Consumer Credit Risk Management

- $3T of consumer credit outstanding as of 8/13
- $840B of it is revolving consumer credit
- Average credit card debt as of 10/13: $15,159
- 46.7% of households carry positive credit card balance as of 12/12
- Current "charge-off" rates are 6.7% (2013Q2), but reached 10.2% in 2010Q1

⇒ Can We Predict These Credit Cycles?
Standard Credit Scores Are Too Insensitive

MIT Laboratory for Financial Engineering:

Tackling the Challenges of Big Data
Big Data Analytics
Applications: Finance
The Challenge of Consumer Credit Risk Management

THANK YOU
Tackling the Challenges of Big Data
Big Data Analytics

Andrew W. Lo
Charles E. and Susan T. Harris Professor
Massachusetts Institute of Technology

© 2014 Massachusetts Institute of Technology

Tackling the Challenges of Big Data
Big Data Analytics
Applications: Finance
Big Data for Consumer Credit

Andrew W. Lo
Charles E. and Susan T. Harris Professor
Massachusetts Institute of Technology

© 2014 Massachusetts Institute of Technology

Anonymized Data from Large U.S. Commercial Bank

Transaction Data

Credit Bureau Data
1% Sample = 10 Tb!

© 2014 Massachusetts Institute of Technology
Extract “Interesting” Features

Tackling the Challenges of Big Data
Big Data Analytics
Applications: Finance
Big Data for Consumer Credit

THANK YOU

Tackling the Challenges of Big Data
Big Data Analytics

Andrew W. Lo
Charles E. and Susan T. Harris Professor
Massachusetts Institute of Technology
Objectives

- For consumer $j$ with characteristics or "features" $X_j$, estimate probability of default or delinquency $P(X_j)$.
- Characteristics include:
  - Individual characteristics, macro factors, interactions between the two

Machine Learning Techniques

- Decision trees (e.g., CART)
- Logistic regression
- Random forests
- Clustering/segmentation (can be used with other models)
- Software:
  - WEKA (machine-learning suite - University of Waikato, NZ)
    http://www.cs.waikato.ac.nz/ml/weka/
  - LIBLINEAR (National Taiwan University)
    http://www.csie.ntu.edu.tw/~cjlin/liblinear/
Model Evaluation Framework

- Prediction made for probability of going 90+ delinquent for individual credit cards over 3-month horizon
- Using non-overlapping data (in time) to calibrate the model:

### Model Evaluation Framework

<table>
<thead>
<tr>
<th>Training Period</th>
<th>Evaluation Period</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Date</td>
<td>Start Date</td>
<td>End Date</td>
</tr>
<tr>
<td>Jan-08</td>
<td>Feb-08</td>
<td>Apr-08</td>
</tr>
<tr>
<td>Feb-08</td>
<td>Mar-08</td>
<td>May-08</td>
</tr>
<tr>
<td>Mar-08</td>
<td>Apr-08</td>
<td>Jun-08</td>
</tr>
<tr>
<td>Apr-08</td>
<td>May-08</td>
<td>Jun-08</td>
</tr>
<tr>
<td>May-08</td>
<td>Jun-08</td>
<td>Jul-08</td>
</tr>
<tr>
<td>Jun-08</td>
<td>Jul-08</td>
<td>Aug-08</td>
</tr>
<tr>
<td>Jul-08</td>
<td>Aug-08</td>
<td>Sep-08</td>
</tr>
<tr>
<td>Aug-08</td>
<td>Sep-08</td>
<td>Oct-08</td>
</tr>
<tr>
<td>Sep-08</td>
<td>Oct-08</td>
<td>Nov-08</td>
</tr>
<tr>
<td>Oct-08</td>
<td>Nov-08</td>
<td>Dec-08</td>
</tr>
</tbody>
</table>

