Dissecting Characteristics Nonparametrically

Freyberger, J., Neuhierl, A., Weber, M.

The Review of Financial Studies, 2020.

雷印如 2021/01/02

Motivation and Contribution

- Motivation

1. Two typical methods (Portfolio sorts and linear regressions) are subject to the curse of dimensionality when the number of characteristics is large
2. The association between characteristics and returns appears nonlinear in many empirical settings

- Contribution: The paper provides a new method and framework that allows for nonlinear relations between characteristics and returns in empirical asset pricing to handle the “multidimensional challenge”

Literatures

- Authors usually consider their proposed return predictor in isolation without conditioning on previously discovered return predictors. (Haugen and Baker, 1996; Lewellen, 2015; Green et al., 2017)
- LASSO application in finance (Rapach et al., 2013; Huang and Shi, 2016; Chinco et al., 2019a; Goto and Xu, 2015)
- Method update (Giglio and Xiu, 2016; Gagliardini et al., 2016; Kelly et al., 2017; Kim et al., 2018; Lettau and Pelger, 2018)

Research Question

1. We study which characteristics provide incremental information for the cross-section of expected returns

   ✓ 13 out of 62 characteristics

2. We compare the out-of-sample performance of the nonparametric model with a linear model

   ✓ Sharpe ratio of 2.75 compared to 1.06 for the linear model

3. Third, we study whether the predictive power of characteristics for expected returns varies over time

   ✓ substantial time variation exists

Comparison

- Current Methods and Nonparametric Models

$$m_t(c_1, \ldots, c_S) = E[R_{it} | C_{1,it-1} = c_1, \ldots, C_{S,it-1} = c_S]$$

- Current Methods

1. Portfolio sorts. We typically sort stocks into 10 portfolios and compare mean returns across portfolios. (Suffer from the curse of dimensionality and cannot detect incremental information)
2. Linear panel regressions. (potential nonlinearities and sensitive to outliers and extreme observations)

Comparison

- Nonparametric Models

$$m_t(c_1, \dots, c_S) = E[R_{it} | C_{1,it-1} = c_1, \dots, C_{S,it-1} = c_S]$$

1. A natural solution in the nonparametric regression framework is to assume an additive model

$$m_t(c_1, \dots, c_S) = \sum_{s=1}^{S} m_{ts}(c_s)$$

- The additive model are insensitive to outliers by rank transformation but does not allow for cross-dependencies between characteristics

Adaptive group LASSO

- We partition the support of each characteristic into L intervals similar to portfolio sorts.

$$\tilde{m}_{ts}(\tilde{c}) \approx \sum_{K=1}^{L+2} \beta_{tsk} p_k(\tilde{c})$$

- In the first step, we obtain estimates of the coefficients as

$$\tilde{\boldsymbol{\beta}}_t = \underset{b_{sk}:s=1,\dots,S;k=1,\dots,L+2}{\operatorname{argmin}} \sum_{i=1}^{N} \left(R_{it} - \sum_{s=1}^{S} \sum_{k=1}^{L+2} b_{sk} p_k(\tilde{C}_{s,it-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{L+2} b_{sk}^2 \right)^{\frac{1}{2}}$$

- choose λ_1 in a data-dependent way to minimize Bayesian information criterion (BIC),

Adaptive group LASSO

- A second step introduces characteristic-specific weights in the LASSO group penalty function as a function of first-step estimates to address this problem.

- We first define the following weights:

$$w_{ts} = \begin{cases} \left(\sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2\right)^{-\frac{1}{2}} & \text{if} \quad \sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2 \neq 0 \\ \infty & \text{if} \quad \sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2 = 0 \end{cases}$$

- In the second step of the adaptive group LASSO, we solve

$$\breve{\boldsymbol{\beta}}_t = \underset{b_{sk}:s=1,\ldots,S;k=1,\ldots,L+2}{\arg\min} \sum_{i=1}^{N} \left(R_{it} - \sum_{s=1}^{S}\sum_{k=1}^{L+2} b_{sk} p_k(\tilde{C}_{s,it-1})\right)^2 + \lambda_2 \sum_{s=1}^{S} \left(\boxed{w_{ts}} \sum_{k=1}^{L+2} b_{sk}^2\right)^{\frac{1}{2}}$$

Empirical Analysis

A. Data

1. Stock return data come from the CRSP
2. Balance sheet data are from the Compustat database require that a firm has at least 2 years of Compustat data.
3. We group 62 characteristics into six categories:

   past return based predictors, investment-related, profitability-related, value-related characteristics, intangibles and trading frictions

- The sample period is 1965.7-2014.6 with 1.6 million observations

Empirical Analysis

- 62 characteristics

Past returns:

