The Value of First Impressions:
Leveraging Acquisition Data for Customer Management

Nicolas Padilla
Eva Ascarza†

September 2017

†Nicolas Padilla is a doctoral candidate in Marketing, Columbia Business School (email: npadilla19@gsb.columbia.edu). Eva Ascarza is the Daniel W. Stanton Associate Professor of Business, Columbia Business School (email: ascarza@gsb.columbia.edu).

Abstract

Firms increasingly have access to richer customer data. What a decade ago was merely a transaction added to a customer database has become a collection of behaviors that a customer engages in while she is making a purchase (e.g., whether her purchase was online or offline, whether she used a tablet or computer, whether she bought a new product or an old best-seller). These data can be used to form the "first impressions" of a customer based on her initial transaction. In this paper we posit that a customer's first impressions carry valuable information for the firm by explaining a large proportion of the heterogeneity in future behavior, both in terms of what the customer is expected to do (i.e., her lifetime value) and how responsive she will be to marketing actions. The latter point is especially relevant for contexts in which firms do not observe many purchases per individual (e.g., retail) or where targeting occurs soon after customer acquisition. In these contexts, models that only rely on unobserved sources of heterogeneity are unable to help the marketer target newly acquired customers with precision.

We develop a latent attrition model of purchase incidence and amount that incorporates observed heterogeneity — capturing the effect of first impressions — both in individual propensity to buy as well as responsiveness to the firm's marketing actions. Applying the model to data from a retail context, we demonstrate that first impressions are very informative for the firm. For example, customers who purchase a newly introduced product on their first purchase with the brand not only tend to buy more frequently in the future than customers who bought existing products, but they also tend to spend more money in those (future) transactions. Interestingly, these customers are less responsive to email marketing campaigns and more responsive to direct marketing. Overall, we show that the focal firm would improve the return on their marketing actions by 148% if it targeted just-acquired customers based on their first impressions.

Keywords: Customer base analysis, latent attrition models, probability models, customer heterogeneity.

My first impressions of people are invariably right.

- Oscar Wilde

1 INTRODUCTION

Customers are different, not only in their preferences for products and services, but also in the way they respond to marketing actions. Understanding customer heterogeneity is crucial for a variety of marketing problems — from obtaining accurate estimates of the value of current and future customers, to deciding which customers should be targeted in the next marketing campaign. Over the last three decades, the marketing literature has provided researchers and analysts with methods to empirically identify unobserved differences across customers — e.g., customers with higher versus lower propensity to buy (e.g., Schmittlein et al. 1987; Fader et al. 2005, 2010), those who are less sensitive to a price increase (e.g., Rossi et al. 1996; Allenby and Rossi 1998), or those who are more receptive to marketing communications (e.g., Ansari and Mela 2003). However, despite the advances in technological capabilities that allow firms to track customers over long periods of time, many firms face the challenge of not having enough observations per customer to precisely estimate these differences across customers. This is especially the case for retail, hospitality services, and non-profit fund raising among other sectors, where a very large proportion of customers are one-time buyers (Fader and Hardie 2002).[1] The lack of repeated observations presents a structural challenge for estimating unobserved differences across customers, precluding firms from leveraging such heterogeneity to differentiate recently acquired customers based on their expected lifetime value and to target marketing actions precisely, which is crucial to secure a second or a third purchase.

Firms in these sectors have traditionally relied on demographics (e.g., age, gender) and/or recency metrics (e.g., how many weeks since your last transaction) to target marketing efforts with limited data (Shaffer and Zhang 1995). These approaches, however, face practical limitations: Recency metrics, for example, do not differentiate among recently acquired customers (as they all were acquired at the same time), and relevant personal information is generally hard to collect.

[1] A recent study among retailers found that between 50% and 80% of customers were one-time shoppers (with an average of 65%), phenomenon commonly known as the curse of the one-time shopper. This problem is aggravated in online channels, where the average for one-time shoppers is 80%(Zodiac-Metrics 2016).

However, thanks to technological advances, firms can now increasingly observe a wider range of behaviors on each customer touch. What might have once been considered simply a transaction added to a customer base is now a collection of behaviors that a customer incurs while she is making such transaction (e.g., is the purchase online or offline, did she use tablet or computer, did she buy a new product or an old best-seller, did she buy on discount or at full price). While some of these factors may be purely coincidental with the moment in which the customer made her first purchase, others may carry important information as they reflect latent customer preferences/attitudes. Thus, whereas firms in the aforementioned contexts might not observe customers in many occasions, they now have many more cues to form customers' "first impressions," which can then be used to better understand heterogeneity across customers.

The purpose of this research is to measure the value of first impressions (i.e., the information gathered when the customer made her first purchase) for customer-base analysis by examining the degree to which these information sources are indicative of customer's future behavior with the firm. Specifically, we examine behaviors that are easily tracked at the first purchase occasion and investigate whether they carry meaningful information that can be used to (1) identify high-value customers, (2) understand heterogeneity in responsiveness to marketing actions, and, ultimately, (3) decide which customers should be contacted in the next marketing campaign.

We investigate the value of first impressions in a retail context. The focal company is an international retailer that sells beauty products in multiple markets, both online and offline. During the period of study, the firm regularly engaged with customers by running email and direct marketing campaigns, which we observe at the customer level, and by introducing new products, which we observe at the market level. We analyze individual-level purchases in six different markets for a period of up to four years after customers are acquired (i.e., after making their first purchase). This rich variation in the data allows as to empirically separate behaviors that are merely coincidental (i.e., temporal shocks when the customer made her first transaction) with what we label "first impressions" (i.e., those that are predictive of future behavior). We show descriptive evidence that first impressions are indeed predictive of future customer behavior.

To quantify the value of first impressions, we develop a model of purchase incidence and amount. We build on recent latent attrition literature (e.g., Schweidel and Knox 2013; Gopalakrishnan et al. 2016; McCarthy et al. 2016) and incorporate observed heterogeneity — capturing customers' first impressions — both in the individual propensity to buy as well as in the responsiveness to the firm's marketing actions. Doing so allows us to *explain* customer heterogeneity with first impressions — data collected by the firm at the very first purchase — , assisting managers in predicting the value of just-acquired customers and targeting their marketing efforts more effectively soon after customers have been acquired.

We show that first impressions add relevant information to the model, even when unobserved heterogeneity is already accounted for. We further demonstrate that first impressions are insightful for the firm, not only in terms of how valuable customers will be in the future, but also on their responsiveness to the firm's marketing actions and on their sensitivity to seasonalities on demand. For example, customers who purchase a newly introduced product (as opposed to a product that has been part of the retailer's product line for a while) when they first purchased from the brand not only tend to buy more frequently in the future than customers who bought existing products, but they also tend to spend more money when making those (future) transactions. Interestingly, these customers are less responsive to email marketing campaigns and more responsive to direct marketing than regular customers. The number of products bought in the first purchase also helps explain differences in future customer value. Specifically, customers who buy multiple units in their first purchase tend to spend more (per transaction) than regular customers; however, their transaction rate is indistinguishable from the customers who purchase fewer units.

Consistent with common beliefs among researchers and practitioners, customers acquired during holiday periods are least valuable to the firm: They are less likely to be retained and spend less (per transaction) than regular customers. However, even though these customers are less likely to make a repeat transaction compared to customers acquired in different times of the year, they are more responsive to marketing tactics such as new product launches and email campaigns. Finally, we find that information such as whether the customer buys discounted items or the average price paid in their first purchase also shows differential effects on both future behavior and responsiveness to marketing actions.

We also analyze the value of first impressions for newly acquired customers. First, we show how the proposed approach can better identify heavy spenders (separately from those who are expected to bring less value to the firm) right after the customer made her first transaction; that is, without the need of observing the customer in multiple occasions. Second, we conduct a set of "what-if" analyses to quantify the incremental value (in terms of transactions and total revenues) that our focal firm would obtain by leveraging the insights obtained from first impressions to target just-acquired customers more effectively. Specifically, we simulate all future purchases of newly acquired customers and compare their behavior under two scenarios: (1) if the firm were to implement its current marketing policy and (2) if the firm were to select target customers based on the insights from our model. We show that the focal firm would significantly improve the return of their marketing actions if it targeted just-acquired customers by their acquisition-related variables — the increase in sales due to a marketing campaign is estimated to be 148% greater under the proposed approach than under the current policy. This increase is mainly driven by the additional value captured from customers who were acquired online and/or those who purchased on discount in their first transaction.

While some of the behavioral insights might be specific to the context we chose to investigate (i.e., retail), the main finding of showing that first impressions contain valuable information is relevant and generalizable to other contexts as well. Because first impressions are the realization of decisions made by a customer (when she made the first transaction), they should be informative as to how that customer will behave in the future. In practice, most of what firms know about a just-acquired customer is the information collected when she made her first transaction. Our research highlights that, by not leveraging this information — already available in their databases — firms are leaving money on the table.

We proceed as follows. We review the recent work on customer-base analysis, highlighting how our research fits within and contributes to that literature, and briefly discuss the literature on acquisition and its relationship with subsequent customer behavior. Next, we describe the context of our empirical setting, the data used in the analysis, and several model-free analyses highlighting the benefits of incorporating first impressions in customer-base analysis. We then present a model that incorporates the differential effect of first impressions on customer behavior, allowing the analyst to differentiate recently acquired customers by their expected lifetime value as well as by their sensitivity to marketing actions. Once we have validated and presented the results of our model, we conduct a set of scenario analyses to quantify the value of first impressions for our focal retailer. We conclude with a discussion of the implications and managerial relevance of our research.

2 RELATED LITERATURE

Our research relates to the literature on customer-base analysis that has provided managers and analysts with tools to understand, forecast, and manage the (heterogeneous) behavior of customers. Methodologically, our research is closer to the work that has incorporated the effect of marketing variables, or more generally, time-varying covariates, in latent attrition models. Building on the foundations of the Beta-Geometric/Beta-Binomial (BG/BB) model (Fader et al. 2010), Schweidel and Knox (2013) and Schweidel et al. (2014) incorporate the effect of covariates — direct marketing activity and past customer activity, respectively — on the latent attrition process and on customer's purchase propensity while alive. Similarly, Knox and van Oest (2014) and Braun et al. (2015) incorporate the effect of covariates — customer's service experience and customer complaints, respectively — on the latent attrition process of the Beta-Geometric/NBD (BG/NBD) model (Fader et al. 2005). Our model generalizes the former set of models by allowing customers to differ in their responsiveness to the firm's marketing actions. Furthermore, we *explain* these differences using customers' first impressions, enabling the model to segment or prioritize recently acquired customers (who have almost identical transaction histories) by their heterogeneous response to specific marketing interventions.

From a more substantive point of view, our work is closely related to Gopalakrishnan et al. (2016), who propose a framework for multi-cohort data that is able to predict behavior of new cohorts of customers for whom the firm has little transactional data. Similarly to their work, we aim to make predictions about customers for which the firm has little information (e.g., just one transaction). However, our work differs from theirs in several ways. Gopalakrishnan and colleagues achieve forecasting accuracy at the cohort level by allowing customers to be inherently different depending on when they were acquired (i.e., *which cohort* they belong to), while capturing the underlying dynamics across cohorts. Alternatively, we posit that such inherent heterogeneity can be explained (at least partially) by observed characteristics collected when the customer made her first purchase (i.e., first impressions). Therefore, extracting this type of observed heterogeneity will allow firms to segment and target customers on the basis of variables/behaviors easily observed at the moment of acquisition. This is especially relevant for contexts where many customers make few transactions (e.g. online retailers); contexts in which models that only include unobserved heterogeneity are underpowered to pin down the individual level parameters, and in cases for which individual-level predictions (not cohort-level predictions) are of interest. Furthermore, we also capture observed heterogeneity in the sensitivity to marketing actions. While the framework introduced by Gopalakrishnan et al. (2016) can be adapted to include the impact of marketing variables, such a framework will not be able to differentiate among just-acquired customers in their sensitivity to the firm's marketing actions. In contrast, our approach differentiates customers' responsiveness by using their first impressions, crucial for the cases in which targeting occurs soon after the customer is being acquired or where securing a second purchase is challenging.

Conceptually, our work also relates to McCarthy et al. (2016), who derive a closed-from expression for V(CLV) (i.e., the variance of customer lifetime value) to explain variation in customers' future profitability. We take a different approach and investigate the variation in future profitability that can be explained by customers' first impressions. The premise that first impressions explain heterogeneity in future behavior, originates from combining findings from the acquisition literature, which has investigated the relationship between acquisition-related information (e.g., channel of acquisition) and subsequent behavior of customers (e.g., lifetime value). For example, Verhoef and Donkers (2005) find that direct-mailing acquired customers present low retention and cross-selling rates, whereas online acquired customers have higher retention rates, Villanueva et al. (2008) find that customers acquired through word-of-mouth remain with the firm longer than regular customers and therefore are more valuable, Chan et al. (2011) find that customers acquired through the Google search advertising have higher purchase incidence, and therefore higher value to the firm, than those acquired by other channels, and Steffes et al. (2011) find that customers acquired through internet and direct mail channels are more profitable than those acquired by telemarketing and direct selling. Another line of research has focused on the impact of various acquisition-related promotions (e.g., free trials, referral programs, discounts) on the value of acquired customers. For example, Schmitt et al. (2011) and Uncles et al. (2013) find that customers acquired through referral have superior retention than regular customers, Lewis (2006) shows that customers acquired through price discounts have lower lifetime value, a finding consistent with that of Datta et al. (2015), who have shown that customers acquired through a free trial have lower lifetime value.

Our work similarly investigates relationships between acquisition-related variables and subsequent customer behavior. However, we differentiate from the aforementioned work in two main ways. First and foremost, our research aims to assist decisions related to the management of already acquired customers (e.g., whom should I target in the next campaign?), rather than designing optimal strategies for customer acquisition (e.g., should I offer free trials to increase customer acquisition?). Therefore, our goal is to extract as much observed heterogeneity as possible from first impressions while controlling for the firms' acquisition activities, rather than estimating the casual impact of these acquisition variables. Second, this literature suggests that customers are inherently different depending on *how* they have been acquired. We broaden the range of acquisition-related behaviors by looking at not only how a customer was acquired (e.g., online vs. offline, trial vs. regular), but also what she did when she was acquired (e.g., what did she buy on her first purchase), hence extracting more information from the initial transaction. The latter aspect is especially relevant for managers/analysts of large retailers and hospitality businesses, among others, as such information not only is easily observed but typically already exists in the firm's database.

3 EMPIRICAL CONTEXT

We turn to a retail context to study the value of the first impressions. Retail is a good context to examine this phenomenon for several reasons. First, the large majority of retailers face the challenge on having many one-time buyers in their customer base (Zodiac-Metrics 2016). Second, firms in this sector increasingly collect transactional data and rely on analytics to better manage their customers (Forbes 2015). Finally, retail represents a large proportion of the total economy, with revenues accounting for 31% for the global GDP (Research and Markets 2016).

We analyze data from an international retailer that sells its own brand of beauty and cosmetic products (e.g., skincare, fragrance, haircare).[2] Customers can only purchase the company's products via owned stores, either offline (the company owns "brick and mortar" stores across many countries) or online (with one online store per country). While the company is present in many countries, marketing and operations are very consistent across markets.

[2]The authors thank the Wharton Customer Analytics Initiative (WCAI) for providing this data set.

3.1 Data

We obtain individual-level transactions for registered customers in the six major markets — USA, UK, Germany, France, Italy, and Spain. We observe customers from the moment they made their first purchase (starting on November of 2010). At the point of purchase, customers are asked to provide their name, email, and address so that they can receive promotions and other marketing communications from the firm.[3] We track their behavior up to 4 years after that date (ending on November of 2014). We have 13,473 customers, with a minimum of 3 and a maximum of 51 periods of individual observations, resulting in 287,584 observations.[4] During this time, we observe a total of 15,985 repeated transactions in which customers spent, on average, 51.77€ per transaction. This quantity is heterogeneous across customers, varying from 1.5€ to 803.79€.

Common to most businesses in transactional (noncontractual) settings, a large proportion of customers do not make a second transaction with the firm. Figure 1a shows the histogram of repeated transactions: While some customers made more than 10 transactions, 54.16% of customers never bought again after they had been acquired. Furthermore, we look at the average (empirical) probability of making a (repeat) transaction overtime (Figure 1b). We observe that the probability of transacting dramatically decreases as the number of periods since acquisition increases. This pattern is a clear sign of latent attrition among the customers in this setting.

— Insert Figure 1 here —

In addition to the behavior of the 13,473 registered customers, we collect data on all purchases made by "anonymous" customers in all six markets — i.e., those who never shared their identity with the firm. While their behavior is not included in our main analysis (the firm can neither track their future behavior nor communicate with them via email or mail), we use these transactional data to control for shocks in distribution channels that affected the timing of the introduction of new products in specific markets.

Customers' first impressions

[3]A customer can choose not to provide any personal details, in which case the company does not add her to the customer list, nor can we observe her future behavior.

[4]A period corresponds to exactly 28 days. We do not use calendar month as our unit of analysis because we want to have the same number of days in all periods.

We compute customers' first impressions using the data from each customer's first transaction — data that had been already collected but not leveraged by the focal firm. In particular, we use six variables to represent a multi-dimensional vector of first impressions. `Holiday` is a dummy variable that equals 1 if customer made her first transaction during the winter holiday period and 0 otherwise,[5] `Online` is a dummy variable that equals 1 if the first transaction was made online, and 0 otherwise; `NewProduct` is a dummy variable that equals 1 if the customer bought a product that had been introduced in the 30 days prior to the purchase, and 0 otherwise; `Quantity` is the total number of units bought at the first transaction; `Avg.Price` is the total amount in euros of the ticket divided by the number of units bought at the first transaction; and `Discount` equals the total amount of discount the customer received in euros for the first transaction. In principle, any variable related to the moment of the first purchase could be added to this multi-dimensional vector (e.g., did the customer ask for help, was the customer buying with others or alone). Our choice of variables was driven by three factors: data availability (i.e., information easily extracted from the already existing transactional data), findings from previous literature (e.g., customers acquired via discount are less likely to buy in the future (Lewis 2006)), and industry-specific purchase patterns (e.g., retail experts believe that customers acquired over the holidays are less profitable in the long run). The model we propose in Section 4 is applicable to other settings as well, allowing the research to adapt the vector of first impressions easily. [6]

— Insert Table 1 and here —

The variation in the data is very rich (Table 1). For example, 22% of the sample was acquired over the holiday period, and 17% was acquired online. The standard deviations of price, number of items purchased, and discount are large, reflecting the heterogeneous behavior of customers across the six markets. We explore the correlations among first impressions (Table 2). It is to be expected that some of these variables will be correlated, as they capture different behaviors incurred by the same customer (e.g., a customer who buys many items might look for cheaper products).

