Economic Modeling with Big Data: *Understanding Consumer Overdrafting at Banks*

Xiao Liu, Alan L. Montgomery
and Kannan Srinivasan

Tepper School of Business
Carnegie Mellon University

Hong Kong University
Ecom-Icomp Experts Address (20 July 2015)

Outline

- Big Data and Economic Modeling
- Banking's Overdrafting Problem
- Economic Modeling
  - Data
  - Model
  - Findings
- Conclusions

Big Data and Economic Modeling

Synergy and Conflict

Consumers generate Big Data

What is Big Data?

- Four basic components:
  - Massive datasets
  - Unstructured data
  - Collected as a by-product from transactions (not for decision making)
  - Populations not samples

Related to Business Analytics, Data Analysis, Data Mining, Data Science, Machine Learning, Statistics

Transaction Bank Data as an Example of Big Data

Time	Payee	Type	Amount
1May2013 : 07:32	Starbucks #56819 3618 Forbes Ave., Pittsburgh	Credit Card	$6.83
1May2013 : 12:16	EatUnique 305 S Craig St., Pittsburgh	Debit Card	$10.21
1May2013	~~Mrs. Smith~~	Check	$20.00
1May2013	~~Mr. Jones~~	Check	$50.00
1May2013	Carnegie Mellon University	Direct Deposit	$2,315.92
2May2013 : 14:45	Mobile Deposit (from ~~Acct 018468290~~)	Deposit	$18.99
2May2013: 18:20	~~alanmontgomery@cmu.edu~~	POPMoney	$25.00
2May2013	Verizon	BillPay	$92.18
2May2013	West Penn Electric	BillPay	$45.89
3May2013: 18:39	ATM Deposit; PNC #2999 4612 Forbes Ave., Pittsburgh	ATM	$100.00

Potential Promise of Big Data

Area	Key Benefits of Big Data
Customer data monetization	Customer Centricity
	Customer Risk Analysis
	Customer Retention
Transactions and operations	New products and services
	Algorithmic trading and analytics
	Organizational intelligence
Risk management and regulatory reporting	Risk management
	Regulatory Reporting

http://www.pwc.com/en_US/us/financial-services/publications/viewpoints/assets/pwc-unlocking-big-data-value.pdf

Potential Big Data Applications in Banking

Customer Data	Transactions	Risk Management
Customer life event analytics	Interactive Voice Analysis	MIS regulatory reporting
Next best offer	B2B Merchant Insight	Disclosure reporting
Real-time location based offerings	Real-time capital analytics	Real-time conversation keyword tracking
Sentiment analysis enabled sales	Log analytics	Anti-money laundering
Micro-segmentation		Indirect risk exposure analytics
Customer gamification		

http://www.pwc.com/en_US/us/financial-services/publications/viewpoints/assets/pwc-unlocking-big-data-value.pdf

Perceptions of Big Data by Financial Institutions

pwc, "*How the financial services industry can unlock the value of Big Data*":

- Financial institutions often mistakenly view Big Data as primarily a technology challenge rather than a business opportunity.
- Many financial institutions are not sure what it will take to translate the flood of information into business insights.
- Others are concerned about whether they have the right analytical skills and technologies in place.
- And those that are ready to join the data management revolution are asking where and how to begin transforming data into insights, intelligence, and ultimately, competitive advantage.

http://www.pwc.com/en_US/us/financial-services/publications/viewpoints/assets/pwc-unlocking-big-data-value.pdf

Perceptions of Big Data by Financial Institutions

"We do see industry leaders actively seeking strategies and solutions that will empower their organizations to comply with differing cross-border business initiatives, become more nimble, seize business opportunities, foster innovation, and improve their position in the marketplace."

pwc

http://www.pwc.com/en_US/us/financial-services/publications/viewpoints/assets/pwc-unlocking-big-data-value.pdf

Economic Modeling

- A mathematical representation of consumer and firm behavior that represents economic process by variables (choice, spending, savings, investments, time allocations, …) and relationships (either logical or quantitative) between these variables.
- Economic models are abstractions, and theory typically guides us to decide what is important or relevant (like price) and can be used to describe, predict or prescribe behavior.
- Most commonly economists assume rational behavior (e.g., utility or profit maximizing), but can incorporate bounded rationality, limited information, and search behavior.

Economic models are powerful tools in understanding economic relationships.

