Nonlinearities Everywhere: Sparse Supervised Learning of Market Anomalies*

Gabriel Kaiser

This version: November 2020
First draft: November 2019

Abstract

This paper studies various regularization terms of nonparametric return forecasting models to identify a reduced set of characteristics associated with expected return spreads in the cross-section. The models employed are outlier-resistant, computationally efficient, and able to handle high dimensional data. The number of discovered firm characteristics range between two and eight, out of a total of 90, taken from a novel market anomaly database. This study finds that in a multivariate setting, nonlinearities as well as variations in time matter for the cross-section of expected returns. Out-of-sample results suggest that Elastic Net regularization terms tend to overfit the data, while MCP and SCAD penalties suggest sparse models with remarkable Sharpe ratios.

JEL Classification: C14, C52, C58, G12.

Keywords: Anomalies, Cross-Sectional Return Predictability, LASSO, Model Selection.

---

*I would like to thank my supervisors Tibor Neugebauer, Julien Penasse and Andrea Tamoni. Further, I am grateful to Arnaud Dufays, Simon Rottke, Ulf von Lilienfeld-Toal, Christian von Drathen, Denis Gromb for insightful conference discussions and seminar and conference participants at Luxembourg School of Finance, the London School of Economics, HEC Liège. The estimations presented in this paper were carried out using the HPC facilities of the University of Luxembourg Varrette, Bouvry, Cartiaux, and Georgatos (2014) – see `https://hpc.uni.lu`. All errors are the sole responsibility of the author.
Correspondence: Department of Finance, LSF, University of Luxembourg, Tel.: (+352) 46 66 44 5948, Email: gabriel.kaiser@uni.lu, Web: `https://gabrielkaiserqfin.github.io`

1 Introduction

In this paper I revisit 90 well-documented market anomalies, see Harvey, Liu, and Zhu (2016), Green, Hand, and Zhang (2017), Freyberger, Neuhierl, and Weber (2020), and select a reduced subset with independent drivers of cross-sectional expected returns. A market anomaly is thereby understood as a predictable deviation of observed market behavior from the prediction of a standard asset pricing model, such as the Fama and French (1992) three factor model (FF3). It typically implies a significant nonzero intercept in the empirical validation of the model. In a perfectly rational market, this intercept should be indistinguishable from zero at least most of the time.

Nonetheless, 50 years of comprehensive research in empirical asset pricing led to a remarkably large number of cross-sectional market anomalies, see Harvey et al. (2016). These market anomalies are based on information such as firm characteristics and stock prices, hereinafter referred to as firm characteristics. However, with an increasing number of explanatory variables, the number of parameter estimates grows, and so does the parameter uncertainty. Eventually, a large variance can lead to inefficiencies and over-fitting. In contrast, reducing the number of explanatory variables might produce a larger bias with higher efficiency, thereby under-fitting the true model. Controlling for this curse of dimensionality has thus become a major objective in the literature focusing on cross-sectional equity returns.

Secondly, in times of financial distress, cross-sectional hedge returns obtained from portfolio sorting on firm characteristics or on price changes have substantial commonality. In these periods, portfolio returns are driven by the same underlying risk factors. This implies a low pricing kernel dimension and hence, a small set of independent firm characteristics proxying for latent risk factors, see Cochrane (2011). Indeed, some market anomalies are highly correlated and merely products of others, as shown in Figure 1.

This paper solves these two shortcomings by employing the statistical learning technique known as regularization or penalization. Regularization selects only non-redundant explanatory variables, which reduces variance and increases bias. This is particularly important when the number of predictors p relative to the number of observations N is large, i.e. $p \gg N$. In the following, I identify a reduced set of characteristics associated with expected return spreads in the cross-section, thereby making use of a variety of regularization terms. These terms differ in their properties regarding variable selection and their degree of bias. I suggest one-stage penalties with a significantly lower bias compared to penalization methods discussed in current literature.

Another main objective of a common methodology used in factor modelling, namely portfolio sorting, is the limitation to the number of simultaneous explanatory variables. I follow the approach of Huang, Horowitz, and Wei (2010) and Freyberger et al. (2020) and use a nonparametric return forecasting model. This model is an econometric technique that overcomes the limitations of portfolio sorting concerning high dimensional data. Additionally, the nonparametric method allows for the retrieval of a flexible nonlinear functional form that maps the cross-sectional heterogeneity of firm characteristics to expected returns without imposing numerous assumptions in contrast to linear Fama-MacBeth regressions. The nonparametric framework expresses itself in the form of additive Basis-splines. These regression splines have several advantages compared to a Generalized Linear Model or Fama-MacBeth regressions. First, allowing for nonlinear relationships between characteristics and expected returns makes any individual transformation of characteristic variables redundant. Second, the use of Basis-splines increases prediction accuracy and decreases the confidence intervals of the forecasts. Third, the additive structure enables tractability and interpretability of individual characteristics similar to a standard ordinary least squares (OLS) model.

Combining regularization terms and predictive regressions that are novel within the finance literature allows me to identify a reduced and independently important information subset of 13 firm characteristics from a new data set of 90 market anomalies. The determined information subset discovered is time-varying and hence state-dependent.

In particular, I show that single-stage models with a Minimax Concave Penalty (MCP) or Smoothly Clipped Absolute Deviation (SCAD) penalty at a group level introduce enough sparsity to answer the multidimensional challenge imposed by the substantial set of return predictors. As alternatives to the adaptive group Least Absolute Shrinkage and Selection Operator (adaptive LASSO) as suggested by Freyberger et al. (2020), MCP and SCAD penalties obtain less biased regression coefficients in sparse models. That is, given a certain threshold parameter, estimates are not penalized at all. The threshold is determined by a tuning parameter, which, in the case of this paper, is determined via 10-fold cross-validation, in contrast to Huang et al. (2010). Furthermore, the suggested methodology allows the estimated coefficients to reach large values quicker than multi-stage models such as the adaptive group LASSO. Important firm characteristics have therefore a greater chance of entering the reduced subset. A positive side effect of a single-stage approach is computational efficiency.

The market anomalies considered in this paper span the period from January 1965 to December 2018, which is, on average, twice the original sample period, as presented in Table 2. Hence, anomalies subject to data mining shall not survive the extended sample.

Mehra and Prescott (1985) suggest an equity premium of 8% and a standard deviation of about 16% for annual US equity returns, which leads to tradeable Sharpe ratios of around 0.5. In this extended sample, 13 of these market anomalies outperform the market in terms of Sharpe ratios, and around 30% show significant intercepts after controlling for the CAPM and the Fama-French three-factor (FF3) model. Therefore, 30 of the market anomalies considered show persistent performance and are not subject to data mining. Focusing on the variable selection, out of these 90 market anomalies, 31 variables are selected in at least one of the in-sample linear models. However, only six characteristics are selected by all nine linear models. Accounting for nonlinear relationships via a non-parametric return forecasting model enhanced with various regularization terms such as adaptive LASSO, SCAD or MCP, leads to a tremendous reduction in variables of two to nine characteristics and an increase in the predictive power within the full sample. However, only three are consistent across various model specifications. As a result, I conduct an out-of-sample comparison to identify the most promising model. Moreover, I present Lewellen-R2 and Betas in addition to the standard performance measures. I also include higher moments of the constructed portfolio because a market anomaly is expected to be an adequate risk premium whenever mean returns line up with covariances. This results in insurance-like payoffs of hedge returns. Distribution-wise, these returns should show a slightly positive mean, a larger median, and a leptokurtic body. Exogenous shocks, such as the Great Depression, will lead to heavy tails on the downside that cause the negative skewness in the joint distribution of returns. Several of the reported portfolios seem to share a feature of risk premium-like payoffs. Throughout specifications, the SCAD model seems to present the highest Lewellen Betas and R2. This regularization term does not penalize critical firm characteristics and therefore allows them to survive the dimension reduction. Furthermore, it turns out that, across various periods, the majority of previously discovered firm characteristics vanish and others emerge, which suggests that the function that maps characteristics to expected stock returns is not only nonlinear but also time-varying. Given the evidence, I continue with a rolling estimation of a variable selection window of 20 years and a forecasting horizon of only one year employing the nonparametric model enhanced with the SCAD penalty. The estimation leads to significantly larger out-of-sample Sharpe ratios than their time-invariant counterparts. Four characteristics, A2ME, REV1M, MOM18M and IndRev1M, persist over time, while 16, such as the Bid-Ask spread, vary tremendously through time.

I structure the remainder of this paper as follows. In section 2, I review related literature. Section 3 introduces the methodology. Section 4 describes the data sources used in this paper. In section 5, I present the in- and out-of-sample results. Finally, section 6 summarizes the findings.

2 Related Literature

Fama and French (1992, 1993) identified specific firm characteristics (C), in particular valuation ratios and market capitalization. Based on these characteristics, they form factor portfolios providing independent risk exposures additional to the market risk factor, leading to the FF3 model. The subsequent extension by Carhart (1997) includes Momentum and forms the Fama-French-Carhart four-factor model (FF4). It has become a successful tool in the academic world as well as in the industry, although it is rejected in explaining average returns due to its significant nonzero intercepts. Consequently, subsequent research started fishing for factors that jointly eliminate any market anomalies.

In the past two decades, empirical asset pricing has led to a fairly large number of anomalies in the cross-section of average equity returns. Harvey et al. (2016) summarize 316 firm characteristics that form market anomalies and typically also factor portfolios in the literature. They also find that 50% of anomalies result from statistical biases in the analysis, a common concern intensively documented in the literature, for example, by Barras, Scaillet, and Wermers (2010). Other contributions that address the emergence of market anomalies are Green et al. (2017), McLean and Pontiff (2016), and many others. Green et al. (2017), for instance, discover that only the minority of characteristics present independent information after adjusting p-values and controlling for micro-cap stocks in their Fama-MacBeth regressions. In contrast, Novy-Marx and Velikov (2016), as well as A. Y. Chen and Velikov (2019), argue that 50% of remaining anomalies are, in fact, subject to transaction costs. Accordingly, these papers suggest that the entire zoo vanishes after publication. Other causes for anomalies can be the limit of arbitrage, in particular deteriorating funding conditions, as suggested by Adrian, Etula, and Muir (2014).

A different approach to overcome these adverse effects of multiple null hypothesis testing is to estimate the rate of type I errors or false discovery rate. However, this approach is only applicable if one possesses information regarding the number of multiple testings. Typically, this information is only known to the author of the market anomaly and can thus rarely be used for testing.

While portfolio sorting is limited to the number of sorts feasible, Fama-MacBeth regressions are limited to the number of factors, as outlined in the introduction 1. Given these difficulties, Freyberger et al. (2020) suggest a nonparametric return forecasting model that enables statistical tests in a high-dimensional setting. In addition, to overcome data mining as well as overfitting, they introduce the adaptive LASSO regularization term that penalizes parameter estimates and proposes sparsity. In general, regularization is capable of extracting independent firm characteristics of the dozens of variables found in the literature that produce expected return spreads. I pick up on the methodology suggested by Freyberger et al. (2020) and extend their model with various regularization terms with different properties.

One common stylized fact within this research field is to assume that "means and covariances are stable functions of characteristics not of individual securities," as examined by Cochrane (2011). We can denote such a relationship by

$$R_{it+1} = g(C_{it}, k) + \epsilon_{it+1}, \tag{1}$$

where $g(.)$ represents an unknown function with parameters k and an individual firm characteristic C_{it}, which formulates the conditional expectation of excess returns $\mathbb{E}(R^e_{t+1}|C_t)$. Individual excess or long-short returns are denoted as R_{it+1}. The functional form relates the multidimensional vector of firm characteristics to individual expected returns and covariances, thereby also allowing for nonlinearities in the cross-section.

Whereas the majority of the abovementioned papers suggest a static model, time-varying expected returns demand dynamic models, leading to a relatively new strand in the finance literature and the involvement of machine learning in asset pricing. This literature of machine learning is essentially motivated by standard theoretical asset pricing models such as Campbell and Cochrane (1999), Bansal and Yaron (2004), and Santos and Veronesi (2004). The common objective of this literature is to recover the function $g(.)$, i.e. which firm characteristics are arguments of this function, in a linear or non-linear way. An example of its empirical implementation is, for instance, a paper by Haddad, Kozak, and Santosh (2018), which identifies the time-varying function of the conditional expectation of excess returns in terms of characteristics using Principal Component Analysis (PCA). L. Chen, Pelger, and Zhu (2019) employ a Deep Neural Network with long-short-term-memory and find that nonlinearities are essential in a multidimensional factor world. Kelly, Pruitt, and Su (2019) suggest using an Instrumented Principal Component Analysis to incorporate both time-varying factor loadings and latent factors. Their suggested five-factor model outperforms previous well-established models. Gu, Kelly, and Xiu (2018) compare a set of techniques including dimension reduction, generalized linear models, random forests, boosted regression trees, and neural networks, and they conclude that such networks and trees outperform the rest.

In contrast, many papers rely on LASSO regularization in order to introduce sparsity in their models: DeMiguel, Martin-Utrera, Nogales, and Uppal (2019) take a portfolio perspective extending Brandt, Santa-Clara, and Valkanov (2009) and include a trading cost term in the objective function adhered by an LASSO regularization term. They find that the number of significant characteristics increases from six to fifteen after including transaction costs due to a cancellation effect in rebalancing. Kozak, Nagel, and Santosh (2019) construct a robust stochastic discount factor (SDF) in a Bayesian setting and find that sparsity (LASSO) is elusive, whereas RIDGE penalization works well. Freyberger et al. (2020) apply a nonparametric adaptive group LASSO on ranked firm characteristics and observe high Sharpe-ratios. From 62 firm characteristics around 13 seem to provide independent information in a nonlinear framework. Chinco, Clark-Joseph, and Ye (2019) successfully predict high-frequency expected returns employing LASSO regressions.

Nevertheless, since LASSO as well as adaptive LASSO have biased estimates, see for instance Zhang et al. (2010) or Fan and Li (2001), I suggest to use an extension of LASSO that excludes large estimates from penalization and hence keeps these estimates unbiased. Therefore, this paper extends by introducing a variety of regularization terms, such as MCP and SCAD, in combination of a nonparametric return forecasting model.

3 Model and Methodology

In general, regularization methods try to optimize the trade-off between model bias and efficiency. Before elaborating on the bias-variance trade-off in the context of this paper, let us define the relationship between expected returns and firm characteristics.

Assumption *Return moments such as mean and covariance are stable and well-behaved functions of firm characteristics.*

$$g(C) = \mathbb{E}[R_{it+1}|C_{1,it} = size_{it}, \ldots, C_{Z,it} = mom_{it}] \tag{2}$$

I denote this set of firm characteristics or anomalies by the matrix $C_t \in \mathbb{R}^{N \times Z}$, where N is the number of firms, and Z is the set of 90 firm characteristics of a novel market anomaly database. Equation (2) relates an unknown function $g(\cdot)$, with characteristics as arguments to conditional expected returns of individual stocks.

Cochrane (2011) argues that portfolio sorts, the most common methodology to study risk premia, are equivalent to nonparametric cross-sectional predictive regressions. Freyberger et al. (2020) establish the theoretical relationship and provide empirical evidence of the superiority of nonparametric regressions relative to independent portfolio sorts and linear regressions. In their empirical section, they employ the adaptive group LASSO approach as in Huang et al. (2010) to select nonlinear mappings of rank-normalized firm characteristics.

I pick up on their findings and extend the nonparametric model with novel regularization terms as discussed in the next subsection. The suggested penalties require only a single-stage estimation and produce unbiased estimates. I benchmark my findings against the adaptive group LASSO approach. To lessen the impact of extreme values, I rank-normalize each firm characteristic across firms, which I denote by $\tilde{C}_{it}$. The predictive regression then becomes:

$$R_{it+1} = \sum_{z=1}^{Z} g_z(\tilde{C}_{it}) + \epsilon_{it+1}. \tag{3}$$

Identifying any unknown function $g(\cdot)$ is a difficult task, specifically if covariates are slightly collinear and thus challenging to separate. Additionally, the signal-to-noise ratio in empirical asset pricing is exceptionally low compared with other disciplines, such as electrical engineering or medicine. Another aspect is that the univariate evidence of nonlinear functional forms that map certain characteristics to expected returns, as in Cattaneo, Crump, Farrell, and Schaumburg (2018), might not hold in a multivariate form. These unfortunate aspects motivate the usage of various feature subset selection techniques to benefit from the contrasting model properties.

3.1 Nonparametric Interpolation

A standard approach in the market anomaly literature is to form characteristic-sorted portfolios and to evaluate the expected returns of these portfolios against some hypothesis, known as independent portfolio sorting. By construction, this methodology is limited to bivariate or in the best case trivariate sortings and thus inappropriate for my case of 90 market anomalies. An alternative to independent portfolio sorts are predictive panel regressions in linear form as in Fama and French (2008):

$$R_{it+1} = \alpha + \sum_{z}^{Z} \beta_z C_{it,z} + \epsilon_{it+1}. \tag{4}$$

This return forecasting model is flexible enough to deal with a multivariate set of characteristics but assumes linearity across the range of each characteristic. For instance, including indicator functions allows for nonlinear relationships as suggested by Cattaneo et al. (2018) for the univariate case. The corresponding pooled time-series cross-sectional regression can be written as:

$$R_{it+1} = \alpha + \sum_{z}^{Z} \beta_{z,k} \mathbb{1}(C_{it,z} \in F^{-1}_{C_{t,z}}(\tau)) + \epsilon_{it+1}, \tag{5}$$

where the indicator function $\mathbb{1}(.)$ is one, whenever characteristic z of stock i in period t is in percentile k of the empirical cumulative distribution function $F_{C_{t,z}}$. Freyberger et al. (2020) already established the equivalence between portfolio sorts and panel regressions of such kind.

The most direct way to present portfolio sorts in a panel regression is via piecewise linear basis functions, also known as regression splines in statistics. To capture nonlinearities beyond the percentile cut points, we can extend the basis function to polynomials. In the context of the cross-section of expected returns, quadratic up to cubic polynomial degrees, seem to be adequate, as shown in Huang et al. (2010). I rank-normalize each characteristic per period t and denote the transformed characteristic $C_{t,z}$ by tilde $\tilde{C}_{t,z}$. Rank-normalizing is equivalent to estimating the empirical cumulative distribution function $F_{C_{t,z}}$ and thus any percentile cutoff equals actual values of $\tilde{C}_{t,z}$. Furthermore, extreme values will not influence the estimation. The piecewise nonlinear basis functions in this paper use quadratic polynomials in an additive form (i.e. 1, $\tilde{C}_{t,z}$, $\tilde{C}^2_{t,z}$) and a truncated quadratic basis function $h(.)$ to imitate portfolio sorts per percentile k:

$$h(\tilde{C}_{t,z}, k) = (\tilde{C}_{t,z} - k)^2_+ = \begin{cases} (\tilde{C}_{t,z} - k)^2 & \text{if } \tilde{C}_{t,z} > k \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where k comprises percentile values of the empirical distribution of the rank-normalized firm characteristics $\tilde{C}_{t,z}$. Therefore, K can be interpreted as the number of portfolio sorts, where K sorts require $K - 1$ cutoff points, typically referred to as knots. The OLS component can be broken down into future returns and a vectorized sum. Hence, I establish the baseline model as:

$$\begin{aligned} R_{it+1} &= \alpha + \sum_{z}^{Z} \beta'_z \cdot g_z(\tilde{C}_{it,z}, k) + \epsilon_{it+1} \\ &= \alpha + \sum_{z=1}^{Z} \sum_{k=1}^{K+2} \beta_{z,k} \cdot g_z(\tilde{C}_{it,z}, k) + \epsilon_{it+1}, \end{aligned} \tag{7}$$

where g is just a mapping of each characteristic, i.e. $g_z(\tilde{C}_{t,z}) : \mathbb{R}^N \mapsto \mathbb{R}^{(K+2)\times N}$ and $\hat{\beta}_z$ is a $K + 2$ vector of estimated coefficients per basis function, respectively portfolio sorting.

