Penalized Utility Estimators in Finance

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

Goldman Sachs ISG
February 12, 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?

Statistics

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

Now, a brief statistics interlude ...

Reasons for variable selection

- ⇒ Because we're scientists and we test hypotheses!
- ⇒ Because fewer variables are faster to compute with!
- ⇒ Because thinking hard about fewer things is easier than thinking hard about many things.

There are probably others. The important point is that they are distinct reasons.

My motivation for sparsity

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)           -8.211e+06  8.519e+04 -96.383  < 2e-16 ***
X.trim350              2.657e+04  1.160e+04   2.290 0.022024 *  
X.trim400             -1.275e+04  1.649e+04  -0.773 0.439459    
X.trim420              4.954e+04  1.178e+04   4.205 2.62e-05 ***
X.trim430              2.662e+04  1.177e+04   2.263 0.023662 *  
X.trim500              2.935e+04  1.177e+04   2.494 0.012623 *  
X.trim550             -4.942e+03  1.078e+04  -0.458 0.646705    
X.trim55 AMG           2.823e+04  1.178e+04   2.397 0.016542 *  
X.trim600              4.477e+03  1.079e+04   0.415 0.678100    
X.trim63 AMG           4.445e+04  1.080e+04   4.117 3.84e-05 ***
X.trim65 AMG           6.142e+03  1.083e+04   0.567 0.570524    
X.trimunsp             2.666e+04  1.081e+04   2.466 0.013657 *  
X.conditionNew         3.513e+04  2.284e+02 153.819  < 2e-16 ***
X.conditionUsed       -4.337e+03  1.993e+02 -21.758  < 2e-16 ***
X.isOneOwnert         -5.043e+02  1.725e+02  -2.924 0.003459 ** 
X.mileage             -1.324e-01  2.522e-03 -52.488  < 2e-16 ***
X.year                 4.103e+03  4.224e+01  97.134  < 2e-16 ***
X.colorBlack          -4.381e+02  6.660e+02  -0.658 0.510685    
X.colorBlue           -6.830e+02  7.000e+02  -0.976 0.329230    
X.colorBronze          3.997e+03  3.460e+03   1.155 0.247937    

