The (adaptive) Lasso in the Zoo
- Firm Characteristic Selection in the Cross-Section of Expected Returns*

Version 1.3.7

Marcial Messmer ‡ Francesco Audrino***
University of St.Gallen

March 9, 2017

Abstract

We find short-term reversal, the twelve-months momentum and research spending scaled by market-value to be the firm characteristics (FC) most robustly selected by the adaptive Lasso in the US cross-section of stock returns. Moreover, the majority of the 68 FC included in our analysis are not considered. Nonetheless, the return process we identify is multi-dimensional, comprising 14 FC. Additionally, our Monte Carlo Simulations indicate that the adaptive Lasso is superior to Lasso and OLS-based selection in panel specifications with a low signal-to-noise ratio. The results are robust to various assumptions. These findings gain support by an empirical out-of-sample factor analysis.

JEL Classification: C52; C53; C58; G12; G17;
Key words: Factor models; cross-section; Lasso; asset pricing; firm characteristics selection;

*We are grateful to workshop participants at the CMStatistics 2016 conference, University of St.Gallen and University of Konstanz. Furthermore, we are thankful for insightful comments from Alberto Martin-Utrera and Paul Söderlind.
‡E-mail: marcial.messmer@gmail.com. Address: Bodanstrasse 6, 9000 St.Gallen, Switzerland.
***E-mail: francesco.audrino@unisg.ch. Address: Bodanstrasse 6, 9000 St.Gallen, Switzerland
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 an essential 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. More precisely, 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 in 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 which needs to be answered is which of the various stock characteristics provides 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 (FC) 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 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.

The study contributes to the literature mainly in the following two ways. First, we suggest the application of the adaptive Lasso in selecting firm characteristics because, as outlined below, the theoretical properties are appealing. Second, the study sheds more light on the performance of Lasso methods in panels for various error specifications by applying Monte Carlo simulations. One recent study, which

---

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.
resembles our approach, is the work of Green et al. (2017). Here the authors digest a large set of FC in a multivariate set-up; we closely follow their data construction and FC pre-selection procedure.

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 if 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 1965 until 2014.

The full pooled panel adaptive Lasso selection characterizes 14 FC of relevance for future differences in stock returns. Most consistent along size and time dimensions are the one month short-term reversal, the standard twelve-months momentum and research and development (R&D) spendings scaled by market capitalization. Moreover, the Fama and French (2014) five-factor model is almost fully represented in these findings, but complemented among others by four purely market-based measures.

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 but only if no relation between the return risk and the FC exists. Hence, FC, which carry information about the return’s risk but do not impact mean returns, are selected too often (relative to OLS).

Due to the inherently stylized character of simulation studies, we evaluate the performance of the two competing methods from an empirical perspective based on an out-of-sample factor analysis. For this purpose we form two factors according to return predictions, in which the set of FC included is determined by the adaptive Lasso and OLS, respectively. The findings reveal in most cases a statistically significant outperformance of the adaptive Lasso factor.

The paper is organized as follows. Section 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 4 covers the simulation study. The data is briefly discussed in part 5. The penultimate section covers the empirical work, including both the out-of-sample analysis as well as the FC selection results. The final part concludes.

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.
2.1 Firm characteristics and their relation to cross-sectional stock returns

Fama and French (1992) proceed in selecting relevant firm characteristics in the following way. In a first step, univariate regressions are run and significant coefficients are kept; 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 parameters \(\beta_t\). Statistical inference is then based on simple moment estimates of the time-series collection of \(\beta_t\) with \(t = 1, \ldots, T\). 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 consecutive 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 until today an interesting 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 phenomena 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, by purely 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. More precisely, he pays attention to the questions of how to deal with the high-dimensionality of characteristics in a way that no relevant information is left out. Central to the research agenda is the conditional expectation of excess returns,

\(^2\)To be precise, the independent variables B/M, ME, A/ME, A/BE are considered in terms of their natural logarithm.

\(^3\)See Table 8 and 9 in the appendix for an overview of the firm characteristics included in this study.
$E(R_{t+1}^e|C_t)$, and its linear interpretation in the form of the following panel regression:

$$R_{t+1,n}^e = x_{t,n}' \beta + \epsilon_{t+1,n}$$

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. 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 inference 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, respectively 24, firm characteristics. 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 in 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 the status of 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. In this way 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 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 accumulate to obtain any useful inference.

The study of Green et al. (2017) approaches the problem of reducing the dimensionality of expected

---

\(^4\)Cochrane (2011) provides the theoretical relation of the two procedures in the appendix.
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 twelve stocks 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, however it achieves higher Sharpe ratios compared to an OLS approach. The study includes 36 FC, of which roughly one-third of them 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 to the adaptive Lasso.

DeMiguel et al. (2017) analyze the FC selection from a portfolio perspective in a mean-variance 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 transactions costs in the portfolio optimization procedure increases this number to 15.

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 methodology in a comprehensive way. A more detailed introduction to estimation methods can be found in Cochrane (2005).

2.2 Lasso methods

The Lasso introduced by Tibshirani (1996) is motivated 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. The Lasso estimator optimizes least squares under, an additional condition involving the total sum of the absolute size of the coefficients (known as the $\ell_1$-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 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.

5 An overview of the family of linear methods can be found in Hastie et al. (2009)
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 a 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. In particular, Medeiros and Mendes (2012) and Audrino and Camponovo (2015) derive consistency properties of the adaptive Lasso in time series 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. 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.

3 Methodology

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

3.1 Notation

Generally, and if not explicitly otherwise stated, we follow the notation that \( n \) refers to stock \( n \) of \( N \) 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 risk-factor exposure with a positive risk-premium are denoted by \( p \) (of a total \( P \)) and define the set \( \mathcal{P} \); FC without a risk-premium, but measuring factor sensitivity, are defined by \( r \) (of a total \( R \), set \( \mathcal{R} \)), and uninformative FC are described by \( u \) (of a total \( U \), set \( \mathcal{U} \)). The total number of FC, \( C \) equals \( P + R + U \). The specification of FC for the simulation is outlined in more detail in section 4. Moreover, return-related parameters such as risk-premia (\( \mu \)) and volatility (\( \sigma \)) are expressed in annualized terms despite the fact that the data frequency is monthly.

---

6Additional literature references for the Lasso in panels are available in Kock (2016b).
7The HAR model describes the dynamics of realized variance as a process of past realized variances; see Corsi (2009) for details.
3.2 Methods

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

$$ Y = X \beta + \epsilon, $$  (1)

where $Y \in \mathbb{R}$ and $X \in \mathbb{R}^p$ and the corresponding response vector $Y_{n \times 1}$, the design matrix $X_{n \times p}$, the parameter vector $\beta_{p \times 1}$ and a vector of residuals $\epsilon_{n \times 1}$. Throughout this work we treat $Y$ and $X$ 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 $A = \{ j : \beta_j \neq 0 \}$ and let a measure of sparsity be defined as $\| \beta \|_0 = \sum_{j=1}^{p} |\beta_j|^0$, which counts all non-zero elements. Generally, we assume for (1) that $\| \beta \|_0 = q \leq p$.

The ordinary least squares (OLS) estimator of (1) is defined as

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

where $\|Y - X \beta\|_2^2 = \sum_{i} (Y_i - X_i \beta)^2$. Moreover, it holds that $\|\hat{\beta}_{\text{ols}}\|_0 = p$.

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

$$ \hat{\beta}_{\text{Lasso}}(\lambda) = \arg\min_{\beta} \left( \frac{\|Y - X \beta\|_2^2}{n} + \lambda \|\beta\|_1 \right), $$  (3)

where $\lambda \geq 0$ is a penalty parameter and $\|\beta\|_1 = \sum_{j} |\beta_j|$. The penalty term $\lambda$ used in (3) is determined by cross-validation (CV) or classical selection criteria; typically five-fold or ten-fold CV or the Bayesian information criterion (BIC). If the neighborhood stability and the beta-min condition are fulfilled, the Lasso is a consistent variable selection method since $\lim_n P(\hat{A}_{\text{Lasso}} = A) \rightarrow 1$ and, hence, $\|\hat{\beta}_{\text{Lasso}}\|_0 = q$.