(1)	r_{2-1}
(2)	r_{6-2}
(3)	r_{12-2}
(4)	r_{12-7}
(5)	r_{36-13}

Investment:

(6)	Investment
(7)	ΔCEQ
(8)	ΔPI2A
(9)	ΔShrout
(10)	IVC
(11)	NOA

Intangibles:

(29)	AOA
(30)	OL
(31)	Tan
(32)	OA

Value:

(33)	A2ME
(34)	BEME
(35)	BEME_{adj}
(36)	C
(37)	C2D
(38)	ΔSO
(39)	Debt2P
(40)	E2P
(41)	Free CF
(42)	LDP
(43)	NOP
(44)	O2P
(45)	Q
(46)	S2P
(47)	Sales_g

Profitability:

(12)	ATO
(13)	CTO
(14)	$\Delta(\Delta$GM-ΔSales)
(15)	EPS
(16)	IPM
(17)	PCM
(18)	PM
(19)	PM_adj
(20)	Prof
(21)	RNA
(22)	ROA
(23)	ROC
(24)	ROE
(25)	ROIC
(26)	S2C
(27)	SAT
(28)	SAT_adj

Trading frictions:

(48)	AT
(49)	Beta
(50)	Beta daily
(51)	DTO
(52)	Idio vol
(53)	LME
(54)	LME_adj
(55)	Lturnover
(56)	Rel_to_high_price
(57)	Ret_max
(58)	Spread
(59)	Std turnover
(60)	Std volume
(61)	SUV
(62)	Total vol

Empirical Analysis

A. Portfolio sorts

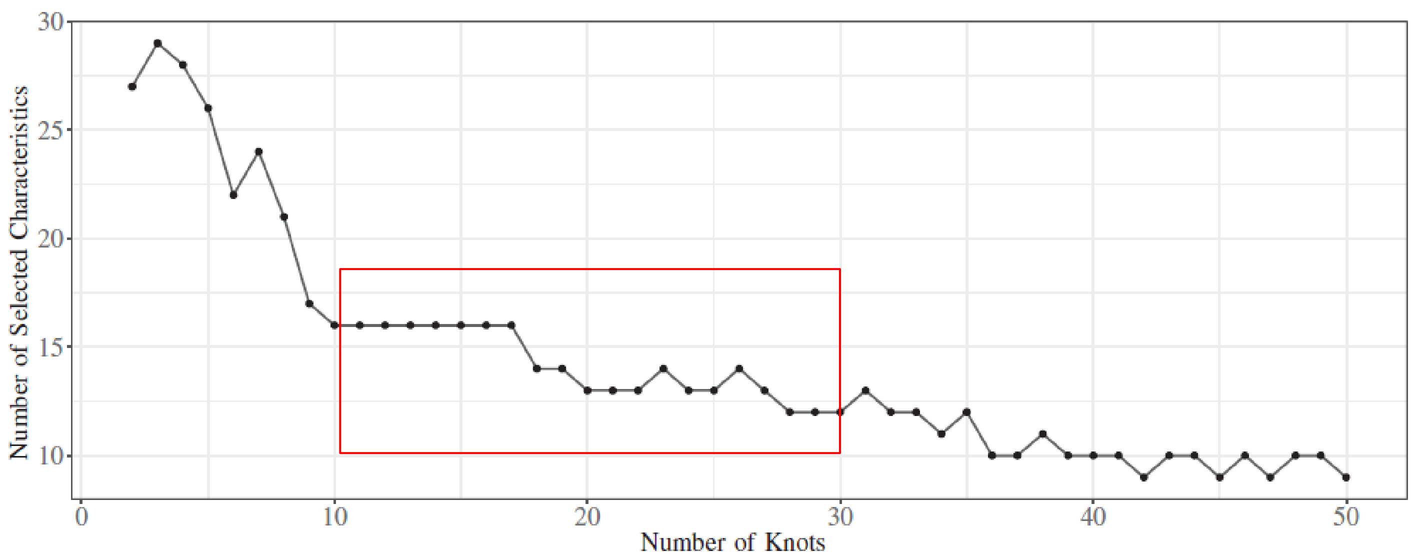
- We use a total of 750 portfolios as test assets.

	Past returns	Trading friction	profitability	investment	value	intangible
Raw return	5	0	7	13	11	0
FF3 Alpha	5	9	5	6	9	4

- Thirty-six characteristics have a t -statistic above 2.
- Correcting for exposure to the Fama-French 3-factor model affects these findings minimally.

