Sparsity in Econometrics and Finance

David Puelz

Thesis defense
April 27, 2018

Outline

Motivating problem

Utility-based selection

Applications

- Portfolio selection
- Seemingly unrelated regression models
- Monotonic function estimation

Motivating problem: Charles' dilemma

He'd like to invest some money in the market.

He's heard passive funds are the way to go.

Motivating problem: Charles' dilemma

He'd like to invest some money in the market.

He's heard passive funds are the way to go.

4 Charts That Explain Why Active Funds Can't Beat **Passive** Funds
Forbes - Jun 14, 2017

OPINION

Index Funds Still Beat 'Active' Portfolio Management

By Burton G. Malkiel June 5, 2017 6:19 pm ET

There is no bette
Appeared in the Ju

Chinese investors pick **passive investing** for long-t
China Daily - Jun 4, 2017
Retail investors are increasingly choosing **passive investin**

Investors Now Have More than $4 Trillion in Exchange-Traded Products

By Ben Eisen May 10, 2017 11:47 am ET

The amount of money in exchange-traded funds and products topped $4 trillion globally last mont
rise of so-called passive investing.

But ... ∃ thousands of passive funds

SPDR
STATE STREET GLOBAL ADVISORS

iShares®
The world's No.1 ETF provider

BLACKROCK

Vanguard®

The context for this talk

This problem (and many others like it!) can be studied using variable selection techniques from statistics to induce *sparsity*.

The context for this talk

This problem (and many others like it!) can be studied using variable selection techniques from statistics to induce *sparsity*.

What's typically done? (broadly speaking)

- **Bayesian:** Shrinkage prior design.
- **Frequentist:** Penalized likelihood methods.

Common theme? Sparsity and inference go hand in hand.

Separating priors from utilities

Our view: Subset selection is a decision problem. We need a suitable loss function, **not** a more clever prior.

Separating priors from utilities

Our view: Subset selection is a decision problem. We need a suitable loss function, **not** a more clever prior.

This leads us to think of selection in a "post-inference world" by comparing models (or in this case, portfolios) based on utility.*

*sparsity and statistical uncertainty play a key role in this *post-inference* exercise.

Utility-based selection: Primitives

Let w_t be a portfolio decision, λ_t be a complexity parameter, Θ_t be a vector of model parameters, and $\tilde{R}_t$ be future data.

1. Loss function $\mathcal{L}(w_t, \tilde{R}_t)$ – measures utility.
2. Complexity function $\Phi(\lambda_t, w_t)$ – measures sparsity.
3. Statistical model $\Pi(\Theta_t)$ – characterizes uncertainty.
4. Regret tolerance κ – characterizes degree of comfort from deviating from a "target decision" (in terms of posterior probability).

Utility-based selection: Procedure

- Optimize $\mathbb{E}[\mathcal{L}(w_t, \tilde{R}_t) + \Phi(\lambda_t, w_t)]$, where the expectation is over $p(\tilde{R}_t, \Theta_t \mid \mathbf{R})$.

- Calculate regret versus a target w_t^* for decisions indexed by λ_t.

  $\rightarrow \quad \rho(w_{\lambda_t}, w_t^*, \tilde{R}_t) = \mathcal{L}(w_{\lambda_t}, \tilde{R}_t) - \mathcal{L}(w_t^*, \tilde{R}_t)$

- Select $w_{\lambda_t}^*$ as the decision satisfying the tolerance.

  $\rightarrow \quad \pi_{\lambda_t} = \mathbb{P}[\rho(w_{\lambda_t}, w_t^*, \tilde{R}_t) < 0]$ (satisfaction probability)

  $\rightarrow$ Select $w_{\lambda_t^*}$ s.t. $\pi_{\lambda_t^*} > \kappa$

What is innovative here?

Portfolio selection literature typically focuses on **one** of the following:

- Modeling inputs $\Theta_t = (\mu_t, \Sigma_t)$: Jobson (1980), Ledoit and Wolf (2007), Garlappi (2007), DeMiguel (2009) ...
- Optimizing in a clever way: Jagananathan (2002), Brodie (2009), Fan (2012), Fastrich (2013) ...

What is innovative here?

Portfolio selection literature typically focuses on **one** of the following:

- Modeling inputs $\Theta_t = (\mu_t, \Sigma_t)$: Jobson (1980), Ledoit and Wolf (2007), Garlappi (2007), DeMiguel (2009) ...
- Optimizing in a clever way: Jagananathan (2002), Brodie (2009), Fan (2012), Fastrich (2013) ...

Utility-based selection incorporates both modeling and optimization through analysis of $\rho(w_{\lambda_t}, w_t^*, \tilde{R}_t)$.

Example I: Long-only ETF investing

- Let $\tilde{R}_t$ be a vector of future ETF returns.
- Let w_t be the portfolio weight vector (decision) at time t.
- We use the log cumulative growth rate for our utility.

Primitives:

1. Loss: $-\log\left(1 + \sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right)$
2. Complexity: Number of funds in portfolio (think $\|w_t\|_0$)
3. Model: DLM for $\tilde{R}_t$ parameterized by $(\mu_t, \Sigma_t \mid D_{t-1})$

Example I: Long-only ETF investing

- Let $\tilde{R}_t$ be a vector of future ETF returns.
- Let w_t be the portfolio weight vector (decision) at time t.
- We use the log cumulative growth rate for our utility.

Primitives:

1. Loss: $-\log\left(1+\sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right)$
2. Complexity: Number of funds in portfolio (think $\|w_t\|_0$)
3. Model: DLM for $\tilde{R}_t$ parameterized by $(\mu_t, \Sigma_t \mid D_{t-1})$

Data: Monthly returns on 25 ETFs from 1992-2016.

Target: Fully invested (dense) portfolio.

