Bayesian Imputation for Anonymous Visits in CRM Data

Julie Novak
IBM T.J. Watson Research Center
jenovak@us.ibm.com

Elea McDonnell Feit
Drexel University
efeit@drexel.edu

Shane T. Jensen
The Wharton School
stjensen@wharton.upenn.edu

Eric T. Bradlow
The Wharton School
ebradlow@wharton.upenn.edu

December 7, 2015

Abstract

Targeting individual consumers has become a hallmark of direct and digital marketing, particularly as it has become easier to identify customers as they interact repeatedly with a company. However, across a wide variety of contexts and tracking technologies, companies find that customers can not be consistently identified which leads to a substantial fraction of anonymous visits in any CRM database. We develop a Bayesian imputation approach that allows us to probabilistically assign anonymous sessions to users, while accounting for a customer's demographic information, frequency of interaction with the firm, and activities the customer engages in. Our approach simultaneously estimates a hierarchical model of customer behavior while probabilistically imputing which customers made the anonymous visits. We present both synthetic and real data studies that demonstrate our approach makes more accurate inference about individual customers' preferences and responsiveness to marketing, relative to common approaches to anonymous visits: nearest-neighbor matching or ignoring the anonymous visits. We show how companies who use the proposed method will be better able to target individual customers, as well as infer how many of the anonymous visits are made by new customers.

Keywords: Bayesian estimation, missing data, imputation, hierarchical modeling, targeted marketing

1 Introduction and Motivation

An important aspect of marketing practice is the targeting of consumers for differential promotional activity (cf. Rossi et al. 1996, Gordon 2010). Recent advancements in digital marketing and loyalty card programs have expanded companies' ability to track customers, thus increasing the popularity of targeted marketing (cf. Mulhern 2009, Winer 2009). Many companies keep extensive customer relationship management (CRM) databases that record their interactions with individual customers and use this past data to target them. However, despite the advancements in tracking technologies, companies still find that a large number of their interactions with their customers can not be matched to a particular customer and remain anonymous (Grover and Vriens 2006, Grabowski 2005). Marketers have long recognized this problem and have established generous incentive programs and other strategies to reduce anonymous visits (Nunes and Dreze 2006). For example, online retailers encourage customers to sign up for loyalty programs in order to receive special promotional emails (Tode 2007, Grant 2008). Yet, with few exceptions, companies consistently report that a large proportion of visits cannot be tracked back to an existing customer.

There are numerous different contexts where anonymous visits arise. A daily frequenter of a coffee shop might often pay with her credit card. However, some days she may prefer to pay with cash, resulting in a record of her purchase in the CRM database that is not tied to her customer ID. Another case where anonymous visits may occur is when an online customer frequents a clothing retailer's website. Although the customer has a user ID and sometimes logs in to the website, on some days she may browse without logging in, resulting in an anonymous visit and a loss of information about the customers preferences. Similarly, a media company with a “freemium” website will often get their paying subscribers to log in, but not always, if the user just wants to view some free content.

When companies compile customers' behavioral patterns over time to inform their direct marketing strategies, they do not typically attempt to link the anonymous visits to their other visits. However, there is a lot of potential information in anonymous visits; the data on anonymous visits still includes the time of visit, as well as the activities that the unknown customer engaged in.

We propose a Bayesian imputation approach that probabilistically assigns anonymous visits to customers based on observed behavioral patterns in the CRM database while simultaneously estimating a hierarchical model of consumer behavior. The model, and therefore the assignment of anonymous visits, accounts for the timing of visits (relative to the timing of each customer's observed visits), as well as the set of activities that the customer engages in during the visit (relative to the activities that all customers have engaged in). Using our approach, companies can better track the behavior of their customers, allowing them to better target those customers in the future. Under some circumstances, our approach could also be used to target a customer during an anonymous visit, based on real-time inference about their identity.

Accounting for the anonymous visits and probabilistically assigning them to known customers allows us to deepen our knowledge of each customer, which we will illustrate increases the precision of targeted advertising to both unidentified and identified customers. In addition, our model allows us to account for the anonymous visits when estimating overall features of a company's customer base. We will show that failing to account for anonymous visits can lead to erroneous inferences about critical business questions like, "How effective is my marketing overall?" and "How many customers do I have?"

We should point out that the systems for tracking users, and hence the prevalence of anonymous visits, varies across different situations. For example, on the web a visit can be associated with an identified customer through cookies, log-in or because the user clicked an ad sent to them by email, whereas in a coffee shop, customers can be tracked through use of a credit card or loyalty card. The method we develop is agnostic to the tracking technology; so long as users regularly interact with the company, which we generically call visits, and engage in activities during those visits, e.g. purchasing or browsing in certain categories, visiting certain webpages, viewing video content, etc.

To illustrate our method's ability to assign anonymous visits and infer customers' responsiveness to marketing, we present several parameter recovery studies: the first with simulated data and the second using data from a specialty retailer where the true visits are known, but we anonymize visits in a non-ignorable way. We compare our approach to the common practice of removing anonymous visits from the analysis by estimating our hierarchical consumer behavior model using only the identified visits without the anonymous imputation step (complete case analysis in the language of missing data). If customers who tend to make anonymous visits are also the customers who are more (or less) responsive to marketing, then the missing information is non-ignorable (Little and Rubin 2002) for the inferential goal of determining the effectiveness of marketing and to the wrong customers. When this happens, ignoring the anonymous visits results in biased effects, potentially leading the firm to engage in too little (or too much) marketing. Our synthetic data studies illustrate this bias and we show that accounting for the anonymous visits using our Bayesian approach obtains more precise estimates of the effects of marketing actions on customers.

Our proposed Bayesian approach simultaneously estimates the model and probabilistically assigns the anonymous visits to customers. This is a more coherent approach to inference than the common practice of imputing the missing data as a first step and then doing model inference as a separate subsequent step typically without accounting for the uncertainty in the imputation (albeit multiple imputation can also reflect this uncertainty). We compare our simultaneous approach to an impute-then-estimate strategy where anonymous visits are first assigned deterministically to customers via nearest-neighbor matching based on the activities in the visit, and then our model of customer behavior is estimated from that imputed data. As we will show, this two-step approach underestimates parameter uncertainty by not accounting for the uncertainty in the imputation and can also lead to bias when the missing data mechanism is non-ignorable.

Like most other applications of Bayesian missing data methods in marketing, we impute missing values using Bayesian data augmentation (Tanner and Wong 1987), where data is imputed simultaneously with the model parameters as part of a Markov chain Monte Carlo sampler (Geman and Geman 1984).

Data augmentation has been used to account for survey non-response (Ying et al. 2006), analyze split questionnaires (Adigüzel and Wedel 2008), handle covariate information that is only available in the aggregate (Musalem et al. 2008), and to address the issue of having some outcomes observed at the individual level and others in the aggregate (Feit et al. 2013). The present work represents a new application of these methods to the important problem of accounting for anonymous visits when analyzing CRM data.

As a motivation for our general model, we provide an example of a typical CRM data set with anonymous visits in Table 1. Each row in this data table represents a customer visit. For the identified visits, the data contains the User ID, the time stamp of the visit, indicators for which activities the customer engaged in during the visit, and potentially some demographic information about the customer. When a customer is not identified, we still have time stamp and the activity indicators; however, we no longer have the customer's ID number or (likely) the demographic information.

In the fourth row of Table 1, there is an anonymous visit where an unknown customer arrived at 13:24:24 on 2010-01-01 and purchased shoes but did not purchase pants[1]. From

[1] Note that we use purchase as an example here, but the idea of an activity can be generalized to include browsing products, consuming media or any other activity that the customer might engage in.

Table 1: Typical CRM Data Table with Anonymous Visits

Time j	User ID U_j	Activity 1: Shoes y_{j1}	Activity 2: Pants y_{j2}	Age Z_{j1}	Gender Z_{j2}
2010-01-01 12:46:49	16	1	0	34	0
2010-01-01 12:50:47	19	1	1	17	1
2010-01-01 13:20:54	3	0	0	19	0
2010-01-01 13:24:24	**?**	1	0	**?**	**?**
2010-01-01 13:25:00	27	0	1	45	1
2010-01-01 13:26:07	5	1	1	20	1
2010-01-01 14:10:09	16	1	0	34	0
2010-01-01 15:12:00	12	0	0	12	0

the first and seventh rows in Table 1, we can see that customer 16 visited twice, purchasing shoes but not pants in both of her visits, which is the same purchasing behavior as the anonymous visitor. This similarity in purchasing behavior as well as the fact that customer 16 visits more frequently than everyone else both increases the probability (based on the likelihood function we will define in the next section) that they are the anonymous visitor.

Note that in CRM data like that depicted in Table 1, there are no visits that are completely missing. All transactions are recorded; the missing data problem arises from not being able to associate some of those transactions with a specific user ID. This makes our problem distinct from that of completely missing records in media panel data described by Goerg et al. (2015), where panelist are supposed to track their own visits, but sometimes don't.

