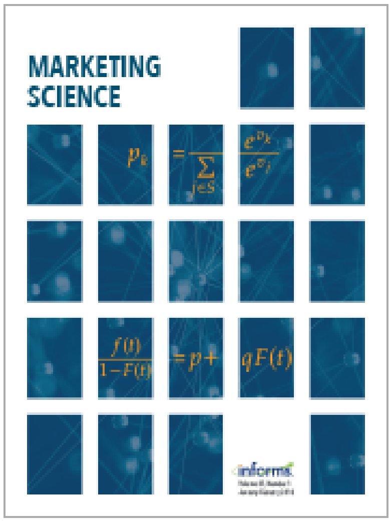


Marketing Science

Analyzing Bank Overdraft Fees with Big Data

Xiao Liu, Alan Montgomery, Kannan Srinivasan

To cite this article:
Xiao Liu, Alan Montgomery, Kannan Srinivasan (2018) Analyzing Bank Overdraft Fees with Big Data. Marketing Science 37(6):855-882. https://doi.org/10.1287/mksc.2018.1106

MARKETING SCIENCE
Vol. 37, No. 6, November–December 2018, pp. 855–882
ISSN 0732-2399 (print), ISSN 1526-548X (online)
http://pubsonline.informs.org/journal/mksc/

Analyzing Bank Overdraft Fees with Big Data

Xiao Liu,[a] **Alan Montgomery,**[b] **Kannan Srinivasan**[b]

[a] Stern School of Business, New York University, New York, New York 10012; [b] Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Contact: xliu@stern.nyu.edu, http://orcid.org/0000-0002-7093-8534 (XL); alanmontgomery@cmu.edu (AM); kannans@cmu.edu, http://orcid.org/0000-0002-5389-8569 (KS)

Received: August 18, 2016
Revised: November 8, 2017; April 19, 2018
Accepted: April 19, 2018
Published Online in Articles in Advance: October 23, 2018

https://doi.org/10.1287/mksc.2018.1106

Abstract. In 2012, consumers paid $32 billion in overdraft fees, representing the single largest source of revenue for banks from demand deposit accounts during this period. Owing to consumer attrition caused by overdraft fees and potential government regulations to reform these fees, financial institutions have become motivated to investigate their overdraft fee structures. Banks need to balance the revenue generated from overdraft fees with consumer dissatisfaction and potential churn caused by these fees. However, no empirical research has been conducted to explain consumer responses to overdraft fees or to evaluate alternative pricing strategies associated with these fees. In this research, we propose a dynamic structural model with consumer monitoring costs and dissatisfaction associated with overdraft fees. We apply the model to an enterprise-level data set of more than 500,000 accounts with a history of 450 days, providing a total of 200 million transactions. We find that consumers heavily discount the future and potentially overdraw because of impulsive spending. However, we also find that high monitoring costs hinder consumers' effort to track their balance accurately; consequently, consumers may overdraw because of rational inattention. The large data set is necessary because of the infrequent nature of overdrafts; however, it also engenders computational challenges, which we address by using parallel computing techniques. Our policy simulations show that alternative pricing strategies may increase bank revenue and improve consumer welfare. Fixed bill schedules and overdraft waiver programs may also enhance social welfare. This paper explains consumer responses to overdraft fees and evaluates alternative pricing strategies associated with these fees.

History: K. Sudhir served as the editor-in-chief and Tat Chan served as associate editor for this article.
Funding: X. Liu acknowledges support from the Dipankar and Sharmila Chakravarti Fellowship.
Supplemental Material: The online appendices are available at https://doi.org/10.1287/mksc.2018.1106.

Keywords: banking • overdraft fees • dynamic programming • big data

1. Introduction

An overdraft occurs when a consumer spends or withdraws an amount of funds from his or her checking account that exceeds the account's available funds. U.S. banks allow consumers to overdraw their account (subject to some restrictions at the bank's discretion) but charge an overdraft fee (OD). Overdraft fees have been a major source of bank revenue since the early 1980s (Topper 2012), when banks started to offer free checking accounts to attract consumers. According to Moebs Services, the total amount of overdraft fees in the United States reached $32 billion in 2012. This is equivalent to an average of $178 for each checking account annually.[1] According to the Center for Responsible Lending, U.S. households spent more on overdraft fees than on fresh vegetables, postage, or books in 2010 (Burns 2010).

Overdraft fees have provoked a storm of consumer dissatisfaction and can induce many consumers who experience these fees to close their accounts.[2] The U.S. government has taken actions to regulate overdraft fees through the Consumer Financial Protection Agency,[3] and it may take more drastic steps in the future.[4] Therefore, there are pressures throughout the industry for banks to reexamine their overdraft fee practices to prevent churn and better prepare for regulations.

A potential solution for both consumers and banks is to leverage financial transaction data to manage overdrafting and offer new services using these financial transaction data. Financial institutions store massive amounts of information about consumers, commonly referred to as *big data*, as a by-product of their transactions. In this research, we show how this information can be harnessed with structural economic theories to predict consumers' overdrafting behavior. The large-scale of our financial transaction panel data allows us to detect rare events of overdrafts, identify rich consumer heterogeneity, and minimize sampling bias to avoid potential financial losses. Consequently, we propose strategies that can increase both consumer welfare and bank revenue. Our goal is to show that the knowledge about consumers contained within their financial transaction data can form the basis for improving customer welfare and increasing profitability for the bank by tapping into this underutilized resource for marketing purposes.

The ultimate objective of this paper is to investigate alternative pricing strategies of overdraft fees. Specifically, we tackle the following questions: Is the current overdraft fee structure optimal? How will bank revenue and consumer welfare change under alternative pricing strategies? The answers to these questions rely on our understanding of the decision process underlying consumers' overdrafting behavior. So, we first address the following research questions: Why do consumers overdraw? How do consumers react to overdraft fees?

To study these problems, we obtained anonymized data from a large U.S. bank. The data include over 500,000 accounts with a history of up to 450 days, amounting to 200 million relevant observations. This enterprise-level data set is much larger than those reported in other research studies. Substantively, we find that some consumers are inattentive in monitoring their balances because of the associated high monitoring costs. In contrast, attentive consumers primarily overdraw because they heavily discount future utilities (they are subject to impulsive spending) or use overdrafts to satisfy short-term liquidity constraint. Consumers who are dissatisfied may then leave their bank after being charged high overdraft fees.

These findings suggest that to evaluate the welfare consequences of new policies, we must carefully model consumer demand by taking into account heavy discounting, inattention, and dissatisfaction behaviors. This drives us to make two methodological contributions. First, we construct a dynamic structural model that incorporates inattention and dissatisfaction into the lifetime consumption model. These modeling components are necessary to accurately predict consumers' reactions to new pricing policies and subsequently assess welfare implications. Structural models have the merit of producing policy-invariant parameters that allow us to conduct counterfactual analyses. However, the inherent computational burden prevents them from being widely adopted by the industry. This leads to our second contribution, whereby we show how to estimate a structural model applied to big data with the help of parallel computing techniques. Our proposed algorithm takes advantage of state-of-the-art parallel computing techniques and estimation methods to lessen the computational burden and reduce the curse of dimensionality to the point where near-real-time results are possible.

In our counterfactual analysis, we show that a percentage fee or a quantity premium fee strategy can achieve higher bank revenue than the current flat per-transaction fee strategy because of the demand expansion effect and second degree price discrimination effect. Consumers also benefit from the lowered overdraft fees by improving their capabilities to smooth out consumption over time and save monitoring costs. Therefore, all the new pricing strategies can improve social welfare. Moreover, a fixed bill payment schedule can also improve social welfare by saving consumers time in monitoring their accounts. Interestingly, an overdraft waiver for high-balance customers can be a win–win for both consumers and the bank.

Consumer choice of financial products is an important and enormous area that has been underresearched in marketing. Providing a strong impetus to new topics for research, this journal's publication policy is favorably disposed to even exploratory research efforts (Sudhir 2016). We believe that our research is on an important and underresearched area, and that our analysis is far beyond what may be deemed as exploratory.

The rest of this paper is organized as follows. In Section 2, we review related research. We report an exploratory data analysis in Section 3 to motivate our model setup. In Section 4, we describe our structural model. We give details about the identification and estimation procedures in Section 5. In Sections 6 and 7, we discuss our estimation results and counterfactual analysis. We conclude in Section 8 with a discussion of our findings and the limitations of our research.

2. Literature Review

An economic approach to explaining overdrafting would assume that consumers are rational and forward-looking with an objective to maximize their total discounted utility by making optimal choices (Modigliani and Brumberg 1954, Hall 1978). Consistent with the rational argument for overdrafting is that consumers heavily discount the future and are willing to pay future overdraft fees to allow consumption today. Although we are sympathetic to full information rational models of consumer behavior, we do not want to overlook potential behavioral explanations of overdrafting behavior. Specifically, we consider two novel arguments that offer behavioral explanations concerning overdrafting: inattention and dissatisfaction.

The inattention argument is present in a large body of literature in psychology and economics, which has found that consumers pay limited attention to relevant information. In their review paper, DellaVigna (2009) summarize findings indicating that consumers pay limited attention to (1) shipping costs, (2) tax (Chetty et al. 2009), and (3) rankings (Pope 2009). Gabaix and Laibson (2006) find that consumers do not pay enough attention to add-on pricing, Grubb (2014) shows that consumers are inattentive to their cell-phone minute balances, and Ching et al. (2009) find that consumers only occasionally attend to prices. Many papers in finance and accounting have documented that investors and financial analysts are inattentive to various types of financial information (e.g., Hirshleifer and Teoh 2003, Peng and Xiong 2006).

Limited attention can also explain overdrafting behaviors, as demonstrated by Stango and Zinman (2014). They state that there is survey evidence that people overdraw accounts because "they 'thought there was enough money in my account'" (p. 996). So they define inattention as incomplete consideration of account balances (realized balance and available balance net of upcoming bills) that would inform choices. Although Stango and Zinman (2014) use a data set similar to ours, their aim is to show that reminding participants about overdraft fees can reduce the likelihood of overdrafts. We adopt this definition of inattention, but we introduce inattention through a structural parameter, the monitoring cost (Reis 2006), which represents the time and effort required for a consumer to know the exact amount of money in his or her checking account.

A second behavioral argument related to overdrafting is that it may cause consumer dissatisfaction. The implied interest rate for an overdraft originated by a small transaction amount implies usurious rates that are much higher than the socially accepted interest rate (Matzler et al. 2006), leading to price dissatisfaction. This is because under current banking practices, consumers pay flat per-transaction fees regardless of the transaction amount. Overdrafting fees may cause consumer dissatisfaction, which is one of the main causes of customer switching behavior (Keaveney 1995, Bolton 1998). We conjecture that consumers are likely to close their account after they pay an overdraft fee and/or if the ratio of the overdraft fee to the overdraft transaction amount is high.

Before posing a formal economic model, we begin with the data and an exploratory data analysis to validate whether there is evidence for high discounting, inattention, and dissatisfaction.

3. Data

We obtain anonymized data from a large U.S. bank. Our data comprise bank transaction data for a sample of more than 500,000 accounts[5] with more than 200 million transactions over a 15-month period (June 2012 to Aug 2013). These data are a by-product of consumers' financial transactions. For each transaction, we know a unique but anonymized account identifier along with the date, channel, amount, and type of the transaction. Table 1 provides a simulated example of the raw information for a consumer. In this example, the consumer makes an ATM withdrawal and starts with a positive balance. On the next day, a check is paid by the bank even though the consumer has insufficient funds, which triggers an overdraft and the corresponding fee. A direct deposit from salary income is received, which brings the consumer's balance to a positive amount. Subsequently, the consumer does a balance check and makes a purchase at the supermarket on the next day. The description in this example is given for illustrative purposes and is not provided in our data set. Each transaction is classified into one of five categories: bills, fees assessed by the bank, income (from deposits and transfers), spending, and balance inquiries.

The bank in the data set provides a comprehensive set of services for consumers to avoid overdrafts, such as automatic transfers, but despite these offerings, a significant number of consumers still overdraw. (For a good review of general overdraft practices in the United States, refer to Stango and Zinman (2014). Online Appendix A1 tabulates the current fee settings of the top U.S. banks.) If a consumer overdraws his or her account with the standard overdraft service, then the bank might cover the transaction and charge a \$31[6] OD or decline the transaction and charge a \$31 nonsufficient fund fee (NSF). The bank can accept or decline the transaction at its discretion. The OD/NSF is applied at a per-item level: if a consumer performs several transactions when his or her account is already overdrawn, each transaction item will incur a fee of \$31. However, within a day, the total number of per-item fees is capped.[7] If the account remains overdrawn for several[8] consecutive calendar days, a continuous overdraft fee of \$6 is assessed. The bank also provides an overdraft protection service, where a checking account can be linked 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. In summary, the overdraft fee structure for the bank, as for most others, is quite complicated. In our empirical analysis, we do not distinguish among different types of overdraft fees, and we assume that consumers care only about the total amount of overdraft fees rather than the underlying pricing structure.

The bank also provides balance-checking services through its branches, ATMs, call centers, and

Table 1. Example of Simulated Transaction Data for an Individual

Date	Description	Channel	Type	+/−	Amount (\$)	Balance (\$)
11/14/2012	ATM withdrawal	ATM	Spending	–	80.00	63.15
11/15/2012	Check cashed for electric payment	ACH	Bill	–	130.41	–67.26
11/15/2012	Overdraft item fee		Fee	–	31.00	–98.26
11/16/2012	Salary from direct deposit	ACH	Income	+	287.42	189.16
11/17/2012	Check balance	ATM	Balance inquiry	O		189.16
11/17/2012	Debit card purchase at supermarket	Debit	Spending	–	97.84	91.32

online/mobile banking service. Consumers can inquire about their available balances and recent activities. There is also a notification service, providing so-called "alerts," that notifies consumers via emails or text messages when certain events take place, such as overdrafts, incidents of insufficient funds, transfers, and deposits. Unfortunately, our data set includes only balance-checking data, not alert data. We 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 nonrecurring point-of-sale (POS) debit card transactions before banks can charge them for paying such transactions.[9] Regulation E was intended to protect consumers from heavy overdraft fees. The change became effective for new accounts on July 1, 2010, and for existing accounts on August 15, 2010. Our data contain both opt-in and opt-out accounts. However, we do not know which accounts have opted in unless we observe an ATM/POS-initiated overdraft incident. We discuss this data limitation in Section 8.

3.1. Descriptive Statistics

In our data set, 15.8% of accounts had at least one overdraft incident. The proportion of accounts with overdrafts is lower than the 27% (across all banks and credit unions) reported by the CFPB in 2012.[10] Table 2 shows that consumers who paid overdraft fees overdrew nearly 10 times and paid $245 on average 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 with one or more overdraft transactions in 2011 totaled $225.[11] 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 the heaviest overdrafters overdrew more than 100 times. A similar skewed pattern is observed for the distribution of overdraft fees. Whereas the median overdraft fee is $77, the top 0.15% of heaviest overdrafters paid more than $2,730 in fees.

The majority of overdrafters have overdrawn less than 40 times. The first panel in Figure 1 depicts the distribution of the overdraft frequency and fees conditional upon the accountholder overdrafting at least once during the sample period. Notice that most consumers (>50%) overdraft only once or twice. The second panel shows the distribution censored at $300 for the overdraft fees paid for each accountholder who has overdrawn. Consistent with the fee structure where the standard per-item overdraft fee is $22 or $31, we see spikes at these two numbers and their multiples.

To better understand what types of transactions trigger an overdraft, we construct a table of the transaction channel that triggers overdrafts. We find (in Table 3) that nearly 50% of overdrafts are caused by debit card purchases with mean transaction amounts of approximately $30. On the other hand, Automated Clearing House (ACH) and check transactions account for 13.77% and 11.68% of overdraft incidents, and 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 approximately $90.

