Machine learning and the cross-section of expected stock returns

DISSERTATION
University of St. Gallen
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Marcial Messmer

from

Germany

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Robert J. Hodrick, PhD
Prof. Dr. Michael Lechner

Dissertation no. 4816

Difo-Druck GmbH, Bamberg, 2018

The University of St. Gallen School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing and opinion on the views herein expressed

St. Gallen, June 15, 2018

The President:

Prof. Dr. Thomas Bieger

To Klara, Eberhard, Anna-Lena, Beatrice and Jana

Acknowledgments

This work has benefited from financial support of the Swiss National Science Foundation. Moreover, this dissertation has profited from open source technology, in particular, from the Python community.

I thank Professor Francesco Audrino for his supervision and support throughout this project. I am glad he gave me the opportunity to write these chapters under his guidance. I am grateful to Professor Robert Hodrick for being co-referee as well as sponsoring my research visit at Columbia University. His support and encouragement was inspiring. I thank Professor Juan-Pablo Ortega for his referee contribution. Furthermore, I would like to thank participants at research and conference seminars for valuable feedback and comments.

I am indebted to Dmitry Borisenko, Daniel Buncic, Kameliya Filipova, Anselm Ivanovas, Constantin Roth and Anastasija Tetereva for numerous discussions, their answers, their ideas not only for research but also beyond — a chapter of friendships I gladly add to this thesis.

The journey of completing this PhD project was time and energy consuming but it mostly provided lots of joy and fun in return. Pursuing this endeavor would not have been possible without the support of friends, my family and my wife. Their support through good and sometimes challenging periods was special and dates back a very long time.

Two women deserve a special note. I thank my mother for all the unconditional support she has provided me at anytime in my life. I thank my wife Jana for her support and love, she was always there for me. I will remain grateful to both of them for this as long as I shall live.

Zurich, August 2018 Marcial Messmer

Contents

Summary

Modeling expected cross-sectional stock returns has a long tradition in asset pricing. My dissertation is motivated by shortcomings of the prevailing and classically used portfolio sorting approach. Consequently, this thesis tackles the task with alternative methodologies. It comprises classical linear models but includes more advanced machine learning algorithms as well. The work contains three chapters. The first chapter uses a linear-shrinkage approach to select relevant firm characteristics (FC), the second leverages a deep-learning architecture to detect non-linear patterns in expected stock returns reviews and the third part compares two established linear models. In short, this thesis covers machine learning based stock picking.

The goal of the first chapter is twofold. First, Francesco Audrino and I show, based on Monte Carlo simulations, that the adaptive Lasso methodology is generally suitable for panel specifications. These findings are robust to various distributional assumptions. Second, the empirical task solves the multivariate problem of selecting a set of FC helpful in describing expected stock returns. We find a large number of FC does not survive this shrinkage procedure, however, we document a highly dimensional linear relationship.

Chapter 2 loosens the linearity restriction and trains a deep-learning algorithm to identify non-linearities. The tedious task for the researcher is to select appropriate hyper-parameters. I show that random search yields promising results when compared to the linear model based on training data. The portfolio exercise reveals that these benefits materialize on a test data set, a linear model hinges behind the non-linear framework.

The final chapter extensively reviews two standard linear estimators. Despite identical objective functions, the two methods exhibit substantial variations with respect to model inference. An investor's perspective shows that these differences lead only to marginal economical discrepancies.

Zusammenfassung

Die Modellierung der erwarteten Renditen von Aktien hat eine lange Tradition im Asset Pricing. Meine Dissertation ist motiviert durch die Unzulänglichkeiten des vorherrschenden und klassisch genutzten Portfolio Sortierverfahrens. Daher wird in dieser Doktorarbeit die Aufgabe mit alternativen Methoden angegangen. Es umfasst klassische lineare Modelle, beinhaltet aber auch fortschrittlichere maschinelle Lernalgorithmen. Die Arbeit umfasst drei Kapitel. Das erste Kapitel verwendet einen linearen Shrinkage Ansatz zur Auswahl relevanter Unternehmensmerkmale (UM), das zweite nutzt eine Deep-Learning Architektur, um nichtlineare Muster in erwarteten Aktienrenditen zu erkennen und der dritte Teil vergleicht zwei etablierte lineare Modelle. Kurz gesagt, beschäftigt sich diese Arbeit mit dem maschinellen Selektieren von Aktien.

Das Ziel des ersten Kapitels ist zweifach. Zunächst zeigen Francesco Audrino und ich anhand von Monte Carlo Simulationen, dass die adaptive Lasso Methode allgemein für Panel Spezifikationen geeignet ist. Diese Ergebnisse sind robust gegenüber verschiedenen Verteilungsannahmen. Zweitens löst die empirische Aufgabe das multivariate Problem der Auswahl einer Menge von UM, die bei der Beschreibung der erwarteten Aktienrenditen hilfreich sind. Wir zeigen, dass eine grosse Anzahl von UM diesen Shrinkage Ansatz nicht überlebt, jedoch dokumentieren wir einen hoch-dimensionalen linearen Zusammenhang.

Kapitel 2 lockert die Linearitätsbeschränkung und trainiert einen Deep-Learning Algorithmus zur Identifizierung von Nichtlinearitäten. Die aufwendige Aufgabe für den Forscher besteht darin, geeignete Hyperparameter auszuwählen. Ich zeige, dass eine auf Zufall basierte Suche im Vergleich zu dem auf Trainingsdaten basierenden linearen Modell vielversprechende Ergebnisse liefert. Die Portfolioübung zeigt, dass diese Vorteile auf einem Testdatensatz Bestand haben, da das lineare Modell hinter dem nichtlinearen Ansatz zurückfällt.

Im letzten Kapitel werden zwei lineare Standardschätzer ausführlich besprochen. Trotz identischer Zielfunktionen zeigen die beiden Verfahren wesentliche Unterschiede in Bezug auf die Modellschlussfolgerung. Eine Investorenperspektive zeigt, dass diese Unterschiede nur zu marginal messbaren ökonomischen Diskrepanzen führen.

Preface

This cumulative dissertation comprises three papers. All three papers are concerned with the estimation of expected cross-sectional stock returns. The thesis is motivated by the challenge to jointly process a multitude of firm characteristics (FC) in an efficient manner. The asset pricing literature traditionally builds on portfolio sorts or classical linear regressions. Portfolio sorts are robust but seem inefficient, but more crucially suffer from the curse of dimensionality. Given 447 documented published FC up to date, which claim to explain differences in expected returns, portfolio sorting becomes an infeasible exercise if one is interested in analyzing their joint dependence. Consequently, alternative methodologies are a requisite for understanding this highly dimensional problem. As a result, this challenge turned into an active area of research in finance and financial econometrics. This dissertation contributes to this strand of the literature in each of the three following chapters.

The data of this thesis is built up based on the CRSP/Compustat database, which contains all stocks traded in the United States. Next to the individual stock returns and other market data, the database includes balance sheet information on the stock level. Based on these market and accounting data, I reconstruct several prominent FC, which are all based on published research, for example, momentum indicators, but also valuation and accounting based metrics like the book-to-market or the earnings-to-price ratio. In total, I consider 68 published FC.

In the first chapter, Francesco Audrino and I approach the problem of selecting relevant FC based on a linear shrinkage estimator. In particular, we suggest applying the adaptive Lasso as a selection tool. This choice is motivated by the variance reduction benefits of L1 shrinkage estimators compared to ordinary least squares (OLS). Despite imposing a small bias, the tradeoff between bias and variance seems promising given the relatively large noise component observed when predicting cross-sectional returns. Moreover, the adaptive Lasso is less restrictive with respect to its required conditions compared to the standard Lasso estimator.

The contribution is threefold. First, we show generally, that the adaptive Lasso is a suitable selection tool for panel data regression problems. This evidence is based on extensive Monte Carlo simulations and holds for various distributional assumptions. In particular, we consider a relatively low signal-to-noise ratio, heteroscedasticity (estimated based on the US stock market), cross-sectional correlation but also a high degree of multicollinearity among the independent variables. The latter is an important aspect given the actual empirical correlation structure of FC. Indeed, the simulation reveals that the Lasso estimator produces a high error rate given high levels of correlation. On the other hand, the adaptive Lasso shows its resilience to these assumptions. Second, our simulation study provides a flexible framework to study highly dimensional data generating processes (DGP), which possess typical characteristics observed in cross-sectional returns. A novel feature is, for example, the simulation of large and realistic correlation matrices of the independent variables. This procedure is introduced from the statistics literature. Third, we estimate the empirical problem at hand. We find short-term reversal, the change in the six-months momentum, research spending scaled by market-value and two value related metrics to be the FC most robustly selected in the US cross-section of stock returns. Moreover, many of the 68 FC included in our analysis are not considered. Nonetheless, the return process we identify is highly multi-dimensional, comprising 37 FC.

The second chapter targets a different objective. Compared to the previous chapter, the focus is less on the identification of relevant FC, but more on achieving a high return prediction accuracy. For this purpose, I drop the linearity restriction and employ a deep learning framework. As the true DGP is unknown, the functional form between FC and expected returns is of unknown shape. Consequently, the deep learning approach provides an agnostic framework to deal with this model structure uncertainty.

In this chapter, I contribute to an emerging literature of non-linear cross-sectional estimation strategies. To the best of my knowledge, the paper is the first approach, which relies on a purely data-driven and algorithmically determined prediction in the context of cross-sectional stock returns. However, using artificial neural networks is not new in the finance literature. Recent innovations in software and hardware technology, reportedly improve classical prediction problems faced by the computer science community. Typically, the statistical properties of financial time series prediction problems deviate from the latter, for example, the signal-to-noise ratio is on an extremely low level. Consequently, the paper investigates if these recent advances are of any use when faced with predicting expected stock returns.

After applying a computationally efficient random search network optimization strategy, I find that DFN long-short portfolios can generate attractive risk-adjusted returns compared to a linear benchmark. These findings underscore the importance of non-linear relationships among FC and expected returns. The results are robust to size, weighting schemes, and portfolio cutoff points. Due to the high turnover of the strategy, a trading cost naive implementation would have historically been hardly profitable. Moreover, I show that price related FC, namely, short-term reversal and the twelve-months momentum, are among the main drivers of the return predictions. The majority of FC play a minor role in the variation of these predictions.

In the final chapter, I provide an empirical comparison of two established linear estimators. In particular, I investigate differences between the Fama and MacBeth (1973) (FM) and pooled OLS based inference. The literature mostly relies on the former procedure. This research question is motivated by two aspects. First, it is important to understand why and how this choice affects the model inference, as both methods aim to solve the exact same objective function. Second, it is unclear which linear model to choose as a benchmark case when comparing alternative methodologies to the linear model.

The contribution of this chapter is mainly of empirical nature. It is the first study providing an explicit comparison of the two estimators of this highly multivariate regression problem. Additionally, it also contributes by analyzing the impact of multiple testing corrections as recently proposed in the literature to curb the problem of a potential publication bias.

I find that the selection of firm-characteristics in the cross-section of expected returns varies depending on the regression method in the narrow case of the classical linear model. These differences can largely be traced back to discrepancies observed at the level of standard error estimation. The simulation-based evidence suggests that FM standard errors are biased downward in situations when T is relatively small. Additionally, this study shows that multiple testing adjustments, do not improve an investor's portfolio returns. A historical out-of-sample perspective indicates a slight performance deterioration.

Chapter 1

The (adaptive) lasso in the zoo - firm characteristic selection in the cross-section of expected returns

Marcial Messmer,
Francesco Audrino

1.1 Introduction

After years of strong growth in the number of published firm characteristics (FC) claiming to explain differences in average cross-sectional returns, some researchers have more recently shifted attention to the fundamental question of which statistical method to employ in selecting these variables; see for example, Harvey et al. (2016), McLean and Pontiff (2016) or Green et al. (2017).[1] Understanding these differences in cross-sectional returns has far-reaching implications for finance theory in general and consequently also for a vast part of the investment management industry. Hence, improving these methods is a requisite for future finance research. This work aims to contribute to the task by showing performance benefits of shrinkage methods combined with the application to the actual empirical problem at hand.

More generally, selecting variables and estimating coefficients is a common problem in finance and economics. Classically, multivariate regressions are run and insignificant coefficients are dropped. An important application in the context of selecting FC is the seminal contribution by Fama and French (1992), where variable selection is performed based on the multivariate regression framework and where insignificant coefficients are discarded. In particular, they regress cross-sectional returns on several firm characteristics to determine the crucial set of criteria explaining differences in returns. Based on this selection procedure, Fama and French (1993) form the well-known Fama-French (FF) three-factor model, which has set the benchmark and raised the bar for detecting new relevant FC. However, these estimates, mostly ordinary least squares (OLS), often suffer from a large variance and, hence, conclusions about the relevance of coefficients come potentially with a high degree of uncertainty.

Therefore, one of the main questions in empirical asset pricing that needs to be answered is which of the various stock characteristics provide independent and relevant information in explaining stock return variation. Despite the success of the FF three-factor model, a rich set of factors, anomalies or "significant" FC has grown and the need for an alternative or new methodological approach in the selection of relevant variables or firm characteristics has been felt. Or, as Cochrane (2011) puts it, "to address these questions in the zoo of new variables, I suspect we will have to use different methods." Statistically, it can be characterized as a variable selection and coefficient estimation problem, for which we suggest the application of the adaptive Lasso (Zou (2006)).

To overcome the high variance problem of classical linear methods, the machine learning literature has introduced alternative methods for variance reduction by tolerating a small bias. In an important contribution Tibshirani (1996) presents the least absolute shrinkage and selection operator (Lasso) method for estimating linear models. It simultaneously performs variable selection and coefficient estimation by shrinkage. However, some conditions required

[1] These firm characteristics are often introduced as factors or anomalies. We do not discriminate between these three and treat all equivalently throughout this paper. However, FC is the preferred wording.

for the Lasso for consistent variable selection prove too restrictive in many situations. To preserve the advantages of absolute shrinkage, Zou (2006) proposes a modified version, the so-called adaptive Lasso, such that consistent variable selection can be achieved even under less stringent conditions.

This study contributes to the literature mainly in the following three ways: First, we suggest the application of the adaptive Lasso in selecting firm characteristics because, as outlined below, the theoretical properties are appealing. Second, we develop an extensive Monte Carlo simulation to approximate a panel of cross-sectional returns. A distinct and novel feature here is the flexible simulation of highly-dimensional FC correlation matrices, which we borrow from the statistics literature. Third, this simulation design allows us to shed more light on the performance of Lasso methods in panels for various error specifications.

The chief goal of the paper is to answer the question of which firm characteristics are relevant in explaining differences in average returns. Secondly, the paper aims to determine whether the adaptive Lasso methods perform better than classical approaches; the simulation study sheds light on this. For the empirical part of our analysis, we focus on the US cross-section. We include 68 published firm characteristics constructed based on the CRSP/Compustat merged database with monthly data starting from 1974 until 2014.

It is important to note that we constrain our selection procedure to the linear model. Specifically, we want to perform a multivariate FC selection based on the character of the FC as originally proposed in the publications. Consequently, any non-linear effect is not of interest here. We document empirically, that univariate portfolio sorts and univariate linear regressions coincide with respect to the selection inference in almost all cases. Thus, the multivariate linear model is an appropriate method to address this joint selection problem. Hence, this study builds on the work of Green et al. (2017) and has to be differentiated from any non-linear statistical procedure. In particular, the authors digest a large set of FC in a linear multivariate Fama and MacBeth (1973) regression; we closely follow their data construction and FC pre-selection procedure. However, instead of relying on a multivariate regression, we employ with the adaptive Lasso a true variable selection technique.

The full pooled panel adaptive Lasso selection characterizes 37 FC of relevance for future differences in stock returns. Most consistent along size are the one month short-term reversal, research and development (R&D) spending scaled by market capitalization, the earning-to-price ratio and sales-to-price ratio. In a rolling regression, the change in the six-months momentum marks the only FC consistently selected in each period by the value-weighted adaptive Lasso. Generally, this contrasts with the findings of Green et al. (2017), as they identify a relatively low-dimensional linear cross-section.

Our simulation results show that the Lasso is inferior to the adaptive Lasso for all simulated cases. In particular, the results suggest that the strong conditions required for the Lasso to perform consistent variable selection are likely to be violated in the FC selection procedure. This holds in particular when high correlations among FC are observed, but even for cases where correlations are assumed to be of modest size, the Lasso struggles to identify irrelevant FC as too many FC get selected.

Whether the performance comparison between the adaptive Lasso and the OLS leads to a similar dominance is not as obvious, but most results suggest that the adaptive Lasso has advantages in almost all cases and specifications. The simulations show that the adaptive Lasso minimizes type II and type I error ratios. The latter holds true only if the FC are spurious and independent of the return process. However, FC linked to unpriced factors are too often selected to be priced (relative to OLS).

The paper is organized as follows. Section 1.2 reviews the relevant literature on cross-sectional returns and Lasso methods. This section is followed by a detailed description of the relevant methodology. Section 1.4 describes the estimation objective and how it relates to a factor structure. This section is followed by the simulation study. The data are briefly discussed in part 1.6. The penultimate section covers the empirical work, including the FC selection results. The final part concludes.

1.2 Related Literature

The first subsection provides an overview of the literature related to firm characteristics, stock market anomalies and methodologies applied in empirical asset pricing. The second part gives a summary of the relevant literature focusing on the main methodological contributions, mainly in the context of the Lasso.

1.2.1 FC and their relation to cross-sectional stock returns

Fama and French (1992) select relevant firm characteristics in the following way. In a first step, they run univariate regressions and keep significant coefficients; in a second step, they analyze the relevance of these variables in multivariate regressions. In their application, they find significant t-statistics for the coefficients of book-to-market (B/M) and market equity (ME), whereas total assets-to-market equity (A/ME) and asset-to-book equity (A/BE) as well as the earning-price (E/P) ratio are not significant in a multivariate analysis and, hence, are dropped.[2] Fama and French (1992) obtain the coefficient estimates and base their inference of the regressions by applying the Fama and MacBeth (1973) (FM) approach. The FM approach estimates at each point t a set of regression coefficients β_t ($r_{n,t} = \beta_t FC_{n,t}$ with

[2] To be precise, the independent variables B/M, ME, A/ME, A/BE are considered in terms of their natural logarithm.

$r_{n,t}$ the excess return and $FC_{n,t}$ the vector of FC of firm n at t). Statistical inference is then based on simple moment estimates of the time-series collection of β_t with $t = 1, ..., T$. Following the findings in Fama and French (1992), Fama and French (1993) present in a seminal contribution the Fama-French (FF) three-factor model, explaining expected return variation of the widely used 25-portfolios constructed by double sorts of B/M and ME. The three factors defined are: market, HML (high-minus-low B/M) and SMB (small- minus big market capitalization). Most subsequent studies condition their search for additional informative FC on these findings. Noteworthy is the insignificant coefficient for the market factor, which in economic terms implies the absence of a premium of equity market risk in cross-sectional returns. This aspect of their results had been documented in earlier studies and remains to this day an interesting topic of discussion in asset pricing; see Frazzini and Pedersen (2014) for a recent contribution.

Despite the higher hurdle imposed by the FF three-factor model, additional variables have been detected, which are helpful in explaining a bigger part of the variation in expected returns. Most prominent among them is momentum (Jegadeesh and Titman (1993)), where selling losers and buying winners evidently delivers a significant return premium (for a more recent overview see Asness et al. (2014)). Consequently, Carhart (1997) proposes a four-factor model, adding momentum to the FF three-factor model. Another example is profitability (Haugen and Baker (1996), Asness et al. (2013), Novy-Marx (2013)), where higher profitability leads to higher average returns. Moreover, the low volatility anomaly shows that stocks with high-historical idiosyncratic volatility tend to deliver exceedingly low returns on average (see Ang et al. (2006), Ang et al. (2009) and Blitz and van Vliet (2007)). A further example, presented in Frazzini and Pedersen (2014), is the low beta phenomenon which delivers superior risk-adjusted returns by buying leveraged low-beta assets and selling high-beta assets. The list of these FC could easily be extended here, but instead we refer the reader to Harvey et al. (2016), Green et al. (2017) and Goyal (2012), where a plethora of factors with references is provided.[3] Hence, a rich set of selection criteria can be defined, purely by selecting criteria applied in academic publications. The selection and construction of our FC mainly follows the work of Green et al. (2017).

Cochrane (2011) provides a comprehensive overview of research questions related, among other things, to firm characteristics and their functional relation to expected returns. In particular, he pays attention to the questions of how to deal with the high-dimensionality of characteristics so that no relevant information is left out. Central to the research agenda is the conditional expectation of excess returns, $E(R^e_{t+1}|C_t)$, and its linear interpretation in the form of the following panel regression:

$$R^e_{t+1,n} = x'_{t,n}\gamma + \epsilon_{t+1,n}$$

[3] See Tables A.3 and A.4 in the appendix for an overview of the firm characteristics included in this study.

where $x_{t,n} = [1\ C_{t,n}]$, for example, with $C_{t,n} = [size_{t,n}, bm_{t,n}, mom_{t,n}, ...]'$ the vector of firm characteristics at a given point in time $t = 1, ..., T$ for firm $n = 1, ..., N$ and $\epsilon_{t+1,n}$ the error.

Fama and French (2008) address the issue of selecting relevant variables by applying classical portfolio sorts and the FM approach to a more recent set of FC. Moreover, they summarize the methodological drawbacks of portfolios and regression approaches.[4] They note that portfolio sorts suffer particularly from the difficulty in disentangling effects in a multivariate context, and regressions might prove sensitive to micro-cap stocks or extreme outliers. However, according to Fama and French (2008) the concern is only of a theoretical nature in their set-up, as inferences with respect to regressions and portfolio sorts are empirically not in contradiction with each other.

The vast selection of FC allows Kogan and Tian (2013) to conduct a data-mining experiment of factor models. In their work, they exhaust all possible three- and four-factor models based on 27 selected stock characteristics. As a result, they construct a total of 351 possible three- and 2925 possible four-factor models. Then, tests on the performance of all factor models are conducted based on the ability to explain the cross-sectional variation of decile-portfolios from the remaining 25 or 24 firm characteristics, respectively. In their findings the authors report that it is relatively easy to form factor models which price a majority of the 25 (24) cross-sectional decile portfolios. In their conclusion they underscore the difficulty of evaluating empirical factor models, especially the difficulty of selecting a factor model based on the pricing performance. The results also show that the performance of the FF three-factor model is only average and this suggests that testing for significant alphas of new factors conditional on the FF three- (or any other arbitrary) factor model might not be the optimal testing procedure. Consequently, we do not condition the FC on any factor model.

Another approach to identifying relevant factors is followed by Harvey et al. (2016). The authors introduce the concepts of family-wise error and the false discovery rate to the finance literature. The procedure imposes higher hurdles in the form of higher t-values required to claim significance. The idea is justified based on the reasonable assumption of an existing "publication bias," implying a large fraction of published factors suffer from type I error. Applying the t-value adjustment reveals that many published factors would lose their status as a significant factor. The approach has the advantage that results are relatively cheap to obtain, requiring only a simple t-value adjustment. However, the method suffers from one shortcoming: it does not explicitly take into account the dependence structure of the FC. For this reason it is difficult to disentangle highly correlated firm characteristics, as they potentially capture a similar dimension of the return process.

McLean and Pontiff (2016) tackle the problem by applying "true" out-of-sample backtests of published factors. Two problems are involved. Factors are tested in isolation, ignoring

[4] Cochrane (2011) provides the theoretical relation of the two procedures in the appendix.

any correlations among the factors. Additionally, it requires some time after publication to perform the tests, as a certain amount of out-of-sample data needs to be accumulated to obtain any useful inference.

Linnainmaa and Roberts (2017) follow a similar strategy. However, instead of exploiting only the post-publication period, the authors collect data reaching back to 1918. The authors show a decrease in alpha, when measuring "out-of-sample" including these additional data-mining bias-free observations. The results are robust when purely conditioned on the pre-sample periods. As a result, seems to suggest that the factor is a product of data mining — as mispricing and unmodeled risk applies to the vintage sample as well.

The study of Green et al. (2017) approaches the problem of reducing the dimensionality of expected cross-sectional returns by including around 100 stock criteria. The authors utilize the FM approach in selecting characteristics. In their findings the authors reveal that 12 stock characteristics are priced in a multivariate FM-regression. Hence, their contribution supports the importance of analyzing the return process precisely in a multivariate setting, as most included FC vanish in their application.

The work of Freyberger et al. (2017) approaches the problem using non-parametric techniques. The authors apply the grouped adaptive Lasso to capture nonlinearities in estimating expected returns. Consequently, the estimated model is less parsimonious, while still achieving higher Sharpe ratios (SR) compared to an OLS approach. The study includes 36 FC, of which roughly one-third are informative.

Another contribution which makes use of shrinkage methods in the context of factor models is the paper by Bryzgalova (2016). It pays particular attention to problems arising from model mispecification. She introduces an alternative adaptive weighting scheme based on partial correlations instead of a two-stage procedure as compared with the adaptive Lasso. DeMiguel et al. (2017) analyze the FC selection from a portfolio perspective in a mean-variance (MV) framework. The analysis considers cases with and without transaction costs. Without transaction costs, 6 of 51 FC are reported to be of relevance; adding a term to capture transaction costs in the portfolio optimization procedure increases this number to 15.

Kozak et al. (2017) investigate the problem from a portfolio perspective in combination with L_2 and L_1 penalties. The authors identify a sparse set of FC. An MV optimized portfolio including 50 anomaly variables yields a CAPM alpha similar to the Fama and French (2014) five-factor model. Furthermore, the authors include two-dimensional interactions between these 50 FC and show a substantial increase in alpha of the MV portfolio compared to the case without interactions.

We skip a detailed review of methods used in (cross-sectional) asset pricing, referring instead to Goyal (2012). The study reviews the long list of empirical asset pricing methodologies in a comprehensive way. A more detailed introduction to estimation methods can be found in Cochrane (2005).

1.2.2 Lasso methods

The Lasso introduced by Tibshirani (1996) is motivated by the desire to improve OLS estimates without the shortcomings of subset selection and ridge regression. Tibshirani (1996) notes that subset selection suffers from high variability, as small data changes can cause subset selection to easily select a different model. Zou (2006) remarks that subset selection can become computationally infeasible if the number of variables is large. Ridge regression on the other hand has no obvious interpretation due to the fact that the coefficients are not set to zero.[5] The Lasso estimator optimizes least squares under, an additional condition involving the total sum of the absolute size of the coefficients (known as the ℓ_1-norm) that cannot be larger than a given tolerance value. The inclusion of a penalty term leads to consistent coefficient estimation and variable selection, if two necessary conditions are fulfilled, as Meinshausen (2006) shows. The necessary conditions are, namely, the neighborhood stability condition (also called the irrepresentable condition in some cases) and the beta-min condition. Generally, the first one is a concern whenever the matrix of independent variables displays a certain degree of collinearity. The beta-min condition requires some lower non-zero bound for the coefficients. A more formal treatment and discussion of the necessary conditions is provided in Bühlmann and Van De Geer (2011). As mentioned above, the conditions imposed are too restrictive in many empirical applications. Zou (2006) modifies the Lasso insofar that the weight of each coefficient in the penalization term is adaptive. This is achieved by scaling the absolute value of each coefficient with a first-stage estimator such that more highly relevant variables are less strongly affected by the penalty. Setting adaptive weights leads to consistent variable selection and coefficient estimation even if the neighborhood stability condition is not fulfilled.

The previously mentioned consistency properties are developed for a cross-sectional set-up with iid errors. Typically, the majority of applications in finance and macroeconomics require the use of time-series or panel data. Moreover, an iid error specification is more the exception than the rule. Consequently, Medeiros and Mendes (2012), Caner and Zhang (2014), Caner and Kock (2014), Kock and Callot (2015), Audrino and Camponovo (2015) and Kock (2016a,b) derive asymptotic properties of the Lasso and the adaptive Lasso in time series and panel settings.[6] In particular, Medeiros and Mendes (2012) and Audrino and Camponovo (2015) derive consistency properties of the adaptive Lasso in time series

[5] An overview of the family of linear methods can be found in Hastie et al. (2009)

[6] Additional literature references for the Lasso in panels are available in Kock (2016b).

environments. Medeiros and Mendes (2012) prove that the oracle properties of the adaptive Lasso are preserved for linear time series models even under non-Gaussian, conditionally heteroscedastic and time-dependent errors. Audrino and Camponovo (2015) show that the adaptive Lasso combines efficient parameter estimation, variable selection and valid finite sample inference for general time series regression models.

The Lasso and the adaptive Lasso have been applied to finance and economics. For example, Audrino and Knaus (2016) apply the Lasso in the field of volatility modeling. In particular, the authors prove that the Lasso has the ability to detect the heterogeneous autoregressive (HAR) model asymptotically, assuming it is the true model.[7] Based on their findings the HAR model is not detected in the underlying data set. Another interesting example can be found in Audrino and Camponovo (2015), where a Taylor rule model is selected by the adaptive Lasso in explaining the behavior of the short-rate set by the US central bank.

1.3 Methodology

This section introduces the notation and presents the underlying estimation methods in more depth. The section explains why the adaptive Lasso is a more useful method for our application than the Lasso, based on the requirements of the methods. Finally, we show why shrinkage approaches are promising for improving statistical inference when selecting relevant FC, based on the mean variance decomposition.

1.3.1 Notation

Generally, if not explicitly otherwise stated, we follow the notation that n refers to stock n of N_t total stocks and t to period (i.e. month) t of T total periods. Moreover, firm characteristics are indexed by c of C total characteristics and belong to set $\mathcal{C}$, where each c belongs to one of the following three groups or subsets: FC measuring factor exposure with priced risk are denoted by p (of a total P) and define the set $\mathcal{P}$; FC measuring sensitivity to an unpriced factor are defined by u (of a total U, set $\mathcal{U}$), and FC independent or spurious with respect to the return process are described by s (of a total S, set $\mathcal{S}$). The total number of FC, $P+U+S=C$. The specification of FC for the simulation is outlined in more detail in section 1.5. Moreover, return-related parameters such as risk-premia (μ) and volatility (σ) are expressed in annualized terms despite the fact that the data frequency is monthly.[8]

[7] The HAR model describes the dynamics of realized variance as a process of past realized variances; see Corsi (2009) for details.

[8] We slightly deviate in this section and denote the regression coeffecient as β (vs. γ). In all other sections we use the term β exclusively as measure of factor exposure, and γ as the regression coefficient we aim to estimate.

1.3.2 Methods

The linear model can be defined in matrix- and vector-notation as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{1.1}$$

where $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$ and the corresponding response vector $\mathbf{Y}_{n\times 1}$, the design matrix $\mathbf{X}_{n\times p}$, the parameter vector $\beta_{p\times 1}$ and a vector of residuals $\epsilon_{n\times 1}$. Throughout this work we treat $\mathbf{Y}$ and $\mathbf{X}$ as standardized matrices, with $\mu = 0$ and $\sigma = 1$, where the standardization is applied column by column. Furthermore, let the true set of non-zero coefficients be denoted by $\mathcal{A} = \{j : \beta_j \neq 0\}$ and let a measure of sparsity be defined as $\|\beta\|_0^0 = \sum_{j=1}^p |\beta_j|^0$, which counts all non-zero elements. Generally, we assume for (1.1) that $\|\beta\|_0^0 = q \leq p$.

The **ordinary least squares** (OLS) estimator of (1.1) is defined as

$$\hat{\beta}_{\text{ols}} = \arg\min_{\beta} \left(\|\mathbf{Y} - \mathbf{X}\beta\|_2^2/n \right), \tag{1.2}$$

where $\|\mathbf{Y} - \mathbf{X}\beta\|_2^2 = \sum_t^T (Y_t - X_t\beta)^2$. Moreover, it holds that $\|\hat{\beta}_{\text{ols}}\|_0^0 = p$.

The **Lasso** estimator (Tibshirani (1996)) is then defined by:

$$\hat{\beta}_{\text{Lasso}}(\lambda) = \arg\min_{\beta} \left(\|\mathbf{Y} - \mathbf{X}\beta\|_2^2/n + \lambda\|\beta\|_1 \right), \tag{1.3}$$

where $\lambda \geq 0$ is a penalty parameter and $\|\beta\|_1 = \sum_j^p |\beta_j|$. The penalty term λ used in (1.3) is determined by cross-validation (CV) or classical selection criteria, typically five-fold or ten-fold CV, the Bayesian information criterion (BIC), or the Akaike information criterion (AIC). If the neighborhood stability and the beta-min condition are fulfilled, the Lasso is a consistent variable selection method since $\lim_n \mathbb{P}(\hat{\mathcal{A}}_{\text{Lasso}} = \mathcal{A}) \to 1$ and, hence, $\|\hat{\beta}_{\text{Lasso}}\|_0^0 = q$. The technical details of the two conditions are presented in the appendix.

The **adaptive Lasso** (Zou (2006)) differs in terms of the penalization term, which allows the weights to vary for each parameter. The assigned individual weights are inversely proportional to a first stage β estimate. Algebraically, we can express the estimator as

$$\hat{\beta}_{\text{adapt}}(\lambda) = \arg\min_{\beta} \left(\|\mathbf{Y} - \mathbf{X}\beta\|_2^2/n + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\hat{\beta}_{\text{init},j}|} \right) \tag{1.4}$$

with $\hat{\beta}_{\text{init}}$ determined in a first stage. Zou (2006) suggests the use of the OLS estimator, $\hat{\beta}_{\text{ols}}$ as $\hat{\beta}_{\text{init}}$, unless collinearity is an issue. Bühlmann and Van De Geer (2011) set $\hat{\beta}_{\text{init}} = \hat{\beta}_{\text{Lasso}}(\lambda)$ and again use CV to determine λ at the second stage. The use of the Lasso as a first-stage estimator is justified by the screening property of the Lasso ($\mathbb{P}(\hat{\mathcal{A}}_{\text{Lasso}} \supseteq \mathcal{A}) \rightarrow 1$, where $\hat{\mathcal{A}}_{\text{Lasso}}$ is the estimated set of active variables by the Lasso), which still allows consistent variable selection of the adaptive Lasso at the second stage ($\mathbb{P}(\hat{\mathcal{A}}_{\text{adapt}} = \mathcal{A}) \rightarrow 1$).

The computational cost of the adaptive Lasso solution is of the same order as that of a single OLS fit; hence, the application of the adaptive Lasso is computationally tractable. Zou (2006) presents a least-angle regression (LARS) algorithm (Efron et al. (2004)) for the adaptive Lasso. The solution path of the adaptive Lasso can then be displayed in dependence on the selection parameter λ at the second stage.

Moreover, the adaptive Lasso enjoys under relaxed conditions the so-called oracle properties, which are:

1. Identification of the right subset: $\lim_n P(\hat{\mathcal{A}}_{\text{adapt},n} = \mathcal{A}) = 1$, where $\mathcal{A}$ is as defined above.
2. Optimal estimation rate (similar to the OLS estimator):
   $\sqrt{N}(\hat{\beta}_{\text{adapt},\mathcal{A}} - \beta_{\mathcal{A}}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathcal{A}})$, with $\Sigma_{\mathcal{A}}$ the covariance matrix of the true subset.

Bühlmann and Van De Geer (2011) show that the optimal $\hat{\lambda}$ based on the BIC evaluation reads as follows:

$$\hat{\lambda}_{BIC} = \arg\min_{\lambda} \left(n \log \left(\frac{\|\mathbf{Y} - \hat{Y}_\lambda\|^2}{n} \right) + \log(n) \|\hat{\beta}_\lambda\|_0^0 \right),$$

and accordingly the AIC,

$$\hat{\lambda}_{AIC} = \arg\min_{\lambda} \left(n \log \left(\frac{\|\mathbf{Y} - \hat{Y}_\lambda\|^2}{n} \right) + 2 \|\hat{\beta}_\lambda\|_0^0 \right).$$

Alternatively, the optimal λ can be estimated by cross-validation. Here we randomly split the samples along time points and never within a given period. Assume we observe T periods ($t = 1, ..., T$) with each containing N stocks $S_{1,t}$, $S_{2,t}$,..., $S_{N,t}$. Consequently, we can randomly select a training and testing set along the time index t. Hence, high cross-sectional correlations cannot cause biased estimates for the optimal λ. Following Hastie et al. (2009) and assuming that we have a functional relation of Y and X as described in (1.1), we can express the OLS bias-variance decomposition for squared error loss for a single stock as follows:

$$\begin{aligned}
\text{Err}(x_0) &= \mathbb{E}[(Y - X'\hat{\beta})^2 | X = x_0] \\
&= \underbrace{\sigma_\epsilon^2}_{\text{Irreducible Error}} + \underbrace{(\mathbb{E}[x_0'\hat{\beta}] - x_0'\beta)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}[x_0'\hat{\beta} - \mathbb{E}[x_0'\hat{\beta}]]^2}_{\text{Variance}}
\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T} \text{Err}(x_t) = \sigma_\epsilon^2 + \frac{1}{T}\sum_{t=1}^{T}(x_t'(\hat{\beta} - \beta))^2 + \frac{p}{T}\sigma_\epsilon^2. \tag{1.5}$$

This decomposition exhibits the tradeoff faced between bias and variance. The influence of the bias increases with larger $|\hat{\beta} - \beta|$; on the other hand, the variance term gains in importance, as p and/or σ_ϵ^2 rises.

Given that shrinkage methods like the ones introduced above have the ability to reduce the variance at the cost of slightly increasing the bias, equation 1.5 reflects the motivation for applying the shrinkage methods to the empirical set-up presented above. First, the ratio of $\frac{p}{T}$ can potentially be high, as we have 400+ presented factors in the literature and in the best case 50 years of monthly data ($T = 600$). Moreover, if some FC are available only for a shorter period of time, we can still perform the regression, as the Lasso methods are feasible even for the case where we have a truly high-dimensional problem ($p > T$), which imposes a constraint for classical OLS. Moreover, the noise component makes up unambiguously a significant proportion of the return process (even when assuming that the efficient market hypothesis is violated). Hence, the noise variance component has an important impact.

1.4 FC and the return generating process

Generally, we assume a Rosenberg (1974) and Daniel and Titman (1997) type cross-sectional return structure. Covariances are determined based on a factor structure and expected returns mark a compensation for factor risk (default assumption). Following Daniel and Titman (1997), we consider FC and their dependence on expected returns in three specifications. Namely,

1.4.1 Model 1

$$r_{n,t} = r_{\mathrm{rf},t} + \underbrace{\beta_{n,t}' f_t}_{\beta_{n,t}' \tilde{f}_t + \beta_{n,t}' \mu} + \eta_{t,n}$$

with $f_t = \mu + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, \Sigma)$, $\tilde{f}_t$ captures only the variance component of the factor and is therefore equivalent to ϵ and has zero mean, μ defines the risk-premia of the factors. Additionally to the factor structure, each stock has its own idiosyncratic component, $\eta_{t,n}$, orthogonal to other factors and other stocks' idiosyncratic component.

1.4.2 Model 2

The model is identical to Model 1 except that it is equipped with a time-varying μ_t, instead of the time constant μ as before.[9]

1.4.3 Model 3

$$r_{n,t} = r_{\text{rf},t} + \beta_{n,t}' \tilde{f}_t + x_{n,t-1}' \delta + \eta_{t,n},$$

where δ measures the sensitivity of FC to expected returns that is not directly compensating for factor risk. Furthermore, model 3 imposes a non-zero covariance between the FC ($x_{n,t-1}$) and some factor in $\tilde{f}_t$, in case the FC are linked to the non-zero elements in δ as described in Daniel and Titman (1997). This model implies asymptotic arbitrage.

1.4.4 Estimation objective

Moreover, we assume a linear functional relation of the FC and factor exposure, $\beta_{n,t} = g(x_{n,t-1})$, i.e.,

$$g(x_{n,t-1}) = a + b' x_{n,t-1} \tag{1.6}$$

for Models 1 and 2. Notice, that a and $r_{\text{rf},t}$ (the risk-free rate) cancels, once we consider de-meaned cross-sectional returns. The linear dependence we aim to measure is of the form $\mathbb{E}[R^e_{t,n}] = x_{n,t-1}' \gamma$. The interpretation of the coefficient is according to Models 1 and 2, $\gamma = b\mu$, in case of Model 3, $\gamma = \delta$.

1.5 Simulation study

In order to analyze the suitability of the adaptive Lasso as selection operator in the previously described context, we propose a simulation of cross-sectional returns and firm characteristics. The simulation is calibrated such that crucial properties of the cross-sectional return data are satisfied. However, we stress that the simulation is highly stylized and cannot be a perfect replication of the underlying data-generating process (DGP). Nonetheless, it is helpful and necessary for gaining insights into model selection performance under different distributional assumptions in our specific setting; see the literature review in Section 1.2.2 on what has

[9]Notice, that this case can collapse to the first model from a statistical perspective. For example, assume the following time-varying process: $\mu_t = \bar{\mu} + \kappa_t$, and, $\kappa_t \sim \mathcal{N}(0, \sigma_\kappa)$. Any variation in the risk-premia would then be absorbed by the error term and cannot be distinguished from it. In a panel regression, we would then simply estimate $\overline{\mu} = \sum_{t=1}^{T} \mu_t$.

already been proved theoretically. Hence, the simulation study includes several specifications to analyze the method's sensitivity to these assumptions.

1.5.1 Calibration

In addition to the aforementioned structure and to establish a base case, we calibrate the simulation of the return-generating process as follows:

- We set the number of priced factors to 6. Additionally, we have 6 factors representing unpriced risk. Even if it were theoretically assumed that unpriced factor risk should not exist, it could empirically still be present. For example, this may be the case if not all factor-exposed FC are fully reflected on the right hand side of the regression equation. Moreover, we assume that the factors are independent of one another and that the stock market factor explains the highest proportion of variance of all priced factors.
- Firm characteristics fall in one of the following three categories/groups: Category 1 measures factor exposure to priced factors and category 2 to unpriced factors. Group 3 measures FC which are independent of the return-generating process. Empirical examples for category 1 are: book-to-market, measuring exposure to value; for 2: the sector classification, measuring sector risks; and for 3: the last letter of the company name, (presumably) irrelevant for the DGP. FC are potentially correlated across groups.
- The signal-to-noise ratio is relatively small, assuming a ratio implying a yearly R^2 of 5% (see the appendix for details on the transformation to monthly R^2s). This is in line with empirically documented R^2's in the case of the linear model; see, for example, Lewellen (2015).
- The stock market factor follows a time-varying volatility process (implying heteroskedasticity for the individual stocks over time as well).
- The simulated return series do not possess any auto-correlation.

Algebraically, we can achieve the desired properties with the following specification of the return-generating process with the following equation (with $b=1$ in equation 1.6):

$$r_{t+1,n} = x'_{t,n} f_{t+1} + \eta_{t+1,n}, \tag{1.7}$$

- f_t defines the vector of factor returns of length C. Each f_t is determined by: $f_{t+1} = \mu + \epsilon_{t+1}$ with $\epsilon_{t+1,c} \sim \mathcal{N}(0, \sigma_f^2)\ \forall c \in \mathcal{P} \cup \mathcal{S} \wedge c \neq 0$ and $\epsilon_{t+1,0} \sim \mathcal{N}(0, \sigma_{t,0}^2)$. FC, $c = 0$,

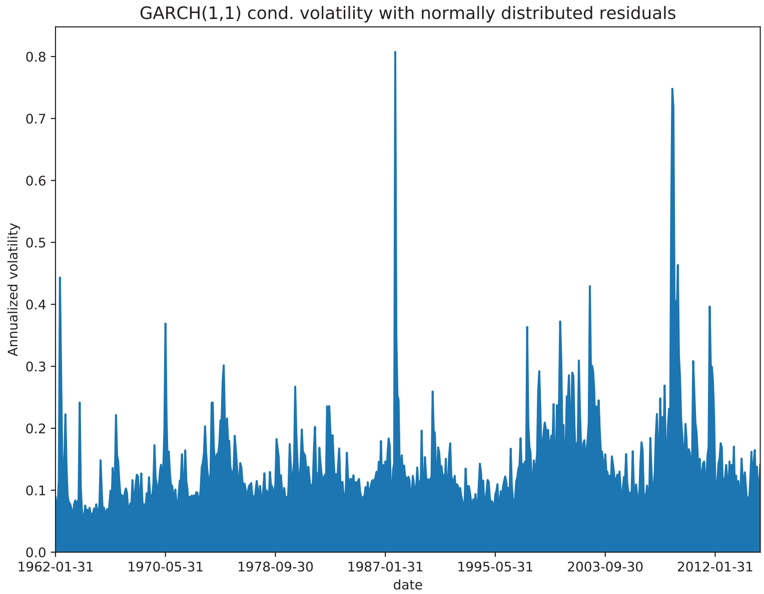

Figure 1.1: The figure shows the estimated GARCH(1,1) process used in the simulation, where $\sigma_t^2 = \alpha + \beta_0\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$ The parameters estimated with normal distributed errors are: $\hat{\alpha} = 0.0079$, $\hat{\beta}_0 = 0.093$ and $\hat{\beta}_1 = 0.9019$. The model is estimated based on daily market factor returns with data from 1962-2014. We observe time-varying stock market volatility. The graph indicates periods of high volatility, for example, the stock market crash in 1987, the sharp decline of stock prices in the early 2000s and the financial crises during the years 2008-09.

represents the stock market return, which follows a latent volatility process $\sigma_{t,0}^2$ (see next item). $\epsilon_{t+1,c}$ is set to 0 $\forall c \in \mathcal{S}$ (spurious FC). The elements of the vector of risk-premia, μ_c, are drawn from $\sim \text{unif}(0.1, 3.5) \forall c \in \mathcal{P} \wedge c \neq 0$ and are set equal to zero $\forall c \in \mathcal{U} \cup \mathcal{S}$. We assume a stock market premium, μ_0, of 5.5% per year. [10]

- $\sigma_{t,0}^2$, the stock market volatility, is estimated by using a GARCH(1,1) process, where the estimated $\hat{\sigma}_t$ are obtained by fitting a GARCH(1,1) model on daily observed US stock market returns over the previous T periods ending on 2014-12-31.[11] Figure 1.1 displays the volatility estimates over time; the estimated coefficients can be found in its caption.

- $\eta_{t,i}$ represents the idiosyncratic stock market component and is drawn from $\sim \mathcal{N}(0, \sigma_{idio}^2)$. Bekaert et al. (2012) show that aggregated idiosyncratic volatility varies over time. Despite this empirical evidence, we choose a parsimonious approach to model idiosyncratic volatility. This is mainly motivated by the statistical properties our DGP already possesses.

- $x_{t,n}$ marks the vector of firm characteristics of stock n at t of length C. We simulate

[10]The risk-premia are drawn only once per case and are kept constant through each simulation. The market premium is the estimate of the Fama French market factor from 1962-2014.

[11]The GARCH(1,1) model captures a large fraction of distribution properties observed in stock returns. Moreover, model performance seems reasonable compared to less parsimonious approaches (see Hansen and Lunde (2005)).

the characteristics x_t from $\sim \mathcal{N}(0, \Sigma)$.

- The correlation matrix of FC, Σ, is obtained following the simulation approach of Hardin et al. (2013) and is a crucial feature of our simulation. It is important, as many FC measure empirically similar variations. The base case refers to the constant correlation structure within groups (algorithm 1) in Hardin et al. (2013). The Σ is drawn only initially and kept constant in each specification. The empirical correlation structure described in Section 1.7.1 shows a handful of cases with pairwise correlations around 0.9 and many between 0.4-0.5. Our simulated correlation pairs reflect this, in order to investigate the impact of this difference. However, this study does not consider a correlation grouping of more than two FC, or any other forms of more involved linear dependencies.

Finally, the collection of all T periods of the simulations can then be stacked together in matrix X and matrix Y. The true coefficients of interest is the vector of μ, which is estimated as the vector of coefficients $\hat{\gamma}$.

The specification allows flexible simulations under different assumptions, analyzing the sensitivity of simulation parameters on performance of the method. For this, we perform several different simulation specifications, each loosening one assumption separately. In general, the following default parameters are set:

- Number of simulations, 2000
- Number of periods, 600, corresponding to 50 years of monthly data
- Number of stocks, 4000 [12]
- Number of firm characteristics, 100 – with $P = 6$, $U = 6$ and $S = 88$
- σ_f^2 and σ_{idio}^2 result from the pre-specified level of R^2, where the noise variance is distributed as described in the appendix.
- The correlation matrix, Σ, is simulated such that we have one high (0.9) and one low (0.4) pairwise correlation between FC from each group. See Figure 1.2 for a visualization of one realization of the specified correlation matrix simulation.

1.5.2 Sensitivity analysis

The behavior of the simulated DGP crucially depends on its calibration defined above. In order to investigate the sensitivity to these choices, we define the following cases:

[12] 4000 corresponds to the average number of stocks used in the empirical part.

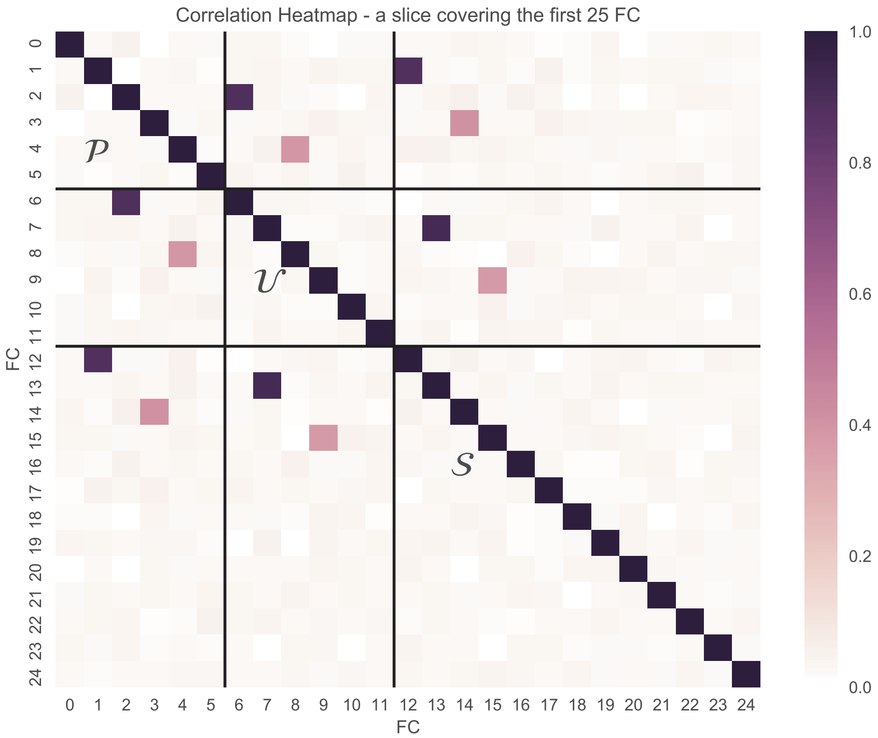

Figure 1.2: The figure displays a fraction of the default correlation matrix used for the simulations. It shows the correlation for the first 25 FC, where FC 0 to 5 refer to the set of FC with positive risk-premia, 6-11 belong to $\mathcal{R}$ and the rest to $\mathcal{U}$. For example, FC 1 (part of $\mathcal{P}$) and FC 12 (part of $\mathcal{U}$) are correlated with each other with a correlation of around 0.9. Note that the missing 75 FC are uninteresting insofar as their off-diagonal elements are close to zero.

Case 1: Base case

Default settings.

Case 2: Time-constant stock market volatility

Default settings, except for the assumption of the underlying latent volatility process of $\hat{\sigma}^2_{t,0}$, which we fix for all t to the long-term volatility estimate of the US stock of 15.8 %.

Case 3: t-distributed stock market returns

Default settings, except $\epsilon_{t+1,0} \sim t$ - with the GARCH(1,1) estimation also based on the t-distributed errors, with an estimated number of degrees of freedom, $\hat{\nu}$, of 7.14. However, differences are not overly strong.