The remainder of the paper is as follows. In Section 2, we develop a model of customer behavior that can be applied to any CRM data set where customers engage in visits and activities, such as repeated website visits, repeated transactions, regular use of a media site, etc. Our model provides both a framework for understanding repeated customer behavior and a basis for imputing anonymous visits. We outline a Bayesian approach to missing data for the specific case where some customer visits are anonymous, and where we also infer how many of the anonymous visits come from new, previously unobserved customers (along with the imputation of demographics for these unobserved customers). In Section 3, we compare our approach to two common alternatives, complete case analysis and impute-then-estimate (via nearest-neighbor matching), and show that our method is better able to recover individual-level parameters by accounting for anonymous visits, which leads to better targeting of individual customers. Our approach performs particularly well relative to the alternatives when there is a substantial number of anonymous visits and the missingness mechanism is non-ignorable, i.e. correlated with other parameters of the model. In Section 4, we apply our methodology to a specialty retailer's dataset where we have artificially anonymized some visits, and show that the method still performs well when the data generating mechanism is not known. In Section 5, we conclude with a discussion of the findings and areas for future study.

2 Bayesian Hierarchical Model for CRM data

2.1 Modeling Customer Activities

The data structure in Table 1 occurs in many contexts, such as repeated website visits, repeated transactions at a retailer or a restaurant, or repeated visits to a service provider like a gym or library. In this section, we lay out a general model for repeated customer behavior that combines an exponential model for the interarrival times of visits by a customer with a multivariate probit model for the activities engaged in by that customer. Both models are embedded in a Bayesian hierarchical framework, where customer demographics can enter as covariates for the arrival rate of each customer as well as each customer's propensity to engage in activities.

As outlined in Table 1, let $j = 1, \ldots, n$ index a set of n observed (but not necessarily identified) customer visits. The variable $U_j \in \{1, \ldots, I\}$ indicates which of the I unique users made visit j. In this section, where we build up our hierarchical model for consumer behavior, we will assume that each U_j is observed; we address unknown U_j for anonymous visits in Section 2.3. At each visit, we observe a set of discrete variables $\boldsymbol{y}_j = (y_{j1}, \ldots, y_{jM})$ indicating which of the M possible activities the customer engaged in during that visit. In our retailer example, y_{jm} takes on values 0 or 1, indicating whether or not the customer purchased items in category m, such as women's shoes, housewares, etc. In other applications, y_{jm} could be ordinal counts or a continuous variable (e.g. dollar spend) in which case we would need to substitute an appropriate link function.

We model the observed indicators $\boldsymbol{y}_j$ for the activities engaged in by the customer U_j during visit j using a multivariate probit regression model (Rossi et al. 2005, Chib and Greenberg 1998),

$$y_{jm} = \begin{cases} 1 & \text{if } y^{\star}_{jm} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where $y^{\star}_{jm}$ is customer U_j's latent underlying utility to engage in activity m on visit j.

In many marketing datasets, there is also the possibility that direct marketing was sent to customers. We build marketing actions into our multivariate probit framework by allowing $\boldsymbol{y}^{\star}_j = (y^{\star}_{j1}, \ldots, y^{\star}_{jM})$ to depend on whether marketing was in-effect during the time of the visit,

$$y^{\star}_{jm} = \nu_{U_j,m} + \boldsymbol{\beta}^T_{U_j,m} \boldsymbol{X}_{jm} + e_{jm} \qquad \text{where} \qquad \boldsymbol{e}_j \sim N(0, \boldsymbol{\Sigma}) \tag{2}$$

where $\boldsymbol{X}_{jm}$ is a vector of length P_x that indicates the number of different marketing actions that the firm took to encourage customers to engage in activity m. In our retail example, if activity m was the potential purchase of shoes then $\boldsymbol{X}_{jm}$ could be a scalar indicator which takes on the value 1 if there was an advertisement for shoes sent to the customer (at some specified time period, discussed later) before visit j and 0 otherwise.

The customer-specific parameters in our multivariate probit model consist of intercepts, $\nu_{U_j,m}$ which characterize customer U_j's overall propensity to engage in each activity m, and coefficients, $\boldsymbol{\beta}_{U_j,m}$, which characterize customer U_j's response to visit-specific marketing actions for each activity m. In our retail example, $\nu_{U_j,\text{shoes}}$ would be the underlying propensity for customer U_j to purchase shoes without any marketing action. If the store sends this customer an advertisement, their underlying utility for purchasing shoes would increase by $\beta_{U_j,\text{shoes,ad}}$. We also have a population-level correlation structure, $\boldsymbol{\Sigma}$, among all the activities, as was done in Manchanda et al. (1999), to accommodate the possibility that some activities tend to occur together during the same visit, e.g., purchasing women's tops and women's skirts.

2.2 Modeling Customer Visits

Note that in Table 1 we also observe a time stamp for each visit. To model the rate of visitation for each customer, we let a_{U_j,t_j} denote the interarrival time between the $t_j - 1^{th}$ visit and the t_j^{th} visit by customer U_j. While j indexes the visits among *all* the customers in the dataset, t_j indexes the visits that correspond to a *specific customer*, U_j. We assume that the times between visits follow a heterogeneous covariate-driven exponential distribution given by

$$a_{U_j,t_j} \sim \text{Exponential}\ (\lambda_{U_j,t_j}) \tag{3}$$

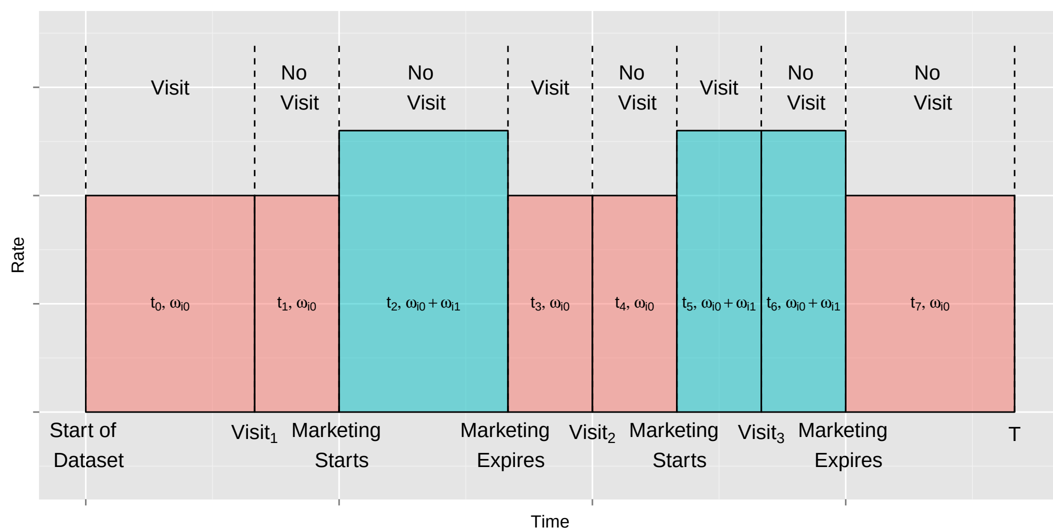
with customer-specific arrival rates λ_{U_j,t_j} that also can change over time due to marketing actions taken by the firm. Specifically, we assume that customer U_j's arrival rate λ_{U_j,t_j} during time period t_j is comprised of two components: (i) $\omega_{U_j,0}$, the baseline arrival rate for that customer (independent of time and marketing activity) and (ii) $\boldsymbol{\omega}_{U_j,1}$, a P_H-dimensional vector of effects on that customer's arrival rate from P_H marketing actions, $\boldsymbol{H}_{U_j,t_j}$,

$$\log \lambda_{U_j,t_j} = \omega_{U_j,0} + \boldsymbol{\omega}_{U_j,1}\boldsymbol{H}_{U_j,t_j} \tag{4}$$

Note that some elements of $\boldsymbol{H}_{U_j,t_j}$ may be the same as $\boldsymbol{X}_{jm}$ if there is a marketing action that affects both customer U_j's arrival rate and propensity to undertake activities simultaneously.

Figure 1: **An Example of a Customer's Rates of Arrival**

We split each customer's lifespan in the dataset into a series of periods with varying arrival rates. These periods can start and end with any of the following: a start of a marketing action, an end of a marketing action, and a visit. We take the product of the likelihood for all such events for each customer to obtain their arrival likelihood.



As shown by an example in Figure 1, we can segment each customer's arrival rates into periods that begin and end with the start of a marketing action, the expiration of a marketing action or a visit. Customer U_j visits at a constant underlying baseline rate $\omega_{U_j,0}$. Upon receiving a marketing action, the customer's underlying visitation rate changes to $\omega_{U_j,0} + \omega_{U_j,1} H_{U_j,t_j}$ and continues at that rate until the marketing action "expires" at which point the customers visitation rate drops back to their baseline, $\omega_{U_j,0}$. In Table 2, we delineate the likelihood function based on these arrival rates for all possible start and end events for a time segment.

Table 2: **Likelihoods for Intervals between Visits**

In the absence of a marketing action, customer U_j has an underlying rate of arrival of $\omega_{U_j,0}$. Upon receiving a marketing action, customer U_j's arrival rate increases to $\omega_{U_j,0} + \omega_{U_j,1}$ for a fixed length of time. F_{exp} is the cumulative distribution function and f_{exp} is the probability density function for the exponential distribution.