Empirical Analysis

- Figure 3 shows how the number of characteristics we select varies with the number of interpolation points.
- We see the number of selected characteristics is stable around 20 interpolation points and varies between 16 when we use only 10 knots and 12 when we use 30 interpolation points



Empirical Analysis

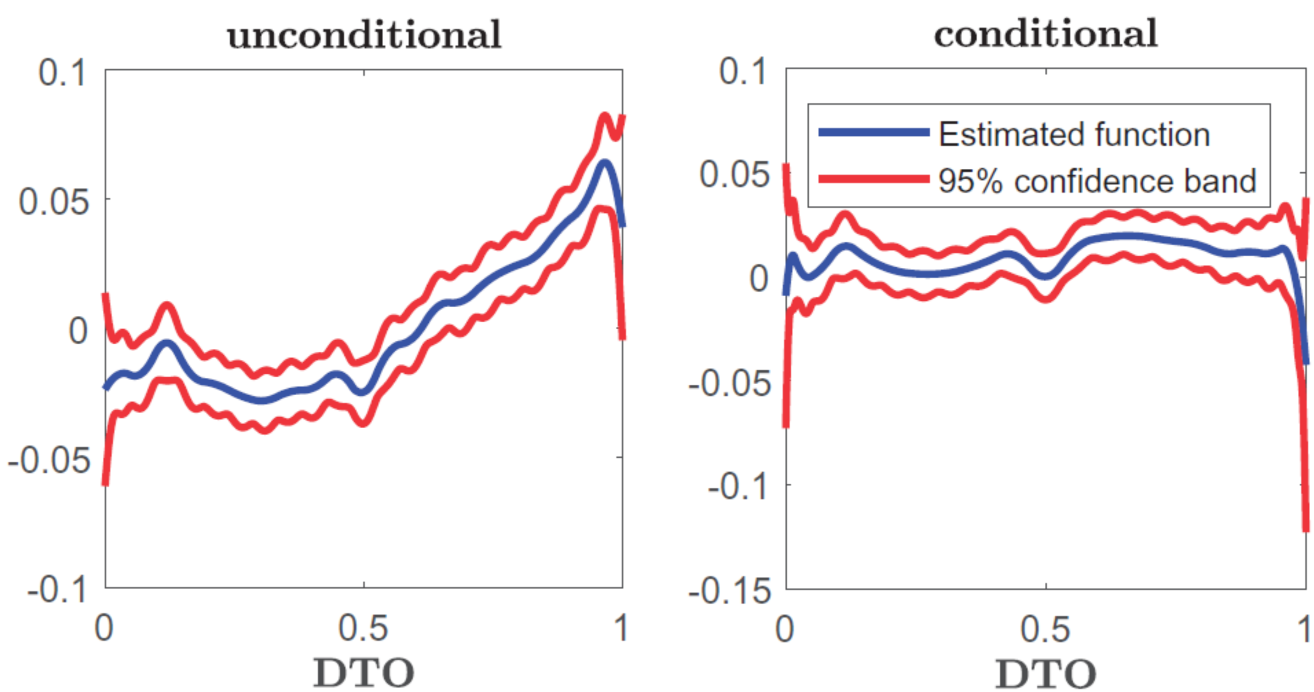
Selected characteristics in nonparametric model

Firms		All	All	All	$Size > q_{10}$	$Size > q_{20}$	$Size > q_{20}$	All
Sample		Full	Full	Full	Full	Full	Full	1965-1990
Knots		20	15	25	15	15	10	15
Sample Size		1,629,155	1,629,155	1,629,155	959,757	763,850	763,850	603,658
# Selected		13	16	13	10	9	11	11
Sharpe Ratio		3.15	3.05	3.16	2.53	2.25	2.37	3.99
Characteristics	# Selected	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BEME	2		BEME					BEME
ΔShrout	8	ΔShrout	ΔShrout	ΔShrout	ΔShrout	ΔShrout	ΔShrout	ΔShrout
ΔSO	7	ΔSO	ΔSO	ΔSO	ΔSO	ΔSO	ΔSO	
Investment	5	Investment	Investment	Investment	Investment			
LDP	1							LDP
LME	5	LME	LME	LME				LME
Lturnover	4	Lturnover	Lturnover	Lturnover				
NOA	2		NOA					
NOP	1							NOP
PM_adj	4	PM_adj	PM_adj	PM_adj				
r_{2-1}	8	r_{2-1}	r_{2-1}	r_{2-1}	r_{2-1}	r_{2-1}	r_{2-1}	r_{2-1}
r_{6-2}	2					r_{6-2}	r_{6-2}	
r_{12-2}	6	r_{12-2}	r_{12-2}	r_{12-2}	r_{12-2}		r_{12-2}	r_{12-2}
r_{12-7}	7	r_{12-7}	r_{12-7}	r_{12-7}	r_{12-7}	r_{12-7}		r_{12-7}
r_{36-13}	3		r_{36-13}				r_{36-13}	
Rel_to_high_price	6	Rel_to_high	Rel_to_high	Rel_to_high			Rel_to_high	Rel_to_high
Ret max	1							Ret max
ROC	7	ROC	ROC	ROC	ROC	ROC	ROC	
S2P	3				S2P	S2P	S2P	
SUV	8	SUV	SUV	SUV	SUV	SUV	SUV	SUV
Total vol	7	Total vol	Total vol	Total vol	Total vol	Total vol	Total vol	

