UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

MFIN 290: Financial Econometrics

Lecture 8-1

This time

- Bridging the gap into Artificial Intelligence

- LASSO Regression

The setup

- We have a set of data X that we want to use to describe the variation in some response variable Y. We suspect that some of the features in X are related to Y, but don't know which ones or how…
- We want to pick between specifications
- $Y = X_1\beta + e$
- $Y = X_2\gamma + u$
- $Y = X_3\delta + v$
- For some $X_1, X_2, \ldots X_n$ that are subsets of X

The problem

- No easy way to do this. We can compare the Adjusted R-squared, we can look at what is significant or not, we can look at performance in hold out samples, we can start small and add things, we can start big and subtract, we can look at all of the models with N regressors.

- All have weaknesses.. Usually around repeat inference or path dependence of the addition/removal process

Example

Regressors can be (highly) correlated, and the point estimates can pull against one another.

Remember, this regression just on the market gave us a beta of ~1.2, which is basically the sum of these two effects

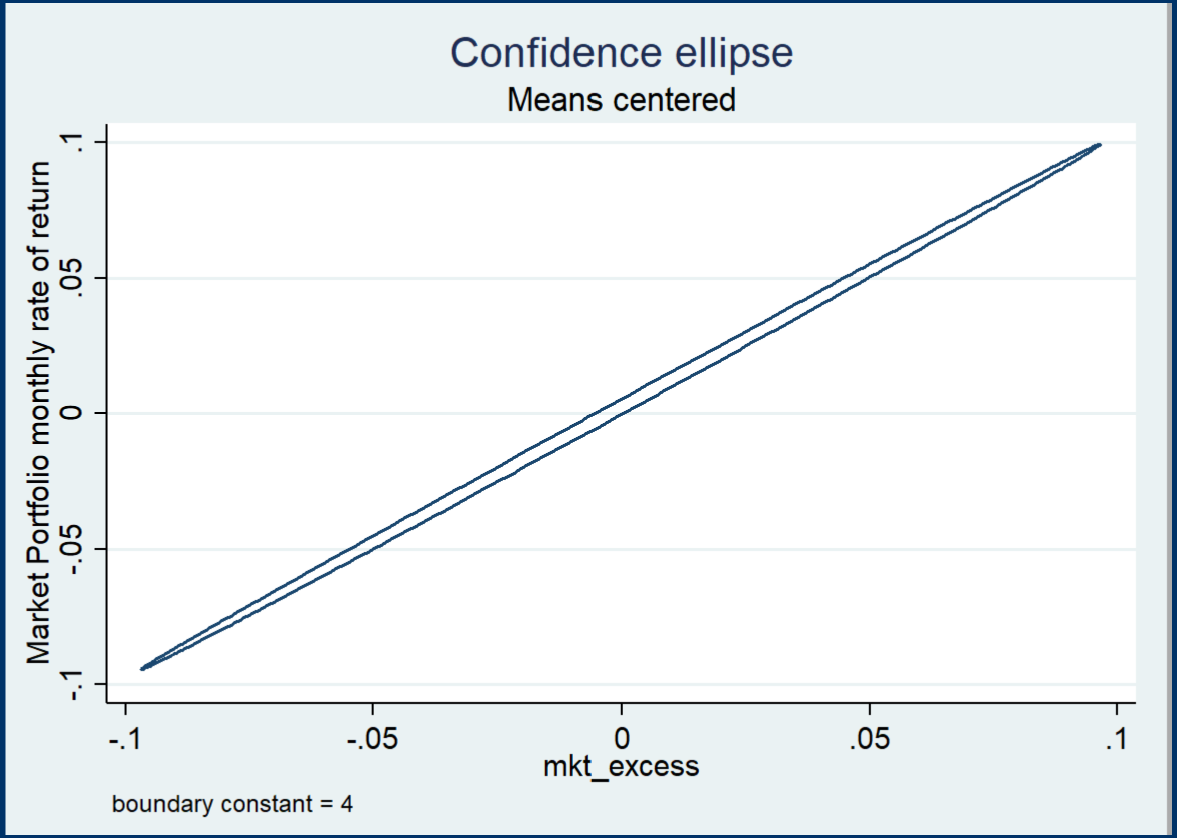
```
. corr mkt mkt_excess
(obs=132)

             |      mkt mkt_ex~s
-------------+------------------
         mkt |   1.0000
  mkt_excess |   0.9996   1.0000
```

```
. gen mkt_excess = mkt - riskfree

. reg ibm mkt mkt_excess

      Source |       SS           df       MS      Number of obs   =       132
-------------+----------------------------------   F(2, 129)       =     44.51
       Model |  .437939033         2  .218969517   Prob > F        =    0.0000
    Residual |  .634621055       129  .004919543   R-squared       =    0.4083
-------------+----------------------------------   Adj R-squared   =    0.3991
       Total |  1.07256009       131  .008187482   Root MSE        =    .07014

------------------------------------------------------------------------------
         ibm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mkt |   3.814218   4.306652     0.89   0.377    -4.706598    12.33503
  mkt_excess |   -2.63178   4.317342    -0.61   0.543    -11.17375    5.910187
       _cons |  -.0016786   .0130404    -0.13   0.898    -.0274793    .0241221
------------------------------------------------------------------------------
```