Step 1: Constructing portfolio decisions

- Portfolio decisions have ≤ 5 funds.
- $\geq 25\%$ in SPY

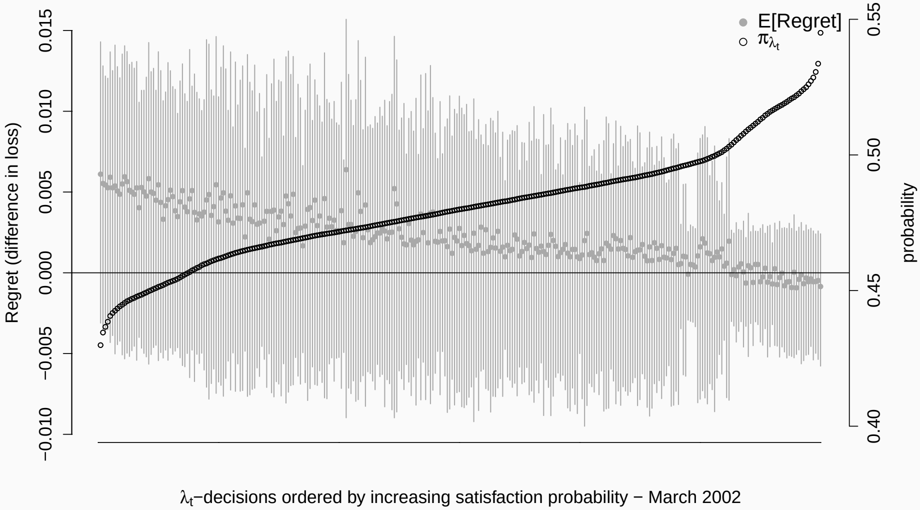
Decisions are found by minimizing expected loss for each time t.
Results in a choice of **12,950** decisions to choose among!!

Step 1: The expected loss

$$
\begin{aligned}
\mathcal{L}(w_t) &= \mathbb{E}_{\Theta_t}\mathbb{E}_{\tilde{R}_t|\Theta_t}\left[-\log(1+\Sigma_{k=1}^{N} w_t^k \tilde{R}_t^k) + \Phi(\lambda_t, w_t)\right] \\
&\approx \mathbb{E}_{\Theta_t}\mathbb{E}_{\tilde{R}_t|\Theta_t}\left[-\Sigma_{k=1}^{N} w_t^k \tilde{R}_t^k + \frac{1}{2}\Sigma_{k=1}^{N}\Sigma_{j=1}^{N} w_t^k w_t^j \tilde{R}_t^k \tilde{R}_t^j + \Phi(\lambda_t, w_t)\right] \\
&= -w_t^T \overline{\mu}_t + \frac{1}{2} w_t^T \overline{\Sigma}_t^{\mathrm{NC}} w_t + \Phi(\lambda_t, w_t).
\end{aligned}
$$

The past returns R_t enter into our utility consideration by defining the posterior predictive distribution.

Step 2: Compute and examine ρ for optimal decisions

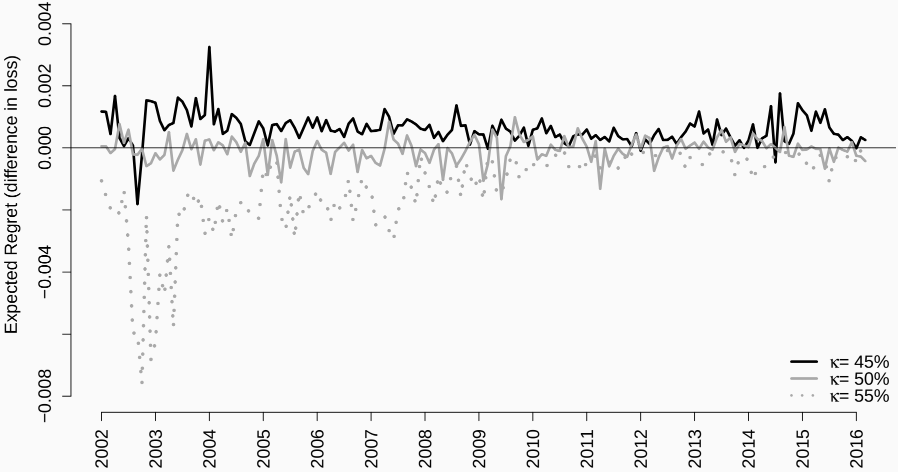


Step 3: Select decisions based on satisfaction threshold κ

Dates	SPY	EZU	EWU	EWY	EWG	EWJ	OEF	IVV	IVE	EFA	IWP	IWR	IWF	IWN	IWM	IYW	IYR	RSP
2003	25	-	58	-	-	-	-	-	-	-	-	-	-	8.3	-	-	-	8.3
2004	25	-	43	-	-	20	-	6.2	-	-	-	-	-	-	-	-	-	6.2
2005	25	-	25	-	6.2	13	-	-	-	-	-	-	-	-	-	-	30	-
2006	62	-	-	-	6.2	19	-	-	-	-	-	-	6.3	-	6.2	-	-	-
2007	75	-	-	25	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2008	44	-	-	-	8.3	21	-	-	-	26	-	-	-	-	-	-	-	-
2009	30	-	-	6.2	-	41	-	-	-	17	6.3	-	-	-	-	-	-	-
2010	75	-	-	8.3	-	-	-	-	-	-	8.3	-	-	-	-	8.3	-	-
2011	58	-	25	-	-	-	-	-	-	-	8.3	-	-	-	-	8.3	-	-
2012	29	8.3	-	-	-	54	-	-	-	-	-	-	-	-	-	8.3	-	-
2013	34	-	-	-	-	49	-	-	-	-	8.3	-	-	-	-	8.3	-	-
2014	25	-	-	-	-	37	26	-	-	6.2	-	6.2	-	-	-	-	-	-
2015	45	-	-	-	-	39	-	-	8.3	-	8.3	-	-	-	-	-	-	-
2016	35	-	-	-	-	40	-	17	-	-	8.3	-	-	-	-	-	-	-

Selected decisions for $\kappa = 45\%$ threshold.

What happens when κ is varied?



Higher satisfaction threshold $\Longrightarrow$ lower expected regret!

Comparing portfolios to their targets *out of sample*

	out-of-sample statistics		
	Sharpe ratio	s.d.	mean return
sparse	0.40	14.98	6.02
dense	0.45	14.41	6.47

Ex ante equivalence appears to carry over ex post.

There appear to be little ex post benefits of diversification.

What about other models / variable selection tasks?

Example II: Seemingly unrelated regressions

$$Y = \beta X + \epsilon, \qquad \epsilon \sim N(0, \Psi)$$

- Y is q length response vector
- X is p length covariate vector
- β is $q \times p$ coefficient matrix
- Ψ is non-diagonal matrix

finance: asset pricing, operations management: supply/demand structural equations, marketing: consumer preferences, economics: capital structure, firm composition, macroeconomic indicators.

