Nonlinearities Everywhere: Sparse Supervised Learning of Market Anomalies

Gabriel Kaiser

University of Luxembourg, LSF

Monday 27th April, 2020

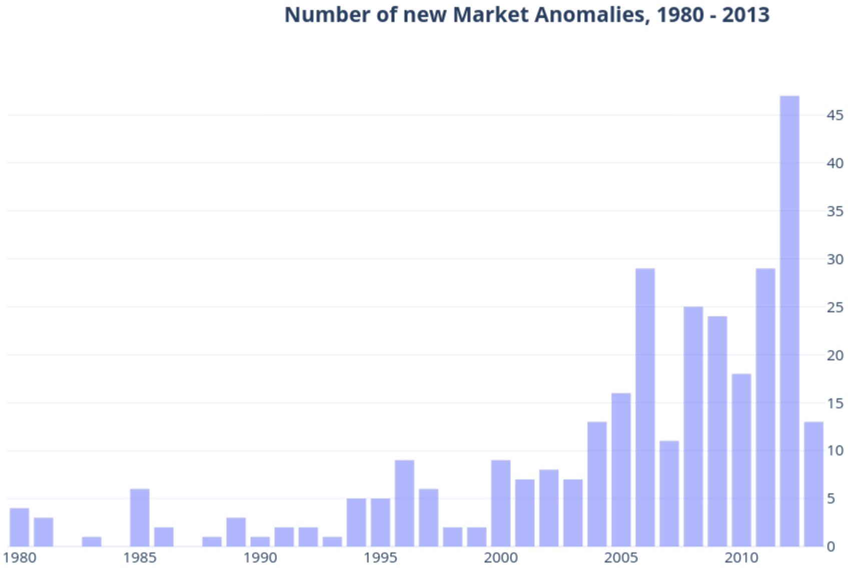
This paper

Can we find a sparse set of characteristics predicting expected hedge returns in the cross-section?

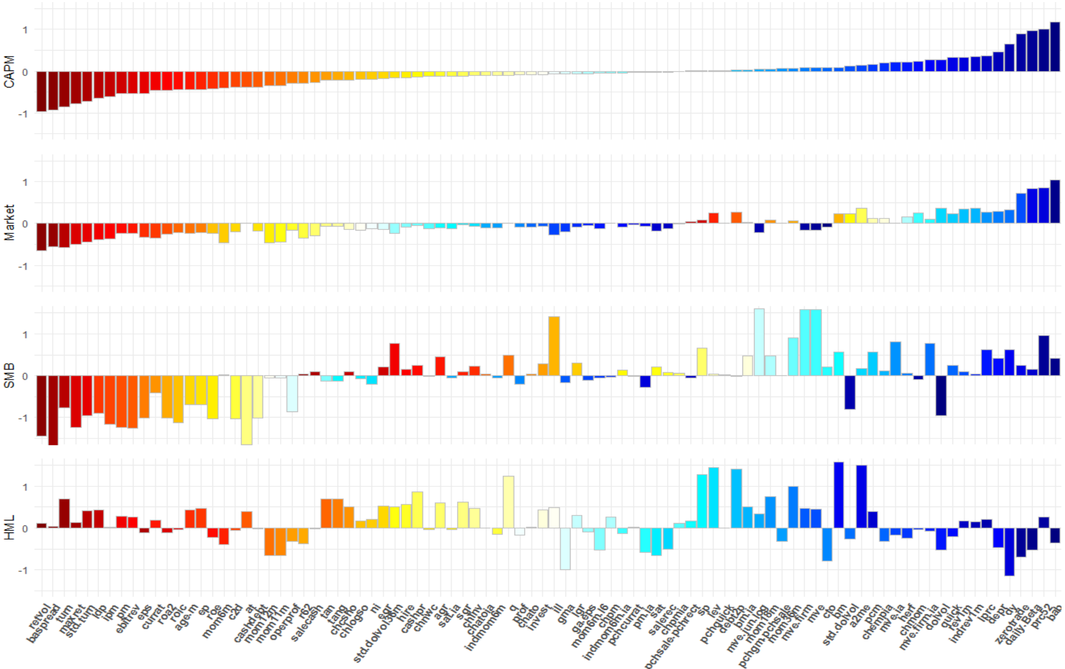
- Replicate 90 cross-sectional anomalies
- Model selection and parameter estimation
- The models select between 2 to 8 characteristics in an in-sample and out-of-sample comparison
- The predictability of characteristics is time varying
- Out of 26 different predictors only 4 are consistent through time