Empirical Analysis

B. Multidimensional challenge

- These characteristics, however, are correlated with other firm characteristics



Empirical Analysis

"size matters, if you control your junk."

B. Specifically, we interact each 61 firm characteristics with firm size for a total of 123 firm characteristics.

Selected Characteristics in nonparametric model: Size interactions

Firms		All	$Size > q_{10}$	$Size > q_{20}$	$Size > q_{20}$
Sample		Full	Full	Full	Full
Knots		20	15	15	10
Sample size		1,629,155	959,757	763,850	763,850
# selected		25	15	9	13
Sharpe ratio		3.33	3.13	2.48	2.72
Characteristics	# selected	(1)	(2)	(3)	(4)
BEME	1	BEME			
ΔShrout	4	ΔShrout	ΔShrout	ΔShrout	ΔShrout
ΔSO	4	ΔSO	ΔSO	ΔSO	ΔSO
DTO	1	DTO			
Investment	1	Investment			
Lturnover	2	Lturnover	Lturnover		
NOA	1		NOA		
PM_adj	1	PM_adj			
r_{2-1}	1	r_{2-1}			
r_{6-2}	2		r_{6-2}		r_{6-2}
r_{12-2}	1		r_{12-2}		
r_{12-7}	4	r_{12-7}	r_{12-7}	r_{12-7}	r_{12-7}
r_{36-13}	3	r_{36-13}	r_{36-13}		r_{36-13}
Rel_to_high_price	2	Rel_to_high_price	Rel_to_high_price		Rel_to_high_price
S2P	3		S2P	S2P	S2P
SUV	4	SUV	SUV	SUV	SUV
Total vol	4	Total vol	Total vol	Total vol	Total vol

Empirical Analysis

“size matters, if you control your junk.”

- We now select a total of 25 of the 123 firm characteristics.
- Once we focus on stocks above the 10%- and 20%-size quantile of NYSE stocks only short-term reversal, momentum, and return over the previous 6 months

Characteristics × Size					
A2ME	1	A2ME			
BEME_adj	1	BEME_adj			
DTO	1	DTO			
EPS	1	EPS			
NOA	1	NOA			
r_{2-1}	4	r_{2-1}	r_{2-1}	r_{2-1}	r_{2-1}
r_{6-2}	4	r_{6-2}	r_{6-2}	r_{6-2}	r_{6-2}
r_{12-2}	4	r_{12-2}	r_{12-2}	r_{12-2}	r_{12-2}
Rel_to_high_price	1	Rel_to_high_price			
Ret max	1	Ret max			
ROC	1				ROC
ROE	1	ROE			
SUV	1	SUV			

Empirical Analysis

- These results are reassuring for previous research, which relied on multivariate regressions to dissect anomalies
- Table 7 shows nonlinearities between characteristics and returns might result in a larger number of selected characteristics in a linear model

