Improving Discrete Time BTYD Model with Covariates and Non-Parametric Priors

By

Zhangyi (David) Fan (W’21, GEN’21)

fzy@wharton.upenn.edu

An Undergraduate Thesis submitted in partial fulfillment of the requirements for the

WHARTON RESEARCH SCHOLARS

Faculty Advisor:

Peter Fader

faderp@wharton.upenn.edu

Frances and Pei-Yuan Chia Professor, Marketing Department

Research Discipline:

Statistics, Marketing

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

MAY 2021

ABSTRACT

Overall, the paper explored the discrete buy-till-you-die process by incorporating covariate effects and non-parametric priors. In doing so, the paper hopes to enrich the space of customer lifetime value modelling. Using a simulation-based approach, the paper found that models with non-parametric priors is capable of adequately picking up simplified parametric distribution shapes without drastically overfitting and offer some predictive improvement in cases where a multi-modal distribution exists. In addition, the paper also found that models missing covariate specification, when such effect is present, may generate systematic upward bias to the parameter values. Such bias will lead to a bad aggregate level model performance even when such covariates are missing in the future while also causing the model to underpredict individual and aggregate conversions when covariates are in fact present. The paper also performed a market simulation that showed how the covariate effect extracted from the models can help firms better perform targeted marketing to improve its ROI.

Keywords: Customer Lifetime Value, Marketing, Dirichlet Process, Covariate Marketing Mixes, Probability Models

INTRODUCTION AND EXISTING STUDIES BACKGROUND

Customer lifetime value has long been discussed both academically and within the industry as it has huge implications across various business functions. Having an accurate customer lifetime value measurement for customers allows us to benchmark the appropriate customer acquisition cost for various cohorts and craft customer targeting strategies. In addition, there has also been a new wave of academic research that seeks to create forecasting based financial models to value a company based on the values of the customer.

The challenge of CLV models is particularly pertinent in cases where customer churn is not directly observable and when the transaction is not contractual in nature. The absence of observable churn meant that the CLV model would have to guess not just when customers are likely to purchase a product but also whether the customer is still active. This paper will focus on the non-contractual set of CLV models. However, within the non-contractual setting, there are two additional case that needs to be considered.

- **Continuous Time Contractual Case:** This refers to transactional events that are not subscription based (it could terminate anytime) and could happen at any time (there is no fix opportunity for the transaction)
    - *Examples:* Online Grocery Purchases, Hotel Visits
- **Discrete Time Non-Contractual Case:** This refers to transactional events that are not subscription based (it could terminate anytime) but could only happen at fixed period (there is fix opportunity for the transaction)
    - *Examples:* Event Attendance, Charity Drives

What underpins all the applicability of CLV models is the inherent ability of the models to predict customer values. This may not strictly be the customer lifetime values per se, since in the non-contractual case, it may not be directly observable. Instead, it could be the customer values in terms of what we expect the customers to do in the next period, the next five periods or later. In lieu of that, there has always been interest in improving the models we have in accurately predicting customer lifetime values to better perform the tasks highlighted earlier.

Classic CLV Model

Different kind of CLV models have been created to tackle these scenarios. For the Continuous Time Non-Contractual case, the known models include Pareto-NBD (Pareto-Negative Binomial Distribution) model (Schmittlein, Morrison and Colombo, 1987), while for the Discrete Time Non-Contractual case, the known models include BGBB (The Beta-Geometric Beta-Binomial) model (Fader, Hardie and Shang, 2009). Both models mentioned model customer behavior from two lenses/processes: the 'Buying' story and the 'Dying' story. On an individual level, both models assume that every customer has some inherent buying propensity. For the discrete case, in each discrete period, the customer is said to have some probability of making a purchase (i.e., Binomial Distribution), whereas in the continuous case, customers are said to have a purchasing rate that governs their purchasing propensity in a unit time (i.e., The Poisson Distribution). Similar, every customer is also assumed to have some inherent dying/churning propensity where in, for the discrete case, at each discrete period, they may end up churning away from the company with some probability (the Geometric Distribution) and for the continuous case, at every single period, they may end up churning based on some churning rate (The Exponential Distribution). It can also be inferred here that the Poisson Distribution and Exponential

Distribution are respectively a generalization of the Binomial Distribution and Geometric distribution, where the number of opportunities becomes infinitely large and the propensity at the single opportunity becomes increasingly small.

To collectively model this for all individuals/the population, the models assume that every single customer draws their respective buying and dying propensity from a beta distribution for the discrete case and the gamma distribution for the continuous case. These two distributions are chosen since the beta distribution is nicely bounded between zero and one (a measure of propensity) and have a conjugacy relationship with the Binomial and Geometric distribution, while the gamma distribution is bounded between zero and infinity (a measure of rate) and have a conjugacy relationship with the Exponential and Poisson distribution. It is important to note here that these propensities are not observable but are in fact latent. These traits can be then later backed out through empirical data.

Together, the buying model (The Beta Binomial model for the discrete case, and the Negative Binomial Model for the continuous case) and the dying model (The Beta Geometric model for the discrete case, and the Pareto II Model for the continuous case) forms a collective model that allows us to predict future collective purchases. Furthermore, through conditional expectation it also allows us to extrapolate individual level inferences.[1]

Pitfall of the Classic Models

[1] Negative Binomial Model is a mixture model made from a Poisson and Gamma mixture; Pareto II is a mixture model made from an Exponential and Gamma mixture

As can be seen earlier, the classic CLV models such as Pareto-NBD and BGBB, along with its other variants: BG/NBD (Fader, Hardie and Lee, 2005) and PDO (Jerath, Fader and Hardie, 2011), focus on extrapolating some latent characteristics about the customers (the dying and buying propensity) whereby each of these characteristics is defined by a distribution. The characterization through a form of predisposed distribution (or what we refer to as the prior distribution) is why this school of model is often referred to as a parametric form of model. While the parameterization (and conjugacy) allows for ease of estimation and straight forward model inferences, it does impose a flexibility constraint on how the latent characteristic of customers may look like. In particular, the distribution chosen to represent the latent traits (Beta Distribution & Gamma Distribution) tends to be either unimodal or bimodal in shape, which may not be a realistic characterization of underlying customer traits.

In addition to the rigidity of prior distribution, classic CLV models also do not have a straightforward integration of covariates. This poses a serious challenge for the adoption of these models as external events such as marketing campaigns or macro level market sentiment, and simple demographic information can affect customer behaviors and hence their value. The integration of marketing mix covariates, such as active promotions, loyalty programs and other methods that aim to boost customer activity and retention, could also help companies quantify the effectiveness of these strategies in terms of boosting customer values.

Existing Work

The two subfields are not entirely novel as existing literatures have made progress around both areas (i.e., non-parametric/semi-parametric priors & covariate integration).

On nonparametric priors, one notable paper focused on utilizing the Gaussian Process prior in customer base analysis (Dew and Ansari, 2018). This paper specifically challenged the lack of flexibility in parametric models and in how it fails to capture potential seasonality and calendar events that may be crucial for forecasting customer behavior. This paper posited that the use of a non-parametric dynamic prior such as Gaussian Process will allow marketers to preserve the principled approach to utilizing latent traits but also incorporate calendar time events. This paper concluded by showing how the use of such a non-parametric distribution was indeed able to improve the model performance when benchmarked against its parametric predecessors.

Another paper that also took advantage of the flexible non-parametric distribution involved using a Dirichlet Process prior in place of the Gamma distribution (present in the Pareto NBD model) for customer lifetime value predictions (Quintana and Marshall, 2014). This paper directly tackled how the commonly used distributions in parametric CLV models assume a rather unrealistic and simplistic view of how customer traits are distributed. To address that, the paper incorporated the Dirichlet Process prior, which does not assume anything about the latent traits of the customer in place. The other assumptions that extended from the Pareto NBD model are still persistent with this new model. The proposed model was compared to the Pareto NBD model in two different datasets and was able to outperform the parametric counterpart in both.

On the covariate side, one paper brought in the use of time invariant covariates into the Pareto NBD model like that of an NBD regression model (Singh, Borle and Jain 2009). The covariate effects, which in this case are constant through time, were brought in directly to the underlying parameters of the Pareto distribution and the NBD distribution. This method allows the model to capture an additional level of observed heterogeneity in addition to the unobserved heterogeneity intrinsic within the prior distribution. The model outperformed the vanilla Pareto-NBD model in both its predictive abilities but also its potential for targeting the most valuable customers

Another notable set of papers took a different but complementary approach to bring in time variant covariates into probability models. One example of such is the paper by Gupta (Gupta, 1991), where he integrated time varying covariates into adoption/inter-purchase processes models such as Pareto II and Erlang 2-Gamma. The crucial assumption laid out in the paper is that the covariate effect is typically fixed for a certain duration until such covariate changes. This allows the modelling process to be easier as modelers can then partition the inter-arrival process to individual periods before the covariates change. It also is logically sound as covariate effects such as marketing mix are not expected to materialize immediately due to for instance lagging exposure to the marketing mix or physical constraint that delays the purchase. The integration of the covariate itself within their fixed interval can be thought of as a stretching of time, where a positive covariate effect is thought to be equivalent to an individual having a longer period to decide whether he/she wants to purchase the product.

This general modelling philosophy around partitioning the inter-purchase process by covariate changes and modelling covariate as a stretching of time is later adopted by Schweidel and Knox (2010). They utilized a similar modelling philosophy except this time bringing such covariates into the both components of the Pareto-NBD model, where the dying process is imagined to be a single Pareto distribution with covariates and the buying process that of one which is analogous to the inter-arrival process described earlier. Both papers found that the integration of covariates yield a significant effect and improved the modelling performance.

Contribution

The above literature, and the wide array of other literature on the area all tend to focus on the Pareto-NBD model (and hence the continuous time non-contractual case). This paper would like to contribute and broaden the scope of existing literature by integrating on-parametric priors and time-varying covariates into the discrete time non-contractual case

While Discrete Time Non-Contractual models may not be as prominent as its continuous counterpart, it can still nonetheless be used in a wide array of settings. These includes two main cases, with the first being inherently discrete transactional activities such as responding to email campaigns (where the opportunity is discrete when the email is sent) and event/concert attendances (where joining the event is a discrete scenario). The second case would be discretized continuous settings such as modelling non-contractual yearly retention and spending. For the discretized continuous case, the pro of using discrete time models is that it produces parameters that are easier to interpret than its continuous counterpart through the forms of probability. It also is generally more convenient to estimate.