Case 4: Small T

Default settings, T is set to 240.

Case 5: Large T

Default settings, T is set to 4200 (and N reduced to 1500 to keep it computationally tractable). This specification requires a longer than available estimated GARCH(1,1) series; the missing $\sigma^2_{t,1}$ are simply simulated based on the GARCH(1,1) parameter estimates described in Figure 1.1.[13]

Case 6: Model 2 — Time varying risk-premia
Default settings, but with time-varying $\mu_{t,0}$ for the stock market factor. We impose that, $\mathbb{E}[\mu_{t,0}]$ is equivalent to μ_0. The stock market risk-premia is replaced with the following AR(1) process:

$$\mu_{t,c} = c_c + \varphi\mu_{t-1,c} + \psi_t,$$

the constant, c_c, is set to $\mu_c(1-\varphi) = c_c$ for the mean constraint to hold and the noise component, $\psi_t \sim \mathcal{N}(0, \sigma^2_{\mu_0})$ and $\varphi = 0.2$. The size of the standard error, σ_{μ_0}, is set such that it absorbs 1/5 of the unconditional variance of the associated stock market factor variance (which is reduced accordingly). The errors are independent of each other.[14]

Case 7: Model 3 — Expected returns: a function of FC instead of factor exposure
Default settings. The premium of the stock market factor is set to zero. Instead, we attach the premium to the FC directly, and impose a correlation of the factor exposure and the FC of 0.9.[15]

Note that each simulation considers a balanced panel. As the actual data consist of an unbalanced panel, we adjust the data as described in Section 1.7.1. Moreover, the simulation ignores potential measurement errors in FC and assumes that they are measured without errors. Empirically, the most common FC suffering from measurement error are, as mentioned above, market betas.

1.5.3 Performance evaluation

After each data simulation is performed, we run seven regressions covering all three methods presented above in the pooled panel framework. Next to OLS, we provide estimates for Lasso- and adaptive Lasso- based BIC, AIC and five-fold cross-validation (CV5) optimized regularization strength. We collect the results of the OLS estimates, the Lasso and the adaptive Lasso. First, we need to set a significance level for the OLS estimates to have a rule determining whether or not a coefficient can be seen as selected — we set the level to the

[13] This case is not a realistic scenario for monthly data, but is insightful for applications with higher frequencies.

[14] $\text{var}(\mu_t) = \frac{\sigma^2_\varepsilon}{1-\varphi^2}$

[15] The label "market factor" might be misleading in this specification, since we decouple the FC (in this case the market beta) from the factor loading which actually is the market beta. As it is just an econometric exercise, we can ignore this label problem.

literature standard of 5%. The t-values are based on the Driscoll and Kraay (1998) robust standard-errors. In case of the Lasso and the adaptive Lasso, the FC selection procedure is straightforward: all non-zero coefficient estimates are considered to be selected. We can then simply calculate the ratios of correctly classified coefficients, providing insights on type I and type II error behavior. Type II errors are calculated for FC 0-5 — a failure to reject the null hypothesis of an unpriced factor FC given the factor is truly priced. Type I errors are measured for FC 6-100 — rejecting the null hypothesis of an unpriced factor exposure given the FC is tied to an unpriced factor (FC 6-11) or the FC is independent of the returns (FC 12-100).

1.5.4 Results

Case 1: Default settings

The simulation results of case 1 are exhibited in Figure 1.3 and in Table 1.1. The results reveal that there are distinct differences between the methods applied. It shows that for the simulated stock market factor, OLS performs the worst, as it displays a type II error rate of over 30%. On the other hand, the Lasso and the adaptive Lasso methods reveal a far better performance with an error rate of around 0% and 5% respectively for the stock market factor (in case of CV5). For the other, significantly less volatile factors carrying a positive risk-premium, the type II error rates are indistinguishable from zero for almost all cases and all three methods. The exceptions are one OLS estimate and all Lasso- and adaptive Lasso-selected FC based on CV5. The type I error cases have to be distinguished in two cases: First, FC belonging to set $\mathcal{U}$ (unpriced factor FC) where OLS has slight advantages over the adaptive Lasso. The Lasso reveals a poor performance, in particular for the case where the correlation is assumed to be .9 (FC #6). This case appears to be an issue as measured by an error rate of well above 80%. The second case of type I errors comprises spurious FC. That the uninformative FC are correlated with any of the two other FC types proves once more to be a problem for the Lasso. For these cases the more restrictive adaptive Lasso reveals a far better selection performance than the Lasso, as the error rate is zero for all cases. Moreover, the OLS type I error rate behaves as expected by varying around the 5% significance level.

Case 2: Time constant stock market volatility

Table 1.1 reveals that assuming homoscedastic errors for the stock market factor only marginally affects the error ratios. The only difference we observe is a slight drop in the error ratio for the stock market factor for the OLS approach.

Case 3: T-distributed stock market returns

Using draws from a Student-t distribution instead of the Normal with corresponding GARCH-(1,1) volatility estimates for the market factor does not have an impact on the performance

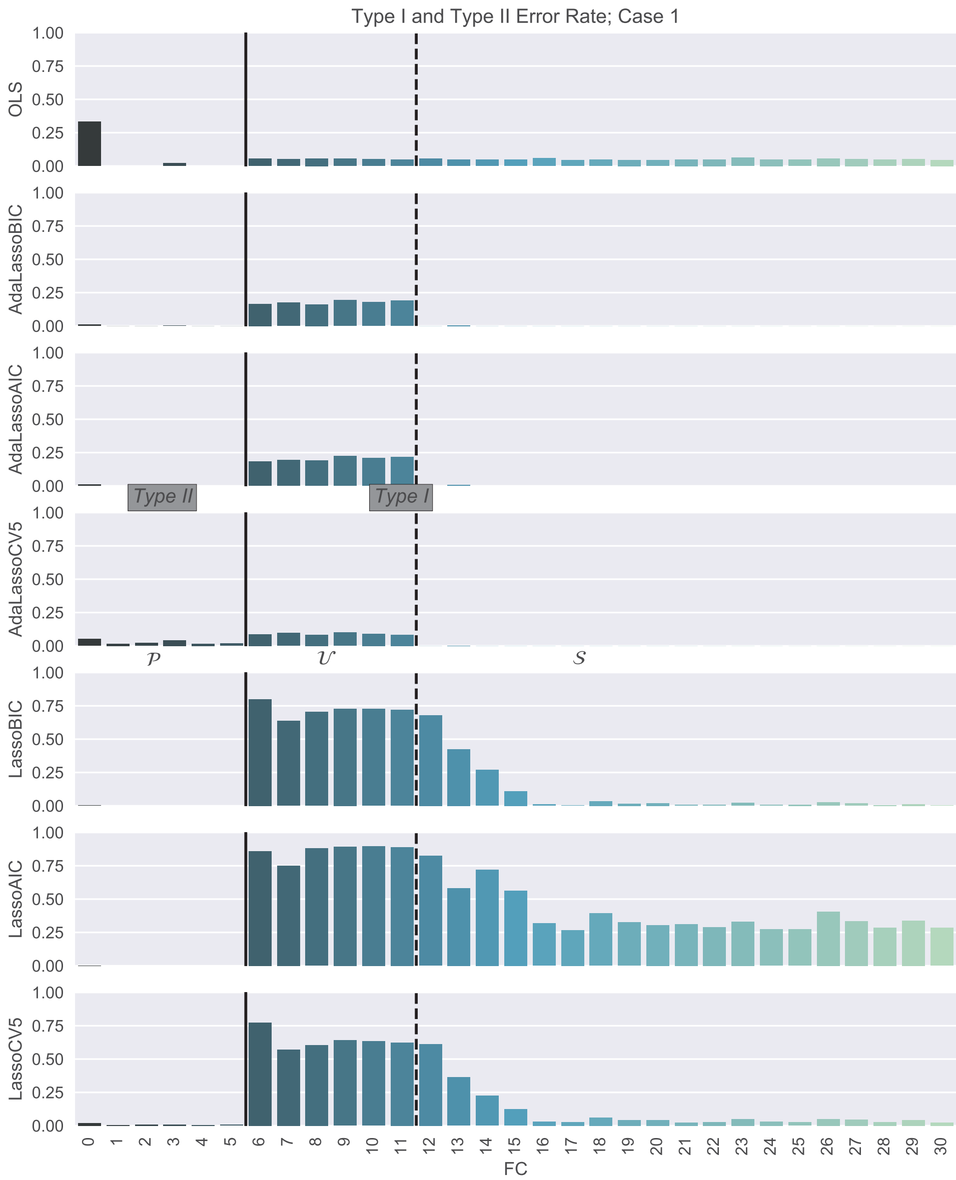

Figure 1.3: The figure displays type I and type II errors for the OLS, adaptive Lasso and Lasso selection procedure. Type II errors are for FC number 0 to 5, and the rest of the FC 6-30 correspond to type I errors. The first sub-figure shows the results of the OLS selection procedure. The following three charts in the middle illustrate the adaptive Lasso error ratios. The last three subplots on the bottom exhibit the Lasso error behavior.

behavior of the three different methods.

Case 4: Small T

Reducing the number of periods in the simulation reveals differences in the performance compared to the default simulation results, as Table 1.1 shows. First, type II error rates rise strongly for OLS estimates, whereas only a slight increase is observed for the adaptive Lasso and practically no changes are visible for the Lasso. The picture also changes when looking at type I error rates for FC in set $\mathcal{U}$, where we observe a jump in error ratios for the adaptive Lasso.

Case 5: Large T
As T grows larger, the error ratios decline as expected. The only remarkable exception is found for the Lasso, where once more the correlated FC remain prone to false inference. A strong indication that the neighborhood stability condition is likely to be violated in cases of higher correlations is that the error rate of FC 6 is 99% and that of FC 12 reaches 84%. Moreover, higher error ratios are also observed for cases with weaker correlations of around .5; see FC cases 7-9 and 12-15.

Case 6: Model 2 — Time varying risk-premia
Assuming an AR(1) process for the mean component of the stock market factor does not influence the inference by much. We document only marginal changes for all methods in this case for FC 0.

Case 7: Model 3 — Direct linear dependence between FC and expected returns
Model 3 yields a performance improvement for all methods, most apparent for OLS, where the classification error comes down from 33% to 23% compared to the default assumption.

Finally, we **briefly summarize** the simulation results. We show that the adaptive Lasso is superior to OLS when type II errors are a concern. A Lasso-based selection reveals for this case only negligible advantages over the adaptive Lasso. The picture changes if we want to minimize type I error behavior. Here we have to differentiate between two distinct scenarios. First, whenever we encounter an entirely uninformative independent variable, we show that the adaptive Lasso performs best. Second, in case we have a relation of the independent variable with an unpriced risk factor of the dependent variable, an OLS approach achieves the best results. We note that correlations are a crucial driver behind these results, where in particular, the Lasso presents problems reaching reasonable type I error ratios when confronted with higher correlations (≈ 0.9). Moreover, altering the optimal λ selection mechanism impacts the results importantly. BIC is favorable over AIC in the specifications under consideration. BIC reduces type I errors without suffering from an increase in type II misclassifications. BIC vs. CV exposes a tradeoff between type I and type II. Assigning equal weight to both error types, BIC-based estimation is the preferable tuning method.

#	Method	$\mathcal{P}$ — Type II						$\mathcal{U}$ — Type I						$\mathcal{S}$ — Type I				Summary FC 16-99			
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	mean	std	min	max
1	OLS	0.33		0.01	0.02		0.00	0.06	0.06	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.04	0.05	0.00	0.04	0.06
	AL BIC	0.02		0.01	0.01			0.18	0.10	0.10	0.10	0.12	0.11	0.00	0.03	0.00	0.00	0.00	0.00		0.00
	AL AIC	0.01		0.00	0.00			0.23	0.16	0.17	0.19	0.20	0.19	0.01	0.05	0.00	0.00	0.00	0.00		0.00
	AL CV5	0.06	0.02	0.03	0.05	0.02	0.02	0.14	0.07	0.07	0.08	0.10	0.08	0.00	0.02			0.00	0.00		0.00
	L BIC	0.01			0.00			0.67	0.30	0.31	0.33	0.34	0.33	0.37	0.13	0.06	0.02	0.01	0.00	0.00	0.02
	L AIC	0.00						0.76	0.55	0.64	0.66	0.65	0.66	0.54	0.32	0.34	0.25	0.23	0.01	0.21	0.27
	L CV5	0.03	0.01	0.01	0.01	0.01	0.01	0.73	0.46	0.50	0.52	0.52	0.52	0.48	0.24	0.21	0.13	0.10	0.01	0.08	0.13
2	OLS	0.34			0.01			0.06	0.05	0.06	0.05	0.06	0.05	0.04	0.05	0.05	0.05	0.05	0.00	0.04	0.07
	AL BIC	0.01			0.00			0.16	0.17	0.16	0.17	0.18	0.19		0.00						
	AL AIC	0.01			0.00			0.17	0.18	0.19	0.20	0.22	0.22		0.00						
	AL CV5	0.06	0.02	0.02	0.04	0.02	0.02	0.08	0.09	0.07	0.09	0.09	0.08		0.00						
	L BIC	0.00						0.81	0.65	0.72	0.74	0.74	0.76	0.71	0.44	0.28	0.12	0.01	0.01	0.00	0.06
	L AIC	0.00						0.87	0.77	0.88	0.90	0.90	0.91	0.85	0.58	0.73	0.56	0.32	0.04	0.28	0.47
	L CV5	0.02	0.01	0.01	0.01	0.01	0.01	0.78	0.57	0.62	0.64	0.63	0.66	0.62	0.37	0.24	0.13	0.04	0.01	0.02	0.09
3	OLS	0.33			0.02			0.06	0.06	0.06	0.05	0.05	0.06	0.04	0.05	0.05	0.05	0.05	0.01	0.03	0.06
	AL BIC	0.01			0.00			0.15	0.19	0.18	0.18	0.17	0.21		0.01						
	AL AIC	0.01			0.00			0.17	0.21	0.21	0.21	0.20	0.24		0.01						
	AL CV5	0.05	0.02	0.02	0.04	0.02	0.02	0.08	0.11	0.09	0.09	0.08	0.12		0.00						
	L BIC	0.00						0.81	0.66	0.71	0.73	0.73	0.76	0.69	0.44	0.26	0.11	0.01	0.01	0.00	0.04
	L AIC	0.00						0.86	0.77	0.89	0.89	0.88	0.91	0.83	0.57	0.71	0.56	0.31	0.04	0.27	0.44
	L CV5	0.03	0.01	0.01	0.01	0.01	0.01	0.77	0.59	0.62	0.63	0.63	0.66	0.61	0.37	0.24	0.13	0.04	0.01	0.02	0.08
4	OLS	0.75	0.00	0.06	0.26	0.00	0.04	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.01	0.04	0.06
	AL BIC	0.04		0.02	0.04		0.00	0.27	0.33	0.33	0.36	0.35	0.35		0.01						
	AL AIC	0.04		0.02	0.04		0.00	0.29	0.37	0.39	0.43	0.41	0.41		0.02						
	AL CV5	0.49	0.42	0.47	0.51	0.42	0.44	0.11	0.10	0.09	0.10	0.11	0.10		0.00						
	L BIC	0.02		0.00	0.02		0.00	0.68	0.64	0.69	0.72	0.70	0.71	0.51	0.29	0.12	0.05	0.01	0.00	0.00	0.03
	L AIC	0.01		0.00	0.01		0.00	0.76	0.78	0.87	0.86	0.87	0.87	0.65	0.45	0.49	0.36	0.27	0.02	0.24	0.34
	L CV5	0.33	0.30	0.31	0.33	0.30	0.30	0.47	0.38	0.40	0.42	0.41	0.41	0.31	0.17	0.09	0.05	0.02	0.00	0.01	0.03
5	OLS							0.05	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.00	0.04	0.06
	AL BIC							0.01	0.00	0.00	0.00	0.00	0.00								
	AL AIC							0.01	0.00	0.00	0.00	0.00	0.00								
	AL CV5							0.00	0.00	0.00	0.00		0.00								
	L BIC							0.99	0.42	0.55	0.54	0.54	0.54	0.84	0.66	0.48	0.21	0.01	0.01	0.00	0.07
	L AIC							1.00	0.56	0.83	0.81	0.83	0.84	0.94	0.84	0.92	0.81	0.37	0.08	0.29	0.62
	L CV5							0.99	0.50	0.69	0.68	0.70	0.70	0.91	0.75	0.78	0.54	0.14	0.06	0.09	0.36
6	OLS	0.32			0.01			0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.00	0.04	0.06
	AL BIC	0.01			0.00			0.18	0.18	0.17	0.19	0.18	0.19		0.00						
	AL AIC	0.01			0.00			0.19	0.20	0.20	0.22	0.21	0.23		0.01						
	AL CV5	0.05	0.02	0.02	0.03	0.02	0.02	0.10	0.09	0.08	0.09	0.09	0.09		0.00						
	L BIC	0.00						0.80	0.66	0.73	0.73	0.73	0.74	0.71	0.43	0.29	0.12	0.01	0.01	0.00	0.06
	L AIC	0.00						0.87	0.77	0.90	0.88	0.88	0.90	0.84	0.58	0.73	0.56	0.32	0.04	0.27	0.48
	L CV5	0.02	0.01	0.01	0.01	0.01	0.01	0.78	0.59	0.63	0.63	0.63	0.64	0.62	0.37	0.25	0.12	0.03	0.01	0.02	0.08
7	OLS	0.23			0.01			0.05	0.05	0.04	0.05	0.04	0.06	0.05	0.05	0.05	0.04	0.05	0.01	0.04	0.06
	AL BIC	0.01			0.00			0.18	0.16	0.16	0.18	0.17	0.20		0.01						
	AL AIC	0.01			0.00			0.18	0.18	0.19	0.22	0.20	0.23		0.01						
	AL CV5	0.04	0.01	0.01	0.02	0.01	0.01	0.09	0.08	0.07	0.09	0.08	0.10		0.00						
	L BIC	0.00						0.81	0.62	0.71	0.72	0.73	0.74	0.69	0.45	0.22	0.09	0.01	0.01	0.00	0.04
	L AIC	0.00						0.86	0.74	0.89	0.89	0.89	0.89	0.82	0.59	0.67	0.50	0.31	0.04	0.26	0.42
	L CV5	0.01	0.00	0.00	0.00	0.00	0.00	0.78	0.55	0.62	0.63	0.64	0.64	0.62	0.39	0.20	0.11	0.04	0.01	0.02	0.07

Table 1.1: Simulation Results The table provides an overview of type II (for priced factor FC, FC 0-5) and type I error (for unpriced factor FC, 6-11, and spurious FC, 12-99) ratio behavior in percentage points for the specified simulation cases. All details of the cases can be found in section 1.5.2. FC cases 12-15 are interesting insofar as they are spurious FC with high correlations to priced and unpriced FC. AL abbreviates adaptive Lasso, L stands for Lasso.

1.6 Data

Our objective is to preserve consistency as much as possible. Therefore the selection, data preparation and the notation of the description of firm characteristics generally follow the approach of Green et al. (2017). As in most studies, the analysis considers only CRSP stocks with share code 10 and 11 which are traded either at NYSE, AMEX or NASDAQ; for an example, see Fama and French (1992). Furthermore, we exclude stocks with missing market capitalization data and/or where book values are unavailable. Compustat data are aligned with a standard lag of six months of the fiscal year end date.[16] CRSP based stock/firm characteristics, such as idiosyncratic volatility, beta, maximum return or six-months momentum are used as of the most recent month end.[17] Additionally, following Green et al. (2017) some selected Compustat accounting data are set to zero if not available; see the appendix for details. In processing larger amounts of data, correcting extreme and often implausible values is mostly unavoidable. Correcting these values on a discretionary basis is not feasible; hence, winsorizing the data is a useful strategy to reduce the problem.[18] Therefore, each FC is winsorized at the 5% and 95% percentile at each point in time. Binary FC like *divi*, *divo*, *rd* and *ipo* are excluded from the winsorizing procedure. In the next step, missing data are replaced by the mean of the winsorized data at each point in time. Only then can the z-score standardization be applied at each calendar point. Note that we do not winsorize the return observations; instead we manually screen the data where the most extreme positive return is 2400%, which is below the elimination threshold of 10000% choosen in Green et al. (2017). Therefore, no observations are excluded because of implausible returns. Moreover, returns are only de-meaned for each period and not corrected by the standard deviations. Finally, the data can be stacked and the pooled regressions applied, as each independent variable has mean zero and variance one given by the property of combining z-scores. Note that this is necessary as the Lasso requires a normalized design matrix as input, as described above.

However, differences in selection of FC are unavoidable. This study employs only FC which are not dependent on Compustat quarterly and I/B/E/S data. A detailed description of each FC included in the empirical part of this study can be found in the appendix. Moreover, the β estimates are obtained by regressing rolling weekly stock returns on the market returns. The literature often employs an alternative procedure whereby stocks are ranked and sorted into portfolios according to their individual market beta; see, for example, Fama and French (1992). The betas assigned to each stock for estimating the equity market premia are obtained by using betas of the corresponding portfolios. Using portfolio beta estimates instead of individual stock betas has been applied to reduce potential errors-in-variable issues in the second stage regression. However, Ang et al. (2016) cast doubt on whether portfolio betas are optimal due to the loss of dispersion in individual betas.[19] More details about the specific CRSP and Compustat data and the corresponding data alignment process can be found in the appendix. The returns used in the prediction regression are the CRSP returns (RET) adjusted by the provided CRSP delisting return ($DLRET$). Additionally and for verification purposes, we benchmark our data for selected FC with the FC portfolio returns provided

[16]For example, the data of a firm with fiscal year end date 12/31 are aligned with data 06/30, predicting monthly returns from 6/30 to 7/31.

[17]For example, for the return prediction from 6/30 to 7/31, the max daily return from the period 5/31-6/30 is used.

[18]Winsorizing uses a sorted vector of data, and replaces the lowest x-percentile of values with the value at the x-percentile level. Equivalently, this is conducted for the highest x-percentile values.

[19]Ang et al. (2016) provide a detailed econometric discussion of using stocks vs. portfolios in empirical asset pricing.

by Kenneth French's Data Library. We find satisfying R^2's, reaching values from $0.99 >$ to about 0.9 for cases where the FC definition of the benchmark data slightly deviates from the one presented in Green et al. (2017). Furthermore, we follow Fama and French (1996) for the size classification definition, where large cap stocks are the 1000 stocks with the highest market capitalization, mid cap stocks rank 1001-2000 and small comprise all stocks with rank > 2000.

1.7 Empirical part

The first part explains the details of how we construct the required normalized matrix X of the unbalanced panel of FC and returns. This subsection is followed by an empirical analysis on the inference overlap between portfolio sorting and linear regressions. The third part covers in detail the discussion of the reduced factor zoo according to the adaptive Lasso selection procedure. It also stresses the differences compared to OLS selection. The final subsection briefly summarizes the results of an out-of-sample test comparing adaptive Lasso and OLS performance, complementing the simulation study to evaluate the benefits of adaptive Lasso selection from a portfolio/investors perspective.

1.7.1 Estimation set-up

As described above, the approach estimates the coefficients based on a pooled panel set-up. However, simply stacking the data causes two problems.

First, since we are interested in cross-sectional differences, we need to **normalize each FC at each point in time** to preserve the cross-sectional information. To illustrate the problem, one can think about the book-to-market ratio of single stocks, which certainly fluctuates partially based on market-wide price movements through time; standardizing along the entire panel would then implicitly change the order as time and cross-sectional information get mixed up.

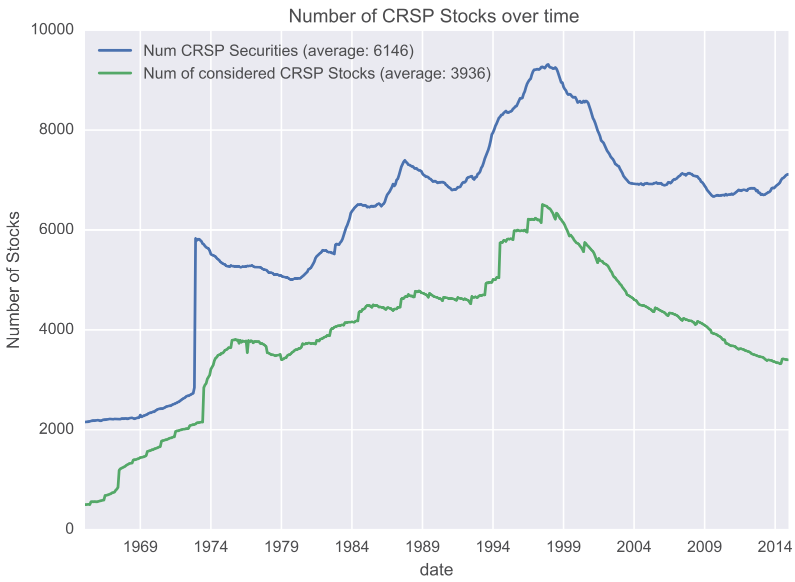

Figure 1.4: The plot displays the number of CRSP securities over time and the number of stocks we use for our analysis. The difference comes from only considering stocks with SHRCD 10 and 11, which are traded at either NYSE, AMEX or NASDAQ and which have positive market capitalization and no missing book value data. We observe an increase in the number of stocks around the year 2000. The jump in the 1970s is due to inclusion of NASDAQ stocks in the database.

Another issue which needs to be addressed is the **unbalanced panel** structure, as it implicitly causes the weights of each period in the regression to vary. Assuming there is no correlation between the returns and the number of stocks, we could ignore this issue, but as Figure 1.4 reveals, we see that prior to the stock-market peak in the beginning of the 2000s we have a much higher number of stocks with unknown return dependence. Therefore, we suggest adjusting the number of stocks in each period to the mean number of stocks per time point. This can be achieved by simply randomly drawing stocks with replacement at each point in time until we have filled the desired sample size.

Figure 1.5 presents the **correlation structure** of the FC. It shows overall only five cases with absolute correlation coefficients greater than 0.9. Even though we cannot achieve precisely the same correlation structure in our simulation, we have considered cases with correlations of around 0.9 and hence capture this feature observed in the data in our simulation as well. As we show in the simulation study in Section 1.5, correlations around 0.9 cause no selection issues for the adaptive Lasso; only a Lasso-based selection appears prone to mis-classification. However, we want to avoid including almost identical FC. Hence, before regressing the returns on the full set of FC included in our dataset, we screen the correlations for cases with an absolute correlation greater than .95. In such cases, we eliminate the more recently published FC of the affected pair from our analysis. Finally, not all FC are included in our FC analysis due to data problems; we drop: *cfp_ia*, *roic*, *pchemp_ia* and *ipo*.

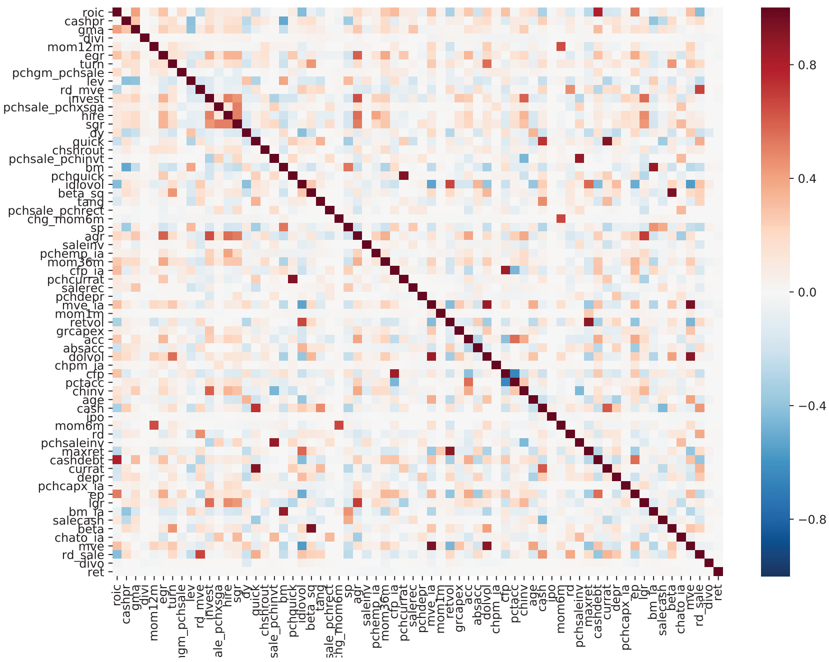

Figure 1.5: Empirical correlation matrix (best seen in color):The figure exhibits the correlation of all stocks and firm characteristics included in our analysis for the years June 1974 until December 2014. We use the normalized, winsorized and pooled FC data (as used in the full sample regression) to calculate the correlation coefficients. The figure shows five extreme correlation pairs (>0.9). For example, the highest absolute correlation is measured for *beta* and *beta_sq* with a coefficient slightly less than 0.95.

1.7.2 Portfolio sorts and the linear model

Portfolio sorting is the prevailing methodology underlying most (if not all) anomaly variables included in this study. However, our final objective is to measure the significance of these FC in a multivariate linear specification. Our concern is, that the linear model might not necessarily capture the effects which are documented in the portfolio specification of the original studies. This would not be a fair evaluation of these studies, as we are not able to differentiate whether an FC becomes irrelevant because it is absorbed in the multivariate analysis or it is simply a non-linear effect we cannot capture. Hence, we want to ensure that the effects we measure based on univariate portfolio sorts are comparable to the inference resulting from a linear regression. For example, the effects of cross-sectional return spreads might only be measurable in the extreme part of the FC distribution (i.e. the decile long-short strategy). One way to test this potential mismatch, is to compare the t-value of the FC coefficient estimate obtained from a regression approach with the t-value of the intercept (the α) of factor regression on the corresponding long-short portfolio returns. We investigate the effect separately conditioned on a univariate, CAPM, FF three-factor and Carhart (1997) four-factor model (the dominant benchmarks at the time). The relevant t-value of the regression stems from a pooled panel OLS, where we add the matching factor FC for each model to the right-hand side. For instance, the CAPM regression right-hand side variables are the FC of interest and *beta*; the three-factor model includes additionally *mve* and bm. The portfolio counterpart is obtained by running a regression of the corresponding long-short hedge returns conditioned on the respective factor returns. For example, in case of the three-factor model, we run the following two regressions,

$$R^e_{t+1} = \beta_c \mathrm{FC}_{c,t} + \gamma_1 beta_t + \gamma_2 mve_t + \gamma_3 bm_t + \epsilon_{t+1}, \tag{1.8}$$

$$r_{\mathrm{pf}_c,t} = \alpha_c + \beta_{\mathrm{mkt}} r_{\mathrm{mkt},t} + \beta_{\mathrm{HML}} r_{\mathrm{HML},t} + \beta_{\mathrm{SMB}} r_{\mathrm{SMB},t} + \kappa_t, \tag{1.9}$$

where we take the t-value associated with coefficient β_c from equation 1.8, and the t-value of the intercept, α_c, of 1.9. Note, R^e_t is the vector of pooled excess returns and $r_{\mathrm{pf}_c,t}$ is the long-short portfolio return constructed based on the specific FC c. We loop over all FC separately and obtain a collection of t-value pairs for each FC. We use these pairs to analyze the degree of matching. We repeat this exercise for the different benchmark models by adjusting the right-hand side variables accordingly.

We measure the commonality in two ways. First, we calculate a significance coverage ratio — it measures the share of FC that are identically classified into either significant or insignificant (at the 5%-level). Second, we look at the correlation between the two t-values. The results are displayed in Table 1.2. The correlation is close to one for the univariate model, and performs reasonably well for the other models. The coverage ratio is slightly lower, but reaches values of about 90%. We conclude that the linear regression captures most of the α measured in long-short portfolios. Furthermore, the table shows that the commonality declines once the results are conditioned on some benchmark factor model.[20]

Model	Uni		CAPM		FF3		Carhart	
Cutoff	FF Style	decile	FF Style	decile	FF Style	decile	FF Style	decile
Coverage	0.89	0.93	0.80	0.89	0.80	0.87	0.75	0.74
Corr	0.97	0.97	0.94	0.95	0.87	0.86	0.85	0.83

Table 1.2: Linear model vs. portfolio sorting The table shows a comparison of portfolio sorts and a linear regression approach (pooled OLS). The row "coverage" indicates the classification overlap. It expresses the ratio of the number of FC that share the same classification (i.e. significant or not) over the total number of FC considered. The row "corr" measures the correlation between the t-value of the coefficient of the regression model and the t-value of the α of the long-short portfolio. The model "Uni" measures the differences between the hedge portfolio and a univariate regression. The column "FF3" measures the α w.r.t to Fama and French (1993) three-factor model. In this case the benchmark is a multivariate regression, where the FC variable is augmented by "beta", "bm" and "mve" on the right-hand side. The other models are adjusted accordingly. The data include all stocks, from 1970-01-01 until 2014-12-31.

1.7.3 Shrinking the zoo of firm characteristics

The analysis considers the years from 1974 to 2014 and primarily emphasizes the selection regression including all stocks. These findings are shown in Table 1.3. We separately look at

[20] The other models are potentially less precise because the data underlying the factor models do not perfectly match, potentially explaining parts of the deviation we observe.

the regression conditioning on large, mid and small cap sized stocks. These results can be found in Tables 1.4 to 1.6. All four tables display the sets of active FC determined by the adaptive Lasso, value-weighted adpative Lasso, OLS and Lasso selection. The discussion in this section stresses mostly the details of the FC selection for the value-weighted adaptive Lasso and the differences from the alternative selection procedures. This reflects the findings of Hou et al. (2017), who document the impact of micro-cap stocks. It is also in line with the work of Green et al. (2017), as they emphasize a value-weighted selection exercise. We are interested in the full sample analysis, i.e. the results of the single pooled regression applied at once over all periods to obtain the set of selected FC. Additionally, we apply a rolling regression with an estimation window of 15 years to better understand differences and robustness over time. Note that we denote the sign of the selected FC in brackets behind each FC when mentioned in the text for the first time and also provide a brief description of the respective FC but for the following subsection only. For the other subsections we refer to the appendix.

FC selection including all stocks

Considering all stocks for the selection analysis, we do not eliminate FC prior to selection regression due to extreme correlations, since no correlation pair exceeds the defined threshold level in Section 1.7.1. Table 1.3 embodies the FC selection results for the equally and value-weighted adaptive Lasso, OLS and Lasso based on **all sample periods**. We find that the value-weighted adaptive Lasso selects two out of five FC associated with the Fama and French (2014) five-factor model and observe consistency with respect to the sign of the coefficients. Specifically, we identify *beta*(+) – market; and *gma*(+) – profitability as part of the set of active FC. The FC not directly represented are *mve*, the *bm* and *agr*; reflecting the size, value and investment factor, respectively. However, the selection comprises *bm_ia*(−) – an industry-adjusted variant of the classical value FC *bm*. Additionally, it includes *ep*(+)— earnings-price ratio; which defines a different value metric. In contrast to the five-factor model, we measure 37 FC characterizing differences in expected cross-sectional returns. Many out of these 37 FC are based exclusively on price information. This includes the most relevant FC, measured by the absolute size of the coefficient, *chg_mom6m*(−) – difference in the six months momentum or a reversion to six months lagged six months trend. Moreover, the value-weighted adaptive Lasso selects the following price-related FC: *ret_vol*(−) – last month return volatility; *mom6m*(+) – the six-months momentum; *maxret*(+) – the max daily return of the previous month; *mom1m*(−) – short-term reversal; *mom36m*(−) – long-term reversal; *idiovol*(−) – last month idiosyncratic volatility; and *mom12m*(+) – the classical twelve-months momentum. Finally, we point out that *dolvol*(−), reflects the dollar trading volume, a liquidity proxy. We skip all other selected FC and refer to Table 1.3 instead. Furthermore, the table expresses the differences between the four selection specifications and underscores the relevance of this choice. The adaptive Lasso selects a set of FC of similar size, but with distinct differences. The impact of small and micro-cap stocks is not limited to the selection process, but is apparent when the size of the coefficients is taken into account. For example, *chg_mom6m*, *mom1m*, *bm*, *rd_mve* and *retvol* differ starkly. This is in line with the results of Green et al. (2017), who document similar tilts in case of the value-weighted regressions. Generally, these across-study comparisons have to be conducted with caution, as the set of FC and the sample periods can differ and hence impact the inference in an unknown way. OLS on the other hand selects a sparser set, whose 16 FC elements are all part of the set identified by the adaptive Lasso.

The table reveals that a Lasso-based procedure would suggest an even higher dimensional relation between FC and returns. We dispense with the discussion here, due to the strong conditions it imposes and the reported under-performing simulation figures (as a consequence of the former). Overall, we find a substantial number of FC inspected do not contain relevant information for predicting returns when considered in a multivariate selection, as 27 of the included 64 FC are not picked by the adaptive Lasso. Further insights into the selection

FC	VWAL	AL	OLS	Lasso	FC	VWAL	AL	OLS	Lasso
chg_mom6m	-7.25	-1.48	-1.30 (-1.25)	-1.46	rd	0.26	0.00	-0.04 (-0.16)	0.00
ep	5.41	0.85	0.97 (1.55)	0.87	mom12m	0.25	2.64	2.88** (2.37)	2.59
retvol	-5.03	-0.51	-0.58 (-0.52)	-0.48	chato_ia	0.24	0.35	0.72** (2.11)	0.60
dolvol	-4.68	0.00	0.04 (0.04)	0.00	saleinv	0.24	0.23	0.30 (1.42)	0.35
mom6m	4.39	0.00	-0.33 (-0.23)	0.00	salecash	-0.22	0.00	-0.16 (-0.61)	-0.12
maxret	4.04	0.00	-0.18 (-0.17)	-0.23	pchsale_pchxsga	-0.16	0.00	-0.15 (-0.48)	-0.07
gma	2.97	1.47	1.53*** (3.55)	1.43	currat	-0.15	-0.99	-1.85** (-2.51)	-0.92
cashpr	-2.24	-0.63	-0.80** (-2.49)	-0.72	lev	0.12	0.44	0.73 (1.29)	0.56
egr	-1.83	0.00	0.05 (0.12)	0.00	mve_ia	0.00	-2.32	-2.06** (-2.07)	-1.98
beta	1.83	0.66	1.01 (1.19)	0.80	agr	0.00	-2.23	-1.63** (-2.46)	-1.61
mom1m	-1.50	-6.28	-6.22*** (-5.82)	-6.16	bm	0.00	2.06	1.43 (1.58)	1.41
idiovol	-1.47	-0.42	-0.79 (-0.87)	-0.58	cash	0.00	1.02	0.92 (1.51)	0.96
chinv	-1.25	0.00	-0.11 (-0.28)	-0.24	cashdebt	0.00	0.82	0.87** (1.97)	0.80
acc	-1.21	-0.63	-0.78 (-1.30)	-0.57	age	0.00	0.79	1.07*** (3.78)	0.91
turn	1.18	-0.60	-0.72 (-1.16)	-0.64	chshrout	0.00	-0.67	-0.71** (-2.19)	-0.68
pctacc	1.10	0.00	0.32 (0.81)	0.06	dy	0.00	-0.65	-0.88* (-1.92)	-0.74
rd_mve	0.98	2.81	2.95*** (5.67)	2.75	pchsale_pchinvt	0.00	0.62	0.63 (1.34)	0.57
pchquick	-0.98	0.00	0.11 (0.17)	0.00	chpm_ia	0.00	0.58	0.70** (2.23)	0.64
bm_ia	0.98	0.00	0.51 (0.71)	0.53	pchcapx_ia	0.00	-0.45	-0.62*** (-2.58)	-0.55
divi	0.93	0.00	-0.26 (-1.33)	-0.20	mve	0.00	-0.36	-1.22 (-0.84)	-0.96
rd_sale	-0.78	0.00	-0.15 (-0.24)	0.00	invest	0.00	-0.23	-0.47 (-1.14)	-0.42
sp	0.75	0.00	-0.09 (-0.17)	0.00	pchcurrat	0.00	0.00	-0.52 (-0.89)	-0.35
pchdepr	0.66	-0.12	-0.44 (-1.64)	-0.36	grcapex	0.00	0.00	0.27 (0.97)	0.17
mom36m	-0.65	0.00	-0.33 (-0.60)	-0.26	depr	0.00	0.00	0.24 (0.68)	0.16
lgr	-0.65	-0.10	-0.57 (-1.37)	-0.50	hire	0.00	0.00	0.33 (1.09)	0.16
pchgm_pchsale	0.59	0.27	0.42* (1.67)	0.41	salerec	0.00	0.00	0.30 (0.86)	0.14
absacc	0.57	0.45	0.68** (2.14)	0.61	pchsale_pchrect	0.00	0.00	-0.18 (-0.76)	-0.10
sgr	-0.51	-0.41	-0.65 (-1.44)	-0.53	tang	0.00	0.00	0.03 (0.09)	0.01
cfp	0.39	0.08	0.41 (0.65)	0.31					

Table 1.3: FC Selection All Stocks: The table exhibits the regression results of OLS, Lasso, adaptive Lasso (AL), and the value-weighted adaptive Lasso (VWAL) for each FC, sorted by the absolute value in the order of the VWAL, AL, Lasso and t-values of the OLS coefficients. The numbers in the brackets of the OLS column represent t-values. The table displays only FC having an abs adaptive Lasso or Lasso coefficient greater than zero, or an abs OLS t-value greater than 1.645. Accordingly, the following FC used in the selection regression are not represented: *quick*, *pchsaleinv*, *divo*. Note the coefficients are scaled by a factor of 1000. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Cells with blue background color indicate selected status. The sample period spans Jun 1974 until Dec 2014.

process and its stability over time can be gained by running the same type of regression in a **15 years rolling window** framework. The results are illustrated graphically in Figure 1.6. Overall the results are in line with the full sample selection, despite instabilities over time. We notice that the two most often selected FC over time are the price-related *chg_mom6m* and *maxret*. However, *maxret* changes the sign of the signal around the year 2000. Moreover, the signal strength of *chg_mom6m* weakens with time. Effects of liquidity and value are visible through all months. Specifically, we observe that *dolvol* as well as *ep* are particularly pronounced in the more recent periods. Furthermore, we notice that three additionally important FC are *acc* – accruals; *cashdebt* – cash flow to debt ratio; and *beta_sq*, especially because of their strengthening signal towards the more recent periods after only scattered initial consideration. The selection of *mom1m* and *gma* in the full period selection seems mostly a result of the first part of the sample, as the signals almost completely vanish for the most recent 10 to 15 years. The least relevant selected FC of the full sample selection, *hire* and *absacc*, appear only in one month (out of 306) in the rolling window selection, indicating only a marginal role in the return-forming process.

FC selection conditioned on size

The selection results conditioned on **large cap stocks** only are depicted in Table 1.4. Note that we drop *beta_sq* prior to the selection regression, due to its high correlation with *beta*, which exceeds .95. Overall, we observe fewer active FC compared to results including all stocks; a total of 18 (vs. 37) are selected. Not overly surprising, given the value-weight, the top-ranked FC resemble the picture described above – *chg_mom6m*$(-)$, *mom6m*$(+)$, *ep*$(+)$, and *dolvol*$(-)$ describe the top four FC. Noteworthy is that no FC directly related to the stock's riskiness is included, as it misses, for example, *beta*, *retvol*$(-)$ or *idiovol*$(-)$. Differences between an equal weighted adaptive Lasso are obvious, but these are pronounced mostly among the least important FC. Moreover, OLS selects once more an even sparser set vs. the adaptive Lasso with less than half as many members as the competing method.

Considering only **mid cap stocks**, we find similarly to the large cap case an above threshold correlation for the *beta* and *beta_sq* correlation pair, which leads to discarding *beta_sq*.[21] Table 1.5 presents the active set of FC conditioned on mid-sized stocks. Strikingly, price-based FC rank highest, with *mom1m*$(-)$, *mom12m*$(+)$ and *retvol*$(-)$ defining the top three. Furthermore, *agr*$(-)$, *rd_mve*$(+)$, *sp*$(+)$, *cashdebt*$(+)$, *ep*$(+)$, *chpm_ia*$(+)$ and *maxret*$(-)$ complete the characterization of mid cap returns from an FC perspective. The value-weighted and equal weighted adaptive Lasso selections mostly overlap, due to more equally distributed market-capitalization in this subsample. The pattern repeats; an OLS-based selection would

[21] It is not surprising that *beta* and *beta_sq* are more highly correlated for large and mid cap stocks, as the beta of these stocks tend to be more centered around one, causing any quadratic transformation to be more correlated compared to a more dispersed *beta* measured among small caps.

FC	VWAL	AL	OLS	Lasso	FC	VWAL	AL	OLS	Lasso
chg_mom6m	-5.29	-3.60	-3.78*** (-2.89)	-3.09	currat	0.00	-0.91	-1.82* (-1.85)	-1.07
mom6m	3.46	2.27	2.45 (1.63)	1.62	mve	0.00	-0.85	-0.67 (-0.74)	-0.65
ep	1.90	1.24	1.19** (2.17)	1.16	mom12m	0.00	0.79	0.72 (0.59)	1.17
dolvol	-1.58	-0.25	-0.92 (-1.17)	-0.51	cashdebt	0.00	0.79	1.19 (1.47)	0.92
gma	1.26	0.18	0.81* (1.72)	0.54	retvol	0.00	-0.79	-1.51** (-2.14)	-0.84
cashpr	-1.22	-0.33	-0.68 (-1.59)	-0.49	dy	0.00	-0.43	-0.90* (-1.77)	-0.62
mom1m	-1.20	-1.85	-2.10** (-2.57)	-1.82	agr	0.00	-0.38	-0.46 (-0.70)	-0.43
egr	-1.18	-1.20	-1.18*** (-3.54)	-1.03	pchsale_pchinvt	0.00	0.29	0.75 (0.92)	0.39
acc	-0.86	0.00	-0.61 (-1.09)	-0.29	chpm_ia	0.00	0.26	0.54 (1.41)	0.42
rd_mve	0.73	1.39	1.64** (2.16)	1.33	pchsale_pchxsga	0.00	-0.19	-0.48 (-1.61)	-0.40
pctacc	0.56	0.00	0.23 (0.58)	0.00	depr	0.00	0.16	0.35 (1.01)	0.33
mom36m	-0.53	0.00	-0.29 (-0.52)	-0.22	lev	0.00	0.00	0.58 (0.90)	0.34
chinv	-0.40	0.00	0.08 (0.19)	0.00	bm	0.00	0.00	0.57 (0.84)	0.33
pchdepr	0.33	0.00	0.31 (1.21)	0.16	divi	0.00	0.00	-0.36** (-2.03)	-0.25
cash	0.26	0.00	0.32 (0.44)	0.24	grcapex	0.00	0.00	-0.20 (-0.78)	-0.09
chato_ia	0.22	0.00	0.43 (1.45)	0.27	beta	0.00	0.00	0.31 (0.40)	0.05
sp	0.11	0.65	0.47 (1.13)	0.53	pchquick	0.00	0.00	-0.19 (-0.34)	-0.03
salerec	0.05	0.00	0.34 (0.86)	0.19					

Table 1.4: FC Selection Large Stocks: The table exhibits the regression results of OLS, Lasso, adaptive Lasso (AL), and the value-weighted adaptive Lasso (VWAL) for each FC, sorted by the absolute value in the order of the VWAL, AL, Lasso and t-values of the OLS coefficients. The numbers in the brackets of the OLS column represent t-values. The table displays only FC having an abs adaptive Lasso or Lasso coefficient greater than zero, or an abs OLS t-value greater than 1.645. Accordingly, the following FC used in the selection regression are not represented: *maxret*, *rd*, *divo*, *quick*, *sgr*, *pchcapx_ia*, *hire*, *invest*, *turn*, *mve_ia*, *salecash*, *age*, *pchsale_pchrect*, *tang*, *chshrout*, *pchcurrat*, *pchsaleinv*, *rd_sale*, *absacc*, *pchgm_pchsale*, *bm_ia*, *cfp*, *idiovol*, *saleinv*, *lgr*. Note the coefficients are scaled by a factor of 1000. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Cells with blue background color indicate selected status. The sample period spans Jun 1974 until Dec 2014.

FC	VWAL	AL	OLS	Lasso	FC	VWAL	AL	OLS	Lasso
mom1m	-2.83	-2.59	-2.54*** (-3.09)	-1.72	dy	0.00	0.00	-1.29*** (-2.79)	0.00
mom12m	2.80	2.65	1.70 (1.25)	1.98	lgr	0.00	0.00	1.02** (2.08)	0.00
retvol	-2.22	-2.25	-2.18*** (-2.91)	-1.16	idiovol	0.00	0.00	-1.39** (-2.07)	0.00
agr	-2.02	-1.98	-2.89*** (-4.10)	-1.37	lev	0.00	0.00	1.16** (2.05)	0.00
rd_mve	1.62	1.85	1.09** (2.00)	0.88	beta	0.00	0.00	1.75* (1.94)	0.00
sp	1.40	1.39	1.00* (1.92)	0.69	age	0.00	0.00	0.57* (1.88)	0.00
cashdebt	0.99	1.61	2.18*** (3.00)	0.61	depr	0.00	0.00	0.68* (1.84)	0.00
ep	0.40	0.00	0.74 (1.21)	0.14	quick	0.00	0.00	1.75* (1.80)	0.00
chpm_ia	0.38	0.56	0.89** (2.47)	0.15	acc	0.00	0.00	-1.11* (-1.80)	0.00
maxret	-0.16	-0.53	-0.53 (-0.73)	-1.01	pchsale_pchinvt	0.00	0.00	0.91* (1.72)	0.00
invest	0.00	0.00	-0.37 (-0.70)	-0.08	pchsale_pchrect	0.00	0.00	-0.42* (-1.65)	0.00
currat	0.00	0.00	-2.97*** (-3.02)	0.00					

Table 1.5: FC Selection Mid Stocks: The table exhibits the regression results of OLS, Lasso, adaptive Lasso (AL), and the value-weighted adaptive Lasso (VWAL) for each FC, sorted by the absolute value in the order of the VWAL, AL, Lasso and t-values of the OLS coefficients. The numbers in the brackets of the OLS column represent t-values. The table displays only FC having an abs adaptive Lasso or Lasso coefficient greater than zero, or an abs OLS t-value greater than 1.645. Accordingly, the following FC used in the selection regression are not represented: *cashpr*, *rd_sale*, *salerec*, *gma*, *chg_mom6m*, *chshrout*, *pchquick*, *pchsaleinv*, *mom6m*, *pchdepr*, *dolvol*, *rd*, *saleinv*, *sgr*, *hire*, *pchsale_pchxsga*, *pchgm_pchsale*, *mve*, *mve_ia*, *grcapex*, *divi*, *tang*, *pchcapx_ia*, *absacc*, *egr*, *pchcurrat*, *salecash*, *cfp*, *cash*, *mom36m*, *chato_ia*, *turn*, *bm*, *bm_ia*, *pctacc*, *divo*, *chinv*. Note the coefficients are scaled by a factor of 1000. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Cells with blue background color indicate selected status. The sample period spans Jun 1974 until Dec 2014.

change the picture significantly, as the active OLS FC set differs substantially.

