Submitted to the Annals of Applied Statistics

ESTIMATING PROMOTION EFFECTS IN EMAIL MARKETING USING A LARGE-SCALE CROSS-CLASSIFIED BAYESIAN JOINT MODEL FOR NESTED IMBALANCED DATA

BY SABYASACHI MUKHOPADHYAY[1], WREETABRATA KAR[2,*] AND GOURAB MUKHERJEE[3,†]

[1]*Department of Quantitative Methods and Information Systems, Indian Institute of Management, sabyasachi.mukhopadhyay@iimu.ac.in*

[2]*Krannert School of Management, Purdue University, [*]wkar@purdue.edu*

[3]*Department of Data Sciences and Operations, University of Southern California, [†]gourab@usc.edu*

We consider a large-scale, *cross-classified nested* (CRON) joint model for modeling customer responses to opening, clicking, and purchasing from promotional emails. Our logistic regression-based joint model contains crossing of promotions and customer effects, and allows estimation of the heterogeneous effects of different promotion emails after adjusting for customer preferences, attributes, and historical behaviors. Using data from an email marketing campaign of an apparel company, we exhibit the varying effects of promotions not only based on the contents of the email but also across the three different stages, viz. open, click, and purchase, of the conversion funnel. We conduct Bayesian estimation of the parameters in the joint model by using a block Metropolis-Hastings algorithm that not only incorporates nested subsampling to tackle the severe imbalance between conversions and no conversions, but also uses additive transformation-based modifications of random walk Metropolis to scale estimation for large numbers of customers. We extend our approach to a *segmented cross-classified nested* (SCRON) joint model that encompasses the possibility of varying promotion effects across different customer segments. The resultant high-dimensional model is estimated using spike-and-slab priors on the promotion and customer type interactions. Our nested joint model captures the correlations in customer preferences across the conversion funnel. Based on the promotion estimates from the model, we demonstrate how marketers can use different priced, non-priced, and combination of price and non-price promotions to increase brand awareness or increase purchases. Comparing estimates from CRON and SCRON models, we display the benefits of targeted marketing by using email promotion lists which are separately optimized for the different customer segments.

1. Introduction. Over the past decade, the digital revolution has changed the paradigm of marketing promotions. According to studies by Forrester Research and eMarketer, in 2020, an average firm was expected to allocate 45% of their total marketing budget to online media (Leone, 2020). Among the different types of digital promotions, email marketing is the most popular, as it yeilds the highest *return on investment* (ROI) (Lee, 2019). The massive popularity of email marketing is primarily due to two factors: its reach, as at least 91% of current internet users participate in digital activities using emails; and secondly, the minimal cost of sending out an email compared to other digital media (Perrin, 2019). Despite such wide usage, promotion campaigns through emails are currently facing the colossal challenge of relevance. In a recent study by Aldighieri (2019), it was reported that, while 55% of marketers

Keywords and phrases: logistic regression, joint modeling, imbalanced data, spike-and-slab priors, bayesian hierarchical model, promotional email, high-dimensional regression.

think that more than half of all the emails sent to customers were relevant, only 14% of customers believe that more than half of the emails from an organization were relevant to them. Promotions lacking relevance cause customer alienation and fatigue that may eventually lead to unsubscription of emails or even substitution by competitors (Forte, 2019). Hence, it is extremely important to understand the role of different promotions, so that relevant, refined, and targeted email campaigns can be launched.

Marketers are increasingly using data-driven techniques to study promotion effects in email campaigns (see Zhang, Kumar and Cosguner, 2017; Sahni, Wheeler and Chintagunta, 2018; Sahni, Zou and Chintagunta, 2017 and the references therein). Here, based on customer responses to a retailer's email marketing campaign (which is the application case described in detail in Section 2), we develop a disciplined statistical framework to accurately estimate varied promotion effects. Promotions can be classified into two fundamentally different groups: (a) *priced*, which includes discount percentages or discount dollars of varied magnitudes, and (b) *non-priced*, such as buy-one-get-one free or free shipping or free samples or free returns. Along with the above types, modern-day retailers also use different combinations of priced and non-priced promotions. Statistically, this gives rise to a high-dimensional estimation problem, where the efficacy of multiple promotions simultaneously used in an email campaign needs to be estimated. These new-age problems are further compounded with nonlinearity, imbalance, heterogeneity, and big-data computational challenges. For this purpose, we develop a flexible, novel Bayesian methodology that conducts efficient estimation in the presence of the aforementioned challenges in a high-dimensional regression model. These features of our method are described in Section 1.3. Next, we briefly describe the chief objectives and the associated responses in digital marketing campaigns.

1.1. *Email Marketing: Promotions and Conversion Funnel.* Most contemporary email marketing campaigns track their effectiveness by evaluating the three metrics: open, click, and purchase rates of promotional emails. The open-click-purchase metrics are not mutually independent but are inter-dependent nested stages constituting the *conversion funnel* (Zhang, Kumar and Cosguner, 2017), which a customer starts by opening the promotional email. These metrics are highly heterogeneous across different campaigns. For the US market, the open rates of email promotions in 2018 varied between 19% and 54%, while the click rate varied between 3% and 10% (Perrin, 2019). Purchase, being the terminal point at the bottom of the funnel, is directly related to revenue. However, it is also equally important to estimate a promotion's effectiveness at the open and click stages, for they reflect brand awareness, customer acquisition, and customer engagement, which are important attributes for tracking daily operations and business growth. Different promotional strategies have disparate impacts on the aforementioned three stages of the conversion funnel. For estimating promotion effectiveness, we develop a joint modeling framework (Rizopoulos, 2012) that incorporates the interdependencies among the customer responses across the conversion funnel.

1.2. *Joint Modeling of Customers' Responses to Promotion Emails.* We model customer responses to an email marketing campaign of an apparel company. Finding relevance of different kinds of promotions in this sector is particularly important, as on average 60% of shoppers currently feel that promotional emails do not cater to their tastes and interests (Kapner, 2017). We use a nested logistic regression setup for the three stages in the conversion funnel. Recently, Zhang, Kumar and Cosguner (2017), using a hidden Markov model, showed that it is critical to control for the correlated nature of the conversion funnel. However, they considered an aggregated setup where conversion denoted one or more purchases made by a customer from any number of emails she might have received in a month. Thus, their model is agnostic about the promotion content. Our joint modeling framework, based upon a generalized linear model (GLM), not only allows for modeling codependencies across the three stages in the conversion funnel, but also allows for measuring customer responses at the disaggregative level for all promotion types and contents. It can generate rich insights into which promotions are more effective at what stage of the conversion funnel, thereby developing actionable insights for managers, depending on what the objective of the promotion is. To the best of our knowledge, this ranks among the first works on joint modeling of the conversion funnel based on promotion content types.

Our proposed GLM-based joint modeling framework predicts customer choices based on promotion effects as well as customer's recency, frequency, and monetary (RFM) values. We entertain nonlinear effects of a customer's recency values, measured in days since she made her last purchase, or days since her latest click or opening of a promotion email. These RFM covariates are described in Section 2. A critical challenge in jointly modeling the three stages is high imbalance between the positive and negative responses, because of low click rates and even lower purchase rates in promotional emails. We develop an imbalance-corrected Bayesian logistic regression framework that also encompasses interactions betweeen RFM values and promotion effects. The resultant model can be used in designing targeted marketing campaigns (as done in Section 5.4 of the application case) that use different promotion contents for different strata of the customer base.

1.3. *Background and Statistical Challenges.* Traditionally, the effectiveness of promotions is studied at the end stage of the conversion funnel based on buy or not buy decisions. While Kamakura and Kang (2007) and Osinga, Leeflang and Wieringa (2010) used sale response model to study purchases at the aggregate level, recent works of Wu, Li and Liu (2018); Zhang, Kumar and Cosguner (2017) demonstrated the benefits of conducting efficiency analysis at the individual customer-identifier level. Using randomized field experiments, Gopalakrishnan and Park (2019) and Sahni, Wheeler and Chintagunta (2018) demonstrated empirical evidence of the heterogeneous behavior of customers across the different stages of the conversion funnel. Here, we use Bayesian joint modeling (Rizopoulos and Ghosh, 2011; Rizopoulos, 2012; McCulloch, 2008; Rizopoulos and Lesaffre, 2014), which provides a disciplined framework for simultaneously studying customer-level responses across the three stages of the conversion funnel. However, email-marketing data containing responses to multiple email promotions from a large number of customers, and the business model associated with marketing funnels pose several statistical challenges and necessitates novel extensions of the conventional Bayesian joint modeling regression framework. We describe the salient features of our approach below.

Evaluating promotion content and a two-way cross classified joint model. While extant academic research has extensively worked on determining the optimal promotion frequency and volume (see Wu, Li and Liu, 2018; Seetharaman and Chintagunta, 2003, and the references therein), very limited research has delved into the content of the promotion. The extent of research on promotion content has been mostly restricted to determining, given a customer segment(s) or engagement propensity, whether a price promotion was better than a non-price promotion. However, given the current variation in offers, the above insight is not quite actionable for a retailer. The retailer needs to know whether a price promotion should be about discount percentage or discount dollars; whether a non-price promotion should be buy-one-get-one free, or free shipping or free samples or free returns, and how much is the difference in their effects, if any; whether it is beneficial to use combinations of priced and non-priced promotions. For that purpose, here we estimate the efficacy of each of the different promotion emails (different subject line or text content) used in the campaign. Statistically, this translates to considering 2-way cross-classified models with crossed effects for customers' personal preferences to email marketing and a promotion email's attractiveness.

We use correlated random effects for modeling individual customer preferences at the different stages of the conversion funnel, and fixed effects for studying his/her responses to promotions. While using a cross-classified model enabled us to estimate promotion effects, we encounter severe imbalance in the responses (between conversion and no conversion), which is controlled in aggregated models and which does not differentiate between promotion content as in Zhang, Kumar and Cosguner (2017), who aggregated responses at the monthly resolution. Also, estimating parameters in a large cross-classified model with random effects is challenging. We describe these issues next.

Imbalanced data. As the conversion rate in email marketing can be quite low, often severe imbalance is observed in the purchase stage of the conversion funnel. It is well known (see Ch. 16 of Kuhn et al., 2013, Fithian and Hastie, 2014, Johndrow et al., 2019 and the references therein) that, in modeling discrete classes, highly disproportional relative frequencies of the classes can have a significant impact on the effectiveness of the model. Subsampling approaches with subsequent correction of the regression estimates (Fithian and Hastie, 2014; Wang, Zhu and Ma, 2018) is one of the popular approaches to tackling imbalance. Here, in Section 4.2.1, we develop a nested two-stage subsampling scheme that can correct imbalance across the different stages of the conversion funnel.

Scalable estimation for Big-data. Unlike, recent works in email marketing (Zhang, Kumar and Cosguner (2017) considered 200 and Wu, Li and Liu (2018) had 2000 focal customers), here we have a very large sample of customers (around 78000). Estimating regression coefficients in such a massive cross-classified *generalized linear mixed model* (GLMM) presents several computational challenges. As such, scaling mixed effects models in massive customer-level data for modeling the customerâĂŹs personal preferences to a product is a vibrant topic in contemporary statistics (Gao et al., 2017; Gao, 2017; Zhang et al., 2016). In a two-way cross-classified balanced linear mixed effect model, Papaspiliopoulos, Roberts and Zanella (2020) have recently established fast convergence of a collapsed Gibbs sampler. We use a related block Metropolis-Hastings (BMH) algorithm (see sec. 4.2.2) for Bayesian estimation of the parameters in our large and complex cross-classified joint modeling setup. To scale estimation, we use additive transformation-based modifications of random walk Metropolis (Dutta and Bhattacharya, 2014; Dey et al., 2016), which greatly increased the acceptance rate in BMH. We establish convergence properties of the BHM algorithm with additive transformations (BHMT). We run BHMT in a parallel computing framework on multiple machines, each based on a subsample of the data, suitably selected, as explained later in sec. 4.2.1, to tackle imbalanced responses. We combine posterior inference from these multiple runs on subsampled data by using the consensus Monte Carlo approach of Scott et al. (2016) and Huang and Gelman (2005).

Nonlinear RFM effects. Customer characteristics and purchases greatly vary as RFM values changes. To estimate the effects of promotion emails, we need to look at residual customer responses after adjusting for the influence of the RFM variables. However, the effect of the RFM variables is not linear (Bult and Wansbeek, 1995; Fader, Hardie and Lee, 2005). Figure 3 shows the marginal distribution of the recency variable. We model its nonlinear effect using cubic splines (See ch. 6 of James et al., 2013). However, incorporating a nonlinear structure that not only well fits the empirical observation but also is tractable in the large-scale joint modeling framework is non-trivial. The flexible generalized cross validation methodology in Ruppert (2002) would be difficult to implement in our large-scale joint modeling setup. In Section 4.1, based on conditional distributions, we implement a pragmatic, computationally cheap, heuristic policy for capturing nonlinear RFM effects.

Customer segmentation and high-dimensional promotion effects. In the marketing literature, it is well documented that promotions may impact different customer segments differently (Rossi, McCulloch and Allenby, 1996; Gopalakrishnan and Park, 2019). Assuming that the customer segments are known from previous market research, we introduce interactions between promotions and customer types to encompass the possibility of varying promotion effects across different customer segments in our cross-classified joint model. Even with a moderate number of segments, the resultant model is high-dimensional. However, most of these interaction effects will not be significantly different from their modal values. In Section 4.4 , we use spike-and-slab prior (Ročková and George, 2018; Castillo, Schmidt-Hieber and Van der Vaart, 2015) on the interaction parameters to discard insignificant effects.

The rest of the paper is organized as follows. In Section 2, we describe our data, which is obtained from email marketing campaigns of an apparel retailer. In Section 3, we present our cross-classified joint modeling framework and we detail the methodology used for estimating the proposed model in Section 4. Using the data set introduced in Section 2, we present inference and business implications of the results obtained by our proposed analysis in Section 5. We demonstrate how marketers should use different promotions for increasing brand awareness or purchases. We also show (see Section 5.5) the benefits of targeted marketing by using email promotion lists separately optimized for the different customer segments. Figures and tables referred in the paper with the prefix S are presented in the supplementary material.

2. Data. Our data set consists of responses from 77986 customers to a series of promotion emails sent to them in a campaign season. Overall, there were 25 promotional emails that were used during this campaign. Out of these, nine are solely priced (P) promotions, seven are solely non-priced (N), and nine are combination (C) promotions having priced as well as non-priced incentives. Our data contain a recipient's responses to opening, clicking (if opened), and purchasing (if any, after clicking) from each of the promotional emails.

Fig 1: Pairwise scatterplots of customer characteristics: avergage yearly web spendings (in USD), past purchase frequency, age (in years) and income (in USD).

