Subsampled Factor Models for Asset Pricing: The Rise of Vasa

Gianluca De Nard, Simon Hediger, Markus Leippold
Working Paper

王健
2021-11-07

Content

- Introduction
  - Background & Motivation
  - Question
  - Research content
  - Related researches
  - Contribution
- Methodology
- Monte Carlo Simulation
- Empirical results
- Conclusion

1. Introduction

Background & Motivation

- Recent research suggests that machine learning models dominate traditional models in predicting cross-sectional stock returns.
- Nonlinear algorithms require a vast amount of data for training. A lack of sufficient data may introduce instability, making simpler methods preferable to more complex ones.
- VASA, provides similar and even superior results than some benchmark neural networks and random forest.

1. Introduction

Question

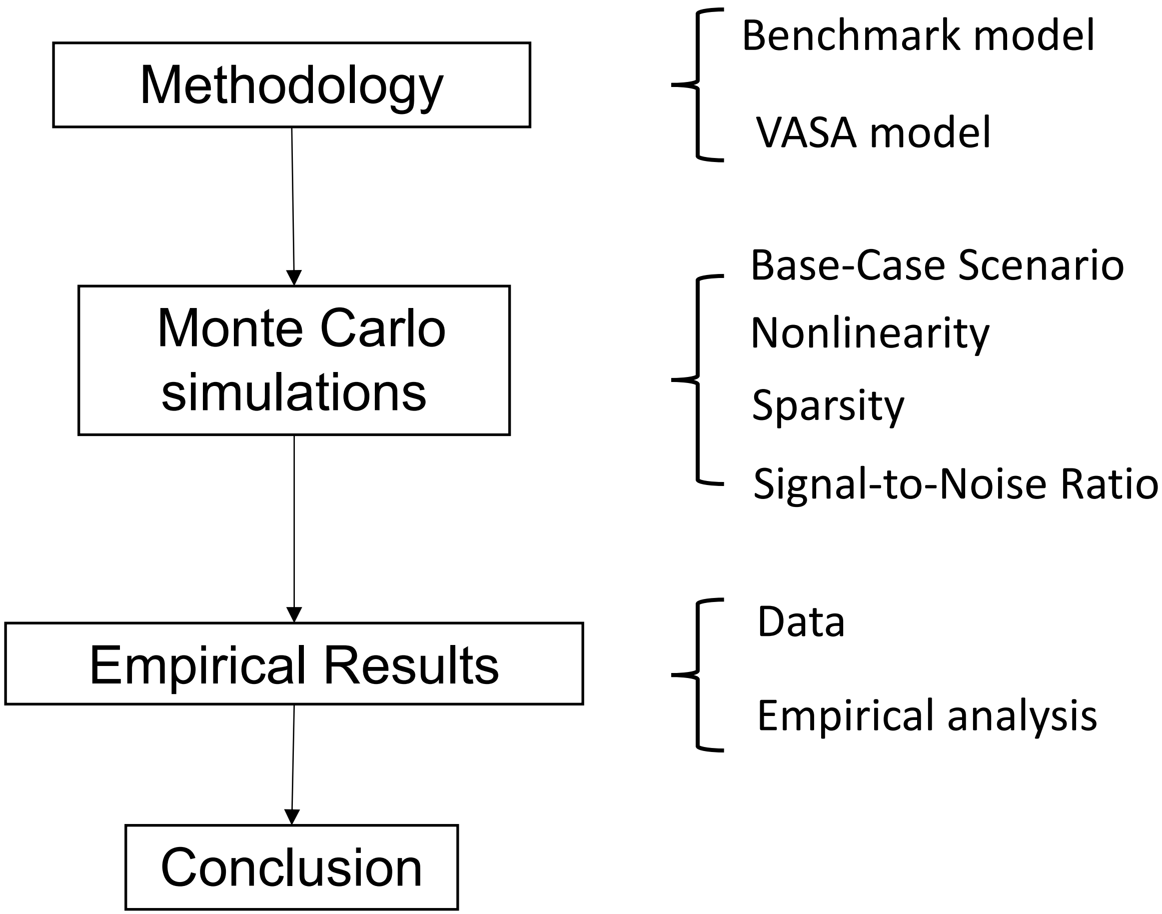
- What is VASA?
  - Variable subsample aggregation.
- Why do we choose VASA?
  - For its simplicity and robustness.
- How does VASA perform?
  - Generally its performance is as good as those of some well-known ML methodologies, but in some places, it has its advantages.

1. Introduction

Research contents

- Section 2 gives a short overview of the benchmark models used in our study and provides the details on our new subsampling framework.
- Section 3 examines and compares the finite-sample behavior of the different methods via Monte Carlo simulations.
- In Section 4, we describe the empirical methodology and present the results of the out-of-sample backtest exercise based on historical stock returns and stock characteristics.
- Section 5 concludes.

