Imperial College London
Business School

Essays on Empirical Asset Pricing

Adam Denny

2021

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy of Imperial College London

Abstract

This thesis examines issues in factor investing, return predictability, and applications of machine learning to empirical asset pricing.

The first chapter provides an empirical evaluation of potential explanations for the value premium by extending the decomposition methodology of Hou and Loh (2016). I find that cross-sectional differences in financial leverage are able to explain up to 60 percent of the value premium in individual stocks and book-to-market and industry sorted portfolios. I do not find support for behavioural explanations, or rational explanations based on adjustment costs, human capital, or operating leverage. I show that most of the explanatory power of financial leverage is concentrated in the short-term, and a factor based on financial leverage is able to partially explain the alpha of the Fama-French HML factor.

The second chapter provides a comprehensive analysis on the use of supervised machine learning methods for multi-asset class return predictability. I extend the existing literature in three dimensions: time, asset class, and geographical coverage. Using the longest available time series of asset returns and macroeconomic fundamentals across multiple countries I show that most machine learning models fail to beat the historical mean benchmark out-of-sample, often by a large margin. However, combining forecasts across models dramatically improves performance, and leads to sizeable gains for mean-variance investors in trading strategies that aim to exploit return predictability. Additionally, I show that forecasting performance is better in bad times and when market volatility is high.

The third chapter documents a new phenomenon I call 'cross-factor momentum.' I show that standard time series momentum strategies applied to long-short equity factors can be enhanced by incorporating predictability from other factors. Exploiting the robust cross-correlation structure between factors can improve Sharpe ratios by up to 1.7 times compared to standard strategies, with lower drawdowns and similar trading costs. This finding holds in international data, cannot be explained by intermediary asset pricing models, and is only partially explained by fund flows.

Declaration of Originality

I herewith declare that this thesis is my own work. All material and sources that do not constitute my own work have been explicitly acknowledged and referenced.

Adam Denny

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Adam Denny

Acknowledgements

I would like to extend my deepest thanks to my supervisors, Robert Kosowski and Pasquale Della Corte. Without their advice and guidance this thesis would not have been possible.

I am also grateful for all comments and feedback from faculty and seminar participants at Imperial College Business School, and especially my thesis examiners Paolo Zaffaroni and Daniele Bianchi. Special thanks go out to Andrea Buraschi, Walter Distaso, and James Sefton, who have greatly influenced the direction of this thesis, both directly and indirectly, through our discussions.

I would like to sincerely thank my fellow PhD students at Imperial in and out of the Finance Department: Can Gao, Zaici Li, Hormoz Ramian, and Oren Schneorson for their exceptional collective knowledge on all aspects of finance and economics from which I benefitted greatly; the MRes Class of 2016 for making my time at Imperial so welcoming; Mobeen Iqbal and Antoine Uettwiller for lively Python discussions; Krystal Lau for our break room chats; and many others.

Finally, I would like to thank my friends and family for their love and support over the last five years—I could not have done this without them.

Contents

1 What Drives the Value Premium? 13

1.1 Introduction . . . 13

1.2 Literature Review . . . 16

1.2.1 Risk-Based Explanations . . . 16

1.2.2 Behavioural Explanations . . . 18

1.3 Data . . . 20

1.3.1 Sample . . . 20

1.3.2 Measurement of Key Variables . . . 21

1.3.2.1 Adjustment Costs . . . 21

1.3.2.2 Earnings Surprise . . . 22

1.3.2.3 Financial Leverage . . . 22

1.3.2.4 Human Capital . . . 23

1.3.2.5 Institutional Ownership . . . 23

1.3.2.6 Operating Leverage . . . 24

1.4 Decomposition Methodology . . . 26

1.4.1 Summary of Hou and Loh (2016) . . . 26

1.4.2 Extension to Multiple First-Stage Regressors . . . 28

1.5 Empirical Analysis . . . 30

1.5.1 Descriptive Statistics . . . 30

1.5.2 Univariate Sorts . . . 31

1.5.3 Decomposing the Value Premium . . . 32

1.5.3.1 Univariate Analysis . . . 32

1.5.3.2 Multivariate Analysis . . . 34

1.5.4 Asset Pricing Tests . . . 34

1.5.5 The Role of Financial Leverage . . . 35
1.5.6 Robustness Tests . . . 36
1.6 Conclusion . . . 38

2 Three Centuries of Machine Learning for Return Predictability 53
2.1 Introduction . . . 53
2.2 Literature Review . . . 55
2.2.1 Out-of-Sample Return Predictability . . . 55
2.2.1.1 Equity Return Predictability . . . 55
2.2.1.2 Bond Return Predictability . . . 56
2.2.1.3 Currency Return Predictability . . . 56
2.2.2 Machine Learning in Finance . . . 57
2.3 Data . . . 59
2.3.1 Equities . . . 59
2.3.2 Bonds . . . 59
2.3.3 Currencies . . . 60
2.3.4 Macroeconomic Fundamentals and Financial Variables . . . 60
2.4 Forecasting Framework . . . 61
2.4.1 Notation . . . 61
2.4.2 Economic Restrictions . . . 61
2.4.3 Forecast Combination . . . 62
2.5 Forecasting with Machine Learning . . . 66
2.5.1 Why Machine Learning? . . . 66
2.5.2 Overview of Machine Learning Models . . . 66
2.5.2.1 Ordinary Least Squares . . . 67
2.5.2.2 Penalised Regression . . . 67
2.5.2.3 Partial Least Squares . . . 68
2.5.2.4 Random Forest . . . 69
2.5.2.5 Gradient Boosting . . . 69
2.5.2.6 K-Nearest Neighbours . . . 70

2.5.2.7 A Note on Neural Networks . . . 71
2.5.3 Hyperparameter Tuning . . . 71
2.6 Statistical Evaluation . . . 73
2.6.1 Forecast Evaluation Methodology . . . 73
2.6.1.1 Out-of-Sample R-Squared . . . 73
2.6.1.2 Henriksson-Merton Test . . . 73
2.6.1.3 Mincer-Zarnowitz Regressions . . . 74
2.6.1.4 Model Comparison . . . 74
2.6.2 Results . . . 75
2.6.2.1 Full-Sample . . . 75
2.6.2.2 Sample Splits . . . 78
2.6.2.3 Variable Importance . . . 78
2.7 Economic Evaluation . . . 80
2.7.1 Dynamic Trading Strategies . . . 80
2.7.2 Performance Measures . . . 81
2.7.3 Results . . . 82
2.8 Conclusion . . . 83

3 Cross-Factor Momentum 106

3.1 Introduction . . . 106
3.2 Time Series Momentum . . . 109
3.3 Data . . . 111
3.4 Factor Momentum . . . 112
3.4.1 Previous Literature . . . 112
3.4.2 Evidence on Factor Momentum . . . 113
3.4.2.1 In-Sample . . . 113
3.4.2.2 Out-of-Sample . . . 115
3.4.3 Time Series Factor Momentum Strategies . . . 117
3.5 Cross-Factor Momentum . . . 119
3.5.1 Evidence of Cross-Factor Correlations . . . 119

3.5.1.1 In-Sample . . . 119
3.5.1.2 Out-of-Sample . . . 121
3.5.2 Cross-Factor Momentum Strategies . . . 121
3.5.3 Robustness Tests and Additional Exercises . . . 122
3.6 What Drives Cross-Factor Predictability? . . . 124
3.6.1 Intermediary Channel . . . 124
3.6.2 Fund Flow Channel . . . 124
3.6.2.1 Fund Flows . . . 125
3.6.2.2 Return-Flow VAR . . . 126
3.7 Conclusion . . . 127

Bibliography **145**

Appendix **155**

List of Figures

1.1 Time series distribution of candidate variables 48
1.2 Cross-sectional distribution of candidate variables 49
1.3 Univariate proportion of the value premium explained 50
1.4 Proportion of the value premium explained over time 51
1.5 Distribution of FLEV across book-to-market quintiles 52

2.1 Illustrative decision tree 98
2.2 Walk forward validation procedure 98
2.3 Variable importance for international equity markets 99
2.4 Variable importance for international government bond markets 100
2.5 Variable importance for real currency returns 101
2.6 Cumulative MSE difference for international equity markets 102
2.7 Cumulative MSE difference for international government bond markets 103
2.8 Cumulative MSE difference for real currency returns 104
2.9 Cumulative returns for trading strategies 105

3.1 Lagged return regression t-statistics 139
3.2 Lagged return sign regression t-statistics 140
3.3 Lagged cross-factor return regression t-statistics 141
3.4 Return-flow VAR impulse response functions 144

List of Tables

1.1 Descriptive statistics. . . . 39
1.2 Pairwise correlation matrix . . . 40
1.3 Characteristics of book-to-market sorted portfolios . . . 41
1.4 Univariate decomposition . . . 42
1.5 Multivariate decomposition . . . 43
1.6 Factor regressions . . . 44
1.7 Short-term vs. long-term financial leverage . . . 45
1.8 Portfolio level decomposition . . . 46
1.9 Comparison of decomposition methodologies . . . 47

2.1 Summary statistics for equity, bond, and currency returns . . . 84
2.2 Summary of hyperparameter optimisation . . . 85
2.3 Out-of-sample R^2 for machine learning models . . . 86
2.4 Out-of-sample R^2 for combined forecasts with economic restrictions . . . 87
2.5 Correlation matrix of return forecasts . . . 88
2.6 Henriksson-Merton directional accuracy tests . . . 89
2.7 Mincer-Zarnowitz regressions for MSE(12) forecasts of equity and bond returns . . . 90
2.8 Mincer-Zarnowitz regressions for MSE(12) forecasts of currency returns . . . 91
2.9 Out-of-sample R^2 by recession subperiod . . . 92
2.10 Out-of-sample R^2 by volatility regime . . . 93
2.11 Out-of-sample R^2 by inflation environment . . . 94
2.12 Out-of-sample R^2 by interest rate environment . . . 95
2.13 Out-of-sample R^2 pre- and post-WW2 . . . 96
2.14 Summary statistics for strategy returns . . . 97

3.1 Summary statistics for monthly factor returns . . . 128

3.2 t-statistics for pooled predictive regressions with factor fixed effects 129
3.3 Out-of-sample performance for time series factor momentum regressions 130
3.4 Lagged return LASSO predictive regression coefficients 131
3.5 Performance of time series factor momentum strategies 132
3.6 Cross-correlation matrix for monthly factor returns 133
3.7 Cross-factor LASSO predictive regression coefficients 134
3.8 Performance of cross-factor momentum strategies 135
3.9 Sharpe ratios of international cross-factor momentum strategies 136
3.10 Alphas from intermediary factor regressions 137
3.11 Return-flow VAR coefficients 138

Chapter 1

What Drives the Value Premium?

1.1 Introduction

The value premium refers to the observation that stocks with high book-to-market ratios (value stocks) tend to outperform stocks with low book-to-market ratios (growth stocks). It can be viewed as anomalous, since conventional wisdom says that growth options are riskier than assets-in-place, thereby generating a *growth* premium. Yet the existence of a value premium appears to be a pervasive feature of modern equity markets, both in the US and internationally (Fama and French, 1993, 2012; Asness, Moskowitz, and Pedersen, 2013).

There have been many attempts to resolve the value anomaly, but there remains no consensus on its fundamental drivers. Existing explanations proposed in the literature typically fall into one of two categories: risk-based or behavioural. Risk-based explanations argue that the value premium is compensation for systematic risk not captured by the market factor of the Sharpe-Lintner capital asset pricing model (CAPM). Behavioural explanations rely on investor irrationality and market inefficiency to generate systematic differences in the returns of value and growth stocks.

In this chapter, I use a novel decomposition methodology to comprehensively examine a range of risk-based and behavioural explanations that claim to be able to explain the value premium. This chapter is motivated by the observation that when a new theory or explanation is proposed, it is typically evaluated in isolation or considering only a handful of alternative explanations, making it difficult to directly compare the relative strength of competing explanations. Moreover, different studies use different methodologies and samples to evaluate their hypotheses, which further muddies a comparison of the results of existing studies. Therefore, the first contribution of this chapter is to

harmonise the literature by evaluating a range of alternative explanations using a consistent dataset and unified framework.

Another contribution is to generalise the decomposition methodology of Hou and Loh (2016) to the case of multiple regressors. In summary, the methodology is used to decompose the coefficient from a cross-sectional regression of stock returns on a variable related to an asset pricing anomaly (e.g. book-to-market ratio) into the contribution of one or more candidate explanations and an unexplained residual component. The contributions of the candidate variables and the residual are constructed to sum to one, and therefore can be interpreted as the percentage of the anomaly explained. This makes a direct comparison of the various candidate explanations straightforward and transparent. I argue that this methodology offers an ideal framework to evaluate competing explanations of the value premium, which typically relate cross-sectional differences in returns and book-to-market ratios to firm-specific characteristics, often based on accounting variables. My extension uses a simple theorem from econometrics to allow the user to control for a large number of other variables related to stock returns but not directly related to the variable(s) of interests, such as firm size, prior returns (momentum), volatility, and profitability.

Using this methodology, I find that cross-sectional differences in financial leverage are able to explain 56 percent of the value premium in a univariate setting, and between 40 and 60 percent in a multivariate setting, after controlling for size and momentum. Value firms are typically more levered than growth firms, and one interpretation is that the increased risk of financial distress generates a positive risk premium on value stocks. This is consistent with the real options model of Ozdagli (2012) and empirical evidence in Cao (2015), who both conclude that the value premium is primarily driven by financial leverage. In contrast to previous studies, I find that risk-based explanations based on adjustment costs, human capital, and operating leverage are not able to explain the value premium, and sometimes predict a *negative* relationship between returns and book-to-market ratios. I also do not find support for behavioural explanations based on earnings surprises or institutional ownership; in fact the evidence suggests earnings surprises are lower and frequently negative for value firms. These results are robust to alternative definitions of financial leverage, and are similar for both individual stocks and portfolios sorted by book-to-market or industry.

Digging deeper into the financial distress theory I find that, consistent with economic intuition,

most of the effect is concentrated in short-term financial leverage (debt maturing in less than one year). Short-term financial leverage explains around 42 percent of the value premium, whilst long-term financial leverage explains only 21 percent, despite long-term debt making up a much larger proportion of total debt on average. This strengthens the interpretation of value as a financial distress risk premium if short-term debt is believed to be riskier than long-term debt (Choe, 1994).

The rest of the chapter proceeds as follows. Section 1.2 surveys the literature on the value premium and its explanations. Section 1.3 describes the data. Section 1.4 details the decomposition methodology. Section 1.5 presents the empirical analysis and robustness tests, and Section 1.6 concludes. Proofs can be found in Appendices A.1 and A.2.

1.2 Literature Review

Theories of the value premium typically fall into one of two categories. Rational risk-based explanations relate the returns on high book-to-market stocks to some fundamental measure(s) of systematic risk. Behavioural explanations deviate from the traditional complete markets rational expectations framework and rely on investor irrationality and market inefficiency to explain the difference in returns between value and growth stocks. In this section, I survey both strands of literature to identify candidate variables to be tested in the empirical analysis.

1.2.1 Risk-Based Explanations

In the models of Zhang (2005) and Cooper (2006), the value premium reflects asymmetric adjustment costs (i.e. costly reversibility) faced by firms when investing and disinvesting in capital. Although these papers differ in technical details, the mechanism generating the value premium is similar. High book-to-market firms derive most of their value from assets-in-place, so when there is a negative shock to aggregate demand, value firms are burdened with unproductive capital that is costly to reduce. Conversely, low book-to-market firms derive most of their value from growth options, which can costlessly remain unexercised in light of an adverse demand shock. Therefore, the returns of value stocks covary more with aggregate economic conditions, generating higher systematic risk. Zhang (2005) calibrates his model and shows that it is able to generate a value premium comparable in magnitude to historical data. In simulations, Cooper (2006) shows that when investment is completely irreversible, his model is able to match the empirical results in Fama and French (1992). However, using various proxies for adjustment costs, Cao (2015) empirically finds that the value premium *decreases* with the degree of investment irreversibility, contrary to the intuition in Zhang (2005) and Cooper (2006).

Several papers document a link between financial leverage, financial distress, and the value premium. Vassalou and Xing (2004) find that the value premium only exists in the two quintiles of firms with the highest default risk. Ozdagli (2012) extends the investment irreversibility model of Abel and Eberly (1996) to incorporate investors' risk preferences, risk-free debt, and debt adjustment costs. He shows that adjusting equity returns to account for leverage in the sense of Modigliani and Miller (1958)

reduces the spread between the top and bottom decile portfolios sorted on book-to-market from 16 percent to 4 percent per annum. Higher financial leverage also indirectly affects returns through its impact on business risk and investment decisions. However, in Ozdagli's model, investment irreversibility remains an important ingredient to improve the model's fit with the data. Doshi, Jacobs, Kumar, and Rabinovich (2019) note that the relationship between financial leverage and cross-sectional variation in stock returns is highly nonlinear, and use structural credit risk models to adjust for leverage and compute unlevered stock returns. The authors find that the value premium disappears when using unlevered equity returns, and this finding is robust to using different credit risk models.

Research as early as Lev (1974) has recognised the link between operating leverage and expected stock returns, and recently operating leverage has been linked to the value premium. In the structural model of Carlson, Fisher, and Giammarino (2004) when demand for a firm's product decreases, growth options become less valuable, which reduces market equity and increases the book-to-market ratio. Assuming fixed costs are proportional to book value (a proxy for the firm's capital stock), the riskiness of returns increases due to higher operating leverage. Similarly, Obreja (2013) theoretically shows when operating leverage is significantly high, value firms have high equity risk premia. Garcia-Feijoo and Jorgensen (2010) and Novy-Marx (2011) provide direct empirical evidence that firms with higher operating leverage earn higher average returns than operationally unlevered firms.

Fama and French (1996) hint that the value premium may be related to human capital. They argue that if an investor has specialised human capital tied to a growth firm, then a negative shock to the growth firm's prospects probably does not reduce the value of the investor's human capital; it may just expand less rapidly. In contrast, a negative shock to a distressed firm more likely implies a reduction in the value of human capital, since the distressed firm is more likely to downsize. This creates an incentive for investors not to hold the stock of distressed firms, and generates a "state-variable risk premium in the expected returns of distressed stocks" (Fama and French, 1996, pg. 77). Sylvain (2014) develops a general equilibrium model where a value premium arises endogenously from risk to human capital. In this model, value firms have relatively more firm-specific human capital, and the positive covariance between asset growth and returns to human capital generates a positive value risk premium. Using data on Swedish firms, Hansson (2004) provides support for

the human capital theory, arguing that dispersion in average wage growth across firms (a proxy for human capital) is able to explain a large portion of the value premium.

Finally, the value premium has been linked to measures of aggregate uncertainty. Uncertainty refers to a situation where investors do not know the probability distribution of outcomes they face. This differs from risk where the distribution is known but the outcome is not. Aboura and Arisoy (2017) use the VVIX as a proxy for aggregate uncertainty, which is an indicator of the 30-day forward price of the VIX index, calculated daily by the CBOE and available from January 2007. They find that portfolios of small and value stocks have negative and significant uncertainty betas, providing an uncertainty-based explanation of the value premium. Due to the relatively short history of the VVIX it is not included in the set of candidate variables, but is discussed here for completeness.

1.2.2 Behavioural Explanations

Some papers do not find support for risk-based explanations. For example, the evidence in Daniel and Titman (1997) suggests there is no discernible risk factor associated with high or low book-to-market firms, and there is no return premium associated with any of the Fama and French (1993) factors. Ong, Fong, and Low (2008) find that value stocks second-order stochastically dominate growth stocks (i.e. a risk-averse investor would prefer the distribution of value returns to the distribution of growth returns), implying the implausibility of risk-based explanations if investors act as rational agents.

Alternative explanations based on investor irrationality and market inefficiency have been proposed to explain the value premium. Lakonishok, Shleifer, and Vishny (1994) (henceforth LSV) argue that strategies that long value stocks and short growth stocks outperform because they are contrarian to naive strategies. Naive strategies include extrapolating past earnings growth too far into the future, overreacting to good or bad news, and assuming recent trends in stock prices are likely to continue in the future. They find value stocks are not fundamentally riskier than growth stocks based on their performance in states of the world where marginal utility of consumption is high (e.g. recessions and extreme bear markets). However, the LSV definition of a value stock is a stock that has performed poorly in the past based on earnings, cash flow, and sales growth, and is expected to perform poorly in the future. This is not necessarily the same definition used by other authors, who

primarily define value in terms of the book-to-market ratio.

La Porta (1996) uses survey data on analyst forecasts of expected five-year earnings growth rates, and shows that such forecasts are often too extreme. He finds that stocks with low expected growth rates outperform stocks with high expected growth rates by over 20 percent per year. La Porta, Lakonishok, Shleifer, and Vishny (1997) study stock price reactions around earnings announcements for value and growth stocks, and find earnings surprises for value stocks are systematically positive, suggesting investors are persistently pessimistic about the prospects of value firms. Both of these papers are consistent with the LSV proposition that the value premium is caused by expectational errors of investors.

Finally, Phalippou (2007a) shows that several prominent risk-based theories that can capture the cross-section of returns on portfolios sorted on size and book-to-market are not able to explain portfolio returns sorted on book-to-market and institutional ownership (IO). IO is relevant to the value premium for two reasons. First, as highlighted by Nagel (2005), the outperformance of value stocks is most pronounced amongst stocks with the lowest IO, since short sale constraints are more likely to bind. Second, Phalippou (2007b) finds that the value premium is heavily concentrated in stocks held by individual investors. He argues that IO provides a parsimonious way to classify stocks by their mispricing likelihood, and the value premium arises purely from mispricing and limits to arbitrage occurring in a small subset of stocks with low IO.

1.3 Data

1.3.1 Sample

Data on the prices, returns, and number of shares outstanding for all stocks with ordinary common equity (share code 10 or 11) listed on the three major US stock exchanges (NYSE, AMEX, and NASDAQ) is obtained from the monthly CRSP file. I exclude financial firms (SIC 6000–6999) because they have different financing structures than non-financial firms leading to a different interpretation of the book-to-market ratio, and utilities (SIC 4900–4999) because their investment and financing activities are more strictly regulated. I exclude stocks with prices less than \$5, and those that have missing market capitalisations. Duplicate share classes are also dropped. Returns are adjusted for the survivorship bias arising from stock delisting following Shumway (1997) and Shumway and Warther (1999).[1] Data on firm fundamentals is from the annual CRSP/Compustat Merged database. The quarterly CRSP/Compustat Merged database is also used to estimate some variables where a longer time series is needed to improve estimation accuracy (see next subsection).

The book-to-market (BM) ratio is calculated as in Fama and French (1992). Specifically, book equity (BE) is calculated as the sum of the book value of common equity (Compustat item CEQ) plus deferred taxes and investment tax credit (item TXDITC).[2] BE for each fiscal year end in calendar year $t-1$ is then matched with returns for July of year t to June of year $t+1$. This guarantees at least a six month gap between the fiscal year end and the monthly return, to ensure that the accounting data used to predict returns was observable to investors at that time. Market capitalisation is taken at the end of December of calendar year $t-1$ to compute the BM ratio. Firms with negative or missing BM are dropped. The other accounting variables are matched to returns in the same way as BE. All variables are winsorised at the 0.5th and 99.5th percentiles to remove the impact of extreme outliers and erroneous data entries.

[1] In particular, if a stock was delisted for performance reasons (CRSP codes 520–587, 589–591) or the reason is not specified (CRSP code 500) and the CRSP delisting return is missing, I use -30 percent and -55 percent as the return in the month of delisting for NYSE/AMEX and NASDAQ firms respectively. Around 0.75 percent of NYSE/AMEX observations and 2.08 percent of NASDAQ observations in my sample are adjusted in this way.

[2] Some authors prefer to also subtract the book value of preferred stock. For example, Gomes and Schmid (2010) define BE as the book value of common equity plus deferred taxes and investment tax credit minus the redemption, liquidation, or par value of preferred stock (items PSTKRV, PSTKL, and PSTK respectively). The results in this chapter are robust to using either definition, since the book value of preferred stock is typically very small (on average less than 2 percent of common equity in my sample).

The final sample contains 1,260,986 firm-month observations across 12,255 unique firms from January 1962 to December 2019. However, not all variables are available in each month because either (i) the accounting data is not reported in Compustat, or (ii) past observations are needed to construct the variable. See Table 1.1 for a summary of the number of observations available for each variable.

1.3.2 Measurement of Key Variables

1.3.2.1 Adjustment Costs

Adjustment costs are not observable, so proxies are needed to evaluate the predictions of Zhang (2005) and Cooper (2006). At least three measures have been proposed in the microeconomics and industrial organisation literature, which I employ in the empirical analysis. Kessides (1990), Farinas and Ruano (2005), and Chirinko and Schaller (2009) use the depreciation-to-fixed assets ratio (DEP), arguing assets that depreciate more quickly are easier to discard if the firm needs to downsize, reducing the degree of adjustment costs. I take depreciation from Compustat as depreciation of tangible fixed assets (item DFXA), and fixed assets are taken as total gross property, plant and equipment (item PPEGT). Kessides (1990) and Farinas and Ruano (2005) also use a firm's rental expense (RENT), since rented assets can more easily be disposed at the end of the rental contract. I scale a firm's rental expense (item XRENT) by the firm's balance sheet fixed assets (item PPEGT) to generate a scale-free measure that is comparable across firms. Campa (1993) uses asset tangibility (TANG)—the ratio of fixed to total assets (item AT)—since it is more costly for fixed asset intensive firms to scale down operations, because of, e.g., fire sales or illiquid second-hand markets.

Finally, following Cao (2015), I aggregate the information contained in DEP, RENT, and TANG into a composite measure of adjustment costs (ADJCO) using principal component analysis (PCA). ADJCO for firm i in month t is constructed using the first principal component loadings of the standardised natural log of one plus DEP, RENT, and TANG for firm i in month t, as in equation (1.1), where the loadings $\ell_{1t}, \ldots, \ell_{3t}$ and standardisation factors are estimated for all stocks over an expanding window using only data up to time t:

$$\text{ADJCO}_{it} \;=\; \ell_{1t}\log(1+\text{DEP}_{it}) + \ell_{2t}\log(1+\text{RENT}_{it}) + \ell_{3t}\log(1+\text{TANG}_{it}). \tag{1.1}$$

The eigenvalues of the three components over the full sample are 1.38, 0.95, and 0.67 respectively, indicating that the first principal component explains about 46% of the variance in the data, followed by 32% and 22% for the second and third respectively. ADJCO typically loads positively on DEP and RENT (i.e. $\ell_1, \ell_2 > 0$), and negatively on TANG ($\ell_3 < 0$), and therefore can be interpreted as the *reversibility* of investment as opposed to the irreversibility, and hence should be negatively related to BM.

1.3.2.2 Earnings Surprise

A commonly used measure of earnings surprise is standardised unexpected earnings (SUE), which assumes quarterly earnings follows a seasonal random walk (e.g. Hou, Xiong, and Peng, 2010; Novy-Marx, 2015). SUE is defined as the most recently available quarterly earnings before extraordinary items (item IBQ) minus earnings four quarters ago, divided by the standard deviation of the difference over the previous eight quarters. If investors are systematically pessimistic about the growth prospects of value stocks as in La Porta (1996) and La Porta, Lakonishok, Shleifer, and Vishny (1997), then SUE should be positive and relatively higher for high BM firms. Conversely, SUE should be negative for low BM firms if investors are irrationally overoptimistic about the prospects of growth stocks.

1.3.2.3 Financial Leverage

Following Welch (2004), Chen and Zhao (2006), and others, the baseline measure of financial leverage (FLEV) is defined as:

$$\text{FLEV} = \frac{\text{CD} + \text{NCD}}{\text{CD} + \text{NCD} + \text{ME}}, \tag{1.2}$$

where CD is current debt (item DLC), NCD is noncurrent debt (item DLTT), and ME is market equity. Essentially, the definition in (1.2) is using book debt as a proxy for (unobservable) market debt.[3] I consider alternative definitions in robustness tests. Previous evidence suggests that value stocks earn higher returns because they are more highly levered than growth stocks. Therefore, FLEV should be positively related to the value premium.

[3] Bhandari (1988), Ozdagli (2012), Cao (2015), and others use total liabilities as a proxy instead of book debt. One issue with this is total liabilities includes non-debt items such as accounts payable, accrued expenses, and accrued taxes, and therefore is not a clean proxy for market debt.

1.3.2.4 Human Capital

Human capital is a measure of the economic value of the skills, experience, and knowledge of a firm's employees. If labour markets are efficient, one would expect employees with relatively more human capital to receive higher labour income. Therefore, following Hansson (2004) and Sylvain (2014), my measure of human capital risk is related to the wage bill of the firm. Specifically, the average salary per employee within a firm is used as a proxy for human capital risk exposure. I take the total wage bill (item XLR), defined as the sum of salaries, wages, pension costs, profit sharing and incentive compensation, payroll taxes and other employee benefits, and scale it by number of employees (item EMP) to calculate salary per worker (SPW). SPW is then adjusted for inflation using the FRED GDP deflator (series GDPDEF), and used as a firm-specific proxy for human capital. It should be positively correlated with the value premium to reflect the fact that value firms typically have more human capital risk exposure.

A second potential proxy is the average wage growth (AWG) measure used by Hansson (2004), which is simply the annual percentage change in SPW. However, SPW is preferred to AWG because it is more intuitive to focus on a variable measured in levels when analysing the cross-section of stock returns as opposed to growth rates which would reflect the *change* in human capital.

1.3.2.5 Institutional Ownership

To examine whether institutional ownership (IO) is able to explain the value premium, I obtain common-stock holdings data from the Thomson Reuters Institutional (13F) Holdings database. This database, described in detail by Gompers and Metrick (2001), contains quarterly stock-level holdings data from Q1 1980 to Q4 2019 for all banks, insurance companies, mutual funds, investment advisors, and university endowments that are required to file a 13F report with the SEC.[4] IO for stock i in month t is calculated as the most recently observable number of shares outstanding held by institutions (assuming a three month reporting lag) divided by the CRSP number of shares outstanding in month t. If a stock is not listed in the Institutional Holdings database in a particular quarter it is assumed to have a zero institutional stock holding.

[4]Only institutions with greater than $100 million of securities under discretionary management are required to submit a 13F, and common-stock holdings less than 10,000 shares or $200,000 are not required to be disclosed.

1.3.2.6 Operating Leverage

Operating leverage comes in two closely related forms. The classical definition of operating leverage (OL) is the ratio of a firm's fixed costs to total costs. If a firm has higher OL, then the marginal profit earned from an increase in sales is high as variable costs represent a low proportion of the firm's cost structure. Following Gorodnichenko and Weber (2016), I calculate fixed costs as the sum of selling, general and administrative expenses (item XSGA), advertising (item XAD), and R&D expenses (item XRD). Total costs are fixed costs plus variable costs, where variable costs are proxied by a firm's cost of goods sold (item COGS). COGS directly depends on the firm's level of production, and therefore should be a reasonable proxy for variable costs.

Similarly, the degree of operating leverage (DOL) directly measures how sensitive a firm's operating profit is to a change in sales in a regression framework. Following Garcia-Feijoo and Jorgensen (2010), to estimate DOL I first detrend EBIT and SALES (item SALEQ) by running the following regressions of each series on a time trend for each stock over a rolling five year window using quarterly data:

$$\log \text{EBIT}_t = \log \text{EBIT}_0 + g_{\text{EBIT}} \times t + u_{t,\text{EBIT}}, \tag{1.3}$$

$$\log \text{SALES}_t = \log \text{SALES}_0 + g_{\text{SALES}} \times t + u_{t,\text{SALES}}, \tag{1.4}$$

where the residuals represent the detrended series. Quarterly EBIT is manually calculated as operating income before depreciation (item OIBDPQ) less depreciation and amortisation over the quarter (item DPQ). Since EBIT can be negative, I apply the standard transformation which treats positive and negative earnings symmetrically:

$$\log \text{EBIT}_t = \begin{cases} \log(1 + \text{EBIT}_t) & \text{if EBIT}_t \geq 0 \\ -\log(1 - \text{EBIT}_t) & \text{otherwise,} \end{cases} \tag{1.5}$$

as in Garcia-Feijoo and Jorgensen (2010). The degree of operating leverage (DOL) is then measured as the absolute value of the slope coefficient from a regression of $u_{t,\text{EBIT}}$ on $u_{t,\text{SALES}}$ (without an

intercept), as in equation (1.6):

$$u_{t,\text{EBIT}} = \text{DOL} \times u_{t,\text{SALES}} + e_t. \tag{1.6}$$

DOL measures the percentage change in EBIT given a one percent change in SALES. High BM firms should have a higher DOL coefficient, reflecting the arguments in Carlson, Fisher, and Giammarino (2004) and Obreja (2013) that the value premium is driven by operating leverage.

Novy-Marx (2011) argues that operating leverage captures two distinct elements: (i) limited operational flexibility (generated by, for example, investment irreversibility), and (ii) total operating costs (e.g. the costs incurred in producing and selling goods). He argues and provides evidence that the total operating cost component of operating leverage is responsible for the value premium. Therefore, to re-examine Novy-Marx's proposition, I also define the total operating cost component of operating leverage (TOCOL) as the sum of cost of goods sold (item COGS) and selling and general administrative expenses (item XSGA), scaled by total assets, as in his paper.

As with adjustment costs, these three proxies are aggregated into a composite measure of operating leverage denoted OLEV using PCA. OLEV is calculated as the standardised (subtract the mean and divide by the standard deviation) first principal component of OL, DOL, and TOCOL using all data up to time t:

$$\text{OLEV}_{it} = \omega_{1t}\text{OL}_{it} + \omega_{2t}\text{DOL}_{it} + \omega_{3t}\text{TOCOL}_{it}. \tag{1.7}$$

The eigenvalues of the three components over the full sample are 1.41, 1.00, and 0.59, explaining 47%, 33%, and 20% of the variance in the data respectively. A higher value of OLEV can be interpreted as the firm having higher operating leverage, and hence should be positively related to the value premium.

1.4 Decomposition Methodology

1.4.1 Summary of Hou and Loh (2016)

Hou and Loh (2016) develop a methodology to evaluate a large number of potential explanations for an asset pricing anomaly. The methodology starts by estimating a set of univariate cross-sectional regressions of returns on some variable related to an asset pricing anomaly (e.g. the book-to-market ratio), and then decomposes the estimated coefficient into contributions from one or more candidate explanatory variables, and an unexplained residual component. They apply this methodology to investigate a range of explanations for the idiosyncratic volatility puzzle documented by Ang, Hodrick, Xing, and Zhang (2006) and others. To illustrate how the methodology can be applied to investigate the value premium, consider the following set of cross-sectional regressions, one for each time $t = 2, \ldots, T$:

$$r_{it} \;=\; \alpha_t + \beta_t \mathrm{BM}_{it-1} + \varepsilon_{it}, \quad i = 1, \ldots, n, \tag{1.8}$$

where r_{it} is the Daniel, Grinblatt, Titman, and Wermers (1997) (DGTW) characteristic-adjusted return on asset i from time $t-1$ to t, BM_{it-1} is the book-to-market ratio of asset i at time $t-1$, and ε_{it} is an error term that is assumed to satisfy the usual Gauss-Markov assumptions. The slope coefficient β_t can be interpreted as a measure of the strength of value premium because it quantifies the predictive relationship between BM and future returns. This first-stage regression captures the dynamic relationship between returns and BM by allowing the regression coefficients to be time-varying. The standard errors can be estimated from the time series distribution of the coefficients, as in Fama and MacBeth (1973).

Now suppose there is a set of k firm-specific candidate variables $C_{it}^1, \ldots, C_{it}^k$ (e.g. adjustment costs, financial leverage) that have been proposed to explain the value premium. The relationship between BM and the candidate variables can also be expressed as a set of $T-1$ cross-sectional regressions for $t = 1, \ldots, T-1$:

$$\mathrm{BM}_{it} \;=\; \gamma_{0t} + \gamma_{1t} C_{it}^1 + \cdots + \gamma_{kt} C_{it}^k + u_{it}, \quad i = 1, \ldots, n. \tag{1.9}$$

The final step decomposes the estimate of β_t into the contribution from each of the candidate variables and a residual component left unexplained. We can write the slope coefficient in the first-stage regression as:

$$\begin{aligned}
\beta_t &= \frac{\text{cov}(r_{it}, \text{BM}_{it-1})}{\text{var}(\text{BM}_{it-1})} \\
&= \frac{\text{cov}(r_{it}, \gamma_{0t-1} + \gamma_{1t-1}C^1_{it-1} + \cdots + \gamma_{kt-1}C^k_{it-1} + u_{it-1})}{\text{var}(\text{BM}_{it-1})} \\
&= \frac{\gamma_{1t-1}\text{cov}(r_{it}, C^1_{it-1}) + \cdots + \gamma_{kt-1}\text{cov}(r_{it}, C^k_{it-1}) + \text{cov}(r_{it}, u_{it-1})}{\text{var}(\text{BM}_{it-1})} \\
&= \gamma_{1t-1}\frac{\text{cov}(r_{it}, C^1_{it-1})}{\text{var}(\text{BM}_{it-1})} + \cdots + \gamma_{kt-1}\frac{\text{cov}(r_{it}, C^k_{it-1})}{\text{var}(\text{BM}_{it-1})} + \frac{\text{cov}(r_{it}, u_{it-1})}{\text{var}(\text{BM}_{it-1})} \\
&\equiv \beta^1_t + \cdots + \beta^k_t + \beta^u_t, \qquad (1.10)
\end{aligned}$$

where the second step substitutes in (1.9), and the third step uses the linearity property of the covariance operator. Dividing the last line of (1.10) by β_t gives the relative proportion of the relationship between returns and BM explained by each candidate variable, as well as the residual component left unexplained. By construction, the proportions sum to one, and hence can be interpreted as a percentage of the value premium explained.

There are two points worth noting about this methodology. First, a candidate variable that is highly correlated with the book-to-market ratio may not necessarily explain a large proportion of the value premium. This is because the part of the candidate variable that is correlated with BM need not necessarily be correlated with stock returns, and it is the relationship between both that dictates the amount of explanatory power each candidate variable has. That is, γ_{jt-1} may be large, but $\text{cov}(r_{it}, C^j_{it-1})$ may not. This is verified in the empirical analysis where a number of variables that are highly correlated with BM explain a small, and in some cases negative, proportion of the value premium. Second, Hou and Loh (2016) show that this methodology is directly related to the conventional approach of evaluating explanations for an anomaly. The conventional approach involves adding the candidate variable as a control to the Fama and MacBeth (1973) cross-sectional regressions, and testing whether the coefficient on BM remains statistically significant. If adding the candidate variable to the regression causes BM to lose significance, then one may conclude it is

able to explain the value premium. The advantage of the decomposition methodology is it tells us 'how much' a candidate can explain, which facilitates a straightforward comparison of competing explanations not possible with the conventional approach.

1.4.2 Extension to Multiple First-Stage Regressors

The Hou and Loh (2016) methodology requires the first-stage cross-sectional regression to have only one independent variable, so that β_t can be easily decomposed using the OLS formula for the slope coefficient in a univariate regression. In this subsection, I appeal to a basic result in econometrics to show that this decomposition methodology works in the case of multiple first-stage regressors. This is useful because it allows us to add a large number of control variables to the first-stage regression, without relying on DGTW characteristic-based benchmarks. For example, if we want to simultaneously control for the size, market, profitability, and investment factors of Fama and French (2015) when estimating the coefficient on BM, then the DGTW portfolios will be more sparsely populated the finer the benchmark portfolios are constructed due to the curse of dimensionality. But if the portfolios are sorted on several variables into, say, terciles, then the accuracy of each benchmark portfolio is diminished. This problem scales with the number of control variables added, so generalising the methodology to allow for multiple regressors in the first-stage regression is practically useful for mitigating omitted variables bias, and allows the methodology to be applied in other contexts where many control variables are necessary.

Consider a general cross-sectional regression of raw (not DGTW-adjusted) returns on lagged BM and a vector $\boldsymbol{x}$ of m lagged control variables (including a constant):

$$r_{it} \;=\; \beta_t \mathrm{BM}_{it-1} + \boldsymbol{\delta}_t \boldsymbol{x}_{it-1} + e_{it}, \quad i = 1, \ldots, n. \tag{1.11}$$

The Frisch-Waugh-Lovell theorem says we can consistently estimate β_t using the following procedure. First, regress r_{it} on $\boldsymbol{x}_{it-1}$ in each month and obtain the residuals, $\tilde{r}_{it}$ for each asset i. Then, regress BM_{it-1} on $\boldsymbol{x}_{it-1}$ in each month and obtain the residuals $\widetilde{\mathrm{BM}}_{it-1}$ for each i. Finally, regress the orthogonalised returns $\tilde{r}_{it}$ on the orthogonalised book-to-market ratio $\widetilde{\mathrm{BM}}_{it-1}$ (without a constant) in each month and obtain a time series of estimates of β_t. The values of β_t estimated with this

procedure are numerically identical to those from estimating model (1.11) directly.[5]

The second-stage cross-sectional regression can be expressed as:

$$\widetilde{\mathrm{BM}}_{it} \;=\; \gamma_{1t}\widetilde{C}^1_{it} + \cdots + \gamma_{kt}\widetilde{C}^k_{it} + \tilde{u}_{it}, \quad i = 1, \ldots, n, \tag{1.12}$$

which follows from orthogonalising (1.9).[6] As before, β_t can be decomposed as:

$$\begin{aligned}
\beta_t &= \frac{\sum_{i=1}^n \widetilde{\mathrm{BM}}_{it-1}\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} \\
&= \frac{\sum_{i=1}^n \left(\gamma_{1t-1}\widetilde{C}^1_{it-1} + \cdots + \gamma_{kt-1}\widetilde{C}^k_{it-1} + \tilde{u}_{it-1}\right)\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} \\
&= \frac{\gamma_{1t-1}\sum_{i=1}^n \widetilde{C}^1_{it-1}\tilde{r}_{it} + \cdots + \gamma_{kt-1}\sum_{i=1}^n \widetilde{C}^k_{it-1}\tilde{r}_{it} + \sum_{i=1}^n \tilde{u}_{it-1}\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} \\
&= \gamma_{1t-1}\frac{\sum_{i=1}^n \widetilde{C}^1_{it-1}\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} + \cdots + \gamma_{kt-1}\frac{\sum_{i=1}^n \widetilde{C}^k_{it-1}\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} + \frac{\sum_{i=1}^n \tilde{u}_{it-1}\tilde{r}_{it}}{\sum_{i=1}^n \widetilde{\mathrm{BM}}^2_{it-1}} \\
&\equiv \tilde{\beta}^1_t + \cdots + \tilde{\beta}^k_t + \tilde{\beta}^u_t.
\end{aligned} \tag{1.13}$$

The proportion of the value premium explained by each candidate variable can then be found by dividing the last line of (1.13) by β_t. For the purpose of inference, the means and standard errors of the proportions $\tilde{\beta}^i_t/\beta_t$ (for $i = 1, \ldots, k, u$) are unobtainable in closed-form. However, since the proportions are simply ratios of two random variables whose distributions are known, their moments can be approximated using the multivariate delta method (see Appendix A.2).

