors with different degrees of loss aversion . . . 172
A.11 Long-only deep portfolio policy for mean-variance and loss-averse investors with different degrees of risk aversion (γ) and loss aversion (l) . . . 173

B.1 Traditional earnings prediction models . . . 175
B.2 Hyperparameters for the machine learning models . . . 176
B.3 Predictor variables for the machine learning models . . . 177
B.4 Implied cost of capital models . . . 188
B.5 Firm life cycle stages . . . 189

B.6 Industry classifications . . . 190
B.7 Median PAFE per life cycle stage . . . 191
B.8 Median PAFE per size tercile . . . 191
B.9 Median PAFE per industry . . . 192

C.1 Average stock characteristics of ATYP-sorted portfolios . . . 202
C.2 Fama-MacBeth cross-sectional regressions with post-earnings-announcement drift . . . 203
C.3 Fama-MacBeth cross-sectional regressions with limits to arbitrage proxies 204
C.4 Fama-MacBeth cross-sectional regressions with attention proxies . . . 205
C.5 Univariate portfolio sorts on ATYP by firm characteristic clusters . . . 206
C.6 Univariate portfolio sorts on industry-adjusted ATYP . . . 207
C.7 Within-industry portfolio sorts on industry-adjusted ATYP . . . 208
C.8 Univariate portfolio sorts on size-adjusted ATYP . . . 209
C.9 Within-size group portfolio sorts on size-adjusted ATYP . . . 210
C.10 Univariate portfolio sorts on ATYP with alternative factor models . . . 211
C.11 Univariate portfolio sorts on ATYP for WRDS financial ratios . . . 212
C.12 Univariate portfolio sorts on ATYP for the complete Jensen et al. (2023) dataset . . . 213
C.13 Univariate portfolio sorts on ATYP for different model specifications . . . 214
C.14 Univariate portfolio sorts on $ATYP_{AE}$ orthogonalized with respect to $ATYP_{PCA}$ . . . 216
C.15 Univariate portfolio sorts on median ATYP . . . 217
C.16 Univariate portfolio sorts on rank ATYP . . . 218
C.17 Univariate portfolio sorts on ATYP using non-missing characteristics . . . 219
C.18 Univariate portfolio sorts on ATYP orthogonalized with respect to imputation intensity . . . 220
C.19 Univariate portfolio sorts on ATYP orthogonalized with respect to extreme observations . . . 221
C.20 Univariate portfolio sorts on ATYP in developed countries (excl. USA) . . . 222
C.21 Univariate portfolio sorts on ATYP in Europe . . . 223

List of Figures

2.1 Risk aversion as economic regularization 19
2.2 Neural network structure 21
2.3 Mean absolute portfolio weight differences by risk aversion 26
2.4 Cumulative performance over time for CRRA preferences 27
2.5 Variable importance for the CRRA preference for the DPPP and the PPP . 29
2.6 Portfolio moments across utility functions 40

3.1 PAFE across out-of-sample periods 62
3.2 R^2 across out-of-sample periods 64
3.3 Portfolio evaluation 65
3.4 Variable importance for the machine learning forecast ensemble 67
3.5 Correlation heatmap for the most important variables 68
3.6 Surrogate models 73
3.7 Partial dependence plots 74
3.8 Accuracy differences per life cycle stage 76
3.9 Accuracy differences per size tercile 78
3.10 Accuracy differences per industry 81

4.1 Distribution and time-series of ATYP 102
4.2 Correlation of ATYP with established proxies of information-processing difficulty 106
4.3 Long-term predictive power 118

A.1 Out-of-sample testing strategy 142
A.2 Loss aversion as economic regularization 145

A.3 Mean returns, standard deviations and Sharpe ratios of one-dimensional portfolio sorts 146
A.4 Marginal contribution of characteristics to portfolio weights in the DPPP 149
A.5 Surrogate R^2 and difference in certainty equivalent in the DPPP 151
A.6 Variable importance for the CRRA including macro variables for the DPPP 155
A.7 Net exposure to clusters of firm characteristics for different preferences . 157
A.8 Cumulative performance over time for benchmark strategies 158
A.9 Top 20 characteristics for Sharpe ratio utility and ML model 158
A.10 Mean-variance utility 159
A.11 Loss aversion 159
A.12 Cumulative performance over time for MV and LA preferences 159

C.1 Distribution of ATYP across bottleneck dimensions 194
C.2 Pairwise correlations of raw characteristics vs. residuals 195
C.3 MSE and bottleneck importance by cluster 196
C.4 Graphical illustration of an autoencoder 197
C.5 Reconstruction accuracy and atypical firms 198
C.6 Cumulative performance over time for ATYP portfolio sorts 199
C.7 Transition probabilities 200
C.8 Distribution of ATYP within size groups 201

Chapter 1

Introduction

The landscape of empirical financial research has undergone a substantial shift over the last decade, driven largely by the exponential growth in available data and the concurrent rise in computational power. For the better part of the twentieth century, the field was dominated by linear paradigms and a focus on parsimony. Asset pricing models, such as the Capital Asset Pricing Model (CAPM) and the Fama-French factor models (e.g., Fama and French, 1993, 2015), relied on the assumption that the cross-section of returns could be spanned by a small number of linear risk factors. Similarly, in corporate finance and accounting, forecasting models for fundamental variables like earnings were typically constrained to linear regressions using a handful of predictors (e.g., Hou et al., 2012; Li and Mohanram, 2014). While these foundational theories provided the bedrock of modern finance, the complexity of financial markets often defies such simple, linear characterizations.

We have now entered an era defined by high-dimensional data. The universe of potential return predictors, the so-called "factor zoo", has expanded dramatically (Harvey et al., 2016), rendering traditional econometric techniques inadequate due to the curse of dimensionality. Simultaneously, the assumption of linearity has been challenged by growing evidence of complex interaction effects and nonlinear dependencies between economic variables and asset prices (Gu et al., 2020). In this context, machine learning (ML) has emerged not merely as a tool for prediction, but as a robust framework for testing economic theories and uncovering structural relationships that remain hidden to standard linear models.

However, the integration of machine learning into finance is not without challenges. The primary critique levied against these methods is their "black box" nature. Researchers

require transparency, i.e., they need to understand *why* a model makes a specific prediction and whether that prediction aligns with economic intuition (Israel et al., 2020). Furthermore, financial data is characterized by low signal-to-noise ratios, making complex models particularly susceptible to overfitting.

This dissertation, titled *Three Essays on Machine Learning in Empirical Finance*, seeks to bridge the divide between sophisticated machine learning techniques and economic theory. Across three self-contained chapters, I demonstrate that the flexibility of deep learning and other nonlinear methods can be harnessed to improve performance while deepening our understanding of economic mechanisms, specifically in the domains of portfolio optimization, earnings forecasting, and asset pricing frictions.

The first essay, presented in Chapter 2, addresses the classic problem of optimal asset allocation. Since Markowitz (1952), the mean-variance framework has served as the standard for portfolio choice. However, implementing this framework in practice is notoriously difficult due to estimation errors in expected returns and covariance matrices (Merton, 1980). As the number of assets increases, the number of parameters to estimate explodes, often resulting in optimized portfolios that perform worse out-of-sample than naive diversification strategies (DeMiguel et al., 2009). Brandt et al. (2009) offered a seminal solution to this dimensionality problem by parameterizing portfolio weights directly as a linear function of firm characteristics. This *Parametric Portfolio Policy* (PPP) reduces the estimation problem to a small set of coefficients, but it imposes a rigid linear structure on the relationship between characteristics and optimal weights.

In this chapter, titled *Deep Parametric Portfolio Policies*, my co-authors and I generalize the PPP approach by replacing the linear specification with a feed-forward neural network (Simon et al., 2025). This methodological innovation, termed the *Deep Parametric Portfolio Policy* (DPPP), allows for nonlinear relationships and complex interactions between firm characteristics to determine portfolio weights directly. While Kelly et al. (2024) argue that model complexity is a virtue in return prediction, we explore this concept in the context of direct utility maximization. We utilize a comprehensive dataset of 157 firm characteristics to train the model, allowing it to exploit a vast information set (Chen and Zimmermann, 2022).

The primary theoretical contribution of this essay is the establishment of a formal link between investor preferences and effective model complexity. We introduce the concept of *economic regularization*. In standard machine learning, regularization (such as

Lasso or Ridge regression) is a statistical tool used to prevent overfitting by penalizing large coefficients. We demonstrate that for an investor maximizing expected utility, the coefficient of risk aversion (γ) naturally performs this regularizing function. As risk aversion increases, the investor places a higher penalty on the variance of portfolio returns. Since estimation error contributes to variance, a highly risk-averse investor effectively penalizes model complexity.

Our theoretical proofs and simulation evidence show that as $\gamma \to \infty$, the optimal weights derived from the complex DPPP converge toward those of the simpler linear PPP. This implies that the optimal degree of model complexity is endogenous to the investor's preferences. An aggressive investor is willing to tolerate the higher estimation risk associated with a flexible nonlinear model in pursuit of higher returns, whereas a conservative investor naturally gravitates toward simpler specifications.

Empirically, we find that the DPPP delivers substantial certainty equivalent gains over the linear benchmark, ranging from 43 to 102 basis points depending on risk aversion. Furthermore, by analyzing variable importance, we show that while return-based signals dominate for low risk aversion, the model shifts toward a balanced mix of accounting and price-based variables as risk aversion rises. This chapter provides a framework where the flexibility of neural networks is disciplined by the economic objective of the investor.

The second essay, found in Chapter 3, shifts the focus to the fundamental drivers of asset value, namely corporate earnings, and is titled *Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data*. This study challenges the dominance of linear models in earnings prediction (Hess et al., 2025). Earnings are a central variable in finance, influencing valuation, executive compensation, and market efficiency. Historically, statistical forecasting models have been parsimonious, relying on simple time-series properties or a small set of predictors like accruals and past earnings (Hou et al., 2012; Li and Mohanram, 2014).

We argue that restricting prediction models to a small set of variables ignores the vast information content available in modern financial reporting. While recent studies have begun to apply machine learning to earnings, they often rely on restricted input sets or focus on earnings *changes* rather than levels (Chen et al., 2022). We address this gap by utilizing the full spectrum of Compustat data, i.e., over 500 predictor variables including lags and differences, to predict earnings per share using an ensemble of machine learning models, including Random Forests, Gradient Boosted Trees, and Neural Networks.

Our results demonstrate that the machine learning ensemble (ENML) significantly outperforms traditional linear benchmarks. For a one-year forecast horizon, the ENML reduces the median price-scaled absolute forecast error (PAFE) by approximately 11% compared to the best linear alternative. This performance advantage persists across forecast horizons extending up to five years. Crucially, we show that this statistical accuracy translates into economic value. A long-short investment strategy based on Implied Cost of Capital (ICC) derived from our machine learning forecasts generates higher returns than strategies based on traditional models.

The core contribution of this chapter lies in its rigorous application of interpretability techniques. We employ SHAP values to open the black box and quantify variable importance (Lundberg and Lee, 2017). This analysis reveals a distinct term structure of accounting information. For short-term forecasts ($t + 1$), income statement variables such as current earnings and EBIT are the dominant predictors, aligning with the intuition of linear models. However, as the forecast horizon extends to five years, the importance of balance sheet information rises significantly, eventually rivaling the income statement in predictive power. This suggests that while flow variables predict immediate future flows, the stock variables representing the firm's asset base and financial health are the true determinants of long-term profitability. Furthermore, using surrogate modeling, we decompose the nonlinearities in the model. We find that while linear approximations work reasonably well for stable firms in the short run, nonlinear interactions are essential for accurate predictions in longer horizons and for firms in volatile life cycle stages (Dickinson, 2011).

The third essay, *Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns*, presented in Chapter 4, investigates the role of information frictions in financial markets (Weibels, 2025). Standard asset pricing theory assumes that information is processed instantaneously. In reality, investors have limited attention and bounded processing capacities (Simon, 1955; Peng and Xiong, 2006). When faced with a large amount of data and limited attention, investors often rely on heuristics and categorical thinking (e.g., Barberis and Shleifer, 2003; Peng and Xiong, 2006; Gabaix, 2014; Kacperczyk et al., 2016).

I posit that firms whose characteristics conform to systematic patterns (e.g., a standard "value" firm or "growth" firm) are easy to process and value. In contrast, firms that exhibit unusual or idiosyncratic combinations of characteristics defy standard categorization

and are therefore hard to process. To operationalize this concept, I introduce a novel, data-driven measure called *ATYP* (Atypicality). Unlike previous proxies for complexity that rely on 10-K readability (Loughran and McDonald, 2024) or segment counts (Cohen and Lou, 2012), ATYP is derived from the underlying economic characteristics of the firm.

I train an autoencoder, i.e., an unsupervised neural network designed to compress high-dimensional data into a latent space, to learn the systematic patterns common to most firms. The reconstruction error, which measures the difference between a firm's actual characteristics and the model's attempt to reproduce them, serves as the proxy for atypicality. A high ATYP score indicates that a firm's profile cannot be explained by the systematic patterns that describe the market, making it cognitively costly to process.

I document a strong negative relationship between ATYP and future stock returns. A portfolio that goes long on low-ATYP (typical) firms and short on high-ATYP (atypical) firms generates a significant alpha, robust to standard risk factors including the Fama-French six-factor model (Fama and French, 2018). The equal-weighted spread is approximately 1.47% per month.

To validate ATYP as a measure of information-processing frictions, I show that high-ATYP firms are associated with higher analyst forecast dispersion, lower forecast accuracy, and higher idiosyncratic volatility. Furthermore, atypical firms exhibit delayed price responses to information, such as stronger post-earnings-announcement drift and sluggish reactions to industry shocks.

Finally, I provide extensive evidence that this premium is driven by mispricing. The negative return premium is concentrated in stocks with high limits to arbitrage (e.g., high idiosyncratic volatility) and low investor attention (e.g., small size, low analyst coverage), suggesting that atypical firms become overpriced due to investor disagreement or slow information diffusion that arbitrageurs fail to correct (Miller, 1977; Hong and Stein, 1999).

Together, these three essays demonstrate the potential of machine learning in empirical finance. They share a common methodological thread in that they use advanced computational techniques not as ends in themselves, but as tools to answer economic questions that were previously intractable. The unifying conclusion of this dissertation is that the integration of machine learning into finance requires a dual focus. While predictive power is undeniable, the true value to the academic community lies in interpretability and alignment with economic logic. By imposing economic constraints,

dissecting variable importance, and using reconstruction errors to proxy for cognitive frictions, this research shows how data-driven methods can enrich our understanding of market efficiency, investor behavior, and asset prices. The following chapters outline these contributions, providing a roadmap for a future in which empirical finance and data science are closely intertwined.

To conclude the introduction, I outline my contributions to the three essays in this thesis. The initial research idea for the first essay was proposed by me. We collaboratively surveyed the literature and refined the concept through multiple discussions. I conducted the majority of the empirical analyses and continuously updated the empirical models based on input from my co-authors. The first draft of the paper was also jointly written. Lastly, we jointly revised the study multiple times based on feedback we got from numerous conferences and seminars that we presented the study on. The idea behind the second essay emerged from several discussions between my co-authors and me. Again, we collaboratively surveyed the literature and refined the concept through multiple discussions. My co-authors did the majority of the empirical analyses and continuously updated the empirical models based on discussions. The first draft of the study was jointly written and revised several times based on feedback we received. The third study is solo-authored. I came up with the research idea, conducted the empirical analysis, and wrote the paper.

Chapter 2

Deep Parametric Portfolio Policies*

2.1 Introduction

Consider the formidable problem of an investor who wants to choose an optimal asset allocation within her equity portfolio. The literature provides her with a few options: She can opt for a traditional Markowitz approach (Markowitz, 1952) which requires estimating expected returns, variances and covariances, with the number of moments to estimate increasing rapidly in the number of assets. At the other end of the spectrum, she might estimate a low-dimensional parametric portfolio policy (PPP) (Brandt et al., 2009) but a linear model might not provide sufficient flexibility. She can also consult a large literature that relates characteristics to expected returns but even studies that consider a multitude of firm-level characteristics (e.g., Gu et al., 2020) only investigate expected returns and do not speak to risk as perceived by different investors' objective functions.

We provide a general solution to the portfolio optimization challenge. In short, we combine the parametric portfolio policy approach that can estimate portfolio weights for any utility function with the flexibility of feed-forward networks from the machine learning literature. The resulting approach that we label *Deep Parametric Portfolio Policy* (DPPP) is well-suited to accommodate flexible nonlinear and interactive relationships

*This chapter is based on Simon et al. (2025). We thank Victor DeMiguel, Christian Fieberg, Bryan Kelly, Alexander Klos, Simon Rottke, Mark Salmon, Fabricius Somogyi (discussant), Bastidon Cécile (discussant), Heiner Beckmeyer (discussant) and seminar participants at the Amsterdam Business School, the Research in Behavioral Finance Conference (RBFC), the Cardiff Fintech Conference, the 2022 New Zealand Finance Meeting (NZFM), the Paris Financial Management Conference (PFMC), the Theory-based Empirical Asset Pricing Research (TBEAR) Network Workshop 2023, the University of Liechtenstein, University Pompeu Fabra, the CEQURA Conference 2023 on Advances in Financial and Insurance Risk Management, the BVI-CFR Event 2023, the 4th Frontiers of Factor Investing 2024 Conference, the Oxford-Man Institute of Quantitative Finance, and the International Monetary Fund (IMF) Brownbag Research Seminar for helpful comments and suggestions.

between portfolio weights and stock characteristics, to integrate different utility functions, to deal with leverage or portfolio weight constraints, and to incorporate transaction costs. Importantly, the model also allows us to study the relationship between model complexity and investor preferences.

The contributions of our paper are fourfold. First, we advance the theoretical literature by formally linking investor risk preferences with effective model complexity through the mechanism of economic regularization. Second, we extend the parametric portfolio policy framework by integrating deep neural networks to capture nonlinear and interactive effects, thereby offering a more flexible approach to portfolio optimization. Third, our empirical results show that the DPPP delivers economically meaningful utility gains over traditional linear models across a variety of settings, objective functions and benchmark models. Notably, investor preferences shape the optimal portfolio allocation in a manner similar to statistical regularization, but they do so with respect to the objectives investors actually care about. Fourth, our analysis of variable importance contributes to a deeper understanding of how different types of firm characteristics influence portfolio construction, particularly under varying degrees of risk aversion and preferences.

To the best of our knowledge, our study is the first to systematically explore how the benefits of a complex and flexible model vary for investors with different levels of risk aversion or different utility functions. A natural concern with parameter-rich models is their potential to overfit historical data. Overfitting leads to less reliable out-of-sample estimates and higher prediction variance. Since our portfolio policy approach maximizes the investor's objective function directly (as opposed to minimizing a statistical objective such as the squared distance between realized and predicted returns (Moritz and Zimmermann, 2016; Gu et al., 2020)), volatility of predictions becomes a systematic part of the economic objective. As risk aversion increases, the variance of portfolio returns becomes more important and leans against overfitting and thus model complexity. We refer to this mechanism as *economic regularization* (in contrast to purely statistically motivated regularization techniques), and present theoretical and simulation-based findings which demonstrate that models with different degrees of complexity converge as risk aversion increases.

Our empirical work represents a significant conceptual departure from linear parametric portfolio policies in two ways: first, by replacing the linear specification with a

neural network, we allow for nonlinearities and interactions in the relationship between firm characteristics and portfolio weights. Research on using machine learning for return prediction shows that such flexibility is relevant to model the relationship between firm characteristics and future returns and can lead to substantial improvements over less flexible specifications (Moritz and Zimmermann, 2016; Freyberger et al., 2020; Gu et al., 2020). It is conceivable that such flexibility will also help to model the relation between portfolio weights and firm characteristics. Second, this flexibility comes at the cost of having to estimate a model with a high-dimensional parameter vector. This is a deviation from the original motivation of the parametric portfolio policy literature which aims to reduce portfolio optimization to a low-dimensional problem with only a small number of coefficients that need to be estimated. Kelly et al. (2024) argue that model complexity is a virtue for return prediction, and our approach can be viewed as an exploration of that point in the context of parametric portfolio policies.

Our empirical investigation further underscores the value of the DPPP approach. Utilizing a comprehensive dataset of firm-level characteristics, we document substantial improvements in investor utility when using the DPPP relative to a standard linear model. Empirically, our complex model significantly improves over a standard linear parametric portfolio policy, with certainty equivalent gains ranging from about 43 basis points to 102 basis points. Further, in line with our theoretical results, we find that the benefit of model complexity decreases when an investor's risk aversion increases. While our benchmark investor is a classical constant relative risk aversion (CRRA) optimizer, our setup easily accommodates other utility functions. We explore portfolio policies with and without transaction costs and short-selling constraints, as well as for different utility functions, such as mean–variance and loss-aversion preferences, and a statistically motivated utility function. We find that economic regularization via risk aversion has similar dampening effects on higher-order portfolio moments as statistical regularization but economic regularization targets precisely the investor's objective function and thereby the risks most relevant to their investment decisions. Overall, complex portfolio policies can be beneficial in all these scenarios, but utility gains are higher for lower risk (or loss) aversion.

Beyond the aggregate performance improvements, our study offers novel insights into the relative importance of different types of firm characteristics. Past return-based stock characteristics turn out to be more important to the portfolio policy than accounting-

based characteristics. However, while prior research has highlighted the dominance of return-based signals in asset pricing, our results indicate that the inclusion of a large set of predictors with both return-based and accounting-based measures leads to a more balanced importance profile as risk aversion increases, extending the existing literature that examines the importance of firm characteristics under economic constraints (DeMiguel et al., 2020; Jensen et al., 2022).

Overall, our work bridges the gap between traditional portfolio optimization and modern machine learning techniques. By directly mapping firm characteristics to portfolio weights through neural networks, we offer a flexible, robust, and economically intuitive framework that adapts to the complexities of real-world investment challenges. This novel approach not only improves upon classical methods in terms of performance but also advances our theoretical understanding of how investor preferences can naturally regulate model complexity in high-dimensional settings.

Related literature

Our work relates to several different streams of the literature. First, we add to a growing literature that explores the potential of machine learning algorithms in finance (e.g., Heaton et al., 2017; Gu et al., 2020; Bianchi et al., 2020; Kelly et al., 2024). Studies in this literature typically consider a prediction task (e.g., predicting stock returns), and minimize a statistical loss function such as the mean squared error (or a related distance metric) between the actual and predicted values. Predicted values are used to construct portfolio weights (e.g., Gu et al., 2020). In contrast, our methodology uses machine learning to provide a direct, one-step mapping from firm characteristics to portfolio weights that explicitly targets the investor's objective rather than a statistical loss function.

Second, our paper serves as a natural extension of the parametric portfolio approach by Brandt et al. (2009). While Brandt et al. (2009) argue that it may be worthwhile to consider nonlinear functions and interactions in weight modeling, subsequent papers that have implemented and extended parametric portfolio policies parameterize portfolio weights as a linear function of firm characteristics (e.g., Hjalmarsson and Manchev, 2012; Ammann et al., 2016). DeMiguel et al. (2020) incorporate transaction costs, a larger set of firm characteristics, and statistical regularization but stay within the linear framework. Our DPPP replaces the linear model with a feed-forward neural network that accounts for both nonlinearity and possible interactions of firm characteristics. In addition, we

use a larger set of firm characteristics than previous studies and explore different utility functions, constraints, and degrees of risk aversion.

Further, we contribute to the literature that employs alternative methods to direct portfolio optimization via machine learning. Particularly relevant approaches in our context include Cong et al. (2021), Chevalier et al. (2022), Jensen et al. (2022), Guijarro-Ordonez et al. (2022), Coulombe and Goebel (2024), and Feng et al. (2024). Each of these differs from ours in one or more aspects. Cong et al. (2021) propose a reinforcement learning-based approach (as opposed to our feed-forward framework) and connect to a related literature in computer sciences that puts additional emphasis on more technical parts of the model implementation. Our study naturally connects to the preceding finance literature, and generalizes the approach of Brandt et al. (2009) to explicitly analyze differences between a linear and nonlinear specification for different utility functions, constraints, and levels of risk aversion, and we derive theoretical results for the convergence of these specifications under economic regularization. Chevalier et al. (2022) derive optimal in-sample weights based on investor preferences and subsequently predict these weights conditional on covariates. This is conceptually different from our approach, primarily because we do not require the preprocessing step of computing the optimal in-sample weights. Jensen et al. (2022) take a different approach. They specifically address the issue of integrating transaction costs into mean-variance portfolio optimization with machine learning. While their focus is the derivation of an efficient frontier including transaction costs, we explicitly analyze how different types of investor preferences and constraints affect the benefit of complexity in portfolio optimization. Guijarro-Ordonez et al. (2022) also employ neural networks for portfolio optimization. Their framework, however, is grounded in statistical arbitrage, whereas we directly map portfolio weights to stock-specific signals. Coulombe and Goebel (2024) propose a machine learning framework for directly optimizing portfolio weights with nonlinear algorithms, building on Lo and MacKinlay's (1997) maximally predictable portfolio approach. Their method aligns with mean-variance utility maximization. In contrast, our framework supports any utility function, offering broader flexibility beyond mean-variance preferences. Liu et al. (2024) propose an alternative one-step optimization approach, mapping predictors to optimal portfolio weights through genetic programming. Our methodology leverages feed-forward neural networks while incorporating a substantially larger set of stock characteristics and practical constraints such as transaction costs and leverage limits.

Feng et al. (2024) employ feed-forward neural networks to estimate portfolio weights by modeling a deep factor, i.e., a long-short factor based on a nonlinear combination of characteristics. They apply their method to bond data.

Our work also connects to the growing literature on machine learning approaches for estimating stochastic discount factors (SDFs). In this stream, Kozak et al. (2020) propose shrinking the cross-section of returns into a parsimonious set of factors that price all assets. Chen et al. (2024) use deep neural networks to construct a flexible, high-dimensional SDF, showing improved explanatory power for cross-sectional returns. Similarly, Bryzgalova et al. (2023) introduce a decision-tree–based method for constructing managed portfolios that span the SDF. Their approach focuses on identifying a parsimonious and interpretable set of characteristic-based portfolios that capture complex nonlinear interactions while remaining tractable and economically meaningful. In contrast, our paper sidesteps the explicit construction of test assets for SDF estimation and instead directly parameterizes and estimates the portfolio-weight function in a one-step framework.

In addition, our work relates to the literature that deals with estimation risk arising from parameter uncertainty (Kirby and Ostdiek, 2012a,b; Lassance et al., 2024) and the literature that explores regularization in this regard through economic mechanisms (Jagannathan and Ma, 2003; Skouras, 2007; Hautsch and Voigt, 2019). Adding to the literature, we explore how risk aversion affects model complexity and uncertainty, as well as how this affects the difference between models of different complexity.

Finally, our paper relates to research that explicitly analyzes how transaction costs and other forms of optimization constraints impact portfolio choice (DeMiguel et al., 2020; Jensen et al., 2022; Detzel et al., 2023). Complementing the literature, we study how nonlinearities contribute to the portfolio optimization, and how risk aversion regularizes optimization on top of and beyond the effects of transaction costs.

2.2 Theory

2.2.1 Expected utility framework and parametric portfolio policies

The starting point of our framework is the parametric portfolio policy model in Brandt et al. (2009). Consider a universe of N_t stocks that an investor can invest in at each month $t \in T$. Following Brandt et al. (2009) and to focus on the rich dynamics of risky asset

allocations, we do not include a risk-free asset.[1] Each stock i is associated with a vector of firm characteristics $x_{i,t}$ and a return $r_{i,t+1}$ from date t to $t+1$. The investor maximizes the conditional expected utility of future portfolio returns $r_{p,t+1}$:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t\left[u(r_{p,t+1})\right] = E_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right], \tag{2.1}$$

where $w_{i,t}$ is the weight of stock i in the portfolio at date t and $u(\cdot)$ denotes the respective utility function.

Instead of directly deriving the weights $w_{i,t}$ (as e.g., following the traditional Markowitz approach), we follow Brandt et al. (2009) and parameterize the weights as a function of firm characteristics $x_{i,t}$, i.e.,

$$w_{i,t} = f(x_{i,t}; \theta), \tag{2.2}$$

where θ is the coefficient vector to be estimated.

The parameter vector θ remains constant across assets i and periods t, i.e., it maximizes the conditional expected utility at every period t. This necessarily implies that θ also maximizes the unconditional expected utility. Hence, one can estimate θ by maximizing the unconditional expected utility via the return distribution's sample analogues:

$$\max_{\theta} \frac{1}{T}\sum_{t=1}^{T} u\left(r_{p,t+1}(\theta)\right) = \frac{1}{T}\sum_{t=1}^{T} u\left(\sum_{i=1}^{N_t} f(x_{i,t};\theta) r_{i,t+1}\right). \tag{2.3}$$

The idea behind parametric portfolio policies is that one may exploit firm characteristics in order to tilt some benchmark portfolio towards stocks that increase an investor's utility, so that $f(\cdot)$ can be expressed as

$$w_{i,t} = b_{i,t} + \frac{1}{N_t} g(x_{i,t}; \theta), \tag{2.4}$$

where $b_{i,t}$ denotes benchmark portfolio weights such as the equally weighted or value weighted portfolio and $x_{i,t}$ denotes the characteristics of stock i, standardized cross-sectionally to have zero mean and unit standard deviation in each cross-section t.[2]

[1] As noted by Brandt et al. (2009), the presence of a risk-free asset is not required for our setting, because it only adds a scalar leverage decision that is orthogonal to the cross-sectional allocation problem.

[2] The $1/N_t$ term is a normalization that allows the portfolio weight function to be applied to a time-varying number of stocks. Without this normalization, an increase in the number of stocks with an otherwise unchanged cross-sectional distribution of characteristics leads to more radical allocations, although the investment opportunities are basically unchanged.

In essence, our model can be interpreted as a generalization of the linear parametric portfolio policy approach, as we allow $x_{i,t}$ to enter the model flexibly. Brandt et al. (2009) and the subsequent literature (e.g., DeMiguel et al., 2020) restrict firm characteristics to affect the portfolio in a linear, additive manner. In contrast, we model $g(\cdot)$ in Equation (2.4) as a feed-forward neural network, arguably one of the most flexible forms. As discussed in the introduction, this represents a significant conceptual deviation from the literature in at least two respects: first, by replacing the linear specification with a neural network, we allow the relationship between firm characteristics and weights to be nonlinear, and we account for potential interactions of firm characteristics, in line with the recent literature that finds that such flexibility can be important to predict returns (Moritz and Zimmermann, 2016; Freyberger et al., 2020; Gu et al., 2020). Here, our approach explores whether such flexibility also helps to model the relationship between *portfolio weights* and firm characteristics. Second, this flexibility comes at the cost of having to estimate a model with a high-dimensional parameter vector. Thus, it departs from the original motivation of the parametric portfolio policy literature, which aimed to reduce portfolio optimization to a low-dimensional problem where only a small number of coefficients need to be estimated. In fact, our benchmark model below has about 5,700 to 5,900 parameters compared to the three parameters that need to be estimated when following Brandt et al. (2009).

2.2.2 Risk aversion as economic regularization

This section establishes the theoretical underpinnings for how risk aversion serves as an economic regularization mechanism in our setting. Our key insight is that risk aversion naturally constrains model complexity when estimation risk is a concern. Intuitively, estimation risk arises from the uncertainty about the parameters of the data generating process. This leads to errors in the estimation of portfolio weights which increases portfolio risk (Kirby and Ostdiek, 2012b; Lassance et al., 2024). As risk aversion increases, the investor places a greater penalty on portfolio return variance, leading to more conservative portfolios with simpler investment strategies.

Formally, when portfolio returns are evaluated under the predictive distribution, the law of total variance implies that estimation uncertainty about model parameters θ enters expected utility as an *additional* source of variance. Risk aversion penalizes not only intrinsic return risk but also the excess variance induced by parameter estimation. For

example, in the mean-variance case with one risky asset, Brandt (2010) shows that the certainty equivalent loss due to estimation uncertainty is proportional to the risk aversion coefficient.

This stands in contrast to previous approaches to regularization in portfolio choice. While Hautsch and Voigt (2019) and Jagannathan and Ma (2003) show that transaction costs and short-selling constraints can serve as economically motivated penalties, our framework demonstrates how investor preferences themselves create a natural regularization mechanism. In the following two subsections, we formalize this idea through two approaches: first, from an economic perspective, and then using results from statistical learning theory.

Economic intuition

To establish economic intuition, consider a CRRA investor maximizing expected utility over portfolio returns as in Equation (2.1). Following Brandt et al. (2009), we express portfolio returns as:

$$r_{p,t+1}(\theta) = b_t^T r_{t+1} + \theta^T X_t^T r_{t+1}/N_t = r_{b,t+1} + \theta^T r_{c,t+1}, \tag{2.5}$$

where $r_{b,t+1}$ is the benchmark return and $r_{c,t+1}$ contains characteristic portfolio returns. As outlined in Didisheim et al. (2023), we can interpret this linear portfolio framework through a neural network lens by replacing our characteristic-sorted portfolios $X_t^T r_{t+1}/N_t$ with transformed portfolios $S_t^T r_{t+1}/N_t$, where S_t represents nonlinear transformations of the original characteristics.

The following proposition shows how the active deviations from the benchmark defined in Equation (2.4) critically depend on risk aversion:

Proposition 1 *Define the optimal active deviations from the benchmarks:*

$$\frac{1}{N_t}\theta^{*T} S_t = \underbrace{\frac{1}{\gamma}\hat{\Sigma}_c^{-1}\hat{\mu}_c^T S_t}_{\text{Risk premium term}} - \underbrace{\hat{\Sigma}_c^{-1}\hat{\sigma}_{bc}^T S_t}_{\text{Risk minimization term}}. \tag{2.6}$$

The optimal deviations from the benchmark portfolio converge as follows:

$$\lim_{\gamma\to\infty} \frac{1}{N_t}\theta^T S_t = -\hat{\Sigma}_c^{-1}\hat{\sigma}_{bc}^T S_t. \tag{2.7}$$

The proof is in Appendix A.1.1.

In particular, if the benchmark is uncorrelated with the active positions (i.e., $\hat{\sigma}_{bc} = 0$), then the active positions converge to zero. If the benchmark portfolio is correlated with active positions, i.e., $\hat{\sigma}_{bc} \neq 0$, then as $\gamma \to \infty$ the optimal active tilt converges to the benchmark-hedge $-\hat{\Sigma}_c^{-1}\hat{\sigma}_{bc}$ rather than zero. Economically, higher γ shuts down premium-seeking tilts but preserves positions that hedge benchmark risk. Only under the special case $\hat{\sigma}_{bc} = 0$ do all active tilts vanish in the limit.

High risk aversion forces both the PPP and the DPPP to prioritize risk minimization, leading to convergence for the mean absolute weight differences between the models. We summarize this in the following proposition, which follows directly and trivially from Proposition 1:

Proposition 2 *Compare two models of different complexity, with corresponding portfolio weights $w_{PPP}(\gamma)$ and $w_{DPPP}(\gamma)$. Let Δ_{RP} and Δ_{RM} be the differences in the risk premium and risk minimization terms, respectively, from Equation* (2.7)*. The absolute portfolio weight difference can be decomposed as*

$$\lim_{\gamma\to\infty} \|w_{PPP}(\gamma) - w_{DPPP}(\gamma)\| = \underbrace{\lim_{\gamma\to\infty} \frac{1}{\gamma}\|\Delta_{RP}\|}_{=0} + \lim_{\gamma\to\infty} \|\Delta_{RM}\|, \tag{2.8}$$

where Δ_{RP} and Δ_{RM} are deterministic quantities independent of γ that depend only on the problem structure (X_t, b_t, and r_{t+1}). If both models build active portfolios orthogonal to the benchmark, then the gap between the PPP and DPPP weights converges to zero, otherwise a constant hedge difference can remain.

In Section A.3 in the Appendix we provide a similar intuition for loss aversion preference.

This convergence in weights is consistent with the classical separation theorem. In our mean–variance setting, all investors share the same risky tangency portfolio and risk aversion only scales overall exposure. As risk aversion increases, the premium-seeking component vanishes and the DPPP weights contract toward those of the PPP, leading to the convergence established in Proposition 2. In this sense, the convergence in weights corresponds to investors moving toward a common region of the efficient set where model flexibility has little incremental effect.

Complexity interpretation

The concept of risk aversion as a regularization mechanism is also grounded in statistical learning theory. Following Skouras (2007), estimation uncertainty directly affects economic decisions through the utility function. This leads to a natural complexity measure known as the *effective degrees of freedom* (EDF), originally developed by Murata et al. (1994) which is defined as:

$$\text{EDF} = \text{tr}(G^{-1}V)/T, \tag{2.9}$$

where G is the hessian of the model with respect to the parameters and V is the outer product of the gradients of the model with respect to the parameters.

This measure has an intuitive interpretation: for linear models like the PPP, the trace reduces to the parameter count p, making EDF interpretable as the number of "effective" parameters in more complex models. This leads to our key result about model complexity

Proposition 3 *As risk aversion increases, the effective model complexity converges to zero*

$$\lim_{\gamma\to\infty} EDF = \lim_{\gamma\to\infty} \frac{p}{\gamma} = 0. \tag{2.10}$$

The proof is in Appendix A.1.2.

This result shows that increasing risk aversion effectively reduces model complexity, providing a theoretical link between risk preferences and model complexity in our portfolio framework. The complexity interpretation provides additional insight into the convergence between the PPP and the DPPP established in the previous section. As $\gamma \to \infty$, the EDF approaches zero for both models, meaning they effectively become less complex regardless of their nominal parameter count. For the PPP, the number of effective parameters p/γ goes to zero. Similarly for the DPPP, despite having a richer nonlinear structure through S_t its effective complexity also converges to zero as risk aversion increases. In essence, high risk aversion forces both models to prioritize risk minimization over exploiting their different parametric structures, leading to their convergence.

Naturally, one might wonder how these theoretical results relate to the trade-off between bias (less complex models might be misspecified) and variance (less complex models might have lower estimation variance) in the traditional statistical sense. While traditional bias-variance analysis focuses on minimizing prediction error, we focus on

maximizing expected utility from decisions, i.e., models are evaluated based on their utility for decision-making rather than statistical criteria. Skouras (2007) shows that when models are potentially misspecified, the standard bias-variance tradeoff is not the most relevant consideration for decision-makers. Instead, models should balance complexity against performance. Expected utility reflects both traditional statistical channels: higher risk aversion amplifies the penalty on estimation variance and thus constrains functional forms. Conversely, low risk aversion allows the model to exploit richer functional forms. Preferences thus endogenously determine the optimal degree of complexity from a decision-theoretic rather than purely statistical perspective. The fact that regularization/shrinkage can lead to improvements in non-statistical objective functions is not new to academic finance (see e.g., Jorion (1986)). In Section 2.4.2, we compare model regularization via risk aversion to statistical model regularization, and find that both have dampening effects on portfolio return volatility.

Simulation evidence

We illustrate our theoretical results through a simulation study featuring two nested parametric portfolio policies that share the same base information set but differ in complexity. We generate a panel of $N = 100$ firms over $T = 200$ months with $K = 10$ base firm characteristics that follow persistent AR(1) processes $x_{i,k,t} = \rho x_{i,k,t-1} + \epsilon_{i,k,t}$, where $\rho = 0.8$ captures the empirically observed persistence in firm characteristics, and $\epsilon_{i,k,t} \sim N(0, 1-\rho^2)$. All characteristics are standardized cross-sectionally. Returns are generated with predictability based on the base characteristics $r_{i,t+1} = x_{i,t}^\top \beta + \eta_{i,t+1}$, where $\beta \sim N(0, 0.1^2 I_K)$ and $\eta_{i,t+1} \sim N(0, 0.15^2)$. Lastly, both models use an equally-weighted portfolio as a benchmark.

Following our theoretical framework with S_t representing nonlinear transformations of the original characteristics, we expand the feature space with random Fourier features. Specifically, we draw random vectors $w^j \overset{\text{iid}}{\sim} N(0, \eta^2 I)$ for $j = 1, ..., p/2$. For each j, we create a pair of new features using sine and cosine transformations, where we generate $p = 10$ features for the simple model (PPP) and $p = 100$ features for the complex model (DPPP). This approach ensures both models operate on transformations of the same

underlying information while differing substantially in their parametric complexity.[3]

Figure 2.1 presents the key findings from estimating both models across a grid of risk aversion values $\gamma \in [1, 100]$. We examine two metrics that directly correspond to our theoretical results: (i) the mean absolute difference in portfolio weights, and (ii) the effective degrees of freedom of each model. Consistent with our theoretical predictions, we find that the weight difference between models decreases and the EDF of both models converge as risk aversion increases, effectively constraining model complexity through the investor's utility function rather than via statistical penalties. Appendix A.3 shows similar results for investors with loss-aversion utility.

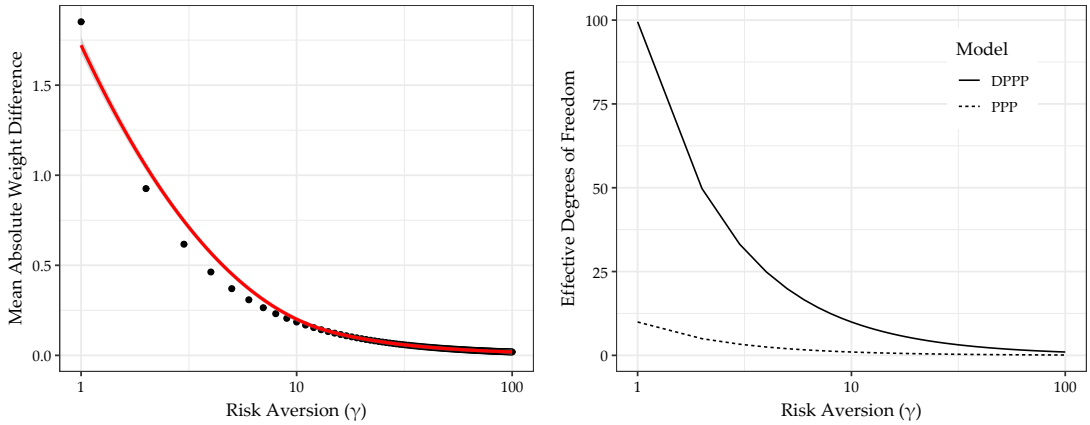


Figure 2.1: Risk aversion as economic regularization
This figure presents simulation evidence demonstrating that risk aversion acts as an economic regularization mechanism. We compare two nested parametric portfolio policies of different complexity: one using 10 characteristics (PPP) and a second one using 100 characteristics (DPPP) constructed through random Fourier transformations of the base characteristics. The left panel shows the mean absolute difference in portfolio weights between models across risk aversion levels. The right panel plots the effective degrees of freedom (EDF) for both models, demonstrating how increasing risk aversion reduces model complexity. All panels use a logarithmic scale (base 10) for risk aversion.

2.3 Estimation and results

2.3.1 Network architecture

We model function $g(\cdot)$ in Equation (2.4) as a feed-forward network. Conceptually, our feed-forward networks are structured to estimate optimal portfolio weights and as such differ from networks used in pure prediction contexts in two important ways.

[3]While return prediction using random Fourier features has become popular following Kelly et al. (2024), Nagel (2025) shows that predictions from such (ridgeless) regressions become a weighted average of past returns in the training window, with weights being functions of simple Gaussian kernels. Our results are invariant to the transformation function and therefore also hold for using simple polynomial features.

First, the objective of our estimation is to maximize expected utility. Standard use of predictive modeling for stock returns (with or without networks) tries to minimize some distance metric (e.g., the mean squared error) between observed and predicted stock returns. For example, Gu et al. (2020) use neural networks to predict stock returns using a penalized mean squared error as the statistical loss function. In contrast, we follow Brandt et al. (2009) and directly estimate portfolio weights. More specifically, we predict portfolio weights by maximizing the unconditional sample analogue of a utility function as given in Equation (2.3). For example, in our base case, the loss function $\mathcal{L}$ that we aim to minimize with respect to θ is the constant relative risk aversion (CRRA) utility:

$$\mathcal{L}(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\left(\frac{(1+r_{p,t+1}(\theta))^{1-\gamma}}{1-\gamma}\right), \tag{2.11}$$

where γ is the relative risk aversion parameter. Note that minimizing Equation (2.11) is equivalent to maximizing CRRA utility.

Second, unlike applications that predict stock-level returns using neural nets, our estimated stock-level portfolio weights are only intermediate outputs of the neural network in that the loss function is based on the *portfolio* return. Hence, we need to aggregate intermediate network outputs (stock-level weights in period t; that are a function of stock-level characteristics in period t) and stock-level returns in period $t+1$ cross-sectionally (see Equations (2.2) - (2.4)).

To operationalize this, we maintain the three-dimensional structure of our data (time / stocks / characteristics) where the three-dimensional input tensor reflects the panel structure of the data. Still, portfolio weights at time t are determined by that period's stock characteristics, maintaining the original spirit of Brandt et al.'s (2009) approach while leveraging the additional flexibility of neural networks to capture cross-sectional nonlinearities. In other words, unlike time series models (such as Recurrent Neural Networks or Long Short-term Memory Networks) that explicitly model sequential dependencies, our network makes independent decisions at each time step based on the concurrent cross-sectional relationships between characteristics and expected returns. This is by design, as our goal is to identify robust cross-sectional patterns rather than temporal dependencies.

Conceptually, our models can be depicted as shown in Figure 2.2. The input data on the left form a cube (or 3D tensor, the three-dimensional structure described above)

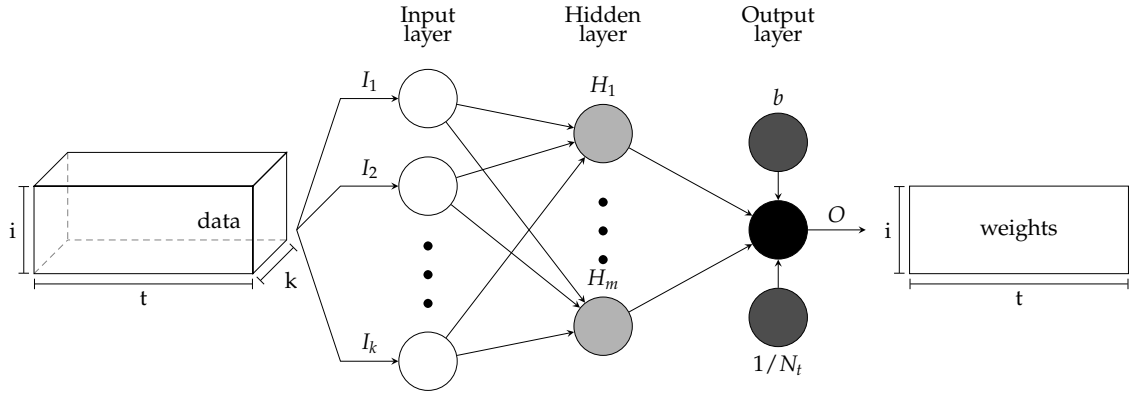


Figure 2.2: Neural network structure
This figure presents the core structure of our neural networks. White circles denote the input layer, grey circles denote the hidden layer and black circles denote the output layer. The data cube on the left depicts the structure of our data, i.e., we have k variables across i cross-sections in t periods. The rectangle on the right depicts our output, i.e., weights across i cross-sections in t periods. The output of the neural network is normalized by $1/N_t$ and added to the benchmark portfolio b. The final output is labeled O.

with dimensions time t, stocks i and input variables k. Input data is fed into networks with different numbers of hidden layers. In line with Equation (2.4), the output of the neural network is then normalized by $1/N_t$ and added to the benchmark portfolio b. The output of the model O is a two-dimensional matrix with dimensions $t \times i$ of portfolio weights for each stock and time period that is then aggregated (as a weighted sum of period $t+1$ stock returns) across all stocks in each time period into a portfolio return that is the input of the loss function in Equation (2.1).

Constructing a neural network requires many design choices, including, e.g., the depth (number of layers) and width (units per layer) of the model, or the activation function for different units and layers,[4] and selecting the optimal network architecture is a challenging task. We simplify the process by tuning the number of hidden layers only, evaluating configurations with three, four, and five layers. In every configuration, the first hidden layer starts with 32 nodes, and each subsequent hidden layer contains half as many nodes as the preceding layer.

As discussed in Section 2.2.1, the network's output needs to be normalized and can then be interpreted as the deviation from a benchmark portfolio. In our application, the benchmark portfolio is the equal-weighted portfolio in all models. A common alternative

[4]The activation function introduces nonlinearity into the model by applying a transformation that isn't simply a straight line. We use the leaky rectified linear unit (ReLU) as activation function throughout all layers to prevent the issue of "dying ReLU", see Appendix A.2. The leaky ReLU is a piecewise linear function: it behaves like a regular ReLU for positive inputs but applies a small, non-zero slope to negative inputs instead of completely zeroing them out. Because of this change in slope, the overall function is not purely linear, which lets the network capture more complex, nonlinear relationships in the data.

would be a value-weighted benchmark portfolio where weights are determined by a stock's market capitalization. We stick to the equal-weighted benchmark because of empirical evidence that it outperforms other benchmarks for longer periods (DeMiguel et al., 2009).

Lastly, we impose two constraints to ensure that the model's performance stems from diversified positions with economically reasonable leverage levels rather than from concentrated bets or excessive leverage. First, we impose an ex-ante upper bound on an individual stock's absolute portfolio weight of $|3\%|$, i.e.,

$$|w_{i,t}| \leq 0.03, \tag{2.12}$$

where $w_{i,t}$ represents the portfolio weight of stock i at time t. Second, we limit leverage to 100% of the portfolio value in any single period during model training.[5] This constraint is formulated for every period t as

$$\sum_{i=1}^{N_t} w_i I(w_i < 0) \geq -1, \tag{2.13}$$

where $I(w_i < 0)$ is an indicator function that equals one if the corresponding portfolio weight is negative and zero otherwise.

Additionally, we maintain the full investment constraint. Due to the nonlinear nature of the model, it is not obvious that the full investment constraint holds (unlike in Brandt et al. (2009)). To make sure that the full investment constraint is satisfied, we standardize the outputs of each unit in the hidden layers cross-sectionally to have zero mean and unit standard deviation across all stocks at date t. Hence, the output of each node in each hidden layer can be interpreted as a deviation from the benchmark portfolio (see Appendix A.2 for details).

We also employ a range of different additional regularization techniques that are standard in the deep learning literature. We give an outline of these techniques and a more detailed description of the structure of the model including its hyperparameters in Appendix A.2.

To estimate our model, we use an expanding-window strategy with a rolling 12-month

[5] Ang et al. (2011) show that the average gross leverage of hedge fund companies amounts to 120% in the period after the financial crisis 2007-2008. We use a slightly more conservative number of a maximum leverage of 100%.

out-of-sample period (details in Appendix A.2).[6] Specifically, we train on the first 20 years, validate on the next 5 years, and then test on the next 12 months. We then roll forward one year at a time, continually re-estimating.

2.3.2 Data

We use the Open Source Asset Pricing dataset of Chen and Zimmermann (2022) for the period from January 1971 to December 2020, as comprehensive accounting data is only sparsely available in prior years. In addition, we only keep common stocks, i.e., stocks with share codes 10 and 11, and stocks that are traded on the NYSE (exchange code equal to 1) to ensure that results are not driven by small or illiquid stocks. We match the data with monthly stock return data from the Center for Research in Security Prices (CRSP). We drop any observation with missing return, size and/or a return of less than -100%. We include continuous firm characteristics from Chen and Zimmermann (2022)'s categories *Price, Trading, Accounting* and *Analyst*, respectively.

Finally, we follow Gu et al. (2020) and replace missing values with the cross-sectional median at each month for each stock, respectively.[7] Additionally, similar to Gu et al. (2020) we rank all stock characteristics cross-sectionally. As in Brandt et al. (2009) and DeMiguel et al. (2020), each predictor is then standardized to have a cross-sectional mean of zero and standard deviation of one. Note that each predictor is signed so that a larger value implies a higher expected return in the original in-sample period.

Our final dataset contains 157 predictors for a total of 5,154 firms. Each month, the dataset contains a minimum of 1,213, a maximum of 1,855 and an average of 1,422 firms. These numbers are consistent with a sample of established, liquid companies rather than the broader universe including small and illiquid stocks. Table A.5 in the Appendix lists the included predictors by original paper. The three columns in the table describe the update frequency of each predictor, the predictor category and the economic category, both taken from Chen and Zimmermann (2022).

[6] We also experimented with a single split of the data into an estimation and a test period but results are significantly worse. This suggests that the relationship between stock weights and characteristics varies over time. Hence, more frequent coefficient updates (either via expanding- or rolling-window strategies) are crucial to achieve promising results.

[7] Chen and McCoy (2024) show that simple median imputation of missing values outperforms more sophisticated methods in the context of machine learning portfolio formation. In fact, they explicitly recommend applying simple median imputation in this context. They argue that there is little to be gained from other methods (and that such methods might even introduce estimation noise) that try to exploit the cross-sectional or time-series structure because a. missingness occurs in blocks and b. non-missing predictors display low cross-sectional correlations.

2.3.3 Performance results for CRRA investors

Table 2.1 compares the results of the optimization process for CRRA investors with different degrees of risk aversion for our DPPP with its linear counterpart.[8] Results reveal substantial economic gains from employing deep learning in portfolio optimization across different levels of risk aversion. The DPPP consistently outperforms the PPP, with the magnitude and statistical significance of improvements varying systematically with investor risk preferences.

The economic significance of the deep learning approach is most evident in the certainty equivalent returns (CE).[9] At a risk aversion level of $\gamma = 2$, the DPPP achieves a CE of 2.97% compared to the PPP's 1.95%, representing a significant enhancement of 102 basis points (p-value = 0.0003).[10] This substantial improvement suggests that capturing nonlinear relationships and interactions between predictors creates meaningful economic value for investors.

Notably, the performance differential between the DPPP and the PPP exhibits a decline with increasing risk aversion. The difference in monthly certainty equivalent narrows to 43-69 basis points at higher levels of risk aversion, with statistical significance declining correspondingly (p-values increase from 0.0003 at $\gamma = 2$ to 0.0278 at $\gamma = 20$). Intriguingly, we also find that the mean absolute weight differences between the PPP and DPPP models decrease with risk aversion, consistent with our analytical results and simulations in Section 2.2.2. Empirically, different from the simulation, constraints and modeling choices influence the composition of risky assets across risk aversion. Yet Figure 2.3 mirrors our simulation results in Figure 2.1 and shows a steady decline of weight differences for the models of Table 2.1 but also for other model specifications considered in the next sections, suggesting that risk aversion serves as an effective economic regularization mechanism.

The portfolio characteristics in Table 2.1 provide further insights into the sources of outperformance. The DPPP exhibits more aggressive position-taking, as evidenced by

[8] To ensure comparability between the linear and the deep parametric portfolio policy we differ slightly from Brandt et al. (2009) in that the linear model includes l_1-regularization and early stopping, similar to the deep model. A more detailed description is given in Appendix A.2.

[9] The certainty equivalent return is the guaranteed monthly return an investor would require to achieve the same expected utility as via following the corresponding estimated portfolio policy.

[10] We follow DeMiguel et al. (2024) and construct one-sided p-values from 10,000 bootstrap samples using the stationary bootstrap method of Politis and Romano (1994) with an average block size of five and the procedure of Ledoit and Wolf (2008). This method is also used when assessing the statistical significance of Sharpe ratio differences between the deep and the linear parametric portfolio policy hereafter.

Table 2.1: Deep portfolio policy for CRRA investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0195	0.0297	0.0163	0.0232	0.0109	0.0152	-0.0006	0.0040
p-value$(CE_{DPPP} - CE_{PPP})$		0.0003		0.0002		0.0338		0.0278
$\sum \lvert w_i \rvert / N_t * 100$	0.1696	0.1907	0.1770	0.1938	0.1769	0.1933	0.1690	0.1729
$max\ w_i * 100$	0.6815	1.1483	0.7221	0.9843	0.7087	0.8305	0.6710	0.4582
$min\ w_i * 100$	-0.6581	-1.2824	-0.6953	-1.2053	-0.6981	-0.9743	-0.6322	-0.7224
$\sum w_i I(w_i < 0)$	-0.7228	-0.8748	-0.7762	-0.8974	-0.7754	-0.8932	-0.7180	-0.7464
$\sum I(w_i < 0)/N_t$	0.3426	0.3400	0.3498	0.3368	0.3490	0.3319	0.3426	0.3202
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	1.3426	2.6342	1.4571	2.6022	1.4224	2.3813	1.2204	1.7516
Mean	0.0220	0.0341	0.0214	0.0305	0.0201	0.0281	0.0179	0.0224
StdDev	0.0492	0.0710	0.0435	0.0550	0.0401	0.0475	0.0372	0.0378
Skew	-0.5991	2.6646	-0.8212	0.8411	-0.8161	-0.2470	-0.7878	-0.5201
Kurt	2.8950	26.4755	2.5283	10.9695	2.1622	4.0705	1.9090	1.9954
Max DD	0.6302	0.4979	0.4467	0.5601	0.3953	0.4662	0.3803	0.3027
Max 1M loss	0.2140	0.2264	0.1855	0.1789	0.1489	0.1838	0.1369	0.1446
CVaR (95%)	0.1044	0.1107	0.0938	0.0978	0.0881	0.0882	0.0803	0.0713
SR	1.5465	1.6607	1.7007	1.9230	1.7404	2.0446	1.6635	2.0491
p-value$(SR_{DPPP} - SR_{PPP})$		0.3709		0.0985		0.0445		0.0042
$FF5 + Mom\ \alpha$	0.0104	0.0232	0.0103	0.0205	0.0097	0.0182	0.0085	0.0130
$StdErr(\alpha)$	0.0012	0.0029	0.0013	0.0024	0.0014	0.0020	0.0014	0.0016

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

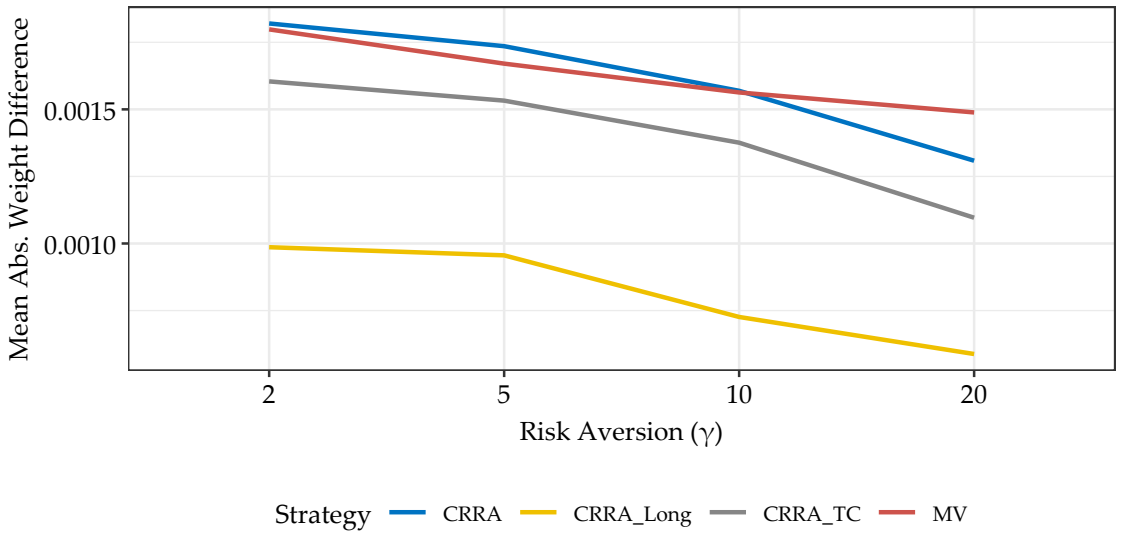


Figure 2.3: Mean absolute portfolio weight differences by risk aversion
The plot shows mean absolute portfolio weight differences between the PPP and DPPP models. Different lines refer to different model specifications, with CRRA (the blue line) the models in Table 2.1, CRRA_TC (grey line) the models in Table 2.3, CRRA_Long (yellow line) the models in Table 2.3 and MV (red line) the models in Table 2.4.

the maximum portfolio weights (1.15% versus 0.68% at $\gamma = 2$) and minimum weights (-1.28% versus -0.66% at $\gamma = 2$), more concentrated portfolios (average absolute weights of 0.19% versus 0.17% at $\gamma = 2$) and higher turnover than the linear portfolio policy (263% versus 134% at $\gamma = 2$).[11] All these differences decline with risk aversion, reflecting the regularizing effect of risk preferences.

Risk metrics offer insights into downside protection, a crucial consideration for many investors that may not be fully accounted for by variance-based measures or standard utility functions. Despite its more aggressive positioning, the DPPP achieves risk control comparable to the PPP with lower maximum drawdowns (49.79% versus 63.02% at $\gamma = 2$), maximum one-month losses (22.64% versus 21.40% at $\gamma = 2$) or Conditional Value at Risk (CVaR) (11.07% versus 10.44% at $\gamma = 2$).[12] Qualitatively similar results hold for all values of risk aversion, demonstrating that the DPPP approach can yield utility benefits without sacrificing practically relevant performance dimensions, even when those dimensions are not explicitly targeted in the optimization.

The improvements in risk-adjusted performance are substantial. Sharpe ratios are consistently higher for the DPPP across all risk aversion levels (1.66 versus 1.52 at $\gamma = 2$), with the differences becoming statistically significant at higher risk aversion levels (p-

[11] See Section 2.3.6 for a formal definition of turnover.
[12] Following Gu et al. (2020), we calculate maximum drawdown based on cumulative log returns.

value = 0.0042 at $\gamma = 20$). The outperformance is robust to controlling for standard risk factors, as evidenced by significant monthly alphas against the Fama-French five-factor model augmented with momentum (2.32% versus 1.04% at $\gamma = 2$, with standard errors of 0.29% and 0.12% respectively).

The main results from Table 2.1 are visually summarized in Figure 2.4, which shows the cumulative performance of portfolio returns over time for both the PPP and DPPP across different degrees of risk aversion. The figure demonstrates several key patterns. First, the DPPP (solid lines) consistently outperforms its linear counterpart (dotted lines) across all risk aversion levels, with the outperformance becoming more pronounced after the 2008 financial crisis. Second, lower risk aversion portfolios ($\gamma = 2$, blue line) achieve higher cumulative returns but exhibit more volatility during periods of market stress. For instance, during the dot-com bubble burst (2000-2002), the global financial crisis (2008-2009), and the COVID-19 market crash (2020), the $\gamma = 2$ portfolio experiences larger drawdowns compared to higher risk aversion portfolios.

Notably, the outperformance of the DPPP over the PPP persists across market environments, though the magnitude varies. The gap between the DPPP and the PPP tends to widen during strong market periods (e.g., 2003-2007 and 2009-2020) and narrows during market stress, suggesting that the benefits of nonlinear modeling are particularly valuable in capturing upside potential while still providing some downside protection.

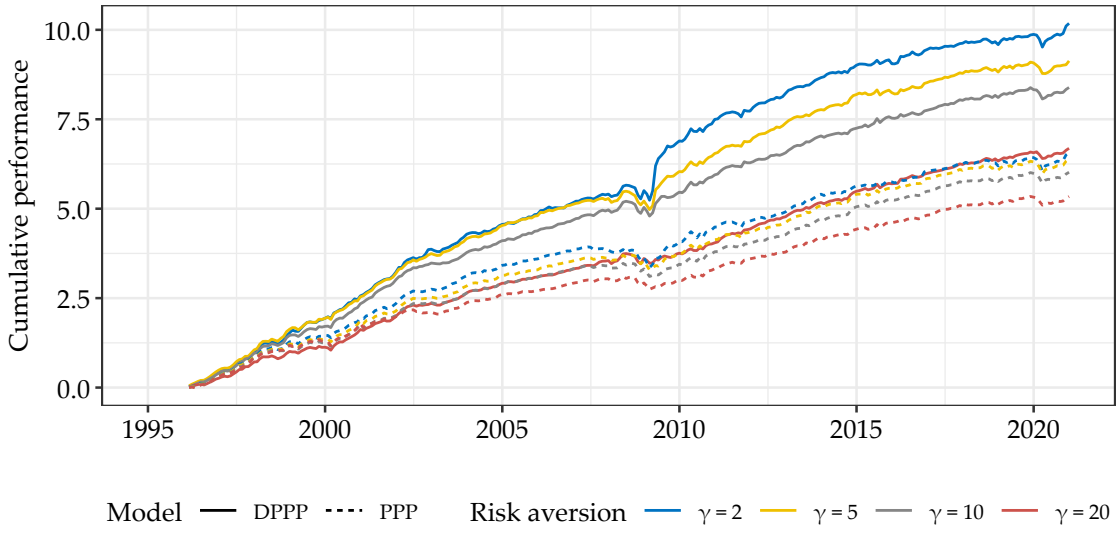


Figure 2.4: Cumulative performance over time for CRRA preferences
The plot shows the cumulative sum of portfolio returns for the DPPP and PPP. We show the results for each of the degrees of relative risk aversion considered and across all out-of-sample periods.

2.3.4 Supporting results

Variable importance

We calculate the importance of the variables in the model as the mean absolute gradient of the model with respect to the input features. That is, for each period, we calculate the gradient of the investor's utility with respect to an input feature, take the absolute value of each value, and then take the average over all values. We repeat this for each feature in every out-of-sample period and take the average across all models. For the sake of comparability, we scale the average utility losses across all variables for each model so that they add up to one. As a result, we are able to rank the variables according to the average absolute gradient.

Figure 2.5 displays the relative importance of the 40 most influential characteristics across different risk aversion levels for both models, measured using absolute average gradients. The variables are ordered according to the importance of the DPPP model optimized for $\gamma = 2$.

Several key patterns emerge from this analysis. First, past return-based characteristics dominate the importance rankings across all specifications, with short-term reversal (STreversal), industry returns of big firms (IndRetBig), and momentum seasonality (MomSeason) consistently appearing among the most influential features, mirroring the findings in Moritz and Zimmermann (2016), Gu et al. (2020) and Chen et al. (2024) for prediction of returns rather than utility maximization. This finding holds for both the DPPP and the PPP, though the relative magnitudes differ substantially.

Second, the DPPP shows more pronounced differentiation in feature importance compared to the PPP, particularly at lower risk aversion levels. For instance, with $\gamma = 2$, short-term reversal exhibits approximately twice the importance in the DPPP compared to the PPP. This suggests that the nonlinear model is better at capturing and exploiting the dynamic nature of return reversal patterns. Importantly, the pattern of feature importance varies systematically with risk aversion. As risk aversion increases, our analysis reveals a more balanced importance across characteristics, particularly in the DPPP, consistent with the results of DeMiguel et al. (2020). Therefore, model flexibility matters precisely when investors' preferences allow them to bear estimation risk in pursuit of complex return premia.

Third, the analysis also reveals interesting differences in how the two models utilize

similar information. While both models draw heavily on momentum-related signals (MomSeason, IntMom, High52), the DPPP appears to extract more nuanced information, as evidenced by the higher importance weights on various momentum components (seasonal, intermediate, and price-based momentum). Notably, characteristics related to fundamental firm information (earnings, analyst forecasts, and balance sheet measures) show relatively stable importance across risk aversion levels, particularly in the DPPP. This suggests that these features provide complementary information that remains valuable even as the portfolio becomes more conservative.