3.2. Exploratory Data Analysis

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' overdrafting behavior and that consumers are dissatisfied because of 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, we conjecture that a consumer may overdraw because of a much greater preference for current consumption than future consumption, this is, the consumer heavily discounts future consumption utility. At the point of sale, such a consumer sharply discounts the future cost of the overdraft fee to satisfy his or her immediate gratification.[12] In such a case, we should observe a steep downward sloping trend in the consumer's spending pattern within a pay period; that is, the consumer will increase spending right after he or she receives a pay check and will then reduce spending over the course of the month. However, because of his or her overspending at the beginning of the month, the consumer will run out of funds at the end of the pay period and have to overdraw.

We test this hypothesis with the following model of spending for consumer i at time t:

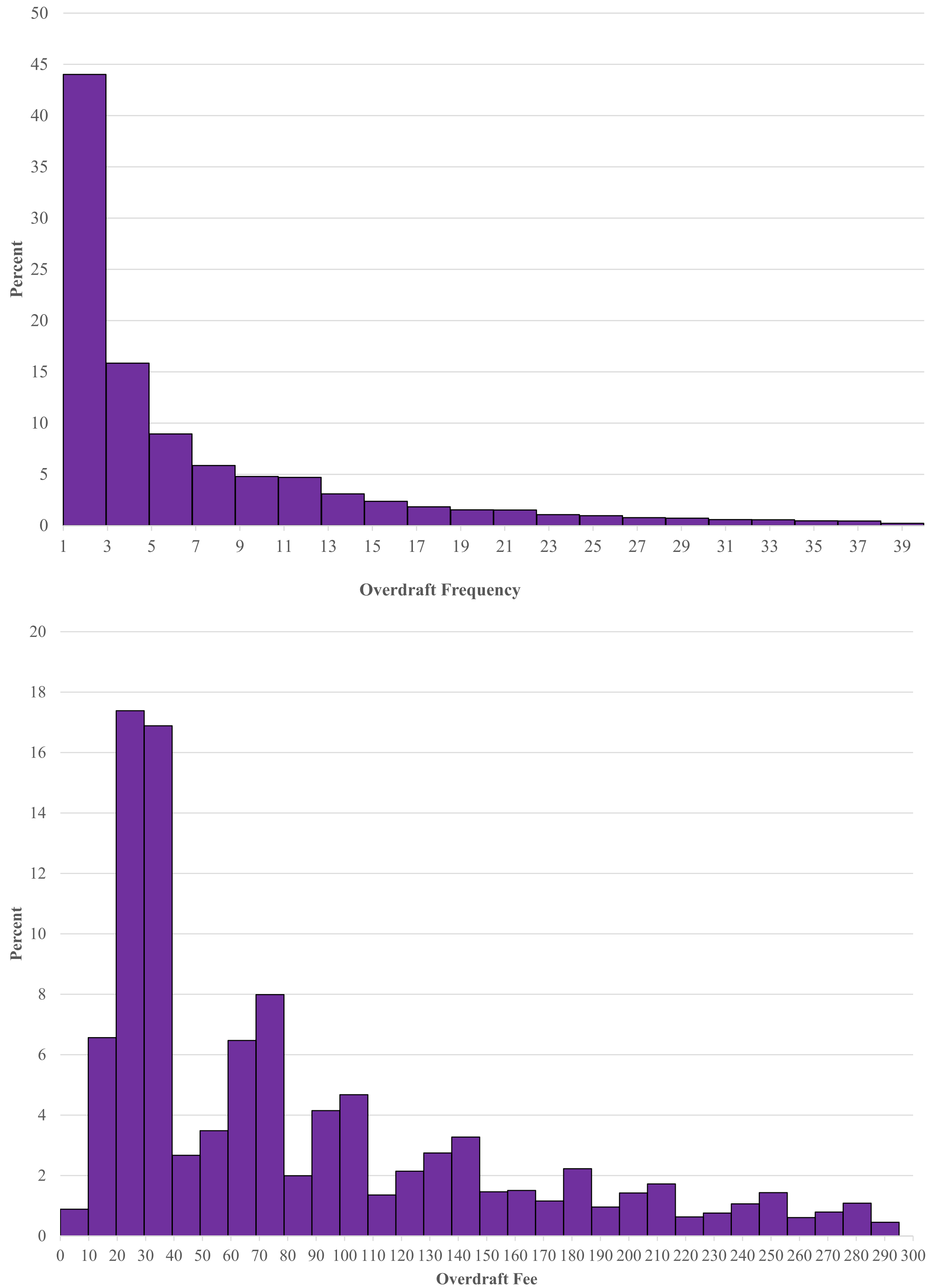
$$Spending_{it} = \beta * LapsedTimeAfterIncome_{it} + \mu_i + v_t + \epsilon_{it},$$

where $LapsedTimeAfterIncome_{it}$ is the number of days after the consumer received income (salary), μ_i is the individual

Table 2. Overdraft Frequency and Fee Distribution for Consumers Who Overdraft

	Mean	Standard deviation	Median	Minimum	99.85 Percentile
OD frequency	9.84	18.74	3	1	>100
OD amount	$245.46	$523.04	$77	$10	>$2,730

Figure 1. (Color online) Overdraft Frequency and Fee Distribution



fixed effect, and v_t is the time (day) fixed effect. To control for the effect that consumers usually pay for their bills (utilities, phone bills, credit card bills, etc.) after receiving their paycheck, we exclude checks and ACH transactions, which are the common choices for bill payments from daily spending and keep only debit card purchases, ATM withdrawals, and person-to-person transfers.

We run this ordinary least squares regression separately for heavy overdrafters (whose overdraft frequencies are in the top 20 percentile of all overdrafters), light overdrafters (whose overdraft frequencies are not in the top 20 percentile), and nonoverdrafters (who do not overdraw during the 15-month sample period).[13] The results are reported in columns (1)–(3) of Table 4.

Table 3. 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

We find that the coefficient of $LapsedTimeAfterIncome_{it}$ is negative and significant for heavy overdrafters but not light overdrafters or nonoverdrafters. This suggests that heavy overdrafters have a steep downward sloping spending pattern during a pay period, whereas light overdrafters or nonoverdrafters have a relatively stable spending stream. The heavy overdrafters are likely to overdraw because of their heavy discounting of future consumption.

3.2.2. Inattention. Next, we consider why light overdrafters may overdraft due to inattention. The idea is that consumers might not always monitor their account balance and may be 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 on our data, but to protect the privacy of the consumer and the merchants, the amounts have been changed. However, the example remains representative of the underlying data.

As shown in Figure 2, the consumer first received his or her biweekly salary on August 17. After a series of expenses, the consumer is left with $21.16 on August 20. The consumer did not check his or her balance but continued spending and overdrew the 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 totaled only $53.38 but caused the consumer to pay four overdraft item fees for total fees of $124. We speculate that this consumer was careless in monitoring his or her account and overestimated his or her balance.[14]

Beyond this example, we find more evidence of inattention in the data. To start off, Table 3 suggests that

Table 4. Spending Decreases with Time in a Pay Cycle

	(1) Heavy overdrafters	(2) Light overdrafters	(3) Nonoverdrafters
Lapsed time after income (β)	−6.8374***	−0.07815	−0.02195
	(0.06923)	(0.06540)	(0.02328)
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

***$p < 0.0001$.

Figure 2. (Color online) Overdraft Due to a Balance Perception Error

DATE	+/-	AMOUNT	BALANCE	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**

debit card purchases cause the most overdraft incidences, perhaps because it is more difficult to check balances when using a debit card than using an ATM or ACH.

Moreover, intuitively, as direct support of our hypothesis regarding inattention, the less frequently a consumer checks his or her balance, the more overdraft fees the consumer will likely incur. To test this hypothesis, we estimate the following specification:

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

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

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 columns (1) and (2) in Table 5.

The result suggests that a higher frequency of balance checking decreases the overdraft payment for light overdrafters but not for heavy overdrafters. We further test this effect by including the overdraft frequency ($ODFreq_{it}$) and an interaction term for balance-checking frequency and overdraft frequency $BCFreq_{it} * ODFreq_{it}$ 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 a high frequency of balance checking decreases overdraft fees only for consumers who overdraw infrequently, not for those who overdraw frequently:

$$\begin{aligned} TotODPmt_{it} = \beta_0 &+ \beta_1 BCFreq_{it} + \beta_2 ODFreq_i \\ &+ \beta_3 BCFreq_{it} * ODFreq_i + v_t + \epsilon_{it}. \end{aligned}$$

Table 5. Frequent Balance Checking Reduces Overdrafts for Light Overdrafters

	(1) Light overdrafters	(2) Heavy overdrafters	(3) All overdrafters
Balance-checking frequency (*BCFreq*, β_1)	−0.5001***	−0.1389	−0.6823***
	(0.0391)	(0.0894)	(0.0882)
Overdraft frequency (*ODFreq*, β_2)			16.0294***
			(0.2819)
BCFreq ∗ *ODFreq* (β_3)			0.278136***
			(0.0607)
Number of observations	1,794,835	593,676	2,388,511
R^2	0.1417	0.1563	0.6742

Notes. Fixed effects at the individual and day levels are shown. Robust standard errors are clustered at the individual level.
***$p < 0.0001$.

The results in column (3) of Table 5 confirm our hypothesis.[15]

Interestingly, we find that consumers' balance perception error accumulates over time in the sense that the greater the time elapsed without checking their balance, the more likely they are to overdraw and consequently pay higher amounts in overdraft fees. Figure 3 exhibits the overdraft probability across the number of days elapsed since the last time a consumer checked his or her balance for light overdrafters (whose overdraft frequency is not in the top 20 percentile). As the figure shows, the overdraft probability increases moderately with the number of days elapsed since the last balance check.

We confirm this relationship with the following two specifications. We assume that the 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

$$\begin{aligned} I(OD)_{it} &= \Phi(\rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} \\ &\quad + \rho_2 BeginBal_{it} + \mu_i + v_t), \\ ODFee_{it} &= \rho_0 + \rho_1 DaysSinceLastBalanceCheck_{it} \\ &\quad + \rho_2 BeginBal_{it} + \mu_i + v_t + \epsilon_{it}, \end{aligned}$$

where Φ is the cumulative distribution function for a standard normal distribution. The term $DaysSinceLastBalanceCheck_{it}$ denotes the number of days that consumer i has not checked his or her balance until time t, and $BeginBal_{it}$ is the beginning balance at time t. We control for the beginning balance because it may be negatively correlated with the days elapsed since last balance check because consumers tend to check their balance when it is low, and a lower balance often leads to an overdraft. Table 6 reports the estimation results, which support our hypothesis that the greater the time elapsed after a balance check, the more likely the consumer is to overdraw and incur higher overdraft fees.

Figure 3. (Color online) Overdraft Likelihood Increases with Time Elapsed Since the Last Balance Check

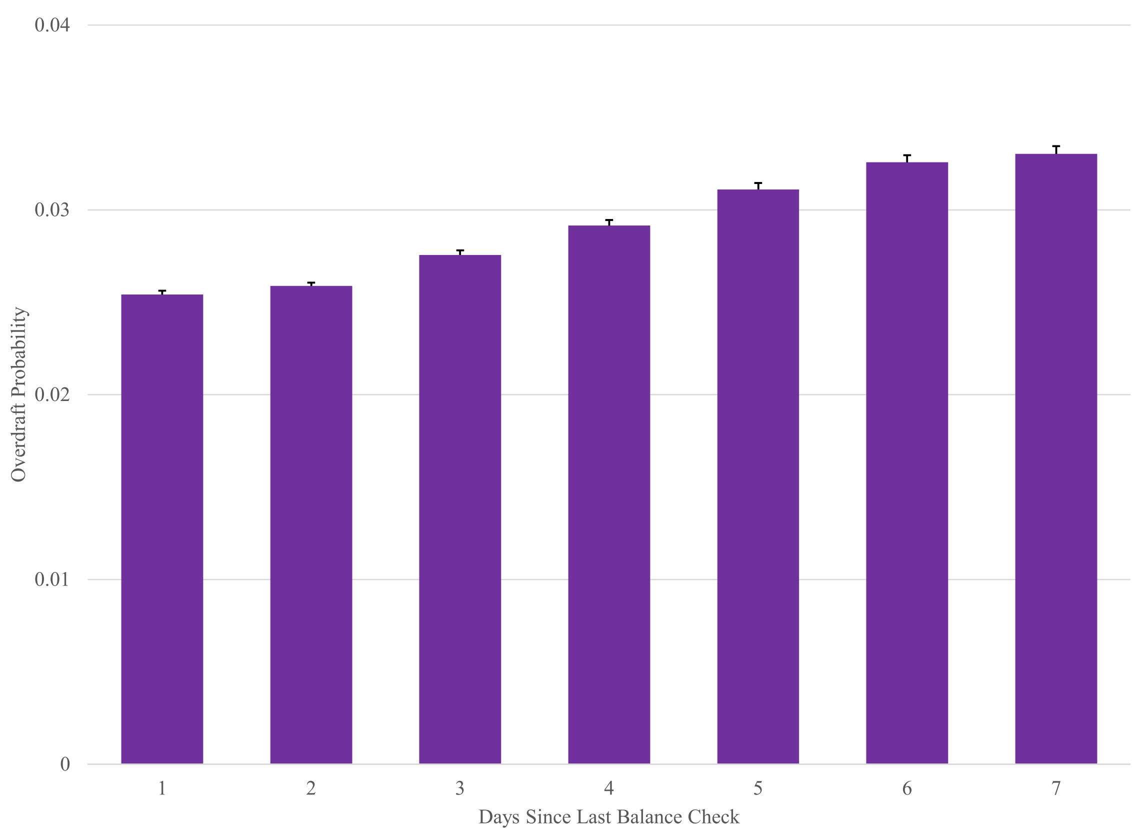


Table 6. Reduced-Form Evidence of the Existence of Monitoring Costs

	I (OD)	ODFee
Days since last balance check (ρ_1)	0.0415***	0.0003***
	(0.0027)	(0.0001)
Beginning balance (ρ_2)	−0.7265***	−0.0439***
	(0.0066)	(0.0038)
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

Notes. The estimation sample includes only overdrafters. Marginal effects for the probit model are shown. Fixed effects are for the individual and day levels. Robust standard errors are clustered at the individual level.

***$p < 0.0001$.

If balance checking can help prevent overdrafts, why do consumers not check their balance more frequently and avoid overdraft fees? We argue that there exists a monitoring cost because monitoring account balance is costly in terms of time, effort and mental resources, which reduces the number of balance checks. In the data, we find that consumers with online banking accounts check their balances more frequently than those without online banking accounts, suggesting that monitoring costs exist and that consumers monitor their account more frequently when these costs are reduced. Please see the details in Online Appendix A2.

3.2.3. Dissatisfaction.

We find that overdrafts might cause consumers to close their accounts (Table 7). Among nonoverdrafters, 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, whereas 10.56% of light overdrafters (whose overdraft frequency is not in the top 20 percentile) closed their accounts.

From the description field associated with each account, we can distinguish the cause of account closure: forced closure by the bank because the consumer is unable or unwilling to pay back the overdrawn balance and fees (in which case the bank executes a charge-off) versus voluntary closure. Among heavy overdrafters, 13.66% closed their accounts voluntarily, and the remaining 86.34% were forced by the bank to close their accounts (Table 8). In contrast, 47.42% of the light overdrafters closed their accounts voluntarily. We conjecture that the higher voluntary closures among light overdrafters may be due to customer dissatisfaction with the bank, as the evidence below shows.

