

University of Massachusetts Boston

**ScholarWorks at UMass Boston**

---

Graduate Doctoral Dissertations

Doctoral Dissertations and Masters Theses

---

8-2021

## **Return Predictability and Market Sentiment: Evidence from Deep Learning**

Usama A. Khan

Follow this and additional works at: [https://scholarworks.umb.edu/doctoral\\_dissertations](https://scholarworks.umb.edu/doctoral_dissertations)



Part of the [Finance and Financial Management Commons](#)

---

RETURN PREDICTABILITY AND MARKET SENTIMENT: EVIDENCE FROM  
DEEP LEARNING

A Dissertation Presented  
by  
USAMA A. KHAN

Submitted to the Office of Graduate Studies,  
University of Massachusetts Boston,  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2021

Business Administration Program

© 2021 by Usama A. Khan  
All rights reserved.

RETURN PREDICTABILITY AND MARKET SENTIMENT: EVIDENCE FROM  
DEEP LEARNING

A Dissertation Presented  
by  
USAMA A. KHAN

Approved as to style and content by:

---

Josephine M. Namayanja, Assistant Professor  
Co-Chairperson of Committee

---

Sangwan Kim, Associate Professor  
Co-Chairperson of Committee

---

H. Zafer Yuksel, Assistant Professor  
Texas A&M University  
Member

---

Atreya Chakraborty, Professor  
Member

---

Lucia Silva-Gao, Program Director  
Accounting and Finance Program

---

Kiran Verma, Chair  
Accounting and Finance Department

## ABSTRACT

### RETURN PREDICTABILITY AND MARKET SENTIMENT: EVIDENCE FROM DEEP LEARNING

May 2021

Usama A. Khan, B.B.A., Bahauddin Zakariya University  
M.S., Sabanci University  
Ph.D., University of Massachusetts Boston

Directed by Professors Josephine Namayanja & Sangwan Kim

Recent studies in asset pricing find that Artificial Neural Networks (also known as Deep Learning models) provide the most accurate firm-level return predictions using a vast set of predictive signals. These models offer high predictive accuracy over long out-of-sample periods, translating into highly profitable trading strategies. In this thesis, I argue that sentiment-driven mispricing is a vital source of the high predictability and the resulting profitability implied by deep learning models. Using a novel Artificial Neural Network (ANN) regression model, I obtain firm-level predictions conditional on 54 firm-level characteristics and on an encoded representation of the macro-economic state. These predictions provide important insights into the sources of overall cross-sectional return predictability. First, the future negative returns are predictable out-of-sample which implies negative expected returns. Such predictability is hard to reconcile with a risk-based explanation. Secondly, the predictability in negative returns is higher following periods of high sentiment and vice versa. This evidence is consistent with the existence of a market-level investor sentiment that drives misvaluations. Third, a long-short strategy based on ANN prediction deciles is more profitable following periods of high sentiment. This disparity in profitability points to arbitrage asymmetry implied by short-sale constraints. Fourth, the predictability in losses and high profitability of the ANN top decile vanishes in estimation horizons longer

than a month. This suggests that mispricing is short-lived and that predictability is realized due to corrections to such misvaluations. These corrections are preceded by high put-to-call(PCR) trading volumes and high implied volatility(VIX). Finally, the short-term and long-term predictions load on different conditioning variables indicating varying sources of predictability across return horizons. Overall, these findings are consistent with the existence of sentiment-driven short-lived mispricing that corrects in longer horizons.

# DEDICATION

*"To my mother Farah Khan. I hope she is watching me from the heavens,  
supporting me with pride and joy like she always did."*

# ACKNOWLEDGEMENTS

This doctoral thesis was written during my time as a doctoral candidate at University of Massachusetts Boston, College of Management. I will always cherish this period as an extraordinary one, encompassing significant personal and professional growth.

Firstly, I would like to thank my supervisor Dr. Josephine Namayanja, for her kind support and guidance. She has been a great help with my understanding of the topic and with the progress of my research. I would especially like to appreciate her patience with my half baked ideas and encouragement during hard times. While granting me the academic freedom to pursue my research interests, Josephine made sure that I stay on track. I would also like to thank my co-supervisor Dr. Sangwan Kim, who has been an extremely supportive colleague throughout my PhD journey. Likewise, I owe a debt of gratitude to my committee member Dr. Zafer Yuksel. Zafer has been a great teacher and an even better friend. It would be fair to say that my humble knowledge of asset pricing is owed to hours and hours of discussions with Zafer.

I am indebted to Dr. Atreya Chakraborty, the Director of PhD programs at UMB. Chuck has been nothing less of a father figure, who went out of his way to make sure that my progress was on track. I would especially like to thank Chuck for his kind support when I suffered a great personal loss during the program. I wish to thank Dr. Lucia Silva-Gao for her keen interest in my progress and her encouragement throughout my time at UMB.

Last but not least, I deeply thank my friends at the Computer Science Department and Mathematics Department at Boston University and University of Missouri. This interdisciplinary endeavor would not be done without their help. The growth, the learning, the memories, the ups and downs, and the kindness of those around me, will be an everlasting source of joy.

Usama A. Khan



# TABLE OF CONTENTS

DEDICATION . . . . .	vi
ACKNOWLEDGEMENTS . . . . .	vii
CHAPTER	Page
1. INTRODUCTION . . . . .	1
2. CROSS-SECTIONAL RETURN PREDICTABILITY: A CASE FOR DEEP LEARNING . . . . .	5
2.1 The Story of Risk: A History of Anomalies . . . . .	6
2.2 Modelling Expected Returns . . . . .	12
2.2.1 Factor Models . . . . .	14
2.2.2 Regression Models . . . . .	20
2.3 Conclusion . . . . .	21
3. DEEP RETURN PREDICTABILITY . . . . .	24
3.1 Artificial Neural Networks . . . . .	24
3.1.1 Regression Models . . . . .	25
3.1.2 Auto-encoders . . . . .	27
3.2 Methodology . . . . .	30
3.2.1 Data . . . . .	30
3.2.2 Sample Splitting and Cross Validation . . . . .	33
3.2.3 Dimensionality Reduction of the Macro-economic State . . . . .	33
3.2.4 Digesting the “ <i>Anomalies</i> ” . . . . .	34
3.2.5 Artificial Neural Network Regression . . . . .	36
3.2.6 Model Selection and Evaluation . . . . .	37
3.3 Conclusion . . . . .	39
4. DEEP RETURN PREDICTABILITY AND MARKET SENTIMENT . . . . .	40
4.1 Introduction . . . . .	40
4.2 Motivation . . . . .	44
4.2.1 Investor Sentiment, Arbitrage Asymmetry, and Expected Returns . . . . .	46
4.3 Results . . . . .	48
4.3.1 Predictive Accuracy and Profitability . . . . .	48
4.3.2 Market Sentiment and Return Predictability . . . . .	56
4.4 Conclusion . . . . .	60
BIBLIOGRAPHY . . . . .	62

# CHAPTER 1

## INTRODUCTION

The asset pricing literature has discovered a vast set of cross-sectional return predictive signals, commonly referred to as the factor zoo<sup>1</sup>. One key debate around the so-called factor zoo in the asset pricing literature stems from the source of these returns predictive signals. What causes returns to be predictable? This debate is at the core of asset pricing literature, and it has equally important implications for both the investors and policymakers.

There are three plausible explanations for this predictability. First, there is risk. It follows that in the presence of no-arbitrage and investor rationality, risk should be the only source of return predictability in informationally efficient markets. As a consequence, the factor zoo only exists as proxies for the underlying risk. Second, some studies attribute the return predictability to systemic behavioral biases that stem from psychological and sociological factors. These biases lead to patterns that can be observed in the data as systemic over and undervaluation. The observation of the return predictive signals or anomalies happens as the market corrects for these biases. Thirdly, return predictability could be a statistical illusion that arises due to extensive data mining. Evolution has rendered humans more prone to type 1 errors because type 1 errors would have helped us evade predators. Such evolutionary tendencies have equipped us to identify patterns and rationalize heuristic simplifications to explain these patterns.

In this thesis, I argue that cross-sectional return predictability is driven by mispricing. I use a novel deep learning based method to find a consistent pricing model for firm-level

---

<sup>1</sup> The term factor zoo is a little misleading as it has an implicit suggestion that the underlying cause is risk.

expected returns. I show that the resulting expected returns have properties that are consistent with a mispricing-based explanation. Deep learning models imply negative expected returns, which are very hard to justify with a risk-based explanation. Additionally, the implied predictability and profitability can be predicted using market sentiment. Moreover, the model loads more characteristics that imply difficult-to-arbitrage stocks, which are more vulnerable to mispricing. All of these findings make a strong case for mispricing as a source of predictability in returns. I complement my main findings in the last chapter with two related discourses in the earlier two chapters. I have two primary considerations in supporting my main case for market sentiment. Firstly, I make a case for deep learning models. I provide my discourse on the evolution of the so-called factor zoo and some remedies that modern quantitative methods (specifically deep learning models) offer. Secondly, I lay out the implementation details of my model, given the complexity of training a neural network. Because of this literature's novel nature, there are no specified procedures, the likes of which we would find for linear factor models and Fama-Macbeth regressions. Therefore, I discuss the methodology elaborately and provide ample implementation details for replication. I have used publicly available data and trained the models on my personal computer (training doesn't require high-performance computers).

My thesis is organized as follows: The first chapter provides a comprehensive introduction to the factor zoo's evolution. This chapter's main focus is to summarize the proposed machine learning solutions for the so-called factor-zoo and, most importantly, the theoretical implications of this literature's results. I start the background discussion from the early risk models and build up to the current debates. These discoveries' empirically driven nature is critical to understand the case for ANN models and, consequently, this thesis. In other words, It is essential to know that the process of discovering these factor models starts with an attempt to explain an observed difference in conditional mean returns based on some predictive signal, a so-called "anomaly". A discussion that naturally follows is the zoo of predictive signals proposed in the literature and presented as a multidimensional challenge by Cochrane 2011. Lastly, this chapter provides a brief review of machine learning and deep learning approaches to risk premia estimation proposed in the literature. To truly appreciate

the potential that deep learning has to offer in asset pricing, one must appreciate the little to none of the out-of-sample predictability implied by traditional linear models (Lewellen 2014).

The second chapter provides a brief introduction to ANN's and how I have trained my model. In the first section, I start with a discussion of ANN's, their universal approximation abilities, and their applications for prediction and dimensionality reduction purposes. The second section provides an example of auto-encoders(ANN's used for dimensionality reduction) and their comparison with principal component analysis(PCA). This section also compares the reconstruction accuracy of PCA and AE applied to financial and macro-economic times series data. This discussion is aimed at explaining the consequence of allowing non-linearity in model estimation. The third section discusses the prediction model that I use for my analysis of expected returns.

In training the model, I fit an ANN to approximate firm-level expected returns following Gu, B. Kelly, and Xiu 2020b. The prediction model utilizes both dimensionality reduction and predictive aspects of deep learning, using firm-level characteristics and a latent representation of the macro-economic state. I extract this latent representation of the macro-economic state separately using an autoencoder model. Compared to Principal Component Analysis (PCA), I show that a denoising autoencoder can better extract this lower-dimensional representation. I also use a novel activation function in the model and a novel hyperparameter optimization approach. I use a Gaussian process (Bayesian) approach for selecting hyper-parameters for the Adam optimizer. The complexity of deep learning models results in various hyper-parameters that can be tuned to adjust the model performance. These commonly include the number of neurons, the number of hidden layers, the choice of activation functions, and the optimizer's learning rate. I show that by training three Adam optimizer parameters (namely learning rate, first-moment decay rate, second-moment decay rate), there is little need to optimize the other hyper-parameters. These results are important because they imply a significant reduction of hyper-parameter search space for asset pricing specific problems and makes fitting ANN models more accessible to academic with little expertise in machine learning.

In the third chapter, I present evidence that the high return predictability implied by deep learning models and the resulting profitability is driven by mispricing. This is contrary to the literature that attributes this predictability to risk (Chen, Pelger, and J. Zhu 2019). I show that short-run losses are predictable, which violates the positive risk premium assumption of rational-expectations-based asset pricing models. The predictability implied by risk should be strictly limited to positive returns which seems to be violated in case of deep learning predictive models. A model that predicts positive returns out-of-sample is interpreted as approximating the risk premia or expected returns. But what should the prediction of losses be attributed to? I find that the 1-month horizon negative returns are even more predictable than the positive returns measured as the out-of-sample R-squared. Additionally, I find that this predictability in losses is mainly driven by periods following high market sentiment, and the positive return predictability is high following low sentiment periods. Furthermore, this predictability in losses disappears for longer return horizons, suggesting that mispricing appears for a short time and is then arbitrated away. An important aspect of longer horizon predictions is that the resulting profitability significantly deteriorates. These findings are consistent with Avramov, Cheng, and Metzker 2020 who find that deep learning models extract their profitability from difficult-to-arbitrage stocks, and incorporating economic restrictions in these models attenuates profitability.

# CHAPTER 2

## CROSS-SECTIONAL RETURN PREDICTABILITY: A CASE FOR DEEP LEARNING

The 2013 Nobel Prize in Economics was awarded to two influential researchers in asset pricing, Eugene Fama and Robert Shiller<sup>1</sup>. While seemingly at odds with each other, their research was interestingly seen by the committee as pivotal to our current understanding of asset prices. The two Nobel laureates have very different explanations to the most central question in asset pricing: *"Why do different assets earn different returns?"* Alternatively, we can ask *"What explains the predictable component in returns?"* Studies in the Fama camp rely on rational-expectation equilibrium models as their baseline framework, where return predictability is seen as a consequence of time-varying risk premia (Cochrane 2008). On the other hand, the Shiller camp provides evidence for systemic mispricing in the markets due to psychological biases. Return predictability, for behaviorists, can arise as these systemic biases correct back to fundamentals.

These two schools-of-thoughts have deeply influenced the risk modeling literature. On one side of the aisle, some studies formulate and empirically validate models of risk based on economic equilibrium conditions. The other side of the debate presents violations of these risk models. A review of the asset pricing reveals that this disagreement has contributed

---

<sup>1</sup> Lars Hansen shared this prize with the two. I mention only Fama and Shiller because of their opposing views highlighting the thesis and anti-thesis debate in asset pricing

significantly towards progressing the literature, with one side providing the thesis and the other providing the anti-thesis. This chapter provides a review of the factor modeling literature. Mainly, I stress the empirically driven nature of our understanding of risk factor models. I provide a brief background on risk models until the more recent phenomena of the so-called factor zoo. My thesis's relevant discussion is how the contemporary literature has attempted to reconcile these "anomalies" with mainstream asset pricing theory using risk-based explanations. Such a discussion warrants an overview of how this problem arises in the literature.

## 2.1 The Story of Risk: A History of Anomalies

A good starting point to understand the cross-section of returns is the risk-return trade-off applied by investors and academics alike. Synonymous to the English phrase "there is no such thing as a free lunch," this fundamental assumption has been extensively explored and is widely accepted in investment circles. To make higher returns, investors must take risky positions. Similarly, to avoid risky investments, they must settle for lower expected returns. However, measuring risk is a highly challenging problem empirically.

A simple measure of risk is the volatility in the assets returns because it signals uncertain future returns. Consider a simplistic assumption that assets with more volatile returns are riskier and investors would rather have positions in less volatile assets than highly volatile ones (for a given level of average returns). This is to say that investors would only accept higher volatility in returns if the average returns were higher. Another important assumption is that investors can choose to invest from a variety of correlated assets and can hold diversified portfolios. These two assumptions combined are key to how we measure risk.

Optimal Portfolio and Efficient Markets

***"In the beginning, there was chaos. Practitioners thought that one only needed to be clever to earn high returns. Then came the CAPM. Every clever strategy to deliver high average returns ended up delivering high market betas as well. Then anomalies erupted,***

*and there was chaos again.” (Cochrane 2011)*

Markowitz 1952 introduced the concept of a mean-variance efficient optimal portfolio, based on the simplistic assumption that investors consider high average returns “desirable” and high variance “undesirable”. In his paper titled “Portfolio Selection”, Markowitz shows how risk-averse investors can construct portfolios to maximize expected returns for a given level of portfolio variance. An equally weighted portfolio of  $N$  correlated assets would have portfolio returns equal to the average returns of the underlying assets; however, the standard deviation of portfolio returns would be lower than the average standard deviation of the underlying securities. Given a set of correlated time series, we can show that different combinations of these series’ yield many possible different risk-to-reward ratios. There exists an optimal choice that maximizes the Sharpe ratio of expected returns. This is an exercise conducted in most investment classes and serves as the foundation of modern portfolio theory.

Markowitz’s optimal portfolio was extended by Sharpe 1964, Litner 1965, and Black 1972, resulting in the widely used Capital Asset Pricing Model (CAPM). CAPM’s central prediction is that in an efficient market, the market portfolio is the optimal portfolio. In the presence of a positive risk-free rate, each investor will have a utility-maximizing portfolio that is a combination of the risk-free portfolio and the optimal portfolio of risky assets. If investors have homogeneous beliefs, they have the same linear efficient set called the Capital Market Line (CML).

The CAPM provided a theoretically motivated model to determine the relative riskiness of assets. According to the CAPM, an asset’s riskiness can be determined by its relative riskiness to the market portfolio. This relative riskiness is traditionally measured by running a linear regression of excess returns (returns in excess of the risk-free rate) of the asset to the market excess returns. The coefficient in this regression is the market beta, the measure of how risky a security compares to the market in risk. The CAPM concludes that the market beta is sufficient to explain the cross-section of stock returns. This optimal portfolio paradigm still dominates much of how we think about risk and return and is still widely used



by practitioners.

The general form of CAPM is presented as the following:

$$r_i = r_f + \beta_i(r_M - r_f) \quad \forall i = 1, 2, \dots, n$$

where,  $r_M$  is the expected rate of return on the market portfolio,  $r_i$  is the return on security  $i$ ,  $\beta_i$  is the market beta of security  $i$ ,  $r_M - r_f$  is the market risk premium,  $r_f$  is the risk-free rate.

The CAPM found great popularity in the 70's and 80's backed by sophisticated theoretical models that placed CAPM as a consequence of underlying economic equilibriums (Sharpe 1964, Litner 1965). However, the anomaly literature pointed toward deficiencies in the theory. The multi-factor models were introduced to reconcile the growing anomaly literature in the 1980s with the market model (Fama and French 1993). These multi-factor models included characteristic sorted portfolios or anomaly portfolios as additional factors in the market model regression. This section briefly summarizes the historical evolution of factor models, which is important to provide the readers with a background.

