Fintech Nudges: Overspending Messages and Personal Finance Management *

Sung K. Lee†

December 30, 2019

Abstract

Using large proprietary money management app data from a major commercial bank in Canada, I study how the app users manage their personal finance upon seeing an overspending message on the mobile app. First, I find that the message recipients reduced spending on the following day by C$8.15, which corresponds to 5.4% of their daily average spending, compared to the non-recipients. Second, these fintech nudges had temporary effect on flow spending and resulted in permanent reduction in cumulative spending. Third, the effects are especially pronounced for the users who are older, have higher liquid wealth, are more finance-savvy, are new to the app experience, or reside in a city with a higher fraction of educated population. Fourth, I find suggestive evidence that these effects could spill over from one app user to another in the same household. On the other hand, the message recipients were less likely to monitor their accounts via log-ins afterward, which is selective inattention known as the ostrich effect.

*I am extremely indebted to Johannes Stroebel for helpful suggestions and guidance. Also, I am very thankful to Andres Liberman, Theresa Kuchler, Anthony Saunders, and Arpit Gupta for constructive comments and discussion. Any errors are my own.

†Stern School of Business, New York University. Email: slee2@stern.nyu.edu

1 Introduction

Individuals often fail to achieve their personal financial goals (Kuchler and Pagel, 2018), and several approaches have been implemented to influence their financial decision making, ranging from direct intervention of government at one extreme (DeFusco et al., 2019) and relatively soft intervention in the form of informational nudges at another (Agarwal et al., 2015). The consequences of nudges as commitment devices have had mixed success; the general consensus is that although this type of intervention indeed alters individual behavior, the magnitude of the effect is minor (Zwane et al., 2011), and the effect dissipates over the long run (Kast et al., 2012). In a similar vein, would the use of fintech nudges via a money management app help users with their spending and personal finance management, especially given that the contents of the nudges are tailored to each individual and that the users can easily have access to their bank account information in real time?

In this paper, I study the effectiveness of one such money management app, which was once a top download one in the App Store and Google Play, in inducing changes in spending patterns. In the context that notifying app users of their overspending could work as a nudge mechanism to cause changes in household behavior, I focus on the following behavioral finance questions specifically: (i) whether technological nudges lead to a decline in spending and if so, what spending categories the users reduce spending on, (ii) how long the effects last and whether nudges are useful for long-term habit formation, (iii) which subgroup of the app users respond to nudges most significantly, (iv) whether effects of nudges imposed on an individual user spread over to another user in the same household, and (v) how differently the app users manage their personal finance after being nudged.

I conduct these empirical analyses, using large proprietary data on mobile app users of a major commercial bank in Canada (hereinafter referred to as the Bank), which has outsourced the management of the app services to a fintech company. The data provide information on daily transactions at the account level, daily app usage at the user level, and month-end financial status and demographics at the user level. Out of the user base

of 1 million, as of January 2018, the data cover 55,586 app users between June 2017 and January 2018, who downloaded the app after the new version containing the overspending message feature was released. The daily frequency of observations at the user level allows me to capture immediate changes in individuals' spending behavior after a nudge is given.

In order to overcome the endogeneity concern that app users receiving a high number of overspending messages could be inherently different from those receiving a low number of messages, I exploit the exogenous variation in the generation of messages within the app by implementing a sharp regression discontinuity design among the app users. More specifically, upon logging in to the app, users get to see an overspending message only if their spending on a day normalized by their historical daily average spend exceeds the predetermined threshold multiple, which is configured by the Bank uniformly across the users.[1] As Panel A of Figure 1 illustrates an example of overspending messages, they alert users of how much their spending on a particular date deviates from their usual spending patterns. Hence, the 219 days of discontinuities around the threshold multiple in the data allow me to estimate the local average treatment effect of the nudging feature on the subsequent spending behavior of the app users.

Using this regression discontinuity design, I find that the users adjust their spending once they become aware of their deviation from the historical spending due to seeing an overspending message. In particular, the users with an overspending message reduced spending on the next day of a log-in by C$8.15, which corresponds to 3.3% of their date average over the past 12 months and 5.4% of their rolling daily average spend over the past 365 days, relative to the users who did not receive a message.[2] Relatedly, the message recipients made an adjustment on the extensive margin by purchasing 0.06 fewer goods and on the intensive margin by spending C$2.07 less on a per-purchase basis. In terms of categorical breakdown, overspending messages are the most effective in deliv-

[1]The Bank has determined the threshold multiple to target a specific frequency of overspending messages per period, based on UX (user experience) principles such as transaction volume and variance of spending per user. Its business strategy is to win trust from customers, retain them, and then introduce new financial products.

[2]The date average refers to the average value of spending on the same calendar date over 12 observational points whereas the rolling daily average refers to the average value of spending over 365 observational points.

ering reduction of C$3.09 in the Shopping category. Therefore, the users responded by adjusting discretionary spending, and the high inter-temporal elasticity of substitution of expenditure can explain why the recipients reduced spending on Shopping as the cost of delaying or accelerating purchasing this category is relatively low.

Next, I vary the spending horizon of the outcome variables to measure the duration of impact from the nudges. There is a delayed response of reduction of C$1.64 by the message recipients relative to the non-recipients on the second day, and there is little spending gap thereafter. As a consequence, the cumulative spending gap widens to C$9.79 over the two day horizon and remains statistically significant around C$10.02 over the four day horizon. Although it is difficult to precisely estimate the spending gap over a very long horizon, overspending messages induce reduction of flow spending in the short term and permanent reduction in cumulative spending in the long term.

I then apply the same technique to different groups of app users to investigate if the effects of nudges differ across them. More to the point, I run the discontinuity regression for subgroups with different demographic characteristics, financial status, app experience, and education level using the following criteria: gender, age, liquid wealth, the fraction of checking account balance to total liquid asset balance, the number of days in use since the registration of the app, and fraction of population with a diploma or a degree in the city the users live in. I find that the effects of nudges are more pronounced for the users who are older, have higher liquid wealth, more finance-savvy, are new to the app experience, or reside in the city with a higher fraction of diploma holders among the population.

Moreover, I test on whether the effects of fintech nudges are contagious among households by running the baseline regression specification on spending of the users who themselves did not log in but whose spouse logged in. More precisely, I identify two individual users as a couple if they share at least one Bank account, are of different gender, have an age difference of at most 15 years, and live in the same city, and I make sure that only one member of a couple logs in on a particular date. Then I compare spending from the non-joint accounts of the users whose spouse has been informed of overspending to that from the users whose spouse has not received any message. I find that the spending gap

is statistically significant at C$7.10. Thus, there is suggestive evidence that changes in spending behavior can be induced and spread at the household level due to overspending messages.

Finally, I examine their behavior on personal finance management once the users read an overspending message. In other words, I again run the regression discontinuity specification using the probability of logging at least once in the future as the outcome variable; this statistical analysis tests whether the users monitor their accounts more regularly when alarmed by the app. Surprisingly, I find that relative to the message non-recipients, the message recipients are 2.3 percentage points *less* likely to log in at least once over the next 7 days and 1.6 percentage points *less* likely to log in at least once over the next 30 days. Since the record of overspending carries a negative connotation, the users selectively pay less attention to their financial status; this unexpected phenomenon exemplifies what is known as the ostrich effect.

Although the effect of a single overall spending message may seem trivial on a daily basis, the back-of-the-envelope calculation suggests that the average user would have saved about C$411 per year. This reduction in spending would have corresponded to a 10% decrease in the size of outstanding credit card debt held by an average Canadian. Likewise, on the assumption that every user of the app has manually downloaded by now the new version that has the overspending message feature, the aggregate impact would have been a yearly reduction in spending about C$400M. Indeed, despite the simple feature that notifies individuals of deviation from their past norm, these informational fintech nudges via a mobile app can function not just as an effective way of reducing spending but also as a tool to impose substantial spending response to the macro-economy.

The effectiveness of fintech nudges stems from the personalization of contents and the high frequency of intervention. Compared to one-size-fits-all generic messages, the specific contents of messages in this paper convey the degree of deviation so that individual users can pay close attention to their unusual behavior. At the same time, since messages are sent right after an unusual pattern occurs, this prompt feedback is conducive to immediate behavioral responses while this deviation is fresh in users' memory. Hence, as long as fintech nudges inform individuals about deviation from their own past behavior

in a timely manner, they can function as a mechanism to maintain self-discipline even in other dimensions of consumer behavior.

The contributions of this paper are four-fold. Most importantly, this paper is the first, to the best of my knowledge, to test the effectiveness of nudges in the space of spending among individuals who have signed up for app services to monitor their finance regularly. The dimensions of financial activity, on which previous studies test the effectiveness of reminders and messages, include savings (Karlan et al., 2016; Kast et al., 2012), debt repayment (Agarwal et al., 2015; Cadena and Schoar, 2011; Bursztyn et al., 2016; Karlan et al., 2015), avoidance of overdraft fees (Stango and Zinman, 2014), and health insurance take-up (Zwane et al., 2012). Unlike these works, which are primarily field experiments and suffer from a lack of statistical power, this paper takes advantage of the sheer size of observational data to make a precise estimate. Furthermore, despite the absence of explicit perks provided by the Bank, overspending messages can indeed serve as fintech nudges to induce reduction in spending.

Relatedly, this paper makes a contribution to the literature documenting unintended consequences associated with salient messages and individuals' tendency to neglect negative financial information. More to the point, informational nudges based on social comparison may lead to a lack of action in retirement savings (Beshears et al., 2015) or disutility in the form of high willingness to pay for energy bills (Allcott and Kessler, 2019), and inducing behavioral change in one dimension with a reminder could result in negative spillover effects in another dimension for forgetful and salient thinkers (Medina, 2017). Similarly, investors check their stock investment accounts less frequently when the market is bearish (Karlsson et al., 2009), and people monitor their bank accounts less frequently when the balance becomes negative (Olafsson and Pagel, 2017). This paper also documents that selective inattention to monitoring finance arises from reading overspending messages, and therefore, it points out that side effects of nudges should also be taken into account.

Furthermore, this paper contributes to a strand of literature on peer effects. It provides suggestive evidence that spending is another dimension of household activity that can be influenced by other peers, in addition to stock market participation (Hong et al.,

2004), strategic default on mortgages (Guiso et al., 2013), foreclosure (Gupta, Forthcoming), and product adoption (Bailey et al., 2019). In addition, peer effects can occur within households, not just across co-workers (Kalda, 2018) or Facebook friends (Bailey et al., Forthcoming). Hence, fintech nudges can be used as a peer effect mechanism to shape others' spending behavior, and their power can be amplified.

Additionally, this paper expands on the growing literature on individual financial activity using large-scale microdata of online services and personal finance management software. Papers using online datasets have enhanced an understanding of household financial decision making, ranging from effects of social interactions on housing investments (Bailey et al., 2018) to the failure of sticking to original debt paydown plans (Kuchler and Pagel, 2018). Also, studies using personal finance management data investigate spending behavior at the time of pay checks (Gelman et al., 2014), consumption responses to income fluctuations among highly indebted households (Baker, 2018), and liquidity and consumption responses before and after a borrowing event among payday borrowers (Olafsson and Pagel, 2016). Since the data in this paper are recorded at the high frequency on a daily basis, this paper examines the immediate response of the users and focuses on the direct usage of the app per se in the fintech era; thus, it further improves an understanding of the use of technology in everyday finance.

The paper's findings suggest that well-implemented technological nudges can promote changes in households' financial behavior in a way that the designer of the nudges intends. One policy implication is that exogenous attention cue in the form of push notifications should be considered even if users do not log in to the app themselves. As the effect of a single message is not hugely amplified over the long run, keeping them up-to-date on a regular basis about deviation from their own past spending patterns will induce closer monitoring of their finance. Furthermore, overspending messages should be framed in an encouraging way so that users are not deterred from using the app altogether. Even an informational message without any frame could carry a negative connotation and impose psychological discomfort, which would mitigate the effect of the intervention. Hence, using more positive language in messages could alleviate the unintended consequence of lower log-in likelihood.

The organization of the paper is as follows. In Section 2, I provide background information about the app and describe the data. In Section 3, I explain the empirical strategy to show how the users of the app have changed their spending behavior and personal finance management. In Section 4, I explain the results of the regression analyses. In Section 5, I carry out placebo tests and robustness checks. Section 6 concludes.

2 Background Information and Data

The fintech company processing the transaction and log-in data in the paper was established in 2010 and has provided money management app services to a variety of banks, and the Bank in this paper has outsourced the services to the fintech company and branded the app under the Bank's name since April 2016. The data come from this Bank, which is a major commercial one in Canada. The introduction of the app is one of the earliest technology adoptions in Canadian digital banking, and the total number of app users, as of January 2018, reached over one million.[3]

2.1 Services of the App

Users can view transactions and manage their accounts through the app. After they register their checking, savings, or credit card accounts to the app, as they swipe the cards for spending, transactions will be reflected in the app in real time. As a result, users can check how much they have spent overall in the current month and how much they have spent on each spending category. The categories are automatically set by the app, and app users can also manually categorize the transaction every time they swipe their card. The categories include Wants and Needs: Wants include Cash (ATM withdrawal), Dining Out, Shopping, Entertainment, Travel, and Fees, and Needs include Groceries, Utilities, Transportation, Education, Health, and Home. Panel B of Figure 1 illustrates how a summary of overall and categorical spending is indicated.

As users spend along, the app provides four different management functions to indi-

[3]This total number of app users represents about 20% of the Bank's mobile banking customers, 7% of the Bank's overall customers, and 2% of the entire Canadian population.

cate where the current spending stands, especially in regard to historical average: Monthly Spending, Spending by Category, Daily Digest, Spending Alerts. In Monthly Spending, users can view the amount of overall spending for the current month and see their status compared to the historical monthly average. Basically, if the current spending is above the past month's spending, the status red is indicated as a warning. If the current spending is below (around) the historical average, then the status green (yellow) is shown to indicate that the user is not overspending. In Spending by Category, users can view the amount of spending by each category for the current month and view their status for each category compared to the historical monthly average. In Daily Digest, users can view a daily summary of spending, which displays the name and amount of every purchase transaction. Spending Alerts is a notification feature in case users overspend. More to the point, they are notified when spending on a particular day is a high multiple of spending on a typical day. Panel A of Figure 1 illustrates an example of the app's alerting app users of overspending.

The focus of the paper among the app's features is Spending Alerts messages, which are generated only when the spending on a particular date is above spending on a typical date multiplied by the threshold multiple. The overall and categorical daily spending average on a date is calculated by dividing the sum of the spending on that date for each month over the last twelve months by the number of observations, where the denominator is twelve at maximum.[4] To illustrate an example, the daily overall spending average on July 30, 2017 is calculated by computing the average of spending for the 30th of each month, i.e. June 30, 2017 through July 30, 2016. Supposing that overall spending on July 30, 2017 is C$201, and the average overall spending for that date is C$100, spending on this date exceeds the threshold of 2 times the average just computed, as a result of which a Spending Alerts message would be shown on July 31, 2017. The threshold multiple for each spending category, above which a Spending Alerts message is generated, is listed in Table 1. As the Spending Alerts feature was introduced in the newer version of the app on June 19, 2017, and users can see the reports upon logging in to the app, I focus only on

[4]If the number of observations on that date is fewer than 3 over the last twelve months either because the user did not spend on that date frequently enough or because the spending history is too short, no Spending Alerts message will be reported. Hence, I drop the observations in this case.

the users who have signed up for the app after this release date and the days on which they logged in, which results in 219 daily quasi-experiments.

2.2 Data Types Included

There are largely three types of information included in the dataset. The first type is daily transaction information at the account level. The information is pertinent to the debit or credit card accounts registered to the app. The amount, date, merchant, and category description of each transaction are included in this information. The second type of data is time stamp information at the user level. From this app usage information, I figure out the days on which the users logged in and calculate the frequency of log-ins in a given month. The last set of data is month-end demographic characteristics and bank account information at the user level. Demographics include gender, age, app registration date, and tenure with the Bank. Bank account information includes the number of accounts for each user and balance by account type (interest-bearing checking accounts, checking accounts with no interest, savings accounts, other investment assets, credit cards, and other liabilities accounts).[5] It is used to sort the users by liquid wealth and financial savviness in subgroup analysis as well as to investigate the behavior of account management after the receipt of a message.

2.3 Advantages and Disadvantages of Data

The data in this paper have several important advantages over aggregated data at the national level and other money management app data. First of all, measurement error is very unlikely as transactions are recorded real-time. As a result, the dataset reflects the timing and amounts of purchases precisely. For instance, unlike the dataset in the paper, the Survey of Household Spending, which is aggregated at the national level and reported by Statistics Canada, imputes missing values and categories. In addition, since this free-of-charge app is marketed by the Bank and was once a top download app in the App Store and Google Play, this dataset gives a somewhat unbiased representation

[5]The month-end information is pertinent to all bank accounts registered with the Bank, not just the ones registered to the app.

of households. For instance, in comparison to other money management apps used by mostly young and educated individuals (Baker, 2018), this app is used by relatively older individuals as well. In this regard, the dataset is advantageous for precise measurement and unbiased representation.

Most notably, another crucial advantage of the data is that the contents of overspending messages are personalized to each app user with respect to his or her past spending patterns; only when spending on a day exceeds a pre-determined threshold, is an overspending message generated on the next day. This personalization is in contrast to many field experiments on messages in which the participants of the treatment group receive the same content.[6] Thus, these tailored messages in this paper, as opposed to one-size-fits-all messages, are more suitable for seeking changes in individual behavior.

