

# 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 interesting/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 dimensional' 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

The screenshot shows the website of the Federal Reserve Bank of Kansas City, specifically the Agricultural Finance Databook section. The header includes the bank's name and locations (Denver, Oklahoma City, Omaha), a search bar, and navigation links. The main content area features a large banner for the 'Ag Finance Databook' with a description and a 'Subscribe' button. Below this, there are two article previews: one about agricultural lending increases from February 2018, and another about data and information from October 2017. A sidebar on the right lists 'Data & Analysis' options like 'About', 'Historical Data', and 'Tables'.

FEDERAL RESERVE BANK of KANSAS CITY • Denver • Oklahoma City • Omaha

Contact Us | Jobs | Museums

Search:  GO

Banking | Research & Data | Community Development | Newsroom | Publications | Education | About Us

Home > Research and Data > Indicators & Data > Agricultural Finance Databook

## Ag Finance Databook

This quarterly report compiles national and regional agricultural finance data.

[Subscribe](#) | [Past issues](#)

### Agricultural Lending Increases, As Do Interest Expenses for Farmers

By *Cortney Cowley and John McCoy*  
February 02, 2018

Lending at agricultural banks increased sharply in the fourth quarter, after appearing to stabilize in previous quarters.

### Data and Information

October 19, 2017

[Historical Data](#)  
[Tables](#)  
[About the Ag Finance Databook](#)

## Data & Analysis

[About](#)  
[Historical Data](#)  
[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-dimensional 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	<i>1977- annual and quarters</i>
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	<i>1987- quarterly</i>
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 Maturity* 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, accruing

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 - Interest rates

## Potential independent variables

*Share of outstanding loans Non-performing total*

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

Share of outstanding loans Non-performing, non-accruing

How many Lags?

1



## High Dimensional 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

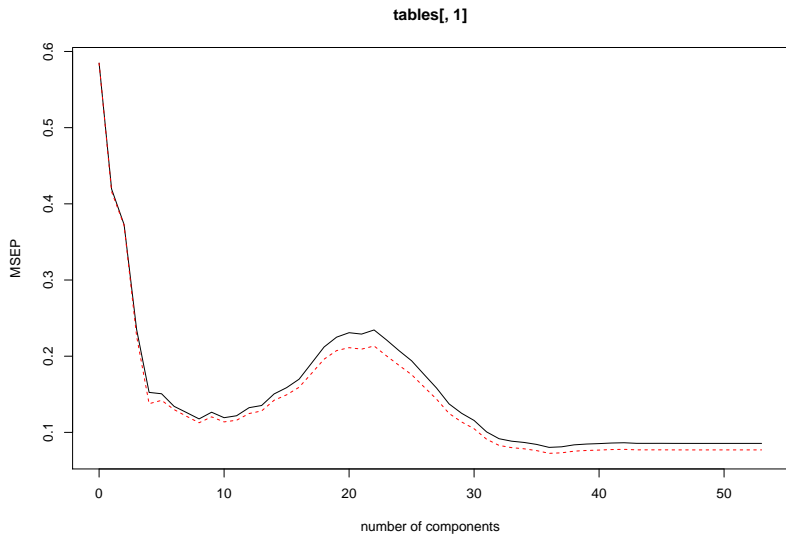
# Econometrics versus Machine Learning: Variable Selection