MODEL FORMATION MATHEMATICAL FORMULATION OF BGBB

Before moving to discuss the novel model formation proposed in this paper, it is important to discuss the mathematical details of the BGBB model, starting from the individual level distribution. As aforementioned, the individual level model of a discrete time non-contractual process includes a ‘buying’ story and ‘dying’ story, represented by a binomial distribution and geometric distribution respectively. This can be concretized with an example. Assuming we observe a customer for ten periods and this customer has made a purchase on fifth and seventh period, from an individual model point of view, there are several cases that are possible to represent such an observable data:

- this customer could have made a purchase on the fifth and seventh period and immediately churned in the next period (eighth period)
- this customer could’ve made a purchase on the fifth and seventh period, stayed with the company but didn’t purchase on the eighth period, and churned on the ninth period
- this customer could’ve made a purchase on the fifth and seventh period, stayed with the company but didn’t purchase on the eighth and ninth period, and churned on the tenth period
- this customer could’ve made a purchase on the fifth and seventh period, stayed with the company but didn’t purchase on the eighth, ninth and tenth period, and didn’t churn at all

Mathematically, if assume this customer has *p* as his purchasing propensity at every period and *u* as his dying propensity every period, the aforementioned scenario can be summed into a likelihood expression:

$$L(p, u \mid purchasing\ 5th\ and\ 7th\ period, observed\ for\ 10\ periods)$$
$$= (1-p)^4 p(1-p)p(1-u)^7 u + (1-p)^4 p(1-p)p(1-p)(1-u)^8 u$$
$$+ (1-p)^4 p(1-p)p(1-p)^2(1-u)^9 u$$
$$+ (1-p)^4 p(1-p)p(1-p)^3(1-u)^{10}$$

$$= (1-p)^8 p^2 (1-u)^{10} + \sum_{i=0}^{2} (1-p)^{5+i} p^2 (1-u)^{7+i} u$$

What we can notice here is that the above likelihood is only affected by the recency and frequency of purchase. Thus, more generally, for a customer who has been observed for *n* periods, who purchased *x* times and who's last purchase is on the *t*-th period, his/her likelihood would be:

$$L(p, u \mid x, t, n) = (1-p)^{n-x} p^x (1-u)^n + \sum_{i=0}^{n-t-1} (1-p)^{t-x+i} p^x (1-u)^{t+i} u$$

Above forms the individual distribution. To then integrate the prior distribution to incorporate heterogeneity across the population, a beta distribution is assumed for both then *p* and *u* propensity of every individual, that is:

$$f(p|a, B) = \frac{p^{a-1}\,(1-p)^{B-1}}{B(a, B)}$$

and

$$f(u|\gamma, \delta) = \frac{u^{\gamma-1}\,(1-u)^{\delta-1}}{B(\gamma, \delta)}$$

a, B,γ, δ are the so called hyperparameters that governs that distribution of the buying and dying propensities, these are also the parameters that needs to be changed in order for us to maximize our likelihood based on the data. Since *p* and *u* are not observable on a per individual basis, the joining of the individual model and the prior distribution would require the integration of the individual likelihood over the prior distribution for both parameters.

$$L(a, B, \gamma, \delta|\ x, t, n) = \int_0^1 \int_0^1 L(p, u|x, t, n) * f(p|a, B) * f(u, \gamma, \delta) dpdu$$

$$= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{\beta(\alpha, \beta)} \frac{B(\gamma, \delta + n)}{B(\gamma, \delta)}$$

$$+ \sum_{i=x}^{n-1} \binom{i}{x} \frac{B(\alpha + x, \beta + i - x)}{\beta(\alpha, \beta)} \frac{B(\gamma + 1, \delta + i)}{B(\gamma, \delta)}$$

Non-Parametric Priors

In principle, the non-parametric priors are supposed to add additional flexibility to the distribution of the latent characteristic to accommodate multi-modal cases. More generally, what the paper poses is that the distribution of latent traits is not something that is immediately obvious a priori – thus the best forms of model to represent such latent traits should be one that can adapt and thereby assimilate various forms of latent trait distribution.

For this reason, the Dirichlet Process prior is chosen. There are several ways that the DP prior can be illustrated mathematically, the formulation chosen here follows from (Neal, 2000). Generally, if we want to apply a DP Prior to data $y_1, y_2 \ldots y_n$, such prior can be thought of as:

$$y_i \mid \theta_i \sim F(\theta_i)$$

$$\theta_i \mid G \sim G$$

$$G \sim DP(G_0, a)$$

Where G_0 is the base distribution and a is the concentration parameter. To further contextualize this distribution, one can analyze such a prior on θ_i as a successive conditional distribution (Blackwell and MacQueen, 1973):

$$\theta_i \mid \theta_i \dots \theta_{i-1} \sim \frac{1}{i-1+a}\sum_{j=1}^{i-1}\delta(\theta_j) + \frac{a}{i-1+a}G_0$$

Where $\delta(\theta_j)$ refers to the distribution concentrated at θ_j**.** Inference from the above formulation sheds light on how the Dirichlet Process Prior operates. What this above formulation indicates is that a successive draw from the Dirichlet Process prior, would be dependent on where the past draws are and what the base distribution indicates. *a,* the concentration parameter, is meant to indicate how close should the distribution lie relative to the base distribution.

An alternative formation of Dirichlet process prior, may further clarify its use in the context of this paper:

$$y_i \mid c_i, \boldsymbol{\phi} \sim F(\phi_{c_i})$$

$$c_i \mid \boldsymbol{p} \sim Discrete(p_1, \dots, p_k)$$

$$\phi_c \sim G_0$$

$$\boldsymbol{p} \sim Dirichlet(\frac{a}{K}, \dots, \frac{a}{K})$$

In this formulation, c_i refers to a latent class that is associated with an observation. The above formulation, after taking *K* to infinity will yield a Dirichlet Process prior (Neal, 2000), illustrating the nature of Dirichlet Process prior as a distribution of distributions, or in other words an infinite mixture. Because of such a characteristic, Dirichlet Process prior have been used in various marketing literature (Kim, Menzefricke and Feinberg, 2004) to approximate distributional forms that are initial unknown.

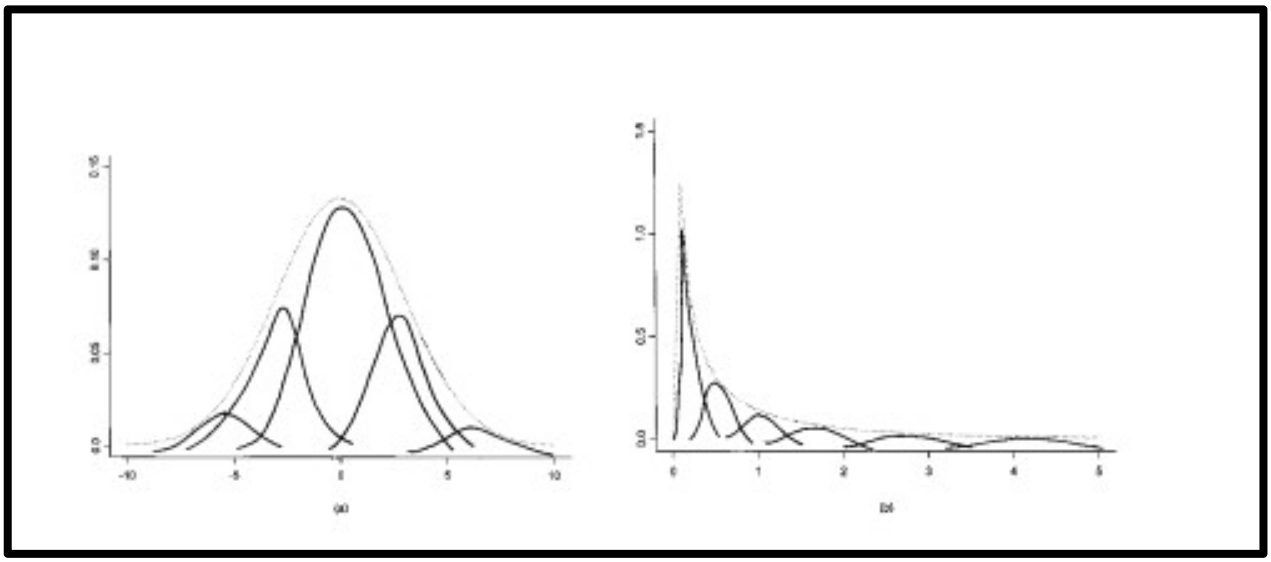


Figure 1: Dirichlet Process Priors ability to approximate various distribution (Taken from Kim, Menzefricke and Feinberg, 2004)

For this paper, the Dirichlet Process formulation attempted will take the form of hierarchical model analogous to the one posted by (Gelman, Carlin, Stern, and Rubin, 2014):

$$x_i, t_i \mid p_i, u_i, n \sim (1 - p_i)^{n - x_i} p_i^{x_i} (1 - u_i)^{n_i} + \sum_{j=0}^{x_i - t_i - 1} (1 - p_i)^{t_i - x_i + j} p_i^{x_i} (1 - u_i)^{t_i + j} u_i$$

$$p_i \sim Beta(a_i, B_j)$$

$$u_i \sim Beta(\gamma_i, \delta_j)$$

$$a_i, B_j, \gamma_i, \delta_j \sim DP(G_0, M)$$

$$G_0 \sim Unif(a_i|a_0, a_1) * Unif(B_i|B_0, B_1) * Unif(\gamma_i|\gamma_0, \gamma_1) * Unif(\delta_i, \delta_0, \delta_1)$$

Where $Unif(a_i|a_0, a_1)$ refers to a uniform distribution bounded by a_0 and a_1. Analogously yet differently for this formulation, what is suggested is that there will be various latent classes of a_i, B_j,γ, δ, such that each latent class with have a different value for a_i, B_j,γ, δ. An individual belonging to one of such latent class, will then use these values to generate its p_i and u_i through the familiar Beta distribution. Simply put, this formulation poses an infinite mixture of Beta distribution for both the dying and buying propensity.

It can be hypothesized that the DP model will do a better job that a BGBB in capturing various shape that the distribution may take but may also be more prone to overfitting due to its flexibility. Regardless, given this change, the parameter optimization, and subsequently the process of obtaining a posterior distribution for the parameters and individual inferences, will no longer yield a solved closed-form solution but will instead need to be maximized through the Monte Carlo Markov Chain. This model will be referred to as the DP-G DP-B model from here on onwards.

Covariates

This paper will primarily examine time-varying covariates as such brings non-stationarity into the observable process (often more crucial to model). Unlike the continuous time model, where the covariates could be introduced to stretch the time period in place (Gupta, 1991; Schweidel and Knox, 2010), the integration of covariates on the discrete time model is equally challenging. Given that in a discrete time case, the action itself can only happen at a specific time period, the stretching of time analogy would cease to make sense.

However, given that in the discrete time case, whatever covariates have occurred between an action interval and the next is only really factored in when the next action is being casted, the covariate values in between action periods can be combined into a lump sum effect that carries over to the next action periods.

This paper poses that this lump sum effect can be created in two different ways. Firstly, it can be brought in as a binary variable (i.e., whether an email campaign is sent) or Secondly in a fashion like a regression formula with a discount factor such that values closer to the action period are weighted more (i.e., how many email campaigns are sent where recent emails are weighted more). For simplicity, the paper will investigate primarily the binarized method. In addition, this effect will only be introduced into the buying story of our model, as the dying/attrition story is truly unobservable and therefore may allow the model to be overly parametrized and thus easily overfit.

This paper hypothesized that the lump sum effect mentioned earlier can be used to scale the individual's buying propensity. One can think of the base buying propensity as the intercept and the lump sum effect as a coefficient in a logistic regression to return a value bounded between 0 and 1. It is important to note that while an individual is still thought to have his own *p*, since the covariates is time varying, the buying propensity per period will no longer be homogenous for a customer and the individual level buying distribution could then no longer be easily represented as a Binomial distribution. Instead, it is more useful to imagine that every single buying choice will be an independent Bernoulli distribution with possibly a different *p* due to the covariate effect.

