Asset Pricing and Machine Learning

Princeton Lectures in Finance
Lecture 1

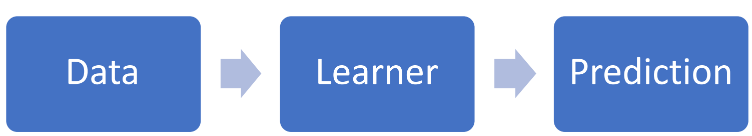
Stefan Nagel[1]

[1]University of Chicago, NBER, CEPR, and CESIfo

May 1, 2019

Machine Learning

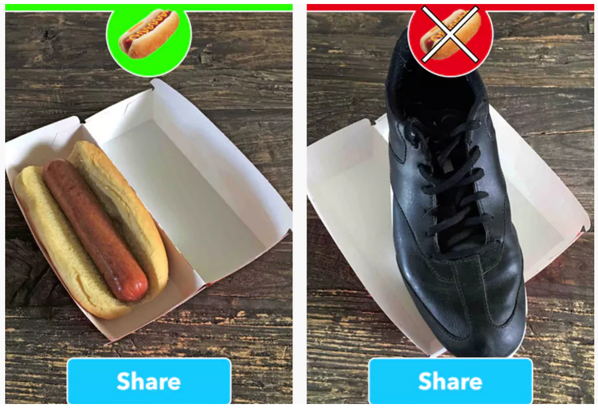
- ML ≈ Automated detection of patterns in data



- ML emphasizes
  - prediction rather than statistical inference
  - algorithms for practical application
  - distribution-free approaches

Example: Image recognition

- Food image classification task: Classify $y_i \in \{$ Hot dog, Not hot dog $\}$
- Features: $\boldsymbol{X}_i$ = matrix of image pixels
- Example



Machine Learning

- **Supervised learning**: Approximate $y_i = f(\boldsymbol{x}_i)$.
  - Regression
  - Classification: Support Vector Machines, Trees, Random Forests,
  - Nonlinearity: Neural Networks, Deep Learning
- **Unsupervised learning**: Summarize $\{\boldsymbol{x}_i\}$, dimension reduction
  - Clustering
  - Principal components analysis
- Recent surge due to
  - data availability ('Big Data')
  - computational power
  - development of algorithms for computationally efficient implementation

Asset Pricing and Machine Learning

- Applicability of ML in asset pricing?
- Prediction central to ML and also essential to asset pricing (AP)
  - Forecasting returns
  - Forecasting cash-flows
  - Forecasting default
  - Forecasting risk exposures
- Machine learning (ML) offers potentially useful toolbox for prediction

High-dimensional prediction problems in asset pricing

- Fundamental asset pricing equation for asset with excess return R and SDF M:

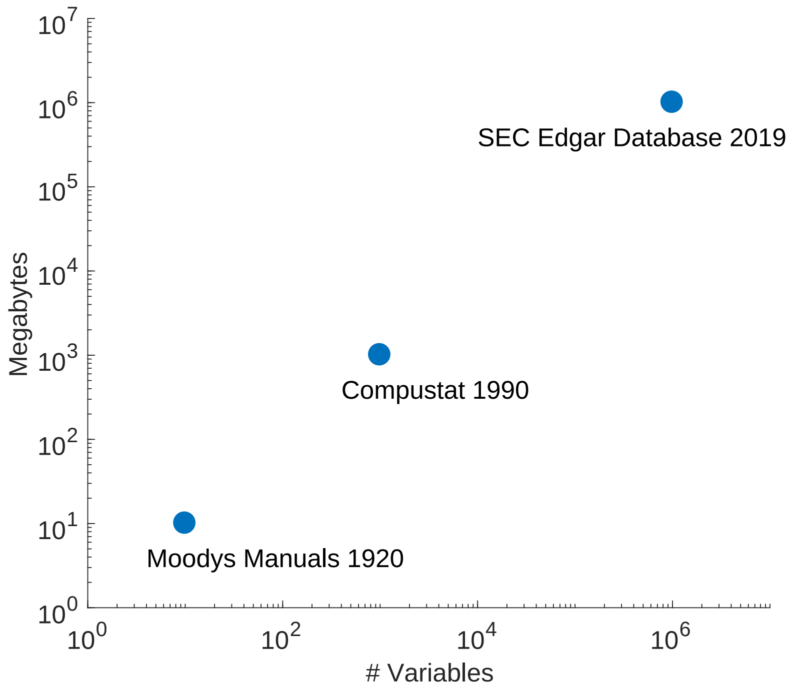
$$\mathbb{E}[R_{t+1}M_{t+1}|\boldsymbol{x}_t] = 0$$

