Marketing Science Institute Working Paper Series 2016
Report No. 16-110

Detection of Customers’ Life Change: Real-time Analysis Using a Control Chart Approach

Yi Zhao, Nuo Xu, and Yingge Qu

Report Summary

Customers may change their purchasing behaviors and develop new needs for products when life changes happen. However, it is difficult for managers to conduct targeted direct marketing based on life changes because consumers will not voluntarily report their life changes.

Yi Zhao, Nuo Xu, and Yingge Qu provide a simple and efficient solution to detect customers' life changes using rich information collected through a customer management system. Drawing from the statistical process control literature, they present a detailed solution for constructing an index based on sequential likelihood ratio test, which can quickly capture a deviation of consumer's behavior from a typical pattern before life change.

They evaluate the ability of their solution to detect life changes using both empirical data and simulation. Both settings show better performance than that of the benchmark model.

Specifically, using data from a Fortune 500 financial services company, they show how the method helps managers detect a specific type of major life change using information about customers' financial activities, their communications with the bank, and individual characteristics such as length of time in the job.

To demonstrate that the proposed solution is managerially and practically useful, the authors conduct a simulation study. Their findings show that, given certain revenue (loss) – false alarm function, a manager can set the threshold of alarm by maximizing profit. Maximum profit based on the proposed solution is 30%-70% higher than that based on a benchmark solution.

The proposed method can also be used to detect other types of change that are of interest to managers, such as detecting changes in customers' preferences in grocery shopping and changes in customers' risk of defaulting on credit card and mortgage payments, etc.

Yi Zhao is Associate Professor of Marketing, J. Mack Robinson College of Business, Georgia State University. Nuo Xu is Assistant Professor of Marketing, Strome College of Business, Old Dominion University. Yingge Qu is Assistant Professor of Marketing, College of Business, Mississippi State University.

Acknowledgements
We thank the Wharton Customer Analytics Initiative and an anonymous company that provided the data.

Introduction

As customers move through life's milestones—including graduation, career change, marriage, parenthood, home ownership, and chronic disease—their needs and aspirations also change. Major life changes ultimately reshape what customers value in a product or service. Such changes also affect the income levels and budget constraints that influence their purchasing decisions. Yet, although life events have important implications for managerial decisions, it is difficult for managers to apply these insights because they lack information on individual customers' life changes. In practice, customers do not update their life events in a firm's database, at least not in a simultaneous and immediate fashion. The result is that managers sit on a pool of outdated customer profiles that were usually collected on the first day of registration. Managers need a way to detect new changes in customers' lives.

Big data makes it feasible to infer life changes from observed information such as customers' activities and individual characteristics. Customer management systems today capture rich information about each customer. Banks, for example, have records on customers' financial portfolios and their contacts with the banks through walk-ins, phone calls, mail, e-mail, and visits to websites. Banks also acquire information about customers' financial activities at competitor banks through third-party data services, such as IXI. These pieces of information, when combined, can provide us with critical inferences on customers' life changes. Consider the case of a family preparing to purchase a home. The family is likely to increase their savings for the down payment, make more frequent visits to financial advisors, and browse information on mortgages and home insurance. In this way, a life change causes a systematic shift in a customer's behavioral pattern. By detecting these behavioral signals in the data, we can infer a family's intention to buy a home. The question is, *how do we quickly detect a shift in a customer's behavioral pattern that signals a life change of interest?*

This task is challenging, for three reasons. The first challenge is that customers typically show large variances in their behavior over time. The systematic shift that can be attributed to the life change of interest, however, is small. It is difficult to detect the signal of a small shift in an individual's behavioral pattern amid the noise of that individual's other behavioral variances. The second challenge is the selection of the optimal window for change detection. Because the shift is small, optimizing the time window for change detection is crucial. In an ideal case, it is most efficient to test for the shift starting from the actual change point. However, one cannot know in advance whether and when a customer's life change will occur. To ensure the efficiency of the test, one must identify the most likely change point based on the data available. The third challenge is that, in real-life application, data arrives every new period. To obtain timely intelligence on a customer's life change, an algorithm needs to be able to incorporate new data as they arrive and produce the most up-to-date results.

The objective of this study is to develop a solution to detect a specific life change of interest while addressing these three challenges. (Note, however, that our solution is not specific to a certain type of life change. This framework can be applied to detect any type of life change that can cause systematic changes in customer behavior.) Drawing from the literature of the field of statistical process control, we develop a sequential test of a life change of interest based on the framework of the CUSUM control chart.[1] In the proposed solution, we construct the problem as one of hypothesis testing, the goal of which is to test for the shift in a customer's behavior that signals the life change of interest. The test statistic accumulates deviation in the direction of interests over time as evidence of life change. The efficiency of the test is enhanced by selecting the optimal window of observations for the testing and by modifying this optimal window dynamically as new data arrive. We thereby render the test statistic sensitive to a shift in behavior pattern.

To the best of our knowledge, our research is the first to introduce the CUSUM control chart for change detection into a general marketing context. The CUSUM algorithm is designed for real-time analysis. It offers a recursive equation describing relationships between the statistic at time t and t-1, which simplifies the computation as new data arrive. Our proposed solution extends the original design of the CUSUM control chart to accommodate the complexity of the customer management context.

It is a straightforward task to measure the parts and calculate their deviations from the designed norm in a typical quality control setting. It is more difficult to gauge deviations in customer behavior, however. In this paper, we present a detailed solution for constructing the likelihood function in order to extract typical pattern before and after life change and to use the likelihood to evaluate deviation from the typical pattern before change. We further extend the test sta-

[1] CUSUM gets its name because it cumulates deviations from the norm and uses their sum as the test statistic.

tistic to adjust for individual and circumstantial differences in the probability of changes in customer behavior. Despite the addition of these complexities to the model, we are still able to maintain the simplicity of the CUSUM method by deriving a recursive formula for the test statistics between time t and time t-1. Success in deriving this recursive formula is crucial for applying the method to the big data scenario. In this way, this study also contributes to the literature of statistical process control by allowing the CUSUM control chart to accommodate more variety in the data.

We demonstrate the applicability of our method using data sponsored by a Fortune 500 financial services company. Managers of the bank wish to detect a specific type of major life change—career change—using information about customers' financial activities, their communications with the bank, and individual characteristics such as their lengths of time in the job. We evaluate the ability of our solution to detect life changes using both empirical data and simulation. In both settings, we are able to show better performance than that of the benchmark model.

In the next section, we review the literature on life changes and models for regime change. We then discuss the empirical context and the data to establish the context for the proposed solution. We then describe the details of the model and present results. Finally, we summarize our methodological and managerial contributions and discuss directions for future research.

Literature Review

The importance of major life changes

Major life events are valuable information for marketing managers. For example, PRIZM, a well-known system for customer segmentation, incorporates information on customers' life cycles with customers' life styles along with geographic information to effectively segment US customers. Existing studies have established the impact of major life events on a wide range of customer behavior, including consumption level (Gourinchas and Parker 2002), brand preferences (Andreasen 1984; Mathur et al. 2008) and financial behaviors, such as investing (Cocco et al. 2005), buying insurance (Wilkes 1995), and loan payment (Baek and Hong 2004). Researchers have also found that life stages classified based on major life events provide meaningful interpretations of customers' consumption patterns (Du and Kamakura 2006; Lansing and

Kish 1957). The occurrence of major life events are found to be related to demographic factors, such as age, education level, family structure, employment opportunities, and economic resources (Benzies et al. 2006; Kreyenfeld 2010).

Our research takes a different path. Unlike previous research, our study does not observe life changes directly. Our goal is to develop an efficient solution to detect life changes after they occur. Based on existing knowledge about the connections between life events and customer behavior, we use changes in customer behavior as indicators of life changes. We also exploit demographic information to account for individual differences in customers' propensities to undergo life changes. Figure 1 presents a conceptual map of the problem, as well as available information. The challenge of this task is that the observed variables, when examined individually, are weak indicators of life changes. Advanced technique is needed in order to quickly detect life changes using rich customer data accumulated over time. (Tables and figures follow References.)

Evaluating and modeling changes

The detection of life changes can be framed as a problem of change point detection or a problem of classification (labeling observations as "no life change" or "life change"). In this broad sense, several methods in the existing literature are related to this problem; they can be categorized into four groups. A summary table of these methods is provided in a web appendix.

The first category contains studies using event study method and studies on structure break (Sood and Tellis 2009; Wiles et al. 2010; Perron 1989). Researchers typically apply these two methods when the event is known to have occurred and the approximate date of the occurrence is also known. The goal of these studies is to understand and evaluate the impact of the focal event. In some cases, the date of the event is difficult to determine. In such a case, a common strategy is to evaluate all possible dates and choose the one that most favors the hypothesis of regime change. This remedy, however, are difficult to implement when the occurrence of the event is uncertain and the possible dates of the event cover a long period of time.

The second category of detection methods contains studies using cluster analysis (Liao 2005), which aims to categorize objects into groups by minimizing within-group distance and maximizing between-group distance. A common feature of clustering algorithms is an iterative process that screens through the entire data set for the ideal partitions. Such algorithms, however, can be time consuming when applied to big data and are not suitable for real-time analysis.

The third category contains studies using logistic regression, discriminate analysis, and machine learning methods, such as decision-tree algorithm (Morrison 1969; Punj and Stewart 1983). In these methods, coefficients, or weights, are estimated for all factors in order to calculate propensity scores for group memberships. Both of these methods are designed for cross-sectional analysis and are not methods for time series data.