1. Introduction



1.Introduction

Related researches

- *Gu et al. (2020)* compare a wide variety of different machine learning methods, ranging from penalized linear models to random forests and neural nets.
- *Jacobsen et al. (2019)* introduce ensemble machine learning for stock return prediction. *Rossi (2018)* follows a similar approach, but uses nonlinear models.
- VASA is close to dropout regressions introduced by *Srivastava et al. (2014a)*, where we set at random some of the elements in the design matrix to zero such that any input dimension is retained.

1.Introduction

Contribution

- We introduce VASA in asset pricing as a simple method for prediction and we show its advantages.
- It becomes evident that the global R^2_{OOS} might not be a sufficient measure to guarantee superior portfolio performance. The distribution of the individual $R^2_{OOS,i}$ plays a crucial role.
- However, VASA is not restricted to linear submodels, and future research should focus on more complex nonlinear base algorithms and aggregation functions.

2. Methodology

1) VASA in General

- We describe an asset (excess) return $r_{i,t+1}$ as:

$$r_{i,t+1} := \mathbb{E}_t[r_{i,t+1}] + \epsilon_{i,t+1} \qquad \mathbb{E}_t[r_{i,t+1}] := g(\boldsymbol{z}_{i,t})$$

- where $\mathrm{z}_{i,t}$ (a vector of P predictors) is a mixture of asset specific factors (characteristics) and macroeconomic variables.
- VASA trains B submodels via a common base algorithm.

$$\hat{r}_{i,t+1} = \hat{g}^{\mathrm{VASA}}(\boldsymbol{z}_{i,t}) := f\left[\hat{g}^{BASE}(\tilde{\boldsymbol{z}}_{i,t,1}), \ldots, \hat{g}^{BASE}(\tilde{\boldsymbol{z}}_{i,t,B})\right]$$

- Instead of taking a bootstrap sample in the observational space, we suggest taking a subset in the predictor space (a vector of K_b predictors).

2. Methodology

1) VASA in General

Algorithm 1 VASA $\leftarrow$ function$(x, \boldsymbol{z}, r, B, K_b, f)$

Require: $x \in \mathbb{R}^{1\times P}$, $\boldsymbol{z} \in \mathbb{R}^{N\times T\times P}$, $r \in \mathbb{R}^{N\times T}$, $B \in \mathbb{N}$, $K_b \in \{1, \ldots, P\}$ $\forall$ $b = 1, \ldots, B$ and an aggregation function $f : \mathbb{R}^B \to \mathbb{R}$

for b in $1 : B$ **do**
  $V_b \leftarrow \text{randsample}(p, K_b, \text{replacement=FALSE})$ // $V_b \sim \text{HGeom}(P, K_b)$
  $\tilde{\boldsymbol{z}}_b \leftarrow \tilde{\boldsymbol{z}}[:, :, V_b]$
  Estimate the b-th submodel using $\tilde{\boldsymbol{z}}_b$ and the response r
return $f\left[\hat{g}^{BASE}(x[V_1]), \ldots, \hat{g}^{BASE}(x[V_1])\right]$

$$\hat{g}^{\text{VASA}}(\boldsymbol{z}_{i,t}) := \sum_{b=1}^{B} \omega_b \hat{g}^{BASE}(\tilde{\boldsymbol{z}}_{i,t,b}) \qquad \tilde{\boldsymbol{z}}_{t,b} := \tilde{\boldsymbol{z}}_t \Lambda(V_b)'$$

- $V_b \in \{0,1\}^{\mathrm{P}}$, Hence, for $V_b = \{v_{b,1}, \ldots, v_{b,P}\}'$ such that $v_b 1 = K_b$, where 1 denotes a vector of ones of dimension $P \times 1$.

2. Methodology

2) VASA with Linear Submodels

$$\hat{g}^{\mathrm{VASA}}(\boldsymbol{z}_{i,t}) := \sum_{b=1}^{B} \omega_b \hat{g}^{\mathrm{OLS}}(\tilde{\boldsymbol{z}}_{i,t,b}) = \sum_{b=1}^{B} \omega_b (\hat{\alpha}_b + \tilde{\boldsymbol{z}}'_{i,t,b} \hat{\boldsymbol{\beta}}_b)$$

- $K \equiv K_b$. Hence, each $z_{i,t,b}$ contains K (pseudo) randomly chosen variables (without replacement) from the P predictors
- If we take equally distributed subsampling probabilities, $q_p = 1/P$ and weights $\omega_b = 1/B$

$$q_p = R_{i,p}^2 / \sum_{j=1}^{P} R_{i,j}^2$$

- where $R_{i,p}^2$ is the in-sample R^2 from regressing r_i on the p_{th} factor.