Finally, we briefly discuss the results including exclusively **small cap stocks**. The correlation pre-screening reveals that all correlation pairs measured are below the defined level and hence all variables enter the regression. The FC selection, as shown in Table 1.6, shows that the price information-driven FC are dominant as in the mid sized selection regression. We find

FC	VWAL	AL	OLS	Lasso	FC	VWAL	AL	OLS	Lasso
retvol	-4.86	0.00	-0.59 (-0.47)	0.00	sp	0.56	0.00	0.37 (0.50)	0.00
mom1m	-4.11	-8.01	-8.09*** (-6.99)	-7.69	pchquick	0.48	0.00	1.30 (1.33)	0.00
mve	-3.71	-3.27	-3.23*** (-3.12)	-2.90	lev	0.45	0.00	-0.44 (-0.60)	0.00
mom12m	3.67	2.59	3.48*** (2.70)	2.21	chato_ia	0.37	0.00	0.46 (0.94)	0.00
rd_mve	3.07	3.10	2.94*** (4.71)	2.70	depr	0.27	0.00	0.48 (1.07)	0.00
idiovol	-2.82	0.00	-0.42 (-0.48)	0.00	cashpr	-0.21	-0.43	-1.13** (-2.38)	-0.29
agr	-2.51	-3.17	-2.33** (-2.36)	-2.63	divo	-0.18	0.00	-0.45 (-1.45)	0.00
beta	1.99	0.56	1.16 (1.45)	0.24	chshrout	-0.18	-0.61	-0.60 (-1.58)	-0.52
currat	-1.90	0.00	-2.70*** (-2.71)	0.00	pchsaleinv	0.13	0.00	0.23 (0.39)	0.05
cash	1.82	1.21	1.00 (1.41)	0.71	chpm_ia	0.07	0.00	0.50 (1.43)	0.16
mom6m	1.63	0.00	-1.38 (-0.88)	0.00	maxret	0.00	-1.13	-0.69 (-0.56)	-0.89
mve_ia	-1.50	-1.26	-1.59** (-1.99)	-1.16	chg_mom6m	0.00	-1.12	-0.28 (-0.27)	-0.81
dy	-1.41	-0.41	-0.94* (-1.94)	-0.28	bm_ia	0.00	1.11	0.74 (0.73)	1.06
ep	1.29	1.13	1.32* (1.87)	0.72	dolvol	0.00	-1.02	-1.06 (-1.32)	-0.81
gma	1.29	1.16	1.24** (2.53)	0.94	pchsale_pchinvt	0.00	0.99	0.75 (1.32)	0.65
bm	1.23	0.87	1.15 (0.93)	0.78	pctacc	0.00	-0.78	-0.72 (-1.48)	-0.52
cashdebt	1.22	0.78	0.94* (1.74)	0.50	mom36m	0.00	-0.35	-0.64 (-1.02)	-0.30
pchgm_pchsale	1.09	1.45	1.40*** (3.80)	1.23	invest	0.00	0.00	-0.15 (-0.25)	-0.32
cfp	0.88	0.00	0.14 (0.17)	0.12	rd	0.00	0.00	0.52 (1.32)	0.12
pchcapx_ia	-0.77	-0.58	-0.80** (-2.41)	-0.40	pchsale_pchrect	0.00	0.00	0.50 (1.43)	0.06
age	0.74	0.87	1.08*** (2.75)	0.63	chinv	0.00	0.00	0.26 (0.47)	-0.03
egr	-0.63	0.00	-0.40 (-0.79)	0.00	salecash	0.00	0.00	-0.77** (-2.07)	0.00
quick	0.58	0.00	2.55** (2.33)	0.00					

Table 1.6: FC Selection Small Stocks: The table exhibits the regression results of OLS, Lasso, adaptive Lasso (AL), and the value-weighted adaptive Lasso (VWAL) for each FC, sorted by the absolute value in the order of the VWAL, AL, Lasso and t-values of the OLS coefficients. The numbers in the brackets of the OLS column represent t-values. The table displays only FC having an abs adaptive Lasso or Lasso coefficient greater than zero, or an abs OLS t-value greater than 1.645. Accordingly, the following FC used in the selection regression are not represented: *pchcurrat*, *sgr*, *grcapex*, *divi*, *pchsale_pchxsga*, *salerec*, *turn*, *pchdepr*, *saleinv*, *absacc*, *acc*, *hire*, *rd_sale*, *tang*, *lgr*. Note the coefficients are scaled by a factor of 1000. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Cells with blue background color indicate selected status. The sample period spans Jun 1974 until Dec 2014.

consistency on which FC occupy the top ranks, considering the magnitude of short-term reversal. The five highest rank FC are: *retvol*(−), *mom1m*(−), *mve*(−), *mom12m*(+), and *rd_mve*(+). OLS and adaptive Lasso deviate but slightly less so in this case compared to the FC selection of large and mid cap stocks.

Condensing the results jointly

Comparing Tables 1.3 and 1.4-1.6 allows us to understand the main results in greater detail. First, we see that the intersection of all sets through all methods (value-weighted adaptive Lasso, adaptive Lasso and OLS) consists of *mom1m* and *rd_mve* (always with the same sign). This across size and methodology consistency provides evidence for short-term price reversals and a positive return benefit from R&D spending. However, we stress that a value dimension should not be simply neglected from this perspective as the value-weighted adaptive Lasso captures two value-related FC in each subsample. In particular, we identify *ep* and *sp* in the selection results including all, only large cap, mid and small cap stocks, respectively. Less robust are a handful of other FC, whose appearance in the active set is not visible in all three size categories. For example, two FC of the five-factor model are not robust conditioned on size. The positive relation between profitability (*gma*) and the returns is only observed for large and small cap stocks, and the investment factor (*agr*) is not relevant for the economically important large caps. Generally, we find that OLS selects much sparser sets of relevant FC, with the exception of mid cap stocks – which is in line, once more, with Green et al. (2017), who document a relatively small set of relevant FC. Three out of the eight selected FC in Green et al. (2017), Table 5 (A), appear in the value-weighted adaptive Lasso selection: *mom1m*, *chg_mom6m* and *retvol*. On the other hand, *cash* and *bm* are not selected; the remaining three FC, *ear*, *nincr* and *zerotrade* are not part of our data set.

1.7.4 Summary long-short portfolio performance

We test the empirical performance by taking an investor's perspective. We form long-short portfolios based on adaptive Lasso and OLS selection procedures. Despite relatively large deviations in the selection, the findings suggest that the differences are not significant when the Sharpe ratio is the measure of interest. For more details we refer to A.1 in the appendix.

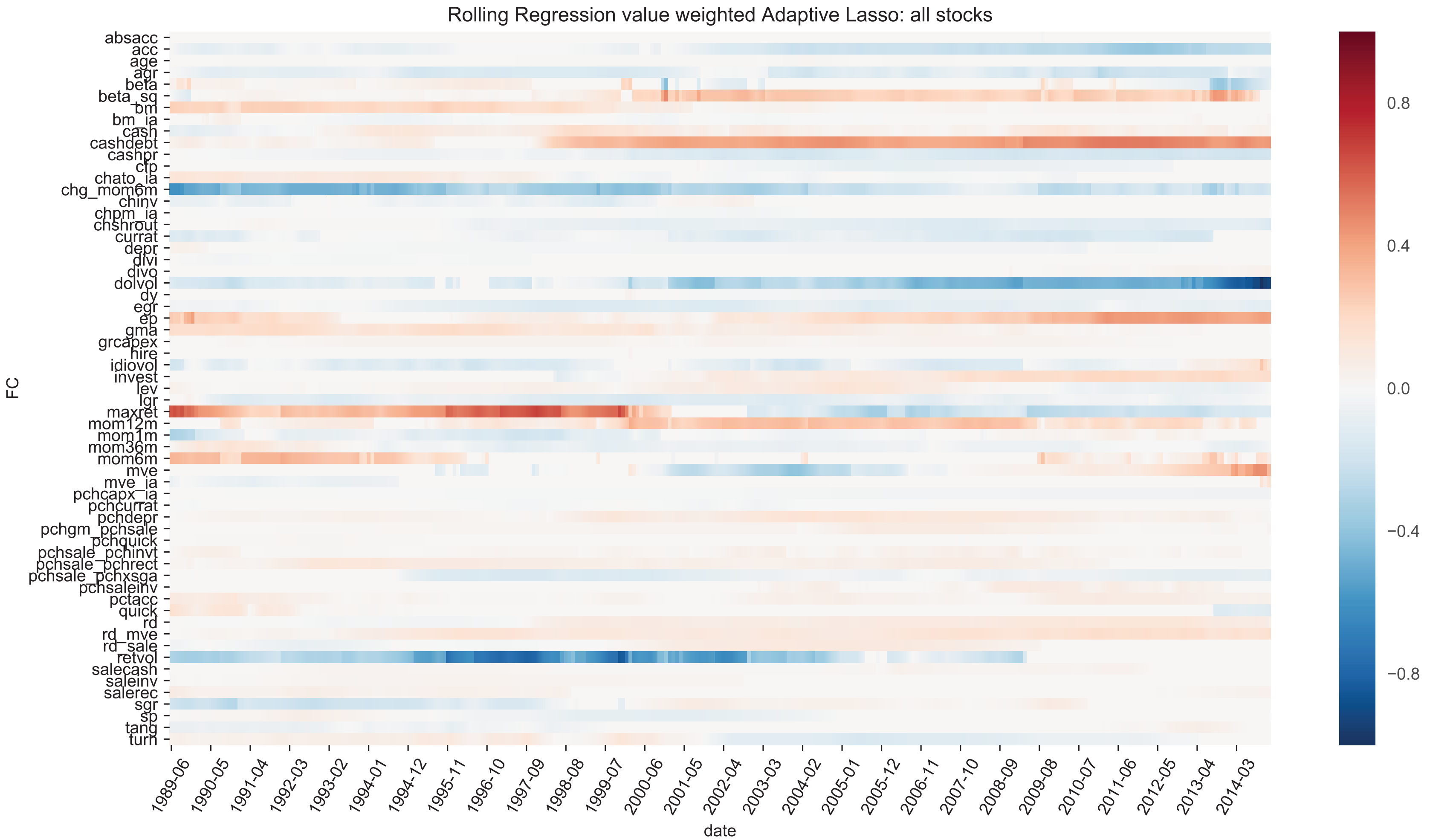

Figure 1.6: Rolling adaptive Lasso selection all stocks (best seen in color): The figure displays the adaptive Lasso estimates on rolling basis with an estimation window of 15 years. The top 5 FC measured over the relative frequency of occurrence are: *chg_mom6m*(100%), *maxret*(93.5%), *dolvol*(92.2%), *gma*(87.6%), *cash*(86.0%) and *acc*(81.4%).

1.8 Conclusion

In this work, we propose the application of the adaptive Lasso for selecting FC in the rich zoo of firm characteristics that purportedly explain differences in the cross-section of expected returns. An adaptive Lasso selection procedure applied to 68 FC included in this paper and constructed based on US stock data from 1965 to 2014 identifies a highly dimensional return process. However, we show that the vast part of published FC are selected when considered in a multivariate analysis simultaneously; we identify 37 FC of relevance. This contrasts with a much sparser set selected by OLS. Among the most consistently selected FC through time and size subsamples are price-related measures. Particularly outstanding are the one-month momentum, well known as the short-term reversal, and the change in the six months momentum. The two are complemented by size-adjusted R&D spending, which appears, like the former two, to be a major driver behind the differences in expected stock returns. Moreover, we robustly identify a value dimension explaining differences in expected returns. Many other FC are sensitive to size groups and/or unstable over time.

Furthermore, this study contributes to a better understanding of the behavior of the adaptive Lasso when applied in panel data settings. Our results are based on a Monte Carlo simulation study. The simulation considers panel data scenarios of low signal-to-noise ratios including heteroscedastic, non-normal and highly cross-sectionally correlated errors. We compare the performance of the adaptive Lasso, Lasso and OLS-based selection. The results show unambiguously that the Lasso is inferior to its adaptive version in most specifications. In particular, a required condition, most apparent in cases of higher correlations, reveals massive shortcomings in the Lasso. In contradistinction, the adaptive Lasso appears promising compared to classical OLS inference, especially at reducing type II error ratios and controlling FC that suffer from a likely publication bias. The only concern revealed by the simulation are cases where FC do not possess explanatory power for differences in average returns, but nonetheless impact the risk dimension of the returns.

Appendix A

A.1 FC selection performance

A stock market simulation study can never be a perfect replication of the true underlying DGP of the cross-section of stock returns. Hence, a further performance test to evaluate the performance of the Adaptive Lasso vs. OLS is to analyze an out-of-sample long-short portfolio performance for each of the two methods. This comparison is a useful tool to analyze the selection quality, as the portfolio constituency is determined based on the return prediction for each stock. The following examples illustrates why. First, assume that a spurious FC is present in the set of predictors for one method but not for the other. This spurious FC simply adds noise to the prediction outcomes and thus deteriorates the quality of the ranking of the affected method. Second, assume that a FC, which contains valuable predictive information, is only selected by one of the two methods. This additional dimension leads to an improvement of the prediction accuracy of the method selecting this FC. Hence, a portfolio performance exercise weights type I and type II according to their impact on the prediction quality. Though the portfolio comparison is not a perfect test, it provides a useful indication of whether these differences are economically meaningful from an investor's perspective and allows conclusions one whether one selection method is preferable over the other.

The respective Adaptive Lasso and OLS portfolios are obtained by proceeding in the following way: The relevant FC of the adaptive Lasso portfolio are selected using the adaptive Lasso as the selection tool, estimated as described above. These selected FC are then used in an OLS regression to obtain coefficient estimates for each FC. Combining these coefficients and the individual FC information, return predictions for each stock can be calculated. This two-step approach helps to separate the selection and coefficient estimation problem of the two methods. The return predictions can then be sorted from highest to lowest at each point in time and hence the classical portfolio return analysis can be applied. The portfolio is long the stocks with the highest expected returns and short the stocks with the lowest expected returns. The estimation and the rebalancing occurs with a monthly frequency. In similar fashion we can construct the OLS portfolios, where we simply replace the FC selection with the OLS selection procedure. Notice, that the estimation and prediction is carried out in an out-of-sample framework such that no future data-points are included at each point in time. These long-short returns can then be used to test differences in Sharpe ratios of the two methods. Specifically, we use the Ledoit and Wolf (2008) test with a HAC robust covariance estimate, including a Parzen kernel and AR automatic bandwith selection as described in Andrews (1991).

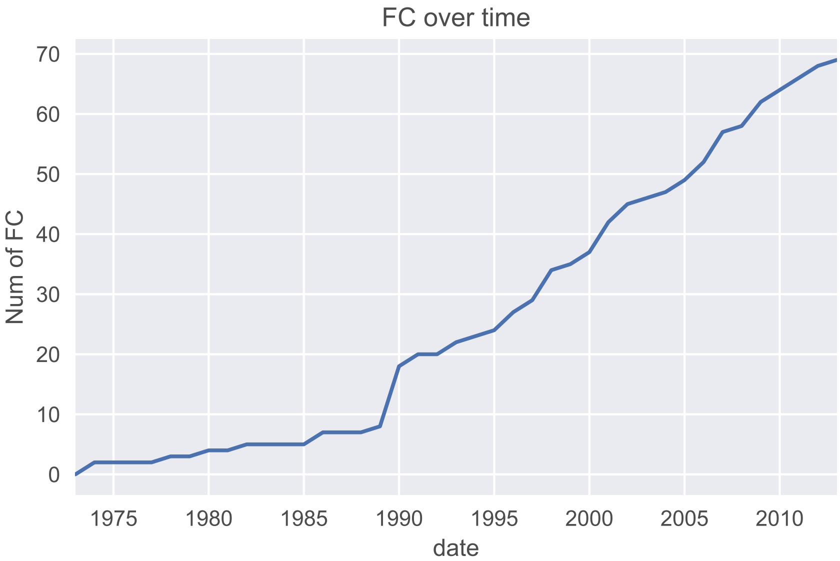

Figure A.1: The graph depicts the number of published FC included in this study as a function of time.

The portfolio return series and consequently the inference based on these returns are often sensitive with respect to the choice parameters. Because of these variations, we exhaust a total of 320 different combinations. In our case, we have to decide on the estimation window type (rolling vs. expanding), the length of the estimation window (10, 15 or 20 years, when using a rolling window), the weighting scheme (value vs. equally weighted) and the factor cutoff (decile vs. Fama and French (1993) style 30/40/30 (FF style)). We look at different size subsets in isolation. Furthermore, we consider winsorizing at the 1% and 99% and at the 5% and 95% percentile separately. Importantly, we perform the analysis based on a "true" out-of-sample test in the spirit of McLean and Pontiff (2016), in which we condition the estimation and prediction only on FC published at the respective dates. In particular, we take the publication date and include the FC for the first time with the start of the next calender year. Figure A.1 illustrates this graphically, showing how FC become available over time only gradually. Additionally, we run the analysis with all FC present ata each point in time. As these choices impact the results in a significant manner, we exhaust all possible combinations.

Table A.1 provides a performance comparison of the 320 specifications. Panel A reveals that the adaptive Lasso selection performs in a large fraction of the cases better than an OLS-

Table A.1: Overview of long-short portfolio performance of adaptive Lasso vs. OLS-based selection
This table provides an overview of 320 different estimation and portfolio specifications; for more details on each of the specifications we refer to the appendix. The return series are long-short portfolio returns, obtained by sorting the prediction of each stock and each point in time. Panel A shows the number of cases in which the adaptive Lasso long-short portfolio realizes a higher mean return, Panel B counts the number of higher Sharpe ratios. "OOS" indicates whether the FC are only included in the post-publication period. Panels C and D indicate how many of these Sharpe ratios are significantly different when tested based on the Ledoit and Wolf (2008) test with different significance levels.

Panel A: Comparison of mean returns (# adaptive Lasso > OLS)

OOS	Large	Mid	Small	Large + Mid	All	Total
Yes	25 of 32 (78%)	20 of 32 (62%)	29 of 32 (91%)	22 of 32 (69%)	26 of 32 (81%)	122 of 160 (76%)
No	32 of 32 (100%)	22 of 32 (69%)	25 of 32 (78%)	29 of 32 (91%)	19 of 32 (59%)	127 of 160 (79%)
Total	57 of 64 (89%)	42 of 64 (66%)	54 of 64 (84%)	51 of 64 (80%)	45 of 64 (70%)	249 of 320 (78%)

Panel B: Sharpe ratio comparison (# adaptive Lasso > OLS)

OOS	Large	Mid	Small	Large + Mid	All	Total
Yes	25 of 32 (78%)	15 of 32 (47%)	31 of 32 (97%)	24 of 32 (75%)	21 of 32 (66%)	116 of 160 (72%)
No	29 of 32 (91%)	12 of 32 (38%)	31 of 32 (97%)	24 of 32 (75%)	21 of 32 (66%)	117 of 160 (73%)
Total	54 of 64 (84%)	27 of 64 (42%)	62 of 64 (97%)	48 of 64 (75%)	42 of 64 (66%)	233 of 320 (73%)

Panel C: Number of significantly higher Sharpe ratios of adaptive Lasso vs. OLS

Level	Large	Mid	Small	Large + Mid	All	Total
10%	2 of 64 (3.1%)	0 of 64 (0.0%)	15 of 64 (23.4%)	1 of 64 (1.6%)	0 of 64 (0.0%)	18 of 320 (5.6%)
5%	0 of 64 (0.0%)	0 of 64 (0.0%)	11 of 64 (17.2%)	0 of 64 (0.0%)	0 of 64 (0.0%)	11 of 320 (3.4%)
1%	0 of 64 (0.0%)	0 of 64 (0.0%)	2 of 64 (3.1%)	0 of 64 (0.0%)	0 of 64 (0.0%)	2 of 320 (0.6%)

Panel D: Number of significantly higher Sharpe ratios of OLS vs. adaptive Lasso

Level	Large	Mid	Small	Large + Mid	All	Total
10%	0 of 64 (0.0%)	2 of 64 (3.1%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	2 of 320 (0.6%)
5%	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 320 (0.0%)
1%	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 64 (0.0%)	0 of 320 (0.0%)

based selection when simply comparing differences in means, as this is true in 249 out of 320 cases. The mean difference is 1.2% in annualized terms (not reported). This result is not driven by any specific size group, as all sub-samples reveal on average higher means for an adaptive Lasso-based selection. The differences between conditioning only on published FC or does not yield much different results overall, although we can observe variations within size groups. This indicates that a potential forward-looking bias affects both methods, if at all, in a similar way. The higher average returns translate into higher Sharpe ratios for the adaptive Lasso-based approach as documented in Panel B of Table A.1. In total, we observe in 73% of the specifications a higher SR, which are only in case of mid caps below 50%. Finally, we test for each specification for a significant difference in SR. The relative frequency of significantly higher SR in case of the adaptive Lasso is exhibited in Panel C, the results of the OLS counterpart in Panel D. It reveals that only very few SR differ significantly. The SR of the adaptive Lasso long-short portfolios are in 5.6% of the cases significantly higher at a 10% level than the comparable OLS measure. Most of the significant tests are a observed for the small cap sample. On the other hand, only 0.6% of the OLS-based portfolios manage to achieve significantly higher SR at the 10% level. For more details on each single portfolio specification, we refer to Tables A.7-A.11.

A.2 Lasso conditions

The **beta-min condition** reads,

$$\min_{j \in \mathcal{A}} |\beta_j| \gg \phi^2 \sqrt{q \log(p)/n},$$

where ϕ^2 is the restricted eigenvalue of matrix $\mathbf{X}$.

Moreover, the **neighborhood stability condition** is,

$$\|\hat{\Sigma}_{\mathcal{A}^c,\mathcal{A}} \hat{\Sigma}_{\mathcal{A}}^{-1} \text{sign}(\beta_1, ..., \beta q)\|_\infty \leq \theta \text{ for some } 0 < \theta < 1,$$

where $\|z\|_\infty = \max_j |z_j|$ and, $\hat{\Sigma}_{\mathcal{A}^c,\mathcal{A}}$ and $\hat{\Sigma}_{\mathcal{A}}$ are the lower and upper left part of the covariance matrix of $\mathbf{X}$. The upper left part represents the covariance matrix of all active variables (dimension $q \times q$) and the lower left the covariance matrix of the active with the inactive variables (dimension $(p-q) \times q$).

A.3 Signal-to-noise ratio

In order to calibrate the simulation to the desired signal-to-noise ratio, we set the volatility of the factors and idiosyncratic volatility as follows. The signal-to-noise ratio (SNR) is defined as:

$$SNR = \frac{\sigma^2_{\text{signal}}}{\sigma^2_{\text{noise}}}$$

and it is related to the r-squared as follows:

$$SNR = \frac{R^2}{1 - R^2} \tag{A.1}$$

Recalling equation (1.7) and ignoring the indices, we can write,

$$r = x'f + \eta,$$

with $f = \mu + \epsilon$
Hence, we can define the variance of the signal as:

$$\begin{aligned}\sigma^2_{\text{signal}} &= \text{Var}(x'\mu) \\ &= \sigma^2_x \mu'\mu\end{aligned}$$

Note that the μ is defined as a uniformly distributed random vector and the realized μ are fixed at the beginning of the simulation and can be treated as deterministic.
Furthermore, the variance of the noise can be stated as follows:

$$\begin{aligned}\sigma^2_{\text{noise}} &= \text{Var}(x'\epsilon) + \text{Var}(\eta) \\ &= \sigma^{\mathbf{2}}_{\mathbf{x}} \sigma^{\mathbf{2}}_{\epsilon} + \sigma^2_\eta \\ &= \sum_c^{|\mathcal{C}|} \sigma^2_{\epsilon,c}\sigma^2_x I_{c\in\mathcal{P}\cup\mathcal{R}} + \sigma^2_\eta \\ &= \sigma^2_{\epsilon,1}\sigma^2_x + (P+R-1)\sigma^2_{\epsilon_f}\sigma^2_x + \sigma^2_\eta\end{aligned}$$

The first line can be simplified according to equation (A.2) below as all terms involving $\text{Cov}(x,\epsilon)$, $E(\epsilon)$ and $E(x)$ collapse to zero. $\sigma_{\epsilon,1}$ is given by the data reflecting the long-term mean of the stock market GARCH volatility. σ_{ϵ_f} and σ_η are calibrated such that each part contributes equally to fit the desired signal-to-noise ratio ($\sigma^2_\eta = (P+R-1)\sigma^2_{\epsilon_f}$). The value of σ^2_η and $\sigma^2_{\epsilon_f}$ of the desired SNR or the desired R^2 follow then straightforwardly.

The variance of the product of two random variables X and Y can be expressed as follows:

$$\begin{aligned}\text{Var}(XY) =& E[X^2Y^2] - [E(XY)]^2 \\ =& \text{Cov}(X^2,Y^2) + E(X^2)E(Y^2) - [E(XY)]^2 \\ =& \text{Cov}(X^2,Y^2) + (\text{Var}(X) + [E(X)]^2)(\text{Var}(Y) + [E(Y)]^2) \\ & - [\text{Cov}(X,Y) + E(X)E(Y)]^2\end{aligned} \tag{A.2}$$

Moreover, we can show that the R^2 of a frequency with length T and a frequency comprising a fraction $\frac{T}{\tau}$ of it are related as follows, assuming that x does not change with the time horizon and all terms in σ^2_{noise} are treated as returns with zero auto-correlation:

$$\begin{aligned}\sigma^2_{\text{signal},T} &= \text{Var}(x'\mu) = \sigma^2_x \mu'\mu \\ \sigma^2_{\text{signal},\frac{T}{\tau}} &= \text{Var}(x'\frac{\mu}{\tau}) = \sigma^2_x \frac{\mu'\mu}{\tau^2} \\ \sigma^2_{\text{noise},\frac{T}{\tau}} &= \frac{1}{\tau}\sigma^2_{\text{noise},T}\end{aligned}$$

and hence,

$$\frac{SNR_T}{SNR_{\frac{T}{\tau}}} = \frac{\sigma^2_{\text{signal},T}}{\sigma^2_{\text{signal},\frac{T}{\tau}}} \frac{\sigma^2_{\text{noise},\frac{T}{\tau}}}{\sigma^2_{\text{noise},y}} = \frac{\mu^2\sigma_x^2}{\frac{1}{\tau^2}\mu^2\sigma_x^2} \frac{\frac{1}{\tau}\sigma^2_{noise,T}}{\sigma^2_{noise,T}} = \tau \tag{A.3}$$

Combining equations (A.3) and (A.1), the following relation holds:

$$\frac{R^2_{\frac{T}{\tau}}}{1-R^2_{\frac{T}{\tau}}}\tau = \frac{R^2_T}{1-R^2_T} \tag{A.4}$$

A.4 Firm characteristics

As described above, the data in the paper utilize the CRSP/Compustat Merged database. In particular, we use the monthly and daily Stock data from CRSP Stock/Security files. Weekly returns are calculated from daily returns provided by CRSP. We always calculate returns using Friday as the last day of the week. Additionally, market and factor returns as well as risk-free rate data are obtained from Kenneth French's Data Library. The firm characteristics constructed follow the methodology used in Green et al. (2017). Generally, we use yearly accounting data. The matching between CRSP and Compustat data is done through the provided link-table. The Compustat data are only aligned from the first valid fiscal year within the valid period of the Compustat-CRSP link. For example, if the link start date (LINKDT) is 10/25/2001 and the fiscal year end is 12/31/2001 (datadate), only data after 12/31/2001 are assigned to particular CRSP return series. We construct the single criteria as described in Tables A.3 and A.4. Some raw data used to construct the FC have missing data; we follow Green et al. (2017) in replacing the missing data points with zeros for the following raw data variables: *xrd, emp, dp, rect, invt, dvt, che, nopi* and *at.*

A.5 Code

Our code is available upon request via bitbucket.org (a Git VCS); please make requests by email. It is all written in Python 3.x and should be compatible on win and on ux systems. Be aware the simulations as specified above are memory/RAM intensive; in order to run the main simulation, at least 90GB of available RAM are required.

A.6 Additional tables

Abbreviation	Descrption	Data source
DLRET	Delisting Return	CRSP
PRC	Closing Price or Bid/Ask Average	CRSP
RET	Holding Period Return	CRSP
SHRCD	Share Code	CRSP
SHROUT	Number of Shares Outstanding	CRSP
SICCD	Standard Industrial Classification Code	CRSP
VOL	Share Volume	CRSP
act	Current assets	Compustat
at	Total assets	Compustat
capx	Capital expenditure	Compustat
ceq	Book value of equity	Compustat
che	Cash and cash equivalents	Compustat
cogs	Costs of goods sold	Compustat
dlc	Debt in Current Liabilities - Total	Compustat
dltt	Long-Term Debt - Total	Compustat
dp	Depreciation and amortization	Compustat
dvt	Total dividends	Compustat
ebit	Earnings Before Interest and Taxes	Compustat
emp	Employees	Compustat
ib	Annual income before extraordinary items	Compustat
invt	Inventories	Compustat
lct	Current liabilities	Compustat
lt	Total liabilities	Compustat
nopi	Nonoperating Income/Expense	Compustat
oancf	Operating Activities Net Cash Flow	Compustat
ppegt	Gross Property, Plant and Equipment	Compustat
rect	Accounts receivable	Compustat
sale	Sales	Compustat
txp	Income Taxes Payable	Compustat
xrd	Research and Development Expense	Compustat
xsga	Selling, General and Administrative Expense	Compustat

Table A.2: Overview of raw data used and the abbreviations with the corresponding description and data source.

ID	Acronym	Name	Description	Reference
1	beta	Beta	Measured based on 3 years (min 52 weeks) weekly excess returns with standard ols ($y = c + \beta x$)	Fama and MacBeth (1973)
2	beta_sq	Beta squared	Simply obtained by squaring the β based on the beta from # 1	Fama and MacBeth (1973)
3	retvol	Volatility	Volatility is measured by the standard deviation of daily returns of the previous months	Ang et al. (2006)
4	maxret	Maximum return	Maximum return is defined over the max of the daily returns in month $t-1$	Bali et al. (2011)
5	idiovol	Idiosyncratic volatility	Calculated based on the residuals of regression in # 1	Ali et al. (2003)
6	mom1m	1-month momentum	Return in month $t-1$	Jegadeesh (1990)
7	mom6m	6-month momentum	Cumulative return over 5 months ending in $t-2$	Jegadeesh and Titman (1993)
8	mom12m	12-month momentum	Cumulative return over 11 months ending in $t-2$	Jegadeesh (1990)
9	mom36m	36-month momentum	Cumulative return over 24 months ending in $t-13$	Bondt and Thaler (1985)
10	mve	Market capitalization (size)	log of (SHROUT$\times$PRC)	Banz (1981)
11	ep	Earnings-to-price	Earnings per share	Basu (1977)
12	dy	Dividends-to-price	Yearly dividends (dvt) divided by market cap at fiscal year	Litzenberger and Ramaswamy (1979)
13	bm	Book-to-market	Book value of equity (ceq) divided by market cap	Rosenberg et al. (1985)
14	lev	Leverage	Total liabilities (lt) divided by market cap	Bhandari (1988)
15	currat	Current ratio	Current assets (act) divided by current liabilities (lct)	Ou and Penman (1989)
16	pchcurrat	Pct change in current ratio	Percentage change in currat from year $t-1$ to t	Ou and Penman (1989)
17	quick	Quick ratio	Current assets (act) minus inventory (invt), divided by current liabilities (lct)	Ou and Penman (1989)
18	pchquick	Pct change in quick ratio	Percentage change in quick from year $t-1$ to t	Ou and Penman (1989)
19	salecash	Sales-to-cash	Annual sales (sale) divided by cash and cash equivalents (che)	Ou and Penman (1989)
20	salerec	Sales-to-receivables	Annual sales (sale) divided by accounts receivable (rect)	Ou and Penman (1989)
21	saleinv	Sales-to-inventory	Annual sales (sale) divided by total invetory (invt)	Ou and Penman (1989)
22	pchsaleinv	Pct change in sales-to-inventory	Percentage change in saleinv from year $t-1$ to t	Ou and Penman (1989)
23	cashdebt	Cashflow-to-debt	Earnings before depreciation and extraordinary items (ib + dp) divided by avg total liabilities (lt)	Ou and Penman (1989)
24	baspread	Illiquidity (bid-ask-spread)	Monthly avg of daily bid-ask spread divided by avg of daily bid-ask spread	Amihud and Mendelson (1989)
25	depr	Depreciation-to-gross PP&E	Depreciation expense (dp) divided by gross PPE (ppegt)	Holthausen and Larcker (1992)
26	pchdepr	Pct change in Depreciation-to-gross PP&E	Percentage change in depr from year $t-1$ to t	Holthausen and Larcker (1992)
27	mve_ia	Industry-adjusted firm size	Log market caps are adjusted by log of the mean of the industry	Asness et al. (2000)
28	cfp_ia	Industry-adjusted cashflow-to-price	Industry adjusted cash flow-to-price ratio equal weighted average	Asness et al. (2000)
29	bm_ia	Industry-adjusted book-to-market	Industry adjusted book-to-market equal weighted average	Asness et al. (2000)
30	sgr	Annual sales growth	Percentage change in sales from year $t-1$ to t	Lakonishok et al. (1994)
31	ipo	IPO	Indicated by 1 if first 12 months available on CRSP monthly file	Loughran and Ritter (1995)
32	divi	Dividend initiation	Indicated by 1 if company pays dividends but did not in prior year.	Michaely et al. (1995)
33	divo	Dividend omission	Indicated by 1 if company does not pay dividends but did in prior year.	Michaely et al. (1995)
34	sp	Sales-to-price	Annual sales (sale) divided by market cap	Barbee Jr et al. (1996)
35	acc	WC accruals	(ib) - (oancf)/(at), if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Sloan (1996)
36	turn	Share turnover	Average monthly trading volume for the three months $t-3$ to $t-1$ divided by SHROUT at $t-1$	Datar et al. (1998)
37	pchsale_pchinvt	Delta pct change sales vs. inventory	Difference of percentage changes in sales (sale) and inventory (invt)	Abarbanell and Bushee (1997)
38	pchsale_pchrect	Delta pct change sales vs. receivables	Difference of percentage changes in sales (sale) and receivables (rect)	Abarbanell and Bushee (1997)
39	pchcapx_ia	CAPEX	Industry adjusted (two digit SIC) fiscal year mean adjusted percentage change in capital expenditures (capx)	Abarbanell and Bushee (1997)
40	pchgm_pchsale	Delta pct gross margin vs. sales	Annual percentage change in gross margin (sale minus cogs) minus percentage change in sales (sale)	Abarbanell and Bushee (1997)

Table A.3: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values.

ID	Acronym	Name	Description	Reference
41	pchsale_pchxsga	Delta pct sales vs. SGaA	Annual percentage change in sales (sale) minus percentage change in SGaA (xsga)	Abarbanell and Bushee (1997)
42	dolvol	Dollar trading volume	Log of trading volume times price per share from month t-2	Chordia et al. (2001)
43	std_dolvol	Volatility trading volume	Monthly standard deviation of daily trading volume	Chordia et al. (2001)
44	std_turn	Volatility turnover	Monthly standard deviation of daily share turnover	Chordia et al. (2001)
45	chinv	Change in inventory	First difference of inventory (invt) divided by total assets	Thomas and Zhang (2002)
46	pchemp_ia	Industry-adjusted pch in employees	Industry adjusted percentage change in employees	Asness et al. (2000)
47	cfp	Cashflow-to-price	Operating cash flows (oancf) scaled by market capitalization (fiscal year end)	Desai et al. (2004)
48	rd	R&D Increase	If annual increase in R&D expenses (xrd) scaled by total assets (at) >0.05, 1, else 0	Eberhart et al. (2004)
49	lgr	Pct change in long-term debt	Annual percentage change in long term debt (lt)	Richardson et al. (2005)
50	egr	Pct change in book equity	Annual percentage change in book equity (ceq)	Richardson et al. (2005)
51	rd_sale	R&D-to-sales	R&D expenses(xrd) scaled by sales (sale)	Guo et al. (2006)
52	rd_mve	R&D-to-market	R&D expenses(xrd) scaled by market cap	Guo et al. (2006)
53	chg_mom6m	change in mom6m	difference of mom6m measured at t and $t-6$	Gettleman and Marks (2006)
54	hire	Pct change in employee	Annual percentage change in employee (emp)	Belo et al. (2014)
55	agr	Asset growth	Annual percentage change in assets (at)	Cooper et al. (2008)
56	cashpr	Cash productivity	Market cap plus long term debt (dltt) minus assets (at) divided by cash (che)	Chandrashekar and Rao (2009)
57	gma	Gross-profitability	Sales (sale) minus costs of goods sold (cogs) divided by one-year lagged assets(at)	Novy-Marx (2013)
58	cash	Cash-to-assets	Cash (che) divided by assets(at)	Palazzo (2012)
59	pctacc	Accruals-to-income	(ib) minus (oancf) divided by abs ((ib)), when (ib) equals 0, it is set to 0.01, if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Hafzalla et al. (2011)
60	absacc	Absolut accruals	Absolute value of acc	Bandyopadhyay et al. (2010)
61	roic	Return on invested capital	Earnings before interest and taxes (ebit) - non-operating income (nopi), divided by non-cash enterprise value (ceq+lt-che)	Brown and Rowe (2007)
62	grcapex	Pct change in two year CAPX	Percentage change in two year capital expenditure (capx)	Anderson and Garcia (2006)
63	tang	Debt capacity-to-firm-tangability	(Cash (che) + 0.715 receivables (rect) + 0.547 inventory(invt) + 0.535 (ppegt))/ total assets (at)	Hahn and Lee (2009)
64	chshrout	Change in shares-outstanding	Yearly percentage change in outstanding shares (SHROUT)	Pontiff and Woodgate (2008)
65	invest	CAPEX and inventory	Yearly difference in gross property, plant and equipment (ppegt) + diff in (invt) / (t-1) total assets (at)	Chen and Zhang (2010)
66	age	Years since CS coverage	Years since first compustat coverage years(datadate - min(datadate))	Jiang et al. (2005)
67	chpm_ia	Industry-adjusted change in profit margin	Industry adjusted (two-digit SIC) change in profit margin (ib/sale)	Soliman (2008)
68	chato_ia	Industriy-adjusted change in asset turnover	Industry adjusted (two-digit SIC) change in asset turnover (sale/at)	Soliman (2008)

Table A.4: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values.

FC	coverage	mean	vw_mean	std	min	5%	50%	95%	max	data_start
absacc	0.99	0.11	0.07	0.38	0.00	0.00	0.06	0.38	105.00	1965-02
acc	0.99	-0.02	0.01	0.40	-105.00	-0.26	-0.01	0.25	59.85	1965-02
age	0.99	10.05	17.69	9.81	0.00	0.00	7.00	31.00	52.00	1965-02
agr	0.90	0.18	0.16	1.49	-1.00	-0.23	0.08	0.76	1071.38	1965-02
beta	0.98	0.90	1.01	0.61	-5.13	0.04	0.85	1.96	10.15	1965-02
beta_sq	0.98	1.18	1.20	1.46	0.00	0.01	0.74	3.85	103.04	1965-02
bm	0.99	0.81	0.55	2.91	-906.64	0.09	0.62	2.24	230.91	1965-02
bm_ia	0.97	0.00	-0.21	2.89	-906.50	-0.82	-0.11	1.22	226.14	1965-02
cash	0.99	0.15	0.11	0.20	-0.07	0.00	0.07	0.62	1.00	1965-02
cashdebt	0.99	-0.22	0.08	2.78	-469.50	-1.45	0.01	0.43	580.33	1965-02
cashpr	0.98	-2.95	17.47	1664.18	-539194.36	-51.49	-0.96	41.72	274867.00	1965-02
cfp	0.99	0.03	0.06	1.40	-187.75	-0.54	0.05	0.53	307.25	1965-02
cfp_ia	0.57	0.00	0.01	0.61	-55.07	-0.35	-0.00	0.34	33.98	1988-01
chato_ia	0.89	-0.00	-0.01	0.73	-175.98	-0.39	0.00	0.37	173.70	1965-02
chg_mom6m	0.97	-0.00	-0.00	0.62	-35.73	-0.77	-0.01	0.78	67.74	1965-02
chinv	0.90	0.01	0.01	0.14	-44.90	-0.06	0.00	0.11	0.96	1965-02
chpm_ia	0.79	-0.02	-0.10	27.11	-5349.32	-0.40	0.00	0.35	7404.14	1965-02
chshrout	0.98	0.12	0.17	0.59	-1.00	-0.06	0.01	0.89	183.67	1965-02
currat	0.85	3.21	1.91	16.04	0.00	0.70	2.08	7.70	4928.57	1965-02
depr	0.89	0.13	0.09	1.52	-0.44	0.03	0.08	0.29	558.00	1965-02
divi	0.99	0.03	0.01	0.17	0.00	0.00	0.00	0.00	1.00	1965-02
divo	0.99	0.03	0.01	0.16	0.00	0.00	0.00	0.00	1.00	1965-02
dolvol	0.92	15.50	19.27	3.00	0.00	10.84	15.37	20.53	26.57	1965-02
dy	0.99	0.02	0.03	0.12	-5.78	0.00	0.00	0.08	15.97	1965-02
egr	0.90	0.14	0.16	10.29	-2196.50	-0.47	0.08	0.91	2325.36	1965-02
ep	0.99	-0.04	0.06	2.36	-917.01	-0.41	0.05	0.21	29.43	1965-02
gma	0.90	0.38	0.40	0.70	-21.39	0.01	0.33	1.02	545.90	1965-02
grcapex	0.74	12.59	232.28	230.68	-16133.00	-34.59	0.27	82.50	12141.00	1965-02
hire	0.87	0.22	0.08	15.70	-1.00	-0.29	0.02	0.60	5665.67	1965-02
idiovol	0.98	0.07	0.04	0.04	0.00	0.03	0.06	0.15	1.63	1965-02
invest	0.82	0.10	0.10	0.63	-7.42	-0.11	0.05	0.40	447.24	1965-02
ipo	0.99	0.01	0.00	0.09	0.00	0.00	0.00	0.00	1.00	1965-02
lev	0.98	2.73	1.77	21.24	0.00	0.05	0.73	10.42	6332.38	1965-02
lgr	0.90	0.41	0.21	13.27	-1.00	-0.31	0.08	1.25	4248.62	1965-02
maxret	0.99	0.08	0.04	0.09	-0.42	0.01	0.05	0.21	19.00	1965-02
mom12m	0.97	0.15	0.18	0.72	-1.00	-0.60	0.05	1.13	53.50	1965-02
mom1m	0.99	0.01	0.01	0.18	-0.97	-0.22	0.00	0.27	24.00	1965-02
mom36m	0.82	0.36	0.45	1.26	-1.00	-0.69	0.15	2.01	145.00	1965-02
mom6m	0.99	0.06	0.08	0.44	-0.99	-0.45	0.02	0.67	66.94	1965-02
mve	0.99	18.49	22.46	2.20	10.56	15.13	18.35	22.30	27.27	1965-02
mve_ia	0.97	-1.80	1.83	2.01	-10.19	-5.03	-1.84	1.56	5.81	1965-02
pchcapx_ia	0.81	0.00	-0.30	41.43	-624.05	-1.62	0.00	1.25	25564.33	1965-02
pchcurrat	0.77	0.16	0.02	11.21	-1.00	-0.51	-0.02	0.79	5151.98	1965-02
pchdepr	0.81	0.15	0.03	4.07	-26.02	-0.36	-0.00	0.59	755.73	1965-02
pchemp_ia	0.84	0.00	-0.04	13.76	-15.40	-0.27	0.00	0.14	5652.65	1965-02
pchgm_pchsale	0.89	-0.51	-0.06	44.77	-11880.89	-0.53	-0.00	0.45	2540.93	1965-02
pchquick	0.77	0.20	0.04	11.30	-1.00	-0.55	-0.02	0.97	5151.98	1965-02
pchsale_pchinvt	0.74	-0.47	-0.49	66.69	-18507.67	-0.85	0.01	0.69	5201.77	1965-02
pchsale_pchrect	0.87	-0.13	0.02	44.32	-6021.57	-0.72	0.00	0.56	11879.56	1965-02
pchsale_pchxsga	0.76	-0.22	-1.20	165.31	-62906.70	-0.35	-0.00	0.40	8842.23	1965-02
pchsaleinv	0.74	0.58	0.19	76.74	-121.04	-0.54	0.01	0.94	28764.03	1965-02
pctacc	0.99	-2.13	-0.53	212.25	-54294.60	-7.85	-0.21	6.12	14369.90	1965-02
quick	0.85	2.48	1.38	15.73	0.00	0.39	1.33	6.70	4928.57	1965-02
rd	0.99	0.16	0.17	0.36	0.00	0.00	0.00	1.00	1.00	1965-02
rd_mve	0.99	0.03	0.02	0.11	-0.03	0.00	0.00	0.15	19.96	1965-02
rd_sale	0.98	1.05	0.13	72.99	-90.38	0.00	0.00	0.26	25684.40	1965-02
ret	1.00	0.01	0.02	0.18	-1.00	-0.22	0.00	0.27	24.00	1965-01
retvol	0.99	0.03	0.02	0.03	0.00	0.01	0.03	0.08	4.08	1965-02
roic	0.95	-0.02	0.13	0.48	-5.14	-0.69	0.07	0.30	0.52	1965-02
salecash	0.99	102.00	51.37	1849.09	-1230.91	0.28	11.38	204.41	282280.67	1965-02
saleinv	0.98	31.90	28.01	923.82	-35.44	0.00	5.96	85.93	295353.00	1965-02
salerec	0.98	14.92	11.21	187.28	-21796.00	0.09	5.84	41.05	47246.00	1965-02
sgr	0.89	0.58	0.23	38.78	-110.64	-0.27	0.10	0.82	11879.50	1965-02
sp	0.98	2.70	1.31	6.16	-7.39	0.08	1.20	9.49	671.35	1965-02
tang	0.90	0.66	0.64	0.22	0.00	0.30	0.67	0.95	27.44	1965-02
turn	0.92	0.10	0.08	0.19	0.00	0.01	0.05	0.34	67.89	1965-02

Table A.5: Data description of raw data before any data adjustments (z-scores, winsorizing etc.) for the period 1965-01-01 to 2014-12-31, based on 2377145 obs. Note also that it is the summary over all periods for each FC. The column vw-mean measures the market value weighted average. For that, we calculate vw averages for each period, which are then used to obtain an equally weighted average over all periods. Data start indicates the date on which the FC was calculated for at least one stock. Note that we display *roic* winsorized at 2.5% to control for data points which are scaled by numbers close to zero.

FC	VWAL	AL	OLS	Lasso	FC	VWAL	AL	OLS	Lasso
chg_mom6m	-6.07	-2.27	-2.84** (-2.48)	-2.02	pchcurrat	-0.04	0.00	-0.22 (-0.51)	0.00
mom6m	3.98	1.33	2.06 (1.55)	1.06	mom12m	0.00	1.71	1.31 (1.08)	1.80
ep	2.64	1.14	1.19** (2.25)	1.06	agr	0.00	-1.22	-1.85*** (-3.32)	-1.05
dolvol	-1.91	-1.55	-1.35* (-1.88)	-1.22	currat	0.00	-0.98	-2.53*** (-2.68)	-1.00
mom1m	-1.51	-2.37	-2.40*** (-2.99)	-2.24	dy	0.00	-0.54	-1.09** (-2.43)	-0.64
egr	-1.37	-0.73	-0.72*** (-2.81)	-0.78	chpm_ia	0.00	0.43	0.68* (1.94)	0.53
acc	-1.33	-0.51	-0.96* (-1.80)	-0.50	depr	0.00	0.41	0.51 (1.55)	0.45
gma	1.22	0.24	0.74* (1.65)	0.52	cash	0.00	0.23	0.28 (0.41)	0.38
cashpr	-1.19	-0.24	-0.63 (-1.56)	-0.41	bm	0.00	0.00	0.51 (0.75)	0.32
rd_mve	1.09	1.68	1.56*** (2.88)	1.57	mve	0.00	0.00	-0.57 (-0.56)	-0.32
pctacc	0.72	0.00	0.17 (0.47)	0.00	pchsale_pchinvt	0.00	0.00	0.69 (1.26)	0.26
cashdebt	0.62	1.28	1.65** (2.24)	1.24	idiovol	0.00	0.00	-0.80 (-1.24)	-0.24
mom36m	-0.59	0.00	-0.18 (-0.34)	-0.10	pchsale_pchxsga	0.00	0.00	-0.36 (-1.37)	-0.21
chinv	-0.58	0.00	-0.00 (-0.01)	-0.10	saleinv	0.00	0.00	0.15 (0.88)	0.16
pchdepr	0.50	0.00	0.35 (1.44)	0.24	maxret	0.00	0.00	-0.02 (-0.03)	-0.12
sp	0.48	0.96	0.78* (1.86)	0.73	chshrout	0.00	0.00	-0.23 (-0.75)	-0.12
chato_ia	0.48	0.00	0.16 (0.64)	0.05	pchgm_pchsale	0.00	0.00	0.16 (0.74)	0.08
beta	0.33	0.84	1.26 (1.58)	0.82	absacc	0.00	0.00	0.18 (0.61)	0.07
sgr	-0.31	0.00	0.06 (0.14)	0.00	grcapex	0.00	0.00	-0.13 (-0.55)	-0.03
retvol	-0.29	-1.91	-1.88*** (-2.81)	-1.68	salecash	0.00	0.00	-0.09 (-0.46)	-0.03
salerec	0.28	0.00	0.60 (1.50)	0.29	invest	0.00	0.00	-0.03 (-0.07)	-0.01
tang	-0.28	0.00	0.05 (0.13)	0.00	lgr	0.00	0.00	0.59* (1.67)	0.00
lev	0.09	0.33	0.93* (1.67)	0.48	quick	0.00	0.00	1.39* (1.67)	0.00
pchcapx_ia	0.06	0.00	0.00 (0.02)	0.00					

Table A.6: FC Selection Large + Mid Stocks: The table exhibits the regression results of OLS, Lasso, adaptive Lasso (AL), and the value-weighted adaptive Lasso (VWAL) for each FC, sorted by the absolute value in the order of the VWAL, AL, Lasso and t-values of the OLS coefficients. The numbers in the brackets of the OLS column represent t-values. The table displays only FC having an abs adaptive Lasso or Lasso coefficient greater than zero, or an abs OLS t-value greater than 1.645. Accordingly, the following FC used in the selection regression are not represented: *hire*, *age*, *pchquick*, *pchsale_pchrect*, *pchsaleinv*, *rd_sale*, *cfp*, *rd*, *divo*, *divi*, *mve_ia*, *turn*, *bm_ia*. Note the coefficients are scaled by a factor of 1000. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. Cells with blue background color indicate selected status. The sample period spans Jun 1974 until Dec 2014.