[5]We compute such a variable for each market separately because the exact calendar time for the holiday period varies across countries. For example, in the USA the holiday "shopping" period covers Thanksgiving week until the last week of December (i.e., the end of Christmas), whereas in Spain the only holiday season corresponds to Christmas, which starts at the end of December and ends after the first week of January.

[6]In turn, while not the main purpose of our study, one could also use our approach to identify which behaviors, among all observed, should be considered first impressions in different contexts.

Some of these correlations might also arise from the market conditions at the moment in which a customer was acquired (e.g., if the company introduces all of its new products during the holiday, customers with `holiday`$=1$ will also have `NewProduct`$=1$ and vice versa). The latter case could be potentially problematic because if we were not able to separate the effect of being a "holiday customer" from being a "new product customer," or if the company were to change its policy in the future (e.g., introducing new products in June), our model inferences about just-acquired customers could be biased. We corroborate that this is not the case in our context. Because we obtained first impressions data during 4 years, from 6 different markets, we have substantial longitudinal variation, ruling out any firm-related systematic relationship among first impressions.

— Insert Table 2 and here —

We indeed find that some first impressions are correlated with each other — e.g., customers who purchased many items paid less per item (correlation$= -0.321$), and those who bought on discount also paid slightly lower than those who paid full price when they were first acquired (correlation$= -0.110$). However, these correlations are not very high, suggesting that while these variables are related, each of them seems to capture something different about a customer's first impressions, enabling us to separately identify the effect of each variable.

Marketing actions

The firm regularly sends emails and direct marketing to registered customers. The content of these promotional activities is set globally (i.e., the same promotional materials are used across countries, translated to the local language), though their intensity is set by market (e.g., the US tend to send more emails than France). In addition to promotional activity, the company uses product innovation as a marketing tool. Like other major brands in this category, the focal retailer regularly adds extensions and/or replacements to their product lines. The sense among the company managers is that such activity not only helps acquiring new customers but also keeps current customers more engaged. When the company introduces a new product, it does so in all markets simultaneously. There is, however, some variation across markets regarding when new products were introduced. Conversations with the company confirmed that such variation is due to differences (and random shocks) in the local distribution channels.

While direct and email marketing are observed at the individual level (we denote them by `DM` and `Email`, respectively), the availability of new products is not observed at a granular level. We create a new product introduction variable (`Introd`) by combining point-of-sale data (at the SKU level) with a firm-provided SKU list of new products. Specifically, we obtain the list of all new products introduced during the period of our study. We identify the SKUs for all products in that list and infer inventory in each market from *all* purchases observed in that particular market (including all 304,497 transactions from "anonymous" customers). We assume that a new product was introduced in a market at the time the first unit of that SKU was sold.[7] We then create a period/market-level variable representing the number of new products that were introduced in each market in each time period. Table 3 shows the summary statistics for the marketing actions summarized across observations and across individuals. For the latter, we summarize individual averages, individual standard deviations, and the individual coefficient of variation. The variation in these data is also very rich, both across customers (capturing mostly differences across markets) and within customers (capturing differences over time).

— Insert Table 3 here —

3.2 Model-free analyses

We first explore the data to see if there is initial evidence that first impressions explain differences in subsequent behavior (and value) across customers. To do so, we select the customers for whom we observe at least 15 periods and compute the total revenue they generate, excluding the first transaction, during those first 15 periods. We further split this metric by computing the number of repeat transactions made by each customer and the average amount spent per transaction (Table 4).

— Insert Table 4 here —

Consistent with common belief in the industry (e.g., Artun 2014; RJMetrics 2016), customers that were acquired during the holiday season are less valuable to the firm. On the other hand, customers that bought a new product on their first transaction generate more revenue during

[7]For this step we include all purchases observed by the firm during our observation window, including those from customers not in our sample (e.g., customers that had been acquired before our observation window or customers who never registered and remained anonymous to the firm).

the first 15 periods than customers who did not. A similar pattern exists for customers who used the online channel and those who bought using discounts on their first transaction. Not surprisingly, customers that bought more units also bring more value to the firm during their first 15 periods. Interestingly, when we decompose the revenue in purchase incident and amount, we learn that customers that bought more products are more valuable not because they purchase more frequently — in fact, the number of future transactions is virtually the same, no matter how much customers bought in their first transaction — but because they spend more per transaction. Moreover, customers who purchased new products in their first purchase will likely generate more value to the firm as they have a higher chance to make both more transactions and spend more per transaction.

While these results provide preliminary evidence that first impressions are informative of future behavior, this simple analysis is not without caveats. First, this approach does not separate the predicting power of one first impression from that of another. As discussed earlier, we want to estimate the predictive power of each first impression separately, as doing so will enable us to predict behavior of newly acquired customers, even as the correlations among first impressions (perhaps due to firms actions or to changes in behavior) change over time. Second, the analysis thus far does not shed any light about customers' response to marketing actions. These results indicate that "holiday" customers are less likely to transact again. However, are they more/less sensitive to the firm's communication? Should the firm send them an email or a DM? These insights are crucial for the manager interested in elevating the value of the customers already acquired. Finally, this analysis corresponds to a sub-sample of customers — those for whom we observe enough periods — in order to have a fair comparison across customers over the same number of periods. Ideally, we would like to include all customers, even those who joined later and therefore were not observed for several years, to better capture all the heterogeneity in our data. In the next section we build a model that addresses these caveats. Such a model will be able to prioritize just-acquired customers by their expected value as well as by their sensitivity to the firm's marketing actions.

4 THE MODEL

We build a latent attrition model for repeat purchase incidence and amount spent that allows customers to differ in their responsiveness to marketing actions.[8] Specifically, we assume that customers become inactive (latently) at some unobserved rate. While alive, the propensity to transact with the company as well as the monetary amount of that purchase vary across customers. Furthermore, purchase incidence and amount (given purchase) are influenced by the firm's marketing efforts (e.g., direct mail, email marketing, the introduction of new products) as well as by seasonal factors (e.g., holiday periods). Most importantly, customers are heterogeneous in the extent to which these covariates affect their individual behavior.

We separate customer heterogeneity (both in the attrition, purchase, and spend propensities as well as in the impact of the time-varying covariates) into two components: (1) heterogeneity driven by idiosyncratic customer characteristics, *unobserved* to the researcher, and (2) heterogeneity explained by *observed* characteristics. We include all the first impressions in the set of observed characteristics, as doing so will allow us to *explain*, at least partially, heterogeneity in customer behavior in terms of purchases and responsiveness to marketing actions.[9] For example, if customers acquired during the holiday period have a systematically lower propensity to purchase than customers acquired in other times during the year, the model would capture (and quantify) such a difference. Moreover, the model also captures whether customers who bought new products in their first purchase have a systematically higher sensitivity to promotional activity than customers who bought already existing products. We acknowledge that a model that only includes unobserved heterogeneity could, in principle, capture some of those differences, provided a flexible enough distribution for modeling unobserved heterogeneity. However, such a model cannot precisely estimate the individual parameters of customers with few observations, preventing the marketer to manage just-acquired customers (by definition, with no purchase history) in the most effective way.

[8]While we could build a simpler model without incorporating the latent attrition component, we chose to do so in order to control for the commonly observed pattern (in retail and similar contexts) of customers ceasing to transact with the company for reasons other than the product offering or the firm's activity. If latent attrition is present, and the model did not account for that, we would likely bias the parameters that capture the effect of the marketing activity.

[9]Note that similar to existing models for customer base analysis (e.g., Fader et al. 2005, 2010; Schweidel and Knox 2013; Braun et al. 2015) we build a model for repeat purchases. Therefore, the information from the first transaction is only used to explain heterogeneity and never as a dependent variable.

4.1 Model components

4.1.1 (Unobserved) attrition

We assume the existence of a latent state that reflects the status of the relationship between the customer and the firm. We denote z_{it} as a latent indicator of customer attrition, with $z_{it} = 0$ if the customer is alive, $z_{it} = 1$ otherwise. (By construction, $z_{i0} = 0$ as a customer is always alive when she is acquired.) In every period, each customer might "die" with a probability θ_i. Attrition is an absorbing process, implying that once a customer "dies," she no longer comes back to life. More formally,

$$P(z_{it} = 1 | z_{it-1}) = \begin{cases} \theta_i & \text{if } z_{it-1} = 0 \\ 1 & \text{if } z_{it-1} = 1 \end{cases} \qquad i \in \{1, \ldots, I\},\ t \in \{1, \ldots, T_i\}, \tag{1}$$

where I represents the total number of customers, T_i denotes the number of periods since the customer was acquired, and θ_i is defined as

$$\theta_i = \frac{e^{\alpha_i^\theta}}{1 + e^{\alpha_i^\theta}}. \tag{2}$$

4.1.2 Purchase incidence

While alive, a customer makes a transaction in period t (i.e., $y_{it} = 1$) with probability p_{it}. Such a probability is affected by seasonalities in demand (e.g., holidays) as well as the firm's marketing actions, both of which change over time. More formally, we assume that

$$P(y_{it} = 1 | z_{it}) = \begin{cases} p_{it} & \text{if } z_{it} = 0 \\ 0 & \text{if } z_{it} = 1 \end{cases}, \qquad i \in \{1, \ldots, I\},\ t \in \{1, \ldots, T_i\}, \tag{3}$$

where p_{it} is defined as

$$p_{it} = \frac{e^{\alpha_i^p + \boldsymbol{\beta}_i^p \mathbf{x}_{it}}}{1 + e^{\alpha_i^p + \boldsymbol{\beta}_i^p \mathbf{x}_{it}}}. \tag{4}$$

The term α_i^p captures the individual-specific propensity to transact[10] and $\boldsymbol{\beta}_i^p$ is a vector containing customer i's responsiveness to the time-varying covariates, $\mathbf{x}_{it}$. With reference to the notation introduced in Section 3, we have:

$$\mathbf{x}_{it} = [\text{DM}_{it}, \text{Email}_{it}, \text{Introd}_{it}, \text{Season}_{it}].$$

4.1.3 Amount spent

In addition to purchase incidence, our model also incorporates purchase amount, a_{it}, conditional on the customer having made a transaction. We measure amount in monetary units (e.g., dollars, euros) as this metric is always available in the firm's database and relates to customer profitability in a more straightforward way. Because amount spent is generally very skewed, we model a_{it} in the log space (Chan et al. 2011; Schweidel and Knox 2013).Therefore, conditional on the customer making a purchase (i.e., $y_{it} = 1$), we assume that

$$\log(a_{it}) = \alpha_i^a + \boldsymbol{\beta}_i^a \mathbf{x}_{it} + \varepsilon_{it}, \tag{5}$$

where ε_{it} is normally distributed with mean 0 and variance σ_a^2.

4.2 Customer heterogeneity (capturing first impressions)

The model incorporates customer heterogeneity in the attrition, purchase, and spend propensities (α_i^θ, α_i^p, and α_i^a) as well as in the responsiveness to marketing actions ($\boldsymbol{\beta}_i^p$, and $\boldsymbol{\beta}_i^a$). We separate this heterogeneity into observed factors (aimed at *explaining* customer differences that will be leveraged in future targeting decisions) and other unobserved factors (aimed at controlling for idiosyncratic differences across customers). Following the notation introduced earlier, we have

$$\alpha_i^b = \tilde{\alpha}_i^b + \alpha_0^b \cdot W_i, \qquad \text{with } b \in \{\theta,\ p,\ a\} \tag{6}$$

$$\boldsymbol{\beta}_i^b = \tilde{\boldsymbol{\beta}}_i^b + \boldsymbol{\beta}_0^b \cdot W_i, \qquad \text{with } b \in \{p,\ a\} \tag{7}$$

where W_i denotes customer i's first impressions, that is,

$$W_i = [\text{Holiday}_i, \text{Online}_i, \text{NewProduct}_i, \text{Quantity}_i, \text{Avg.Price}_i, \text{Discount}_i].$$

[10] In our data there are merely 1,139 occasions in which a customer makes two or more transactions in a given period. Because it only accounts for 5.55% of the observations, we treat those observations as if the customer made one transaction. If the data were such that multiple transactions happen in the same time period, one would easily change the binomial specification for purchase incidence to a Poisson specification for the number of transactions in a given period.

We further assume that $\widetilde{\alpha}_i^b \sim N(\mu_{\alpha^b}, \sigma^2_{\alpha^b})$ and $\widetilde{\boldsymbol{\beta}}_i^b \sim N(\mu_{\beta^b}, \sigma^2_{\beta^b})$. This heterogeneity specification is consistent with that of previous literature (Rossi et al. 1996), with the difference that the vector of observed variables (W_i) includes information easily collected by the firm at the moment of first purchase. As a result, the model would allow to quantify the extent to which first impressions *explain* heterogeneity in expected value as well as in responsiveness to marketing variables. Moreover, such a model would help the marketer target just-acquired customers on the basis of information easily obtained when she made her first purchase (W_i).

Finally, to capture the possibility that customers in different markets could be intrinsically different in their propensities to stay alive, transact, and spend, we also include a country-specific parameter (fixed effects) in all model specifications.

4.3 Possible endogeneity of the model components

We now discuss the extent to which potential omitted variables that affect simultaneously demand and other components of our model could occur in our setting.

Marketing actions

According to the managers of the focal firm, and consistent with what we observe in the data, marketing actions are decided in two steps. First, the firm chooses periods in which it will engage in promotional activity (i.e., run a marketing campaign). This decision is made from the headquarters and affects all markets simultaneously. Second, managers in each focal market choose the set of customers who will receive each campaign, with the proportion of customers not being determined consistently. The introduction of new products follows a similar process — i.e., the decision being made globally, the implementation affected also by local factors — with the main difference being that the second step does not vary across customers of the same market.

Reflecting on our model specification, identification of the sensitivities to marketing action comes mainly from two sources. First, our model accounts for unobserved heterogeneity in sensitivities (β_i^p and β_i^a); thus, identification of those parameters arises from customer-level differences in purchase rates between periods with and without marketing actions, accounting for the probabilistic "death" of a customer due to latent attrition. Second, the model also pulls information across customers with similar first impressions, across periods, and across markets, capturing systematic differences in their responsiveness to marketing actions.

Therefore, given the business nature of our application and our model specification, we argue that the potential endogenous nature of the marketing actions is not affecting our results. The main reasons are twofold. First, we observe a rich variation on marketing actions across periods as well as across markets (Section 3.1). In order to bias our estimates, omitted variables would need to affect both demand and firm's actions simultaneously *in all markets*. As we noted above, the schedule of campaigns/product introductions are decided at the firm level, but idiosyncratic demand shocks are unlikely to affect customers simultaneously in different markets, other than the winter holiday, which our model already controls for. Second, our model accounts for unobserved heterogeneity on purchase frequency and amount (α_i^p and α_i^a), which alleviates endogeneity concerns arising from potential correlation between the firm's targeting policies and customers' level of activity.

To summarize, potential omitted variables that affect demand and marketing actions simultaneously would need to be both period- and customer-specific determined in order to affect our results. That being said, in other applications where these conditions do not hold (due to strategic behavior by the firm or because not multiple markets are available), the model could be extended to account for the firm's targeting decisions using individual level responsiveness (Manchanda et al. 2004) or adding correlations between firm decisions and unobserved demand shocks through copulas (Park and Gupta 2012), depending on how these actions are determined by the firm.

First impressions

First impressions are the realization of decisions made, almost simultaneously, when a customer purchased for the first time (e.g., buying online or at the store, buying an expensive product or a cheap item). As a result, we should expect to observe some correlation (i.e., not full independence) among those behaviors. As shown in Section 3.1, there is variation in our data that will allow us to estimate the effect of each first impression separately. Furthermore, the extent to which we observe some of these first impressions would likely be influenced by the firm's actions in the period in which the customer was acquired (e.g., periods of promotional campaigns will likely attract more discount-seekers). However, note that those marketing actions, omitted in our model, are not simultaneous to the demand (our dependent variable), but rather occur several periods earlier.

Therefore, while it is very unlikely that such an omission could bias the parameters in the model, we acknowledge that such a pattern might raise the question of whether our results will generalize to customers that will be acquired in the future if, for example, the firm stopped running those promotional campaigns. We argue, however, that this is not a concern for our analysis. Indeed, we will likely observe more discount-seekers in periods of heavy discounts compared to in periods where most items are at full price. However, this does not mean that the model's inference about what discount-seekers will do in the future is erroneous, provided we have enough variation of discount-seekers in the data.

4.4 Bringing it all together

In sum, we propose a latent attrition model for discrete-time transactions in noncontractual settings. At the core of our model there is a variant of the BG/BB model where we model not only purchase incidence but also amount (given purchase). We relax some of the distributional assumptions of the BG/BB to incorporate the *heterogeneous effect* of the covariates on the propensity to purchase as well as on the propensity to spend. In modeling customer differences in the responsiveness to those actions, we allow for unobserved heterogeneity (via $\widetilde{\boldsymbol{\beta}}_i^p$) — to capture idiosyncratic differences (unobserved to the researcher) in the responsiveness to the marketing variables — as well as observed heterogeneity (via $\boldsymbol{\beta}_0^b \cdot W_i$) — which enables the firm to identify individual factors that help targeting customers with further precision.

4.5 The likelihood function

For each individual and for each time period, we observe a bivariate vector of behavior $\boldsymbol{B}_{it} = [\,Y_{it}, A_{it}\,]$ with realization $b_{it} = [\,y_{it}, a_{it}\,]$. These two behaviors are not only conditional on each other — we only observe expenditure when there is a transaction — but also conditional on the latent variable, z_{it}, that captures the nature of the relationship between the customer and the firm. While this relationship is active (i.e., $z_{it} = 0$), we observe realizations of behavior; once a customer has ended her relationship with the firm, we only observe the absence of transactions.