The Hou and Loh (2016) methodology is nested in this methodology. To see this, note that when $\boldsymbol{x}_{it-1} = 1$, i.e. the vector of controls contains only an intercept, then the orthogonalisation procedure simply demeans the returns and book-to-market ratio, and the first line of (1.13) collapses to the well-known formula in the first line of (1.10). In the empirical analysis, I apply this generalised decomposition methodology to evaluate competing explanations of the value premium controlling for firm size and past returns (momentum).

[5] See Lovell (2008) for a simple proof of the theorem.
[6] See Appendix A.1 for a proof.

1.5 Empirical Analysis

1.5.1 Descriptive Statistics

Table 1.1 provides descriptive statistics for all of the main variables used in this chapter, and Figures 1.1 and 1.2 plot the time series and cross-sectional distributions of these variables. The average BM ratio is 0.73, and exhibits wide cross-sectional variation as illustrated by the spread in the 5th (0.11) and 95th (1.97) percentiles, and the standard deviation of 0.61. The average ME is around \$3.4bn, but there is substantial heterogeneity in firm size, with firms ranging from microcaps to the largest S&P 500 constituents. The distribution of firm size is naturally positively skewed and exhibits very high kurtosis. The average return is positive at 2.00 percent per month, with a standard deviation of 13 percent.

The interpretation of ADJCO and OLEV is not straightforward as they represent the first principal component of three adjustment cost and three operating leverage proxies respectively. The average firm has financial leverage (FLEV) of 22 percent, and the distribution is slightly positively skewed. The average firm has around 50 percent of its shares held by institutional investors, but many stocks in the sample are fully held by individual investors and have zero IO. The mean of salary per worker (SPW) is around \$30,500. Standardised unexpected earnings (SUE) is positive on average at 0.25, implying earning surprises are typically good news for investors. Due to the availability of data, the number of observations for some variables is significantly lower than the full sample of 1,260,986. Particularly ADJCO, OLEV, and SPW have a smaller number of observations, since data on the depreciation, advertising expenses, and the wage bill are not widely reported in Compustat.

Table 1.2 shows the pairwise correlations for each set of variables. Consistent with intuition, BM is negatively correlated with ADJCO as this variable proxies for investment reversibility, which should be lower for high BM firms. The correlation between BM and FLEV is strong at 0.58, which is a well established finding in the capital structure literature (e.g. Rajan and Zingales, 1995; Chen and Zhao, 2006). BM is negatively correlated with IO, implying institutional investors typically hold growth stocks with lower book-to-market ratios. Surprisingly BM is highly negatively correlated with OLEV with a correlation of -0.31, suggesting value firms typically have lower operating leverage than growth firms. SPW and SUE are negatively correlated with BM, but the values are small in

magnitude. The correlations between returns and the candidate variables are all less than 0.06 in absolute value.

1.5.2 Univariate Sorts

As a precursor to the main analysis, I sort the sample by BM into value- and equally-weighted quintile portfolios, and calculate the value of each variable for each portfolio. These univariate sorts can help shed light on the relationship between the book-to-market ratio, average returns, and each of the candidate explanatory variables by reducing the noise inherent in pairwise correlations.

Both panels of Table 1.3 verify the existence of a value premium in this sample, as average returns are monotonically increasing in the book-to-market ratio for both the value- and equally-weighted portfolios. The difference in time series means between the highest and lowest BM quintiles is 1.07 percent per month (12.84 percent per year) for the value-weighted portfolios and 1.03 percent per month (12.36 percent per year) for the equally-weighted portfolios. Both of the differences are highly significant using Newey and West (1987) standard errors. Interestingly, ME is monotonically decreasing in BM for both the value- and equally-weighted portfolios, implying high BM firms are typically smaller in terms of market capitalisation. This motivates the importance of controlling for firm size in the empirical analysis, to ensure the results are not contaminated by the size effect of Banz (1981). MOM is lower for higher BM quintiles, since firms that have had high past 12 month cumulative returns will have higher market capitalisations and lower book-to-market ratios (all else equal), and therefore will move to a lower BM quintile when the portfolios are rebalanced.

The sorts also reveal interesting patterns in the candidate variables. ADJCO is typically decreasing as one moves from the growth to the value portfolio, consistent with the intuition that having a low book value of assets decreases adjustment costs. FLEV is strictly increasing across the portfolios, suggesting that high BM firms are more financially levered, as predicted by theory. However, SUE is monotonically decreasing in the BM quintiles, implying that firms with higher BM ratios have lower and often *negative* earnings surprises. This provides initial suggestive evidence against behavioural explanations of the value premium. OLEV is also lower for high BM firms. There appears to be a flat relationship between IO and BM, as there is no clear pattern across the quintiles. All of the differences in time series means of the candidate variables are highly significant, barring SPW.

1.5.3 Decomposing the Value Premium

In this section, I apply the generalised decomposition methodology of Section 1.4.2 to decompose the value premium. First, I examine each candidate variable in isolation, after which I turn to a multivariate decomposition. In all of the regressions, I include the natural log of a stock's market capitalisation in the previous month to control for the size effect of Banz (1981), and the previous 12 month cumulative return to control for the momentum effect of Jegadeesh and Titman (1993).

1.5.3.1 Univariate Analysis

Table 1.4 shows the results of the univariate decomposition, broken into the three stages. The orthogonalisation procedure in stage one is similar for all of the variables, and the numbers only differ because each regression has a different sample size. Regressing returns on the control variables highlights the negative relation between returns and size, which is significant in all cases. The coefficient on MOM is consistently not significantly different from zero. The negative relationship between log(BM) and size and momentum is mechanical, since all else equal, higher previous 12 month returns increases ME, mechanically increasing the denominator of the book-to-market equity ratio. The final part of stage one confirms the value premium persists after controlling for size and momentum, as the slope from regressing $\tilde{r}$ on $\widetilde{\log(\text{BM})}$ is positive and significant in all samples.

In the second stage, the candidate variable is regressed on the control variables, and then the orthogonalised log BM ratio is regressed on the orthogonalised candidate variable. All of the candidate variables considered are significantly related to log firm size, and some appear to be related to MOM. There is evidence of a relationship between the orthogonalised BM and the orthogonalised candidate variables in all cases.

In the third stage, the coefficient from regressing orthogonalised returns on the orthogonalised book-to-market ratio is decomposed into the contribution from the candidate variable and the residual component. This is the main result of the table, and is also plotted in Figure 1.3. The decomposition results do not offer support for most of the variables considered in this chapter. ADJCO, SPW, and SUE are all insignificant, and IO, OLEV, and SUE even have the wrong sign. However, the failure of using ADJCO based on DEP and RENT is consistent with the evidence in Cao (2015), who finds that using DEP, RENT, and TANG as proxies for adjustment costs supports

the existence of a *growth* premium. Taken together, the evidence does not support an adjustment cost explanation of the value premium.

SPW is able to explain 9 percent of the value premium, suggesting there may be a weak role for human capital, but it is far from the dominant factor. IO, OLEV, and SUE explain -12, -26, and -3 percent of the value premium respectively. A negative contribution implies there should be a return premium on growth stocks, which contradicts the well documented evidence of a value premium. Ultimately, the data suggests that behavioural stories, and stories based on human capital or operating leverage are inadequate explanations of the value premium.

The most interesting result in Table 1.4 is that FLEV is able to explain a significant proportion of the value premium. In this univariate decomposition, FLEV can explain 56 percent of the value premium, which implies that cross-sectional differences in financial leverage are a primary driver of the difference in returns between value and growth stocks. This finding lends support to the theory of Ozdagli (2012), who theoretically shows that financial leverage is the dominant factor, although the results in Table 1.4 suggest it is not able to fully capture the value premium.

To gain further insight into the explanatory power of each candidate variable, Figure 1.4 plots the evolution of the percentage of the value premium explained over time for each of the candidate variables. To reduce noise, a 20 year rolling period is used to compute the time series means and standard errors, which rolls forward on a monthly basis. The shaded areas represent 95% confidence intervals around the 20 year time series mean. The figures confirm that the proportion explained is relatively stable over time for most variables. ADJCO and IO fluctuate by less than around 10 percent over the entire period. FLEV has a minimum of around 40 percent, and a maximum of slightly above 60 percent, confirming the persistence in the explanatory power of financial leverage. OLEV exhibits significant variation, with a high of close to zero, and a low of -100 percent in 2009. OLEV has very wide confidence bands, so it is never significantly different from zero. SPW exhibits a downwards trend, starting at a promising 30 percent but progressively moving closer to zero in later time periods. Interestingly, the proportion explained by SUE was negative until 2000 when there were spikes which brought SUE greater than zero in later years.

1.5.3.2 Multivariate Analysis

The previous section showed that, of all the variables considered, only FLEV was able to explain a large proportion of the value premium in a univariate setting. In Table 1.5, I consider FLEV together with each other variable, along with two models including several candidate variables, to see if the results in Table 1.4 are robust to relationships between the candidate variables themselves.[7] To save space, only the third-stage decomposition results are reported.

The results confirm the robustness of FLEV, which is always significant and ranges from 38 to 58 percent depending on the model used. The only other variables that are marginally significant alongside FLEV are IO and OLEV. In the final models (6) and (7) which include multiple candidate variables, only FLEV, OLEV, and SUE are significant. Including FLEV, IO, OLEV, and SUE jointly brings the proportion of the value premium FLEV can explain down to its lowest level at 38 percent. These results are expected since FLEV is not highly correlated with the other candidate variables, so including them in the analysis does not greatly affect its explanatory power. Overall, the results suggest that no combination of variables considered in this chapter are able to fully explain the value premium, but the most promising explanation is one based on financial leverage.

1.5.4 Asset Pricing Tests

In this section I examine explanations of the value premium using an asset pricing factor approach. For each candidate variable I sort the universe of stocks in the sample into quintile portfolios, and compute the value-weighted average return for each portfolio. The factors go long the portfolio with the theoretically higher return (i.e. high financial leverage, low institutional ownership, etc.) and short the portfolio with the theoretically lower return in a dollar-neutral fashion. Portfolios are rebalanced monthly. The tests run time series regressions of the value factor (HML) on combinations of the other factors to see whether the alpha of HML can be fully or partially explained:

$$\text{HML}_t \;=\; \alpha + \sum_{c \in C} \beta_c \times f_{c,t} + \varepsilon_{c,t}, \tag{1.14}$$

[7] All candidate variables cannot be included in a single model, because the number of firms with nonmissing data for all variables is too small to yield meaningful results. Therefore, a combination of ADJCO, OLEV, and SPW have been dropped as these variables have the least number of observations.

for a set of candidate factors (f_c) defined by C. If an alpha in these regressions is small and insignificant there is evidence that the candidate variable factor is able to explain the value premium.

Table 1.6 reports the coefficients from these regressions. Model (1) is HML regressed on a constant, and represents the raw average return to the HML factor without controlling for other factors. The HML factor earns an average return of around 12 percent per annum, which acts as a benchmark to compare the alphas in the other regressions. Models (2)–(7) add each candidate variable as an additional factor. With the exception of model (3) which adds the FLEV factor, the alphas for the other univariate models all remain significantly positive, suggesting the other factors are not able to fully explain the alpha of the value factor. When FLEV is included, the alpha drops from 12 percent in model (1) to an insignificant 4 percent. In the final model when all factors are included, the alpha remains at an insignificant 4 percent, and the loadings suggest most of the explanatory power comes from the FLEV factor due to its higher slope coefficient.

1.5.5 The Role of Financial Leverage

Given the relative importance of financial leverage I delve deeper into this explanation in an attempt to shed further light on the financial distress hypothesis and its relation to the value premium. Given the high correlation between BM and FLEV, a natural question arises as to whether these variables are capturing the same phenomenon. That is to say, are BM and FLEV mechanically related, in the sense that value firms always have high financial leverage, and growth firms always have low FLEV? Figure 1.5 plots overlapping histograms of the distribution of FLEV for the five BM portfolios described in Section 1.5.4. Each of the portfolios has companies spanning the spectrum of FLEV, especially Q5 which has the largest spread. Therefore it is not the case that high BM firms necessarily have higher financial leverage, suggesting both variables are not capturing the same thing.

Moreover, leverage can be decomposed into short-term and long-term components depending on the maturity of the debt. If FLEV is capturing financial distress risk, then short-term leverage should be more impactful than long-term leverage, all else equal, because it is maturing sooner and requires more immediate cash flow liquidity. Therefore, firms with relatively more short-term debt are riskier from the perspective of stockholders, and should command a higher risk premium.

Choe (1994) finds that an increase in short-term debt that displaces the same amount of long-term debt increases a stock's risk because of the substitution effect from debtholders to stockholders. I therefore construct two further measures of FLEV denoted short-term financial leverage (STFLEV) and long-term financial leverage (LTFLEV) as:

$$\text{STFLEV} = \frac{\text{CD}}{\text{CD} + \text{NCD} + \text{ME}}, \tag{1.15}$$

$$\text{LTFLEV} = \frac{\text{NCD}}{\text{CD} + \text{NCD} + \text{ME}}, \tag{1.16}$$

where the variables are defined as in equation (1.2). Based on standard accounting conventions the maturity breakpoint separating short-term and long-term debt is one year. By definition, the sum of STFLEV and LTFLEV equals FLEV.

Table 1.7 details selected summary statistics and abbreviated results from applying the decomposition methodology of Section 1.4.2 to STFLEV and LTFLEV, along with the previous results for FLEV for comparison. On average most financial leverage is long-term at 18% compared to 4% for short-term, but most of the explanatory power of FLEV for the value premium is driven by the short-term component. In total FLEV explains 56% of the value premium, STFLEV explains 42%, and LTFLEV explains only 21%. These results are consistent with the financial distress hypothesis, since the results for STFLEV are stronger than for LTFLEV.

1.5.6 Robustness Tests

Portfolio level analysis. The baseline results in this chapter use a sample of individual stocks as the test assets to retain as much information as possible. But individual stocks are subject to idiosyncratic risk and measurement error, and individual stock returns are noisy. One solution is to sort stocks into portfolios, which diversifies away the idiosyncratic risk, but leads to a loss of information through reducing the cross-sectional variation in the candidate variables (Ang, Liu, and Schwarz, 2019). I consider four sorted portfolios: 25 and 50 BM sorted portfolios, and industries defined by both the 2-digit and 3-digit SIC codes.[8] Table 1.8 presents the third-stage decomposition

[8]The SIC code is a standardised system for grouping companies into industries. The 2-digit code defines the 'major' industry group, and the 3-digit code defines a more granular industry groups. In my sample there are 67 2-digit industry groups and 268 3-digit groups.

results for each of the candidate variables and sorted portfolios. The results are quantitatively very similar to those for individual stocks, with FLEV explaining between 45 and 67 percent of the value premium, and no other variables explaining a significant amount.

Alternative variable definitions. I also consider alternative definitions for some variables that can be measured in different ways. Specifically, I repeat the analysis using each component of ADJCO and OLEV separately (as opposed to the first principal component), replacing ME in FLEV with BE, and defining financial leverage as total debt (CD + NCD) divided by total assets. I also consider an alternative version of BE which subtracts the book value of preferred stock. In unreported results I verify that the results are qualitatively and quantitatively robust to these alternative definitions.

Comparison of methodologies. I compare the results of my generalised decomposition methodology with the methodology of Hou and Loh (2016) using DGTW-adjusted and raw returns and show there can be small differences in the amount of the anomaly explained. The DGTW characteristic-based benchmarks are constructed by conditionally sorting into $5 \times 5 \times 5$ size-BM-momentum portfolios, and equally weighting returns within each of the 125 portfolios. Stocks are then matched to a benchmark by size, BM, and momentum, and an adjusted return is calculated as the stock's return in excess of its characteristic-based benchmark. Table 1.9 summarises univariate decomposition results using the three methodologies. The methodology of Hou and Loh (2016) tends to overstate the proportion explained relative to the generalised decomposition methodology. For example, the amount explained by FLEV is up to 15 percent higher when using raw returns at 71 percent. Other variables show a similar pattern, albeit not as high. Overall the results suggest it is worth using the generalised decomposition methodology due to its advantages and quantitative differences over the Hou and Loh (2016) approach.

1.6 Conclusion

In this chapter, I generalise the decomposition methodology of Hou and Loh (2016) and apply it to evaluate competing explanations of the value premium. Building on previous theoretical and empirical literature, I identify six potential explanations that claim to be able to explain the value premium: adjustment costs, financial leverage, institutional ownership, operating leverage, human capital, and investor optimism. I first evaluate whether there is a relationship between these candidate variables and the book-to-market ratio by sorting the sample into quintiles and examining the characteristics of the resulting portfolios. Consistent with theory, adjustment costs, financial leverage, and returns are larger for high book-to-market portfolios. Operating leverage and earnings surprises are lower for value firms, and there is no strong relationship between institutional ownership or human capital and the book-to-market ratio.

I then apply the generalised decomposition methodology to evaluate the explanatory power of each of the candidate variables in a univariate and multivariate setting. The results suggest that of all the variables considered in this chapter, only financial leverage is able to explain a significant proportion of the value premium, ranging from around 40 to 60 percent depending on the model used. This result holds in univariate and multivariate decompositions, and for individual stocks and book-to-market and industry sorted portfolios. A factor based on financial leverage is able to explain the alpha of a high-minus-low book-to-market factor, but other factors cannot. Finally, I show that most of the explanatory power of financial leverage is concentrated in short-term leverage, implying the value premium may manifest from financial distress risk. Overall the results in this chapter suggest that the value premium puzzle has not been fully resolved, but an explanation based on financial leverage is the most promising.

Table 1.1: Descriptive statistics. This table reports descriptive statistics for all of the main variables used in this chapter. ADJCO proxies for adjustment costs and is the first principal component of a firm's depreciation ratio, rental expense, and asset tangibility ratio. BM is the book-to-market equity ratio. FLEV is financial leverage, defined as current and noncurrent debt, divided by current and noncurrent debt plus ME. IO is institutional ownership, defined as the percentage of total shares outstanding held by institutions. ME is market capitalisation (in \$bn). MOM is the cumulative return for months $t-1$ to $t-12$. OLEV measures operating leverage and is the first principal component of a firm's fixed-to-total cost ratio, degree of operating leverage as in Garcia-Feijoo and Jorgensen (2010), and total operating cost component of operating leverage as in Novy-Marx (2011). RET is the monthly CRSP return. SPW is inflation-adjusted salary per worker (in \$000). SUE is standardised unexpected earnings, defined as earnings in quarter t less earnings four quarters ago, divided by the standard deviation of the difference over the previous eight quarters. Number of observations represents the number of firm-level months. All variables are winsorised at the 0.5th and 99.5th percentiles.

	ADJCO	BM	FLEV	IO	ME	MOM	OLEV	RET	SPW	SUE
No. observations	366,518	1,260,986	1,123,504	991,279	1,260,986	1,126,858	233,692	1,260,986	134,372	1,025,865
Start date	197007	196201	196301	198006	196201	196301	196507	196201	196501	196410
Mean	0.04	0.73	0.22	0.50	3.35	0.21	−0.02	0.02	30.53	0.25
Std. deviation	1.14	0.61	0.21	0.27	18.09	0.51	1.13	0.13	30.68	1.72
Skewness	0.57	1.82	0.80	−0.06	19.95	1.75	0.14	0.60	3.01	−0.31
Kurtosis	3.36	7.26	2.81	1.65	666.41	8.23	2.98	4.82	16.08	9.17
Min	−2.46	0.02	0.00	0.00	0.00	−0.68	−2.53	−0.35	1.92	−10.80
5th percentile	−1.69	0.11	0.01	0.06	0.02	−0.42	−1.93	−0.18	6.44	−2.33
25th percentile	−0.74	0.30	0.08	0.26	0.11	−0.11	−0.81	−0.05	11.30	−0.36
50th percentile	−0.06	0.56	0.20	0.51	0.38	0.12	0.01	0.01	20.37	0.20
75th percentile	0.68	0.96	0.38	0.73	1.39	0.41	0.75	0.08	39.08	0.92
95th percentile	2.18	1.97	0.66	0.92	11.57	1.16	1.78	0.24	85.38	2.85
Max	3.87	4.09	0.86	1.00	1,187.46	3.33	4.11	0.60	273.52	8.94

Table 1.2: Pairwise correlation matrix. This table reports pairwise correlations for all of the main variables used in this chapter. ADJCO proxies for adjustment costs and is the first principal component of a firm's depreciation ratio, rental expense, and asset tangibility ratio. BM is the book-to-market equity ratio. FLEV is financial leverage, defined as current and noncurrent debt, divided by current and noncurrent debt plus ME. IO is institutional ownership, defined as the percentage of total shares outstanding held by institutions. ME is market capitalisation (in \$bn). MOM is the cumulative return for months $t-1$ to $t-12$. OLEV measures operating leverage and is the first principal component of a firm's fixed-to-total cost ratio, degree of operating leverage as in Garcia-Feijoo and Jorgensen (2010), and total operating cost component of operating leverage as in Novy-Marx (2011). RET is the monthly CRSP return. SPW is inflation-adjusted salary per worker (in \$000). SUE is standardised unexpected earnings, defined as earnings in quarter t less earnings four quarters ago, divided by the standard deviation of the difference over the previous eight quarters. All variables are winsorised at the 0.5th and 99.5th percentiles.

	ADJCO	BM	FLEV	IO	ME	MOM	OLEV	RET	SPW	SUE
ADJCO	1.00									
BM	−0.17	1.00								
FLEV	−0.10	0.58	1.00							
IO	0.01	−0.22	0.11	1.00						
ME	−0.22	0.58	−0.05	−0.15	1.00					
MOM	0.07	−0.07	0.00	−0.04	−0.03	1.00				
OLEV	0.31	−0.31	0.07	0.15	−0.39	0.01	1.00			
RET	0.02	0.05	−0.01	−0.03	0.03	−0.01	0.01	1.00		
SPW	−0.14	−0.15	0.23	0.42	−0.07	0.03	0.25	0.01	1.00	
SUE	0.01	−0.05	0.01	−0.01	−0.03	0.06	−0.01	−0.00	−0.03	1.00

Table 1.3: Characteristics of book-to-market sorted portfolios. This table reports value- and equally-weighted returns (in percent) and time series averages of the main variables for BM sorted quintile portfolios. Portfolios are sorted on BM in June of year t and are held for twelve months until June of year $t+1$. The equally-weighted portfolios are rebalanced each month. Value − growth is the difference in time series means between the highest book-to-market quintile and the lowest book-to-market quintile. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses.***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	BM quintile						
	Growth	2	3	4	Value	Value − growth	t(Value − growth)
Panel A: Value-weighted							
ADJCO	0.11	−0.36	−0.71	−0.79	−0.81	−0.95***	(−11.14)
BM	0.18	0.39	0.60	0.86	1.45	1.27***	(28.19)
FLEV	0.08	0.15	0.22	0.29	0.41	0.33***	(33.01)
IO	0.47	0.46	0.47	0.45	0.42	−0.05***	(−6.39)
ME	5.61	5.27	3.62	3.03	1.99	−3.64***	(−7.01)
MOM	0.25	0.18	0.16	0.15	0.11	−0.14***	(−6.55)
OLEV	0.51	0.42	0.37	0.33	0.26	−0.25***	(−27.31)
RET	1.03	1.17	1.16	1.36	2.09	1.07***	(3.90)
SPW	46.16	46.81	46.13	48.21	46.38	0.22	(0.11)
SUE	0.92	0.43	0.35	0.15	−0.18	−1.10***	(−17.47)
Panel B: Equally-weighted							
ADJCO	0.49	0.05	−0.09	−0.15	−0.32	−0.72***	(−19.09)
BM	0.19	0.40	0.61	0.88	1.60	1.41***	(33.07)
FLEV	0.10	0.15	0.21	0.26	0.38	0.29***	(38.16)
IO	0.40	0.44	0.43	0.42	0.38	−0.02***	(−4.67)
ME	4.78	3.55	2.39	1.63	0.94	−3.68***	(−8.34)
MOM	0.33	0.23	0.19	0.16	0.13	−0.20***	(−9.43)
OLEV	0.50	0.43	0.39	0.36	0.33	−0.17***	(−42.93)
RET	1.45	1.47	1.62	1.84	2.44	1.03***	(5.68)
SPW	39.21	35.98	38.03	37.12	37.47	−1.75	(−1.24)
SUE	0.84	0.46	0.24	0.04	−0.21	−1.05***	(−24.42)

Table 1.4: Univariate decomposition. This table reports the results from applying the generalised decomposition methodology with a single candidate variable. Log(BM) is the log book-to-market ratio. Log(ME) is the natural log of the firm's market capitalisation (in \$bn) in month $t-1$. MOM is the cumulative return for months $t-1$ to $t-12$. ADJCO proxies for adjustment costs and is the first principal component of a firm's depreciation ratio, rental expense, and asset tangibility ratio. FLEV is financial leverage, defined as current and noncurrent debt, divided by current and noncurrent debt plus ME. IO is institutional ownership, defined as the percentage of total shares outstanding held by institutions. OLEV measures operating leverage and is the first principal component of a firm's fixed-to-total cost ratio, degree of operating leverage as in Garcia-Feijoo and Jorgensen (2010), and total operating cost component of operating leverage as in Novy-Marx (2011). SPW is inflation-adjusted salary per worker (in \$000). SUE is standardised unexpected earnings, defined as earnings in quarter t less earnings four quarters ago, divided by the standard deviation of the difference over the previous eight quarters. All variables are winsorised at the 0.5th and 99.5th percentiles. Coefficients from regressions involving returns are multiplied by 100. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

		ADJCO	FLEV	IO	OLEV	SPW	SUE
First stage							
Return on controls	Intercept	1.05^{***}	1.07^{***}	1.41^{***}	1.32^{***}	1.15^{***}	1.10^{***}
		(3.06)	(4.04)	(4.27)	(3.85)	(4.06)	(4.08)
	Log(ME)	-0.31^{***}	-0.33^{***}	-0.39^{***}	-0.30^{***}	-0.22^{***}	-0.32^{***}
		(−3.83)	(−4.45)	(−4.71)	(−3.70)	(−3.68)	(−4.37)
	MOM	0.26	0.16	−0.01	0.00	0.10	0.14
		(1.22)	(1.12)	(−0.09)	(−0.02)	(0.65)	(0.97)
Log(BM) on controls	Intercept	-0.56^{***}	-0.75^{***}	-0.86^{***}	-0.85^{***}	-0.58^{***}	-0.74^{***}
		(−3.15)	(−3.77)	(−3.52)	(−3.66)	(−3.65)	(−3.75)
	Log(ME)	-0.12^{***}	-0.14^{***}	-0.13^{***}	-0.15^{***}	-0.12^{***}	-0.15^{***}
		(−3.08)	(−3.79)	(−3.42)	(−3.63)	(−3.56)	(−3.80)
	MOM	-0.17^{***}	-0.23^{***}	-0.20^{***}	-0.21^{***}	-0.29^{***}	-0.23^{***}
		(−2.58)	(−3.58)	(−3.37)	(−3.44)	(−3.38)	(−3.52)
$\widetilde{\text{R}}$ on $\widetilde{\log(\text{BM})}$	Slope	0.33^{**}	0.22^{***}	0.16^{*}	0.24^{**}	0.17^{**}	0.22^{**}
		(2.27)	(2.66)	(1.90)	(2.47)	(1.98)	(2.51)
Second stage							
Candidate on controls	Intercept	-0.36^{***}	0.19^{***}	0.46^{***}	0.35^{***}	36.99^{***}	0.25^{***}
		(−2.73)	(3.81)	(3.48)	(3.36)	(3.41)	(2.68)
	Log(ME)	-0.14^{***}	-0.01^{***}	0.06^{***}	0.00^{**}	2.63^{***}	0.10^{***}
		(−3.04)	(−2.94)	(3.58)	(2.20)	(3.05)	(3.68)
	MOM	0.04	-0.02^{**}	-0.04^{***}	0.01	0.32	0.92^{***}
		(0.68)	(−2.42)	(−3.32)	(1.35)	(0.23)	(3.72)
$\widetilde{\text{Log(BM)}}$ on $\widetilde{C}$	Slope	-0.19^{***}	1.79^{***}	0.27^{***}	-1.27^{***}	0.02^{**}	-0.06^{***}
		(−3.15)	(3.90)	(3.24)	(−3.60)	(2.57)	(−3.66)
Third stage							
Decompose coefficient	Candidate	0.05	0.56^{***}	-0.12^{*}	-0.26^{*}	0.09	−0.03
		(0.83)	(6.38)	(−1.86)	(−1.70)	(1.41)	(−0.96)
	Residual	0.95^{***}	0.44^{***}	1.12^{***}	1.26^{***}	0.91^{***}	1.03^{***}
		(15.73)	(5.02)	(17.65)	(8.13)	(14.94)	(29.70)
No months		312	658	473	568	657	658
Avg. no. firms per month		1,174	1,707	2,096	479	205	1,559

Table 1.5: Multivariate decomposition. This table reports the results from applying the generalised decomposition methodology with multiple candidate variables. To conserve space, only the third-stage decomposition is reported. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ADJCO	0.04						
	(1.22)						
FLEV	0.54***	0.53***	0.39***	0.54***	0.58***	0.46***	0.38***
	(5.91)	(4.97)	(5.22)	(3.20)	(5.60)	(6.20)	(3.59)
IO		−0.11*				−0.02	−0.01
		(−1.92)				(−1.61)	(−0.42)
OLEV			−0.17*				0.09***
			(−1.72)				(2.59)
SPW				0.06			
				(1.40)			
SUE					−0.03	0.03	0.04*
					(−0.94)	(1.52)	(1.82)
Residual	0.42***	0.58***	0.78***	0.40**	0.44***	0.53***	0.56***
	(4.66)	(6.75)	(6.83)	(2.20)	(4.85)	(6.31)	(5.55)
No. months	312	473	568	657	658	473	473
Avg. no. firms per month	1,159	1,689	477	204	1,555	1,180	351

Table 1.6: Factor regressions. This table reports the results from regressing the value factor on long-short factors created from the individual candidate variables. α is the regression intercept, and the variable row names denote the regression slopes on those variables. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
α	0.12^{***}	0.09^{**}	0.04	0.05^{**}	0.12^{***}	0.12^{***}	0.13^{***}	0.04
	(2.99)	(2.49)	(1.09)	(2.05)	(3.13)	(2.99)	(3.45)	(0.95)
ADJCO		0.55^{***}						0.22^{*}
		(5.76)						(1.88)
FLEV			0.78^{***}					0.47^{***}
			(10.51)					(8.89)
IO				0.26^{**}				0.22^{**}
				(2.29)				(2.03)
OLEV					-0.31^{***}			-0.37^{***}
					(-3.65)			(-2.73)
SPW						0.10		-0.08^{*}
						(1.41)		(-1.70)
SUE							-0.78^{***}	-0.17
							(-7.16)	(-1.21)
Adj. R^2	0.00	0.18	0.55	0.02	0.07	0.00	0.15	0.53

Table 1.7: Short-term vs. long-term financial leverage. This table reports descriptive statistics and decomposition results for financial leverage related variables. FLEV is financial leverage, calculated as the sum of short-term and long-term debt divided by total debt plus market equity. STFLEV (LTFLEV) is short-term (long-term) financial leverage, calculated similar to FLEV but with short-term (long-term) debt only in the numerator. All variables are winsorised at the 0.5th and 99.5th percentiles. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

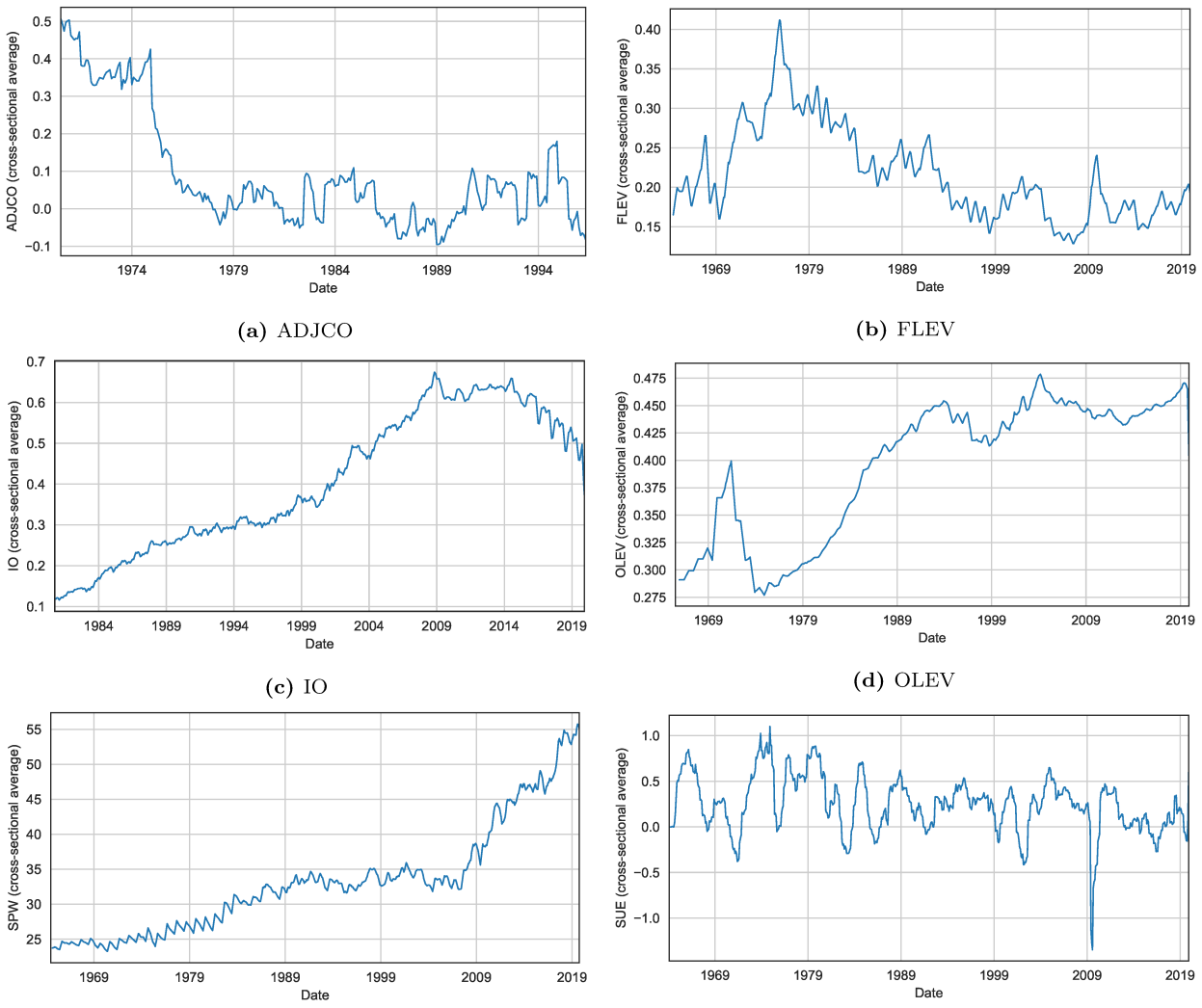
	Mean	Std. deviation	Prop. value premium explained	t-statistic
FLEV	0.22	0.21	0.56^{***}	(6.38)
STFLEV	0.04	0.07	0.42^{***}	(6.21)
LTFLEV	0.18	0.18	0.21^{***}	(4.84)

Table 1.8: Portfolio level decomposition. This table reports the results from applying the generalised decomposition methodology to 25 or 50 book-to-market sorted portfolios, and to industry portfolios defined by either the 2 or 3-digit SIC code. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	25 BM sorted portfolios		50 BM sorted portfolios		2-digit SIC industries		3-digit SIC industries	
	Prop. explained	Residual	Prop. explained	Residual	Prop. explained	Residual	Prop. explained	Residual
ADJCO	0.11	0.89***	0.13	0.87***	0.07	0.93***	0.08	0.92***
	(0.58)	(4.37)	(0.99)	(7.16)	(0.53)	(7.69)	(1.60)	(8.38)
FLEV	0.67***	0.33	0.57***	0.44	0.58***	0.42	0.45*	0.55
	(2.75)	(1.75)	(3.80)	(1.08)	(3.89)	(1.51)	(1.83)	(0.20)
IO	−0.08	1.08***	−0.10	1.10***	−0.09	1.09***	−0.03	1.03***
	(−0.24)	(3.13)	(−0.55)	(6.20)	(−0.55)	(5.78)	(−0.63)	(18.25)
OLEV	0.14	0.86**	0.15*	0.79***	0.22*	0.78***	−0.06	1.06***
	(1.13)	(2.33)	(1.69)	(6.88)	(1.76)	(6.39)	(−0.83)	(15.46)
SPW	0.05	0.95***	0.08	0.88***	0.12	0.88***	−0.03	1.03***
	(0.25)	(5.23)	(1.22)	(8.69)	(1.33)	(8.72)	(−0.56)	(17.13)
SUE	0.08	0.92***	0.09	0.89***	0.11	0.89***	0.00	1.00**
	(0.45)	(4.93)	(1.00)	(7.77)	(1.13)	(7.98)	(0.01)	(2.27)

Table 1.9: Comparison of decomposition methodologies. This table reports univariate decomposition results using the methodology of Hou and Loh (2016) with and without adjusted returns using Daniel, Grinblatt, Titman, and Wermers (1997) (DGTW) characteristic-based benchmarks, and the generalised decomposition methodology derived in this chapter. Newey and West (1987) heteroskedasticity and autocorrelation robust t-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	Hou and Loh (2016) (raw)		Hou and Loh (2016) (DGTW)		Generalised	
	Prop. explained	Residual	Prop. explained	Residual	Prop. explained	Residual
ADJCO	0.11	0.89***	0.08	0.92***	0.05	0.95***
	(1.54)	(12.90)	(1.41)	(13.33)	(0.83)	(15.73)
FLEV	0.71***	0.29**	0.67***	0.33**	0.56***	0.44***
	(11.98)	(2.49)	(9.87)	(2.62)	(6.38)	(5.02)
IO	0.01	0.99***	−0.10	1.10***	−0.12*	1.12***
	(0.35)	(16.92)	(−1.55)	(21.88)	(−1.86)	(17.65)
OLEV	−0.19*	1.19***	−0.20*	1.20***	−0.26*	1.26***
	(−1.92)	(11.97)	(−1.73)	(10.04)	(−1.70)	(8.13)
SPW	0.09	0.91***	0.08	0.92***	0.09	0.91***
	(1.38)	(15.81)	(1.21)	(17.05)	(1.41)	(14.94)
SUE	−0.05	1.05***	−0.05	1.05***	−0.03	1.03***
	(−0.68)	(24.69)	(−0.54)	(29.48)	(−0.96)	(29.70)



(a) ADJCO

(b) FLEV

(c) IO

(d) OLEV

(e) SPW

(f) SUE

Figure 1.1: Time series distribution of candidate variables. This figure plots cross-sectional averages on each month for each of the candidate variables.

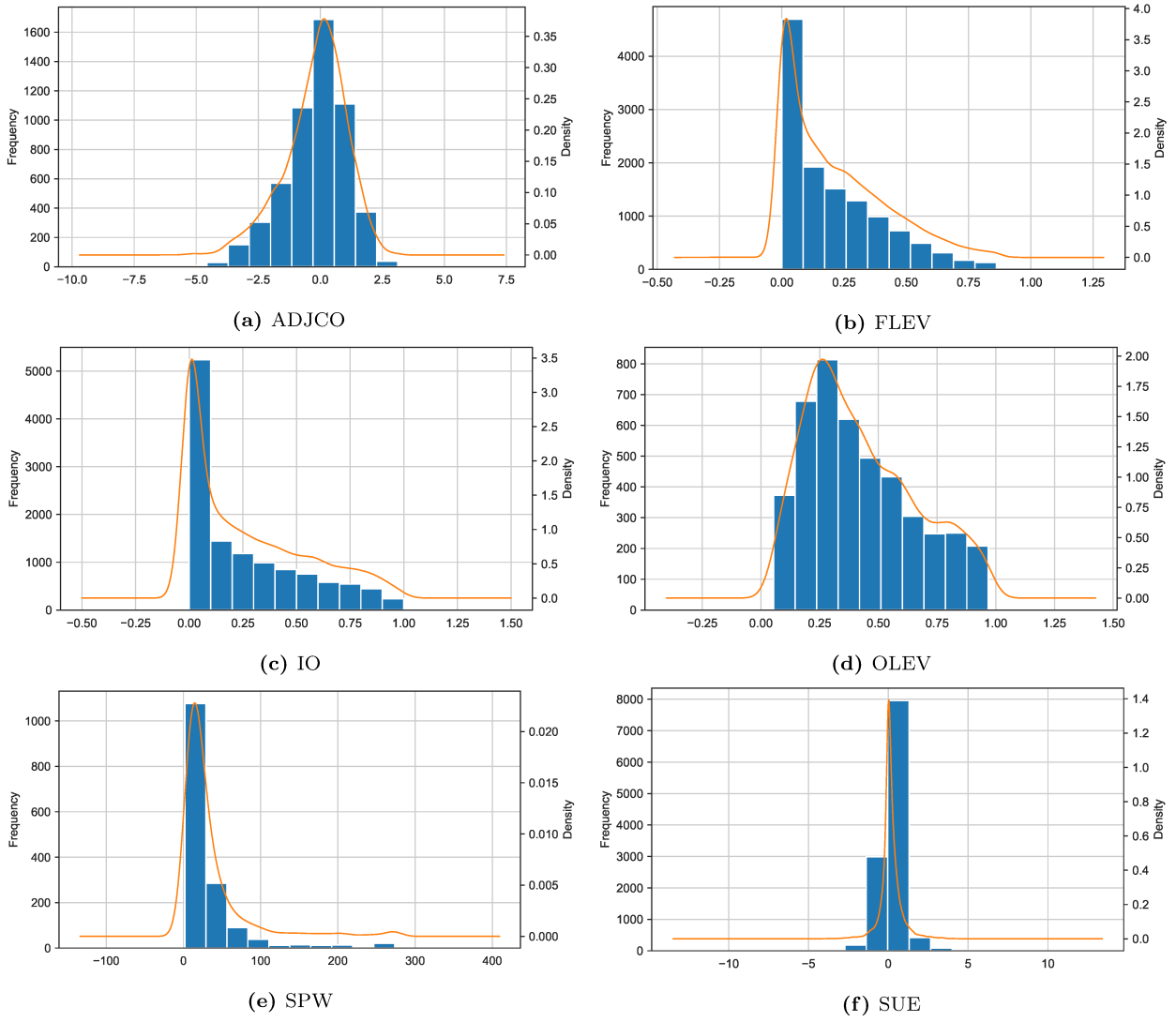


Figure 1.2: Cross-sectional distribution of candidate variables. This figure plots histograms and kernel densities of time series averages across each firm for each of the candidate variables. The left axis represents the histogram frequency, whilst the right axis represents the non-parametric Gaussian density with automatic bandwidth selection following Scott's rule.

