Overhaul Overdraft Fees: Creating Pricing and Product Design Strategies with Big Data*

Xiao Liu, Alan Montgomery, Kannan Srinivasan

October 12, 2014

Abstract

In 2012, consumers paid an enormous $32 billion overdraft fees. Consumer attrition and potential government regulations to shut down the overdraft service urge banks to come up with financial innovations to overhaul the overdraft fees. However, no empirical research has been done to explain consumers' overdraft incentives and evaluate alternative pricing and product strategies. In this paper, we build a dynamic structural model with consumer monitoring cost and dissatisfaction. We find that on one hand, consumers heavily discount the future and overdraw because of impulsive spending. On the other hand, a high monitoring cost makes it hard for consumers to track their finances therefore they overdraw because of rational inattention. In addition, consumers are dissatisfied by the overly high overdraft fee and close their accounts. We apply the model to a big dataset of more than 500,000 accounts for a span of 450 days. Our policy simulations show that alternative pricing strategies may increase the bank's revenue. Sending targeted and dynamic alerts to consumers can not only help consumers avoid overdraft fees but improve bank profits from higher interchange fees and less consumer attrition. To alleviate the computational burden of solving dynamic programming problems on a large scale, we combine parallel computing techniques with a Bayesian Markov Chain Monte Carlo algorithm. The Big Data allow us to detect the rare event of overdraft and reduce the sampling error with minimal computational costs.

1 Introduction

An overdraft occurs when a consumer attempts to spend or withdraw funds from her checking accounts in an amount exceeding the account's available funds. In the US, banks allow consumers to overdraw their accounts (subject to some restrictions at banks' discretion) and charge an overdraft fee. Overdraft fees have become a major source of bank revenues since banks started to offer free checking accounts to attract consumers. In 2012, the total amount of overdraft fees in the US reached $32 billion, according to Moebs Services[1]. This is equivalent to an average of $178 for each checking account annually[2]. According to the Center for Responsible Lending, US households spent more on overdraft fees than on fresh vegetables, postage and books in 2010.[3]

*We acknowledge support from the Dipankar and Sharmila Chakravarti Fellowship. All errors are our own.

[1] http://www.moebs.com

[2] According to Evans, Litan, and Schmalensee 2011, there are 180 million checking accounts in the US.

[3] http://www.blackenterprise.com/money/managing-credit-3-ways-overdraft-fees-will-still-haunt-you/

The unfairly high overdraft fee has provoked a storm of consumer outrage and therefore caused many consumers to close the account. The US government has taken actions to regulate these overdraft fees through the Consumer Financial Protection Agency[4] and may potentially shut down the overdraft service[5]. Without overhauling the current overdraft fee, banks encounter the problem of losing valuable customers and possibly totally losing the revenue source from overdrafts.

Financial institutions store massive amounts of information about consumers. The advantages of technology and Big Data enable banks to reverse the information asymmetry (Kamenica, Mullainathan, and Thaler 2011) as they may be able to generate better forecasts about a consumer's financial state than consumers themselves can. In this paper, we extract the valuable information embedded in the Big Data and harness it with structural economic theories to explain consumers' overdraft behavior. The large scale financial transaction panel data allows us to sort through consumers' financial decision making processes and discover rich consumer heterogeneity. As a consequence, we come up with individually customized strategies that can increase both consumer welfare and bank revenue.

In this paper, we aim to achieve two substantive goals. First, we leverage rich data about consumer spending and balance checking to understand the decision process for consumers to overdraw. We address the following research questions. Are consumers fully attentive in monitoring their checking account balances? How great is the monitoring cost? Why do attentive consumers also overdraw? Are consumers dissatisfied because the overdraft fee?

Second, we investigate pricing and new product design strategies that overhaul overdraft fees. Specifically, we tackle these questions. Is the current overdraft fee structure optimal? How will the bank revenue change under alternative pricing strategies? More importantly, what new revenue model can make the incentives of the bank and consumers better aligned? Can the bank benefit from helping consumers make more informed financial decisions, like sending alerts to consumers? If so, what's the optimal alert strategy? How can the bank leverage its rich data about consumer financial behaviors to reverse information asymmetry and create targeted strategies?

We estimate the dynamic structural model using data from a large commercial bank in the US. The sample size is over 500,000 accounts and the sample length is up to 450 days. We find that some consumers are inattentive in monitoring their finances because of a substantially high monitoring cost. In contrast, attentive consumers overdraw because they heavily discount future utilities and are subject to impulsive spending. Consumers are dissatisfied to leave the bank after being charged the unfairly high overdraft fees. In our counterfactual analysis, we show that a percentage fee or a quantity premium fee strategy can achieve higher bank revenue compared to the current flat per-transaction fee strategy. Enabled by Big Data, we also propose an optimal targeted alert strategy. The bank can benefit from sending alerts to let consumers spend their unused balances so that the bank can earn more interchange fees. Helping consumers make more informed decisions will also significantly reduce consumer attrition. The targeted dynamic alerts should be sent to consumers with higher monitoring costs and both when they are underspending and overspending.

Methodologically, our paper makes two key contributions. First, we build a dynamic structural model that incorporates inattention and dissatisfaction into the life-time consumption model. Although we apply it to the overdraft context, the model framework can be generalized to ana-

[4]http://banking-law.lawyers.com/consumer-banking/consumers-and-congress-tackle-big-bank-fees.html
[5]http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf

lyze other marketing problems regarding consumer dynamic budget allocation, like electricity and cellphone usage.

Second, we estimate the model on Big Data with the help of parallel computing techniques. Structural models have the merit of producing policy invariate parameters that allow us to conduct counterfactual analysis. However, the inherent computational burden prevents it from being widely adopted by industries. Moreover, the data size in a real setting is typically much larger than what's used for research purposes. Companies, in our case a large bank, need to have methods that are easily scalable to generate targeted solutions for each consumer. Our proposed algorithm takes advantage of state-of-the-art parallel computing techniques and estimation methods that alleviate computational burden and reduce the curse of dimensionality.

The rest of the paper is organized as follows. In section 2 we first review related literature. Then we show summary statistics in section 3 which motivate our model setup. Section 4 describes our structural model and we provide details of identification and estimation procedures in section 5. Then in sections 6 and 7 we show estimation results and counterfactual analysis. Section 8 concludes and summarizes our limitations.

2 Related Literature

A variety of economic and psychological models can explain overdrafts, including full-information pure rational models and limited attention, as summarized by Stango and Zinman (2014). However, no empirical paper has applied these theories to real consumer spending data. Although Stango and Zinman (2014) had a similar dataset to ours, their focus was on testing whether taking related surveys can reduce overdrafts. We develop a dynamic structural model that incorporates theories of heavy discounting, inattention and dissatisfaction in a comprehensive framework. The model is flexible to address various overdraft scenarios, thus it can be used by policy makers and the bank to design targeted strategies to increase consumer welfare and bank revenue.

Our model inherits from the traditional lifetime consumption model but adds two novel features, inattention and dissatisfaction. First of all, a large body of literature in psychology and economics has found that consumers pay limited attention to relevant information. In the review paper by Card, DellaVigna and Malmendier (2011), they summarize findings indicating that consumers pay limited attention to 1) shipping costs, 2) tax (Chetty et. al. 2009) and 3) ranking (Pope 2009). Gabaix and Laibson (2006) find that consumers don't pay enough attention to add-on pricing and Grubb (2014) shows consumers' inattention to their cell-phone minute balances. Many papers in the finance and accounting domain have documented that investors and financial analysts are inattentive to various financial information (e.g., Hirshleifer and Teoh 2003, Peng and Xiong 2006). We follow Stango and Zinman (2014) to define inattention as incomplete consideration of account balances (realized balance and available balance net of coming bills) that would inform choices. We further explain inattention with a structural parameter, monitoring cost, which represents the time and effort to know the exact amount of money in the checking account. With this parameter estimated, we are able to quantify the economic value of sending alerts to consumers and provide guidance for the bank to set its pricing strategy. We also come up with policy simulations about alerts because we think a direct remedy for consumers' limited attention is to make information more salient (Card, DellaVigna and Malmendier 2011). Past literature also finds that reminders (Karlan et. al. 2010), mandatory disclosure (Fishman and Hagerty 2003), and penalties (Haselhuhn et al. 2012) all serve the purpose of increasing salience and thus mitigating the negative consequences of inattention.

Second, as documented in previous literature, unfairly high price may cause consumer dissatisfaction which is one of the main causes of customer switching behavior (Keaveney 1995, Bolton 1998). We notice that consumers are more likely to close the account after paying the overdraft fee and when the ratio of the overdraft fee over the overdraft transaction amount is high. This is because given the current banking industry practice, a consumer pays a flat per-transaction fee regardless of the transaction amount. Therefore, the implied interest rate for an overdraft originated by a small transaction amount is much higher than the socially accepted interest rate (Matzler, Wurtele and Renzl 2006), leading to price dissatisfaction.

We aim to estimate this infinite horizon dynamic structural model on a large scale of data and obtain heterogeneous best response for each consumer to prepare targeted marketing strategies. After searching among different estimation methods, including the nested fixed point algorithm (Rust 1987), the conditional choice probability estimation (Arcidiacono and Miller 2011) and the Bayesian estimation method developed in Imai, Jain and Ching (2009) (IJC), we finally choose the IJC method for the following reasons. First of all, the hierarchical Bayes framework fits our goal of obtaining heterogeneous parameters. Second, in order to apply our model to a large scale of data, we need to estimate the model with Bayesian MCMC so that we can implement a parallel computing technique. Third, IJC is the state-of-the art Bayesian estimation algorithm for infinite horizon dynamic programming models. It provides two additional benefits in tackling the computational challenges. One is that it alleviates the computational burden by only evaluating the value function once in each MC iteration. Essentially, the algorithm solves the value function and estimates the structural parameters simultaneously. So the computational burden of a dynamic problem is reduced by an order of magnitude similar to those computational costs of a static model. The other is that the method reduces the curse of dimensionality by allowing state space grid points to vary between estimation iterations. On the other hand, as our sample size is huge, traditional MCMC estimation may take a prohibitively, if not impossibly, long time, since for N data points, most methods must perform O(N) operations to draw a sample. A natural way to reduce the computation time is to run the chain in parallel. Past methods of Parallel MCMC duplicate the data on multiple machines and cannot reduce the time of burn-in. We instead use a new technique developed by Neiswanger, Wang and Xing (2014) to solve this problem. The key idea of this algorithm is that we can distribute data into multiple machines and perform IJC estimation in parallel. Once we obtain the posterior Markov Chains from each machine, we can algorithmically combine these individual chains to get the posterior chain of the whole sample.

3 Background and Model Free Evidence

We obtained data from a major commercial bank in the US. During our sample period in 2012 and 2013, overdraft fees accounted for 47% of the revenue from deposit account service charges and 9.8% of the operating revenue.

The bank provides a comprehensive overdraft solution to consumers. (For general overdraft practices in the US, please refer to Stango and Zinman (2014) for a good review. Appendix A.1 tabulates current fee settings in top US banks.) In the standard overdraft service, if the consumer overdraws her account, the bank might cover the transaction and charge \$31[6] Overdraft Fee (OD) or decline the transaction and charge a \$31 Non-Sufficient-Fund Fee (NSF). Whether the transaction is accepted or declined is at the bank's discretion. The OD/NSF fee is at a per-item level. If a consumer performs several transactions when the account is already overdrawn, each transaction item will incur a fee of 31 dollars. Within a day, a maximum of four per-item fees can be charged. If the account remains overdrawn for five or more consecutive calendar days, a Continuous Overdraft Fee of \$6 will be assessed up to a maximum of \$84. The bank also provides an Overdraft Protection Service where the checking account can link to another checking account, a credit card or a line of credit. In this case, when the focal account is overdrawn, funds can be transferred to cover the negative balance. The Overdraft Transfer Balance Fee is \$9 for each transfer. As you can see, the fee structure for the bank is quite complicated. In the empirical analysis below, we don't distinguish between different types of overdraft fees and assume that money is fungible so that the consumer only cares about the total amount of overdraft fee rather than the underlying pricing structure.

The bank also provides balance checking services through branch, automated teller machine (ATM), call center and online/mobile banking. Consumers can inquire about their available balances and recent activities. There's also a notification service to consumers via email or text message, named "alerts". Consumers can set alerts when certain events take place, like overdrafts, insufficient funds, transfers, deposits, etc. Unfortunately, our dataset only includes the balance checking data but not the alert data. We'll discuss this limitation in section 8.

In 2009, the Federal Reserve Board made an amendment to Regulation E (subsequently recodified by the Consumer Financial Protection Bureau (CFPB)) which requires account holders to provide affirmative consent (opt in) for overdraft coverage of ATM and non-recurring point of sale (POS) debit card transactions before banks can charge for paying such transactions[7]. This Regulation E aimed to protect consumers against the heavy overdraft fees. The change became effective for new accounts on July 1, 2010, and for existing accounts on August 15, 2010. Our sample contains both opt-in and opt-out accounts. However, we don't know which accounts have opted in unless we observe an ATM/POS initiated overdraft occasion. We also discuss this data limitation in section 8.

3.1 Summary Statistics

Our data can be divided into two categories, checking account transactions and balance inquiry activities. In our sample, there are between 500,000 and 1,000,000[8] accounts, among which 15.8% had at least one overdraft incidence during the sample period between June 2012 and Aug 2013. The proportion of accounts with overdraft is lower than the 27% (across all banks and credit unions) reported by the CFPB in 2012[9]. In total, all the counts performed more than 200 million transactions, including deposits, withdrawals, transfers, and payments etc. For each transaction, we know the account number, transaction date, transaction amount, and transaction description. The transac-

[6] All dollar values in the paper have been rescaled by a number between .85 and 1.15 to help obfuscate the exact amounts without changing the substantive implications. The bank also sets the first time overdraft fee for each consumer at \$22. All the rest overdraft fees are set at \$31.