3.2 Regularization

After establishing the baseline nonparametric panel model, we consider that a large number of characteristics increases parameter uncertainty and is likely to produce an overfitted model. To mitigate the influence of each component and to perform variable selection, I extend the baseline model (7) with regularization terms. The extended model is flexible enough to allow for various regularization terms with distinctive properties. In particular, I consider five distinct penalties for the linear case: LASSO, adaptive LASSO, SCAD, Elastic Net, and MCP. For the nonlinear case, I utilize their grouped analogs: **Group LASSO, Adaptive Group LASSO, Group SCAD, Group Elastic Net, Group MCP**, where the word 'group' indicates that penalization is applied at the group level. Without loss of generality, this section describes all regularization terms at the group level to cluster interpolations of each characteristic. It is straightforward to obtain the original form by setting $(K+2) = 1$.

First, the group LASSO algorithm utilizes the ℓ_1 norm, which controls the complexity of the model. Although the ℓ_1 norm is convex, it is non-differentiable around zero and therefore harder to optimize via quadratic programming. Assuming that all variable pairs are not perfectly collinear, a unique solution exists. Hence, highly correlated covariates are problematic. In general, when $N \ll p$, its solution is not uniquely defined and can select at most N variables before it saturates due to the nature of ℓ_1 optimization. In other words, it tends to select fewer correlated variables and not necessarily the optimal subset. Nonetheless, if a unique solution exists, it pushes the entries towards zero and enforces sparsity. As suggested by Tibshirani (1996), it is therefore an appropriate methodology for variable subset selection:

$$R_{\text{LASSO}} = \|\boldsymbol{\beta}_z\|_{2,1} = \left(\sum_{k=1}^{K+2} \beta_{z,k}^2\right)^{\frac{1}{2}}, \tag{8}$$

where the number of groups is defined by the number of knots and basis functions (i.e. $K+2$). All coefficients are squared and thereafter aggregated per characteristic z before the square root is taken. In the linear LASSO model, the sum of equation (8) is dropped, and the absolute value of the coefficient β_z is taken.

A disadvantage of LASSO is that it does not satisfy the weak Oracle properties, which implies that the parameter estimates obtained from variable selection are not unbiased. We can employ adaptive LASSO as a countermeasure, which leads us to the second method. The adaptive component realizes after estimating a standard LASSO. In this second stage, additional weights w_z (one per group z) are introduced in the regularization term leading to the following equation:

$$R_{\text{Adaptive}} = w_z \cdot \hat{R}_{\text{LASSO}} = \begin{cases} \left(\sum_{k=1}^{K+2} \hat{\beta}_{zk}^2\right)^{-\frac{1}{2}} & \text{if } \sum_{k=1}^{K+2} |\hat{\beta}_{zk}| \neq 0 \\ \infty & \text{if } \sum_{k=1}^{K+2} |\hat{\beta}_{zk}| = 0. \end{cases} \tag{9}$$

These weights are just the inverse of the root of the squared sum per group of step one estimated coefficients. Hence, adaptive LASSO penalizes coefficients with lower initial estimates more and in the extreme case of zero predictive power per group, with infinity. This is because the regularization term pushes all coefficients of an entire group of each characteristic to zero whenever the predictive power is little, thereby introducing sparsity. The oracle procedure ensures an optimal estimation rate and is thus more efficient as illustrated in Zou (2006). Although parameter estimates are improved, they are still biased.

A third alternative, the SCAD penalty, is employed to compare the results. It is a single step approach that extends the concept of LASSO to a certain threshold and sets the penalization term to zero thereafter. Substantial coefficients are, therefore, unbiased, as described in more detail in Fan and Li (2001):

$$R_{\text{SCAD}} = p_z^{SCAD} \|\boldsymbol{\beta}_z\|_{2,1}. \tag{10}$$

The switch from the penalized to unpenalized form is smooth and is accomplished by a quadratic spline function with knots at λ and $\gamma\lambda$:

$$p_z^{SCAD}(\beta; \lambda, \gamma) = \begin{cases} \lambda \|\boldsymbol{\beta}_z\|_{2,1} & \text{if} \quad \|\boldsymbol{\beta}_z\|_{2,1} \leq \lambda \\ \frac{\|\boldsymbol{\beta}_z\|_{2,1}^2 - 2\gamma\lambda\|\boldsymbol{\beta}_z\|_{2,1} + \lambda^2}{2(1-\gamma)} & \text{if} \quad \lambda < \|\boldsymbol{\beta}_z\|_{2,1} \leq \gamma\lambda \\ \frac{(\gamma+1)\lambda^2}{2} & \text{if} \quad \|\boldsymbol{\beta}_z\|_{2,1} > \gamma\lambda, \end{cases} \tag{11}$$

where $\|\boldsymbol{\beta}_z\|_{2,1} = \left(\sum_{k=1}^{K+2} \beta_{z,k}^2\right)^{\frac{1}{2}}$ is again aggregated at the group level, as in equation (8). In the case of the linear SCAD model the $\ell_{2,1}$-Norm of equations (10), and (11) is replaced by the ℓ_1-Norm. I set the knot parameter of the quadratic spline function of the SCAD term in equation 11 to three, $\lambda = 3$, a common choice in the machine learning literature.

The next method introduces shrinkage in the form of a RIDGE penalty in combination with the LASSO regularization term. Zou and Hastie (2005) argue that it is more stable than LASSO with regards to feature selection. Furthermore, it yields improved predictions compared with LASSO whenever variables are correlated. The intuition behind Elastic Net is that unknown groups of variables, for instance, related genes in microarrays, are grouped while producing sparse solutions:

$$R_{\text{Elastic Net}} = (1 - \alpha)\|\boldsymbol{\beta}_z\|_2 + \alpha\|\boldsymbol{\beta}_z\|_{2,1}. \tag{12}$$

The Minimax Concave Penalty (MCP) as described in Zhang et al. (2010) is the fifth and final method employed in this paper. Its variable selection is nearly unbiased and leads to consistent variable selection even in the case of $p \gg N$ in contrast to LASSO:

$$R_{\text{MCP}} = p_z^{MCP} \|\boldsymbol{\beta}_z\|_{2,1}. \tag{13}$$

The switch from the penalized to the unpenalized form is non-smooth and discontinuous with knots at $\gamma\lambda$:

$$p_z^{MCP}(\beta; \lambda, \gamma) = \begin{cases} \lambda\|\boldsymbol{\beta}_z\|_{2,1} + (2\gamma)^{-1}\|\boldsymbol{\beta}_z\|_2 & \|\boldsymbol{\beta}_z\|_{2,1} \leq \gamma\lambda \\ (2)^{-1}\lambda^2\gamma & \|\boldsymbol{\beta}_z\|_{2,1} > \gamma\lambda. \end{cases} \tag{14}$$

The $\ell_{2,1}$ norm controls the complexity of the nonlinearities in the model. It pushes the entries towards zero and thus enforces sparsity at the group level. It drops or selects the smoothing spline of each characteristic simultaneously. Hence, it encourages sparsity in the number of characteristics included, rather than in the number of nonlinear expansions of characteristics.

After defining the regularization terms, I can combine a generalized penalty with the baseline model of equation (7) used for Tables (5)-(8). Minimizing the sum of the squared residuals leads to:

$$\hat{\beta}_s = \underset{\beta_{s,z}}{\arg\min} \sum_{i=1}^{N} \left(R_{it+1} - \sum_{z=1}^{Z} \beta'_{s,z} \cdot g_z(\tilde{C}_{it,z}, k) \right)^2 + \lambda_s \sum_{z=1}^{Z} R_{type}(\beta_{s,z}), \tag{15}$$

for each stage $s \in \{1, 2\}$, where $\hat{\beta}_s$ is a $Z \times (K+2)$ matrix of estimated coefficients with row columns $\beta_{s,z}$ of dimension $(K+2) \times 1$ and the hyperparameter λ_s. The subscript s of λ and β indicates the stage of the model because, the adaptive group model has a second stage and thus a second tuning parameter λ_2 exists. Each group z is necessary to cluster smooth functions of a particular characteristic C_z together. In both stages and with all regularization types, I select the penalization parameter based on the minimum 10-fold cross-validated Bayesian Information Criterion following Yuan and Lin (2006).

3.3 Time-variation

Another question that arises is whether this set of variables is time-invariant or not. McLean and Pontiff (2016) for instance, test 97 cross-sectional anomalies of 79 studies and find that long-short portfolio returns are 32% smaller out-of-sample due to publication-informed trading, implying a significant variation in the informative set of variables over time. Penasse (2017) goes one step further and shows that ignoring time variation, whenever present, can even lead to false discoveries.

Given the ambiguity of many dynamics found to impact the time variation of financial market anomalies such as publication dates or funding constraints, this paper aims at identifying a subset of characteristics with predictive power per time point. I do so by generalizing the baseline model of equation 15, to a time-varying as follows:

$$\hat{\beta}_{t,s} = \underset{\beta_{t,s,z}}{\arg\min} \sum_{i=1}^{N} \left(R_{it+1} - \sum_{z=1}^{Z} \beta'_{t,s,z} \cdot g_{t,z}(\tilde{C}_{it,z}, k) \right)^2 + \lambda_{t,s} \sum_{z=1}^{Z} R_{type}(\beta_{t,s,z}), \tag{16}$$

where $\hat{\beta}_t$ is a $T \times Z(K+2)$ matrix of estimated coefficients with row columns $\beta_{t,z}$ of dimension $T \times (K+2)$. For this reason, Table 9 presents a rolling window model for a constant estimation window of 20 years and a holding period of one year.

4 Data and Market Anomalies

I obtain daily, monthly, and annual equity returns from the Center for Research in Security Prices database. Following standard conventions, I use domestically incorporated ordinary common stocks listed on NYSE, AMEX, or NASDAQ, and I include delisting returns following the methodology of Shumway and Warther (1999).
Firm characteristics include quarterly, and annual balance sheet information from the Standard & Poor's Capital IQ Compustat industrial database. To mitigate a potential survivorship bias due to backfilling, I exclude firms with less than three years of Compustat data. I delete duplicates based on the global company identifier, fiscal year and stock level. The time horizon spans from January 1965 to December 2018 and is a result of data availability. These restrictions lead to an average of 1.4 million observations per market anomaly or 127 million data points in total.
The risk-free rate is the 1-month TBill return from Ibbotson and Associates. EDHEC Risk Institute provides the long-short equity benchmark, which is the weighted average of the following six hedge fund indexes: CSFB, HFR, HF Net, Barclay, CISDM, and Altvest.
In the appendix, Table 2 includes the identifier, anomaly abbreviation, authors and year of publication of the corresponding anomaly that I analyze in this paper. Subsection C in the appendix states the details regarding the factor construction following the previously mentioned literature. I do not investigate any arbitrage opportunities that result from violations of the law of one price or any parity violations because these rare events are either short-lived or occur in different asset classes.
Successive anomaly portfolio construction uses the cross-sectional heterogeneity in specific factor characteristics. This dispersion allows me to compute decile factor portfolios from $P1$ to $P10$ for each anomaly factor. Accordingly, the difference in the first and the last

decile forms a long-short self-financing portfolio. I invert the factor signal whenever the original paper considers the lowest factor characteristic to construct the long position. Hence, the long leg of the portfolio includes the most undervalued and the short leg the most overvalued securities. All portfolios are value-weighted if not specified else to ensure that illiquid firms do not drive results. Table 3 reports these portfolio decile returns.
To generate comparative statics I measure these returns per month and adjust annual or quarterly accounting data as if they were monthly. For quarterly or yearly observations I do not rebalance monthly, but express the returns per month. Table 1 presents summary statistics of the joint distribution of both value-weighted and equally-weighted, dollar-neutral portfolio returns, and benchmark portfolio returns.

5 Empirical Results

I start the analysis by reporting monthly average returns for 10 value-weighted portfolios for each market anomaly, as seen in Table 3. Parentheses indicate standard errors and significant average returns at a confidence interval of 95% are printed in bold. All 90 market anomalies explain expected returns in the long leg (P10). On the short leg, around 90% are significant. Moreover, approximately half of the hedge returns show a t-statistic above two, and eight have a monthly average return above one percent.

5.1 Identification of Mispricing

As this paper focuses on the identification of a subset of truly predictive characteristics and, thus, a reduction of the anomaly zoo, it is important to distinguish them from anomalies that arise through other sources such as statistical biases, risk factors, or limits to arbitrage. Harvey et al. (2016) suggest a multiple testing framework, such as estimating the rate of type I errors, false discovery rate, or to consider a new t-statistic threshold of 3. As I am unable to verify the truthful number of repetitions per anomaly that I need in such a multiple testing framework, I estimate Jensen's alphas with respect to a benchmark risk model using an extended sample period.
Unconditional Results: Anomalies identified through data mining or risk premiums should resolve in insignificant intercepts. Table 4 presents unconditional information ratios as well as spanning tests, using two common factor-mimicking portfolios (FMP). The statistics are reported on unadjusted hedge returns and in excess of the CAPM or FF3 factor model according to:

$$\alpha_z = \mathbb{E}[R_z^e] - \sum_{f=1}^{F} \beta_{f,z} \cdot \mathbb{E}[\mathrm{FMP}_f^e]. \tag{17}$$

The approximation $\sqrt{1}2 \cdot \hat{\alpha}/\hat{\sigma}_\epsilon$ annualizes Sharpe ratios and Information ratios. The standard errors of the Information ratios are estimated by moving blocks bootstrapping the data 10,000 times per month to retain the autocorrelation structure in the returns.
What is surprising, the majority of anomalies does not outperform the market in terms of Sharpe ratios, implying a statistical bias or time variation of expected returns. Depending on the periods measured, the value-weighted buy-and-hold Sharpe ratio for US stocks is around 0.5, given an equity premium of 8% and market volatility of 16% per year. Moreover, the t-statistic threshold of 3 is exceeded in 13 out of 89 cases. Considering risk-adjusted returns and thus Information ratios, it seems that more than half of all anomalies have market or factor loadings of negligible size. Figure 4 presents these estimated factor betas per anomaly sorted by CAPM betas to compare factor loadings across risk factor models. The anomaly "Illiquidity" (ill) is the perfect example of the joint hypothesis problem. It lacks a CAPM beta but significantly loads on the SMB and HML factors. This indicates that returns are higher due to greater risk premiums on small or illiquid firms, and thus, "ill" is not (entirely) a mispricing.
Collinearity: However, the big challenge is to understand the multivariate counterparts of univariate regressions as Cochrane (2011) advocates in his presidential address. This indicates that we have to move from univariate to multivariate regressions with the ability to filter for the important covariates, given that they are multicollinear as suggested by the pairwise correlation plot, see Figure 1. These heat maps present pairwise Pearson correlations between rank-normalized firm characteristics. Dark blue (red) indicates positively (negatively) correlated features, and uncorrelated variables are transparent. Features are ordered by complete hierarchical clustering. The correlation across characteristics is relatively strong for some clusters. Even after controlling for micro stocks at the NYSE 10% level in Figure 2 as well as at the NYSE 20% level in Figure 3, these co-movements seem even to increase. I try to solve this problem by applying penalized multivariate regression as discussed in the next subsection.

5.2 In-Sample Comparison

The baseline setting includes all firms of the Compustat North America data. In addition, I restrict the set of firms to a subset with a market capitalization above the NYSE 10%, and NYSE 20% percentile, to exclude effects that manifest due to the limits of arbitrage. Hou, Mo, Xue, and Zhang (2019), for instance, replicate a set of well-established anomalies and find that most returns are concentrated in micro-caps or illiquid stocks.
Conditional Linear Mean Function: Table 5 reports selected characteristics filtered by the penalized regressions from a total universe of 90 market anomalies. The columns report model selection results for different steps, firms, and regularization terms. Charac-

teristics are rank-normalized per month and the predictive model is estimated via pooling, omitting the time index as suggested by Cochrane (2011).

I present four different regularization approaches that turned out to be useful in the asset pricing literature: LASSO, Elastic Net, adaptive LASSO, and SCAD. I set the α parameter to 0.5 which weights the LASSO and the RIDGE penalty, and hence equally weights sparsity and shrinkage. For robustness, I also set α to 0.25 (0.75) and thus allow for more (less) shrinkage than sparsity. However, the results are similar, and the Elastic Net converges to LASSO already at an α level of 0.5, given the low signal-to-noise ratio. As a consequence, I omit the LASSO results in this table and present LASSO in the online appendix. Additionally, but not represented in this paper, I find that MCP penalties result in almost identical feature selections as the SCAD penalties. This finding is intuitive given their similarities within their functional form.

Models (1) and (3) yield similar in-sample Sharpe ratios with less selected features compared with Model (2). Surprisingly, once controlled for micro stocks, as suggested by Asness, Frazzini, Israel, Moskowitz, and Pedersen (2018), Sharpe ratios increase across all models. Moreover, in terms of Sharpe ratios the SCAD and the adaptive LASSO model exceed the Elastic Net model although the number of covariates is 40% smaller.

Furthermore, by comparing the multivariate in-sample results with the univariate regressions of Table 4, I observe that 38 characteristics have significant intercepts after controlling for the FF3 model. Out of those, 31 variables are selected in at least one of the in-sample linear models. However, only six characteristics are selected by all nine linear models, namely: **Net Working Capital Changes, Dividend-to-Price ratio, Earnings per Share, Industry Relative Reversals, Net Issuance and Return Volatility**. This circumstance once more highlights the difference between a univariate analysis and a multivariate analysis that conditions on multiple firm characteristics enhanced with regularization.

Overall, I conclude that the Elastic Net, as well as LASSO, does not introduce enough sparsity to sufficiently reduce features within individual stock forecasting models. Furthermore, controlling for micro-cap stocks within both thresholds increases the in-sample explanatory power measured by the Sharpe ratio independent of the number of selected covariates.

Conditional Nonlinear Mean Function: Given the evidence that sparser models have greater explanatory power once micro stocks are excluded, I account for nonlinearities in the cross-section of firm characteristics to further increase model accuracy as well as sparsity as represented in Table 6. Section 3 describes the spline approach in more detail and states the equivalence to local regressions. The main intuition behind spline interpolation is to split the cross-section into subsets and to model each subset individually. Therefore, extreme values in certain subsets do not affect other areas of the cross-section and thus reduce biases. This methodology is the regression equivalent to multivariate independent portfolio sortings, but it allows for presenting results of more than three characteristics. Knots are the percentile cutoffs that are necessary to form portfolio sorts. Given nine knots, for instance, implies sorting into 10 portfolios. I employ 19 cutoffs or knots in the in-sample analysis which is equivalent to sorts into 20 portfolios. For robustness, I increase the number of knots to 24. The full set of results can be found in the online appendix. The magnitude of knots depends on the sample size. Because I omit the time index, each firm-time observation becomes an independent cross-sectional observation. Hence, pooling the entire sample period allows for a detailed nonparametric estimation of the cross-section with a larger number of knots. In the choice of the magnitude of knots, I follow Freyberger et al. (2020).