FC OOS	Window	Type	Buckets	Weights	Win	Mean		SR		SR Test	StdErr	
						AdaLasso	OLS	AdaLasso	OLS	AdaLasso-OLS	AdaLasso	OLS
No	120	rolling	Decile	EQ	0.01	37.59	36.81	2.22	2.01	1.19	16.95	18.30
					0.05	33.80	33.73	2.07	2.15	-0.48	16.33	15.68
				VW	0.01	12.99	13.56	0.65	0.62	0.17	20.01	21.86
					0.05	15.72	14.16	0.75	0.72	0.15	21.00	19.62
			FF Style	EQ	0.01	20.44	19.48	1.78	1.58	1.19	11.51	12.30
					0.05	19.64	18.08	1.77	1.73	0.24	11.08	10.44
				VW	0.01	7.95	7.33	0.60	0.54	0.42	13.19	13.56
					0.05	7.27	7.78	0.53	0.64	-0.62	13.68	12.22
	180	expanding	Decile	EQ	0.01	38.29	38.35	2.11	2.14	-0.40	18.12	17.94
					0.05	34.48	34.16	2.06	1.97	1.09	16.71	17.38
				VW	0.01	11.79	13.12	0.58	0.64	-0.69	20.41	20.47
					0.05	12.14	10.70	0.59	0.54	0.46	20.43	19.99
			FF Style	EQ	0.01	20.20	19.88	1.71	1.72	-0.15	11.82	11.56
					0.05	20.05	19.60	1.75	1.64	1.48	11.46	11.92
				VW	0.01	7.44	5.68	0.56	0.43	1.33	13.17	13.18
					0.05	7.29	7.66	0.54	0.55	-0.16	13.61	13.84
	180	rolling	Decile	EQ	0.01	35.14	37.04	2.13	2.10	0.18	16.47	17.66
					0.05	33.25	35.10	2.08	2.03	0.25	16.02	17.27
				VW	0.01	13.19	15.33	0.66	0.76	-0.53	19.94	20.12
					0.05	12.60	11.04	0.60	0.51	0.53	21.03	21.64
			FF Style	EQ	0.01	19.72	19.54	1.74	1.64	0.53	11.34	11.92
					0.05	19.72	19.40	1.80	1.65	0.94	10.95	11.75
				VW	0.01	7.00	6.44	0.50	0.50	-0.04	14.12	12.77
					0.05	7.97	7.08	0.55	0.51	0.34	14.47	14.01
	240	rolling	Decile	EQ	0.01	38.61	39.35	2.11	1.99	0.70	18.27	19.80
					0.05	32.61	34.07	1.90	1.89	0.09	17.14	18.02
				VW	0.01	17.05	15.96	0.80	0.74	0.32	21.43	21.68
					0.05	15.79	18.54	0.75	0.88	-0.72	21.10	21.10
			FF Style	EQ	0.01	20.21	20.71	1.69	1.55	0.91	11.98	13.36
					0.05	19.46	19.28	1.67	1.59	0.63	11.65	12.10
				VW	0.01	7.64	7.55	0.49	0.56	-0.34	15.51	13.54
					0.05	9.29	10.56	0.64	0.75	-0.69	14.60	14.07
Yes	120	rolling	Decile	EQ	0.01	25.32	24.70	1.36	1.34		18.55	18.40
					0.05	22.64	21.03	1.33	1.29		17.02	16.26
				VW	0.01	7.00	3.33	0.32	0.16	1.08	22.03	20.88
					0.05	9.28	6.67	0.45	0.35	0.60	20.75	19.09
			FF Style	EQ	0.01	13.01	13.02	1.03	1.04		12.66	12.51
					0.05	13.12	12.61	1.15	1.16		11.39	10.86
				VW	0.01	3.07	2.36	0.20	0.15	0.31	15.19	15.61
					0.05	3.24	2.18	0.23	0.17	0.42	13.78	12.56
	180	expanding	Decile	EQ	0.01	29.43	27.40	1.38	1.32	0.76	21.39	20.80
					0.05	26.44	26.11	1.40	1.40	-0.04	18.94	18.65
				VW	0.01	7.64	3.37	0.35	0.16	1.23	22.06	21.44
					0.05	8.57	6.47	0.40	0.29	0.79	21.44	22.10
			FF Style	EQ	0.01	14.66	13.15	1.09	1.00	0.99	13.43	13.10
					0.05	15.25	14.67	1.20	1.17	0.35	12.69	12.53
				VW	0.01	4.44	3.04	0.29	0.21	0.51	15.29	14.75
					0.05	4.98	3.04	0.35	0.21	1.44	14.12	14.81
	180	rolling	Decile	EQ	0.01	28.61	29.01	1.50	1.55	-0.46	19.09	18.67
					0.05	24.08	24.21	1.47	1.46	0.10	16.36	16.63
				VW	0.01	8.70	8.33	0.41	0.43	-0.13	21.39	19.34
					0.05	8.23	7.90	0.41	0.38	0.18	20.31	21.01
			FF Style	EQ	0.01	15.29	15.27	1.24	1.28	-0.28	12.30	11.92
					0.05	14.40	15.19	1.32	1.37	-0.36	10.93	11.06
				VW	0.01	4.43	4.37	0.29	0.33	-0.21	15.18	13.22
					0.05	3.93	5.13	0.27	0.38	-0.93	14.29	13.45
	240	rolling	Decile	EQ	0.01	33.91	32.25	1.69	1.58	1.35	20.04	20.36
					0.05	28.72	28.55	1.56	1.49	0.62	18.45	19.11
				VW	0.01	13.43	10.46	0.58	0.45	0.87	23.26	23.13
					0.05	9.19	12.93	0.40	0.55	-0.99	22.80	23.64
			FF Style	EQ	0.01	17.27	16.55	1.35	1.30	0.61	12.77	12.77
					0.05	17.17	17.04	1.44	1.36	0.72	11.94	12.49
				VW	0.01	6.31	6.15	0.38	0.39	-0.11	16.71	15.70
					0.05	5.95	4.11	0.37	0.27	0.99	15.86	15.29

Table A.7: Long-short portfolio performance, All stocks specifications. The table shows selected portfolio characteristics in annualized figures. "FC OOS" indicates whether the FC are only included in the post-publication period. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test. The sample period ranges from Jan 1972 until Dec 2014. DiFF Styleerent window lengths are not forced on the same out-of-sample start point; hence, each time window has different return series lengths. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

FC OOS	Window	Type	Buckets	Weights	Win	Mean		SR		SR Test	StdErr	
						AdaLasso	OLS	AdaLasso	OLS	AdaLasso-OLS	AdaLasso	OLS
No	120	rolling	Decile	EQ	0.01	11.90	8.01	0.56	0.47	0.52	21.30	17.14
					0.05	12.96	8.49	0.58	0.45	0.88	22.17	18.67
				VW	0.01	8.70	5.20	0.43	0.30	0.60	20.33	17.29
					0.05	9.90	5.36	0.47	0.28	1.13	21.15	18.95
			FF Style	EQ	0.01	7.71	3.60	0.56	0.31	1.46	13.86	11.65
					0.05	7.80	4.18	0.56	0.35	1.52	13.90	12.08
				VW	0.01	6.10	1.70	0.48	0.15	1.67*	12.69	11.00
					0.05	6.14	3.05	0.49	0.27	1.48	12.42	11.32
	180	expanding	Decile	EQ	0.01	12.75	9.14	0.67	0.58	0.33	18.98	15.77
					0.05	12.40	10.06	0.64	0.57	0.33	19.42	17.60
				VW	0.01	13.39	8.42	0.71	0.51	0.70	18.82	16.51
					0.05	13.05	9.59	0.67	0.52	0.66	19.49	18.28
			FF Style	EQ	0.01	8.15	6.13	0.66	0.61	0.17	12.31	9.98
					0.05	7.66	7.16	0.58	0.61	-0.16	13.24	11.73
				VW	0.01	8.36	6.58	0.68	0.61	0.25	12.32	10.73
					0.05	7.78	7.23	0.60	0.59	0.05	13.06	12.34
	180	rolling	Decile	EQ	0.01	12.03	7.18	0.55	0.33	1.18	21.88	22.06
					0.05	10.46	6.75	0.46	0.35	0.62	22.89	19.54
				VW	0.01	12.22	4.81	0.57	0.22	1.61	21.42	21.61
					0.05	11.17	4.88	0.51	0.26	1.20	22.00	18.75
			FF Style	EQ	0.01	5.60	3.19	0.38	0.22	0.91	14.74	14.50
					0.05	5.46	3.54	0.37	0.26	0.62	14.80	13.39
				VW	0.01	4.76	1.36	0.36	0.10	1.42	13.40	14.13
					0.05	5.30	1.50	0.39	0.12	1.31	13.52	12.25
	240	rolling	Decile	EQ	0.01	11.71	4.74	0.54	0.26	1.19	21.56	18.03
					0.05	13.01	11.19	0.60	0.61	-0.02	21.61	18.48
				VW	0.01	13.99	4.25	0.63	0.22	1.55	22.04	19.30
					0.05	14.80	10.24	0.70	0.51	0.83	21.10	19.91
			FF Style	EQ	0.01	6.59	3.55	0.46	0.30	0.75	14.48	11.87
					0.05	6.87	6.28	0.47	0.49	-0.11	14.77	12.88
				VW	0.01	7.01	3.18	0.50	0.24	1.17	14.01	13.07
					0.05	6.83	6.19	0.49	0.49	0.00	13.94	12.64
Yes	120	rolling	Decile	EQ	0.01	7.57	7.36	0.35	0.37	-0.13	21.90	20.09
					0.05	7.57	5.07	0.35	0.25	0.61	21.60	20.01
				VW	0.01	7.85	6.03	0.38	0.30	0.41	20.68	20.02
					0.05	6.38	2.84	0.30	0.14	0.96	20.99	19.70
			FF Style	EQ	0.01	4.35	4.20	0.30	0.32	-0.13	14.47	13.00
					0.05	5.22	3.57	0.37	0.26	0.71	14.30	13.59
				VW	0.01	4.89	3.71	0.37	0.30	0.45	13.08	12.39
					0.05	5.08	2.85	0.39	0.22	1.17	12.89	12.96
	180	expanding	Decile	EQ	0.01	7.93	6.58	0.43	0.37	0.31	18.66	17.64
					0.05	8.33	7.53	0.45	0.41	0.27	18.42	18.39
				VW	0.01	9.38	9.07	0.47	0.47	0.03	19.78	19.34
					0.05	10.13	8.05	0.52	0.40	0.67	19.67	20.07
			FF Style	EQ	0.01	4.97	4.71	0.41	0.39	0.09	12.23	11.99
					0.05	4.96	5.06	0.40	0.40	-0.03	12.47	12.58
				VW	0.01	6.85	6.89	0.55	0.51	0.24	12.49	13.45
					0.05	5.65	5.19	0.46	0.40	0.38	12.37	13.13
	180	rolling	Decile	EQ	0.01	5.75	7.19	0.26	0.33	-0.55	22.34	21.74
					0.05	7.12	5.67	0.34	0.27	0.49	20.74	21.28
				VW	0.01	7.30	7.59	0.34	0.34	-0.02	21.76	22.48
					0.05	8.03	1.66	0.38	0.08	1.74*	21.38	21.23
			FF Style	EQ	0.01	2.40	3.71	0.16	0.25	-0.67	14.94	14.81
					0.05	3.40	3.72	0.24	0.26	-0.13	14.45	14.52
				VW	0.01	2.64	2.87	0.19	0.19	0.01	13.69	15.03
					0.05	3.35	1.84	0.25	0.13	0.70	13.56	14.29
	240	rolling	Decile	EQ	0.01	10.46	8.22	0.51	0.38	0.88	20.49	21.48
					0.05	11.15	9.88	0.55	0.50	0.32	20.29	19.91
				VW	0.01	13.47	8.48	0.60	0.37	1.38	22.41	22.82
					0.05	11.51	7.98	0.54	0.38	0.99	21.21	21.16
			FF Style	EQ	0.01	6.24	4.59	0.46	0.32	1.03	13.70	14.56
					0.05	6.02	5.45	0.43	0.39	0.31	13.94	14.09
				VW	0.01	7.86	5.32	0.57	0.35	1.46	13.81	15.13
					0.05	6.46	4.93	0.47	0.33	0.86	13.68	14.88

Table A.8: Long-short portfolio performance, Large cap stocks specifications. The table shows selected portfolio characteristics in annualized figures. "FC OOS" indicates whether the FC are only included in the post-publication period. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test. The sample period ranges from Jan 1972 until Dec 2014. Different window lengths are not forced on the same out-of-sample start point; hence, each time window has different return series lengths. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

FC OOS	Window	Type	Buckets	Weights	Win	Mean		SR		SR Test	StdErr	
						AdaLasso	OLS	AdaLasso	OLS	AdaLasso-OLS	AdaLasso	OLS
No	120	rolling	Decile	EQ	0.01	19.19	16.74	0.93	1.03	-0.43	20.69	16.31
					0.05	19.48	18.71	0.90	0.98	-0.39	21.63	19.09
				VW	0.01	18.78	16.21	0.92	1.01	-0.40	20.33	15.97
					0.05	19.17	18.57	0.89	0.97	-0.39	21.52	19.16
			FF Style	EQ	0.01	11.93	9.51	0.89	0.91	-0.11	13.48	10.43
					0.05	11.93	10.13	0.86	0.84	0.13	13.82	12.10
				VW	0.01	11.65	9.19	0.88	0.89	-0.03	13.18	10.33
					0.05	11.76	10.15	0.86	0.83	0.19	13.61	12.29
	180	expanding	Decile	EQ	0.01	17.89	18.47	0.89	1.04	-1.01	20.15	17.73
					0.05	16.90	17.48	0.81	0.84	-0.26	20.90	20.92
				VW	0.01	17.04	17.47	0.85	0.98	-0.88	20.13	17.85
					0.05	16.00	17.40	0.77	0.83	-0.63	20.90	21.02
			FF Style	EQ	0.01	10.70	12.11	0.77	1.06	-1.95*	13.83	11.46
					0.05	10.80	10.21	0.78	0.73	0.58	13.91	13.92
				VW	0.01	10.23	11.43	0.76	1.00	-1.66*	13.47	11.47
					0.05	10.46	10.17	0.77	0.75	0.32	13.57	13.62
	180	rolling	Decile	EQ	0.01	18.04	17.00	0.83	0.88	-0.27	21.83	19.24
					0.05	17.00	18.81	0.72	0.87	-0.73	23.58	21.58
				VW	0.01	16.55	16.36	0.76	0.87	-0.51	21.91	18.84
					0.05	17.06	18.36	0.72	0.86	-0.70	23.66	21.23
			FF Style	EQ	0.01	10.06	9.19	0.69	0.75	-0.28	14.50	12.19
					0.05	9.99	10.39	0.67	0.75	-0.38	14.92	13.93
				VW	0.01	9.48	8.93	0.66	0.75	-0.43	14.34	11.87
					0.05	9.54	10.54	0.65	0.77	-0.54	14.62	13.75
	240	rolling	Decile	EQ	0.01	17.38	13.85	0.73	0.62	0.65	23.71	22.17
					0.05	18.01	15.08	0.73	0.62	0.81	24.77	24.49
				VW	0.01	16.47	13.47	0.71	0.63	0.48	23.11	21.37
					0.05	17.43	14.75	0.72	0.62	0.65	24.26	23.74
			FF Style	EQ	0.01	10.26	7.20	0.65	0.52	0.84	15.75	13.90
					0.05	10.84	9.29	0.67	0.60	0.44	16.19	15.45
				VW	0.01	9.68	7.08	0.64	0.53	0.69	15.05	13.36
					0.05	10.45	8.97	0.67	0.60	0.45	15.61	15.01
Yes	120	rolling	Decile	EQ	0.01	9.40	9.75	0.44	0.51	-0.41	21.25	19.29
					0.05	12.59	11.08	0.60	0.56	0.34	20.95	19.87
				VW	0.01	9.59	9.92	0.45	0.51	-0.42	21.41	19.44
					0.05	12.57	11.38	0.60	0.57	0.20	21.05	19.92
			FF Style	EQ	0.01	4.88	4.76	0.33	0.36	-0.17	14.76	13.29
					0.05	6.35	6.67	0.45	0.49	-0.28	14.19	13.75
				VW	0.01	5.14	4.84	0.35	0.36	-0.04	14.57	13.48
					0.05	6.48	6.67	0.46	0.49	-0.22	14.10	13.59
	180	expanding	Decile	EQ	0.01	12.16	14.25	0.60	0.75	-1.38	20.28	18.87
					0.05	14.59	11.33	0.70	0.56	1.45	20.73	20.16
				VW	0.01	11.31	13.15	0.55	0.69	-1.36	20.50	19.05
					0.05	13.67	10.61	0.66	0.52	1.42	20.79	20.21
			FF Style	EQ	0.01	7.21	7.97	0.54	0.66	-0.90	13.30	12.16
					0.05	8.52	6.62	0.61	0.51	1.23	13.88	12.96
				VW	0.01	7.26	7.88	0.55	0.65	-0.86	13.24	12.14
					0.05	8.38	6.84	0.61	0.54	0.99	13.66	12.78
	180	rolling	Decile	EQ	0.01	11.07	11.43	0.52	0.56	-0.34	21.09	20.42
					0.05	13.11	11.78	0.59	0.51	0.60	22.08	23.01
				VW	0.01	10.18	10.17	0.47	0.49	-0.19	21.46	20.55
					0.05	13.22	11.72	0.60	0.51	0.62	22.16	22.89
			FF Style	EQ	0.01	6.77	5.89	0.44	0.41	0.34	15.23	14.51
					0.05	6.75	6.58	0.44	0.40	0.32	15.29	16.47
				VW	0.01	6.56	5.41	0.44	0.37	0.55	15.04	14.52
					0.05	6.38	6.24	0.42	0.38	0.31	15.10	16.40
	240	rolling	Decile	EQ	0.01	13.13	13.79	0.59	0.65	-0.56	22.17	21.33
					0.05	14.48	14.28	0.63	0.62	0.03	23.08	22.94
				VW	0.01	13.20	13.52	0.59	0.64	-0.43	22.23	21.23
					0.05	14.66	14.05	0.64	0.61	0.16	23.08	23.02
			FF Style	EQ	0.01	8.11	7.41	0.53	0.52	0.07	15.24	14.14
					0.05	9.06	8.96	0.57	0.60	-0.18	15.88	15.01
				VW	0.01	8.29	7.70	0.56	0.56	-0.01	14.83	13.77
					0.05	8.98	9.03	0.58	0.62	-0.23	15.40	14.66

Table A.9: Long-short portfolio performance, Mid cap stocks specifications. The table shows selected portfolio characteristics in annualized figures. "FC OOS" indicates whether the FC are only included in the post-publication period. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test. The sample period ranges from Jan 1972 until Dec 2014. Different window lengths are not forced on the same out-of-sample start point; hence, each time window has different return series lengths. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

FC OOS	Window	Type	Buckets	Weights	Win	Mean AdaLasso	Mean OLS	SR AdaLasso	SR OLS	SR Test AdaLasso-OLS	StdErr AdaLasso	StdErr OLS
No	120	rolling	Decile	EQ	0.01	15.70	13.90	0.81	0.88	-0.34	19.34	15.84
					0.05	16.80	12.96	0.81	0.70	0.74	20.75	18.62
				VW	0.01	11.74	11.50	0.63	0.69	-0.26	18.55	16.78
					0.05	13.11	10.98	0.65	0.56	0.54	20.16	19.73
			FF Style	EQ	0.01	9.06	7.03	0.73	0.67	0.33	12.36	10.52
					0.05	9.61	6.26	0.73	0.53	1.19	13.23	11.75
				VW	0.01	6.85	5.62	0.59	0.52	0.37	11.53	10.79
					0.05	7.50	5.25	0.62	0.45	1.01	12.19	11.76
	180	expanding	Decile	EQ	0.01	15.97	15.15	0.84	0.97	-0.75	19.08	15.59
					0.05	15.59	14.18	0.78	0.76	0.28	19.90	18.65
				VW	0.01	14.17	13.15	0.72	0.74	-0.07	19.56	17.79
					0.05	15.58	16.03	0.76	0.83	-0.50	20.57	19.37
			FF Style	EQ	0.01	9.88	10.03	0.82	1.00	-1.04	12.04	10.00
					0.05	9.59	9.25	0.74	0.74	0.01	12.93	12.47
				VW	0.01	8.12	8.21	0.65	0.69	-0.19	12.57	11.86
					0.05	8.32	8.14	0.62	0.63	-0.08	13.50	12.95
	180	rolling	Decile	EQ	0.01	14.93	12.67	0.76	0.66	0.48	19.64	19.17
					0.05	14.46	10.66	0.70	0.53	0.93	20.73	19.95
				VW	0.01	13.08	10.41	0.70	0.53	0.81	18.66	19.70
					0.05	13.74	9.36	0.66	0.46	1.05	20.96	20.14
			FF Style	EQ	0.01	7.78	6.24	0.60	0.48	0.63	13.00	13.05
					0.05	8.30	5.54	0.59	0.39	1.28	13.97	14.14
				VW	0.01	6.02	4.03	0.50	0.31	0.82	12.13	13.19
					0.05	5.80	4.96	0.46	0.36	0.65	12.48	13.95
	240	rolling	Decile	EQ	0.01	15.42	12.35	0.75	0.67	0.50	20.59	18.51
					0.05	15.71	13.61	0.73	0.65	0.53	21.49	20.89
				VW	0.01	15.15	10.96	0.74	0.58	0.73	20.58	18.93
					0.05	15.81	13.55	0.74	0.62	0.66	21.44	21.79
			FF Style	EQ	0.01	8.87	5.62	0.65	0.44	1.21	13.59	12.76
					0.05	9.56	8.28	0.67	0.57	0.75	14.33	14.57
				VW	0.01	8.19	3.07	0.61	0.24	1.53	13.34	12.94
					0.05	8.16	7.08	0.59	0.49	0.59	13.93	14.55
Yes	120	rolling	Decile	EQ	0.01	7.92	8.72	0.41	0.47	-0.43	19.44	18.40
					0.05	10.11	9.26	0.50	0.48	0.11	20.23	19.18
				VW	0.01	9.40	9.64	0.47	0.49	-0.12	19.84	19.53
					0.05	9.90	7.70	0.48	0.39	0.59	20.75	19.84
			FF Style	EQ	0.01	4.68	5.47	0.35	0.42	-0.40	13.44	13.18
					0.05	5.35	4.92	0.39	0.37	0.11	13.86	13.29
				VW	0.01	5.02	5.20	0.40	0.40	-0.05	12.65	12.84
					0.05	5.35	3.97	0.42	0.31	0.78	12.61	12.93
	180	expanding	Decile	EQ	0.01	10.42	11.00	0.56	0.60	-0.32	18.44	18.27
					0.05	11.24	10.47	0.59	0.55	0.43	19.13	19.11
				VW	0.01	10.70	10.15	0.51	0.51	0.00	21.05	20.00
					0.05	11.53	11.82	0.55	0.59	-0.26	20.94	20.18
			FF Style	EQ	0.01	6.46	6.80	0.54	0.57	-0.28	12.04	11.93
					0.05	6.32	6.20	0.49	0.47	0.19	13.00	13.17
				VW	0.01	8.09	6.16	0.61	0.46	1.05	13.16	13.49
					0.05	7.00	6.02	0.52	0.44	0.72	13.42	13.73
	180	rolling	Decile	EQ	0.01	7.86	9.13	0.40	0.44	-0.37	19.60	20.67
					0.05	9.48	8.12	0.46	0.39	0.62	20.51	20.61
				VW	0.01	8.43	8.48	0.43	0.38	0.31	19.77	22.26
					0.05	8.56	5.30	0.41	0.25	1.20	20.84	21.56
			FF Style	EQ	0.01	4.62	4.93	0.35	0.34	0.05	13.30	14.43
					0.05	5.45	5.03	0.38	0.34	0.39	14.27	14.71
				VW	0.01	4.37	3.54	0.35	0.25	0.68	12.33	14.25
					0.05	4.56	3.65	0.34	0.25	0.69	13.32	14.58
	240	rolling	Decile	EQ	0.01	12.55	11.35	0.62	0.55	0.49	20.31	20.53
					0.05	13.14	11.93	0.62	0.59	0.25	21.02	20.35
				VW	0.01	14.36	11.73	0.66	0.52	0.82	21.72	22.57
					0.05	14.42	13.40	0.64	0.61	0.17	22.42	21.78
			FF Style	EQ	0.01	7.40	6.44	0.54	0.46	0.53	13.80	14.05
					0.05	7.63	6.92	0.53	0.51	0.12	14.47	13.54
				VW	0.01	8.45	5.43	0.60	0.37	1.32	14.01	14.67
					0.05	8.23	4.66	0.57	0.32	1.78*	14.31	14.68

Table A.10: Long-short portfolio performance, Large + Mid cap stocks specifications. The table shows selected portfolio characteristics in annualized figures. "FC OOS" indicates whether the FC are only included in the post-publication period. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test. The sample period ranges from Jan 1972 until Dec 2014. Different window lengths are not forced on the same out-of-sample start point; hence, each time window has different return series lengths. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

FC OOS	Window	Type	Buckets	Weights	Win	Mean		SR		SR Test	StdErr	
						AdaLasso	OLS	AdaLasso	OLS	AdaLasso-OLS	AdaLasso	OLS
No	120	rolling	Decile	EQ	0.01	53.54	51.04	2.75	2.55	1.10	19.50	19.99
					0.05	50.19	46.97	2.66	2.43	1.61	18.84	19.35
				VW	0.01	32.27	30.58	1.63	1.55	0.52	19.74	19.69
					0.05	31.16	29.31	1.67	1.52	1.09	18.62	19.27
			FF Style	EQ	0.01	28.22	25.75	2.21	1.95	1.58	12.79	13.21
					0.05	28.44	26.27	2.29	2.08	1.44	12.40	12.64
				VW	0.01	17.05	16.15	1.29	1.23	0.48	13.24	13.17
					0.05	17.35	16.72	1.34	1.26	0.65	12.91	13.31
	180	expanding	Decile	EQ	0.01	55.78	54.86	2.34	2.15	3.14***	23.79	25.47
					0.05	50.99	51.27	2.46	2.32	1.29	20.76	22.14
				VW	0.01	27.66	28.93	1.24	1.18	0.76	22.29	24.52
					0.05	27.83	28.54	1.35	1.28	0.64	20.63	22.35
			FF Style	EQ	0.01	26.88	26.78	1.84	1.74	2.04**	14.58	15.42
					0.05	27.81	25.95	2.06	1.84	2.60***	13.48	14.10
				VW	0.01	14.43	14.73	1.00	0.96	0.60	14.43	15.35
					0.05	16.26	14.61	1.19	1.01	2.25**	13.65	14.46
	180	rolling	Decile	EQ	0.01	55.47	53.92	2.46	2.29	1.62	22.52	23.53
					0.05	50.81	50.06	2.68	2.60	0.68	18.94	19.26
				VW	0.01	29.18	29.80	1.30	1.31	-0.12	22.40	22.66
					0.05	29.77	29.88	1.56	1.53	0.22	19.12	19.52
			FF Style	EQ	0.01	28.19	26.57	1.95	1.74	2.07**	14.49	15.29
					0.05	28.47	26.68	2.35	2.15	1.52	12.14	12.39
				VW	0.01	15.33	14.11	1.05	0.92	1.52	14.61	15.39
					0.05	15.95	15.77	1.26	1.22	0.38	12.62	12.89
	240	rolling	Decile	EQ	0.01	51.64	50.48	2.23	2.06	2.19**	23.18	24.47
					0.05	49.56	47.75	2.36	2.20	1.51	20.96	21.68
				VW	0.01	25.91	24.61	1.16	1.03	1.48	22.39	23.78
					0.05	27.79	26.13	1.33	1.23	0.74	20.89	21.19
			FF Style	EQ	0.01	26.19	26.05	1.84	1.74	1.38	14.21	14.97
					0.05	26.82	24.77	1.98	1.76	2.19**	13.53	14.07
				VW	0.01	13.75	14.40	0.94	0.94	0.07	14.56	15.32
					0.05	14.91	12.82	1.06	0.88	1.92*	14.07	14.62
Yes	120	rolling	Decile	EQ	0.01	43.16	39.48	1.94	1.73		22.21	22.82
					0.05	40.01	35.92	1.99	1.77	1.64	20.13	20.29
				VW	0.01	21.23	18.26	0.98	0.82	1.45	21.66	22.19
					0.05	21.52	18.92	1.10	0.99	0.95	19.58	19.08
			FF Style	EQ	0.01	22.15	20.06	1.57	1.36		14.09	14.71
					0.05	23.01	19.41	1.75	1.49	2.21**	13.16	12.99
				VW	0.01	9.76	9.35	0.70	0.65	0.65	13.96	14.49
					0.05	11.57	9.46	0.88	0.74	1.38	13.11	12.72
	180	expanding	Decile	EQ	0.01	49.22	48.97	2.01	1.95	0.98	24.53	25.16
					0.05	45.87	46.30	1.97	1.96	0.19	23.24	23.61
				VW	0.01	21.83	22.43	0.94	0.95	-0.06	23.17	23.69
					0.05	22.41	22.48	1.00	0.99	0.16	22.46	22.80
			FF Style	EQ	0.01	23.45	23.05	1.57	1.50	1.86*	14.91	15.34
					0.05	23.92	23.46	1.62	1.56	1.48	14.77	15.03
				VW	0.01	10.27	10.01	0.71	0.67	0.91	14.50	14.98
					0.05	11.50	11.07	0.79	0.75	0.92	14.59	14.84
	180	rolling	Decile	EQ	0.01	48.58	47.67	2.06	1.92	1.55	23.54	24.77
					0.05	43.52	42.56	2.14	1.95	2.10**	20.32	21.79
				VW	0.01	24.26	23.19	1.03	0.97	0.67	23.51	24.01
					0.05	22.01	21.86	1.06	1.04	0.33	20.69	21.10
			FF Style	EQ	0.01	24.13	23.18	1.62	1.49	1.53	14.92	15.60
					0.05	23.83	22.77	1.78	1.61	2.34**	13.41	14.15
				VW	0.01	10.85	9.78	0.74	0.66	0.76	14.66	14.71
					0.05	11.17	10.35	0.83	0.74	1.08	13.50	13.91
	240	rolling	Decile	EQ	0.01	50.70	49.98	2.14	2.00	1.81*	23.70	24.95
					0.05	45.95	45.02	1.99	1.89	1.18	23.15	23.81
				VW	0.01	23.35	23.08	1.04	0.97	0.80	22.52	23.84
					0.05	22.57	21.29	1.03	0.92	1.16	21.97	23.15
			FF Style	EQ	0.01	24.14	23.63	1.67	1.55	2.16**	14.42	15.24
					0.05	24.51	23.80	1.66	1.54	1.78*	14.75	15.45
				VW	0.01	11.52	11.26	0.80	0.75	0.65	14.37	14.93
					0.05	11.23	10.97	0.75	0.72	0.44	14.96	15.29

Table A.11: Long-short portfolio performance, Small cap stocks specifications. The table shows selected portfolio characteristics in annualized figures. "FC OOS" indicates whether the FC are only included in the post-publication period. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test. The sample period ranges from Jan 1972 until Dec 2014. Different window lengths are not forced onthe same out-of-sample start point, hence, each time windowhas different return series lengths. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

A.7 Additional figures

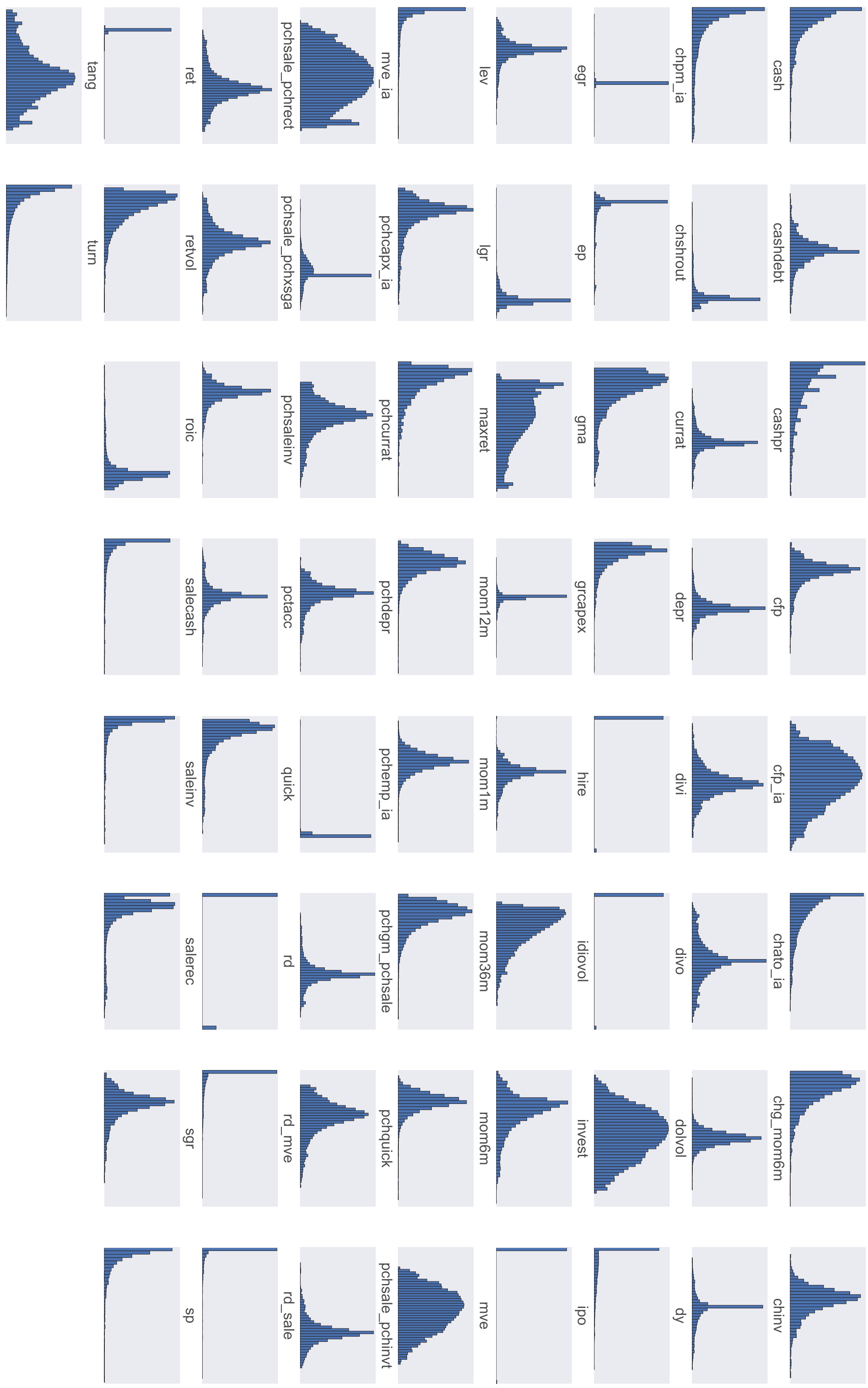

Figure A.2: The figure exhibits the histogram for each FC. The data used are datapoints over all periods; the only adjustment procedure applied is the winsorizing (at each calender point) at the 5% for better visualization, as outliers potentially lead to scaling problems.

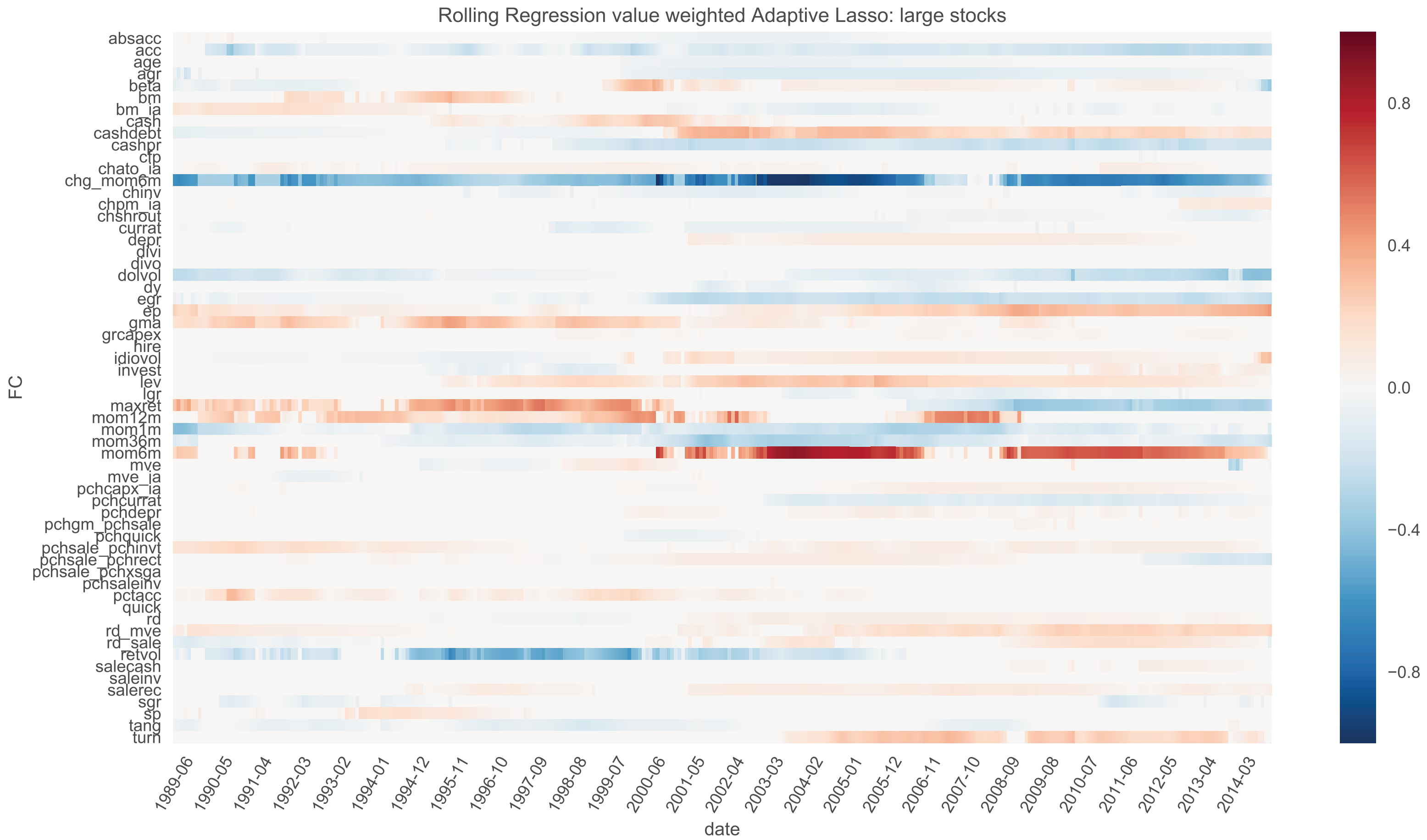

Figure A.3: Rolling adaptive Lasso selection large cap stocks (best seen in color): The figure displays the adaptive Lasso estimates on rolling basis with an estimation window of 15 years for large cap stocks. The top 5 FC measured over the relative frequency of occurrence are: *chg_mom6m*(98.7%), *mom1m*(84.7%), *acc*(82.7%), *maxret*(73.9%) and *egr*(71.7%).

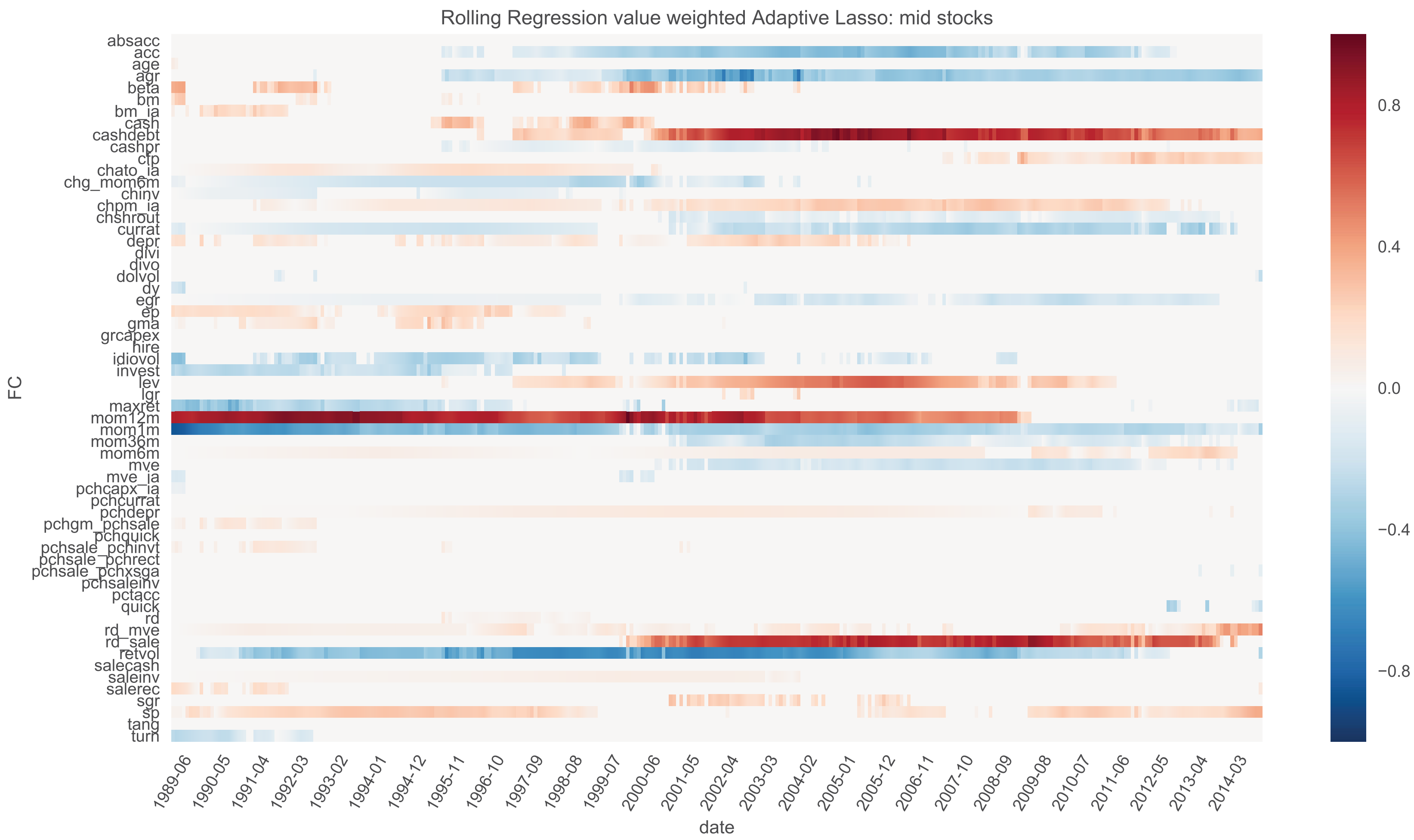

Figure A.4: Rolling adaptive Lasso selection mid cap stocks (best seen in color): The figure displays the adaptive Lasso estimates on rolling basis with an estimation window of 15 years for mid cap stocks. The top 5 FC measured over the relative frequency of occurrence are: *mom1m*(98.7%), *retvol*(87.9%), *mom12m*(78.8%), *agr*(75.6%), and *cashdebt*(67.4%). Overall, we observe consistent results over time compared with the full period selection regression; see Table 1.5 for details.

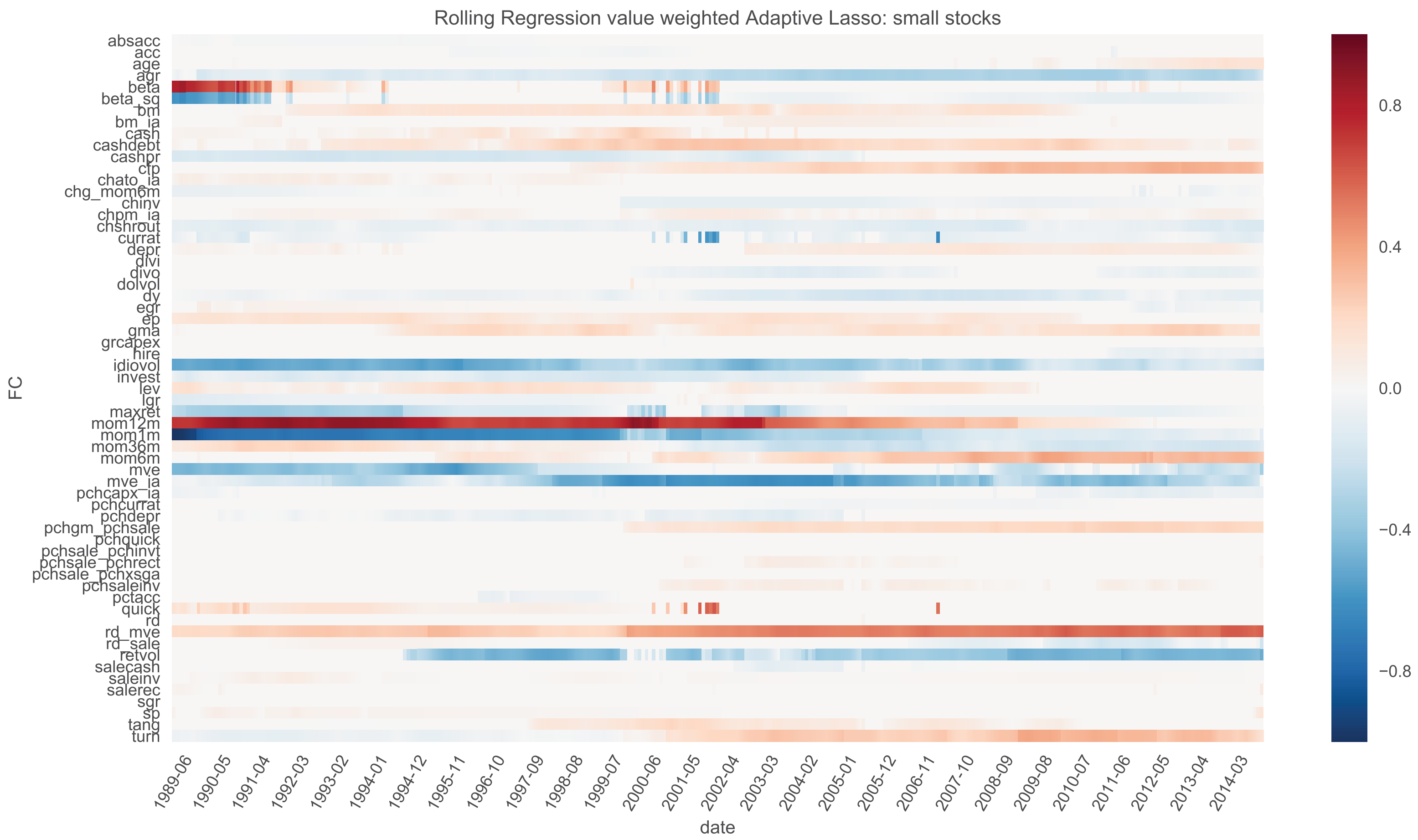

Figure A.5: Rolling adaptive Lasso selection small cap stocks (best seen in color): The figure displays the adaptive Lasso estimates on rolling basis with an estimation window of 15 years for small cap stocks. The top 5 FC measured over the relative frequency of occurrence are: *idiovol*(100.0%), *rd_mve*(100.0%), *chshrout*(99.7%), *agr*(99.0%) and *mve_ia*(98.0%). The rolling selection mirrors the picture measured over all periods in Table 1.6, as the four FC described are most frequently chosen.

Chapter 2

Deep learning and the cross-section of expected returns

Marcial Messmer

2.1 Introduction

The rapid development of software and hardware technology in computer science in combination with the access to a large amount of data enables the training of large and complex models. Particularly, advances in machine learning and artificial intelligence allow researchers to employ self-learning algorithms in many different areas where faced with prediction problems. In particular, deep learning algorithms show promising results in improving the prediction accuracy of regression and classification problems encountered in other areas of science, for example, image processing, speech recognition and medical drug discovery, see Krizhevsky et al. (2012). This work aims to answer the question if and how recent innovations in deep learning (DL) can improve the prediction of stock returns in a cross-section.

In general, DL refers to a rich set of neural network models. However, the deep feedforward network (DFN) specification is the workhorse model of DL applications.[1] The latter marks the focus of this contribution. An appealing, but at the same time intimidating core characteristic stems from the universal approximation theorem, see Hornik et al. (1989) and Cybenko (1989). More explicitly, Hornik et al. (1989) states that a "multilayer feedforward networks with as few as one hidden layer using arbitrary squashing functions are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently many hidden units are available." This powerful statement implies that we have an extremely flexible framework which allows us to gauge the dependence of some input to some output data without any prior assumption regarding the functional form. Ultimately, this means that the model is purely determined by an algorithm. The key challenge for the researcher is then to train the DFN such that it provides a generalization of the data generating process (DGP).

The finance literature has recently addressed the question which firm characteristics (FC) provide independent information in explaining differences in expected cross-sectional returns - see, for example, Harvey et al. (2016), McLean and Pontiff (2016), Green et al. (2017) and Messmer and Audrino (2017). This literature mainly builds on linear or linear-like models, notable exceptions are Freyberger et al. (2017) using non-parametric techniques, and Moritz and Zimmermann (2016) using a CART-like approach to predict cross-sectional returns. Linear and other parsimonious models and methods are appealing for various reasons, for example, transparency and computational efficiency, where the former allows for a clear interpretation of the results and the latter for low implementation costs. However, this simplicity potentially misses important non-linear aspects of the true underlying DGP, which often results in a poor performance when evaluated from a prediction perspective. A classical example in the computer science literature is the XOR problem, which the linear model is incapable to capture despite its low complexity. Applying DL contributes only marginally in

[1] Often DFNs are referred to as feedforward neural networks or multilayer perceptrons.

learning more about the underlying forces of the DGP, however, it can help to understand to which degree we still have to learn from a pure data-driven approach. Besides the loss of a straightforward model intuition, fitting a DL model is typically computationally much more challenging than a linear model.

Training a deep neural network for stock picking is at least partly motivated by a recent important contribution to the open source software community. The so-called "Tensorflow" (Abadi et al. (2015)) library provides a highly scalable and flexible machine learning framework, allowing an efficient usage of DL networks and is the core implementation of Google's artificial intelligence (AI), unit which is under active development. Moreover, the US cross-section of returns is a relative data rich environment. In this application, roughly 2.1 million observations provide a fertile ground for these parameter-rich networks. On the other hand, it is hard to assess ex-ante if these methods are suitable for predicting stock returns, due to the inherently different statistical character they possess — for example, the signal-to-noise ratio of a stock return process is a tiny fraction compared to the processes typically encountered in computer science.

The main contribution of this work lies in the investigation if recent developments in artificial intelligence are of any use predicting cross-sectional stock returns. Applying artificial neural networks (ANN) in finance is not new. Hence, past attempts have to be distinguished. This study can be seen as an extension to earlier attempts of applying ANN to predict stock returns, with the difference of having access to additional regularization techniques, better computational resources, and more data. Additionally, it is, to best of my knowledge, the first study which investigates the cross-section and its relation to a rich set of published FC by exploiting a purely data-driven algorithm without any prior assumption on the functional relation between FC and expected return spreads.

Precisely, this paper aims at answering the following main three research questions: First, how can one efficiently employ a DFN framework for the purpose of return predictions? Second, do DFN based predictions add additional economic value compared to a parsimonious linear approach utilizing the same information set? Third, which set of FC drive the prediction results and how far do they differ from recent findings in the literature of FC selection.

The first question is related to the selection problem of the optimal DFN design for this exercise. I address this question by stating the problem as an outer optimization problem. This computational intensive task is tackled by utilizing a random search algorithm as proposed in Bergstra and Bengio (2012) in combination with a one-dimensional grid search for learning rate tuning. The procedure reveals that many network designs fail to deliver reasonable numerical behavior. Despite a relatively high failure rate, I identify architectures which show promising improvements compared to the linear benchmarks based on a validation data set.

The short answer whether economically measurable improvements can be achieved, is yes. I find significant and robust factor α's, which are consistently higher compared to the parsimonious linear benchmark. In many (but not in all) cases I document significant higher Sharpe Ratios (SR). No specification favors the linear model, irrespectively of which performance measure is considered. However, a naive strategy is sensitive to trading cost adjustments for both approaches. Nonetheless, I show that a simple rebalancing frequency adjustment leads to stark improvements. An explicit rebalancing optimization is not carried out and can be seen as a limitation of this work. Over the sample period, I document that DFN based portfolios perform much weaker during high volatility periods compared to times of calmer markets, a phenomenon which is characteristical for momentum-based strategies. Controlling for momentum exposure during these times levels the α's significantly into the positive domain.

The answer to the question which FC drive the predictions points unambiguously at price based information, predominantly short-term reversal (providing an explanation for the turnover intensity) and the twelve-months momentum. However, I study the impact purely by looking at prediction changes arising from variation in the input data. As a result, it can not be seen as a perfect measure, but a computational trivial way in gaining model insights at this stage.

