Why are firms slow to adopt profitable business practices? Evidence on the roles of present bias, forgetfulness, and overconfidence about memory

Pre-Analysis Plan*

Paul Gertler Sean Higgins Ulrike Malmendier Waldo Ojeda

October 6, 2020

Abstract

Why are micro, small, and medium enterprises slow to adopt profitable business practices? We test the role of three behavioral biases: present bias, limited memory, and overconfidence about memory. In partnership with a payments financial technology (FinTech) provider in Mexico, we randomly offer businesses that are already users of the payments technology the opportunity to be charged a lower merchant fee for each payment they receive. We randomly vary whether the firms face a deadline to register for this lower fee, whether they receive a reminder, and whether the reminder is anticipated (i.e., whether we tell them in advance that they will receive a reminder on a certain date).

1 Introduction

In this document we provide additional details not included in the pre-registration submitted September 28, 2020 to the AEA RCT Registry (AEARCTR-0006540). The intervention, primary and secondary outcomes, experimental design, experiment characteristics (such as sample size, how the randomization was conducted, and power calculations), and IRB details were all submitted September 28, 2020, which was one day prior to the start of the RCT. Here we include those details as well as additional details uploaded prior to accessing data on experimental outcomes from our implementing partners.

*We thank the CEGA-Visa Financial Inclusion Lab for research funding and Noah Forougi, Miguel Angel Jimenez, Alexandra Wall, and Tiange Ye for research assistance. IRB approvals: UC Berkeley IRB 2018-02-10796 and 2020-03-13091.

2 Design

2.1 Intervention

We randomly offer a cost-saving measure to firms who are already users of a financial technology (FinTech) to process electronic payments by debit and credit card: through the FinTech company that processes their payments, we offer to lower the merchant fee they are charged for each sale they make through the technology. This fee reduction intervention is offered through an email and SMS campaign, where firms can complete a short form to obtain the fee reduction. Firms are further randomized into groups that obtain different versions of the emails and SMS messages with some combination of a one-week deadline, one-day deadline, or no deadline, and an anticipated reminder, unanticipated reminder, or no reminder. The firms assigned to an anticipated or unanticipated reminder group will receive a follow-up reminder email the day before the deadline. Each of the emails will be complemented by two SMS text messages that contain similar information in a condensed format.

The initial emails are sent out on September 29, 2020 at 10am (all times are Mexico City time). The group with a one-day deadline has all of September 29 to take up the offer. The group with a one-week deadline must activate it by October 6. For groups with a reminder, the reminders are received on October 5 at 10am.

Merchants are currently charged either a 3.75% or 3.50% fee for each transaction, measured as a percent of the sale amount. Their reduced fee is randomized to be either 3.00% or 2.75%. (Thus, the fee reduction ranges from 50 basis points—for those reduced from 3.50% to 3.00%–to 100 basis points—for those reduced from 3.75% to 2.75%. Part of this reduction is randomized based on their new fee offer, and part is endogenous based on whether they currently have a 3.75% or 3.50% fee.) The lower fee lasts until March 31, 2021, after which point the firm's rate will return to their current (pre-intervention) rate. All of this information is included in the e-mail they receive.

To activate the lower fee, the merchant must click a link and fill out a short form that takes approximately one minute to complete. The email informs the user that the form will only take one minute to complete.

2.2 Primary outcomes

Our primary outcome is a dummy variable that indicates whether a firm took up the lower merchant fee. We will also use this outcome to estimate the model in Ericson (2017) to quantify the relative importance and interaction of different behavioral biases in preventing firms from adopting profitable business practices.

2.3 Secondary outcomes

Through link tracking, we will be able to tell which firms clicked the link. Thus, a secondary outcome will be whether a firm clicked the link, regardless of whether they actually completed the form or not. (Note that for the email links, we are able to track the specific user that clicked the link. For the SMS links, we can only track the number of clicks by treatment arm by day but not merge the clicks to individual users.)

We will also measure the elasticity of firm sales with respect to the experimental change in fee. Thus we will use asinh(sales) and asinh(number of transactions) as secondary outcomes, where asinh() is the inverse hyperbolic sine transformation, a log-like transformation that can deal with 0s. We will also use levels as robustness checks of these outcomes: sales (winsorized at 5%), number of transactions (winsorized at 5%), and a dummy for whether the firm made any sale in the time period.