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 $\beta$ estimate. Algebraically, we can express the estimator as

$$ \hat{\beta}_{\text{adapt}}(\lambda) = \arg\min_{\beta} \left( \frac{\|Y - X \beta\|_2^2}{n} + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\hat{\beta}_{\text{init},j}|} \right), $$  (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 use again CV to determine $\lambda$ at the second stage. The use of the Lasso as a first stage estimator is justified by the screening property of the Lasso ($P(\hat{A}_{\text{Lasso}} \supseteq A) \rightarrow 1$, where $\hat{A}_{\text{Lasso}}$ is the estimated set of active variables by the Lasso), which still allows consistent variable selection of the adaptive Lasso at the second stage ($P(\hat{A}_{\text{adapt}} = A) \rightarrow 1$).

The computational cost of the adaptive Lasso solution is of the same order as 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 $\lambda$ 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: \( \lim_{n} P(\hat{A}_{\text{adapt},n} = A) = 1 \), where \( A \) is as defined above.

2. Optimal estimation rate (similar to the OLS estimator): \( \sqrt{N}(\hat{\beta}_{\text{adapt},A} - \beta_A) \xrightarrow{d} N(0, \Sigma_A) \), with \( \Sigma_A \) the covariance matrix of the true subset.

Following Hastie et al. (2009) and assuming that we have a functional relation of \( Y \) and \( X \) as described in (1), we can express the OLS bias-variance decomposition for squared error loss as follows:

\[
Err(x_0) = E[(Y - X'\hat{\beta})^2 | X = x_0] = \underbrace{\sigma^2} + \underbrace{\frac{1}{T} \sum_{t=1}^{T} (x'_t(\hat{\beta} - \beta))^2}_{\text{Bias}^2} + \underbrace{p \sigma^2}_{\text{Variance}}
\]

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 \( \sigma^2 \) 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 5 reflects the motivation of 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 200-300 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.

4 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 2.2 on what has already been proved theoretically. Hence, the simulation study includes several specifications to analyze the method’s sensitivity to these assumptions.
4.1 Set-up

Generally, we assume that the following properties potentially apply to the return-generating process of cross-sectional stock returns and which we also use as guidance to establish the base case simulation:

- Cross-sectional returns follow a factor structure.

- Only a small fraction of factors carry a risk-premium. Moreover, we assume that the factors are independent of each other and that the stock market factor explains the highest proportion of variance of all factors.

- Firm characteristics fall in one of the following three categories/groups: Categories 1 and 2 both measure factor exposure, with the difference being that only factors of category 1 carry a positive risk-premium. 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 first letter of the company name, (presumably) irrelevant for the DGP. Furthermore, we assume that category 1 marks the smallest fraction of FC, followed by category 2, and category 3 marks by assumption the largest fraction in the set of FC. FC are potentially correlated across and within the groups.

- Additionally, to the factor structure, each stock has its own idiosyncratic component, orthogonal to other factors and other stocks’ idiosyncratic component.

- 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 $R^2$s).

- 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 equations:

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

- $f_t$ defines the vector of factor returns of length $C$. Each $f_t$ is determined by: $f_{t+1} = \mu + \epsilon_{t+1}$ with $\epsilon_{t+1,c} \sim \mathcal{N}(0,\sigma_f^2)$ $\forall c \in \mathcal{P} \cup \mathcal{R} \land c \neq 1$ and $\epsilon_{t+1,1} \sim \mathcal{N}(0,\sigma_{t,1}^2)$. FC, $c = 1$, represents the stock market return, which follows a latent volatility process $\sigma_{t,1}^2$ (see next item). $\epsilon_{t+1,c}$ is set to 0 $\forall c \in \mathcal{U}$ (uninformative characteristics with no return dependence). The elements of the vector of risk-premia, $\mu_c$, are drawn from $\sim \text{unif}(0.1,3.5)\forall c \in \mathcal{P} \land c \neq 1$ and are set equal to zero $\forall c \in \mathcal{R} \cup \mathcal{U}$. We assume a stock market premium, $\mu_1$, of 5.5% per year.  

---

8 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 between 1962-2014.
The parameters estimated with normal distributed errors are: $\hat{\alpha} = 0.0079$, $\hat{\beta}_0 = 0.093$ and $\hat{\beta}_1 = 0.9019$. And for the t-distributed process: $\hat{\alpha} = 0.006$, $\hat{\beta}_0 = 0.084$, $\hat{\beta}_1 = 0.914$ and $\nu = 7.14$. The models are 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.

- $\sigma_{t,1}^2$, 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.\(^9\) Figure 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 $\sim \mathcal{N}(0, \sigma_{idio}^2)$.

- $x_{t,n}$ marks the vector of firm characteristics of stock $n$ at $t$ of length $C$. We simulate the characteristics $x_t$ from $\sim \mathcal{N}(0, \Sigma)$.

- The correlation matrix of FC, $\Sigma$, is obtained following the simulation approach of Hardin et al. (2013). The base case refers to the constant correlation structure within groups (algorithm 1) in Hardin et al. (2013). The $\Sigma$ is drawn only initially and kept constant in each specification. The empirical correlation structure described in Section 6.1 shows a handful of cases with pair-wise 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 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 $\mu$, which is estimated as the vector

\(^9\)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)).
of coefficients $\hat{\beta}$.

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

- Number of simulations, 500
- Number of periods, 600, corresponding to 50 years of monthly data
- Number of stocks, 4800 $^{10}$
- Number of firm characteristics, 100 – with $P = 6$, $R = 6$ and $U = 88$
- $\sigma^2_f$ and $\sigma^2_{idio}$ result from pre-specified level of $R^2$, where the noise variance is distributed as described in the appendix.
- The correlation matrix, $\Sigma$, is simulated such that we have one high (0.9) and one low (0.5) pairwise correlation between FC from each group. See Figure 2 for a visualization of one realization of the specified correlation matrix simulation.

The following five cases are defined:

Case 1: Approximation of Return Data
Default settings.

Case 2: Time-constant stock market volatility
Default settings, except for the assumption of the underlying latent volatility process of $\hat{\sigma}^2_{t,1}$, 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 $\epsilon_{t+1,1} \sim t$ - 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 - Figure 1 shows the estimated process.

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 500 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.

$^{10}$4800 corresponds to the average number of stocks used in the empirical part in an earlier version of this paper, which compares to roughly 4000 stocks used in this version.
Figure 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 FCs are uninteresting insofar as their off-diagonal elements are close to zero.
Note that each simulation considers a balanced panel. As the actual data consists of an unbalanced panel, we adjust the data as described in Section 6.1. 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.

### 4.2 Performance evaluation

After each data simulation is performed, we run three regressions covering all three methods presented above in the pooled panel framework. We collect the results of the OLS estimates, the Lasso and the adaptive Lasso. Unfortunately, a more formal performance comparison is not available. Hence, we discuss the results on an informal basis. 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 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.

### 4.3 Results

#### Case 1: Default settings

The simulation results of case 1 are exhibited in Figure 3 and in Table 1. The results reveal that there are distinct differences between the three 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 reveal a far better performance with an error rate of 1.4% and 2.8% respectively for the stock market factor. For the other, significantly less volatile factors carrying a positive risk-premium, the type II error rates are indistinguishable from zero for all cases and all three methods. The type I error cases have to be distinguished in two cases: First, FC belonging to set $\mathcal{R}$, where OLS has slight advantages over the adaptive Lasso in particular for the case where the correlation is assumed to be $0.9$ (FC #6). This case appears to be an issue especially for the Lasso as measured by an error rate of well above 80%. Generally, it is notable that the Lasso performs extremely weak for cases with FC being part of set $\mathcal{R}$. The second case of type I errors comprises uninformative 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 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 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 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 $\mathcal{R}$, where we observe a jump in error ratios for the adaptive Lasso.