In order to properly integrate the covariates in a logistic regression set up, a transformation is first needed. That is, we first need to transform an individual's original *p* into a logit link set up:

$$p = \frac{1}{1 + e^{-(B_0)}}$$

$$-\log\left(\frac{1}{p} - 1\right) = B_0$$

We can then introduce the covariates into the purchasing propensity, let us define *Z,* whether an individual purchases in the next period, as a Bernoulli random variable, where *p* is his/her inherent buying propensity:

$$Z \sim Bernoulli(p_i)$$

$$p_i \sim \frac{1}{1 + e^{-\left(-\log\left(\frac{1}{p} - 1\right) + \boldsymbol{B} * \boldsymbol{X}\right)}}$$

B, above, is a vector of coefficient corresponding to the number covariates present and **X** is the vector of the lump sum effect derived from the covariates observed between the previous period and this period. Notice that if **B** equals to 0 then p_i will collapse back to p. The formulation of **X** in a binary case would then simply be 1 when a marketing action is casted and 0 otherwise.

The extension from this single period formation to the individual level likelihood is straightforward but cumbersome. As opposed to being able to roll up the distribution by only accounting for recency and frequency, the integration of covariates means that one would need to account for the per period observations. Taking the example above:

$$
\begin{aligned}
&L(p, u \mid purchasing\ 5th\ and\ 7th\ period, observed\ for\ 10\ periods) \\
&\quad = (1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)p_7(1-u)^7 u \\
&+(1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)p_7(1-p_8)(1-u)^8 u \\
&+(1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)p_7(1-p_8)(1-p_9)(1-u)^9 u \\
&+(1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)p_7(1-p_8)(1-p_9)(1-p_{10})(1-u)^{10}
\end{aligned}
$$

$$
= (\prod_{i=1}^{10} I_i + ((-1 * I_i) * p_i)) * (1-u)^{10} + \sum_{j=0}^{2} \prod_{i=1}^{10-j} I_i + ((-1 * I_i) * p_i)) * (1-u)^{7+j} u
$$

$$
p_i \sim \frac{1}{1 + e^{-(-\log\left(\frac{1}{p}-1\right) + \boldsymbol{B} + \boldsymbol{X_i})}}
$$

$$
I_i \sim \begin{cases} 1\ if\ period\ 5\ and\ 7 \\ 0\ if\ not \end{cases}
$$

Where I_i indicates if an individual has a made a purchase in this period and p_i the purchasing propensity at period *i.* Such expansion is necessary due to the non-homogenous nature of the buying process. However as one would expect the dying process remains unchanged. More generally:

$$L(p, u \mid x, t, n)$$

$$= (\prod_{i=1}^{n} I_i + \left((-1 * I_i) * p_i\right)) * (1-u)^n + \sum_{j=0}^{n-t-1} \prod_{i=1}^{n-j} I_i + ((-1 * I_i) * p_i)) * (1-u)^{t+j} u$$

$$p_i \sim \frac{1}{1 + e^{-(-\log\left(\frac{1}{p}-1\right) + \boldsymbol{B} + \boldsymbol{X_i})}}$$

$$I_i \sim \begin{cases} 1\ if\ purchased\ at\ period\ i \\ 0\ if\ not \end{cases}$$

As opposed to the classic model where a closed form result can be nicely integrated, this method would also require the use of Monte Carlo Markov Chain to account for the heterogeneity indicated by:

$$f(p|a, B) = \frac{p^{a-1}\,(1-p)^{B-1}}{B(a, B)}$$

and

$$f(u|\gamma,\delta) = \frac{u^{\gamma-1}\,(1-u)^{\delta-1}}{B(\gamma,\delta)}$$

Non-Parametric Priors and Covariates

The final model proposed, DP-G DP-B with covariates that includes both non-parametric finite beta prior and covariate would can be seen as

$$x_i, t_i \mid p_{i,} u_i, n$$

$$= (\prod_{z=1}^{n} I_z + \left((-1 * I_z) * p_z\right)) * (1-u_i)^n + \sum_{j=0}^{n-t-1} \prod_{z=1}^{n-j} I_z + ((-1 * I_z) * p_z)) * (1-u_i)^{t+j} u_i$$

$$p_z \sim \frac{1}{1 + e^{-(-\log\left(\frac{1}{p_i}-1\right) + \boldsymbol{B} + \boldsymbol{X_z})}}$$

$$I_z \sim \begin{cases} 1\ if\ purchased\ at\ period\ z \\ 0\ if\ not \end{cases}$$

$$p_i \sim Beta\left(a_i, B_j\right)$$

$$u_i \sim Beta\left(\gamma_i, \delta_j\right)$$

$$a_i, B_j, \gamma_i, \delta_i \sim DP(G_0, M)$$

$$G_0 \sim Unif(a_i|a_0, a_1) * Unif(B_i|B_0, B_1) * Unif(\gamma_i|\gamma_0, \gamma_1) * Unif(\delta_i|\delta_0, \delta_1)$$

METHOD

The paper will implement an iterative modelling process, where the paper will show the baseline performance of the classic BGBB, BGBB with covariates, DP-G DP-B with a non-parametric prior, and end with DP-G DP-B with a non-parametric prior and covariates.

The DP formulation will utilize Algorithm 8 from Neal (2000) due to its lack of conjugacy. The algorithm is a generalization of the MCMC algorithm that utilizes the Chinese Restaurant Process sampler. The fitting of the model will be done in ***Stan.***

To gauge the efficacy of adding nonparametric priors and covariates into the model, a simulation analysis is used with three goals in mind:

- First, to understand whether or not the model in question is capable of recovering the 'true parameters' that are known a-priori. In a business context, this is useful in understanding the population distribution of consumer propensity, providing firms with a best guess of how valuable a new customer will be, and projecting out macro adoption trends.
- Second, should the model be mis-specified or underspecified, how bad would it do in terms of forecasting the individual future behavior of a customer. This in a business context is useful as it gauges the efficacy of the models in providing individual level estimates and subsequently the ability of the model to target specific customers.
- Thirdly, to understand how the estimated covariate effect of the model can be used to design a more effective marketing campaign scheme.

The paper solicited several scenarios that are to be simulated:

- A near homogenous scenario where all individuals to be simulated have the same buying and dying propensity with no covariate effect.
- A heterogeneous scenario where individuals to be simulated each have a different buying and dying propensity drawn from a unified beta distribution with no covariate effect.
- A heterogenous scenario where individuals to be simulated each have a different buying and dying propensity drawn from a unified beta distribution with a homogenous covariate effect.
- A heterogenous scenario where individuals to be simulated each have a different buying and dying propensity drawn from a unified beta distribution with a heterogenous covariate effect drawn from a normal distribution.
- A heterogeneous scenario where individuals to be simulated each belong to 1 or 2 segments and have a different buying and dying propensity drawn from their respective segment's beta distribution with no covariate effect.
- A heterogeneous scenario where individuals to be simulated each belong to 1 or 2 segments and have a different buying and dying propensity drawn from their respective segment's beta distribution with heterogeneous covariate effect drawn from a normal distribution.

200 customers over 20 periods are being simulated from these datasets to ensure computational efficiency. The covariates to be simulated can be imagined as a company-initiated marketing campaign. Two such covariates will be simulated with varying degrees of effects on the

consumer. During the training period, the covariates are assigned in a random Bernoulli process to represent a ‘test marketing’ phase, where the companies randomly assign marketing campaigns to gauge its efficacy.

The detailed parameters used to simulate the dataset can be found through the table below:

	Buying propensity (q)	Dying propensity (p)	Covariate effect (b)
Homogenous Scenario	$q \sim Beta\ (150,150)$	$p \sim Beta\ (50,350)$	$B_1 \sim 0$ $B_2 \sim 0$
Heterogenous Scenario	$q \sim Beta\ (1.5,1.5)$	$p \sim Beta\ (0.5,3.5)$	$B_1 \sim 0$ $B_2 \sim 0$
Heterogenous Scenario with Homogenous Covariate Effect	$q \sim Beta\ (1.5,1.5)$	$p \sim Beta\ (0.5,3.5)$	$B_1 \sim 0.3$ $B_2 \sim 0.6$
Heterogenous Scenario with Heterogeneous Covariate effect	$q \sim Beta\ (1.5,1.5)$	$p \sim Beta\ (0.5,3.5)$	$B_1 \sim Normal\ (0.3,0.2)$ $B_2 \sim Normal\ (0.6,0.2)$
Heterogenous Segmented Scenario	$q \mid seg\ 1 \sim Beta\ (4,8)$ $q \mid seg\ 2 \sim Beta\ (10,2)$ $P\ (seg\ 1) \sim 0.45$	$p \mid seg\ 1 \sim Beta\ (0.5,3.5)$ $p \mid seg\ 2 \sim Beta\ (3.5,7)$ $P\ (seg\ 1) \sim 0.45$	$B_1 \sim 0$ $B_2 \sim 0$
Heterogenous Segmented Scenario	$q \mid seg\ 1 \sim Beta\ (4,8)$ $q \mid seg\ 2 \sim Beta\ (10,2)$ $P\ (seg\ 1) \sim 0.45$	$p \mid seg\ 1 \sim Beta\ (0.5,3.5)$ $p \mid seg\ 2 \sim Beta\ (3.5,7)$ $P\ (seg\ 1) \sim 0.45$	$B_1 \sim Normal\ (0.3,0.2)$ $B_2 \sim Normal\ (0.6,0.2)$

with Heterogeneous Covariate Effect			

Figure 2: Parameters for the Simulated Dataset

In addition to the proposed model illustrated in the earlier sections, additional models were also run to better benchmark the results. The list of models included the following:

- Homogenous Geometric Binomial Model
- Vanila Beta-Geometric Beta-Binomial Model
- Beta-Geometric Beta-Binomial Model with homogenous covariate effect
- Beta-Geometric Beta-Binomial Model with heterogeneous covariate effect
- Vanila DP-Geometric DP-Binomial Model
- DP-Geometric DP-Binomial Model with homogenous covariate effect
- DP-Geometric DP-Binomial Model with heterogeneous covariate effect

The models listed above were fitted on the first ten periods of the data using an MCMC of 500 iterations to obtain the posterior mean and interval.

RESULTS

The parameter values obtained from the list of models under the different scenarios can be found in the appendix (Figure 3).

To reference the objectives listed earlier, the model results will be analyzed in three different fashions.

Firstly, the paper will compare the estimated distribution of the buying and dying propensity from each of the models to the expected distribution of such propensities using the actual parameters of the simulations. To further concretize any deviation, the paper will analyze the expected per period conversion and the expected spread of conversion per individual as suggested by the model. Mathematically, the two can be illustrated as follow:

$$X_{it} = Whether\ Individual_i\ converted\ at\ period_t$$

$$I = Total\ number\ of\ customers$$

$$N = Total\ number\ of\ periods$$

$$expected\ conversion\ at\ period\ T = \frac{\sum_{i=1}^{I} P(X_{it}|t = T)}{I}$$

$$Conversion\ for\ individual\ Q = \frac{\sum_{t=1}^{N} P(X_{it}|i = Q)}{N}$$

The calculations for the above quantities were done in a Monte-Carlo simulation of 200 iterations for each model. The output of the expected per period conversion can be viewed as an incremental tracking plot over 20 periods, whereas the expected spread of conversion per individual can be seen as a histogram bounded by the number of opportunities (20 periods). Note that this estimation does not require the individual level posterior distribution but only the aggregate level parameters.