The fourth category contains two time-series methods: the survival model (Helsen and Schmittlein 1999) and the hidden Markov model (HMM). These methods have two merits: (a) both incorporate time-varying variables to infer the propensity of an event, and (b) both provide simple computation schemes that allow newly arrived data to be easily incorporated in the analysis. The HMM framework provides a more flexible way than the survival model to simultaneous model the different relationships on how influential factors impact the transition process and how behavior reflects changes in the underlying states (Fader et al. 2004; Schwartz et al. 2014; Schweidel et al. 2014; Wedel 2000; Schweidel 2011).

The hidden Markov model is the state-of-the-art for modeling underlying processes (Netzer et al. 2008). It has also been widely applied in the marketing literature to model unobserved processes that guide customer behavior, such as the status of customers' relationships with firms (Netzer et al. 2008) and competitors' actions (Moon et al. 2007). However, the HMM is less sensitive to the change because it utilizes all past data to recover underlying states. The major disadvantage of using all past data is that the test statistics will take in all previous behavior variations in detecting the current life-change event. Consequently, we may end up with a low thus unconvincing probability in the life-change state because of the dilution from the previous behaviors variations. In other terms, in order for us to confirm the focal shift in the test, it will either require stronger signal to even up the prior variation or take longer time to detect the shift.

In sum, no extant method is able to dynamically select optimal observation windows for change detection while remaining simple and feasible for real-time analysis. Our solution, based on the CUSUM control chart, fills this gap.

Empirical Context

This study was conducted in the context of a Fortune 500 financial institution. (We thank the Wharton Customer Analytic Center as well as the sponsoring company for offering this data set.) Its managers are interested in detecting a specific type of major change in customers' career trajectories, a change that has great implications for managing a customer's portfolio. For reasons of confidentiality, we cannot reveal the specific career change of interest and the name of the bank in this case. Examples of this type of career change include leaving a previous job to attend graduate school; leaving a previous job to start a new company; and retiring. Such major career changes can fundamentally alter a person's financial situation, resulting in new needs for financial products. The solution is developed under the following data conditions, which are also generalizable to other customer management settings.

The type of life change to be detected is given. In this setting, managers have identified the type of life change that is important for the business. We can therefore extract the typical behavior, before and after life changes, from the historic data and use these patterns as signals of life changes.

The shift in behavior due to life change is small compared to the size of variances in behavior. Customers show large variance in behavior, both among customers and within a single customer's data. Customers tend to conduct many of the same activities at different times for different purposes, and these purposes are not necessarily related to the focal life events. A record of a customer buying a baby play yard, for example, is not a strong indicator of parenthood because the customer can purchase the same play yard for his or her friend's baby shower. A record of a series of purchases of items such as baby formulas, diapers, and toys over a month, however, *is* a strong indicator of parenthood. Consistent and systematic changes over a wide range of a given customer's behavior effectively distinguish a major life event from a one-time event. An effective solution, therefore, should utilize holistic behavioral patterns and accumulated evidence over time.

It is not known whether and when a change will occur in a customer's life. In an ideal case, it is most efficient to test for changes in behavior starting from the change point. In our setting, however, the change point is not known when conducting the detection, and the temporal range when the change point might occur can span over one or two decades. Because the shift is small, it is critical to select the optimal observation window for life change detection. While a short observation window might not contain enough behavioral evidence to confirm a life change, a long observation window might include observations before a life change. Lumping behavior that precedes a life change together with behavior following a life change can dilute evidence of change. An efficient solution should dynamically select an appropriate window tailored for each individual customer.

The analysis should be able to incorporate new data as they arrive and generate actionable intelligence in real time. The marketing data, such as those from customer management systems and social media, are generated continuously. It is desirable for the company to obtain the most up-to-date intelligence about customers. The algorithm therefore needs to be scalable for application to real-time analysis.

The data contain different types of factors that are indicative of a consumer's propensity to undergo the life change of interest. In the customer management context, we observe two types of factors that are related to life changes. One type of factors is behaviors; changes in behaviors reflect changes in customers' lives. The other type of factors is the conditions that influence the propensity for life changes. Examples of this factor are individual characteristics such as age, gender, and work experience. Analysis can exploit both factors to detect life changes.

Observed behavior data contain both continuous and discrete variables. Customer data contain variables of different types, including continuous and categorical variables. In order to capture the holistic pattern of behavior before and after life changes, analysis needs to account for correlations among variables of different types, as well as auto-correlations of behavior over time.

Historic data are available and contain information on the actual time of the life change for purposes of validation. The data set is then divided into two. A sample of the historic data can be used for calibration to capture behavior pattern before and after life changes. Another sample of the historic data can be used for testing to evaluate the performance of the proposed solution.

Data

The anonymized dataset contains observations on 98,088 randomly selected customers over seventeen months from January 2012 to May 2013. Only 12,982 customers remain in the study; the rest are excluded from analysis because of missing information. The majority

(81.44%) of the excluded samples lack information about customers' career changes.[2] This high percentage indicates that the company's managers have very little knowledge of customers' career changes, even though they considered this knowledge to be critical.[3]

In this data set, we observe a wide range of customer behavior on a monthly basis. These observations can be categorized into two types. One considers whether customers possess financial products at the bank; the other considers the number of customer contacts with the bank regarding financial products. We further group the financial products by their functionality: basic banking products, investments, loans, and insurances. We thereby obtain eight variables on customers' ownership and communications regarding each type of banking product. We single out the possession of auto insurance and checking accounts as two variables because these are the two most popular products and attract more than half of the bank's customers.

Table 1 presents the descriptive statistics of these variables before and after change. It shows a vivid feature of the data: the variances in customers' financial activities are large in comparison with the small shifts in behavior that result from career changes. For example, the mean frequency for customers to contact the bank is 1.579 before career changes. This number drops to 1.442 after a career change, showing a .107 decrease in frequency. However, the variance within the group of customers before a career change is 1.465 and after a career change is 1.469. No single variable, therefore, can serve as a strong indicator of customers' career changes. The challenge is to extract information from all the weak behavior indicators and accumulate the evidence over time to create an effective indicator of customers' career changes. This requires an advanced technique.

The data also provide information on dates when customers first started their original career and changed their career. Based on the data, the marginal probability of a career change at different times of their career is calculated and presented in Figure 2. As we can see in the diagram, customers at different stages of their careers have vastly different propensities for career

---

[2] Other missing information includes dates when customers start their careers and dates when they first become customers of the bank.
[3] Our data sponsors have put in great efforts to gather information on the career change statuses of their customers. A reasonable guess, therefore, is that the percentage of missing data on career change status is even larger than what we observed in this data set.

change. In particular, the probability of leaving the original career trajectory peaks in the fourth, fifth, sixth, and eighth years. Overall, the probability of a career change decreases as a customer's time in the career increases. Customers who stay on their original career path for more than fifteen years are very likely to stay on the same path until their retirement. The length of time since first taking the career can be therefore considered to be a factor that influences career change.

Detecting Life Changes Through a Control Chart Approach

Engineers solve a problem in the field of quality control similar to our task of detecting small systematic shift in customers' behavior. While machines produce parts with random errors, engineers need to detect consistent, small shifts away from the design standard to avoid deterioration in quality.[4] The CUSUM control chart is considered to be one of the most efficient tools for this problem. Because of its efficiency and simplicity, the CUSUM chart is also wildly applied in computer science (Lu and Tong 2009) and public health (Chandola et al. 2013) to monitor massive data for abnormal turmoil. For example, the CUSUM control chart is used in the Real-time Outbreak and Surveillance System (RODS) in Pennsylvania and Utah for public health surveillance. Its task is to monitor data from hospitals for anomalous patterns of syndromes outbreak (Tsui et al. 2003).

The CUSUM control chart is built on the sequential probability ratio test (SPRT) (Wald 1945). Unlike traditional hypothesis testing, in which the number of observations is determined in advance, SPRT allows the test statistic to be updated as new data become available. Given the level of type I and type II error, SPRT has been shown to be an optimal test because it requires the smallest expected number of observations (Wald and Wolfowitz 1948). This makes SPRT particularly fit for real-time analysis. The CUSUM control chart, based on SPRT, further improves its sensitivity of change detection by modeling a change point in the likelihood. This change point is unobserved; it is estimated from the data. The beauty of the CUSUM test is that this complicated formula eventually reduces to a simple scheme. We describe a simple example

[4] The essence of the task is to identify deviation from the norm, which can be either deterioration in quality or improvement in quality. Any control chart for the detection of deterioration can also be used to detect improvement. In this paper, we use only quality deterioration as an example.

from the context of manufacturing to provide a concrete view of the CUSUM control chart and its underlying logic.

A Univariate Example of the CUSUM Control Chart

Suppose a machine is designed to punch a hole one centimeter in diameter, but the machine produces holes with small errors. While random errors are inevitable, one-sided deviations are undesirable because they indicate a change in the machine's condition that requires corrective attention. To detect one-sided deviations, products are constantly sampled and the holes are measured. Table 2 presents two sequences of results produced by two machines, respectively.

Both sequences have the same results until period 7. Five of the first seven observations are larger than one, indicating possibility of one-sided deviation. The procedure of testing this hypothesis by CUSUM is as follows. Let μ be the mean of the holes. The goal is to test the null hypothesis, $\mu = 1$, meaning the holes are produced as designed, against the alternative hypothesis, $\mu > 1$, meaning the holes are larger than designed. Let x_t be the deviation from the design of observation at time t (measure of the hole -1), and the test statistic in the CUSUM control chart, Q_t, is as follows.

$$Q_t = \frac{\max\limits_{1 \le k \le t} \prod_{l=1}^{k-1} L_0(x_t) \prod_{l=k}^{t} L_1(x_t)}{\prod_{l=1}^{t} L_0(x_t)}$$

Here, $L_0(\cdot)$ is the likelihood function for x_t before change and $L_1(\cdot)$ is the likelihood function after change. The change point is represented by k. Because k is estimated by choosing the one time point among all past time points that maximizes the likelihood function under the alternative hypothesis. Assuming that the observation x_t follows identical independent normal distribution, the test statistic becomes:

$$Q_t = \max\{Q_{t-1} + x_t, 0\}$$

When t = 0, $Q_0 =$. Once Q_t is larger than a predetermined threshold, the alternative hypothesis is accepted; otherwise, monitoring of the production continues. Figure 3 presents the plot of the CUSUM control chart.