[7] http://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-43.html

[8] For the sake of privacy, we can't disclose the exact number.

[9] http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf

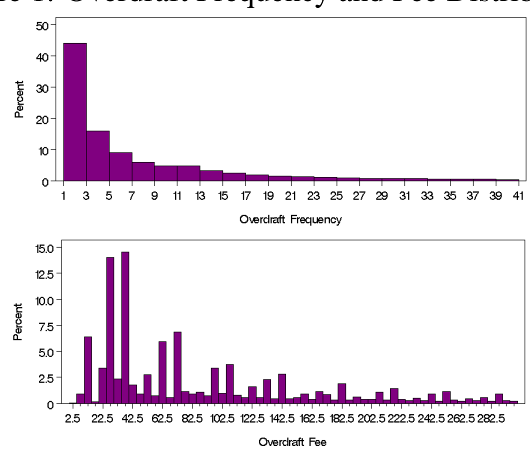
tion description tells us the type of transaction (e.g., ATM withdrawal or debit card purchase) and location/associated institution of the transaction, like merchant name or branch location. The description helps us identify the cause of the overdraft, for instance whether it's due to an electricity bill or due to a grocery purchase.

Table 1: Overdraft Frequency and Fee Distribution

	Mean	Std	Median	Min	99.85 Percentile
OD Frequency	9.84	18.74	3	1	>100
OD Fee	245.46	523.04	77	10	>2730

As shown in Table 1, consumers who paid overdraft fees, on average, overdrew nearly 10 times and paid $245 during the 15 month sample period. This is consistent with the finding from the CFPB that the average overdraft- and NSF-related fees paid by all accounts that had one or more overdraft transactions in 2011 were $225[10]. There is significant heterogeneity in consumers' overdraft frequency and the distribution of overdraft frequency is quite skewed. The median overdraft frequency is three and more than 25% of consumers overdrew only once. In contrast, the top 0.15% of heavy overdrafters overdrew more than 100 times. A similar skewed pattern applies to the distribution of overdraft fees. While the median overdraft fee is $77, the top 0.15% of heaviest overdrafters paid more than $2,730 in fees.

Figure 1: Overdraft Frequency and Fee Distribution



Now let's zoom in to take a look at the behavior of the majority overdrafters that have overdrawn less than 40 times. The first panel in Figure 1 depicts the distribution of overdraft frequency for those accounts. Notice that most consumers (> 50%) only overdrew less than three times. The second panel shows the distribution of the paid overdraft fee for accounts that have overdrawn less than $300. Consistent with the fee structure where the standard per-item overdraft fee is $22 or $31, we see spikes on these two numbers and their multiples.

[10]http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf

Table 2: Types of Transactions That Cause Overdraft

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

What types of transactions cause overdraft? We find that nearly 50% of overdrafts are caused by debit card purchases with mean transaction amounts around $30. On the other hand, ACH (Automated Clearing House) and Check transactions account for 13.77% and 11.68% of overdraft occasions. These transactions are generally for larger amounts, $294.57 and $417.78, respectively. ATM withdrawals lead to another 3.51% of the overdraft transactions with an average amount of around $90.

3.2 Model Free Evidence

This section presents some patterns in the data that suggest the causes and effects of overdrafts. We show that heavy discounting and inattention may drive consumers' overdraft behaviors. And consumers are dissatisfied because of the overdraft fees. The model free evidence also highlights the variation in the data that will allow for the identification of the discount factor, monitoring cost and dissatisfaction sensitivity.

3.2.1 Heavy Discounting

First of all, we argue that a consumer may overdraw because she prefers current consumption much more than future consumption, i.e. she heavily discounts future consumption utility. At the point of sale, the consumer sharply discounts the future cost of the overdraft fee to satisfy immediate gratification [11]. If that's the case, then we should observe a steep downward sloping trend in the spending pattern within a pay period. That is, the consumer will spend a lot right after getting a pay check and then reduce spending during the course of the month. But because of overspending at the beginning, the consumer is going to run out of budget at the end of the pay period and has to overdraw.

We test this hypothesis with the following model specification. We assume that the spending for consumer i at time t $Spending_{it}$ can be modeled as

$$Spending_{it} = \beta * LapsedTimeAfterIncome_{it} + \mu_i + v_t + \varepsilon_{it}$$

where $LapsedTimeAfterIncome_{it}$ is the number of days after the consumer received income (salary), μ_i is the individual fixed effect and v_t is the time (day) fixed effect. To control for the

[11] We also considered hyperbolic discounting with two discount factors, a short term present bias parameter and a long term discount factor. With more than three periods of data within a pay period, hyperbolic discount factors can be identified (Fang and Silverman 2009). However, our estimation results show that the present bias parameter is not significantly different from 1. Therefore we only keep one discount factor in the current model. Estimation results with hyperbolic discount factors are available upon requests.

effect that consumers usually pay for their bills (utilities, phone bills, credit card bills, etc) after getting the pay check, we exclude checks and ACH transactions which are the common choices for bill payments from the daily spendings and only keep debit card purchases, ATM withdrawals and person-to-person transfers.

We run this OLS regression for heavy overdrafters (whose overdraft frequency is in the top 20 percentile among all overdrafters), light overdrafters (whose overdraft frequency is not in the top 20 percentile among all overdrafters) and non-overdrafters (who didn't overdraw during the 15 months sample period) separately. The results are reported in column (1) (2) and (3) of Table 3.

Table 3: Spending Decreases with Time in a Pay Cycle

	(1)	(2)	(3)
	Heavy Overdrafters	Light Overdrafters	Non-Overdrafters
Lapsed Time after Income (β)	-6.8374^{***}	-0.00007815	-0.00002195
	(0.00006923)	(0.00006540)	(0.00002328)
Fixed Effect	Yes	Yes	Yes
Number of Observations	17,810,276	53,845,039	242,598,851
R^2	0.207	0.275	0.280

Note: *p<0.01;**p<0.001;***p<0.0001

We find that the coefficient of $LapsedTimeAfterIncome_{it}$ is negative and significant for heavy overdrafters but not light overdrafters or non-overdrafters. This suggests that heavy overdrafters have a steep downward sloping spending pattern during a pay period while light overdrafters or non-overdrafters have a relatively stable spending stream. The heavy overdrafters are likely to overdraw because they heavily discount their future consumptions.

3.2.2 Inattention

Next we explain the overdraft incentives for the light overdrafters with inattention. The idea is that consumers might be inattentively monitoring their checking accounts so that they are uncertain about the exact balance amount. Sometimes the perceived balance can be higher than the true balance and this might cause an overdraft. We first present a representative example of consumer inattention. The example is based upon our data, but to protect the privacy of the consumer and the merchants, amounts have been changed. However, the example remains representative of the underlying data.

Figure 2: Overdraft due to Balance Perception Error

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**

As shown in figure 2, the consumer first received her salary on August 17th. After a series of expenses she was left with \$21.16 on August 20th. As she had never checked her balance, she continued spending and overdrew her account for several small purchases, including a \$25 restaurant bill, a \$17.12 beauty purchase, a \$6.31 game and a \$4.95 coffee purchase. These four transactions added up to \$53.38 but caused her to pay four overdraft item fees, a total of \$124. We speculate that this consumer was careless in monitoring her account and overestimated her balance.

Beyond this example, we find more evidence of inattention in the data. Intuitively, a direct support of inattention is that the less frequent a consumer checks her balance, the more overdraft fee she pays. To test this hypothesis, we estimate the following specification:

$$TotODPmt_{it} = \beta_0 + \beta_1 BCFreq_{it} + \mu_i + v_t + \varepsilon_{it}$$

where for consumer i at time t (month), $TotODPmt_{it}$ is the total overdraft payment, $BCFreq_{it}$ is the balance checking frequency.

We estimate this model on light overdrafters (whose overdraft frequency is not in the top 20 percentile) and heavy overdrafters (whose overdraft frequency is in the top 20 percentile) separately and report the result in the column (1) and (2) in Table 4.

Table 4: Frequent Balance Checking Reduces Overdrafts for Light Overdrafters

	(1)	(2)	(3)
	Light Overdrafters	Heavy Overdrafters	All Overdrafters
Balance Checking	-0.5001^{***}	-0.00001389	-0.6823^{***}
Frequency (BCFreq, β_1)	(0.00000391)	(0.00000894)	(0.00000882)
Overdraft Frequency			16.0294^{***}
(ODFreq, β_2)			(0.00002819)
BCFreq$\times$ODFreq (β_3)			27.8136^{***}
			(0.00000607)
Number of Observations	53,845,039	17,810,276	71,655,315
R^2	0.1417	0.1563	0.6742

Note: Fixed effects at individual and day level; Robust standard errors, clustered at individual level.*p<0.01;**p<0.001;***p<0.0001

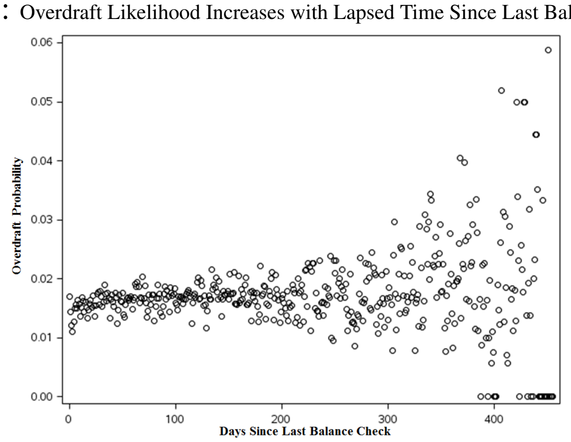
The result suggests that more balance checking decreases overdraft payment for light overdrafters but not for heavy overdrafters. We further test this effect by including overdraft frequency ($ODFreq_{it}$) and an interaction term of balance checking frequency and overdraft frequency $BCFreq_{it} \times ODFreq_{it}$ in the equation below. The idea is that if the coefficient for this interaction term is positive while the coefficient for balance checking frequency ($BCFreq_{it}$) is negative, then it implies that checking balances more often only decreases the overdraft payment for consumers who overdraw infrequently but not for those who do it with high frequency.

$$\begin{aligned} TotODPmt_{it} =& \beta_0 + \beta_1 BCFreq_{it} + \beta_2 ODFreq_{it} + \beta_3 BCFreq_{it} \times ODFreq_{it} \\ & + \mu_i + \nu_t + \varepsilon_{it} \end{aligned}$$

The result in column (3) of Table 4 confirms our hypothesis.

Interestingly, we find that a consumer's balance perception error accumulates overtime in the sense that the longer a consumer hasn't check balances, the more likely that she is going to overdraw and pay higher amount of overdraft fees. Figure 3 below exhibits the overdraft probability across number of days since a consumer checked balance last time for light overdrafters (whose overdraft frequency is not in the top 20 percentile). It suggests that the overdraft probability increases moderately with the number of days since the last balance check.

Figure 3: Overdraft Likelihood Increases with Lapsed Time Since Last Balance Check



We confirm this relationship with the following two specifications. We assume that overdraft incidence $I(OD)_{it}$ (where $I(OD)_{it} = 1$ denotes overdraft and $I(OD)_{it} = 0$ denotes no overdraft) and overdraft fee payment amount $ODFee_{it}$ for consumer i at time t can be modeled as:

$$I(OD)_{it} = \Phi\left(\rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} + \rho_2 BeginBal_{it} + \mu_i + v_t\right)$$
$$ODFee_{it} = \rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} + \rho_2 BeginBal_{it} + \mu_i + v_t + \varepsilon_{it}$$

where Φ is the cumulative distribution function for standard normal distribution. The term $DaysSinceLastBalanceCheck_{it}$ denotes the number of days consumer i hasn't checked her balance until time t and $BeginBal_{it}$ is the beginning balance at time t. We control for the beginning balance because it can be negatively correlated with the days since last balance check due to the fact that consumers tend to check when the balance is low and a lower balance usually leads to an overdraft.

Table 5: Reduced Form Evidence of Existance of Monitoring Cost

	$I(OD)$	$ODFee$
Days Since Last Balance Check (ρ_1)	0.0415***	0.0003***
	(0.00000027)	(0.00000001)
Beginning Balance (ρ_2)	−0.7265***	−0.0439***
	(0.00000066)	(0.00000038)
Individual Fixed Effect	Yes	Yes
Time Fixed Effect	Yes	Yes
Number of Observations	53,845,039	53,845,039
R^2	0.5971	0.6448

Note: The estimation sample only includes overdrafters. Marginal effects for the Probit model; Fixed effects at individual and day level; robust standard errors, clustered at individual level.*p<0.01;**p<0.001;***p<0.0001.

Table 5 reports the estimation results which support our hypothesis that the longer a consumer hasn't checked balance, the more likely she overdraws and the higher overdraft fee she pays.

Since checking balances can effectively help prevent overdrafts, why don't consumers do it often enough to avoid overdraft fees? We argue that it's because monitoring the account is costly in terms of time, effort and mental resources. Therefore, a natural consequence is that if there's a means to save consumers' time, effort or mental resources, the consumer will indeed check balances more frequently. We find such support from the data about online banking ownership. Specifically, for consumer i we estimate the following specification:

$$CheckBalFreq_i = \beta_0 + \beta_1 OnlineBanking_i + \beta_2 LowIncome_i + \beta_3 Age_i + \varepsilon_i$$

where $CheckBalFreq_i$ is the balance checking frequency, $OnlineBanking_i$ is online banking ownership (1 denotes the consumer has online banking while 0 denotes otherwise), $LowIncome_i$ is whether the consumer belongs to the low income group (1 denotes yes and 0 denotes no) and Age_i is age (in years).