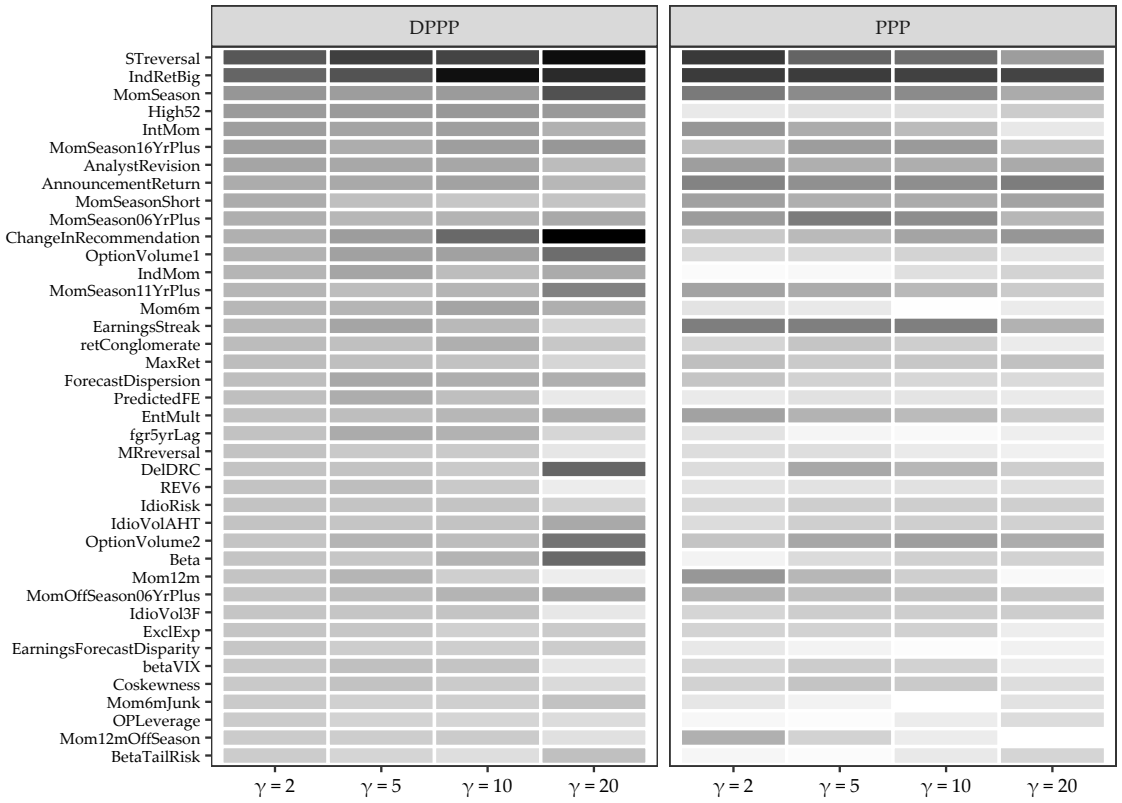


Figure 2.5: Variable importance for the CRRA preference for the DPPP and the PPP
Variable importance for the 40 most influential variables in the PPP and DPPP across model specifications and risk aversions, respectively. Variable importance is computed as the average absolute gradient over all training samples and normalized to sum to one within each model. The darker the color gradient, the higher the respective importance. The variables are ordered according to the importance of the DPPP model optimized for $\gamma = 2$.

In Appendix A.4.1, we examine the marginal contribution (partial dependence) of characteristics to portfolio weights in the DPPP and we find that nonlinear modeling is beneficial for capturing the complex relationship between firm characteristics and portfolio allocations. Key variables such as short-term reversal, book-to-market, and various momentum measures exert nonlinear and risk-aversion–dependent effects on

portfolio weights. For instance, short-term reversal shows a strong, varying impact across its range, particularly under low risk aversion, which aligns well with the risk-return profiles observed in decile portfolios. In contrast, other characteristics exhibit more subdued or context-specific influences, with higher risk aversion generally dampening these effects.

To analyze further the extent to which nonlinearity plays a role, we fit linear surrogate models to explain the portfolio weights of the DPPP. The findings indicate that 30–60% of the characteristic–weight relationship is linear, an additional 20–30% is explained by interactions, and the remaining 10–50% is due to higher-order nonlinearities. Moreover, the economic significance of these nonlinear components, measured via certainty equivalent differences, is most pronounced for lower risk aversion levels, underscoring that the flexibility of nonlinear models adds substantial value. It also provides empirical support for our understanding that risk aversion acts as an economic regularization parameter. See Appendix A.4.1 for detailed results.

Comparison to benchmark models

What do investors gain by optimizing their utility function directly rather than optimizing standard benchmarks such as, e.g., the Sharpe ratio? Table 2.2 comprehensively evaluates DPPP portfolios against several important benchmarks, including machine learning-based return forecasting, Sharpe ratio optimization, a factor portfolio, and traditional passive strategies.

Results indicate that involved methods such as directly optimizing the Sharpe ratio, or a two-step strategy that uses machine learning to forecast returns and subsequently builds portfolios based on those forecasts, have similar overall portfolio characteristics in terms of leverage, turnover or maximal one-month losses to the DPPP strategies, especially those DPPP strategies optimized for higher levels of risk aversion. DPPP strategies generally deliver higher out-of-sample alphas against involved benchmarks and also passive benchmarks such as equal- or value-weighted index portfolios. Consequently, DPPP strategies also deliver higher cumulative returns than the benchmark strategies (see Figure A.8 in the Appendix).

Interestingly, the DPPP strategies deliver higher out-of-sample Sharpe ratios than even the strategies that target maximization of the Sharpe ratios directly. In Table A.4 in the Appendix, we check whether the reverse (i.e., whether benchmark strategies deliver

higher out-of-sample utility than strategies that target utility maximization directly) is also true, and find that this is not the case: strategies optimized for a specific level of risk aversion deliver higher out-of-sample utility than benchmark models or strategies optimized for other levels of risk aversion. This outperformance holds across different risk aversion levels. The relative advantage becomes increasingly pronounced as risk aversion increases. This is particularly evident when compared to strategies that do not explicitly incorporate portfolio risk in the optimization objective, underscoring the critical importance of integrating risk considerations directly into the optimization process.

Figure A.9 in the Appendix shows variable importance rankings for the involved benchmark models. We find that, similar to the DPPP strategies, Sharpe ratio-maximizing strategies and machine learning-based return forecasting put high weight on past return-based characteristics such as short-term reversal or industry returns of big firms (IndRetBig). PPP-based maximization of the Sharpe ratio leads to a more even distribution across variable importance scores than DPPP-based maximization which mirrors our findings for variable importance for utility maximization above.

These results collectively reinforce three key findings. First, the value of deep learning in portfolio optimization extends beyond simple return prediction to the direct optimization of investor utility. Second, the benefits of preference-aligned optimization are robust across different utility specifications. Finally, sophisticated modeling approaches consistently outperform traditional passive strategies.

2.3.5 Robustness

We examine a number of alternative and extended model specifications. For the sake of brevity, the results are presented in the Appendix A.4 and we only discuss the main take-aways here.

Our main results are based on an expanding-window framework that uses successively more data for model estimation (see Section 2.3.1). Rolling-window estimation that uses a fixed number of months for training might be able to adapt more readily to potential structural changes in the data by discarding older observations. Results in Appendix A.4.2 show that rolling-window estimation does not consistently lead to better (or worse) results than expanding-window estimation. In fact, for high levels of risk aversion, the stability provided by a longer estimation sample can be crucial in achieving robust portfolio outcomes, aligning with our broader argument that model simplicity

Table 2.2: Comparing deep portfolio policies to benchmark strategies

	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$	SR_{PPP}	SR_{DPPP}	ML	Factor	EW	VW
$\sum \lvert w_i \rvert / N_t * 100$	0.1907	0.1938	0.1933	0.1729	0.1868	0.1951	0.1802	0.1128	0.0694	0.0694
$max\ w_i * 100$	1.1483	0.9843	0.8305	0.4582	0.8462	0.6434	0.4230	0.3987	0.0704	0.1113
$min\ w_i * 100$	-1.2824	-1.2053	-0.9743	-0.7224	-0.7137	-0.7960	-0.2801	-0.3262	0.0704	0.0410
$\sum w_i I(w_i < 0)$	-0.8748	-0.8974	-0.8932	-0.7464	-0.8466	-0.9068	-0.7987	-0.3132	0.0000	0.0000
$\sum I(w_i < 0) / N_t$	0.3400	0.3368	0.3319	0.3202	0.3654	0.3566	0.2009	0.2558	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	2.5891	2.5616	2.3473	1.7311	1.5163	1.8824	1.8640	0.4587	0.0829	0.0659
Mean	0.0341	0.0305	0.0281	0.0224	0.0202	0.0192	0.0193	0.0115	0.0110	0.0105
StdDev	0.0710	0.0550	0.0475	0.0378	0.0354	0.0354	0.0667	0.0416	0.0587	0.0552
Skew	2.6646	0.8411	-0.2470	-0.5201	-0.7604	-0.6342	-0.0544	-1.0598	-0.3716	-0.5039
Kurt	26.4755	10.9695	4.0705	1.9954	1.8370	1.8973	7.2504	3.2232	3.6591	3.3455
Max DD	0.4979	0.5601	0.4662	0.3027	0.2607	0.4356	1.0025	0.7836	0.9026	0.8585
Max 1M loss	0.2264	0.1789	0.1838	0.1446	0.1304	0.1336	0.3531	0.1924	0.2556	0.2398
CVaR (95%)	0.1107	0.0978	0.0882	0.0713	0.0732	0.0693	0.1466	0.1037	0.1351	0.1286
SR	1.6607	1.9230	2.0446	2.0491	1.9783	1.8815	1.0042	0.9590	0.6461	0.6609
$FF5 + Mom\ \alpha$	0.0353	0.0310	0.0284	0.0224	0.0201	0.0190	0.0198	0.0116	0.0114	0.0109
$StdErr(\alpha)$	0.0031	0.0024	0.0021	0.0016	0.0015	0.0015	0.0029	0.0018	0.0025	0.0024

This table presents out-of-sample performance estimates for various portfolio strategies using data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Strategies include CRRA-based deep parametric portfolio policies for different risk aversions, Sharpe ratio optimization approaches (SR) for linear and deep parametric portfolio policies, machine learning portfolios (ML), factor-based CRRA portfolio (Factor), and benchmark strategies (EW, VW). The DPPP strategy represents our baseline strategy for different risk aversions. The SR strategy is a PPP and DPPP optimized for Sharpe ratio preference. ML is the portfolio of a machine learning model trained to predict expected returns. The factor strategy is the linear PPP of a simple Fama-French five-factor model plus momentum. EW and VW are passive equal-weighted and value-weighted strategies. The first set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion) and portfolio turnover. The second set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), and annualized Sharpe ratios. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

and regularization often yield more reliable results.

An important question concerns the interaction of cross-sectional characteristics and the state of the macroeconomy in the portfolio weight function. To study the impact of macroeconomic variables, we expand our baseline model with 8 macroeconomic variables from Welch and Goyal (2008) as in Gu et al. (2020), and we interact each macroeconomic variable with each cross-sectional characteristic for a total of 1,413 covariates. Results in Appendix A.4.3 show that models including macroeconomic variables do not lead to higher investor utility than models that do not include macroeconomic variables, for all levels of risk aversion.

2.3.6 Market frictions

In the benchmark setting, average turnover and leverage are economically high for both the PPP and the DPPP. Next, we compare both approaches in a more economically feasible scenario that explicitly accounts for market frictions by imposing a transaction cost penalty and using a long-only constraint in the optimization task. Note that both frictions act as regularization mechanisms (Jagannathan and Ma, 2003; Hautsch and Voigt, 2019) on top of regularization via risk aversion. We therefore expect nonlinear and linear models to be closer for all levels of risk aversion in these scenarios, making it harder to isolate the risk aversion channel.

To account for transaction costs, we follow DeMiguel et al. (2020) and add the following penalty term to the optimization problem:

$$TC = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{N_t}|\kappa_{i,t}(w_{i,t} - w^{+}_{i,t-1})|\right], \tag{2.14}$$

where $\kappa_{i,t}$ are transaction costs for stock i at time t and $w^{+}_{i,t-1}$ is the portfolio weight before rebalancing, that is,

$$w^{+}_{i,t-1} = \frac{w_{i,t-1}(1 + r_{i,t})}{1 + \sum_{j=1}^{N_t} w_{j,t-1} r_{j,t}}. \tag{2.15}$$

Our transaction cost estimates come from Chen and Velikov (2023).[13] Thus, we define transaction costs $\kappa_{i,t}$ as the effective half bid-ask spread.

An important consideration when incorporating transaction costs into portfolio optimization is the inherently dynamic nature of the problem. In a multi-period setting,

[13]We thank the authors for making an updated version of the data available.

optimal portfolio weights at time t depend not only on current characteristics but also on expected future optimal positions and the associated trading costs. Jensen et al. (2022) formalize this intuition by deriving a closed-form solution in the mean-variance case that explicitly accounts for these dynamic effects. Our approach, following DeMiguel et al. (2020), instead incorporates transaction costs through a penalty term in the objective function (2.14). While this simplifies the dynamic aspect of the problem, it maintains tractability when dealing with a large cross-section of assets and characteristics while still capturing the first-order effects of trading frictions on portfolio choice. As our empirical results demonstrate, this formulation effectively constrains turnover and generates economically reasonable portfolios.

A large majority of equity portfolios face restrictions on short selling. We incorporate short-sale constraints as in Brandt et al. (2009), i.e., we restrict portfolios weights to be non-negative in the optimization problem of Equation (2.1) (and still keep the cap of 3% per stock). In particular, to make sure that portfolio weights still sum up to one, we add the following portfolio rebalancing term to our optimization process:

$$w_{i,t}^* = \frac{max[0, w_{i,t}]}{\sum_{j=1}^{N_t} max[0, w_{j,t}]}. \tag{2.16}$$

Table 2.3 shows separately the results of the optimization process with the transaction cost penalty and the long-only constraint for CRRA investors with different degrees of risk aversion. We show a selected set of results compared to Table 2.1, but provide similar tables with all results in Table A.6 for transaction cost and Table A.7 for long-only in the Appendix.

The first panel of Table 2.3 shows that even with transaction costs, the DPPP outperforms the PPP across all risk aversion levels, with monthly certainty equivalent differences ranging from 10 to 39 basis points (certainty equivalents are reported net of transaction costs). This suggests that, like risk aversion, the transaction cost penalty acts as an economic regularizer that reduces model complexity. Consequently, both models exhibit lower certainty equivalents and smaller, less significant differences, as supported by the reduced mean absolute weight differences in Figure 2.3. This is in line with the results of Hautsch and Voigt (2019), who show that a transaction cost penalty is analogous to a ridge penalty and thus acts as a natural economic regularization. As risk aversion increases, the significance of these differences declines with $\gamma = 10$ showing no

Table 2.3: Long-only & transaction costs constrained deep portfolio policy for CRRA investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
Transaction costs	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0155	0.0194	0.0129	0.0157	0.0077	0.0087	-0.0029	-0.0006
p-value($CE_{DPPP} - CE_{PPP}$)		0.0118		0.0620		0.3195		0.1555
$\sum w_i I(w_i < 0)$	-0.7139	-0.8877	-0.7612	-0.8973	-0.7638	-0.8798	-0.6588	-0.6756
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.8441	2.0257	0.8794	1.9002	0.8593	1.5947	0.7754	1.1407
Mean	0.0179	0.0225	0.0178	0.0221	0.0170	0.0182	0.0144	0.0157
StdDev	0.0482	0.0551	0.0427	0.0498	0.0397	0.0412	0.0360	0.0349
Max 1M loss	0.2228	0.2280	0.1812	0.2015	0.1559	0.1546	0.1303	0.1513
SR	1.2851	1.4123	1.4453	1.5370	1.4823	1.5296	1.3805	1.5552
p-value($SR_{DPPP} - SR_{PPP}$)		0.2090		0.2962		0.3852		0.0447
Long-only								
CE	0.0118	0.0164	0.0076	0.0107	0.0011	0.0020	-0.0157	-0.0104
p-value($CE_{DPPP} - CE_{PPP}$)		0.0001		0.0143		0.3308		0.0114
$\sum w_i I(w_i < 0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.5883	1.3508	0.6426	1.3417	0.5000	1.0914	0.3274	0.7656
Mean	0.0150	0.0215	0.0147	0.0216	0.0137	0.0174	0.0114	0.0135
StdDev	0.0566	0.0713	0.0510	0.0647	0.0459	0.0490	0.0406	0.0390
Max 1M loss	0.2483	0.2603	0.2171	0.2667	0.1968	0.2260	0.1832	0.1780
SR	0.9213	1.0418	0.9996	1.1580	1.0342	1.2262	0.9717	1.1974
p-value($SR_{DPPP} - SR_{PPP}$)		0.0119		0.0077		0.0006		0.0001

This table presents out-of-sample performance estimates for deep portfolio policies with the transaction costs penalty from Equation (2.14) and including a long-only constraint using 157 firm characteristics separately, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. Results in the first panel are reported net of transaction costs. For each panel the first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including leverage and portfolio turnover. The third set of rows displays the return distribution characteristics: the first two moments, maximum monthly loss, annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP.

significant difference at the 5% level while the constraints reduce turnover to 78–84% for the PPP and 114–203% for the DPPP. Despite higher turnover, the DPPP delivers notably larger net returns and higher Sharpe ratios.

The second panel of Table 2.3 presents long-only portfolio optimization results for CRRA investors. Here, the DPPP again outperforms the PPP (monthly certainty equivalent differences from 9 to 53 basis points), although the benefits of model complexity diminish more rapidly as risk aversion increases. The long-only constraint, like risk aversion, acts as an economic regularizer that reduces complexity, as evidenced by lower certainty equivalents and minimal weight differences in Figure 2.3. This is consistent with the results of Jagannathan and Ma (2003), who show that short-selling restrictions can also be interpreted as a form of regularization that implicitly shrinks the set of possible weights and prevents extreme allocations. Therefore, it also leads to more concentrated positions, with DPPP turnover ranging from 76.56-135.08% versus 32.74-58.83% for the PPP. Additionally, the DPPP achieves significantly higher Sharpe ratios at the 5% level across all risk aversion levels.

Similar patterns are observed when both constraints are applied jointly (see Table A.8 in the Appendix).

2.4 Alternative investor utility functions

2.4.1 Mean-variance and loss aversion

We explore results for different investor types by changing the utility function that we use to optimize the models. In particular, we consider linear and deep portfolio policies for an investor with mean-variance utility defined as

$$u(r_{p,t+1}) = r_{p,t+1} - \frac{\gamma}{2}\left(r_{p,t+1} - \frac{1}{T}\sum_{t=1}^{T} r_{p,t+1}\right)^2, \tag{2.17}$$

where γ is the absolute risk aversion of the investor, and for a loss-averse investor (Tversky and Kahneman, 1992) with utility defined as

$$u(r_{p,t+1}) = \begin{cases} -l(\overline{W} - (1 + r_{p,t+1}))^b & \text{if } (1 + r_{p,t+1}) < \overline{W} \\ ((1 + r_{p,t+1}) - \overline{W})^b & \text{otherwise} \end{cases}, \tag{2.18}$$

where $\overline{W}$ is a reference wealth level determined in the editing stage, the parameter l measures the investor's loss aversion and the parameter b captures the degree of risk seeking over losses and risk aversion over gains. For simplicity, we fix the parameters $\overline{W}$ and b at one and only change the loss aversion parameter l. We include the constraints specified in Section 2.3.1 in the optimization process for both preferences.

Table 2.4 shows separately the results of the optimization process for the mean-variance investors with different degrees of risk aversion and loss-averse investors with different degrees of loss aversion. We show a selected set of results compared to Table 2.1, but provide similar tables with all results in Table A.9 for mean-variance preference and Table A.10 for loss-aversion preference in the Appendix.

The first panel of Table 2.4 shows that for a mean-variance investor, the deep portfolio policy yields higher certainty equivalent returns than the linear policy across all risk aversion levels. While the DPPP's results (certainty equivalents, Sharpe ratios, and weight characteristics) are similar to those for a CRRA investor, the linear model performs relatively better in the mean-variance setting, reducing the monthly certainty equivalent difference to 23–86 basis points. The mean-variance utility function perfectly illustrates that the degree of absolute risk aversion determines the strength of the penalty on the variance of portfolio returns, i.e., the strength of regularization, since portfolio return variance is an explicit part of the utility function (see Section 2.2.2). Figure 2.3 illustrates the convergence of mean absolute weight differences between the two models with increasing risk aversion.

The second panel of Table 2.4 reports results for a loss-averse investor. Here, the DPPP outperforms the PPP at all levels of loss aversion, with improvements ranging from 11 to 123 basis points-differences significant at the 1% level for $l = 1.5$ and $l = 2$, at 5% for $l = 3$, and insignificant for $l = 4$. Because a loss-averse investor values the tail behavior of returns more than the mean–variance trade-off, both models show higher skewness compared to mean–variance or CRRA optimizations. Notably, the DPPP produces significantly higher (right) skewness, which explains its higher certainty equivalent. In line with our theoretical results in Appendix A.3, increasing loss aversion l does indeed penalize negative outcomes more severely.

Leverage of the PPP and DPPP strategies using different utility functions has about the same magnitude as in the CRRA benchmark results in Table 2.1. In Table A.11 in the Appendix, we show that certainty-equivalent returns are lower for long-only than for

Table 2.4: Deep portfolio policy for mean-variance and loss-averse investors with different degrees of risk aversion (γ) and loss aversion (l)

Mean-variance preference	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0201	0.0287	0.0184	0.0217	0.0143	0.0170	0.0065	0.0088
p-value($CE_{DPPP} - CE_{PPP}$)		0.0001		0.0292		0.0291		0.0849
$\sum w_i I(w_i < 0)$	-0.7602	-0.8882	-0.8059	-0.9070	-0.8093	-0.8899	-0.7879	-0.8925
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	1.5185	2.6428	1.7406	2.5648	1.6789	2.4174	1.4693	2.2676
Mean	0.0225	0.0319	0.0232	0.0281	0.0224	0.0276	0.0205	0.0254
StdDev	0.0492	0.0566	0.0435	0.0505	0.0402	0.0459	0.0373	0.0407
Skew	-0.6239	-0.1348	-0.8530	-0.6631	-0.8516	-0.4331	-0.7727	-0.5940
SR	1.5843	1.9506	1.8438	1.9259	1.9317	2.0786	1.9007	2.1596
p-value($SR_{DPPP} - SR_{PPP}$)		0.0019		0.2768		0.1185		0.0171
Loss-aversion preference	$l = 1.5$		$l = 2$		$l = 3$		$l = 4$	
CE	0.0188	0.0311	0.0147	0.0235	0.0082	0.0137	0.0025	0.0036
p-value($CE_{DPPP} - CE_{PPP}$)		0.0002		0.0015		0.0247		0.3014
$\sum w_i I(w_i < 0)$	-0.7929	-0.8918	-0.7980	-0.8833	-0.8090	-0.8823	-0.8083	-0.8702
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	1.6336	2.6846	1.5951	2.6742	1.6887	2.5599	1.7273	2.4745
Mean	0.0235	0.0361	0.0227	0.0319	0.0226	0.0306	0.0227	0.0275
StdDev	0.0494	0.0751	0.0442	0.0580	0.0412	0.0548	0.0395	0.0485
Skew	-0.6194	1.9765	-0.7339	0.5481	-0.7651	0.3536	-0.7475	-0.2204
SR	1.6475	1.6666	1.7793	1.9049	1.9052	1.9331	1.9905	1.9677
p-value($SR_{DPPP} - SR_{PPP}$)		0.4931		0.2424		0.4498		0.4400

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for mean-variance investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20 in the first panel and loss-averse investors with loss aversion (l) of 1.5, 2, 3, and 4 in the second panel. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including leverage and portfolio turnover. The third set of rows displays the return distribution characteristics: the first three moments, annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP.

leveraged strategies for mean-variance or loss aversion utility, mirroring our results for CRRA utility (Table 2.3). The PPP–DPPP return gap is of comparable magnitude to that with leverage and remains statistically significant, except under very high loss aversion, consistent with the theoretical framework and simulations. Figure A.12 in the Appendix further illustrates that the DPPP consistently outperforms the PPP over time, with the degree of outperformance varying with the investor's risk or loss aversion.

Finally, it is instructive to compare portfolio return moments across utility functions and portfolio policies in Table 2.4. Under mean-variance utility, increasing risk aversion is associated with lower portfolio return variance for both the linear (PPP) and complex (DPPP) models. In contrast, with loss aversion utility, higher loss aversion leads to lower skewness in absolute terms. This distinction highlights the relevance of capturing optimal portfolio return distributions when the investor's objective function is sensitive to higher moments. We study this in more detail in the next section.

2.4.2 Portfolio moments across utility functions and models

To highlight the economic implications of different utility functions, Figure 2.6 reports the first four moments of optimal portfolios across investor preferences, model classes, and regularization schemes. The figure compares CRRA, mean–variance, and loss-averse preferences, as well as a statistical objective function (StatReg) that maximizes mean return subject to an ℓ_1 penalty on model parameters with strength λ. The StatReg baseline provides a useful benchmark for understanding how purely statistical regularization compares to preference-induced regularization.

Three patterns are evident. First, increasing risk aversion or loss aversion systematically reduces volatility and higher-order moment exposures. Both CRRA and loss-averse investors display a monotonic decline in variance and kurtosis as parameters γ and λ increase, confirming the theoretical mechanism that preferences penalize estimation variance in a manner analogous to statistical shrinkage. This comparison is reinforced by the StatReg benchmark: purely statistical lasso regularization produces similar dampening of higher-order moments, but the utility-based approach targets precisely those risks that the investor's preferences place weight on.

Second, the different utility functions emphasize distinct moment trade-offs. CRRA portfolios initially exhibit very high mean return and positive skewness, but these come with extreme kurtosis that vanishes as risk aversion rises. Loss-averse portfolios are

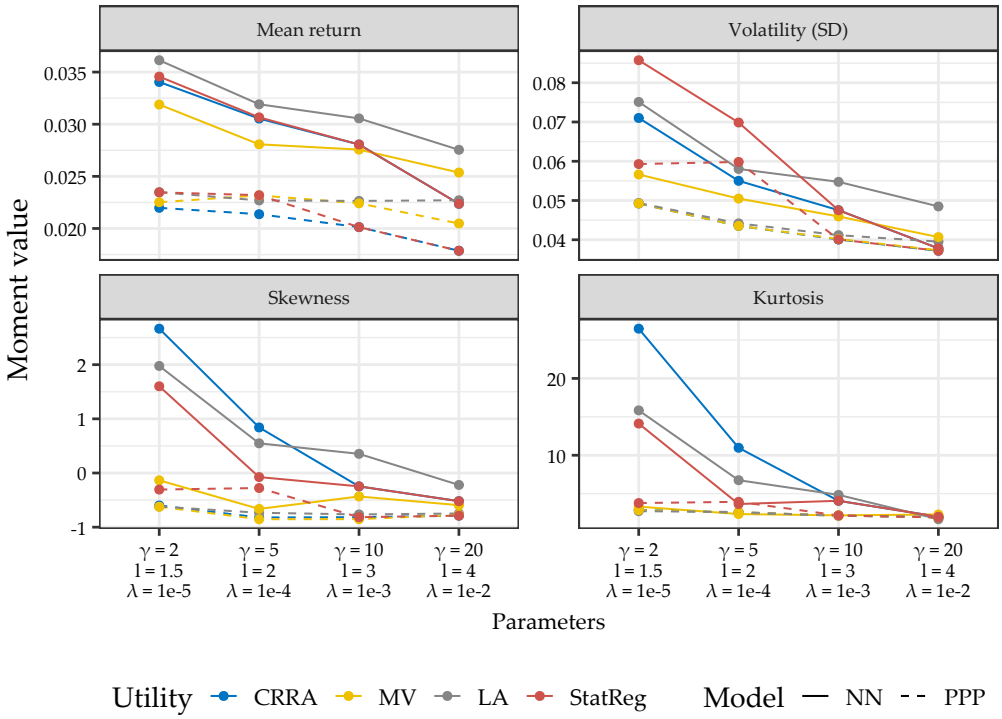


Figure 2.6: Portfolio moments across utility functions
This figure shows portfolio moments (mean return, volatility, skewness, and kurtosis) across utility functions, preference parameters, and model classes. The utility functions are CRRA (blue), mean–variance (yellow), loss aversion (gray) and StatReg (red). Solid lines denote DPPP; dashed lines denote PPP. StatReg optimizes mean return with an ℓ_1 penalty (lasso).

particularly distinctive: as loss aversion strengthens, these portfolios generate the highest skewness of any utility specification, but this comes jointly with high variance. In other words, loss aversion leads to portfolios that sacrifice stability in exchange for both higher upside and protection against extreme downside states. In contrast, the StatReg portfolio with the highest regularization yields the lowest mean return and relatively muted higher moments. This reflects the effect of the lasso penalty, which shrinks exposures but does not respond to investor preferences.

Third, a comparison of different model classes shows that the DPPP generally delivers higher means and stronger moment tilts at low levels of risk tolerance, and it converges more strongly. In contrast, the PPP exhibits a less exposed profile across preferences. As risk aversion or loss aversion increases, the gap between the DPPP and the PPP closes, reflecting the theoretical result that preferences shrink the range of possible outcomes and dampen the role of nonlinear interactions.

In Appendix A.5, we analyze the net exposure of different portfolios to different stock characteristics. While characteristics exposure is similar across utility functions and

risk/loss aversion levels for most clusters of characteristics, some intuitive differences emerge: As risk or loss aversion increases, investors decrease net exposure to short-term reversal and firm size. Since predictors are signed in the Chen and Zimmermann (2022) database to have positive mean return in the original in-sample periods (and smaller firms are associated with higher returns), this implies for size that more risk or loss averse investors load *more* on larger firms, presumably because such firms are less volatile. For short-term reversal, the decline in exposure is least pronounced for loss-averse portfolios, suggesting that loss-averse investors, unlike CRRA and mean-variance investors, are more willing to invest in stocks that have recently underperformed even for higher degrees of loss aversion. See Appendix A.5 for a more comprehensive discussion.

2.5 Conclusion

Building on the seminal work of Brandt et al. (2009) and the extensive literature on portfolio optimization and machine learning, we develop a novel Deep Parametric Portfolio Policy (DPPP) that integrates the structural advantages of traditional parametric portfolio policies with the flexibility of deep neural networks. Our approach maps a large set of firm characteristics to optimal portfolio weights in a nonlinear, interactive manner while directly incorporating market frictions and investor-specific utility functions such as CRRA, mean–variance, or loss aversion into the optimization process.

A key contribution of our work is the introduction of the concept of *economic regularization*. We provide a theoretical framework demonstrating how an investor's risk aversion naturally limits effective model complexity. As risk aversion increases, the incentive to exploit nonlinear relationships is tempered by the heightened penalty on return variance. Our simulations and analytical derivations show that, under higher risk aversion, the benefits of additional complexity diminish, leading the flexible DPPP to converge toward its linear counterpart. The results provide diagnostic guidance: if an investor is conservative or faces strong frictions, then a linear PPP is close to optimal; if the investor is aggressive and frictions are light, then the DPPP offers meaningful gains. In other words, one can match model complexity to investor preferences.

Our empirical investigation reinforces the theoretical insights. We document substantial improvements in investor utility when adopting the DPPP relative to standard linear models, across a variety of settings with different constraints or objective functions.

Across all settings, certainty-equivalent gains over linear models decrease systematically with increasing risk aversion. Further, our analysis of variable importance reveals that while past return-based signals dominate for low risk aversions, a more balanced mix of return-based and accounting-based characteristics emerges as risk aversion increases.

Overall, our approach puts forward a comparably simple and flexible neural network-based model that enables practitioners and researchers alike to create reasonable portfolio allocations based on firm characteristics and preferences, highlighting the growing role of machine learning and nonlinear models in finance. Moreover, the built-in economic regularization mechanism selects the optimal level of model complexity based on investor preferences directly. As such, it provides an intuitive alternative to purely statistical model regularization.

Chapter 3

Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data[†]

3.1 Introduction

Future earnings play a crucial role in determining the intrinsic value of assets (e.g., Monahan, 2018). Analysts, in particular, use earnings forecasts to derive buy/sell recommendations for stocks (e.g., Schipper, 1991; Brown, 1993). Earnings also play a key role in corporate decision-making, serving as a primary financial metric for external stakeholders (Graham et al., 2005). Lastly, as demonstrated by Ball and Brown (1968), earnings are directly related to stock returns. Return expectations can even be derived directly from predicted earnings in the form of the implied cost of capital (ICC) of a company (e.g., Gordon and Gordon, 1997; Claus and Thomas, 2001; Gebhardt et al., 2001; Easton, 2004; Ohlson and Juettner-Nauroth, 2005).

Earnings forecasts are generally obtained from either analysts or statistical models. Traditionally, statistical approaches have been primarily simple and linear in nature (e.g.,

[†]This chapter is based on Hess et al. (2025). We thank Tobias Bauckloh, Mario Hendriock, Alexander Kempf, Cédric Lesage, Roman Liesenfeld, Maximilian Müller, Michael Overesch, Tom Zimmermann and participants of the EAA Meeting 2024 as well as the CAAA Annual Conference 2024 for helpful comments and suggestions.

Hou et al., 2012; Li and Mohanram, 2014).[1] With the advent of more advanced statistical models, i.e., machine learning approaches in particular, various more flexible approaches have been proposed by researchers (e.g., Cao and You, 2021; Jones et al., 2023).

As outlined by Gu et al. (2020) and Israel et al. (2020), machine learning introduces flexibility along three dimensions: first, it accommodates complex functional forms. That is, in contrast to the traditional linear approaches that have been used to predict earnings, machine learning allows for complex nonlinear functional forms. Second, machine learning enables the use of large conditioning information sets, allowing researchers to uncover previously undetected relationships. Third, machine learning incorporates advanced optimization techniques, such as regularization, to prevent overfitting.

To the best of our knowledge, research on earnings prediction has primarily focused on the first dimension. Put differently, researchers have proposed the use of different types of machine learning algorithms to approximate the possibly complex functional form that relates predictors and future earnings (e.g., Cao and You, 2021). In general, more complex machine learning models are likely to yield more accurate earnings predictions. In fact, Kelly et al. (2024) demonstrate that increasing complexity, defined as the ratio of model parameters to data, consistently improves out-of-sample prediction performance in return forecasting. They explicitly recommend to use rich nonlinear model specifications rather than simple linear ones. We argue that further research along this dimension thus bears little additional insights. Moreover, extant studies on earnings prediction typically only assess a rather limited set of predictor variables. An important exception is the study by Chen et al. (2022) which exploits the entirety of Extensible Business Reporting Language (XBRL) data. Yet, they restrict their analysis to the prediction of earnings changes and do not predict earnings per se.[2] The fact that machine learning allows for the use of large conditioning information sets as mentioned by Israel et al. (2020) is rather unexploited in this context so far.

We address this gap by predicting annual earnings per share using a comprehensive set of variables – namely, the full set of financial statement variables from Compustat. This allows us to thoroughly analyze how fundamental accounting-based information

[1] Hereafter, we refer to these models as traditional models.

[2] In fact, as an extension to their main analysis, they also predict earnings levels. However, they only use a set of 24 variables for this and not the high-dimensional XBRL data set which they use for their main analysis. Surprisingly, their machine learning approach does worse than a simple random walk model. They conclude that earnings levels are hard to predict and use this as an argument for the fact that they focus on earnings changes in their primary analysis.

drives and relates to future earnings. We employ a selection of prominent, flexible machine learning models, i.e., a random forest model (RF), a gradient boosted tree model (GBT), a gradient boosted tree model with dropout (DART), a feed-forward neural network (NN) and an ensemble of the aforementioned models (ENML).[3] We further show how our approach relates to the most widely used traditional linear approaches, namely a simple model only including earnings as a predictor (L) (Gerakos and Gramacy, 2012), the HVZ-model (Hou et al., 2012), the EP-model (Li and Mohanram, 2014), the RI-model (Li and Mohanram, 2014) and an ensemble of the aforementioned models (ENTD).

A key contribution of our study is extensive model interpretation. Understanding the inner workings of prediction models is a fundamental requirement in asset management applications (Israel et al., 2020). However, due to their complexity, machine learning models are hard to interpret. In the machine learning earnings prediction case, model interpretation has thus far been restricted to metrics of variable importance and partial dependencies. More so, these metrics have usually been applied to a predetermined, restrictive set of predictor variables as mentioned above. Put differently, researchers typically choose or construct a set of predictor variables that they deem important before estimating the model. After estimating the models, they derive the extent to which variables from this predetermined set contribute to the predictions and assess the partial dependencies of predicted earnings in regards to them.

This study aims to broaden the limited scope of model interpretation found in the existing literature on earnings forecasts: first, we derive the relative importance of variables using SHAP (SHapley Additive exPlanations) values, an approach based on cooperative game theory (Lundberg and Lee, 2017).[4] Since we do not select or construct variables beforehand as done in comparable studies (e.g., Hansen and Thimsen, 2020; Cao and You, 2021), we are able to holistically infer which out of all the financial statement variables are important from a statistical perspective. We also derive the relative importance of different groups of financial statement variables, such as cash flow statement (CF/S) variables, income statement (I/S) variables and balance sheet (B/S) variables. Importantly, we conduct this analysis for forecast horizons of up to five years. This allows us to assess how different (groups of) variables might vary in terms of their

[3]The model selection is based on Bali et al. (2023).

[4]SHAP values are a way of explaining the results of any machine learning model. They are based on a game-theoretic approach that measures the contribution of each player to the final outcome. In machine learning, each variable is assigned an importance value that represents its contribution to the model's outcome (Lundberg and Lee, 2017).

predictive power, depending on the forecast horizon considered. Second, we analyze nonlinearity in the context of earnings prediction. In addition to partial dependencies that extant studies focus on in that context (e.g., Cao and You, 2021; Chen et al., 2022), we infer the degree to which different types of nonlinearity play a role. More precisely, we show the degree to which interaction effects across financial statement variables and other forms of nonlinearities, i.e., nonlinearity of the functional form, are important by means of surrogate modeling. This approach is completely transparent, intuitive and easy to replicate, irrespective of the model or software used. Again, we conduct these analyses for forecast horizons of up to five years. The last component of our model interpretation consists of stratifying the sample into different subsamples. Specifically, we group firms-years according to the life cycle stage, the size tercile, and the industry which the respective firm is in. We then assess differences in accuracy, variable importance as well as the impact of interactions and nonlinearities across groups. This enables us to more thoroughly understand how future earnings are related to current fundamental data and to infer which types of firms require more nuanced and complex earnings predictions models than others.

Our results can be summarized as follows: first, in line with previous research on model averaging, we find that ensemble models improve predictive accuracy relative to their individual components across both traditional and machine learning approaches. For instance, in one-year-ahead predictions, the machine learning ensemble further improves accuracy beyond the best-performing individual machine learning model, achieving an additional 1% gain. Given that the best machine learning model already substantially outperforms the best traditional model, this additional improvement underscores the value of ensemble modeling. Notably, the machine learning ensemble consistently outperforms all component models across all forecast horizons

In terms of bias, the traditional and the machine learning approaches yield very similar results. Both types of models yield mostly unbiased predictions for forecast horizons of up to three years. An interesting exception are the traditional models who tend to underestimate one-year-ahead earnings. For forecasts horizons of four and five years, the models considered begin to systematically overestimate earnings. However, the machine learning models are consistently less biased in terms of levels as compared to their traditional analogues.

Mirroring previous findings on machine learning earnings predictions, we further

show that our machine learning approaches constantly outperform traditional linear approaches in terms of accuracy (e.g., Cao and You, 2021). For the one-year forecast horizon, the best performing machine learning model, i.e., the ENML is around 11% more accurate than the best performing traditional model, i.e., the ENTD. Even for long forecast horizons of five years, we find that the ENML beats the ENTD in terms of accuracy by around 6%. Furthermore, assessing accuracy differences across out-of-sample periods shows that model performance converges in the periods following the financial crisis, and diverges in favor of the machine learning approaches afterwards.

Assessing the degree to which the models are able to explain out-of-sample variation in earnings makes an even more convincing case for the ML approaches. In fact, the average out-of-sample R^2 (OOS R^2) of the ENML is between 15% and 23% higher than that of the ENTD for the forecast horizons considered.

Lastly, we show that the more accurate predictions of the machine learning approach translate into more profitable portfolios based on ICC. An investment strategy which invests into the top ICC quintile based on ENML forecasts each year outperforms an analogous strategy which utilizes ICC based on ENTD forecasts. Moreover, we show that both ICC strategies outperform a value-weighted market portfolio in terms of cumulative returns.

Turning to model interpretation of the machine learning approach, we find that current I/S variables, particularly current earnings, play a dominant role in short-term earnings predictions. This aligns with traditional models, which also heavily rely on lagged earnings as a key predictor. However, a key novel finding is that with increasing forecast horizon, variable importance shifts significantly, becoming more balanced among financial statement types. For one-year-ahead predictions, I/S variables contribute around 66% while B/S variables contribute around 19% to total importance. Over longer horizons, this imbalance diminishes: for $t + 5$-predictions, I/S variables contribute around 47%, while B/S variables contribute around 37%. CF/S variables consistently contribute around 15% to total importance throughout the forecast horizons considered. Put differently, the longer the forecast horizon, the more important B/S information becomes. This shift suggests that while short-term earnings primarily reflect income statement dynamics, longer-term earnings expectations are increasingly shaped by firms' balance sheet fundamentals.

Further disentangling the effects of different components of financial statement

information reveals that certain pieces of financial statement information dominate others. For example, debt and supplemental information resemble the most important pieces of B/S information. Turning to the CF/S, we find that variables related to the operating cash flow are much more relevant than variables related to either the investing cash flow or the financing cash flow. Lastly, turning to the I/S, we find that especially EBIT and net income are important, contributing around 18% and 25% to total importance for one-year-ahead forecasts, respectively. Interestingly, while the importance of net income declines to approximately 12% for five-year forecasts, EBIT maintains a consistent level of significance. This suggests that information which is less exposed to accounting manipulation gains relevance for longer-term forecasts.

We further show that for one-year-ahead predictions, a linear surrogate model is able to explain around 90-94% of the variation in machine learning earnings predictions across out-of-sample periods. Interactions and nonlinearities contribute somewhat equally to explaining the remaining variation. In fact, adding interactions increases the R^2 by around 3-4%. We attribute the remaining unexplained variation to other types of nonlinearities, which are not captured by interactions, i.e., nonlinearity of the functional form. As the forecast horizon increases, the linear surrogate model provides a progressively less accurate approximation of the relationship between predictions and inputs, suggesting that longer-term earnings forecasts involve increasingly complex functional dependencies that require more flexible models to capture effectively.

The last part of our model interpretation consists of stratifying the sample into different types of firms, as outlined above. We find that generally, the ENML dominates the ENTD in terms of accuracy for each firm subsample considered. However, significant differences in accuracy (improvements) across subsamples exist. Most notably, this can be seen for firms in different life cycle stages. Earnings of firms in the beginning and the end of their life cycle are forecasted with substantially lower accuracy than those of firms in the middle of their life cycle. Importantly, it is precisely for these firms – where earnings predictability is inherently more difficult – that the ENML achieves the most significant improvements over the ENTD. Moreover, we find that for these firms, interactions account for a larger portion of the ENML predictions than for firms in the middle of their life cycle. Again, this suggests that when the prediction task is more complex – whether due to longer forecast horizons or greater firm heterogeneity – models that allow for richer interactions and nonlinear structures provide a distinct advantage.

The remainder of this paper is structured as follows: In Section 3.1 we outline the relevant literature that we are contributing to. In Section 3.2 we describe our empirical approach. We evaluate and compare our approach in Section 3.3. In Section 3.4 we provide extensive interpretation. Finally, Section 3.5 concludes the study.

Related Literature

Our work relates to three strands of literature in particular. First, we contribute to the literature on machine learning applications in finance. Machine learning methods have become the prevalent way of conducting prediction exercises, primarily due to their superiority in terms of flexibility as compared to traditional econometric methods and their efficacy in regards to large sets of input data (e.g., Israel et al., 2020; Kelly et al., 2024). For example, Gu et al. (2020) show how different machine learning approaches perform in terms of predicting stock returns and Bali et al. (2023) apply machine learning to the task of predicting option returns. Our study is similar, in the sense that we predict another financial variable, i.e., earnings, with machine learning.

Second, we apply state-of-the-art techniques to interpret our machine learning predictions, thereby explicitly responding to the "need for interpretability" of financial machine learning models as formulated by Israel et al. (2020). We hence also contribute to the literature that aims to foster transparency and understanding of machine learning methods for prediction in finance research, such as e.g., Bali et al. (2023) who extensively assess machine learning option return predictions.

Third, we contribute to the literature on model-based earnings forecasts. Traditionally, researchers have suggested predicting earnings using time-series regression models (e.g., Ball and Watts, 1972; Albrecht et al., 1977; Watts and Leftwich, 1977), cross-sectional regression models (e.g., Hou et al., 2012; Li and Mohanram, 2014; Harris and Wang, 2019) or even simple random walk models (e.g., Li and Mohanram, 2014). More recently, however, several machine learning earnings prediction approaches have been implemented in the research (e.g., Hansen and Thimsen, 2020; Cao and You, 2021; Chen et al., 2022; Hendriock, 2022; Campbell et al., 2023; Jones et al., 2023; Van Binsbergen et al., 2023). However, the extant literature differs from our study in several key aspects which are outlined in the following.

Hansen and Thimsen (2020) also estimate a range of machine learning methods. They

use a more high-dimensional input vector than other studies on predicting level earnings. However, it is still restrictive in the sense that it is based on prior research and not as high-dimensional as our data. Moreover, in contrast to our study, no model interpretation is provided.

The study of Cao and You (2021) also encompasses different machine learning models. However, they use a more restrictive set of input variables than us and provide only limited model interpretation. Moreover, Cao and You (2021) validate their models using traditional cross-validation (and a limited hyperparameter space). This, however, destroys the temporal structure of the observations and introduces information leakage (Gu et al., 2020). We preserve the temporal ordering of the observations by using fixed training, validation and test intervals.

Chen et al. (2022) use a single model (gradient boosted trees) as opposed to our multi-model approach. Furthermore, they predict binary earnings changes, while we focus on predicting level earnings.

Hendriock (2022) suggests predicting earnings by predicting the complete conditional density function. However, he restricts his input variable space to the one as defined by traditional linear models. Moreover, he does not provide model interpretation.

Campbell et al. (2023) benchmark an extensive range of machine learning model specifications with the aim of identifying the ones which compare the best to analyst forecasts. Apart from the fact that their model choices differ from ours, they use a different, smaller set of inputs, i.e., the Wharton Research Data Services (WRDS) Financial Suite Ratios extended by some additional variables like e.g., the stock return. Furthermore, they provide limited model interpretation as their primary focus is on the aforementioned horse-race between model specifications.

Jones et al. (2023) use a single model, i.e., a gradient boosted tree model algorithm, as opposed to our ensemble approach. In contrast to our study, their target variable is return on net operating assets and they use a set of six ratios as their predictors. Another crucial difference is that Jones et al. (2023) exclusively forecast earnings (changes) in $t+1$, while we forecast earnings (per share) for horizons $t+1$ to $t+5$. Finally, to the best of our knowledge, as the only other study in this context, they assess the impact of interactions. However, their method of doing so and their predictor variables differ from ours. Our surrogate modeling approach is easily applicable to any type of model, in any software and further allows us to explicitly determine the effect of interaction effects and

nonlinearity in parameters. Interestingly, our findings in regards to interactions differ strongly from the ones provided by Jones et al. (2023). They find substantial importance of interactions, while we find that interactions among financial statement variables are irrelevant.

Van Binsbergen et al. (2023) use a random forest model to predict earnings conditional on financial ratios, similar to Campbell et al. (2023). In contrast, our approach involves estimating a spectrum of machine learning models individually and in ensemble configurations. Moreover, we utilize an entirely distinct set of input data, specifically the comprehensive Compustat financial statement dataset. Finally, unlike their study, which primarily assesses analyst biases, our analysis encompasses detailed explanations of model predictions.

Summing up, to the best of our knowledge, we are the first to predict level earnings per share for forecast horizons of up to five years using the entirety of available Compustat financial statement variables. Furthermore, we contribute novel guidance for future research on earnings (per share) predictions by thoroughly interpreting our state-of-the-art machine learning approaches using model agnostic and easily applicable methods, something that has been not done extensively thus far.

3.2 Empirical approach

3.2.1 General setup

We express earnings E of firm i in period $t+\tau$ as the expectation in period t plus an error ϵ:

$$E_{i,t+\tau} = \mathbb{E}_t[E_{i,t+\tau}] + \epsilon_{i,t+\tau}. \tag{3.1}$$

Every earnings prediction model can be viewed as an attempt to estimate $\mathbb{E}_t[E_{i,t+\tau}]$, the expected value. More precisely, we assume that expected earnings in $t+\tau$ are a function of a vector of inputs $\boldsymbol{X}$ of firm i known at time t:

$$\mathbb{E}_t[E_{i,t+\tau}] = f(\boldsymbol{X}_{i,t}). \tag{3.2}$$

It becomes evident that modeling expected earnings consists of three crucial parts. First, one has to determine which inputs enter the model, that is, how $\boldsymbol{X}$ is defined.

Second, one has to decide which functional form $f(\cdot)$ takes on. This corresponds to the decision about which empirical model to choose from. Lastly, one has to decide on how to estimate $f(\cdot)$, i.e., which statistical loss function to minimize. The latter is not the explicit focus of this study and thus we keep it simple. We follow the original implementations of the traditional models and estimate them using the mean squared error (MSE). In case of the machine learning models, we follow Gu et al. (2020) and estimate the models both using the MSE and the mean absolute error (MAE) and report the predictions based on the loss function that leads to more accurate forecasts according to the price scaled absolute forecast error (PAFE) at the 1-year horizon.[5]

Thus, apart from the decision regarding the loss-function, the two contrasting *extreme* approaches to predicting earnings are: (1) actively making the decision which variables and which functional form to assume ex ante, and (2) letting the data speak by selecting a model that permits flexible functional forms and providing it with the entire data (or at least a very large set of variables) available. The former corresponds to the traditional prediction approaches suggested by e.g., Hou et al. (2012). To the best of our knowledge, the latter has not yet been implemented for earnings per se, a noteworthy exception being Chen et al. (2022) who employ this approach for (binary) earnings *changes*. There are approaches that fall in between (1) and (2). In these approaches, either the functional form or the input vector is restricted significantly (e.g., Hendriock, 2022; Van Binsbergen et al., 2023).

Our study focuses on assessing the second, flexible approach (2) and comparing it to traditional approaches (1). We more thoroughly elaborate on the choice of the input vector and of the model in 3.2.2 and 3.2.3, respectively.

3.2.2 Data

US annual financial statement data is obtained from Compustat. Our sample period ranges from 1988 to 2023. This is due to the fact that CF/S-data is only sparsely available prior to 1988. To conduct the ICC portfolio evaluation, we add price and return data from CRSP to the Compustat data used for model estimation. Moreover, we drop observations with missing prices, prices smaller than 1$, missing or zero common shares outstanding or missing earnings. This results in a final sample for model estimation that consists of

[5]To be precise, Gu et al. (2020) choose either the MSE or the Huber loss, depending on which performs better. We choose either the MSE or the MAE, depending on which performs better.

158,809 observations.

For our machine learning models, we use the Compustat financial statement items as predictors. We drop variables with more than 50% of observations missing or no observations in any of the cross-sections (i.e., estimation years), yielding 193 variables.[6] An overview over these variables is given in Table B.3 in the Appendix. Analogous to Chen et al. (2022), we include lags and first-order differences of these variables, resulting in a set of 579 predictor variables in total. For the traditional models, we construct input variables according to the respective models. An overview over these variables is given in Table B.1 in the Appendix. All our variables, including our target variable, are scaled by common shares outstanding and winsorized at the 1%- and 99%-level, respectively. Finally, since neural networks are sensitive to scale differences across input variables, we standardize each variable.

3.2.3 Models

We estimate two groups of models. The first group consists of popular simple linear models that have been introduced in the literature thus far. All of these models assume a linear additive relation between earnings and some low-dimensional input vector, i.e.,

$$\mathbb{E}_t[E_{i,t+\tau}] = \boldsymbol{\beta}' \boldsymbol{X_{i,t}}, \tag{3.3}$$

where $\boldsymbol{\beta}$ denotes a vector of coefficients. The models differ in terms of which variables the input vector consists of. A more detailed description is given in Appendix B.1. A difference to be noted is that the models also slightly differ in how they define the output. While the RI and the EP model use earnings per share, the HVZ model uses earnings. In this study, we define earnings as "income before extraordinary items" (Compustat variable: ib). As mentioned above, we consistently scale our output as well as our input variables by common shares outstanding in all of the models. Our forecast horizon, both for this and the following group of models spans from $\tau = 1$ to $\tau = 5$.

The second group of models consists of flexible models that are able to approximate arbitrary complex functional forms. Put differently, we do not assume any specific type of functional form when estimating these models. Analogous to Bali et al. (2023), we

[6] We also drop variables already scaled by shares. The reason for that is, that we scale all our variables by shares as mentioned below and hence these variables are redundant. However, this only pertains to five variables in our study.

estimate a random forest model (RF), a gradient boosted tree model with (DART) and without dropout (GBT) and a neural net (NN). Importantly, we try to restrict our input vector as little as possible. Specifically, we feed the models the 579 variables as outlined above, including the lags and first-level differences. We argue that this input vector corresponds to a proxy for the entirety of (relevant) financial statement input variables. Note that extending the input vector even further implies more computational effort.[7]

Since studies in other realms of financial forecasting have shown that using an ensemble of models may prove superior to using single models (e.g., Bali et al., 2023), we derive the equally weighted average prediction for both groups of models, i.e., we derive two ensemble model predictions:

$$\mathbb{E}_t^{(En)}[E_{i,t+\tau}] = \frac{1}{J}\sum_{j=1}^{J}\mathbb{E}_t^{(j)}[E_{i,t+\tau}], \tag{3.4}$$

where $j \in J$ denotes the respective single model and En denotes the respective ensemble model.

The ensemble of the traditional models is denoted by ENTD and the ensemble of the fully flexible machine learning prediction models is denoted by ENML.

3.2.4 Out-of-sample approach

We employ a rolling window strategy to obtain our out-of-sample prediction results. Specifically, for the machine learning models, we divide our data into training, validation and test sets. For each forecast horizon τ, the process for generating forecasts as of t proceeds as follows: we train our models using earnings from $t-11$ to $t-2$ as output and corresponding financial statement data lagged by τ as predictors. Next, we tune the machine learning models using earnings from $t-1$ to t as output and lagged financial statement data by τ as predictors. Tuning involves determining the optimal hyperparameter values for the model, as detailed in Table B.2 in the Appendix. Subsequently, earnings predictions for $t+\tau$ are derived by inputting variables from t into the optimized models. This process is repeated recursively, advancing one year at a time. We follow the approach of Hou et al. (2012) and Li and Mohanram (2014), estimating models at the end of June each year, under the assumption of a reporting lag of three to

[7] We could have also included other variables like price, analyst forecasts, etc. (e.g., Campbell et al., 2023; Van Binsbergen et al., 2023). However, the focus of our study is the thorough analysis of the relationship between fundamental accounting-based information and future earnings.

fourteen months for financial statements.[8]

In contrast, traditional linear approaches do not require a tuning window. Hence, we only partition the data into training and test sets when estimating these models. The subsequent steps of the procedure remain unchanged. More precisely, for each forecast horizon τ, models are trained using earnings from $t-11$ to t as output and corresponding lagged financial statement data by τ as predictors. Earnings predictions for $t+\tau$ are then derived by utilizing predictor variables from t.

3.2.5 Evaluation

We evaluate the predictive performance of the models across a range of evaluation metrics. First, we compute the error metrics that are common in the earnings prediction literature (e.g., Hou et al., 2012). For each forecast horizon $\tau \in [1,2,3,4,5]$, these include the price scaled forecast error (PFE) or bias:

$$PFE_{i,t+\tau} = \frac{E_{i,t+\tau} - \hat{E}_{i,t+\tau}}{Price_{i,t}}, \tag{3.5}$$

and the price scaled absolute forecast error (PAFE) or accuracy:

$$PAFE_{i,t+\tau} = \frac{|E_{i,t+\tau} - \hat{E}_{i,t+\tau}|}{Price_{i,t}}, \tag{3.6}$$

where $E_{i,t+\tau}$ denotes actual earnings for firm i in period $t+\tau$, $\hat{E}_{i,t+\tau}$ denotes the respective forecast and $Price_{i,t}$ is the firm's stock price at the end of June in the respective estimation year.

Second, we assess the out-of-sample R^2 ($OOSR^2$) of each individual as well as the ensemble forecasts, i.e., for every out-of-sample period we calculate

$$OOS\,R^2_{t+\tau} = 1 - \frac{\Sigma_{i=1}^{n_{t+\tau}}(E_{i,t+\tau} - \hat{E}_{i,t+\tau})^2}{\Sigma_{i=1}^{n_{t+\tau}}(E_{i,t+\tau} - \bar{E}_{t+\tau})^2}, \tag{3.7}$$

for each model. Here, $\bar{E}_{i,t+\tau}$ denotes average earnings of firms in period $t+\tau$. Albeit not commonly used in the earnings prediction literature (e.g., Hendriock (2022) being an exception), this evaluation metric is of particular importance for the typical use case of earnings predictions, i.e., long-short ICC portfolios. In this context, predicting cross-sectional variation is much more important than accurately predicting earnings per se,

[8] Specifically, data from April of year $t-1$ to March of year t is considered the most recent fiscal year-end data available as of June in year t, capturing the information as of t.

since an investor goes long (short) the stocks which's ICC is high (low) in cross-sectional comparison.

Lastly, we derive the ICC based on the two ensemble forecasts. We follow the literature and calculate ICC following the methods of Gordon and Gordon (1997), Claus and Thomas (2001), Gebhardt et al. (2001), Easton (2004) and Ohlson and Juettner-Nauroth (2005).[9] More precisely, our ICC estimates are derived as the average of the five aforementioned methods using both the traditional ensemble and the machine learning ensemble earnings predictions. We then construct equally weighted long-short zero investment portfolios based on ICC and assess their average performance across the out-of-sample periods.

3.2.6 Interpretation

A primary contribution of this study is the comprehensive interpretation of the machine learning approach, addressing a key issue in machine learning applications in finance and accounting (Israel et al., 2020). Our study fills a gap in the existing literature, which either lacks model interpretation entirely or provides only limited insights, as discussed above. More precisely, we derive the variable importance and the degree to which different types of nonlinearity play a role for our best performing machine learning model, i.e., the machine learning ensemble. Further, we assess differences in accuracy, variable importance as well as the degree of nonlinearity across different subsamples of firms.

Variable importance

We determine the importance of the different variables with respect to the earnings prediction. To do so, we compute their SHAP values, a state-of-the-art approach for assessing the importance of input variables which is based on cooperative game theory (Lundberg and Lee, 2017). In essence, SHAP values approximate how a model's prediction changes when knowing the value of a respective input variable. The approach is model-agnostic and allows us to evaluate the importance of input variables irrespective of the model used. We conduct these analyses at both the individual variable level and the grouped-variable level. Specifically, we determine the relative importance of predictors grouped into balance sheet, cash flow statement, and income statement data as well as predictors grouped into current, lagged, and difference variables. Furthermore, we

[9] A description of the models is provided in Appendix B.3.

provide an in-depth accounting perspective on which specific types of financial statement information are important by breaking the financial statements down into schematic components. Importantly, we conduct these analyses per forecast horizon.

Nonlinearity

In addition to deriving the variable importance, we also evaluate the extent to which nonlinearity plays a role in our machine learning model.

First, we assess the degree to which nonlinearity in terms of variables and nonlinearity in terms of functional form play a role. In a first step, we regress predicted earnings on the 20 most important input variables using a linear regression model.[10] In a second step, we add all possible two-way interactions to the surrogate regression model from the prior step. Assuming there is some degree of nonlinearity present in the fully flexible model, this allows us to disentangle the degree to which nonlinearity in terms of variables (i.e., interactions across inputs) and nonlinearity in terms of functional form play a role. More specifically, we assess the adjusted in-sample R^2s of the two surrogate models. The R^2 of the first-step surrogate model indicates the degree to which the predictions are linear. The difference between the R^2 of the first-step surrogate model and the second-step surrogate model indicates the degree to which nonlinearity in terms of variables, i.e., (two-way) interactions across financial statement variables, plays a role. We attribute the portion that remains unexplained by the second-step surrogate model to nonlinearity in functional form.[11]

Second, we assess the partial dependence of earnings with respect to the most important input variables. As typically done in the literature, we evaluate the partial dependencies graphically via so-called partial dependence plots. To do so, we fit a non-parametric lowess model (locally weighted linear regression) to the SHAP values of a predictor value of interest (the output) and the associated predictor values (the input) and plot the result. This allows us to approximate the effect of the respective predictor variable on future earnings.

Again, we conduct the surrogate modeling and the partial dependence analyses per

[10]We only use the 20 most important variables, because otherwise, including all possible two-way interactions in the second step requires an excessive amount of computing power. More so, we find in undocumented results that adding more variables does not significantly change the results.

[11]Theoretically, the unexplained portion also includes effects of interaction terms of order three and higher. However, we assume that these can be neglected and find evidence for this assumption in undocumented analyses.

forecast horizon.

Subsample analysis

The last part of our model interpretation comprises a subsample analysis. More precisely, we briefly assess differences in accuracy, variable importance as well as non-linearity across different subsets of firm-year observations. Inter alia, this assessment allows us to determine whether the superiority of the machine learning model stems from specific subsamples specifically and whether there exist differences in how current fundamental information is related to future earnings across different subsets firms.

We consider three ways of stratifying the sample: first, we stratify predictions according to the life cycle stage the respective firm is in in the period in which the forecast is conducted. We define life cycle stages following Dickinson (2011).[12] Second, we stratify predictions according to the firm size, measured via market capitalization as of the period in which the forecast is conducted. Lastly, we stratify predictions according to the industry, defined via the Fama-French 5 industries, a firm is in as of forecast date.[13]

3.3 Evaluation

3.3.1 Accuracy and bias

Chen et al. (2022) report that their machine learning approach does worse than a simple random walk type model when predicting level earnings. They explicitly state that level earnings are hard to predict and thus resort to the prediction of earnings changes in their main analysis. In contrast, mirroring findings by e.g., Cao and You (2021) our flexible machine learning models for earnings (per share) level prediction outperform the traditional linear models by a significant margin.

Price scaled forecast error

Table 3.1 shows the time-series averages of the median PFE for the four traditional, the four machine learning and the two ensemble models. The PFE provides insight into whether the estimated earnings are systematically over- or underestimated (biased)

[12]The classification scheme is summarized in Table B.5 in the Appendix.

[13]Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) website. The classification scheme is summarized in Table B.6 in the Appendix.

relative to actual earnings.

In general, the two ensemble models appear to be the least biased models for each group of models considered. Both models yield insignificant biases in most cases. This is in line with extant studies on earnings prediction, which report that earnings predictions by statistical models do not exhibit biases as opposed to those by analysts (Hou et al., 2012). In fact, only the traditional ensemble yields significant biases for two out of five forecast horizons considered. Interestingly, it systematically underestimates earnings in $t+1$. Both ensemble models tend to systematically overestimate earnings with increasing forecast horizons. However, the machine learning ensemble achieves between 44.80-87.50% lower biases, depending on the forecast horizon.

The results further indicate that at the non-ensemble level across forecast horizons $t+1$ to $t+3$, some models yield statistically significant PFEs, with the most biased machine learning model for $t+1$ being the RF model (0.0059) and the most biased traditional model for $t+1$ being the RI model (0.0063). Interestingly, the three more sophisticated traditional models systematically underestimate earnings in $t+1$, reflecting the aforementioned results regarding the ensembles. For $t+4$ no single model yields statistically significant PFEs, while for $t+5$, most non-ensemble models yield PFEs that are significantly different from zero. Overall, our findings suggest that non-ensemble machine learning models generally yield lower PFEs compared to their traditional counterparts.

In conclusion, machine learning models generally exhibit lower bias. Additionally, earnings prediction models tend to systematically overestimate earnings as the forecast horizon increases, regardless of whether traditional linear or machine learning methods are employed.

Price scaled absolute forecast error

Turning to the next evaluation metric, Table 3.2 reports the time-series averages of the median PAFEs for the four traditional, the four machine learning and the two ensemble models. The PAFE is a measure for the accuracy of a model, with values closer to zero indicating higher accuracy.

First, we observe a positive effect of model stacking. Overall, among the traditional models, the best-performing one is the traditional ensemble, while among the machine learning models, the best-performing one is the machine learning ensemble. Specifically,

Table 3.1: Median PFE

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.0007	−0.0001	−0.0027	−0.0083	−0.0146**
HVZ^{MSE}	0.0039***	0.0033	0.0006	−0.0054	−0.0129***
EP^{MSE}	0.0056***	0.0055**	0.0033	−0.0027	−0.0091*
RI^{MSE}	0.0063***	0.0059**	0.0028	−0.0032	−0.0103**
ENTD	0.0043***	0.0035	0.0008	−0.0055	−0.0125**
RF^{MSE}	0.0059***	0.0062***	0.0043*	0.0001	−0.0052
GBT^{MAE}	−0.0011	−0.0019	−0.0019	−0.0040	−0.0083*
$DART^{MAE}$	−0.0003	−0.0011	−0.0020	−0.0038	−0.0061
NN^{MAE}	−0.0006	−0.0010	−0.0019	−0.0043	−0.0081*
ENML	0.0014	0.0009	0.0001	−0.0029	−0.0070

This table reports the time-series averages of the median price scaled forecasting errors (PFEs) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript MAE (MSE) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t+1$ horizon). The PFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

the machine learning ensemble outperforms each of its individual component models for every forecast horizon. The same is true for the traditional ensemble for the horizons $t+1$ to $t+2$. Especially in the traditional case, this result is surprising, since the models differ very little in terms of the predictor variables. For forecast horizons $t+3$ to $t+5$, the traditional ensemble performs slightly worse than the best performing traditional component model(s) (0.0396 vs 0.0394 in $t+3$, 0.0445 vs 0.0438 in $t+4$ and 0.0505 vs 0.0494/0.0504 in $t+5$). Second, we find that in most cases, all machine learning models, including the ensemble, outperform all traditional models, including the ensemble, for all prediction horizons. However, for forecast horizons $t+3$ to $t+5$, the RI as well as the ENTD model are more accurate than some of the machine learning component models. This stresses the benefit of model averaging for the nonlinear machine learning models specifically. The difference in accuracy between the machine learning and the linear ensemble is at a statistically significant level between -0.0020 and -0.0032. This translates into a relative difference of 11.44% for earnings in $t+1$ to 6.34% for earnings in $t+5$. The

ENML thus provides not only statistically significant, but also economically meaningful gains in accuracy over the traditional models.