Along with customer responses to the promotional emails, we have the following additional information regarding each customer: (a) *age*, which is binned as a categorical variable with five levels: less than 17 years; 18–35 years; 36–49 years; 50–64 years; 65 plus years;

(b) income, binned as a categorical variable with nine levels: under \$15K, \$15–25K, \$25–35K, \$35–50K, \$50–75K, \$75–100K, \$100–120, \$120–149K, \$150K plus; (c) three recency variables pertaining to the number of days since the customer last opened (*Days Since Last Opened*), clicked (*Days Since Last Clicked*), and purchased (*Days Since Last Purchased*) from a promotion email; (d) frequency of purchases ((Past Pur. Freq.), tabulated as the number of orders made in the last two years; (e) average order values separately tabulated for in-store (*Ave. Retail Spend*) and internet (*Ave. Web Spend*) purchases.

Figure 1 shows the pairwise distributions of these customer-identifier level attributes; income and age buckets were judiciously clubbed in the figure for presentational ease. For each promotion, Table S.2 presents the aggregated values of these statistics summarized across the recipients, wherein we observe that the average customer characteristics pertaining to the frequency and monetary values (Ave. Retail Spend and Ave. Web Spend) are very similar across all the different promotions. Figure 2 (right plot) shows the average customer responses for the different promotion emails; the average responses vary significantly, not only across promotions but also across the stages of the conversion funnel.

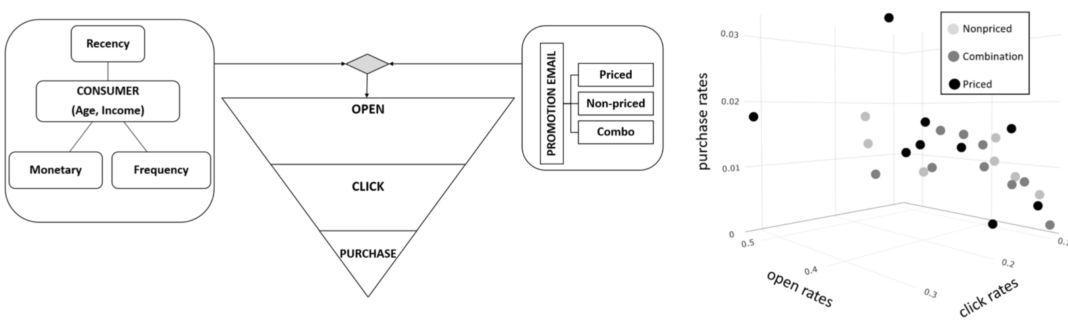


Fig 2: Left: Schematic diagram showing different components of the data; Customer responses at different stages of the conversion funnel are jointly modeled based on customer characteristics and the content of a promotion email. Right: Scatterplot of average customer responses to different promotion emails.

Figure 2 (left diagram) shows a schematic demonstrating the different components of the data. We model the customer responses at the three different stages of the conversion funnel based on his/her age, income, recency, frequency, and monetary characteristics as well as the contents of the promotion email. We have an imbalanced design as not all the customers received every email. On average, an arbitrary customer received around 21 emails. Table S.1 shows the distribution of the number of emails received while Figure S.1 shows the coverage of each of the promotional emails. From the above data, we randomly select $T = 10000$ customers as test data and use the data corresponding to the remaining $N = 67986$ customers as our training data set.

3. Model.

3.1. *Cross-classified Nested Nonlinear Model (CRON).* Let the binary variables O_{ik}, C_{ik} and P_{ik} denote the response of customer i at the open, click, and purchase stage, respectively, to promotion k, where $i = 1, \ldots, N$ and $k = 1, \ldots, K$; positive values of O_{ik}, C_{ik}, and P_{ik} indicate conversion at the respective stages. Here, $K = 25$. We develop a conditional hierarchical model with its base at the open stage of the conversion funnel. Consider the following

2-way cross-classified model:

$$\texttt{logit}\{P(O_{ik}=1)\} = \boldsymbol{U}_{ik}^T \boldsymbol{\beta}^{(1)} + f_1(\boldsymbol{r}_{ik}) + \nu_k^{(1)} + b_i^{(1)}, \tag{1}$$

with $\boldsymbol{b}_i^{(1)}$ being the preferences of customer i and $\boldsymbol{\nu}_k^{(1)}$ being the effect of the kth promotion email. The suffix `1` denotes that it is the model for the first stage (open) of the conversion funnel. $\boldsymbol{\beta}^{(1)}$ is invariant across both promotions and customers; the matrix $\boldsymbol{U}$ contains the covariates values corresponding to Age, Income, Frequency, Ave. Retail Spend and Ave. Web Spend. $\boldsymbol{r}_{ik}$ is the 3-dimensional vector of recency variable (last opened, clicked, purchased) that is allowed to have nonlinear effects. In the click stage, note that, $C_{it}=0$ if $O_{it}=0$ and so, we model:

$$\texttt{logit}\{P(C_{ik}=1|O_{ik}=1)\} = \boldsymbol{U}_{ik}^T \boldsymbol{\beta}^{(2)} + f_2(\boldsymbol{r}_{ik}) + \nu_k^{(2)} + b_i^{(2)}. \tag{2}$$

Note that, $P(C_{ik}=0) = P(O_{ik}=0) + (1-P(C_{ik}=1|O_{ik}=1)) \cdot (1-P(O_{ik}=0))$. In this second-stage model, the parameters denote the conditioned (additional) effects, based on the first-stage effects. Similarly, for the purchase stage we consider:

$$\texttt{logit}\{P(P_{ik}=1|C_{ik}=1)\} = \boldsymbol{U}_{ik}^T \boldsymbol{\beta}^{(3)} + f_3(\boldsymbol{r}_{ik}) + \nu_k^{(3)} + b_i^{(3)}. \tag{3}$$

We impose an hierarchical structure on the customer preferences $\boldsymbol{b}_i = (b_i^{(1)}, b_i^{(2)}, b_i^{(3)})$ and consider $\{\boldsymbol{b}_i : 1 \le i \le N\}$ are independent and identically distributed (i.i.d.) random variables from $N(0,\Sigma)$. Thus, the outcomes O_{ik}, C_{ik}, P_{ik} are allowed to be correlated across the different stages of the funnel.

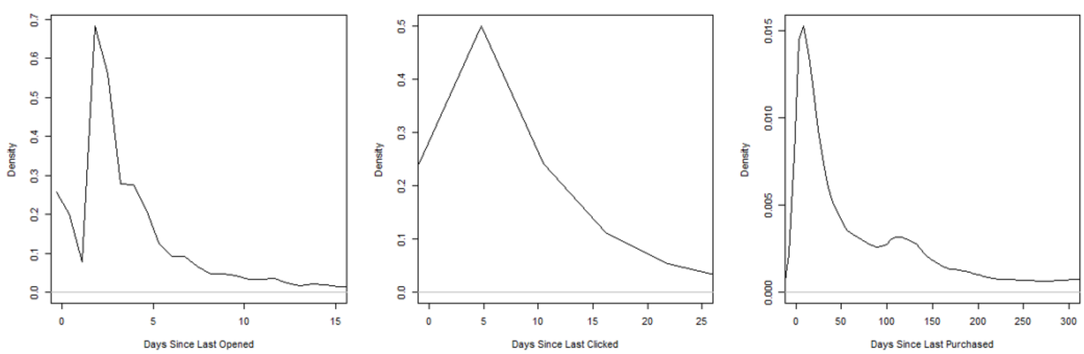


Fig 3: Marginal density of the days since last (a) opened (b) clicked (c) purchased respectively from left to right.

Figure 3 shows the marginal distribution of the recency variables in the data; the marginal densities are multimodal with greatly varying intermodal distances, suggesting that the recency variables might only have nonlinear but also local effects. Thus, we use quadratic splines to model those. f_1, f_2, f_3 are modeled by additive splines $f_j(\boldsymbol{r}_{ik}) = \sum_{l=1}^{3} \pi_{jl} f(r_{ikl}; \boldsymbol{\mu}^{(jl)}, \boldsymbol{t}^{(jl)})$, for $j,\, l = 1,2,3$. The weights π_{jl} are binary; some of them were set to zero, based on model diagnostics to eliminate cases of multicollinearity between the the last opened ($r_{..1}$) and last clicked ($r_{..2}$) covariates. We use quadratic splines in f as:

$$f(x; \boldsymbol{\mu}, \boldsymbol{t}) = \sum_{i=1}^{2} \mu_i x^i + \sum_{j=1}^{m} \mu_{2+j} (x - t_j)_+^2 \text{ where } m = \text{length}(\boldsymbol{t}). \tag{4}$$

The coefficients ($\boldsymbol{\mu}$), the number of knots (m), as well as the location of the knots ($\boldsymbol{t}$) are chosen separately for each stage and each recency variable $r_{..l}$. Henceforth, we call the model in Equations (1)-(4) CRON (Cross-classified Nested Nonlinear model). Table S.6 lists its coefficient estimates.

To understand the impact of local effects of recency variables, and contentwise promotions in CRON, we compare CRON to the following two non-classified (NC) nonlinear submodels: (a) Non-classified Global (NCG) which does not have contentwise promotion, but only global effects of the recency variables based on quadratic polynomials. NCG corresponds to imposing the following constraints $\nu_k^{(j)} = \nu^{(j)}$ for all k and $j = 1, 2, 3$, and $\boldsymbol{t} = \emptyset$ in Equations (1)-(4).
(b) Non-classified Local (NCL) which has non-linear local recency variables, as in CRON, but no contentwise promotion effects. Note, that NCG $\subseteq$ NCL $\subseteq$ CRON. Table 1 presents the key features of these models when compared to existing methods.

TABLE 1
Comparison of SCRON model with existing methods.

Model	Customer ID resolution	Joint Model of conversion funnel	Locally estimated recency effects	Heterogeneous Promo effects based on email content	Varying Promo effects on customer segments
Kamakura and Kang (2007)					
Osinga, Leeflang and Wieringa (2010)	✗	✗	✗	✗	✗
Zhang, Kumar and Cosguner (2017)					
Wu, Li and Liu (2018)	✓	✗	✓	✗	✗
NCG	✓	✗	✗	✗	✗
NCL	✓	✓	✗	✗	✗
CRON	✓	✓	✓	✓	✗
SCON	✓	✓	✓	✓	✓

3.2. *Segmented Cross-classified Nested Nonlinear Model (SCRON).* We next consider an extension of CRON to encompass scenarios where the promotion effects are allowed to vary with customer type. Suppose the customers are pre-classified in $g = 1, \ldots, G$ segments and the promotions effects can vary between different segments. Instead of $\nu_k^{(j)}$ we use $\nu_{kg}^{(j)}$ in Equations (1)-(3). As it contains interaction terms between promotion content and customer types, we call it Segmented Cross-classified Nested Non-linear Model (SCRON).

Due to these interactions, even for moderate values of G the GLMs in each stage of SCRON are very high-dimensional. However, a lot of the promotion effects $\{\nu_{kg}^{(j)} : k = 1, \ldots, K, g = 1, \ldots, G\}$ do not differ significantly from the baseline effect. We use spike-and-slab priors (see Ročková and George, 2018; Ishwaran et al., 2005; Castillo, Schmidt-Hieber and Van der Vaart, 2015, and the references therein) on $\nu_{kg}^{(j)}$. This helps to discard the insignificant promotion effects. Further details are presented in Section 5.4.

4. Estimation: Algorithmic Details and Properties. We next describe the algorithm used for estimating the CRON and the SCRON models. There are several statistical challenges associated with the estimation procedure, as the data set is not only large, needing a scalable algorithm, but also the design is incomplete (missing (i, k) pairs, see Table S.1) and imbalanced. Our proposed algorithm contains several key features which we explain subsequently.

4.1. *Calibrating the Nonlinear Effects.* We use a greedy, marginal method for appropriately calibrating the shape of the nonlinear effects of the recency variables. Substitute $b_i = 0$ and $\nu_k^{(j)}$ for all i, k, j in Equations (1)–(3). We fit splines separately for the open, click, and purchase stages. We illustrate the procedure for the open stage first. Let $\boldsymbol{R}^o = \{R^o_{ik} : 1 \le i \le n, 1 \le k \le K\}$ be all the different values of days since last opened in the data. Correspondingly, let $\boldsymbol{R}^c$ and $\boldsymbol{R}^p$ denote the vectors of days since last clicked and purchased respectively. Consider placing knots at the deciles of $\boldsymbol{R}^o$, $\boldsymbol{R}^c$ and $\boldsymbol{R}^p$. Fit logistic regression model of (1) (with $b_i, \nu_k^{(1)} = 0$) on a subsample of the training data that contains all the conversions and an equal number of randomly chosen non-conversions. This subsampling is done to correct the imbalance (preponderance of non-conversions) in the data (details is presented in subsection 4.2.1 later in a different context). Construct receiver operating characteristic (ROC) curve on the test data based on the above fit. Drop knots iteratively (one at a time) from the center of the data, unless the ROC curve significantly deteriorates. The process is repeated for click and purchase stages with Equations (2) and (3).

4.2. *Block Metropolis-Hastings with Subsampling and Transformations.* We first illustrate the log-likelihood involving of the CRON model. Let $B = \{\boldsymbol{b}_i : i = 1, \dots, N\}$ contain all the latent customer preferences and Θ be the set containing all the other parameters used in Equations (1)–(4). Thus, Θ includes the promotion effects $\boldsymbol{\nu} = \{\nu_k^{(j)}, k = 1, \dots, K; j = 1, 2, 3\}$, the $\boldsymbol{\beta} = \{\boldsymbol{\beta}^{(j)} : j = 1, 2, 3\}$ as well as the spline coefficients $\mathfrak{m} = \{\boldsymbol{\mu}^{(jl)} : j, l = 1, 2, 3\}$. Denote the log-odds ratio $\ell^o_{ik} := \text{logit}(P_{(\Theta,B)}(O_{ik} = 1)$, the conditional log-odds ratios $\ell^c_{ik} = \text{logit}(P_{(\Theta,B)}(C_{ik} = 1|O_{ik} = 1)$ and $\ell^p_{ik} = \text{logit}(P_{(\Theta,B)}(P_{ik} = 1|C_{ik} = 1)$ as given in Equations (1)–(3). Let Ω be the set of (i, k) pairs in our training data $\mathcal{D}$, .i.e., $\Omega = \{(i, k) \in \mathcal{D} : i = 1, \dots, n;\ k = 1, \dots, K\}$. Then, the log-likelihood $l(\Theta, B|\boldsymbol{o}, \boldsymbol{c}, \boldsymbol{p})$ is given by

$$\sum_{(i,k)\in\Omega} \Big[o_{ik}\ell^o_{ik} + c_{ik}\ell^c_{ik} + p_{ik}\ell^p_{ik} - \log(1 + \exp(\ell^o_{ik})) - o_{ik}\log(1 + \exp(\ell^c_{ik})) - c_{ik}\log(1 + \exp(\ell^p_{ik})) \Big] \tag{5}$$

On Θ, we use a independent normal prior $g_1 = N(\boldsymbol{\theta_0}, \text{diag}(\boldsymbol{\sigma}_0^2))$. The hyper-parameters $\boldsymbol{\beta_0}$, $\boldsymbol{\sigma}_0^2$ are set based on GLM estimates fitted separately for the three stages. On B, we use prior $g_2(B|\Sigma) = \prod_{i=1}^n \phi_3(\boldsymbol{b}_i|\boldsymbol{0}, \Sigma)$ where $\phi_3(\ |\boldsymbol{0}, \Sigma)$ is a trivariate normal distribution with mean $\boldsymbol{0}$ and variance Σ. An Inverse-Wishart prior g_3 is imposed on Σ. We generate Markov chain Monte Carlo (MCMC) samples from the logarithm of the posterior density:

$$d(\Theta, B) := l(\Theta, B|\boldsymbol{o}, \boldsymbol{c}, \boldsymbol{p}) + \log g_1(\Theta) + \log g_2(B|\Sigma) + \log g_3(\Sigma). \tag{6}$$

4.2.1. *Subsampling and Imbalance Data.* As issue with sampling directly from $d(\Theta, B)$ is that, in these applications, there are often much lower numbers of conversions (cases) than non-conversions (controls). Although, this marginal imbalance in the data is pervasive in all three stages of the conversion funnel, it is acute in the purchase stage, and $\sum_{(i,k)\in\Omega} p_{ik} / \sum_{(i,k)\in\Omega}(1 - p_{ik})$ as well as $\sum_{(i,k)\in\Omega} p_{ik} / \sum_{(i,k)\in\Omega} c_{ik}$ are usually quite small. Traditional frequentist as well as Bayesian approaches for estimating logistic regression models in imbalanced data sets lead to erroneous results (see Owen, 2007; Johndrow et al., 2019 and the references therein). A popular fix (Anderson, 1972; Prentice and Pyke, 1979; Chawla, Japkowicz and Kotcz, 2004) is to conduct *Case-Control subsampling* (Fithian and Hastie, 2014) so that the ratio of conversions to non-conversions increases in the sample, and imbalance is reduced.

