Can Machine Learning Techniques Predict Non-performance of Farm Non-Real Estate Loans in the Ag Finance Databook

"Mindy Mallory, Todd Kuethe, and Todd Hubbs"
"University of Illinois"

April 6, 2018

Outline of Talk

- Basic motivation of why it is intersting/important to predict non-performing loans in the farm sector. If there's gonna be another 1980's, want to see it coming
- Problem is inherently 'high dimentional' few observations, many (probably colinear) possible predictors
- Can the shiny new toy (machine learning techniques) help with this problem?
- Discussion of basic similarities and differences between ML and typical econometrics
- Introduce (in spirit) the ML models test run in this talk
- Results individually
- Results compare across
- Not clear what if any could be an econometric straw man? What would be the most natural traditional linear model to put these up against?

Research Question

- Try to predict farm non-performing loans in the Agricultural Finance Databook
- Selected ‘A’ and ‘B’ tables

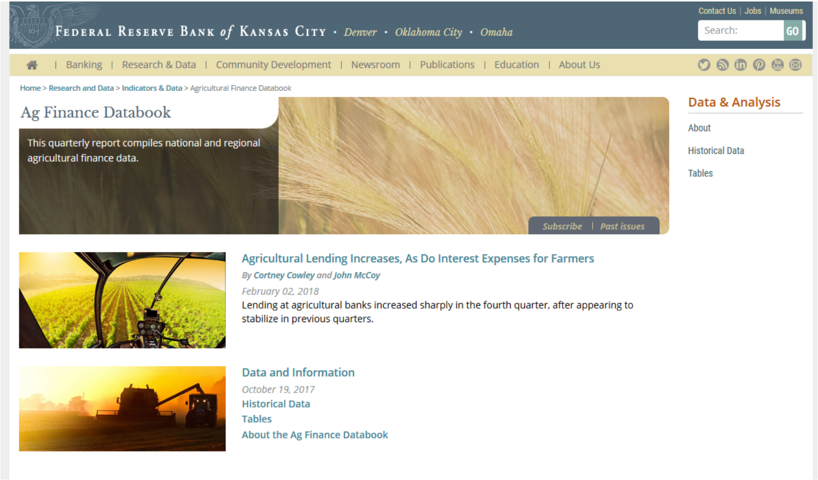


Figure 1

Research Question

- Try to predict farm non-performing loans in the Agricultural Finance Databook

Why??

Research Question

- Try to predict farm non-performing loans in the Agricultural Finance Databook

Why??

- See ‘the next 1980’s crisis coming’
- Even if not 1980’s crisis levels $\Rightarrow$ Still good to predict what non-performance will look like next quarter.

Research Question

- Try to predict farm non-performing loans in the Agricultural Finance Databook

Why??

- Less access to bank-level data
- Can machine learning techniques help us make sense of aggregate-level data?
- Can machine learning techniques help us handle high-dimentional data?
- Like in the Ag Finance Databook

Historical Coverage: The 'A' Tables

Section A of the Agricultural Finance Databook are from quarterly sample surveys for non-real-estate farm loans of $1,000 or more made by commercial banks. a1-a7 'attempt to show estimates that are comparable to those that have been presented for a number of years.'

A table	Range
A1-A6	*1977- annual and quarters*
A7	1987- quarters (interest rates - lots of NA's before 1990's)
A8-A13	1999Q4- quarters

Historical Coverage: The 'B' Tables

Section B of the AFD are quarterly reports of condition and income for commercial banks.

B table	Range
*B1	1973- Q4, 1986-quarterly*
B2	*1987- quarterly*
B3	1984- annual
B4	1991- annual and quarters
B5	1991- annual and quarters
B6	1987- quarterly*
B7	1988- quarterly, 1977- annual
B8	1975- quarterly
B9	1981- quarterly

Historical Coverage: The 'C' Tables

Section C of the AFD are quarterly surveys of agricultural credit conditions and farmland values at commercial banks.