Table 3.2: Median PAFE

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.0297***	0.0380***	0.0439***	0.0499***	0.0568***
HVZ^{MSE}	0.0282***	0.0360***	0.0416***	0.0473***	0.0534***
EP^{MSE}	0.0280***	0.0356***	0.0407***	0.0451***	0.0504***
RI^{MSE}	0.0278***	0.0350***	0.0394***	0.0438***	0.0494***
ENTD	0.0271***	0.0345***	0.0396***	0.0445***	0.0505***
RF^{MSE}	0.0268***	0.0344***	0.0398***	0.0452***	0.0494***
GBT^{MAE}	0.0266***	0.0341***	0.0394***	0.0449***	0.0499***
$DART^{MAE}$	0.0243***	0.0325***	0.0383***	0.0444***	0.0489***
NN^{MAE}	0.0247***	0.0339***	0.0380***	0.0444***	0.0481***
ENML	0.0240***	0.0317***	0.0372***	0.0425***	0.0473***
ENML-ENTD	−0.0031***	−0.0028***	−0.0024***	−0.0020***	−0.0032***

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript MAE (MSE) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t+1$ horizon). The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Figure 3.1 illustrates the median PAFE of the ENML and ENTD for each out-of-sample year and for forecast horizons of $t+1$ and $t+5$, respectively. The plots illustrate that the overall PAFE levels strongly increase with increasing forecast horizon. Furthermore, they reveal that the machine learning ensemble is more accurate in all years and for both forecast horizons, except for 2009, in which the ENTD yields slightly more accurate $t+5$ forecasts than the ENML. In general, the difference between the ENML-PAFE and the ENTD-PAFE varies across out-of-sample periods.

3.3.2 Out-of-sample R^2

The next metric considered is the out-of-sample R^2 ($OOS\ R^2$). The results are reported in Table 3.3. The $OOS\ R^2$ allows us to assess how the models perform in terms of explaining

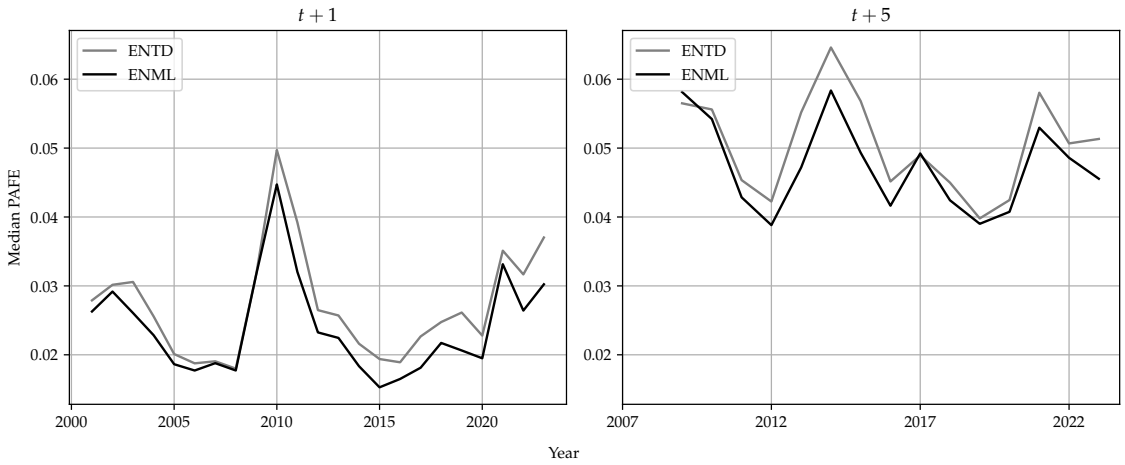


Figure 3.1: PAFE across out-of-sample periods
This figure shows the median price scaled absolute forecast errors (PAFEs) of the machine learning ensemble (ENML) and the traditional ensemble (ENTD) per out-of-sample period for forecast horizons $t+1$ and $t+5$.

out-of-sample variation in future earnings.

As expected, the *OOS* R^2 decreases with increasing forecast horizon. Moreover, the results generally confirm the positive effect of model stacking. Our analysis demonstrates that the machine learning ensemble consistently outperforms its individual components across all forecast horizons considered. In contrast, there are instances in which the traditional ensemble exhibits slightly lower performance than the RI and the EP model for specific forecast horizons throughout forecast horizons $t+1$ to $t+3$. Nevertheless, the traditional ensemble consistently outperforms all of its component models for the forecast horizons $t+4$ to $t+5$. This outperformance is particularly strong for $t+5$ forecasts, amounting to 28.30% when comparing it to the second-best performing traditional model.

Furthermore, our findings validate the notion that machine learning approaches surpass traditional linear models in terms of predictive performance. To be more specific, when we compare the *OOS* R^2 of the machine learning models, including the ensemble, against that of the traditional models, including the ensemble, we find that the machine learning models outperform the traditional ones in most cases. In fact, just assessing the best-performing models, i.e., the ensembles, we find that the machine learning ensemble beats its traditional counterpart for every forecast horizon. The difference in *OOS* R^2 between the ensemble models is statistically significant at the 1% level for forecast horizons $t+1$ to $t+3$. Further, while the relative PAFE difference between the ensembles decreases with increasing forecast horizon, the difference in *OOS* R^2 increases

from 14.91% for $t+1$ predictions to 23.40% for $t+5$ predictions, in relative terms.

Table 3.3: Average out-of-sample R^2

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
L^{MSE}	0.3631	0.1932	0.1417	0.0856	0.0262
HVZ^{MSE}	0.4198	0.2875	0.2315	0.1596	0.0901
EP^{MSE}	0.4232	0.2843	0.2285	0.1554	0.0934
RI^{MSE}	0.4326	0.3022	0.2501	0.1713	0.0986
ENTD	0.4220	0.2891	0.2440	0.1807	0.1265
RF^{MSE}	0.4659	0.3147	0.2582	0.1792	0.1293
GBT^{MAE}	0.4417	0.2985	0.2458	0.1673	0.1146
$DART^{MAE}$	0.4760	0.3237	0.2486	0.1653	0.1154
NN^{MAE}	0.4625	0.2857	0.2527	0.1560	0.1222
ENML	0.4849	0.3447	0.2872	0.2100	0.1561
ENML-ENTD	0.0629***	0.0556**	0.0433**	0.0293	0.0296

This table reports the time-series averages of the out-of-sample R^2s (*OOS* R^2s) for all models. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. L is a model with only current earnings as a predictor, HVZ is the model by Hou et al. (2012), EP and RI are the models by Li and Mohanram (2014), ENTD is an equally weighted ensemble of L, HVZ, EP, and RI, RF is a random forest model, GBT and DART are gradient boosted tree models without and with dropout, NN is a neural net, and ENML is an equally weighted ensemble of RF, GBT, DART, and NN. The superscript MAE (MSE) indicates that the respective model is estimated using the mean absolute error (mean squared error) as its loss function. We follow the literature and estimate the traditional models using the MSE. In untabulated results we find that our results are robust to estimating the traditional models using the MAE. We decide on which loss function to report for the ML models depending on which one yields more accurate predictions (as indicated by the price scaled absolute forecast error (PAFE) for the $t+1$ horizon). ***, **, and * denote statistical significance at the 10%, the 5% and the 1% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years. We only test for statistical significance of the difference between the ensemble models (ENML - ENTD).

Figure 3.2 plots the *OOS* R^2 for the ENML and ENTD for each year and for $t+1$ and $t+5$, respectively. Again, it is evident that the machine learning ensemble outperforms the traditional ensemble in the majority of years and for both forecast horizons. Remarkably, the *OOS* R^2 of both ensemble models exhibits a noticeable dip around 2009, particularly pronounced in forecasts for $t+5$. Notably, for $t+5$ forecasts, it takes some years for the *OOS* R^2 to rebound. This result underscores the delayed integration of new information, such as the financial crisis in this case, into longer-term forecasts. Such delayed adaptation is inherent in the rolling window approach we employ.

3.3.3 Implied cost of capital

A common application of earnings forecasts is the derivation of the implied cost of capital (ICC), for which earnings predictions serve as a crucial input. We restrict this analysis to

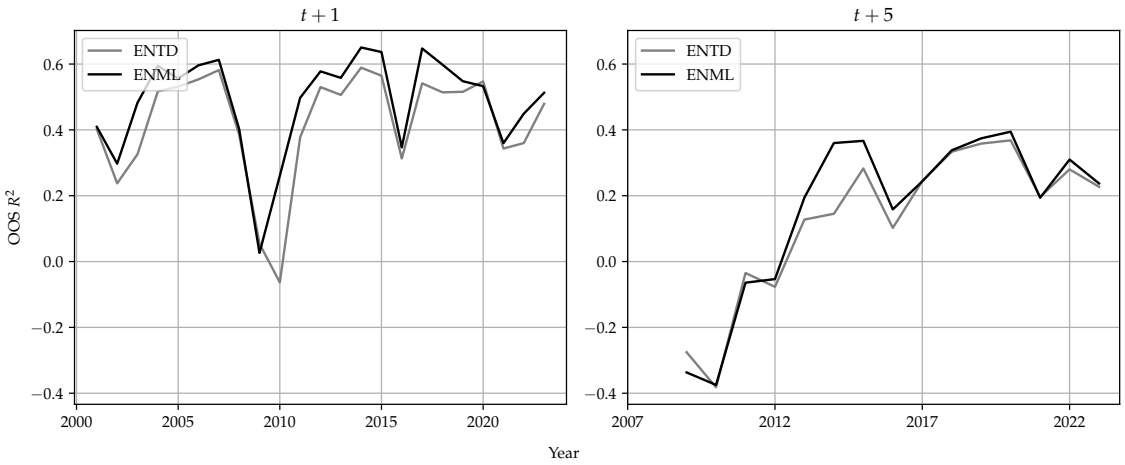


Figure 3.2: R^2 across out-of-sample periods
This figure shows the OOS R^2 of the machine learning ensemble (ENML) and the traditional ensemble (ENTD) per out-of-sample period for forecast horizons $t+1$ and $t+5$.

the best performing traditional earnings forecast model, i.e., the traditional ensemble, as well as the best performing machine learning earnings forecast model, i.e., the machine learning ensemble.

We evaluate the return predictions via ICC as follows. At the end of June in a given year t we estimate the ICC as an average of five commonly used methods.[14] We then sort stocks into quintiles according to the estimated ICC and buy the top quintile.[15] We then evaluate the return from July in t to June in $t+1$ and repeat the process. This long-only strategy corresponds to a theoretically feasible and easily implementable strategy. In Figure 3.3 we show how an initial investment of 1$ would have developed over the out-of-sample periods. We plot the $-portfolio value based on the ENTD-ICC and the ENML-ICC. As an additional benchmark, we also plot the value of a value-weighted market portfolio for the same period.[16]

The plot reveals two key insights. First, (long-only) ICC investment strategies seem to work well in general. Both ICC portfolios outperform the value-weighted benchmark portfolio that we include. Second, the ENML-ICC portfolio outperforms the ENTD-ICC portfolio by a significant margin. More precisely, at the end of the out-of-sample periods, the initial investment of the ENML-ICC portfolio grew to a value of around $ 6, outperforming the ENTD-ICC portfolio by around $ 1.2.

[14] An overview over these methods is provided in Appendix B.4.

[15] Note that this quintile corresponds to the stocks for which the return expectation is the highest based on the ICC.

[16] Value-weighted returns are retrieved from CRSP. More specifically, we annualize the monthly value-weighted market returns including dividends.

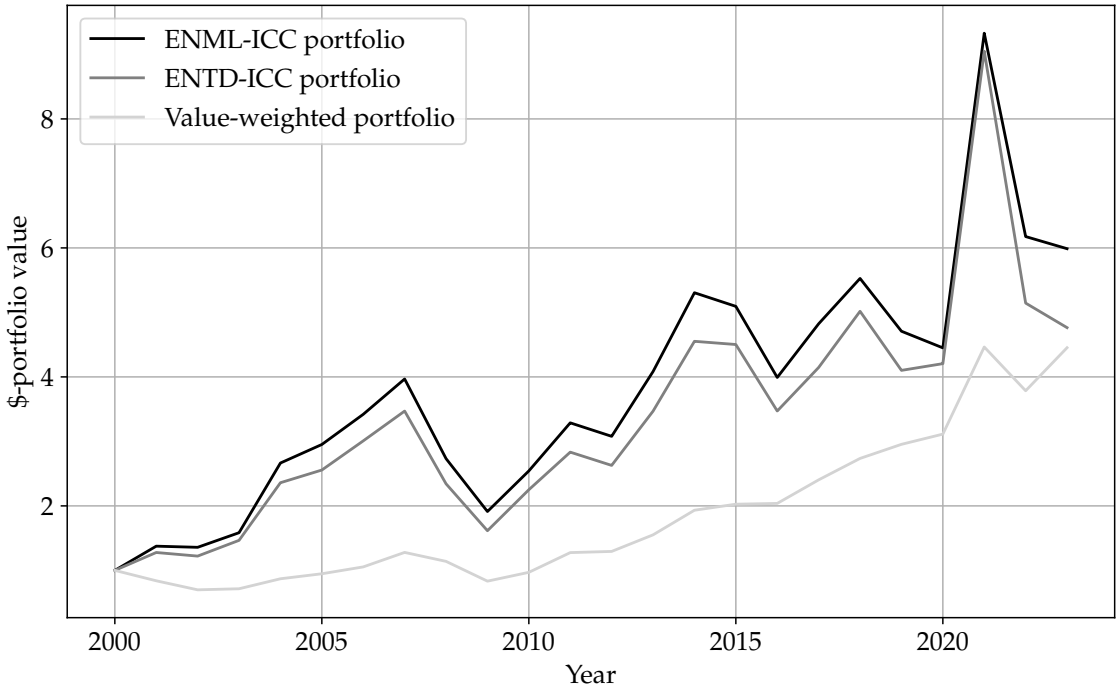


Figure 3.3: Portfolio evaluation
This figure shows the development of an initial investment of 1$ investment across out of sample periods. We assess three strategies. First, we invest based on implied cost of capital (ICC). At the end of June of a respective year t, we sort stocks based on their ICC and buy the top quintile. We then collect the returns from July in t to June in $t+1$ and repeat the process. We do this based on ICC derived via ENML forecasts (ENML-ICC) and based on ICC derived via ENTD forecasts (ENTD-ICC). Second, we invest into the value weighted market portfolio. Monthly value-weighted market portfolio returns are directly derived via CRSP.

We conclude that the improved accuracy of machine learning predictions translates into more profitable investment strategies, thereby stressing the practical importance of earnings prediction accuracy.

3.4 Interpretation

3.4.1 Variable importance

As outlined above, we assess the degree to which financial statement variables matter in the machine learning ensemble by computing their SHAP values.[17] More precisely, we compute SHAP values per out-of-sample period and derive their respective averages for each variable.

Figure 3.4 shows the mean absolute SHAP values of the ENML, averaged over all out-of-sample periods and scaled so that variable importance per forecast horizon sums

[17]We focus on the ensemble model as it is the best performing machine learning approach.

to one. We show the twenty most important variables for predicting earnings in $t+1$ and sort them according to their importance.[18] The higher the SHAP value, the more important the variable. For $t+1$ predictions, *ib*, i.e., current earnings, is the most important variable by far.[19] This comes as no surprise, considering that a simple model including only current earnings as a predictor performs comparably well in predicting future earnings.

A striking finding is that the remaining 19 most important variables are primarily different definitions of earnings. For example, the second most important variable *oiadp* resembles "operating income after depreciation" and the third most important variable *ebit* resembles "earnings before interest and taxes". The only top-twenty variables that do not originate from the income statement are *oancf* and $fopo$.

In general, few variables dominate across forecast horizons $t+1$ to $t+5$. Another finding regarding the different forecast horizons is that the significance of *ib* gradually diminishes with increasing horizon. Instead, *oiadp*, another earnings variable, emerges as the most important variable. Moreover, one of the variables not stemming from the income statement, i.e., *oancf*, becomes increasingly important with increasing forecast horizon.

Lastly, the results suggest that current data is more important than lagged data or first order differences. We revisit this claim below.

We now explore whether the most important predictor variables act as substitutes or complements. This assessment is conducted by examining the absolute Pearson correlation coefficients of the top-twenty variables reported in Figure 3.5. If the variables are substitutes, one would expect high coefficient values. In general, we find mixed results. *pi*, *ibcom*, *ibadj*, *ni*, *ibc* and *niadj* are correlated quite strongly with *ib* and each other. All of these are variations of earnings definitions which do not differ strongly from each other. This observation leads us to consider these variables as substitutes for *ib*, indicating that they do not necessarily possess significant independent predictive power. Other variables, that are not as closely related to *ib*, either because they explicitly exclude major income statement items, such as *oiadp*, or because they are not income statement items at all, such as *oancf*, do not correlate as strongly with earnings. We interpret this as

[18] The variable definitions are provided in Table B.3 in the Appendix.

[19] Note that this is the earnings definition that we use as our target variable. Further note that all of our variables are scaled by common shares outstanding. Thus, strictly speaking, we refer to earnings per share when talking about *ib*.

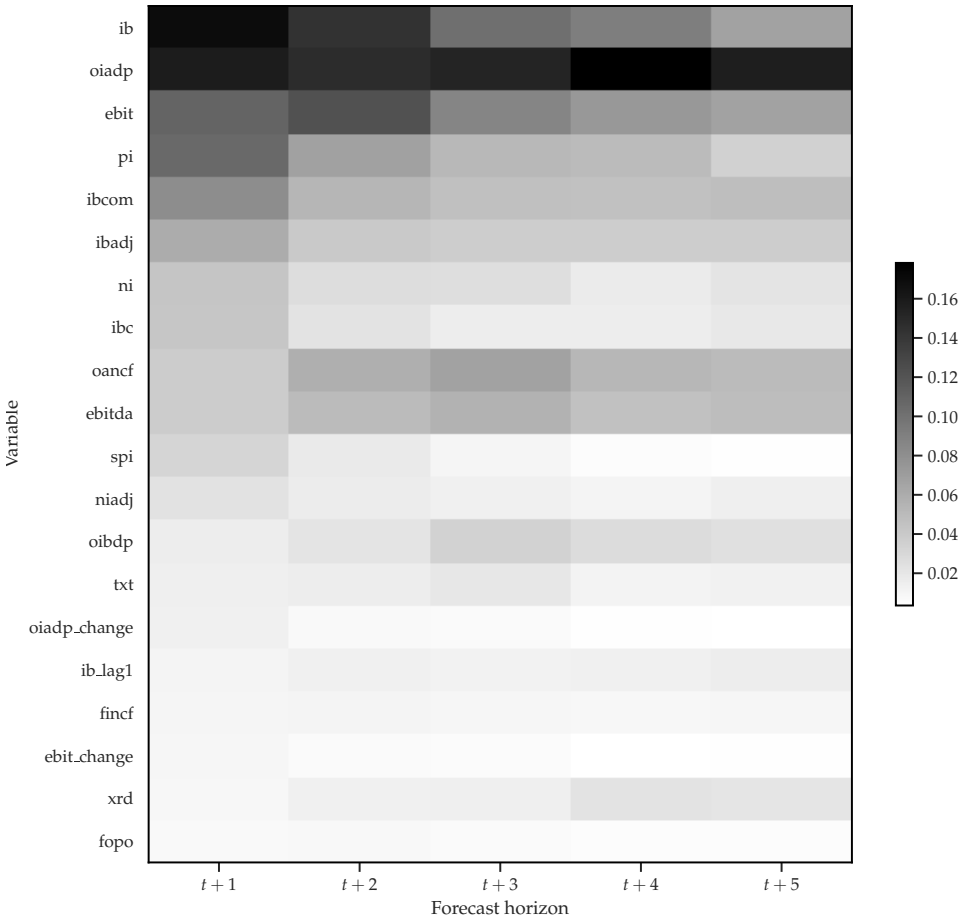


Figure 3.4: Variable importance for the machine learning forecast ensemble
This figure depicts the absolute SHAP values of the 20 most important variables for the machine learning ensemble, averaged over out-of-sample periods and scaled so they sum up to one within each forecast horizon. In this context, importance is defined as the ranking of the respective variable according to the aforementioned metric for forecast horizon $t + 1$.

evidence that these items are complements to *ib* and hence possess stand-alone predictive power.

3.4.2 Group importance

Table 3.4 reports variable importance per group. More precisely, we group the variables according to the financial statement they originate from (Panel A), whether they are current, lagged or change information (Panel B), and according to the two aforementioned categories (Panel C). Grouping the variables according to the financial statement

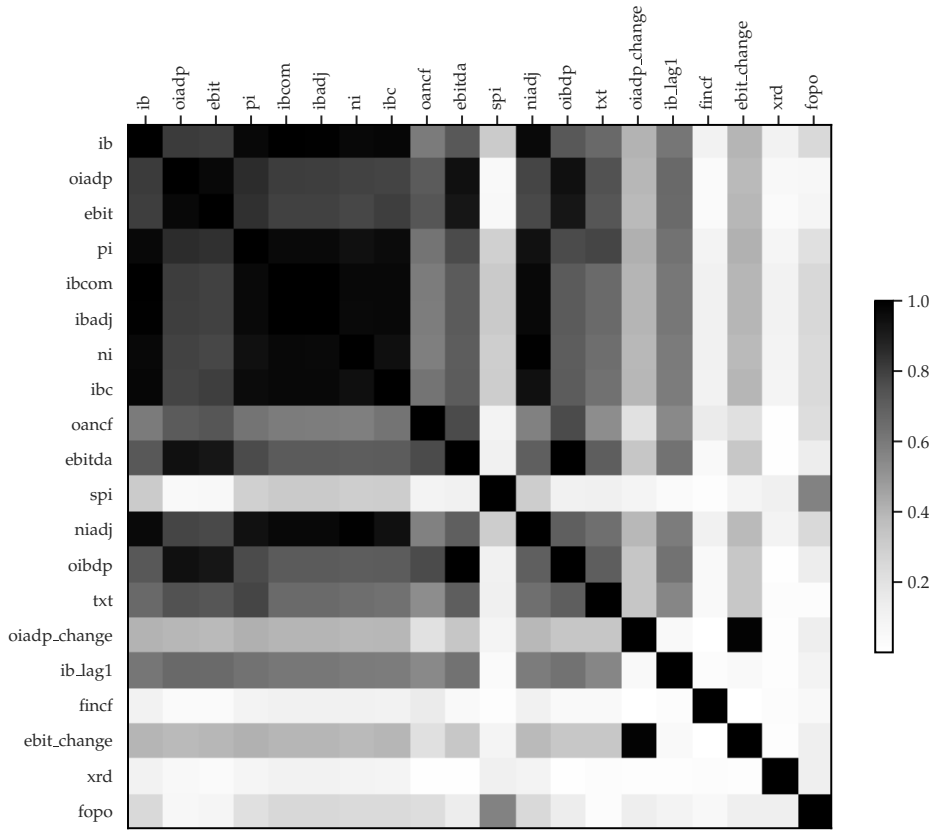


Figure 3.5: Correlation heatmap for the most important variables
This figure shows the absolute Pearson correlation coefficients for the 20 most important variables for the machine learning ensemble.

they originate from reveals that income statement (I/S) variables are the most important variables. On average, for one-year-ahead predictions, I/S variables contribute approximately 67% to the total importance, while balance sheet (B/S) variables and cash flow statement (CF/S) variables contribute around 19% and 15%, respectively. This finding aligns with the analysis of the most important variables, indicating that I/S variables significantly outweigh others in importance for predicting earnings. However, we also find that I/S variables become less important with increasing forecast horizon. In fact, the importance of I/S variables decreases to around 47% for $t + 5$ forecasts. In contrast, B/S variables become more important with increasing forecast horizon (around 37% for

$t+5$ predictions) while CF/S variables stay at a constant level.

Table 3.4: Variable importance by groups

Panel A: Financial statement type

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
B/S	0.1928	0.2599	0.3128	0.3361	0.3719
CF/S	0.1503	0.1609	0.1641	0.1529	0.1543
I/S	0.6569	0.5792	0.5231	0.5110	0.4738

Panel B: Variable type

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Current	0.7352	0.6875	0.6506	0.6492	0.6207
Lagged	0.1252	0.1734	0.2080	0.2167	0.2492
Change	0.1396	0.1391	0.1413	0.1341	0.1301

Panel C: Financial statement type $\times$ variable type

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
B/S current	0.0874	0.1255	0.1563	0.1727	0.1971
B/S lagged	0.0434	0.0677	0.0878	0.0963	0.1069
B/S change	0.0621	0.0668	0.0687	0.0671	0.0679
CF/S current	0.0954	0.0974	0.0924	0.0829	0.0774
CF/S lagged	0.0280	0.0360	0.0430	0.0386	0.0491
CF/S change	0.0269	0.0275	0.0287	0.0314	0.0277
I/S current	0.5525	0.4646	0.4020	0.3936	0.3462
I/S lagged	0.0537	0.0697	0.0772	0.0818	0.0933
I/S change	0.0506	0.0448	0.0439	0.0356	0.0344

Panel A reports the relative variable importance per financial statement group. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement, respectively. The variables are grouped according to Table B.3 in the Appendix. Panel B reports the relative variable importance per variable type. Panel C reports the relative variable importance per financial statement type $\times$ variable type group. Importance per group in each Panel is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Grouping variables according to whether they are current, lagged or difference variables in Panel B reveals that current data is by far the most important group out of these categories, contributing around 74% of total importance for $t+1$ predictions. Lagged and difference data each contribute around $13-14\%$ to total importance for $t+1$ predictions. Again, importance becomes more evenly distributed among the groups with increasing forecast horizon. More precisely, current variables become less important while lagged variables become more important. This might be the case because short-term forecasts are heavily influenced by current information due to their sensitivity to recent developments. Longer-term forecasts benefit from a combination of current and lagged information to capture the interplay of short-term dynamics and longer-term

trends.

Further breaking down the groups according to the two aforementioned categories stresses the findings above. Overall, current I/S variables contribute around 55% to total importance for $t+1$ forecasts and hence represent the most important group of variables by a significant margin. This is intuitive and supports the finding that simple earnings forecasts models only considering current earnings items, like the L model or the EP model, perform comparably well in predicting future earnings.

We conclude our variable importance assessment by more thoroughly analyzing the variable importance per financial statement type in Table 3.5. More precisely, we assess the importance of each financial statement type per schematic financial statement component, such as e.g., current assets, fixed assets or equity, in the B/S case. This analysis provides an intuitive accounting perspective on which components of financial statements are important and how the importance might change across forecast horizons.

The table provides several key insights: first, assessing the B/S, we find that the debt and supplemental items are the most important pieces of B/S information, with both contributing around 5-6% to total importance for $t+1$ forecasts. Moreover, all pieces of B/S information consistently increase in importance with increasing forecast horizon.

Second, variables associated with the operating cash flow resemble the most important category of CF/S variables. This comes as no surprise, since the operating cash flow closely relates to earnings. The relevance of the investing cash flow slightly increases with increasing forecast horizon. However, overall, the differences are minor.

Third, different definitions of earnings, i.e., EBITDA, EBIT, EBT and net income, resemble the most important I/S categories. Out of these categories, EBIT and net income are the most important categories with around 18% and 25% share in total importance, respectively. Interestingly, the importance of EBIT only slightly declines with increasing forecast horizon, whereas the importance of the net income consistently strongly declines with increasing forecast horizon to around 12% for $t+5$ forecasts. This dynamic might be attributable to the fact that net income is more strongly exposed to accounting manipulation than EBIT and hence less reliable in the long-term. Moreover, we find that while sales contribute very little overall, they consistently increase in importance with increasing forecast horizon. This supports the notion that items which are less exposed to discretionary accounting gain predictive value when considering longer forecast horizons. Revisiting the aforementioned finding that operating cash flow

Table 3.5: Variable importance: financial statements

Panel A: Balance sheet

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Current assets	0.0283	0.0410	0.0465	0.0495	0.0595
Fixed assets	0.0199	0.0213	0.0270	0.0287	0.0294
Total assets	0.0037	0.0090	0.0123	0.0159	0.0196
Debt	0.0467	0.0635	0.0743	0.0785	0.0929
Equity	0.0336	0.0438	0.0532	0.0596	0.0605
Total debt & equity	0.0029	0.0052	0.0087	0.0101	0.0122
Supplemental	0.0577	0.0761	0.0908	0.0939	0.0979
Sum B/S importance	0.1928	0.2599	0.3128	0.3361	0.3719

Panel B: Cash flow statement

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Operating cash flow	0.1080	0.1081	0.1084	0.0949	0.0976
Investing cash flow	0.0134	0.0184	0.0208	0.0217	0.0211
Financing cash flow	0.0247	0.0288	0.0292	0.0311	0.0302
Total cash flow	0.0042	0.0057	0.0056	0.0052	0.0054
Sum CF/S importance	0.1503	0.1609	0.1641	0.1529	0.1543

Panel C: Income statement

	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Sale	0.0072	0.0129	0.0169	0.0211	0.0219
Operating expenses	0.0217	0.0310	0.0355	0.0482	0.0470
EBITDA	0.0380	0.0440	0.0517	0.0432	0.0420
Depr. & Amort.	0.0071	0.0076	0.0072	0.0085	0.0108
EBIT	0.1766	0.1625	0.1397	0.1426	0.1240
Interest expenses	0.0542	0.0462	0.0406	0.0328	0.0325
EBT	0.0656	0.0416	0.0311	0.0299	0.0208
Tax expenses	0.0318	0.0395	0.0438	0.0406	0.0455
Net Income	0.2460	0.1818	0.1430	0.1313	0.1174
Dividends	0.0087	0.0120	0.0136	0.0128	0.0118
Sum I/S importance	0.6569	0.5792	0.5231	0.5110	0.4738

This table reports the relative variable importance per financial statement group. B/S, CF/S and I/S denote balance sheet, cash flow statement and income statement. EBITDA denotes earnings before interest, taxes and depreciation and amortization. EBIT denotes earnings before interest and taxes. EBT denotes earnings before taxes. The variables are grouped according to Table B.3 in the Appendix. Importance per financial statement component is defined as the fraction that the respective component contributes to total importance, measured as the sum of absolute SHAP values.

variables maintain consistent importance across forecast horizons further reinforces this notion. Unlike earnings, cash flows include no discretionary accrual items and are hence

not exposed to earnings management (e.g., Jones, 1991). Consequently, their predictive value does not decrease for longer-term forecasts. In summary, these findings suggest that the variations in importance across forecast horizons are primarily driven by the presence of earnings management. Future research endeavors could offer additional insights into these dynamics.

3.4.3 Nonlinearity

We now approximate the degree to which nonlinearity of the functional form and nonlinearity of variables, i.e., interactions among financial statement variables considered, play a role in predicting earnings.

Surrogate model

We find that for our flexible ENML approach, around 94% of the variation in predicted earnings can be explained by the linear surrogate model for $t+1$ predictions on average. This indicates that the ENML predictions can still be approximated quite accurately by a linear model. More so, it indicates that interaction effects as well as nonlinearity of the functional form do not play a major role, on average. In fact, including two-way interactions increases the average adjusted R^2 by around 2 percentage points, on average. The remaining unexplained portion of variance of the ENML predictions is attributable to the nonlinear functional form of the ENML.[20] We depict this graphically in Figure 3.6. The figure shows the adjusted in-sample R^2 per out-of-sample period, derived by regressing predicted ENML earnings on a linear surrogate model and a linear surrogate model including two-way interactions. The figure also includes the surrogate models for predictions for $t+5$. With increasing horizon, a slightly larger portion of the earnings-predictor relation can be attributed to interaction effects and nonlinearity of the functional form. Nonetheless, the linear surrogate model still explains around 90% of the $t+5$ predictions on average.

Partial dependence plots

We now turn to how the aforementioned degree of nonlinearity is expressed at the variable-level. Figure 3.7 shows the partial dependence plots for *ib*, *oiadp*, *ebit* and

[20] Theoretically, it can also be attributed to higher-order interactions and variables which we did not include. However, in undocumented results, we find that this is not the case.

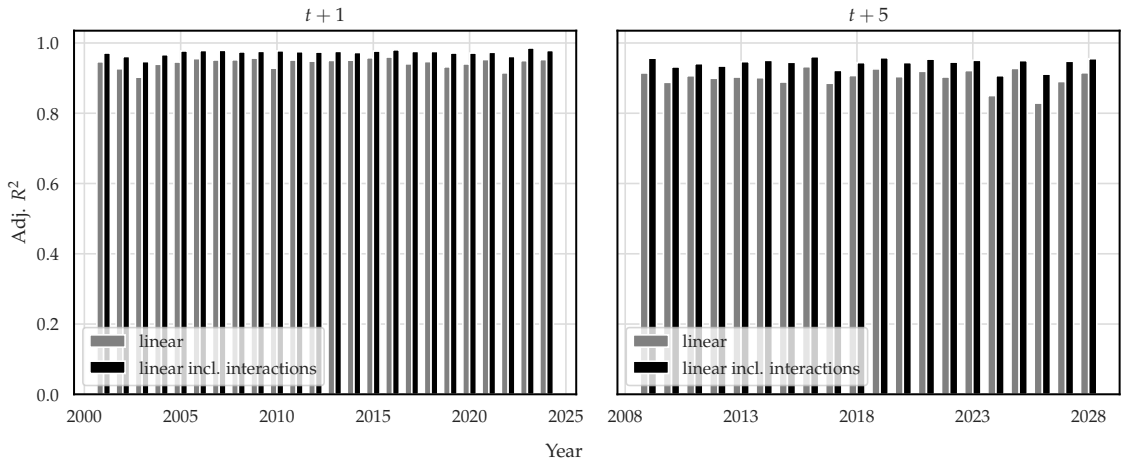


Figure 3.6: Surrogate models
This figure shows the adjusted R^2 of the surrogate models that we fit to our machine learning ensemble predictions for forecast horizons $t+1$ and $t+5$. The linear model (linear) is a simple linear model in which we regress the respective predictions on the 20 most important predictor variables according to their average absolute SHAP values for $t+1$ forecasts. The linear model including interactions (linear incl. interactions) is a linear model in which we use the same set of predictors as well as all possible two-way interactions.

oancf for all forecast horizons.[21] The partial dependence measures the sensitivity of the predicted earnings to the individual financial statement variables.

The upper-left panel shows the effect of *ib* on the model output. Remarkably, the sensitivity appears to be linear for both positive and negative values of *ib*. However, there is a distinction in the slope of the line for positive and negative values, suggesting varying sensitivities of future earnings to current earnings for profit and loss firms, respectively. This may explain why the EP model and the RI model by Li and Mohanram (2014) yield comparably good forecasting results, especially for short forecast horizons. The two models include a dummy for negative earnings and the interaction between earnings and the negative earnings dummy, which essentially allows for different slopes of *ib* for profit and loss firms.

For *ebit* in the lower-left panel we find a similar trend as for *ib*. For *oiadp*, there is also a difference in slopes between profit and loss firms, especially for longer forecast horizons. However, the kink appears slightly below zero. Also, interestingly, for $t+1$ forecasts, the change relationship appears to be concave. In contrast, *oancf* is essentially linearly related to future earnings across all forecast horizons.

[21] We plot the partial dependence of the three most important variables and the most important variable not stemming from the income statement.

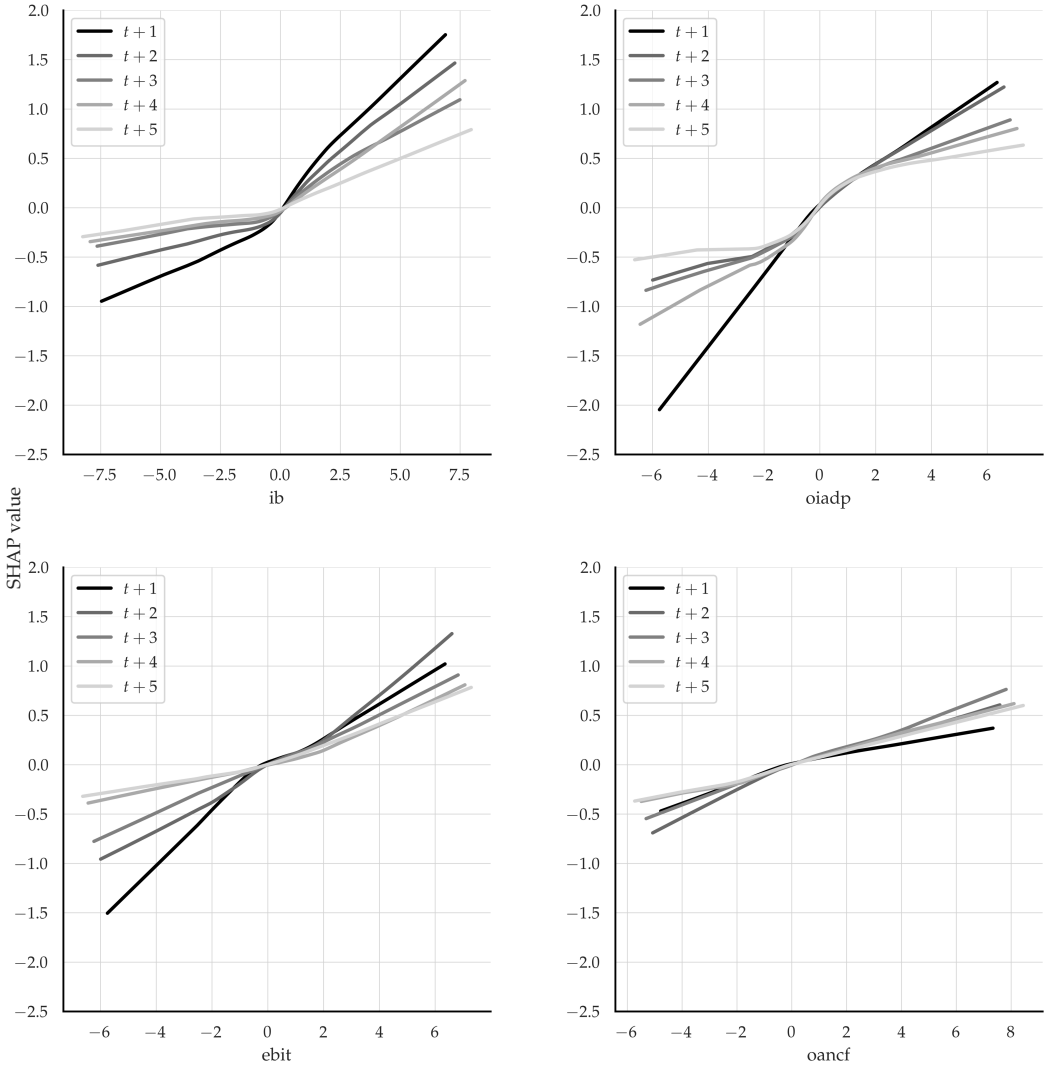


Figure 3.7: Partial dependence plots
The panels show the sensitivity of expected future earnings to the respective variable for all forecast horizons. More specifically, we fit a nonparametric lowess model (locally weighted linear regression) to the SHAP values of the respective variable.

3.4.4 Subsamples

We now turn to the subsample analysis. More precisely, we assess the ENML predictions per life cycle stage, firm size tercile and industry. This analysis allows an assessment of where exactly overall accuracy improvements stem from and whether the relation between current fundamental data and future earnings differs for different subsets of firms. For the sake of comprehensiveness, we restrict this analysis to the two ensemble models and focus on the key evaluation and interpretation metrics, i.e., the median PAFE, variable importance as well as the surrogate models.

Life cycle

First, similar to Easton et al. (2024), we stratify predictions according to the life cycle stage that a firm is in as of the estimation date. Importantly, life cycle classifications are only used for evaluation, not for estimation – meaning that the ML models do not directly "see" which life cycle stage a firm belongs to. Despite this, the ML models produce much better predictions for firms at the beginning and at the end of their life cycle, suggesting that they implicitly detect patterns in the underlying financial statement data that distinguish these firms.

Assessing accuracy differences using the median PAFE reveals interesting patterns. First, ENML outperforms ENTD in every life cycle stage and for every forecast horizon. Second, firms at the beginning of their life cycle (labelled "Intro") and at the end (labelled "Decline") exhibit the lowest overall accuracy by far. However, these are also the stages where ML provides the largest improvement: for $t+1$ forecasts, ENML reduces median PAFE by 0.0110 and 0.0086 for firms in their "Intro" and "Decline" stages, respectively – while improvements for "Growth," "Mature," and "Shake-out" firms range from 0.0022 to 0.0034. These results are visualized in Figure 3.8.[22]

Turning to variable importance, we find no significant differences across life cycle stages. As summarized in Table 3.6, the importance of financial statement information in forecasting future earnings remains consistent across life cycle stages.[23] More precisely, for firms in every life cycle stage, income statement data is the most important group of predictors for $t+1$ forecasts, while balance sheet data becomes more important for longer horizons.

Lastly, we examine the degree to which interaction effects and nonlinearities play a role for different life cycle stages. Table 3.7 reports the results. In general, we find a decrease in the linear surrogate model's R^2 with increasing forecast horizon, indicating that nonlinearities and interactions become more important over time. This decrease is observed for all life cycle stages. Nevertheless, the effects differ strongly across life cycle stages. For "Intro" and "Decline" firms, a simple linear surrogate model without interactions explains only 86-87% of ENML predictions, whereas for "Growth," "Mature," and "Shake-out" firms, it explains 94-95%. However, once we introduce interaction terms, the

[22]The results for all forecast horizons are summarized in Table B.7 in the Appendix.

[23]In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

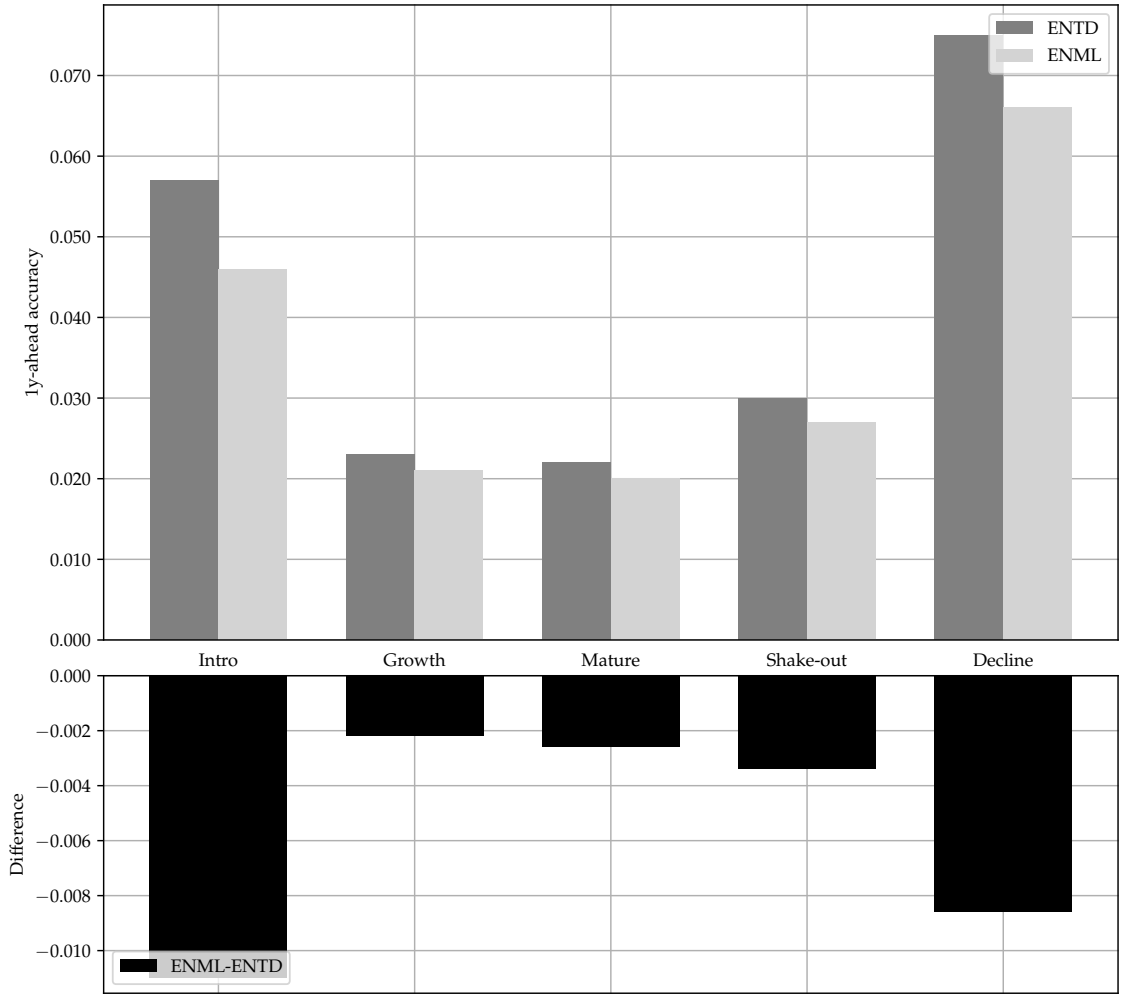


Figure 3.8: Accuracy differences per life cycle stage
This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per life cycle stage for the two ensemble models and for $t + 1$ forecasts. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table B.5 in the Appendix. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

surrogate model explains 97-99% of the ENML predictions for all life cycle stages. This suggests that interaction effects are particularly important for firms at the beginning and end of their life cycle, which aligns with the stronger accuracy improvements observed for these firms. Importantly, these effects are uncovered by the ENML despite the fact that life cycle stages are not provided as an input. Instead, it autonomously captures complex patterns in financial statement data that distinguish these firms and enhance earnings predictability.

Table 3.6: Variable importance by financial statement type per life cycle stage

Firm size	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Intro	ENTD	0.2199	0.2810	0.3330	0.3742	0.4062
	ENML	0.1578	0.1657	0.1538	0.1479	0.1418
	ENML - ENTD	0.6222	0.5532	0.5133	0.4779	0.4520
Growth	ENTD	0.1975	0.2558	0.3107	0.3384	0.3775
	ENML	0.1554	0.1617	0.1559	0.1524	0.1596
	ENML - ENTD	0.6471	0.5825	0.5334	0.5093	0.4629
Mature	ENTD	0.2156	0.2943	0.3468	0.3902	0.4089
	ENML	0.1506	0.1538	0.1534	0.1458	0.1510
	ENML - ENTD	0.6338	0.5519	0.4998	0.4640	0.4401
Shake-out	ENTD	0.2091	0.2718	0.3237	0.3470	0.3773
	ENML	0.1515	0.1604	0.1590	0.1497	0.1420
	ENML - ENTD	0.6394	0.5678	0.5174	0.5033	0.4807
Decline	ENTD	0.2066	0.2844	0.3558	0.3933	0.4378
	ENML	0.1706	0.1757	0.1703	0.1678	0.1605
	ENML - ENTD	0.6228	0.5399	0.4739	0.4388	0.4017

This table reports the relative variable importance per financial statement and per life cycle stage. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table B.5 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S , CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.3 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 3.7: Surrogate R^2 per life cycle stage

Life cycle stage	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Intro	No	0.8682	0.8461	0.8245	0.8226	0.8274
	Yes	0.9811	0.9671	0.9447	0.9565	0.9492
Growth	No	0.9466	0.9205	0.9176	0.9173	0.9138
	Yes	0.9798	0.9683	0.9609	0.9602	0.9606
Mature	No	0.9464	0.9246	0.9243	0.9180	0.9058
	Yes	0.9799	0.9676	0.9612	0.9593	0.9513
Shake-out	No	0.9447	0.9181	0.9144	0.9141	0.9009
	Yes	0.9864	0.9781	0.9680	0.9666	0.9629
Decline	No	0.8604	0.8341	0.8002	0.7979	0.8021
	Yes	0.9711	0.9183	0.8573	0.8952	0.9025

This table reports the average adjusted R^2s of the surrogate models per life cycle stages. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table B.5 in the Appendix. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

Firm size

Next, we stratify predictions according to the size of the firm as of estimation date. In line with the literature, we find that the larger the firms, the higher the prediction accuracy (Li and Mohanram, 2014). Nonetheless, the ENML outperforms the ENTD for each size subsample considered. Further, we find that the accuracy improvements increase with decreasing firm size. More precisely, for $t+1$ forecasts, the ENML yields a median PAFE improvement of 0.0073 for small firms and of 0.0030-0.0033 for medium and large firms. We visualize these findings in Figure 3.9.[24]

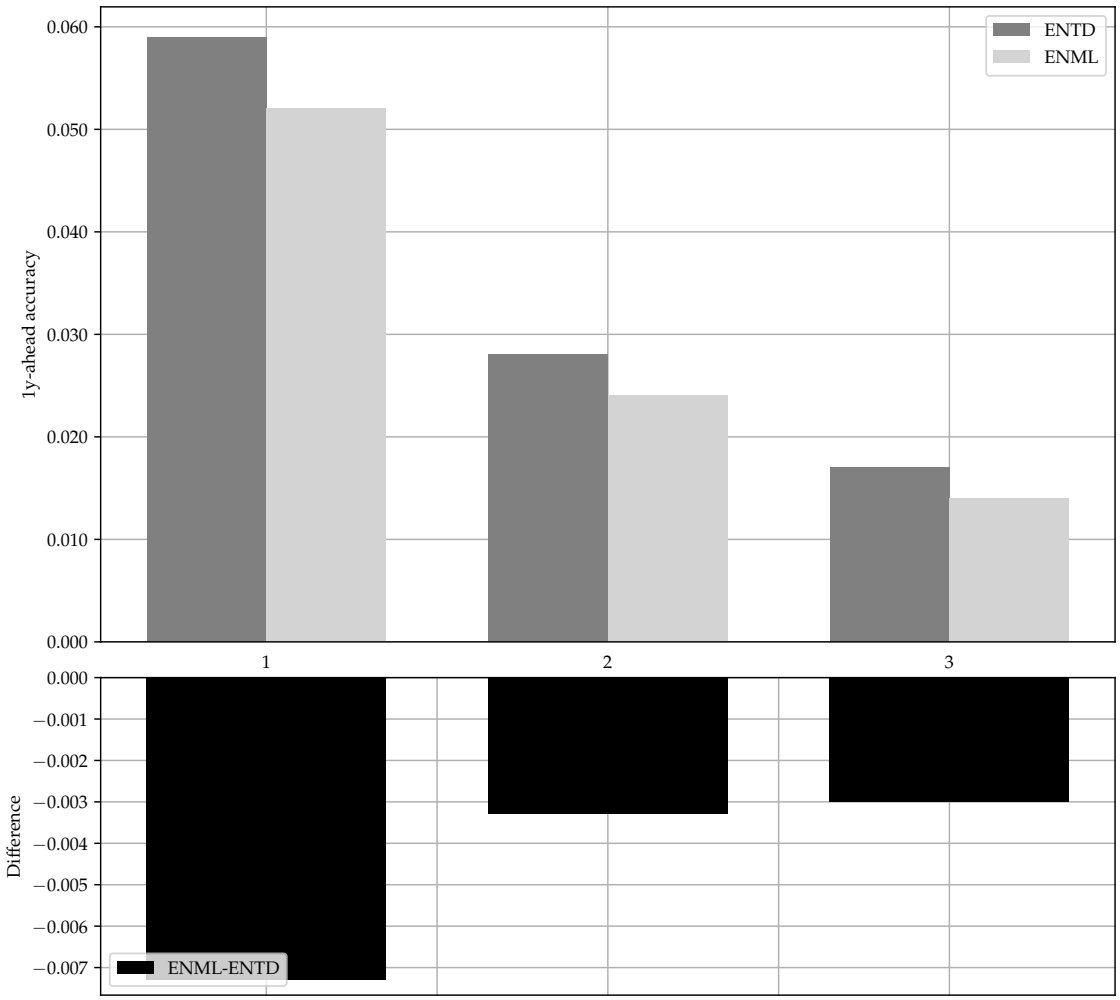


Figure 3.9: Accuracy differences per size tercile
This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per size tercile for the two ensemble models and for $t+1$ forecasts. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

[24]The results for all forecast horizons are summarized in Table B.8 in the Appendix.

In terms of variable importance, we find no significant differences between the findings regarding the full sample and the sample stratified by size. We summarize the results with respect to the financial statements in Table 3.8.[25]

Table 3.8: Variable importance by financial statement type per size tercile

Firm size	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Small	B/S	0.2181	0.3016	0.3508	0.3791	0.4213
	CF/S	0.1703	0.1653	0.1579	0.1614	0.1598
	I/S	0.6115	0.5331	0.4913	0.4595	0.4189
Medium	B/S	0.2015	0.2659	0.3196	0.3333	0.3992
	CF/S	0.1545	0.1588	0.1543	0.1420	0.1421
	I/S	0.6440	0.5753	0.5262	0.5246	0.4587
Large	B/S	0.2151	0.2818	0.3265	0.3528	0.3915
	CF/S	0.1536	0.1493	0.1494	0.1503	0.1426
	I/S	0.6313	0.5690	0.5241	0.4968	0.4659

This table reports the relative variable importance per financial statement and per size tercile. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S , CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.3 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Turning to the surrogate models, we again find a decrease in linear surrogate R^2, i.e., an increasing influence of nonlinearities and interactions in the ENML, with increasing forecast horizon. Moreover, as in the life cycle analysis, the surrogate R^2s match the differences in accuracy improvements across subsamples. The ENML achieves the highest accuracy improvements for the small firm subsample and it is this firm subsample for which adding interactions leads to the largest gains in surrogate R^2. Again, the effect of nonlinearities, which we infer from the portion of the variance in ENML predictions which is not explained by the surrogate model including interactions, is constant across subsamples. The results are summarized in Table 3.9.

Industry

Lastly, we stratify the predictions according to the industry the respective firm is in as of the estimation date. We find that the ENML outperforms the ENTD for nearly all forecast horizons and industries. In fact, the ENML improve the median PAFE by 0.0032-0.0050, depending on the industry. The aforementioned range highlights the

[25] In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

Table 3.9: Surrogate R^2 per size tercile

Firm size	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
Small	No	0.8927	0.8686	0.8489	0.8529	0.8496
	Yes	0.9769	0.9611	0.9478	0.9527	0.9456
Medium	No	0.9361	0.9016	0.8947	0.8925	0.8778
	Yes	0.9763	0.9608	0.9494	0.9472	0.9427
Large	No	0.9155	0.8913	0.8971	0.8910	0.8817
	Yes	0.9755	0.9616	0.9530	0.9510	0.9449

This table reports the average adjusted R^2s of the surrogate models per industry. Size is defined as the market capitalization as of estimation date. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

substantial variations in accuracy enhancements across different industries. The results are visualized in Figure 3.10.[26]

Turning to variable importance, we again find that stratification of the sample does not lead to results that differ from that for the full sample. We report the importance per financial statement in Table 3.10.[27]

Interestingly, the surrogate R^2s do not match the results regarding accuracy as in the two cases above. The surrogate modeling results are reported in Table 3.11. For example, the ENML achieves the highest accuracy gains over the ENTD for firms within industry "4". Yet, the degree to which interaction terms and nonlinearities play a role for the ENML predictions are not the highest for this industry, as indicated by the surrogate R^2s. We conclude that when stratifying by industries, it is primarily the additional variables which we feed to the ENML which lead to differences in predictive performance. Again, the effect of the nonlinear functional form appears to be modest and slightly lower than the effect of interactions, in general.

[26]The results for all forecast horizons are summarized in Table B.9 in the Appendix.

[27]In untabulated results, we find that this is also true for importance on the variable as well as the financial statement component level.

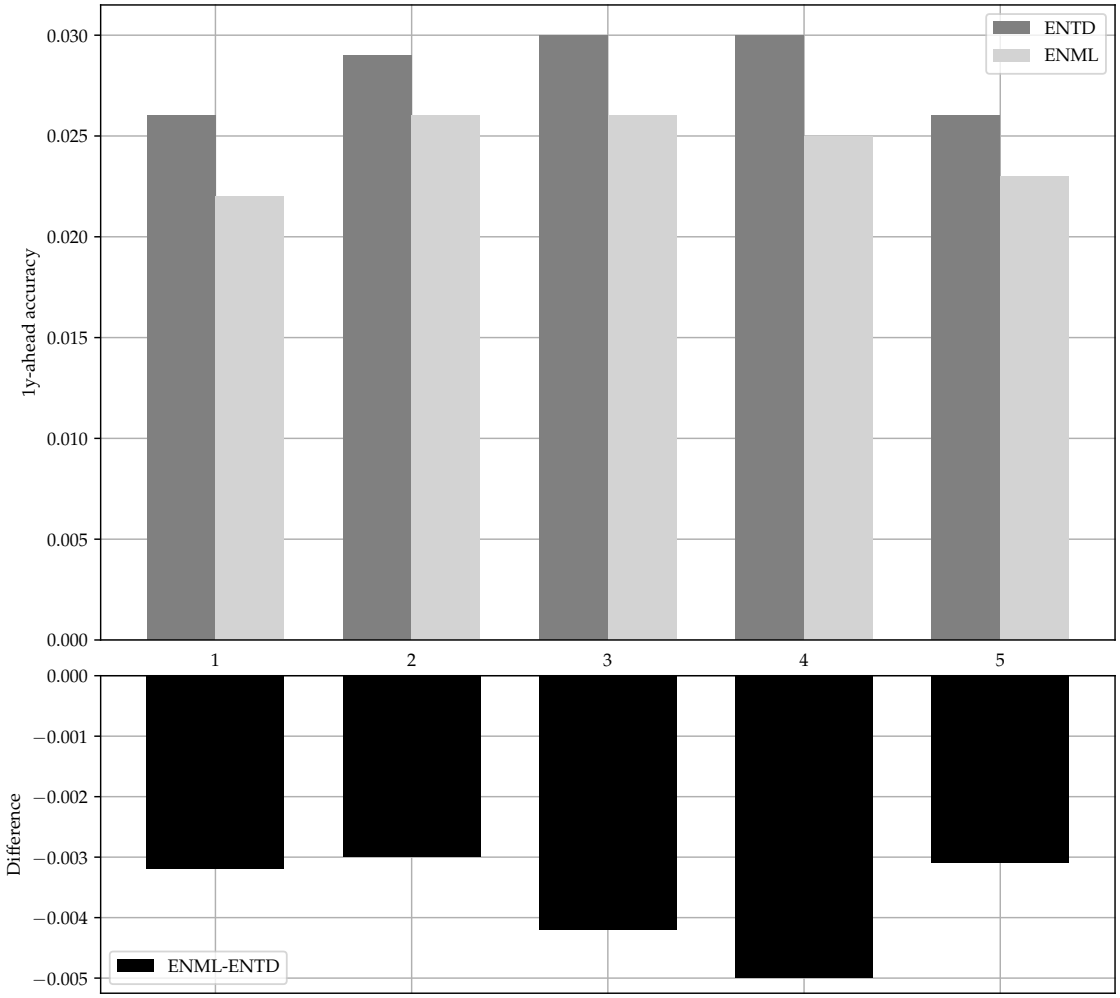


Figure 3.10: Accuracy differences per industry
This figure reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per industry for the two ensemble models and for $t + 1$ forecasts. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`), and as of estimation date. An overview is given in B.6 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year.

Table 3.10: Variable importance by financial statement type per industry

Industry	Financial statement	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
1	B/S	0.2189	0.2922	0.3427	0.3702	0.4115
	CF/S	0.1592	0.1649	0.1511	0.1507	0.1449
	I/S	0.6219	0.5429	0.5062	0.4792	0.4436
2	B/S	0.2070	0.2804	0.3385	0.3568	0.4089
	CF/S	0.1648	0.1736	0.1702	0.1510	0.1461
	I/S	0.6282	0.5460	0.4912	0.4922	0.4450
3	B/S	0.2284	0.2974	0.3441	0.3654	0.4027
	CF/S	0.1535	0.1578	0.1493	0.1363	0.1435
	I/S	0.6181	0.5448	0.5067	0.4983	0.4538
4	B/S	0.2165	0.2870	0.3258	0.3470	0.4020
	CF/S	0.1565	0.1713	0.1583	0.1521	0.1448
	I/S	0.6270	0.5418	0.5159	0.5008	0.4533
5	B/S	0.2132	0.3024	0.3478	0.4010	0.4388
	CF/S	0.1742	0.1911	0.1878	0.1578	0.1484
	I/S	0.6126	0.5065	0.4644	0.4412	0.4127

This table reports the relative variable importance per financial statement and per industry. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`), and as of estimation date. An overview is given in B.6 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. B/S , CF/S and I/S denote balance sheet, cash flow statement and income statement respectively. The variables are grouped according to Table B.3 in the Appendix. Importance per financial statement type is defined as the fraction that the respective group contributes to total importance, measured as the sum of absolute SHAP values.

Table 3.11: Surrogate R^2 per industry

Industry	Interactions	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
1	No	0.9439	0.9226	0.9161	0.9110	0.8984
	Yes	0.9849	0.9759	0.9697	0.9652	0.9632
2	No	0.9484	0.9177	0.9202	0.9180	0.9048
	Yes	0.9803	0.9689	0.9627	0.9594	0.9535
3	No	0.9443	0.9205	0.9107	0.9156	0.9161
	Yes	0.9830	0.9717	0.9600	0.9636	0.9620
4	No	0.9593	0.9479	0.9386	0.9263	0.9209
	Yes	0.9910	0.9823	0.9753	0.9731	0.9757
5	No	0.9362	0.9117	0.9036	0.9035	0.9001
	Yes	0.9792	0.9659	0.9552	0.9556	0.9511

This table reports the average adjusted R^2s of the surrogate models per industry. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`), and as of estimation date. An overview is given in B.6 in the Appendix. The surrogate model without (with) interactions is a linear model which regresses the ENML predictions on the twenty most important variables (plus all possible two-way interactions) as of estimation date. Importance is defined as the fraction that the respective variable contributes to total importance, measured as the sum of absolute SHAP values. For each forecast horizon and industry, we repeat this regression per ENML out-of-sample prediction and report the average adjusted R^2 across surrogates. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings.

3.5 Conclusion

We show that earnings per share predictions based on state-of-the-art machine learning approaches using high-dimensional financial statement data are more accurate than those based on traditional linear approaches. These improvements hold across all evaluation metrics assessed, i.e., commonly used error metrics, the *OOS* R^2 as well as the performance of long-short ICC portfolios based on the predictions.

Importantly, we provide an intuitive breakdown of how important the different pieces of fundamental accounting information are for predicting earnings. We find that current I/S variables, especially current earnings, are the most important predictors. However, with increasing forecast horizon, variable importance becomes more balanced. More precisely, B/S information becomes much more important whereas I/S information becomes less important with increasing forecast horizon. Thoroughly disentangling the different financial statements suggests that this dynamic may be attributable to earnings management.

As the first study to thoroughly decompose the effects of nonlinearity in the earnings prediction context we find that especially for short term-horizons, the relationship as approximated by the best performing machine learning model, i.e., the machine learning ensemble, can still be described by a linear surrogate model to a large extent. More precisely, we find that on average around 94% of the variance in machine learning ensemble predictions can be explained by a linear model, depending on the forecast horizon. Interactions and nonlinearity of the functional form somewhat equally contribute to the remaining small unexplained portion of predictions. As the forecast horizon increases, the linear surrogate R^2 decreases slightly.

We conclude our empirical analysis by stratifying firms into different subsets and assessing differences in accuracy, variable importance and nonlinearity. Overall, the ENML beats the ENTD for every subset of firms considered. However, significant differences in terms of accuracy and improvements of accuracy of the ENML as compared to the ENTD exist across firm subsets. Moreover, the degree to which the ENML exploits interactions across fundamental variables appears to vary across life cycle stages and firm size terciles.

Our findings provide important guidance for future research. First, we show that machine learning approaches are an excellent tool for earnings predictions. We hence

argue that research which uses (model) earnings predictions in some way or another should resort to machine learning methods, if high accuracy is desired. Second, we show which financial statement variables and groups thereof are important. Future research may build upon that when building models and deciding which variables to include. Importantly, this includes the differences in terms of variable importance across forecast horizons. For example, if one is interested in an earnings prediction model including only a small number of variables for computation-related reasons, employing distinct (small) sets of variables for different forecast horizons might be beneficial. Lastly, we show that model-based forecast performance as well as the degree to which complex forecast models are necessary differs between subsets of firms. Additional research that further aims to disentangle these relations and their causes is an interesting avenue for future research and could ultimately lead to further forecast model improvements.

Chapter 4

Hard to Process: Atypical Firms and the Cross-Section of Expected Stock Returns[‡]

4.1 Introduction

Information is fundamental to the functioning of financial markets and the allocation of capital. Standard asset pricing models typically assume that investors rapidly collect and interpret all available information, ensuring that prices fully reflect fundamental value. In practice, however, not all information is equally easy to process. Investors have bounded processing capacities and limited attention, preventing frictionless information processing. To cope, investors do not process every piece of information independently. Instead, they think in patterns and categories (e.g., Barberis and Shleifer, 2003; Peng and Xiong, 2006; Gabaix, 2014; Kacperczyk et al., 2016).

Difficulties arise when a firm presents information in unusual or idiosyncratic ways. I refer to such firms as *atypical firms*, meaning companies whose joint realization of observable information differs from that of most other firms. The key idea is that when a firm's information does not fit familiar patterns, it imposes higher processing difficulty, making its firm-specific information more likely to be neglected or incorporated only gradually. A central question in asset pricing is whether such information-processing

[‡]This chapter is based on Weibels (2025). I thank Tom Zimmermann for his valuable feedback. I also thank seminar participants at the University of Tübingen, University of Cologne, and the 4th International Econometrics PhD Conference for their helpful comments and suggestions.

frictions have return predictive power.

Despite the centrality of information-processing frictions in theory, empirical work has lacked a general-purpose measure of how hard it is to process a firm's information at any given time. This gap is particularly striking given the prominence of information frictions in theoretical models of asset pricing. Models of bounded rationality, gradual information diffusion, and disagreement all rest on the premise that some information is more costly to process than others.[1]

This paper addresses this gap by proposing a new approach to capture information-processing frictions. Rather than treating processing difficulty as a property of disclosures or organizational scope, I define it as a statistical property of how well a firm fits within the joint distribution of characteristics, which represent the information investors use when forming expectations and making valuation decisions. I develop a data-driven measure, ATYP, that captures how atypical a firm appears relative to the systematic patterns that describe most firms. An unsupervised machine learning model learns these patterns across a wide range of firm characteristics, and ATYP quantifies the residual component that the model cannot reconstruct. In other words, it captures the firm-specific component that systematic patterns fail to explain. This approach conceptually connects to the idea that investors rely on categorical and pattern-based reasoning when navigating high-dimensional information environments. Consequently, High-ATYP firms present unusual or idiosyncratic combinations of characteristics that are more difficult and therefore costly for investors to process.

Existing empirical proxies for information-processing and valuation difficulty capture only small slices of this broader concept.[2] Measures such as textual readability, segment counts, or accounting intricacy capture important aspects of information-processing frictions, but they are inherently narrow, infrequent, and tied to reporting choices. More fundamentally, they measure properties of disclosure or organizational scope rather than properties of the underlying economic reality that investors must interpret. ATYP differs in three important ways. First, whereas existing complexity measures focus on the reporting of information, ATYP focuses on the information itself. It therefore captures complexity in the underlying economic structure of the firm rather than in its

[1] Examples include Simon (1955); Miller (1977); Hong and Stein (1999); Peng and Xiong (2006); Van Nieuwerburgh and Veldkamp (2010); Gabaix (2014); Kacperczyk et al. (2016).

[2] Important contributions include Li (2008); Loughran and McDonald (2014, 2024); Cohen and Lou (2012); García and Norli (2012); Barinov et al. (2024); Peterson (2012); Hoitash and Hoitash (2018).

disclosures. Second, it is fully data-driven and independent of managerial reporting choices. Third, it is available at high frequency, over a long sample, and for a broad set of firms, including international stocks. These features make ATYP a flexible tool for studying how processing frictions shape asset prices.

Empirically, I document a large and robust association between ATYP and future returns. Firms with higher ATYP, i.e., those for which information is harder to process, earn significantly lower subsequent returns. More specifically, a decile spread portfolio that sells high-ATYP firms and buys low-ATYP firms earns 1.47% per month when portfolios are equal-weighted and 0.82% when they are value-weighted. These results cannot be explained by standard risk factor models and are significantly stronger among firms with high limits to arbitrage and low investor attention, precisely where hard-to-process information is most likely to be neglected. Therefore, return predictability may be driven by mispricing rather than by compensation for risk.