Table 7. Account Closure Frequency for Overdrafters vs. Nonoverdrafters

	% Closed
Heavy overdrafters	23.36%
Light overdrafters	10.56%
Nonoverdrafters	7.87%

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	—
Nonoverdrafters	—	—	100.00

Notes. Nonoverdrafters might voluntarily close for reasons such as dwelling location change and account consolidation. Because of a lack of data, we cannot explain why nonoverdrafters are more likely to voluntarily close than light overdrafters.

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

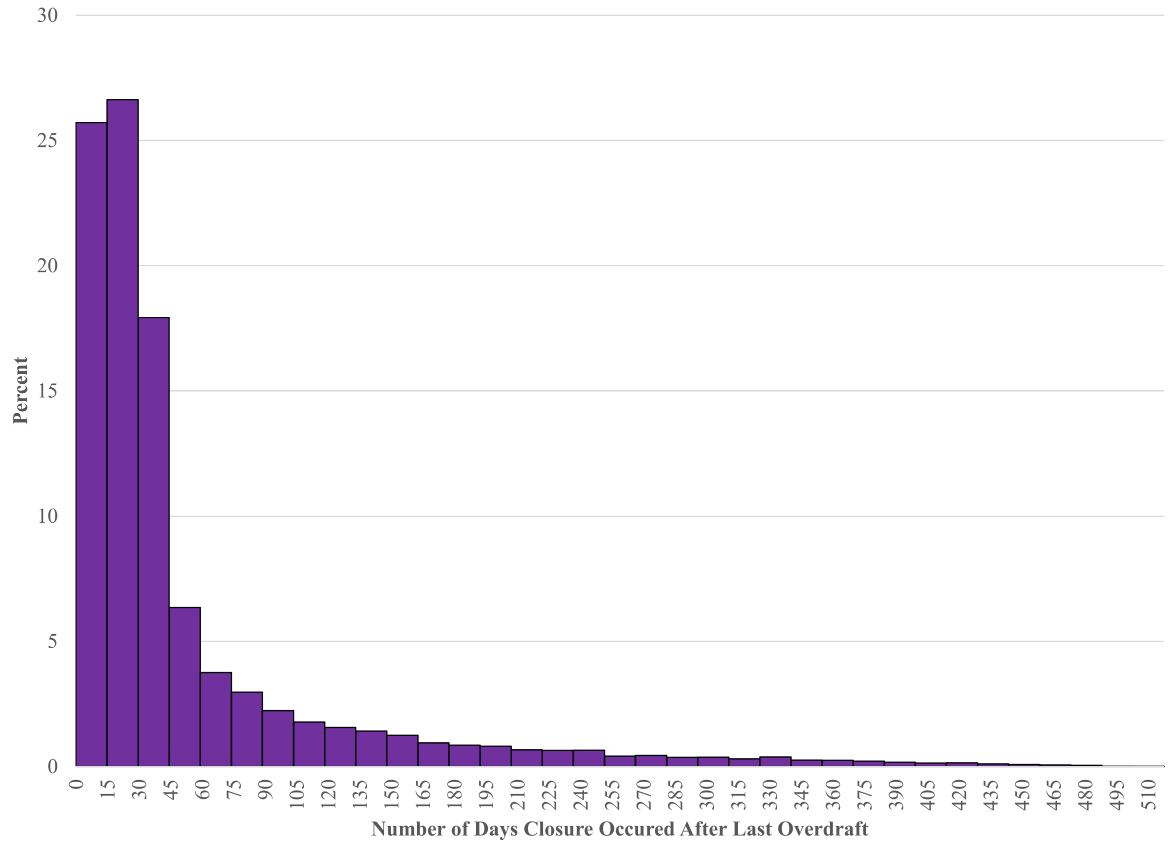
Second, light overdrafters are also more likely to close their accounts when the ratio of the overdraft fee to the transaction amount that caused the overdraft fee is higher. In other words, the higher the ratio of the overdraft amount, the more likely a consumer will be to close his or her account. We show this pattern in the left panel of Figure 5. However, this effect does not seem to be present for heavy overdrafters (right panel of Figure 5).

4. Model

Our exploratory data analysis indicates that heavy discounting and inattention can help explain consumers' overdrafting behavior and that consumers' dissatisfaction due to overdraft fees contributes to their attrition. Ultimately, our goal is to predict the overdraft incidence for each consumer on a daily basis. To do so, we develop a structural model that incorporates discounting, inattention, and dissatisfaction. This is a dynamic model in which consumers make daily decisions about how to spend their funds, whether to check their balance, and whether to close their account. We assume that consumers are rational and forward looking,[16] with an objective to maximize their total discounted utility by making optimal choices. As by-products of this model, we make a prediction about the likelihood of an overdraft on any given day for each consumer as well as the consumer's tenure and profitability with the bank.

The timing of the events for our model is illustrated in Figure 6. On each day for every consumer, the model has seven steps. First, the consumer receives income (if any). Second, the consumer's bills arrive (if any).

Figure 4. (Color online) Days to Closure After Last Overdraft



Third, the consumer decides whether to check his or her balance. If the consumer inquiries about his or her balance, then the beginning balance and bills for the day are known with complete certainty; otherwise, the consumer forms an estimate of both the beginning balance and the bill amount. Fourth, the consumer makes discretionary spending decisions and spends or consumes this amount. At this stage, the consumer chooses consumption (C) to maximize the total discounted utility (V) if the consumer chooses to check the balance or chooses it to maximize the expected discounted utility ($E[V]$) if the balance is not checked. Fifth, an overdraft fee is assessed by the bank if the ending balance is below zero. Sixth, the bank decides whether to force the account to close. Based on the bank's risk assessment, if the account is associated with high risk activities, it will be forced to close and the consumer will get zero utility (normalized). The bank will incur a default cost, which equals the negative balance of the account, in this case. Seventh, the consumer decides whether to close

Figure 5. (Color online) Percentage of Accounts Closed Increases with Fee/Transaction Amount Ratio

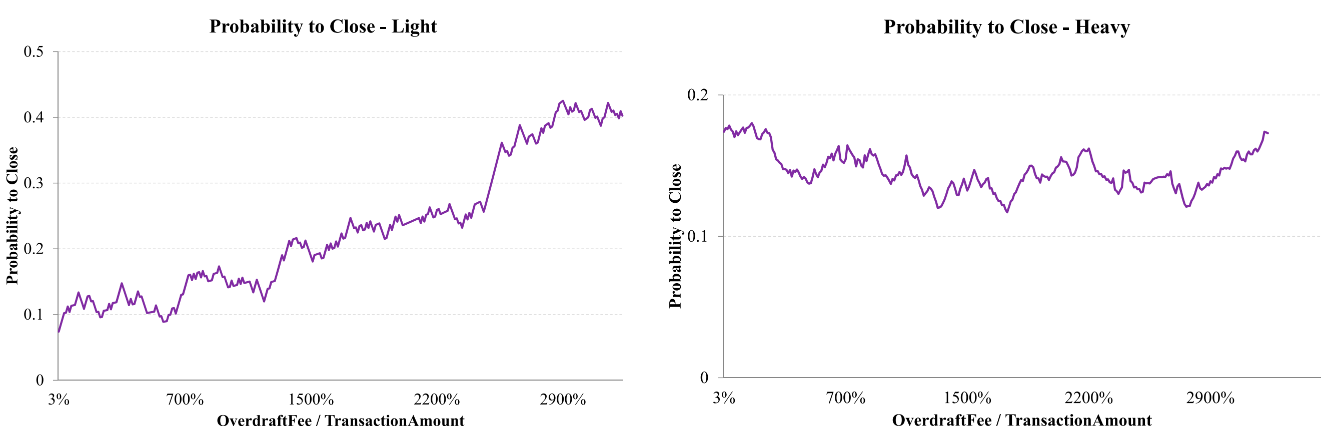
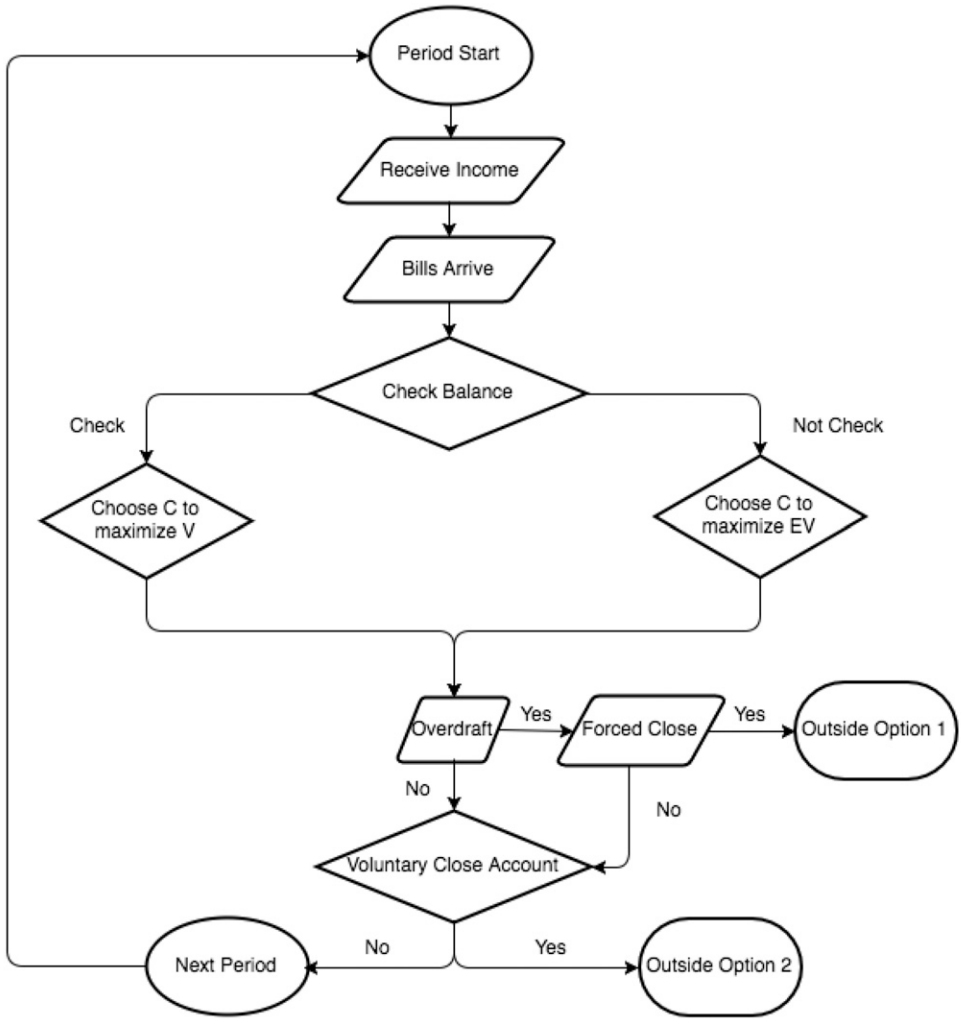


Figure 6. Timing of Events Within Our Model for Each Day



the account (after paying any overdraft fees). If the consumer closes his or her account, then an outside option is received and the model ends. Otherwise, in the eighth step, the balance is updated, and the cycle of events is repeated daily.[17]

To fully implement this multistage model, we have to specify a number of components concerning utility. To make it easier for the reader to follow the specification of the model, we do not start with its full specification, but instead build up the model starting with the utility from consumption in Section 4.1. We then incorporate adjustments to utility of not knowing the balance with certainty, which incorporates monitoring costs that capture inattention in Section 4.2. Finally, we introduce dissatisfaction with overdraft fees into utility in Section 4.3 so that we can predict not only when overdrafts occur but also when consumers will close their accounts. The full dynamic programming problem that considers the net present value of utility for an individual consumer is presented in Sections 4.4 and 4.5. Finally, we consider how heterogeneity across consumers can be specified in Section 4.6 to complete the full specification of the model. The estimation of the model is discussed in Section 5.

4.1. Consumption Model

At the core of our model is the need to predict consumers' daily decisions about how much to consume today versus in the future. Following the lifetime consumption literature (Modigliani and Brumberg 1954, Hall 1978), we assume that consumer i's per-period consumption utility at time t is determined by a constant relative risk averse (CRRA) utility function (Arrow 1963):

$$u_C(C_{it}) = \frac{C_{it}^{1-\theta_{it}}}{1-\theta_{it}}, \tag{1}$$

where C_{it}[18] is consumer i's consumption at period t, and θ_{it} is the coefficient of relative risk aversion. We choose the CRRA utility function because it has the merits of empirical support (Friend and Blume 1975), analytical convenience (Merton 1992), and is commonly used in the economics literature. The coefficient of relative risk aversion is always positive, and its inverse $\frac{1}{\theta_{it}}$ is the intertemporal substitution elasticity between consumption in any two adjacent periods. Higher values for θ_{it} imply greater utility from each marginal unit of consumption and a lower willingness to substitute today's consumption for future consumption.

Given that consumers might overdraw because of emergency expenses, for example, medical bills, car repairs, or expensive group dinners, we allow θ_{it} to vary each period according to a random shock term ε_{it} to capture these unexpected needs for consumption. Large positive values of ε_{it} will result in increased marginal utilities of consumption, representing days when urgent expenses are due. Specifically, we allow θ_{it} to follow a log-normal distribution with a time-invariant location θ_i and a random shock ε_{it}. The shock ε_{it} follows a normal distribution with mean zero and variance ς_i^2 (Yao et al. 2012):

$$\theta_{it} = \exp(\theta_i + \varepsilon_{it}),$$
$$\varepsilon_{it} \sim N(0, \varsigma_i^2).$$

The consumption plan captured by C_{it} depends on the consumer's budget constraint, which is a function of the consumer's current balance B_{it}, income Y_{it}, and bills Ψ_{it}. Bills represent preauthorized spending that relates to medium- or long-run consumption expenditures such as loan, rent, or utility payments. We model bills separately[19] from consumption because preauthorized spending is difficult to change on a daily basis after it is authorized, whereas consumption is more likely to be the result of consumers' day-to-day decisions. Consumers' budget constraints are as follows:

$$B_{it+1} = B_{it} - C_{it} - OD_{it} * I(B_{it} - C_{it} < 0) + Y_{it+1} - \psi_{it+1}. \quad (2)$$

The next day's available balance B_{it+1} is equal to the current balance B_{it} minus current consumption C_{it} and overdraft fees OD_{it} (if the balance becomes negative, denoted by $(B_{it} - C_{it} < 0)$) plus the next day's net income after bills, given by the difference of Y_{it+1} and ψ_{it+1}. Note that because we model consumers' spending decisions at the daily level rather than the transaction level, we aggregate all overdraft fees paid and assume that consumers know the per-item fee structure[20] stated in Section 3 when deciding their daily consumption; that is, OD_{it} is not the per-item overdraft fee ($31 for our focal bank), but the daily sum of all per-item overdraft fees. Thus, OD_{it} can take a value such as $62, $93, or $124.

Because our focus is overdrafting behavior, we make a number of assumptions to render our problem tractable. First, consumption is not observed in our data; therefore, we make the assumption that spending is equivalent to consumption in terms of generating utility. Hereafter, we use consumption and spending interchangeably. Second, we abstract away from the complexity associated with our data and assume that the consumer's income and bills are exogenously determined. In our data set, we are able to distinguish bills from spending using their transaction channels, as illustrated in Table 1. For example, bills are associated with checks and ACH and bill payments, whereas spending is associated with debit cards and cash withdrawals. Note that although credit card spending is discretionary, we treat it as a bill because it affects the checking account balance only when the consumer pays the bill rather than when the consumer swipes the credit card each time. Thus, credit card spending does not cause any immediate overdrafts, whereas debit card purchases may. It is for this reason that we treat credit card and debit card spending differently. The main focus of our paper is to examine overdrafts. Thus, the transactions are modeled according to the extent that they affect overdrafts. Third, we assume that bills are not within consumers' daily discretion but that spending (or, more precisely, nonpreauthorized spending) can be adjusted daily. In summary, we model consumers' consumption decisions, where consumption is nonpreauthorized spending[21] from their checking accounts.