- Empirical implementation involves function approximation

$$\boldsymbol{x}_t \mapsto \text{(Co-)moments of} \quad R_{t+1}, M_{t+1}$$

  i.e., a supervised learning problem
  - plus dimension reduction in joint distribution of (R_{t+1}, M_{t+1}) may be useful: unsupervised learning
- Pre-ML literature: $\boldsymbol{x}_t$ typically low-dimensional, but little real-world justification for this

Big Data in asset pricing: Example of corporate financial reports data



Application of ML methods in asset pricing research

Two ways in which ML can help us think about

$$\mathbb{E}[R_{t+1}M_{t+1}|\boldsymbol{x}_t] = 0$$

1. Econometrician as outside observer of financial market data: Approximate $\mathbb{E}[R_{t+1}M_{t+1}|\boldsymbol{x}_t]$
    - understand determinants of expected returns
    - construct an SDF that summarizes investment opportunities
    - link result to economic models
2. Modeling of investors inside the financial market: Investors as using ML to form $\mathbb{E}[R_{t+1}M_{t+1}|\boldsymbol{x}_t]$

Outline

1. More on ML techniques relevant for asset pricing
2. ML used by econometrician outside the market: SDF extraction in high-dimensional setting
3. ML used by investors inside the market: Rethinking market efficiency in the age of Big Data (tomorrow)
4. Conclusion: Agenda for further research (tomorrow)

Prior knowledge in ML

- On one hand, ML emphasizes distribution-free, flexible, data-driven approach, but there are limits
- **No free lunch theorem** of ML (Wolpert 1996) $\approx$ A universal learner does not exist
  - Without prior knowledge, we can't have confidence that generalization beyond training data set will work
- Therefore, some prior knowledge required, e.g,
  - parametric distributional assumptions
  - restricting set of entertained models, hypotheses
  - temporal stability assumptions

Prior knowledge $\Rightarrow$ Regularization

- Consider supervised learning problem: find $y_i = f(\boldsymbol{x}_i)$ where $i = 1, 2, ..., N$ and $\boldsymbol{x}_i$ has dimension $J \times 1$.
- When $\boldsymbol{x}_i$ high-dimensional (e.g, $J > N$), standard methods (e.g., OLS) would horribly overfit in-sample $\Rightarrow$ bad out-of-sample prediction performance
- Regularization: Penalize estimation results that are regarded as implausible based on prior knowledge
  - Example: If big magnitudes of regression coefficients a priori unlikely, penalize big coefficient estimates

A framework that encompasses much of ML

- Many ML methods can be derived as penalized estimators

$$\hat{\boldsymbol{\theta}} = \arg\min_{\theta} \sum_i L\left\{y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})\right\} + \lambda R(\boldsymbol{\theta})$$

  for loss function $L(.)$ and penalty function $R(.)$.
  - $R(\theta) = ||\boldsymbol{\theta}||_1$: Lasso
  - $R(\theta) = ||\boldsymbol{\theta}||_2^2$: Ridge regression
  - $R(\theta) = \alpha||\boldsymbol{\theta}||_1 + (1-\alpha)||\boldsymbol{\theta}||_2^2$: Elastic net
- Penalty forces regularization: Well-behaved estimates, useful for prediction, even if $J > N$
  - Penalty is crucial for prediction performance

Bayesian foundations

- Penalty function and penalty weight are a way to express prior knowledge about θ
- Examples
  - Laplace prior + normal likelihood $\Rightarrow$ Posterior mode = Lasso
  - Normal prior + normal likelihood $\Rightarrow$ Posterior mean = Ridge regression
- In asset pricing (AP) applications, economic reasoning my yield informative prior specification and penalty choice (rather than 'off-the-shelf' ML method)

ML in asset pricing

- How do typical AP applications compare to typical ML applications?
- Consider return prediction problem: Cross-section of stocks $i = 1, ..., N$, with $J \times 1$ characteristics vector (observable predictors) $\boldsymbol{x}_{i,t}$.

$$\mathbb{E}[r_{i,t+1}|\boldsymbol{x}_{i,t}] = f(\boldsymbol{x}_{i,t}, \theta)$$

- Observations $\boldsymbol{r}_t = (r_{1,t}, ..., r_{N,t})'$ for $t = 1, ..., T$.
- Could ML techniques be useful?