2.4 Experimental design

Messages with deadlines will either have a one-day deadline or a one-week deadline. For the one-day deadline, they will state that the link must be clicked and form completed "with a deadline of today September 29." For the one-week deadline, they will state that the form must be completed "with a deadline of October 6." Messages with a reminder will receive a reminder on October 5; in the groups with both a deadline and reminder, the reminder will also remind the user of the deadline. Messages with an unanticipated reminder will receive a reminder to activate the lower fee on October 5. Messages with an anticipated reminder will receive a reminder to access the link one week after and will also be told in the initial email and SMS message that they will receive a reminder on October 5 if they have not yet activated the lower fee. As a result, we will have 15 treatment arms:

1. Control, no messages
2. Messages with no deadline, no reminder, 3.00% offer
3. Messages with no deadline, anticipated reminder, 3.00% offer
4. Messages with no deadline, unanticipated reminder, 3.00% offer
5. Messages with one-week deadline, no reminder, 3.00% offer
6. Messages with one-week deadline, anticipated reminder, 3.00% offer
7. Messages with one-week deadline, unanticipated reminder, 3.00% offer
8. Messages with one-day deadline, no reminder, 3.00% offer

9. Messages with no deadline, no reminder, 2.75% offer
10. Messages with no deadline, anticipated reminder, 2.75% offer
11. Messages with no deadline, unanticipated reminder, 2.75% offer
12. Messages with one-week deadline, no reminder, 2.75% offer
13. Messages with one-week deadline, anticipated reminder, 2.75% offer
14. Messages with one-week deadline, unanticipated reminder, 2.75% offer
15. Messages with one-day deadline, no reminder, 2.75% offer

2.5 Randomization method

Randomization was conducted using the R programming language to assign firms to one of the fifteen treatment arms. No rerandomization was conducted to ensure balance. While no rerandomization was conducted to ensure balance, we did test balance on the variables that we stratified on (business type and baseline sales quartiles; Tables 5–6) as well as other baseline variables we could measure in the administrative data (tax status of firm, gender of merchant, above/below median time with the technology; Tables 7–9).

2.6 Randomization unit and sample size

The randomization unit is the firm level and we use stratified randomization. There are 34,010 firms in our sample.

2.7 Stratification

We use two variables in our stratification:

1. Average monthly electronic sales from July 2019 to August 2020, or since the firm registered with the payments processing company if it was after July 2019 (4 quartiles)
2. Business type (6 categories: Beauty, Clothing, Professionals, Restaurants, Small Retailers and Other)

In total we have 21 cells, because some of the interaction cells between business type and baseline sales quartile were empty. For example, among the "Beauty" business types, every firm was in the 3rd or 4th quartile of baseline sales, so the "Beauty" $\times$ 1st quartile and "Beauty" $\times$ 2nd quartile strata were empty.

We stratified on these variables since we will test for heterogeneous treatment effects by each of these variables (sales and business type). We will also test for heterogeneity based on several other firm characteristics including tax registration status, gender, length of time using the technology, and a measure of Covid-19's impact on firm sales prior to the start of our RCT.

2.8 Experiment characteristics

Sample. The FinTech company has two types of rates: a fixed rate that is independent of the amount of monthly sales made by the firm, and a "smart rate" that is fixed if the firm makes up to 20,000 pesos in sales in a particular month, and begins decreasing if they make over 20,000 pesos in sales. To define our sampling frame we identified firms in a certain range of monthly sales. Specifically, we set a maximum of monthly sales at 20,000 pesos for firms that had the smart rate (since their status quo rate begins to fall if they have higher sales), and no maximum for firms that had the fixed rate. For the minimum, we are informed by a randomized pilot we did with 11,755 firms in May 2019 where we offered a smaller fee reduction from 3.75% to 3.50%. In that pilot, we found that the take-up rate of the lower fee was increasing in baseline sales (see Table 1), and that the elasticity of sales through the technology with respect to the fee (which is what our partner is interested in, as offering a lower fee may or may not be profitable depending on the elasticity) was statistically significant only for the fourth quartile in baseline sales (see Tables 2 and 3). Thus, we use the 75th percentile of baseline monthly sales from firms in that randomized pilot (i.e., the minimum sales among firms in the fourth quartile) as the minimum.

As a result, our sampling frame for this experiment consists of all firms who are users of this FinTech that had (i) fixed rate and August 2020 sales greater than 1400 pesos with no maximum or (ii) smart rate and August 2020 sales greater than 1400 pesos and less than 20,000 pesos. The sampling frame was made up of 34,010 users.