	Interval End Event		
Interval Start Event	Marketing Starts	Marketing Expires	Visit
Marketing Starts	$F_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$	$F_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$	$f_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$
Marketing Expires	$F_{exp}(t_j, \omega_{U_j,0})$	$F_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$	$f_{exp}(t_j, \omega_{U_j,0})$
Visit	$F_{exp}(t_j, \omega_{U_j,0})$	$F_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$	$f_{exp}(t_j, \omega_{U_j,0})$, if not within marketing period $f_{exp}(t_j, \omega_{U_j,0} + \omega_{U_j,1})$ if within marketing period

For the example in Figure 1, we can construct the arrival likelihood $L_{\lambda_{U_j}}$ for customer U_j by taking a product over all the consecutive periods between the start of the dataset and terminal time point T.

$$\begin{aligned} L_{\lambda_{U_j}} &= (\omega_{U_j,0}\exp[-\omega_{U_j,0}t_0]) \times (1-\exp[-\omega_{U_j,0}t_1]) \times (1-\exp[-(\omega_{U_j,0}+\omega_{U_j,1})t_2]) \\ &\times (\omega_{U_j,0}\exp[-\omega_{U_j,0}t_3]) \times (1-\exp[-\omega_{U_j,0}t_4]) \\ &\times ([\omega_{U_j,0}+\omega_{U_j,1}]\exp[-[\omega_{U_j,0}+\omega_{U_j,1}]t_5]) \\ &\times (1-\exp[-(\omega_{U_j,0}+\omega_{U_j,1})t_6]) \times (1-\exp[-\omega_{U_j,0}t_7]) \end{aligned} \tag{5}$$

We note that our framework assumes that all direct marketing communications last for a fixed period of time (similar to Sahni et al. 2014), but not following a Koyck-like decay function that is sometimes utilized in the literature. We have also not accounted for any cumulative effect of receiving multiple emails in a single time period. Our focus is on accounting for the presence of marketing action when we make our anonymous visit imputation and a more complicated advertising response model is outside our scope, but within the current modeling framework given in equations (4) and (5).

With the likelihood for activities defined in equations 1-2, and the likelihood for arrivals times defined in equations 3-4, we can now write the joint likelihood function for customer visits,

$$P(\boldsymbol{y},\boldsymbol{A},\boldsymbol{U}|\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\Sigma},\boldsymbol{y}^{\star}) = \prod_{j=1}^{n}\prod_{U_j=1}^{I}[(\int_{G_{U_j,M}}\dots\int_{G_{U_j,1}} \Phi_M\{\boldsymbol{y^\star}_j|\boldsymbol{\nu_{U_j}}+\boldsymbol{\beta_{U_j}^T}\boldsymbol{X_j},\boldsymbol{\Sigma}\}d\boldsymbol{y^\star}_j)L_{\lambda_{U_j}}]^{I(U_j)} \tag{6}$$

Note that arrival times are censored given that no arrivals are observed after a terminal time point T (which enters the likelihood through the survival $L_{\lambda_{U_j}}$ term).

In our Bayesian approach, we specify priors for all customer-specific parameters $\boldsymbol{\theta}_i^T = (\boldsymbol{\nu}_i, \boldsymbol{\beta_i}, \omega_{i0}, \boldsymbol{\omega_{i,1}})$ as a function of both customer-specific demographic covariates $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2}, \dots, Z_{iS})$ and population-level regression coefficients, $\boldsymbol{\Gamma}$. Specifically, we model each customer's parameter vector, $\boldsymbol{\theta}_i$, with a hierarchical multivariate regression,

$$\boldsymbol{\theta}_i = \begin{pmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\beta}_i \\ \omega_{i0} \\ \boldsymbol{\omega}_{i1} \\ \text{logit } \delta_i \end{pmatrix} \sim MVN(\boldsymbol{\Gamma Z_i}, \boldsymbol{\Omega}) \tag{7}$$

which allows for correlation between different customer-specific parameters. We impose conjugate multivariate normal and inverse Wishart prior distributions on the population-level model parameters, $\boldsymbol{\Gamma}, \boldsymbol{\Omega}$, and $\boldsymbol{\Sigma}$ (Gelman et al. 2003) which is also more fully described in Appendix A. The extra (and yet to be described) customer-specific parameter δ_i specifies the propensity for customer i to be anonymous, which we address by extending our model as given next in Section 2.3.

2.3 Modeling Anonymous Visits

We next address the primary goal of our approach: modeling the anonymous visits in CRM data. As given in Table 1, $U_j = i$ represents the user ID for visit j, but we must now account for the fact that U_j can be unknown. We define a missing data indicator V_j = 1 if the user for visit j is unknown, and 0 otherwise. Let $\boldsymbol{U^{obs}}$ be the subset of $\boldsymbol{U}$ when $V_j = 0$, and let $\boldsymbol{U^{mis}}$ be the subset of $\boldsymbol{U}$ when $V_j = 1$.

We also let δ_i be the probability that user i will be anonymous during a visit, i.e., δ_i is the probability that V_j=1 conditional on $U_j = i$. Note that δ_i is specific to each customer which allows for some customers to be more likely to make anonymous visits than others.

We will simultaneously estimate both the anonymous customers IDs in $\boldsymbol{U^{mis}}$ and the parameters of the model using the aforementioned Bayesian data augmentation approach where we estimate the joint posterior distribution,

$$P(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{\Sigma}, \boldsymbol{U^{mis}} | \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{U^{obs}}) \propto \prod_{j=1}^{n} \prod_{i=1}^{I} [(\int_{G_{U_j,M}} \cdots \int_{G_{U_j,1}} \Phi_M \{ \boldsymbol{y^\star}_j | \boldsymbol{\nu_{U_j}} + \boldsymbol{\beta_{U_j}^T X_j}, \boldsymbol{\Sigma} \} d\boldsymbol{y^\star}_j)$$
$$\times L_{\lambda_{U_j}} \times \delta_{U_j}^{(V_j=0)} (1 - \delta_{U_j})^{(V_j=1)}]^{I_{(U_j=i)}} P(\boldsymbol{\theta}, \boldsymbol{Z}, \boldsymbol{\Sigma}) \tag{8}$$

2.4 Model Estimation

We estimate the joint posterior distribution of all model parameters, including the customer IDs for any anonymous visits, using a Gibbs sampler, a Markov chain Monte Carlo method (Geman and Geman 1984, Rubin and Schenker 1986). Pseudo-code for our Gibbs sampler is as follows:

1. Sample a specific user for each missing U_j from a multinomial distribution.

2. Sample $\boldsymbol{y}^{\star}$ from a truncated multivariate normal distribution.

3. Sample $\boldsymbol{\theta}_i = (\boldsymbol{\nu}_i, \boldsymbol{\beta}_i, \omega_{i,0},)^T$ for each customer.

    a. Sample $(\boldsymbol{\nu}_i, \boldsymbol{\beta}_i)$ from a conjugate multivariate normal distribution.

    b. Sample $\omega_{i,0}$ and $\boldsymbol{\omega}_{i,1}$ using a Metropolis-Hastings step (Hastings 1970).

    c. Sample logit δ_i using a Metropolis-Hastings step (Hastings 1970).

4. Sample $\boldsymbol{\Gamma}$ from a conjugate multivariate normal distribution.

5. Sample $\boldsymbol{\Sigma}$ from a conjugate inverse-Wishart distribution.

6. Sample $\boldsymbol{\Omega}$ from a conjugate inverse-Wishart distribution.

7. Sample $\boldsymbol{Z}_i$ for each anonymous customer from a conjugate multivariate normal distribution.

In step 1, we sample a specific user for each anonymous U_j from a multinomial distribution where the probability of visit j being made by user k (conditional on current values of the other model parameters) is:

$$P(U_j^{mis} = k | \boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{y}^{\star}) = \tag{9}$$

$$\frac{(\int_{G_{kM}}\cdots\int_{G_{k1}}\Phi_M\{\boldsymbol{y^\star}_j|\boldsymbol{\nu_{U_j,k}}+\boldsymbol{\beta^T_{U_j,k}X_{jk}},\boldsymbol{\Sigma}\}d\boldsymbol{y^\star}_j)L_{\lambda_{U_k}}\delta_k^{(V_j=1)}(1-\delta_k)^{(V_j=0)}}{\sum_{i=1}^{I}(\int_{G_{iM}}\cdots\int_{G_{i1}}\Phi_M\{\boldsymbol{y^\star}_j|\boldsymbol{\nu_{U_j,i}}+\boldsymbol{\beta^T_{U_j,i}X_{ji}},\boldsymbol{\Sigma}\}d\boldsymbol{y^\star}_j)L_{\lambda_{U_i}}\delta_i^{(V_j=1)}(1-\delta_i)^{(V_j=0)}}$$

Using the probabilities given in (9), we probabilistically select customers who have the highest probability of making the anonymous visit based on the time of arrival, their demographic information, and the targeted advertisements they received. Once all $\boldsymbol{U}_j^{mis}$ are sampled, we sample the other parameters from their full conditional distributions in steps 2-7, which are outlined in detail in Appendix 5 . In this way, we simultaneously obtain draws from the posterior of $\boldsymbol{U}_j^{mis}$ and the model parameters. Thus, we can incorporate the anonymous visits in our model estimation in a way that utilizes all the information from both observed and anonymous visits.

2.5 Allowing for Additional Customers among Anonymous Visits

A key issue in the context of CRM data with anonymous visits is accounting for new customers who have never been identified in the data because all of their visits are anonymous. This information is valuable to firms for gauging their total customer base as well as distinguishing between new and repeat customers in measuring customer lifetime value, churn rate, and company value.

To infer the number of new customers, we create R *potential* new customers that could be assigned to the anonymous visits in pseudo-code Step 1. Within our Gibbs sampling scheme, we allow for each anonymous visit to be assigned to either an identified customer or to one of the potential new customers, depending on the distributions of arrival rates and visit propensities inferred from the data. The true number of unique customers can be no more than I, the total number of anonymous visits plus the total number of previously observed customers.