Size, Value, and the Multi-factor Model

***"If assets are priced rationally, our results suggest that stock risks are multidimensional. One dimension of risk is proxied by size, ME. Another dimension of risk is proxied by BE/ME, the ratio of the book value of common equity to its market risk." (Fama and French 1992)***

Soon after the CAPM was introduced, researchers documented patterns in the equity markets that were not explained by the market factor. These deviations from the CAPM came to be known as market anomalies and suggested further inquiry. For example, Banz 1981 documented that smaller firms have higher CAPM risk-adjusted return than larger firms. A portfolio with a long position in the smallest firms and a short position in the largest firms would have positive CAPM alphas on average. The firms' CAPM beta did not capture this "size effect", and hence, Banz 1981 concluded that the CAPM was misspecified.

The size effect was among the first "anomalies" documented in the 1980s. Some other prominent anomalies that followed were the "value effect" (Stattman 1980, Rosenberg, Reid, and Lanstein 1985), the "leverage effect" (Bhandari 1988), and the Earning to Price Ratio effect (Ball 1978) among others. These so-called anomalies contradicted one of CAPM's central predictions because the market beta should have captured these effects if the CAPM specification were true. Moreover, many of these anomalies could generate a positive CAPM alpha in a long-short portfolio.

To reconcile these anomalies with a risk-based explanation, Fama and French 1993 suggested that risk is multidimensional, and the so-called anomalies are capturing dimensions of risk that the market beta is unable to capture. The additional risk factors were approximated as characteristic-sorted long-short portfolios. This interpretation of anomalies is still predominant in risk modeling and gives rise to the term factor zoo.

Fama and French 1993 proposed a 3-factor model to accommodate a multidimensional representation of common risk. Fama and French included a decile-based long-short portfolio sorted on size and another one sorted on book-to-market ratio as the two additional factors. They show that including size and value portfolios as additional factors subsumed many of the prominent anomalies at the time.

The additional two factors' theoretical motivation was derived from the Arbitrage Pricing Theory (APT) of Ross 1976. The APT, unlike CAPM, does not assume perfectly efficient markets and is flexible in incorporating multiple factors

The general form of APT can be written as:

$$r_i = r_f + \beta_i^1(f_1) + \beta_i^2(f_2) + \beta_i^3(f_3) + \dots + \beta_i^k(f_k) \quad \forall i = 1, 2, \dots, n$$

Where  $f_k$  is the  $k$ th factor, and  $\beta_i^k$  is relative loading of the factor.

As a special case, the CAPM can be viewed as a single-factor model in the APT since the latter does not specify which factors should be used. However, like the CAPM, the APT assumes that a factor model can be used to describe the correlation between risk and return. The APT suggests that asset returns are linearly related to multiple variables that

capture systematic risk, viewed as hedges for different states in nature. The central idea in the APT is that there exists no arbitrage opportunity among well-diversified portfolios, and if such an opportunity arises, it is exploited away by rational investors (arbitrageurs). The CAPM adjusted differential in security returns when sorted on characteristics like size and value should not exist in the APT framework unless these portfolios load on systematic risk. If these anomaly portfolios load on some underlying common risk, only then can the arbitrageurs not exploit these portfolios. This would explain the CAPM adjusted differential on the anomaly portfolios.

Fama and French 1993 suggest that the size and value portfolios have partial exposure to unobserved state variables that affect the risk premiums in equity markets. Along with the market portfolio and the risk-free rate, these two factors span the efficient linear set. An important consideration here is that the size and value effects are not factors themselves but have partial loadings on the underlying systematic variable. This interpretation is very important for understanding the key arguments of this chapter. Fama and French 3-Factor Model can be written as:

$$r_i = r_f + \beta_i^M (r_M - r_f) + \beta_i^{SMB} (SMB) + \beta_i^{HML} (HML) \quad \forall i = 1, 2, \dots, n$$

Where,  $\beta_i^M$  is the market beta for asset  $i$ ,  $\beta_i^{HML}$  is the value beta capturing the sensitivity of the asset  $i$ 's return to the high-minus-low value factor  $HML$ ,  $\beta_i^{SMB}$  is the size beta that captures the asset  $i$ 's sensitivity to the size factor  $SMB$ .

***”At the time of our 1993 paper , these were the two well-known patterns in average returns left unexplained by the CAPM of Sharpe (1964) and Lintner (1965).”(Fama and French 2015)***

Despite the multidimensional approach of Fama and French 1993, more anomalies were introduced in the following literature and are being introduced to date. The 3-factor model did not explain much of these new effects. Recently, Fama and French 2015 proposed a

5-factor model with two additional factors, namely an investment factor and a profitability factor. This incorporated the gross profitability premium (Novy-Marx 2013) and the investment premium (Aharoni, Grundy, and Zeng 2013). The interpretation for these factors is the same in that these additional four factors are noisy signals for underlying systematic risks. In addition to these five factors, researchers also incorporate a momentum factor based on the momentum effect first documented by Jegadeesh and Titman 1993.

## The Factor Zoo

***”Alas, the world is once again descending into chaos. Expected return strategies have emerged that do not correspond to market, value, and size betas.”(Cochrane 2011)***

Despite the increasing dimensionality of the linear factor models, hundreds of new ”factors” that predict asset returns have been proposed in the literature. Green, Hand, and Zhang 2013 count 330 such stock- level predictive signals. Harvey, Liu, and H. Zhu 2016 study 316 signals, which include 213 firm-level characteristics and 113 common factors to describe stock return behavior. They note that this is only a subset of those studied in the literature.

These predictive signals can be broadly classified as firm-specific signals and common risk factors. The firm-specific signals, also referred to as firm characteristics, provide information on the firm’s financial health. Since these characteristics are firm-specific, they do not necessarily result in return co-movement and are diversifiable.

On the other hand, common risks are affecting the whole cross-section of returns and result in return co-variance. Although these factors are approximated using characteristics sorted portfolios, they approximate a common risk factor that eventually results in observed co-movements in returns. This co-movement is the anomaly that Fama and French 1993 seek to address with their multi-factor approach. For example, if size is a firm-level characteristic that predicts the firm’s financial state, why do stocks in the same size bucket move together? This co-movement of returns is what is typically referred to as a factor structure. These factor structures are indicative of common underlying risks for a set of stocks. It is important to note that the predictive characteristics and characteristics-based co-variances

are not necessarily mutually exclusive.

From this perspective, a predictive signal for stock returns does not necessarily constitute a factor. There exists asset pricing literature that deals with factor identification and distinguishing between characteristics and co-variances. For this chapter's purpose, characteristic-based portfolios are assumed to have partial loadings on the systematic factor and not a firm-level predictor variable.

A characteristic-sparse factor model (Fama and French 5-factor model) considers only a subset of the factors discovered in the literature. Besides, as Fama and French 1992 suggests, these factors can be interpreted as part loading on the true systematic factors and are at best approximations of true underlying factors. Given the large number of potential factors in the literature, emphasizing only a few of these partially loaded portfolios might not be optimal. Moreover, unlike the CAPM market factor, the Fama and French 5-factor model's additional factors do not have any meaningful interpretation beyond the APT and, eventually, the law of one price. Therefore, a factor model that relies on many potential factors can better estimate the underlying risk compared to a characteristic sparse model.

The following section discusses recent literature that provides an alternative to the characteristic-sparse representation of the common risk. The methodologies discussed incorporate a larger set of potential factors to approximate the true underlying systematic risk factors.

## 2.2 Modelling Expected Returns

***“Portfolio sorts are really the same thing as non-parametric cross-sectional regressions, using non-overlapping histogram weights.” (Cochrane 2011)***

In the last decade or so machine learning has emerged as a novel paradigm in risk modelling. This trend is credited to the multidimensional nature of the problem, formally summarized by Cochrane 2011. Studies published in the years since show vast potential in marrying machine learning and asset pricing. This literature can be broadly categorized into two groups based on how they formulate their problems and this distinction can be viewed

from the lens of the famous characteristics vs co-variance debate in the 1990's (Fama and French 1993, Daniel and Titman 1998).

First, there are the papers that tilt toward the co-variance side of the debate and seek to explore latent factor representation of common risk. This approach has been extensively explored since the Arbitrage Pricing Theory (APT) of Ross 1976. Fama and French 1993 and Fama and French 2015, for instance, interpret characteristics sorted long-short portfolios as approximations for unobservable common risk factors. This is a theoretically elegant approach, with a well-established no-arbitrage equilibrium condition and a risk-based explanation. However, the existence of a large number of latent risk factors are not plausible and hence not consistent with the factor zoo if we consider each characteristic to approximate a unique risk. With the assumption that the characteristic-sorted portfolios have an underlying low-dimensional latent structure, feature extraction tools are well suited to marrying this approach with the growing list of factors. We can apply feature extraction on the factor zoo only under the assumption that there are a small number of underlying risk factors that explain much of the cross section of stock returns. There is literature that explores such feature extraction techniques offered by machine learning. It is useful in extracting lower-dimensional latent representations of multi-dimensional macroeconomic and financial time-series (Kozak, Nagel, and Santosh 2018, B. T. Kelly, Pruitt, and Su 2019). This literature is in the spirit of the co-variance view where long-short portfolios are used to approximate latent risk factors (Fama and French 1993).

Second, another stream of the literature approximates risk premia using predictive models to obtain an estimate of firm-level expected returns (Lewellen 2014, Gu, B. Kelly, and Xiu 2020b). This stream of literature is in line with the characteristics view that uses regression models to provide firm-level expected return estimates (Daniel and Titman 1998, Lewellen 2014). Artificial Neural Networks (ANN'S) commonly referred to as deep learning models, have shown to be the most promising among all the candidate methods considered by these studies. Deep learning models have been applied by the literature in both dimensionality reduction (Gu, B. Kelly, and Xiu 2020a) as well as prediction problems (Gu, B. Kelly, and Xiu 2020b, Chen, Pelger, and J. Zhu 2019). The predictive gains from these

models are attributed to their incorporation of flexible functional forms and their allowance for non-linear interactions among variables.

### 2.2.1 Factor Models

Recent studies find that once the latent factors are accounted for, the characteristics-sorted long-short portfolios have little to no predictive power (B. T. Kelly, Pruitt, and Su 2019). This is an important finding since it substantiates the claim that the factor zoo is a manifestation of an underlying low dimensional factor structure. Kozak, Nagel, and Santosh 2018 demonstrate that a small number of principal components of anomaly portfolios are able to account for the alphas of these anomalies. This is an important result since ideally the alphas should be zero in the APT. The principal components are designed to maximize the factor variance. Huang, Li, and Zhou 2019 propose a reduced rank approach designed to reduce the dimensions of the factor zoo to explain the cross section of returns. This is more in line with theory and a five-factor version of this approach outperforms the equivalent Fama and French 5- factor model for pricing target portfolios. Over the last half-decade, a variety of techniques has been explored to address the factor zoo. This section discusses two prominent dimensionality reduction approaches proposed in the literature and their empirical performance. One approach is a PCA inspired approach, which allows for time-varying loadings while the other is a deep learning inspired autoencoder. These two approaches are selected for discussion because they provide an interesting contrast of linear and non-linear methods available to cater to this problem.

#### Re-evaluating Factor Models

*"We are going to have to repeat Fama and French's anomaly digestion, but with many more dimensions. We have a lot of questions to answer. First, which characteristics really provide independent information about average returns? Which are subsumed by others? Second, does each new anomaly variable also correspond to a new factor formed on those same anomalies? Third, how many of these new factors are really important? Can we again account for  $N$  independent dimensions of expected returns with  $K < N$*

*factor exposures? Can we account for accruals return strategies by betas on some other factor, as with sales growth?” (Cochrane 2011)*

To address this “multidimensional challenge” of Cochrane 2011, machine learning offers a variety of dimensionality reduction techniques. One potential solution to the factor zoo is feature selection. Feature selection, as the name suggests, incorporates selecting most informative features from a given set of features and filtering out redundant ones, based on a certain selection criterion. Fama and French 2015 5-factor model is one example of important features selected manually over the past two decades. Alternatively, one could simply identify groups of factors that are correlated with each other, and then select one feature from each group that explains most of the expected returns. Kogan and Tian 2015 conduct a similar factor mining exercise to select appropriate features. There are other studies that explore more sophisticated dynamic feature selection techniques (DeMiguel et al. 2017, Freyberger, Neuhierl, and Weber 2017, Feng, Giglio, and Xiu 2017). However, despite an initial consideration of more characteristics, information loss is inevitable when relying on a small subset. Financial data is notorious for noise and the noise to signal ratio in these characteristic sorted portfolios is very high.

On the other hand, feature extraction extracts new features from the initial features that are meant to be informative and non-redundant. These techniques are especially effective in estimating the underlying latent structure in the data in presence of noise. Hence, feature extraction is well suited to the factor zoo since the observed “factors” are interpreted as noisy signals for the underlying latent state variables (Fama and French 1993, Fama and French 2015). Feature extraction remedies the noise concern given a large enough feature space and observations. Additionally, feature selection applied to the factor zoo would lead to a characteristic-sparse model and could potentially lead to spurious over-fitting (i.e. fitting the noise instead of the true signal).

Recent literature has explored a wide range of feature extraction methods for factor mining (B. T. Kelly, Pruitt, and Su 2019, Kozak, Nagel, and Santosh 2018). The following subsections are organized as follows. First subsection uses a simple example to provide the



readers with a general intuition for feature extraction and its possible benefits in presence of noisy signals. The next sub-section demonstrates the main idea using a simple PCA applied to characteristics sorted portfolios. The third section briefly discusses the modern literature and two major approaches to the factor zoo problem.

#### A Case for Feature Extraction

***”These (factor) models are reduced-form because they are not derived from assumptions about investor beliefs, preferences, and technology that prescribe which factors should appear in the SDF. What interpretation should one give such a model if it works well empirically?” Kozak, Nagel, and Santosh 2018***

There could be many potential ways of extracting the feature from a given set of anomaly returns. Many feature extraction algorithms are applied in the literature for tackling this problem. In addition, studies have altered dimensionality reduction methods to incorporate asset pricing theory. However, the fundamental case for feature extraction can be made using a vanilla Principal Component Analysis (PCA), compared to the characteristic-sparse Fama and French 5-factor model.

The Principal Component Analysis (PCA) is the most common feature extraction technique that is widely applied in finance (Baker and Wurgler 2006, Kozak, Nagel, and Santosh 2018). PCA is used to reduce a large number of possibly correlated variables into a small number of uncorrelated variables such that the first principal component (PC) captures most of the variance in the data; the second PC captures most of the remaining variance and so on. Hence, PCA is geared to represent a higher dimensional data in lower dimensions such that maximum possible variance is captured by the low dimensional representation.

More formally, PCA takes the  $N$ -dimensional feature set and finds  $M$  orthogonal directions in which the data have the most variance. Where  $N$  is the number of features in the original feature set and  $M$  is its lower dimensional representation. PCA reduces the dimensions of a given feature set by orthogonally transforming the data into a set of principal components such that the first principal component has the most variance, the second princi-

pal component has the second most variance and so on. The first few principal components capture the most variance and are treated as the “extracted” features in PCA.

If the goal were simply to explain the maximum variation in characteristics managed portfolios, a Principal Component Analysis (PCA) would do a better job compared to the Fama and French multi-factor model. This is demonstrated in Figure 2.1. The first five Principal Components (PC’s) of 50 long-short anomaly portfolio daily returns used by Kozak, Nagel, and Santosh 2020, along with the market factor, can explain the anomaly portfolio returns better than the Fama, French, and Cahart 6-factor model.

Following Kozak, Nagel, and Santosh 2020, the characteristic portfolios are orthogonalized to the market returns to take care of any correlation with the market factor, which is the level factor. The first 5 PC’s of the orthogonalized characteristics managed portfolios are used. The resulting model is a 6-PC-factor model with the 5 PC’s and one market factor included in the regressions. Then we run 50-pooled regressions with the characteristics managed portfolios as the dependent variables and, 5 PC’s and market excess returns as independent variables. For comparison, our baseline model is the Fama, French and Cahart 6-factor model, which includes the momentum factor in addition to the five factors in the Fama and French 5-factor model.

Figure 2.1 reports the regression R-Squares from the pooled regressions. The 6-PC-factor model outperforms the Fama, French, and Cahart 6-factor model in terms of the model fit (i.e. the portfolio return variation is better captured by the extracted features as compared to the characteristic-sparse model). An important trend to note here is that the PC model captures substantial variation among most of the characteristic portfolios. This suggests that there is substantial commonality among these portfolios and indicates the existence of an underlying latent structure. The characteristics sparse model, however, does not do as well in explaining the portfolio returns.

These results, however, are only to demonstrate the common variation in these anomaly portfolios and that the extracted features can be used to explain portfolio returns. This issue has been discussed in more detail in Kozak, Nagel, and Santosh 2018. Recent studies published in notable finance journals explore the issue with more scrutiny. The next section

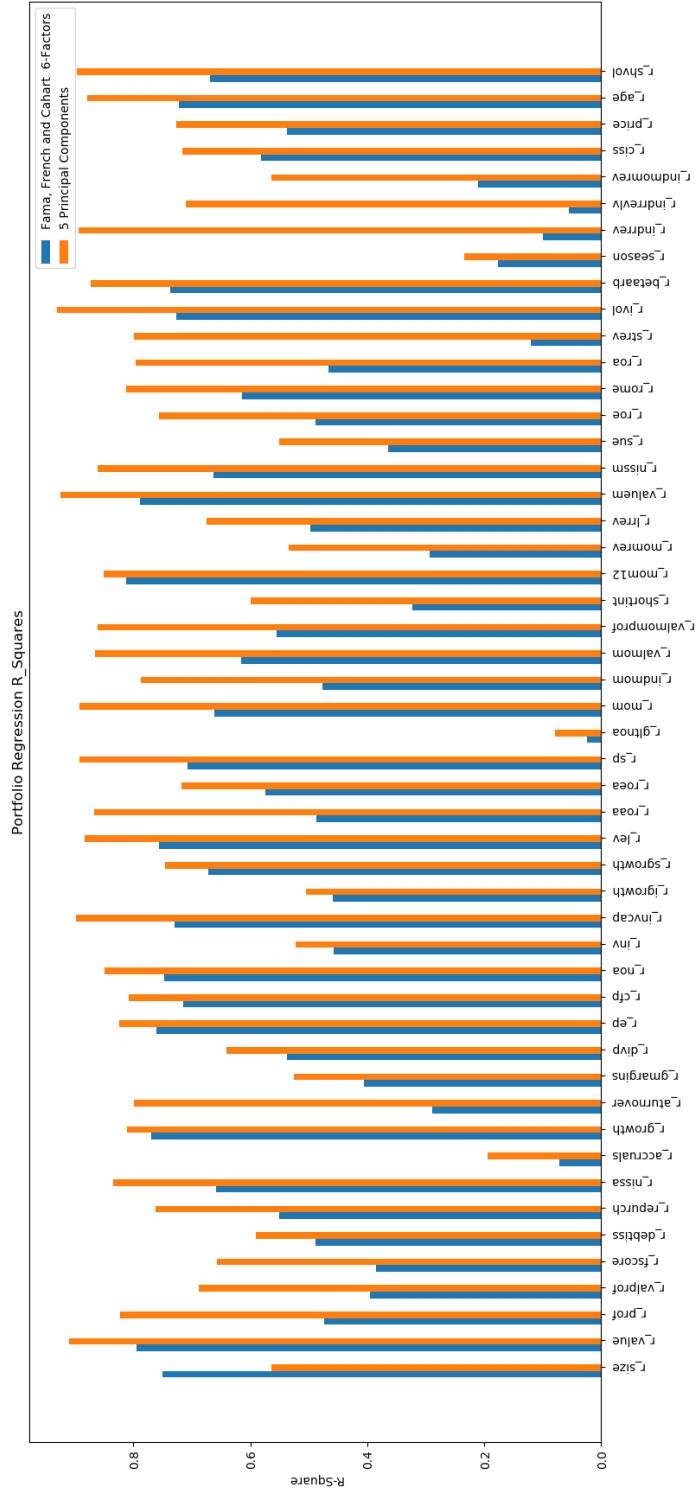


Figure 2.1: **Common Variation in Anomaly Returns**  
Shows the regression R-Squares where each characteristic managed portfolio is regressed on the 5 PC + Market Factor, as well as the Fama, French and Cahart 6-Factor Model

provides a brief review of these papers along with regression models that fundamentally explore the same issue.