2. Methodology

3) Sample Splitting and Performance Evaluation

- A training sample, comprising of the first 30% observations.
- A validation sample, retaining the successive 20% of observations (the number of subsampled factor models B and their dimension K for VASA)
- A testing sample containing the next (last) twelve months of data.

$$R^2_{OOS,i} := 1 - \frac{\sum_{t\in\mathcal{T}}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{t\in\mathcal{T}} r^2_{i,t+1}}$$

3. Monte Carlo Simulation

Data Generating Process

$$\bar{c}_{i,p,t} := \rho_p \bar{c}_{i,p,t-1} + e_{i,p,t} \qquad c_{i,p,t} := \frac{2}{N+1}\mathrm{CSRank}(\bar{c}_{i,p,t}) - 1$$

- $\bar{c}_{i,p,t} = 0$, $\rho_p \sim U(0.9,1)$, $e_{i,p,t} \sim N(0,1-\rho_p^2)$, We then use this auxiliary variable to generate the cross section and time series of all the characteristics.

$$x_t := \rho x_{t-1} + u_t \qquad \boldsymbol{z}_{i,t} := (1, x_t)' \otimes \boldsymbol{c}_{i,t}$$

- we simulate a time series x_t representing the (macro-) economic environment, and $x_0 = 0$, $\rho = 0.95$, $u_t \sim N(0,1-\rho^2)$

$$r_{i,t+1} := g^*(\boldsymbol{z}_{i,t}) + \epsilon_{i,t+1} \qquad \epsilon_{i,t+1} := \boldsymbol{v}_{t+1}' \boldsymbol{\beta}_{i,t}^* + \varepsilon_{i,t+1}$$

- we define a latent K^*-factor model to generate (excess) returns, and $v_{t+1} \sim N(0,0.005^2 * 1_3)$, $\varepsilon_{i,t+1} \sim t_5(0,0.005^2)$

3. Monte Carlo Simulation

Data Generating Process

- we suggest to introduce sparsity by simulating a three-factor model with $\beta_{i,t}^{*} = (c_{i,1,t}, c_{i,2,t}, c_{i,3,t})'$ and we use two cases for the functional form $g^{*}(z_{i,t})$

$$g^*(\boldsymbol{z}_{i,t}) := (c_{i,1,t}, c_{i,2,t}, c_{i,3,t} \times x_t)\theta_0 = (\boldsymbol{\beta}_{i,t}^* \circ (1, 1, x_t)')'\theta_0$$

- For $\beta_{i,t}^{*} = (0.02, 0.02, 0.02)'$

$$g^*(\boldsymbol{z}_{i,t}) := (c_{i,1,t}^2, c_{i,1,t} \times c_{i,2,t}, \mathrm{sign}(c_{i,3,t} \times x_t))\theta_0$$

- For $\beta_{i,t}^{*} = (0.04, 0.03, 0.012)'$
- Finally, we set $N = 100$, $\mathrm{T} = 480$, $\mathrm{P_c} = 100$

3. Monte Carlo Simulation

Base-case Setting: Comparison of Machine Learning Methods								
	"Oracle"	Average	OLS	LASSO	RIDGE	VASA	RF	NNET
MED	4.61	0.02	4.33	**5.13**	4.40	4.73	4.12	3.04
AV	4.46	0.00	3.93	**4.67**	4.01	4.51	3.81	3.33
SD	4.62	**0.11**	4.82	4.45	4.60	4.63	3.68	4.21
P10	−1.48	**−0.15**	−2.60	−1.33	−2.40	−1.63	−1.30	−2.23
R^2_{OOS}	4.53	0.00	4.01	**4.74**	4.08	4.58	3.87	3.43

Nonlinear Setting: Comparison of Machine Learning Methods								
	"Oracle"	Average	OLS	LASSO	RIDGE	VASA	RF	NNET
MED	9.62	3.98	4.41	5.13	4.91	5.39	**8.79**	4.63
AV	11.09	3.30	4.62	5.19	4.80	5.37	**9.75**	5.20
SD	6.74	4.75	5.13	4.54	4.82	**4.50**	6.46	5.20
P10	3.48	−3.14	−1.46	−1.23	−1.10	−0.61	**2.56**	−0.46
R^2_{OOS}	11.66	3.75	5.08	5.63	5.25	5.80	**10.28**	5.72