Example II: Seemingly unrelated regressions

$$Y = \beta X + \epsilon, \qquad \epsilon \sim N(0, \Psi)$$

- Y is q length response vector
- X is p length covariate vector
- β is $q \times p$ coefficient matrix
- Ψ is non-diagonal matrix

finance: asset pricing, operations management: supply/demand structural equations, marketing: consumer preferences, economics: capital structure, firm composition, macroeconomic indicators.

We are interested in the structure of β!

Meat science

This paper has been published in *Meat Science* 92, 2012, p. 548–553
DOI: 10.1016/j.meatsci.2012.05.0257.

The Use of Seemingly Unrelated Regression (SUR) to Predict the Carcass Composition of Lambs

V.A.P. Cadavez[1], A. Henningsen[2]

Abstract

The aim of this study was to develop and evaluate models for predicting the carcass composition of lambs. Forty male lambs were slaughtered and their carcasses were cooled for 24 hours. The subcutaneous fat thickness was measured between the 12th and 13th rib and the total breast bone tissue thickness was taken in the middle of the second sternebrae. The left side of carcasses was dissected into five components and the proportions of lean meat (LMP), subcutaneous fat (SFP),

Factor selection for asset pricing

The Factor Zoo (Cochrane, 2011) – many possible factors ...

- Market
- Size
- Value
- Momentum
- Short and long term reversal
- Betting against β
- Direct profitability
- Dividend initiation
- Carry trade
- Liquidity
- Quality minus junk
- Investment
- Leverage
- ...

Example II: Factor selection for asset pricing

Let the return on test assets be R, and the return on factors be F.

$$R = \gamma F + \epsilon,\ \epsilon \sim N(0, \Psi)$$

Primitives:

1. Loss: $\mathcal{L}(\gamma, \tilde{R}, \tilde{F}) = -\log p(\tilde{R}|\tilde{F})$
2. Complexity: $\Phi(\lambda, \gamma) = \lambda \|\gamma\|_1$.
3. Model: $R|F$ with normal errors and conjugate g-priors and F via gaussian linear latent factor model.
4. Regret tolerance: Let's consider several κ's.

Example II: Factor selection for asset pricing

Let the return on test assets be R, and the return on factors be F.

$$R = \gamma F + \epsilon,\ \epsilon \sim N(0, \Psi)$$

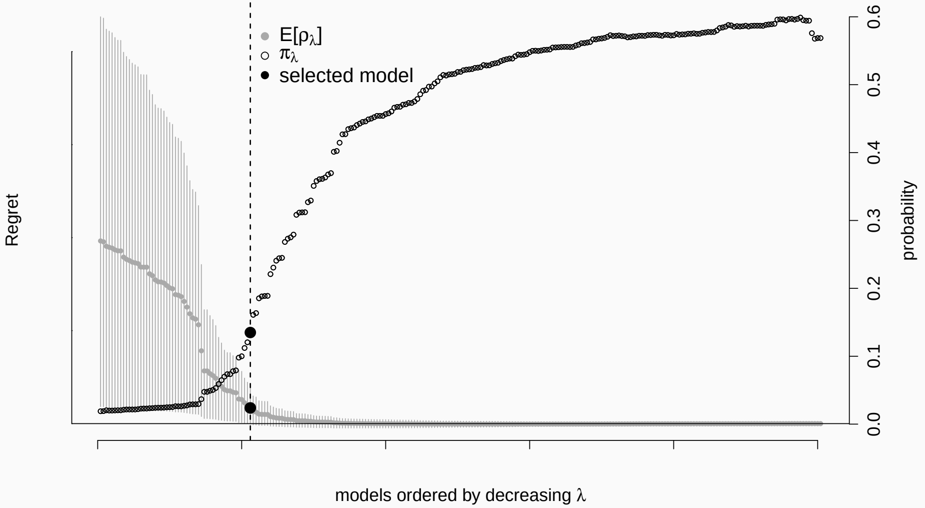
Primitives:

1. Loss: $\mathcal{L}(\gamma, \tilde{R}, \tilde{F}) = -\log p(\tilde{R}|\tilde{F})$
2. Complexity: $\Phi(\lambda, \gamma) = \lambda \left\|\gamma\right\|_1$.
3. Model: $R|F$ with normal errors and conjugate g-priors and F via gaussian linear latent factor model.
4. Regret tolerance: Let's consider several κ's.

Data: R**: 25 Fama-French portfolios,** F**: 10 factors from finance literature**

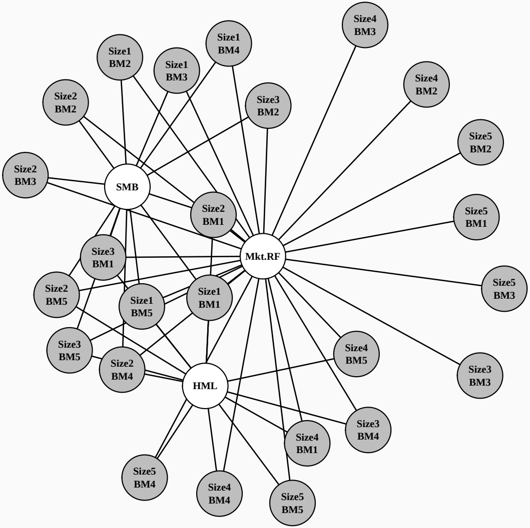
Targets: The $\lambda = 0$ model, i.e.: the fully dense graph