Outline

The 'Zoo of Anomalies'



Datasource: Harvey, Liu, and Zhu 2016


CAPM
Market
SMB
HML
1
0
-1
retvol baspread turn maxret std.turn idp ipm pm ebitrev eps currat roa2 roic age.m ep roe mom6m c2d at cashdebt mom12m mom11m operprof r62 salecash tan tang chcsho chlogso ni egr std.dolvol36m hire cashpr chnwc agr sat.ia sgr chinv chatoia indmom6m q prof chato invest ill gma lgr ga.eps mom6m.l6 chpm indmom6m.ia pchcurrat pm.ia sat salerec chpmia pchsale.pchrect sp lev pchquick debt2p bm.ia mve.jun.log mom18m pchgm.pchsale mom36m mve.firm mve cto bm std.dolvol a2me pcm chempia mve.ia herf chmom mve.firm.ia dolvol quick rev1m indrev1m lprc dept dy zerotrade daily.Beta prc52 bab

anomaly [/a'nom.a.li/]

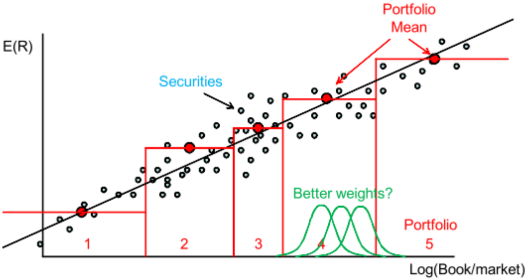
- Dictionary
  - a deviation from the common rule, type, arrangement, or form.
  - an inconsistency

anomaly [/a'nom.a.li/]

- Dictionary
  - a deviation from the common rule, type, arrangement, or form.
  - an inconsistency
- Asset Pricing:
  - Synonym: **Market anomaly**
  - Distortions in returns that contradict the **efficient market hypothesis** (EMH)
  - A stock is priced differently to how a model (eg.: CAPM) predicts it will be priced
  - Any **arbitrage** opportunity to earn excess returns undermines the assumptions of market efficiency
  - Past price information, such as **Momentum** violate the **weak** form EMH
  - Fundamental firm characteristics, such as **Value** violate the **semi-strong** form EMH

Characteristics and Returns

- Characteristics seems to matter for future returns, i.e. $\mathbb{E}_t[R_{it}|C_{1,it-1},\ldots,C_{M,it-1}]$
- Literature suggests around 500 firm characteristics
- Does a subset in a multivariate mapping exist?
- Is the relationship persistent or time-varying?
- How does the functional form look like?
  - linear, non-linear, interactions, state dependent



Graph from Cochrane 2011;

Classic approaches in empirical asset pricing

- Time series regressions: Black, Jensen, and Scholes (1972) suggest N timeseries regressions with T observations on K factors, i.e.

$$R^e_{i,t} = \alpha_i + \beta_{i,M} R_{M,t} + \beta_{i,SMB} R_{SMB,t} + \beta_{i,HML} R_{HML,t} + \beta_{i,LIQ} R_{LIQ,t} + \epsilon_{i,t}$$

- Cross sectional regressions: are N time-series, followed by 1 cross-sectional regressions on sample averages. Now, the factors do not need to be return portfolios anymore.

$$\mathbb{E}_T[R^e_i] = \alpha_i + \hat{\beta}_{i,M}\lambda_M + \hat{\beta}_{i,SMB}\lambda_{SMB} + \hat{\beta}_{i,HML}\lambda_{HML} + \hat{\beta}_{i,LIQ}\lambda_{LIQ} + \nu_i$$

- Fama-MacBeth regression: is a two-pass approach that combines N time-series and T cross-sectional regressions, where sample avarages define risk prices $\mathbb{E}_T[\hat{\lambda}]$

$$R^e_{i,t} = \lambda_{0,t} + \hat{\beta}_{i,M}\lambda_{M,t} + \hat{\beta}_{i,SMB}\lambda_{SMB,t} + \hat{\beta}_{i,HML}\lambda_{HML,t} + \hat{\beta}_{i,LIQ}\lambda_{LIQ,t} + \xi_{i,t}$$