The formulation of this test statistic is in line with practical heuristics. Deviations from design are cumulated and summed over time as evidence of an upward shift in mean. In this way, deviations due to random errors in production cancel out, leaving the test statistic approximate to zero in the long run. In contrast, a consistent shift in mean will produce deviations consistently larger than zero and a test statistic larger than zero. Random errors are thereby distinguished from a consistent shift from the mean. Because cumulative sum is used as the test statistic, this method is named cumulative sum control chart. Furthermore, the formulation of the statistic allows an automatic inference of the most likely point at which the machine's condition had changed. Because the test statistic is the maximum between $Q_{t-1} + x_t$ and zero, any evidence supporting a downward deviation is discarded. In this way, the test statistic dynamically determines the time point when the evidence should start to be accumulated. Last but not least, although the observation windows are dynamically selected, the calculation of the CUSUM statistic remains simple. When new data arrive, a test statistic can be calculated based on the new data and the test statistic from the previous period. This feature is a great fit for real-time analysis as data continuously arrive.

As illustrated in this example, the CUSUM chart was originally developed for monitoring a single feature in the manufacturing process, which typically follows an independent and identically normal distribution. Its ability to accommodate rich observations and to adjust for individual or circumstantial factors that influence changes is therefore limited. These limitations largely constrain its applicability to customer management. While deviation from design can be measured in a manufacturing setting, it is unclear how to transfer various customer activities into a measure of deviation. Furthermore, although individual differences are seldom an issue in the manufacturing field, these factors can be informative in a customer management context as illustrated in the previous conceptual map (Figure 1), because many life changes (such as career change and marriage) are related to factors such as education and age.

The Proposed Solution

In this section, we provide a detailed description of how the test statistic can be extended to accommodate the problems faced in customer management settings. We first show how the module of likelihood function can be extended to extract behavioral patterns from a mix of continuous and categorical variables, such as binary and count variables. Second, we show how factors that influence life changes can be incorporated.

Likelihood of multivariate observations

Suppose we observe customer i's activities A_{il} at each of the time points l. Without loss of generality, suppose A_{il} contain three variables: A_{il}^1, A_{il}^2 and A_{il}^3. To capture the behavioral pattern before and after a life change, they are modeled as follows.

A_{il}^1 is a continuous variable, and is modeled using the simple linear regression.

$$A_{il}^1 = \alpha^1 + \beta^1 x_{il} + \gamma^1 A_{il-1}^1 + \varepsilon_{il}^1$$

A_{il}^2 is a binary variable, and is modeled using the probit model.

$$U_{il}^2 = \alpha^2 + \beta^2 x_{il} + \gamma^2 A_{il-1}^2 + \varepsilon_{il}^2$$

$$A_{il}^2 = \begin{cases} 0, & U_{il}^2 < 0 \\ 1, & U_{il}^2 \geq 0 \end{cases}$$

A_{il}^3 is a count variable and we can group the value of A_l^3 into s categories based on the distribution of A_l^3. The variable is then modeled using the ordered probit model.

$$U_{il}^3 = \alpha^3 + \beta^3 x_{il} + \gamma^3 A_{il-1}^3 + \varepsilon_{il}^3$$

$$A_{il}^3 = \begin{cases} 0, & U_{il} < 0 \\ 1, & 0 < U_{il} < \theta_1 \\ & \dots \\ s, & U_{il} \geq \theta_s \end{cases}$$

We account for the auto-correlation between A_{it} and A_{it-1} by adding A_{t-1} as a covariate of A_{it}. We can also control for the heterogeneity in customers' tendencies to conduct an activity by adding personal characteristics, x_{il}, as covariates of A_{it}. To allow for correlations between different activities, we assume that ε_{il}^1, ε_{il}^2, and ε_{il}^3 follow multivariate normal distribution.

$$\begin{pmatrix} \varepsilon_{il}^1 \\ \varepsilon_{il}^2 \\ \varepsilon_{il}^3 \end{pmatrix} = N(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{\Sigma}$ is a 3×3 variance covariance matrix.

The likelihood is:

$$L_i = f(\varepsilon_{il}^1) \cdot \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\varepsilon_{il}^2, \varepsilon_{il}^3 \mid \varepsilon_{il}^1)\, d\varepsilon_{il}^2 d\varepsilon_{il}^3$$

This integral can be calculated through the GHK simulator.

The model that we describe here provides a simple and intuitive way to capture customers' behavioral patterns. It does not assume any structural constraints on customers' behavior specific to the context of financial activities. Although we illustrate this using a three-variable case, our method can be easily extended to cases with more variables. This model thereby provides a general framework for modeling a set of customer behavior of different types. For specific problems, other models might capture customer behavior more precisely. In those cases, we can simply replace the specification of the likelihood with a specification from a better model, without affecting other parts of the proposed solution.

Sequential test of life change given only information of customers' behavior

Suppose we only observe information on customer behavior. The data structure is presented in Figure 4. The goal of the test is to determine whether a change has occurred by the current time point, t. Let C_i represent the time when a life change occurs for customer, i. We define the null hypothesis and alternative hypothesis as follows.

H_0: a life change has not occurred until current time point, t, i.e., $C_i > t$.

H_1: a life change has occurred before current time point, t, i.e., $C_i \leq t$.

For ease of representation, at time l when no change occurs, we define the probability distribution of A_{il} as $P_0(A_{il} \mid l < C_i)$. After a life change occurs to the consumer at time, C_i, the activity pattern changes and we define the new probability distribution of A_{il} as $P_1(A_{il} \mid l \geq C_i)$. Therefore, the maximum likelihood of customers' behavior under the null hypothesis is:

$$\prod_{l=1}^{t} P_0(A_{il} \mid C_i > t)$$

Under the alternative hypothesis, the life change occurs at time, k. The change point k is unknown from the data and is estimated by selecting the time point that maximizes the likelihood under the alternative hypothesis. The maximum likelihood under the alternative hypothesis is:

$$\max_{1 \leq k \leq t} \prod_{l=1}^{k-1} P_0(A_{il} \mid C_i = k) \prod_{l=k}^{t} P_1(A_{il} \mid C_i = k)$$

The ratio of the two likelihoods, Λ^{it}, is defined as:

$$\Lambda^{it} = \frac{\max_{1 \leq k \leq t} \prod_{l=1}^{k-1} P_0(A_{il} \mid C_i = k) \prod_{l=k}^{t} P_1(A_{il} \mid C_i = k)}{\prod_{l=1}^{t} P_0(A_{il} \mid C_i > t)}$$

The terms before time k cancel out. Λ^{t} becomes:

$$\Lambda^{it} = \max_{1\le k\le t} \prod_{l=k}^{t} \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} \mid C_i > t)}$$

In this way, the time window for the test statistic is dynamically selected: evidence of a change is accumulated from the time point at which the change most likely had occurred, and previous records are excluded from future analysis. If a change occurred at time, k, the observations can generally be better described by the post-change model $P_1(\cdot \mid C_i = k)$ than the pre-change model $P_0(\cdot \mid C_i > t)$, making Λ^{it} is greater than 1. Following the same line of reasoning, when a change did not occur before time, t, Λ^{it} is smaller than 1. This allows us to create a test statistic S_{it} as follows:

$$S_{it} = \max\{\ln\Lambda^{it}, 0\}$$

We are able to derive a recursive equation to describe the relationship between S_{it} and S_{it-1} that allow for agile computation when new data come. The derivation is presented in Appendix 1.

$$S_{it} = \max\left\{S_{it-1} + ln\frac{P_1(A_{it} \mid C_i = t)}{P_0(A_{it} \mid C_i > t)}, 0\right\}$$

The resulting recursive equation of the test statistic resembles that of the CUSUM control chart. Once a negative number appears, $S_{it-1} + ln\frac{P_1(A_{it} \mid C_i = t)}{P_0(A_{it} \mid C_i > t)}$ is immediately replaced by zero. This means that the data before time t show no tendency to shift upward and are discarded for the purpose of this test. In this way, the formula dynamically changes the window of observations for the evaluation of the alternative hypothesis with new data added to the test statistic each period, ensuring agile detection of a life change once it happens. The derivation of this recursive formula is critical for the implementation of the solution in big data and real-time analysis. Without this recursive computation scheme, the computation of the statistic would be a tedious job that requires the comparison of test statistics using different time points as the change point.

