Penalized Utility Estimators in Finance

David Puelz (UT Austin)
Carlos M. Carvalho (UT Austin)
P. Richard Hahn (Chicago Booth)

SBIES 2016
April 30, 2016

Two problems

1. Investing: Among thousands of choices, which passive funds should I invest in?

Two problems

1. Investing: Among thousands of choices, which passive funds should I invest in?

2. Asset pricing: Which risk factors matter?

Two problems

1. Investing: Among thousands of choices, which passive funds should I invest in?

2. Asset pricing: Which risk factors matter?

How are these connected?

An answer: Both can be studied using variable selection techniques from statistics.

How is variable selection (sparsifying) typically done?

$\Rightarrow$ Frequentist / Penalized likelihood: Forward/backward stepwise selection. LARS, LASSO, Group Lasso, Ridge.

$\Rightarrow$ Bayesian: Priors forcing irrelevant coefficients to zero.

How is variable selection (sparsifying) typically done?

⇒ Frequentist / Penalized likelihood: Forward/backward stepwise selection. LARS, LASSO, Group Lasso, Ridge.

Challenges: What stopping criterion? What penalty parameter (λ)?

⇒ Bayesian: Priors forcing irrelevant coefficients to zero.

Challenges: Mixing inference with desire for sparsity. How should inclusion probabilities be interpreted / used?

Intelligently summarizing the posterior

We can overcome these challenges with a two-step approach.[1]

[1] "*Decoupling Shrinkage and Selection in Bayesian Linear Models*." Hahn and Carvalho. Journal of the American Statistical Association, 2015.

Intelligently summarizing the posterior

We can overcome these challenges with a two-step approach.[1]

Given a loss function dependent on parameters θ and action γ to be taken by the scientist:

$$\mathcal{L}(\gamma) = f(\gamma, \theta) + \lambda * \text{penalty}(\gamma).$$

[1] "*Decoupling Shrinkage and Selection in Bayesian Linear Models*." Hahn and Carvalho. Journal of the American Statistical Association, 2015.

Intelligently summarizing the posterior

We can overcome these challenges with a two-step approach.[1]

Given a loss function dependent on parameters θ and action γ to be taken by the scientist:

$$\mathcal{L}(\gamma) = f(\gamma, \theta) + \lambda * \text{penalty}(\gamma).$$

1. Characterize uncertainty: $p(\theta | Data)$.
2. Optimize $\mathcal{L}(\gamma)$ integrated over this uncertainty.

2a. Examine solution path to choose level of sparsity.

[1] "*Decoupling Shrinkage and Selection in Bayesian Linear Models*." Hahn and Carvalho. Journal of the American Statistical Association, 2015.

The main results...

1. Build high Sharpe ratio, simple ETF portfolios.

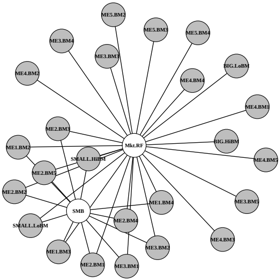
ETF	IWR	RSP	IYR	IYW
weight	56%	21.5%	13.9%	8.6%
style	*mid-cap*	*equal weight*	*real estate*	*tech*

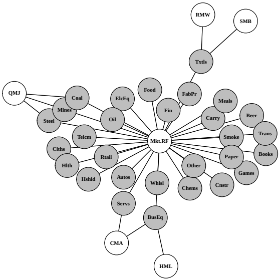
The main results...

1. Build high Sharpe ratio, simple ETF portfolios.

ETF	IWR	RSP	IYR	IYW
weight	56%	21.5%	13.9%	8.6%
style	*mid-cap*	*equal weight*	*real estate*	*tech*

2. Select risk factors for asset pricing.





Passive Investing

The mean-variance setup

- Action: Portfolio weights w.
- Loss: Given future asset returns, $\tilde{R}$:

  $\mathcal{L}(w, \tilde{R}) = -\sum_{k=1}^{N} w_k \tilde{R}_k + \frac{1}{2} \sum_{k=1}^{N} \sum_{j=1}^{N} w_k w_j \tilde{R}_k \tilde{R}_j + \lambda \left\| w \right\|_1$
- *Goal:* Maximize Sharpe ratio subject to finding a a sparse representation of w.

Where is the uncertainty?