3. Monte Carlo Simulation

Sparsity: Comparison of Machine Learning Methods								
	"Oracle"	Average	OLS	LASSO	RIDGE	VASA	RF	NNET
	$K^* = 1$ Driving Covariate							
MED	4.41	0.00	1.97	3.99	1.95	**4.03**	3.18	2.48
AV	4.26	0.00	2.03	**3.98**	2.16	3.97	3.28	2.53
SD	3.04	**0.04**	4.02	2.77	3.48	2.89	2.29	2.62
P10	0.41	−0.05	−3.23	**0.43**	−2.46	0.35	0.22	−0.50
R^2_{OOS}	4.25	0.00	2.05	**3.96**	2.16	3.95	3.29	2.47
	$K^* = 3$ Driving Covariates							
MED	4.61	0.02	4.33	**5.13**	4.40	4.73	4.12	3.04
AV	4.46	0.00	3.93	**4.67**	4.01	4.51	3.81	3.33
SD	4.62	**0.11**	4.82	4.45	4.60	4.63	3.68	4.21
P10	−1.48	**−0.15**	−2.60	−1.33	−2.40	−1.63	−1.30	−2.23
R^2_{OOS}	4.53	0.00	4.01	**4.74**	4.08	4.58	3.87	3.43
	$K^* = 10$ Driving Covariates							
MED	8.22	0.00	6.36	8.01	7.11	**7.60**	5.08	6.39
AV	8.54	0.00	7.25	**8.49**	7.92	8.34	5.73	7.26
SD	6.64	**0.06**	6.66	6.24	6.13	6.49	4.40	5.91
P10	0.55	−0.08	−1.59	**1.44**	0.63	0.78	0.19	0.19
R^2_{OOS}	9.00	0.00	7.73	**8.89**	8.33	8.78	5.99	7.67

3. Monte Carlo Simulation

Signal-to-Noise Ratio: Comparison of Machine Learning Methods

	"Oracle"	Average	OLS	LASSO	RIDGE	VASA	RF	NNET
				$\theta_0 = 0.02$				
MED	4.61	0.02	4.33	**5.13**	4.40	4.73	4.12	3.04
AV	4.46	0.00	3.93	**4.67**	4.01	4.51	3.81	3.33
SD	4.62	**0.11**	4.82	4.45	4.60	4.63	3.68	4.21
P10	−1.48	**−0.15**	−2.60	−1.33	−2.40	−1.63	−1.30	−2.23
R^2_{OOS}	4.53	0.00	4.01	**4.74**	4.08	4.58	3.87	3.43
				$\theta_0 = 0.05$				
MED	25.15	0.00	25.48	**25.65**	25.34	25.13	23.91	25.42
AV	26.06	0.00	25.66	**26.23**	25.74	26.03	24.44	25.05
SD	12.61	**0.19**	12.69	12.52	12.55	12.60	12.35	12.76
P10	10.11	−0.24	9.88	**11.25**	10.22	10.44	8.44	10.15
R^2_{OOS}	27.95	0.00	27.56	**28.11**	27.63	27.92	26.20	26.97
				$\theta_0 = 0.1$				
MED	62.15	0.00	61.42	62.04	61.26	**62.15**	57.55	61.08
AV	63.00	0.00	62.42	62.89	62.20	**63.03**	58.95	61.98
SD	9.09	**0.32**	9.29	9.07	9.14	9.07	9.29	9.24
P10	51.51	−0.47	50.67	**51.53**	50.44	51.53	46.95	49.89
R^2_{OOS}	65.03	0.00	64.47	64.91	64.21	**65.06**	60.92	64.02

4. Monte Carlo Simulation

1) Stock Data

- Resource: CRSP and Compustat
- Time interval: January 1977 and ends in December 2016 (monthly)
- Risk-free rate: Treasury-bill rate

2) Characteristic Data

- 94 stock-level predictive characteristics used by Gu et al. (2020), industry dummies and eight macroeconomic predictors in Welch and Goyal (2008)

3) Sample pool

- we restrict our sample to stocks that have a complete return and stock-level characteristics history for the entire 40 years. In doing so, the number of stocks in our sample reduces to 501.