When assigning anonymous visits to new customers, we note that these unobserved customers (likely) have missing demographic information that can also be inferred from the customer's behavior on the visit. This inference is based on the frequency of visits and the activities during those visits and how those correlate with demographics among the identified customers. Estimating their user-specific characteristics may in some cases provide the company with a more accurate assessment of the demographics of their customer base, helping them optimize their assortment of products, create appealing advertising, etc.

To draw the demographics for new customers, we take advantage of the relationship $\boldsymbol{\theta}_i \sim MVN(\boldsymbol{\Gamma Z}_i, \boldsymbol{\Omega})$ given in equation 7. Following the usual approach to missing regressors by treating the matrix $\boldsymbol{\Gamma}$ as the regressors, and the $\boldsymbol{Z}_i$ as the parameter vector, we simply switch what we consider to be the covariates and regression coefficients. We sample $\boldsymbol{Z}_i$ for each new customer as follows:

$$\boldsymbol{Z}_i | \boldsymbol{\Omega}, \boldsymbol{U}, \boldsymbol{\Gamma}, \boldsymbol{\theta}_i \sim MVN(\hat{\boldsymbol{Z}}_\star, \boldsymbol{V}_{\boldsymbol{Z}_\star}) \tag{10}$$

where $\hat{\boldsymbol{Z}}_\star = (\boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma} + \boldsymbol{P_0})^{-1} (\boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\theta}_i + \boldsymbol{P_0} \boldsymbol{\xi_0})$ and $\boldsymbol{V}_{\boldsymbol{Z}_\star} = (\boldsymbol{\Gamma}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Gamma} + \boldsymbol{P_0})^{-1}$, with $\boldsymbol{P_0}$ and $\boldsymbol{\xi}_0$ being prior parameters that can either be set given a researchers knowledge of demographic frequencies in their purchasing population, or set as non-informative if appropriate.

2.6 Alternative Approaches to Anonymous Visits

We will compare our Bayesian model for imputing anonymous visits to two common alternative approaches for missing customers in CRM data. The most common way to analyze CRM data with anonymous visits is *complete-case analysis*: remove the anonymous visits from the dataset and then proceed with the analysis. Although complete-case analysis is clearly straightforward, we will demonstrate that it results in a loss of efficiency (i.e. greater uncertainty in our inference) and can result in bias when the missingness mechanism is non-ignorable, i.e. correlated with other model parameters (Rubin 1976). In these situations, the complete cases can not be relied upon as a representative sample of all possible data (Little and Rubin 2002, Graham et al. 1994).

A more sophisticated alternative would be impute-then-estimate: use a matching algorithm to make a deterministic imputation and then proceed with the analysis. We illustrate this approach on CRM data with anonymous visits by using nearest-neighbor matching (Hastie et al. 2009) to match anonymous visits to the closest identified customer visit based on the observed activities in the visits.

In order to match anonymous visits to the closest observed visit, we define a distance between each anonymous visit i and each visit k with a known customer based on the Mahalanobis distance between the vector of activities $\boldsymbol{y}_j$ in the anonymous visit and the vectors of activities $\boldsymbol{y}_k$ from each visit k by a known customer:

$$d(j,k) = (\boldsymbol{y}_j - \boldsymbol{y}_k)^T S_{yy}^{-1} (\boldsymbol{y}_j - \boldsymbol{y}_k) \tag{11}$$

where S_{yy} is an estimate of the covariance matrix of y_j across all anonymous and identified visits. Incorporating the covariance matrix means that categories which have high variability will carry less weight when finding potential candidates to match to the anonymous visits. For each anonymous visit k, we then deterministically impute the customer k for that visit that has the smallest distance $d(j,k)$, randomly breaking ties if needed. We also note that there are other approaches that could be done to match anonymous visits to observed visits, such as matching to an observed customer based upon an average across all their visits (as opposed to matching each anonymous visit to an observed visit).

Once the matching routine is complete, one can proceed with any type of analysis on the now-complete CRM data. As we show in the next section, this impute-then-estimate approach works better than the complete case analysis, but still fails to match customers as well as our Bayesian imputation model.

3 Synthetic Data Evaluation

We consider several synthetic data settings to explore how our proposed Bayesian imputation method performs relative to the alternatives described in Section 2.6. In each case, we generate data from the model described in Section 2. In our first synthetic data setting, we generate CRM datasets with anonymous visits where the missingness of the anonymous visits is ignorable, i.e. customer IDs for visits are missing-at-random. In this situation, we expect that the complete-case analysis and impute-then-estimate alternative methods will perform relatively well.

In our second synthetic data setting, we generate CRM datasets with anonymous visits in a non-ignorable way such that there is a high proportion of anonymous visits and also a strong correlation between a customer's propensity to be anonymous and their responsiveness to a marketing action. In this setting, we expect that our Bayesian imputation, which accounts for non-ignorable missing data mechanisms, will be better at recovering the underlying parameters of the CRM data compared to the alternatives.

In our third synthetic data setting, we generate CRM datasets with anonymous visits in the same non-ignorable fashion as our second setting, but we reduce the overall proportion of anonymous visits. With a reduced number of anonymous visits, we expect that the competitive advantage of our Bayesian imputation approach over the alternatives will be reduced.

The other settings (characteristics) for our synthetic data were chosen to emulate characteristics of our speciality retailer application in Section 4. We use a large number of activities ($M = 15$) in order to improve inference about which customers made the anonymous visit. Each generated dataset containes $N = 400$ customers with a range of anywhere from 8 to 22 visits per customer, as there is heterogeneity in the visitation rates between customers. Customers receive a single marketing action that varies in timing and frequency across customers. The marketing action has a large effect ω_{i1} on the arrival rate, with customers visiting roughly every second period when not affected by marketing and visiting roughly 4 times per period when affected by marketing.

3.1 Evaluation of Parameter Recovery

In Table 3, we present results for the recovery of the population-level average effect of marketing on the customers propensity to visit, a parameter one would expect to be biased in the data settings where missingness is non-ignorable. The true population value used to generate the data is reported in the last row of Table 3.

In the first setting, when the missingness mechanism is ignorable, each method is unbiased even with a large proportion of anonymous visits (45%). We note that the other model parameters were also recovered by all three methods in this setting. This result suggests that if the mechanism that leads to anonymous visits is unrelated to other aspects of consumer behavior, not surprisingly, then most reasonable methods will perform well.

In the second setting, we have the same large proportion of anonymous visits (45%) as in the first setting, but now there is a non-ignorable missingness mechanism where a high correlation (0.9) is induced between a customer's propensity to be anonymous and their propensity to visit in response to a marketing action. In other words, customers who have the largest increase in arrival rate in response to marketing are also much more likely to make an anonymous visit. We see in the middle column of Table 3 that both complete case analysis and impute-then-estimate have downward biased estimates of the population-level effect of the marketing action, along with posterior intervals that do not cover the true value. This downward bias might lead a company to reduce or eliminate marketing that is

Table 3: **Parameter Recovery in Synthetic Data where Missingness is Correlated with the Effect of a Marketing Action on the Arrival Rate of Customers**

Recovery of $\mathbf{\Gamma}_{\omega_1}$, the population-level propensity to visit in response to a marketing action in a setting where missingness (δ_i) is correlated with the individual-level propensity to visit in response to a marketing action (ω_{i1}). Gray indicates that the true parameter was covered by the estimated posterior interval for $\mathbf{\Gamma}_{\omega_1}$. The last row in each cell indicates the percent bias.

	Setting 1 Ignorable	Setting 2 Non-Ignorable	Setting 3 Non-Ignorable
Anonymous Visit Rate	45%	45%	30%
Correlation between missingness δ_i and marketing effect ω_{i1}	0.0	0.9	0.9
Estimates of Population Average Effect of Marketing on Visits ($\hat{\mathbf{\Gamma}}_{\omega_1}$)			
Bayesian Imputation	3.86	3.67	3.73
	(3.55,4.12)	(3.44,3.93)	(3.52,4.04)
	0%	4%	2%
Complete Case Analysis	3.96	3.51	3.84
	(3.69,4.23)	(3.19,3.81)	(3.52,4.12)
	3%	8%	0%
Impute-then-Estimate	3.75	3.56	3.63
	(3.53,3.98)	(3.30,3.82)	(3.42,3.84)
	2%	7%	5%
True value ($\mathbf{\Gamma}_{\omega_1}$)	3.83	3.83	3.83

actually effective, ultimately reducing profits.

In contrast, our Bayesian imputation method which accounts for potential correlations between missingness and marketing actions, shows little bias in this second setting along with a posterior interval that covers the true value.

The third setting has the same non-ignorable missingness mechanism but the overall proportion of anonymous visits is reduced (30%). In this setting with less overall missingness, each method is able to obtain coverage of the true parameter value. As we would expect, as missingness goes to zero, there is less need to properly account for it in making inference. However, accounting for the missingness doesn't hurt; across all three cases, we find that the Bayesian imputation reliably produces correct estimates.

3.2 Evaluation of Targeted Marketing

In Table 3, we focused on inference for a population-level parameter, but our Bayesian imputation approach can also lead to improved inference for individual-level parameters. Companies frequently use CRM databases to identify which customers would be most responsive to direct marketing, since sending marketing only to those most responsive customers improves marketing efficiency. In the context of our framework, companies would want to target customers with the highest expected response to the marketing action on arrival rates, i.e. customers with the largest values of $(\omega_{i0} + \omega_{i1})$.