#### Feature Extraction in the Literature

B. T. Kelly, Pruitt, and Su 2019 propose a novel Instrumented Principal Component Analysis (IPCA), which allows for latent factors as well as time-varying loadings. Their IPCA technique is flexible to accommodate characteristics effects in the characteristic managed portfolios. If the main driver of the characteristic/return relation is a common risk, then the IPCA identifies the corresponding risk factor. Whereas, if there exists no such factor, IPCA infers it as a characteristic effect and allocates it to the anomaly intercept. This ensures that the latent structure extracted with IPCA is truly the common risk. The authors show that this IPCA model outperforms the Fama, French and Cahart 6-factor model in explaining the cross-section of stock returns. In addition, this method outperforms PCA in explaining test portfolios that are different from the characteristic portfolios used to estimate the latent factors: suggesting that it captures the true systematic risk better. The main findings of this study show that most of the so-called anomaly characteristics have predictive power for stock returns because of time-varying exposures to underlying unobservable factors. They compare the performance of the conditional IPCA model compared to Fama, French and Cahart 6-factor models. The so-called anomaly-return alphas see a significant reduction in anomaly portfolio alphas. This is a very important result, which suggests that the so-called factor zoo can be explained by time-varying exposure to real unobservable factors. This common latent structure is also evident in the higher regression R-squares of the first 5 principal components.

The IPCA approach described in the above section has very important implications for the so-called factor zoo. However, this model is restrictive because it imposes a linear mapping of the characteristic anomaly returns to the latent factors. There is no theoretical guidance to such a restriction in dimensionality reduction. An autoencoder is well suited to non-linear dimensionality reduction. Gu, B. Kelly, and Xiu 2020a use an autoencoder to model the latent factors as nonlinear functions of the characteristic assets. Their model

is sophisticated as it models the conditional exposures as well as imposes a no-arbitrage condition. While the model architecture is out of scope of this chapter, the main results of this study shows the promise of this technique when applied to the factor zoo problem. The IPCA outperforms all model candidates for out-of-sample R-Squares. However, for out-of-sample predictive performance, the autoencoder performs better compared to all other models. This suggests that incorporating non-linearity can aid in better generalizations of the underlying process.

The literature in such dimensionality reduction methodologies is in its early stages and recent developments in machine learning offers many other viable methodologies to explore the issue. For example, most of the approaches introduced in the field seek a latent representation, which is linearly related to the expected returns. This is consistent with the prior literature on factor modelling and the resulting approaches use linear or non-linear dimensionality reduction approaches. However, this approach is restrictive even while using non-linear mapping functions like the conditional autoencoder example discussed in the chapter, because these models impose a limiting functional form on the SDF.

### 2.2.2 Regression Models

From a prediction standpoint, studies have documented that nonlinear prediction algorithms perform well in modelling stock returns using firm level characteristics as well as macroeconomic states (including characteristic portfolios). Using a deep learning model not as a dimensionality reduction tool (autoencoder) but as a predictive model increases the prediction accuracy many-fold, both in, in-sample and out-of-sample (Gu, B. Kelly, and Xiu 2020b, Chen, Pelger, and J. Zhu 2019). Additionally, it makes more sense theoretically not to restrict the functional specification of the SDF. The higher out-of-sample prediction accuracy suggests that the models are estimating robust risk prices and the SDF estimates are extremely consistent over time. Although the literature is new, the results in these studies are potentially paradigm shifting in risk modelling and is definitely something that would improve and gain more consensus in the risk modelling literature in the future.

Expected returns approximation is essentially a prediction function of the form:

$$E_t(r_{(i,t+1)}) = g^*(X_{(i,t)}, I_t) \quad (2.1)$$

where  $E_t(\cdot)$  is the expectation at time  $t$ .  $r_{(i,t+1)}$  is the excess return for asset  $i$  at time  $t + 1$ .  $g^*(\cdot)$  is a real-valued deterministic function that provides the conditional expected returns given the conditioning information. This function takes a firm-level input vector  $X_{(i,t)}$ , and a macro-level input vector  $I_t$ , both reported at time  $t$ .

The multidimensionality of the problem and the noise in the data renders the traditionally popular linear econometric regressions obsolete for prediction purposes, consequently providing suitable settings to leverage the advances in machine learning. Recently, this paradigm has gained popularity in the asset pricing literature, where studies show that using high dimensional methods that allow for nonlinear signal interactions significantly improves predictive performance. Artificial Neural Networks (ANN's) have shown to be the best performing models in this regard. In this thesis, my main focus is on this approach of firm-level prediction regressions. As pointed out by Cochrane 2011, the two approaches are almost equivalent in terms of the common cross-sectional variability that they capture. Regression models have less restrictive theoretical assumptions and allow better out-of-sample empirical performance.

## 2.3 Conclusion

In this chapter, I present a discourse on the evolution of the common risk factor models. The main point that I stress in this chapter is the empirically driven nature of this paradigm. The risk-return paradigm suggests that returns should be unpredictable and any empirical predictability should be due to risk. Any predictive signal or anomaly is thus attributed to risk with the underlying assumption that the risk model is misspecified. Hence, the focus in this literature has been to estimate risk premia better by incorporating most of the observed anomalies. In such studies, the view that characteristics based long-short portfolio returns are “anomalous” is challenged. These studies contend that these portfolio returns are actu-

ally manifestations of an underlying factor structure that approximates the true Stochastic Discount Factor (SDF). Modern feature extraction tools applied to the factor zoo provide us with very good approximation of the true latent factors. This approach has the theoretical elegance of APT inspired factor models and incorporates much of the factor zoo.

There are two main methods employed by the empirical literature to evaluate return predictive variables (anomalies). The first method to evaluate a proposed variable is to form decile portfolios and compare their mean returns or alphas. It is widely accepted that decile portfolios formed on conditional variables such as firm size, or book-to-market ratios earn different returns which are not explained by the portfolio market betas. Back in the 1980's when these predictive variables were discovered, they stood in clear contradiction of the CAPM and were presented as market anomalies. In the seminal papers of Fama and French 1992, they argue that this difference in mean returns is orthogonal to the market beta and that a multidimensional representation of the cross-section is warranted. They presented decile based long-short portfolios as approximations of these "risk factors" and proposed a 3-factor model that explain the cross-section of stock return better. Recent literature presents evidence of substantial commonality in the characteristic portfolio returns, and shows that once this commonality is extracted, there is little to no predictive power left in the characteristics portfolio. This empirical fact alone is suggestive that these portfolio returns are not anomalous and has important implications for discussion surrounding risk and market sentiment. However, this discussion evades the scope of this chapter. The authors seek to present this empirical pattern in data along with a discussion to provide readers with a perspective.

The second method which the literature explore is firm-level regressions. These regressions provide us with a direct measure of expected returns. The challenge is again the multidimensionality of the problem and this is where machine learning methods shine. Gu, B. Kelly, and Xiu 2020b is a good benchmark study which shows that machine learning methods have far superior predictive performance compared to the traditional regression-based methods used in the literature. What is far more surprising and previously unheard of in asset pricing is the out-of-sample predictive accuracy that these models exhibit. They find that shallow artificial neural networks have the best predictive performance among all the

candidate machine learning models that they consider. In my thesis, I use a novel method to estimate an ANN model of firm-level expected returns. I find that the monthly firm-level return predictions follow a market-wide sentiment. This suggests that at least part of the return predictability stems from mispricing and not risk. Chapter 2 provides the implementation details of my model along with a brief introduction of ANN models for prediction and dimensionality reduction purposes. Chapter 3 presents my main findings.



# CHAPTER 3

## DEEP RETURN PREDICTABILITY

This chapter discusses the training approach I have used to train a deep learning model. Starting with a brief introduction of the method, I draw a parallel between a linear regression and an ANN regression. I build on this by introducing the dimensionality reduction specification of ANNs known as auto-encoders. Second, I discuss the model I use to obtain the firm-level expected return predictions. My model uses both dimensionality reduction and prediction models. In addition, I use a novel Bayesian optimization technique to select the optimal specifications for the gradient descent algorithm used to train the ANN model. I divide this chapter in two sections, with one section providing an introduction to artificial neural network models and the second section discussing the methodology.

### 3.1 Artificial Neural Networks

An artificial neural network (ANN) is an information processing systems that is used to obtain a prediction model, mapping one or more inputs to one or more output. An ANN learns from data to alter the mathematical model with an objective to minimize the prediction errors. This is done iteratively until the prediction errors cannot be minimized any further (the model converges). ANN's have gained popularity in the last decade and currently lie at the core of applications in computer vision and natural language processing.

ANN's have found applications in a variety of tasks including prediction, dimensionality reduction and classification tasks. For the purpose of my thesis, I use prediction and dimensionality reduction properties of ANN's. To predict firm-level returns given a high

$$y = b_1 + w_1(x_1) + w_2(x_2) + w_3(x_3)$$

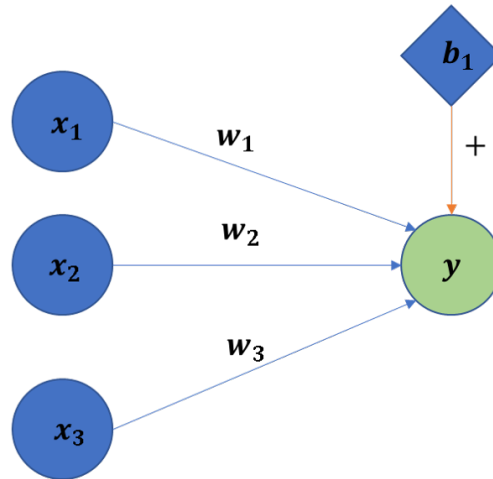


Figure 3.1: **Multivariate Linear Regression as a Neural Network**

This figure presents a pictorial representation of the regression presented in equation 3.1.

dimensional input data, I use an ANN regression. This problem is similar to that of a multivariate regression. For the dimensionality reduction of high dimensional macro-economic data, I use the auto-encoder specification of ANN's to obtain a low dimensional representation of the macro-economic state. This problem is comparable to taking the first few principal component of the data.

### 3.1.1 Regression Models

In this section, I draw a parallel between the commonly used linear regressions and ANN regression models. My aim is to provide a high level introduction of the structure of a simple ANN function to an audience with prior knowledge of linear regressions. I start with the visual representation of a multivariate regression and then I build on top of it to arrive at an ANN function with non-linear activations.

Consider the following multivariate linear regression:

$$y = b_1 + w_1(x_1) + w_2(x_2) + w_3(x_3) \tag{3.1}$$

This function can be represented by figure 3.1 where  $x_1, x_2$ , and  $x_3$  are model inputs,  $w_1, w_2$ , and  $w_3$  are their relative coefficients and  $b_1$  is the intercept. In ANN terminology, we refer to the intercept  $b_1$  as the bias, the coefficients  $w_1, w_2$ , and  $w_3$  as the weights, and

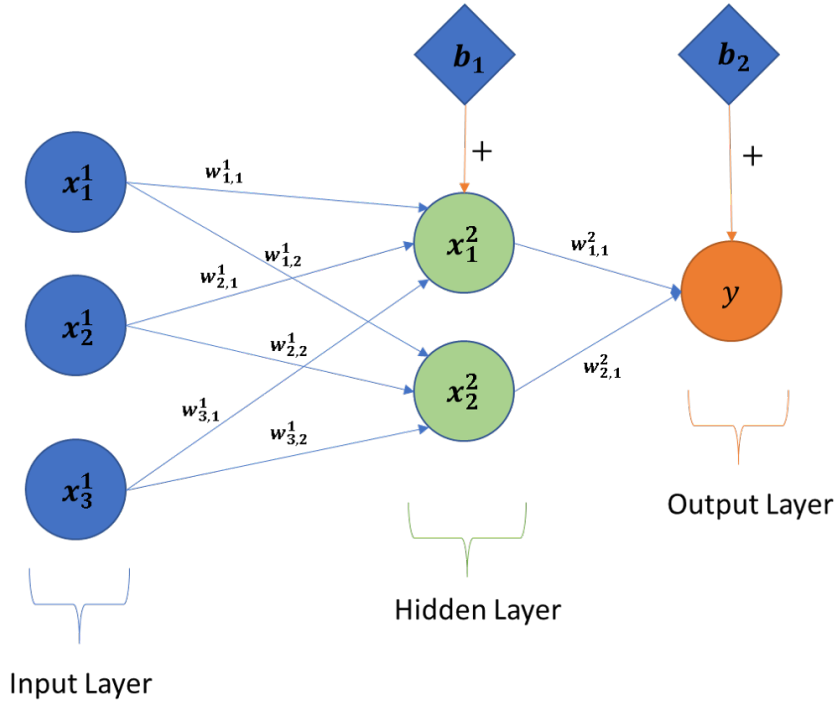


Figure 3.2: **Neural Network Regression with one hidden layer**

This figure presents a pictorial representation of the ANN regression presented in equation 3.2.

the inputs  $x_1, x_2,$  and  $x_3$  as neurons. These neurons connected through weights serve as the building blocks of the ANN models.

To allow for non-linearity in variable interactions the output  $y$  of the regression function 3.1 goes through a non-linear activation function  $f_{act}(\cdot)$  as follows:

$$y = f_{act}(b_1 + w_1(x_1) + w_2(x_2) + w_3(x_3))$$

For simplicity, I write the equation as:

$$y = f_{act}(b_1 + X_1 W_1) \tag{3.2}$$

where  $W_1 \in R^{3 \times 1}$  is the weights vector  $X_1 \in R^{1 \times 3}$  is the input vector and  $b_1 \in R^{1 \times 1}$  is the bias term.

The model in equation 3.2 is a neural network with no hidden layers, 3 input variables and one output. We can further expand this by adding an additional layer.

$$X_2 = f_{act}(b_1 + X_1 W_1)$$

$$y = f_{act}(b_2 + X_2W_2) \quad (3.3)$$

where  $W_1 \in R^{3 \times 2}$ , and  $W_2 \in R^{2 \times 1}$  are the weights vectors,  $X_2 \in R^{1 \times 2}$  are the hidden neurons and  $b_2 \in R^{1 \times 1}$  is the bias term for the output layer.

Figure 3.2 provides a visual representation of the model in equation 3.3. The weights and biases are the trainable model parameters that provide the model with it degrees of freedom. For model training, these parameters are generated from a random distributions and are altered iteratively using a Stochastic Gradient Descent algorithms to minimize the prediction errors.

Figure 3.2 presents a very simple ANN for visualization purposes. Artificial Neural network can have many hidden layers with each layer having many neurons. There are various ANN architectures based on the types of neurons and how they are connected together which are customized to solve a variety of problems. For the purpose of this thesis, I use a simple Feed Forward Network (FFN) as shown in figure 3.2 with four hidden layers which is detailed in the following methodology section.

### 3.1.2 Auto-encoders

An autoencoder is a neural network architecture used for dimensionality reduction. This is achieved by having a bottleneck hidden layers with a lower dimensionality than that of the input layer. The output of the auto-encoder is the input and the objective of this function is to replicate the input as well as possible. As a result, an auto-encoder learns a lower-dimensional representation of the higher dimensional data. It can discover hidden structures in the data while developing a compressed representation of the input. An autoencoder is comparable with the familiar Principal Component Analysis(PCA) in that both methods are used to extract lower dimensional representations for the data.