4. Empirical results

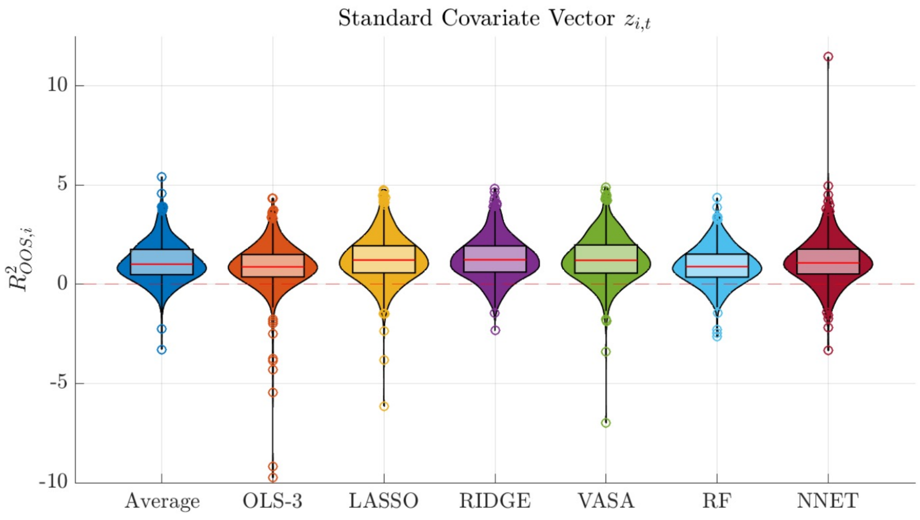
The global R^2_{OOS} and $R^2_{OOS,i}$

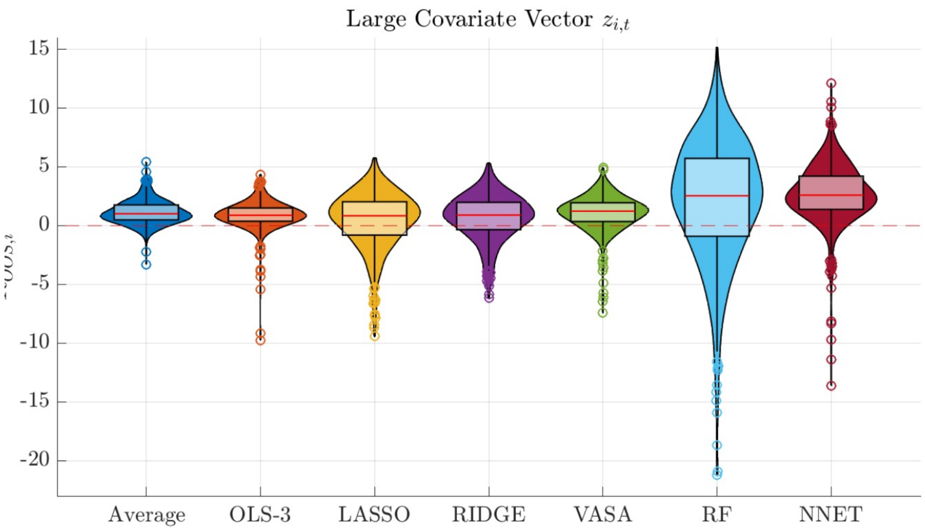
Monthly Out-Of-Sample Stock-level Prediction Performance (in %)

	Average	OLS	OLS-3	LASSO	RIDGE	VASA	RF	NNET
				$\boldsymbol{z}_{i,t}^{\text{standard}}$				
MED	1.02	1.03	0.88	1.22	1.24	**1.26**	0.89	1.09
AV	1.12	1.03	0.89	1.28	1.29	**1.33**	0.95	1.16
SD	1.01	1.49	1.24	1.22	1.07	1.22	**0.95**	1.20
P10	−0.04	−0.53	−0.17	−0.09	**−0.00**	−0.04	−0.14	−0.23
R^2_{OOS}	0.81	0.87	0.77	0.95	0.95	**0.97**	0.77	0.82
				$\boldsymbol{z}_{i,t}^{\text{large}}$				
MED	1.02	−39.64	0.88	0.84	0.91	1.23	2.55	**2.59**
AV	1.12	−513	0.89	0.46	0.70	1.07	2.00	**2.66**
SD	**1.01**	2080	1.24	2.39	1.93	1.46	5.24	2.59
P10	−0.04	−607	−0.17	−2.54	−1.97	−0.56	−4.53	**0.36**
R^2_{OOS}	0.81	−194	0.77	0.93	1.07	1.10	**2.53**	2.30

