DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

# Bayesian Inference for Assessing Effects of Email Marketing Campaigns

**The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters**

| | |
|---|---|
| Citation | Wu, Jiexing, Kate J. Li, and Jun S. Liu. 2016. "Bayesian Inference for Assessing Effects of Email Marketing Campaigns." Journal of Business & Economic Statistics (January 20): 1–40. doi:10.1080/07350015.2016.1141096. |
| Published Version | doi:10.1080/07350015.2016.1141096 |
| Citable link | http://nrs.harvard.edu/urn-3:HUL.InstRepos:28552972 |
| Terms of Use | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP |

ACCEPTED MANUSCRIPT

# Bayesian Inference for Assessing Effects of Email Marketing Campaigns

Jiexing Wu
Department of Statistics, Harvard University

Kate J. Li
Department of Information Systems and Operations Management
Sawyer Business School, Suffolk University

Jun S. Liu
Department of Statistics, Harvard University

Downloaded by [Jun Liu] at 05:43 20 January 2016

**Abstract**

Email marketing has been an increasingly important tool for today's businesses. In this paper, we propose a counting-process-based Bayesian method for quantifying the effectiveness of email marketing campaigns in conjunction with customer characteristics. Our model explicitly addresses the seasonality of data, accounts for the impact of customer characteristics on their purchasing behavior, and evaluates effects of email offers as well as their interactions with customer characteristics. Using the proposed method, together with a propensity-score-based unit-matching technique for alleviating potential confounding, we analyze a large email marketing data set of an online ticket marketplace to evaluate the short- and long-term effectiveness of their email campaigns. It is shown that email offers can increase customer purchase rate both immediately and during a longer term. Customers' characteristics such as length of shopping history, purchase recency, average ticket price, average ticket count, and number of genres purchased also affect customers' purchase rate. A strong positive interaction is uncovered between email offer and purchase recency, suggesting that customers who have been inactive recently are more likely to take advantage of promotional offers.

*Keywords:* Hazard function, Markov chain Monte Carlo, Normal approximation, Propensity score matching, Purchase rate, Survival process

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

# 1 Introduction

Email marketing is to directly market a commercial message to people using email. It is significantly cheaper and faster than traditional marketing vehicles and is widely used today. U.S. firms spent $400 million on email marketing in 2006, and $1.5 billion in 2012, compared to $64 billion on TV, $34 billion on print ads, and $39 billion on Internet advertising. The estimated return on investment (ROI) is 4325% (VanBoskirk 2007). Due to the increasing popularity of email marketing, various surveys have been conducted to understand consumers' response to it. In a survey conducted by Direct Marketing Association, 66% of consumers have made online purchases as a result of an email marketing message. According to a ChoozOn Corporation survey, 70% of consumers made use of a discount coupon from a marketing email in the week prior to the survey. It is well-established that email marketing is useful at the overall level. However, evaluating the effectiveness of an individual company's email marketing campaigns is crucial and still challenging.

The effectiveness of promotions has been evaluated at both aggregate and individual levels. At the aggregate level, sales response models are widely used (e.g., Kamakura and Kang (2007) and Osinga et al. (2010)). Our data set contains the information of individual customers of an online ticket marketplace; therefore, we focus on individual-level analysis and use purchase rate as the response variable. In the literature, three basic types of models have been proposed to model purchase rate or purchase time. The first type uses a probability distribution to directly model purchase time. Allenby et al. (1999) developed a dynamic model, assuming that purchase time follows a generalized gamma distribution whose parameters are specified to allow for both cross-sectional heterogeneity and temporal dynamics. Boatwright et al. (2003) proposed a hierarchical Bayes approach that assumes the Conway-Maxwell-Poisson distribution for purchase time and models the distribution of purchase quantity conditional on purchase time. Such models are not suitable for our data because there are no known dynamic distributional models that can account for seasonality and time-dependent individual-specific covariates simultaneously.

The second type of models employs logit or probit models and their extensions to treat the buy/not buy decision during a time interval (e.g., a week). Bucklin and Lattin (1991) developed and tested a two-state, probabilistic model of purchase incidence and brand choice for frequently-purchased consumer products. Chintagunta (1993) developed a joint model of purchase proba-

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

bility, brand choice and purchase quantity to assess the impact of marketing variables, including feature advertisement, special display and temporary price cut. Similar models were applied to study various marketing issues of frequently-purchased products (e.g., Ailawadi and Neslin (1998), Zhang and Krishnamurthi (2004), Chan et al. (2008)). This type of models is appropriate for situations where customers make regular and frequent purchases. However, in our context event ticket purchases are much less frequent and therefore the number of periods with no purchases is considerably higher.

The third type of models utilizes survival processes and hazard functions to take advantage of their capability of handling right-censored data, which is prevalent in duration times data, and time-dependent covariates. Gupta (1988) studied the impact of marketing variables on consumer decisions about when, what, and how much to buy, using an Erlang-2 purchase time model, a multinomial logit model of brand choice, and a cumulative logit model of purchase quantity, respectively. Following Gupta (1988), many studies used the proportional hazard model (PHM) to characterize household purchase time. In these studies the construct of interest is a household's instantaneous probability of making a purchase in a product category, conditional on the elapsed time since the household's previous purchase in that category. This conditional probability is the hazard function. Jain and Vilcassim (1991) was the first to formally decompose the hazard function into the baseline hazard, which captures a household's intrinsic temporal purchase pattern, a covariate function, which represents the influence of marketing variables, and effects of unobserved heterogeneity. Empirical studies showed that purchase time could not be adequately described by probability distributions such as exponential, Erlang-2, or Weibull; and the capture of unobserved heterogeneity was essential. Studies built upon Jain and Vilcassim (1991) include Gonul and Srinivasan (1993), Chintagunta and Haldar (1998), Seetharaman and Chintagunta (2003), and Manchanda et al. (2006), among many others.

The goal of our study is to evaluate the effectiveness of email marketing and to understand factors that impact a customer's purchase rate using transaction and email marketing data from an online ticket marketplace. The survival process framework serves our purpose well for modeling customer purchase rate. However, existing models such as those based on PHMs and those summarized by Bijwaard et al. (2006) cannot be applied directly to adequately reflect important

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

characteristics of our data. In order to accommodate our data, we employ the following strategies. First, because event ticket purchase exhibits strong seasonality, we model the baseline hazard as a periodic function. Thus, the partial likelihood approach employed by Bijwaard et al. (2006) is not applicable. We express this periodic function by a Fourier series with up to $M$ terms, where $M$ can be determined by the Bayes factor computation or Bayesian information criterion (BIC). Second, instead of employing the EM algorithm described in Bijwaard et al. (2006), we conduct a full Bayesian analysis of the data. By integrating over all individual-specific random effects, we greatly reduce the dimension of the posterior distribution (from tens of thousands to fewer than 20), which results in an efficient Markov chain Monte Carlo (MCMC) algorithm. We have also developed a robust Normal approximation method, which is used to estimate the Bayes factor effectively. Third, we employ a propensity-score-based unit-matching method to eliminate potential biases and confounding factors in the analysis.

Using the aforementioned approach, we evaluate short- and long-term effectiveness of the company's email marketing campaigns. We also examine how customer characteristics affect their purchase rate and investigate whether certain customer segments are more responsive to email promotions than others. We find that among the five major event genres (e.g., concerts, Major League Baseball (MLB), National Basketball Association (NBA), National Football Association (NFL), and National Hockey League (NHL)), NHL purchases are the most responsive to email offers both immediately and during a longer term. Email offers also show a certain degree of effectiveness on MLB and NFL purchases. As regards customer characteristics, we discover that customer purchase rate decreases as the length of shopping history or purchase recency increases, indicating that customer retention is a serious challenge to the company. Average ticket price reflects whether a customer is a value or a budget customer; while the average ticket count per order suggests the size of group with whom a customer attends events. These two attributes affect purchase rate of different genres differently. Another customer characteristic that we consider is the number of genres that a customer has purchased from the company. Consistent with the cross-selling literature, we find that the more genres bought, the higher the purchase rate. Last but not least, we find that offer and purchase recency has positive interaction for NBA and NHL purchases, which suggests that offers are more effective for encouraging NBA and NHL customers who have been inactive

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

for a longer time to purchase again.

The rest of the paper is organized as follows. Section 2 provides a detailed description of the data set and the propensity-score-based unit-matching mechanism. Section 3 formally introduces the model and its Bayesian inference procedure. Section 4 reports estimation and prediction results, and discusses the managerial implications. Section 5 concludes with some comments. Technical details are presented in the Appendix.

## 2 The Email Marketing Campaign Data

The data set was provided by a large online ticket marketplace ("the company") that offers tickets of all major sports and live entertainment events. During each month from February 2007 to February 2011, a random sample of 2,000 customers were selected from those who made their first purchase in that month. Starting from July 2009, the company successively conducted promotional offer campaigns, in which coupon codes were sent to customers via email. Each coupon entitled customers to a certain discount towards their purchases made before the expiration date of the coupon, which normally was two to four weeks after the coupon was issued. The discounts were in the form of percentage-off (ranging from 10-15%) or free-shipping. Due to the homogeneous nature of the coupons, we do not distinguish among them; instead, we focus on analyzing the average effect of email offers.

The data set contains the entire transaction and email offer history of the randomly-selected customers. Each transaction record includes customer ID, purchase time, event, number of tickets, and ticket price. Additional information about the customers includes their zip codes, genre preferences, and email preference. A customer has the option to opt out of email communication. We include only opted-in customers accounting for 45.7% of the customers in our data set. Table 1 provides key summary statistics for the five major genres with a total of 79,757 transactions. The data also exhibits cross-genre purchasing behavior of the customers. Numbers of customers who have purchased two, three, four, and five different genres are 5,319, 996, 209, and 28, respectively, representing 14.7% of all customers.

When launching an offer campaign, the company intended to target at a specific group of cus-

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

tomers and randomly assigned customers from the group to treatment or control group. However, the data reveals that the randomized design was not properly implemented: customers in the treatment and control groups differ significantly and the sizes of treatment and control groups of each campaign are severely imbalanced (with control groups having fewer than 200 customers while treatment groups having 2∼3,000 customers). As a remedy, we define a treatment group to include all customers who received offers from the company during the time period under consideration, whereas all other opted-in customers belong to the corresponding control group. In order to ensure the robustness of statistical inference to model misspecification, we match customers in the two groups to improve covariate balance so that the resulting data set resembles one generated from a properly randomized experiment (Imbens and Rubin, 2015).