This study includes cross-sectional stock data from the CRSP/Compustat database from 1970-2014. In total, I use 68 published FC, constructed based on accounting and market data. The focus of the analysis lies solely on large and mid-cap stocks, to prevent a potential contamination arising from economically unimportant small and micro-cap stocks, as recently documented in Hou et al. (2017). Notably, the latter study includes more than 400 FC, I instead rely on a much smaller subset, however, I include many prominent FC, such as the market beta, twelve-months momentum, book-to-market ratio, idiosyncratic volatility, and profitability.

The paper is structured as follows. Section 2.2 reviews the relevant literature on finance applications of machine learning and briefly presents important contributions to the field of deep learning. Section 2.3 gives a detailed description of the deep-learning methodology utilized throughout the study. This part is followed by a brief description of the data. Section 2.5 includes the network design optimization. The penultimate section covers the portfolio mapping and the analysis corresponding to the long-short returns of these portfolios. The final part concludes.

2.2 Related literature

In two seminal contributions, Fama and French (1992, 1993) present the three-factor model, relating expected cross-sectional returns to a market, size and value factor. The model would mark the benchmark for the years to come. Irrespective of the presence of these three factors, a huge amount of new FC have been identified explaining additional variation in cross-sectional expected returns. Prominent examples are, momentum (Jegadeesh and Titman (1993)), quality (Novy-Marx (2013)), low-beta (Frazzini and Pedersen (2014)) and low-volatility (Ang et al. (2006)). Harvey et al. (2016) identify 300+ and, more recently, Hou et al. (2017) document and replicate a total of 447 such anomaly FC. This large number of FC is likely according to these studies a consequence of a publication bias. Consequently, a number of studies address the question of digesting this "factor zoo" (Cochrane (2011)). Notable contributions are Harvey et al. (2016), McLean and Pontiff (2016) and Green et al. (2017). From a methodical point of view, Harvey et al. (2016) introduce the concept of family-wise error rates, McLean and Pontiff (2016) follow a true out-of-sample strategy, and Green et al. (2017) focus on the multivariate linear digestion. Common to all three papers are the findings that many of the published factors are indeed not robust to those alternative testing procedures.

Central to all of these papers is the understanding of the following relationship:

$$R^e_{t+1,n} = f(x_{t,n}, \theta) + \epsilon_{t+1,n}$$

where $f(\cdot)$ defines a generic function with parameters θ, $x_{t,n} = [size_{t,n}, bm_{t,n}, mom_{t,n}, ...]$ the vector of FC and $R^e_{t+1,n}$ the corresponding excess return and $\epsilon_{t+1,n}$ the error at a given point in time $t = 1, ..., T$ for stock $n = 1, ..., N$.

The following studies can be grouped into a rather young strand of the literature on machine learning and asset pricing — mostly shrinkage based approaches in the cross-section of stock returns. Moritz and Zimmermann (2016) use a CART (Breiman et al. (1984)) inspired methodology and find significant performance improvements. DeMiguel et al. (2017) take a portfolio perspective and optimize the problem from an investor's view, with an explicit trading cost term in the objective function. The findings show that only a limited number of FC are important, but trading diversification is an important aspect and leads eventually to a larger set of optimal FC. Kozak et al. (2017) use L1 and L2 norms and estimate the cross-section based on a stochastic discount factor specification, tying FC more to a risk-return relation. Their findings suggest the importance of interactions and characterize a non-sparse set of FC, explaining the cross-section of expected returns. Messmer and Audrino (2017) use the adaptive Lasso (Zou (2006)) to identify a sparser set of FC. The application is motivated based on an extensive cross-sectional simulation study, which indicates that the adaptive Lasso enjoys advantages over Lasso-based shrinkage due to less stringent conditions imposed connected to the covariance structure of FC. Freyberger et al. (2017) tackle the problem by using a non-parametric model in combination with the grouped adaptive Lasso. Their approach is motivated by the classical portfolio sorting methodology and robust to extreme values.

The early work of artificial neural networks dates back to McCulloch and Pitts (1943), Hebb (1949) and Widrow et al. (1960). Moreover, two seminal contributions to the field are the studies by Parker (1985) and Williams and Hinton (1986), which facilitate the usage of DFN by advances of the back-propagation algorithm. Recently, the work of Bengio et al. (2007), Hinton (2007), Poultney et al. (2007) mark important achievements of artificial intelligence. For more details on the history of the deep learning and AI literature I refer the reader to Goodfellow et al. (2016).

Not surprisingly, the idea of applying DFNs to financial prediction problems is not new, see, for example, Trippi and DeSieno (1992) or Leung et al. (2000). However, the early work has not been too well received because of various reasons. The main problem connected to DFN is that they are prone to over-fitting. Consequently, the attention which is paid to DFNs is only marginal in the finance literature. Most closely related to this work is a paper by Takeuchi and Lee (2013), who analyze cross-sectional momentum strategies. The authors use a restricted Boltzmann machine approach and show performance improvements compared to the classical counterparts. In contrast to their work, this paper relies on a much wider set of FC and follows a different methodology by applying DFN. Recently, Dixon et al. (2015) show an application of deep learning to predict the returns of currency and commodities futures contracts based on high-frequency data. The work is mainly focused on technical aspects, but it reports improvements in predicting market directions.

2.3 Methods

This section provides a brief introduction into deep learning, however, I focus on concepts mainly relevant for this work, for a detailed overview I refer the reader to Goodfellow et al. (2016), which I use as a main reference as well as a guidance for the notation throughout the paper. Deep learning describes a field of machine learning, which comprises a variety of different neural network model architectures. It aims to provide a generic, trainable framework capable of capturing almost any arbitrary functional dependence the DGP may be subject to. The ultimate goal for the researcher is to find a specific approximation of this specific function f, which expresses some output y as a function of an input vector $\mathbf{x}$. In my case the goal is to capture the functional relationship of expected cross-sectional returns based on a set of FC (which I indicate by k of K total FC). Let $y_{n,t+1}$ denote a prediction target, which is here, an excess return of firm n (of a total of N) at point in time t (of a total of T periods). The expression $f(x_{n,t};\theta)$ defines this function, where $x_{n,t}$ defines the input vector of FC, and θ the set of model parameters. The shape of θ, as well as the model capacity and hence, the richness of the set of functions is a result of the set of hyper-parameters, which has to be determined by the researcher and is denoted by λ. An example of θ is a vector of model coefficients, whereas the number of layers in the DFN is part of the set λ. Hyper-parameter selection is an optimization problem of its own, with the ultimate objective to minimize generalization error, subject to the computational constraints. This important problem is addressed in detail in section 2.5. First, I present an overview of the neural network anatomy and corresponding optimization problem.

2.3.1 Deep learning anatomy

The first step when specifying a neural network learning task is to select an appropriate network architecture. Possible families and combinations of thereof are DFN, convolutional neural networks (CNN) or recurrent neural networks (RNN). I skip a detailed discussion on the technical details and how and which data environment they can be used in. Instead, I focus on the fully connected deep-feedforward neural network design, because its application is straightforward and does not require to manipulate, extend or order the input data. The other model families might potentially work as well, but would require non-trivial data extensions and transformations to accommodate the specific structural requirements.[2]

The basic mechanism of the DFN can be described as follows. The network follows a hierarchical structure and consists of three different layers: the **input layer**, which represents the input data, **the hidden layers**, connecting various inputs in a nested and hidden structure, and an **output layer**, reflecting the prediction target — which are ordered in the sequence just described. Figure 2.1 illustrates the basic structure and flow of a very small network. However, I use this simple example to provide more meaning to the generic function $f(x_{n,t};\theta)$ mentioned before. The Figure illustrates the dependence of x to h, the first hidden layer, which is in this case simply Wx. Typically, it reads as $h = g(b + W'x)$ with bias term b, a vector of length 2, weight matrix W, of dimension 2×2, and activation function $g(\cdot)$. Accordingly, this simple network function reads,

$$\hat{y} = b + w'g(\mathbf{b} + W'x).$$

Hence, the set of θ includes $W, \mathbf{b}, w, b$ and requires 9 model parameters to be estimated. Applying the same logic, wider and deeper models can be expressed in similar fashion, which then only results in a much more nested function $f(x_{n,t};\theta)$ with many more parameters to be estimated.

The activation function $g(\cdot)$ is applied element-wise and can take various forms. Common choices are the more traditional sigmoid activation ($\sigma(x) = \frac{1}{1+\exp(-x)}$) and the popular rectified linear unit. The rectified linear unit, $g(z) = max(0,z)$, has computational advantages over the other candidate functions and is, therefore, often the preferred choice for DFN designs, see Jarrett et al. (2009), Nair and Hinton (2010) and Glorot et al. (2011) for more

[2]For example, CNN would require a reasonable topology. However, it is not obvious how to structure the FC input data in comparison to image processing, where the ordering of pixels in 2D or 3D come naturally.

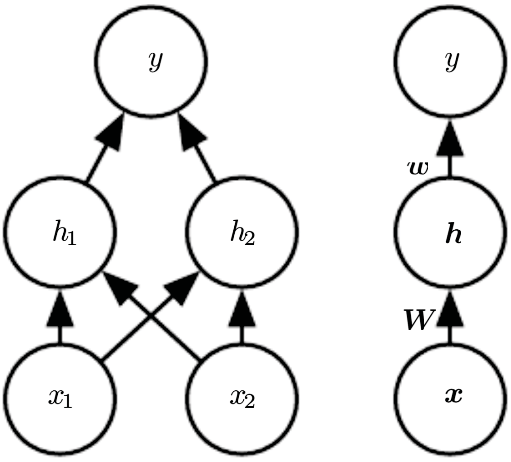

Figure 2.1: Computational Graph DFN: Base anatomy of a fully connected deep neural network with one hidden layer and two units. The input variables x_1 and x_2 mark the input layer, h_1 and h_2 and two hidden units in the only hidden layer, y marks the output. W and w defines the weight matrix. Illustration from Goodfellow et al. (2016), page 169.

details.

2.3.2 Optimizing and training the network

Objective Function

A key task of deep learning is model training, which requires an objective function as a starting point. The ultimate objective of the researcher is to minimize a loss function evaluated with respect to the true DGP. Algebraically, it is expressed as follows,

$$J(\theta) = \mathbb{E}_{\mathbf{x},y\sim p_{\mathrm{DGP}}} L(f(\mathbf{x},\theta);y) \tag{2.1}$$

However, practically the true DGP is mostly unknown, like in the application of this work. Consequently, the empirical counterpart becomes the actual problem at hand. According to equation 2.1 the empirical objective function can be written as,

$$\hat{J}(\theta) = \mathbb{E}_{\mathbf{x},y\sim p_{\mathbf{x},y}} L(f(\mathbf{x},\theta);y) = \frac{1}{TN}\sum_{t}^{T}\sum_{n}^{N} L(f(\mathbf{x_{n,t}},\theta);y_{n,t+1}), \tag{2.2}$$

where I simply replace the underlying DGP by the empirical distribution. Furthermore, I solely deploy a standard mean squared error loss ($L(a,b) = (a-b)^2$) throughout this paper. This is in line with the literature, see for example Green et al. (2017) and Freyberger et al. (2017).

Stochastic-Gradient Descent

In contrast to standard convex optimization problems, DFN optimization is almost exclusively non-convex and, hence, builds heavily on numerical optimization routines. Stochastic-gradient descent (SGD) is the standard optimization principle underlying almost all neural network optimization routines. The idea behind gradient-based approaches is to solely iterate towards a low value of the cost function, often the algorithms terminate prior to reaching the (local) minimum of the training cost function. The SGD algorithm requires mainly some initialization strategy for the set of parameters to be estimated. Additionally, exact gradient computations are available for wide and deep DFN choices based on back-propagation, which itself relies heavily on the application of the chain rule, which is required because of the nested functional structure introduced above. The basic algorithm is described in 1. However, many variants of this basic SGD exist, all addressing different problems arising of misbehaved gradient functions. Examples are Momentum based algorithms (Polyak (1964)), which consider not only the most recent gradient for updating the parameters but past gradient values, with a decaying weight function. Other strategies vary the learning rate adaptively with respect to the parameters, examples are, AdaGrad, Duchi et al. (2011), or Adam, Kinga and Adam (2015). Schaul et al. (2013) provide a framework of testing these optimization routines, however, no general statement can be made on which SGD variant is the most robust. More details on SGD and its variants can be found in Saad (1998) and Ruder (2016).

Algorithm 1 Stochastic Gradient Descent

1: $\hat{\theta} \leftarrow$ random draw ▷ Initialize parameters
2: $\{(y_{\text{valid}}, X_{\text{valid}}), (y_{\text{train}}, X_{\text{train}})\} \leftarrow \text{split}(X,\, y)$ ▷ Split data into training and validation set
3: $\kappa \leftarrow$ assign learning rate
4: **for** $i = 1$ to K **do** ▷ K max number of iterations
5: $\quad \hat{g} \leftarrow \frac{1}{J}\nabla_\theta \sum_j^J L(f(x_j;\theta), y_j)$ with $x_j \in X_{\text{train}}$ and $y_j \in y_{\text{train}}$ ▷ Gradient estimate
6: $\quad \hat{\theta} \leftarrow \hat{\theta} - \kappa\hat{g}$
7: **return** $\hat{\theta}$ ▷ The parameter set estimate

Mini-batch

Mini-batching is a data splitting strategy for gradient computations. Instead of calculating gradients for the entire training set, gradients are computed and parameters updated based on only a small subsample. This serves two purposes. First, computations are eased due to the much smaller data size and, second, according to Wilson and Martinez (2003), mini-batch training enjoys a regularization benefit. The study by Goyal et al. (2017) investigates the tradeoff between validation error and mini-batch size/computational efficiency, where they document a positive relationship between mini-batch size and validation error. Consequently, I carry out the optimization including mini-batching.[3]

2.3.3 Regularization

It is easy to imagine that a wide and deep enough DFN design can fit any finite training sample almost always perfectly. Consequently, the machine learning community has developed a range of tools, which aim to find an optimal balance between over- and under-fitting. Goodfellow et al. (2016) present several techniques, of which I utilize early stopping, dropout and norm penalties (L1 and L2), which are described in detail in the appendix. However, not all available tools are applicable in this context. For example, commonly used data-augmentation strategies cannot simply be adapted to financial data sets. For more details on regularization, I refer the reader to Goodfellow et al. (2016).

2.4 Data

2.4.1 Returns and FC construction

The cross sectional FC and returns used in this study are based on the CRSP/Computstat database. The FC construction and notation is identical as in Messmer and Audrino (2017) and, therefore, mainly consistent with the study of Green et al. (2017). Details on the FC construction can be found in the appendix. Like the aforementioned contributions, this work considers only CRSP stocks with share code 10 and 11 which are traded either on the NYSE, AMEX or NASDAQ. Additionally, stocks with missing market capitalization data and/or where book values are unavailable are dropped from the analysis. The alignment of accounting data is performed with a standard lag of six months of the fiscal year end date.[4] Return based FC, such as one-month momentum, beta, maximum return, or six-months momentum are used as of the end of the most recent month.[5] Handling extreme values is a sensitive procedure. Here the study deviates from Messmer and Audrino (2017). Extreme FC observations are controlled for by using relative cross-sectional ranks at each point in time instead of relying on winsorizing the extremes. In the next step, missing data are replaced by the median value of 0.5 at each point in time. Note that the return observations are not adjusted for any extreme values. However, returns are de-meaned for each period

[3] Mini-batch size becomes more critical from a computational perspective once GPUs are in use and should be configured according to the exact hardware specification.

[4] For instance, the Compustat data of a firm with fiscal year end date 12/31 are aligned with data 06/30, predicting monthly returns from 6/30 to 7/31.

[5] For example, for the return prediction from from 6/30 to 7/31, the max daily return from the period 5/31-6/30 is used.

to preserve the cross-sectional information. The last step requires the pooling of FC and returns over time, such that the networks can be estimated with the stacked matrix $\mathbf{X}$ and the output return vector $\mathbf{y}$. As in Fama and French (1996), stocks qualify as large if their market capitalization ranks among the top 1000, accordingly mid-cap stocks have to fall into the range of rank 1001-2000, and small-caps comprise all stocks with rank > 2000.

2.4.2 Training, validation and test data

The trade-off between over-fitting vs. under-fitting is the major concern of training complex machine learning algorithms. The financial econometrics literature typically splits the data into an in-sample and out-of-sample data set to evaluate model performance. However, model complexity and capacity are not a key concern in this field, as the researcher typically has a prior belief of the intrinsic nature of the DGP, which is often derived from economic theory. Hence, model capacity restrictions are implicitly imposed, therefore, parameter tuning is redundant. Pure data learning approaches lack these prior model restrictions and hence, require a more careful approach to tackle model complexity explicitly. This is typically achieved by splitting the data-set into a training, validation and test set. The training set is used to train the model, the validation set is used to evaluate the estimated model on an independent data-set. Cross-validation (CV) is a commonly used specification of validation error evaluation, it provides an estimate of the expected prediction error. Moreover, CV is unbiased if the validation set is independent of the training set, see Bishop (1995a). Commonly used are k-fold CV, which splits the training data into k-equal sized parts. However, cross-validation is computationally costly, as it requires k-optimization routines on $\frac{k-1}{k}$ of the training data. Typically, stock returns are highly cross-sectionally correlated and if at all, only weakly correlated across time. Consequently, cross-validation is performed along the time axis, such that each training and validation data set does not share any stocks from the same period. The training and validation data set spans the period from 1970-01 until 1981-12. It is exclusively used for model selection. As a result, the test data set begins with the first month of 1982 and ends in December of 2014. Note that, I still use a training and validation strategy for estimation purposes.

2.5 Hyper-parameter optimization

Section 2.3 introduces important concepts and tools to minimize naive training. Most of these require explicit hyper-parameters which have to be supplied. Hyper-parameters are, for example, the number of hidden layers or the learning rate of the optimization routine. In particular, the hyper-parameter space is typically of high dimension and optimal hyper-parameter tuning an area of research of its own, see, for example, Bergstra and Bengio (2012) or Larochelle et al. (2007).

2.5.1 Hyper-parameter set

I can define the set of hyper-parameters, λ, which the researcher has to provide to obtain a particular instance of the aforementioned model design. Specifically, the complete set of parameters underlying this work looks as follows: *number of hidden layers (HL), number of units per layer (UN), dropout probability (Drop), L1 regularization strength (L1), L2 regularization strength (L2), the optimization routine (Opt), learning rate (LR), epoch bound (EB), patience (PA), mini-batch size (Mini), the k in K-Fold-CV* and finally, *the type of activation function.*

2.5.2 Outer optimization problem

A priori it is not known what the optimal combination of these hyper-parameters looks like. Hence, hyper-parameter selection can be seen as an outer optimization problem. Some variables can be fixed due to computational considerations, for others, I need to provide bounds to keep the problem tractable. Specifically, choosing between five or ten-fold CV, I use 5-Fold CV to estimate the validation error, mainly because it more than halves the computational costs. Furthermore, I set the activation function to the popular rectified linear unit (ReLu), motivated by computational advantages over alternative activation functions — due to its trivial gradient calculations. The mini-batch size is set to 128 — a value, which typically provides a good balance of robustness and efficiency. Finally, I set an epoche bound of 200, to limit the training time for each specification.[6] Hence, I am left to determine optimal values for the following set of hyper-parameters: $\lambda := \{\mathrm{HL}, \mathrm{UN}, \mathrm{Drop}, \mathrm{L1}, \mathrm{L2}, \mathrm{Opt}, \mathrm{LR}, \mathrm{PA}\}$. CV provides training ($y_{\text{train}}$, X_{train}) and validation (y_{valid}, X_{valid}) data to compute loss function values. Consequently, the optimization problem can be formulated as follows:

$$\lambda_{opt} = \underset{\lambda \in \Lambda}{\arg\min} \sum_{i=1}^{k} \mathcal{L}(y_{\text{valid},i}, X_{\text{valid},i}, \hat{\theta}_{X_{\text{train},i},\lambda}, \lambda). \tag{2.3}$$

where $\hat{\theta}_{X_{\text{train},\lambda,i}}$ are estimated model parameters from the i-th training data sample given the hyper-parameter λ, k refers to the number of CV folds. Hence, we evaluate a finite number of trials, t, to determine $\hat{\lambda}_{opt}$:

$$\hat{\lambda}_{opt} \equiv \underset{\lambda \in \{\lambda_1, \ldots, \lambda_t\}}{\arg\min} \sum_{i=1}^{k} \mathcal{L}(y_{\text{valid},i}, X_{\text{valid},i}, \hat{\theta}_{X_{\text{train},\lambda,i}}, \lambda).$$

[6]One epoche defines one complete optimization run through the entire training data set.

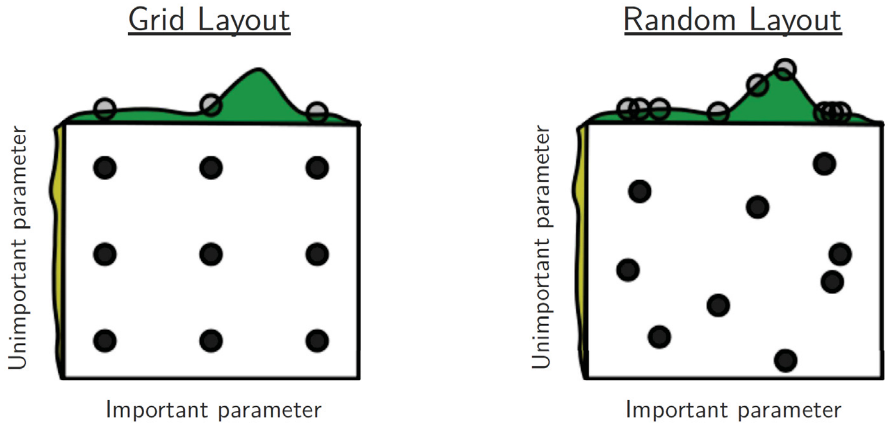

Figure 2.2: The figure is taken from Bergstra and Bengio (2012). It illustrates the difference between a grid and random search algorithm. The example assumes that the objective value is almost insensitive to changes along one parameter dimension (the y-axis).

2.5.3 Curse of dimensionality

Optimizing equation (2.3) is a non-convex high-dimensional problem. Grid-search or brute-force is one possible option to solve the problem. However, it suffers from the curse of dimensionality, since the optimization problem of this application is carried out in eight dimensions. Assuming a lower bound of grid-points in each dimension of three (a relatively sparse grid), the routine would need to train $3^8 = 6561$ neural networks. Instead of performing an expensive grid-search, I utilize random search in the fashion of Bergstra and Bengio (2012) to determine an optimal set of hyper-parameters. However, Goodfellow et al. (2016) note that the learning rate is the decisive factor in many applications. Hence, I further tune the best network found in the random search along the learning rate dimension.

2.5.4 Random search

Bergstra and Bengio (2012) show that random-search has advantages over grid-search in situations, where the effective-dimensionality of the problem is lower than implied by the length of the set of hyper-parameters. This implies that only a smaller subset of hyper-parameters is crucial of minimizing validation error. The intuition behind this is that grid search wastes resources moving along insensitive dimensions too often instead of varying all dimensions in each trial. Figure 2.2 illustrates this graphically.

Specification

For each trial, I draw a random set of hyper-parameters. I mainly follow Bergstra and Bengio (2012) and Goodfellow et al. (2016) by drawing the respective hyper-parameters independently from the following distribution — however, the following specification is slightly adjusted to reflect computational limitations at the application at hand, the largest model can, hence, be 14 layers deep and approximately 100 units wide:

- HL: random draw from integers ranging from 2 to 14.
- UN: $round(\exp(\eta))$, where $\eta \sim U(log(A), log(B))$ and A and B is set to 0.5 times and 1.5 times, respectively, the number of FC.
- Drop: $\sim U(0, 0.75)$.
- L1: 10^a where $a \sim U(-5, 0)$
- L2: 10^a where $a \sim U(-5, 0)$
- Opt: random draw form the following list of optimizers available in Abadi et al. (2015): 'RMSprop', 'Nadam', 'Adagrad', 'Adadelta', 'Adam', 'Adamax' and 'Nesterov'.
- LR: 10^a where $a \sim U(-5, -1)$
- PA: random draw from integers ranging from 20 to 40.

Empirical results

The random search conducted in this study includes three different data specifications: large and mid cap separately and the joint set of large and mid caps. For each data set, I run 200 trials for a period of 12 years, using data from 1/1/1970 until 12/31/1981. Applying five-fold cross-validation leads to 9.6 years of monthly training data and 2.4 years of validation data. This specification requires the training of 3000 networks. For each trial, I calculate the loss function as described in Equation 2.3. In order to ensure comparability across the trials and networks, the same training and validation data splitting is used in each trial. The motivation behind using five fold cross-validation is the intrinsic random character of the validation score. Five-fold cross-validation provides five (cross-sectionally) independent estimates, and hence, helps to reduce the variance of the estimate. Consequently, time dependencies are neglected.

Table 2.1 shows the random search results for the joint set of large and mid cap stocks. The search identifies 8 networks, which achieve higher R^2's evaluated with the validation sample. The best network design reaches a value of 7.31% vs. 4.5% with the linear specification. The architecture of the best performing network for large cap stocks has 62 hidden units, 3 layers, and a learning rate of 0.02×10^{-3}; for more details, I refer to Table 2.2. Hence, this network is not deep. However, the annualized R^2 of 5.06% shows an improvement compared to the linear benchmark of 4.34%. However, only the two best performing designs achieve lower objective function values in comparison to the linear model. The optimal network discovered for mid cap stocks comprises a deeper architecture, with 53 hidden units and 8 layers as Table

Rank	R^2 DFN	R^2 Lin	LR	L2	L1	dropout	UN	HL	Opt	PA
1	7.31	4.50	0.01	3.14	0.05	0.02	78	7	Nadam	35
2	6.80	4.50	0.01	0.07	0.04	0.07	51	14	Nadam	33
3	6.30	4.50	0.02	158.76	0.01	0.06	67	2	RMSprop	20
4	6.00	4.50	0.28	44.44	0.04	0.15	46	3	Adam	25
5	5.51	4.50	1.22	0.51	5.22	0.09	91	12	Adamax	30
6	5.32	4.50	1.02	37.81	0.05	0.07	61	5	Adamax	38
7	4.78	4.50	0.02	61.17	1.15	0.01	45	8	Adam	34
8	4.76	4.50	0.24	0.33	0.18	0.29	47	2	Adam	33
9	4.33	4.50	0.01	0.02	0.21	0.26	76	6	Nadam	36
10	4.24	4.50	0.13	3.31	0.88	0.27	47	5	Adam	29

Table 2.1: The Table displays the random search results for **Large + Mid Cap Stocks:**. It shows the top ten results of the 200 trials performed, sorted by R^2s. The R^2s shown are annualized based on monthly observations. R^2 are given %-points. The columns "LR", "L2", and "L1" are scaled by a factor of 1000. In total 89 out of 200 models show an R^2 below zero, moreover, 61 out of 200 models fail to provide any numerical results.

Rank	R^2 DNN	R^2 Lin	LR	L2	L1	dropout	UN	HL	Opt	PA
1	5.06	4.34	0.02	0.06	0.26	0.18	62	3	Nadam	37
2	4.39	4.34	4.21	0.05	0.04	0.26	78	2	Adagrad	32
3	3.67	4.34	0.02	41.65	0.18	0.06	33	10	RMSprop	22
4	3.38	4.34	0.03	0.24	2.97	0.07	82	13	Adamax	34
5	3.20	4.34	2.45	138.46	0.02	0.24	57	4	Adamax	38
6	3.19	4.34	0.04	0.57	7.28	0.01	38	5	Nadam	24
7	3.05	4.34	0.67	0.01	0.17	0.37	32	2	Adamax	27
8	3.01	4.34	0.09	258.04	22.41	0.08	37	4	Adam	36
9	2.88	4.34	0.73	0.02	0.01	0.27	57	7	Adamax	28
10	2.57	4.34	40.27	94.13	0.83	0.11	48	9	Adadelta	40

Table 2.2: The Table displays the random search results for **Large Cap Stocks:**. It shows the top ten results of the 200 trials performed, sorted by R^2s. The R^2s shown are annualized based on monthly observations. R^2 are given %-points. The columns "LR", "L2", and "L1" are scaled by a factor of 1000. In total 75 out of 200 models show an R^2 below zero, moreover, 83 out of 200 models fail to provide any numerical results.

2.3 reveals. Furthermore, and not surprisingly, the overall R^2's are slightly more elevated compared to the case of large cap stocks. The best DNN model outperforms with an R^2 of 6.18% vs. the linear hurdle of 4.71%. Furthermore, a total of 9 out of 200 models indicate an edge over the parsimonious alternative.

Evidently, training DNNs is numerically a tedious challenge as in both cases a majority of the models produce either below zero R^2's or no numerical results at all. It is technically relatively difficult to understand the source of these errors, potential sources are poor starting values or inappropriate learning-rates. A more detailed analysis of this behavior is beyond the scope of this work.

Rank	R^2 DNN	R^2 Lin	LR	L2	L1	dropout	UN	HL	Opt	PA
1	6.18	4.71	0.08	0.22	1.91	0.11	53	8	Adam	22
2	5.42	4.71	1.55	0.06	0.04	0.18	46	5	Adamax	36
3	5.36	4.71	0.03	0.09	0.05	0.18	86	10	Nadam	33
4	5.36	4.71	3.51	1.38	0.02	0.17	70	2	Adam	33
5	5.25	4.71	3.50	0.38	18.29	0.17	66	2	Adamax	25
6	5.01	4.71	3.50	0.38	18.29	0.17	66	2	Adamax	25
7	4.96	4.71	0.06	0.11	0.05	0.19	77	3	RMSprop	30
8	4.92	4.71	0.09	0.24	0.82	0.30	55	2	Adamax	34
9	4.76	4.71	9.21	0.45	3.06	0.15	67	6	Adamax	30
10	4.65	4.71	1.68	0.74	0.43	0.25	80	4	Adagrad	28

Table 2.3: The Table displays the random search results for **Mid Cap Stocks:**. It shows the top ten results of the 200 trials performed, sorted by R^2s. The R^2s shown are annualized based on monthly observations. R^2 are given %-points. The columns "LR", "L2", and "L1" are scaled by a factor of 1000. In total 92 out of 200 models show an R^2 below zero, moreover, 75 out of 200 models fail to provide any numerical results.

2.5.5 Learning rate tuning

Goodfellow et al. (2016) underscore the sensitivity of DNN optimization with respect to the learning rate. I use the optimal hyper-parameters from the random search conducted in subsection 2.5.4 and define a grid along the learning rate dimension to further optimize the given network architectures. Precisely, I use 20 equally spaced points (log-based) and optimize the model accordingly. Figure 2.3 shows the results, which confirm a strong sensitivity of learning rate variation for the problem at hand. Despite different network structures the optimal learning rate is in a similar region for both large and mid cap stocks. The annualized R^2 improves for the large cap sample around 17% (5.06% vs. 5.94%), compared to the optimization result without learning-rate tuning. The gain in case for mid cap stocks is only moderate with an increase of roughly 1% (6.18% vs. 6.25%).

2.5.6 Network details

Random search has shown that many architectures suffer from numerical failure, hence, I investigate briefly the behavior of the optimal network selected. In this case, a numerical failure is defined when the optimization algorithm returns errors or extreme and error-like numerical predictions. First, Figure 2.4a presents the development of the loss function evaluated at each iteration step for both the validation and the training data sets. It behaves as expected, with an initial decline in both function values, followed by a slow increase of the validation score, which causes the optimization to terminate after 97 iterations. Moreover, figure 2.4b shows a smooth distribution of the predicted cross-sectional returns after the final step of the optimization, which looks from a numerical point of view unsuspicious. Initially, the output concentrates a large mass around zero. Over time the output distribution becomes

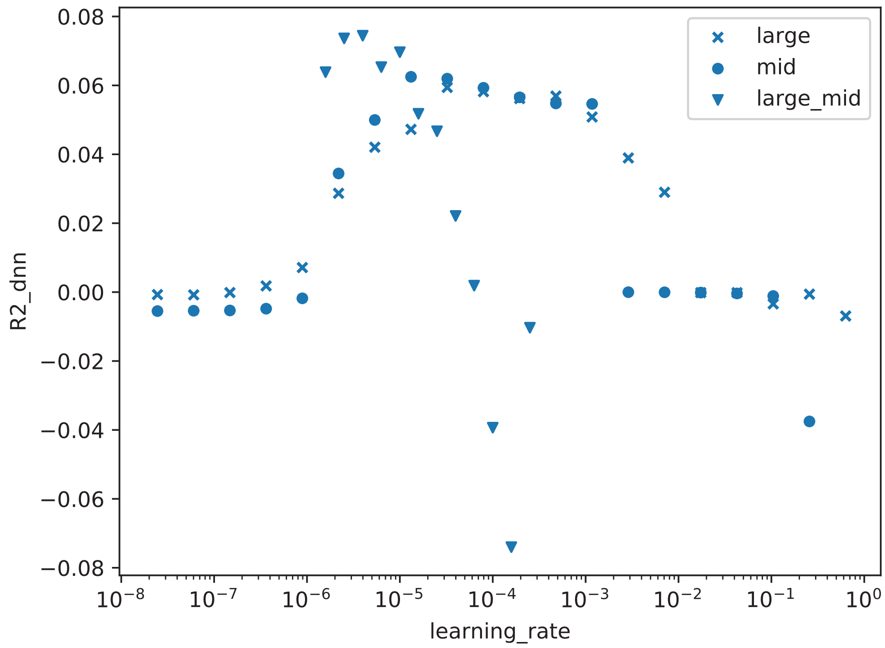

Figure 2.3: The figure shows the effect of learning-rate variation for DNN optimization. The x-axis captures the variation in the learning rate, the y-axis displays the resulting (annualized) R^2s. Strikingly, the variation is quite strong. Learning rates outside the range of 10^{-6}-10^{-2} produce mainly useless estimates. The optimal learning rate for large caps is roughly $10^{-4.5}$ and for mid caps $10^{-4.9}$. The maximum R^2 values are 5.94 % and 6.25% for the large and mid cap samples.

wider and more smooth. Towards the end of the optimization procedure, the predictions center more around 3, however, with a relative heavy left tail. Ill-behaved networks often show visually odd output distributions, for example, discontinued densities at multiple points or an extreme centering with hardly any variation at a specific point.

(a) Validation loss

(b) Output density function

Figure 2.4: Training Graphs: Figure 2.4a shows the training and validation loss as a function of the iteration steps, based on a random instance of the optimal DNN determined in 2.5.5. The loss functions behave as expected, as the training loss is mostly declining from left to right, whereas the validation score plateaus after around 45 steps (with a minimum at step 59) before it rises slightly towards the end. The optimization terminates after 97 steps as it reaches the maximum patience. Figure 2.4b depicts the histograms of the output values (cross-sectional returns) of the training data over all 97 iteration steps, based on the same instance of the optimal DNN determined in 2.4a. The distribution in the very front represents the final output prediction, the very first reflects the output values after the first iteration is completed. Note, the Graph is taken from Abadi et al. (2015) visualization tool "Tensorboard".

2.6 DFN portfolios

This section analyzes portfolio returns constructed based on deep-learning networks. It contains two different construction approaches, where each of them deals differently with the estimation instability of neural networks. The first subsection describes the estimation and portfolio construction mechanics. The second part contains a standard performance analysis of the portfolio returns, including an analysis conditioning on volatility and trading costs. Finally, I discuss the crucial FC of the prediction outcomes.

2.6.1 Estimation and portfolio construction

Table 2.2 and 2.3 reveal that training deep neural networks on cross-sectional returns is numerically not stable in this specific application. Hence, the researcher, or the investor, has to explicitly address the issues arising from it. I follow two different approaches. One way, which I call the **single prediction approach**, simply trains one network and continues with the return prediction and portfolios construction. Alternatively, a **forecast combination approach** can be applied with the objective of variance reduction at the prediction stage. Model estimation uncertainty, despite identical network hyper-parameters, arises because of variation in the training and validation data sets, but also because of the randomness in starting values of the numerical optimization.

The starting point for both portfolio approaches is the optimal network design found in the learning rate tuning in section 2.5.5. I follow an expanding window approach starting initially with twelve years of data (1/1/1970-12/31/1981). I split the data randomly along time points in a training (4/5) and a validation (1/5) data set and optimize the network by evaluating once more the validation results. I use these parameters for the network for the next twelve months for predicting one-month ahead returns, before re-estimating the same model with additional observations. The estimation frequency of twelve months is the result of the available computational resources. However, return predictions and portfolio weights are calculated on a monthly base by updating the FC. For example, the first prediction is obtained for 1/31/1982 as of 12/31/1981, using the trained network and FC input from 12/31/1981. The prediction for the other 11 months of 1982 all use the same network, with monthly FC input updates. Accordingly, the next network optimization expands the data window by one year, re-trains the network, and provides the first prediction for the first month in the next year. As a result, the predictions are based on past information only. Motivated by the estimation uncertainty, for the reasons mentioned above, this prediction exercise is repeated multiple times with the same network ingredients. It is to expect that each portfolio path realizes a different time-series of long-short returns. Ultimately, this leads to several predictions for each single excess return.[7] These predictions serve two purposes. First, it allows the investigation of the severity of the model uncertainty. Second, the predictions can be recycled for the alternative prediction exercise.

The two approaches differ at the prediction stage. The construction of the **single** prediction based portfolio simply uses the predicted return spreads and goes on with the portfolio formation. In contrast to the latter, the **combination** approach combines several single prediction outcomes. Given the 150 return predictions from the estimation and prediction exercise, one can form a forecast combination prediction straightforwardly. Forecast combination is popular technique, which is often successfully applied in the machine learning literature (Szegedy et al. (2013)), but it has also been employed to some classical financial prediction problems, see Rapach et al. (2010) for more details.[8] Primarily, it aims at variance reduction and is in the basic form an equal-weighted average of several prediction models. The idea can be related to portfolio diversification. Most prediction contains some idiosyncratic noise component, which is diversified away through averaging, like the idiosyncratic volatility of stocks in a diversified stock market index. The common systematic prediction component shared among all single predictions remains. Hence, I construct once more 150 portfolios, each based on an average of, for example, 75 randomly drawn predictions from the previous 150 series of naive $\hat{\mathbf{y}}$'s. To be precise the prediction for a particular stock n at time t used to form one

[7] I set the number to 150 for the large and mid cap stock sample and to 82 for the joint set of large and mid stocks. These limits are motivated by the computational resources available. The joint set of large and mid cap is much more computationally intensive as the the learning rate is lower and the data is twice the size compared to the other samples.

[8] Typically, forecast combinations pools predictions based over different models. The interpretation here is that each different prediction is a result of the same model with different parameters, which can be seen as ensemble of different models.

combination portfolio is,

$$\hat{y}_{n,t}^{\text{comb}} = \frac{1}{75}\sum_{i=1}^{75} \hat{y}_{i,n,t}^{*\text{single}}$$

where $\hat{y}_{i,n,t}^{*\text{single}} \in \mathcal{Y}$, whose 75 elements are randomly drawn (with replacement) from the set of the 150 single predictions obtained before.

The predictions of the linear benchmark are based on Fama and MacBeth (1973) regressions, where the variable selection is performed as in Green et al. (2017) (without correcting for a family-wise error). The re-estimation window and prediction intervals are identical as for the DFN approach.

Once the return predictions are collected, I form value-weighted (VW) and equal-weighted (EW) portfolios and calculate standard performance measures without considering any trading costs at first. I use a decile style cutoff procedure, by going long the upper 10% and short the lower 10% of the predicted return series.[9]

2.6.2 Performance analysis

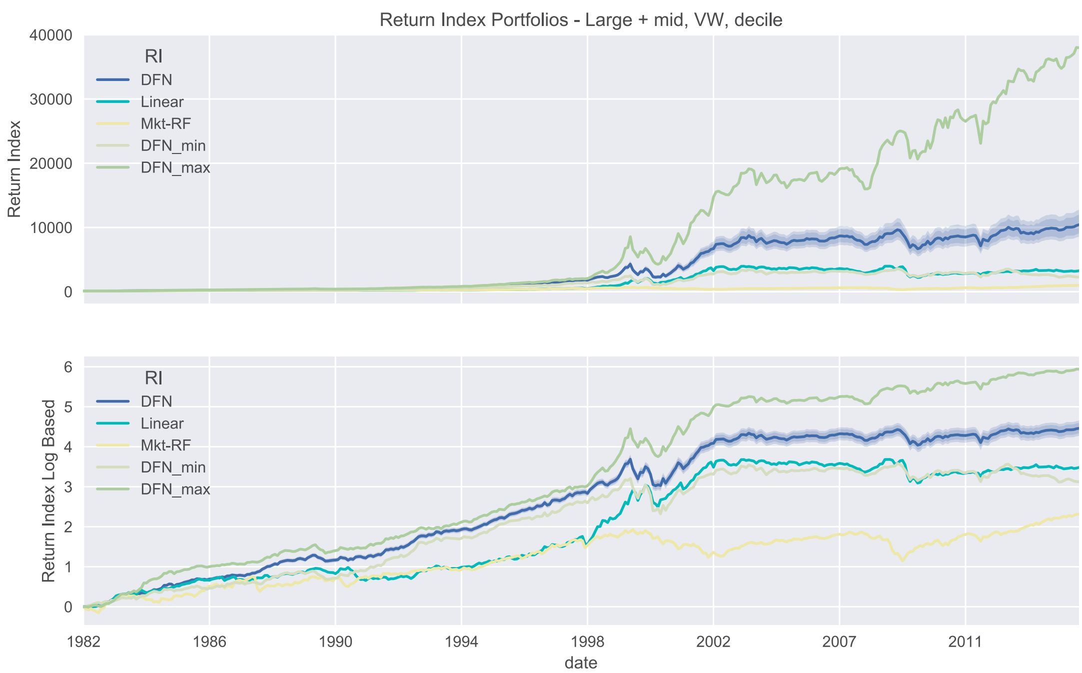

Figure 2.5: The figure exhibits selected paths of the 82 portfolios using large and mid cap stocks with value weighted decile long-short portfolios for the years 1982-2014. "DFN" reflects the average of the 82 portfolio paths, which is embedded into the corresponding 1% and 10%-confidence bounds.

[9]The appendix includes the results using a 30-70 or a Fama and French (1993) style (FF-style) cutoff procedure.

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	15.26	18.21	0.84	15.0***(4.57)	11.7***(3.46)	0.67
Min			11.16	17.90	0.62	10.4***(3.51)	7.1** (2.27)	-0.20
Max			19.68	17.55	1.12	17.6***(5.59)	14.5***(5.02)	1.75*
Linear			12.28	18.22	0.67	13.4***(3.82)	9.0***(2.95)	
Median	EW	single	17.96	15.88	1.13	14.5***(5.25)	10.7***(4.01)	1.07
Min			13.78	16.66	0.83	11.2***(3.54)	7.3** (2.50)	-0.03
Max			21.48	14.89	1.44	18.9***(7.07)	15.1***(6.41)	2.53**
Linear			14.49	17.36	0.83	17.1***(4.83)	12.6***(4.35)	
Median	VW	Comb 75	18.30	18.79	0.97	17.3***(5.17)	13.5***(3.92)	1.25
Min			17.12	18.70	0.92	15.9***(4.81)	12.3***(3.66)	0.99
Max			19.27	18.94	1.02	18.5***(5.51)	14.9***(4.28)	1.39
Linear			12.28	18.22	0.67	13.4***(3.82)	9.0***(2.95)	
Median	EW	Comb 75	19.88	17.01	1.17	17.6***(5.90)	13.2***(4.83)	1.22
Min			18.96	16.79	1.13	16.5***(5.67)	12.2***(4.51)	1.07
Max			20.27	16.87	1.20	17.8***(6.01)	13.4***(4.97)	1.31
Linear			14.49	17.36	0.83	17.1***(4.83)	12.6***(4.35)	

Table 2.4: Large + mid Cap Stocks, Decile Cutoffs, Sample Period 1982-01 until 2014-12: The table exhibits performance measures for several portfolio schemas. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 82 DFN portfolios are considered.

Table 2.4 reports the performance characteristics of the decile long-short portfolio strategies. Strikingly, the 82 single prediction approach portfolios realize a wide range of portfolio outcomes, as a result of the expected estimation uncertainty. Sharpe ratio (SR) between the "Min" and "Max" portfolio range from 0.62 to 1.12 and 0.83 to 1.44 for the value-weighted and equally-weighted portfolio, respectively. The median portfolio realizes a higher SR compared to the linear model in both cases, however, these differences are statistically not significant. Finally, the table includes two α estimates, one estimated with respect to the Fama and French (2014) 5-factor model and the other with a version extended by the classical momentum factor. Except of the minimum DFN realizations, all DFN portfolio paths realize higher α's compared to the linear benchmark. Despite the large performance variation, the worst performing portfolio over the sample period still competes reasonably well, compared to the linear benchmark based on the α's measures. Furthermore, you can see in Table 2.4 that the EW portfolios perform slightly better compared to the VW version, which holds for both the DFN and the linear model. This is consistent with the findings of Freyberger et al. (2017) and Lewellen (2015), as both report higher mean returns for equally weighted portfolios.

Furthermore, Figure 2.5 illustrates the corresponding value-weighted return indexes graphically. It emphasizes the economic relevance of the estimation uncertainty, as you can observe by the spread of the "Min" and the "Max" return index values. The linear as well as the mean

DFN portfolio realizes gains mostly prior to the year 2004. Furthermore, the figure visualizes that the outperformance relative to the linear model does not stem from a distinct period or event but rather accumulates gradually over time. In the last years from the beginning of 2004 to the end of 2014, both strategies flatten out and realize only marginal gains, partly a result of the losses suffered during the financial crisis. However, the DFN approach still realizes a higher return compared to the benchmark. Details on the sub-periods 1982-2003 and 2004-2014 can be found in the appendix.

One obvious problem associated with the DFN portfolios is the high uncertainty with respect to the potential return realizations. Most investors would prefer a less volatile return distribution, and would rather not be exposed to economically important risk arising from model optimization instability. The combination approach mitigates much of this risk, as you can see in the lower half of Table 2.4. As desired, one can observe a less volatile performance behavior for both VW and EW, simply measured by the difference of the "Min" and the "Max" realizations. Clearly, the objective of a variance reduction is achieved. As a result, the median portfolio shows a higher mean return, which results in a higher SR. The factor α's are on a similar level compared to the single predictions used before. Irrespective of the improvement, the Ledoit and Wolf (2008) test does not reject the null hypothesis of no difference in SR for the VW portfolio vs. the linear model. On the other hand, all Factor model α's associated with the DFN approach dominate the linear benchmark.

For both approaches, you can observe that a large fraction of the five-factor α's are absorbed by momentum exposure, as the difference between the regression α's is relatively high. Despite this difference, the momentum adjusted DFN α's are still mostly significant at the 1% level.

The same exercise can be repeated for **large** and **mid cap stocks** separately. The findings are documented in Table 2.5. The results are in line with the previously reported findings. The size separation reveals that the differences are stronger for the mid cap sample. Table 2.5 Panel A shows that for large cap strategies the median DFN combination approach yields an annualized return of 13.59% which compares to 9.74% for the linear model. This higher return is associated with a higher risk, consequently, the Sharpe ratio difference is only marginal and statistically not significant. However, once factor risks are controlled for the DFN dominate the competitor strategy in terms of α's. Generally, the mid cap performance statistics of both the linear and the DFN portfolio look impressive, as Table 2.5 Panel B exhibits. Differences between VW and EW are small, with slight advantages for the EW schema. The single based predictions consistently realize higher SR's compared to the benchmark. These differences are, except in case of the min DFN portfolio realizations, significant. The higher SR's are a consequence of higher means, as the volatility measures are comparable. Most DFN portfolios generate higher factor α's than the benchmark, once more the min portfolios fall short of the linear model. Forecast combination portfolios master the linear hurdle in all cases, as we can observe in Table 2.5 Panel B. All SR differences are significantly favoring the DFN approach at the 1% level in case of the max and median portfolios and at the 5% level worst performing DFN portfolio.

2.6.3 Volatility regimes

The DFN portfolios seem to be prone to losses during volatile periods as Figur B.3 indicates graphically. A more informative measure than visual inspection is the analysis of returns conditioning on the volatility state. For this purpose, I split the sample into two parts — a low- and high-volatility regime. The volatility classification into high and low is based on a standard GARCH(1,1) volatility estimation. Hansen and Lunde (2005) show that the parsimonious GARCH(1,1) is a reasonable choice to proxy the latent volatility process when compared to more sophisticated approaches. The low-volatility regime are all months which the volatility estimate belongs to the lowest 85% of the estimates, the high-volatility regime is then defined by the remaining months. We can see the conditional performance measures in Table 2.6. It documents large variations in average returns for the DFN strategy for all three samples, which perform better during less volatile times. For example, the difference in mean of the median portfolio for large cap stocks is 29.8 in annualized terms! The opposite holds for the linear approach in case of the large and mid sample, as the mean is higher during more uncertain periods. As a result, both strategies have a much different character during episodes of high volatility. This is reflected in the SR's, which are higher (lower) in the low (high) volatility state for the DFN strategy compared to the linear model. In case of the joint set of large and mid cap stocks, differences between the DFN and the linear approach are only notable during the low volatility periods.

Moreover, despite the negative mean, a large α is measured in case of the momentum augmented five-factor model in the high volatility regime. As pure momentum-strategies are known to be prone to crashes, see for example Daniel and Moskowitz (2016), multi-signal based long-short portfolios load momentum exposure but indicate more resilience to momentum drawdowns.

2.6.4 Trading costs

The previous section shows attractive return characteristics of both strategies considered. Naturally, the question arises how much do the performance measures deteriorate once trading costs are accounted for. Hence, I use the same portfolios obtained in the previous section, but explicitly correct them for trading costs. The definition of the trading cost function is as in Brandt et al. (2009), Hand and Green (2011) and DeMiguel et al. (2017). The approach

Table 2.5: Long-Short PF Summary, Decile Cutoffs, Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for several portfolio schemas. Panel A reflects purely large cap stocks, Panel B only mid cap stocks. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 150 DFN portfolios are considered.