Table 6: Reduced Form Evidence of Existance of Monitoring Cost

Dependent variable	Check Balance Frequency
Online Banking (β_1)	58.4245***
	(0.5709)
Low Income (β_2)	3.3812***
	(0.4178)
Age (β_3)	0.6474***
	(0.0899)
Number of Observations	$602,481$
R^2	0.6448

*p<0.01;**p<0.001;***p<0.0001.

Table 6 shows that after controlling for income and age, consumers with online banking accounts check the balance more frequently than those without, which suggests that monitoring costs exist and when they are reduced, consumers monitor more frequently.

3.2.3 Dissatisfaction

Table 7: Account Closure Frequency for Overdrafters vs Non-Overdrafters

Total	% Closed
Heavy Overdrafters	23.36%
Light Overdrafters	10.56%
Non-Overdrafters	7.87%

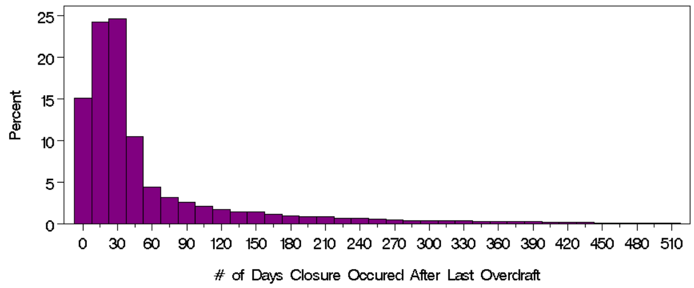
We also find that overdrafters are more likely to close their accounts (Table 7). Among non-overdrafters, 7.87% closed their accounts during the sample period. This ratio is much higher for overdrafters. Specifically, 23.36% of heavy overdrafters (whose overdraft frequency is in the top 20 percentile) closed their accounts, while 10.56% of light overdrafters (whose overdraft frequency is not in the top 20 percentile) closed their accounts.

Table 8: Closure Reasons

	Overdraft Forced Closure	Overdraft Voluntary Closure	No Overdraft Voluntary Closure
Heavy Overdrafters	86.34%	13.66%	–
Light Overdrafters	52.58%	47.42%	–
Non-Overdrafters	–	–	100.00%

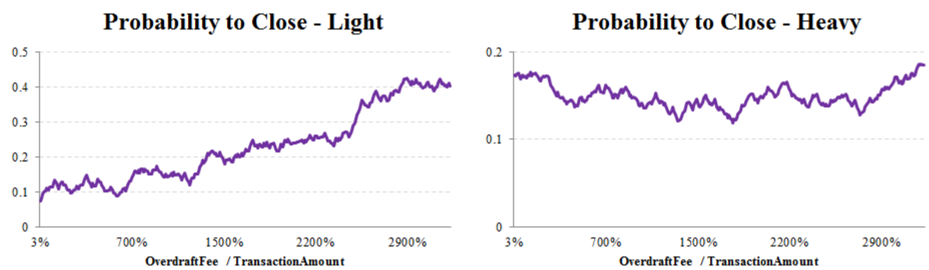
From the description field in the data, we can distinguish the cause of account closure: forced closure by the bank because the consumer is unable or unwilling to pay back the negative balances and the fee (charge-off) or voluntary closure by the consumer. Among heavy overdrafters, 13.66% closed voluntarily and the rest (86.34%) were forced to close by the bank (Table 8). In contrast, 47.42% of the light overdrafters closed their accounts voluntarily. We conjecture that the higher voluntary closures may be due to customer dissatisfaction with the bank, with evidence shown below.

Figure 4: Days to Closure After Last Overdraft



First, we find that overdrafters who closed voluntarily were very likely to close soon after the overdraft. In Figure 4 we plot the histogram of number of days it took the account to close after its last overdraft occasion. It shows that more than 60% of accounts closed within 30 days after the overdraft occasion.

Figure 5: Percentage of Accounts Closed Increases with Fee/Transaction Amount Ratio



Second, light overdrafters are also more likely to close their accounts when the ratio of overdraft fee over the transaction amount that caused the overdraft fee is higher. In other words, the more unfair the overdraft fee (higher ratio of overdraft fee over the transaction amount that caused the overdraft fee), the more likely it is that she will close the account. We show this pattern in the left panel of Figure 5. However, this effect doesn't seem to be present for heavy overdrafters (right panel of Figure 5).

The model free evidence indicate that consumer heavy discounting and inattention can help explain consumers' overdraft behaviors as consumers might be dissatisfied after being charged the overdraft fees. Below we'll build a structural model that incorporates consumer heavy discounting, inattention and dissatisfaction.

4 Model

We model a consumer's daily decision about non-preauthorized spending in her checking account. Alternatively we could describe this non-preauthorized spending as immediate or discretionary; not discretionary in the sense that economists traditionally use the term, but in the sense that immediate spending likely could have been delayed. To focus on rationalizing the consumer's overdraft behavior, we make the following assumptions. First, we abstract away from the complexity associated with our data and assume that the consumer's income and preauthorized spendings are exogenously given. We refer to preauthorized spending to mean those expenses for which the spending decision was made prior to payment. For example, a telephone bill or a mortgage due are usually arranged before the date that the actual payment occurs. We assume that decisions for preauthorized spending are hard to change on a daily basis after they are authorized and more likely to be related to consumption that has medium or long-run consequences. In contrast, non-preauthorized spending involves a consumer's frequent day-to-day decisions and the consumer can adjust the spending amount flexibly. We make this distinction because non-preauthorized spending is at the consumer's discretion and thus affects the overdraft outcome directly. To ease explanation, we use "coming bills" to represent preauthorized spending for the rest of the paper. Second, we allow the consumer to be inattentive to monitoring her account balance and coming bills. But she can decide whether to check her balance. When a consumer hasn't checked the balance, she comes up with an estimate of the available balance and forms an expectation about coming bills. If she makes a wrong estimate or expectation, she faces the risk of overdrawing her account. Last, as consumption is not observed in the data, we make a bold assumption that spending is equivalent to consumption in terms of generating utility. That is, the more a consumer spends, the more she consumes, the higher utility she obtains. In what follows, we use consumption and spending interchangeably.

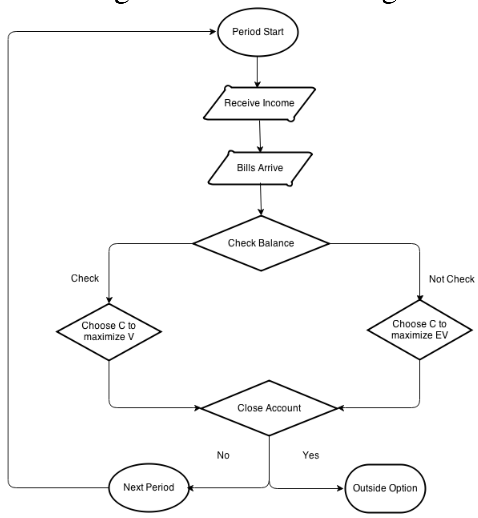
We'll describe the model in the next four parts: (1) timing, (2) basic model (3) inattention and balance checking and (4) dissatisfaction and account closing.

4.1 Timing

The timing of the model is as follows (Figure 6). On each day:

1. The consumer receives income, if there is any.
2. Her bills arrive if there is any.
3. Balance checking stage (CB): She decides whether to check her balance. If she checks, she incurs a cost and knows today's beginning balance and the bill amount. If not, she recalls an estimate of the balance and bill amount.
4. Spending stage (SP): She makes the discretionary spending decision (Choose C) to maximize total discounted utility V (or expected total discounted utility EV if she didn't check balance)for today and spends the money.
5. Overdraft fee is charged if the ending balance is below zero.
6. Account closing stage (AC): She decides whether to close the account (after paying the overdraft fee if there's any). If she closes the account, she receives an outside option. If she doesn't chose the account, she goes to 7.
7. Balance updates and the next day comes.

Figure 6: Model Timing



4.2 Basic Model

We assume the consumer's per-period consumption utility at time t is a constant relative risk averse utility (Arrow 1963):

$$u_C(C_t) = \frac{C_t^{1-\theta_t}}{1-\theta_t} \tag{1}$$

where θ_t is the relative risk averse coefficient which represents the consumer's preference about consumption. The higher θ_t, the higher utility the consumer can derive from a marginal unit of consumption.

$$\begin{aligned} \theta_t &= exp(\theta + \varepsilon_t) \\ \varepsilon_t &\sim N\left(0, \varsigma^2\right) \end{aligned}$$

As consumers' preference for consumption might change over time and the relative risk averse coefficient is always positive, we allow θ_t to follow a log-normal distribution. Essentially, θ_t is the exponential of the sum of a time-invariant mean θ and a random shock ε_t. The shocks capture unexpected needs for consumption and follow a normal distribution with mean 0 and variance ς^2 (Yao et. al. 2012).

Notice that the consumption plan C_t depends on the consumer's budget constraint, which further depends on her current balance B_t, income Y_t and future bills Ψ_t. For example, when the coming bill is for a small amount, the consumption can be higher than when the bill is for a large amount.

4.3 Inattention and Balance Checking

In practice, the consumer may not be fully attentive to her financial well-being. Because monitoring her account balance takes time and effort, she may not check her balance frequently. As a consequence, instead of knowing the exact (available) balance B_t [12], she recalls a perceived balance $\widetilde{B}_t$. Following Mehta, Rajiv and Srinivasan (2003), we allow the perceived balance $\widetilde{B}_t$ to be the sum of the true balance B_t and a perception error $\eta_t \omega_t$. The first component of the perception error η_t is a random draw from the standard normal distribution[13] and the second component is the standard deviation of the perception error, ω_t. So $\widetilde{B}_t$ follows a normal distribution

$$\widetilde{B}_t \sim N\left(B_t + \eta_t \omega_t, \omega_t^2\right)$$

The variance of the perception error ω_t^2 measures the extent of uncertainty. Based on the evidence from section 3.2.2, we allow this extent of uncertainty to accumulate through time which implies that the longer the consumer goes without checking her balance, the more inaccurate her perceived balance is. That is,

$$\omega_t^2 \quad = \quad \rho \Gamma_t \tag{2}$$

where Γ_t denotes the lapsed time since the consumer last checked her balance, and ρ denotes the sensitivity to lapsed time as shown in the equation (2) above[14]. Notice that the expected utility is decreasing in the variance of the perception error ω_t^2. This is true because the larger the variance of the perception error, the less accurate the consumer's estimate of her true balance, and the more likely she is going to mistakenly overdraw, which lowers her utility.

We further assume that the consumer is sophisticated inattentive[15] in the sense that she is aware of her own inattention (Grubb 2014). Sophisticated inattentive consumers are rational in that they choose to be inattentive due to the high cost of monitoring her balances from day-to-day. We also model the consumer's balance checking behavior. We denote the balance checking choice as $Q_t \in \{1,0\}$ where 1 means check and 0 otherwise. If a consumer checks her balance, she incurs a monitoring cost but knows exactly what her balance is. So the perception error is reduced to zero and she can make her optimal spending decision with all information. In mathematics form, her consumption utility function changes to

$$u_t = \frac{C_t^{1-\theta_t}}{1-\theta_t} - Q_t \xi + \chi_{tQ_t} \tag{3}$$

where ξ is her balance checking cost and χ_{Q_t} is the idiosyncratic shock that affects her balance checking cost. The shock χ_{tQ_t}can come from random events like a consumer checks balance because she's also performing other types of transactions (like online bill payments) or she is on vacation without access to any bank channels so it's hard for her to check balances. The equation

[12]Available balance means the initial balance plus income minus bills. For the ease of exposition, we omit the word "available" and only use "balance".

[13]The mean balance perception error $\bar{\eta}$ cannot be separately identified from the variance parameters ρ because the identification sources both come from consumers' overdraft fee payment. Specifically, the high overdraft payment for a consumer can be either explained by a positive balance perception error or large perception error variance caused by large ρ. So we fix $\bar{\eta}$ at zero, i.e. the perception error is assumed to be unbiased.

[14]We considered other specifications for the relationship between perception error variance and lapsed time since last balance check. Results remain qualitatively unchanged

[15]Consumers can also be naively inattentive, but we don't allow it here. See discussion in Grubb 2014.

implies that if the consumer checks her balance, then her utility decreases by a monetary equivalence of $[(1-\theta_t)\xi]^{\frac{1}{1-\theta_t}}$. We assume that χ_{tQ_t} are iid and follow a type I extreme value distribution.

If she doesn't check, she recalls her balance $\widetilde{B}_t$ with the perception error η_t. So her perceived balance is

$$\widetilde{B}_t \sim Q_t B_t + (1-Q_t) N\left(B_t + \eta_t \omega_t, \omega_t^2\right)$$

She forms an expected utility based on her knowledge about the distribution of her perception error. The optimal spending will maximize her "expected" utility after integrating out the balance perception error, which is

$$u_t = \int_{\widetilde{B}_t} \int_{\eta_t} u_t\left(C_t; \widetilde{B}_t\right) dF(\eta_t)\, dF\left(\widetilde{B}_t\right)$$

4.4 Dissatisfaction and Account Closing

We assume that the consumer also has the option of closing the account (e.g., an "outside option"). If she chooses to close the account, she might switch to other competing banks or become unbanked. With support from section 3.1, we make an assumption that consumers are sensitive to the ratio of the overdraft fee to the overdraft transaction amount and we use$\Xi_{it}$to denote this ratio as a state variable. We assume that the higher the ratio, the more likely it is that the consumer will be dissatisfied to close the account because the forward-looking consumer anticipates that she's going to accumulate more dissatisfaction (as well as lost consumption utility due to overdrafts) in the future so that it's not beneficial for her to keep the account open any more. Furthermore, we assume that consumers keep updating her belief of the ratio and only remembers the highest ratio that has ever incurred. That is, if we use Δ_t to denote the per-period ratio then

$$\Delta_t = \frac{OD_t}{|B_t - C_t|}$$

and

$$E\left[\Xi_{t+1} | \Xi_t\right] = max\left(\Xi_t, \Delta_t\right)$$

This assumption reflects a consumer's learning behavior over time in the sense that after experiencing many overdrafts, a consumer realizes how costly (or dissatisfied) it could be for her to keep the account open. When she learns that the ratio can be high enough so that it's not beneficial for her to keep the account open any more, she'll choose to close the account. Specifically, we add the dissatisfaction effect to the per-period utility function where

$$U_t = u_t - \Upsilon * \Delta_t * I[B_t - C_t < 0]$$

In the above equation, u_t is defined in equation 3 and Υ is the dissatisfaction sensitivity, i.e., the impact of charging an overdraft fee on a consumer's decision to close the account.