Case-control subsampling is usually conducted by incorporating all the cases and exactly ρ times as many controls for some fixed ρ, such as $\rho = 1, 2, 5$. We consider a similar procedure based on acceptâĂŞreject sampling. Consider the following nested subsampling scheme. Instead of using Ω for model fitting, we use the following nested subsets $\Omega \supseteq \Omega_o \supseteq \Omega_c \supseteq \Omega_p$ for estimation in the open, click, and purchase stages. The case-to-control ratio in the open stage is $1:\rho_1$. Conditioned on an email being opened, the case-to-control ratio in the click stage is $1:\rho_2$, and conditioned on a clicked email the case-to-control ratio in the purchase stage is $1:\rho_3$. The nested subsampling scheme is:

- For all $(i,k) \in \Omega$, generate independent Bernoulli random variable Z^o_{ik} such that
$$P(Z^o_{ik}=1) = o_{ik} + \rho_1(1-o_{ik}) \sum_{(i,k)\in\Omega_0} o_{ik} \bigg/ \Bigg(|\Omega_0| - \sum_{(i,k)\in\Omega_0} o_{ik}\Bigg),$$
where, $|\Omega_0| = \sum_{1\le i\le N; 1\le k\le K} I\{(i,k)\in\Omega_0\}$. Define a sub-sample Ω_o of Ω as $\Omega_o = \{(i,k)\in\Omega : Z^o_{ij}=1\}$. It contains all the conversions. The expected proportion of conversions to non-conversions in the sample is ρ_1.
- For all $(i,k)\in\Omega_o$, generate independent Bernoulli random variable Z^c_{ik} such that $P(Z^c_{ik} = 1) = c_{ik} + \rho_2(1-c_{ik})\sum_{(i,k)\in\Omega_o} c_{ik}/(|\Omega_o| - \sum_{(i,k)\in\Omega_o} c_{ik})$ where $|\Omega_o|$ denotes the cardinality of Ω_o. Define $\Omega_c = \{(i,k)\in\Omega_o : Z^c_{ik}=1\}$.
- For all $(i,k)\in\Omega_c$, generate independent Bernoulli random variable Z^p_{ik} such that $P(Z^p_{ik} = 1) = p_{ik} + \rho_3(1-p_{ik})\sum_{(i,k)\in\Omega_c} p_{ik}/(|\Omega_c| - \sum_{(i,k)\in\Omega_c} p_{ik})$. Define $\Omega_p = \{(i,k)\in\Omega_c : Z^p_{ik}=1\}$.

On the sub-sampled data, the dimension of the B can decrease. Set, $B^S = \{b_i : (i,k)\in\Omega_o \text{ for some } k\}$. Then, in the subsampled data, the logarithm of the posterior density from Equation (6) reduces to:

$$d_S(\Theta, B^S) := l_S(\Theta, B^S) + \log g_1(\Theta) + \log g_2(B^S|\Sigma) + \log g_3(\Sigma), \tag{7}$$

where, the $l_S(\Theta, B^S)$ is the log-likelihood based on the sampled data:

$$\begin{aligned} l_S(\Theta, B^S) = &\sum_{(i,k)\in\Omega_o} o_{ik}\ell^o_{ik} + \sum_{(i,k)\in\Omega_c} c_{ik}\ell^c_{ik} + \sum_{(i,k)\in\Omega_p} p_{ik}\ell^p_{ik} \\ &- |\Omega_o|\log(1+\exp(\ell^o_{ik})) - |\Omega_c|\log(1+\exp(\ell^c_{ik})) - |\Omega_p|\log(1+\exp(\ell^p_{ik}))\,. \end{aligned} \tag{8}$$

Note that $l_S(\Theta, B^S)$ is not an unbiased estimator of $l(\Theta, B|\boldsymbol{o}, \boldsymbol{c}, \boldsymbol{p})$ from Equation (5), and therefore we need to adjust the posterior estimates produced by sampling from $d_S(\Theta, B^S)$ instead of $d(\Theta, B)$. This is done at the end of the Metropolis-Hastings algorithm, below. Also, note that here we have implemented a uniform down sampling of the controls. Fithian and Hastie (2014) showed that logistic regression subsampling policies that adjusts the class balance locally in feature space can outperform uniform down-sampling. However, such policies use proximity measures in the feature space and cannot be directly extended to our application, due to the high-dimensionality of the feature space and joint modeling across 3 nested stages.

4.2.2. *Block Metropolis-Hastings with Additive Transformations (BMHT).* We next provide updates for sampling from the posterior distribution. These updates are done in two major blocks. We initialize Θ Âăby $\hat\Theta_0$ which is selected by running GLM (with b_is set to 0) separately on the three stages in CRON after the Âăknots locations are chosen, based on the prescription in the previous subsection. We set $\hat B_0 = 0$ and present updates for $(\hat\Theta_t, \hat B_t)$ as iteration step t increases.

Block 1: Update Θ keeping B fixed. This block is updated sequentially for the three stages of the conversion funnel. Write $\hat{\Theta}$ as $\{\hat{\Theta}^{(j)} : j = 1, 2, 3\}$ consisting of parameters from the open, click, and purchase stages respectively. Denote $\hat{\Theta}^{(-j)} = \hat{\Theta} \setminus \hat{\Theta}^{(j)}$. Âă Note that, given B, $\hat{\Theta}^{(j)}$ is conditionally independent of $\hat{\Theta}^{(-j)}$ in Equation (6). Thus, at iteration t, separately for each j we update from $\hat{\Theta}_t^{(j)}$ to $\hat{\Theta}_{t+1}^{(j)}$ while $\hat{B}_t$ is fixed. Now, noting that, $g_1(\hat{\Theta}) = g_{11}(\hat{\Theta}^{(j)})g_{12}(\hat{\Theta}^{(j)})g_{13}(\hat{\Theta}^{(j)})$ for $j = 1, 2, 3$:

(a) Draw $\hat{\mathbf{\Theta}}^{(j)}$ by perturbing $\hat{\Theta}_t^{(j)}$ with Gaussian noise, the variance of which is calibrated by looking at mixing in the chain. Then, calculate,

$$r_{t+1}^{(j)} = d_S(\hat{\Theta}^{(j)}, \hat{\Theta}_t^{(-j)}, \hat{B}_t^S) - d_S(\hat{\Theta}_t^{(j)}, \hat{\Theta}_t^{(-j)}, \hat{B}_t^S)$$

(b) Generate U from Uniform$[0, 1]$. If $r_{t+1}^{(j)} \leq U$ then update $\hat{\Theta}_{t+1}^{(j)} = \hat{\Theta}^{(j)}$; else $\hat{\Theta}_{t+1}^{(j)} = \hat{\Theta}_t^{(j)}$.

Block 2: Update customer-specific preferences B keeping Θ fixed As cardinality of B^S is large, convergence of the naive MCMC samplers can take super-linear time (Gao et al., 2017). Unless great care is taken to properly scale the proposal distribution, the random walk Metropolis (RWM) algorithm can have poor convergence properties. As B^S is large, an issue in updating all $\boldsymbol{b}_i \in B^S$ simultaneously using i.i.d. draws from $g_2 \cdot g_3$ is that there is always a positive probability that some coordinate of the multiple random draws will be ill-proposed; in that case, the acceptance ratio will be extremely small, and with high probability will lead to the rejection of the entire move. To enhance the acceptance rate, we implement the Transformation-based Markov Chain Monte Carlo (TMCMC) algorithm (Dutta and Bhattacharya, 2014; Dey et al., 2019) that simultaneously updates all components in B^S using appropriate move types as defined by deterministic transformation of a single draw from $g_2 \cdot g_3$. Generate Σ from g_3 and $\hat{\boldsymbol{b}}$ from $\phi(\cdot|\mathbf{0}, \Sigma)$. Consider $\hat{B}_{t+1}^S = \hat{B}_t^S + \rho Z_i \hat{\boldsymbol{b}}$ where Z_is are i.i.d. from Bernoulli(0.5) and $\rho > 0$ is a fixed small number. Generate U from Uniform$[0, 1]$. Now, if $d_S(\hat{\Theta}_{t+1}, \hat{B}_{t+1}^S) - d_S(\hat{\Theta}_t, \hat{B}_t^S) \leq U$, then $\hat{B}_t^S = \hat{B}_{t+1}^S$.

The scale ρ in the TMCMC step is chosen based on the convergence diagnostics as prescribed in Dey et al. (2019). Papaspiliopoulos, Roberts and Zanella (2020) have shown fast convergence of collapsed samplers in two-way cross-classified setups albeit in a balanced linear model. Geometric ergodicity of TMCMC methods has been established by Dey et al. (2016). Thus, we expect our proposed algorithm to converge. We show that the above-mentioned block Metropolis-Hastings with additive TMCMC modifications (BMHT) converges. The proof, presented in the appendix, follows by showing that BMHT satisfies the conditions in Theorem 2.1 of Dey et al. (2016), which are necessary and sufficient conditions for geometric convergence of the TMCMC induced Markov chains.

We need the following assumptions on the design matrix. Let $X_S^{(j)}$ denote the covariate matrix corresponding to $\Theta^{(j)}$ for the subsample $S = (\Omega_o, \Omega_c, \Omega_p)$. For any fixed K, any subsample S of Σ generated following the scheme in sec. 4.2.1 obeys the conditions:

Assumption 1: the data matrix $X^{(j)}$ has full column rank for all $j = 1, 2, 3$.

Assumption 2: ρ_1, ρ_2, ρ_3 are bounded above by a constant C_0. Also, the imbalance at any stage of the conversion funnel for any promotion does not lead to degeneracy, i.e., there exists $\epsilon > 0$ such that

$$\min_{1 \leq k \leq K} \min\left(\sum_i a_{ik} I\{(i,k) \in A\}, \sum_i (1 - a_{ik}) I\{(i,k) \in A\}\right) > \epsilon|\Omega|$$

holds when o_{ik}, c_{ik} and p_{ik} are substituted for a_{ik}.

These assumptions are very standard in regression setups and can be checked for each subsample. We have the following result on convergence of our proposed algorithm. The proof is provided in supplement sec S.3.

PROPOSITION 1. *Under assumptions A1 and A2, the BMHT Markov chain is geometrically ergodic.*

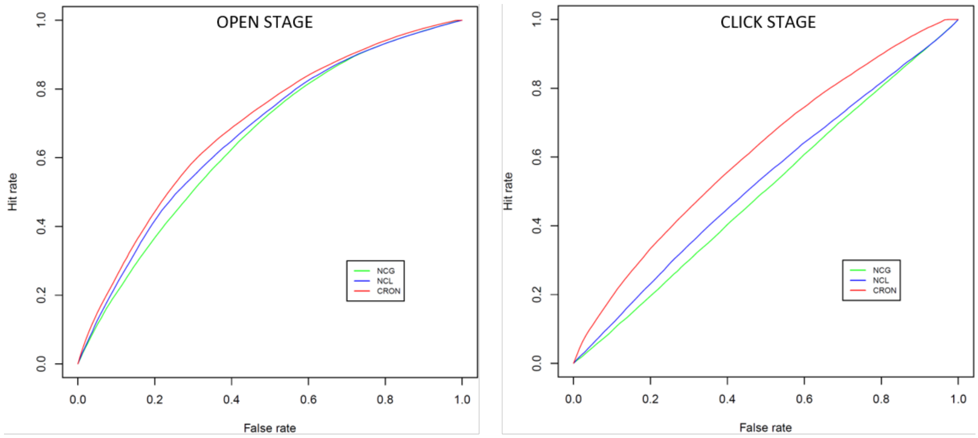
4.2.3. *Adjusting Estimates from BHMT for Case-Control Subsampling.* After the chain has converged, the posterior distribution from the burnt-in chain based on the subsampled data can be converted to a consistent posterior distribution of the model parameter based on the original data, by simply conducting only a location shift to the intercepts $\alpha^{(j)}$: update $\hat{\alpha}^{(j)}$ to $\hat{\alpha}^{(j)} - \log \rho_j$ for $j = 1, 2, 3$. Let $\hat{\Theta}^S$ denote the posterior mean (after correction) and $\hat{\Sigma}^S$ be the estimates for the variance components obtained from BHMT, based on subsample S and satsifying assumptions A1–A2. The plug-in estimate of predictive distribution (Geisser, 1993) using $\hat{\Theta}$ and $\hat{\Sigma}$ converges in probability to the true distribution and can be used for out-of-sample predictions (see supplement sec. S.4).

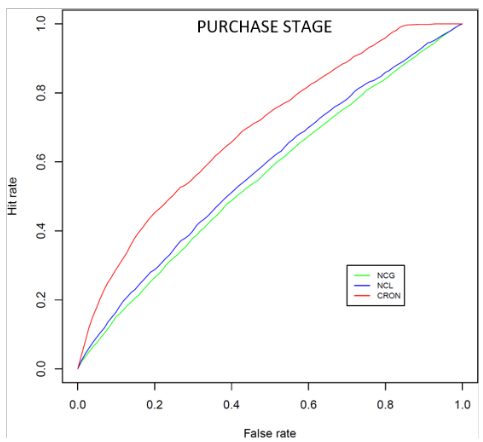
4.3. *Aggregating Inference from Multiple Subsamples.* We repeat the estimation process in sections 4.2.1 and 4.2.2 for L different subsamples by running Monte Carlo algorithms in separate machines. Akin to Scott et al. (2016), after the chains have burnt-in, we combine the draws from the chains based on fractional priors by using weighted averages with each weight being the reciprocal of its marginal posterior variance. A problem with this approach is that the invariant distribution is not preserved across samples. Recent works of Sen et al. (2020); Sachs et al. (2020) develop a novel methodolgy that uses subsampling based unbiased estimates of the gradient of the log-likelihood to arrive at the exact invariant distribution. However, as we have very large sample sizes here, the distributions from the separate machines will be reasonably close and we can expect the combined distribution of the TMCMC output from the subsampled chains, each of which are geoetrically ergodic, to coverge to the target distribution (see Theorem 2.2 of Alquier et al. (2016)). Combining information from different correlated chains can increase estimation efficiency, as eventually more data is used compared to inference from a separate chain. However, note that, if subsampling is used in the open stage, then there are different parameters in B for the individual chains and we conduct a careful tabulation for aggregating results from the different subsampled TMCMC chains.