Microeconomic Theory

- Consumers will choose the bundle of goods (q) that maximizes their utility (U) given prices (p) and budget constraint (x)

$$\max_{q} U(q) \text{ s.t. } p \cdot q \leq x$$

- Many extensions:
  - Random utility
  - Multiple time periods
  - Savings and Investment
  - Changing utility models
  - Multiple consumers

Criticisms

- Incorrect assumptions
- Consumers do not really optimize
- Information is costly to gather and process
- Mathematical models may not yield intractable solutions that are not intuitive
- Lack of psychological connections

"It is better to be roughly right than precisely wrong"

Synergies and Conflicts

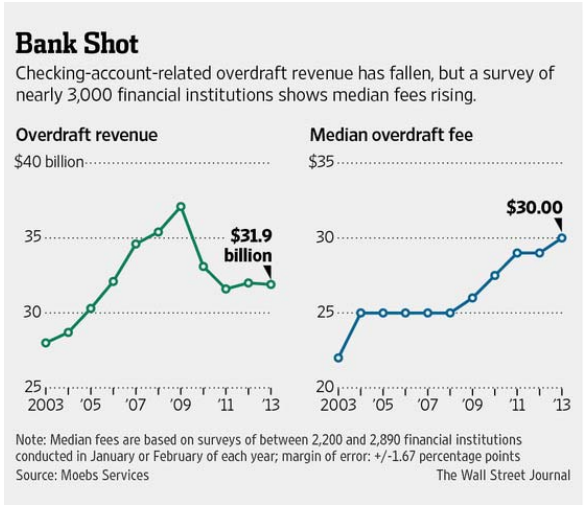
- Data science is a empirical science, and is well suited to absorb large scale data sets to derive inferences
  - But many machine learning techniques ignore theory and prior information, instead tending to use the data to find these patterns
- Our argument is that economic modeling can lend rigor and discipline to our data mining
  - Economic models provide an excellent paradigm for modeling consumer behavior

Banking Overdrafting Problem

Mobile Banking is Transformational

- Transaction data provides a rich resource for understanding and interacting with customers
- Mobile banking provides a new mechanism to interact with consumers at the time of purchase through mobile alerts and apps
- Large physical infrastructure needs to reorganized due to mobile adoption
- Potential to help consumers make better financial decisions

Industry Background



Industry Background

Overdraft Fees at Top U.S. Banks

Bank	Overdraft Fee (Per-transaction)
Bank of America	$35
BB&T	$36
Capital One	$35
Capital One 360	$0
Chase	$34
Citibank	$34
PNC	$36
SunTrust	$36
TD Bank	$35
US Bank	$36
Wells Fargo	$35

Source: http://www.nerdwallet.com/blog/banking/overdraft-fees-what-banks-charge/

Industry Background

Overdraft Fees at Top U.S. Banks

Bank	Overdraft Fee	Max Fees per Day	Overdraft Protection Transfer	Continuous Fee	Grace Period
Bank of America	$35	4	$10.00	$35	5
BB&T	$36	6	$12.50	$36	5
Capital One	$35	4	$10.00		
Capital One 360	$0	N/A	N/A		
Chase	$34	3	$10.00	$15	5
Citibank	$34	4	$10.00		
PNC	$36	4	$10 .00	$7	5
SunTrust	$36	6	$12.50	$36	7
TD Bank	$35	5	$10.00	$20	10
US Bank	$36	4	$12.50	$25	7
Wells Fargo	$35	4	$12.50		

Industry Background

Competitive Pressures

- Overdraft Free Checking Account aimed at low- and middle-income families
- No credit rating
- No overdraft fees
- No maintenance fee with $500/month deposit

Bank's Perspective

- Overdrafts are expensive and while fees are a major source of revenue, overdrawn accounts represent a significant liability
- Overdrafts represent high-risk, unsecured loans for which the costs of recovery likely exceed its value
- Fees serve to discourage customers from overdrafting

Research Questions

Pricing

- Is the current overdraft fee optimal?
- How will the revenue change under alternative pricing strategies?

Product Design

- How to both satisfy consumers and improve the bank profit? Alerts?
- How to design optimal alerts?