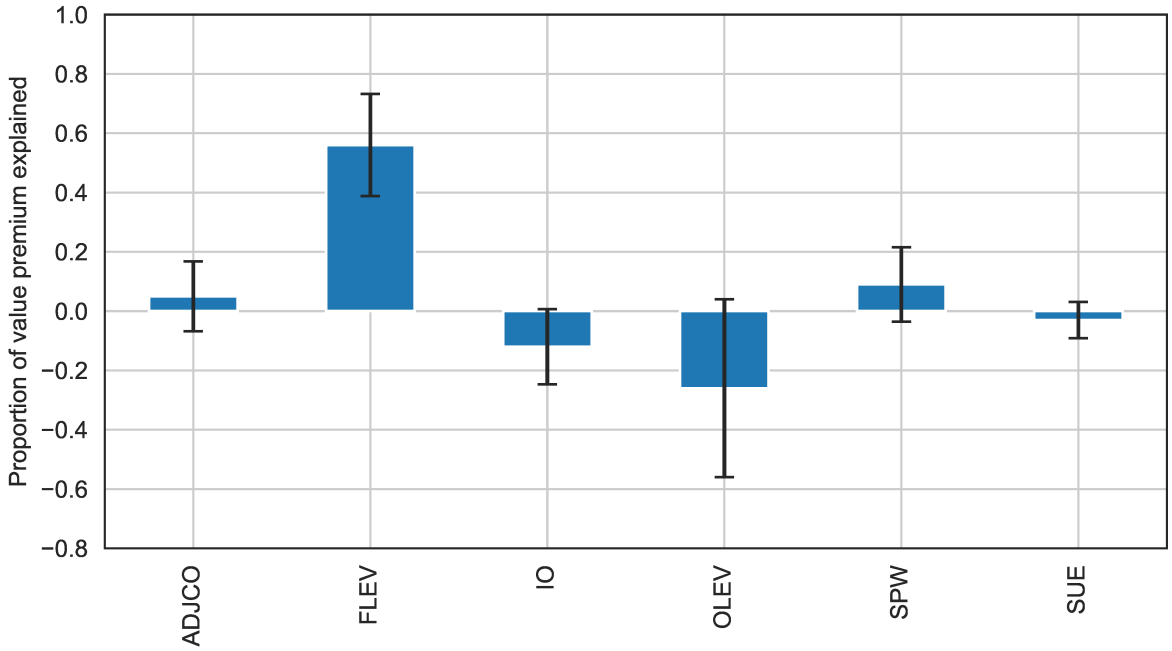
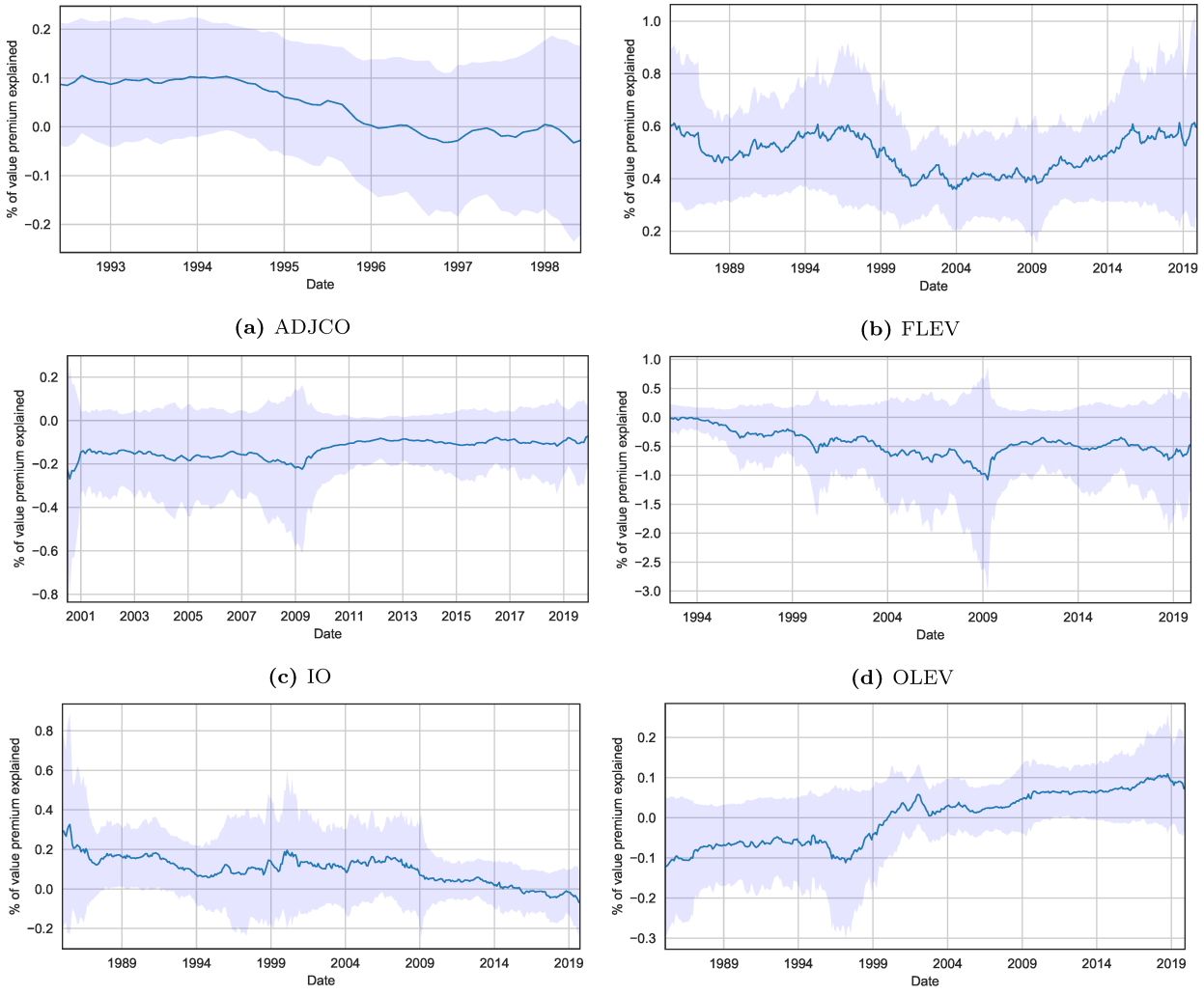


Figure 1.3: Univariate proportion of value premium explained. This figure plots the proportion of the value premium explained from the univariate decompositions for each of the six candidate variables. The black bars represent 95% confidence intervals around the estimated proportions.



(a) ADJCO

(b) FLEV

(c) IO

(d) OLEV

(e) SPW

(f) SUE

Figure 1.4: Proportion of the value premium explained over time. This figure plots the proportion of the value premium explained over time for each candidate variable. The solid blue line represents the time series mean over a 20 year rolling period, and the shaded areas are 95% confidence intervals around the time series mean.

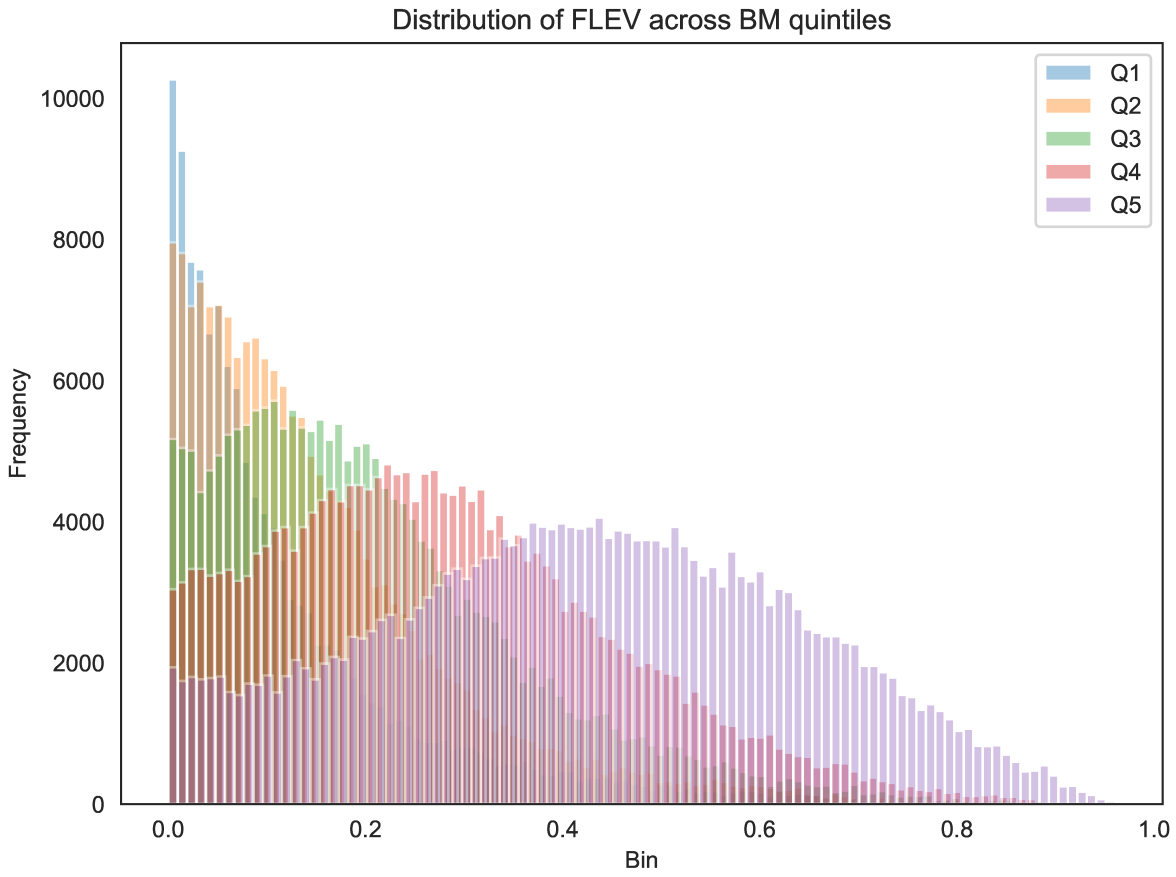


Figure 1.5: Distribution of FLEV across book-to-market quintiles. This figure plots overlapping histograms of financial leverage (FLEV) for five portfolios sorted by book-to-market ratio (BM). Q1 is the lowest BM quintile and Q5 is the highest BM quintile. By construction FLEV is capped between 0 and 1.

Chapter 2

Three Centuries of Machine Learning for Return Predictability

2.1 Introduction

The popularity of machine learning (ML) in finance has exploded in recent decades. Financial asset return predictability in particular has received a lot of attention, as it represents an apt training ground for ML algorithms due to their focus on out-of-sample (OOS) performance rather than in-sample (IS) statistical significance, and due to the wealth of potential predictors available to market participants at any point in time. The literature on return predictability with ML is diffuse, with different authors examining different methods, predictors, time frames, and asset classes. The aim of this chapter is to provide a comprehensive analysis of ML for return predictability, synthesising the literature and providing a set of benchmark results to guide future research and users of ML models.

I extend previous literature in three dimensions: time, asset classes, and geographical coverage. In the time dimension I utilise the longest possible time series of returns and macroeconomic variables available, with the earliest forecasts starting in the late 17th century for the UK, and the late 18th century for the US. This adds an additional 100+ years to typical sample sizes used in the literature, and allows me to capture many more states of nature, including business cycle turning points, and rare events such as wars or natural disasters. In the asset class dimension, I consider the three major asset classes: equities, bonds, and currencies. Previous work has primarily focused on equity markets, with limited attention being paid to bonds and currencies, despite the size and importance of these markets globally. Finally, I analyse eight countries with the longest stock and bond market

histories, and nine of the most heavily traded currency pairs. This broadens the scope of the current literature which has typically focused on the US. The ML algorithms I consider cover the most popular and widely used methods, and are fairly simple to understand and implement. There are many more sophisticated algorithms, such as those in the fields of deep or reinforcement learning, that are beyond the scope of this chapter and are left for future work.[1]

I show that, over the very long run, most ML models fail to outperform the historical mean. Most individual forecasts generate negative OOS R^2, implying the historical mean return is a better predictor of future returns than forecasts from the ML models. However, combining individual model forecasts into a composite forecast lowers prediction variability and performs better than individual forecasts, generating sizeable OOS gains. Imposing weak economic restrictions on the forecasts strengthens these results further. I consider several popular forecast weighting schemes, and show that composite forecasts weighted by each forecast's inverse mean squared error (MSE) over the recent past tend to perform the best.

Next I construct a dynamic trading strategy that aims to exploit return predictability, and show that material improvements in Sharpe ratios can be achieved when using the composite forecasts. The ML-based trading strategy outperforms the $1/N$ portfolio and the mean-variance optimal (MVO) portfolio using historical data, generating a Sharpe ratio of 1.04. Finally I report a plethora of robustness tests including sample splits by business cycle period, interest rate and inflation regime, market volatility, and pre- and post-WW2. The results suggest predictability is stronger in bad times and when market volatility is high, consistent with prior literature.

The chapter proceeds as follows. Section 2.2 reviews the main literature on return predictability and ML in finance. Section 2.3 describes the data. Section 2.4 outlines the forecasting framework and notation. Section 2.5 describes the ML models used in the analysis. Section 2.6 describes and presents results for the statistical evaluation measures used. Section 2.7 details the economic evaluation and implementation of the trading strategies. Finally, Section 2.8 concludes. Appendix A.3 contains additional information on the predictors used in this chapter.

[1] Heaton, Polson, and Witte (2016) review some applications of deep learning in quantitative finance. Kolm and Ritter (2020) give an overview of solving intertemporal choice problems in finance using reinforcement learning.

2.2 Literature Review

This chapter contributes to two strands of literature. The first focuses on OOS return predictability across multiple asset classes in the US and international markets, and the second on applying ML methods to finance.

2.2.1 Out-of-Sample Return Predictability

2.2.1.1 Equity Return Predictability

The seminal study on OOS equity premium predictability in the US is Goyal and Welch (2008), who comprehensively examine a range of predictors, including well-known variables like the price-dividend and book-to-market ratios. They document both poor IS and OOS performance of these predictors. They also highlight the instability of the predictions, with many variables seemingly only performing well due to the 1973–75 oil shock, and note that investors would not have been able to profitably time the market with these variables. One limitation of their analysis is they run all predictive tests using OLS time series regressions, which impose a strong linear structure of the form of the predictive relationship. This restriction can be overcome using nonlinear methods such as those commonly used in ML. Another limitation is they consider each forecast individually, whereas Rapach, Strauss, and Zhou (2010) advise combining model forecasts to reduce volatility and improve prediction accuracy. Doing so generates statistically and economically significant OOS gains.

A potential way to improve forecasts is through regularisation of model coefficients, and/or imposing additional structure on the model motivated by economic theory. Campbell and Thompson (2008) impose economic restrictions on the signs of coefficients and return forecasts, and show that a mean-variance investor can generate material improvements in utility even if OOS explanatory power is low.[2] Pettenuzzo, Timmermann, and Valkanov (2014) and Li and Tsiakas (2017) impose similar constraints, as well as bounds on the conditional equity Sharpe ratio, and find comparable improvements.

[2]The restrictions they consider include setting coefficients to zero if the sign is different from the theoretically expected sign, or setting equity premium forecasts to zero if they are negative.

2.2.1.2 Bond Return Predictability

Thornton and Valente (2012) investigate OOS bond return predictability using short-term and long-term forward rates, as in Fama and Bliss (1987) and Cochrane and Piazessi (2005). They show that in a mean-variance asset allocation exercise investors cannot realise higher Sharpe ratios or economic utility relative to a no-predictability expectations hypothesis benchmark. Sarno, Schneider, and Wagner (2016) reach a similar conclusion, and also find that predictability is weaker in times of low macroeconomic uncertainty. Finally, this conclusion is also echoed in Fulop, Li, and Wan (2019), who adopt a Bayesian approach to learn in real-time about model parameters and hidden states.

Conversely, Gargano, Pettenuzzo, and Timmermann (2019) find that by combining forward rates with the Lettau and Ludvigson (2009) macro factor investors can earn significant risk-adjusted portfolio returns, even after accounting for estimation error and model uncertainty. This differs from Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) primarily due to the inclusion of a composite macro factor, and allowing for stochastic volatility which leads to notable gains in performance for many forecasting models. Huang, Li, Wang, and Zhou (2020) use real-time (unrevised) macroeconomic data and show that even though real-time data is noisier, it renders standard forecasts insignificant, generates significant statistical and economic gains, and is able to predict future macroeconomic conditions.

2.2.1.3 Currency Return Predictability

In currency markets, Della Corte and Tsiakas (2012) review the main methods and results for predicting exchange rates with economic fundamentals. They consider six empirical models: exchange rates follow a random walk, uncovered interest parity (UIP), purchasing power parity (PPP), monetary fundamentals, and two variants of the Taylor rule.[3] They find that models based on UIP, PPP, and the Taylor rule outperform the random walk benchmark both statistically and economically, and model averaged predictions outperform individual predictions OOS. This is consistent with Molodtsova and Papell (2009) who also find that Taylor rule models have the best OOS performance.

[3]The random walk hypothesis is motivated by the seminal work of Meese and Rogoff (1983) who find exchange rates are not predictable at short horizons.

2.2.2 Machine Learning in Finance

Gu, Kelly, and Xiu (2020) is a seminal paper on applying a wide variety of supervised ML methods to empirical asset pricing. They focus on predicting individual stock returns in a panel setting using firm characteristics and macroeconomic variables as predictors, but also present results for equity factor portfolios. They find that supervised learning typically outperforms a preselected OLS benchmark, and in particular nonlinear methods that incorporate multiple predictor interactions such as trees and neural networks are the most promising in terms of OOS performance. Finally they find that the dominant firm characteristics that drive returns are variations of momentum, liquidity, and volatility. These results are consistent with Messmer (2017), who trains deep feedforward neural networks using a set of 68 firm characteristics to predict the cross-section of US stock returns, and finds that short-term reversals and 12-month momentum are the dominant firm characteristics.

Alberg and Lipton (2018) argue that forecasting firm fundamentals is less noisy than forecasting returns, and form factor portfolios based on neural network predicted book-to-market and EBIT-to-enterprise value ratios. In backtests, these predicted fundamentals generate a higher risk-adjusted return than standard factor construction methods. Chinco, Clark-Joseph, and Ye (2019) use the least absolute shrinkage and selection operator (LASSO; see Section 2.5.2.2) at high frequency to generate rolling one-minute-ahead return forecasts at the stock level using the entire cross-section of lagged stock returns as predictors. The LASSO tends to identify unexpected and short-lived predictors as stocks with meaningful news about fundamentals. In a similar exercise Leung and Tam (2018) coin the R^2 from a penalised regression of a stock's return against all other stocks as 'asset synchronicity,' and show that unique stocks earn higher returns than ubiquitous stocks because they are less valuable as factor risk hedges. Rapach and Zhou (2019) use a regularised regression technique called the elastic net which shrinks coefficients on uninformative predictors in a predictive regression towards zero (see Section 2.5.2.2 for a detailed description), and claim that this generates "one of the best market excess return forecasts to date."

Outside of equities, Huang and Shi (2019) use an adaptive group LASSO to identify macroeconomic drivers of bond returns. Their new factor combines information from employment, housing, financial, and inflation factors, has strong OOS predictive power for excess bond returns, and subsumes existing macro-based return factors. Bianchi, Buchner, and Tamoni (2021) conduct a similar exercise

to Gu, Kelly, and Xiu (2020) but for bond risk premia using macroeconomic predictors, and also trace predictive gains to nonlinearities. Li, Tsiakas, and Wang (2015) apply the elastic net to a 'kitchen sink' regression of exchange rate returns on macroeconomic fundamentals, and show that this can generate material OOS predictability through shrinkage of the coefficients on uninformative predictors. Finally, Jia, Liu, Wu, and Yan (2020) apply the OLS-post-LASSO technique to predict excess currency returns using the lagged forward discounts on other currencies as predictors, and show that a long-short portfolio of the highest minus lowest predicted excess returns generates a premium of 1.36% per month. In comparison to equities, the literature on applying ML to bond or currency markets is relatively scarce.

Other recent applications of ML in finance outside of return predictability include identifying and testing asset pricing models (Feng, Polson, and Xu, 2019; Gu, Kelly, and Xiu, 2019; Chen, Pelger, and Zhu, 2020; Feng, Giglio, and Xiu, 2020; Kozak, Nagel, and Santosh, 2020), pricing and hedging derivatives (Halperin, 2017, 2018; Buehler, Gonon, Teichmann, and Wood, 2018), and estimating default probabilities or creditworthiness in consumer finance (Khandani, Kim, and Lo, 2010; Butaru, Chen, Clark, Das, Lo, and Siddique, 2016; Sirignano, Sadhwani, and Giesecke, 2018).

2.3 Data

The historical data is obtained from Global Financial Data (GFD), who provide historical financial and economic data series for around 200 countries. Many of the GFD series have been manually extended using historical archives to generate series that go back further in time than standard indices. I obtain data for major equity, bond, and currency indices, as well as additional macroeconomic and financial variables to use as predictors in the ML models. The following subsections outline each data category and describe useful summary statistics.

2.3.1 Equities

I obtain monthly total equity index returns (capital gain plus dividends) for eight countries with the longest sample periods: Australia, Canada, Denmark, France, Germany, Japan, the UK, and the US. In Panel A of Table 2.1 I summarise the earliest start date, as well as the mean, standard deviation, skewness, and kurtosis of monthly holding period returns. Note that returns are not excess.

The earliest market with a full history is the UK, which starts in 1692, and the latest is Japan starting in 1886. All of the markets begin in the 19th century or earlier, which is much longer than typically used in the return predictability literature.[4] The mean returns are in line with typical post-WW2 equity premium estimates of 5–7% per annum, as are the standard deviations which range from 12–23%. Most countries exhibit low levels of skewness implying the return distribution is roughly symmetric, with the exception of Japan which has a large positive skewness of 1.51. Moreover, many of the return distributions unsurprisingly exhibit fatter tails than a normal distribution, with kurtosis greater than 3. In particular Germany and Japan have the fattest tailed distributions, whilst Denmark and the US have the thinnest.

2.3.2 Bonds

I also obtain monthly total returns for each country's government bond index, which are summarised in Panel B of Table 2.1. These indices are a composite of sovereign bonds of all maturities, and provide an indication of overall bond market performance regardless of maturity. The UK has the

[4] E.g. Campbell and Thompson (2008) use US data from the late 1800s, Goyal and Welch (2008) start in the early 1900s.

earliest bond market starting in 1700, and Japan has the latest starting in 1870. Unsurprisingly the mean bond index returns are all lower than the equity returns, ranging from 3–7% per annum, but some countries exhibit similar or greater bond return volatility. Most bond return distributions have low skewness, with the exceptions of Germany (−2.68) and the UK (1.84), which have highly skewed distributions. The kurtosis of bond returns all exceed that of a normal distribution, particularly Germany (23.23) and the UK (10.64).

2.3.3 Currencies

Exchange rates are treated differently to the other asset classes, as currencies are only free floating from 1970 onwards. Prior to this I construct the real exchange rate as the nominal exchange rate multiplied by the ratio of each country's CPI inflation index. These series are non-tradable, but can be viewed as a proxy for the exchange rate risk premium after adjusting for inflation.

I consider the G10 currencies in the analysis: AUD, CAD, CHF, EUR (average of DEM, FFR, and ITL prior to 1999), GBP, JPY, NOK, SEK, and NZD, all relative to USD. Quotes are expressed in USD per foreign currency. Panel C of Table 2.1 summarises the monthly returns of these series. Most of the currencies have average returns close to zero, with standard deviations in the range of 8–15% per annum. All series start in the early-mid 19th century, with the exception of NZD which begins in 1914. All the currencies have low skewness except EUR (1.37), which is primarily driven by devaluation of the Deutsche Mark in the 1940s, and JPY (1.73). Those currencies also exhibit high excess kurtosis greater than 10.

2.3.4 Macroeconomic Fundamentals and Financial Variables

I augment the list of predictors with the following macroeconomic fundamentals for each country: real GDP, CPI inflation, net exports, and unemployment rates. I also obtain additional financial market variables including 10-year constant maturity government bond yields and commodity price indices. Macroeconomic variables are conservatively lagged by one quarter to account for publication delay, but do not account for data revisions. All series are transformed through (log) differencing to ensure they are stationary. In total I have just over 100 predictors across all countries, which are detailed in Appendix A.3.

2.4 Forecasting Framework

2.4.1 Notation

Each asset is denoted $i \in \{1, \ldots, I\}$, and each forecast model is denoted $j \in \{1, \ldots, J\}$. An asset's return is described as an additive prediction error model:

$$r_{it} \;=\; \mathbb{E}_{t-1}(r_{it}) + \varepsilon_{it}, \tag{2.1}$$

where the conditional expected return $\mathbb{E}_{t-1}(r_{it})$ is a flexible function of the $k \times 1$ vector of lagged predictors $\boldsymbol{x}_{t-1}$:

$$\mathbb{E}_{t-1}(r_{it}) \;=\; g_i(\boldsymbol{x}_{t-1}). \tag{2.2}$$

Note that the function $g_i(\cdot)$ depends on i, which allows different assets to have different functional forms. It also does not depend on t, so the functional form is fixed over time (although the function's parameters may change). The heart of the forecasting exercise is to estimate a function for $g_i(\cdot)$ that maximises OOS predictive power for future returns.

The objective of this chapter is to examine the performance of a variety of ML models, which essentially amounts to estimating and evaluating different forms of $g_i(\cdot)$. Denote the estimated function using model j as $\hat{g}_{ij}(\cdot)$, then the forecasted return for asset i in month t is:

$$\hat{r}_{ijt} \;=\; \hat{g}_{ij}(\boldsymbol{x}_{t-1}). \tag{2.3}$$

Finally, note that the set of predictors is the same for every asset, and depends only on the month t.

2.4.2 Economic Restrictions

Campbell and Thompson (2008) show that imposing sensible economic restrictions on the signs of coefficients and forecasts can improve OOS performance of linear return forecasting models. It makes less sense to restrict the coefficients of ML models because of the complex interactions between predictors. For example, whilst economic theory may imply a positive or negative coefficient for a

single variable, it rarely suggests how multiple variables should interact with each other. For this reason I focus on restricting the actual forecasts. The restricted forecasts are defined as:

$$\hat{r}_{ijt}^{\text{res}} = \begin{cases} \hat{r}_{ijt} & \text{if } \hat{r}_{ijt} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The motivation underlying this restriction is based on rational models of asset pricing. If investors are risk-averse, the conditional expected return on any risky asset should always be positive as compensation for bearing risk (Merton, 1980). Moreover, as noted by Boudoukh, Richardson, and Smith (1993) a positive risk premium on any asset that positively correlates with the market is a necessary condition for the market portfolio to be conditionally mean-variance efficient. Therefore, in classical asset pricing models with risk-averse preferences it is not possible to generate a negative risk premium and it makes sense to impose a non-negativity restriction on return forecasts.[5]

2.4.3 Forecast Combination

One issue with examining individual forecasts is model uncertainty. Each model is only an estimate of the true unobservable data generating process, and individual forecasts are exposed to model specific assumptions and estimation error. Therefore to mitigate model uncertainty I combine the individual forecasts into a composite forecast for each asset. As noted by Rapach, Strauss, and Zhou (2010) and others, combining forecasts reduces variance and improves forecast accuracy.

The combined forecasts are computed by weighting the individual forecasts from model j with a specific weighting scheme ω:

$$\hat{r}_{it}^{c} = \sum_{j=1}^{J} \omega_{ijt} \hat{r}_{ijt}, \tag{2.5}$$

and the restricted combined forecasts are computed similarly to the individual restricted forecasts:

$$\hat{r}_{it}^{c,\text{res}} = \begin{cases} \hat{r}_{it}^{c} & \text{if } \hat{r}_{it}^{c} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

[5]Non-classical models are able to generate negative risk premia. E.g. Barlevy and Veronesi (2003) focus on the role of rational but uninformed traders, and Barberis, Greenwood, Jin, and Shleifer (2018) examine the roles of extrapolation and price bubbles.

I consider six forecast combination weighting schemes. First is a simple average scheme with $\omega_{ijt}^{\text{EW}} = 1/J$ that equally weights all forecasts regardless of their OOS performance. Simple forecast averaging is favourable as it often outperforms more sophisticated methods that exploit the correlation between forecasts to estimate theoretically optimal combination weights (Timmermann, 2006).

Alternative combined forecasts can be computed accounting for individual model performance. Stock and Watson (2004) define discounted mean squared error (DMSE) weights, which weight individual forecasts by the inverse of their discounted MSE over an expanding window:

$$\omega_{ijt}^{\text{DMSE}} = \frac{m_{ijt}^{-1}}{\sum_{j=1}^{J} m_{ijt}^{-1}}, \tag{2.7}$$

$$m_{ijt} = \sum_{s=\tau}^{t-1} \delta^{t-s-1} \left(r_{s+1} - \hat{r}_{ijs+1}\right)^2, \tag{2.8}$$

for some discount factor $0 < \delta \leq 1$. The DMSE weights account for the full history of forecast errors, but assign relatively more weight to recent errors. A higher (lower) δ puts relatively more (less) weight on older forecast errors. The case of no discounting ($\delta = 1$) corresponds to the Bates and Granger (1969) optimal weighting scheme when individual forecasts are uncorrelated.

The next three schemes are variations of the Della Corte and Tsiakas (2012) 'most recently best' MSE(κ) weights:

$$\omega_{ijt}^{\text{MSE}(\kappa)} = \frac{\text{MSE}(\kappa)_{ijt}^{-1}}{\sum_{j=1}^{J} \text{MSE}(\kappa)_{ijt}^{-1}}, \tag{2.9}$$

$$\text{MSE}(\kappa)_{ijt} = \sum_{s=1}^{\kappa} \left(r_{t-s} - \hat{r}_{ijt-s}\right)^2. \tag{2.10}$$

These weights are similar to the DMSE weights but use no discounting and weight forecasts by the inverse MSE over the past $\kappa = \{12, 36, 60\}$ months. The MSE(κ) weights are preferred when the best forecasting model changes over time, or if the investor does not care about forecast errors greater than κ months ago.

The final scheme is based on the heuristic thick modelling approach of Granger and Jeon (2004). In a thick model averaging (TMA) approach, several models are pooled together after removing outliers

or the worst performing models. The surviving models are then typically assigned equal weights. Pesaran, Schleicher, and Zaffaroni (2009) apply this approach to the estimation of volatility models across multiple asset classes, and show that it outperforms individual and Bayesian models. Their implementation of TMA focuses on the top p percentile of models ranked by Akaike information criterion (AIC) and Schwarz Bayesian information criterion (SBIC). Since this ranking is not feasible for most of the models considered in this chapter due to the lack of a likelihood function to compute the information criteria, I rank each model by its IS R^2 over the estimation period, and take an equally-weighted average of the top n model forecasts:

$$\omega_{ijt}^{\text{TMA}} = \begin{cases} \frac{1}{n} & \text{if } R^2_{\text{IS},ijt} \in \mathcal{N}_n \\ 0 & \text{otherwise,} \end{cases} \tag{2.11}$$

where $\mathcal{N}_n$ is the set of the top n performing models ranked by R^2_{IS}. In the empirical analysis I set $n = 3$, but the results are robust to choosing 4 or 5 instead. This weighting scheme can be interpreted as an asymmetrically trimmed mean, and should generate superior performance by excluding the worst performing models from the forecast.

Bayesian model averaging. An alternative model averaging paradigm not considered in this chapter is Bayesian model averaging (BMA). In the BMA approach, the weights used to combine forecasts are based on posterior model probabilities in a Bayesian framework. First, a prior distribution for each model is specified, and the observed data leads to a posterior for each model which can be used to combine forecasts. BMA has successfully been used in the return predictability literature by Avramov (2002), Cremers (2002), Faust, Gilchrist, Wright, and Zakrajssek (2013), Bianchi and McAlinn (2020), and many others.

The effect of bias in individual forecasts. As discussed by Min and Zellner (1993) and Timmermann (2006), problems can arise for forecast combinations under certain conditions when one or more of the individual forecasts is biased. To see this, consider two forecasts $\hat{y}_1$ and $\hat{y}_2$ of a random variable y. Let the forecast errors be $e_1 = y - \hat{y}_1 \sim (0, \sigma^2)$ and $e_2 = y - \hat{y}_2 \sim (\theta, \sigma^2)$, such that $\hat{y}_1$ is unbiased and $\hat{y}_2$ has a bias of θ. Assume $\text{cov}(e_1, e_2) = \sigma_{12} = \rho_{12}\sigma^2$. The MSE of $\hat{y}_1$ is $\mathbb{E}[(y - \hat{y}_1)^2] = \sigma^2$, whilst the MSE of $\hat{y}_2$ is $\mathbb{E}[(y - \hat{y}_2)^2] = \sigma^2 + \theta^2$. The MSE of the linearly combined

forecast $\hat{y}_c = \omega\hat{y}_1 + (1-\omega)\hat{y}_2$ is therefore $\text{MSE}(\hat{y}_c) = \omega^2\sigma^2 + (1-\omega)^2(\sigma^2+\theta^2) + 2\omega(1-\omega)\rho_{12}\sigma^2$. It follows that the MSE of the combined forecast in excess of the unbiased forecast ($\hat{y}_1$) is:

$$\text{MSE}(\hat{y}_c) - \text{MSE}(\hat{y}_1) \;=\; (1-\omega)\sigma^2\left((1-\omega)\left(\frac{\theta}{\sigma}\right)^2 - 2\omega(1-\rho_{12})\right). \tag{2.12}$$

Hence, $\text{MSE}(\hat{y}_c) > \text{MSE}(\hat{y}_1)$ if:

$$\left(\frac{\theta}{\sigma}\right)^2 \;>\; \frac{2\omega}{1-\omega}(1-\rho_{12}), \tag{2.13}$$

which is always satisfied if the forecast errors are perfectly correlated (i.e. as $\rho_{12} \to 1$), and may be satisfied if the magnitude of the bias is sufficiently large. This is particularly important in the context of ML, as many algorithms trade off increased bias for lower variance, and must be kept in mind when interpreting the performance of the combined forecasts.

2.5 Forecasting with Machine Learning

2.5.1 Why Machine Learning?

A natural question to ask is why use ML for return forecasting? What are the advantages over widely used methods such as linear forecasting models? As noted by Gu, Kelly, and Xiu (2020), there are at least three main reasons. First, the fundamental goal of ML is OOS prediction, which is well aligned with the goals of an investor seeking to exploit return predictability through active trading. Second, there are a large number of potentially highly correlated predictors that have been proposed in the academic literature, ranging from macroeconomic to financial predictors, of various countries and lags.[6] Many ML algorithms have built-in regularisation and dimensionality reduction, which reduces the risk of IS overfitting and mitigates collinearity. Finally, many ML methods have flexible functional forms that are able to incorporate nonlinearities and predictor interactions in ways that are infeasible for simple linear models due to the curse of dimensionality.

That said, ML is not a panacea. ML algorithms often require large amounts of data to be trained effectively, and positive OOS performance is not guaranteed. The field of return predictability is especially sensitive to this, as the signal-to-noise ratio is exceptionally low. Many ML algorithms were developed for applications such as image recognition or spam email filtering where the ratio of signal-to-noise is much higher (Hastie, Tibshirani, and Friedman, 2009). The data issue is somewhat mitigated by using long horizon data as in this chapter, which provides a greater training sample and is able to capture more states of nature and economic regimes to better train the ML models.

2.5.2 Overview of Machine Learning Models

In this section I outline in detail the specifics of the ML algorithms used throughout this chapter, highlighting the main idea and pros and cons. The algorithms have been chosen because they cover the most popular and widely used methods, whilst remaining relatively simple to understand and implement.

[6] E.g. Goyal and Welch (2008) examine almost 20 predictors for the aggregate US equity market return, whilst Harvey, Liu, and Zhu (2016) document over 300 'significant' predictors of the cross-section of individual stock returns.

2.5.2.1 Ordinary Least Squares

OLS is the most popular and widely used tool in empirical finance, and serves as the benchmark against which other models can be compared. In terms of notation, the function $g(\boldsymbol{x}_{t-1})$ is linear in the vector of predictors (where $\boldsymbol{x}$ includes a constant):

$$g(\boldsymbol{x}_{t-1}) \;=\; \beta' \boldsymbol{x}_{t-1}, \tag{2.14}$$

which leads to the standard predictive regression:

$$\begin{aligned} r_{it} &= \beta' \boldsymbol{x}_{t-1} + \varepsilon_{it} && (2.15)\\ &= \beta_0 + \beta_1 x_{1t-1} + \cdots + \beta_k x_{kt-1} + \varepsilon_{it}. && (2.16) \end{aligned}$$

The forecasted return is therefore:

$$\hat{r}_{it} \;=\; \hat{\beta}_0 + \hat{\beta}_1 x_{1t-1} + \cdots + \hat{\beta}_k x_{kt-1}. \tag{2.17}$$

The advantages of OLS are its simplicity and ease of computation, and its disadvantages are the restrictive linearity assumptions and lack of coefficient regularisation which makes OLS very prone to IS overfitting.

2.5.2.2 Penalised Regression

The main idea behind penalised regression is to add a penalty to the OLS loss function that penalises models with too many parameters or parameters that are too large. With a penalised regression, IS performance is intentionally sacrificed to create a more parsimonious model that hopefully performs better OOS. Specifically the penalised regression parameters are estimated by minimising the sum of squared errors (SSE) loss function plus a penalty function:

$$\underbrace{\sum_{t=2}^{T} (r_{it} - \beta' \boldsymbol{x}_{t-1})^2}_{\text{SSE loss function}} \;+\; \underbrace{\lambda p \sum_{j=1}^{k} |\beta_k| + \frac{1}{2}\lambda(1-p)\sqrt{\sum_{j=1}^{k} \beta_k^2}}_{\text{Penalty function}}. \tag{2.18}$$

This penalty function nests the three most popular penalised regression techniques: LASSO, ridge, and elastic net (EN). LASSO corresponds to the case where $p = 1$, and the penalty collapses to the ℓ_1-norm of the parameters: $||\beta||_1 = \sum_{j=1}^{k} |\beta_k|$. The LASSO is advantageous because the geometry of the ℓ_1 penalty sets some coefficients to exactly zero, thereby performing variable selection by removing uninformative variables. The ridge regression corresponds to the case where $p = 0$, where the penalty collapses to the ℓ_2-norm of the parameters: $||\beta||_2 = \sqrt{\sum_{j=1}^{k} \beta_k^2}$. It performs regularisation by stopping coefficients from being too large by shrinking them towards zero, but does not set any parameters to exactly zero as in the LASSO. Finally, the EN penalty is a weighted average of the ℓ_1 (LASSO) and ℓ_2 (ridge) penalties, and therefore performs both variable selection and shrinkage at the cost of an additional parameter (p).

The strength of regularisation is captured by the hyperparameter λ. A higher λ penalises coefficients to a greater extent, as it acts as a multiplier to the penalty function. When $\lambda = 0$ the loss function reduces to the OLS loss function, and as $\lambda \to \infty$ the penalty term dominates and the coefficients on all predictors are set to zero. In the latter case, only the intercept remains in the regression, and the historical mean provides the best forecast. Details on how to choose λ and p are described in Section 2.5.3.

2.5.2.3 Partial Least Squares

Partial least squares (PLS) is a technique for reducing the dimensionality of the predictor set to mitigate overfitting. It weights predictors by their univariate covariance with the target variable in the training period, reducing a large set of potentially correlated predictors to a smaller set with a size specified by the user. Predictors that covary more with the target receive higher weights, and completely uninformative predictors receive a weight of zero.[7]

The specific steps of the PLS algorithm are as follows. For each predictor $j = 1, \ldots, k$, estimate its partial sensitivity to the target variable (φ_j) over the training sample using OLS:

$$r_{it} \;=\; \phi_j + \varphi_j x_{jt-1} + \varepsilon_{it}. \tag{2.19}$$

[7] As noted by Gu, Kelly, and Xiu (2020), PLS is similar to principal component regression (PCR), except it has the advantage of performing dimensionality reduction with the ultimate goal of forecasting in mind through estimation of the partial covariances.

Next, compute the first PLS component x_{1t}^{PLS} as the average of all predictors weighted by their partial sensitivities:

$$x_{1t}^{\text{PLS}} = \sum_{j=1}^{k} \varphi_j x_{jt}. \tag{2.20}$$

Subsequent PLS components (x_{2t}^{PLS}, x_{3t}^{PLS}, etc.) are found by orthogonalising the target and predictors with respect to previous components, and iterating the previous steps until the desired number of components is achieved. The number of PLS components is the only hyperparameter of the PLS algorithm.

2.5.2.4 Random Forest

The random forest (RF) is a fully nonparametric algorithm that builds on the idea of a regression tree to incorporate multiway predictor interactions. At the most basic level, a regression tree groups together observations with similar characteristics and makes predictions using the average of all observations within a group. Figure 2.1 illustrates this with a simple example. There are two predictors: GDP growth and inflation. Observations are first sorted by GDP growth, with the branches in the tree corresponding to negative and positive GDP growth respectively. Observations with positive GDP growth are then split again based on whether inflation < 0.02. The square on the right provides an alternative visual depiction for the case of two predictors.