4.4. *Estimating SCRON with Spike and Slab Prior.* While estimating SCRON, instead of normal prior we impose $\{\nu_{kg}^{(j)} : k = 1, \ldots, K; g = 1, \ldots, G; j = 1, 2, 3\}$ i.i.d. form mixture normal prior $\pi_\eta = (1 - \eta)N(0, \tau_1) + \eta N(0, \tau_2)$ where the mixing weight $\eta \sim \text{Beta}(a_1, a_2)$. The hyper-parameters τ_1 and τ_2 were chosen as random quantities with τ_1 following Inverse Gamma distribution with parameters a_σ and b_σ and $\tau_2 = \tau_1/1000$, respectively. For open stage, these hyperparameters are taken as a_σ = 2.5; b_σ = 35; for click stage a_σ = 2.5; b_σ = 35; and for purchase stage a_σ = 2.5; b_σ = 87.5. The mixing weight a_1 and a_2 are set to be $(3, 2)$ for open stage; $(2, 2)$ for click stage and $(7, 3)$ for purchase stage. Promotions parameters with more than 0.5 probability of belonging to the spiked class were deleted from the model, and the resultant model was refitted again using normal priors on the remaining promotion coefficients to produce the final model estimates. Following Bai, Rockova and George (2020), the estimated model inherits decision theoretic guarantees for consistent variable selection.

5. Results and Business Implications.

5.1. *ROC Curves: Test vs Training Errors.* We fitted the training data using the models described in the previous Section. We set $\rho_3 = 4$ and did not use any subsampling in the open and click stages, for those conversion percentages were not very imbalanced in our data. Figure 4 shows the receiver operations characteristics (ROC) curves for the open, click,





AREA UNDER THE CURVE				
Data Type	Model	Stages		
		Open	Click	Purchase
Test	NCG	0.6550	0.5014	0.5568
	NCL	0.6717	0.5278	0.5769
	CRON	0.6921	0.6143	0.6909
Test	NCG	0.6984	0.5170	0.5808
	NCL	0.6903	0.5456	0.5995
	CRON	0.7054	0.6201	0.7033

Fig 4: ROC curves on the test data based on NCG, NCL, and CRON model estimates from the training set. Starting from top-left, in the subplots, clockwise we have the curves for open, click, and purchase stages respectively. The table in the bottom-right panel presents the area under the curves for the concerned models in the training and the test data.

and purchase stages of the conversion funnel. The ROC shows the plots of the proportion of correct predictions of conversion (hit rate) vs the proportion of incorrect predictions of the no-conversions (false positive rate). Table S.10 in the supplement documents the hit rates from the curves as the false positive proportion is allowed to increase. We observed that the results in the test data are in accordance with those in the training sample. Based on the areas under the ROC curves in the open stage (see Figure 4), we observe NCL to improve on the NCG by 2.5%, and CRON to further improve on NCL by 3.03% in the test data; the improvements are more pronounced in the click and purchase stages, with CRON improving on its nearest competitor by 16.39% and 19.76%, respectively. Tables S.4, S.5, and S.6 present the coefficient estimates from these models.

5.2. *Impact of Promotions on Open Rates.* Table S.6 shows the estimates of the promotions in the CRON model (with respect to the baseline Promo ID 1). We observe that the estimated logodds of opening an email increases around 7 times in the best promotion as compared to the worst, provided the other factors are kept constant. The results show that it is extremely important to accurately identify the hig-performing promotions. The promotion with the highest conditional effect had “Secret Sale" in the subject of the email; it is wellknown in the marketing literature that such keywords produces significant hedonic appeal which leads to consumption impulses (Moore and Lee, 2012). Based on estimates, we observe that the top promotions were mostly non-priced promotions; these had promotion codes for "Cash Refund" or "Buy-One-Get-One Free" in the subject line. The dominance of non-priced promotions, is in accordance with findings that exist in the marketing lietrature (Park, Park and Schweidel, 2018).

To validate the effectiveness of these promotion-effect estimates $\{\hat{\nu}_k^{(1)} : k = 1, \ldots, K\}$ for the open stage, we look at their opening conversion rates in the test data. Based on the estimates from the CRON model (see Table S.6), we further divide the promotions into three baskets of equal sizes containing top-seven, mid-seven and bottom-seven promotions. Table 2 presents their respective conversion rates in the test data. We witness that the rank order of the promotions provided by the estimates is well-maintained in the conversion efficiency in the test data. We find while the top promos produce on average 46% conversions, the mid promos yield 40%, and the bottom promos 36%.

5.3. *Promotion Effects across the Conversion Funnel.* Using the estimated conditional promotion effects for the click and purchase stages of the CRON model from Table S.6, we divide the promotions in the click and purchase stages into three baskets of equal sizes as before containing top, mid, and bottom promotions. On the test data, we find that promotions in the top basket produce on average 48% click conversions among the opened emails and 13% purchases from the clicked emails. The mid basket produces 34% click conversions and 8% purchases, whereas the bottom basket produces 31% and 5%, respectively. Thus, choosing top promotion basket can lead to 17% more clicks on opened emails and 8% more purchases from the clicked emails than the bottom basket. Table 2 presents these detailed conversion results for each promotion on the test data. We observe in this Table that the ordering of the estimated promotion effects from CRON agree with their performance (rank) in the test data.

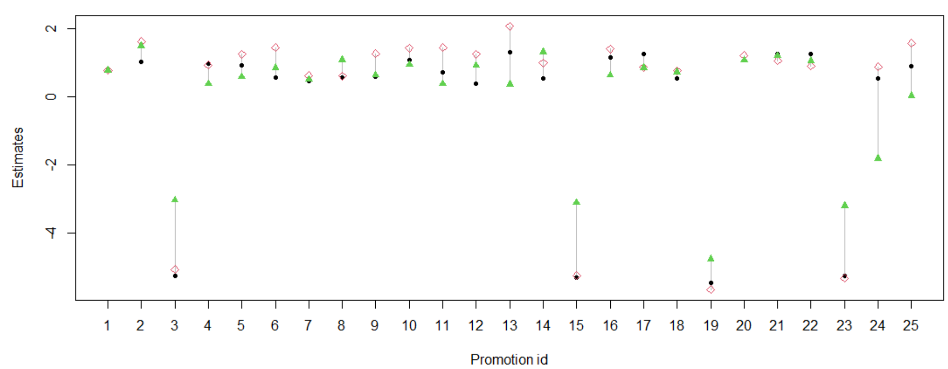


Fig 5: Estimates of promotion effects in the CRON model across open (circles), click (paralleleograms), and purchase (uptriangles) stages plotted with promotion ID 1 being the baseline.

Tables 2 A and B show that the comparative efficiency of any particular promotion can greatly vary across the conversion funnel. Figure 5 plots these estimates. For example, *Promo* 13 (with the subject line "Secret Sale") estimates are the highest in open and click stages, but its conditional impact in the purchase stage is very low (ranked 18 out of 25). Such promotion will be very useful for brand awareness but may not be very effective for purchase coversion. Thus, CRON estimates will be helpful for marketers to determine which promotions to use for different functional purposes.

TABLE 2

Conversion rates of the promotions ranked in the top, mid, and bottom basket categories from the CRON model. Here, No. of Recipients in the click and purchase stages are the counts of customers who opened and clicked the promotions respectively.

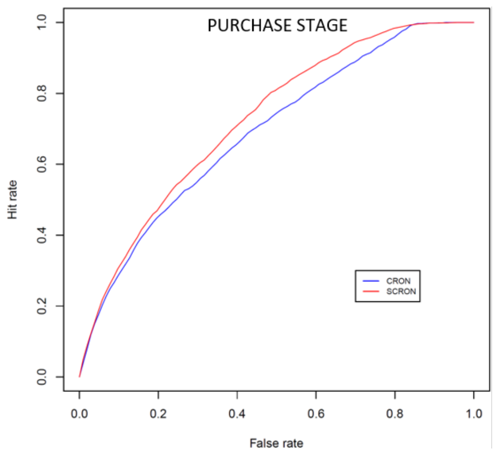
Panel A: Open stage of the conversion funnel

Promo. Efficiency	Open Stage				
	Promo. No.	No. of Recipients	Opened	Open Rate (%)	Ave. Open Rate (%)
Top 7 Promotions	13	9980	5244	0.53	46
	17	5070	2642	52	
	21	9814	3591	37	
	22	9885	5145	52	
	16	9638	4137	43	
	10	2019	817	40	
	20	4659	1866	40	
Mid 7 Promotions	5	7670	3408	44	40
	25	9836	4493	46	
	1	9665	3507	36	
	11	9885	4051	41	
	9	9770	3566	36	
	6	9763	3910	40	
	8	5036	1962	39	
Bottom 7 Promotions	24	6670	2511	38	36
	7	9725	3329	34	
	12	9762	3621	37	
	3	10000	3805	38	
	23	4842	1775	37	
	15	4968	1660	33	
	19	9758	3222	33	

Panel B: Click and Purchase stages of the conversion funnel

Promo. Efficiency	Click Stage					Purchase Stage				
	Promo. No.	No. of Recipients	No. of Clicks	Click Rate (%)	Ave. Click Rate (%)	Promo. No.	No. of Recipients	No. of Purchases	Pur. Rate (%)	Ave. Pur. Rate (%)
Top 7 Promotions	13	5244	3172	60	48	2	2056	327	16	13
	2	4266	2056	48		14	1122	172	15	
	25	4493	2101	47		21	1238	150	12	
	6	3910	1726	44		8	534	46	09	
	11	4051	1790	44		20	712	76	11	
	10	817	367	45		22	1624	162	10	
	16	4137	1790	43		17	829	54	06	
Mid 7 Promotions	5	3408	1329	39	34	6	1726	161	09	08
	20	1866	712	38		1	982	73	07	
	21	3591	1238	34		18	1009	84	09	
	14	3515	1122	32		16	1790	137	08	
	4	4362	1430	33		9	1431	113	08	
	22	5145	1624	32		5	1329	79	06	
	24	2511	762	30		7	857	65	08	
Bottom 7 Promotions	18	3439	1009	29	31	13	3172	177	06	05
	7	3329	857	26		25	2101	90	04	
	8	1962	534	27		24	762	4	01	
	3	3805	1659	44		3	1659	131	08	
	15	1660	535	32		15	535	21	04	
	23	1775	583	33		23	583	39	07	
	19	3222	838	26		19	838	15	02	

5.4. *Variation in Promotion Effects across Customer Types.* Based on traditional marketing practice and without using the customer responses to focal promotions in the current data (Zhang, Kumar and Cosguner, 2017; Gopalakrishnan and Park, 2019), the customers are binned in $G = 3$ groups: low, mid, and highly active types based on a customer's engagement history with the firm. Table S.3 documents the average attributes from these three customer segments. We fit the SCRON model (coefficient estimates are presented in supplemental tables S.7, S.8, S.9) that allows promotion effects to vary not only within but also between these customer segments. While the performance of SCRON with respect to CRON was similar in the open and click stages, the SCRON performed significantly better in the purchase stage and had 4.91% higher area under the curve in the test data as compared to CRON (as shown in Figure 6).



AREA UNDER THE CURVE				
Data Type	Model	Stages		
		Open	Click	Purchase
Test	CRON	0.6921	0.6143	0.6909
	SCRON	0.6923	0.6171	0.7248
Training	CRON	0.7054	0.6201	0.7033
	SCRON	0.7055	0.6222	0.7210

Fig 6: Comparative performance of SCRON and CRON. On left, test-data ROC curves for purchase stage are plotted. The table on the right presents the area under the curves for the concerned models in the training and test datasets.

Targeted marketing aims to design high performing advertisements and promotions suitably tailored for different customer groups. Figure S.2 shows how the promotion-effect estimates vary across the segments. It also shows that the promotion effects from the high engagement segment are consistently higher than those from mid and low segments particularly in the open stage (in the other two stages, we have conditional effects based on the open stage). Based on these promotion estimates from SCRON, top-performing promotions in different customer segments can be listed. Marketers can use these lists for targeted marketing. Next, we present an application demonstrating that significant gains can be achieved by targeted marketing strategies based on the SCRON model.

5.5. *Profitability Analysis using Targeted Promotions.* In this marketing campaign, customers received on average around 21 emails in a very short campaign window (see Table S.1). Marketing research shows that such high frequency of emails can lead to customer fatigue (Statista, 2019). Highly frequent email promotional campaigns often lose relevance and are treated as spam emails. Reducing the number of emails sent to each customers during a campaign is important. We next consider sending only five promotional emails to customers. Based on the promotion-effect estimates from the SCRON model, we construct three different lists of the top five promotions for the low, mid, and highly engaged customer segments. In Table 3A, we report their conversion percentages in the test data. We compare the results to the non-targeted marketing policy that uses a list of top five promotions based on the CRON model, uniformly for all customers. Compared to the unsegmented CRON model, using SCRON we witness marginal improvements in open (1%) and click (3%) stages, and significant gains (29%) in the purchase stage (see Table 3B for details).

6. Discussion. We developed a large-scale, cross-classified, nested joint model for modeling customer responses to opening, clicking, and purchasing from promotion emails, thereby estimating the efficacy of these marketing interventions for customers with varying characteristics and engagement history. We estimate the parameters in the joint model by using a block Metropolis-Hastings algorithm that uses nested subsampling to tackle the severe

TABLE 3
Conversion Rates of SCORN and CORN based promotion campaign with only 5 emails

Panel A: Conversion rates in the test data for SCRON based personalized email campaign and CRON based unsegmented email promotions. The last row shows the lift of the former method based on the later.

Segments	Open Rate (%)	Click Rate (%)	Purchase Rate (%)
SCRON in High Engagement group	48.25	44.56	12.95
SCRON in Mid Engagement group	45.39	38.94	05.32
SCRON in Low Engagement group	44.55	41.99	03.03
SCRON combined across groups	47.19	42.88	10.17
CRON with no segments	46.77	41.68	07.86
Lift in SCRON vs. CRON	00.90	02.88	29.44

Panel B: Detailed conversion rates of the different promotions chosen by the SCRON model in the three different consumer segments and by the CRON model across all consumers. These conversion rates on aggregation produce the table in panel A.