Secondly, the paper will compare the conditional expectation of each model for the 200 individuals that were simulated. The conditional expectation computed will be using the posterior distributions obtained by the model when it was fitted in the using the data points from the first 10 period. The conditional expectation will then be compared to the actual number of conversions that an individual made in the 11th to 20th period. The calculation of the conditional expectation can be represented as follow:

$$Conditional\ Expectation\ for\ individual\ Q = P(Q\ Alive\ at\ t = 10) * \sum_{t=10}^{N} P(X_{it}|i = Q)$$

Thirdly, the paper will demonstrate how such covariate effect can be used to boost marketing ROI. In particular, the paper will attempt to simulate a marketing policy, starting the 10th period, based on the model specific posterior distribution of the 200 individuals and the model estimated covariate effect. The paper will do so primarily for the homogenous covariate case but will also provide a guideline as to how can such effect be extended to the heterogenous case.

Population-Wide Propensity

Homogenous Population

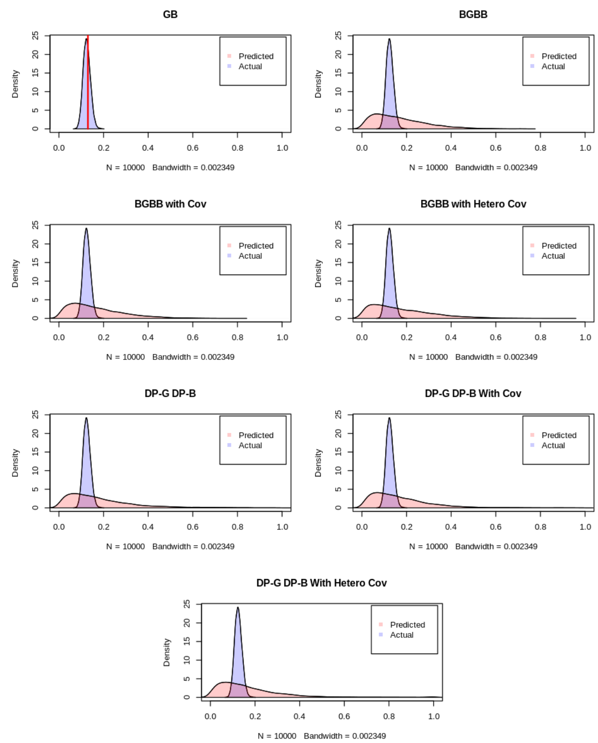


Figure 5: Estimated Dying Propensity for Homogeneous Population

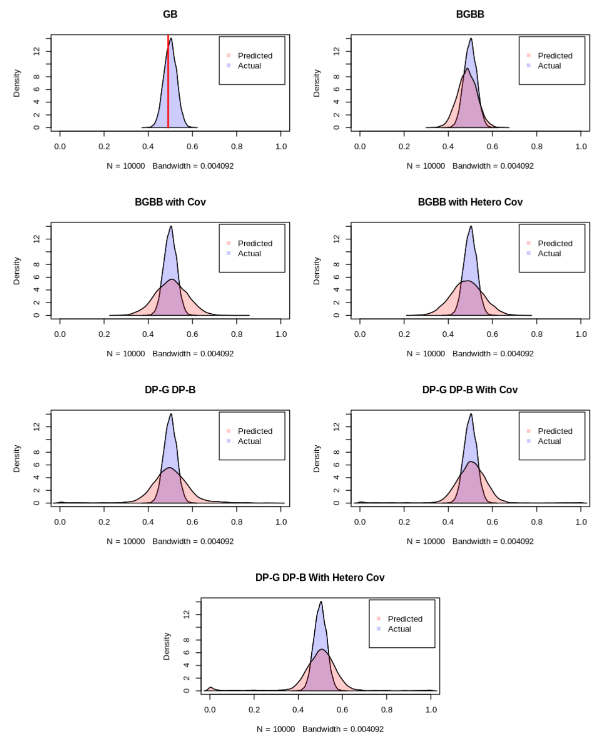


Figure 6: Estimated Buying Propensity for Homogeneous Population

Comparing the results of true underlying buying propensity and dying propensity, one could note that all the flexible models (BGBB to DP-G DP-B) seem to over inflate the heterogeneity in the dying propensity, whereas the GB homogenous model, as expected, understates it. The flexible models however were better at picking up the shape of the buying propensity with the BGBB model particularly fitting the supposed shape most closely.

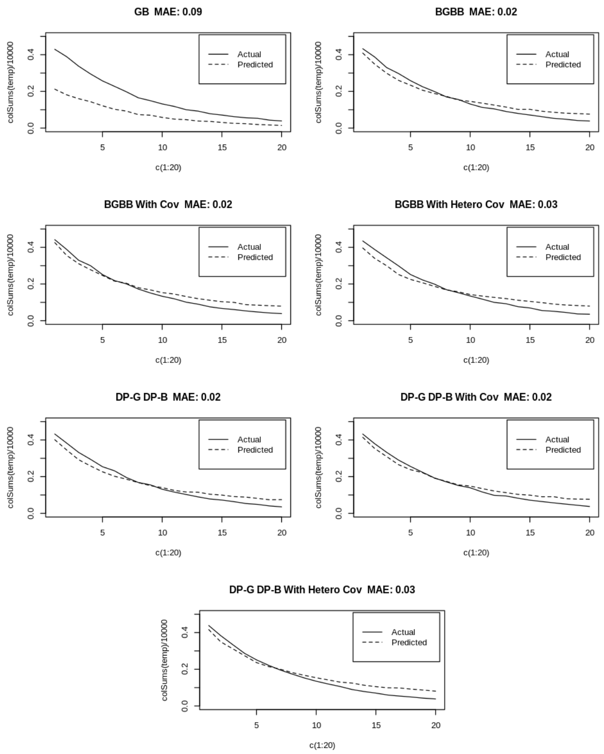


Figure 7: Incremental Plots for Homogeneous Population

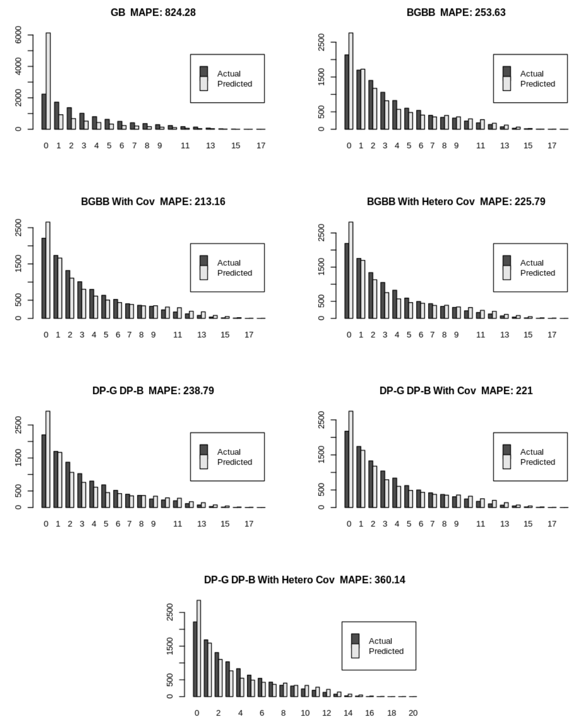
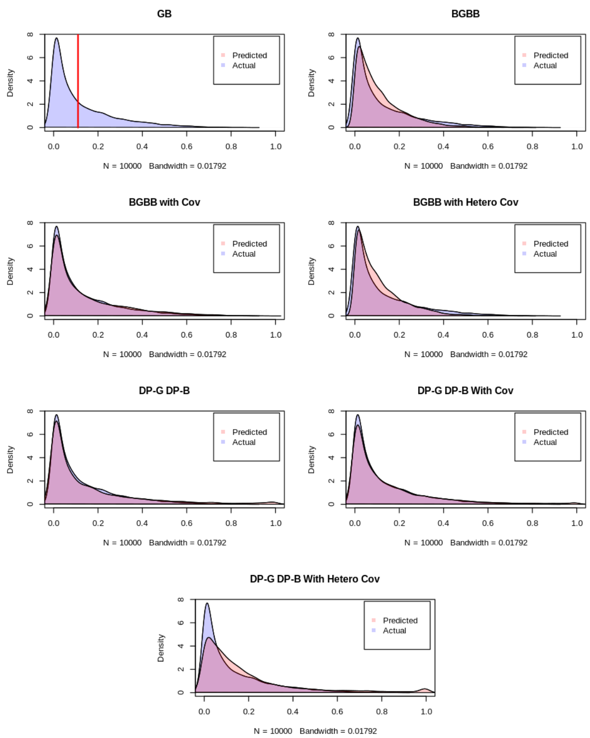


Figure 8: Histogram for Homogeneous Population

The above trend is further confirmed by the incremental plots and histogram. However, while the flexible model beats its homogenous counterpart, it also caused the incremental plots to slightly deviate from the actual result and the histogram over-stretched out due to its overstated heterogeneity.

Heterogenous Population

The results from the heterogenous are roughly like the homogenous case, with the exception that the more flexible models can perform even better in this dataset. Noting from the graphics below, the flexible models are now able to capture the span of the churning propensity rather well in addition to the buying propensity as well.



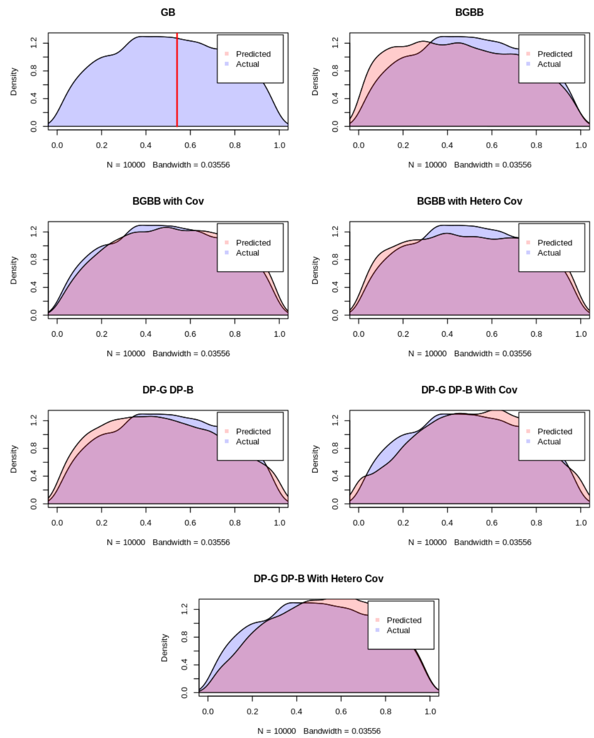


Figure 9: Estimated Dying Propensity for Heterogenous Population

Figure 10: Estimated Buying Propensity for Heterogenous Population

In a similar fashion, the incremental plots and histograms generated showed a much better fit. It was, however, worth noting that the vanilla BGBB model and the DP-G DP-B model with heterogenous covariate both showed slight deviation from expectation in the incremental plot.