Residual standard error: 10740 on 39391 degrees of freedom
Multiple R-squared:  0.9429,Adjusted R-squared:  0.9428
```

Vast literature on variable selection (a.k.a. sparsifying)

$\Rightarrow$ Frequentist: forward/backward stepwise selection.

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

$\Rightarrow$ Penalized likelihood: LARS, LASSO, Group Lasso, Ridge.

Vast literature on variable selection (a.k.a. sparsifying)

⇒ Frequentist: forward/backward stepwise selection.

Issue: What stopping criterion?

⇒ Bayesian: Priors forcing irrelevant coefficients to zero.

Issue: Confusion of inference and utility?

⇒ Penalized likelihood: LARS, LASSO, Group Lasso, Ridge.

Issue: What penalty parameter (λ)?

A new paradigm

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

A new paradigm

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

- Argues for a two-step approach using a loss function, $\mathcal{L}(\gamma) = f(\gamma, \theta) + \text{penalty}(\gamma)$. γ is the "action."

A new paradigm

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

- Argues for a two-step approach using a loss function, $\mathcal{L}(\gamma) = f(\gamma, \theta) + \text{penalty}(\gamma)$. γ is the "action."

1. Characterize uncertainty in the problem.

A new paradigm

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

- Argues for a two-step approach using a loss function, $\mathcal{L}(\gamma) = f(\gamma, \theta) + \text{penalty}(\gamma)$. γ is the "action."

1. Characterize uncertainty in the problem.
2. Optimize $\mathcal{L}(\gamma)$ integrated over this uncertainty.

Passive Investing

The setup

- Action: Portfolio weights w.
- Define loss by "portfolio variance - portfolio mean + penalty".
- *Goal:* find a sparse representation of w while simultaneously maximizing Sharpe ratio.

A loss function for the mean-variance investor

Given future asset returns $\tilde{R}$, we define a loss function balancing Sharpe ratio and portfolio simplicity.

$$\mathcal{L}(w, \tilde{R}) = -\sum_{k=1}^{N} w_k \tilde{R}_k + \frac{1}{2} \sum_{k=1}^{N} \sum_{j=1}^{N} w_k w_j \tilde{R}_k \tilde{R}_j + \lambda \|w\|_1$$

We are in search of the highest Sharpe ratio, simplest portfolios!

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 LASSO

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