Economic Modeling

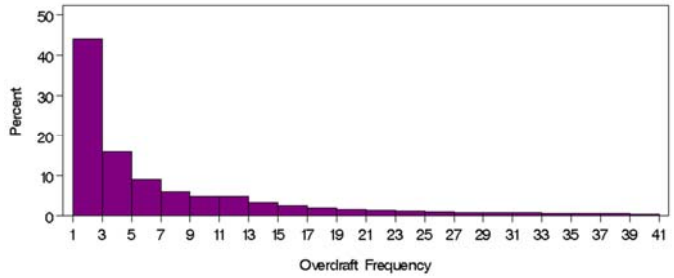
Exploratory Data Analysis

Data

- Large US bank
- 500k+ accounts
- 200m+ transactions
- June 2012-Aug 2013
- 16% have at least one overdraft (If overdraft average is 10 with $245 fees)

Data Snapshot

DATE	+/-	AMOUNT	BALANCE	DESCRIPTION
11/15/2012	+	287.42	301.57	SALARY
11/15/2012	-	130.41	171.16	BILLPAY TV CABLE
11/15/2012	-	97.84	73.32	PURCHASE GROCERY
11/16/2012				CHECK BALANCE
11/19/2012	-	80	-6.68	ATM WITHDRAWAL
11/21/2012	-	31	-37.68	OVERDRAFT ITEM FEE



What transactions cause overdrafts?

Type	Frequency	Percentage	Amount
Debit Card Purchase	946,049	48.65%	$29.50
ACH Transaction	267,854	13.77%	$294.57
Check	227,128	11.68%	$417.78
ATM Withdrawal	68,328	3.51%	$89.77

Consumer Characteristics

Characteristics for Overdrafters and Non-overdrafters

	Non-Overdrafter		Overdrafter
Share	84.2%		15.8%
Age	46.71 (18.23)	>	43.50 (15.10)
Low Income	19.25% (39.43%)	<	25.72% (43.71%)
Mean Balance	8325.68 (34387.61)	>	1257.93 (13769.75)
Certificate of Dep	0.11 (0.72)	>	0.02 (0.25)
Credit Card	0.27 (0.53)	>	0.14 (0.39)
Online Transfer	0.05 (0.40)	>	0.04 (0.39)
Debit Card	1.66 (1.33)	<	1.71 (1.25)
Total Txns(Month)	46.85 (43.68)	<	55.97 (46.00)
Debit Card Txns	42.44 (41.51)	<	50.92 (43.75)
ATM Txns	3.24 (4.62)	<	4.53 (5.63)
Branch Txns	1.42 (2.12)	<	1.62 (2.37)

Overdraft Example

DATE	+/-	AMOUNT	BALANC	DESCRIPTION
8/17/2012	+	734.11	1705.34	SALARY
8/17/2012	-	535	1170.34	BILLPAY RENT
8/17/2012	-	96.85	1073.49	PURCHASE DEPARTMENT STORE
8/17/2012	-	87	986.49	PURCHASE ELECTRONICS
8/18/2012	-	56.99	929.5	PURCHASE CLOTHING
8/18/2012	-	15.23	914.27	PURCHASE RESTAURANT
8/18/2012	-	585.05	329.22	BILLPAY MORTGAGE
8/19/2012	-	106.3	222.92	PURCHASE HOME
8/19/2012	-	92.52	130.4	PURCHASE GROCERY
8/20/2012	-	38.59	91.81	PURCHASE RESTAURANT
8/20/2012	-	37.13	54.68	PURCHASE ONLINE SPORTS
8/20/2012	-	33.52	21.16	PURCHASE CLOTHING
8/21/2012	-	25	-3.84	PURCHASE RESTAURANT
8/21/2012	-	17.12	-20.96	PURCHASE BEAUTY
8/22/2012	-	6.31	-27.27	PURCHASE GAME STORE
8/22/2012	-	4.95	-32.22	PURCHASE COFFEE
8/23/2012	+	180	147.78	ATM DEPOSIT
8/23/2012	-	31	116.78	OVERDRAFT FEE
8/23/2012	-	31	85.78	OVERDRAFT FEE
8/23/2012	-	31	54.78	OVERDRAFT FEE
8/23/2012	-	31	23.78	OVERDRAFT FEE

Rational explanations of Overdrafting Behavior

- High Discount Rates

  Consumers value current consumption so much that they are willing to pay exorbitant interest rates
- Inattention

  Consumers have to "guess" at their balance and perceptual errors can cause overdrafts
- Monitoring Costs

  Consumers incur a cost to check their balances. This is an opportunity cost and not a cost imposed by the bank.
- Dissatisfaction