Among these 34,010 firms, 4,010 firms were randomly assigned to the pure control group that receives no offer; the size of this arm was based on the FinTech company's desire to cap the number of firms receiving an offer at 30,000. The remaining 30,000 firms were assigned to one of the fourteen other groups. The firms were first randomly assigned to one of seven groups combining deadlines and reminders: (i) no deadline, no reminder; (ii) no deadline, anticipated reminder; (iii) no deadline, unanticipated reminder; (iv) one-week deadline, no reminder; (v) one-week deadline, anticipated reminder; (vi) one-week deadline, unanticipated reminder; (vii) one-day deadline, no reminder. The sample size in each of these seven groups receiving an offer was determined based on power calculations using the results from our May 2019 randomized pilot, as we describe in more detail below. As described above, this randomization is stratified on business type and business sales quartiles. Then among firms assigned to one of these 7 groups receiving an offer, we cross-randomize the rate they are offered, either 3.00% or 2.75%; this cross-randomization is stratified based on initial group assignment, as well as business type and baseline sales quartiles. This leaves us with 15 total treatment arms.

Sample size by treatment arm.

1. Control (no messages): 4010

2. No deadline, no reminder with 3.00% offer: 2230
3. No deadline, anticipated reminder with 3.00% offer: 1838
4. No deadline, unanticipated reminder with 3.00% offer: 2227
5. One-week deadline, no reminder with 3.00% offer: 2311
6. One-week deadline, anticipated reminder with 3.00% offer: 1755
7. One-week deadline, unanticipated reminder with 3.00% offer: 2316
8. One-day deadline, no reminder with 3.00% offer: 2324
9. No deadline, no reminder with 2.75% offer: 2230
10. No deadline, anticipated reminder with 2.75% offer: 1838
11. No deadline, unanticipated reminder with 2.75% offer: 2229
12. One-week deadline, no reminder with 2.75% offer: 2311
13. One-week deadline, anticipated reminder with 2.75% offer: 1753
14. One-week deadline, unanticipated reminder with 2.75% offer: 2316
15. One-day deadline, no reminder with 2.75% offer: 2322

Power calculations. We conduct two types of power calculations: first, we use observed effect sizes from our pilot in May 2019, in addition to theoretical predictions from Ericson (2017) for the cases we did not test in our pilot, to determine the minimum sample size needed in each arm for each pairwise comparison of interest. We use these power calculations to determine how to optimally allocate our fixed total sample (30,000 firms to be allocated to arms 2 through 15). Second, after conducting the randomization, we take as given the sample size in each arm and calculate the minimum detectable effect for each pairwise comparison of interest. Note that for the measuring the effect of a deadline, anticipated reminder, and unanticipated reminder, in the main results we will pool the 2.75% and 3.00% cross-randomized groups, as both of these are relevant potential fee reductions. Thus, the treatment effects will be a weighted average of the effect of a deadline or reminder with a 2.75% fee offer and the effect of a deadline or reminder with a 3.00% fee offer.

Minimum sample size per arm: In the pilot we estimated the take-up rates of merchants with a 3.75% merchant fee to accept a 3.50% merchant fee reduction offer. Compared to the control group that received no offer, we estimated treatment effects of merchants accepting the offer with an email with no deadline and with a 24-hour deadline. We did not test different length deadlines, as we will do in the RCT. Thus, for the expected effect of both a one-week and one-day deadline we use the effect from our pilot of this 24-hour deadline.

After three days, we also sent unanticipated reminders to merchants, but did not randomize these reminders. We thus estimate the expected effect of a reminder by comparing take-up right before the reminder to 24 hours after the reminder. With these groups, we use the equation for minimum sample size from equation 8 in List et al. (2011), plugging in the observed P0 and P1 from the experiment to calculate n*. Doing so, we calculate the following minimum sample sizes per arm for each pairwise comparison of interest, where take-up in the first group listed in each comparison is P0 and take-up in the second group listed is P1.

Because we will be pooling the 2.75% and 3.00% fee cross-randomization within each treatment arm defined by the combination of deadline type and reminder type, for the description below we consider the following 8 (pooled) treatment arms:

- T1) Control
- T2) No deadline, no reminder
- T3) No deadline, anticipated reminder
- T4) No deadline, unanticipated reminder
- T5) One-week deadline, no reminder
- T6) One-week deadline, anticipated reminder
- T7) One-week deadline, unanticipated reminder
- T8) One-day deadline, no reminder

The take-up rates in the pilot by pairwise comparison and the necessary sample size to detect such differences are as follows (using the treatment group numbers above). For the reminders which were not randomized in our pilot, we estimate P1 as cumulative take-up of the offer 24 hours after the reminder was sent and P0 as cumulative take-up immediately before the reminder was sent.