Table 6 again represents the four main regularization models allowing for nonlinear relationships in the cross-section. Despite the use of different model specifications, information sets selected show very little variation. Column 1 represents the SCAD model, which selects only three variables: **52-Week High, Log of stock price** and **Industry Relative Reversals**. This finding is identical to the NYSE 10% restriction, which is therefore omitted from the table. The former two characteristics are selected across all models, linear or nonlinear. In 14 out of 15 cases, the variable 52-Week High is selected. However, the last column presents the SCAD model for the most restricted sample, which excludes the feature Industry Relative Reversals. Although this model consists of only two variables, its in-sample Sharpe ratio is only slightly lower. In comparison, the linear SCAD model of table 5 selects 65 characteristics with only a slightly higher Sharpe ratio. Columns (3)-(5) have identical results for NYSE 10% restrictions and are thus only presented in the online appendix. As in the linear case, the Elastic Net selects far more variables without significantly increasing the explanatory power in the form of in-sample Sharpe ratios.

Another interesting fact is the difference between the nonlinear LASSO model and its adaptive counterpart. The only variable that is excluded through the second and more comprehensive estimation procedure is the Earnings-to-Price ratio, with an only slightly lower Sharpe ratio.

The linear in-sample model comparison suggests that sufficiently sparser models explain almost as much as large (non-sparse) models. This finding is even more pronounced for the nonlinear model in-sample results. In summary, between two and nine variables are enough to explain the in-sample variation and are hence a tremendous reduction of features.

The literature suggests an upper bound of around 13 risk factors, which seems to be in line with the set of features found by the conditional models presented in this subsection. Nevertheless, to understand whether the excluded characteristics lead to over-fitting as theoretically predicted by the curse of dimensionality, we need to consider out-of-sample analysis, which we discuss in the next section.

5.3 Out-of-Sample Comparison

While the previous subsection selected firm characteristics are independently meaningful for the entire sample period, the out-of-sample comparison should provide information about estimation uncertainty. The previous model fits might be subject to overfitting.
For this reason, I perform a model selection for a certain period, as, for instance, from 1965 to 1990. I re-estimate the coefficients for the selected set of significant variables within the last 10 years of the sample period to obtain unpenalized parameter estimates. With these beta coefficients and new data starting from January 1991, for instance, I predict individual returns for the next month. Subsequently, I estimate the coefficients rolling forward one month and forecast using the updated values of the previously selected information set, therefore conducting expanding return prediction. I sort predicted returns in 10 portfolios and form value-weighted hedge portfolios going long the stocks in the highest sort and shorting the stocks of the lowest sort. The different out-of-sample periods, number of selected characteristics and regularization terms are presented for the linear case in Table 7. Table 8 presents the results for the nonlinear model. Both tables also report the first four moments of monthly percentage returns, annualized Sharpe ratios, turnover, and predictive slopes. They also report the R^2 for the hedge portfolios in Panel A, the long legs in Panel B and the short legs in Panel C. To calculate turnover, I follow Koijen, Moskowitz, Pedersen, and Vrugt (2018) for both turnover measures, which is defined as $TO_t = \frac{1}{4}\sum_t^{N_t} |w_{it} - w_{it-1}|$ or as $TO_{R_t} = \frac{1}{4}\sum_t^{N_t} |w_{it} - w_{it-1}(1 + r_{it})|$, where w_{it} is the value-weight of stock i at time t.
Conditional Linear Mean Function: In the first to third columns of Table 7, the model selection period ranges from 1965 to 1982. Thereafter, given the selected variables, out-of-sample predictions are made for the next successive 120 months. The SCAD model selects 19 variables within this period, 17 less than the Elastic Net model and five more compared with the two step LASSO penalization. However, its out-of-sample long-short portfolio suggests a monthly return of 2.9% and a Sharpe ratio of 2.52, which is slightly higher than its competitors.
In the same period, I also estimate a pure LASSO model, which as in the in-sample period yields identical selected features as the Elastic Net and is thus omitted from the table. Furthermore, I omit the presentation of the Elastic Net for further estimation periods due to its overfitting tendency and its consequent poor out-of-sample performance.

Comparing the model fits for the remaining estimation periods reveals a decaying out-of-sample performance. Indeed, the adaptive LASSO model of column (6) yields an out-of-sample Sharpe ratio similar to a market portfolio of 0.5, if we assume an equity premium of 8% and market volatility around 16%. Overall, the skewness is mostly positive for hedge and long positions, whereas it is negative for the short leg.
To further validate the quality of individual return predictions, I use average betas and R^2, as suggested by Lewellen (2014). I estimate these measures by regressing individual ex-post returns on individual predicted returns per month. In both cases, a value of 1 represents a perfect one-to-one relationship. R^2 ranges between 3% and 5%, indicating that the SCAD penalty has slightly greater predictive power.
Decomposing hedge returns into the long and short returns reveals that most of their performance comes from the long leg. Indeed, besides column (3), the SCAD model, returns are almost 0 or even positive, leading to a smaller long-short portfolio return. On the other hand, volatility is always larger on the long leg.
Overall, the SCAD model seems to select the right features more often than its competing regularization terms. Despite the use of expanding prediction for the remaining 10 years, average percentage returns and Sharpe ratios of Panel A are surprisingly large and above the mean-variance efficient tangency portfolio of 0.5.
Conditional Nonlinear Mean Function: In the nonlinear case, Table 8 presents the out-of-sample results. I chose identical data sample restrictions as for the linear model to compare both approaches. We can observe that once we allow for quadratic terms, the model selects far fewer characteristics. In the most recent period, both the adaptive LASSO as well as the SCAD model have higher Sharpe ratios with only 3, respectively 2 features: **Industry Relative Reversals, Log of stock price, and Lagged Momentum,** compared to 18 and 30 active characteristics. Moreover, out-of-sample betas and R^2 are larger throughout the nonparametric estimation given similar turnover. For the various estimation periods, the Group SCAD model selects 1, 4 and 2 characteristics, respectively 2, 3 and 3 for the adaptive group LASSO model. As a result, we can conclude that the predictive power of market anomalies and the selected subset of the first 20 years relative to the entire sample varies across time.
Rolling Conditional Nonlinear Mean Function: Consequently, I conduct a rolling estimation with a variable selection window of 20 years. Thereafter, I predict returns for each consecutive month, fixing the selected variable set for one year before I redo variable selection. Table 9 shows the out-of-sample results for rolling variable selection. All various model specifications lead to significantly larger out-of-sample Sharpe ratios than their time-invariant complement of Table 8. As the number of selected variables per column changes from estimation to estimation, I present the selected characteristics per year of

the SCAD regularization term in Figure 5. Four characteristics seem to persist over time, while 16 such as the Bid-Ask spread vary tremendously through time.

In sum, differences across models are substantial. The number of variables selected ranges from 1 to 30 out of 90 characteristics. Nonetheless, a stable intersection between the characteristics selected across models exists. SCAD regularization terms are more stable in terms of parameter choices than their counterparts. It also selects the sparest model with economically significant out-of-sample Sharpe ratios. Adaptive group LASSO seems to be a consistent alternative, whereas the Elastic Net tends to select many variables, and the out-of-sample results suggest that it overfits the model. Quadratic terms seem to matter substantially performance-wise as well as in terms of explanatory power.

6 Conclusion

The large number of firm characteristics emerging from recent literature and its multivariate combinations eventually created a multidimensional challenge, as documented by Cochrane (2011). In this paper, I suggest an econometric technique that overcomes the curse of dimensionality and dissects the independent information content that firm characteristics provide for expected returns. In particular, I employ a single-stage regression enhanced by a smoothly clipped absolute deviations (SCAD) penalty at the group level. This regularization term does not penalize critical firm characteristics and therefore allows them to survive the dimension reduction.

Starting with a novel database of 90 market anomalies, 38 characteristics have significant intercepts after controlling for the FF3 model. Out of those, 31 variables are selected in at least one of the in-sample linear models. However, only six characteristics are selected by all nine linear models. Applying a nonparametric return forecasting model enhanced with various regularization terms, such as adaptive LASSO, SCAD, or MCP, as described in subsection 3.2, leads to a variable selection of two to nine characteristics within the full sample. By accounting for nonlinear relationships between expected returns and rank-normalized firm characteristics, I can improve the prediction power even further and achieve a tremendous reduction in variables. I find that three characteristics are consistent across various model specifications with different penalization terms and number of stages.

To mitigate overfitting, I then conduct an out-of-sample comparison. For various estimation periods, the average number of characteristics selected by the nonlinear models vary between in- and out-of-sample. Given this evidence, we can conclude that the predictive power of market anomalies as well as the selected subset of the first 20 years relative to the entire sample alternates massively across time.

As a result, I move forward by conducting a rolling estimation with a variable selection window of 20 years and a forecasting horizon of only one year. I apply this rolling estimation on the model with nearly unbiased estimates, which is the model with the SCAD penalty as defined by equation (10). The estimation leads to significantly larger out-of-sample Sharpe ratios than their time-invariant counterpart. As the number of selected variables per column changes from estimation to estimation, I present the selected characteristics of the SCAD regularization term per year. Four characteristics namely A2ME, REV1M, MOM18M, and IndRev1M persist over time, while 16, such as the Bid-Ask spread vary tremendously through time.

This paper studied various regularization terms of nonparametric return forecasting models to identify a reduced set of characteristics associated with expected return spreads in the cross-section. I find that these models can overcome the challenge imposed by multidimensional firm characteristics. The models employed are outlier-resistant as they use the empirical cumulative distribution function of firm characteristics, computationally efficient, and able to handle high-dimensional data. Out of the five tested model variations applying different regularization terms, the SCAD penalty is more stable than their counterparts. Additionally, it has the highest economic significance with remarkable out-of-sample Sharpe ratios. Adaptive group LASSO and MCP seem to be consistent alternatives, although their Sharpe ratios are smaller and estimates are less consistent. In contrast, the Elastic Net tends to select too many variables, and the out-of-sample results suggest overfitting. Furthermore, quadratic terms seem to have a substantial impact on performance as well as on explanatory power. We can hence conclude that nonlinearities matter for the cross-section of expected returns and that firm characteristics are not consistent over time.

References

Adrian, T., Etula, E., & Muir, T. (2014). Financial Intermediaries and the Cross-Section of Asset Returns. *Journal of Finance*, *69*(6), 2557–2596. doi: 10.1111/jofi.12189

Asness, C., Frazzini, A., Israel, R., Moskowitz, T. J., & Pedersen, L. H. (2018). Size matters, if you control your junk. *Journal of Financial Economics*, *129*(3), 479–509.

Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The journal of Finance*, *59*(4), 1481–1509.

Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, *65*(1), 179–216.

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

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of political Economy*, *107*(2), 205–251.

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

Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Schaumburg, E. (2018). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics*, 1–47.

Chen, A. Y., & Velikov, M. (2019). Accounting for the anomaly zoo: A trading cost perspective. *Finance Down Under 2019 Building on the Best from the Cellars of Finance*.

Chen, L., Pelger, M., & Zhu, J. (2019). Deep learning in asset pricing. *Available at SSRN 3350138*.

Chinco, A., Clark-Joseph, A. D., & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, *74*(1), 449–492.

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

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., & Uppal, R. (2019). A transaction-cost perspective on the multitude of firm characteristics. *Available at SSRN 2912819*.

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

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

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

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*.

Green, J., Hand, J. R. M., & Zhang, X. F. (2017). The Characteristics that Provide

Independent Information about Average US Monthly Stock Returns. *Review of Financial Studies*, *30*(12), 4389–4436. doi: 10.1093/rfs/hhx019

Gu, S., Kelly, B., & Xiu, D. (2018). Empirical asset pricing via machine learning. *working paper*.

Haddad, V., Kozak, S., & Santosh, S. (2018). The Economics of Factor Timing. *WP*.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hou, K., Mo, H., Xue, C., & Zhang, L. (2019). Which factors? *Review of Finance*, *23*(1), 1–35.

Huang, J., Horowitz, J. L., & Wei, F. (2010). Variable selection in nonparametric additive models. *Annals of statistics*, *38*(4), 2282.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*.

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2018). Carry. *Journal of Financial Economics*, *127*(2), 197–225.

Kozak, S., Nagel, S., & Santosh, S. (2019). Shrinking the cross-section. *Journal of Financial Economics*.

Lewellen, J. (2014). The cross section of expected stock returns. *Forthcoming in Critical Finance Review*.

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

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of monetary Economics*, *15*(2), 145–161.

Novy-Marx, R., & Velikov, M. (2016). A Taxonomy of Anomalies and Their Trading Costs. *Review of Financial Studies*, *29*(1), 104–147. doi: 10.1093/rfs/hhv063

Penasse, J. (2017). Understanding anomaly decay. *WP*.

Santos, T., & Veronesi, P. (2004). Conditional betas. *working paper*.

Shumway, T., & Warther, V. A. (1999). The delisting bias in crsp's nasdaq data and its implications for the size effect. *The Journal of Finance*, *54*(6), 2361–2379.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Varrette, S., Bouvry, P., Cartiaux, H., & Georgatos, F. (2014, July). Management of an academic hpc cluster: The ul experience. In *Proc. of the 2014 intl. conf. on high performance computing & simulation (hpcs 2014)* (pp. 959–967). Bologna, Italy: IEEE.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.

Zhang, C.-H., et al. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, *38*(2), 894–942.

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

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

Appendix

A Tables

Table 1: **Summary Statistics**
This table reports summary statistics of characteristics and characteristic sorted portfolio returns. Panel A presents average value-weighted monthly hedge returns in percentage. Stocks are sorted into extreme deciles based on their rank-normalized characteristic at the end of each month. Panel B presents the mean and the standard deviation of the underlying characteristics.

Panel A: Long-Short Portfolio Returns

Anomalies	Mean	Median	Std. Dev.	Skewness	Kurtosis	Min	Max	Obs.
Full Sample	0.33	0.30	4.95	0.03	3.87	-24.43	24.42	57,978
Expansion	0.32	0.29	4.71	0.05	4.22	-23.20	23.47	50,539
Depression	0.41	0.45	6.30	-0.02	1.29	-17.55	18.12	7,439
Pre 2003	0.42	0.42	5.11	0.01	3.86	-23.56	23.66	40,790
Post 2003	0.12	0.06	4.44	0.06	2.14	-15.87	15.93	17,188
CRSP-Rf	0.70	1.02	5.24	-0.56	2.17	-27.54	20.48	1,417,151

Panel B: Descriptive Statistics of Characteristics

Anomalies	Mean	StDev	Anomalies	Mean	StDev	Anomalies	Mean	StDev
a2me	2.71	8.06	eps	1.16	2.53	pchgm_pchsale	-0.20	28.17
agr	0.14	0.51	ga_eps	-0.04	2.54	pchquick	0.11	2.29
at	4.14	40.82	gma	0.39	0.34	pchsale_pchrect	-0.09	32.09
bab	1.01	0.36	herf	0.08	0.07	pcm	-0.52	98.08
baspread	0.04	0.05	hire	0.11	1.63	pm	-0.93	101.22
bm	0.80	0.87	ill	0.00	0.00	pm_ia	2.55	101.22
bm_ia	-2.84	8.50	indmom6m	0.00	0.02	prc52	0.81	0.20
c2d	0.12	0.99	indmom6m_ia	0.10	0.17	prof	1.06	18.74
cashdebt	0.13	0.85	indrev1m	-0.00	0.14	ps	25.00	437.62
cashpr	-3.29	2152.32	invest	0.08	0.22	q	0.70	1.60
chaTO	-0.01	0.30	ipm	-0.97	99.62	quick	3.77	191.80
chatoia	-0.04	2.47	ldp	0.43	0.96	r62	0.06	0.40
chcsho	0.10	0.39	lev	1.75	5.18	retvol	0.03	0.03
chempia	-0.16	1.95	lgr	0.24	1.97	rev1m	-0.01	0.16
chinv	0.01	0.06	lprc	2.33	1.27	roa2	0.02	0.19
chlogso	0.03	0.14	maxret	0.07	0.08	roe	0.06	2.96
chmom	-0.00	0.62	mom11m	0.14	0.66	roic	0.01	1.34
chnwc	0.01	0.10	mom12m	0.16	0.70	salecash	0.12	2.24
chpm	-0.11	94.49	mom18m	-0.08	0.43	salerec	14.92	136.20
chpmia	-0.04	95.83	mom36m	0.35	1.22	sat	1.14	0.91
cto	1.28	1.13	mom6m	0.07	0.44	sat_ia	-0.06	1.85
currat	5.43	461.58	mom6m_l6	0.08	0.44	sgr	0.32	25.94
daily_Beta	0.92	0.85	mve	5.40	2.21	sp	2.27	4.27
debt2p	0.69	2.12	mve_firm	2.81	15.60	std_dolvol	0.80	0.40
depr	0.31	3.96	mve_firm_ia	0.73	15.24	std_dolvol36m	1.41	8.19
dolvol	11.65	2.99	mve_ia	0.69	14.10	std_turn	4.06	14.61
dy	0.02	0.05	mve_jun_log	5.42	2.18	tan	0.53	0.14
ebitrev	-0.93	101.22	ni	0.05	0.93	tang	0.53	0.14
egr	0.24	8.00	operprof	0.32	17.56	turn	1.08	1.76
ep	-0.01	0.42	pchcurrat	0.08	2.07	zerotrade	1.00	2.85

Table 2: **Defining Market Anomalies**
This table reports the name, author, and sample period of the original study of the 90 anomalies constructed in this paper.