Furthermore, a customer being alive at period t implies the customer was also alive at all previous periods. Therefore, if we denote tx_i as the period of the last observed transaction (with $tx_i = 0$ if no transaction other than acquisition was observed), we know with certainty that the customer was alive in all periods prior to tx_i. As a result, we can separate the likelihood function in two blocks: the early periods in which the customer was certainly alive — given that she made a transaction at time tx_i — and the later periods (with no activity) in which the customer might have become inactive — given that she had not made any transaction after tx_i.

Therefore, integrating all the model components, and denoting $\boldsymbol{\eta_0} = [\alpha_0^\theta, \alpha_0^p, \alpha_0^a, \boldsymbol{\beta}_0^p, \boldsymbol{\beta}_0^a, \sigma_a]$, $\boldsymbol{\eta_i} = [\widetilde{\alpha}_i^\theta, \widetilde{\alpha}_i^p, \widetilde{\alpha}_i^a, \widetilde{\boldsymbol{\beta}}_i^p, \widetilde{\boldsymbol{\beta}}_i^a]$, and $\boldsymbol{b_i} = [b_{i1}, ..., b_{iT_i}]$, we can write the individual likelihood function as

$$\mathscr{L}_i(\boldsymbol{\eta_0}, \boldsymbol{\eta_i} \,|\, \boldsymbol{b_i}) = \left(\prod_{\tau=1}^{tx_i} (p_{i\tau}\phi_{i\tau})^{y_{i\tau}} (1 - p_{i\tau})^{1-y_{i\tau}} (1 - \theta_{i\tau}) \right) \times \sum_{t=tx_i+1}^{T_i+1} \left(\theta_{it}^{I\{t \leqslant T_i\}} \prod_{\tau=tx_i+1}^{t-1} (1 - \theta_{i\tau})(1 - p_{i\tau}) \right), \tag{8}$$

where $\phi_{it} = \phi\left(\frac{\log(a_{it}) - (\alpha_i^a + \boldsymbol{\beta}_i^a \mathbf{x_{it}})}{\sigma_a}\right)$ is the probability density function (pdf) of the amount purchased, with $\phi(x)$ denoting a standard normal pdf. We describe the derivation of the individual likelihood function in the Appendix.

4.6 Model estimation

We estimate the model using full Bayesian statistical inference with MCMC sampling. We sample the parameters from the posterior distribution which is proportional to the joint,

$$p\left(\{\boldsymbol{b_i}\}_{i=1}^I, \boldsymbol{\eta_0}, \boldsymbol{\eta_i}\right) = \left[\prod_{i=1}^I \mathscr{L}_i(\boldsymbol{\eta_0}, \boldsymbol{\eta_i} \,|\, \boldsymbol{b_i})\right] \cdot \left[\prod_{i=1}^I p(\boldsymbol{\eta_i}|\mu_\eta, \Sigma_\eta)\right] \cdot p(\boldsymbol{\eta_0}) \cdot p(\mu_\eta) \cdot p(\Sigma_\eta), \tag{9}$$

where μ_η and Σ_η are the mean and covariance matrix of the unobserved individual parameters.[11]

In particular, we use the Hamiltonian Monte Carlo (HMC) algorithm. Compared to the random-walk type of exploration from traditional Metropolis-Hastings algorithms, HMC methods allows us to efficiently explore the posterior distribution by exploiting the gradient of the log of the posterior probability. Our algorithm, described in Web Appendix A, recursively computes the log likelihood at the individual level, $\log(\mathscr{L}_i(\boldsymbol{\eta_0}, \boldsymbol{\eta_i} \,|\, \boldsymbol{b_i}))$, allowing for a flexible mixing distribution for the unobserved heterogeneous parameters, as well as in the specification of the incidence and amount distributions.[12] Moreover, the algorithm computes the likelihood function in the log space

[11] All details about the mixing distribution $p(\boldsymbol{\eta_i}|\mu_\eta, \Sigma_\eta)$, and the prior distributions $p(\boldsymbol{\eta_0})$, $p(\mu_\eta)$, and $p(\Sigma_\eta)$ are presented in the Appendix.

[12] For example, our approach does not require closed-form expressions of the estimated quantities (e.g., Braun et al. 2015), allowing for a more general set of functional forms.

offering a more robust method than models estimated based on HMM likelihood functions (e.g., Schweidel and Knox 2013).

Furthermore, given that the parameters from the different behaviors of a latent attrition model are highly correlated in the posterior distribution through the likelihood function, specifically those governing attrition and incidence probabilities, HMC is a more efficient procedure to obtain samples from the posterior distribution. Specifically, we use the No U-Turn Sampling (NUTS), implemented in the Stan probabilistic programming language (Hoffman and Gelman 2014; Carpenter et al. 2016). We choose Stan for several reasons. First, it allows to easily build models and modify their different components (e.g., presence of observed and unobserved heterogeneity on the different behaviors, prior distributions, behaviors affected by time-varying covariates). Second, we can rapidly implement the HMC algorithm to obtain a sample from the posterior distribution. Finally, Stan is freely available, which facilitates the use of this model among researchers and practitioners.[13]

Using the NUTS algorithm, we obtain 2,000 draws of the posterior distribution after a burn-in period of 2,000 iterations. We assess convergence by analyzing the $\hat{R}$ statistic (Gelman and Rubin 1992).

4.7 Estimated models

In addition to the proposed model, we estimate three nested specifications where we simplify the assumptions about heterogeneity in the model components. We do so for two main reasons. First, we want to quantify the gain in model fit when adding different sources of heterogeneity. Second, we also want to test the robustness of the insights regarding first impressions when estimating different model specifications. We estimate the following models:

- **[M1] Full heterogeneity**, as described above.
- **[M2] Unobserved heterogeneity only**, in which only unobserved heterogeneity is in the model. This corresponds to M1 without the first impressions.
- **[M3] Observed heterogeneity only**, in which only observed heterogeneity is included. That is, heterogeneity is captured by first impressions only.

[13] The code is available from the authors.

- **[M4] No heterogeneity**, in which all customers have the same propensity to transact, to churn (latently) and also have the same responsiveness to the time-varying covariates

4.8 Model comparison

We randomly split the 13,473 customers into a calibration sample (80% of customers, used to estimate each model) and validation sample (the remaining 20% of customers, left aside to evaluate the out-of-sample predictions as well as to quantify the value of targeting customers based on first impressions). When comparing individual-level models, it is common practice to split the data longitudinally using the first observations per customer to estimate the model and then predict behavior for subsequent observations for the same set of customers. We split the data across customers instead because the main purpose of our model is to uncover observed sources of heterogeneity that can be used to target just-acquired customers with more precision. Therefore, a measure of how the model performs on a new cohort of customers compared to existing models is highly in line with its potential use for managerial purposes (we present these analyses in Section 5.4).

As measures of model fit, we compute the posterior mean of log likelihood, the log marginal density (LMD) and the Watanabe-Akaike Information Criterion (WAIC) in the calibration sample. It is worth noting that, unlike the posterior log likelihood and the LMD, the WAIC (Watanabe 2010) is a fully Bayesian information criterion that takes into account the posterior draws as well as any prior information. It computes model fit based on the log pointwise predictive density adjusted by a penalty for the number of parameters, and is asymptotically similar to Bayesian leave-one-out (LOO) cross validation criterion (Gelman et al. 2014). In other words, the WAIC not only estimates cross-validation expectations, but also adds a correction for effective number of parameters to adjust for overfitting.

With reference to Table 5, all three measures of fit favor the model with full heterogeneity (M1). (For a reference, we also include the fit of a BG/BB model (Fader et al. 2010), combined with a log-normal distribution to capture purchase amount. The main difference between such a model and M2 is that the baseline model does not include the impact of the time-varying covariates). The model fit clearly improves when adding unobserved heterogeneity to the model components (i.e., improvement from M4 to M2); this is expected, as M2 allows for a flexible heterogeneity structure that captures observed differences in behavior across customers. Similarly, the model improves its fit when adding observed heterogeneity (i.e., model improvement from M4 to M3). Interestingly, there is a value to adding first impressions into the model, even when unobserved sources of heterogeneity are already incorporated (i.e., model improvement from M2 to M1). This latter point is noteworthy, as one could expect that, provided a flexible heterogeneity structure is embedded in the model, it might not be necessary to add observed heterogeneity. Our analysis shows that even such a case, first impressions confer additional value. We also show that, not surprisingly, customers with fewer purchases are those who benefit the most from including first impressions in the model (See Web Appendix C for the analysis).

— Insert Table 5 here —

To summarize, the full heterogeneity model (M1) provides the best fit of the data, even when using measures that penalize model complexity. In what follows, we present the results for the model with full heterogeneity (M1). Nevertheless, the insights regarding observed heterogeneity in customer expected value and responsiveness to marketing actions are equivalent for models M1 and M3. (See Web Appendix B for the results of both specifications and a discussion of the differences and similarities.)

5 RESULTS

We now present the parameter estimates for all model components. We first discuss the findings regarding customers' propensities to incur each of the three behaviors. Then we turn to analyze the effect of the time-varying covariates on customer behavior. We start by describing the overall effects across the population (i.e., the impact of marketing actions for the "average" customer) and then proceed to analyze the observed heterogeneity in those effects. Finally, we demonstrate the external validity of the model showing the value of first impressions in newly acquired customers.

5.1 Propensity to transact, amount spend, and latent attrition

Table 6 shows the parameters governing the base propensities (transact, spend, and churn) as well as the effect of the acquisition-related variables on those propensities. Several findings are noteworthy. Customers who were acquired during the holiday periods have lower transaction rates and tend to spend less than customers acquired during the rest of the year. Consistent with the model-free evidence presented in Section 3.1, these customers tend to be less profitable to the company in the long run. In contrast, heavy buyers (i.e., those who purchase more products when acquired) will likely be more profitable in the long run, as they exhibit higher transaction propensities and tend to spend more than regular customers. On the other hand, these customers have higher propensity to leave the firm. In Section 5.3 we will compute the expected lifetime value per customer and compare these metrics among types of customers.

Interestingly, customers who purchase a new product when they were acquired tend to purchase more often and spend more money (given purchase), while their propensity to leave the brand is no different than that of other, "regular," customers. Customers who made their first purchase online tend to spend more money, but they do not differ from offline customers in their propensity to transact and to become permanently inactive. Regarding the price attributes of the first purchase ever made, the results are mixed. Customers who paid higher prices (on average) and those who bought items that were more heavily discounted tend to spend more in the future (if they purchase); however, their propensity to churn (latently) is relatively higher.

— Insert Table 6 here —

While this analysis is useful to identify which first impressions differentiate customers in their future behavior, the magnitude of these differences are all relative to the base probability to incur in each of the behaviors. To provide a clearer interpretation of the results of the model, we compute the base rates (purchase probability, "death" probability, and expected amount given transaction) in a non-holiday period without marketing actions, for different types of customers. Specifically, we simulate the three rates for seven prototypical customers ("personas"), each of them representing a customer with different first impressions.

We define the first "persona" as *Default*, and the other types are defined as deviations in first impressions. We assume the *Default* customer was acquired in a non-holiday period, her first transaction was made offline, did not buy a new product, purchased the average number of units, at an average price per unit (i.e., equal to the mean across first purchasers), and received an average discount. In other words, the *Default* customer is a customer with $W_i = 0$ for all dummy acquisition variables (Holiday, Online, and New Product) and the average $W_i = \overline{W}$ for all continuous acquisition variables (Amount, Avg. Price, and Discount). In addition, we simulate the remaining customer types by changing one first impression at a time and keeping the remaining first impressions as the *Default*. *Holiday* represents a customer acquired in Holiday ($\text{Holiday}_i = 1$);

Online represents a customer which her first purchase was online ($\text{Online}_i = 1$); *New Product*, a customer that bought a new product on her first purchase ($\text{NewProduct}_i = 1$). The remaining types *Quantity*, *Avg. Price*, and *Discount* represent customers with values equal to one standard deviation greater than the average for their first impression variables Quantity_i, AvgPrice_i and Discount_i.

Table 7 shows the simulated posterior mean, and the 95% posterior interval of purchase probability, "death" probability, and amount given purchase for all seven types of customers, for a non-holiday period with no marketing actions. From these results we conclude that *New Product* customers seems to be the most valuable type of customers for the firm. These customers purchase more frequently (3.90% purchase incidence and 1.99% death rate) and with higher amounts (36.47€) than *Default* customers (with 3.09%, 2.23% and 31.28€, respectively). On the other hand, and not surprisingly, *Holiday* type customers purchase less frequently (lowest purchase rate, 2.15%), spend lower amounts (lowest amount, 28.67€), and are also more likely to leave and not return (highest death rate, 2.63%). While the frequency of purchases is almost equivalent for *Online* and *Quantity* customers — their "death" propensity is almost identical and their transaction rates very similar —, *Online* customer tend to spend more than *Default* customers (42.71€ versus 31.28€).[14]

— Insert Table 7 here —

5.2 The effect of time-varying covariates

Table 8 shows the average effects of the time-varying covariates — that is, the extent to which marketing variables and holiday seasonalities impact sales (both in purchase incidence and amount) for an "average" customer. Not surprisingly, there is a positive and significant effect of Holiday on purchase frequency and amount. Not only are there more purchases during the holiday season, but the average amount of each transaction is higher than during the rest of the year. Regarding the three marketing activities, there are different effects of marketing actions in customer behavior. The effect of email is positive and significant on purchase incidence and amount. Surprisingly, direct marketing efforts have a negative effect on purchase incidence but a positive effect on purchase amount. Our interpretation of this finding is that DM might have saturated some customers,

[14]In Section 5.3 we combine these insights with those regarding the effect of marketing actions to compare the CLV across customer types.

resulting in lower number of overall transactions; however, among those who make a transaction, DM seems to encourage bigger purchases. Finally, the introduction of new products has a positive effect on purchase incidence but no effect on purchase amount.

— Insert Table 8 here —

Heterogeneity in the effect of time-varying covariates

The inclusion of first impressions in the model not only helps explain customer heterogeneity in transaction behavior but is also a way to identify *which marketing actions* are more or less effective to *which customers*. These insights are managerially relevant because the firm can use such information to increase the effectiveness of its individually-targeted marketing efforts. It is important to highlight that even though these insights could be achieved by estimating a model with (individual-level) unobserved heterogeneity, most firms in noncontractual settings do not have many observations per customer, implying that the majority of the individual-level parameters will be estimated without precision.

For ease of interpretation, we present the effects of the time-varying covariates on purchase probability and amount for each of the seven customer types.[15] Similar to the analysis presented in Section 5.1, we use (7) to simulate the effects of the time-varying covariates for each prototypical customer. Table 9 shows the posterior means and 95% CPIs of the individual responsiveness to each of three marketing variables as well as to the holiday season period. The effect on purchase probability is expressed in units of log-odds, and the effect on amount is expressed in units of log-euros.

With reference to Table 9, the first column shows the parameter estimates for a *Default* consumer, the second column corresponds to customers acquired during the holidays, and so forth. (Note that the parameters for the effects reported in Table 8 are not equivalent to a Default consumer, given that the Default consumer has the discount acquisition variable as average, whereas the values in Table 8 correspond to a consumer with discount set to zero). Comparing the effects of each specific activity across columns (i.e., across customer types), we highlight the following findings.

[15]The full set of results (with the posterior estimates of all model parameters) are reported in Web Appendix B.

First, the effect of emails in purchase incidence is especially pronounced for customers acquired online. This result is not surprising as these customers might be more internet/online oriented and hence more attentive to email communications. That being said, emails do not necessarily affect the amount spent in each purchase, an effect that is present in other types of customers. One group of customers — those who bought new products when they were acquired — seems to be not responsive to emails at all: Not only is their sensitivity to emails lower compared to all other types but its posterior interval contains zero.

— Insert Table 9 here —

Second, while the effect of direct marketing is overall negative (and statistically different from zero) there is one group of customers for whom this effect is not relevant in magnitude. These are the *Holiday* customers, whose transaction propensity is not influenced by direct marketing activities. While most types of customers increase their expenditure (given purchase) in periods when they have been targeted, such an effect is not significant for customers acquired online and those who purchased new products in their first transaction.

Third, the introduction of new products has an overall positive effect in purchase incidence but not on the amount spent. We find that *online* customers are not affected by the introduction of new products. This suggests that the firm could do more on emphasizing the introduction of new products in their online store.[16] *Holiday* customers are also less responsive to new product introductions, whereas those who purchased higher quantities (when acquired) are the most responsive to the introduction of new products in future periods.

Finally, the impact of holiday seasonalities in purchases and amounts is heterogeneous across customers.[17] As we would expect, customers acquired in the holiday period are most sensitive to future holiday periods (compared to customer acquired in non-holiday periods) and also tend to increase their expenditure the most. While the holiday season increases transactions almost equally among all other types of customers, those who purchased new products in their first transaction tend to increase their expenditure the most when purchasing during holiday periods. On the contrary,

[16] This suggestion is only speculative as not all customers acquired via online channel keep buying online in all their future purchases.

[17] Even though the marketer cannot affect the holiday season, understanding this source of heterogeneity can be useful designing holiday campaigns.

customers who purchased large quantities in their first transaction do not seem to increase their total expenditure (given transaction) during the holiday periods.