![Figure 3: Type I and Type II Error Rate; Case: "default"](image_url)

**Figure 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 error. The first sub-figure shows the results of the OLS selection procedure. We can see an elevated level for type II errors for the identification of the positive market premia (FC 0). The other return premia relevant to FC 1-5 are mostly correctly classified as active. FC 6-30 exhibit the expected behavior with an error ratio distributed around the 5%-level. The chart in the middle illustrates the adaptive Lasso error ratios. Mostly the error ratios are low, the only exception being the FC belonging to set $\mathcal{R}$. The subplot on the bottom reveals that the Lasso has the most elevated type II error ratios for FC 6-11, and additionally, classification issues for cases with significant correlations (FC 12-15). Type II errors on the other are lowest compared to the other two classifiers.
Table 1: The Table provides an overview of type II (FC 0-5) and type I error (6-99) ratio behavior in percentage points for the specified simulation cases. FC cases 12-15 are interesting in so far, as they are uninformative FC with higher correlations with FC of set P and R.

**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 98% and that of FC 12 reaches 57%. Moreover, higher error ratios are also observed for cases with weaker correlations of around .5; see FC cases 7-9 and 12-15.

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 pictures 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 the risk dimension 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).

**5 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 follow generally 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 were unavailable. Compustat data is aligned with a standard lag of six months of the fiscal year end date.\(^\text{11}\) 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. 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. Therefore, each FC is winsorized at the 5% 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 \( \beta \) 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. 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 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 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 \( > 2000 \).

\(^{12}\)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.

\(^{13}\)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.

\(^{14}\)Ang et al. (2016) provide a detailed econometric discussion of using stocks vs. portfolios in empirical asset pricing.
Figure 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.

6 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 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 prospective. The final subsection 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.

6.1 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.

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 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
Figure 5: The figure exhibits the correlation of all firm characteristics included in our analysis for the years 1964-2014. We use the normalized, winsorized and pooled FC data (as used in the entire 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.

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 5 presents the correlation structure of the FC. It shows overall only five cases with absolute correlations 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 4, 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.

6.2 Out-of-sample evidence

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 factor performance for each of the two methods. The respective Adaptive Lasso and OLS factors are obtained by proceeding in the following way. The
relevant FC of the adaptive Lasso factor 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.\textsuperscript{15} The return predictions can then be sorted from highest to lowest at each point in time and hence the classical portfolio/factor return analysis can be applied. The factor 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 factor, where we simply replace the FC selection with the OLS selection procedure. These factor returns can then be used to test differences in Sharpe ratios of the two methods. Specifically, we use the Jobson and Korkie (1981) test including the Memmel (2003) correction:

$$z = \frac{\text{SR}_1 - \text{SR}_2}{\sqrt{\frac{1}{T} \left( 2(1 - \rho_{1,2}) + \frac{1}{2}(\text{SR}_1^2 - \text{SR}_2^2 - \text{SR}_1\text{SR}_2(1 + \rho_{1,2})) \right)}}$$

where, \text{SR}_1 corresponds to the Sharpe Ratio of the Adaptive Lasso factor portfolio, and \text{SR}_2 is associated with the OLS factor portfolio.

Factor return series and consequently the inference based on these returns are often sensitive with respect to the choice parameters. 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 (market cap vs. equally weighted) and the factor cutoff (decile vs. Fama and French (1993) style 30/40/30). As these choices impact the results in a significant manner, we exhaust all possible combinations. Table 2 shows the test statistics of these results for factors constructed based on all stocks, including only large, mid and small cap stocks, respectively. Column ”all” reveals that the adaptive Lasso selection performs in a large fraction of the cases significantly better than an OLS-based selection. In 12 out of 16 cases we observe a significantly higher Sharpe Ratio at the 1%-level with the remaining four cases showing no difference between the two methods. The picture remains similar when considering only large cap stocks; however, we notice more cases with no significant difference. In total we have nine cases where the adaptive Lasso returns a significantly better Sharpe ratio vs. only one case favoring OLS. Moreover, in case of mid and small cap stocks, we can see that the results are more balanced. Considering all cases with OLS estimation (64 combinations), we find that in 66% of the cases, the adaptive Lasso factor achieves a higher SR than the OLS counterpart. In more detail, 45% of the differences favor significantly (at the 5%-level) adaptive Lasso over OLS; OLS only reaches a level of 14% in comparison. Overall, we can summarize that our out-of-sample analysis reveals that the performance of the adaptive Lasso factor dominates the OLS factor measured in terms of Sharpe Ratios. Furthermore, using adaptive Lasso coefficient estimates has only a minor impact compared to the OLS coefficient estimation approach; for more details we refer to the appendix.

\textsuperscript{15}For completion, we provide additional tests using the Adaptive Lasso coefficient estimates for the return prediction in the appendix - the overall performance differences are negligible.
Table 2: Sharpe Ratio Tests - OLS coef estimator, i.e. using OLS for coefficient estimation. The columns all, large, mid and small represent the test-statistics for differences in Sharpe Ratios using the Jobson and Korkie (1981) test including the Memmel (2003) correction. A negative test statistic indicates a higher SR ratio for the adaptive Lasso factor compared to the OLS factor. The table displays the results conditioned on the four cases we consider: for each case we have 16 combinations, which results in a total of 64 combinations. Considering only the results of all stocks in the analysis, we see that in 81.25% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 75.00% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 0.00% of the cases. No significant difference is observed in 25.00% of the cases. The column of the large cap stocks in the analysis shows that in 87.50% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 56.25% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 6.25% of the cases. No significant difference is observed in 37.50% of the cases. The results conditioned on only mid cap stocks reveals that in 31.25% of the cases we observe a significantly higher Sharpe ratio at the 5% level in 56.25% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 31.25% of the cases. No significant difference is observed in 56.25% of the cases. Finally, the results with respect to the small caps stocks show that in 62.50% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 37.50% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 18.75% of the cases. No significant difference is observed in 43.75% of the cases. Over all 64 combinations we can summarize that in 65.62% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 45.31% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 14.06% of the cases. No significant difference is observed in 40.62% of the cases.
6.3 Shrinking the zoo of firm characteristics

The analysis considers the years from 1965 to 2014 and primarily emphasizes the selection regression including all stocks. These findings are shown in Table 3. We separately look at the regression conditioning on large, mid and small cap sized stocks. These results can be found in Tables 4 to 6. All four tables display the sets of active FC determined by the adaptive Lasso, OLS and Lasso selection. The discussion in this section stresses mostly the details of the FC selection for the adaptive Lasso and the differences with OLS. Mostly, 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.