4.2. Inattention and Monitoring Costs

Our reduced-form evidence in Section 3.2.2 suggests that because of monitoring costs, consumers are inattentive to their financial well-being. This is consistent with the theory of rational inattention (Sims 1998, 2003) that individuals have many things to think about and limited time, and they can devote only limited intellectual resources to these tasks of data gathering and analysis. Because monitoring an account balance[22] takes time and effort, a consumer may not check his balance frequently enough to avoid overdrafts. To capture this effect, we assume that consumers are rational inattentive[23] in the sense that they are aware of their own inattention and may choose to be inattentive if monitoring costs are high (Grubb 2014). Specifically, we model consumers' balance-checking behavior as a binary choice: $Q_{it} \in \{1,0\}$, where 1 denotes the decision to check the balance and 0 denotes the decision not to check.

The balance-checking activity affects the consumer's balance perception $\widetilde{B_{it}}$. As shown in Equation (2), the available balance B_{it+1} is the income net of preauthorized bills ψ_{it+1}. We assume that consumers are not fully aware of the bill arrival timing and amount. For example, because of the business day schedule, a credit card bill might arrive on any day in the last few days of a month or the first few days of a month. And the amount of a utility bill may vary by season. Therefore consumers have uncertainty of the exact bill, and henceforth the available balance. On the one hand, if a consumer checks his balance by incurring a monitoring cost (to be explained later), then the available balance B_{it}

will be known with certainty. On the other hand, if the consumer does not check his balance, he will recall a perceived balance, which gives a noisy measure of true balance;[24] that is,

$$\begin{aligned} \widetilde{B_{it}} &= B_{it} && \text{if } Q_{it} = 1, \\ \widetilde{B_{it}} &\sim N(B_{it} + \eta_{it}\omega_{it}, \omega_{it}^2) && \text{if } Q_{it} = 0. \end{aligned} \tag{3}$$

Following Mehta et al. (2004), we allow the perceived balance $\widetilde{B_{it}}$ when the balance is not checked to be a normally distributed random variable. The mean of $\widetilde{B_{it}}$ is the sum of the true balance B_{it} and a perception error: $\eta_{it}\omega_{it}$. The first component of the perception error η_{it} is a random draw from the standard normal distribution,[25] and the second component is the standard deviation of the perception error, ω_{it}. The variance of $\widetilde{B_{it}}$ is ω_t^2, which measures the extent of uncertainty.

For notational convenience, we introduce a variable Γ_{it} that measures the time (number of days) elapsed since the consumer last checked the balance. By definition, $\Gamma_{it+1} = (1 + \Gamma_{it})(1 - Q_{it})$; that is, if a consumer checks the balance ($Q_{it} = 1$), then the time lapsed since last balance check is 0, but if the consumer does not check the balance ($Q_{it} = 0$), the time elapsed increases by one day. Based on the evidence from our exploratory data analysis, we allow the extent of uncertainty of the perceived balance to accumulate over time, which implies that the longer a consumer goes without checking his balance, the more inaccurate the perceived balance will be. We formulate this statement as

$$\omega_{it}^2 = \rho_i \Gamma_{it}, \tag{4}$$

where Γ_t denotes the time elapsed since the consumer last checked his balance and ρ denotes the sensitivity to the time elapsed since the last balance check.[26]

Recall that a consumer incurs the monitoring cost to check the balance. The monitoring cost is an opportunity cost, not an explicit cost charged by the bank. Formally, we calculate the consumption utility based on this monitoring cost:

$$\overline{u_t}(C_{it}, Q_{it}, \widetilde{B_{it}}) = u_C(C_{it}, \widetilde{B_{it}}) - Q_{it}\xi_i + \chi_{itQ_{it}}, \tag{5}$$

where ξ_i is the consumer's monitoring cost, and $\chi_{itQ_{it}}$ is the idiosyncratic shock that affects his monitoring cost. The shock $\chi_{itQ_{it}}$ can capture idiosyncratic events such as vacations, during which it is difficult for consumers to monitor their balances, or it could capture increased awareness about consumers' financial states from other events such as online bill payments, which automatically report their balances. The equation implies that if a consumer checks his balance, then the utility decreases by a monetary equivalence of $|(1 - \theta_{it})\xi_i|^{\frac{1}{1-\theta_{it}}}$. We assume that $\chi_{itQ_{it}}$ are independent and identically distributed (i.i.d.) and follow a type I extreme value distribution.

The consumer does not know the balance perception error (η_{it} and $\widetilde{B_{it}}$), so he forms an expected utility based on his knowledge about the distribution of their perception error. The optimal spending will maximize his expected utility, which is calculated by integrating over the balance perception error:

$$u_{it} = \iint \overline{u_t}(C_{it}, Q_{it}, \widetilde{B_{it}})dF(\eta_{it})dF(\widetilde{B_{it}}). \tag{6}$$

The expected utility u_{it} is decreasing with the variance in the perception error ω_{it}^2 (through $\widetilde{B_{it}}$; see Equation (4)). This relationship arises because greater variance in the perception error decreases the accuracy of the consumer's estimate of his true balance and thus increases the likelihood that he will mistakenly overdraw his account, which lowers his utility. The derivation is shown in a technical report available from the authors.

4.3. Forced Closure

As Table 9 suggests, the bank often forces an account to close once an overdraft occurs. Therefore, when a consumer overdraws, he faces the risk of being forced out of the bank. To take this into account, we add the probability of forced closure into the model and denote it by π_f. Specifically, after a consumer decides how much to consumer or spend (C), if an overdraft occurs, he faces a lottery of forced closure. With probability π_f he will be forced to leave the bank and obtain the outside option 1 (to be differentiated from the other outside option 2 after voluntary closure; we normalize the outside option 1 to be 0), whereas with probability $1 - \pi_f$ the bank will keep the account open and the consumer will make the next decision of whether to voluntarily close the account. We found the data evidence that an account is more likely to be forced to close when the overdraft payment due is high.[27] Therefore, we model $\pi_f = \kappa_i * OD_{it}$,[28] and κ_i is the parameter to be estimated.

Table 9. Model Comparison

	A: No forward looking	B: No inattention	C: No heterogeneity	D: Proposed
log-marginal density	−2,937.50	−3,625.43	−2,749.95	−1,741.91
Hit rate: overdraft	0.539	0.385	0.516	0.882
Hit rate: check balance	0.425	0.242	0.647	0.846
Hit rate: close account	0.670	0.746	0.681	0.782

4.4. Dissatisfaction and Account Closing

The exploratory analysis in Section 3.2.3 suggests that overdrafts trigger consumer dissatisfaction and attrition. We model attrition as consumers choosing an outside option of closing their account and switching to a competing bank or becoming unbanked. Based on the data pattern in Figure 5, we make an assumption that consumers are sensitive to the ratio of the overdraft fee to the overdraft transaction amount, and we use Ξ_{it} to denote this ratio as a state variable. We assume that a larger ratio indicates a higher likelihood that the consumer will be dissatisfied, because the ratio is essentially the implicit price (interest rate) of overdrafts, and prior research (Keaveney 1995, Bolton 1998) has documented that a high price may cause consumer dissatisfaction. Forward-looking consumers anticipate the accumulation of dissatisfaction (as well as lost consumption utility due to overdrafts) in the future and will become more likely to close their accounts. Furthermore, we assume that consumers formulate their belief of the ratio for a future period based on the highest ratio they have personally incurred;[29] that is, if we use Δ_{it} to denote the per-period ratio, then

$$\Delta_{it} = \frac{OD_{it}}{|\widetilde{B_{it}} - C_{it}|},$$

and

$$E[\Xi_{it+1}|\Xi_{it}] = \max(\Xi_{it}, \Delta_{it}). \tag{7}$$

This assumption is made based on the findings from Tversky and Kahneman (1973), Nwokoye (1975), Monroe (1990), and Fiske and Taylor (1991) that extremely high prices are comparatively distinct, more salient, and easier to retrieve from memory, so that they are more likely to be used as anchors in memory-based tasks. This assumption also reflects consumers' learning behavior over time. Consider a consumer who experiences many overdrafts; our model captures the idea that his dissatisfaction grows with each overdraft.

To introduce dissatisfaction from overdrafts into the model, we calculate the per-period utility:

$$\overline{U_{it}} = u_{it} - \Upsilon_{\mathrm{i}} * \max(\Xi_{it}, \Delta_{it}) * I[\widetilde{B_{it}} - C_{it} < 0].$$

In the above equation, u_{it} is defined as in Equation (6), and Υ_{i} is the dissatisfaction sensitivity, that is, the impact of charging an overdraft fee on a consumer's decision to close the account. As in Equation (6), we integrate out the balance perception error to get the expected utility:

$$U_{it} = \iint \overline{U_{it}}(C_{it}, Q_{it}, \widetilde{B_{it}}) dF(\eta_{it}) dF(\widetilde{B_{it}}). \tag{8}$$

We assume that the decision to close the account is a terminal decision. Once a consumer chooses to close his or her account, their value function (or total discounted utility function) equals an outside option with a mean value of α_i, normalized to be the same across states for identification purposes.[30] If the consumer keeps the account open, continuation values from future per-period utility functions will continue to be received. More specifically, let W_i denote a consumer's choice to close his or her account, where $W_i = 1$ denotes the decision to close the account before the period starts, and $W_i = 0$ denotes the decision to keep the account open for this period. Then, the value function for the consumer becomes

$$V_{it} = \begin{cases} \{U_{it} + \varpi_{it0} + \beta_i\ E[V_{it+1}\ |S_{it}]\}\{(1-\pi_f) \\ \qquad * I[\widetilde{B_{it}} - C_{it} < 0] + I[\widetilde{B_{it}} - C_{it} \geq 0]\}, & W_{it} = 0 \\ \alpha_i + \varpi_{it1}, & W_{it} = 1 \end{cases} \tag{9}$$

where ϖ_{it0} and ϖ_{it1} are the idiosyncratic shocks that determine a consumer's account-closing decision. Sources of shocks may include events such as when the consumer moves out of town or when a competing bank enters the market. We assume that these shocks follow a type I extreme value distribution. Note that even when a consumer decides to keep the account open, the bank might force the account to close with probability π_f when the account is overdrawn. So, in expectation, the consumer's value function is $\{U_{it} + \varpi_{it0} + \beta_i\ E[V_{it+1}\ |S_{it}]\}$ $\{(1-\pi_f) * I[\widetilde{B_{it}} - C_{it} < 0] + I[\widetilde{B_{it}} - C_{it} \geq 0]\}$.

4.5. State Variables

In this subsection, we formalize the statistical properties associated with the state variables so that we can complete the specification by stating consumers' expectations about their future state.

Income. Consumer accounts tend to have regular spikes in deposits that correspond with monthly, weekly, or biweekly periods. Specifically, we assume that the distribution for income is

$$Y_{it} = Y_i * I(DL_{it} = PC_i),$$

where Y_i is the stable periodic (monthly/weekly/biweekly) income, DL_{it} is the number of days left until the next payday, and PC_i is the length of the pay cycle. The transition process of DL_i is deterministic $DL_{it+1} = DL_{it} - 1 + PC_i * I(DL_{it} = 1)$, decreasing by one for each period ahead and returning to the full length when one pay cycle ends.

Overdraft fee. The state variable OD_{it} is assumed to be i.i.d. over time[31] and to follow a discrete distribution with the support vector and probability vector $\{X_i, p_i\}$. The support vector contains multiples of the per-item overdraft fee.

Bills. Our data suggests that the bills arrival dates vary considerably. For example, depending on the business day schedule, monthly credit card bills can arrive on any day between the 27th day of the month and 3rd day of the consecutive month. In addition, even when the date is fixed, the amount varies. For example, in many areas, utility bills in the winter are much higher than those in the summer. Given the stochastic nature of the timing and amount of bills, we treat bill arrival dates as a random variable, which we model as a Poisson process. Specifically, bills are assumed to be i.i.d. draws from a compound Poisson distribution with arrival rate ϕ_i and with a jump size distribution G_i: $\Psi_{it} \sim CP(\phi_i, G_i)$. 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).[32] This model setup is consistent with the lifetime consumption model with *unanticipated and transitory* income shocks.[33] When consumers move money from a savings account to a checking account, this is recorded as a negative bill, which increases the available budget. Savings (and ATM deposits) are modeled in the same way as bills. This is because in the data, savings are recorded as either cross-institution ACH transactions or within-bank transfers, similarly to bills. Moreover, we treat savings similarly to bills because, although both generate consumption utility for consumers, they are usually preauthorized and do not cause any immediate overdrafts. People usually do not make retirement and other long-term savings decisions on a daily basis. However, overdrafts result from negative balances on a daily level. The main purpose of our paper is to examine overdrafts. Thus, transactions are modeled based on the extent to which they affect overdrafts.

Dissatisfaction. The ratio of the overdraft fee to the overdraft transaction amount evolves by keeping the maximum amount over time (see Equation (7)).

Open status. The account status is denoted by OP_{it}. If $OP_{it} = 1$, then the account is open. If $OP_{it} = 0$, then the account is closed. The transition of this state variable is deterministic:

$$f(OP_{it+1}|OP_{it}, W_{it}) = \begin{cases} 0, & \text{if } OP_{it} = 0, \\ 1, & \text{if } W_{it} = 0 \text{ and } OP_{it} = 1, \\ 0, & \text{if } W_{it} = 1 \text{ and } OP_{it} = 1. \end{cases}$$

Random errors. The shocks ε_{it}, χ_{it} and ϖ_{it} are all assumed to be i.i.d. over time.

In summary, the whole state space for consumers is $S_{it} = \{\widetilde{B_{it}},$[34] $\Psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \mathrm{OP}_{it}, \varepsilon_{it}, \chi_{it}, \varpi_{it}\}$. In our data set, we observe $\widehat{S_{it}} = \{B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, \mathrm{OP}_{it}\}$, and our unobservable state variables are $\widetilde{S_{it}} = \{\widetilde{B_{it}}, \eta_{it}, \varepsilon_{it}, \chi_{it}, \varpi_{it}\}$. So $S_{it} = \widehat{S_{it}} \cup \widetilde{S_{it}} \backslash \{B_{it}, \psi_{it}\}$. Notice here that consumers also have unobserved states B_{it} and ψ_{it} due to inattention. If a consumer checks his balance, then the true balance (B_{it}) and bill amount (ψ_{it}) are known; otherwise, a perceived balance ($\widetilde{B_{it}}$) and expected bill (Ψ_{it}) are known.