- Control (no offer) vs. No Deadline (T1 vs. T2)
  - P0 = 0.01, P1 = 0.18. Minimum sample size per arm to detect this effect = 46
- Control (no offer) vs. Deadline (T1 vs. T5)

- P0 = 0.01, P1 = 0.28. Minimum sample size per arm to detect this effect = 26

- No Deadline vs. Deadline (T2 vs. T5)
  - P0 = 0.18, P1 = 0.28. Minimum sample size per arm to detect this effect = 277
- No Deadline vs. No Deadline with Unanticipated Reminder (T2 vs. T4)
  - P0 = 0.10, P1 = 0.14. Minimum sample size per arm to detect this effect = 1472.
- Deadline vs. Deadline with Unanticipated Reminder (T5 vs. T7)
  - P0 = 0.20, P1 = 0.24. Minimum sample size per arm to detect this effect = 1529.

In the pilot there was no anticipated reminder treatment group. To obtain an estimate of expected effect size of the anticipated reminder and perform power calculations, we benchmark results from the pilot with model simulations based on the model in Ericson (2017) assuming standard magnitudes for present-bias and forgetfulness from the literature, and assuming full naïveté about present-bias but accurate beliefs about memory (using Ericson's terminology, beta = 0.9, $\widehat{\beta}$ = 1, ρ = 0.95, $\widehat{\rho}$ = 0.95). The model simulations predict ratios of the difference in take-up between the groups with unanticipated and anticipated reminders over the difference in take-up between the groups with unanticipated and no reminders. In our pilot we use the difference in take-up between groups with unanticipated and no reminders and then apply the ratio to scale an estimated take-up rate for groups with an anticipated reminder. The ratio from the model is measured in the period when the reminder is sent, one period before the deadline. With these simulated take-up rates of groups with an anticipated reminder, we get a treatment effect ratio of 1.23, which means the necessary sample size to detect differences in relevant pairwise comparisons are:

- No Deadline with Unanticipated Reminder vs. No Deadline with Anticipated Reminder (T4 vs. T3)
  - P0 = 0.14, P1 = 0.18. Minimum sample size per arm to detect this effect = 1222.
- Deadline with Unanticipated Reminder vs. Deadline with Anticipated Reminder (T7 vs. T6)
  - P0 = 0.24, P1 = 0.29. Minimum sample size per arm to detect this effect = 1152.

To obtain a minimum sample size per arm for our study, we select the largest sample size needed for each group depending on its relevant pairwise power calculations above. For group 8, we assume we need the same number of observations as for group 5, since both include a deadline and no reminder. We calculate the following minimum sample sizes per arm to detect the expected effect sizes based on our pilot and simulations:

- T1) Control: 46
- T2) No deadline, no reminder: 1472
- T3) No deadline, anticipated reminder: 1222
- T4) No deadline, unanticipated reminder: 1472
- T5) One-week deadline, no reminder: 1529
- T6) One-week deadline, anticipated reminder: 1152
- T7) One-week deadline, unanticipated reminder: 1529
- T8) One-day deadline, no reminder: 1529

Thus, we need in total 9,951 observations across all treatment arms to statistically detect the expected differences in take-up between treatment groups of interest, based on outcomes in our pilot and simulations of the Ericson (2017) model. As our available sample size for the experiment is 34,010 firms, with 4,010 assigned to control per our partner's preferences and the remaining 30,000 available to allocate between T2 and T8, which is much larger than the needed 9,951, we adjust sample sizes of each treatment arm T2-T8 proportionally to arrive at the sample sizes per arm in our study shown under "Sample size by treatment arm."

Minimum Detectable Effect. For each pairwise comparison of interest between the arms T1 through T8 and the allocation of merchants across treatment arms listed above under "Sample size by treatment arm", we can calculate the minimum detectable effect we are powered to detect. To do so, we use the formula for minimum detectable effect from equation 5 in List et al. (2011), where we can plug in $\sigma_0^2 = P0 * (1–P0)$ and $\sigma_1^2 = P1 * (1–P1)$ We thus plug in n_0 and n_1 as the sample size for each arm of the comparison, plug in $P0$ from our pilot data for the relevant group, and solve for $P1$ and thus for the MDE $= P1–P0$. We express the MDE in percentage points and also divide by the standard deviation to obtain the MDE in standard deviations. The standard deviation we divide by is the standard deviation for all merchants in the comparison of interest, i.e. $\sqrt{\bar{p}(1-\bar{p})}$, where $\bar{p} = (P0 + P1)/2$.