ρ distributions for different sparse graphs

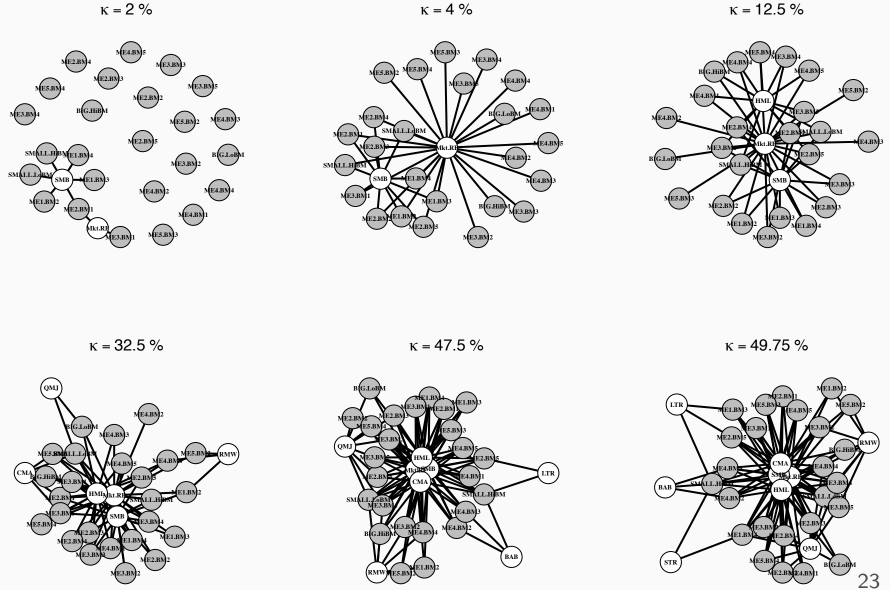


Factor selection graph, $\kappa = 12.5\%$

R: 25 Fama-French portfolios, *F*: 10 factors from finance literature



Selected graphs under different satisfaction tolerances κ



Example III: Monotonic function estimation

Goal: Describe expected returns with firm characteristics or accounting measures (size, book-to-market, momentum, ...).

$$\mathbb{E}[R_{it} \mid \mathrm{X}_{it-1}] = f(\mathrm{X}_{it-1})$$

R_{it}: excess return of firm i at time t

X_{it-1}: vector of characteristics of firm i at time t

Example III: Monotonic function estimation

Goal: Describe expected returns with firm characteristics or accounting measures (size, book-to-market, momentum, ...).

$$\mathbb{E}[R_{it} \mid \mathsf{X}_{it-1}] = f(\mathsf{X}_{it-1})$$

R_{it}: excess return of firm i at time t
X_{it-1}: vector of characteristics of firm i at time t

We would like to learn f !

Portfolio sorts are one way to understand f ...

Jegadeesh and Titman (2001)

Table I

Momentum Portfolio Returns

This table reports the monthly returns for momentum portfolios formed based on past six-month returns and held for six months. P1 is the equal-weighted portfolio of 10 percent of the stocks with the highest returns over the previous six months, P2 is the equal-weighted portfolio of the 10 percent of the stocks with the next highest returns, and so on. The "All stocks" sample includes all stocks traded on the NYSE, AMEX, or Nasdaq excluding stocks priced less than $5 at the beginning of the holding period and stocks in the smallest market cap decile (NYSE size decile cutoff). The "Small Cap" and "Large Cap" subsamples comprise stocks in the "All Stocks" sample that are smaller and larger than the median market cap NYSE stock respectively. "EWI" is the returns on the equal-weighted index of stocks in each sample.

	All Stocks			Small Cap			Large Cap		
	1965–1998	1965–1989	1990–1998	1965–1998	1965–1989	1990–1998	1965–1998	1965–1989	1990–1998
P1 (Past winners)	1.65	1.63	1.69	1.70	1.69	1.73	1.56	1.52	1.66
P2	1.39	1.41	1.32	1.45	1.50	1.33	1.25	1.24	1.27
P3	1.28	1.30	1.21	1.37	1.42	1.23	1.12	1.10	1.19
P4	1.19	1.21	1.13	1.26	1.34	1.05	1.10	1.07	1.20
P5	1.17	1.18	1.12	1.26	1.33	1.06	1.05	1.00	1.19
P6	1.13	1.15	1.09	1.19	1.26	1.01	1.09	1.05	1.20
P7	1.11	1.12	1.09	1.14	1.20	0.99	1.09	1.04	1.23
P8	1.05	1.05	1.03	1.09	1.17	0.89	1.04	1.00	1.17
P9	0.90	0.94	0.77	0.84	0.95	0.54	1.00	0.96	1.09
P10 (Past losers)	0.42	0.46	0.30	0.28	0.35	0.08	0.70	0.68	0.78
P1–P10	1.23	1.17	1.39	1.42	1.34	1.65	0.86	0.85	0.88
t statistic	6.46	4.96	4.71	7.41	5.60	5.74	4.34	3.55	2.59
EWI	1.09	1.10	1.04	1.13	1.19	0.98	1.03	1.00	1.12

Challenges and a solution

- X_{it-1} is multidimensional.
- Even if we had only 12 characteristics and sorted into quintiles along each dimension, that requires constructing $5^{12} =$ **244140625** portfolios!

We propose modeling the CEF as an additive quadratic spline model (with monotonicity constraints *and* time variation):

$$\boxed{\mathbb{E}[R_{it} \mid X_{it-1}] = \alpha_t + \sum_{k=1}^{K} g_{kt}(x_{ki,t-1})}$$

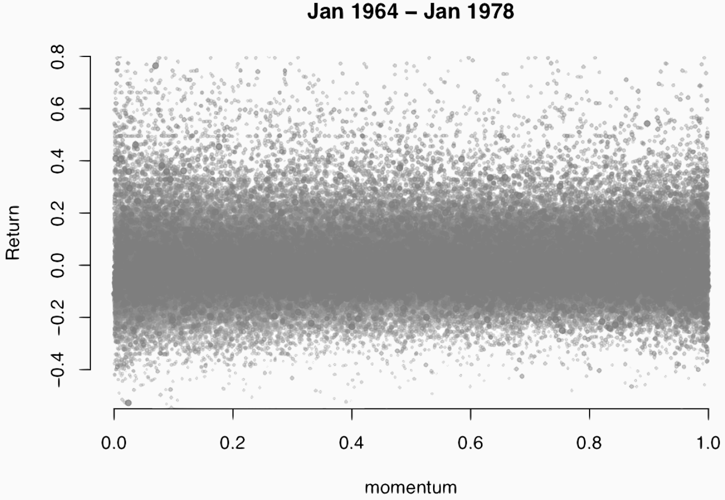
Why monotonicity?

Finance theory often tells us that expected returns *increase* or *decrease* in each characteristic. Ex: past high-performing firms have higher returns than past weak-performing firms, on average.