We assume that closing the account is a termination decison. Once a consumer chooses to close the account, her value function (or total discounted utility function) equals an outside option with a value normalized to 0 for identification purposes.[16]If the consumer keeps the account open, she'll receive continution values from future per-period utility functions. More specifically, let W denote the choice to close the account, where $W = 1$ is closing the account and $W = 0$ is keeping the account open. Then the value function for the consumer becomes

$$V_t = \begin{cases} U_t + \varpi_{t0} + \beta E\left[V_{t+1}|S_t\right] & \text{if } W_t = 0 \\ U_t + \varpi_{t1} & \text{if } W_t = 1 \end{cases}$$

where ϖ_{t0} and ϖ_{t1} are the idiosyncratic shocks that determine a consumer's account closing decision. Sources of the shocks may include (1) the consumer moved address; (2) competing bank entered the market, and so on. We assume these shocks follow a type I extreme value distribution.

4.5 State Variables and the Transition Process

We have explained the following state variables in the model: (beginning) balance B_t, income Y_t, coming bill ψ_t, lapsed time since last balance check Γ_t, overdraft fee OD_t, ratio of overdraft fee to the overdraft transaction amount Ξ_t, preference shock ε_t, balance checking cost shock χ_t and account closure utility shock ϖ_t. The other state variable to be introduced later, DL_t, is involved in the transition process.

For (available) balance B_t, the transition process satisfies the consumer's budget constraint, which is

$$B_{t+1} = B_t - C_t - OD_t * I\left(B_t - C_t < 0\right) + Y_{t+1} - \psi_{t+1}$$

where OD_t is the overdraft fee. As we model the consumer's spending decision at the daily level rather than transaction level, we aggregate all overdraft fees paid and assume the consumer knows the per-item fee structure stated in section 3. This assumption is realistic in our setting because we have already distinguished between inattentive and attentive consumers. The argument that a consumer might not be fully aware of the per-item fee is indirectly captured by the balance perception error in the sense that the uncertain overdraft fee is equivalent to the uncertain balance because they both tighten the consumer's budget constraint. As for the attentive consumer who overdraws because of heavy discounting, she should be fully aware of the potential cost of overdraft. So in both cases we argue that the assumption of a known total overdraft fee is reasonable.

The state variable OD_t is assumed to be iid over time and to follow a discrete distribution with support vector and probability vector $\{X, p\}$. The support vector contains multiples of the per-item overdraft fee.

Consistent with our data, we assume an income distribution as follows

$$Y_t = Y * I\left(DL_t = PC\right)$$

where Y is the stable periodic (monthly/weekly/biweekly) income, DL_t is the number of days left until the next payday and PC is the length of the pay cycle. The transition process of DL is

[16]Although the outside option is normalized to zero for all consumers, the implicit assumption is that we allow for heterogeneous utility of the outside option. The heterogeneity is reflected by the other structural parameters, including the dissatisfaction sensitivity.

deterministic

$$DL_{t+1} = DL_t - 1 + PC * I\left(DL_t = 1\right)$$

where it decreases by one for each period ahead and goes back to the full length when one pay cycle ends.

The coming bills are assumed to be iid draws from a compound Poisson distribution with arrival rate ϕ and jump size distribution G, $\Psi_t \sim CP(\phi, G)$. This distribution can capture the pattern of bills arriving randomly according to a Poisson process and bill sizes are sums of fixed components (each separate bill).

The time since last checking the balance also evolves deterministically based on the balance checking behavior. Formally, we have

$$\Gamma_{t+1} = 1 + \Gamma_t\left(1 - Q_t\right)$$

which means that if the consumer checks her balance in the current period, then the lapsed time goes back to 1 but if she doesn't check, the lapsed time accumulates by one more period.

The ratio of the overdraft fee to the overdraft transaction amount evolves by keeping the maximum amount over time.

$$E\left[\Xi_{t+1} | \Xi_t\right] = max\left(\Xi_t, \Delta_t\right)$$

The shocks ε_t, χ_t and ϖ_t are all assumed to be iid over time.

In summary, the whole state space for consumer is $S_t = \left\{\widetilde{B}_t, \Psi_t, Y_t, DL_t, OD_t, \Gamma_t, \Xi_t, \varepsilon_t, \chi_t, \varpi_t\right\}$. In our dataset, we observe $\hat{S}_t = \{B_t, \psi_t, Y_t, DL_t, OD_t, \Gamma_t, \Xi_t\}$ and our unobservable state variables are $\widetilde{S}_t = \left\{\widetilde{B}_t, \eta_t, \varepsilon_t, \chi_t, \varpi_t\right\}$. $S_t = \hat{S}_t \cup \widetilde{S}_t \cap \{B_t, \psi_t\}$. Notice here that consumers also have unobserved states B_t and ψ_t due to inattention, which means that the consumer doesn't know the true balance (B_t) or the bill amount (ψ_t) if she doesn't check her balance but only the perceived balance ($\widetilde{B}_t$) and expected bill (Ψ_t).

4.6 The Dynamic Optimization Problem and Intertemporal Tradeoff

The consumer chooses an infinite sequence of decision rules $\{C_t, Q_t, W_t\}_{t=1}^{\infty}$in order to maximize the expected total discounted utility:

$$\max_{\{C_t, Q_t, W_t\}_{t=0}^{\infty}} E_{\{S_t\}_{t=1}^{\infty}} \left\{U_0\left(C_0, Q_0, W_0; S_0\right) + \sum_{t=1}^{\infty} \beta^t U_t\left(C_t, Q_t, W_t; S_t\right) | S_0\right\}$$

where

$$U_t\left(C_t, Q_t, W_t; S_t\right) = \left[\int_{\widetilde{B}_t}\int_{\eta_t}\left\{\frac{C_t^{1-\theta_t}}{1-\theta_t} - Q_t\xi + \chi_{tQ_t}\right\}dF\left(\eta_t\right)dF\left(\widetilde{B}_t\right) - \Upsilon\frac{OD_t * I[B_t - C_t < 0]}{|B_t - C_t|} + \varpi_{t0}\right]\left(1 - W_t\right) + W_t\varpi_{t1}$$

.

Let $V\left(S_t\right)$ denote the value function:

$$V\left(S_t\right) = \max_{\{C_\tau, Q_\tau, W_\tau\}_{\tau=t}^{\infty}} E_{\left\{\{S_\tau\}_{\tau=t+1}^{\infty}\right\}} \left\{U_t\left(S_t\right) + \sum_{\tau=t+1}^{\infty} \beta^{\tau-t} U_\tau\left(S_\tau\right) | S_t\right\} \tag{4}$$

according to Bellman (1957), this infinite period dynamic optimization problem can be solved through the Bellman Equation

$$V(S_t) = \max_{C,Q,W} E_{S_{t+1}} \{U(C,Q,W;S_t) + \beta V(S_{t+1}) | S_t\} \tag{5}$$

In the infinite horizon dynamic programming problem, the policy function doesn't depend on time. So we can eliminate the time subscript. Then we have the following choice specific value function:

$$\begin{aligned}
& v\left(C,Q,W;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\right) \\
&= \begin{cases}
u_C(C) - \xi + \chi_1 - \Upsilon \frac{OD*I[B-C<0]}{|B-C|} + \varpi_0 & \\
+\beta E_{S_{+1}}\left[V\left(\widetilde{B}_{+1},\Psi_{+1},Y_{+1},DL_{+1},OD_{+1},1,\Xi_{+1}\varepsilon_{+1},\chi_{+1},\varpi_{+1}\right)\right] & \text{if } Q=1\&W=0 \\
\int_{\widetilde{B}_t}\int_{\eta_t}\left[u_C(C)+\chi_0\right]dF(\eta_t)\,dF\left(\widetilde{B}_t\right) - \Upsilon \frac{OD*I[B-C<0]}{|B-C|} + \varpi_0 & \\
+\beta E_{S_{+1}}\left[V\left(\widetilde{B}_{+1},\Psi_{+1},Y_{+1},DL_{+1},OD_{+1},\Gamma+1,\Xi_{+1},\varepsilon_{+1},\chi_{+1},\varpi_{+1}\right)\right] & \text{if } Q=0\&W=0 \\
\varpi_1 & \text{if } W=1
\end{cases}
\end{aligned} \tag{6}$$

where subscript$+1$ denotes the next time period. So the optimal policy is given by the following solution

$$\{C^*,Q^*,W^*\} = \arg\max v\left(C,Q,W;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\right)$$

One thing that's worth noticing is that there's a distinction between this dynamic programming problem and traditional ones. Because of the perception error, the consumer observes $\widetilde{B}_t = B_t + \eta_t \omega_t$ but doesn't know B_t or η_t. She only knows the distribution $N(B_t + \eta_t\omega_t, \omega_t^2)$. The consumer makes a decision $C^*\left(\widetilde{B}_t\right)$ based on the perceived balance $\widetilde{B}_t$. But as researchers, we don't know the realized perception error η_t. We observe the true balance B_t and the consumer's spending $C^*\left(\widetilde{B}_t\right)$. So we can only assume $C^*\left(\widetilde{B}_t\right)$ maximizes the "expected ex-ante value function". Later we look for parameters such that the likelihood for $C^*\left(\widetilde{B}_t\right)$ maximizes the expected ex-ante value function attains maximum. Following Rust (1987), we obtain the ex-ante value function which integrates out the cost shocks, preference shocks, account closing shocks and unobserved mean balance error.

$$EV(B,\psi,Y,DL,OD,\Gamma,\Xi) = \int_{\varpi}\int_{\chi}\int_{\varepsilon}\int_{\eta} v\left(C^*,Q^*,W^*;\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,\varepsilon,\chi,\varpi\right) d\eta\, d\varepsilon\, d\chi\, d\varpi$$

Consumers' intertemporal trade-offs are associated with the three dynamic decisions. First of all, given the budget constraint, a consumer will evaluate the utility of spending (or consuming) today versus tomorrow. The higher amount she spends today, the lower amount she can spend tomorrow. So spending is essentially a dynamic decision and the optimal choice for the consumer is to smooth out consumption over the time. Second, when deciding when to check balance, the consumer will compare the monitoring cost with the expected gain from avoiding the overdraft fee. She'll only check when the expected overdraft fee is higher than her monitoring cost. As the consumer's balance perception error might accumulate with time, the consumer's overdraft probability also increases with the lapse time since the last balance check. As a result, the consumer will wait until the overdraft probability reaches the certain threshold (when the expected overdraft fee equals the monitoring cost) to check the balance. Finally, the decision to close the account is an optimal stopping problem. The consumer will compare the total discounted utility of keeping the account with the utility from the outside option to decide when to close the account. When expecting too much overdraft fees as well as the accompanied dissatisfaction, the consumer will find it more attractive to take the outside option and close the account.

4.7 Heterogeneity

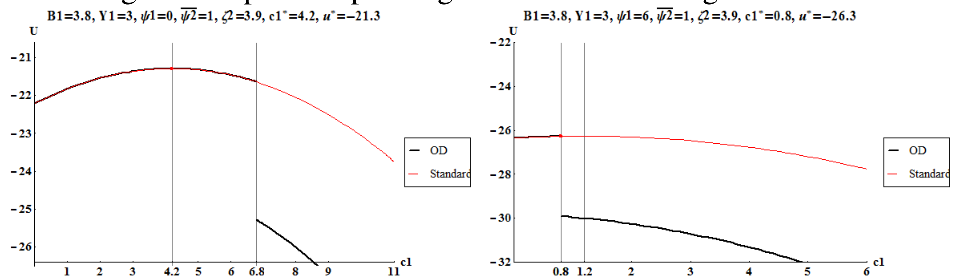
In our data, consumers exhibit different responses to their state conditions. For example, some consumers have never checked their balances and frequently overdraw while other consumers frequently check their balances and rarely overdraw. We hypothesize that it's due to their heterogeneous discount factors and monitoring costs. Therefore, our model needs to account for unobserved heterogeneity. We follow a hierarchical Bayesian framework (Rossi, McCulloch and Allenby 2005) and incorporate heterogeneity by assuming that all parameters: β_i (discount factor), ς_i (standard deviation of risk averse coefficient),ξ_i (monitoring cost), ρ_i (sensitivity of error variance to lapsed time since last checking balance) and Υ_i (dissatisfaction sensitivity) have a random coefficient specification. For each of these parameters, $\vartheta \in \{\beta_i, \varsigma_i, \lambda_i, \xi_i, \rho_i\}$, the prior distribution is defined as $\vartheta \sim N\left(\mu_\vartheta, \sigma_\vartheta^2\right)$. The hyper-prior distribution is assumed to be diffuse.