Sequential test of life change with additional information on influential factors

In some cases, companies also observe factors that influence the probability of a consumer experiencing a life change, such as age and education. A visualization of the data structure is presented in Figure 5. The relationships between these influential factors and life changes are typically modeled using a hazard model. Let Z_{il} represent a vector of factors that influence individual i's life changes.

$$H(Z_{il}) = P(C_i = l | C_i > l-1; Z_{il})$$

The structure of this model is different from the way we model the relationship between life changes and behavior, and it cannot be directly incorporated into the CUSUM framework. To exploit information on influential factors and behavior, we derive the test statistic as follows.

$$\Lambda^{it} = \max_{1\le k\le t} \frac{P(A_{il}, Z_{il}; l = 1, \dots, t |\ C_i = k)}{P(A_{il}, Z_{il}; l = 1, \dots t |\ C_i > t)}$$

$$= \max_{1\le k\le t} \prod_{l=k}^{t} \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} |\ C_i > t)} \times \frac{P(Z_{il}; l = 1, \dots, t | C_i = k)}{P(Z_{il}; l = 1, \dots, t | C_i > t)}$$

$$= \max_{1\le k\le t} \prod_{l=k}^{t} \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} | C_i > t)} \times \frac{P(C_i = k | Z_{il}; l = 1, \dots, t)\ P(Z_{il}; l = 1, \dots, t)/P(C_i = k)}{P(C_i > t | Z_{il}; l = 1, \dots, t)\ P(Z_{il}; l = 1, \dots, t)/P(C_i > t)}$$

Because $P(C_i = k | Z_{il}, l = 1, \dots, t) = \prod_{l=1}^{t-1}(1 - H(Z_{il})) \times H(Z_{it})$, and $P(C_i > t | Z_{il}, l = 1, \dots, t) = \prod_{l=1}^{t}(1 - H(Z_{il}))$. We further derive the statistic as follows.

$$= \max_{1\le k\le t} \prod_{l=k}^{t} \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} | C_i > t)} \times \frac{H(Z_{ik})\ /P(C_i = k)}{\prod_{l=k}^{t}(1 - H(Z_{il}))\ /P(C_i > t)}$$

The test statistic therefore becomes:

$$S_{it} = max\left\{\max_{1\le k\le t} ln \prod_{l=k}^{t} \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} | C_i > t)} \cdot \frac{H(Z_{ik})\ /P(C_i = k)}{\prod_{l=k}^{t}(1 - H(Z_{il}))\ /P(C_i > t)}, 0\right\}$$

$$= max\left\{\max_{1\le k\le t} \left\{\sum_{l=k}^{t} ln \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} | C_i > t)(1 - H(Z_{il}))} + \ln \frac{H(Z_{ik})\ P(C_i > t)}{P(C_i = k)}\right\}, 0\right\}$$

We are able to derive the recursive relationship between the test statistics S_{it} and S_{it-1} with the additional component of individual characteristics. The derivation is presented in the appendix 2. Define S_{it}^* as $S_{it}^* = \max_{1\le k\le t}\left\{\sum_{l=k}^{t} ln \frac{P_1(A_{il} \mid C_i = k)}{P_0(A_{il} | C_i > t)(1 - H(Z_{il}))} + \ln \frac{H(Z_{ik})\ P(C_i > t)}{P(C_i = k)}\right\}$.

The recursive relationship is as follows. This recursive relationship simplifies the computation scheme, making the proposed solution fit for real-time analysis in big data.

$$S_{it}^* = \max\left\{S_{it-1}^*, ln \frac{H(Z_{it})\ P(C_i > t-1)}{P(C_i = t)}\right\} + ln \frac{P_1(A_{it} | C_i \le t)\ P(C_i > t)}{P_0(A_{it} | C_i > t)\ (1 - H(Z_{it}))\ P(C_i > t-1)}$$

Test procedure

The implementation of our solution requires three steps. First, typical patterns of customers' behavior before and after life changes are extracted from a calibration sample. Second, the test statistic is calculated using the recursive formula derived above. Third, the test statistic is compared to a predetermined threshold, h. The decision rules are as follows.

If $S_{it} \geq h$, reject the null hypothesis and report a life change;

If $S_{it} < h$, do not reject the null hypothesis and continue monitoring for a change.

A given value of the threshold corresponds to a pair of true positive and false positive rates. A higher threshold h makes it easier to detect a life change, but induces a higher risk of falsely reporting a life change. Inversely, a lower threshold h reduces the risk of falsely reporting a life change while also making it more difficult to detect a life change.

We have now provided a general framework for the detection of major life changes in the context of customer management. This framework dynamically selects the optimal window for accumulating information from customers' behavior and individual characteristics for the detection of major life changes. More importantly, the framework simplifies the computation scheme so that the resulting test statistic fits within a big data scenario. In real-time analysis, where data is continuously flowing in and agile decision support is needed, the recursive formula that we derive allows us to effortlessly compute the updated statistic at time t given the test statistic from the previous period.

Empirical Analysis

Parameter estimation

Our test statistic is composed of three parts and their parameters are estimated respectively. The first part extracts behavioral patterns before and after the career change. We jointly model the ten behavioral indicators listed in Table 1 using the framework proposed previously. Particularly, the four count variables are transformed into ordered categorical variables of three levels. The top level contains the possession of two or more products in the category because most people own no products or one product in a product category, as illustrated in Table 3.

A sample of 3,035 customers are set aside as a calibration sample. Two sets of estimates are obtained based on observations from before and after a life change, each describing behavioral patterns before or after a career change. The estimates of the parameters are presented in the Appendix. Similarly to what we observed in Table 1, the difference in observations from before and from after a career change regarding each particular activity is small. We use the remaining records of 9,947 customers as a test sample to validate the method.

The second part of our test statistic captures the hazard of change at different times in a customer's career. As we observed in Figure 2, the relationship between the length of time in the career and the probability of change is quite complex and is hard to describe using simple models such as a linear or log linear relationship. Without further knowledge of the context, we estimated a non-parametric hazard model so that we do not need to impose any assumption when describing this relationship.

The third part of our test statistic is the parameters regarding the marginal probability of career change each month. In this study, we assume that customers change their career paths at the same rate, and model it as the average monthly probability of change. In cases where the change rates are different across months, the differences can be incorporated in the test statistic by modeling the marginal probability in either a parametric or non-parametric way.

Benchmark model

As stated before, we choose the HMM as the benchmark model because it is considered the state-of-the-art framework for modeling behavior and its unobserved processes. Although the HMM has not previously been applied to infer customers' life changes, a popular approach in practice is to use logistic regression, and logistic regression can be considered a simple special case of the HMM. With additional flexibility in its structure, the HMM will perform better than the logistic regression. We therefore consider the HMM to be a good benchmark model and apply it in this study to evaluate the performance of the proposed solution. Using the HMM, life states (no change vs. change) are modeled as the unobserved states that guide customers' financial behavior. Given the life states, the calibration of the HMM reduces to the modeling of two independent components of the binary transition process and the conditional likelihood. The conditional likelihood describes customers' behavior before and after life changes. The binary transition process is modeled by the non-parametric model, which describes the relationship between the tendency to change and time in career. In this way, the HMM is applied on the same data structure and uses the same information as the proposed model does. The unobserved states are then inferred by calculating the probabilities of being in different life states (see details about the inference in Netzer et al. 2008). The inferred probability is then compared with a predetermined threshold. Customers with probabilities higher than the cut-off point are considered as having changed their careers. Similarly, when using the HMM, the determination of the threshold is also a trade-off between type I and type II errors.

The performance of the HMM and the proposed sequential test are only comparable when the transition process between life stages are modeled. The proposed solution is still able to function without modeling the transition process. In contrast, the HMM requires imposing additional assumptions in order to model the transitional process. In this case, the efficiency of the model would be harmed by wrong assumptions about the transition process.

Performance comparison

To compare the performance of the proposed sequential test with that of the HMM, we create receiver operating characteristic (ROC) curves for both methods. ROC curves have been widely applied in the statistics field as a tool to evaluate the performance of a binary classification system when the threshold varies (Hanley 2005; Metz 1978). To create an ROC curve, the true positive rate is plotted against the false positive rate at various levels of the threshold.

Figure 6 presents the results. The line of no discrimination—the dotted diagonal line stretching from the left bottom to the top right corners—represents results from random guesses, regardless of the base rate of life changes. A curve above this line indicates better performance than a random guess, whereas a curve below the line indicates a worse performance. The results from the HMM are plotted using dashed lines, showing a substantial improvement compared to random guess. This fine performance is expected from the HMM since it is the state-of-the-art model for understanding the unobserved underlying process of customer behavior. With the efficient design of the test statistic, the performance of the proposed solution surpasses that of the HMM. The area between the curve of the proposed solution and the HMM is .0556, representing an 8.39% increase in performance.

Simulation Study

Our data is limited in both its length and variety of observations. Consequently, we cannot assess the performance of our solution in different contexts. We therefore resort to the simulation experiments. The simulation starts when customers become members of the bank and their behavior is recorded. To conduct the simulation, we first generate customers' length of experience at work at the time they became customers of the bank. Based on these individual differences, we generate the time when customers undergo life changes. Finally, we generate customers' behavior before and after change based on the parameters from the empirical data.

Assessing the effectiveness of the proposed sequential test of life change

The purpose of the first simulation is to validate the results using simulation experiments and to assess the performance of the proposed solution in the long term. We simulated the behavior of 1,000 customers at the monthly level and monitored their behavior for indicators of life changes over five years.

The results are presented in Figure 7. In the plot, the curve of the proposed solution is above that of the HMM, meaning that given the same level of false positive rate, the proposed model correctly detects more life changes than the HMM. The area between the curve of the proposed solution and the HMM is .134, representing an 18.55% increase in performance. This result validates that the proposed solution performs better than the HMM in detecting life changes. To further demonstrate the difference in the performance of the proposed model and the HMM, we provide examples of two cohorts of customers in Figure 8. The first cohort is composed of 101 customers who undergo life changes at the fortieth period of the surveillance in the simulation study (Figure 8-a). We observe that fourteen customers are reported by the proposed solution as changed before period 40. These false detections take place only a few periods before the actual change. In contrast, thirty-seven customers are reported by the HMM as having changed and the false detections take place long before the actual change as early as the second period of the surveillance. After the occurrence of a life change, the proposed sequential test quickly detects the changes within nine periods. It takes the HMM, however, another thirty-five periods to detect all changes. A similar pattern is observed in Figure 8-b, which features customers who undergo life changes at period 70.