Customer behavior, including customers' transaction records, responses to marketing activities, and Internet browsing records are commonly used in quantitative models for direct marketing (Venkatesan et al., 2007). It has also been shown that customers' cross-genre purchasing behavior is an important aspect of customer relationship management (Ngobo, 2001; Kumar et al., 2008). Hence, given literature support and data availability, we use four main aspects of customer behavior as covariates in our analysis: (1) length of shopping history, which measures the length of time since a customer's first purchase; (2) average price of tickets purchased, which suggests whether a customer is a value- or budget-customer; (3) average number of tickets per order, which indicates the size of group with whom a customer usually attends events; and (4) number of genres purchased, which captures a customer's cross-genre purchasing behavior. We employ a propensity score matching approach by modeling treatment status as a logistic function of the four covariates (Rosenbaum and Rubin, 1985). The unit-matching approach consists of the following steps.

- Step 1: Estimate the propensity score of each customer, find the overlapping range $R$ of scores of treatment and control groups, and remove customers whose propensity scores lie out of the range. Let $N_t$ and $N_c$ denote the numbers of remaining customers in treatment and control groups, respectively.

- Step 2: Stratify the range $R$ into 5 strata and determine the maximum number of customers from each group within each stratum under the constraint that the size ratio is $N_t : N_C$. Randomly select the determined number of customers from each group when needed. Combine

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

all the selected data.

As shown in Table 2, covariate imbalance has been significantly reduced for most of the covariates after the matching. For the rest of the analysis, we use data of the matched groups and consider that the two groups differ only in whether they have received an email offer from the company. We also conducted the same analysis using all data without matching and the results were consistent with that from the matched data. Note that the full-data analysis is only valid if our specified analysis model is completely correct.

# 3 The Model For Purchase Events

## 3.1 Model formulation

Let $t_{ij}$, $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, N_i$ represent the time of the $i$th customer's $j$th purchase. We use a survival process to model each customer's purchase history, in which a "death" event corresponds to a purchase. Conditioning on the previous transaction time, the occurrence of the next purchase is governed by a hazard rate function $\lambda_i(t)$, so that $\lambda_i(t)\Delta t$ is the probability of having the next purchase during the infinitesimal time period $(t, t + \Delta t]$ conditional on having had no purchases till time $t$. We emphasize that $\lambda_i(t)$ is allowed to depend on all available information up to time $t$, including the transaction time before. For simplicity, we omit all the conditioning information in the notation of $\lambda_i(t)$. Table 3 lists the main notations. Throughout the paper, we use "hazard rate" and "purchase rate" interchangeably. Since each customer enters the company's database after his/her first purchase, we model subsequent transactions conditioning on the first one by concatenating a series of survival processes. The occurrence of one purchase symbolizes the termination of the current process and initiation of a new one. All customers' purchase processes are censored at time $T$. Note that, if $\int \lambda_i(t)\mathrm{d}t$ is bounded, the probability that one will never make another purchase is nonzero, and can be estimated from the data.

Conditioning on the time of the first purchase $t_{i1}$ and the censoring time $T$, the probability distribution function for intermediate purchases $\{t_{ij}\}_{j=2}^{N_i}$ can be written as (Fleming and Harrington

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

(2011)):

$$f_i(\{t_{ij}\}_{j=2}^{N_i}|t_{i1}, T) = \exp\left(-\int_{t_{i1}}^{T} \lambda_i(t)\mathrm{d}t\right)\prod_{j=2}^{N_i} \lambda_i(t_{ij}). \tag{1}$$

When $N_i = 1$, Equation (1) is reduced to only the exponential part. For simplicity, we suppress the explicit dependence of $\lambda_i(t)$ on the purchase history before $t$.

As discussed in Section 1, survival process has been used in the marketing literature to model transaction data. These models usually allow the rate function to depend on covariates, but do not model seasonality explicitly. Our hazard function consists of three multiplicative components:

$$\lambda_i(t) = \lambda_{10}(t)\lambda_{2i}(t)\lambda_{3i}(t), \tag{2}$$

where $\lambda_{10}(t)$ is a periodic function that depicts seasonality and serves as a baseline purchase rate, $\lambda_{2i}(t)$ is an individual-specific factor that uses covariates to explain purchase rate, and $\lambda_{3i}(t)$ is the random effects term that captures unobserved heterogeneity.

## 3.2 Three components of the purchase rate function

**Seasonal effect** $\lambda_{10}(t)$. Figure 1 plots the empirical purchase rates for the five genres that demonstrate the need of a periodic function to model seasonality. These empirical purchase rates were estimated as follows. Each year was partitioned into 52 equal-length time window. For each window $w$ of year $n$, we let $I_{nw}$ denote the number of customers whose first purchase occurred before the time window, and record $y_{nw}$, the number of transactions made by those $I_{nw}$ customers during this time window. Then the customers' annual purchase rate for that time interval is estimated as $52(y_{nw}/I_{nw})$. Its confidence interval can be obtained by assuming that $y_{nw}$ follows a Poisson distribution with the estimated purchase rate. Although less apparent for concerts, the sales of sports tickets display strong periodic patterns that match well with their corresponding sports seasons.

To reflect the strong seasonal effect, we let $\lambda_{10}(t)$ be a nonnegative periodic function that serves as the baseline purchase rate for all customers. Since any periodic function can be approximated well by its Fourier expansion up to $M$'th term, we express $\lambda_{10}(t)$ as:

$$\lambda_{10}(t) = \max\left\{0, \alpha_0 + \sum_{m=1}^{M} [\alpha_{2m-1}\sin(m\omega t) + \alpha_{2m}\cos(m\omega t)]\right\}, \tag{3}$$

ACCEPTED MANUSCRIPT



Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

where $\omega$ is $2\pi$ since we use one year as the time unit for the function, and $M$ typically ranges from 1 to 15, whose exact value will be determined by our full Bayesian model detailed in Section 4.2. Although modeling $\log \lambda_{10}(t)$ directly can ensure nonnegative rate, we choose the current form to allow $\lambda_{10}(t)$ to be zero and for easier theoretical derivations.

**Covariate effect $\lambda_{2i}(t)$.** Taking into account the four covariates used for propensity score matching and information about email offers, we postulate the following model to capture the impact of customer characteristics and email offer on purchase rate:

$$\begin{aligned}\log \lambda_{2i}(t) = {} & \beta_1 o_i(t) + \beta_2 eo_i(t) + \beta_3(t - t_{i1}) + \beta_4(t - lt_i(t) - \tau) + \beta_5 p_i \\ & + \beta_6 tct_i + \beta_7 g_i + \beta_8 o_i(t)(t - t_{i1}) + \beta_9 o_i(t)(t - lt_i(t) - \tau) \\ & + \beta_{10} o_i(t) p_i + \beta_{11} o_i(t) tct_i + \beta_{12} o_i(t) g_i, \end{aligned} \tag{4}$$

where $o_i(t)$ indicates whether customer $i$ has a valid offer at time $t$, which reflects the immediate offer effect, and $eo_i(t)$ indicates whether customer $i$ has an offer that expired less than two months ago at time $t$ , which reflects the longer term effect of an offer. To avoid double-counting of offer exposure, we set $eo_i(t) = 0$ when $o_i(t) = 1$.

$\beta_3$ captures the effect of shopping history. $lt_i(t)$ is the time of customer $i$'s most recent purchase and $\tau$ is the average interpurchase time of all customers up to time $t$, which will be estimated by the sample mean *a priori*. There are two reasons for subtracting $\tau$ from $t$ to centralize the term representing recency. First, it reduces the collinearity between main effect terms and their interaction terms to almost zero. Second, after subtracting $\tau$, $\beta_1$ reflects the "average" offer effect, whereas $\beta_9$ represents additional offer effect when a customer's purchase recency is distinct from the population average. Nonlinear relationship between recency and purchase rate has been examined in the literature. For example, Gonul and Shi (1998) and Gonul and Hofstede (2006) include a quadratic and a logarithmic term of recency, respectively. Our preliminary analysis showed that such nonlinear terms were insignificant and added computational burden substantially. Therefore, only linear effect of recency is considered in the model. $\beta_5$, $\beta_6$, and $\beta_7$ represent the effects of average ticket prices, average ticket count per order, and cross-genre purchase on purchase rate, respectively. $\beta_8$ to $\beta_{12}$ are the coefficients of the interaction terms that indicate the types of customers who are more responsive to promotional offers.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

**Random effects $\lambda_{3i}(t)$.** Figure 2 shows distributions of inter-purchase time for the five genres, all exhibiting a wide range of values. A random effects term, $\lambda_{3i}(t)$, is necessary for capturing unobserved heterogeneity among customers, which may include age, gender, income level, occupation, marriage status, etc. We assume that $\lambda_{3i}(t)$ is person-specific and time-invariant and will be estimated from the data: $\lambda_{3i}(t) = \delta_i$.

# 4 Posterior Inference With the Model

## 4.1 The posterior distribution and marginalization

Functions $o_i(t)$ and $eo_i(t)$ in $\lambda_{2i}(t)$ are step functions of $t$. Their switching points, together with purchase time $\{t_{ij}\}_{j=1}^{N_i}$, partition the entire shopping history of customer $i$ into a set of disjoint intervals, denoted by $L_{i1}, L_{i2}, \ldots, L_{iS_i}$, where $S_i$ is the total number of such intervals of customer $i$ and $\cup_{s=1}^{S_i} L_{is} = [t_{i1}, T)$. Within each interval $L_{is}$, time of the last purchase, offer status, and post-offer status are all constants. Let $bt_{is}$, $et_{is}$, $lt_{is}$, $o_{is}$, and $eo_{is}$ denote the beginning time (left boundary), ending time (right boundary), last purchase time, offer status, and post-offer status associated with interval $L_{is}$, respectively.