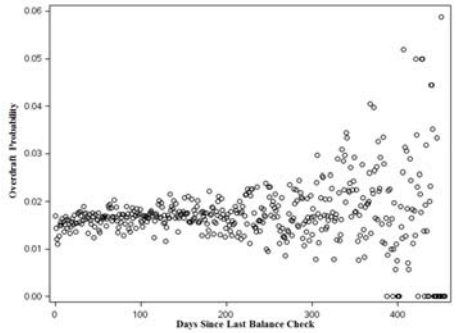
  Consumers get irritated with banks over overdrafts fees and close their accounts

Consumer Discounting

- Consumer wants to make a purchase may sharply discount the future cost of overdraft fee to satisfy immediate consumption
- Empirical support: Consumers spend more after getting a pay check and then reduce spending during the course of the month
- Overspending at the beginning, the consumer is going to run out of budget at the end of the pay period and has to overdraft. Strong support for this behavior amongst heavy overdrafters

Inattention

- Consumers might be inattentively monitoring their checking accounts so that they are uncertain about the exact balance amount.
- The perceived balance may be higher than the true balance which might result in an inadvertent overdraft
- Empirical evidence: overdrafting behavior for infrequent overdrafters and find that those who check their balances more frequently. Heavy overdrafters do not.

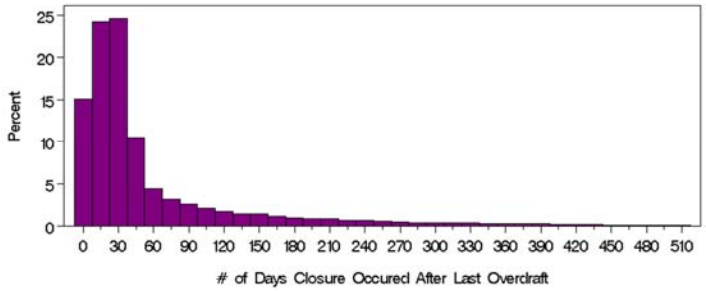


Monitoring Costs

- If making a balance inquiry lessens overdrafts then why don't consumers always check their balance?
- We argue that there are opportunity costs (time, effort, and mental efforts) which consumers incur and limit their balance inquiries

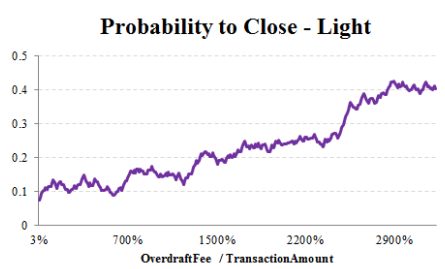
Dissatisfaction

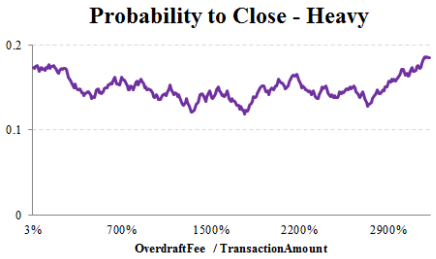
- Consumers get upset with banks when they incur overdraft fees – especially those that are associated with small overdrafts
- Empirical evidence: Overdrafts who voluntarily close their account are very likely to close it soon after the overdraft



Dissatisfaction

- If the ratio of overdraft fees to transaction amounts are large then consumers are more likely to close their accounts. (Not true for heavy overdrafters)