- Effect of the offer (conditional on no deadline).
  - T2 vs T1 : MDE = 0.73 percentage points. MDE (SD) = 0.07 SD.
- Effect of the offer (conditional on deadline).
  - T5 vs T1 : MDE = 0.72 percentage points. MDE (SD) = 0.07 SD.

- Effect of a deadline (conditional on no reminder).
  - T5 vs T2 : MDE = 2.32 percentage points. MDE (SD) = 0.06 SD. (Note the MDE in SD went down relative to the last comparison even though the MDE in percentage points went up, since $\bar{p}$ is higher and hence SD is higher for this comparison.)
- Effect of an unanticipated reminder (conditional on no deadline).
  - T4 vs T2 : MDE = 1.88 percentage points. MDE (SD) = 0.06 SD.
- Effect of an anticipated reminder (conditional on reminder, no deadline).
  - T3 vs T4 : MDE = 2.22 percentage points. MDE (SD) = 0.06 SD.
- Effect of an unanticipated reminder (conditional on deadline).
  - T7 vs T5 : MDE = 2.37 percentage points. MDE (SD) = 0.06 SD.
- Effect of an anticipated reminder (conditional on reminder, deadline).
  - T6 vs T7 : MDE = 2.74 percentage points. MDE (SD) = 0.06 SD.

3 Analysis Plan

3.1 Reduced form

Our experimental design allows to compare in reduced form the effects of a reminder or deadline on the probability of taking up the lower fee. Our primary results will be from the following regression:

$$y_i = \lambda_{s(i)} + \sum_{k=2}^{8} \beta_k T_i^k + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest, $\lambda_{s(i)}$ are strata fixed effects for the 21 stratification blocks defined above (which also absorb the constant), T_i is a vector of indicator variables denoting treatment assignment, and ε_i are heteroskedasticity-robust standard errors (not clustered since the randomization unit is the individual firm).

Note that for one of our secondary outcomes, clicking the SMS link, we are unable to merge with data other than treatment assignment since the SMS link clicks are tracked separately; thus for that outcome we will estimate (1) without strata fixed effects and with a constant α in their place.

Our main estimates of the effect of a reminder, deadline, and anticipated vs. unanticipated reminder will pool together the fee groups into the 8 treatment arms and will be estimated as follows:

- To estimate the effect of a deadline conditional on having an anticipated reminder, the comparison of interest will be between the coefficients on T5 and T3
- To estimate the effect of a deadline conditional on having an unanticipated reminder, the comparison of interest will be between the coefficients on T7 and T4
- To estimate the effect of a deadline conditional on having no reminder, the comparison of interest will be between the coefficients on T5 and T2
- To estimate the effect of an anticipated reminder conditional on not having a deadline, the comparison of interest will be between T3 and T2
- To estimate the effect of an anticipated reminder conditional on having a deadline, the comparison of interest will be between T6 and T5
- To estimate the effect of an unanticipated reminder conditional on not having a deadline, the comparison of interest will be between T4 and T2
- To estimate the effect of an unanticipated reminder conditional on having a deadline, the comparison of interest will be between T7 and T5

We will estimate the above (i) for applications aggregated over the 8-day period between the initial messages and the deadline; (ii) for overall applications from the beginning of the experiment to the latest date in which we receive data from the FinTech company (beyond one week); (iii) we will estimate daily effects over the eight-day experiment period by estimating separate regressions for each day where the outcome is a dummy for taking up the lower fee; and (iv) weekly effects by estimating separate regressions for each week since the beginning of the experiment. On the first day of the experiment, we use the period from 10:00 AM to midnight as the day to compute the outcomes, and for every other day, the day ends at midnight. The reasoning for this is that the initial messages were sent out at 10 AM on September 29th and the offer is available till midnight on October 6th. For items iii and iv we will use two specifications: one with *cumulative* take-up by that day or week and another with take-up just in that day or week as the outcome. The cumulative application indicator for a particular day will be equal to 1 if that firm took up the fee reduction that day or any of the previous days, and similarly for the weekly cumulative application indicator. The daily application indicator will be an indicator equal to 1 if the firm took up the fee reduction on that specific day, and similarly for the weekly application indicator.

Finally, to view take-up over time across treatment arms we will graph cumulative distribution functions of take-up by treatment arm. We will also estimate a proportional hazards model where the dependent variable is the hazard rate of our outcomes of interest.[1] Coefficients on the treatment assignment variables when estimating this model will reveal how each treatment varies the proportion of take-up over time.