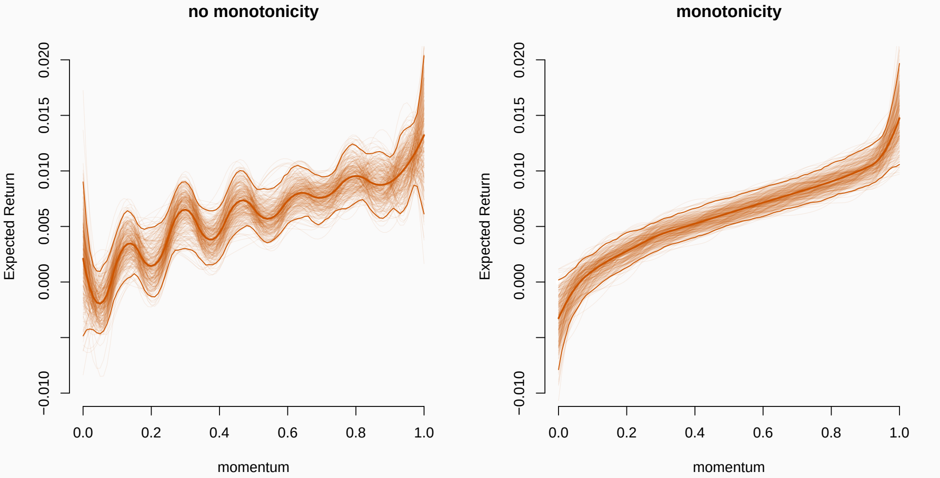
Using this information is statistically advantageous!

Why monotonicity?

Finance data is noisy – any bias aids in more precise estimation.

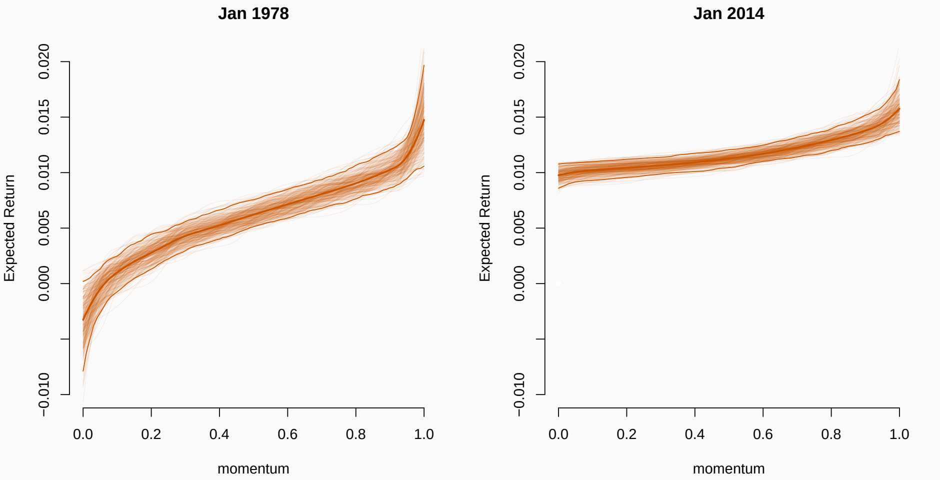


Estimated functions at January 1978



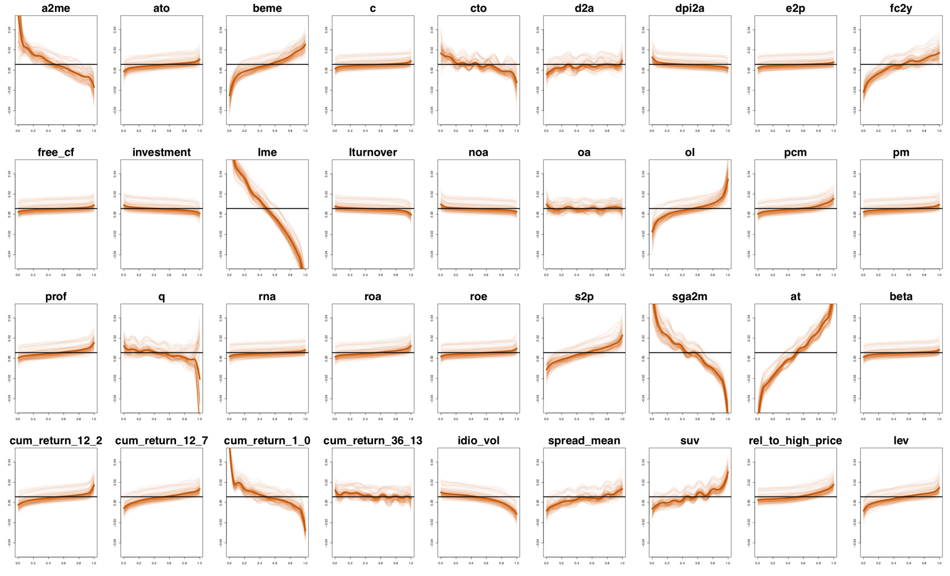
monotonicity is enforced by linear constraints on spline coefficients

How does the function vary over time?

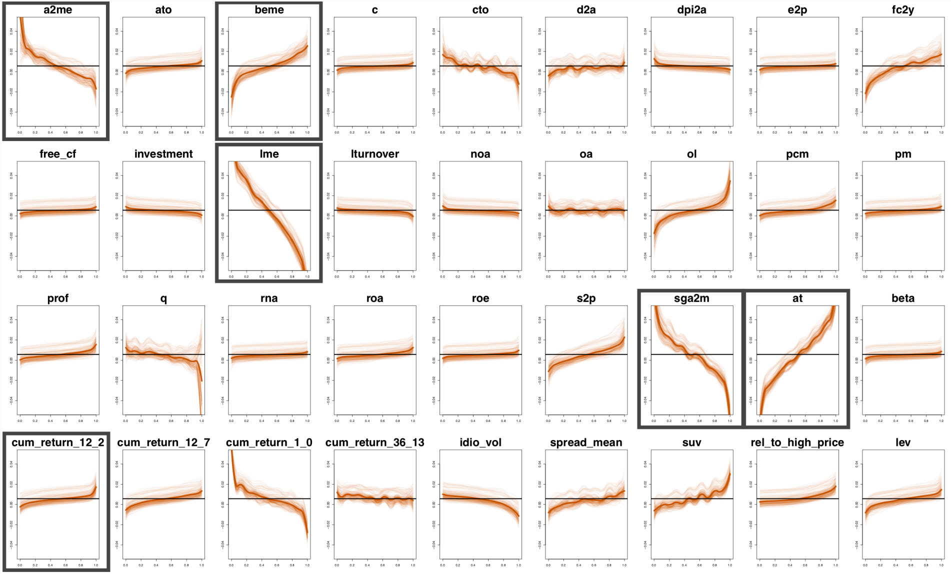


dynamics are modeled by likelihood discounting, McCarthy and Jenson (2016)

A model with 36 characteristics - January 1978



A model with 36 characteristics - January 1978



Future work

Where to go from here?