Finally, transactions are recorded at the daily frequency, and log-in information is verifiable. More to the point, prior studies of field experiments on messages suffer from measurement error in that it is difficult to determine whether the treatment group actually read the messages. Furthermore, since data from these studies were recorded at the monthly frequency, it is difficult to pin down on exactly when the effect started to occur, at a level more granular than monthly. On the other hand, in my case, it is assured that the app users read an overspending message if they log in, and I can observe how the message recipients differentially spend relative to the non-recipients on a daily basis. Hence, I can establish the causal effect of a message on spending behavior with the verification of message check and precisely measure when the effect starts to emerge thanks to the high frequency of the data.

On the other hand, the dataset also has a few caveats. First, as the dataset contains information about spending tied to this Bank's accounts only, it may not provide a holistic view of individuals' financial situations. Yet, to alleviate this problem, I only consider users with at least one credit card for analysis. Especially, as the Bank of Canada reports that the average number of credit cards per Canadian adult is 2.0 in 2017, confining to app users with at least one credit card will capture a substantial fraction of their total

[6] One exception is the inclusion of names of loan officers in text messages sent to microloan borrowers in Karlan et al (2015). Indeed, mentioning the names resulted in a lower likelihood of late repayment among the repeated borrowers.

expenditure. In this regard, this way of filtering a specific group of app users will remove some concern regarding a comprehensive understanding of their finance.

Another caveat is that the dataset does not include direct cash transactions. However, the concern is somewhat alleviated by the app's feature and Canadians' payment methods. First of all, the app includes the Cash category, which refers to ATM withdrawal. Even though I do not observe what the users spend on with cash after withdrawing it from an ATM machine, this feature is a good approximation of cash transactions. Moreover, the use of cash in Canada is comparable to that in the United States, where most payments do not occur by cash; the share of cash payment by value in 2017 is 15% in Canada according to the Bank of Canada whereas the Federal Reserve Bank of San Francisco reports that the counterpart in the United States is 9%. Thus, most transactions occur by non-cash payment. More to the point, the credit card use especially has been on the rise. Since the share of credit card payment by value in Canada increased to 56% by 2017, which is an increase of 10 percentage points from 2013, and the proportion of spending from debit and credit card payment is significant at 82% by value in 2017, the dataset captures the rising trend of card spending as a method of payment.

2.4 Data Summary and External Validity

My data cover 55,586 app users, who have signed up for the app after the release of the new version including the overspending message feature in June, 2017. In order to show that the dataset is good for external validity, I include a summary statistics of the app users considered in this paper, especially as Baker (2018) points out that the characteristics of users of personal finance software might be different from those of the nation, indicating greater presence of male and the young. Panel A of Table 2 shows the proportion of male is near 50%; thus, the user base is not concentrated on men. Similarly, the mean age of the users in the dataset is 36.3 years old, and the age ranges from 22 at the 10th percentile to 57 at the 90th percentile; the user base is not particularly youth-oriented. In this respect, the dataset is fairly good in terms of unbiased representation for external validity.

In terms of app usage, Panel B indicates that the monthly log-in frequency is fairly

low, and as a result, the number of messages seen per month, which is defined to be the case where a user logs in on the date when an overspending message is generated, is even smaller. Panel C shows that the median user has liquid wealth, which is defined to be the difference between the balance of checking and savings accounts and that of credit card accounts, less than the median monthly spending. Panel D indicates that the average app user spends C$5,055 per month[7]; the average spending is mostly concentrated on Cash, Shopping, and Home, and the median spending is concentrated on Shopping.

3 Research Design

I explain in this section how I implement empirical strategy to investigate the role of technological nudges in shaping consumer behavior. In addition to the basic scheme of examining whether the app users reduce overall spending and how they make categorical adjustments after receiving and seeing an overspending message, I examine their patterns over a longer time horizon and conduct cross-sectional analyses by analyzing users with different demographic characteristics, financial experience, financial status, app experience, and education level. Furthermore, I investigate among couples whether the users whose spouse is informed of overspending indicate reduction in spending relative to the users whose spouse did not receive any overspending message. Last, I examine whether the users get to differentially manage and monitor their accounts once alarmed by overspending messages.

3.1 Choice of Research Design

A major identification challenge with documenting the causal effect of an overspending message on changes in spending is that using the number of messages seen as the main explanatory variable in OLS is not appropriate due to endogeneity concerns regarding confounding factors. For instance, it is very likely that those who read a low number of messages are inherently different from those who read a high number of messages.

[7] As the average expenditure per household in Canada, including income taxes, is C$86,070 in 2017, according to Statistics Canada, the spending data on the app users capture a high fraction of household spending.

Since the data in this paper do not involve randomized field experiments as in Medina (2017) and Karlan et al (2016), in which a nudge message is sent randomly to the customers, I need to focus on the quasi-random treatment of a message in the app around the threshold; the overspending message threshold is exogenously chosen by the Bank and uniform across the app users, thus providing a random source of variation around the cutoff.

In particular, I exploit a sharp regression discontinuity approach in investigating whether spending patterns are affected by a nudging feature, following Imbens and Lemieux (2008). The running variable in this setup is spending multiple compared to historical average. As explained in Section 2, the spending multiple is computed by dividing the spending on a particular date by the calendar date average over the past 12 month, and if the spending multiple on a particular date exceeds the predetermined threshold, the app users receive an overspending message on the following date. Since whether this message is reported or not is deterministic depending on the spending multiple, the treatment of messages on the app users is also deterministic; as a result, a sharp regression discontinuity method is suitable in this case.

The baseline regression discontinuity equation would be the following:

$$Y_{i,t+1} = \alpha + \beta * 1[X_{i,t} > cutoff] + \gamma_1 * 1[X_{i,t} > cutoff] * f(X_{i,t}) + \gamma_2 * 1[X_{i,t} \leq cutoff] * f(X_{i,t}) + \epsilon_{i,t+1} \tag{1}$$

where for each app user i on date t, X is the daily spending multiple as the running variable, and f(·) is a local linear function that fits the observations. For the outcome variable Y, I consider the following five measures of spending on the next day of a log-in: spending amount in Canadian dollars, spending amount scaled by the daily average of the same calendar date over the last 12 months, spending amount scaled by the daily average over the last 365 days, number of purchases, and dollar amount spending per purchase. This basic model will estimate how an average user's spending pattern changes around the cutoff over a day.

For the sake of readability and visualization, I pool the 219 daily quasi-experiments across the dates. For the baseline, I run Regression 1 with individuals' overall spend-

ing as the outcome variable to investigate whether they reduce overall spending after receiving an overspending message on overall spending. In addition, I run the regression again with categorical spending to measure how they make spending adjustments for each spending category. In Section 5, Placebo Tests and Robustness, I also investigate the effects of categorical overspending messages on categorical spending by running Regression 1 by using categorical spending multiples as the running variable instead of overall spending multiples.[8]

3.2 Additional Analyses

In addition to the baseline model, I investigate the duration of impact from nudges by varying the spending horizon. In other words, I vary the spending horizon from one day in the baseline to a different number of days after the receipt of a message. Using spending on each of the days since the receipt as the outcome variable in Regression 1 will measure the dynamic impact of a message, and using the sum of spending over a time horizon after the receipt as the outcome variable will measure the cumulative impact. Since the user characteristics will still be continuous at the cutoff, these tests will consequently measure how long the effects of nudges on shaping spending patterns last and whether they lead to the formation of new spending habits.

Furthermore, I investigate the heterogeneous effects of the app feature by running the same Regression 1 among different subgroups of the users.[9] First, I test the differential effects of nudges on demographic characteristics, which include gender and age. These tests examine whether a particular gender group or a group with differing financial incentives along different stages of life reacts more significantly to nudges. Then, I run the same regression on users with different financial status, namely liquid wealth, which is the difference between the balance of checking and savings accounts and the balance of credit card accounts, and the fraction of checking account balance to the sum of the bal-

[8]Because only a small number of categorical overspending messages is present in the data, I mostly focus on overall overspending messages.

[9]In order to hold other macro-economic effects fixed and only make cross-sectional comparison, I sort the users for each day of the sample by their user characteristics and pool the observations of spending across dates. Since the information of the sorting criteria is only available at the monthly level, I use user characteristics from the previous month.

ances of the checking and savings accounts. The former criterion tests whether the users who are more liquidity-constrained make greater spending adjustments, and the latter criterion tests whether more financial savvy users react more significantly to the messages as a small fraction of deposits in non-interest bearing accounts indicates a greater level of financial risk taking in investment. Moreover, I carry out heterogeneity analysis by app experience by sorting the users in terms of how long the users have used the app. This test examines whether users freshly exposed to the app service are more aware of their finance and thus change their spending patterns. I also sort the users by the fraction of population aged 15 or above with a degree or a diploma in the city they live in by matching my dataset with the Survey data of Statistics Canada. This criterion is used as a proxy for financial knowledge and literacy to investigate whether the users living in a relatively well-educated region reacts more strongly to nudges.

In addition, I examine the contagion of nudges in households. In particular, I focus on couples and consider only the case where only one of them logged in; I test whether the spouse of a user who has received and seen an overspending message spends less relative to the spouse of a user who has logged in but is without an overspending message. Because I do not directly observe the spouse relationship, I identify two users as a couple if their age difference is equal to or less than 15, they are of different genders, they reside in the same city, and they share at least one Bank account. Through this filtering, I identify 4,199 unique couples and run the test of peer effects based on spending from the non-joint accounts of spouses who did not log in.

Finally, I examine how the way the users manage their personal finance differs once they are reminded of overspending. In other words, I try to detect if the overspending message recipients sign up for a higher number of new Bank services, cancel some of their existing services, or monitor their accounts more frequently, relative to the non-recipients. In order to establish such relation, I again run Regression 1 and use the following as the outcome variables: monthly changes in the number of total asset accounts, liquid asset accounts, total debt accounts, and liquid debt accounts, the probability of logging in to the app at least once, and the number of log-ins over the next 7 days or 30 days.

3.3 Identification Assumptions for Sharp Regression Discontinuity

The most basic assumption to be met is that a discontinuous jump in the outcome variable should be observed near the cutoff, and the discontinuity assumption will be illustrated in detail in Section 4. Because each quasi-experiment is carried out on a daily basis since the introduction of the app feature in June, 2017, graphically illustrating all the discontinuities is not efficient. Especially, in this research design, app users belonging to a particular bin of spending multiples at a time may belong to a different bin at some other point in time. This is different from the standard regression discontinuity design, in which groups belonging to a particular bin stay the same throughout the sample period. Therefore, I follow the methodology of Agarwal et al (2018) and Howell (2017),[10] by pooling observations across dates and putting them into bins of spending multiples, on which the generation of overspending messages is based; this method will allow me to estimate precisely the local average treatment effect of a single overspending message.

To validate the sharp regression discontinuity design, several identification assumptions must be satisfied. First, the density distribution of the running variable, or spending multiples for overall spending, must be smooth at the cutoff. The smoothness in the relative frequency of users for each bin of spending multiples will indicate random treatment on the app users around the threshold exogenously chosen by the Bank. As Panel A of Figure 2 indicates, the density of spending multiples for overall spending is smooth around the cutoff of 2. For concreteness, I run the McCrary test on the continuity of the distribution, and the test statistics falls between -0.121 and 0.010, failing to reject the null hypothesis of zero change in density around the threshold at 95% significance level.

Closely related to this assumption, the assumption that app users would not have manipulated spending multiples at the cutoff must be also satisfied. It is doubtful that app users want to spend just above or below the threshold as there is no explicit punishment or incentive for doing so. For instance, no special reward such as bonus cash back perks is given by the Bank to users for spending just below the threshold. Especially, since

[10] Agarwal et al (2018) pool 743 quasi-experiments showing a discontinuous jump in borrowing and spending around different FICO scores, and Howell (2017) pools competition results of different years with a discontinuous jump in the success measure around different rankings.

the cutoff number is exogenously chosen by the Bank, and the users do not know the algorithm as to how overspending messages are generated, there is little reason to suspect that they would have manipulated their spending multiples precisely at the cutoff.

Even though there was no manipulation of spending multiples, one potential worry is that the change in spending behavior on the following day of a message occurs due to mean-reversion; the users who spent just above (below) the cutoff on a given day could have simply come down (up) to a daily average, causing discontinuity in spending on the next day, instead of responding to a message. In order to rule out this hypothesis, I show in Panel B of Figure 2 that the spending amount on the day of message generation is continuous at the threshold multiple. Hence, if there is differential in spending behavior on the following day, then it must stem from the treatment of messages, not from mean-reversion.

Finally, observable user characteristics need to be continuous around the cutoff. Only when this assumption is met, will the exogenous treatment of messages cause a differential response in the outcome variables. The following observables are included for the continuity test: gender, age, enrollment in the app, tenure with the Bank, the unconditional probability of logging in on a day, and the probability of at least one log-in over the past 7 days conditional on logging in on a day. As Figure 3 displays, the observable characteristics of the users categorized by spending multiples are continuous near the threshold. Notably, the continuity of both the unconditional log-in probability and the conditional past log-in probability at the threshold suggests that there is little correlation between the likelihood of logging in and that of receiving a message. In this regard, the identifying assumptions for the sharp regression discontinuity are met, and the research design is validated.

4 Results

4.1 Baseline Results

As the research design of regression discontinuity aims to achieve, Figure 4 indicates a discontinuous downward jump of outcome variables at the threshold. Panels A, B, and C indicate that among the users who logged in, those who spent just above the threshold and received an overspending message responded by reducing overall spending on the following day compared to the users who spent just below the threshold and did not receive one; the former group spent C$8.15, which corresponds to a spending multiple of 0.033 on a daily basis over the last 12 months and a spending multiple of 0.054 on a rolling daily basis over the last 365 days, less relative to the latter group. Panels D and E indicate that the message recipients made adjustments on the extensive margin by reducing the spending frequency by 0.06 as well as on the intensive margin by spending C$2.07 less on a per-purchase basis.[11, 12]

In order to test if the effectiveness of overspending messages differs at different points in time, I run Regression 1 using dollar amount of spending on the next day for the following cases: weekday vs. weekend and dates up to the time when the users read a message for the first time vs. dates after the first check of the message. Then, I compare the magnitudes, as illustrated in Figure 5. Panel A indicates that there is a statistically significant spending gap on both weekdays and weekends. Although the coefficients are not different from each other, since there is more discretionary spending on weekends, the coefficient is larger on weekends. Panel B illustrates that the message is effective in delivering spending reduction on dates up to the first time the users read a message. Although the coefficient on dates after the first check of a message is not statistically significant due

[11] As the Bank of Canada reports that the mean amount of a credit card transaction and a debit card transaction in 2017 is C$62 and C$44, respectively, the average amount of a purchase in this paper is in line with these statistics.

[12] A possible form of spending adjustments on the intensive margin can be found in Stroebel and Vavra (2019), where falling home prices reduce the wealth of homeowners and increase their demand elasticity and price sensitivity, as a result of which they reduce nominal spending, buy more goods with a coupon, and buy a higher fraction of generic goods and items on sale.

to somewhat smaller sample size,[13] and most of the baseline result comes from seeing the message for the first time, the recipients still spent about C$5.84 less than the non-recipients even if they had seen a message before. Hence, overspending messages are effective regardless of the time point at which the users receive them.[14]

Table 3 illustrates the composition of spending categories on which the message recipients made spending adjustments out of the C$8.15 difference. Panel A displays that they made adjustments on both discretionary spending and necessary spending for everyday life, and the adjustment margin is very similar. Panels B and C depict a finer categorical spending breakdown. The users reduced spending on the Shopping category among the Wants category, with the statistically significant reduction of C$3.09 while they did not make statistically significant adjustments for the Needs category. It is interesting that the users reduced spending on every category of spending for necessity, but not significantly on any of them.

This phenomenon that Shopping is the first category to cut down on spending is most consistent with high inter-temporal elasticity of substitution of expenditure for Shopping as in Parker (1999); since there is little to no cost associated with accelerating or delaying purchasing somewhat durable goods, Shopping is the first category which the users started to reduce spending on. For instance, when the U.S. government implemented federal income tax rebates in 2001, the largest cumulative response came from spending on apparel (Johnson et al., 2006), and when it distributed economic stimulus payments in 2008, the recipients indicated higher marginal propensity to consume on durable goods including vehicles and apparel instead of that on nondurable goods (Parker et al., 2013). In this regard, the behavior of the app users is consistent with making spending adjustments on durable goods in the literature.

[13] A substantial number of app users never received an overspending message during the sample period, and the observations of these people are only included in the category of dates up to the first message.

[14] This is similar to Karlan et al (2015), in which the timing of when the borrowers received a text message reminder of microloan repayment, relative to the due date, had no effect on the likelihood of repayment.

4.2 Duration of Impact

Figure 6 shows the spending gap between the overspending message recipients and non-recipients over time horizons longer than one day. Panel A indicates the dollar amount spending gap on each of the days after the receipt. There is a delayed response on the second day, and then there is no more significant spending gap afterward. As a result, Panel B shows that the cumulative spending gap actually widened to C$9.79 over the two day horizon and is stabilized after the fourth day. Although the long-term effect cannot be measured precisely after the fourth day due to high standard errors,[15] there is no sign of reversal in the cumulative spending gap, and the effects of messages are definitely present and do not dissipate; messages had temporary effect on flow spending in the short term and permanent reduction in cumulative spending in the long term.

Appendix Figure A.1 also shows the long-term effect of an overspending message on the spending gap in terms of dollars normalized by daily average spending over the past 365 days. The effect is statistically significant up to the second day after the receipt, and it is also stabilized after the second day. Figure A.2 indicates the results of placebo tests on the long-term effect for the app users who did not log in. The spending gap in terms of dollar amount and rolling multiple is statistically insignificant on each of the days after receipt of a message, and as a result, the cumulative spending gap is also statistically insignificant.

4.3 Heterogeneous Effects - Subgroup Analysis

Figure 7 shows differential responses to overspending messages by different subgroups of the users. Panel A shows that there is no differential effect of gender[16]; the messages are effective both for the male recipients, who spent C$8.61 less than the male non-recipients, and the female recipients, who spent C$7.94 less than the female non-recipients. Panel B indicates spending responses grouped by different ages. As the the

[15] Higher variability of cumulative spending over longer horizon within spending multiples attributes to higher standard errors.