Confidence ellipses

A higher dimensional confidence region can show this (harder with even more than two!)

But that doesn't really help us here – most of these issues come from repeat inference. Even if I knew that I shouldn't use IG and HY spread together in this case for SURE, which one I remove first will determine the subsequent path I take.



LASSO

- “Least absolute shrinkage and selection operator”
- Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the lasso".
- Sometimes referred to as “Sparse Regression” (we will see why shortly)

LASSO Intuition

- Large coefficients pull you in opposing directions and don't add a lot of value

- What is Large? Coefficient size is a function of variance in x and y…
  - LASSO works on standardized data. Some (not all) packages do this automatically.

$$x^* = \frac{x - \bar{x}}{\sigma_x}$$
$$y^* = \frac{y - \bar{y}}{\sigma_y}$$

LASSO Objective

- Instead of choosing beta to minimize the sum of squared errors, we choose beta to minimize a different objective
- $\min e'e = (y - x\beta)'(y - x\beta)$ to
- $\min (y - x\beta)'(y - x\beta) + \boldsymbol{\lambda 1}'|\beta|$
- We add to the objective function the weighted (λ) sum of the absolute value of the coefficients

LASSO Objective

- $\min (y - x\beta)'(y - x\beta) \ + \boldsymbol{\lambda 1}'|\beta|$
- We add to the objective function the weighted (λ) sum of the absolute value of the coefficients
- It's very costly now to have a big positive coefficient offset by a big negative coefficient..
- In fact, what will the optimal betas look like compared to the OLS betas?

LASSO Solution

- Absolute values are not the best things to have in an objective function since they have equal derivatives everywhere. We can't use our Calc 101 tools to find the answer here reliably.
- Use different algorithms here (Least Angle Regression, Shooting Algorithms)

LASSO Mechanics

- $\hat{\beta}_{ols} = [\beta_1, \beta_2, \dots, \beta_k]$

- $\hat{\beta}_{lasso} = [b_1\ , b_2, \dots, b_k]$
  What's going to be the difference?
- Absolute value weighting will push some to zero (which ones depend on how large the unconstrained betas are and how large the lambda is)
- Will "shrink" the others towards zero

LASSO as variable selection

- The first effect acts as a variable selection mechanism. Some of the candidate variables will be set to zero in lasso (how many will vary with lambda), and I can then run an OLS specification on these.

- We almost never want to use the lasso betas directly, but the post-lasso coefficients run on the selected variables.

How to choose lambda

- Cross validation

- This works, but should narrow the neighborhood rather than be decisive.

How to choose lambda

- Objective function modification

- Square root lasso
- Belloni, Chernozhukov, Wang (2010)

- Instead of minimizing the sum of squares, we minimize the square root of the sum of squares.
- The derivatives of this have nice properties, and lets us assign a value of lambda as a function of things that we know (number of observations, number of candidate variables, etc).

How to choose lambda

- Thought process:

1. How does the set of selected coefficients change as I tighten the constraint on the sum of the absolute value of the betas?
2. Which set of variables do we have strong priors about?
3. When do I start to lose performance in a meaningful way (this is what CV does, but we can be more precise about "performance" if we think this through)
4. Sometimes there might be a maximum number of variables you can tolerate…

LASSO example

- Unlike a few things in this course, this is relatively new and is still being implemented and tinkered with.

- It actually is in the most recent versions of Stata, but not the one on my machine. But it IS in MATLAB. That's a bit more awkward syntax wise (python is best here IMO but I won't spring that on you halfway through the course).