5.3 Expected value of customers

Once we have understood the "observed" differences across customers based on their first impressions, our next step is to compute the value that the firm would expect to capture from each of these types of customers over a long time horizon. Such a value will depend on the customer's characteristics, the time-varying covariates (i.e., marketing variables and seasonalities) that she will face over her lifetime, and her sensitivities to those market conditions. Accordingly, we incorporate the parameter estimates for each of the behaviors (Table 6) and for the sensitivities to the time-varying covariates (Tables 8 and 9) to our model specification and compute, via simulations, the discounted value of a customer's transactions for a time horizon of 51 periods in which marketing activities as well as seasonalities occur at the same level and frequency as we observe in the data. Using our data generating process, we replicate this simulation for each of the seven customer types and employ a discount rate of 1% per period, which corresponds to an annual discount rate of approximately 12.7%.[18] Figure 2 shows the CLV of each type of acquired customer.[19]

— Insert Figure 2 here —

We highlight several findings. As expected (given the results in Section 5), there is a large degree of observed heterogeneity in the value of customers. For example, a customer acquired during the holidays is expected to generate 11€ less (corresponding to a 36% lower CLV) than an otherwise identical customer who was acquired during the year. Customers acquired online are about 4€ more valuable than customers who first came to the offline store. Contrasting this difference with the model results regarding transaction, expenditure, and "death" rates (Table 7), whereas *Online* customers spend 11.43€ more than *Default* customers (given transaction) because

[18]Given that new product introductions and seasonalities are market-level covariates, we selected a representative customer from the calibration sample with 51 periods of available data and used the new product introductions and seasonalities that she observed in the market. Regarding email and DM, and given that those variables are individual/time-specific, we use the following procedure: For each period t, we compute the average email and DM activity that customers received after t periods from acquisition, and replicate this policy to all seven customer types. Further details about the simulation exercise are provided in Web Appendix D.

[19]Note that this is the CLV of each customer type calculated on the moment she was acquired. That is, this value does not include the revenue generated in the first transaction.

the former type of customer has a lower purchase rate and a higher death rate, the overall value of a customer acquired online is only 4€ higher (34.8€ vs. 30.9€). Furthermore, customers who buy larger quantities or pay higher prices on their first transaction are also expected to be significantly more valuable to the firm in the long run (39.8€ and 40.4€, respectively, versus 30.9€ for the *Default* case).

Comparing across the seven customer prototypes, the *Avg.Price* customer provides the highest expected value to the firm, followed by the *Quantity* and *NewProduct* groups. A priori, this finding seems at odds with the results presented in Table 7, in which customers who bought new products upon acquisition (i.e., the *NewProduct* group) had the potential to be of greatest value to the firm — they had the highest transaction rates and lowest death rates among all groups of customers. How is it possible that a group that was found to be most preferable for the firm does not provide the highest value in the long-run? While *NewProduct* customers would be the most profitable in the absence of emails and DM campaigns, their sensitivity to marketing actions is lower compared to other types of customers. (Recall from Table 9 that, for a *NewProduct* customer, the increase in purchase rate due to an email is only 0.2 versus the 0.9 and 0.7 for a *Quantity* and *Avg.Price* customer, respectively. The same reasoning applies for increase in expenditure, with 1.2€ increase for the *NewProduct* customer, compared to 1.9€ and 2.1€ for the other two groups.) Hence, provided that the firm continues some level of marketing activity, it is expected that the other two groups (i.e., the groups that have been found to be the most "sensitive" to the marketing activities) will provide higher value to the firm.

Herein lies a word of caution for the researcher and analyst: When valuing customers, not only should one incorporate expected future revenues in the calculation, but it is also important to take into account the fact that customers respond differently to market circumstances. And to the extent that those changes can be projected, as is the case for holidays and the firm's marketing activities, failing to account for customer heterogeneity in the sensitivity to those factors might mask the real value of customers.

Putting all the pieces together, we have demonstrated that first impressions are indeed informative for the firm. Such information, in addition to capturing heterogeneity on transaction and expenditure propensities, also *explains* differences in customers' sensitivity to changes in market conditions such as marketing activities and seasonalities. In the next section we show how this information can (and should) be leveraged by marketers.

5.4 First impressions of newly acquired customers

We turn to analyze the value of first impressions for newly acquired customers. We do so by leveraging the information from the 20% of customers (N= $2,736$) who were randomly selected as a hold-out sample, and whose behavior was not included in the model estimation. First, we investigate the out-of-sample predictive accuracy of the model. In doing so, we also show how the proposed approach can identify heavy spenders right after the customer made her first transaction. Second, we run a set of "what-if" analyses illustrating how the marketer can increase revenues by targeting just-acquired customers based on their first impressions.

5.4.1 (Out-of-sample) Predictive accuracy of first impressions

Table 10 presents the predictive accuracy of the estimated models on the sample of newly acquired customers. As we did for the in-sample predictions, we also present the fit of the BG/BB model (with log-normal distribution for amount). The out-of-sample accuracy of the BG/BB model has been well-validated in a variety of applications (e.g., Fader et al. 2010; Schweidel and Knox 2013; Schwartz et al. 2014; Gopalakrishnan et al. 2016), therefore the fit of such a model can serve a good baseline.

We first evaluate the accuracy of the model at predicting behavior of the full set of new customers. We aggregate the predictions both at the period- and at the individual-level, computing the root mean squared error at the period (individual) level and then aggregating across periods(individuals). As shown in Table 10, the RSME of all model specifications are comparable to the baseline (demonstrating the models' ability to predict the level of purchase rates and the heterogeneity across the population), with a sizeable improvement in model fit due to the inclusion of time-varying covariates.[20] Furthermore, the differences in RMSE show that incorporating first impressions (models M1 and M3) makes the biggest difference on predicting out-of-sample behavior accurately. The model with full heterogeneity (M1) provides the most accurate out-of-sample

[20]Because the BG/BB model incorporates unobserved heterogeneity in all components, comparing M2 with the baseline essentially captures the improve in model predictions due to the inclusion of time-varying covariates.

predictions at the individual level while the model that incorporates observed heterogeneity only (M3) provides best fit at the period level.

— Insert Table 10 here —

Most importantly, we want to test the model's ability to identify high value customers (separately from those who are expected to bring less value to the firm), without the need of observing the customer in multiple occasions. If, as we posit, first impressions carry important information about customers' heterogeneity, models M1 and M3 (which incorporate first impressions) should be able to classify/sort customers based on their expected future behavior (e.g., discounted expected value, number of transactions) soon after they have been acquired. Note that existing models for customer base analysis, which rely on unobserved heterogeneity only, are unable to differentiate between two customers right after they have made their first purchase with the firm. In contrast, provided first impressions are valuable to the firm, our model should be able to provide those insights.

We compute, for each model specification, the proportion of high value customers correctly identified/classified if the analyst were to use customers' first impression as only the information available. More specifically, using the data from the first transaction, we compute the expected revenue generated by each customer during the next 15 periods, and flag the customers who belong to the top 10% and to the top 20% percentiles. We then contrast those predictions with the actual data based on the revenue generated by each customer over the same period of time. We compare those figures with what a random classifier would predict, shown in the last row of the Table 10. As expected, the models that do not include first impressions cannot outperform a random classifier. In contrast, M1 and M3 significantly outperform the random chance of identifying high value customers (e.g., 28.3% and 29.7% of top 20% customers correctly classified, compared with a 20% baseline). These results demonstrate that incorporating first impressions in the model can help the manager uncovering customer heterogeneity in future spent (very early on) with greater precision.

5.4.2 The value of (targeting based on) first impressions

While the previous analysis is very insightful for acquisition decisions and for customer valuation analyses, it does not shed light on to *how to manage* existing customers. The next question we ask is, how a firm could use these insights to *increase the value* of its already acquired customers? We investigate this question by conducting a set of what-if analyses on the holdout sample.[21]

Using our proposed model for purchase incidence and amount, we simulate the behavior of these customers (for up to 51 periods) under three scenarios: (1) *No marketing*, assuming the company does not send email or DM campaigns to these customers. (2) *Current*, assuming the company sends email and DM campaigns to these customers with similar frequency as we observed in the calibration data, and (3) *Proposed*, assuming the company runs the same email and DM campaigns as in the *Current* scenario — i.e., reaching the same number of customers, thus incurring the exact same costs — but prioritizing the recipients for these campaigns based on the results obtained from estimating the model in the calibration sample.[22]

Note that we evaluate the scenarios on a different set of customers than those whose behavior was analyzed to determine the targeting policy. This distinction is important for two reasons. First, this approach does not suffer from potential selection bias. By construction, if a policy is derived from insights that "empirically worked" for a sample of customers, evaluating the impact of that policy on the same set of customers would likely overestimate the effect. On the contrary, evaluating the policy on a new set of customers mitigates such a potential bias as we are calculating the "out of sample" impact of such policy. Second, it would be unrealistic to believe that the firm can "test" the new policy on the exact same customers, given that their behavior needs be observed in order to derive the insights. By evaluating the different scenarios on a cross-sectional sample of customers, we aim to simulate a realistic situation in which the firm transforms the already obtained insights into targeting rules that are applied to a new set of just-acquired customers.

We start by summarizing the changes in monthly sales under the different scenarios. For ease of comparison, we look at *differences in demand* between each of the marketing scenarios (*Current* and *Proposed*) and the *No marketing* case. Figure 3a shows increments in the number of transactions whereas Figure 3b shows increments in total sales. First, we look at the incremental effect of the *Current* scenario (represented by the dotted line) to corroborate that the results discussed

[21]Ideally, one would set a field experiment in which the firm implements a new targeting rule (based on first impressions) in a randomly selected group of customers, and compare their behavior with that of the rest of the sample. Unfortunately, we could not implement such an experiment with the focal company; thus we rely on the out-of-sample customers to quantify the impact of altering the targeting rules.

[22]The details about the simulation analysis are presented in Web Appendix E.

in Section 5 are generalizable to customers outside our estimation sample. Briefly, emails have a positive effect on demand (both in terms of transactions and expenditures) and DM decreases transactions while increases total sales. Moreover, we observe that the incremental effect of the marketing actions decline overtime. This result highlights that even when we assumed a constant effect of the time-varying covariates at the individual level, the overall effect of either type of marketing campaign diminishes over time due to the latent attrition among customers.

Second, we quantify the value of targeting based on first impressions by comparing the *Current* and *Proposed* scenarios. As Figure 3 shows, the focal firm could substantially increase the impact of its marketing actions on purchases and revenues if it targeted customers based on first impressions. Not only is the proposed policy always more beneficial than the current one, but the magnitude of the difference between the two policies is considerable. For example, consider period 1, in which we simulated the case of running one email campaign. Comparing the increase in sales under *Current* and *Proposed* scenarios we find that targeting based on first impressions would increase the impact of the campaign by 68.3% — sales would increase by 355€ under the *Proposed* scenario, compared to the 211€ under the *Current* scenario. Aggregating the sales over all the periods in which a marketing campaign was run, we find that the average increase in marketing effectiveness is 148% (which corresponds to an average increase of 69.5€ under the *Proposed* rule versus the 28.0€ increase under the *Current* rule).

— Insert Figure 3 here —

One question that naturally arises is, where the value of first impressions is coming from. In other words, *who* are the customers for whom revenues increases the most when the firm targets based on first impressions? To answer this question, we compute the CLV of each of the 2,736 newly acquired customers under each of the three scenarios. Similar to the calculations described in Section 5.3, CLV is computed right after the customer has been acquired (thus excluding the revenue from the first transaction) and assuming a monthly discount rate of 1%. We then compute, for each customer, the difference in CLV if the firm used the current marketing policy vs. if it targeted based on first impressions (i.e., $\text{CLV}_i(\textit{Proposed}) - \text{CLV}_i(\textit{Current})$). Figure 4 shows the histogram of the change in CLV across the 2,736 customers. Customers for whom CLV would increase are colored in gray whereas those for whom CLV would decrease are colored in black.

— Insert Figure 4 here —

As expected, targeting based on first impressions increases the value for the large majority of customers. In particular, CLV *increases* for 72.9% of the customers. A priori, it might seem contradictory that CLV would decrease for 27.1% of customers — recall from Figure 3b that the *Proposed* scenario increases sales with respect to the *Current* scenario in every period in which a campaign was run. However, *some* decrease of CLV (for specific customers) should be expected due to the re-allocation of marketing resources. Targeting on first impressions means that the marketing communications (either email or DM) will be directed to the customers with the highest expected increase in sales. Therefore, because we keep the costs constant across current and proposed scenarios, targeting based on first impressions will sometimes come at the cost of not sending the communication to a customer for whom the communication would have been less effective yet beneficial (these are the customers for whom the *Proposed* policy results in lower CLV than the *Current* policy).

Finally, we quantify where most of value of first impressions is coming from by running a regression of the ΔCLV_i (i.e., $\text{CLV}_i(\textit{Proposed}) - \text{CLV}_i(\textit{Current})$) on the actual values of the first impressions for each of the 2,736 newly acquired customers. Following the notation introduced in Section 4.1, we run the following linear regression

$$\Delta\text{CLV}_i = \lambda_0 + \boldsymbol{\lambda}_1 W_i + \epsilon_i, \tag{10}$$

where W_i include the fist impressions and ϵ_i is normally distributed with mean 0 and variance σ^2_{clv}. Table 11 shows the result of such analysis.

— Insert Figure 11 here —

Given that our main objective is to compare variance explained across the different variables, and because not all first impressions are on the same scale (e.g., online is binary whereas discount is a continuous variable), we base our discussion on the t-values of the regression coefficients. The type of customers who seem to be driving most of the value of targeting based on first impressions are those who were acquired online (t-value= 13.880), followed by those who bought heavily discounted items (t-value= 8.177) and those who bought in the holidays (t-value= 4.019) when they were acquired. These results are consistent with the insights obtained from the main model (Table 9) where online-type customers were found to have the highest sensitivity to email. Hence, the proposed approach prioritizes these customers in email campaigns. Furthermore, the model identified that both the discount- and holiday-type customers were those for whom DM seemed to be most beneficial because the decrease in transaction rate was lower and the increase in purchase amount was higher than all other customer types. Thus, the proposed approach prioritizes these customers when sending DM campaigns.

Finally, the t-statistic for Avg.Price is negative (t-value$= -3.001$), implying that customers who purchase higher-priced items on their first purchase are those whose CLV is expected to decrease under the *Proposed* scenario. This finding is also in line with the insights obtained from the full model. While the impact of sending email to Avg.Price customers was positive, such an increase in sales was not as large as for Online customers. This also suggests that the firm could further increase the value of its acquired customers by adjusting the number of customers targeted in each campaign. While we fixed the number of customers receiving the marketing communication constant, the same type of analysis could be done to explore alternative scenarios — for example, to calculate the optimal number of recipients of each campaign.

6 GENERAL DISCUSSION

To quantify the value of first impressions, we have developed a latent attrition model of purchase incidence and amount. The novel aspect of the model is that it incorporates customers' first impressions in their propensity to transact, spend, and (latently) churn, as well as in the responsiveness to the firm's marketing actions. Such a modeling framework allows firms to not only identify valuable customers very early on but also to target just-acquired customers with larger degree of precision. This is particularly relevant for firms in retail, hospitality services, and other sectors who face the challenge of populating their customer base with one-time buyers or customers with only a few purchases. In these settings, designing targeting rules based on first impressions (i.e., behaviors easily observed at the first purchase) is especially valuable in helping managers secure a second transaction.

We apply our model to a retail context and demonstrate that first impressions are indeed very informative for the firm. For instance, we find that customers who purchase recently introduced products when they first purchased from the brand buy more frequently and greater volumes than customers who bought existing products. We also find that these "new product" customers are not as responsive to emails as other customers, implying that the firm will not be able to elevate their value through email campaigns. On the other hand, customers whose first purchase was made online have higher sensitivity to email campaigns, making them better targets for future marketing interventions.

Using a holdout sample of customers, we demonstrate the predictive power of first impressions both in terms of customer value as well as in her responsiveness to marketing actions. In addition to demonstrate the forecasting accuracy of the model relative to existing, well-validated, models for customer base analysis, we show how a firm can identify high-value customers (separately from low-value customers) right after the customer's first purchase. Furthermore, by leveraging the insights of our model, we derive a set of targeting rules, purely based on first impressions, that the firm should use when running email and DM campaigns. We evaluate the impact of applying such targeting rules on a new set of just-acquired customers and estimate that the focal firm would improve the return of their marketing actions by 149% if it targeted just-acquired customers by their first impressions.

All in all, the above findings suggest that firms are leaving value on the table by not fully leveraging information available in their databases. Our research aims to encourage managers and analysts to explore the already collected information and identify the variables that are most valuable for managing customers more effectively. To that end, we provide a flexible model that can easily incorporate these available data to derive new insights about customer behavior as well as to provide estimates of customer's long-term value. Related to the latter point, our research also highlights that when valuating customers, it is important to account for customer heterogeneity in the sensitivity to marketing actions, as failing do so may disguise the actual value of customers.

Our research can be extended in several ways. We investigated six forms of first impressions— those available in our empirical application. Future research should consider additional behaviors and identify which first impressions are most valuable for which types of firms and in which contexts. Relatedly, future work should also explore additional variables that firms should collect. For example, there might be first impressions that firms are not currently collecting (e.g., whether the customer visited the store alone and with family) but that could be very valuable to identify which marketing actions are most likely to increase sales in the future.

Finally, while we developed our model for noncontractual settings (as these are the contexts where information about customers is most limited), our research is potentially relevant for contractual businesses as well. When customers sign up for a contract, they might also exhibit different behaviors that are likely to be predictive of different sensitivity to future marketing interventions. Our model could be adapted for those settings and should be used to identify behaviors that could inform managers how to better manage customer retention in the early stages of the relationship. We hope that future research will address these and other related issues.

References

Allenby, G. M. and Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1-2):57–78.

Ansari, A. and Mela, C. F. (2003). E-customization. *Journal of Marketing Research*, 40(2):131–145.

Artun, O. (2014). What are those new holiday customers worth? [Online; accessed 5-February-2017] https://www.internetretailer.com/2014/12/19/what-are-those-new-holiday-customers-worth.

Braun, M., Schweidel, D. A., and Stein, E. (2015). Transaction attributes and customer valuation. *Journal of Marketing Research*, 52(6):848–864.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–29.

Chan, T. Y., Wu, C., and Xie, Y. (2011). Measuring the lifetime value of customers acquired from Google Search Advertising. *Marketing Science*, 30(5):837–850.

Datta, H., Foubert, B., and Van Heerde, H. J. (2015). The challenge of retaining customers acquired with free trials. *Journal of Marketing Research*, 52(2):217–234.