6.3.1 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 6.1. Table 3 embodies the FC selection results for the adaptive Lasso, OLS and Lasso based on all sample periods. We find that the adaptive Lasso selects three 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 identifyagr(−) – investment; bm(+) – value; andgma(+) – profitability as part of the set of active FC. The FC not directly represented aremve and the beta, reflecting the size and market factor, respectively. However, the adaptive Lasso selection comprises mve_ia(−) – an industry-adjusted variant of the classical size FC mve. In contrast to the five-factor model, we measure 14 FC characterizing differences in expected cross-sectional returns. Four out of these 14 FC are based exclusively on price information. This includes the most relevant FC, measured by the absolute size of the coefficient, mom1m(−), reflecting a strong short-term reversal effect in cross-sectional returns. Moreover, the adaptive Lasso selects mom12m(+) – the classical twelve-months momentum; chg_mom6m(−) – difference in the six months momentum or a reversion to six months lagged six months trend; andmaxret(−) – the max daily return of the previous month. Additionally, the following six FC are part of the selection: rd_mve(+) – R&D spending relative to market value; chpm_ia(+) – industry adjusted (IA) change in profit margin; invest(−) – change in investment relative to total assets; pchdepr(−) – change in depreciations divided by gross property, plant and equipment; pchcapx_ia(−) – IA changes in capital expenditures; and hire(−) – relative change in employees. The latter three are the least relevant measured by the absolute size of the coefficient. Table 3 expresses the differences between the two (three) selection tools and underscores the relevance of this choice. Specifically, we observe that eight FC are jointly selected, whereas the OLS-based selection picks four different FC not appearing in the FC set of the adaptive Lasso. The table reveals that a Lasso-based procedure would suggest a higher dimensional relation between FC and returns. We discard 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 most FC inspected do not contain

---

16The beta is part of the FF three factor model, despite insignificant regression coefficients; see Fama and French (1992, 1993) for details.
Table 3: FC Selection All Stocks: The table exhibits the regression results of OLS, Lasso, and adaptive Lasso for each FC, sorted by the absolute value in the order of the adaptive Lasso, 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, chato_ia, cash, grcapex, absacc, salecash, divo, pchcurrat, pchsale, beta, beta_sq, bm_ia, pctacc, pchquick, dolvol, dy, idiovol, saleinv, lgr, chinu, retvol, pchsale_pchxsga, lev, salerc, cp, pchsale_pchinv, rd, egr, src, rd_sale, mom6m, divi, cfp. 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.

<table>
<thead>
<tr>
<th>FC</th>
<th>AdaLasso</th>
<th>OLS</th>
<th>Lasso</th>
<th>FC</th>
<th>AdaLasso</th>
<th>OLS</th>
<th>Lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>mom1m</td>
<td>-6.33</td>
<td>-6.14***</td>
<td>-5.62</td>
<td>mve</td>
<td>0.00</td>
<td>-2.08</td>
<td>-1.21</td>
</tr>
<tr>
<td>agr</td>
<td>-2.80</td>
<td>-2.02**</td>
<td>-2.23</td>
<td>acc</td>
<td>0.00</td>
<td>-1.41*</td>
<td>-1.90</td>
</tr>
<tr>
<td>mve_ia</td>
<td>-2.31</td>
<td>-1.93(-1.64)</td>
<td>-1.49</td>
<td>sp</td>
<td>0.00</td>
<td>0.01(0.02)</td>
<td>0.30</td>
</tr>
<tr>
<td>rd_mve</td>
<td>2.28</td>
<td>2.36***</td>
<td>4.04</td>
<td>cashpr</td>
<td>0.00</td>
<td>-0.95**</td>
<td>-2.35</td>
</tr>
<tr>
<td>mom12m</td>
<td>2.18</td>
<td>2.45*</td>
<td>1.94</td>
<td>pchgm_pchsale</td>
<td>0.00</td>
<td>0.68* (1.86)</td>
<td>0.24</td>
</tr>
<tr>
<td>bm</td>
<td>1.87</td>
<td>2.09**</td>
<td>2.02</td>
<td>cashdeb</td>
<td>0.00</td>
<td>1.42** (2.67)</td>
<td>0.23</td>
</tr>
<tr>
<td>gma</td>
<td>1.21</td>
<td>1.27**</td>
<td>2.53</td>
<td>turn</td>
<td>0.00</td>
<td>-0.92 (-1.26)</td>
<td>-0.19</td>
</tr>
<tr>
<td>chg_mom6m</td>
<td>-1.16</td>
<td>-1.34 (-1.25)</td>
<td>-0.79</td>
<td>pchsaleinv</td>
<td>0.00</td>
<td>0.74 (0.97)</td>
<td>0.16</td>
</tr>
<tr>
<td>maxret</td>
<td>-0.74</td>
<td>-1.50 (-1.22)</td>
<td>-0.50</td>
<td>mom36m</td>
<td>0.00</td>
<td>-0.46 (-0.78)</td>
<td>-0.16</td>
</tr>
<tr>
<td>chpm_ia</td>
<td>0.73</td>
<td>0.96**</td>
<td>2.46</td>
<td>tang</td>
<td>0.00</td>
<td>0.23 (0.48)</td>
<td>0.06</td>
</tr>
<tr>
<td>invest</td>
<td>-0.63</td>
<td>-0.55 (-0.97)</td>
<td>-0.62</td>
<td>age</td>
<td>0.00</td>
<td>1.02*** (2.86)</td>
<td>0.00</td>
</tr>
<tr>
<td>pchdepr</td>
<td>-0.50</td>
<td>-1.21**</td>
<td>-3.22</td>
<td>pchrm</td>
<td>0.00</td>
<td>-1.47 (-1.47)</td>
<td>-0.15</td>
</tr>
<tr>
<td>pchcapx_ia</td>
<td>-0.33</td>
<td>-0.99***  (-3.03)</td>
<td>-0.36</td>
<td>currat</td>
<td>0.00</td>
<td>-2.15*** (-2.62)</td>
<td>0.00</td>
</tr>
<tr>
<td>hire</td>
<td>-0.26</td>
<td>-0.80*</td>
<td>-1.87</td>
<td>depr</td>
<td>0.00</td>
<td>0.76* (1.83)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Further insights into the selection 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 6. Overall the results are in line with the full sample selection. We notice that the two most often selected FC over time are the price-related mom1m and mom12m. However, the signal strength of the two FC weakens with time, with the effect particularly pronounced with the latter. Effects of size and value are visible through all points in time, but are each not represented by a unique FC. Specifically, we observe that bm and bm_ia as well as dolvol – dollar trading volume and mve_ia complement each other over time. Furthermore, we notice that two additionally important FC over time are agr and rd_mve, especially because of their strengthening signal towards the more recent periods after only scattered initial consideration. The selection of chg_mom6m, gma, invest in the full period selection is 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, does not appear at all in the rolling window selection, indicating it plays only a marginal role in the return-forming process.

6.3.2 FC selection conditioned on size

The selection results conditioned on large cap stocks only are depicted in Table 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 a higher dimensional relationship of FC and returns compared to results including all stocks. Strikingly, price-based FC dominate the overall selection as the top four coefficients can purely be linked
Figure 6: Rolling adaptive Lasso selection all stocks: 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: \textit{mom1m}(100.0\%), \textit{mom12m}(83.3\%), \textit{rd}_mve(77.8\%), \textit{agr}(59.3\%), \textit{bm}(59.3\%) and \textit{chg}_mom6m(58.1\%).
to past price information. Noteworthy is the selection of \( \text{retvol}(-) \), \( \text{dolvol}(-) \) and \( \text{ep}(+) \), FC representing low-volatility, size and value dimensions. Moreover, OLS selects a subset of the adaptive Lasso FC with less than half as many members as the competing method.

Considering only \textbf{mid cap stocks}, we find similarly to the large cap case an above threshold correlation for the \( \text{beta} \) and \( \text{beta\textunderscore sq} \) correlation pair, which leads to discarding \( \text{beta\textunderscore sq} \).\(^{17}\) Table 5 presents the active set of FC conditioned on mid sized stocks. The pattern repeats, price-based FC rank highest, with \( \text{mom12m}(+) \) on top and \( \text{retvol}(-) \) second. Furthermore, \( \text{agr}(-) \), \( \text{rd\textunderscore mve}(+) \), \( \text{cashdebt}(+) \), \( \text{mom1m}(-) \), \( \text{sp}(+) \), \( \text{beta}(+) \), \( \text{maxret}(-) \), \( \text{currat}(-) \) and \( \text{cfp}(+) \) complete the characterization of mid cap returns from an FC perspective. Interestingly, an OLS-based selection would change the picture in this case significantly, as OLS includes many more FC and not all those picked by the adaptive Lasso.