- Independent portfolio sorting
  - simple and robust non-parametric method
  - Fama and French exclusively use value-weighted portfolio sorting

Shortcomings of standard approaches

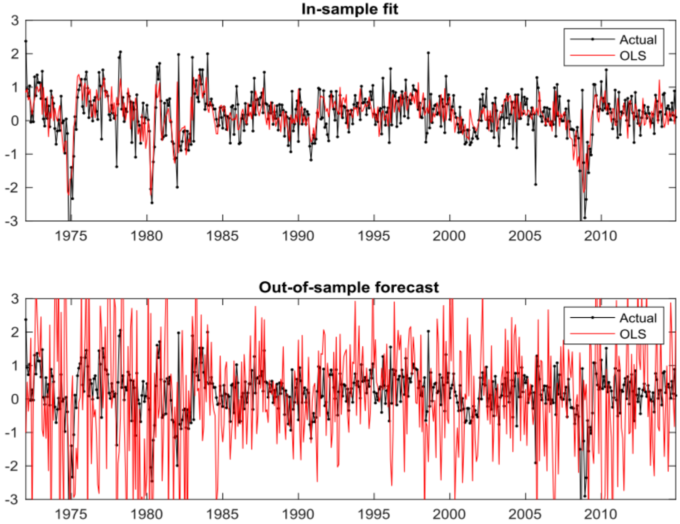
- Big Data problem: $500 \cdot \frac{500 \cdot 499}{2} \cdot 30 \cdot 68/3 \approx 42 \cdot 10^9$ possible parameter estimates
  - 500 characteristics that forecast returns
  - Possibility of dual interactions: $500 \cdot 499 \ / \ 2$
  - Possibility of non-linearities: Cubic splines with 10 knots
  - Possibility of time variation: 68 years (Compustat starts 1950)

$$\Rightarrow NT \ll p$$

- Fama MacBeth regression & Co and Independent portfolio sorts suffer from:
  - High-dimensional issue or the number of sorts is restricted or multicollinearity
  - No model selection technique
  - Assumes constant cross-sectional expected returns per percentile

$$\mathbb{E}(R_i | C_i \in \mathbb{1}_p) = \mathbb{E}_t(R_i | C_i \in \mathbb{1}_p)$$

Curse of Dimensionality



This Paper

- The predictive model tries to approximate $\mathbb{E}_t[R_{it}|C_{1,it-1},\ldots,C_{Z,it-1}]$:

$$R_t = C_{t-1}\beta + \epsilon_t$$

  R_t is a N x 1 vector of returns and C_{t-1} is a N x Z matrix of characteristics
- Minimizing the quadratic function: $\hat{\beta} = (X^TX)^{-1}X^Ty$ is problematic if $Z >> N$
- **Multicollinearity:** singular (X^TX) and $\hat{\beta}$ is not unique $\Rightarrow$ Drop redundant variables
- **Overfitting:** Gauss-Markov Th states that the MSE estimator is BLUE

$$\downarrow \mathbb{E}\left[\left(y-\hat{f}(X)\right)^2\right] = \uparrow \text{Bias}\left[\hat{f}(X)\right]^2 + \downarrow \text{Var}\left[\hat{f}(X)\right]$$

This Paper

- The predictive model tries to approximate $\mathbb{E}_t[R_{it}|C_{1,it-1},\ldots,C_{Z,it-1}]$:

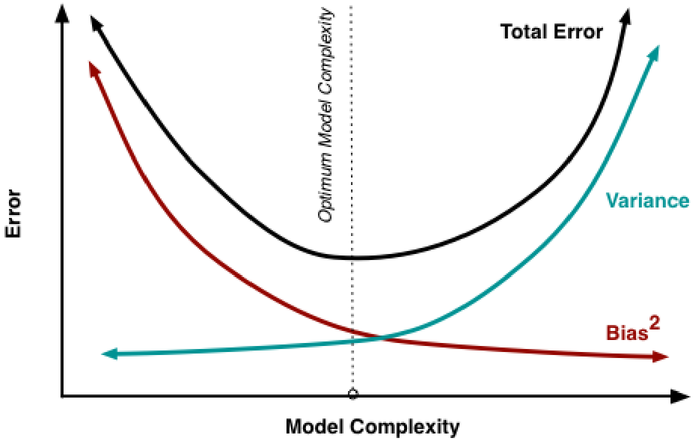
$$R_t = C_{t-1}\beta + \epsilon_t$$

  R_t is a N x 1 vector of returns and C_{t-1} is a N x Z matrix of characteristics
- Minimizing the quadratic function: $\hat{\beta} = (X^TX)^{-1}X^Ty$ is problematic if $Z >> N$
- **Multicollinearity:** singular (X^TX) and $\hat{\beta}$ is not unique $\Rightarrow$ Drop redundant variables
- **Overfitting:** Gauss-Markov Th states that the MSE estimator is BLUE

$$\downarrow \mathbb{E}\left[\left(y-\hat{f}(X)\right)^2\right] = \uparrow \text{Bias}\left[\hat{f}(X)\right]^2 + \downarrow \text{Var}\left[\hat{f}(X)\right]$$

$\Rightarrow$ **Regularization**

CoD: Bias - Efficiency Tradeoff



Literature on Asset Pricing

1. **Cochrane 2011**
   - Which characteristics really matter given the multidimensional vector of characteristics?
2. **Fama and French 2018**
   - Use maximum Sharpe ratio metrics to rank factor models between non- and nested models
3. **Feng, Giglio, and Xiu 2019**
   - Develop a two-pass cross-sectional Lasso regression.
4. **Kozak, Nagel, and Santosh 2019**
   - Find a robust SDF that is not sparse by testing Lasso and Ridge regressions.
5. **Cattaneo et al. 2018**
   - Develop a univariate methodology of non-parametric partitioned regression to overcome the shortfalls of portfolio sorting.
6. **Freyberger, Neuhierl, and Weber 2020**
   - Estimate a non-parametric function in an adaptive group lasso framework.

Literature on Signal Processing and Statisitcal Learning

1. **Tibshirani 1996**
   - Least Absolute Shrinkage and Selection Operator (LASSO)
2. **Zou 2006**
   - Adaptive LASSO and derives Oracle properties
3. **Fan and Li 2001**
   - Smoothly Clipped Absolute Deviation (SCAD)
4. **Zou and Hastie 2005**
   - Elastic Net that linearly combines the L_1 and L_2 penalties
5. **Zhang 2010**
   - Minimax concave penalty (MCP)
6. **Huang, Horowitz, and Wei 2010**
   - Variable selection in nonparametric additive models

Model

- Additivity assumption: $\mathbb{E}_t[R_{it+1}|C_{1,it},\dots,C_{Z,it}] = \mathbb{E}_t[R_{it+1}|C_{1,it}] + \dots + \mathbb{E}_t[R_{it+1}|C_{Z,it}]$
- Rank-normalized firm characteristics $\tilde{C}_{t,z}$, i.e. $\mathbb{R}^{Nx1} \mapsto [0,1]$
- General Model:

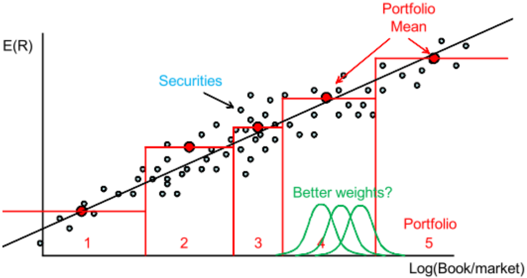
$$\hat{\beta}_{s,z,k} = \underset{\beta_{s,z,k}}{\text{argmin}} \sum_{i=1}^{N} \Bigg(R_{i,t+1} - \underbrace{\sum_{z=1}^{Z} \beta'_{s,z} \cdot g_z(\tilde{C}_{i,t,z},\theta)}_{\sum_{z=1}^{Z}\sum_{k=1}^{K+2} \beta_{s,z,k} \cdot g_z(\tilde{C}_{i,t,z},\theta)} \Bigg)^2 + \lambda_s \sum_{z=1}^{Z} R_{type}(\beta_{s,z}) \qquad (1)$$