The RF is an extension that uses many trees (the 'forest') to generate more stable predictions. Using predictions from many trees corrects for the tendency of individual decision trees to overfit the data in the training sample. A natural question of how to choose the predictors and values to split the branches of the tree arises. The RF algorithm uses a method called dropout, which randomly samples predictors from the full set to reduce the correlation of forecasts from each tree. A large number of trees are generated, and the forecasts are averaged across all trees. The depth of each tree and the number of trees are the hyperparameters of the RF algorithm.

2.5.2.5 Gradient Boosting

The main idea behind the gradient boosting (GB) algorithm of Friedman (2002) is to recursively fit shallow trees to the residuals from previous trees to create an ensemble forecast. Shallow trees (i.e.

regression trees with a depth of one) are by construction weak learners, meaning individual trees have very little predictive power, but combining multiple weak learners can lead to a strong learner.

The steps of the GB algorithm are as follows. First, a shallow tree is fit to the training period data and the prediction errors are obtained. Then a second shallow tree is fit to the residuals of the first tree, and the forecasts are added together to form an ensemble forecast. Subsequent shallow trees are fit to the residuals from the previous tree until the desired number of trees is reached. The second and subsequent predictions are allocated a shrinkage factor $\theta \in (0, 1)$ to prevent the model from overfitting the residuals (Gu, Kelly, and Xiu, 2020). The maximum depth and shrinkage factor θ are the two hyperparameters of GB.

2.5.2.6 K-Nearest Neighbours

The final method is the k-nearest neighbours (KNN) regressor. KNN is a simple nonparametric algorithm that uses the k 'closest' observations from the training sample to predict the target variable in the testing period. Mathematically the KNN return prediction can be expressed as:

$$\hat{r}_t = \frac{1}{k} \sum_{\boldsymbol{x}_\tau \in N_k(\boldsymbol{x}_{t-1})} r_\tau, \tag{2.21}$$

where $N_k(\boldsymbol{x}_{t-1})$ is the neighbourhood of $\boldsymbol{x}_{t-1}$ defined by the k closest points in the training sample (Hastie, Tibshirani, and Friedman, 2009). The distance between two sets of predictors is typically measured as the Euclidian distance, which is the approach adopted in this chapter. Alternative metrics such as the Manhattan distance can also be used.

An alternative way of stating equation (2.21) is to take the predictors $\boldsymbol{x}_{t-1}$ and measure the distance in each month of the training sample. The KNN return prediction is the average of the k months with the lowest distances, where k is the only hyperparameter of the KNN algorithm. Intuitively the idea behind KNN is instances of the past that most closely resemble the present are the most useful for prediction. KNN tends to generate better predictions when the set of predictors is small, and requires the predictors to be standardised to the same scale to generate sensible predictions.

2.5.2.7 A Note on Neural Networks

Neural networks have been successfully used in the literature to predict the cross-section of stock returns (Gu, Kelly, and Xiu, 2020), bond returns (Bianchi, Buchner, and Tamoni, 2021), firm fundamentals (Alberg and Lipton, 2018), amongst other applications. Neural networks thrive with large datasets, as they have many parameters that need estimating, and the number of parameters scales with the number of layers in the network and the number of neurons in each layer. The focus of this chapter is on univariate time series prediction as opposed to pooled predictions, so although the time series length in this chapter is long, the total amount of data (number of assets $\times$ time series length) is not sufficient to adequately train neural networks in this context.

2.5.3 Hyperparameter Tuning

A key element of any ML model is the concept of hyperparameter tuning. Hyperparameters are the parameters of a model that are set before the model starts learning (i.e. before any data is brought to the model), such as the strength of the LASSO penalty, maximum tree depth in RF, or number of neighbours in KNN. Tuning a hyperparameter means finding its 'optimal' value, typically the value that minimises forecast errors in a held-out data sample after the model has been estimated. This process is necessary because economic theory rarely provides guidance on choosing the hyperparameters of a ML model.

The standard way to tune hyperparameters is through a technique called k-fold cross-validation. In k-fold cross-validation the sample is first split into two parts: training and testing. The training sample is then split into k 'folds' (i.e. bins), typically of equal size and after shuffling the training sample so successive observations are not necessarily in the same fold. The model is then estimated using all but one of the folds, and is evaluated using the remaining fold. This step is repeated but a different fold is excluded until all the folds have been used, and the evaluation metric is averaged across all of the k iterations. This process is repeated again for different values of the hyperparameters, leading to a set of hyperparameter and average evaluation metric pairs. The optimally tuned hyperparameters are those that generate the lowest forecast errors (typically MSE) averaged across all folds.

However, k-fold cross-validation is not valid in a time series context because it randomly samples from the entire distribution, ignoring time dependence in the predictors and target variables. Therefore, I apply a method called walk forward (also known as expanding window) validation to tune the hyperparameters of the ML models. In a walk forward validation procedure the sample is chronologically split into three parts: training, validation, and testing. The hyperparameters are first fixed at certain values, the model is estimated using the training sample, and a forecast is generated for the next n periods of the validation sample. The training sample is then expanded by n periods, the model is re-estimated, and another set of n-period forecasts are generated. Typically $n = 1$, so the training sample is expanded one period at a time. This process is iterated until forecasts have been generated for the entire length of the validation sample, and for a range of hyperparameters. The optimal hyperparameters are those that minimise forecast errors in the validation sample. Figure 2.2 provides a visual illustration of the walk forward validation procedure.

Table 2.2 summarises the hyperparameters and tuning methods I use in the empirical analysis. For methods that only have a single hyperparameter such as LASSO or KNN, I adopt a grid search procedure, using a fine grid of possible hyperparameters for the search. For methods that have multiple hyperparameters such as EN or GB, the curse of dimensionality becomes a problem, so I use a more computationally efficient method called random search.[8] In a random search procedure a distribution is specified for each hyperparameter, and hyperparameters are randomly sampled from the distribution many times, hopefully capturing a wide variety of possible combinations. For example, for the EN algorithm I specify the penalty and weighting parameters as both coming from a uniform distribution, and jointly sample 100 observations from the distributions and iterate the walk forward validation procedure over these random samples.

A final point to note is the models are retuned every year to account for possible variation in the optimal hyperparameters over time.

[8]The curse of dimensionality refers to the phenomenon that as the number of possible parameters increases the number of parameter combinations grows exponentially, making an exhaustive grid search of all combinations computationally infeasible without using a coarse grid.

2.6 Statistical Evaluation

2.6.1 Forecast Evaluation Methodology

This section outlines the statistical evaluation methodology, focusing on the OOS forecast evaluation metrics, and tests of forecast optimality.

2.6.1.1 Out-of-Sample R-Squared

A standard metric for forecast evaluation is the OOS R-squared (OOS R^2). The OOS R^2 compares the MSE of the forecast model to a benchmark model, typically the historical mean (Campbell and Thompson, 2008; Goyal and Welch, 2008):

$$R^2_{\text{OOS}} \;=\; 1 - \frac{\sum_{t=\tau}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=\tau}^{T}(r_t - \bar{r}_t)^2}, \tag{2.22}$$

where τ is the first date of the testing period, r_t is the actual return in period t, and $\hat{r}_t$ is the forecasted return for period t. $\bar{r}_t$ is the historical mean computed using all data up to but excluding period t (so no information from the testing period is used to compute the historical mean):[9]

$$\bar{r}_t \;=\; \frac{1}{t-1}\sum_{s=1}^{t-1} r_s. \tag{2.23}$$

If a model outperforms the historical mean, its MSE will be lower and the OOS R^2 will be positive. Note that the OOS R^2 can be negative, as a model can perform arbitrarily worse than the historical mean.

2.6.1.2 Henriksson-Merton Test

Henriksson and Merton (1981) develop a nonparametric directional accuracy test that examines whether the *signs* of the forecasts are correct. The HM statistic is defined as the proportion of

[9]Some authors evaluate forecasts relative to a benchmark of zero instead of the historical mean, e.g. Gu, Kelly, and Xiu (2020). This is appropriate when forecasting individual stock returns, because the historical mean is so noisy it artificially lowers the bar for model evaluation.

forecasts with the correct sign:

$$\text{HM} \;=\; \frac{1}{T-\tau}\sum_{t=\tau}^{T} \mathbb{1}_{\text{sign}(\hat{r}_t)=\text{sign}(r_t)}, \tag{2.24}$$

where $\mathbb{1}_{\{\cdot\}}$ is a function that returns 1 if the condition in the subscript is true, and 0 otherwise. The HM statistic follows a binomial distribution with variance:

$$\text{var}(\text{HM}) \;=\; \frac{\text{HM}(1-\text{HM})}{T-\tau}. \tag{2.25}$$

For large samples the binomial distribution can be approximated by a normal distribution, and a standard normal Z-statistic provides a good approximation. The null hypothesis is that the forecasting model is correct 50 percent of the time, versus the one-sided alternative that the model is better than an equally likely random guess.

2.6.1.3 Mincer-Zarnowitz Regressions

I also estimate Mincer-Zarnowitz optimal forecasting regressions by regressing the realisation of the target variable on the contemporaneous predicted value:

$$r_{t+1} \;=\; \alpha + \beta\hat{r}_{t+1} + u_{t+1}. \tag{2.26}$$

If the forecast is optimal α and β would be jointly zero and one respectively, i.e. in expectation the forecast would exactly equal the realisation with no bias. Therefore, in addition to the regression parameters I also report p-values for an F-test of the joint null hypothesis that $\alpha = 0$ and $\beta = 1$, against the alternative that $\alpha \neq 0$ or $\beta \neq 1$ or both.

2.6.1.4 Model Comparison

Finally, I compute metrics that can be used to compare models in statistical sense, i.e. does model X significantly outperform model Y? The classic method of model comparison is the Diebold and Mariano (1995) (DM) test. The DM test is a pairwise test of two non-nested models computed from

each model's forecasting errors:

$$\text{DM} = \frac{\bar{d}}{\sqrt{\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k/n}}, \tag{2.27}$$

$$\bar{d} = \frac{1}{T-\tau}\sum_{t=\tau}^{T} d_t, \tag{2.28}$$

$$\gamma_k = \frac{1}{n}\sum_{t=\tau+k+1}^{T}(d_t - \bar{d})(d_{t-k} - \bar{d}), \tag{2.29}$$

where $d_t = |e_{1,t}| - |e_{2,t}|$ is the difference in absolute (or alternatively, squared) forecast errors, and γ_k is a measure of the autocovariance of the difference at lag k. The DM statistic is asymptotically standard normal under the null hypothesis that both models have the same predictive power.

The DM statistic is invalid in the case of nested models, such OLS or penalised regressions nesting the historical mean. Clark and West (2006, 2007) (CW) develop an inference procedure for nested models which accounts for the fact that additional parameters add noise into the forecasting exercise, and therefore the richer model having higher MSE is not necessarily evidence in favour of the benchmark model. The CW statistic is estimated by first defining a variable f as:

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - (r_{t+1} - \hat{r}_{t+1})^2 + (\bar{r}_{t+1} - \hat{r}_{t+1})^2, \tag{2.30}$$

and regressing f on a constant. The significance of the CW statistic is the t-statistic from that regression under the null hypothesis that the models are equal, and the alternative that the forecasting model is better than the historical mean benchmark.

2.6.2 Results

2.6.2.1 Full-Sample

Table 2.3 reports the OOS R^2 for each country's equity and bond market, and for the currency pairs for each ML model over the full three century sample. Panel A reports results for the equity indices, Panel B for the bond indices, and Panel C for the currencies. Significance is determined by either

the DM or CW test depending on whether the historical mean is nested in the model or not.

The results are overwhelmingly poor for the individual forecasts. Almost every forecast underperforms the historical mean in terms of OOS R^2, often by a large margin. Unsurprisingly OLS generates disastrous OOS forecasts, because there is no regularisation to prevent it from overfitting in-sample, and this problem is exacerbated when the number of predictors is large. Ridge regression tends to generate similar forecasts to OLS. One reason for this is because it does not regularise the coefficients enough compared to LASSO or EN, which both tend to generate more reasonable OOS R^2s by setting coefficients on redundant variables to zero. The optimally tuned EN tends to load more heavily on the LASSO penalty than the ridge penalty, so LASSO and EN tend to generate heavily correlated predictions.[10] PLS also performs very poorly in all cases. The tree based methods-–RF and GB—do not perform well, but RF tends to consistently perform better than GB. Finally, KNN performs consistently poorly in all cases. The only significantly positive OOS R^2s are for Canadian equities, where LASSO and EN generate modest 3.25–3.29% OOS R^2s on a monthly basis, and Japanese bonds where RF generates an OOS R^2 of 1.01%. French equities generate positive OOS R^2s over 1% using LASSO and EN, but these are not statistically significant. The currency results paint a similar picture to the equity and bond market results, exhibiting the same patterns and poor performance. CHF is marginally significantly positive when using the LASSO, at 2.88% per month.

Table 2.4 reports the same results for the combined forecasts using the six weighting schemes, with and without economic restrictions. The results are much better than the individual forecasts, especially for equity and bond returns. There are a few notable patterns. First, the MSE(12) weighting scheme appears to almost consistently dominate the other schemes, suggesting the best forecasting model varies over time, and at short horizons. Second, the simple weighting scheme performs moderately well despite its simplicity, especially in bond markets. Third, DMSE and MSE(60) are generally the worst performing weighting schemes, suggesting longer term forecast errors are less relevant for choosing an optimal return forecasting model. Fourth, currencies remain very difficult to predict OOS, as the OOS R^2s are mostly negative and none are significantly positive. Finally, imposing economic constraints almost always improves forecasting performance, most notably for German bonds where the large negative OOS R^2 becomes positive and mostly significant.

[10]For reference, the average ℓ_1 weight for the EN (p) is around 0.80 compared to the ℓ_2 weight of 0.20.

Table 2.5 reports the correlations between the ML model forecasts averaged across all assets. In general, the forecasts across models seem to be fairly independent with low absolute correlations mostly less than 0.30. However the LASSO and EN forecasts tend to be highly correlated, because the optimal EN loads heavily on the LASSO penalty. Moreover, the tree-based methods are on average highly correlated with a coefficient of 0.65. The generally low correlations support the improvements gained from combining forecasts.

For brevity and due to the significantly better performance, the rest of the results will focus on the combined forecasts, with economic restrictions where relevant. Table 2.6 reports results of the HM directional accuracy test with significance levels. All of the equity return forecasts significantly outperform the 50 percent benchmark, with accuracy rates in the region of 53–60 percent, except for the simple scheme in Japan. Similarly, all of the bond return forecasts are highly significant and exceed 54 percent. These numbers may seem low, but even a small edge can accumulate into large gains over long horizons. With the exception of EUR, the currency return forecasts are all 50 percent or lower, and hence underperform a random guess.

Tables 2.7 and 2.8 report the Mincer-Zarnowitz regression coefficients and F-test p-values for the MSE(12) combined forecasts (results for other schemes are available on request). Recall if the forecast is optimal the intercept should be close to zero and the slope should be close to one. For all markets except Australian and Danish equities, and French and German bonds, the null hypothesis of an optimal forecast cannot be rejected at the 5% level. All of the intercepts are close to zero, and most of the slopes are below one, implying the forecasts tend to systematically underestimate the true return. For currencies, the null hypothesis cannot be rejected for CHF, EUR, and JPY, but is strongly rejected for the others. Overall the results suggest the MSE(12) combined forecasts perform well in equity and bond markets, but not in currency markets.

Finally, Figures 2.6–2.8 plot the cumulative SSE difference between the historical mean benchmark and the MSE(12) forecasts, along with a 95% confidence interval. An increase in the line indicates better performance of the forecast model, and vice versa for a decrease. The plots highlight steady increases in the lines over time, implying the forecasts tend to consistently outperform the benchmark. Some assets exhibit spikes around major historical events such as recessions or major wars.

2.6.2.2 Sample Splits

It is well documented that asset returns are more predictable in recessions and times of high economic uncertainty, possibly due to the countercyclical nature of risk premiums and time-variation in the dynamics of the predictors (Rapach, Strauss, and Zhou, 2010; Henkel, Martin, and Nardari, 2011; Dangl and Halling, 2012; Sarno, Schneider, and Wagner, 2016; Gargano, Pettenuzzo, and Timmermann, 2019). To examine whether there is systematic variation in the forecasting power of ML across economic regimes, I split the sample by five variables: a recession indicator, asset volatility, inflation, interest rates, and a post-WW2 indicator. The recession indicator is based on NBER classified recessions, extended to 1802 using hand-collected data from Gordon (1986) for the US, and defined as two consecutive quarters of negative GDP growth for the other countries. Volatility, inflation, and interest rates are split into low and high periods by their full-sample medians.

Tables 2.9–2.13 summarise the results in terms of OOS R^2 for each forecast combination method. Some notable patterns emerge. First, forecasting performance is generally better in recessions for all assets, consistent with prior research. Second, forecasts are better in periods where asset volatility is high, implying it is easier to accurately predict future returns in volatile markets. Third, there is no strong evidence that forecasting power varies with inflation or the level of interest rates, except bond returns which are easier to predict in low interest rate environments. Lastly, forecast accuracy seems to be equally strong in the pre- and post-WW2 periods, suggesting the results are not concentrated in a particular period of time.

2.6.2.3 Variable Importance

A potential criticism of ML models is they can be a 'black box', i.e. it is difficult to understand what is going on inside the model and what the main driving factors are. To alleviate these concerns, I construct a measure of variable importance (VI) for each model and asset. VI measures how important a predictor is in a model by how much performance changes when it is excluded. Specifically, for each model j and variable k, the model is re-estimated excluding that variable, and the OOS R^2s are compared: the reduction in OOS R^2 when excluding variable k scaled by all other

reductions is the measure of VI. Mathematically:

$$\text{VI}_{jk} = \frac{R^2_{\text{OOS}} - R^2_{\text{OOS},-k}}{\sum_{i=1}^{K}\left(R^2_{\text{OOS}} - R^2_{\text{OOS},-i}\right)}, \tag{2.31}$$

where $R^2_{\text{OOS},-k}$ is the OOS R^2 when variable k is excluded from the model but all other variables remain. A higher VI implies the variable is more important for generating accurate predictions, and assists with understanding the economic drivers of return predictability.

Figures 2.3–2.5 plot the VI for the top 5 most important variables averaged across all models for each asset. A full list of variables with abbreviations is detailed in Appendix A.3. In many cases there are no clear driving variables, e.g. for Australian, Danish, German, or Japanese equities the VIs of the top 5 variables are all very similar and low in magnitude. However, for Canadian equities a large amount of predictability comes from commodity markets, lagged Canadian equity returns, and US 10-year government bond yields. For French equities the most dominant variable is the 10-year French government bond yield. In the UK, commodity and lagged UK equity returns are the most important predictors. The lagged US equity return is the dominant predictor of US equities, implying some degree of autocorrelation in returns. For bond markets, arguably UK bond returns drive US bond returns. Otherwise there are no discernible relationships. Finally, currency returns are primarily driven by commodities, lagged other currency returns, and to some extent bond yields, bond returns, and net exports.

2.7 Economic Evaluation

2.7.1 Dynamic Trading Strategies

To evaluate the economic significance of using ML for return predictability, I implement four dynamic trading strategies. The first is a naive $1/N$ portfolio that assigns equal weight to each asset. This portfolio maximises diversification without the need for a sophisticated asset allocation model. The second is a variation of the $1/N$ portfolio called $1/N$-neutral that assigns an equal weight to each asset class (one-third each) and then equally-weights assets *within* an asset class, i.e. it is asset class neutral. This accounts for the fact there are more currencies than equity and bond indices. The third and fourth strategies use Markowitz mean-variance analysis to create an optimal portfolio that seeks to maximise risk-adjusted returns. The dynamic portfolio weights for the MVO portfolios are computed as the solution to the maximisation problem:

$$\begin{aligned} \max_{\boldsymbol{w}_t} \quad & \frac{\boldsymbol{w}_t' \mathbb{E}_t(\boldsymbol{r}_{t+1})}{\sqrt{\boldsymbol{w}_t' \Sigma_{t+1} \boldsymbol{w}_t}} \\ \text{s.t.} \quad & \mathbb{1}' \boldsymbol{w}_t = 1 \\ & w_{it} \geq 0 \quad \forall\ i, \end{aligned} \tag{2.32}$$

where Σ_{t+1} is the conditional covariance matrix, $\mathbb{E}_t(\boldsymbol{r}_{t+1})$ is a vector of conditional expected excess returns, and $\mathbb{1}$ is a vector of ones of appropriate dimension. The second constraint imposes a no short-selling restriction, which is more realistic as it would not have been feasible to short many of the assets throughout the sample period. Jagannathan and Ma (2003) show that imposing this constraint is equivalent to shrinking the elements of covariance matrix. The standard MVO portfolio uses historical returns and covariances as inputs into the analysis, i.e.:

$$\mathbb{E}_t(\boldsymbol{r}_{t+1}) = \frac{1}{t} \sum_{s=1}^{t} \boldsymbol{r}_s, \tag{2.33}$$

$$\Sigma_{t+1} = \frac{1}{t} \sum_{s=1}^{t} (\boldsymbol{r}_s - \mathbb{E}_t(\boldsymbol{r}_{t+1}))^2. \tag{2.34}$$

The ML MVO portfolio (MVO-ML) uses the MSE(12) return forecasts with economic restrictions as

the conditional expected returns, as opposed to the historical means. This chapter has not attempted to model or predict covariances, so the historical covariance matrix is used to find the MVO-ML weights. Modelling volatility with ML in addition to expected returns would be an interesting extension.

2.7.2 Performance Measures

To evaluate the performance of the strategies I compute standard performance metrics like moments of returns, Sharpe and Sortino ratios, and maximum drawdowns. I also compute the breakeven transaction cost τ^{be} following Han (2006) and Della Corte and Tsiakas (2012). τ^{be} is the transaction cost (in monthly basis points) that would make an investor indifferent between two strategies, computed as:

$$\tau_t = \frac{1}{N}\sum_{i=1}^{N}\left|w_{it} - w_{it-1}\left(\frac{1+r_{it}}{1+r_{pt}}\right)\right|, \tag{2.35}$$

$$\tau^{\text{be}} = \frac{\bar{r}_1 - \bar{r}_2}{\bar{\tau}_1 - \bar{\tau}_2}. \tag{2.36}$$

If an investor pays transaction costs lower than τ^{be} they will prefer strategy 1 to strategy 2. This measure is advantageous because it accounts for two potential issues that would make accurately computing transaction costs for the strategies difficult: transaction costs have decreased significantly over time, and it is difficult to get accurate historical transaction cost information far into the past. The breakeven transaction cost estimates are computed for both MVO strategies relative to the $1/N$ benchmark, and for the MVO-ML strategy relative to the standard MVO portfolio.

Additionally, Goetzmann, Ingersoll, Spiegel, and Welch (2007) develop a manipulation-proof performance measure to measure the utility of realised returns:

$$\hat{\Theta} = \frac{1}{1-\delta}\ln\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1+r_{pt}}{1+r_{ft}}\right)^{1-\delta}\right], \tag{2.37}$$

where δ is the degree of relative risk aversion. $\hat{\Theta}$ is a measure of the strategy's return premium after adjusting for risk. Another way to interpret $\hat{\Theta}$ is "a portfolio has the same score as a risk-free

asset whose continuously-compounded return exceeds the risk-free rate by $\hat{\Theta}$" (Goetzmann, Ingersoll, Spiegel, and Welch, 2007, pg. 1507). Hence the difference $\mathcal{P} = \hat{\Theta}_1 - \hat{\Theta}_2$ represents the performance fee one would be willing to pay to get the returns of strategy 1 compared to strategy 2. In the empirical analysis I assume a risk aversion of $\delta = 2$, and use the GFD T-bill total return index as a proxy for the risk-free rate.

2.7.3 Results

Table 2.14 details summary statistics for the four strategies, and Figure 2.9 plots the cumulative returns over the full sample on a log scale. Both MVO strategies dominate the naive benchmarks across most measures, generating sizeable improvements in Sharpe ratios with lower drawdowns and tail risk (as measured by kurtosis). This is surprising in the context of DeMiguel, Garlappi, and Uppal (2009), Duchin and Levy (2009), and others, who find that the $1/N$ portfolio typically outperforms more sophisticated optimisation methods. This may be an artifact of the data, or an advantage caused by estimating the inputs over a long window which generates significantly more stable weights.

The MVO-ML strategy compares favourably to the standard MVO strategy across most measures: it has moderately higher returns, similar standard deviation, lower downside deviation and maximum drawdown, and higher Sharpe and Sortino ratios. The breakeven transaction cost estimates relative to the $1/N$ strategy are considerable at over 500 basis points per month, implying an investor facing fees lower than this can benefit from implementing a MVO strategy. The MVO-ML incurs a breakeven transaction cost of 79 basis points above the standard MVO portfolio. Finally, the MVO-ML strategy generates the highest manipulation-proof performance measure at 10.18% per annum, exceeding the standard MVO portfolio by almost 2%, and dramatically outperforming the naive strategies. Overall there is strong evidence that investors can exploit ML return predictability through simple trading strategies.

2.8 Conclusion

This chapter provides a comprehensive analysis on the use of ML methods for return predictability across international equity, bond, and currency markets. I use up to three centuries of data to train and evaluate the models, capturing multiple business cycle turning points and rare events. The ML algorithms cover the most popular and widely used methods, and are fairly simple to understand and implement.

I show that individual ML forecasts perform very poorly OOS, often underperforming the historical mean by a large margin. Methods that perform shrinkage or variable selection fare better than those that do not, by mitigating the large degree of IS overfitting. Combining the forecasts of multiple independent models together generates significantly positive OOS performance, and the performance improves further when imposing weak economically motivated restrictions on the signs of the forecasts. A forecast combination scheme using the inverse of the most recent 12 months of forecast errors is by far the best, suggesting the optimal forecasting model varies significantly over time and at short horizons. I also split the sample by various variables and show that forecasting performance is better in bad times and when market volatility is high, consistent with prior literature. There is no systematic difference in performance before and after WW2, suggesting the results are not concentrated in a particular period of time.

In terms of limitations, there are three main areas that could be improved in future work. First, the predictors used in these chapter are primarily low frequency macroeconomic variables, so broadening the scope to include other relevant variables such as the equity price-dividend ratio may be beneficial (subject to data availability). Second, I focus on simple and easily implementable ML models in this chapter, but there is potential to implement more sophisticated models, most notably those in deep and reinforcement learning. Finally, broadening the scope to more international markets or alternative asset classes like commodities or real estate may be of interest to some market participants.

Table 2.1: Summary statistics for equity, bond, and currency returns. This table reports summary statistics for monthly equity and bond returns for eight international markets, and monthly real exchange rate returns for the G10 currencies relative to USD. Prior to 1999 EUR is the equally-weighted average of DEM, FFR, and ITL. Start date is the YYYY-MM date when the first observation is recorded. Means and standard deviations (SD) of returns are in annualised percent.

	Start date	Mean	SD	Skew	Kurtosis
Panel A: Equity returns					
Australia	1825-02	10.15	17.93	0.37	6.12
Canada	1825-03	8.83	18.64	−0.20	4.71
Denmark	1873-02	7.55	13.38	0.15	2.77
France	1802-01	9.15	18.50	0.26	3.81
Germany	1835-06	8.53	20.78	−0.32	8.41
Japan	1886-01	10.88	23.18	1.51	11.16
UK	1692-02	7.35	11.51	0.07	3.78
US	1792-02	8.73	12.17	−0.35	1.97
Panel B: Bond returns					
Australia	1857-06	5.55	10.29	−0.29	4.86
Canada	1853-02	5.44	7.98	0.11	4.52
Denmark	1819-05	6.71	12.10	0.46	8.66
France	1800-02	5.84	14.55	0.34	6.13
Germany	1832-08	2.77	17.10	−2.68	23.23
Japan	1870-06	6.09	11.67	0.05	3.39
UK	1700-08	6.96	13.88	1.84	10.64
US	1786-08	5.52	7.33	0.73	4.66
Panel C: Currency returns					
AUD	1862-01	0.16	9.41	−0.11	5.93
CAD	1800-02	−0.15	8.34	−0.15	8.16
CHF	1819-08	0.72	10.26	0.44	9.61
EUR	1800-03	1.84	12.99	1.37	14.89
GBP	1801-01	−0.09	8.83	0.00	8.17
JPY	1869-01	2.89	14.90	1.73	18.80
NOK	1819-05	0.46	10.66	0.42	7.50
SEK	1813-12	0.17	9.02	−0.04	7.50
NZD	1914-08	0.74	10.09	−0.03	5.59

Table 2.2: Summary of hyperparameter optimisation. This table lists the key hyperparameters for each machine learning algorithm, along with a tuning method (either grid search or random search). OLS is ordinary least squares, LASSO is the least absolute shrinkage and selection operator, EN is elastic net, PLS is partial least squares, RF is random forest, GB is gradient boosting, and KNN is k-nearest neighbours. The grid search syntax is [start, stop, step], and the random search syntax is the distribution which the parameter is sampled from. The continuous uniform distribution uniform(a, b) randomly samples value between a and b with equal probability. The discrete uniform distribution duniform(n, m) selects a random integer between n and m inclusive.

Method	**Hyperparameters**	**Tuning method**
OLS	–	–
LASSO	ℓ_1 penalty (λ)	Grid search: [0.001, 0.1, 100]
Ridge	ℓ_2 penalty (λ)	Grid search: [0.001, 0.1, 100]
EN	Joint ℓ_1, ℓ_2 penalty (λ)	Random search: uniform$(0.001, 0.5)$
	ℓ_1 ratio weight (p)	Random search: uniform$(0, 1)$
PLS	Number of components	Grid search: [1, 10, 1]
RF	Maximum depth	Grid search: [1, 5, 1]
	Number of trees	Grid search: [100, 200, 100]
GB	Maximum depth	Random search: duniform(1, 5)
	Shrinkage factor (θ)	Random search: uniform(0, 1)
KNN	Number of neighbours	Grid search: [1, 5, 1]

Table 2.3: Out-of-sample R^2 for machine learning models. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns, using various machine learning models over the full sample. OLS is ordinary least squares, LASSO is the least absolute shrinkage and selection operator, EN is elastic net, PLS is partial least squares, RF is random forest, GB is gradient boosting, and KNN is k-nearest neighbours. '–' means the value is less than −100%. Significance is based on the Diebold and Mariano (1995) test when the historical mean model is not nested in the ML model, and the Clark and West (2006, 2007) test otherwise. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	OLS	LASSO	Ridge	EN	PLS	RF	GB	KNN
Panel A: Equity returns								
Australia	–	−0.37	–	−0.31	–	−3.24	−14.16	−20.40
Canada	–	3.25***	–	3.29***	–	1.64	−9.86	−17.07
Denmark	–	–	–	−3.00	–	−3.40	−14.59	−9.46
France	–	1.57	–	1.37	–	−0.47	−10.94	−20.66
Germany	–	−0.02	–	0.03	–	−2.27	−27.76	−10.70
Japan	–	−2.38	–	−2.21	–	−5.62	−40.04	−25.64
UK	–	−0.53	–	−0.53	–	−1.00	−8.61	−21.84
US	–	0.06	–	0.25	–	−1.11	−4.38	−15.70
Panel B: Bond returns								
Australia	–	0.05	–	−0.09	–	0.49	−8.41	−14.27
Canada	–	−0.45	–	−0.26	–	0.48	−8.04	−11.53
Denmark	–	0.18	–	0.24	–	1.29	−14.08	−8.07
France	–	−2.11	–	−1.72	–	−0.46	−10.74	−21.36
Germany	–	–	–	–	–	−2.72	−19.17	−6.04
Japan	–	−0.03	–	0.31	–	1.01*	−5.37	−11.71
UK	–	−0.14	–	−0.05	–	−0.45	−5.82	−15.96
US	–	0.51	–	−0.02	–	0.80	−7.23	−24.28
Panel C: Currency returns								
AUD	–	−0.35	–	−0.05	–	−2.84	−19.09	−16.85
CAD	–	−0.03	–	−0.02	–	−0.67	−7.76	−19.95
CHF	–	2.88*	–	1.98	–	−0.44	−7.27	−19.51
GBP	–	−0.71	–	−0.98	–	−1.75	−7.66	−21.36
EUR	–	–	–	–	–	−3.61	−20.49	−21.23
JPY	–	−3.57	–	−2.59	–	−1.90	−23.05	−22.34
NOK	–	−0.85	–	−1.02	–	−1.23	−11.77	−22.05
SEK	–	−0.32	–	−0.60	–	−2.62	−23.33	−17.25
NZD	–	−0.66	−84.09	−0.42	−8.75	−1.69	−20.96	−5.46

Table 2.4: Out-of-sample R^2 for combined forecasts with economic restrictions. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns, using various forecast averaging methods over the full sample. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. The economic restrictions replace negative forecasts with zero. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	Unrestricted						Restricted					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	−0.11	−0.24	1.16	0.41	0.12	0.87	−0.08	0.12	−1.02	−0.35	−0.16	−0.30
Canada	2.30***	3.80***	5.80***	4.79***	4.08***	5.01***	2.19***	3.41***	5.24***	4.25***	3.69***	4.37***
Denmark	−0.80	0.17	2.08	0.86	0.39	0.72	−0.29	1.03	2.51	1.67	1.39	1.49
France	0.65	0.09	1.69	0.64	−0.40	1.22	0.56	0.62	1.79	1.14	0.34	1.54
Germany	0.32	1.00	2.28	1.01	0.42	1.81	0.61	1.78	2.95***	2.17	1.87	2.31*
Japan	0.03	−1.22	0.93	−0.29	−0.56	0.42	0.06	−0.70	1.02	−0.13	−0.38	0.57
UK	0.00	−0.01	0.77	0.14	−0.01	0.38	0.25	0.31	1.18	0.66	0.53	0.71
US	0.07	0.24	1.05*	0.56	0.39	0.88	0.06	0.37	1.14	0.69	0.53	0.89
Panel B: Bond returns												
Australia	0.20*	0.20	1.00	0.20	−0.00	0.54	0.34*	0.51	1.43*	0.71	0.44	0.81
Canada	0.36*	0.36	1.17	−0.07	−0.50	0.19	0.96*	0.48	1.19	0.33	−0.18	0.52
Denmark	0.83**	1.14	2.74**	1.86*	1.53	1.94*	0.68**	1.00	2.21***	1.37*	1.24	1.76**
France	0.13	−0.48	1.06	0.31	0.04	0.84	0.13	0.35	1.42**	0.88	0.61	1.12*
Germany	−0.00	−14.30	−3.21	−10.17	−12.96	−4.98	0.47*	0.95*	1.50**	1.19*	1.05	1.22*
Japan	0.97**	1.33	2.63**	1.68	1.20	2.19*	0.93**	1.15	2.17**	1.44	1.01	1.86*
UK	0.00	1.18	2.19**	1.54	1.22	1.97*	0.43	1.67*	2.60***	2.09**	1.83*	2.36***
US	0.95***	0.42	1.73	1.26	0.92	1.37	0.95***	1.30*	2.82***	2.15***	1.85**	2.23***
Panel C: Currency returns												
AUD	−0.54	−1.23	−0.85	−2.01	−2.23	−1.32	0.12	−0.05	0.10	−0.67	−0.86	−0.38
CAD	0.07	−1.83	−0.80	−1.34	−1.17	−0.99	0.09	−0.96	−0.34	−0.64	−0.58	−0.51
CHF	0.61	1.41	3.18	2.33	1.97	2.13	0.61	0.87	1.78	1.30	1.09	0.93
GBP	−0.11	−1.39	−0.37	−0.78	−1.43	−0.88	0.08	−0.25	0.36	−0.02	−0.18	−0.09
EUR	0.00	1.42	3.33	2.37	1.92	−1.15	0.55	1.99	3.18	2.59	2.38	2.63
JPY	1.56	0.50	2.67	1.73	1.50	1.98	1.60	1.22	2.51	1.87	1.63	1.87
NOK	−0.17	−0.98	0.37	−0.30	−0.49	0.06	0.04	−1.13	−0.48	−0.92	−1.11	−0.30
SEK	−0.20	−1.93	−0.62	−1.50	−1.78	−1.24	−0.12	−1.63	−0.94	−1.43	−1.53	−1.33
NZD	−0.21	−0.80	0.34	−0.57	−0.75	−0.15	0.17	0.51	1.31	0.72	0.60	0.88

Table 2.5: Correlation matrix of return forecasts. This table reports the pairwise correlation between return forecasts for different machine learning models, averaged across all assets. HM is the historical mean computed over an expanding window, OLS is ordinary least squares, LASSO is the least absolute shrinkage and selection operator, EN is elastic net, PLS is partial least squares, RF is random forest, GB is gradient boosting, and KNN is k-nearest neighbours.

	HM	OLS	LASSO	Ridge	EN	PLS	RF	GB
HM	1.00	0.00	0.29	0.01	0.30	0.01	0.20	0.10
OLS	0.00	1.00	−0.01	0.01	−0.01	0.00	−0.01	0.00
LASSO	0.29	−0.01	1.00	0.08	0.86	0.08	0.35	0.23
Ridge	0.01	0.01	0.08	1.00	0.09	0.26	0.04	0.03
EN	0.30	−0.01	0.86	0.09	1.00	0.08	0.35	0.24
PLS	0.01	0.00	0.08	0.26	0.08	1.00	0.04	0.03
RF	0.20	−0.01	0.35	0.04	0.35	0.04	1.00	0.65
GB	0.10	0.00	0.23	0.03	0.24	0.03	0.65	1.00
KNN	0.11	0.00	0.15	0.04	0.15	0.05	0.19	0.20

Table 2.6: Henriksson-Merton directional accuracy tests. This table reports the proportion of correctly forecasted return signs for each forecast combination method. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	Simple	DMSE	MSE(12)	MSE(36)	MSE(60)
Panel A: Equity returns					
Australia	0.62***	0.60***	0.60***	0.60***	0.59***
Canada	0.59***	0.57***	0.58***	0.57***	0.57***
Denmark	0.58***	0.55***	0.55***	0.54***	0.54***
France	0.61***	0.58***	0.58***	0.57***	0.57***
Germany	0.55***	0.53***	0.54***	0.53***	0.53***
Japan	0.51	0.53**	0.54***	0.54***	0.55***
UK	0.60***	0.58***	0.58***	0.58***	0.58***
US	0.63***	0.60***	0.60***	0.60***	0.60***
Panel B: Bond returns					
Australia	0.66***	0.60***	0.61***	0.60***	0.60***
Canada	0.62***	0.58***	0.59***	0.58***	0.57***
Denmark	0.62***	0.58***	0.59***	0.59***	0.58***
France	0.59***	0.55***	0.56***	0.55***	0.55***
Germany	0.58***	0.57***	0.57***	0.56***	0.56***
Japan	0.65***	0.59***	0.59***	0.57***	0.57***
UK	0.56***	0.54***	0.55***	0.54***	0.54***
US	0.59***	0.57***	0.58***	0.57***	0.57***
Panel C: Currency returns					
AUD	0.48	0.45	0.46	0.45	0.44
CAD	0.48	0.44	0.45	0.44	0.45
CHF	0.48	0.49	0.50	0.49	0.49
GBP	0.47	0.47	0.48	0.48	0.47
EUR	0.53***	0.51	0.52**	0.51*	0.50
JPY	0.47	0.46	0.47	0.46	0.45
NOK	0.49	0.46	0.45	0.45	0.45
SEK	0.48	0.47	0.47	0.47	0.46
NZD	0.49	0.48	0.48	0.47	0.46

Table 2.7: Mincer-Zarnowitz regressions for MSE(12) forecasts of equity and bond returns. This table reports results for the regression $r_{t+1} = \alpha + \beta \hat{r}_{t+1} + u_{t+1}$, where r_{t+1} is the realised return on a country's equity or bond market index, and $\hat{r}_{t+1}$ is the forecasted value, averaged across models using the Della Corte and Tsiakas (2012) 'most recently best' MSE weights over the previous 12 months. $p(F)$ is the p-value for an F-test of the joint null hypothesis H_0: $\alpha = 0$ and $\beta = 1$.

	Equity return			**Bond return**		
Country	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{p(F)}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{p(F)}$
Australia	0.00	0.64	0.04	0.00	0.68	0.20
Canada	−0.00	1.14	0.66	0.00	0.65	0.11
Denmark	0.00	0.70	0.03	0.00	0.92	0.87
France	0.00	0.79	0.56	0.00	0.51	0.01
Germany	0.00	0.79	0.45	0.00	0.41	0.00
Japan	0.00	0.56	0.07	0.00	0.75	0.12
UK	0.00	0.71	0.07	0.00	0.94	0.98
US	0.00	0.72	0.15	0.00	0.73	0.06

Table 2.8: Mincer-Zarnowitz regressions for MSE(12) forecasts of currency returns. This table reports results for the regression $r_{t+1} = \alpha + \beta \hat{r}_{t+1} + u_{t+1}$, where r_{t+1} is the realised real return on a currency pair, and $\hat{r}_{t+1}$ is the forecasted value, averaged across models using the Della Corte and Tsiakas (2012) 'most recently best' MSE weights over the previous 12 months. $p(F)$ is the p-value for an F-test of the joint null hypothesis H_0: $\alpha = 0$ and $\beta = 1$.