Segment	Promo ID	Count	Open	Click	Purchase	Open Rate (%)	Click Rate (%)	Purchase Rate (%)
SCRON: High Engagement	13	6104	3345	2107	152	55	63	07
	17	2872	1539	489	37	54	32	08
	22	6047	3215	1017	132	53	32	13
	2	5907	2719	1416	314	46	52	22
	21	6009	2181	763	115	36	35	15
SCRON: Mid Engagement	21	2714	1001	335	29	37	34	09
	13	2763	1350	763	20	49	57	03
	22	2736	1384	425	25	51	31	06
	17	1484	739	231	15	50	31	07
	20	1358	544	200	15	40	37	07
SCRON: Low Engagement	21	1091	409	140	06	38	34	04
	16	1023	452	201	05	44	44	02
	13	1113	549	302	05	49	55	02
	20	616	247	100	07	40	40	07
	22	1102	546	182	05	50	33	03
CRON: No Segments	13	9980	5244	3172	177	53	60	06
	17	5070	2642	829	54	52	31	07
	21	9814	3591	1238	150	37	34	12
	22	9885	5145	1624	162	52	32	10
	16	9638	4137	1790	137	43	43	08

imbalance between conversions and no conversions. We implemented standard case-control subsampling. It will be beneficial to use relatively sophisticated subsampling schemes akin to Fithian and Hastie (2014) that preferentially select examples, the responses of which are conditionally rare, given their features. Such schemes will be more immune to model misspecification; however, as the feature space is high dimensional, further exploration is needed to implement such subsampling.

As future work it will be useful to explore whether our estimation algorithm for the logistic-regression-based joint modeling framework in cross-classified models can be made faster by using data-augmentation with Polya-Gamma latent variables, as in Polson, Scott and Windle (2013). In this context, it will be particularly interesting to explore the recent approach in Sen et al. (2020) that uses sub-sample based unbiased estimator of the gradient of the log-likelihood and has the advantange of working with the exact invariant distribution.

We proved geometric ergodicity of our proposed MCMC-based estimation algorithm. Following Bissiri, Holmes and Walker (2016); Kleijn et al. (2012), it will be interesting to introspect posterior concentration of the estimates, particularly under misspecification. Our joint modeling framework do not involve interactions between successive promotional emails as we assumed that the impacts of these emails follows a memoryless process, as theorized in Zhang, Kumar and Cosguner (2017). It will be interesting to estimate the impact of a chain/sequence of promotional emails in future data sets where such sequences, which can be parametrized in our framework by higher order interactions among promotional emails, are estimable.

Acknowledgements. We would like to thank Professor Edward George for numerous stimulating discussions and several helpful comments that greatly improved the quality of the paper. The research was partly supported by NSF grants DMS-1811866.

SUPPLEMENTARY MATERIAL

Supplement to "Estimating Promotion Effects in Email Marketing using a Large-scale Cross-Classified Bayesian Joint Model for Nested Imbalanced Data"
The supplement contains several figures and tables. Their numbers are prefixed by S. Lists of tables and figures are provided in the first two pages of the supplement. The proof of proposition 1 is also presented in the supplement. All R codes developed for conducting the analysis described in the paper are available along with their documentation from the GitHub repository https://github.com/gmukherjee/emarketing.

REFERENCES

ALDIGHIERI, R. (2019). Marketer email tracker 2019. https://dma.org.uk/uploads/misc/marketers-email-tracker-2019.pdf.

ALQUIER, P., FRIEL, N., EVERITT, R. and BOLAND, A. (2016). Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Statistics and Computing* **26** 29–47.

ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59** 19–35.

BAI, R., ROCKOVA, V. and GEORGE, E. I. (2020). Spike-and-Slab Meets LASSO: A Review of the Spike-and-Slab LASSO. *arXiv preprint arXiv:2010.06451*.

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical methodology* **78** 1103.

BULT, J. R. and WANSBEEK, T. (1995). Optimal selection for direct mail. *Marketing Science* **14** 378–394.

CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *The Annals of Statistics* **43** 1986–2018.

CHAWLA, N. V., JAPKOWICZ, N. and KOTCZ, A. (2004). Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter* **6** 1–6.

DEY, K. K., BHATTACHARYA, S. et al. (2016). On geometric ergodicity of additive and multiplicative transformation-based Markov Chain Monte Carlo in high dimensions. *Brazilian Journal of Probability and Statistics* **30** 570–613.

DEY, K. K., BHATTACHARYA, S. et al. (2019). A brief review of optimal scaling of the main MCMC approaches and optimal scaling of additive TMCMC under non-regular cases. *Brazilian Journal of Probability and Statistics* **33** 222–266.

DUTTA, S. and BHATTACHARYA, S. (2014). Markov chain Monte Carlo based on deterministic transformations. *Statistical Methodology* **16** 100–116.

FADER, P. S., HARDIE, B. G. and LEE, K. L. (2005). âĂIJCounting your customersâĂİ the easy way: An alternative to the Pareto/NBD model. *Marketing science* **24** 275–284.

FITHIAN, W. and HASTIE, T. (2014). Local case-control sampling: Efficient subsampling in imbalanced data sets. *Annals of statistics* **42** 1693.

FORTE, D. (2019). Consumers Want More Personalized Emails, But Retailers Aren't Delivering. https://multichannelmerchant.com/marketing/consumers-want-personalized-emails-retailers-arent-delivering/.

GAO, K. (2017). Scalable Estimation and Inference for Massive Linear Mixed Models With Crossed Random Effects, PhD thesis, STANFORD UNIVERSITY.

GAO, K., OWEN, A. et al. (2017). Efficient moment calculations for variance components in large unbalanced crossed random effects models. *Electronic Journal of Statistics* **11** 1235–1296.

GEISSER, S. (1993). *Predictive inference* **55**. CRC press.

GOPALAKRISHNAN, A. and PARK, Y.-H. (2019). The Impact of Coupons on the Visit-to-Purchase Funnel: Theory and Empirical Evidence. *Available at SSRN 3206159*.

HUANG, Z. and GELMAN, A. (2005). Sampling for Bayesian computation with large datasets. *Available at SSRN 1010107*.

ISHWARAN, H., RAO, J. S. et al. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics* **33** 730–773.

JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. (2013). *An introduction to statistical learning* **112**. Springer.

JOHNDROW, J. E., SMITH, A., PILLAI, N. and DUNSON, D. B. (2019). MCMC for imbalanced categorical data. *Journal of the American Statistical Association* **114** 1394–1403.

KAMAKURA, W. A. and KANG, W. (2007). Chain-wide and store-level analysis for cross-category management. *Journal of Retailing* **83** 159–170.

KAPNER, S. (2017). Retailers' Emails Are Misfires for Many Holiday Shoppers. https://www.wsj.com/articles/retailers-emails-are-misfires-for-many-holiday-shoppers-1511778600.

KLEIJN, B. J. K., VAN DER VAART, A. W. et al. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics* **6** 354–381.

KUHN, M., JOHNSON, K. et al. (2013). *Applied predictive modeling* **26**. Springer.

LEE (2019). Digital vs Traditional Media âĂŞ Where Should You Invest Your Marketing Dollars? https://hookdpromotions.com/digital-vs-traditional-media-where-should-you-invest-your-marketing-dollars/.

LEONE, C. (2020). How Much Should You Budget For Marketing In 2020.

MCCULLOCH, C. (2008). Joint modelling of mixed outcome types using latent variables. *Statistical methods in medical research* **17** 53–73.

MOORE, D. J. and LEE, S. P. (2012). How advertising influences consumption impulses. *Journal of Advertising* **41** 107–120.

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** 173–185.

OWEN, A. B. (2007). Infinitely imbalanced logistic regression. *Journal of Machine Learning Research* **8** 761–773.

PAPASPILIOPOULOS, O., ROBERTS, G. O. and ZANELLA, G. (2020). Scalable inference for crossed random effects models. *Biometrika* **107** 25–40.

PARK, C. H., PARK, Y.-H. and SCHWEIDEL, D. A. (2018). The effects of mobile promotions on customer purchase dynamics. *International Journal of Research in Marketing* **35** 453–470.

PERRIN, N. (2019). Email Marketing 2019: Still a Leading Touchpoint for Marketers and Consumers Alike. https://content-na1.emarketer.com/email-marketing-2019.

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association* **108** 1339–1349.

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411.

RIZOPOULOS, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press.

RIZOPOULOS, D. and GHOSH, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine* **30** 1366–1380.

RIZOPOULOS, D. and LESAFFRE, E. (2014). Introduction to the special issue on joint modelling techniques. *Statistical methods in medical research* **23** 3–10.

ROSSI, P. E., MCCULLOCH, R. E. and ALLENBY, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science* **15** 321–340.

ROČKOVÁ, V. and GEORGE, E. I. (2018). The spike-and-slab lasso. *Journal of the American Statistical Association* **113** 431–444.

RUPPERT, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics* **11** 735–757.

SACHS, M., SEN, D., LU, J. and DUNSON, D. (2020). Posterior computation with the gibbs zig-zag sampler. *arXiv preprint arXiv:2004.04254*.

SAHNI, N. S., WHEELER, S. C. and CHINTAGUNTA, P. (2018). Personalization in email marketing: The role of noninformative advertising content. *Marketing Science* **37** 236–258.

SAHNI, N. S., ZOU, D. and CHINTAGUNTA, P. K. (2017). Do targeted discount offers serve as advertising? Evidence from 70 field experiments. *Management Science* **63** 2688–2705.

SCOTT, S. L., BLOCKER, A. W., BONASSI, F. V., CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management* **11** 78–88.

SEETHARAMAN, P. and CHINTAGUNTA, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business & Economic Statistics* **21** 368–382.

SEN, D., SACHS, M., LU, J. and DUNSON, D. B. (2020). Efficient posterior sampling for high-dimensional imbalanced logistic regression. *Biometrika* **107** 1005–1012.

STATISTA (2019). Dossier on e-mail marketing in the U.S. and Worldwide.

WANG, H., ZHU, R. and MA, P. (2018). Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association* **113** 829–844.

WU, J., LI, K. J. and LIU, J. S. (2018). Bayesian inference for assessing effects of email marketing campaigns. *Journal of Business & Economic Statistics* **36** 253–266.

ZHANG, X., KUMAR, V. and COSGUNER, K. (2017). Dynamically managing a profitable email marketing program. *Journal of marketing research* **54** 851–866.

ZHANG, X., ZHOU, Y., MA, Y., CHEN, B.-C., ZHANG, L. and AGARWAL, D. (2016). Glmix: Generalized linear mixed models for large-scale response prediction. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 363–372.

Supplement to "Estimating Promotion Effects in Email Marketing using a Large-scale Cross-Classified Bayesian Joint Model for Nested Imbalanced Data"

Sabyasachi Mukhopadhyay, Indian Institute of Management
Wreetabrata Kar, Purdue University
Gourab Mukherjee, University of Southern california.

The supplement contains several figures and tables that are referenced in the main paper. We first present an exhaustive list of tables and figures spread across the main paper and the supplement. The tables and figures in the supplement (whose numbers are prefixed by S) are presented in the order in which they are referenced in the main paper. The page numbers in the supplementary materials are prefixed by `S`. The proofs of the results stated in the main paper are also presented here.

List of Tables

List of Figures

S.1 Data Summary

Table S.1: Distribution of the number of promotion emails received by customers during the concerned campaign.

No of Promotions	**No.of Recipients**
4	2
5	4
6	18
7	18
8	25
9	40
10	66
11	191
12	392
13	668
14	713
15	450
16	363
17	619
18	2196
19	7034
20	15127
21	20518
22	17631
23	9160
24	2461
25	290
Total	77986

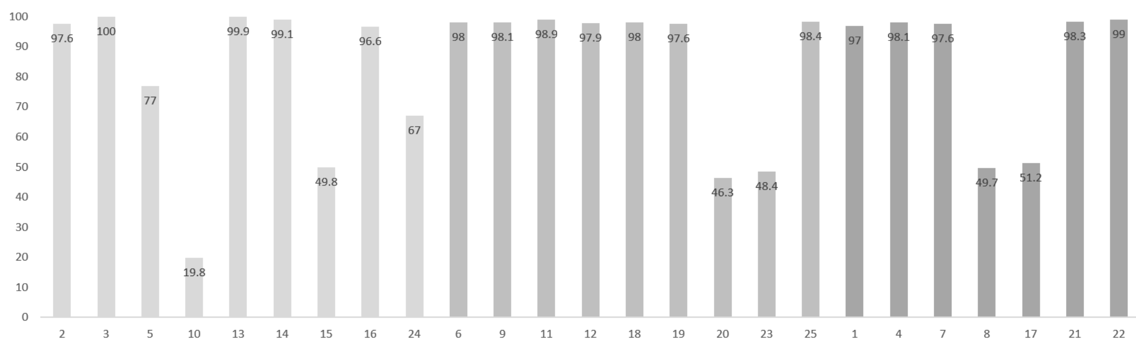


Figure S.1: Percentage of customers reached by different promotion emails. The promotion number are in congruence with those in table S.2. In light gray, black and dark gray respectively are 9 priced (in lightgray), 7 non-priced (in darkgray) and 9 combination (in gray) promotions.

Table S.2: Aggregate statistics averaged across recipients for different promotion emails used in the campaign