Example: Predicting returns with past returns

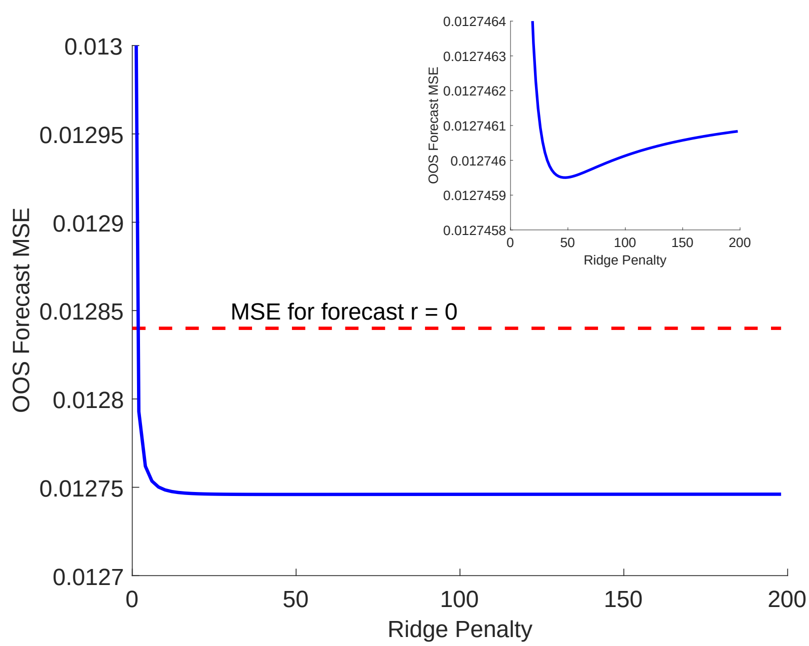
- Predict monthly return of individual U.S. stocks where $\boldsymbol{x}_{i,t}$ contains
  - 120 lags of monthly returns, $r_{i,t}, r_{i,t-1}, \ldots, r_{i,t-120}$
  - 120 lags of monthly squared returns, $r_{i,t}^2, r_{i,t-1}^2, \ldots, r_{i,t-120}^2$

  where all returns are cross-sectionally demeaned each month (i.e., cross-sectional focus) and $\boldsymbol{x}_{i,t}$ is standardized.
- Estimate during 1980 - 2000. Evaluate forecasts OOS during 2001 - 2019.
- Ridge regression

$$\hat{\theta} = \arg\min_{\theta} \sum_i (r_{i,t+1} - \boldsymbol{\theta}' \boldsymbol{x}_{i,t})^2 + \lambda \boldsymbol{\theta}' \boldsymbol{\theta}$$

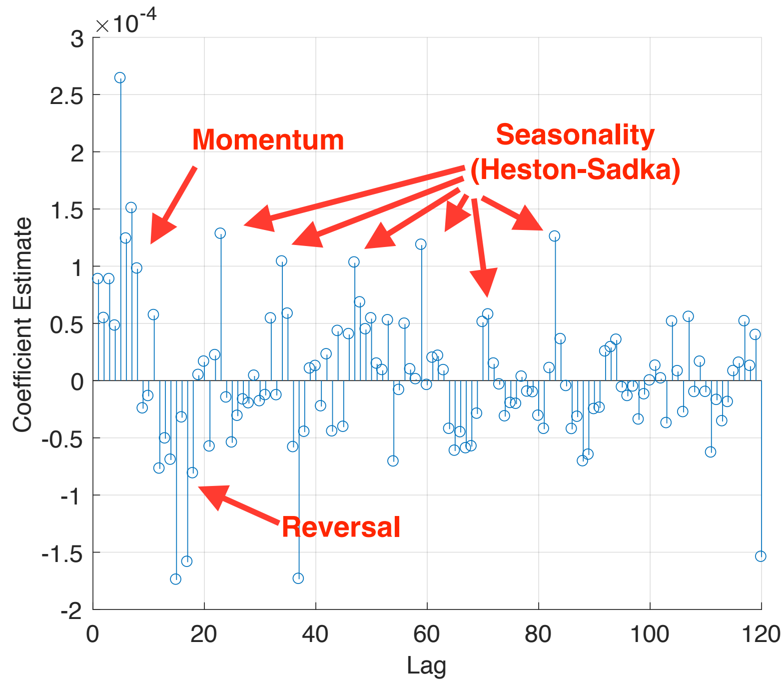
where $\lambda = 0$ implements OLS.