Currency	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{p(F)}$
AUD	0.00	0.17	0.00
CAD	−0.00	0.04	0.00
CHF	−0.00	0.93	0.97
GBP	−0.00	0.32	0.00
EUR	0.00	0.93	0.97
JPY	0.00	0.71	0.42
NOK	0.00	0.38	0.00
SEK	0.00	0.32	0.00
NZD	0.00	0.55	0.09

Table 2.9: Out-of-sample R^2 by recession subperiod. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns. The sample is split by an NBER recession indicator, extended to 1802 for the US using hand-collected data from Gordon (1986), and defined as two consecutive quarters of negative GDP growth for the other countries. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. '–' means the value is less than −100%. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	No recession						Recession					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	−0.33	−0.21	1.15	0.37	0.04	0.45	−0.31	−0.46	2.05	0.73	0.54	0.54
Canada	1.86**	2.07	2.41**	1.82**	1.65*	1.41	2.82**	6.32**	7.81**	6.79**	6.07**	1.38
Denmark	−0.63	0.39	2.69*	1.43	0.77	1.86	0.60	1.01	2.98*	1.75	1.28	1.89
France	−0.11	−1.32	−0.67	−0.55	−0.67	−0.52	1.38	0.66	2.16*	1.07	−0.28	1.65
Germany	0.83	2.79**	4.74**	3.41**	2.60*	4.04**	−0.71	−1.86	−1.55	−2.85	−3.06	−2.43
Japan	−1.01	−1.74	0.42	−0.15	−0.58	0.18	0.06	−1.09	0.99	−0.37	−0.53	0.31
UK	−1.18	−1.59	−0.89	−1.09	−0.99	−0.91	1.27*	0.92	2.18**	1.77	0.99	2.02*
US	−0.31	−0.27	0.34	0.52	0.16	0.28	0.13	0.28	1.74	1.12	0.51	1.55
Panel B: Bond returns												
Australia	−0.17	−0.23	−0.02	−0.19	−0.47	−0.13	0.54	0.78	1.33	0.67	0.12	0.38
Canada	1.33*	0.41	0.61	−0.82	−1.53	0.67	0.11	0.60	2.54*	1.67	1.70	2.03*
Denmark	1.91*	1.08	2.73*	1.72	1.37	1.98	0.93	1.49	3.21**	2.45*	2.14*	2.59*
France	−1.61	−2.87	−1.25	−2.32	−2.76	−2.21	0.61	1.03	2.70**	1.97*	1.69	2.08*
Germany	3.70	5.89	6.82	5.34	4.52	5.49	–	−62.34	−30.23	−48.23	−53.92	−38.39
Japan	0.37	0.52	1.03	0.45	−0.05	0.39	0.69	1.55	3.31**	2.41*	1.92	2.48*
UK	0.98*	0.65	1.30*	1.08	0.83	1.04	0.04	0.94	2.12**	1.33*	0.61	1.30
US	0.05	0.86	2.47**	1.48	1.13	1.81*	1.68*	1.46	3.10**	2.26*	1.84*	2.36*
Panel C: Currency returns												
AUD	−0.91	−1.12	−1.56	−2.13	−2.11	−1.68	−0.33	−0.40	0.25	−0.09	−0.08	−0.00
CAD	0.04	−0.89	−1.28	−1.71	−1.37	−1.39	−0.08	−1.34	−0.61	−0.78	−0.92	−0.71
CHF	1.81	2.34	3.87	3.43	3.12	3.48	1.01	−1.08	0.38	1.01	0.43	0.69
GBP	−0.21	−1.13	−1.34	−1.29	−1.84	−1.40	0.74	0.81	0.38	0.18	−0.40	0.22
EUR	–	−1.23	0.16	−0.37	−0.98	0.24	5.83	6.49	8.32	7.62	7.35	7.83
JPY	−0.07	−0.19	1.76	1.13	1.07	1.22	−0.04	0.43	2.89	1.47	1.22	2.30
NOK	−0.43	−1.37	−0.29	−0.70	−1.39	−0.45	−0.18	−0.30	1.33	0.61	0.22	0.96
SEK	−0.15	−1.99	−1.45	−1.82	−2.34	−1.91	−0.35	−0.47	1.44	0.32	−0.21	0.49
NZD	−0.05	−1.00	0.19	−0.83	−1.08	−0.35	−1.02	0.99	1.80	1.22	1.05	0.94

Table 2.10: Out-of-sample R^2 by volatility regime. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns. The sample is split by the median of each asset's own historical volatility. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. '–' means the value is less than −100%. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	No recession						Recession					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	−0.25	−0.95	0.34	−0.79	−1.12	0.05	−0.31	0.25	2.21*	1.49	1.18	1.25
Canada	2.27**	2.98**	5.28**	4.44**	3.95**	4.82**	2.52**	4.85**	7.01**	5.62**	4.55**	6.94**
Denmark	−0.53	−0.07	1.53	0.64	0.38	1.19	−0.72	0.99	3.07*	2.00	1.12	2.37
France	0.40	−0.94	0.80	−0.15	−1.01	0.58	1.39	1.54	2.88	1.79	0.95	2.18
Germany	0.30	0.10	0.83	−0.84	−1.49	0.57	0.28	2.16	4.18**	3.13*	2.57	2.06
Japan	0.59	−0.25	0.56	−0.25	−0.40	−0.18	−1.64	−2.12	1.05	−0.62	−0.92	−0.08
UK	0.38	0.27	0.14	0.19	−0.08	0.23	1.32**	0.71	1.54**	1.00	0.88	1.48*
US	−0.13	−0.04	0.63	0.48	0.54	0.58	0.18	0.36	1.43	0.86	0.47	1.22
Panel B: Bond returns												
Australia	0.18	−0.33	0.08	−0.56	−0.45	−0.38	0.68**	0.73	2.15*	1.32	0.68	1.18
Canada	0.98	0.54	0.05	0.29	−1.11	−0.58	0.58*	1.39	2.45*	1.67	1.32	1.33
Denmark	0.83	0.47	1.23	1.38	0.96	0.87	1.19	2.10	4.06**	2.93**	2.60*	2.65*
France	−0.15	−0.69	1.32	0.43	0.14	0.76	−0.10	0.36	1.26	0.73	0.49	0.89
Germany	5.04	9.52	11.28	10.27	9.88	10.92	–	−38.66	−17.54	−30.63	−36.26	−32.39
Japan	0.35	0.68	1.39	0.54	0.47	0.90	0.68	2.12*	3.79**	3.44*	2.61	3.41*
UK	−0.09	0.06	1.17	0.62	0.39	0.54	1.05	1.75	2.81*	2.06	1.65	2.18
US	0.47*	−0.01	1.87*	1.02	0.61	1.09	0.93	1.78	3.02*	2.67*	2.31	2.44
Panel C: Currency returns												
AUD	−0.15	−1.88	−1.36	−2.42	−2.78	−1.84	−0.54	−0.87	0.37	0.02	−1.10	−0.17
CAD	−0.08	−1.57	−1.29	−1.91	−2.06	−1.87	0.21	0.38	−0.06	−0.31	−0.47	−0.45
CHF	−0.45	−1.02	−0.76	−1.32	−1.27	−1.01	2.07	2.71	5.98	5.33	4.72	5.18
GBP	−0.09	−1.48	−0.43	−1.01	−1.29	−0.99	−0.62	−1.43	−0.41	−0.80	−1.63	−0.78
EUR	–	1.65	4.29	2.59	2.09	3.68	−1.08	0.37	1.39	0.54	0.16	1.28
JPY	−0.13	−0.24	1.03	1.12	0.74	1.21	0.17	1.14	2.08	1.82	1.57	2.28
NOK	−0.32	−1.71	−0.65	−1.32	−1.46	−1.17	−0.42	−1.95	−0.12	−0.94	−1.22	−0.37
SEK	−0.17	−1.35	−0.95	−1.78	−1.96	−1.39	−0.08	−0.63	0.13	0.04	−0.16	−0.01
NZD	−0.49	−1.50	0.33	−0.86	−1.17	0.21	−0.09	0.32	1.03	0.25	0.17	0.88

Table 2.11: Out-of-sample R^2 by inflation environment. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns. The sample is split by the median of each country's CPI inflation. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. '–' means the value is less than −100%. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	Low inflation						High inflation					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	0.06	−0.19	0.59	−0.11	−0.33	0.36	−0.43	−0.04	1.51	0.69	0.40	1.06
Canada	1.87**	4.01**	6.07**	4.96**	3.94**	5.24**	2.13**	3.00**	4.98**	3.89**	3.11**	4.14**
Denmark	−5.20	−0.33	0.46	−0.49	−0.93	0.11	−0.94	0.54	2.91*	1.54	0.96	2.03
France	2.38	2.02	3.51	2.11	−0.07	1.87	1.06	−0.03	1.67	0.49	−0.71	0.90
Germany	2.06	4.51*	7.14**	5.04*	3.92*	6.18**	0.33	0.94	2.22	0.84	0.18	1.99
Japan	−3.64	−1.71	0.79	−0.02	−0.26	0.38	−0.26	−1.64	0.43	−0.96	−1.20	0.27
UK	−0.33	−0.15	0.61	−0.13	−0.30	0.43	0.29	0.34	1.20**	0.53	0.35	0.90*
US	0.18	−0.20	0.45	−0.00	−0.14	0.03	−0.05	0.25	1.15*	0.59	0.39	0.48
Panel B: Bond returns												
Australia	0.46	0.61	0.82	0.34	0.37	0.41	0.23	0.05	0.85	−0.04	−0.29	−0.13
Canada	1.30	0.19	−0.14	−1.30	−1.84	−0.69	0.96	1.38	2.06	0.82	0.34	1.40
Denmark	2.03**	2.69*	3.84**	3.08*	2.92*	3.18*	1.17*	0.50	2.24*	1.31	0.88	1.74
France	−0.28	−0.90	0.87	0.02	−0.35	0.66	−0.26	−0.00	1.63	0.88	0.54	1.06
Germany	–	−0.18	8.04	4.07	2.35	6.73	–	−16.67	−4.93	−12.42	−15.58	−8.32
Japan	0.98	−0.10	0.68	0.28	−0.35	0.34	1.08**	2.18**	3.48**	2.40*	1.97	3.02*
UK	0.60	0.97	2.00	1.31	1.00	1.92	0.66	1.14	2.20**	1.52	1.18	1.66
US	0.95	0.61	2.01	1.63	1.31	1.81	0.68	0.08	1.26	0.79	0.46	0.75
Panel C: Currency returns												
AUD	−0.12	1.46	1.41	0.70	0.58	0.98	−1.01	−1.78	−1.18	−2.50	−2.79	−1.83
CAD	−0.05	−1.95	−0.66	−1.59	−1.43	−0.71	0.09	−1.98	−0.78	−1.41	−1.10	−1.18
CHF	3.57	5.14	8.11	7.07	6.61	7.48	1.68	1.93	3.73	2.81	2.47	3.04
GBP	−0.77	−1.64	−0.63	−1.46	−1.79	−0.80	−0.49	−1.52	−0.50	−0.92	−1.69	−1.13
EUR	–	−6.11	−4.27	−5.38	−5.75	−4.84	–	3.37	5.60	4.57	3.98	4.81
JPY	−2.80	−6.67	−4.69	−5.65	−5.90	−5.22	0.83	2.11	4.64	3.53	3.31	3.77
NOK	−0.57	−0.90	0.08	−0.41	−0.52	−0.38	−0.56	−1.04	0.46	−0.38	−0.59	0.12
SEK	0.11	−1.76	−0.01	−1.47	−1.73	−0.91	−0.22	−2.02	−0.61	−1.61	−1.94	−1.04
NZD	−0.74	0.71	1.53	1.00	0.85	1.39	−0.12	−0.96	0.32	−0.75	−0.96	0.09

Table 2.12: Out-of-sample R^2 by interest rate environment. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns. The sample is split by the median of each country's 10-year government bond yield. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. '–' means the value is less than −100%. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	Low interest rates						High interest rates					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	1.31*	1.03	1.71	1.35	1.12	1.48	−0.36	−0.23	1.22	0.44	0.15	0.38
Canada	−4.66	−10.09	−10.25	−10.66	−10.92	−10.82	2.49**	3.96**	6.06**	5.01**	4.28**	5.54**
Denmark	0.20	−3.84	−1.41	−2.97	−3.45	−2.25	−20.16	0.41	2.42	1.17	0.68	1.85
France	0.44	−1.12	0.54	−0.53	−0.46	0.07	1.00	0.11	1.70	0.64	−0.43	0.98
Germany	−0.05	−2.74	−1.54	−2.70	−2.99	−1.88	0.29	1.20	2.58	1.26	0.65	2.04
Japan	0.02	0.52	3.12	3.25	1.85	2.96	−0.82	−1.40	0.73	−0.50	−0.74	0.31
UK	−0.38	0.26	0.89	−0.00	−0.26	−0.01	0.41	0.10	1.01*	0.44	0.31	0.87*
US	−0.33	−1.77	−1.03	−1.33	−1.45	−1.28	0.03	0.18	1.02*	0.51	0.33	0.93*
Panel B: Bond returns												
Australia	1.49*	1.16	1.92	1.10	1.23	1.49	0.35*	0.16	1.01	0.18	−0.03	0.22
Canada	1.84**	0.28	1.33	0.38	0.53	0.83	0.85	0.47	1.27	0.07	−0.36	0.18
Denmark	−0.20	−5.74	−3.55	−4.64	−4.70	−4.04	1.59*	1.33	2.95**	2.06*	1.70	2.13*
France	0.08	−3.75	−3.01	−3.46	−3.68	−3.19	−0.22	−0.35	1.33	0.49	0.18	0.51
Germany	−0.44	−7.63	−8.37	−9.72	−9.81	−8.99	–	−15.21	−3.49	−10.70	−13.79	−7.38
Japan	0.03	3.76	4.41	3.60	2.93	3.04	0.96	1.28	2.61**	1.64	1.15	2.13*
UK	2.13	2.91	3.46	2.86	2.49	2.57	−0.51	−0.32	1.15*	0.44	0.13	0.74
US	0.08	2.90	6.17**	3.18	3.42	4.14	0.74	0.28	1.65	1.17	0.81	1.12
Panel C: Currency returns												
AUD	0.20*	−1.95	−2.33	−3.01	−2.99	−2.93	−0.79	−1.30	−0.85	−2.04	−2.30	−1.94
CAD	−0.03	−1.49	−1.10	−1.37	−1.36	−1.26	0.04	−1.95	−0.91	−1.45	−1.25	−1.18
CHF	0.05	−1.06	0.35	−1.08	−0.98	−0.03	1.23	1.36	3.15	2.27	1.91	2.35
GBP	0.01	−0.99	−0.37	−0.77	−0.93	−0.48	−0.38	−1.49	−0.49	−0.93	−1.59	−0.53
EUR	0.18	−3.36	−1.98	−2.68	−2.75	−2.13	–	1.09	3.16	2.12	1.60	2.40
JPY	0.34	−0.18	0.68	−0.41	−0.83	0.38	0.04	0.12	2.69	1.55	1.31	1.87
NOK	−0.01	−2.13	−1.84	−2.04	−2.12	−1.90	−0.38	−1.08	0.36	−0.38	−0.59	0.07
SEK	−0.10	−1.63	−0.50	−1.33	−1.40	−0.16	−0.14	−1.82	−0.39	−1.31	−1.62	−0.84
NZD	0.05	−3.71	−3.28	−3.47	−3.63	−3.79	−0.28	−0.42	0.79	−0.19	−0.39	−0.41

Table 2.13: Out-of-sample R^2 pre- and post-WW2. This table reports out-of-sample R^2s in percent relative to the historical mean for international equity, bond, and currency returns. The sample is split by an indicator variable that equals 1 after September 1945. Simple is the equally-weighted average of all forecasts. DMSE uses the Stock and Watson (2004) discounted mean squared error weights. MSE(κ) uses the Della Corte and Tsiakas (2012) 'most recently best' weights over the previous κ months. TMA is an equally-weighted average of the three best performing models in the estimation period. '–' means the value is less than −100%. Significance is based on the Clark and West (2006, 2007) test. ***, **, * denote significance at the 1%, 5%, and 10% level respectively.

	Pre-WW2						Post-WW2					
	Simple	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**	**Simple**	**DMSE**	**MSE(12)**	**MSE(36)**	**MSE(60)**	**TMA**
Panel A: Equity returns												
Australia	0.02	−0.74	0.56	−0.32	−0.67	0.19	−0.60	0.37	1.95	1.30	1.07	1.42
Canada	2.83**	5.09**	7.12**	6.08**	5.28**	6.59**	1.78*	1.36	3.49**	2.48*	1.89	2.91**
Denmark	−0.17	0.13	2.75	1.35	0.52	1.79	−0.39	0.39	2.12	0.93	0.60	1.18
France	1.62	0.73	2.15	0.96	−0.60	1.82	0.20	−0.72	1.10	0.18	−0.22	0.65
Germany	0.70	1.43	2.06	0.76	0.06	1.78	−0.31	0.76	3.24**	1.88	1.44	2.04
Japan	−1.76	−2.19	1.00	−0.01	−0.23	0.37	−0.21	−0.82	0.58	−0.80	−1.05	−0.09
UK	−0.26	−0.24	0.57	−0.16	−0.39	0.28	0.75	0.65	1.50	0.92	0.83	1.19
US	0.49	−0.09	0.58	0.14	0.00	0.27	−0.56	0.46	1.52*	0.90	0.69	1.13
Panel B: Bond returns												
Australia	0.20	−0.38	−0.03	−0.66	−0.90	−0.28	0.50	0.75	2.12*	1.09	0.92	1.47
Canada	1.29	−0.09	0.14	−1.50	−2.09	−0.39	0.46	1.07	2.49	1.76	1.54	1.88
Denmark	2.01*	1.15	2.84*	2.00	1.60	2.39*	0.74	1.52	3.02**	2.01	1.74	2.56*
France	−0.24	−0.27	1.45	0.61	0.26	1.09	−0.14	−0.77	0.72	−0.08	−0.24	0.48
Germany	–	−19.52	−5.33	−14.12	−17.86	−10.91	−0.07	−0.00	2.88	1.26	0.50	0.26
Japan	1.19**	1.89	3.68*	2.85	2.13	3.05*	0.77	0.89	1.84	0.77	0.44	1.18
UK	0.73	1.41	2.46*	1.73	1.36	2.49*	0.02	−0.11	1.09	0.62	0.40	1.20
US	0.87	0.42	1.80	1.43	1.06	1.50	0.46	0.06	1.47	0.62	0.31	1.33
Panel C: Currency returns												
AUD	−0.60	−0.95	−1.11	−2.60	−2.81	−1.93	−0.87	−1.65	−0.70	−1.60	−1.88	−1.59
CAD	0.07	−1.88	−0.81	−1.36	−1.11	−1.01	−0.13	−2.22	−1.41	−1.89	−1.96	−1.52
CHF	2.04	2.70	4.95	4.09	3.69	4.39	−0.38	−1.38	−0.51	−1.45	−1.69	−0.89
GBP	−0.49	−1.34	−0.41	−0.76	−1.58	−0.57	−0.15	−1.79	−0.63	−1.30	−1.56	−0.86
EUR	–	2.47	4.50	3.45	2.93	3.79	−0.45	−3.92	−1.75	−2.75	−3.25	−2.12
JPY	−3.33	−1.86	1.50	0.34	0.13	1.08	3.10	2.40	4.02	2.90	2.61	3.03
NOK	−0.53	−0.71	0.82	0.09	−0.14	0.29	−0.01	−2.23	−1.09	−1.83	−1.99	−1.31
SEK	−0.21	−1.82	−0.25	−1.34	−1.73	−0.74	−0.03	−1.82	−0.61	−1.26	−1.43	−1.01
NZD	−0.86	−1.07	−1.69	−1.92	−1.85	−1.72	−0.11	−0.38	1.46	0.25	−0.05	1.38

Table 2.14: Summary statistics for strategy returns. This table reports summary statistics for returns of four multi-asset trading strategies. $1/N$ is an equal-weighted portfolio. $1/N$-neutral allocates equal weight to each asset class, and then equally-weights assets within an asset class. MVO is the mean-variance optimal portfolio using historical data, and MVO-ML is the mean-variance optimal portfolio using the historical covariance matrix but machine learning predictions for expected returns. Means, standard deviations, downside deviations, Sharpe ratios, and Sortino ratios are annualised. τ_{BE} is the break-even transaction cost relative to the benchmark in brackets. $\hat{\Theta}$ is the Goetzmann, Ingersoll, Spiegel, and Welch (2007) manipulation-proof performance measure in annualised percent.

	$1/N$	$1/N$-neutral	MVO	MVO-ML
Mean	8.97	10.18	12.93	14.86
Std dev.	11.58	11.78	10.73	10.74
Downside dev.	8.99	8.65	7.32	6.98
Sharpe	0.46	0.56	0.87	1.04
Sortino	1.00	1.18	1.77	2.12
Max. drawdown	−74.69	−73.61	−53.06	−30.92
Min	−21.54	−17.23	−11.38	−12.19
Median	0.73	0.86	1.10	1.11
Max	21.48	14.29	13.76	18.92
Skew	0.09	−0.09	−0.07	0.47
Kurtosis	6.36	2.72	1.61	3.14
τ_{BE} $(1/N)$	–	–	531	617
τ_{BE} (MVO)	–	–	–	79
$\hat{\Theta}$	3.84	4.91	8.25	10.18

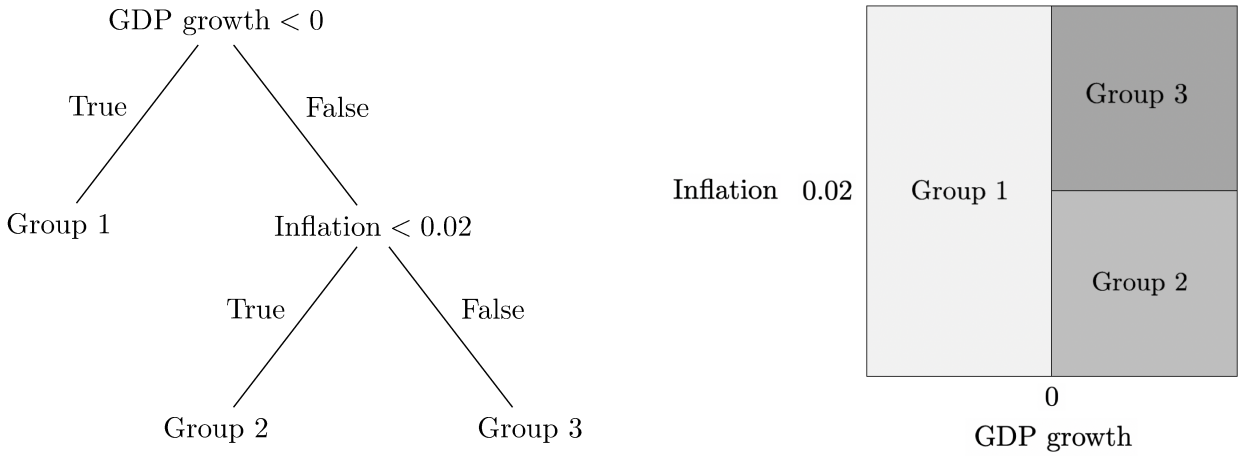


Figure 2.1: Illustrative decision tree.

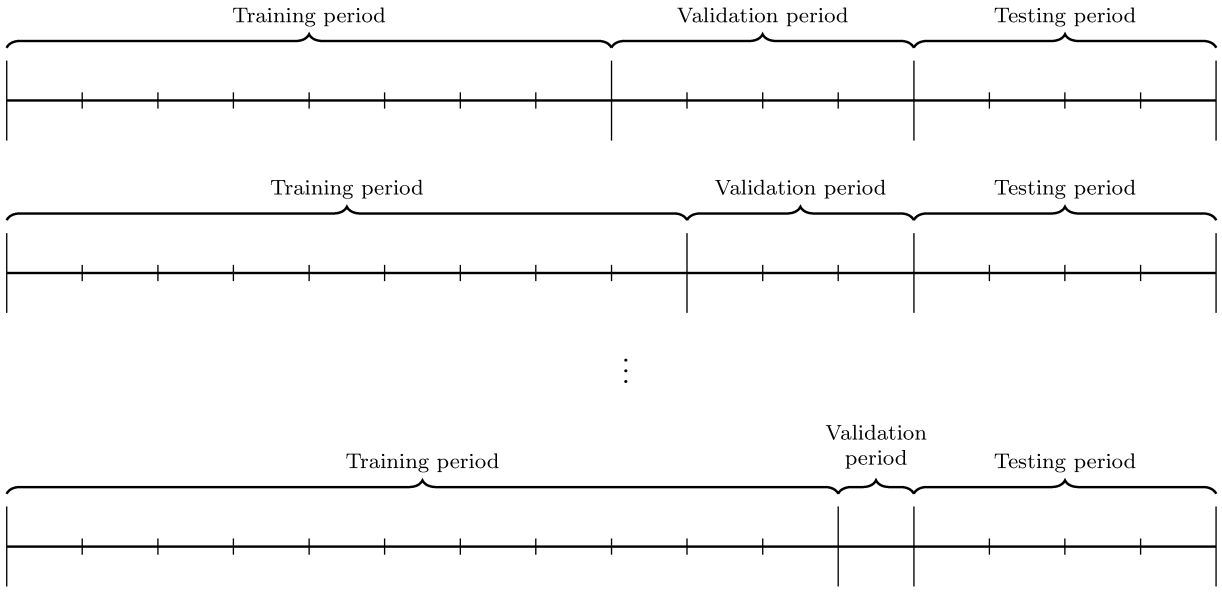
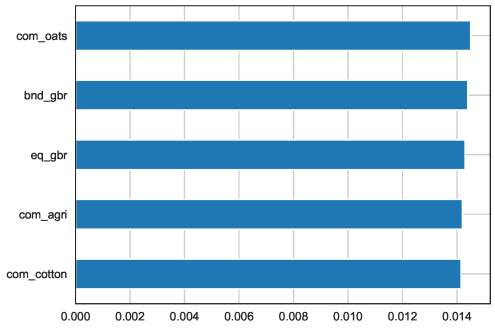
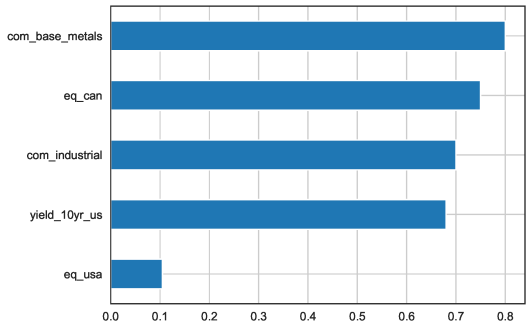


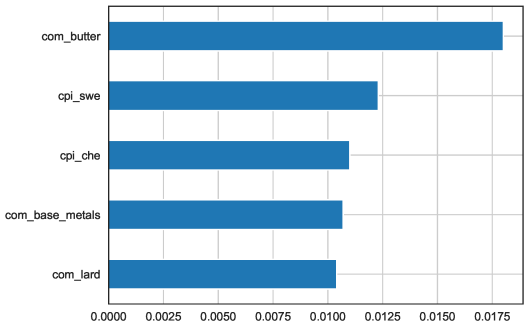
Figure 2.2: Walk forward validation procedure.



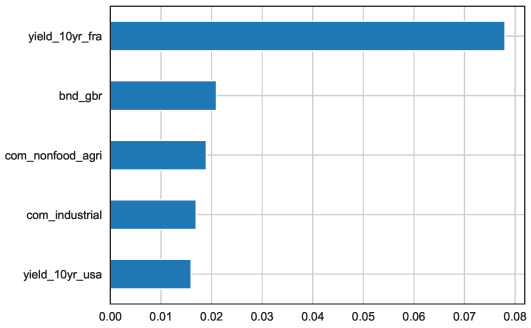
(a) Australia



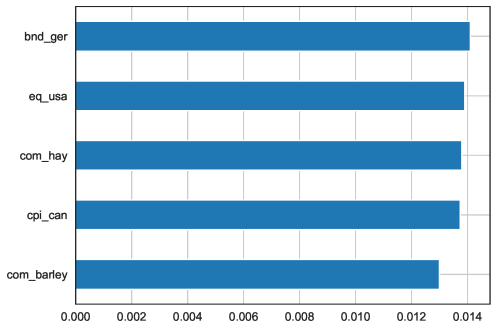
(b) Canada



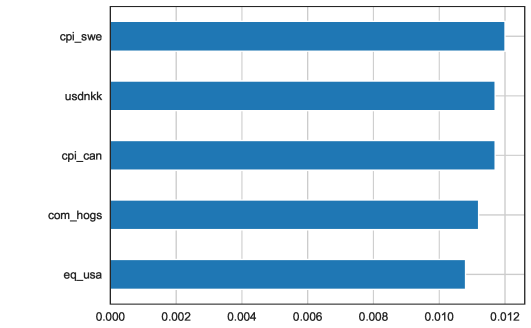
(c) Denmark



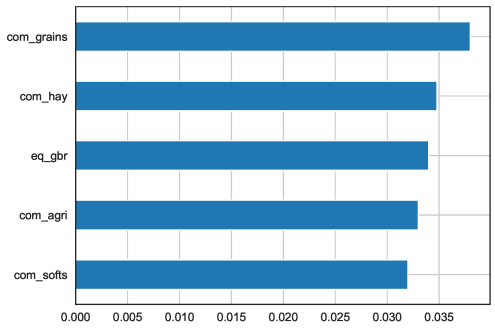
(d) France



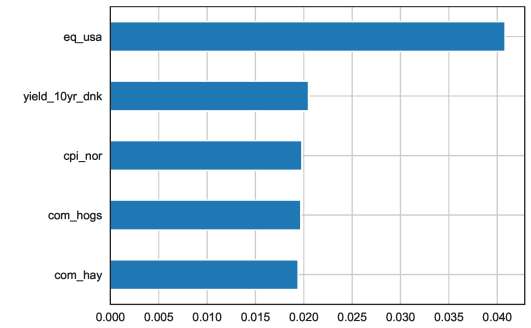
(e) Germany



(f) Japan

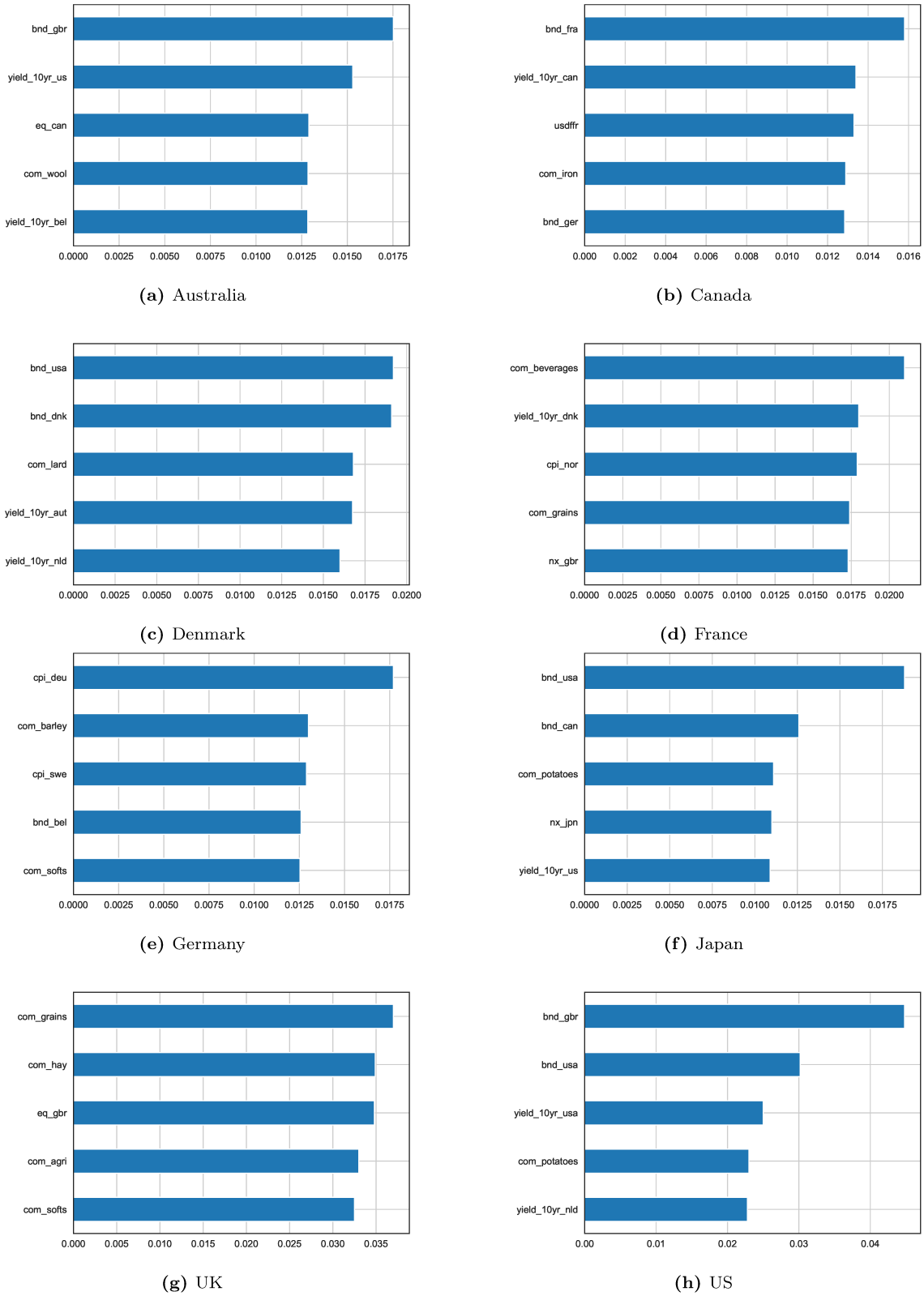


(g) UK



(h) US

Figure 2.3: Variable importance for international equity markets. This figure plots the variable importance (VI) for the top 5 most important variables averaged across all models. VI is calculated as the change in out-of-sample R^2 when the variable is excluded from the model, scaled by the sum of these changes across all variables.



(a) Australia **(b)** Canada

(c) Denmark **(d)** France

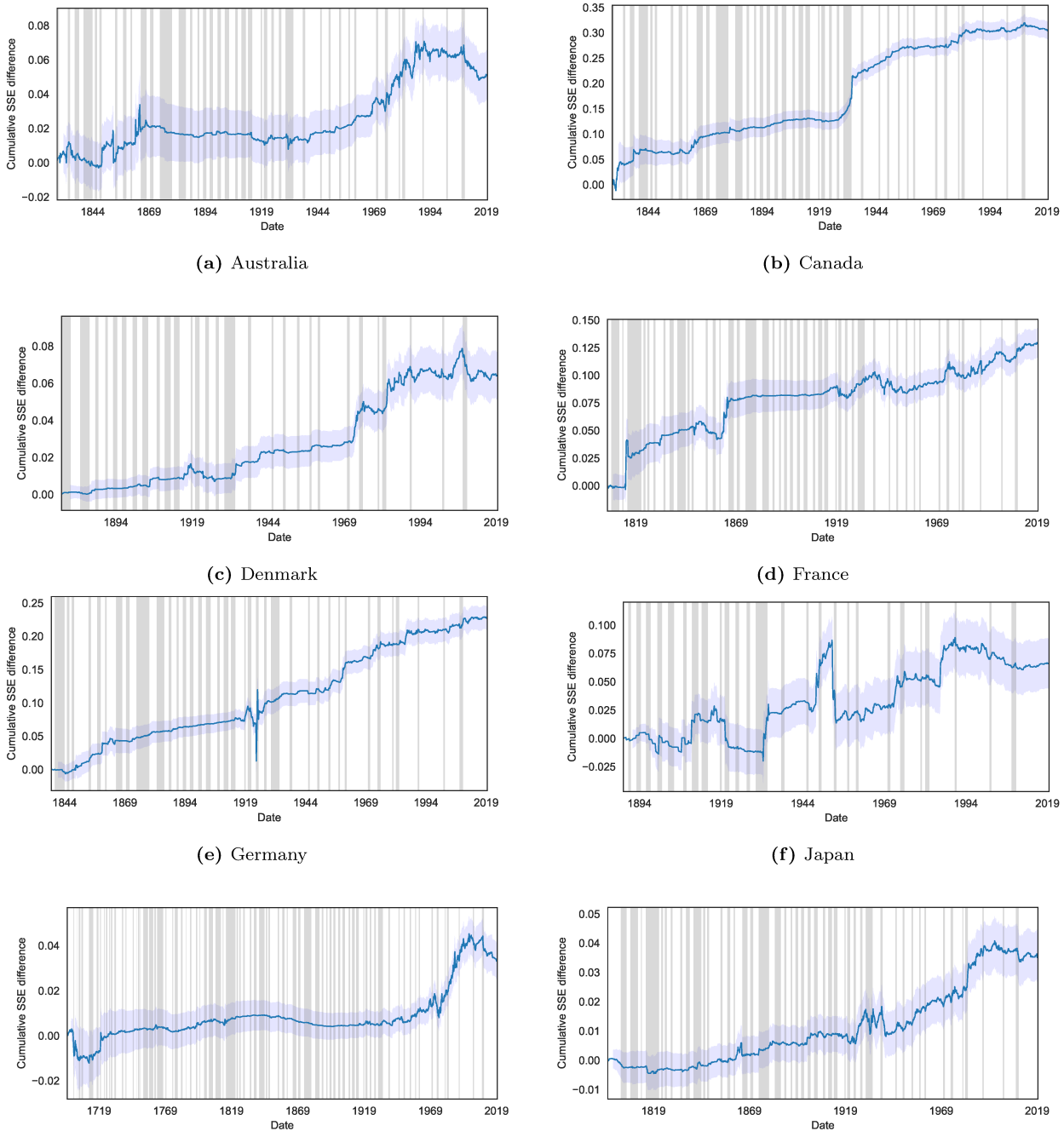
(e) Germany **(f)** Japan

(g) UK **(h)** US

Figure 2.4: Variable importance for international government bond markets. This figure plots the variable importance (VI) for the top 5 most important variables averaged across all models. VI is calculated as the change in out-of-sample R^2 when the variable is excluded from the model, scaled by the sum of these changes across all variables.



Figure 2.5: Variable importance for real currency returns. This figure plots the variable importance (VI) for the top 5 most important variables averaged across all models. VI is calculated as the change in out-of-sample R^2 when the variable is excluded from the model, scaled by the sum of these changes across all variables.



(a) Australia

(b) Canada

(c) Denmark

(d) France

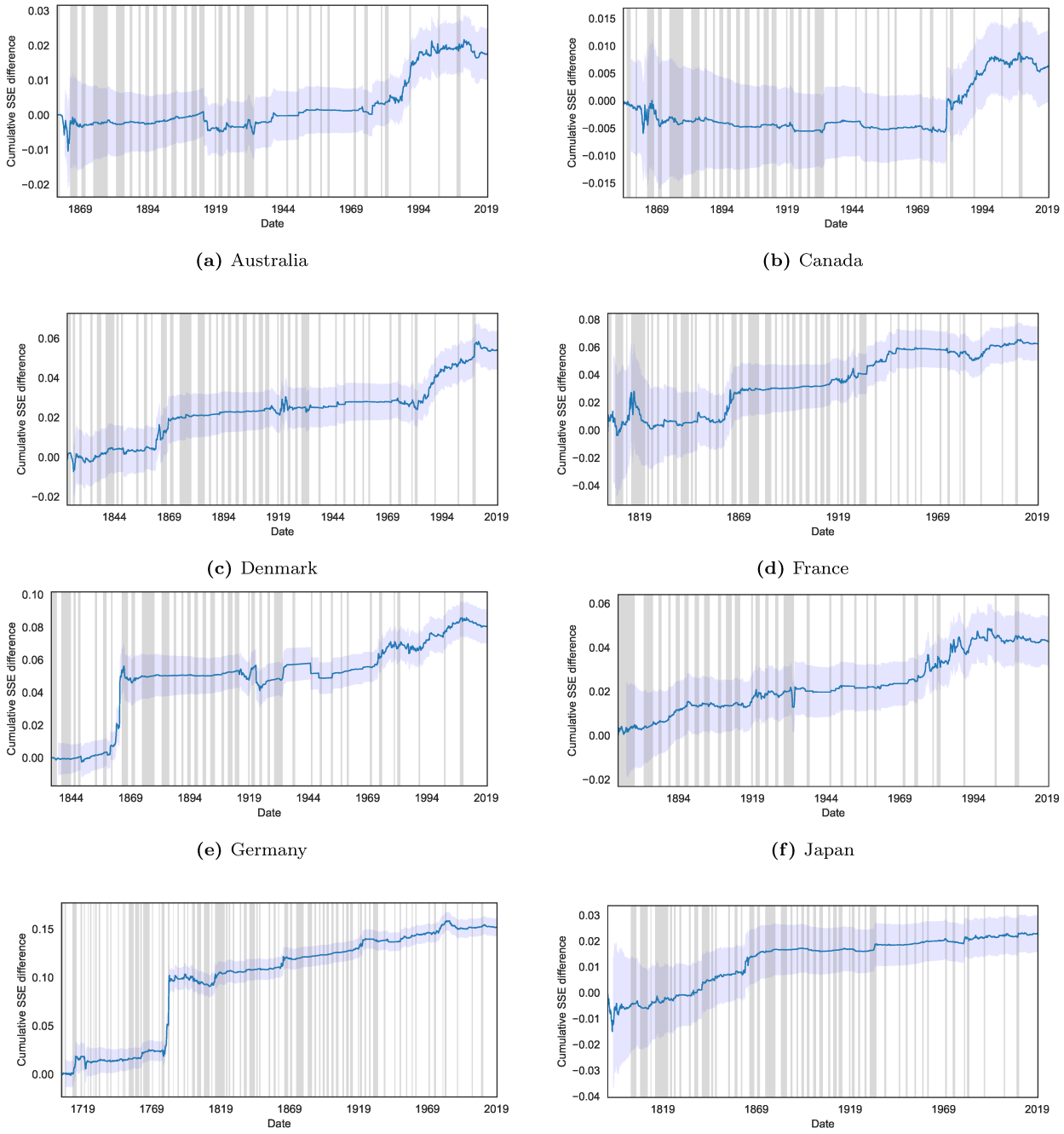
(e) Germany

(f) Japan

(g) UK

(h) US

Figure 2.6: Cumulative MSE difference for international equity markets. This figure shows the cumulative squared prediction errors of the historical mean benchmark model minus the cumulative squared prediction errors of the MSE(12) combined forecast model. An increase in the line indicates better performance of the forecast model, and vice versa for a decrease. The grey bars represent recessions. The blue bands represent a 95% two-sided confidence interval based on the Clark and West (2006, 2007) statistic.