Finally, we briefly discuss the results including exclusively \textbf{small cap stocks}. The correlation prescreening 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 6, shows that the dominance of price information-driven FC is not as strong as in the previous three selection regressions but still present. We find consistency for which FC occupy the top ranks, considering the magnitude of short-term reversal. The five highest rank FC are: \( \text{mom1m}(-) \), \( \text{mve\textunderscore ia}(-) \), \( \text{rd\textunderscore mve}(+) \), \( \text{mve}(-) \) and \( \text{bm}(+) \). OLS and adaptive

\(^{17}\)It is not surprising that \( \text{beta} \) and \( \text{beta\textunderscore sq} \) 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.

<table>
<thead>
<tr>
<th>FC</th>
<th>AdaLasso</th>
<th>OLS</th>
<th>Lasso</th>
<th>FC</th>
<th>AdaLasso</th>
<th>OLS</th>
<th>Lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>chg\textunderscore mom6m</td>
<td>-2.33</td>
<td>-2.98***</td>
<td>-2.68</td>
<td>cash</td>
<td>0.00</td>
<td>0.27 (0.44)</td>
<td>0.32</td>
</tr>
<tr>
<td>mom12m</td>
<td>1.72</td>
<td>1.15 (1.09)</td>
<td>1.74</td>
<td>cphm\textunderscore ia</td>
<td>0.00</td>
<td>0.46 (1.41)</td>
<td>0.32</td>
</tr>
<tr>
<td>mom1m</td>
<td>-1.53</td>
<td>-1.85***</td>
<td>-2.84</td>
<td>salerec</td>
<td>0.00</td>
<td>0.52 (1.40)</td>
<td>0.30</td>
</tr>
<tr>
<td>mom6m</td>
<td>1.37</td>
<td>2.21* (1.73)</td>
<td>1.15</td>
<td>dy</td>
<td>0.00</td>
<td>-0.83* (-1.75)</td>
<td>-0.30</td>
</tr>
<tr>
<td>egr</td>
<td>-1.32</td>
<td>-1.26***</td>
<td>-4.06</td>
<td>agr</td>
<td>0.00</td>
<td>-0.45 (-0.75)</td>
<td>-0.25</td>
</tr>
<tr>
<td>rd\textunderscore mve</td>
<td>1.19</td>
<td>1.30* (1.86)</td>
<td>1.12</td>
<td>saleinv</td>
<td>0.00</td>
<td>0.22 (0.93)</td>
<td>0.23</td>
</tr>
<tr>
<td>retvol</td>
<td>-0.95</td>
<td>-1.72***</td>
<td>-2.80</td>
<td>mom36m</td>
<td>0.00</td>
<td>-0.31 (-0.62)</td>
<td>-0.23</td>
</tr>
<tr>
<td>ep</td>
<td>0.93</td>
<td>0.99** (2.03)</td>
<td>0.95</td>
<td>grcapex</td>
<td>0.00</td>
<td>-0.38* (-1.72)</td>
<td>-0.22</td>
</tr>
<tr>
<td>sp</td>
<td>0.84</td>
<td>0.68* (1.73)</td>
<td>0.63</td>
<td>acc</td>
<td>0.00</td>
<td>-0.44 (-0.84)</td>
<td>-0.19</td>
</tr>
<tr>
<td>cashdebt</td>
<td>0.83</td>
<td>1.26* (1.86)</td>
<td>0.88</td>
<td>pchdepr</td>
<td>0.00</td>
<td>0.35 (1.53)</td>
<td>0.18</td>
</tr>
<tr>
<td>dolvol</td>
<td>-0.68</td>
<td>-1.02 (-1.53)</td>
<td>-0.71</td>
<td>divi</td>
<td>0.00</td>
<td>-0.27* (-1.75)</td>
<td>-0.14</td>
</tr>
<tr>
<td>currat</td>
<td>-0.47</td>
<td>-1.84** (-2.27)</td>
<td>-0.75</td>
<td>pchs\textunderscore pchxsga</td>
<td>0.00</td>
<td>-0.29 (-1.06)</td>
<td>-0.14</td>
</tr>
<tr>
<td>pchs\textunderscore pchxsga</td>
<td>0.26</td>
<td>0.59 (0.79)</td>
<td>0.38</td>
<td>chato\textunderscore ia</td>
<td>0.00</td>
<td>0.22 (0.78)</td>
<td>0.09</td>
</tr>
<tr>
<td>cashpr</td>
<td>-0.14</td>
<td>-0.69* (-1.77)</td>
<td>-0.47</td>
<td>lev</td>
<td>0.00</td>
<td>0.35 (0.62)</td>
<td>0.05</td>
</tr>
<tr>
<td>depr</td>
<td>0.08</td>
<td>0.43 (1.36)</td>
<td>0.33</td>
<td>cfp</td>
<td>0.00</td>
<td>0.06 (0.10)</td>
<td>0.03</td>
</tr>
<tr>
<td>gma</td>
<td>0.00</td>
<td>0.69 (1.63)</td>
<td>0.39</td>
<td>chinv</td>
<td>0.00</td>
<td>-0.10 (-0.27)</td>
<td>-0.02</td>
</tr>
<tr>
<td>bm</td>
<td>0.00</td>
<td>0.92 (1.42)</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textbf{Table 4: FC Selection Large Stocks}: The table exhibits the regression results of OLS, Lasso, and adaptive Lasso for each FC, sorted by the absolute value in the order of the adaptive Lasso, 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: \( \text{quick} \), \( \text{maxret} \), \( \text{bm\textunderscore ia} \), \( \text{invest} \), \( \text{age} \), \( \text{pchquick} \), \( \text{divo} \), \( \text{idiovol} \), \( \text{pchcurrat} \), \( \text{chshrout} \), \( \text{pctace} \), \( \text{rd\textunderscore sale\textunderscore cash} \), \( \text{turn} \), \( \text{sgr} \), \( \text{hire} \), \( \text{lgr} \), \( \text{pechcapx\textunderscore ia} \), \( \text{absacc} \), \( \text{rd\textunderscore sale} \), \( \text{beta} \), \( \text{pchsale\textunderscore pchrect} \), \( \text{pchsale\textunderscore in} \), \( \text{mve} \), \( \text{pchgm\textunderscore pchsale} \), \( \text{mve\textunderscore ia} \), \( \text{tang} \). 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.
Lasso deviate but less so in this case compared to the FC selection of large and mid cap stocks.

6.3.3 Condensing the results jointly

Comparing Tables 3 and 4-6 allows us to understand the main results in greater detail. First, we see that the intersection of all four active adaptive Lasso FC sets consists (always with the same sign) of

Table 5: FC Selection Mid Stocks: The table exhibits the regression results of OLS, Lasso, and adaptive Lasso for each FC, sorted by the absolute value in the order of the adaptive Lasso, 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: pchquick, cashpr, pchsale_pchrect, pchgm_pchsale, quick, gma, rd, acc, chg_mom6m, mom6m, sgr, salecash, chshrt, pchsale_pchxsga, pchcapx_ia, hire, pchdepr, grcepx, rd_sale, tang, turn, saleinv, pchcurrat, chato_ia, mve_ia, divi, bm, absacc, cash, dolvol, ep, egr, mom36m, mve, divo, bm_ia, pctacc. 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.