[16] The basis on which this analysis is conducted is to test if women are more responsive to overspending messages as Shoham and Brencic (2003) point out that they have tendency to make unplanned purchases.

spending gap between message recipients and non-recipients is statistically significant for Quartiles 2, 3, and 4, effects of messages come from older users. Due to the absence of information on the income of the users, age is a proxy for income, and the users with higher income made spending adjustments.

Panels C and D show how the users with a differential degree of financial status responded to overspending messages. Panel C shows results for different quartiles of liquid wealth measured by the account balance of Bank-wide accounts, in addition to the accounts registered to the app, on the last day of the previous month. More specifically, it measures the difference between the balance of checking and savings accounts and the balance of credit card accounts, and the statistical test examines whether the effect of fintech nudges is more likely to come from behavioral bias or liquidity constraints. The result indicates that the effect mostly comes from the messages recipients of the third and fourth quartile groups, who have relatively higher liquid wealth and reduced spending by C$13.99 and C$18.61, respectively. In other words, these relatively cash-poor users were already constrained and had little room to further decrease their spending after the receipt of a message. Panel D shows spending responses based on how financially savvy the users are. It is inversely measured by the fraction of checking account balance to the sum of balances of checking and savings accounts at the end of the previous month, as the lower the fraction is, the higher proportion of liquid assets the users put in to the money market to earn interest. As the bottom three quartile groups indicate a significant behavioral change for dollar amounts, these users probably have a better understanding of financial awareness, especially with respect to their own finance, and learned to save by reducing overall spending.

Panels E and F show differential responses based on the app experience and education level. Panel E indicates behavior of groups with different number of days since the registration of the app. The tendency to pay more attention to messages and reduce spending is present for those who recently signed up for the app; the bottom quartile group reduced C$12.18, and the second quartile group reduced C$13.45. Panel F shows how the level of education is differentially associated with changes in spending, and the education level refers to the fraction of population aged 15 or above with a diploma or a degree at the city

level and is used as a proxy for financial knowledge and literacy.[17] The third and fourth quartile groups reduced C$15.76 and C$13.86, respectively. Thus, the users who reside in a relatively highly educated region tend to respond to overspending messages, and this result could arise from a lower degree of the present bias, which individuals with low education levels are more susceptible to.

4.4 Spillover Effects

Table 4 indicates the results for the contagion of nudges among couples. More precisely, since I made sure that only one spouse of a couple logged in and the other spouse did not log in, it indicates the spending gap between the users who themselves did not log in but whose spouse logged in and read an overspending message and the users who did not log in and whose spouse logged in but did not receive a message. Because the user population of couples is substantially smaller than that of individual users, the small number of observations in Table 4 leads to a somewhat imprecise estimate. Nevertheless, the spending gap between the two groups is statistically significant at C$7.10 Furthermore, compared to the effect of messages at the individual level in the baseline, which is C$8.15, the spending gap caused by spillovers is not that different from the direct effect. Hence, there is suggestive evidence that a message received by one spouse could affect the spending of the other spouse in the same household and the magnitude of spillover effects could be non-trivial.

These results indicate that spending is a dimension of household behavior that is subject to peer effects and thus nudges can have amplified effects. Just as individuals are more likely to invest in the stock market (Hong et al., 2004), strategically default on mortgages (Guiso et al., 2013), choose to do foreclosure (Gupta, Forthcoming), and adopt the same type of mobile phones (Bailey et al., 2019) when their peers do, reducing overall spending can also occur due to peer effects. Furthermore, as this contagious action can happen within households, not just across co-workers (Kalda, 2018) or online friends (Bai-

[17] The basis on which this analysis is conducted is to test if these fintech nudges are more salient for financially illiterate people as Stango and Zinman (2014) document that effects of nudges on avoiding overdraft fees by means of exposure to a higher fraction of questions related to overdraft fees on a survey are more pronounced for the respondents who are at low literacy and education levels.

ley et al., Forthcoming), these spillover effects suggest that spouses make unitary financial decisions, possibly through communication about their financial situations, and the impact of nudges can be quite substantial beyond the direct impact on the users who have received an overspending message.

Table 4 also shows the magnitude of spillover effects by gender. The responses are somewhat more pronounced when a female spouse in a household receives a message and her male spouse makes spending adjustments as opposed to when a male spouse receives a message and his female spouse makes spending adjustments. This result is in contrast to spending response on an individual basis in Panel A of Figure 7, in which the reduction of spending is similar for both men and women. On the other hand, as far as intra-household financial decision making is concerned, decisions on everyday spending are dominantly made by women, as in Mader and Schneebaum (2013).

4.5 Personal Finance Management

This subsection investigates whether the app users managed their personal finance differently in the longer run after receiving an overspending message. First, I run Regression 1, using the month-to-month change in the number of Bank accounts as the outcome variable; the rationale is to test if the alarm raised by the overspending messages led the recipients to save and invest more by signing up for new checking, savings, or investment accounts, or if annoyance caused by the negative connotation of the overspending messages led them to terminate the Bank's product services.

Panel A of Figure 8 indicates that the recipients did not sign up for a higher number of new asset accounts, which include checking, savings, and investment accounts, after receiving a message, relative to the non-recipients. Panel B indicates that they did not differentially increase the number of checking and savings accounts, which are relatively easy to open. Similarly, Panel C indicates that the recipients did not cancel their existing debt accounts, which include credit card or mortgage accounts, either, and Panel D indicates that they did not close down their credit card accounts, which again are relatively easy to terminate. Thus, the receipt of an overspending message did not result in different behavior with respect to altering the Bank's services, and they did not necessarily lose

trust in the Bank.

The phenomenon that the effects of a nudge on a behavioral dimension in which it aims to induce changes do not spill over to another behavioral dimension is consistent with Bursztyn et al (2016). More to the point, in that study, a higher proportion of the treatment group made minimum repayment on their credit card debt after receiving a reminder text message, which included moral appeal, compared to the control group, who only received a simple reminder. Despite the presence of the differential in repayment, there was no differential in the domain of card usage. As a matter of fact, most of the survey respondents wanted to receive a reminder text message going forward. Therefore, although the messages in this paper included a negative connotation, they did not result in dis-satisfaction with the Bank, which in turn would have led to switching to services of another bank. Hence, no unintended behavior with respect to changes in the number of Bank accounts in this paper is consistent with no spillover effects of nudges to a different dimension.

Next, I examine how the recipients of messages pay attention to their accounts by replacing the outcome variable with the probability of logging in to the app at least once over the longer horizon. Panel A of Figure 9 shows the result over the next 7 days of a log-in. The message recipients were actually *less likely*, instead of more likely, to check their app afterward relative to the non-recipients; the gap in the log-in probability is 2.3 percentage points relative to the baseline probability of 60 percentage points, or 3.8% decrease. Panel B shows that even when the horizon is extended to 30 days after a log-in, the recipients logged in less likely with the gap being 1.6 percentage points relative to the baseline probability of 80 percentage points, or 2.0% decrease. Therefore, viewing an overspending message yielded substantial behavioral change in monitoring their accounts, and this tendency persists over time.

Even though this empirical result is in stark contrast with the notion that those lacking self-discipline would pay more attention to their finance with the aid of nudges, the ostrich effect provides some insight into this phenomenon. The ostrich effect is referred to as ignoring adverse financial information purposely to seek psychological comfort, noted by Galai and Sade (2006) and Karlsson et al (2009). In close relation to this paper, Olafsson

and Pagel (2017) document that when money is going out of their account or when the account balance becomes negative, people log in to their personal financial management software less frequently so as to forget about financial problems. Similarly, in this paper, the app users, who self-selected to manage spending via use of the app, choose to avoid looking at their finance as the signal of overspending may yield psychological discomfort. Therefore, although daily overspending messages help the app users reduce overall spending upon receipt, in the longer run, they choose to view their financial status less likely so that they would not be disturbed by their overspending patterns.

One potential alternative hypothesis as to why the message recipients reduced the probability of checking the app is that they ran down on liquidity. In order to rule out this hypothesis, I examine whether the recipients continued to use the registered cards to make a purchase as well as whether they had enough liquidity in the accounts. Panels A and B of Appendix Figure A.3 show the probability of making at least one purchase over the next 7 and 30 days, respectively, suggesting that the recipients did not stop using the registered cards. Moreover, Panel C shows that the month-end liquid wealth is higher for the recipients with a statistically insignificant gap of C$160.66. Hence, the recipients had enough funds in the accounts to make purchases in the subsequent period, and the ostrich effect originates from psychological discomfort, instead of the exhaustion of liquidity.

While Panels A and B of Figure 9 indicate that the recipients reduce the likelihood of logging in to the app on the extensive margin, relative to the non-recipients, Panel C shows that among the app users who made at least one log-in over the subsequent 7 days, the message recipients rather increased the frequency of checking the app by 0.047 if anything. Panel D illustrates that the frequency gap widened to 0.16 over the next 30 days although the gap is again not statistically significant.[18] Hence, these results suggest that the message recipients paid slightly more attention to their finance, conditional on keeping using the app in the subsequent period.

Panels E and F illustrate how the negative adjustment on the extensive margin and the positive adjustment on the intensive margin affect the overall future frequency of log-ins.

[18]The log-in frequency is normalized by the monthly average log-in frequency in Appendix Figure A.4, and the results are similar.

Panel E indicates the unconditional number of log-ins over the next 7 days of a receipt, including the users who stopped looking at the app. The frequency gap is 0.013, and it is statistically indistinguishable from 0. Similarly, Panel F indicates that the frequency gap over the next 30 days is 0.030, which again is statistically not significant; this little gap in the number of log-ins over the subsequent period could provide explanation as to why the effect of the messages does not amplify hugely over the long run as in Panel B of Figure 6. Therefore, the negative extensive margin effect and the positive intensive margin effect offset each other in the longer term.

In order to look at the log-in behavior more closely, I re-run Regression 1 with the log-in probability over the next 7 days after a receipt as the outcome variable, but I now group users by their user experience. Panel A of Figure 10 indicates the point estimate measuring the log-in probability gap between the message recipients and the non-recipients for users with different lengths of user experience, which is measured by the number of days since the registration of the app. It shows that even when the app experience gets longer and the users become more accustomed to using the app, the ostrich effect persists. Panel B groups the users by those who saw the message for the first time and those who saw the messages multiple times. Interestingly, the ostrich effect is very weak for the first message, but it is very strong for the second message seen and beyond. Hence, the ostrich effect persists over the longer run, and is stronger for the users who are familiar with how the messaging system works, probably because the feeling of guilt emerges every time they read a message.

4.6 Potential Mechanisms & Policy Implications

There are potentially two mechanisms through which the users reduce spending upon seeing an overspending message. The first is the attention channel; since these individuals are most likely to have signed up for the app to monitor their finance, the messages should remind them of their discipline to manage their spending. The second is the guilt channel; the wording overspending carries a negative connotation, and this psychological discomfort could pose a cue to the stoppage of spending an extra dollar.

Although it is not possible to completely disentangle the effect of one channel from

the other, I can at least gauge which channel dominates the other at different times. More to the point, when the users see an overspending message for the first time, the attention shock should be relatively greater than guilt, based on the greater effect in reducing spending as in Panel B of Figure 5 and no effect in reducing log-in probability in Panel B of 10. Conversely, the smaller effect in reducing spending and the appearance of the ostrich effect when the users see the message for multiple times suggest that the emotional channel should dominate the attention shock in this case.

These mechanisms lead to suggestions of policy implications. First, exogenous attention shock should be provided from time to time, and it could be implemented in the form of push notifications. The effectiveness of messages gets reduced when the users become accustomed to seeing the messages, and the long-run effectiveness does not amplify hugely over the long run. Hence, keeping the users informed of their past spending patterns should serve as a more effective discipline device. Second, more encouraging language in the messages should be used. The feeling of guilt would dominate the salience of messages and mitigate the effectiveness of messages via lower probability of monitoring the accounts. Hence, framing messages in a more positive manner could inhibit the users from stopping using the app altogether.

4.7 Back-of-Envelope Calculation

In this subsection, I carry out a back-of-the-envelope calculation to gauge the aggregate effect of overspending messages. The direct effect of a single message on an individual is about C$8 per user, and an average user receives four messages in a month. [19] The indirect or contagious effect is about C$7 for each of the married users, and the proportion of married users to the total user base is about 4,000 out of 50,000. Upon assuming that an average app user read all the messages sent, the user would have reduced overall spending by C$411 per year, relative to a counter-factual person, who would not have received any message. Compared to an average Canadian individual's outstanding credit card debt of C$3,954 in the third quarter of 2016, according to TransUnion Industry Insights Report, this reduction in spending corresponds to 10% decrease of the debt size.

[19] I take into account for the finding that the direct effect is smaller for the high order messages.

Accordingly, the aggregate effect on the 1 million app user base would have been a yearly reduction of spending about C$400M. Given that the marginal cost of implementing this nudge feature is very low, this saved amount could be meaningful.

5 Placebo Tests and Robustness

In this section, to confirm that overspending messages indeed played a role in inducing spending changes, I run three types of placebo tests. First, I examine the spending behavior of the users who did not log in and as a result did not read the message despite the possibility of having received one.[20] Figure 11 graphically shows that there is no discontinuity in spending around the threshold when the users did not log in; the level of spending is closely aligned for the non-recipients regardless of whether they logged in, but the spending gap is present only for the recipients who logged in. Therefore, this placebo test confirms that the effects of fintech nudges come from the users who have actually read the messages and taken action.

Second, I re-run the baseline analysis at different false numerical values of threshold as an additional placebo test. Figure 12 indicates that for all of the five outcome variables, the sharpest and statistically significant gap occurs at the exact threshold of 2. This result confirms that I exploit a sharp discontinuity at which overspending messages are generated. Consequently, the sharp regression discontinuity design establishes the causal effect of fintech nudges on changes in spending behavior around the cutoff.

Third, I examine the spending response of the existing users, who signed up for the app service before the newer version of the app came out with the overspending message feature. To do so, I select on the existing users whose level of spending matches with that of the users in the baseline analysis and run the baseline specification when they logged in during the sample period. Figure 13 indicates that the level of spending is closely aligned for the users who spent the below the cutoff multiple regardless of whether they signed up for the app before or after the release of the newer version, but the spending gap is only present for the users with the new app. Hence, it confirms that the effects are not

[20] For this placebo test, I use exactly the same group of users as in the baseline analysis.

present for the existing users, who most likely did not have the overspending message feature in the app.

Although I cannot identify who downloaded the newer version among the existing users, as time went by, a greater fraction of them would have downloaded it and responded to overspending messages. To exploit the gradual adoption of the newer version, I again apply the baseline specification to the existing users when they logged in within the most recent month of January 2018. Appendix Figure A.5 shows that the existing users somewhat reduced spending in the most recent month, indicated by statistically insignificant but slight negative discontinuities in the outcome variables. Therefore, it provides weak evidence that they gradually adopted the technology and responded to the overspending messages.

Moreover, I present the results of the baseline regressions for categorical overspending messages in Table 5. As I drop observations on a date if the count of categorical spending on the date of the month over the past 12 months is fewer than 3, the number of observations is significantly lowered. Furthermore, since the thresholds at which categorical overspending messages are generated are higher than that for overall overspending messages, the relative fraction of observations on the right of the threshold is even smaller. Nevertheless, Table 5 provides suggestive evidence that just as overall overspending messages lead to reduction in overall spending around the threshold, categorical overspending messages also lead to reduction in categorical spending around the threshold; the recipients spent C$1.75 on Dining Out, C$8.01 on Shopping, and C$5.93 on Groceries less relative to the non-recipients. Similar to findings in Panel B of Table 3, the users made the largest spending adjustments on the Shopping category.

Furthermore, I conduct two different types of robustness checks. First, I check the robustness of the baseline research design by including the app users who only have debit cards whereas the baseline results are confined to the app users who have at least one credit card for the sake of external validity. Panel A of Table 6 indicates the results for all the app users, including both debit card holders and credit card holders, and all five outcome variables are statistically different from zero. As debit card holders tend to be less wealthier than credit card holders, the spending gap in dollar amounts reduced

slightly to C$7.59 compared to C$8.15 in the baseline, but the spending multiple gaps are also similar as in baseline results. Overall, the effects of fintech nudges are present not only for credit card holders but also for debit card holders.

Finally, I show that the local average treatment effect of a message on spending behavior is robust to different bandwidth selection. An optimal bandwidth for the baseline case is chosen to minimize the mean-square error of the estimator to take into consideration tradeoff between precision and bias. Panel B of Table 6 shows that the results are robust when the bandwidth is reduced by half or doubled; the magnitude is greater for a narrower bandwidth, and it is smaller for a wider bandwidth. The estimates are statistically significant for most of the cases, and hence, overspending messages led to behavioral changes in spending near the cutoff.

6 Conclusion

In this paper, I examine whether technological nudges in a money management app can induce changes in spending patterns among the app users. Overspending messages indeed lead to differential spending behavior. Using a regression discontinuity design, I find that the app users who received and read an overspending message reduced a non-trivial amount of overall spending on the following day, compared to the users who did not receive one. They adjusted the spending margin both extensively and intensively, and especially on the Shopping category. The overall cumulative spending gap between the recipients and the non-recipients widened over a two-day horizon, remained statistically significant over a four-day horizon, and remained negative over longer time horizons. The effects of nudges are more pronounced for the users who are older, have higher liquid wealth, are more finance-savvy, are newly introduced to the app, or live in a city with high education levels. I also find suggestive evidence that the effects of nudges could spread over from one app user to another in the same household. Despite the change that the recipients reduced spending after reading an overspending message, they indicated a lower likelihood of logging in to the app in the longer run, which is not a reaction the app aims to induce.