(a) Australia

(b) Canada

(c) Denmark

(d) France

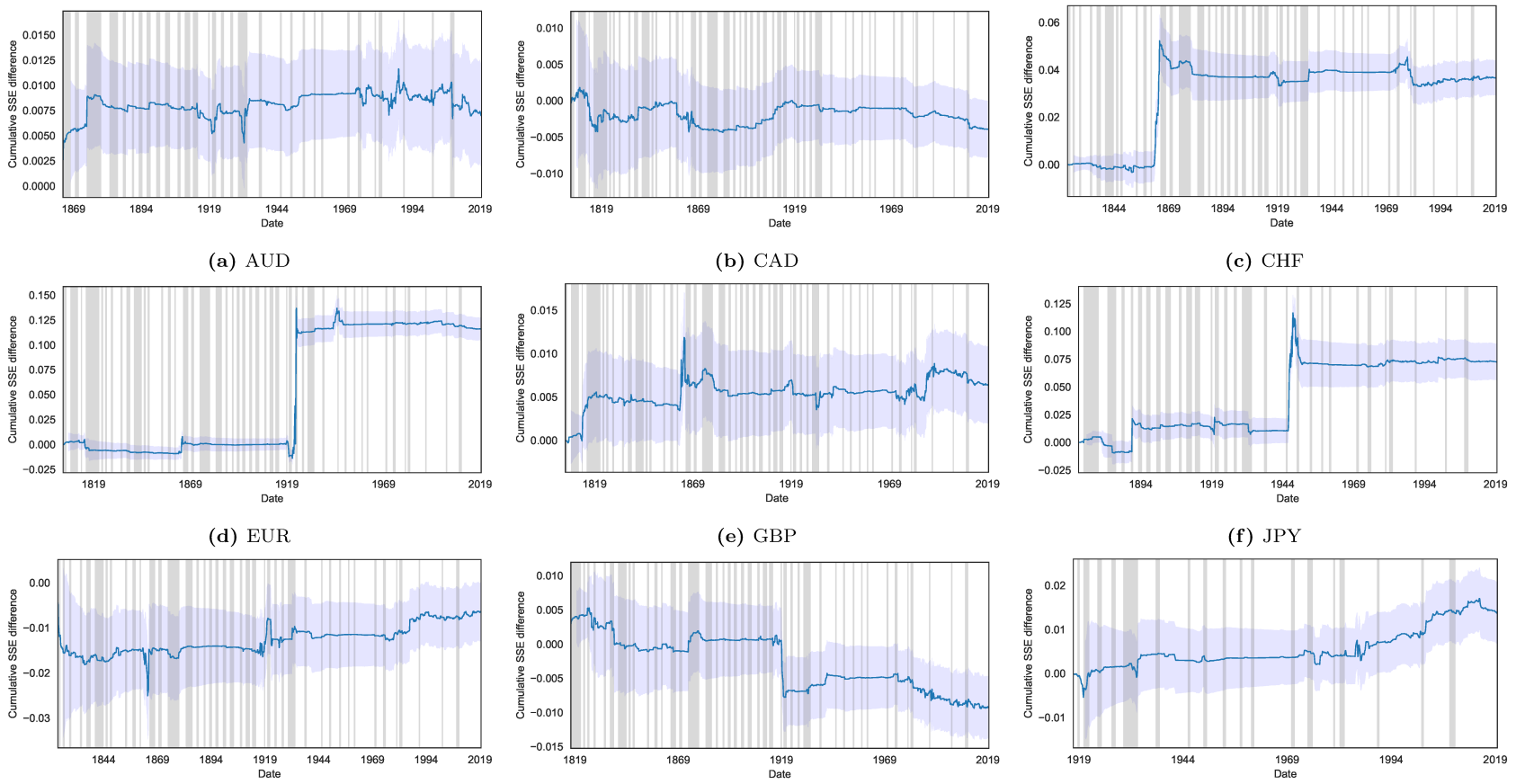
(e) Germany

(f) Japan

(g) UK

(h) US

Figure 2.7: Cumulative MSE difference for international government bond markets. This figure shows the cumulative squared prediction errors of the historical mean benchmark model minus the cumulative squared prediction errors of the MSE(12) combined forecast model. An increase in the line indicates better performance of the forecast model, and vice versa for a decrease. The grey bars represent recessions. The blue bands represent a 95% two-sided confidence interval based on the Clark and West (2006, 2007) statistic.



(a) AUD **(b)** CAD **(c)** CHF

(d) EUR **(e)** GBP **(f)** JPY

(g) NOK **(h)** SEK **(i)** NZD

Figure 2.8: Cumulative MSE difference for real currency returns. This figure shows the cumulative squared prediction errors of the historical mean benchmark model minus the cumulative squared prediction errors of the MSE(12) combined forecast model. An increase in the line indicates better performance of the forecast model, and vice versa for a decrease. The grey bars represent recessions. The blue bands represent 95% two-sided confidence intervals based on the Clark and West (2006, 2007) statistic.

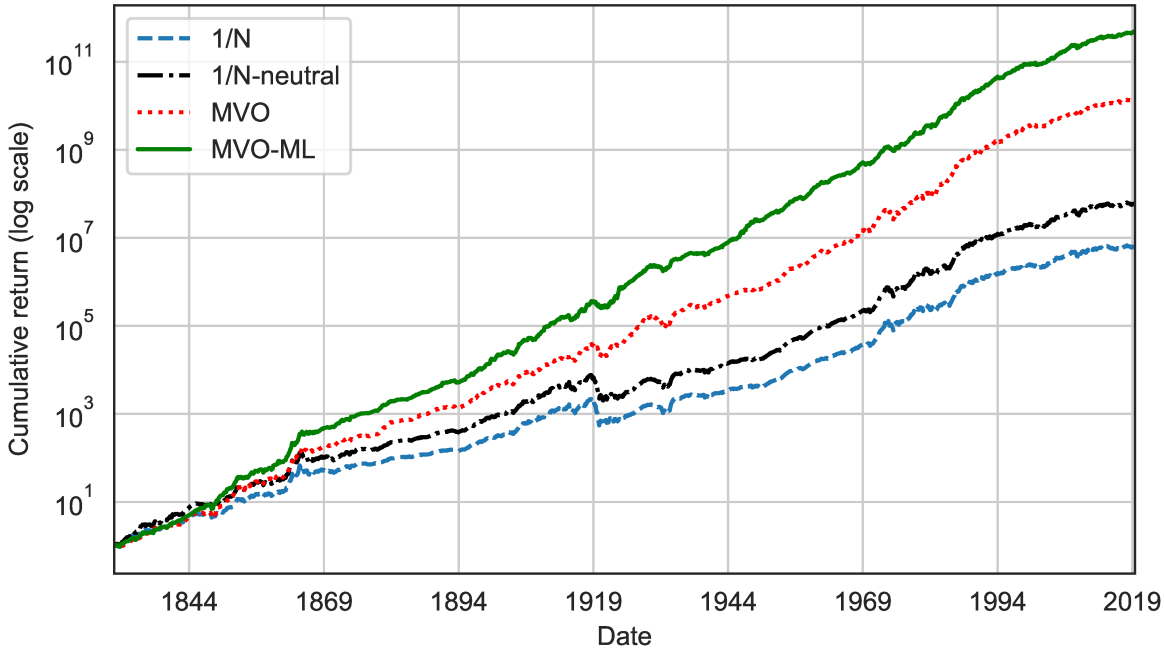


Figure 2.9: Cumulative returns for trading strategies. This figure plots the cumulative returns on a log scale for the four trading strategies. $1/N$ is an equal-weighted portfolio, $1/N$-neutral allocates equal weight to each asset class, and then equally-weights assets within an asset class. MVO is the mean-variance optimised portfolio using historical data, and MVO-ML is the mean-variance optimised portfolio using the historical covariance matrix but machine learning predictions for expected returns.

Chapter 3

Cross-Factor Momentum

3.1 Introduction

Factor momentum—or more generally, time series momentum (TSM)—strategies have received increasing attention in practice and academia in recent years. Moskowitz, Ooi, and Pedersen (2012) document significant TSM in 58 liquid futures written on equity indices, currencies, commodities, and bonds. Gupta and Kelly (2019) find similar results for 65 characteristic-based global equity factors. These papers consider TSM in a univariate sense: the strategies are constructed independently for each asset using that asset's own past returns. In this chapter, I build on this literature by examining whether factor momentum strategies can be enhanced by incorporating time series predictability from other sources beyond an asset's own lagged return.

I begin by reviewing the in-sample (IS) and out-of-sample (OOS) evidence on time series factor momentum (TSFM) for a set of widely used US equity factors. Consistent with previous results I find significant IS predictability for factor returns using the level and sign of the factor's own lagged returns, which is strongest at a one month lag and quickly decays to zero at higher lags. This finding is remarkably consistent across factors in univariate regressions and pooled regressions with factor fixed effects. Moreover, in light of recent evidence that the OOS performance of TSM in asset class futures is poor (Huang, Li, Wang, and Zhou, 2020), I also conduct OOS predictive regressions for each factor using OLS and a robust shrinkage technique called LASSO.[1] In contrast to asset class futures, the OOS performance of lagged returns applied to factors is much stronger, with positive OOS R^2s and several non-zero LASSO coefficients. Finally, I construct TSFM strategies in the style

[1]LASSO stands for least absolute shrinkage and selection operator, and is widely used in machine learning for prediction and feature selection.

of Gupta and Kelly (2019) and find mixed evidence that these strategies outperform the raw factors in my sample. TSFM generates improvements in the Sharpe ratio of only two of the seven factors (size and beta), and similar or lower Sharpe ratios for the remaining factors.

In the next section I document material cross-correlations between factors, implying past returns of other factors may be informative for predicting future factor returns.[2] Most of the factors I consider have significant first-order cross-correlations with at least one or two other factors, with absolute values ranging from 0.06 to 0.22. These magnitudes are comparable to, or exceed, many of the factors' first-order autocorrelations. IS multiple regressions also generate significant loadings on other factors, with large t-statistics in both level and sign specifications. Moreover, the OOS evidence is also promising, with positive OOS R^2s in the range of 1–4 percent, and several non-zero LASSO coefficients. Overall the results support the existence of a strong and robust short-horizon cross-correlation structure in US equity factors.

Motivated by these findings, I construct *cross-factor momentum* (CFM) strategies that use a combination of the lagged returns on all factors to inform directional trading decisions for each individual factor. The CFM strategies are constructed similarly to standard TSM or TSFM strategies such as those in Moskowitz, Ooi, and Pedersen (2012) and Gupta and Kelly (2019), but the trading signal is a weighted average of the lagged returns of all factors for a given weighting scheme. In the analysis I consider four schemes to weight the lagged returns: an equally-weighted '$1/N$' average, historical pairwise correlations, OLS regression coefficients, and LASSO regression coefficients. I show that a historical correlation or OLS coefficient weighted signal outperforms the standard univariate TSFM signals in terms of risk-adjusted returns, generating Sharpe ratios that are up to 1.7 times higher. The CFM strategies also exhibit lower drawdowns and have similar trading costs to the standard strategies.

Finally, I examine some potential explanations for why cross-factor predictability exists. In this section I show that my results cannot be explained by shocks to the capital ratio of financial intermediaries. The alphas generated by the CFM strategies are not spanned by the He, Kelly, and Manela (2017) tradable intermediary asset pricing factor, or the five-factor model of Fama and

[2]Cross-correlations are widely used in signal processing and measure the correlation of a series with lagged values of another series, i.e. $\text{corr}(x_t, y_{t-k})$ for some lag k.

French (2015). A channel based on factor ETF flows offers a partial explanation, but cannot fully account for the findings, adding an extra hurdle for further theoretical work on asset pricing factors.

The rest of the chapter proceeds as follows. Section 3.2 reviews the literature on TSM and highlights its differences with cross-sectional momentum, Section 3.3 describes the sample, Section 3.4 reviews the literature and evidence on TSFM, Section 3.5 documents the results on cross-factor predictability and constructs the CFM strategies, Section 3.6 examines some potential explanations for CFM, and Section 3.7 concludes.

3.2 Time Series Momentum

In this section I review the literature on TSM in general (not just applied to factors), and highlight its differences with cross-sectional momentum. This section is mostly self-contained and provides additional context for the factor momentum strategies.

TSM (sometimes called trend following) broadly refers to a trading strategy that buys an asset when its price is trending up, and (short) sells an asset when its price is trending down. Typically horizons of up to 12 months are used to measure the trend. The rationale for TSM stems from the observed small positive autocorrelation in asset returns, generating some degree of short-term return continuation (Lo and MacKinlay, 1990). This contrasts with cross-sectional momentum, which is a strategy that buys a portfolio of assets that have appreciated the most over some horizon (the 'winners'), and shorts a portfolio of assets that depreciated the most (the 'losers') over the same horizon. Hence the main difference between time series and cross-sectional momentum is time series is purely univariate, whilst cross-sectional considers relative performance across multiple assets.

Lo and MacKinlay (1990) find that cross-sectional momentum may be driven by autocorrelation in stock returns, lead-lag relations among stocks, or cross-sectional dispersion in unconditional means, explicitly linking cross-sectional momentum to time series momentum. Moreover, Goyal and Jegadeesh (2018) empirically examine the differences between time series and cross-sectional momentum strategies, and note that TSM strategies are on average net long, whilst cross-sectional strategies are zero net investment. After adjusting for the net long position, time series and cross-sectional momentum strategies generally earn similar excess returns.

The seminal paper on TSM is Moskowitz, Ooi, and Pedersen (2012), who document significant return continuation in 58 liquid futures written on equity indices, currencies, commodities, and bonds. Returns on these contracts persist for one to twelve months, before partially reversing over longer horizons. They build TSM strategies for each instrument individually, and show that each generates a positive Sharpe ratio before fees, and a diversified portfolio of all TSM strategies generates a significant four-factor alpha of 1.58 percent per month. Hurst, Ooi, and Pedersen (2017) extend the sample to 1880 and study the performance of TSM strategies across asset classes. Using this longer sample, they find that TSM has delivered positive average returns in each decade, with

low correlations with traditional asset classes. Furthermore, they highlight the robustness of TSM, showing it has performed well in eight of the 10 largest crisis periods, as well as in low and high interest rate and inflation environments.

Menkhoff, Sarno, Schmeling, and Schrimpf (2012) examine TSM in currencies, and find that whilst most of the strategies are individually profitable, an equally-weighted portfolio of the strategies underperforms a cross-sectional strategy on the same assets with a Sharpe ratio half the size. In commodity futures markets, Dewally, Ederington, and Fernando (2013) show that TSM profits are large and positive, and are primarily driven by hedging pressure.

In terms of the theoretical literature, He and Li (2015) propose a continuous-time heterogeneous agent model of fundamental trading and TSM. They show that when the number of momentum traders with short investing horizons is high, momentum trading has a stabilising effect on the market, and TSM is profitable because the rest of the market underreacts in the short-term. Andrei and Cujean (2017) offer an alternative theory of TSM and long-term reversal based on informational percolation. They show that, in a rational expectations framework, word-of-mouth communication flows across investors and generates short-term price run-ups, causing TSM which reverses at longer horizons.

There have been some criticisms of the aforementioned papers. Kim, Tse, and Wald (2016) argue that most of the performance of the TSM strategies in Moskowitz, Ooi, and Pedersen (2012) is driven by volatility scaling, and the unscaled strategies have similar alphas to the buy-and-hold assets. Huang, Li, Wang, and Zhou (2020) show the t-statistics from pooled predictive regressions of returns on lagged returns are not statistically reliable due to assets having different Sharpe ratios, and the performance of TSM is similar to a strategy based on the historical mean with no predictability.

3.3 Data

The sample consists of some of the most widely studied factors in empirical asset pricing: the Fama and French (1992, 1993) market (RMRF), size (SMB), and value (HML) factors, the Fama and French (2015) profitability (RMW) and investment (CMA) factors, the winners-minus-losers cross-sectional momentum strategy (WML), and the Frazzini and Pedersen (2014) betting-against-beta (BAB) factor. The baseline results focus on the US, and results for international factors are shown in additional robustness tests. All factors are obtained up to the end of 2019 from Ken French's data library, except the BAB factor which is obtained from AQR.[3,4] The advantages of focusing on these factors are their properties have been well studied, and the results in this chapter can easily be replicated and compared to the existing literature.

The factor returns are summarised in Table 3.1. On average all factors earn a positive premium of between 0.99 and 8.27 percent per annum, with volatilities ranging from 6.81 to 16.96 percent per annum. The subpar performance of the SMB factor since its discovery by Banz (1981) has been well documented in the literature (e.g. Alquist, Isreal, and Moskowitz, 2018; Blitz and Hanauer, 2020). With the exception of SMB, the factors earn high risk-adjusted returns with unconditional Sharpe ratios in the range of 0.37 to 0.78. Finally the first-order autocorrelations all exceed 0.06, reaching up to 0.19 for RMW, with the exception of WML that has a negligibly negative autocorrelation.

[3]Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[4]AQR: https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Daily

3.4 Factor Momentum

In this section I review the literature on TSFM, and review its IS and OOS evidence by replicating the main results from the literature for my set of factors.

3.4.1 Previous Literature

The idea behind TSFM is analogous to Moskowitz, Ooi, and Pedersen (2012) TSM applied to long-short equity factors: if the recent past performance of a factor was positive, a TSFM strategy would go long the factor, and short if the cumulative past return was negative. Equity factors represent an interesting case compared to traditional assets because they are long-short portfolios of individual stocks, and therefore need not exhibit the same return persistence characteristics as equity indices or other asset classes. Indeed, as noted by Gupta and Kelly (2019) TSFM decays faster than traditional assets, typically lasting only one month instead of up to 12.

Two similar papers are Flögel, Schlag, and Zunft (2019) and Gupta and Kelly (2019), who both document material first-order autocorrelations in equity factor returns. Flögel, Schlag, and Zunft (2019) focus on a small set of US factors, whilst Gupta and Kelly (2019) examine a wider set of 65 US and international factors. They both construct variants of TSFM strategies and show these strategies consistently generate higher Sharpe ratios than the underlying factors. A portfolio of TSFM strategies generates a Sharpe ratio of up to 0.84, representing a material improvement over a portfolio of the underlying factors.

Arnott, Clements, Kalesnik, and Linnainmaa (2019) link cross-sectional factor momentum to industry momentum: factor momentum is transmitted into the cross-section of industry returns via variation in industries' factor loadings. Ehsani and Linnainmaa (2020) link TSFM to momentum in individual stock returns. Cross-sectional stock momentum indirectly times factors, profiting when factor autocorrelations are high, and crashing when the autocorrelations break down. Finally, Falck, Rej, and Thesmar (2020) also link factor to individual stock momentum, and show that TSFM is distinct but only at short lags. They rationalise these findings in a simple theoretical model that predicts a correlation between factor and stock momentum that is independent of the lag and holding period used.

3.4.2 Evidence on Factor Momentum

3.4.2.1 In-Sample

Individual OLS. I begin by examining time series predictability of individual factors using their own lagged returns at various horizons with IS regressions. Specifically, for each factor f and lag $k \in \{1, \ldots, 24\}$ months I estimate the univariate time series regression:

$$\frac{r_t^f}{\sigma_{t-1}^f} = \alpha + \beta_k \frac{r_{t-k}^f}{\sigma_{t-k-1}^f} + \varepsilon_t^f, \tag{3.1}$$

where r_t^f is the return on factor f in month t, and σ_t^f is the volatility of factor f in month t, calculated as the standard deviation of daily returns within that month. I also consider alternative volatility models, but consistent with Moskowitz, Ooi, and Pedersen (2012) I find that they make little difference both qualitatively and quantitatively.[5]

A simpler way to examine return persistence is to use the *sign* of lagged returns as opposed to the level. In a regression context this can be captured by the specification:

$$\frac{r_t^f}{\sigma_{t-1}^f} = \alpha + \beta_k \text{sign}\left(r_{t-k}^f\right) + \varepsilon_t^f, \tag{3.2}$$

where $\text{sign}(x) = 1$ if $x > 0$, -1 if $x < 0$, and zero if $x = 0$. Since volatility is strictly positive it is not necessary to scale returns on the right-hand side of equation (3.2). The left-hand side returns are always scaled by volatility to make parameter estimates comparable across factors.

Pooled OLS. I also report results when pooling all factors together in the same regression. Because the factors have different means and volatilities, I control for factor fixed effects to capture differences in the intercept for each factor. The pooled OLS with fixed effects specification is:

$$\frac{r_{it}^f}{\sigma_{it-1}^f} = \alpha + \frac{\mu_i}{\sigma_i} + \beta_k \frac{r_{it-k}^f}{\sigma_{it-k-1}^f} + \varepsilon_{it}^f, \tag{3.3}$$

[5]The alternative models include historical volatility calculated over 2, 3, 6, and 12 months periods using daily returns, and expected volatility using a GARCH(1,1) forecasting model. The choice of model makes little difference due to the highly persistent nature of volatility.

or

$$\frac{r^f_{it}}{\sigma^f_{it-1}} = \alpha + \frac{\mu_i}{\sigma_i} + \beta_k \text{sign}\left(r^f_{it-k}\right) + \varepsilon^f_{it} \tag{3.4}$$

for the sign specification. The fixed effects are captured by the factor-specific intercepts μ_i/σ_i, which represent the ratio of the unconditional mean of factor i to its unconditional volatility. A potential issue with this specification is the fixed effects demean the variables using the mean over the full sample, whereas in a predictive regression we should not use forward looking data that was not available to investors at the time. This induces a bias in the standard errors of the coefficients due to a non-zero correlation between the lagged value of the demeaned regressors and the error (Hjalmarsson, 2010; Huang, Li, Wang, and Zhou, 2020).

To solve this issue, I follow Huang, Li, Wang, and Zhou (2020) and conduct a pairs bootstrap exercise. Specifically, I sample T pairs (where T is the length of the sample) of the dependent and independent variables in the regression with replacement 1,000 times, and run a pooled regression for each of the 1,000 samples. This generates a distribution of simulated t-statistics for $\hat{\beta}_k$, of which the 97.5th percentile can be computed to be used as a critical value for the actual regression (assuming a two-tailed test). The estimated coefficient is significant if the absolute value of the t-statistic in the actual pooled regression exceeds the 97.5th percentile bootstrapped t-statistic.

Figures 3.1 and 3.2 plot the t-statistics from these regressions. Figure 3.1 corresponds to the level regressions in (3.1) and (3.3), whilst Figure 3.2 corresponds to the sign regressions in (3.2) and (3.4). The individual factor t-statistics are calculated using Newey and West (1987) HAC robust standard errors with automatic lag selection. Both regressions paint a similar picture: predictability is generally strongest at a one month lag, and decays quickly for most factors. For the market and value factors there is some evidence of longer-term reversals evidenced by the significant negative coefficients at 13–14 month lags, but it is not as strong or consistent as the short-term momentum effect. Interestingly, the size factor exhibits very strong predictability at the 12 month lag, with a coefficient larger than at 1 month. It is not clear why this is the case. The BAB factor exhibits consistent predictability up to 12 months, and is the only factor with this property. Evidence of predictability for the WML and CMA factors is weak at all lags.

Regarding the pooled OLS regressions, Table 3.2 reports the coefficient estimates for lags of 1–12 months, along with the t-statistics and bootstrapped critical values. The standard errors in these regressions are clustered by time to allow for correlated errors across factors that are independent over time. The coefficients are all positive, ranging from 0.02 to 0.16 in magnitude. The only significant coefficients at the 5% level are at lag 1, as the t-statistics of 9.15 and 8.83 exceed the bootstrapped critical values of 8.54 and 8.06 for the level and sign regressions respectively. To summarise, the IS evidence is broadly consistent with the findings in Gupta and Kelly (2019).

3.4.2.2 Out-of-Sample

Huang, Li, Wang, and Zhou (2020) have recently questioned the existence of TSM in traditional assets. They show that the OOS performance of lagged return regressions is poor, and the performance of TSM is similar to a strategy based on the historical mean with no OOS predictability. In this section I show that the OOS predictability of lagged *factor* returns remains strong, and outperforms the historical sample mean.

OLS. The OOS testing procedure for the OLS models is as follows. For each factor, the sample is split 70:30 into a training and a test set. An OLS regression of factor returns on one month lagged factor returns is estimated over the training period, generating a set of training period coefficients that define the initial predictive model. In the first OOS test ('fixed window'), the training period coefficients are fixed and are used to generate predictions for each month in the testing period. In the second OOS test ('expanding window'), each month in the testing period a new OLS model is estimated using all past data, i.e. the sample grows over an expanding window with a minimum number of observations equal to the length of the initial training period. The final outputs are a set of predicted returns for each month in the testing period for each factor. The results are robust to using alternative training-test splits, provided the training sample is sufficiently long to generate reliable forecasts.

A standard metric to evaluate a forecasting model is the OOS R^2, which compares the mean squared error (MSE) of a forecast to some benchmark model, typically an expanding window historical mean

(Campbell and Thompson, 2008; Goyal and Welch, 2008):

$$R^2_{\text{OOS}} \;=\; 1 - \frac{\sum_{t=\tau}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=\tau}^{T}(r_t - \bar{r}_t)^2}, \tag{3.5}$$

where τ is the first date of the testing period, r_t is the actual return in period t, $\hat{r}_t$ is the value predicted using a regression estimated over the training period, and $\bar{r}_t$ is the historical mean computed using all data up to but excluding period t (so no information from the testing sample is used to compute the historical mean). Naturally if incorporating lagged returns into a regression is informative it will reduce MSE relative to the historical mean, and generate a positive OOS R^2. Note that OOS R^2 can be negative, as a model can perform arbitrarily worse than the historical mean.

Table 3.3 summarises the full-sample regression slope at lag one (β_1), the full IS adjusted R^2 (R^2_{IS}), and the OOS R^2 for each factor using both evaluation methods. The OOS R^2s are positive and comparable in magnitude to the IS R^2s for all factors except SMB, and are close to zero for RMRF and WML. The rest of the factors exhibit moderately positive OOS R^2s ranging from 1–4% per month, implying lagged returns do contain predictive information beyond the historical mean. These equity factor results sharply contrast Huang, Li, Wang, and Zhou (2020), who find consistently negative OOS R^2s for a large cross-section of long-only assets. Moreover, as noted by Campbell and Thompson (2008) and Gu, Kelly, and Xiu (2020), even a small improvement in predictive R^2 can lead to large utility gains for a mean-variance investor, as the Sharpe ratio improvement can be directly quantified as:

$$SR^* \;=\; \sqrt{\frac{SR^2 + R^2}{1 - R^2}}. \tag{3.6}$$

Hence given an average monthly (annualised) Sharpe ratio of 0.15 (0.52), a monthly R^2 of 0.03 increases this Sharpe ratio to 0.23 (0.81), which represents a material improvement.

LASSO selection. An alternative way to examine OOS return predictability is to use a variable selection model like the LASSO to tell us which lags are the most important for predicting returns OOS.[6] The LASSO is a form of penalised regression that shrinks the coefficients of uninformative

[6]For recent applications of the LASSO in finance see Chinco, Clark-Joseph, and Ye (2019) and Freyberger, Neuhierl, and Weber (2020).

regressors to zero in the hope of generating a more parsimonious model with better OOS performance. The strength of penalisation is dictated by a hyperparameter λ; a higher value of λ corresponds to greater penalisation, and more of the coefficients will be shrunk to zero. There exists an optimal value of λ which must be found through a process called hyperparameter tuning.[7]

The optimal degree of penalisation is tuned through time series validation. As with the OOS OLS regressions, the sample is split 70:30 into two periods: training and validation. I then estimate a LASSO regression for a fixed degree of penalisation for each factor over the training period, and generate a forecast for the first month in the validation period. I then extend the training period by one month, re-estimate the LASSO for each factor, and generate another one-month-ahead forecast. This process is iterated until forecasts have been generated for each month in the validation period. It is then iterated again for different levels of penalisation from a fine grid of possible values, and an MSE metric is computed for each penalisation level. The optimal degree of penalisation is the one that generates the lowest MSE, and is used in the final LASSO model.

As detailed in Table 3.4, most of the LASSO coefficients are zero, implying returns at those lags are not useful for predicting returns OOS. Notably for all factors except SMB and WML (which do not exhibit the strongest lag one IS predictability) the coefficient at lag one is non-zero and positive. There is also sporadic evidence that returns at higher lags are useful for OOS prediction for different factors, often at lag 12 but with smaller magnitudes. Finally, consistent with the IS evidence, most of the coefficients for the BAB factor are non-zero. Overall, there is significant evidence of OOS time series predictability for factors using lagged returns.

3.4.3 Time Series Factor Momentum Strategies

In this section I construct simple trading strategies that exploit the time series predictability documented in the previous sections. I consider two variants of the simple TSFM strategy: in the first the signal is proportional to the sign of the previous one month return as in Moskowitz, Ooi, and Pedersen (2012), whilst in the second is it proportional to the level (after scaling by volatility)

[7] Optimal in the sense of minimising OOS forecast errors.

as in Gupta and Kelly (2019). The general notation for the TSFM strategies is:

$$r_{t+1}^{\text{TSFM}} = s_t \times r_{t+1}, \tag{3.7}$$

where the signal s_t is constructed as either:

$$s_t = \text{sign}(r_t), \tag{3.8}$$

$$s_t = \min\left(\max\left(\frac{r_t}{\sigma_{t-1}}, -1\right), 2\right). \tag{3.9}$$

Note that the second signal is capped between ± 2 to limit extreme weights, whilst the first signal is by definition either -1, 0, or 1. The basic intuition underlying these strategies is to go long when past one month returns are positive, short when they are negative, and take more extreme positions when past returns are more extreme (relative to volatility).

The Sharpe ratios of these strategies, along with the raw factors for comparison, are reported in Table 3.5. All of the Sharpe ratios are positive, but it is not the case that the TSFM strategies always improve performance. The biggest improvements are for the SMB and BAB factors, whose Sharpe ratios increase by about 2.6–4 and 1.2–1.4 times respectively. RMRF and HML have no material change, and RMA, CMA, and WML actually generate lower Sharpe ratios. There is no clear winner when choosing between using the sign or level of past returns as the signal.

To examine the effect of trading costs, I also calculate the turnover of the TSFM strategies. Turnover is defined as the average absolute monthly change in the signal:

$$\text{Turnover} = \frac{1}{T}\sum_{t=2}^{T} \mid s_t - s_{t-1} \mid . \tag{3.10}$$

All TSFM strategies exhibit similarly high turnovers in the range of 0.8–1.0 times per month. Unsurprisingly the level-based strategies have higher turnover, because the weights have greater variability (being capped between ± 2 rather than ± 1). Momentum strategies are high turnover by nature, so the high average turnover of these strategies is not surprising, and compare similarly to previous literature (e.g. Gupta and Kelly, 2019).

3.5 Cross-Factor Momentum

In the following sections I examine whether there exists a lead-lag relationship in equity factor returns. The motivation for this exercises stems in part from Pitkäjärvi, Suominen, and Vaittinen (2020), who document two-way predictability of equity and bond returns in 20 countries: lagged equity returns predict future bond returns, and lagged bond returns predict future equity returns. They construct cross-asset TSM strategies that exploit this predictability, and generate Sharpe ratios that are 45% higher than standard TSM strategies. Similarly, Jia, Liu, Wu, and Yan (2020) employ the OLS-post-LASSO to examine cross-asset predictability in the foreign exchange market. They run predictive regressions of currency excess returns on the lagged forward discounts of all currencies, and document strong OOS predictability driven by information spillovers across currencies. Finally, Chinco, Clark-Joseph, and Ye (2019) examine high frequency return predictability using the LASSO and the entire cross-section of stocks as candidate predictors. They find the LASSO is able to identify predictors as those with news about fundamentals, which spills over into other assets. The exercise in this section is similar to these papers in spirit, but applied to equity factors.

3.5.1 Evidence of Cross-Factor Correlations

3.5.1.1 In-Sample

In this section I document novel IS evidence on cross-factor predictability. I begin by examining the cross-correlations of pairs of factors. Cross-correlations are widely used in signal processing, and measure the correlation of a series with lagged values of another series, i.e. $\text{corr}(x_t, y_{t-k})$ for some lag k, and represent a useful preliminary exercise for investigating cross-factor predictability. The cross-correlations are estimated over the full sample, and can be interpreted similarly to standard pairwise correlations but in a predictive context.

Table 3.6 presents the contemporaneous and cross-correlation matrices for the seven factors, where the lagged series is represented in the columns, e.g. the second element of the the first row in Panel B (0.03) represents the correlation between RMRF returns and one period lagged SMB returns. Bold values are significant at the 5% level given the sample size of the data and the magnitude of

the correlation:

$$t_r \;=\; r\sqrt{\frac{n-2}{1-r^2}} \;\geq\; 1.96. \tag{3.11}$$

There is strong evidence of non-zero cross-correlations in monthly US factor returns. In particular, market returns are significantly correlated with lagged market and investment returns, with (absolute) correlations in the range of 0.07–0.09. In some cases the cross-factor correlations are stronger than the factor's autocorrelations themselves, implying it might be possible to enhance TSM strategies using the returns of other factors in addition to a factor's own historical returns. For example, the autocorrelation of SMB is around 0.06, but its cross-correlations with RMRF and CMA are 2–3 times larger at 0.20 and −0.14 respectively. Notably BAB has significant cross-correlations with 5 of the 7 factors, all with magnitudes exceeding 0.10. The diagonal elements of Panel B represent the first-order autocorrelation, and are significant for all factors except WML.

Next I examine cross-factor predictability in a regression context to jointly account for correlations in lagged factor returns. For each factor i, I regress its volatility-scaled return on the one month lagged (scaled) returns of all factors, in both levels and signs:

$$\frac{r_{it}}{\sigma_{it-1}} \;=\; \alpha + \sum_{k=1}^{n} \beta_j \frac{r_{jt-1}}{\sigma_{jt-2}} + \varepsilon_{it-1}, \tag{3.12}$$

$$\frac{r_{it}}{\sigma_{it-1}} \;=\; \alpha + \sum_{j=1}^{n} \beta_j \text{sign}\left(r_{jt-1}\right) + \varepsilon_{it-1}. \tag{3.13}$$

These regressions are analogous to equations (3.1) and (3.2) but include additional regressors, and only use one month lagged returns as the independent variables. Equation (3.12) is equivalent to a first-order vector autoregression estimated on monthly factor returns, and can be interpreted as such.

The Newey and West (1987) t-statistics for the estimated coefficients from these regressions are plotted for each factor in Figure 3.3. Consistent with the cross-correlation matrix, there is strong evidence of cross-factor predictability for most of the factors considered. RMRF loads on itself, SMB, and HML although the significance is borderline. SMB is strongly predicted by the lagged market and itself. HML is predicted by lagged value and the signs of profitability and BAB. WML

mostly only loads on lagged WML returns and somewhat RMRF. RMW is predicted by lagged RMW and the sign of HML. CMA loads on the signs of RMW and CMA. Lastly BAB loads heavily on all factors except RMW, CMA, and WML.

3.5.1.2 Out-of-Sample

A stricter test of predictability focuses on OOS performance. As in Section 3.4.2.2 I run LASSO regressions to examine which factors have non-zero coefficients. The hyperparameter tuning procedure is identical to Section 3.4.2.2, and the regressions only differ in the independent variables used: in this section one-month lagged returns on all factors are used instead of multiple lags of the factor's own return.

Table 3.7 presents the coefficients from the OOS LASSO regressions. The diagonal elements correspond to OOS standard momentum coefficients, and the off-diagonal elements correspond to OOS cross-factor predictability coefficients. With the exception of HML, all factors have at least one non-zero loading on other factors, implying there exists OOS cross-factor predictability. RMW and BAB especially, as they load on 5 of 7 and all of the factors respectively, with low degrees of penalisation. Overall, these results suggest cross-factor predictability exists OOS as well as IS.

3.5.2 Cross-Factor Momentum Strategies

Building on the findings in the previous sections, I construct CFM strategies by augmenting the signals used in the standard TSFM strategies with additional components related to the other factor returns. Construction of the strategies is similar to Section 3.4.3, differing only in the construction of the signal:

$$r_{t+1}^{\text{CFM}} = s_t \times r_{t+1}, \tag{3.14}$$

$$s_t = \sum_{j \in J} w_{jt} \times s_{jt}, \tag{3.15}$$

$$s_{jt} = \min\left(\max\left(\frac{r_{jt}}{\sigma_{jt-1}}, -1\right), 2\right), \tag{3.16}$$

where the signal s_t is a weighted average of the s_{jt}s, which are the volatility-scaled one month

lagged returns on each factor j (capped between ± 2). I consider four schemes for w_{jt}: (i) signed $1/N$, which weights all factors equally and disregards the strength of the predictive relationship but accounts for the sign, (ii) pairwise correlation weights, which assigns greater weight to factors with stronger cross-correlations, (iii) OLS coefficient weights, which uses multiple regression coefficients and accounts for correlation between lagged predictors simultaneously unlike the correlation weights, and (iv) LASSO coefficient weights, which only includes factors that have been selected by the LASSO to generate superior OOS performance. To avoid using forward looking data when constructing the strategies, the correlations, and OLS and LASSO coefficients are re-estimated each month over an expanding window using only past data.

Table 3.8 reports detailed statistics on the performance of the CFM strategies, as well as the raw factor and vanilla TSFM strategies for comparison. Panel A presents the (annualised) Sharpe ratios. In general, the $1/N$ CFM strategies perform very poorly, rarely outperforming the raw factor. However the correlation, OLS, and LASSO weighted strategies generally perform much better, providing greater risk-return characteristics than the raw factor and the vanilla TSFM strategies. The OLS CFM strategies in particular perform well for all factors except WML. In terms of turnover (Panel B), the CFM strategies generally have similar or lower turnover than the TSFM strategies, implying there are no additional trading costs required to exploit cross-factor predictability. Finally, Panel C reports the maximum drawdown, defined as the maximum peak-to-trough loss.[8] In general, the correlation, OLS, and LASSO CFM strategies have significantly lower drawdowns than the TSFM strategies, making them more attractive from a risk management perspective. In summary, the CFM strategies have higher Sharpe ratios, similar or lower turnovers, and lower drawdowns than both the raw factors and the TSFM strategies, making them more appealing to investors seeking to earn higher risk-adjusted returns.

3.5.3 Robustness Tests and Additional Exercises

International evidence. I verify the robustness of CFM by extending the results to cover international factors from four regions: Asia Pacific ex. Japan, Europe, Global, and Japan. The factor definitions are identical to the baseline US factors so the results are comparable across regions.

[8] Formally, maximum drawdown (MDD) is calculated as: $\text{MDD} = \max_{t \in (0,T)} \left[\max_{t \in (0,\tau)} P(t) - P(\tau) \right]$.

Table 3.9 reports the Sharpe ratios for the raw factors, the level and sign TSFM strategies, and the four CFM strategies using each weighting scheme. The results are similar to the US results: CFM strategies generally outperform both the raw factors and the TSFM strategies, and the OLS and LASSO weighting schemes consistently perform the best.

Lookback period. All of the momentum strategies in the main results of this chapter use a one-month lookback period, motivated by the evidence that factor autocorrelations and cross-correlations are strongest at lag one. In unreported results I reconstruct all strategies using 3-, 6-, and 12-month cumulative lagged returns as signals, and an equally-weighted average of these signals. The results are unsurprisingly weaker using longer lookback periods, but not by much. CFM strategies continue to outperform the raw factors and standard TSFM strategies at all lookbacks.

3.6 What Drives Cross-Factor Predictability?

3.6.1 Intermediary Channel

One potential channel driving cross-factor predictability is the intermediary channel. Financial intermediaries are sophisticated investors often with access to better information than the market as a whole. Shocks to financial intermediaries may therefore have spillover effects from one factor to others: if the intermediary is long or short a particular factor and receives a negative funding shock, they may be forced to reduce exposure and/or reallocate to other less risky factors. This may generate a predictable lead-lag relationship in factor returns, explaining the existence of CFM.

To test this hypothesis, I regress the returns of the CFM strategies on the He, Kelly, and Manela (2017) tradable intermediary asset pricing factor (IAPF). This factor is constructed from the returns to a value-weighted portfolio of NY Fed primary dealers' publicly traded holding companies. As shown by He, Kelly, and Manela (2017), this factor performs similar to a factor based on shocks to intermediary's capital ratios, but has the advantage of being tradable. If shocks to intermediary leverage are able to explain CFM, the alpha (i.e. the intercept) in this regression would be small and insignificantly different from zero. I also control for the seven factors used throughout this chapter, to account for the possibility that the CFM strategies load on the raw factors themselves.

The results in Table 3.10 show that the alphas generally remain significant after controlling for the IAPF for all four weighting schemes. Moreover CFM is not spanned by the Fama-French five-factor model augmented with WML and BAB because the alphas in the first four columns are mostly large and significant. Including the IAPF generally has no material effect on the magnitude of alphas either, suggesting financial intermediaries are not able to account for the existence of CFM.

3.6.2 Fund Flow Channel

Another potential channel driving cross-factor predictability is the fund flow channel. If investors dynamically change allocations to different factors over time in response to changes in the economic environment, there may be a natural spillover effect where flows in and out of one factor fund affects returns on other factor funds. Pitkäjärvi, Suominen, and Vaittinen (2020) find this to be the dominant mechanism driving cross-asset predictability in equity and bond markets. To test this

hypothesis, I gather data on factor-based ETF flows and run a return-flow vector autoregression (VAR) to examine whether there exists two-way predictability for flows and returns.

3.6.2.1 Fund Flows

The first step is to approximate aggregate flows into a factor. This is a difficult exercise because most factors are implemented differently across funds, such as using different measures of value (book-to-market, price-to-earnings, etc.) or profitability (profit margins, return on assets, etc.). Moreover, most funds are not truly pure-play factor funds (funds with a unit beta on one factor and zero on all others), as they will often have non-zero exposures to other factors, whether intentional or not. Therefore, to conduct this exercise a number of simplifying assumptions must be made.

To begin, I search the Morningstar fund database for all long-short factor ETFs, and manually classify them by their main factor exposure. For example, if a fund primarily invests in small cap stocks and shorts large cap stocks, it would be classified as an SMB fund. If the fund name is ambiguous or nondescript it is not included in the analysis. Monthly flows are then backed out from returns and assets-under-management using the relation:

$$Flow_t \;=\; Assets_t - Assets_{t-1}(1 + R_t). \tag{3.17}$$

Flows are aggregated for all funds within each factor group to construct an approximate factor flow index for each factor. Note that I only focus on the three most popular factors in the finance industry—size, value, and cross-sectional momentum—because very few funds are pure investment or profitability funds (such funds typically embed these variables as part of a 'quality' or multi-factor strategy), and there are very few pure BAB ETFs. Despite this limitation, the VAR is still instructive as a preliminary exercise to further understand the drivers of CFM.

It is important to note that fund flows are highly persistent. Warther (1995) shows that the first-order autocorrelation of equity fund flows exceeds 0.55, and decays slowly to 0.30 at lag 6. Therefore to remove the predictable component of flows, I follow the literature and fit an AR(2) model to each aggregate flow series, and use the residuals as a measure of 'unexpected' flows.[9] This means the flow

[9] Warther (1995) finds no material difference between using an AR(2) or AR(3) model, but an AR(1) model is not sufficient to capture the full autocorrelation structure of fund flows.

series use forward looking information, but for the purpose of this exercise it is sufficient because it is not being used for forecasting, but rather to try and explain the relationship.