Id	Abbrev	Anomaly	Paper	Period
1	a2me	A2ME	Bhandari	1946-1981
2	age.m	Firm Age	Barry, Brown	1931-1982
3	agr	Asset growth	Cooper et al.	1963-2003
4	at	Total Assets	Gandhi, Lustig	1970-2013
5	bab	Beta	Frazzini, Pedersen	1984-2012
6	baspread	Bid-Ask spread	Amihud, Mendelson	1961-1980
7	bm	Book-to-market	Rosenberg et al.	1963-1990
8	bm.ia	Ind.-adj. book-to-market	Asness et al.	1963-1998
9	c2d	C2D	Casey, Bartczak	1971-1982
10	cashdebt	Cash Flow to debt	Ou, Penman	1965-1983
11	cashpr	Cash productivity	Chandrashekar, Rao	1962-2003
12	chato	Asset Turnover	Soliman	1984-2002
13	chatoia	Ind.-adj. Chg. in asset turnover	Soliman	1984-2002
14	chcsho	Chg. in shares outstanding	Pontiff, Woodgate	1970-2003
15	chempia	Ind.-adj. Chg. in employees	Asness et al.	1963-1998
16	chinv	Chg. in inventory	Thomas, Zhang	1970-1997
17	chlogso	ΔSO	Fama, French	1963-2005
18	chmom	Chg. in 6-month momentum	Gettleman, Marks	1926-2003
19	chnwc	Chg. in Net Working Capital	Soliman	1984-2002
20	chpm	Chg. in Profit	Soliman	1984-2002
21	chpmia	Ind.-adj. Chg. in profit margin	Soliman	1984-2002
22	cto	CTO	Baker	1983-1993
23	currat	Current ratio	Ou, Penman	1965-1983
24	daily.Beta	Beta daily	Lewellen, Nagel	1964-2001
25	debt2p	Debt2P	Ramaswamy	1936-1977
26	depr	Depreciation divided PP&E	Holthausen, Larcker	1978-1988
27	dolvol	Dollar trading volume	Chordia et al.	1966-1995
28	dy	Dividend to price	Litzenberger et al.	1931-1951
29	ebitrev	Profit Margin	Soliman	1984-2002
30	egr	Growth in common equity	Richardson et al.	1962-2001
31	ep	Earnings to price ratio	Basu	1957-1971
32	eps	Earnings per share	Basu	1957-1971
33	ga.eps	Earnings Consistency	Alwathainani	1971-2002
34	gma	Gross profitability	Novy-Marx	1963-2010
35	herf	Industry sales concentration	Hou, Robinson	1963-2001
36	hire	Employee growth rate	Bazdresch, Belo, Lin	1965-2010
37	ill	Illiquidity	Amihud	1964-1997
89	indmom6m	Industry Momentum	Moskowitz, Grinblatt	1963-1995
38	indmom6m.ia	Industry Momentum	Grinblatt et al.	1963-1995
39	indrev1m	Industry Relative Reversals	Da, Liu, Schaumurg	1982-2009
40	invest	CAPEX and inventory	Chen, Zhang	1972-2006
41	ipm	IPM	NA	NA-NA
42	ldp	LDP	Litzenberger et al.	1936-1977

Id	Abbrev	Anomaly	Paper	Period
43	lev	Leverage	Bhandari	1946-1981
44	lgr	Growth in long-term debt	Richardson et al.	1962-2001
45	lprc	Price	Blume, Husic	1932-1971
46	maxret	Maximum daily return	Bali et al.	1962-2005
47	mom11m	r12-2	Fama, French	1963-1993
48	mom12m	Cash Flow volatility	Jegadeesh	1964-1989
49	mom18m	Momentum-Reversal	Jegadeesh, Titman	1964-1989
50	mom36m	36-month momentum	Jegadeesh, Titman	1926-1982
51	mom6m	Momentum-Volume	Jegadeesh, Titman	1964-1989
52	mom6m.l6	Lagged Momentum	Novy-Marx	1926-2010
53	mve	Size	Banz	1926-1975
54	mve.firm	LME	Fama, French	1963-1990
55	mve.firm.ia	LME Ind.-adj.	Asness et al.	1963-1998
56	mve.ia	Ind.-adj. size	Asness et al.	1963-1998
57	mve.jun.log	Size	Banz	1926-1975
58	ni	Net Issuance	Fama, French	1973-2002
59	operprof	Operating profitability	Fama, French	1977-2003
60	pchcurrat	Chg. in current ratio	Ou, Penman	1965-1983
61	pchgm.pchsale	Chg. in gross margin - Chg. in sales	Abarbanell, Bushee	1974-1988
62	pchquick	Chg. in quick ratio	Ou, Penman	1965-1983
63	pchsale.pchrect	Chg. in sales - Chg. in A/R	Abarbanell, Bushee	1974-1988
64	pcm	PCM	Gorodnichenko, Weber	1994-2009
65	pm	PM	Soliman	1984-2002
66	pm.ia	PM Ind.-adj.	Soliman	1984-2002
67	prc52	52-Week High	George, Hwang	1963-2001
68	prof	Prof	Ball et al.	1963-2013
69	q	Tobin's Q	Wernerfelt et al.	1960-1977
70	quick	Quick ratio	Ou, Penman	1965-1983
71	r62	r6-2	Jegadesh, Titman	1964-1989
72	retvol	Return volatility	Ang et al.	1986-2000
73	rev1m	Short-Term Reversal	Jegadeesh	1934-1987
74	roa2	ROA	Balakrishnan et al.	1976-2005
75	roe	Return-on-Equity	Haugen, Baker	1979-1993
76	roic	Return on invested capital	Brown, Rowe	1970-2005
77	salecash	Sales to cash	Ou, Penman	1965-1983
78	salerec	Sales to receivables	Ou, Penman	1965-1983
79	sat	SAT	Soliman	1984-2002
80	sat.ia	SAT Ind.-adj.	Soliman	1984-2002
81	sgr	Sales growth	Lakonishok et al.	1963-1990
82	sp	Sales/Price	Barbee et al	1979-1991
83	std.dolvol	Volatility of liquidity	Chordia et al.	1966-1995
84	std.dolvol36m	Volume Variance	Chordia et al.	1966-1995
85	std.turn	Volatility of liquidity	Chordia et al.	1966-1995
86	tan	Tan	Hahn, Lee	1973-2001
87	tang	Debt capacity/firm tangibility	Almeida, Campello	1973-2001
88	turn	Share turnover	Datar et al.	1962-1991
90	zerotrade	Zero trading days	Liu	1960-2003

Table 3: **Portfolio Returns per Market Anomaly**

This table reports monthly average returns for 10 value-weighted portfolios per market anomaly. Parentheses indicate standard errors and significant average returns at a confidence interval of 95% are printed in bold type. The sample period spans from January 1965 to December 2018.

Anomaly	P1	s.e.	P2	s.e.	P3	s.e.	P4	s.e.	P5	s.e.	P6	s.e.	P7	s.e.	P8	s.e.	P9	s.e.	P10	s.e.	P10-P1	s.e.
a2me	**0.86**	(0.21)	**0.88**	(0.18)	**0.87**	(0.18)	**0.98**	(0.17)	**0.97**	(0.18)	**1.01**	(0.17)	**1.00**	(0.18)	**1.12**	(0.20)	**1.04**	(0.22)	**1.13**	(0.26)	0.26	(0.21)
age.m	**1.02**	(0.30)	**0.94**	(0.27)	**1.06**	(0.26)	**1.04**	(0.26)	**1.21**	(0.26)	**1.12**	(0.25)	**1.02**	(0.24)	**1.02**	(0.21)	**1.05**	(0.21)	**0.99**	(0.18)	-0.03	(0.21)
agr	**0.73**	(0.24)	**0.90**	(0.22)	**0.89**	(0.19)	**0.93**	(0.17)	**0.92**	(0.17)	**0.97**	(0.16)	**0.95**	(0.16)	**1.08**	(0.18)	**1.11**	(0.19)	**1.12**	(0.23)	**0.38**	(0.15)
at	**0.59**	(0.32)	**0.89**	(0.30)	**1.09**	(0.28)	**1.10**	(0.27)	**1.10**	(0.25)	**1.13**	(0.24)	**1.03**	(0.22)	**1.08**	(0.21)	**0.92**	(0.19)	**0.86**	(0.16)	0.27	(0.26)
bab	**0.86**	(0.15)	**0.98**	(0.13)	**1.03**	(0.14)	**0.96**	(0.15)	**0.94**	(0.16)	**0.91**	(0.18)	**0.86**	(0.20)	**0.89**	(0.22)	**0.80**	(0.26)	**0.61**	(0.33)	-0.25	(0.30)
baspread	0.48	(0.43)	0.41	(0.36)	**0.71**	(0.31)	**0.83**	(0.28)	**1.14**	(0.25)	**1.05**	(0.21)	**1.04**	(0.19)	**0.90**	(0.17)	**0.96**	(0.15)	**0.83**	(0.14)	0.34	(0.40)
bm	**0.90**	(0.21)	**0.87**	(0.19)	**0.91**	(0.18)	**0.94**	(0.18)	**0.89**	(0.17)	**0.98**	(0.18)	**0.91**	(0.18)	**1.19**	(0.20)	**1.03**	(0.23)	**1.24**	(0.26)	0.34	(0.22)
bm.ia	**0.90**	(0.19)	**0.79**	(0.21)	**0.85**	(0.19)	**0.84**	(0.19)	**1.05**	(0.19)	**0.97**	(0.19)	**0.99**	(0.19)	**1.03**	(0.18)	**1.00**	(0.19)	**1.24**	(0.22)	**0.33**	(0.15)
c2d	**0.62**	(0.34)	**0.85**	(0.26)	**0.80**	(0.23)	**0.95**	(0.20)	**0.92**	(0.17)	**0.94**	(0.17)	**0.91**	(0.18)	**0.99**	(0.17)	**0.91**	(0.18)	**0.91**	(0.21)	0.29	(0.23)
cashdebt	0.55	(0.34)	**0.83**	(0.25)	**0.85**	(0.23)	**0.94**	(0.20)	**0.83**	(0.17)	**1.00**	(0.17)	**0.91**	(0.18)	**0.94**	(0.17)	**0.93**	(0.18)	**0.93**	(0.22)	0.37	(0.23)
cashpr	**0.75**	(0.19)	**0.93**	(0.18)	**0.92**	(0.20)	**1.06**	(0.20)	**1.23**	(0.20)	**1.07**	(0.19)	**1.08**	(0.19)	**0.99**	(0.20)	**1.01**	(0.19)	**1.10**	(0.18)	**0.35**	(0.13)
chato	**0.83**	(0.23)	**0.85**	(0.21)	**0.85**	(0.19)	**0.86**	(0.17)	**0.89**	(0.17)	**0.92**	(0.18)	**0.99**	(0.18)	**1.08**	(0.18)	**1.00**	(0.20)	**1.00**	(0.22)	0.17	(0.15)
chatoia	**0.90**	(0.23)	**0.79**	(0.20)	**0.84**	(0.18)	**0.83**	(0.18)	**0.97**	(0.17)	**0.93**	(0.18)	**0.99**	(0.17)	**1.01**	(0.18)	**0.97**	(0.20)	**1.06**	(0.22)	0.16	(0.13)
chcsho	**0.60**	(0.20)	**0.82**	(0.20)	**0.83**	(0.20)	**0.96**	(0.20)	**1.08**	(0.19)	**0.89**	(0.18)	**0.97**	(0.17)	**0.98**	(0.17)	**1.07**	(0.17)	**1.24**	(0.17)	**0.64**	(0.11)
chempia	**0.79**	(0.20)	**0.89**	(0.19)	**0.92**	(0.18)	**0.93**	(0.18)	**1.00**	(0.17)	**0.93**	(0.18)	**0.91**	(0.17)	**0.86**	(0.18)	**1.07**	(0.21)	**0.83**	(0.23)	0.04	(0.13)
chinv	**0.83**	(0.24)	**0.87**	(0.21)	**0.79**	(0.20)	**0.95**	(0.18)	**0.93**	(0.17)	**0.99**	(0.17)	**0.88**	(0.18)	**0.95**	(0.17)	**1.16**	(0.19)	**1.26**	(0.23)	**0.43**	(0.14)
chlogso	**0.48**	(0.21)	**0.72**	(0.20)	**0.83**	(0.21)	**1.04**	(0.20)	**0.99**	(0.19)	**0.91**	(0.18)	**0.90**	(0.17)	**0.98**	(0.17)	**1.05**	(0.17)	**1.29**	(0.18)	**0.81**	(0.12)
chmom	**0.87**	(0.24)	**0.83**	(0.20)	**0.86**	(0.18)	**0.76**	(0.18)	**0.96**	(0.18)	**0.97**	(0.17)	**1.00**	(0.18)	**1.09**	(0.20)	**1.11**	(0.23)	**1.09**	(0.28)	0.22	(0.21)
chnwc	**0.59**	(0.25)	**0.80**	(0.21)	**0.86**	(0.20)	**0.90**	(0.18)	**0.90**	(0.17)	**0.99**	(0.16)	**0.94**	(0.16)	**0.94**	(0.18)	**1.15**	(0.21)	**1.09**	(0.23)	**0.50**	(0.13)
chpm	**0.83**	(0.25)	**0.93**	(0.20)	**0.88**	(0.18)	**0.91**	(0.18)	**0.98**	(0.17)	**1.05**	(0.17)	**0.84**	(0.17)	**0.92**	(0.18)	**0.85**	(0.19)	**0.73**	(0.24)	-0.11	(0.14)
chpmia	**0.91**	(0.20)	**0.88**	(0.20)	**0.89**	(0.18)	**0.75**	(0.19)	**0.99**	(0.17)	**0.91**	(0.18)	**0.85**	(0.19)	**1.07**	(0.20)	**1.08**	(0.21)	**1.03**	(0.21)	0.11	(0.17)
cto	**0.85**	(0.21)	**0.79**	(0.20)	**0.81**	(0.18)	**0.92**	(0.17)	**0.92**	(0.19)	**1.05**	(0.18)	**0.84**	(0.19)	**0.97**	(0.20)	**1.04**	(0.20)	**1.07**	(0.21)	0.22	(0.17)
currat	**0.77**	(0.22)	**1.01**	(0.22)	**0.90**	(0.22)	**1.07**	(0.21)	**0.86**	(0.20)	**0.86**	(0.19)	**1.00**	(0.18)	**0.95**	(0.17)	**0.93**	(0.17)	**0.85**	(0.15)	0.08	(0.15)
daily.Beta	**0.67**	(0.17)	**0.76**	(0.14)	**0.85**	(0.15)	**0.88**	(0.15)	**0.97**	(0.16)	**1.00**	(0.17)	**0.99**	(0.19)	**1.01**	(0.22)	**0.99**	(0.26)	**0.90**	(0.32)	0.23	(0.28)
debt2p	**0.98**	(0.23)	**0.88**	(0.21)	**0.92**	(0.18)	**0.98**	(0.18)	**1.02**	(0.18)	**0.95**	(0.17)	**0.92**	(0.17)	**0.92**	(0.17)	**0.94**	(0.20)	**1.03**	(0.25)	0.05	(0.21)
depr	**0.80**	(0.15)	**0.84**	(0.16)	**0.88**	(0.17)	**0.95**	(0.18)	**0.92**	(0.18)	**0.93**	(0.19)	**1.02**	(0.21)	**1.01**	(0.22)	**1.00**	(0.23)	**0.91**	(0.24)	0.11	(0.18)
dolvol	**1.17**	(0.19)	**1.23**	(0.19)	**1.21**	(0.20)	**1.13**	(0.20)	**1.11**	(0.20)	**1.17**	(0.20)	**1.08**	(0.19)	**1.07**	(0.19)	**1.02**	(0.18)	**0.85**	(0.17)	**-0.32**	(0.17)
dy	**0.90**	(0.18)	**1.01**	(0.16)	**0.99**	(0.17)	**1.04**	(0.17)	**0.93**	(0.17)	**0.90**	(0.18)	**0.89**	(0.19)	**0.75**	(0.20)	**0.85**	(0.23)	**1.10**	(0.27)	0.20	(0.23)
ebitrev	0.53	(0.33)	**1.08**	(0.29)	**0.98**	(0.23)	**1.07**	(0.20)	**0.88**	(0.19)	**0.98**	(0.19)	**0.94**	(0.18)	**0.88**	(0.17)	**0.89**	(0.17)	**0.86**	(0.17)	0.33	(0.25)
egr	**0.72**	(0.24)	**0.83**	(0.21)	**0.95**	(0.19)	**0.96**	(0.17)	**0.90**	(0.17)	**0.92**	(0.16)	**0.94**	(0.17)	**1.05**	(0.17)	**1.07**	(0.19)	**1.11**	(0.22)	**0.39**	(0.15)
ep	**0.89**	(0.32)	**0.63**	(0.26)	**0.79**	(0.23)	**0.86**	(0.20)	**0.91**	(0.18)	**0.99**	(0.17)	**1.00**	(0.17)	**1.08**	(0.17)	**1.14**	(0.18)	**1.31**	(0.20)	**0.42**	(0.24)
eps	**0.97**	(0.29)	**0.84**	(0.27)	**0.89**	(0.26)	**0.92**	(0.24)	**0.79**	(0.23)	**0.88**	(0.21)	**0.83**	(0.19)	**0.92**	(0.18)	**0.89**	(0.17)	**0.96**	(0.16)	0.00	(0.21)
ga.eps	**0.84**	(0.24)	**0.82**	(0.20)	**0.76**	(0.20)	**0.80**	(0.18)	**0.99**	(0.16)	**0.95**	(0.16)	**0.94**	(0.17)	**0.97**	(0.19)	**0.91**	(0.21)	**1.00**	(0.22)	0.16	(0.14)
gma	**0.79**	(0.23)	**0.91**	(0.21)	**0.88**	(0.17)	**0.87**	(0.18)	**0.93**	(0.19)	**0.92**	(0.18)	**0.84**	(0.19)	**0.95**	(0.19)	**0.98**	(0.18)	**1.08**	(0.21)	**0.29**	(0.17)
herf	**0.82**	(0.16)	**1.01**	(0.23)	**0.94**	(0.21)	**0.85**	(0.22)	**0.97**	(0.21)	**0.91**	(0.20)	**1.02**	(0.18)	**0.94**	(0.21)	**0.90**	(0.22)	**0.92**	(0.17)	0.10	(0.14)
hire	**0.87**	(0.23)	**0.92**	(0.22)	**0.96**	(0.20)	**0.92**	(0.18)	**0.87**	(0.17)	**0.89**	(0.17)	**1.00**	(0.17)	**0.97**	(0.17)	**1.03**	(0.19)	**1.05**	(0.21)	0.18	(0.13)
ill	**0.85**	(0.17)	**1.01**	(0.19)	**1.04**	(0.20)	**1.10**	(0.21)	**1.11**	(0.21)	**1.18**	(0.22)	**1.27**	(0.22)	**1.30**	(0.22)	**1.15**	(0.23)	**1.24**	(0.26)	**0.39**	(0.23)
indmom6m	**0.75**	(0.19)	**0.85**	(0.20)	**0.89**	(0.19)	**1.01**	(0.19)	**1.09**	(0.19)	**1.14**	(0.19)	**1.19**	(0.18)	**1.04**	(0.18)	**0.96**	(0.17)	**0.84**	(0.17)	0.09	(0.13)
indmom6m.ia	**0.88**	(0.23)	**0.86**	(0.21)	**0.87**	(0.20)	**1.04**	(0.20)	**0.94**	(0.20)	**0.94**	(0.20)	**0.88**	(0.20)	**0.96**	(0.20)	**1.02**	(0.20)	**1.07**	(0.21)	0.19	(0.21)
indrev1m	**0.54**	(0.23)	**0.62**	(0.19)	**0.65**	(0.17)	**0.85**	(0.17)	**0.93**	(0.17)	**1.11**	(0.17)	**1.12**	(0.18)	**1.12**	(0.20)	**1.25**	(0.23)	**1.30**	(0.29)	**0.76**	(0.21)