Panel A: **Large Cap Stocks**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	12.47	18.20	0.69	15.0***(4.81)	11.2***(4.03)	0.0788
Min			8.45	18.12	0.47	10.5***(3.26)	7.2** (2.26)	-0.64
Max			17.28	17.99	0.96	19.0***(5.69)	15.0***(5.35)	0.987
Linear			9.74	14.72	0.66	4.9* (1.93)	3.3 (1.28)	
Median	EW	single	12.12	16.39	0.74	13.8***(4.84)	9.8***(3.95)	0.48
Min			8.27	18.12	0.46	7.2** (2.03)	4.1 (1.19)	-0.441
Max			16.42	16.77	0.98	18.0***(6.10)	13.8***(5.88)	1.12
Linear			9.07	15.60	0.58	3.9 (1.54)	2.3 (0.89)	
Median	VW	Comb 75	13.59	18.82	0.72	14.6***(4.72)	10.7***(3.54)	0.212
Min			12.38	18.74	0.66	12.9***(4.11)	8.8***(3.10)	-0.00352
Max			14.57	19.34	0.75	15.3***(4.66)	11.0***(3.75)	0.321
Linear			9.74	14.72	0.66	4.9* (1.93)	3.3 (1.28)	
Median	EW	Comb 75	13.03	17.34	0.75	12.2***(4.05)	8.0***(2.96)	0.538
Min			12.50	17.51	0.71	12.3***(4.10)	8.0***(2.97)	0.415
Max			13.57	17.37	0.78	13.0***(4.42)	8.8***(3.39)	0.626
Linear			9.07	15.60	0.58	3.9 (1.54)	2.3 (0.89)	

Panel B: **Mid Cap Stocks**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	20.01	16.67	1.20	16.9***(6.28)	12.8***(4.57)	1.89*
Min			16.72	17.16	0.97	12.9***(4.41)	8.7***(3.10)	0.752
Max			23.30	16.81	1.39	23.1***(7.58)	18.4***(6.78)	2.48**
Linear			13.43	16.84	0.80	15.6***(6.64)	12.0***(5.18)	
Median	EW	single	20.26	16.93	1.20	18.9***(6.77)	14.3***(5.74)	2.14**
Min			17.23	16.49	1.04	16.2***(5.70)	12.3***(4.53)	1.19
Max			23.37	16.33	1.43	22.0***(7.54)	17.4***(7.42)	2.82***
Linear			13.47	17.69	0.76	16.2***(6.72)	12.5***(5.54)	
Median	VW	Comb 75	22.65	16.65	1.36	19.7***(7.08)	15.2***(5.87)	2.62***
Min			21.93	16.65	1.32	18.6***(6.75)	14.2***(5.53)	2.34**
Max			23.25	16.42	1.42	20.1***(7.35)	15.7***(6.21)	2.89***
Linear			13.43	16.84	0.80	15.6***(6.64)	12.0***(5.18)	
Median	EW	Comb 75	22.57	16.62	1.36	19.9***(7.11)	15.7***(6.20)	2.71***
Min			22.09	16.84	1.31	19.4***(6.82)	15.0***(5.91)	2.54**
Max			23.03	16.51	1.39	20.8***(7.46)	16.6***(6.54)	2.8***
Linear			13.47	17.69	0.76	16.2***(6.72)	12.5***(5.54)	

Size	Volatility	Portfolio	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
l+m	high	DFN	-0.16	31.90	-0.00	5.9 (0.46)	32.7***(3.15)	-0.04
		Linear	0.55	30.00	0.02	2.4 (0.14)	31.1***(2.99)	
	low	DFN	21.57	15.15	1.42	17.6***(5.33)	9.1***(3.22)	1.96**
		Linear	14.38	15.17	0.95	16.0***(3.93)	5.4* (1.73)	
large	high	DFN	-11.64	31.25	-0.37	1.2 (0.11)	28.9***(3.18)	-1.5
		Linear	15.15	22.57	0.67	12.6 (1.07)	29.7***(3.30)	
	low	DFN	18.22	15.56	1.17	16.2***(4.65)	6.7** (2.38)	1.6
		Linear	8.77	12.84	0.68	2.8 (1.01)	-0.7 (-0.26)	
mid	high	DFN	3.33	26.23	0.13	-8.8 (-0.64)	23.2***(3.31)	-0.784
		Linear	16.67	29.57	0.56	-5.5 (-0.38)	22.2***(3.02)	
	low	DFN	26.06	14.13	1.84	23.2***(7.59)	14.2***(5.92)	4.05***
		Linear	12.85	13.42	0.96	16.8***(6.57)	9.8***(4.89)	

Table 2.6: Volatility Regimes, Decile Cutoffs, VW, Sample Period 1982-01 until 2014-12: The table depicts performance measures conditional on the market volatility regime. The regime "high" are all months in which the GARCH(1,1) volatility estimate belongs to the 15% highest estimates, "low" accordingly to the lowest 85%. The DFN portfolio refers to the median forecast combination portfolio ("Comb 75"). All performance measures are displayed in annual terms. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark.

explicitly factors in the stock size (decreasing in size) and through time (decreasing from year to year). The function is defined as follows:

$$c_{n,t} = \underbrace{\left(1+3\frac{\max\left(\Delta_{\text{days}}(date(t),1/1/2002),0\right)}{\Delta_{\text{days}}(1/1/1974,1/1/2002)}\right)}_{\text{time effect}}\underbrace{(0.006-0.0025\times(1-z(\text{ME}_{n,t})))}_{\text{size effect}}, \qquad (2.4)$$

where $\Delta_{\text{days}}(a,b)$ counts the difference in days (date b minus date a), and $z(\text{ME}_{n,t}))$ measures the normalized market capitalization of stock n at time t.[10] Hence, the highest trading cost in the analysis is observed for the smallest stock at the beginning of our portfolio formation period at 1982/1/1 of around 190 basis points (BP) and 35 BP for the largest stock post the time trend end date on 1/1/2002.

Table 2.7 shows the adjusted performance characteristics. Not surprisingly, given a monthly average turnover (TO) of 246% and 324%, respectively. Consequently, none of the two strategies realizes positive mean returns when portfolio weights are updated monthly and trading costs are accounted for (the TO impact for the single strategy is very close to the combination approach, hence, they are not reported). All realize a TO of significant magnitude, notable, the linear model realizes the highest TO for all frequencies considered. A trivial TO reduction measure is to reduce the trading frequency. Hence, I consider several rebalancing policies, varying the trading activity monthly ranging from one to six months. This

[10]Large cap stock and mid cap stock market capitalization are always calculated w.r.t to the entire cross-section.

trade-off is reported in table 2.7. For example, the TO decreases around 45-50% for the two portfolios when switching from a monthly rebalancing to only one every 2nd month. Despite an almost monotonic decline for both strategies the mean returns do not fall dramatically, hence, reducing the rebalancing frequencies offers a more attractive cost-return balance. In this example, the maximum cost adjusted return for the DFN long-short portfolio is achieved with a frequency of five months for both the DFN and the linear model.

Freq	Model	Costs	Mean	Std	SR	TO-%	FF-5 α	FF-5 + Mom α	SR Test
1	Comb	No	15.15	18.16	0.83		15.0***(4.57)	11.6***(3.46)	0.65
		Yes	-13.33	18.11	-0.74	246	-13.2***(-4.11)	-16.4***(-5.07)	2.69***
	Linear	No	12.28	18.22	0.67		13.4***(3.82)	9.0***(2.95)	
		Yes	-24.88	18.34	-1.36	324	-23.2***(-6.33)	-27.5***(-8.76)	
2	Comb	No	14.74	17.58	0.84		14.9***(4.31)	11.9***(3.37)	0.46
		Yes	-0.89	18.27	-0.05	136	-0.5 (-0.14)	-3.3 (-0.95)	1.30
	Linear	No	12.49	17.46	0.72		13.2***(3.59)	9.5***(3.09)	
		Yes	-6.34	18.18	-0.35	165	-5.2 (-1.41)	-8.8***(-2.83)	
3	Comb	No	11.23	17.85	0.63		12.6***(3.65)	9.9***(2.58)	-0.04
		Yes	0.05	18.75	0.00	98	1.5 (0.43)	-1.3 (-0.33)	0.31
	Linear	No	11.14	17.38	0.64		13.6***(3.78)	11.3***(3.13)	
		Yes	-1.38	18.53	-0.07	111	1.3 (0.34)	-1.2 (-0.32)	
4	Comb	No	7.90	17.97	0.44		8.7***(2.65)	5.8* (1.85)	-0.03
		Yes	-0.86	18.62	-0.05	77	-0.1 (-0.02)	-2.8 (-0.88)	0.04
	Linear	No	8.41	18.84	0.45		10.3***(2.66)	7.5** (2.18)	
		Yes	-1.10	19.37	-0.06	84	0.7 (0.17)	-1.9 (-0.53)	
5	Comb	No	11.96	17.37	0.69		14.6***(4.45)	12.5***(3.68)	0.76
		Yes	4.82	18.30	0.26	64	7.2** (2.10)	5.0 (1.41)	0.61
	Linear	No	10.13	20.02	0.51		15.9***(4.05)	13.3***(3.27)	
		Yes	2.72	20.65	0.13	66	8.3** (2.04)	5.6 (1.33)	
6	Comb	No	8.05	18.35	0.44		8.8***(2.76)	6.0* (1.83)	-0.13
		Yes	1.84	18.91	0.10	55	2.6 (0.80)	-0.2 (-0.06)	0.09
	Linear	No	7.47	15.76	0.47		10.4***(3.09)	8.6***(2.90)	
		Yes	1.21	16.52	0.07	56	4.3 (1.24)	2.5 (0.81)	

Table 2.7: Trading Cost Adjusted Performance, Large + Mid Cap Stocks, Decile Cutoffs, VW Weighted: The table shows the performance impact once trading costs are explicitly accounted for. The column "Costs" compares the results with (Yes) and without (No) trading costs. Trading costs are calculated as in Brandt et al. (2009). TO-% indicates the average of the monthly portfolio turnover (the upper bound of a long-short portfolio is 400%). The column "Freq" refers to the rebalancing frequency in months.

2.6.5 Which FC drive the prediction?

Another important aspect is to understand which FC are the fundamental drivers behind the predictions. It is intrinsically hard to assess the relationship between a FC and the predicted returns as the DFN has no clear interpretation available, comparable to a regression coefficient in the standard linear model. The output, a result of a nested structure, can theoretically be traced back to the input level, however, it is a rather challenging endeavor. Alternatively, I can manipulate the input matrix. Consequently, I separately replace each FC input vector with zeros and measure the changes compared to the predicted output when all FC are included. The measure I construct is a mean-squared deviation (MSD). Precisely, the MSD is calculated as follows:

$$\mathrm{MSD}_{t,k,p} = \frac{1}{N}\sum_{n=1}^{N}(\hat{f}_p(\mathbf{x}_{t,n}) - \hat{f}_p(\mathbf{x}_{t,n})|_{x_{t,n,k}=0})^2 \tag{2.5}$$

where $\hat{f}_p$ is the function estimate of run p. Alternatives, like a subsample analysis, is computationally prohibitive but would, in general, be preferable. Hence, the MSD approach allows model insights at almost zero computational costs.[11]

Table 2.8 displays the average MSD for the top 10 FC (from highest to lowest) for the three samples of interest.[12] The two predominant FC for the three subsamples are the short-term reversal and the twelve months momentum. Moreover, price based FC have by far the largest impact on the prediction outcomes, as the top five and the top eight FC, respectively, are purely reflecting return information. Figures 2.6 - 2.8 exhibit this behavior over time. Strikingly, the pictures are dominated by only very few FC over time, indicating that predictions are insensitive to the majority of FC included in this work. Notable, short-term reversal becomes slightly less relevant in case of large cap stocks towards the end of the sample period, the dominating role is taken over by the twelve months momentum. Another interesting observation is the impact of the stock market correction around the beginning of the 2000s, which leads to a short distortion of the FC sensitivities. Finally, towards the end of the sample period, a slight increase of other FC can be documented, as the colors partially darken in the upper half of the FC.

[11] A thorough subsample approach takes sequentially one FC out and repeats the procedures presented in this work. Consequently, this would consume approximately 68 times the resources required to conduct the sensitivity check.

[12] The average is calculated as follows: $M\bar{S}D_k = \frac{1}{TP}\sum_{t=1}^{T}\sum_{p=1}^{P}\mathrm{MSD}_{t,k,p}$.

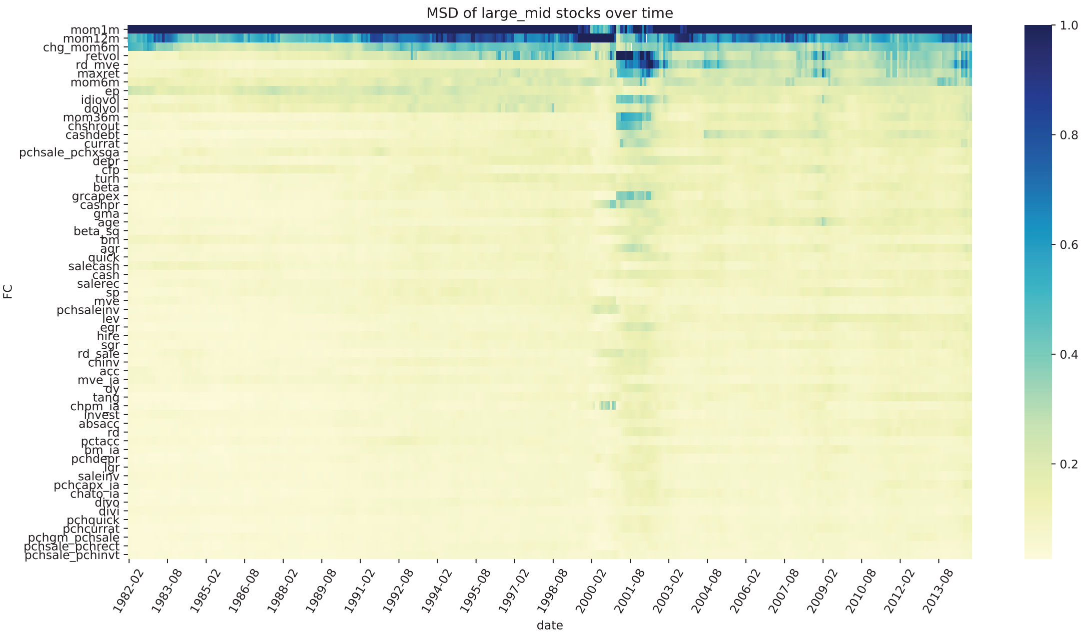

Figure 2.6: This sample includes the joint set of **large and mid cap stocks** and data from 1982-2014. The figure shows the expanding window of the MSD as calculated in equation 2.5 for all FC (sorted from top to bottom by the pooled average, from highest to lowest). The values are cross-sectionally normalized to range between 0 and 1. The highest value is obtained for *mom1m*, followed by *mom12m*. Moreover, the top five FC are all solely based on past price information.

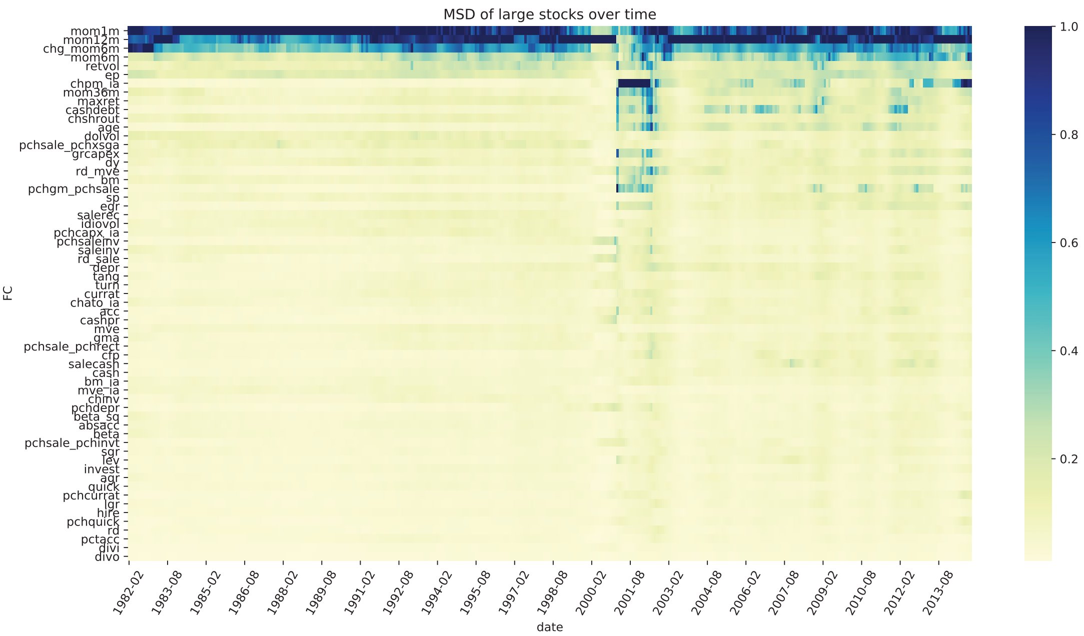

Figure 2.7: This sample includes **large cap stocks** and data from 1982-2014. The figure shows the expanding window of the MSD as calculated in equation 2.5 for all FC (sorted from top to bottom by the pooled average, from highest to lowest). The values are cross-sectionally normalized to range between 0 and 1. The highest value is obtained for *mom1m*, followed by *mom12m*. Moreover, the top five FC are all solely based on past price information.

	Size	1	2	3	4	5	6	7	8	9	10
FC	l+m	mom1m	mom12m	chg_mom6m	retvol	maxret	rd_mve	mom6m	ep	idiovol	dolvol
MSD		1.00	0.66	0.35	0.28	0.21	0.20	0.20	0.19	0.18	0.15
FC	large	mom1m	mom12m	chg_mom6m	mom6m	retvol	ep	chpm_ia	mom36m	maxret	cashdebt
MSD		1.00	1.00	0.60	0.32	0.20	0.20	0.18	0.14	0.13	0.13
FC	mid	mom1m	mom12m	maxret	retvol	mom6m	idiovol	chg_mom6m	age	rd_mve	depr
MSD		1.00	0.80	0.38	0.30	0.22	0.15	0.14	0.12	0.11	0.10

Table 2.8: The table presents the MSD pooled over time and predictions conditioned on the respective size sample. The MSD is calculated asin equation 2.5. Row "l+m" represents the joint sample of large and mid caps. The values are normalized with respect to the maximum MSD value in each subsample.

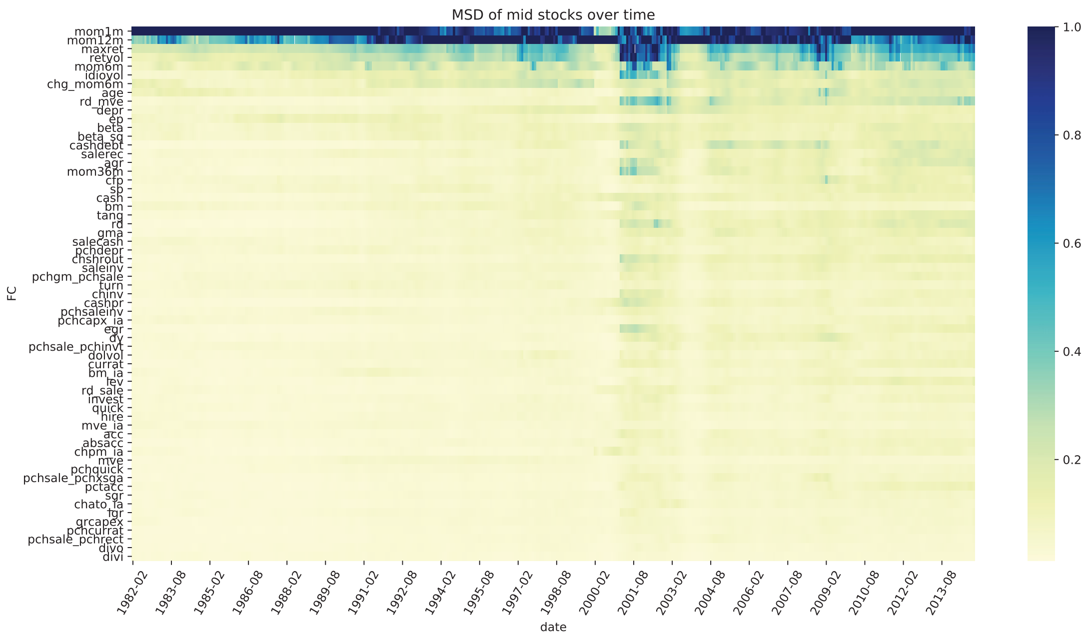

Figure 2.8: This sample includes **mid cap stocks** and data from 1982-2014. The figure shows the expanding window of the MSD as calculated in equation 2.5 for all FC (sorted from top to bottom by the pooled average, from highest to lowest). The values are cross-sectionally normalized to range between 0 and 1. The highest value is obtained for *mom1m*, followed by *mom12m*. Moreover, the top five FC are all solely based on past price information.

2.7 Conclusion

The study shows that deep feedforward neural networks provide a framework which prediction improvements of cross-sectional returns can be achieved when compared to a linear benchmark. One of the key tasks here is the identification of an optimal network design, which is robust to numerical deficiencies and competitive in terms of its prediction accuracy. I document that the majority of randomly determined architectures fail numerically, others perform poorly relative to the parsimonious benchmark or suffer from a stark estimation instability. Once these shortcomings are carefully bridged, long-short portfolios with attractive risk-adjusted returns can be constructed. I document that most predictions are driven by price-based FC. Moreover, a large fraction of FC have a negligible impact on the predictions. However, one cannot detangle whether these market inefficiencies stem from mispricing or are a result of computational and technological constraints faced by investors at the time.

This work is basic by design, and there are many ways to improve the deep-learning application. Moreover, slight deviations in the objective functions might cater other purposes more appropriately. Other obvious limitations are the specific selection of the underlying FC, which can be altered arbitrarily.

Furthermore, the implications from an investors perspective are manifold. Generally, deep-learning appears to be an attractive approach to stock selection. However, many problems need to be addressed first. For example, the rebalancing strategy needs to be optimized as a result of the prohibitive trading costs associated with a plain portfolio sorting strategy. Second-moments need to be accounted for and the risk capacity during crises periods has to be managed. These are all open questions, which are left for future research.

Finally, this study displays that non-linearities are important determinants of expected cross-sectional return spreads. There is no claim that deep-learning is the best way of exploiting these non-linearities. However, a distinct feature of this work is that it does not impose any prior on the functional relation between FC and the expected return process. The machine learns this non-linear dependence structure purely by itself.

Appendix B

B.1 Regularization tools

B.1.1 Early stopping

Early stopping is an algorithm to reduce over-fitting. It prevents the training process of the network from purely memorizing the training data, by evaluating the validation error. Goodfellow et al. (2016), Bishop (1995b) and Sjöberg and Ljung (1995) show the regularization property of early stopping, which displays a similar character as L^2 regularization. It implicitly constrains the parameter space, due to the limited distance gradient-descent can travel during training. The algorithm terminates training the network after the validation error couldn't be improved for a certain amount of iterations. The number of iterations allowed, without seeing improvements, is called patience. As a side effect it limits the number of training iterations, and hence, increases computational efficiency.

B.1.2 Dropout

Srivastava et al. (2014) introduce Dropout regularization to the machine learning literature. The idea underlying the method is to randomly drop units during training. It can be interpreted as a form of bagging.[1] Figure B.1 illustrates graphically the working mechanism applied to the basic network example illustrated in figure 2.1. Mechanically one can attach an independent Bernoulli random variable to each of the nodes with shared (Dropout) probability p - a hyper-parameter. The Bernoulli random variables are then drawn prior to each mini-batch process, hence, during neural network training a large number of subnetwork is actually optimized. More details can be found in Srivastava et al. (2014).

[1] Bagging is an ensemble strategy which combines several predictions models to compute a weighted prediction output, profiting from variance reduction.

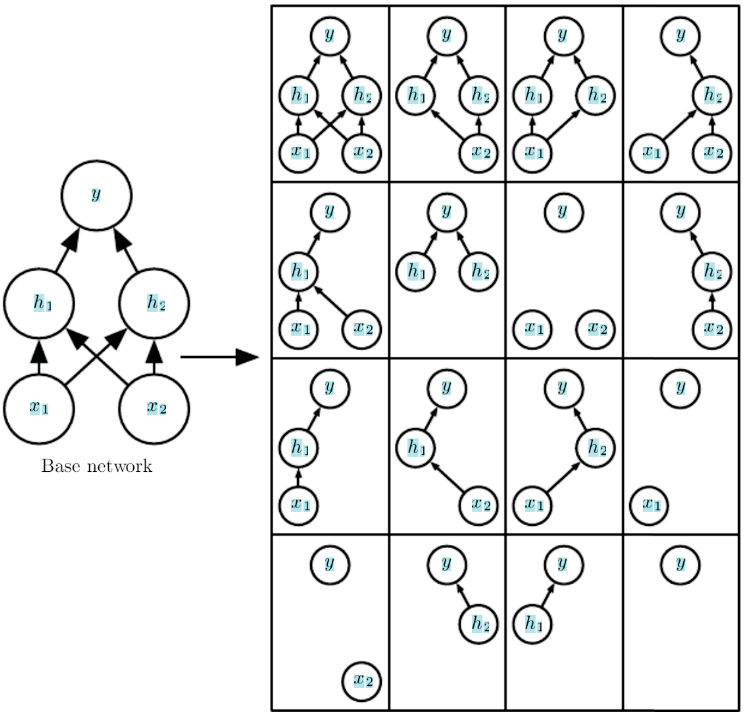

Figure B.1: Dropout DFN: Principal working mechanism of dropout regulariation anatomy of a fully connected deep neural network with one hidden layer and two units. The input variables x_1 and x_2 mark the input layer, h_1 and h_2 and two hidden units in the only hidden layer, y marks the output. W and w defines the weight matrix. Illustration from Goodfellow et al. (2016), page 256.

B.1.3 L1 and L2 regularization

Norm regularization aim at variance reduction through explicit penalization of the size of the model parameters, which shrink towards zero, the heavier the penalization is pronounced. Specifically L1 and L2 constraints can be imposed, which are known in the linear specification as Lasso, Tibshirani (1996), and Ridge Regression methods. These shrinkage approaches can simply be extended to neural networks in general and specifically to the optimization problem above by adding the corresponding terms to the cost function in equation 2.2,

$$\tilde{J}(\theta) = \hat{J}(\theta) + \alpha_1 \|\mathbf{W}\|_1 + \alpha_2 \|\mathbf{W}\|_2^2, \tag{B.1}$$

where $\|W\|_1 = \sum_j^J \|w_j\|$ and $\|W\|_2^2 = \sum_j^J w_j^2$ with J the total number of weight parameters. Note, the bias terms are usually not included into the penalization forms.

B.2 Computational remarks

Despite computational benefits of random search, optimizing hyper-parameters can still be a computationally exhaustive task - ultimately depending on the number of trials performed and the allowed model complexity. Access to large CPU/GPU resources are crucial. This work is feasible only through access to a high performance computing grid and cloud computing resources. RAM is typically not the critical factor. I note that Abadi et al. (2015) provides two implementations, one for standard CPU designs and the other for modern GPU architecture. GPUs offer often significant performance benefits over classical CPUs design, and hence, can be a crucial factor in training and model capacity. This work is exclusively trained on CPU machines.

B.3 Additional tables

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	8.95	10.21	0.88	8.2***(4.18)	7.2***(3.51)	0.91
Min			6.02	11.21	0.54	6.6***(3.47)	4.8** (2.43)	-0.29
Max			10.91	10.38	1.05	10.8***(6.26)	9.6***(5.49)	1.57
Linear			7.66	12.53	0.61	8.8***(3.98)	6.1***(2.94)	
Median	EW	single	11.27	8.64	1.30	10.1***(6.41)	8.1***(5.03)	1.72*
Min			9.08	10.83	0.84	8.2***(4.23)	5.8***(3.08)	-0.04
Max			13.53	8.74	1.55	11.0***(6.69)	9.0***(5.53)	2.59***
Linear			9.84	11.59	0.85	11.3***(4.86)	8.6***(4.28)	
Median	VW	Comb 75	10.98	11.07	0.99	11.1***(6.29)	9.5***(5.10)	1.46
Min			10.32	11.01	0.94	10.4***(5.86)	8.9***(4.70)	1.24
Max			11.55	11.10	1.04	11.7***(6.36)	10.0***(5.18)	1.68*
Linear			7.66	12.53	0.61	8.8***(3.98)	6.1***(2.94)	
Median	EW	Comb 75	12.98	10.19	1.27	11.6***(6.59)	9.1***(5.21)	1.51
Min			12.73	10.03	1.27	11.2***(6.43)	8.7***(5.09)	1.52
Max			13.29	10.25	1.30	11.9***(6.79)	9.4***(5.35)	1.60
Linear			9.84	11.59	0.85	11.3***(4.86)	8.6***(4.28)	

Table B.1: Large + mid Cap Stocks, FF-style Cutoffs, Sample Period 1982-01 until 2014-12: The table exhibits performance measures for several portfolio schemas. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 82 DFN portfolios are considered.

Table B.2: Long-Short PF Summary, FF-Style Cutoffs, Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for several portfolio schemas. Panel A reflects purely large cap stocks, Panel B only mid cap stocks. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 150 DFN portfolios are considered.

Panel A: **Large Cap Stocks**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	7.36	10.02	0.73	7.7***(4.09)	6.8***(3.61)	0.621
Min			4.69	11.13	0.42	4.3** (2.21)	2.8 (1.41)	-0.417
Max			10.22	9.42	1.08	10.2***(6.32)	8.7***(5.80)	1.93*
Linear			5.93	10.96	0.54	2.1 (1.14)	1.1 (0.58)	
Median	EW	single	7.85	9.54	0.82	7.3***(4.33)	5.7***(3.25)	0.821
Min			5.47	10.41	0.53	5.4***(2.81)	3.7** (1.98)	-0.0673
Max			10.28	8.18	1.26	9.0***(6.35)	7.3***(5.59)	2.33**
Linear			5.80	10.59	0.55	2.0 (1.12)	1.1 (0.59)	
Median	VW	Comb 75	8.29	10.60	0.78	8.5***(4.91)	6.8***(3.97)	0.846
Min			7.82	10.61	0.74	7.5***(4.32)	5.9***(3.33)	0.704
Max			9.03	10.69	0.84	8.8***(5.13)	7.1***(4.21)	1.06
Linear			5.93	10.96	0.54	2.1 (1.14)	1.1 (0.58)	
Median	EW	Comb 75	8.70	10.12	0.86	7.9***(4.48)	5.9***(3.34)	0.924
Min			8.47	10.37	0.82	7.9***(4.38)	6.0***(3.23)	0.783
Max			9.03	10.07	0.90	8.3***(4.80)	6.3***(3.67)	1.01
Linear			5.80	10.59	0.55	2.0 (1.12)	1.1 (0.59)	

Panel B: **Mid Cap Stocks**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	11.68	10.97	1.06	10.5***(5.72)	7.6***(4.37)	1.06
Min			9.71	11.37	0.85	9.6***(5.13)	6.8***(3.62)	0.204
Max			14.21	10.47	1.36	12.3***(6.92)	9.4***(6.14)	2.2**
Linear			9.10	11.34	0.80	9.5***(5.97)	7.0***(4.25)	
Median	EW	single	11.73	10.80	1.09	11.3***(6.42)	8.4***(5.24)	1.49
Min			9.69	11.22	0.86	9.8***(5.23)	7.0***(3.90)	0.354
Max			14.39	10.34	1.39	12.6***(7.12)	9.8***(6.84)	2.57**
Linear			9.11	11.73	0.78	10.0***(6.19)	7.5***(4.80)	
Median	VW	Comb 75	12.14	10.60	1.15	10.6***(6.10)	7.8***(4.53)	1.74*
Min			11.87	10.70	1.11	10.4***(5.97)	7.6***(4.35)	1.54
Max			12.51	10.67	1.17	10.9***(6.21)	8.0***(4.68)	1.82*
Linear			9.10	11.34	0.80	9.5***(5.97)	7.0***(4.25)	
Median	EW	Comb 75	12.26	10.66	1.15	11.2***(6.33)	8.3***(5.13)	1.95*
Min			11.93	10.68	1.12	10.8***(6.16)	8.0***(4.94)	1.83*
Max			12.67	10.53	1.20	11.3***(6.56)	8.5***(5.43)	2.25**
Linear			9.11	11.73	0.78	10.0***(6.19)	7.5***(4.80)	

Table B.3: Large + Mid Cap Stocks Long-Short PF Summary, Decile Cutoffs, Subperiods: The tables exhibit performance measures for several portfolio schemas. Panel A reflects purely large cap stocks, Panel B only mid cap stocks. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 82 DFN portfolios are considered.

Panel A: **Sample Period 1982-01 until 2003-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	21.25	17.82	1.19	25.2***(6.72)	20.8***(5.10)	0.64
Min			16.37	18.34	0.89	14.6***(3.53)	9.5** (2.30)	-0.20
Max			26.91	18.81	1.43	28.7***(8.26)	23.7***(5.81)	1.39
Linear			18.67	19.46	0.96	21.5***(4.50)	14.5***(3.55)	
Median	EW	single	24.53	16.58	1.48	22.8***(6.55)	17.2***(4.73)	0.79
Min			19.84	18.51	1.07	18.4***(4.37)	12.3***(3.14)	-0.30
Max			30.09	15.83	1.90	28.1***(7.90)	22.7***(5.92)	1.86*
Linear			21.61	18.27	1.18	26.1***(5.44)	19.4***(4.94)	
Median	VW	Comb 75	25.45	19.84	1.28	26.8***(6.53)	20.9***(4.58)	1.00
Min			24.17	19.57	1.24	25.2***(6.35)	19.5***(4.34)	0.89
Max			26.81	20.18	1.33	27.9***(6.65)	21.9***(4.67)	1.16
Linear			18.67	19.46	0.96	21.5***(4.50)	14.5***(3.55)	
Median	EW	Comb 75	27.52	18.46	1.49	26.0***(6.76)	19.1***(5.22)	0.80
Min			26.37	18.18	1.45	24.8***(6.58)	18.1***(4.91)	0.69
Max			27.99	18.07	1.55	26.5***(7.11)	19.8***(5.34)	0.94
Linear			21.61	18.27	1.18	26.1***(5.44)	19.4***(4.94)	

Panel B: **Sample Period 2004-01 until 2014-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	2.42	14.44	0.17	-1.8 (-0.45)	-2.0 (-0.51)	0.41
Min			-5.54	14.62	-0.38	-7.5* (-1.69)	-7.9** (-2.00)	-1.02
Max			7.90	13.96	0.57	5.5 (1.40)	5.1 (1.44)	1.22
Linear			-0.43	14.84	-0.03	-0.8 (-0.21)	-1.3 (-0.35)	
Median	EW	single	3.96	10.78	0.37	4.3 (1.42)	3.8 (1.51)	1.11
Min			-0.39	13.32	-0.03	-0.1 (-0.02)	-0.7 (-0.21)	-0.23
Max			8.37	10.38	0.81	7.1** (2.53)	6.7***(2.82)	2.13**
Linear			0.35	14.62	0.02	2.0 (0.55)	1.3 (0.39)	
Median	VW	Comb 75	4.16	15.78	0.26	0.4 (0.10)	0.0 (0.00)	0.82
Min			2.55	15.71	0.16	-1.1 (-0.25)	-1.5 (-0.37)	0.54
Max			6.18	15.65	0.39	3.0 (0.68)	2.6 (0.63)	1.15
Linear			-0.43	14.84	-0.03	-0.8 (-0.21)	-1.3 (-0.35)	
Median	EW	Comb 75	4.61	12.46	0.37	3.1 (0.90)	2.5 (0.87)	1.38
Min			3.51	13.06	0.27	2.2 (0.60)	1.6 (0.51)	0.98
Max			5.56	12.33	0.45	4.0 (1.19)	3.4 (1.24)	1.62
Linear			0.35	14.62	0.02	2.0 (0.55)	1.3 (0.39)	

Table B.4: Large Cap Stocks Long-Short PF Summary, Decile Cutoffs, Subperiods: The tables exhibit performance measures for several portfolio schemas. Panel A reflects purely large cap stocks, Panel B only mid cap stocks. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 150 DFN portfolios are considered.

Panel A: **Sample Period 1982-01 until 2003-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	18.52	20.12	0.92	18.6***(4.68)	14.0***(3.04)	0.0509
Min			13.07	20.63	0.63	16.3***(4.16)	10.8** (2.29)	-0.709
Max			23.30	18.47	1.26	23.3***(6.31)	18.3***(4.02)	0.944
Linear			13.87	15.39	0.90	8.7***(2.64)	7.2** (2.29)	
Median	EW	single	18.53	18.72	0.99	20.5***(5.42)	13.7***(4.49)	0.293
Min			12.72	19.26	0.66	14.7***(3.69)	10.0** (2.25)	-0.502
Max			22.65	17.68	1.28	24.6***(6.56)	17.6***(5.33)	0.992
Linear			14.40	16.81	0.86	8.3***(2.78)	7.1** (2.50)	
Median	VW	Comb 75	20.06	21.06	0.95	23.6***(6.00)	16.6***(4.20)	0.139
Min			18.75	20.09	0.93	20.9***(5.35)	14.3***(3.84)	0.0853
Max			21.14	20.40	1.04	23.7***(6.24)	17.1***(4.36)	0.358
Linear			13.87	15.39	0.90	8.7***(2.64)	7.2** (2.29)	
Median	EW	Comb 75	19.80	19.05	1.04	21.2***(5.69)	14.2***(3.90)	0.406
Min			19.17	19.08	1.01	20.7***(5.53)	13.6***(3.77)	0.333
Max			20.61	19.26	1.07	21.7***(5.67)	14.5***(3.90)	0.477
Linear			14.40	16.81	0.86	8.3***(2.78)	7.1** (2.50)	

Panel B: **Sample Period 2004-01 until 2014-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	0.71	13.92	0.05	-1.6 (-0.36)	-2.2 (-0.57)	-0.16
Min			-5.64	14.50	-0.39	-7.4* (-1.77)	-7.7** (-1.97)	-1.36
Max			6.30	12.17	0.52	3.6 (0.93)	3.3 (0.91)	0.863
Linear			1.51	13.00	0.12	-0.1 (-0.02)	-0.5 (-0.15)	
Median	EW	single	-0.22	11.17	-0.02	-4.5 (-1.57)	-4.8* (-1.82)	0.253
Min			-4.58	11.73	-0.39	-6.1* (-1.93)	-6.5** (-2.32)	-0.822
Max			4.51	10.06	0.45	2.5 (0.84)	2.1 (0.85)	1.33
Linear			-1.51	12.38	-0.12	-0.4 (-0.12)	-0.9 (-0.26)	
Median	VW	Comb 75	0.99	14.43	0.07	-1.3 (-0.28)	-1.7 (-0.43)	-0.121
Min			-0.49	15.05	-0.03	-3.6 (-0.78)	-4.1 (-1.01)	-0.402
Max			3.17	14.69	0.22	0.9 (0.21)	0.5 (0.12)	0.248
Linear			1.51	13.00	0.12	-0.1 (-0.02)	-0.5 (-0.15)	
Median	EW	Comb 75	-0.48	12.09	-0.04	-2.1 (-0.62)	-2.6 (-0.87)	0.243
Min			-1.20	12.44	-0.10	-2.9 (-0.81)	-3.4 (-1.12)	0.0765
Max			0.19	11.94	0.02	-1.1 (-0.32)	-1.6 (-0.52)	0.409
Linear			-1.51	12.38	-0.12	-0.4 (-0.12)	-0.9 (-0.26)	

Table B.5: Mid Cap Stocks Long-Short PF Summary, Decile Cutoffs, Subperiods: The tables exhibit performance measures for several portfolio schemas. Panel A reflects purely large cap stocks, Panel B only mid cap stocks. It contains value-weighted (VW) and equally-weighted (EW) portfolios, both for single and forecast combination based predictions ("Comb 75", combines 75 single model predictions). All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which are measured against the linear benchmark. The portfolio classification into "Median", "Min" and "Max" portfolios are determined based on the the final return index values. 150 DFN portfolios are considered.

Panel A: **Sample Period 1982-01 until 2003-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	21.25	17.82	1.19	25.2***(6.72)	20.8***(5.10)	0.64
Min			16.37	18.34	0.89	14.6***(3.53)	9.5** (2.30)	-0.20
Max			26.91	18.81	1.43	28.7***(8.26)	23.7***(5.81)	1.39
Linear			18.67	19.46	0.96	21.5***(4.50)	14.5***(3.55)	
Median	EW	single	24.53	16.58	1.48	22.8***(6.55)	17.2***(4.73)	0.79
Min			19.84	18.51	1.07	18.4***(4.37)	12.3***(3.14)	-0.30
Max			30.09	15.83	1.90	28.1***(7.90)	22.7***(5.92)	1.86*
Linear			21.61	18.27	1.18	26.1***(5.44)	19.4***(4.94)	
Median	VW	Comb 75	25.45	19.84	1.28	26.8***(6.53)	20.9***(4.58)	1.00
Min			24.17	19.57	1.24	25.2***(6.35)	19.5***(4.34)	0.89
Max			26.81	20.18	1.33	27.9***(6.65)	21.9***(4.67)	1.16
Linear			18.67	19.46	0.96	21.5***(4.50)	14.5***(3.55)	
Median	EW	Comb 75	27.52	18.46	1.49	26.0***(6.76)	19.1***(5.22)	0.80
Min			26.37	18.18	1.45	24.8***(6.58)	18.1***(4.91)	0.69
Max			27.99	18.07	1.55	26.5***(7.11)	19.8***(5.34)	0.94
Linear			21.61	18.27	1.18	26.1***(5.44)	19.4***(4.94)	

Panel B: **Sample Period 2004-01 until 2014-12:**

Portfolio	Weights	Type	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
Median	VW	single	5.18	14.13	0.37	2.3 (0.62)	1.6 (0.52)	0.92
Min			0.34	14.99	0.02	-1.2 (-0.28)	-1.9 (-0.58)	-0.11
Max			9.56	13.62	0.70	7.8** (2.10)	7.2** (2.11)	1.65*
Linear			1.02	17.50	0.06	6.2* (1.83)	5.5 (1.55)	
Median	EW	single	5.57	14.48	0.38	4.2 (0.98)	3.5 (1.01)	1.02
Min			1.90	15.20	0.12	2.2 (0.58)	1.5 (0.42)	0.207
Max			10.23	12.61	0.81	8.3** (2.31)	7.8** (2.37)	1.99**
Linear			1.18	17.99	0.07	6.5* (1.90)	5.7* (1.70)	
Median	VW	Comb 75	7.75	14.19	0.55	6.3 (1.61)	5.6* (1.73)	1.5
Min			6.89	14.59	0.47	5.6 (1.45)	4.9 (1.53)	1.3
Max			8.70	14.31	0.61	7.4* (1.91)	6.7** (2.03)	1.66*
Linear			1.02	17.50	0.06	6.2* (1.83)	5.5 (1.55)	
Median	EW	Comb 75	7.60	13.58	0.56	6.4 (1.62)	5.7* (1.74)	1.57
Min			6.40	13.82	0.46	5.3 (1.32)	4.6 (1.35)	1.27
Max			8.52	13.61	0.63	7.4* (1.86)	6.7** (1.99)	1.75*
Linear			1.18	17.99	0.07	6.5* (1.90)	5.7* (1.70)	

Freq	Model	Costs	Mean	Std	SR	TO-%	FF-5 α	FF-5 + Mom α	SR Test
1	Comb	No	12.48	18.20	0.69		15.0***(4.81)	11.2***(4.04)	0.0794
		Yes	-12.73	17.93	-0.71	218	-10.2***(-3.31)	-13.8***(-5.12)	3.13***
	Linear	No	9.74	14.72	0.66		4.9* (1.93)	3.3 (1.28)	
		Yes	-26.71	14.81	-1.80	324	-31.0***(-11.54)	-32.6***(-12.31)	
2	Comb	No	12.44	17.92	0.69		14.0***(4.52)	10.3***(3.80)	0.167
		Yes	-1.69	18.08	-0.09	124	0.1 (0.04)	-3.4 (-1.29)	1.52
	Linear	No	9.73	15.31	0.64		3.4 (1.17)	2.9 (1.06)	
		Yes	-8.87	16.00	-0.55	165	-14.6***(-4.95)	-15.0***(-5.37)	
3	Comb	No	10.77	18.79	0.57		14.4***(4.48)	10.8***(3.59)	0.285
		Yes	0.70	19.53	0.04	90	4.4 (1.34)	0.6 (0.21)	1.12
	Linear	No	7.15	15.04	0.48		2.0 (0.72)	2.1 (0.71)	
		Yes	-5.21	15.96	-0.33	111	-10.1***(-3.46)	-10.2***(-3.31)	
4	Comb	No	9.42	18.69	0.50		13.0***(4.34)	9.4***(3.45)	0.506
		Yes	1.10	18.97	0.06	74	4.6 (1.52)	1.2 (0.44)	1.1
	Linear	No	5.03	14.92	0.34		1.5 (0.53)	1.5 (0.53)	
		Yes	-4.36	15.55	-0.28	84	-8.0***(-2.87)	-7.8***(-2.78)	
5	Comb	No	9.33	17.61	0.53		13.8***(4.42)	11.2***(3.69)	0.491
		Yes	2.60	18.27	0.14	61	6.8** (2.11)	4.1 (1.30)	0.825
	Linear	No	5.72	15.43	0.37		4.4 (1.40)	4.5 (1.30)	
		Yes	-1.73	16.13	-0.11	68	-3.3 (-1.01)	-3.4 (-0.95)	
6	Comb	No	7.92	19.80	0.40		11.9***(3.64)	9.0***(2.62)	0.0143
		Yes	2.13	20.16	0.11	52	6.2* (1.85)	3.2 (0.91)	0.428
	Linear	No	5.66	14.34	0.39		1.6 (0.64)	1.9 (0.72)	
		Yes	-0.70	15.02	-0.05	57	-4.6* (-1.76)	-4.4* (-1.67)	

Table B.6: Large Cap Stocks, Decile Cutoffs, VW Weighted: The table shows the performance impact once trading costs are explicitly accounted for. The column "Costs" compares the results with (Yes) and without (No) trading costs. Trading costs are calculated as in Brandt et al. (2009). TO-% indicates the average of the monthly portfolio turnover (the upper bound of a long-short portfolio is 400%). The column "Freq" refers to the rebalancing frequency in months.

Freq	Model	Costs	Mean	Std	SR	TO-%	FF-5 α	FF-5 + Mom α	SR Test
1	Comb	No	20.02	16.67	1.20		16.9***(6.28)	12.9***(4.57)	1.89*
		Yes	-8.73	16.06	-0.54	232	-11.5***(-4.51)	-15.3***(-5.87)	3.3***
	Linear	No	13.43	16.84	0.80		15.6***(6.64)	12.0***(5.18)	
		Yes	-22.53	16.20	-1.39	299	-20.2***(-8.95)	-23.6***(-10.88)	
2	Comb	No	16.35	16.33	1.00		13.9***(5.27)	9.9***(3.73)	1.06
		Yes	1.03	17.01	0.06	127	-1.1 (-0.43)	-5.0** (-1.97)	2.16**
	Linear	No	12.45	15.97	0.78		13.6***(5.60)	10.4***(4.90)	
		Yes	-5.79	16.87	-0.34	153	-4.3* (-1.81)	-7.4***(-3.52)	
3	Comb	No	14.31	16.19	0.88		12.3***(4.49)	8.6***(3.46)	1.65*
		Yes	3.13	17.05	0.18	94	1.3 (0.46)	-2.6 (-1.05)	2.27**
	Linear	No	8.60	16.06	0.54		9.5***(4.35)	7.3***(3.22)	
		Yes	-3.88	17.08	-0.23	107	-2.8 (-1.25)	-5.2** (-2.23)	
4	Comb	No	11.96	16.11	0.74		9.9***(3.94)	6.4***(2.70)	1.08
		Yes	2.82	16.54	0.17	77	0.7 (0.28)	-2.6 (-1.11)	1.55
	Linear	No	7.99	15.41	0.52		9.3***(4.75)	7.3***(3.47)	
		Yes	-2.02	15.97	-0.13	85	-0.8 (-0.38)	-2.6 (-1.21)	
5	Comb	No	12.38	15.55	0.80		9.5***(3.86)	6.4***(2.87)	2.25**
		Yes	4.56	16.21	0.28	67	1.5 (0.60)	-1.8 (-0.79)	2.24**
	Linear	No	5.95	16.45	0.36		8.2***(3.74)	5.7** (2.30)	
		Yes	-2.22	17.20	-0.13	70	-0.2 (-0.09)	-2.9 (-1.11)	
6	Comb	No	8.90	16.06	0.55		7.6***(2.92)	4.3* (1.69)	0.814
		Yes	1.93	16.68	0.12	60	0.7 (0.28)	-2.5 (-0.95)	1.01
	Linear	No	5.70	15.10	0.38		7.1***(3.50)	5.2** (2.45)	
		Yes	-1.42	15.70	-0.09	61	0.1 (0.05)	-1.8 (-0.85)	

Table B.7: Mid Cap Stocks, Decile Cutoffs, VW Weighted: The table shows the performance impact once trading costs are explicitly accounted for. The column "Costs" compares the results with (Yes) and without (No) trading costs. Trading costs are calculated as in Brandt et al. (2009). TO-% indicates the average of the monthly portfolio turnover (the upper bound of a long-short portfolio is 400%). The column "Freq" refers to the rebalancing frequency in months.

Size	Measure	Name	Top	1	2	3	4	5	6	7	8	Bottom
l + m	Δ Mean	DFN - Linear	0.80	1.14	3.60	2.54	0.79	0.81	-0.67	0.35	-1.35	-5.36
	Mean	DFN	19.89	17.57	18.04	16.73	13.98	13.20	12.35	10.29	7.08	1.45
		Linear	19.09	16.43	14.44	14.20	13.19	12.40	13.02	9.94	8.43	6.81
	SR	DFN	0.87	0.92	1.01	1.01	0.94	0.91	0.83	0.66	0.44	0.06
		Linear	1.00	0.90	0.85	0.89	0.83	0.80	0.82	0.62	0.49	0.35
	Std	DFN	22.93	19.08	17.86	16.52	14.82	14.50	14.89	15.55	16.20	22.28
		Linear	19.19	18.23	16.91	15.93	15.84	15.59	15.97	16.11	17.17	19.24
large	Δ Mean	DFN - Linear	1.06	2.22	-0.11	1.68	1.27	-1.16	1.36	0.18	-2.01	-3.36
	Mean	DFN	18.13	17.01	15.05	16.67	15.19	12.80	14.23	10.61	8.76	3.96
		Linear	17.06	14.79	15.16	14.99	13.92	13.96	12.87	10.42	10.76	7.33
	SR	DFN	0.76	0.91	0.85	1.05	1.00	0.87	0.98	0.71	0.54	0.19
		Linear	0.93	0.83	0.89	0.90	0.87	0.91	0.82	0.67	0.63	0.40
	Std	DFN	23.71	18.78	17.75	15.83	15.14	14.72	14.49	14.97	16.08	20.67
		Linear	18.28	17.84	16.96	16.61	15.96	15.32	15.63	15.57	17.09	18.23
mid	Δ Mean	DFN - Linear	3.06	0.88	1.34	3.24	2.36	-0.25	1.87	-0.29	0.50	-7.40
	Mean	DFN	20.72	17.91	16.83	18.49	15.39	12.57	13.15	10.51	8.47	-3.17
		Linear	17.66	17.02	15.50	15.25	13.03	12.83	11.28	10.80	7.97	4.23
	SR	DFN	0.88	0.88	0.90	1.01	0.88	0.69	0.71	0.50	0.36	-0.11
		Linear	1.30	0.98	0.89	0.85	0.71	0.68	0.58	0.53	0.37	0.17
	Std	DFN	23.65	20.28	18.67	18.33	17.59	18.21	18.61	20.85	23.60	29.30
		Linear	13.54	17.28	17.32	18.03	18.31	18.78	19.54	20.54	21.35	24.50

Table B.8: Performance summary decile, VW portfolios. The sample period is 1982-2014: The table shows selected performance characteristics of the decile portfolios. "Top" includes the highest ranked stocks according to the respective prediction, "Bottom" the lowest ranked stocks. The statistics are based on monthly returns but expressed in annualized scale.

Type	l+m	large	mid
Comb 10	17.56	14.19	23.58
Comb 25	17.82	14.16	23.49
Comb 50	18.28	14.18	23.42
Comb 75	**18.30**	**13.59**	**22.65**
Comb 100		14.03	23.53
Comb 150		14.07	23.60

Table B.9: Forecast Combination Sensitivity, Decile Cutoffs, Sample Period 1982-01 until 2014-12: The table shows the sensitivity with respect to the choice parameter of how many prediction results to include. It represents the mean return of the median of 30 forecast combination portfolios. The bold marked row, Comb 75, is in use throughout this work.