4.8 Numerical Example

Here we use a numerical example to show that inattention can explain the observed overdraft occasions in the data. More importantly, we display an interesting case in which an unbiased perception can make the consumer spend less than the desired level. In this example, there are two periods, $t \in \{1,2\}$. The consumer chooses the optimal consumption to maximize the expected total discounted utility. In order to obtain an analytical solution for the optimal spending, we assume a CARA utility $u_C(C_t) = \frac{1}{\theta} exp(-\theta C_t)$ and the coming bill following a normal distribution $\Psi_2 \sim N\left(\bar{\psi}_2, \zeta_2^2\right)$. The initial balance is B_1 and the consumer receives income Y_1 and Y_2. As period 2 is the termination period, the consumer will spend whatever is left from period 1, i.e., $C_2 = B_1 + Y_1 - \psi_1 - C_1 - OD * (B_1 + Y - \psi_1 - C_1 - \psi_2) + Y_2 - \psi_2$. So the only decision is how much to spend for period 1: C_1. Let $\theta = 0.07$, $B_1 = 3.8$, $Y_1 = 3$, $Y_2 = 3$, $\overline{\psi_2} = 1$,$\zeta_2 = 3.9$,$\beta = 0.99$, $OD = 3.58$ (The values seem small compared to spending in reality because we apply log to all monetary values).

4.8.1 Effect of Overdraft

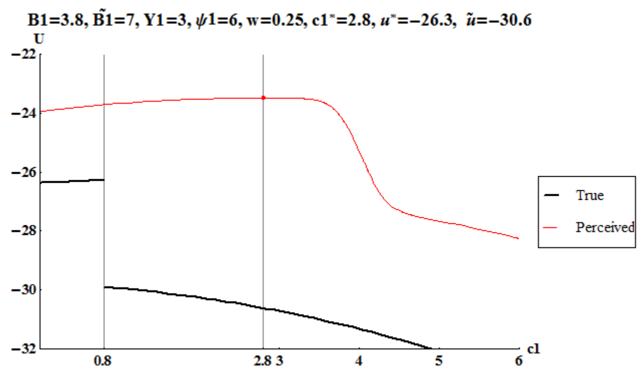
Figure 7: Optimal Spending with Neutral vs Negative Shock



In this example in Figure 7, when there's no bill to pay in the first period ($\psi_1 = 0$ in the left panel), the total budget for the consumer is 6.8 and she would like to spend 4.2 to attain the maximum utility. However, when she has to pay for a bill of 6 (right panel), she is left with only 0.8. Her optimal choice is to spend 0.8 and just clear the budget because the disutility of overdraft (utility function with overdraft is the black line labeled as OD) is too high. This example shows that since the overdraft fee is equivalent to an extremely high interest rate short-term loan, the consumer wouldn't want to overdraw her account.

4.8.2 Effect of Inattention–Overdraft

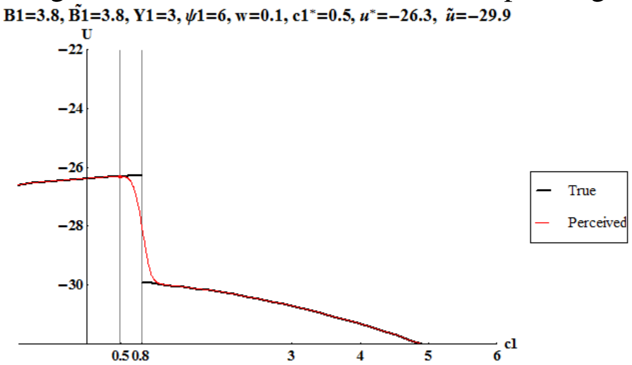
Figure 8: Inattention Leads to Overdraft–Balance Error $\widetilde{B_1} > B_1$



In a different scenario (Figure 8), if the consumer overestimates her balance to be 7 (her true balance is 3.8), i.e., she has a positive perception error regarding her true balance, then she would spend 2.8 which is the optimal amount based on this misperception. This perception error leads her to an overdraft.

4.8.3 Effect of Inattention–Error Constraints Spending

Figure 9: Inattention Leads to Underspending



Finally, we discover an interesting case where inattention may cause the consumer to spend less than her optimal spending level. This happens because the consumer knows that she is inattentive, i.e., she might overestimate her effective balance to run into overdraft. In order to prevent this, the consumer tends to constrain her spending. As shown in Figure 9, though the optimal spending is 0.8 as in the previous example (section 4.8.1), the inattentive consumer chooses to spend 0.5 to prevent overdraft. This example suggests a new revenue source for the bank. If the bank provides automatic alerts to consumers to inform them of their exact balances, the consumers won't have to take precautions to avoid overdrafts. As a consequence, consumers will spend more and the bank can benefit from the increased interchange fees.

5 Identification and Estimation

We now discuss the identification of the parameters and the estimation procedure.

5.1 Identification

The unknown structural parameters in the model include $\{\theta, \beta, \varsigma, \xi, \rho, \Upsilon\}$ where θ is the logarithm of the mean risk averse coefficient, β is the discount factor, ς is the standard deviation of the risk averse coefficient, ξ is the monitoring cost, ρ is the sensitivity of balance error variance to the lapsed time since last balance checking, and Υ is the dissatisfaction sensitivity. Next we provide an informal rationale for identification of each parameter.

First of all, as we know from Rust (1987), the discount factor β cannot be separately identified from the static utility parameter, which in our case, the risk aversion coefficient θ. The reason is that lowering θ tends to increase consumption/spending, an effect which can also be achieved by lowering β. As we are more interested in the consumers' time preference rather than risk preference, we fix the risk averse coefficient θ, which allows me to identify the discount factor[17].

[17] We also tried to fix the discount factor (at 0.9998) and estimate the risk averse coefficients. Other structural parameter estimates are not significantly unaffected under this specification. Our results confirm that the risk averse coefficient and the discount factor are mathematically substitutes (Andersen et al. 2008). Estimation results with fixed

This practice is also used in Gopalakrishnan, Iyengar, Meyer 2014. As to the risk averse coefficient, we choose $\theta = 0.74$, following the latest literature by Andersen et al. (2008) where they jointly elicit risk and time preferences[18]. After fixing θ, β_i can be well identified by the sequences of consumption (spending) within a pay period. A large discount factor (close to 1) implies a stable consumption stream while a small discount factor implies a downward sloping consumption stream. Because a discount factor is constrained above by 1, we do a transformation to set $\beta_i = \frac{1}{1+exp(\lambda_i)}$ and estimate λ_i instead.

Second, the standard deviation of risk averse coefficient ς_i is identified by the variation of consumptions on the same day of the pay period but across different pay periods.

Moreover, according to the intertemporal tradeoff, the longer the consumer goes without checking her balance, the more likely she will be to overdraw due to the balance error. The observed data pattern of more overdraft fees paid longer after a balance checking inquiry can help pin down the structural parameters ρ_i.

Intuitively, the monitoring cost ξ_i is identified by the expected overdraft payment amount. Recall that the tradeoff regarding balance checking is that a consumer only checks balance when ξ_i is smaller than the expected overdraft payment amount. In the data we observe the balance checking frequency. Combining this with the calculated ρ_i we can compute the expected overdraft probability and further the expected overdraft payment amount, which is the identified ξ_i. Given ρ_i, a consumer with few balance checking inquiries must have a higher balance checking cost ξ_i.

Lastly, the dissatisfaction sensitivity parameter Υ_i can be identified by the data pattern that consumers' account closure probability varies with the ratio of overdraft fee over the overdraft transaction amount, as shown in section 3.1.

Note that aside from these structural parameters, there is another set of parameters that govern the transition process. These parameters can be identified prior to structural estimation from the observed state variables in our data. The set includes $\{\phi, G, X, p\}$.

In sum, the structural parameters to be estimated include $\{\lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i\}$.

5.2 Likelihood

The full likelihood function is

$$\begin{aligned} & L\left(\left\{\left\{C_{it}, Q_{it}, W_{it}; \hat{S_{it}}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) \\ = & \left(\left\{L\left\{C_{it}, Q_{it}, W_{it}; \hat{S_{it}}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{f\left\{\hat{S_{it}}|\hat{S}_{it-1}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right) L\left(\left\{\hat{S}_{i0}\right\}_{i=1}^{I}\right) \end{aligned}$$

where $\hat{S}_{it} = \{B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_t\}$. As the likelihood for the optimal choices and that for the state transition process are additively separable when we apply log to the likelihood function, we can first estimate the state transition process from the data, then maximize the likelihood for the optimal choices. The likelihood function for the optimal choice is

discount factor are available upon requests.

[18] We also tried other values for the relative risk averse coefficient θ, the estimated discount factor β values change with different θ's, but other structural parameter values remain the same. The policy simulation results are also robust with different values of θ's.

$$
\begin{aligned}
&L\left(\left\{\left\{C_{it},Q_{it},W_{it};\hat{S}_{it}\right\}_{t=1}^{T}\right\}_{i=1}^{I}\right)\\
&=\prod_{i=1}^{L}\prod_{t=1}^{T}L\left(C_{it};\hat{S}_{it}\right)L\left(Q_{it};\hat{S}_{it}\right)L\left(W_{it};\hat{S}_{it}\right)\\
&=\prod_{i=1}^{L}\prod_{t=1}^{T}f\left(\varepsilon_{it}|C_{it}\right)Pr\left(\chi_{it}|Q_{it},C_{it}\right)Pr\left(\varpi_{it}|W_{it},Q_{it},C_{it}\right)
\end{aligned}
$$

where $f(\varepsilon_{it}|C_{it})$ is estimated from the normal kernel density estimator to be explained in section 5.3.1, $Pr(\chi_{it}|C_{it},Q_{it})$ and $Pr(\varpi_{it}|C_{it},Q_{it},W_{it})$ follow the standard logit model given the choice specific value function in equation 6. In specific,

$$
Pr\left(Q_{it}=1;\hat{S}_{it}\right)=\int_{\varpi_{it}}\int_{\varepsilon_{it}}\int_{\eta_{it}}\frac{exp\left\{v\left(C_{it},Q_{it}=1,W_{it};\hat{S}_{it}\right)\right\}}{\sum_{Q_{it}}exp\left\{v\left(C_{it},Q_{it},W_{it};\hat{S}_{it}\right)\right\}}
$$

$$
Pr\left(W_{it}=1;\hat{S}_{it}\right)=\int_{\chi_{it}}\int_{\varepsilon_{it}}\int_{\eta_{it}}\frac{exp\left\{v\left(C_{it},Q_{it},W_{it}=1;\hat{S}_{it}\right)\right\}}{\sum_{W_{it}}exp\left\{v\left(C_{it},Q_{it},W_{it};\hat{S}_{it}\right)\right\}}
$$

5.3 Estimation: Imai, Jain and Ching (2009)

5.3.1 Modified IJC

We use the Bayesian estimation method developed by Imai, Jain and Ching (2009) to estimate the dynamic choice problem with heterogeneous parameters. As our model involves a continuous choice variable, spending, we adjust the IJC algorithm [19] to obtain the choice probability through kernel density estimation. We now show the details of the estimation procedure. The whole parameter space is divided into two sets ($\Omega=\{\Omega_1,\Omega_2\}$), where the first one contains hyper-parameters in the distribution of the heterogeneous parameters ($\Omega_1=\left\{\mu_\lambda,\mu_\varsigma,\mu_\xi,\mu_\rho,\mu_\Upsilon,\sigma_\lambda,\sigma_\varsigma,\sigma_\xi,\sigma_\rho,\sigma_\Upsilon\right\}$), and the second set contains heterogeneous parameters ($\Omega_2=\{\lambda_i,\varsigma_i,\xi_i,\rho_i,\Upsilon_i\}_{i=1}^{I}$) . We allow all heterogeneous parameters (represented by ϑ_i) to follow a normal distribution with parameters mean μ_ϑ and standard deviation σ_ϑ. Let the observed choices be $O^d=\left\{O_i^d\right\}_{i=1}^{I}=\left\{C_i^d,Q_i^d,W_i^d\right\}$ where $C_i^d\equiv\left\{C_{it}^d,\forall t\right\}$, $Q_i^d\equiv\left\{Q_{it}^d,\forall t\right\}$ and $W_i^d\equiv\left\{W_{it}^d,\forall t\right\}$.

Each MCMC iteration mainly consists of two blocks.

(i) Draw Ω_1^r , that is, draw $\mu_\vartheta^r\sim f_{\mu_\vartheta}(\vartheta|\sigma_\vartheta^{r-1},\Omega_2^{r-1})$ and $\sigma_\vartheta^r\sim f_{\sigma_\vartheta}(\sigma_\vartheta|\mu_\vartheta^r,\Omega_2^{r-1})$ ($\vartheta\in\{\lambda,\varsigma,\xi,\rho,\Upsilon\}$, the parameters that capture the distribution of ϑ for the population) where f_{μ_ϑ} and f_{σ_ϑ} are the conditional posterior distributions.

(ii) Draw Ω_2^r , that is, draw individual parameters $\vartheta_i\sim f_i\left(\vartheta_i|O_i^d,\Omega_1^r\right)$ by the Metropolis-Hastings (M-H) algorithm.

More details of the estimation algorithm is presented in Appendix A.2.

[19] The IJC method is designed for dynamic discrete choice problems. Zhou (2012) also applied it to a continuous choice problem.

5.3.2 Parallel Computing: Neiswanger, Wang and Xing (2014)

We adopt the parallel computing algorithm by Neiswanger, Wang and Xing (2014) to estimate our model with data from more than 500,000 consumers. The logic behind this algorithm is that the full likelihood function is a multiplicative of the individual likelihood.

$$p\left(\vartheta|x^N\right) \;\propto\; p\left(\vartheta\right)p\left(x^N|\vartheta\right) = p\left(\vartheta\right)\prod_{i=1}^{N} p\left(x_i|\vartheta\right)$$

So we can partition the data onto multiple machines, and then perform MCMC sampling on each using only the subset of data on that machine (in parallel, without any communication). Finally, we can combine the subposterior samples to algorithmically construct samples from the full-data posterior.

In details, the procedure is:

(1) Partition data x^N into M subsets $\{x^{n_1},...,x^{n_M}\}$.