Anomaly	P1	s.e.	P2	s.e.	P3	s.e.	P4	s.e.	P5	s.e.	P6	s.e.	P7	s.e.	P8	s.e.	P9	s.e.	P10	s.e.	P10-P1	s.e.
invest	**0.75**	(0.23)	**0.81**	(0.21)	**0.92**	(0.19)	**0.90**	(0.18)	**0.89**	(0.16)	**0.93**	(0.17)	**0.98**	(0.18)	**1.07**	(0.20)	**1.03**	(0.20)	**1.26**	(0.22)	**0.51**	(0.12)
ipm	**0.67**	(0.34)	**1.04**	(0.25)	**0.92**	(0.24)	**1.09**	(0.21)	**0.94**	(0.19)	**0.93**	(0.19)	**0.95**	(0.18)	**0.93**	(0.17)	**0.83**	(0.17)	**0.91**	(0.17)	0.24	(0.25)
ldp	**1.01**	(0.27)	**0.93**	(0.27)	**1.02**	(0.24)	**0.89**	(0.22)	**0.89**	(0.21)	**0.86**	(0.20)	**0.83**	(0.19)	**1.00**	(0.17)	**0.91**	(0.16)	**0.88**	(0.15)	-0.13	(0.19)
lev	**0.81**	(0.23)	**0.98**	(0.19)	**0.85**	(0.18)	**0.91**	(0.17)	**0.97**	(0.17)	**0.96**	(0.17)	**0.99**	(0.17)	**1.02**	(0.18)	**1.02**	(0.21)	**1.05**	(0.25)	0.24	(0.21)
lgr	**0.73**	(0.23)	**0.97**	(0.21)	**0.88**	(0.19)	**0.90**	(0.17)	**0.97**	(0.17)	**0.95**	(0.16)	**0.97**	(0.17)	**0.98**	(0.17)	**0.95**	(0.20)	**0.90**	(0.22)	0.18	(0.11)
lprc	**0.46**	(0.17)	**0.86**	(0.17)	**0.99**	(0.18)	**1.18**	(0.18)	**1.31**	(0.19)	**1.36**	(0.21)	**1.52**	(0.23)	**1.56**	(0.24)	**1.94**	(0.26)	**2.27**	(0.31)	**1.81**	(0.24)
maxret	0.41	(0.32)	**0.77**	(0.30)	**0.86**	(0.27)	**0.95**	(0.24)	**1.03**	(0.22)	**1.01**	(0.21)	**0.93**	(0.19)	**0.94**	(0.17)	**0.94**	(0.15)	**0.94**	(0.14)	**0.53**	(0.28)
mom11m	0.11	(0.36)	0.44	(0.27)	**0.64**	(0.23)	**0.85**	(0.20)	**0.83**	(0.18)	**0.84**	(0.18)	**0.94**	(0.17)	**1.09**	(0.18)	**1.19**	(0.20)	**1.56**	(0.26)	**1.45**	(0.32)
mom12m	0.11	(0.36)	**0.51**	(0.28)	**0.69**	(0.24)	**0.81**	(0.20)	**0.85**	(0.18)	**0.87**	(0.18)	**0.92**	(0.17)	**1.08**	(0.18)	**1.13**	(0.20)	**1.45**	(0.26)	**1.34**	(0.32)
mom18m	**0.71**	(0.25)	**0.91**	(0.21)	**0.94**	(0.18)	**0.90**	(0.17)	**0.99**	(0.17)	**1.05**	(0.18)	**1.05**	(0.18)	**0.98**	(0.20)	**1.08**	(0.23)	**1.30**	(0.28)	**0.58**	(0.22)
mom36m	**0.84**	(0.25)	**0.95**	(0.20)	**0.99**	(0.18)	**1.02**	(0.17)	**0.96**	(0.17)	**0.99**	(0.18)	**1.09**	(0.19)	**1.16**	(0.21)	**1.18**	(0.25)	**1.11**	(0.31)	0.27	(0.25)
mom6m	0.31	(0.35)	**0.81**	(0.26)	**0.87**	(0.22)	**0.93**	(0.20)	**0.90**	(0.18)	**0.97**	(0.17)	**0.91**	(0.17)	**0.95**	(0.18)	**0.95**	(0.20)	**1.38**	(0.25)	**1.08**	(0.31)
mom6m.l6	0.23	(0.29)	0.32	(0.23)	**0.65**	(0.20)	**0.80**	(0.18)	**0.81**	(0.18)	**0.87**	(0.17)	**0.96**	(0.17)	**1.01**	(0.19)	**1.30**	(0.21)	**1.56**	(0.27)	**1.33**	(0.25)
mve	**0.86**	(0.17)	**1.06**	(0.19)	**1.13**	(0.21)	**1.18**	(0.23)	**1.19**	(0.24)	**1.17**	(0.25)	**1.17**	(0.25)	**1.16**	(0.25)	**1.22**	(0.27)	**2.02**	(0.31)	**1.16**	(0.27)
mve.firm	**0.86**	(0.17)	**1.06**	(0.19)	**1.12**	(0.21)	**1.18**	(0.23)	**1.19**	(0.24)	**1.16**	(0.25)	**1.18**	(0.25)	**1.18**	(0.25)	**1.20**	(0.27)	**2.01**	(0.31)	**1.15**	(0.27)
mve.firm.ia	**0.84**	(0.17)	**1.04**	(0.20)	**1.03**	(0.21)	**1.15**	(0.21)	**1.16**	(0.22)	**1.13**	(0.21)	**1.19**	(0.20)	**1.20**	(0.20)	**1.24**	(0.22)	**1.40**	(0.24)	**0.56**	(0.14)
mve.ia	**0.84**	(0.17)	**1.04**	(0.20)	**1.15**	(0.21)	**1.11**	(0.22)	**1.20**	(0.22)	**1.11**	(0.21)	**1.11**	(0.19)	**0.99**	(0.19)	**1.05**	(0.21)	**1.20**	(0.23)	**0.35**	(0.14)
mve.jun.log	**0.85**	(0.17)	**1.07**	(0.19)	**1.09**	(0.21)	**1.14**	(0.22)	**1.15**	(0.23)	**1.23**	(0.25)	**1.24**	(0.25)	**1.29**	(0.26)	**1.19**	(0.26)	**1.36**	(0.28)	**0.51**	(0.24)
ni	**0.57**	(0.22)	**0.80**	(0.21)	**0.86**	(0.20)	**1.03**	(0.20)	**1.00**	(0.19)	**0.88**	(0.18)	**0.88**	(0.17)	**0.91**	(0.17)	**1.03**	(0.17)	**1.38**	(0.19)	**0.81**	(0.12)
operprof	**0.53**	(0.27)	**0.87**	(0.26)	**0.91**	(0.22)	**0.85**	(0.19)	**0.88**	(0.18)	**0.93**	(0.17)	**0.86**	(0.17)	**0.96**	(0.17)	**0.89**	(0.17)	**0.93**	(0.19)	**0.40**	(0.17)
pchcurrat	**0.70**	(0.21)	**0.89**	(0.18)	**0.87**	(0.18)	**0.92**	(0.18)	**1.00**	(0.17)	**0.94**	(0.18)	**0.92**	(0.17)	**0.96**	(0.18)	**0.90**	(0.19)	**0.81**	(0.20)	0.12	(0.09)
pchgm.pchsale	**0.68**	(0.22)	**0.79**	(0.20)	**0.96**	(0.19)	**0.86**	(0.18)	**0.86**	(0.17)	**0.97**	(0.17)	**0.96**	(0.17)	**1.02**	(0.18)	**0.93**	(0.20)	**0.91**	(0.22)	**0.23**	(0.12)
pchquick	**0.85**	(0.21)	**0.94**	(0.20)	**0.82**	(0.18)	**0.95**	(0.17)	**0.97**	(0.17)	**0.99**	(0.18)	**0.92**	(0.18)	**0.88**	(0.18)	**0.94**	(0.19)	**0.90**	(0.21)	0.05	(0.10)
pchsale.pchrect	**0.73**	(0.22)	**1.10**	(0.20)	**1.08**	(0.18)	**0.95**	(0.17)	**1.00**	(0.17)	**0.88**	(0.17)	**0.82**	(0.17)	**0.91**	(0.19)	**0.89**	(0.20)	**0.86**	(0.22)	0.13	(0.11)
pcm	**0.96**	(0.18)	**0.86**	(0.19)	**0.86**	(0.17)	**0.92**	(0.18)	**0.92**	(0.18)	**0.91**	(0.19)	**0.91**	(0.19)	**0.93**	(0.18)	**0.92**	(0.19)	**0.97**	(0.22)	0.01	(0.14)
pm	0.51	(0.32)	**1.10**	(0.29)	**0.99**	(0.23)	**1.02**	(0.20)	**0.94**	(0.19)	**0.96**	(0.19)	**0.89**	(0.18)	**0.90**	(0.17)	**0.92**	(0.17)	**0.85**	(0.17)	0.34	(0.25)
pm.ia	**1.00**	(0.19)	**0.88**	(0.18)	**0.93**	(0.18)	**0.97**	(0.20)	**0.93**	(0.20)	**0.93**	(0.20)	**1.02**	(0.20)	**0.80**	(0.19)	**0.88**	(0.20)	**1.01**	(0.19)	0.01	(0.13)
prc52	**0.86**	(0.16)	**0.86**	(0.16)	**0.86**	(0.17)	**0.94**	(0.18)	**1.01**	(0.20)	**0.96**	(0.22)	**0.97**	(0.24)	**0.78**	(0.29)	**0.80**	(0.33)	0.69	(0.42)	-0.17	(0.36)
prof	**0.72**	(0.21)	**0.80**	(0.18)	**0.73**	(0.17)	**0.86**	(0.18)	**0.85**	(0.19)	**0.92**	(0.19)	**0.95**	(0.19)	**1.01**	(0.19)	**1.08**	(0.19)	**1.14**	(0.19)	**0.43**	(0.14)
q	**0.88**	(0.21)	**0.86**	(0.18)	**0.83**	(0.18)	**0.94**	(0.17)	**0.98**	(0.18)	**0.89**	(0.18)	**1.09**	(0.18)	**1.12**	(0.20)	**1.11**	(0.21)	**1.21**	(0.21)	**0.32**	(0.18)
quick	**0.96**	(0.16)	**1.00**	(0.16)	**0.89**	(0.17)	**0.91**	(0.18)	**0.91**	(0.18)	**0.89**	(0.19)	**0.93**	(0.20)	**1.03**	(0.21)	**0.95**	(0.22)	**0.81**	(0.22)	-0.15	(0.13)
r62	0.24	(0.32)	**0.73**	(0.25)	**0.85**	(0.22)	**0.91**	(0.19)	**0.97**	(0.18)	**1.01**	(0.17)	**0.92**	(0.17)	**0.89**	(0.18)	**1.00**	(0.20)	**1.37**	(0.25)	**1.13**	(0.28)
retvol	-0.01	(0.38)	0.47	(0.33)	**0.76**	(0.30)	**1.02**	(0.27)	**1.13**	(0.24)	**1.02**	(0.21)	**0.97**	(0.19)	**0.95**	(0.17)	**0.97**	(0.15)	**0.87**	(0.13)	**0.89**	(0.34)
rev1m	**0.59**	(0.25)	**0.79**	(0.19)	**0.84**	(0.18)	**0.93**	(0.17)	**0.90**	(0.17)	**1.00**	(0.18)	**1.10**	(0.19)	**1.19**	(0.21)	**1.06**	(0.24)	**1.07**	(0.31)	**0.48**	(0.26)
roa2	**0.61**	(0.33)	**0.79**	(0.25)	**0.95**	(0.23)	**0.92**	(0.21)	**0.95**	(0.18)	**0.98**	(0.17)	**0.88**	(0.17)	**0.97**	(0.17)	**0.89**	(0.17)	**0.93**	(0.19)	0.32	(0.24)
roe	**0.68**	(0.32)	**0.88**	(0.25)	**0.99**	(0.21)	**0.90**	(0.19)	**0.94**	(0.18)	**0.88**	(0.17)	**0.90**	(0.17)	**0.91**	(0.18)	**0.91**	(0.18)	**0.93**	(0.20)	0.25	(0.23)
roic	**0.66**	(0.33)	**0.85**	(0.27)	**0.86**	(0.23)	**0.85**	(0.22)	**0.84**	(0.18)	**0.85**	(0.18)	**0.95**	(0.18)	**0.94**	(0.18)	**0.96**	(0.17)	**0.99**	(0.19)	0.34	(0.24)
salecash	**0.89**	(0.24)	**0.96**	(0.21)	**0.95**	(0.20)	**0.95**	(0.19)	**0.89**	(0.18)	**1.02**	(0.18)	**0.88**	(0.18)	**0.91**	(0.18)	**0.93**	(0.17)	**0.93**	(0.18)	0.05	(0.16)
salerec	**0.84**	(0.21)	**0.79**	(0.22)	**0.86**	(0.21)	**0.86**	(0.19)	**0.95**	(0.19)	**1.01**	(0.18)	**0.97**	(0.16)	**0.93**	(0.17)	**0.89**	(0.17)	**1.10**	(0.19)	**0.26**	(0.14)
sat	**0.79**	(0.23)	**0.78**	(0.20)	**0.84**	(0.18)	**0.86**	(0.18)	**0.92**	(0.19)	**0.97**	(0.18)	**1.01**	(0.18)	**0.95**	(0.19)	**1.12**	(0.19)	**1.15**	(0.20)	**0.37**	(0.17)
sat.ia	**0.85**	(0.20)	**0.93**	(0.21)	**0.88**	(0.20)	**0.77**	(0.18)	**0.97**	(0.17)	**0.92**	(0.19)	**0.90**	(0.19)	**0.90**	(0.17)	**1.02**	(0.18)	**1.13**	(0.19)	**0.27**	(0.12)
sgr	**0.81**	(0.23)	**0.93**	(0.22)	**0.95**	(0.20)	**0.91**	(0.18)	**0.93**	(0.17)	**0.98**	(0.16)	**0.91**	(0.17)	**0.95**	(0.17)	**1.05**	(0.20)	**0.86**	(0.22)	0.05	(0.14)
sp	**0.70**	(0.22)	**0.85**	(0.17)	**0.86**	(0.17)	**1.01**	(0.17)	**1.03**	(0.18)	**1.03**	(0.19)	**1.13**	(0.19)	**1.13**	(0.20)	**1.24**	(0.21)	**1.29**	(0.24)	**0.59**	(0.20)
std.dolvol	**1.15**	(0.18)	**1.15**	(0.19)	**1.10**	(0.19)	**1.02**	(0.19)	**0.98**	(0.19)	**1.02**	(0.19)	**0.98**	(0.19)	**0.96**	(0.19)	**0.89**	(0.18)	**0.85**	(0.17)	**-0.31**	(0.13)
std.dolvol36m	**0.78**	(0.17)	**0.99**	(0.17)	**0.97**	(0.17)	**1.07**	(0.18)	**1.05**	(0.18)	**1.06**	(0.18)	**1.09**	(0.18)	**1.13**	(0.18)	**1.13**	(0.18)	**1.24**	(0.17)	**0.46**	(0.13)
std.turn	**0.94**	(0.28)	**1.06**	(0.25)	**1.08**	(0.23)	**1.05**	(0.21)	**1.06**	(0.20)	**1.07**	(0.19)	**0.84**	(0.17)	**0.90**	(0.17)	**0.86**	(0.16)	**0.79**	(0.15)	-0.14	(0.23)
tan	**1.20**	(0.26)	**0.94**	(0.23)	**1.04**	(0.20)	**0.92**	(0.20)	**0.99**	(0.19)	**0.90**	(0.17)	**0.85**	(0.17)	**0.90**	(0.16)	**0.94**	(0.17)	**0.92**	(0.19)	-0.27	(0.18)
tang	**1.20**	(0.26)	**0.94**	(0.23)	**1.04**	(0.20)	**0.92**	(0.20)	**0.99**	(0.19)	**0.90**	(0.17)	**0.85**	(0.17)	**0.90**	(0.16)	**0.94**	(0.17)	**0.92**	(0.19)	-0.27	(0.18)
turn	**0.88**	(0.31)	**0.97**	(0.25)	**0.96**	(0.21)	**0.91**	(0.19)	**0.97**	(0.18)	**0.91**	(0.16)	**0.89**	(0.16)	**0.87**	(0.16)	**0.94**	(0.15)	**0.90**	(0.15)	0.01	(0.26)
zerotrade	**0.97**	(0.15)	**0.94**	(0.17)	**0.87**	(0.16)	**0.87**	(0.16)	**0.84**	(0.15)	**0.96**	(0.17)	**0.96**	(0.19)	**0.98**	(0.21)	**1.05**	(0.24)	**0.94**	(0.30)	-0.03	(0.24)

Table 4: **Jensen's Alpha and Information Ratio**

This table reports average monthly portfolio returns, Jensen's alphas and Information ratios in percentage. The statistics are reported on unadjusted returns and in excess of the CAPM or Fama and French three factor model. The null hypothesis suggests that the intercepts are less equal zero and is thus a one-sided t-test. Annualized Sharpe and Information ratios are approximated by $\sqrt{12} \cdot \hat{\alpha}/\hat{\sigma_\epsilon}$. In the case of Sharpe ratios, $\hat{\alpha}$ is the average return of a long-short anomaly portfolio. Standard errors are reported in parentheses. Significant average returns at a single critical value of 5% are printed in bold type. The standard errors of the Information ratios are estimated by moving block bootstrapping the data 10,000 times per month to retain the autocorrelation structure in the returns.

	Return				**CAPM**				**FF3**			
Anomaly	Avg	s.e.	SR	s.e.	$\hat{\alpha}$	s.e.	IR	s.e.	$\hat{\alpha}$	s.e.	IR	s.e.
a2me	0.26	(0.21)	0.17	(0.14)	0.19	(0.21)	0.11	(0.14)	-0.44	(0.14)	-0.39	(0.14)
age.m	-0.03	(0.21)	-0.02	(0.16)	0.27	(0.20)	0.21	(0.16)	0.1	(0.16)	0.09	(0.16)
agr	**0.38**	(0.15)	**0.35**	(0.14)	**0.44**	(0.15)	**0.40**	(0.14)	0.15	(0.13)	0.17	(0.14)
at	0.27	(0.26)	0.14	(0.14)	0.47	(0.25)	0.25	(0.14)	**0.5**	(0.14)	**0.45**	(0.14)
bab	-0.25	(0.30)	-0.11	(0.14)	-0.85	(0.22)	-0.55	(0.14)	-0.75	(0.21)	-0.5	(0.14)
baspread	0.34	(0.4)	0.12	(0.14)	**0.82**	(0.37)	**0.31**	(0.14)	**1.03**	(0.30)	**0.47**	(0.14)
bm	0.34	(0.22)	0.21	(0.14)	0.29	(0.22)	0.18	(0.14)	-0.41	(0.14)	-0.42	(0.14)
bm.ia	**0.33**	(0.15)	**0.31**	(0.14)	**0.32**	(0.15)	**0.28**	(0.14)	0.05	(0.13)	0.05	(0.14)
c2d	0.29	(0.23)	0.17	(0.14)	**0.49**	(0.22)	**0.29**	(0.14)	**0.64**	(0.19)	**0.48**	(0.14)
cashdebt	0.37	(0.23)	0.22	(0.14)	**0.57**	(0.22)	**0.33**	(0.14)	**0.69**	(0.19)	**0.53**	(0.14)
cashpr	**0.35**	(0.13)	**0.36**	(0.14)	**0.43**	(0.13)	**0.43**	(0.14)	0.05	(0.09)	0.06	(0.14)
chato	0.17	(0.15)	0.16	(0.14)	0.22	(0.15)	0.2	(0.14)	0.2	(0.15)	0.19	(0.14)
chatoia	0.16	(0.13)	0.17	(0.14)	0.22	(0.13)	0.22	(0.14)	0.21	(0.13)	0.23	(0.14)
chcsho	**0.64**	(0.11)	**0.76**	(0.14)	**0.74**	(0.11)	**0.95**	(0.14)	**0.53**	(0.10)	**0.76**	(0.14)
chempia	0.04	(0.13)	0.04	(0.14)	-0.06	(0.13)	-0.06	(0.14)	0.07	(0.12)	0.07	(0.14)
chinv	**0.43**	(0.14)	**0.41**	(0.14)	**0.48**	(0.14)	**0.48**	(0.14)	**0.27**	(0.13)	**0.28**	(0.14)
chlogso	**0.81**	(0.12)	**0.93**	(0.14)	**0.91**	(0.11)	**1.13**	(0.14)	**0.86**	(0.11)	**1.04**	(0.14)
chmom	0.22	(0.21)	0.14	(0.14)	0.09	(0.21)	0.07	(0.14)	0.12	(0.21)	0.09	(0.14)
chnwc	**0.5**	(0.13)	**0.51**	(0.14)	**0.56**	(0.13)	**0.6**	(0.14)	**0.58**	(0.13)	**0.61**	(0.14)
chpm	-0.11	(0.14)	-0.1	(0.14)	-0.08	(0.14)	-0.08	(0.14)	-0.18	(0.14)	-0.2	(0.14)
chpmia	0.11	(0.17)	0.09	(0.14)	0.12	(0.17)	0.09	(0.14)	0.07	(0.17)	0.05	(0.14)
cto	0.22	(0.17)	0.18	(0.14)	0.18	(0.17)	0.15	(0.14)	**0.47**	(0.14)	**0.45**	(0.14)
currat	0.08	(0.15)	0.07	(0.14)	**0.31**	(0.12)	**0.34**	(0.14)	**0.28**	(0.11)	**0.38**	(0.14)
daily.Beta	0.23	(0.28)	0.11	(0.14)	-0.26	(0.22)	-0.16	(0.14)	-0.07	(0.21)	-0.04	(0.14)
debt2p	0.05	(0.21)	0.03	(0.14)	0.03	(0.21)	0.02	(0.14)	-0.54	(0.15)	-0.48	(0.14)
depr	0.11	(0.18)	0.08	(0.14)	-0.12	(0.16)	-0.11	(0.14)	0.02	(0.14)	0.02	(0.14)
dolvol	-0.32	(0.17)	-0.25	(0.14)	-0.45	(0.17)	-0.39	(0.14)	-0.13	(0.12)	-0.14	(0.14)
dy	0.2	(0.23)	0.12	(0.14)	-0.14	(0.20)	-0.09	(0.14)	**0.25**	(0.13)	**0.28**	(0.14)
ebitrev	0.33	(0.25)	0.18	(0.14)	**0.6**	(0.23)	**0.39**	(0.14)	**0.65**	(0.18)	**0.54**	(0.14)
egr	**0.39**	(0.15)	**0.37**	(0.14)	**0.49**	(0.14)	**0.47**	(0.14)	0.25	(0.13)	0.26	(0.14)
ep	0.42	(0.24)	0.24	(0.14)	**0.64**	(0.23)	**0.41**	(0.14)	**0.53**	(0.21)	**0.38**	(0.14)
eps	0	(0.21)	0	(0.14)	0.27	(0.19)	0.18	(0.14)	**0.43**	(0.15)	**0.38**	(0.14)
ga.eps	0.16	(0.14)	0.16	(0.14)	0.19	(0.14)	0.18	(0.14)	0.24	(0.14)	0.24	(0.14)
gma	0.29	(0.17)	0.23	(0.14)	0.33	(0.17)	**0.27**	(0.14)	**0.75**	(0.13)	**0.76**	(0.14)
herf	0.1	(0.14)	0.1	(0.14)	-0.01	(0.14)	-0.01	(0.14)	0.08	(0.13)	0.08	(0.14)
hire	0.18	(0.13)	0.19	(0.14)	**0.26**	(0.13)	**0.27**	(0.14)	0.02	(0.11)	0.02	(0.14)
ill	0.39	(0.23)	0.23	(0.14)	0.42	(0.23)	**0.27**	(0.14)	0.06	(0.16)	0.05	(0.14)
indmom6m	0.09	(0.13)	0.09	(0.14)	0.14	(0.13)	0.14	(0.14)	0.21	(0.13)	0.24	(0.14)