In Table 4, we compare each method in terms of the ability to identify the top 100 customers (out of 400 total customers) with the highest true value of $(\omega_{i0} + \omega_{i1})$. Table 4 includes results from both setting 1 and setting 2 which have the same high proportion of anonymous visits (45%).

Our Bayesian imputation approach does an excellent job of identifying those customers who are most responsive to marketing, selecting 75 of the true top 100 customers in the

Table 4: **Ability to Identify Responsive Customers in Synthetic Data**

We rank order customers in terms of highest propensity to visit in response to receiving a marketing action ($\omega_{i0} + \omega_{i1}$) using estimates from each method and then select the top 100 customers. We report in this table the number of these top 100 selected customers that are consistent with the true top 100 customers.

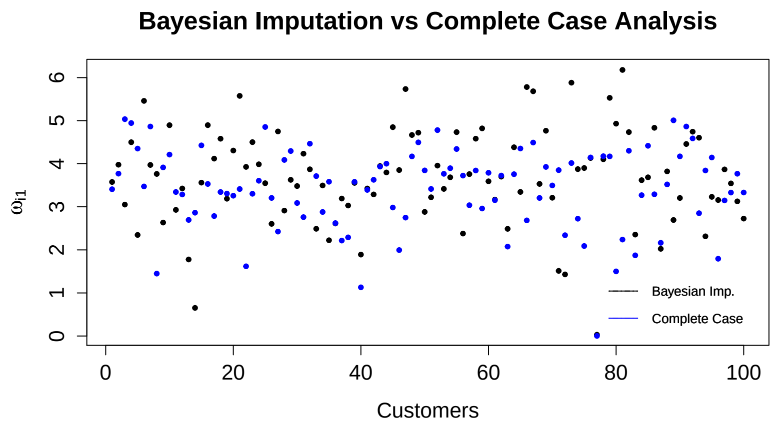
	Setting 1	Setting 2
Bayesian Imputation	72	75
Complete Case Analysis	40	46
Impute then Estimate	60	36

correlated setting 2 and 72 of the top 100 in the uncorrelated setting 1. In contrast, when the anonymous visits are removed from the CRM data (complete case analysis), the reduced information about each customer leads to identification of only 40 of the top 100 customers in setting 1 and 46 of the top 100 customers in setting 2. The impute-then-estimate method uses some information in the anonymous visits but does not do as good of a job as our Bayesian imputation approach, especially in setting 2 with the presence of correlation between responsiveness and the propensity to make an anonymous visits.

In order to understand why complete case analysis performs so poorly in Setting 1, we plot the individual-level responsiveness to marketing ($\omega_{i,1}$), in Figure 2. We see that in the complete case analysis, these estimates are being pulled closer to the population mean than in the Bayesian imputation method. This results in a "scrambling" in the rankings of the individual customers in the complete case analysis, hindering the method's ability to identify customers with the highest propensity to visit in response to marketing.

Figure 2: **Individual-level Estimates of $\omega_{i,1}$**

Estimates of the posterior means of $\omega_{i,1}$ for customers from our Bayesian imputation model and the complete case analysis model for synthetic data setting 1. The posterior means from our Bayesian imputation model are indicated by black points and from the complete case analysis model by blue points.



3.3 Estimating the Size of the Customer Base

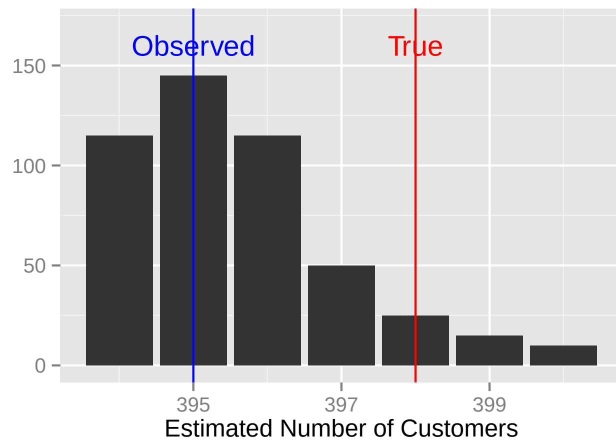
As discussed in Section 2.5, firms also want to know the number of unique customers in their CRM database, accounting for the fact that some anonymous visits may be new customers who only visit anonymously. The two alternative methods, complete case analysis and impute-then-estimate have no way of addressing this possibility, as they both assume that the total number of customers is the number of observed customers.

We evaluate the ability of our Bayesian imputation approach to recover the true number of customers in synthetic data setting 2, where there were 400 customers generated but only 398 of those customers made at least one visit, so the true customer base is 398 customers. Of those 398 customers, three customers were anonymous for all of their visits leading to 395 observed customers.

As described in Section 2.5, our Bayesian imputation model allows for anonymous visits to be assigned to new customers, which allows us to estimate a potentially larger customer base than the number of observed customers. The estimated posterior distribution of the number of customers from our Bayesian imputation model is shown in Figure 3. We see that the estimated posterior distribution covers the true size of the customer base.

Figure 3: **Estimated Size of the Customer Base**

Estimated posterior distribution from our Bayesian imputation model of the number of customers in synthetic data setting 2. The number of observed customers is indicated by a blue line and the true size of the customer base is indicated by the red line.



3.4 Summary of Synthetic Data Study

Our synthetic data settings have illustrated the importance of accounting for anonymous visits when analyzing CRM data. Complete case analysis is commonly used in practice but performs much worse at identifying the customers most responsive to marketing, which is one of the main motivations for keeping and analyzing CRM data. In addition, complete case analysis can not be used to estimate the number of additional anonymous customers when evaluating the total size of the customer base in CRM data.

This synthetic study, where we can manipulate the data generating process, also allows us to illustrate a point that is well-understood in the missing data literature, but perhaps less prevalent in marketing. When missingness is unrelated to other parameters of interest (missing-at-random), then complete case analysis results in unbiased but less efficient inference. However, when missingness is related to other parameters, then analyzing only the complete cases can produce biased estimates, as in setting 2 in Section 3.1.

The more sophisticated impute-then-estimate alternative method of imputing anonymous visits deterministically (and then proceeding with model inference) performs substantially worse than our Bayesian imputation method at identifying customers who are most responsive to marketing. In order to assess the performance of these methods in the case where the data does not perfectly conform to a known parametric model, we apply all three methods to a real CRM dataset next in Section 4.

4 Application to a Retail CRM Database

In this section, we examine a dataset describing customers' online and in-store transactions with an international speciality retailer (which prefers to remain anonymous) that is typical of most retail CRM data. The data consists of 24,000 identified customers' transactions over a two-year period. We will consider each transaction as a visit in our terminology from Section 2. Customers are identified by their payment method (i.e., credit or debit card number) and/or the name and address (provided for shipping of online orders).

For each transaction, we define the customer activities as binary indicators of which product categories that were purchased by the customer during that visit. For each transaction, a customer may purchase in any of 21 product categories (e.g. accessories, holiday, home furnishings, women's tops, etc.). On average, customers purchased from 2.15 categories per visit. This vector of activities gives us a way to identify different patterns of customer behavior during visits.

The data also includes a small number of customer characteristics including age, gender, whether the customer has created a wishlist with the retailer, and distance from nearest store to place of residence. The mean age of observed customers is 38 and the median age is 34 years old. 85% of the customers are women and 20% of customers have a wishlist. The company also regularly sends emails to their customers which promote new product lines and offer promotional discounts. In applying our model to this data, these emails are taken as the marketing action. Based on discussion with the retailer, we assume that the effect of an email lasts one week[2].

For the purposes of illustrating the imputation method, we focus our analysis on 100 customers in the dataset who a) made a purchase within the first three weeks of the observation period, which allows us to assume they were active from the beginning of the observation period, and b) have at least two transactions during a period when an email was in effect and at least two transactions when an email was not in effect so that the effect of marketing is well-identified for each customer.

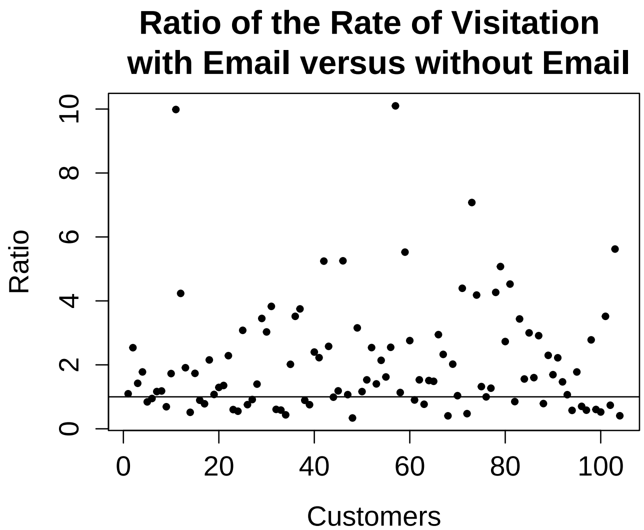
We present exploratory evidence of the effect of email on arrival rates in Figure 4, which plots for each customer the ratio of the arrival rate when emails are in effect (i.e., within a week after they were received) versus when emails are not in effect. Most customers are above the line of ratio equal to 1, indicating that they transact more often when they have received an email within the past week than when they have not. A binomial test confirms a positive effect of emails ($p < 0.001$), which we will model more explicitly using our hierarchical model of customer behavior.