- Assume future asset returns follow $\tilde{R} \sim \Pi(\mu, \Sigma)$.
- The parameters $\theta = (\mu, \Sigma)$ are uncertain, too!
- Our expected loss is derived by integrating over $p(\tilde{R}|\theta)$ followed by $p(\theta|R)$, the posterior distribution over θ.

Integrating over uncertainty

$$
\begin{aligned}
\mathcal{L}(w) &= \mathbb{E}_{\theta}\mathbb{E}_{\tilde{R}|\theta}\left[-\sum_{k=1}^{N} w_k \tilde{R}_k + \frac{1}{2}\sum_{k=1}^{N}\sum_{j=1}^{N} w_k w_j \tilde{R}_k \tilde{R}_j + \lambda \left\|w\right\|_1\right] \\
&= \mathbb{E}_{\theta}\left[-w^T \mu + \frac{1}{2} w^T \Sigma w\right] + \lambda \left\|w\right\|_1 \\
&= -w^T \overline{\mu} + \frac{1}{2} w^T \overline{\Sigma} w + \lambda \left\|w\right\|_1 .
\end{aligned}
$$

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

Formulating as a convex penalized optimization

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

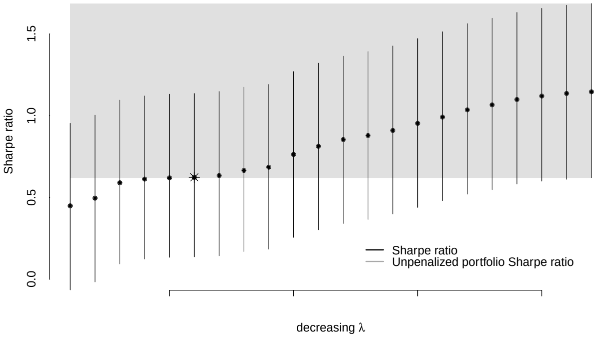
$$\begin{aligned}\mathcal{L}(w) &= -w^T\overline{\mu} + \frac{1}{2}w^T\overline{\Sigma}w + \lambda\left\|w\right\|_1 \\ &= \frac{1}{2}\left\|L^T w - L^{-1}\overline{\mu}\right\|_2^2 + \lambda\left\|w\right\|_1.\end{aligned}$$

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

Application to ETF investing

- Data: Returns on 25 ETFs from 1992-2015.
- Model: Assume returns follow a latent factor model.
- Question: Optimal portfolio of a small number of ETFs?

Posterior summary plot



ETF	IWR	RSP	IYR	IYW
weight	56%	21.5%	13.9%	8.6%
style	*mid-cap*	*equal weight*	*real estate*	*tech*

Find the smallest portfolio such that with probability 99% I give up less than (blank) in Sharpe ratio.

Which risk factors matter?

The Factor Zoo (Cochrane, 2011)

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

The Factor Zoo (Cochrane, 2011)

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

A loss function for determining important factors

- Test assets: R, Factors: $\mathrm{R} = \gamma \mathrm{F} + \epsilon$, $\epsilon \sim N(0, \Psi)$.
- Define loss by conditional likelihood, $p(R|F)$.
- *Goal:* find a sparse representation of γ, where γ is a matrix relating R and F.

Integrating conditional likelihood over $p(\tilde{R}, \tilde{F}|\theta)$ and $p(\theta|R, F)$ gives another convex penalized objective function!

After integration, the loss function is:

$$\mathcal{L}(\gamma) = -\frac{1}{2} \left\| \left[[L^T \otimes \mathbb{I}] \mathbf{vec}(\gamma) - \mathbf{vec}(fL^{-1}) \right] \right\|_2^2 + \lambda \left\| \mathbf{vec}(\gamma) \right\|_1$$

where:

$$LL^T = \overline{\Sigma_f} + \Sigma_{\mu_f} + \overline{\mu_f}\,\overline{\mu_f}^T, \quad f = \overline{\boldsymbol{\beta}\Sigma_f} + \Sigma_{\mu_f \mu_r} + \overline{\mu_r}\,\overline{\mu_f}^T$$