	Return				**CAPM**				**FF3**			
Anomaly	Avg	s.e.	SR	s.e.	$\hat{\alpha}$	s.e.	IR	s.e.	$\hat{\alpha}$	s.e.	IR	s.e.
indmom6m.ia	0.19	(0.21)	0.12	(0.14)	0.21	(0.21)	0.14	(0.14)	0.25	(0.21)	0.17	(0.14)
indrev1m	**0.76**	(0.21)	**0.48**	(0.14)	**0.58**	(0.21)	**0.39**	(0.14)	**0.51**	(0.21)	**0.35**	(0.14)
invest	**0.51**	(0.12)	**0.57**	(0.14)	**0.56**	(0.12)	**0.64**	(0.14)	**0.35**	(0.11)	**0.46**	(0.14)
ipm	0.24	(0.25)	0.13	(0.14)	**0.55**	(0.23)	**0.35**	(0.14)	**0.68**	(0.18)	**0.47**	(0.14)
ldp	-0.13	(0.19)	-0.09	(0.14)	0.2	(0.16)	0.18	(0.14)	0.13	(0.10)	0.18	(0.14)
lev	0.24	(0.21)	0.16	(0.14)	0.23	(0.21)	0.15	(0.14)	-0.36	(0.14)	-0.36	(0.14)
lgr	0.18	(0.11)	0.21	(0.14)	0.21	(0.11)	0.25	(0.14)	0.05	(0.11)	0.07	(0.14)
lprc	**1.81**	(0.24)	**1.04**	(0.14)	**1.62**	(0.23)	**1.09**	(0.14)	**1.46**	(0.22)	**0.96**	(0.14)
maxret	0.53	(0.28)	0.26	(0.14)	**0.93**	(0.24)	**0.54**	(0.14)	**1.02**	(0.19)	**0.74**	(0.14)
mom11m	**1.45**	(0.32)	**0.61**	(0.14)	**1.62**	(0.32)	**0.70**	(0.14)	**1.89**	(0.32)	**0.80**	(0.14)
mom12m	**1.34**	(0.32)	**0.58**	(0.14)	**1.52**	(0.31)	**0.70**	(0.14)	**1.8**	(0.31)	**0.79**	(0.14)
mom18m	**0.58**	(0.22)	**0.37**	(0.14)	**0.56**	(0.22)	**0.32**	(0.14)	0.2	(0.20)	0.15	(0.14)
mom36m	0.27	(0.25)	0.15	(0.14)	0.23	(0.25)	0.13	(0.14)	-0.28	(0.21)	-0.17	(0.14)
mom6m	**1.08**	(0.31)	**0.47**	(0.14)	**1.28**	(0.31)	**0.57**	(0.14)	**1.44**	(0.31)	**0.71**	(0.14)
mom6m.l6	**1.33**	(0.25)	**0.74**	(0.14)	**1.35**	(0.25)	**0.77**	(0.14)	**1.57**	(0.24)	**0.91**	(0.14)
mve	**1.16**	(0.27)	**0.59**	(0.14)	**1.12**	(0.27)	**0.55**	(0.14)	**0.75**	(0.20)	**0.61**	(0.14)
mve.firm	**1.15**	(0.27)	**0.58**	(0.14)	**1.11**	(0.27)	**0.56**	(0.14)	**0.74**	(0.20)	**0.58**	(0.14)
mve.firm.ia	**0.56**	(0.14)	**0.56**	(0.14)	**0.43**	(0.13)	**0.47**	(0.14)	**0.37**	(0.09)	**0.56**	(0.14)
mve.ia	**0.35**	(0.14)	**0.35**	(0.14)	0.25	(0.13)	0.26	(0.14)	**0.22**	(0.09)	**0.37**	(0.14)
mve.jun.log	**0.51**	(0.24)	**0.28**	(0.14)	**0.49**	(0.25)	**0.29**	(0.14)	0.16	(0.16)	0.14	(0.14)
ni	**0.81**	(0.12)	**0.89**	(0.14)	**0.91**	(0.12)	**0.96**	(0.14)	**0.85**	(0.12)	**1.05**	(0.14)
operprof	**0.40**	(0.17)	**0.32**	(0.14)	**0.55**	(0.17)	**0.46**	(0.14)	**0.78**	(0.14)	**0.81**	(0.14)
pchcurrat	0.12	(0.09)	0.17	(0.14)	0.13	(0.09)	0.19	(0.14)	0.13	(0.10)	0.2	(0.14)
pchgm.pchsale	**0.23**	(0.12)	**0.27**	(0.14)	0.2	(0.12)	0.23	(0.14)	**0.33**	(0.11)	**0.42**	(0.14)
pchquick	0.05	(0.10)	0.07	(0.14)	0.04	(0.10)	0.05	(0.14)	0.04	(0.10)	0.05	(0.14)
pchsale.pchrect	0.13	(0.11)	0.15	(0.14)	0.12	(0.11)	0.13	(0.14)	0.06	(0.11)	0.08	(0.14)
pcm	0.01	(0.14)	0.01	(0.14)	-0.08	(0.13)	-0.07	(0.14)	-0.3	(0.11)	-0.36	(0.14)
pm	0.34	(0.25)	0.19	(0.14)	**0.62**	(0.23)	**0.38**	(0.14)	**0.64**	(0.17)	**0.49**	(0.14)
pm.ia	0.01	(0.13)	0.01	(0.14)	0.02	(0.14)	0.02	(0.14)	**0.29**	(0.12)	**0.33**	(0.14)
prc52	-0.17	(0.36)	-0.06	(0.14)	-0.68	(0.32)	-0.31	(0.14)	-0.9	(0.30)	-0.44	(0.14)
prof	**0.43**	(0.14)	**0.41**	(0.14)	**0.48**	(0.14)	**0.46**	(0.14)	**0.57**	(0.14)	**0.58**	(0.14)
q	0.32	(0.18)	0.24	(0.14)	**0.37**	(0.18)	**0.27**	(0.14)	-0.19	(0.12)	-0.22	(0.14)
quick	-0.15	(0.13)	-0.16	(0.14)	-0.32	(0.12)	-0.37	(0.14)	-0.27	(0.11)	-0.32	(0.14)
r62	**1.13**	(0.28)	**0.56**	(0.14)	**1.28**	(0.28)	**0.6**	(0.14)	**1.42**	(0.28)	**0.77**	(0.14)
retvol	**0.89**	(0.34)	**0.36**	(0.14)	**1.38**	(0.30)	**0.67**	(0.14)	**1.5**	(0.25)	**0.75**	(0.14)
rev1m	0.48	(0.26)	0.26	(0.14)	0.31	(0.25)	0.15	(0.14)	0.23	(0.25)	0.13	(0.14)
roa2	0.32	(0.24)	0.18	(0.14)	**0.55**	(0.23)	**0.37**	(0.14)	**0.71**	(0.20)	**0.53**	(0.14)
roe	0.25	(0.23)	0.15	(0.14)	**0.46**	(0.22)	**0.28**	(0.14)	**0.68**	(0.18)	**0.51**	(0.14)
roic	0.34	(0.24)	0.19	(0.14)	**0.56**	(0.23)	**0.40**	(0.14)	**0.71**	(0.19)	**0.51**	(0.14)
salecash	0.05	(0.16)	0.04	(0.14)	0.18	(0.16)	0.16	(0.14)	0.18	(0.16)	0.17	(0.14)
salerec	0.26	(0.14)	0.25	(0.14)	0.27	(0.14)	0.26	(0.14)	**0.46**	(0.13)	**0.48**	(0.14)
sat	**0.37**	(0.17)	**0.29**	(0.14)	**0.38**	(0.17)	**0.33**	(0.14)	**0.62**	(0.16)	**0.58**	(0.14)
sat.ia	**0.27**	(0.12)	**0.31**	(0.14)	**0.33**	(0.12)	**0.40**	(0.14)	**0.36**	(0.12)	**0.39**	(0.14)
sgr	0.05	(0.14)	0.05	(0.14)	0.11	(0.14)	0.11	(0.14)	-0.15	(0.13)	-0.17	(0.14)
sp	**0.59**	(0.20)	**0.41**	(0.14)	**0.58**	(0.20)	**0.42**	(0.14)	-0.01	(0.13)	-0.01	(0.14)
std.dolvol	-0.31	(0.13)	-0.31	(0.14)	-0.36	(0.13)	-0.4	(0.14)	-0.17	(0.09)	-0.27	(0.14)
std.dolvol36m	**0.46**	(0.13)	**0.49**	(0.14)	**0.55**	(0.13)	**0.62**	(0.14)	**0.25**	(0.08)	**0.46**	(0.14)
std.turn	-0.14	(0.23)	-0.08	(0.14)	0.22	(0.20)	0.15	(0.14)	0.17	(0.15)	0.15	(0.14)
tan	-0.27	(0.18)	-0.21	(0.14)	-0.16	(0.18)	-0.13	(0.14)	-0.43	(0.16)	-0.38	(0.14)
tang	-0.27	(0.18)	-0.21	(0.14)	-0.16	(0.18)	-0.13	(0.14)	-0.43	(0.16)	-0.38	(0.14)
turn	0.01	(0.26)	0.01	(0.14)	**0.45**	(0.22)	**0.27**	(0.14)	0.26	(0.18)	0.21	(0.14)
zerotrade	-0.03	(0.24)	-0.02	(0.14)	-0.48	(0.19)	-0.36	(0.14)	-0.23	(0.17)	-0.19	(0.14)

Table 5: **Linear Model In-Sample Results**

This table reports selected characteristics obtained from penalized regressions from a total universe of 90 market anomalies. The columns report model selection results for different samples, regularization models and number of steps. The in-sample Sharpe ratios summarize the performance of equally-weighted hedge portfolios going long the decile of stocks with highest predicted returns and shorting the decile of stocks with lowest predicted returns. The sample period spans from January 1965 to December 2018 unless otherwise specified.

Firms	All	All	All	NYSE 10	NYSE 10	NYSE 10	NYSE 20	NYSE 20	NYSE 20
Period	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018
Sample Size	1,417,151	1,417,151	1,417,151	861,982	861,982	861,982	690,545	690,545	690,545
Regularization	SCAD	Elastic Net	Adaptive LASSO	SCAD	MCP	Adaptive LASSO	SCAD	MCP	Adaptive LASSO
# Steps	1	1	2	1	1	2	1	1	2
# Selected	65	86	59	42	75	45	37	62	38
IS Sharpe ratio	2.95	2.91	2.88	3.47	3.47	3.48	3.27	3.22	3.29
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
a2me	a2me	a2me	a2me	a2me	a2me	a2me	a2me	a2me	a2me
age_m	age_m	age_m	age_m	age_m	age_m	age_m	age_m	age_m	age_m
agr	agr	agr	agr	agr	agr				
at		at							
bab	bab	bab	bab						bab
baspread	baspread	baspread	baspread	baspread	baspread	baspread			baspread
bm	bm	bm	bm	bm	bm	bm	bm	bm	bm
bm_ia		bm_ia							
c2d		c2d	c2d			c2d	c2d		
cashdebt		cashdebt		cashdebt	cashdebt				cashdebt
cashpr	cashpr	cashpr							
chato	chato	chato		chato	chato	chato	chato	chato	chato
chatoia		chatoia	chatoia						
chcsho		chcsho							
chempia		chempia	chempia	chempia	chempia	chempia	chempia	chempia	chempia
chinv	chinv	chinv	chinv	chinv	chinv	chinv			chinv
chlogso		chlogso							
chmom	chmom	chmom	chmom	chmom	chmom	chmom			chmom
chnwc	chnwc	chnwc	chnwc	chnwc	chnwc	chnwc	chnwc	chnwc	chnwc
chpm	chpm	chpm							
chpmia	chpmia	chpmia							
cto	cto	cto	cto			cto	cto		cto
currat	currat	currat	currat						
daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta	daily_Beta
debt2p	debt2p	debt2p	debt2p			debt2p			debt2p
depr	depr	depr	depr	depr	depr	depr	depr	depr	depr
dolvol	dolvol	dolvol	dolvol			dolvol			dolvol
dy	dy	dy	dy	dy	dy	dy	dy	dy	dy
ebitrev	ebitrev	ebitrev	ebitrev						
egr	egr	egr	egr	egr	egr	egr	egr	egr	egr
ep	ep	ep	ep						
eps	eps	eps	eps	eps	eps				
ga_eps		ga_eps							
gma	gma	gma	gma						
herf	herf	herf	herf	herf	herf	herf	herf	herf	herf
hire		hire							

ill	ill	ill	ill	ill	ill	ill			
indmom6m	indmom6m	indmom6m	indmom6m	indmom6m	indmom6m	indmom6m	indmom6m	indmom6m	indmom6m
indmom6m_ia	indmom6m_ia	indmom6m_ia	indmom6m_ia						
indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m
invest		invest							
ipm		ipm							
ldp	ldp	ldp	ldp	ldp	ldp	ldp	ldp	ldp	ldp
lev	lev	lev	lev						
lgr		lgr							
lprc	lprc	lprc	lprc	lprc	lprc	lprc	lprc	lprc	lprc
maxret	maxret	maxret	maxret						
mom11m	mom11m	mom11m	mom11m			mom11m			mom11m
mom12m	mom12m	mom12m	mom12m			mom12m			mom12m
mom18m	mom18m	mom18m	mom18m	mom18m	mom18m	mom18m	mom18m	mom18m	mom18m
mom36m	mom36m	mom36m		mom36m			mom36m	mom36m	mom36m
mom6m	mom6m	mom6m	mom6m				mom6m	mom6m	
mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6	mom6m_l6
mve		mve	mve						
mve_firm		mve_firm	mve_firm						
mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia	mve_firm_ia
mve_ia	mve_ia	mve_ia	mve_ia	mve_ia	mve_ia	mve_ia	mve_ia	mve_ia	mve_ia
mve_jun_log	mve_jun_log	mve_jun_log	mve_jun_log	mve_jun_log	mve_jun_log	mve_jun_log			mve_jun_log
ni	ni	ni	ni	ni	ni	ni	ni	ni	ni
operprof	operprof	operprof	operprof						
pchcurrat	pchcurrat	pchcurrat	pchcurrat						
pchgm_pchsale		pchgm_pchsale							
pchquick		pchquick							
pchsale_pchrect	pchsale_pchrect	pchsale_pchrect	pchsale_pchrect						
pcm		pcm							
pm	pm	pm	pm						
pm_ia	pm_ia	pm_ia	pm_ia	pm_ia	pm_ia	pm_ia	pm_ia	pm_ia	pm_ia
prc52	prc52	prc52	prc52	prc52	prc52	prc52	prc52	prc52	prc52
prof		prof		prof	prof	prof	prof	prof	prof
q		q							
quick		quick	quick						
r62	r62	r62	r62						
retvol	retvol	retvol	retvol	retvol	retvol	retvol	retvol	retvol	retvol
rev1m	rev1m	rev1m	rev1m	rev1m	rev1m	rev1m	rev1m	rev1m	rev1m
roa2		roa2	roa2						
roe		roe	roe						roe
roic	roic	roic		roic	roic	roic	roic	roic	roic
salecash	salecash	salecash	salecash	salecash	salecash	salecash	salecash	salecash	salecash
salerec	salerec	salerec		salerec	salerec	salerec	salerec	salerec	salerec
sat		sat	sat						
sat_ia	sat_ia	sat_ia	sat_ia	sat_ia	sat_ia	sat_ia	sat_ia	sat_ia	sat_ia
sgr		sgr							
sp	sp	sp	sp	sp		sp	sp		sp
std_dolvol	std_dolvol	std_dolvol	std_dolvol	std_dolvol	std_dolvol	std_dolvol			std_dolvol
std_dolvol36m		std_dolvol36m							
std_turn	std_turn	std_turn	std_turn	std_turn	std_turn	std_turn	std_turn		std_turn
tan		tan							
tang		tang							
turn	turn	turn	turn	turn	turn	turn	turn	turn	turn
zerotrade	zerotrade	zerotrade	zerotrade	zerotrade	zerotrade	zerotrade	zerotrade	zerotrade	zerotrade

Table 6: **Nonlinear Model In-Sample Results**

This table reports selected characteristics obtained from penalized regressions from a total universe of 90 market anomalies. The columns report model selection results for different samples, regularization models and number of steps. The in-sample Sharpe ratios summarize the performance of equally-weighted hedge portfolios going long the decile of stocks with highest predicted returns and shorting the decile of stocks with lowest predicted returns. The sample period spans from January 1965 to December 2018 unless otherwise specified.