Within time interval $L_{is} = [bt_{is}, et_{is})$, we replace time of the last purchase, offer status, and post-offer status in Equation (4) by their corresponding interval-specific constants. Equation (4) can be rewritten as

$$\begin{aligned}\log \lambda_{2i}(t) = {} & \beta_1 o_{is} + \beta_2 eo_{is} + \beta_3(lt_{is} - t_{i1} + \tau) + \beta_5 p_i + \beta_6 tct_i + \beta_7 g_i + \beta_8 o_{is}(lt_{is} - t_{i1} + \tau) \\ & + \beta_{10} o_{is} p_i + \beta_{11} o_{is} tct_i + \beta_{12} o_{is} g_i + (\beta_3 + \beta_4)(t - lt_{is} - \tau) \\ & + (\beta_8 + \beta_9) o_{is}(t - lt_{is} - \tau). \end{aligned} \tag{5}$$

More succinctly, we have

$$\log \lambda_{2i}(t) = \boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \tau)$$

for customer $i$ at $t \in L_{is}$, and the customer's overall purchase rate for $t \in L_{is}$ is:

$$\lambda_{is}(t) \equiv \lambda_i(t)|_{\{t \in L_{is}\}} = \max(0, \boldsymbol{X}(t)'\boldsymbol{\alpha}) \exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \tau))\delta_i, \tag{6}$$

ACCEPTED MANUSCRIPT



Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{2M})'$, $\boldsymbol{\beta}_1 = (\beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \beta_8, \beta_{10}, \beta_{11}, \beta_{12})'$, $\boldsymbol{\beta}_2 = (\beta_3 + \beta_4, \beta_8 + \beta_9)'$, and

$$\begin{aligned}
\mathbf{X}(t) &= (1, \sin(\omega t), \cos(\omega t), \ldots, \sin(M\omega t), \cos(M\omega t))' \,, \\
\mathbf{Y}_{is} &= (o_{is}, eo_{is}, lt_{is} - t_{i1} + \tau, p_i, tct_i, g_i, o_{is}(lt_{is} - t_{i1} + \tau), o_{is}p_i, o_{is}tct_i, o_{is}g_i)' \,, \\
\mathbf{Z}_{is} &= (1, o_{is})'.
\end{aligned}$$

Using $s_{ij}$ to index the unique interval among $\{L_{is}\}_{s=1}^{S_i}$ that covers transaction time $t_{ij}$, i.e., $t_{ij} \in L_{is_{ij}}$, we can write the likelihood function for customer $i$ as:

$$f_i(\{t_{ij}\}_{j=2}^{N_i} | t_{i1}, T) = \exp\left(-\sum_{s=1}^{S_i} \int_{L_{is}} \lambda_{is}(t)\mathrm{d}t\right) \prod_{j=2}^{N_i} \lambda_{is_{ij}}(t_{ij}). \tag{7}$$

In light of the left-skewedness of the log-inter-purchase time distributions shown in Figure 2, we assume that random effects $\delta_i$'s follow a Gamma distribution and are independent of each other *a priori*:

$$\delta_i \overset{iid}{\sim} \text{Gamma}(\gamma, \gamma), \;\; i = 1, \ldots, I. \tag{8}$$

This distribution has mean 1 and variance $1/\gamma$, implying that the random effect does not impact the expected purchase rate but contributes to the population variance. Another advantage of the Gamma prior is that it enables us to integrate out $\delta_i$ analytically.We assume that $\pi(\gamma) \propto (\gamma + v)^{-2}\mathbb{1}_{\gamma>0}$.

We further assume a Gaussian prior for other parameters: $(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T \sim N(\mathbf{0}, u^2\mathbf{I})$. Hyper-parameters $u$=100 and $v$=1 are so chosen that the prior distributions are sufficiently diffuse, making the inference sufficiently dependent on data. Combining the likelihood function and the priors, the

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

posterior distribution of all parameters of interest is:

$$
\begin{aligned}
&\log f(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma, \{\delta_i\}_{i=1}^I \mid \{t_{ij}\}_{i,j}, T) = \\
&- \boldsymbol{\alpha}' \left( \sum_{i=1}^I \delta_i \left[ \sum_{s=1}^{S_i} \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \boldsymbol{A}_{is}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right) \right] \right) \\
&+ \sum_{i=1}^I \sum_{j=2}^{N_i} \log(\boldsymbol{X}(t_{ij})' \boldsymbol{\alpha}) + \left( \sum_{i=1}^I \sum_{j=2}^{N_i} \boldsymbol{Y}_{is_{ij}}' \right) \boldsymbol{\beta}_1 + \left( \sum_{i=1}^I \sum_{j=2}^{N_i} \boldsymbol{Z}_{is_{ij}}' (t_{ij} - lt_{is_{ij}}) \right) \boldsymbol{\beta}_2 \\
&+ \sum_{i=1}^I (N_i - 1) \log(\delta_i) - \frac{1}{2u^2} (\boldsymbol{\alpha}' \boldsymbol{\alpha} + \boldsymbol{\beta}_1' \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2' \boldsymbol{\beta}_2) + (\gamma - 1) \sum_{i=1}^I \log(\delta_i) \\
&- \gamma \sum\nolimits_{i=1}^I \delta_i + I\gamma \log(\gamma) - I \log \Gamma(\gamma) - 2 \log(\gamma + v) + \text{Const}
\end{aligned} \tag{9}
$$

Integrating out all $\{\delta_i\}_{i=1}^I$ yields the posterior distribution of main parameters of interest that has much lower dimensions:

$$
\begin{aligned}
&\log f(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma \mid \{t_{ij}\}_{i,j}, T) = \\
&- \sum_{i=1}^I (N_i + \gamma - 1) \log \left( \gamma + \boldsymbol{\alpha}' \left[ \sum_{s=1}^{S_i} \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \boldsymbol{A}_{is}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right) \right] \right) \\
&+ \sum_{i=1}^I \sum_{j=2}^{N_i} \log(\boldsymbol{X}(t_{ij})' \boldsymbol{\alpha}) + \left( \sum_{i=1}^I \sum_{j=2}^{N_i} \boldsymbol{Y}_{is_{ij}}' \right) \boldsymbol{\beta}_1 + \left( \sum_{i=1}^I \sum_{j=2}^{N_i} \boldsymbol{Z}_{is_{ij}}' (t_{ij} - lt_{is_{ij}}) \right) \boldsymbol{\beta}_2 \\
&- \frac{1}{2u^2} (\boldsymbol{\alpha}' \boldsymbol{\alpha} + \boldsymbol{\beta}_1' \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2' \boldsymbol{\beta}_2) \\
&+ \sum\nolimits_{i=1}^I \log \Gamma(N_i + \gamma - 1) + I\gamma \log(\gamma) - I \log \Gamma(\gamma) - 2 \log(\gamma + v) + \text{Const}
\end{aligned} \tag{10}
$$

## 4.2 Bayesian computation via Markov chain Monte Carlo

To draw inferences, we implement a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution $f(\boldsymbol{\Theta})$. In order to improve computational efficiency, we first use a modified Newton-Raphson algorithm to find the posterior mode $\hat{\boldsymbol{\Theta}}$ and the corresponding inverse Hessian matrix $\mathbf{H}(\hat{\boldsymbol{\Theta}})^{-1}$ (see Appendix 7.3 for details). We then initialize the sampler with $\boldsymbol{\Theta}^{(0)} = \hat{\boldsymbol{\Theta}}$ and implement a random-walk Metropolis (RWM) algorithm (Liu, 2008): At step $t+1$, generate

$$
\boldsymbol{\Theta}_{prop}^{(t+1)} = \boldsymbol{\Theta}^{(t)} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N\left(0, -\sigma^2 \mathbf{H}(\hat{\boldsymbol{\Theta}})^{-1}\right), \tag{11}
$$

ACCEPTED MANUSCRIPT



Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

then let $\Theta^{(t+1)} = \Theta^{(t+1)}_{prop}$ with probability $p_a \equiv \min\left\{1, f(\Theta^{(t+1)}_{prop})/f(\Theta^{(t)})\right\}$ and $\Theta^{(t+1)} = \Theta^{(t)}$ with probability $1 - p_a$.

Importance sampling and independent Metropolis-Hastings (IMH) algorithm can take advantage of Normal approximation to the posterior distribution as well (Liu, 1996, 2008). However, even when normal approximation is reasonably close to the true posterior, the RWM proposal described by (11) can still lead to very sticky algorithm in high-dimensions, and the corresponding importance sampling and IMH algorithms can only be worse – leading to unstable approximations (Liu, 1996). A good alternative is to combine Gibbs sampling with Metropolis type moves, especially given that we have observed from the inverse-Hessian matrix that no pair of parameters are too highly correlated. More precisely, since $\Theta$ can be represented as $\{\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma\}$, we cycle through each of the four elements with conditional updates, i.e., updating $\Theta_d$ by draws from $f(\Theta_d \mid \Theta^{(t)}_{[-d]})$ for $d = 1, 2, 3, 4$. Although at a lower dimension, each conditional distribution still evades exact sampling. We thus make $l = 10$ steps of RWM moves within each conditional move, with the proposal covariance matrix being $\Sigma_d = 2.38^2(-\mathbf{H}_{dd})^{-1}/dim(\Theta_d)$, where $\mathbf{H}_{dd}$ stands for the sub-matrix of $\mathbf{H}$ corresponding to $\Theta_d$. The step size was recommended by Gelman et al. (1996) for Gaussian-like target distributions to achieve maximum efficiency. As shown in Section 5, for our data set the posterior samples obtained by the MCMC strategy coincide very well with the asymptotic normal distribution derived in the Supplementary Materials.

Downloaded by [Jun Liu] at 05:43 20 January 2016

## 4.3 Bayesian model selection

As discussed in Section 3.2, we model the baseline hazard as a periodic function with $M$ Fourier expansion terms. To find a proper $M$ for each genre, we employ a Bayesian model selection strategy in which we give each model a prior, compute model likelihood (i.e., $P(\text{data} \mid M)$, where $M$ represents a model), and choose the model with the highest posterior probability. The posterior probability of a model $M$ can be written as

$$p(M|\{t_{ij}\}_{i,j}) \propto p_0(M) \int f(\{t_{ij}\}_{i,j}|\Theta_M, M)\pi(\Theta_M)\mathrm{d}\Theta_M, \tag{12}$$

where $p_0(M)$ is a prior belief of $M$ that is assumed to be uniform.

The model likelihood, also known as the normalizing constant for the posterior distribution, is

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

usually not in analytically tractable form and difficult to estimate. For our data, because the normal approximation to the posterior distribution is reasonably accurate, we use the following importance sampling procedure to approximate the integral in (12):

- Step 1: Obtain $K$ independent samples from $N(\hat{\boldsymbol{\Theta}}, -\mathbf{H}(\hat{\boldsymbol{\Theta}})^{-1})$, denoted by $\{\boldsymbol{\Theta}^{(k)}\}_{k=1}^{K}$.
- Step 2: Compute $w^{(k)} = f(\{t_{ij}\}_{i,j}|\boldsymbol{\Theta}^{(k)}, M)\pi(\boldsymbol{\Theta}^{(k)})/\phi(\boldsymbol{\Theta}^{(k)}; \hat{\boldsymbol{\Theta}}, -\mathbf{H}(\hat{\boldsymbol{\Theta}})^{-1})$ for $k = 1, \ldots, K$, where $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; and use the average of $\{w^{(k)}\}_{k=1}^{K}$ as an estimate for the integral in (12).

A computationally cheaper strategy for model selection is to choose the model that minimizes the Bayesian Information Criterion (BIC), which can be viewed as a monotone function of the approximated Bayes factor (Schwarz, 1978). In our model, BIC takes the form of:

$$BIC = -2\left[\log f(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\gamma} \mid \{t_{ij}\}_{i,j}, T) + \frac{1}{2u^2}(\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}} + \hat{\boldsymbol{\beta}}_1'\hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\beta}}_2'\hat{\boldsymbol{\beta}}_2) + 2\log(\hat{\gamma} + v)\right]$$
$$+ [2M + 2 + dim(\hat{\boldsymbol{\beta}}_1) + dim(\hat{\boldsymbol{\beta}}_2)] \log\left(\sum_{i=1}^{I} N_i - I\right). \tag{13}$$

Figure 3 tracks the posterior probability of each model and its corresponding BIC as the model size grows. The two model selection criteria clearly agree on the same optimal model for all genres except NFL, whose optimal and sub-optimal models are however quite close in terms of either criterion. For concerts, the optimal $M$ is found at 1, which is consistent with the fact that the periodic pattern of concert ticket sales is the weakest among the five genres. MLB requires a much larger $M$ ($M = 10$), while the other three genres all favor a moderate $M$ of around 5.