ID	Acronym	Name	Description	Reference
1	beta	Beta	Measured based on 3 years (min 52 weeks) weekly excess returns with standard ols ($y = c + \beta x$)	Fama and MacBeth (1973)
2	beta_sq	Beta squared	Simply obtained by squaring the β based on the beta from # 1	Fama and MacBeth (1973)
3	retvol	Volatility	Volatility is measured by the standard deviation of daily returns of the previous months	Ang et al. (2006)
4	maxret	Maximum return	Maximum return is defined over the max of the daily returns in month $t-1$	Bali et al. (2011)
5	idiovol	Idiosyncratic volatility	Calculated based on the residuals of regression in # 1	Ali et al. (2003)
6	mom1m	1-month momentum	Return in month $t-1$	Jegadeesh (1990)
7	mom6m	6-month momentum	Cumulative return over 5 months ending in $t-2$	Jegadeesh and Titman (1993)
8	mom12m	12-month momentum	Cumulative return over 11 months ending in $t-2$	Jegadeesh (1990)
9	mom36m	36-month momentum	Cumulative return over 24 months ending in $t-13$	Bondt and Thaler (1985)
10	mve	Market capitalization (size)	log of (SHROUT$\times$PRC)	Banz (1981)
11	ep	Earnings-to-price	Earnings per share	Basu (1977)
12	dy	Dividends-to-price	Yearly dividends (dvt) divided by market cap at fiscal year	Litzenberger and Ramaswamy (1979)
13	bm	Book-to-market	Book value of equity (ceq) divided by market cap	Rosenberg et al. (1985)
14	lev	Leverage	Total liabilities (lt) divided by market cap	Bhandari (1988)
15	currat	Current ratio	Current assets (act) divided by current liabilities (lct)	Ou and Penman (1989)
16	pchcurrat	Pct change in current ratio	Percentage change in currat from year $t-1$ to t	Ou and Penman (1989)
17	quick	Quick ratio	Current assets (act) minus inventory (invt), divided by current liabilities (lct)	Ou and Penman (1989)
18	pchquick	Pct change in quick ratio	Percentage change in quick from year $t-1$ to t	Ou and Penman (1989)
19	salecash	Sales-to-cash	Annual sales (sale) divided by cash and cash equivalents (che)	Ou and Penman (1989)
20	salerec	Sales-to-receivables	Annual sales (sale) divided by accounts receivable (rect)	Ou and Penman (1989)
21	saleinv	Sales-to-inventory	Annual sales (sale) divided by total invetory (invt)	Ou and Penman (1989)
22	pchsaleinv	Pct change in sales-to-inventory	Percentage change in saleinv from year $t-1$ to t	Ou and Penman (1989)
23	cashdebt	Cashflow-to-debt	Earnings before depreciation and extraordinary items (ib + dp) divided by avg total liabilities (lt)	Ou and Penman (1989)
24	baspread	Illiquidity (bid-ask-spread)	Monthly avg of daily bid-ask spread divided by avg of daily bid-ask spread	Amihud and Mendelson (1989)
25	depr	Depreciation-to-gross PP&E	Depreciation expense (dp) divided by gross PPE (ppegt)	Holthausen and Larcker (1992)
26	pchdepr	Pct change in Depreciation-to-gross PP&E	Percentage change in depr from year $t-1$ to t	Holthausen and Larcker (1992)
27	mve_ia	Industry-adjusted firm size	Log market caps are adjusted by log of the mean of the industry	Asness et al. (2000)
28	cfp_ia	Industry-adjusted cashflow-to-price	Industry adjusted cash flow-to-price ratio equal weighted average	Asness et al. (2000)
29	bm_ia	Industry-adjusted book-to-market	Industry adjusted book-to-market equal weighted average	Asness et al. (2000)
30	sgr	Annual sales growth	Percentage change in sales from year $t-1$ to t	Lakonishok et al. (1994)
31	ipo	IPO	Indicated by 1 if first 12 months available on CRSP monthly file	Loughran and Ritter (1995)
32	divi	Dividend initiation	Indicated by 1 if company pays dividends but did not in prior year.	Michaely et al. (1995)
33	divo	Dividend omission	Indicated by 1 if company does not pay dividends but did in prior year.	Michaely et al. (1995)
34	sp	Sales-to-price	Annual sales (sale) divided by market cap	Barbee Jr et al. (1996)
35	acc	WC accruals	(ib) - (oancf)/(at), if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Sloan (1996)
36	turn	Share turnover	Average monthly trading volume for the three months $t-3$ to $t-1$ divided by SHROUT at $t-1$	Datar et al. (1998)
37	pchsale_pchinvt	Delta pct change sales vs. inventory	Difference of percentage changes in sales (sale) and inventory (invt)	Abarbanell and Bushee (1997)
38	pchsale_pchrect	Delta pct change sales vs. receivables	Difference of percentage changes in sales (sale) and receivables (rect)	Abarbanell and Bushee (1997)
39	pchcapx_ia	CAPEX	Industry adjusted (two digit SIC) fiscal year mean adjusted percentage change in capital expenditures (capx)	Abarbanell and Bushee (1997)
40	pchgm_pchsale	Delta pct gross margin vs. sales	Annual percentage change in gross margin (sale minus cogs) minus percentage change in sales (sale)	Abarbanell and Bushee (1997)

Table B.10: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values. The table is taken from Messmer and Audrino (2017).

ID	Acronym	Name	Description	Reference
41	pchsale_pchxsga	Delta pct sales vs. SGaA	Annual percentage change in sales (sale) minus percentage change in SGaA (xsga)	Abarbanell and Bushee (1997)
42	dolvol	Dollar trading volume	Log of trading volume times price per share from month t-2	Chordia et al. (2001)
43	std_dolvol	Volatility trading volume	Monthly standard deviation of daily trading volume	Chordia et al. (2001)
44	std_turn	Volatility turnover	Monthly standard deviation of daily share turnover	Chordia et al. (2001)
45	chinv	Change in inventory	First difference of inventory (invt) divided by total assets	Thomas and Zhang (2002)
46	pchemp_ia	Industry-adjusted pch in employees	Industry adjusted percentage change in employees	Asness et al. (2000)
47	cfp	Cashflow-to-price	Operating cash flows (oancf) scaled by market capitalization (fiscal year end)	Desai et al. (2004)
48	rd	R&D Increase	If annual increase in R&D expenses (xrd) scaled by total assets (at) >0.05, 1, else 0	Eberhart et al. (2004)
49	lgr	Pct change in long-term debt	Annual percentage change in long term debt (lt)	Richardson et al. (2005)
50	egr	Pct change in book equity	Annual percentage change in book equity (ceq)	Richardson et al. (2005)
51	rd_sale	R&D-to-sales	R&D expenses(xrd) scaled by sales (sale)	Guo et al. (2006)
52	rd_mve	R&D-to-market	R&D expenses(xrd) scaled by market cap	Guo et al. (2006)
53	chg_mom6m	change in mom6m	difference of mom6m measured at t and $t-6$	Gettleman and Marks (2006)
54	hire	Pct change in employee	Annual percentage change in employee (emp)	Belo et al. (2014)
55	agr	Asset growth	Annual percentage change in assets (at)	Cooper et al. (2008)
56	cashpr	Cash productivity	Market cap plus long term debt (dltt) minus assets (at) divided by cash (che)	Chandrashekar and Rao (2009)
57	gma	Gross-profitability	Sales (sale) minus costs of goods sold (cogs) divided by one-year lagged assets(at)	Novy-Marx (2013)
58	cash	Cash-to-assets	Cash (che) divided by assets(at)	Palazzo (2012)
59	pctacc	Accruals-to-income	(ib) minus (oancf) divided by abs ((ib)), when (ib) equals 0, it is set to 0.01, if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Hafzalla et al. (2011)
60	absacc	Absolut accruals	Absolute value of acc	Bandyopadhyay et al. (2010)
61	roic	Return on invested capital	Earnings before interest and taxes (ebit) - non-operating income (nopi), divided by non-cash enterprise value (ceq+lt-che)	Brown and Rowe (2007)
62	grcapex	Pct change in two year CAPX	Percentage change in two year capital expenditure (capx)	Anderson and Garcia (2006)
63	tang	Debt capacity-to-firm-tangability	(Cash (che) + 0.715 receivables (rect) + 0.547 inventory(invt) + 0.535 (ppegt))/ total assets (at)	Hahn and Lee (2009)
64	chshrout	Change in shares-outstanding	Yearly percentage change in outstanding shares (SHROUT)	Pontiff and Woodgate (2008)
65	invest	CAPEX and inventory	Yearly difference in gross property, plant and equipment (ppegt) + diff in (invt) / (t-1) total assets (at)	Chen and Zhang (2010)
66	age	Years since CS coverage	Years since first compustat coverage years(datadate - min(datadate))	Jiang et al. (2005)
67	chpm_ia	Industry-adjusted change in profit margin	Industry adjusted (two-digit SIC) change in profit margin (ib/sale)	Soliman (2008)
68	chato_ia	Industriy-adjusted change in asset turnover	Industry adjusted (two-digit SIC) change in asset turnover (sale/at)	Soliman (2008)

Table B.11: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values.The table is taken from Messmer and Audrino (2017).

B.4 Additional figures

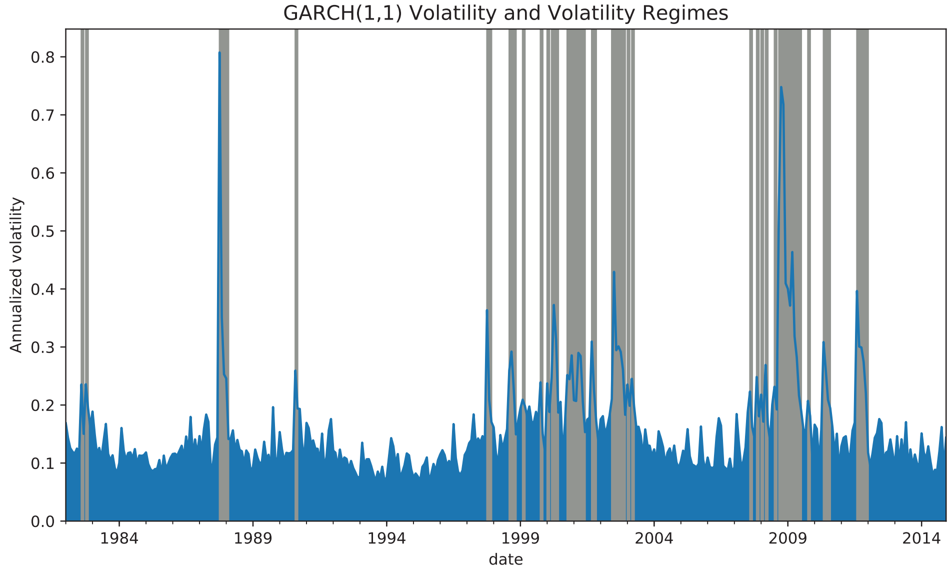

Figure B.2: The graph depicts the estimated GARCH(1,1) process. Precisely, $\sigma_t^2 = \alpha + \beta_0 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, The parameters estimated with normal distributed errors are: $\hat{\alpha} = 0.0079$, $\hat{\beta}_0 = 0.093$ and $\hat{\beta}_1 = 0.9019$. The model is estimated based on daily market factor returns with data from 1962-2014. The figure shows the corresponding month end values from 1982-2014. The area shaded in grey marks the highest 15% of the volatility estimates for the monthly subsample.

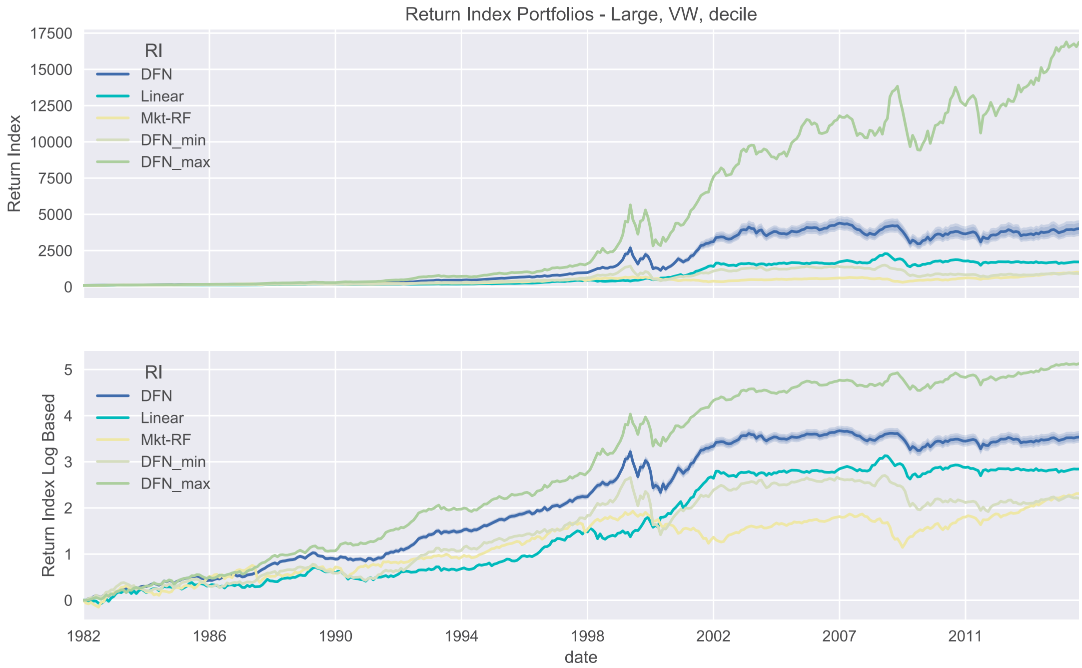

Figure B.3: The figure exhibits selected paths of the 150 portfolios using large cap stocks with value weighted decile long-short portfolios for the years 1982-2014. "DFN" reflects the average of the 150 portfolio paths, which is embedded into the corresponding 1% and 10%-confidence bounds.

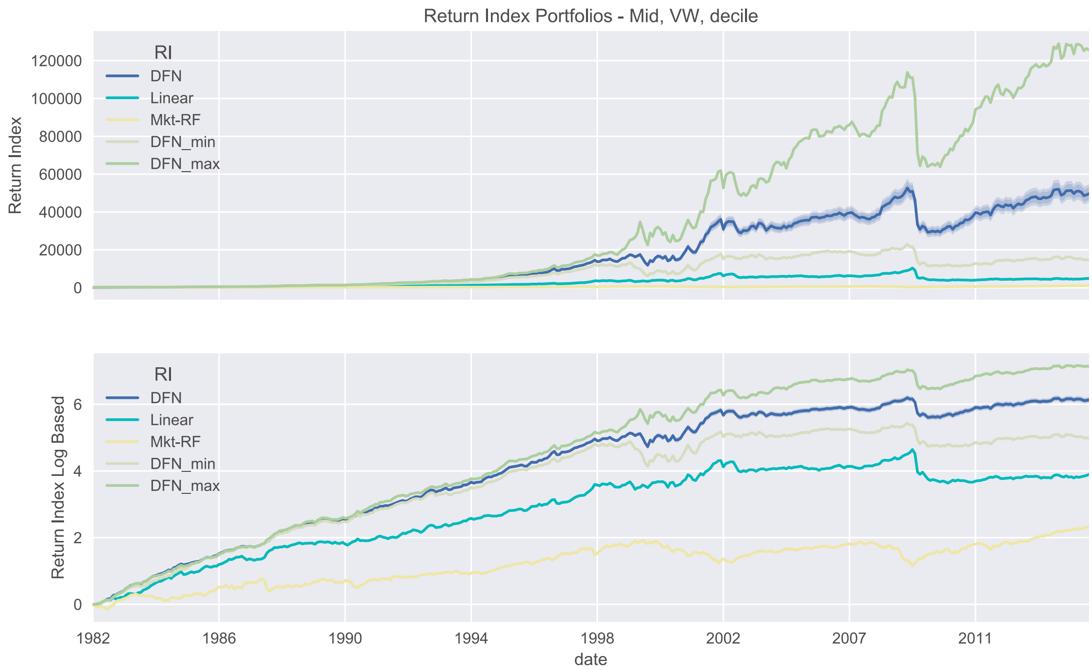

Figure B.4: The figure exhibits the selected paths of 150 estimated portfolios using mid cap stocks for the years 1982-2014, using value weighted portfolios. "DFN" reflects the average of the 150 portfolio paths, which is embedded into the corresponding 1% and 10%-confidence bounds.

Chapter 3

The cross-section of expected returns: Fama-MacBeth vs. pooled OLS

Marcial Messmer

3.1 Introduction

The selection of firm characteristics (FC) which explain deviations in expected cross-sectional stock return spreads is a major task in finance. Academically it is important, mainly because financial theory requires robust empirical results which are unbiased. Moreover, this topic has a high practical relevance as the asset management industry these days funnels billions of dollars into products, which are based on various FC and often promise attractive return premia.[1] The offering advocates and justifies the usage of new "factors" often by results of academic research. This is a concern, especially as Hou et al. (2017) document that out of 400+ published FC, 85% do not survive a (conservative) replication exercise when analyzed univariately.

Typically, these lower-dimensional empirical tests are performed with the help of long-short portfolios (constructed based on FC sorts) or linear regression models. Due to the curse of dimensionality, portfolio sorting becomes infeasible in higher dimensions, and, therefore, leaves the multivariate linear model as the only traditionally used tool left. As a result, Green et al. (2017) perform a multivariate Fama and MacBeth (1973) (FM) regression. Alternatively, Harvey et al. (2016) propose to apply p-value adjustments, which control for multiple testing biases to curb the problem. As a consequence, this study investigates these propositions in more detail. I focus mainly on testing the appropriateness of FM with an in-depth comparison to pooled ordinary-least-squares (POLS) estimation.[2] The multiple testing corrections are solely evaluated from a portfolio perspective.

POLS is a natural benchmark, as the two methods share an identical objective function. The comparison is interesting because both methods have two distinct solutions, which affect both, the coefficient and the corresponding standard error estimates. Consequently, this work examines the resulting differences. The investigation concentrates on three questions. First, how does a FM based FC inference evaluated at the empirical problem at hand alternate when compared to the benchmark? How do these variations impact the performance of a long-short portfolio strategy? Finally and crucially, why do FM and POLS lead to different results?

Moreover, the estimation methodology is not the only choice variable left. The other source of variation, which is of interest for this research, is the option to adjust p-values — a choice between family-wise error ratio (FWER), false discovery rates (FDR) adjustments or no adjustments at all. Naturally, differences in the selection of FC emerge. My objective here is to analyze the impact of these adjustments from an investors perspective. This

[1] These products are often univariate based portfolios, but exist in multi-FC variants or strategies as well. For an introduction into the practical aspects of this topic, I refer the read to Ang (2014).

[2] Throughout this work I refer to POLS as POLS with robust standard errors estimates. These estimates are based on Driscoll and Kraay (1998), as they account for cross-sectional dependence but also are robust to heteroscedasticity and autocorrelation (HAC).

contributes to answer this open question, especially as Harvey et al. (2016, p. 13) state when confronted with selecting among these adjustment procedures: "*This creates difficulty in choosing between FWER and FDR. Given this difficulty, we do not take a stand on the relative appropriateness of these two measures*".

Hence, this study provides a detailed examination of this specific empirical set-up. This work is most closely related to an article by Petersen (2009), who comprehensively revisits the options of standard error estimation in finance panel data sets (including a comparison of POLS and FM). In contrast to his paper, this work focuses mainly on the empirical differences of this specific selection problem. Thus the contribution of this paper is of empirical nature, as it provides guidance on choosing an appropriate linear estimation technique. Apart from the POLS vs. FM discussion, this paper contributes as it studies the performance influence of multiple testing corrections from an investor's perspective.

I document that the multivariate selection of FC explaining variation in expected cross-sectional stock returns (ECSR), varies substantially depending on whether the inference is based on FM or POLS regressions. This finding is remarkable, as both approaches aim to estimate the exact same model. These differences are mainly driven by deviations in standard error (SE) estimates, which are on average 13%, 29% and 30%, respectively, higher for the POLS approach in the three stock size categories under consideration. Hence, the choice of the estimation methodology is an important decision to be made by researchers. Most often the finance literature relies on Fama and MacBeth (1973) regressions, a relevant example of this paper is the work of Green et al. (2017). These differences lead to variations in portfolio returns. However, these variations are of limited magnitude as the formal testing of differences in Sharpe ratio reveals. The analysis shows that POLS is more robust to adjustments in the estimation and prediction procedure. Moreover, a Monte Carlo simulation indicates that SE deviations are a consequence of biased SE estimates of the FM methodology. The bias is most pronounced in small T settings and almost disappears as T grows large.

In addition to the multivariate estimation, the exercise is repeated for the univariate case. The empirical results show that FM and POLS hardly deviate from a selection point of view.

The application of multiple testing adjustments does not lead to improvements in portfolio performance, as a prediction and portfolio sorting exercise in the spirit of McLean and Pontiff (2016) demonstrates. Specifically, the results point in the opposite direction. I find that in many specification investors would have been harmed if applied historically. However, for some instances of FM specifications the adjustment leads to performance improvements. Given the documented possibility of downward biased FM SE estimates, these observed improvements are consistent, as the adjustments manipulate the p-value cutoffs and potentially correct the FM biases in the right direction.

This paper relies on stock data obtained from the CRSP/Compustat database for a period from 1970 until 2014 and purely focuses on the US stock market. I borrow the 68 published FC used in this work from Messmer and Audrino (2017), which are build based on accounting and market data. Hou et al. (2017) document the impact of economically irrelevant small and micro-cap stocks. Thus the results of this paper stem mostly from large and mid samples.

The paper is structured as follows. Section 3.2 discusses the relevant literature. Section 3.3 briefly introduces the FM and POLS estimators (including the standard error estimation) and the multiple testing concepts of the family-wise error rate and false discovery rate. Section 3.4 describes the data. In Section 3.5 the empirical results are presented, including the FC selection and the portfolio perspective. This part is followed by a simulation exercise. In the end, a conclusion is provided.

3.2 Related literature

This paper builds on several strands of literature. First, this paper relates to the literature that studies estimation methodologies in asset pricing and finance. Starting with the work of Fama and MacBeth (1973), which introduces the FM estimator. The estimator is appealing for empirical asset pricing mainly for the following reasons. The FM regression accounts for cross-sectional dependence and time-series correlations can be addressed easily by a Newey and West (1987) type variance correction. Finally, unbalanced panels can straightforwardly be dealt with, which often exposes challenges for pooled regression approaches in particular when the distributional character of the errors deviate from standard cases. In some cases the computational aspect might be relevant, FM requires typically less RAM than POLS (nowadays only a concern with huge panel data sets) and is also trivially parallelized. However, shortcomings have also been documented which are mostly linked to errors-in-variable issues, see Shanken (1992) for more details. A detailed analytical representation of the FM estimator is available in Jagannathan et al. (2010). More broadly, Cochrane (2005) and Goyal (2012) provide a detailed discussions on a variety of methodical topics. For instance, a thorough review of cross-sectional, time-series and FM regressions and their relationship to each other is provided. They also show that under specific assumptions the approaches collapse to the same estimator. A study closely linked to this work is Petersen (2009). It illustrates the relevance of appropriate standard error corrections in settings encountered in asset pricing and corporate finance. He provides an example from cross-sectional asset-pricing, which uses a (small) set of FC to predict cross-sectional returns. This empirical example shows that correcting SE for cross-sectional dependence (he calls it a time-effect) eliminates by far the single most significant source causing a bias when dealing with a cross-section of stock returns. Precisely, they report SE which are 2.0 to 3.4 times larger once cross-sectional correlations are corrected for.

Second, the paper builds on recent studies that investigate expected cross-sectional returns and their relationship to a potentially rich set of FC. A recent study by Hou et al. (2017) find a total of 447 studies or FC published. As a consequence of this abundance of available published FC, which supposedly explain differences in expected cross-sectional return spreads, Cochrane (2011) states in his presidential address the need for new and alternative methodologies to curb the dimensionality of the problem. Following his call, Harvey et al. (2016), McLean and Pontiff (2016) and Green et al. (2017) provide different answers. These three studies are particularly relevant for this work, as I borrow parts of their methodology. Harvey et al. (2016) propose to adopt a multiple testing procedures from other areas of science. Multiple testing in the form of family-wise error rates and false discovery ratios enforces much smaller p-value cutoff points. Thus a much smaller set of relevant FC survives, the majority fails to meet these higher requirements. McLean and Pontiff (2016) follow a true out-of-sample strategy, by controlling for the publication date of the respective FC. They find that many FC are no longer significant. Last, the work of Green et al. (2017) focus on a multivariate linear digestion. They apply univariate and multivariate FM regressions. The selection exercise also includes a false discovery ratio adjustment, limiting the number of active FC even further. Common to all three papers are the findings that many of the published factors are indeed not robust to these alternative testing procedures. Messmer and Audrino (2017) compare the FC selection of POLS vs. the adaptive Lasso (Zou (2006)). A novel cross-sectional simulation and a portfolio perspective reveal advantages of this shrinkage approach over POLS. Another interesting approach is the work by Linnainmaa and Roberts (2017), who take a historical perspective, as they use data as early as 1917 to investigate the relation of FC and cross-sectional returns. Their findings are similar as before, the significance of many FC disappears. Hou et al. (2017) mark so far the most comprehensive empirical review, as their study replicates 447 anomaly characteristics. Moreover, an emerging branch of this literature investigates the non-linear relationship between FC and expected cross-sectional returns. These studies include, for example, Moritz and Zimmermann (2016), Kozak et al. (2017), Freyberger et al. (2017) and Messmer (2017).

More generally, this paper contributes to the vast literature of cross-sectional return studies. Most prominently here, are the seminal contributions by Fama and French (1992, 1993), which introduce the famous three-factor model, itself partly obtained by a FM regression of return spreads on FC. Once established, most of the consecutive anomaly (or factor) studies relied on a portfolio sorting methodology. Well-known examples of these are, momentum (Jegadeesh and Titman (1993), Carhart (1997)), capital investments (Titman et al. (2004), Chen and Zhang (2010)), low-volatility (Ang et al. (2006)), quality (Novy-Marx (2013)) and low-beta (Jensen et al. (1972), Frazzini and Pedersen (2014)).

3.3 Methods

This section describes the notation and the methods. The notation throughout this work is as follows. Let the scalar $y_{n,t}$ denote an excess return of stock n (of a total of N stocks) at point in time t (of a total of T periods). The vector of FC is defined as, $x_{n,t}$. It comprises, for example, the book-to-market ratio (bm), the twelve-months momentum ($mom12m$) or short-term reversal ($mom1m$). The vector $x_{n,t}$ is of length P and contains all FC which are indexed by p. The linear model I aim to estimate is of the following form,

$$y_{n,t+1} = x'_{n,t}\gamma + \epsilon_{n,t+1}. \tag{3.1}$$

Note, that the regression coefficient γ is time-constant. Furthermore, X_t reflects a stacked matrix of vectors of FC for each stock included at point in time t. Equivalently, y_t corresponds to a stacked vector of excess return scalars. Finally, we can concatenate the monthly observations into one pooled representation, $\mathbf{X}$, and y, which are of dimension $TN \times P$ and $TN \times 1$, respectively.

POLS and FM collapse under certain conditions numerically to the same estimator, for more details I refer to Cochrane (2005).

3.3.1 Pooled OLS

Given $\mathbf{X}$ and y, equation 3.1 can be expressed as $y = \mathbf{X}\gamma + \epsilon$ in matrix notation. The POLS estimator reads,

$$\hat{\gamma}_{\text{POLS}} = \arg\min_{\gamma} \left(\|y - \mathbf{X}\gamma\|_2^2 / TN \right), \tag{3.2}$$

which is solved by,

$$\hat{\gamma}_{\text{POLS}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'y), \tag{3.3}$$

where $\hat{\gamma}_{\text{POLS}}$ represents an estimate.

3.3.2 Fama-MacBeth regression

The FM estimator is a two-step procedure.

In the **first step**, cross-sectional coefficients are estimated by OLS at each point in time. Given 3.3, these estimates are,

$$\hat{\lambda}_t = (X'_t X_t)^{-1}(X'_t y_t), \tag{3.4}$$

which yields a time series collection of the vector of estimates, λ_t (of dimension $P \times 1$).

In the **second step**, γ_{FM} is obtained by averaging over the time-series values estimated in 3.4,

$$\hat{\gamma}_{p,\mathrm{FM}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\lambda}_{p,t}. \tag{3.5}$$

3.3.3 Standard error estimates

In case of POLS, I use Driscoll and Kraay (1998) robust standard errors, which are robust to cross-sectional correlations, heteroscedasticity and auto-correlation. The SE estimator reads,

$$\hat{\mathrm{SE}}(\hat{\gamma}_{\mathrm{POLS}}) = \mathrm{diag}\left((\mathbf{X}'\mathbf{X})^{-1}\frac{T}{T-1}\frac{TN}{TN-P}\mathbf{S}_{\mathrm{POLS}}(\mathbf{X}'\mathbf{X})^{-1}\right)^{1/2}, \tag{3.6}$$

with,

$$\mathbf{S}_{\mathrm{POLS}} = \frac{1}{T^2}\sum_{j=-K}^{K}\left(1-\frac{|j|}{K+1}\right)\hat{\mathbf{\Gamma}}_{\mathrm{POLS}}(j),$$

where K denotes the number of Newey and West (1987) (NW) lags, and,

$$\hat{\mathbf{\Gamma}}_{\mathrm{POLS}}(j) = \begin{cases} \sum_{t=j+1}^{T} h_t h'_{t-j} & \text{if} \quad j \geq 0, \\ \sum_{t=-j+1}^{T} h_{t+j} h'_t & \text{if} \quad j < 0, \end{cases}$$

with $h_t = \frac{1}{N}\sum_{n=1}^{N} x_{n,t}\epsilon_{n,t+1}$. Notice that $\mathbf{S}_{\mathrm{POLS}}$ is of dimension $P \times P$, and h_t of $P \times 1$.

The FM SE estimator of the coefficient is as used in Green et al. (2017) and of NW type, i.e.,

$$\hat{\mathrm{SE}}(\hat{\gamma}_{\mathrm{FM}}) = \mathrm{diag}\left(\frac{T}{T-1}\mathbf{S}_{\mathrm{FM}}\right)^{1/2}, \tag{3.7}$$

where,

$$\mathbf{S}_{\mathrm{FM}} = \frac{1}{T^2}\sum_{j=-K}^{K}\left(1-\frac{|j|}{K+1}\right)\hat{\mathbf{\Gamma}}_{\mathrm{FM}}(j),$$

and,

$$\hat{\mathbf{\Gamma}}_{\mathrm{FM}}(j) = \begin{cases} \sum_{t=j+1}^{T} \tilde{\lambda}_t \tilde{\lambda}'_{t-j} & \text{if} \quad j \geq 0, \\ \sum_{t=-j+1}^{T} \tilde{\lambda}_{t+j} \tilde{\lambda}'_t & \text{if} \quad j < 0, \end{cases}$$

and finally, given 3.4 and 3.5, $\tilde{\lambda}_t = \hat{\lambda}_t - \hat{\gamma}_{\mathrm{FM}}$.

3.3.4 Family-wise error rate and false discovery rate

Three different p-value adjustment procedures have been presented by Harvey et al. (2016). More broadly, these multiple tests can be categorized into two different classes. The category of family-wise error rate adjustments or the false discovery rate approaches.

Following the notation in Harvey et al. (2016), **FWER** controls for the probability that at least one type I occurs:

$$\text{FWER} = Pr(N_{0|r} \geq 1), \tag{3.8}$$

where $N_{0|r}$, counts the number of type I errors. As a result, the significance level, α, reflects the probability of making at least one type I error. Two out of the three proposed adjustments, the Bonferroni and Holm adjustments, belong to the class of FWER. On the other hand, the Benjamini, Hochberg, and Yekutieli's (BHY) adjustment controls for the FDR. In contrast to FWER, the **FDR** measures the expected relative frequency of false discoveries among all significant variables, i.e.,

$$\text{FDR} = \begin{cases} \mathbb{E}\left[\frac{N_{0|r}}{R}\right] & \text{if} \quad R > 0, \\ 0 & \text{if} \quad R = 0. \end{cases} \tag{3.9}$$

A detailed introduction to the three different adjustments can be found in Harvey et al. (2016). Generally, FWER's are more stringent than FDR's. Consequently, the inference based on equation 3.8 or 3.9 can lead to substantially different conclusions. The most conservative among the three is the Bonferroni correction, followed by the Holm adjustment, which leaves the BHY as the least restrictive. As noted in Harvey et al. (2016), there is no clear guidance in which situation the FWER is preferable over the FDR or vice versa.[3]

3.4 Data

The empirical part relies on the data used in Messmer and Audrino (2017), which is build on the CRSP/Compustat database. Consequently, I adapt the abbreviations and definitions of FC of their study, which itself follows the work of Green et al. (2017). All abbreviations and definitions are enclosed in the appendix. As a result, the analysis comprises all CRSP stocks (share code 10 and 11) traded either at NYSE, AMEX or NASDAQ. In line with the vast part of the literature, observations missing book values and/or market capitalization are excluded from all samples. In contrast to some studies, no price based filtering is conducted, hence, stocks with low prices are present. However, the difference should not be of importance, since I focus on the 2000 stocks with the highest market capitalization. The data frequency

[3]More precisely, it depends on the number of included variables. For a very large number of included variables a FDR approach is recommended, for small numbers a FWER is considered more appropriate. However, it is not known how this threshold is determined.

is monthly. Moreover, accounting data is aligned with a standard lag of six months of the fiscal year end date. Return based FC, such as the short-term reversal, twelve-months momentum, market beta or maximum return, are used as of the end of the most recent month.[4] Extreme values (except of returns) are controlled for by winsorizing at the 1% and 99% percentile at each point in time.[5] Exceptions are binary FC like *divi*,*divo*, *rd*, and *ipo*, which are not corrected. Next, missing data are handled by inserting the cross-sectional mean at each point. Once these adjustments are performed, cross-sectionally normalized (mean=0, standard deviation=1) FC are formed. As mentioned before, return observations are not transformed except that each return is de-meaned in each period such that the cross-sectional information is preserved before concatenated over time. Pooling the observation over time is then straightforward, by stacking the vector and matrices together. As in Fama and French (1996), stocks qualify as large if their market capitalization ranks among the top 1000, accordingly mid cap stocks have to fall into the range of rank 1001-2000. I focus on the joint set of these two subsamples and the subsamples itself. Hence, unbalanced panel structures are not a concern here, since each of the three datasets is balanced through time.

3.5 Empirical study

This section includes the empirical analysis of the two different estimation and inference strategies. The first section presents the differences in the FC selection between the two methods in case of a univariate and a multivariate regression. The second part investigates if these differences have an economically meaningful impact from an investors perspective. This objective is achieved by inspecting differences in long-short portfolio returns.

An important choice for classifying FC as active or inactive is the significance level. Throughout this work, I stick to literature standard cutoff level of 5%. In case of multiple testing, the level is held constant at the same value.

3.5.1 Differences in the FC selection

The univariate case

In univariate case defines single regressions of cross-sectional excess-returns on each FC in isolation, i.e.,

$$y_{n,t+1} = x_{p,n,t}\gamma_p + \epsilon_{p,n,t+1},$$

[4]For example, for the return prediction from 6/30 to 7/31, the max daily return for the period 5/31-6/30 is used.

[5]As a robustness check, I consider a 5%/95% winsorizing-procedure as well.

for each p separately. I estimate this relationship for both, POLS and FM, estimators as defined in Equations 3.2-3.5, the standard errors are calculated according to Equations 3.6 and 3.7, with a NW lag of 12 months (as in Green et al. (2017)).

The differences in case of the univariate regressions are shown in Tables 3.1 and 3.2. Table 3.1 indicates that the standard errors (SE) of POLS on average about 4%-8% lower when each FC is considered in isolation. The sum of the absolute value of the coefficient estimates varies around 2%-3%. As a result, the t-value deviations are not large either. Furthermore, the FC selection results coincide for the most part. For example, the joint sample of large and mid cap stocks match in 25 out 26/27 (FM/POLS) cases. Similar are the results of the mid cap based selection, which reveal an overlap in all 31 cases. Only in case of large cap stocks, deviations are more pronounced. Table 3.1 provides more details on the specific FC selected. The single selected FC include around half of the FC used in this study. The sets selected by both consists of a mix of pure price-based FC, for example, *mom1m* or *mom12m*, but it also comprises FC constructed based on balance sheet information, for instance, *bm_ia* or *invest*.

	large+mid	large	mid
SE Ratio Mean	0.94	0.92	0.97
SE Ratio Median	0.94	0.91	0.96
Abs Coef Ratio	0.97	0.98	0.97
t-value Ratio	1.03	1.06	1.01
FC POLS|FM	27|26	23|20	31|31
JOINT|POLS|FM	25|2|1	19|4|1	31|0|0

Table 3.1: Meta statistics FC selection, univariate regressions, Sample Period 1970-01 until 2014-12. This table shows a meta comparison of the FC selection by POLS and FM. The ratio rows always express the POLS over the FM quantities. The row "SE Ratio Mean" measures the average SE Ratio of all FC, "SE Ratio Median" accordingly the median. "Abs Coef Ratio" exhibits the ratio of the sum of all absolute coef sizes. Accordingly, the "t-value Ratio" expresses the ratio of the sum of absolute t-values. "FC POLS|FM" counts the number of FC selected by each method. "JOINT|POLS|FM" indicates the selection commonality, by counting the number of jointly selected FC, the FC only selected by POLS, and last, the FC solely chosen by FM.

The multivariate case

Analogously, the multivariate case can be defined as follows,

$$y_{n,t+1} = x'_{n,t}\gamma + \epsilon_{n,t+1},$$

where $x_{n,t}$ includes all P FC of interest. The results of the comparison between the two estimation techniques are documented in Tables 3.3 and 3.4. Strikingly, the SE's are on average 16% to 33% higher when estimated with robust POLS standard errors, as you can see in Table 3.3. On the other hand, the absolute size of the coefficients is inflated between 8%-19% compared to the FM estimates. This leads to a total t-value deflation of 9% to 16%.

	JOINT FC	POLS only	FM only
large+mid	bm_ia, sgr, chato_ia, pchsaleinv, chinv, mom36m, hire, mom12m, lgr, retvol, grcapex, bm, acc, sp, pchcurrat, rd_mve, maxret, invest, mom1m, pchgm_pchsale, cashpr, agr, pchsale_pchinvt, egr, pchcapx_ia	dolvol, ep	pchsale_pchrect
large	mom12m, lgr, sp, bm_ia, chg_mom6m, pchcurrat, rd_mve, sgr, chato_ia, agr, pchsale_pchinvt, pchsaleinv, egr, chinv, grcapex, invest, bm, hire, mom1m	divi, cashpr, mve_ia, ep	pchsale_pchrect
mid	bm_ia, sgr, chato_ia, pchsaleinv, chinv, mom36m, hire, turn, mom12m, lgr, rd, retvol, grcapex, bm, acc, sp, pctacc, pchcurrat, maxret, invest, mve, age, mom1m, pchgm_pchsale, mom6m, cfp, cashpr, agr, pchsale_pchinvt, egr, pchcapx_ia		

Table 3.2: FC selection, univariate regressions, Sample Period 1970-01 until 2014-12. This table gives an overview of the selected FC by POLS and FM. The column "JOINT FC" exhibits the FC selected by both POLS and FM. Accordingly, "POLS only" ("FM only") presents the FC included solely in the POLS (FM) selection procedure.

Consequently, POLS selects slighter fewer FC for all three samples, as these averages are not driven by a few outliers. Overall roughly one sixth to a fourth of the FC survive a multivariate-based selection. More importantly, the selected sets of FC vary now substantially. In case of the large and mid cap sample, only 5 out of 14 (11) active POLS (FM) FC are jointly significant. It marginally changes once only large cap stocks are considered, the ratios are $\frac{1}{2}$ and $\frac{5}{11}$, respectively. However, the overlap is higher conditioned only on mid caps. Table 3.4 illustrates the selected FC in detail. Across samples and methods, *mom1m* and *retvol* are the only FC always included. Consistently selected by POLS across samples is only *currat*. FM singles out *mom12m*.

As expected, the multivariate regressions eliminate more FC than the univariate based alternative, in the presence of a significant multi-collinearity among FC. Moreover, two interesting differences can be observed. First, the relative standard errors behave quite differently, they POLS SE are much higher in the multivariate setting. Second, the difference in the selection variation is extreme. Given these distinct selection results, one can expect differences in expected return spreads.

3.5.2 Return predictions and portfolio construction

The previous section has revealed a stark contrast between POLS and FM based FC selection. This section studies the economic implications from an investors perspective. Precisely, I look at performance characteristics of long-short portfolios, which are formed based on return

	large+mid	large	mid
SE Ratio Mean	1.33	1.31	1.16
SE Ratio Median	1.25	1.24	1.11
Abs Coef Ratio	1.19	1.16	1.08
t-value Ratio	0.89	0.84	0.91
FC POLS|FM	14|11	10|11	13|15
JOINT|POLS|FM	5|9|6	5|5|6	10|3|5

Table 3.3: Meta statistics FC selection, multivariate regressions, Sample Period 1970-01 until 2014-12. This table shows a meta comparison of the FC selection by POLS and FM. The ratio rows always express the POLS over the FM quantities. The row "SE Ratio Mean" measures the average SE Ratio of all FC, "SE Ratio Median" accordingly the median. "Abs Coef Ratio" exhibits the ratio of the sum of all absolute coef sizes. Accordingly, the "t-value Ratio" expresses the ratio of the sum of absolute t-values. "FC POLS|FM" counts the number of FC selected by each method. "JOINT|POLS|FM" indicates the selection commonality, by counting the number of jointly selected FC, the FC only selected by POLS, and last, the FC solely chosen by FM.

	JOINT FC	POLS only	FM only
large+mid	agr, mom1m, rd_mve, retvol, cashdebt	mom6m, acc, ep, egr, beta, rd_sale, chg_mom6m, currat, gma	depr, turn, mom12m, idiovol, dy, maxret
large	rd_sale, mom1m, chg_mom6m, ep, retvol	mom6m, agr, currat, egr, bm	pchquick, acc, mom12m, divi, pchcurrat, maxret
mid	cashpr, depr, rd_mve, mom1m, idiovol, cashdebt, beta, retvol, agr, beta_sq	chg_mom6m, currat, pchgm_pchsale	mve_ia, tang, dy, mom12m, turn

Table 3.4: FC selection, multivariate regressions, Sample Period 1970-01 until 2014-12. This table gives an overview of the selected FC by POLS and FM. The column "JOINT FC" exhibits the FC selected by both POLS and FM. Accordingly, "POLS only" ("FM only") presents the FC included solely in the POLS (FM) selection procedure.

predictions of the respective methods.

The procedure is straightforward. The start of the sample period is 1970. Moreover, I require at least 12 years of monthly data to obtain the first prediction. Hence, the first return predictions are available as of January 1982. I apply an expanding window strategy, with an estimation and prediction frequency of one month. Consequently, the FC selection is repeated every month. The prediction is a result of a two-stage procedure. Stage one defines a selection regression with the purpose of eliminating insignificant coefficients. At stage two, the coefficients are estimated based on a second regression including only the set of selected FC from the previous stage. I also look at the case with only regression, in which it directly uses the significant coefficients at the prediction stage. As in Messmer and Audrino (2017), I differentiate two cases with respect to the available FC. Once, I include the entire set of FC in the selection regressions from the beginning of the sample period in 1982. Alternatively, I control for a potential forward-looking bias by allowing only FC to enter the set of available predictors, which are published as of the prediction date. This is motivated by the empirical findings of McLean and Pontiff (2016), who find important differences in significance once the publication date is controlled for. This is particularly crucial, as any look-ahead biased out-of-sample evaluation would favor a downward biased SE estimator.

Given the selection results presented in Section 3.5.1, this possibility can not be ruled out here. Figure 3.1 shows the number of available FC at each point in time, you can see that the number FC rises gradually over time. Hence, the information used for the estimation and prediction is markedly different at each point in time. Additionally, I analyze the portfolio performance impact of multiple testing adjustments. Thus using forward-looking bias-free information sets is essential for a fair comparison, as higher significant levels would impose a disadvantage by construction.

In Table 3.5 we can see the differences between FM and POLS for various specifications for the joint sample of large and mid cap stocks. First, I point to the differences correcting for a potential publication bias or not. The differences are visible when the following rows are compared: (1) and (5), (2) and (6), (3) and (7) and (4) and (8). In all four cases, we see that the mean is lower if we control for a look-ahead bias. These differences are highest when FM regressions are employed. Note, that these results are not sufficient to conclude that the missing FC suffer from a publication bias.[6] Furthermore, I document that the second stage regression, improves the long-short portfolio performance, as you can see by comparing, for example, row (1) with (3). Once more the differences are largest in case of FM. The sensitivity with respect to multiple testing adjustments are negative. The performance of FM and POLS worsens. For example, in case of the BY adjustment of FM the portfolio performance comes down by around 1.66% p.a., (1) vs. (9). The average POLS return declines from 10.21% to 6.18%. Overall the economic differences between the two estimators are rather small, as none of the Share ratio tests indicate significant differences. Only the factor α's tend to be higher for the POLS based approach. Multiple testing leads to slight advantages of FM, unadjusted predictions yield a small edge for a POLS approach.

Table 3.6 Panel A shows the identical specifications, but is conditioned on large cap stocks only. Once more it documents differences between a publication bias-free selection and the potentially contaminated procedure. The results are in line with the previous findings that constraining FC leads to performance declines. Performance deviations between the FM and POLS portfolios are visible, but like in Table 3.5, Table 3.6 Panel A shows that no significant SR differences are visible. While multiple testing does not affect the portfolios by much, as Sharpe ratios are only slighly reduced for both the FM and the POLS regression based portfolios.

Finally, Table 3.6 Panel B displays the performance characteristics for the mid cap sample. The impact of restricting the FC is clearly visible for mid cap stocks, as documented by a substantial decline in average returns, compare, for instance, row (1) and (5). Applying a second stage regression elevates the means for both portfolios. Overall, the POLS reaches

[6]Some FC might contribute to the prediction in the pre- and post-publication period, but are excluded during the pre-publication times. Hence, this only ensures I exclude FC which predictive power is limited to the pre-publication sample.

higher SR's in all seven specifications, but these differences have no statistically significant meaning. Adjusting p-values harms the performance in all instances considered — compare, row (1) with (9), (11) or (13) and row (2) with (10), (12) or (14).

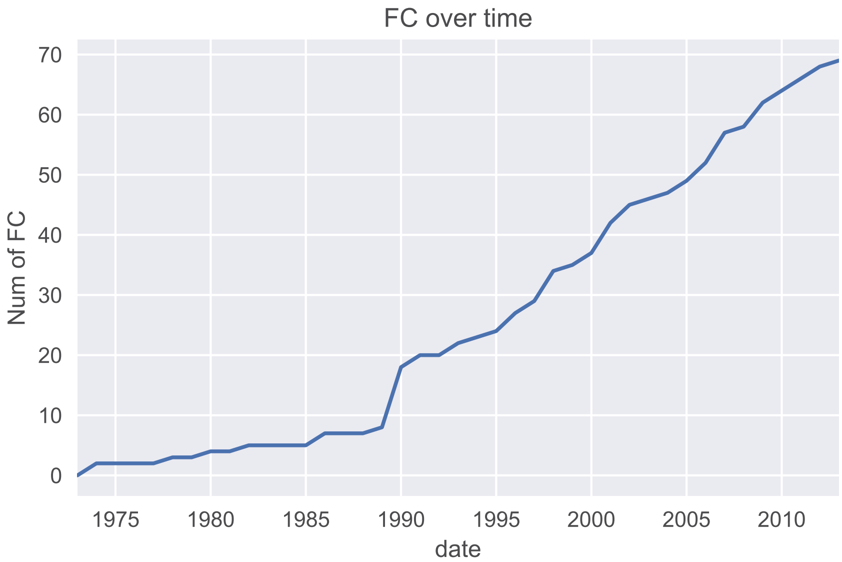

Figure 3.1: The graph depicts the number of published FC as a function of time. The figure is taken from Messmer and Audrino (2017).

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	9.01	18.92	0.48	10.6***(2.97)	6.1** (2.14)	-0.44
2	POLS				10.21	19.19	0.53	11.1***(2.98)	6.4** (2.15)	
3	FM	Yes	-	No	6.75	19.26	0.35	8.0** (2.13)	4.4 (1.32)	-0.55
4	POLS				8.87	20.07	0.44	8.7** (2.18)	4.8 (1.31)	
5	FM	No	-	Yes	12.87	18.84	0.68	10.7***(3.44)	6.2** (2.34)	-0.35
6	POLS				12.90	17.13	0.75	12.4***(3.86)	9.5***(2.83)	
7	FM	No	-	No	10.24	19.02	0.54	6.0** (2.08)	3.9 (1.31)	-0.80
8	POLS				11.38	16.01	0.71	6.5** (2.19)	6.2* (1.91)	
9	FM	Yes	BY	Yes	7.35	20.34	0.36	9.7***(2.80)	4.7* (1.67)	0.27
10	POLS				6.18	19.19	0.32	10.5***(2.91)	6.2** (2.09)	
11	FM	Yes	Holm	Yes	7.17	19.51	0.37	9.2***(2.72)	4.5 (1.60)	-0.01
12	POLS				7.10	19.24	0.37	11.3***(3.17)	6.9** (2.39)	
13	FM	Yes	Bonf	Yes	8.77	19.79	0.44	10.7***(3.13)	5.9** (2.06)	0.72
14	POLS				6.73	19.24	0.35	10.8***(3.03)	6.3** (2.21)	

Table 3.5: Large + mid Cap Stocks, Decile Cutoffs, VW weighted, Sample Period 1982-01 until 2014-12: The table exhibits performance measures for FM and POLS regressions. It contains value-weighted (VW) long short portfolios. The column "OOS FC" indicates if the set of FC included at each point in time reflects only post-publication FC. Column "Adj" shows if p-values are adjusted for multiple testing, if empty no adjustment is performed — "Bonf" refers to Bonferroni, "BY" to Benjamini/Yekutieli. Column "2nd Stage" indicates if two regressions are run or not — first, a selection regression and second, a coefficient estimation with the subset of selected FC. All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which tests the FM vs. the POLS estimation seperately for each specification. The input data are winsorized at the 0.01 level and standardized w.r.t to the cross-section.

Table 3.6: Decile Cutoffs, VW weighted, Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for FM and POLS regression portfolios. Panel A shows the results for large cap stocks, Panel B the results for mid caps. Each panel contains value-weighted (VW) long short portfolios. The column "OOS FC" indicates if the set of FC included at each point in time reflects only post-publication FC. Column "Adj" shows if p-values are adjusted for multiple testing, if empty no adjustment is performed — "Bonf" refers to Bonferroni, "BY" to Benjamini/Yekutieli. Column "2nd Stage" indicates if two regressions are run or not — first, a selection regression and second, a coefficient estimation with the subset of selected FC. All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which tests the FM vs. the POLS estimation seperately for each specification. The input data are winsorized at the 0.01 level and standardized w.r.t to the cross-section.