Example: Forecast MSE in predicting month-ahead return



⇒ OLS ($\lambda = 0$) produces terrible forecasts

Example: Ridge Regression coefficients estimates for past returns



ML in asset pricing: Issues

	Typical ML application	Asset pricing
Signal-to-noise	**Outcome observable** e.g. { hotdog, not hotdog }	**Very noisy observation** of outcome e.g. {high $\mathbb{E}[r]$, low $\mathbb{E}[r]$}
Big Data dimensions	**N and J big**	**J big, N not so much**
Sparsity	**Often sparse** e.g., some regions of image irrelevant	**Unclear**
Lucas critique	**Often not an issue** e.g. hotdogs don't change shape in response to image classification	**Investors learn from data and adapt**

Adapting ML for asset pricing

- Low signal-to-noise ratio, high $J/N \Rightarrow$ Imposition of prior knowledge more important than in typical ML application
  - Functional forms and distributional assumptions
  - Penalty function
- Views about what is implausible should be guided by economic reasoning $\Rightarrow$ e.g., express absence of near-arbitrage in prior
- Not obvious that methods striving for sparsity (e.g., Lasso) are necessarily appropriate

Outline

1. More on ML techniques relevant for asset pricing
2. ML used by econometrician outside the market: SDF extraction in high-dimensional setting
   - Based on joint work with Serhiy Kozak and Shrihari Santosh: Kozak, S., S. Nagel, and S. Santosh, 2019, Shrinking the Cross-Section, *Journal of Financial Economics*, forthcoming.
3. ML used by investors inside the market: Rethinking market efficiency in the age of Big Data
4. Conclusion: Agenda for further research

Factor models and the cross-section of expected stock returns

- Multi-decade quest: Describe cross-section of N excess stock returns, $\mathbb{E}[\boldsymbol{r}]$, with small number ($K$) of factor excess returns $\boldsymbol{f}$ in SDF

$$M_t = 1 - \boldsymbol{b}'(\boldsymbol{f}_t - \mathbb{E}\,\boldsymbol{f}_t) \qquad \text{with} \quad \mathbb{E}[M_t \boldsymbol{r}_t] = 0$$

  where factors are returns on portfolios constructed based on firm characteristics (size, B/M, momentum, ...).
- Popular factor models are sparse in characteristics, e.g: Fama-French 3-, 4-, 5-factor models
- But can a characteristics-sparse representation of the SDF be adequate?
  - Taking into account all anomalies that have been discovered
  - Plus potentially hundreds or thousands of additional stock characteristics, including interactions
  - High-dimensional problem!

Seemingly simple back then...

THE JOURNAL OF FINANCE • VOL. LI, NO. 1 • MARCH 1996

Multifactor Explanations of Asset Pricing Anomalies

EUGENE F. FAMA and KENNETH R. FRENCH*

ABSTRACT

Previous work shows that average returns on common stocks are related to firm characteristics like size, earnings/price, cash flow/price, book-to-market equity, past sales growth, long-term past return, and short-term past return. Because these

Complex today...

Table 1
Factor classification

Risk type		Description	Examples
Common (113)	**Financial** (46)	Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, among others	Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns
	Macro (40)	Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, among others	Breeden (1979): consumption growth; Cochrane (1991): investment returns
	Microstructure (11)	Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, among others	Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume
	Behavioral (3)	Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing	Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing
	Accounting (8)	Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, among others	Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow
	Other (5)	Proxy for aggregate movements that do not fall into the above categories, including momentum, investors' beliefs, among others	Carhart (1997): return momentum; Ozoguz (2009): investors' beliefs
Characteristics (202)	**Financial** (61)	Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, among others	Ang et al. (2006): idiosyncratic volatility; Bali, Cakici, and Whitelaw (2011): extreme stock returns
	Microstructure (28)	Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, among others	Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs
	Behavioral (3)	Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, among others	Diether, Malloy, and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage
	Accounting (87)	Proxy for firm-level accounting variables, including PE ratio, debt-to-equity ratio, among others	Basu (1977): PE ratio; Bhandari (1988): debt-to-equity ratio
	Other (24)	Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, among others	Cooper, Gulen, and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles

The numbers in parentheses represent the number of factors identified. See Table 6 and http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.