A dimensionality reduction model's objective is to obtain a compressed representation of the data while retaining maximum information from the original high-dimensional data. In other words, we need an  $N$ -dimensional representation of an  $M$  dimensional series, where  $N < M$ . In the case of PCA, this is achieved by taking the first  $N$  principal components

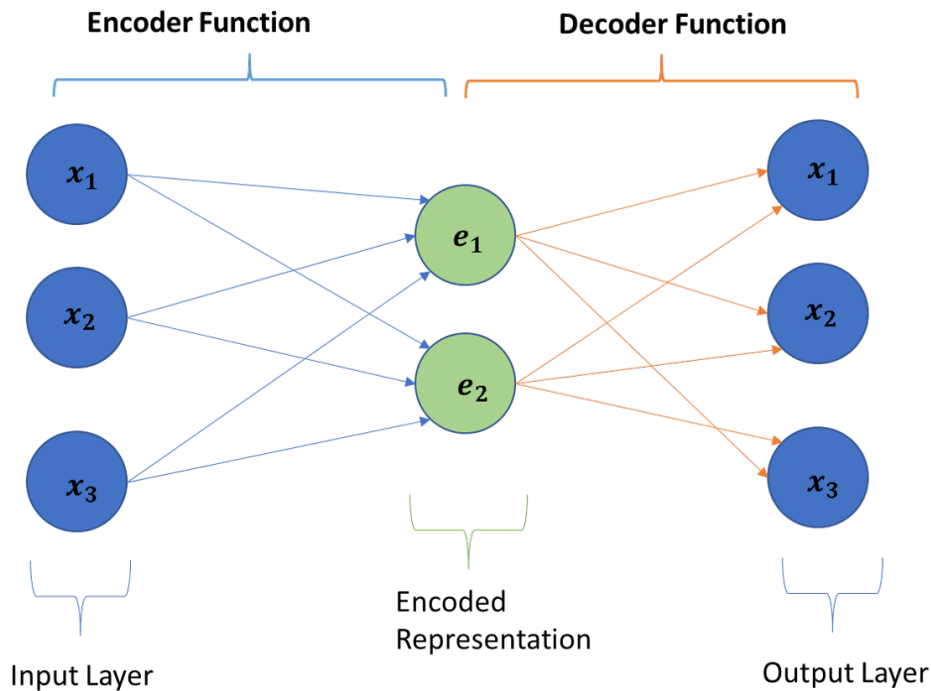


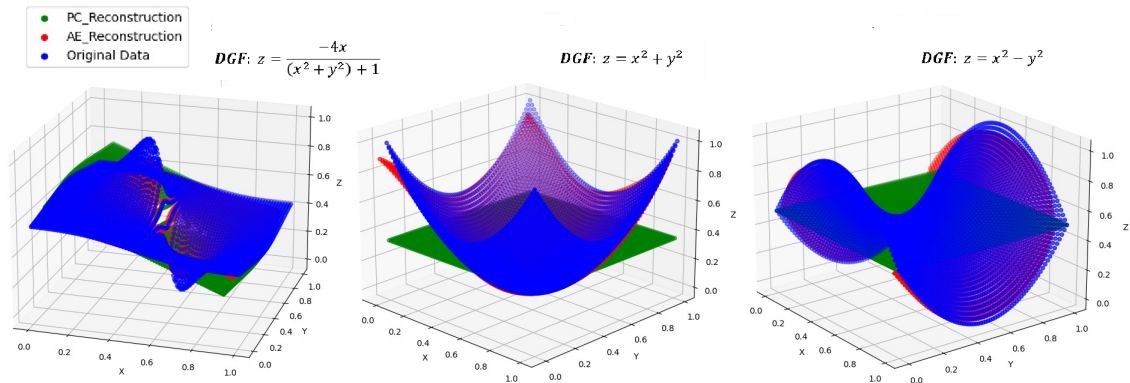
Figure 3.3: A Simple Auto-encoder

This figure demonstrates a simple autoencoder with one hidden layer, where  $M = 3$  and the "bottleneck"  $N = 2$ .

as the compressed representation of the data. An equivalent autoencoder is given the task to reproduce the input of dimension  $M$  and is given a bottleneck layer with  $N$  neurons. In this neural network, the output is the same as the input, and there is a "bottleneck" layer in the hidden layers. This forces the autoencoder to learn meaningful structures in the data to obtain compressed representation while allowing for non-linearity. Figure 3.3 provides an example of a simple autoencoder with one hidden layer where  $M = 3$  and  $N = 2$ .

### PCA Vs Auto-encoders

The Principal Component Analysis (PCA) is a dimensionality reduction technique that has been widely used in finance (Baker and Wurgler 2006, Kozak, Nagel, and Santosh 2018). This section introduces the effectiveness of simple auto-encoders contrasting with that of the PCA. The key focus is to demonstrate the ability of auto-encoders to approximate the underlying Data Generating Function (DGF). For a visual comparison of autoencoders and PCA, consider the example presented in Figure 3.4. Given a 2-dimensional grid of  $x$  and  $y$ , a 3rd dimension  $z$  is generated using the Data Generating Function (DGF)  $z = f(x, y)$ .



**Figure 3.4: Auto-encoder vs PCA: The role of Non-Linearity**

A 2-dimensional representation of the 3-dimensional original data (blue) is extracted using PCA and an Auto-encoder. The figure shows the reconstruction from both the PCA (green) and the Auto-encoder (red). The Auto-encoder captures the non-linearity that PCA fails to incorporate.

This 3-dimensional data is compressed into a 2-dimensional representation using PCA and autoencoders. Figure 3.4 presents a 3-dimensional visualization of the original data (blue) along with the reconstructed data from autoencoder (red) and PCA (green). Autoencoders can better compress the data compared to linear methods like PCA. As Figure 3.4 shows, when the data lies on or near curved manifolds, linear techniques such as PCA lose a lot of information in dimensionality reduction. While PCA is a quick brute force dimensionality reduction solution, it does not capture the underlying data generating function. In contrast, autoencoders are flexible in estimating curved manifolds, essentially learning the underlying function in learning a compressed representation of the data.

### Auto-encoder Training

ANN training starts with a random generation of the trainable parameters. The parameters are then updated to minimize the loss function (for example the mean squared error). This is similar to the Ordinary Least Square estimation of a linear regression, however, for a neural network this is done in iterations. With each iteration, the model learns a little bit more about the data.

Consider the autoencoder training example presented in figure 3.5, which is a continuation of the auto-encoder visualizations presented earlier. Initially the reconstructed output from the auto-encoder produces random data without much variance compared to the orig-

inal data. This is because at this stage the model parameters are randomly generated and the model training has not started. After 25% of model training, the reconstructed data appears as a plane and the loss is visibly decreased<sup>1</sup>. As the training approaches 75%, the reconstructed data begins to show curvature and with 100% training, the model is able to fully reconstruct the data.

## 3.2 Methodology

### 3.2.1 Data

In my analysis, I use three types of data:

1. The monthly cross-sectional ranking of each firm on 54 firm-level predictor characteristics (for model input  $X_{(i,t)}$ ).
2. A multivariate representation the macroeconomic state each month (for model input  $I_t$ ).
3. Market Sentiment (investor optimism and/or pessimism) every month.

For the monthly cross-sectional predictor data, I use the publicly available data-set of all U.S. CRSP firms (excluding micro-caps), used by Kozak 2019<sup>2</sup>. This data has two desirable properties: the high dimensionality in the firm state and the preexisting normalization of variables. This high dimensional data of 54 firm-level predictive signals is perfect for my analysis since it incorporates variables loading on both risk and presumed mispricing<sup>3</sup>. Additionally, the data excludes market caps below 0.01% of the total market equity at a given time<sup>4</sup>. Moreover, each characteristic signal is normalized as the relative cross-sectional rank of the firm  $i$  on that characteristic, for a given time  $t$ <sup>5</sup>. Such normalization is essential for an

---

<sup>1</sup> In this example, the auto-encoder was trained for 2000 epochs, the training percentage is calculated accordingly.

<sup>2</sup> <https://sites.google.com/site/serhiykozak/data?authuser=0><https://sites.google.com/site/serhiykozak/data?authuser=0>

<sup>3</sup> 14 of the predictors load on the four risk factors of Fama and French 2015, while the other 40 do not, hence the later is categorized as "anomalies"

<sup>4</sup> to exclude micro-caps that are more prone to mispricing and could potentially drive the results.

<sup>5</sup> For details see equations (23) and (24) in Kozak 2019

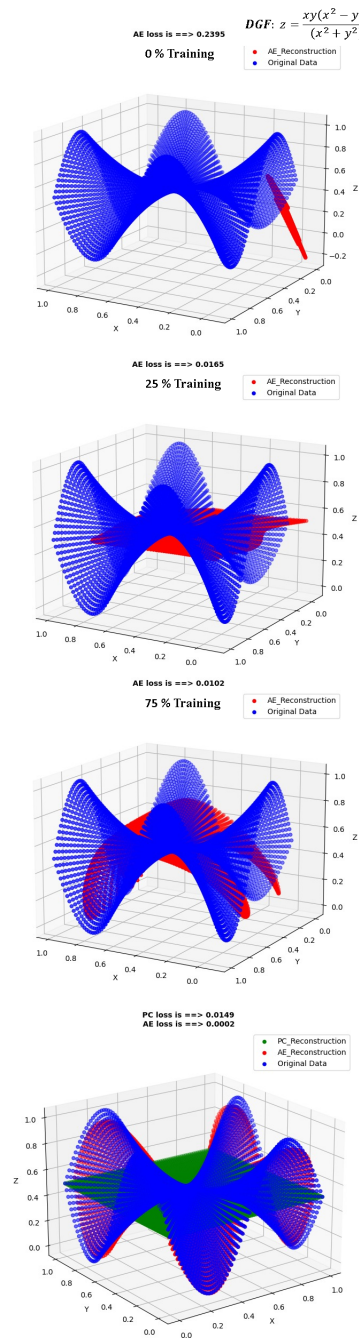


Figure 3.5: This figures shows the auto-encoder reconstruction at various stages of training. At the beginning, the model parameters are randomly generated and hence there is no variance in the reconstructed outputs. As the model trains more, the reconstructed output gets close to the original data



ANN regression, and this data comes ready to be used as such. I replace the missing values with 0 following Gu, B. Kelly, and Xiu 2020b and retain the firm month observations.

For the macro-economic data, I use the Fred MD database of 124 monthly macroeconomic predictors. I apply transformations to these variables as suggested in McCracken and Ng 2016. After this transformation, I scale the data between -1 and 1 using Min-Max normalization<sup>6</sup> given by the equation:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.4)$$

This, however, is not the input to my main regression. I use a low-dimensional latent representation of this data (detailed in the methodology section). The model input is a 10-dimensional series to represent the macroeconomic state.

In my primary analysis, I measure market sentiment using the market-based monthly sentiment index constructed by Baker and Wurgler 2006. For robustness check, I use the related sentiment index of Huang, Jiang, et al. 2015. This data is available from the authors' website. I also use two other measures from the options market that have been proposed as sentiments proxies: the Implied Volatility Index (VIX) and the put-to-call ratio (PCR). This data is reported at the daily frequency by the Chicago Board Options Exchange (CBOE). To extract a monthly sentiment measure, I use the average daily observation in the last ten days of the estimation month  $t$ .

For the output variable of excess returns, I use the CRSP MSF file. I winsorize the excess returns at 99% to curtail the effects of extreme observations. My total sample starts in 1963 and ends in 2019, where I use the period 1963-1986 to train my model, the years 1987-1991 to select an optimal model specification. The period from 1992-2019 to evaluate the model performance. The need for this sample splitting scheme is discussed in the following methodology section.

---

<sup>6</sup> This normalization is done separately for the training and testing data-sets

### 3.2.2 Sample Splitting and Cross Validation

The ANN models have many trainable parameters, which allows them extremely high degrees of freedom compared to linear regression. This makes over-fitting a valid concern and an essential consideration in model training. To mitigate this concern, I divide my sample across time into three sub-samples, each for model training(1963-1986), validation(1987-1991), and testing(1992-2019).

A pitfall in training an ANN is that the model may over-fit the noise in the training data. This means that the model may memorize a function specific to the training data that does not generalize to unseen observations. To ensure that the model is learning a generalized underlying process, I select a model that offers the best predictive accuracy on the validation data (i.e., select optimal hyper-parameters  $h^*$  that maximizes prediction accuracy for the validation data). This step includes running several models to see which model offers the best OOS performance for the validation period 1987-1991. The performance on the validation data suffers from selection bias and cannot gauge true model performance. Once a model is selected, I evaluate the model performance over the testing period 1992-2019. This data is not included in model training or selection and hence serves as a truly OOS sample. I refer to this training scheme as "fixed estimation".

For cross-validation purposes, I employ another training scheme for the selected model that includes recent observations. Once a model is selected, I re-train a new model with hyper-parameters  $h^*$  every year, using all the previous years. For example, for the year 1992, I train a model on the years 1963-1991. For the year 1993, I train a model again using data from 1963-1992 and so on. This ensures that training data from a specific period do not drive the model performance. I refer to this scheme as "incremental training" in this thesis.

### 3.2.3 Dimensionality Reduction of the Macro-economic State

To extract a lower-dimensional representation of the high-dimensional macroeconomic state, I use a denoising autoencoder.

For the 124 dimensional Fred MD macro-economic time series, I extract a 10-dimensional representation of the economic state using a denoising autoencoder. Figure

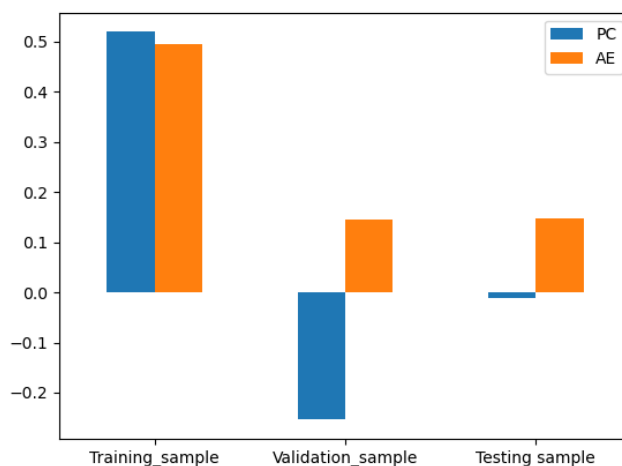


Figure 3.6: **Dimensionality Reduction Auto-encoder and PCA Reconstruction Accuracy**

The figure reports the reconstruction R-squares of 124 macro-economic variables from 10 encoded dimensions compressed using PCA and auto-encoder(AE). Whereas the PCA fails to generate any reconstruction OOS, an auto-encoder captures the variance in both validation and testing periods.

3.6 presents the reconstruction  $R^2$  from my autoencoder model as well as PCA. The PCA fails to generate any reconstruction of the original data OOS. In comparison, autoencoders are able to reconstruct almost 20 percent variance in the data. Unreported results show that for return prediction purposes, this is the only variation we need and a lower dimensional macro-economic state helps with the model convergence<sup>7</sup>. The resulting encoded macro-economic state from the autoencoder is the input  $I_t$  in the prediction model.

### 3.2.4 Digesting the “Anomalies”

The main assertion of this study rests on one core pillar: a good predictive model. It is only when a model offers (to some degree of satisfaction) a generalization of the underlying process that we can take its output to the operating table. ANN’s offer us that generalization of the data as suggested by an unprecedented OOS predictive performance over 28 years of unseen data. This superior OOS predictive performance is attributed to the exceptional ability of ANN’s to estimate non-linear prediction functions in high-dimensional settings<sup>8</sup>. In other words, the ANN model captures structures in the data that persist OOS for around

<sup>7</sup> as suggested by the identical performance of two models, one given the 124 macroeconomic data and one given the 10-dimensional encoded representation of that data.

<sup>8</sup> see for example, Gu, B. Kelly, and Xiu 2020b, Chen, Pelger, and J. Zhu 2019, Avramov, Cheng, and Metzker 2020

three decades, hence is less prone to data-mining bias.

Training an ANN is an elaborate undertaking. This is because the model needs to be optimized for high accuracy and low over-fitting. To understand this study's main discourse, such implementation details are perhaps not as crucial<sup>9</sup>. It is the predictive accuracy, the resulting profitability, and the fact that the model holds over large OOS periods, that are important to establish the goodness of the model. The following equations lay out the fundamental problem of expected return estimation and where to conceptually place ANN models in that context.

Expected returns approximation is essentially a prediction function of the form:

$$E_t(r_{(i,t+1)}) = g^*(X_{(i,t)}, I_t) \quad (3.5)$$

where  $E_t(\cdot)$  is the expectation at time  $t$ .  $r_{(i,t+1)}$  is the excess return for asset  $i$  at time  $t + 1$ .  $g^*(\cdot)$  is a real-valued deterministic function that provides the conditional expected returns given the conditioning information. This function takes a firm-level input vector  $X_{(i,t)}$ , and a macro-level input vector  $I_t$ , both reported at time  $t$ .

From an empirical standpoint we write equation 3.5 as:

$$r_{(i,t+1)} = g^*(X_{(i,t)}, I_t) + \epsilon_{(i,t+1)} \quad (3.6)$$

where  $\epsilon_{(i,t+1)}$  are the forecast errors realized at  $t + 1$  given by .

$$\epsilon_{(i,t+1)} = |r_{(i,t+1)} - \hat{r}_{(i,t+1)}| \quad (3.7)$$

and,

$$\hat{r}_{(i,t+1)} = g^*(X_{(i,t)}, I_t) \quad (3.8)$$

I estimate the conditional expectations function  $g^*(\cdot)$ , using the ANN function  $\psi(\cdot)$ , hence rewriting the empirical equation as:

---

<sup>9</sup> For replication of the results, the implementation details are provided in the appendix

$$\hat{r}_{(i,t+1)}^\psi = \psi(X_{(i,t)}, I_t) \quad (3.9)$$

In this study, the expected returns estimate is the forecast  $\hat{r}_{(i,t+1)}^\psi$ , approximated using the ANN model  $\psi(\cdot)$ , with inputs  $X_{(i,t)}$  and  $I_t$ .

### 3.2.5 Artificial Neural Network Regression

The excess return prediction function takes two input: a firm level input  $X_{(i,t)}$  of 54 dimensions and a macro level input  $I_t$  of 10 dimensions<sup>10</sup>. I denote the combination of the input vectors  $X_{(i,t)}$  and  $I_t$  as  $z_{(i,t)}$ .

To obtain the expected returns estimate  $\hat{r}_{(i,t+1)}^\psi$ , I use a Feed Forward Network(FFN) as a regression function with four hidden layers<sup>11</sup>. The following equations provide a map from the inputs  $z_{(i,t)}$  to the output  $\psi(z_{(i,t)})$ , in terms of hidden layer outputs  $Y_l$ , the weights  $W_l$ , and biases  $b_l$ , for layers  $l$ .

$$\hat{r}_{(i,t+1)}^\psi = \psi(z_{(i,t)}) = \lambda(Y_5 W_5 + b_5) \quad (3.10)$$

$$Y_5 = \kappa(Y_4 W_4 + b_4)$$

$$Y_4 = \kappa(Y_3 W_3 + b_3)$$

$$Y_3 = \kappa(Y_2 W_2 + b_2)$$

$$Y_2 = \kappa(Y_1 W_1 + b_1)$$

$$Y_1 = \kappa(z_{i,t} W_0 + b_0)$$

where  $W_0 \in R^{64 \times 64}$ ,  $W_1 \in R^{64 \times 32}$ ,  $W_2 \in R^{32 \times 16}$ ,  $W_3 \in R^{16 \times 4}$ ,  $W_4 \in R^{4 \times 4}$ , and  $W_5 \in R^{4 \times 1}$  are model parameters  $\Theta_h$  that are optimized using Adam optimizer to find optimal parameters  $\Theta_h^*$ .  $\kappa(\cdot)$  and  $\lambda(\cdot)$  are activation functions for each layer<sup>12</sup>.