3.2 Heterogeneity

We will estimate

$$y_i = \lambda_{s(i)} + H_i + \sum_{k=2}^{8} \left[\beta_k T_i^k + \gamma_k T_i^k \times H_i \right] + \varepsilon_i \tag{2}$$

where H_i is a vector of dummy variables that measure heterogeneity and γ_k is the vector of coefficients of interest. In the specifications where we estimate heterogeneity based on the variables for which we stratified, there will no simple (non-interacted) H_i term in (2) because it is absorbed by the strata fixed effects (since each strata is an interaction of the variables we will use for heterogeneity tests). Our main heterogeneity tests will include the variables we stratified on and other variables. The variables we stratified on are:

- Baseline sales quartiles (4 categories). (For increased power, we will also use above/below-median sales in a separate regression.)
- Business type (6 categories).

The additional variables are:

- Tax registration status (3 categories: individual, self-employed or limited company)
- Impact of Covid-19 on sales (3 categories: above median impact, below median impact, or undefined impact)
- Gender (3 categories for the gender of the firm's owner: male, female or missing)
- Length of time using the technology (2 categories: new users (below median length of time using the technology) and older users (above median length of time))
- New fee (3.00% or 2.75%)
- Expected gain: we compute estimated gain of the fee reduction by multiplying baseline sales * fee reduction * 6 months; we will construct quartiles and an above/below-median split based on these variables and use each in heterogeneity tests.

We will estimate (2) for the primary and secondary outcomes.

[1] The specification is: $\lambda(t|T_i) = \lambda_0(t)exp(\beta T_i)$ where $\lambda(t|T_i)$ is the hazard rate of our outcomes of interest, $\lambda_0(t)$ is the baseline hazard function and T_i is a vector of indicator variables denoting treatment assignment.

3.3 Structural model

We will follow Ericson (2017) in our estimation method to structurally estimate forgetfulness and present-bias parameters based on the experiment outcomes. In the model, an agent makes a decision to perform a task that is beneficial in the future but has an immediate cost.

The agent has present-biased preferences and possibly naïveté: $U = u_0 + \beta\left(\sum_{t=1}^{\infty}\delta^t u_t\right)$, where δ is the discount factor, β is the present bias parameter, and the agent has beliefs $\hat{\beta}$ about β. The model also incorporates imperfect memory. There is a probability of remembering the task in period t conditional on remembering it in period $t-1$, measured by the parameter ρ_t (with $\rho_0 = 1$). The agent will only be able to perform the task if they remember it. Agents have beliefs $\hat{\rho}_t$ and are overconfident about their memory if $\hat{\rho}_t > \rho_t$. Reminders about the task raise ρ_t in the period they are sent. However, only an anticipated reminder in period t (when the agent is told about a reminder that it will receive in a future period t) increases the agent's expectations in earlier periods of $\hat{\rho}_t$.

In each period, the agent draws a cost c_t from a known distribution of costs $F(c)$ and would receive benefit y next period if they complete the task. Thus, the agent decides to act based on the current value function

$$W_t = \begin{cases} \beta\delta y - c_t, & \text{if act,} \\ \hat{\rho}_{t+1}\beta\delta E_t[V_{t+1}], & \text{if do not act,} \end{cases}$$

where $E_t[V_{t+1}]$ is the perceived continuation value of not doing the task in the current period (and potentially doing the task in a future period). Note that present bias leads the current value function W_t, which is a function of β, to differ from the perceived continuation value V_t, which is a function of $\hat{\beta}$. At the deadline, the continuation value is zero as the opportunity to perform the task in future periods is removed. By backwards induction from the deadline, the model leads to a cutoff strategy where the agent adopts in period t if the cost draw c_t is below a threshold c_t^*.

Specifically, by backwards induction we obtain a recursive set of expressions that implicitly define the cost threshold:

$$\begin{aligned} c_t^* &= \beta\delta\left(y - \hat{\rho}_{t+1}E_t\left[V_{t+1}\right]\right) \qquad (3) \\ E_{t-1}\left[V_t\right] &= F\left(\hat{c}_t^*\right)\left[\delta y - E\left[\hat{c}|\text{act}\right]\right] + \left(1 - F\left(\hat{c}_t^*\right)\right)\delta\hat{\rho}_{t+1}E_t\left[V_{t+1}\right] \\ E\left[\hat{c}|\text{act}\right] &= \int_0^{\hat{c}_t^*} c\,dF(c) \\ \hat{c}_t^* &= \hat{\beta}\delta\left(y - \hat{\rho}_{t+1}E_t\left[V_{t+1}\right]\right) \end{aligned}$$

The probability of adopting at period t is:

$$Pr\left(\text{adopt at }t\right) = \underbrace{\prod_{j=1}^{t}\rho_j}_{Pr(\text{remember})} \underbrace{\prod_{k=0}^{t-1}\left(1-F\left(c_k^*\right)\right)}_{Pr(\text{didn't already adopt})} \tag{4}$$

Thus, integrating over individual firms (whose i subscript was excluded above for ease of notation)—which can have heterogeneous costs—provides a set of moments giving the fraction of individual firms that adopt in period t, with one moment for each of the T periods (where T is the period in which the deadline occurs).