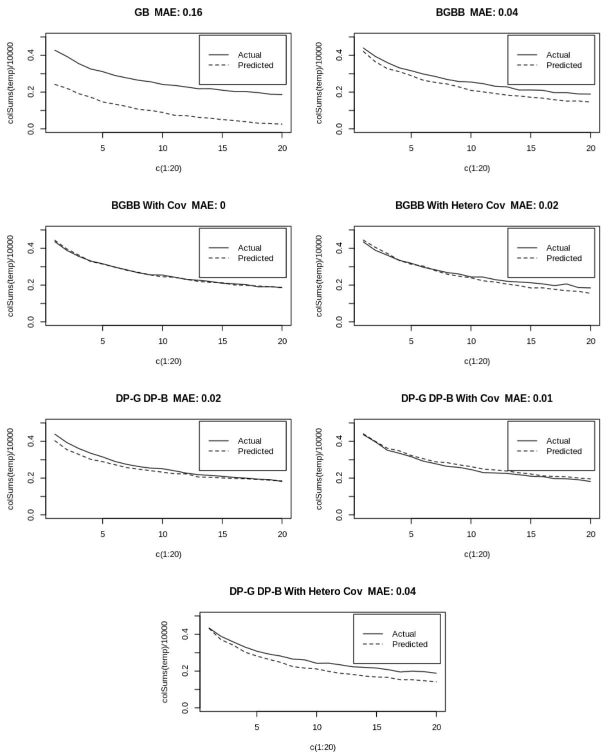


Figure 11: Incremental Tracking Plots for Heterogenous Population

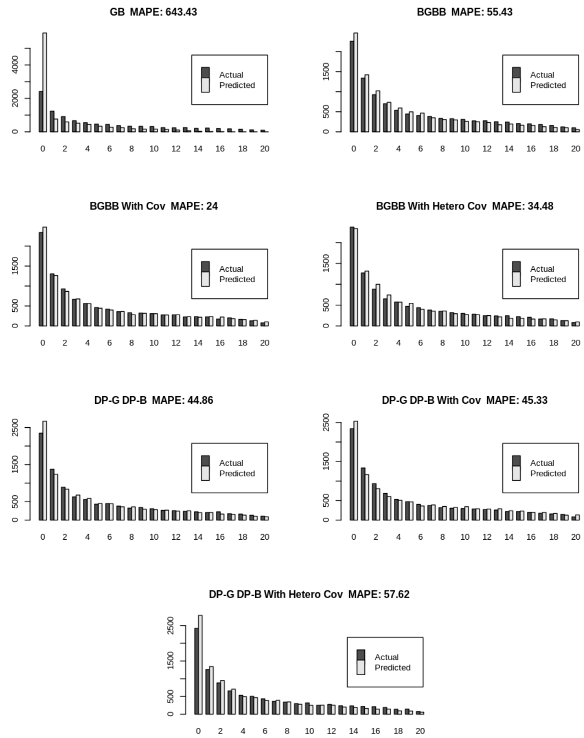


Figure 12: Histogram for Heterogenous Population

Heterogeneous Population with Homogenous Covariate Effect

Model misspecification when a covariate effect is present did seem to introduce some bias to the estimation of the buying propensity. In the graphics below, one can clearly see that for the models without covariate effect (in particular, the BGBB and DP-G DP-B model), the estimated distribution of the buying propensity is inflated upward, whereas for the models where a covariate effect is estimated, the shape is more moderate and closer to the expected spread of the parameter.

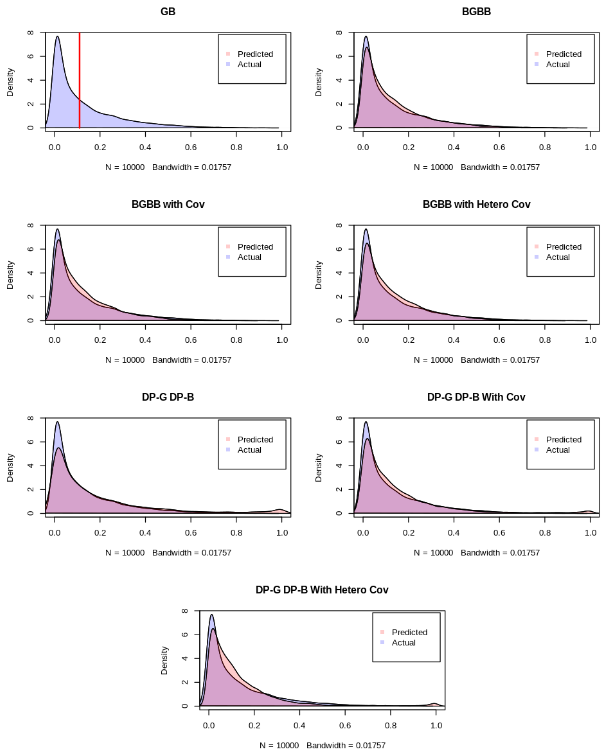


Figure 13: Estimated Dying Propensity for Heterogenous Population with Homogenous Covariate

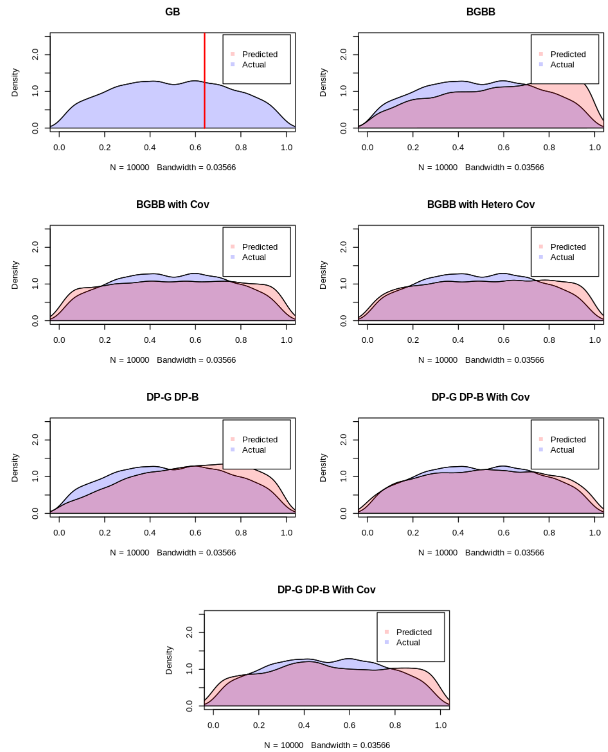


Figure 14: Estimated Buying Propensity for Heterogenous Population with Homogenous Covariate

To gauge how the deviation plays out, two scenarios were simulated using the incremental plot and histogram. The first is where every period is assumed to have both covariate effects. Under such conditions, the incremental plots of the model without the covariate effect showed systematic underestimation of conversion, which naturally also translated to the histograms being overly left-skewed.

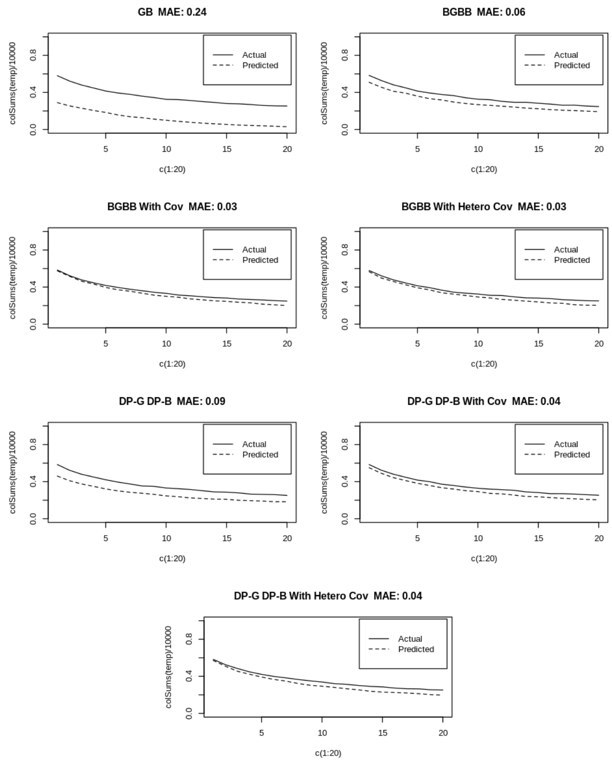


Figure 15: Incremental Tracking Plot for Heterogenous Population with Homogenous Covariate (All Periods Assumed to Have Covariates)

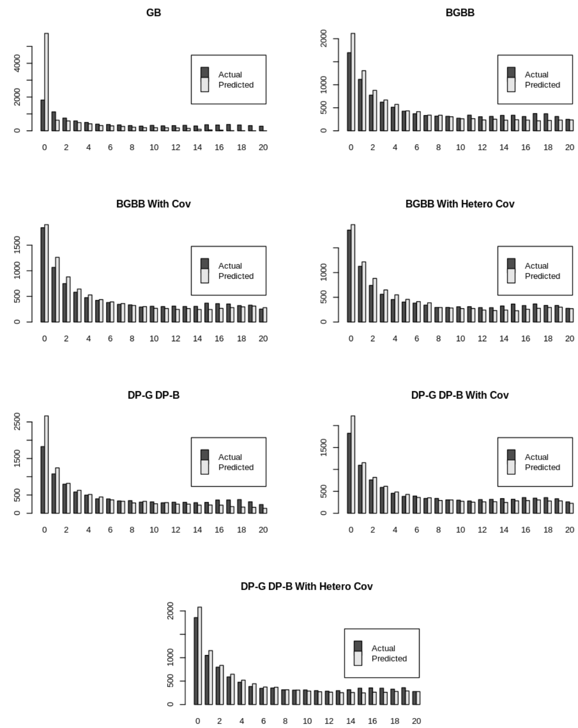


Figure 16: Histogram for Heterogenous Population with Homogenous Covariate (All Periods Assumed to Have Covariates)

The second scenario assumes that the covariate effect is completely taken away. The simulation is now solely dependent on how good the models were at stripping away the non-stationarity introduced by the covariate.