Promotions			**Responses**			**Days: Last Open**		**Days: Last Clicked**		**Days: Last Purchased**		**Past Order Count**		**AOV Retail**		**AOV Web**	
ID No	**Type**	**Count**	**OpenRate**	**ClickRate**	**PurchaseRate**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
1	N	75661	0.3589	0.1031	0.0084	9.73	20.41	27.34	91.31	93.38	102.58	7.69	8.63	13.96	30.36	52.08	33.82
2	P	76103	0.4343	0.2113	0.0326	9.35	20.31	26.66	90.06	92.53	101.49	7.70	8.64	13.96	30.34	52.07	33.81
3	P	77986	0.3828	0.1623	0.0132	0.81	8.20	3.67	35.80	68.56	89.06	7.83	8.66	13.92	30.26	52.30	33.50
4	N	76492	0.4451	0.1449	0.0084	8.45	19.25	23.55	84.52	87.61	88.99	7.69	8.63	13.96	30.34	52.04	33.81
5	P	60048	0.4448	0.1742	0.0119	6.56	15.14	17.19	62.24	85.82	90.67	7.79	8.64	14.04	30.11	52.30	33.75
6	C	76398	0.3971	0.1770	0.0157	5.80	14.11	14.46	53.80	84.26	90.84	7.79	8.66	14.02	30.36	52.27	33.70
7	N	76136	0.3404	0.0881	0.0057	6.98	16.04	18.25	66.89	86.30	90.29	7.78	8.64	14.01	30.35	52.28	33.70
8	N	38782	0.3843	0.1009	0.0105	7.35	17.74	20.61	77.11	88.83	89.60	7.70	8.71	13.81	30.08	52.28	34.07
9	C	76470	0.3666	0.1475	0.0103	8.12	18.39	22.71	79.84	87.47	89.20	7.69	8.63	13.96	30.34	52.04	33.81
10	P	15432	0.4128	0.1812	0.0169	8.70	18.67	21.84	66.88	90.81	90.11	7.68	8.54	14.00	30.77	52.17	32.73
11	C	77140	0.4064	0.1783	0.0100	5.05	9.81	10.15	37.07	79.44	87.85	7.87	8.68	13.93	30.27	52.34	33.43
12	C	76385	0.3685	0.1472	0.0136	5.49	13.51	13.43	52.29	85.25	90.83	7.80	8.66	14.02	30.36	52.27	33.70
13	P	77874	0.5260	0.3129	0.0177	5.60	12.18	12.21	46.89	83.68	90.50	7.83	8.65	13.92	30.26	52.32	33.50
14	P	77323	0.3555	0.1169	0.0159	5.59	11.82	12.21	45.54	83.23	90.60	7.83	8.66	13.92	30.27	52.32	33.49
15	P	38808	0.3326	0.1066	0.0045	0.85	5.80	2.81	17.94	70.61	87.68	8.12	8.80	13.99	30.39	52.73	33.54
16	P	75321	0.4220	0.1780	0.0134	6.06	9.64	11.02	36.40	80.79	87.64	7.95	8.66	13.99	30.36	52.50	33.18
17	N	39951	0.5214	0.1573	0.0126	5.04	7.64	9.41	24.13	90.35	96.80	6.08	6.50	12.31	28.41	50.77	32.94
18	C	76403	0.3508	0.0993	0.0077	5.26	9.18	9.74	32.20	78.28	87.30	7.99	8.72	13.98	30.33	52.47	33.14
19	C	76145	0.3293	0.0883	0.0013	0.31	1.65	1.32	3.30	65.46	90.01	8.16	8.80	13.97	30.16	52.32	32.78
20	C	36076	0.3927	0.1447	0.0150	5.83	6.10	10.17	13.77	87.01	96.07	6.72	7.79	12.12	28.30	51.62	33.63
21	N	76633	0.3696	0.1226	0.0145	6.62	5.74	10.81	11.26	79.85	91.22	8.06	8.75	13.98	30.26	52.43	33.10
22	N	77177	0.5212	0.1621	0.0176	4.90	7.11	8.96	21.06	80.88	90.96	7.97	8.71	13.97	30.30	52.42	33.21
23	C	37755	0.3575	0.1119	0.0073	0.62	2.60	2.64	5.06	65.30	86.60	8.30	8.82	13.98	29.82	52.45	32.55
24	P	52242	0.3780	0.1136	0.0007	4.56	6.46	8.94	13.19	82.67	92.16	6.89	7.19	14.93	31.40	53.78	34.12
25	C	76761	0.4546	0.2103	0.0087	4.54	6.10	8.70	12.15	79.57	91.13	8.05	8.74	13.98	30.25	52.41	33.07

Table S.3: Summary statistics corresponding to the three different segments of customers

Type	**ID_Type**	**Count**	**Recipients**	**OpenRate**	**ClickRate**	**PurchaseRate**	**Last Open**		**Last Clicked**		**Last Purchased**		**Past Order Count**		**AOV Retail**		**AOV Web**	
							Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Low	Non Price	7	51686	0.3871	0.1164	0.0041	10.25	25.15	26.02	74.31	270.19	85.21	3.38	2.98	10.46	27.26	49.37	36.57
	Price	9	60125	0.3855	0.1559	0.0041	7.85	22.85	20.51	66.52	268.34	83.15	3.39	2.97	10.64	27.61	49.56	36.62
	Combined	9	66218	0.3695	0.1443	0.0039	6.81	19.88	16.90	57.09	262.22	94.61	3.40	2.97	10.48	27.20	49.47	36.52
Mid	Non Price	7	128000	0.3974	0.1173	0.0057	8.42	16.31	21.17	71.34	118.43	48.42	5.63	5.15	12.84	29.78	53.05	36.24
	Price	9	152399	0.3979	0.1612	0.0051	6.36	14.75	16.21	61.53	117.58	48.07	5.63	5.14	13.01	30.01	53.33	36.46
	Combined	9	152399	0.3812	0.1446	0.0051	5.63	12.91	13.59	51.67	116.05	54.46	5.67	5.16	12.80	29.83	53.21	36.22
High	Non Price	7	281146	0.4287	0.1304	0.0147	5.97	10.97	13.87	56.32	37.81	52.61	9.39	9.86	14.88	30.75	52.21	31.57
	Price	9	338613	0.4284	0.1918	0.0215	4.31	8.96	9.69	43.77	32.45	47.52	9.48	9.84	15.13	30.99	52.60	31.56
	Combined	9	374612	0.3832	0.1488	0.0129	3.95	7.63	8.59	35.53	30.71	37.61	9.64	9.98	14.95	30.73	52.34	31.31

S.2 Results

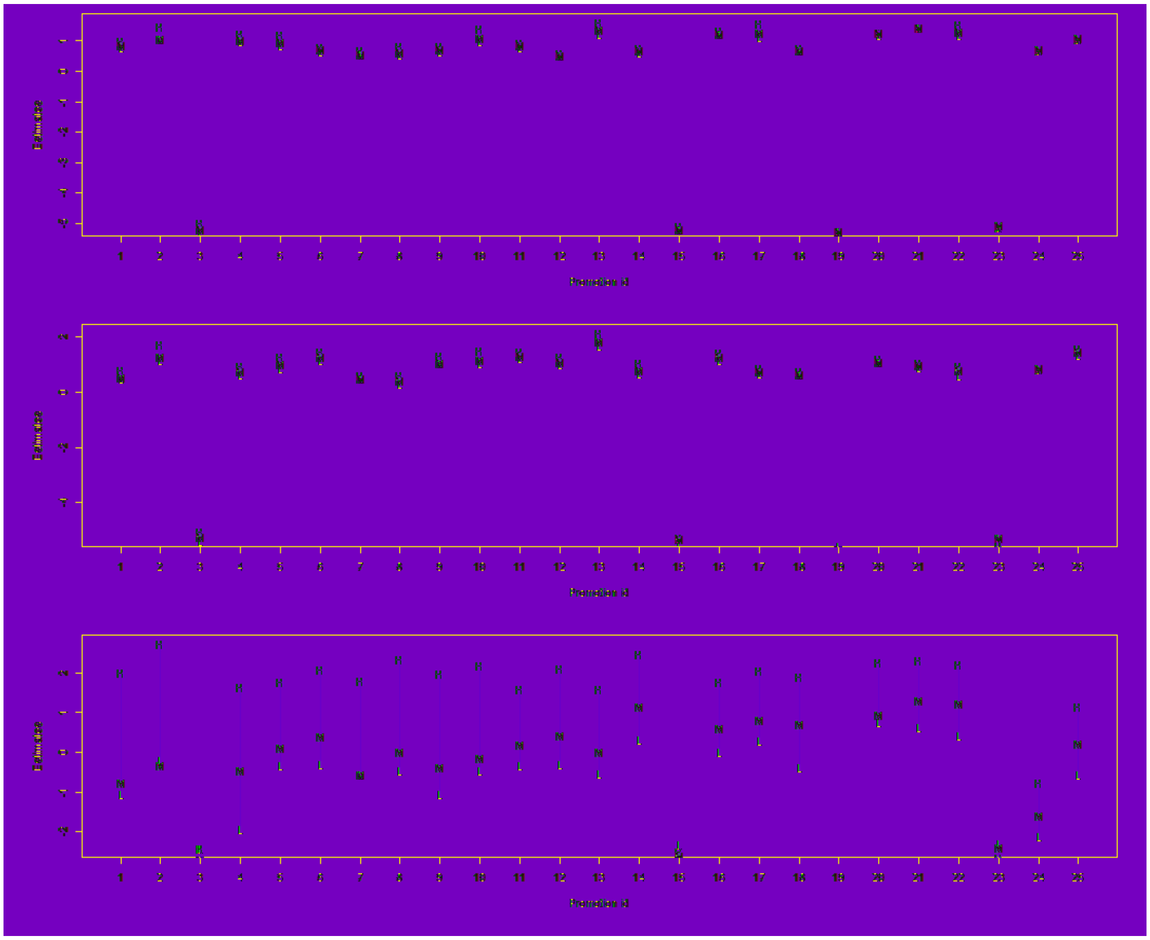


Figure S.2: Plot of estimates of the promotion effects from SCRON model. In the subplots from top to bottom, we have estimates from the open, click, and purchase stages respectively. H, M, and L, respectively, denote the high, mid, and low engaged customer segments.

Table S.4: Stage-wise model estimates of NCG

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	-0.8818	0.0014	-1.7628	0.0033	-4.4131	0.0053
Ave. Retail Spend	0.0008	0.0001	0.0007	0.0002	0.0031	0.0004
Ave. Web Spend	-0.0003	0.0001	0.0017	0.0004	0.0144	0.0007
Past Pur. Freq.	0.0041	0.0003	0.0161	0.0012	0.0611	0.0027
Age	0.0865	0.0019	0.1038	0.0171	0.1676	0.0284
Income: 15K - 25K	-0.0024	0.0008	0.0006	0.0016	-0.0122	0.0018
Income: 25K - 35K	0.0131	0.0014	-0.0281	0.0012	-0.1038	0.0018
Income: 35K - 50K	0.0239	0.0028	-0.0342	0.0016	-0.0308	0.0014
Income: 50K - 75K	0.0307	0.0013	-0.044	0.0022	-0.0723	0.002
Income: 75K - 100K	0.0552	0.0010	-0.0664	0.0011	-0.1107	0.0015
Income: 100K - 120K	0.0728	0.001	-0.0686	0.0017	-0.1203	0.0029
Income: 120K - 149K	0.0893	0.0009	-0.0689	0.0006	0.019	0.0012
Income: 150K plus	0.1156	0.0018	-0.0809	0.0028	0.0832	0.0013
Days Since Last Opened	0.3576	0.0011				
Days Since Last Opened^2	-0.2961	0.0045				
Days Since Last Clicked			-0.0765	0.0064		
Days Since Last Clicked^2			0.0322	0.002		
Days Since Last Purchased	0.1443	0.0026	0.1228	0.014	-0.2327	0.0095
Days Since Last Purchased^2	-0.0147	0.0005	0.006	0.005	0.0725	0.0048

Table S.5: Stage-wise model estimates of NCL

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	-0.8212	0.0028	-1.7773	0.0013	-4.2676	0.0049
Ave. Retail Spend	0.0006	0.0002	0.0004	0.0002	0.0009	0.0004
Ave. Web Spend	-0.0003	0.0002	0.0007	0.0002	0.0092	0.0005
Past Pur. Freq.	0.0032	0.0008	0.0124	0.0006	0.0423	0.002
Age	0.0813	0.0012	0.0739	0.0051	0.2916	0.0228
Income: 15K - 25K	0.0026	0.0022	-0.0046	0.0016	-0.0367	0.0038
Income: 25K - 35K	0.024	0.002	-0.0225	0.0016	-0.1065	0.002
Income: 35K - 50K	0.0272	0.0023	-0.0422	0.0008	-0.0226	0.0029
Income: 50K - 75K	0.0385	0.0035	-0.0454	0.0009	-0.0595	0.0015
Income: 75K - 100K	0.0518	0.002	-0.0617	0.001	-0.1378	0.002
Income: 100K - 120K	0.073	0.0016	-0.0671	0.0026	-0.1319	0.002
Income: 120K - 149K	0.0733	0.002	-0.0775	0.0013	0.0068	0.0042
Income: 150K plus	0.0978	0.0019	-0.0726	0.0012	0.0882	0.0013
Days Since Last Opened	1.6208	0.0049				
Days Since Last Opened^2	-0.6053	0.007				
Open: Knot 1	-1.4695	0.0052				
Open: Knot 2	2.9845	0.0074				
Open: Knot 3	-1.0986	0.0044				
Open: Knot 4	0.6465	0.0048				
Open: Knot 5	-1.0992	0.0046				
Open: Knot 6	0.8947	0.0039				
Days Since Last Clicked			-0.7544	0.0069		
Days Since Last Clicked^2			1.3947	0.0047		
Click: Knot 1			-2.584	0.0044		
Click: Knot 2			1.3811	0.0094		
Click: Knot 3			0.4162	0.0043		
Click: Knot 4			-1.231	0.0063		
Click: Knot 5			0.4002	0.0067		
Click: Knot 6			0.2666	0.0039		
Days Since Last Purchased	-0.5232	0.0031	0.0656	0.0088	-1.6554	0.0162
Days Since Last Purchased^2	0.0706	0.0052	-0.244	0.0049	1.5163	0.0096
Purchase: Knot 1	0.4417	0.006	0.8402	0.0102	-1.5476	0.0216
Purchase: Knot 2	-0.6094	0.0056	-0.7288	0.0086	-0.1684	0.0136
Purchase: Knot 3	0.1934	0.0034	-0.0431	0.0054	0.2146	0.0141
Purchase: Knot 4	-0.1801	0.0068	0.2862	0.0101	1.4211	0.0213
Purchase: Knot 5	0.0346	0.0064	-0.4838	0.01	-4.764	0.0151
Purchase: Knot 6	0.0358	0.0072	0.3721	0.0136	3.9507	0.0045