The HMM performs worse than the proposed model because of fundamental differences in how the two statistics are constructed. The first difference is that while the HMM employs all past history to infer the change, the proposed solution dynamically selects the time window to test changes in behavior. Because the shift in patterns of behavior after the life change is small, when the HMM is applied, the evidence of change is vulnerable to dilution by observations before a change. The second difference is that while the proposed solution uses the hazard of change to gauge the baseline probability of change at a given time point, the HMM uses the cumulative probability of change. In this specific case, the hazard of change at the forty-eighth month of a career on average is thirty times that of any previous periods. However, in terms of cumulative probability, the overall probability of change increases from 20.29% to 28.76% at period 40, which is a much weaker signal of change compared to that of the marginal probability. These differences distinguish the two methods in their ability to detect changes based on the same information.

Assessing the performances when the probability of change is low

We then evaluate the performance of the proposed solution at different levels of probability of change. In these simulation experiments, customers are monitored for 240 periods, and their marginal probabilities of life change per period are the same. We adjust the marginal probability of change to 1/300 and 1/2500 respectively. We present the results in Figure 9.

When the marginal probability is 1/300 (Figure 9-a), the area between curves is .0464. This area increases to .1586 when the marginal probability of change decreases to 1/2500 (Figure 9-b). The area between the two curves becomes bigger as the marginal probability of change decreases. The result shows that the performance gained from choosing the proposed solution over the HMM becomes larger when the probability of change decreases.

These simulations demonstrate the advantage of the proposed solution in a customer management setting. In general, in such a setting, the probability of life change at a given time point is low, and the possible time for life changes span over a long period of time. It is in such a setting that conducting real-time analysis and extracting information from big data becomes valuable. Otherwise, in extreme cases where the possibility of life change can be narrowed down to one or a few given time points, there is no need to constantly monitor customer behavior and dynamically identify the window for detection.

Assessing the performances given a larger shift in behavior patterns due to life change

We continue to evaluate the performance of the proposed solution when shifts in behavioral pattern due to life changes become large. In these experiments, we vary the shift in behavior patterns by increasing the differences in the estimates of the intercepts before and after changes, while holding other parameters the same. The difference in intercept is increased to .05 and .30 respectively. The results are shown in Figure 10.

The area between curves is .0913 given the original set of parameters. This area reduces to .0471 when the difference is increased by .015. The area continues to shrink to .0266 when the difference is increased to .3. As the shifts in behavioral patterns increase, the area between the two curves becomes smaller.

This set of simulations demonstrates the advantage of the proposed solution when the shift in behavioral patterns is small. Small shifts in behavior are more difficult to detect when records from before and after a life change are lumped together. This becomes less of a problem when shifts in behavior become large. Strong behavioral signals improve the performance of both methods. In such cases, the performance of the HMM is already good, leaving little room for further improvement. The performances of the two methods thus converge.

Managerial Insights

We conduct an additional simulation to demonstrate the application of the proposed solution to assisting managerial decision making. The purpose is twofold. First, we evaluate the gain in profit when the proposed method is applied. Second, we demonstrate how managers can decide whether to apply the proposed solution to detect life changes and how managers can determine the optimal value of the threshold in a specific problem.

In these simulations, we simulate 10,000 customers and monitor their life changes for 240 periods. The marginal probability of undergoing a life change is 1/1000 for each period. Suppose the companies take actions immediately after receiving alerts of life changes, and each action costs one dollar. We conduct two sets of simulations by manipulating the loss function in two ways.

In the first set of these simulations, we evaluate the proposed solution under different shapes of loss function. We assume the maximum revenue is one hundred dollars and can be obtained if changes are detected once they occur. As the gap between the detection time and the actual change point increases, the revenue decreases. We vary the rate that the return decreases and obtain two different scenarios, which are illustrated in Figure 11. In scenario 1 (Figure 11-a), the opportunity window after life change is long, while taking action before life changes does not earn much reward. In contrast, in scenario 2 (Figure 11-b), the opportunity window soon close after life change, but taking actions before life change can earn more reward.

To obtain the optimal value of the threshold, we first conduct simulations under different values of the threshold and then identify the optimal value through grid search. The results from the scenarios, including the maximum profits and their corresponding false positive rates, are summarized in Table 4. The results show that in both scenarios, applying the proposed solution brings in more profit than the HMM method. Compared to the HMM method, the proposed solution brings an increase of 76.98% in profit in scenario 1 and an increase of 53.44% in scenario 2. Table 4 also shows how the value of the threshold should be adjusted based on different loss function. In scenario 1, where taking action before life change receives little reward and the opportunity window is long after life change, the threshold should be set higher to avoid type I error. In scenario 2, where taking action before a life change is rewarding and the opportunity window closes soon after life change, the threshold should be set lower to avoid type II error.

We further illustrate the relationship between the threshold and the profit in Figure 12. Figure 12 shows how the change in the threshold influences the expected profit given the loss function in scenario 1. The relationships between the false positive rate and the expected profit show an Inverted U-shaped pattern. For the HMM, the profit reaches its peak at 23,659 when the false positive rate is set at .35. For the proposed sequential test, the profit reaches its peak at 41,872 when the false positive rate is set at .36. In sum, each value of the threshold represents a different trade-off between type I and type II error, and managers can identify the optimal value of the threshold based on the structure of cost and revenue in their specific applications.

In the second set of simulations, we evaluate the value of the proposed solution given different levels of returns. The setting is similar to the revenue structure in scenario 1, in which the revenue decreases to zero in ninety periods before change and thirty periods after change. We then manipulate the maximum revenue from one dollar to twenty dollars, while the cost of actions remains one dollar.

The results are shown in Figure 13. When the maximum revenue is as high as twenty dollars, the total profit from the pool of 10,000 customers reaches 2,709 dollars. The profit decreases as the expected revenue from the opportunities of life changes decreases. When the maximum revenue per case further decreases to twelve dollars, the maximum profit, which is zero, is reached when the false positive rate is zero. This means that under this cost and revenue structure, the costs of false identifications is larger than the benefits of correct identifications, making it unprofitable to apply the method to detect changes.

In sum, managers need to weigh the costs and benefits in order to determine whether to use a data-driven approach to identify potential business opportunities. Based on the revenue and cost structure, managers must then decide the optimal trade-off between false positive and true positive rates. Generally, in customer management, the return is high compared to the cost of action and the proposed solution is useful.

Discussion

Advanced big data analytics are becoming a critical driver of growth in customer value. Big data enables us to transform massive quantities of customer information into useful customer intelligence in real time to assist managerial decisions and thereby capture valuable business opportunities. Our study contributes to big data analytics by transforming copious customer data into critical information on customers' life changes. To achieve this goal, we propose a sequential test based on the framework of the CUSUM control chart and extend the test statistic in order to apply it to customer management. This solution provides a general framework for the problem of life change detection. In doing so, this paper introduces to the marketing field a new perspective in the area of change detection: a scalable approach for dynamically optimizing the window for detecting change at the individual level. We make this solution applicable to the context of customer management by accounting for the impact of individual factors in life transitions, as well as by illustrating ways of modeling multiple behavioral indicators of different types along with their correlations and autocorrelations. We thereby also contribute to the field of statistical process control by providing a way to adjust for individual and circumstantial differences in the test statistic. Despite the additional complexity, we are still able to derive a recursive formula for the test statistic, making the proposed solution particularly fit for application to big data. Our solution offers superior performance compared with the benchmark model. We also illustrate the value of its improved performance compared to the HMM under different loss functions.

Our study benefits managers by providing a practical tool for monitoring customers' major life changes in real time. This intelligence opens up precious opportunities for managers. Knowing, for example, that the customer has a newborn baby in the house, a grocery store can target the customer for diapers and formula, which are often "destination" products that drive traffic to the store. This intelligence is even more important for durable goods, such as cribs and car seats. Households purchase these products only once or twice in a lifetime and there is hardly any historical data for retailers to use to customize their marketing efforts at the individual level. The beauty of our solution is that it builds on the intuitive idea of the control chart, a notion that is easy for managers to grasp. Its easy computation scheme also makes it friendly for real-life application. Beyond detecting life changes, the proposed method can also be used to detect other types of change that are of interest to managers. Examples include detecting changes in customers' preferences in grocery shopping and changes in customers' risk of defaulting on credit card and mortgage payments.

In the specific context of banking data, the performance of our model can be further improved in two ways. One way would be to obtain more data, including a longer observation window and more detailed information on customers' activities. Our data lasts only seventeen months, which is a short window compared with customers' lifetimes at the bank. It limits our ability to incorporate other dynamics in customers' behavior when modeling behavior before and after a career change. Our data on customers' financial products is also limited to the possession of products in a given month. For example, in real life, banks observe a much richer set of information, including account balance and activities such as deposits, withdrawals, and transfers. Such additional information is also indicative of career changes. The other way is to include more existing knowledge about the financial products and specific career changes when modeling customers' behavior. It is not our aim, however, to dig deep into the data on customers' financial behavior. We keep the modeling of customers' behavior simple because our goal is to provide a general framework applicable to the type of problem that requires detection of a major life change. In a real-life application, our framework can be extended to incorporate managerial insights as well as rich findings from existing literatures to improve the performance of the test.