Table 6: FC Selection Small Stocks: The table exhibits the regression results of OLS, Lasso, and adaptive Lasso for each FC, sorted by the absolute value in the order of the adaptive Lasso, 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, pchquick, lev, cash, pctsacc, acc, cashpr, grcepx, turn, chato_ia, pchsale_pchrect, salecash, tang, dolvol, pchgm_pchsale, egr, rd_sale, ep, dy, hire, salerec, mom6m, divo, rd, pchsaleinv, depr, saleinv, chg_mom6m, lgr, chinv, beta, pchsale_pchxsga, pchcapx_ia, hire, pchdepr, beta_sq, absacc, sgr, pchsale_pchxsga. 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.
mom1m, rd_mve and mom12m. This across-size consistency provides strong evidence for short-term price reversals, a positive return benefit from R&D spending and price continuation measured over a twelve month period. In comparison, the OLS procedure yields mom1m and currat. However, we stress that a value dimension should not be simply neglected from this perspective as it is captured through different FC in each sub-selection. To be precise, we identify $bm$, $sp$ and $ep$, and $bm$ in the selection results including all, only large cap, mid and small caps 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, the negative relation between the stock return volatility and the returns is only observed for mid and large cap stocks.\footnote{This is consistent with the literature as the low volatility effect fades when measured in an equal weighting setting (and thus implicitly giving more weight to small cap stocks); see, for example, Bali and Cakici (2008).}

The final remark is an interesting but expected side result. We observe that the sum of all coefficients over the absolute value of the adaptive Lasso increases from large (14.63), mid (15.01) to small (28.5) cap stocks. This indicates a more efficiently priced market for large cap stocks, at least before any transaction costs are accounted for. A more in-depth analysis of this observation and an explicit correction for trading costs is beyond the scope of this work.\footnote{Implicit and explicit transaction costs are considered to be higher for smaller stocks.}

## 7 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 not selected when considered in a multivariate analysis simultaneously; we identify 14 FC of relevance. Among the most consistently selected FC are price-related measures. Particularly outstanding are the one-month momentum, well known as the short-term reversal, and the standard twelve-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 find four out of five Fama and French (2014) factors at least partially identified in the adpative Lasso selection. The market factor is the only factor not represented here.

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 posses explanatory power for differences in average returns, but nonetheless impact the risk dimension of the returns.
We find additional support for an adaptive Lasso-based selection vs. an OLS procedure based on an empirical out-of-sample test. We construct two competing factors for each of the two methods and compare their portfolio performance in terms of Sharpe ratios. The findings indicate the adaptive Lasso’s superior selection abilities.
8 Bibliography


28


A Appendix

A.1 Lasso Conditions

The beta-min condition reads,

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

where $\phi^2$ is the restricted eigenvalue of matrix $X$.

Moreover, the neighborhood stability condition is,

$$\|\hat{\Sigma}_{A^c \setminus A} \hat{\Sigma}_{A}^{-1} \text{sign}(\beta_1, \ldots, \beta_q)\|_{\infty} \leq \theta \text{ for some } 0 < \theta < 1,$$

where $\|z\|_{\infty} = \max_j |z_j|$ and, $\hat{\Sigma}_{A^c \setminus A}$ and $\hat{\Sigma}_{A}$ are the lower and upper left part of the covariance matrix of $X$. 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.2 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:

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

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

$$SNR = \frac{R^2}{1 - R^2} \quad (7)$$

Recalling equation (6) and ignoring the indices, we can write,

$$r = x^T f + \eta,$$

with $f = \mu + \epsilon$

Hence, we can define the variance of the signal as:

$$\sigma^2_{signal} = \text{Var}(x^T \mu) = \sigma_x^2 \mu^T \mu$$

Note that the $\mu$ is defined as a uniformly distributed random vector and the realized $\mu$ 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:

\[ \sigma_{\text{noise}}^2 = \text{Var}(x' \epsilon) + \text{Var}(\eta) \]

\[ = \sigma_x^2 \sigma_\epsilon^2 + \sigma_\eta^2 \]

\[ = \sum c \sigma_{\epsilon,c}^2 \sigma_x^2 + \sigma_\eta^2 \]

\[ = \sigma_{\epsilon,1}^2 \sigma_x^2 + (P + R - 1) \sigma_\epsilon^2 \sigma_x^2 + \sigma_\eta^2 \]

The first line can be simplified according to equation (8) 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. \( \sigma_\epsilon \) and \( \sigma_\eta \) are calibrated such that each part contributes equally to fit the desired signal-to-noise ratio (\( \sigma_\eta^2 = (P + R - 1) \sigma_\epsilon^2 \)). The value of \( \sigma_\eta^2 \) and \( \sigma_\epsilon^2 \) 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:

\[ \text{Var}(XY) = E[X^2Y^2] - [E(XY)]^2 \]

\[ = \text{Cov}(X^2, Y^2) + E(X^2)E(Y^2) - [E(XY)]^2 \]

\[ = \text{Cov}(X^2, Y^2) + (\text{Var}(X) + [E(X)]^2)(\text{Var}(Y) + [E(Y)]^2) - [\text{Cov}(X, Y) + E(X)E(Y)]^2 \] (8)

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 \( \sigma_{\text{noise}}^2 \) are treated as returns with zero auto-correlation:

\[ \sigma_{\text{signal},T}^2 = \text{Var}(x' \mu) = \sigma_x^2 \mu' \mu \]

\[ \sigma_{\text{signal},\frac{T}{\tau}}^2 = \text{Var}(x' \frac{\mu}{\tau}) = \sigma_x^2 \frac{\mu' \mu}{\tau^2} \]

\[ \sigma_{\text{noise},\frac{T}{\tau}}^2 = \frac{1}{\tau^2} \sigma_{\text{noise},T}^2 \]

and hence,

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal},T}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{\mu^2 \sigma_x^2}{\tau \sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal},T}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

\[ \frac{\text{SNR}_T}{\text{SNR}_{\frac{T}{\tau}}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{signal},\frac{T}{\tau}}^2} \cdot \frac{\sigma_{\text{noise},\frac{T}{\tau}}^2}{\sigma_{\text{noise},T}^2} = \frac{1}{\tau} \]

Combining equations (9) and (7), the following relation holds:

\[ \frac{R^2_{\frac{T}{\tau}}}{1 - R^2_{\frac{T}{\tau}}} = \frac{R^2_T}{1 - R^2_T} \] (10)
A.3 Firm Characteristics

As described above, the data of the paper utilizes 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 is assigned to particular CRSP return series. We construct the single criteria as described in Tables 8 and 9. 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.4 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 30GB of available RAM are required.

A.5 Additional Figures
Figure 7: 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 calendar point) at the 5% for better visualization, as outliers potentially lead to scaling problems.
Figure 8: Rolling adaptive Lasso selection large cap stocks: 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: mom12m(99.28%), mom1m(81.62%), chg_mom6m(76.13%), egr(65.63%), ep(58.00%) and rd_mve(53.22%). The rolling window analysis displayed shows consistent with previous findings in Table 4 that mom12m and mom1m are strong drivers behind returns. Additionally, we observe frequent occurrences (>50%) of the following FC: chg_mom6m, egr, ep, rd_mve, agr and retvol.
Figure 9: Rolling adaptive Lasso selection mid cap stocks: 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: agr(81.7%), mom12m(78.8%), retvol(76.5%), cashdebt(75.8%), rd_mve(73.2%), mom1m(68.3%). Overall, we observe consistent results over time compared with the full period selection regression; see Table 5 for details.
Figure 10: Rolling adaptive Lasso selection small cap stocks: 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: $mom_{1m}(100.0\%)$, $mve_{ia}(93.8\%)$, $rd_{mve}(82.0\%)$, $invest(63.4\%)$, $mom_{36m}(54.9\%)$, $cashpr(46.1\%)$. The rolling selection mirrors the picture measured over all periods in Table 6, as the four FC described are most frequently chosen.
## A.6 Additional Tables