- New utility specifications: value-at-risk and simulation based. Analyzing other properties of the regret distribution.
- New models: multinomial regression and classification models, nonlinear and nonparametric models.
- New application areas: corporate finance, marketing, macroeconomics.

Existing papers:

Regret-based selection for sparse dynamic portfolios. submitted (2017). Thesis ch. 2.

Variable selection in SUR models with random predictors. Bayesian Analysis (2017). Thesis ch. 3.

Monotonic effects of characteristics on returns. working paper (2018). Thesis ch. 3.5.

Concluding thoughts, and thanks!

- Passive investing, SUR model selection, and monotonic function estimation approached using new feature selection technique.
- **Utility functions can enforce inferential preferences that are not prior beliefs.**
- Statistical uncertainty should be used as a guide to avoid overfitting.

Extra slides

Treatment effect estimation

Suppose we are trying to estimate the treatment effect of dietary kale on cholesterol level. But ... we only have observational data.

$$Y_i = \beta_0 + \alpha Z_i + \epsilon_i$$

- Y_i is cholesterol level
- Z_i is amount of kale eaten.

Problem: Gym rats tend to eat more kale!

In other words, exercise is predictive of cholesterol *and* kale intake! This leads to omitted variable bias.

$$Y_i = \beta_0 + \alpha Z_i + \epsilon_i$$

Because $\text{cov}(Z_i, \epsilon_i) \neq 0$ we can write:

$$Y_i = \beta_0 + \alpha Z_i + wZ_i + \tilde{\epsilon}_i$$

with $\text{cov}(Z_i, \tilde{\epsilon}_i) = 0$, **we mis-estimate α as $\alpha + w$!**

Solution: "Adjust" for weekly exercise

By controlling for weekly exercise X_i in the regression

$$Y_i = \beta_0 + \alpha Z_i + \beta X_i + \epsilon_i$$

we can "clear out" the confounding.

Conditional on X_i, $\text{cov}(Z_i, \epsilon_i) = 0$ and we're all set!

But what if X_i is a big vector, and we don't know which covariates to control for? (Enter sparsity).

Regularized treatment effect estimation

Consider the model with no intercept and **many** covariates X_i:

$$Y_i = \alpha Z_i + \mathrm{X}_i^T \boldsymbol{\beta} + \epsilon_i$$

We can induce sparsity with a ridge prior on $\boldsymbol{\beta}$ and leaving α unpenalized. This injects bias into treatment effect estimate:

$$\mathrm{bias}(\hat{\alpha}_{\mathrm{ridge}}) = -(\mathrm{Z}^T\mathrm{Z})^{-1}\mathrm{Z}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda\mathbb{I}_p - \mathbf{X}^T\hat{\mathbf{X}}_{\mathrm{Z}}\right)^{-1}\lambda\beta \neq 0$$

where $(\mathrm{Z}^T\mathrm{Z})^{-1}\mathrm{Z}^T\mathbf{X}$ is a p-length vector of coefficients from p univariate regressions of each X^j on Z and $\hat{\mathbf{X}}_{\mathrm{Z}} = \mathrm{Z}(\mathrm{Z}^T\mathrm{Z})^{-1}\mathrm{Z}^T\mathbf{X}$ are the predicted values from these regressions.

This nonzero bias is referred to regularization-induced confounding (RIC).

A different approach eliminates RIC

Consider the model where a likelihood is included for Z:

$$\text{Selection equation:} \quad Z_i = \mathrm{X}_i^T \gamma + \epsilon_i$$
$$\text{Response equation:} \quad Y_i = \alpha Z_i + \mathrm{X}_i^T \boldsymbol{\beta} + \nu_i$$

- Extract propensity from selection equation: $\hat{\mathrm{Z}} \approx \mathbf{X}\hat{\gamma}$
- Augment covariates with propensity $\hat{\mathbf{X}}_{\text{new}} = \begin{pmatrix} \mathrm{Z} & \hat{\mathrm{Z}} & \mathbf{X} \end{pmatrix}$
- Ridge estimate with Z and $\hat{\mathrm{Z}}$ unpenalized mitigates RIC

Regularization and confounding in linear regression for treatment effect estimation.

Bayesian Analysis (2017).

A different approach eliminates RIC

The bias of the treatment effect becomes:

$$\text{bias}(\hat{\alpha}_{\text{ridge}}) = -\{(\tilde{Z}^T\tilde{Z})^{-1}\tilde{Z}^T\mathbf{X}\}_1 \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbb{I}_p - \mathbf{X}^T\hat{\mathbf{X}}_Z\right)^{-1} \lambda\beta \approx 0$$

where $\tilde{Z} = \begin{pmatrix} Z & \hat{Z} \end{pmatrix}$ and $\{\cdot\}_1$ corresponds to the top row of the matrix $\{\cdot\}$. $\{(\tilde{Z}^T\tilde{Z})^{-1}\tilde{Z}^T\mathbf{X}\}_1$ are the coefficients on Z in the two variable regressions of each X_i on $\begin{pmatrix} Z & \hat{Z} \end{pmatrix}$.

Controlling for the propensity of the treatment wipes out regularization-induced confounding (RIC) in the treatment effect estimate.

Next steps

$$\text{Selection equation:} \quad Z_i = \mathrm{X}_i^T \gamma + \epsilon_i$$
$$\text{Response equation:} \quad Y_i = \alpha Z_i + \mathrm{X}_i^T \boldsymbol{\beta} + \nu_i$$

- Develop fast empirical Bayes approach to regularize two equation system.
- Account for clustered observations using block boostrapping.
- Many application in social science, including micro/macroeconomics and corporate finance.
- RIC still exists even in nonlinear, statistical learning based models! Why? Because they especially need to be regularized. Extend this approach to random forests.