- But this is a **great** example of the need to understand the algorithms and verify what is coming out of the tool is what you need…. This comes up all the time with newer tools and techniques (like ML)

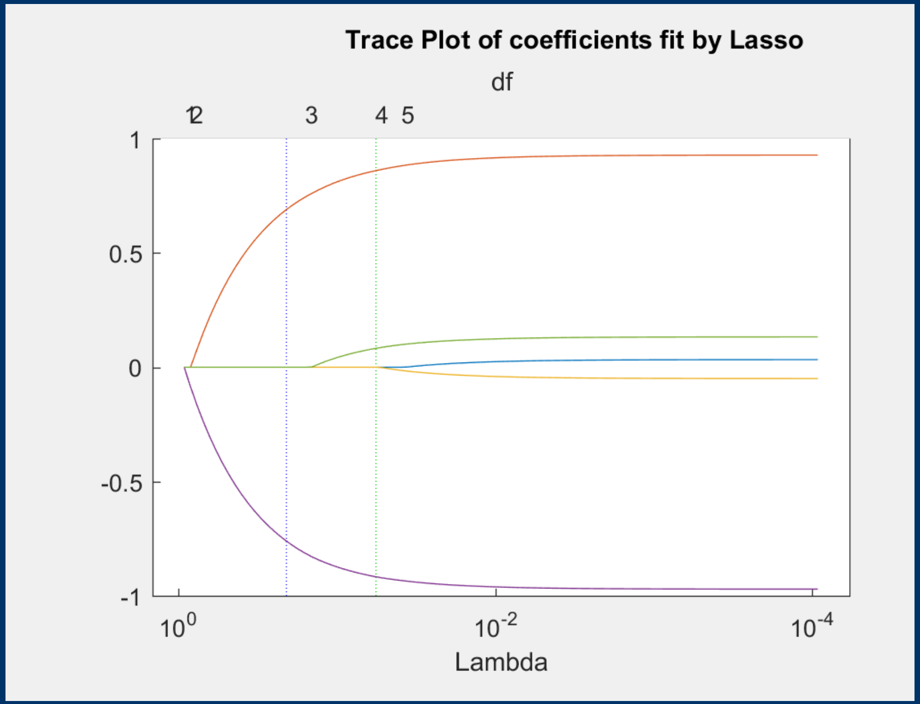
LASSO example

```
clear

rng(12345,'twister') % set the random number seed - code is reproducable
X = zeros(200,5);
for i = 1:5
    X(:,i) = mvnrnd(0,1,200);
end

coef = [0;1;0;-1;0];
u = mvnrnd(0,1,200);
Y = X*coef + u;

%fit a lasso, choose lambda through 10 fold cross validation
[b fitinfo] = lasso(X,Y,'CV',10)
```

LASSO example

```
%plot the betas against the lambda
lassoPlot(b,fitinfo,'PlotType','Lambda','XScale','log')
```

You would look at this picture and pick two betas



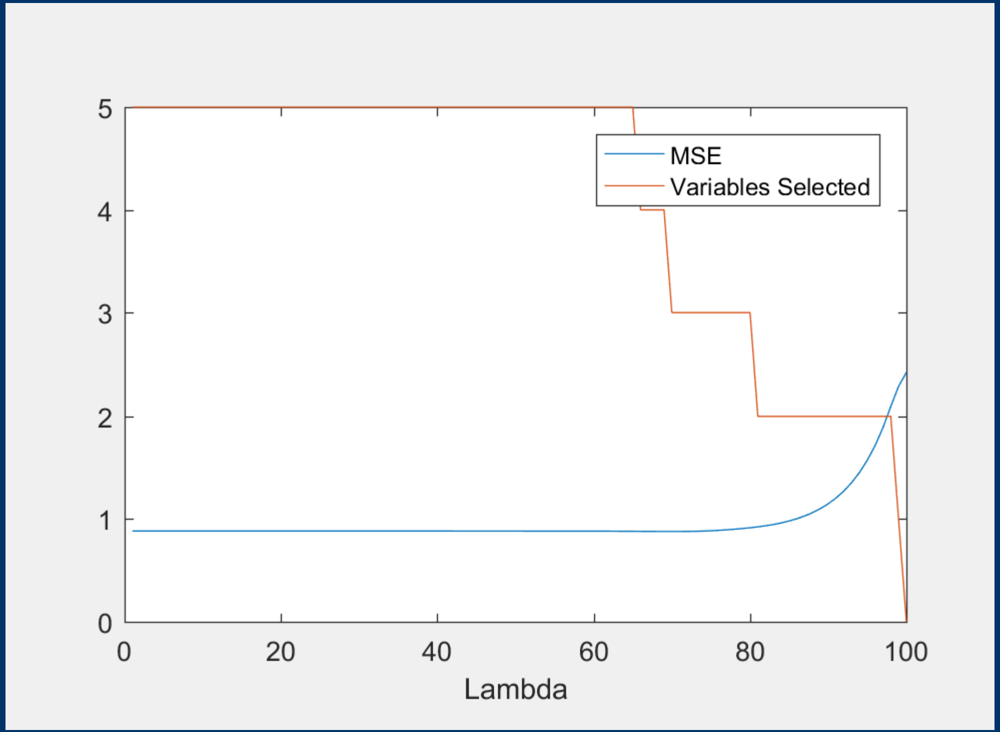
LASSO example

```
%which lambda minimized MSE?
fitinfo.Lambda(fitinfo.IndexMinMSE)
fitinfo.LambdaMinMSE

%how many variables did that pick?
fitinfo.DF(fitinfo.IndexMinMSE)
%what are the coefficients?
%Are they constrained or post-lasso?
b(:,fitinfo.IndexMinMSE)
```

```
ans =

    0.0566


ans =

    0.0566


ans =

     3


ans =

         0
    0.8599
         0
   -0.9168
    0.0825
```