- Non-linear part: $g(\tilde{C}_{t,z}) : \mathbb{R}^N \mapsto \mathbb{R}^{(K+2)\times N}$ and in particular g_k:

$$g_1(x) = 1,\ g_2(x) = x,\ g_3(x) = x^2,\ \underbrace{g_k(x) = max\{x - \frac{k-3}{K-2}, 0\}}_{\text{Portfolio-Sorting}} \quad \forall k \in [4, K]$$

Portfolio-Sorting

$$g_k(\tilde{C}_{t,z}) = max\{\tilde{C}_{t,z} - \frac{k}{5}, 0\} \quad \forall \text{ Knots } k \in [1,4]$$



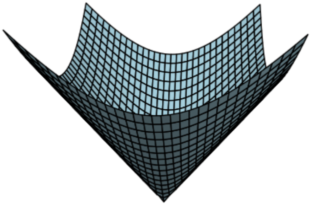
- Portfolio Mean:

$$\mathbb{E}(R|\tilde{C} \in \mathbb{1}_p) = \int_{-1}^{\infty} R d\mathbb{P}(R|\phi(C)) = \int_{-1}^{\infty} \mathbb{1}_{(p-1,p]} R dR$$

Regularization Type 1

- **Group LASSO:** Least Absolute Shrinkage and Selection Operator

$$R_{\text{LASSO}} = \|\beta_{s,G_z}\|_{2,1} = \frac{1}{|G_z|}\left(\sum_{k=1}^{K+2}\beta_{s,k,z}^2\right)^{\frac{1}{2}}$$

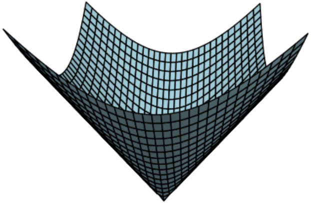


Regularization Type 2

- **Adaptive Group LASSO:**

$$R_{\text{Adaptive}} = w_z \cdot R_{\text{LASSO}}$$

$$w_z = \begin{cases} \frac{1}{|G_z|}\left(\sum_{k=1}^{K+2} \hat{\beta}^2_{1,k,z}\right)^{-\frac{1}{2}} & \text{if } \sum_{k=1}^{K+2} |\hat{\beta}_{1,k,z}| \neq 0 \\ \infty & \text{if } \sum_{k=1}^{K+2} |\hat{\beta}_{1,k,z}| = 0 \end{cases}$$

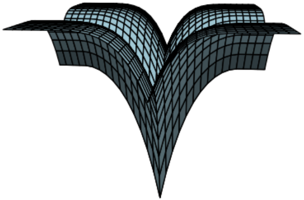


Regularization Type 3

- **Group SCAD:** Smoothly Clipped Absolute Deviation

$$R_{\text{SCAD}} = p^{SCAD}_{\lambda,\gamma}\left(\|\boldsymbol{\beta}_{G_z}\|_{2,1}\right)$$

$$p^{SCAD}_{\lambda,\gamma}(\beta) = \begin{cases} \lambda|\beta| & \text{if} \quad |\beta| \leq \lambda \\ \frac{|\beta|^2 - 2\gamma\lambda|\beta| + \lambda^2}{2(1-\gamma)} & \text{if} \quad \lambda < |\beta| \leq \gamma\lambda \\ \frac{(\gamma+1)\lambda^2}{2} & \text{if} \quad |\beta| \geq \gamma\lambda. \end{cases}$$

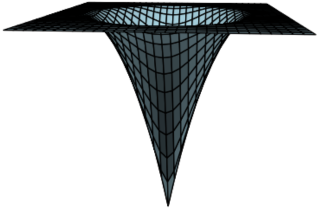


Regularization Type 4

- **Group MCP:** Minimax concave penalty

$$R_{\text{MCP}} = p_{\lambda,\gamma}^{MCP}\left(\|\boldsymbol{\beta}_{G_z}\|_{2,1}\right)$$

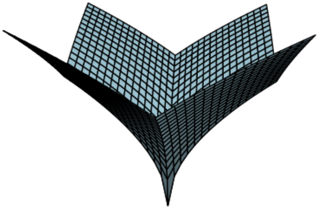
$$p_{\lambda,\gamma}^{MCP}(\beta) = \begin{cases} \lambda|\beta| + \frac{\beta^2}{2\gamma} & \text{if} \quad |\beta| \leq \gamma\lambda \\ \frac{\gamma\lambda^2}{2} & \text{if} \quad |\beta| > \gamma\lambda. \end{cases}$$