A dynamic regression model giving moments (μ_t, Σ_t)

$$\underbrace{\begin{aligned}
&\tilde{R}_t^i = (\beta_t^i)^T \tilde{R}_t^F + \epsilon_t^i, \quad \epsilon_t^i \sim N(0, 1/\phi_t^i), \; \beta_t^i = \beta_{t-1}^i + w_t^i, \quad w_t^i \sim \mathrm{T}_{n_{t-1}^i}(0, W_t^i), \\
&\quad \beta_0^i \mid D_0 \sim \mathrm{T}_{n_0^i}(m_0^i, C_0^i), \quad \phi_0^i \mid D_0 \sim \mathrm{Ga}(n_0^i/2, d_0^i/2), \\
&\quad \beta_t^i \mid D_{t-1} \sim \mathrm{T}_{n_{t-1}^i}(m_{t-1}^i, R_t^i), \quad R_t^i = C_{t-1}^i/\delta_\beta, \\
&\quad \phi_t^i \mid D_{t-1} \sim \mathrm{Ga}(\delta_\epsilon n_{t-1}^i/2, \delta_\epsilon d_{t-1}^i/2), \\
&\tilde{R}_t^F = \mu_t^F + \nu_t, \quad \nu_t \sim \mathrm{N}(0, \Sigma_t^F), \quad \mu_t^F = \mu_{t-1}^F + \Omega_t \quad \Omega_t \sim \mathrm{N}(0, W_t, \Sigma_t^F), \\
&\quad (\mu_0^F, \Sigma_0^F \mid D_0) \sim \mathrm{NW}_{n_0}^{-1}(m_0, C_0, S_0), \\
&\quad (\mu_t^F, \Sigma_t^F \mid D_{t-1}) \sim \mathrm{NW}_{\delta_F n_{t-1}}^{-1}(m_{t-1}, R_t, S_{t-1}), \quad R_t = C_{t-1}/\delta_c
\end{aligned}}$$

$$\mu_t = \beta_t^T \mu_t^F$$

$$\Sigma_t = \beta_t \Sigma_t^F \beta_t^T + \Psi_t$$

$\rightarrow$ Moments are used in the expected loss minimization

$\rightarrow$ Predictive distribution is used to compute ρ

Formulating as a convex penalized optimization

Define $\overline{\Sigma} = LL^T$.

$$\begin{aligned}\mathcal{L}(w) &= -w^T\overline{\mu} + \frac{1}{2}w^T\overline{\Sigma}w + \lambda \|w\|_1 \\ &= \frac{1}{2}\left\|L^T w - L^{-1}\overline{\mu}\right\|_2^2 + \lambda \|w\|_1 .\end{aligned}$$

Now, we can solve the optimization using existing algorithms, such as `lars` of Efron et. al. (2004).

Example: Gross exposure complexity function

- Let $\tilde{R}_t$ be a vector of N future asset returns.
- Let w_t be the portfolio weight vector (decision) at time t.
- We use the log cumulative growth rate for our utility.

Primitives:

1. Loss: $-\log\left(1 + \sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right)$
2. Complexity: $\lambda_t \left\| w_t \right\|_1$
3. Model: DLM for $\tilde{R}_t$ parameterized by (μ_t, Σ_t)
4. Regret tolerance: Let's consider several κ's.

Example: Gross exposure complexity function

- Let $\tilde{R}_t$ be a vector of N future asset returns.
- Let w_t be the portfolio weight vector (decision) at time t.
- We use the log cumulative growth rate for our utility.

Primitives:

1. Loss: $-\log\left(1 + \sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right)$
2. Complexity: $\lambda_t \left\| w_t \right\|_1$
3. Model: DLM for $\tilde{R}_t$ parameterized by (μ_t, Σ_t)
4. Regret tolerance: Let's consider several κ's.

Assume the target is fully invested (dense) portfolio.

Example: Gross exposure complexity function

- Let $\tilde{R}_t$ be a vector of N future asset returns.
- Let w_t be the portfolio weight vector (decision) at time t.
- We use the log cumulative growth rate for our utility.

Primitives:

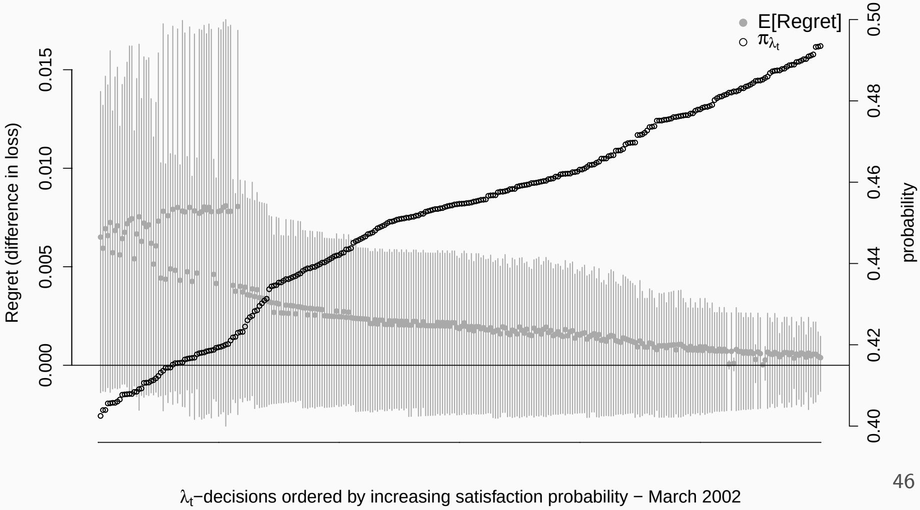
1. Loss: $-\log\left(1 + \sum_{k=1}^{N} w_t^k \tilde{R}_t^k\right)$
2. Complexity: $\lambda_t \left\| w_t \right\|_1$
3. Model: DLM for $\tilde{R}_t$ parameterized by (μ_t, Σ_t)
4. Regret tolerance: Let's consider several κ's.

Assume the target is fully invested (dense) portfolio.

Data: Returns on 25 ETFs from 1992-2016.

Optimal decisions lined up for a snapshot in time

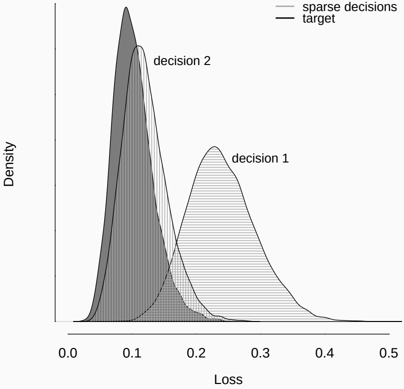
After optimizing expected loss for 500 λ_t's, we compute regret $\rho(w_{\lambda_t}, w_t^*, \tilde{R}_t)$ (left axis) and π_{λ_t} (right axis).