This paper documents that although they are not as coercive as penalty, fintech nudges can still function in the households' space of financial decision making. Unlike reminders for repaying debts on time or avoiding overdraft fees, which involve direct incentives of saving bank service fees, the benefits of reducing overall spending are not imminent, but this paper finds that the app users react to the messages and make different spending decisions. Thus, when implemented properly, nudges can shape households' behavior in the way the designer of the nudges intends.

In addition, nudges can be easily implemented with the help of technology. In particular, app use is easily accessible to a wide range of population and can be a a relatively cheap method of implementing nudges. More importantly, people can continuously monitor their financial situations via a mobile app. Furthermore, nudges can appeal to a variety of app users, including those who may not know much about financial services. Thus, the power of nudges can be amplified when used in conjunction with technological advancement.

However, the use of nudges should be handled with care. When the nuisance of the nudges is negative, they could also result in unintended behavior; the participants may react to them by selectively paying less attention to their financial status from feeling guilty of overspending. Consequently, policy makers need to contemplate the types of unintended behavior of the participants as well as how best to induce intended changes in the participants' behavior.

7 References

Addoum, Jawad M., Stefanos Delikouras, George M. Korniotis. 2015. "License to Spend: Consumption-Income Sensitivity and Portfolio Choice."

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2018. "Do Banks Pass Through Credit Expansions to Consumers Who Want to Borrow?" *Quarterly Journal of Economics*, 133(1): 129-190.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2015. "Regulating Consumer Financial Products: Evidence from Credit Cards." *Quarterly Journal of Economics*, 130(1).

Alan, Sule, Mehmet Cemalcilar, Dean Karlan, and Jonathan Zinman. 2015. "Unshrouding Effects on Demand for a Costly Add-on: Evidence from Bank Overdrafts in Turkey." National Bureau of Economic Research, Working Paper.

Allcott, Hunt and Judd B. Kessler. 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics*, 11(1): 236-76.

Bailey, Michael, Drew Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2019. "Peer Effects in Product Adoption." Working Paper.

Bailey, Michael, Eduardo Davila, Theresa Kuchler, and Johannes Stroebel. Forthcoming. "House Price Beliefs and Mortgage Leverage Choice." *Review of Economic Studies*.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*, 126(6).

Baker, Scott. 2018. "Debt and Response to Household Income Shocks: Validation and Application of Linked Financial Account Data." *Journal of Political Economy*, 126(4): 1504-1557.

Baugh, Brian, Itzhak Ben-David, and Hoonsuk Park. 2013. "Disentangling Financial Constraints, Precautionary Savings, and Myopia: Household Behavior Surrounding Federal Tax Returns." National Bureau of Economic Research Working Paper.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine

L. Milkman. 2015. "The Effect of Providing Peer Information on Retirement Savings Decisions." *Journal of Finance*, 70(3): 1161-1201.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2015. "Memory, Attention and Choice." Working Paper.

Bursztyn, Leonardo, Stefano Fiorin, Daniel Gottlieb, and Martin Kanz. 2016. "Moral Incentives: Experimental Evidence from Repayments of an Islamic Credit Card." National Bureau of Economic Research Working Paper.

Cadena, Ximena and Antoinette Schoar. 2011. "Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments." National Bureau of Economic Research Working Paper.

D'Acunto, Francesco, Alberto G. Rossi, and Michael Weber. 2019. "Crowdsourcing Financial Information to Change Spending Behavior." Working Paper.

DeFusco, Anthony A., Stephanie Johnson, and John Mondragon. 2019. "Regulating Household Leverage." Working Paper.

Galai, Dan and Orly Sade. 2006. "The Ostrich Effect and the Relationship between the Liquidity and the Yields of Financial Assets." *The Journal of Business*, 79(5), 2741-2759.

Gelman, Michael, Shachar Kariv, Matthew D. Shapiro, Dan Silverman, and Steven Tadelis. 2014. "Harnessing Naturally Occurring Data to Measure the Response of Spending to Income." *Science*: Vol. 345, Issue 6193, pp. 212-215.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2013. "The Determinants of Attitudes toward Strategic Default on Mortgages." *Journal of Finance*, 68(4): 14731515.

Gupta, Arpit. Forthcoming. "Foreclosure Contagion and the Neighborhood Spillover Effects of Mortgage Defaults." *Journal of Finance*.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein. 2004. "Social Interaction and Stock-Market Participation." *Journal of Finance*, 59 (1): 137163.

Howell, Sabrina. 2017. "Financing Innovation: Evidence from R&D Grants." *American Economic Review*, 107(4): 1136-1164.

Imbens, Guido and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142 (2): 615-635.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household

Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589-1610.

Kalda, Ankit. 2018. "Peer Financial Distress and Individual Leverage: Evidence from 30 Million Individuals."

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science*, 62(12): 3393-3411.

Karlan, Dean, Melanie Morten, and Jonathan Zinman. 2015. "A Personal Touch: Text Messaging for Loan Repayment." *Behavioral Science and Policy Association*, 1(2): 25-31.

Karlsson, Niklas, George Loewenstein, and Duane Seppi. 2009. "The Ostrich Effect: Selective Attention to Information." *Journal of Risk and Uncertainty*, 38(2), 95-115.

Kast, Felipe, Stephan Meier, and Dina Pomeranz. 2012. "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device." Harvard Business School Working Paper.

Kuchler, Theresa and Michaela Pagel. 2018. "Sticking to Your Plan: The Role of Present Bias for Credit Card Paydown."

Kueng, Lorenz. 2018. "Excess Sensitivity of High-Income Consumers." *The Quarterly Journal of Economics*, 133(4): 1-59.

Mader, Katharina and Alyssa Schneebaum. 2013. "The Gendered Nature of Intra-Household Decision Making in and across Europe." Working Paper.

McCrary, Justin. 2008. "Manipuation of the Running Variable in the Regression Discontinuity Design: A density Test." *Journal of Econometrics*, 142, 698-714.

Medina, Paolina. 2017. "Selective Attention in Consumer Finance: Evidence from a Randomized Intervention in the Credit Card Market."

Olafsson, Arna and Michaela Pagel. 2016. "Payday Borrower's Consumption: Revelation of Self-Control Problems?" Working Paper.

Olafsson, Arna and Michaela Pagel. 2016. "The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software." *Review of Financial Studies*.

Olafsson, Arna and Michaela Pagel. 2017. "The Ostrich in Us: Selective Attention to

Financial Accounts, Income, Spending, and Liquidity." Working Paper.

Parker, Jonathan A. 1999. "The Reaction of Household Consumption to Predictable Changes in Social Security Taxes." *American Economic Review*, 89(4): 959-73.

Parker, Jonathan A., Nicholas S. Souleles, Davis D. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, 103(6): 2530-2553.

Shoham, Aviv and Makovec Brencic. 2003. "Compulsive Buying Behavior." *The Journal of Consumer Marketing*, 20(2): 127-138.

Stango, Victor and Jonathan Zinman. 2014. "Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Bank Overdraft Fees." *Review of Financial Studies*, 27(4): 990-1030.

Stroebel, Johannes and Joseph Vavra. 2019. "House Prices, Local Demand, and Retail Prices." *Journal of Political Economy*, 127(3).

Thaler, Richard H. and Cass R. Sunstein. 2009. "Nudge: Improving Decisions About Health, Wealth, and Happiness." *Yale University Press*.

Thaler, Richard H. and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy*, vol. 122, no. 1, pt.2.

TransUnion Industry Insights Report. 2016.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S. Karlan, Richard Hornbeck, Xavier Gine, Esther Duflo, Florencia Devoto, Bruno Crepon, and Abhijit Banerjee. 2011. " Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences*, 108(5):1821-6.

Table 1: Overspending Message Thresholds

Overall Spending		2x
Wants		
	Cash	5x
	Dining Out	4x
	Shopping	4x
	Entertainment	4x
	Travel	4x
	Fees	4x
Needs		
	Groceries	6x
	Utilities	5x
	Transportation	8x
	Education	5x
	Health	6x
	Home	7x

Table 2: Summary Statistics

	N	Mean	Std	10th	25th	50th	75th	90th
				Panel A: Demographics				
Gender (Proportion of Male)	55,566	0.52	NA	NA	NA	NA	NA	NA
Age (Years Old)	55,586	36.3	13.3	22	27	32	47	57
Tenure (Years with the Bank)	55,586	9.1	3.8	2.5	7.5	12	12	12
				Panel B: App Usage				
Number of Days since Registration	55,586	127	59	39	83	135	178	202
Log-in Frequency per Month	55,586	1.9	2.7	0.3	0.5	1	2.1	4.3
Number of Messages Received per Month	55,586	4.1	1.8	2.0	3.0	4.0	5.1	6.4
Number of Messages Seen per Month	55,586	0.3	0.6	0	0	0	0.3	0.7
				Panel C: Finance				
Number of Credit Cards Registered to App	55,586	1.2	0.4	1	1	1	1	2
Number of Debit Cards Registered to App	55,586	2.0	1	1	1	2	2	3
Liquid Assets (C$)	55,586	17,244	59,665	50	853	4,029	13,853	37,656
Liquid Debt (C$)	55,586	2,457	25,785	0	130	827	2,774	6,525
Liquid Wealth (C$)	55,586	14,787	64,980	-3,739	-407	2,619	12,218	35,782
				Panel D: Monthly Spending				
Overall Spending	55,586	5,055	7,218	1,428	2,238	3,723	6,230	9,727
Cash	55,586	896	2,159	13	111	380	1,001	2,106
Dining Out	55,586	300	558	50	115	224	387	614
Shopping	55,586	922	1,430	159	319	607	1,105	1,922
Entertainment	55,586	100	327	0	5	34	102	232
Travel	55,586	316	915	0	0	45	282	834
Fees	55,586	80	741	5	21	48	80	147
Groceries	55,586	387	420	42	116	269	525	879
Utilities	55,586	364	1,076	12	87	240	483	783
Transportation	55,586	381	586	23	96	233	482	865
Education	55,586	180	700	0	0	15	97	359
Health	55,586	229	1,754	5	30	97	240	503
Home	55,586	899	5,135	0	48	286	1,100	2,380

Table 3: Adjustments by Spending Category

	Panel A: Wants & Needs Breakdown	
	Wants	Needs
Dollar	-3.82**	-4.32**
	(1.75)	(1.83)
Multiple	-0.018	-0.015
	(0.013)	(0.012)
Rolling Multiple	-0.032**	-0.023*
	(0.015)	(0.014)
N	180,523	180,523
Mean Spending (C$)	64.23	69.90

	Panel B: Wants					
	Cash	DiningOut	Shopping	Entertainment	Travel	Fees
Dollar	-1.27	0.17	-3.09***	-0.07	0.23	0.21
	(0.96)	(0.49)	(1.15)	(0.38)	(0.61)	(0.30)
Multiple	-0.006	0.002	-0.018**	0.001	0.001	0.003
	(0.007)	(0.005)	(0.009)	(0.003)	(0.004)	(0.003)
Rolling Multiple	-0.014*	0.004	-0.024**	-0.003	0.000	0.004
	(0.008)	(0.006)	(0.010)	(0.003)	(0.004)	(0.003)
N	180,523	180,523	180,523	180,523	180,523	180,523
Mean Spending (C$)	14.54	12.61	26.96	3.12	4.20	2.81

	Panel C: Needs					
	Groceries	Utilities	Transportation	Education	Health	Home
Dollar	-0.78	-0.82	-0.22	-0.26	-0.43	-1.82
	(0.70)	(0.79)	(0.69)	(0.33)	(0.55)	(1.15)
Multiple	-0.003	-0.006	0.001	0.000	-0.001	-0.006
	(0.006)	(0.006)	(0.006)	(0.003)	(0.004)	(0.007)
Rolling Multiple	-0.003	-0.008	0.002	-0.001	-0.001	-0.012
	(0.006)	(0.007)	(0.006)	(0.003)	(0.005)	(0.007)
N	180,523	180,523	180,523	180,523	180,523	180,523
Mean Spending (C$)	17.21	12.57	13.07	2.08	6.78	18.19

Note: Table illustrates the spending gap for each spending category between recipients of an overall over-spending message and non-recipients on the next day of receipt. Panel A indicates regression coefficients of Regression 1 for the Wants and Needs categories. Panels B and C indicate regression coefficients for the subcategories of Wants and Needs, respectively. Dollar represents the dollar amount spending gap, Multiple represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, and Rolling Multiple represents the dollar amount spending gap scaled by daily average spend over the last 365 days. Standard errors are in parentheses. Significance level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Spillover Effects of Nudges

	Spouse	Gender Split	
		Male Spouse	Female Spouse
Dollar	-7.10*	-10.63*	-2.88
	(3.95)	(6.37)	(4.46)
Multiple	-0.089*	-0.099*	-0.029
	(0.046)	(0.057)	(0.048)
Rolling Multiple	-0.027	-0.063	-0.012
	(0.046)	(0.099)	(0.036)
N	16,301	8,229	8,072

Note: Table illustrates the gap in overall spending from non-joint accounts between the users who themselves did not log in but whose spouse logged in and read an overspending message and the users who did not log in and whose spouse logged in but did not receive a message. The first column pools the observations of spouses. The second and third columns split the observations by gender. Male Spouse corresponds to the case where female spouses logged in but male spouses did not log in. Female Spouse corresponds to the case where male spouses logged in but female spouses did not log in. Dollar represents the dollar amount spending gap, Multiple represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, and Rolling Multiple represents the dollar amount spending gap scaled by daily average spend over the last 365 days. Standard errors are in parentheses. Significance level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Effects of Categorical Overspending Messages

	Panel A: Wants Category					
	Cash	DiningOut	Shopping	Entertainment	Travel	Fees
Dollar	-4.08	-1.75*	-8.01*	7.44	-0.43	1.09
	(14.58)	(0.97)	(4.55)	(10.57)	(19.84)	(1.30)
Multiple	0.242	-0.028	-0.086	-0.037	0.744	0.044
	(0.203)	(0.084)	(0.078)	(0.162)	(0.981)	(0.083)
Rolling Multiple	0.231	-0.250	-0.522	3.474	0.112	0.074
	(0.392)	(0.305)	(0.459)	(3.374)	(1.041)	(0.231)
N	8,009	87,965	41,681	1,756	2,504	11,895
	Panel B: Needs Category					
	Groceries	Utilities	Transportation	Education	Health	Home
Dollar	-5.93**	1.44	0.63	-6.83	1.59	-4.52
	(2.41)	(5.38)	(4.28)	(12.98)	(5.41)	(23.94)
Multiple	-0.057	0.091	0.093	-0.398	0.314	-0.302
	(0.119)	(0.202)	(0.219)	(1.030)	(0.293)	(0.331)
Rolling Multiple	0.035	0.037	1.514	-2.047	1.563	0.271
	(0.389)	(0.292)	(3.828)	(1.467)	(0.998)	(0.592)
N	45,721	11,460	25,127	728	4,767	13,090

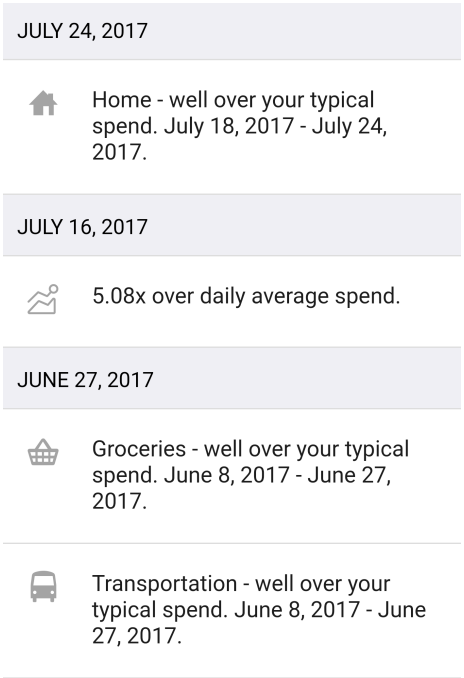
Note: Table illustrates the spending gap for each spending category between recipients of a categorical overspending message and non-recipients on the next day of receipt. Panel A indicates regression coefficients of Regression 1 for the Wants category, and Panel B indicates regression coefficients for the Needs category. Dollar represents the dollar amount spending gap, Multiple represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, and Rolling Multiple represents the dollar amount spending gap scaled by daily average spend over the last 365 days. Standard errors are in parentheses. Significance level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Robustness Checks

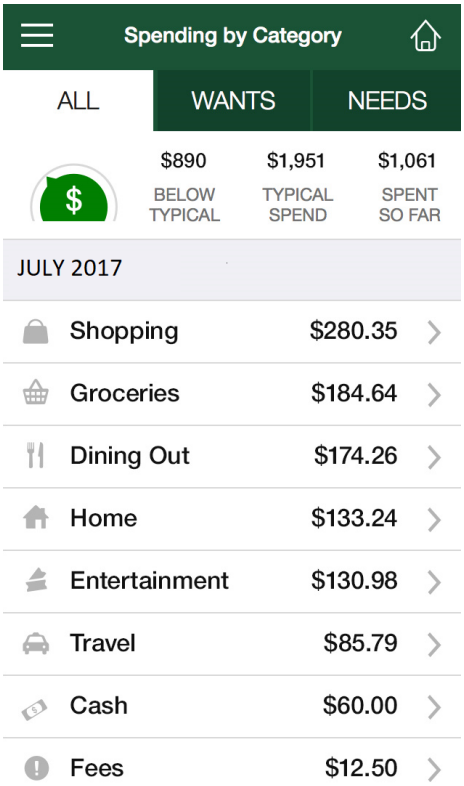
	Panel A: Inclusion of Users wihtout Credit Cards				
	Dollar	Multiple	Rolling Multiple	Number of Purchases	Spending per Purchase
Estimate	-7.59***	-0.031*	-0.054***	-0.066**	-1.62*
	(2.28)	(0.016)	(0.020)	(0.031)	(0.89)
N	237,295	237,295	237,295	237,295	237,295
	Panel B: Different Choices of Bandwidth				
	Dollar	Multiple	Rolling Multiple	Number of Purchases	Spending per Purchase
Baseline Bandwidth x 0.5	-12.89***	-0.0560**	-0.0748***	-0.0765*	-3.587**
	(3.630)	(0.0264)	(0.0272)	(0.0452)	(1.575)
Baseline Bandwidth	-8.151***	-0.0333*	-0.0535***	-0.0597*	-2.065*
	(2.540)	(0.0187)	(0.0193)	(0.0320)	(1.098)
Baseline Bandwidth x 2	-5.445***	-0.0553***	-0.0319**	-0.0644**	-1.171
	(2.049)	(0.0141)	(0.0143)	(0.0250)	(0.867)