<sup>10</sup> encoded representations extracted using a denoising autoencoder

<sup>11</sup> I refer to the same model as ANN regression throughout the thesis

<sup>12</sup> In my estimation I choose the activation functions  $\kappa(\cdot) = ELU(\alpha = 0.3)$  and  $\lambda(\cdot) = Tanh$

### 3.2.6 Model Selection and Evaluation

Training an ANN is a complex task involving a multitude of modeling choices. The model contains many moving parts(hyper-parameters) that can affect the model performance. These include the number of neurons in each layer, the number of layers, batch size, learning rate, etc. Usually, these hyper-parameters are selected using a trial and error approach since there limited theoretical guidance as to what would work for a particular problem. This presents an optimization problem in model training where the underlying function is unknown.

To formally lay out the process of model selection, consider the following representation of equation 3.10:

$$\hat{r}_{(i,t+1)}^\psi = \psi(z_{(i,t)}; \Theta_h) \quad (3.11)$$

where  $\Theta_h$  are the parameter given the hyperparameter set  $h$ . The objective for training the ANN function  $\psi(\cdot)$  is to find optimal parameters  $\Theta_h^*$  that minimize the loss  $L(\epsilon; \Theta_h)$ <sup>13</sup> given by:

$$\Theta_h^* = L(\epsilon; \Theta_h) \quad (3.12)$$

where,  $\epsilon = \frac{\epsilon_{(i,t+1)}}{N}$ , and  $N$  is the total number of training observations.

To find the optimal parameters  $\Theta_h^*$ , I use a variation of the Stochastic Gradient Descent called the Adam optimizer. This process is what is generally referred to as "training" the model. So far, we have only discussed the optimization of the model weights and biases that allow the model its degrees of freedom. However, the optimal convergence of the Adam optimizer depends on the choice of hyper-parameters  $h$  for both the optimizer<sup>14</sup> and the model. For this "black-box" optimization to find the optimal hyper-parameters  $h$ , I use Bayesian optimization. Bayesian optimization refers to a sequential strategy to find global minimums of black-box functions and is well suited to situations where sampling is

---

<sup>13</sup> I define loss as the  $l1$  distance between the observed returns  $r_{(i,t+1)}$  and predicted returns  $\hat{r}_{(i,t+1)}$ , i.e. the mean absolute error.

<sup>14</sup> like learning rate, first-moment decay, second-moment decay, etc.

expensive.

I use Bayesian optimization to obtain a set of optimum hyper-parameters  $h^*$  defined as:

$$h^* = L(\epsilon; h) \quad (3.13)$$

Since the function  $L(\epsilon; h)$  is unknown and the sampling is expensive (i.e. requires fully training an ANN), I use Bayesian optimization with a Gaussian process posterior to solve for the optimum hyper-parameter set  $h^*$ . This concludes the model selection process.

Once I have the optimal hyper-parameters  $h^*$ , I train an ensemble of 5 models for my predictions. This means that I train 5 models using  $h^*$  and the final output is the average of the individual outputs of the 5 models. This technique is called ensemble learning and is useful when dealing with noisy estimations. The final estimate  $\hat{r}_{(i,t+1)}^\psi$  can thus be written down in its final form as:

$$\hat{r}_{(i,t+1)}^\psi = \frac{1}{5} \sum_{e=1}^5 \psi_e(z_{(i,t)}) \quad (3.14)$$

Once model predictions are obtained, I evaluate the predictive accuracy following the literature<sup>15</sup> defined as:

$$R_{oos}^2 = 1 - \frac{\sum_{(i,t) \in \tau} (r_{(i,t+1)} - \hat{r}_{(i,t+1)}^\psi)^2}{\sum_{(i,t) \in \tau} r_{(i,t+1)}^2} \quad (3.15)$$

where  $\tau$  indicates that the model fit is only assessed on a testing sub-sample that is not used in either training or tuning the model.  $R_{oos}^2$  pools prediction errors across all testing time periods and firms. One notable aspect in 3.15 is that the denominator is not demeaned. The underlying idea is that since a naive prediction of zero outperforms the mean, demeaning would add noise to the predictive accuracy.

<sup>15</sup> see for example: Gu, B. Kelly, and Xiu 2020a, Chen, Pelger, and J. Zhu 2019, Kozak 2019

### 3.3 Conclusion

This section covers two main topics. First, I provide a basic introduction to ANN's for a finance audience. Second, I provide a detailed explanation of my methodology. The main purpose of this chapter is to aid in replicability of my results. Given that this research is still in its early stages and the results have extremely important fundamental implications, I find replicability a key consideration. Although I have had similar results for other more elaborate data-sets used Gu, B. Kelly, and Xiu 2020b, I have chosen an alternate return predictive signal data-set from Kozak 2019, which in my opinion is more compact and clean. The main advantage of my methodology is that it does not require high performance computers to replicate the results. My hope is that future researchers in the field find this chapter a helpful stepping stone.



# CHAPTER 4

## DEEP RETURN PREDICTABILITY AND MARKET SENTIMENT

### 4.1 Introduction

Does return predictability arise due to market sentiment or due to underlying risk factors? This question has had fundamental importance in asset pricing ever since the anomaly literature started. With the advancements in quantitative modeling, we are better equipped to provide a comprehensive answer to this question. In this chapter, I use an Artificial Neural Network (ANN) regression model from the previous chapter to estimate conditional expected returns<sup>1</sup>, given conditioning information from 54 firm-level predictive characteristics and a latent representation of the macro-economic state. Using this novel estimate of expected returns, I argue that market sentiment-driven mispricing plays a substantial role in the observed cross-sectional return predictability in the short-run.

Market efficiency implies that returns should be unpredictable in equilibrium. The underlying assumption in this implication is that only unforeseeable information shocks lead to price movements. This was a widely held belief in academic circles up until the 1980s when the first anomalies started to appear in the literature<sup>2</sup>. Although a vast number of return

---

<sup>1</sup> Anomaly predictive signals provide conditioning information to form our conditional expectations of returns. Expected returns are measured as the output of the prediction problem. Hence my use of the term "return predictability" refers to the predictive accuracy or the accuracy of the expected return estimate. These terms, however, are void of any implications regarding the source of the predictability.

<sup>2</sup> The size effect of Banz 1981 was among the first "anomalies" documented in the 1980s. Some other prominent anomalies that followed were the "value effect" (Stattman 1980, Rosenberg, Reid, and Lanstein 1985),

predictive signals have been discovered in the four decades since the early anomalies, the literature is divided on what causes returns to exhibit such predictable behavior<sup>3</sup>. The three most popular propositions are risk, mispricing, and data mining. Studies provide compelling evidence for each proposition, and the common understanding is that possibly all three of the proposed reasons lead to the observed predictability. With the pretext that the empirical predictability stems from multiple sources and given the expansive list of proposed predictors, researchers pursue several interesting questions. For example, which predictors are factors and how to identify them (Pukthuanthong, Roll, and Subrahmanyam 2019)? How to have effective statistical barriers against false identification of random patterns in the data (Harvey and Liu 2015)? Is part of the anomaly performance due to mispricing (McLean and Pontiff 2016)? Do anomaly returns partially reflect mispricing channeled through market sentiment (Stambaugh, Yu, and Yuan 2012)? Answers to such questions have direct implications for the collective discourse on the sources of return predictability.

My methodology allows me to pursue a direct line of inquiry in studying the sources of return predictability. I make my assertions based on out-of-sample (OOS) predictions while incorporating a vast number of predictive characteristics that include both "factors" and "anomalies". This enables me to consider all three of the proposed reasons simultaneously. I argue the following: *Could we obtain an accurate model of expected returns given all available conditioning variables?* Well, yes, although it is a challenging problem<sup>4</sup> but an empirical problem nonetheless. This empirical challenge presents itself as a modeling opportunity for modern quantitative methods<sup>5</sup>. It is an extremely well-suited problem to deep learning models (Gu, B. Kelly, and Xiu 2020b). The follow-up question that I ask is: *"Does market sentiment explain patterns in the overall predictability of this model?"* For return predictability to be driven at least partially by corrections to sentiment-induced mispricing, the corrections should follow the market sentiment in the right direction.

---

the "leverage effect" (Bhandari 1988), and the Earning to Price Ratio effect (Ball 1978) among others.

<sup>3</sup> The term "factor zoo" is used for this growing list of return predictor variables

<sup>4</sup> One problem is the multidimensionality itself, and another challenge is the noise. Given that most of the variation in returns is generated by information shocks, there is added noise in the predictive signals with very high noise-to-signal ratios.

<sup>5</sup> Recently, a novel stream of literature has started to explore high-dimensional modeling approaches (See for example Kozak, Nagel, and Santosh 2020, B. T. Kelly, Pruitt, and Su 2019).

In the previous chapter, I use an ANN regression to approximate a model of firm-level expected returns. In this chapter, I use the approximated expected returns to gain insights into the fundamental sources of return predictability. I make my arguments about misvaluation based on these expected returns and their interaction with market sentiment. In other words, I explore the sentiment-related distortions in expected returns, where expectations are conditioned on multiple conditioning variables simultaneously. My main results are the following:

Firstly, I show that in addition to positive returns, losses (negative returns) are also predictable over long OOS periods. This violates the positive risk premium assumption of rational-expectations-based asset pricing models. As Merton 1980 notes: *"in estimating models of the expected return on the market, the non-negativity restriction of the expected excess return should be explicitly included as part of the specification"*. Furthermore, Boudoukh, Richardson, and Smith 1993 point out that a positive risk premium on any asset with positive betas is a necessary condition for a mean-variance efficient market portfolio. This, they note, is implied directly by the CAPM (and indirectly by multi-factor models). The predictability implied by risk should be strictly limited to positive returns, which seems to be violated in the case of deep learning predictive models. Another puzzling aspect of results is the extremely high profitability implied by the ANN model. With monthly 6-factor alphas around 2%, this profitability is hard to justify as compensation for risk.

Secondly, I present evidence that the accuracy in approximating expected returns follows a market-wide investor sentiment. I find that cross-sectional predictability in negative returns is mainly driven by periods following high market sentiment and the positive return predictability is higher following low sentiment periods. The resulting expected returns imply a highly profitable trading strategy which is not explained by multi-factor models (with up to six factors). This is in line with studies like Baker and Wurgler 2006, which argue for a market-wide sentiment that drives prices in the short-run. Moreover, my results are robust to three additional measures of market sentiment. First, I use the aligned investor sentiment index. Huang, Jiang, et al. 2015 propose a return 'aligned' index to capture the market sentiment based on the same variables as Baker and Wurgler 2006. Also, I use two sentiment

measures from the options market, the put-to-call trading volume ratio (PCR) and the implied volatility (VIX)<sup>6</sup>. I find that higher PCR and higher VIX "Fears" index are followed by high predictive accuracy in negative returns. This is consistent with my main argument that corrections to mispricing partially drive much of the monthly horizons' observed return predictability.

Thirdly, I find a higher degree of overvaluation than undervaluation, which is consistent with arbitrage asymmetry. As pointed out by Miller 1977, constraints to short selling significantly hamper an arbitrageur's ability to correct for overpricing, resulting in arbitrage asymmetry. My results are consistent with this arbitrage asymmetry in that the profitability implied by overpricing is twice that implied by underpricing. These results are closely related to the finding of Stambaugh, Yu, and Yuan 2012 who find that anomaly-based long-short strategies are more profitable following periods of high sentiment.

Finally, I find that this mispricing is short-lived and is arbitrated away in the longer horizons. This is firstly indicated by the fact that deep return predictability<sup>7</sup> in negative returns vanishes in longer horizons. In addition, the puzzling high profitability implied by the deep learning models significantly decreases in longer horizons. The long-only portfolios from longer-term estimations are barely able to beat the market. In contrast, the long-only portfolio from monthly estimation beats the market by roughly a factor of 55 in terms of cumulative excess returns over 28 years of OOS period.

The results have important implications for our understanding of systemic misvaluations and deep return predictability. I argue that the short-term predictability in stock returns is observed at least partially due to corrections in sentiment-driven systemic misvaluations. This is perhaps most closely related with the overpricing argument of Stambaugh, Yu, and Yuan 2012, who show that the short legs of 11 anomaly portfolios are driven by investor sentiment. The key difference in my findings is that I present evidence supporting both underpricing as well as overpricing. I find that short-lived underpricing is a strong prospect given the profitability from my model's long leg. Despite the underpricing, I find evidence for asymmetric effects of arbitrage due to short-sale constraints. Another study that re-

---

<sup>6</sup> see for example, Bandopadhyaya, Jones, et al. 2008

<sup>7</sup> I use the term "deep return predictability" to refer to the predictions from ANN regressions.

lates closely to this study is McLean and Pontiff 2016, who share similar conclusions about systemic misvaluations using high dimensional predictor data. This study is the first to my knowledge that explores the direct relation between market sentiment and deep return predictions. The novel methodology allows exploration of this fundamental question in previously uncharted domains. My approach allows the inclusion of multiple characteristics, provides accurate generalizations OOS and deals directly with firm-level expected returns in addressing the research question. The OOS predictive performance on almost three decades of unseen data is novel to this literature and adds to the model's reliability.

These findings also have implications for the financial machine learning literature that models expected returns with deep learning regressions. Some of the recent machine learning papers attribute this predictability to the stochastic discount factor (Chen, Pelger, and J. Zhu 2019, Kozak, Nagel, and Santosh 2020). On the contrary, I find that sentiment-driven mispricing is a strong driver for the observed return predictability. My results are consistent with those reported in Avramov, Cheng, and Metzker 2020, who find that incorporating economic restrictions in deep learning models significantly decreases the resulting profitability.

This chapter is organized as follows: In the motivation section, I lay out the problem of measuring expected returns using deep learning models and what properties one should expect from the resulting predictions if they were driven by mispricing. Given these properties, I develop my hypotheses on the interaction of market sentiment and mispricing-driven return predictability in the presence of arbitrage asymmetry<sup>8</sup>. The results section presents the empirical results, and the following section concludes.

## 4.2 Motivation

The return predictive variables are referred to as anomalies because they were initially presented as violations of what we (then) understood to be risk. Before the discoveries, the notion of return predictability was wholly dismissed. However, upon such a discovery, the first question that naturally arises is *"why is return predictability observed in the first*

---

<sup>8</sup> The implementation details are provided in the second chapter of my thesis

*place?*”. The literature provides three distinct reasons for such observations. One, the predictive signals represent varying exposure to multidimensional risk. Two, the signals are noisy proxies for some market-wide misvaluations that arise due to a market-wide sentiment. Predictability is observed as a consequence of corrections to such misvaluations. Three, the signals are statistical illusions that arise as a result of extensive data mining. One can further classify these three based on the agents under consideration. The argument of data mining brings the behavior of the econometrician under consideration, citing the somewhat evolutionary tendencies in humans to identify patterns where there are none. Similarly, the debate on mispricing and/or risk is essentially a debate on investor behavior. Mispricing and risk are more closely related arguments in that aspect, and hence they take the spotlight in my discourse on predictability.

Risk is the only plausible explanation for return predictability, assuming that investors are rational and that arbitrage opportunities are rare. Any predictability that does not correspond with a risk factor should be arbitrated away. This point of view is encouraged further when we consider the substantial common variance in predictors, suggesting common underlying drivers to this zoo of characteristics. It is quite possible that two or more predictors give similar information about returns. One can argue that this common information relates to risk<sup>9</sup>. This argument lies at the core of the risk literature, starting from the earlier “anomaly digestion exercise” of Fama and French 1993, to the more recent work of Kozak, Nagel, and Santosh 2018 and B. T. Kelly, Pruitt, and Su 2019.

From the behavioral perspective, a variety of sociological and psychological factors may manifest in the data. The literature has documented various patterns in the market that coincide with seemingly irrational human behaviors otherwise perfectly explainable. For example, consider the puzzle that the sin stock premium presents (Fabozzi, Ma, and Oliphant 2008). Why would investors demand a premium on sin stocks given a rational objective of wealth maximization? For behaviorists, this seemingly anomalous pattern has a convenient explanation that it represents an underlying social stigma. Regarding the common variation

---

<sup>9</sup> Kozak, Nagel, and Santosh 2018 make the same point when they note about factor models that “*These (factor) models are reduced-form because they are not derived from assumptions about investor beliefs, preferences, and technology that prescribe which factors should appear in the SDF. What interpretation should one give such a model if it works well empirically?*”

in predictive signals, one can always argue that it is driven by a wide market sentiment that leads to systemic over or undervaluations<sup>10</sup>. Investor optimism and pessimism lead to asset values deviating from their fundamental values. From the mispricing perspective, return predictability is observed because of predictable corrections to mispriced securities, which is what I argue in this thesis.

The existence of predictive signals and their common variance presents an empirical reality. There exists a difference in conditional mean returns, and such differences could be exploited to formulate profitable trading strategies. It is this cross-sectional difference that shows up in both predictive regressions as well as portfolio sorts<sup>11</sup>. From a return predictability perspective, a challenge that arises from this common variation in predictors is the multicollinearity when using linear regression models. An ANN model, however, is robust to high dimensionality and multicollinearity in the estimation problem. A prediction model's objective is to get the highest predictive accuracy, regardless of whether it comes from common or idiosyncratic variation in the predictor set. Higher accuracy would mean that the model explains a higher portion of the observed cross-sectional variation in firm-level returns. Predictive regressions that offer high OOS predictability essentially capture a stable function for the expected returns.

#### 4.2.1 Investor Sentiment, Arbitrage Asymmetry, and Expected Returns

Many studies in the literature have considered the possibility that a market-wide sentiment can cause prices to deviate from their fundamental values. This understanding is deeply rooted in psychology and sociology and an intuitively understandable tenant of human behavior. On the other hand, there is the classical argument of no-arbitrage. As long as some investors with access to cheap credit are rational enough to identify this mispricing, they can profit from such opportunities while correcting the market back to the fundamentals. If market sentiment leads to misvaluations that arbitrageurs eventually correct, such corrections

---

<sup>10</sup> Stambaugh and Yuan 2017 for example, construct "mispricing" factors capturing the common variation in anomaly returns as approximations of misvaluation.

<sup>11</sup> A predictive model is equivalent to the popular characteristic portfolio sorts as noted by Cochrane 2011: *"Portfolio sorts are the same thing as non-parametric cross-sectional regressions, using non-overlapping histogram weights"*

should be predictable given market sentiment. In other words, If predictive characteristics partially reflect mispricing, then the predictability in returns should vary with the market sentiment. Additionally, arbitrageurs face constraints against short-selling, making overvaluations more difficult to correct (compared to undervaluations). Given this asymmetric exposure to arbitrage, the sentiment effect in predictability should be higher in negative returns (corrections for overvaluation) than positive returns (corrections for underpricing). Alternatively, in the case of solely risk-driven predictability, negative returns should not be predictable at all to be consistent with strictly positive risk premiums. Besides, no relationship between market sentiment, arbitrage asymmetry, and cross-sectional predictability should exist.