Fader, P. S. and Hardie, B. G. S. (2002). A note on an integrated model of customer buying behavior. *European Journal of Operational Research*, 139(3):682–687.

Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005). Counting your customers? the easy aay: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2):275–284.

Fader, P. S., Hardie, B. G. S., and Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29(6):1086–1108.

Forbes (2015). Big data: A game changer in the retail sector. [Online; accessed 23-September-2017] https://www.forbes.com/sites/bernardmarr/2015/11/10/big-data-a-game-changer-in-the-retail-sector/.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC Boca Raton, FL, USA, third edition.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–511.

Gopalakrishnan, A., Bradlow, E. T., and Fader, P. S. (2016). A cross-cohort changepoint model for customer-base analysis. *Marketing Science*, (December).

Hoffman, M. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Knox, G. and van Oest, R. (2014). Customer complaints and recovery effectiveness: A customer base approach. *Journal of Marketing*, 78(5):42–57.

Lewis, M. (2006). Customer acquisition promotions and customer asset value. *Journal of Marketing Research*, 43(2):195–203.

Manchanda, P., Rossi, P. E., and Chintagunta, P. K. (2004). Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, 41(4):467–478.

McCarthy, D., Fader, P. S., and Hardie, B. G. (2016). V(clv): Examining variance in models of customer lifetime value.

Park, S. and Gupta, S. (2012). Handling Endogenous Regressors by Joint Estimation Using Copulas. *Marketing Science*, 31(4):567–586.

Research and Markets (2016). Overview & evolution of the global retail industry. [Online; accessed 23-September-2017] https://www.researchandmarkets.com/research/tqh2xb/.

RJMetrics (2016). The ecommerce holiday customer benchmark. [Online; accessed 5-February-2017] https://rjmetrics.com/resources/reports/the-ecommerce-holiday-customer-benchmark/.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340.

Schmitt, P., Skiera, B., and Van den Bulte, C. (2011). Referral programs and customer value. *Journal of Marketing*, 75(1):46–59.

Schmittlein, D. C., Morrison, D. G., and Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, 33(1):1–24.

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

Schweidel, D. A. and Knox, G. (2013). Incorporating direct marketing activity into latent attrition models. *Marketing Science*, 32(3):471–487.

Schweidel, D. A., Park, Y.-h., and Jamal, Z. (2014). A multiactivity latent attrition model for customer base analysis. *Marketing Science*, 33(2):273–286.

Shaffer, G. and Zhang, Z. J. (1995). Competitive coupon targeting. *Marketing Science*, 14(4):395–416.

Steffes, E. M., Murthi, B. P. S., and Rao, R. C. (2011). Why are some modes of acquisition more profitable? A study of the credit card industry. *Journal of Financial Services Marketing*, 16(2):90–100.

Uncles, M. D., East, R., and Lomax, W. (2013). Good customers: The value of customers by mode of acquisition. *Australasian Marketing Journal*, 21(2):119–125.

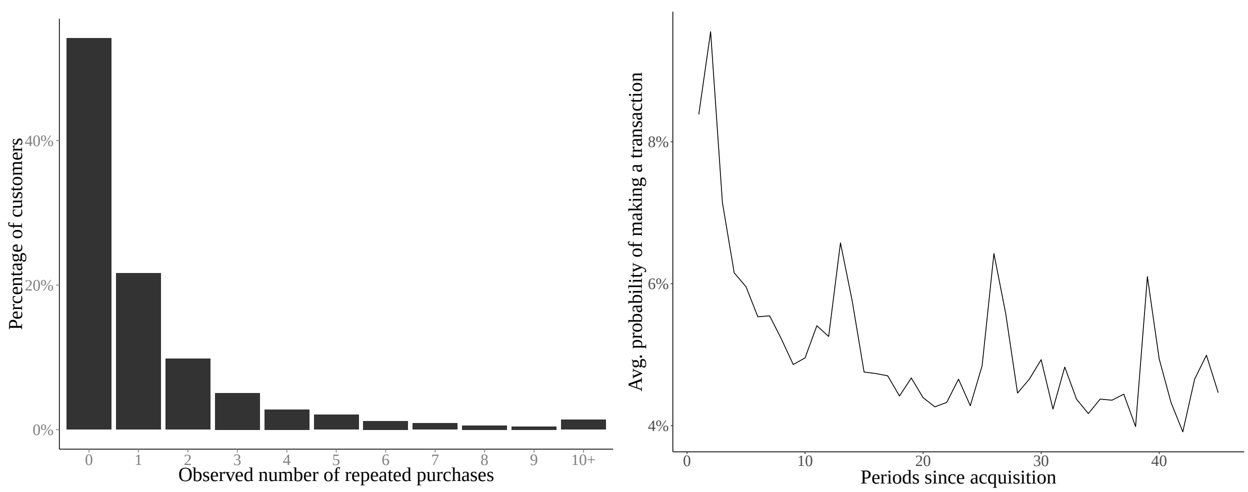
Verhoef, P. C. and Donkers, B. (2005). The effect of acquisition channels on customer loyalty and cross-buying. *Journal of Interactive Marketing*, 19(2):31–43.

Villanueva, J., Yoo, S., and Hanssens, D. M. (2008). The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *Journal of Marketing*, 45(1):48–59.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594.

Zodiac-Metrics (2016). Why you should invest in repeat customers, not one-time shoppers. [Online; accessed 5-February-2017] https://www.zodiacmetrics.com/customer-behavior/repeat-customers-vs-one-time-shoppers/.

FIGURES



(a) Histogram, by customer **(b)** Average probability, over time

Figure 1: Repeated transactions.

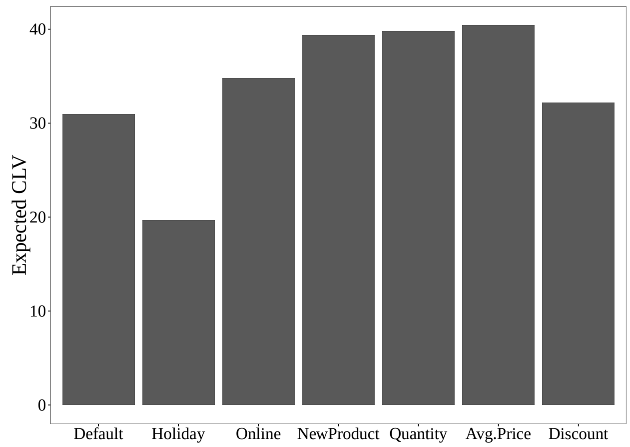
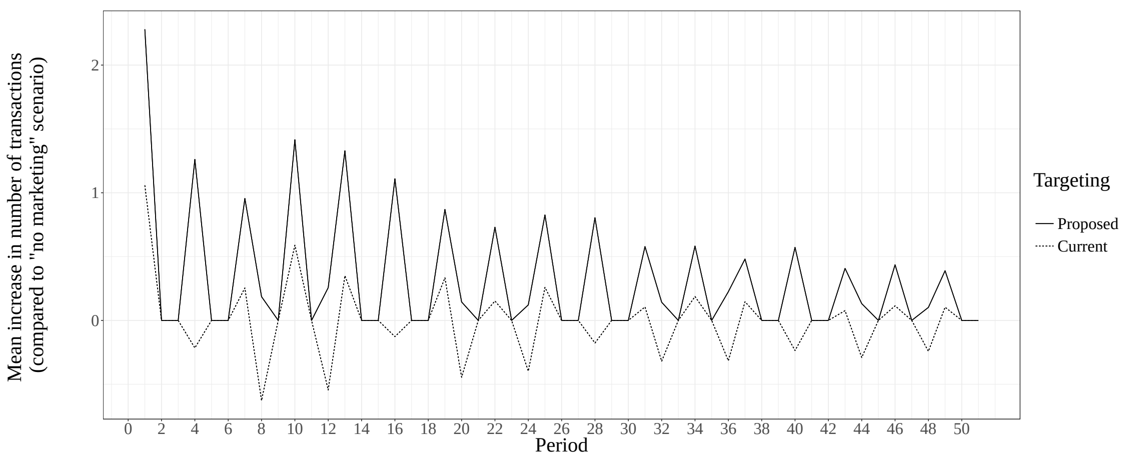
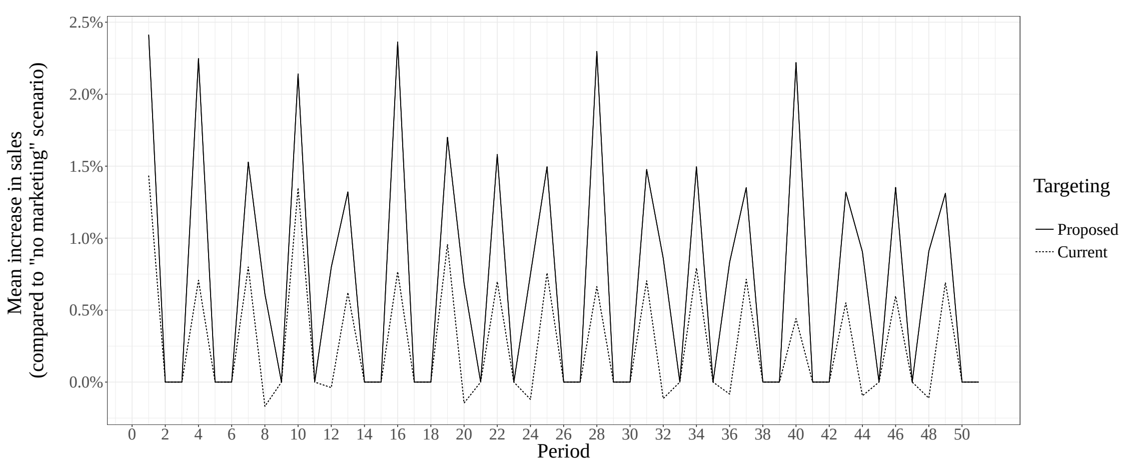


Figure 2: Expected CLV per customer type.



(a) Incremental # transactions



(b) Incremental revenues

Figure 3: Comparing each targeting policy with the scenario of no marketing actions. Email marketing campaigns are run in periods 1, 4, 7, 10, . . . 49. DM campaigns are run in periods 4, 8, 12, 16, . . . 48.

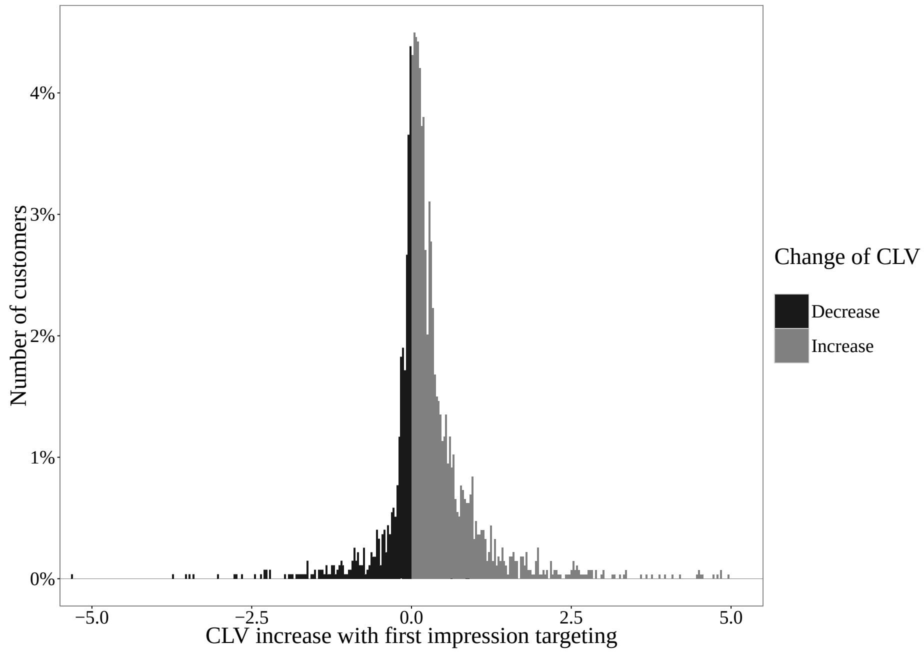


Figure 4: Distribution of difference of CLV between *Proposed* and *Current* scenarios. (CLV increase= $\text{CLV}_i(\textit{Proposed}) - \text{CLV}_i(\textit{Current})$.)

TABLES

Variable	**Description**	**Mean**	**SD**	**N**
Holiday	Whether was acquired during the Holiday	0.220	$--$	13,473
Online	Whether transaction was online	0.176	$--$	13,473
New Product	Whether a new product was bought	0.431	$--$	13,473
Quantity	Total number of units bought	4.934	5.298	13,473
Average Price (€)	Average price per unit in euros	11.642	10.237	13,473
Discount (€)	Total discount in euros	3.266	8.173	13,473

Table 1: Summary statistics of first impressions.

	Holiday	Online	New product	Avg. quantity	Avg. price
Online	0.029				
New product	0.027	0.096			
Avg. quantity	0.138	0.422	0.276		
Avg. price	−0.060	−0.227	−0.018	−0.321	
Discount	0.065	0.022	0.101	0.375	−0.110

Table 2: Correlations among first impressions.

Marketing action	**Statistic**	**Mean**	**SD**	**N**
Email	Across observations	3.267	4.686	287,584
	Indiv. average	4.272	3.612	13,473
	Indiv. st. dev.	3.404	1.790	13,473
	Indiv. coeff. of variation	1.425	1.082	13,336
Direct Marketing	Across observations	1.006	1.889	287,584
	Indiv. average	1.329	1.018	13,473
	Indiv. st. dev.	1.731	0.769	13,473
	Indiv. coeff. of variation	2.031	1.205	13,455
Products introduced	Across observations	0.923	1.264	287,584
	Indiv. average	0.657	0.532	13,473
	Indiv. st. dev.	0.755	0.534	13,473
	Indiv. coeff. of variation	1.354	0.478	11,927

Table 3: Summary of time-varying marketing actions.

First impressions		Total revenue		Transactions		Avg. amount	
		Mean	SE	Mean	SE	Mean	SE
Holiday	No	50.42	1.63	1.02	0.02	20.9	0.46
	Yes	38.10	2.51	0.82	0.04	18.51	0.90
New product	No	40.70	1.84	0.92	0.03	17.71	0.51
	Yes	55.99	2.14	1.05	0.03	23.35	0.63
Online	No	45.52	1.50	0.99	0.02	19.37	0.42
	Yes	61.67	3.91	0.94	0.05	26.01	1.22
Discount	No	46.50	1.61	0.95	0.02	19.74	0.45
	Yes	52.67	2.88	1.07	0.04	22.41	0.88
Avg. quantity	0% - 25%	36.82	2.42	0.92	0.03	16.00	0.62
	26% - 50%	45.70	2.78	1.00	0.04	19.25	0.74
	51% - 75%	49.53	2.75	0.99	0.04	20.86	0.84
	76% - 100%	62.96	3.31	1.02	0.04	26.73	1.05
Avg. price	0% - 25%	34.82	2.25	0.84	0.03	15.15	0.63
	26% - 50%	48.47	2.80	1.02	0.04	20.45	0.78
	51% - 75%	54.19	2.87	1.04	0.04	23.4	0.90
	76% - 100%	55.24	3.25	1.03	0.04	22.88	0.90
Population average		48.17	1.41	0.98	0.02	20.46	0.41

Table 4: Observed repeated transactions and revenue as a function of customers' first impressions.

	Post. mean log likelihood	LMD	WAIC
M1 - Full heterogeneity	**−52,517**	**−52,927**	**111,866**
M2 - Unobserved heterogeneity only	−52,569	−53,005	112,142
M3 - Observed heterogeneity only	−60,814	−60,836	121,796
M4 - No heterogeneity	−61,430	−61,436	122,918
Baseline (BG/BB/LogNormal)	−56,830	−57,023	115,902

Table 5: Model fit for different specifications of the heterogeneity structure. As a baseline we use a variation of the BG/BB model that incorporates a LogNormal distribution for purchase amount.

	Purchase			Amount			Latent churn		
	Post. Mean	95% CPI		Post. Mean	95% CPI		Post. Mean	95% CPI	
Parameters determining mean propensities									
μ_α	**−3.351**	−3.457	−3.241	**3.466**	3.421	3.509	**−3.851**	−4.049	−3.658
Observed heterogeneity – differences based on first impressions (α_0)									
Holiday	**−0.377**	−0.498	−0.254	**−0.069**	−0.121	−0.015	0.165	−0.053	0.379
Online	−0.088	−0.219	0.052	**0.309**	0.247	0.369	0.090	−0.156	0.314
NewProduct	**0.173**	0.082	0.267	**0.076**	0.037	0.114	−0.114	−0.291	0.071
Quantity	**0.092**	0.038	0.150	**0.127**	0.102	0.153	**0.102**	0.025	0.174
Avg.Price	0.110	0.066	0.154	**0.144**	0.123	0.163	**0.117**	0.036	0.195
Discount	0.005	−0.001	0.011	**0.004**	0.001	0.007	**0.010**	0.001	0.019
Unobserved heterogeneity									
σ_α	**1.194**	1.150	1.238	**0.414**	0.397	0.434	**0.304**	0.028	0.694

Table 6: Posterior mean and 95% central posterior interval (CPI) for parameters governing the individual propensities to transact, amount spend (given transaction) and churn (latently). Numbers in bold indicate that the 95% CPI does not contain zero.

	Type of customer, defined by first impresssions (W_i)						
	Default	*Holiday*	*Online*	*NewProduct*	*Quantity*	*Avg.Price*	*Discount*
Purchase (%)	3.09 [2.81,3.38]	2.15 [1.85,2.46]	2.82 [2.40,3.30]	3.90 [3.52,4.29]	3.23 [2.83,3.67]	3.34 [2.99,3.69]	3.21 [2.88,3.56]
Death (%)	2.23 [1.91,2.54]	2.63 [2.09,3.20]	2.45 [1.89,3.06]	1.99 [1.70,2.27]	2.46 [2.06,2.90]	2.50 [2.13,2.90]	2.42 [2.04,2.79]
Amount (€)	31.28 [29.90,32.73]	28.67 [26.69,30.82]	42.71 [39.28,46.38]	34.19 [32.71,35.69]	34.87 [32.56,37.48]	35.74 [33.93,37.70]	32.22 [30.51,33.99]

Table 7: Purchase rate, death rate and amount (given purchase), per type of customer, based on first impressions.