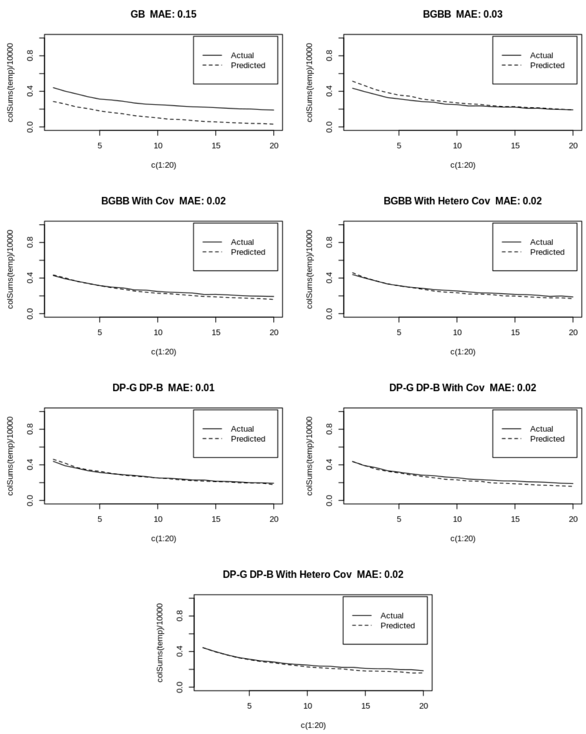


Figure 17: Incremental Tracking Plot for Heterogenous Population with Homogenous Covariate (All Periods Assumed to not Have Covariates)

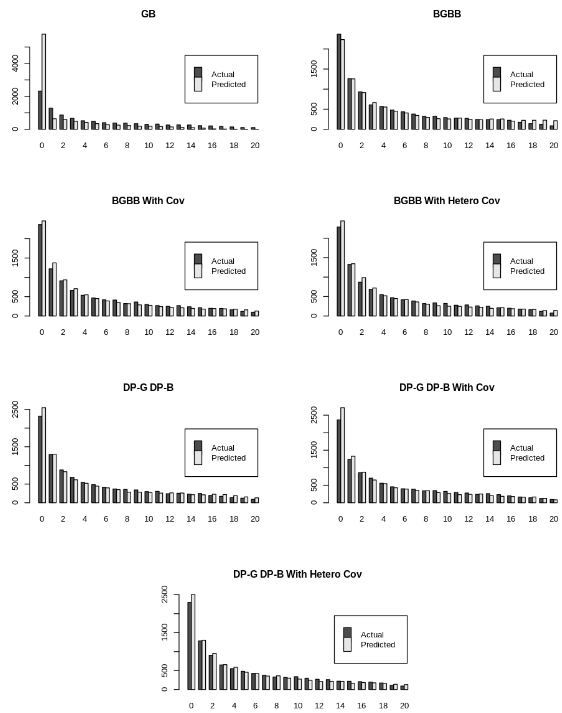
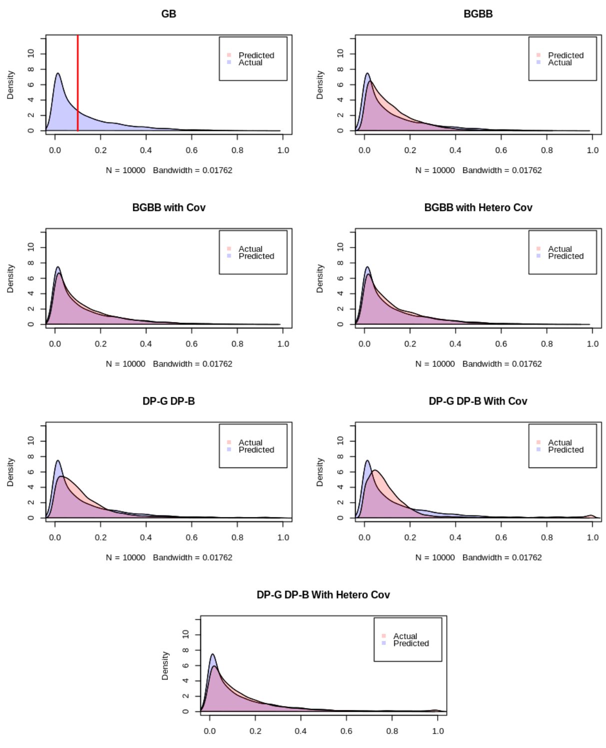


Figure 18: Histogram for Heterogenous Population with Homogenous Covariate (All Periods Assumed to not Have Covariates)

Interestingly, apart from the initial bump for the BGBB and DP-G DP-B model along with the slight trace of over-estimation towards the tail of the histogram, both models were rather adequate at recovering the base adoption levels and spreads albeit not having directly accounted for the covariate effect. That said, the best performing model overall even after the covariate effect is removed still appears to be the BGBB model with a homogenous covariate effect.

Heterogeneous Population with Homogenous Covariate Effect

The overarching picture when a heterogenous covariate effect is present is also rather analogous to the homogenous case. The models without the covariate effect once again shifted the buying propensity estimation unnecessarily to the right.



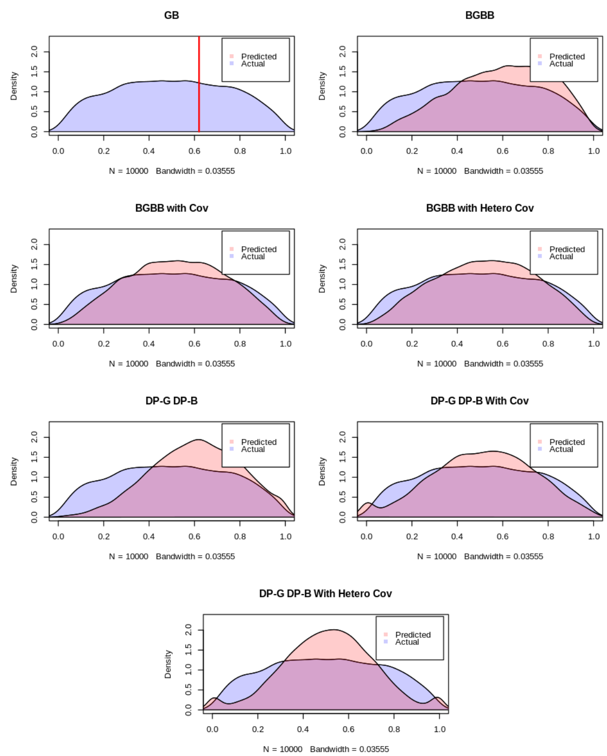


Figure 19: Estimated Dying Propensity for Heterogenous Population with Heterogenous Covariate

Figure 20: Estimated Buying Propensity for Heterogenous Population with Heterogenous Covariate

Using a similar procedure indicated in the homogenous case, the incremental plots and histogram for a scenario where a covariate is assumed to exist throughout or when a covariate is assumed to not exist at all, generally reflected a similar phenomenon. Where, in the former, the model without the covariate effect appeared to be underestimating consistently and, in the latter, the under-specified model performed better but still not as good as the properly specified model.

In this case, one may also additionally compare the BGBB with homogenous covariate model and the BGBB with heterogenous covariate model. While both models are similar when no-covariate effects are present, the BGBB model with heterogenous effect can capture the spread of the histogram better than its counterpart. In other words, it is capable of accounting for the marketing lovers whose propensity moves drastically with marketing action. Similarly, this can be said for the DP-G DP-B model with heterogenous covariate effect.

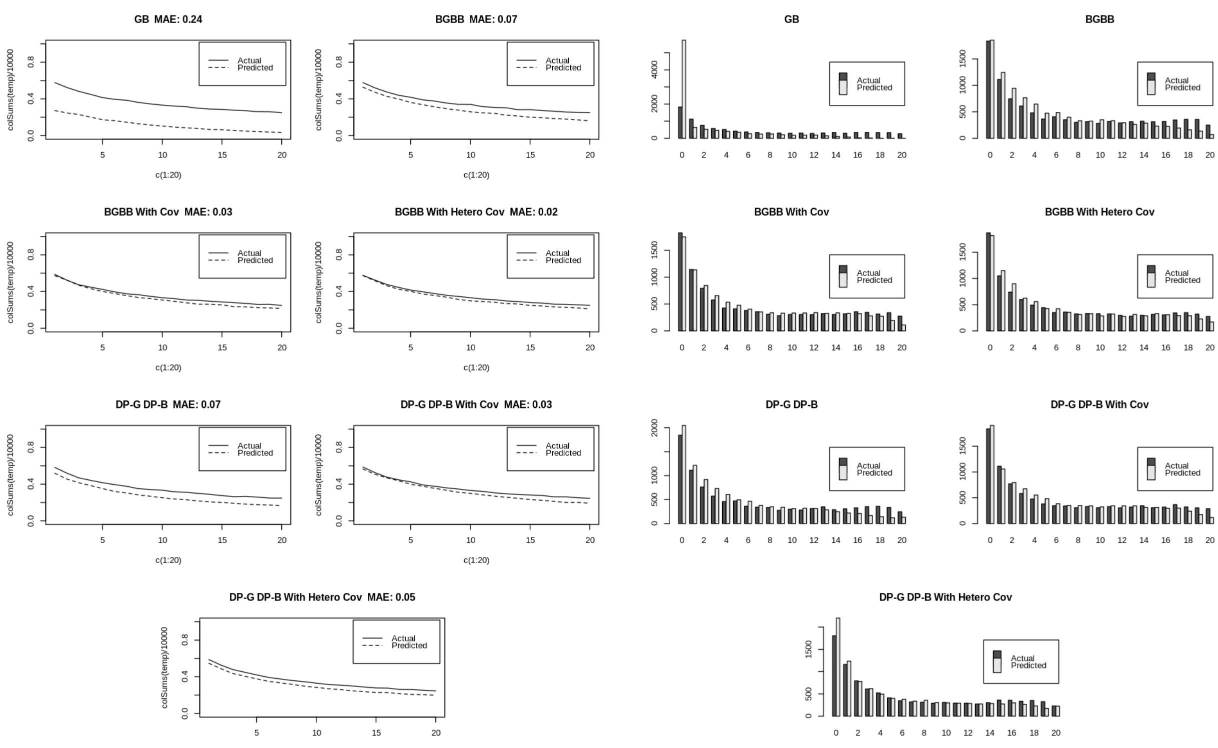
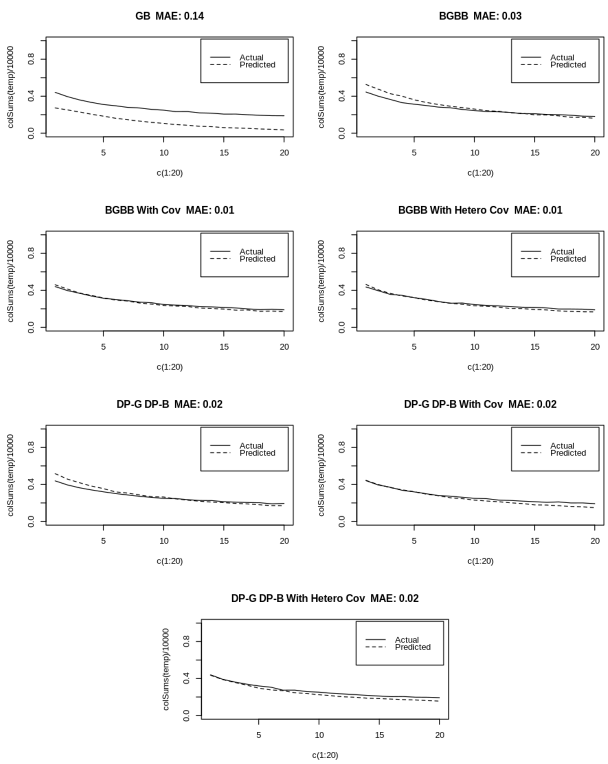


Figure 21: Incremental Tracking Plot for Heterogenous Population with Heterogenous Covariate (All Periods Assumed to Have Covariates)