When applied to specific problems, the proposed solution is subject to two limitations. One limitation is in its ability to capture high dimensional data, containing thousands of variables. In such cases, a Bayesian method can be used to capture the correlations among the large number of variables. Another limitation is that we assume that customers take on a new behavioral pattern at the time of life change. In real life, customers might gradually migrate to the new pattern or present some abnormal patterns before settling down with a new pattern. Future research can account for these patterns to further improve the solution's performance.

APPENDIX 1:
DERIVATION OF THE RECURSIVE FORM GIVEN ONLY OBSERVATIONS ON CUSTOMERS' BEHAVIOR

$$\max_{1\le k\le t} \sum_{l=k}^{t} ln \frac{P_1(A_{il}|\ C_i = k)}{P_0(A_{il}|C_i > t)}$$

$$= \max\left\{\max_{1\le k\le t-1} \sum_{l=k}^{t} ln \frac{P_1(A_{il}|\ C_i = k)}{P_0(A_{il}|C_i > t)},\ ln \frac{P_1(A_{it}|C_i = t)}{P_0(A_{it}|C_i > t)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1} \sum_{l=k}^{t-1} ln \frac{P_1(A_{il}|\ C_i = k)}{P_0(A_{il}|C_i > t)} + ln \frac{P_1(A_{it}|\ C_i = t)}{P_0(A_{it}|C_i > t)},\ ln \frac{P_1(A_{it}|\ C_i = t)}{P_0(A_{it}|C_i > t)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1} \sum_{l=k}^{t-1} ln \frac{P_1(A_{il}|\ C_i = k)}{P_0(A_{il}|C_i > t)},\ 0\right\} + ln \frac{P_1(A_{it}|\ C_i = t)}{P_0(A_{it}|C_i > t)}$$

$$= S_{it-1} + ln \frac{P_1(A_{it}|\ C_i = t)}{P_0(A_{it}|C_i > t)}$$

APPENDIX 2:
DERIVATION OF THE RECURSIVE FORM GIVEN BOTH OBSERVATIONS ON CUSTOMERS' BEHAVIOR AND INFLUENTIAL FACTORS

$$S_{it}^{*} = \max_{1\le k\le t}\left\{\sum_{l=k}^{t} ln \frac{P_1(A_{il}\ |\ C_i = k)}{P_0(A_{il}\ |C_i > t)(1 - H(Z_{il}))} + \ln \frac{H(Z_{ik})\ P(C_i > t)}{P(C_i = k)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1}\left\{\sum_{l=k}^{t} ln \frac{P_1(A_{il}|\ C_i = k)}{P_0(A_{il}|C_i > t)\ (1 - H(Z_{il}))} + \ln \frac{H(Z_{ik})\ P(C_i > t)}{P(C_i = k)}\right\}, ln \frac{P_1(A_{it}|C_i = t)\quad H(Z_{it})\qquad P(C_i > t)}{P_0(A_{it}|C_i > t)(1 - H(Z_{it})\ )P(C_i = t)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1}\left\{\sum_{l=k}^{t-1} ln\frac{P_1(A_{il}|\, C_i=k)}{P_0(A_{il}|\, C_i>t)\,(1-H(Z_{il}))} + \; ln\frac{P_1(A_{it}|C_i=k)}{P_0(A_{it}|C_i>t)\,(1-H(Z_{it}))}\right.\right.$$

$$\left.\left. + \ln\frac{H(Z_{ik})P(C_i>t-1)P(C_i>t)}{P(C_i=k)P(C_i>t-1)}\right\}, ln\frac{P_1(A_t|C_i=t) \quad H(Z_{it}) \quad P(C_i>t)}{P_0(A_t|C_i>t)(1-H(Z_{it})\,)P(C_i=t)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1}\left\{\sum_{l=k}^{t-1} ln\frac{P_1(A_{il}|\, C_i=k)}{P_0(A_{il}|C>t)\,(1-H(Z_{il}))} + \ln\frac{H(Z_{ik})\,P(C_i>t-1)}{P(C_i=k)}\right\} + \ln\frac{P(C_i>t)}{P(C_i>t-1)}\right.$$

$$\left. + \; ln\frac{P_1(A_{it}|C_i\le t)}{P_0(A_{it}|C_i>t)\,(1-H(Z_{it}))}\;, ln\frac{P_1(A_{it}|C_i=t) \quad H(Z_{it}) \quad P(C_i>t)}{P_0(A_{it}|C_i>t)(1-H(Z_{it})\,)P(C_i=t)}\right\}$$

$$= \max\left\{\max_{1\le k\le t-1}\left\{\sum_{l=k}^{t-1} ln\frac{P_1(A_{il}|\, C_i=k)}{P_0(A_{il}|C_i>t)\,(1-H(Z_{il}))}\right.\right.$$

$$\left.\left. + \ln\frac{H(Z_{ik})\,P(C_i>t-1)}{P(C_i=k)}\right\}, ln\frac{H(Z_{it})\,P(C_i>t-1)}{P(C_i=t)}\right\}$$

$$+ \; ln\frac{P_1(A_{it}|C_i\le t)}{P_0(A_{it}|C_i>t)\,(1-H(Z_{it}))} + \ln\frac{P(C_i>t)}{P(C_i>t-1)}$$

$$= \max\left\{S^*_{it-1}, ln\frac{H(Z_{it})\,P(C_i>t-1)}{P(C_i=t)}\right\} + \; ln\frac{P_1(A_{it}|C_i\le t)\;\; P(C_i>t)}{P_0(A_{it}|C_i>t)\,(1-H(Z_{it}))\,P(C_i>t-1)}$$

References

Andreasen, A. (1984), "Life Status Changes and Changes in Consumer Preferences and Satisfaction," *Journal of Consumer Research*, 11 (3), 784–94.

Baek, E. and G. Hong (2004), "Effects of Family Life-Cycle Stages on Consumer Debts," *Journal of Family and Economic Issues*, 25 (3), 359–85.

Benzies, K., S. Tough, K. Tofflemire, C. Frick, A. Faber, and C. Newburn-Cook (2006), "Factors Influencing Women's Decisions About Timing of Motherhood," *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 35 (5), 625–33.

Chandola, V., S. Sukumar, and J. Schryver (2013), "Knowledge discovery from massive healthcare claims data," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. Chicago, Illinois, USA: ACM.

Cocco, F., F. Gomes, and P. Maenhout (2005), "Consumption and Portfolio Choice over the Life Cycle," *Review of Financial Studies*, 18 (2), 491–533.

Du, Y. and W. Kamakura (2006), "Household Life Cycles and Lifestyles in the United States," *Journal of Marketing Research*, 43 (1), 121–32.

Fader, P., B. Hardie, and H. Chun-Yao (2004), "A Dynamic Changepoint Model for New Product Sales Forecasting," *Marketing Science*, 23 (1), 50–65.

Gourinchas, P. and J. Parker (2002), "Consumption over the Life Cycle," *Econometrica*, 70 (1), 47–89.

Hanley, A. (2005), "Receiver Operating Characteristic (ROC) Curves," in Encyclopedia of Biostatistics. John Wiley & Sons, Ltd.

Helsen, K., Schmittlein D (1999), "Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models," *Marketing Science*, 11(4) 395-414.

Kreyenfeld, M. (2010), "Uncertainties in Female Employment Careers and the Postponement of Parenthood in Germany," *European Sociological Review*, 26 (3), 351–66.

Lansing, J. and Leslie K. (1957), "Family Life Cycle as an Independent Variable," *American Sociological Review*, 22 (5), 512–19.

Liao, T. (2005), "Clustering of time series data—a survey," *Pattern Recognition*, 38(11) 1857-1874.

Lu, W. and H. Tong (2009), "Detecting Network Anomalies Using CUSUM and EM Clustering," in Advances in Computation and Intelligence, Zhihua Cai and Zhenhua Li and Zhuo Kang and Yong Liu, eds. Vol. 5821: Springer Berlin Heidelberg.

Mathur, A., G. Moschis, and E. Lee (2008), "A longitudinal study of the effects of life status changes on changes in consumer preferences," *Journal of the Academy of Marketing Science*, 36 (2), 234–46.

Metz, C. (1978), "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, 8 (4), 283–98.

Moon, S., W. Kamakura, and J. Ledolter (2007), "Estimating Promotion Response When Competitive Promotions Are Unobservable," *Journal of Marketing Research*, 44 (3), 503–15.

Morrison, D. (1969), "On the Interpretation of Discriminant Analysis," *Journal of Marketing Research*, 6 (2), 156–63.

Netzer, O., J. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27 (2), 185–204.
Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41 (1/2), 100–15.

Perron, P., (1989), "The great crash, the oil price shock and the unit root hypothesis," *Econometrica*, 57, 1361-1401.

Punj, G. and D. Stewart (1983), "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, 20 (2), 134–48.

Schwartz, E., E. Bradlow, and P. Fader (2014), "Model Selection Using Database Characteristics: Developing a Classification Tree for Longitudinal Incidence Data," *Marketing Science*, 33 (2), 188–205.

Schweidel, D., E. Bradlow, P. Fader. (2011), "Portfolio Dynamics for Customers of a Multiservice Provider," *Management Science*, 57(3) 471-486.

Schweidel, D., P. Young-Hoon, and Z. Jamal (2014), "A Multiactivity Latent Attrition Model for Customer Base Analysis," *Marketing Science*, 33 (2), 273–86.

Sood, A., G. Tellis (2009), "Do Innovations Really Pay Off? Total Stock Market Returns to Innovation," *Marketing Science*, 28(3) 442-456.

Tsui, F., J. Espino, V. Dato, P. Gesteland, J. Hutman, and M. Wagner (2003), "Technical Description of RODS: A Real-time Public Health Surveillance System," *Journal of the American Medical Informatics Association*, 10 (5), 399–408.