Harvey, Liu, and Zhu (2015)

The high-dimensionality challenge

- Why not just throw in hundreds of factor portfolio returns into vector $\boldsymbol{f}$ and estimate price-of-risk coefficients $\boldsymbol{b}$ in

$$M_t = 1 - \boldsymbol{b}'\boldsymbol{f}_t \quad ?$$

- Naive approach:
  - With population moments, using $\mathbb{E}[M_t \boldsymbol{r}_t] = 0$, we can solve for

$$\boldsymbol{b} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_f$$

  - Estimate with sample equivalent

$$\hat{\boldsymbol{b}} = \widehat{\boldsymbol{\Sigma}}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t,$$

- Naive approach would result in extreme overfitting of noise $\Rightarrow$ Terrible out-of-sample performance.

Regularization through economically motivated priors

- Consider normally distributed factor returns

$$\boldsymbol{f}_t | \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where we assume that $\boldsymbol{\Sigma}$ is known.

- To regularize, we need to effectively tell the estimator which aspects of the data to regard as implausible and likely spurious
- Prior: Economic restriction that links first ($\boldsymbol{\mu}$) and second moments ($\boldsymbol{\Sigma}$)

$$\boldsymbol{\mu} \sim \mathcal{N}\left(0, \kappa \boldsymbol{\Sigma}^2\right)$$

Regularization through economically motivated priors

- To understand intuition: Prior implies that Sharpe ratios of principal component (PC) portfolios are distributed

$$\frac{\mu_i}{\sigma_i} \sim \mathcal{N}(0, \kappa \sigma^2)$$

- Reflects two economic restrictions
  1. absence of near-arbitrage opportunities (extremely high Sharpe Ratios)
  2. high Sharpe Ratios more likely to originate from high- than from low-variance principal components

  $\iff$ bounded MVE portfolio weights
- These restrictions hold in
  - rational asset pricing models with "macro" factors
  - behavioral model with sentiment-driven investors co-existing with arbitrageurs (Kozak, Nagel, Santosh 2018)

Posterior

- With sample of mean factor returns $\bar{\boldsymbol{f}}$, we get the posterior mean

$$\hat{\boldsymbol{b}} = (\boldsymbol{\Sigma} + \gamma \boldsymbol{I}_K)^{-1} \bar{\boldsymbol{f}}$$

where $\gamma = \frac{1}{\kappa T}$

- Equivalently, we can rewrite this as

$$\hat{\boldsymbol{b}} = \left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma} + \gamma \boldsymbol{I}_K\right)^{-1} \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{f}}$$

i.e., a GLS version of cross-sectional ridge regression of $\bar{\boldsymbol{f}}$ on columns of $\boldsymbol{\Sigma}$

- Economically motivated prior based on properties of Sharpe Ratios lead us to estimator that differs from off-the-shelf ML estimators

Penalized regression representation

- Equivalently, we can rewrite this as minimizing the Hansen-Jagannathan distance subject to an L_2-norm penalty:

$$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \left\{ \left(\bar{\boldsymbol{f}} - \boldsymbol{\Sigma}\boldsymbol{b}\right)' \boldsymbol{\Sigma}^{-1} \left(\bar{\boldsymbol{f}} - \boldsymbol{\Sigma}\boldsymbol{b}\right) + \gamma \boldsymbol{b}'\boldsymbol{b} \right\}$$

- Penalty parameter γ to be chosen by cross-validation or based on prior views about the maximum squared Sharpe Ratio