Figure 22: Histogram for Heterogenous Population with Heterogenous Covariate (All Periods Assumed to Have Covariates)



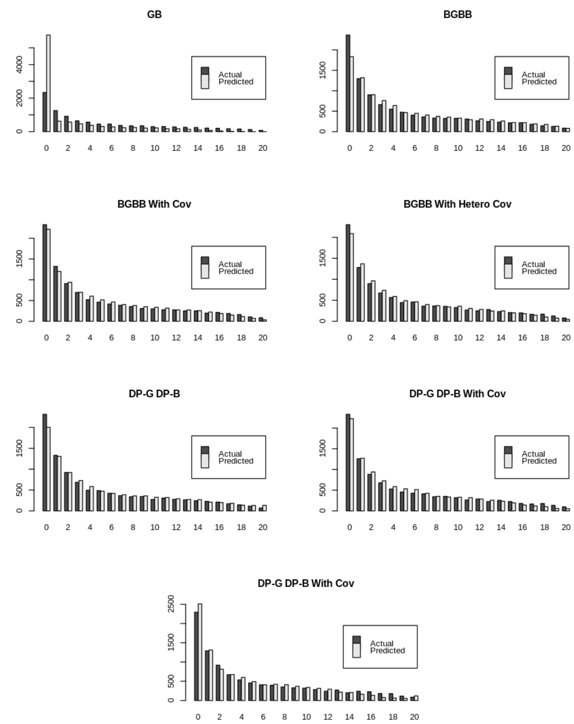


Figure 23: Incremental Tracking Plot for Heterogenous Population with Heterogenous Covariate (All Periods Assumed to not Have Covariates)

Figure 24: Histogram for Heterogenous Population with Heterogenous Covariate (All Periods Assumed to not Have Covariates)

Finite Mixture Population

Given the flexibility of the DP-G DP-B model, the paper further attempts to gauge its efficacy when it is being applied to a population that has bi-modal propensities (often called finite mixtures). Graphically, compared to earlier models such as the BGBB model, the DP-G DP-B model was able to infer the existence off bi-modalities in the data and does an adequate job extracting the location of the modes. Visibly however, the inference is not perfect, while the DP-G DP-B model can infer the modal location, it does a poor job inferring the heights or the volume at the modes.

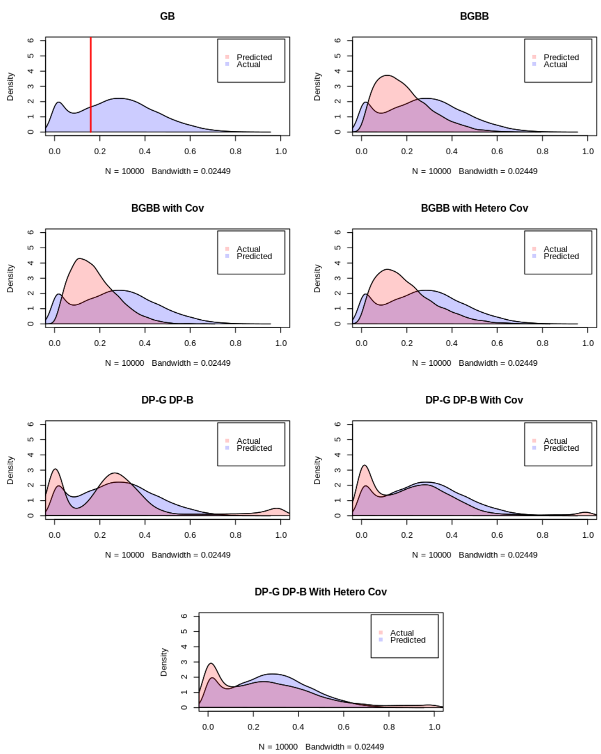


Figure 25: Estimated Dying Propensity for Finite Mixture Population

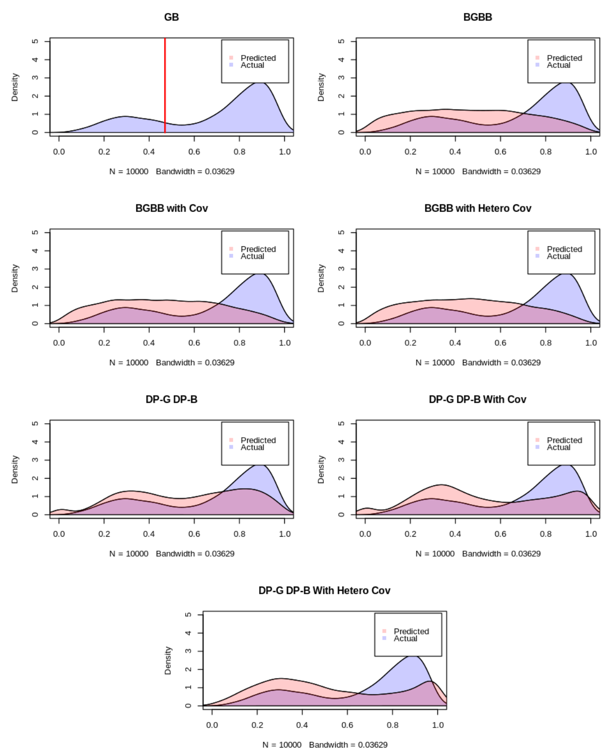
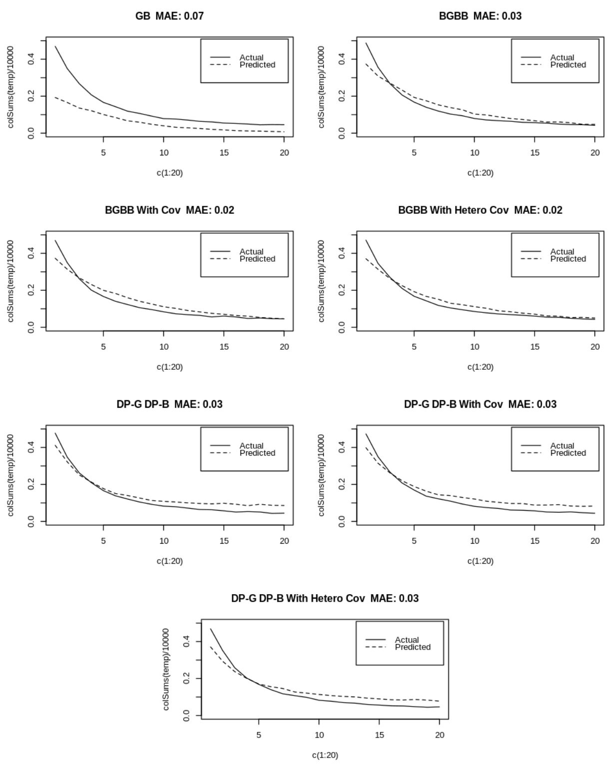


Figure 26: Estimated Buying Propensity for Finite Mixture Population

More interestingly, while the graphs above may suggest that the DP-G DP-B model should offer a superior performance, at least compared to the BGBB models, it actually does not. Examining the incremental plots and histograms suggests that the DP-B DP-G model is generally at par with the BGBB model in terms of aggregate level predictions.



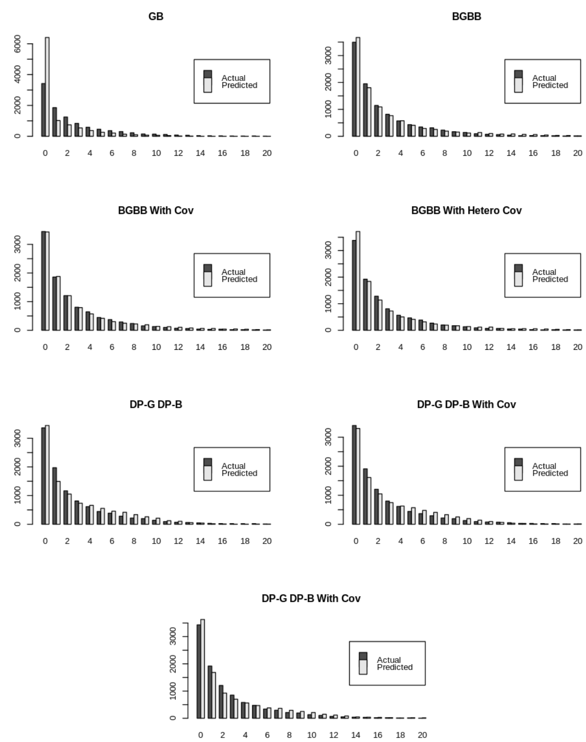


Figure 27: Incremental Tracking Plots for Finite Mixture Population

Figure 28: Histogram for Finite Mixture Population

Breaking the tracking plots down can shed further insights into the root of the issue. One can note that, in the earlier parts of the tracking plot, the DP-G DP-B model appears to perform better, however in the later parts it appears to overestimate the number of purchases that exists.

This can be explained by the fact the DP-G DP-B model appears to all have an over-estimated mode towards the end of distributions (specifically it estimates more individuals to have a low dying propensity).

Finite Mixture Population with Heterogenous Covariate Effect

Adding in covariates appears to impose additional difficulty for the model to infer the varying modality in the dataset. As seen below, while the DP-G DP-B model can still do a satisfactory job in distinguishing modality within the dying propensities, it fails to do so in the buying propensity (where the covariates are being introduced). That said, the model with the lack of covariate specification still causes the entire buying propensity distribution to shift rightward as seen before.