4. Empirical results

The global R^2_{OOS} and $R^2_{OOS,i}$





4. Empirical results

Long-Short Portfolio

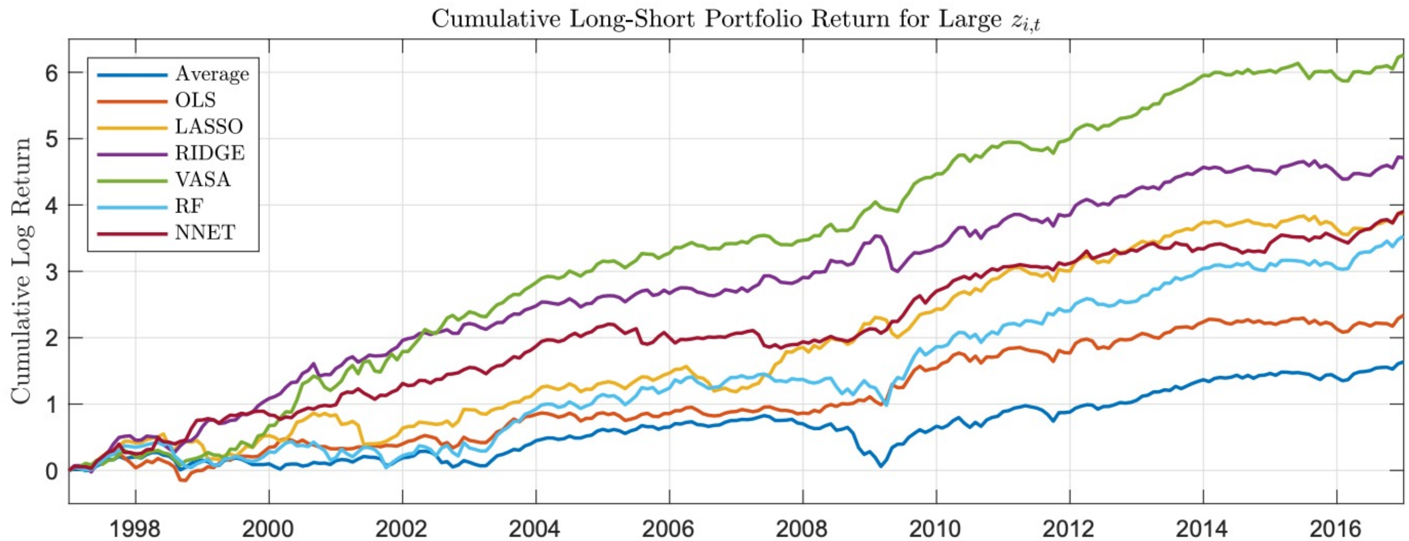
Long-Short Portfolio Analysis (Equal-Weighted)								
	Average	OLS	OLS-3	LASSO	RIDGE	VASA	RF	NNET
	$z_{i,t}^{\text{standard}}$							
Value	5.129	41.271	14.816	75.957	56.204	**91.181**	29.535	52.434
AV	9.504	20.025	16.354	23.892	23.273	**24.790**	19.975	22.843
SD	**16.104**	16.158	24.037	20.320	24.492	20.260	24.339	24.368
SR	0.590	**1.239**	0.680	1.176	0.950	1.224	0.821	0.937
Skew	−0.330	0.145	**0.554**	−0.234	0.087	−0.024	0.250	0.329
Kurt	2.433	1.975	2.973	0.745	3.338	**0.070**	1.500	0.984
	$z_{i,t}^{\text{large}}$							
Value	5.129	10.377	14.816	47.378	111.362	**522.675**	34.154	49.571
AV	9.504	13.346	16.354	21.861	26.191	**34.271**	20.580	21.424
SD	**16.104**	18.097	24.037	21.966	21.903	23.156	24.049	18.848
SR	0.590	0.737	0.680	0.995	1.196	**1.480**	0.856	1.137
Skew	−0.330	**0.567**	0.554	−0.340	−0.353	0.416	0.492	0.002
Kurt	2.433	2.410	2.973	1.862	1.945	**0.481**	2.079	1.298