By establishing ATYP as a predictive state variable, this paper makes four contributions. First, I develop and propose ATYP, a novel, general-purpose measure of processing frictions grounded in the statistical atypicality of a firm's characteristics. Second, I provide extensive validation showing that ATYP is not just a statistical curiosity but captures a meaningful, and previously unmeasured, dimension of information-processing frictions. Third, I document the pricing power of ATYP for the cross-section of stocks, where I find that stocks with higher ATYP earn significantly lower future returns. Fourth, I show that these return patterns are concentrated among firms with higher limits to arbitrage and lower investor attention.

A key contribution of my study is validating that ATYP captures information-processing frictions, rather than repackaging existing measures. I show that ATYP is distinct from and only marginally negatively correlated with existing measures of firm complexity and information-processing frictions. Firms with many segments, long filings, or complex disclosures tend to be large, established companies with strong information environments. Despite their organizational complexity, such firms often exhibit low ATYP because they follow familiar characteristic patterns.

In contrast, ATYP is higher in firms with weak information environments. Consistent with this, we find that it is positively correlated with proxies of information uncertainty. However, ATYP remains conceptually and empirically distinct. Crucially, it has significant incremental explanatory power for outcomes linked to processing frictions. High-ATYP

firms are associated with stronger analyst disagreement, higher return volatility, lower analyst accuracy, and stronger absolute earnings surprises.

Perhaps most compellingly, I document that high-ATYP firms experience slower price discovery following information events. Following earnings announcements, high-ATYP firms exhibit stronger and more persistent post-earnings-announcement drift. I find a similar pattern following industry information shocks according to Cohen and Lou (2012), where I compare how quickly firms respond to news that has already been reflected in the prices of their simpler industry peers. High-ATYP firms adjust more slowly, consistent with investors requiring more time and effort to process information about atypical firms and incorporate it into valuations. This slow adjustment is difficult to reconcile with alternative explanations such as risk or data mining and instead points to information-processing frictions.

Having established that ATYP captures information-processing frictions, I turn to its implications for asset pricing. To the best of my knowledge, I am the first to demonstrate the direct cross-sectional pricing effects of processing difficulty induced by atypical information patterns. Previous studies by Cohen and Lou (2012) on organizational complexity and Zhang (2006) on information uncertainty have documented delayed reactions to information shocks, but they have not systematically explored the unconditional return predictability of information-processing frictions in the cross-section.

Empirically, sorting stocks into deciles by ATYP produces a striking monotonic pattern. Firms with high-ATYP earn significantly lower subsequent returns. The equal-weighted spread between the lowest and highest decile is -1.47% per month with a t-statistic of -4.54, while the value-weighted spread is -0.82% with a t-statistic of -2.50. Importantly, the return patterns remain highly robust after controlling for standard factors. Under the six factor model of Fama and French (2018), the equal-weighted spread continues to deliver significant alphas of -0.84% per month with a t-statistic of -4.52, while the value-weighted results weaken to -0.28%, consistent with the effect being concentrated in smaller firms, where investor attention and arbitrage possibilities are more limited.

Dependent double sorts and Fama-MacBeth regressions show that the ATYP premium remains large and significant after controlling for other commonly studied characteristics, including size, investment, profitability, momentum, reversal, illiquidity, analyst disagreement, and volatility. In fully saturated Fama-MacBeth specifications that include more than a dozen established predictors, ATYP continues to load significantly at -0.72 with a

t-statistic of -3.21. The effect extends to different peer group definitions and holds within different clusters of characteristics, demonstrating that it is not confined to particular information domains. The pricing power extends to international markets, particularly in Europe where the equal-weighted spread is -1.21% per month, and applies robustly to alternative factor models including the q-factor model and behavioral factor frameworks. The results remain robust even when I change the measurement approach, use alternative data sources such as WRDS financial ratios, vary model specifications across 48 different hyperparameter combinations, or modify the sample construction. This pervasive robustness underscores that ATYP is not a repackaging of existing variables but an independent dimension of information frictions in asset pricing.

The negative pricing of ATYP raises a fundamental question: why do firms that are harder to process earn lower returns? The negative sign of the ATYP premium, combined with its concentration in settings where arbitrage and attention are limited, suggests a mispricing interpretation. Two well-established theoretical mechanisms explain why information-processing frictions might generate systematic overpricing followed by slow corrections.

High ATYP increases valuation uncertainty by making the mapping from fundamentals to value more ambiguous. Under such uncertainty, behavioral biases such as overconfidence become stronger (Daniel et al., 1998, 2001), which amplifies heterogeneous interpretations of signals and raises disagreement (Zhang, 2006). Combined with short-sale constraints, this disagreement shifts prices toward optimistic investors and creates overpricing among high-ATYP firms (Miller, 1977). At the same time, limited attention leads investors to focus on systematic patterns rather than fully processing firm-specific information (Peng and Xiong, 2006). This selective processing causes signals about atypical firms to diffuse only gradually across investors, and in the spirit of Hong and Stein (1999), the resulting slow adjustment generates systematic underreaction and more persistent return predictability.

My empirical evidence strongly supports this interpretation. I trace the economic channels of the ATYP premium to settings characterized by high limits to arbitrage and low investor attention. Conditioning on proxies for limits to arbitrage reveals that the ATYP premium is concentrated precisely where arbitrage is most difficult. For instance, among firms with high idiosyncratic volatility, the six-factor adjusted spread is -1.51% per month with a *t*-statistic of -–6.77, compared with an insignificant 0.13%

among low-volatility firms. Parallel evidence arises when sorting firms into quintiles based on investor-attention proxies. The ATYP spread is large and highly significant in low-attention groups but shrinks drastically among firms with higher attention. For example, among small firms the six-factor adjusted spread is -1.54% per month (t = -6.90), while among large firms it is economically negligible at 0.20%.

The asymmetry of returns provides further evidence in favor of the mispricing interpretation. The return spread is driven almost entirely by the underperformance of high-ATYP firms, whereas low-ATYP firms earn only slightly above-average returns. High-ATYP firms deliver six-factor alphas of -0.76% per month in the equal-weighted specification, while low-ATYP firms deliver alphas of only 0.08%. Consistent with this view, double sorts with the mispricing factor of Stambaugh et al. (2015) show that the ATYP premium is concentrated among firms ex-ante classified as prone to mispricing. High-ATYP firms exhibit significantly higher mispricing scores on average, and the ATYP spread is most pronounced in the highest-mispricing quintile, with a six-factor alpha of -0.88% and a *t*-statistic of -3.98. Taken together, these results reinforce the conclusion that the ATYP premium reflects mispricing rather than compensation for risk.

Related literature

My work relates to five strands of literature: (i) on the measurement of information-processing difficulty, where I provide an alternative to disclosure- or segment-based proxies, (ii) on the use of machine learning in finance, (iii) on the literature that catalogs cross-sectional predictors, (iv) on the literature examining the cross-sectional predictability of returns from information-processing frictions, and (v) on limited attention and behavioral biases in asset pricing.

First, my measure relates to other studies on information-processing frictions and firm complexity. The literature focuses on three big classes of proxies. Text-based approaches (e.g., Li, 2008; Loughran and McDonald, 2014, 2024) use readability or linguistic complexity of 10-K filings, finding that harder-to-process disclosures are associated with slower price adjustment, greater analyst forecast dispersion, and higher audit fees. Organizational measures focus on business segments or conglomerate structure (e.g., Cohen and Lou, 2012; Barinov et al., 2024), showing that multi-segment firms react more slowly to industry news and may be mispriced. Geographic dispersion has also been linked to delayed price reactions (e.g., García and Norli, 2012). Finally, accounting-based

proxies emphasize the intricacy of financial reporting, such as the complexity of revenue-recognition rules or the breadth of accounting variables disclosed (e.g., Peterson, 2012; Hoitash and Hoitash, 2018). While these proxies capture important forms of disclosure or organizational scope, they do not necessarily reflect the complexity of a firm's underlying economic profile. A firm can have complex filings yet follow simple economic patterns. In contrast, ATYP is data-driven and captures information-processing difficulty in the underlying economic information that investors must interpret.

Second, recent work applies ML methods to capture high-dimensional and nonlinear relationships in returns. For instance, Gu et al. (2020) show that machine learning algorithms outperform linear models in out-of-sample return prediction, while Gu et al. (2021) develop a conditional asset pricing model based on autoencoder-derived latent factors. Closest to my approach, Bali et al. (2025) use an ensemble of random forest models to construct a measure of machine forecast disagreement (MFD), a proxy for investor disagreement. My study differs in two key respects. First, I introduce ATYP, defined as the reconstruction error from an unsupervised machine learning model trained on firm characteristics, capturing the degree to which a firm deviates from the common, low-dimensional manifold of the market. Second, whereas MFD reflects the dispersion of beliefs after they have formed, ATYP is an ex-ante state variable describing the information environment before beliefs form. By increasing the cost of information-processing, ATYP interacts with investor attention and frictions, setting the stage for disagreement to persist. This upstream positioning in the information-processing channel, coupled with an unsupervised learning design, yields a persistent signal with pricing implications not subsumed by disagreement measures or standard predictors.

Third, a vast literature documents that firm characteristics predict the cross-section of expected returns. Early evidence has motivated large-scale compilations (e.g., Harvey et al., 2016; Hou et al., 2018; Chen and Zimmermann, 2022; Jensen et al., 2023) that catalog hundreds of predictors that investors actually use in creating coherent economic valuations. My contribution is to move beyond marginal predictors by examining how unusual a firm's entire profile of characteristics appears, and to show that such atypicality is priced. This approach treats ATYP as a state variable that spans the joint distribution of characteristics, rather than as another stand-alone anomaly.

Fourth, to the best of my knowledge, only few studies have looked at the direct cross-sectional pricing implications of information-processing difficulty. Cohen and

Lou (2012) look at the pricing effects of information shocks where they compare easy- and hard-to-process firms and find that the latter reacts more slowly to shocks. Yet, they do not explore the direct predictive power of their segment count measure in the cross-section of stocks. Barinov (2020) extend this analysis and finds that segment counts are weakly negatively priced in cross-section. Similarly, Zhang (2006) look at the cross-sectional predictions of several proxies of information uncertainty and also find a weak negative association between these proxies and future stock returns. However, they also focus on delayed price responses to news. ATYP however is a more direct proxy of information-processing frictions and a notably stronger predictor of stock returns. Also, I am the first to explore the pricing power across a variety of settings and specifications.

Fifth, my work builds on theoretical models of limited attention and bounded rationality (e.g., Simon, 1955; Peng and Xiong, 2006; Van Nieuwerburgh and Veldkamp, 2010; Gabaix, 2014; Kacperczyk et al., 2016), which emphasize that investors optimally allocate scarce cognitive resources across information signals. In these frameworks, costly or idiosyncratic information is more likely to be neglected or only partially processed. My study also relates to behavioral theories showing that when the information environment is uncertain or difficult to interpret, investors overweight their own signals and become more overconfident, leading to greater belief dispersion and disagreement (e.g., Daniel et al., 1998, 2001; Zhang, 2006; Peng and Xiong, 2006). Empirically, limited attention slows the incorporation of firm-specific and related-firm information into prices (Hirshleifer and Teoh, 2003; DellaVigna and Pollet, 2009; Cohen and Frazzini, 2008; Engelberg and Parsons, 2011), and information uncertainty amplifies behavioral biases in ways that generate disagreement and return predictability (Zhang, 2006). I contribute to this literature by operationalizing the idea that investors rely on systematic patterns and neglect difficult-to-process firm-specific information. More specifically, I show that atypicality in the joint characteristic space systematically interacts with attention and arbitrage capacity.

The remainder of the paper proceeds as follows. Section 4.2 introduces the construction of ATYP. Section 4.3 describes the data and model specifications. Section 4.4 validates ATYP as a measure of information-processing difficulty. Section 4.5 presents the main empirical asset pricing results. Section 4.6 analyzes the economic channels through which ATYP affects prices. Section 4.7 runs a series of robustness checks. Section 4.8 concludes.

4.2 Constructing a measure of firm atypicality

My starting point is the idea that financial markets are populated by investors with limited attention and bounded information-processing capacity. In such an environment, firms differ not only in their characteristics, but also in how easy or hard it is to process them. A stock whose firm characteristics align with familiar valuation patterns is easier for investors to interpret than one whose characteristics combine in an atypical or unfamiliar way.

For instance, a company may be long established and liquid, yet simultaneously display sharply negative profitability, weak sales and asset growth, high leverage, and persistently negative cash flows. It may invest heavily in research and development, but without corresponding improvements in earnings or operating performance. Such a mix of stable features and distressed fundamentals, combined with volatile performance indicators and repeated negative surprises, does not align with the systematic patterns investors typically rely on. Profiles of this kind are inherently harder to process through standard heuristics, which illustrates the type of joint atypicality I aim to measure.

The challenge is to operationalize this idea in a way that remains faithful to its joint nature yet empirically tractable in large samples. To this end, I draw on machine learning methods for modeling high-dimensional data. I view the sample of firm characteristics observed over time as a cloud of points in a high-dimensional space, where most firms cluster around typical patterns in their characteristics. To capture these patterns, I employ an autoencoder, an unsupervised neural network that learns to compress firm characteristics into a compact latent representation and then reconstruct them. The model learns the systematic patterns that characterize a typical firm, and ATYP reflects how well these patterns explain a given firm's profile. The greater the model's difficulty in explaining a firm's characteristics, the more atypical, and thus harder to process, the firm's profile is.

To estimate these latent patterns and assess how well they explain each firm's characteristics, I formalize the intuition using an autoencoder framework. Let i index firms and t index months in my sample. For each firm-month observation, let $\mathbf{X}_{i,t} \in \mathbb{R}^p$ denote the p-dimensional vector of standardized firm characteristics. An autoencoder consists of two parameterized functions: an encoder $f_\theta : \mathbb{R}^p \to \mathbb{R}^k$ with $k \ll p$, and a decoder $g_\phi : \mathbb{R}^k \to \mathbb{R}^p$. The encoder maps the high-dimensional characteristic vector into a

k-dimensional latent representation,

$$\mathbf{z}_{i,t} = f_\theta(\mathbf{X}_{i,t}), \tag{4.1}$$

which summarizes the key patterns in the data. The decoder maps this latent vector z back into the original p-dimensional space to produce a reconstruction,

$$\hat{\mathbf{X}}_{i,t} = g_\phi(\mathbf{z}_{i,t}) = g_\phi(f_\theta(\mathbf{X}_{i,t})). \tag{4.2}$$

The parameters θ and ϕ are estimated jointly by solving

$$\min_{\theta,\phi} \sum_t \sum_i \|\mathbf{X}_{i,t} - g_\phi(f_\theta(\mathbf{X}_{i,t}))\|_2^2, \tag{4.3}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. This training criterion minimizes the sum of squared reconstruction errors across all observations. This forces the encoder to learn the systematic, low-dimensional patterns of the data and the decoder to accurately map these back into the original space.[3]

The reconstruction $\hat{\mathbf{X}}_{i,t}$ can be interpreted as the projection of the firm's characteristics onto the learned patterns of typical configurations. The reconstruction residual

$$\mathbf{R}_{i,t} = \mathbf{X}_{i,t} - \hat{\mathbf{X}}_{i,t} \tag{4.4}$$

is the residual part of the firm's characteristic vector that is not explained by these patterns. If the autoencoder captures the main nonlinear relationships among characteristics, $\mathbf{R}_{i,t}$ represents the firm-specific deviation from what is typical.

A key question is how to collapse this residual vector into a single scalar measure of atypicality. I define ATYP as the root mean squared reconstruction error (RMSE):

$$ATYP_{i,t} = \sqrt{\frac{1}{p}\sum_{j=1}^{p}\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)}\right)^2}. \tag{4.5}$$

Intuitively, ATYP reflects the extent to which a firm's characteristics are not captured by the patterns learned by the autoencoder. A low score implies that the firm's characteristics conform to typical co-movements, whereas a high score signals an unusual or

[3]Figure C.4 in the Appendix provides a graphical illustration of an autoencoder model.

idiosyncratic combinations. Because the autoencoder is trained to reproduce systematic patterns, it will assign large errors to combinations it has rarely seen.[4] Additionally, because I cross-sectionally standardize all input features to zero mean and unit variance, each component of the error vector $\mathbf{R}_{i,t}$ is expressed in units of one standard deviation.[5] Therefore, ATYP quantifies the average standardized deviation per feature between an observed firm and its reconstructed counterpart. This makes the measure directly interpretable in the same units as the input data and consistent with standard RMSE conventions in statistics.

Economically, ATYP captures the rarity of a firm's configuration of characteristics in a way that is inherently joint and context-dependent. Put differently, it measures the amount of non-redundant information an investor would need to process to fully understand the firm. This interpretation is directly linked to the bounded processing capacities. If processing costs increase with ATYP, then more attention is needed to fully incorporate firm-specific details into valuations. To further build intuition for the concept of processing difficulty, Figure C.5 in the Appendix visualizes the reconstruction accuracy of the autoencoder across all firms. The scatterplot contrasts realized characteristic values with their reconstructed counterparts. Firms that lie close to the diagonal represent typical firms whose characteristic patterns are well captured by the model. In contrast, atypical firms deviate more strongly from the diagonal, indicating that their joint configuration of characteristics cannot be easily reproduced by the systematic patterns learned by the model.

My motivation for employing a nonlinear framework rests on the idea that systematic combinations of firm characteristics are inherently nonlinear. The economic meaning of a given variable often depends on the configuration of others, in ways that cannot be captured by additive linear models. For example, a high book-to-market ratio signals financial distress when combined with low profitability and weak investment, but may reflect a traditional value firm when paired with strong profitability and conservative investment (Fama and French, 1992, 2015). Similarly, different industries may present different systematic patterns. Ignoring such interactions and conditional effects means that firms with superficially similar attributes may be treated alike in linear models, even though they imply very different valuation profiles once context is taken into account.

To complement this, Stein (2009) documents that investors are becoming increasingly

[4] I consider alternative measurements of ATYP in Section 4.7.8.
[5] See Section 4.3.1 for more information on feature preprocessing.

sophisticated. In my context, this means that they can understand nonlinear relationships between characteristics and factor them into their valuations. In fact, Li and Rossi (2021) show that the performance of mutual funds is nonlinearly related to the characteristics of the firms in which they invest. This demonstrates that sophisticated investors select stocks based on nonlinear combinations of characteristics.

So, while linear dimensionality-reduction methods such as PCA (e.g., Connor and Korajczyk, 1986; Kelly et al., 2019; Lettau and Pelger, 2020) provide a powerful way to summarize common variation across firms, they are inherently restricted to linear patterns. Autoencoders build on this idea. A linear autoencoder with mean squared error loss collapses to PCA, but by introducing nonlinear encoder and decoder functions, the autoencoder generalizes PCA by capturing nonlinear and heterogeneous interactions (Hinton and Salakhutdinov, 2006). This flexibility allows the model to better capture context-dependent patterns in firms.

Autoencoders are widely applied in anomaly detection, from identifying fraud to detecting irregular physiological signals in medicine. These successes highlight their suitability for identifying firms with unusual financial profiles that standard linear methods fail to flag. Recent work in finance shows that nonlinear ML methods capture economically meaningful interactions missed by linear approaches (Gu et al., 2020; Bryzgalova et al., 2023), and that replacing PCA with nonlinear architectures improves factor extraction (Gu et al., 2021). In my setting, the autoencoder provides a natural foundation for defining ATYP, measuring how far a firm's joint configuration of accounting and market variables lies from the manifold of typical firms.

Finally, my design is deliberately unsupervised. The construct I seek is information-processing frictions, not a new return forecaster. By separating the construction of ATYP from future returns, I ensure that any subsequent pricing power reflects the market's reaction to information-processing frictions, rather than mechanical overfitting to realized returns. Pricing tests then ask whether markets efficiently price these frictions, in the same spirit that Bali et al. (2025) first model beliefs and only subsequently study their pricing implications.

4.3 Data and model

4.3.1 Data

I use the dataset from Jensen et al. (2023), a publicly available dataset of stock returns and characteristics. The underlying monthly return data are from the Center for Research in Security Prices (CRSP) and accounting information from Compustat. I restrict my sample on stocks trading at the NYSE, AMEX and NASDAQ. I also exclude financial and utility firms.[6] Finally, to reduce the effect of very small and illiquid stocks, I exclude low-priced stocks trading below $1 per share. My sample covers the period from 1971 to 2023 of monthly data.

These characteristics are natural building blocks for my proposed measure because they summarize the economic fundamentals and market signals that investors are known to process when forming expectations (e.g., Harvey et al., 2016; Hou et al., 2018; Chen and Zimmermann, 2022; Jensen et al., 2023). Using these characteristics as inputs ensures that my measure is grounded in variables with demonstrated relevance for both investors and asset pricing models, rather than statistical artifacts.[7]

For firm characteristics, I start with the full set of 153 characteristics. Jensen et al. (2023) group these characteristics into 13 clusters. My information set drops characteristics related to *Seasonality*, *Short-Term Reversal* and *Low Risk*.[8] I exclude short-horizon return-based risk measures (e.g., idiosyncratic volatility, lottery proxies, and short-term reversal) from the construction of the score to keep the measure focused on fundamentals and valuations. This choice keeps the economic interpretation of the score as information-processing difficulty of the firm's information set rather than ex-post return risk. Thus, in the baseline specification the input set comprises accounting and market-ratio variables spanning the clusters *Investment*, *Value*, *Profitability*, *Profit Growth*, *Accruals*, *Quality*, *Low Leverage*, *Momentum*, *Debt Issuance* and *Size*. The 117 characteristics make up my complete information set $\mathbf{X}$. Lastly, I require each observation to have at least 50% non-missing features.

I follow standard cross-sectional preprocessing of characteristics. First, I impute

[6] More specifically, I exclude financial firms (with one-digit SIC = 6) and utility firms (with two-digit SIC = 49).

[7] As a robustness check, I replicate my analysis using financial ratios from WRDS, which provide a more granular but less curated set of accounting-based predictors, in Section 4.7.5.

[8] For robustness I also replicate my results with the full information set in Section 4.7.5.

missing values with the median as suggested by Chen and McCoy (2024). I then winsorize all characteristics at the 1% level for both tails to mitigate the effect of outliers and data errors. Finally, I standardize all features to have mean zero and unit standard deviation.

4.3.2 Control variables

To rigorously isolate the role of my main variable of interest in the cross-sectional return predictability framework, I include a comprehensive set of control variables that are widely recognized in the asset pricing literature for their ability to explain the cross-section of expected stock returns. These variables are constructed following established methodologies to ensure consistency with prior work and robustness of inference. The choice of variables follows partly Bali et al. (2025). All controls are taken from either Jensen et al. (2023) or, if not available there, from Chen and Zimmermann (2022).

The CAPM Beta (*BETA*) is calculated as the coefficient of a 60-month rolling window regression of monthly stock returns minus the risk-free rate on market return minus the risk-free rate as in Fama and MacBeth (1973). Firm size (*SIZE*) is defined as market capitalization at the end of month $t-1$, computed as the stock price times shares outstanding. The book-to-market ratio (*BM*) equals the book value of equity from the most recent fiscal year divided by market capitalization, both constructed following Fama and French (2008). To capture return dynamics, I include short-term reversal (*STR*) as the stock's return in month $t-1$ and momentum (*MOM*) measured as the cumulative return from months $t-12$ to $t-2$, consistent with Jegadeesh and Titman (1993). Profitability (*OP*) is operating profits over book equity, as in Fama and French (2015), and asset growth (*AG*) is the annual percentage change in total assets, following Cooper et al. (2008). Liquidity is proxied by turnover (*TURN*), measured as share volume over the past 126 trading days divided by shares outstanding as in Datar et al. (1998), and illiquidity (*ILLIQ*), the average ratio of absolute daily return to dollar volume over the same window, following Amihud (2002). Idiosyncratic volatility (*IVOL*) is the standard deviation of residuals from a daily market model regression over the past year, following Ali et al. (2003). Earnings surprises are captured via standardized unexpected earnings (*SUE*), computed as the earnings surprise scaled by the standard deviation of forecast errors, in line with Foster et al. (1984). To proxy for lottery-like payoffs, I include *MAX*, the average of the five highest daily returns in month $t-1$, as in Bali et al. (2011). Firm age (*AGE*) is

the age of a firm in month, following Jiang et al. (2005). To control for disagreement, I use the forecast dispersion (*AFD*) measure of Diether et al. (2002) defined as the standard deviation of analysts' earnings estimates scaled by the mean earnings estimate. Lastly, I measure attention as the number of analysts following a firm (#*ANA*) as in Elgers et al. (2001).

For all variables I impute missing values with the cross-sectional median. Then, all variables are cross-sectionally winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

Table 4.1 shows the descriptive statistics for excess returns in $t+1$ and the specified control variables. Descriptives are calculated in each cross-section and displayed as the time-series average.

Table 4.1: Descriptive statistics

	Mean	Sd	10th	Q1	Q2	Q3	90th
Ret_{t+1}	0.01	0.16	-0.15	-0.07	0.00	0.07	0.16
BETA	1.21	0.61	0.50	0.86	1.14	1.47	2.02
SIZE	2.69	7.94	0.03	0.09	0.35	1.44	5.58
BM	0.69	0.62	0.15	0.29	0.52	0.88	1.41
STR	0.01	0.14	-0.14	-0.07	0.00	0.08	0.17
MOM	0.14	0.54	-0.40	-0.18	0.05	0.33	0.75
OP	0.10	0.52	-0.27	0.04	0.18	0.29	0.44
AG	0.29	0.85	-0.15	-0.03	0.08	0.25	0.74
TURN	0.64	0.79	0.10	0.22	0.43	0.75	1.31
ILLIQ	2.94	9.39	0.00	0.04	0.18	1.14	6.40
IVOL	0.03	0.02	0.02	0.02	0.03	0.04	0.06
SUE	-0.11	1.61	-1.70	-0.57	-0.01	0.53	1.53
MAX	0.07	0.06	0.03	0.04	0.06	0.09	0.14
AGE	2.39	1.92	0.62	1.00	1.77	3.12	5.11
AFD	-0.13	0.29	-0.24	-0.08	-0.06	-0.04	-0.02
#ANA	5.79	6.46	0.23	1.40	3.58	7.58	14.94

The table reports the descriptive statistics for the cross-sectional variables. Statistics are calculated in each cross-section and displayed as the time-series average. The variables are defined in Section 4.3.2. The mean, standard deviation (Sd), 10th percentile (10th), first up to third quartil (Q1-Q3), and the 90th percentile (90th) are shown. The sample is from 1981 to 2023.

4.3.3 Autoencoder model

My goal is to measure how atypical a firm's configuration of characteristics is relative to the cross-section at a point in time. I operationalize atypicality with an autoencoder trained only on lagged characteristics and evaluated out of sample month by month. The model learns the low-dimensional patterns of typical co-movements, and the reconstruc-

tion residual reflects how well those patterns reproduce each firm's characteristics. I keep the design simple, regularized, and fully rolling to avoid look-ahead.

The autoencoder is trained on a rolling basis using the past 10 years of data. More specifically, at the end of month $t-1$ I construct ATYP for month t to $t+12$ using a model estimated only on the past 120 months $[t-120,\ldots,t-1]$. I re-estimate the model every 12 months so that the learned patterns can adapt over time. This is to ease the computational burden and is in line with other approaches in return prediction (e.g., Gu et al., 2020; Chen and McCoy, 2024). Therefore, I conduct out-of-sample tests for the period 1981 to 2023.

I implement a symmetric, fully connected autoencoder with ReLU activations in all hidden layers and a linear output layer. I begin with three hidden layers arranged in a pyramid structure, as described in Gu et al. (2020), such that the number of units is halved with each hidden layer. The bottleneck dimension in the baseline specification is set to $k = 16$, which corresponds to roughly 10–20% of the 117 input characteristics. Economically, larger k values absorb firm-specific noise and attenuate the cross-sectional variation in information-processing difficulty, while smaller k values blur typical structure and inflate noise. The selected k thus corresponds to the minimal latent dimensionality capturing the typical patterns of firm characteristics.[9]

This choice is further motivated by evidence that a relatively small number of latent factors is sufficient to organize the cross-section of stock returns. For instance, Barillas and Shanken (2018) show that most anomalies can be captured by a small set of benchmark factors, while Lettau and Pelger (2020) demonstrate that latent factor models extracted from a large panel of characteristics achieve high explanatory power with only a handful of factors. My results are robust to alternative bottleneck dimensions in the range $\{12, 16, 20\}$. The network is trained to minimize mean squared reconstruction error using the Adam optimizer with a fixed learning rate of 0.001. To guard against overfitting, I apply L_2 regularization and dropout before the first hidden layer. For my baseline specification, I select an L_2 penalty of 0.0001 and a dropout probability of 0.1.

Accounting data are noisy and asynchronously updated. To make the statistic robust, I use a denoising variant of the objective specified in Equation (4.3). More specifically, for

[9] I provide direct evidence for this choice by analyzing under- and overfitting in the next Section C.1.

each feature I add a small noise term, so that

$$\mathbf{X} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma), \tag{4.6}$$

which is known to help the model learn the true underlying patterns in the data and to filter idiosyncratic measurement noise. This choice echoes the concern in Bali et al. (2025) that prediction difficulty due to statistical noise can masquerade as information-processing difficulty, which they address by isolating estimation uncertainty. My denoising step plays the analogous role on the characteristics side. For my baseline I start with $\mathcal{N}(0, 0.1)$.[10]

To ensure that ATYP captures genuine processing difficulty, I perform a series of validation exercises on the autoencoder fit detailed in Section C.1 in the Appendix. First, I analyze the trade-off between model flexibility and parsimony, confirming that my choice of bottleneck dimension balances the need to capture nonlinear patterns while avoiding the reproduction of noise. Second, I demonstrate that the network effectively filters out systematic economic patterns. While raw characteristics exhibit strong co-movements, the reconstruction residuals are virtually uncorrelated, indicating that ATYP isolates idiosyncratic firm-specific departures rather than common firm characteristics. Finally, I examine the economic drivers of the measure and find that conditional characteristics, such as *Investment* and *Profit Growth*, are the most difficult to reconstruct and most influential for the latent representation, whereas unconditional attributes like *Size* are easily captured. This confirms that ATYP reflects the difficulty of processing complex, context-dependent economic profiles rather than simply acting as a proxy for firm size or obscure reporting.

4.3.4 Descriptive statistics of ATYP

Figure 4.1 provides a first look at ATYP. The left panel plots its distribution across all firm-month observations. The measure is strictly nonnegative, highly right-skewed, and concentrated near zero. Most firms are well represented by the common patterns, while a nontrivial fraction exhibit unusually high ATYP values, reflecting atypical combinations of fundamentals. The dashed line marks the mean of 0.24, indicating that the bulk of

[10]The selection of suitable parameters may influence the results and therefore merits careful consideration. However, in Section 4.7.6, I perform a series of sensitivity analyses that suggest that my results are robust to a variety of hyperparameter selections.

firms cluster tightly around low values with a long right tail of hard-to-process firms.

To place ATYP in a broader context, the right panel plots the 12-month rolling mean of ATYP from 1981 to 2023. The measure varies meaningfully over time, rising during well-known periods of market stress such as the early 2000s tech bubble and the 2008 financial crisis, and declining in calmer periods. This pattern illustrates that ATYP is not a static characteristic but rather a relative measure of how atypical a firm's features are compared to the most other firms at a given time. The underlying patterns evolve in response to macroeconomic conditions, technological advancements, and regulatory shifts that shape the economic landscape. For instance, book leverage or inventory measures may be central in one decade while intangibles or R&D expenditures may dominate in another. As characteristic relevance shifts, what was once typical becomes atypical and vice versa. Therefore, the time variation in ATYP partly reflects the natural evolution of the economy and the accounting landscape, reinforcing the notion that processing difficulty is an inherently dynamic construct.

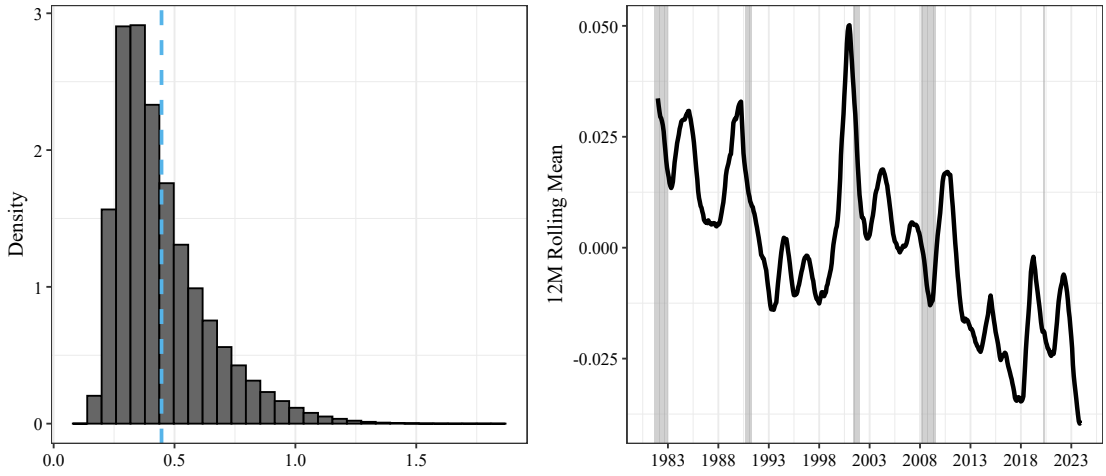


Figure 4.1: Distribution and time-series of ATYP
The left panel shows the cross-sectional distribution of ATYP across all firm-month observations. The y-axis shows density values. The dashed line indicates the mean. The right panel plots the 12-month rolling mean of ATYP, with shaded areas denoting NBER recessions. The sample is from 1981 to 2023.

Next, I examine which firms exhibit high ATYP values. To this end, Table 4.2 reports cross-sectional correlations of ATYP with one-month-ahead returns and standard firm characteristics.[11] The first row shows that the average correlation between ATYP and future excess returns is negative. Additionally, ATYP is moderately positively correlated

[11]The correlations reported in my tables and figures are calculated using Pearson's correlation coefficient.

with idiosyncratic volatility (0.49), and maximum daily return (0.32), and negatively correlated with size (-0.14) as well as firm age (-0.25). Taken together, these patterns indicate that high-ATYP firms tend to be small, volatile, lottery-like. At the same time, correlations with book-to-market, 12–1 momentum, and short-term reversal are weak, suggesting that ATYP does not simply replicate standard anomaly variables.

These correlations are economically consistent with the interpretation of ATYP as a fundamental source of information uncertainty. Firms with atypical characteristics are inherently more difficult to model because they deviate from the standard patterns investors use to interpret economic signals. Because high ATYP firms do not fit these patterns, investors neglect the firm specific details. This neglect increases information uncertainty. This model uncertainty naturally manifests as higher idiosyncratic volatility because market participants struggle to agree on a precise valuation for firms with ambiguous profiles.

Furthermore the results suggest that high ATYP firms operate in a weaker information environment. The negative association with analyst coverage suggests that information intermediaries allocate fewer resources to firms that are costlier to process and harder to explain to clients. Furthermore the link to younger and smaller firms aligns with the business lifecycle where early stage companies have not yet converged to the stable characteristic configurations of mature incumbents. Consequently, ATYP identifies the statistical root cause of the ambiguity found in high uncertainty environments.

These patterns align with prior evidence that investors face greater difficulty when processing these types of stocks (e.g., Hong et al., 2000; Baker and Wurgler, 2006; Bali et al., 2011). Importantly, the dispersion across percentiles shows that these correlations vary across the cross-section but remain economically modest, reinforcing that ATYP captures a distinct dimension of information-processing frictions rather than duplicating known anomalies.

To further characterize the types of firms driving the effect, I examine average firm characteristics across ATYP deciles. Again, high-ATYP firms are smaller, less profitable, more illiquid, and exhibit higher idiosyncratic volatility and lottery-like return profiles. For completeness, these univariate characteristic sorts are reported in Table C.1 in the Appendix.

Table 4.2: Cross-sectional correlations to ATYP

	Mean	Sd	10th	Q1	Q2	Q3	90th
Ret_{t+1}	-0.04	0.09	-0.16	-0.10	-0.04	0.03	0.08
BETA	0.13	0.08	0.04	0.07	0.13	0.17	0.23
SIZE	-0.14	0.03	-0.17	-0.16	-0.15	-0.13	-0.11
BM	0.00	0.09	-0.11	-0.08	-0.03	0.07	0.12
STR	-0.03	0.11	-0.17	-0.10	-0.03	0.03	0.09
MOM	-0.05	0.12	-0.21	-0.13	-0.05	0.02	0.11
OP	-0.37	0.04	-0.41	-0.40	-0.38	-0.35	-0.31
AG	0.28	0.10	0.16	0.24	0.31	0.35	0.38
TURN	0.14	0.12	0.00	0.05	0.12	0.21	0.31
ILLIQ	0.13	0.04	0.08	0.10	0.12	0.15	0.18
IVOL	0.49	0.09	0.37	0.44	0.50	0.56	0.60
SUE	-0.02	0.05	-0.08	-0.06	-0.02	0.01	0.04
MAX	0.32	0.06	0.24	0.28	0.31	0.36	0.39
AGE	-0.25	0.04	-0.30	-0.27	-0.24	-0.22	-0.20
AFD	0.07	0.03	0.03	0.05	0.06	0.09	0.12
#ANA	-0.23	0.06	-0.29	-0.28	-0.25	-0.19	-0.14

The table reports summary statistics on the cross-sectional correlations of multiple variables with ATYP. Correlation is measured using Pearson's correlation coefficient. Statistics are calculated in the time-series of correlations. The variables are defined in Section 4.3.2. The mean, standard deviation (Sd), 10th percentile (10th), first up to third quartil (Q1-Q3), and the 90th percentile (90th) are shown. The sample is from 1981 to 2023.

4.4 ATYP as a measure of information-processing difficulty

In this section, I empirically validate ATYP as an ex-ante measure of information-processing difficulty. I begin by examining the relationship with some of the most notable existing proxies of processing difficulty. I further demonstrate that ATYP predicts future manifestations of processing difficulty. Finally, I document that firms with higher ATYP exhibit slower price discovery.

4.4.1 Relation to other proxies of information-processing difficulty

A central requirement for any proposed proxy of information-processing difficulty is that it captures a distinct underlying dimension rather than rebranding existing measures. To examine this, I systematically compare ATYP with four widely used proxies of processing difficulty or firm complexity: the textual complexity of 10-K filings, the number of business segments, the number of geographic segments, and the accounting reporting complexity (Cohen and Lou, 2012; García and Norli, 2012; Hoitash and Hoitash, 2018; Loughran and McDonald, 2024). Each of these measures reflects a different conceptual foundation. Textual complexity captures the readability and

syntactic difficulty of disclosure, which proxies for the cognitive burden of parsing firm narratives. Segment counts capture organizational scope, which approximates the breadth of activities and the heterogeneity of cash-flow sources that investors must process. Accounting reporting complexity captures disclosure of more accounting items, which proxies for processing difficulty because it requires greater knowledge of authoritative accounting standards. Additionally, processing more items requires more resources. These proxies have proven empirically useful but remain confined to specific reporting dimensions and depend heavily on managerial disclosure choices. In contrast, ATYP is an outcome-based construct derived from firm characteristics rather than disclosures or organizational form. It measures the atypicality of a firm's information set and therefore embodies a statistical rather than representational notion of complexity.

To assess the empirical relation between these measures, I compute cross-sectional correlations between ATYP and each other proxy in every month and summarize the distribution of these correlations in Figure 4.2. The correlations are modest in magnitude and frequently negative. The mean correlation between ATYP and textual complexity is close to zero, while the correlations with business and geographic segment counts are slightly negative on average. These patterns show that information-processing difficulty, as captured by ATYP, is not merely a reformulation of disclosure or diversification breadth. Instead, it reflects a qualitatively distinct information-processing friction.

The weak negative association is theoretically coherent. Large, diversified firms tend to exhibit high segment counts or lengthy disclosures, which inflate traditional measures of complexity. Yet their fundamentals typically follow well-understood patterns with stable characteristics that the model reconstructs easily, resulting in low ATYP scores. Hence, they have a strong information environment. Conversely, some firms that appear simple by conventional standards, because they have few segments or concise, readable filings, can exhibit unusual combinations of firm characteristics. The distinction highlights the difference between organizational complexity and information atypicality: the former concerns how many activities a firm undertakes, whereas the latter concerns how difficult its economic profile is to interpret given the prevailing structure of information in the market.

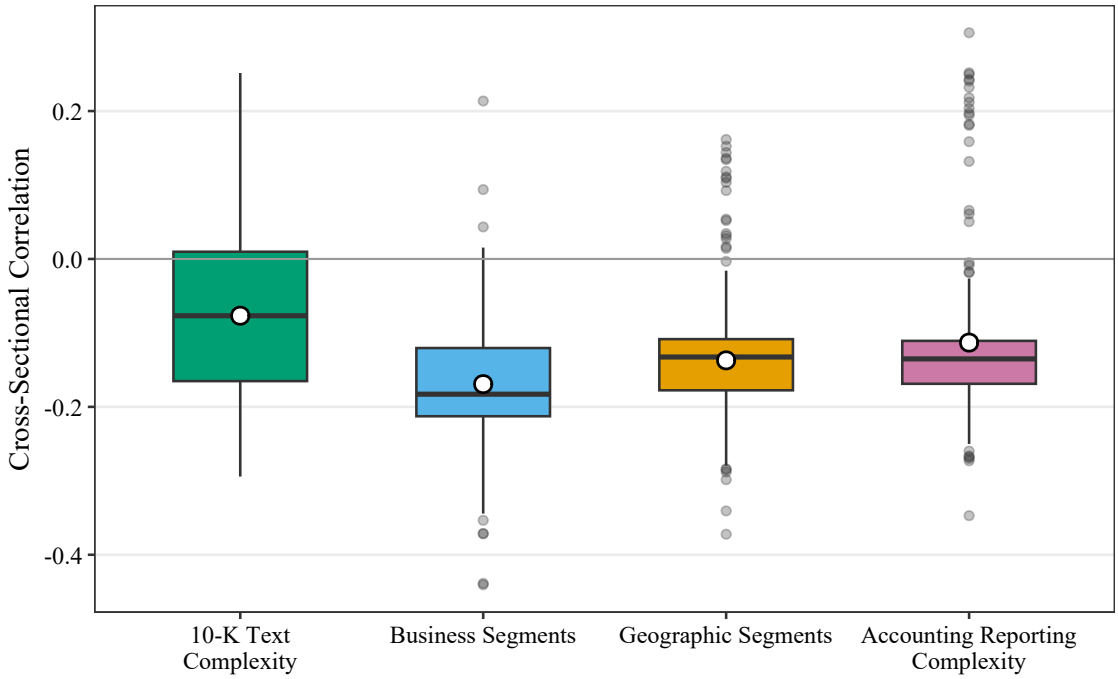


Figure 4.2: Correlation of ATYP with established proxies of information-processing difficulty The figure plots the distribution of cross-sectional correlations between ATYP and established proxies of information-processing difficulty. I consider four measures: 10-K text complexity, the number of business segments, the number of geographic segments, and accounting reporting complexity. For each month, I compute the Pearson correlation between ATYP and each proxy across firms. The boxplot summarizes the distribution of these correlations over time. The white circle marks the mean correlation for each proxy.

4.4.2 Incremental explanatory power

I next examine whether ATYP captures dimensions of information-processing difficulty that manifest in observable market outcomes following Loughran and McDonald (2014), Loughran and McDonald (2024) and Bali et al. (2025). Prior research establishes that processing difficulty is associated with greater disagreement among investors, lower forecast accuracy, higher sensitivity of prices to news, and larger deviations between realized and expected earnings. For example, Diether et al. (2002) and Johnson (2004) show that forecast dispersion rises with information uncertainty. Pástor and Veronesi (2003) and Brogaard and Detzel (2015) link uncertainty and disagreement to elevated return volatility. DellaVigna and Pollet (2009) and Hirshleifer et al. (2009) document that processing difficulty and limited attention increase the likelihood of earnings surprises. Building on these insights, I investigate whether ATYP systematically predicts forecast dispersion, forecast accuracy, return volatility, and absolute earnings surprises. To test these associations, I regress each outcome in $t + 1$ on ATYP, controlling for the standard

set of firm characteristics and including firm and month fixed effects. Standard errors are double-clustered by firm and month.

Table 4.3 reports the results. Consistent with the information-processing interpretation, ATYP is strongly and positively related to analyst forecast dispersion and subsequent return volatility. A one-unit increase in ATYP predicts a 0.161 rise in forecast dispersion and a 0.017 rise in return volatility in the following month. ATYP also loads positively and significantly on absolute earnings surprises, indicating that firms with high ATYP are more likely to deliver outcomes that deviate sharply from expectations. For forecast accuracy, I document a strong negative association, as analysts make more errors for high-ATYP firms. This strengthens the argument that ATYP is an ex-ante state variable that describes the information environment before beliefs are formed. By raising the cost of processing information, ATYP paves the way for persistent disagreement and mispricing. Note, that these effects are highly significant after controlling for known determinants, suggesting that ATYP provides incremental explanatory power beyond established predictors.

Table 4.3: Explaining forecast dispersion, forecast accuracy, return volatility, and abs. earnings surprise

	Dispersion	Accuracy	Volatility	Abs. SUE
Model:	(1)	(2)	(3)	(4)
Variables				
ATYP	0.087***	-1.49***	0.005***	1.01***
	(0.009)	(0.376)	(0.0002)	(0.058)
Controls	Yes	Yes	Yes	Yes
Fixed-effects				
Month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,139,347	1,282,279	1,747,739	1,599,628
R^2	0.24139	0.56263	0.55949	0.07316

The table reports panel regressions of forecast dispersion, forecast accuracy, return volatility, and absolute standardized earnings surprise (SUE) in t+1 on ATYP. Specifications include firm and month fixed effects, with controls. The controls considered are defined in Section 4.3.2 excluding AFD and SUE. Standard errors are double-clustered by firm and month. *, **, and *** denote significance at the 10%, 5%, and 1% levels. The sample is from 1981 to 2023.

4.4.3 Post-earnings-announcement drift

If ATYP captures an information-processing friction that impedes timely price adjustment, its pricing effects should be especially pronounced around major information events. The literature on post-earnings-announcement drift (PEAD) documents that stock prices underreact to earnings news, generating predictable returns in subsequent months (e.g., Bernard and Thomas, 1989; Engelberg et al., 2018). PEAD is commonly interpreted as evidence of limited attention and gradual information diffusion (Hirshleifer and Teoh, 2003; Hirshleifer et al., 2009). If high-ATYP firms are more difficult to process, investors may require more time to incorporate earnings signals into prices, producing stronger and more persistent drift (Barinov et al., 2024).

I test this prediction using both bivariate portfolio sorts and Fama–MacBeth cross-sectional regressions. Table 4.4 reports dependent double sorts in which firms are first grouped into ATYP quintiles and then into earnings-surprise quintiles. I then compute the H–L earnings-surprise return spread over different horizons. Panel A shows that in month $t+1$, the spread is 0.27% ($t = 3.23$) among low-ATYP firms but rises to 0.89% ($t = 6.30$) among high-ATYP firms. Panels B and C extend the horizon to months $t+2$ and $t+3$, respectively. In both cases, the spreads remain large and significant, with the strongest effects concentrated in the high-ATYP quintile. This persistence indicates that earnings signals are incorporated more gradually for atypical firms.

To corroborate, Table C.2 in the Appendix presents Fama–MacBeth regressions of post-announcement returns on earnings surprise, ATYP, and their interaction, controlling for standard firm characteristics. Across horizons $t+1$ and $t+2$, the interaction term is positive and significant. These regressions confirm the portfolio evidence in a continuous setting with controls.

4.4.4 Industry information diffusion

To examine whether the price discovery implications of ATYP extend beyond firm-specific earnings announcements, I also study lead-lag effects of industry information shocks following the approach of Hou (2007) and Cohen and Lou (2012). In their setting, a common industry shock affects two sets of firms: one that processes information straightforwardly, and another that requires more complex analysis to incorporate the same information into prices. In their study, they use the number of firm segments to

Table 4.4: ATYP and post-earnings-announcement drift

	Low	2	3	4	High	H-L	t-stat	FF6	t-stat
Panel A: Bivariate Sort on ATYP — $t+1$									
Low	0.77	0.92	0.90	1.03	1.04	0.27***	3.23	0.24***	2.99
2	0.71	0.88	0.77	1.11	1.14	0.42***	5.44	0.34***	4.33
3	0.50	0.85	0.55	0.96	1.07	0.57***	4.77	0.38***	4.05
4	0.19	0.57	0.16	0.86	0.88	0.69***	4.90	0.47***	3.89
High	-0.43	-0.18	-0.71	-0.16	0.46	0.89***	6.30	0.71***	5.06
H-L	-1.20	-1.10	-1.62	-1.19	-0.58	0.62***	4.36	0.47***	3.27
Panel B: Bivariate Sort on ATYP — $t+2$									
Low	0.88	0.95	0.97	1.01	1.06	0.18**	2.31	0.17**	1.99
2	0.71	0.89	0.76	0.97	1.10	0.39***	4.37	0.40***	3.87
3	0.60	0.81	0.58	1.04	1.05	0.45***	4.47	0.44***	3.90
4	0.26	0.61	0.17	0.72	0.92	0.66***	5.22	0.64***	5.03
High	-0.42	-0.06	-0.75	-0.16	0.36	0.79***	5.55	0.74***	4.80
H-L	-1.31	-1.02	-1.72	-1.17	-0.69	0.61***	4.27	0.57***	3.70
Panel C: Bivariate Sort on ATYP — $t+3$									
Low	0.88	1.00	0.93	1.01	1.03	0.15*	1.73	0.14	1.44
2	0.73	0.86	0.81	0.99	1.02	0.29***	3.64	0.30***	3.33
3	0.62	0.86	0.44	0.94	1.00	0.38***	3.49	0.35***	2.83
4	0.33	0.66	0.13	0.62	0.84	0.51***	4.20	0.36***	3.04
High	-0.36	-0.00	-0.83	-0.24	0.32	0.67***	4.56	0.59***	3.84
H-L	-1.24	-1.01	-1.76	-1.25	-0.71	0.53***	3.41	0.44***	2.68

The table reports bivariate portfolio sorts examining the relation between ATYP and earnings surprises, linking post-earnings-announcement drift to information-processing difficulty. Panels A, B, and C show dependent double sorts, where firms are first partitioned into quintiles by ATYP and subsequently into quintiles by earnings surprises within each quintile. Panel A shows returns in month $t+1$, Panel B reports cumulative returns over months $t+1$ to $t+2$, and Panel C extends to months $t+1$ to $t+3$. Reported are high-minus-low return spreads (H–L), Newey and West (1987) adjusted t-statistics in parentheses, and Fama and French (2018) six-factor (FF6) adjusted alphas. *, **, and *** denote significance at the 10%, 5%, and 1% levels. The sample is from 1981 to 2023.

characterize hard-to-process firms.

Analogously, I test whether ATYP influences firms' response to industry-wide signals using a two-step approach. First, I construct industry shocks as the average prior-month return of the simplest firms (lowest ATYP quintile) within each 2-digit SIC industry. I then examine how the most atypical firms (highest ATYP quintile) within the same industries respond to these shocks. If atypical firms process information more slowly, their prices should adjust more gradually to signals already reflected in their simpler peers. To test this, I sort high-ATYP firms into decile portfolios based on the industry shock and evaluate return patterns.

Table 4.5 shows that my findings mirror those of Cohen and Lou (2012). I find that high-ATYP firms exhibit a strong and predictable delayed response to information that

has already been incorporated into the prices of their low-ATYP industry peers. In the equal-weighted specification in Panel A the portfolio of high-ATYP firms linked to the most negative industry shocks (Decile 1) generates a statistically significant six factor alpha of -1.54% ($t = -6.50$) while the spread between the positive and negative shock portfolios yields a monthly alpha of 1.62% ($t = 5.11$). The return pattern is monotonic confirming that information-processing difficulty impedes the timely incorporation of industry wide news. These results hold firmly in value-weighted portfolios in Panel B where the high-minus-low spread remains economically large at 1.10% per month ($t = 2.47$) after controlling for the six factor model. Collectively these estimates provide compelling evidence that ATYP captures a dimension of complexity that slows price discovery causing high-ATYP firms to lag behind the fundamental information revealed by their simpler peers.

4.5 Atypical firms and the cross-section of expected returns

In this section, I conduct different tests to assess the predictive power of ATYP over future stock returns. First, I show univariate and bivariate portfolio sorts. Second, I present firm-level Fama-MacBeth cross-sectional regression results. Third, I look at the persistence of the predictive power.

4.5.1 Univariate portfolio sorts

I begin my empirical analysis with univariate portfolio sorts. At the end of each month from 1981 to 2023, I sort all stocks into deciles according to their ATYP. Within each decile, I compute the one-month-ahead excess return, measured relative to the one-month Treasury bill rate, for both equal-weighted and value-weighted portfolios. To capture the return differential associated with information-processing difficulty, I form a long-short (H–L) portfolio that takes a long position in the highest-ATYP decile and a short position in the lowest-ATYP decile. This H–L spread provides a direct measure of the pricing impact of ATYP.

In addition to excess returns, I report alphas relative to standard factor models. Specifically, I estimate alphas under the capital asset pricing model (CAPM), the Fama and French (1993) three-factor model (FF3: market, size, and value), the Fama and French (2015) five-factor model (FF5: market, size, value, profitability, and investment), and the

Table 4.5: Univariate portfolio sorts of high-ATYP firms on industry information shocks

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	-0.91**	(-2.07)	-1.84***	(-6.54)	-1.87***	(-7.35)	-1.75***	(-6.64)	-1.54***	(-6.50)
2	-0.62	(-1.32)	-1.57***	(-5.26)	-1.49***	(-6.06)	-1.14***	(-4.08)	-0.94***	(-3.58)
3	-0.30	(-0.71)	-1.29***	(-4.98)	-1.15***	(-5.40)	-0.84***	(-3.52)	-0.67***	(-3.17)
4	0.03	(0.06)	-0.95***	(-3.18)	-0.83***	(-3.37)	-0.38	(-1.57)	-0.23	(-0.93)
5	-0.07	(-0.15)	-1.01***	(-3.27)	-0.90***	(-3.99)	-0.64***	(-2.99)	-0.50**	(-2.42)
6	-0.18	(-0.41)	-1.11***	(-3.78)	-0.90***	(-4.04)	-0.50**	(-2.16)	-0.40	(-1.62)
7	-0.33	(-0.74)	-1.26***	(-4.41)	-1.14***	(-5.31)	-0.86***	(-4.12)	-0.69***	(-3.40)
8	0.17	(0.41)	-0.76***	(-2.71)	-0.63***	(-2.70)	-0.33	(-1.55)	-0.21	(-0.94)
9	0.54	(1.21)	-0.39	(-1.41)	-0.26	(-1.14)	0.01	(0.03)	0.12	(0.55)
10	0.62	(1.41)	-0.28	(-0.95)	-0.25	(-1.11)	-0.06	(-0.26)	0.08	(0.33)
H-L	1.53***	(5.09)	1.56***	(5.24)	1.62***	(5.19)	1.69***	(5.09)	1.62***	(5.11)
Panel B: Value-Weighted										
1	-0.08	(-0.20)	-1.05***	(-3.83)	-1.02***	(-3.54)	-0.96***	(-3.22)	-0.88***	(-3.13)
2	-0.00	(-0.01)	-1.03***	(-3.96)	-0.90***	(-3.50)	-0.61**	(-2.22)	-0.54**	(-2.03)
3	0.35	(0.83)	-0.71**	(-2.46)	-0.60**	(-2.04)	-0.29	(-0.95)	-0.13	(-0.45)
4	0.42	(1.07)	-0.51*	(-1.75)	-0.37	(-1.28)	0.03	(0.10)	-0.01	(-0.05)
5	0.28	(0.63)	-0.76**	(-2.35)	-0.54**	(-2.08)	-0.23	(-1.07)	-0.16	(-0.80)
6	0.40	(1.07)	-0.54**	(-2.06)	-0.35	(-1.53)	-0.16	(-0.66)	-0.17	(-0.70)
7	0.04	(0.09)	-0.93***	(-2.82)	-0.74**	(-2.51)	-0.42	(-1.59)	-0.33	(-1.23)
8	0.51	(1.38)	-0.45	(-1.61)	-0.37	(-1.34)	-0.19	(-0.72)	-0.07	(-0.25)
9	0.96**	(2.41)	0.05	(0.18)	0.27	(1.07)	0.62**	(2.36)	0.59**	(2.23)
10	0.87**	(2.34)	0.04	(0.16)	0.14	(0.52)	0.19	(0.69)	0.22	(0.74)
H-L	0.95**	(2.39)	1.09***	(2.68)	1.15***	(2.65)	1.16**	(2.50)	1.10**	(2.47)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts of high-ATYP firms based on industry information shocks following Cohen and Lou (2012). At the end of each month t, high-ATYP stocks are sorted into deciles according to returns of their corresponding low-ATYP industry peers in the previous month. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Fama and French (2018) six-factor model (FF6: FF5 plus momentum). By controlling for these benchmarks, I assess whether the return patterns associated with ATYP are distinct from known risk factors.

Table 4.6 presents the results. A clear monotonic pattern emerges, with average returns declining steadily from the lowest- to the highest-ATYP decile. The equal-weighted H–L portfolio earns an average return of -1.47% per month ($t = -4.54$), while the value-weighted spread delivers -0.82% per month ($t = -2.50$).[12] After controlling for risk factors, the spreads remain large and statistically significant. For example, under the FF6 model, the equal-weighted H–L alpha is –0.84% per month ($t = -4.52$). The value-weighted FF6 alpha is smaller in magnitude at –0.33% per month, with weaker significance ($t = -1.49$). This contrast indicates that the ATYP effect is concentrated in smaller firms, where attention frictions are most binding and information-processing is less efficient.

Importantly, the negative association between ATYP and abnormal future returns is concentrated in the high-ATYP leg. The most atypical firms earn strongly negative alphas, while low-ATYP firms yield moderate positive alphas. This asymmetry suggests that investors systematically overvalue firms that are hard to process, leading to predictable reversals. The evidence is more consistent with mispricing than with compensation for risk, since high-ATYP firms appear overpriced relative to their fundamentals.

As an additional validation of the univariate portfolio sorts, Figure C.6 in the Appendix plots the cumulative performance of the low-minus-high (L–H) ATYP portfolios over the full sample period. Both equal- and value-weighted portfolios yield persistently positive returns, with the equal-weighted series showing particularly strong and steady outperformance of low-ATYP firms relative to high-ATYP firms. Importantly, this pattern is stable across decades and robust to different market environments, including recessions. This time-series evidence reinforces the cross-sectional sorts by showing that the premium associated with ATYP is a persistent and recurring feature of financial markets rather than a temporary anomaly.

[12]The t-statistics reported in my tables are Newey and West (1987) adjusted with six lags to control for heteroskedasticity and autocorrelation.

Table 4.6: Univariate portfolio sorts on ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.25)	0.25*	(1.94)	0.18***	(2.64)	0.02	(0.38)	0.08	(1.52)
2	0.93***	(3.96)	0.21*	(1.67)	0.18***	(3.06)	0.07	(1.27)	0.12**	(2.29)
3	0.94***	(3.77)	0.19	(1.52)	0.18***	(3.10)	0.13**	(2.11)	0.19***	(3.44)
4	0.91***	(3.45)	0.14	(1.05)	0.15**	(2.44)	0.15**	(2.36)	0.22***	(3.74)
5	0.79***	(2.77)	-0.01	(-0.06)	0.02	(0.27)	0.06	(0.84)	0.15**	(2.02)
6	0.78***	(2.59)	-0.06	(-0.38)	-0.02	(-0.24)	0.09	(1.04)	0.20**	(2.40)
7	0.65*	(1.96)	-0.20	(-1.20)	-0.14	(-1.43)	-0.01	(-0.10)	0.10	(1.19)
8	0.41	(1.16)	-0.48**	(-2.51)	-0.40***	(-3.29)	-0.19*	(-1.73)	-0.06	(-0.54)
9	0.12	(0.31)	-0.82***	(-3.76)	-0.68***	(-4.91)	-0.33***	(-2.75)	-0.20	(-1.61)
10	-0.53	(-1.17)	-1.53***	(-5.47)	-1.35***	(-6.82)	-0.91***	(-5.21)	-0.76***	(-4.36)
H-L	-1.47***	(-4.54)	-1.78***	(-5.99)	-1.54***	(-6.50)	-0.93***	(-5.17)	-0.84***	(-4.52)
Panel B: Value-Weighted										
1	0.80***	(4.35)	0.15	(1.53)	0.10	(1.19)	-0.09	(-1.18)	-0.05	(-0.67)
2	0.66***	(3.65)	0.02	(0.23)	0.00	(0.03)	-0.15**	(-2.29)	-0.14**	(-2.09)
3	0.72***	(3.64)	0.06	(0.99)	0.07	(1.13)	-0.04	(-0.67)	-0.04	(-0.61)
4	0.77***	(3.77)	0.09	(1.20)	0.12*	(1.70)	0.06	(0.79)	0.06	(0.77)
5	0.72***	(3.16)	-0.00	(-0.02)	0.02	(0.28)	0.03	(0.42)	0.04	(0.54)
6	0.79***	(3.31)	0.02	(0.22)	0.08	(1.03)	0.15*	(1.81)	0.18**	(2.03)
7	0.85***	(3.21)	0.06	(0.57)	0.14	(1.56)	0.23***	(2.62)	0.26***	(2.78)
8	0.48*	(1.79)	-0.36***	(-3.04)	-0.25**	(-2.39)	-0.11	(-1.06)	-0.09	(-0.92)
9	0.38	(1.17)	-0.51***	(-2.73)	-0.37**	(-2.34)	-0.08	(-0.52)	-0.05	(-0.35)
10	-0.02	(-0.04)	-1.06***	(-4.21)	-0.82***	(-3.91)	-0.35**	(-2.00)	-0.33*	(-1.92)
H-L	-0.82**	(-2.50)	-1.21***	(-3.91)	-0.92***	(-3.65)	-0.26	(-1.38)	-0.28	(-1.49)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

4.5.2 Bivariate portfolio sorts

Next, I evaluate whether the predictive value of ATYP is subsumed by other well-established predictors of return. To do so, I perform dependent double sorts. At the start of each month, firms are first sorted into quintiles according to a given control characteristic, such as size, book-to-market ratio, 12–1 momentum, profitability, or volatility. Within each quintile, firms are further sorted into ATYP deciles. This procedure yields 50 portfolios (5 × 10), which are conditioned jointly on the control variable and ATYP. The H–L return spread is formed by combining the lowest-ATYP portfolios across quintiles into the low leg and the highest-ATYP portfolios into the high leg. For each control variable, I compute the H–L spread and its alpha under the FF6 model using equal- and value-weighted returns.

The findings shown in Table 4.7 are compelling. To keep the presentation concise, I focus on the H–L portfolio spreads and the Fama and French (2018) six-factor alphas. The evidence demonstrates that ATYP continues to generate substantial economic returns and maintains strong statistical significance even when accounting for various control factors. When portfolios are equally weighted, the six-factor alphas consistently fall between –0.57% and –0.89% monthly, accompanied by robust t-statistics ranging from –3.63 to –5.91. While value-weighted results show somewhat reduced magnitudes, they nonetheless remain statistically meaningful, with alphas spanning –0.13% to –0.74% per month and corresponding t-statistics between –0.89 and –5.08. The persistence of these effects indicates that the return premium associated with ATYP cannot be attributed to cross-sectional return predictors such as firm size, book-to-market ratios, momentum, or other established return predictors.

4.5.3 Fama–MacBeth regressions

I complement the portfolio sorts with Fama and MacBeth (1973) cross-sectional regressions, which allow us to quantify the pricing implications of ATYP while simultaneously controlling for a broad set of firm characteristics. Specifically, at the end of each month I regress one-month-ahead excess stock returns on ATYP and standard predictors, and then average the coefficients over time.

Table 4.8 reports the results for five different specifications with increasing number of controls. Across all specifications, the loading on ATYP is negative, large in magnitude,

Table 4.7: Bivariate portfolio sorts

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
BETA	-1.28***	-0.78***	-0.63***	-0.23
	(-4.68)	(-4.54)	(-2.67)	(-1.32)
SIZE	-1.31***	-0.79***	-1.28***	-0.76***
	(-4.71)	(-5.40)	(-4.53)	(-5.15)
BM	-1.34***	-0.72***	-0.85***	-0.28*
	(-4.49)	(-3.97)	(-2.87)	(-1.74)
STR	-1.42***	-0.89***	-0.62**	-0.16
	(-4.81)	(-4.96)	(-2.14)	(-0.89)
MOM	-1.21***	-0.74***	-0.89***	-0.47***
	(-4.40)	(-4.53)	(-3.28)	(-2.99)
OP	-1.05***	-0.77***	-0.57**	-0.22
	(-4.83)	(-4.84)	(-2.42)	(-1.32)
AG	-1.17***	-0.69***	-0.48**	-0.13
	(-4.71)	(-4.35)	(-2.32)	(-0.93)
TURN	-1.32***	-0.72***	-0.72***	-0.24
	(-4.65)	(-4.24)	(-2.93)	(-1.59)
ILLIQ	-1.33***	-0.81***	-1.18***	-0.74***
	(-4.60)	(-5.15)	(-4.38)	(-5.08)
IVOL	-0.96***	-0.67***	-0.70***	-0.39**
	(-6.30)	(-5.91)	(-3.09)	(-2.43)
SUE	-1.33***	-0.73***	-0.81***	-0.33*
	(-4.28)	(-3.90)	(-2.66)	(-1.80)
MAX	-1.00***	-0.57***	-0.80***	-0.38**
	(-4.26)	(-3.63)	(-3.21)	(-2.21)
AGE	-1.26***	-0.81***	-0.66***	-0.25
	(-4.71)	(-4.88)	(-2.66)	(-1.50)
AFD	-1.26***	-0.70***	-0.89***	-0.37**
	(-4.43)	(-4.39)	(-3.07)	(-2.11)
#ANA	-1.22***	-0.65***	-1.03***	-0.59***
	(-4.02)	(-3.98)	(-3.57)	(-3.22)

The table reports average monthly return spreads (H–L) and Fama and French (2018) six-factor (FF6) alphas from bivariate portfolio sorts on ATYP and standard firm characteristics. In each month, stocks are independently sorted into quintiles by a given control variable and, within each quintile, further sorted into deciles by ATYP. Reported values are the average return differences between the highest and lowest ATYP deciles, averaged across the control quintiles. Panel columns display results for equal-weighted and value-weighted portfolios. H–L denotes the excess return spread, while FF6 reports the intercept from regressions of portfolio excess returns on the six-factor model. Newey and West (1987) t-statistics with six lags are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

and statistically significant. In the univariate specification, the ATYP coefficient is –2.33 ($t = -4.55$), implying that a one-unit increase in ATYP predicts a 2.33% lower excess return in the subsequent month. Importantly, the coefficient remains strongly negative even after sequentially including controls for size, value, momentum, profitability, investment, trading activity, illiquidity, idiosyncratic volatility, earnings surprises, firm

age, disagreement, and attention. In the most saturated specification, which includes more than a dozen established predictors, ATYP continues to load significantly at –0.72 ($t = -3.21$). This demonstrates that the predictive power of ATYP is not subsumed by other well-known return determinants. Most importantly, the negative association is robust to characteristics related to illiquidity, idiosyncratic volatility, lottery-like return profiles and disagreement, suggesting that ATYP captures a distinct dimension of information-processing frictions.