4.6. The Dynamic Optimization Problem and Intertemporal Trade-off

We can now state the complete optimization problem facing each consumer. Each consumer chooses an infinite sequence of decision rules $\{C_{it}, Q_{it}, W_{it}\}_{t=1}^{\infty}$ to maximize the expected total discounted utility:

$$\max_{\{C_{it}, Q_{it}, W_{it}\}_{t=0}^{\infty}} E_{\{S_{it}\}_{t=0}^{\infty}} \left\{ \sum_{t=0}^{\infty} \beta_i^t U_t(C_{it}, Q_{it}, W_{it}, S_{it}) | \right.$$
$$\left. C_{i0}, Q_{i0}, W_{i0}, S_{i0} \right\},$$

where

$$\begin{aligned} U_t(C_{it}, Q_{it}, W_{it}, S_{it}) = \Bigg[\iint \Bigg\{ \Bigg[\Bigg(& \frac{C_{it}^{1-\theta_{it}}}{1-\theta_{it}} - Q_{it}\xi_i + \chi_{itQ_{it}} \Bigg) \\ & - \Upsilon_\mathrm{i} * \max(\Xi_{it}, \Delta_{it}) \\ & * I[\widetilde{B_{it}} - C_{it} < 0] + \varpi_{it0} \Bigg] \\ & * \{(1-\pi_f) * I[\widetilde{B_{it}} - C_{it} < 0] \\ & + I[\widetilde{B_{it}} - C_{it} \geq 0]\} \Bigg\} dF(\eta_{it}) dF(\widetilde{B_{it}}) \\ & * (1 - W_{it}) + (\alpha_i + \varpi_{it1}) W_{it} \Bigg] * OP_{it}. \end{aligned}$$

Let $V(S_{it})$ denote the value function:

$$\begin{aligned} V(S_{it}) = \max_{\{C_{i\tau}, Q_{i\tau}, W_{i\tau}\}_{\tau=t}^{\infty}} E_{\{S_{i\tau}\}_{\tau=t+1}^{\infty}} \Bigg\{ & U_t(C_{it}, Q_{it}, W_{it}, S_{it}) \\ & + \sum_{\tau=t+1}^{\infty} \beta_i^{\tau-t} U_\tau(C_{i\tau}, Q_{i\tau}, W_{i\tau}, S_{i\tau}) | \, C_{it}, Q_{it}, W_{it}, S_{it} \Bigg\}. \end{aligned} \tag{10}$$

This infinite period dynamic optimization problem can be solved through the Bellman (1957) equation:

$$\begin{aligned} V(S_{it}) = \max_{C_i, Q_i, W_i} E_{S_{it+1}} \{ & U(C_i, Q_i, W_i, S_{it}) \\ & + \beta V(S_{it+1}) | C_i, Q_i, W_i, S_{it} \}. \end{aligned} \tag{11}$$

In the infinite horizon dynamic programming problem, the policy function does not depend on time. We can thus eliminate the time subscript.

Consequently, we have the following choice-specific value function:[35]

$$v(C,Q,W,\tilde{B},\Psi,Y,DL,OD,\Gamma,\Xi,OP,\varepsilon,\chi,\varpi)$$
$$=\begin{cases}\{(1-\pi_f)*I[B-C<0]+I[B-C\geq 0]\} \\ \quad\cdot\Big\{u_C(C)-\xi+\chi_1-\Upsilon*\max\Big(\Xi,\dfrac{OD*I[B-C<0]}{|B-C|}\Big) \\ \quad+\varpi_0+\beta\,E_{S_{+1}}\Big[V(\tilde{B}_{+1},\Psi_{+1},Y_{+1},DL_{+1},OD_{+1},1,\Xi_{+1}, \\ \quad 1,\varepsilon_{+1},\chi_{+1},\varpi_{+1})|C,Q,W,\tilde{B},\Psi,Y, \\ \quad DL,OD,\Gamma,\Xi,1,\varepsilon,\chi,\varpi\Big]\Big\}, \\ \qquad\qquad\qquad\qquad \text{if } Q=1 \text{ and } W=0 \\ \iint\{(1-\pi_f)*I[\tilde{B}-C<0]+I[\tilde{B}-C\geq 0]\} \\ \quad\cdot\Big\{u_C(C)+\chi_0-\Upsilon*\max\Big(\Xi,\dfrac{OD*I[\tilde{B}-C<0]}{|\tilde{B}-C|}\Big) \\ \quad+\varpi_0+\beta\,E_{S_{+1}}\Big[V(\tilde{B}_{+1},\Psi_{+1},Y_{+1},DL_{+1},OD_{+1},\Gamma+1, \\ \quad \Xi_{+1},1,\varepsilon_{+1},\chi_{+1},\varpi_{+1})|C,Q,W,\tilde{B},\Psi,Y, \\ \quad DL,OD,\Gamma,\Xi,1,\varepsilon,\chi,\varpi\Big]\Big\}dF(\eta)dF(\tilde{B}), \\ \qquad\qquad\qquad\qquad \text{if } Q=0 \text{ and } W=0 \\ a+\varpi_1, \qquad\qquad\qquad \text{if } W=1\end{cases}$$

where subscript "+1" denotes the next time period. Therefore, the optimal policy is

$$\{C_i^*,Q_i^*,W_i^*\}=\arg\max\, v(C_i,Q_i,W_i,\widetilde{B_\iota},\Psi_i,Y_i,DL_i,OD_i,\Gamma_i,\Xi_i,OP_i,\varepsilon_i,\chi_i,\varpi_i).$$

We note that a distinction exists between this dynamic programming problem and traditional ones. Because of the perception error, a consumer observes $\tilde{B}_{it}=B_{it}+\eta_{it}\omega_{it}$ but does not know B_{it} or η_{it}. The consumer knows only the distribution $(B_{it}+\eta_{it}\omega_{it},\omega_{it}^2)$ and makes a decision $C_i^*(\tilde{B}_{it})$ based on the perceived balance $\tilde{B}_{it}$. However, we—as analysts—do not know the realized perception error η_{it}. We observe the true balance B_{it} and the consumer's spending $C_i^*(\tilde{B}_{it})$. Therefore, we can assume only that $C_i^*(\tilde{B}_{it})$ maximizes the *expected ex ante value function*. Later, we look for parameters that make the likelihood of $C_i^*(\tilde{B}_{it})$, which maximizes the expected ex ante value function, reach its maximum. Following Rust (1987), we obtain the ex ante value function that integrates out the cost shocks, preference shocks, account-closing shocks, and unobserved mean balance error:

$$EV(B_i,\psi_i,Y_i,DL_i,OD_i,\Gamma_i,\Xi_i,OP_i)$$
$$=\iiiint v(C_i^*,Q_i^*,W_i^*,\widetilde{B_\iota},\Psi_i,Y_i,DL_i,OD_i,\Gamma_i,\Xi_i,OP_i,\varepsilon_i,\chi_i,\varpi_i)\,d\eta_i d\varepsilon_i d\chi_i d\varpi_i.$$

In summary, consumers' intertemporal trade-offs are associated with three dynamic decisions. First, given the budget constraint, a consumer will evaluate the utility of spending (or consuming) today versus tomorrow. Higher spending today implies lower spending in the future. Therefore, spending is a dynamic decision, and the optimal choice for the consumer is to smooth out his or her consumption over time. Second, when deciding when to check his or her balance, the consumer compares the monitoring cost with the expected gain from avoiding an overdraft fee. The consumer checks his or her balance only when the expected overdraft fee is higher than the monitoring cost. Because the consumer's balance perception error might accumulate over time, the consumer's overdraft probability also increases as more time elapses since the last balance check. As a result, the consumer waits until the overdraft probability reaches a certain threshold (when the expected overdraft fee equals the monitoring cost) before checking 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 close the account. When the consumer expects to incur many overdraft fees and the accompanying dissatisfaction, the consumer finds it more attractive to take the outside option and close his or her account.

4.7. Heterogeneity

In our data, consumers exhibit different responses to their state conditions. For example, some consumers have never checked their balance and frequently overdraw, whereas other consumers frequently check their balance and rarely overdraw. We hypothesize that these differences result from their heterogeneous discount factors and monitoring costs. Therefore, our model needs to account for unobserved heterogeneity. We follow a hierarchical Bayesian framework (Rossi et al. 2005) and incorporate heterogeneity into the model by assuming that all parameters, namely, θ_i (mean relative risk averse coefficient), β_i (discount factor), ς_i (standard deviation of the coefficient of risk aversion), ξ_i (monitoring cost), ρ_i (sensitivity of the error variance to the time elapsed since the last balance check), Υ_i (dissatisfaction sensitivity), and α_i (mean value of the outside option) each have a random coefficient specification. The subscript i denotes the consumer, and the previous models are understood to be defined by the customer. For each of these parameters, $\theta \in \{\theta_i, \beta_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i, \alpha_i, \kappa_i\}$, the prior distribution is defined as $\theta \sim N(\mu_\vartheta, \sigma_\vartheta^2)$. The hyper-prior distribution is assumed to be diffuse.

5. Identification and Estimation

5.1. Identification

The unknown structural parameters in the model include $\{\theta_i, \beta_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i, \alpha_i\}$, where θ_i is the logarithm of the mean of the coefficient of risk aversion, β_i is the discount factor, ς_i is the standard deviation of the coefficient of risk aversion, ξ_i is the monitoring cost, ρ_i is the sensitivity of the balance error variance to the time elapsed since the last balance check, Υ_i is the dissatisfaction sensitivity, α_i is the mean value of the outside option, and κ_i is the forced closure probability sensitivity to overdraft fee payment. We provide the rationale for the identification of each parameter.

We know from Rust (1987) that the discount factor β_i cannot be separately identified from the static utility parameter, which is the risk aversion coefficient in our case. The reason is that lowering θ_i tends to increase consumption/spending, an effect that can also be achieved by lowering β_i. Because we are more interested in consumers' time preference than their risk preference, we fix the risk aversion coefficient, which allows us to identify the discount factor.[36] This practice is also used by Gopalakrishnan et al. (2014). Following the latest research by Andersen et al. (2008), who jointly elicit risk and time preferences, we choose $\theta_i = \theta = 0.74$ for the coefficient of risk aversion.[37] After we fix θ_i, 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, whereas a small discount factor implies a downward-sloping consumption stream. Because a discount factor is constrained above by 1, we take a logit transformation, namely, $\beta_i = \frac{1}{1+\exp(\lambda_i)}$, and estimate the transformed parameter λ_i instead.

The standard deviation of the coefficient risk aversion 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 trade-off, the longer the consumer goes without checking his or her balance, the more likely an overdraft is to occur because of a balance error. Therefore, the observed data pattern of higher overdraft fees paid for a longer period after the balance is checked can inform the structural parameter ρ_i.

Intuitively, the monitoring cost ξ_i is identified by the expected overdraft payment amount. Recall that the trade-off regarding balance checking is that a consumer checks his or her balance only when ξ_i is smaller than the expected overdraft payment amount. In the data, we observe the balance-checking frequency. By combining this with the calculated ρ_i, we can compute the expected overdraft probability and then 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. The dissatisfaction sensitivity parameter Υ_i can be identified by the data pattern where consumers' account closure probability varies with the ratio of the overdraft fee to the overdraft transaction amount, as shown in our exploratory data analysis (Section 3.2.3). Last, the mean value of the outside option α_i can be identified by the average voluntary account closing probability, and κ_i can be identified by the variation of the forced closure probability with respect to the variation of overdraft fee payment.

Note that aside from these structural parameters, another set of parameters governs the transition process. These parameters can be identified prior to the structural estimation from the observed state variables in our data. The set includes $\{\phi_i, G_i, X_i, p_i\}$.

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

5.2. Likelihood

The full likelihood function is

$$\begin{aligned} Likelihood = {} & L(\{\{C_{it}, Q_{it}, W_{it} | \widehat{S_{it}}\}_{t=1}^T\}_{i=1}^I) \\ & L(\{f\{\widehat{S_{it}} | \widehat{S_{it-1}}, C_{it-1}, Q_{it-1}, W_{it-1}\}_{t=1}^T\}_{i=1}^I) \\ & \cdot L(\{\widehat{S_{i0}}\}_{i=1}^I), \end{aligned}$$

where $\widehat{S_{it}} = \{B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, OP_{it}\}$. Because the likelihood for the optimal choice and that for the state transition process are additively separable when we apply a log transformation to the likelihood function, we can first estimate the state transition process from the data and then maximize the likelihood for the optimal choice. The likelihood function for the optimal choice is

$$\begin{aligned} L(\{\{C_{it}, Q_{it}, W_{it} | \widehat{S_{it}}\}_{t=1}^T\}_{i=1}^I) &= \prod_{i=1}^I \prod_{t=1}^T L\{C_{it}, Q_{it}, W_{it} | \widehat{S_{it}}\} \\ &= \prod_{i=1}^I \prod_{t=1}^T f\{C_{it} | Q_{it}, W_{it} \widehat{S_{it}}\} \Pr\{Q_{it} | W_{it}, \widehat{S_{it}}\} \Pr\{W_{it} | \widehat{S_{it}}\}, \end{aligned}$$

where $f\{C_{it} | Q_{it}, W_{it}, \widehat{S_{it}}\}$ is estimated from the normal kernel density estimator (explained in Online Appendix A3), and $\Pr\{Q_{it} | W_{it}, \widehat{S_{it}}\}$ and $\Pr\{W_{it} | \widehat{S_{it}}\}$ can be written as

$$\begin{aligned} \Pr(Q_{it} | W_{it} = 0, \widehat{S_{it}}) \\ = EV_{Q_{it}}(B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, OP_{it}),^{38} \end{aligned}$$

$$\Pr(W_{it} | \widehat{S_{it}}) = EV_{W_{it}}(B_{it}, \psi_{it}, Y_{it}, DL_{it}, OD_{it}, \Gamma_{it}, \Xi_{it}, OP_{it}),$$

where $EV_{Q_{it}}$ and $EV_{W_{it}}$ denote the partial derivatives of EV with respect to $U_t(C^*_{it}, Q_{it}, W^*_{it}, S_{it})$ and $U_t(C^*_{it}, Q^*_{it}, W_{it}, S_{it})$, respectively.[39]

5.3. Initial Conditions

For each consumer i, we simulate the model for 60 initial periods to derive the initial state variables. Then, we proceed to construct the likelihood increment for consumer i.

5.4. Estimation Using the Imai et al. (2009) Algorithm

We aim to estimate our infinite horizon dynamic structural model on a large data set, and we want to obtain individual responses so that we can recommend targeted marketing strategies. We investigate several 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 by Imai et al. (2009) (hereafter, IJC). We adopt the IJC method for the following reasons. First, the hierarchical Bayes framework fits our goal of obtaining heterogeneous parameters. Second, we apply the model to a large data set, so the estimation is computationally challenging. Fortunately, Bayesian Markov chain Monte Carlo (MCMC) can be combined with a parallel computing technique to reduce the computational burden. Third, the IJC algorithm is the state-of-the-art Bayesian estimation algorithm for infinite horizon dynamic programming models. The IJC algorithm provides two additional benefits in tackling the computational challenges. One is that it alleviates the computational burden by evaluating the value function only once in each epoch. Essentially, the algorithm solves the value function and estimates the structural parameters simultaneously. Thus, the computational burden of a dynamic problem is reduced by an order of magnitude with computational costs similar to 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.

Given the massive size of our data set, a traditional MCMC estimation may take a prohibitively long time, because most methods must perform $O(N)$ operations for N data points. 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 et al. (2014) (hereafter, NWX) to address this problem. The key idea of this algorithm is that the data can be distributed into multiple machines, and the IJC estimation can be performed in parallel. Once we obtain the posterior Markov Chains from each machine, we can algorithmically combine these individual chains to obtain the posterior chain of the whole sample.

5.4.1. Modified IJC Algorithm.

Our model involves a continuous choice variable, spending. Therefore, we modify the IJC algorithm[40] to obtain the choice probability through kernel density estimation. We provide a sketch of our estimation procedure and refer the reader to Online Appendix A3 for more details. The whole parameter space is divided into two sets ($\Omega = \{\Omega_1, \Omega_2\}$), where the first one contains the set of hyperparameters ($\Omega_1 = \{\mu_\lambda, \mu_\varsigma, \mu_\xi, \mu_\rho, \mu_\Upsilon, \mu_\alpha, \sigma_\lambda, \sigma_\varsigma, \sigma_\xi, \sigma_\rho, \sigma_\Upsilon, \sigma_\alpha\}$), and the second set contains the set of heterogeneous parameters ($\Omega_2 = \{\lambda_i, \varsigma_i, \xi_i, \rho_i, \Upsilon_i, \alpha_i\}_{i=1}^I$). We allow all the heterogeneous parameters (represented by ϑ_i) to follow a normal distribution with mean μ_ϑ and standard deviation σ_ϑ for the parameters. Let the observed choices be $O^d = \{O_i^d\}_{i=1}^I = \{C_i^d, Q_i^d, W_i^d\}$, where $C_i^d \equiv \{C_{it}^d, \forall\, t\}$, $Q_i^d \equiv \{Q_{it}^d, \forall\, t\}$, and $W_i^d \equiv \{W_{it}^d, \forall\, t\}$.