## 4.4 Prediction of future events

Predicting customers' future purchase behavior is of practical interest of the company. Specifically, given customer $l$'s purchase history up to time $T$, we want to forecast the number of orders that she will place in the next $\Delta T$ years. It can be answered via Monte Carlo simulation of her purchases during $(T, T+\Delta T]$ given that her offer status during $(T, T+\Delta T]$ is known. First we sample $\{\boldsymbol{\Theta}^{(k)}\}_{k=1}^{K}$ as described in Section 4.3 using the MCMC algorithm. For each $k$, we draw $\delta_l^{(k)}$ from Gamma$(N_l + \gamma^{(k)} - 1, (\boldsymbol{\alpha}^{(k)})' \sum_{s=1}^{S_l} \boldsymbol{B}_{is}^{(k)} + \gamma^{(k)})$ and sample new purchases iteratively. Given the customer's latest

ACCEPTED MANUSCRIPT



Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

transaction time $t_{lj}$ (either recorded or sampled) for the period $(t_{lj}, \min\{T + \Delta T, t_{l,j+1}\}]$, we can construct intervals $\{L_{ls}^{(j)}\}_{s=1}^{S_l^{(j)}}$ and $\{\tilde{L}^{(j)}(\boldsymbol{\alpha}^{(k)})\}_{ls}\}_{s=1}^{S_l^{(j)}}$, and vectors $\boldsymbol{Y}_{ls}^{(j)}$ and $\boldsymbol{Z}_{ls}^{(j)}$ in the same way as described before. The next purchase time $t_{l,j+1}$ is governed by the hazard rate:

$$\lambda_{ls}^{(kj)}(t) = \max(0, \boldsymbol{X}(t)'\boldsymbol{\alpha}^{(k)}) \exp((\boldsymbol{Y}_{ls}^{(j)})'\boldsymbol{\beta}_1^{(k)} + (\boldsymbol{Z}_{ls}^{(j)})'\boldsymbol{\beta}_2^{(k)}(t - t_{lN_l} - \tau))\delta_l^{(k)}.$$

The process of $t_{l,j+1}$ can be simulated accordingly. An algorithm to sample from a survival process with an inhomogeneous hazard rate is provided by Cinlar (2013). Every time after a new purchase time is sampled, the rate function is updated by treating the current purchase as the last purchase input for the new process. We repeat the procedure until the newest sampled purchase time surpasses $T + \Delta T$. Then we estimate the expected sales by taking the average of the resulting number of purchases for each simulated process. The prediction result is discussed in Section 5.3.

# 5 Results and Interpretations

## 5.1 Offer effects

Employing the model discussed in Section 3, we assess the effects of email offer, customer characteristics, and their interactions on customers' purchase rate of event tickets. Table 4 summarizes posterior means and standard deviations for coefficients of interest for the five genres. Figure 4 shows histograms of posterior samples obtained via the MCMC algorithm described in Section 4.2 overlaid with their normal approximations.

NHL goers appear to be the most responsive to promotional offers. On average, an offer increases the purchase rate of NHL tickets by 34.0% and 15.3% in the short and long terms, respectively. The purchase rate of NFL tickets increases by 24.2% when an offer is valid; while the purchase rate of MLB tickets increases by 17.2% during the two months after a customer receives an offer. Although the promotional offers do not show significant immediate effect on genres other than NFL and NHL, the data reveal that the offer campaigns launched by the company did successfully promote sales to some extent. It is worth noting that the offers exhibit long-term effect as well, i.e., the customer purchase rate is elevated even after an offer expires. This may be due to the situation where a customer does not have an immediate need to buy tickets, but receiving an

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

offer reminds the customer about the company and then the customer recalls and purchases from it when a need arises. The company can benefit from such behaviors because it increases purchase rate without compromising the profit margin.

We also investigate interactions between the promotional offer and customer characteristics. Except the one between offer and recency, we do not discover other significant interactions. A positive interaction exists between offer and recency for NBA and NHL. NHL has shown positive immediate and long-term effects of promotional offers. This positive interaction further reveals that customers who have been inactive for a while are more inclined to take advantage of an offer. When recency is large, offers serve as an effective catalyst for attracting customers back to make a purchase.

Downloaded by [Jun Liu] at 05:43 20 January 2016

## 5.2 Customer characteristics effects

Customer characteristics have some interesting impacts on purchase rate. Length of shopping history, defined as the length of time from a customer's first purchase until the time point under consideration, has a negative impact on the purchase rates of all genres except concerts. For instance, the purchase rate of an MLB ticket buyer decreases by $1 - \exp(-0.041) = 4.0\%$ each year from the time she joins the company when the other covariates stay unchanged. The percentage decreases are 10.9%, 9.2% and 6.3% for NBA, NFL and NHL, respectively. The result suggests that customers lose interest in purchasing from the company over time. Recency, defined as time elapsed since the last purchase, has been studied extensively in the literature and is shown to be a good predictor of customers' purchase behavior (Elsner et al. (2004)). For regularly-consumed products, such as clothes, food, and office supplies, empirical data showed that customers with smaller recency are more likely to respond to marketing activities, i.e., recency has a negative impact on response rate (e.g., Bitran and Mondschein (1996), Gonul and Shi (1998)). This phenomenon can be explained by the dynamics that when considering frequently-purchased products, the fact that a customer does not buy in a period time usually means that she is buying elsewhere. The longer the time a customer has not purchased from a company, the lower the probability that she will come back to the company. On the other hand, positive effect of recency is observed as well. For example, the purchase probability of durable goods may increase with recency because consumers

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

need to replace their product when approaching the end of the product's life cycle (e.g., Roberts and Berger (1989) and Ansari et al. (2008)). Our result shows that recency negatively affects the purchase rates of all genres. It suggests that in terms of recency's impact, event ticket purchase resembles the behavior of frequently-purchased products, although its purchase frequency is much lower. In essence, the likelihood of a customer making another purchase decreases as time passes. This, along with the effect of the length of shopping history, makes customer retention a challenging task for the company, which in fact validates the company's concern over it. In order to keep customers interested and active, the company needs to engage customers as early and frequent as possible by using a variety of marketing activities.

Average ticket price significantly affects the purchase rate of all genres except NBA. However, the signs are inconsistent: it is positive for concerts and NFL, while negative for MLB and NHL. The opposite signs may be explained by the price level and fan base of different genres. Figure 5 provides box plots showing the differences in average ticket prices of the five genres. Also as shown in Table 1, the average ticket prices of concerts and NFL are $146.80 and $150.62, respectively, while those of MLB and NHL are $70.07 and $89.17, respectively. The result indicates that at a higher price level, customers who buy more expensive tickets tend to buy more often; while at a lower price level, it is the opposite: customers who buy cheaper tickets tend to buy more often.

The effect of average ticket count per order on purchase rate is significant for all genres, except concerts. The signs are also mixed: NBA, NFL, and NHL have positive coefficients, while MLB has a negative one. The former three genres have smaller average ticket count per order (2.92, 2.93, and 2.79, respectively) than MLB (3.22), which is statistically significant (see Figure 5). In essence, when ticket count per order is small, purchase rate increases with ticket count; the effect reverses when ticket count is large. Notably low ticket count per order may imply that a person attends an event alone to try it out. An increasing ticket count suggests that a stable group of people attend events together, which tend to make it a more regular activity. When ticket count is considerably high, it may indicate a one-time gathering with a large group which does not happen often. This implies that there is a certain range of average ticket count that leads to the highest purchase rate.

Summarizing the effects of average ticket price and average ticket count per order by genre

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

provides useful insights on customer segmentation. For concerts and NFL, characterized by high ticket price and low ticket count per order, customers who highly value good seats (reflected by more expensive tickets) and attend with a relatively large group (reflected by higher ticket count per order) are the ones who have high purchase rate, which may be the most valuable customers to the company. These customers are devoted fans to the genre who are likely to enjoy it with their family and/or close friends. For MLB, characterized by low ticket price and high ticket count per order, its customers exhibit opposite behavior: the ones who buy cheaper tickets and attend with a relatively small group tend to purchase more frequently. This seems to suggest that people who enjoy MLB games as a small-scale social event (evidenced by attending with an extended, but not too large, group of family and friends) and do not care too much about having good seats, tend to go more often. Lastly, for NBA and NHL, characterized by low ticket price and low ticket count per order, customers who buy less expensive tickets and attend with a larger group are likely to purchase frequently.

Number of genres that a customer has purchased from the company represents the degree of cross-buying of the customer, which has been associated with higher levels of customer retention, revenue generation, and loyalty (Reinartz et al. (2008)). Our result clearly shows that as the purchased number of genres increases, purchase rate increases for all five genres. It is consistent with the findings in the literature and accentuates the importance of promoting cross-selling to the company. Although the observed association may not imply causation, it is worthwhile for the company to send targeted informational or promotional emails introducing multiple genres to a new customer to raise her awareness and stimulate cross-buying behavior.

Downloaded by [Jun Liu] at 05:43 20 January 2016

## 5.3 Prediction results

To validate model fitting and illustrate the prediction power of our proposed model, we choose an intermediate time point $t_0$ and use transaction records prior to $t_0$ to predict transactions occurred after it. We set $t_0$ to be January 1, 2010. Figure 6 illustrates the comparison between our model prediction and the actual sales for the five genres. The apparent shape difference between training and test periods is because only existing customers' purchases are predicted during the test period.

We evaluate the prediction accuracy by root mean squared errors (RMSE) of the total predicted

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

sales on the logarithm scale during the test period, which is calculated as:

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\widehat{\log pc}^{(h)} - \log pc)^2}$$

where $\widehat{\log pc}^{(h)} = \left(\sum_{k=1}^{K} \widehat{\log pc}^{(hk)}\right)/n$ and $pc$ is the actual purchase count during the test period. That is, we replicate $H$ independent predictions of log-purchase count and each prediction $\widehat{\log pc}^{(h)}$ is estimated by the average of $K$ simulated purchase counts as explained in Section 4.4 (aggregated over all customers and on the logarithmic scale). RMSEs with $K = 1000$ for concerts, MLB, NBA, NFL, and NHL are 0.134, 0.071, 0.104, 0.048 and 0.057, respectively, which indicates satisfactory prediction capability.

Downloaded by [Jun Liu] at 05:43 20 January 2016

# 6 Concluding Remarks

We have employed the survival analysis framework developed in biostatistics and adopted recently by marketing researchers to model recurring purchase events and to examine how email marketing campaigns and personal characteristics affect customers' purchasing behaviors. Our model is a generalized proportional hazard model that enables us to accomplish two goals. First is to connect an individual customer's purchase likelihood to a variety of factors, including customer characteristics and marketing variables. Second is to predict future purchases made by existing customers based on their historical transactions and email offer information.