To guide this empirical inquiry, I propose three hypotheses. The first hypothesis is that negative returns are predictable. If the expected returns reflect mispricing, then corrections for overpricing would make negative returns predictable. Negative returns should not be predictable if only risk leads to cross-sectional predictability (assuming positive risk premiums).

The second hypothesis is that the predictability in negative returns is higher following periods of high market sentiment. The predictability in positive returns is higher following periods of low sentiment. This is because high sentiment periods lead to systemic overvaluations that make negative returns (downward corrections) more predictable and vice versa. If solely risk underlies predictability, then we should not observe such a pattern in predictive accuracy.

The third hypothesis is that a long-short strategy based on extreme deciles of expected returns is more profitable following high sentiment periods. Short-sale constraints would imply that overvaluation is more prevalent than undervaluation. This disparity should be observed in the long-short profitability of high and low sentiment periods. Suppose the underlying driver of predictability is mispricing, and overvaluations tend to be more prevalent due to short-sale impediments. In that case, a strategy that exploits overvaluation should be more profitable than exploiting undervaluation.

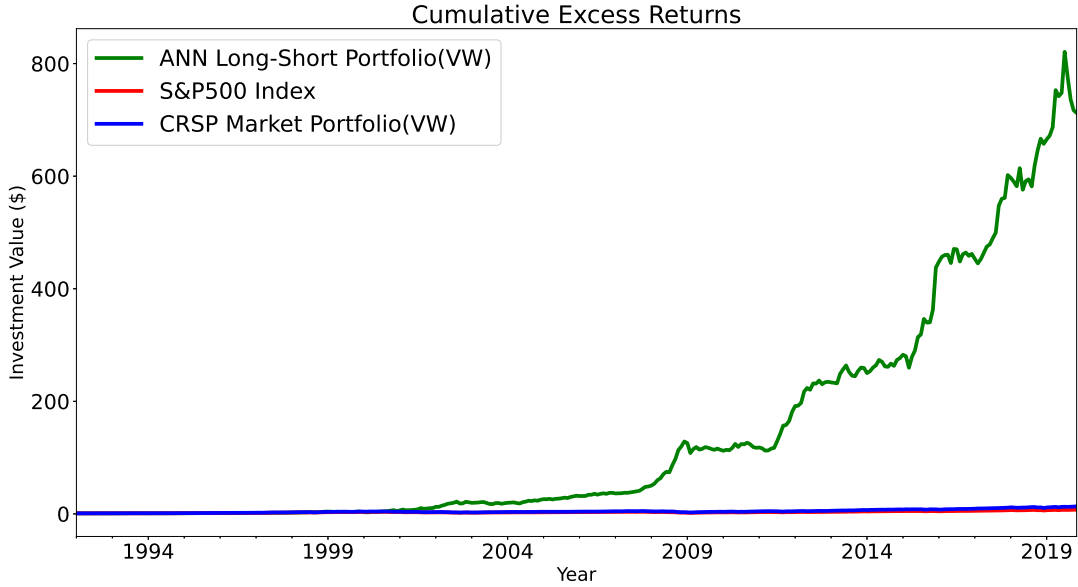
My results support mispricing driven by investor sentiment, the eventual correction to



**Table 4.1: Firm Level Predictive R-Squares**

The table shows the predictive R-squares for training and validation periods (calculated as equation 3.15). In addition, the out-of-sample predictive accuracy  $R_{oos}^2$  for the testing period is reported for both incremental and fixed training schemes. The training period spans years 1963-1986, the validation period spans 1987-1991, and the testing period spans the years 1992-2019

	Training	Validation	Testing
<b>Fixed Estimation</b>	13.63%	10.03%	6.60%
<b>Incremental Estimation</b>	—	—	6.93%



**Figure 4.1: ANN Prediction Based Long-Short Portfolio Performance**

This figure plots the OOS performance of the ANN VW long-short portfolio (green) with that of SP500 index (red) and a VW market portfolio (blue). The profitability of the long-short portfolio is incredibly high compared to a market benchmark, with \$1 invested in 1992 going up to more than \$700 at the end of 2019.

fundamentals, and the asymmetric effects of arbitrage on misvaluation given short-sale constraints.

## 4.3 Results

### 4.3.1 Predictive Accuracy and Profitability

In my analysis, the  $R_{oos}^2$  comes out to be 6.6%. In other words, the model can predict more than  $1/20^{th}$  of the variability in future returns compared to a benchmark prediction of zero. Although this is a tiny portion of the total variability in  $r_{(i,t+1)}$  nonetheless, it shows that future returns are predictable over long out-of-sample periods. I train the model in equation

3.9 using a fixed scheme and incremental scheme as explained in the methodology section. This is to cross-validate the model and mitigate concerns that the results are driven by a particular estimation period. Additionally, the model does not learn from recent data in the fixed estimation, whereas incremental estimation incorporates observations until the most recent year before the prediction year. In both estimation schemes, the predictions are out-of-sample (i.e., the prediction data is never included in the model training or validation). Table 4.1 provides the in-sample and out-of-sample predictive accuracy of the ANN model. Although the predictive accuracy is higher for training and validation sub-samples, what is noteworthy is that the model retains a big chunk of its predictive accuracy in the testing period.

This seemingly small percentage of explained variance can lead to a highly profitable long-short strategy that outperforms the market by a factor north of 50, yielding a 6-factor monthly alpha of 2%. Table 4.2 provides the average returns and Sharpe ratios of decile portfolios based on prediction deciles. Every month, each firm  $i$  is allocated into a decile based on the predicted returns  $\hat{r}_{(i,t+1)}^\psi$ , predicted at time  $t$ . For each decile every month, value-weighted(VW) and equal-weighted(EW) portfolios comprise all the firms in each decile. Table 4.2 shows the OOS time-series averages and Sharpe ratios of realized returns for each decile portfolio in the subsequent month  $t + 1$ . Despite the low  $R_{OOS}^2$ , the decile portfolios exhibit a monotonic decreasing pattern from top to bottom deciles. The decile with the highest predictions  $\hat{r}_{(i,t+1)}^\psi$ , have the highest average returns and Sharpe ratios. This pattern is evident for both incremental and fixed estimations, and these results hold for both equal-weighted (EW) and value-weighted (VW) portfolios. The last row of table 4.2 presents the ANN decile-based long-short portfolio's average monthly returns. The incremental estimation yields a VW average long-short return of 2.09% and a Sharpe ratio of 1.23. This is slightly higher compared to the mean long-short return of 1.86% and Sharpe ratio of 1.04. However, the long-short portfolios from both these estimations beat a portfolio of all stocks in my sample by a wide margin in terms of both mean returns and Sharpe ratios.

Figure 4.1 plots the cumulative excess returns for the ANN decile long-short portfolio for the testing period along with the S&P500 and VW CRSP market portfolio. The long-short strategy's

**Table 4.2: Monthly Average Returns and Sharpe Ratios of Decile Portfolios**

The table reports the time series mean returns and Sharpe ratios for equal weighted (EW) and values weighted decile portfolios for the testing period 1992-2019. The deciles are based on the expected return estimation of the ANN regression model  $\hat{r}_{(i,t+1)}^\psi$  estimated at time  $t$ . The returns and Sharpe ratios reported are for the subsequent month  $t + 1$ .

<b>Panel A. Incremental Estimation</b>				
	<b>Mean (EW)</b>	<b>Mean (VW)</b>	<b>SR(EW)</b>	<b>SR(VW)</b>
<b>All Stocks</b>	0.01*** (3.01)	0.01*** (3.17)	0.57	0.60
<b>Top</b>	1.60*** (6.58)	1.61*** (6.30)	1.25	1.19
<b>D9</b>	1.28*** (5.68)	1.22*** (5.37)	1.08	1.02
<b>D8</b>	1.22*** (5.26)	1.13*** (4.87)	1.00	0.92
<b>D7</b>	0.96*** (4.17)	0.80*** (3.53)	0.79	0.67
<b>D6</b>	0.90*** (3.84)	0.72*** (2.92)	0.73	0.55
<b>D5</b>	0.67*** (2.77)	0.51** (2.12)	0.53	0.40
<b>D4</b>	0.52** (2.21)	0.41 (1.63)	0.42	0.31
<b>D3</b>	0.43 (1.55)	0.30 (1.07)	0.29	0.20
<b>D2</b>	0.09 (0.30)	-0.12 (-0.39)	0.06	-0.07
<b>Bottom</b>	-0.44 (-1.25)	-0.48 (-1.32)	-0.24	-0.25
<b>Top-Bottom</b>	2.05*** (7.04)	2.09*** (6.50)	1.33	1.23

<b>Panel B. Fixed Estimation</b>				
	<b>Mean (EW)</b>	<b>Mean (VW)</b>	<b>SR (EW)</b>	<b>SR (VW)</b>
<b>All Stocks</b>	0.01*** (3.01)	0.01*** (3.17)	0.57	0.60
<b>Top</b>	1.57*** (6.88)	1.63*** (6.81)	1.30	1.29
<b>D9</b>	1.22*** (5.57)	1.17*** (5.31)	1.06	1.01
<b>D8</b>	1.06*** (4.79)	0.95*** (4.21)	0.91	0.80
<b>D7</b>	0.94*** (4.15)	0.85*** (3.83)	0.79	0.73
<b>D6</b>	0.84*** (3.68)	0.61*** (2.62)	0.70	0.50
<b>D5</b>	0.69** (2.86)	0.47* (1.92)	0.54	0.36
<b>D4</b>	0.66** (2.55)	0.57** (2.23)	0.48	0.42
<b>D3</b>	0.43 (1.54)	0.32 (1.10)	0.29	0.21
<b>D2</b>	0.12 (0.35)	-0.09 (-0.29)	0.07	-0.06
<b>Bottom</b>	-0.26 (-0.69)	-0.23 (-0.61)	-0.13	-0.12
<b>Top-Bottom</b>	1.83*** (5.83)	1.86*** (5.47)	1.11	1.04

t-values in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01

relatively high profitability is evident from the figure where the S&P500 and VW market returns appear as an almost flat line. \$1 invested in the long-short portfolio in 1992 yields an investment value of over \$700 at the end of 2019. This is tremendously high compared to the market portfolio, which yields a value of about \$13 at the end of 2019. For the ANN predictability to be justified by risk-based explanations, one would also have to justify extremely high risk premiums. Such a description would also imply a highly inefficient market in terms of mean-variance efficiency.

This high profitability survives factor adjustment for up to 6 factors. Table 4.3 provides results for long-short portfolio adjustments for various specifications of single and multi-factor models. The results show that the profitability in the ANN long-short portfolio survives factor adjustment for both estimation schemes. The alphas in both Panel A and B of table 4.3 are very close to the mean long-short returns reported in 4.2 across all model specifications. Also, the long-short returns do not exhibit much risk factor loadings. Assuming that risk models are correctly specified, the pure alpha in the long-short returns points to a source other than risk.

### Short-Term Mispricing

In addition to the high profitability, the predictive accuracy in negative returns presents another puzzle. The existence of negative cross-sectional risk premiums is hard to justify. One possible explanation is that negative risk premiums could exist for securities that provide hedging against some macroeconomic states. However, in such a case, negative returns should be predictable no matter the estimation horizon, which does not seem to be the case in my results.

Consider the first hypothesis that negative returns should be predictable in the short horizons. Table 4.4 presents the  $R_{OOS}^2$  for return predictive ANN regression where returns are defined over different time horizons. For each horizon, I re-estimate the model to predict future excess returns over different future windows. This involves both model selection and training done separately for each horizon. I report results separately for incremental and fixed training regimes and for negative and positive returns. In both estimations, the deep return predictability in negative returns vanishes in longer horizons, with  $R_{OOS}^2$  diving deeply into the negative territory. The positive return predictability is higher for longer horizons which is reconcilable with a risk-based explanation. These results hold for both fixed and incremental estimation schemes. These results show that losses are predictable but over short (monthly) horizons only.

With the predictability in negative returns, the profitability implied by the ANN model also seems

**Table 4.3: Long-Short Portfolio Factor Regressions**

The table reports long-short portfolio regressions on various specifications of risk factor models for the testing period 1992-2019. Panel A provides the long-short portfolios using predictions from incrementally trained model and Panel B reports long-short portfolios from the fixed estimation.

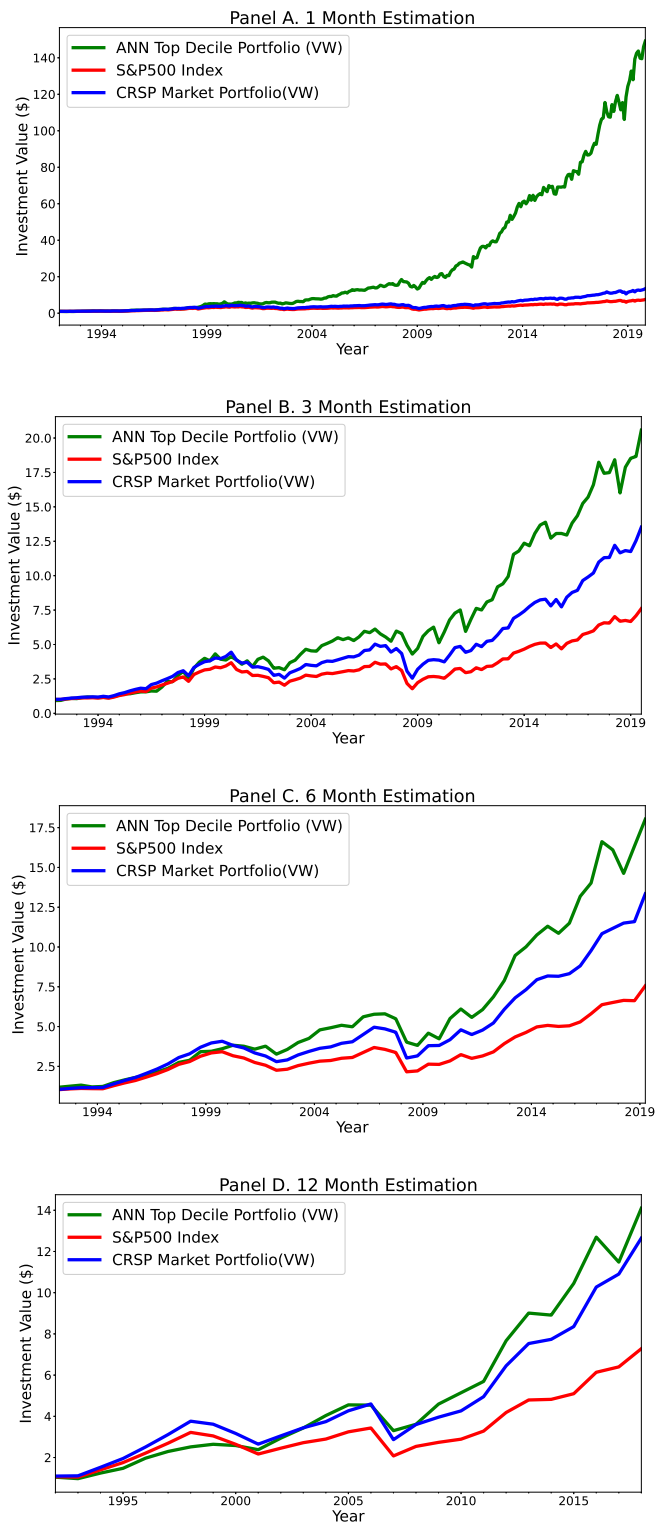
<b>Panel A. Incremental Estimation</b>					
	(1)	(2)	(3)	(4)	(5)
	CAPM	FF3	FF3+MOM	FF5	FF5+MOM
<b>Alpha</b>	<b>2.21***</b>	<b>2.16***</b>	<b>2.29***</b>	<b>1.95***</b>	<b>2.07***</b>
	(0.33)	(0.33)	(0.33)	(0.34)	(0.31)
<b>MKTRF</b>	-0.07	-0.02	-0.08	0.08	0.01
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)
<b>HML</b>		-0.03	0.15	0.03	-0.06
		(-0.03)	(0.11)	(0.15)	(0.14)
<b>SMB</b>		-0.17	-0.15	-0.06	-0.05
		(0.11)	(0.11)	(0.12)	(0.11)
<b>UMD</b>			-0.17**		-0.18***
			(0.07)		(0.06)
<b>CMA</b>				0.18	0.07
				(0.22)	(0.20)
<b>RMW</b>				0.38**	0.35**
				(0.16)	(0.15)
<b>R2</b>	0.20%	2.30%	4.10%	4.00%	6.20%
<b>Adj_R2</b>	-0.10%	1.40%	2.90%	2.50%	4.40%

<b>Panel B. Fixed Estimation</b>					
	(1)	(2)	(3)	(4)	(5)
	CAPM	FF3	FF3+MOM	FF5	FF5+MOM
<b>Alpha</b>	<b>1.88***</b>	<b>1.82***</b>	<b>1.90***</b>	<b>1.66***</b>	<b>1.74***</b>
	(0.35)	(0.35)	(0.35)	(0.36)	(0.34)
<b>MKTRF</b>	0.04	0.09	0.05	0.17*	0.06
	(0.08)	(0.08)	(0.10)	(0.10)	(0.09)
<b>HML</b>		0.23**	0.19	0.12	0.01
		(0.12)	(0.12)	(0.16)	(0.16)
<b>SMB</b>		-0.18	-0.17	-0.05	-0.04
		(0.12)	(0.13)	(0.13)	(0.12)
<b>UMD</b>			-0.10		-0.12*
			(0.08)		(0.07)
<b>CMA</b>				-0.02	0.01
				(0.24)	(0.22)
<b>RMW</b>				0.39**	0.40**
				(0.17)	(0.16)
<b>R2</b>	0.10%	2.10%	2.70%	3.70%	4.40%
<b>Adj_R2</b>	0.00%	1.20%	1.50%	2.20%	2.60%

Standard errors in parentheses.

\* p<.1, \*\* p<.05, \*\*\*p<.01



**Figure 4.2: Profitability for Longer Horizon Estimation**

This figure plots the OOS profitability for longer term estimations of ANN VW Top Decile Portfolio (green) with that of SP500 index (red) and a VW market portfolio (blue). The high profitability of the top ANN decile vanishes for 3, 6, and 12 month estimations.

to disappear. Figure 4.2 plots the cumulative excess returns on the long-only portfolios for 1-month, 3-month, 6-month, and 12-month horizons, along with the cumulative excess returns S&P500 and VW market indices. The profitability gap between the top ANN decile and the market indices shrinks in estimation horizons higher than one month. The significant drop in profitability for longer horizon estimations brings the implied profitability close enough to the market risk premium to be evaluated as compensation for risk.