Table S.6: Stage-wise model estimates of CRON

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	5.2852	0.0012	5.9916	0.0006	2.4787	0.0027
Ave. Retail Spend	0.0007	0.0001	-0.0003	1.00E-04	-0.0006	0.0002
Ave. Web Spend	-0.0003	1.00E-04	-0.0007	1.00E-04	0.0034	0.0002
Past Pur. Freq.	0.0035	0.0005	0.006	0.0003	0.0233	0.0005
Age	0.0797	0.0009	-0.036	0.0008	0.0877	0.0016
Income: 15K - 25K	-0.0078	0.0022	-0.0041	0.0017	-0.0877	0.0036
Income: 25K - 35K	0.0222	0.0008	-0.0621	0.0011	-0.0433	0.0014
Income: 35K - 50K	0.0346	0.0014	-0.0857	0.0009	0.0005	0.0019
Income: 50K - 75K	0.031	0.0014	-0.1053	0.0006	0.0002	0.002
Income: 75K - 100K	0.0581	0.0007	-0.1453	0.0008	0.039	0.0022
Income: 100K - 120K	0.0812	0.0006	-0.1805	0.001	0.0464	0.0031
Income: 120K - 149K	0.0806	0.0007	-0.196	0.0009	0.0936	0.0048
Income: 150K plus	0.1052	0.0006	-0.2229	0.0008	0.0845	0.0049
Days Since Last Opened	-14.62	0.0016				
Days Since Last Opened^2	10.6154	0.0028				
Open: Knot 1	-12.3773	0.0016				
Open: Knot 2	2.8874	0.002				
Open: Knot 3	-1.7181	0.0015				
Open: Knot 4	1.3061	0.002				
Open: Knot 5	-1.5721	0.0017				
Open: Knot 6	1.1711	0.002				
Days Since Last Clicked			-17.1054	0.0015		
Days Since Last Clicked^2			11.8874	0.0011		
Click: Knot 1			-11.6616	0.0014		
Click: Knot 2			-0.1183	0.0009		
Click: Knot 3			0.5473	0.0012		
Click: Knot 4			-1.4043	0.0012		
Click: Knot 5			0.762	0.0017		
Click: Knot 6			-0.0432	0.0011		
Days Since Last Purchased	-0.4596	0.0025	-0.3902	0.0013	-13.9191	0.002
Days Since Last Purchased^2	-0.0439	0.0008	0.0298	0.0015	10.091	0.0015
Purchase: Knot 1	0.6753	0.0011	0.4361	0.0012	-10.1418	0.0024
Purchase: Knot 2	-0.7802	0.0007	-0.6298	0.0033	-0.0325	0.002
Purchase: Knot 3	0.3035	0.0022	0.1602	0.0013	0.1334	0.0018
Purchase: Knot 4	-0.2311	0.001	0.1646	0.0009	0.4342	0.0019
Purchase: Knot 5	0.0058	0.0011	-0.4595	0.0015	-1.8808	0.0044
Purchase: Knot 6	0.06	0.0014	0.3763	0.0012	1.8666	0.0011
Promo 2	0.2537	0.0025	0.8472	0.0007	0.7154	0.0012
Promo 3	-6.0246	0.001	-5.8443	0.0012	-3.8129	0.0012
Promo 4	0.1981	0.0009	0.1563	0.0011	-0.389	0.0022
Promo 5	0.1501	0.0011	0.4743	0.0008	-0.1915	0.0016
Promo 6	-0.2022	0.001	0.6781	0.0006	0.0814	0.0012
Promo 7	-0.304	0.0015	-0.1519	0.0008	-0.2666	0.0012
Promo 8	-0.2167	0.0009	-0.1637	0.0006	0.3096	0.0011
Promo 9	-0.1851	0.0013	0.4961	0.0014	-0.1414	0.0017
Promo 10	0.318	0.0019	0.6597	0.0005	0.165	0.0025
Promo 11	-0.0628	0.0011	0.6766	0.0022	-0.3893	0.0023
Promo 12	-0.3803	0.001	0.4773	0.0008	0.1363	0.0013
Promo 13	0.5475	0.001	1.3002	0.0011	-0.4129	0.0018
Promo 14	-0.2244	0.0011	0.2209	0.0015	0.5437	0.0016
Promo 15	-6.0804	0.0014	-6.0259	0.0008	-3.8896	0.0008
Promo 16	0.3712	0.0012	0.6348	0.0006	-0.1371	0.002
Promo 17	0.4937	0.0016	0.0979	0.0011	0.082	0.0025
Promo 18	-0.2202	0.0009	-0.0051	0.0022	-0.0492	0.0034
Promo 19	-6.2357	0.001	-6.4296	0.0011	-5.5452	0.0018
Promo 20	0.3035	0.0008	0.4395	0.0007	0.3001	0.0014
Promo 21	0.4908	0.0014	0.2896	0.0012	0.4204	0.0014
Promo 22	0.4742	0.0013	0.1394	0.0009	0.2866	0.002
Promo 23	-6.0384	0.0009	-6.0974	0.0015	-3.975	0.0033
Promo 24	-0.2334	0.0024	0.1113	0.0007	-2.5912	0.0013
Promo 25	0.1151	0.0017	0.7958	0.0008	-0.7408	0.0014

Table S.7: Stage-wise model estimates of SCRON for low engagement customers

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	5.1055	0.0023	5.7398	0.001	0.0189	0.0024
Ave. Retail Spend	0.0006	0.0002	-0.0003	0.0002	-0.0007	0.0003
Ave. Web Spend	-0.0002	0.0002	-0.0007	0.0002	0.004	0.0003
Past Pur. Freq.	0.0032	0.0006	0.0056	0.0008	0.0236	0.0009
Age	0.0759	0.0012	-0.0354	0.0009	0.1212	0.0027
Income: 15K - 25K	-0.0026	0.0018	-0.0102	0.0015	-0.0563	0.0012
Income: 25K - 35K	0.0204	0.0017	-0.0647	0.0019	-0.0461	0.0029
Income: 35K - 50K	0.0294	0.0014	-0.0862	0.0007	0.0133	0.0015
Income: 50K - 75K	0.0415	0.002	-0.098	0.0023	0.0229	0.0029
Income: 75K - 100K	0.0576	0.0024	-0.1431	0.0021	0.071	0.0009
Income: 100K - 120K	0.0744	0.0019	-0.176	0.001	0.0266	0.0012
Income: 120K - 149K	0.0823	0.0018	-0.1888	0.0015	0.1005	0.0015
Income: 150K plus	0.0999	0.0009	-0.2264	0.0006	0.0806	0.0015
Days Since Last Opened	-14.6378	0.0092				
Days Since Last Opened^2	10.6483	0.0114				
Open: Knot 1	-12.4159	0.0055				
Open: Knot 2	2.8864	0.0068				
Open: Knot 3	-1.7323	0.01				
Open: Knot 4	1.3113	0.0092				
Open: Knot 5	-1.5789	0.0065				
Open: Knot 6	1.1663	0.0058				
Days Since Last Clicked			-17.101	0.0097		
Days Since Last Clicked^2			11.8783	0.0059		
Click: Knot 1			-11.6513	0.0079		
Click: Knot 2			-0.1214	0.0029		
Click: Knot 3			0.5692	0.0042		
Click: Knot 4			-1.3804	0.0021		
Click: Knot 5			0.6875	0.0056		
Click: Knot 6			-0.0211	0.006		
Days Since Last Purchased	-0.4478	0.0053	-0.4946	0.0043	-15.3012	0.0258
Days Since Last Purchased^2	-0.0714	0.0038	0.0923	0.0029	11.1235	0.0158
Purchase: Knot 1	0.7259	0.0063	0.3957	0.0031	-11.4026	0.0243
Purchase: Knot 2	-0.8141	0.0049	-0.6642	0.0049	-0.1895	0.0193
Purchase: Knot 3	0.3235	0.0054	0.2242	0.0039	-0.0839	0.0043
Purchase: Knot 4	-0.2684	0.0151	0.0393	0.0059	4.6886	0.0133
Purchase: Knot 5	0.2064	0.0033	-0.137	0.0059	-7.8951	0.0356
Purchase: Knot 6	-0.1359	0.0046	0.09	0.0037	3.7391	0.0189
Promo 2	0.2631	0.0028	0.6849	0.0007	0.8665	0.0011
Promo 3	-6.0722	0.0027	-5.87	0.0012	-1.3692	0.0011
Promo 4	0.2027	0.0014	0.1776	0.001	-0.8826	0.0021
Promo 5	0.0641	0.0015	0.3807	0.0012	0.7319	0.0012
Promo 6	-0.1297	0.0021	0.6755	0.001	0.7572	0.0009
Promo 7	-0.2311	0.0011	0	0	0.5025	0.0011
Promo 8	-0.2476	0.0016	-0.1639	0.0011	0.608	0.0025
Promo 9	-0.1492	0.0009	0.5648	0.0016	0	0
Promo 10	0.199	0.0027	0.5603	0.001	0.5918	0.0026
Promo 11	0	0	0.7477	0.0013	0.734	0.0034
Promo 12	-0.2903	0.0014	0.5212	0.0015	0.7696	0.001
Promo 13	0.4271	0.0014	1.1998	0.0009	0.526	0.0013
Promo 14	-0.1899	0.0014	0.2088	0.001	1.3859	0.0008
Promo 15	-5.9961	0.0014	-5.9491	0.0015	-1.2487	0.0014
Promo 16	0.4458	0.0012	0.6753	0.001	1.0777	0.0015
Promo 17	0.331	0.0018	0.204	0.0012	1.3672	0.0027
Promo 18	-0.0938	0.0015	0.1729	0.0006	0.6785	0.0045
Promo 19	-6.0752	0.0009	-6.04	0.0005	-3.179	0.001
Promo 20	0.4016	0.0028	0.5986	0.0013	1.8092	0.0012
Promo 21	0.6212	0.0018	0.4186	0.0017	1.7045	0.0014
Promo 22	0.4019	0.0038	0.1374	0.001	1.4751	0.0032
Promo 23	-5.9312	0.001	-5.7605	0.0011	-1.2423	0.0028
Promo 24	-0.1178	0.0014	0.3373	0.0007	-1.0473	0.0014
Promo 25	0.2477	0.0017	0.8613	0.0009	0.5016	0.0019

Table S.8: Stage-wise model estimates of SCRON for mid engagement customers

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	5.1055	0.0023	5.7398	0.001	0.0189	0.0024
Ave. Retail Spend	0.0006	0.0002	-0.0003	0.0002	-0.0007	0.0003
Ave. Web Spend	-0.0002	0.0002	-0.0007	0.0002	0.004	0.0003
Past Pur. Freq.	0.0032	0.0006	0.0056	0.0008	0.0236	0.0009
Age	0.0759	0.0012	-0.0354	0.0009	0.1212	0.0027
Income: 15K - 25K	-0.0026	0.0018	-0.0102	0.0015	-0.0563	0.0012
Income: 25K - 35K	0.0204	0.0017	-0.0647	0.0019	-0.0461	0.0029
Income: 35K - 50K	0.0294	0.0014	-0.0862	0.0007	0.0133	0.0015
Income: 50K - 75K	0.0415	0.002	-0.098	0.0023	0.0229	0.0029
Income: 75K - 100K	0.0576	0.0024	-0.1431	0.0021	0.071	0.0009
Income: 100K - 120K	0.0744	0.0019	-0.176	0.001	0.0266	0.0012
Income: 120K - 149K	0.0823	0.0018	-0.1888	0.0015	0.1005	0.0015
Income: 150K plus	0.0999	0.0009	-0.2264	0.0006	0.0806	0.0015
Days Since Last Opened	-14.6378	0.0092				
Days Since Last Opened^2	10.6483	0.0114				
Open: Knot 1	-12.4159	0.0055				
Open: Knot 2	2.8864	0.0068				
Open: Knot 3	-1.7323	0.01				
Open: Knot 4	1.3113	0.0092				
Open: Knot 5	-1.5789	0.0065				
Open: Knot 6	1.1663	0.0058				
Days Since Last Clicked			-17.101	0.0097		
Days Since Last Clicked^2			11.8783	0.0059		
Click: Knot 1			-11.6513	0.0079		
Click: Knot 2			-0.1214	0.0029		
Click: Knot 3			0.5692	0.0042		
Click: Knot 4			-1.3804	0.0021		
Click: Knot 5			0.6875	0.0056		
Click: Knot 6			-0.0211	0.006		
Days Since Last Purchased	-0.4478	0.0053	-0.4946	0.0043	-15.3012	0.0258
Days Since Last Purchased^2	-0.0714	0.0038	0.0923	0.0029	11.1235	0.0158
Purchase: Knot 1	0.7259	0.0063	0.3957	0.0031	-11.4026	0.0243
Purchase: Knot 2	-0.8141	0.0049	-0.6642	0.0049	-0.1895	0.0193
Purchase: Knot 3	0.3235	0.0054	0.2242	0.0039	-0.0839	0.0043
Purchase: Knot 4	-0.2684	0.0151	0.0393	0.0059	4.6886	0.0133
Purchase: Knot 5	0.2064	0.0033	-0.137	0.0059	-7.8951	0.0356
Purchase: Knot 6	-0.1359	0.0046	0.09	0.0037	3.7391	0.0189
Promo 1	0.0600	0.0022	0.0660	0.0011	0.2869	0.0015
Promo 2	0.2631	0.0028	0.8079	0.0032	0.7316	0.0032
Promo 3	-6.0018	0.0042	-5.7259	0.0022	-1.6511	0.0030
Promo 4	0.2370	0.0036	0.2756	0.0023	0.6119	0.0038
Promo 5	0.1710	0.0028	0.5393	0.0035	1.1760	0.0018
Promo 6	-0.0739	0.0036	0.8038	0.0029	1.4671	0.0036
Promo 7	-0.2311	0.0011	0.0000	0.0000	0.5025	0.0011
Promo 8	-0.1854	0.0039	-0.0561	0.0028	1.0623	0.0054
Promo 9	-0.0918	0.0031	0.5648	0.0016	0.6738	0.0014
Promo 10	0.2966	0.0039	0.6717	0.0017	0.9202	0.0035
Promo 11	0.0669	0.0009	0.8296	0.0029	1.2507	0.0059
Promo 12	-0.2903	0.0014	0.6151	0.0030	1.4947	0.0022
Promo 13	0.5607	0.0026	1.3398	0.0017	1.0655	0.0031
Promo 14	-0.1246	0.0043	0.3207	0.0023	2.2080	0.0023
Promo 15	-5.9961	0.0014	-5.7910	0.0020	-1.5469	0.0033
Promo 16	0.4458	0.0012	0.7847	0.0026	1.6643	0.0037
Promo 17	0.4797	0.0039	0.2891	0.0024	1.8743	0.0039
Promo 18	-0.0938	0.0015	0.1729	0.0006	1.7653	0.0059
Promo 19	-6.0752	0.0009	-6.1621	0.0018	-3.1790	0.0010
Promo 20	0.4662	0.0047	0.5986	0.0013	1.9910	0.0024
Promo 21	0.6212	0.0018	0.4997	0.0038	2.3634	0.0031
Promo 22	0.4947	0.0059	0.3144	0.0027	2.2773	0.0049
Promo 23	-5.8418	0.0033	-5.7605	0.0011	-1.3279	0.0044
Promo 24	-0.0652	0.0036	0.3816	0.0020	-0.5405	0.0023
Promo 25	0.3012	0.0046	0.9940	0.0024	1.2706	0.0031