The Fama–MacBeth results reinforce the portfolio evidence that stocks with atypical characteristic profiles systematically earn lower future returns, even after accounting for risk factors and anomaly controls. This robustness strengthens the interpretation of ATYP as an independent state variable linked to investor attention frictions rather than as a proxy for existing characteristics.

4.5.4 Return persistence

I conclude my main results by examining the persistence of the ATYP premium. Since ATYP captures fundamental economic atypicality rather than transient noise, I expect both the signal and its associated return predictability to persist over longer horizons.

First, I verify the stability of the measure itself. Because ATYP is largely derived from fundamental accounting variables that evolve gradually, firms typically do not experience sharp month-to-month changes in their atypicality status. Figure C.7 in the Appendix confirms this stability. Transitions across ATYP deciles are infrequent, particularly at the extremes, where nearly half of the firms remain in the same decile after 12 months. This persistence suggests information-processing difficulty is a durable characteristic of a firm's information environment.

Next, we examine whether this persistent signal translates into long-lasting return predictability. Figure 4.3 plots the average high-minus-low (H–L) decile portfolio spread for horizons ranging from one to twelve months ahead. We report results for the full sample and distinguish between equal-weighted and value-weighted portfolios to assess the role of firm size, and for both raw returns and alphas.

The return spread remains economically large and persistent across horizons. For equal-weighted portfolios, the H–L spread is consistently negative and lies close to -1.5% per month at short horizons, with only modest variation over time. Although there is some attenuation as the horizon increases, the spread remains around -1.25% even

Table 4.8: Fama-MacBeth cross-sectional regressions

	Excess Return	Excess Return	Excess Return	Excess Return	Excess Return
(Intercept)	1.63***	0.87***	0.60***	1.04***	0.91***
	(7.57)	(4.08)	(2.85)	(4.58)	(3.93)
ATYP	-2.33***	-2.19***	-1.43***	-0.81***	-0.72***
	(-4.55)	(-5.02)	(-4.04)	(-3.51)	(-3.21)
BETA		0.10	0.09	0.22**	0.20**
		(0.94)	(0.88)	(2.29)	(2.18)
SIZE		0.00	-0.00	-0.00*	-0.00***
		(0.07)	(-0.42)	(-1.70)	(-3.35)
BM		0.63***	0.57***	0.41***	0.43***
		(5.54)	(5.25)	(4.23)	(4.62)
MOM		0.97***	0.91***	0.82***	0.83***
		(6.58)	(6.33)	(5.58)	(5.74)
AG			-0.35***	-0.33***	-0.34***
			(-6.09)	(-6.46)	(-6.67)
OP			0.53***	0.45***	0.44***
			(5.13)	(4.95)	(4.99)
STR				-2.09***	-2.13***
				(-5.73)	(-5.86)
TURN				-44.92***	-55.19***
				(-3.44)	(-4.10)
ILLIQ				0.13*	0.12*
				(1.96)	(1.95)
IVOL				-11.94***	-9.97**
				(-2.71)	(-2.38)
SUE				0.07***	0.07***
				(5.92)	(5.79)
MAX				-4.28***	-4.16***
				(-6.14)	(-6.16)
AGE					-0.00
					(-0.61)
AFD					-0.20**
					(-2.41)
#ANA					0.02***
					(3.88)

The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is included either individually or jointly with other predictors. The controls considered are defined in Section 4.3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

twelve months ahead. The corresponding alpha estimates closely track the raw returns, indicating that standard risk factor adjustments do not materially alter the magnitude or persistence of the effect. Value-weighted portfolios display a smaller but still robust pattern. The H–L spread starts at roughly -0.8% to -0.9% per month and declines to approximately -1.0% at intermediate horizons before partially rebounding toward the end of the year. As in the equal-weighted case, the alpha series closely mirrors the raw

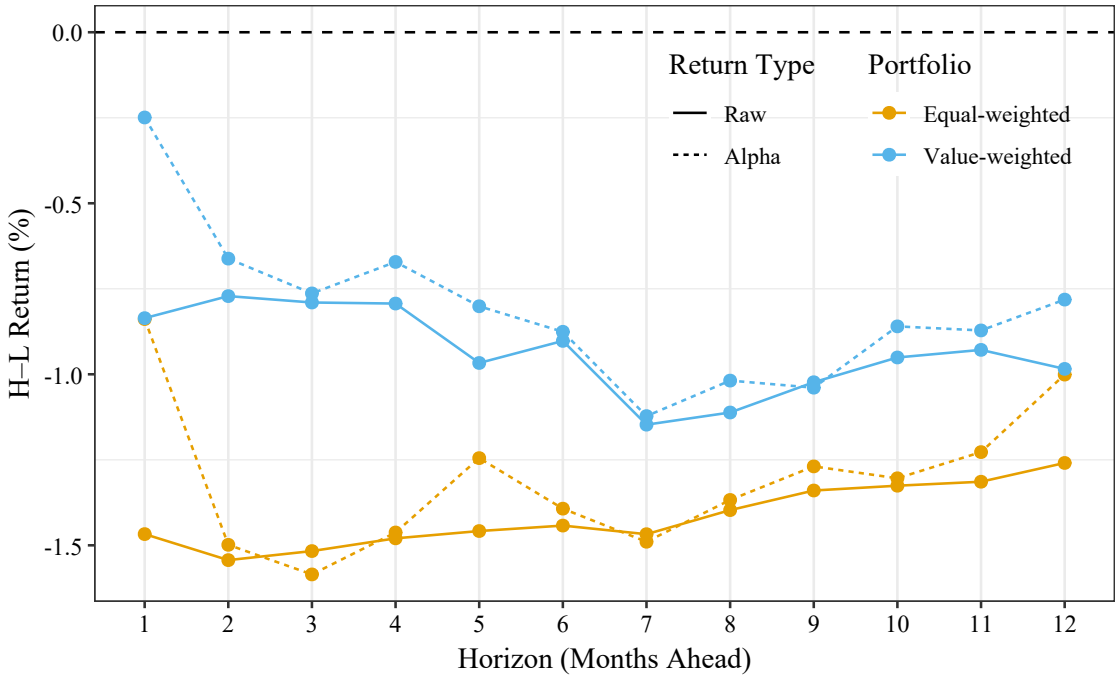


Figure 4.3: Long-term predictive power
The figure plots average high-minus-low (H–L) returns of ATYP-sorted decile portfolios across different horizons up to 12 months ahead. At each month t, stocks are sorted into ATYP deciles, and the return spread between the highest and lowest deciles is computed for horizons $t+1$ through $t+12$. The figure reports results separately for equal-weighted and value-weighted portfolios, and for both raw returns and alphas. Solid lines represent raw spreads, and dotted lines represent alphas. The yellow line corresponds to equal-weighted portfolios and the blue line corresponds to value-weighted portfolios.

spread, suggesting that the negative return differential is not explained by conventional risk exposure. Overall, the figure indicates that the mispricing associated with atypical firms is slow to dissipate and is substantially stronger in equal-weighted portfolios, consistent with the interpretation that the effect is driven primarily by smaller firms and is amplified by limits to arbitrage and constrained investor attention.

4.6 Channels of return predictability

After documenting the negative relationship between ATYP and average stock returns, I examine the economic mechanisms that may explain this pattern. Based on the theoretical literature and the evidence presented thus far, I argue that information-processing difficulty may lead to mispricing. First, I examine whether ATYP covaries with established proxies of mispricing. Then, I explore potential channels. Specifically, I consider limits to arbitrage arising from short-sale constraints and limited investor attention.

4.6.1 Mispricing vs. risk

The negative pricing of ATYP is best understood as a mispricing premium rather than compensation for risk. Two interacting mechanisms explain why information-processing difficulty leads to systematic overpricing and subsequent low returns.

First, high ATYP increases the ambiguity surrounding a firm's valuation, which raises information uncertainty. In environments with greater uncertainty, behavioral biases such as overconfidence become more pronounced, leading investors to overweight their own noisy and imperfect signals (Daniel et al., 1998, 2001; Zhang, 2006). This amplified confidence in private assessments generates stronger divergence of opinion across market participants. When such disagreement interacts with short-sale constraints, market prices increasingly reflect the valuations of optimistic investors rather than the full distribution of beliefs, resulting in systematic overpricing of firms with high ATYP (Miller, 1977). Consistent with this mechanism, Section 4.4.2 shows that ATYP is associated with higher forecast dispersion, forecast errors, and return volatility, all of which indicate greater disagreement and uncertainty in the valuation of atypical firms.

Second, limited attention affects how quickly information about atypical firms is incorporated into prices. Investors with constrained processing capacity optimally allocate attention toward broad systematic patterns and devote less effort to complex firm-specific details (Peng and Xiong, 2006). Because atypical firms deviate from these coarse patterns, their fundamentals receive less immediate scrutiny, and new information is only gradually absorbed across the investor population. In the spirit of Hong and Stein (1999), such slow diffusion produces systematic underreaction, where only a subset of sophisticated investors processes the news promptly and others update their valuations with delay. This mechanism explains the strong post-earnings-announcement drift and the sluggish response to industry shocks documented in Section 4.4.3 and Section 4.4.4. Moreover, limits to arbitrage, such as high idiosyncratic volatility and limited liquidity, prevent informed traders from correcting these pricing errors quickly (Pontiff, 2006), allowing mispricing in high-ATYP firms to persist for extended periods.

If the negative pricing of ATYP indeed reflects mispricing, it should covary with established mispricing measures. I therefore benchmark ATYP against the composite mispricing score of Stambaugh et al. (2015), where higher values indicate greater overpricing. Panel A of Table 4.9 shows that firms with high ATYP also have significantly

higher mispricing scores, consistent with the idea that they are more prone to overpricing. Panel B reports dependent double sorts, where firms are first grouped by the mispricing score and then sorted into ATYP quintiles. The ATYP high-minus-low spread is most pronounced in the highest mispricing quintile, with a six-factor alpha of -0.88% ($t = -3.98$). These empirical results reinforce my interpretation that the ATYP premium reflects mispricing rather than compensation for risk and is concentrated in the most overpriced firms.

Table 4.9: Mispricing and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Average Mispricing									
	0.42	0.44	0.47	0.51	0.59	0.17***	47.54		
Panel B: Bivariate Sort on Mispricing									
Low	1.12	1.26	1.36	1.50	1.55	0.43***	2.64	0.56***	4.29
2	1.02	1.06	1.15	1.18	1.10	0.08	0.45	0.30**	2.39
3	0.82	0.92	0.97	0.88	0.67	-0.14	-0.65	0.23	1.50
4	0.55	0.52	0.45	0.35	0.18	-0.37	-1.42	0.01	0.08
High	-0.24	-0.48	-0.67	-0.87	-1.49	-1.25***	-4.74	-0.88***	-3.98
H-L	-1.36	-1.74	-2.03	-2.37	-3.04	-1.68***	-8.33	-1.44***	-7.60

The table reports univariate and bivariate portfolio sorts analyzing the interaction between ATYP and the mispricing score of Stambaugh et al. (2015). Panel A displays average values of the mispricing score across ATYP quintiles. Panel B shows dependent double sorts in which firms are first grouped into quintiles based on the mispricing score and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, associated t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

4.6.2 Limits to arbitrage

When investors disagree, frictions that limit arbitrage can lead to overpricing. Since ATYP involves ex-ante information-processing frictions that increase investor disagreement, I should observe a similar effect with ATYP under high limits to arbitrage. The literature highlights several such frictions. First, high idiosyncratic volatility raises the residual risk of arbitrage positions, deterring arbitrageurs from betting against overpriced firms (Pontiff, 2006; Stambaugh et al., 2015). Second, illiquidity raises trading costs and price impact, further discouraging large corrective trades (Amihud, 2002; Pástor and Stambaugh, 2003). Third, young firms have shorter track records, higher uncertainty, and greater information asymmetry, making them especially difficult to value (Pástor and

Veronesi, 2003).

Table 4.10 explores these mechanisms by conditioning the ATYP premium on proxies for arbitrage frictions. The conditional sorts show that the negative ATYP premium is concentrated in the high-friction groups. For instance, in the high-IVOL group, the FF6-adjusted H–L return is -1.51% per month ($t = -6.77$), compared to a statistically insignificant 0.13% ($t = 1.90$) in the low-IVOL group. Similarly, the spread is -1.22% ($t = -6.18$) in illiquid stocks vs. 0.07% ($t = 0.73$) in liquid stocks, and –1.18% ($t = -5.50$) in young firms vs. -0.06% ($t = -0.48$) in older firms. In all cases, the difference-in-spreads is large and highly significant.

Table 4.10: Limits to arbitrage and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Bivariate Sort on Idiosyncratic Volatility									
Low	0.91	0.90	0.89	0.87	0.88	-0.03	-0.41	0.13*	1.90
2	0.97	0.99	0.91	0.83	0.70	-0.27***	-2.76	-0.15	-1.60
3	0.98	0.87	0.75	0.48	-0.02	-1.00***	-5.70	-0.71***	-5.45
4	0.88	0.87	0.82	0.36	-0.06	-0.93***	-5.07	-0.63***	-4.88
High	0.72	0.50	0.08	-0.18	-1.06	-1.78***	-7.59	-1.51***	-6.77
H-L	-0.19	-0.40	-0.81	-1.05	-1.95	-1.76***	-7.56	-1.64***	-7.30
Panel B: Bivariate Sort on Illiquidity									
Low	0.83	0.83	0.78	0.71	0.45	-0.38*	-1.77	0.07	0.73
2	0.96	0.96	0.82	0.54	0.05	-0.91***	-3.56	-0.40***	-2.69
3	0.97	0.80	0.70	0.27	-0.28	-1.26***	-4.55	-0.66***	-3.63
4	1.05	0.96	0.65	0.33	-0.34	-1.39***	-4.66	-0.91***	-4.62
High	1.01	0.96	0.76	0.48	-0.41	-1.41***	-5.64	-1.22***	-6.18
H-L	0.17	0.13	-0.02	-0.22	-0.86	-1.03***	-4.81	-1.29***	-5.89
Panel C: Bivariate Sort on Firm Age									
Low	0.81	0.61	0.20	-0.18	-0.91	-1.73***	-6.20	-1.18***	-5.50
2	0.83	0.76	0.53	0.34	-0.36	-1.20***	-4.30	-0.67***	-3.50
3	1.05	1.02	0.94	0.72	0.17	-0.88***	-3.26	-0.41**	-2.49
4	0.97	1.03	0.91	0.84	0.30	-0.67***	-3.17	-0.31**	-2.22
High	0.91	0.88	0.88	0.91	0.68	-0.22	-1.26	-0.06	-0.48
H-L	0.09	0.27	0.68	1.10	1.60	1.51***	6.45	1.12***	5.23

The table presents bivariate portfolio sorts examining the relation between ATYP and proxies for limits to arbitrage. Panels A, B and C show dependent double sorts in which firms are first partitioned into quintiles based on each limits to arbitrage proxy and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, corresponding t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

These findings strongly support a mispricing interpretation. If ATYP represented a priced risk factor, the premium should persist regardless of arbitrage frictions, or potentially be stronger where arbitrage is easier. Instead, the effect is entirely absent

in low-friction groups and survives standard risk factor controls only where limits to arbitrage are binding. I confirm this conclusion using Fama-MacBeth regressions in Table C.3 in the Appendix, where the interaction terms between ATYP and arbitrage proxies remain significant after controlling for firm characteristics. This evidence suggests that information-processing difficulty generates mispricing that persists precisely because arbitrageurs are unable or unwilling to correct it.

4.6.3 Limited attention

Limited investor attention leads to a slower correction of mispricing, because constrained investors rely on simplified heuristics rather than full information processing. As Peng and Xiong (2006) demonstrate, investors cope with cognitive limits by allocating attention to broad systematic patterns while neglecting firm-specific information. Because high-ATYP firms do not fit these standard categories, investors neglect their idiosyncratic details. In the spirit of Hong and Stein (1999), this neglect prevents signals from being immediately incorporated into prices. Instead, information diffuses only gradually across the investor population, leading to underreaction and persistent mispricing. However, this friction should vanish when attention is abundant. In high-attention environments, sophisticated investors rapidly process all available information regardless of its complexity and ensure that prices reflect fundamental value. To test this prediction, I condition the ATYP premium on three proxies for the information environment: firm size, analyst coverage, and short interest. Prior work shows that larger firms attract more informed attention Hong et al. (2000), low analyst coverage reflects weaker information efficiency Elgers et al. (2001), and high short interest signals a high share of sophisticated investors Boehmer and Wu (2012).

Table 4.11 reports the results of dependent double sorts. The findings show that the pricing effect of ATYP is sharply concentrated in low-attention environments, where investors are most likely to rely on coarse patterns. For example, within the low-size group, the six-factor adjusted high-minus-low return is -1.54% per month ($t = -6.90$), whereas in the high-size group the spread is economically and statistically negligible. I observe an identical pattern for analyst coverage, with a spread of -1.02% in low-coverage firms versus 0.27% in high-coverage firms, and for short interest with -1.39% versus -0.20%. In every specification, the difference in spreads between the low- and high-attention groups is large and statistically significant, ranging from 1.19% to 1.74% per

month.

Table 4.11: Attention constraints and ATYP

	Low	2	3	4	High	H-L	t-stat	FF6	tstat
Panel A: Bivariate Sort on Size									
Low	1.13	0.87	0.71	0.28	-0.60	-1.73***	-6.43	-1.54***	-6.90
2	1.07	0.84	0.70	0.24	-0.39	-1.46***	-5.31	-0.94***	-5.00
3	1.00	0.94	0.72	0.38	-0.14	-1.13***	-4.10	-0.60***	-3.77
4	0.95	0.90	0.85	0.61	0.13	-0.82***	-3.41	-0.35***	-2.99
High	0.80	0.80	0.77	0.71	0.55	-0.26	-1.17	0.20*	1.96
H-L	-0.33	-0.07	0.06	0.42	1.15	1.48***	6.28	1.74***	7.90
Panel B: Bivariate Sort on #Analyst									
Low	0.92	1.04	0.69	0.28	-0.36	-1.29***	-4.62	-1.02***	-4.96
2	0.99	0.87	0.69	0.12	-0.67	-1.66***	-5.26	-1.20***	-5.81
3	1.00	0.88	0.84	0.53	-0.00	-1.00***	-3.75	-0.47***	-2.71
4	0.91	0.91	0.70	0.53	0.13	-0.78***	-2.98	-0.20	-1.25
High	0.86	0.82	0.88	0.72	0.56	-0.30	-1.23	0.27**	2.35
H-L	-0.06	-0.22	0.19	0.44	0.92	0.99***	4.25	1.29***	6.31
Panel C: Bivariate Sort on Short-interest									
Low	0.94	0.74	0.54	0.20	-0.63	-1.58***	-5.95	-1.39***	-6.41
2	0.95	1.03	0.97	0.47	0.01	-0.94***	-3.43	-0.50***	-2.81
3	1.07	0.98	0.93	0.59	0.12	-0.95***	-3.63	-0.50***	-3.23
4	0.88	0.77	0.55	0.38	-0.13	-1.01***	-3.66	-0.51***	-2.86
High	0.86	0.85	0.82	0.69	0.25	-0.60**	-2.46	-0.20	-1.47
H-L	-0.09	0.11	0.28	0.48	0.89	0.97***	4.09	1.19***	4.87

The table presents bivariate portfolio sorts examining the relation between ATYP and proxies for investor attention constraints. Panels A, B and C show dependent double sorts in which firms are first partitioned into quintiles based on each attention proxy and then sorted into quintiles by ATYP within each quintile. Reported are average excess returns, high-minus-low (H–L) spreads, corresponding t-statistics, and Fama and French (2018) six-factor (FF6) alphas. Newey and West (1987) adjusted t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.4 in the Appendix confirms these results using Fama-MacBeth cross-sectional regressions. I regress returns on ATYP, attention measures, and their interactions while controlling for standard characteristics. The interaction coefficients are consistently significant, indicating that the negative pricing of ATYP is conditional on the scarcity of investor attention.

These findings strongly support the attention-based mechanism. In low-attention segments of the market, pattern-based traders dominate. Because they process information through standard patterns, as modeled by Peng and Xiong (2006), they struggle to value atypical firms correctly. Consequently, information diffuses slowly, generating the systematic underreaction predicted by Hong and Stein (1999). In contrast, when attention is high, sophisticated investors allocate sufficient cognitive resources to decipher complex

signals. In these environments, atypical information is incorporated efficiently and the return predictability disappears. Together with my evidence on limits to arbitrage, this confirms that information-processing difficulty generates overpricing precisely when investors lack the bandwidth to look beyond systematic patterns.

4.7 Robustness checks

Having established that ATYP is a valid measure of information-processing difficulty and that there is a negative cross-sectional relationship between ATYP and average stock returns that is driven by limited attention, limits to arbitrage and potentially mispricing, I next run a series of robustness checks to challenge the aforementioned findings.

4.7.1 Firm characteristic clusters

An important question is whether the ATYP premium generalizes across distinct domains of firm information or arises from a concentrated subset of characteristics. To address this, I re-calculate ATYP within clusters of related firm characteristics and examine whether the return predictability persists across these different information domains. The clusters are taken from Jensen et al. (2023). Within each cluster, I compute firm-specific reconstruction errors based solely on the variables in that group and form decile portfolios on the resulting ATYP measure.

Table C.5 in the Appendix reports the high-minus-low return spreads from these sorts, along with factor-adjusted alphas from the Fama and French (2018) six-factor model. The equal-weighted results show that the ATYP premium is present in nearly all clusters. For example, firms that are hard to process with respect to investment, profitability, and value characteristics exhibit economically large and statistically significant underperformance, with spreads ranging from -0.74% to -1.53% per month. The effect is somewhat weaker for the value-weighted portfolios, but still evident in investment and leverage-related clusters.

4.7.2 Within-industry construction

So far, my results have taken the systematic patterns for the ATYP computation from the whole market. In this section, I examine whether the pricing of ATYP also emerges when the measure is estimated within industries. ATYP is re-estimated separately for each

one digit SIC industry, and all preprocessing steps outlined in Section 4.3.1, including imputation, winsorization, and standardization, are performed within each industry in every month. The model is then trained separately for each industry, allowing it to learn the prevailing information structure within that group and to identify firms that are atypical relative to their immediate peers. The resulting industry-adjusted ATYP values are then combined across industries to form the complete sample.

Table C.6 in the Appendix presents the results. Panel A reports equal-weighted and Panel B value-weighted portfolio sorts for the full cross-section using the industry adjusted measure. The pattern of returns is clear and monotonic. Firms with higher ATYP continue to earn lower subsequent returns. The equal-weighted high-minus-low spread is -1.40 percent per month ($t = -4.73$) with an FF6 alpha of -0.93 percent ($t = -5.43$). Value-weighted portfolios show a smaller but consistent effect. The overall magnitude and statistical strength of the spreads remain comparable to the baseline results, showing that the pricing of information-processing difficulty persists even when ATYP is estimated entirely within industries.

Table C.7 in the Appendix further reports high-minus-low spreads for portfolio sorts conducted separately within each one digit SIC industry. The pricing effect of ATYP remains negative and significant across nearly all industries. Equal-weighted alphas range between -0.39 and -1.28 percent, with the strongest effects in *Public Administration*, *Services*, and *Trade*. Value-weighted results follow the same pattern. The persistence of the pattern across nearly all sectors demonstrates that information-processing difficulty is priced within the industrial structure of the economy.

4.7.3 Within-size group construction

Next, I examine whether the pricing of ATYP varies across the size distribution. To ensure that the measure captures atypicality relative to firms with similar information environments, I re-estimate ATYP separately within each size group.[13] At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. All stages of the ATYP construction, including preprocessing and the model training, are performed within each size group. This procedure removes any

[13]The size groups are taken from Jensen et al. (2023). These groups are non-overlapping, and the breakpoints are based on the market equity of NYSE stocks at the end of each month. Specifically, mega caps are stocks with a market cap above the 80th percentile of NYSE stocks, large caps are all remaining stocks above the 50th percentile, small caps are all remaining stocks above the 20th percentile, micro caps are all remaining stocks above the 1st percentile, and nano caps are all remaining stocks.

systematic relation between firm size and the ATYP measure. The boxplots in Figure C.8 confirm that the distribution of ATYP is nearly identical across the five size groups, and that the unconditional correlation between size and ATYP, as documented in Section 4.3.4, effectively disappears under this design.

Table C.8 presents the full sample portfolio sorts based on size-adjusted ATYP. Panel A reports equal-weighted portfolios and Panel B reports value-weighted portfolios. The return pattern across deciles is clear and monotonic. Firms in the lowest ATYP decile earn the highest subsequent returns, while firms in the highest ATYP decile earn substantially lower returns. The equal-weighted high-minus-low spread is -1.33 percent per month ($t = -4.81$), with a corresponding FF6 alpha of -0.80 percent ($t = -5.49$). As in the baseline results, value-weighted portfolios show a similar but somewhat attenuated pattern, consistent with the concentration of information-processing difficulty among smaller firms. Overall, the magnitude and statistical strength of the spreads remain comparable to the baseline results, showing that the pricing of processing difficulty persists even when ATYP is estimated entirely within size groups.

Despite the absence of a direct mechanical relation between ATYP and size, the return implications of the measure vary substantially across the size dimension. Table C.9 reports the high-minus-low portfolio spreads computed separately within each size group. The pricing effect is concentrated almost entirely among the smaller firms. For nano and micro firms, the equal-weighted high-minus-low spreads are large and statistically significant. The spreads remain strongly negative after controlling for standard priced risk factors. For example, among nano firms, the FF6 alpha of the high-minus-low portfolio is approximately -1.59 percent per month with a t-statistic of -4.12. The pattern is nearly identical for micro firms. Small firms also exhibit a negative spread, though the magnitude is slightly smaller. In contrast, the pricing effect attenuates sharply among large and mega firms. For these firms, the high-minus-low spreads are close to zero, and the corresponding alphas are statistically indistinguishable from zero. Value-weighted results follow the same pattern and further underscore that the pricing of ATYP is concentrated among firms with lower size, consistent with interpretations related to investor attention and limits to arbitrage.

4.7.4 Alternative factor models

In this section, I examine whether the estimated ATYP premium survives when controlling for alternative factor models beyond the baseline Fama–French six-factor model. To do so, I re-estimate portfolio alphas using several widely employed frameworks that encompass both rational and behavioral sources of return variation. Specifically, I compare the equal- and value-weighted portfolio spreads under the six-factor model of Fama and French (2018), the four-factor q-model of Hou et al. (2014) that includes the market, size, investment, and profitability factors, the behavioral factor model (DHS) of Daniel et al. (2019) that employs market, post-earnings-announcement-drift, and financing factors, and the characteristic-efficient portfolio factor model (DMRS) of Daniel et al. (2020) which builds on the FF5 framework while neutralizing unpriced covariance structures. The corresponding results are reported in Table C.10 in the Appendix.

Panel A reports equal-weighted portfolio returns. When controlling for risk using the FF6 model, the alpha remains economically and statistically significant at -0.84 percent per month ($t = -4.52$). The magnitude of the alpha is only slightly reduced when applying the q-factor model (-0.73 percent, $t = -3.24$) or the behavioral DHS model (-0.67 percent, $t = -2.56$). The DMRS specification yields a comparable result (-0.93 percent, $t = -3.34$). The persistence of negative and statistically significant alphas across all model specifications indicates that the return differential associated with ATYP cannot be explained by standard investment-based or behavioral factor structures.

The somewhat smaller alpha under the DHS model is consistent with the interpretation that information-processing difficulty affects prices through behavioral channels related to limited attention and delayed information processing. Since the DHS framework explicitly includes a post–earnings-announcement-drift factor that captures investor underreaction to new information, it absorbs part of the same behavioral component that the information-processing difficulty measure identifies. Consequently, the weaker alpha under the DHS model reinforces the notion that the pricing effect of ATYP reflects potential mispricing arising from cognitive and attention-related frictions rather than compensation for systematic risk.

Panel B reports value-weighted results. The alphas across all factor models are smaller in magnitude and statistically weaker (-0.28, -0.23, -0.17, and -0.11 percent for the FF6, q, DHS, and DMRS models, respectively), which is expected given that high-ATYP

firms are typically smaller and less liquid, causing value-weighted portfolios to dilute their influence. Nonetheless, the signs of all estimated alphas remain consistent with the equal-weighted results, and no specification fully eliminates the return differential between high- and low-ATYP firms.

4.7.5 Alternative data sources

To test the robustness of my findings to data construction choices, I replicate the main portfolio analyses using alternative sources of firm-level information. For each dataset, I apply the same data processing as in the baseline specification in Section 4.3.1: missing data are imputed with the cross-sectional median, variables are winsorized at the 1% level on both sides, standardized, and used as inputs in the autoencoder to construct ATYP as the firm-level reconstruction error. This approach ensures that any differences in results are due to the underlying data rather than methodological choices. The model is also specified as the baseline specification in Section 4.3.3.

As a first replication, I employ the WRDS Financial Ratio database, which provides a standardized set of 71 accounting and market-based variables. Table C.11 in the Appendix reports the results. The equal-weighted high-minus-low ATYP portfolio spread is large and statistically significant, with an average difference of –1.42% per month ($t = -5.68$) in excess returns and –0.95% ($t = -6.48$) after controlling for the Fama and French (2018) six-factor model. Value-weighted sorts yield a weaker but still meaningful spread of –0.50% ($t = -2.15$). The return pattern is robust across alternative factor models, including the CAPM, the Fama and French (1993) three-factor model, and the Fama and French (2015) five-factor model.

As a second replication, I expand the input space to the full set of 153 firm characteristics included in the dataset of Jensen et al. (2023) without excluding any clusters. Table C.12 in the Appendix presents the results. Equal-weighted sorts again show a pronounced ATYP effect, with the high-minus-low portfolio spread of –1.42% per month in excess returns ($t = -4.48$) and –0.83% ($t = -4.48$) after controlling for the Fama and French (2018) six-factor model. The spread remains highly significant across alternative factor models, including the three- and five-factor specifications. Value-weighted portfolios deliver somewhat weaker results, with a spread of –0.57% ($t = -1.82$) in excess returns and –0.94% ($t = -3.27$) relative to the CAPM. These findings show that the predictive power of ATYP is not an artifact of excluding seasonal or short-horizon

return-based risk measures. Instead, ATYP continues to price the cross-section even in a maximally inclusive specification, reinforcing the robustness of the measure to dataset construction choices.

4.7.6 Different model specifications

The predictive power of ATYP may hinge on the choice of hyperparameters in the autoencoder or on the specific model architecture employed. To address this, I systematically vary the main tuning dimensions of the autoencoder, including the size of the bottleneck layer, the number of hidden layers, the number of units per layer, and the degree of regularization through L_2, dropout, and input noise added.

Table C.13 in the Appendix summarizes the results. Across all 48 hyperparameter choices, the equal-weighted high-minus-low ATYP portfolio spread remains large in magnitude and statistically significant, with t-statistics typically above four in both raw and factor-adjusted returns. The value-weighted spreads are smaller, but they are consistently negative and retain statistical significance for many specifications. This robustness underscores that the ATYP premium is not an artifact of a particular configuration of the network. Even when the dimensionality of the bottleneck or the strength of regularization is altered substantially, the return patterns associated with ATYP remain stable.

4.7.7 Autoencoder vs. PCA

In Section 4.2, I motivate the use of an autoencoder by arguing that systematic combinations of firm characteristics are inherently nonlinear. A potential concern, however, is whether the predictive power of ATYP truly stems from these nonlinear interactions or simply from linear deviations that could be captured by standard dimensionality reduction techniques like PCA. If ATYP is merely a proxy for linear atypicality, the additional complexity of the neural network would be unnecessary.

To isolate the contribution of nonlinear information processing, I construct a linear benchmark measure, ATYP_{PCA}, using a PCA model with the same bottleneck dimension ($k = 16$) as my baseline autoencoder. I then regress my original autoencoder-based measure, ATYP_{AE}, on ATYP_{PCA} for each firm-month and extract the orthogonal residual. This residual captures the component of information-processing difficulty that arises strictly from nonlinear deviations, i.e., information that the linear PCA model considers typical or cannot effectively compress.

Table C.14 in the Appendix reports the results of univariate portfolio sorts based on this orthogonalized measure. The equal-weighted H–L spread remains economically large and statistically significant at -0.53 percent per month ($t = -4.55$). Crucially, the spread retains a significant alpha of -0.42 percent ($t = -4.45$) under the Fama and French six-factor model, indicating that the nonlinear component of atypicality carries independent pricing information. While the value-weighted results are weaker, the persistence of a strong equal-weighted effect confirms that the autoencoder identifies a distinct set of hard-to-process firms. Those are likely cases with complex, interactive characteristic profiles that a linear model fails to capture. These findings validate my choice of a nonlinear architecture and suggest that information-processing difficulty is, at least in part, a function of complex, non-additive economic patterns.

4.7.8 Different measurement specifications

In constructing the ATYP measure as the root mean of squared reconstruction errors, it is possible that the measure places substantial weight on a small number of extreme observations within each cluster. To assess the influence of such observations, I develop alternative formulations of the ATYP measure that reduce sensitivity to outliers and evaluate the robustness of my findings across these specifications. I then replicate my univariate portfolio sorts using the alternative constructions in place of the mean-based ATYP definition in Equation (4.5).

First, I calculate *MedATYP* as

$$\text{MedATYP}_{i,t} \;=\; \text{median}_j \left[\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)} \right)^2 \right]. \tag{4.7}$$

By construction, MedATYP is less sensitive to extreme deviations in individual characteristics and instead reflects the typical level of reconstruction error across dimensions. This construction ensures that the measure captures systematic mismatches between the observed and reconstructed firm profiles rather than being disproportionately affected by a single variable.

Second, I construct a rank-based measure, *RankATYP*, which normalizes reconstruction errors dimension by dimension. More specifically, for each month t and characteristic j, I assign firms to deciles based on the squared reconstruction error $\left(\mathbf{X}_{i,t}^{(j)} - \hat{\mathbf{X}}_{i,t}^{(j)} \right)^2$. I then

compute each firm's RankATYP score as the average decile rank across all characteristics,

$$\text{RankATYP}_{i,t} \;=\; \frac{1}{p}\sum_{j=1}^{p}\text{RankDecile}_{i,t}^{(j)}\,. \tag{4.8}$$

This construction has two advantages: it eliminates the influence of scale differences across characteristics and ensures that no single feature with very large errors dominates the aggregate score. RankATYP thus provides a robust, distribution-free measure of how atypical a firm appears in the cross-section.

Tables C.15 and C.16 in the Appendix show the results for MedATYP and RankATYP, respectively. Both specifications yield highly consistent results with my baseline measure. The decile spreads remain large, negative, and statistically significant, with equal-weighted excess returns of approximately –1.4% per month and significant FF6 alphas. The value-weighted spreads are somewhat weaker but follow the same monotonic pattern. These robustness exercises demonstrate that my findings are not artifacts of the specific aggregation method. Whether ATYP is computed as a mean, median, or average rank across reconstruction errors, the evidence consistently points to the same underlying mechanism.

4.7.9 Missing data and imputation

Section 4.3.1 explains how I handle missing data. In this section, I examine whether the ATYP measure reflects genuine information structure or mechanical effects arising from missing data and imputation procedures. Firms with less complete information could mechanically exhibit higher reconstruction errors because the autoencoder relies on imputed rather than observed characteristics. To address this concern, I implement two robustness checks designed to isolate any influence of missing or imputed characteristics on the estimated measure.

First, I recalculate the ATYP score using only the non-missing firm characteristics in each observation of $\mathbf{X}_{i,t}$. For every firm-month, I exclude imputed entries from the reconstruction and compute the reconstruction error in Equation (4.5) based on the number of available characteristics, denoted $p_{i,t}$. This procedure ensures that the measure reflects reconstruction difficulty for the actual information set reported by each firm rather than artifacts of imputation. The resulting portfolio sorts, reported in Table C.17 in the Appendix, show that the return pattern remains largely unchanged. The equal-weighted

high-minus-low (H–L) spread is -1.54 percent per month in excess returns ($t = -4.79$) and remains economically and statistically significant at -0.90 percent ($t = -4.89$) under the FF6 model. The value-weighted results are smaller in magnitude (-0.77 percent, $t = -2.31$ in excess returns) but display the same monotonic decline across deciles. These results demonstrate that the negative relation between ATYP and future returns is not driven by the imputation of missing data.

Second, I regress the original ATYP measure on the share of imputed characteristics for each firm-month and extract the orthogonal residual as a new measure that is purged of any linear association with data sparsity. This residual measure isolates the component of ATYP unrelated to missingness in the underlying features. Portfolio sorts based on this residual measure, presented in Table C.18 in the Appendix, yield qualitatively similar results. The equal-weighted H—L spread remains -1.10 percent per month ($t = -4.24$) and -0.70 percent ($t = -4.06$) under the FF6 specification, while the value-weighted spread is -0.40 percent ($t = -1.49$) in excess returns and directionally consistent with the baseline estimates. The persistence of economically meaningful and statistically significant return differentials after controlling for the share of imputed features confirms that the ATYP measure captures genuine variation in firms' information environments rather than artifacts of data completeness. Overall, both exercises show that the pricing effect of information-processing difficulty is robust to alternative treatments of missing data.

4.7.10 Impact of extreme observations

I argued in Section 4.2 that a firm may have entirely unremarkable values for each individual characteristic, yet the combination of those values may be so rare compared to other firms that it defies the patterns investors expect. A potential concern with that is that the ATYP measure may reflect the presence of extreme firm characteristics rather than genuine atypical combinations of otherwise regular features. Firms with highly unusual individual characteristics are often more difficult to process, and this difficulty may mechanically increase reconstruction errors even when the joint configuration of characteristics is not itself unusual. To examine this possibility, I test whether the pricing power of ATYP persists after removing the influence of extreme observations in the cross-section of characteristics.

For each month and characteristic, I assign firms to deciles based on their realized

values. I then count the number of characteristics for each firm that fall into either the bottom or top decile of their respective cross-sectional distributions. This count represents the degree to which a firm displays extreme realizations of observable characteristics. I then follow the procedure in Section 4.7.9 and regress the original ATYP measure on this count and extract the residual component. This residual measure isolates the variation in ATYP that is orthogonal to extremeness in the underlying characteristics. By construction, it captures the atypicality that does not arise simply because a firm lies in extreme tails of individual characteristics, but rather reflects the distinctiveness of the joint configuration of these characteristics.

Table C.19 in the Appendix presents univariate portfolio sorts based on this orthogonalized ATYP measure. Panel A displays results for equal-weighted portfolios and Panel B for value-weighted portfolios. The return differentials decline in magnitude compared to the baseline results, which is expected because firms with many extreme characteristics are indeed harder to process and naturally fit into the ATYP measure. Nevertheless, the pattern of returns remains monotonic and economically meaningful. The equal-weighted high-minus-low spread is -0.81 percent per month in excess returns ($t = -2.53$) and remains significant after controlling for standard factor models. Under the Fama and French six factor specification, the spread is -0.47 percent ($t = -3.67$). The value-weighted results also show a negative and statistically meaningful return spread, though with smaller magnitudes consistent with the baseline findings.

Interestingly, after orthogonalizing ATYP with respect to the share of extreme characteristics, the statistical significance of the value-weighted spreads increases relative to the baseline. This pattern arises because the orthogonalization reduces the influence of very small firms, where both information-processing difficulty and extremeness are concentrated. As a result, variation in the residual ATYP measure becomes more pronounced among larger firms, increasing the ability of value-weighted portfolios to detect the pricing effect. The results indicate that firms with atypical combinations of otherwise moderate characteristics continue to exhibit strong predictive power in the cross-section, consistent with a broader information-processing frictions interpretation.

4.7.11 International robustness

To examine whether the pricing of ATYP is a uniquely U.S. phenomenon or a broader feature of global markets, I extend the analysis to developed countries outside the U.S.

This test provides an important check on the external validity of my findings, as market structure, disclosure regimes, and investor composition differ across regions. Global stock returns and firm characteristics are obtained from Jensen et al. (2023), and regional factor portfolios are sourced from the Kenneth R. French Data Library. I follow the Library's country classifications for both developed and European countries.[14] I begin my sample in 1990, since many firm characteristics are unavailable prior to that year for several developed markets. All data are processed following the baseline specification in Section 4.3.1, and I form portfolios and compute alphas exactly as in the U.S. baseline in Section 4.5.1.

Table C.20 in the Appendix reports results for all developed markets excluding the U.S. Equal-weighted sorts reveal a clear and statistically significant negative relation between ATYP and future returns. The high–low portfolio earns -0.65% per month in excess returns ($t = -1.96$) and remains significant after adjusting for standard factors, with CAPM and FF3 alphas of -0.99% ($t = -4.72$) and -0.88% ($t = -4.33$). Although magnitudes decline under FF5 and FF6, the spreads remain economically meaningful at -0.44% per month. Value-weighted portfolios exhibit the same pattern with smaller coefficients. When Japan is excluded, the ATYP premium strengthens, producing alphas comparable to those observed in the U.S. sample. This pattern aligns with prior evidence that several established anomalies, such as momentum (Asness, 2011) and profitability or investment effects (Fama and French, 2017), are notably weaker in Japan than in other developed markets.

Because Japan's distinct market structure can obscure cross-country patterns, I next examine Europe separately. Table C.21 shows that ATYP remains strongly priced across European equities. The equal-weighted high–low spread averages about -1.21% per month in excess returns ($t = -4.40$) and retains statistical significance across all factor models, with FF6 alphas of roughly -0.55% ($t = -3.22$). Value-weighted portfolios yield similarly robust results. The European evidence thus confirms that ATYP commands a negative return premium across advanced markets, and that the weaker developed-market aggregate primarily reflects Japan's idiosyncratic dynamics rather than a failure of the ATYP measure. Overall, these findings underscore that the pricing of information-processing difficulty is not confined to U.S. data but represents a pervasive global pattern.

[14] See the French Data Library documentation for details on factor and portfolio construction.

4.8 Conclusion

This paper introduces ATYP, a new statistical measure of firm-level atypicality grounded in the joint structure of firm characteristics. ATYP is constructed as the root mean squared reconstruction error from an autoencoder trained to learn the systematic patterns that describe most firms. By capturing the extent to which a firm deviates from these patterns, ATYP provides a continuous and data-driven proxy for information-processing frictions that is distinct from disclosure-based or organizational measures.

My validation results show that ATYP reflects meaningful informational challenges. High-ATYP firms exhibit greater analyst disagreement, higher return volatility, larger absolute earnings surprises, and lower analyst accuracy—patterns consistent with greater uncertainty and more difficult interpretation. These firms also display slower price discovery, including stronger and more persistent post-earnings-announcement drift and delayed reactions to industry shocks, indicating that atypical information diffuses more slowly into prices.

I document a strong and robust cross-sectional relation between ATYP and future stock returns. A portfolio long in low-ATYP firms and short in high-ATYP firms earns 1.47% per month when equal-weighted and 0.82% when value-weighted. These return differences remain large after controlling for standard risk factors and established return predictors. In multivariate Fama–MacBeth regressions with an extensive set of controls, ATYP continues to load negatively and significantly, and the results hold across data sources, model specifications, and alternative constructions of the measure.

The evidence points to mispricing as the source of the ATYP premium. Atypical firms are harder to interpret, which increases disagreement and amplifies behavioral biases such as overconfidence. Combined with short-sale constraints, this disagreement pushes prices upward. Limited attention further slows the incorporation of firm-specific signals, while limits to arbitrage restrict the ability of sophisticated investors to correct the resulting mispricing. Consistent with this mechanism, the ATYP premium is concentrated in settings with high limits to arbitrage and low investor attention.

Taken together, these results indicate that ATYP is not merely a background feature of information environments but a systematic driver of prices. ATYP provides a unified way to capture information-processing difficulty in the cross-section, linking theoretical ideas about bounded information processing to empirical patterns in returns. For researchers,

it offers a flexible tool for studying how the structure of firm information shapes investor behavior. For investors, it highlights a strong relation between atypicality and expected returns. For policymakers, it emphasizes that informational frictions have tangible consequences for capital allocation.

Appendix A

Appendix to Chapter 2

A.1 Proofs

A.1.1 Proof of Proposition 1

For this proof consider a second-order Taylor expansion of the CRRA utility function around the expected portfolio return:

$$\mathbb{E}[u(r_{p,t+1}(\theta))] \approx \mathbb{E}[r_{p,t+1}(\theta)] - \frac{\gamma}{2}\mathbb{E}[r_{p,t+1}(\theta)^2]. \tag{A.1}$$

This approximation is well-known and shows that the CRRA utility framework naturally reduces to mean-variance preferences, where γ represents risk aversion and directly scales the penalty on portfolio return variance. We can express returns as specified in Equation (2.5). For mean-variance utility, this yields the optimization problem:

$$\max_{\theta} \theta^T \hat{\mu}_c - \frac{\gamma}{2}\theta^T \hat{\Sigma}_c \theta - \gamma \theta^T \hat{\sigma}_{bc}, \tag{A.2}$$

with first-order condition:

$$\theta^* = \frac{1}{\gamma}\hat{\Sigma}_c^{-1}\hat{\mu}_c - \hat{\Sigma}_c^{-1}\hat{\sigma}_{bc}. \tag{A.3}$$

Plugging in the optimal coefficients θ^* into Equation (2.4) yields Proposition 1.

A.1.2 Proof of Proposition 3

For this proof consider the definition of the EDF in Equation (2.9). We define the two key matrices as:

$$G = \mathbb{E}\left[\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\right], \tag{A.4}$$

and

$$V = \mathbb{E}\left[\frac{\partial L(\theta)}{\partial\theta}\frac{\partial L(\theta)}{\partial\theta'}\right]. \tag{A.5}$$

Under mean-variance utility from Equation (A.2):

$$G = \frac{1}{T}\gamma\hat{\Sigma}_c, \tag{A.6}$$

$$\frac{\partial L(\theta)}{\partial\theta} = r_{c,t+1} - \gamma\hat{\Sigma}_c\left(\frac{1}{\gamma}\hat{\Sigma}_c^{-1}\hat{\mu}_c - \hat{\Sigma}_c^{-1}\hat{\sigma}_{bc}\right) - \gamma\hat{\sigma}_{bc} = r_{c,t+1} - \hat{\mu}_c, \tag{A.7}$$

and

$$V = \frac{1}{T}\left(r_{c,t+1} - \hat{\mu}_c\right)^T\left(r_{c,t+1} - \hat{\mu}_c\right) = \hat{\Sigma}_c. \tag{A.8}$$

Therefore, our EDF measure simplifies to

$$\text{EDF} = \text{tr}(G^{-1}V)/T = \frac{1}{\gamma}\text{tr}(\hat{\Sigma}_c^{-1}\hat{\Sigma}_c) = \frac{p}{\gamma}, \tag{A.9}$$

where p denotes the number of characteristics. This leads to our key result about model complexity in Proposition 3.

A.2 Neural network configuration

Our benchmark model consists of an input layer, three to five hidden layers and an output layer. We apply the geometric pyramid rule (Masters, 1993), i.e., the first hidden layer consists of 32 nodes, and for each subsequent layer, the number of units is halved.

At each node of the network, a linear transformation of the preceding outputs is fed into an activation function. We choose to use the leaky rectified linear unit (leaky ReLU) activation function at every node:

$$R(z) = \begin{cases} z & \text{if } z > 0 \\ \alpha z & \text{otherwise} \end{cases}, \tag{A.10}$$

where z denotes the input and α denotes some small non-zero constant, in our case 0.01. ReLU is the most popular activation function because it is cheap to compute, converges fast and is sparsely activated. The disadvantage of transforming all negative values to zero is a problem called "dying ReLU". A ReLU neuron is "dead" if it is stuck in the negative range and always outputs zero. Since the slope of ReLU in the negative range is also zero, it is unlikely that a neuron will recover once it goes negative. Such neurons play no role in discriminating inputs and are essentially useless. Over time, a large part of the network may do nothing. Leaky ReLU fixes this problem because it has small slope for negative values instead of a flat slope. Moreover, we shift the activation function at every node in every hidden layer by adding a constant. This is commonly referred to as bias in the machine learning literature.

Our benchmark network is estimated by minimizing the loss function (utility function) given in Equation (2.11). To do so, we apply the commonly used ADAM stochastic gradient descent optimization technique developed by Kingma and Ba (2014). One technical detail in our implementation involves handling missing firm-date observations in our three-dimensional input tensor (see Figure 2.2). Since not every firm is observed at every point in time, some entries in the tensor are missing. To maintain a consistent tensor shape for computational purposes, we fill these missing entries with zeros. However, because these zeros do not represent actual data, we add a masking layer to our network. This layer ensures that any firm missing in a particular month is excluded from the utility calculation, so that only real observations contribute to the model's performance.

To control for the nonlinearity and heavy parametrization of the model, we employ different regularization techniques to prevent overfitting: first, as mentioned in Section 2.3.1, we impose constraints on portfolio weights, i.e., the absolute portfolio weight of a single stock cannot exceed $|3\%|$ and the portfolio leverage in any period cannot exceed 100%. The weight constraint is imposed via an additional penalty term in the objective function. Specifically, we augment the utility function with a term of the form

$$\lambda \sum max(0, |w_i| - 0.03), \tag{A.11}$$

where λ is a large positive constant. This penalty becomes positive when any weight w_i exceeds ± 0.03. Effectively, this forces the network to keep $|w_i|$ below 0.03 in order to avoid incurring a large penalty in the objective function.

In practice, we set λ to be sufficiently large so that the violation of the constraint becomes very costly, while still allowing the optimizer enough flexibility to search for the best feasible solution. Hence, although our implementation is not a "hard" constraint in the sense of truncating outputs directly, in our estimations the model learns to remain strictly within $|w_i| < 0.03$. We verified empirically that none of the weights exceed this bound during optimization or at convergence. We implement the constraint on leverage analogously.

Second, we add a lasso (l_1) penalty term to the loss function to be minimized. Adding the penalty implies a potential shrinkage of coefficients towards zero. This in turn reduces the variance of the prediction, i.e., prevents overfit of the model.

Third, we employ early stopping on the validation data. Early stopping refers to a very general regularization technique. At each new iteration, predictions are estimated for the validation sample, and the loss (utility) is constructed. The optimization is terminated when the validation sample loss starts to increase by some small specified number (tolerance) over a specified number of iterations (patience). Typically, the termination occurs before the loss is minimized in the training sample. Early stopping is a popular regularization tool because it reduces the computational cost.

Fourth, we implement a dropout layer before the first hidden layer (Srivastava et al., 2014). The basic idea of dropout is to randomly remove units (and their connections) from the neural network during training. This prevents the units from becoming too similar. During training, samples are taken from an exponential number of different thinned networks. At test time, it is easy to approximate the effect of averaging the predictions of all these thinned networks by simply using a single, unthinned network with smaller weights. The combination of a dropout layer, l_1-regularization and early stopping tremendously helps to reduce overfitting and model complexity.

Finally, we adopt our own version of a batch normalization algorithm (Ioffe and Szegedy, 2015). In general, training deep neural networks is complicated by the fact that the distribution of inputs to each layer changes during training as the parameters of the previous layers change. This phenomenon is referred to as internal covariate shift and can be remedied by normalizing the layer inputs. The strength of this method is that normalization is part of the model architecture and is performed for each training mini-batch. Batch normalization allows much higher learning rates to be used and less care to be taken in initialization. Brandt et al. (2009) standardize characteristics

cross-sectionally to have zero mean and unit standard deviation across all stocks at date t. Hence, the model predictions represent deviations from the benchmark portfolio. However, applying the aforementioned activation function destroys this structure. In our model each observation can be interpreted as a complete cross-section (e.g., a batch size of 12 refers to 12 complete cross-sections of data). However, the model of Brandt et al. (2009) requires normalization on a cross-sectional level instead of a batch level. Thus, we employ our own version of cross-sectional normalization after applying the activation function in each hidden layer, such that the output of each node in the hidden layer is standardized cross-sectionally to have zero mean and unit standard deviation across all stocks at date t. Hence, the output of each node in each hidden layer can also be interpreted as a deviation from the benchmark portfolio.

We provide a summary of the relevant hyperparameters in Table B.2. Models are estimated using TensorFlow via Keras. The estimation of a single DPPP model including hyperparameter tuning takes about six hours with a standard retail GPU.

Table A.1: Hyperparameters

	PPP	DPPP
L1 penalty	$\lambda \in \{0, 10^{-5}, 10^{-3}\}$	$\lambda \in \{0, 10^{-5}, 10^{-3}\}$
Learning Rate	0.001	0.001
Dropout	0	$D \in \{0, 0.2, 0.4\}$
Batch Size	60	60
Epochs	200	200
Patience	30	30
Hidden Layers	–	$H \in \{3, 4, 5\}$
Leaky ReLU	–	0.01

This table gives the hyperparameters that we tune. The first column shows the hyperparameters for the linear Parametric Portfolio Policy (PPP). The second column shows the hyperparameters for the Deep Parametric Portfolio Policy (DPPP). For the DPPP, we start with 32 units in the first hidden layer, and for each subsequent layer, the number of units is halved.

For our estimation strategy, we follow Brandt et al. (2009) and Gu et al. (2020) and use an expanding window strategy to generate out-of-sample results. More specifically, we split our data into a training sample used to estimate the model, a validation sample used to tune the hyperparameters of the model, and a test sample used to evaluate the out-of-sample performance of the model.

We initially train the model on the first 20 years of the dataset, validate it on the following five years, and evaluate its out of-sample-performance on the 12 months

following the validation window. We then recursively increase the training sample by one year. Each time the training sample is increased, we refit the entire model while holding the size of the validation and test window fixed. The result is a sequence of out-of-sample periods corresponding to each expanding window, in our case 25 in total. This corresponds to a total out-of-sample period of 300 months. Note that this approach ensures that the temporal ordering of the data is maintained. The testing strategy is depicted graphically in Figure A.1.

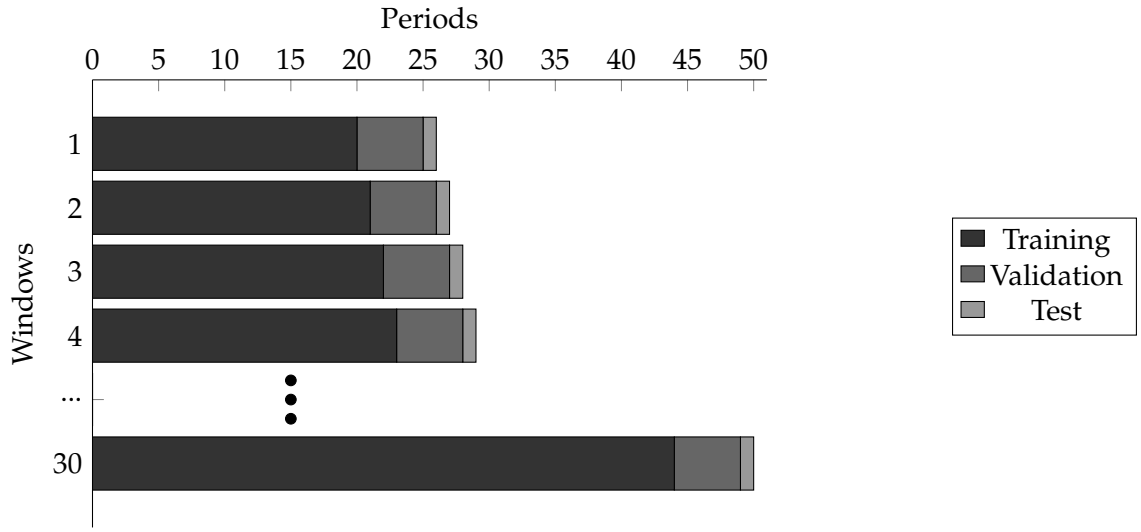


Figure A.1: Out-of-sample testing strategy
This figure presents our out-of-sample testing strategy. We recursively increase our training window, presented by the black portion of each bar, while holding the validation and the test window constant, presented by the grey portions of each bar.

A.3 Loss aversion as economic regularization

Building on Tversky and Kahneman (1992), consider a loss-averse investor with the piecewise-linear utility function as in section 2.4:

$$u(r_{p,t+1}) = \begin{cases} -l(\overline{W} - r_{p,t+1})^b & \text{if } r_{p,t+1} < \overline{W} \\ (r_{p,t+1} - \overline{W})^b & \text{otherwise} \end{cases}, \tag{A.12}$$

where l measures the strength of loss aversion relative to gains, and $\overline{W}$ is a reference return (here $\overline{W} = 0$) and the parameter b captures the degree of risk seeking over losses and risk aversion over gains (here $b = 1$).

When l is large, negative deviations $1 + r_{p,t+1} < \overline{W}$ are penalized heavily. This parallels the way high risk aversion γ suppresses active exposures in mean-variance and CRRA frameworks: high l shrinks portfolios toward safer strategies that minimize downside realizations. Hence, loss aversion naturally acts as an economic regularizer against overfitting predictive signals, just as high γ penalizes variance in mean-variance or CRRA utility.

Consider an investor's optimization problem

$$\max_{\theta} \ \mathbb{E}\big[u_(r_{p,t+1}(\theta))\big], \tag{A.13}$$

where $r_{p,t+1}(\theta)$ denotes the portfolio return. We decompose the expected utility:

$$\mathbb{E}[u_l(r_{p,t+1}(\theta))] = \mathbb{E}\big[(r_{p,t+1} - \overline{W})\,\mathbf{1}\{r_{p,t+1} \geq \overline{W}\}\big] \ + \ l\,\mathbb{E}\big[(r_{p,t+1} - \overline{W})\,\mathbf{1}\{r_{p,t+1} < \overline{W}\}\big]. \tag{A.14}$$

As $l \to \infty$, the second term dominates unless the probability of shortfalls $\{r_{p,t+1} < \overline{W}\}$ is driven toward zero. Consequently, the optimal portfolio θ^* converges to a solution that minimizes downside risk.

If a near riskless benchmark b_t with return $r_{b,t+1} \approx \overline{W}$ exists, active deviations eventually go to zero, mirroring the high-γ limit in Section 2.2.2. Formally,

$$\lim_{l \to \infty} \theta^* \ \to \ \arg\min_{\theta} \ \mathbb{E}\Big[(r_{p,t+1}(\theta) - \overline{W})\,\mathbf{1}\{r_{p,t+1}(\theta) < \overline{W}\}\Big]. \tag{A.15}$$

Hence, infinite loss aversion collapses the portfolio to a baseline strategy that nearly

eliminates downside deviations, thereby reducing active complexity in the limit.

If $r_{b,t+1}$ is a benchmark with limited downside, then the optimal deviation from b_t converges to zero in the loss-aversion limit:

$$\lim_{l\to\infty} \left\| \theta^* \right\| = 0, \tag{A.16}$$

and the investor asymptotically holds the benchmark.

The above limit behavior implies an effective shrinkage of parameters in parametric portfolio policies or more flexible deep portfolio policies. In analogy to the CRRA and mean-variance case (see Section 2.2.2), one can interpret l as scaling the penalty on negative outcomes. Large l forces the model to reduce the probability of downside, thereby reducing the susceptibility to overfitting in high-dimensional or nonlinear representations.

Although the piecewise function is not globally differentiable (making a simple closed-form effective degrees of freedom (EDF) derivation more involved), the economic intuition is clear: as $l \to \infty$, any parameter that increases downside risk is curtailed. This downside penalty channels into the Hessian of the loss, compressing active exposures in a manner analogous to L2 regularization, but founded on investor preferences rather than a purely statistical criterion.

A simple simulation parallel to Section 2.2.2 confirms that as l increases, we see a convergence in portfolio weights. The results are depicted in Figure A.2. Thus, loss aversion plays a role akin to a built-in regularization term that trims complexity to mitigate estimation risk.

Our findings complement the results in Section 2.2.2: while risk aversion γ penalizes variance, loss aversion l penalizes negative deviations, yet both yield parallel shrinkage outcomes in the respective limits $\gamma \to \infty$ or $l \to \infty$. Thus, behavioral preferences such as loss aversion can be viewed as alternative economic mechanisms that align well with the notion of regularization against overfitting.

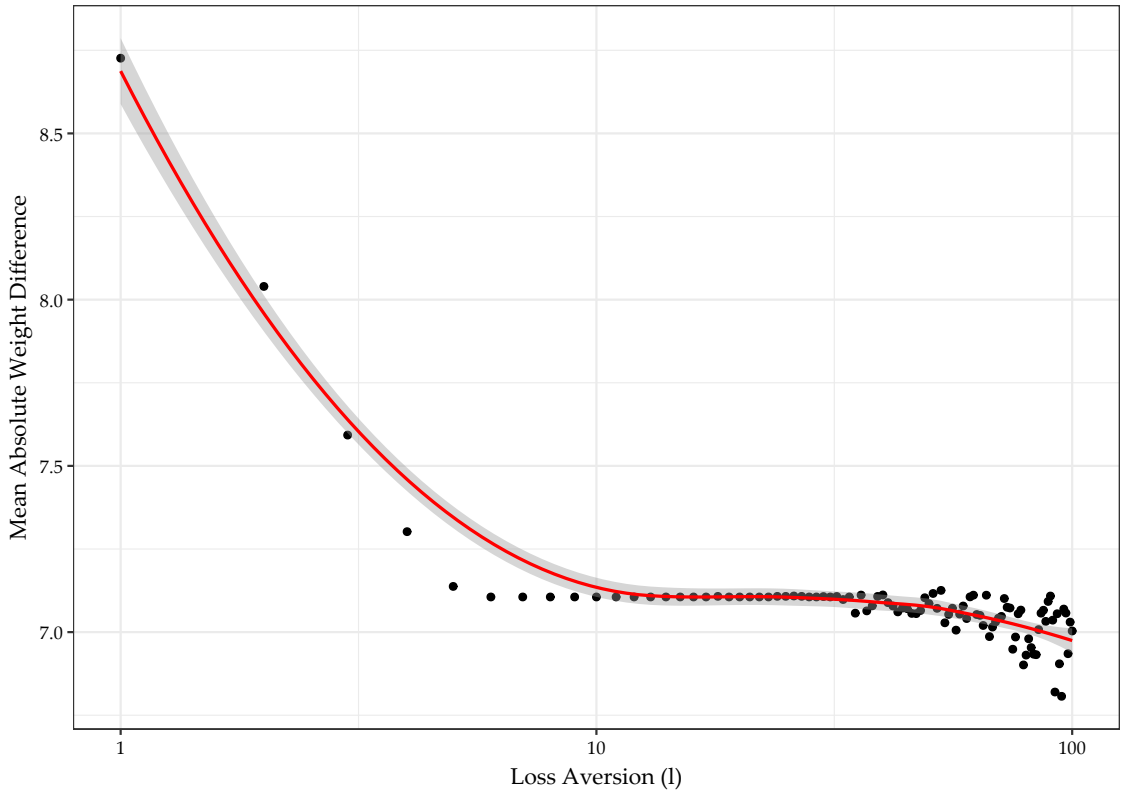


Figure A.2: Loss aversion as economic regularization
This figure presents simulation evidence demonstrating that risk aversion acts as an economic regularization mechanism. The simulation compares two nested parametric portfolio policies of different complexity: one using 10 characteristics (PPP) and another using 100 characteristics (DPPP) constructed through random Fourier transformations of the base characteristics. The figure shows the mean absolute difference in portfolio weights between models across loss aversion levels. All panels use a logarithmic scale (base 10) for loss aversion.

A.4 Additional results and extensions

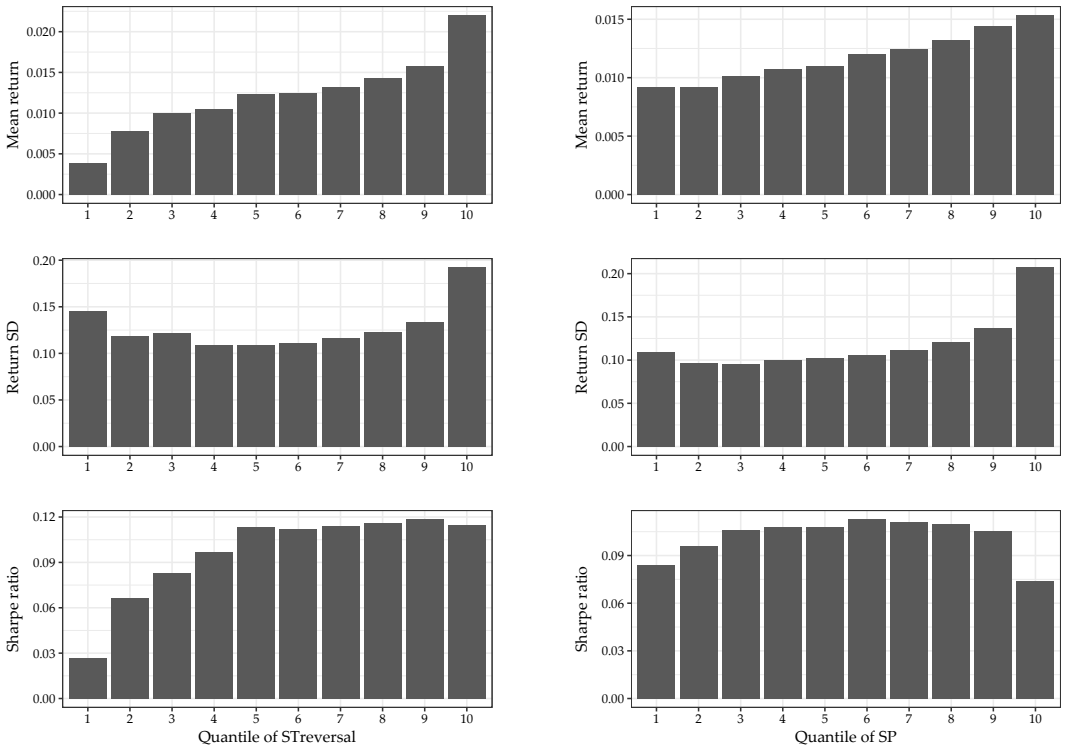
A.4.1 Partial dependence and surrogate models

The importance of nonlinear modeling of portfolio weights becomes evident when considering an investor who trades off mean return against return volatility. The investor uses standard one-dimensional portfolio sorting techniques as pictured in Figure A.3. Decile portfolios formed on short-term reversal or sales-to-price display monotonically increasing mean return.[1] At the same time, the standard deviations of decile portfolios are nonlinear in deciles, with top and bottom decile portfolios having high standard deviations. This leads to extreme portfolios having comparatively low Sharpe ratios relative to decile portfolios in the middle of the distribution. A (long-only) investor would

[1]We picked these two variables for illustrative purposes as these variables are the most important return- and fundamental-based variables in Gu et al. (2020).

therefore potentially be indifferent between investing in any portfolio in the upper half of the short-term reversal distribution, and she would prefer to invest in portfolios in the middle of the sales-to-price distribution rather than investing in the extreme portfolios. nonlinear portfolio policies are able to capture these kinds of relationships.

In our application, understanding the estimated relation between input (firm characteristics) and output (estimated portfolio weights) is essential in order to shed light on the relation between firm characteristics and utility. Moreover, such an understanding allows us to compare our results to the existing literature. We provide two additional ways of interpreting the nonlinearity in our models.