**Table 4.4: OOS Predictive Accuracy Over Longer Horizons**

This table reports predictive  $R_{oos}^2$  for the testing period 1992-2019, across multiple return horizon. For each horizon, equation 3.10 is re-estimated using ANN regression for different definitions of  $R_{(i,t+1)}$  computed over 1 month, 3 months, 6 months, and 12 months horizons. Both panels report separately the predictive accuracy for positive and negative realized returns  $R_{(i,t+1)}$ .

<b>Panel A. Incremental Estimation</b>				
	<b>1month</b>	<b>3months</b>	<b>6months</b>	<b>12months</b>
<b>Positive Returns</b>	8.44%	12.36%	13.11%	14.49%
<b>Negative Returns</b>	6.06%	-12.48%	-15.43%	-21.93%
<b>All Returns</b>	6.93%	2.23%	2.96%	4.08%

<b>Panel B. Fixed Estimation</b>				
	<b>1month</b>	<b>3months</b>	<b>6months</b>	<b>12months</b>
<b>Positive Returns</b>	5.57%	10.73%	16.91%	15.97%
<b>Negative Returns</b>	7.87%	-10.13%	-20.59%	-23.03%
<b>All Returns</b>	6.60%	2.22%	3.58%	4.82%

Moreover, I find that shorter-term predictability stems from different sources compared to longer-term estimation. To evaluate how much the expected returns estimate  $\hat{r}_{(i,t+1)}^\psi$  loads on a particular predictor, I compare the average of each predictor in ANN deciles. For each predictive signal, I take the absolute difference in the mean predictor values of extreme ANN deciles<sup>12</sup>. Figure 4.3 plots this absolute mean difference for 1 month (blue), 6 month(orange), and 12 month(green) estimations. These results show that 1-month estimation loads on different variables compared to longer-term models. Consistent with Chen, Pelger, and J. Zhu 2019 and Gu, B. Kelly, and Xiu 2020b, I find that 1-month estimation on ANN loads on price trends and liquidity variables such as momentum, reversal, and volatility. For longer-term estimations, the model loads more heavily on fundamentals such as earning, book-to-market, and profitability.

The results presented in table 4.2, table 4.3, and figure 4.1 are for the OOS testing period 1992-2019. These results establish that stable estimates of conditional return means exist that prevail for long OOS periods and yield highly profitable strategies. The results presented in table 4.4, figure 4.3 and figure 4.2 cast doubt on risk as the source of this short term predictability. Next, I discuss

<sup>12</sup>These values are comparable since all the predictive signals are normalized between -1 and 1

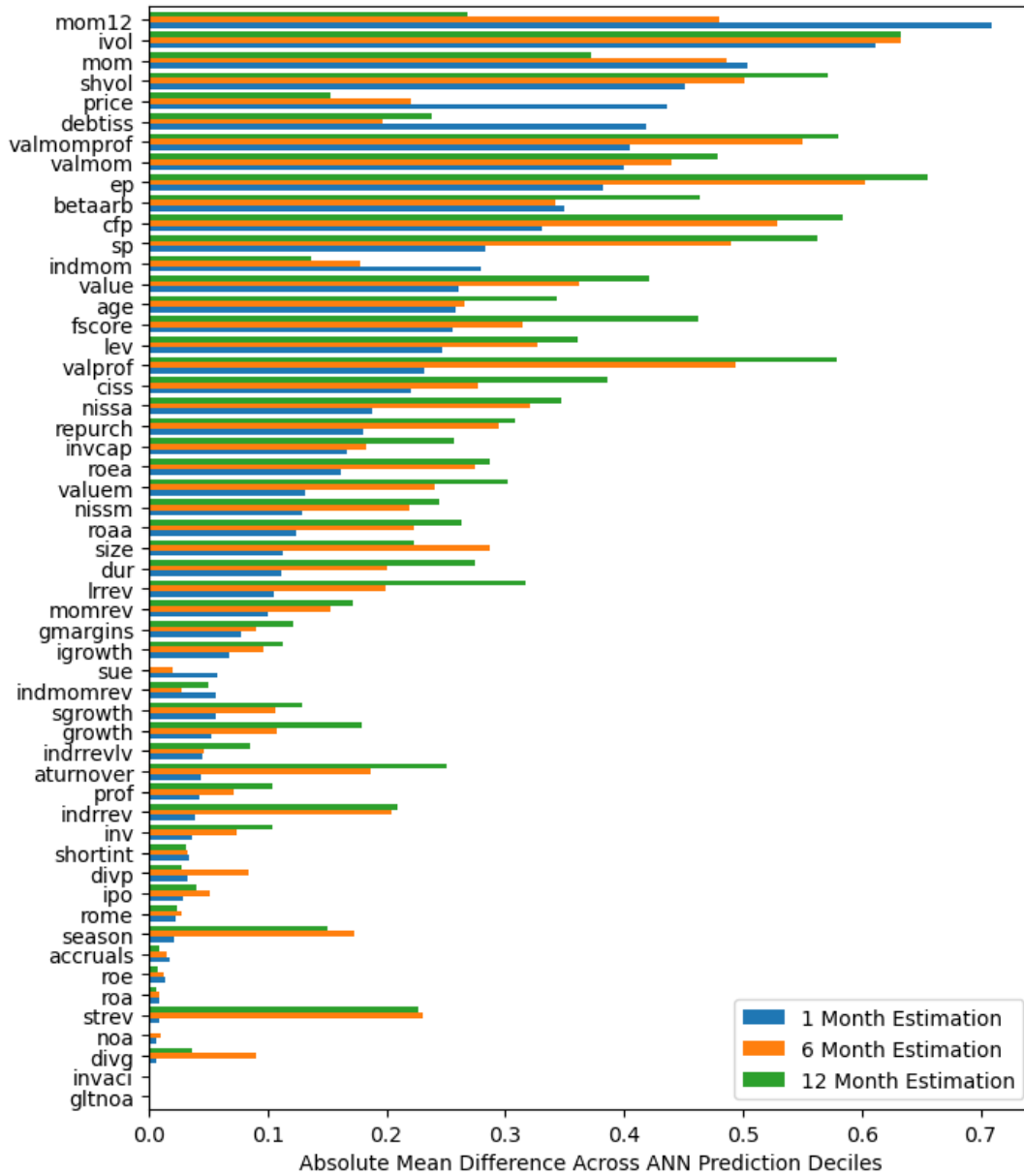


Figure 4.3: **Characteristic Loadings of the ANN Model**

This figure plots the OOS loading of the expected returns  $\hat{r}_{(i,t+1)}^\psi$  on predictive signals  $X_{(i,t)}$ , defined as the absolute difference in the mean of each predictor between extreme prediction deciles. The plot shows that the 1 month estimation loads on different characteristics compared to the 6 and 12 month estimations.



what explains the existence of such structures in return predictability. I argue that corrections to mispricing are a major source for this short-term predictability in returns. Market sentiment plays an important role in the mispricing, hence predicting eventual corrections in the subsequent periods.

### 4.3.2 Market Sentiment and Return Predictability

**Table 4.5: OOS Predictability and Market Sentiment**

This table reports predictive  $R_{oos}^2$  for the OOS period 1992-2014, following high and low sentiment months. The predictive accuracy is reported separately for positive and negative realized returns  $r_{(i,t+1)}$ . Market sentiment is measured using the market based sentiment index of Baker and Wurgler 2006, where a value greater than 0 is considered high sentiment and vice versa. The OOS predictability in negative returns is higher following periods of high sentiment (and vice versa) which is consistent with the argument that at least part of return predictability is observed due to predictable corrections to mispricing.

<b>Panel A. Incremental Estimation</b>		
	<b>Low Sentiment</b>	<b>High Sentiment</b>
<b>Positive Return Predictability</b>	8.7%	4.5%
<b>Negative Return Predictability</b>	0.0%	10.3%

<b>Panel B. Fixed Estimation</b>		
	<b>Low Sentiment</b>	<b>High Sentiment</b>
<b>Positive Return Predictability</b>	5.4%	4.4%
<b>Negative Return Predictability</b>	2.4%	11.7%

If return predictability is observed due to sentiment-driven mispricing, then market sentiment should precede return predictability in the right direction. High sentiment leads to systemic overpricing and consequently negative corrections in the subsequent period due to arbitrage. Similarly, low sentiment leads to underpricing, making positive corrections in the successive period more likely. Table 4.5 reports the predictive accuracy of the ANN model across negative and positive returns, following periods of high and low sentiment where sentiment is defined following Baker and Wurgler 2006.

The results in table 4.5 are consistent with the second hypothesis across both fixed and incremental estimation schemes. The predictability in negative returns is much higher following high sentiment periods. In the incremental estimation reported in Panel A of the table, almost none of the losses are predictable following periods of low sentiment given the  $R_{OOS}^2$  of 0.0%. In fixed estimation, this negative predictability is very low ( $R_{OOS}^2$  of 2.4%) following low sentiment periods. This disparity in negative return predictability across market sentiment is a central prediction of this thesis.

In addition, the same sentiment effects can be observed in positive returns. The predictability in positive returns is higher following periods of low sentiment and vice versa. This effect is more

pronounced for incremental estimation of the model and is weaker overall compared to the sentiment effects in negative return predictability. To complement these results, I use the firm-level mispricing score used in Stambaugh, Yu, and Yuan 2015. Table 4.6 Panel B reports a monotonic decrease in the mispricing score across prediction deciles. The most profitable decile tends to include stocks that are difficult to arbitrage and hence more prone to mispricing. These findings support undervaluation-driven positive return predictability reported in table 4.5.

The disproportionate sentiment effects in negative and positive return predictability are expected given arbitrage asymmetry. Since overpricing is more common compared to underpricing, the sentiment effects in the short-legs of anomaly strategies are expected to be more pronounced. Overall, these results are consistent with mispricing-driven return predictability.

### Market Sentiment and Long-Short Profitability

**Table 4.6: OOS Profitability, Market Sentiment, and Mispricing**

Panel A. reports the time series averages of decile portfolio returns for the OOS period 1992-2014 split across high and low sentiment periods. Panel B. reports the average mispricing score for firms within each decile for the OOS period 1992-2016 following Stambaugh, Yu, and Yuan 2015 .

	<b>Panel A.</b>		<b>Panel B.</b>
	<b>High Sentiment</b>	<b>Low Sentiment</b>	<b>Mispricing Score</b>
<b>All Stocks</b>	0.00	0.01	46.92
<b>Top</b>	1.41	1.83	56.18
<b>D9</b>	1.24	1.22	50.85
<b>D8</b>	1.12	1.17	48.69
<b>D7</b>	0.45	0.95	47.43
<b>D6</b>	0.37	0.75	46.54
<b>D5</b>	0.36	0.46	45.64
<b>D4</b>	-0.24	0.47	44.93
<b>D3</b>	-0.84	0.60	44.02
<b>D2</b>	-1.22	0.20	43.39
<b>Bottom</b>	-1.80	-0.12	42.26
<b>Top-Bottom</b>	3.22	1.95	

The imbalance in overvaluation versus undervaluation could also be observed in the ANN long-short strategy's profitability across high and low sentiment periods. The results reported in table 4.6 panel A support the third hypothesis that profits from the ANN long-short strategy should be higher following periods of high sentiment. The last row of 4.6 panel A shows the time-series average of long-short portfolio returns following high and low sentiment periods. Consistent with arbitrage asymmetry, the long-short profitability following high sentiment periods is almost twice the profitability following low sentiment periods.

Additionally, 4.6 panel A reports that the bottom deciles yield highly negative average returns

following high sentiment periods compared to the average returns following low sentiment periods. This effect is not as strong for the top deciles, where only the top decile produces a higher average. This is also expected given arbitrage asymmetry. The results in 4.6 panel A show that profits from mispricing induced overvaluations disproportionately drive predictability compared to undervaluations.

### Alternative Measures of Market Sentiment

For robustness of the results, I verify the findings in 4.5 using three alternate market sentiment measures. First, I use the "aligned" market sentiment index of Huang, Jiang, et al. 2015. In addition, I use two sentiment measures from the options market proxies for market sentiment, the implied volatility index (VIX) and put-to-call trading volume ratio (PCR). I find that the disparity in prediction accuracy reported in table 4.5 holds for these additional market sentiment measures.

**Table 4.7: OOS Predictability and Return Aligned Sentiment Index**

This table reports predictive  $R_{oos}^2$  for the OOS period 1992-2014, following high and low sentiment months. The predictive accuracy is reported separately for positive and negative realized returns  $r_{(i,t+1)}$ . Market sentiment is measured using the market based monthly sentiment index of Huang, Jiang, et al. 2015, where a value greater than 0 is considered high sentiment and vice versa. The OOS predictability in negative returns is higher following periods of high sentiment (and vice versa) which is consistent with the argument that at least part of return predictability is observed due to predictable corrections to mispricing.

	<b>Panel A. Incremental Estimation</b>	
	<b>Low Sentiment</b>	<b>High Sentiment</b>
<b>Positive Return Predictability</b>	7.6%	5.3%
<b>Negative Return Predictability</b>	2.3%	10.7%

	<b>Panel B. Fixed Estimation</b>	
	<b>Low Sentiment</b>	<b>High Sentiment</b>
<b>Positive Return Predictability</b>	5.6%	3.8%
<b>Negative Return Predictability</b>	4.3%	12.2%

Table 4.7 reports the results for the aligned sentiment index of Huang, Jiang, et al. 2015. They provide an alternate market sentiment measure based on the same market-based indicators as Baker and Wurgler 2006. They use a different dimensionality reduction technique to extract the market sentiment that explains maximum variance in future returns, terming it the aligned market sentiment index. The conclusions from this table are the same as those from 4.5. Higher sentiment periods are followed by higher predictability in negative returns, and low sentiment periods are followed by higher predictability in positive returns. Moreover, the disparity in these sentiment effects between positive and negative returns also holds for this sentiment index.

VIX has been popularly termed as the "Investor Fear Gauge" (see, for example, Whaley 2000).

**Table 4.8: OOS Predictability and Implied Volatility**

This table reports predictive  $R_{oos}^2$  for the OOS period 1996-2019 separated by the implied volatility on S&P100 Index (VIX) in the period preceding the prediction month. The predictive accuracy is reported separately for positive and negative realized returns  $r_{(i,t+1)}$ . A month is considered high VIX sentiment if the median daily PCR in the last 10 days of that month is higher than the sample mean VIX. The OOS predictability in negative returns is higher following periods of high sentiment (and vice versa) which is consistent with the argument that at least part of return predictability is observed due to predictable corrections to mispricing.

<b>Panel A. Incremental Estimation</b>		
	<b>High Volatility</b>	<b>Low Volatility</b>
<b>Positive Return Predictability</b>	6.4%	9.2%
<b>Negative Return Predictability</b>	8.8%	2.1%

<b>Panel B. Fixed Estimation</b>		
	<b>High Volatility</b>	<b>Low Volatility</b>
<b>Positive Return Predictability</b>	5.3%	4.8%
<b>Negative Return Predictability</b>	10.1%	6.2%

It is a forward-looking index that measures the investor's expectation of market volatility over the next month, as implied by the skew of S&P100 options. Similarly, PCR is the daily volume of put options traded as a ratio of daily call volume. PCR captures the options market consensus on the expected market direction in the future. A low PCR would mean that more calls are traded compared to puts, and hence the options market expects the prices to rise and vice versa. These two sentiment measures are unique in that they come from options markets data. If the return corrections due to mispricing are predictable, then the option market data should exhibit consistent behavior.

Table 4.8 reports the predictive accuracy for negative and positive returns following periods of low and high VIX. I define monthly VIX as the median value of the VIX index in the last ten days of the month  $t$ . A period is defined as high volatility if the monthly VIX for that month is higher than the average VIX in the sample. Table 4.8 shows that the predictive accuracy in negative returns is higher following high volatility, which is consistent with the main findings in table 4.5. The results in panel A. report high predictive accuracy in positive returns following low volatility periods. VIX is considered a barometer for investor fear and not investor optimism<sup>13</sup>. Hence, the disproportionate sentiment effect in the predictability of positive and negative returns reported in tables 4.5 and 4.7 is expected to be higher for the VIX.

Similarly, the dichotomy in sentiment effects in predictability is expected to be higher for PCR as well. This is by design since these sentiment measures come from the options market, which is a dedicated market for hedging and arbitrage. Thus these sentiment measures capture overvaluations disproportionately more than undervaluations. To obtain a monthly PCR measure from the daily

<sup>13</sup> see, for example, Whaley 2009

**Table 4.9: OOS Predictability and Put-to-Call Ratio**

This table reports predictive  $R_{oos}^2$  for the OOS period 1996-2019 separated by the put-to-call ratios (PCR) in the period preceding the prediction month. The predictive accuracy is reported separately for positive and negative realized returns  $r_{(i,t+1)}$ . A month is considered high PCR sentiment if the median daily PCR in the last 10 days of that month is higher than the sample mean PCR of 0.85. The OOS predictability in negative returns is higher following periods of high sentiment (and vice versa) which is consistent with the argument that at least part of return predictability is observed due to predictable corrections to mispricing.

<b>Panel A. Incremental Estimation</b>		
	<b>High PCR</b>	<b>Low PCR</b>
<b>Positive Return Predictability</b>	8.1%	6.9%
<b>Negative Return Predictability</b>	8.5%	5.6%

<b>Panel B. Fixed Estimation</b>		
	<b>High PCR</b>	<b>Low PCR</b>
<b>Positive Return Predictability</b>	2.0%	7.3%
<b>Negative Return Predictability</b>	13.2%	6.6%

index, I follow the same scheme as for VIX. A monthly PCR is calculated as the median PCR index value in the last 10 days of the month  $t$ . A month is considered high PCR if the monthly PCR value is higher than the mean PCR in the sample<sup>14</sup>. Table 4.9 reports the predictive accuracy in negative and positive returns. The overvaluation effects are evident in that the predictability in negative returns is higher following periods where more puts are traded than calls. Panel B of table 4.9 reports that for fixed estimation, the predictive accuracy is higher following periods where more calls are traded compared to puts. Overall, I find that my results in 4.5 hold across the other investor sentiment measures that I consider.

## 4.4 Conclusion

Market-wide variations in investor sentiment lead to systemic over and undervaluations for many stocks during that period. As arbitrageurs correct for such misvaluations, the prices revert to the fundamentals. These corrections are partially responsible for the observed cross-sectional predictability in equity returns. With constraints to short-selling, overvaluations become more challenging to eradicate. Hence the mispricing-driven predictability is more pronounced for corrections to overvaluations.