(2) For $m = 1,...,M$ (in parallel):

(a) Sample from the subposterior p_m, where $p_m\left(\vartheta|x^{n_m}\right) \propto p\left(\vartheta\right)^{\frac{1}{M}} p\left(x^{n_m}|\vartheta\right)$

(3) Combine the subposterior samples to produce samples from an estimate of the subposterior density product $p_1...p_M$, which is proportional to the full-data posterior, i.e. $p_1...p_M\left(\vartheta\right) \propto p\left(\vartheta|x^N\right)$.

Given T samples $\{\vartheta_t\}_{t=1}^{T}$ from a subposterior p_m, we can write the kernel density estimator as $\hat{p_m}\left(\vartheta\right)$,

$$\begin{aligned}
\hat{p_m}\left(\vartheta\right) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{h^d}K\left(\frac{||\vartheta-\vartheta_t||}{h}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(2\pi h^2\right)^{-\frac{d}{2}}|I_d|^{-\frac{1}{2}}exp\left\{-\frac{1}{2h^2}\left(\vartheta-\vartheta_t\right)' I_d^{-1}\left(\vartheta-\vartheta_t\right)\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}N\left(\vartheta|\vartheta_t,h^2 I_d\right)
\end{aligned}$$

where we have used a Gaussian kernel with bandwidth parameter h. After we have obtained the kernel density estimator $\hat{p_m}\left(\vartheta\right)$ for M subposteriors, we define our nonparametric density product estimator for the full posterior as

$$\begin{aligned}
& \widehat{p_1\cdots p_m}\left(\vartheta\right) \\
=\; & \hat{p}_1\cdots\hat{p_m}\left(\vartheta\right) \\
=\; & \frac{1}{T^M}\sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}\prod_{m=1}^{M}N\left(\vartheta|\vartheta_{t_m}^m,h^2 I_d\right) \\
\propto\; & \sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}N\left(\vartheta|\bar{\vartheta}_{t\cdot},\frac{h^2}{M}I_d\right)\prod_{m=1}^{M}N\left(\vartheta_{t_m}^m|\bar{\vartheta}_{t\cdot},h^2 I_d\right) \\
\propto\; & \sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}w_{t\cdot}N\left(\vartheta|\bar{\vartheta}_{t\cdot},\frac{h^2}{M}I_d\right)
\end{aligned}$$

This estimate is the probability density function (pdf) of a mixture of TM Gaussians with unnormalized mixture weights $w_{t\cdot}$. Here, we use $t\cdot = \{t_1, \ldots, t_M\}$ to denote the set of indices for the M samples $\left\{\vartheta_{t_1}^1, \ldots, \vartheta_{t_M}^M\right\}$ (each from one machine) associated with a given mixture component, and let

$$\begin{aligned} w_{t\cdot} &= N\left(\vartheta | \bar{\vartheta}_{t\cdot}, \frac{h^2}{M} I_d\right) \\ \bar{\vartheta}_{t\cdot} &= \frac{1}{M} \sum_{m=1}^{M} \vartheta_{t_m}^m \end{aligned}$$

(4) Given the hierarchical Bayes framework, after obtaining the posterior distribution of the population parameter ϑ, use M-H algorithm once more to obtain the individual parameters (details in Appendix A.2 Step 4)

The sampling algorithm is presented in Appendix A.3.

6 Results

6.1 Model Comparison

Table 9: Model Comparison

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

We compare our model against the other four benchmark models in order to investigate the contribution of each element of the structural model. Models A to C are our proposed model without forward-looking, inattention and unobserved heterogeneity respectively and model D is our proposed model. Table 9 shows the log-marginal density (Kass and Raftery 1995) and the hit rate for overdraft, check balance and close account incidences (we only consider when these events happen because no event takes place the majority of the time). All four measures show that our proposed model significantly outperforms the benchmark models. Notably inattention contributes the most to model fit which is consistent with our conjecture in section 3.

6.2 Value of Parallel IJC

Table 10: Estimation Time Comparison

Size\Method (seconds)	**Parallel IJC**	IJC	CCP	FIML
1,000	518	1579	526	5,010
10,000	3,199	12,560	4,679	54,280
100,000	4,059	14,0813	55,226	640,360
>500,000	5,308	788,294	399,337	3,372,660
	(1.5 hr)	(9 days)	(5 days)	(39 days)

Table 11: Monte Carlo Results when N=100,000

Var	True Value		Parallel IJC	IJC	CCP	FIML
μ_β	0.9	Mean	0.878	0.883	0.851	0.892
		Std	0.041	0.039	0.036	0.025
μ_ς	1.5	Mean	1.505	1.502	1.508	1.501
		Std	0.131	0.124	0.199	0.103
μ_ξ	0.5	Mean	0.482	0.507	0.515	0.502
		Std	0.056	0.039	0.071	0.044
μ_ρ	1	Mean	1.006	1.003	1.015	1.002
		Std	0.027	0.022	0.026	0.019
μ_Υ	5	Mean	5.032	5.011	4.943	4.987
		Std	0.023	0.010	0.124	0.008
σ_β	0.1	Mean	0.113	0.095	0.084	0.104
		Std	0.016	0.014	0.015	0.010
σ_ς	0.3	Mean	0.332	0.318	0.277	0.309
		Std	0.024	0.015	0.029	0.021
σ_ξ	0.1	Mean	0.112	0.091	0.080	0.090
		Std	0.055	0.029	0.025	0.025
σ_ρ	0.1	Mean	0.107	0.107	0.085	0.105
		Std	0.008	0.006	0.010	0.006
σ_Υ	0.1	Mean	0.092	0.109	0.111	0.100
		Std	0.014	0.013	0.021	0.009

We report the computational performance of different estimation methods in Table 10. All the experiments are done on a server with an Intel Xeon CPU, 144 cores and 64 GB RAM. The first column is the performance of our proposed method, IJC with parallel computing. We compare it with the original IJC method, the Conditional Choice Probability (CCP) method by Arcidiacono and Miller (2011) [20] and the Full Information Maximum Likelihood (FIML) method by Rust (1987) (or Nested Fixed Point Algorithm) [21]. As the sample size increases, the comparative advan-

[20]We use the finite mixture model to capture unobserved heterogeneity and apply the EM algorithm to solve for the unobserved heterogeneity. More details of the estimation results can be obtained upon requests.

[21]We use the random coefficient model to capture unobserved heterogeneity. More details of the estimation results can be obtained upon requests.

tage of our proposed method is more notable. To run the model on the full dataset with more than 500,000 accounts takes roughly 1.5 hours compared to 9 days with the original IJC method. The reason for the decrease in computing time is that our method takes advantage of multiple machines that run in parallel. We further run a simulation study to see if the various methods are able to accurately estimate all parameters. Table 11 shows that different methods produce quite similar estimates and all mean parameter estimates are within two standard errors of the true values. The Parallel IJC method is slightly less accurate than the original IJC method.

The parallel IJC is almost 600 times faster than FIML. This happens because the full solution method solves the dynamic programming problem at each candidate value for the parameter estimates, whereas this IJC estimator only evaluates the value function once in each iteration.

6.3 Parameter Estimates

Table 12: Structural Model Estimation Results

Var	Interpretation	Mean (μ_ϑ)	Standard deviation (σ_ϑ)
β_i	Discount factor	0.9997	0.362
		(0.00005)	(0.058)
ς_i	Standard deviation of relative risk aversion	0.257	0.028
		(0.014)	(0.003)
ξ_i	Monitoring cost	0.708	0.255
		(0.084)	(0.041)
ρ_i	Inattention Dynamics–lapsed time	7.865	0.648
		(0.334)	(0.097)
Υ_i	Dissatisfaction Sensitivity	5.479	1.276
		(1.329)	(0.109)

Table 12 presents the results of the structural model. We find that the higher the age, the more risk averse the consumer is. The monitoring cost is estimated to be 0.708. Using the risk averse coefficient, we can evaluate the monitoring cost in monetary terms. It turns out to be \$2.03. We also obtained the cost measure for each individual consumer.

The variance of the balance perception error increases with the lapsed time since the last time to check balance and with the mean balance level. Notably the variance of the balance perception error is quite large. If we take the average number of days to check the balance from the data, which is 9, then the standard deviation is $7.865 * 9 = 70.79$. This suggests a very widely spread distribution of the balance perception error.

The estimated dissatisfaction sensitivity parameter confirms our hypothesis that consumers can be strongly affected by the bank fee and close the account as a consequence of dissatisfaction. If we consider an average overdraft transaction amount at \$33, then the relative magnitude of the effect of dissatisfaction is comparable to \$171. This suggests that unless the bank would like to offer a \$171 compensation to the consumer, the dissatisfied consumer will close the current account and switch. Moreover, consistent with the evidence in Figure 5, the dissatisfaction sensitivity is stronger for light overdrafters (whose average is 5.911) than for heavy overdrafters (whose average is 3.387).

And keeping the average overdraft transaction amount as fixed, a 1% increase in the overdraft fee can increase the closing probability by 0.12%.

7 Counterfactuals

7.1 Pricing

The structurally estimated model allows us to examine the effect of changing the pricing structure on consumers' spending pattern and more importantly, their overdraft behavior. We test three alternative pricing schemes: a reduced per-item flat fee, a percentage fee, and a quantity premium.

Table 13: Overdraft Fee under Alternative Pricing

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%

Notice here that the underlying assumption for all these simulations is fungibility, i.e., consumers' reaction only depends on the fee amount rather than the fee structure. If two different fee structures result in the same fee amount, then the consumer should respond in the same fashion.

In the first scenario, we keep the per-item flat fee scheme but reduce it to \$29.27 per item. Because of law of demand, there's a negative relationship between the per-item overdraft fee and overdraft frequency. So we further pursue an optimization task where we try to solve the optimal per-item fee. As we aggregate data to the daily level, we calculate the average transaction amount for each item, which is \$44, and use it to derive the total overdraft fee. For example, if a consumer overspent \$170, then the consumer had to pay four overdraft item fees. The optimization is a nested algorithm where in the outer loop we search for the per-item overdraft fee, and in the inner loop we solve the consumer's best response, including optimal spending, balance checking and account closing given the fee size. We found that the optimal per-item overdraft fee is \$29.27 under which the bank's revenue will increase by 3.26%. This suggests that the current overdraft fee is too high because the bank fails to take into account consumer's negative reaction to the overdraft fee, which results in huge loss in the consumers' lifetime value (I calculate the lifetime value of a consumer in a conservative way by multiplying the accounts spendings by the interchange rate).

In the second scenario, the per-item flat fee is changed to a percentage fee of 15.8% (optimized in a similar way as described in the first scenario). This is lower than the 17% calculated from the ratio of the total fee paid over the total transaction amount that caused the fees in the data. Again this suggests that the bank might be charging a too high fee currently. Intuitively, the percentage structure should encourage consumers to overdraw on transactions of a small amount but deter them from overdrawing on transactions of a large amount. As there are more transactions of a small amount than transactions of a large amount, the total fees generated soars by 7.12%. Therefore, the percentage overdraft fee invites more consumers to use the overdraft service. It is this market expansion effect that increases the bank's overdraft revenue.

In the last scenario, a quantity premium structure is employed, where when a consumer overdraws no more than 10 times, she pays a 8.5% percentage fee and if she overdraws more than 10 times, she pays a flat fee at $31. This quantity premium can increase the bank's revenue by 8.81%, because the quantity premium uses the second degree price discrimination to segment two types of overdrafters. The bank will earn more overdraft fee from the heavy overdrafters who are willing to pay for the flat fee while retaining the lifetime value for the light overdrafters who prefer the percentage fee (due to the high dissatisfaction sensitivity).

7.2 Alerts Benefit Consumers And the Bank

Although the changed pricing strategies can help the bank improve revenue, the bank is still exploiting consumer inattention and may exacerbate consumer attrition. In this counterfactual, we propose a new product design strategy (specific design to be introduced in section 7.3) to help consumers prevent overdrafts: sending automatic alerts to inform consumers about their balances. As alerts eliminate consumers' balance perception error, the total amount of overdraft fee paid by consumers decreases by 49.53% (Table 14. This is in comparison to the overdraft revenue under the optimal Quantity Premium pricing strategy in Table 13).

Table 14: Effect of Alerts on Bank's Revenue

	Amount	Percentage Change
Overdraft revenue	$10,243,529	-49.53%
Interchange revenue from increased spendings	$1,997,488	+9.84%
Lifetime value from retained consumers	$8,430,424	+41.53%
Total	$20,671,441	+1.84%

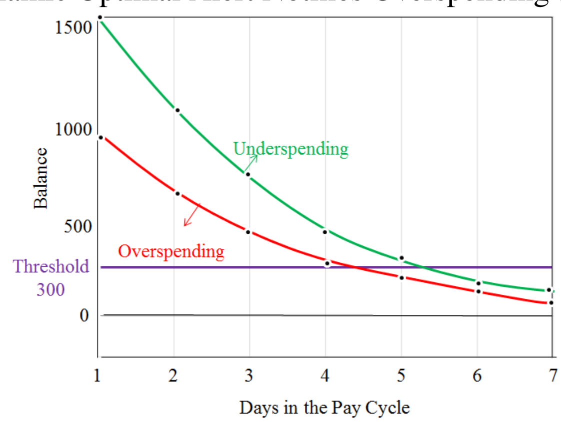
Although alerts benefit consumers by helping them avoid the high overdraft fees, the bank might not have incentives to send out alerts as its objective is to earn more revenue. However, we find that alerts can benefit the bank too for two reasons. First of all, as shown in section 4.8.3, due to inattention consumers are constraining spendings to prevent overdrafts. With alerts, consumers' precautionary motive is relieved so that they will increase spendings. As a result, the bank can gain more interchange fees. We calculate this gain of more interchange fee from the increased amount of spending by multiplying the increased spending with an average interchange fee rate of 0.8%[22]. We find that sending alerts to consumers can offset 9.84% of the loss in overdraft fees because of the gain in the interchange fees. Moreover, without being dissatisfied by the overdraft fee, consumers are less likely to close their accounts. We find that alerts reduce the number of closed accounts from 16.37% to 8.25% which increases the bank's revenue by getting the lifetime value from these retained consumers. As shown in Table 14, the increased lifetime value from retained consumers and the increase in interchange fee from increased spendings not only offset the loss in overdraft revenue but increase it by 1.84%.

