Building a Habit: How Initial Saving Activity Predicts Long-term Account Engagement

Mirpourian, Mehrdad

Women's World Banking

30 July 2020

Building a Habit:

How Initial Saving Activity Predicts Long-term Account Engagement

Mehrdad Mirpourian

Women's World Banking
New York, NY

Abstract

In this paper, time-to-event analysis is used to predict the risk of dormancy within financial institution accounts. The hypothesis tested was that a customer's behavior in the first month of account ownership holds clues to future dormancy, an idea supported by behavioral science literature. In many situations, the initial behavior of an individual can predict future behavior. Individual-level transaction data on a group of customers from one of the largest microfinance banks in Mexico was used to conduct a survival analysis using the Stratified Cox model. While adjusting for two other financial indicators, the team studied frequency of account usage during the first month after account opening and found that customers who use their account more often during that first month have a significantly lower risk of account dormancy than their counterparts.

Keywords: *account dormancy, survival analysis, financial inclusion, risk, savings*

1. Introduction

Accumulating savings at the individual customer level is an important component of economic empowerment and international development more broadly. Saving protects individuals and households from economic shocks, increases security, and has a significantly positive impact on people's lives overall (Ashraf et al., 2006). The 2017 Global Financial Inclusion Database (Findex) shows that 69 percent of the world's adults have bank or mobile money accounts in which to save safely. The picture is only slightly less inclusive in developing countries, where 63 percent of the population have a savings account. The development community is still trying to understand how to make use of these accounts, however, particularly since many of these accounts remain dormant. As Figure 1 illustrates, account dormancy is a global problem (Demirgüç-Kunt, 2018; Rhyne and Kelly, 2018).

Account dormancy could signify either that customers are not saving or that they are not using their account to save. This inactivity poses a puzzle for researchers: What barriers do customers experience when using their savings account? What communications materials, trainings, or other activities could financial services providers (FSPs) introduce in order to encourage increased usage of savings accounts for their customers, especially low-income women?

Everyone has the capacity to save. In fact, among populations for whom income is volatile, savings provide an essential safety net that eases household economic shocks (Collins et al., 2009). High rates of dormant accounts in formal financial institutions, such as those featured in global demand-side surveys, illustrate a disconnect between people's savings activity and their engagement with formal financial services.