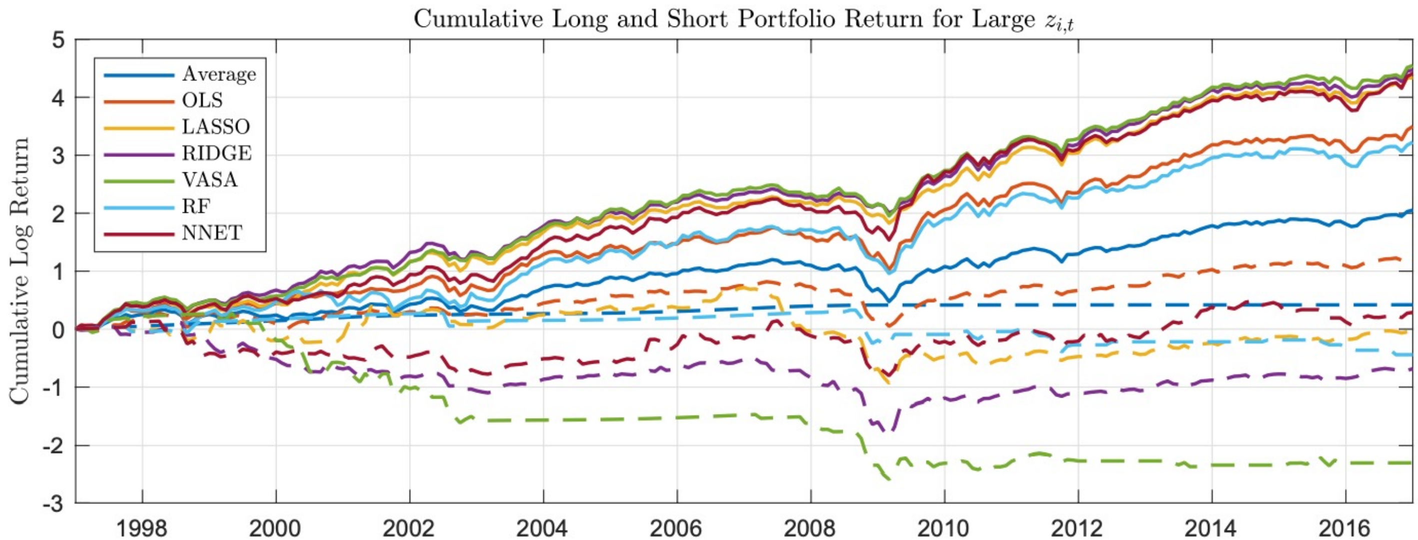
4. Empirical results

Long-Short Portfolio

Portfolio Analysis (Value-Weighted)								
	Average	OLS	OLS-3	LASSO	RIDGE	VASA	RF	NNET
				$z_{i,t}^{\text{standard}}$				
Value	6.520	10.841	54.507	18.030	97.577	68.906	**146.833**	45.408
AV	10.350	13.519	23.852	17.052	26.283	24.149	**29.597**	23.151
SD	**13.692**	17.678	27.646	22.503	25.947	24.052	30.176	28.553
SR	0.756	0.766	0.863	0.758	**1.013**	1.004	0.981	0.811
Skew	−0.448	0.075	0.422	−0.028	**1.302**	0.380	0.552	0.494
Kurt	1.095	1.892	1.602	**0.512**	10.335	1.221	4.153	1.123
				$z_{i,t}^{\text{large}}$				
Value	6.520	8.320	54.507	46.927	104.961	**416.884**	19.612	94.298
AV	10.35	12.851	23.852	22.462	26.414	**33.577**	18.761	26.030
SD	**13.692**	21.382	27.646	24.768	24.146	25.081	28.155	25.379
SR	0.756	0.601	0.863	0.907	1.094	**1.339**	0.666	1.026
Skew	−0.448	0.654	0.422	−0.249	−0.226	0.276	**0.903**	0.659
Kurt	1.095	2.314	1.602	1.312	3.784	**0.652**	6.139	3.355

4. Empirical results





4. Empirical results

Decile Portfolios: average return and the Sharpe ratio

Equal-Weighted Decile Portfolios								
	Average	OLS	OLS-3	LASSO	RIDGE	VASA	RF	NNET
AV								
High	11.61	20.51	15.33	24.11	24.77	**25.36**	20.35	25.01
D9	11.61	15.66	11.69	16.21	17.46	18.23	15.25	16.32
D8	11.61	14.91	10.72	13.43	16.02	14.50	14.00	13.48
D7	11.61	12.16	11.82	13.99	13.42	12.81	12.30	12.35
D6	11.61	10.57	11.93	11.12	10.11	11.76	11.04	10.46
D5	11.61	8.43	13.19	9.63	11.33	10.31	9.94	9.92
D4	11.61	8.84	12.77	8.32	7.56	8.60	10.64	9.06
D3	11.61	10.10	10.05	6.25	6.99	7.39	6.76	8.38
D2	11.61	7.89	10.77	7.61	5.48	5.27	9.40	7.37
Low	11.61	7.14	8.11	5.62	3.42	**2.69**	6.52	3.70
SR								
High	0.72	0.84	0.86	1.12	1.16	**1.21**	0.84	0.99
D9	0.72	0.76	0.76	0.85	0.94	1.03	0.79	0.84
D8	0.72	0.79	0.73	0.77	0.90	0.83	0.85	0.76
D7	0.72	0.70	0.75	0.81	0.79	0.77	0.77	0.72
D6	0.72	0.61	0.73	0.67	0.63	0.71	0.70	0.70
D5	0.72	0.50	0.70	0.60	0.71	0.64	0.64	0.67
D4	0.72	0.57	0.72	0.53	0.46	0.53	0.68	0.61
D3	0.72	0.68	0.55	0.40	0.42	0.47	0.42	0.60
D2	0.72	0.61	0.59	0.46	0.34	0.31	0.58	0.48
Low	0.72	0.55	0.39	0.31	0.20	**0.16**	0.38	0.22

4. Empirical results

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\widehat{\Sigma}_{t+1}\boldsymbol{w}$$

$$\mathbb{E}_t[\boldsymbol{r}_{t+1}]'\boldsymbol{w} = \mathbb{E}_t[\boldsymbol{r}_{t+1}]'\boldsymbol{w}^{\mathrm{EW}}_{t+1}$$