[22] http://www.federalreserve.gov/paymentsystems/regii-average-interchange-fee.htm

7.3 Optimal Alert Strategy

Finally, we explain how we design the optimal alert that can help the bank increase its revenue in section 7.2. We show the effect of the proposed alert with an example in Figure 10. Consider a consumer who receives a weekly wage of \$2000. This consumer's discount factor is 0.8.[23] She sets a threshold alert at \$300 originally thus will only receive the alert when the account balance is below \$300. But our proposed alert will be triggered both when the consumer is overspending and underspending. As shown in the figure, as long as the consumer's spending falls out of the range between the overspending and underspending lines, an alert will be received. So when the consumer's balance is below \$700 on day 2, she will receive an alert although the threshold is not reached yet. The optimal alert is earlier than the threshold alert to give the consumer more time to adjust her spending rather than to wait until the last moment when she can hardly make any immediate change. On the other hand, if the consumer's balance is below \$300 on day 5, the threshold alert will be triggered while the consumer is still in a safe zone. Receiving the threshold alert doesn't help consumers because her perception error accumulates too fast to make day 6 and 7 danger days prone to overdrafts again. Therefore, the dynamic alert can correct the defects of the threshold alerts of being either too late or too early.

Figure 10: Dynamic Optimal Alert Notifies Overspending and Underspending



Another imbedded feature of the dynamic alert is that it accounts for consumers' disutility to receive too many alerts. In reality, consumers dislike frequent alerts that spam their mailboxes. We incorporate this alert-averse effect into an optimization task where we choose the optimal timing to send the alerts given the estimated structural parameters. The objective function is as follows

$$
\begin{aligned}
max_{\{A_{it}\}} \quad & \sum_{i=1}^{N}\sum_{t=1}^{\infty}\beta^{t-1}\left[U_{it}\left(C_{it}^{*},W_{it}^{*};\widehat{S_{it}}\right)-\kappa_i\right] \\
\widehat{S_{it}} \quad &= \quad A_{it}S_{it}+(1-A_{it})\widetilde{S_{it}}
\end{aligned}
$$

where A_{it} is a binary choice of whether to send an alert to the consumer i at time t. The second equation means that if the alert is sent, the consumer knows the exact balance and coming

[23] For the ease of exposition, we choose a relatively small discount factor.

bills, denoted as the true state variable S_{it}; if not, the consumer only knows the distribution of the perceived balance and coming bills, denoted as $\widetilde{S_{it}}$. The consumers' disutility of receiving the alert is summarized by a time invariate the parameter κ_i. We solve the optimization problem in a nested algorithm where in the outer loop we test for all combinations of alert opportunities, and in the inner loop we solve the consumer's best response, including optimal spending and account closing given the alert profile. (We assume that consumers don't have to make the balance checking decision because of the automatic alerts.)

We first test the optimal alert strategy assuming that all consumers have the same structural parameters (we use the posterior mean of the hyper-distribution parameters). We set this disutility as the inverse of the estimated monitoring cost (μ_ξ) because the consumer who incurs a high monitoring cost might not know how to use online banking or call centers so automatic message alerts are favored. As Table 15 reports, this alert service increases total consumer utility by 1.11% when the threshold rule of $300 is applied and 2.85% when the dynamic rule is applied.

We further allow all structural parameters to be heterogeneous across consumers and solve the optimal alert timing specific to each individual. We find that targeted alerts can increase consumer utilities six times more than the uniform threshold alert (6.65%).

Table 15: Utility Impact of Different Types of Alerts

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

8 Contributions and Limitations

The $32 billion dollar annual overdraft fee has caused consumer attrition and may induce potentially tighter regulation. However there is little quantitative research on consumers' financial decision making processes that explains their overdraft behaviors. The lack of well-calibrated models prevent financial institutions from designing pricing strategies and improving financial products. With the aid of Big Data associated with consumers' spending patterns and financial management activities, banks can use adverse targeting (Kamenica, Mullainathan, and Thaler 2011) to help consumers know themselves better and make better financial decisions.

In this paper we build a dynamic structural model of consumer daily spending that incorporates inattention to rationalize consumers' overdraft behavior. We quantify the discount factor, monitoring cost and dissatisfaction sensitivity for each consumer and use these to design new strategies. First we compare the current pricing scheme with several alternative pricing strategies. We find that a percentage fee structure can increase the bank's revenue through market expansion and the quantity premium structure can increase the bank's revenue because of second degree price discrimination. More importantly, we propose an alert strategy to make the incentive of the bank and the incentive of the consumers better aligned. The optimal alert can be sent to the right consumer at the right time to prevent overdrafts. This customized dynamic alert product can be six times more effective than a uniform threshold alert. Not only does this alert benefits consumers, it can also benefit the bank through increased interchange fees and lower consumer attrition.

We calibrated our model at an individual level on a sample of more than 500,000 accounts. This Big Data provide great value for our analysis. First of all, an overdraft is still a relatively rare event compared to numerous other transactions. Without a large amount of data, we cannot detect these rare but detrimental events, let alone their diverse causes. Second, as summarized by Einav and Levin (2014), Big Data contain rich micro-level variation that can used to identify novel behavior and develop predictive models that are harder with smaller samples, fewer variables, and more aggregation. We leverage the variation in consumer daily spending and balance checking behaviors to evaluate the effect of heterogeneous policy instruments. These evaluations can be useful for bank managers to design new products and policy makers to create new regulation rules at a much more refined fashion than before.

In order to estimate a complicated structural model with Big Data, we adopt parallel computing techniques in combination with the Bayesian estimation algorithm developed by Imai, Jain and Ching (2009). This new method significantly reduces the computation burden and could be used for other researchers and marketers who would like to use structural models to solve real-world large-scale problems.

There are several limitations of the current study that call for future work. First, we don't observe consumers' existing alert settings. Some consumers may have already received alerts to help them make financial decisions. In our policy simulations, we made bold assumptions about consumers' disutility for reading alerts. These assumptions could be tested if we had the alerts data. The current alerts are set by consumers who might fail to consider their spending dynamics. Future field experiments are needed to test the effect of our proposed alert strategy. Second, we don't have the data about consumers' decision on whether to opt-in for overdraft protection by ATM/POS transactions. We only know that if ATM/POS transactions caused an overdraft, then the consumer must have opted-in. If no such transactions happened, we do not know the consumer's opt-in status. Had we known this information, we could have provided an informative prior in our the Bayesian model. The logic is that a consumer who has opted in probably has stronger needs for short term liquidity due to fluctuations in the size and arrival time of income and expenditures. Finally, we only model consumers' non-preauthorized spending in the checking account. In reality, consumers usually have multiple accounts, like savings, credit cards and loans, with multiple financial institutions. A model to capture consumers' decisions across all accounts for both short-term and long-term finances will provide a more complete picture of consumers' financial management capabilities and resources so that the bank can design more customized products.

References

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3), 583-618.

Arcidiacono, P., & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6), 1823-1867.

Arrow, K. J. 1963. Liquidity preference. Lecture VI in *Lecture Notes for Economics 285, The Economics of Uncertainty*, pp. 33-53, undated, Stanford University.

Bellman, Richard 1957. *Dynamic Programming*, Princeton, NJ: Princeton University Press.

Bolton, R. N. (1998). A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction. *Marketing science*, 17(1), 45-65.
Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4), 1145-1177.
Card, D., DellaVigna, S., & Malmendier, U. (2011). The role of theory in field experiments. *The Journal of Economic Perspectives*, 25(3), 39-62.
Einav, L., & Levin, J. D. (2014). The data revolution and economic analysis. *Innovation Policy and the Economy*, forthcoming.
Evans, D., Litan, R., & Schmalensee, R. (2011). Economic Analysis of the Effects of the Federal Reserve Board's Proposed Debit Card Interchange Fee Regulations on Consumers and Small Businesses. *Available at SSRN 1769887*.
Fang, H., & Silverman, D. (2009). Time-inconsistency and welfare program participation: Evidence from the nlsy. *International Economic Review*, 50(4), 1043-1077.
Fishman, M. J., & Hagerty, K. M. (2003). Mandatory versus voluntary disclosure in markets with informed and uninformed customers. *Journal of Law, Economics, and Organization*, 19(1), 45-63.
Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, 121(2), 505-540.
Gopalakrishnan, A., Iyengar, R., & Meyer, R. J. (2014). Consumer dynamic usage allocation and learning under multi-part tariffs: theory and empirical evidence. Working paper, Wharton School, University of Pennsylvania.
Grubb, M. D. (2014). Consumer inattention and bill-shock regulation. *The Review of Economic Studies, forthcoming.*
Grubb, M. D., & Osborne, M. (2014) Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock. *American Economic Review, forthcoming*.
Haselhuhn, M. P., Pope, D. G., Schweitzer, M. E., & Fishman, P. (2012). The impact of personal experience on behavior: Evidence from video-rental fines. *Management Science*, 58(1), 52-61.
Hirshleifer, D., & Hong Teoh, S. (2003). Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management*, 9(1), 25-66.
Imai, S., Jain, N., & Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, 77(6), 1865-1899.
Jiang, L. (2012). The welfare effects of "bill shock" regulation in mobile telecommunication markets. Working Paper, University of British Columbia.
Kahneman, D., J. Knetsch, and R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy,* 98(6), 1325–1348.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.
Kamenica, E., Mullainathan, S., & Thaler, R. (2011). Helping consumers know themselves. *The American Economic Review*, 101(3), 417-422.
Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). Getting to the top of mind: How reminders increase saving (No. w16205). National Bureau of Economic Research.
Kass, R., and A. Raftery. (1995). Bayes Factors. *Journal of the American Statistical Association* 90 (430): 773-95.
Keaveney, S. M. (1995). Customer switching behavior in service industries: an exploratory study. *The Journal of Marketing*, 71-82.

Magnac, T., & Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2), 801-816.
Matzler, K., Würtele, A., & Renzl, B. (2006). Dimensions of price satisfaction: a study in the retail banking industry. *International Journal of Bank Marketing*, 24(4), 216-231.
Mehta, N., Rajiv, S., & Srinivasan, K. (2003). Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing science*, 22(1), 58-84.
Neiswanger, W., Wang, C., & Xing, E. (2014). Asymptotically exact, embarrassingly parallel MCMC. *Proceedings of the 30th International Conference on Conference on Uncertainty in Artificial Intelligence.*
Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563-602.
Pope, D. G. (2009). Reacting to rankings: Evidence from "America's Best Hospitals". *Journal of Health Economics*, 28(6), 1154-1165.
Rossi, Peter E., Allenby, G. M., & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. J. Wiley & Sons.
Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica: *Journal of the Econometric Society*, 999-1033.
Stango, V., & Zinman, J. (2014). Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *Review of Financial Studies, forthcoming*.
Yao, S., Mela, C. F., Chiang, J., & Chen, Y. (2012). Determining consumers' discount rates with field studies. *Journal of Marketing Research*, 49(6), 822-841.
Zhou, Y. (2012). Failure to Launch in Two-Sided Markets: A Study of the US Video Game Market. Working paper, SUNY at Stony Brook.

Appendix

A.1 Overdraft Fees at Top US Banks

Table 16: Overdraft Fees at Top U.S. Banks

Bank	Overdraft Fee	Max Fees per Day	Overdraft Protection Transfer	Continuous Overdraft 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		

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

A.2 Estimation Algorithm: Modified IJC

Detailed steps

1. Suppose that we are at iteration r. We start with $H^r = \left\{ \left\{ \widetilde{S}_i^k, \widetilde{V}^k \left(\hat{S}_i^k, \widetilde{S}_i^k; \vartheta_i^k \right) \right\}_{i=1}^{I} \right\}_{k=r-N}^{r-1}$ where I is the number of consumers; N is the number of past iterations used for the expected future value approximation; $\vartheta_i = \{\lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i\}$.

2. Draw μ_ϑ^r (population mean of ϑ_i) from the posterior density (normal) conditional on σ_ϑ^{r-1} and $\left\{\vartheta_i^{r-1}\right\}_{i=1}^{I}$. $\mu_\vartheta^r \sim N\left(\frac{\sum_{i=1}^{I} \vartheta_i^{r-1}}{I}, \sigma_\vartheta^{r-1}\right)$

3. Draw σ_ϑ^r (population variance of ϑ_i) from the posterior density (inverted gamma) conditional on μ_ϑ^r and $\left\{\vartheta_i^{r-1}\right\}_{i=1}^{I}$.$\sigma_\vartheta^r \sim IG\left(\frac{I}{2}, \frac{\sum_{i=1}^{I}\left(\vartheta_i^{r-1}-\mu_\vartheta^r\right)^2}{2}\right)$

4. For each $i = 1, ..., I$, draw ϑ_i^r from its posterior distribution conditional on $(C_i^d, Q_i^d, W_i^d, \mu_\vartheta^r, \sigma_\vartheta^r)$, which is

$$f_i\left(\vartheta_i | C_i^d, Q_i^d, W_i^d, \mu_\vartheta^r, \sigma_\vartheta^r\right) \propto \pi\left(\vartheta_i | \mu_\vartheta^r, \sigma_\vartheta^r\right) \rho_i\left(C_i^d | \vartheta_i\right) \rho_i\left(Q_i^d | \vartheta_i\right) \rho_i\left(W_i^d | \vartheta_i\right)$$

Since there is no easy way to draw from this posterior, we use the M-H algorithm.