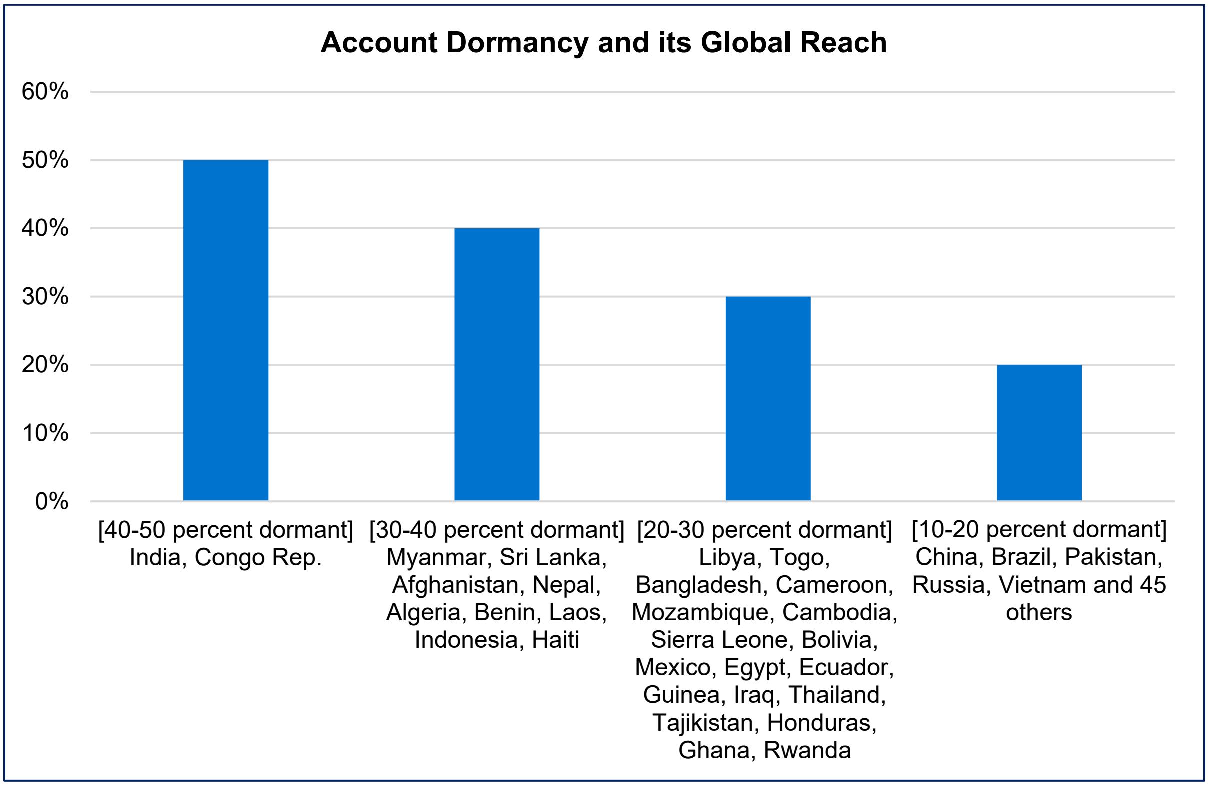

Figure 1. Percentage of dormant accounts in different countries

This paper explores saving activity among low-income Mexicans. We hypothesize that a customer's behavior in the first month of account ownership holds clues to future dormancy. Our findings have implications for how financial institutions engage with these populations in order to better tailor products and institutional engagement to meet their needs. In this section, we introduced the concept of savings and shared the global problem. Next, we discuss the literature and the related theoretical frameworks in behavioral science that prove relevant to this question. After these sections, we outline the model we use to test the hypothesis, then we share the results. Our analysis lends support to the idea that saving in formal institutions continues for a longer period of time if a customer takes particular actions in the first month of account ownership. The analysis has implications for financial institutions and contributes to a broader understanding of how to build financial capability among low-income consumers.

2. Related Literature

Household level surveys show that many low-income people have some surplus and tend to save. They often save in small amounts week after week (Collins et al., 2009; Banerjee and Duflo, 2007; Berry, 2006; Barr, 2004). When there are no formal savings services, or when formal saving is not a viable option, they may use informal savings methods. They may save at home, in informal savings groups, or in assets like livestock or gold. Therefore, the poor can save, but they do so in their own way (Mullainathan, 2009). Evidence shows that low-income people do not have many safe savings options and tend to make savings decisions based on relative risk

Although members of the low-income population are exposed to higher risks, there are fewer financial tools available for them to solve their financial problems (Collins et al., 2009). In fact, one study in Uganda showed that the cost of saving in the informal sector was more than 20 percent of the value of the savings (Wright & Mutesasira, 2001).

Access to formal financial services is vital for low-income people, and it is a fundamental step towards poverty alleviation. Savings accounts are not only safe places to save, but also serve as a helpful tool that can increase financial health if used frequently (Ladha et al., 2017). The financial inclusion industry holds that the key to economic empowerment is translating access to well-designed products into usage of those products. Karlan et al. (2016, 2014) studied some of the reasons that the poor prefer informal savings to formal savings. They also provided a list of the main reasons why the low-income population generally has a very low activity rate with formal savings: These include market frictions, transaction costs, lack of trust, and increased regulation. Many products that were designed to encourage savings among the poor had a low uptake of 20-40 percent, and the usage rates were too low. Metrics used to measure usage rates are also minimal in order to indicate if an account is active or dormant, i.e. “at least one deposit during the study period” (Karlan, McConnell, et al., 2016).

The Pradhan Mantri Jan Dhan Yojana (PMJDY) program, launched by the Indian government in 2014 as a national savings policy targeting Indians who did not have a bank account, illustrates clearly the aforementioned gap between access to and usage of savings accounts. The PMJDY program provided the target population with a free debit card, financial literacy programs, access to credit, and insurance. Although PMJDY experienced great success in providing access to finance, customers' usage rates remained low. Research showed different reasons for low engagement among program participants, including limited knowledge about the functionality of the account combined with rudimentary financial literacy (Verma and Garg, 2016).

Low-income people may also choose not to use their savings accounts due to poor infrastructure or to cultural norms. Long commutes to the bank, high expenses to maintain an account, lack of trust in financial services, religious motives, insufficient funds or surplus, reliance on another family member who already uses an account, or lack of need for and interest in financial services are other possible explanations for why people may not use their accounts.

There are two reasons that these explanations may be unsatisfactory for financial institutions. The first challenge is that limited knowledge of the account and low financial literacy may drive inactivity, but addressing these directly may not be the best way to improve account use. Financial literacy and financial education do not have a strong link to long-term changes in financial behavior (Fernandes et al., 2014). The second reason why these explanations may be unsatisfactory is that they involve changing the infrastructure or the culture. The infrastructure constraints on account use are often too capital intensive for a financial institution to address at scale. Cultural norms are intractable in this context. These explanations for dormant accounts, therefore, do not provide much hope to financial institutions engaging in increasing savings among low-income customers.