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

**Table 7:** Overview of raw data used and the abbreviations with the corresponding description and data source.
<table>
<thead>
<tr>
<th>ID</th>
<th>Acronym</th>
<th>Name</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>beta_sq</td>
<td>Beta squared</td>
<td>Simply obtained by squaring the $\beta$ based on the beta from # 1</td>
<td>Fama and MacBeth (1973)</td>
</tr>
<tr>
<td>2</td>
<td>beta_sq</td>
<td>Beta squared</td>
<td>Measured based on 3 years (min 52 weeks) weekly excess returns with standard ols ($y = c + \beta x$)</td>
<td>Fama and MacBeth (1973)</td>
</tr>
<tr>
<td>3</td>
<td>retvol</td>
<td>Volatility</td>
<td>Volatility is measured by the standard deviation of daily returns of the previous months</td>
<td>Ang et al. (2006)</td>
</tr>
<tr>
<td>4</td>
<td>maxret</td>
<td>Maximum return</td>
<td>Maximum return is defined over the max of the daily returns in month $t-1$</td>
<td>Bali et al. (2011)</td>
</tr>
<tr>
<td>5</td>
<td>ldivo</td>
<td>Idiosyncratic volatility</td>
<td>Calculated based on the residuals of regression in # 1</td>
<td>Ali et al. (2003)</td>
</tr>
<tr>
<td>6</td>
<td>mom1m</td>
<td>1-month momentum</td>
<td>Return in month $t-1$</td>
<td>Jegadeesh (1990)</td>
</tr>
<tr>
<td>7</td>
<td>mom6m</td>
<td>6-month momentum</td>
<td>Cumulative return over 6 months ending in $t-2$</td>
<td>Jegadeesh and Titman (1993)</td>
</tr>
<tr>
<td>8</td>
<td>mom12m</td>
<td>12-month momentum</td>
<td>Cumulative return over 11 months ending in $t-2$</td>
<td>Jegadeesh (1990)</td>
</tr>
<tr>
<td>9</td>
<td>mom30m</td>
<td>30-month momentum</td>
<td>Cumulative return over 24 months ending in $t-13$</td>
<td>Bondt and Thaler (1985)</td>
</tr>
<tr>
<td>10</td>
<td>nve</td>
<td>Market capitalization (size)</td>
<td>Log of ($SHROUT \times PRC$)</td>
<td>Banz (1981)</td>
</tr>
<tr>
<td>11</td>
<td>ep</td>
<td>Earnings-to-price</td>
<td>Earnings per share</td>
<td>Basu (1977)</td>
</tr>
<tr>
<td>12</td>
<td>dy</td>
<td>Dividends-to-price</td>
<td>Yearly dividends (dvt) divided by market cap at fiscal year</td>
<td>Litzenberger and Ramaswamy (1979)</td>
</tr>
<tr>
<td>13</td>
<td>bm</td>
<td>Book-value-to-market</td>
<td>Book value of equity (ceq) divided by market cap</td>
<td>Rosenberg et al. (1985)</td>
</tr>
<tr>
<td>14</td>
<td>lev</td>
<td>Leverage</td>
<td>Total liabilities (lt) divided by market cap</td>
<td>Bhandari (1988)</td>
</tr>
<tr>
<td>15</td>
<td>currat</td>
<td>Current ratio</td>
<td>Current assets (act) divided by current liabilities (lct)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>16</td>
<td>pchcurrat</td>
<td>Pct change in current ratio</td>
<td>Percentage change in currat from year $t-1$ to $t$</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>17</td>
<td>quick</td>
<td>Quick ratio</td>
<td>Current assets (act) minus inventory (invt), divided by current liabilities (lct)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>18</td>
<td>pchquick</td>
<td>Pct change in quick ratio</td>
<td>Percentage change in quick from year $t-1$ to $t$</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>19</td>
<td>sale</td>
<td>Sales-to-cash</td>
<td>Annual sales (sale) divided by cash and cash equivalents (che)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>20</td>
<td>salerec</td>
<td>Sales-to-receivables</td>
<td>Annual sales (sale) divided by accounts receivable (rect)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>21</td>
<td>saleinv</td>
<td>Sales-to-inventory</td>
<td>Annual sales (sale) divided by total inventory (invt)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>22</td>
<td>pchsaleinv</td>
<td>Pct change in sales-to-inventory</td>
<td>Percentage change in salesinv from year $t-1$ to $t$</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>23</td>
<td>cashdebt</td>
<td>Cashflow-to-debt</td>
<td>Earnings before depreciation and extraordinary items (ib + dp) divided by avg total liabilities (lt)</td>
<td>Ou and Penman (1989)</td>
</tr>
<tr>
<td>25</td>
<td>dep</td>
<td>Depreciation-to-gross PP&amp;E</td>
<td>Depreciation expense (dp) divided by gross PP&amp;E (ppegpt)</td>
<td>Holthausen and Larcker (1992)</td>
</tr>
<tr>
<td>26</td>
<td>pchdepr</td>
<td>Pct change in Depreciation-to-gross PP&amp;E</td>
<td>Percentage change in dep in year $t-1$ to $t$</td>
<td>Holthausen and Larcker (1992)</td>
</tr>
<tr>
<td>27</td>
<td>nve_ia</td>
<td>Industry-adjusted firm size</td>
<td>Log market caps are adjusted by log of the mean of the industry</td>
<td>Asness et al. (2000)</td>
</tr>
<tr>
<td>30</td>
<td>sgr</td>
<td>Annual sales growth</td>
<td>Percentage change in sales from year $t-1$ to $t$</td>
<td>Lakonishok et al. (1994)</td>
</tr>
<tr>
<td>31</td>
<td>ipo</td>
<td>IPO</td>
<td>Indicated by 1 if first 12 months available on CRSP monthly file</td>
<td>Loughran and Ritter (1995)</td>
</tr>
<tr>
<td>32</td>
<td>divi</td>
<td>Dividend initiation</td>
<td>Indicated by 1 if company pays dividends but did not in prior year.</td>
<td>Michaely et al. (1995)</td>
</tr>
<tr>
<td>33</td>
<td>divo</td>
<td>Dividend omission</td>
<td>Indicated by 1 if company does not pay dividends but did in prior year.</td>
<td>Michaely et al. (1995)</td>
</tr>
<tr>
<td>34</td>
<td>sp</td>
<td>Sales-to-price</td>
<td>Annual sales (sale) divided by market cap</td>
<td>Barber Jr et al. (1996)</td>
</tr>
<tr>
<td>35</td>
<td>acc</td>
<td>WC accruals</td>
<td>((ib) + (oancf) - (at), if (oancf) is missing then (ib) - (delta_act) - (delta_che) - (delta_lct) + (delta_dlc) + (txp-dp))</td>
<td>Sloan (1996)</td>
</tr>
<tr>
<td>36</td>
<td>turn</td>
<td>Share turnover</td>
<td>Average monthly trading volume for the three months $t-3$ to $t-1$ divided by SHROUT at $t-1$</td>
<td>Datar et al. (1998)</td>
</tr>
<tr>
<td>37</td>
<td>pchsal_pchinv</td>
<td>Delta pct change sales vs. inventory</td>
<td>Difference of percentage changes in sales (sale) and inventory (invt) where each item 0 if missing</td>
<td>Datar et al. (1998)</td>
</tr>
<tr>
<td>38</td>
<td>pchsales_pchrc</td>
<td>Delta pct change sales vs. receivables</td>
<td>Difference of percentage changes in sales (sale) and receivables (rect)</td>
<td>Datar et al. (1998)</td>
</tr>
<tr>
<td>39</td>
<td>pchcgx_pchpx</td>
<td>CAPEX</td>
<td>Industry adjusted (two digit SIC) fiscal year mean adjusted percentage change in capital expenditures (capex)</td>
<td>Datar et al. (1998)</td>
</tr>
<tr>
<td>40</td>
<td>pchgm_pchsa</td>
<td>Delta pct gross margin vs. sales</td>
<td>Annual percentage change in gross margin (sale minus cogs) minus percentage change in sales (sale)</td>
<td>Datar et al. (1998)</td>
</tr>
</tbody>
</table>