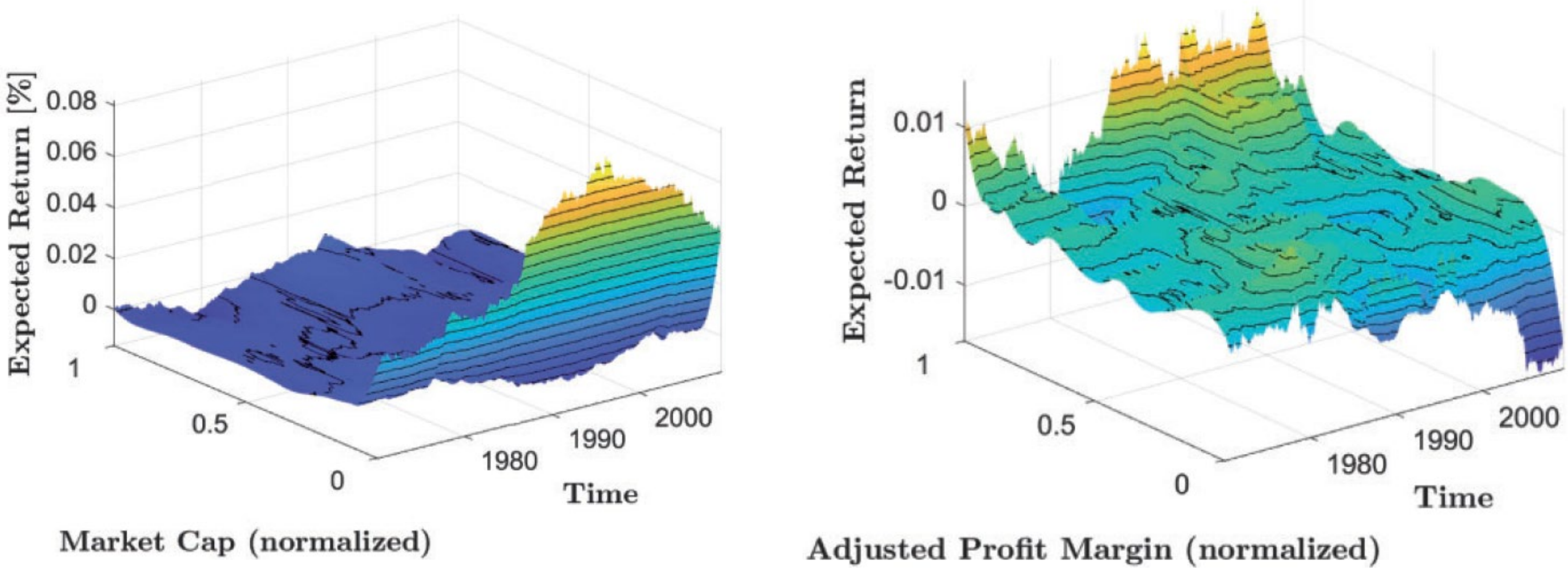
Table 7
Selected characteristics in linear model

Firms	All	All	All
Model	Linear model	Linear model: Rank normalized	Linear model: FDR
Sample	Full	Full	Full
Sample size	1,629,155	1,629,155	1,629,155
# selected	24	35	32
Sharpe ratio	1.47	2.52	1.64

Empirical Analysis

C. Time variation in return predictors

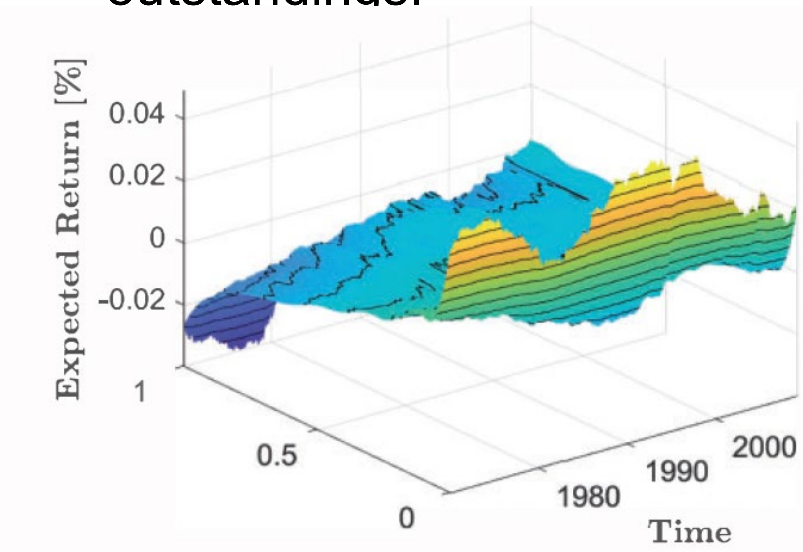
- The conditional mean function is non-constant throughout the sample period for lagged market cap and adjusted profit margin.



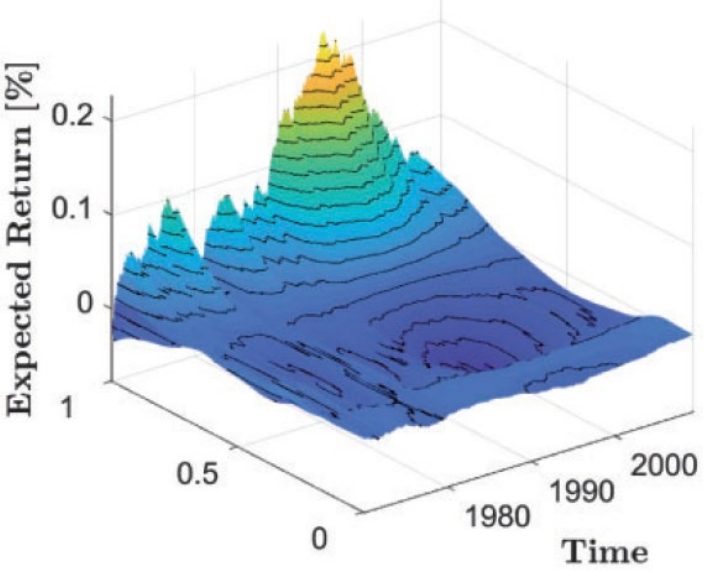
Empirical Analysis

C. Time variation in return predictors

- The conditional mean function is non-constant throughout the sample period for 1-month lagged return and changes in share outstandings.