Wald, A. (1945), "Sequential Tests of Statistical Hypotheses," *The Annals of Mathematical Statistics*, 16 (2), 117–86.

Wald, A. and J. Wolfowitz (1948), "Optimum Character of the Sequential Probability Ratio

Test," *The Annals of Mathematical Statistics*, 19 (3), 326–39.

Wedel, M. (2000), *Market segmentation: Conceptual and methodological foundations*. Springer Science & Business Media.

Wiles, M., S. Jain, S. Mishra, C. Lindsey (2010), "Stock Market Response to Regulatory Reports of Deceptive Advertising: The Moderating Effect of Omission Bias and Firm Reputation," *Marketing Science*, 29(5) 828-845.

Wilkes, R. (1995), "Household Life-Cycle Stages, Transitions, and Product Expenditures," *Journal of Consumer Research*, 22 (1), 27–42.

Figure 1: Conceptual Framework for Proposed Detection of Life Change

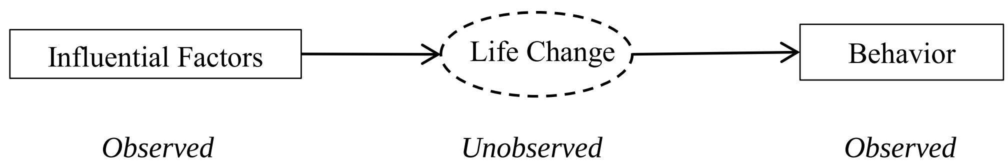


Figure 2: Marginal Probability of Career Change over the Length of the Career

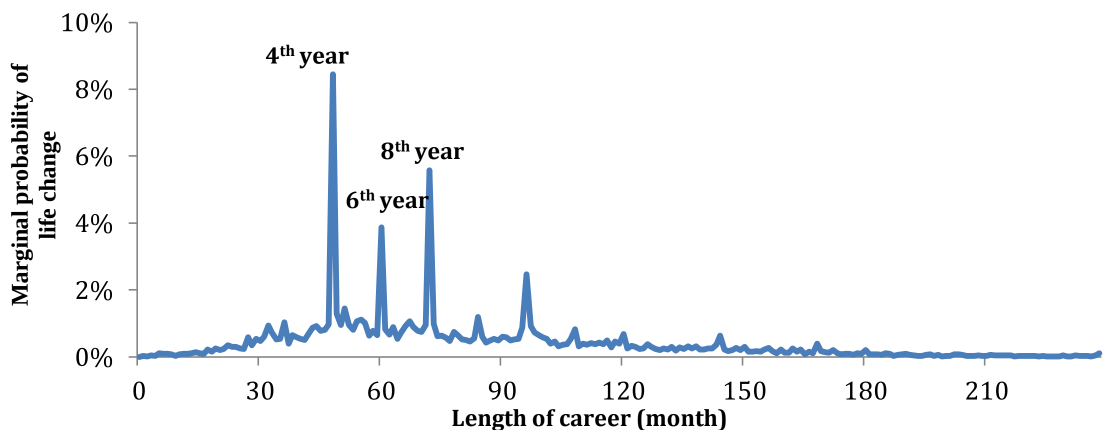


Figure 3: Control Charts for the Two Sequences of Observations

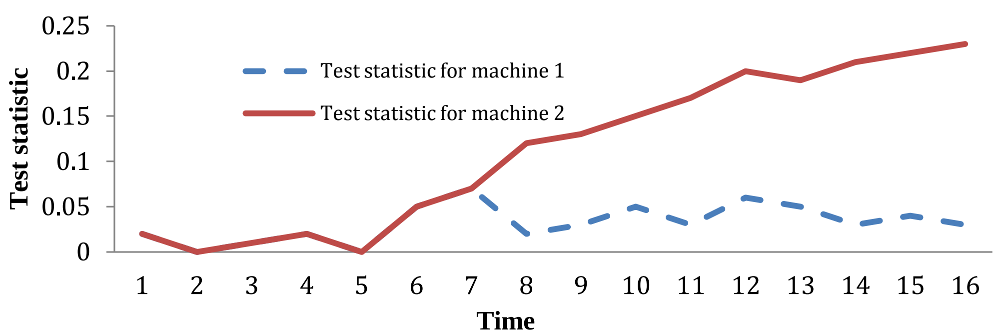


Figure 4: Data Structure given Only Information of Customer Behavior

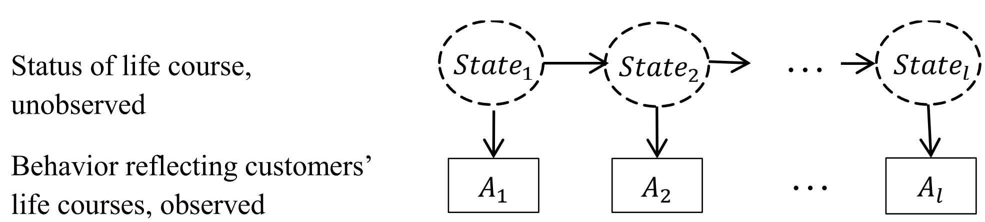


Figure 5: Data Structure given Both Influential Factors and Customers' Behaviors

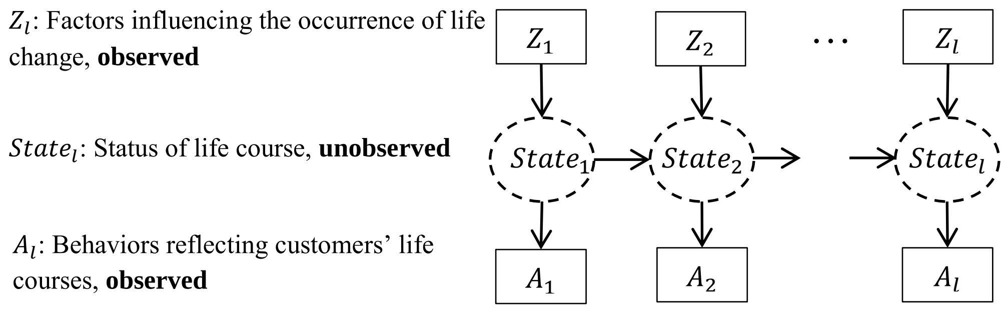
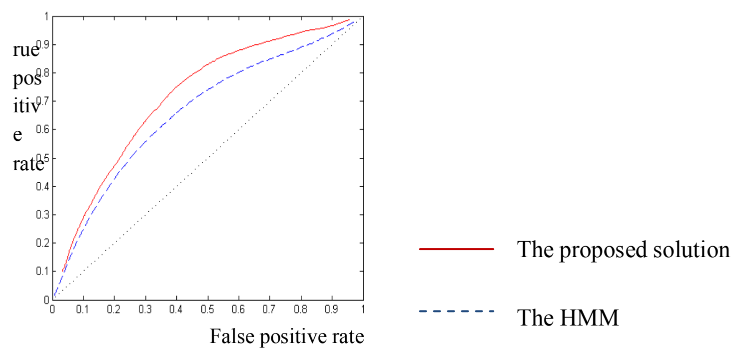


Figure 6: Comparing the Performance of the Proposed Sequential Test with the Hidden Markov Framework using Empirical Da-



ta

Figure 7: Comparing the Performance of the Proposed Solution with the Hidden Markov Model in Simulation Experiment

