Machine Learning and Applications in Finance and Macroeconomics

Discussion by
Amit Seru
University of Chicago

Macro Financial Modeling
2016 Winter Meeting
Overview

• Machine Learning
  – Mortgage delinquency risk
  – Credit card delinquency risk
  – Matching datasets

• Comments
  – Where might we need this…
  – Predictions in a changing world…
Where might we need this?

• Better risk assessment
  – Investments
  – Regulation/Capital charges

• Regulation/Supervision
  – At what frequency are we running these models?
    ➢ Every supervision cycle?
    ➢ Every stress test?
    ➢ Does it matter if we can predict better over a short horizon at high frequencies?

  – Where did we need these predictions in supervision and regulation?
• How well do we do when the market conditions change?
  – Train for 12 years (1999-2011) and test for 2012-2014
Machine Learning: Interpreting the “Black Box”

- Combine several observable factors to “improve” predictions
  - How do we interpret the results from the black box?
    ➢ Would matter for what policy intervention is designed.
Credit Card Defaults Paper

• Why is risk management different across banks?
  – Selection of consumers different. (Citi v/s Capital One)
  – Selection across regions different (Citi urban v/s BofA rural)
  – Rewards program different (gas, airlines, stores, etc etc)
  – Risk management practices different -- some estimate at the account level; others at the portfolio level.
  – Attrition rates could be different
Machine Learning: Interpreting the “Black Box”

• Combine several observable factors to “improve” predictions
  – How do we interpret the black box?

• Not modeling primitives $\Rightarrow$ could potentially miss key quantitatively important aspects. Would this matter?
  – Incentives of agents that influence data generating process
    $\Rightarrow$ Meaning of observables can change over time
### Incentives: Reliance on Hard Information

In the context of financial markets, hard information refers to data that is verifiable and objective, whereas soft information is subjective and can be influenced by factors such as market sentiment. The relationship between the rate of return ($r_t$) and hard information indicators is often modeled using a linear regression framework:

$$r_t = \alpha + \beta_{FICO} \times FICO_t + \beta_{LTV} \times LTV_t + \epsilon_t.$$ 

<table>
<thead>
<tr>
<th>Year</th>
<th>$\beta_{FICO}$</th>
<th>$\beta_{LTV}$</th>
<th>$R^2$ (in %)</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>-0.004***</td>
<td>0.030***</td>
<td>3</td>
<td>24067</td>
</tr>
<tr>
<td></td>
<td>(.0002)</td>
<td>(.0013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>-0.007***</td>
<td>0.035***</td>
<td>7</td>
<td>60094</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>-0.007***</td>
<td>0.020***</td>
<td>8</td>
<td>104847</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>-0.010***</td>
<td>0.035***</td>
<td>14</td>
<td>116778</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>-0.012***</td>
<td>0.038***</td>
<td>20</td>
<td>136483</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>-0.011***</td>
<td>0.071***</td>
<td>18</td>
<td>162501</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>-0.012***</td>
<td>0.079***</td>
<td>32</td>
<td>318866</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>-0.010***</td>
<td>0.097***</td>
<td>40</td>
<td>610753</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>-0.009***</td>
<td>0.110***</td>
<td>48</td>
<td>793725</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>-0.011***</td>
<td>0.115***</td>
<td>50</td>
<td>614820</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.0001)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Dramatic increase in the $R^2$:** About 3% in 1997, goes up to almost 50%.
  - $\beta_{FICO} < 0$, $\beta_{LTV} > 0$.
- In the low securitization regime, the hard information variables explain very little variation in interest rates.
  - Omitted variables are particularly important.
  - Soft information, by its very nature, is one of the omitted variables.
Incentives: Actual versus Predicted Defaults
Machine Learning: Interpreting the “Black Box”

- Combine several observable factors to “improve” predictions
  - How do we interpret the black box?

- Not modeling primitives ➔ could potentially miss key quantitatively important aspects. Would this matter?
  - Incentives of agents that influence data generating process
    ➔ Meaning of observables can change over time

Information Disclosed

Second Lien

First Mortgage

Equity

Identifying Criteria

Information Disclosed
- Loan \( i \) has NO second lien
- \( CLTV=LTV \)

Information in Equifax
- There is second lien on property at same time

Actual Information

First Mortgage

Second Lien

True Equity

Lower “Financial Equity”
Truthfully Reported No Second Lien
DGP

Quarters from Origination

Truthfully Reported No Second Lien

Misreported Second Lien
DGP

Quarters from Origination

Truthfully Reported No Second Lien

Misreported Second Lien

Truthfully Reported Second Lien
Machine Learning: Interpreting the “Black Box”

- Combine several observable factors to “improve” predictions

- Not modeling primitives ➔ could potentially miss key quantitatively important aspects
  - Policy interventions at times when “value of prediction” the most
    - ➢ Foreclosure behavior different before and after HAMP
    - ➢ Foreclosure behavior different before and after HARP
    ➔ accounting for timing and eligibility of some borrowers important

- Institutional factors could change data generation
  - ➢ Different incentives to foreclose depending on ownership status
  - ➢ Organizational ability of servicers to pass through government subsidies
Conclusion

• Nice and interesting set of papers
  – Machine learning can help improve predictions
  – Better data would be better

• Going forward
  – What are we using these models for?
    ➢ Regulation/Supervision/Investments?
  – Do these models do better on the Lucas critique?
    ➢ Incentives/Institutional Factors/Government Interventions
  – How would one interpret the model with…
    ➢ …Other agencies also fitting same data (FICO, Zillow, TransUnion)
    ➢ …Human capital/political constraints in supervision