(a) Draw ϑ_i^{*r} from the proposal distribution $q\left(\vartheta_i^{r-1}, \vartheta_i^{*r}\right)$ (e.g., $\vartheta_i^{*r} \sim N\left(\vartheta_i^{r-1}, \sigma^2\right)$ where ϑ_i^{*r} is a candidate value of ϑ_i^r .

(b) Compute the pseudo-likelihood for consumer i at ϑ_i^{*r}, i.e., $\rho_i^r\left(C_i^d | \vartheta_i^{*r}\right)$, $\rho_i^r\left(Q_i^d | \vartheta_i^{*r}\right)$ and $\rho_i\left(W_i^d | \vartheta_i^{*r}\right)$. Since there is no closed form solution to the optimal strategy profile, a likelihood function based on observed C_{it} becomes infeasible. Instead, we implement a numerical approximation method to establish a simulated likelihood function for estimation. For each C_{it} observed in the data and its corresponding state point $\hat{S}_{it}$, we use the following steps to simulate its density:

i. First assume the unobserved state variables are $\widetilde{S}_{it} = \{\varepsilon_{it}, \eta_{it}, \chi_{it}, \varpi_{it}\}$. Draw nr=1000 random shocks $\widetilde{S}_{it} = \{\varepsilon_{it}, \eta_{it}, \chi_{it}, \varpi_{it}\}$ from

$$\eta_{it} \sim N(0, \omega_i^2),\ \ \varepsilon_{it} \sim N(0,1), \chi_{it}\text{~EVI}^{24}, \varpi_{it}\text{~EVI};$$

i. For each balance checking decision $Q = \{1,0\}$and account closing decision $W = \{1,0\}$, each random draw of $\widetilde{S}_{it}$ and the observed $\hat{S}_{it}$, calculate the optimal consumption by solving the following equations

$$\begin{aligned} C_{it}^*\left(\hat{S}_{it}, \widetilde{S}_{it}|Q,W\right) &= arg\max_{C_{it}} \widetilde{v}^r\left(Q,W;\hat{S}_{it},\widetilde{S}_{it},\vartheta_i^{*r}\right) \\ &= arg\max_{C_{it}} U\left(C_{it},Q,W;\hat{S}_{it},\widetilde{S}_{it},\vartheta_i^{*r}\right) + \beta\hat{E}^r_{S_{it+1}}\left\{V\left(\hat{S}_{it+1},\widetilde{S}_{it+1};\vartheta_i^{*r}\right)|\hat{S}_{it},\widetilde{S}_{it}\right\} \end{aligned}$$

iii. Using the calculated nr = 1000 optimal $C_{it}^*\left(\hat{S}_{it}, \widetilde{S}_{it}\right)$, simulate $\rho_i^r\left(C_{it}^d|\vartheta_i^{*r}\right)$, the density of the observed C_{it}^d, using a Gaussian kernel density estimator. (This simulation borrows an idea from Yao, Mela, Chiang and Chen (2012)) . Moreover,

$$\rho_i\left(Q_{it}^d|\vartheta_i^{*r}\right) = \frac{1}{nr}\sum_{\eta,\varepsilon,\varpi}\frac{exp\left\{\overline{v}_{\chi_{it}}\left(C_{it}^*, Q=Q_{it}, W_{it}^*;\widetilde{B}_{it},R_{it},\Psi_{it},Y_{it},DL_{it},D_{it},OD_{it},\Gamma_{it},\Xi_{it},\varepsilon_{it},\varpi_{it},\vartheta_i^{*r}\right)\right\}}{\sum_{Q\in\{1,0\}}exp\left\{\overline{v}_{\chi_{it}}\left(C_{it}^*, Q, W_{it}^*;\widetilde{B}_{it},R_{t},\Psi_{it},Y_{it},DL_{it},D_{it},OD_{it},\Gamma_{it},\Xi_{it},\varepsilon_{it},\varpi_{it},\vartheta_i^{*r}\right)\right\}}$$

and

$$\rho_i\left(W_{it}^d|\vartheta_i^{*r}\right) = \frac{1}{nr}\sum_{\eta,\varepsilon,\chi}\frac{exp\left\{\overline{v}_{\varpi}\left(C_{it}^*, Q_{it}^*, W=W_{it};\widetilde{B}_{it},R_{it},\Psi_{it},Y_{it},DL_{it},D_{it},OD_{it},\Gamma_{it},\Xi_{it},\varepsilon_{it},\chi_{it},\vartheta_i^{*r}\right)\right\}}{\sum_{W\in\{1,0\}}exp\left\{\overline{v}_{\varpi}\left(C_{it}^*, Q_{it}^*, W;\widetilde{B}_{it},R_{t},\Psi_{it},Y_{it},DL_{it},D_{it},OD_{it},\Gamma_{it},\Xi_{it},\varepsilon_{it},\chi_{it},\vartheta_i^{*r}\right)\right\}}$$

$$\begin{aligned} \rho_i\left(O_i^d|\vartheta_i^{*r}\right) &= \rho_i\left(C_i^d|\vartheta_i^{*r}\right)\rho_i\left(Q_i^d|\vartheta_i^{*r}\right)\rho_i\left(W_i^d|\vartheta_i^{*r}\right) \\ &= \prod_{t=1}^{T}\rho_i\left(C_{it}^d|\vartheta_i^{*r}\right)\rho_i\left(Q_{it}^d|\vartheta_i^{*r}\right)\rho_i\left(W_{it}^d|\vartheta_i^{*r}\right) \end{aligned}$$

.

To obtain $\widetilde{v}^r\left(\hat{S}_{it}, \widetilde{S}_{it}, \vartheta_i^{*r}\right)$, we need $\hat{E}^r_{S'}\left\{V\left(\hat{S}_i', \widetilde{S}_i'; \vartheta_i^{*r}\right)|\hat{S}_i, \widetilde{S}_i\right\}$, which is obtained by a weighted average of $\left\{\widetilde{V}^k\left(\hat{S}_i^k, \widetilde{S}_i^k; \vartheta_i^{*k}\right)\right\}_{k=r-N}^{r-1}$, treating ϑ_i as one of the parameters when computing the weights. In the case of independent kernels, for all $\hat{S}_i = \{B_i, \psi_i, Y_i, DL_i, OD_i, \Gamma_i, \Xi_i\}$, because B_i,Ξ_i are continuous and evolves deterministically, ψ_i and OD_i are continuous and evolve stochastically, and Y_i, DL_i, Γ_i are discrete so

$$\begin{aligned} &\hat{E}^r_{S'}\left\{V\left(B_i',\psi_i',Y_i',DL_i',OD_i',\Gamma_i',\Xi_i',\widetilde{S}_i';\vartheta_i^{*r}\right)|B_i,\psi_i,Y_i,DL_i,OD_i,\Gamma_i,\Xi_i,\widetilde{S}_i\right\} \\ = &\sum_{k=r-N}^{r-1}\widetilde{V}^k\left(B_i^k,\psi_i^k,Y_i',DL_i',OD_i^k,\Gamma_i',\Xi_i',\widetilde{S}_i^k;\vartheta_i^{*k}\right)\frac{K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*k}\right)K_{hs}\left(B_i'-B_i^k\right)f\left(\psi_i^k|\phi_i,G_i\right)f\left(OD_i^k|X_i,p_i\right)K_{hs}\left(\Xi_i'-\Xi_i^k\right)}{\sum_{l=r-N}^{r-1}K_{h\vartheta}\left(\vartheta_i^{*r}-\vartheta_i^{*l}\right)K_{hs}\left(B_i'-B_i^l\right)f\left(\psi_i^l|\phi_i,G_i\right)f\left(OD_i^l|X_i,p_i\right)K_{hs}\left(\Xi_i'-\Xi_i^l\right)} \end{aligned}$$

We repeat the same step and obtain the pseudo-likelihood $(\rho_i^r\left(O_i^d|\vartheta_i^{r-1}\right))$ conditional on $\left(\vartheta_i^{r-1}\right)$. Then, we determine whether or not to accept ϑ_i^{*r}. The acceptance probability, Λ, is given by

$$\Lambda = min\left(\frac{\pi\left(\vartheta_i^{*r}|\mu_\vartheta^r,\sigma_\vartheta^r\right)\rho_i^r\left(O_i^d|\vartheta_i^{*r}\right)q\left(\vartheta_i^{*r},\vartheta_i^{r-1}\right)}{\pi\left(\vartheta_i^{r-1}|\mu_\vartheta^r,\sigma_\vartheta^r\right)\rho_i^r\left(O_i^d|\vartheta_i^{r-1}\right)q\left(\vartheta_i^{r-1},\vartheta_i^{*r}\right)},1\right)$$

[24] Type I Extreme Value Distribution

where $\pi(\cdot)$ denotes the prior distribution.

(c) Repeat (a) & (b) for all i.

5. Computation of the pseudo-value function, $\left\{\widetilde{V}^r\left(\hat{S}_i^r, \widetilde{S}_i^r; \vartheta_i^{*r}\right)\right\}_{i=1}^{I}$

(a) Make one draw of the unobserved state variables $\widetilde{S}_i^r$ from $\eta_i \sim N(0, \omega_i^2)$, $\varepsilon_i \sim N\left(0, \varsigma_i^2\right)$, χ_i~EVI[25], and ϖ_i~EVI;

(b) Compute the pseudo expected future value at ϑ_i^{*r}.

$$\hat{E}_{S_i'}^r\left\{V\left(\hat{S}_i', \widetilde{S}_i'; \vartheta_i^{*r}\right)|\hat{S}_i^r, \widetilde{S}_i^r\right\} = \sum_{k=r-N}^{r-1} \widetilde{V}^k\left(B_i^k, \psi_i^k, Y_i', DL_i', OD_i^k, \Gamma_i', \Xi_i', \widetilde{S}_i^k; \vartheta_i^{*k}\right) \frac{K_{h\vartheta}\left(\vartheta_i^{*r} - \vartheta_i^{*k}\right) K_{hs}\left(B_i^r - B_i^k\right) f\left(\psi_i^k|\phi_i, G_i\right) f\left(OD_i^k|X_i, p_i\right) K_{hs}\left(\Xi_i^r - \Xi_i^k\right)}{\sum_{l=r-N}^{r-1} K_{h\vartheta}\left(\vartheta_i^{*r} - \vartheta_i^{*l}\right) K_{hs}\left(B_i^r - B_i^l\right) f\left(\psi_i^l|\phi_i, G_i\right) f\left(OD_i^l|X_i, p_i\right) K_{hs}\left(\Xi_i^r - \Xi_i^l\right)}$$

(c) Compute $\widetilde{V}^r\left(\hat{S}_i^r, \widetilde{S}_i^r; \vartheta_i^{*r}\right)$, using the pseudo expected future values computed in (b) and the optimal choices C_i^*, Q_i^*, W_i^*.

$$\widetilde{V}^r\left(\hat{S}_i^r, \widetilde{S}_i^r; \vartheta_i^{*r}\right) = U\left(C_i^*, Q_i^*, W_i^*; \hat{S}_i^r, \widetilde{S}_i\right) + \beta \hat{E}_{S_i'}^r\left\{V\left(\hat{S}_i', \widetilde{S}_i'; \vartheta_i^r\right)|\hat{S}_i^r, \widetilde{S}_i^r\right\}$$

where C_i^*, Q_i^*, W_i^* satisfy

$$\widetilde{V}^r\left(\hat{S}_i^r, \widetilde{S}_i^r; \vartheta_i^{*r}\right) = \max_{C_i, Q_i, W_i} U\left(C_i^*, Q_i^*, W_i^*; \hat{S}_i^r, \widetilde{S}_i^r\right) + \beta \hat{E}_{S'}^r\left\{V\left(\hat{S}_i', \widetilde{S}_i'; \vartheta_i^r\right)|\hat{S}_i^r, \widetilde{S}_i^r\right\}$$

(d) Repeat (a-c) for all i.

6. Go to iteration r + 1.

A.3 Parallel MCMC Sampling Algorithm

Table 17: Algorithm: Asymptotically Exact Sampling via Nonparametric Density Product Estimation

Input: Subposterior samples, $\left\{\vartheta_{t_1}\right\}_{t_1=1}^{T} \sim p_1(\vartheta), \ldots, \left\{\vartheta_{t_M}\right\}_{t_M=1}^{T} \sim p_M(\vartheta)$

Output: Posterior samples (asymptotically, as $T \to \infty$), $\{\vartheta_i\}_{i=1}^{T} \sim p_1 \ldots p_M(\vartheta) \propto p\left(\vartheta|x^N\right)$

1: Set $h = 1$.	10: Draw $u \sim Unif([0,1])$.
2: Draw $t\cdot = \{t_1, \ldots, t_M\} \overset{iid}{\sim} Unif(\{1, .., T\})$	11: if $u < \frac{w_{t\cdot}}{w_{c\cdot}}$ then
3: Set $c\cdot = t\cdot$.	12: Draw $\vartheta_t \sim N\left(\bar{\vartheta}_{t\cdot}, \frac{h^2}{M} I_d\right)$.
4: Draw $\vartheta_1 \sim N\left(\bar{\vartheta}_{t\cdot}, \frac{h^2}{M} I_d\right)$.	13: Set $c\cdot = t\cdot$.
5: for i = 2 to T do	14: else
6: for m = 1 to M do	15: Draw $\vartheta_t \sim N\left(\bar{\vartheta}_{c\cdot}, \frac{h^2}{M} I_d\right)$.
7: Set $t\cdot = c\cdot$.	16: end if
8: Draw $t_m \sim Unif(\{1, .., T\})$	17: end for
9: Set $h = i^{-\frac{1}{(4+d)}}$.	18: end for

[25]Type I Extreme Value Distribution