(a) Short-Term Reversal **(b)** Sales-To-Price

Figure A.3: Mean returns, standard deviations and Sharpe ratios of one-dimensional portfolio sorts
Mean returns, standard deviations and Sharpe ratios of decile portfolios sorted on short-term reversal (left panel) and sales-to-price ratio (right panel). Data is from Chen and Zimmermann (2022) and spans from 1925 to 2021.

Partial dependence

We evaluate the sensitivity of the model output to each variable. Typically, partial dependence plots provide an assessment of the variables of interest over a range of values. At each value of the variable, the model is evaluated while the remaining variables remain

unchanged, and the results are then averaged across the cross-section. However, since the sum of all weights in each cross-section is equal to one and thus the mean weight prediction is always the same, applying this method to parametric portfolio policies does not yield reasonable results. To address this, we apply our own algorithm: when assessing the sensitivity with respect to variable k, we compute two sets of predictions, one with all features and another where feature k is set to zero (equivalent to its mean in our standardized setting). The difference between these predictions represents the marginal contribution of variable k to the portfolio weights. We then plot these marginal contributions against the actual values of feature k to understand how the feature's impact varies across its range. We interpret this difference in predicted weights conditional on values of k as the marginal sensitivity of weights (i.e., its partial dependence) with respect to k, which allows us to assess both the magnitude and direction of each variable's influence on the portfolio allocation decision.

Figure A.4 depicts how different characteristics contribute to the DPPP's portfolio allocation decisions. We assess this by computing the marginal contribution of each characteristic, defined as the difference between predictions using all features and predictions where the respective feature is set to zero (its mean in our standardized setting). We examine the sensitivity with respect to three fundamental variables, namely the book-to-market ratio (BM), liquid assets (cash), and quarterly return on assets (roaq), as well as an analyst variable, namely earnings forecast revisions per share (AnalystRevision), and four past return-based variables, namely 12-month momentum (Mom12m), short-term reversal (STreversal), seasonal momentum (MomSeason), and intermediate momentum (IntMom). Recall that each predictor is signed, so that a larger value implies a higher expected return. To assess whether the marginal association of the deep model is more in line with the actual risk and return associated with each characteristic than a linear model, we include the overall Sharpe ratio for each decile portfolio sorted on each of the characteristics.

The results reveal distinct patterns in how the DPPP utilizes different characteristics. In line with the findings in regards to importance, short-term reversal exhibits the most pronounced marginal effect, as indicated by the steepness of the depicted relationship, with its impact on portfolio weights varying substantially across its range and risk aversion levels. This effect is particularly strong for low risk aversion ($\gamma = 2$), where extreme values of STreversal trigger the largest portfolio weight adjustments. The strong

response aligns with the monotonically increasing Sharpe ratios across STreversal deciles, suggesting the model effectively captures this signal's risk-return profile.

Most characteristics exhibit nonlinear marginal contributions, though their magnitude and patterns differ notably. BM, for instance, shows minimal impact in lower deciles but increasingly affects portfolio weights in higher deciles, particularly under lower risk aversion. This pattern partially reflects the underlying Sharpe ratio profile of BM-sorted portfolios. In contrast, momentum variables (Mom12m, MomSeason, IntMom) display more modest marginal effects, suggesting they serve a more supplementary role in the portfolio allocation decision.

Risk aversion systematically influences how the DPPP incorporates characteristic information. Higher risk aversion levels generally lead to more muted marginal effects, as evidenced by the flatter curves for $\gamma = 10$ and $\gamma = 20$. This finding suggests that as risk aversion increases, the model takes a more conservative approach to characteristic-based tilts, consistent with theoretical expectations about risk-return tradeoffs. It also confirms the reasoning that increasing risk aversion leads to a decrease in model complexity.

The relationship between Sharpe ratios of characteristic-sorted portfolios and marginal contribution patterns varies across characteristics. While some characteristics like STreversal show strong alignment between Sharpe ratios and marginal contributions, others exhibit more complex relationships. This suggests that the DPPP captures both direct risk-return relationships and potentially more sophisticated interactions between characteristics in its portfolio allocation decisions.

Surrogate model

We evaluate the extent to which nonlinearity contributes to the estimated DPPP. Put differently, we assess the extent to which different forms of nonlinearity play a role when optimizing portfolios conditional on firm characteristics. To do so, we estimate a sequence of increasingly complex surrogate models. First, we regress the out-of-sample weight predictions from the DPPP on all firm characteristics in a linear model. This allows us to assess the extent to which simple linear relationships explain the predicted weights. In a next step, we estimate a second surrogate model that includes all possible two-way interactions between variables. The incremental explanatory power of this model captures the importance of variable interactions in the DPPP's portfolio allocation decisions. We attribute the remaining unexplained portion of predicted DPPP weights to nonlinearities in the functional form. To assess the economic significance of these

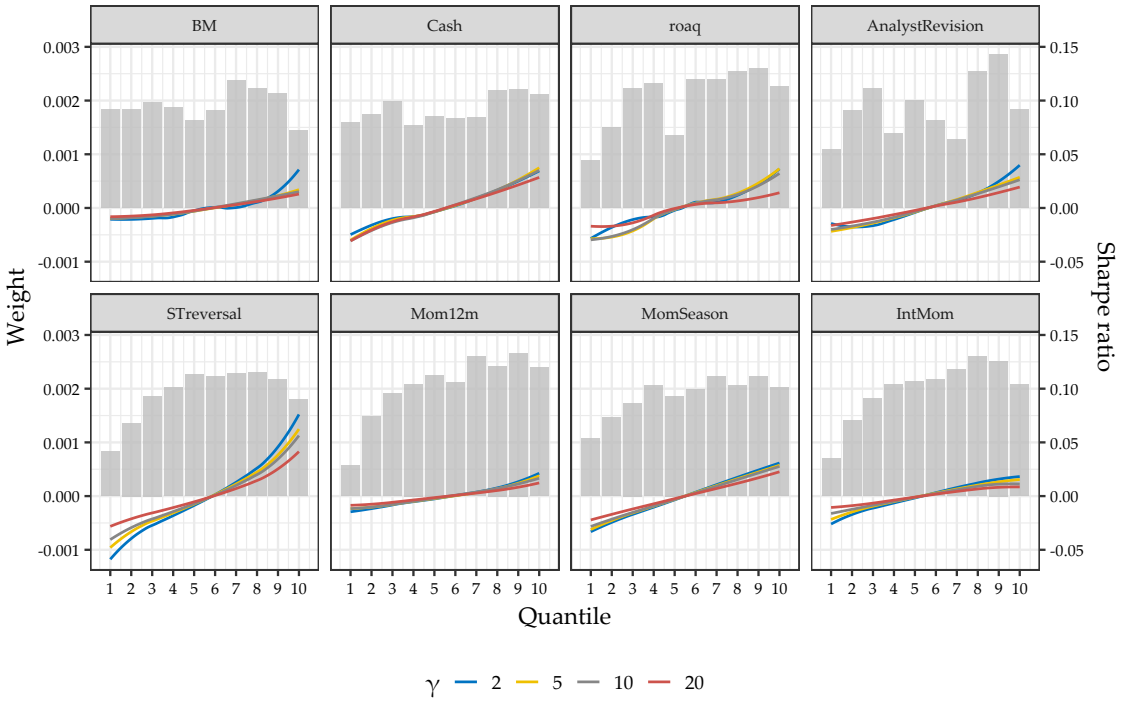


Figure A.4: Marginal contribution of characteristics to portfolio weights in the DPPP
This figure shows the sensitivity of predicted weights (left vertical axis) with respect to values of the respective variable (horizontal axis) across different risk aversions γ. The marginal contribution is computed as the difference between predictions with all features and predictions where the respective characteristic is set to zero. The aforementioned relationship is depicted by curves, fitted via local polynomial regressions. The figure also includes bars, depicting the Sharpe ratio (right vertical axis), per variable decile (horizontal axis).

nonlinearities, we additionally examine the certainty equivalent differences of the ex-post fitted surrogate models compared to the actual model during the out-of-sample periods.

Figure A.5 shows both the adjusted R^2s of linear surrogate models for the out-of-sample predicted weights and the resulting differences in certainty equivalents across different levels of risk aversion. For each risk aversion level, we estimate two surrogate models: a simple linear model and an extended version that includes all possible two-way interactions between the 50 most important characteristics.

The results strongly support our theoretical findings that higher risk aversion reduces model complexity and acts as an economic regularization parameter, i.e., that the importance of nonlinearity varies with the degree of risk aversion. The simple linear surrogate model explains about 30-40% of the variation in predicted portfolio weights for $\gamma = 2$, while the R^2 ranges between 40-60% for higher degrees of risk aversion. This underscores that risk aversion acts as an economic regularization parameter, reducing model complexity. Adding interactions substantially improves the model fit. The R^2 increases by

approximately 20-30 percentage points across all degrees of risk aversion when including two-way interactions. Moreover, we observe that models with interactions show more stable explanatory power over time, as evidenced by less fluctuation in R^2 across periods.

The economic significance of these nonlinearities is assessed through the certainty equivalent differences between the DPPP and the surrogate models. The right panel of Figure A.5 reveals that the economic impact of nonlinearities is most pronounced for lower levels of risk aversion. For $\gamma = 2$ and $\gamma = 5$, we observe certainty equivalent differences of up to 200-300 basis points in some periods, indicating substantial economic value from the DPPP's nonlinear portfolio allocation decisions. This effect diminishes with increasing risk aversion, as shown by the smaller and more stable certainty equivalent differences for $\gamma = 10$ and $\gamma = 20$. Adding interactions to the surrogate models generally reduces these certainty equivalent differences, suggesting that a significant portion of the DPPP's economic value comes from capturing interaction effects between characteristics.

Based on these findings, we can decompose the DPPP's portfolio allocation decisions as follows: approximately 30-60% of the underlying characteristic-weight relationship is linear in nature, depending on the degree of risk aversion. An additional 20-30% can be captured by two-way interactions, while the remaining 10-50% can be attributed to higher-order nonlinearities in the DPPP model. The economic significance of these nonlinear components, as measured by certainty equivalent differences, is most pronounced for lower risk aversion levels and during periods of market stress, suggesting that the DPPP's flexibility in capturing complex relationships becomes particularly valuable under these conditions. This empirical evidence strongly supports our theoretical framework showing how risk aversion serves as an economic regularization mechanism that naturally constrains model complexity.

A.4.2 Rolling window estimation

For additional robustness, we consider a rolling-window estimation procedure of fixed length 20 years (240 months). At each date t, we estimate our model parameters using the most recent 240 months of data, keeping five years (60 months) of validation data and then form one-period-ahead forecasts of portfolio weights. Compared to the baseline expanding-window approach, this rolling scheme is designed to adapt more readily to potential structural changes in the data by discarding older observations.

Table A.2 presents out-of-sample results for the DPPP under both rolling- and

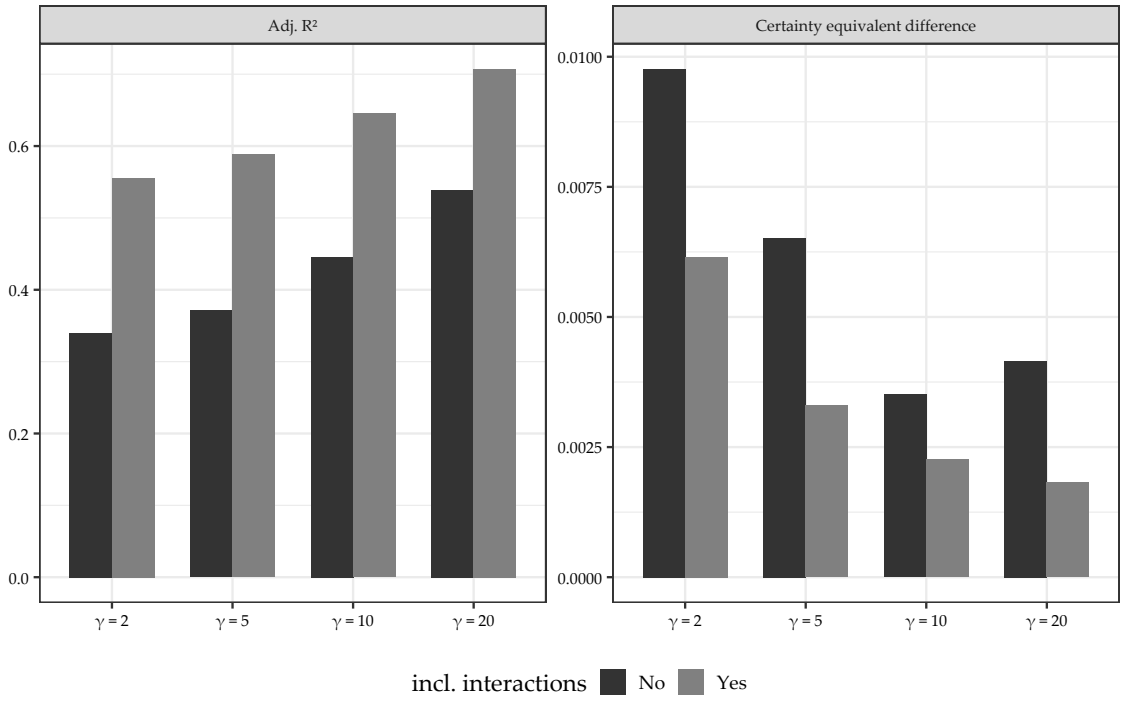


Figure A.5: Surrogate R^2 and difference in certainty equivalent in the DPPP
This figure depicts two key metrics evaluating the complexity and economic significance of the DPPP's portfolio allocation decisions. The left panel shows the adjusted R^2 of linear surrogate models fitted to the DPPP's weight predictions and the right panel displays the certainty equivalent differences between the DPPP and the surrogate models. More specifically, the bars show the R^2 and certainty equivalent differences for a linear surrogate model of the estimated weights by the deep models on the 50 most important variables in each model for all out-of-sample periods and across different risk aversion levels. Interactions include all possible two-way interactions between the variables.

expanding-window estimation, using the same 157 firm-level characteristics. The table reports certainty equivalent returns for investors with CRRA preference coefficients $\gamma \in \{2, 5, 10, 20\}$. While the rolling-window approach can be more responsive to recent market conditions, our results indicate that it does not substantially outperform the expanding-window method in terms of certainty equivalent return. Moreover, for the most risk-averse investors ($\gamma = 20$), the baseline expanding-window approach achieves statistically higher certainty equivalent return compared to rolling-window estimation. We hypothesize that this pattern arises because the rolling procedure reduces the effective sample size, thereby increasing estimation uncertainty precisely in those regimes where the portfolio is most sensitive to parameter estimates.

Other performance metrics exhibit similar behavior. For instance, Sharpe ratios under rolling estimation are generally on par with those from the baseline method, showing little evidence of consistent improvement. Maximum drawdown and conditional value at

risk measures also do not favor one approach conclusively, reinforcing the notion that the benefits of a rolling scheme can be offset by the increased estimation variability.

In short, while a rolling-window procedure may be appealing in settings prone to regime shifts, our empirical findings suggest that its advantages over the expanding-window baseline are limited in this particular application. In fact, for high levels of risk aversion, the stability provided by a longer estimation sample can be crucial in achieving robust portfolio outcomes, aligning with our broader argument that model simplicity and regularization often yield more reliable results.

A.4.3 Adding macroeconomic variables

To investigate how our portfolio policies interact with the state of the economy, we augment our models with macroeconomic variables. Specifically, we construct eight macro predictors based on Welch and Goyal (2008): the dividend-price ratio (dp), earnings-price ratio (ep), book-to-market ratio (bm), net equity expansion (ntis), the Treasury-bill rate (tbl), the term spread (tms), the default spread (dfy), and stock variance (svar).

Following Gu et al. (2020), we introduce a transformation layer in our network that generates a new set of interaction variables by multiplying stock-level characteristics with these macro predictors. Formally, for each stock i at time t,

$$z_{i,t} = x_{i,t} \otimes c_t, \tag{A.17}$$

where $x_{i,t}$ is a $P_c \times 1$ matrix of characteristics for each stock i, and c_t is a $P_x \times 1$ vector of macroeconomic predictors (including a constant). Hence, $z_{i,t}$ is a $P \times 1$ ($P = P_c P_x$) that captures the interaction between stock-level characteristics and macro-level factors. This yields a total of $157 \times (8 + 1) = 1,413$ covariates in our model.

Table A.3 compares the results of our baseline DPPP model with those of the augmented model that incorporates the macro variables. The difference in certainty equivalent returns is statistically insignificant for common levels of risk aversion ($\gamma \in \{2, 5, 10\}$). However, for the highest risk aversion ($\gamma = 20$), the baseline model outperforms the augmented model at a statistically significant level. We attribute this result to overfitting: the large number of parameters introduced by the macro interaction terms can degrade performance in high-risk-aversion scenarios, where a simpler model better aligns with our economic regularization theory. These findings are corroborated by the Sharpe ratio

Table A.2: Deep portfolio policy for CRRA investors with different degrees of risk aversion with expanding vs. rolling window estimation

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	Expanding	Rolling	Expanding	Rolling	Expanding	Rolling	Expanding	Rolling
CE	0.0297	0.0285	0.0232	0.0209	0.0152	0.0163	0.0040	0.0002
p-value$(CE_{Expanding} - CE_{Rolling})$		0.2759		0.1405		0.2230		0.0071
$\sum \lvert w_i \rvert / N_t * 100$	0.1907	0.1914	0.1938	0.1930	0.1933	0.1907	0.1729	0.1420
$max\ w_i * 100$	1.1483	1.0082	0.9843	0.7808	0.8305	0.6637	0.4582	0.3675
$min\ w_i * 100$	-1.2824	-1.1808	-1.2053	-1.1702	-0.9743	-0.8930	-0.7224	-0.4995
$\sum w_i I(w_i < 0)$	-0.8748	-0.8795	-0.8974	-0.8912	-0.8932	-0.8751	-0.7464	-0.5234
$\sum I(w_i < 0)/N_t$	0.3400	0.3247	0.3368	0.3160	0.3319	0.3340	0.3202	0.3095
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	2.6342	2.4667	2.6022	2.3367	2.3813	2.1576	1.7516	1.2103
Mean	0.0341	0.0327	0.0305	0.0275	0.0281	0.0268	0.0224	0.0181
StdDev	0.0710	0.0655	0.0550	0.0507	0.0475	0.0432	0.0378	0.0350
Skew	2.6646	0.3897	0.8411	-0.2837	-0.2470	-0.5179	-0.5201	-0.8981
Kurt	26.4755	5.5252	10.9695	2.7665	4.0705	2.3334	1.9954	2.6755
Max DD	0.4979	0.5745	0.5601	0.5388	0.4662	0.5192	0.3027	0.3812
Max 1M loss	0.2264	0.2753	0.1789	0.2087	0.1838	0.1779	0.1446	0.1623
CVaR (95%)	0.1107	0.1157	0.0978	0.0911	0.0882	0.0815	0.0713	0.0717
SR	1.6607	1.7277	1.9230	1.8819	2.0446	2.1501	2.0491	1.7856
p-value$(SR_{Expanding} - SR_{Rolling})$		0.2777		0.4351		0.1309		0.0018
$FF5 + Mom\ \alpha$	0.0232	0.0199	0.0205	0.0159	0.0182	0.0165	0.0130	0.0089
$StdErr(\alpha)$	0.0029	0.0024	0.0024	0.0021	0.0020	0.0017	0.0016	0.0012

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics, as specified in Equation (2.1). The results show the DPPP for two different forecasting methods, namely expanding window (baseline) and rolling window. The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.3: Deep portfolio policy for CRRA investors with different degrees of risk aversion including macro variables

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	Base	+Macro	Base	+Macro	Base	+Macro	Base	+Macro
CE	0.0297	0.0273	0.0232	0.0210	0.0152	0.0156	0.0040	-0.0022
p-value$(CE_{Base} - CE_{+Macro})$		0.1058		0.1022		0.3727		0.0001
$\sum \lvert w_i \rvert / N_t * 100$	0.1907	0.1880	0.1938	0.1864	0.1933	0.1863	0.1729	0.1500
$max\ w_i * 100$	1.1483	1.2595	0.9843	1.0513	0.8305	0.8609	0.4582	0.4140
$min\ w_i * 100$	-1.2824	-1.2296	-1.2053	-1.0027	-0.9743	-0.9916	-0.7224	-0.6849
$\sum w_i I(w_i < 0)$	-0.8748	-0.8555	-0.8974	-0.8436	-0.8932	-0.8433	-0.7464	-0.5814
$\sum I(w_i < 0)/N_t$	0.3400	0.3447	0.3368	0.3464	0.3319	0.3304	0.3202	0.2854
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	2.6342	2.6562	2.6022	2.4281	2.3813	2.3192	1.7516	1.5513
Mean	0.0341	0.0308	0.0305	0.0290	0.0281	0.0271	0.0224	0.0198
StdDev	0.0710	0.0612	0.0550	0.0567	0.0475	0.0464	0.0378	0.0401
Skew	2.6646	1.0584	0.8411	0.8840	-0.2470	-0.0614	-0.5201	-0.4935
Kurt	26.4755	8.7724	10.9695	13.8198	4.0705	2.9006	1.9954	2.7715
Max DD	0.4979	0.5091	0.5601	0.6141	0.4662	0.4275	0.3027	0.4104
Max 1M loss	0.2264	0.2101	0.1789	0.2460	0.1838	0.1686	0.1446	0.1577
CVaR (95%)	0.1107	0.1045	0.0978	0.1052	0.0882	0.0826	0.0713	0.0814
SR	1.6607	1.7409	1.9230	1.7712	2.0446	2.0208	2.0491	1.7116
p-value$(SR_{Base} - SR_{+Macro})$		0.1901		0.0801		0.4413		0.0016
$FF5 + Mom\ \alpha$	0.0232	0.0221	0.0205	0.0178	0.0182	0.0166	0.0130	0.0102
$StdErr(\alpha)$	0.0029	0.0025	0.0024	0.0022	0.0020	0.0018	0.0016	0.0014

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics and eight macro variables, as well as their interactions, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

comparisons.

All other performance metrics exhibit a similar pattern, suggesting that macro variables do not substantially enhance our model's predictive power. Nevertheless, we further examine the mean absolute gradient of each macro predictor in Figure A.6. We observe that the importance of net equity expansion (ntis) increases monotonically with higher levels of risk aversion, while the other macro variables become comparatively less influential. This is notably different from the behavior of firm-level characteristics, whose relative importance tends to be more evenly distributed as risk aversion increases.

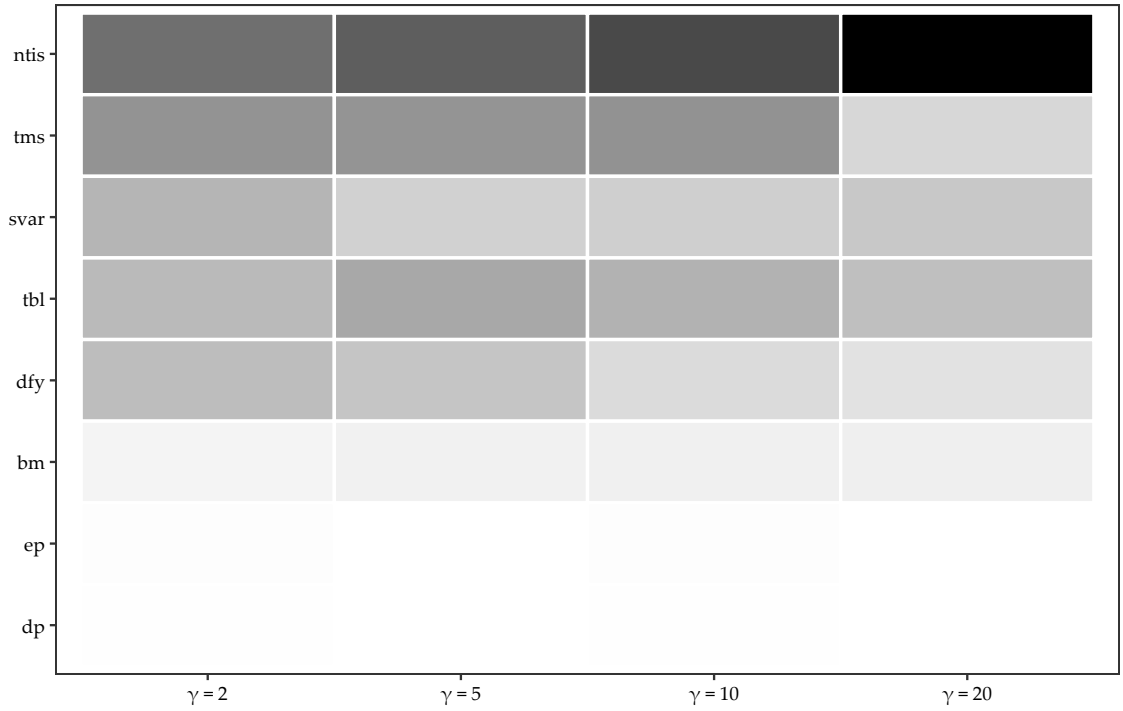


Figure A.6: Variable importance for the CRRA including macro variables for the DPPP
Variable importance for the eight macro variables in the DPPP across risk aversions. Variable importance is computed as the average absolute gradient over all training samples and normalized to sum to one within each model. The darker the color gradient, the higher the respective importance. The variables are ordered according to the importance of the DPPP model optimized for $\gamma = 2$.

Welch and Goyal (2008) show that net equity expansion acts as a market-timing signal that benefits investors with moderate risk aversion ($\gamma \approx 3$) but results in negative certainty equivalent returns for those with higher risk aversion. In our portfolio optimization framework, the increased importance of net equity expansion for higher risk aversion can be interpreted as the model relying more on net equity expansion to adjust portfolio weights in order to mitigate the risks and trading costs associated with excessive leverage.

A.5 Comparison of portfolio weights and variable importance

To further analyze the economic differences between CRRA, mean-variance and loss-averse portfolio policies, we examine their exposure to characteristics. For each topical cluster of characteristics k and portfolio p, we calculate the portfolio's exposure as

$$E_{p,k,t} = \sum_{i=1}^{N_t} w_{i,t}^{p} X_{i,k,t}, \tag{A.18}$$

where $w_{i,t}^{p}$ represents the portfolio weight of stock i at time t in portfolio p, and $X_{i,k,t}$ is the standardized value of characteristic k for stock i at time t. Since we allow for short-selling, $w_{i,t}^{p}$ can be negative, implying that positive characteristic exposures can arise from either long positions in stocks with positive characteristic values or short positions in stocks with negative characteristic values. Conversely, negative exposures result from long positions in stocks with negative characteristic values or short positions in stocks with positive characteristic values. We assess the economic significance of these net exposures by examining the time-series average

$$\overline{E}_{p,k} = \frac{1}{T} \sum_{t=1}^{T} E_{p,k,t}. \tag{A.19}$$

Figure A.7 shows the time-series averages of net exposures $\overline{E}_{p,k}$ for a set of eight clusters. Across the panels, distinct patterns emerge in the exposure of portfolios to firm characteristics, with some notable differences across risk (or loss) aversions, while exposures are largely similar across utility functions.

The short-term reversal and size clusters exhibit a declining trend in net exposure as aversion increases, suggesting that more risk-averse investors allocate less positive or even negative weight to stocks associated with these characteristics. Since predictors are signed in the Chen and Zimmermann (2022) database to have positive mean return in the original in-sample periods (and smaller firms had higher returns in Banz (1981)), this implies for size that more risk or loss averse investors load *more* on larger firms, presumably because such firms are less volatile. For short-term reversal, the decline in exposure is least pronounced for loss-averse portfolios, suggesting that loss-averse investors, unlike CRRA and mean-variance investors, are more willing to invest in stocks that have recently underperformed even for higher degrees of loss aversion.

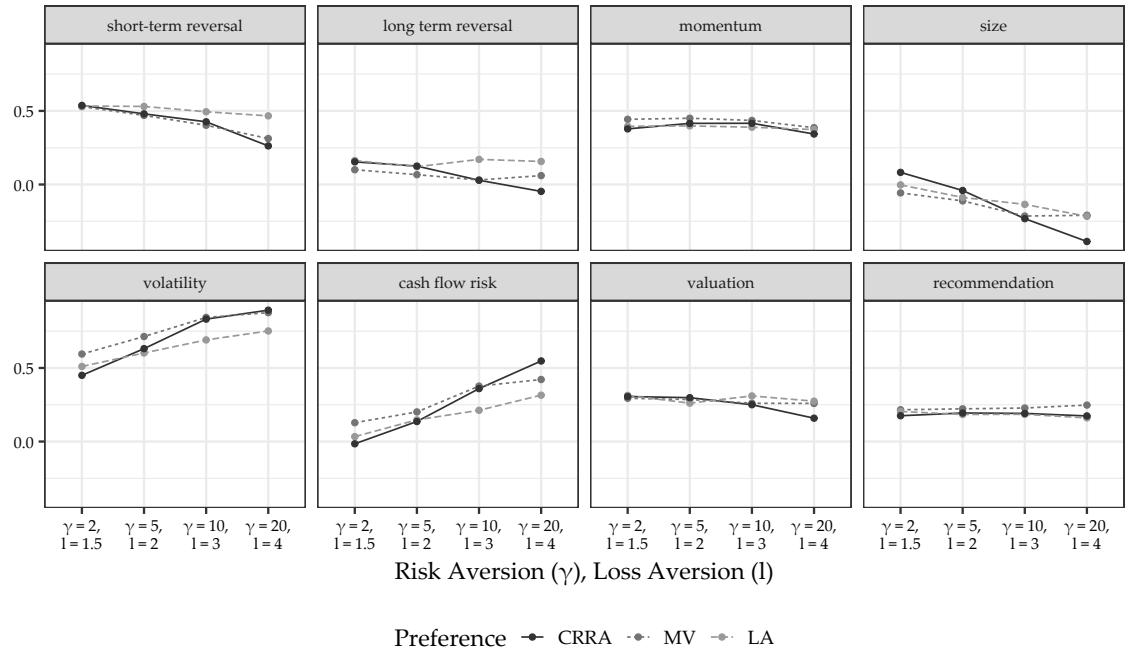


Figure A.7: Net exposure to clusters of firm characteristics for different preferences
This figure shows the net expose to clusters for CRRA, mean-variance (MV) and loss-averse (LA) portfolio policies across different levels of risk and loss aversion. We group the characteristics into clusters according to the economic category specified in the Open Source Asset Pricing data set by Chen and Zimmermann (2022). Each panel presents time-series averages of net exposures to a given cluster for a specific risk aversion level γ and corresponding loss aversion parameter l.

In contrast, the volatility and cash flow risk clusters show increasing exposure with higher aversions, particularly for CRRA and mean-variance investors, who appear willing to allocate more weight to firms with higher risk in pursuit of potential returns. However, loss-averse portfolios consistently exhibit lower exposure to these clusters, indicating that risk-related characteristics are less important, likely since volatility is not a specific part of the loss function.

For the long-term reversal, momentum, valuation and recommendations clusters, exposures remain relatively stable across aversion levels for all three preference types, implying that these characteristics play a more neutral role in portfolio construction. Nevertheless, the long-term reversal cluster displays slight variations across investor types, with loss-averse portfolios consistently maintaining higher exposures than CRRA and mean-variance portfolios, especially for higher aversion levels. This suggests that loss-averse investors are more willing to take contrarian positions on previously underperforming stocks.

A.6 Supplementary figures

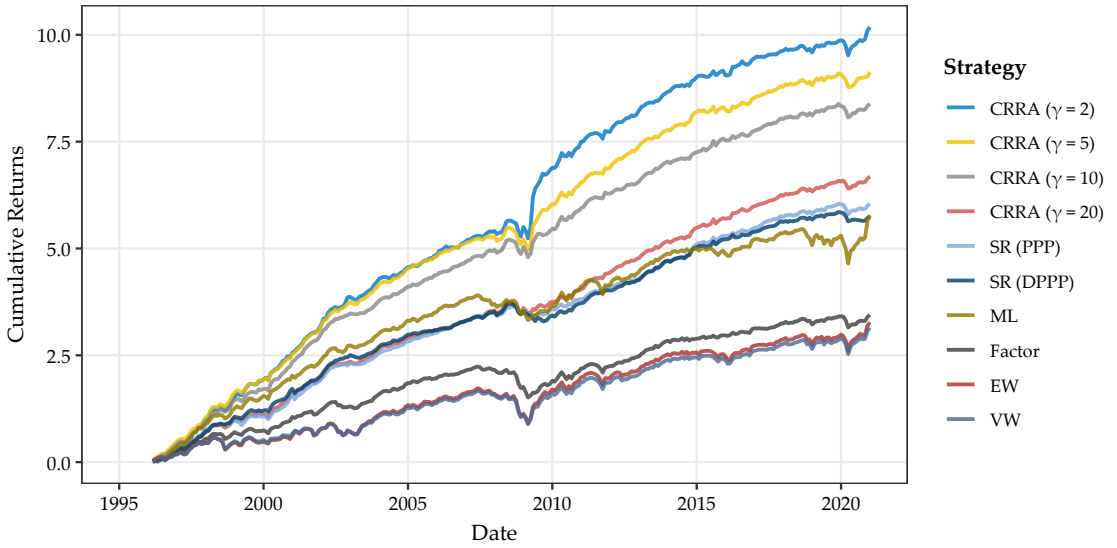


Figure A.8: Cumulative performance over time for benchmark strategies
The plot shows the cumulative sum of portfolio returns for different benchmark strategies studied in Section 2.3.4. We show the results for each of the strategies considered and across all out-of-sample periods.

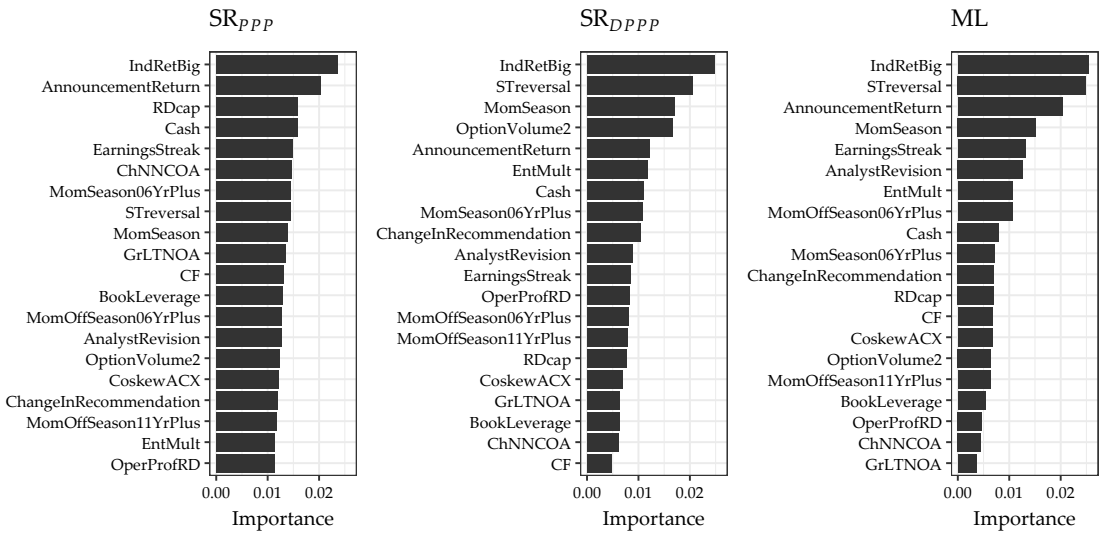


Figure A.9: Top 20 characteristics for Sharpe ratio utility and ML model
The plot shows the 20 most important characteristics for complex benchmark models discussed in section 2.3.4. SR$_{PPP}$ and SR$_{DPPP}$ refer to direct optimization of the Sharpe ratio via a linear (PPP) or network-based (DPPP) model. ML refers to network-based modeling of expected returns as in, e.g., Gu et al. (2020). Variable importance is measured by the average absolute gradient (see section 2.3.4).

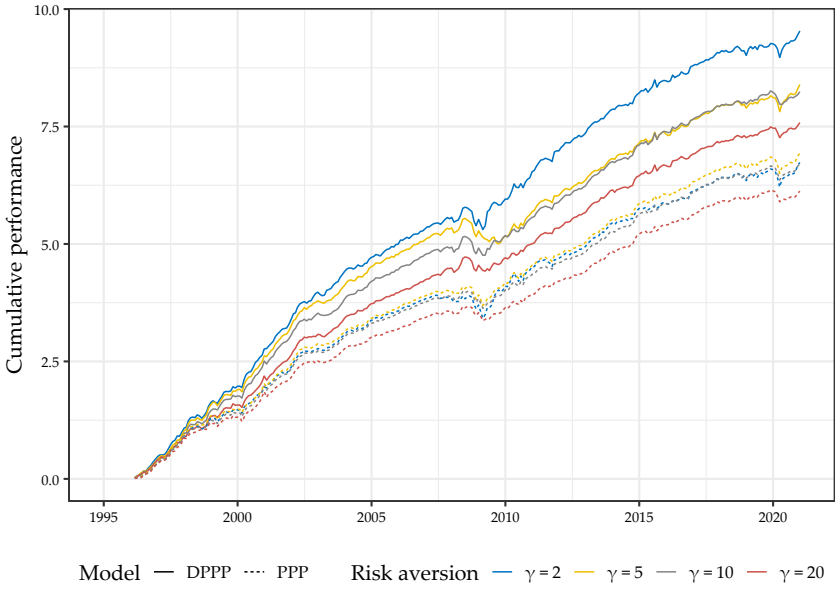


Figure A.10: Mean-variance utility

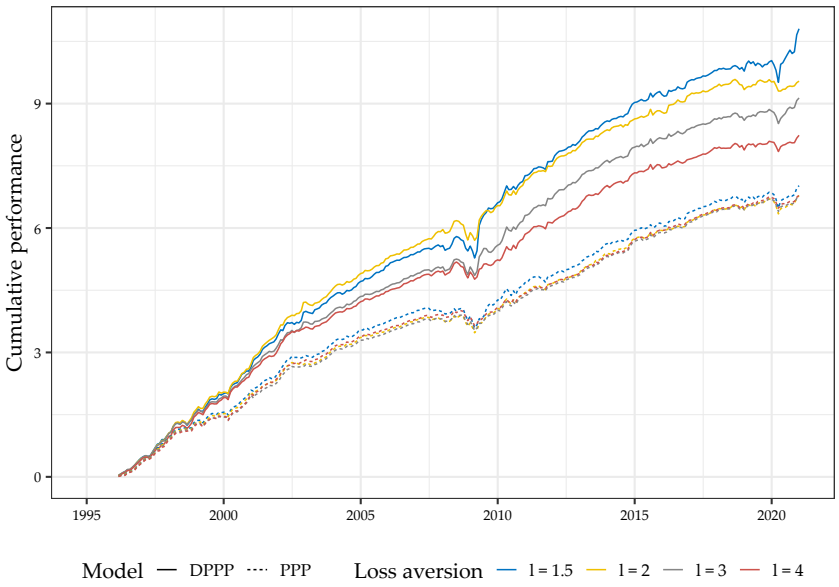


Figure A.11: Loss aversion

Figure A.12: Cumulative performance over time for MV and LA preferences
Panel (a) shows the cumulative sum of portfolio returns for the DPPP and PPP of investors with mean-variance preferences. Panel (b) shows the cumulative sum of portfolio returns net of the DPPP and PPP of investors with loss-aversion preferences. We show the results for each of the degrees of absolute risk aversion and loss aversion considered and across all out-of-sample periods.

A.7 Supplementary tables

Table A.4: Cross-evaluation of portfolio strategies against different utility preferences

	CRRA			
Strategy	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 20$
DPPP ($\gamma = 2$)	**1.0000**	0.9998	0.9663	0.5374
DPPP ($\gamma = 5$)	0.9981	**1.0000**	0.9980	0.7937
DPPP ($\gamma = 10$)	0.9963	0.9956	**1.0000**	0.8207
DPPP ($\gamma = 20$)	0.9915	0.9819	0.9944	**1.0000**
SR (PPP)	0.9896	0.9752	0.9829	0.9991
SR (DPPP)	0.9887	0.9715	0.9750	0.9831
ML	0.9899	0.9595	0.8725	0.4247
Factor	0.9806	0.9366	0.8788	0.6576
EW	0.9784	0.9173	0.7887	0.3067
VW	0.9784	0.9197	0.8057	0.3707

This table presents the out-of-sample utility for various investment strategies evaluated for a CRRA investor across different risk aversion levels. The utility of each strategy is normalized by the maximum utility within each risk aversion category, so that the best performing strategy is set to one (100%), and all other values are expressed as a fraction of that optimum. Bold values indicate the best performing strategy for each preference. The DPPP strategy represents our baseline strategy for different risk aversions. The SR strategy is a PPP and a DPPP optimized for Sharpe ratio preference. ML is the portfolio of a machine learning model trained to predict expected returns. The factor strategy is the linear PPP of a simple Fama-French five-factor model plus momentum. EW and VW are passive equal-weighted and value-weighted strategies.

Table A.5: Predictor variables for the (D)PPP

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
ChInvIA	Change in capital inv (ind adj)	Abarbanell and Bushee	1998, AR	yearly	Accounting	investment growth
GrSaleToGrInv	Sales growth over inventory growth	Abarbanell and Bushee	1998, AR	yearly	Accounting	sales growth
GrSaleToGrOverhead	Sales growth over overhead growth	Abarbanell and Bushee	1998, AR	yearly	Accounting	sales growth
IdioVolAHT	Idiosyncratic risk (AHT)	Ali, Hwang, and Trombley	2003, JFE	monthly	Price	volatility
EarningsConsistency	Earnings consistency	Alwathainani	2009, BAR	yearly	Accounting	earnings
Illiquidity	Amihud's illiquidity	Amihud	2002, JFM	monthly	Trading	liquidity
BidAskSpread	Bid-ask spread	Amihud and Mendelsohn	1986, JFE	monthly	Trading	liquidity
grcapx	Change in capex (two years)	Anderson and Garcia-Feijoo	2006, JF	yearly	Accounting	investment growth
grcapx3y	Change in capex (three years)	Anderson and Garcia-Feijoo	2006, JF	yearly	Accounting	investment growth
betaVIX	Systematic volatility	Ang et al.	2006, JF	monthly	Price	volatility
IdioRisk	Idiosyncratic risk	Ang et al.	2006, JF	monthly	Price	volatility
IdioVol3F	Idiosyncratic risk (3 factor)	Ang et al.	2006, JF	monthly	Price	volatility
CoskewACX	Coskewness using daily returns	Ang, Chen and Xing	2006, RFS	monthly	Price	risk
Mom6mJunk	Junk Stock Momentum	Avramov et al	2007, JF	monthly	Price	momentum
OrderBacklogChg	Change in order backlog	Baik and Ahn	2007, Other	yearly	Accounting	accruals
roaq	Return on assets (qtrly)	Balakrishnan, Bartov and Faurel	2010, JAE	quarterly	Accounting	profitability
MaxRet	Maximum return over month	Bali, Cakici, and Whitelaw	2010, JF	monthly	Price	volatility
ReturnSkew	Return skewness	Bali, Engle and Murray	2015, Book	monthly	Price	risk
ReturnSkew3F	Idiosyncratic skewness (3F model)	Bali, Engle and Murray	2015, Book	monthly	Price	risk
CBOperProf	Cash-based operating profitability	Ball et al.	2016, JFE	yearly	Accounting	profitability
OperProfRD	Operating profitability R&D adjusted	Ball et al.	2016, JFE	yearly	Accounting	profitability
Size	Size	Banz	1981, JFE	monthly	Price	size
SP	Sales-to-price	Barbee, Mukherji and Raines	1996, FAJ	yearly	Accounting	valuation

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
EP	Earnings-to-Price Ratio	Basu	1977, JF	monthly	Price	valuation
InvGrowth	Inventory Growth	Belo and Lin	2012, RFS	yearly	Accounting	profitability
BrandInvest	Brand capital investment	Belo, Lin and Vitorino	2014, RED	yearly	Accounting	investment
Leverage	Market leverage	Bhandari	1988, JFE	monthly	Price	leverage
ResidualMomentum	Momentum based on FF3 residuals	Blitz, Huij and Martens	2011, JEmpFin	monthly	Price	momentum
Price	Price	Blume and Husic	1972, JF	monthly	Price	other
NetPayoutYield	Net Payout Yield	Boudoukh et al.	2007, JF	monthly	Price	valuation
PayoutYield	Payout Yield	Boudoukh et al.	2007, JF	monthly	Price	valuation
NetDebtFinance	Net debt financing	Bradshaw, Richardson, Sloan	2006, JAE	yearly	Accounting	external financing
NetEquityFinance	Net equity financing	Bradshaw, Richardson, Sloan	2006, JAE	yearly	Accounting	external financing
XFIN	Net external financing	Bradshaw, Richardson, Sloan	2006, JAE	yearly	Accounting	external financing
DolVol	Past trading volume	Brennan, Chordia, Subra	1998, JFE	monthly	Trading	volume
FEPS	Analyst earnings per share	Cen, Wei, and Zhang	2006, WP	monthly	Analyst	profitability
AnnouncementReturn	Earnings announcement return	Chan, Jegadeesh and Lakonishok	1996, JF	monthly	Price	earnings
REV6	Earnings forecast revisions	Chan, Jegadeesh and Lakonishok	1996, JF	monthly	Analyst	earnings
AdExp	Advertising Expense	Chan, Lakonishok and Sougiannis	2001, JF	monthly	Accounting	R&D
RD	R&D over market cap	Chan, Lakonishok and Sougiannis	2001, JF	monthly	Accounting	R&D
CashProd	Cash Productivity	Chandrashekar and Rao	2009, WP	yearly	Accounting	profitability
std_turn	Share turnover volatility	Chordia, Subra, Anshuman	2001, JFE	monthly	Trading	liquidity
VolSD	Volume Variance	Chordia, Subra, Anshuman	2001, JFE	monthly	Trading	liquidity
retConglomerate	Conglomerate return	Cohen and Lou	2012, JFE	monthly	Price	delayed processing
RDAbility	R&D ability	Cohen, Diether and Malloy	2013, RFS	yearly	Accounting	other
AssetGrowth	Asset growth	Cooper, Gulen and Schill	2008, JF	yearly	Accounting	investment

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
EarningsForecastDisparity	Long-vs-short EPS forecasts	Da and Warachka	2011, JFE	monthly	Analyst	earnings
CompEquIss	Composite equity issuance	Daniel and Titman	2006, JF	monthly	Accounting	external financing
IntanBM	Intangible return using BM	Daniel and Titman	2006, JF	yearly	Accounting	long term reversal
IntanCFP	Intangible return using CFtoP	Daniel and Titman	2006, JF	yearly	Accounting	long term reversal
IntanEP	Intangible return using EP	Daniel and Titman	2006, JF	yearly	Accounting	long term reversal
IntanSP	Intangible return using Sale2P	Daniel and Titman	2006, JF	yearly	Accounting	long term reversal
ShareIss5Y	Share issuance (5 year)	Daniel and Titman	2006, JF	monthly	Accounting	external financing
LRreversal	Long-run reversal	De Bondt and Thaler	1985, JF	monthly	Price	long term reversal
MRreversal	Medium-run reversal	De Bondt and Thaler	1985, JF	monthly	Price	long term reversal
EquityDuration	Equity Duration	Dechow, Sloan and Soliman	2004, RAS	yearly	Price	valuation
cfp	Operating Cash flows to price	Desai, Rajgopal, Venkatachalam	2004, AR	yearly	Accounting	valuation
ForecastDispersion	EPS Forecast Dispersion	Diether, Malloy and Scherbina	2002, JF	monthly	Analyst	volatility
ExclExp	Excluded Expenses	Doyle, Lundholm and Soliman	2003, RAS	quarterly	Analyst	composite accounting
ProbInformedTrading	Probability of Informed Trading	Easley, Hvidkjaer and O'Hara	2002, JF	yearly	Trading	liquidity
OrgCap	Organizational capital	Eisfeldt and Papanikolaou	2013, JF	yearly	Accounting	R&D
sfe	Earnings Forecast to price	Elgers, Lo and Pfeiffer	2001, AR	monthly	Analyst	valuation
GrLTNOA	Growth in long term operating assets	Fairfield, Whisenant and Yohn	2003, AR	yearly	Accounting	investment
AM	Total assets to market	Fama and French	1992, JF	yearly	Accounting	valuation
BMdec	Book to market using December ME	Fama and French	1992, JPM	yearly	Accounting	valuation
BookLeverage	Book leverage (annual)	Fama and French	1992, JF	yearly	Accounting	leverage
OperProf	operating profits / book equity	Fama and French	2006, JFE	yearly	Accounting	profitability
Beta	CAPM beta	Fama and MacBeth	1973, JPE	monthly	Price	risk
EarningsSurprise	Earnings Surprise	Foster, Olsen and Shevlin	1984, AR	quarterly	Analyst	earnings

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
AnalystValue	Analyst Value	Frankel and Lee	1998, JAE	monthly	Analyst	valuation
AOP	Analyst Optimism	Frankel and Lee	1998, JAE	monthly	Analyst	other
PredictedFE	Predicted Analyst forecast error	Frankel and Lee	1998, JAE	monthly	Accounting	earnings
FR	Pension Funding Status	Franzoni and Marin	2006, JF	monthly	Accounting	composite accounting
BetaFP	Frazzini-Pedersen Beta	Frazzini and Pedersen	2014, JFE	monthly	Price	other
High52	52 week high	George and Hwang	2004, JF	monthly	Price	momentum
IndMom	Industry Momentum	Grinblatt and Moskowitz	1999, JFE	monthly	Price	momentum
PctAcc	Percent Operating Accruals	Hafzalla, Lundholm, Van Winkle	2011, AR	yearly	Accounting	accruals
PctTotAcc	Percent Total Accruals	Hafzalla, Lundholm, Van Winkle	2011, AR	yearly	Accounting	accruals
tang	Tangibility	Hahn and Lee	2009, JF	yearly	Accounting	asset composition
Coskewness	Coskewness	Harvey and Siddique	2000, JF	monthly	Price	risk
RoE	net income / book equity	Haugen and Baker	1996, JFE	yearly	Accounting	profitability
VarCF	Cash-flow to price variance	Haugen and Baker	1996, JFE	monthly	Accounting	cash flow risk
VolMkt	Volume to market equity	Haugen and Baker	1996, JFE	monthly	Trading	volume
VolumeTrend	Volume Trend	Haugen and Baker	1996, JFE	monthly	Trading	volume
AnalystRevision	EPS forecast revision	Hawkins, Chamberlin, Daniel	1984, FAJ	monthly	Analyst	earnings
Mom12mOffSeason	Momentum without the seasonal part	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomOffSeason	Off season long-term reversal	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomOffSeason06YrPlus	Off season reversal years 6 to 10	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomOffSeason11YrPlus	Off season reversal years 11 to 15	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomOffSeason16YrPlus	Off season reversal years 16 to 20	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomSeason	Return seasonality years 2 to 5	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomSeason06YrPlus	Return seasonality years 6 to 10	Heston and Sadka	2008, JFE	monthly	Price	momentum

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
MomSeason11YrPlus	Return seasonality years 11 to 15	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomSeason16YrPlus	Return seasonality years 16 to 20	Heston and Sadka	2008, JFE	monthly	Price	momentum
MomSeasonShort	Return seasonality last year	Heston and Sadka	2008, JFE	monthly	Price	momentum
NOA	Net Operating Assets	Hirshleifer et al.	2004, JAE	yearly	Accounting	asset composition
dNoa	change in net operating assets	Hirshleifer, Hou, Teoh, Zhang	2004, JAE	yearly	Accounting	investment
EarnSupBig	Earnings surprise of big firms	Hou	2007, RFS	quarterly	Accounting	delayed processing
IndRetBig	Industry return of big firms	Hou	2007, RFS	monthly	Price	delayed processing
PriceDelayRsq	Price delay r square	Hou and Moskowitz	2005, RFS	monthly	Price	delayed processing
PriceDelaySlope	Price delay coeff	Hou and Moskowitz	2005, RFS	monthly	Price	delayed processing
PriceDelayTstat	Price delay SE adjusted	Hou and Moskowitz	2005, RFS	monthly	Price	delayed processing
STreversal	Short term reversal	Jegadeesh	1989, JF	monthly	Price	short-term reversal
RevenueSurprise	Revenue Surprise	Jegadeesh and Livnat	2006, JFE	quarterly	Accounting	sales growth
Mom12m	Momentum (12 month)	Jegadeesh and Titman	1993, JF	monthly	Price	momentum
Mom6m	Momentum (6 month)	Jegadeesh and Titman	1993, JF	monthly	Price	momentum
ChangeInRecommendation	Change in recommendation	Jegadeesh et al.	2004, JF	monthly	Analyst	recommendation
OptionVolume1	Option to stock volume	Johnson and So	2012, JFE	monthly	Trading	volume
OptionVolume2	Option volume to average	Johnson and So	2012, JFE	monthly	Trading	volume
BetaTailRisk	Tail risk beta	Kelly and Jiang	2014, RFS	monthly	Price	risk
fgr5yrLag	Long-term EPS forecast	La Porta	1996, JF	monthly	Analyst	earnings
CF	Cash flow to market	Lakonishok, Shleifer, Vishny	1994, JF	monthly	Accounting	valuation
MeanRankRevGrowth	Revenue Growth Rank	Lakonishok, Shleifer, Vishny	1994, JF	yearly	Accounting	sales growth
RDS	Real dirty surplus	Landsman et al.	2011, AR	yearly	Accounting	composite accounting
Tax	Taxable income to income	Lev and Nissim	2004, AR	yearly	Accounting	tax

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
RDcap	R&D capital-to-assets	Li	2011, RFS	yearly	Accounting	asset composition
zerotrade	Days with zero trades	Liu	2006, JFE	monthly	Trading	liquidity
zerotradeAlt1	Days with zero trades	Liu	2006, JFE	monthly	Trading	liquidity
zerotradeAlt12	Days with zero trades	Liu	2006, JFE	monthly	Trading	liquidity
ChEQ	Growth in book equity	Lockwood and Prombutr	2010, JFR	yearly	Accounting	investment
EarningsStreak	Earnings surprise streak	Loh and Warachka	2012, MS	monthly	Accounting	earnings
NumEarnIncrease	Earnings streak length	Loh and Warachka	2012, MS	quarterly	Accounting	earnings
GrAdExp	Growth in advertising expenses	Lou	2014, RFS	yearly	Accounting	investment
EntMult	Enterprise Multiple	Loughran and Wellman	2011, JFQA	monthly	Accounting	valuation
CompositeDebtIssuance	Composite debt issuance	Lyandres, Sun and Zhang	2008, RFS	yearly	Accounting	external financing
InvestPPEInv	change in ppe and inv/assets	Lyandres, Sun and Zhang	2008, RFS	yearly	Accounting	investment
Frontier	Efficient frontier index	Nguyen and Swanson	2009, JFQA	yearly	Accounting	valuation
GP	gross profits / total assets	Novy-Marx	2013, JFE	yearly	Accounting	profitability
IntMom	Intermediate Momentum	Novy-Marx	2012, JFE	monthly	Price	momentum
OPLeverage	Operating leverage	Novy-Marx	2010, ROF	yearly	Accounting	other
Cash	Cash to assets	Palazzo	2012, JFE	quarterly	Accounting	asset composition
BetaLiquidityPS	Pastor-Stambaugh liquidity beta	Pastor and Stambaugh	2003, JPE	monthly	Price	liquidity
BPEBM	Leverage component of BM	Penman, Richardson and Tuna	2007, JAR	monthly	Accounting	leverage
EBM	Enterprise component of BM	Penman, Richardson and Tuna	2007, JAR	monthly	Accounting	valuation
NetDebtPrice	Net debt to price	Penman, Richardson and Tuna	2007, JAR	monthly	Accounting	leverage
PS	Piotroski F-score	Piotroski	2000, AR	yearly	Accounting	composite accounting
ShareIss1Y	Share issuance (1 year)	Pontiff and Woodgate	2008, JF	monthly	Accounting	external financing
DelDRC	Deferred Revenue	Prakash and Sinha	2012, CAR	yearly	Accounting	investment

Continued on next page

Table A.5: Predictor variables for the machine learning model (continued)

Acronym	Description	Author(s)	Year, Journal	Frequency	Cat. data	Cat. economic
OrderBacklog	Order backlog	Rajgopal, Shevlin, Venkatachalam	2003, RAS	yearly	Accounting	sales growth
DelCOA	Change in current operating assets	Richardson et al.	2005, JAE	yearly	Accounting	investment
DelCOL	Change in current operating liabilities	Richardson et al.	2005, JAE	yearly	Accounting	external financing
DelEqu	Change in equity to assets	Richardson et al.	2005, JAE	yearly	Accounting	investment
DelFINL	Change in financial liabilities	Richardson et al.	2005, JAE	yearly	Accounting	external financing
DelLTI	Change in long-term investment	Richardson et al.	2005, JAE	yearly	Accounting	investment
DelNetFin	Change in net financial assets	Richardson et al.	2005, JAE	yearly	Accounting	investment
TotalAccruals	Total accruals	Richardson et al.	2005, JAE	yearly	Accounting	investment
BM	Book to market using most recent ME	Rosenberg, Reid, and Lanstein	1985, JF	monthly	Accounting	valuation
Accruals	Accruals	Sloan	1996, AR	yearly	Accounting	accruals
ChAssetTurnover	Change in Asset Turnover	Soliman	2008, AR	yearly	Accounting	sales growth
ChNNCOA	Change in Net Noncurrent Op Assets	Soliman	2008, AR	yearly	Accounting	investment
ChNWC	Change in Net Working Capital	Soliman	2008, AR	yearly	Accounting	investment
ChInv	Inventory Growth	Thomas and Zhang	2002, RAS	yearly	Accounting	investment
ChTax	Change in Taxes	Thomas and Zhang	2011, JAR	quarterly	Accounting	tax
Investment	Investment to revenue	Titman, Wei and Xie	2004, JFQA	yearly	Accounting	investment
realestate	Real estate holdings	Tuzel	2010, RFS	yearly	Accounting	asset composition
AbnormalAccruals	Abnormal Accruals	Xie	2001, AR	yearly	Accounting	accruals
FirmAgeMom	Firm Age - Momentum	Zhang	2004, JF	monthly	Price	momentum

The table shows all available characteristics used, the author(s), the year and the journal of publication. In addition, this table shows the update frequency, the data category as well as the economic category.

Table A.6: Transaction cost constrained deep portfolio policy for CRRA investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0155	0.0194	0.0129	0.0157	0.0077	0.0087	-0.0029	-0.0006
p-value$(CE_{DPPP} - CE_{PPP})$		0.0118		0.0620		0.3195		0.1555
$\sum \lvert w_i \rvert / N_t * 100$	0.1684	0.1925	0.1750	0.1938	0.1753	0.1914	0.1607	0.1631
$max\ w_i * 100$	0.6553	0.7983	0.6721	0.6514	0.6595	0.4899	0.5866	0.3813
$min\ w_i * 100$	-0.6305	-1.0150	-0.6774	-0.9952	-0.6812	-1.0398	-0.5735	-0.6183
$\sum w_i I(w_i < 0)$	-0.7139	-0.8877	-0.7612	-0.8973	-0.7638	-0.8798	-0.6588	-0.6756
$\sum I(w_i < 0)/N_t$	0.3267	0.3296	0.3417	0.3267	0.3440	0.3185	0.3319	0.3367
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.8441	2.0257	0.8794	1.9002	0.8593	1.5947	0.7754	1.1407
Mean	0.0179	0.0225	0.0178	0.0221	0.0170	0.0182	0.0144	0.0157
StdDev	0.0482	0.0551	0.0427	0.0498	0.0397	0.0412	0.0360	0.0349
Skew	-0.6222	-0.1700	-0.8459	0.0763	-0.9009	-0.5885	-0.8134	-0.7482
Kurt	3.1259	4.9340	2.4793	7.0330	2.4114	1.9342	1.8024	2.0122
Max DD	0.6062	0.7288	0.4937	0.5344	0.4224	0.5714	0.4020	0.3975
Max 1M loss	0.2228	0.2280	0.1812	0.2015	0.1559	0.1546	0.1303	0.1513
CVaR (95%)	0.1037	0.1164	0.0937	0.0981	0.0891	0.0873	0.0794	0.0727
SR	1.2851	1.4123	1.4453	1.5370	1.4823	1.5296	1.3805	1.5552
p-value$(SR_{DPPP} - SR_{PPP})$		0.2090		0.2962		0.3852		0.0447
$FF5 + Mom\ \alpha$	0.0065	0.0112	0.0071	0.0100	0.0069	0.0082	0.0056	0.0070
$StdErr(\alpha)$	0.0013	0.0021	0.0013	0.0019	0.0013	0.0017	0.0014	0.0014

This table presents out-of-sample performance estimates for deep portfolio policies with the transaction costs penalty from Equation (2.14) using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. All results are reported net of transaction costs. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.7: Long-only deep portfolio policy for CRRA investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0118	0.0164	0.0076	0.0107	0.0011	0.0020	-0.0157	-0.0104
p-value$(CE_{DPPP} - CE_{PPP})$		0.0001		0.0143		0.3308		0.0114
$\sum \lvert w_i \rvert / N_t * 100$	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694
$max\ w_i * 100$	0.3543	1.9711	0.3761	1.9329	0.3588	1.2718	0.3608	0.7288
$min\ w_i * 100$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum w_i I(w_i < 0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum I(w_i < 0)/N_t$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.5883	1.3508	0.6426	1.3417	0.5000	1.0914	0.3274	0.7656
Mean	0.0150	0.0215	0.0147	0.0216	0.0137	0.0174	0.0114	0.0135
StdDev	0.0566	0.0713	0.0510	0.0647	0.0459	0.0490	0.0406	0.0390
Skew	-0.4486	0.1148	-0.6717	0.3396	-0.5582	-0.6934	-0.9488	-0.9849
Kurt	3.3286	3.9959	3.0863	7.3953	3.6061	3.9808	3.0547	2.8460
Max DD	0.7942	0.7985	0.7257	0.8282	0.6778	0.6520	0.6011	0.5064
Max 1M loss	0.2483	0.2603	0.2171	0.2667	0.1968	0.2260	0.1832	0.1780
CVaR (95%)	0.1266	0.1472	0.1168	0.1295	0.1037	0.1093	0.0964	0.0898
SR	0.9213	1.0418	0.9996	1.1580	1.0342	1.2262	0.9717	1.1974
p-value$(SR_{DPPP} - SR_{PPP})$		0.0119		0.0077		0.0006		0.0001
$FF5 + Mom\ \alpha$	0.0040	0.0105	0.0045	0.0114	0.0042	0.0076	0.0023	0.0047
$StdErr(\alpha)$	0.0007	0.0017	0.0008	0.0016	0.0008	0.0010	0.0008	0.0009

This table presents out-of-sample performance estimates for deep portfolio policies including a long-only constraint using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.8: Long-only & transaction cost constrained deep portfolio policy for CRRA investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0101	0.0128	0.0063	0.0067	0.0003	0.0015	-0.0157	-0.0087
p-value$(CE_{DPPP} - CE_{PPP})$		0.0079		0.4253		0.0245		0.0001
$\sum \lvert w_i \rvert / N_t * 100$	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694	0.0694
$max\ w_i * 100$	0.2776	1.7957	0.3111	1.6695	0.3280	1.1258	0.3575	0.5347
$min\ w_i * 100$	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum w_i I(w_i < 0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum I(w_i < 0)/N_t$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.3104	1.1833	0.3388	1.1537	0.2910	0.8621	0.2509	0.4959
Mean	0.0130	0.0172	0.0127	0.0153	0.0120	0.0137	0.0106	0.0113
StdDev	0.0536	0.0658	0.0481	0.0559	0.0439	0.0457	0.0401	0.0363
Skew	-0.5117	0.1447	-0.6868	-0.5080	-0.6450	-0.3657	-0.9713	-1.0731
Kurt	3.7273	5.0898	3.6745	4.2885	3.7214	4.1127	3.0701	2.8108
Max DD	0.7754	0.8689	0.7044	0.8739	0.6672	0.6905	0.6061	0.5223
Max 1M loss	0.2419	0.2587	0.2192	0.2513	0.1899	0.1843	0.1808	0.1617
CVaR (95%)	0.1229	0.1383	0.1119	0.1255	0.1019	0.1020	0.0960	0.0858
SR	0.8413	0.9042	0.9130	0.9496	0.9432	1.0342	0.9152	1.0760
p-value$(SR_{DPPP} - SR_{PPP})$		0.1453		0.2827		0.0238		0.0001
$FF5 + Mom\ \alpha$	0.0019	0.0059	0.0023	0.0045	0.0022	0.0041	0.0014	0.0024
$StdErr(\alpha)$	0.0007	0.0014	0.0007	0.0012	0.0007	0.0010	0.0008	0.0009

This table presents out-of-sample performance estimates for deep portfolio policies with the transaction costs penalty from Equation (2.14) and including a long-only constraint using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for Constant Relative Risk Aversion (CRRA) investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. All results are reported net of transaction costs. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.9: Deep portfolio policy for mean-variance investors with different degrees of risk aversion

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0201	0.0287	0.0184	0.0217	0.0143	0.0170	0.0065	0.0088
p-value$(CE_{DPPP} - CE_{PPP})$		0.0001		0.0292		0.0291		0.0849
$\sum \lvert w_i \rvert / N_t * 100$	0.1748	0.1926	0.1811	0.1952	0.1816	0.1928	0.1786	0.1932
$max\ w_i * 100$	0.7115	1.0449	0.7675	0.9073	0.7761	0.7741	0.7586	0.6787
$min\ w_i * 100$	-0.6847	-1.3109	-0.7234	-1.2246	-0.7227	-1.1856	-0.6987	-1.0437
$\sum w_i I(w_i < 0)$	-0.7602	-0.8882	-0.8059	-0.9070	-0.8093	-0.8899	-0.7879	-0.8925
$\sum I(w_i < 0)/N_t$	0.3488	0.3250	0.3564	0.3193	0.3560	0.3227	0.3517	0.3381
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	1.5185	2.6428	1.7406	2.5648	1.6789	2.4174	1.4693	2.2676
Mean	0.0225	0.0319	0.0232	0.0281	0.0224	0.0276	0.0205	0.0254
StdDev	0.0492	0.0566	0.0435	0.0505	0.0402	0.0459	0.0373	0.0407
Skew	-0.6239	-0.1348	-0.8530	-0.6631	-0.8516	-0.4331	-0.7727	-0.5940
Kurt	2.8505	3.3104	2.5837	2.3437	2.1502	2.1743	1.9302	2.2886
Max DD	0.5478	0.5284	0.4404	0.5957	0.3857	0.4392	0.3219	0.3290
Max 1M loss	0.2151	0.2134	0.1867	0.1980	0.1551	0.1837	0.1235	0.1620
CVaR (95%)	0.1039	0.1035	0.0921	0.0979	0.0871	0.0833	0.0801	0.0761
SR	1.5843	1.9506	1.8438	1.9259	1.9317	2.0786	1.9007	2.1596
p-value$(SR_{DPPP} - SR_{PPP})$		0.0019		0.2768		0.1185		0.0171
$FF5 + Mom\ \alpha$	0.0108	0.0198	0.0122	0.0158	0.0121	0.0171	0.0108	0.0151
$StdErr(\alpha)$	0.0013	0.0022	0.0014	0.0020	0.0014	0.0020	0.0014	0.0017

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for mean-variance investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.10: Deep portfolio policy for loss-averse investors with different degrees of loss aversion

	$l = 1.5$		$l = 2$		$l = 3$		$l = 4$	
	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0188	0.0311	0.0147	0.0235	0.0082	0.0137	0.0025	0.0036
p-value$(CE_{DPPP} - CE_{PPP})$		0.0002		0.0015		0.0247		0.3014
$\sum \lvert w_i \rvert / N_t * 100$	0.1793	0.1931	0.1801	0.1919	0.1816	0.1917	0.1815	0.1901
$max\ w_i * 100$	0.7510	1.0462	0.7490	1.1024	0.7652	1.0074	0.7625	0.9322
$min\ w_i * 100$	-0.7062	-1.3205	-0.7093	-1.2292	-0.7215	-1.1298	-0.7172	-1.1071
$\sum w_i I(w_i < 0)$	-0.7929	-0.8918	-0.7980	-0.8833	-0.8090	-0.8823	-0.8083	-0.8702
$\sum I(w_i < 0)/N_t$	0.3537	0.3219	0.3522	0.3471	0.3559	0.3353	0.3566	0.3326
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	1.6336	2.6846	1.5951	2.6742	1.6887	2.5599	1.7273	2.4745
Mean	0.0235	0.0361	0.0227	0.0319	0.0226	0.0306	0.0227	0.0275
StdDev	0.0494	0.0751	0.0442	0.0580	0.0412	0.0548	0.0395	0.0485
Skew	-0.6194	1.9765	-0.7339	0.5481	-0.7651	0.3536	-0.7475	-0.2204
Kurt	2.7196	15.8393	2.6054	6.7495	2.1086	4.8422	1.9761	1.6728
Max DD	0.5395	0.5974	0.4395	0.5395	0.4019	0.4461	0.3955	0.4560
Max 1M loss	0.2115	0.2903	0.1901	0.2144	0.1588	0.1786	0.1358	0.1604
CVaR (95%)	0.1032	0.1183	0.0920	0.1039	0.0874	0.0971	0.0828	0.0926
SR	1.6475	1.6666	1.7793	1.9049	1.9052	1.9331	1.9905	1.9677
p-value$(SR_{DPPP} - SR_{PPP})$		0.4931		0.2424		0.4498		0.4400
$FF5 + Mom\ \alpha$	0.0116	0.0235	0.0117	0.0223	0.0122	0.0198	0.0131	0.0167
$StdErr(\alpha)$	0.0014	0.0029	0.0014	0.0026	0.0014	0.0023	0.0014	0.0019

This table presents out-of-sample performance estimates for deep portfolio policies using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for loss-averse investors with loss aversions coefficients (l) of 1.5, 2, 3, and 4. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including absolute weights, maximum and minimum weights, negative weight metrics (sum and proportion), and portfolio turnover. The third set of rows displays the return distribution characteristics: the first four moments, risk metrics (maximum drawdown, maximum monthly loss, and conditional value at risk), annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP. The bottom set of rows reports the alphas and their standard errors relative to the Fama-French five-factor model augmented with the momentum factor.