A predictive model that captures a good amount of cross-sectional variability in returns out-of-sample also tends to exhibit trends in predictive accuracy consistent with a mispricing-based explanation. Because overpricing is more prevalent than underpricing, this effect is more evident in negative

<sup>14</sup>Historically, the mean PCR comes out to be around 0.80. I divide my sample across the sample mean of 0.85

returns. Similarly, a long-short portfolio based on prediction deciles shows higher profitability following high sentiment periods than low sentiment periods. These patterns in return predictability are observed for alternate measures of investor sentiment.

This study does not aim to discuss each predictive signal separately, nor does it claim to thoroughly explain the predictability in equity returns. I consider a broad set of anomalies altogether that are known proxies for both risk and mispricing. Given this comprehensive set of predictors, I obtain a prediction model with satisfactory out-of-sample performance. The objective is to explore the possibility that sentiment-driven mispricing drives part of the observed return predictability. While the novel approach in this study reveals evidence consistent with sentiment-driven mispricing and its eventual correction, more exploration is warranted on the topic given the novel nature of financial machine learning literature.

# BIBLIOGRAPHY

- [ACM20] Doron Avramov, Si Cheng, and Lior Metzker. “Machine learning versus economic restrictions: Evidence from stock return predictability”. In: *Available at SSRN 3450322* (2020).
- [AGZ13] Gil Aharoni, Bruce Grundy, and Qi Zeng. “Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis”. In: *Journal of Financial Economics* 110.2 (2013), pp. 347–357.
- [Bal78] Ray Ball. “Anomalies in relationships between securities’ yields and yield-surrogates”. In: *Journal of financial economics* 6.2-3 (1978), pp. 103–126.
- [Ban81] Rolf W Banz. “The relationship between return and market value of common stocks”. In: *Journal of financial economics* 9.1 (1981), pp. 3–18.
- [Bha88] Laxmi Chand Bhandari. “Debt/equity ratio and expected common stock returns: Empirical evidence”. In: *The journal of finance* 43.2 (1988), pp. 507–528.
- [BJ+08] Arindam Bandopadhyaya, Anne Leah Jones, et al. “Measures of investor sentiment: A comparative analysis put-call ratio vs. volatility index”. In: *Journal of Business & Economics Research (JBER)* 6.8 (2008).
- [Bla72] Fischer Black. “Capital market equilibrium with restricted borrowing”. In: *The Journal of business* 45.3 (1972), pp. 444–455.
- [BRS93] Jacob Boudoukh, Matthew Richardson, and Tom Smith. “Is the ex ante risk premium always positive?: A new approach to testing conditional asset pricing models”. In: *Journal of Financial Economics* 34.3 (1993), pp. 387–408.

- [BW06] Malcolm Baker and Jeffrey Wurgler. “Investor sentiment and the cross-section of stock returns”. In: *The Journal of Finance* 61.4 (2006), pp. 1645–1680.
- [Coc08] John H Cochrane. “The dog that did not bark: A defense of return predictability”. In: *The Review of Financial Studies* 21.4 (2008), pp. 1533–1575.
- [Coc11] John H Cochrane. “Presidential address: Discount rates”. In: *The Journal of Finance* 66.4 (2011), pp. 1047–1108.
- [CPZ19] Luyang Chen, Markus Pelger, and Jason Zhu. “Deep learning in asset pricing”. In: *Available at SSRN 3350138* (2019).
- [DeM+17] Victor DeMiguel et al. “A portfolio perspective on the multitude of firm characteristics”. In: (2017).
- [DT98] Kent Daniel and Sheridan Titman. “Characteristics or covariances”. In: *Journal of Portfolio Management* 24.4 (1998), pp. 24–33.
- [FF15] Eugene F Fama and Kenneth R French. “A five-factor asset pricing model”. In: *Journal of financial economics* 116.1 (2015), pp. 1–22.
- [FF92] Eugene F Fama and Kenneth R French. “The cross-section of expected stock returns”. In: *the Journal of Finance* 47.2 (1992), pp. 427–465.
- [FF93] Eugene F Fama and Kenneth R French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of financial economics* 33.1 (1993), pp. 3–56.
- [FGX17] Guanhao Feng, Stefano Giglio, and Dacheng Xiu. “Taming the factor zoo”. In: *Chicago Booth research paper 17-04* (2017).
- [FMO08] Frank J Fabozzi, KC Ma, and Becky J Oliphant. “Sin stock returns”. In: *The Journal of Portfolio Management* 35.1 (2008), pp. 82–94.
- [FNW17] J Freyberger, A Neuhierl, and M Weber. *Dissecting characteristics nonparametrically* (No. w23227). 2017.
- [GHZ13] Jeremiah Green, John RM Hand, and X Frank Zhang. “The superview of return predictive signals”. In: *Review of Accounting Studies* 18.3 (2013), pp. 692–730.



- [GKX20a] Shihao Gu, Bryan Kelly, and Dacheng Xiu. “Autoencoder asset pricing models”. In: *Journal of Econometrics* (2020).
- [GKX20b] Shihao Gu, Bryan Kelly, and Dacheng Xiu. “Empirical asset pricing via machine learning”. In: *The Review of Financial Studies* 33.5 (2020), pp. 2223–2273.
- [HL15] Campbell R Harvey and Yan Liu. “Lucky factors”. In: *Available at SSRN* 2528780 (2015), p. 980.
- [HLZ16] Campbell R Harvey, Yan Liu, and Heqing Zhu. “... and the cross-section of expected returns”. In: *The Review of Financial Studies* 29.1 (2016), pp. 5–68.
- [HLZ19] Dashan Huang, Jiaen Li, and Guofu Zhou. “Shrinking factor dimension: A reduced-rank approach”. In: *Available at SSRN* 3205697 (2019).
- [Hua+15] Dashan Huang, Fuwei Jiang, et al. “Investor sentiment aligned: A powerful predictor of stock returns”. In: *The Review of Financial Studies* 28.3 (2015), pp. 791–837.
- [JT93] Narasimhan Jegadeesh and Sheridan Titman. “Returns to buying winners and selling losers: Implications for stock market efficiency”. In: *The Journal of finance* 48.1 (1993), pp. 65–91.
- [KNS18] Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. “Interpreting factor models”. In: *The Journal of Finance* 73.3 (2018), pp. 1183–1223.
- [KNS20] Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. “Shrinking the cross-section”. In: *Journal of Financial Economics* 135.2 (2020), pp. 271–292.
- [Koz19] Serhiy Kozak. “Kernel trick for the cross-section”. In: *Available at SSRN* 3307895 (2019).
- [KPS19] Bryan T Kelly, Seth Pruitt, and Yinan Su. “Characteristics are covariances: A unified model of risk and return”. In: *Journal of Financial Economics* 134.3 (2019), pp. 501–524.

- [KT15] Leonid Kogan and Mary H Tian. “Firm characteristics and empirical factor models: A model-mining experiment”. In: *FRB International Finance Discussion Paper* (2015).
- [Lew14] Jonathan Lewellen. “The cross section of expected stock returns”. In: *Forthcoming in Critical Finance Review, Tuck School of Business Working Paper* 2511246 (2014).
- [Lit65] J Litner. “The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets, Review of Economic Statistics 47”. In: (1965).
- [Mar52] Harry Markowitz. “Portfolio Selection, Journal of Finance”. In: *Markowitz HM—1952.—№* (1952), pp. 77–91.
- [Mer80] Robert C Merton. “On estimating the expected return on the market: An exploratory investigation”. In: *Journal of financial economics* 8.4 (1980), pp. 323–361.
- [Mil77] Edward M Miller. “Risk, uncertainty, and divergence of opinion”. In: *The Journal of finance* 32.4 (1977), pp. 1151–1168.
- [MN16] Michael W McCracken and Serena Ng. “FRED-MD: A monthly database for macroeconomic research”. In: *Journal of Business & Economic Statistics* 34.4 (2016), pp. 574–589.
- [MP16] R David McLean and Jeffrey Pontiff. “Does academic research destroy stock return predictability?” In: *The Journal of Finance* 71.1 (2016), pp. 5–32.
- [Nov13] Robert Novy-Marx. “The other side of value: The gross profitability premium”. In: *Journal of financial economics* 108.1 (2013), pp. 1–28.
- [PRS19] Kuntara Pukthuanthong, Richard Roll, and Avanidhar Subrahmanyam. “A protocol for factor identification”. In: *The Review of Financial Studies* 32.4 (2019), pp. 1573–1607.
- [Ros76] Stephen A Ross. “Options and efficiency”. In: *The Quarterly Journal of Economics* 90.1 (1976), pp. 75–89.

- [RRL85] Barr Rosenberg, Kenneth Reid, and Ronald Lanstein. “Persuasive evidence of market inefficiency”. In: *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.
- [Sha64] William F Sharpe. “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3 (1964), pp. 425–442.
- [Sta80] Dennis Stattman. “Book values and stock returns”. In: *The Chicago MBA: A journal of selected papers* 4.1 (1980), pp. 25–45.
- [SY17] Robert F Stambaugh and Yu Yuan. “Mispricing factors”. In: *The Review of Financial Studies* 30.4 (2017), pp. 1270–1315.
- [SYY12] Robert F Stambaugh, Jianfeng Yu, and Yu Yuan. “The short of it: Investor sentiment and anomalies”. In: *Journal of Financial Economics* 104.2 (2012), pp. 288–302.
- [SYY15] Robert F Stambaugh, Jianfeng Yu, and Yu Yuan. “Arbitrage asymmetry and the idiosyncratic volatility puzzle”. In: *The Journal of Finance* 70.5 (2015), pp. 1903–1948.
- [Wha00] Robert E Whaley. “The investor fear gauge”. In: *The Journal of Portfolio Management* 26.3 (2000), pp. 12–17.
- [Wha09] Robert E Whaley. “Understanding the VIX”. In: *The Journal of Portfolio Management* 35.3 (2009), pp. 98–105.

# BIBLIOGRAPHY

- [ACM20] Doron Avramov, Si Cheng, and Lior Metzker. “Machine learning versus economic restrictions: Evidence from stock return predictability”. In: *Available at SSRN 3450322* (2020).
- [AGZ13] Gil Aharoni, Bruce Grundy, and Qi Zeng. “Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis”. In: *Journal of Financial Economics* 110.2 (2013), pp. 347–357.
- [Bal78] Ray Ball. “Anomalies in relationships between securities’ yields and yield-surrogates”. In: *Journal of financial economics* 6.2-3 (1978), pp. 103–126.
- [Ban81] Rolf W Banz. “The relationship between return and market value of common stocks”. In: *Journal of financial economics* 9.1 (1981), pp. 3–18.
- [Bha88] Laxmi Chand Bhandari. “Debt/equity ratio and expected common stock returns: Empirical evidence”. In: *The journal of finance* 43.2 (1988), pp. 507–528.
- [BJ+08] Arindam Bandopadhyaya, Anne Leah Jones, et al. “Measures of investor sentiment: A comparative analysis put-call ratio vs. volatility index”. In: *Journal of Business & Economics Research (JBER)* 6.8 (2008).
- [Bla72] Fischer Black. “Capital market equilibrium with restricted borrowing”. In: *The Journal of business* 45.3 (1972), pp. 444–455.
- [BRS93] Jacob Boudoukh, Matthew Richardson, and Tom Smith. “Is the ex ante risk premium always positive?: A new approach to testing conditional asset pricing models”. In: *Journal of Financial Economics* 34.3 (1993), pp. 387–408.

- [BW06] Malcolm Baker and Jeffrey Wurgler. “Investor sentiment and the cross-section of stock returns”. In: *The Journal of Finance* 61.4 (2006), pp. 1645–1680.
- [Coc08] John H Cochrane. “The dog that did not bark: A defense of return predictability”. In: *The Review of Financial Studies* 21.4 (2008), pp. 1533–1575.
- [Coc11] John H Cochrane. “Presidential address: Discount rates”. In: *The Journal of Finance* 66.4 (2011), pp. 1047–1108.
- [CPZ19] Luyang Chen, Markus Pelger, and Jason Zhu. “Deep learning in asset pricing”. In: *Available at SSRN 3350138* (2019).
- [DeM+17] Victor DeMiguel et al. “A portfolio perspective on the multitude of firm characteristics”. In: (2017).
- [DT98] Kent Daniel and Sheridan Titman. “Characteristics or covariances”. In: *Journal of Portfolio Management* 24.4 (1998), pp. 24–33.
- [FF15] Eugene F Fama and Kenneth R French. “A five-factor asset pricing model”. In: *Journal of financial economics* 116.1 (2015), pp. 1–22.
- [FF92] Eugene F Fama and Kenneth R French. “The cross-section of expected stock returns”. In: *the Journal of Finance* 47.2 (1992), pp. 427–465.
- [FF93] Eugene F Fama and Kenneth R French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of financial economics* 33.1 (1993), pp. 3–56.
- [FGX17] Guanhao Feng, Stefano Giglio, and Dacheng Xiu. “Taming the factor zoo”. In: *Chicago Booth research paper 17-04* (2017).
- [FMO08] Frank J Fabozzi, KC Ma, and Becky J Oliphant. “Sin stock returns”. In: *The Journal of Portfolio Management* 35.1 (2008), pp. 82–94.
- [FNW17] J Freyberger, A Neuhierl, and M Weber. *Dissecting characteristics nonparametrically* (No. w23227). 2017.
- [GHZ13] Jeremiah Green, John RM Hand, and X Frank Zhang. “The superview of return predictive signals”. In: *Review of Accounting Studies* 18.3 (2013), pp. 692–730.

- [GKX20a] Shihao Gu, Bryan Kelly, and Dacheng Xiu. “Autoencoder asset pricing models”. In: *Journal of Econometrics* (2020).
- [GKX20b] Shihao Gu, Bryan Kelly, and Dacheng Xiu. “Empirical asset pricing via machine learning”. In: *The Review of Financial Studies* 33.5 (2020), pp. 2223–2273.
- [HL15] Campbell R Harvey and Yan Liu. “Lucky factors”. In: *Available at SSRN* 2528780 (2015), p. 980.
- [HLZ16] Campbell R Harvey, Yan Liu, and Heqing Zhu. “... and the cross-section of expected returns”. In: *The Review of Financial Studies* 29.1 (2016), pp. 5–68.
- [HLZ19] Dashan Huang, Jiaen Li, and Guofu Zhou. “Shrinking factor dimension: A reduced-rank approach”. In: *Available at SSRN* 3205697 (2019).
- [Hua+15] Dashan Huang, Fuwei Jiang, et al. “Investor sentiment aligned: A powerful predictor of stock returns”. In: *The Review of Financial Studies* 28.3 (2015), pp. 791–837.
- [JT93] Narasimhan Jegadeesh and Sheridan Titman. “Returns to buying winners and selling losers: Implications for stock market efficiency”. In: *The Journal of finance* 48.1 (1993), pp. 65–91.
- [KNS18] Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. “Interpreting factor models”. In: *The Journal of Finance* 73.3 (2018), pp. 1183–1223.
- [KNS20] Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. “Shrinking the cross-section”. In: *Journal of Financial Economics* 135.2 (2020), pp. 271–292.
- [Koz19] Serhiy Kozak. “Kernel trick for the cross-section”. In: *Available at SSRN* 3307895 (2019).
- [KPS19] Bryan T Kelly, Seth Pruitt, and Yinan Su. “Characteristics are covariances: A unified model of risk and return”. In: *Journal of Financial Economics* 134.3 (2019), pp. 501–524.

- [KT15] Leonid Kogan and Mary H Tian. “Firm characteristics and empirical factor models: A model-mining experiment”. In: *FRB International Finance Discussion Paper* (2015).
- [Lew14] Jonathan Lewellen. “The cross section of expected stock returns”. In: *Forthcoming in Critical Finance Review, Tuck School of Business Working Paper* 2511246 (2014).
- [Lit65] J Litner. “The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets, Review of Economic Statistics 47”. In: (1965).
- [Mar52] Harry Markowitz. “Portfolio Selection, Journal of Finance”. In: *Markowitz HM—1952.—№* (1952), pp. 77–91.
- [Mer80] Robert C Merton. “On estimating the expected return on the market: An exploratory investigation”. In: *Journal of financial economics* 8.4 (1980), pp. 323–361.
- [Mil77] Edward M Miller. “Risk, uncertainty, and divergence of opinion”. In: *The Journal of finance* 32.4 (1977), pp. 1151–1168.
- [MN16] Michael W McCracken and Serena Ng. “FRED-MD: A monthly database for macroeconomic research”. In: *Journal of Business & Economic Statistics* 34.4 (2016), pp. 574–589.
- [MP16] R David McLean and Jeffrey Pontiff. “Does academic research destroy stock return predictability?” In: *The Journal of Finance* 71.1 (2016), pp. 5–32.
- [Nov13] Robert Novy-Marx. “The other side of value: The gross profitability premium”. In: *Journal of financial economics* 108.1 (2013), pp. 1–28.
- [PRS19] Kuntara Pukthuanthong, Richard Roll, and Avanidhar Subrahmanyam. “A protocol for factor identification”. In: *The Review of Financial Studies* 32.4 (2019), pp. 1573–1607.
- [Ros76] Stephen A Ross. “Options and efficiency”. In: *The Quarterly Journal of Economics* 90.1 (1976), pp. 75–89.

- [RRL85] Barr Rosenberg, Kenneth Reid, and Ronald Lanstein. “Persuasive evidence of market inefficiency”. In: *The Journal of Portfolio Management* 11.3 (1985), pp. 9–16.
- [Sha64] William F Sharpe. “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3 (1964), pp. 425–442.
- [Sta80] Dennis Stattman. “Book values and stock returns”. In: *The Chicago MBA: A journal of selected papers* 4.1 (1980), pp. 25–45.
- [SY17] Robert F Stambaugh and Yu Yuan. “Mispricing factors”. In: *The Review of Financial Studies* 30.4 (2017), pp. 1270–1315.
- [SYY12] Robert F Stambaugh, Jianfeng Yu, and Yu Yuan. “The short of it: Investor sentiment and anomalies”. In: *Journal of Financial Economics* 104.2 (2012), pp. 288–302.
- [SYY15] Robert F Stambaugh, Jianfeng Yu, and Yu Yuan. “Arbitrage asymmetry and the idiosyncratic volatility puzzle”. In: *The Journal of Finance* 70.5 (2015), pp. 1903–1948.
- [Wha00] Robert E Whaley. “The investor fear gauge”. In: *The Journal of Portfolio Management* 26.3 (2000), pp. 12–17.
- [Wha09] Robert E Whaley. “Understanding the VIX”. In: *The Journal of Portfolio Management* 35.3 (2009), pp. 98–105.