More empowering to consumers and more enabling to FSPs is the idea that there are other, less intractable explanations for dormant accounts. An analysis of the behavioral science literature focusing on account engagement, particularly the endowment effect, is helpful in order to understand why low-income people may "under-save." The endowment effect theory argues that when people receive cash, it is more difficult for them to hand it back to the bank and deposit the savings into their account (Spantig, 2017). Prelec et al. (1998) focused on how savers feel when they deposit their money into their bank account and found that even if customers have a willingness to save, depositing their money in their savings account could present a psychologically painful experience for many of them (Karlan, Ratan, et al., 2014; Thaler and Benartzi, 2004). Strömbäck (2017) tried to explain savings behavior using behavioral life cycle theory (BLC). According to BLC, "people act as if there within every person is an ongoing conflict between a 'planner' who thinks about the long-run and a 'doer' who is more concerned about the current situation." This theory argues that people believe it is costly to save for the future based on the time it takes to save. For example, it is easier to spend your income now, while it may seem more costly to save it for a future need and put it aside.

Based on BLC, long-term success in savings is highly correlated with the subject's abilities to control impulses. Other scholars like Gathergood (2012) evoke the self-control theory for explaining savings behavior. Gathergood claims that saving behavior and tendency to save have a positive correlation with self-control. Similarly, research by Biljanovska and Palligkinis (2015) studied the lack of self-control among households and its correlation with lower wealth accumulation. Choi et al. (2011) discussed the theory that those who have low self-control are less likely to save enough for their retirement. In all of these theories, there is a common theme: The subjects' propensity toward the aforementioned behavioral biases will drive their future behavior. In other words, prior behavior influences future decision-making.

Some academics argue that basic human traits can be a useful predictive tool in understanding future behavior (Harris et al., 2016). To help customers overcome traits that discourage saving, financial institutions could consider designing products in ways that encourage savings behavior. The possibility of introducing design shifts—by adding "moments" into the customer journey, for instance—is empowering both for institutions and for individuals.

3. Research Idea and Question

Behavioral science theories provide explanations for habits such as under-saving and account dormancy. However, these theories may not always accurately predict individuals' behavior (Yarkoni et al., 2017). Although explanation and prediction may seem compatible philosophically, they are often in statistical and pragmatic tension with one another (Hagerty & Srinivasan, 1991; Shmueli, 2010; Wu, Harris, & Mcauley, 2007; James, Witten, Hastie, & Tibshirani, 2013). Traditionally, behavioral scientists preferred explanation rather than prediction. However, in many cases, programs that focus on prediction and consider explanation as a secondary goal are more successful in the short and long term. Yarkoni et al. (2017) suggested that focusing on prediction instead of explanation could lead to greater understanding of behavior. Some other studies recommend integrating components of behavioral theories as features in building predictive models, to help improve the accuracy of predictions (Plonsky et al., 2019; Rosenfeld et al., 2012).

Given the practical implications of this research, however, the team chose to pursue prediction with the intention that this approach would also provide some explanation. In doing so, the team emulated studies that focus on prediction while considering explanation as the secondary goal. Therefore, when building the statistical model and feature selection, learnings and insights from behavioral economics were incorporated.

Data on the initial traits of first-time microfinance customers contains valuable information on their habits, conveying how much difficulty and pain a customer might experience when handing in his/her cash as a savings deposit to a bank, and how an individual's financial behavior might look in the future. In this case, it might be a predictor of account dormancy. Therefore, among many possible options, we chose to look at how first-month financial behavior compares against future account activity.

To test our hypothesis, the team studied a novel dataset to probe whether a customer's first month of financial behavior predicts his or her risk of account dormancy within the first year of account ownership. We collected the data from a large microfinance bank based in Mexico. This paper follows a particular set of financial indicators during the first month after a customer opens an account, in order to study the extent to which these factors can predict the risk of account dormancy within a one-year time window after opening the account. This is the first study to use survival analysis to assess account dormancy, and the first to incorporate into its statistical model a customer's first-month financial behavior as a factor that might contribute to account dormancy.

If FSPs can identify customers who are at high risk of dormancy one month after opening their account, they can provide those customers with customized services and support. By doing so, they can transform access to savings into usage of savings. In addition, FSPs will be able to more accurately project their portfolio cash flow and estimate what percentage of their customers may experience account dormancy within the first year of ownership. Lower account dormancy allows providers the opportunity to increase their cash flow and decrease the cost of lending and credit products, directly helping their customers.

4. The Empirical Strategy

Survival analysis, also called "time to event analysis," is a collection of statistical techniques that study the expected duration of time until one or more events happen.

There are two key time-related terminologies in survival analysis: survival time and censoring. Survival time is the time period that an individual has survived during the follow-up period. In other words, the term "survived" refers to any individual who did not experience a so-called death event and survived for the duration of the study period. Based on the context of each study, "event" refers to different experiences. For example, when an application user deletes an application, deleting the application is the event. The moment that a person dies is the event in a public health study. Breaking down a machine in the production line of a factory is the event in engineering research.