Table A.11: Long-only deep portfolio policy for mean-variance and loss-averse investors with different degrees of risk aversion (γ) and loss aversion (l)

	$\gamma = 2$		$\gamma = 5$		$\gamma = 10$		$\gamma = 20$	
Mean-variance preference	PPP	DPPP	PPP	DPPP	PPP	DPPP	PPP	DPPP
CE	0.0127	0.0166	0.0083	0.0114	0.0036	0.0058	-0.0038	-0.0008
p-value($CE_{DPPP} - CE_{PPP}$)		0.0003		0.0010		0.0036		0.0001
$\sum w_i I(w_i < 0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.8041	1.3588	0.7186	1.3401	0.5924	1.2230	0.4725	1.0690
Mean	0.0163	0.0217	0.0153	0.0203	0.0138	0.0177	0.0124	0.0155
StdDev	0.0599	0.0716	0.0527	0.0597	0.0452	0.0489	0.0403	0.0404
Skew	-0.2458	0.2325	-0.5465	-0.1201	-0.8497	-0.9253	-0.9430	-0.9499
SR	0.9414	1.0497	1.0046	1.1785	1.0584	1.2569	1.0701	1.3272
p-value($SR_{DPPP} - SR_{PPP}$)		0.0185		0.0027		0.0011		0.0001
Loss-aversion preference	$l = 1.5$		$l = 2$		$l = 3$		$l = 4$	
CE	0.0092	0.0161	0.0024	0.0090	-0.0029	-0.0008	-0.0046	-0.0032
p-value($CE_{DPPP} - CE_{PPP}$)		0.0020		0.0024		0.2229		0.0330
$\sum w_i I(w_i < 0)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum \lvert w_{i,t} - w^+_{i,t-1} \rvert$	0.7649	1.3785	0.7080	1.3539	0.6649	1.2797	0.6124	1.2232
Mean	0.0162	0.0243	0.0152	0.0226	0.0145	0.0208	0.0141	0.0194
StdDev	0.0589	0.0794	0.0528	0.0692	0.0480	0.0585	0.0455	0.0535
Skew	-0.2387	0.7594	-0.5982	0.9525	-0.7169	0.4273	-0.7146	0.2861
SR	0.9522	1.0618	0.9971	1.1306	1.0456	1.2322	1.0755	1.2583
p-value($SR_{DPPP} - SR_{PPP}$)		0.0241		0.0231		0.0004		0.0028

This table presents out-of-sample performance estimates for deep portfolio policies including a long-only constraint using 157 firm characteristics, as specified in Equation (2.1). The analysis employs a feed-forward neural network model and data from the Open Source Asset Pricing Dataset spanning January 1971 to December 2020. Results are shown for mean-variance investors with relative risk aversion coefficients (γ) of 2, 5, 10, and 20 in the first panel and loss-averse investors with loss aversion (l) of 1.5, 2, 3, and 4 in the second panel. The first set of rows reports the certainty equivalent for each investor type, along with bootstrapped one-sided p-values comparing the certainty equivalents between the Deep Parametric Portfolio Policy (DPPP) and the Parametric Portfolio Policy (PPP). The second set of rows presents time-averaged portfolio weight statistics, including leverage and portfolio turnover. The third set of rows displays the return distribution characteristics: the first three moments, annualized Sharpe ratios, and bootstrapped one-sided p-values comparing Sharpe ratios between the DPPP and the PPP.

Appendix B

Appendix to Chapter 3

B.1 Traditional earnings prediction models

Table B.1: Traditional earnings prediction models

Panel A: Traditional model specifications

Name	Model	Source
L	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t}$	Gerakos and Gramacy (2012)
HVZ	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 A_{i,t} + \beta_3 D_{i,t} + \beta_4 DD_{i,t} + \beta_5 NegE_{i,t} + \beta_6 ACC_{i,t}$	Hou et al. (2012)
EP	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 NegE_{i,t} + \beta_3 NegE_{i,t} E_{i,t}$	Li and Mohanram (2014)
RI	$\mathbb{E}_t[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 NegE_{i,t} + \beta_3 NegE_{i,t} E_{i,t} + \beta_4 B_{i,t} + \beta_5 ACC_{i,t}$	Li and Mohanram (2014)

Panel B: Variable definitions

Variable	Definition
E	Income before extraordinary items (ib) / Common shares outstanding (csho)
A	Total assets (at) / csho
D	Dividends total (dvt) / csho
DD	1 if dvt > 0; 0 else
$NegE$	1 if ib < 0; 0 else
ACC	(Income before extraordinary items (ib) - Operating activities - net cash flow (oancf)) / csho
B	Common/Ordinary equity- total (ceq) / csho

Panel A reports the traditional earnings models estimated. $\mathbb{E}_t[E_{i,t+\tau}]$ denotes the expectation for earnings E of firm i in period $t+\tau$ as of t. β_0-β_5 are the model coefficients. Panel B reports the variable definitions for the traditional models. Compustat variable names are provided in parentheses. Note the slight changes as opposed to the original papers. More precisely, we scale all variables by common shares outstanding and use a consistent earnings as well as accruals definition.

B.2 Machine learning earnings prediction models

Table B.2: Hyperparameters for the machine learning models

	RF, GBT & DART	
	Maximum number of trees	512
	Learning rate	$\in [0.001, 0.01, 0.1, 1]$
	Maximum depth	$U^{int}(2, 10)$
	Maximum number of leaves	$U^{int}(2, 512)$
	L1-regularization	$U(0, 0.1)$
	L2-regularization	$U(0, 0.1)$
	Feature fraction	$U(0.25, 1)$
	Bagging fraction	$U(0.25, 1)$
	Bagging frequency	$\in (1, 10, 50)$
DART	Dropout rate	$\in (0.05, 0.1, 0.15)$
DART	Probability of skipping dropout	$\in (0.25, 0.5)$
	NN	
	Learning rate	$\in [0.001, 0.01, 0.1, 1]$
	L1-regularization	$U(0, 0.1)$
	Dropout	$U(0, 0.5)$
	Number of hidden layers	$\in [1, 2, 3, 4, 5]$
	First layer size	$\in [32, 64, 128]$
	Batch size	$\in [2^{11}, 2^{12}, 2^{13}, 2^{14}]$

This table gives the hyperparameters that we tune and their respective boundaries. U (U^{int}) means drawing from a uniform (integer-wise uniform) distribution. Our choice of hyperparameters and their respective boundaries is based on Bali et al. (2023). We use the *Ray* Python framework to efficiently optimize the hyperparameters (Liaw et al., 2018).

Table B.3: Predictor variables for the machine learning models

	Variable	Compustat description	Financial statement	Component
1	aco	Current assets - other - total	Balance sheet	Current assets
2	acox	Current assets - other - sundry	Balance sheet	Current assets
3	act	Current assets - total	Balance sheet	Current assets
4	am	Amortization of intangibles	Income statement	Depreciation and amortization
5	ao	Assets - other	Balance sheet	Fixed assets
6	aoloch	Assets and liabilities - other - net change	Cash flow statement	Operating cash flow
7	aox	Assets - other - sundry	Balance sheet	Fixed assets
8	ap	Accounts payable - trade	Balance sheet	Liabilities
9	apalch	Accounts payable and accrued liabilities - increase/(decrease)	Cash flow statement	Operating cash flow
10	aqc	Acquisitions	Cash flow statement	Investing cash flow
11	aqi	Acquisitions - income contribution	Income statement	Interest and other
12	aqs	Acquisitions - sales contribution	Income statement	Sales
13	at	Assets - total	Balance sheet	Total assets
14	caps	Capital surplus/share premium reserve	Balance sheet	Equity
15	capx	Capital expenditures	Cash flow statement	Investing cash flow
16	capxv	Capital expend property, plant and equipment schd v	Cash flow statement	Investing cash flow
17	ceq	Common/ordinary equity - total	Balance sheet	Equity
18	ceql	Common equity - liquidation value	Balance sheet	Supplemental
19	ceqt	Common equity - tangible	Balance sheet	Supplemental

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
20	ch	Cash	Balance sheet	Current assets
21	che	Cash and short-term investments	Balance sheet	Current assets
22	chech	Cash and cash equivalents - increase/(decrease)	Cash flow statement	Total cash flow
23	cld2	Capitalized leases - due in 2nd year	Balance sheet	Supplemental
24	cld3	Capitalized leases - due in 3rd year	Balance sheet	Supplemental
25	cld4	Capitalized leases - due in 4th year	Balance sheet	Supplemental
26	cld5	Capitalized leases - due in 5th year	Balance sheet	Supplemental
27	cogs	Cost of goods sold	Income statement	Operating expenses
28	cstk	Common/ordinary stock (capital)	Balance sheet	Equity
29	cstkcv	Common stock-carrying value	Balance sheet	Supplemental
30	cstke	Common stock equivalents - dollar savings	Income statement	Interest and other
31	dc	Deferred charges	Balance sheet	Fixed assets
32	dclo	Debt - capitalized lease obligations	Balance sheet	Liabilities
33	dcpstk	Convertible debt and preferred stock	Balance sheet	Supplemental
34	dcvsr	Debt - senior convertible	Balance sheet	Liabilities
35	dcvsub	Debt - subordinated convertible	Balance sheet	Liabilities
36	dcvt	Debt - convertible	Balance sheet	Liabilities
37	dd	Debt - debentures	Balance sheet	Liabilities
38	dd1	Long-term debt due in one year	Balance sheet	Liabilities
39	dd2	Debt - due in 2nd year	Balance sheet	Liabilities

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
40	dd3	Debt - due in 3rd year	Balance sheet	Liabilities
41	dd4	Debt - due in 4th year	Balance sheet	Liabilities
42	dd5	Debt - due in 5th year	Balance sheet	Liabilities
43	dlc	Debt in current liabilities - total	Balance sheet	Liabilities
44	dltis	Long-term debt - issuance	Cash flow statement	Financing cash flow
45	dlto	Other long-term debt	Balance sheet	Liabilities
46	dltp	Long-term debt - tied to prime	Balance sheet	Liabilities
47	dltr	Long-term debt - reduction	Cash flow statement	Financing cash flow
48	dltt	Long-term debt - total	Balance sheet	Liabilities
49	dm	Debt - mortgages and other secured	Balance sheet	Liabilities
50	dn	Debt - notes	Balance sheet	Liabilities
51	do	Discontinued operations	Income statement	Interest and other
52	dp	Depreciation and amortization	Income statement	Depreciation and amortization
53	dpact	Depreciation, depletion and amortization (accumulated)	Balance sheet	Fixed assets
54	dpc	Depreciation and amortization (cash flow)	Cash flow statement	Operating cash flow
55	dpvieb	Depreciation (accumulated) - ending balance (schedule vi)	Balance sheet	Supplemental
56	ds	Debt-subordinated	Balance sheet	Liabilities
57	dudd	Debt - unamortized debt discount and other	Balance sheet	Liabilities
58	dv	Cash dividends (cash flow)	Cash flow statement	Financing cash flow
59	dvc	Dividends common/ordinary	Income statement	Dividends

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
60	dvp	Dividends - preferred/preference	Income statement	Dividends
61	dvpa	Preferred dividends in arrears	Balance sheet	Supplemental
62	dvt	Dividends - total	Income statement	Dividends
63	dxd2	Debt (excl capitalized leases) - due in 2nd year	Balance sheet	Supplemental
64	dxd3	Debt (excl capitalized leases) - due in 3rd year	Balance sheet	Supplemental
65	dxd4	Debt (excl capitalized leases) - due in 4th year	Balance sheet	Supplemental
66	dxd5	Debt (excl capitalized leases) - due in 5th year	Balance sheet	Supplemental
67	ebit	Earnings before interest and taxes	Income statement	EBIT
68	ebitda	Earnings before interest	Income statement	EBITDA
69	esub	Equity in earnings - unconsolidated subsidiaries	Income statement	Interest and other
70	esubc	Equity in net loss - earnings	Cash flow statement	Operating cash flow
71	exre	Exchange rate effect	Cash flow statement	Total cash flow
72	fatb	Property, plant, and equipment - buildings at cost	Balance sheet	Supplemental
73	fatc	Property, plant, and equipment - construction in progress at cost	Balance sheet	Supplemental
74	fate	Property, plant, and equipment - machinery and equipment at cost	Balance sheet	Supplemental
75	fatl	Property, plant, and equipment - leases at cost	Balance sheet	Supplemental
76	fatn	Property, plant, and equipment - natural resources at cost	Balance sheet	Supplemental
77	fato	Property, plant, and equipment - other at cost	Balance sheet	Supplemental
78	fatp	Property, plant, and equipment - land and improvements at cost	Balance sheet	Supplemental
79	fiao	Financing activities - other	Cash flow statement	Financing cash flow

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
80	fincf	Financing activities - net cash flow	Cash flow statement	Financing cash flow
81	fopo	Funds from operations - other	Cash flow statement	Operating cash flow
82	gp	Gross profit	Income statement	Operating expenses
83	ib	Income before extraordinary items	Income statement	Net income
84	ibadj	Income before extraordinary items - adjusted for common stock equivalents	Income statement	Net income
85	ibc	Income before extraordinary items (cash flow)	Cash flow statement	Operating cash flow
86	ibcom	Income before extraordinary items - available for common	Income statement	Net income
87	icapt	Invested capital - total	Balance sheet	Supplemental
88	idit	Interest and related income - total	Income statement	Interest and other
89	intan	Intangible assets - total	Balance sheet	Fixed assets
90	intc	Interest capitalized	Income statement	Interest and other
91	intpn	Interest paid - net	Cash flow statement	Operating cash flow
92	invch	Inventory - decrease (increase)	Cash flow statement	Operating cash flow
93	invfg	Inventories - finished goods	Balance sheet	Current assets
94	invo	Inventories - other	Balance sheet	Current assets
95	invrm	Inventories - raw materials	Balance sheet	Current assets
96	invt	Inventories - total	Balance sheet	Current assets
97	invwip	Inventories - work in process	Balance sheet	Current assets
98	itcb	Investment tax credit (balance sheet)	Balance sheet	Liabilities
99	itci	Investment tax credit (income account)	Income statement	Taxes

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
100	ivaco	Investing activities - other	Cash flow statement	Investing cash flow
101	ivaeq	Investment and advances - equity	Balance sheet	Fixed assets
102	ivao	Investment and advances - other	Balance sheet	Fixed assets
103	ivch	Increase in investments	Cash flow statement	Investing cash flow
104	ivncf	Investing activities - net cash flow	Cash flow statement	Investing cash flow
105	ivst	Short-term investments - total	Balance sheet	Current assets
106	ivstch	Short-term investments - change	Cash flow statement	Investing cash flow
107	lco	Current liabilities - other - total	Balance sheet	Liabilities
108	lcox	Current liabilities - other - sundry	Balance sheet	Liabilities
109	lct	Current liabilities - total	Balance sheet	Liabilities
110	lifr	Lifo reserve	Balance sheet	Supplemental
111	lo	Liabilities - other - total	Balance sheet	Liabilities
112	lse	Liabilities and stockholders equity - total	Balance sheet	Total liabilities and equity
113	lt	Liabilities - total	Balance sheet	Liabilities
114	mib	Noncontrolling interest (balance sheet)	Balance sheet	Liabilities
115	mii	Noncontrolling interest (income account)	Income statement	Interest and other
116	mrc1	Rental commitments - minimum - 1st year	Balance sheet	Supplemental
117	mrc2	Rental commitments - minimum - 2nd year	Balance sheet	Supplemental
118	mrc3	Rental commitments - minimum - 3rd year	Balance sheet	Supplemental
119	mrc4	Rental commitments - minimum - 4th year	Balance sheet	Supplemental

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
120	mrc5	Rental commitments - minimum - 5th year	Balance sheet	Supplemental
121	mrct	Rental commitments - minimum - 5 year total	Balance sheet	Supplemental
122	msa	Marketable securities adjustment	Balance sheet	Supplemental
123	ni	Net income (loss)	Income statement	Net income
124	niadj	Net income adjusted for common/ordinary stock (capital) equivalents	Income statement	Net income
125	nopi	Nonoperating income (expense)	Income statement	Interest and other
126	nopio	Nonoperating income (expense) - other	Income statement	Interest and other
127	np	Notes payable - short-term borrowings	Balance sheet	Liabilities
128	oancf	Operating activities - net cash flow	Cash flow statement	Operating cash flow
129	oiadp	Operating income after depreciation	Income statement	EBIT
130	oibdp	Operating income before depreciation	Income statement	EBITDA
131	pi	Pretax income	Income statement	EBT
132	ppegt	Property, plant and equipment - total (gross)	Balance sheet	Fixed assets
133	ppent	Property, plant and equipment - total (net)	Balance sheet	Fixed assets
134	ppeveb	Property, plant, and equipment - ending balance (schedule v)	Balance sheet	Supplemental
135	prstkc	Purchase of common and preferred stock	Cash flow statement	Financing cash flow
136	pstk	Preferred/preference stock (capital) - total	Balance sheet	Equity
137	pstkc	Preferred stock - convertible	Balance sheet	Equity
138	pstkl	Preferred stock - liquidating value	Balance sheet	Supplemental
139	pstkn	Preferred/preference stock - nonredeemable	Balance sheet	Equity

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
140	pstkr	Preferred/preference stock - redeemable	Balance sheet	Equity
141	pstkrv	Preferred stock - redemption value	Balance sheet	Supplemental
142	re	Retained earnings	Balance sheet	Equity
143	rea	Retained earnings - restatement	Balance sheet	Supplemental
144	reajo	Retained earnings - other adjustments	Balance sheet	Supplemental
145	recch	Accounts receivable - decrease (increase)	Cash flow statement	Operating cash flow
146	recco	Receivables - current - other	Balance sheet	Current assets
147	recd	Receivables - estimated doubtful	Balance sheet	Current assets
148	rect	Receivables - tota	Balance sheet	Current assets
149	recta	Retained earnings - cumulative translation adjustment	Balance sheet	Supplemental
150	rectr	Receivables - trade	Balance sheet	Current assets
151	reuna	Retained earnings - unadjusted	Balance sheet	Equity
152	revt	Revenue - total	Income statement	Sales
153	sale	Sales/turnover (net)	Income statement	Sales
154	seq	Stockholders equity - parent	Balance sheet	Equity
155	siv	Sale of investments	Cash flow statement	Investing cash flow
156	spi	Special items	Income statement	Interest and other
157	sppe	Sale of property	Cash flow statement	Operating cash flow
158	sppiv	Sale of property, plant and equipment and investments - gain (loss)	Cash flow statement	Operating cash flow
159	sstk	Sale of common and preferred stock	Cash flow statement	Financing cash flow

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
160	tlcf	Tax loss carry forward	Balance sheet	Supplemental
161	tstk	Treasury stock - total (all capital)	Balance sheet	Equity
162	tstkc	Treasury stock - common	Balance sheet	Equity
163	tstkp	Treasury stock - preferred	Balance sheet	Equity
164	txach	Income taxes - accrued - increase/(decrease)	Cash flow statement	Operating cash flow
165	txc	Income taxes - current	Income statement	Taxes
166	txdb	Deferred taxes (balance sheet)	Balance sheet	Liabilities
167	txdc	Deferred taxes (cash flow)	Income statement	Operating cash flow
168	txdfed	Deferred taxes-federal	Income statement	Taxes
169	txdfo	Deferred taxes-foreign	Income statement	Taxes
170	txdi	Income taxes - deferred	Income statement	Taxes
171	txditc	Deferred taxes and investment tax credit	Balance sheet	Liabilities
172	txds	Deferred taxes-state	Income statement	Taxes
173	txfed	Income taxes - federal	Income statement	Taxes
174	txfo	Income taxes - foreign	Income statement	Taxes
175	txo	Income taxes - other	Income statement	Taxes
176	txp	Income taxes payable	Balance sheet	Liabilities
177	txpd	Income taxes paid	Cash flow statement	Operating cash flow
178	txr	Income tax refund	Balance sheet	Current assets
179	txs	Income taxes - state	Income statement	Taxes

Continued on next page

Table B.3: Predictor variables for the machine learning models (continued)

	Variable	Compustat description	Financial statement	Component
180	txt	Income taxes - total	Income statement	Taxes
181	txw	Excise taxes	Income statement	Taxes
182	wcap	Working capital (balance sheet)	Balance sheet	Supplemental
183	xacc	Accrued expenses	Balance sheet	Liabilities
184	xi	Extraordinary items	Income statement	Interest and other
185	xido	Extraordinary items and discontinued operations	Income statement	Interest and other
186	xidoc	Extraordinary items and discontinued operations (cash flow)	Cash flow statement	Operating cash flow
187	xint	Interest and related expense - total	Income statement	Interest and other
188	xopr	Operating expenses - total	Income statement	Operating expenses
189	xpp	Prepaid expenses	Balance sheet	Current assets
190	xpr	Pension and retirement expense	Income statement	Operating expenses
191	xrent	Rental expense	Income statement	Operating expenses
192	xsga	Selling, general and administrative expense	Income statement	Operating expenses

This table reports the input variables used in our machine learning models. We also report the Compustat description, the financial statement group as well as the financial statement component group we assign to the respective variable. EBITDA denotes earnings before interest, taxes, depreciation and amortization. EBIT denotes earnings before interest and taxes. EBT denotes earnings before taxes. We scale all variables by common shares outstanding.

B.3 Implied cost of capital models

Table B.4: Implied cost of capital models

Name	Model/Description
GLS	$P_t = B_t + \Sigma_{\tau=1}^{11} \frac{\mathbb{E}_t[(ROE_{t+\tau} - ICC_{GLS}) \cdot B_{t+\tau-1}]}{(1+ICC_{GLS})^{\tau}} + \frac{\mathbb{E}_t[ROE_{t+12} - ICC_{GLS}) \cdot B_{t+11}]}{(1+ICC_{GLS})^{11} \cdot ICC_{GLS}}$ This model is given by Gebhardt et al. (2001). P_t denotes the stock price as of the estimation date in t, ICC_{GLS} denotes the implied cost of capital (ICC), B_t denotes the book value of equity per share in t and ROE_t is the return on equity in t. B_t is calculated using the clean surplus relation following Hou et al. (2012). ROE_t is calculated by dividing earnings per share (forecasts) E_t by B_{t-1}. For $ROE_{t+\tau}$ up to $\tau = 3$ we use the respective models earnings per share forecast. Afterwards, we assume $ROE_{t+\tau}$ to revert to the historical industry median by $\tau = 11$ (e.g., Hou et al., 2012). The industry median of ROE is derived using 10 years of data while excluding loss firms (e.g., Gebhardt et al., 2001). We expect ROE to be constant after $\tau = 11$.
CT	$P_t = B_t + \sum_{\tau=1}^{5} \frac{\mathbb{E}_t[(ROE_{t+\tau} - ICC_{CT}) \cdot B_{t+\tau-1}]}{(1+ICC_{CT})^{\tau}} + \frac{\mathbb{E}_t[(ROE_{t+5} - ICC_{CT}) \cdot B_{t+4}] \cdot (1+g)}{(1+ICC_{CT})^{5} \cdot (ICC_{CT} - g)}$ This model is given by Claus and Thomas (2001). P_t denotes the stock price as of the estimation date in t, ICC_{CT} denotes the implied cost of capital (ICC), B_t denotes the book value of equity per share in t, ROE_t is the return on equity in t and g is the perpetuity growth rate. B_t is calculated using the clean surplus relation following Hou et al. (2012). ROE_t is calculated by dividing earnings per share (forecasts) E_t by B_{t-1}. g is calculated as the current risk-free rate minus 3% (e.g., Hou et al., 2012).
OJ	$P_t = \frac{\mathbb{E}_t[E_{t+1}] \cdot (g_{st} - (\gamma - 1))}{(R-A)^{-} A^2}$, with $A = 0.5((\gamma - 1)\frac{\mathbb{E}_t[E_{t+1}] \cdot payout}{P_t})$, $g_{st} = 0.5(\frac{\mathbb{E}_t[E_{t+3}] - \mathbb{E}_t[E_{t+2}]}{\mathbb{E}_t[E_{t+2}]} - \frac{\mathbb{E}_t[E_{t+5}] - \mathbb{E}_t[E_{t+4}]}{\mathbb{E}_t[E_{t+4}]})$ This model is given by Ohlson and Juettner-Nauroth (2005). P_t denotes the stock price as of the estimation date in t, ICC_{OJ} denotes the implied cost of capital (ICC), $E_{t+\tau}$ is the earnings forecast for $t + \tau$, g_{st} is the short-term growth rate, γ is the perpetual growth rate and *payout* is the current payout ratio. g_{st} is calculated as the mean of forecasted earnings growth in $\tau = 3$ and $\tau = 5$ (e.g., Hou et al., 2012). γ is the current risk-free rate minus 3% (e.g., Hou et al., 2012). *payout* is calculated as dividends divided by earnings for profit firms and as dividends divided by $0.06 \cdot$ total assets for loss firms (e.g., Hou et al. (2012)).
MPEG	$P_t = \frac{\mathbb{E}_t[E_{t+2}] + (ICC_{MPEG} \cdot payout - 1) \cdot \mathbb{E}_t[E_{t+1}]}{ICC_{MPEG}^2}$ This model is given by Easton (2004). P_t denotes the stock price as of the estimation date in t, ICC_{MPEG} denotes the implied cost of capital (ICC) and $E_{t+\tau}$ denotes the earnings per share forecast for $t + \tau$. *payout* is derived as dividends divided by earnings for profit firms and as dividends divided by $0.06 \cdot$ total assets for loss firms (e.g., Hou et al., 2012).
GG	$P_t = \frac{\mathbb{E}_t[E_{t+1}]}{ICC_{GG}}$ This model is given by Gordon and Gordon (1997). P_t denotes the stock price as of the estimation date in t, ICC_{GG} denotes the implied cost of capital (ICC) and E_{t+1} denotes the earnings per share forecast for $t + 1$.

This table gives implied cost of capital (ICC) models that we base our composite ICC on. For simplicity, we drop the firm index i. The composite ICC that we use is derived as the average of these ICC.

B.4 Subsample analysis

Table B.5: Firm life cycle stages

Cash flows from	Intro	Growth	Mature	Shake-out	Shake-out	Shake-out	Decline	Decline
operating activities	-	+	+	+	+	-	-	-
investing activities	-	-	-	+	+	+	-	-
financing activities	+	+	-	+	-	+	+	-

This table provides the firm life cycle stages based on the respective firm's cash flows. A plus sign indicates a positive cash flow and a minus sign indicates a negative cash flow. The classification follows Dickinson (2011).

Table B.6: Industry classifications

No.	Industry Name	SIC Code Ranges
1	Consumer Durables, Nondurables, Wholesale, Retail, and Some Services	0100-0999, 2000-2399, 2700-2749, 2770-2799, 3100-3199, 3940-3989, 2500-2519, 2590-2599, 3630-3659, 3710-3711, 3714-3714, 3716-3716, 3750-3751, 3792-3792, 3900-3939, 3990-3999, 5000-5999, 7200-7299, 7600-7699
2	Manufacturing, Energy, and Utilities	2520-2589, 2600-2699, 2750-2769, 2800-2829, 2840-2899, 3000-3099, 3200-3569, 3580-3621, 3623-3629, 3700-3709, 3712-3713, 3715-3715, 3717-3749, 3752-3791, 3793-3799, 3860-3899, 1200-1399, 2900-2999, 4900-4949
3	Business Equipment, Telephone and Television Transmission	3570-3579, 3622-3622, 3660-3692, 3694-3699, 3810-3839, 7370-7372, 7373-7373, 7374-7374, 7375-7375, 7376-7376, 7377-7377, 7378-7378, 7379-7379, 7391-7391, 8730-8734, 4800-4899
4	Healthcare, Medical Equipment, and Drugs	2830-2839, 3693-3693, 3840-3859, 8000-8099
5	Other – Mines, Construction, Building Materials, Transportation, Hotels, Business Services, Entertainment, Finance	Remaining

This table provides industry classifications based on SIC code ranges. We follow the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`).

Table B.7: Median PAFE per life cycle stage

Life cycle stage	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
	ENTD	0.0573***	0.0685***	0.0737***	0.0832***	0.0915***
Intro	ENML	0.0464***	0.0597***	0.0696***	0.0778***	0.0799***
	ENML - ENTD	−0.0110***	−0.0087***	−0.0041	−0.0054	−0.0116*
	ENTD	0.0228***	0.0308***	0.0358***	0.0413***	0.0465***
Growth	ENML	0.0206***	0.0283***	0.0345***	0.0410***	0.0450***
	ENML - ENTD	−0.0022***	−0.0024***	−0.0013**	−0.0003	−0.0015
	ENTD	0.0222***	0.0289***	0.0331***	0.0369***	0.0417***
Mature	ENML	0.0196***	0.0261***	0.0306***	0.0343***	0.0392***
	ENML - ENTD	−0.0026***	−0.0028***	−0.0026***	−0.0026***	−0.0024***
	ENTD	0.0301***	0.0374***	0.0421***	0.0501***	0.0568***
Shake-out	ENML	0.0267***	0.0339***	0.0397***	0.0459***	0.0511***
	ENML - ENTD	−0.0034***	−0.0035***	−0.0024	−0.0042**	−0.0058***
	ENTD	0.0748***	0.0894***	0.0936***	0.1002***	0.1091***
Decline	ENML	0.0662***	0.0824***	0.0935***	0.0957***	0.1035***
	ENML - ENTD	−0.0086***	−0.0070***	−0.0001	−0.0044	−0.0056

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per life cycle stage for the two ensemble models. Life cycle stages are defined following Dickinson (2011) and as of estimation date. An overview is given in Table B.5 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table B.8: Median PAFE per size tercile

Firm size	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
	ENTD	0.0594***	0.0695***	0.0773***	0.0851***	0.0958***
Small	ENML	0.0522***	0.0651***	0.0731***	0.0780***	0.0837***
	ENML - ENTD	−0.0073***	−0.0045**	−0.0042*	−0.0071***	−0.0121***
	ENTD	0.0277***	0.0361***	0.0411***	0.0466***	0.0525***
Medium	ENML	0.0244***	0.0330***	0.0387***	0.0447***	0.0500***
	ENML - ENTD	−0.0033***	−0.0031***	−0.0024***	−0.0020***	−0.0025***
	ENTD	0.0167***	0.0224***	0.0260***	0.0294***	0.0330***
Large	ENML	0.0138***	0.0197***	0.0240***	0.0287***	0.0324***
	ENML - ENTD	−0.0030***	−0.0027***	−0.0020***	−0.0006	−0.0006

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per size tercile for the two ensemble models. Size is defined as the market capitalization as of estimation date. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Table B.9: Median PAFE per industry

Industry	Model	E_{t+1}	E_{t+2}	E_{t+3}	E_{t+4}	E_{t+5}
	ENTD	0.0257***	0.0324***	0.0375***	0.0414***	0.0457***
1	ENML	0.0225***	0.0290***	0.0342***	0.0387***	0.0431***
	ENML - ENTD	−0.0032***	−0.0034***	−0.0033***	−0.0028***	−0.0026***
	ENTD	0.0287***	0.0340***	0.0376***	0.0410***	0.0444***
2	ENML	0.0257***	0.0317***	0.0348***	0.0382***	0.0415***
	ENML - ENTD	−0.0030***	−0.0023***	−0.0028***	−0.0028***	−0.0028**
	ENTD	0.0300***	0.0382***	0.0424***	0.0454***	0.0519***
3	ENML	0.0258***	0.0333***	0.0371***	0.0406***	0.0461***
	ENML - ENTD	−0.0042***	−0.0049***	−0.0053***	−0.0049***	−0.0059***
	ENTD	0.0296***	0.0378***	0.0439***	0.0515***	0.0569***
4	ENML	0.0246***	0.0351***	0.0400***	0.0471***	0.0504***
	ENML - ENTD	−0.0050***	−0.0026***	−0.0039***	−0.0043*	−0.0065**
	ENTD	0.0263***	0.0340***	0.0397***	0.0468***	0.0555***
5	ENML	0.0233***	0.0321***	0.0401***	0.0474***	0.0539***
	ENML - ENTD	−0.0031***	−0.0019*	0.0004	0.0006	−0.0016

This table reports the time-series averages of the median price scaled absolute forecasting errors (PAFEs) per industry for the two ensemble models. Industries are defined following the Fama-French 5 industry portfolios classification, available on Kenneth French's website (`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`), and as of estimation date. An overview is given in B.6 in the Appendix. E_{t+1} to E_{t+5} denote one- to five-year ahead earnings. ENTD denotes the traditional ensemble and ENML denotes the machine learning ensemble. The PAFE is calculated as the difference between actual and forecasted earnings per share, scaled by price at the end of June of the respective estimation year. ***, **, and * denote statistical significance at the 1%, the 5% and the 10% level, respectively. Standard errors used for deriving statistical significance are adjusted following Newey and West (1987) assuming a lag length of three years.

Appendix C

Appendix to Chapter 4

C.1 Autoencoder fit

The bottleneck determines how much information the autoencoder compresses when capturing joint patterns in firm characteristics. A very small bottleneck underfits and produces large reconstruction errors, while a very large one overfits and begins to reproduce noise. The model must therefore balance flexibility and parsimony to avoid both under- and overfitting. To illustrate this tradeoff, I compare models with four, sixteen, and sixty four bottleneck dimensions. To make a more accurate comparison, I train the models without regularization and with only one hidden layer of 128 units. Figure C.1 displays the distribution of ATYP across specifications. The dashed vertical lines mark the 10th and 90th percentile values of each distribution. The four dimensional model fails to capture the heterogeneity in the data and generates broad, right skewed errors, whereas the sixty four dimensional model nearly memorizes the data and yields errors close to zero. The intermediate model with sixteen dimensions achieves the most balanced performance. Its reconstruction errors fall in a moderate and symmetric range, indicating that it captures the essential nonlinear patterns without absorbing noise. This specification provides a practical and economically meaningful basis for the ATYP measure.

For my measure to be valid, it is crucial to verify that the network learns the latent economic structure rather than simply memorizing the inputs. Figure C.2 plots the pairwise correlations of the raw characteristics against the correlations of the corresponding residuals. In the raw data, characteristics exhibit strong systematic comovements that reflect standard economic forces. However, the figure shows that the residual correlations

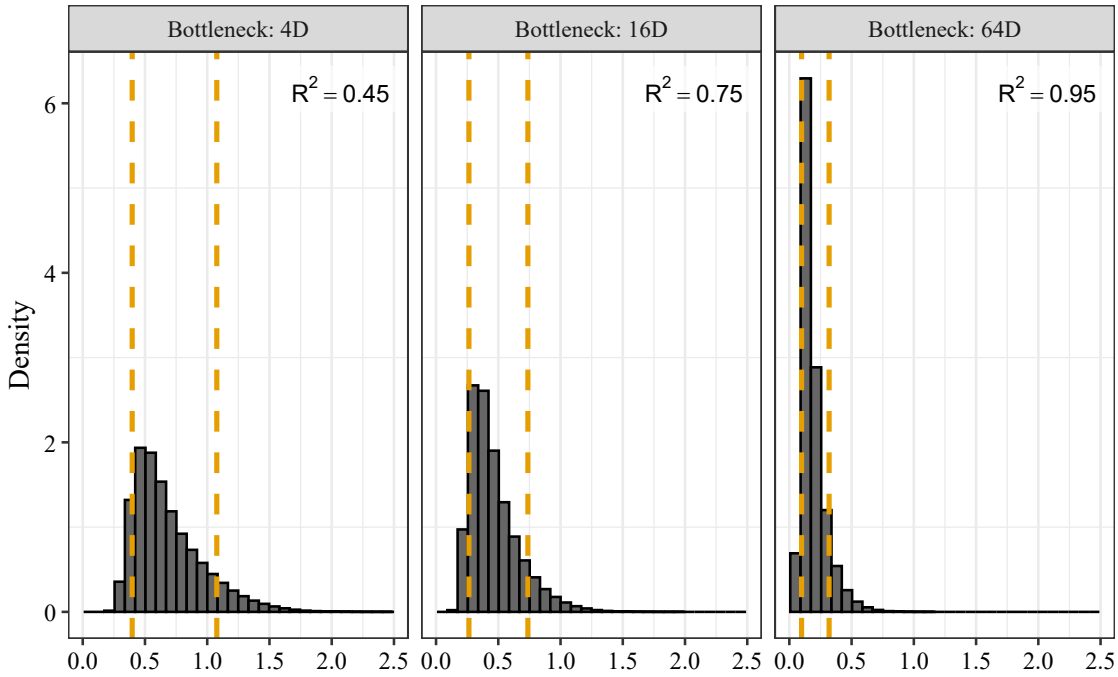


Figure C.1: Distribution of ATYP across bottleneck dimensions
The figure plots the density of ATYP for autoencoders with four, 16, and 64 latent factors. Dashed vertical lines mark the 10th and 90th percentiles. The R^2 is the measure of overall reconstruction fit.

collapse toward zero and the fitted slope is nearly flat. This transformation demonstrates that the network effectively filters out the common variation present in the original data. The result confirms that the residual space contains almost no systematic structure. Consequently, ATYP isolates idiosyncratic firm specific departures from the typical patterns the network has learned. This distinction ensures that my measure captures unique processing difficulty rather than exposure to common firm characteristics.

I also examine which inputs are harder to reconstruct and which inputs most influence the bottleneck, i.e., the latent factors. At the feature level I rank characteristics by mean squared reconstruction error (MSE) and by importance in changing the latent features, defined as the mean absolute gradient of the bottleneck with respect to each input. The most informative view is at the cluster level. Figure C.3 summarizes mean squared error and importance by cluster. *Investment*, *Profit Growth*, *Value* and *Quality* are all relatively hard to reconstruct and highly influential for the latent representation, while *Size* and *Accruals* are easier to reconstruct and less influential. I observe that unconditional characteristics, most notably *Size*, exhibit the lowest reconstruction errors and bottleneck importance. This indicates that market capitalization acts as a primary,

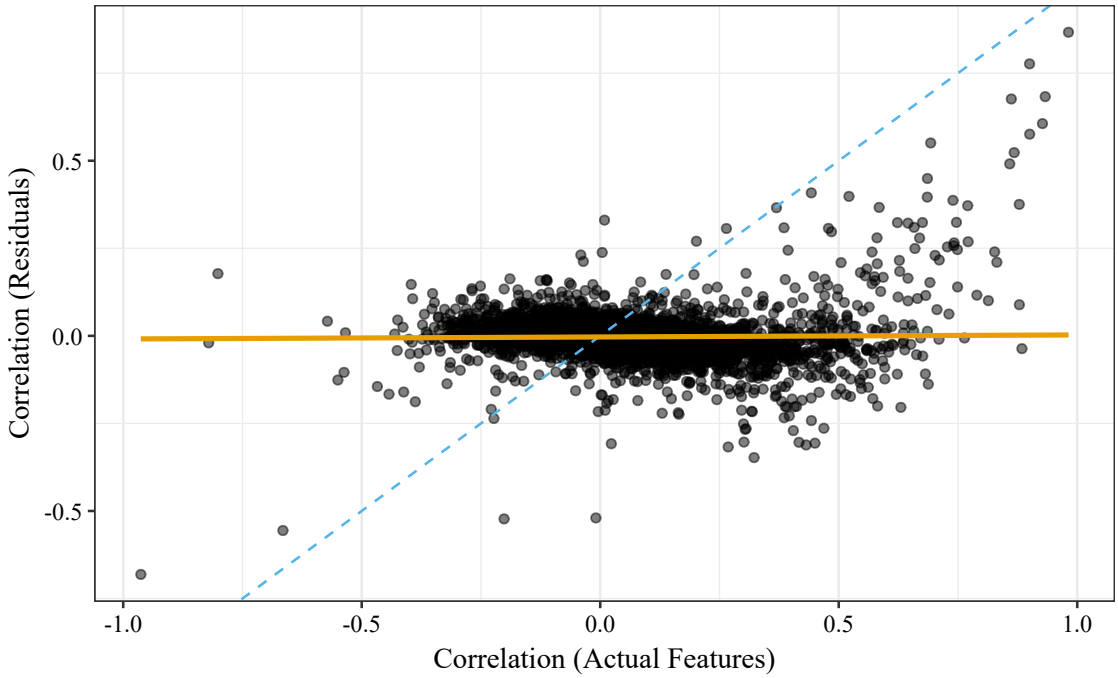


Figure C.2: Pairwise correlations of raw characteristics vs. residuals
Pairwise correlations of raw characteristics versus correlations of residuals, i.e., the reconstructed characteristics. Correlation is measured using Pearson's correlation coefficient. The 45-degree dashed line indicates equality; the fitted line is the trend. The sample is from 1981 to 2023.

stable sorting category that follows linear, predictable patterns easily captured by the model. In sharp contrast, the model exhibits the greatest difficulty in reconstructing conditional characteristics, such as *Investment*, *Profit Growth* and *Value*. This distinction is economically fundamental, because the valuation implication of a variable like asset growth is inherently context-dependent, i.e., high growth may imply value creation for a highly profitable firm but financial distress or agency costs for a firm with poor cash flows. That the autoencoder assigns the highest importance and error to these features confirms that ATYP successfully isolates firms with complex, nonlinear economic profiles, rather than simply identifying small or obscure firms.

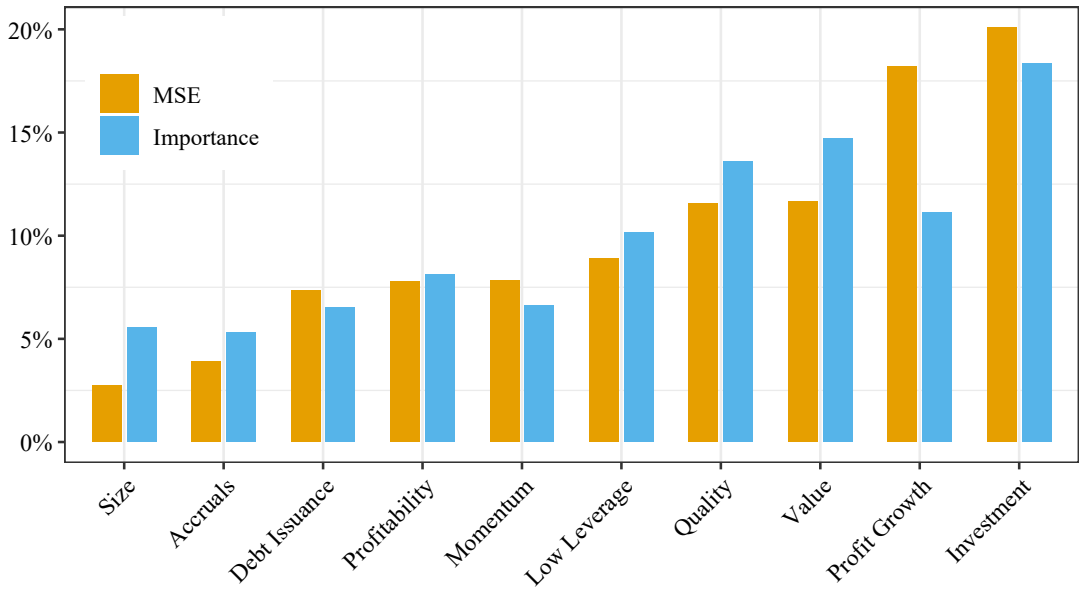


Figure C.3: MSE and bottleneck importance by cluster
The figure summarizes mean squared error and importance by cluster. Mean squared error is calculated for the residuals, i.e., the reconstructed characteristics compared to the actual characteristics. The importance is measured as the mean absolute gradient of the characteristics w.r.t. the bottleneck layer, i.e., the output of the encoder in Equation (4.1). The clusters are taken from Jensen et al. (2023).

C.2 Supplementary figures

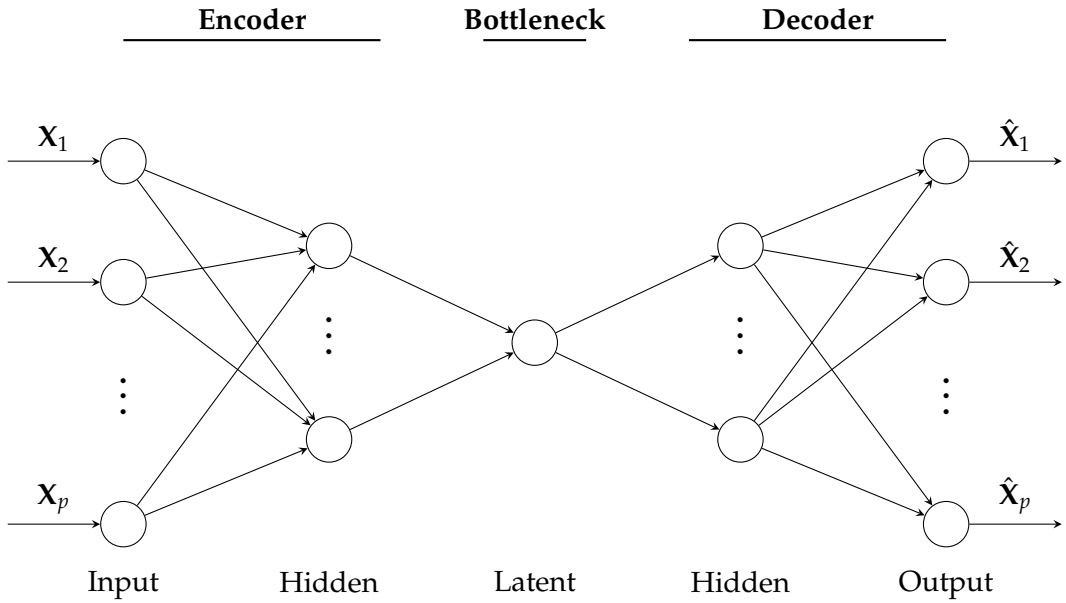


Figure C.4: Graphical illustration of an autoencoder
This figure illustrates an autoencoder model graphically. The encoder compresses the inputs $\mathbf{X}$ into a small latent bottleneck space in multiple nonlinear hidden layers. The decoder then attempts to reconstruct the original inputs, $\mathbf{X}$, from the latent space by predicting $\hat{\mathbf{X}}$.

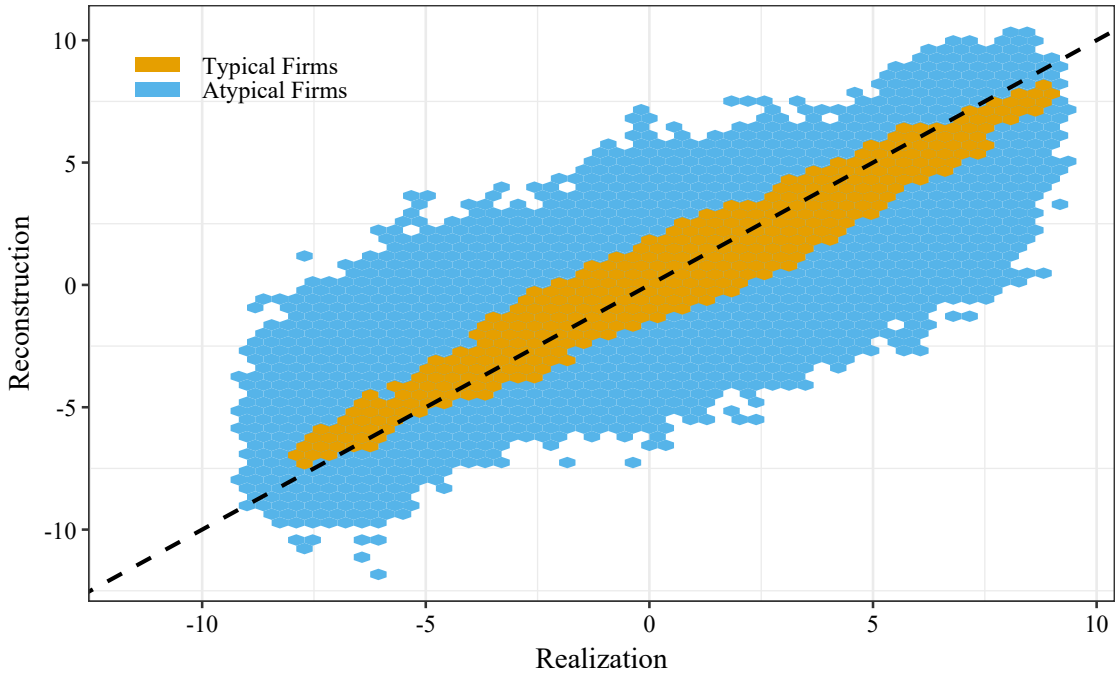


Figure C.5: Reconstruction accuracy and atypical firms
The figure plots realized firm characteristics on the horizontal axis and their reconstructed counterparts on the vertical axis. Each point represents a firm–characteristic observation. Firms that the autoencoder reconstructs accurately (lowest ATYP decile) are shown in orange and labeled as typical firms, while those with large reconstruction errors (highest ATYP decile) are shown in blue and labeled as atypical firms. The 45-degree dashed line indicates perfect reconstruction.

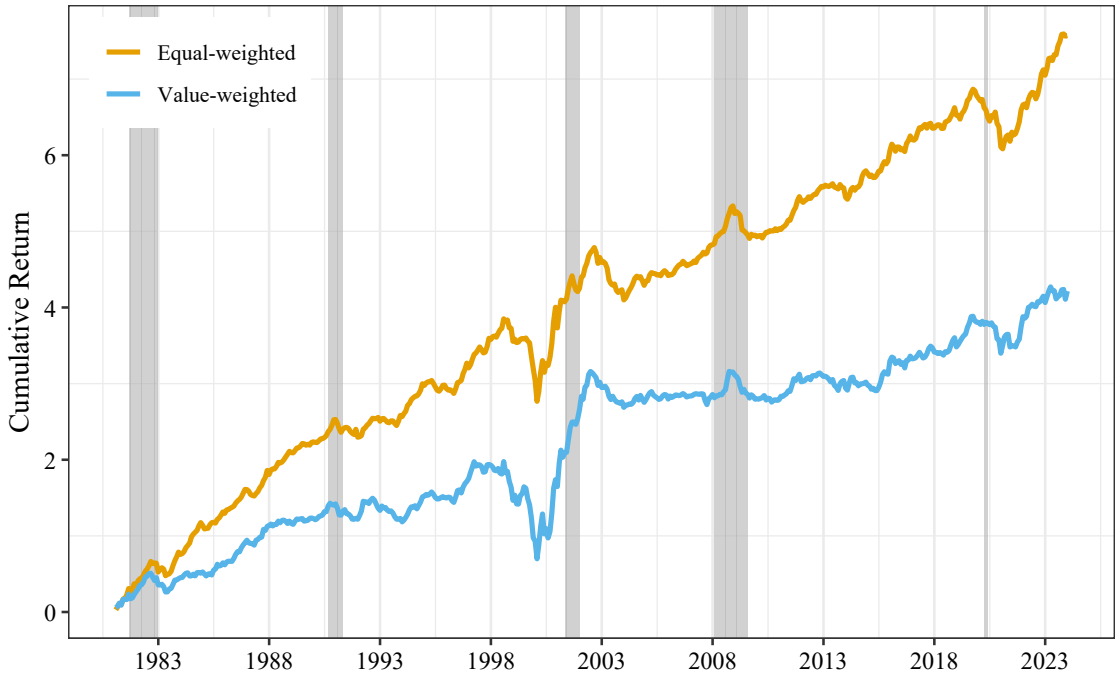


Figure C.6: Cumulative performance over time for ATYP portfolio sorts
The plot shows the cumulative sum of the equal- and value-weighted decile spread (L–H) portfolios' returns over time. Shaded areas represent NBER-dated recessions.

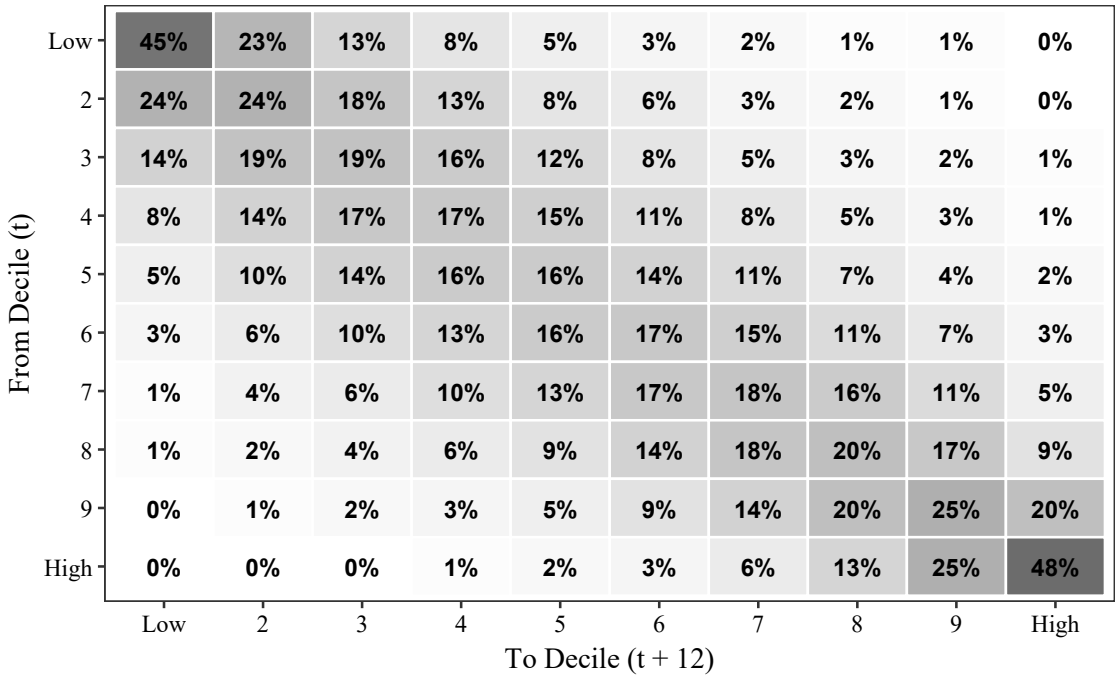


Figure C.7: Transition probabilities
The figure displays transition probabilities for ATYP deciles over a 12-month horizon. Each month, all firms are ranked by ATYP and assigned into decile portfolios from Low (lowest ATYP) to High (highest ATYP). The same sorting procedure is applied 12 months later. The entries show the average percentage of firms that move from a given ATYP decile at month t (rows) to each ATYP decile at month $t + 12$ (columns). Probabilities are time-series averages over the full sample period.

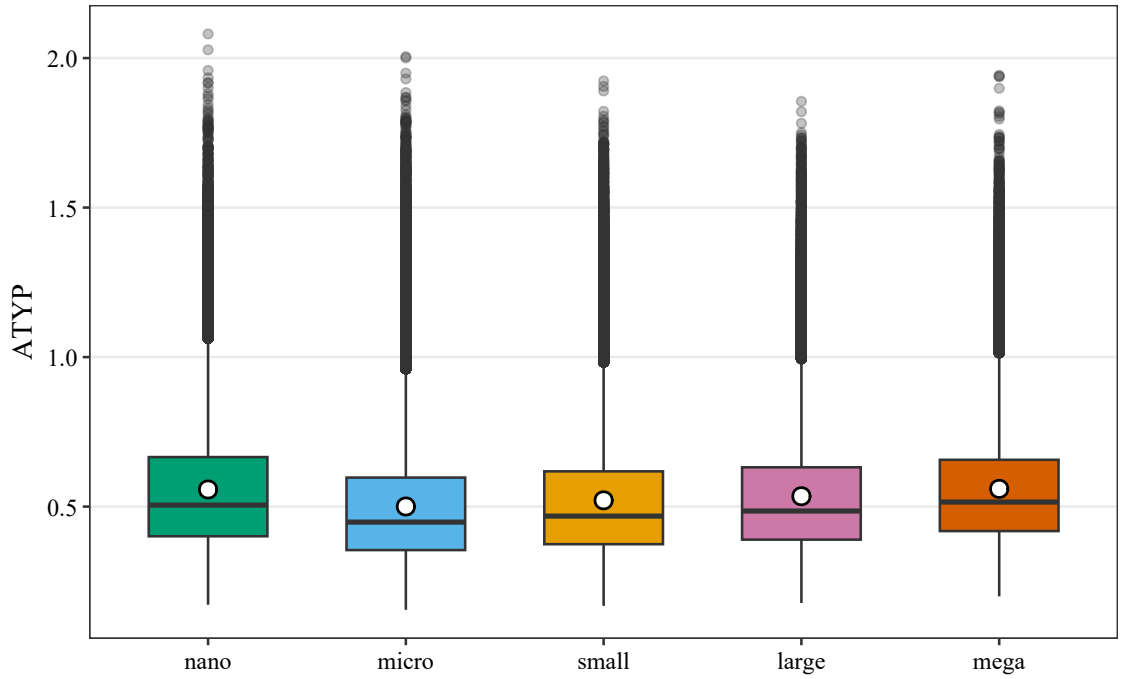


Figure C.8: Distribution of ATYP within size groups
The figure presents the distribution of ATYP scores across five size groups: nano, micro, small, large, and mega. ATYP is re-estimated within each size group following the full preprocessing and model training procedure. This construction removes any mechanical relation between firm size and ATYP. The boxplots show the distribution of ATYP across the size groups. The white circle denotes the mean ATYP value in each group.

C.3 Supplementary tables

Table C.1: Average stock characteristics of ATYP-sorted portfolios

	Low	2	3	4	5	6	7	8	9	High	H-L	t-stat
BETA	1.05	1.08	1.11	1.13	1.14	1.14	1.14	1.15	1.15	1.15	0.1**	(2.2)
SIZE	1.30	0.96	0.73	0.55	0.40	0.30	0.21	0.16	0.13	0.08	-1.22	(-0.77)
BM	0.53	0.55	0.55	0.56	0.56	0.55	0.53	0.52	0.47	0.40	-0.13***	(-3)
STR	0.01	0.01	0.01	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.03***	(-11.83)
MOM	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.02	0.00	-0.04	-0.15***	(-4.04)
OP	0.24	0.23	0.21	0.20	0.18	0.18	0.16	0.13	0.06	-0.05	-0.29***	(-12.37)
AG	0.06	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.10	0.21	0.15*	(1.78)
TURN	0.39	0.42	0.43	0.43	0.43	0.43	0.43	0.45	0.48	0.54	0.16	(0.6)
ILLIQ	0.06	0.09	0.13	0.16	0.18	0.20	0.21	0.23	0.27	0.34	0.28*	(1.71)
IVOL	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.03***	(3.87)
SUE	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.01	-0.02	0	(-0.19)
MAX	0.04	0.05	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.08	0.04***	(14.14)
AGE	2.74	2.38	2.20	2.05	1.87	1.73	1.58	1.42	1.25	1.10	-1.65***	(-18.71)
AFD	-0.04	-0.05	-0.05	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.02***	(-5.32)
#ANA	6.08	5.46	4.89	4.38	3.82	3.42	3.08	2.69	2.39	1.78	-4.3***	(-7.33)

The table reports the time-series averages of the monthly cross-sectional median for stock characteristics across decile portfolios sorted on ATYP. Reported characteristics are defined in Section 4.3.2. The last two columns present high-minus-low differences and Newey and West (1987) adjusted t-statistics with six lags. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.2: Fama-MacBeth cross-sectional regressions with post-earnings-announcement drift

	Excess Return $t+1$	Excess Return $t+2$	Excess Return $t+3$
ATYP	-0.75***	-0.94***	-1.01***
	(-3.31)	(-4.51)	(-4.96)
SUE	-0.02	0.00	0.02
	(-0.62)	(0.11)	(0.92)
ATYP x SUE	0.18***	0.12**	0.04
	(3.44)	(2.46)	(0.84)
Controls	Yes	Yes	Yes

The table reports results from Fama–MacBeth cross-sectional regressions linking post-earnings-announcement drift to information-processing difficulty. The dependent variable is the cumulative excess return over horizons $t+1$, $t+1$ to $t+2$, and $t+1$ to $t+3$. Independent variables include ATYP, standardized unexpected earnings (SUE), and their interaction. The controls considered are defined in Section 4.3.2. Standard errors are Newey and West (1987) adjusted with six lags. t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.3: Fama-MacBeth cross-sectional regressions with limits to arbitrage proxies

	Excess Return	Excess Return	Excess Return
ATYP	0.40	-0.72***	-1.35***
	(1.07)	(-3.09)	(-4.84)
ATYP x IVOL	-29.65***		
	(-4.22)		
ATYP x ILLIQ		-0.05	
		(-0.80)	
ATYP x AGE			0.29***
			(4.17)
Controls	Yes	Yes	Yes

The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is interacted with idiosyncratic volatility, illiquidity, and firm age, respectively. The controls considered are defined in Section 4.3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.4: Fama-MacBeth cross-sectional regressions with attention proxies

	Excess Return	Excess Return	Excess Return
ATYP	-0.85***	-1.27***	-0.96***
	(-3.64)	(-5.04)	(-4.25)
ATYP x SIZE	2.23***		
	(3.35)		
ATYP x #ANA		0.13***	
		(5.49)	
ATYP x SHORT			0.02***
			(4.16)
Controls	Yes	Yes	Yes

The table reports the results of monthly Fama and MacBeth (1973) regressions of excess stock returns on ATYP and a broad set of firm characteristics. ATYP is interacted with firm size, analyst coverage, and short interest, respectively. The controls considered are defined in Section 4.3.2. Reported coefficients represent average slopes across months, with Newey and West (1987) adjusted t-statistics (six lags) shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.5: Univariate portfolio sorts on ATYP by firm characteristic clusters

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
Accruals	-0.74***	-0.48***	-0.07	0.19
	(-4.24)	(-3.67)	(-0.36)	(1.21)
Debt Issuance	-0.79***	-0.43***	-0.27	0.02
	(-5.07)	(-4.36)	(-1.62)	(0.20)
Investment	-1.53***	-0.96***	-0.80***	-0.34**
	(-5.41)	(-6.18)	(-2.93)	(-2.12)
Low Leverage	-0.84***	-0.42***	-0.43**	-0.23
	(-3.76)	(-2.84)	(-2.27)	(-1.44)
Momentum	-0.76***	-0.44***	-0.08	0.26
	(-3.37)	(-3.07)	(-0.35)	(1.31)
Profit Growth	-1.18***	-0.69***	-0.57**	-0.24
	(-4.77)	(-4.60)	(-2.34)	(-1.55)
Profitability	-1.28***	-0.78***	-0.28	0.07
	(-4.45)	(-4.60)	(-1.19)	(0.42)
Quality	-1.08***	-0.57***	-0.34	0.01
	(-4.22)	(-3.51)	(-1.34)	(0.04)
Size	0.22	0.45***	-0.03	0.16*
	(1.61)	(3.67)	(-0.21)	(1.96)
Value	-0.85***	-0.47***	-0.32	-0.06
	(-3.15)	(-2.69)	(-1.26)	(-0.41)

The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts on cluster-specific squared reconstruction errors used to construct ATYP. Each row corresponds to a firm characteristic cluster, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.6: Univariate portfolio sorts on industry-adjusted ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.91***	(4.09)	0.21*	(1.72)	0.17***	(2.62)	0.06	(0.90)	0.11**	(2.07)
2	0.93***	(3.88)	0.18	(1.59)	0.18***	(3.06)	0.11*	(1.85)	0.18***	(3.43)
3	0.86***	(3.39)	0.10	(0.84)	0.11**	(2.03)	0.10*	(1.77)	0.15***	(2.93)
4	0.86***	(3.20)	0.06	(0.51)	0.09	(1.40)	0.11	(1.54)	0.18***	(2.72)
5	0.78***	(2.75)	-0.02	(-0.17)	0.00	(0.00)	0.06	(0.87)	0.16**	(2.32)
6	0.77**	(2.53)	-0.07	(-0.44)	-0.03	(-0.35)	0.07	(0.75)	0.17**	(2.05)
7	0.61*	(1.88)	-0.25	(-1.54)	-0.19**	(-2.18)	-0.06	(-0.68)	0.05	(0.53)
8	0.55	(1.57)	-0.34*	(-1.89)	-0.26**	(-2.52)	-0.07	(-0.67)	0.06	(0.61)
9	0.22	(0.57)	-0.70***	(-3.24)	-0.60***	(-4.26)	-0.31**	(-2.44)	-0.18	(-1.42)
10	-0.49	(-1.11)	-1.48***	(-5.49)	-1.33***	(-6.91)	-0.97***	(-5.87)	-0.81***	(-4.91)
H-L	-1.40***	(-4.73)	-1.69***	(-6.42)	-1.50***	(-6.89)	-1.02***	(-6.18)	-0.93***	(-5.43)
Panel B: Value-Weighted										
1	0.74***	(4.22)	0.12	(1.22)	0.08	(0.85)	-0.12	(-1.64)	-0.09	(-1.26)
2	0.74***	(4.11)	0.11	(1.44)	0.09	(1.14)	-0.09	(-1.26)	-0.09	(-1.33)
3	0.71***	(3.60)	0.04	(0.74)	0.06	(1.09)	-0.03	(-0.48)	-0.02	(-0.38)
4	0.70***	(3.13)	-0.02	(-0.32)	0.03	(0.43)	0.06	(0.68)	0.06	(0.70)
5	0.79***	(3.57)	0.05	(0.70)	0.08	(1.13)	0.04	(0.53)	0.10	(1.33)
6	0.73***	(3.16)	-0.03	(-0.37)	0.03	(0.41)	0.06	(0.80)	0.07	(0.91)
7	0.72***	(2.69)	-0.10	(-0.97)	-0.02	(-0.17)	0.11	(1.17)	0.10	(1.06)
8	0.71**	(2.54)	-0.12	(-0.93)	-0.04	(-0.33)	0.09	(0.76)	0.12	(1.11)
9	0.49	(1.50)	-0.39**	(-2.38)	-0.24*	(-1.77)	-0.02	(-0.19)	-0.01	(-0.09)
10	0.30	(0.74)	-0.74***	(-3.23)	-0.54***	(-2.66)	-0.12	(-0.66)	-0.12	(-0.65)
H-L	-0.45	(-1.40)	-0.86***	(-2.99)	-0.62**	(-2.46)	-0.00	(-0.00)	-0.03	(-0.14)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the industry-adjusted measure of ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score estimated within one-digit SIC industries using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.7: Within-industry portfolio sorts on industry-adjusted ATYP

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
Mining & Construction	-1.17***	-0.95***	-0.33	-0.19
	(-3.54)	(-2.91)	(-0.91)	(-0.57)
Manufacturing (Food, Textile, etc.)	-0.91**	-0.39	-0.92**	-0.53*
	(-2.38)	(-1.38)	(-2.39)	(-1.73)
Manufacturing (Machinery, Transport, etc.)	-1.60***	-1.03***	-0.87**	-0.21
	(-4.58)	(-4.37)	(-2.21)	(-0.72)
Transportation, Communications	-1.11***	-0.52*	-0.38	-0.09
	(-2.77)	(-1.68)	(-1.13)	(-0.25)
Wholesale & Retail Trade	-1.64***	-1.08***	-0.10	0.24
	(-4.72)	(-3.98)	(-0.27)	(0.74)
Services	-1.51***	-1.22***	-0.70*	-0.20
	(-4.29)	(-4.43)	(-1.66)	(-0.55)
Public Administration	-1.62***	-1.28***	-1.00**	-0.65*
	(-3.83)	(-3.69)	(-2.02)	(-1.65)

The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts conducted within each industry based on the industry-adjusted ATYP measure. Each row corresponds to a one-digit SIC industry, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.8: Univariate portfolio sorts on size-adjusted ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	1.01***	(4.09)	0.33**	(2.23)	0.28***	(4.07)	0.17**	(2.46)	0.23***	(3.72)
2	0.92***	(3.58)	0.20	(1.41)	0.18***	(2.67)	0.10	(1.49)	0.18***	(2.79)
3	0.90***	(3.32)	0.14	(1.04)	0.13**	(2.06)	0.11	(1.56)	0.18***	(2.76)
4	0.87***	(3.26)	0.11	(0.78)	0.11*	(1.86)	0.11	(1.60)	0.20***	(3.20)
5	0.78***	(2.69)	-0.01	(-0.04)	0.01	(0.12)	0.05	(0.56)	0.14*	(1.70)
6	0.67**	(2.24)	-0.17	(-1.19)	-0.14*	(-1.77)	-0.07	(-0.91)	0.03	(0.39)
7	0.53*	(1.67)	-0.33**	(-2.02)	-0.28***	(-2.94)	-0.14	(-1.48)	-0.03	(-0.36)
8	0.46	(1.39)	-0.43**	(-2.53)	-0.34***	(-3.45)	-0.14	(-1.48)	-0.01	(-0.11)
9	0.11	(0.29)	-0.83***	(-4.31)	-0.70***	(-5.56)	-0.41***	(-3.77)	-0.29***	(-2.77)
10	-0.31	(-0.74)	-1.33***	(-5.37)	-1.13***	(-6.51)	-0.69***	(-4.91)	-0.57***	(-4.11)
H-L	-1.33***	(-4.81)	-1.66***	(-6.31)	-1.41***	(-6.98)	-0.86***	(-5.98)	-0.80***	(-5.49)
Panel B: Value-Weighted										
1	0.82***	(4.35)	0.20*	(1.94)	0.15*	(1.68)	-0.08	(-0.90)	-0.06	(-0.66)
2	0.78***	(4.15)	0.14	(1.49)	0.10	(1.25)	-0.10	(-1.54)	-0.09	(-1.25)
3	0.68***	(3.71)	0.06	(0.91)	0.03	(0.49)	-0.11*	(-1.74)	-0.13**	(-2.04)
4	0.68***	(3.47)	0.02	(0.30)	0.02	(0.30)	-0.07	(-1.06)	-0.07	(-1.05)
5	0.77***	(4.07)	0.12	(1.59)	0.13	(1.64)	0.01	(0.19)	-0.00	(-0.04)
6	0.74***	(3.54)	0.05	(0.58)	0.06	(0.80)	0.04	(0.46)	0.04	(0.48)
7	0.63***	(2.77)	-0.10	(-1.23)	-0.04	(-0.49)	-0.05	(-0.71)	-0.04	(-0.47)
8	0.90***	(3.93)	0.13*	(1.69)	0.19**	(2.28)	0.24***	(3.06)	0.28***	(3.44)
9	0.55*	(1.93)	-0.33***	(-3.03)	-0.21**	(-2.17)	-0.03	(-0.25)	0.02	(0.16)
10	0.53	(1.49)	-0.41*	(-1.88)	-0.18	(-1.10)	0.23*	(1.78)	0.21*	(1.64)
H-L	-0.30	(-1.03)	-0.61**	(-2.09)	-0.33	(-1.50)	0.31*	(1.80)	0.26	(1.59)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the size-adjusted measure of ATYP. At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. Then, at the end of each month t, all stocks are sorted into deciles according to their ATYP score estimated within size groups using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.9: Within-size group portfolio sorts on size-adjusted ATYP

	Equal-Weighted		Value-Weighted	
	H-L	FF6	H-L	FF6
nano	-1.84***	-1.59***	-1.80***	-1.42***
	(-4.85)	(-4.12)	(-4.12)	(-2.93)
micro	-1.85***	-1.32***	-1.63***	-0.99***
	(-5.94)	(-6.41)	(-4.97)	(-4.97)
small	-1.03***	-0.41***	-0.99***	-0.36**
	(-3.58)	(-2.70)	(-3.33)	(-2.34)
large	-0.49	0.06	-0.42	0.13
	(-1.54)	(0.34)	(-1.37)	(0.67)
mega	-0.33	0.31*	-0.01	0.58***
	(-1.11)	(1.93)	(-0.04)	(3.24)

The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts conducted within each size group based on the size-adjusted ATYP measure. At the end of each month, firms are sorted into five groups based on their market equity: nano, micro, small, large, and mega. Each row corresponds to a size group, with high-minus-low (H–L) decile return spreads reported for equal-weighted and value-weighted portfolios. FF6 denotes the intercept from a regression of portfolio excess returns on the Fama and French (2018) six-factor model. Newey and West (1987) adjusted t-statistics with six lags are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The sample is from 1981 to 2023.