Firms	All	All	NYSE 20	NYSE 20	NYSE 20	NYSE 20
Period	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018	1965-2018
Sample Size	1,417,151	1,417,151	690,545	690,545	690,545	690,545
Regularization	SCAD	Adaptive LASSO	LASSO	Elastic Net	Adaptive LASSO	SCAD
# Steps	1	2	1	1	2	1
Knots	19	19	19	19	19	19
# Selected	3	6	9	27	8	2
IS Sharpe ratio	2.35	2.57	2.77	2.94	2.82	1.89
Models	(1)	(2)	(3)	(4)	(5)	(6)
a2me		a2me		a2me		
bm				bm		
bm_ia				bm_ia		
chaTO				chaTO		
chcsho				chcsho		
chlogso				chlogso		
chnwc				chnwc		
daily_Beta				daily_Beta		
depr				depr		
eps			eps	eps	eps	
herf				herf		
indmom6m				indmom6m		
indmom6m_ia		indmom6m_ia				
indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m	indrev1m
ldp			ldp	ldp	ldp	
lprc	lprc	lprc	lprc	lprc	lprc	lprc
mom11m			mom11m	mom11m	mom11m	
mom18m		mom18m	mom18m	mom18m	mom18m	
mom6m_l6			mom6m_l6	mom6m_l6	mom6m_l6	
ni				ni		
pm_ia				pm_ia		
prc52	prc52	prc52	prc52	prc52	prc52	
prof				prof		
retvol				retvol		
sp				sp		
std_turn				std_turn		
ep			ep	ep		
roic				roic		

Table 7: **Linear Model Out-of-Sample Prediction**
This table reports out-of-sample, value-weighted hedge portfolio returns, going long the stocks in the highest predicted return decile and shorting the stocks in the lowest predicted return decile for different estimation periods and regularization terms. The table reports the first four moments of monthly percentage returns, turnover, and predictive slopes and R^2 for the hedge portfolios in Panel A, as well as for the long legs in Panel B and the short legs in Panel C. The model selection and estimation period spans from January 1965 to December of the year before start of the 10 year out-of-sample period.

Firms	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20
Period	1965-1982	1965-1982	1965-1982	1965-1990	1965-1990	1983-2000	1983-2000
Sample Size	138,733	138,733	138,733	227,802	227,802	258,044	258,044
Regularization	Elastic Net	Adaptive LASSO	SCAD	Adaptive LASSO	SCAD	Adaptive LASSO	SCAD
# Steps	1	2	1	2	1	2	1
Knots	9	9	9	9	9	9	9
# Selected	36	14	19	14	22	18	30
IS Sharpe ratio	3.86	3.84	3.77	3.97	4.08	4.06	4.34
Panel A: Long-Short Portfolio							
Mean	2.24	2.72	2.9	2.01	2.09	0.85	1.11
Std. Dev.	4.04	4.94	3.99	5.51	5.49	6.26	6.12
Sharpe ratio	1.92	1.91	2.52	1.27	1.32	0.47	0.63
Skewness	0.39	1.1	0.37	0.17	-0.09	-0.22	-0.24
Kurtosis	-0.01	3.08	0.98	2.1	2.01	3.84	4.46
β	0.5	0.51	0.53	0.86	0.81	0.62	0.61
R^2	0.03	0.03	0.03	0.04	0.04	0.04	0.04
TO	0.36	0.42	0.4	0.39	0.41	0.33	0.34
TO_R	0.37	0.42	0.4	0.4	0.41	0.33	0.34
Panel B: Long Leg							
Mean	2.3	2.65	2.39	2.18	2.27	0.87	1.08
Std. Dev.	6.16	7.47	5.92	5.39	5.78	7.96	8.07
Sharpe ratio	1.29	1.23	1.4	1.4	1.36	0.38	0.46
Skewness	-0.4	-0.84	-0.67	0.09	-0.06	0.1	0.13
Kurtosis	1.87	5.07	4.01	1.14	0.68	1.49	1.43
β	0.44	0.48	0.45	0.99	0.92	0.71	0.66
R^2	0.07	0.06	0.06	0.08	0.08	0.1	0.09
TO	0.2	0.2	0.2	0.2	0.21	0.21	0.21
TO_R	0.21	0.21	0.21	0.2	0.21	0.22	0.22
Panel C: Short Leg							
Mean	0.06	-0.07	-0.51	0.16	0.18	0.02	-0.03
Std. Dev.	4.45	5.46	5.44	5.2	5.08	4.42	4.42
Sharpe ratio	0.05	-0.04	-0.32	0.11	0.12	0.01	-0.02
Skewness	-0.74	-0.89	-1.1	-0.22	-0.33	-0.22	-0.28
Kurtosis	2.8	5.88	4.83	0.02	0.55	1.05	1.03
β	0.33	0.38	0.41	0.08	0.09	0.24	0.24
R^2	0.06	0.05	0.05	0.06	0.06	0.08	0.08
TO	0.16	0.21	0.19	0.19	0.19	0.11	0.12
TO_R	0.16	0.21	0.19	0.19	0.19	0.11	0.11

Table 8: **Nonlinear Model Out-of-Sample Prediction**

This table reports out-of-sample, value-weighted hedge portfolio returns, going long the stocks in the highest predicted return decile and shorting the stocks in the lowest predicted return decile for different estimation periods and regularization terms. The table reports the first four moments of monthly percentage returns, turnover, and predictive slopes and R^2 for the hedge portfolios in Panel A, as well as for the long legs in Panel B and the short legs in Panel C. The model selection and estimation period spans from January 1965 to December of the year before start of the 10 year out-of-sample period.

Firms	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20	NYSE 20
Period	1965-1982	1965-1982	1965-1982	1965-1990	1965-1990	1983-2000	1983-2000
Sample Size	138,733	138,733	138,733	227,802	227,802	258,044	258,044
Regularization	Elastic Net	Adaptive LASSO	SCAD	Adaptive LASSO	SCAD	Adaptive LASSO	SCAD
# Steps	1	2	1	2	1	2	1
Knots	9	9	9	9	9	9	9
# Selected	16	2	1	3	4	3	2
IS Sharpe ratio	3.57	2.65	2.12	3.12	3.07	2.83	2.36
Panel A: Long-Short Portfolio							
Mean	1.93	1.18	0.67	2.21	2.8	1.47	1.5
Std. Dev.	3.69	4.04	4.2	6.56	6.02	6.5	7.49
Sharpe ratio	1.81	1.01	0.56	1.17	1.61	0.78	0.69
Skewness	0.78	0.34	0.59	0.26	0.14	0.14	-0.28
Kurtosis	2.56	0.93	0.99	2.11	1.77	0.53	0.91
β	0.83	0.75	0.75	0.95	0.97	0.65	0.68
R^2	0.03	0.05	0.05	0.05	0.05	0.05	0.05
TO	0.4	0.52	0.52	0.44	0.43	0.43	0.44
TO_R	0.41	0.52	0.52	0.45	0.43	0.43	0.44
Panel B: Long Leg							
Mean	2.2	1.55 2	1.3	2.63	3.11	1.32	1.39
Std. Dev.	6.41	6.68	6.79	8.05	7.96	9.9	10.38
Sharpe ratio	1.2	0.8	0.66	1.13	1.35	0.46	0.47
Skewness	-0.63	-0.85	-0.78	-0.09	0.07	-0.23	-0.46
Kurtosis	2.86	3.1	2.89	1.11	1.63	0.49	0.44
β	0.9	0.77	0.76	1.01	1.06	0.74	0.75
R^2	0.08	0.08	0.08	0.09	0.09	0.1	0.1
TO	0.21	0.24	0.24	0.22	0.22	0.22	0.22
TO_R	0.22	0.25	0.25	0.23	0.23	0.23	0.23
Panel C: Short Leg							
Mean	0.29	0.37	0.63	0.42	0.31	-0.14	-0.1
Std. Dev.	5.13	5.54	5.46	3.98	4.46	5.21	4.51
Sharpe ratio	0.2	0.23	0.4	0.37	0.24	-0.1	-0.08
Skewness	-1.15	-1.5	-1.45	-0.65	-0.3	-0.83	-0.69
Kurtosis	4.44	6.88	6.09	1.75	0.16	1.38	0.78
β	0.17	0.87	-9.2	0.12	0.28	0.17	0.02
R^2	0.08	0.03	0.04	0.06	0.06	0.11	0.1
TO	0.19	0.24	0.24	0.2	0.19	0.18	0.19
TO_R	0.19	0.24	0.24	0.19	0.18	0.18	0.19

Table 9: **Nonlinear Model Out-of-Sample Rolling Prediction**

This table reports out-of-sample, value-weighted hedge portfolio returns, going long the stocks in the highest predicted return decile and shorting the stocks in the lowest predicted return decile for different out-of-sample periods and regularization terms. The rolling estimation window is 20 years long and starts 1970. The table reports the first four moments of monthly percentage returns and predictive slopes and R^2 for the hedge portfolios in Panel A, as well as for the long legs in Panel B and the short legs in Panel C. The model selection and estimation period spans from January 1965 to December of the year before start of the 10 year out-of-sample period.

OoS Period	1991-2018	1991-2018	1991-2018	1991-2018	1991-2018	1991-2014	1991-1999
Firms	ALL	ALL	ALL	ALL	ALL	NYSE 10	NYSE 20
Sample Size	1,010,876	1,010,876	1,010,876	1,010,876	1,010,876	866,105	347,456
Regularization	MCP	SCAD	Adaptive LASSO	MCP	SCAD	SCAD	SCAD
# Steps	1	1	2	2	2	1	1
Avg. # Selected	11	13	10	10	10	13	13
IS Sharpe ratio	3	3.01	2.91	2.96	2.95	3.01	3.01
Panel A: Long-Short Portfolio							
Mean	2.63	2.92	2.53	2.53	2.53	2.96	3.87
Std. Dev.	5.02	5.45	5.08	5.08	5.08	5.22	4.37
Sharpe ratio	1.82	1.86	1.73	1.73	1.73	1.96	3.07
Skewness	1.86	2.35	1.84	1.84	1.84	1.82	1.67
Kurtosis	6.06	9.75	6.21	6.21	6.21	5.88	3.9
β	0.47	0.5	0.47	0.47	0.47	0.51	0.57
R^2	0.04	0.04	0.04	0.04	0.04	0.05	0.08
Panel B: Long Leg							
Mean	3.04	3.29	3.03	3.03	3.03	3.42	4.46
Std. Dev.	8.48	8.53	8.53	8.53	8.53	8.9	7.12
Sharpe ratio	1.24	1.34	1.23	1.23	1.23	1.33	2.17
Skewness	0.69	0.88	0.68	0.68	0.68	0.6	0.2
Kurtosis	3.29	3.81	3.28	3.28	3.28	2.98	1.84
β	0.59	0.61	0.59	0.59	0.59	0.66	0.8
R^2	0.1	0.09	0.1	0.1	0.1	0.1	0.08
Panel C: Short Leg							
Mean	0.4	0.37	0.5	0.5	0.5	0.47	0.6
Std. Dev.	5.96	5.81	5.91	5.91	5.91	6.11	5.22
Sharpe ratio	0.23	0.22	0.29	0.29	0.29	0.27	0.4
Skewness	-0.13	-0.1	-0.09	-0.09	-0.09	-0.05	-0.87
Kurtosis	2.25	2.28	2.06	2.06	2.06	2	2.68
β	0.06	-0.02	0.04	0.04	0.04	0.06	-0.21
R^2	0.08	0.08	0.08	0.08	0.08	0.08	0.06

B Figures

Figure 1: **Correlation between Firm Characteristics, 1970-2018**
This heat map presents pairwise Pearson correlations between rank-normalized firm characteristics. Dark blue (red) indicates positively (negatively) correlated features and uncorrelated variables are transparent. Features are ordered by complete hierarchical clustering.

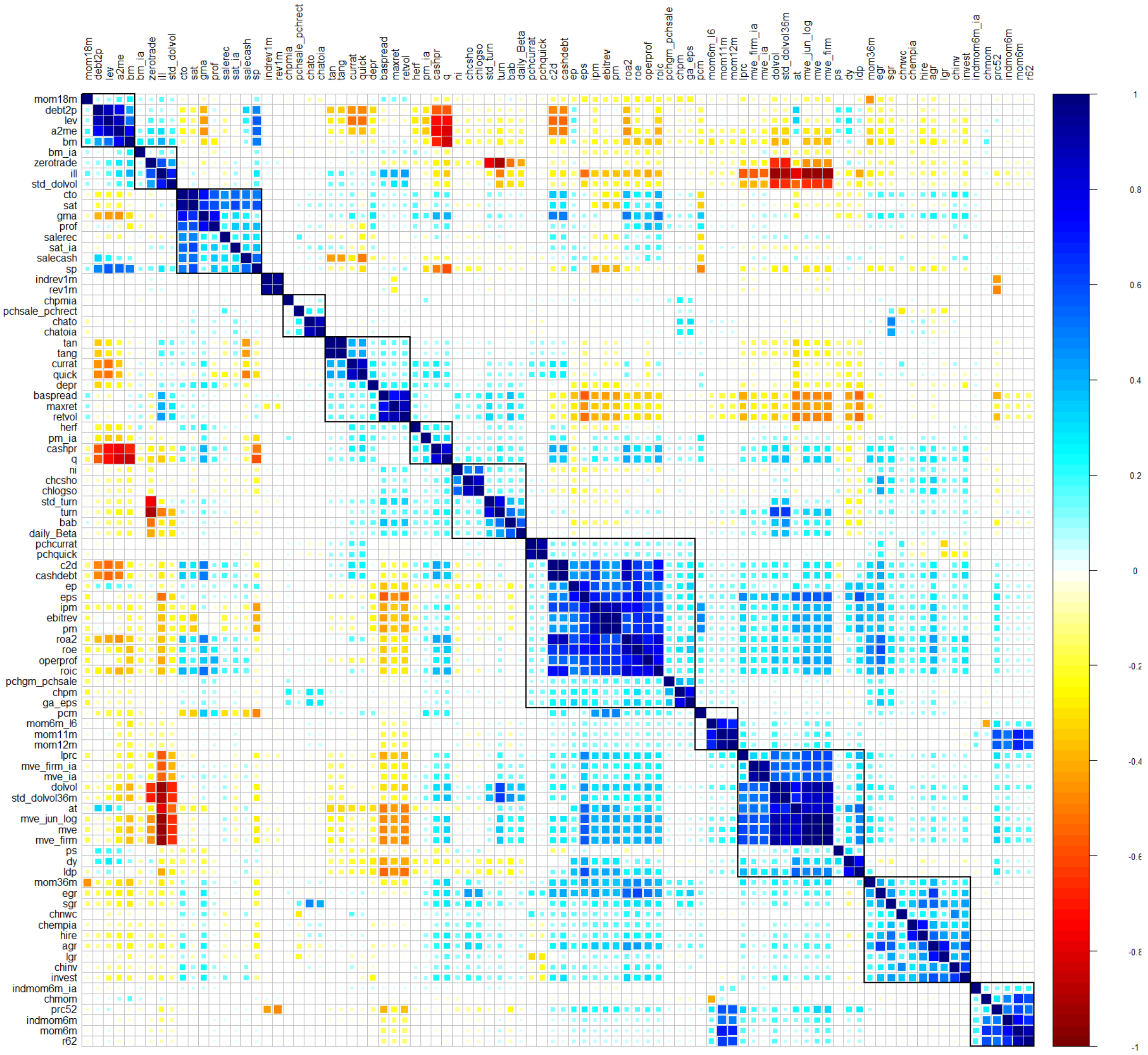

Figure 2: **Correlation between Firm Characteristics NYSE 10%, 1970-2018**
This heat map presents pairwise Pearson correlations between rank-normalized firm characteristics of all stocks that are larger than the first decile of NYSE listed stocks. Dark blue (red) indicates positively (negatively) correlated features and uncorrelated variables are transparent. Features are ordered by complete hierarchical clustering.

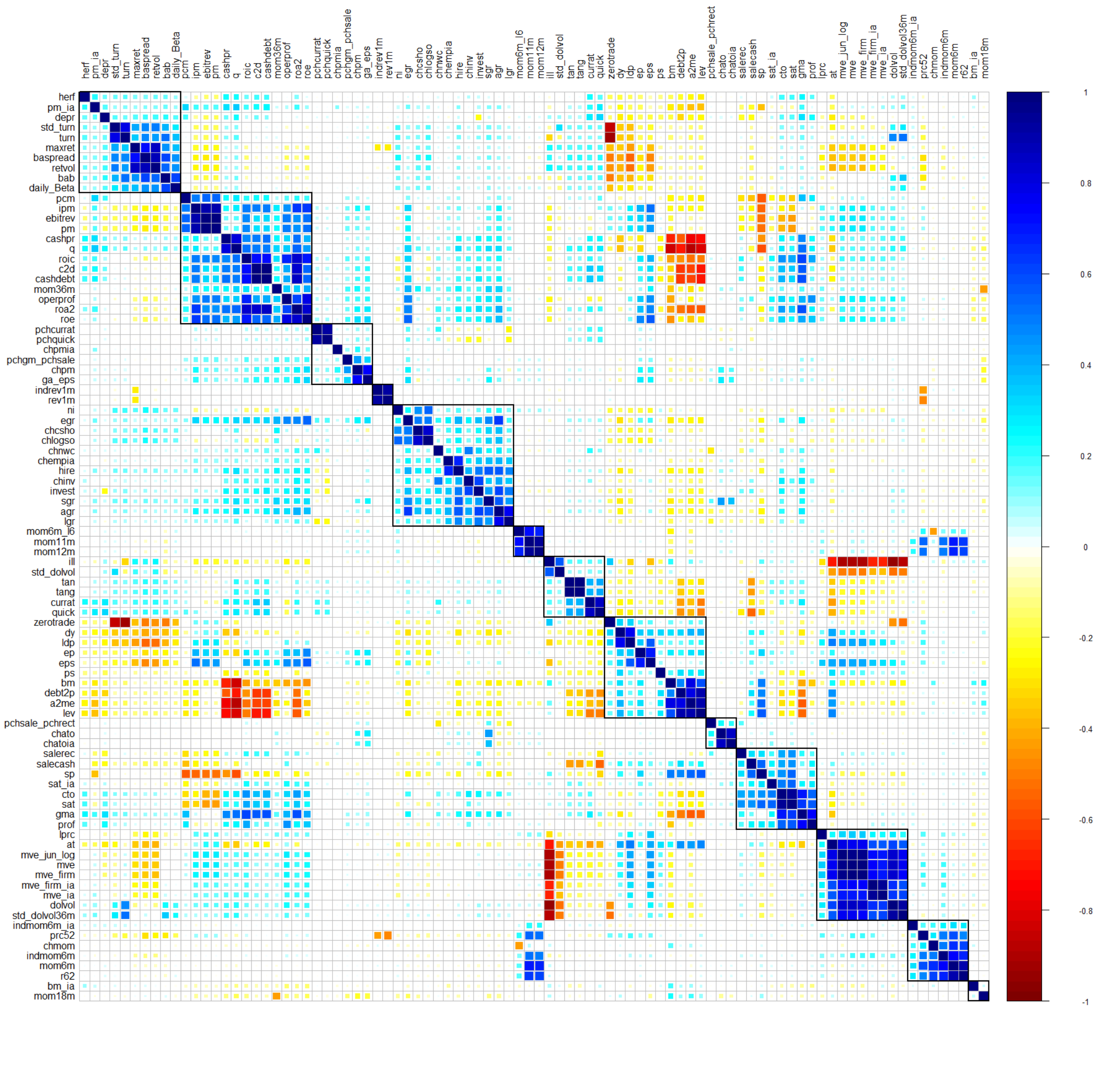

Figure 3: **Correlation between Firm Characteristics NYSE 20%, 1970-2018**
This heat map presents pairwise Pearson correlations between rank-normalized firm characteristics of all stocks that are larger than the second decile of NYSE listed stocks. Dark blue (red) indicates positively (negatively) correlated features and uncorrelated variables are transparent. Features are ordered by complete hierarchical clustering.