C table	Range
C1	1991-quarterly
C2	1991-quarterly
C3	1991-quarterly
C4	1991-quarterly
C5	2001- quarterly
C6	1991-quarterly
C7	2001- quarterly

This Study: Selected Variables from the A and B Tables

A1-A6
Feeder livestock
Other livestock Other current operating expenses
Farm machinery and equipment
Other
1 to 9
10 to 24
25 to 99
100 and over
Small or mid-size
Large

This Study: Selected Variables from the A and B Tables

A.1. *Number* of Non-Real Estate Bank Loans
A.2. *Average Size* of Non-Real Estate Bank Loans
A.3. *Volume* of Non-Real Estate Bank Loans
A.4. *Average Maturit*y of Non-Real Estate Bank Loans
A.5. *Average Effective Interest Rate* on Non-Real Estate Bank Loans A.6. *Share* of Non-Real Estate Bank Loans *with a Floating Interest Rate*

This Study: Selected Variables from the A and B Tables

B1 Loan volumes

This Study: Selected Variables from the A and B Tables

B2
Estimated volume - Total
Estimated volume - Past due 30 to 89 days, accruing
Estimated volume - Nonperforming - Total
Estimated volume - Nonperforming - Past due 90 days, accruing
Estimated volume - Nonperforming - Non-accruing
Share of outstanding loans - Total
Share of outstanding loans - Past due 30 to 89 days, accuring
Share of outstanding loans - Nonperforming - Total
Share of outstanding loans - Nonperforming - Past due 90 days, accruing
Share of outstanding loans - Nonperforming - Non-accruing

This Study: Selected Variables from the A and B Tables

B6 - Interst rates

Potential independent variables

Share of outstanding loans Non-performing total

Share of outstanding loans Non-performing, 90 days past due, accrueing

Share of outstanding loans Non-performing, non-accrueing

How many Lags?

1

High Dimentional Problem

- 122 observations, and 88 potential ‘predictors’. Many are most probably highly colinear.

Econometrics versus Machine Learning

- ML more flexible on model specification
- Use cross validation to keep honest wrt overfitting

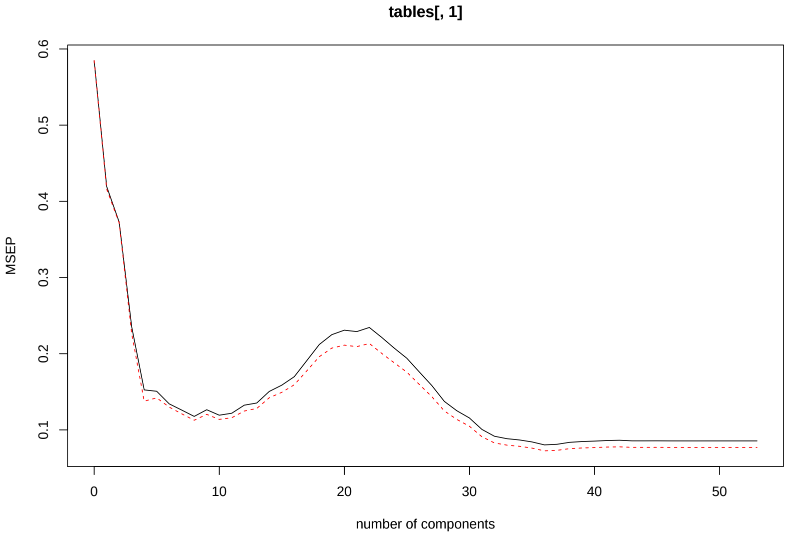
Econometrics versus Machine Learning: Variable Selection

- ML has variable selection baked into most procedures
- Econometrics strategies to reduce high dimentionality
- Theory
- Make and index
- Ad hoc
- General-to-Specific
- Specific-to-General

Partial Least Squares

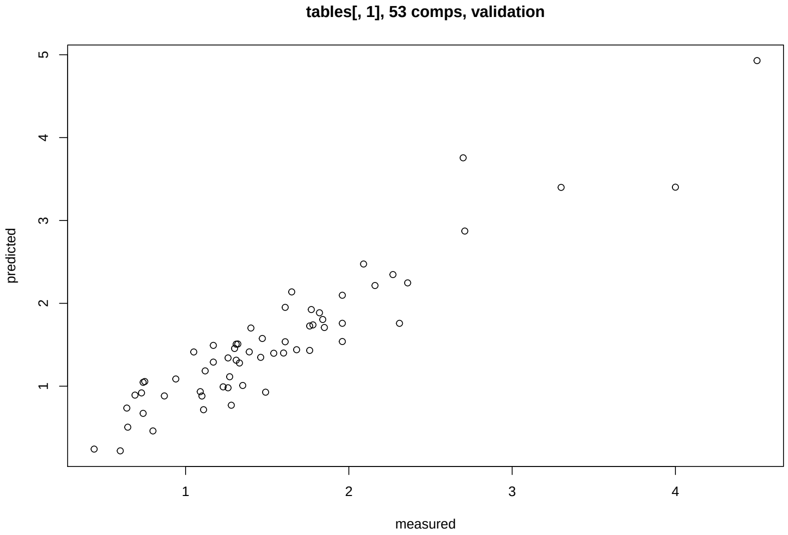
1. Fit OLS $Y = \beta^1 X + \epsilon^1$
2. Set $Z_1 = \beta^1 X$
3. Fit OLS $\epsilon^1 = \beta^2 X + \epsilon^2$
4. Set $Z_2 = \beta^2 X$ . . .
5. Fit OLS $Y = \gamma Z + \eta$