Regularization Type 5

- **Group Elastic Net:**

$$R_{\text{Elastic Net}} = \alpha \|\boldsymbol{\beta}_{G_z}\|_{2,1} + (1-\alpha)\|\boldsymbol{\beta}_{G_z}\|_2$$
$$= \frac{1}{|G_z|}\left(\alpha\left(\sum_{k=1}^{K+2}\beta_{s,k,z}^2\right)^{\frac{1}{2}} + (1-\alpha)\sum_{k=1}^{K+2}\beta_{s,k,z}^2\right)$$



Data

- Daily, monthly and annual data from CRSP
- COMPUSTAT industrial database
- Time-horizon is from January 1965 to December 2017
- Domestic ordinary common stocks listed on NYSE, AMEX or NASDAQ (19000 stocks)
- Exclude firms with prices below $5 a share and/or that are in the bottom NYSE size decile
- Deciles as breakpoints for the long-short self-financing portfolio
- All portfolios are value-weighted (decreasing alpha) and monthly re-balanced

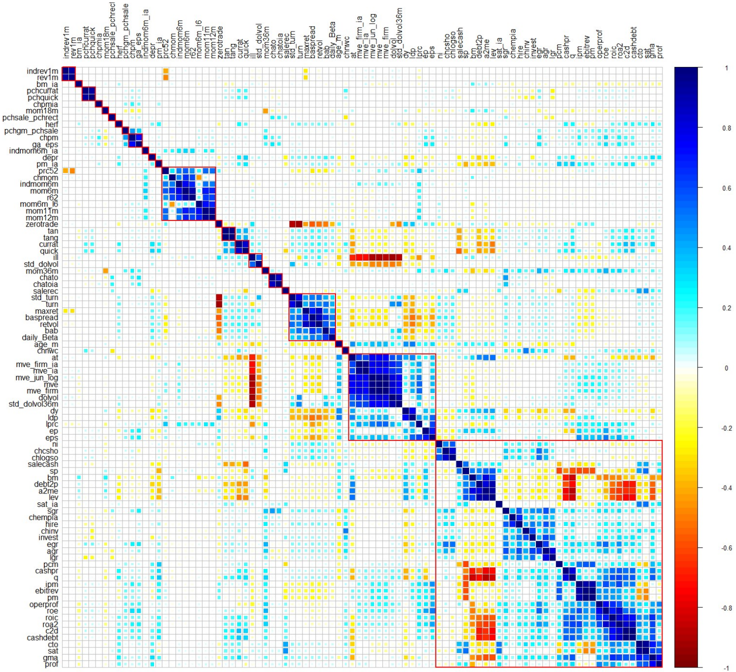
Long/Short portfolio returns

Table: **Summary Statistics**

This table reports characteristic sorted 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.

Anomalies	Mean	Median	Std. Dev.	Skew	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

Heatmap



Linear Model In-Sample Results

Table: **Linear Model In-Sample Results**
This table reports selected characteristics obtained from penalized regressions from a total universe of 90 market anomalies. 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.

Firms	All	All	NYSE 10	NYSE 10	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	861,982	861,982	690,545	690,545
Regularization	SCAD	Adaptive LASSO	SCAD	Adaptive LASSO	SCAD	Adaptive LASSO
# Steps	1	2	1	2	1	2
# Selected	65	59	42	45	37	38
IS Sharpe Ratio	2.95	2.88	3.47	3.48	3.27	3.29

Non-Linear Model In-Sample Results

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

This table reports selected characteristics obtained from penalized regressions from a total universe of 90 market anomalies. 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.

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

Out-of-Sample Selection

Table: **Linear Model Out-of-Sample Selection**
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	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad. LASSO	SCAD
# Steps	2	1	1	2	1	2	1
Knots	9	9	9	9	9	9	9
# Selected	14	36	19	14	22	18	30
IS Sharpe Ratio	3.84	3.86	3.77	3.97	4.08	4.06	4.34

Out-of-Sample Return Prediction

Table: **Linear Model: Long-Short Portfolio**

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.