Economic Modeling

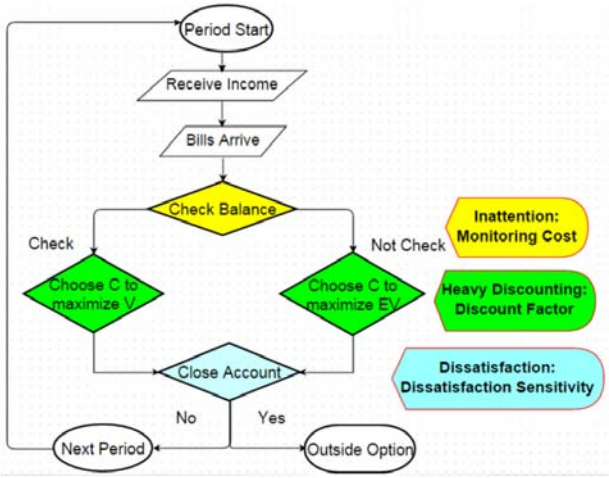
Model and Results

Model Input: Tag Data

- Classify each transaction as income, bill, fee or spending

Date	+/-	Amount	Balance	Description	Tag
11/15/2012	+	287.42	301.57	SALARY	Income
11/15/2012	-	130.41	171.16	BILLPAY TV CABLE	Bill
11/15/2012	-	97.84	73.32	PURCHASE GROCERY	Spending
11/16/2012				CHECK BALANCE	
11/19/2012	-	80	-6.68	ATM WITHDRAWAL	Spending
11/21/2012	-	31	-37.68	OVERDAFT ITEM FEE	Fee

Model Overview



Mathematical Formulation

Objective	$\max_{\{W_{it},Q_{it},C_{it}\}_{t=1}^{\infty}} E\left[\sum_{t=1}^{\infty}\beta_i^{t-1}U_{it}\left(W_{it},Q_{it},C_{it};S_{it}\right)\right]$
1) Basic Model	$u_C(C_{it})=\frac{C_{it}^{1-\theta_{it}}}{1-\theta_{it}}$ $\theta_{it}=exp(\theta+\varepsilon_{it}),\varepsilon_{it}\sim N(0,\varsigma_i^2)$
2) Inattention	$\widetilde{B}_{it}\sim Q_{it}B_{it}+(1-Q_{it})N(B_{it}+\eta_{it}\omega_{it},\omega_{it}^2)$ $u_{it}=\int_{\widetilde{B}_{it}}\int_{\eta_{it}}[u_C(C_{it})-Q_{it}\xi_i+\chi_{itQ_{it}}]dF(\eta_{it})dF(\widetilde{B}_{it})$ $\eta_{it}\sim N(0,1),\omega_{it}^2=\rho_i\Gamma_{it}$
3) Dissatisfaction	$U_{it}=[u_{it}-\Upsilon_i*\Delta_{it}*I[B_{it}-C_{it}<0]+\varpi_{it0}](1-W_{it})+W_{it}\varpi_{it1}$

Model Method: Dynamic Programming

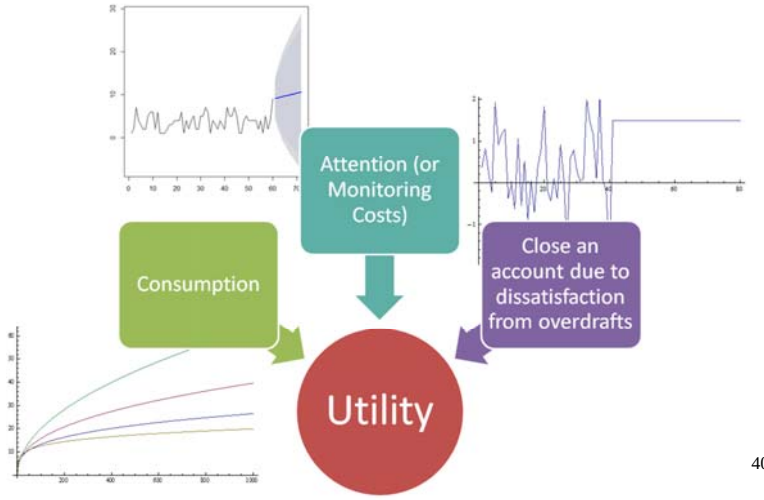
- Consumer make daily decisions to optimize NPV of utility:

	1	2	3	4	5	6	7	...
Spending	$138	$68	$20	$32	$73	$54	$38	...
Check Balance?	Yes	No	No	No	Yes	No	No	...
Close Account?	No	No	No	No	No	Yes	Yes	...

- On a daily basis consumers update their balance

	1	2	3	4	5	6	7	...
Fee				$31				...
Income		$200			$310			
Bills			$80		$50			
Balance	$100	$162	$14	-$6	$191	$118	$64	...