In each study, the researcher defines what the event is. In our study, "event" refers to the time when a customer stops using his/her account. When a customer stops using his/her account for a specific period, we call his/her account dormant. Therefore, dormancy is the event in our study, and individuals who stayed active in using their accounts are the survived cases. Looking at the literature and at FSP conventions, we did not find unanimous agreement about how long an account should be inactive in order to be considered dormant. Therefore, we came up with our own definition of account dormancy, which has its roots in our expertise in the field of financial inclusion. In our definition, an account is considered dormant if the account holder does not use it for at least 90 days out of 12 consecutive weeks. As mentioned earlier, this period may vary among different FSPs, but based on our experience in financial inclusion, we defined this period as 90 days.

The other time-related concept in survival analysis is censoring. Censoring happens when we do not know the exact survival time of an individual; however, we might have some information about it. In our study, data is censored due to one of the following reasons: a customer's account does not become dormant, or customer information gets lost in the database, or a customer closes his/her account during the study period. To make this point more clearly, we show one censored and one survived case in Figure 2. This illustration describes the experience of two customers we followed over time. The **X** denotes an account that became dormant (received the event). Customer A used his account for the first three weeks, and after that he stopped using it. Therefore, at the end of week 15 he received the event and became dormant. Since this customer received the event, his data is not censored. Customer B is followed from the time she opened her account until the end of week 52. During this time, she never became inactive for 12 consecutive weeks and did not get the event. So, her data is censored.

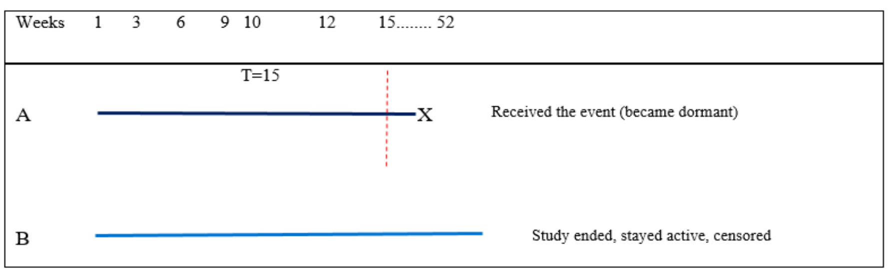

Figure 2. Censored vs. failed cases

Source: Authors' elaboration

To study how much the initial behavior of customers predicts a future risk of account dormancy, we use four explanatory variables that measure customers' financial behavior during the first month after account opening. The variables are: time between the first and second transaction; number of times a customer uses the account in the first month; and first-time transaction amount after account opening. We selected these four variables because they convey the most useful information about the first-month financial behavior of a customer after account opening.

Among these four variables, the number of times a customer uses his/her account (frequency of usage) is the main variable of interest. Using this variable, we set out to test whether or not first month high-engaged vs. low-engaged customers show the same level of engagement in the long term. We did not use frequency of usage directly in our model. First, we calculated the 20^{th} and 80^{th} percentile of this variable. Customers who fell under the 20^{th} percentile were considered low-engaged, and those who fell in the 80^{th} percentile or above were considered high-engaged. Using this approach, we created a two-level categorical variable that takes 1 if the customer is high-engaged and takes 0 if he/she is low-engaged. Customers who did not fall into one of these two categories were excluded from our study. Therefore, we used the frequency of usage for two purposes. First, we used it to detect low- and high-engaged customers so we could include only those two groups in our model. Second, we used it as the main variable of interest in studying account dormancy.

Finally, to minimize the effect of macroeconomic factors as well as seasonality, we only included customers who opened their accounts in May 2017.

5. Mathematical Modeling and Terminology

5.1. Mathematical Notation

To explain the mathematical formulation of the survival model used in this study, we will start by introducing the notations. Capital *T* shows the time in which a customer stayed active (survival time), and small *t* shows the value of interest for the random variable capital *T*. Small letter *d* is a binary random variable indicating either failure or censorship. If the event occurs in the course of the study period, *d* = 1 and otherwise *d* = 0. *S(t)* denotes the survivor function. Survivor function gives the probability that a customer does not experience dormancy by a specified time *t,* and we showed its definition using the following formula:

$$S(t) = P(T > t)$$
(1)

The notation *h(t)* denotes the hazard function. Hazard function is the potential that the event occurs, given that an individual has survived up to time *T*. Formula 2 gives the mathematical definition of hazard function.

$$h(t) = \lim_{\Delta t \to \infty} \frac{P(t \leq T < t + \Delta t \,|T \geq t)}{\Delta t}$$
(2)