Model	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad. LASSO	SCAD
Mean	2.72	2.24	2.9	2.01	2.09	0.85	1.11
Std. Dev.	4.94	4.04	3.99	5.51	5.49	6.26	6.12
Sharpe Ratio	1.91	1.92	2.52	1.27	1.32	0.47	0.63
Skewness	1.1	0.39	0.37	0.17	-0.09	-0.22	-0.24
Kurtosis	3.08	-0.01	0.98	2.1	2.01	3.84	4.46
β	0.51	0.5	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	0.42	0.36	0.4	0.39	0.41	0.33	0.34
TO_R	0.42	0.37	0.4	0.4	0.41	0.33	0.34

Out-of-Sample Selection

Table: **Non-Linear Model Out-of-Sample Selection**
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	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad.e LASSO	SCAD
# Steps	2	1	1	2	1	2	1
Knots	9	9	9	9	9	9	9
# Selected	2	16	1	3	4	3	2
IS Sharpe Ratio	2.65	3.57	2.12	3.12	3.07	2.83	2.36

Out-of-Sample Return Prediction

Table: **Non-Linear Model: Long-Short Portfolio**

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.

Regularization	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad.e LASSO	SCAD
Mean	1.18	1.93	0.67	2.21	2.8	1.47	1.5
Std. Dev.	4.04	3.69	4.2	6.56	6.02	6.5	7.49
Sharpe Ratio	1.01	1.81	0.56	1.17	1.61	0.78	0.69
Skewness	0.34	0.78	0.59	0.26	0.14	0.14	-0.28
Kurtosis	0.93	2.56	0.99	2.11	1.77	0.53	0.91
β	0.75	0.83	0.75	0.95	0.97	0.65	0.68
R^2	0.05	0.03	0.05	0.05	0.05	0.05	0.05
TO	0.52	0.4	0.52	0.44	0.43	0.43	0.44
TO_R	0.52	0.41	0.52	0.45	0.43	0.43	0.44

Out-of-Sample Return Prediction

Table: **Non-Linear Model: Decomposition in Long Portfolio**
This table reports out-of-sample, value-weighted long-only portfolio returns, going long the stocks in the highest predicted return decile for different estimation periods and regularization terms.

Model	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad.e LASSO	SCAD
Mean	1.55	2.22	1.3	2.63	3.11	1.32	1.39
Std. Dev.	6.68	6.41	6.79	8.05	7.96	9.9	10.38
Sharpe Ratio	0.8	1.2	0.66	1.13	1.35	0.46	0.47
Skewness	-0.85	-0.63	-0.78	-0.09	0.07	-0.23	-0.46
Kurtosis	3.1	2.86	2.89	1.11	1.63	0.49	0.44
β	0.77	0.9	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	0.24	0.21	0.24	0.22	0.22	0.22	0.22
TO_R	0.25	0.22	0.25	0.23	0.23	0.23	0.23

Out-of-Sample Return Prediction

Table: **Non-Linear Model: Decomposition in Short Portfolio**
This table reports out-of-sample, value-weighted long-only portfolio returns, going long the stocks in the lowest predicted return decile for different estimation periods and regularization terms.