We also explored potential effect of emails on the propensity to purchase in specific categories (the "activities" in this application) and found no significant effect. However, this is not surprising as emails from this retailer generally do not promote specific categories,

[2]Other research that also uses data from this retailer confirms that the effect of an email lasts approximately one week (Zantedeschi et al. 2015).

Figure 4: **Visitation Rates with Emails Versus without Emails**

For each customer, we plot the ratio of the arrival rate (number of visits divided by total time) to the arrival rate when email was not in effect. Horizontal line at ratio=1 indicates equality between the arrival rate when email is in effect versus when email is not in effect.



but rather focus on blanket offers, e.g., “free shipping”.

To evaluate each method against a ground truth, we estimate our hierarchical model on the full data to obtain an initial set of “true” parameter estimates in the absence of anonymous visits. With this complete dataset, the estimated population-level baseline rate of arrival (without having received a discount email) is 0.62 with a posterior interval of [0.56,0.68], which is quite similar to the synthetic data in Section 3. The estimated population average effect of emails on the propensity to visit is 2.16 with a posterior interval of [2.06,2.33] which means that, on average a customer visits approximately once every second week, whereas upon receiving an email, the customer will visit approximately twice in one week.

We create a test-bed for our Bayesian imputation model (vs. the alternative methods) by artificially creating anonymous visits in the data via the removal of customer IDs from some transactions. As in our synthetic data study (Section 3), we consider three different settings for creating anonymous visits. In setting 1, we anonymize visits completely at random whereas in setting 2, we induce a correlation between customers' propensity to be missing and the estimated effect of marketing on propensity to visit, i.e. $\mathrm{cor}(\delta_i, \omega_{i1}) = 0.9$. Setting 3 also induces a correlation, but has a lower overall rate of anonymous visits.

4.1 Evaluation of Parameter Recovery

In Table 5, we provide estimates and posterior intervals for the population average effect of emails on the propensity to visit ($\hat{\mathbf{\Gamma}}_{\omega_1}$) from each method in all three data settings. We compare these estimates to the "true" value of $\mathbf{\Gamma}_{\omega_1} = 2.16$ with a posterior interval of [2.06, 2.33] given in the last row of Table 5, which we estimated from the complete data without anonymous visits.

We see similar trends in Table 5 as in our synthetic data studies in Section 3. Our Bayesian imputation obtains coverage of the true complete data estimate in each setting, whereas complete case analysis underestimates the effect of the marketing action in both setting 1 and setting 2 where there is a high proportion of anonymous visits. At the lower missingness level (setting 3), complete case analysis is able to recover the complete data estimate. The impute-then-estimate method is unable to recover the true complete data estimate in setting 2 where there is a high proportion of anonymous visits and there is a correlation between the propensity to be anonymous and the propensity to visit in response to marketing.

In Table 6, we provide estimates and posterior intervals for the population average baseline visit rate ($\hat{\mathbf{\Gamma}}_{\omega_0}$) from each method in all three data settings. We compare these estimates to the true value of $\mathbf{\Gamma}_{\omega_0} = 0.62$ with a posterior interval of [0.56, 0.68] given in

Table 5: **Estimating Effect of Marketing Action in Retail CRM Data with Artificial Anonymous Visits**

Recovery of the population average propensity to visit in response to a marketing action ($\hat{\mathbf{\Gamma}}_{\omega_1}$) for all three methods in three data settings. Gray indicates that the true parameter was covered by the posterior interval. The last row in each cell indicates the percent difference from the "true" complete data estimate.

	Setting 1 Ignorable	Setting 2 Non-Ignorable	Setting 3 Non-Ignorable
Proportion of Anonymous Visits	45%	45%	30%
Correlation between missingness and marketing effect (δ_i, ω_{i1})	0	0.9	0.9
Estimates of Population Average Effect of Emails on Visits ($\hat{\mathbf{\Gamma}}_{\omega_1}$)			
Bayesian Imputation	2.05 (1.79,2.51) 5%	2.18 (1.99,2.37) 0%	2.19 (1.95,2.45) 1%
Complete Case Analysis	1.97 (1.87,2.06) 9%	1.83 (1.69,1.98) 15%	2.17 (1.83,2.47) 0%
Impute then Estimate	2.15 (1.95,2.35) 0%	1.97 (1.86,2.10) 9%	2.05 (1.90,2.19) 5%
Complete Data ("True")	2.16	2.16	2.16

the last row of Table 6, which we estimated from the complete data before anonymizing some of the visits.

Our Bayesian imputation obtains coverage of the true complete data estimate in each setting, whereas complete case analysis consistently underestimates the baseline arrival rate, since it discards all of the anonymous visits. The complete case analysis estimates only gets closer to the complete data truth in setting 3, where the proportion of anonymous visits is lower.

In contrast to complete case analysis, the impute-then-estimate method is better able to recover the baseline visit rate since it does not ignore the anonymous visits. However, referring back to Table 5, impute-then-estimate was not able to recover the effect of the marketing action since it matches those anonymous visits to the wrong customers and does not account for the uncertainty in those matches.

4.2 Evaluation of Targeted Marketing

We next evaluate the ability of each missing data method to identify customers who would be most responsive to direct marketing based on the estimated individual-level parameters of those customers. Similar to our synthetic data analysis in Section 3.2, we rank customers by their visit rate when exposed to email estimated using each of the three methods. We then compare the top customers identified by each method to the "true" top customers as ranked by the complete data analysis.

In Table 7, we compare each method in terms of the ability to identify the top 25 customers (out of 100 total customers) with the highest true value of $(\omega_{i0} + \omega_{i1})$ estimated from the complete data. Table 4 includes results from both setting 1 and setting 2 which have the same high proportion of anonymous visits (45%).

As in Section 3.2, our Bayesian imputation method performs substantially better than

Table 6: **Estimating Baseline Visit Rate in Retail CRM Data with Artificial Anonymous Visits**

Recovery of the population average propensity to visit in response to a marketing action ($\hat{\mathbf{\Gamma}}_{\omega_0}$) for all three methods in three data settings. Gray indicates that the true parameter was covered by the posterior interval. The last row in each cell indicates the percent difference from the "true" complete data estimate.

	Setting 1 ignorable	Setting 2 non ignorable	Setting 3 non ignorable
Proportion of Anonymous Visits	45%	45%	30%
Correlation between missingness and marketing effect (δ_i, ω_{i1})	0	0.9	0.9
Estimates of Population Average Baseline Visit Rate($\hat{\mathbf{\Gamma}}_{\omega_0}$)			
Bayesian Imputation	0.61	0.69	0.66
	(0.53,0.69)	(0.62,0.75)	(0.58,0.74)
	2%	11%	6%
Complete Case Analysis	0.53	0.51	0.57
	(0.46,0.60)	(0.44,0.57)	(0.51,0.64)
	15%	18%	8 %
Impute then Estimate	0.56	0.58	0.58
	(0.48,0.63)	(0.50,0.66)	(0.52,0.65)
	10%	6%	6%
Complete Data ("True")	0.62	0.62	0.62

Table 7: **Ability to Identify Responsive Customers in Retail CRM Data**

We rank order customers in terms of highest propensity to visit in response to receiving a marketing action ($\omega_{i0} + \omega_{i1}$) using estimates from each method and then select the top 25 customers. We report in this table the number of these top 25 selected customers that are consistent with the true top 25 customers (estimated from the complete data).

	Setting 1	Setting 2
Bayesian Imputation	15	18
Complete Case Analysis	12	2
Impute then Estimate	4	5

either alternative method at targeted marketing, which suggests that our model does a better job at recovering individual-level parameters. Complete case analysis performs reasonably well in setting 1 where visits were anonymized completely at random, but performs poorly in setting 2 where the propensity to be anonymous is correlated with marketing responsiveness across customers. Somewhat surprisingly, the impute-then-estimate method performs even worse than complete case analysis in setting 2.

The implication of this analysis is that targeting the top 25 customers based on the estimates using Bayesian imputation (from Setting 2) would result in 13 more visits per week (among 100 customers) than by using the complete case analysis estimates, which could lead to substantially more revenue. More specifically, when targeting the top 35 using the Bayesian imputation estimates from Setting 2, you would expect 121 visits that week from the 100 customers, 107 visits using the complete case analysis, and 110 visits using the impute-then-estimate method.

4.3 Summary of Application to Retail CRM Data

Our application to retail CRM data largely confirms the findings of our synthetic data analysis: our Bayesian imputation model is better suited to estimating the overall effects of marketing and identifying customers who are highly-responsive to marketing, particularly when the process that generates anonymous visits is correlated with other parameters of interest. Although not reported here, we also find that the Bayesian imputation performs well relative to case complete case analysis when customers propensity to be anonymous is correlated with other parameters such as the propensity to engage in a particular activity. This application to real data suggests that our Bayesian imputation method is robust to minor differences between the model and the true data generating process (as is the case with any real data.)

5 Conclusion

Nearly every company that tracks customer behavior over time faces the challenge of anonymous visits. Regardless of the consumer behavior being tracked or the technology used to track it, media providers, retailers and service providers who build CRM databases to record customer interactions all find that they have unidentified, anonymous visits in their database.