1-month lagged return (normalized)



Change in Shares Outstanding (normalized)

Empirical Analysis

C. Out-of-sample performance and model comparison

- We find a substantial increase in out-of-sample Sharpe ratios relative to linear model when employing the nonparametric model

Out-of-sample return prediction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Firms	All	All	All	All	All	All	$Size > q_{10}$	$Size > q_{10}$	$Size > q_{20}$	$Size > q_{20}$
oos period	1991-2014	1991-2014	1991-2014	1991-2014	1973-2014	1973-2014	1991-2014	1991-2014	1991-2014	1991-2014
Knots	10		10		10		10		10	
Sample size	1,025,497	1,025,497	1,025,497	1,025,497	1,541,922	1,541,922	959,757	959,757	763,850	763,850
Model	NP	Linear	NP	Linear	NP	Linear	NP	Linear	NP	Linear
# selected	11	30	30	11	12	30	9	24	9	24
Model for selection	NP	Linear	Linear	NP	NP	Linear	NP	Linear	NP	Linear
Sharpe ratio	2.75	1.06	2.61	1.09	3.11	1.41	1.22	0.13	0.89	0.06
A: Long-Short Portfolio										
Mean Return (monthly)	3.82	1.95	3.59	2.09	4.36	2.17	1.55	0.19	1.20	0.09
SD (monthly)	4.81	6.37	4.75	6.63	4.85	5.31	4.40	4.92	4.64	5.22
Sharpe ratio	2.75	1.06	2.61	1.09	3.11	1.41	1.22	0.13	0.89	0.06
Sharpe ratio_*adj*	1.56	0.29	1.50	0.33	1.05	0.05	0.01	−0.70	−0.20	−0.63
Transaction costs	1.71	1.54	1.58	1.36	2.87	2.09	1.54	1.18	1.47	1.04
Skewness	2.77	2.27	1.54	3.14	3.59	2.12	0.54	1.18	0.74	−0.51
Kurtosis	19.56	19.21	7.69	29.84	34.07	22.34	8.45	20.36	10.21	16.92
Turnover1	69.26	55.24	65.04	62.17	73.46	55.47	74.29	55.57	73.77	50.68
Turnover2	33.11	25.72	31.07	29.48	35.51	25.96	36.17	26.32	35.94	23.85
β	0.78	0.38	0.56	0.45	0.88	0.39	0.51	0.10	0.44	0.03
R^2	1.95%	1.37%	1.78%	1.19%	2.78%	1.60%	2.12%	1.64%	2.38%	2.27%

Empirical Analysis

C. Out-of-sample performance and model comparison

- We find a substantial increase in out-of-sample Sharpe ratios relative to linear model when employing the nonparametric model

Out-of-sample predictions (Rolling selection)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Firms	All	All	All	All	$Size > q_{10}$	$Size > q_{10}$	$Size > q_{20}$	$Size > q_{20}$
oos period	1991-2014	1991-2014	1991-2014	1991-2014	1991-2014	1991-2014	1991-2014	1991-2014
Knots	10		10		10		10	
Sample size	1,025,497	1,025,497	1,025,497	1,025,497	959,757	959,757	763,850	763,850
Model	NP	Linear	NP	Linear	NP	Linear	NP	Linear
Average # selected	14.13	26.58	30	11	10.75	27.21	10.75	28.92
Model for selection	NP	Linear	Linear	NP	NP	Linear	NP	Linear
				A: Long-Short Portfolio				
Mean return (monthly)	4.26	2.37	3.72	2.45	1.76	0.63	1.29	0.58
SD (monthly)	5.65	5.75	5.35	6.06	3.83	4.37	4.30	4.37
Sharpe ratio	2.61	1.43	2.41	1.40	1.60	0.50	1.04	0.46
Skewness	3.53	2.61	5.02	2.28	0.30	0.14	−0.50	−0.48
Kurtosis	28.74	21.57	52.79	19.61	7.80	10.39	13.06	10.99
Turnover1	69.70	56.05	71.27	62.24	72.22	49.58	74.48	47.62
Turnover2	33.37	26.13	34.32	29.48	35.09	23.19	36.29	22.25
β	0.87	0.44	0.72	0.52	0.55	0.18	0.45	0.14
R^2	2.16%	1.26%	1.86%	1.18%	2.00%	1.63%	2.37%	2.07%