In our Bayesian model, a customer's purchase rate is characterized by a hazard rate function consisting of three components: a baseline function, a function of covariates, and a random effect term capturing unobserved heterogeneity. Different from existing models in the literature, for the baseline function, we postulate a periodic function to model the seasonality of event ticket purchases. Our approach is fully Bayesian, making the inference less dependent on asymptotic approximations. We are able to analytically integrate over the hazard rate function and the random effects so as to derive a workable posterior distribution with much lower dimensions, for which both a Newton-Raphson mode-finding algorithm and an efficient MCMC algorithm are implemented. Another attractive aspect of our Bayesian approach is that it can be conveniently used to provide purchase predictions via Monte Carlo simulations.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Our results reveal that email offers have different degrees of effectiveness for different event genres. Therefore, a one-size-fits-all email marketing strategy is unlikely to be effective. Instead, the use of genre-specific offers should be encouraged and explored. The effects of email offers, purchase recency, and their positive interaction suggest that there is an optimal timing for using offers to revive inactive customers. Examining the impact of average ticket price and average ticket count per order on purchase rate allows customer segmentation based on the two factors. Email offers can then be used to shape customer behavior and habit into the types that are the most valuable to the company.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

# 7 Appendix

## 7.1 Deriving the log-likelihood function

Define $\boldsymbol{D}(\boldsymbol{\alpha}) = \{t \in (-\infty, \infty) : \boldsymbol{X}(t)'\boldsymbol{\alpha} <= 0\}$. The region $\boldsymbol{D}(\boldsymbol{\alpha}) \cap [t_{i1}, T)$ can be decomposed into $R_i$ disjoint sets $\{\tilde{L}(\boldsymbol{\alpha})_{is}\}_{s=1}^{S_i}$ where $\tilde{L}(\boldsymbol{\alpha})_{is} = \boldsymbol{D}(\boldsymbol{\alpha}) \cap L_{is}$. The boundary points of $\tilde{L}(\boldsymbol{\alpha})_{is}$ are either ones of $L_{is}$ or roots of the equation $\boldsymbol{X}(t)'\boldsymbol{\alpha} = 0$ within $[t_{i1}, T)$. Apparently values of the offer status, post-offer status and last purchase time within $\tilde{L}(\boldsymbol{\alpha})_{is}$ remain the same of those of $L_{is}$.

It is easy to see that $\tilde{L}(\boldsymbol{\alpha})_{is}$ is a set of disconnected intervals (also can be empty set). Let $\tilde{L}(\boldsymbol{\alpha})_{is} = \cup_{r=1}^{R_{is}}[\tilde{bt}(\boldsymbol{\alpha})_{isr}, \tilde{et}(\boldsymbol{\alpha})_{isr}]$.

Expand the exponent part of equation (7).

$$
\begin{aligned}
\sum_{s=1}^{S_i}\int_{L_{is}} \lambda_{is}(t)dt &= \sum_{s=1}^{S_i}\int_{L_{is}} \max(0, \boldsymbol{X}(t)'\boldsymbol{\alpha})\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\delta_i \mathrm{d}t \\
&= \sum_{s=1}^{S_i}\int_{L_{is}} \boldsymbol{X}(t)'\boldsymbol{\alpha}\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\delta_i \mathrm{d}t \\
&\quad - \sum_{s=1}^{S_i}\int_{\tilde{L}(\alpha)_{is}} \boldsymbol{X}(t)'\boldsymbol{\alpha}\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\delta_i \mathrm{d}t \\
&= \delta_i\Big[\sum_{s=1}^{S_i}\int_{bt_{is}}^{et_{is}} \boldsymbol{X}(t)'\boldsymbol{\alpha}\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\mathrm{d}t \\
&\quad - \sum_{s=1}^{S_i}\sum_{r=1}^{R_{is}}\int_{\tilde{bt}(\alpha)_{isr}}^{\tilde{et}(\alpha)_{isr}} \boldsymbol{X}(t)'\boldsymbol{\alpha}\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1 + \boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\mathrm{d}t\Big]
\end{aligned}
$$

The integration of $\lambda_{is}(t)\mathrm{d}t$ over interval $[a, b]$ is

$$
\begin{aligned}
\int_a^b \lambda_{is}(t)\mathrm{d}t &= \int_a^b \delta_i \exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1)\boldsymbol{\alpha}'\boldsymbol{X}(t)\exp(\boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\mathrm{d}t \\
&= \delta_i \exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1)\boldsymbol{\alpha}'\int_a^b \boldsymbol{X}(t)\exp(\boldsymbol{Z}_{is}'\boldsymbol{\beta}_2(t - lt_{is} - \overline{\Delta t}))\mathrm{d}t \\
&= \delta_i \exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1)\boldsymbol{\alpha}'\int_{a-lt_{is}-\overline{\Delta t}}^{b-lt_{is}-\overline{\Delta t}} \boldsymbol{X}(t + lt_{is} + \overline{\Delta t})\exp(\boldsymbol{Z}_{is}'\boldsymbol{\beta}_2 t)\mathrm{d}t
\end{aligned}
$$

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Define

$$
\begin{aligned}
A_a^b(\omega,\phi,\theta) &= \int_a^b \sin(\omega t+\phi)\exp(\theta t)dt \\
&= \begin{cases} (b-a)\sin\phi & \text{if } \omega=\theta=0 \\ \frac{1}{\theta^2+\omega^2}(\theta\sin(\omega t+\phi)-\omega\cos(\omega t+\phi))\exp(\theta t)\Big|_{t=a}^{b} & \text{otherwise} \end{cases}
\end{aligned}
$$

And define

$$
\begin{aligned}
\boldsymbol{A}_{is}(\theta) = \Big(& A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(0,\pi/2,\theta), \\
& A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(\omega,\omega\cdot(lt_{is}+\overline{\Delta t}),\theta), A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(\omega,\omega\cdot(lt_{is}+\overline{\Delta t})+\pi/2,\theta), \\
& A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(2\omega,2\omega\cdot(lt_{is}+\overline{\Delta t}),\theta), A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(2\omega,2\omega\cdot(lt_{is}+\overline{\Delta t})+\pi/2,\theta), \\
& \cdots, \\
& A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(M\omega,M\omega\cdot(lt_{is}+\overline{\Delta t}),\theta), A_{bt_{is}-lt_{is}-\overline{\Delta t}}^{et_{is}-lt_{is}-\overline{\Delta t}}(M\omega,M\omega\cdot(lt_{is}+\overline{\Delta t})+\pi/2,\theta)\Big)'
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha},\theta) = \Big(& A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(0,\pi/2,\theta), \\
& A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(\omega,\omega(lt_{is}+\overline{\Delta t}),\theta), A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(\omega,\omega(lt_{is}+\overline{\Delta t})+\pi/2,\theta), \\
& A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(2\omega,2\omega(lt_{is}+\overline{\Delta t}),\theta), A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(2\omega,2\omega(lt_{is}+\overline{\Delta t})+\pi/2,\theta), \\
& \cdots, \\
& A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(M\omega,M\omega(lt_{is}+\overline{\Delta t}),\theta), A_{\tilde{bt}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}^{\tilde{et}(\alpha)_{isr}-lt_{is}-\overline{\Delta t}}(M\omega,M\omega(lt_{is}+\overline{\Delta t})+\pi/2,\theta)\Big)'
\end{aligned}
$$

We obtain a simplified form of $\log f_i$:

$$
\begin{aligned}
& \log f_i(\{t_{ij}\}_{j=2}^{N_i}|\boldsymbol{\alpha},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\gamma,\{\delta_i\}_{i=1}^{I},t_{i1},T) = \\
& -\delta_i\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\left[\exp(\boldsymbol{Y}_{is}'\boldsymbol{\beta}_1)\left(\boldsymbol{A}_{is}(\boldsymbol{Z}_{is}'\boldsymbol{\beta}_2)-\sum_{r=1}^{R_{is}}\tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha},\boldsymbol{Z}_{is}'\boldsymbol{\beta}_2)\right)\right] \\
& +\sum_{j=2}^{N_i}\log(\boldsymbol{X}(t_{ij})'\boldsymbol{\alpha})+\left(\sum_{j=2}^{N_i}\boldsymbol{Y}_{is_{ij}}'\right)\boldsymbol{\beta}_1+\left(\sum_{j=2}^{N_i}\boldsymbol{Z}_{is_{ij}}'(t_ij-lt_{is_{ij}})\right)\boldsymbol{\beta}_2 \\
& +(N_i-1)\log(\delta_i)
\end{aligned}
$$

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Aggregating $\log f_i$ for all $i$'s yields the total density function for all observations.

$$
\begin{aligned}
&\log f(\{t_{ij}\}_{j=2}^{N_i}{}_{i=1}^{I} \big| \boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma, \{\delta_i\}_{i=1}^{I}, T) = \\
&- \boldsymbol{\alpha}' \left( \sum_{i=1}^{I} \delta_i \left[ \sum_{s=1}^{S_i} \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \boldsymbol{A}_{is}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right) \right] \right) \\
&+ \sum_{i=1}^{I} \sum_{j=2}^{N_i} \log(\boldsymbol{X}(t_{ij})' \boldsymbol{\alpha}) + \left( \sum_{i=1}^{I} \sum_{j=2}^{N_i} \boldsymbol{Y}_{is_{ij}}' \right) \boldsymbol{\beta}_1 + \left( \sum_{i=1}^{I} \sum_{j=2}^{N_i} \boldsymbol{Z}_{is_{ij}}' (t_{ij} - lt_{is_{ij}}) \right) \boldsymbol{\beta}_2 \\
&+ \sum_{i=1}^{I} (N_i - 1) \log(\delta_i)
\end{aligned}
$$

## 7.2 Gradient and Hessian matrix of the log-posterior

Gradient: $\nabla \log f = \left( \frac{\partial \log f}{\partial \boldsymbol{\alpha}}, \frac{\partial \log f}{\partial \boldsymbol{\beta}_1}, \frac{\partial \log f}{\partial \boldsymbol{\beta}_2}, \frac{\partial \log f}{\partial \gamma} \right)$,

Hessian Matrix: $\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \log f}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \log f}{\partial \alpha \partial \beta_1'} & \frac{\partial^2 \log f}{\partial \alpha \partial \beta_2'} & \frac{\partial^2 \log f}{\partial \alpha \partial \gamma} \\ \frac{\partial^2 \log f}{\partial \beta_1 \partial \alpha'} & \frac{\partial^2 \log f}{\partial \beta_1 \partial \beta_1'} & \frac{\partial^2 \log f}{\partial \beta_1 \partial \beta_2'} & \frac{\partial^2 \log f}{\partial \beta_1 \partial \gamma} \\ \frac{\partial^2 \log f}{\partial \beta_2 \partial \alpha'} & \frac{\partial^2 \log f}{\partial \beta_2 \partial \beta_1'} & \frac{\partial^2 \log f}{\partial \beta_2 \partial \beta_2'} & \frac{\partial^2 \log f}{\partial \beta_2 \partial \tau} \\ \frac{\partial^2 \log f}{\partial \gamma \partial \alpha'} & \frac{\partial^2 \log f}{\partial \gamma \partial \beta_1'} & \frac{\partial^2 \log f}{\partial \gamma \partial \beta_2'} & \frac{\partial^2 \log f}{\partial \gamma \partial \gamma} \end{pmatrix}$.