Note: Table shows that the baseline results are robust to the inclusion of users without credit cards and the use of different bandwidths. Panel A illustrates the overall spending gap on the next day of receipt between recipients of an overall overspending message and non-recipients among the users who only have debit cards as well as the users with credit cards. The difference between Table 6 and the baseline result in Figure 4 is that in this case, some of the users do not own a credit card. Panel B illustrates the spending gap when the bandwidth is halved or doubled relative to the optimal bandwidth in the baseline. Dollar represents the dollar amount spending gap, Multiple represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Rolling Multiple represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Number of Purchases represents the gap in the number of purchasing transactions, and Spending per Purchase represents the dollar amount per-purchase spending gap. Standard errors are in parentheses. Significance level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: App Layout

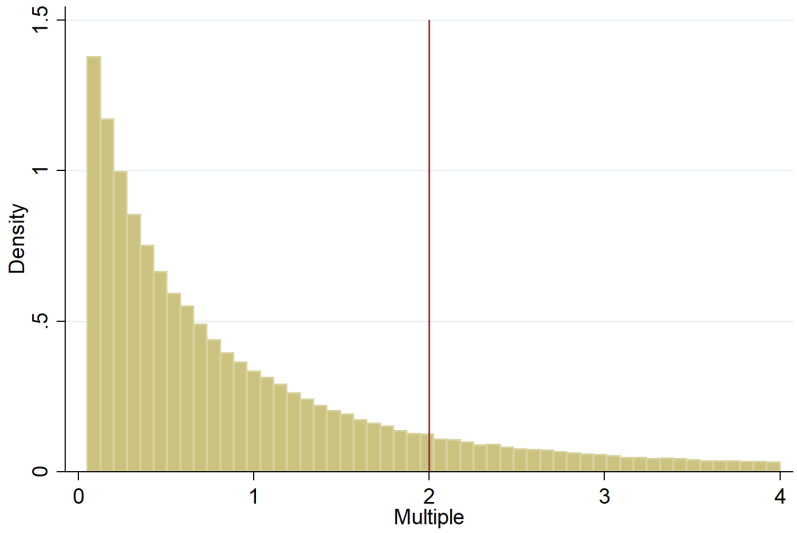


(A) Example of Overspending Messages

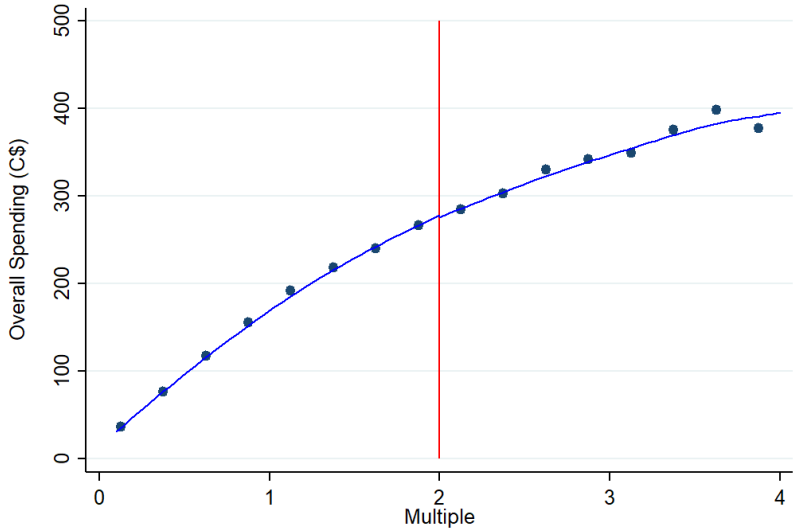


(B) Summary of Overall and Categorical Spending

Figure 2: Smooth Density of Running Variable



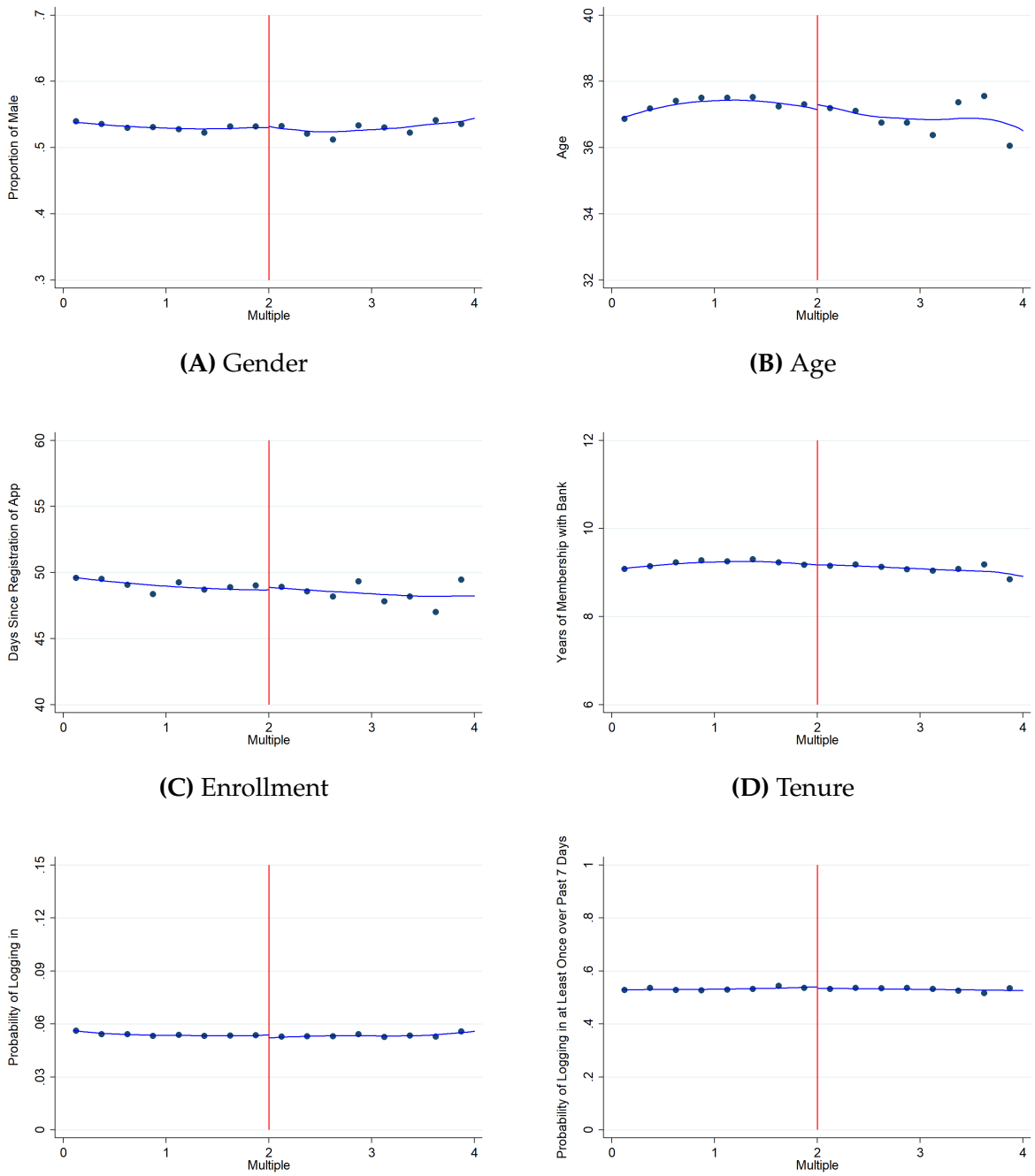
(A) Histogram of Spending Multiples



(B) Spending on Days of Messages

Note: Figure illustrates the smooth density of the running variable. Panel A illustrates the histogram of the spending multiple of app users, which is defined to be dollar amount spending on a given day scaled by the calendar date average of spending over the past 12 months. Panel B illustrates the continuity of dollar amount of spending on days on which messages were generated.

Figure 3: App User Characteristics



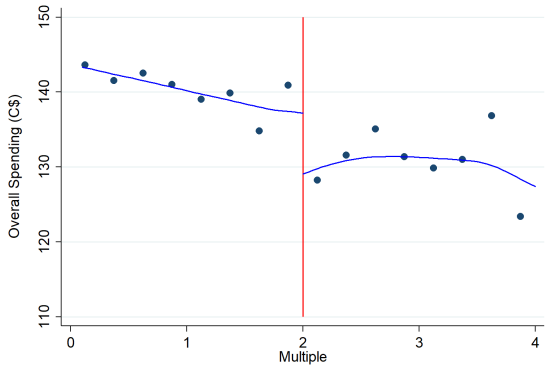
(A) Gender **(B)** Age

(C) Enrollment **(D)** Tenure

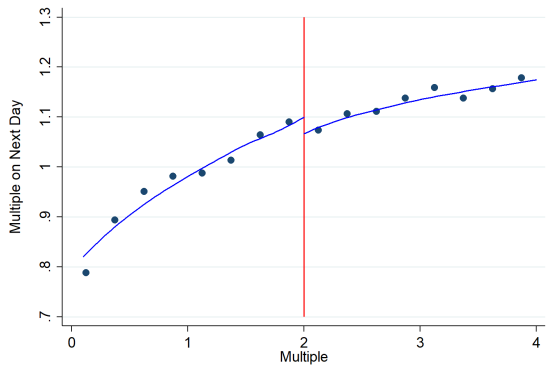
(E) Unconditional Log-in Probability **(F)** Conditional Past Log-in Probability

Note: Figure illustrates the continuous user characteristics at the threshold. Panel A represents the proportion of male users, Panel B represents the age of the users in years, Panel C represents the number of days since the registration of the app, Panel D represents the number of years of membership with the Bank, Panel E represents the unconditional probability of logging in on a day, and Panel F represents the probability of logging in at least once over the past 7 days, conditional on logging in on a day.

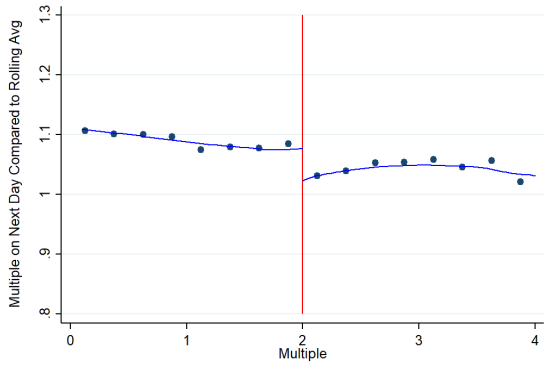
Figure 4: Baseline Results



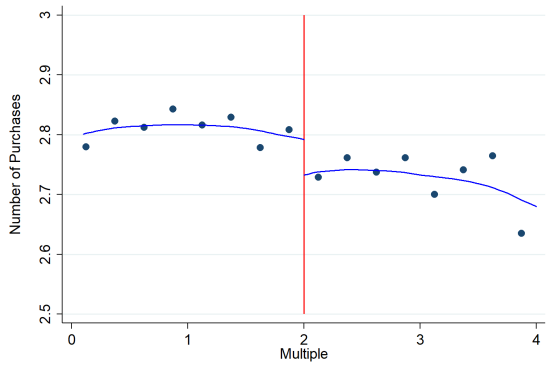
(A) Spending in Dollar Amount



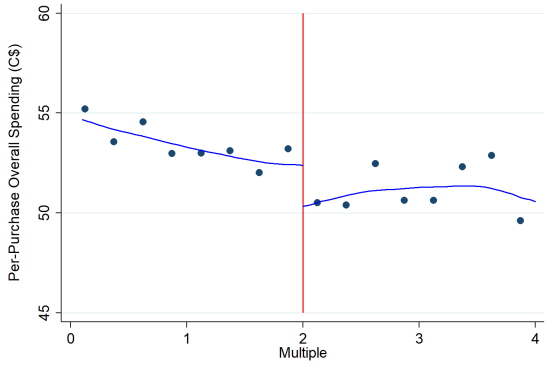
(B) Multiple



(C) Rolling Multiple



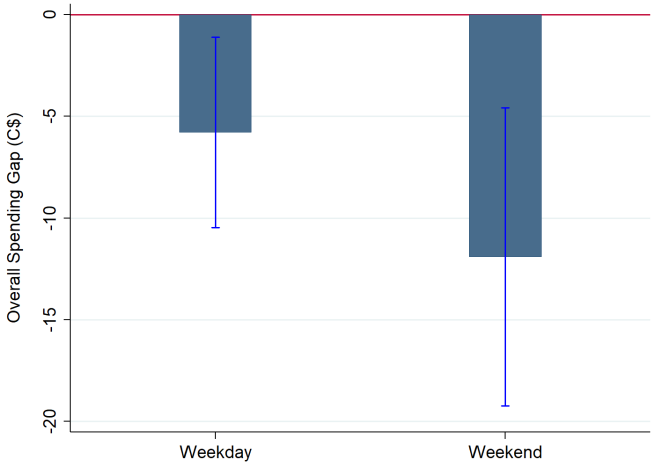
(D) Number of Purchases



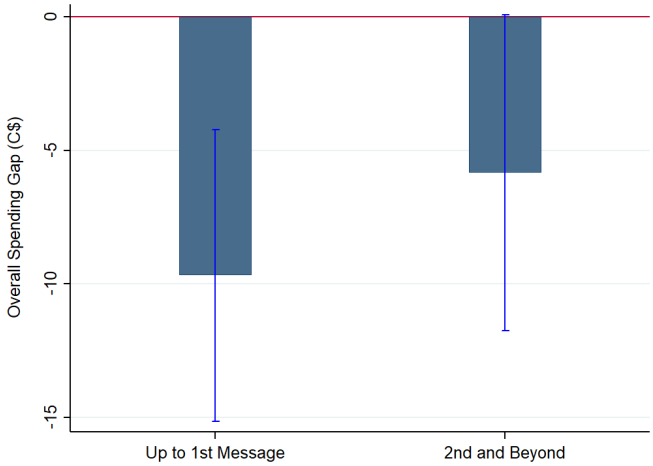
(E) Spending per Purchase

Note: Figure illustrates the overall spending gap between recipients of an overall overspending message and non-recipients on the next day of a log-in. Panel A represents the dollar amount spending gap, Panel B represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Panel C represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Panel D represents the gap in the number of purchases, and Panel E represents the dollar amount per-purchase spending gap.

Figure 5: Effectiveness at Different Time



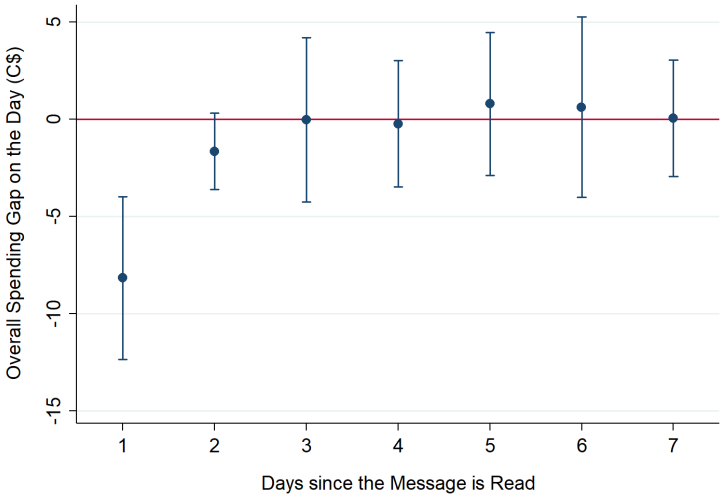
(A) Weekday vs. Weekend



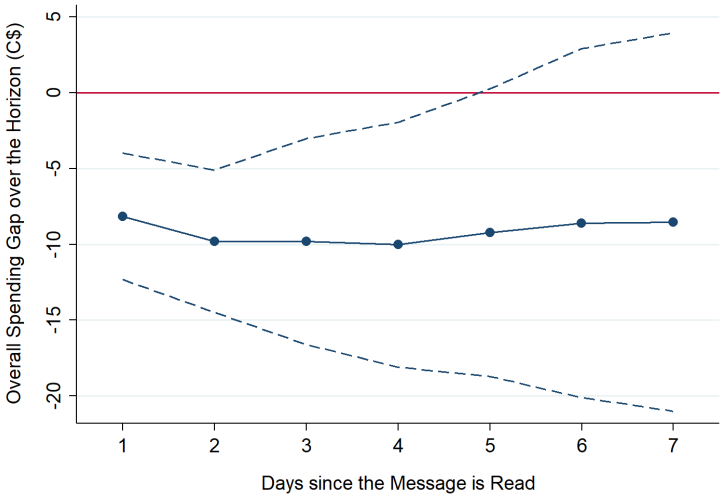
(B) First Message vs. Later Messages

Note: Figure illustrates the overall spending gap between recipients of an overall overspending message and non-recipients on the next day of a log-in at different time. Panel A represents the dollar amount spending gap on a weekday vs. a weekend. Panel B represents the dollar amount spending gap on dates up to reading the first overspending message vs. dates after the first check. The significance level is at 90%.