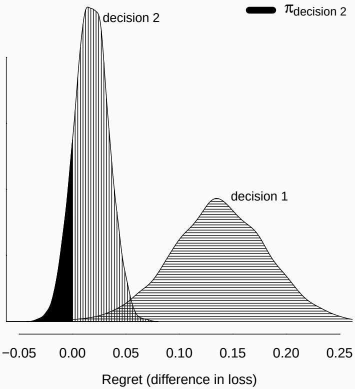


Regret-based selection: Illustration

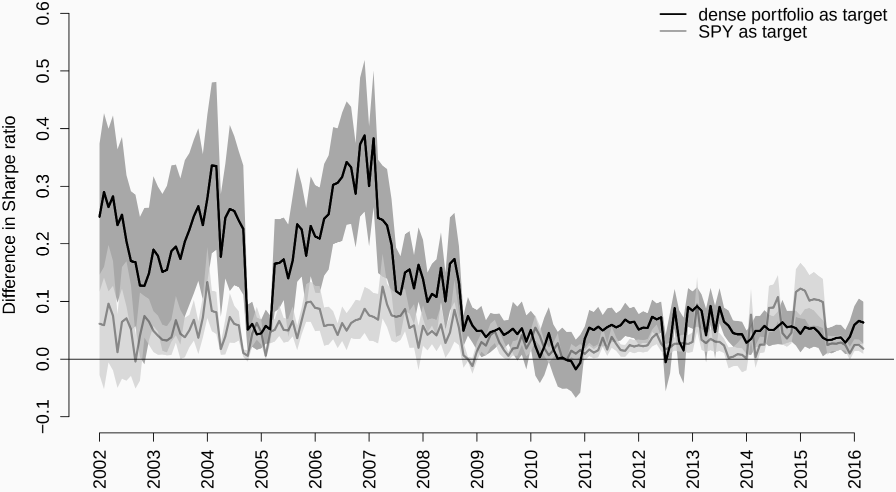
d_λ : sparse decisions, d^* : target decision.

$\pi_\lambda = \mathbb{P}[\rho(d_\lambda, d^*, \tilde{Y}) < 0]$: probability of not regretting λ-decision.





Ex ante $SR_{\textbf{target}} - SR_{\textbf{decision}}$ evolution



UBS for Monotonic function estimation

The regression model is:

$$R_{it} = \alpha_t + \sum_{k=1}^{K} f_{kt}(x_{ki,t-1}) + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2)$$

Insight – with quadratic splines for all f_{kt}, this can be written as a predictive regression:

$$\mathrm{R}_t \sim \mathrm{N}\left(\mathbb{X}_{t-1}\mathbf{B}_t,\ \sigma_t^2 \mathbb{I}_{n_t}\right)$$

where

$$\mathbb{X}_{t-1} = \begin{bmatrix} \mathbf{1}_{n_t} & \mathbf{X}_{t-1} \end{bmatrix}, \quad \mathbf{B}_t = \begin{bmatrix} \alpha_t & \boldsymbol{\beta}_t \end{bmatrix}$$

$\mathbf{X}_{t-1}$ is matrix of size $n_t \times K(m+2)$, $\boldsymbol{\beta}_t$ is vector of size $K(m+2)$. Therefore, each firm is given a row in $\mathbf{X}_{t-1}$, and each $m+2$ block of $\boldsymbol{\beta}_t$ corresponds to the coefficients on the spline basis for a particular characteristic, k.

UBS for Monotonic function estimation

We can now proceed as Hahn and Carvalho (2015). The loss function is the negative log density of the regression plus a penalty function Φ with parameter λ_t. Also, let the “sparsified action” for the coefficient matrix $\mathbf{A}_t$.

$$\mathcal{L}_t(\tilde{\mathrm{R}}_t, \mathbf{A}_t, \Theta_t) = \frac{1}{2}(\tilde{\mathrm{R}}_t - \mathbb{X}_{t-1}\mathbf{A}_t)^T(\tilde{\mathrm{R}}_t - \mathbb{X}_{t-1}\mathbf{A}_t) + \Phi(\lambda_t, \mathbf{A}_t).$$

After integrating over $p(\tilde{\mathrm{R}}_t, \Theta_t)$, we obtain:

$$\boxed{\mathcal{L}_{\lambda_t}(\mathbf{A}_t) = \left\|\mathbb{X}_{t-1}\mathbf{A}_t - \mathbb{X}_{t-1}\overline{\mathbf{B}}_t\right\|_2^2 + \Phi(\lambda_t, \mathbf{A}_t)}$$

Modeling Time-dynamics: McCarthy and Jensen (2016)

- Power-weighted likelihoods let information decay over time
- To estimate parameters at time τ, let $\delta_t = 0.99^{\tau - t}$, such that $\delta_1 \leq \delta_2 \leq ... \leq \delta_\tau = 1$, the likelihood at time $\tau \in \{1, ..., T\}$ is

$$p(R_1, ..., R_\tau | \Theta_\tau) = \prod_{t=1}^{\tau} p(R_t | \Theta_\tau)^{\delta_t}.$$

Model Summary

$$
\begin{aligned}
R_t|\cdot &\sim N\left(\alpha_t 1_{n_t} + \sum_{k=1}^{K} f_{kt}(x_{k,t-1}),\ \sigma_t^2 I_n\right)^{\delta_t} \\
f_{kt}(x_{k,t-1}) &= X_{k,t-1}\beta_{kt} = X_{k,t-1}L^{-1}L\beta_{kt} = W_{kt}\gamma_{kt} \\
\alpha_t &\sim N(0, 10^{-2}) \\
\sigma_t^2 &\sim U(0, 10^3) \\
(\gamma_{jkt}|I_{jkt} = 1, \sigma_t^2) &\sim N_+(0, c_k\sigma_t^2) \\
(\gamma_{jkt}|I_{jkt} = 0) &= 0 \\
I_{jkt} &\sim Bn(p_{jk} = 0.2).
\end{aligned}
$$

Data

Freyberger, Neuhierl, and Weber (2017)'s dataset:

- CRSP monthly stock returns for most US traded firms
- 36 characteristics from Compustat and CRSP, including size, momentum, leverage, etc.
- July 1962 - June 2014

Presence and direction of monotonicity is determined by important paper in the literature