Table C.10: Univariate portfolio sorts on ATYP with alternative factor models

	Excess Return	t-stat	FF6	t-stat	Q5	t-stat	DHS	t-stat	DMRS	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.25)	0.08	(1.52)	0.10	(1.07)	0.38***	(3.48)	0.42*	(1.72)
2	0.93***	(3.96)	0.12**	(2.29)	0.16**	(2.01)	0.44***	(3.72)	0.46*	(1.68)
3	0.94***	(3.77)	0.19***	(3.44)	0.23***	(2.94)	0.47***	(3.60)	0.51*	(1.82)
4	0.91***	(3.45)	0.22***	(3.74)	0.27***	(3.49)	0.54***	(3.77)	0.51*	(1.68)
5	0.79***	(2.77)	0.15**	(2.02)	0.22**	(2.51)	0.50***	(2.96)	0.44	(1.36)
6	0.78***	(2.59)	0.20**	(2.40)	0.28***	(3.03)	0.49***	(2.76)	0.47	(1.37)
7	0.65*	(1.96)	0.10	(1.19)	0.19**	(2.04)	0.46**	(2.35)	0.37	(1.02)
8	0.41	(1.16)	-0.06	(-0.54)	0.05	(0.43)	0.32	(1.45)	0.23	(0.60)
9	0.12	(0.31)	-0.20	(-1.61)	-0.09	(-0.63)	0.22	(0.89)	0.04	(0.11)
10	-0.53	(-1.17)	-0.76***	(-4.36)	-0.63***	(-3.26)	-0.29	(-0.95)	-0.51	(-1.11)
H-L	-1.47***	(-4.54)	-0.84***	(-4.52)	-0.73***	(-3.24)	-0.67**	(-2.56)	-0.93***	(-3.34)
Panel B: Value-Weighted										
1	0.80***	(4.35)	-0.05	(-0.67)	-0.04	(-0.45)	0.11	(1.35)	0.34	(1.53)
2	0.66***	(3.65)	-0.14**	(-2.09)	-0.12	(-1.62)	-0.02	(-0.22)	0.23	(1.05)
3	0.72***	(3.64)	-0.04	(-0.61)	-0.02	(-0.36)	0.03	(0.40)	0.32	(1.35)
4	0.77***	(3.77)	0.06	(0.77)	0.08	(1.08)	0.02	(0.20)	0.45*	(1.72)
5	0.72***	(3.16)	0.04	(0.54)	0.07	(0.93)	0.00	(0.01)	0.42*	(1.71)
6	0.79***	(3.31)	0.18**	(2.03)	0.22***	(2.65)	0.12	(1.37)	0.54*	(1.94)
7	0.85***	(3.21)	0.26***	(2.78)	0.36***	(3.77)	0.38***	(3.51)	0.64**	(2.05)
8	0.48*	(1.79)	-0.09	(-0.92)	-0.05	(-0.45)	-0.02	(-0.17)	0.40	(1.33)
9	0.38	(1.17)	-0.05	(-0.35)	-0.04	(-0.27)	0.17	(1.25)	0.46	(1.39)
10	-0.02	(-0.04)	-0.33*	(-1.92)	-0.27	(-1.25)	-0.06	(-0.32)	0.22	(0.47)
H-L	-0.82**	(-2.50)	-0.28	(-1.49)	-0.23	(-0.92)	-0.17	(-0.84)	-0.11	(-0.35)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the Fama and French (2018) six-factor model including momentum (FF6), the Hou et al. (2014) q-factor model (HXZ), the Daniel et al. (2019) behavioral factor model (DHS), and the Daniel et al. (2020) characteristic-efficient factor model (DMRS). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023 (December 2018 in case of DHS).

Table C.11: Univariate portfolio sorts on ATYP for WRDS financial ratios

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	1.16***	(4.73)	0.42***	(2.97)	0.36***	(4.95)	0.25***	(3.26)	0.33***	(4.91)
2	1.07***	(4.19)	0.31**	(2.18)	0.28***	(3.92)	0.20**	(2.51)	0.29***	(4.39)
3	0.93***	(3.55)	0.15	(1.08)	0.14**	(2.20)	0.13*	(1.70)	0.23***	(3.55)
4	0.89***	(3.19)	0.10	(0.67)	0.11	(1.57)	0.13*	(1.66)	0.24***	(3.22)
5	0.84***	(2.90)	0.03	(0.18)	0.04	(0.59)	0.09	(1.06)	0.20***	(2.80)
6	0.68**	(2.27)	-0.15	(-0.99)	-0.10	(-1.21)	-0.03	(-0.36)	0.08	(1.02)
7	0.65**	(1.99)	-0.20	(-1.24)	-0.15*	(-1.72)	-0.01	(-0.06)	0.12	(1.28)
8	0.51	(1.54)	-0.35**	(-2.03)	-0.26***	(-2.69)	-0.11	(-1.10)	0.01	(0.14)
9	0.22	(0.59)	-0.68***	(-3.36)	-0.55***	(-4.39)	-0.29***	(-2.84)	-0.21*	(-1.93)
10	-0.25	(-0.64)	-1.19***	(-5.12)	-1.00***	(-6.38)	-0.66***	(-5.13)	-0.62***	(-4.58)
H-L	-1.42***	(-5.68)	-1.61***	(-6.68)	-1.37***	(-7.08)	-0.91***	(-6.06)	-0.95***	(-6.48)
Panel B: Value-Weighted										
1	0.75***	(3.62)	0.06	(0.74)	0.03	(0.41)	-0.10	(-1.35)	-0.05	(-0.73)
2	0.85***	(4.30)	0.19**	(2.30)	0.19**	(2.42)	0.05	(0.59)	0.07	(0.91)
3	0.76***	(4.15)	0.09	(1.34)	0.12*	(1.79)	0.03	(0.44)	0.07	(0.92)
4	0.76***	(4.05)	0.10	(1.39)	0.11	(1.61)	0.02	(0.28)	0.01	(0.19)
5	0.66***	(2.99)	-0.03	(-0.42)	0.02	(0.21)	-0.01	(-0.11)	-0.00	(-0.02)
6	0.67***	(2.79)	-0.07	(-0.58)	-0.02	(-0.20)	-0.01	(-0.14)	-0.01	(-0.05)
7	0.71***	(3.12)	-0.01	(-0.14)	0.03	(0.32)	0.02	(0.27)	0.05	(0.55)
8	0.72***	(2.85)	-0.06	(-0.63)	0.02	(0.23)	0.12	(1.26)	0.14	(1.47)
9	0.65**	(2.33)	-0.18	(-1.38)	-0.06	(-0.58)	0.04	(0.42)	0.01	(0.09)
10	0.25	(0.74)	-0.66***	(-3.57)	-0.47***	(-2.99)	-0.19	(-1.56)	-0.28**	(-2.38)
H-L	-0.50**	(-2.15)	-0.72***	(-3.03)	-0.51**	(-2.51)	-0.09	(-0.59)	-0.22	(-1.52)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP as a robustness check using the WRDS Financial Ratio dataset. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.12: Univariate portfolio sorts on ATYP for the complete Jensen et al. (2023) dataset

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.95***	(4.19)	0.25**	(1.97)	0.21***	(2.92)	0.08	(1.15)	0.13**	(2.24)
2	0.89***	(3.70)	0.16	(1.27)	0.14**	(2.10)	0.04	(0.62)	0.10*	(1.69)
3	0.88***	(3.52)	0.12	(1.00)	0.11*	(1.80)	0.06	(0.86)	0.13**	(2.21)
4	0.80***	(3.02)	0.01	(0.09)	0.02	(0.35)	0.01	(0.20)	0.09	(1.48)
5	0.84***	(2.93)	0.03	(0.21)	0.05	(0.65)	0.09	(1.16)	0.18**	(2.45)
6	0.77**	(2.47)	-0.06	(-0.42)	-0.03	(-0.41)	0.07	(0.76)	0.18**	(2.29)
7	0.67**	(2.10)	-0.19	(-1.18)	-0.12	(-1.39)	0.02	(0.18)	0.13	(1.53)
8	0.48	(1.36)	-0.41**	(-2.18)	-0.32***	(-2.73)	-0.11	(-1.02)	0.02	(0.19)
9	0.15	(0.39)	-0.77***	(-3.52)	-0.64***	(-4.50)	-0.31***	(-2.69)	-0.20*	(-1.68)
10	-0.47	(-1.04)	-1.47***	(-5.18)	-1.29***	(-6.41)	-0.86***	(-5.00)	-0.70***	(-4.05)
H-L	-1.42***	(-4.48)	-1.72***	(-5.94)	-1.50***	(-6.36)	-0.94***	(-5.27)	-0.83***	(-4.48)
Panel B: Value-Weighted										
1	0.80***	(4.29)	0.14	(1.52)	0.10	(1.16)	-0.07	(-0.89)	-0.03	(-0.41)
2	0.67***	(3.58)	0.03	(0.45)	-0.00	(-0.05)	-0.17***	(-2.67)	-0.14**	(-2.29)
3	0.76***	(4.20)	0.11*	(1.66)	0.11	(1.59)	-0.01	(-0.18)	-0.01	(-0.20)
4	0.70***	(3.42)	0.03	(0.49)	0.05	(0.87)	0.01	(0.19)	0.01	(0.16)
5	0.78***	(3.57)	0.08	(1.04)	0.12	(1.61)	0.12	(1.62)	0.13*	(1.65)
6	0.92***	(3.66)	0.12	(1.45)	0.18**	(2.18)	0.17**	(2.12)	0.22**	(2.56)
7	0.71***	(2.71)	-0.11	(-0.95)	-0.03	(-0.31)	0.08	(0.74)	0.11	(1.10)
8	0.72**	(2.53)	-0.12	(-0.87)	0.02	(0.19)	0.14	(1.13)	0.18	(1.45)
9	0.30	(0.97)	-0.58***	(-3.15)	-0.43***	(-2.87)	-0.17	(-1.24)	-0.15	(-1.14)
10	0.23	(0.58)	-0.79***	(-3.36)	-0.57***	(-2.71)	-0.17	(-0.88)	-0.17	(-0.90)
H-L	-0.57*	(-1.82)	-0.94***	(-3.27)	-0.67***	(-2.62)	-0.10	(-0.45)	-0.14	(-0.64)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP as a robustness check using the complete dataset of Jensen et al. (2023). At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.13: Univariate portfolio sorts on ATYP for different model specifications

								Equal-Weighted				Value-Weighted			
Model	Bottleneck	#Layers	Units	L2	Dropout	Noise	R^2	H–L	t-stat	FF6	t-stat	H–L	t-stat	FF6	t-stat
Autoencoder	12	2	64	0	0	0	0.66	-1.48	-4.58	-0.86	-4.78	-0.59	-1.79	-0.13	-0.67
Autoencoder	12	3	128	0	0	0	0.69	-1.46	-4.42	-0.86	-4.73	-0.66	-2.03	-0.21	-1
Autoencoder	16	2	64	0	0	0	0.71	-1.42	-4.38	-0.78	-4.25	-0.68	-2.1	-0.15	-0.78
Autoencoder	16	3	128	0	0	0	0.74	-1.46	-4.4	-0.84	-4.45	-0.63	-1.9	-0.15	-0.71
Autoencoder	20	2	64	0	0	0	0.74	-1.48	-4.52	-0.83	-4.58	-0.75	-2.32	-0.22	-1.18
Autoencoder	20	3	128	0	0	0	0.79	-1.53	-4.63	-0.9	-4.93	-0.54	-1.6	-0.08	-0.37
Autoencoder	12	2	64	0.1	0	0	0.66	-1.43	-4.46	-0.8	-4.5	-0.57	-1.61	-0.1	-0.47
Autoencoder	12	3	128	0.1	0	0	0.69	-1.45	-4.39	-0.83	-4.52	-0.61	-1.88	-0.15	-0.75
Autoencoder	16	2	64	0.1	0	0	0.7	-1.46	-4.48	-0.83	-4.48	-0.71	-2.16	-0.22	-1.14
Autoencoder	16	3	128	0.1	0	0	0.74	-1.45	-4.4	-0.81	-4.35	-0.66	-2.01	-0.17	-0.9
Autoencoder	20	2	64	0.1	0	0	0.74	-1.5	-4.57	-0.85	-4.74	-0.87	-2.69	-0.4	-2.03
Autoencoder	20	3	128	0.1	0	0	0.78	-1.48	-4.46	-0.82	-4.28	-0.68	-1.91	-0.2	-0.9
Autoencoder	12	2	64	0	1e-04	0	0.66	-1.42	-4.38	-0.8	-4.37	-0.54	-1.58	-0.07	-0.34
Autoencoder	12	3	128	0	1e-04	0	0.7	-1.45	-4.4	-0.85	-4.6	-0.7	-2.13	-0.22	-1.12
Autoencoder	16	2	64	0	1e-04	0	0.71	-1.43	-4.37	-0.81	-4.3	-0.69	-2.18	-0.27	-1.37
Autoencoder	16	3	128	0	1e-04	0	0.75	-1.48	-4.52	-0.85	-4.59	-0.62	-1.82	-0.15	-0.73
Autoencoder	20	2	64	0	1e-04	0	0.74	-1.47	-4.46	-0.82	-4.54	-0.67	-1.97	-0.15	-0.79
Autoencoder	20	3	128	0	1e-04	0	0.79	-1.5	-4.54	-0.85	-4.71	-0.65	-1.92	-0.18	-0.84
Autoencoder	12	2	64	0.1	1e-04	0	0.66	-1.45	-4.55	-0.84	-4.73	-0.68	-2.12	-0.16	-0.86
Autoencoder	12	3	128	0.1	1e-04	0	0.69	-1.49	-4.49	-0.88	-4.74	-0.67	-2.06	-0.2	-1.01
Autoencoder	16	2	64	0.1	1e-04	0	0.7	-1.47	-4.5	-0.84	-4.62	-0.64	-1.96	-0.12	-0.63
Autoencoder	16	3	128	0.1	1e-04	0	0.74	-1.48	-4.55	-0.85	-4.57	-0.8	-2.42	-0.3	-1.55
Autoencoder	20	2	64	0.1	1e-04	0	0.74	-1.43	-4.33	-0.79	-4.24	-0.67	-1.98	-0.21	-1.02
Autoencoder	20	3	128	0.1	1e-04	0	0.78	-1.46	-4.42	-0.81	-4.4	-0.6	-1.8	-0.09	-0.48
Autoencoder	12	2	64	0	0	0.1	0.66	-1.43	-4.43	-0.81	-4.46	-0.59	-1.76	-0.13	-0.65
Autoencoder	12	3	128	0	0	0.1	0.69	-1.42	-4.31	-0.82	-4.54	-0.66	-2	-0.19	-0.91
Autoencoder	16	2	64	0	0	0.1	0.71	-1.44	-4.38	-0.8	-4.33	-0.6	-1.79	-0.13	-0.66
Autoencoder	16	3	128	0	0	0.1	0.75	-1.46	-4.44	-0.84	-4.53	-0.67	-1.96	-0.17	-0.81
Autoencoder	20	2	64	0	0	0.1	0.74	-1.48	-4.49	-0.83	-4.58	-0.72	-2.08	-0.18	-0.85
Autoencoder	20	3	128	0	0	0.1	0.79	-1.49	-4.56	-0.84	-4.53	-0.67	-2.08	-0.15	-0.77

Continued on next page

Table C.13: Univariate Portfolio Sorts on ATYP for Different Model Specifications *(continued)*

								Equal-Weighted				Value-Weighted			
Model	Bottleneck	#Layers	Units	L2	Dropout	Noise	R^2	H–L	t-stat	FF6	t-stat	H–L	t-stat	FF6	t-stat
Autoencoder	12	2	64	0.1	0	0.1	0.66	-1.44	-4.51	-0.81	-4.69	-0.62	-1.9	-0.13	-0.7
Autoencoder	12	3	128	0.1	0	0.1	0.69	-1.45	-4.42	-0.84	-4.56	-0.58	-1.73	-0.06	-0.29
Autoencoder	16	2	64	0.1	0	0.1	0.7	-1.45	-4.43	-0.82	-4.46	-0.53	-1.56	-0.03	-0.17
Autoencoder	16	3	128	0.1	0	0.1	0.74	-1.48	-4.53	-0.84	-4.57	-0.76	-2.31	-0.31	-1.48
Autoencoder	20	2	64	0.1	0	0.1	0.74	-1.5	-4.54	-0.86	-4.68	-0.73	-2.16	-0.24	-1.2
Autoencoder	20	3	128	0.1	0	0.1	0.78	-1.47	-4.44	-0.84	-4.49	-0.74	-2.28	-0.26	-1.29
Autoencoder	12	2	64	0	1e-04	0.1	0.66	-1.43	-4.45	-0.81	-4.49	-0.54	-1.59	-0.01	-0.05
Autoencoder	12	3	128	0	1e-04	0.1	0.7	-1.46	-4.37	-0.85	-4.54	-0.59	-1.75	-0.1	-0.48
Autoencoder	16	2	64	0	1e-04	0.1	0.71	-1.47	-4.54	-0.85	-4.67	-0.56	-1.68	-0.1	-0.48
Autoencoder	16	3	128	0	1e-04	0.1	0.75	-1.49	-4.56	-0.85	-4.65	-0.63	-1.8	-0.15	-0.72
Autoencoder	20	2	64	0	1e-04	0.1	0.74	-1.45	-4.41	-0.81	-4.48	-0.71	-2.06	-0.21	-1.05
Autoencoder	20	3	128	0	1e-04	0.1	0.79	-1.49	-4.53	-0.84	-4.63	-0.64	-1.88	-0.17	-0.83
Autoencoder	12	2	64	0.1	1e-04	0.1	0.66	-1.45	-4.54	-0.83	-4.74	-0.56	-1.65	-0.06	-0.33
Autoencoder	12	3	128	0.1	1e-04	0.1	0.69	-1.45	-4.39	-0.85	-4.61	-0.64	-1.86	-0.14	-0.67
Autoencoder	16	2	64	0.1	1e-04	0.1	0.7	-1.44	-4.38	-0.81	-4.37	-0.53	-1.64	-0.08	-0.42
Autoencoder	16	3	128	0.1	1e-04	0.1	0.74	-1.52	-4.64	-0.89	-4.83	-0.82	-2.5	-0.29	-1.56
Autoencoder	20	2	64	0.1	1e-04	0.1	0.74	-1.47	-4.41	-0.83	-4.44	-0.82	-2.51	-0.34	-1.69
Autoencoder	20	3	128	0.1	1e-04	0.1	0.78	-1.52	-4.53	-0.85	-4.51	-0.7	-2.09	-0.18	-0.9

The table reports average monthly excess returns and factor-adjusted alphas from univariate portfolio sorts on ATYP constructed using different model specifications. For autoencoders, we vary the bottleneck dimension, number of layers, number of hidden units, regularization (L2), dropout, and noise parameters. For PCA, we vary the number of components (Bottleneck). R^2 indicates the overall fit of each specification. Reported returns include equal-weighted and value-weighted high-minus-low (H–L) decile spreads, along with corresponding Fama and French (2018) six-factor (FF6) alphas and Newey and West (1987) adjusted t-statistics with six lags. The sample is from 1981 to 2023.

Table C.14: Univariate portfolio sorts on $ATYP_{AE}$ orthogonalized with respect to $ATYP_{PCA}$

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.58*	(1.89)	-0.24*	(-1.65)	-0.18*	(-1.85)	-0.06	(-0.58)	0.07	(0.77)
2	0.65**	(2.29)	-0.14	(-1.04)	-0.12*	(-1.65)	-0.07	(-0.88)	0.03	(0.41)
3	0.77***	(2.82)	-0.01	(-0.07)	0.01	(0.08)	0.02	(0.30)	0.12*	(1.83)
4	0.78***	(2.90)	-0.00	(-0.03)	0.01	(0.21)	0.01	(0.09)	0.08	(1.37)
5	0.76***	(2.74)	-0.04	(-0.30)	-0.03	(-0.45)	-0.00	(-0.02)	0.08	(1.34)
6	0.74**	(2.58)	-0.07	(-0.48)	-0.05	(-0.70)	0.01	(0.11)	0.09	(1.32)
7	0.69**	(2.30)	-0.13	(-0.85)	-0.09	(-1.11)	0.01	(0.13)	0.10	(1.34)
8	0.54*	(1.71)	-0.32**	(-1.96)	-0.26***	(-3.12)	-0.13	(-1.43)	-0.03	(-0.35)
9	0.39	(1.16)	-0.49***	(-2.69)	-0.41***	(-4.12)	-0.24**	(-2.31)	-0.14	(-1.47)
10	0.04	(0.12)	-0.88***	(-4.05)	-0.76***	(-5.45)	-0.46***	(-3.75)	-0.35***	(-2.77)
H-L	-0.53***	(-4.55)	-0.63***	(-5.53)	-0.58***	(-5.89)	-0.41***	(-4.45)	-0.42***	(-4.45)
Panel B: Value-Weighted										
1	0.72***	(3.49)	0.04	(0.45)	0.11**	(2.13)	0.13**	(2.39)	0.14**	(2.53)
2	0.66***	(3.32)	0.00	(0.04)	0.01	(0.19)	-0.08	(-1.25)	-0.08	(-1.21)
3	0.71***	(3.36)	0.00	(0.08)	0.00	(0.04)	-0.12*	(-1.83)	-0.09	(-1.56)
4	0.72***	(3.48)	0.01	(0.11)	0.00	(0.02)	-0.09	(-1.59)	-0.09	(-1.52)
5	0.76***	(3.46)	0.01	(0.15)	0.02	(0.25)	-0.04	(-0.48)	-0.01	(-0.19)
6	0.68***	(3.01)	-0.10	(-1.23)	-0.09	(-1.25)	-0.10	(-1.48)	-0.07	(-1.18)
7	0.72***	(3.01)	-0.07	(-0.78)	-0.04	(-0.51)	-0.05	(-0.58)	-0.03	(-0.42)
8	0.61**	(2.47)	-0.19**	(-2.16)	-0.12*	(-1.79)	-0.05	(-0.75)	-0.02	(-0.35)
9	0.62**	(2.33)	-0.22**	(-2.14)	-0.14*	(-1.83)	-0.04	(-0.50)	-0.02	(-0.29)
10	0.49	(1.62)	-0.39***	(-2.61)	-0.28**	(-2.25)	-0.02	(-0.15)	-0.01	(-0.10)
H-L	-0.22	(-1.22)	-0.43***	(-2.63)	-0.38***	(-3.03)	-0.14	(-1.22)	-0.15	(-1.20)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on $ATYP_{AE}$ that has been orthogonalized with respect to $ATYP_{PCA}$. $ATYP_{PCA}$ is the ATYP measure derived by a linear PCA model with the same bottleneck size of 16. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.15: Univariate portfolio sorts on median ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.97***	(4.42)	0.29**	(2.41)	0.23***	(3.77)	0.07	(1.26)	0.11**	(2.33)
2	0.89***	(3.92)	0.19	(1.49)	0.15***	(2.61)	0.04	(0.74)	0.09*	(1.71)
3	0.97***	(3.91)	0.23*	(1.79)	0.22***	(3.52)	0.15**	(2.43)	0.22***	(3.61)
4	0.87***	(3.28)	0.11	(0.79)	0.11*	(1.80)	0.13*	(1.89)	0.20***	(3.06)
5	0.83***	(2.89)	0.02	(0.17)	0.05	(0.65)	0.09	(1.24)	0.19***	(2.59)
6	0.75**	(2.51)	-0.07	(-0.48)	-0.03	(-0.42)	0.05	(0.55)	0.14*	(1.76)
7	0.62*	(1.90)	-0.26	(-1.58)	-0.19**	(-2.19)	-0.02	(-0.25)	0.10	(1.11)
8	0.43	(1.19)	-0.49***	(-2.69)	-0.39***	(-3.50)	-0.17	(-1.60)	-0.03	(-0.34)
9	0.05	(0.13)	-0.89***	(-3.95)	-0.76***	(-5.03)	-0.43***	(-3.64)	-0.29**	(-2.43)
10	-0.43	(-0.93)	-1.44***	(-4.94)	-1.26***	(-6.06)	-0.82***	(-4.41)	-0.67***	(-3.51)
H-L	-1.39***	(-4.27)	-1.72***	(-5.85)	-1.49***	(-6.26)	-0.89***	(-4.56)	-0.78***	(-3.86)
Panel B: Value-Weighted										
1	0.73***	(4.09)	0.10	(1.28)	0.04	(0.65)	-0.15**	(-2.53)	-0.13**	(-2.15)
2	0.73***	(4.08)	0.11	(1.51)	0.08	(1.17)	-0.08	(-1.08)	-0.05	(-0.77)
3	0.79***	(4.02)	0.12*	(1.88)	0.13**	(2.00)	0.08	(1.11)	0.05	(0.80)
4	0.79***	(4.03)	0.12*	(1.84)	0.14**	(2.21)	0.07	(1.03)	0.07	(0.99)
5	0.74***	(3.12)	-0.00	(-0.02)	0.06	(0.69)	0.06	(0.76)	0.07	(0.76)
6	0.67***	(2.63)	-0.13	(-1.55)	-0.07	(-0.80)	-0.05	(-0.47)	0.01	(0.09)
7	0.68**	(2.58)	-0.13	(-1.20)	-0.02	(-0.25)	0.11	(1.12)	0.12	(1.15)
8	0.69**	(2.29)	-0.23*	(-1.85)	-0.14	(-1.11)	-0.02	(-0.15)	0.01	(0.06)
9	0.40	(1.21)	-0.56***	(-3.24)	-0.37***	(-2.62)	-0.09	(-0.65)	-0.03	(-0.25)
10	0.14	(0.33)	-0.92***	(-3.54)	-0.66***	(-3.20)	-0.15	(-0.83)	-0.16	(-0.91)
H-L	-0.58*	(-1.77)	-1.02***	(-3.29)	-0.70***	(-3.00)	-0.00	(-0.01)	-0.03	(-0.16)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the *median* ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.16: Univariate portfolio sorts on rank ATYP

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.94***	(4.44)	0.27**	(2.30)	0.21***	(3.26)	0.04	(0.72)	0.09*	(1.71)
2	0.94***	(4.10)	0.24*	(1.86)	0.20***	(3.34)	0.07	(1.30)	0.12**	(2.29)
3	1.00***	(4.05)	0.26**	(2.01)	0.24***	(4.04)	0.18***	(2.86)	0.24***	(4.18)
4	0.84***	(3.16)	0.08	(0.58)	0.08	(1.36)	0.09	(1.42)	0.17***	(2.76)
5	0.83***	(2.95)	0.04	(0.27)	0.06	(0.83)	0.11	(1.46)	0.20***	(2.66)
6	0.74**	(2.42)	-0.09	(-0.57)	-0.04	(-0.55)	0.05	(0.60)	0.16*	(1.79)
7	0.62*	(1.91)	-0.25	(-1.52)	-0.19**	(-2.11)	-0.05	(-0.56)	0.07	(0.77)
8	0.38	(1.06)	-0.54***	(-2.87)	-0.43***	(-3.76)	-0.17*	(-1.67)	-0.04	(-0.41)
9	0.08	(0.19)	-0.88***	(-3.95)	-0.74***	(-4.97)	-0.40***	(-3.34)	-0.25**	(-2.14)
10	-0.44	(-0.95)	-1.45***	(-4.96)	-1.27***	(-6.10)	-0.84***	(-4.64)	-0.69***	(-3.75)
H-L	-1.38***	(-4.15)	-1.72***	(-5.72)	-1.48***	(-6.16)	-0.88***	(-4.64)	-0.78***	(-3.94)
Panel B: Value-Weighted										
1	0.70***	(3.97)	0.07	(0.90)	0.01	(0.17)	-0.17***	(-2.66)	-0.13**	(-2.09)
2	0.71***	(3.97)	0.11	(1.41)	0.08	(1.07)	-0.10	(-1.60)	-0.09	(-1.52)
3	0.79***	(3.91)	0.11*	(1.78)	0.12**	(2.01)	0.04	(0.54)	0.02	(0.38)
4	0.75***	(3.82)	0.06	(0.88)	0.09	(1.41)	0.03	(0.51)	0.03	(0.40)
5	0.73***	(3.08)	-0.01	(-0.18)	0.04	(0.52)	0.06	(0.67)	0.08	(0.86)
6	0.71***	(2.78)	-0.06	(-0.62)	0.02	(0.20)	0.08	(0.74)	0.09	(0.85)
7	0.80***	(2.87)	-0.07	(-0.59)	0.02	(0.25)	0.16	(1.64)	0.17*	(1.78)
8	0.64**	(2.10)	-0.26**	(-2.09)	-0.14	(-1.39)	0.00	(0.02)	0.06	(0.63)
9	0.35	(1.05)	-0.62***	(-3.56)	-0.43***	(-2.91)	-0.11	(-0.76)	-0.07	(-0.49)
10	0.15	(0.36)	-0.91***	(-3.61)	-0.66***	(-3.28)	-0.17	(-1.03)	-0.18	(-1.16)
H-L	-0.55*	(-1.75)	-0.98***	(-3.35)	-0.67***	(-2.96)	-0.00	(-0.03)	-0.05	(-0.31)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on the *rank* ATYP. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.17: Univariate portfolio sorts on ATYP using non-missing characteristics

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.93***	(4.25)	0.24*	(1.93)	0.18***	(2.64)	0.02	(0.37)	0.08	(1.51)
2	0.96***	(4.11)	0.25*	(1.95)	0.21***	(3.61)	0.10*	(1.80)	0.15***	(2.95)
3	0.92***	(3.71)	0.17	(1.37)	0.16***	(2.71)	0.10*	(1.71)	0.16***	(2.91)
4	0.94***	(3.52)	0.17	(1.25)	0.17***	(2.77)	0.18***	(2.68)	0.25***	(3.94)
5	0.80***	(2.80)	0.00	(0.00)	0.02	(0.29)	0.06	(0.79)	0.16**	(2.09)
6	0.75**	(2.49)	-0.08	(-0.54)	-0.04	(-0.49)	0.06	(0.77)	0.18**	(2.17)
7	0.68**	(2.06)	-0.18	(-1.06)	-0.12	(-1.23)	0.01	(0.14)	0.12	(1.40)
8	0.42	(1.18)	-0.48**	(-2.50)	-0.38***	(-3.30)	-0.17	(-1.57)	-0.03	(-0.32)
9	0.13	(0.34)	-0.81***	(-3.66)	-0.68***	(-4.66)	-0.32***	(-2.65)	-0.19	(-1.59)
10	-0.60	(-1.35)	-1.60***	(-5.75)	-1.42***	(-7.15)	-0.96***	(-5.54)	-0.82***	(-4.69)
H-L	-1.54***	(-4.79)	-1.85***	(-6.21)	-1.60***	(-6.81)	-0.99***	(-5.56)	-0.90***	(-4.89)
Panel B: Value-Weighted										
1	0.79***	(4.38)	0.15	(1.57)	0.10	(1.23)	-0.08	(-1.06)	-0.04	(-0.52)
2	0.67***	(3.68)	0.03	(0.37)	0.01	(0.09)	-0.15**	(-2.30)	-0.14**	(-2.11)
3	0.73***	(3.74)	0.06	(1.03)	0.07	(1.20)	-0.03	(-0.58)	-0.03	(-0.47)
4	0.75***	(3.65)	0.07	(0.98)	0.11	(1.45)	0.06	(0.75)	0.06	(0.73)
5	0.75***	(3.19)	0.02	(0.28)	0.06	(0.79)	0.08	(1.04)	0.09	(1.25)
6	0.78***	(3.27)	0.00	(0.03)	0.08	(0.99)	0.14	(1.63)	0.17*	(1.91)
7	0.89***	(3.43)	0.09	(0.97)	0.17**	(1.96)	0.27***	(3.08)	0.29***	(3.17)
8	0.44	(1.61)	-0.40***	(-3.36)	-0.30***	(-2.79)	-0.17	(-1.58)	-0.14	(-1.39)
9	0.39	(1.18)	-0.53***	(-2.83)	-0.37**	(-2.39)	-0.05	(-0.37)	-0.03	(-0.20)
10	0.03	(0.06)	-1.03***	(-3.96)	-0.78***	(-3.64)	-0.30*	(-1.68)	-0.29	(-1.62)
H-L	-0.77**	(-2.31)	-1.18***	(-3.78)	-0.89***	(-3.49)	-0.23	(-1.18)	-0.25	(-1.32)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP recalculated using only non-missing firm characteristics in each observation of $\mathbf{X}i, t$. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.18: Univariate portfolio sorts on ATYP orthogonalized with respect to imputation intensity

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.72***	(2.83)	-0.03	(-0.19)	0.01	(0.08)	0.00	(0.04)	0.05	(0.63)
2	0.84***	(3.46)	0.11	(0.89)	0.09	(1.56)	0.02	(0.33)	0.08	(1.49)
3	0.85***	(3.43)	0.10	(0.85)	0.10*	(1.88)	0.05	(0.98)	0.11**	(2.37)
4	0.88***	(3.38)	0.11	(0.92)	0.11*	(1.89)	0.10	(1.51)	0.16***	(2.94)
5	0.88***	(3.14)	0.08	(0.55)	0.09	(1.25)	0.12	(1.45)	0.21***	(2.70)
6	0.69**	(2.35)	-0.12	(-0.83)	-0.09	(-1.20)	-0.03	(-0.33)	0.07	(0.88)
7	0.70**	(2.22)	-0.15	(-0.98)	-0.10	(-1.23)	0.02	(0.28)	0.14*	(1.79)
8	0.51	(1.45)	-0.38**	(-2.01)	-0.30**	(-2.56)	-0.11	(-1.03)	0.01	(0.09)
9	0.26	(0.67)	-0.67***	(-3.18)	-0.56***	(-4.14)	-0.27**	(-2.28)	-0.13	(-1.11)
10	-0.38	(-0.86)	-1.37***	(-5.12)	-1.22***	(-6.46)	-0.82***	(-4.85)	-0.66***	(-3.92)
H-L	-1.10***	(-4.24)	-1.35***	(-5.84)	-1.23***	(-5.95)	-0.83***	(-4.80)	-0.70***	(-4.06)
Panel B: Value-Weighted										
1	0.71***	(3.22)	-0.04	(-0.43)	-0.03	(-0.41)	-0.14*	(-1.81)	-0.12*	(-1.66)
2	0.78***	(4.19)	0.12	(1.40)	0.11	(1.29)	-0.04	(-0.57)	-0.02	(-0.32)
3	0.63***	(3.23)	-0.03	(-0.39)	-0.04	(-0.46)	-0.17**	(-2.34)	-0.15**	(-2.20)
4	0.73***	(3.84)	0.07	(1.34)	0.08	(1.50)	-0.02	(-0.35)	-0.02	(-0.33)
5	0.74***	(3.55)	0.05	(0.83)	0.08	(1.26)	0.01	(0.13)	0.02	(0.21)
6	0.76***	(3.24)	0.02	(0.27)	0.04	(0.55)	0.07	(0.87)	0.08	(1.01)
7	0.81***	(3.21)	0.03	(0.35)	0.12	(1.37)	0.21**	(2.41)	0.25***	(2.63)
8	0.76***	(2.88)	-0.04	(-0.43)	0.05	(0.58)	0.16*	(1.91)	0.19**	(2.26)
9	0.44	(1.51)	-0.43***	(-3.06)	-0.32***	(-2.60)	-0.12	(-1.08)	-0.10	(-0.97)
10	0.31	(0.81)	-0.69***	(-3.02)	-0.50**	(-2.54)	-0.08	(-0.44)	-0.05	(-0.28)
H-L	-0.40	(-1.49)	-0.65**	(-2.54)	-0.47**	(-2.06)	0.07	(0.36)	0.07	(0.42)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP that has been orthogonalized with respect to the share of imputed characteristics. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.19: Univariate portfolio sorts on ATYP orthogonalized with respect to extreme observations

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.75**	(2.44)	-0.09	(-0.54)	-0.03	(-0.29)	-0.01	(-0.15)	0.08	(0.81)
2	0.66**	(2.38)	-0.12	(-0.85)	-0.10	(-1.42)	-0.10	(-1.30)	-0.02	(-0.21)
3	0.77***	(2.84)	-0.01	(-0.09)	0.00	(0.02)	-0.01	(-0.10)	0.08	(1.17)
4	0.80***	(2.95)	0.01	(0.08)	0.02	(0.34)	0.03	(0.42)	0.12*	(1.95)
5	0.75***	(2.66)	-0.05	(-0.32)	-0.04	(-0.49)	-0.00	(-0.00)	0.09	(1.12)
6	0.76***	(2.69)	-0.03	(-0.22)	-0.01	(-0.18)	0.07	(0.91)	0.15**	(2.19)
7	0.70**	(2.35)	-0.13	(-0.88)	-0.09	(-1.32)	-0.01	(-0.13)	0.08	(1.19)
8	0.51	(1.62)	-0.33**	(-2.08)	-0.27***	(-3.15)	-0.12	(-1.21)	-0.01	(-0.10)
9	0.30	(0.87)	-0.57***	(-3.15)	-0.50***	(-4.58)	-0.26**	(-2.50)	-0.14	(-1.42)
10	-0.06	(-0.16)	-1.00***	(-4.38)	-0.86***	(-5.52)	-0.50***	(-3.55)	-0.39***	(-2.85)
H-L	-0.81***	(-5.03)	-0.91***	(-5.28)	-0.83***	(-5.83)	-0.49***	(-3.69)	-0.47***	(-3.67)
Panel B: Value-Weighted										
1	0.87***	(3.91)	0.17*	(1.81)	0.28***	(3.85)	0.21***	(2.60)	0.17**	(2.37)
2	0.70***	(3.75)	0.05	(0.62)	0.06	(0.80)	-0.03	(-0.35)	-0.02	(-0.25)
3	0.74***	(3.70)	0.05	(0.74)	0.04	(0.66)	-0.05	(-0.88)	-0.03	(-0.46)
4	0.76***	(3.64)	0.06	(0.78)	0.06	(0.73)	0.03	(0.32)	0.05	(0.66)
5	0.65***	(2.99)	-0.09	(-1.03)	-0.09	(-1.18)	-0.13	(-1.56)	-0.11	(-1.41)
6	0.61***	(2.69)	-0.13	(-1.62)	-0.11	(-1.45)	-0.08	(-1.14)	-0.05	(-0.64)
7	0.57**	(2.37)	-0.20**	(-2.43)	-0.19**	(-2.49)	-0.19**	(-2.57)	-0.18**	(-2.28)
8	0.58**	(2.14)	-0.22*	(-1.91)	-0.16	(-1.52)	-0.05	(-0.51)	-0.01	(-0.10)
9	0.41	(1.49)	-0.41***	(-3.20)	-0.32***	(-2.89)	-0.10	(-1.02)	-0.05	(-0.54)
10	0.27	(0.82)	-0.64***	(-3.40)	-0.49***	(-3.08)	-0.13	(-1.02)	-0.10	(-0.79)
H-L	-0.60***	(-2.97)	-0.81***	(-4.14)	-0.77***	(-4.18)	-0.33**	(-2.25)	-0.27*	(-1.87)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP that has been orthogonalized with respect to the share of extreme observations. Extreme observations are those that fall within the bottom or top decile of each characteristic in the cross-section. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample is from 1981 to 2023.

Table C.20: Univariate portfolio sorts on ATYP in developed countries (excl. USA)

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.62**	(2.34)	0.34**	(2.11)	0.27**	(1.99)	0.06	(0.45)	0.07	(0.49)
2	0.72**	(2.50)	0.38**	(2.51)	0.30***	(2.60)	0.09	(0.82)	0.10	(0.88)
3	0.69**	(2.26)	0.31**	(2.20)	0.23**	(2.28)	0.02	(0.17)	0.02	(0.25)
4	0.70**	(2.16)	0.30**	(2.15)	0.21**	(2.24)	0.03	(0.32)	0.03	(0.38)
5	0.75**	(2.20)	0.32**	(2.29)	0.23**	(2.48)	0.06	(0.62)	0.06	(0.68)
6	0.69*	(1.93)	0.24*	(1.69)	0.15*	(1.66)	-0.01	(-0.08)	0.02	(0.18)
7	0.69*	(1.76)	0.20	(1.32)	0.12	(1.24)	0.00	(0.05)	0.02	(0.27)
8	0.56	(1.36)	0.04	(0.25)	-0.03	(-0.28)	-0.10	(-1.17)	-0.08	(-0.97)
9	0.46	(1.03)	-0.10	(-0.60)	-0.15	(-1.34)	-0.13	(-1.20)	-0.11	(-1.01)
10	-0.03	(-0.05)	-0.64***	(-3.21)	-0.61***	(-4.27)	-0.37***	(-2.83)	-0.37***	(-2.77)
H-L	-0.65*	(-1.96)	-0.99***	(-4.72)	-0.88***	(-4.33)	-0.43**	(-2.12)	-0.44**	(-2.11)
Panel B: Value-Weighted										
1	0.43	(1.55)	0.08	(0.63)	0.00	(0.01)	-0.22**	(-2.07)	-0.21**	(-1.98)
2	0.51*	(1.68)	0.11	(1.01)	0.04	(0.38)	-0.18*	(-1.87)	-0.19**	(-1.98)
3	0.51*	(1.69)	0.07	(0.70)	0.02	(0.16)	-0.22**	(-2.21)	-0.25**	(-2.50)
4	0.47	(1.49)	0.02	(0.16)	-0.03	(-0.31)	-0.21**	(-2.30)	-0.23***	(-2.60)
5	0.39	(1.23)	-0.08	(-0.79)	-0.11	(-1.23)	-0.23**	(-2.50)	-0.25***	(-2.70)
6	0.41	(1.29)	-0.05	(-0.52)	-0.07	(-0.71)	-0.22**	(-2.19)	-0.23**	(-2.17)
7	0.47	(1.36)	-0.02	(-0.20)	-0.03	(-0.32)	-0.10	(-0.86)	-0.12	(-1.01)
8	0.24	(0.64)	-0.26**	(-2.25)	-0.24**	(-2.19)	-0.24**	(-2.01)	-0.27**	(-2.33)
9	0.44	(1.12)	-0.12	(-0.91)	-0.06	(-0.54)	-0.02	(-0.20)	-0.03	(-0.24)
10	-0.11	(-0.23)	-0.69***	(-3.09)	-0.60***	(-3.09)	-0.44**	(-2.30)	-0.44**	(-2.36)
H-L	-0.53*	(-1.95)	-0.77***	(-3.34)	-0.61***	(-2.80)	-0.22	(-0.99)	-0.24	(-1.06)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP using stocks from developed countries (excl. USA). At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the local one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, the Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). The regional factor portfolios and country classifications for developed markets are taken from the Kenneth R. French Data Library. Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample period is from 2000 to 2023.

Table C.21: Univariate portfolio sorts on ATYP in Europe

	Excess Return	t-stat	CAPM	t-stat	FF3	t-stat	FF5	t-stat	FF6	t-stat
Panel A: Equal-Weighted										
1	0.88**	(2.26)	0.47***	(3.83)	0.35***	(5.23)	0.21***	(2.97)	0.22***	(2.88)
2	0.77**	(1.98)	0.36***	(3.29)	0.23***	(4.49)	0.15***	(2.79)	0.16***	(2.67)
3	0.80**	(2.01)	0.39***	(3.44)	0.26***	(5.11)	0.21***	(4.26)	0.24***	(4.57)
4	0.76*	(1.88)	0.34***	(3.04)	0.21***	(3.75)	0.21***	(3.46)	0.24***	(3.96)
5	0.63	(1.51)	0.20*	(1.72)	0.07	(1.41)	0.15**	(2.54)	0.20***	(3.61)
6	0.58	(1.34)	0.15	(1.18)	0.02	(0.28)	0.11	(1.52)	0.17**	(2.38)
7	0.45	(1.00)	0.01	(0.04)	-0.11	(-1.54)	0.06	(0.85)	0.12*	(1.80)
8	0.33	(0.71)	-0.12	(-0.76)	-0.23**	(-2.50)	-0.01	(-0.09)	0.06	(0.68)
9	0.11	(0.24)	-0.35**	(-1.98)	-0.43***	(-3.84)	-0.08	(-0.69)	-0.02	(-0.17)
10	-0.33	(-0.61)	-0.81***	(-3.31)	-0.86***	(-5.44)	-0.36**	(-2.53)	-0.34**	(-2.28)
H-L	-1.21***	(-4.40)	-1.28***	(-4.94)	-1.21***	(-6.38)	-0.57***	(-3.49)	-0.55***	(-3.22)
Panel B: Value-Weighted										
1	0.75**	(2.17)	0.36***	(3.32)	0.34***	(3.58)	0.12	(1.07)	0.10	(0.92)
2	0.59*	(1.77)	0.20**	(2.54)	0.19***	(2.69)	-0.01	(-0.13)	-0.04	(-0.47)
3	0.56	(1.64)	0.15*	(1.73)	0.15*	(1.77)	-0.06	(-0.67)	-0.08	(-0.96)
4	0.56	(1.62)	0.14	(1.43)	0.19*	(1.82)	0.02	(0.19)	0.01	(0.08)
5	0.59*	(1.66)	0.17*	(1.85)	0.16*	(1.74)	0.07	(0.78)	0.08	(0.86)
6	0.60	(1.64)	0.20**	(2.09)	0.20**	(1.98)	0.14	(1.20)	0.14	(1.14)
7	0.48	(1.40)	0.06	(0.64)	0.08	(0.82)	0.12	(1.02)	0.09	(0.72)
8	0.56	(1.49)	0.13	(1.15)	0.18*	(1.85)	0.19	(1.54)	0.21	(1.50)
9	0.19	(0.44)	-0.29**	(-2.05)	-0.24*	(-1.70)	-0.15	(-0.94)	-0.18	(-1.07)
10	0.04	(0.10)	-0.43**	(-2.49)	-0.40**	(-2.54)	-0.29*	(-1.85)	-0.29*	(-1.82)
H-L	-0.71***	(-3.72)	-0.78***	(-4.23)	-0.74***	(-4.31)	-0.42**	(-2.27)	-0.40**	(-2.11)

The table presents average monthly excess returns and factor-adjusted alphas for univariate portfolio sorts based on ATYP using stocks from European countries. At the end of each month t, all stocks are sorted into deciles according to their ATYP score measured using information available up to month $t-1$. Panel A reports results for equal-weighted portfolios, while Panel B reports value-weighted portfolios. Excess Return is defined relative to the local one-month Treasury bill rate. The remaining columns report alphas from time-series regressions of portfolio excess returns on the CAPM, the Fama and French (1993) three-factor model (FF3), the Fama and French (2015) five-factor model (FF5), and the Fama and French (2018) six-factor model including momentum (FF6). The regional factor portfolios and country classifications for European markets are taken from the Kenneth R. French Data Library. Reported t-statistics are Newey and West (1987) adjusted with six lags. Statistical significance at the 10%, 5%, and 1% levels is denoted by *, **, and ***, respectively. The sample period is from 2000 to 2023.

Bibliography

Albrecht, W. S., L. L. Lookabill, and J. C. McKeown (1977). The time-series properties of annual earnings. *Journal of Accounting Research 15*(2), 226–244.

Ali, A., L.-S. Hwang, and M. A. Trombley (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics 69*(2), 355–373.

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

Ammann, M., G. Coqueret, and J.-P. Schade (2016). Characteristics-based portfolio choice with leverage constraints. *Journal of Banking & Finance 70*, 23–37.

Ang, A., S. Gorovyy, and G. B. van Inwegen (2011). Hedge fund leverage. *Journal of Financial Economics 102*(1), 102–126.

Asness, C. (2011). Momentum in japan: The exception that proves the rule. *The Journal of Portfolio Management 37*(4), 67–75.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance 61*(4), 1645–1680.

Bali, T. G., H. Beckmeyer, M. Moerke, and F. Weigert (2023). Option return predictability with machine learning and big data. *The Review of Financial Studies 36*(9), 3548–3602.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Bali, T. G., B. T. Kelly, M. Mörke, and J. Rahman (2025). Machine forecast disagreement. *Working Paper*.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*(2), 159–178.

Ball, R. and R. Watts (1972). Some time series properties of accounting income. *The Journal of Finance 27*(3), 663–681.

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

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

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

Barinov, A. (2020). Firm complexity and conglomerates expected returns. *Working Paper*.

Barinov, A., S. Park, and Yıldızhan (2024). Firm complexity and post-earnings announcement drift. *Review of Accounting Studies 29*, 527–579.

Bernard, V. L. and J. K. Thomas (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research 27*, 1–36.

Bianchi, D., M. Büchner, and A. Tamoni (2020). Bond risk premiums with machine learning. *The Review of Financial Studies 34*(2), 1046–1089.

Boehmer, E. and J. J. Wu (2012). Short selling and the price discovery process. *The Review of Financial Studies 26*(2), 287–322.

Brandt, M. W. (2010). Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques*, pp. 269–336. Elsevier.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies 22*(9), 3411–3447.

Brogaard, J. and A. Detzel (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science 61*(1), 3–18.

Brown, L. D. (1993). Earnings forecasting research: its implications for capital markets research. *International Journal of Forecasting 9*(3), 295–320.

Bryzgalova, S., M. Pelger, and J. Zhu (2023). Forest through the trees: Building cross-sections of stock returns. *Working Paper*.

Campbell, J., H. Ham, Z. G. Lu, and K. Wood (2023). Expectations matter: When (not) to use machine learning earnings forecasts. *Working Paper*.

Cao, K. and H. You (2021). Fundamental analysis via machine learning. *Working Paper*.

Chen, A. Y. and J. McCoy (2024). Missing values handling for machine learning portfolios. *Journal of Financial Economics 155*, 103815.

Chen, A. Y. and M. Velikov (2023). Zeroing in on the expected returns of anomalies. *Journal of Financial and Quantitative Analysis 58*(3), 968–1004.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review 27*(2), 207–264.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Chen, X., Y. H. T. Cho, Y. Dou, and B. Lev (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research 60*(2), 467–515.

Chevalier, G., G. Coqueret, and T. Raffinot (2022). Supervised portfolios. *Quantitative Finance 22*(12), 2275–2295.

Claus, J. and J. Thomas (2001). Equity premia as low as three percent? evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance 56*(5), 1629–1666.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Finance 63*(4), 1977–2011.

Cohen, L. and D. Lou (2012). Complicated firms. *Journal of Financial Economics 104*(2), 383–400.

Cong, L., K. Tang, J. Wang, and Y. Zhan (2021). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Working Paper*.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

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

Coulombe, P. G. and M. Goebel (2024). Maximally machine-learnable portfolios. *Working Paper*.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance 53*(6), 1839–1885.

Daniel, K., D. Hirshleifer, and L. Sun (2019). Short- and long-horizon behavioral factors. *The Review of Financial Studies 33*(4), 1673–1736.

Daniel, K., L. Mota, S. Rottke, and T. Santos (2020). The cross-section of risk and returns. *The Review of Financial Studies 33*(5), 1927–1979.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance 56*(3), 921–965.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies 22*(5), 1915–1953.

DeMiguel, V., A. Martin-Utrera, and R. Uppal (2024). A multifactor perspective on volatility-managed portfolios. *Journal of Finance 79*(6), 3859–3891.

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies 33*(5), 2180–2222.

Detzel, A., R. Novy-Marx, and M. Velikov (2023). Model comparison with transaction costs. *Journal of Finance 78*(3), 1743–1775.

Dickinson, V. (2011). Cash flow patterns as a proxy for firm life cycle. *The Accounting Review 86*(6), 1969–1994.

Didisheim, A., S. B. Ke, B. T. Kelly, and S. Malamud (2023). Complexity in factor pricing models. *Working Paper*.

Diether, K. B., C. J. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance 57*(5), 2113–2141.

Easton, P. D. (2004). Pe ratios, peg ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review 79*(1), 73–95.

Easton, P. D., M. M. Kapons, S. J. Monahan, H. H. Schütt, and E. H. Weisbrod (2024). Forecasting earnings using k-nearest neighbors. *The Accounting Review 99*(3), 115–140.

Elgers, P. T., M. H. Lo, and R. J. Pfeiffer (2001). Delayed security price adjustments to financial analysts' forecasts of annual earnings. *The Accounting Review 76*(4), 613–632.

Engelberg, J., R. D. McLean, and J. Pontiff (2018). Anomalies and news. *The Journal of Finance 73*(5), 1971–2001.

Engelberg, J. E. and C. A. Parsons (2011). The causal impact of media in financial markets. *The Journal of Finance 66*(1), 67–97.

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

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

Fama, E. F. and K. R. French (2008). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

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

Fama, E. F. and K. R. French (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics 123*(3), 441–463.

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

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

Feng, G., L. Jiang, J. Li, and Y. Song (2024). Deep tangency portfolio. *Working Paper*.

Foster, G., C. Olsen, and T. Shevlin (1984). Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review 59*(4), 574–603.

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

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics 129*(4), 1661–1710.

García, D. and Norli (2012). Geographic dispersion and stock returns. *Journal of Financial Economics 106*(3), 547–565.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001). Toward an implied cost of capital. *Journal of Accounting Research 39*(1), 135–176.

Gerakos, J. J. and R. B. Gramacy (2012). Regression-based earnings forecasts. *Working Paper*.

Gordon, J. R. and M. J. Gordon (1997). The finite horizon expected return model. *Financial Analysts Journal 53*(3), 52–61.

Graham, J. R., C. R. Harvey, and S. Rajgopal (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics 40*(1), 3–73.

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

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

Guijarro-Ordonez, J., M. Pelger, and G. Zanotti (2022). Deep learning statistical arbitrage. *Working Paper*.

Hansen, J. W. and C. Thimsen (2020). Forecasting corporate earnings with machine learning. *Working Paper*.

Harris, R. D. F. and P. Wang (2019). Model-based earnings forecasts vs. financial analysts' earnings forecasts. *British Accounting Review 51*(4), 424–437.

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

Hautsch, N. and S. Voigt (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics 212*(1), 221–240.

Heaton, J. B., N. G. Polson, and J. H. Witte (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry 33*(1), 3–12.

Hendriock, M. (2022). Forecasting earnings with predicted, conditional probability distribution density functions. *Working Paper*.

Hess, D., F. Simon, and S. Weibels (2025). Interpretable machine learning for earnings forecasts: Leveraging high-dimensional financial statement data. *Working Paper*.

Hinton, G. E. and R. R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science 313*(5786), 504–507.

Hirshleifer, D., S. S. Lim, and S. H. Teoh (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance 64*(5), 2289–2325.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1), 337–386.

Hjalmarsson, E. and P. Manchev (2012). Characteristic-based mean-variance portfolio choice. *Journal of Banking & Finance 36*(5), 1392–1401.

Hoitash, R. and U. Hoitash (2018). Measuring accounting reporting complexity with xbrl. *The Accounting Review 93*(1), 259–287.

Hong, H., T. Lim, and J. C. Stein (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance 55*(1), 265–295.

Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance 54*(6), 2143–2184.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *The Review of Financial Studies 20*(4), 1113–1138.

Hou, K., M. A. van Dijk, and Y. Zhang (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics 53*(3), 504–526.

Hou, K., C. Xue, and L. Zhang (2014). Digesting anomalies: An investment approach. *The Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2018). Replicating anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Ioffe, S. and C. Szegedy (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning 37*, 448–456.

Israel, R., B. Kelly, and T. Moskowitz (2020). Can machines learn finance? *Journal of Investment Management 18*(2).

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1683.

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

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

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen (2022). Machine learning and the implementable efficient frontier. *Working Paper*.

Jiang, G., C. M. C. Lee, and Y. Zhang (2005). Information uncertainty and expected returns. *Review of Accounting Studies 10*, 185–221.

Johnson, T. C. (2004). Forecast dispersion and the cross section of expected returns. *The Journal of Finance 59*(5), 1957–1978.

Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of Accounting Research 29*(2), 193–228.

Jones, S., W. J. Moser, and M. M. Wieland (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research 40*(4), 2643–2672.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative analysis 21*(3), 279–292.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Kelly, B., S. Malamud, and K. Zhou (2024). The virtue of complexity in return prediction. *The Journal of Finance 79*(1), 459–503.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *Working Paper*.

Kirby, C. and B. Ostdiek (2012a). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis 47*(2), 437–467.

Kirby, C. and B. Ostdiek (2012b). Optimizing the performance of sample mean-variance efficient portfolios. *Working Paper*.

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

Lassance, N., A. Martín-Utrera, and M. Simaan (2024). The risk of expected utility under parameter uncertainty. *Management Science 70*(11), 7644–7663.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

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

Li, B. and A. G. Rossi (2021). Selecting mutual funds from the stocks they hold: A machine learning approach. *Working Paper*.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics 45*(2), 221–247.

Li, K. K. and P. Mohanram (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies 13*(3), 1152–1185.

Liaw, R., E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica (2018). Tune: A research platform for distributed model selection and training. *Working Paper*.

Liu, Y., G. Zhou, and Y. Zhu (2024). Maximizing the sharpe ratio: A genetic programming approach. *Working Paper*.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *The Journal of Finance 69*(4), 1643–1671.

Loughran, T. and B. McDonald (2024). Measuring firm complexity. *Journal of Financial and Quantitative Analysis 59*(6), 2487–2514.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 4765–4774.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. Academic Press Professional, Inc.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics 8*(4), 323–361.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance 32*(4), 1151–1168.

Monahan, S. J. (2018). Financial statement analysis and earnings forecasting. *Foundation and Trends in Accounting 12*, 105–215.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Working Paper*.

Murata, N., S. Yoshizawa, and S. Amari (1994). Network information criterion-determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks 5*(6), 865–872.

Nagel, S. (2025). Seemingly virtuous complexity in return prediction. *Working Paper*.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3).

Ohlson, J. A. and B. E. Juettner-Nauroth (2005). Expected eps and eps growth as determinantsof value. *Review of Accounting Studies 10*, 349–365.

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563–602.

Peterson, K. (2012). Accounting complexity, misreporting, and the consequences of misreporting. *Review of Accounting Studies 17*, 72–95.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics 42*(1), 35–52.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Pástor, L. and P. Veronesi (2003). Stock valuation and learning about profitability. *The Journal of Finance 58*(5), 1749–1789.

Schipper, K. (1991). Analysts' forecasts. *Accounting Horizons 5*(4), 105–121.

Simon, F., S. Weibels, and T. Zimmermann (2025). Deep parametric portfolio policies. *Working Paper*.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics 69*(1), 99–118.

Skouras, S. (2007). Decisionmetrics: A decision-based approach to econometric modelling. *Journal of Econometrics 137*, 414–440.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*(56), 1929–1958.

Stambaugh, R. F., J. Yu, and Y. Yuan (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance 70*(5), 1903–1948.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance 64*(4), 1517–1548.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man vs. machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies 36*(6), 2361–2396.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and underdiversification. *The Review of Economic Studies 77*(2), 779–805.

Watts, R. L. and R. W. Leftwich (1977). The time series of annual accounting earnings. *Journal of Accounting Research 15*(2), 253–271.

Weibels, S. (2025). Hard to process: Atypical firms and the cross-section of expected stock returns. *Working Paper*.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance 61*(1), 105–137.

Lebenslauf

Name:	Sebastian Weibels
Adresse:	Melchiorstraße 23, 50670 Köln
Geburtstag:	09.09.1994
Geburtsort:	Oberhausen, Deutschland

08/2005 – 07/2013:	Josef-Albers-Gymnasium, Bottrop Abschluss: Abitur
10/2014 – 01/2018:	Studium der Betriebswirtschaftslehre, Heinrich-Heine-Universität Düsseldorf Abschluss: Bachelor of Science
10/2018 – 07/2020:	Studium der Betriebswirtschaftslehre, Universität zu Köln Abschluss: Master of Science
09/2020 – 07/2023:	Wissenschaftlicher Mitarbeiter am Seminar für allgemeine BWL und Bankbetriebslehre, Universität zu Köln
08/2023 – aktuell:	Wissenschaftlicher Mitarbeiter am Institut für Ökonometrie und Statistik, Universität zu Köln Thema der Dissertation: *Three Essays on Machine Learning in Empirical Finance* Abschluss: Promotion

Ort, Datum

Unterschrift

Eidesstattliche Erklärung

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich geholfen:

Prof. Dr. Dieter Hess, Prof. Dr. Tom Zimmermann und Dr. Frederik Simon (gemeinsame Projektarbeit für die Inhalte in Kapitel 2 und Kapitel 3).

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Ort, Datum

Unterschrift