Allowing for sparsity: Characteristics-sparse SDF

- A large collection of stock characteristics may contain some that are "useless" or redundant
- Calls for prior that reflects possibility that many elements of the $\boldsymbol{b}$ vector may be zero: Sparsity in characteristics
  - Laplace prior $\Rightarrow L_1$ norm penalty
- Two-penalty Specification

$$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \left(\bar{\boldsymbol{f}} - \boldsymbol{\Sigma b}\right)' \boldsymbol{\Sigma}^{-1} \left(\bar{\boldsymbol{f}} - \boldsymbol{\Sigma b}\right) + \gamma_1 \boldsymbol{b}'\boldsymbol{b} + \gamma_2 \sum_{i=1}^{H} |\boldsymbol{b}_i|$$

  - Similar to elastic net, but with different loss function

Alternative: PC-sparse SDF

- Not clear that sparsity necessarily in terms of characteristics
- Recall our prior: High Sharpe Ratios more likely for high-variance principal components (PCs)
  - i.e., the PCs $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ from eigendecomposition of covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{QDQ}' \qquad \text{with} \quad \boldsymbol{Q} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N)$$

- Plus empirical fact : Typical sets of asset returns have covariance matrix dominated by a few PCs with high variance

$\Rightarrow$ A PC-sparse SDF with first few (K) PC-factor portfolio returns

$$M_t = 1 - b_1 \boldsymbol{q}_1' \boldsymbol{r}_t - b_2 \boldsymbol{q}_2' \boldsymbol{r}_t - \ldots - b_k \boldsymbol{q}_K' \boldsymbol{r}_t$$

should capture most risk risk premia.

Empirical application

- Dimension reduction prior to analysis: aggregate into characteristics-based stock portfolios
  - Conditional moments should relate to characteristics of a firm, not its "name"
- Stock characteristics portfolios
  - 50 anomaly characteristics portfolios
  - 1,375 portfolios based on powers and pairwise interactions of 50 anomaly characteristics
- Two sets of analyses for each set of characteristics
  - Characteristics-weighted portfolio returns
  - Principal component portfolio returns
- Questions
  - Can we find an SDF sparse in characteristics?
  - Can we find an SDF sparse in PCs?

List of 50 Characteristics

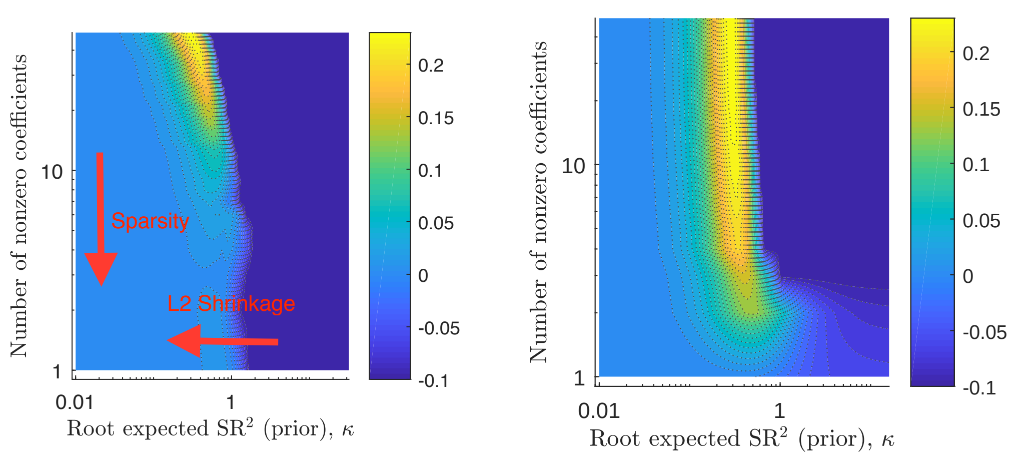
Size	Investment	Long-term Reversals
Value	Inv/Cap	Value (M)
Profitability	Investment Growth	Net Issuance (M)
Value-Profitability	Sales Growth	SUE
F-score	Leverage	Return on Book Equity
Debt Issuance	Return on Assets (A)	Return on Market Equity
Share Repurchases	Return on Equity (A)	Return on Assets
Net Issuance (A)	Sales/Price	Short-term Reversals
Accruals	Growth in LTNOA	Idiosyncratic Volatility
Asset Growth	Momentum (6m)	Beta Arbitrage
Asset Turnover	Industry Momentum	Seasonality
Gross Margins	Value-Momentum	Industry Rel. Reversals
D/P	Value-Prof-Momentum	Industry. Rel. Rev. (LV)
E/P	Short Interest	Industry Momentum-Rev
CF/P	Momentum (12m)	Composite Issuance
Net Operating Assets	Momentum-Reversals	Stock Price