Effect of x_{it} *on:*	Purchase			Amount		
	Post.Mean	97.5% CPI		Post.Mean	97.5% CPI	
Email	**0.126**	0.082	0.171	**0.028**	0.002	0.052
Direct Marketing	**−0.087**	−0.129	−0.045	**0.030**	0.009	0.052
Product Introductions	**0.081**	0.040	0.121	0.001	−0.025	0.026
Seasonality	**0.598**	0.464	0.722	**0.125**	0.047	0.203

Table 8: Average effect of the time-varying covariates on purchase incidence and amount.

	Type of customer, defined by first impresssions (W_i)						
	Default	*Holiday*	*Online*	*NewProduct*	*Quantity*	*Avg.Price*	*Discount*
Effect of x_{it} on **purchase incidence**							
Email	**0.12**	**0.09**	**0.20**	0.04	**0.16**	**0.13**	**0.10**
	[0.08,0.16]	[0.02,0.16]	[0.12,0.28]	[−0.01,0.09]	[0.10,0.23]	[0.08,0.18]	[0.05,0.15]
DM	**−0.08**	−0.02	−0.06	**−0.07**	**−0.09**	**−0.06**	**−0.06**
	[−0.12,−0.04]	[−0.09,0.04]	[−0.14,0.01]	[−0.14,−0.02]	[−0.15,−0.03]	[−0.11,−0.02]	[−0.11,−0.01]
Introd	**0.08**	0.07	0.01	**0.06**	**0.11**	**0.10**	**0.09**
	[0.04,0.12]	[−0.01,0.13]	[−0.06,0.07]	[0.02,0.09]	[0.05,0.17]	[0.06,0.15]	[0.04,0.14]
Season	**0.61**	**1.08**	**0.63**	**0.60**	**0.57**	**0.56**	**0.65**
	[0.49,0.73]	[0.90,1.25]	[0.42,0.84]	[0.48,0.72]	[0.38,0.74]	[0.41,0.70]	[0.51,0.80]
Effect of x_{it} on **amount, given purchase**							
Email	**0.03**	**0.04**	0.03	0.02	**0.03**	**0.04**	**0.03**
	[0.00,0.05]	[0.00,0.08]	[−0.01,0.07]	[−0.01,0.05]	[0.00,0.07]	[0.01,0.06]	[0.00,0.06]
DM	**0.03**	**0.04**	0.01	0.01	**0.04**	**0.03**	**0.04**
	[0.01,0.05]	[0.01,0.08]	[−0.02,0.05]	[−0.02,0.03]	[0.01,0.07]	[0.01,0.06]	[0.01,0.07]
Introd	0.00	0.01	0.01	0.00	0.02	0.00	−0.01
	[−0.02,0.02]	[−0.04,0.05]	[−0.03,0.05]	[−0.02,0.02]	[−0.02,0.06]	[−0.02,0.03]	[−0.03,0.02]
Season	**0.13**	**0.21**	0.03	**0.21**	0.08	**0.09**	**0.14**
	[0.06,0.21]	[0.10,0.32]	[−0.10,0.16]	[0.14,0.28]	[−0.04,0.20]	[0.01,0.18]	[0.05,0.24]

Table 9: Effect of marketing actions and holiday season on purchase incidence and amount, per customer type.

	RMSE of actual vs. predicted revenues		Proportion of top spenders identified by the model	
	Per period	Per individual	Top 10%	Top 20%
M1 - Full heterogeneity	18.20	2.45	**0.157**	**0.283**
M2 - Unobserved heterogeneity only	18.90	2.46	0.084	0.193
M3 - Observed heterogeneity only	18.04	2.55	**0.185**	**0.297**
M4 - No heterogeneity	18.89	2.56	0.129	0.224
Baseline (BG/BB/LogNormal)	19.35	2.88	0.129	0.207
Mean (confidence interval) if top spenders were chosen at random			0.100 (0.062,0.140)	0.200 (0.165,0.238)

Table 10: Out-of-sample model fit for different specifications of the heterogeneity structure. The proportion of top spenders is computed by predicting total expenditure (per customer) over the first 15 periods, then selecting customers with highest predicted values. Numbers in bold mean that the model identifies customers of interest significantly better than chance.

	Parameter	St. Error	t-value	p-value
Intercept (λ_0)	0.093	0.030	**3.060**	0.002
Holiday	0.189	0.047	**4.019**	0.000
Online	0.757	0.055	**13.880**	0.000
NewProduct	−0.051	0.039	−1.320	0.187
Quantity	−0.017	0.021	−0.821	0.412
Avg.Price	−0.060	0.020	**−3.001**	0.003
Discount	0.020	0.002	**8.177**	0.000

Table 11: Results of regressing ΔCLV_i on first impressions. The t-values corresponding to parameters significantly different from zero are highlighted in bold.

Appendix: Likelihood derivation and prior specification

[1] Likelihood derivation

We are interested in computing the individual likelihood $\mathscr{L}_i(\boldsymbol{\eta_0}, \boldsymbol{\eta_i} \,|\, \boldsymbol{b_i})$ as a function of the transformed parameters $\left\{(p_{it},\ \phi_{it})_{t=1}^{T_i},\ \theta_i\right\}$. That is,

$$\mathscr{L}_i(\boldsymbol{\eta_0}, \boldsymbol{\eta_i} \,|\, \boldsymbol{b_i}) = p(b_{i1}, \ldots, b_{iT_i}),$$

where $b_{it} = (y_{it}, a_{it})$. Given the latent attrition variable z_{it}, we can identify five cases for the likelihood function,

$$p(b_{it}|z_{it}) = \begin{cases} 1 & \text{if } z_{it} = 1 \ \wedge y_{it} = 0 \ \wedge a_{it} = 0, \\ 0 & \text{if } z_{it} = 1 \ \wedge y_{it} = 1, \\ p_{it}\phi_{it} & \text{if } z_{it} = 0 \ \wedge y_{it} = 1 \ \wedge a_{it} > 0, \\ (1 - p_{it}) & \text{if } z_{it} = 0 \ \wedge y_{it} = 0 \ \wedge a_{it} = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

To simplify notation we will use T instead of T_i, however recall that the number of periods is a individual level specific quantity. In addition, we use $b_{i\,t_1:t_2}$ to denote $(b_{it_1}, \ldots, b_{it_2})$, and analogously, $z_{i\,t_1:t_2}$ to denote $(z_{it_1}, \ldots, z_{it_2})$. We will show that

$$\begin{aligned} p(b_{i\,1:T}) &= \sum_{z_{i\,1:T}} p(b_{i\,1:T}, z_{i\,1:T}) \\ &= \left(\prod_{\tau=1}^{tx_i} (p_{i\tau}\phi_{i\tau})^{y_{i\tau}} \, (1 - p_{i\tau})^{1-y_{i\tau}} (1 - \theta_i) \right) \cdot \sum_{t=tx_i+1}^{T+1} \left(\theta_i^{I\{t \leqslant T\}} \prod_{\tau=tx_i+1}^{t-1} (1 - \theta_i)(1 - p_{i\tau}) \right) \end{aligned}$$

Let $d_i \in \{1, \ldots, T_i + 1\}$ be the moment the customer leaves the firm (i.e., $z_{it} = 0 \ \forall t < d_i$, and $z_{it} = 1 \ \forall t \geqslant d_i$). We denote $d_i = T_i + 1$ if customer does not leave the firm during the observed period (i.e., $z_{it} = 0, \forall t \in \{1, \ldots, T_i\}$). Note that there is a bijection between a feasible full sequence of latent attrition variables $z_{i\,1:T} = (z_{i1}, \ldots, z_{iT})$ and the period of leaving the firm d_i. Thus, we can replace $z_{1:T}$ by d_i in the joint probability and marginalize out d_i to compute the marginal probability of the observed sequence of purchases $b_{i\,1:T_i} = (b_{i1}, \ldots, b_{iT_i})$.

We follow three stepts to compute the individual likelihood. First, we write the marginal of $(b_{i\,1:T})$ from the joint $p(b_{i\,1:T}, z_{i\,1:T}) = p(b_{i\,1:T}, d_i)$, using the change of variable $d_i = f(z_{i1}, \ldots, z_{iT})$). Then, we split the sum before and after the last observed purchase at period tx_i. Finally, we compute the likelihood using the specification of the model.

Step 1

Note that, by definition, there is a purchase at tx_i ($b_{i\,tx_i=1}$). Moreover, given that no purchase can be made after the death period d_i, the probability that death occurred before (including) the last observed purchase tx_i is zero (i.e., $p(d_i = t) = 0 \ \forall t \leqslant tx_i$). This property simplifies the calculations on the likelihood function, which can be written as follows

$$\begin{aligned}
p(b_{i\,1:T}) &= \sum_{z_{i\,1:T}} p(b_{i\,1:T}, z_{i\,1:T}) \\
&= \sum_{t=1}^{T+1} p(b_{i\,1:T} | d_i = t) \cdot p(d_i = t) \\
&= \sum_{t=1}^{T} p(b_{i\,1:T} | d_i = t) \cdot p(d_i = t) + p(b_{i\,1:T} | d_i = T+1) \cdot p(d_i = T+1) \\
&= \sum_{t=1}^{tx_i} p(b_{i\,1:T} | d_i = t) \cdot p(d_i = t) + \sum_{t=tx_i}^{T} p(b_{i\,1:T} | d_i = t) \cdot p(d_i = t)) \\
&\qquad + p(b_{i\,1:T} | d_i = T+1) \cdot p(d_i = T+1) \\
&= \sum_{t=tx_i+1}^{T} p(b_{i\,1:T} | d_i = t) \cdot p(d_i = t) + p(b_{i\,1:T} | d_i = T+1) \cdot p(d_i = T+1). \qquad (12)
\end{aligned}$$

Step 2

Now, splitting the sequence of observations $(b_{i\,1:T})$ between those that occurred before the last observed purchase $(b_{i\,1:tx_i})$ and those that occurred after $(b_{i\,tx_i:T})$, we apply Bayes rule[23] and re-write the joint probability as the product of the (marginal) probably of purchases before tx_i times the (conditional) probability of future purchases given past purchases. However, given the death

[23] $p(x, y) = p(x)p(y|x)$.

period, d_i, purchases are independent across periods. Therefore, we can drop $b_{i\,1:tx_i}$ from the conditional probability and write $p(b_{i\,1:T}|d_i = t)$ as follows,

$$\begin{aligned} p(b_{i\,1:T}|d_i = t) &= p(b_{i\,1:tx_i}, b_{i\,tx_i+1:T}|d_i = t) \\ &= p(b_{i\,1:tx_i}|d_i = t) \cdot p(b_{i\,tx_i+1:T}|d_i = t, b_{i\,1:tx_i}) \\ &= p(b_{i\,1:tx_i}|d_i = t) \cdot p(b_{i\,tx_i+1:T}|d_i = t). \end{aligned}$$

The information $d_i = t$ is equivalent to $z_{1:t-1} = 0$ and $z_{t:T} = 1$. Notice that as $t > tx_i$, for the purchases up to tx_i, the only relevant information from $d_i = t$ is that the $z_{1:tx_i} = 0$. Therefore, we can now write

$$p(b_{i\,1:T}|d_i = t) = p(b_{i\,1:tx_i}|z_{i\,1:tx_i} = 0) \cdot p(b_{i\,tx_i+1:T}|d_i = t) \tag{13}$$

Now, replacing (13) in (12), we obtain

$$\begin{aligned} p(b_{i\,1:T}) = &\sum_{t=tx_i+1}^{T} p(b_{i\,1:tx_i}|z_{i\,1:tx_i} = 0) \cdot p(b_{i\,tx_i+1:T}|d_i = t) \cdot p(d_i = t) \\ &+ p(b_{i\,1:T}|d_i = T+1) \cdot p(d_i = T+1) \\ = &\, p(b_{i\,1:tx_i}|z_{i\,1:tx_i} = 0) \cdot \sum_{t=tx_i+1}^{T} p(b_{i\,tx_i+1:T}|d_i = t) \cdot p(d_i = t) \\ &+ p(b_{i\,1:T}|d_i = T+1) \cdot p(d_i = T+1). \end{aligned} \tag{14}$$

We need to compute the following terms:

$$\begin{aligned} C &= p(b_{i\,1:tx_i}|z_{i\,1:tx_i} = 0), \\ D_t &= p(b_{i\,tx_i+1:T}|d_i = t), \\ E_t &= p(d_i = t), \\ F &= p(b_{i\,1:T}|d_i = T+1) \cdot p(d_i = T+1). \end{aligned}$$

with $t \in \{tx_i + 1, \ldots, T\}$.

Computation of C is straightforward. Using (11), given that customer i is alive, this probability corresponds to

$$\begin{aligned} & C = p(b_{i\,1:tx_i}|z_{i\,1:tx_i} = 0) \\ \implies & C = \prod_{\tau=1}^{tx_i} (p_{i\tau}\phi_{i\tau})^{y_{i\tau}} (1 - p_{i\tau})^{1-y_{i\tau}} \end{aligned} \tag{15}$$

By definition, there are no purchases after tx_i, implying that $b_{i\,tx_i+1:T} = 0$. In addition, there can only be purchase occasions before death at t. Notice that if $t = tx_i + 1$, it implies that there are no purchase occasions thus we observe $b_{i\,tx_i+1:T} = 0$ with probability 1. On the other hand if $t > tx_i + 1$, the customer has $t - (tx_i + 1)$ purchase occasions at which she did not purchase. As a result, we can write D_t as follows

$$D_t = p(b_{i\,tx_i+1:T}|d_i = t) = \begin{cases} \prod_{\tau=tx_i+1}^{t-1}(1 - p_{i\tau}) & \text{if } t > tx_i + 1 \\ 1 & \text{if } t = tx_i + 1. \end{cases} \tag{16}$$

Regarding E_t, note that the customer is alive from period 1 to tx_i and from $tx + 1$ to $t - 1$ ($z_{i\,1:t-1} = 0$), then she 'dies' at t ($z_{it} = 1$). This sequence of events corresponds to a non-homogeneous geometric distribution probability and can be written as follows

$$E_t = p(d_i = t) = \begin{cases} \left(\prod_{\tau=1}^{tx_i}(1 - \theta_i)\right)\theta_i \prod_{\tau=tx_i+1}^{t-1}(1 - \theta_i) & \text{if } t > tx_i + 1 \\ \left(\prod_{\tau=1}^{tx_i}(1 - \theta_i)\right)\theta_i & \text{if } t = tx_i + 1 \end{cases} \tag{17}$$

We highlight two properties from (16) and (17). The first parenthesis on (17) does not depend on t, therefore we can take it outside the sum of (14). In addition, the different cases in (16) and (17) are the same, therefore we can write the sum $\sum_{t=tx_i+1}^{T} D_t E_t$ as follows

$$\sum_{t=tx_i+1}^{T} D_t E_t = \left(\prod_{\tau=1}^{tx_i}(1 - \theta_i)\right) \cdot \sum_{t=tx_i+1}^{T}\left[\theta_i \prod_{\tau=tx_i+1}^{t-1}(1 - \theta_i)(1 - p_{i\tau})\right]. \tag{18}$$

Note that if $t = tx_i + 1$, the product term $\prod_{\tau=tx_i+1}^{t-1}(\cdot)$ does not contain any terms.

Finally, we note that the term F can be expressed in terms of C, D_{T+1}, and E_{T+1}. That is,

$$\begin{aligned} F &= p(b_{i\,1:T}|d_i = T + 1) \cdot p(d_i = T + 1) = p(b_{i\,1:tx_i}|d_i = T + 1) \cdot p(b_{i\,tx_i+1:T}|d_i = T + 1) \cdot p(d_i = T + 1) \\ &= C \cdot D_{T+1} \cdot E_{T+1} \\ &= C\left(\prod_{\tau=1}^{tx_i}(1 - \theta_i)\right) \cdot \left(\prod_{\tau=tx_i+1}^{T}(1 - \theta_i)(1 - p_{i\tau})\right). \end{aligned}$$

Step 3

Finally, we write the likelihood function (from (14)) as follows

$$
\begin{aligned}
p(b_{i\,1:T}) &= C \cdot \sum_{t=tx_i+1}^{T} D_t E_t + F = C\left(\sum_{t=tx_i+1}^{T} D_t E_t + C \cdot B_{T+1} \cdot E_{T+1}\right) \\
&= \left(\prod_{\tau=1}^{tx_i} \left(p_{i\tau}\phi_{i\tau}\right)^{y_{i\tau}} \left(1-p_{i\tau}\right)^{1-y_{i\tau}} (1-\theta_i)\right) \cdot \left[\sum_{t=tx_i+1}^{T} \left(\theta_i \prod_{\tau=tx_i+1}^{t-1} (1-\theta_i)(1-p_{i\tau})\right)\right. \\
&\qquad\qquad \left. + \prod_{\tau=tx_i+1}^{T} (1-\theta_i)(1-p_{i\tau})\right] \\
&= \left(\prod_{\tau=1}^{tx_i} \left(p_{i\tau}\phi_{i\tau}\right)^{y_{i\tau}} \left(1-p_{i\tau}\right)^{1-y_{i\tau}} (1-\theta_i)\right) \cdot \sum_{t=tx_i+1}^{T+1} \left(\theta_i^{I\{t\leqslant T\}} \prod_{\tau=tx_i+1}^{t-1} (1-\theta_i)(1-p_{i\tau})\right).
\end{aligned}
\tag{19}
$$

[2] Prior specification

We now specify the mixing distribution $p(\boldsymbol{\eta_i}|\mu_\eta, \Sigma_\eta)$ and the prior distributions $p(\eta_0)$, $p(\mu_\eta)$, and $p(\Sigma_\eta)$.

Mixing distribution

In order to simplify the sampling of the parameters of the model, the parameters that account for unobserved heterogeneity are assumed to be independent for each behavior and marketing action. In other words,

$$\boldsymbol{\eta_i}|\mu_\eta, \Sigma_\eta \sim \mathcal{N}(\mu_\eta, \Sigma_\eta),$$

where Σ_η is a diagonal matrix.

