Risk and Risk Management in the Credit Card Industry

F. Butaru, Q. Chen, B. Clark, S. Das, A. W. Lo and A. Siddique

Discussion by
Richard Stanton
Haas School of Business

MFM meeting
January 28–29, 2016

University of California
Berkeley
Haas School of Business

© Richard Stanton, 2016

What does the paper do?

- Compares state-of-the-art machine-learning methods in their ability to classify credit card accounts at 6 large banks as **good** vs. **bad**.
  - **Bad** = predicted to go 90 days past due within next 2 (3, 4) quarters.
  - Three techniques:
    - **Decision trees** (using C4.5 algorithm).
    - **Random forest**.
    - **Logistic regression**.
- Also analyzes **risk management** at these firms.
  - Measures ratio of % credit-line decreases on accounts that go delinquent to % credit-line decreases for all accounts.

© Richard Stanton, 2016

Lots of data!

- Internal account data from 6 large banks, monthly over six years starting Jan. 2008.
- Merged with consumer data from credit bureau, quarterly from 2009Q1, plus macro data from BLS and FHFA.
- 87 explanatory variables altogether:
  - **Account characteristics**, e.g., balance, utilization rate, purchase volume.
  - **Borrower characteristics**, e.g., # accounts, # other delinquent accounts, credit score.
  - **Macro variables**, e.g., home prices, income, unemployment.
- Final merged dataset: **5.7 million** accounts in 2009Q4, increasing to **6.6 million** in 2013Q4.

© Richard Stanton, 2016

Analysis

- Models compared out-of-sample, semiannually from 2010Q4 to 2013Q4.
- Standard metrics include
  - **Precision**: Proportion of positives identified to true positives (false positives).
  - **Recall**: Proportion of positives that is correctly identified (false negatives).
  - **F-measure**: Harmonic mean of precision and recall.
  - **Kappa**: Performance relative to random benchmark.
- Also estimates "value-added" of each classifier, the total savings that each model would have generated.

© Richard Stanton, 2016

Summary of results

- **Prediction**
  - All models do well at prediction.
  - Decision tree/random forest methods outperform logistic regression.
- **Risk management**
  - Measure: Ratio of % credit-line decreases on accounts that become delinquent over a forecast horizon to % credit line decreases for all accounts.
  - Ranges from below 1 (targeting “good” accounts) to 1.3.
  - Rankings across firms remain constant, but there is heterogeneity across firms.

© Richard Stanton, 2016

Performance vs. credit score: Logistic regression

© Richard Stanton, 2016

The big picture I

- Algorithms are rather a "black box," searching sequentially for (combinations of) variables to add/drop from analysis.
  - Trying to avoid **"curse of dimensionality."**
  - We don't know we've ended up with the optimum combination of variables.
  - What are statistical properties of these estimators?
- For stress testing and corporate risk management, we also want to know about *correlations* between defaults.
  - Go in waves, even after conditioning on all observable information.
- Independent variables are not all exogenous; forecast horizon is short.
  - Yes, I agree a loan is more likely to be 90 days delinquent in next 2 quarters if it's already 89 days delinquent!
  - What can we infer from this about the bank's performance over the next year? 5 years?

© Richard Stanton, 2016

The big picture II

- If trying to value the bank, conduct stress tests, or set contract terms, we need long-term (conditional) forecasts.
  - How do we simulate RHS variables over time?
  - Under risk-neutral measure?
  - How do we handle the fact that estimated models change over time?
- Short period, does not cover multiple business cycles.
  - Models only tested semiannually from 2010Q4 to 2013Q4.