Table S.9: Stage-wise model estimates of SCRON for high engagement customers

Coefficients	Open stage		Click stage		Purchase stage	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	5.1055	0.0023	5.7398	0.001	0.0189	0.0024
Ave. Retail Spend	0.0006	0.0002	-0.0003	0.0002	-0.0007	0.0003
Ave. Web Spend	-0.0002	0.0002	-0.0007	0.0002	0.004	0.0003
Past Pur. Freq.	0.0032	0.0006	0.0056	0.0008	0.0236	0.0009
Age	0.0759	0.0012	-0.0354	0.0009	0.1212	0.0027
Income: 15K - 25K	-0.0026	0.0018	-0.0102	0.0015	-0.0563	0.0012
Income: 25K - 35K	0.0204	0.0017	-0.0647	0.0019	-0.0461	0.0029
Income: 35K - 50K	0.0294	0.0014	-0.0862	0.0007	0.0133	0.0015
Income: 50K - 75K	0.0415	0.002	-0.098	0.0023	0.0229	0.0029
Income: 75K - 100K	0.0576	0.0024	-0.1431	0.0021	0.071	0.0009
Income: 100K - 120K	0.0744	0.0019	-0.176	0.001	0.0266	0.0012
Income: 120K - 149K	0.0823	0.0018	-0.1888	0.0015	0.1005	0.0015
Income: 150K plus	0.0999	0.0009	-0.2264	0.0006	0.0806	0.0015
Days Since Last Opened	-14.6378	0.0092	0	0	0	0
Days Since Last Opened^2	10.6483	0.0114	0	0	0	0
Open: Knot 1	-12.4159	0.0055	0	0	0	0
Open: Knot 2	2.8864	0.0068	0	0	0	0
Open: Knot 3	-1.7323	0.01	0	0	0	0
Open: Knot 4	1.3113	0.0092	0	0	0	0
Open: Knot 5	-1.5789	0.0065	0	0	0	0
Open: Knot 6	1.1663	0.0058	0	0	0	0
Days Since Last Clicked	0	0	-17.101	0.0097	0	0
Days Since Last Clicked^2	0	0	11.8783	0.0059	0	0
Click: Knot 1	0	0	-11.6513	0.0079	0	0
Click: Knot 2	0	0	-0.1214	0.0029	0	0
Click: Knot 3	0	0	0.5692	0.0042	0	0
Click: Knot 4	0	0	-1.3804	0.0021	0	0
Click: Knot 5	0	0	0.6875	0.0056	0	0
Click: Knot 6	0	0	-0.0211	0.006	0	0
Days Since Last Purchased	-0.4478	0.0053	-0.4946	0.0043	-15.3012	0.0258
Days Since Last Purchased^2	-0.0714	0.0038	0.0923	0.0029	11.1235	0.0158
Purchase: Knot 1	0.7259	0.0063	0.3957	0.0031	-11.4026	0.0243
Purchase: Knot 2	-0.8141	0.0049	-0.6642	0.0049	-0.1895	0.0193
Purchase: Knot 3	0.3235	0.0054	0.2242	0.0039	-0.0839	0.0043
Purchase: Knot 4	-0.2684	0.0151	0.0393	0.0059	4.6886	0.0133
Purchase: Knot 5	0.2064	0.0033	-0.137	0.0059	-7.8951	0.0356
Purchase: Knot 6	-0.1359	0.0046	0.09	0.0037	3.7391	0.0189
Promo 1	0.2054	0.0015	0.3237	0.0022	3.0569	0.0025
Promo 2	0.6518	0.0035	1.2511	0.0019	3.8000	0.0028
Promo 3	-5.7769	0.0061	-5.5277	0.0021	-1.3692	0.0011
Promo 4	0.4282	0.0026	0.4634	0.0019	2.6974	0.0033
Promo 5	0.3987	0.0036	0.8169	0.0028	2.8405	0.0024
Promo 6	-0.0197	0.0033	0.9983	0.0020	3.1506	0.0022
Promo 7	-0.1136	0.0023	0.1235	0.0018	2.8620	0.0041
Promo 8	0.0118	0.0036	0.1383	0.0030	3.4015	0.0049
Promo 9	0.0136	0.0023	0.8328	0.0031	3.0284	0.0010
Promo 10	0.5951	0.0045	1.0205	0.0021	3.2504	0.0046
Promo 11	0.1377	0.0012	0.9701	0.0042	2.6549	0.0047
Promo 12	-0.2012	0.0041	0.7815	0.0030	3.1772	0.0021
Promo 13	0.8168	0.0030	1.6711	0.0031	2.6507	0.0041
Promo 14	-0.0296	0.0024	0.5644	0.0027	3.5224	0.0038
Promo 15	-5.8775	0.0023	-5.7710	0.0040	-1.4460	0.0028
Promo 16	0.5469	0.0038	0.9328	0.0020	2.8437	0.0029
Promo 17	0.7729	0.0040	0.3807	0.0039	3.1104	0.0039
Promo 18	-0.0374	0.0030	0.2739	0.0034	2.9723	0.0056
Promo 19	-6.0752	0.0009	-6.2444	0.0016	-3.0993	0.0027
Promo 20	0.4483	0.0045	0.7262	0.0020	3.3156	0.0029
Promo 21	0.6212	0.0018	0.5576	0.0030	3.3863	0.0029
Promo 22	0.7453	0.0072	0.4492	0.0025	3.2637	0.0054
Promo 23	-5.9312	0.0010	-5.9169	0.0018	-1.5900	0.0044
Promo 24	-0.1178	0.0014	0.3373	0.0007	0.3053	0.0031
Promo 25	0.2477	0.0017	1.0814	0.0033	2.2192	0.0037

Table S.10: For different levels of false positive rates, we report the correct detection rates of conversions in the test data from the ROC curves of four models in Figure 4.

Data Type	Stage	Model	False Positive Rate							
			0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Training	Open	NCG	0.47	0.58	0.68	0.77	0.83	0.89	0.93	0.97
		NCL	0.44	0.57	0.68	0.77	0.83	0.89	0.93	0.97
		CRON	0.47	0.6	0.7	0.78	0.84	0.9	0.94	0.98
		SCRON	0.47	0.6	0.7	0.78	0.85	0.9	0.94	0.98
	Click	NCG	0.21	0.32	0.42	0.52	0.62	0.72	0.82	0.91
		NCL	0.27	0.37	0.47	0.56	0.65	0.74	0.82	0.91
		CRON	0.34	0.46	0.56	0.66	0.75	0.83	0.89	0.96
		SCRON	0.35	0.46	0.57	0.66	0.75	0.83	0.9	0.96
	Purchase	NCG	0.29	0.41	0.51	0.61	0.7	0.79	0.87	0.94
		NCL	0.31	0.43	0.54	0.64	0.73	0.81	0.88	0.95
		CRON	0.44	0.56	0.67	0.76	0.85	0.92	0.97	0.998
		SCRON	0.46	0.58	0.7	0.8	0.88	0.95	0.98	0.998
Test	Open	NCG	0.36	0.5	0.62	0.73	0.82	0.88	0.93	0.97
		NCL	0.42	0.55	0.65	0.74	0.82	0.88	0.93	0.97
		CRON	0.45	0.59	0.69	0.77	0.84	0.9	0.94	0.98
		SCRON	0.44	0.59	0.69	0.77	0.84	0.9	0.94	0.98
	Click	NCG	0.2	0.3	0.4	0.5	0.61	0.7	0.8	0.9
		NCL	0.23	0.35	0.45	0.55	0.65	0.72	0.81	0.9
		CRON	0.33	0.45	0.55	0.65	0.74	0.83	0.89	0.96
		SCRON	0.33	0.45	0.55	0.65	0.75	0.83	0.9	0.97
	Purchase	NCG	0.26	0.38	0.49	0.58	0.67	0.76	0.84	0.92
		NCL	0.29	0.4	0.51	0.61	0.7	0.78	0.86	0.93
		CRON	0.45	0.55	0.65	0.74	0.82	0.89	0.96	0.998
		SCRON	0.47	0.6	0.71	0.81	0.88	0.94	0.98	0.998

S.3 Proof of Proposition 1

The proof follows by checking the conditions of Theorem 2.1 of Dey et al. (2016) (henceforth abbreviated as DB16) that ensures geometric convergence of Markov chains with transformation-MCMC modifications (TMCMC). Let $T = (\Theta, B)$ and dimension$(T) = p$. A distribution h is said to be from a super-exponential family if

$$\lim_{||T|| \to \infty} ||T||^{-1} \sum_{j=1}^{p} T_j \, \frac{\partial \log h(T)}{\partial T_j} = -\infty. \tag{9}$$

Theorem 2.1 of DB16 showed that if the target density $h(T)$ is from a super exponential family and does not have contours parallel to $\{T : |T_1| = |T_2| = \ldots = |T_p|\}$ then the corresponding TMCMC chain is geometric ergodic. We next check that our proposal density defined in Section 4 is super-exponential. The prior distribution $g(T)$ used here is continuous and differentiable. Decompose the logarithm of the proposal density as $\log h(T) = l_S(T) + \log g(T)$ where $l_S(T)$ is the log-likelihood based on sample S defined in Equation (8) of the main paper.

As done in Equation (6) of the main paper, we decompose T in Θ, B, and Σ and $\log g(T) = \log g_1(\Theta) + \log g_2(B) + \log g_3(\Sigma)$. Also, note that Θ is actually the conglomeration of $(\boldsymbol{\beta}, \mathfrak{m}, \boldsymbol{\nu})$.

Recall, that the parameters are different for the different stages of the conversion funnel. For e.g., $\beta_j^{(l)}$ varied over $l = 1, 2, 3$ stages of the funnel. For convenience, we would use the notation o, c, and p for stages 1, 2 and 3 of the conversion funnel. From its definition in the main paper, it follows that for $a, l \in \{o, c, p\}$:

$$\begin{aligned} &\frac{\partial}{\partial \beta_j^{(l)}} \ell_{ik}^{(a)}(T) = U_{ikj} I\{l = a\}, \qquad &&\frac{\partial}{\partial \mu_j^{(l)}} \ell_{ik}^{(a)}(T) = r_{ikj}^{(a)} I\{l = a\}, \text{ and },\\ &\frac{\partial}{\partial \nu_j^{(l)}} \ell_{ik}^{(a)}(T) = I\{j = i, l = a\}, \qquad &&\frac{\partial}{\partial b_j^{(l)}} \ell_{ik}^{(a)}(T) = I\{j = i, l = a\} \; . \end{aligned} \tag{10}$$

Next, note that by definition Equation (8) it follows that

$$
\begin{aligned}
-\frac{\partial}{\partial T_j} l_S(T) = & \sum_{(i,k)\in\Omega} \left(1 - o_{ik} - \frac{1}{1+e^{\ell^o_{ik}}}\right) \frac{\partial \ell^o_{ik}}{\partial T_j} \\
& + \sum_{(i,k)\in\Omega_c} \left(1 - c_{ik} - \frac{1}{1+e^{\ell^c_{ik}}}\right) \frac{\partial \ell^c_{ik}}{\partial T_j} \\
& + \sum_{(i,k)\in\Omega_p} \left(1 - p_{ik} - \frac{1}{1+e^{\ell^p_{ik}}}\right) \frac{\partial \ell^p_{ik}}{\partial T_j} \ .
\end{aligned} \tag{11}
$$

Since we have used independent Gaussian priors with fixed variances for each of the parameters in Θ, we have:

$$
-\sum_{T_j\in\Theta} T_j \frac{\partial}{\partial T_j} \log g(T) = \sum_{T_j\in\Theta} w_j T_j^2 \tag{12}
$$

where w_j are positive constants. Also, for the consumer specific effects we have:

$$
-\frac{\partial}{\partial b_i^{(l)}} \log g(T) = \sum_{j\neq l} \bar{\Sigma}_{lj} b_i^{(j)} + \bar{\Sigma}_{ll} b_i^{(l)} \text{ for } i = 1, \ldots, N \text{ where } \bar{\Sigma} = \Sigma^{-1}. \tag{13}
$$

Finally, note that for the $\bar{\Sigma}$ parameter, the contributions are from g_2 and g_3:

$$
\frac{\partial}{\partial \bar{\Sigma}} \log g(T) = \frac{\partial}{\partial \bar{\Sigma}} \log g_2(T) + \frac{\partial}{\partial \bar{\Sigma}} \log g_3(T).
$$

The first term is the derivative with respect to $\bar{\Sigma}$ of multivariate normal density of B. The second term is from the Wishart prior $\log g_3(T) = (d+4)/2 \log|\bar{\Sigma}| - \mathrm{tr}(\bar{\Sigma})/2 + \text{constant}$ where d is the degrees of freedom of the prior. Here, the covariance of the prior is set to I. Thus,

$$
-\frac{\partial}{\partial \bar{\Sigma}} \log g(T) = -(N + d/2 + 2)\bar{\Sigma} + 2^{-1} \sum_{i=1}^{N} \boldsymbol{b}_i \boldsymbol{b}_i' \ . \tag{14}
$$

Now, from Equation (11) it follows that $\sum_{j=1}^{p} T_j \frac{\partial}{\partial T_j} l_S(T)$ is bounded above by

$$
\sum_{(i,k)\in\Omega} o_{ik} T_j \frac{\partial \ell^o_{ik}}{\partial T_j} + \sum_{(i,k)\in\Omega_c} c_{ik} T_j \frac{\partial \ell^c_{ik}}{\partial T_j} + \sum_{(i,k)\in\Omega_p} p_{ik} T_j \frac{\partial \ell^p_{ik}}{\partial T_j}.
$$

From Equation (10) it follows that the above does not involve any quadratic terms of the parameters. Equations (12)–(14) show that all the parameters that appear in the above sum are also involved quadratically in $\sum_{j=1}^{p} T_j \frac{\partial}{\partial T_j} \log g(T)$. Thus, for any fixed N, $\sum_{j=1}^{p} T_j \frac{\partial}{\partial T_j} l_S(T)$ is dominated by $\sum_{j=1}^{p} T_j \frac{\partial}{\partial T_j} \log g(T)$. Note, both assumptions A1 and A2 are essential to ensure this domination. Moreover, the quadratic terms have negative coefficients in $\sum_{j=1}^{p} T_j \frac{\partial}{\partial T_j} \log g(T)$ and therefore, we arrive at Equation (9) which completes the proof.

S.4 Discussion on Predictive Optimality of BHMT

The adjustments, prescribed in Section 4.2.3 of the main paper, results in consistent estimates of the model parameters. Based on this consistent-estimation scheme, our proposed method can be used for predicting the response of future customers. In Section 5 we have provided the ROC curves for test data sets based on models estimated by our proposed BHMT procedure. Following the analysis framework developed in Fithian and Hastie (2014), asymptotic optimality of the BHMT methodology can be established. For any fixed covariate matrix X_{te}, let $Y_{\mathsf{te}} = \{(O_{ik}, C_{ik}, P_{ik}) : i = 1, \ldots, M; k = 1, \ldots, K\}$ responses from m new customers generated from Equations (1)-(4) with true parameters being Θ_0 and $B_{\mathsf{te}} = \{b_i^{\mathsf{te}} : i = 1, \ldots, M\}$ being i.i.d. from $N_3(0, \Sigma_0)$, for any positive definite Σ_0. Denote the true distribution of Y_{te} by $q(Y_{\mathsf{te}}|X_{\mathsf{te}}, \Theta_0, B_{\mathsf{te}})$.

Let X_{tr} be the fixed training matrix and Y_{tr} is generated from (1)-(4) using X_{tr}, Θ_0 and $B_{\mathsf{tr}} = \{b_i^{\mathsf{tr}} : i = 1, \ldots, N\}$ being i.i.d from $N_3(0, \Sigma_0)$. Consider the plugin estimate $q(Y_{\mathsf{te}}|X_{\mathsf{te}}, \hat{\Theta}^S, \hat{B}^S)$ for the predictive distribution Y_{te} where $\hat{B}^S$ contains i.i.d. vectors from $N(0, \hat{\Sigma}^S)$. Note, that $\hat{\Theta}^S, \hat{B}^S$ are functions of $(Y_{\mathsf{tr}}, X_{\mathsf{tr}})$. The analysis in Fithian and Hastie (2014) can be extented to provide weak consistency of the estimated paramters $\hat{\Theta}^S \xrightarrow{P} \Theta_0$ and $\hat{\Sigma}^S \xrightarrow{P} \Sigma_0$. These when substituted in (5) yield the following result on good asymptotic performance by the BMHT based plugin predictive distribution.

Proposition 2. *For any fixed K, as $M, N \to \infty$ the plugin predictive distribution using BHMT estimates from a subsample S that satisfies assumptions A1 and A2 converges to the true distribution in Kullback-Leibler loss:*

$$\frac{1}{M} \int \int q(Y_{te}|X_{te}, \Theta_0, B_{te}) \left\{ \log \frac{q(Y_{te}|X_{te}, \hat{\Theta}^S, \hat{\Sigma}^S)}{q(Y_{te}|X_{te}, \Theta_0, B_{te})} \right\} dY_{te}\, dY_{tr} \xrightarrow{P} 0.$$