Define

$$
\begin{aligned}
\boldsymbol{B}_{is} &= \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \boldsymbol{A}_{is}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \tilde{\boldsymbol{A}}_{isr}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right), \\
\boldsymbol{B}_{is}^{(1)} &= \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \frac{\partial \boldsymbol{A}_{is}}{\partial \theta}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \frac{\partial \tilde{\boldsymbol{A}}_{isr}}{\partial \theta}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right), \\
\boldsymbol{B}_{is}^{(2)} &= \exp(\boldsymbol{Y}_{is}' \boldsymbol{\beta}_1) \left( \frac{\partial^2 \boldsymbol{A}_{is}}{\partial \theta^2}(\boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) - \sum_{r=1}^{R_{is}} \frac{\partial^2 \tilde{\boldsymbol{A}}_{isr}}{\partial \theta^2}(\boldsymbol{\alpha}, \boldsymbol{Z}_{is}' \boldsymbol{\beta}_2) \right).
\end{aligned}
$$

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Then

$$\frac{\partial \log f}{\partial \boldsymbol{\alpha}} = -\sum_{i=1}^{I} \frac{(N_i+\gamma-1)\sum_{s=1}^{S_i} \boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i} \boldsymbol{B}_{is}} + \sum_{i=1}^{I}\sum_{j=2}^{N_i} \frac{\boldsymbol{X}(t_{ij})}{\boldsymbol{X}(t_{ij})'\boldsymbol{\alpha}} - \frac{\boldsymbol{\alpha}}{u^2}$$

$$\frac{\partial \log f}{\partial \boldsymbol{\beta}_1} = -\sum_{i=1}^{I} \frac{(N_i+\gamma-1)\boldsymbol{\alpha}'\sum_{s=1}^{S_i} \boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i} \boldsymbol{B}_{is}} + \sum_{i=1}^{I}\sum_{j=2}^{N_i} \boldsymbol{Y}_{is_{ij}} - \frac{\boldsymbol{\beta}_1}{u^2}$$

$$\frac{\partial \log f}{\partial \boldsymbol{\beta}_2} = -\sum_{i=1}^{I} \frac{(N_i+\gamma-1)\boldsymbol{\alpha}'\sum_{s=1}^{S_i} \boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i} \boldsymbol{B}_{is}} + \sum_{i=1}^{I}\sum_{j=2}^{N_i} \boldsymbol{Z}'_{is_{ij}}(t_{ij}-lt_{is_{ij}}) - \frac{\boldsymbol{\beta}_2}{u^2}$$

$$\begin{aligned}\frac{\partial \log f}{\partial \gamma} &= -\sum_{i=1}^{I} \log\left(\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\right) - \sum_{i=1}^{I}\frac{N_i+\gamma-1}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} \\ &\quad + \sum\nolimits_{i=1}^{I}\varphi(N_i+\gamma-1) - I\varphi(\gamma) + I\log(\gamma) + I - \frac{2}{\gamma+v},\end{aligned}$$

where $\varphi(x) = \frac{\mathrm{d}\log\Gamma(x)}{\mathrm{d}x}$ is the digamma function.

$$\begin{aligned}\frac{\partial^2 \log f}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}'} &= -\sum_{i=1}^{I}(N_i+\gamma-1)\Bigg[\frac{1}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} \\ &\quad \cdot\Bigg(\sum_{s=1}^{S_i}\sum_{r:\alpha' X(\tilde{et}(\alpha)_{isr})=0} \frac{\exp(\boldsymbol{Y}'_{is}\boldsymbol{\beta}_1 + (\tilde{et}(\boldsymbol{\alpha})_{isr} - lt_{is} - \overline{\Delta t})\boldsymbol{Z}'_{is}\boldsymbol{\beta}_2)}{\boldsymbol{\alpha}'\frac{\partial X}{\partial t}(\tilde{et}(\boldsymbol{\alpha})_{isr})}\boldsymbol{X}(\tilde{et}(\boldsymbol{\alpha})_{isr})\boldsymbol{X}(\tilde{et}(\boldsymbol{\alpha})_{isr})' \\ &\quad - \sum_{s=1}^{S_i}\sum_{r:\alpha' X(\tilde{bt}(\alpha)_{isr})=0} \frac{\exp(\boldsymbol{Y}'_{is}\boldsymbol{\beta}_1 + (\tilde{bt}(\boldsymbol{\alpha})_{isr} - lt_{is} - \overline{\Delta t})\boldsymbol{Z}'_{is}\boldsymbol{\beta}_2)}{\boldsymbol{\alpha}'\frac{\partial X}{\partial t}(\tilde{bt}(\boldsymbol{\alpha})_{isr})}\boldsymbol{X}(\tilde{bt}(\boldsymbol{\alpha})_{isr})\boldsymbol{X}(\tilde{bt}(\boldsymbol{\alpha})_{isr})'\Bigg) \\ &\quad - \frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\Bigg] - \sum_{i=1}^{I}\sum_{j=2}^{N_i}\frac{\boldsymbol{X}(t_{ij})\boldsymbol{X}(t_{ij})'}{\left(\boldsymbol{X}(t_{ij})'\boldsymbol{\alpha}\right)^2} - \frac{1}{u^2}\mathbf{I}_{dim(\alpha)}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \log f}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}'_1} &= -\sum_{i=1}^{I}(N_i+\gamma-1)\left[\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}\boldsymbol{Y}'_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\right] \\ &\quad - \frac{1}{u^2}\mathbf{I}_{dim(\beta_1)}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \log f}{\partial \boldsymbol{\beta}_2\partial \boldsymbol{\beta}'_2} &= -\sum_{i=1}^{I}(N_i+\gamma-1)\left[\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(2)}\boldsymbol{Z}_{is}\boldsymbol{Z}'_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\right] \\ &\quad - \frac{1}{u^2}\mathbf{I}_{dim(\beta_2)}\end{aligned}$$

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

$$\frac{\partial^2 \log f}{\partial\gamma\partial\gamma} = -\sum_{i=1}^{I}\left[\frac{2}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{N_i+\gamma-1}{\left(\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\right)^2}\right] + \sum_{i=1}^{I}\varphi_1(N_i+\gamma-1) - I\varphi_1(\gamma) + \frac{I}{\gamma} + \frac{2}{(\gamma+v)^2}$$

where $\varphi_1(x) = \frac{\mathrm{d}\varphi(x)}{\mathrm{d}x}$ is the trigamma function.

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}_1'} = -\sum_{i=1}^{I}(N_i+\gamma-1)\left[\frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}'}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\right]$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}_2'} = -\sum_{i=1}^{I}(N_i+\gamma-1)\left[\frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}'}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\right]$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\alpha}\partial\gamma} - \sum_{i=1}^{I}\left[\frac{\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{(N_i+\gamma-1)\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}{\left(\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\right)^2}\right]$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'} = -\sum_{i=1}^{I}(N_i+\gamma-1)\left[\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}\boldsymbol{Z}_{is}'}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\left(\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}}\right)'\right]$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\beta}_1\partial\gamma} = -\sum_{i=1}^{I}\left[\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{(N_i+\gamma-1)\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\boldsymbol{Y}_{is}}{\left(\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\right)^2}\right]$$

$$\frac{\partial^2 \log f}{\partial\boldsymbol{\beta}_2\partial\gamma} = -\sum_{i=1}^{I}\left[\frac{\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}} - \frac{(N_i+\gamma-1)\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}^{(1)}\boldsymbol{Z}_{is}}{\left(\gamma+\boldsymbol{\alpha}'\sum_{s=1}^{S_i}\boldsymbol{B}_{is}\right)^2}\right]$$

## 7.3 Posterior mode finding and normal approximation

Let $\boldsymbol{\Theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma\}$. For convenience, we abbreviate the posterior distribution in (10) as $f(\boldsymbol{\Theta})$. Note that $\log f(\boldsymbol{\Theta})$ is differentiable everywhere and twice-differentiable almost everywhere. We thus adopt a modified Newton-Raphson (NR) algorithm with adaptive step-size in each iteration so as to avoid the chaotic behavior of the classic NR update (Amrein and Wihler (2014)). Formulas for the gradient and Hessian matrix of (10) are provided in Appendix 7.2.

We initialize $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with vectors of zeros, and $\gamma$ with a small value, say 0.01, to allow for a large heterogeneity. The initial value of the seasonal effect $\boldsymbol{\alpha}$ is most crucial, and is set to be a preliminary estimate from the following procedure, which applies the Fourier transform to an

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

empirically estimated purchase rate function:

1. Partition each year into $W = 52$ equal-length time windows. With a simple Poisson model, we estimate the average purchase rate $\hat{\mu}_w$ of the $w$th interval as $Wy_w/I_w$ and its standard error $s(\hat{\mu}_w)$ as $W\sqrt{y_w}/I_w$, where $y_w$ is the total number of purchases that occurred within the $w$th interval of all years and $I_w$ the number of potential buyers aggregating over all years for that interval. The empirical purchase rate function $\hat{P}(t)$ is estimated as being equal to $\hat{\mu}_w$ in the $w$th time window of the year.

2. Compute the Fourier transform of the empirical purchase rate function to obtain:

$$\begin{aligned}
\hat{\alpha}_0 &= \frac{\sum_{w=1}^{W} \hat{\mu}_w / s(\hat{\mu}_w)^2}{\sum_{w=1}^{W} 1/s(\hat{\mu}_w)^2} = W \frac{\sum_{w=1}^{W} I_w}{\sum_{w=1}^{W} I_w^2/y_w} \\
\hat{\alpha}_{2m-1} &= \frac{2W}{\sum_{w=1}^{W} I_w^2/y_w} \sum_{w=1}^{W} \sin\left(2\pi m \frac{w-.5}{W}\right) I_w, \ m = 1, \ldots, M; \\
\hat{\alpha}_{2m} &= \frac{2W}{\sum_{w=1}^{W} I_w^2/y_w} \sum_{w=1}^{W} \cos\left(2\pi m \frac{w-.5}{W}\right) I_w, \ m = 1, \ldots, M.
\end{aligned}$$

3. Check if the resulting $\hat{\boldsymbol{\alpha}}$ satisfies $\boldsymbol{X}(t_{ij})'\hat{\boldsymbol{\alpha}} > 0$ for all $i$ and $j$'s. Otherwise, increase $\hat{\alpha}_0$ by $0.001 - \min_{i,j} \boldsymbol{X}(t_{ij})'\hat{\boldsymbol{\alpha}}$ to satisfy this requirement.