5.2. Kaplan-Meier Curve and Survival Probabilities

To estimate the survival probability at a specific moment in time, we can use the Kaplan-Meier (KM) method. KM uses a concept called risk set. Risk set contains all the customers who have not received the event and are at risk of receiving the event. This approach includes all the information on censored customers up to time of censorship and does not simply discard the information on censored cases. The general formula for a KM survival probability at failure time *t(f)* is as follows:

$S(t_{(f)}) = S(t_{(f-1)}) * \Pr(T > t_{(f)} \mid T >= t_{(f)})$

Which $S(t_{(f-1)}) = \Pi_{i=1}^{f-1} (\Pr(T > t_{(i)} \mid T >= t_{(i)}))$
(3)

5.3. Cox Proportional Hazards Model

While the KM curve provides invaluable insights on the survival probabilities of high-engaged versus low-engaged customers, it does not do the following:

- Test for the significance of the engagement-level variable, adjusted for other variables included in the model.
- Obtain a point estimate of the effect of the engagement-level variable, adjusted for other variables.
- Obtain a confidence interval for variables included in the model.

Answers to these questions can be found in the Cox Proportional Hazards (PH) model. As we can see in Formula 4, the Cox PH model is usually written in terms of the hazard model.

$$h(t, \boldsymbol{X}) = h_0(t)\, e^{\sum_{i=1}^{P} BiXi} \qquad (4)$$

$\boldsymbol{X} = (X1, X2 \dots Xp)$

This model gives estimates for the hazard at time *t* for a customer with a given set of explanatory variables, and $\boldsymbol{X}$ denotes this set of explanatory variables. $h_0(t)$ shows the baseline hazard function, which is an unspecified function of *t.* Since $h_0(t)$ is an unspecified function, the Cox PH model is a semiparametric method. The second component of this model is an exponential term. This term is a linear function of time-independent predictors.

The Cox PH model is a robust model, and because of this property, the results from using the Cox PH model will closely approximate the results for the correct parametric model. This beneficial property was our main motive for using the Cox PH model. Another point that we need to answer at this stage is why we chose the survival model when we could have used other methods such as logistic regression or some other sorts of classification. The answer is mainly inherent in the fact that in cases in which the survival time information is available and censored data is present, the Cox PH model has advantages over classification methods. This method uses more information compared to classification methods that use only a binary outcome variable but do not use the information inherent in survival times and censoring.

5.4. Assessing PH Assumption

The most important assumption of the Cox PH model requires that the hazard rate (HR) stays constant over time. As we showed in Figure 3, the PH assumption says that the hazard for one individual compared to the hazard for any other individual should be constant over time (time independent).

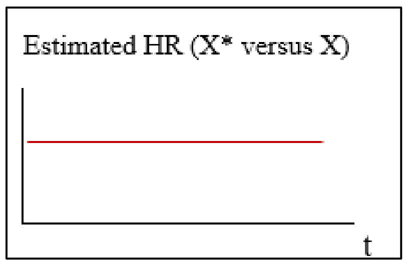

Figure 3. Cox PH assumption

There are three common approaches to checking whether this PH assumption is satisfied or violated. These three methods are the graphical method, the goodness of fit (GOF) test, and the use of a time-dependent variable in the model. In this study, we used the GOF approach. GOF provides a test statistic for assessing the Cox PH assumption for all the predictors. Therefore, the decision-making process is free of any subjectivity. There are different GOF tests developed by statisticians. Among those, we used the Harrel and Lee test, which is a variation of the Schoenfeld test. This is one of the most widely used tests for assessing the PH assumption.

5.5. Stratified Cox (SC) Model

There are two common methods for adjusting a Cox PH model that violates PH assumption. These two methods are the Stratified Cox (SC) model and the Extended Cox model. In our analysis, the PH assumption was violated, and to address this issue, we used the SC model. Therefore, we only provide the mathematical notation for an Extended Cox model. The SC model controls for the predictor that violates the PH assumption, while other variables that satisfy the PH assumption are included in the model as explanatory variables. The hazard function formula for a stratified Cox model is shown formula 5. Subscript g in formula 5 denotes the g^{th} stratum in which strata are different categories of the stratification variable Z*.

$h_g(t, X) = h_{0g}(t) \exp [B_1X_1 + B2X2 + ... + BpXp]$

$g = 1, 2\ldots k^*$, strata defined from Z^*

Z^* has k^* categories

$X_1, X_2 ... X_p$ satisfy PH (5)

6. Results and discussion

6.1. Kaplan-Meier Curve and Cox PH Model

In this section, we apply all of the steps we explained in the mathematical modeling section on customers' individual-level transaction data. These results are derived by using R Studio statistical package version 3.6.3.