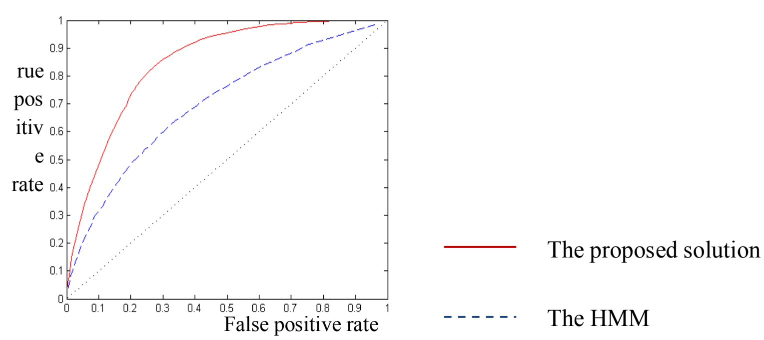


Figure 8: Illustrative Cases

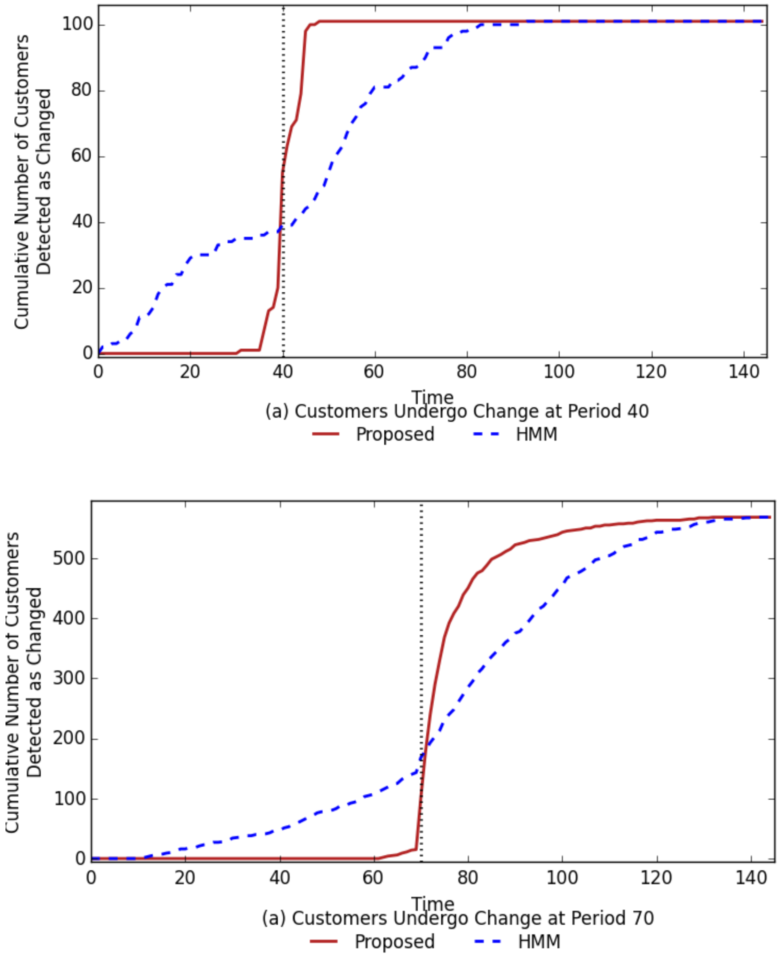
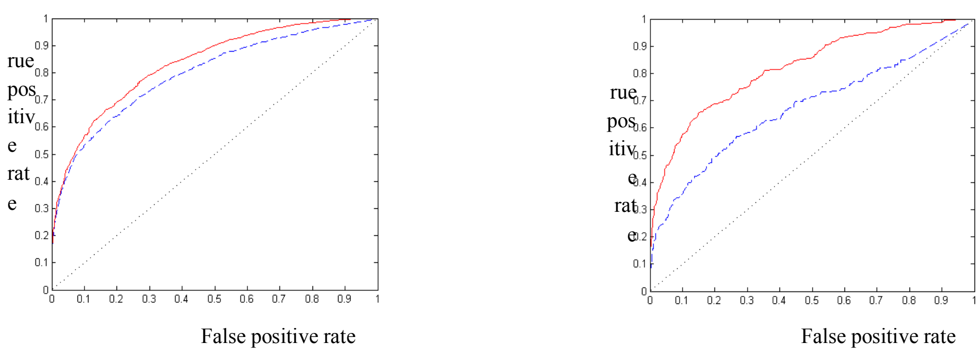
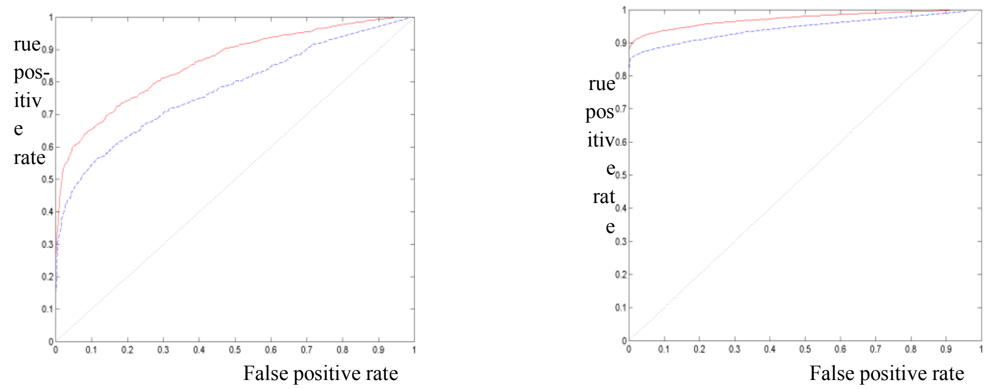


Figure 9: ROC Curves When Probabilities of Change are Different



(a) Marginal Probability of Change = 1/300 **(b) Marginal Probability of Change = 1/2500**

Figure 10: ROC Curves When Shifts in Behavior Patterns are Different



(a) +.05 shift **(b) +.30 Shift**

Figure 11: Loss Functions in Different Scenarios

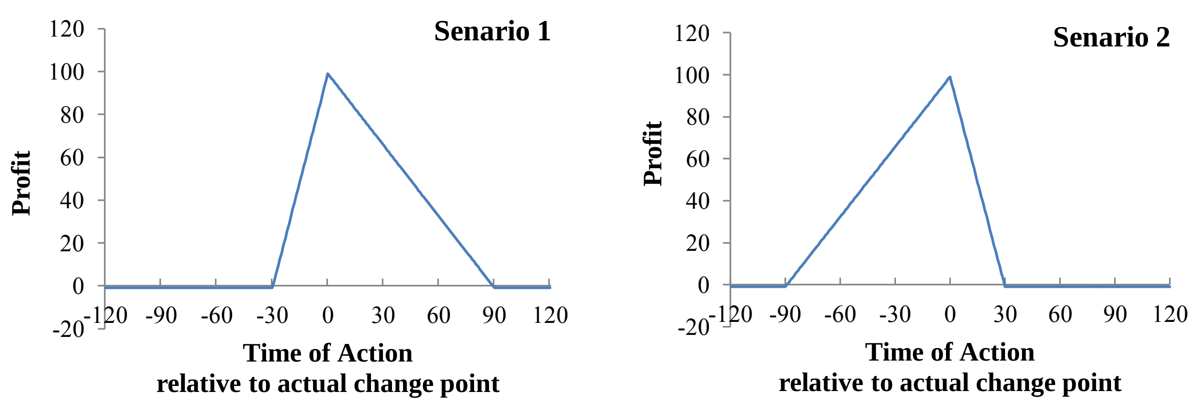


Figure 12: Profit at Different Levels of False Positive Rate under the Loss Function in Scenario1

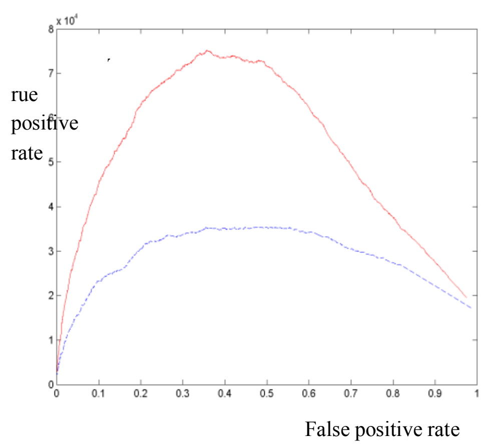


Figure 13: Total Profits from a Pool of 10,000 Customers

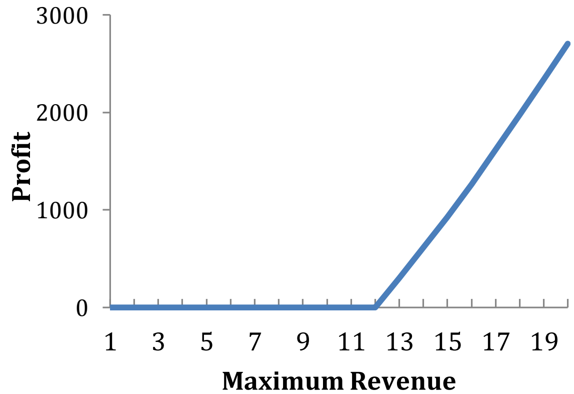


Table 1: Descriptive Statistics of Observed Customer Activities

	Description	Before Change	After Change	Difference in Mean
Products owned by customers	Whether the customer owns any auto insurance product at the bank	**.695** *(.460)*	**.588** *(.492)*	.107
	Whether the customer owns any checking account at the bank	**.588** *(.492)*	**.517** *(.499)*	.071
	Whether the customer owns any basic financial product at the bank	**.652** *(.736)*	**.597** *(.725)*	.055
	Whether the customer owns any investment product at the bank	**.149** *(.453)*	**.131** *(.436)*	.018
	Whether the customer owns any loan product at the bank	**.159** *(.412)*	**.136** *(.730)*	.023
	Whether the customer owns any insurance product at the bank	**.637** *(.824)*	**.552** *(.884)*	.085
Contacts between customers and the bank	Frequency of contacts regarding basic financial service	**1.579** *(1.465)*	**1.442** *(1.469)*	.137
	Frequency of contacts regarding investment	**.355** *(.856)*	**.324** *(.827)*	.031
	Frequency of contacts regarding insurance	**.276** *(.730)*	**.257** *(.711)*	.019
	Frequency of contacts regarding loans	**.529** *(.884)*	**.563** *(.884)*	-.034

Note: the numbers in bold font represent means; the numbers in brackets represent standard deviations.

Table 2: Examples of Two Sequences of Observations from Two Machines

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Machine 1	1.02	.95	1.01	1.01	.97	1.05	1.02	.95	1.01	1.02	.98	1.03	.99	.98	1.01	.99
Machine 2	1.02	.95	1.01	1.01	.97	1.05	1.02	1.05	1.01	1.02	1.02	1.03	.99	1.02	1.01	1.01

Table 3: Number of Products Owned by Customers in Each Category

	Number of products owned in the category	Percentage
Basic products	0	51.47%
	1	33.55%
	2	14.98%
Investment	0	89.4%
	1	7.51%
	2	3.09%
Loan	0	86.3%
	1	12.25%
	2	1.45%
Insurance	0	58.88%
	1	26.05%
	2	15.07%

Table 4: Profit under Different Loss Functions

(a) Scenario 1

	The proposed solution	The hidden Markov model
Optimal threshold	1.57	.67
False positive rate	.36	.35
True positive rate	.80	.67
Maximum profit ($)	41,872	23,659

(b) Scenario 2

	The proposed solution	The hidden Markov model
Optimal threshold	1.15	.5
False positive rate	.57	.77
True positive rate	.9	.9
Maximum profit ($)	32,132	20,941