With these initial values, the modified NR algorithm converges typically in fewer than 20 iterations for all the five datasets. Around its mode $\hat{\boldsymbol{\Theta}}$, the posterior distribution $f(\boldsymbol{\Theta})$ can be approximated well by a normal distribution $N(\hat{\boldsymbol{\Theta}}, -\mathbf{H}(\hat{\boldsymbol{\Theta}})^{-1})$ when sample size is large enough (O'Hagan et al. (2004)), where $H(\boldsymbol{\Theta})$ is the Hessian matrix.

# 8 Supplemental Material

It details the MCMC posterior sampling procedures introduced in Section 4.2, together with the results for our ticket transaction data (a PDF file), available at the journal's website.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

## References

Ailawadi, K. L. and Neslin, S. A. (1998). The effect of promotion on consumption: Buying more and consuming it faster. *Journal of Marketing Research*, 35(3):390–8.

Allenby, G., Leone, R., and Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, 94(446):365–74.

Amrein, M. and Wihler, T. P. (2014). An adaptive newton-method based on a dynamical systems approach. *Communications in Nonlinear Science and Numerical Simulation*, 19(9):2958–2973.

Ansari, A., Mela, C., and Neslin, S. (2008). Customer channel migration. *Journal of Marketing Research*, 45(February):60–76.

Bijwaard, G. E., Franses, P. H., and Paap, R. (2006). Modeling purchases as repeated events. *Journal of Business and Economic Statistics*, 24(4):487–502.

Bitran, G. R. and Mondschein, S. V. (1996). Mailing decisions in the catalog sales industry. *Management Science*, 42(9):1364–81.

Boatwright, P., Borle, S., and Kadane, J. (2003). A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association*, 98(463):564–72.

Bucklin, R. E. and Lattin, J. M. (1991). Two-state model of purchase incidence and brand choice. *Marketing Science*, 10(1):24–39.

Chan, T., Narasimhan, C., and Zhang, Q. (2008). Decomposing promotional effects with a dynamic structural model of flexible consumption. *Journal of Marketing Research*, 45(4):487–98.

Chintagunta, P. and Haldar, S. (1998). Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research*, 35(1):45–53.

Chintagunta, P. K. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2):184–208.

Cinlar, E. (2013). *Introduction to stochastic processes*. Courier Dover Publications.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Elsner, R., Krafft, M., and Huchzermeier, A. (2004). Optimizing rhenania's direct marketing business through dynamic multilevel modeling (dmlm) in a multicatalog-brand environment. *Marketing Science*, 23(2):192–206.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.

Gelman, A., Roberts, G., and Gilks, W. (1996). Efficient metropolis jumping rules. *Bayesian Statistics 5, JM Bernardo, JO Berger, AP Dawid, and AFM Smith (eds)*, pages 599–607.

Gonul, F. and Shi, M. Z. (1998). Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Science*, 44(9):1249–62.

Gonul, F. and Srinivasan, K. (1993). Consumer purchase behavior in a frequently bought product category. *Journal of the American Statistical Association*, 88(88):1219–27.

Gonul, F. F. and Hofstede, F. T. (2006). How to compute optimal catalog mailing decisions. *Marketing Science*, 25(1):65–74.

Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, 25(4):342–55.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Jain, D. and Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, 10(10):1–23.

Kamakura, W. and Kang, W. (2007). Chain-wide and store-level analysis for cross-category management. *Journal of Retailing*, 83(2):159–70.

Kumar, V., George, M., and Pancras, J. (2008). Cross-buying in retailing: Drivers and consequences. *Journal of Retailing*, 84(1):15–27.

Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing* Springer Science & Business Media.

Manchanda, P. J. D., Goh, K., and Chintagunta, P. (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108.

Ngobo, P. (2001). Drivers of customers' cross-buying intentions. *European Journal of Marketing*, 38(9/10):1129–57.

O'Hagan, A., Forster, J., and Kendall, M. G. (2004). *Bayesian inference*. Arnold London.

Osinga, E. C., Leeflang, P. S., and Wieringa, J. E. (2010). Early marketing matters: A time-varying parameter approach to persistence modeling. *Journal of Marketing Research*, 47(February):173–85.

Reinartz, W., Thomas, J. S., and Bascoul, G. (2008). Investigating cross-buying and customer loyalty. *Journal of Interactive Marketing*, 22(1):5–20.

Roberts, M. L. and Berger, P. D. (1989). *Direct Marketing Management*. Prentice-Hall, Inc, Englewood Cliffs, New Jersey.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Seetharaman, P. B. and Chintagunta, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business and Economic Statistics*, 21(3):368–82.

VanBoskirk, S. Us interactive marketing forecast, 2007 to 2012. `https://member-west.express-scripts.com/images/pdf/Previewtest.pdf`.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Venkatesan, R., Kumar, V., and Bohling, T. (2007). Optimal customer relationship management using bayesian decision theory: An application for customer selection. *Journal of Marketing Research*, 44(4):579–94.

Zhang, J. and Krishnamurthi, L. (2004). Customizing promotions in online stores. *Marketing Science*, 23(4):561–78.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

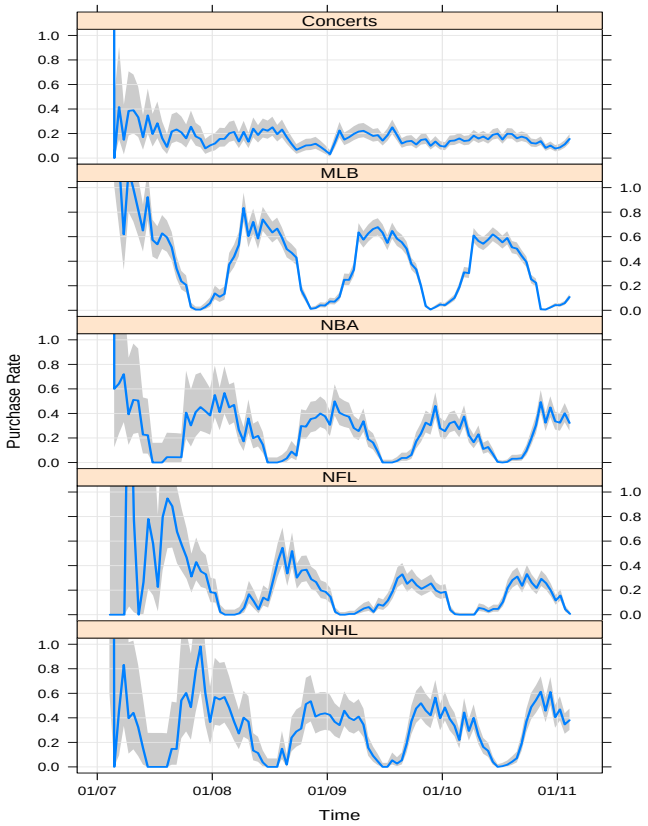


Figure 1: Empirical purchase rate curve and its 95% confidence band (gray) for each genre.

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

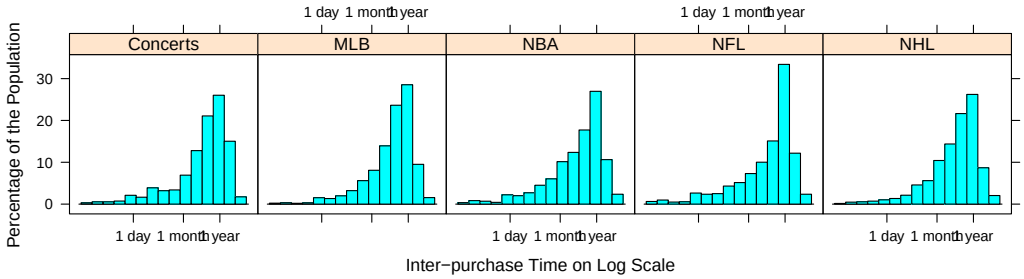


Figure 2: Histograms of average inter-purchase time on the log scale for each genre.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

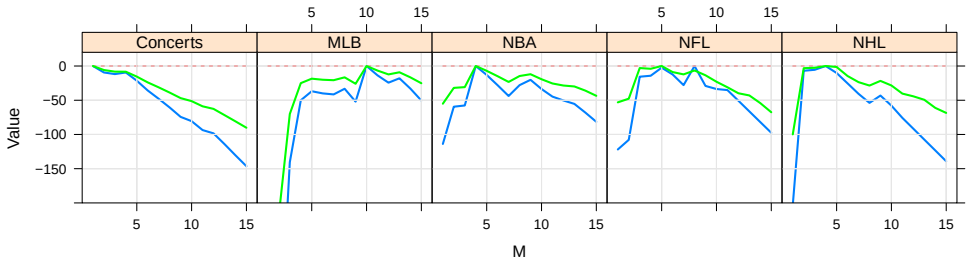


Figure 3: Model selection using log-posterior probability (green curve) and $-BIC$ (blue curve). Values are adjusted by constant terms so that they all share the same maximum of 0. The peak of a curve indicates the optimal choice of $M$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

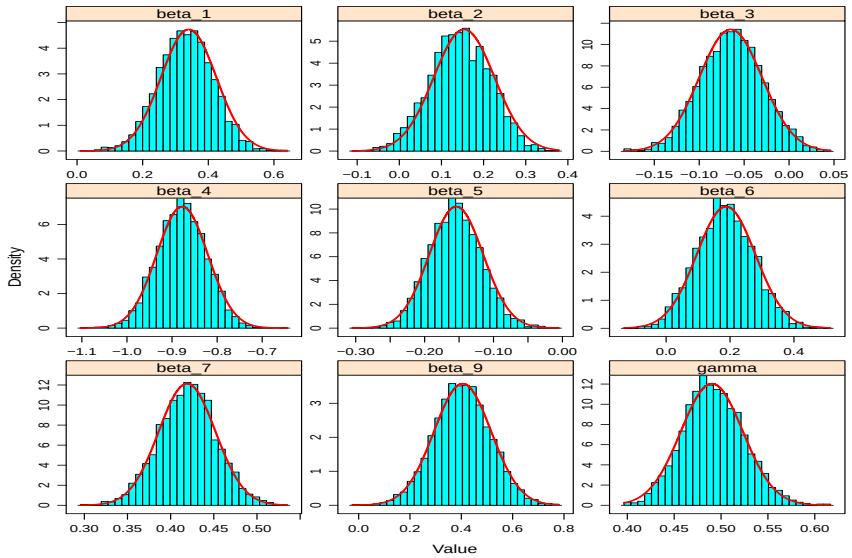


Figure 4: Histograms of posterior samples obtained by the MCMC algorithm overlaid with the asymptotic normal density curves for the NHL ticket sales data. Results for other genres are similar (provided in Supplementary Materials).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