Risk Factor Selection

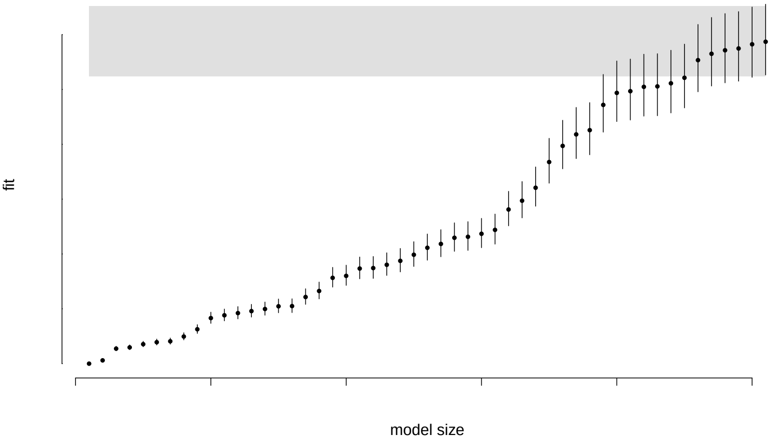
- Test Assets: Fama-French 25 Portfolios and 30 Industry Portfolios from 1963-2015.
- Factors: 10 factors proposed in finance literature.

- Model $p(R|F)$ with normal errors and conjugate g-priors.
- Model $p(F)$ via gaussian linear latent factor model.[2]

- Question: Which factors are most important for pricing?

[2] Taking advantage of compositional representation of the joint: $p(R, F) = p(R|F)p(F)$

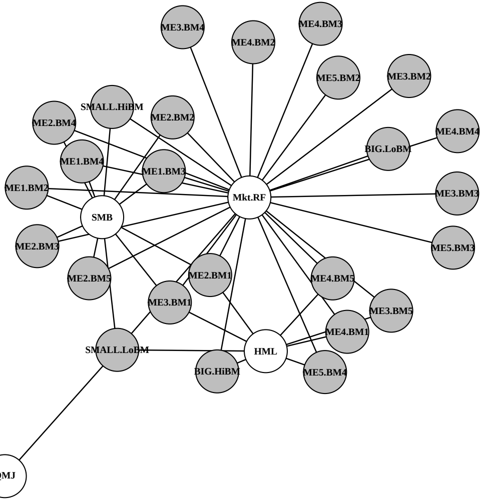
Posterior summary plot



Model size here refers to nonzero entries of γ, or equivalently, edges of graph representing γ.

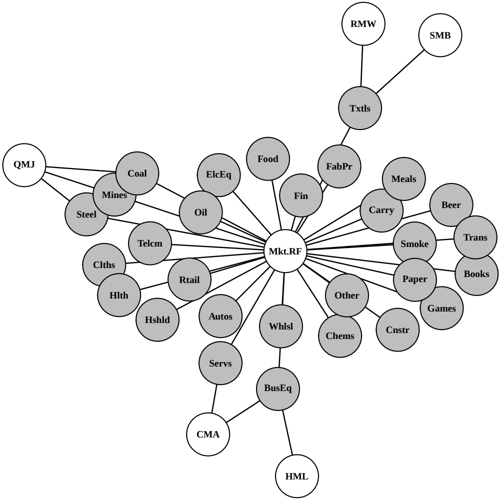
Factor selection graph

R: Fama-French 25 Portfolios, F: 10 factors



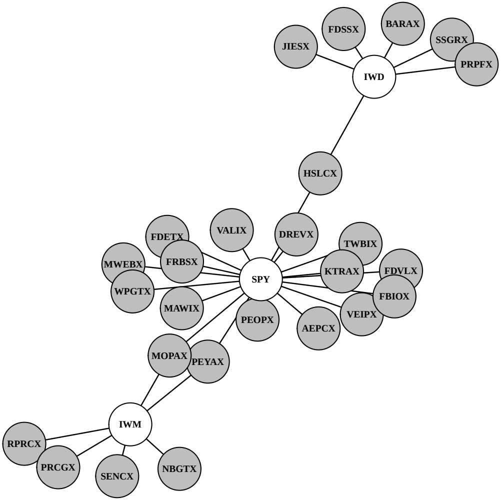
Factor selection graph

R: 30 Industry Portfolios, F: 10 factors



Another application: ETF selection

R: 100 Mutual funds, F: 25 ETFs



Concluding thoughts

- Passive investing and factor selection for asset pricing models approached using new variable selection technique.
- Utility functions can enforce inferential preferences that are not prior beliefs.
- Ideas presented are generalizable and *scalable*. There is more work to be done ..
- Thanks!