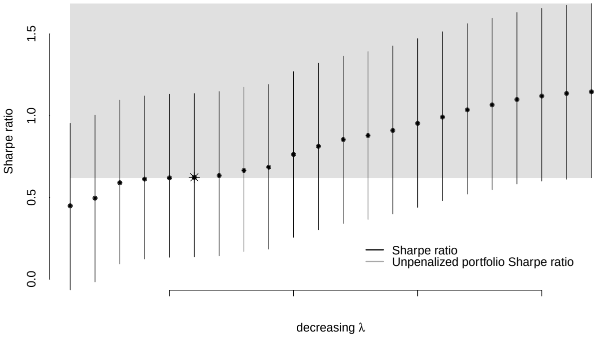
$$\begin{aligned}\mathcal{L}(w) &= -w^T\overline{\mu} + \frac{1}{2}w^T\overline{\Sigma}w + \lambda \left\|w\right\|_1 \\ &= \frac{1}{2}\left\|L^T w - L^{-1}\overline{\mu}\right\|_2^2 + \lambda \left\|w\right\|_1 .\end{aligned}$$

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?

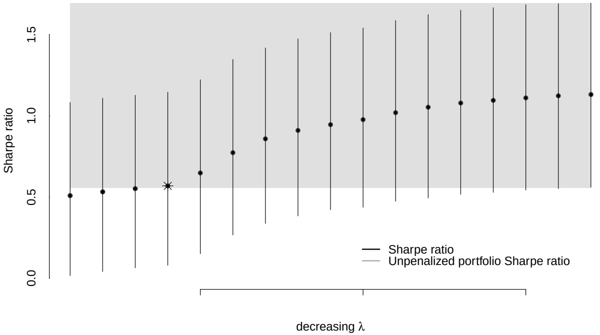
Posterior summary plot



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

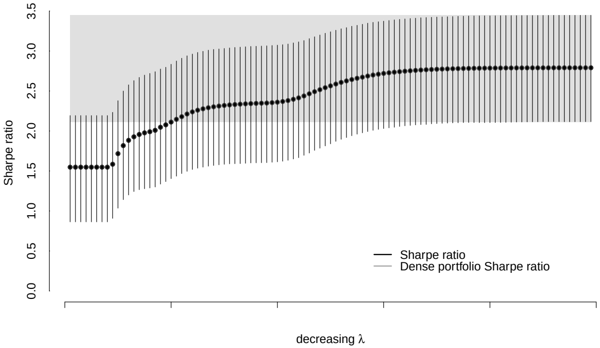
Find the smallest portfolio such that with probability 99% I give up less than (blank) in Sharpe ratio.

Including a market ETF



ETF	SPY	EWJ	IWO	IYR	EWY
weight	108.2	-22.9	-2.6	17.2	0.1
style	*market*	*Japan*	*small growth*	*real estate*	*South Korea*

Choosing among 100 stocks



ETF	AIG	WDC	YRCW	ESCR	ASEI	ARTW
weight	467.7	-51.8	-525.5	-92.7	270.8	31.6

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)

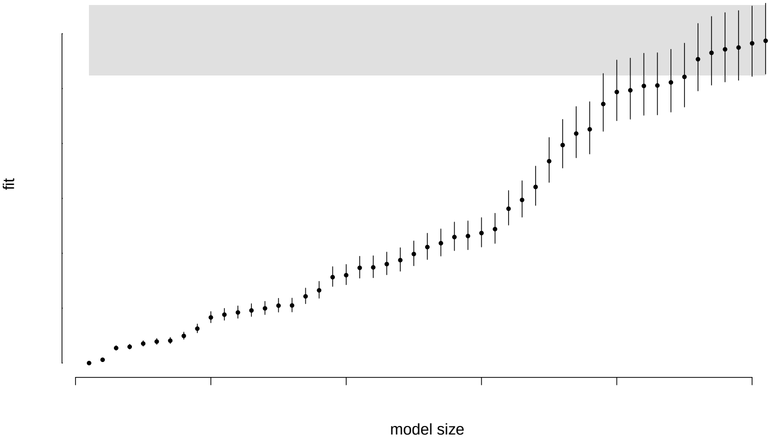
- Market
- Size
- Value
- Momentum
- Short and long term reversal
- Betting against β
- Direct profitability
- Dividend initiation
- Carry trade
- Liquidity
- Quality minus Junk
- Investment
- Leverage
- ...

A loss function for determining important factors

- Test assets: R, Factors: F.
- Define loss by conditional likelihood, $R|F \sim N(\gamma F, D^{-1})$.
- *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 LASSO loss function!

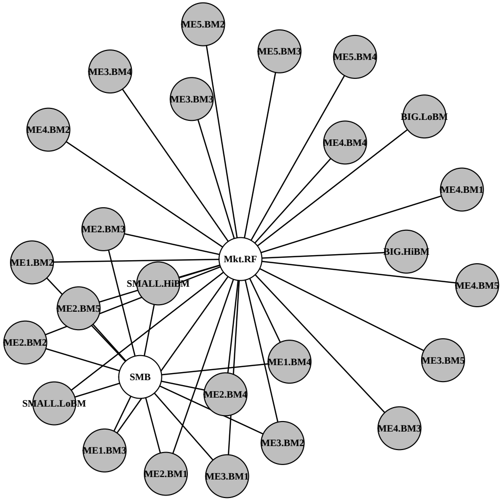
Posterior summary plot



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

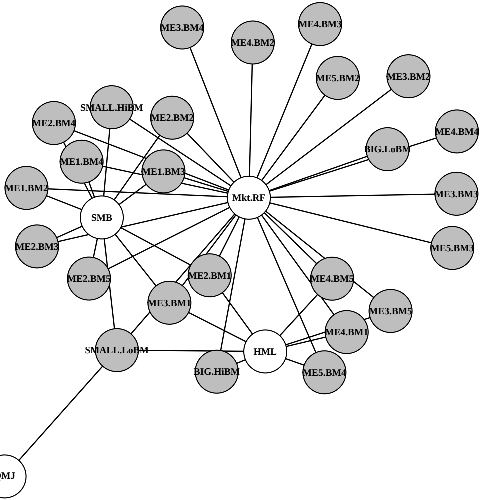
Factor selection graph

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



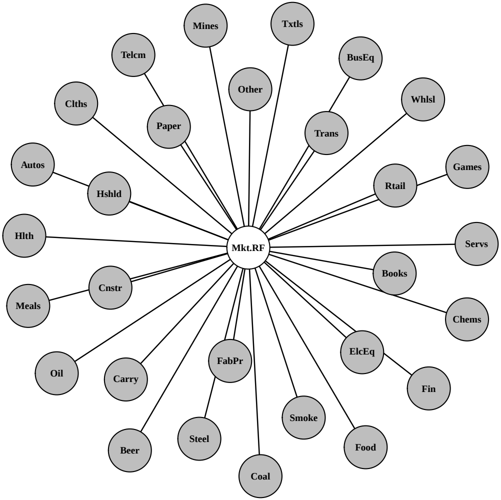
Factor selection graph

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



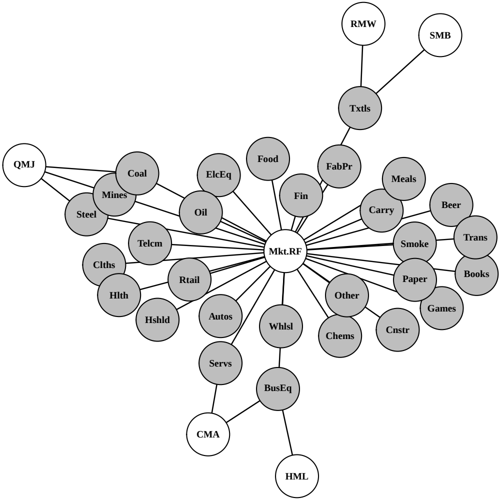
Factor selection graph

R: 30 Industry Portfolios, F: 10 factors, strong shrinkage



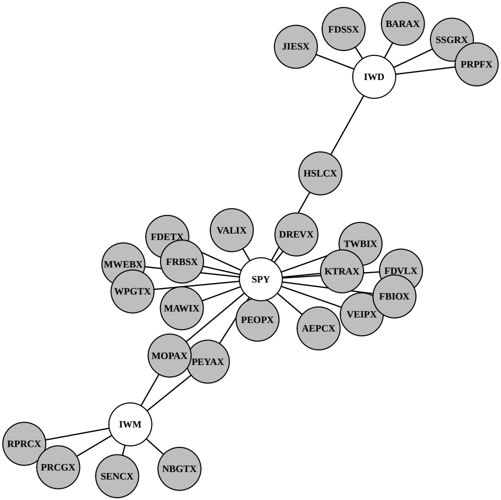
Factor selection graph

R: 30 Industry Portfolios, F: 10 factors, weak shrinkage



Another application: ETF selection

R: 100 Mutual funds, F: 25 ETFs, strong shrinkage



Concluding thoughts

- Passive investing and factor selection for asset pricing models approached using new *DSS* 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!