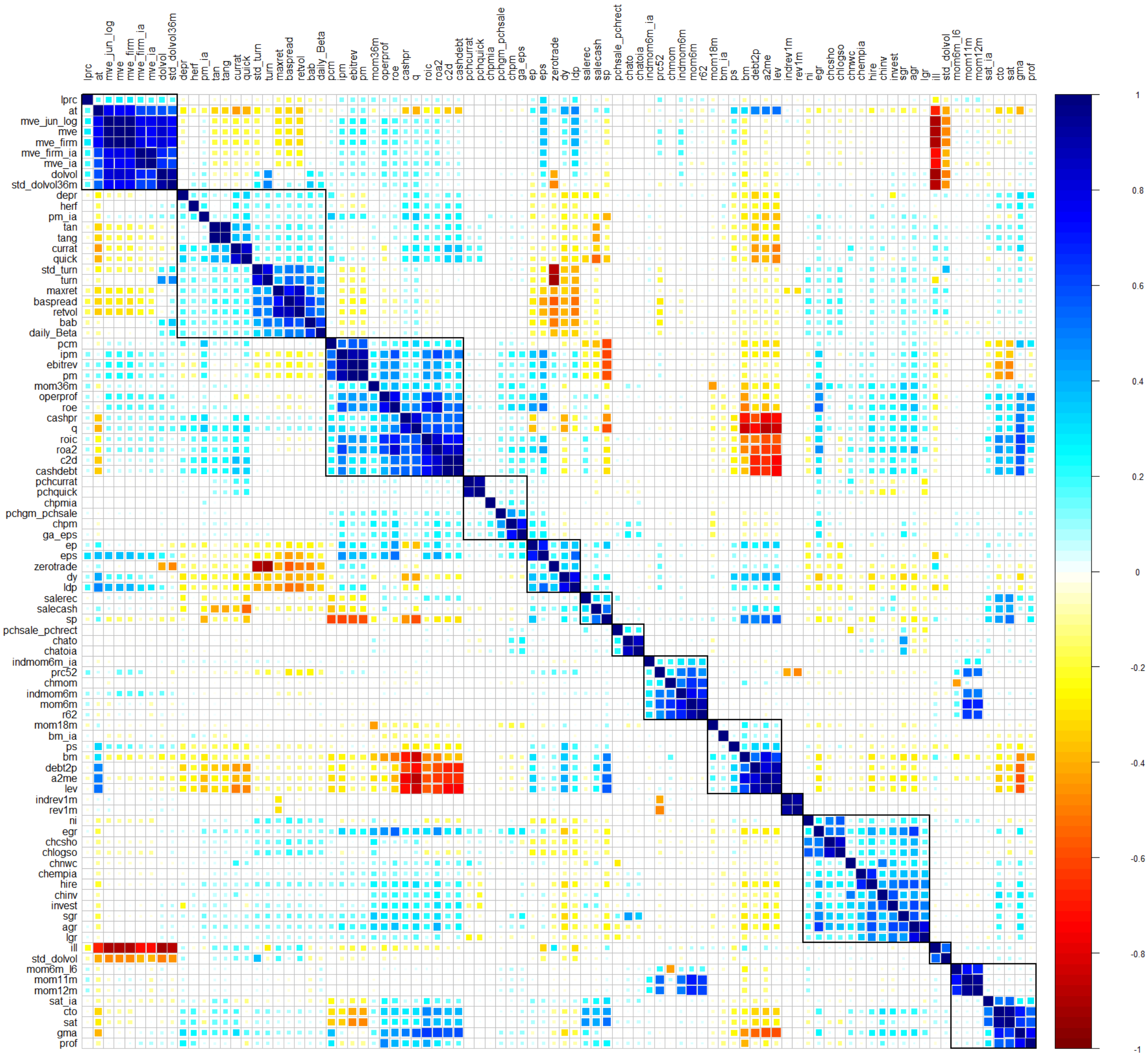

Figure 4: **Anomaly Covariances on CAPM and Fama French 3 Factors, 1965-2018**

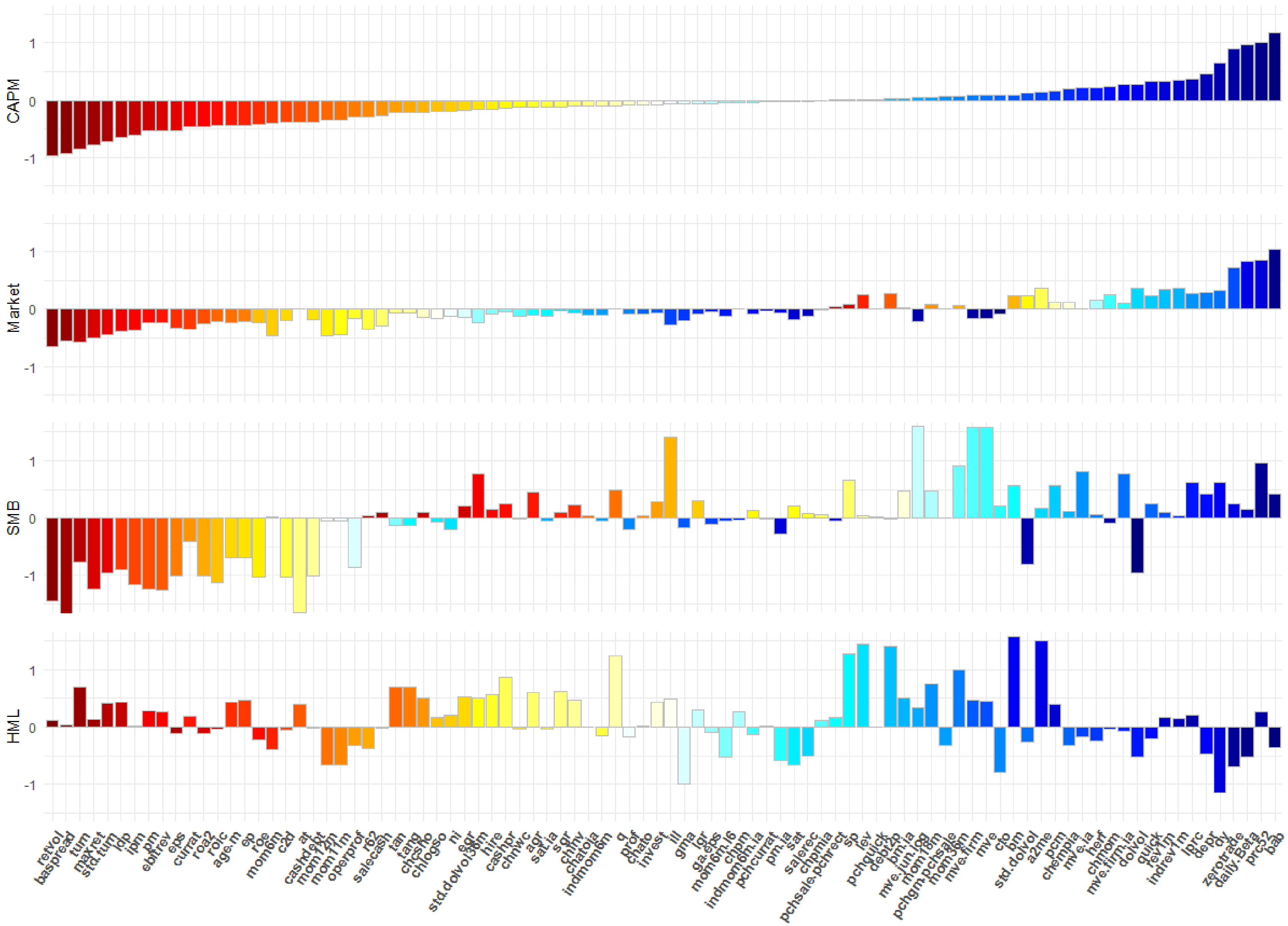

Figure 5: **Time-Varying Feature Subset Selection (Rolling Selection), 1991-2018**

C Anomaly Definition

1. **A2ME:** I follow Bhandari (1988) and define assets-to-market cap as total assets (AT) over market capitalization as of December t-1. Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).

2. **Firm Age:** The number of months that a firm has been listed in the CRSP database. Updated monthly. I exclude stocks with a price below $5.

3. **Asset growth:** Annual percent change in total assets (at).

4. **Total Assets:** Total assets (AT) as in Gandhi and Lustig (2015).

5. **Beta:** We follow Frazzini and Pedersen (2014) and define the CAPM beta as product of correlations between the excess return of stock i and the market excess return and the ratio of volatilities. We calculate volatilities from the standard devi ations of daily log excess returns over a one-year horizon requiring at least 120 observations. We estimate correlations using overlapping three-day log excess returns over a five-year period requiring at least 750 non-missing observations.

6. **Bid-Ask spread:** Monthly average of daily bid-ask spread divided by average of daily spread.

7. **Book-to-market:** Book value of equity (ceq) divided by end of fiscal-year-end market capitalization.

8. **Ind.-adj. book-to-market:** Industry adjusted book-to-market ratio.

9. **C2D:** Cash flow to price is the ratio of income and extraordinary items (IB) and depreciation and amortization (dp) to total liabilities (LT).

10. **Cash Flow to debt:** Earnings before depreciation and extraordinary items (ib+dp) divided by avg. total liabilities (lt).

11. **Cash productivity:** Fiscal year end market capitalization plus long term debt (dltt) minus total assets (at) divided by cash and equivalents (che).

12. **Asset Turnover:** Asset Turnover (t) = Sales (t)/ ((Net Operating Assets (t) + Net Operating Assets (t-1)) / 2) Net Operating Assets = Receivables + Total Inventory + Other Current Assets + PP&E + Intangibles - Payables - Other Current Liabilities - Other Liabilities. Updated annually. I exclude stocks with a price below $5.

13. **Ind.-adj. Chg. in asset turnover:** 2-digit SIC - fiscal-year mean adjusted change in sales (sale) divided by average total assets (at).

14. **Chg. in shares outstanding:** Annual percent change in shares outstanding (csho).

15. **Ind.-adj. Chg. in employees:** Industry-adjusted change in Number of employees.

16. **Chg. in inventory:** Change in inventory (inv) scaled by average total assets (at).

17. **ΔSO:** Log change in the split adjusted shares outstanding as in Fama and French (2008). Split adjusted shares outstanding are the product of Compustat shares outstanding (CSHO) and the adjustment factor (AJEX).

18. **Chg. in 6-month momentum:** Cumulative returns from months t-6 to t-1 minus months t-12 to t-7.

19. **Chg. in Net Working Capital:** Yearly change in net working capital scaled by total assets. Net working capital is measured as current assets minus current liabilities. Current assets are measured as total current assets minus cash and cash equivalents. Current liabilities are measured as total current liabilities minus debt in current liabilities.

20. **Chg. in Profit:** Profit Margin (t) - Profit Margin (t-1). Updated annually. Data from year t are used to forecast returns for 12 months beginning in April of year t+1. I exclude stocks with a price below $5.

21. **Ind.-adj. Chg. in profit margin:** 2-digit SIC - fiscal-year mean adjusted change in income before extraordinary items (ib) divided by sales (sale).

22. **CTO:** We follow Haugen and Baker (1996) and define capital turnover as ratio of net sales (SALE) to lagged total assets (AT).

23. **Current ratio:** Current assets divided by current liabilities.

24. **Beta daily:** Beta daily is the sum of the regression coefficients of daily excess returns on the market excess return and one lag of the market excess return as in Lewellen and Nagel (2006).

25. **Debt2P:** Debt to price is the ratio of long-term debt (DLTT) and debt in current liabilities (DLC) to the market capitalization as of December t-1 as in Litzenberger and Ramaswamy (1979). Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).

26. **Depreciation divided PP&E:** Depreciation divided by PP&E.

27. **Dollar trading volume:** Natural log of trading volume times price per share from month t-2.

28. **Dividend to price:** Total dividends (dvt) divided by market capitalization at fiscal year-end.

29. **Profit Margin:** EBIT / Revenues. Updated annually. I exclude stocks with a price below $5.

30. **Growth in common equity:** Annual percent change in book value of equity (ceq).

31. **Earnings to price ratio:** Annual income before extraordinary items (ib) divided by end of fiscal year market cap.

32. **Earnings per share:** We follow Basu (1977) and define earnings per share as the ratio of income before extraordinary items (IB) to shares outstanding (SHROUT) as of December t-1..

33. **Earnings Consistency:** Geometric average of Earnings Growth from t-1 to t-5. Earnings growth is: Earnings Per Share (t) – Earnings Per Share (t-1) / ((Absolute Value of Earnings Per Share (t-1) + Absolute Value of Earnings Per Share (t-2) / 2). I exclude stocks with a price below $5 and an absolute value of growth greater than 6 if growth is positive this year, but negative last year, or vice versa. Updated annually.

34. **Gross profitability:** Revenues (revt) minus cost of goods sold (cogs) divided by lagged total assets (at).

35. **Industry sales concentration:** 2-digit SIC - fiscal-year sales concentration (sum of squared percent of sales in industry for each company).

36. **Employee growth rate:** Percent change in Number of employees (emp).

37. **Illiquidity:** Average of daily (absolute return / dollar volume).

38. **Industry Momentum:** Value-weighted return from t-6 to t-1 within each industry. Industry is measured with two-digit SIC code. Updated monthly.

39. **Industry Relative Reversals:** In each month, firms are sorted based on the difference between their prior month's return and the prior month's return of their industry based on the Fama and French 49 industries.

40. **CAPEX and inventory:** Annual change in gross property, plant, and equipment (ppegt) + annual change in inventories (invt) all scaled by lagged total assets (at).

41. **IPM:** We define pre-tax profit margin as ratio of pre-tax income (PI) to sales (SALE).

42. **LDP:** We follow Litzenberger and Ramaswamy (1979) and define the dividend-price ratio as annual dividends over last months price (PRC). We measure annual dividends as the sum of monthly dividends over the last 12 months. Monthly dividends are the scaled difference between returns including dividends (RET) and returns excluding dividends (RETX).

43. **Leverage:** Total liabilities (lt) divided by fiscal year end market capitalization.

44. **Growth in long-term debt:** Annual percent change in total liabilities (lt).

45. **Maximum daily return:** Maximum daily return from returns during calendar month t-1.

46. **r12-2 :** We define momentum as cumulative return from 12 months before the return predifftion to two months before as in Fama and French (1996).

47. **Momentum-Reversal:** Buy and hold returns from t-18 to t-13. Updated monthly.

48. **36-month momentum:** Cumulative returns from months t-36 to t-13.

49. **Momentum-Volume:** Buy- and- hold returns from t-6 through t-1. We limit the sample to high trading volume stocks, i.e., stocks in the highest quintile of average monthly trading volume measured over the past six months. NYSE and AMEX only. Updated monthly.

50. **Lagged Momentum:** Buy-and-hold returns from t-13 through t-8. Updated monthly.

51. **Size:** Natural log of market capitalization at end of month t-1.

52. **Ind.-adj. size:** 2-digit SIC industry-adjusted fiscal year-end market capitalization.

53. **Size:** Natural log of market capitalization at end of month t-1.

54. **LME:** Size is the total market capitalization of the previous month defined as price (PRC) times shares outstanding (SHROUT) as in Fama and French (1992).

55. **LME Ind.-adj.:** Industry-adjusted-size is the total market capitalization of the previous month defined as price (PRC) times shares outstanding (SHROUT) minus the average industry market capitalization at the Fama-French 48 industry level as in Asness et al. (2000).

56. **Ind.-adj. size:** 2-digit SIC industry-adjusted fiscal year-end market capitalization.

57. **Size:** The log of market value of equity of June. Updated monthly.

58. **Net Issuance:** Net issuance is the year-over-year percent change in adjusted shares outstanding, CFACSHR SHROUT, where CFACSHR is the monthly CRSP split adjustment factor and SHROUT is common shares outstanding. Rebalanced annually, uses the recent period.

59. **Operating profitability:** Revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders equity.

60. **Chg. in current ratio:** Percent change in currat.

61. **Chg. in gross margin - Chg. in sales:** Percent change in gross margin (sale-cogs) minus percent change in sales (sale).

62. **Chg. in quick ratio:** Percent change in quick.

63. **Chg. in sales - Chg. in A/R:** Annual percent change in sales (sale) minus annual percent change in receivables (rect).

64. **PCM:** The price-to-cost margin is the difference between net sales (SALE) and, costs of goods sold (COGS) diffidef by net sales (SALE) as in Gorodnichenko and Weber, (2016) and D'Acunto, Liu, P?ueger, and Weber (2017)

65. **PM:** The profit margin is operating income after defreciation (OIADP) over sales, _x000D_ (SALE) as in Soliman (2008)

66. **PM Ind.-adj.:** The adjusted profit margin is operating income after defreciation, (OIADP) over net sales (SALE) minus the average profit margin at the Fama-French 48, industry level as in Soliman (2008)

67. **52-Week High:** Price scaled by the highest price or bid/ask average during the last 12 months. Updated monthly.

68. **Prof:** We follow Ball, Gerakos, Linnainmaa, and Nikolaev (2015) and define, profitability as gross profitability (GP) diffidef by the book value of equity as defined, above

69. **Tobin's Q:** Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC), minus cash and short-term investments (CEQ), minus deferred taxes (TXDB) scaled by, _x000D_ total assets (AT)

70. **Quick ratio:** (current assets - inventory) / current liabilities.

71. **r6-2 :** We define r6-2 as cumulative return from 6 months before the return prediﬀtion, to two months before as in Jegadefsh and Titman (1993)

72. **Return volatility:** Standard deviation of daily returns from month t-1.

73. **Short-Term Reversal:** Return in month t. Updated monthly. I exclude stocks with a price below $5.

74. **ROA:** Return-on-assets is income before extraordiﬀary items (IB) to lagged total, assets (AT) following Balakrishnan, Bartov, and Faurel (2010)

75. **Return-on-Equity:** Net income scaled by book value of equity. Exclude if price<\$5. Updated annually.

76. **Return on invested capital:** Annual earnings before interest and taxes (ebit) minus non-operating income (nopi) divided by non-cash enterprise value (ceq+lt-che).

77. **Sales to cash:** Annual sales divided by cash and cash equivalents.

78. **Sales to receivables:** Annual sales divided by accounts receivable.

79. **SAT:** We follow Soliman (2008) and define asset turnover as the ratio of sales, _x000D_ (SALE) to total assets (AT)

80. **SAT Ind.-adj.:** We follow Soliman (2008) and define adjusted asset turnover as the ratio, of sales (SALE) to total assets (AT) minus the average asset turnover at the Fama-French, 48 industry level

81. **Sales growth:** Annual percent change in sales (sale).

82. **Sales/Price:** Total revenues divided by stock price. Updated annually.

83. **Volatility of liquidity:** Monthly standard deviation of daily dollar trading volume.

84. **Volume Variance:** Standard deviation of monthly trading volume over the last 36 months. NYSE only. Updated monthly.

85. **Volatility of liquidity:** Monthly standard deviation of daily share turnover.

86. **Tan:** We follow Hahn and Lee (2009) and define tangibility as (0.715 x total, receivables (RECT) + 0.547 x inventories (INVT) + 0.535 x property, plant and, equipment (PPENT) + cash and short-term investments (CHE)) / total assets (AT)

87. **Debt capacity/firm tangibility:** Cash holdings + 0.715 receivables + 0.547 in-

ventory + 0.535 PPE/total assets.

88. **Share turnover:** Average monthly trading volume for most recent 3 months scaled by Number of shares outstanding in current month.

89. **Industry Momentum:** In each month, the Fama and French 49 industries are sorted on their value-weighted firms in decile 10 (from the 5 winner industries) form the value-weighted portfolio. Rebalanced monthly.

90. **Zero trading days:** Turnover weighted Number of zero trading days for most recent 1 month.