In this paper, we develop a Bayesian imputation method to link anonymous visits to either previously observed or new customers in a company's database. We have shown that by incorporating the additional information in the anonymous visits, companies can get more accurate estimates of critical parameters such as customer's responsiveness to marketing, which can lead to more accurate targeting and greater marketing efficiencies. Our imputation method is built around a generic hierarchical model of repeated customer behavior, including customers' times of arrival, the activities the customers engages in while visiting, the customer's demographic information (when observed), and information on direct marketing that the customer receives. Our estimation procedure probabilistically imputes which customers made each anonymous visit (based on the similarity of observed behavior) while simultaneously estimating all model parameters.

We compare our Bayesian imputation method to two common practices, complete case analysis and impute-then-estimate, and show that our approach performs substantially better in terms of recovering population-level parameters and recovering customer-level parameters that can be used for targeted marketing. In addition, unlike these alternative approaches, our model accounts for the fact that there may be some customers represented in the data who always remain anonymous.

Our studies highlight a fact that is well-understood in the missing data literature (beginning with Rubin 1978), but perhaps less-well understood in marketing: when the process that generates missingness is correlated with other model parameters, then the process that generates the missingness must be modeled or else parameter estimates may be biased. In our context, it is necessary to model each customers propensity to be anonymous and relate that missingness to customer preferences and demographics in our hierarchical model. Our model performs substantially better at recovering population-level parameter particularly when there are correlations between the propensity to be anonymous and other customer characteristics such as responsiveness to marketing. Since we can't know whether missingness is correlated with other important customer characteristics *a priori*, it makes sense to adopt the Bayesian imputation approach whenever there is a substantial rate of anonymous visits. Our synthetic data study also shows that complete case analysis performs reasonably well when the rate of anonymous visits is lower (even when the missingness process is non-ignorable), suggesting that firms with modest rates of anonymous visits can safely ignore them in their analyses, albeit they will pay a price in efficiency.

As with any method, there are some caveats for those who want to apply our Bayesian imputation approach to their own data. Fitting any model of repeat customer behavior is dependent on observing customers repeatedly. In our testing, we found that the model required at least 3-4 observed visits for a large proportion of the customer base (which for complete case analysis means 3-4 visits *after* the anonymous visits are removed). For customer behaviors that occur daily (such as media consumption or purchase of daily consumables), this is not a difficult requirement but may be more challenging for products with long purchase cycles such as electronics or cars.

Our imputation approach is dependent on an assumed parametric model, so it is important that the model fits well to real data. We have shown that the model fits reasonably well to a typical retail transaction data set, but the fit of the exponential inter-arrival times and the probit model for activities should be checked for any specific application and adapted if necessary.

We should also caution that there are additional hurdles to be overcome to scale this method to larger data sets. The computational needs of our MCMC estimation procedure increases as the data set increases in size, though this could be overcome with parallel computation or variational Bayes methods. More importantly, as the data set increases in size, the posterior distribution of potential choices for missing customer IDs becomes more diffuse and more difficult to estimate. However, this is a challenge facing all imputation methods and, in the extreme, no imputation method works well when a large number of customers make all their visits anonymously. If a large proportion of the total customer base is entirely unobserved, none of the methods will have enough signal to infer the individual-level behavior of the unobserved customers (who are a large proportion of the customer base), resulting in poor estimates for the parameters of interest to the firm.

Acknowledgement

This research was supported by a grant from the Baker Retail Center at the Wharton School. The authors would also like to thank the Wharton Customer Analytics Initiative for providing the retail data used in Section 4.

Appendices

Prior Distributions on Global Parameters

(1) The first prior to the population-level regression coefficients, Γ, is

$$\boldsymbol{\Gamma}_h | \boldsymbol{\Gamma_0}, \boldsymbol{\gamma}_0 \sim MVN(\boldsymbol{\gamma}_0, \boldsymbol{\Gamma_0}) \tag{5.0.12}$$

where $h = 1, \ldots, S$ indexes a row of $\boldsymbol{\Gamma}$ and where $\boldsymbol{\gamma}_0$ and $\boldsymbol{\Gamma}_0$ are fixed hyperparameters.

(2) The prior to the population-level variance-covariance matrix that characterizes heterogeneity across the customers, $\boldsymbol{\Omega}$, is

$$\boldsymbol{\Omega} \sim InvWish_{\eta_0}(\boldsymbol{\Lambda}_0) \tag{5.0.13}$$

where η_0, and $\boldsymbol{\Lambda}_0$ are fixed hyperparameters.

(3) The prior on the global correlations amongst the activities within visits, $\boldsymbol{\Sigma}$, is

$$\boldsymbol{\Sigma} \sim InvWish_{\eta_0}(\boldsymbol{T}_0) \tag{5.0.14}$$

for fixed hyperparameters η_0 and $\boldsymbol{T_0}$.

Gibbs Sampler Steps 2 through 7

(2) Sample $\boldsymbol{y}^\star$ for all activities M and all rows n from a truncated multivariate normal distribution,

$$y^{\star}_{jm}|\boldsymbol{\theta_i}, \boldsymbol{\Sigma}, \boldsymbol{U}, \boldsymbol{y^{\star}_{j(-m)}} \sim e^{(-\frac{1}{2}(\mu_i^{\star})'(\Sigma^{\star})^{-1}\mu_i^{\star})}$$
$$\times \{I(y^{\star}_{jm} > 0)I(y_{jm} = 1) + I(y^{\star}_{jm} < 0)I(y_{jm} = 0)\} \qquad (5.0.15)$$

where

$$\mu_i^{\star} = (\boldsymbol{\nu_{U_j}} + \boldsymbol{\beta^T_{U_j} X_j})^{\star} = (\mu_i)_m + \boldsymbol{\Sigma_{12}\Sigma_{22}}^{-1}(\boldsymbol{y^{\star}}_{j(-m)} - \boldsymbol{\mu_i}_{(-m)}) \qquad (5.0.16)$$

and

$$\Sigma^{\star} = \Sigma_{11} - \boldsymbol{\Sigma_{12}\Sigma_{22}}^{-1}\boldsymbol{\Sigma_{21}} \qquad (5.0.17)$$

We use the "star" notation to mean the Schur compliment. For example, $\Gamma^{\star}$ and $\Omega^{\star}$ for the mth page would mean

$\Gamma^{\star} = (\Gamma)_m + \boldsymbol{\Omega_{12}\Omega_{22}}^{-1}(\boldsymbol{\theta_{i(-m)}} - \boldsymbol{(\Gamma)_{(-m)}})$

$\Omega^{\star} = \Omega_{11} - \boldsymbol{\Omega_{12}\Omega_{22}}^{-1}\boldsymbol{\Omega_{21}}$,

and

$\boldsymbol{\Gamma} = \begin{pmatrix} (\Gamma)_m \\ \boldsymbol{(\Gamma)_{(-m)}} \end{pmatrix}$, $(\Gamma)_m$ is 1×1, $\boldsymbol{(\Gamma_{(-m)})}$ is $(M+1) \times 1$

and $\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \boldsymbol{\Omega_{12}} \\ \boldsymbol{\Omega_{21}} & \boldsymbol{\Omega_{22}} \end{pmatrix}$ with size $\begin{pmatrix} 1 \times 1 & 1 \times (M+1) \\ (M+1) \times 1 & (M+1) \times (M+1) \end{pmatrix}$

where we denote Ω_{11} as the variance for the m^{th} entry.

(3) Sample user specific parameters, $\boldsymbol{\theta}_i$. This consists of three parts. First we sample the $\boldsymbol{\beta}_i$'s and ν_i's,

$$\boldsymbol{\beta_i}, \nu_i | \boldsymbol{Z_{U_j=i}}, \boldsymbol{\Gamma^\star}, \boldsymbol{\Omega^\star}, \boldsymbol{\Sigma} \sim MVN(\hat{\boldsymbol{\beta}}_\star, \boldsymbol{V_{\beta\star}}) \tag{5.0.18}$$

where $(\boldsymbol{y}^\star)_\star = \begin{pmatrix} \boldsymbol{y^\star}_{U_j} \\ [\Gamma Z_i]^\star \end{pmatrix}$ $\boldsymbol{X}_\star = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{I}_p \end{pmatrix}$, and $\Sigma_\star = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \Omega^\star \end{pmatrix}$

and $\hat{\boldsymbol{\beta}}_\star = (X_\star^T \Sigma_\star^{-1} X_\star)^{-1} X_\star^T \Sigma_\star^{-1} (y^\star)_\star$,

$\boldsymbol{V_{\beta\star}} = (X_\star^T \Sigma_\star^{-1} X_\star)^{-1}$.

We use the notation $[\Gamma Z_i]^\star$ and $\Omega^\star$ as we did above.

Next, for $(\theta_{i,(M+M\times P+1)}, (\theta_{i,(M+M\times P+2)}) = (\omega_{i,0}, \omega_{i,1})$, we do two Metropolis steps since we have non-standard distributions.

First for $\omega_{i,0}$, we have a proposal,

$$\omega'_{i0} \sim N(\omega_{i0}, \zeta^2) \tag{5.0.19}$$

where ζ^2 is a tuning parameter and do a Metropolis steps with

$$P(\lambda_i | \boldsymbol{\lambda_{-i}}, \boldsymbol{y}, \boldsymbol{U}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}) \propto \prod_{j=1}^{n} L_{\lambda_{i,t_j}} \times e^{-\frac{1}{2}(\theta_{(M+M\times P+1)i} - ((\Gamma Z_i)^\star)'(\Omega^\star)^{-1}(\theta_{(M+M\times P+1)i} - (\Gamma Z_i)^\star)} \tag{5.0.20}$$

where $\log(\lambda_{i,t_j}) = \omega_{i,0} + \omega_{i,1} H_{i,t_j}$, and $L_{\lambda_{i,t_j}}$ is the product of the parts of the likelihood that correspond to user i.