Model Overview



Estimation Challenges

- Scale of Data overwhelm conventional estimation techniques
- Employ two techniques: parallelization and a new Bayesian technique to solve the dynamic programming problem

Estimation Time Comparison

Size\Method(seconds)	**Parallel IJC**	IJC	CCP	FIML
1,000	518	1,579	526	5,010
10,000	3,199	12,560	4,679	54,280
100,000	4,059	140,813	55,226	640,360
>500,000	5,308	788,294	399,337	3,372,660
	(1.5 hr)	(9 days)	(5 days)	(39 days)

Findings

- Some consumers have really low discount rates
- Monitoring cost is equivalent to $2.03
- Online banking reduces monitoring costs by $0.87
- 1% increase in overdraft fee increases closing probability by 0.12%
- Light overdrafters have a higher dissatisfaction sensitivity

Alternative Model Comparisons

	A: No Forward-Looking	B: No Inattention	C: No Dissatisfaction	D: No Heterogeneity	E: Proposed
Log-Marginal Density	-2943.28	-3636.59	-2837.75	-2764.56	-1758.33
Hit Rate: Overdraft	0.499	0.351	0.502	0.504	0.870
Hit Rate: Check Balance	0.405	0.226	0.441	0.632	0.841
Hit Rate: Close Account	0.660	0.727	0.438	0.696	0.758

Value of Big Data

- Discovery of rare events
  - 0.6% of all transactions are overdrafts – without large datasets could not detect these events (would look like outliers)
- Rich micro-level variation
  - Daily spending and balance checking
- Reduce sampling error at minimal computational cost
  - Revenue loss: 3% ≈ $0.6m for a 10% random sample

Economic Modeling

Findings

Pricing

Overdraft Revenue under Alternative Pricing Strategies

Pricing	Current	Reduced Flat	Percentage	Quantity Premium
	$31	$29.27	15.8%	8.5% *I ($OD \leq 10$) + $31 *I ($OD > 10$)
Overdraft Revenue	$18,654,510	$19,262,647	$19,982,711	$20,297,972
Overdraft Freq	544,997	590,093	610,288	631,325
% △Revenue	–	+3.26%	+7.12%	+8.81%
% △Freq	–	+2.77%	+11.98%	+15.84%
% △Check Balance	–	-3.58%	+2.83%	+3.31%
% △Close Account	–	-1.01%	-1.35%	-1.94%

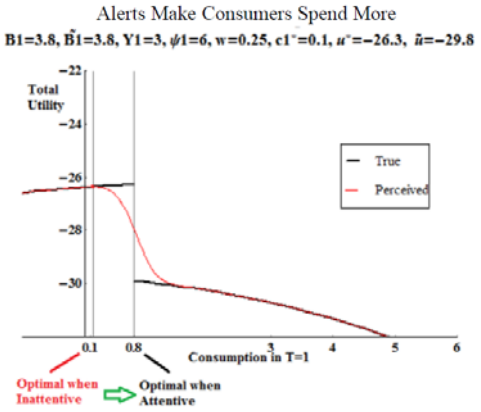
Pricing Implications

- Current per-transaction fee may be too high
- Percentage fee: market expansion
- Quantity premium: second degree price discrimination

Alerts Creates New Revenue Model

Bank's new revenue source:

Increased interchange fees from increased spendings



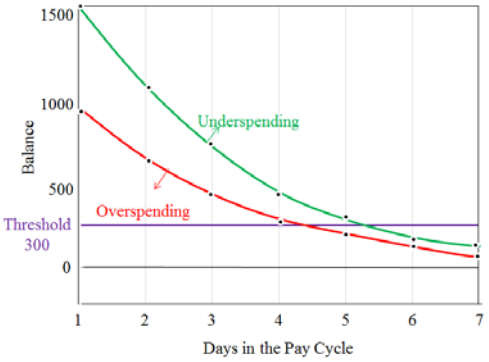
Bank Benefits from Alerts

Gain in interchange fee and consumer LTV could offset loss in overdraft revenue

Dynamic Alerts

- Optimal alerts could be targeted and dynamic
- Alert timing: both overspending and underspending warnings, avoids spamming

Alert Type	Alert Timing	Utility Gain
Uniform	Threshold	1.11%
	Dynamic	2.85%
Targeted	Threshold	4.39%
	Dynamic	6.65%



Conclusions

Conclusions

- Demonstrate the potential for using economic modeling with Big Data
  - Great theories of consumer behavior with structural models
  - More "realistic" and complex models can be estimated
  - Need for parallel computing
- Results show
  - Banks can use transactional data to both improve consumer satisfaction and increase LTV
  - Current per-transaction overdraft fees may be too high (but not huge reductions)
  - Alerts can be intelligent, targeted, dynamic and leverage bank's transaction history about consumers