Each MCMC iteration consists of two blocks:

1. Draw Ω_1^r; that is, draw $\mu_\vartheta^r \sim f_{\mu_\vartheta}(\vartheta|\sigma_\vartheta^{r-1}, \Omega_2^{r-1})$ and $\sigma_\vartheta^r \sim f_{\sigma_\vartheta}(\sigma_\vartheta|\mu_\vartheta^r, \Omega_2^{r-1})(\vartheta \in \{\lambda, \varsigma, \xi, \rho, \Upsilon, \alpha\}$, the parameters that capture the distribution of ϑ for the population), where f_{μ_ϑ} and f_{σ_ϑ} are the conditional posterior distributions.
2. Draw Ω_2^r; that is, draw individual parameters $\vartheta_i \sim f_i(\vartheta_i|O_i^d, \Omega_1^r)$ by the Metropolis–Hastings (MH) algorithm.

5.4.2. Parallel Computing Following Neiswanger et al. (2014).

We adopt the parallel computing algorithm by Neiswanger et al. (2014) to estimate our model with data for more than 500,000 consumers. The logic behind this algorithm is that the full likelihood function is the product of the individual likelihoods:

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

Therefore, we can partition the data onto multiple machines and then perform MCMC sampling on each machine by 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.

Our procedure is outlined below (for additional details, see Online Appendix A4):

1. Partition data x^N into M subsets $\{x^{n_1}, \ldots, x^{n_M}\}$.
2. For $m = 1, \ldots, M$ (in parallel), sample from the subposterior p_m, where $p_m(\vartheta|x^{n_m}) \propto p(\vartheta)^{\frac{1}{M}} p(x^{n_m}|\vartheta)$.
3. Combine the subposterior samples to produce samples from an estimate of the subposterior density product $p_1 \ldots p_M$, which is proportional to the full-data posterior, that is, $p_1 \ldots p_M(\vartheta) \propto p(\vartheta|x^N)$.

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

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

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

$$\begin{aligned}&\widehat{p_1\cdots p_m}(\vartheta)\\ &= \widehat{p_1}\cdots\widehat{p_m}(\vartheta)\\ &= \frac{1}{T^M}\sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}\prod_{m=1}^{M}N(\vartheta|\vartheta_{t_m}^{m}, h^2 I_d)\\ &\propto \sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}N\left(\vartheta|\overline{\vartheta_{t\cdot}}, \frac{h^2}{M}I_d\right)\prod_{m=1}^{M}N(\vartheta_{t_m}^{m}|\overline{\vartheta_{t\cdot}}, h^2 I_d)\\ &= \sum_{t_1=1}^{T}\cdots\sum_{t_M=1}^{T}w_t * N\left(\vartheta|\overline{\vartheta_{t\cdot}}, \frac{h^2}{M}I_d\right).\end{aligned}$$

This estimate is the probability density function of a mixture of T^M 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 $\{\vartheta_{t_1}^{1}, \ldots, \vartheta_{t_M}^{M}\}$ (each from one machine) associated with a given mixture component, and we let

$$w_{t\cdot} = \prod_{m=1}^{M}N(\vartheta_{t_m}^{m}|\overline{\vartheta_{t\cdot}}, h^2 I_d),$$

$$\overline{\vartheta_{t\cdot}} = \frac{1}{M}\sum_{m=1}^{M}\vartheta_{t_m}^{m}.$$

Given the hierarchical Bayes framework, after the posterior distribution of the population parameter is obtained, we use the MH algorithm once more to obtain the individual parameters (see the details in Online Appendix A3, Step 4).

6. Results

6.1. Model Comparison

We compare our model against four other benchmark models to investigate the contribution of each element of the structural model. Models A, B, and C are special cases of our proposed model without forward looking, inattention, and unobserved heterogeneity, respectively. Model D is our proposed model. Table 9 shows the log-marginal density (Kass and Raftery 1995) and the hit rate for incidents of overdrafting, balance checking, and account closing. (We only compare these events because nonincidents are so prevalent. The hit rates for nonincidents are given in Online Appendix A5.) All four measures show that our proposed model significantly outperforms the benchmark models. Notably, inattention contributes the most to the model fit, which is consistent with our conjecture in Section 3.2.

6.2. Computational Gains from the Parallel IJC Method

We report the computational performance of different estimation methods in Table 10.

All the experiments were conducted on a server with a quad-core Intel E7-8860v3 processor (64 cores) and 1 TB RAM. The first column shows the performance of our proposed method, the IJC method with the parallel computing technique developed by Neiswanger et al. (2014). We compare this method with the traditional parallel computing method M-IJC (Murray 2010), the original IJC method, the conditional choice probability method by Arcidiacono and Miller (2011),[41] and the full information maximum likelihood (FIML) method by Rust (1987) (or nested fixed point algorithm).[42] As the sample size increases, the comparative advantage of our proposed method is more notable. Running the model on the full data set with more than 500,000 accounts takes approximately 1.5 hours, whereas running the original IJC method takes nine days.[43] Our method takes less time because it takes advantage of multiple cores that run in parallel, whereas the other algorithms have not been designed to run in parallel and use only one core. The NWX-IJC method is also nearly four times faster than the traditional parallel computing method (Murray 2010) because the traditional parallel computing method is running on the entire data set. Once all the chains converge, we can pick samples from all the chains instead

Table 10. Estimation Time Comparison

Size\method (seconds)	NWX-IJC	M-IJC	IJC	Parallel FIML	FIML
1,000	569	1,150	1,602	665	5,911
10,000	3,317	10,105	13,086	4,905	59,662
100,000	4,106	15,971	161,648	17,339	719,181
>500,000	5,344	20,720	835,774	33,293	3,583,565
	1.5 hours	5.8 hours	10 days	9.2 hours	41 days

Table 11. Monte Carlo Results When N = 100,000

Variable	True value		NWX-IJC	M-IJC	IJC	FIML
μ_β	0.9	Mean	0.838	0.839	0.846	0.929
		Standard deviation	0.042	0.038	0.038	0.028
μ_ς	0.5	Mean	0.513	0.506	0.509	0.498
		Standard deviation	0.135	0.128	0.129	0.106
μ_ξ	5	Mean	4.809	4.903	5.057	5.002
		Standard deviation	0.581	0.179	0.044	0.049
μ_ρ	5	Mean	5.212	5.141	5.094	5.075
		Standard deviation	0.054	0.047	0.048	0.025
μ_Υ	8	Mean	8.179	7.953	8.152	8.005
		Standard deviation	0.075	0.070	0.069	0.013
μ_α	20	Mean	19.522	19.563	19.765	20.788
		Standard deviation	0.275	0.227	0.211	0.104
μ_κ	0.01	Mean	0.007	0.006	0.008	0.009
		Standard deviation	0.003	0.004	0.003	0.002
σ_β	0.1	Mean	0.110	0.106	0.097	0.110
		Standard deviation	0.016	0.016	0.017	0.013
σ_ς	0.2	Mean	0.217	0.217	0.211	0.218
		Standard deviation	0.033	0.033	0.029	0.026
σ_ξ	0.1	Mean	0.114	0.086	0.087	0.087
		Standard deviation	0.058	0.053	0.033	0.028
σ_ρ	1	Mean	1.112	1.084	1.099	1.081
		Standard deviation	0.028	0.027	0.025	0.016
σ_Υ	1	Mean	1.255	1.190	1.166	1.117
		Standard deviation	0.066	0.057	0.057	0.045
σ_α	2	Mean	1.798	2.157	2.186	1.950
		Standard deviation	0.079	0.077	0.070	0.040
σ_κ	0.005	Mean	0.0056	0.0053	0.0053	0.0052
		Standard deviation	0.0005	0.0005	0.0004	0.0004

of a single chain to speed up sampling. However, the speedup occurs only after convergence. The time of burn-in cannot be reduced. Instead, the NWX parallel computing method runs only on a subset of the data on each machine. So, the burn-in time can also be saved (in addition to the speedup after convergence). The NWX-IJC method is almost 600 times faster than the FIML method. This is because the full solution FIML method solves the dynamic programming problem at each candidate value for the parameter estimates, whereas this IJC estimator only evaluates the value function once for each iteration.

We further run a simulation study to determine whether the various methods are able to accurately estimate all parameters. Table 11 shows that the different methods produce quite similar estimates and that all the mean parameter estimates are within two standard

Table 12. Structural Model Estimation Results

Variable	Interpretation	Mean (μ_ϑ)	Standard deviation (σ_ϑ)
β_i	Discount factor	0.9997 (0.00005)	0.365 (0.063)
ς_i	Standard deviation of relative risk aversion	0.258 (0.017)	0.029 (0.006)
ξ_i	Monitoring cost	4.613 (0.085)	0.260 (0.045)
ρ_i	Inattention dynamics—lapsed time	7.861 (0.335)	0.650 (0.106)
Υ_i	Dissatisfaction sensitivity	5.626 (1.362)	1.291 (0.120)
α_i	Mean value of outside option	18.793 (1.734)	2.408 (0.829)
κ_i	Forced closure probability	0.012 (0.005)	0.023 (0.009)

Notes. As a robustness check, we also ran the model using only the sample of consumers without a credit card or savings account. Please find the result in Online Appendix A10. Standard errors of the estimates are in parentheses.

errors of the true values. The parallel IJC method is slightly less accurate than the original IJC method.

6.3. Parameter Estimates

Table 12 presents the results of the structural model.

We find that the daily discount factor is approximately 0.9997. This is equivalent to a yearly discount factor of 0.89, which is largely consistent with the literature (Hartmann and Nair 2010, Fang and Wang 2015). The standard deviation of the discount factor is 0.363. This suggests that some consumers have quite low discount factors—consistent with our heavy discounting hypothesis. The monitoring cost is estimated to be 4.605. Using the coefficient of risk aversion, we can evaluate the monitoring cost in monetary terms to be \$2. In other words, consumers behave as if checking their balance costs them \$2. (We can also obtain the cost measure for each individual consumer.)

The variance of the balance perception error increases with the time elapsed since the last balance check and with the mean balance level. Notably, the variance in 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 $7.860 \times 9 = 70.74$. This suggests that the balance perception error has a diffuse distribution.

The estimated dissatisfaction sensitivity parameters confirm our hypothesis that consumers can be strongly affected by overdraft fees and close their accounts because of dissatisfaction. If we consider an average overdraft transaction amount of \$33, then the relative magnitude of the effect of dissatisfaction is comparable to \$171. This suggests that unless the bank would like to offer \$171 in compensation to consumers, dissatisfied consumers will close their current accounts and choose the outside option. Moreover, consistent with our exploratory data analysis (Figure 5), the dissatisfaction sensitivity is stronger for light overdrafters (whose average is 5.911) than for heavy overdrafters (whose average is 3.387). Keeping the average overdraft transaction amount fixed, a 1% increase in the overdraft fee can increase the closing probability by 0.12%.

6.4. Smaller Sample Sizes

This large data set used in the analysis is necessary for several reasons. First, compared with numerous other types of transactions, overdrafts are relatively rare events. Without a large amount of data, we cannot detect these rare but detrimental events, let alone understand and predict their diverse causes. Second, as shown in Section 3 of this paper, we find that consumers exhibit substantial heterogeneity in their spending behavior, cause of overdraft, monitoring cost, and dissatisfaction sensitivity. Because consumer heterogeneity is high-dimensional, the big data allow us to capture this rich consumer heterogeneity in a much more refined fashion. Third, a selected subset might suffer from sampling error or sample bias. Because consumer behaviors and characteristics are high-dimensional, it is difficult to collect an accurate, random and representative sample. To illustrate the potential sampling error, we present the estimated parameters for randomly selected samples of 1,000 (Table 13) and 50,000 (Table 14) consumers.

The comparison indicates that without the full sample, the majority of the coefficients will become statistically insignificant, and heterogeneity will been underestimated by to very large extent.

7. Counterfactual Studies

7.1. Alternative Pricing Strategies

Our structural model allows us to examine counterfactual studies that consider the effect of changing the pricing structure on consumers' spending patterns and, more importantly, their overdrafting behavior. We

Table 13. Structural Model Estimation Results When N = 1,000

Variable	Interpretation	Mean (μ_ϑ)	Standard deviation (σ_ϑ)
β_i	Discount factor	0.9991 (0.028)	0.158 (0.571)
ς_i	Standard deviation of relative risk aversion	0.170 (0.258)	0.029 (0.166)
ξ_i	Monitoring cost	2.335 (1.973)	0.268 (0.906)
ρ_i	Inattention dynamics—lapsed time	3.621 (4.845)	0.470 (2.000)
Υ_i	Dissatisfaction sensitivity	3.808 (17.991)	1.163 (1.894)
α_i	Mean value of outside option	9.833 (21.882)	2.009 (9.123)
κ_i	Forced closure probability	0.017 (0.089)	0.022 (0.379)

Note. Standard errors of the estimates are in parentheses.

Table 14. Structural Model Estimation Results When N = 50,000

Variable	Interpretation	Mean (μ_ϑ)	Standard deviation (σ_ϑ)
β_i	Discount factor	0.9996 (0.0003)	0.296 (0.202)
ς_i	Standard deviation of relative risk aversion	0.205 (0.043)	0.033 (0.021)
ξ_i	Monitoring cost	3.887 (0.298)	0.248 (0.123)
ρ_i	Inattention dynamics—lapsed time	6.351 (1.354)	0.473 (0.290)
Υ_i	Dissatisfaction sensitivity	5.893 (4.092)	1.096 (0.377)
α_i	Mean value of outside option	12.566 (6.034)	2.821 (2.665)
κ_i	Forced closure probability	0.009 (0.019)	0.027 (0.061)

Note. Standard errors of the estimates are in parentheses.

test the effect of three alternative pricing schemes, namely, a reduced per-item flat fee, a percentage fee, and a quantity premium. We provide the results in Table 15. We make two assumptions for all these simulations. One is fungibility, that is, a consumer's reaction depends only on the fee amount rather than the fee structure. If two different fee structures result in the same fee amount, then consumers should respond in the same fashion. The other is the universal dissatisfaction effect. We assume that consumers' dissatisfaction effect is proportional to the ratio of the overdraft fee to the transaction amount, despite the fee structure. It is possible that consumers will be less dissatisfied with alternative pricing schemes, such as a percentage fee, because they may be perceived as more fair (Bolton et al. 2003). However, we argue that rational consumers discover the commonality between different pricing strategies and react to only the essential factor (implicit price) that triggers dissatisfaction.

In the first scenario, we keep the per-item flat fee scheme but reduce it to $29.27 per item. Consistent with the law of demand, there is a negative relationship between the per-item overdraft fee and the overdraft frequency. Our choice of $29.27 is the solution of the optimal per-item fee to maximize the sum of the expected profit.

Expected profit is revenue minus cost. Revenue is the sum of revenue across our entire sample for a one-year period, and it includes overdraft fees and consumers' lifetime value.[44] For convenience, we define this sum as the total revenue. Because we aggregate data to the daily level, we calculate the average transaction amount for each item, which is $44, and we use it to derive the total overdraft fee amount. For example, if a consumer overspent by $170, then the consumer would incur four overdraft item fees. Cost is defined as the default cost, which is the negative balances owed by consumers who defaulted. We ignore other costs, like the labor costs to handle consumer complaints, costs for the customer service team to waive fees for some transactions, and so forth, because they are relatively small compared with the default cost and hard to measure. The profit calculation is a nested algorithm that searches for the per-item overdraft fee in the outer loop, and that solves the consumer's best response in terms of optimal spending, balance checking, and account closing, given the fee size in the inner loop. We found that the optimal per-item overdraft fee is $29.27, which would increase the bank's revenue by 2.35%. This suggests that the current overdraft fee is too high, because the bank fails to take into account consumers' negative reactions to overdraft fees, which results in a huge loss in consumers' lifetime value. In other words, at the original per-transaction price, demand is in the elastic region. Reducing it could increase the total revenue. Notice that for nonoverdrafters, the account closing frequency increases because some of them start paying overdraft fees and, hence, become dissatisfied. Therefore, the interest revenue decreases under the reduced flat fee.