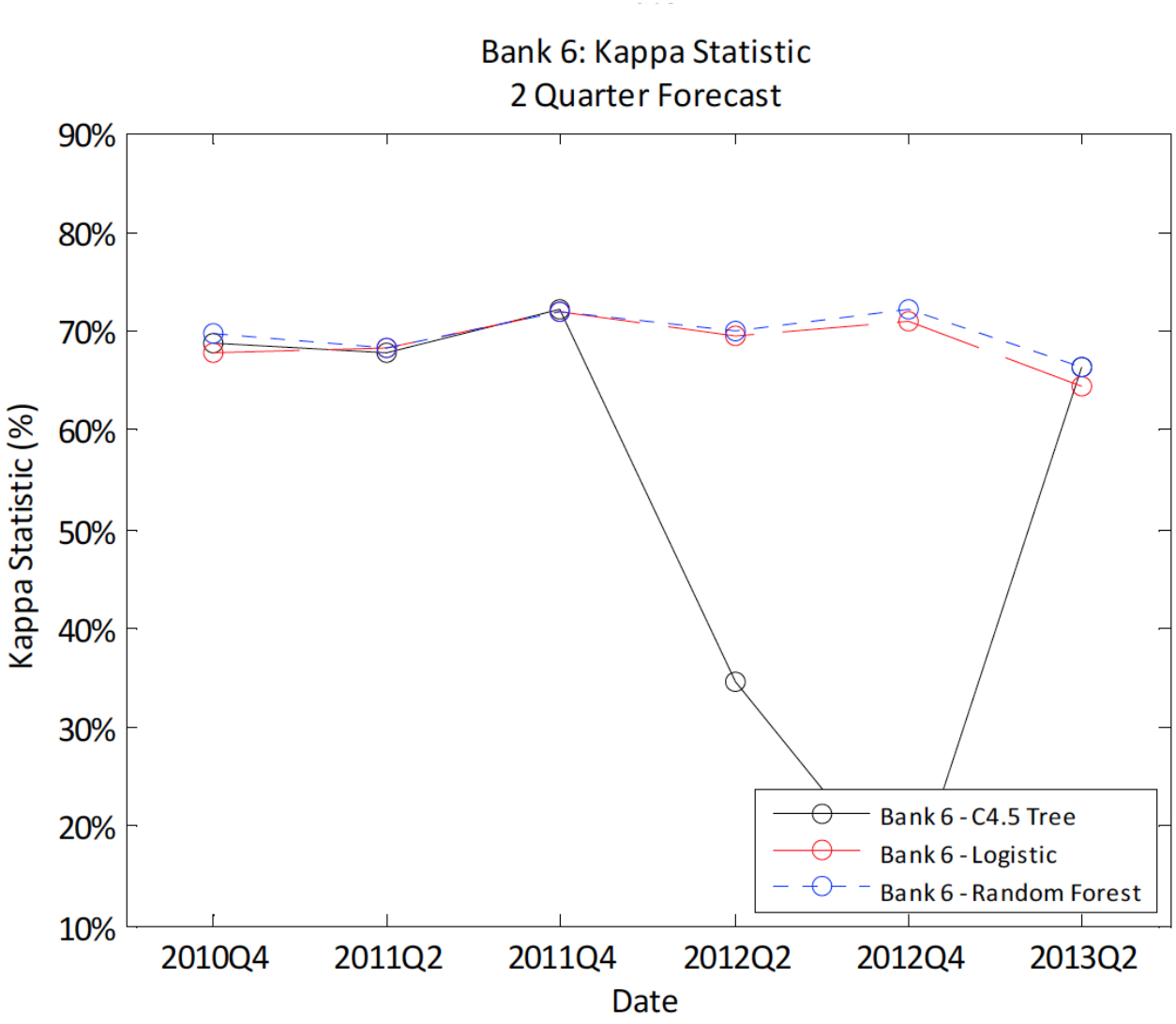
© Richard Stanton, 2016

Other Comments

- Income data, etc.: are they representative of the people actually in the sample?
- Credit-bureau records are not unique. There is one per account.
  - Can you look for identical records to merge accounts?
- Random sampling oversamples those with lots of accounts.
- Isn't home-ownership status important?
- Models generally perform similarly, but there are big outliers for some banks for some models for some periods. Why? For example. . .