3.6.2.2 Return-Flow VAR

Given return and flow series for the size, value, and cross-sectional momentum factors, I fit a VAR(2) model of the form:

$$\mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t, \tag{3.18}$$

which can be expressed compactly in VAR(1) notation:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}_0 \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 \\ \mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ 0 \end{pmatrix}, \tag{3.19}$$

where

$$\mathbf{y}_t = \begin{pmatrix} r_t^{\text{SMB}} & r_t^{\text{HML}} & r_t^{\text{WML}} & f_t^{\text{SMB}} & f_t^{\text{HML}} & f_t^{\text{WML}} \end{pmatrix}' \tag{3.20}$$

is a column vector stacking the return and flow variables, $\boldsymbol{\phi}_0$, $\boldsymbol{\phi}_1$, and $\boldsymbol{\phi}_2$ are vectors or matrices of VAR parameters of appropriate dimension, $\mathbf{I}$ is the identity matrix, and $\boldsymbol{\varepsilon}_t$ is a vector of i.i.d. shocks. I use two lags to account for any potential information delay in the publication of flow data.

Table 3.11 presents the sum of the first and second lag coefficients, as well as p-values for a test of the joint hypothesis that both coefficients are zero. Including flows in the VAR leads to some of the return-return coefficients across factors to become insignificant, supporting the idea that CFM can be at least partially explained by flows across factors. The own-factor return-return coefficients for SMB and HML are still significant at -0.16 and 0.18 respectively, implying fund flows cannot explain standard TSFM. Figure 3.4 plots the impulse response functions from this VAR for each variable, along with a 95% confidence interval using asymptotic standard errors. The subfigures show that shocks die out very quickly for most variables, suggesting any predictive relationship is very short-lived. Overall the evidence that CFM is driven by fund flows is mixed and weak.

3.7 Conclusion

In this chapter I re-examine TSFM strategies and document new results on CFM. I begin by reviewing the IS and OOS evidence on TSM in US equity factors, and find evidence of short-term predictability in factor returns. Predictability is strongest at a one month lag and quickly decays to zero at higher lags. I construct TSFM strategies following Gupta and Kelly (2019) and find mixed evidence these strategies outperform the raw factors, suggesting time series predictability may be difficult to capture for investors in real time.

I then find evidence of material cross-correlations in factor returns. Most of the factors considered correlate highly with at least one or two other lagged factors, with magnitudes often exceeding the factor's own first-order autocorrelations. Motivated by these findings, I construct CFM strategies that use the lagged returns on multiple factors to dictate directional trading positions, as opposed to just the factor's own short-term past return. These strategies improve Sharpe ratios by up to 1.7 times compared to standard factor momentum strategies, with lower drawdowns and similar trading costs. In robustness tests I show CFM remains profitable in international data, and holds at longer lookback periods.

Finally I examine different hypotheses for why cross-factor predictability exists, and find inconclusive evidence. Two channels based on financial intermediaries and ETF flows are not able to fully account for the profitability of CFM strategies. The results in this chapter therefore provide another hurdle for theories of factor-based asset pricing to overcome, and offer several interesting directions for future work.

Table 3.1: Summary statistics for monthly factor returns. This table reports summary statistics for monthly factor returns expressed in percent, where means, standard deviations, downside deviations, and Sharpe ratios are annualised. Start is the YYYY-MM date of the factor's first return. AC(1) is the first-order autocorrelation coefficient.

	RMRF	SMB	HML	RMW	CMA	WML	BAB
Start	1926-07	1926-07	1926-07	1963-07	1963-07	1926-11	1930-12
Mean	7.49	0.99	4.00	3.41	3.37	7.78	8.27
Std dev.	16.96	9.68	10.78	6.81	6.73	14.24	10.59
Downside dev.	12.74	6.04	6.53	4.17	3.62	11.28	8.82
Sharpe	0.44	0.10	0.37	0.50	0.50	0.55	0.78
Sortino	0.59	0.16	0.61	0.82	0.93	0.69	0.94
Max. drawdown	−82.44	−59.08	−40.88	−30.84	−19.24	−55.95	−54.76
Min	−18.73	−8.30	−9.53	−6.16	−5.25	−17.50	−12.73
Median	1.04	0.04	0.14	0.21	0.10	0.68	0.75
Max	16.40	9.47	13.96	9.08	6.86	15.67	11.93
Skew	−0.52	0.17	0.63	0.55	0.42	−0.41	−0.51
Kurtosis	1.95	0.98	3.14	3.29	0.98	3.72	4.01
AC(1)	0.07	0.06	0.14	0.19	0.13	−0.00	0.14

Table 3.2: *t*-statistics for pooled predictive regressions with factor fixed effects. This table reports estimated coefficients, t-statistics, and 95% bootstrapped t-statistics for pooled predictive regressions of factor returns on lagged factor returns. The bootsrapped t-statistics are calculated using a pairs bootstrap following Huang, Li, Wang, and Zhou (2020). All t-statistics are calculated using standard errors clustered by month.

	Level			**Sign**		
Lag	Coef	t(Coef)	Bootstrapped t	Coef	t(Coef)	Bootstrapped t
1	0.16	9.15	8.54	0.06	8.83	8.06
2	0.06	3.47	3.93	0.02	3.16	3.69
3	0.04	2.12	2.84	0.01	1.92	2.72
4	0.02	1.45	2.31	0.02	2.79	3.48
5	0.04	2.39	3.06	0.02	2.81	3.38
6	0.05	2.62	3.37	0.02	2.40	3.18
7	0.03	1.91	2.74	0.01	1.71	2.63
8	0.02	1.10	2.12	0.01	0.72	1.82
9	0.03	1.53	2.56	0.01	1.80	2.62
10	0.03	2.03	2.83	0.01	1.12	2.07
11	0.05	2.61	3.22	0.02	2.85	3.41
12	0.09	4.92	5.17	0.03	4.17	4.47

Table 3.3: Out-of-sample performance for time series factor momentum regressions. This table reports full-sample regression slopes (β_1), full-sample in-sample R^2s (R^2_{IS}), and out-of-sample R^2s calculated over both a fixed and expanding window for predictive regressions of factor returns using one-month lagged returns.

	β_1	R^2_{IS}	R^2_{OOS} (fixed)	R^2_{OOS} (expanding)
RMRF	0.07	0.00	0.00	0.00
SMB	0.06	0.00	−0.02	−0.02
HML	0.14	0.02	0.02	0.02
RMW	0.19	0.04	0.04	0.04
CMA	0.13	0.02	0.03	0.01
WML	0.00	0.00	−0.00	0.00
BAB	0.15	0.02	0.02	0.02

Table 3.4: Lagged return LASSO predictive regression coefficients. This table reports coefficient estimates from LASSO regressions of factor returns on returns lagged from 1 to 12 months. '–' means the coefficient is zero. λ is the degree of penalisation (multiplied by 10^3), calculated using a walk-forward validation approach. IS R^2 is the in-sample R^2 from a regression over the training period.

Lag	RMRF	SMB	HML	RMW	CMA	WML	BAB
1	0.01	–	0.09	0.15	0.03	–	0.12
2	–	–	–	–	–	–	0.01
3	–	–	–	–	–	–	0.09
4	–	–	–	–	–	–	0.06
5	0.03	–	–	–	–	–	0.01
6	–	–	–	–	–	0.05	0.02
7	–	–	0.01	0.05	–	–	0.03
8	–	–	–	–	–	–	0.08
9	–	–	0.01	–	–	–	–
10	–	–	–	0.01	–	–	0.04
11	–	–	–	0.05	–	–	0.05
12	–	0.04	0.02	–	–	0.07	0.02
λ	0.18	0.07	0.07	0.02	0.05	0.08	0.01
IS R^2	−0.01	0.01	0.00	0.05	−0.02	−0.02	0.03

Table 3.5: Performance of time series factor momentum strategies. This table reports Sharpe ratios and turnovers for three strategies: the raw factor, and two time series factor momentum (TSFM) strategies based on the sign and level of past one-month returns. Sharpe ratios are annualised. Turnover is calculated as the average absolute monthly change in the signal.

		TSFM (sign)		TSFM (level)	
	Sharpe (raw)	Sharpe	Turnover	Sharpe	Turnover
RMRF	0.44	0.45	0.89	0.35	1.06
SMB	0.10	0.39	0.92	0.26	1.05
HML	0.37	0.31	0.89	0.37	0.97
RMW	0.50	0.39	0.88	0.41	0.96
CMA	0.50	0.31	0.89	0.33	1.02
WML	0.55	0.32	0.96	0.31	1.01
BAB	0.78	1.06	0.82	0.92	0.93

Table 3.6: Cross-correlation matrix for monthly factor returns. This table reports full-sample first-order cross-correlations calculated as $\text{corr}(r_t^i, r_{t-k}^j)$, where i is the row and j is the column. Panel A corresponds to $k = 0$ (comtemporaneous correlations), and Panel B corresponds to $k = 1$ (cross-correlations at lag one). Bold means the value is significant at the 5% level.

	RMRF	SMB	HML	RMW	CMA	WML	BAB
Panel A: Contemporaneous							
RMRF	1.00						
SMB	**0.32**	1.00					
HML	**0.12**	**0.06**	1.00				
RMW	**−0.25**	**−0.33**	0.02	1.00			
CMA	**−0.40**	**−0.18**	**0.69**	**−0.06**	1.00		
WML	**−0.25**	**−0.14**	**−0.24**	**0.18**	−0.01	1.00	
BAB	**−0.11**	−0.05	0.01	**0.26**	**0.30**	**0.22**	1.00
Panel B: First lag							
RMRF	**0.07**	0.03	0.00	−0.06	**−0.09**	−0.02	−0.00
SMB	**0.20**	**0.06**	0.02	−0.05	**−0.14**	−0.02	−0.01
HML	0.02	−0.02	**0.14**	−0.02	**0.12**	0.02	−0.06
RMW	−0.04	0.04	0.02	**0.19**	0.02	0.01	0.03
CMA	−0.01	−0.01	**0.12**	−0.02	**0.13**	0.00	0.05
WML	−0.04	0.05	−0.02	−0.01	−0.00	−0.00	0.03
BAB	**0.19**	**0.13**	**0.10**	0.03	−0.02	**−0.10**	**0.14**

Table 3.7: Cross-factor LASSO predictive regression coefficients. This table reports coefficient estimates from LASSO regressions of factor returns on one month lagged returns on all factors. '–' means the coefficient is zero. λ is the degree of penalisation (multiplied by 10^3), calculated using a walk-forward validation approach. IS R^2 is the in-sample R^2 from a regression over the training period.

	Dependent variable						
	RMRF	SMB	HML	RMW	CMA	WML	BAB
RMRF	0.03	0.05	–	–	–	–	0.17
SMB	–	–	–	0.09	–	–	0.18
HML	0.01	–	0.09	0.01	0.04	–	0.19
RMW	–	–	–	0.24	–	–	0.21
CMA	–	–	–	0.05	–	–	−0.05
WML	–	–	–	–	–	–	−0.04
BAB	–	–	–	−0.01	–	–	0.14
λ	0.07	0.19	0.06	0.01	0.04	0.08	0.00
IS R^2	−0.01	0.03	0.01	0.04	−0.01	0.00	0.08

Table 3.8: Performance of cross-factor momentum strategies. This table reports Sharpe ratios, turnovers, and maximum drawdowns for seven strategies: the raw factor, two time series factor momentum (TSFM) strateges based on the sign and level of past one month returns, and four cross-factor momentum (CFM) strategies using different weighting schemes to construct the signal. $1/N$ is an equally-weighted average, and Corr, OLS, and LASSO weight the signal by lag one cross-correlations, OLS coefficients, and LASSO coefficients respectively. Sharpe ratios are annualised. Turnover is calculated as the average absolute monthly change in the signal. Maximum drawdown is the maximum peak-to-trough loss. Bold values are more favourable than the raw factor (higher Sharpe ratio or lower drawdown). † means the CFM strategy is more favourable than the best performing TSFM strategy.

		TSFM		CFM			
	Raw factor	Sign	Level	$1/N$	Corr	OLS	LASSO
Panel A: Sharpe ratio							
RMRF	0.44	**0.45**	0.35	0.00	**0.47**†	**0.59**†	**0.47**†
SMB	0.10	**0.39**	**0.26**	**0.53**†	**0.63**†	**0.85**†	**0.79**†
HML	0.37	0.31	0.37	0.23	**0.45**†	**0.65**†	**0.44**†
RMW	0.50	0.39	0.41	0.18	**0.53**†	**0.63**†	0.48†
CMA	0.50	0.31	0.33	0.36†	**0.51**†	**0.78**†	**0.67**†
WML	0.55	0.32	0.31	0.06	0.17	0.41†	0.18
BAB	0.78	**1.06**	**0.92**	0.02	0.65	**0.93**	**0.95**
Panel B: Turnover							
RMRF	–	0.89	1.06	0.99	0.95	1.03	0.74†
SMB	–	0.92	1.05	0.96	0.93	0.93	0.74†
HML	–	0.89	0.97	0.95	0.94	0.95	0.68†
RMW	–	0.88	0.96	1.01	1.10	0.70†	0.56†
CMA	–	0.89	1.02	0.88	0.90	1.05	0.82†
WML	–	0.96	1.01	1.06	0.92†	1.02	0.88†
BAB	–	0.82	0.93	0.88	0.92	0.74†	0.74†
Panel C: Maximum drawdown							
RMRF	−82.44	**−78.05**	**−67.52**	−85.80	**−77.89**	**−46.00**†	−83.29
SMB	−59.08	**−32.68**	−59.10	**−42.88**	**−44.86**	**−21.45**†	**−22.28**†
HML	−40.88	**−39.57**	**−37.84**	−41.43	**−38.46**	**−40.03**	−49.39
RMW	−30.84	**−28.81**	−40.14	**−29.61**	**−17.77**†	**−13.93**†	**−19.11**†
CMA	−19.24	−23.03	**−15.30**	−22.29	−19.30	**−18.57**	**−10.32**†
WML	−55.95	**−47.72**	−57.76	−70.57	−65.15	**−42.96**†	**−51.60**
BAB	−54.76	**−25.88**	**−32.94**	−55.91	**−39.46**	**−17.30**†	**−17.33**†

Table 3.9: Sharpe ratios of international cross-factor momentum strategies. This table reports Sharpe ratios for seven strategies: the raw factor, two time series factor momentum (TSFM) strateges based on the sign and level of past one month returns, and four cross-factor momentum (CFM) strategies using different weighting schemes to construct the signal. $1/N$ is an equally-weighted average, and Corr, OLS, and LASSO weight the signal by lag one cross-correlations, OLS coefficients, and LASSO coefficients respectively. Bold values are higher Sharpe ratios than the raw factors. † means the CFM Sharpe is higher than the best performing TSFM strategy.

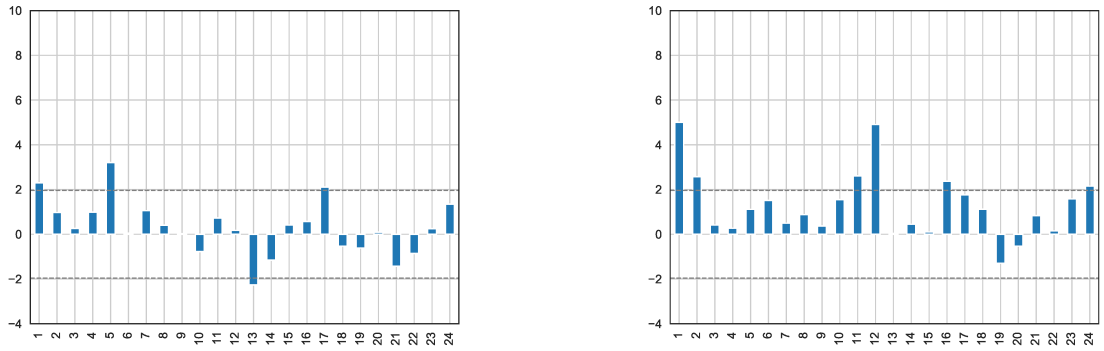
		TSFM		CFM			
	Raw factor	Sign	Level	$1/N$	Corr	OLS	LASSO
Panel A: Asia Pacific ex. Japan							
RMRF	0.44	0.34	0.29	−0.21	0.13	**0.49**†	0.08
SMB	−0.28	**0.01**	**−0.00**	−0.32	**0.78**†	**0.59**†	**1.04**†
HML	0.69	0.29	0.32	−0.02	−0.05	**0.71**†	0.56†
RMW	0.43	**0.50**	0.41	−0.30	**0.79**†	**0.75**†	**0.70**†
CMA	0.53	0.38	0.40	0.29	0.26	0.39	0.48†
WML	0.80	0.42	0.53	0.01	0.17	**1.14**†	**1.01**†
BAB	0.75	0.31	0.73	0.29	0.41	**1.17**†	**1.16**†
Panel B: Europe							
RMRF	0.39	**0.67**	**0.71**	0.06	0.09	0.52	0.55
SMB	−0.04	**0.23**	**0.18**	**0.53**†	**0.55**†	**0.94**†	**1.06**†
HML	0.43	**0.81**	**0.91**	0.07	0.58	**1.15**†	**1.16**†
RMW	0.88	0.79	0.70	−0.25	−0.27	**0.95**†	0.51
CMA	0.34	0.28	**0.34**	**0.59**†	**0.42**†	**0.88**†	**0.54**†
WML	0.75	**0.84**	**0.88**	0.43	0.38	**1.22**†	**1.06**†
BAB	0.85	**1.18**	**1.14**	0.82	0.60	**1.12**	**1.46**†
Panel C: Global							
RMRF	0.44	**0.57**	**0.62**	−0.18	−0.04	**0.49**	0.21
SMB	0.09	**0.43**	**0.41**	**0.48**†	**0.47**†	**0.96**†	**0.86**†
HML	0.43	**0.86**	**0.82**	0.39	**0.53**	**0.90**†	**0.51**
RMW	0.86	0.83	0.72	−0.20	−0.05	**0.91**†	0.41
CMA	0.39	0.21	0.07	**0.62**†	**0.51**†	**0.85**†	**0.61**†
WML	0.53	**0.73**	**0.66**	0.42	0.32	**0.77**†	**0.63**
BAB	1.02	**1.23**	**1.24**	0.83	0.63	**1.04**	**1.13**
Panel D: Japan							
RMRF	0.07	**0.30**	**0.21**	−0.13	−0.15	**0.36**†	**0.47**†
SMB	0.01	**0.43**	**0.41**	**0.01**	**0.11**	**0.41**	**0.32**
HML	0.36	0.17	0.25	**0.53**†	**0.64**†	**0.95**†	**0.39**†
RMW	0.23	**0.41**	0.06	**0.48**†	**0.36**	**0.58**†	**0.50**†
CMA	0.06	**0.28**	**0.25**	**0.38**†	**0.37**†	**1.05**†	**0.50**†
WML	0.01	**0.02**	−0.03	−0.02	−0.10	**0.38**†	**0.32**†
BAB	0.29	**0.43**	**0.47**	**0.36**	**0.38**	**0.34**	**0.49**†

Table 3.10: Alphas from intermediary factor regressions. This table reports the intercepts (in percent, annualised) from comptemparaneous time series regressions of cross-factor momentum (CFM) strategy returns on the raw factor returns with and without the He, Kelly, and Manela (2017) intermediary asset pricing factor (IAPF). * and ** denote significance at the 1% and 5% level respectively using Newey and West (1987) robust standard errors with automatic lag selection.

	Without IAPF				With IAPF			
	$1/N$	Corr	OLS	LASSO	$1/N$	Corr	OLS	LASSO
RMRF	6.53**	5.35**	12.11**	0.69	6.51**	5.33**	12.25**	0.67
	(3.49)	(3.61)	(4.23)	(1.62)	(3.52)	(3.58)	(4.33)	(1.61)
SMB	4.25**	4.75**	9.85**	8.42**	4.34**	4.85**	9.82**	8.47**
	(3.10)	(3.33)	(5.44)	(4.67)	(3.17)	(3.42)	(5.49)	(4.67)
HML	0.90	3.71**	4.85**	1.94*	0.80	3.70**	4.83**	2.00*
	(0.80)	(3.36)	(2.74)	(2.12)	(0.71)	(3.33)	(2.72)	(2.11)
RMW	0.53	2.42*	2.61*	1.21	0.57	2.45*	2.59*	1.16
	(0.52)	(2.33)	(2.18)	(1.00)	(0.58)	(2.36)	(2.15)	(0.97)
CMA	1.89	2.39*	2.53**	2.13*	1.90	2.36*	2.47*	2.10*
	(1.67)	(2.24)	(2.72)	(2.19)	(1.68)	(2.22)	(2.62)	(2.16)
WML	2.01	3.50*	5.46*	0.98	2.06	3.66*	5.78*	1.29
	(1.12)	(2.16)	(2.11)	(0.54)	(1.19)	(2.33)	(2.23)	(0.72)
BAB	−1.33	4.53**	9.12**	9.15**	−1.49	4.38**	9.03**	9.06**
	(−1.05)	(3.12)	(3.38)	(3.38)	(−1.16)	(3.00)	(3.31)	(3.31)

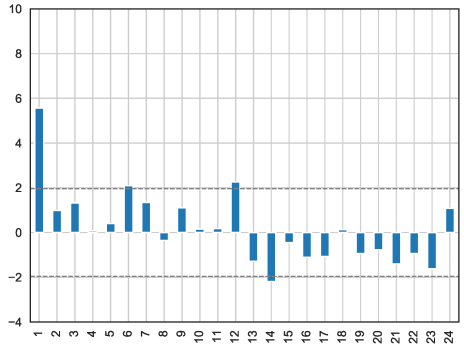
Table 3.11: Return-flow VAR coefficients. This table reports the sum of coefficient estimates from VAR(2) model fit to factor returns and unexpected flows. Unexpected flows are the residuals from an AR(2) model fit to flows. p is the p-value from a test of the joint hypothesis that both coefficients are zero. ***, **, and * denote significance at the 1%, 5%, and 10% level respectively.

	Return						Unexpected flow					
	SMB		HML		WML		SMB		HML		WML	
Dep. var.	Coef	p	Coef	p	Coef	p	Coef	p	Coef	p	Coef	p
Return												
SMB	-0.16^{**}	(0.05)	0.13	(0.30)	-0.12	(0.99)	0.00^{*}	(0.05)	0.00	(0.30)	0.00	(0.99)
HML	-0.10	(0.13)	0.18^{***}	(0.00)	0.16	(0.17)	-0.00	(0.13)	-0.01^{***}	(0.00)	-0.00	(0.17)
WML	0.01	(0.80)	0.08^{*}	(0.08)	0.15	(0.48)	-0.00	(0.80)	-0.00^{*}	(0.08)	-0.00	(0.48)
Unexpected flow												
SMB	15.66	(0.54)	-6.77^{**}	(0.04)	56.69^{*}	(0.06)	0.34	(0.54)	0.60^{**}	(0.04)	0.14^{*}	(0.06)
HML	-8.21	(0.45)	18.08^{**}	(0.04)	-79.14^{*}	(0.05)	-0.81	(0.45)	-1.55^{**}	(0.04)	-0.29^{*}	(0.05)
WML	-70.38	(0.50)	-16.78	(0.12)	-71.39	(0.22)	0.32	(0.50)	0.92	(0.12)	0.03	(0.22)

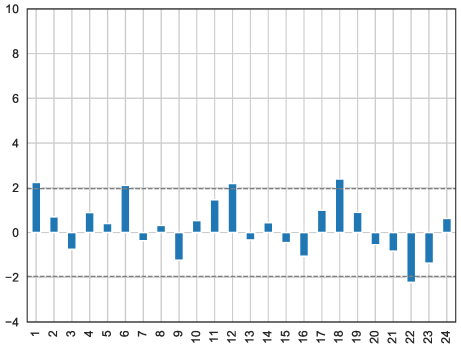


(a) RMRF

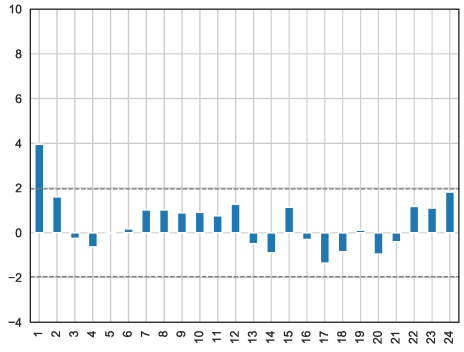
(b) SMB



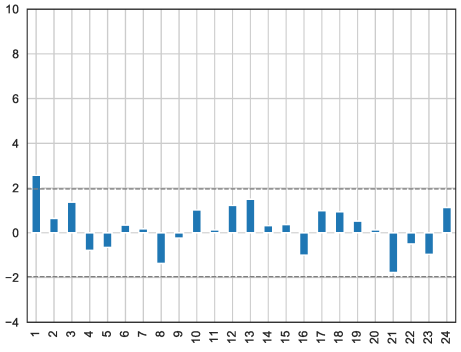
(c) HML



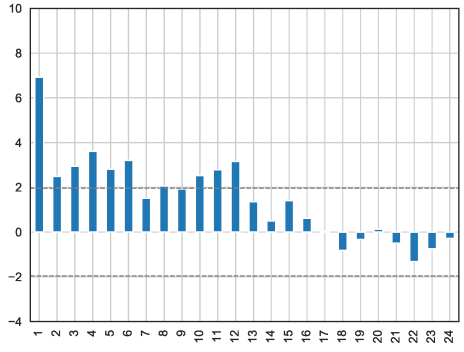
(d) WML



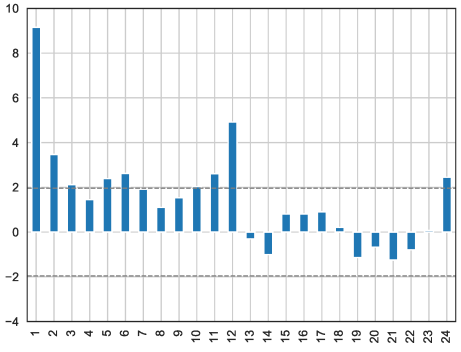
(e) RMW



(f) CMA

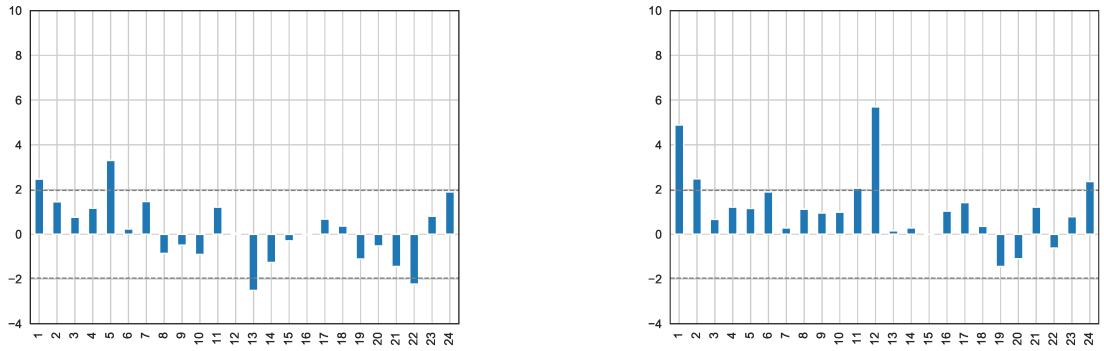


(g) BAB



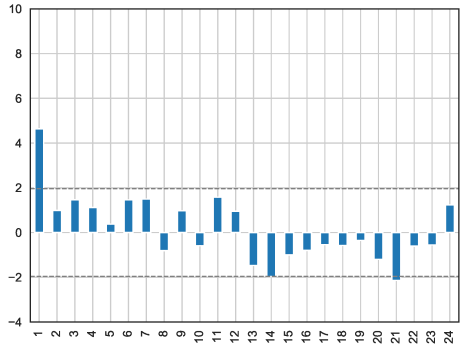
(h) Pooled

Figure 3.1: Lagged return regression *t*-statistics. This figure shows Newey and West (1987) *t*-statistics from regressions of factor returns on each factor's own return lagged from 1 to 24 months. Standard errors are clustered by time in the pooled regressions. The dashed lines represent 5% critical values.

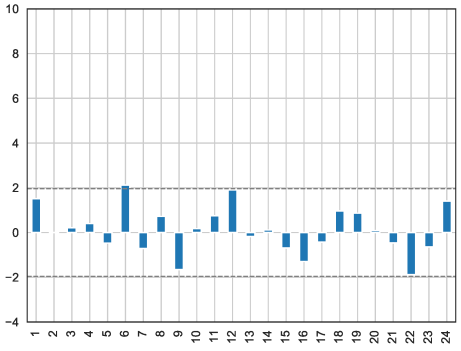


(a) RMRF

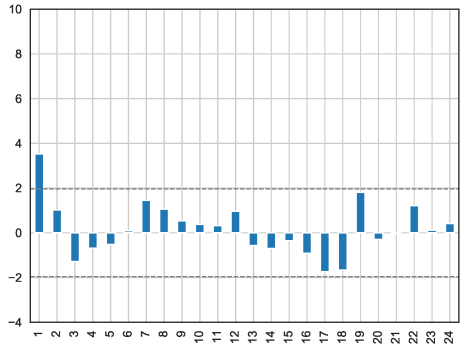
(b) SMB



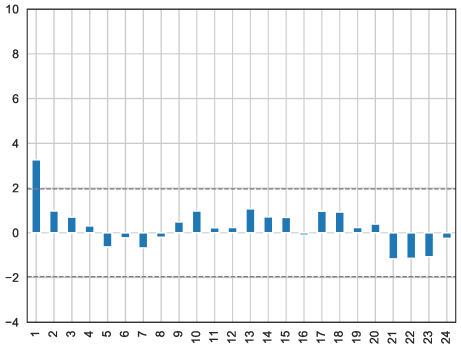
(c) HML



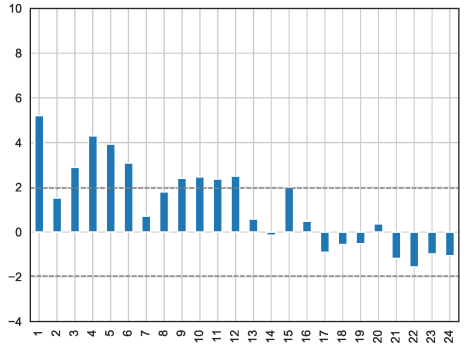
(d) WML



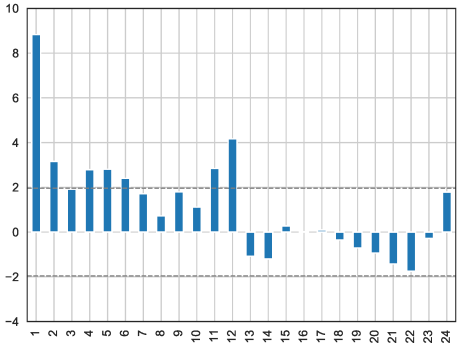
(e) RMW



(f) CMA

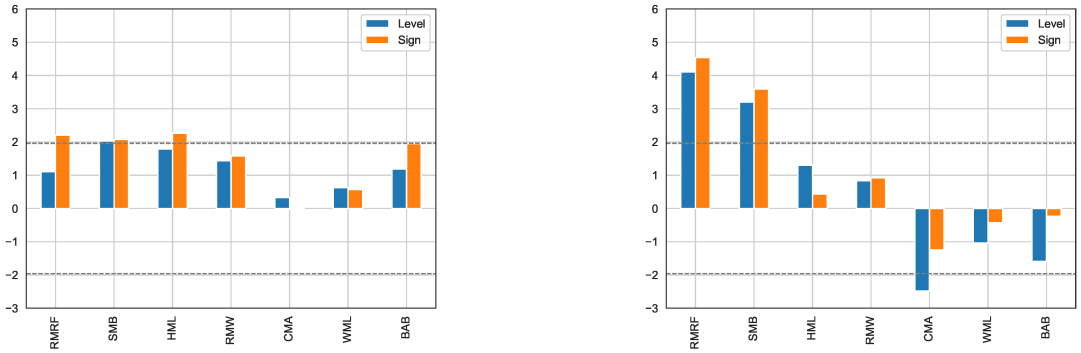


(g) BAB



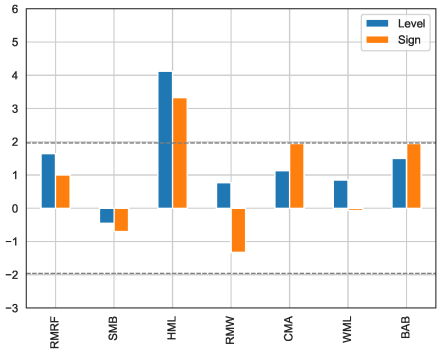
(h) Pooled

Figure 3.2: Lagged return sign regression *t*-statistics. This figure shows Newey and West (1987) *t*-statistics from regressions of factor returns on the sign each factor's own return lagged from 1 to 24 months. Standard errors are clustered by time in the pooled regressions. The dashed lines represent 5% critical values.

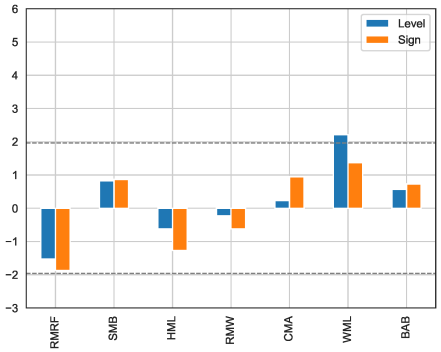


(a) RMRF

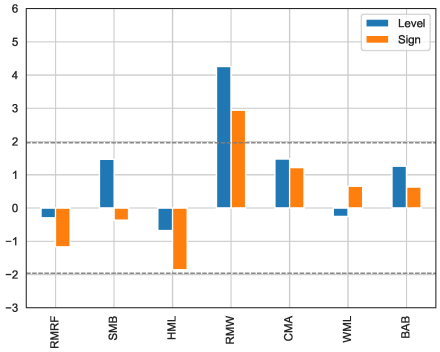
(b) SMB



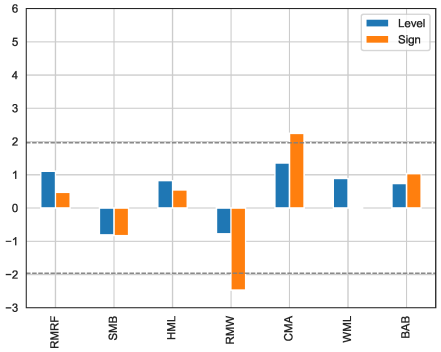
(c) HML



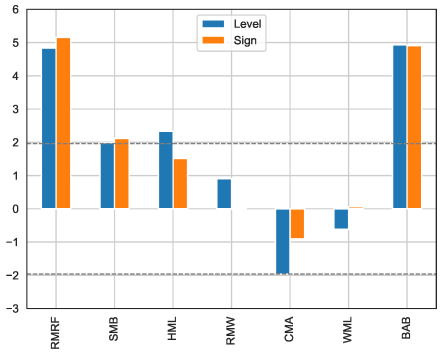
(d) WML



(e) RMW

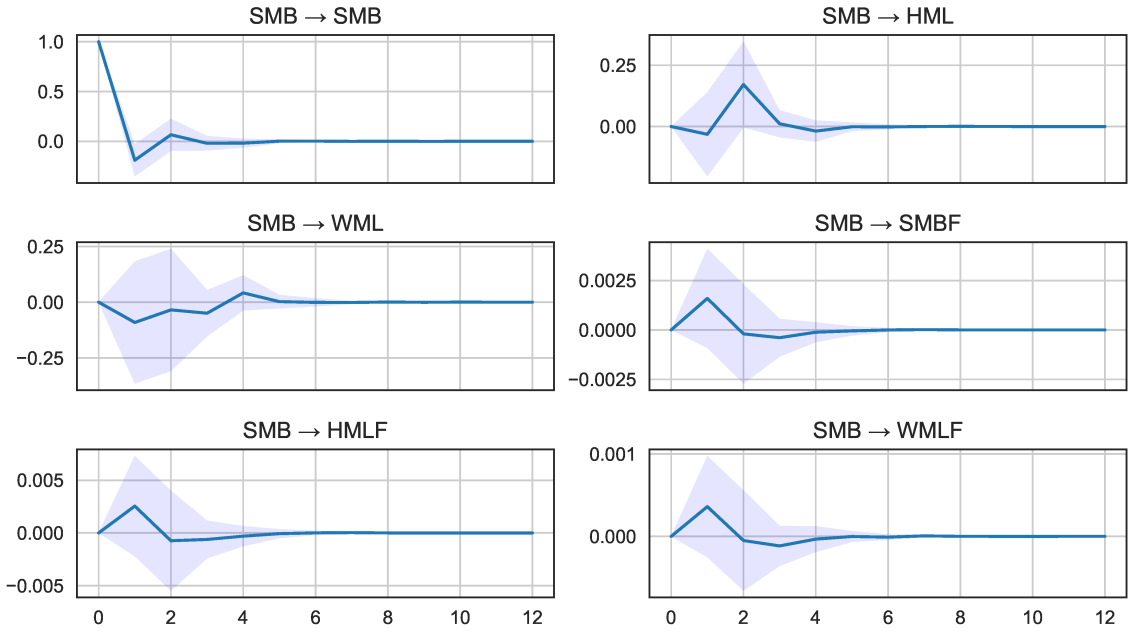


(f) CMA

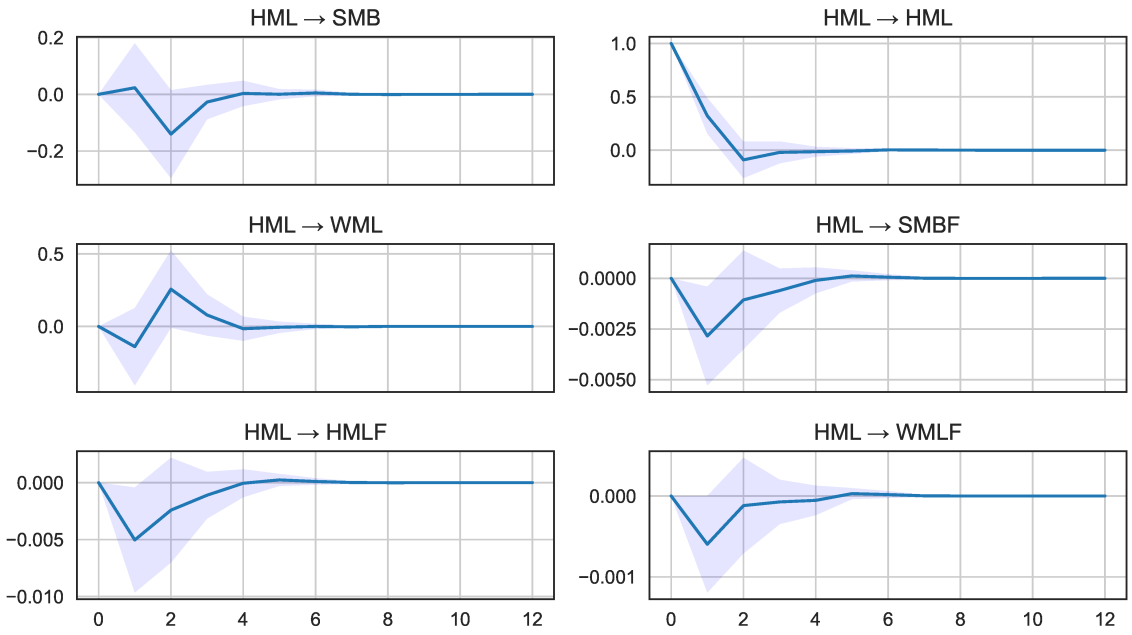


(g) BAB

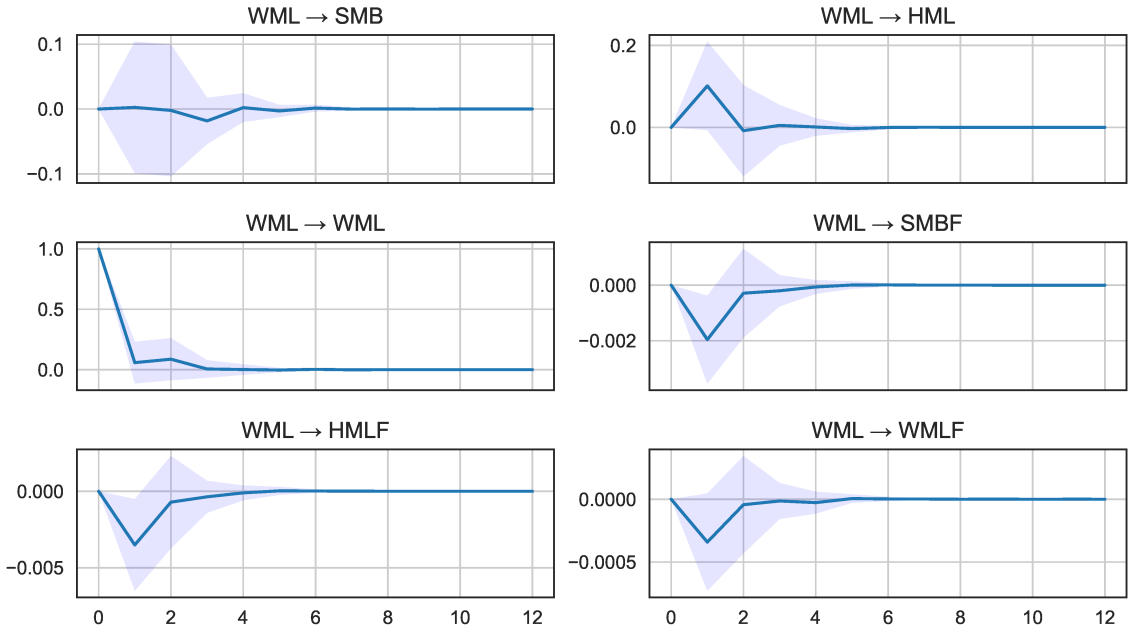
Figure 3.3: Lagged cross-factor return regression *t*-statistics. This figure shows t-statistics from regressions of factor returns on the level or sign of one period lagged returns on all factors. The dashed lines represent 5% critical values.