In our experimental setting, the benefit for firms is a lower merchant fee and the task is clicking the link in the email and filling out a short form. The time period is a day. We set our deadline to be in one week (midnight on day 8 of the experiment) for all treatment groups that have a deadline except group 8. If assigned, the anticipated or unanticipated reminder is sent on the day before the deadline, October 5th. For each treatment group, we will obtain take-up rates for eight periods that will be used as moments to estimate forgetfulness and present-bias parameters. To isolate variation in costs from the probability of forgetting, we add a treatment group with an immediate one-day deadline (group 8), so that this group only faces a one-period decision and (under the assumption that each period is a day) does not risk forgetting. Our experiment thus allows us to estimate a set of moment equations of the form (4) to estimate β, $\hat{\beta}$, ρ, $\hat{\rho}$, and heterogeneous costs, where each treatment arm provides T moments.

4 Tables

Table 1: Average Takeup (Pooled T3 and T4)

Sales Quartile	Average Takeup
1	6.11
2	13.17
3	15.68
4	22.04

Table 2: Sales Volume Regressions

	Log(sales + 1)	arcsinh(sales)
I(T3T4)	0.048**	0.052**
	(0.021)	(0.023)
Num.Obs.	239614	239614
R2	0.348	0.347
R2 Adj.	0.314	0.313
Firm FE	Yes	Yes
Month FE	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Outcome variable is log(sales + 1) in column 1 and arcsinh(sales) in column 2. The independent variable is an indicator for if the firm is in treatment group 3 or 4. Robust standard errors are in parentheses.

Table 3: Sales Volume Regressions by Baseline Sales Quartile

	log(sales + 1)				arcsinh(sales)			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
I(T3T4)	0.007	-0.013	0.064	0.132**	0.007	-0.014	0.069	0.145**
	(0.021)	(0.037)	(0.048)	(0.056)	(0.023)	(0.040)	(0.052)	(0.061)
Num.Obs.	60311	60934	60628	57741	60311	60934	60628	57741
R2	0.192	0.203	0.223	0.340	0.195	0.205	0.226	0.341
R2 Adj.	0.150	0.163	0.184	0.305	0.153	0.165	0.187	0.306
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ The outcome variable is log(sales + 1) in columns 1 through 4, and arcsinh(sales) for columns 5 - 8. Each regression specification is filtered for the sales quartile denoted in the column title. Robust standard errors are included in parentheses. All specifications include firm fixed effects and month fixed effects.

Table 5: Balance Test by Sales Quartile

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
T2	0.002	0.002	0.001	-0.006
	(0.009)	(0.009)	(0.009)	(0.009)
T3	0.002	0.000	-0.000	-0.001
	(0.010)	(0.010)	(0.010)	(0.010)
T4	-0.000	0.002	-0.001	-0.001
	(0.009)	(0.009)	(0.009)	(0.009)
T5	-0.000	0.000	-0.002	0.003
	(0.009)	(0.009)	(0.009)	(0.009)
T6	0.000	-0.001	0.002	-0.002
	(0.010)	(0.010)	(0.010)	(0.010)
T7	-0.002	-0.002	0.002	0.001
	(0.009)	(0.009)	(0.009)	(0.009)
T8	-0.002	-0.001	-0.001	0.004
	(0.009)	(0.009)	(0.009)	(0.009)
Num.Obs.	34010	34010	34010	34010
R2	0.000	0.000	0.000	0.000
R2 Adj.	-0.000	-0.000	-0.000	-0.000
P-value of F-Statistic	0.9999	0.9998	0.9996	0.9839

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Each outcome variable is an indicator for each of the respective sales quartiles. Sales quartiles are defined by the average monthly sales from July 2019 to August 2020. Robust standard errors in parentheses.