### Summary Statistics

<table>
<thead>
<tr>
<th>Starting Date</th>
<th>Ending Date</th>
<th>Total Credit Card Count</th>
<th>Customers Going 60 Days Delinquent</th>
<th>Customers MTF Going 60 Days Delinquent</th>
</tr>
</thead>
<tbody>
<tr>
<td>May-08</td>
<td>Jul-08</td>
<td>575,973</td>
<td>12,859</td>
<td>2.4</td>
</tr>
<tr>
<td>Jun-08</td>
<td>Aug-08</td>
<td>646,480</td>
<td>16,172</td>
<td>2.2</td>
</tr>
<tr>
<td>Aug-08</td>
<td>Sep-08</td>
<td>726,285</td>
<td>16,711</td>
<td>2.0</td>
</tr>
<tr>
<td>Sep-08</td>
<td>Oct-08</td>
<td>796,186</td>
<td>17,291</td>
<td>2.1</td>
</tr>
<tr>
<td>Oct-08</td>
<td>Nov-08</td>
<td>861,289</td>
<td>17,871</td>
<td>2.1</td>
</tr>
<tr>
<td>Nov-08</td>
<td>Dec-08</td>
<td>928,423</td>
<td>18,504</td>
<td>2.8</td>
</tr>
<tr>
<td>Dec-08</td>
<td>Jan-09</td>
<td>995,763</td>
<td>19,189</td>
<td>2.6</td>
</tr>
</tbody>
</table>
Tackling the Challenges of Big Data
Big Data Analytics
Applications: Finance
Machine Learning Techniques for Analyzing Big Data

THANK YOU
Empirical Results

- Type I and Type II error tradeoffs can be controlled by varying the threshold of the model.

<table>
<thead>
<tr>
<th>Classifier Threshold</th>
<th>Model Predict</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td>10%</td>
<td>1.5%</td>
<td>8.5%</td>
</tr>
<tr>
<td>20%</td>
<td>1.0%</td>
<td>8.0%</td>
</tr>
<tr>
<td>30%</td>
<td>2.5%</td>
<td>7.5%</td>
</tr>
<tr>
<td>40%</td>
<td>2.0%</td>
<td>7.0%</td>
</tr>
</tbody>
</table>

- Receiver Operating Characteristic (ROC) Curve
- Summarizes the trade-off noted on last slide
- True and false positive rate is calculated for different level of threshold

The threshold level can be optimized based on:
- Business objectives
- Risk appetite
- Capital requirements
- Employment cycle
- Etc.
Empirical Results

- Comparison with traditional credit scores

Tackling the Challenges of Big Data
Big Data Analytics
Applications: Finance
Empirical Results for a Commercial Bank's Credit Card Division

THANK YOU

Tackling the Challenges of Big Data
Big Data Analytics

Andrew W. Lo
Charles E. and Susan T. Harris Professor
Massachusetts Institute of Technology
Macro Forecasts of Credit Losses

- Forecasts of future credit losses may be used to construct an early warning system (12 months ahead!) for emerging problems in consumer credit

Measuring Value-Added of Forecasts

Assume that:
- In the beginning, both good and bad consumers will have the same average running balance
- Bad consumers will incur certain rate of "run-up" in their balance before default (we use 10%, 20%, 30% and 50% in our analysis)
- Credit card interest rate and lender’s funding cost rate are fixed at 5%
- Time horizon over which consumers amortize their credit card balance is fixed (we use 3, 5, and 10 years)
- The estimated value-added ranges from 6% to 24%, depending on the assumed parameters and client type (see next slide)
Measuring Value-Added of Forecasts

- Type I clients have "thin" files (very few transactions), Type IV clients have very "thick" files (many transactions)

These results show that the availability of features makes a big difference in forecast power and value-added.

Conclusion

- Big data can be used to construct better consumer credit risk forecasts
- Machine-learning techniques can add value
- High dimensionality of the data is both a blessing and a curse
- Key aspects are feature-vector construction and nonlinear interactions
- Many practical applications are possible

Tackling the Challenges of Big Data

Big Data Analytics

Applications: Finance

Gauging the Practical Value of Big Data for Consumer Credit Risk Management

THANK YOU
Tackling the Challenges of Big Data
Big Data Analytics

Andrew W. Lo
Charles E. and Susan T. Harris Professor
Massachusetts Institute of Technology

© 2014 Massachusetts Institute of Technology