In the second scenario, the per-item flat fee is converted into a percentage fee of 15.8% (solved 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 to the total transaction amount that caused the fee in the data. Again, this suggests that the bank's current overdraft fees might be too high. Intuitively, the percentage structure should encourage consumers to overdraw on transactions with small amounts but deter them from overdrawing on transactions with large amounts. Because there are more transactions with small amounts than transactions with large amounts, the total revenue generated soars by 3.11%. 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. At the segment level, the

Table 15. Welfare Analysis Under Alternative Pricing

Segment	Behavior	Current ($31)	Reduced flat ($29.27)	Percentage (15.80%)	Quantity premium (8.5% × I (OD ≤ 10) + $31 × I (OD > 10))	Fixed bills ($31)	Waive high balance (0% × I (OD ≤ 1) + $31 × I (OD > 1))	Waive all (0% × I (OD ≤ 1) + $31 × I (OD > 1))
Nonoverdrafter	Overdraft frequency	—	705	35,231	56,350	0	0	0
	Overdraft revenue	—	19,020	358,830	308,795	0	0	0
	ΔOverdraft frequency		0.09%	4.71%	7.52%	0.00%	0.00%	0.00%
	ΔOverdraft revenue		0.09%	1.62%	1.39%	0.00%	0.00%	0.00%
	Interest	6,093,611	6,092,275	6,090,978	6,091,632	8,297,670	6,520,164	6,520,164
	ΔInterest		−0.02%	−0.04%	−0.03%	36.17%	7.00%	7.00%
	Total revenue	6,093,611	6,111,295	6,449,808	6,400,427	8,297,670	6,520,164	6,520,164
	Default cost	0	3.60	179.68	287.39	0.00	0.00	0.00
	Profit	6,093,611	6,111,291	6,449,629	6,400,140	8,297,670	6,520,164	6,520,164
	ΔProfit		0.29%	5.84%	5.03%	36.17%	7.00%	7.00%
	Balance-checking frequency	62.3	60.3	55.3	51.1	19.3	53.7	53.7
	Voluntary closure frequency	7.87%	7.83%	7.85%	7.81%	3.86%	7.02%	7.02%
	Forced closure frequency	0.00%	0.01%	0.01%	0.01%	0.00%	0.00%	0.00%
Light overdrafter	Overdraft frequency	181,113	195,668	575,163	1,069,159	131,307	167,530	143,804
	Overdraft revenue	5,316,813	5,440,829	6,113,207	6,253,353	4,070,515	4,918,052	4,221,550
	ΔOverdraft frequency		8.04%	217.57%	490.33%	−27.50%	−7.50%	−20.60%
	ΔOverdraft revenue		2.33%	14.98%	17.61%	−23.44%	−7.50%	−20.60%
	Interest	153,601	154,097	155,888	159,431	158,084	197,684	203,675
	ΔInterest		0.32%	1.16%	2.27%	2.92%	28.70%	32.60%
	Total revenue	5,470,413	5,594,926	6,269,095	6,412,784	4,228,599	5,115,737	4,425,224
	Default cost	387,265	387,487	592,550	540,890	233,457	331,354	244,146
	Profit	5,083,148	5,207,439	5,676,545	5,871,894	3,995,142	4,784,382	4,181,078
	ΔProfit		2.45%	11.67%	15.52%	−21.40%	−5.88%	−17.75%
	Balance-checking frequency	56.2	51.5	51.3	50.7	40.4	52.0	50.3
	Voluntary closure frequency	5.30%	4.99%	3.90%	1.72%	4.37%	2.80%	2.54%
	Forced closure frequency	52.58%	48.70%	25.33%	12.44%	43.72%	48.64%	41.75%
Heavy overdrafter[a]	Overdraft frequency	568,439	620,116	314,149	579,824	344,474	566,165	527,511
	Overdraft revenue	16,917,234	17,442,921	16,643,035	17,255,576	10,445,900	16,849,565	15,699,193
	ΔOverdraft frequency		9.09%	−44.73%	2.00%	−39.40%	−0.40%	−7.20%
	ΔOverdraft revenue		3.11%	−1.62%	2.00%	−38.25%	−0.40%	−7.20%
	Interest	24,213	24,233	24,094	24,219	24,274	28,700	32,114
	ΔInterest		0.08%	−0.49%	0.02%	0.25%	18.53%	32.63%
	Total revenue	16,941,448	17,467,154	16,667,129	17,279,795	10,470,173	16,878,265	15,731,308
	Default cost	5,987,641	6,064,481	3,479,250	6,136,567	2,191,227	5,892,546	5,487,669
	Profit	10,953,807	11,402,674	13,187,879	11,143,228	8,278,946	10,985,719	10,243,638
	ΔProfit		4.10%	15.66%	−15.50%	−24.64%	−1.41%	23.73%
	Balance-checking frequency	58.4	58.4	58.3	58.4	31.6	57.1	52.4
	Voluntary closure frequency	4.00%	3.92%	4.46%	4.04%	2.18%	3.96%	3.76%
	Forced closure frequency	86.34%	80.16%	90.78%	86.75%	52.14%	85.31%	85.27%

Table 15. (Continued)

Segment	Behavior	Current ($31)	Reduced flat ($29.27)	Percentage (15.80%)	Quantity premium (8.5% × I (OD ≤ 10) + $31 × I (OD > 10))	Fixed bills ($31)	Waive high balance (0% × I (OD ≤ 1) + $31 × I (OD > 1))	Waive all (0% × I (OD ≤ 1) + $31 × I (OD > 1))
Total	Overdraft frequency	749,551	816,489	924,543	1,705,333	475,781	733,695	671,315
	Overdraft revenue	22,234,047	22,902,771	23,115,073	23,817,725	14,516,414	21,767,617	19,920,743
	ΔOverdraft freq		8.93%	23.35%	127.51%	−36.52%	−2.12%	−10.44%
	ΔOverdraft revenue		3.01%	3.96%	7.12%	−34.71%	−2.10%	−10.40%
	Interest	6,271,426	6,270,605	6,270,960	6,275,282	8,480,028	6,746,548	6,755,953
	ΔInterest		−0.01%	−0.01%	0.06%	35.22%	7.58%	7.73%
	Total revenue	28,505,473	29,173,376	29,386,033	30,093,007	22,996,442	28,514,165	26,676,696
	Default cost	6,374,906	6,451,971	4,071,980	6,677,745	2,424,684	6,223,900	5,731,815
	Profit	22,130,566	22,721,405	25,314,053	23,415,261	20,571,759	22,290,265	20,944,880
	ΔProfit $		590,839	3,183,487	1,284,696	−1,558,807	159,699	−1,185,685
	ΔProfit %		2.67%	14.39%	5.81%	−7.04%	0.72%	−5.36%
	Δ Consumer surplus $	5,050,668,331	1,166,431	1,337,297	515,681	3,055,197	1,532,084	1,937,923
	Δ Consumer surplus %	5,072,798,896	0.02%	0.03%	0.01%	0.06%	0.03%	0.04%
	Δ Total surplus		0.035%	0.089%	0.035%	0.029%	0.033%	0.015%
	ΔBalance-checking frequency		−3.65%	−10.55%	−16.42%	−63.69%	−12.68%	−13.31%
	ΔVoluntary closure frequency	7.41%	−0.99%	−2.24%	−6.41%	48.03%	−13.69%	14.22%
	ΔForced closure frequency	9.61%	−7.22%	−31.89%	−49.49%	−24.78%	−5.30%	−13.86%

[a]The membership of nonoverdrafters (who do not overdraw during the 15-month sample period), light overdrafters (whose overdraft frequencies are not in the top 20 percentile), and heavy overdrafters (whose overdraft frequencies are in the top 20 percentile of all overdrafters) are based on the observed data instead of the counterfactuals.

overdraft frequency for nonoverdrafters and light overdrafters increases, whereas the overdraft frequency for heavy overdrafters decreases. This is because nonoverdrafters and light overdrafters are mainly overdrawing with transactions of small amounts, which benefit from a lower cost under the percentage fee. In contrast, heavy overdrafters suffer from a higher fee because they overdraw primarily with large transaction amounts.

In the last scenario, a quantity premium structure is employed. Specifically, when a consumer overdraws fewer than 10 times, the consumer pays an 8.5% percentage fee, but once 10 overdrafts occur, each overdraft incurs a flat fee of $31. This quantity premium increases the bank's revenue by 5.59%. The motivation for this fee structure is to charge a quantity premium after second-degree price discrimination in which we segment light and heavy overdrafters. The bank would earn more in overdraft fees from the heavy overdrafters, who are willing to pay for the flat fee, but retain the lifetime value for the light overdrafters, who prefer the percentage fee (because of their high dissatisfaction sensitivity). Different from the percentage fee, under the quantity premium, the overdraft revenue from heavy overdrafters also increases because the fee is capped for them.

Interestingly, across the three strategies, the balance-checking frequency of heavy overdrafters remains largely unchanged. This is because they overdraw primarily because of heavy discounting rather than inattention due to their low monitoring cost. For nonoverdrafters and light overdrafters, the three new strategies all decrease their balance-checking frequency, because when their fees become lower, there is less incentive for them to check their balances. Furthermore, the reduced flat fee and percentage fee both lead to a drop of interest revenue, but the quantity premium leads to a rise in interest revenue because the light overdrafters are not likely to close their accounts under this pricing strategy. As to welfare effects, both the consumer surplus and the social surplus increase under all three strategies. The increase in consumer welfare comes from reduced fees as well as lower monitoring costs incurred by consumers.[45]

7.2. Fixed Bills

In our model, the income uncertainty primarily comes from the volatility of bills because their varied arrival dates and amounts. In fact, consumers can avoid these unnecessary uncertainties by fixing their bill arrival dates and amounts.[46] In this counterfactual, we test how welfare is changes when all consumers have a fixed bill arrival date and the same amount in each pay period. The exact arrival date and bill amount is specific for each consumer based on the observed data. As you can see from Table 15, when we fix the bills, the overdraft frequency for all overdrafters decreases, as does their balance-checking frequency and account closure frequency. All of these measures boost consumer surplus but reduce the bank's overdraft revenue. Total surplus rises because the gain to consumers outweighs the loss to the bank. From a public policy perspective, the government might enforce a policy to fix the bill arrival date and/or amount to help consumers with the overdraft problem.

7.3. Waiver

In practice, some banks waive overdraft fees when consumers complain about them. The welfare impact of this strategy is not clear. On one hand, the waived overdraft fee is a direct cost for the bank. But on the other hand, consumer attrition might be lower and the bank could earn more customer lifetime value. To understand this trade-off, we test two counterfactual scenarios: where the bank waives the overdraft fee once a year for consumers with high balances (balances above the median) and for all its consumers. When the waiver is offered to only the high-worth consumers, we find that retention and profitability both increase, as does consumer welfare. The latter is because the waiver reduces the monitoring costs of high-balance consumers, as they no longer have to check their balances very frequently. This increases high-net-worth consumers' utility in remaining with the bank; hence, retention and profitability rise. In contrast, when the bank grants the waiver to all consumers instead of only the high-balance ones, the bank's profit declines owing to lower overdraft revenue. Nevertheless, the social surplus is higher with the waiver than without it. Therefore, although the bank might have an incentive to implement the waiver for high balance customers, from a policy perspective, the government might also consider implementing the waiver for all consumers.

8. Contributions and Limitations

The $32 billion in overdraft fees assessed by banks in 2012 has increased consumer attrition and drawn regulators' attention to this issue. However, there is little quantitative research on consumers' financial decision-making processes that explains their overdrafting behavior. The lack of well-calibrated models prevents financial institutions from designing appropriate pricing strategies and improving their financial products. With the aid of big data that capture consumers' spending patterns, banks can improve overdraft fees.

In this paper, we build a dynamic structural model of consumer daily spending that incorporates inattention to rationalize consumers' overdrafting behavior. We quantify the discount factor, monitoring cost, and dissatisfaction sensitivity for each consumer and use these variables to design new strategies. In comparing the current pricing scheme with several alternative pricing strategies, we find that a percentage fee structure can increase the studied bank's revenue through market expansion and that the quantity premium structure can increase the bank's revenue because of second-degree price discrimination. New fee structures can also improve consumer welfare because of reduced monitoring costs.

We calibrated our model at an individual level on a sample of more than 500,000 accounts. The large sample is necessary because of overdraft being a rare event and the rich unobserved consumer heterogeneity. Besides, the sampling error can cause detrimental effect. To illustrate, we randomly select 10% of the sample and redo the analysis. The predicted revenue in the counterfactual of the quantity premium strategy is 6%, or $0.3 million less than the value obtained from results using the entire data set. In other words, the sampling error might lead to an incorrect strategy and a significant loss to the bank. Therefore, we argue that, given the sizable sampling error, if the computational burden of estimating a dynamic structural model on a large data set is minimal, as demonstrated by our parallel IJC algorithm, using the full data set is preferred.

To estimate a complicated structural model with big data, we adopt parallel computing techniques in combination with the Bayesian estimation algorithm developed by Imai et al. (2009). This new method significantly reduces the computation burden, and it could be used by other researchers and marketers who would like to use structural models to solve real-world large-scale problems. Although we apply it to the overdraft context, the model framework can be generalized to analyze other marketing problems in which consumers have similar dynamic budget allocation decisions (e.g., utility accounts and mobile phones).

Several limitations of the current study call for future work. First, we do not observe consumers' existing alert settings. Some consumers may have already received alerts to help them make financial decisions. This might cause an overestimation of the monitoring cost and an underestimation of the inattention sensitivity to lapsed time. But given the prevalence of overdrafts, we conjecture that many consumers are not using alerts. And because our counterfactual only alters the pricing strategies, we cannot think of any reason that our welfare analysis results can be different after accounting for alerts. Future studies might consider incorporating alert data or investigating counterfactual scenarios where alerts reduce monitoring costs. Second, we do not have data on consumers' decision to opt in for overdraft protection by ATM/POS transactions. We only know that if ATM/POS transactions caused overdrafts, then the consumer must have opted in. However, if we do not observe ATM/POS-transaction-caused overdrafts, this does not mean that the consumer has opted out. Sometimes, an opt-in consumer can be cautious enough (low monitoring cost) to avoid overdrafts from using ATM/POS transactions. Had we known the opt-in status information, we could have provided an informative prior within our Bayesian model, as consumers who opt in probably have stronger needs for short-term liquidity owing to fluctuations in the size and arrival time of their income and spending. Moreover, it is possible that consumers might forgive the bank after a period of time. Future research could examine whether consumer dissatisfaction declines over time. Finally, we model only consumers' nonpreauthorized spending from their checking accounts with one bank. In reality, consumers usually have multiple accounts, such as savings, credit cards,[47] and loans, with multiple financial institutions. A model that captures consumers' decisions across all accounts for both short-term and long-term finances would provide a more complete picture of consumers' financial constraints, management capabilities, and resources so that the bank can design more customized products.

Endnotes

[1] According to Evans et al. (2011), there are 180 million checking accounts in the United States.

[2] Examples can be found in *Consumers Union* (2013) Sick of unfair overdraft fees? Tell us about it! (June 18), http://consumersunion.org/2013/06/rep-carolyn-maloney-pursues-consumer-protections-on-overdraft-fees/ and Hellman J (2014) Overdraft fee abuse. *Aljazeera America* (January 21), http://america.aljazeera.com/watch/shows/real-money-with-alivelshi/Real-Money-Blog/2014/1/21/overdraft-fee-abuse.html.