Certainly, a full covariance matrix could be incorporated, extending the model to allow for (1) correlations between base behaviors (e.g., incidence and attrition, incidence and amount); (2) correlation between base behavior and responsiveness to marketing actions (e.g., base incidence with amount sensitivity to email); and (3) correlation between responsiveness to marketing actions (e.g., incidence and amount sensitivities to direct marketing).

Prior distributions

The first set of parameters is $\eta_0 = [\alpha_0^\theta, \alpha_0^p, \alpha_0^a, \boldsymbol{\beta}_0^p, \boldsymbol{\beta}_0^a, \sigma_a]$. We assume independence between these components. That is,

$$p(\eta_0) = p(\alpha_0^\theta) \cdot p(\alpha_0^p) \cdot p(\alpha_0^a) \cdot p(\boldsymbol{\beta}_0^p) \cdot p(\boldsymbol{\beta}_0^a) \cdot p(\sigma_a).$$

Note that α_0^b for behaviors $b \in \{\theta, p, a\}$ are row vectors for which each component l represents how acquisition-related variable l explains heterogeneity in the base propensity of behavior b. On the other hand, $\boldsymbol{\beta}_0^b$ for behaviors $b \in \{p, a\}$ are matrices for which its $k - l$ component stands for how the l acquisition-related variable, explains differences in the sensitivity of behavior b to marketing action k. We assume independent prior distribution for all components of the mentioned vectors and matrices:

$$\alpha_{0l}^b \sim \text{Cauchy}(0, 0.8) \qquad \forall l, \forall b \in \{\theta, p, a\},$$

$$\beta_{0kl}^b \sim \text{Cauchy}(0, 0.8) \qquad \forall l, \forall k, \forall b \in \{p, a\}.$$

For σ_a, the standard deviation of the residual of the log-amount regression, we also assume Cauchy priors:

$$\sigma_a \sim \text{Cauchy}(0, 0.8).$$

The second set of parameters are the mean and covariance of the mixing distribution μ_η and Σ_η. Given that the parameters of incidence and death are in logit scale, and the parameters for amount are in the log scale, we use independent Cauchy distributions with parameter location 0, and scale 0.8 for all components of the mean parameters:

$$\mu_{\eta j} \sim \text{Cauchy}(0,\ 0.8)\ \forall j,$$

which generates a fairly uninformed prior on the log scale. Using the same argument, we restrict the space for Σ_η using a diagonal matrix with j component equal to $\sigma_{\eta j}^2$, for which

$$\sigma_{\eta j} \sim U[0.01,\ 4]\ \forall j.$$

As a result, the variance Σ_η on each of its components is allowed to take values ranging from 0.0001 to 16 on the logit scale.

First Impressions Count:
Leveraging Acquisition Data for Customer Management

Web Appendices

Web Appendix A: Log-likelihood computation

Equation (8) involves products and sums of probabilities, similar to the computation of likelihood for hidden Markov models (Schweidel and Knox 2013). Particularly, the sum on that equation involves a product of terms that repeat themselves for the different indexes of the sum. The goal of this appendix is to specify how we compute the individual likelihood function in the log scale for a robust implementation in Stan.

First, we denote LSE to the log-sum-exp function:

$$\underset{t=1}{\overset{T}{\mathrm{LSE}}}(x_t) := \log\left(\sum_{t=1}^{T}\exp(x_t)\right).$$

Taking log on Equation (8), we have that the individual log likelihood can be written as

$$\begin{aligned}\log(p(b_{i1},\ldots,b_{iT})) = &\sum_{\tau=1}^{tx_i} y_{i\tau}\log(p_{i\tau}\phi_{i\tau}) + (1-y_{i\tau})\log(1-p_{i\tau}) + \log(1-\theta_i)\\ &+ \underset{t=tx_i+1}{\overset{T+1}{\mathrm{LSE}}}\left(I\{t\leqslant T\}\log(\theta_i) + \sum_{\tau=tx_i+1}^{t-1}\log(1-\theta_i)+\log(1-p_{i\tau})\right).\end{aligned}$$

The first term is a sum of log probabilities. However, the second term involves two iterations, an outer one for the log-sum-exp function, and a inner sum of terms that share terms for different values of t.

Defining $\gamma_{it} = \sum_{\tau=tx_i+1}^{t-1}\log(1-\theta_i)+\log(1-p_{i\tau})$ for $t = tx+2,\ldots,T+1$, and $\gamma_{itx+1} = 0$, we can write γ_{it} recursively in the following manner

$$\gamma_{it} = \begin{cases} 0 & \text{for } t = tx+1,\\ \gamma_{it-1} + \log(1-\theta_i) + \log(1-p_{it-1}) & \text{for } t = tx+2,\ldots,T+1.\end{cases}$$

Finally, we define

$$\alpha_{it} = \begin{cases} \log(\theta_i) + \gamma_{it} & \text{if } tx + 1 \leqslant t \leqslant T \\ \gamma_{iT+1} & \text{if } t = T + 1, \end{cases}$$

and write the log likelihood as

$$\log(p(b_{i1}, \ldots, b_{iT})) = \sum_{\tau=1}^{tx_i} (y_{i\tau} \log(p_{i\tau}\phi_{i\tau}) + (1 - y_{i\tau}) \log(1 - p_{i\tau}) + \log(1 - \theta_i)) + \underset{t=tx_i+1}{\overset{T+1}{\mathrm{LSE}}} (\alpha_{it}) \,. \tag{A1}$$

Finally, using the above notation, we compute the log-likelihood using Algorithm 1:

input : $\{p_{it}\}$: purchase incidence probabilities,
$\{\theta_i\}$: death probabilities,
$\{\phi_{it}\}$: log amount density,
$\boldsymbol{b_i}$: observed behavior

output: LL_i: individual log-likelihood

for $i \leftarrow 1$ **to** I **do**
  $LL_i \leftarrow 0$;
  for $\tau \leftarrow 1$ **to** tx_i **do**
    $LL_i \leftarrow LL_i + y_{i\tau} \log(p_{i\tau}\phi_{i\tau}) + (1 - y_{i\tau}) \log(1 - p_{i\tau}) + \log(1 - \theta_i)$;
  end
  $\gamma_{i,tx_i+1} \leftarrow 0$;
  for $t \leftarrow tx_i + 2$ **to** T_i **do**
    $\gamma_{it} = \gamma_{it-1} + \log(1 - \theta_i) + \log(1 - p_{it-1})$;
    $\alpha_{it} = \log(\theta_i) + \gamma_{it}$;
  end
  $\gamma_{i,T_i+1} = \gamma_{i,T_i} + \log(1 - \theta_i) + \log(1 - p_{iT_i})$;
  $\alpha_{i,T_i+1} = \log(\theta_i) + \gamma_{it}$;
  $LL_i \leftarrow LL_i + \mathrm{LSE}_{t=tx_i+1}^{T_i+1} (\alpha_{it})$;
end

Algorithm 1: Log-likelihood computation

Web Appendix B: Results from additional models (M2 — M4)

In this appendix we present the full set of results for the four model specifications (M1 through M4) described in Section 3, demonstrating that the model parameters are robust to more restricted structures in the heterogeneity components. Table B1 shows the parameter estimates for the variables governing the base probabilities of all three processes (i.e., α_0^b, μ_α^b, and σ_α^b for $b \in \{\theta, p, a\}$). All parameter estimates are very consistent across specifications.

Regarding the effect of the time-varying covariates, Table B2 shows the parameters governing the effect of the time-varying covariates on purchase propensity (i.e., $\boldsymbol{\beta}_0^\theta$, μ_β^θ, and σ_β^θ), and Table B3 shows the results corresponding to the effect of the time-varying covariates on amount spent, given purchase (i.e., $\boldsymbol{\beta}_0^a$, μ_β^a, and σ_β^a).

	M1			M2			M3			M4		
	Posterior Mean	95% CPI		Posterior Mean	95% CPI		Posterior Mean	95% CPI		Posterior Mean	95% CPI	
Purchase												
μ_α	-3.351	-3.457	-3.241	-3.351	-3.458	-3.254	-3.320	-3.418	-3.224	-3.328	-3.427	-3.228
σ_α	1.194	1.150	1.238	1.190	1.148	1.232	1.180	1.133	1.231	–	–	–
Observed heterogeneity – First impressions (α_0)												
Holiday	-0.377	-0.498	-0.254	-0.299	-0.417	-0.195	-0.344	-0.483	-0.196	-0.295	-0.405	-0.190
Online	-0.088	-0.219	0.052	-0.120	-0.239	0.007	-0.117	-0.238	0.027	-0.117	-0.246	0.005
New Product	0.173	0.082	0.267	0.175	0.088	0.263	0.178	0.096	0.260	0.177	0.091	0.267
Quantity	0.092	0.038	0.150	0.098	0.046	0.148	0.103	0.039	0.163	0.098	0.044	0.153
Avg.Price	0.110	0.066	0.154	0.106	0.063	0.150	0.112	0.068	0.152	0.107	0.062	0.148
Discount	0.005	-0.001	0.011	0.005	-0.001	0.011	0.005	-0.001	0.011	0.005	-0.001	0.011
Latent churn												
μ_α	-3.851	-4.049	-3.658	-3.839	-4.040	-3.649	-3.834	-3.992	-3.651	-3.814	-4.009	-3.634
σ_α	0.304	0.028	0.694	0.296	0.022	0.684	0.413	0.027	0.793	–	–	–
Observed heterogeneity – First impressions (α_0)												
Holiday	0.165	-0.053	0.379	0.194	-0.014	0.398	0.203	-0.038	0.416	0.187	-0.012	0.386
Online	0.090	-0.156	0.314	-0.032	-0.279	0.203	0.052	-0.167	0.319	-0.025	-0.249	0.192
New Product	-0.114	-0.291	0.071	-0.092	-0.253	0.066	-0.116	-0.268	0.048	-0.090	-0.242	0.067
Quantity	0.102	0.025	0.174	0.102	0.028	0.170	0.114	0.035	0.195	0.102	0.027	0.168
Avg.Price	0.117	0.036	0.195	0.117	0.045	0.186	0.119	0.049	0.194	0.118	0.050	0.184
Discount	0.010	0.001	0.019	0.011	0.002	0.019	0.011	0.002	0.018	0.011	0.002	0.019
Amount												
μ_α	3.466	3.421	3.509	3.459	3.416	3.502	3.472	3.426	3.504	3.461	3.417	3.502
σ_α	0.414	0.397	0.434	0.413	0.394	0.431	0.422	0.399	0.440	–	–	–
Observed heterogeneity – First impressions (α_0)												
Holiday	-0.069	-0.121	-0.015	-0.055	-0.104	-0.009	-0.079	-0.124	-0.024	-0.055	-0.100	-0.010
Online	0.309	0.247	0.369	0.293	0.238	0.349	0.306	0.251	0.361	0.291	0.236	0.346
New Product	0.076	0.037	0.114	0.088	0.052	0.124	0.070	0.036	0.120	0.087	0.051	0.126
Quantity	0.127	0.102	0.153	0.119	0.095	0.143	0.125	0.100	0.149	0.120	0.098	0.145
Avg.Price	0.144	0.123	0.163	0.139	0.120	0.157	0.141	0.124	0.162	0.138	0.119	0.156
Discount	0.004	0.001	0.007	0.005	0.002	0.007	0.004	0.001	0.006	0.005	0.002	0.007

Table B1: Posterior mean and 95% central posterior interval (CPI) for parameters governing the individual propensities to purchase, latent churn, and amount spend (given transaction), for all model specifications.

Time-varying covariate		M1 Post. Mean	M1 95% CPI		M2 Post. Mean	M2 95% CPI		M3 Post. Mean	M3 95% CPI		M4 Post. Mean	M4 95% CPI	
Email	μ_β	0.126	0.082	0.171	0.091	0.059	0.120	0.141	0.104	0.183	0.111	0.084	0.138
	σ_β	0.218	0.148	0.283	0.217	0.136	0.284	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	-0.029	-0.104	0.043	–	–	–	-0.025	-0.091	0.037	–	–	–
	Online	0.081	-0.002	0.161	–	–	–	0.061	-0.022	0.145	–	–	–
	New Product	-0.081	-0.141	-0.021	–	–	–	-0.070	-0.122	-0.021	–	–	–
	Quantity	0.043	0.009	0.081	–	–	–	0.043	0.006	0.075	–	–	–
	Avg.Price	0.011	-0.016	0.038	–	–	–	0.009	-0.015	0.034	–	–	–
	Discount	-0.003	-0.006	0.001	–	–	–	-0.002	-0.005	0.001	–	–	–
Direct Marketing	μ_β	-0.087	-0.129	-0.045	-0.061	-0.089	-0.035	-0.063	-0.103	-0.024	-0.046	-0.068	-0.024
	σ_β	0.130	0.039	0.190	0.127	0.035	0.190	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.056	-0.005	0.114	–	–	–	0.043	-0.020	0.103	–	–	–
	Online	0.013	-0.059	0.083	–	–	–	0.017	-0.044	0.082	–	–	–
	New Product	0.012	-0.035	0.060	–	–	–	0.005	-0.039	0.050	–	–	–
	Quantity	-0.010	-0.041	0.023	–	–	–	-0.007	-0.040	0.023	–	–	–
	Avg.Price	0.015	-0.008	0.040	–	–	–	0.011	-0.010	0.034	–	–	–
	Discount	0.003	-0.001	0.006	–	–	–	0.002	-0.002	0.006	–	–	–
Product Introductions	μ_β	0.081	0.040	0.121	0.052	0.028	0.075	0.093	0.047	0.143	0.055	0.033	0.077
	σ_β	0.059	0.012	0.124	0.060	0.013	0.127	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	-0.018	-0.085	0.049	–	–	–	-0.020	-0.079	0.041	–	–	–
	Online	-0.077	-0.149	-0.008	–	–	–	-0.089	-0.176	-0.011	–	–	–
	New Product	-0.026	-0.075	0.019	–	–	–	-0.034	-0.084	0.016	–	–	–
	Quantity	0.028	-0.007	0.064	–	–	–	0.034	-0.005	0.071	–	–	–
	Avg.Price	0.021	-0.002	0.043	–	–	–	0.021	-0.003	0.041	–	–	–
	Discount	0.001	-0.003	0.004	–	–	–	0.001	-0.003	0.004	–	–	–
Seasonality	μ_β	0.598	0.464	0.722	0.700	0.608	0.784	0.635	0.522	0.731	0.723	0.654	0.790
	σ_β	0.374	0.056	0.629	0.363	0.042	0.636	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.469	0.283	0.642	–	–	–	0.447	0.250	0.631	–	–	–
	Online	0.021	-0.194	0.232	–	–	–	0.028	-0.152	0.207	–	–	–
	New Product	-0.008	-0.138	0.133	–	–	–	-0.020	-0.138	0.110	–	–	–
	Quantity	-0.044	-0.146	0.050	–	–	–	-0.043	-0.151	0.045	–	–	–
	Avg.Price	-0.056	-0.127	0.018	–	–	–	-0.050	-0.128	0.033	–	–	–
	Discount	0.004	-0.006	0.015	–	–	–	0.005	-0.005	0.014	–	–	–

Table B2: Posterior mean and 95% central posterior interval (CPI) for parameters governing the individual-level impact of the time-varying covariates on the propensity to purchase, for all model specifications.

Time-varying covariate		M1			M2			M3			M4		
		Post. Mean	95% CPI		Post. Mean	95% CPI		Post. Mean	95% CPI		Post. Mean	95% CPI	
Email	μ_β	0.028	0.002	0.052	0.028	0.012	0.044	0.029	0.004	0.056	0.029	0.012	0.044
	σ_β	0.035	0.011	0.073	0.033	0.011	0.069	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.012	-0.032	0.054	–	–	–	0.013	-0.030	0.051	–	–	–
	Online	0.001	-0.043	0.044	–	–	–	0.001	-0.038	0.045	–	–	–
	New Product	-0.006	-0.040	0.025	–	–	–	-0.010	-0.039	0.024	–	–	–
	Quantity	0.005	-0.012	0.023	–	–	–	0.006	-0.011	0.022	–	–	–
	Avg.Price	0.009	-0.007	0.025	–	–	–	0.009	-0.007	0.024	–	–	–
	Discount	0.000	-0.002	0.002	–	–	–	0.000	-0.002	0.002	–	–	–
Direct Marketing	μ_β	0.030	0.009	0.052	0.021	0.009	0.035	0.032	0.008	0.054	0.020	0.007	0.033
	σ_β	0.042	0.013	0.073	0.042	0.013	0.071	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.010	-0.026	0.046	–	–	–	0.016	-0.019	0.047	–	–	–
	Online	-0.019	-0.061	0.021	–	–	–	-0.022	-0.060	0.019	–	–	–
	New Product	-0.021	-0.049	0.004	–	–	–	-0.022	-0.046	0.005	–	–	–
	Quantity	0.004	-0.015	0.023	–	–	–	0.006	-0.013	0.030	–	–	–
	Avg.Price	0.000	-0.014	0.014	–	–	–	0.001	-0.012	0.013	–	–	–
	Discount	0.001	-0.001	0.003	–	–	–	0.001	-0.002	0.003	–	–	–
Product Introductions	μ_β	0.001	-0.025	0.026	0.002	-0.012	0.015	0.005	-0.021	0.028	0.003	-0.010	0.017
	σ_β	0.030	0.011	0.059	0.030	0.011	0.060	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.006	-0.039	0.048	–	–	–	0.012	-0.033	0.055	–	–	–
	Online	0.010	-0.033	0.051	–	–	–	0.006	-0.035	0.050	–	–	–
	New Product	0.004	-0.024	0.033	–	–	–	0.002	-0.022	0.029	–	–	–
	Quantity	0.018	-0.006	0.042	–	–	–	0.020	-0.004	0.041	–	–	–
	Avg.Price	0.006	-0.008	0.019	–	–	–	0.007	-0.006	0.020	–	–	–
	Discount	-0.001	-0.003	0.002	–	–	–	-0.001	-0.003	0.001	–	–	–
Seasonality	μ_β	0.125	0.047	0.203	0.171	0.127	0.214	0.135	0.064	0.197	0.172	0.130	0.215
	σ_β	0.304	0.213	0.378	0.298	0.211	0.379	–	–	–	–	–	–
	Observed heterogeneity – First impressions (β_0)												
	Holiday	0.080	-0.040	0.196	–	–	–	0.085	-0.021	0.182	–	–	–
	Online	-0.098	-0.233	0.040	–	–	–	-0.112	-0.218	0.005	–	–	–
	New Product	0.079	-0.017	0.168	–	–	–	0.071	-0.011	0.155	–	–	–
	Quantity	-0.047	-0.115	0.020	–	–	–	-0.030	-0.091	0.032	–	–	–
	Avg.Price	-0.036	-0.085	0.012	–	–	–	-0.032	-0.076	0.011	–	–	–
	Discount	0.002	-0.005	0.008	–	–	–	0.001	-0.005	0.007	–	–	–

Table B3: Posterior mean and 95% central posterior interval (CPI) for parameters governing the individual-level impact of the time-varying covariates on the amount spend (given purchase), for the model with time-specific effects.