Simulation

- The aim of this section is to discuss some of the tuning parameters and compare the adaptive group LASSO to alternative model selection methods.
- We consider the following selection methods:
  1. Conventional t -statistic cutoff of 2
  2. t -statistic cutoff of 3 for multiple testing (Harvey et al. 2016)
  3. The FDR p-value adjustment of Green et al. (2017)
  4. Linear single-step LASSO

Simulation

- We also employ different LASSO methods in addition to the adaptive group LASSO
  1. Linear adaptive LASSO
  2. Nonlinear group LASSO
  3. Nonlinear adaptive group LASSO
- Regarding the choice of penalty parameter, we consider:
  1. Akaike information criterion (AIC)
  2. Bayesian information criterion (BIC)
  3. BIC as in Yuan and Lin (2006)
  4. Tenfold cross-validation.

Simulation

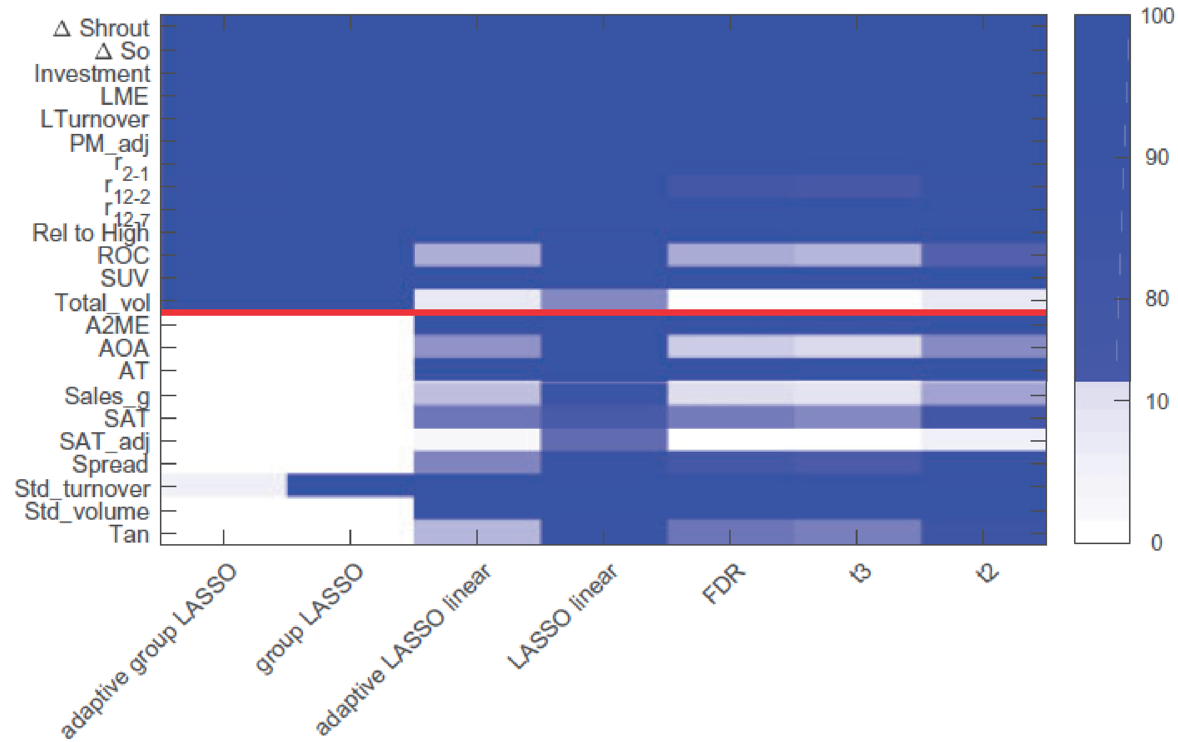
- Our simulation then proceeds in the following steps:
  1. Take the full data set of 62 characteristics
  2. Focus on a sample from 1965 to 2012
  3. Assume the 13 characteristics of are the "true" predictors
  4. Transform all characteristics to be standard normal distributed
  5. Fit a fifth-order polynomial on the true characteristics to estimate $g_s(C_{s,it-1})$ for each characteristic
  6. Generate returns according to: $r_{it} \ = \sum_{s=1}^{13} g_s(C_{s,it-1}) + \varepsilon_{it}$

Simulation

- Our simulation then proceeds in the following steps:
  7. ε_{it} is resampled with replacement from the empirical residuals preserving industry structure
  8. Estimate nonparametric model on rank-transformed data with 20 knots
  9. Estimate linear model on data from step 4
  10. Redo steps 6 to 9500 times