Figure 6: Duration of Impact



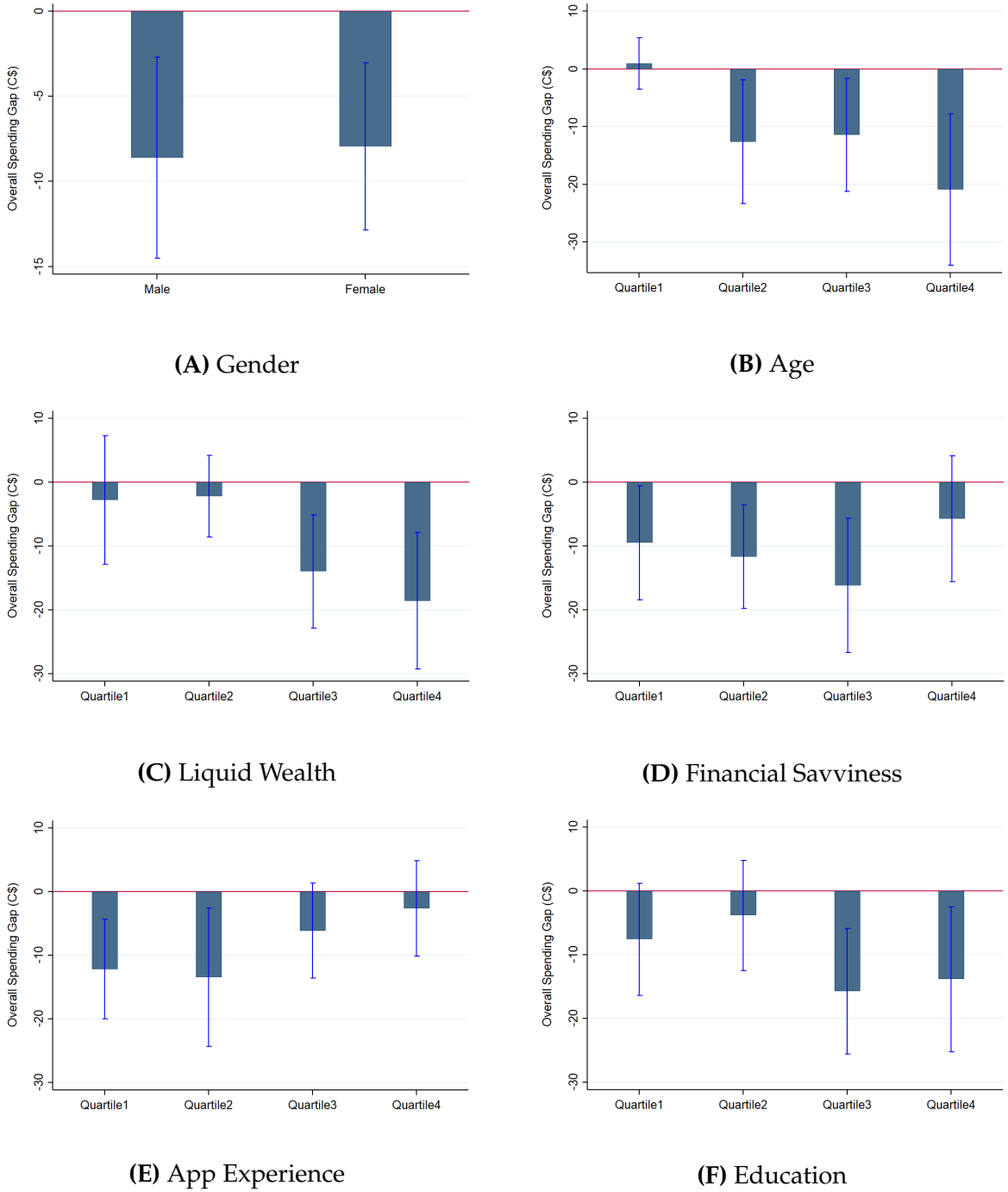
(A) Spending Gap on the Day



(B) Cumulative Spending Gap

Note: Figure illustrates the cumulative gap in overall spending between the message recipients and the non-recipients over the longer horizon. Panel A represents the dollar amount spending gap on each of the days after the receipt, and Panel B represents the cumulative dollar amount spending gap after the receipt. The significance level is at 90%.

Figure 7: Heterogeneity across Users



(A) Gender

(B) Age

(C) Liquid Wealth

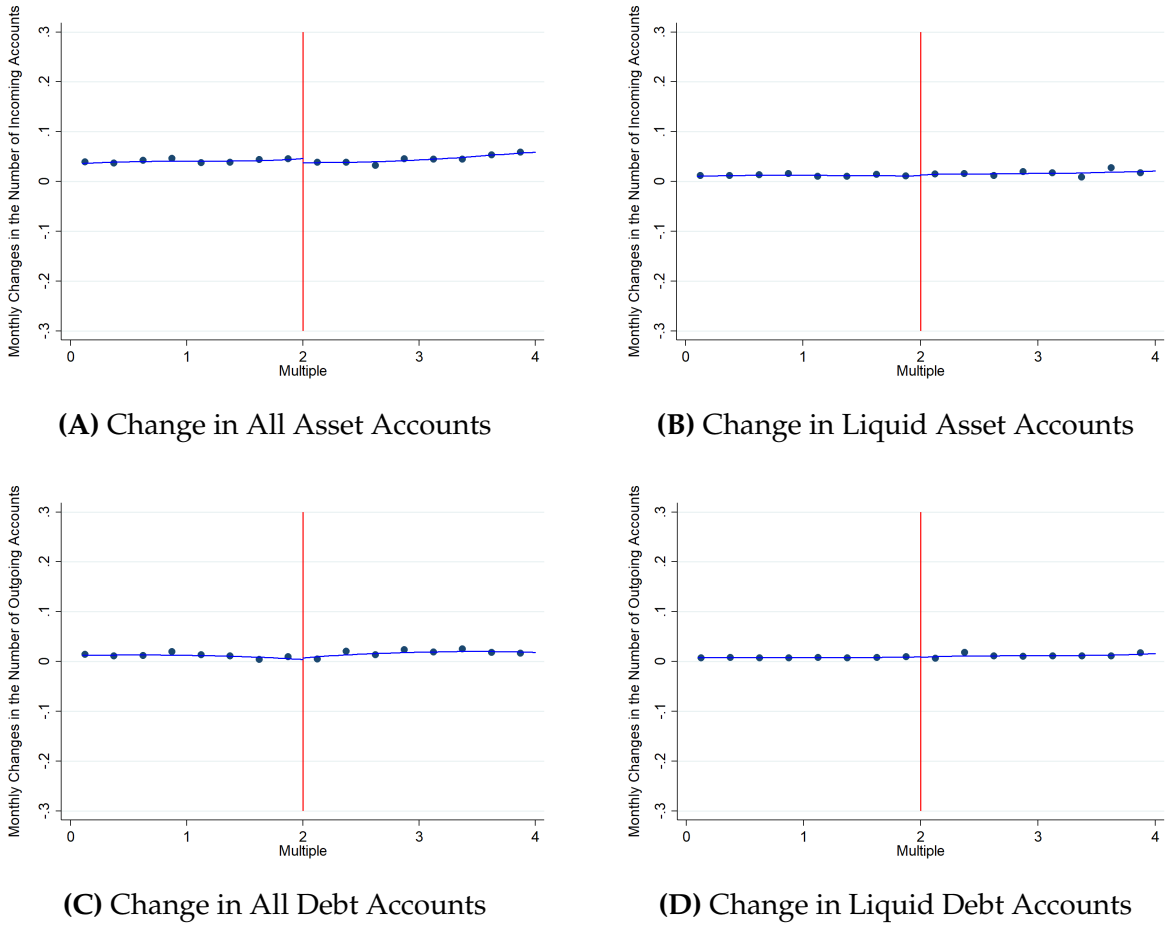
(D) Financial Savviness

(E) App Experience

(F) Education

Note: Figure illustrates the overall spending gap in dollar amounts between recipients of an overall overspending message and non-recipients for each of the subgroups sorted by their user characteristics. In Panel A, the users are sorted by gender. In Panel B, the users are sorted by age. In Panel C, the users are sorted by liquid wealth, which is defined to be the difference between the balance of checking and savings accounts and the balance of credit card accounts. In Panel D, the users are sorted by the fraction of checking account balance to the sum of balances of checking and savings accounts. In Panel E, the users are sorted by the number of days of app usage. In Panel F, the user are sorted by the fraction of population aged 15 or above with a degree or a diploma in the city they live in. The significance level is at 90%.

Figure 8: Personal Finance Management - Account Management



(A) Change in All Asset Accounts

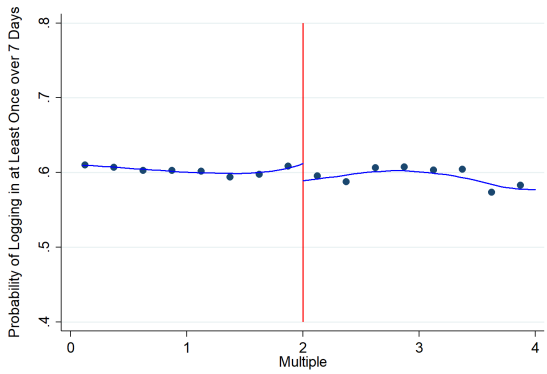
(B) Change in Liquid Asset Accounts

(C) Change in All Debt Accounts

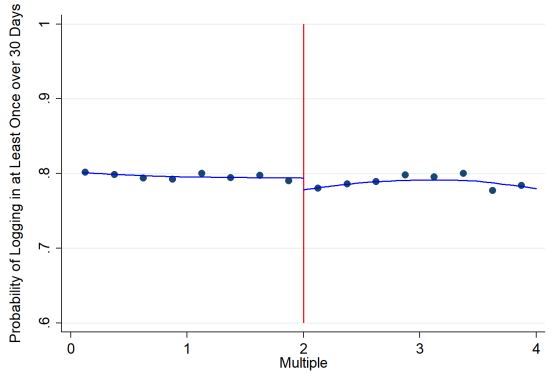
(D) Change in Liquid Debt Accounts

Note: Figure illustrates how recipients of an overall overspending message and non-recipients differ along the dimension of personal finance management after the receipt, regarding the management of accounts. Panel A represents the monthly change in the number of checking, savings, and investment accounts, Panel B represents the monthly change in the number of checking and savings accounts, Panel C represents the monthly change in the number of credit card and mortgage accounts, and Panel D represents the monthly change in the number of credit card accounts.

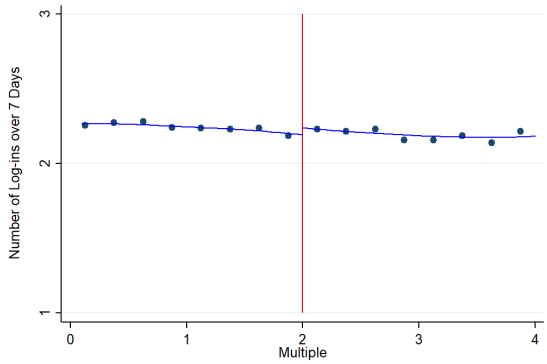
Figure 9: Personal Finance Management - Monitoring of Accounts



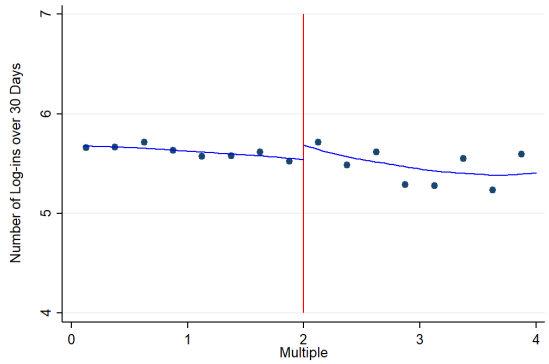
(A) Log-in Probability over 7 Days



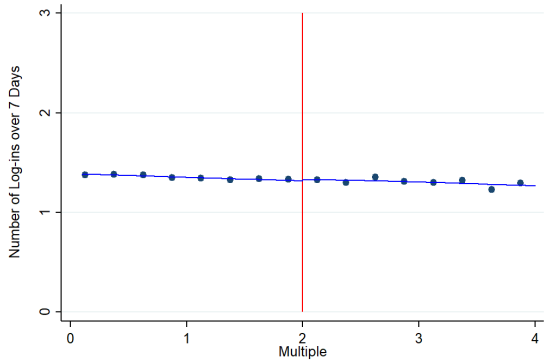
(B) Log-in Probability over 30 Days



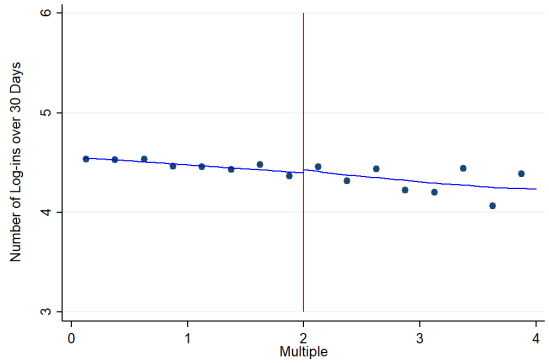
(C) Conditional Log-in Frequency over 7 Days



(D) Conditional Log-in Frequency over 30 Days



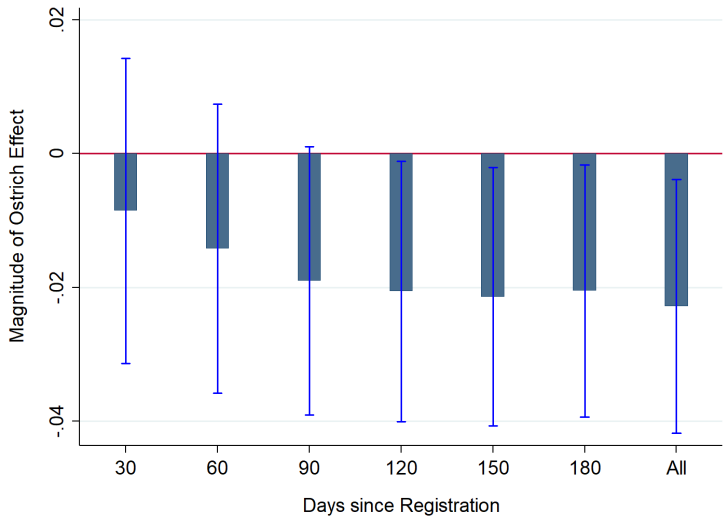
(E) Unconditional Log-in Frequency over 7 Days



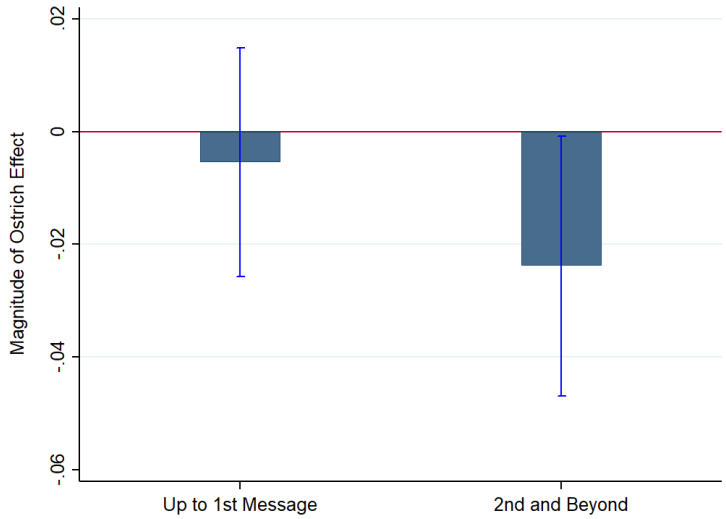
(F) Unconditional Log-in Frequency over 30 Days

Note: Figure illustrates how recipients of an overall overspending message and non-recipients differ along the dimension of personal finance management after the receipt, regarding monitoring of accounts. Panels A and B represent the probability of logging in at least once over the next 7 and 30 days, respectively, after a log-in. Panels C and D represent the number of log-ins for users who logged in at least once over the next 7 and 30 days, respectively, after a log-in on a given day. Panels E and F represent the number of log-ins over the next 7 and 30 days, respectively, after a log-in on a given day.

Figure 10: Relationship between Log-in Behavior and User Experience



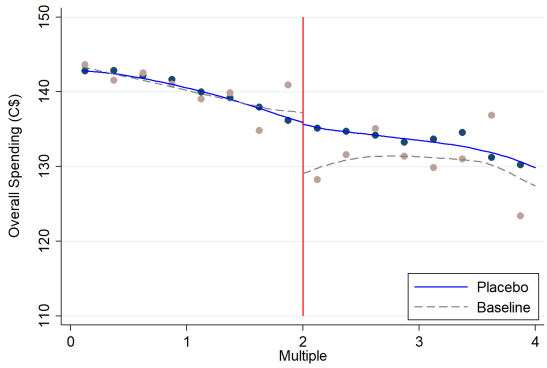
(A) Number of Days of App Usage



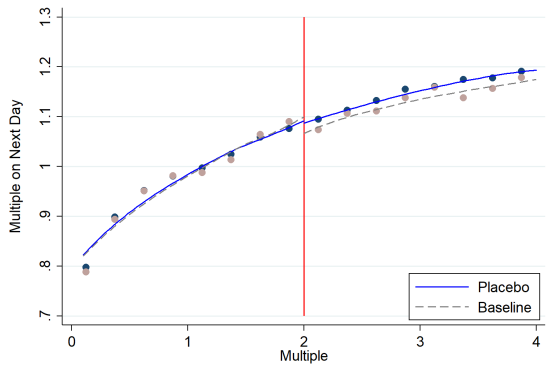
(B) First Message vs. Later Messages

Note: Figure illustrates the sensitivity of the gap in the log-in probability between overspending message recipients and non-recipients to their user experience. Panel A reports the gap in the probability of logging in at least once over the next 7 days of a log-in between the message recipients and the non-recipients for each group of users, for whom I vary the days since the registration of the app. Panel B reports the gap in the probability of logging in at least once over the next 7 days of a log-in between the message recipients and the non-recipients for dates up to the first time of seeing a message and for dates after the first message check. The significance level is at 90%.