As the initial step of survival analysis, we plot the KM curve of low-engaged versus high-engaged customers. As we can see in Figure 4, the survival curve of high-engaged customers (the dark green dotted line – strata 1) falls above the survival curve of low-engaged customers (the purple line – strata 0). This means that during the 52 weeks of this study, high-engaged customers had a higher chance of survival and of not becoming dormant compared to their low-engaged counterparts. To check whether the gap between these two groups is statistically significant or not, we use log rank statistic. Log–rank is a chi-square test based on a large sample that provides an overall comparison of the KM curves. In our study, the p-value of log-rank test became 2e-16. Therefore, at the significance level of 0.05, we can reject the null hypothesis of this test. This result is the equivalent of rejecting the hypothesis that claims "customers who had low engagement in the first month have the same survival probability as those who had high engagement." Therefore, it is valid to conclude that high-engaged customers have a significantly lower chance of becoming dormant during the first year after opening an account compared to those who had a first-month low engagement level.

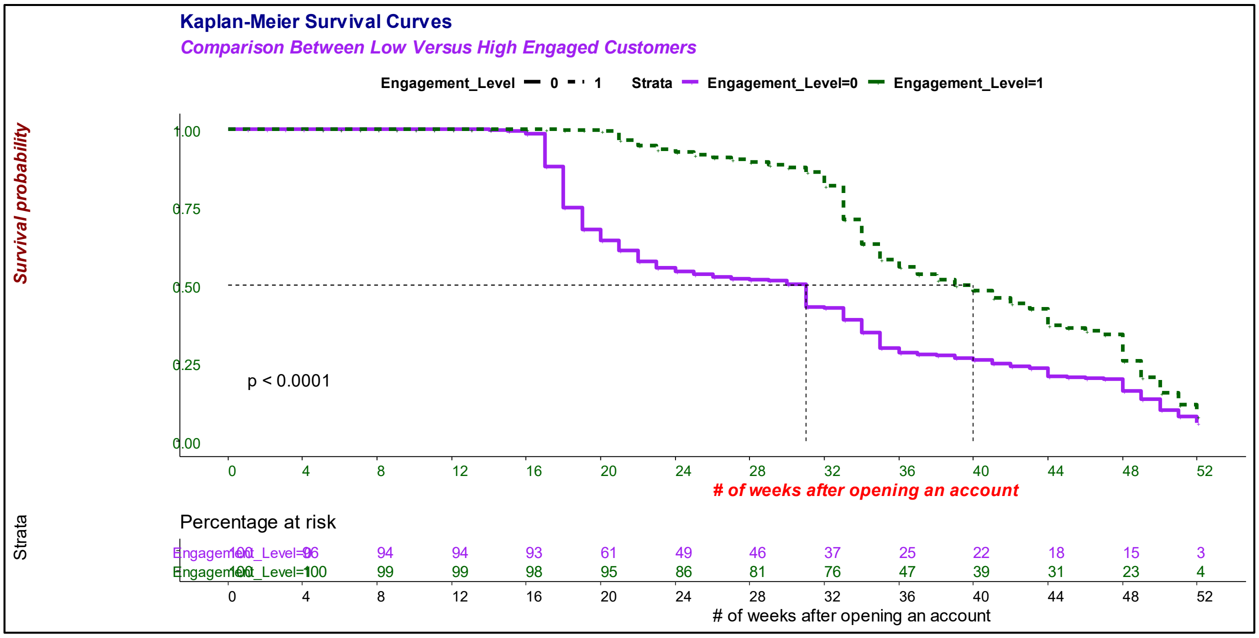

Figure 4. KM survival curve

As we explained earlier in the mathematical modeling section, the KP curve cannot provide point estimates on the effect of engagement level while adjusting for other covariates. One of the methods that can provide this information is the Cox PH model. Using the four aforementioned predictors and the time to becoming dormant as the survival time, we built a Cox PH model. The summary of this model is shown in Table 1. Looking into this table, we find the same conclusion about the effect of first-month engagement level on account dormancy. However, the Cox PH model provides us with more information. Based on the model output in Table 1, we see that customers who had low engagement in the first month after account opening have an almost 50 percent higher risk of having dormant accounts compared to those who were in the high-engagement category. Although the other three explanatory variables are significant, their predictive power for account dormancy is very small and negligible. All of the interpretations we provided here are valid, if PH assumption is not violated.