[3] See "Consumers and Congress Tackle Big Bank Fees," http://banking-law.lawyers.com/consumer-banking/consumers-and-congress-tackle-big-bank-fees.html.

[4] See the Consumer Financial Protection Bureau study of overdraft programs at http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf.

[5] For the sake of confidentiality, we cannot disclose the exact number, but it is a representative sample from the banks' customers and is within the range of 500,000 to 1,000,000 accounts.

[6] All dollar values in the paper have been rescaled by a number between 0.85 and 1.15 to help obfuscate the exact amounts and preserve the anonymity of the customers, but this factor does not change the substantive implications. Using our rescaled values the bank sets the first-time overdraft fee for each consumer at $22, and all subsequent overdraft fees are set at $31.

[7] The maximum number of per-item overdraft fees is between three and six.

[8] The grace period, which is the number of days before the continuous overdraft fee applies is between 3 and 10.

[9] See the Electronic Fund Transfer Act (http://www.federalreserve.gov/bankinforeg/caletters/Attachment_CA_13-17_Reg_E_Examination_Procedures.pdf).

[10] See the Consumer Financial Protection Bureau study of overdraft programs at http://files.consumerfinance.gov/f/201306_cfpb_whitepaper_overdraft-practices.pdf.

[11] See endnote 7.

[12] 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 keep only one discount factor in the current model. Estimation results with hyperbolic discount factors are available upon request.

[13] We separate the analyses for heavy overdrafters, light overdrafters, and nonoverdrafters because each segment shows distinct behavioral

patterns. The segment is defined by the number of overdraft occurrences in the sample period.

[14] Another explanation of this phenomenon is that this consumer forgot the exact amount of balance. Our notion of inattention is consistent with forgetting (Mullainathan 2002, Mehta et al. 2004, Ching and Ishihara 2010).

[15] We also find that light overdrafters are more likely to check balances after overdraft fees are charged than heavy overdrafters. This also provides evidence for inattention. Furthermore, we find that light overdrafters fail to learn from past experiences. Specifically, we find a positive correlation between the time elapsed since overdraft and the time gap between two balance-checking times. This suggests that immediately after overdrafting, consumers might start monitoring their checking accounts. As time goes by, however, they become inattentive again.

[16] We also test this assumption by estimating a myopic model and a bounded forward-looking model. The results can be found in Online Appendix A6.

[17] Our focal bank practices "nightly batch processing," a process of posting transactions (credits and debits) to the account after the close of business each day, following the industry standard. However, the posting order is based on the exact transaction time (to an accuracy of each second). Preauthorized transactions, like income deposits (direct deposit) or bill payments, are always posted before nonpreauthorized transactions. Our model timing is consistent with this posting order. Moreover, our model assumes that balance checking happens before spending. In the scenario when the consumer spends before checking his or her balance (our data will show that the debit transaction occurs before balance checking), we assume that the consumer does not know the initial balance of that day.

[18] Consumption C_{it} must be nonnegative. When applied to the data, there are days when consumers just receive income (e.g., deposit money) without any consumption (spending). In this case, we set $C_{it} = 0$ but update the budget equation with the "negative spending" discussed in Section 4.5.

[19] We discuss how we differentiate bills and consumption in the data in the last paragraph of Section 4.1.

[20] This assumption is realistic in our setting because we distinguish 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 (which we explain in the next subsection) in the sense that the uncertain overdraft fee is equivalent to the uncertain balance because both of these tighten the consumer's budget constraint. As for attentive consumers who overdraw because of heavy discounting, such consumers would be fully aware of the potential costs of overdrafting. Thus, in both cases, we argue that the assumption of a known total overdraft fee is reasonable.

[21] Alternatively, we could describe this nonpreauthorized spending as immediate or discretionary spending. We avoid the term *discretionary spending* to avoid confusion with the usual economic definition. Economists traditionally use the term *discretionary* as the amount of income left after spending on necessities such as food, clothing, and housing, whereas in our problem, we are thinking about immediate spending that could have been delayed.

[22] We choose to assume that consumers are inattentive to balances instead of the overdraft fee for several reasons. First, the mean overdraft frequency is 10 times, and the median overdraft frequency is three times. This implies that most consumers have paid overdraft fees multiple times. According to the bank's policy, once a consumer overdraws, he or she will receive both an email and mail notification immediately, and the overdraft payment will be reflected in the account statement. During the sample period, the bank did not change its overdraft fee. Therefore, provided that a consumer overdrew once and paid the overdraft fee, he or she should be aware of the overdraft charge. Moreover, according to the Regulation E amendment, after 2009, consumers must be informed of the overdraft fee policy, including its amount, for the bank to obtain affirmative consent (opt in) for overdraft coverage from consumers. As our sample period falls after 2010, all consumers have been informed of the overdraft charge. Finally, the model-free evidence in Table 5 and Figure 3 suggests that balance-checking frequency significantly affects overdraft probability. Therefore, uncertainty regarding balances appears to be a major driver of overdrafts.

[23] Consumers can also be naively inattentive, but we do not allow for this here. See the discussion in Grubb (2014).

[24] Conditional on her consumption choices today, C_{it}, based on Equation (2), the consumer is still not perfectly informed about the budget constraint at the beginning of the next period, that is, B_{it+1}, because the preauthorized bills, ψ_{it+1}, is unknown.

[25] 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 payments. Specifically, a high overdraft payment for a consumer can be explained by either a positive balance perception error or a large perception error variance caused by a large ρ. Thus, we fix $\bar{\eta}$ at zero; that is, the perception error is assumed to be unbiased.

[26] We considered other specifications for the relationship between the perception error variance and the time elapsed since the last balance check. The results remained qualitatively unchanged.

[27] We also tested whether the forced closure probability is associated with the overdraft transaction amount, but we failed to find a statistically significant relationship.

[28] When estimating this linear probability model, we set the lower bound and upper bound of π_f at 0 and 1.

[29] We also consider two other models with dissatisfaction specified as the sum of past ratios $\Delta_{it} = \frac{OD_{it}}{|B_{it}-C_{it}|}$ and allow consumers to forget the ratios except the past three ones. However, both the log marginal density and the hit rate of these alternative models are worse than our proposed model.

[30] Please find a discussion of the normalization in Online Appendix A7.

[31] The correlation between the overdraft fee and the overdraft amount is actually very small (0.02), so we assume that the overdraft fee is not an increasing function of the overdraft amount but i.i.d. over time.

[32] A compound Poisson distribution is the probability distribution of the sum of a number of independent identically distributed random variables, where the number of terms to be added is itself a Poisson-distributed variable. In our model, each independent bill, for example, a mortgage loan interest or credit card payment, is a random variable. Because the total number of bills that arrive each day is Poisson-distributed, the sum of the bills becomes a compound Poisson distribution. We use G to characterize the discrete distribution of the size of each individual bill, and G is an empirical distribution. Suppose in the entire sample, that one consumer has three utility bills with amounts \$30, \$50, and \$80 as well as one cell phone bill of the amount \$30. Then, G is $G(x) = \begin{cases} 0.5 & \text{if } x = 30, \\ 0.25 & \text{if } x = 50, \\ 0.25 & \text{if } x = 80. \end{cases}$ The arrival rate parameter ϕ in the Poisson distribution is estimated using the maximum likelihood method. We estimate both G and ϕ from the data heterogeneously for each individual before estimating the structural model. They are used as inputs in the structural estimation.

[33] More explanations on our choice of the income and bill process can be found in Online Appendix A9.

[34] The transition process for the perceived balance $\widetilde{B_{it}}$ is jointly determined by Equations (2) and (4).

[35] For the sake of simplicity, we have omitted the subscript i.

[36] We also tried to fix the discount factor (at 0.9997) and estimate the coefficients of risk aversion. The posterior mean of the estimated

coefficient of relative risk aversion is 0.72. Other structural parameter estimates are not significantly affected under this specification. Our results confirm that the coefficient of risk aversion and the discount factor are mathematically substitutes (Andersen et al. 2008). Estimation results with a fixed discount factor are available upon request.

[37] We also tried other values for the coefficient of relative risk aversion θ. The estimated discount factor β values change when we use different values of θ, but other structural parameter values remain the same. The policy simulation results are also robust to the use of different values of θ.

[38] $\Pr(Q_{it}|W_{it}=1,\widehat{S_{it}})=\emptyset$.

[39] These partial derivatives do not have the logit closed-form solution because of the specification for forced closure.

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

[41] We use the finite mixture model to capture unobserved heterogeneity and apply the Expectation Maximization algorithm to solve for the unobserved heterogeneity. More details of the estimation results can be obtained upon request.

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

[43] We keep a total of 2,000 MCMC iterations and use the first 500 as burn-in. Convergence was assessed visually by using plots of the parameters. We chose a store of N=100 past pseudo–value functions. The bandwidth parameter is set to $h=0.01$.

[44] We calculate the lifetime value of a consumer by multiplying the average balance and the interest rate, accounting for the life expectancy (closing the account). This is the source of the interbank interest rate (the U.S. Board of Governors of the Federal Reserve System Effective Federal Funds Rate (FEDFUNDS), accessed March 14, 2016, https://research.stlouisfed.org/fred2/series/FEDFUNDS).

[45] A similar effect is documented by Chen and Yao (2016), where consumers incur lower search costs with refinement tools.

[46] See Forbes T (2016) Change your payment due dates. *Penny Pinchin Mom* (January 18), https://www.pennypinchinmom.com/budget-tip-change-your-payment-due-dates/.

[47] Credit card spending can be modeled as a random variable, distributed log normally with its variance depending on the average amount of the credit card bill by individual.

References

Andersen S, Harrison GW, Lau MI, Rutström EE (2008) Eliciting risk and time preferences. *Econometrica* 76(3):583–618.

Arcidiacono P, Miller RA (2011) Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79(6):1823–1867.

Arrow KJ (1963) Liquidity preference. *The Economics of Uncertainty*, Lecture Notes for Economics, Vol. 285 (Stanford University, Stanford, CA), 33–53.

Bellman R (1957) *Dynamic Programming* (Princeton University Press, Princeton, NJ).

Bolton RN (1998) A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* 17(1):45–65.

Bolton LE, Warlop L, Alba JW (2003) Consumer perceptions of price (un)fairness. *J. Consumer Res.* 29(4):474–491.

Burns R (2010) Managing credit: 3 ways overdraft fees will still haunt you. *Black Enterprise* (August 9), http://www.blackenterprise.com/money/managing-credit-3-ways-overdraft-fees-will-still-haunt-you/

Chen Y, Yao S (2016) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 63(12):4345–4365.

Chetty R, Looney A, Kroft K (2009) Salience and taxation: Theory and evidence. *Amer. Econom. Rev.* 99(4):1145–1177.

Ching A, Ishihara M (2010) The effects of detailing on prescribing decisions under quality uncertainty. *Quant. Marketing Econom.* 8(2):123–165.

Ching A, Erdem T, Keane M (2009) The price consideration model of brand choice. *J. Appl. Econom.* 24(3):393–420.

DellaVigna S (2009) Psychology and economics: Evidence from the field. *J. Econom. Literature* 47(2):315–372.

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. Working paper, University of Chicago Law School, Chicago.

Fang H, Silverman D (2009) Time-inconsistency and welfare program participation: Evidence from the NLSY. *Internat. Econom. Rev.* 50(4):1043–1077.

Fang H, Wang Y (2015) Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *Internat. Econom. Rev.* 56(2):565–596.

Fiske ST, Taylor SE (1991) *Social Cognition*, 2nd ed. (McGraw-Hill, New York).

Friend I, Blume M (1975) The demand for risky assets. *Amer. Econom. Rev.* 65(5):900–922.

Gabaix X, Laibson D (2006) Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quart. J. Econom.* 121(2):505–540.

Gopalakrishnan A, Iyengar R, Meyer RJ (2014) Consumer dynamic usage allocation and learning under multipart tariffs. *Marketing Sci.* 34(1):116–133.

Grubb MD (2014) Consumer inattention and bill-shock regulation. *Rev. Econom. Stud.* 82(1):219–257.

Hall RE (1978) Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *J. Political Econom.* 86(6):971–987.

Hartmann WR, Nair HS (2010) Retail competition and the dynamics of demand for tied goods. *Marketing Sci.* 29(2):366–386.

Hirshleifer D, Teoh HS (2003) Herd behaviour and cascading in capital markets: A review and synthesis. *Eur. 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.

Kass R, Raftery A (1995) Bayes factors. *J. Amer. Statist. Assoc.* 90(430):773–795.

Keaveney SM (1995) Customer switching behavior in service industries: An exploratory study. *J. Marketing* 59(2):71–82.

Matzler K, Würtele A, Renzl B (2006) Dimensions of price satisfaction: A study in the retail banking industry. *Internat. J. Bank Marketing* 24(4):216–231.

Mehta N, Rajiv S, Srinivasan K (2004) Role of forgetting in memory-based choice decisions: A structural model. *Quant. Marketing Econom.* 2(2):107–140.

Merton R (1992) *Continuous Time Finance* (Blackwell Publishing, Cambridge, UK).

Modigliani F, Brumberg R (1954) Utility analysis and the consumption function: An interpretation of cross-section data. Kurihara KK, ed. *Post Keynesian Economics* (Rutgers University Press, New Brunswick, NJ), 388–436.

Monroe KB (1990) *Pricing: Making Profitable Decisions* (McGraw-Hill, New York).

Murray L (2010) Distributed Markov chain Monte Carlo. Working paper, CSIRO Mathematics, Perth, Australia.

Mullainathan S (2002) A memory-based model of bounded rationality. *Quart. J. Econom.* 117(3):735–774.

Neiswanger W, Wang C, Xing E (2014) Asymptotically exact, embarrassingly parallel MCMC. *Proc. 30th Internat. Conf. Uncertainty Artificial Intelligence* (AUAI, Quebec City, Quebec, Canada).

Nwokoye NG (1975) Subjective judgments of price: The effects of price parameters on adaptation levels. *Proc. Fall Conf. Amer. Marketing Assoc*. (American Marketing Association, Chicago).

Peng L, Xiong W (2006) Investor attention, overconfidence and category learning. *J. Financial Econom.* 80(3):563–602.

Pope DG (2009) Reacting to rankings: Evidence from "America's Best Hospitals." *J. Health Econom.* 28(6):1154–1165.

Reis R (2006) Inattentive producers. *Rev. Econom. Stud*. 73(3):793–821.

Rossi PE Allenby GM, McCulloch RE (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, Hoboken, NJ).

Rust J (1987) Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55(5):999–1033.

Sims CA (1998) Stickiness. Carnegie-Rochester Conference Series on Public Policy, Vol. 49 (North Holland, Amsterdam), 317–356.

Sims CA (2003) Implications of rational inattention. *J. Monetary Econom.* 50(3):665–690.

Stango V, Zinman J (2014) Limited and varying consumer attention evidence from shocks to the salience of bank overdraft fees. *Rev. Financial Stud*. 27(4):990–1030.

Sudhir K (2016) Editorial—The exploration-exploitation tradeoff and efficiency in knowledge production. *Marketing Sci*. 35(1):1–9.

Topper S (2012) The history and evolution of the free checking account. *Free Checking Information* (March 18), http://freecheckinginformation.com/financial-writers/the-history-and-evolution-of-the-free-checking-account/

Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probable cause. *Cognitive Psych.* 5(September):207–232.

Yao S, Mela CF, Chiang J, Chen Y (2012) Determining consumers' discount rates with field studies. *J. Marketing Res*. 49(6):822–841.

Zhou Y (2012) Failure to launch in two-sided markets: A study of the US video game market. Working paper, State University of New York, Stony Brook.