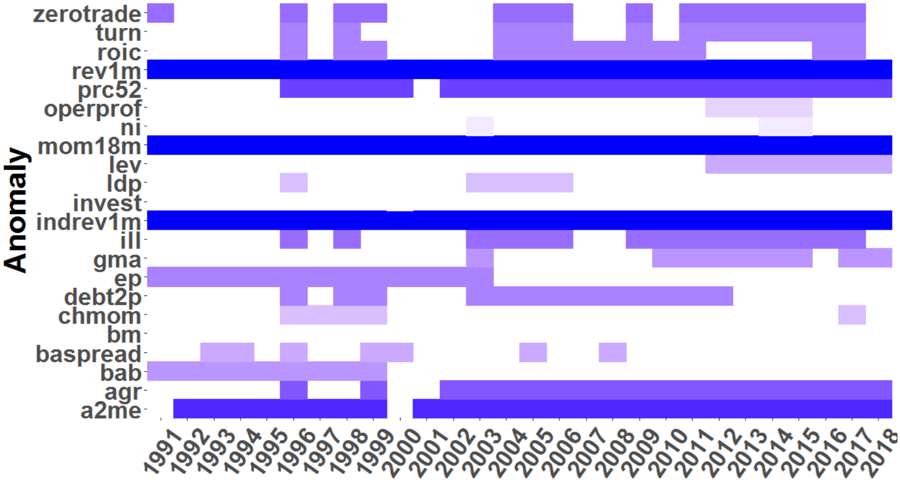
Model	Ad. LASSO	Elastic Net	SCAD	Ad. LASSO	SCAD	Ad.e LASSO	SCAD
Mean	0.37	0.29	0.63	0.42	0.31	-0.14	-0.1
Std. Dev.	5.54	5.13	5.46	3.98	4.46	5.21	4.51
Sharpe Ratio	0.23	0.2	0.4	0.37	0.24	-0.1	-0.08
Skewness	-1.5	-1.15	-1.45	-0.65	-0.3	-0.83	-0.69
Kurtosis	6.88	4.44	6.09	1.75	0.16	1.38	0.78
β	0.87	0.17	-9.2	0.12	0.28	0.17	0.02
R^2	0.03	0.08	0.04	0.06	0.06	0.11	0.1
TO	0.24	0.19	0.24	0.2	0.19	0.18	0.19
TO_R	0.24	0.19	0.24	0.19	0.18	0.18	0.19

Out-of-Sample Rolling Selection

Table: **Non-Linear Model Out-of-Sample Rolling Prediction**
The rolling estimation window is 20 years long and starts 1970.

OoS Period	1991-2018	1991-2018	1991-2018	1991-2018	1991-2018	1991-2014	1991-1999
Firms	ALL	ALL	ALL	ALL	ALL	ALL	ALL
Sample Size	1,010,876	1,010,876	1,010,876	1,010,876	1,010,876	866,105	347,456
Regularization	MCP	SCAD	Ad. 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

Rolling Selection - SCAD



Out-of-Sample Rolling Return Prediction

Table: **Non-Linear Model: Rolling Long-Short Portfolio**
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 a prediction window of 1 year.

Model	MCP	SCAD	Ad. LASSO	MCP	SCAD	SCAD	SCAD
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

Conclusion

- A novel database of 90 market anomalies
- Suggest a technique that manages the curse of dimensionality
- An independent and sparse information subset of firm characteristics exists
- Modelwise I suggest single-stage regression enhanced by a smoothly clipped absolute deviations (SCAD) penalty on group level
- Estimate a multivariate function that links a sparse subset of characteristics to expected returns
- Univariate effects differ from multivariate effects whenever characteristics are collinear
- The predictability of characteristics is time varying
- Non-linearities in the cross-section improve the mapping between characteristics and expected returns

References I

- Cattaneo, Matias D et al. (2018). "Characteristic-sorted portfolios: Estimation and inference". In: *Review of Economics and Statistics*, pp. 1–47.
- Cochrane, John H (2011). "Presidential address: Discount rates". In: *The Journal of finance* 66.4, pp. 1047–1108.
- Fama, Eugene F and Kenneth R French (2018). "Choosing factors". In: *Journal of Financial Economics* 128.2, pp. 234–252.
- Fan, Jianqing and Runze Li (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties". In: *Journal of the American statistical Association* 96.456, pp. 1348–1360.
- Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2019). *Taming the factor zoo: A test of new factors*. Tech. rep. National Bureau of Economic Research.
- Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020). "Dissecting characteristics nonparametrically". In: *Review of Financial Studies*.
- Harvey, Campbell R, Yan Liu, and Heqing Zhu (2016). "... and the cross-section of expected returns". In: *The Review of Financial Studies* 29.1, pp. 5–68.

References II

Huang, Jian, Joel L Horowitz, and Fengrong Wei (2010). "Variable selection in nonparametric additive models". In: *Annals of statistics* 38.4, p. 2282.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh (2019). "Shrinking the cross-section". In: *Journal of Financial Economics*.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288.

Zhang, Cun-Hui et al. (2010). "Nearly unbiased variable selection under minimax concave penalty". In: *The Annals of statistics* 38.2, pp. 894–942.

Zou, Hui (2006). "The adaptive lasso and its oracle properties". In: *Journal of the American statistical association* 101.476, pp. 1418–1429.

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2, pp. 301–320.

Thank you for your attention.