SMB → SMB
SMB → HML
SMB → WML
SMB → SMBF
SMB → HMLF
SMB → WMLF

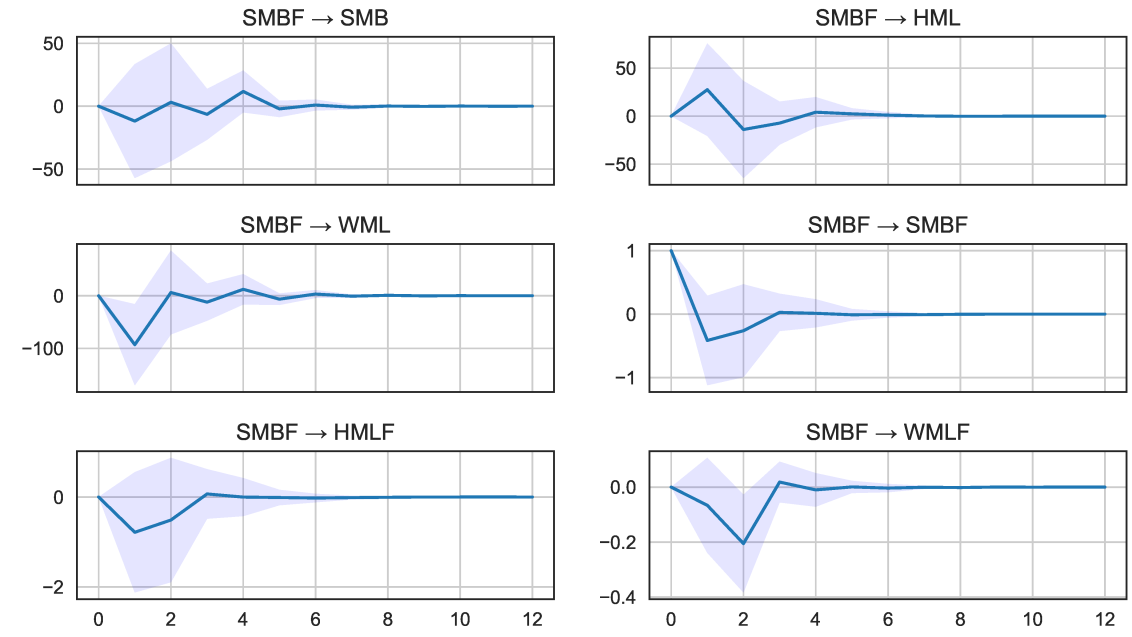
(a) SMB


HML → SMB
HML → HML
HML → WML
HML → SMBF
HML → HMLF
HML → WMLF

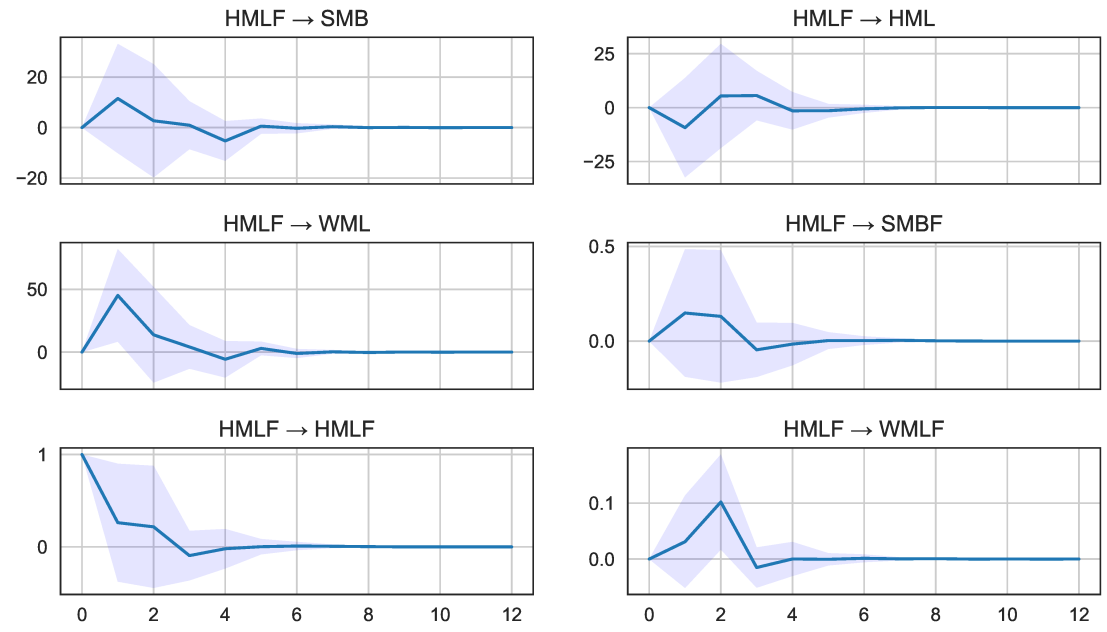
(b) HML


WML → SMB
WML → HML
WML → WML
WML → SMBF
WML → HMLF
WML → WMLF

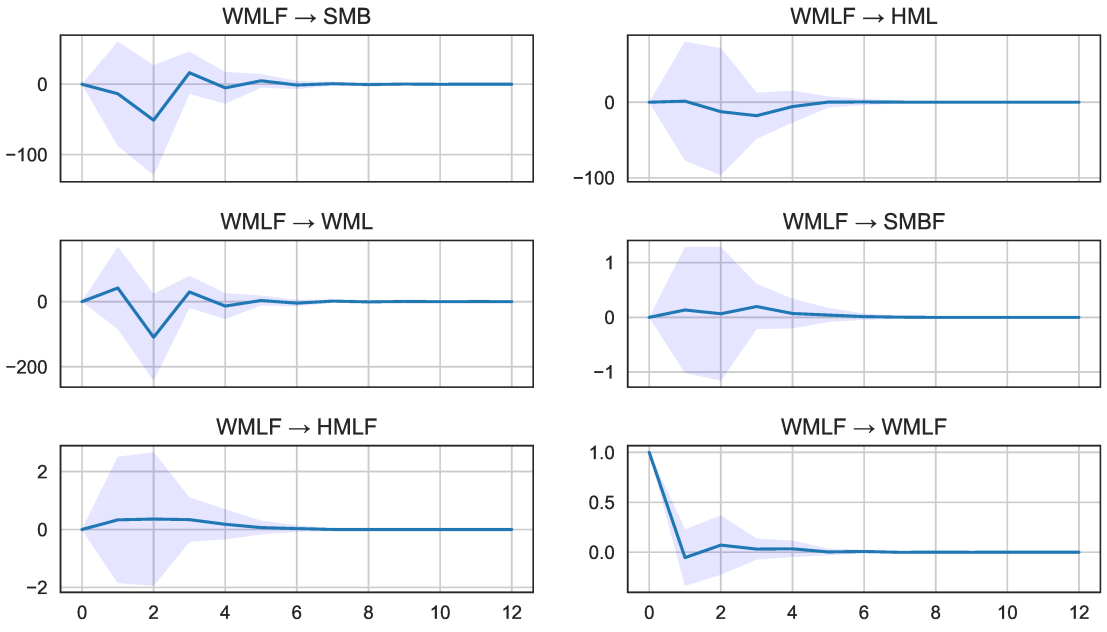
(c) SMB


SMBF → SMB
SMBF → HML
SMBF → WML
SMBF → SMBF
SMBF → HMLF
SMBF → WMLF

(d) HML



(e) SMB



(f) HML

Figure 3.4: Return-flow VAR impulse response functions. This figure shows the effect of a unit shock to a variable on all variables. SMB, HML, and WML are the Fama-French size, value, and cross-sectional momentum factors; SMBF, HMLF, and WMLF are aggregate ETF flows into size, value, and cross-sectional momentum funds respectively. The shaded area represents a 95% confidence interval using asymptotic standard errors.

Bibliography

Abel, A. B., and J. C. Eberly, 1996, Optimal investment with costly reversibility, *Review of Economic Studies* 63, 581–593.

Aboura, S., and Y. E. Arisoy, 2017, Does aggregate uncertainty explain size and value anomalies?, *Applied Economics* 49, 3214–3230.

Alberg, J., and Z. C. Lipton, 2018, Improving factor-based quantitative investing by forecasting company fundamentals. Working Paper. Available at: https://arxiv.org/abs/1711.04837.

Alquist, R., R. Isreal, and T. Moskowitz, 2018, Fact, fiction, and the size effect, *The Journal of Portfolio Management* 45, 34–61.

Andrei, D., and J. Cujean, 2017, Information percolation, momentum and reversal, *Journal of Financial Economics* 123, 617–645.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang, 2006, The cross-section of volatility and expected returns, *The Journal of Finance* 61, 259–299.

Ang, A., J. Liu, and K. Schwarz, 2019, Using stocks or portfolios in tests of factor models, *Journal of Financial and Quantitative Analysis* 55, 709–750.

Arnott, R. D., M. Clements, V. Kalesnik, and J. T. Linnainmaa, 2019, Factor momentum. Working Paper. Available at: https://ssrn.com/abstract=3116974.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen, 2013, Value and momentum everywhere, *The Journal of Finance* 68, 929–985.

Avramov, D., 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Banz, R. W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer, 2018, Extrapolation and bubbles, *Journal of Financial Economics* 129, 203–227.

Barlevy, G., and P. Veronesi, 2003, Rational panics and stock market crashes, *Journal of Economic Theory* 110, 234–263.

Bates, J. M., and C. W. J. Granger, 1969, The combination of forecasts, *Operational Research Quarterly* 20, 451–468.

Bhandari, L. C., 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *The Journal of Finance* 43, 507–528.

Bianchi, D., M. Buchner, and A. Tamoni, 2021, Bond risk premiums with machine learning, *The Review of Financial Studies* 34, 1046–1089.

Bianchi, D., and K. McAlinn, 2020, Divide and conquer: Financial ratios and industry returns predictability. Working Paper. Available at: https://ssrn.com/abstract=3136368.

Blitz, D., and M. X. Hanauer, 2020, Settling the size matter. Working Paper. Available at: https://ssrn.com/abstract=3686583.

Boudoukh, J., M. Richardson, and T. Smith, 1993, Is the ex ante risk premium always positive? A new approach to testing conditional asset pricing models, *Journal of Financial Economics* 32, 387–408.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood, 2018, Deep hedging. Working Paper. Available at: https://arxiv.org/abs/1802.03042.

Butaru, F., Q. Chen, B. Clark, S. Das, A. W. Lo, and A. Siddique, 2016, Risk and risk management in the credit card industry, *Journal of Banking & Finance* 72, 218–239.

Campa, J. M., 1993, Entry by foreign firms in the United States under exchange rate uncertainty, *The Review of Economics and Statistics* 75, 614–622.

Campbell, J. Y., and S. B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *The Review of Financial Studies* 21, 1509–1531.

Cao, V. N., 2015, What explains the value premium? The case of adjustment costs, operating leverage and financial leverage, *Journal of Banking & Finance* 59, 350–366.

Carlson, M., A. Fisher, and R. Giammarino, 2004, Corporate investment and asset price dynamics: Implications for the cross-section of returns, *The Journal of Finance* 59, 2577–2603.

Chen, L., M. Pelger, and J. Zhu, 2020, Deep learning in asset pricing. Working Paper. Available at: https://ssrn.com/abstract=3350138.

Chen, L., and X. Zhao, 2006, On the relation between the market-to-book ratio, growth opportunity, and leverage ratio, *Finance Research Letters* 3, 253–266.

Chinco, A., A. D. Clark-Joseph, and M. Ye, 2019, Sparse signals in the cross-section of returns, *The Journal of Finance* 74, 449–492.

Chirinko, R., and H. Schaller, 2009, The irreversibility premium, *Journal of Monetary Economics* 56, 390–408.

Choe, Y. S., 1994, The substitution effects of short-term debt for long-term debt on the expected returns of common stocks, *Asia Pacific Journal of Management* 11, 187–203.

Clark, T. E., and K. D. West, 2006, Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* 135, 155–186.

——— , 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Cochrane, J. H., and M. Piazessi, 2005, Bond risk premia, *American Economic Review* 95, 138–160.

Cooper, I., 2006, Asset pricing implications of nonconvex adjustment costs and irreversibility of investment, *The Journal of Finance* 61, 139–170.

Cremers, M. K. J., 2002, Stock return predictability: A Bayesian model selection perspective, *The Review of Financial Studies* 15, 1223–1249.

Dangl, T., and M. Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of Finance* 52, 1035–1058.

Daniel, K., and S. Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *The Journal of Finance* 52, 1–33.

Della Corte, P., and I. Tsiakas, 2012, Statistical and economic methods for evaluating exchange rate predictability, *Handbook of Exchange Rates* pp. 221–263.

DeMiguel, V., L. Garlappi, and R. Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *The Review of Financial Studies* 22, 1915–1953.

Dewally, M., L. H. Ederington, and C. S. Fernando, 2013, Determinants of trader profits in commodity futures markets, *The Review of Financial Studies* 26, 2648–2683.

Diebold, F. X., and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 3, 253–263.

Doshi, H., K. Jacobs, P. Kumar, and R. Rabinovich, 2019, Leverage and the cross-section of equity returns, *The Journal of Finance* 74, 1431–1471.

Duchin, R., and H. Levy, 2009, Markowitz versus the Talmudic portfolio diversification strategies, *The Journal of Portfolio Management* 35, 71–74.

Ehsani, S., and J. T. Linnainmaa, 2020, Factor momentum and the momentum factor. Working Paper. Available at: https://ssrn.com/abstract=3014521.

Falck, A., A. Rej, and D. Thesmar, 2020, Is factor momentum more than stock momentum? Working Paper. Available at: https://arxiv.org/abs/2009.04824.

Fama, E. F., and R. R. Bliss, 1987, The information in long-maturity forward rates, *The American Economic Review* 77, 680–692.

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

——— , 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

——— , 1996, Multifactor explanations of asset pricing anomalies, *The Journal of Finance* 51, 55–84.

——— , 2012, Size, value, and momentum in international stock returns, *Journal of Financial Economics* 105, 457–472.

——— , 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, E R.., and J. D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Farinas, J. C., and S. Ruano, 2005, Firm productivity, heterogeneity, sunk costs and market selection, *International Journal of Industrial Organization* 23, 505–534.

Faust, J., S. Gilchrist, J. H. Wright, and E. Zakrajssek, 2013, Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach, *The Review of Economics and Statistics* 95, 1501–1519.

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

Feng, G., N. Polson, and J. Xu, 2019, Deep learning in asset pricing. Working Paper. Available at: https://arxiv.org/abs/1805.01104v3.

Flögel, V., C. Schlag, and C. Zunft, 2019, Momentum-managed equity factors. Working Paper. Available at: https://ssrn.com/abstract=3423287.

Frazzini, A., and L. H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

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

Friedman, J. H., 2002, Stochastic gradient boosting, *Computational Statistics & Data Analysis* 38, 367–378.

Fulop, A., J. Li, and R. Wan, 2019, Real-time Bayesian learning and bond return predictability. Working Paper. Available at: https://ssrn.com/abstract=2979332.

Garcia-Feijoo, L., and R. D. Jorgensen, 2010, Can operating leverage be the cause of the value premium?, *Financial Management* 39, 1127–1154.

Gargano, A., D. Pettenuzzo, and A. Timmermann, 2019, Bond return predictability: Economic value and links to the macroeconomy, *Management Science* 65, 508–540.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch, 2007, Portfolio performance manipulation and manipulation-proof performance measures, *The Review of Financial Studies* 20, 1503–1546.

Gomes, J. F., and L. Schmid, 2010, Levered returns, *The Journal of Finance* 65, 467–494.

Gompers, P. A., and A. Metrick, 2001, Institutional investors and equity prices, *The Quarterly Journal of Economics* 116, 229–259.

Gordon, R. J., 1986, *The American Business Cycle: Continuity and Change* (University of Chicago Press).

Gorodnichenko, Y., and M. Weber, 2016, Are sticky prices costly? Evidence from the stock market, *American Economic Review* 106, 165–199.

Goyal, A., and N. Jegadeesh, 2018, Cross-sectional and time-series tests of return predictability: What is the difference?, *The Review of Financial Studies* 31, 1784–1824.

Goyal, A., and I. Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *The Review of Financial Studies* 21, 1455–1508.

Granger, C. W. J., and Y. Jeon, 2004, Thick modeling, *Economic Modelling* 21, 323–343.

Gu, S., B. T. Kelly, and D. Xiu, 2019, Autoencoder asset pricing models. Yale ICF Working Paper No. 2019–04. Available at: https://ssrn.com/abstract=3335536.

——— , 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Gupta, T., and B. Kelly, 2019, Factor momentum everywhere, *The Journal of Portfolio Management* 45, 13–36.

Halperin, I., 2017, QLBS: Q-learner in the Black-Scholes (-Merton) worlds. Working Paper. Available at: https://ssrn.com/abstract=3087076.

——— , 2018, The QLBS Q-learner goes NuQLear: Fitted Q iteration, inverse RL, and option portfolios. Working Paper. Available at: https://ssrn.com/abstract=3102707.

Han, Y., 2006, Asset allocation with a high dimensional latent factor stochastic volatility model, *The Review of Financial Studies* 19, 237–271.

Hansson, B., 2004, Human capital and stock returns: Is the value premium an approximation for return on human capital?, *Journal of Business Finance & Accounting* 31, 333–358.

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

Hastie, T., R. Tibshirani, and J. H. Friedman, 2009, *The elements of statistical learning: Data mining, inference, and predictions* (Springer) 2 edn.

He, X., and K. Li, 2015, Profitability of time series momentum, *Journal of Banking & Finance* 53, 140–157.

He, Z., B. Kelly, and A. Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Heaton, J. B., N. G. Polson, and J. H. Witte, 2016, Deep learning for finance: Deep portfolios, *Applied Stochastic Models in Business and Industry* 33, 3–12.

Henkel, S. J., J. S. Martin, and F. Nardari, 2011, Time-varying short-horizon predictability, *Journal of Financial Economics* 99, 560–580.

Henriksson, R. D., and R. C. Merton, 1981, On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *The Journal of Business* 54, 513–533.

Hjalmarsson, R., 2010, Predicting global stock returns, *Journal of Financial and Quantitative Analysis* 45, 49–80.

Hou, K., and R. Loh, 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* 121, 167–194.

Hou, K., W. Xiong, and L. Peng, 2010, A tale of two anomalies: The implications of investor attention for price and earnings momentum. Working Paper. Available at: https://ssrn.com/abstract=976394.

Huang, D., J. Li, L. Wang, and G. Zhou, 2020, Time series momentum: Is it there?, *Journal of Financial Economics* 135, 774–794.

Huang, J., and Z. Shi, 2019, Determinants of bond risk premia. Working Paper. Available at: https://ssrn.com/abstract=1573186.

Hurst, B., Y. H. Ooi, and L. H. Pedersen, 2017, A century of evidence on trend-following investing, *The Journal of Portfolio Management* 44, 15–29.

Jagannathan, R., and T. Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* 58, 1651–1683.

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

Jia, Y., Y. Liu, Y. Wu, and S. Yan, 2020, Cross predictability of currency returns: An analysis via machine learning. Working Paper. Available at: https://ssrn.com/abstract=3320199.

Kessides, I. N., 1990, Market concentration, contestability and sunk costs, *The Review of Economics and Statistics* 72, 614–622.

Khandani, A. E., A. J. Kim, and A. W. Lo, 2010, Consumer credit-risk models via machine-learning algorithms, *Journal of Banking & Finance* 34, 2767–2787.

Kim, A. Y., Y. Tse, and J. K. Wald, 2016, Time series momentum and volatility scaling, *Journal of Financial Markets* 30, 103–124.

Kolm, P. N., and G. Ritter, 2020, Modern perspectives on reinforcement learning in finance, *The Journal of Machine Learning in Finance* 1.

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

La Porta, R., 1996, Expectations and the cross-section of stock returns, *The Journal of Finance* 51, 1715–1742.

——— , J. Lakonishok, A. Shleifer, and R. Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *The Journal of Finance* 52, 859–874.

Lakonishok, J., A. Shleifer, and R. Vishny, 1994, Contrarian investment, extrapolation, and risk, *The Journal of Finance* 49, 1541–1578.

Lettau, M., and S. Ludvigson, 2009, Macro factors in bond risk premia, *The Review of Financial Studies* 22, 5027–5067.

Leung, R. C. W., and Y. Tam, 2018, The cross-section of asset synchronicity by elastic-net. Working Paper. Available at: https://ssrn.com/abstract=3300769.

Lev, B., 1974, On the association between operating leverage and risk, *The Journal of Financial and Quantitative Analysis* 9, 627–641.

Li, J., and I. Tsiakas, 2017, Equity premium prediction: The role of economic and statistical constraints, *Journal of Financial Markets* 36, 56–75.

——— , and W. Wang, 2015, Predicting exchange rates out of sample: Can economic fundamentals beat the random walk?, *Journal of Financial Econometrics* 13, 293–341.

Lo, A. W., and C. MacKinlay, 1990, When are contrarian profits due to stock market overreaction?, *The Review of Financial Studies* 3, 175–205.

Lovell, M. C., 2008, A simple proof of the FWL theorem, *The Journal of Economic Education* 39, 88–91.

Meese, R. A., and K. Rogoff, 1983, Empirical exchange rate models of the seventies: Do they fit out of sample?, *Journal of International Economics* 14, 3–24.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf, 2012, Currency momentum strategies, *Journal of Financial Economics* 106, 660–684.

Merton, R. C., 1980, On the microeconomic theory of investment under uncertainty, *Handbook of Mathematical Economics* 2, 601–669.

Messmer, M., 2017, Deep learning and the cross-section of expected returns. Working Paper. Available at: https://ssrn.com/abstract=3081555.

Min, C., and M. Zellner, 1993, Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates, *Journal of Econometrics* 56, 89–118.

Modigliani, F., and M. H. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review* 48, 261–297.

Molodtsova, T., and D. H. Papell, 2009, Out-of-sample exchange rate predictability with Taylor rule fundamentals, *Journal of International Economics* 77, 167–180.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Nagel, S., 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

Newey, W., and K. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Novy-Marx, R., 2011, Operating leverage, *Review of Finance* 15, 103–134.

——— , 2015, How can a q-theoretic model price momentum? NBER Working Paper No. w20985. Available at: https://ssrn.com/abstract=2572144.

Obreja, I., 2013, Book-to-market equity, financial leverage, and the cross-section of stock returns, *The Review of Financial Studies* 26, 1146–1189.

Ong, D., W. Fong, and C. Low, 2008, Investigating the risk argument for the value premium: A stochastic dominance approach. Working Paper. Available at: http://ssrn.com/abstract=1344011.

Ozdagli, A. K., 2012, Financial leverage, corporate investment, and stock returns, *The Review of Financial Studies* 25, 1033–1069.

Pettenuzzo, D., A. Timmermann, and R. Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 114, 517–553.

Pesaran, M. H., C. Schleicher, and P. Zaffaroni, 2009, Model averaging in risk management with an application to futures markets, *Journal of Empirical Finance* 16, 280–305.

Phalippou, L., 2007a, Can risk-based theories explain the value premium?, *Review of Finance* 11, 143–166.

——— , 2007b, Institutional ownership and the value premium. Working Paper. Available at: http://ssrn.com/abstract=360760.

Pitkäjärvi, A., M. Suominen, and L. Vaittinen, 2020, Cross-asset signals and time series momentum, *Journal of Financial Economics* 136, 63–85.

Rajan, R., and L. Zingales, 1995, What do we know about capital structure? Some evidence from international data, *The Journal of Finance* 50, 1421–1460.

Rapach, D., and G. Zhou, 2019, Time-series and cross-sectional stock return forecasting: New machine learning methods. Working Paper. Available at: https://ssrn.com/abstract=3428095.

Rapach, D. E., J. K. Strauss, and G. Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *The Review of Financial Studies* 23, 821–862.

Sarno, L., P. Schneider, and C. Wagner, 2016, The economic value of predicting bond risk premia, *Journal of Empirical Finance* 37, 247–267.

Shumway, T., 1997, The delisting bias in CRSP data, *The Journal of Finance* 52, 327–340.

——— , and V. A. Warther, 1999, The delisting bias in CRSP's Nasdaq data and its implications for the size effect, *The Journal of Finance* 54, 2361–2379.

Sirignano, J., A. Sadhwani, and K. Giesecke, 2018, Deep learning for mortgage risk. Working Paper. Available at: https://arxiv.org/abs/1607.02470.

Stock, J. H., and M. W. Watson, 2004, Combination forecasts of output growth in a seven-country data set, *Journal of Forecasting* 23, 405–430.

Sylvain, S., 2014, Does human capital risk explain the value premium? Working Paper. Available at: http://ssrn.com/abstract=2400593.

Thornton, D. L., and G. Valente, 2012, Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective, *The Review of Financial Studies* 25, 3141–3168.

Timmermann, A., 2006, Chapter 4 forecast combinations, *Handbook of Economic Forecasting* 1, 135–196.

Vassalou, M., and Y. Xing, 2004, Default risk in equity returns, *The Journal of Finance* 59, 831–868.

Warther, V. A.., 1995, Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39, 209–235.

Welch, I., 2004, Capital structure and stock returns, *Journal of Political Economy* 112, 106–131.

Zhang, L., 2005, The value premium, *The Journal of Finance* 60, 67–103.

Appendix

A.1 Derivation of the Second Stage Regression

In section 1.4.2, I claimed the second-stage cross-sectional regression was

$$\widetilde{\mathrm{BM}}_{it} \;=\; \gamma_{1t}\widetilde{C}^1_{it} + \cdots + \gamma_{kt}\widetilde{C}^k_{it} + \tilde{u}_{it}, \quad i = 1, \ldots, n. \tag{A.1}$$

In this Appendix, I derive this equation by orthogonalising equation (1.9). To do this, it is useful to define the annihilator matrix $\mathbf{M}_{Xt} = \mathbf{I}_n - \mathbf{X}_t \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1} \mathbf{X}_t$ which generates the residuals from a linear projection onto the column space of $\mathbf{X}_t$, where $\mathbf{X}_t$ is an $n \times m$ matrix that vertically stacks the $\boldsymbol{x}_{it}$s for all i in a given month. Similarly, vertically stacking equation (1.9) for all i gives $\mathbf{BM}_t = \gamma_{0t}\mathbf{1}_n + \gamma_{1t}\mathbf{C}^1_t + \ldots \gamma_{kt}\mathbf{C}^k_t + \mathbf{u}_t$, where $\mathbf{1}_n$ is an $n \times 1$ vector of ones. Premultiplying this by $\mathbf{M}_{Xt}$ gives:

$$\mathbf{M}_{Xt}\mathbf{BM}_t \;=\; \gamma_{0t}\mathbf{M}_{Xt}\mathbf{1}_n + \gamma_{1t}\mathbf{M}_{Xt}\mathbf{C}^1_t + \ldots \gamma_{kt}\mathbf{M}_{Xt}\mathbf{C}^k_t + \mathbf{M}_{Xt}\mathbf{u}_t, \tag{A.2}$$

which can be written as:

$$\widetilde{\mathbf{BM}}_t \;=\; \gamma_{0t}\widetilde{\mathbf{1}}_n + \gamma_{1t}\widetilde{\mathbf{C}}^1_t + \ldots \gamma_{kt}\widetilde{\mathbf{C}}^k_t + \widetilde{\mathbf{u}}_t. \tag{A.3}$$

The term $\widetilde{\mathbf{1}}_n$ represents the residual from a regression of a column of ones on $\mathbf{X}_t$. As long as $\mathbf{X}_t$ contains a column of ones itself, then the system of equations defining the regression is exactly identified, and the regression residuals will be zero. By assumption, $\mathbf{X}_t$ contains a constant, so $\widetilde{\mathbf{1}}_n = \mathbf{0}$, and hence there is no intercept in equation (1.12).

A.2 Moments of the Proportions

This Appendix closely follows Hou and Loh (2016), but I explicitly show how the means and variances

are derived using a first-order Taylor series expansion. Recall from equation (1.13) that we can decompose the slope coefficient from a regression of orthogonalised returns on the orthogonalised BM ratio as follows:

$$\begin{aligned}
\beta_t &= \frac{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1} \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} \\
&= \frac{\sum_{i=1}^{n} \left(\gamma_{1t-1} \widetilde{C}_{it-1}^1 + \cdots + \gamma_{kt-1} \widetilde{C}_{it-1}^k + \tilde{u}_{it-1} \right) \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} \\
&= \frac{\gamma_{1t-1} \sum_{i=1}^{n} \widetilde{C}_{it-1}^1 \tilde{r}_{it} + \cdots + \gamma_{kt-1} \sum_{i=1}^{n} \widetilde{C}_{it-1}^k \tilde{r}_{it} + \sum_{i=1}^{n} \tilde{u}_{it-1} \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} \\
&= \gamma_{1t-1} \frac{\sum_{i=1}^{n} \widetilde{C}_{it-1}^1 \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} + \cdots + \gamma_{kt-1} \frac{\sum_{i=1}^{n} \widetilde{C}_{it-1}^k \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} + \frac{\sum_{i=1}^{n} \tilde{u}_{it-1} \tilde{r}_{it}}{\sum_{i=1}^{n} \widetilde{\mathrm{BM}}_{it-1}^2} \\
&\equiv \tilde{\beta}_t^1 + \cdots + \tilde{\beta}_t^k + \tilde{\beta}_t^u. \qquad \text{(A.4)}
\end{aligned}$$

We can divide this by β_t to get the percentage of the relationship explained by each candidate variable and the residual component:

$$\frac{\beta_t^1}{\beta_t} + \cdots + \frac{\beta_t^k}{\beta_t} + \frac{\beta_t^u}{\beta_t} = 1. \qquad \text{(A.5)}$$

For the purpose of inference, the means and variances of these proportions are unobtainable in closed form. However, since the proportions are simply ratios of two random variables whose distributions are known, their moments can be approximated using the multivariate delta method. Consider a first-order Taylor series expansion of β_t^i/β_t where $i \in \{1, \ldots, k, u\}$ around the means $(\bar{\beta}_t^i, \bar{\beta}_t)$:

$$\frac{\beta_t^i}{\beta_t} \approx \frac{\bar{\beta}_t^i}{\bar{\beta}_t} + \frac{1}{\bar{\beta}_t} \left(\beta_t^i - \bar{\beta}_t^i \right) - \frac{\bar{\beta}_t^i}{\bar{\beta}_t^2} \left(\beta_t - \bar{\beta}_t \right). \qquad \text{(A.6)}$$

Then the mean and variance of β_t^i/β_t are approximately:

$$\mathbb{E}\left(\frac{\beta_t^i}{\beta_t} \right) \approx \mathbb{E}\left(\frac{\bar{\beta}_t^i}{\bar{\beta}_t} + \frac{1}{\bar{\beta}_t} \left(\beta_t^i - \bar{\beta}_t^i \right) + \frac{\bar{\beta}_t^i}{\bar{\beta}_t^2} \left(\beta_t - \bar{\beta}_t \right) \right)$$

$$
\begin{aligned}
&= \frac{\bar{\beta}_t^i}{\bar{\beta}_t} + \frac{1}{\bar{\beta}_t}\mathbb{E}\left(\beta_t^i - \bar{\beta}_t^i\right) + \frac{\bar{\beta}_t^i}{\bar{\beta}_t^2}\mathbb{E}\left(\beta_t - \bar{\beta}_t\right) \\
&= \frac{\bar{\beta}_t^i}{\bar{\beta}_t}, \qquad \text{(A.7)}
\end{aligned}
$$

$$
\begin{aligned}
\text{var}\left(\frac{\beta_t^i}{\beta_t}\right) &\approx \text{var}\left(\frac{\bar{\beta}_t^i}{\bar{\beta}_t} + \frac{1}{\bar{\beta}_t}\left(\beta_t^i - \bar{\beta}_t^i\right) - \frac{\bar{\beta}_t^i}{\bar{\beta}_t^2}\left(\beta_t - \bar{\beta}_t\right)\right) \\
&= \text{var}\left(\frac{1}{\bar{\beta}_t}\beta_t^i - \frac{\bar{\beta}_t^i}{\bar{\beta}_t^2}\beta_t\right) \\
&= \frac{1}{\bar{\beta}_t^2}\text{var}\left(\beta_t^i\right) + \left(\frac{\bar{\beta}_t^i}{\bar{\beta}_t^2}\right)^2 \text{var}\left(\beta_t\right) - 2\frac{\bar{\beta}_t^i}{\bar{\beta}_t^3}\text{cov}(\beta_t^i, \beta_t) \\
&= \left(\frac{\bar{\beta}_t^i}{\bar{\beta}_t}\right)^2 \left(\frac{\text{var}\left(\beta_t^i\right)}{\bar{\beta}_t^{i\,2}} + \frac{\text{var}\left(\beta_t\right)}{\bar{\beta}_t^2} - 2\frac{\text{cov}(\beta_t^i, \beta_t)}{\bar{\beta}_t^i \bar{\beta}_t}\right). \qquad \text{(A.8)}
\end{aligned}
$$

In the empirical analysis, I use the sample analogues of (A.7) and (A.8) to estimate the moments of the proportions, where:

$$
\begin{aligned}
\bar{\beta}_t^i &= \frac{1}{T-1}\sum_{t=2}^{T}\beta_t^i \qquad \text{(A.9)} \\
\bar{\beta}_t &= \frac{1}{T-1}\sum_{t=2}^{T}\beta_t \qquad \text{(A.10)} \\
\text{var}(\beta_t^i) &= \frac{1}{T-1}\sum_{t=2}^{T}(\beta_t^i - \bar{\beta}_t^i)^2 \qquad \text{(A.11)} \\
\text{var}(\beta_t) &= \frac{1}{T-1}\sum_{t=2}^{T}(\beta_t - \bar{\beta}_t)^2 \qquad \text{(A.12)} \\
\text{cov}(\beta_t^i, \beta_t) &= \frac{1}{T-1}\sum_{t=2}^{T}(\beta_t^i - \bar{\beta}_t^i)(\beta_t - \bar{\beta}_t) \qquad \text{(A.13)}
\end{aligned}
$$

are computed from the time series of β_t^i and β_t.

Appendix: Tables

A.3 List of Predictors and Abbreviations

GFD Index Name	Abbreviation	GFD Index Name	Abbreviation
Inflation			
Australia Consumer Price Index Inflation Rate	cpi_aus	Italy Consumer Price Index Inflation Rate	cpi_ita
Canada Consumer Price Index Inflation Rate	cpi_can	Japan Consumer Price Index Inflation Rate	cpi_jpn
Switzerland Consumer Price Index Inflation Rate	cpi_che	Norway Consumer Price Index Inflation Rate	cpi_nor
Germany Consumer Price Index Inflation Rate	cpi_deu	New Zealand Consumer Price Index Inflation Rate	cpi_nzl
Denmark Consumer Price Index Inflation Rate	cpi_dnk	Sweden Consumer Price Index Inflation Rate	cpi_swe
France Consumer Price Index Inflation Rate	cpi_fra	United States BLS Consumer Price Index Inflation Rate NSA	cpi_usa
United Kingdom Consumer Price Index	cpi_gbr	Euro Consumer Price Index Inflation Rate	cpi_eur
GDP			
Australia Real GDP in 2007-2008 Dollars	gdp_aus	Great Britain Real GDP in 2008 Pounds	gdp_gbr
Canada Real GDP in 2007 Dollars	gdp_can	Italy Real GDP in 2010 Euros	gdp_ita
Switzerland Real GDP in 2010 Francs	gdp_che	Japan Real GDP in 2010 Yen	gdp_jpn
Germany Real GDP in 2010 Euros	gdp_ger	Norway Real GDP in 2010 Krone	gdp_nor
Denmark Real GDP in 2010 Kroner	gdp_dnk	Sweden Real GDP in 2010 Kroner	gdp_swe
France Real GDP in 2010 Euros	gdp_fra	United States Real GDP in 2012 Dollars	gdp_usa
Commodities			
GFD Indices USA Agricultural Commodity Index	com_agri	GFD Indices USA Livestock Commodity Sub-Index	com_livestock
GFD Indices USA Barley Commodity Index	com_barley	GFD Indices USA Milk Commodity Index	com_milk
GFD Indices USA Base Metals Sub-Index	com_base_metals	GFD Indices USD Commodity Composite Index	com_composite
GFD Indices USA Beef Commodity Index	com_beef	GFD Indices USA Natural Gas Commodity Index	com_natural_gas
GFD Indices USA Beverages Commodity Sub-Index	com_beverages	GFD Indices USA Non-Food Agriculturals Sub-Index	com_nonfood_agri
GFD Indices USA Butter Commodity Index	com_butter	GFD Indices USA Oats Commodity Index	com_oats
GFD Indices USA Coal Commodity Index	com_coal	GFD Indices USA Oil Commodity Index	com_oil
GFD Indices USA Coffee Commodity Index	com_coffee	GFD Indices USA Oils and Meal Commodity Sub-Index	com_meal
GFD Indices USA Corn Commodity Index	com_corn	GFD Indices USA Potatoes Commodity Index	com_potatoes
GFD Indices USA Cotton Commodity Index	com_cotton	GFD Indices USA Precious Metals Sub-Index	com_prec_metals
GFD Indices USA Energy Commodity Index	com_energy	GFD Indices USA Rice Commodity Index	com_rice
GFD Indices USA Grains Commodity Sub-Index	com_grains	GFD Indices USA Softs Commodity Sub Index	com_softs
GFD Indices USA Hay Commodity Index	com_hay	GFD Indices USA Sugar Commodity Index	com_sugar
GFD Indices USA Hogs Commodity Index	com_hogs	GFD Indices USA Wheat Commodity Index	com_wheat
GFD Indices USA Industrial Commodity Index	com_industrial	GFD Indices USA Lumber Wood Commodity Index	com_lumber

GFD Indices USA Iron Commodity Index	com_iron	GFD Indices USA Wool Commodity Index	com_wool
GFD Indices USA Lard Commodity Index	com_lard		
Bond Returns			
GFD Indices Australia USD Bond Return Index	bnd_aus	GFD Indices France USD Bond Return Index	bnd_fra
GFD Indices Belgium USD Bond Return Index	bnd_bel	GFD Indices United Kingdom USD Bond Return Index	bnd_gbr
GFD Indices Canada USD Bond Return Index	bnd_can	GFD Indices Italy USD Bond Return Index	bnd_ita
GFD Indices Germany USD Bond Return Index	bnd_ger	GFD Indices Japan USD Bond Return Index	bnd_jpn
GFD Indices Denmark USD Bond Return Index	bnd_dnk	GFD Indices United States Bond Return Index	bnd_usa
GFD Indices Spain USD Bond Return Index	bnd_esp	GFD Indices USA Total Return T-Bill Index	rf
Equity Returns			
GFD Indices Australia ASX Accumulation Index-All Ordinaries	eq_aus	GFD Indices France CAC All-Tradable Total Return Index	eq_fra
GFD Indices Brussels All-Share Return Index	eq_bel	GFD Indices Italy BCI Global Return Index	eq_ita
GFD Indices Canada S&P/TSX-300 Total Return Index	eq_can	GFD Indices Japan Topix Total Return Index	eq_jpn
GFD Indices Germany CDAX Total Return Index	eq_ger	GFD Indices UK Top 100 Total Return Index-Daily	eq_gbr
GFD Indices OMX Copenhagen All-Share Gross Index	eq_dnk	GFD Indices USA Top 100 Total Return Index-Daily	eq_usa
GFD Indices Barcelona SE-30 Return Index (USD)	eq_esp		
Yield (10YR)			
Australia 10-year Government Bond Yield	yield_10yr_aus	India 10-Year Government Bond Yield	yield_10yr_ind
Austria 10-year Government Bond Yield	yield_10yr_aut	Ireland 10-year Government Bond Yield	yield_10yr_irl
Belgium 10-year Government Bond Yield	yield_10yr_bel	Italy 10-year Government Bond Yield	yield_10yr_ita
Canada 10-year Government Bond Yield	yield_10yr_can	Japan 10-year Government Bond Yield	yield_10yr_jpn
Switzerland 10-year Government Bond	yield_10yr_che	Netherlands 10-year Government Bond Yield	yield_10yr_nld
Germany 10-year Benchmark Bond	yield_10yr_ger	Norway 10-year Government Bond Yield	yield_10yr_nor
Denmark 10-year Government Bond Yield	yield_10yr_dnk	New Zealand 10-year Government Bond Yield	yield_10yr_nzl
Spain 10-year Government Bond Yield	yield_10yr_esp	Portugal 10-year Government Bond Yield	yield_10yr_prt
Finland 10-year Government Bonds	yield_10yr_fin	Sweden 10-year Government Bond Yield	yield_10yr_swe
France 10-year Government Bond Yield	yield_10yr_fra	USA 10-year Bond Constant Maturity Yield	yield_10yr_usa
United Kingdom 10-year Government Bond Yield	yield_10yr_gbr	South Africa 10-Year Bond Yield	yield_10yr_zaf
Unemployment			
Australia Unemployment Rate	unem_aus	Great Britain Unemployment Rate	unem_gbr
Belgium Unemployment Rate	unem_bel	Italy Unemployment Rate	unem_ita
Germany Unemployment Rate	unem_ger	Japan Unemployment Rate	unem_jpn
Denmark Unemployment Rate	unem_dnk	United States U-3 Total unemployed[a]	unem_usa
France Unemployment Rate	unem_fra		
Net Exports			
Australia Net Exports	nx_aus	Germany Net Exports	nx_ger

Belgium Net Exports	nx_bel	Italy Net Exports	nx_ita
Canada Net Exports	nx_can	Japan Net Exports	nx_jpn
Denmark Net Exports	nx_dnk	Spain Net Exports	nx_esp
France Net Exports	nx_fra	United Kingdom Net Exports	nx_gbr
Currencies			
Real US Dollars per Australian Dollar	usdaud	Real US Dollars per Euro	usdeur
Real US Dollars per France Franc	usdffr	Real US Dollars per Canadian Dollar	usdcad
Real US Dollars per Germany Deutschemark	usddeu	Real US Dollars per Swiss Franc	usdchf
Real US Dollars per Italy Lira	usditl	Real US Dollars per100 Japanese Yen	usdjpy
Real US Dollars per New Zealand Dollar	usdnzd	Real United States Dollar per Norway Krone	usdnkk
Real US Dollars per British Pound	usdgbp	Real United States Dollar per Sweden Krona	usdsek

[a] This series is expressed as a percent of the civilian labour force.