PLS Results

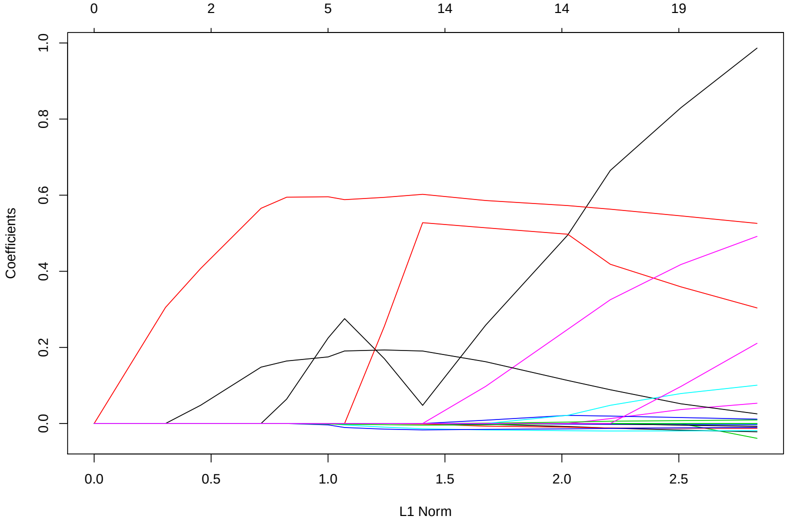


tables[, 1]
MSEP
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40
50
number of components

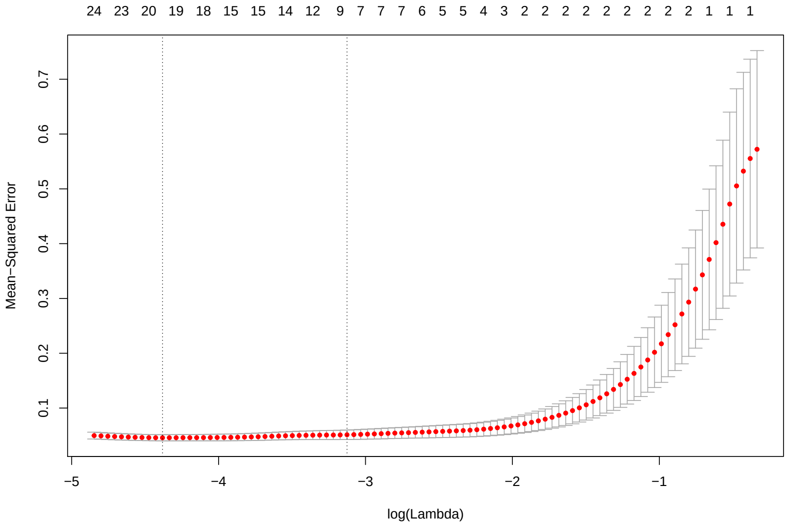
PLS Results


tables[, 1], 53 comps, validation
predicted
measured
1
2
3
4
5
1
2
3
4

Lasso Regression Results



Lasso Regression Results



Lasso Regression Results

```
## 89 x 1 sparse Matrix of class "dgCMatrix"
##
## (Intercept)
## Feeder livestock.x
## Other livestock.x
## Other current operating expenses.x
## Farm machinery and equipment.x
## Other.x
## 1 to 9.x
## 10 to 24.x
## 25 to 99.x
## 100 and over.x
## Small or mid-size.x
## Large.x
## All loans.x
## Feeder livestock.y
## Other livestock.y
## Other current operating expenses.y
```

Boosting Regression Trees

```
## [1] 1.411112
```