Panel A: **Large Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	9.73	18.73	0.52	11.7***(3.44)	7.0***(2.75)	-0.04
2	POLS				10.06	19.12	0.53	14.8***(4.09)	10.5***(3.45)	
3	FM	Yes	-	No	7.93	17.75	0.45	11.6***(3.68)	8.5***(2.82)	-0.08
4	POLS				9.11	19.74	0.46	14.3***(4.06)	10.8***(3.24)	
5	FM	No	-	Yes	10.25	17.80	0.58	9.5***(3.13)	5.3** (2.19)	0.11
6	POLS				8.73	15.93	0.55	7.2** (2.44)	6.9** (2.42)	
7	FM	No	-	No	8.81	16.71	0.53	8.8***(3.26)	6.6** (2.45)	0.15
8	POLS				9.14	18.59	0.49	8.6** (2.57)	7.9** (2.37)	
9	FM	Yes	BY	Yes	8.61	19.05	0.45	12.0***(3.44)	8.1***(2.78)	-0.05
10	POLS				9.48	20.51	0.46	12.1** (2.55)	7.2* (1.78)	
11	FM	Yes	Holm	Yes	8.37	19.19	0.44	11.7***(3.30)	7.6***(2.67)	-0.33
12	POLS				9.68	19.62	0.49	14.4***(3.56)	9.8***(3.00)	
13	FM	Yes	Bonf	Yes	8.79	19.30	0.46	12.3***(3.38)	8.1***(2.85)	-0.29
14	POLS				9.83	19.44	0.51	14.5***(3.61)	9.9***(3.07)	

Panel B: **Mid Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	10.45	16.35	0.64	9.9***(3.38)	5.7** (2.27)	-1.31
2	POLS				12.82	15.98	0.80	12.6***(4.16)	8.5***(3.36)	
3	FM	Yes	-	No	9.57	15.22	0.63	8.8***(3.28)	5.5** (2.17)	-0.33
4	POLS				10.94	16.10	0.68	11.4***(3.58)	7.2***(2.71)	
5	FM	No	-	Yes	18.20	15.93	1.14	14.9***(5.37)	10.6***(4.36)	-0.17
6	POLS				17.00	14.54	1.17	14.0***(5.17)	10.8***(4.37)	
7	FM	No	-	No	16.24	15.50	1.05	11.1***(4.59)	8.4***(3.11)	-0.27
8	POLS				15.53	14.24	1.09	10.0***(4.30)	7.7***(3.16)	
9	FM	Yes	BY	Yes	9.20	16.18	0.57	9.2***(2.98)	5.5** (2.05)	-0.82
10	POLS				10.27	15.88	0.65	11.0***(3.50)	6.9***(2.66)	
11	FM	Yes	Holm	Yes	9.16	16.07	0.57	9.3***(3.02)	5.6** (2.12)	-0.97
12	POLS				10.63	16.05	0.66	11.0***(3.43)	6.9***(2.61)	
13	FM	Yes	Bonf	Yes	8.63	16.23	0.53	9.0***(2.92)	5.3** (2.00)	-0.81
14	POLS				9.75	16.08	0.61	10.3***(3.26)	6.3** (2.43)	

3.6 Simulation

The empirical analysis has revealed that FM and POLS regressions lead to substantially and meaningful deviations in the selection of FC. However, it is unclear why we observe these differences from an econometric perspective. Hence, this section investigates potential sources based on a Monte Carlo simulation.

Precisely, I follow the simulation specification introduced in Messmer and Audrino (2017). Their framework is flexible in many dimensions. The data-generating process (DGP) comprises, for example, heteroscedasdicity, cross-sectional dependence, and also cases of autocorrelation. Importantly, it allows the modeling of a multivariate set-up, which is empirically the case in which the inference of the two linear estimators leads to substantially different conclusions. Moreover, the DGP assumes a relative low signal-to-noise ratio, as typically observed in stock markets. For a detailed description of the simulation set-up, I refer the reader to Messmer and Audrino (2017).

I adopt the same calibration, but I focus purely on FM and POLS based inference and study different cases of interest. In addition to the cases 1 and 5 in Messmer and Audrino (2017), I inspect several more cases.[7]

I add cases in which T, the number of months in the panel, is set to 120 (10 years) and 60 (5 years). This is motivated by its empirical relevance. For example, Green et al. (2017) run a sub-sample analysis with relatively few monthly observations (132 months). Moreover, the portfolio prediction part of this work relies initially on estimates which are based on a relatively small number of months. However, I note that T always relates to the underlying variances of the DGP, hence, this specification is highly stylized and does not allow for a general definition of when T should be considered small or large.

Moreover, FM and POLS both come with explicit standard error estimates. Thus it is easy to calculate and report the relative size of the coefficient and standard error estimate biases.

Table 3.7 presents the simulation results. Generally, the columns (0)-(5) reflect FC which are associated with positive coefficient estimates. Columns (6)-(15), the mean, std, min and max column reflect FC which coefficient is zero. Hence, (0)-(5) relate to type II errors — non-rejection of the null hypothesis given it is truly false. The remaining columns are of type I — false positives. *Case 1* refers to the default case of Messmer and Audrino (2017) and it reflects 600 time periods or 50 years of monthly data. As the first block of rows shows, the differences between OLS and FM are small. Though, the mean type I error for FC 16-99 is with 5.66% above the 5.00%, which corresponds to the expected level of significance. This is a result of slightly biased SE estimates, which are on average 1.43% below the estimated true value. As you can see from *case 2* this bias vanishes almost completed once I increase T, in this case to 350 years or 4200 time points. *Case 3* now reduces the number of years compared to *case 1* to 10 years. As you can see from column "mean" in case 3, the FM type I inference mismatches the targeted significance level by 3.28%. This mismatching can be

[7]Cases 2 and 3, which vary assumptions about the volatility process of the errors do not impact the inference in any measurable way (the results are not reported).

traced back to a SE estimate bias, which underestimates the magnitude by around 6.5%. This bias worsens when the number of years is further reduced, as *case 4* depicts. Here I set the number of years to 5, which is 1/10 of the original panel's time length. As you can see, the bias becomes more severe, 13% below the actual variation measured. The cases so far included the correlation structure of the FC imposed by Messmer and Audrino (2017). In *Case 5* these correlations are entirely eliminated, FC are perfectly orthogonal to each other. It illustrates that the correlation among FC does not affect the estimation of FM or OLS in a measurable way, as the results are almost equivalent to *case 4*.

#	Method	Measure	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	mean	std	min	max
1	OLS	Err	32.45			1.00		0.05	4.05	4.25	5.10	5.20	5.20	4.75	4.85	4.85	5.40	4.30	5.00	0.44	3.70	6.05
	FM		31.10			1.05		0.05	4.40	5.75	5.40	5.25	6.30	5.40	5.80	5.20	6.20	5.55	5.66	0.49	4.55	6.90
	OLS	Rel Coef	-0.45	0.09	0.05	0.05	-0.28	0.45														
	FM		-0.42	0.13	-0.17	-0.04	-0.28	0.44														
	OLS	Rel SE	1.53	1.60	0.83	1.85	-1.04	-2.45	3.03	0.41	-0.06	0.47	-2.65	-0.58	0.00	-0.53	-1.75	1.47	-0.49	1.58	-5.12	3.29
	FM		-0.18	0.57	-0.23	0.05	-2.35	-3.85	2.01	-2.03	-1.65	-0.41	-3.47	-1.92	-2.47	-0.24	-2.70	-2.05	-1.43	1.61	-5.62	2.88
2	OLS	Err							4.70	5.35	4.55	4.60	5.35	4.95	5.75	5.50	5.45	5.25	5.02	0.44	3.70	6.20
	FM								4.85	5.60	5.20	3.95	5.00	5.30	5.95	4.90	5.60	4.60	5.11	0.51	3.90	6.10
	OLS	Rel Coef	0.16	-0.01	0.09	0.01	-0.27	-0.05														
	FM		0.18	-0.10	0.07	-0.01	-0.23	-0.09														
	OLS	Rel SE	-1.21	3.69	3.04	0.87	-1.33	0.07	0.79	-1.06	1.53	2.40	-1.38	0.32	-1.15	-1.24	-1.26	-1.11	-0.27	1.60	-3.88	4.86
	FM		-1.40	4.29	0.51	1.45	-2.03	-0.15	0.90	-0.65	1.32	2.66	-1.56	0.40	0.09	-0.82	-0.64	0.82	-0.29	1.50	-3.63	3.97
3	OLS	Err	85.10	9.68	28.24	53.90	8.94	24.14	5.36	5.66	5.40	6.08	5.64	5.52	5.48	5.54	5.14	5.52	5.01	0.30	4.40	5.66
	FM		80.60	6.38	21.58	48.34	7.88	21.20	8.12	8.36	8.02	8.94	8.56	8.08	9.18	8.26	8.10	7.88	8.28	0.39	7.28	9.28
	OLS	Rel Coef	-3.02	0.13	0.78	-0.28	-0.33	1.16														
	FM		-3.01	0.23	0.62	-0.29	-0.32	1.08														
	OLS	Rel SE	-1.07	-0.71	-0.09	-0.74	-0.62	-2.40	-0.58	-1.56	-1.16	-2.96	-2.98	-1.05	-3.83	-2.75	-3.07	-5.22	-2.93	0.89	-4.70	-0.70
	FM		-8.27	-6.13	-5.88	-6.74	-6.53	-8.10	-6.46	-7.23	-6.59	-8.53	-8.14	-6.81	-7.00	-6.76	-6.02	-6.27	-6.54	0.91	-8.32	-3.86
4	OLS	Err	86.44	32.56	54.72	72.12	32.76	52.26	6.14	5.94	5.94	6.04	6.24	6.14	5.52	6.08	5.48	4.86	5.84	0.35	4.98	6.52
	FM		78.28	24.74	44.02	63.58	25.96	43.38	11.18	12.42	11.86	12.40	12.20	12.08	12.28	11.70	11.76	12.24	11.73	0.47	10.72	12.92
	OLS	Rel Coef	-2.69	0.38	0.60	0.42	0.58	0.81														
	FM		-2.63	0.39	0.46	0.48	0.50	0.69														
	OLS	Rel SE	-4.14	-1.44	-1.05	-0.12	0.84	-3.27	-0.71	-2.30	-2.26	-2.40	-3.06	-1.47	-3.91	-5.24	-2.80	-1.42	-3.74	1.11	-6.71	-0.82
	FM		-15.52	-12.43	-12.20	-11.61	-11.18	-14.56	-11.71	-14.01	-13.34	-13.72	-14.03	-13.43	-13.27	-13.49	-12.45	-13.10	-12.75	0.99	-15.21	-10.54
5	OLS	Err	86.40	30.42	53.28	72.26	32.60	52.36	5.54	6.22	6.10	6.24	6.26	5.96	5.40	6.22	5.80	5.68	5.76	0.33	5.16	6.76
	FM		78.24	24.90	44.00	63.58	25.88	43.34	11.12	12.32	11.70	12.28	12.32	11.94	11.48	11.50	11.44	11.14	11.75	0.50	10.70	13.02
	OLS	Rel Coef	-2.65	0.37	0.50	0.49	0.55	0.56														
	FM		-2.64	0.37	0.47	0.44	0.49	0.72														
	OLS	Rel SE	-4.08	-0.74	-0.82	-0.38	0.55	-3.03	-0.22	-2.44	-2.22	-2.57	-3.04	-1.93	-1.32	-4.43	-5.16	-4.59	-3.53	1.00	-5.83	-1.11
	FM		-15.52	-12.49	-12.19	-11.68	-11.20	-14.53	-11.69	-13.98	-13.30	-13.75	-14.03	-13.41	-12.52	-12.40	-13.17	-12.80	-12.76	0.92	-15.26	-10.56

Table 3.7: The Table shows the simulation results. Rows market with "Err" contain type II (columns 0-5) and type I error (6-11, 12-15,) ratio behavior in percentage points. The columns "mean", "std", "min", and "max" summarize the results of FC 16-99, which are errors of type I. The rows "Rel Coef" display the relative deviations of the coefficient estimates, for cases in which the FC truly contains predictive information. The relative size of the standard error biases are visible in the "Rel SE" rows. All details of the simulation specification can be found in Messmer and Audrino (2017). FC 6-11 are linked to the risk dimension of the returns. FC cases 12-15 are interesting in so far, as these are FC with higher correlations to other FC. In case 1 T=600, case 2 T=4200, case 3 T=120, case 4 T=60 in case 5 T=60. Case 5 removes the correlation among FC completed. The number of simulations is set to 2000 for case 1 and 2, and equals 5000 for cases 3-5.

3.7 Conclusion

This paper documents major empirical differences between FM and POLS based FC selection in the US cross-section of stock returns. I find simulation-based evidence that FM standard errors are biased downward and explain, at least partially, the apparent contrasts in the sets of selected FC. I do not provide rigorous theoretical insights of the estimators, which would help to understand these discrepancies in more depth from another important angle. However, the simulation allows to replicate downward biased SE estimates in a scenario in which T is relatively small, the bias vanishes almost completely as T grows large. Moreover, it would be insightful to analyze a richer or different set of FC or repeat the selection procedure with a non-US cross-section. All of these aspects are open for future studies to explore.

Another key finding of this paper is that multiple testing adjustments, for instance, controlling for a family-wise error, does not yield any additional return benefits for investors. The portfolio performance results provide suggestive evidence rather in the opposite direction, as the majority of specifications realize a slight decline in average returns.

Appendix C

C.1 Appendix

C.1.1 Additional tables

	large+mid	large	mid
SE Ratio Mean	1.39	1.35	1.21
SE Ratio Median	1.28	1.26	1.16
Abs Coef Ratio	1.16	1.19	1.06
t-value Ratio	0.83	0.82	0.89
FC POLS|FM	17|13	7|12	15|14
JOINT|POLS|FM	9|8|4	5|2|7	12|3|2

Table C.1: Meta statistics FC selection, multivariate regressions, Sample Period 1970-01 until 2003-12. This table shows a meta comparison of the FC selection by POLS and FM. The ratio rows always express the POLS over the FM quantities. The row "SE Ratio Mean" measures the average SE Ratio of all FC, "SE Ratio Median" accordingly the median. "Abs Coef Ratio" exhibits the ratio of the sum of all absolute coef sizes. Accordingly, the "t-value Ratio" expresses the ratio of the sum of absolute t-values. "FC POLS|FM" counts the number of FC selected by each method. "JOINT|POLS|FM" indicates the selection commonality, by counting the number of jointly selected FC, the FC only selected by POLS, and last, the FC solely chosen by FM.

	JOINT FC	POLS only	FM only
large+mid	agr, depr, idiovol, cashdebt, retvol, mom1m, beta, rd_mve, chg_mom6m	mom6m, quick, currat, bm, lgr, ep, gma, cashpr	mom12m, maxret, chato_ia, beta_sq
large	chg_mom6m, ep, retvol, mom1m, rd_sale	agr, currat	pchquick, depr, chato_ia, acc, mom12m, maxret, pchcurrat
mid	agr, quick, currat, depr, lgr, idiovol, cashdebt, retvol, mom1m, beta, cashpr, beta_sq	rd_mve, chshrout, gma	mom12m, chato_ia

Table C.2: FC selection, multivariate regressions, Sample Period 1970-01 until 2003-12. This table gives an overview of the selected FC by POLS and FM. The column "JOINT FC" exhibits the FC selected by both POLS and FM. Accordingly, "POLS only" ("FM only") presents the FC included solely in the POLS (FM) selection procedure.

	large+mid	large	mid
SE Ratio Mean	1.29	1.31	1.20
SE Ratio Median	1.21	1.24	1.14
Abs Coef Ratio	1.13	1.22	1.09
t-value Ratio	0.87	0.99	0.83
FC POLS|FM	5|8	4|2	5|11
JOINT|POLS|FM	4|1|4	0|4|2	4|1|7

Table C.3: Meta statistics FC selection, multivariate regressions, Sample Period 2003-12 until 2014-12. This table shows a meta comparison of the FC selection by POLS and FM. The ratio rows always express the POLS over the FM quantities. The row "SE Ratio Mean" measures the average SE Ratio of all FC, "SE Ratio Median" accordingly the median. "Abs Coef Ratio" exhibits the ratio of the sum of all absolute coef sizes. Accordingly, the "t-value Ratio" expresses the ratio of the sum of absolute t-values. "FC POLS|FM" counts the number of FC selected by each method. "JOINT|POLS|FM" indicates the selection commonality, by counting the number of jointly selected FC, the FC only selected by POLS, and last, the FC solely chosen by FM.

	JOINT FC	POLS only	FM only
large+mid	turn, sp, pchsale_pchrect, pchdepr	lgr	dy, retvol, cashdebt, mve_ia
large		pchquick, rd_sale, invest, cfp	retvol, age
mid	sp, cashdebt, turn, pchsaleinv	pchsale_pchrect	rd_mve, tang, pchsale_pchinvt, retvol, pchdepr, mve_ia, dy

Table C.4: FC selection, multivariate regressions, Sample Period 2003-12 until 2014-12. This table gives an overview of the selected FC by POLS and FM. The column "JOINT FC" exhibits the FC selected by both POLS and FM. Accordingly, "POLS only" ("FM only") presents the FC included solely in the POLS (FM) selection procedure.

Table C.5: Large and mid cap Stocks Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for FM and POLS regression portfolios for the joint set of large and mid cap stocks. Panel A shows the results for decile cutoffs and equally weighted stocks, Panel B the results value-weighted FF-style cutoffs. The column "OOS FC" indicates if the set of FC included at each point in time reflects only post-publication FC. Column "Adj" shows if p-values are adjusted for multiple testing, if empty no adjustment is performed — "Bonf" refers to Bonferroni, "BY" to Benjamini/Yekutieli. Column "2nd Stage" indicates if two regressions are run or not — first, a selection regression and second, a coefficient estimation with the subset of selected FC. All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which tests the FM vs. the POLS estimation seperately for each specification. The input data are winsorized at the 0.01 level and standardized w.r.t to the cross-section.

Panel A: **EQ, decile**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	8.99	17.31	0.52	8.8***(3.06)	4.3* (1.82)	-0.96
2	POLS				10.89	16.94	0.64	10.7***(3.55)	6.2***(2.63)	
3	FM	Yes	-	No	8.01	17.42	0.46	8.0***(2.66)	3.9 (1.53)	-0.74
4	POLS				10.50	17.95	0.58	9.4***(2.74)	5.7* (1.79)	
5	FM	No	-	Yes	15.09	17.21	0.88	14.0***(4.76)	9.2***(3.80)	-0.81
6	POLS				15.06	14.56	1.03	13.5***(5.03)	10.9***(3.90)	
7	FM	No	-	No	13.19	17.87	0.74	9.0***(3.43)	6.4** (2.19)	-1.23
8	POLS				15.19	14.63	1.04	9.2***(3.60)	8.9***(3.18)	
9	FM	Yes	BY	Yes	6.75	18.59	0.36	8.2***(2.66)	3.2 (1.29)	-0.76
10	POLS				7.98	17.15	0.47	11.5***(3.60)	7.0***(2.82)	
11	FM	Yes	Holm	Yes	7.04	17.81	0.40	8.0***(2.67)	3.3 (1.34)	-0.51
12	POLS				7.79	17.00	0.46	11.2***(3.56)	6.6***(2.76)	
13	FM	Yes	Bonf	Yes	7.75	17.97	0.43	9.0***(2.98)	4.2* (1.70)	-0.18
14	POLS				7.82	17.24	0.45	11.2***(3.53)	6.6***(2.71)	

Panel B: **VW, FF-style**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	5.55	12.31	0.45	6.8***(3.28)	3.8** (2.22)	-0.35
2	POLS				6.41	12.94	0.50	8.3***(3.37)	5.3** (2.56)	
3	FM	Yes	-	No	4.57	12.51	0.37	6.0***(2.76)	3.5* (1.79)	-0.63
4	POLS				6.13	12.77	0.48	7.3***(2.82)	4.7** (1.97)	
5	FM	No	-	Yes	7.06	12.61	0.56	6.4***(3.08)	3.2* (1.89)	-0.95
6	POLS				8.28	11.43	0.73	9.2***(4.37)	7.4***(3.08)	
7	FM	No	-	No	6.51	12.83	0.51	4.4** (2.21)	2.8 (1.34)	-1.57
8	POLS				8.35	10.18	0.82	6.8***(3.41)	6.4***(2.82)	
9	FM	Yes	BY	Yes	4.11	13.35	0.31	6.2***(2.83)	2.8 (1.55)	-0.29
10	POLS				4.71	13.70	0.34	7.9***(3.15)	4.9** (2.31)	
11	FM	Yes	Holm	Yes	4.34	12.98	0.33	5.9***(2.78)	2.8 (1.56)	0.03
12	POLS				4.44	13.43	0.33	7.7***(3.16)	4.6** (2.32)	
13	FM	Yes	Bonf	Yes	5.11	13.14	0.39	7.0***(3.23)	3.6** (2.08)	0.71
14	POLS				4.27	13.62	0.31	7.6***(3.06)	4.4** (2.18)	

Table C.6: Decile Cutoffs, EW weighted, Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for FM and POLS regression portfolios for the joint set of large and mid cap stocks. Panel A shows the results for large cap stocks, Panel B the results for mid caps. Each panel contains equal-weighted (EW) long short portfolios. The column "OOS FC" indicates if the set of FC included at each point in time reflects only post-publication FC. Column "Adj" shows if p-values are adjusted for multiple testing, if empty no adjustment is performed — "Bonf" refers to Bonferroni, "BY" to Benjamini/Yekutieli. Column "2nd Stage" indicates if two regressions are run or not — first, a selection regression and second, a coefficient estimation with the subset of selected FC. All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which tests the FM vs. the POLS estimation seperately for each specification. The input data are winsorized at the 0.01 level and standardized w.r.t to the cross-section.

Panel A: **Large Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	7.11	17.21	0.41	8.1***(2.69)	3.7 (1.51)	-0.21
2	POLS				7.62	16.95	0.45	11.0***(3.61)	7.2***(2.66)	
3	FM	Yes	-	No	6.04	16.66	0.36	8.1***(2.76)	4.9* (1.79)	-0.00
4	POLS				6.33	17.44	0.36	9.9***(3.20)	6.5** (2.26)	
5	FM	No	-	Yes	11.09	16.97	0.65	10.0***(3.40)	5.8** (2.35)	0.13
6	POLS				9.04	14.54	0.62	7.9***(2.89)	7.5***(2.85)	
7	FM	No	-	No	9.16	15.61	0.59	9.2***(3.59)	6.7***(2.63)	0.15
8	POLS				8.88	16.09	0.55	8.5***(2.87)	7.6***(2.63)	
9	FM	Yes	BY	Yes	5.83	16.82	0.35	7.9***(2.63)	4.4 (1.63)	-0.54
10	POLS				8.61	18.77	0.46	10.4** (2.50)	5.9* (1.65)	
11	FM	Yes	Holm	Yes	6.33	17.32	0.37	9.0***(2.76)	4.9* (1.88)	-0.50
12	POLS				8.04	17.86	0.45	11.7***(3.31)	7.4** (2.57)	
13	FM	Yes	Bonf	Yes	6.38	17.25	0.37	9.0***(2.82)	5.0* (1.92)	-0.60
14	POLS				8.32	17.63	0.47	11.8***(3.37)	7.6***(2.65)	

Panel B: **Mid Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	9.53	16.72	0.57	9.5***(3.26)	5.3** (2.13)	-1.77*
2	POLS				12.67	15.93	0.80	12.8***(4.38)	8.8***(3.64)	
3	FM	Yes	-	No	8.65	15.70	0.55	8.3***(3.15)	5.1** (1.99)	-0.76
4	POLS				10.76	16.05	0.67	11.3***(3.68)	7.3***(2.85)	
5	FM	No	-	Yes	17.71	16.50	1.07	15.3***(5.46)	11.1***(4.48)	-0.63
6	POLS				17.18	14.62	1.17	15.0***(5.57)	11.8***(4.78)	
7	FM	No	-	No	15.64	16.66	0.94	11.3***(4.53)	8.6***(3.11)	-0.69
8	POLS				15.45	14.70	1.05	10.5***(4.51)	8.4***(3.40)	
9	FM	Yes	BY	Yes	8.60	16.51	0.52	9.1***(2.98)	5.4** (2.06)	-1.18
10	POLS				10.11	15.83	0.64	11.2***(3.66)	7.2***(2.87)	
11	FM	Yes	Holm	Yes	8.37	16.49	0.51	9.1***(2.98)	5.4** (2.06)	-1.44
12	POLS				10.33	15.88	0.65	11.0***(3.54)	7.0***(2.75)	
13	FM	Yes	Bonf	Yes	7.96	16.66	0.48	8.9***(2.92)	5.2** (1.98)	-1.11
14	POLS				9.26	15.98	0.58	10.3***(3.33)	6.3** (2.52)	

Table C.7: FF-style Cutoffs, EW weighted, Sample Period 1982-01 until 2014-12: The tables exhibit performance measures for FM and POLS regression portfolios for the joint set of large and mid cap stocks. Panel A shows the results for large cap stocks, Panel B the results for mid caps. Each panel contains equal-weighted (EW) long short portfolios. The column "OOS FC" indicates if the set of FC included at each point in time reflects only post-publication FC. Column "Adj" shows if p-values are adjusted for multiple testing, if empty no adjustment is performed — "Bonf" refers to Bonferroni, "BY" to Benjamini/Yekutieli. Column "2nd Stage" indicates if two regressions are run or not — first, a selection regression and second, a coefficient estimation with the subset of selected FC. All performance measures are displayed annualized. The values in brackets of the α-columns indicate t-values, which are based on standard deviation estimates corrected for HAC, using a lag of one period. "FF 5" abbreviates the Fama and French (2014) five-factor model, "FF 5 + mom" the momentum augmented version of the latter. The Sharpe ratio test column reflects t-values of the Ledoit and Wolf (2008) test, which tests the FM vs. the POLS estimation seperately for each specification. The input data are winsorized at the 0.01 level and standardized w.r.t to the cross-section.

Panel A: **Large Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	6.04	12.57	0.48	8.3***(3.65)	5.2***(2.77)	-0.37
2	POLS				7.04	13.18	0.53	10.8***(4.57)	8.0***(3.94)	
3	FM	Yes	-	No	5.64	12.52	0.45	8.3***(3.79)	5.8***(3.00)	-0.32
4	POLS				6.79	13.41	0.51	11.3***(4.76)	8.8***(4.14)	
5	FM	No	-	Yes	7.26	12.32	0.59	6.5***(3.03)	3.5** (2.12)	-0.32
6	POLS				6.84	10.37	0.66	7.0***(3.63)	6.7***(3.33)	
7	FM	No	-	No	7.10	11.66	0.61	7.0***(3.85)	5.3***(2.94)	0.04
8	POLS				7.11	11.85	0.60	8.4***(3.96)	7.5***(3.41)	
9	FM	Yes	BY	Yes	5.45	12.67	0.43	8.3***(3.71)	5.6***(2.91)	0.20
10	POLS				5.18	13.33	0.39	7.1** (2.47)	4.1* (1.67)	
11	FM	Yes	Holm	Yes	5.68	12.89	0.44	8.6***(3.51)	5.5***(2.94)	0.19
12	POLS				5.28	13.00	0.41	8.3***(3.27)	5.5***(2.64)	
13	FM	Yes	Bonf	Yes	5.67	13.11	0.43	8.4***(3.33)	5.3***(2.80)	0.02
14	POLS				5.53	12.87	0.43	8.5***(3.34)	5.7***(2.73)	

Panel B: **Mid Cap Stocks**

#	Portfolio	OOS FC	Adj	2nd Stage	mean	std	SR	FF-5 α	FF-5 + Mom α	SR Test
1	FM	Yes	-	Yes	5.95	9.98	0.60	5.2***(3.15)	2.7* (1.68)	-0.96
2	POLS				7.36	10.10	0.73	7.2***(3.82)	4.5***(2.90)	
3	FM	Yes	-	No	6.19	9.20	0.67	5.1***(3.27)	3.3** (2.16)	0.25
4	POLS				6.30	9.95	0.63	6.6***(3.33)	4.0** (2.40)	
5	FM	No	-	Yes	11.01	10.46	1.05	8.8***(5.25)	6.0***(4.03)	-1.00
6	POLS				11.43	9.43	1.21	9.8***(5.73)	7.6***(5.05)	
7	FM	No	-	No	11.50	10.06	1.14	8.1***(5.44)	6.6***(4.02)	0.78
8	POLS				9.73	9.43	1.03	6.2***(4.18)	4.8***(3.19)	
9	FM	Yes	BY	Yes	5.74	9.84	0.58	5.6***(3.15)	3.5** (2.15)	-1.50
10	POLS				7.00	9.57	0.73	7.2***(4.00)	4.7***(3.26)	
11	FM	Yes	Holm	Yes	5.54	9.99	0.55	5.4***(3.02)	3.2* (1.96)	-2.03**
12	POLS				7.23	9.73	0.74	7.2***(4.00)	4.7***(3.21)	
13	FM	Yes	Bonf	Yes	5.62	10.05	0.56	5.5***(3.09)	3.3** (2.02)	-1.58
14	POLS				6.86	9.80	0.70	7.1***(3.93)	4.7***(3.08)	

ID	Acronym	Name	Description	Reference
1	beta	Beta	Measured based on 3 years (min 52 weeks) weekly excess returns with standard ols ($y = c + \beta x$)	Fama and MacBeth (1973)
2	beta_sq	Beta squared	Simply obtained by squaring the β based on the beta from # 1	Fama and MacBeth (1973)
3	retvol	Volatility	Volatility is measured by the standard deviation of daily returns of the previous months	Ang et al. (2006)
4	maxret	Maximum return	Maximum return is defined over the max of the daily returns in month $t-1$	Bali et al. (2011)
5	idiovol	Idiosyncratic volatility	Calculated based on the residuals of regression in # 1	Ali et al. (2003)
6	mom1m	1-month momentum	Return in month $t-1$	Jegadeesh (1990)
7	mom6m	6-month momentum	Cumulative return over 5 months ending in $t-2$	Jegadeesh and Titman (1993)
8	mom12m	12-month momentum	Cumulative return over 11 months ending in $t-2$	Jegadeesh (1990)
9	mom36m	36-month momentum	Cumulative return over 24 months ending in $t-13$	Bondt and Thaler (1985)
10	mve	Market capitalization (size)	log of (SHROUT×PRC)	Banz (1981)
11	ep	Earnings-to-price	Earnings per share	Basu (1977)
12	dy	Dividends-to-price	Yearly dividends (dvt) divided by market cap at fiscal year	Litzenberger and Ramaswamy (1979)
13	bm	Book-to-market	Book value of equity (ceq) divided by market cap	Rosenberg et al. (1985)
14	lev	Leverage	Total liabilities (lt) divided by market cap	Bhandari (1988)
15	currat	Current ratio	Current assets (act) divided by current liabilities (lct)	Ou and Penman (1989)
16	pchcurrat	Pct change in current ratio	Percentage change in currat from year $t-1$ to t	Ou and Penman (1989)
17	quick	Quick ratio	Current assets (act) minus inventory (invt), divided by current liabilities (lct)	Ou and Penman (1989)
18	pchquick	Pct change in quick ratio	Percentage change in quick from year $t-1$ to t	Ou and Penman (1989)
19	salecash	Sales-to-cash	Annual sales (sale) divided by cash and cash equivalents (che)	Ou and Penman (1989)
20	salerec	Sales-to-receivables	Annual sales (sale) divided by accounts receivable (rect)	Ou and Penman (1989)
21	saleinv	Sales-to-inventory	Annual sales (sale) divided by total invetory (invt)	Ou and Penman (1989)
22	pchsaleinv	Pct change in sales-to-inventory	Percentage change in saleinv from year $t-1$ to t	Ou and Penman (1989)
23	cashdebt	Cashflow-to-debt	Earnings before depreciation and extraordinary items (ib + dp) divi ded by avg total liabilities (lt)	Ou and Penman (1989)
24	baspread	Illiquidity (bid-ask-spread)	Monthly avg of daily bid-ask spread divided by avg of daily bid-ask spread	Amihud and Mendelson (1989)
25	depr	Depreciation-to-gross PP&E	Depreciation expense (dp) divided by gross PPE (ppegt)	Holthausen and Larcker (1992)
26	pchdepr	Pct change in Depreciation-to-gross PP&E	Percentage change in depr from year $t-1$ to t	Holthausen and Larcker (1992)
27	mve_ia	Industry-adjusted firm size	Log market caps are adjusted by log of the mean of the industry	Asness et al. (2000)
28	cfp_ia	Industry-adjusted cashflow-to-price	Industry adjusted cash flow-to-price ratio equal weighted average	Asness et al. (2000)
29	bm_ia	Industry-adjusted book-to-market	Industry adjusted book-to-market equal weighted average	Asness et al. (2000)
30	sgr	Annual sales growth	Percentage change in sales from year $t-1$ to t	Lakonishok et al. (1994)
31	ipo	IPO	Indicated by 1 if first 12 months available on CRSP monthly file	Loughran and Ritter (1995)
32	divi	Dividend initiation	Indicated by 1 if company pays dividends but did not in prior year.	Michaely et al. (1995)
33	divo	Dividend omission	Indicated by 1 if company does not pay dividends but did in prior year.	Michaely et al. (1995)
34	sp	Sales-to-price	Annual sales (sale) divided by market cap	Barbee Jr et al. (1996)
35	acc	WC accruals	(ib) - (oancf)/(at), if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Sloan (1996)
36	turn	Share turnover	Average monthly trading volume for the three months $t-3$ to $t-1$ divided by SHROUT at $t-1$	Datar et al. (1998)
37	pchsale_pchinvt	Delta pct change sales vs. inventory	Difference of percentage changes in sales (sale) and inventory (invt)	Abarbanell and Bushee (1997)
38	pchsale_pchrect	Delta pct change sales vs. receivables	Difference of percentage changes in sales (sale) and receivables (rect)	Abarbanell and Bushee (1997)
39	pchcapx_ia	CAPEX	Industry adjusted (two digit SIC) fiscal year mean adjusted percentage change in capital expenditures (capx)	Abarbanell and Bushee (1997)
40	pchgm_pchsale	Delta pct gross margin vs. sales	Annual percentage change in gross margin (sale minus cogs) minus percentage change in sales (sale)	Abarbanell and Bushee (1997)

Table C.8: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values. The table is taken from Messmer and Audrino (2017).

ID	Acronym	Name	Description	Reference
41	pchsale_pchxsga	Delta pct sales vs. SGaA	Annual percentage change in sales (sale) minus percentage change in SGaA (xsga)	Abarbanell and Bushee (1997)
42	dolvol	Dollar trading volume	Log of trading volume times price per share from month t-2	Chordia et al. (2001)
43	std_dolvol	Volatility trading volume	Monthly standard deviation of daily trading volume	Chordia et al. (2001)
44	std_turn	Volatility turnover	Monthly standard deviation of daily share turnover	Chordia et al. (2001)
45	chinv	Change in inventory	First difference of inventory (invt) divided by total assets	Thomas and Zhang (2002)
46	pchemp_ia	Industry-adjusted pch in employees	Industry adjusted percentage change in employees	Asness et al. (2000)
47	cfp	Cashflow-to-price	Operating cash flows (oancf) scaled by market capitalization (fiscal year end)	Desai et al. (2004)
48	rd	R&D Increase	If annual increase in R&D expenses (xrd) scaled by total assets (at) >0.05, 1, else 0	Eberhart et al. (2004)
49	lgr	Pct change in long-term debt	Annual percentage change in long term debt (lt)	Richardson et al. (2005)
50	egr	Pct change in book equity	Annual percentage change in book equity (ceq)	Richardson et al. (2005)
51	rd_sale	R&D-to-sales	R&D expenses(xrd) scaled by sales (sale)	Guo et al. (2006)
52	rd_mve	R&D-to-market	R&D expenses(xrd) scaled by market cap	Guo et al. (2006)
53	chg_mom6m	change in mom6m	difference of mom6m measured at t and $t-6$	Gettleman and Marks (2006)
54	hire	Pct change in employee	Annual percentage change in employee (emp)	Belo et al. (2014)
55	agr	Asset growth	Annual percentage change in assets (at)	Cooper et al. (2008)
56	cashpr	Cash productivity	Market cap plus long term debt (dltt) minus assets (at) divided by cash (che)	Chandrashekar and Rao (2009)
57	gma	Gross-profitability	Sales (sale) minus costs of goods sold (cogs) divided by one-year lagged assets(at)	Novy-Marx (2013)
58	cash	Cash-to-assets	Cash (che) divided by assets(at)	Palazzo (2012)
59	pctacc	Accruals-to-income	(ib) minus (oancf) divided by abs ((ib)), when (ib) equals 0, it is set to 0.01, if (oancf) is missing then (ib)-(delta_act)-(delta_che) -(delta_lct) + (delta_dlc) + (txp-dp) where each item 0 if missing	Hafzalla et al. (2011)
60	absacc	Absolut accruals	Absolute value of acc	Bandyopadhyay et al. (2010)
61	roic	Return on invested capital	Earnings before interest and taxes (ebit) - non-operating income (nopi), divided by non-cash enterprise value (ceq+lt-che)	Brown and Rowe (2007)
62	grcapex	Pct change in two year CAPX	Percentage change in two year capital expenditure (capx)	Anderson and Garcia (2006)
63	tang	Debt capacity-to-firm-tangability	(Cash (che) + 0.715 receivables (rect) + 0.547 inventory(invt) + 0.535 (ppegt))/ total assets (at)	Hahn and Lee (2009)
64	chshrout	Change in shares-outstanding	Yearly percentage change in outstanding shares (SHROUT)	Pontiff and Woodgate (2008)
65	invest	CAPEX and inventory	Yearly difference in gross property, plant and equipment (ppegt) + diff in (invt) / (t-1) total assets (at)	Chen and Zhang (2010)
66	age	Years since CS coverage	Years since first compustat coverage years(datadate - min(datadate))	Jiang et al. (2005)
67	chpm_ia	Industry-adjusted change in profit margin	Industry adjusted (two-digit SIC) change in profit margin (ib/sale)	Soliman (2008)
68	chato_ia	Industriy-adjusted change in asset turnover	Industry adjusted (two-digit SIC) change in asset turnover (sale/at)	Soliman (2008)

Table C.9: The table displays the firm characteristics used. Most definitions are taken from Green et al. (2017). If not otherwise stated, accounting ratios always refer to fiscal year end values.The table is taken from Messmer and Audrino (2017).

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abarbanell, J. S. and B. J. Bushee (1997). Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research*, 1–24.

Ali, A., L.-S. Hwang, and M. A. Trombley (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics 69*(2), 355–373.

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

Anderson, C. W. and L. Garcia, Feijóo (2006). Empirical evidence on capital investment, growth options, and security returns. *The Journal of Finance 61*(1), 171–194.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Ang, A. (2014). *Asset management: A systematic approach to factor investing.* Oxford University Press.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The Journal of Finance 61*(1), 259–299.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics 91*(1), 1–23.

Ang, A., J. Liu, and K. Schwarz (2016). Using stocks or portfolios in tests of factor models. *Working Paper*.

Asness, C., A. Frazzini, R. Israel, and T. Moskowitz (2014). Fact, fiction, and momentum investing. *The Journal of Portfolio Management 40*(5), 75–92.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2013). Quality minus junk. Available at SSRN: https://ssrn.com/abstract=2312432.

Asness, C. S., R. B. Porter, and R. L. Stevens (2000). Predicting stock returns using industry-relative firm characteristics. *Available at SSRN 213872*. Available at SSRN: https://ssrn.com/abstract=213872.

Audrino, F. and L. Camponovo (2015). Oracle properties, bias correction, and inference of the adaptive lasso for time series extremum estimators. Available at SSRN: https://ssrn.com/abstract=2340030.

Audrino, F. and S. D. Knaus (2016). Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews 35*(8-10), 1485–1521.

Bali, T. G., N. Cakici, and R. F. Whitelaw (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics 99*(2), 427–446.

Bandyopadhyay, S. P., A. G. Huang, and T. S. Wirjanto (2010). The accrual volatility anomaly. Technical report, working paper, University of Waterloo.

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

Barbee Jr, W. C., S. Mukherji, and G. A. Raines (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal 52*(2), 56–60.

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

Bekaert, G., R. J. Hodrick, and X. Zhang (2012). Aggregate idiosyncratic volatility. *Journal of Financial and Quantitative Analysis 47*(6), 1155–1185.

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

Bengio, Y., Y. LeCun, et al. (2007). Scaling learning algorithms towards ai. *Large-scale kernel machines 34*(5), 1–41.

Bergstra, J. and Y. Bengio (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research 13*(Feb), 281–305.

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

Bishop, C. M. (1995a). *Neural networks for pattern recognition*. Oxford university press.

Bishop, C. M. (1995b). Regularization and complexity control in feed-forward networks. In *Proceedings International Conference on Artificial Neural Networks ICANNâ€™95*, pp. 807–814.

Blitz, D. C. and P. van Vliet (2007). The volatility effect. *The Journal of Portfolio Management 34(1)*, 102–113.

Bondt, W. F. and R. Thaler (1985). Does the stock market overreact? *The Journal of Finance 40*(3), 793–805.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies 22*(9), 3411–3447.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and regression trees*. CRC press.

Brown, D. P. and B. Rowe (2007). The productivity premium in equity returns. Available at SSRN: https://ssrn.com/abstract=993467.

Bryzgalova, S. (2016). Spurious factors in linear asset pricing models. Working Paper.

Bühlmann, P. and S. Van De Geer (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer.

Caner, M. and A. B. Kock (2014). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *Working Paper*.

Caner, M. and H. H. Zhang (2014). Adaptive elastic net for generalized methods of moments. *Journal of Business & Economic Statistics 32*(1), 30–47.

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

Chandrashekar, S. and R. K. Rao (2009). The productivity of corporate cash holdings and the cross-section of expected stock returns. *Working Paper*.

Chen, L. and L. Zhang (2010). A better three-factor model that explains more anomalies. *Journal of Finance 65*(2), 563–595.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001). Trading activity and expected stock returns. *Journal of Financial Economics 59*(1), 3–32.

Cochrane, J. H. (2005). *Asset Pricing*, Volume 41. Princeton University Press Princeton.

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

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

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS) 2*(4), 303–314.

Daniel, K. and T. J. Moskowitz (2016). Momentum crashes. *Journal of Financial Economics 122*(2), 221–247.

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

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2017). A portfolio perspective on the multitude of firm characteristics. Available at SSRN: https://ssrn.com/abstract=2912819.

Desai, H., S. Rajgopal, and M. Venkatachalam (2004). Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review 79*(2), 355–385.

Dixon, M., D. Klabjan, and J. H. Bang (2015). Implementing deep neural networks for financial market prediction on the intel xeon phi. pp. 6.

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics 80*(4), 549–560.

Duchi, J., E. Hazan, and Y. Singer (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research 12*(Jul), 2121–2159.

Eberhart, A. C., W. F. Maxwell, and A. R. Siddique (2004). An examination of long-term abnormal stock returns and operating performance following r&d increases. *The Journal of Finance 59*(2), 623–650.

Efron, B., T. Hastie, I. Johnstone, R. Tibshirani, et al. (2004). Least angle regression. *The Annals of Statistics 32*(2), 407–499.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *The Journal of Finance 47*(2), 427–465.

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

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (2008). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and K. R. French (2014). A five-factor asset pricing model. *Journal of Financial Economics*.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 607–636.

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

Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting characteristics non-parametrically. Available at SSRN: https://ssrn.com/abstract=2820700.

Gettleman, E. and J. M. Marks (2006). Acceleration strategies. *Working Paper*.

Glorot, X., A. Bordes, and Y. Bengio (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Goyal, A. (2012). Empirical cross-sectional asset pricing: a survey. *Financial Markets and Portfolio Management 26*(1), 3–38.

Goyal, P., P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He (2017). Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*.

Green, J., J. R. Hand, and F. Zhang (2017). The characteristics that provide independent information about average us monthly stock returns. *Review of Financial Studies*. Forthcoming.

Guo, R.-J., B. Lev, and C. Shi (2006). Explaining the short-and long-term ipo anomalies in the us by r&d. *Journal of Business Finance & Accounting 33*(3-4), 550–579.

Hafzalla, N., R. Lundholm, and E. Matthew Van Winkle (2011). Percent accruals. *The Accounting Review 86*(1), 209–236.

Hahn, J. and H. Lee (2009). Financial constraints, debt capacity, and the cross-section of stock returns. *The Journal of Finance 64*(2), 891–921.

Hand, J. R. and J. Green (2011). The importance of accounting information in portfolio optimization. *Journal of Accounting, Auditing & Finance 26*(1), 1–34.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics 20*(7), 873–889.

Hardin, J., S. R. Garcia, D. Golan, et al. (2013). A method for generating realistic correlation matrices. *The Annals of Applied Statistics 7*(3), 1733–1762.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning*, Volume 2. Springer.

Haugen, R. A. and N. L. Baker (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics 41*(3), 401–439.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons.

Hinton, G. E. (2007). To recognize shapes, first learn to generate images. *Progress in brain research 165*, 535–547.

Holthausen, R. W. and D. F. Larcker (1992). The prediction of stock returns using financial statement information. *Journal of Accounting and Economics 15*(2), 373–411.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer feedforward networks are universal approximators. *Neural networks 2*(5), 359–366.

Hou, K., C. Xue, and L. Zhang (2017). Replicating anomalies. Available at SSRN: https://ssrn.com/abstract=2961979.

Jagannathan, R., G. Skoulakis, and Z. Wang (2010). The analysis of the cross section of security returns. *Handbook of Financial Econometrics 2*, 73–134.

Jarrett, K., K. Kavukcuoglu, Y. LeCun, et al. (2009). What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 2146–2153. IEEE.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance 45*(3), 881–898.

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

Jensen, M. C., F. Black, and M. S. Scholes (1972). The capital asset pricing model: Some empirical tests.

Jiang, G., C. M. Lee, and Y. Zhang (2005). Information uncertainty and expected returns. *Review of Accounting Studies 10*(2-3), 185–221.

Kinga, D. and J. B. Adam (2015). A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

Kock, A. B. (2016a). Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions. *Econometric Theory 32*(01), 243–259.

Kock, A. B. (2016b). Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models. *Journal of Econometrics 195*(1), 71–85.

Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics 186*(2), 325–344.

Kogan, L. and M. Tian (2013). Firm characteristics and empirical factor models: a data-mining experiment. *Working Paper*.

Kozak, S., S. Nagel, and S. Santosh (2017). Shrinking the cross section. Available at SSRN: https://ssrn.com/abstract=2945663.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance 49*(5), 1541–1578.

Larochelle, H., D. Erhan, A. Courville, J. Bergstra, and Y. Bengio (2007). An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th international conference on Machine learning*, pp. 473–480. ACM.

Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance 15*(5), 850–859.

Leung, M. T., H. Daouk, and A.-S. Chen (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting 16*(2), 173–190.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review 4*(1), 1–44.

Linnainmaa, J. T. and M. R. Roberts (2017). The history of the cross section of stock returns. Available at SSRN: https://ssrn.com/abstract=2897719.

Litzenberger, R. H. and K. Ramaswamy (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*(2), 163–195.

Loughran, T. and J. R. Ritter (1995). The new issues puzzle. *The Journal of Finance 50*(1), 23–51.

McCulloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics 5*(4), 115–133.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

Medeiros, M. C. and E. Mendes (2012). Estimating high-dimensional time series models. *Working Paper*.

Meinshausen, N. (2006). Quantile regression forests. *The Journal of Machine Learning Research 7*, 983–999.

Messmer, M. (2017). Deep learning and the cross-section of expected returns.

Messmer, M. and F. Audrino (2017). The (adaptive) lasso in the zoo - firm characteristic selection in the cross-section of expected returns. Available at SSRN: https://ssrn.com/abstract=2930436.

Michaely, R., R. H. Thaler, and K. L. Womack (1995). Price reactions to dividend initiations and omissions: Overreaction or drift? *The Journal of Finance 50*(2), 573–608.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN*.

Nair, V. and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Ou, J. A. and S. H. Penman (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics 11*(4), 295–329.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics 104*(1), 162–185.

Parker, D. B. (1985). Learning logic.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies 22*(1), 435–480.

Polyak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics 4*(5), 1–17.

Pontiff, J. and A. Woodgate (2008). Share issuance and cross-sectional returns. *The Journal of Finance 63*(2), 921–945.

Poultney, C., S. Chopra, Y. L. Cun, et al. (2007). Efficient learning of sparse representations with an energy-based model. In *Advances in neural information processing systems*, pp. 1137–1144.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23 (2)*, 821–862.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics 39*(3), 437–485.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.

Rosenberg, B., K. Reid, and R. Lanstein (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management 11*(3), 9–16.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Saad, D. (1998). Online algorithms and stochastic approximations. *Online Learning 5*.

Schaul, T., I. Antonoglou, and D. Silver (2013). Unit tests for stochastic optimization. *arXiv preprint arXiv:1312.6055*.

Shanken, J. (1992). On the estimation of beta-pricing models. *The Review of Financial Studies 5*(1), 1–33.

Sjöberg, J. and L. Ljung (1995). Overtraining, regularization and searching for a minimum, with application to neural networks. *International Journal of Control 62*(6), 1391–1407.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?(digest summary). *Accounting Review 71*(3), 289–315.

Soliman, M. T. (2008). The use of dupont analysis by market participants. *The Accounting Review 83*(3), 823–853.

Srivastava, N., G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*(1), 1929–1958.

Szegedy, C., W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Takeuchi, L. and Y.-Y. A. Lee (2013). Applying deep learning to enhance momentum trading strategies in stocks. *Working Paper*.

Thomas, J. K. and H. Zhang (2002). Inventory changes and future returns. *Review of Accounting Studies 7*(2-3), 163–187.

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

Titman, S., K. J. Wei, and F. Xie (2004). Capital investments and stock returns. *Journal of financial and Quantitative Analysis 39*(4), 677–700.

Trippi, R. R. and D. DeSieno (1992). Trading equity index futures with a neural network. *The Journal of Portfolio Management 19*(1), 27–33.

Widrow, B., M. E. Hoff, et al. (1960). Adaptive switching circuits. *IRE WESCON convention record 4*(1), 96–104.

Williams, D. and G. Hinton (1986). Learning representations by back-propagating errors. *Nature 323*(6088), 533–538.

Wilson, D. R. and T. R. Martinez (2003). The general inefficiency of batch training for gradient descent learning. *Neural Networks 16*(10), 1429–1451.

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

Person

Name	Marcial Messmer
Date of birth	August 25th, 1986
Nationality	German
Address	Grütstrasse 39, 8134 Adliswil, Switzerland
Email	marcial.messmer@gmail.com
Phone	+41 (0) 76 650 4082

Education

2013-2018	**University of St. Gallen, Switzerland** Ph.D. in Economics and Finance with specialization in Finance
2017	**Columbia University, NY, USA** Visiting Scholar at Columbia Business School
2011-2012	**University of St. Gallen, Switzerland** Master in Quantitative Economics and Finance
2010	**Boston University, MA, USA** Mathematical Finance
2006-2009	**University of Mannheim, Germany** Bachelor in Economics
2008	**University of California Los Angeles, CA, USA** Summer Program
2001-2006	**Merz-Schule Stuttgart, Germany** German high school (Abitur)

Professional experience

Apr 2018 - current	**UBS Asset Management, Zurich, Switzerland: Quantitative Analyst**
	– Research for systematic and index investments
Jan 2015 - current	**FactorStrategies.com, Internet: Founder and co-owner**
	– Research platform for factor investing and theory
Mar 2012 - Mar 2018	**Kraus Partner Investment Solutions AG, Zurich, Switzerland: Senior Quantitative Analyst (60%)**
	– Management of a systematic global equity fund – Model development for return- and risk-predictions
Apr 2016 - Jan 2017	**University of St. Gallen, Switzerland: Assistant to Prof. Audrino**
	– Support of research activities – Teaching undergraduate Statistics (exercises)
Oct 2011 - Jul 2012	**University of St. Gallen, Switzerland: Research Assistant Institute of Management**
	– IT/data support of research activities
Jul 2011 - Sep 2011	**Allianz Global Investors, Frankfurt, Germany: Internship Portfolio Strategy Alternative Investment**
	– Programmed tools for quantitative analytics (VBA)
Oct 2009 - Jun 2010	**Fellbacher Bank eG, Fellbach, Germany: Internship Treasury**
	– Implemented and systematized credit research

Languages

German	Native Speaker
English	Fluent
Russian	Basics

IT

Python	advanced+
Matlab	advanced
VBA	advanced
Git & SVN	intermediate
SQL	intermediate
R	intermediate
AWS	intermediate
Linux	intermediate
Mathematica	basics
HTML	basics
CSS	basics
JS	basics
EViews	basics