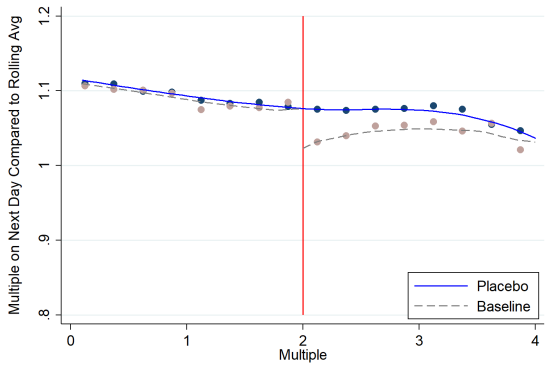
Figure 11: Placebo Tests on Days of No Log-in



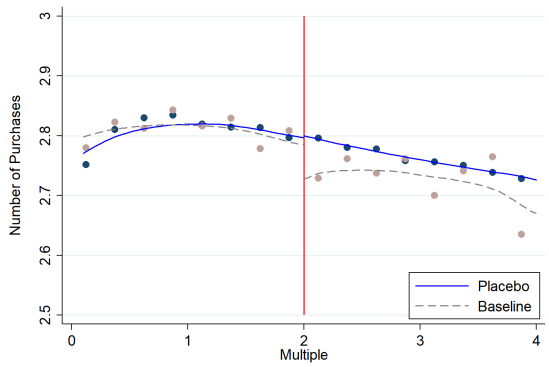
(A) Spending in Dollar Amount



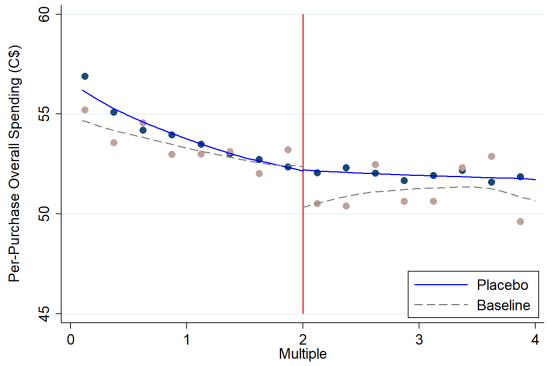
(B) Multiple



(C) Rolling Multiple



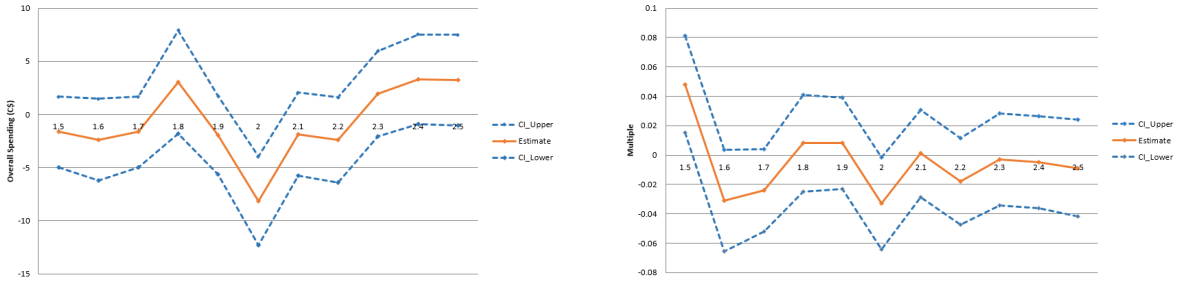
(D) Number of Purchases



(E) Spending per Purchase

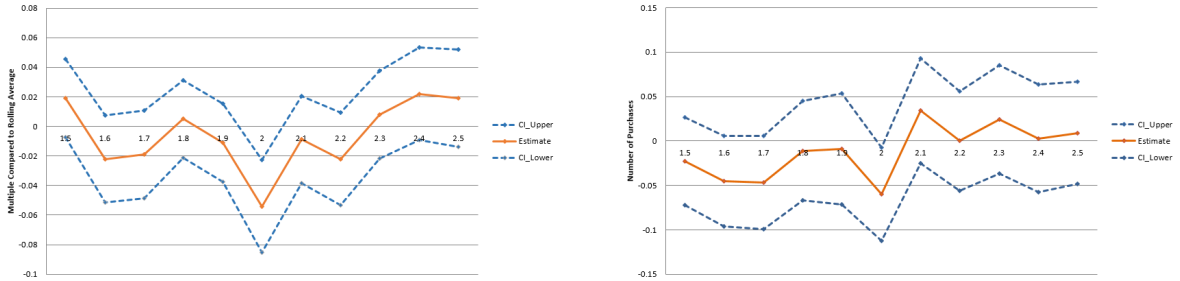
Note: Figure illustrates the overall spending gap between recipients of an overall overspending message and non-recipients on the next day of receipt when they did not log in. The blue solid lines represent spending on the following day of no log-in, and the gray dashed lines represent the spending on the following day of a log-in. Panel A represents the dollar amount spending gap, Panel B represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Panel C represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Panel D represents the gap in the number of purchases, and Panel E represents the per-purchase spending gap.

Figure 12: Test of Thresholds



(A) Spending in Dollar Amount

(B) Multiple



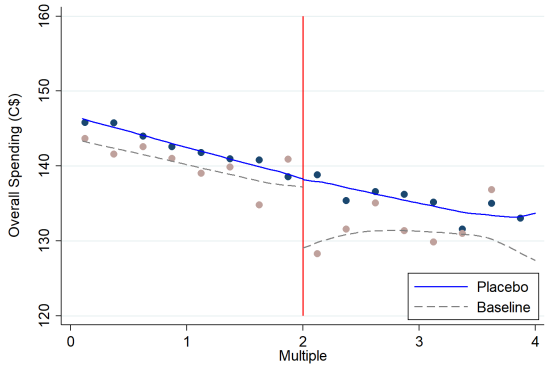
(C) Rolling Multiple

(D) Number of Purchases

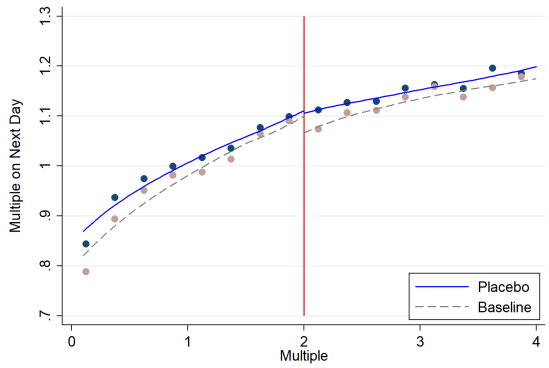
(E) Spending per Purchase

Note: Figure illustrates the overall spending gap between recipients of an overall overspending message and non-recipients when Regression 1 is run at different "false" thresholds compared to the actual threshold of 2. Panel A represents the dollar amount spending gap, Panel B represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Panel C represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Panel D represents the gap in the number of purchases, and Panel E represents the dollar amount per-purchase spending gap. The significance level is at 90%.

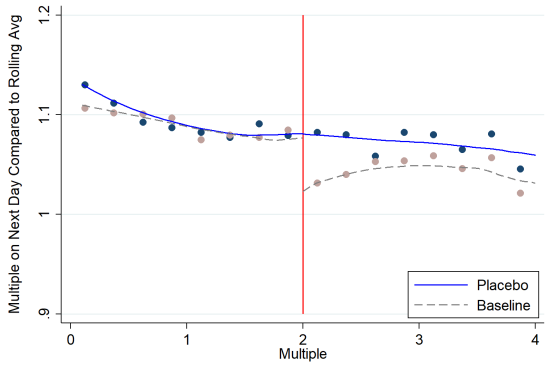
Figure 13: Placebo Tests on Users with Old App



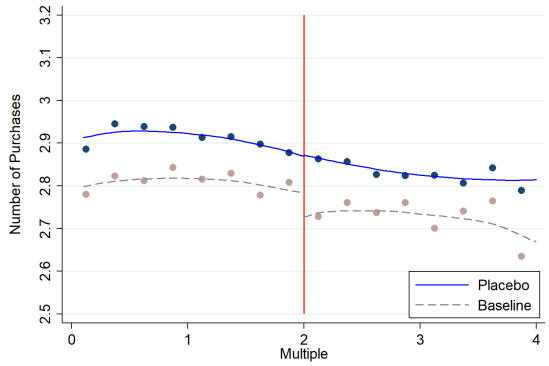
(A) Spending in Dollar Amount



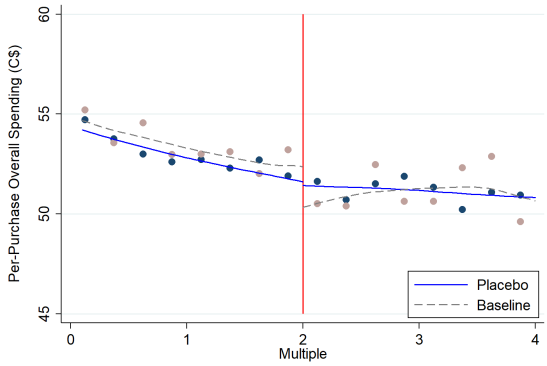
(B) Multiple



(C) Rolling Multiple



(D) Number of Purchases



(E) Spending per Purchase

Note: Figure illustrates the overall spending gap on the following day of a log-in for the users with the old version of the app, which does not have the overspending message feature. The blue solid lines represent spending on the following day in the sample period for the users with the old app, and the gray dashed lines represent the spending on the following day for the baseline user group with the newer version, which has the overspending feature. Panel A represents the dollar amount spending gap, Panel B represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Panel C represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Panel D represents the gap in the number of purchases, and Panel E represents the per-purchase spending gap.

A Appendix

Additional Results of Tests on Long-Term Effects and Placebo Tests This subsection reports additional results of tests on long-term effects and placebo tests. Appendix Figure A.1 indicates the long-term effect of an overspending message on the spending gap in terms of dollars normalized by daily average spending over the past 365 days whereas Panels A and B of Figure 6 in the main text use the dollar amount spending gap as the outcome variable. Here, the effect is statistically significant up to the second day after the receipt, and it is also stabilized after the second day.

Appendix Figure A.2 indicates the results of placebo tests on the long-term effect for the app users who did not log in. Panels A and B of Appendix Figure A.2 are counterparts to Panels A and B of Figure 6, and Panels C and D of Appendix Figure A.2 are counterparts to Panels A and B of Appendix Figure A.1. The spending gap in terms of dollar amount and rolling multiple is statistically insignificant on each of the days after receipt of a message, and as a result, the cumulative spending gap is also statistically insignificant.

Alternative Hypothesis to Ostrich Effect This subsection tests whether the rundown of liquidity can explain the message recipients' lower likelihood of logging in over the subsequent period. Panels A and B of Appendix Figure A.3 show that almost all of the recipients and non-recipients continued to use the registered cards in the subsequent 7 and 30 days, respectively, after a log-in on a given day. Hence, the lower probability of logging in did not arise from stopping using the financial products of this Bank. Furthermore, Panel C shows that the month-end liquid wealth is slightly higher, albeit statistically insignificant, for the recipients. Therefore, I can rule out the hypothesis that the recipients kept using the registered cards by running down on liquidity and borrowing with credit cards. Consequently, the lower likelihood of logging in is mainly due to the ostrich effect.

Additional Results of Tests on Effects on Future Attention This subsection reports additional results of tests on effects on future attention. Appendix Figure A.4 presents the

effects of an overspending message on log-in behavior when the number of log-ins in the subsequent period is normalized by the average monthly number of log-ins. Panels A and B of Appendix Figure A.4 are normalized versions of Panels C and D of Figure 9, and they show that among the app users who made at least one log-in over the next 7 and 30 days, the message recipients increased the normalized frequency of checking the app by 0.31% and 1.4%, respectively, although the gap is not statistically significant. Panels C and D of Appendix Figure A.4 are normalized versions of Panels E and F of Figure 9, and they show that the unconditional gap in the normalized log-in frequency is 0.28% and 1.0% higher, respectively, over the next 7 and 30 days. Hence, these results suggest that the message recipients paid slightly more attention to their finance, conditional on keeping using the app in the subsequent period whereas the overall normalized frequency was not substantially different across the message recipients and non-recipients.

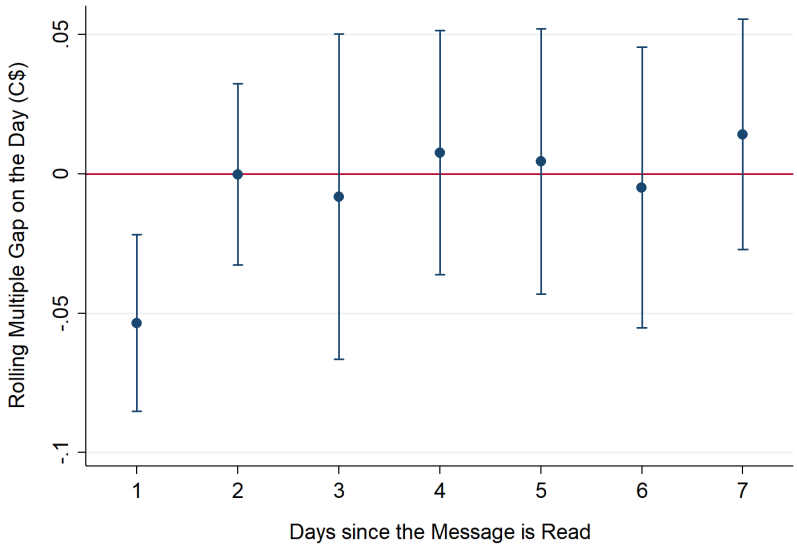
Gradual Adoption of Newer Version of the App This subsection illustrates the gradual adoption of the newer version of the app among the existing users with the old app. As indicated in Figure 13, there is not much spending gap in the entire sample period because very few of them would have downloaded the new app manually. On the other hand, with the passage of time, a greater fraction of the users with the old app would have adopted the newer version of the app, which includes the overspending message feature. Indeed, Appendix Figure A.5 shows that the spending gap started to emerge in the most recent month in the sample period among the users with the old app. Hence, it provides weak evidence that the existing users gradually adopted the newer version and responded to overspending messages on the app.

Types of Users Above and Below the Threshold This subsection tests whether the app users above the threshold and below the threshold are different types of population, not just around the cutoff, but also far away from the cutoff. Appendix Figure A.6 illustrates the the distribution of cumulative dollar amount of spending over the next 7 days of a log-in, normalized by the daily average spend for both the message recipients and non-recipients. Note that all the users whose spending multiple on a day is above 2 are

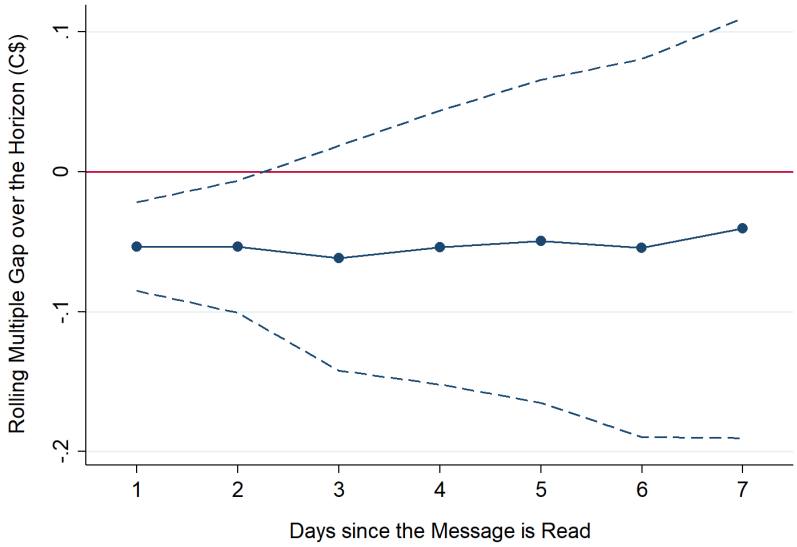
considered to be message recipients, and the rest of the users are non-recipients; in other words, I also include the users far away from the cutoff. As can be seen, the density of the rolling multiple for the message non-recipients, which is represented by the bars colored in green, and that for the message recipients, which is represented by the empty bars with solid black lines, are very similar. Indeed, when I run the two-sample Kolmogorov-Smirnov test on these two distributions, the hypothesis that there is any difference in the distributions is rejected with the p-value of 0.000012. Therefore, even though the recipients received a message due to high spending on a particular date, they are not inherently different from the non-recipients in terms of spending patterns.

In addition, the recipients and non-recipients are not inherently different in terms of log-in patterns, either. As can be seen in Figure 3, the characteristics of the users are flat across the spending multiple, and especially the unconditional and conditional log-in probabilities are also flat. It suggests that the spending multiple fluctuates across different days for individual app users, but they tend to log in to the app, regardless of whether they are spending at a high multiple of their daily average or at a low multiple of the daily average. Similarly, Figure 9 indicates the flat log-in probability and frequency across the spending multiple after the receipt for a given type of users; the users that are spending far away from the cutoff and the users that are spending near the cutoff indicate similar log-in behavior as long as they are categorized as either recipients or non-recipients. Hence, the phenomenon that the likelihood and frequency of logging in are not correlated with where their standing multiple stands persists over the longer run.