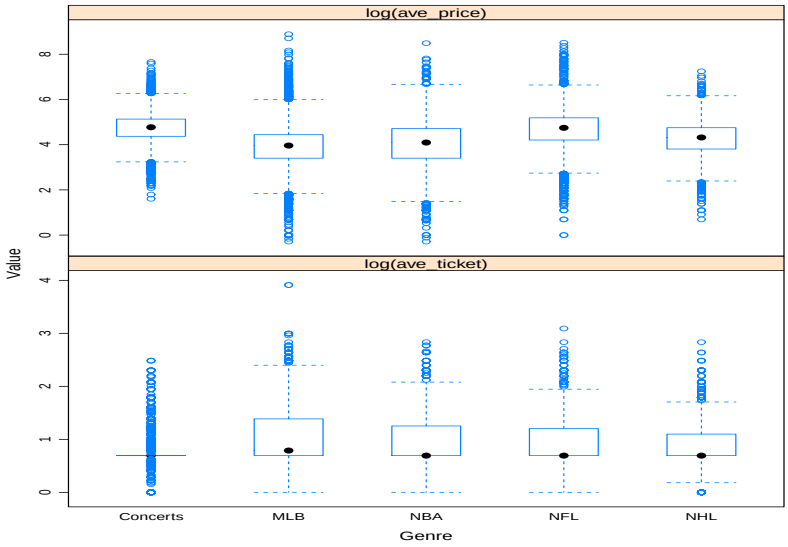


Figure 5: Box plots of average ticket price and average ticket count per order

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

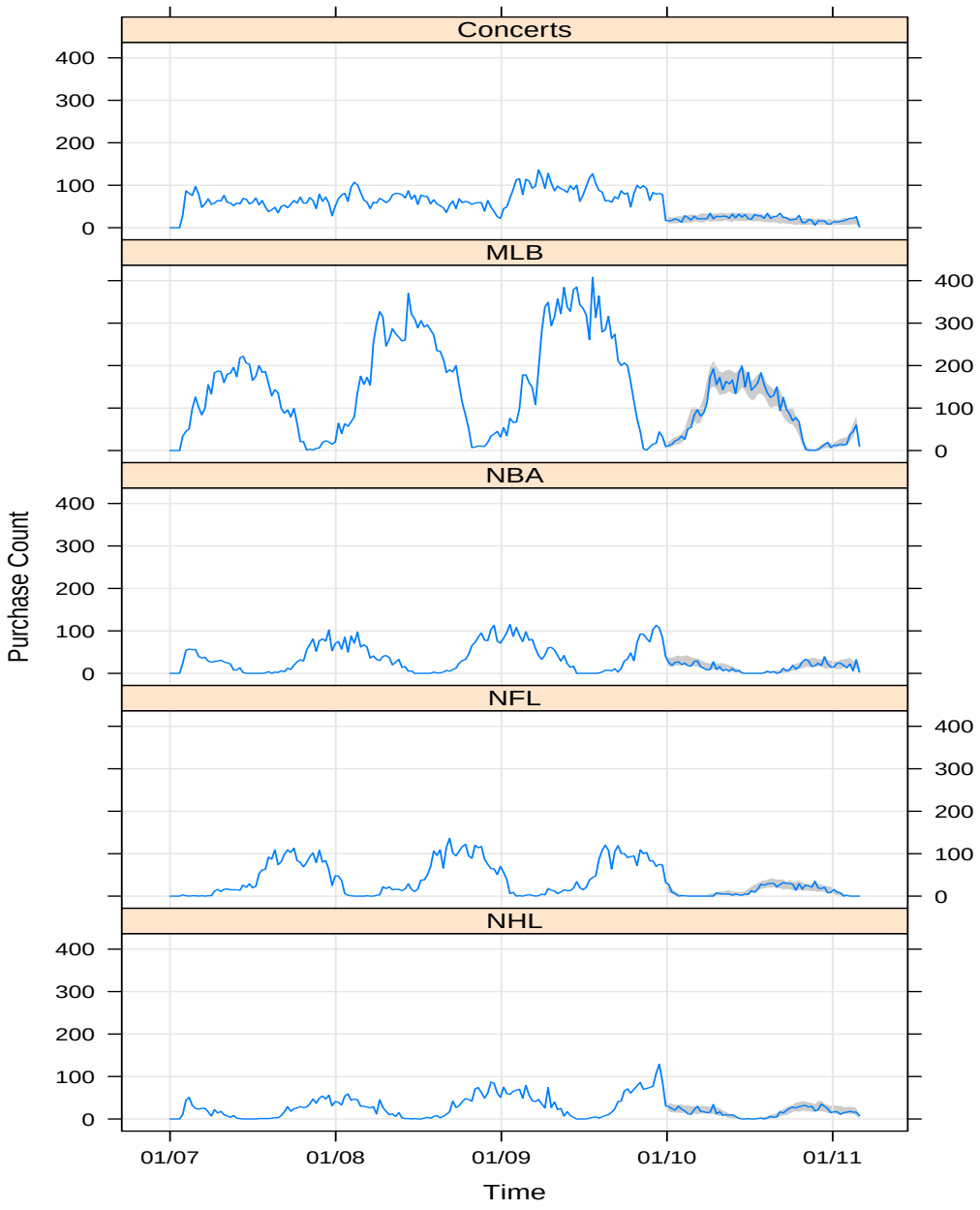


Figure 6: Prediction validation. The figures overlay the actual weekly sales (blue) and the 95% prediction bands (gray).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

Table 1: Summary statistics of the email marketing data

| | Genre | | | | |
|---|---|---|---|---|---|
| | Concerts | MLB | NBA | NFL | NHL |
| # of customers | 12213 | 21066 | 6719 | 7617 | 4974 |
| # of transactions | 15985 | 36362 | 9604 | 10064 | 7742 |
| Average ticket price | 146.80 | 70.07 | 89.99 | 150.62 | 89.17 |
| Average # of tickets per order | 2.45 | 3.22 | 2.92 | 2.93 | 2.79 |
| # of customers who received offer(s) | 10588 | 6861 | 4862 | 3675 | 4199 |
| # of customers who never received offers | 1625 | 14205 | 1857 | 3942 | 775 |

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2: Results of propensity score matching (B – before; A – after)

| | | Genre | | | | |
|---|---|---|---|---|---|---|
| | | Concerts | MLB | NBA | NFL | NHL |
| Treatment group size | B | 10588 | 6861 | 4862 | 3675 | 4199 |
| | A | 7758 | 3511 | 4278 | 2934 | 2532 |
| Control group size | B | 1625 | 14205 | 1857 | 3942 | 775 |
| | A | 1191 | 7270 | 1635 | 3148 | 467 |
| Size ratio | B | 6.516 | 0.483 | 2.618 | 0.932 | 5.418 |
| | A | 6.514 | 0.483 | 2.617 | 0.932 | 5.422 |
| Difference in average length of history | B | -0.392 | -1.037 | 0.081 | 0.156 | -0.652 |
| | A | 0.006 | 0.102 | -0.110 | -0.153 | -0.003 |
| Difference in average log(ticket price) | B | 0.041 | 0.102 | 0.003 | -0.001 | 0.129 |
| | A | -0.005 | 0.038 | 0.003 | 0.025 | -0.056 |
| Difference in average log(ticket count) | B | 0.002 | 0.009 | 0.024 | -0.011 | 0.014 |
| | A | -0.001 | -0.008 | -0.019 | -0.004 | 0.001 |
| Difference in average number of genres | B | -0.089 | 0.319 | 0.191 | 0.297 | -0.086 |
| | A | -0.039 | 0.176 | 0.033 | 0.023 | -0.090 |

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3: List of main notations

| Notation | Definition |
|---|---|
| $T$ | Censoring time |
| $I$ | Total number of customers |
| $i$ | Customer index |
| $j$ | Transaction index |
| $N_i$ | Total number of transactions (up to $T$) made by customer $i$ |
| $t_{ij}$ | Time of the $j$th transaction of customer $i$ |
| $\lambda_i(t)$ | Purchase rate of customer $i$ at time $t$ |
| $o_i(t)$ | Indicator of whether customer $i$ has an offer at time $t$ |
| $eo_i(t)$ | Indicator of whether customer $i$ has an offer expired within the last two months at time $t$ |
| $lt_i(t)$ | Time of the last purchase of customer $i$ dated back from $t$ |
| $p_i$ | Average ticket price of customer $i$ |
| $tct_i$ | Average number of tickets that customer $i$ purchased per order |
| $g_i$ | Number of genres that customer $i$ has purchased |
| $\tau$ | Average inter-purchase time of all customers |

Downloaded by [Jun Liu] at 05:43 20 January 2016

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Jun Liu] at 05:43 20 January 2016

Table 4: Posterior means standard deviations for parameters of interest

| Coefficient | Term | Genre | | | | |
|---|---|---|---|---|---|---|
| | | Concerts | MLB | NBA | NFL | NHL |
| $\beta_1$ | Offer | -0.023 | -0.035 | 0.061 | 0.242 | 0.340 |
| | | (.103) | (.128) | (.096) | (.111)* | (.084)** |
| $\beta_2$ | Post offer | -0.009 | 0.172 | 0.009 | 0.164 | 0.153 |
| | | (.078) | (.084)* | (.084) | (.100) | (.072)* |
| $\beta_3$ | Length of history | 0.026 | -0.041 | -0.115 | -0.096 | -0.065 |
| | | (.033) | (.014)** | (.037)** | (.039)* | (.037)· |
| $\beta_4$ | Recency | -0.846 | -0.734 | -0.807 | -0.684 | -0.877 |
| | | (.045)** | (.022)** | (.054)** | (.056)** | (.058)** |
| $\beta_5$ | Average ticket price | 0.373 | -0.107 | -0.038 | 0.067 | -0.154 |
| | | (.037)** | (.016)** | (.027) | (.032)* | (.039)** |
| $\beta_6$ | Average ticket count | 0.110 | -0.135 | 0.152 | 0.191 | 0.188 |
| | | (.067)· | (.035) ** | (.079)* | (.069)** | (.093)** |
| $\beta_7$ | Number of genres | 0.506 | 0.545 | 0.533 | 0.497 | 0.419 |
| | | (.029)** | (.018)** | (.031)** | (.032)** | (.032)** |
| $\beta_9$ | Offer * Recency | 0.219 | -0.027 | 0.606 | 0.137 | 0.405 |
| | | (.138) | (.131) | (.108)** | (.157) | (.111)** |
| $\gamma$ | 1/Heterogeneity | 0.689 | 0.661 | 0.451 | 0.569 | 0.490 |
| | | (.043)** | (.018)** | (.027)** | (.038)** | (.030)** |

– Posterior standard deviations in parentheses. Significance revealed by 99% credible interval is denoted by (**), 95% credible interval by (*), and 90% credible interval by (·).
– $\beta_8, \beta_{10}, \beta_{11}$, and $\beta_{12}$ are never statistically significant, thus omitted from the table.

ACCEPTED MANUSCRIPT