Web Appendix C: Improvement in fit using first impressions

Where is model improvement coming from?

Reflecting on the value of adding first impressions to the model (i.e., the fact that the WCAI clearly favors M1, even though M2 already included unobserved heterogeneity), a natural question to ask is why this is the case. We posit that the gain mostly comes from customers with very few purchases — customers for whom unobserved individual sensitivities cannot be estimated with precision. In turn, when enough observations per customers are available, unobserved individual level parameters should be recovered accurately, thus diminishing the value of adding first impressions. However, when the researcher/analyst does not observe enough purchases per customer, those individual parameter estimates will mere approximate the population distribution. To corroborate this, we compute the individual-level "incremental benefit" of adding observed heterogeneity in customers' sensitivity to the marketing actions. We operationalize such benefit as the difference in WAIC (for each individual customer) between models M1 and M2. That is,

$$\text{Benefit}_i = \left(\text{WAIC}_i^{\text{M2}} - \text{WAIC}_i^{\text{M1}}\right).$$

We then regress this quantity on the number of purchases per individual, controlling for the total number of periods we observe each customer. Table B4 shows the results from such regression. As expected, the parameters for number of purchases is negative, implying that customers with fewer purchases during the observation period are those who benefit the most from including acquisition variables in the model. This result not only corroborates our intuition, but also implies that the use of acquisition information (i.e., first impressions) will be particularly useful to precisely estimate the sensitivity to marketing actions for customers who have been recently acquired and for whom the firm has not observed purchases yet; matter that we further explore in Section 5.4.2.

	Parameter	St. Error	t-value	p-value
Intercept	0.024	0.011	2.193	0.028
Number of purchases	−0.004	0.002	−1.721	0.085
Number of observations	0.000	0.000	0.662	0.508

Table B4: Results of regressing individual decrease in WAIC on the number of purchases per customer.

Web Appendix D: Details about the CLV calculation

We detail how we compute CLV for each type of acquired customer (Section 5.3). We denote by I_{type} = {*Default*, *Holiday*, *Online*, *NewProduct*, *Quantity*, *Avg.Price*, *Discount*}, the set of types of acquired customers. We compute CLV by drawing purchases from the data generation process (DGP) of the model. As these "customers" do not actually exist in our data, we do not have draws for the unobserved parameters for those type of customers. Thus, we assume all unobserved parameters are equal to the population mean (i.e. $\mathbb{E}(\eta_i|\text{Calibration}) = \widehat{\mu_\eta}$).

We set first impressions W_i, and marketing actions x_{it} according to what is described in Section 5.1. For each draw m from the posterior, we compute the individual level parameters as follows

$$\widehat{\alpha^b_{im}} = \widehat{\mu^b_{\alpha,m}} + \widehat{\alpha^b_0} \cdot W_i \qquad \forall i \in I_{type}, b \in \{p, \theta, a\}$$
$$\widehat{\boldsymbol{\beta}^b_{im}} = \widehat{\mu^b_{\beta,m}} + \widehat{\boldsymbol{\beta}^b_0} \cdot W_i \qquad \forall i \in I_{type}, b \in \{p, a\}$$

Finally, we use algorithm 2 to draw each type of CLV from the DGP.

input : I_{type}: set of types of acquired customers
$\{\mathbf{x}_{it}\}$: marketing actions
M: number of draws from posterior distribution
$\widehat{\eta_{0m}}$: posterior draw m of population parameters

output: $\widehat{y_{itm}}$: draw m of transaction made by customer i at period t
$\widehat{A_{itm}}$: draw m of amount bought by customer i at period t
$\widehat{\text{CLV}}_{im}$: draw m of CLV of customer type i

```
Initialize all variables to zero;
ŷ_itm ← 0, Â_itm ← 0;
for m ← 1 to M do
    for i ∈ I_type do
        d_im ~ Geometric(α̂^θ_im);
        for t ← 1 to min{d_im − 1, 51} do
            u ~ U[0,1];
            ε ~ N(0, σ̂²_am);
            p_itm ← logit^{-1}(α̂^p_im + β̂^p_im · x_it);
            ŷ_itm ← 1{u ≤ p_itm};
            if ŷ_itm = 1 then
                Â_itm ← exp(α̂^a_im + β̂^a_im · x_it + ε);
            else
                Â_itm ← 0;
            end
        end
        CLV̂_im ← Σ_{t=1}^{51} δ^t · Â^s_itm;
    end
end
```

$$\widehat{y}_{itm} \leftarrow 0,\ \widehat{A}_{itm} \leftarrow 0;\quad d_{im} \sim \text{Geometric}(\widehat{\alpha^{\theta}_{im}});\quad \varepsilon \sim N\left(0, \widehat{\sigma^2_{am}}\right);\quad p_{itm} \leftarrow \text{logit}^{-1}\left(\widehat{\alpha^{p}_{im}} + \widehat{\boldsymbol{\beta}^{p}_{im}} \cdot \mathbf{x}_{it}\right);\quad \widehat{y_{itm}} \leftarrow \mathbf{1}\{u \leqslant p_{itm}\};\quad \widehat{A_{itm}} \leftarrow \exp\left(\widehat{\alpha^{a}_{im}} + \widehat{\boldsymbol{\beta}^{a}_{im}} \cdot \mathbf{x}_{it} + \varepsilon\right);\quad \widehat{\text{CLV}}_{im} \leftarrow \sum_{t=1}^{51} \delta^t \cdot \widehat{A^s_{itm}};$$

Algorithm 2: CLV simulation

Web Appendix E: Details about the what-if analyses

In Section 5.4.2 we simulate behavior for a sample of 2,736 customers under different scenarios. The main goal of that analysis is to show how the use of first impressions can be leveraged to better target email and DM efforts to (recently-acquired) customers who are more likely to respond to those marketing actions. For simplicity, we choose the average frequency of emails and DM campaigns observed in the calibration data — every 3 periods for email and every 4 periods for DM, with intensity equal to one email and one DM for the periods each marketing communication is sent. Regarding the remaining time-varying covariates, new product introduction and seasonality, we assume the same values as observed in the data.

The what-if analyses involve two main steps. First, we specify the targeting policies. That is, we determine which newly-acquired customers will receive an email and/or a DM in each marketing campaign. For this step we use the insights obtained when estimating the model on the calibration data (i.e., the population-level parameters μ_β^b and $\boldsymbol{\beta}_0^b$) and the observed first impressions (i.e., W_i) of the newly acquired customers. Second, we simulate the behavior on each new customer depending on whether she receives the marketing action. In order to simulate customer behavior as accurately as possible,[D1] we separatelty estimate our model for these new customers, and use their individual-level estimates to predict behavior.

Note that we use the parameter estimates fro those new customers only to simulate behavior and not for setting the targeting policy. Doing so not only results in unbiased estimates of the effect of each policy (as we run an "out ot sample" validation) but also better approximates the targeting policy that the firm would be able to implement.

Targeting rules

We now describe how we set the targeting rules, and consequently marketing actions (x_{it}), for each one of the three scenarios: (1) *No marketing*, (2) *Current*, and (3) *Proposed.* Recall that we have four time-varying variables, of which, email and DM vary at the individual level (thus can be targeted) and product introduction and seasonality are market-level variables (common across all

[D1]Ideally one would conduct an experiment with newly acquired customers. However we do not have that option. Alternatively, we use our model to simulate behavior.

customers in each market). Therefore, to keep our scenarios realistic, we vary email and DM across all three scenarios, while fixing the pattern of new product introduction and seasonality. More specifically, we randomly select one customer from the calibration data, for which we observe 51 periods, and we use that particular history of observed new product introductions and seasonality for each new customer.

In the *No marketing* scenario, we set emails and DM to zero[D2]. For the other scenarios, we define a schedule of "campaigns," in which emails and DM ($CAMPAINGS_{em}$ and $CAMPAIGNS_{DM}$, respectively) are sent to selected customers. Email campaigns are active every three periods, starting at period 1; i.e., $CAMPAINGS_{em} = \{1, 4, 7, \ldots, 49\}$. DM campaigns on the other hand are active every 4 periods, starting at period 4. i.e., $CAMPAINGS_{DM} = \{4, 8, 12, \ldots, 48\}$. Everytime a campaign is active, only a fixed proportion of customers receive the corresponding email or DM (π_{em} and π_{DM}, respectively). We set $\pi_{em} = 0.5$ and $\pi_{DM} = 0.25$. If a customer is "chosen" to receive the marketing communication, we simulate the firm sending one email or DM respectively that particular period.[D3] Table D1 summarizes the differences in email and DM across the three scenarios.

	Campaign type		Proportion of customers being targeted		Targeting rule
	Email	DM	Email	DM	
(1) *No marketing*	None	None	0	0	–
(2) *Current*	Every 3 periods	Every 4 periods	0.50	0.25	Random per campaign
(3) *Proposed*	Every 3 periods	Every 4 periods	0.50	0.25	Highest purchase sensitivity

Table D1: Marketing actions for each what-if scenario.

Under the *Current* targeting policy, customers are randomly assigned to the target set. This target set may vary across period, this is, the randomization on who gets targeted the marketing communication is done at each period such that the campaign is active. On the other hand, under the *Proposed* targeting policy, the firm chooses the customers with highest (expected) sensitivity to each marketing action. As mentioned before, we operationalize such a behavior by combining the insights obtained when estimating the model on the calibration data with the first impressions observed from the newly acquired customers in their first purchase. In particular, we use the pos-

[D2]As variables are standardized before entering the model, we set values to $-\mu/\sigma$ for emails and DMs.

[D3]As variables are standardized, we set values to $(1-\mu)/\sigma$ for emails and DMs, when a customer is assigned to receive that particular marketing communication.

terior mean of the parameters obtained from the estimation using the calibration data to compute the expected individual level sensitivities of behavior b to marketing actions. We reference to the notation introduced in Section 4.1, for each customer i in the new set of customers, we compute

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{\beta}_i^b|\text{Calibration Data}\right) &= \mathbb{E}\left(\widetilde{\boldsymbol{\beta}_i^b}|\text{Calibration Data}\right) + \mathbb{E}\left(\boldsymbol{\beta}_0^b|\text{Calibration Data}\right) \cdot W_i \\ &= \widehat{\mu_\beta^b} + \widehat{\boldsymbol{\beta}_0^b} \cdot W_i,\end{aligned} \tag{D1}$$

where $\widehat{\mu_\beta^b}$ is the posterior mean vector of sensitivities of behavior b, $\widehat{\boldsymbol{\beta}_0^b}$ is the posterior mean matrix which captures idiosyncratic differences in sensitivities explained by first impressions, and W_i are the first impressions, characteristics observed for the new customers on their moment of acquisition.

As a conservative approximation to what the firm may be able to implement, we choose to target based on the effect of marketing actions on purchase incidence. More sophisticated targeting policies can be employed by including the effect on amount spend as well. Then, for each marketing action (i.e., email or DM), we propose targeting the customers with highest expected sensitivity to that specific marketing, as computed in (D1).

input : I_{oos}: set of out of sample customers
$\widehat{\mu^{p}_{\beta,em}}$: population mean purchase incidence sensitivity to email
$\widehat{\mu^{p}_{\beta,DM}}$: population mean purchase incidence sensitivity to DM
$\widehat{\boldsymbol{\beta}^{p}_{0,em}}$: acquisition-component purchase incidence sensitivity to email
$\widehat{\boldsymbol{\beta}^{p}_{0,DM}}$: acquisition-component purchase incidence sensitivity to DM
π_{em}: proportion of customers to target emails
π_{DM}: proportion of customers to target DM
Z_i: acquisition variables
output: PTARGET_{em}: proposed email target set
TARGET_{DM}: proposed DM target set

for $i \leftarrow 1$ **to** I_{oos} **do**
  $\widehat{\boldsymbol{\beta}^{em}_{i}} \leftarrow \widehat{\mu^{p}_{\beta,em}} + \widehat{\boldsymbol{\beta}^{p}_{0,em}} \cdot Z_i$;
  $\widehat{\boldsymbol{\beta}^{DM}_{i}} \leftarrow \widehat{\mu^{p}_{\beta,DM}} + \widehat{\boldsymbol{\beta}^{p}_{0,DM}} \cdot Z_i$;
end
$q_{em} \leftarrow \text{quantile}\left(\{\widehat{\boldsymbol{\beta}^{em}_{i}}\}_i, 1 - \pi_{em}\right)$;
$q_{DM} \leftarrow \text{quantile}\left(\{\widehat{\boldsymbol{\beta}^{DM}_{i}}\}_i, 1 - \pi_{DM}\right)$;
$\text{PTARGET}_{em} \leftarrow \varnothing$;
$\text{PTARGET}_{DM} \leftarrow \varnothing$;
for $i \leftarrow 1$ **to** I_{oos} **do**
  if $\widehat{\boldsymbol{\beta}^{em}_{i}} > q_{em}$ **then**
    $\text{PTARGET}_{em} \leftarrow \text{PTARGET}_{em} \cup \{i\}$;
  end
  if $\widehat{\boldsymbol{\beta}^{DM}_{i}} > q_{DM}$ **then**
    $\text{PTARGET}_{DM} \leftarrow \text{PTARGET}_{DM} \cup \{i\}$;
  end
end

Algorithm 3: Proposed targeting rules

Simulation algorithm

As discussed above, we simulate behavior using the individual-level parameters of each customer. Hence, we first estimate the model and obtain a sample of the posterior distribution of the parameters for the new customers. We then use these estimates to simulate behavior under three scenarios, only changing the value of the marketing actions (x_{it}).

No avoid simulation noise when comparing the targeting policies, we fix unobserved shocks across scenarios. For example, in order to draw y_{it}, a purchase occasion (if alive), we compute the purchase incidence probability p_{it} for each scenario and draw $u \sim U[0,1]$. That way, different scenarios may have different incidences probabilities, but the random shock u is constant across the three scenarios. In other words, at each observation, differences in purchases can only be a consequence of an increase in purchase incidence and not a consequence of random noise. Details about how we draw transactions and amount spend for the new customers are presented in Algorithm 4.

We aggregate the purchase and amount draws across individuals and obtain the total transactions and sales presented in Figure 3

$$\widehat{\text{Transactions}}^{s}_{tm} = \sum_{i=1}^{I_{oos}} \widehat{y^{s}_{itm}} \qquad \forall s \in S, t \in \{1, \ldots, 51\}, m \in \{1, \ldots, M\}$$

$$\widehat{\text{Sales}}^{s}_{tm} = \sum_{i=1}^{I_{oos}} \widehat{A^{s}_{itm}} \qquad \forall s \in S, t \in \{1, \ldots, 51\}, m \in \{1, \ldots, M\},$$

and compute posterior mean of individual CLV by aggregating across periods.

input : I_{oos}: set of out of sample customers
$S = \{No\ marketing, Current, Proposed\}$: set of scenarios
$\{\mathbf{x}_{it}^s\}$: marketing actions for scenario $s \in S$
M: number of draws from posterior distribution
$\widehat{\eta_{im}}$: posterior draw m of individual i parameters
$\widehat{\eta_{0m}}$: posterior draw m of population parameters
output: $\widehat{y_{itm}^s}$: draw m of transaction made by customer i at period t under scenario s
$\widehat{A_{itm}^s}$: draw m of amount bought by customer i at period t under scenario s
$\widehat{\mathrm{CLV}}_{im}^s$: draw m of CLV of customer i under scenario s

Initialize all variables to zero;
$\widehat{y}_{itm}^s \leftarrow 0$, $\widehat{A}_{itm}^s \leftarrow 0$;
for $m \leftarrow 1$ **to** M **do**
    for $i \leftarrow 1$ **to** I_{oos} **do**
        $d_{im} \sim \text{Geometric}(\widehat{\alpha_{im}^{\theta}})$;
        for $t \leftarrow 1$ **to** $\min\{d_{im} - 1, 51\}$ **do**
            $u \sim U[0, 1]$;
            $\varepsilon \sim N\left(0, \widehat{\sigma_{am}^2}\right)$;
            for $s \in S$ **do**
                $p_{itm}^s \leftarrow \text{logit}^{-1}\left(\widehat{\alpha_{im}^p} + \widehat{\boldsymbol{\beta}_{im}^p} \cdot \mathbf{x}_{it}^s\right)$;
                $\widehat{y_{itm}^s} \leftarrow \mathbf{1}\{u \leqslant p_{itm}^s\}$;
                if $\widehat{y_{itm}^s} = 1$ **then**
                    $\widehat{A_{itm}^s} \leftarrow \exp\left(\widehat{\alpha_{im}^a} + \widehat{\boldsymbol{\beta}_{im}^a} \cdot \mathbf{x}_{it}^s + \varepsilon\right)$;
                else
                    $\widehat{A_{itm}^s} \leftarrow 0$;
                end
            end
        end
        for $s \in S$ **do**
            $\widehat{\mathrm{CLV}}_{im}^s \leftarrow \sum_{t=1}^{51} \delta^t \cdot \widehat{A_{itm}^s}$;
        end
    end
end

Algorithm 4: Scenario simulation