Simulation

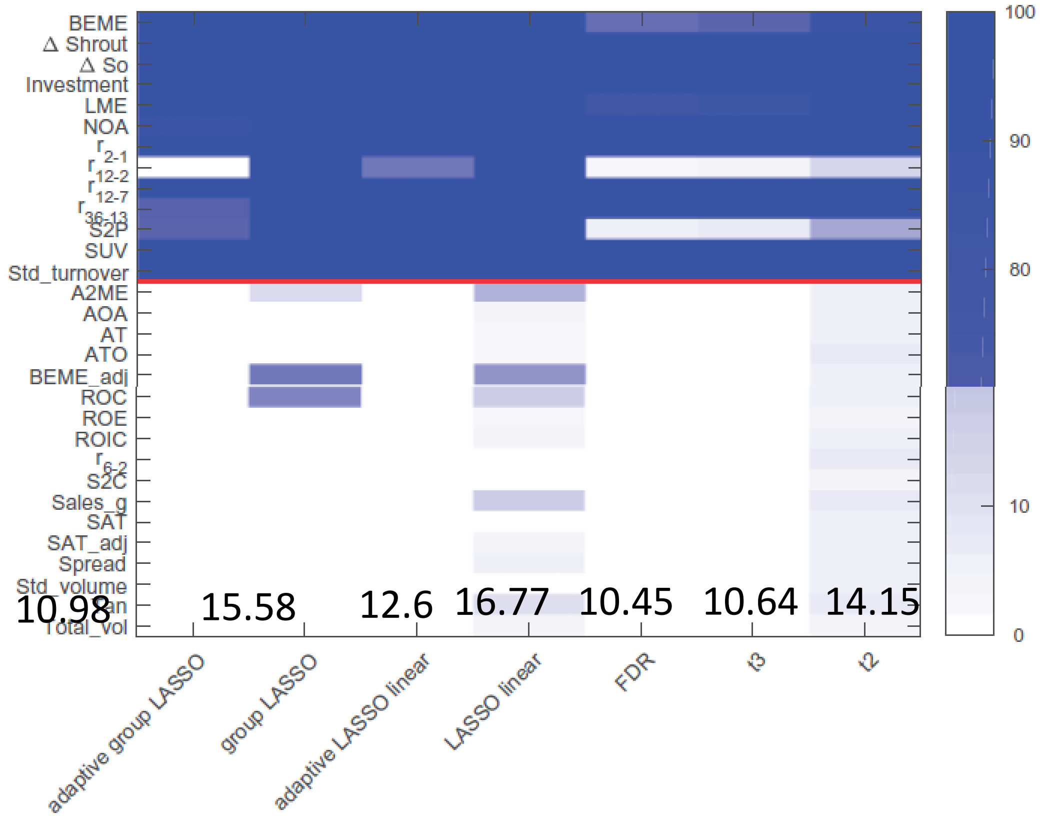
- The average number of selected characteristics



12.99 16.89 29.36 47.72 27.3 26.4 22.75

Simulation

- Instead of a fifth-order polynomial relationship on the true characteristics, we can also assume the true data-generating process is linear and simulate returns under this assumption.



Simulation

- All of the linear models do substantially worse predicting returns out of sample when we follow the true data-generating process.

Out-of-sample predictability in simulation

	(Relative) R^2 (1)	(Relative) RMSPE (2)
	A: Nonlinear Data-Generating Process	
True parametric model	0.0160	0.1204
True nonparametric model	88.64%	0.091%
Adaptive group LASSO	88.61%	0.092%
Group LASSO	87.08%	0.106%
Adaptive LASSO linear	57.42%	0.328%
LASSO linear	57.61%	0.327%
FDR	57.65%	0.326%
t3	57.54%	0.327%
t2	58.03%	0.323%
	B: Linear Data-Generating Process	
True parametric model	0.0088	0.1213
True nonparametric model	94.09%	0.028%
Adaptive group LASSO	93.64%	0.030%
Group LASSO	92.76%	0.035%
Adaptive LASSO linear	99.92%	0.000%
LASSO linear	99.74%	0.001%
FDR	97.23%	0.012%
t3	97.52%	0.011%
t2	99.09%	0.004%

Conclusion

1. Of 62 characteristics, only 9 to 16 provide incremental information depending on the number of interpolation points
2. Substantial time variation is present in the predictive power of characteristics.
3. The nonparametric model selects fewer characteristics than the linear model in-sample and has a Sharpe ratio

Consideration

- Did not consider the joint influence other than size, and the correlation between the factors was not considered sufficiently
- Time varying problems need to be investigated further