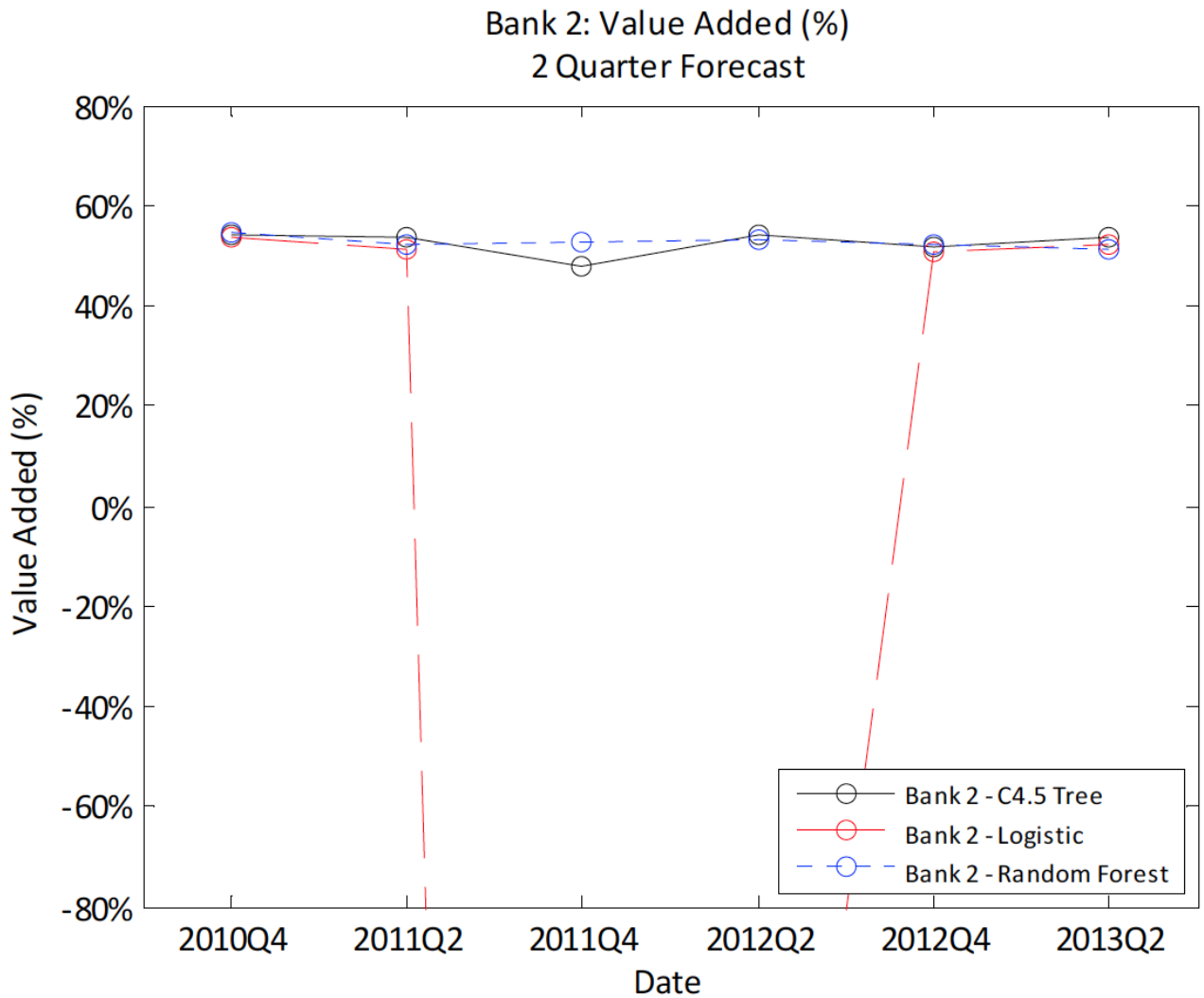
© Richard Stanton, 2016

Bank 6 kappa (figure 5)



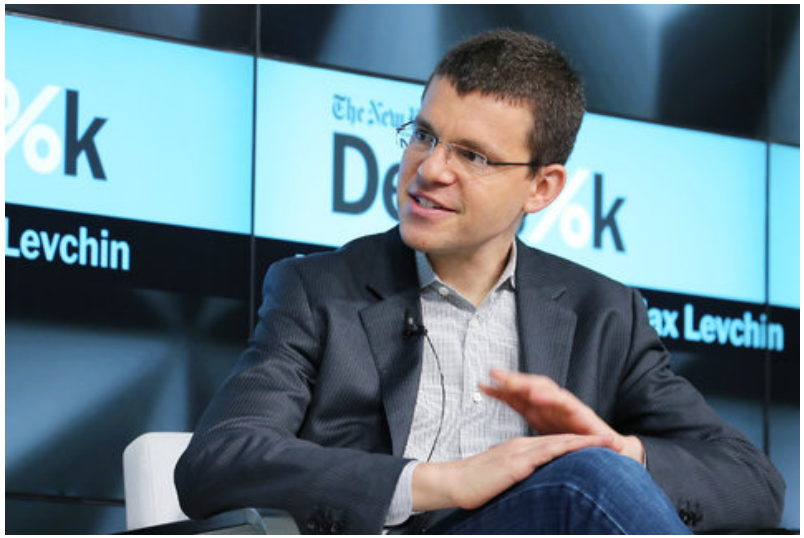
© Richard Stanton, 2016

Bank 2 value added (figure 6)



© Richard Stanton, 2016

How we *should* be doing it . . .



From the *New York Times*, 1/19/16:

> *Why would Affirm . . . make the seemingly snap judgment that Mr. Jimenez was a solid credit risk . . . ?*
>
> *"I wouldn't know," replied Max Levchin, Affirm's chief executive.* ***"Our math model says he's O.K."***

© Richard Stanton, 2016

How we *should* be doing it, contd.

From the *New York Times*, 1/19/16:

> *The traditional approach to risk-scoring relies on a person's credit history.*
>
> *The newcomers crunch all kinds of additional data including* ***social network profiles, bill payment histories, public records, online communications, even how applicants fill out forms on the web.***
>
> ***"It's not magic, it's math,"*** *said Mr. Levchin, a co-founder of PayPal.*

© Richard Stanton, 2016

Summary

- A fascinating look at the details of credit card default.
- Shows how useful machine-learning algorithms can be in high-dimensional settings.
- Raises lots of questions to be addressed in future research.

© Richard Stanton, 2016