Table 6: Balance Test by Business Type

	Small Retailers	Restaurants	Professionals	Beauty	Clothing	Other
T2	-0.002	0.001	0.001	-0.002	0.002	0.001
	(0.010)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)
T3	0.002	-0.001	-0.001	-0.001	-0.002	0.003
	(0.010)	(0.007)	(0.010)	(0.006)	(0.006)	(0.009)
T4	-0.001	0.001	-0.001	-0.000	0.001	0.001
	(0.010)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)
T5	0.003	-0.002	0.001	0.002	-0.001	-0.002
	(0.009)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)
T6	-0.003	0.002	-0.001	0.001	0.001	-0.002
	(0.010)	(0.008)	(0.010)	(0.007)	(0.007)	(0.009)
T7	0.001	0.001	0.000	-0.000	-0.000	-0.002
	(0.009)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)
T8	0.002	-0.000	0.000	-0.001	-0.001	0.000
	(0.009)	(0.007)	(0.009)	(0.006)	(0.006)	(0.009)
Num.Obs.	34010	34010	34010	34010	34010	34010
R2	0.000	0.000	0.000	0.000	0.000	0.000
R2 Adj.	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
P-value of F-Statistic	0.9991	0.9992	1	0.9988	0.9989	0.9989

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Each outcome variable is an indicator for one of the business types. Business types are originally recorded by the FinTech company and then are mapped to one of the 6 general business types, indicated by the column titles. Robust standard errors in parentheses.

Table 7: Balance Test by Tax Registration Status

	Persona Moral	Persona Fisica	Individual
T2	-0.001	0.005	-0.003
	(0.005)	(0.006)	(0.003)
T3	0.002	-0.002	-0.000
	(0.005)	(0.006)	(0.003)
T4	-0.001	-0.001	0.002
	(0.005)	(0.006)	(0.003)
T5	0.005	-0.004	-0.000
	(0.005)	(0.006)	(0.003)
T6	0.004	-0.003	-0.001
	(0.006)	(0.006)	(0.003)
T7	0.005	-0.002	-0.002
	(0.005)	(0.006)	(0.003)
T8	0.004	-0.003	-0.000
	(0.005)	(0.006)	(0.003)
Num.Obs.	34010	34010	34010
R2	0.000	0.000	0.000
R2 Adj.	-0.000	-0.000	-0.000
P-value of F-Statistic	0.8456	0.8734	0.8048

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Each outcome variable is an indicator for one of the tax registation statuses. Tax registration status is an indicator variable that is recorded by the FinTech company. Robust standard errors in parentheses.

Table 8: Balance Test by Gender

	Male	Female	Gender_Missing
T2	0.002	-0.002	-0.002
	(0.011)	(0.011)	(0.005)
T3	0.003	-0.003	0.005
	(0.012)	(0.012)	(0.006)
T4	-0.002	0.002	0.000
	(0.011)	(0.011)	(0.005)
T5	0.004	-0.004	0.005
	(0.011)	(0.011)	(0.005)
T6	-0.010	0.010	0.005
	(0.012)	(0.012)	(0.006)
T7	0.006	-0.006	0.006
	(0.011)	(0.011)	(0.005)
T8	-0.006	0.006	0.003
	(0.011)	(0.011)	(0.005)
Num.Obs.	31725	31725	34010
R2	0.000	0.000	0.000
R2 Adj.	-0.000	-0.000	-0.000
P-value of F-Statistic	0.9007	0.9007	0.7382

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Each of the outcome variables is an indicator for the gender of the firm owner if it is recorded. Otherwise, it is classified as missing. Robust standard errors in parentheses.

Table 9: Balance Test by Length of Usage

	Older	Younger
T2	0.000	-0.000
	(0.011)	(0.011)
T3	0.003	-0.003
	(0.011)	(0.011)
T4	-0.002	0.002
	(0.011)	(0.011)
T5	0.008	-0.008
	(0.011)	(0.011)
T6	0.003	-0.003
	(0.012)	(0.012)
T7	0.005	-0.005
	(0.011)	(0.011)
T8	0.009	-0.009
	(0.011)	(0.011)
Num.Obs.	34010	34010
R2	0.000	0.000
R2 Adj.	-0.000	-0.000
P-value of F-Statistic	0.9597	0.9597

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The outcome variables are indicators for length of usage. The median length of usage for users is approximately 2 years. Newer users are defined as those who started using the technology within the past two years, and older users are those who started using it before that. Robust standard errors in parentheses.

References

Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association 15*(3), 692–719. ISBN: 1542-4766 Publisher: Oxford University Press.

List, J. A., S. Sadoff, and M. Wagner (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics 14*(4), 439. ISBN: 1386-4157 Publisher: Springer.