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

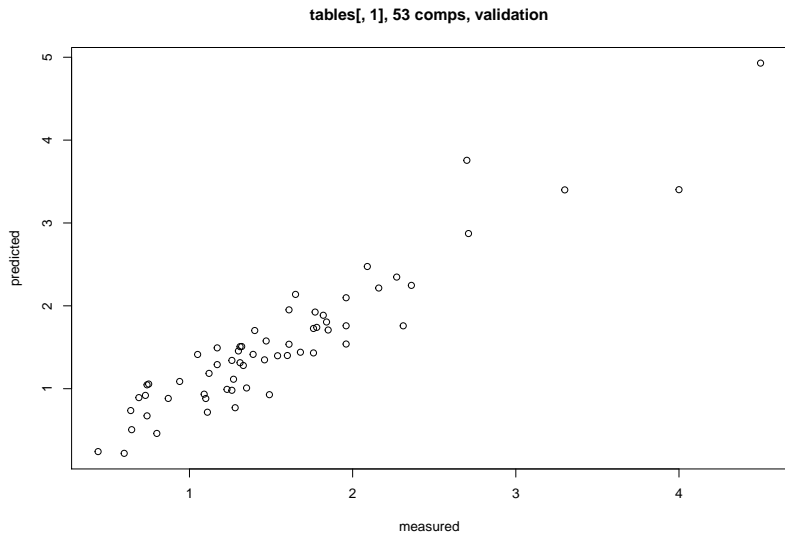
# Partial Least Squares

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  $Z_2 = \beta^2 X \dots$
5. Fit OLS  $Y = \gamma Z + \eta$

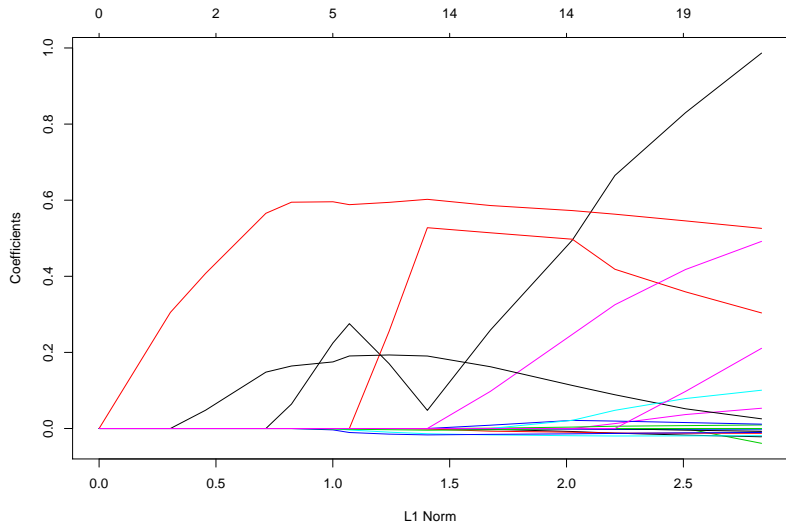
# PLS Results



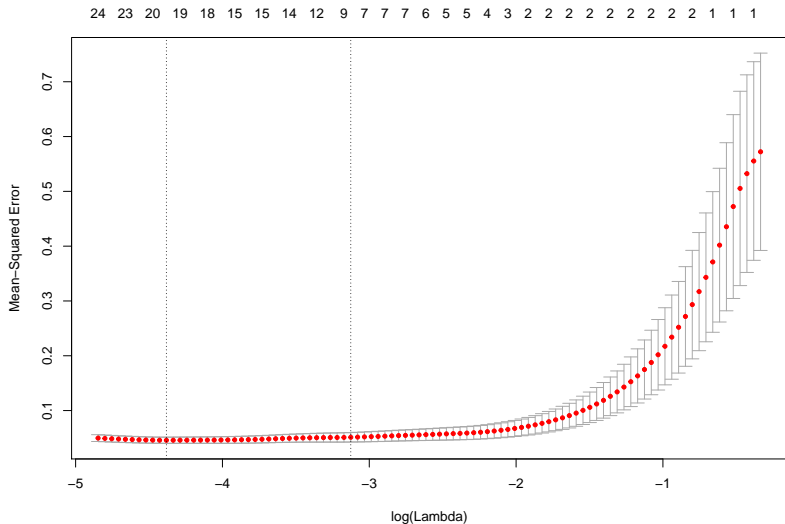
# PLS Results



# 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
```