Empirical implementation

- Daily returns on characteristics-based factor portfolios 1974 - 2017; all in excess of the market index return
- Drop very small stocks (market caps $< 0.01\%$ of agg. market cap.)
- 3-fold cross-validation
  - Sample split in 3 blocks
  - 2 blocks used for estimation of $\boldsymbol{b}$
  - Remaining block used for out-of-sample evaluation: R^2 in explaining average returns of test assets with fitted value $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{b}}\boldsymbol{\Sigma}$.
  - Reshuffle blocks and repeat

$\Rightarrow$ we report average R^2 across 3 validation blocks as CV-R^2

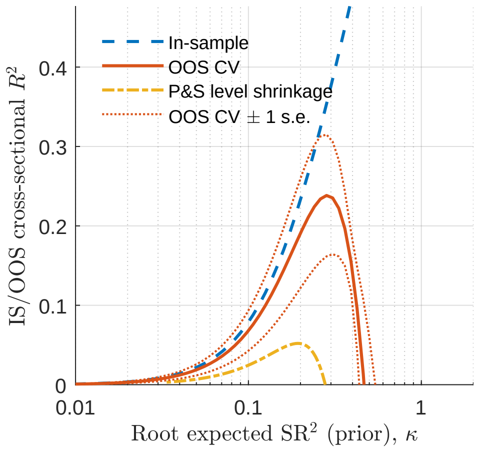
50 anomalies: CV R^2 from dual-penalty specification



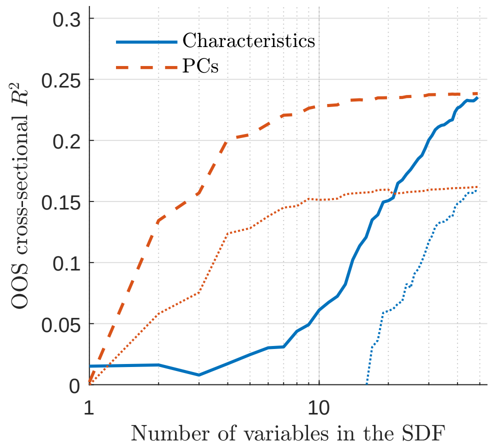
(a) Raw 50 anomaly portfolios

(b) PCs of 50 anomaly portfolios

50 anomalies: L^2 shrinkage and Sparsity



(c) L^2 shrinkage

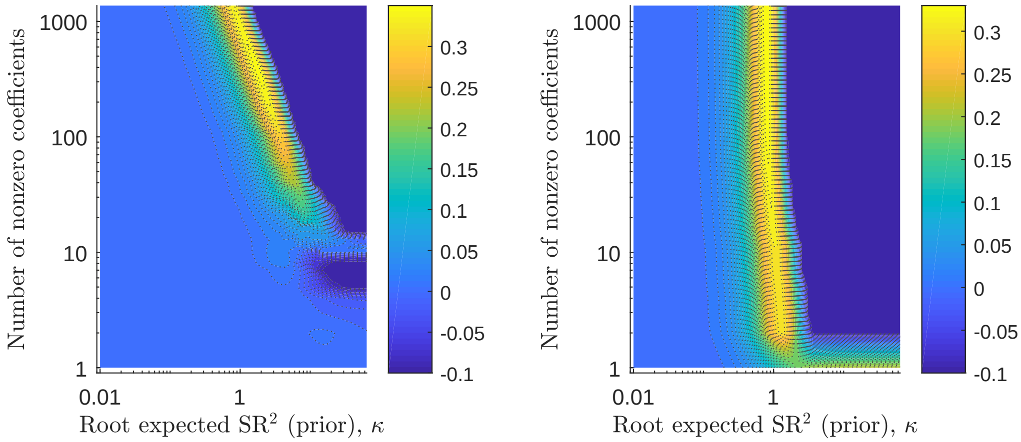


(d) Sparsity

50 anomalies: Interpretation

- Summary of key results
  1. Shrinkage is extremely important
  2. Very little redundancy in original characteristics space: Characteristics-sparse SDF not achievable
  3. But PC-sparse SDF based on a few (high-variance) PCs prices well
- Result (2) could be partly a consequence of looking at a set of data-mined anomalies
- Could there be more characteristics-sparsity if we include some unexplored factors, or factors that are not known to be associated with return premia?
  - Kozak, Nagel, and Santosh (2019) look at 80 WRDS financial ratios that were (partly) not used previously in c.s. empirical AP
  - Interaction portfolios: except for a few like value and size, value and momentum most interactions unexplored in the literature