Next, for $\theta_{i,(M+M\times P+2)} = \omega_{i,1}$, we use the same density function for the Metropolis step (as for $\omega_{i,0}$), except we now hold $\theta_{i(M+M\times P+1)}$ fixed. We use the same tuning parameter, ζ^2, and draw a proposal

$$w'_{i,1} \sim N(w_{i,1}, \zeta^2) \tag{5.0.21}$$

For $\theta_{i(M+M\times P+3)} = \text{logit } \delta_i$, we also do a Metropolis step. We use η^2 for the tuning parameter, and draw

$$\delta_i' \sim N(\delta_i, \eta^2) \tag{5.0.22}$$

and do a Metropolis step with

$$\begin{aligned} P(\delta_i|\boldsymbol{\delta_{-i}}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}) \propto \prod_{j=1}^{n} [\delta_i^{(V_j=1)}(1-\delta_i)^{(V_j=0)} \\ \times e^{-\frac{1}{2}(\theta_{(M+M\times P+3)i}-((\Gamma Z_i)^\star)'(\Omega^\star)^{-1}(\theta_{(M+M\times P+3)i}-(\Gamma Z_i)^\star)} \end{aligned} \tag{5.0.23}$$

where $V_j = 0$ if user i is known at visit j, and $V_j = 1$ is user i is anonymous.

(4) Sample $\boldsymbol{\Gamma}$,

$$\boldsymbol{\Gamma}|\boldsymbol{\Omega}, \boldsymbol{U}, \boldsymbol{\theta} \sim \text{MVN}(\hat{\boldsymbol{\Gamma}}_\star, \boldsymbol{V}_{\boldsymbol{\Gamma}\star}) \tag{5.0.24}$$

where $\theta_\star = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_I \\ \Gamma_0^1 \\ \vdots \\ \Gamma_0^S \end{pmatrix}$ $X_\star = \begin{pmatrix} Z_1 & 0 & 0 \\ 0 & Z_1 & 0 \\ 0 & 0 & Z_1 \\ \vdots & & \\ Z_I & 0 & 0 \\ 0 & Z_I & 0 \\ 0 & 0 & Z_I \\ & I_{(M+M\times P+3)*S} & \end{pmatrix}$, and $\Omega_\star = \begin{pmatrix} \Omega & 0 \\ 0 & \Gamma_0 \end{pmatrix}$

and $\hat{\boldsymbol{\Gamma}}_\star = (X_\star^T \Omega_\star^{-1} X_\star)^{-1} X_\star^T \Omega_\star^{-1} \theta_\star$,

$\boldsymbol{V}_{\boldsymbol{\Gamma}\star} = (X_\star^T \Omega_\star^{-1} X_\star)^{-1}$.

(5) Sample $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma}|\boldsymbol{U},\boldsymbol{\Gamma},\boldsymbol{\Omega} \sim InvWish(\eta_0 + n, \boldsymbol{S}) \tag{5.0.25}$$

where $\boldsymbol{S} = \boldsymbol{T_0} + \sum_{j=1}^{n}(\boldsymbol{y^{\star}}_j - \boldsymbol{\mu_{U_j}})(\boldsymbol{y^{\star}}_j - \boldsymbol{\mu_{U_j}})^T$.

(6) Sample $\boldsymbol{\Omega}$,

$$\boldsymbol{\Omega}|\nu_0, \boldsymbol{\Lambda_0}, \kappa_0, \boldsymbol{\Gamma}, \boldsymbol{\theta} \sim InvWish(\nu_0 + I, \boldsymbol{\Lambda}_n) \tag{5.0.26}$$

where $\boldsymbol{\Lambda}_n = \boldsymbol{\Lambda_0} + \sum_{i=1}^{I}(\boldsymbol{\theta}_i - \boldsymbol{\Gamma}\boldsymbol{Z}_i)(\boldsymbol{\theta}_i - \boldsymbol{\Gamma}\boldsymbol{Z}_i)^T$

(7) Sample $\boldsymbol{Z}_i$,

$$\boldsymbol{Z}_i|\boldsymbol{\Omega},\boldsymbol{U},\boldsymbol{\theta}_i \sim MVN(\hat{\boldsymbol{Z}}_{\star}, \boldsymbol{V_{Z_{\star}}}) \tag{5.0.27}$$

where $\hat{\boldsymbol{Z}}_{\star} = (\boldsymbol{\Gamma}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma} + \boldsymbol{P_0})^{-1}(\boldsymbol{\Gamma}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\theta}_i + \boldsymbol{P_0}\boldsymbol{\xi_0})$

and $\boldsymbol{V_{Z_{\star}}} = (\boldsymbol{\Gamma}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma} + \boldsymbol{P_0})^{-1}$

and where $\boldsymbol{P_0}$ and $\boldsymbol{\xi}_0$ are the prior parameters.

References

Adigüzel, Feray, Michel Wedel. 2008. Split questionnaire design for massive surveys. *Journal of Marketing Research* **45**(5) 608–617.

Chib, S., E. Greenberg. 1998. Analysis of multivariate probit models. *Biometrica* **85**(85) 347–361.

Feit, E. M., P. Wang, E. T. Bradlow, P. S. Fader. 2013. Fusing aggregate and disaggregate data with an application to multiplatform media consumption. *Journal of Marketing Research* **50**(3) 348–364.

Gelman, A., J. B. Carlin, Hal S. Stern, D. B. Rubin. 2003. *Bayesian Data Analysis*. CRC Press, Boca Raton, FL.

Geman, S., D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6** 721–741.

Goerg, Georg M., Yuxue Jin, Nicolas Remy, Jim Koehler. 2015. How many millenials visit youtube? estimating unobserved events from incomplete panel data conditioned on demographic covariates. Tech. rep., Google, Inc.

Gordon, K. T. 2010. The newest insights into loyalty programs reveal the best ways to engage customers. URL `http://www.entrepreneur.com/marketing/marketingcolumnistkimtgordon/article205508.html`. [Online; posted 12-March-2010].

Grabowski, Mariusz. 2005. Handling missing values in marketing research using som. Daniel Baier, Klaus-Dieter Wernecke, eds., *Innovations in Classification, Data Science, and Information Systems*. Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin Heidelberg, 322–329.

Graham, J. W., S. M. Hofer, A. M. Piccinin. 1994. Analysis with missing data in drug prevention research. L. Collins, L. Seitz, eds., *National Institute on Drug Abuse Research Monograph Series*, vol. 142. National Institute on Drug Abuse, Washington, D.C., 201–213.

Grant, K. B. 2008. Retailers expand customer-loyalty programs. URL http://www.smartmoney.com/spend/family-money/retailers-expand-customer-loyalty-programs-22801. [Online; posted 31-March-2008].

Grover, R., M. Vriens. 2006. *The Handbook of Marketing Research: Uses, Misuses, and Future Advancements*. Sage Publications, Thousand Oaks, CA.

Hastie, T., R. Tibshirani, J. Friedman. 2009. *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY.

Hastings, W.K. 1970. Monte carlo samping methods using markov chains and their applications. *Biometrika* **57** 97–109.

Little, R. J. A., D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics, Hoboken, NJ.

Manchanda, P., A. Ansari, S. Gupta. 1999. 'the shopping basket': A model for multicategory purchase incidence decisions. *Marketing Science* **18**(2) 95–114.

Mulhern, F. 2009. Integrated marketing communications: From media channels to digital connectivity. *Journal of Marketing Communications* **15**(2) 85–101.

Musalem, A., E. T. Bradlow, J. S. Raju. 2008. Who's got the coupon: Estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research* **45**(2) 715–730.

Nunes, J. C., X. Dreze. 2006. Your loyalty program is betraying you. *Harvard Business Review* **84**(4) 124–131.

Rossi, P. E., G. M. Allenby, R. McCulloch. 2005. *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistic, Hoboken, NJ.

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

Rubin, D. B. 1976. Inference and missing data. *Biometrika* **63**(3) 581–592.

Rubin, D. B. 1978. *Multiple Imputations in Sample Surveys- A Phenomenological Bayesian Approach to NonResponse*. Proceedings of the International Statistical Institute, Manila.

Rubin, D. B., N. Schenker. 1986. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* **81**(1) 366–374.

Sahni, Navdeep, Dan Zou, Pradeep K Chintagunta. 2014. Effects of targeted promotions: Evidence from field experiments. *Available at SSRN 2530290* .

Tanner, M. A., W. H. Wong. 1987. The calculation of posterior distributions by data augementation. *Journal of the American Statistical Association* **82**(398) 528–540.

Tode, C. 2007. Nordstrom upgrades loyalty program experience. URL http://www.dmnews.com/nordstrom-upgrades-loyalty-program-experience/article/95464/. [Online; posted 04-May-2007].

Winer, R. S. 2009. New communications approaches in marketing: Issues and research directions. *Journal of Interactive Marketing* **23**(2) 108–117.

Ying, Yuanping, Fred Feinberg, Michel Wedel. 2006. Leveraging missing ratings to improve online recommendation systems. *Journal of marketing research* **43**(3) 355–365.

Zantedeschi, D., E. M. Feit, E. T. Bradlow. 2015. Measuring multi-channel advertising effectiveness using consumer-level advertising response data.