LASSO example

```
plot([fitinfo.MSE',fitinfo.DF'])
legend('MSE','Variables Selected')
xlabel('Lambda’) % the index
```



Would you really pick 3 variables here???

If you looked at the MSE tradeoff, you would pick 2

LASSO example

```
%seems to flatten out fast after we are down to 2, and that's consistent
%with the betas vs lambda plot
b(:,81)

%This is low. could be the CONSTRAINED BETA (not Post-lasso)?
 % Let's find the post lasso value by hand:

X_new = X(:,[2 4]);

beta_post= inv(X_new'*X_new)*X_new'*Y
%this is the post-LASSO beta. Very close to the truth
```

```
ans =

         0
    0.7454
         0
   -0.8148
         0


beta_post =

    0.9156
   -0.9812
```

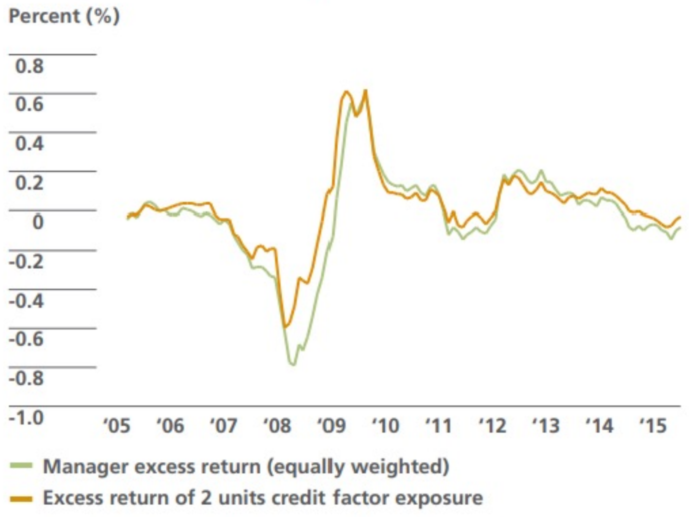
LASSO Applications

- Finding risk factor weights
- Questions:

1. Discovering manager styles
2. What do you get for active fees?
3. Hedge fund strategies
4. How do people generate “alpha”? Am I getting what the prospectus says?

LASSO Applications

Figure 5. Average fixed income fund excess returns (over benchmark) vs. two years of passive credit factor exposure (rolling 12-months, January 2005 – November 2015)



Source: PIMCO, Morningstar, Barclays and Bloomberg.
As of 30 November 2015

Most managers use their tracking error budget to go slightly risk on.

You can replicate this with passive indices and save the fees.

You can't diversify that pattern away by adding managers

Can help you determine how to construct a portfolio

Building your own model

- 1) Select risk factors:

Equities
- US Equities
- DM Equities
- EM Equities
- Value vs Growth?
- Sectors?

Bonds
- 10Y Yields
- 2Y Yields (slope)
- IG Spread
- HY Spread
- EM Spread
- Sovereign Spreads
- MBS/Structured?

Commodities
- Oil
- Metals
- Ag

Other
- Trend Following
- Carry
- Value
- Momentum
- Profitability
- Currency (DM, EM)

Building your own model

- 2) Run LASSO of fund returns on your universe

Questions:

Do we have any reason to think things changed in this sample?

Should we do rolling applications to check?

Building your own model

- 3) Extract selected factors

- Have these changed over time? Do we like how aggressive we are setting that lambda?

Building your own model

- 4) Run post lasso regressions on de-standardized data, plot coefficients over time

- Profit!