1,375 Interactions and powers of 50 anomalies: CV R^2



(e) 50 anomalies (f) PCs of 50 anomalies

⇒ Much more characteristics-sparsity; PC-sparse model still works well

Out-of-sample test with fully-withheld sample

- Analysis so far: Evaluation of SDFs with OOS data not used in estimation, but penalty parameters picked to maximize OOS performance
- Our anomaly data set likely includes data-mined anomalies and ones that have deteriorated due to learning (McLean and Pontiff 2016)
- Statistical significance of claim that SDF is not characteristics-sparse?

⇒ Conduct estimation of SDF coefficients & penalty choice based on pre-2006 data; use post-2005 for evaluation

Out-of-sample test with fully-withheld sample

- Constructing SDF is equivalent to finding the MVE portfolio
- Pre-2006 data yield SDF coefficient estimates $\hat{b}$ = MVE portfolio weights
- Apply $\hat{b}$ to post 2005 returns: OOS MVE portfolio return
- Test alphas relative to restricted benchmarks
  - Construct MVE portfolio weights from sparse characteristics-based models (e.g., FF 5-factor model) in pre-2006 data and apply weights to post-2005 returns
  - Regress OOS MVE portfolio return on MVE portfolio return of sparse characteristics-based factors to estimate OOS abnormal return

Out-of-sample test with fully-withheld sample

MVE portfolio's annualized OOS α (%) in the withheld sample (2005-2017)

Annualized OOS alpha in % (s.e.)

SDF factors \ Benchmark	CAPM	FF 6-factor	Char.-sparse	PC-sparse
50 anomaly portfolios	12.35 (5.26)	8.71 (4.94)	9.55 (3.95)	4.60 (2.22)
1,375 interactions of anomalies	25.00 (5.26)	22.79 (5.18)	21.68 (5.03)	12.41 (3.26)

SDF extraction in high-dimensional setting: Summary

- Characteristics-sparse SDF elusive
  - Not much redundancy among "anomaly" characteristics
  - ⇒ Debate whether we need 3, 4, 5, or 6 characteristics-based factors in SDF seems moot
- Instead, construct an SDF directly, or extract few factors that aggregate information from all characteristics
- E.g., principal component factors
  - Risk premia earned mostly by major sources of return covariance
  - ⇒ Makes economic sense both in asset pricing models with rational and in models with imperfectly rational investors
- PCs could be used as test assets and to look for correlations with macro risk factors, sentiment, intermediary risk factors, ...

Examples from recent literature on ML methods in cross-sectional AP

	Regularization	Assets	Nonlinearity
SDF models			
Kozak, Nagel, Santosh (2019)	elastic net	char. portfolios PC portfolios	interactions
Kozak (2019)	elastic net	char. portfolios PC portfolios	kernels
Giglio, Feng, and Xiu (2019)	Lasso	char. portfolios	-
DeMiguel et al. (2019)	Lasso	char. portfolios	-
Beta models			
Kelly, Pruitt, Su (2018)	PCA cutoff	indiv. stocks	-
Gu, Kelly and Xiu (2019)	Lasso	char. portfolios	autoencoder neural nets
Return prediction models			
Freyberger, Neuhierl, Weber (2018)	Group lasso	indiv. stocks	splines
Moritz and Zimmerman (2016)	Random forest	indiv. stocks	interactions
Gu, Kelly, Xiu (2018)	many	indiv. stocks	many

ML in cross-sectional AP: Summary

- ML well-suited to address challenges in cross-sectional AP
- Given the existing factor zoo, there is little point in analyzing a few return predictors in isolation: ML methods here to stay
- Outcomes much noisier than typical ML application: bringing in informed priors (economic restrictions & motivations) important
- Sparsity depends on rotation: can be quite sparse with PCs, but less so with original characteristics-based factors