Figure A.1: Additional Results for Long-Term Effect



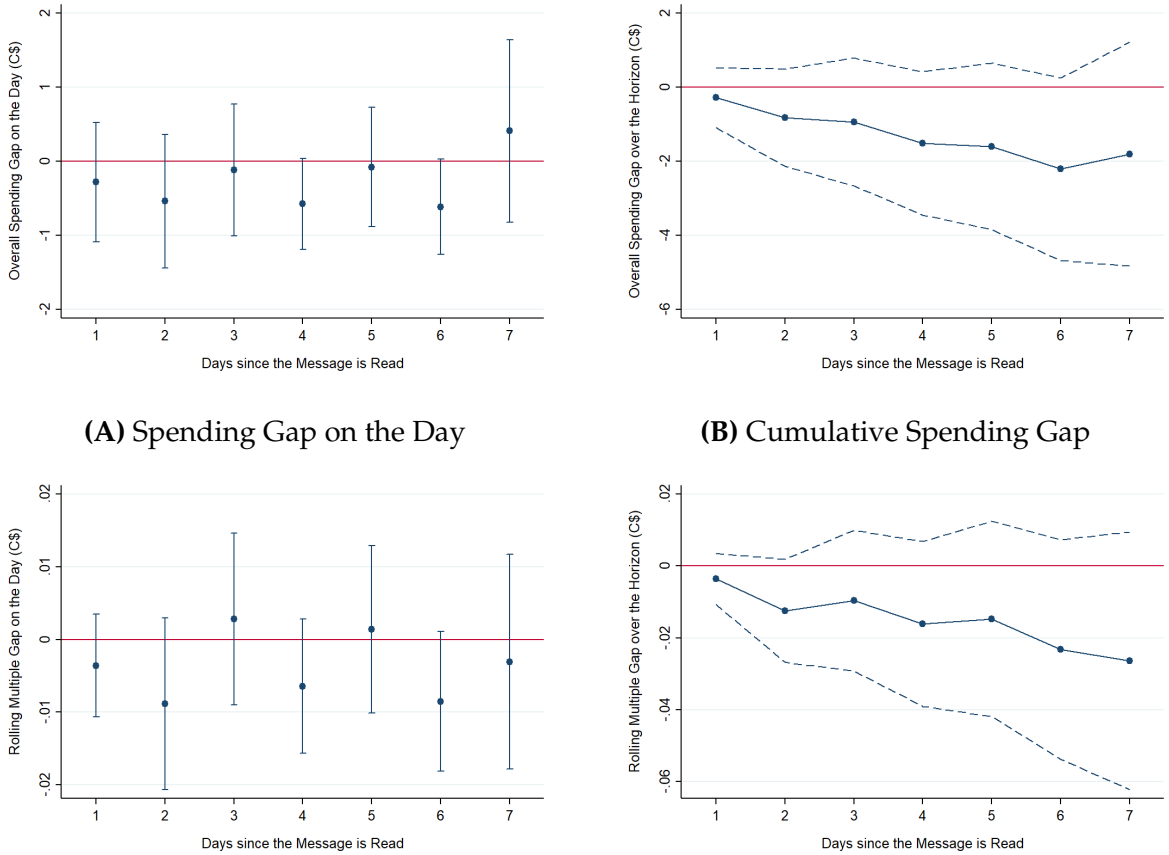
(A) Rolling Multiple Gap on the Day



(B) Cumulative Rolling Multiple Gap

Note: Figure illustrates the cumulative gap in overall spending in terms of rolling multiple between the message recipients and the non-recipients over the longer horizon. Panel A represents the dollar amount spending gap scaled by daily average spending on each of the days after the receipt, and Panel B represents the cumulative dollar amount spending gap scaled by daily average spending after the receipt. The significance level is at 90%.

Figure A.2: Placebo Tests for Long-Term Effect



(A) Spending Gap on the Day

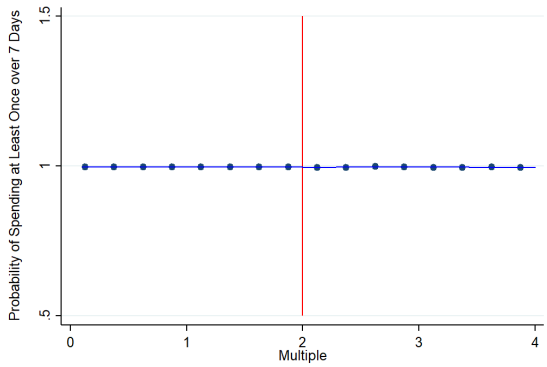
(B) Cumulative Spending Gap

(C) Rolling Multiple Gap on the Day

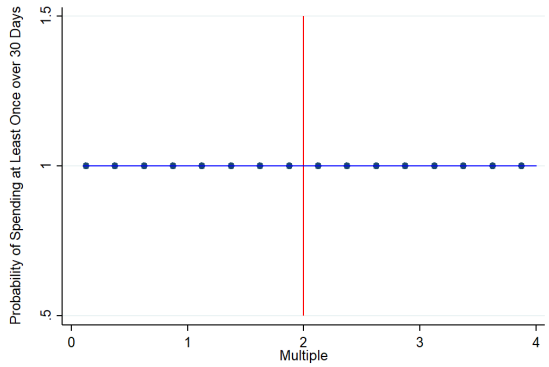
(D) Cumulative Rolling Multiple Gap

Note: Figure illustrates the cumulative gap in overall spending over the longer horizon between the message recipients and the non-recipients, when both of the groups did not log in and therefore did not see a message. Panel A represents the dollar amount spending gap on each of the days after the receipt, and Panel B represents the cumulative dollar amount spending gap after the receipt. Panel C represents the dollar amount spending gap scaled by daily average spending on each of the days after the receipt, and Panel D represents the cumulative dollar amount spending gap scaled by daily average spending after the receipt. The significance level is at 90%.

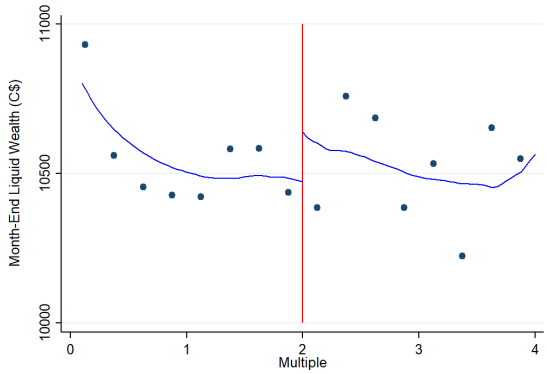
Figure A.3: Alternative Hypothesis of Lower Attention



(A) Probability of Purchase over 7 Days



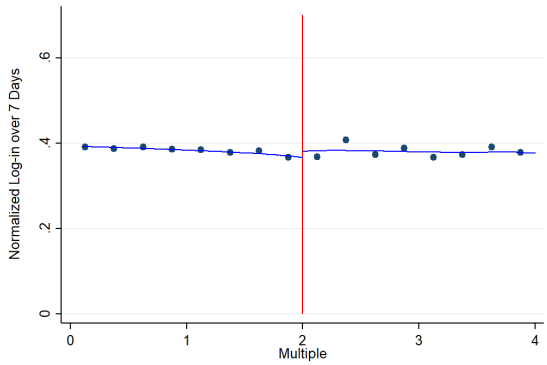
(B) Probability of Purchase over 30 Days



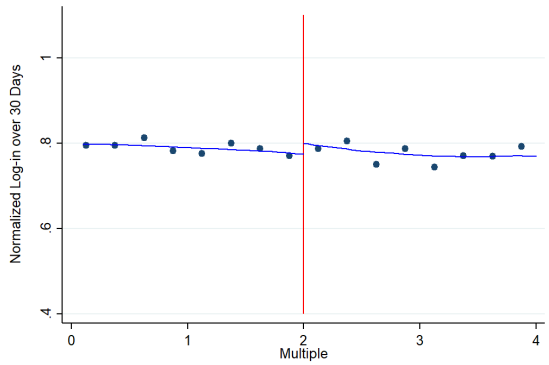
(C) Month-End Liquid Wealth

Note: Figure illustrates how recipients of an overall overspending message and non-recipients differ after the receipt, regarding monitoring of accounts. Panels A and B represent the probability of making at least one purchase over the next 7 and 30 days, respectively, after a log-in on a given day. Panel C represents the gap in the month-end liquid wealth between the recipients and non-recipients.

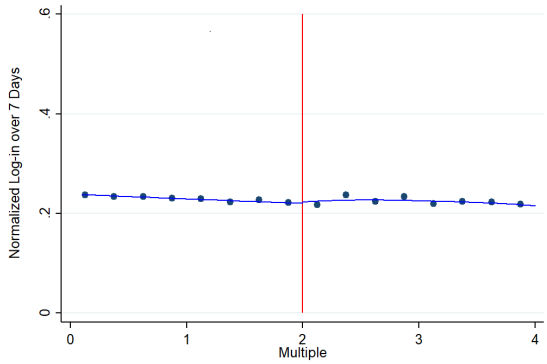
Figure A.4: Additional Results for Log-in Behavior



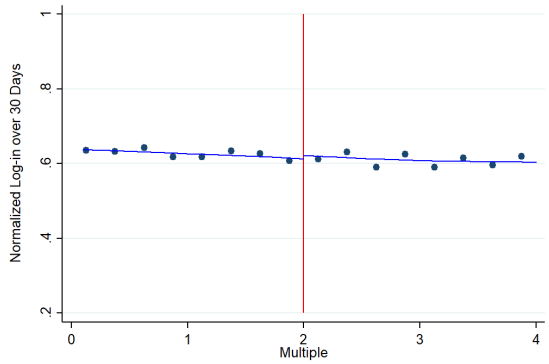
(A) Normalized Conditional Log-in Frequency over 7 Days



(B) Normalized Conditional Log-in Frequency over 30 Days



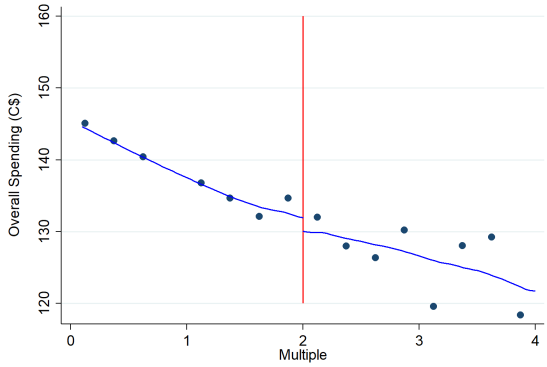
(C) Normalized Unconditional Log-in Frequency over 7 Days



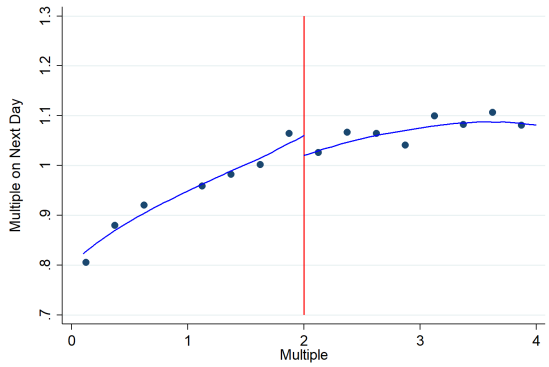
(D) Normalized Unconditional Log-in Frequency over 30 Days

Note: Figure illustrates how recipients of an overall overspending message and non-recipients differ after the receipt, regarding monitoring of accounts. Panels A and B represent the number of log-ins normalized by the monthly average log-in frequency for users who logged in at least once over the next 7 and 30 days, respectively, after a log-in on a given day. Panels C and D represent the number of log-ins over the next 7 and 30 days after a log-in on a given day, respectively, regardless of whether the users made at least one log-in.

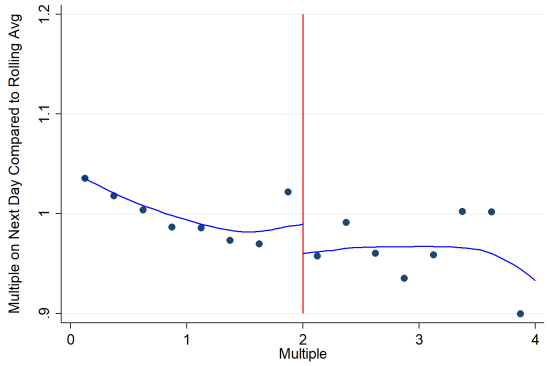
Figure A.5: Placebo Tests on Users with Old App in Most Recent Month



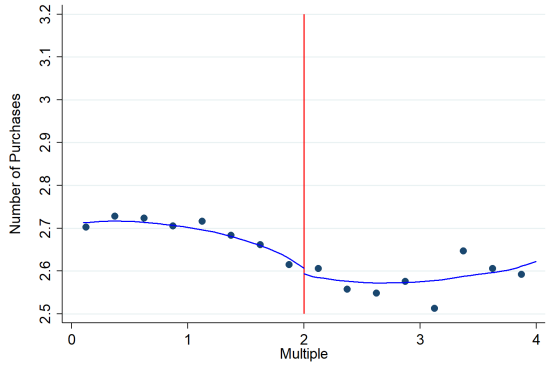
(A) Spending in Dollar Amount



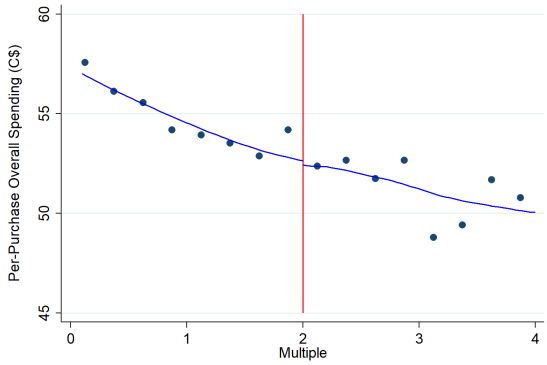
(B) Multiple



(C) Rolling Multiple



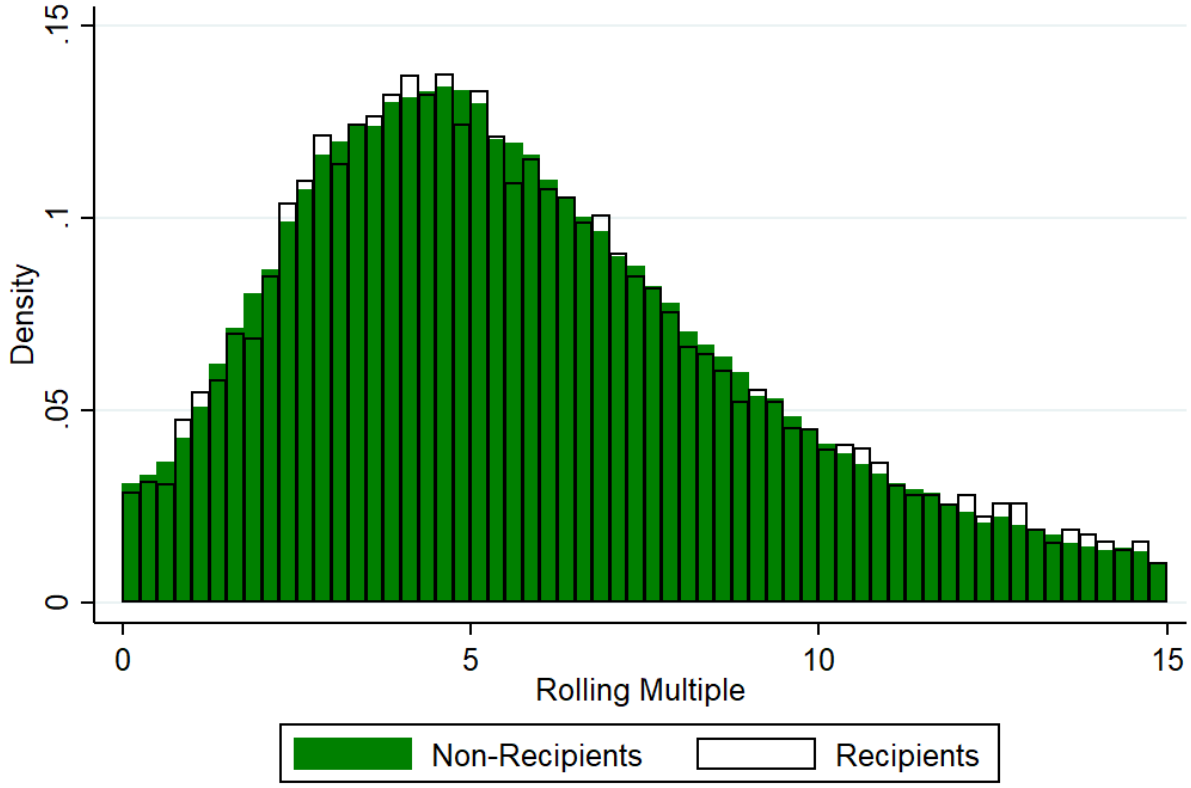
(D) Number of Purchases



(E) Spending per Purchase

Note: Figure illustrates the overall spending gap on the following day of a log-in in the most recent month of the sample period for the users with the old version of the app, which does not have the overspending message feature. Panel A represents the dollar amount spending gap, Panel B represents the dollar amount spending gap scaled by average spend of the same calendar date over the last 12 months, Panel C represents the dollar amount spending gap scaled by daily average spend over the last 365 days, Panel D represents the gap in the number of purchases, and Panel E represents the per-purchase spending gap.

Figure A.6: Distribution of Rolling Multiple over 7 Days



Note: Figure illustrates the distribution of rolling multiple over the next 7 days after a log-in. The bars colored in green indicate the density for the message non-recipients, and the empty bars with solid black lines indicate the density for the message recipients.