Table 1. Cox PH model summary

COX PH Model Summary					
Predictor variable	Coef	se(coef)	exp(coef)	z	Pr(>|z|)
Engagement level_80 percentile	-0.59	0.02	0.55	-30.84	0.00
First transaction amount	0.00	0.00	1.00	-3.30	0.00
Time between first and second transaction	-0.02	0.01	0.98	-2.66	0.01
Likelihood ratio test=1030 on 3 df, p=< 2.2e-16					
n = 14929, number of events = 12057					

6.2. Stratified Cox Model

Using cox.zph built in function in R, we checked the PH assumption. Using this test, we can study the correlation between Schoenfeld residuals and ranked survival time. If the correlation is zero, it is a support for the proportional hazards assumption (the null hypothesis). If there exists a correlation between the Schoenfeld residuals and ranked survival time, we can reject the null hypothesis in favor of the alternative. Table 2 provides the model assumption assessment. As we can see, Schoenfeld's global test indicates a lack of fit of the model. There is a significant violation of the PH assumption for the engagement level.

Table 2. Cox PH model assumption assessment

Predictor variable	chisq	df	p-value
Engagement_Level_80_Percentile	1900.00	1.00	<2e-16
First transaction amount	0.03	1.00	0.86
Time between first and second transaction	1.62	1.00	0.20
Global	1900.00	3.00	<2e-16

In addition to the information in Table 2, we provided a visual tool to assess the PH assumption. Figure 5 shows Schoenfeld residuals for engagement level. The non-linear shape of residuals is a clear indication for violation of PH assumption.

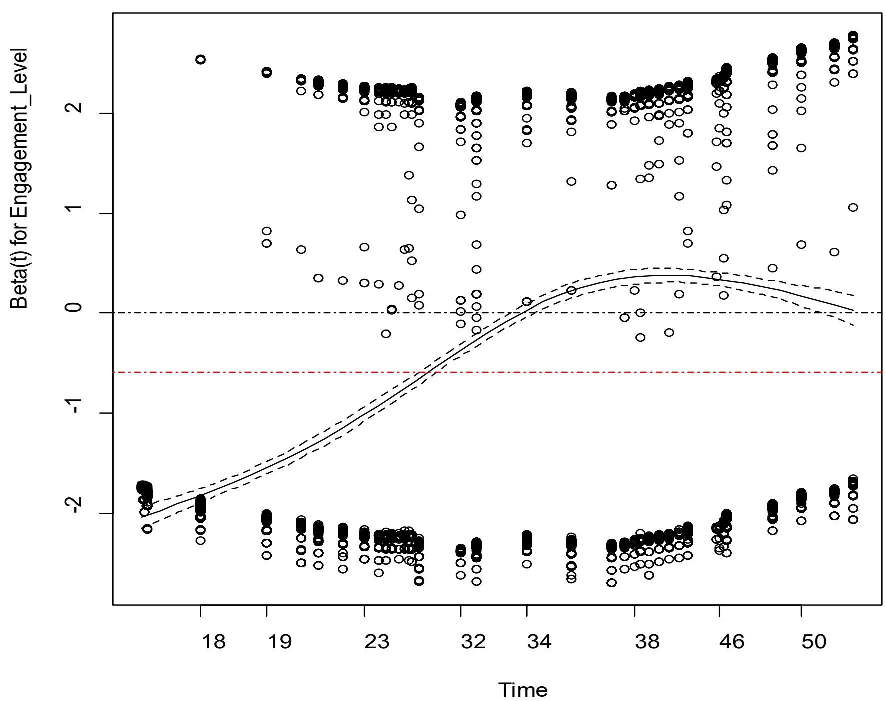

Figure 5. Assessing PH assumption

Since PH assumption is violated for explanatory variable engagement level, we need to adjust the model and address this issue. As explained earlier, there are two common ways to resolve this issue: the Stratified Cox model and the Extended Cox model. Between these two methods, we chose the former and built a new model by stratifying on engagement level. The summary of the Stratified Cox model is provided in Table 3.

Table 3. Model Summary - Stratified Cox Model

Stratified COX Model Summary					
Predictor variable	Coef	se(coef)	exp(coef)	z	Pr(>|z|)
First transaction amount	-1.59E-06	4.81E-07	1.00E+00	-3.299	0.000969
Time between first and second transaction	-2.52E-02	9.07E-03	9.75E-01	-2.78	0.005431
Likelihood ratio test=19.93 on 2 df, p=4.703e-05					
n= 14929, number of events= 12057					

Initially, we had included three variables in the Cox PH model. After stratifying on engagement level, we ended up with a model with six covariates. As we can see in Table 3, the first transaction amount variable and the time between first and second transaction variable are both statistically significant. Although using SC model does not allow us to study engagement level, the KM curve and log rank test showed us that engagement level is a significant variable for predicting account dormancy.

7. Conclusion

This paper analyzed the effects of the engagement level of customers in using their accounts. In addition, the first-time transaction value and the time between the first and second transaction were also included as explanatory variables to study the risk of customers' account dormancy.