Table 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.
<table>
<thead>
<tr>
<th>ID</th>
<th>Acronym</th>
<th>Name</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>pchsale_pchxsga</td>
<td>Delta pct sales vs. SGaA</td>
<td>Annual percentage change in sales (sale) minus percentage change in SGaA (xsga)</td>
<td>Abarbanell and Bushee (1997)</td>
</tr>
<tr>
<td>42</td>
<td>dolvol</td>
<td>Dollar trading volume</td>
<td>Log of trading volume times price per share from month t-2</td>
<td>Chordia et al. (2001)</td>
</tr>
<tr>
<td>43</td>
<td>std_dolvol</td>
<td>Volatility trading volume</td>
<td>Monthly standard deviation of daily trading volume</td>
<td>Chordia et al. (2001)</td>
</tr>
<tr>
<td>44</td>
<td>std_turn</td>
<td>Volatility turnover</td>
<td>Monthly standard deviation of daily share turnover</td>
<td>Chordia et al. (2001)</td>
</tr>
<tr>
<td>45</td>
<td>chinv</td>
<td>Change in inventory</td>
<td>First difference of inventory (invt) divided by total assets</td>
<td>Thomas and Zhang (2002)</td>
</tr>
<tr>
<td>46</td>
<td>pchemp_ia</td>
<td>Industry-adjusted pch in employees</td>
<td>Industry adjusted percentage change in employees</td>
<td>Asness et al. (2000)</td>
</tr>
<tr>
<td>47</td>
<td>cfp</td>
<td>Cashflow-to-price</td>
<td>Operating cash flows (oancf) scaled by market capitalization (fiscal year end)</td>
<td>Desai et al. (2004)</td>
</tr>
<tr>
<td>48</td>
<td>rdl</td>
<td>R&amp;D Increase</td>
<td>If annual increase in R&amp;D expenses (xrd) scaled by total assets (at) &gt;0.05, 1, else 0</td>
<td>Eberhart et al. (2004)</td>
</tr>
<tr>
<td>49</td>
<td>lgr</td>
<td>Pct change in long-term debt</td>
<td>Annual percentage change in long term debt (lt)</td>
<td>Richardson et al. (2005)</td>
</tr>
<tr>
<td>50</td>
<td>egr</td>
<td>Pct change in book equity</td>
<td>Annual percentage change in book equity (ceq)</td>
<td>Richardson et al. (2005)</td>
</tr>
<tr>
<td>51</td>
<td>rd_sale</td>
<td>R&amp;D-to-sales</td>
<td>R&amp;D expenses(xrd) scaled by sales (sale)</td>
<td>Guo et al. (2006)</td>
</tr>
<tr>
<td>52</td>
<td>rd_mve</td>
<td>R&amp;D-to-market</td>
<td>R&amp;D expenses(xrd) scaled by market cap</td>
<td>Guo et al. (2006)</td>
</tr>
<tr>
<td>53</td>
<td>chg_mon6m</td>
<td>change in mom6m</td>
<td>difference of mom6m measured at t and t-6</td>
<td>Gettleman and Marks (2006)</td>
</tr>
<tr>
<td>54</td>
<td>hire</td>
<td>Pct change in employee</td>
<td>Annual percentage change in employee (emp)</td>
<td>Belo et al. (2014)</td>
</tr>
<tr>
<td>55</td>
<td>agr</td>
<td>Asset growth</td>
<td>Annual percentage change in assets (at)</td>
<td>Cooper et al. (2008)</td>
</tr>
<tr>
<td>56</td>
<td>cashpr</td>
<td>Cash productivity</td>
<td>Market cap plus long term debt (dltt) minus assets (at) divided by cash (che)</td>
<td>Chandrashekar and Rao (2009)</td>
</tr>
<tr>
<td>57</td>
<td>gma</td>
<td>Gross-profitability</td>
<td>Sales (sale) minus costs of goods sold (cogs) divided by one-year lagged assets(at)</td>
<td>Novy-Marx (2013)</td>
</tr>
<tr>
<td>58</td>
<td>cash</td>
<td>Cash-to-assets</td>
<td>Cash (che) divided by assets(at)</td>
<td>Palazzo (2012)</td>
</tr>
<tr>
<td>59</td>
<td>pctacc</td>
<td>Accruals-to-income</td>
<td>(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</td>
<td>Hafzalla et al. (2011)</td>
</tr>
<tr>
<td>60</td>
<td>absacc</td>
<td>Absolut accruals</td>
<td>Absolute value of acc:</td>
<td>Bandyopadhyay et al. (2010)</td>
</tr>
<tr>
<td>61</td>
<td>roic</td>
<td>Return on invested capital</td>
<td>Earnings before interest and taxes (ebit) - non-operating income (nopi), divided by non-cash enterprise value</td>
<td>Brown and Rowe (2007)</td>
</tr>
<tr>
<td>62</td>
<td>gcapex</td>
<td>Pct change in two year CAPX</td>
<td>Percentage change in two year capital expenditure (capex)</td>
<td>Anderson and Garcia (2006)</td>
</tr>
<tr>
<td>63</td>
<td>tang</td>
<td>Debt-capacity-to-firm-tangibility</td>
<td>(Cash (che) + 0.715 receivables (rect) + 0.547 inventory(invt) + 0.535 (ppegt)) / total assets (at)</td>
<td>Hahn and Lee (2009)</td>
</tr>
<tr>
<td>64</td>
<td>chshrout</td>
<td>Change in shares-outstanding</td>
<td>Yearly percentage change in outstanding shares (SHROUT)</td>
<td>Pontiff and Woodgate (2008)</td>
</tr>
<tr>
<td>65</td>
<td>invest</td>
<td>CAPEX and inventory</td>
<td>Yearly difference in gross property, plant and equipment (ppegt) + diff in (invt) / (t-1) total assets (at)</td>
<td>Chen and Zhang (2010)</td>
</tr>
<tr>
<td>66</td>
<td>age</td>
<td>Years since CS coverage</td>
<td>Years since first compustat coverage years(data date - min(data date))</td>
<td>Jiang et al. (2005)</td>
</tr>
<tr>
<td>67</td>
<td>chpm_ia</td>
<td>Industry-adjusted change in profit margin</td>
<td>Industry adjusted (two-digit SIC) change in profit margin (ib/sale)</td>
<td>Soliman (2008)</td>
</tr>
<tr>
<td>68</td>
<td>chato_ia</td>
<td>Industry-adjusted change in asset turnover</td>
<td>Industry adjusted (two-digit SIC) change in asset turnover (sale/at)</td>
<td>Soliman (2008)</td>
</tr>
</tbody>
</table>

Table 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.
Table 10: 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.
Table 11: Sharpe Ratio Tests - first stage coef estimator, i.e. also using the respective first stage selection method for the coefficient estimation. The columns all, large, mid and small represent the test-statistics for differences in Sharpe Ratios using the Jobson and Korkie (1981) test including the Memmel (2003) correction. A negative test statistic indicates a higher SR ratio for the adaptive Lasso factor compared to the OLS factor. The table displays the results conditioned on the four cases we consider: for each case we have 16 combinations, which results in a total of 64 combinations. Considering only the results of all stocks in the analysis, we see that in 87.50% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 75.00% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 6.25% of the cases. No significant difference is observed in 18.75% of the cases. The column of the large cap stocks in the analysis shows that in 81.25% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 50.00% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 37.50% of the cases. No significant difference is observed in 25.00% of the cases. Finally, the results with respect to the small caps stocks only show that in 81.25% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 37.50% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 12.50% of the cases. No significant difference is observed in 50.00% of the cases. Finally, the results with respect to the small caps stocks only show that in 81.25% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 37.50% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 12.50% of the cases. No significant difference is observed in 50.00% of the cases. Finally, the results with respect to the small caps stocks only show that in 81.25% of the cases we observe a higher Sharpe Ratio for the Adaptive Lasso factor vs. the OLS factor. The Adaptive Lasso factor has a significantly higher Sharpe ratio at the 5% level in 37.50% of the cases. The OLS factor has a significantly higher Sharpe ratio at the 5% level in 12.50% of the cases. No significant difference is observed in 50.00% of the cases.