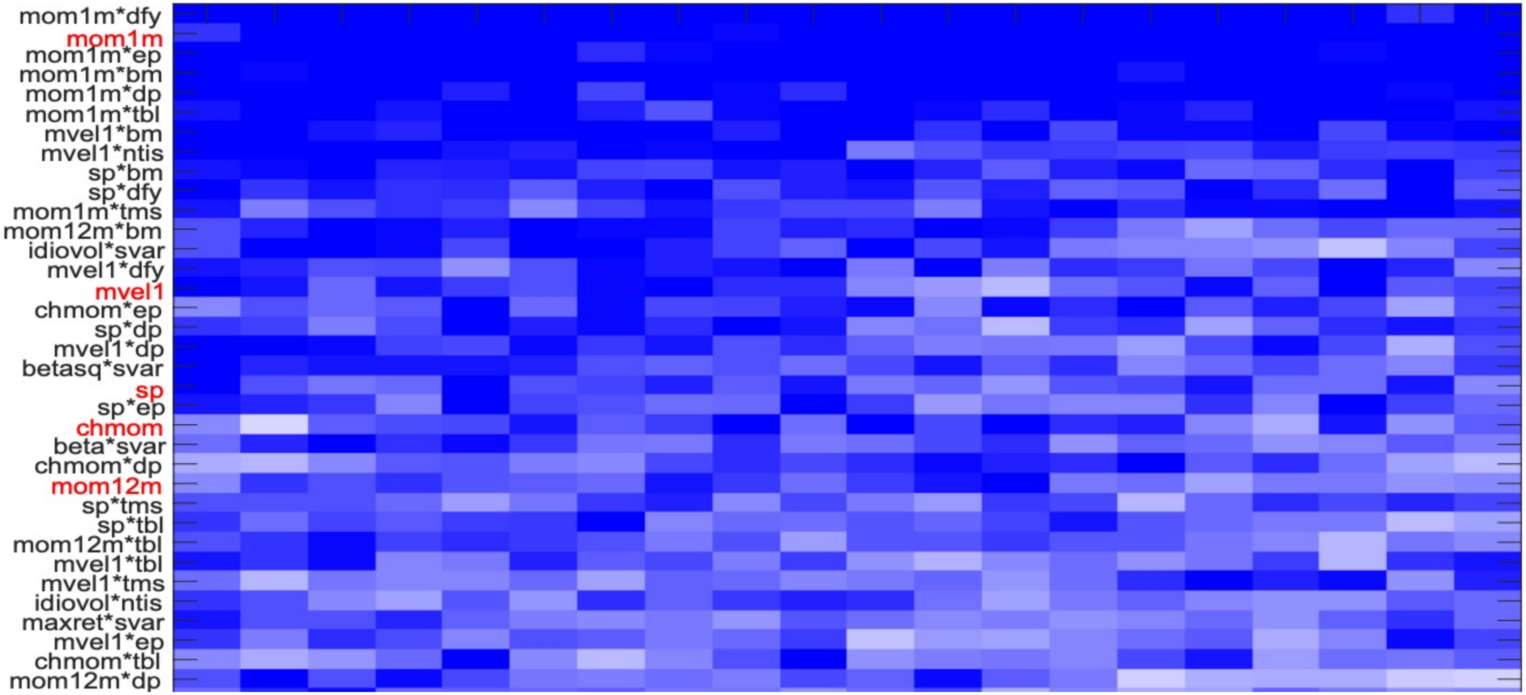
$$\sum_{w_i<0} |w_i| = \sum_{w_i>0} |w_i| = 1 \ .$$

Portfolio Analysis (Efficient Sorting)

	Average	OLS	OLS-3	LASSO	RIDGE	VASA	RF	NNET
				$\boldsymbol{z}_{i,t}^{\text{standard}}$				
Value	7.802	22.889	2.009	41.230	36.100	**60.319**	28.174	36.352
AV	11.608	16.122	4.310	19.386	18.719	**21.426**	18.08	18.796
SD	16.055	**8.660**	12.778	11.547	11.547	12.495	16.395	11.967
SR	0.723	**1.862**	0.337	1.679	1.617	1.715	1.103	1.571
Skew	−0.365	0.461	−0.253	0.746	0.427	0.785	**1.342**	0.600
Kurt	2.458	2.546	**1.307**	3.879	3.294	3.282	4.004	3.116
				$\boldsymbol{z}_{i,t}^{\text{large}}$				
Value	7.802	2.759	2.009	31.133	47.593	**86.393**	5.161	50.233
AV	11.608	5.460	4.310	17.972	20.028	**23.215**	9.110	20.552
SD	16.055	**8.719**	12.778	11.617	10.69	12.166	13.355	12.974
SR	0.723	0.626	0.337	1.547	1.874	**1.908**	0.682	1.584
Skew	−0.365	0.525	−0.254	0.427	0.178	0.768	0.507	**0.941**
Kurt	2.458	2.084	1.307	**0.482**	0.503	1.681	0.839	3.436

4. Empirical results

VASA’s Factor Choice



4. Conclusion

- We demonstrate that more sophisticated algorithms like random forest and neural networks do not necessarily beat simpler linear models.
- We confirm that high variability in $R^2_{OOS,i}$'s is detrimental for long-short portfolios sorted according to predicted returns, due to the higher risk of misclassifying the stocks in the wrong deciles.
- As for characteristic selection, momentum turns out to be the most influential factor and most of the submodels in VASA are driven by characteristics with interaction terms.