Empirical results were obtained based on a novel data set from Mexican microsavers, specifically collected for the purpose of this work. The results presented above show that all variables used in this analysis are statistically significant. Among all of them, the hazard rate and the effect of the variable "engagement level" are stronger compared to the other variables used in the model.

Using the first-month financial behavior of customers as an explanatory variable of a survival model to study account dormancy has practical implications for financial services providers, and offers new insights not provided in previous works in this field of research.

References

Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, *121*(2), 635-672.

Banerjee, A. V., & Duflo, E. (2007). The economic lives of the poor. *Journal of Economic Perspectives*, *21*(1), 141-168.

Barr, M. S. (2004). Banking the poor. *Yale J. on Reg*, *21*, 121.

Berry, C. (2006). To bank or not to bank? A survey of low-income households. *Building Assets, Building Credit: Creating Wealth in Low-income Communities*, 47.

Biljanovska, N., & Palligkinis, S. (2018). Control thyself: Self-control failure and household wealth. *Journal of Banking & Finance*, *92*, 280-294.

Choi, J. J., Laibson, D., & Madrian, B. C. (2011). $100 bills on the sidewalk: Suboptimal investment in 401 (k) plans. *Review of Economics and Statistics*, *93*(3), 748-763.

Collins, D., Morduch, J., Rutherford, S., & Ruthven, O. (2009). *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton University Press.

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. The World Bank

Fernandes, D., Lynch Jr., J. G., & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, *60*(8), 1861-1883.

Gathergood, J. (2012). Self-control, financial literacy, and consumer over-indebtedness. *Journal of Economic Psychology*, *33*(3), 590-602.

Hagerty, M. R., & Srinivasan, V. (1991). Comparing the predictive powers of alternative multiple regression models. *Psychometrika*, *56*(1), 77-85.

Harris, L., Lee, V. K., Thompson, E. H., & Kranton, R. (2016). Exploring the generalization process from past behavior to predicting future behavior. *Journal of Behavioral Decision Making*, *29*(4), 419-436.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 112, p. 18). New York: Springer.

Karlan, D., Ratan, A. L., & Zinman, J. (2014). Savings by and for the poor: A research review and agenda. *Review of Income and Wealth*, *60*(1), 36-78.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, *62*(12), 3393-3411.

Kleinbaum, D. G., & Klein, M. (2010). *Survival Analysis*. Springer.

Ladha, T., Asrow, K., Parker, S., Rhyne, B., & Kelly, S. (2017). Beyond financial inclusion: Financial health as a global framework. *Center for Financial Services Innovation (CFSI)*.

Mullainathan, S., & Shafir, E. (2009). Savings policy and decision-making in low-income households. *Insufficient Funds: Savings, Assets, Credit, and Banking Among Low-income Households*, *121*, 140-142.

Plonsky, O., Apel, R., Ert, E., Tennenholtz, M., Bourgin, D., Peterson, J. C., & Cavanagh, J. F. (2019). Predicting human decisions with behavioral theories and machine learning. *arXiv preprint arXiv:1904.06866*.

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science*, *17*(1), 4-28.

Rhyne, E., & Kelly, S. E. (2018). Financial inclusion hype vs. reality: Deconstructing the 2017 Findex results. *Center for Financial Inclusion/ACCION*.

Rosenfeld, A., Zuckerman, I., Azaria, A., & Kraus, S. (2012). Combining psychological models with machine learning to better predict people's decisions. *Synthese*, *189*(1), 81-93.

Rutherford, S. (2000). *The Poor and Their Money*. New Delhi: Oxford University Press.

Shmueli, G. (2010). To explain or to predict?. *Statistical Science*, *25*(3), 289-310.

Spantig, L. (2017). Mental accounting and savings decisions of the poor. *AEA RCT Registry February*.

Strömbäck, C., Lind, T., Skagerlund, K., Västfjäll, D., & Tinghög, G. (2017). Does self-control predict financial behavior and financial well-being?. *Journal of Behavioral and Experimental Finance*, *14*, 30-38.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(S1), S164-S187.

Verma, Y., & Garg, P. (2016). Pradhan Mantri Jan Dhan Yojana (PMJDY): A step towards eradicating financial untouchability. *Indian Journal of Finance*, *10*(1), 56-64.

Women's World Banking. (2019). *Deconstructing Findex. Insights Into Our Priority Markets.* https://www.womensworldbanking.org/findex-data/

The World Bank. (2018). *The Little Data Book on Financial Inclusion 2018.* The World Bank, Washington, D.C.

Wright, G. A., & Mutesasira, L. K. (2001). The relative risks to the savings of poor people. *Small Enterprise Development*, *12*(3), 33-45.

Wu, S., Harris, T. J., & McAuley, K. B. (2007). The use of simplified or misspecified models: Linear case. *The Canadian Journal of Chemical Engineering*, *85*(4), 386-398.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.