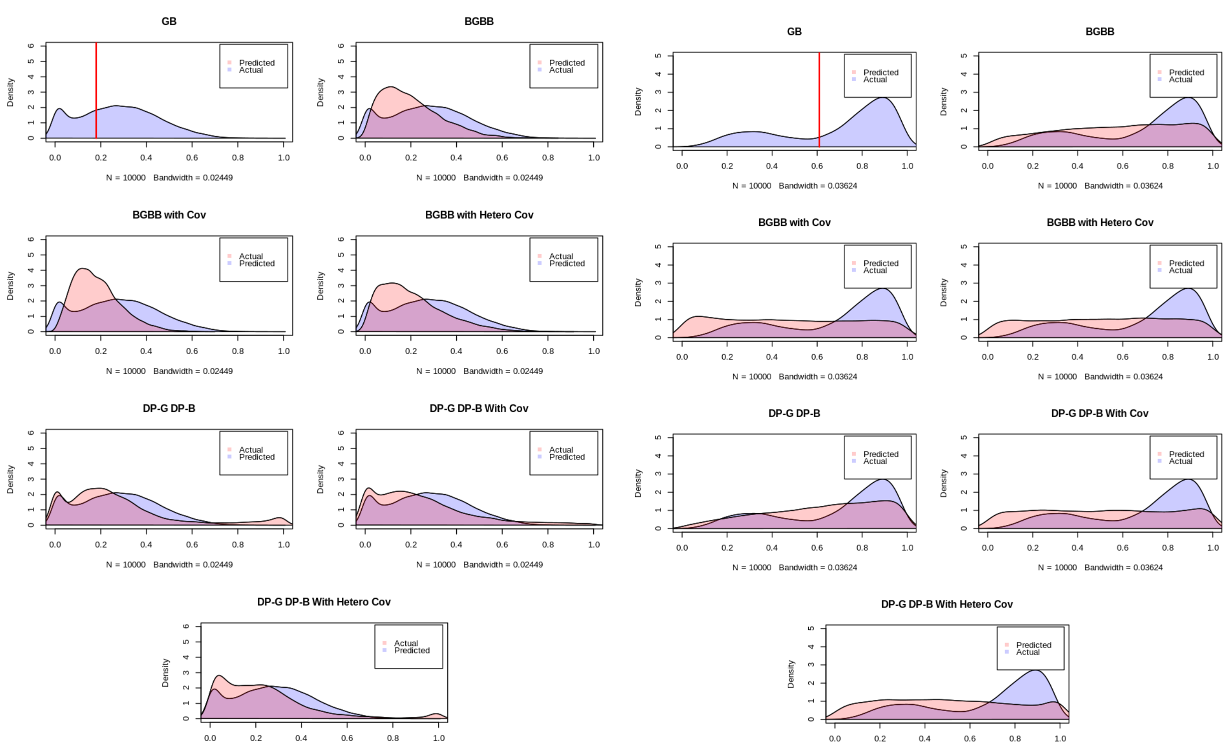
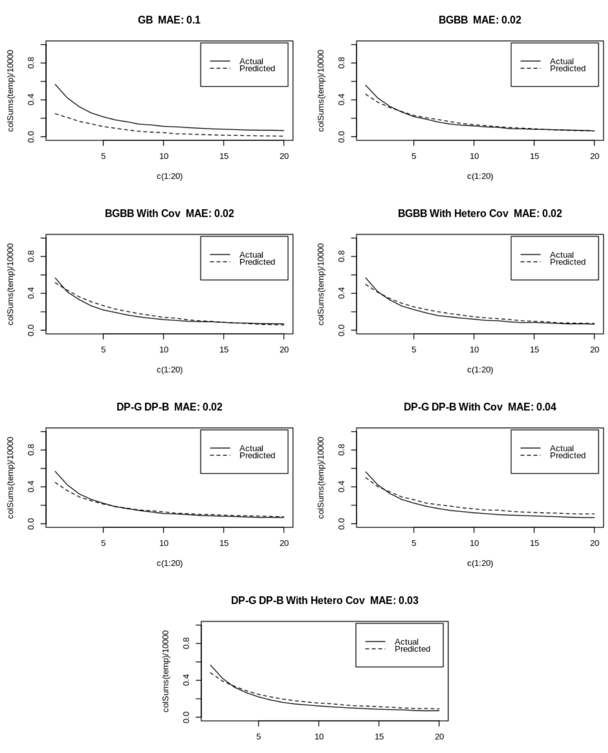


Figure 29: Estimated Dying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

Figure 30: Estimated Buying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

Like the previous case where a covariate effect is not included. The DP-G DP-B model does not seem to offer much direct benefits in terms of the aggregate level forecasting both when the covariate effect exists or presumed to not exist in the holdout. The same problem of over-estimation towards the tail-end of the distribution does seem to exist. While the DP-G DP-B model with covariate specification outperforms its counterpart (vanilla DP-G DP-B model) when covariate is included, it does not seem to be able to outperform the BGBB model with covariates.



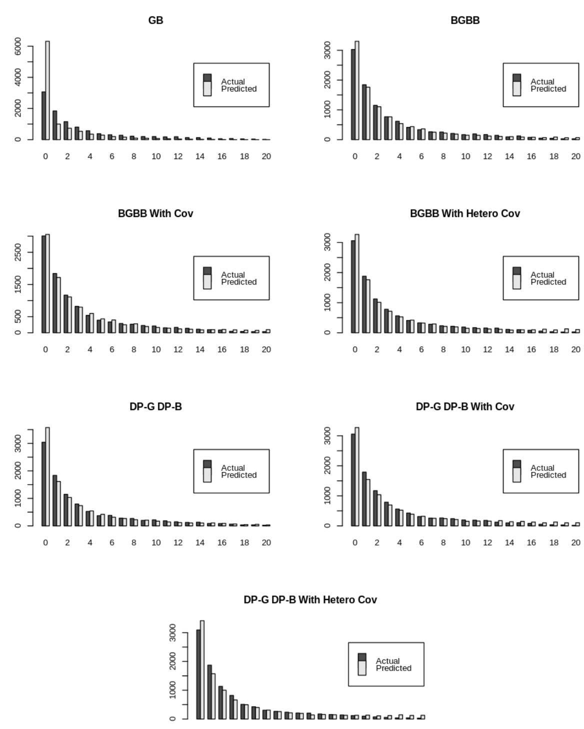


Figure 31: Estimated Dying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

Figure 32: Estimated Buying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

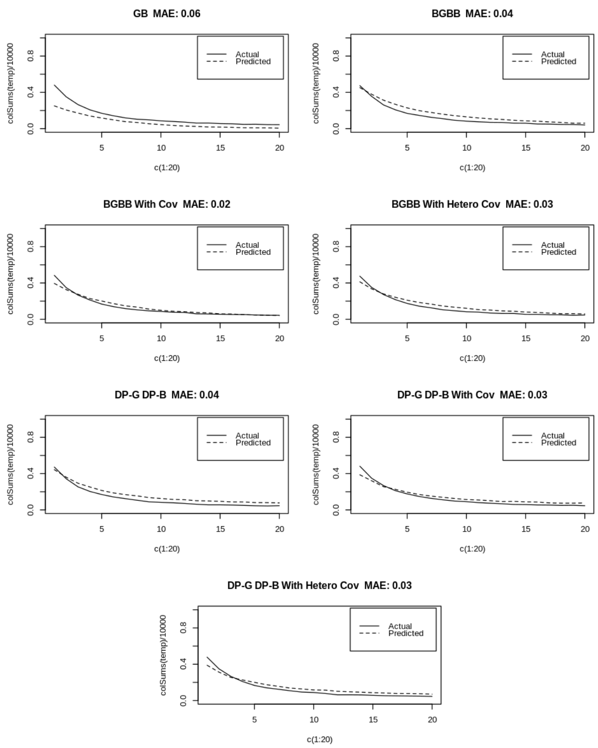


Figure 33: Estimated Dying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

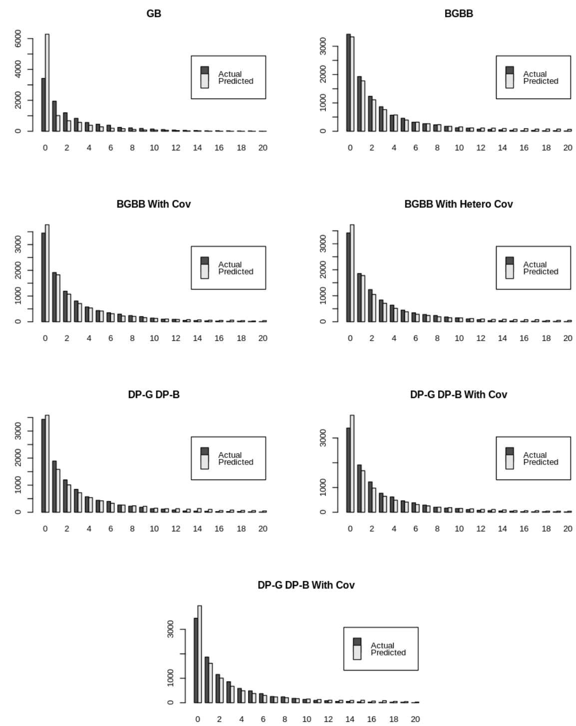


Figure 34: Estimated Buying Propensity for a Finite Mixture Population with Heterogenous Covariate Effect

Conditional Expectation

Moving from aggregated metrics to individualistic metrics, a table of the RMSE and Correlation between actual purchases of an individual and his/her conditional expectations from the different models were calculated.

Homogeneous Population

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	1.719366	1.440016	1.412318	1.360742	1.365648	1.457408	1.407426
Correlation	0.8480558	0.8691005	0.8785887	0.8799905	0.8745955	0.866822	0.8768652

Figure 35: Conditional Expectation Homogenous Population

Looking at the conditional expectation table, the results show that the more flexible models performed better than the homogenous model. Across all flexible models, the performance was rather alike. The added parameters required by the DP models and the models with covariate did not appear to hurt the model's ability to extract individual estimates.

Heterogeneous Population

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	2.442325	1.411642	1.514979	1.581499	1.294719	1.513064	1.561295
Correlation	0.8591819	0.9412351	0.9374091	0.9369156	0.9390811	0.9398021	0.9397667

Figure 36: Conditional Expectation Heterogenous Population

The conditional expectations in a heterogenous case show a similar trend whereby the flexible model in general was able to cut down prediction error by about 40% compared to the homogenous model. The DP model once again had the smallest error in this case, but all flexible models are similar in terms of their correlation.

Heterogeneous Population with Homogenous Covariate Effect

The two-part approach highlighted earlier is once again deployed to analyze how models with and without the covariate effects lines up with the actual results.

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	2.843113	2.319876	2.141549	2.142731	2.292693	2.113183	2.16437
Correlation	0.8953664	0.8639988	0.8715751	0.8746352	0.8569674	0.8701586	0.8668317

Figure 37: Conditional Expectation Heterogeneous Population with Homogeneous Covariate Effect (All Periods Assumed to Have Covariates)

In the case where all periods are assumed to have a covariate, we can see that the conditional expectation of a model with covariate is much better. This is consistent with what was shown in the incremental plots and histograms.

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	2.130536	1.659542	1.778731	1.728162	1.652276	1.743756	1.801642
Correlation	0.8645569	0.8663232	0.8717127	0.8783012	0.8610834	0.8703882	0.8672899

Figure 38: Conditional Expectation Heterogeneous Population with Homogeneous Covariate Effect (All Periods Assumed to not Have Covariates)

However, in the case where all periods are assumed to not have a covariate, the conditional expectation trend reverses. The paper notes that the models where a covariate effect is explicitly included, performed worse at extrapolating the conditional expectation.

This seems at odds with the results from earlier, however such trends could be explained as a fashion of the model shrinkage. When a regression was performed using the conditional expectation derived from the vanilla BGBB model on the conditional expectation derived from BGBB model with covariates, the coefficient on the former is 0.899, with an intercept of -0.049. This shows that as a result of the covariate, a greater deal of shrinkage was applied to the individual posterior distribution.

This can be further confirmed as one map out the RMSE of the conditional expectation to the number of actual purchases an individual has made.

Actual Purchase	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
0	0.1201433	0.1765802	0.1237080	0.1271203	0.1778210	0.1267892	0.1267261
1	0.6328762	0.6191557	0.5245666	0.5266561	0.6363961	0.5536761	0.5950663
2	1.4639951	1.4666284	1.2692820	1.2603956	1.5407170	1.3107225	1.2461962
3	0.6622704	1.2831748	1.1807558	1.1584580	1.3073565	1.1360962	1.1236677
4	1.1720362	1.2620026	1.4566571	1.3882110	1.2335151	1.4978200	1.2687121
5	2.0087548	1.6194656	1.7914512	1.7611822	1.7448578	1.7314956	1.7786899
6	2.6888763	1.9190891	2.3203265	2.3294735	1.8849135	2.2206763	2.3637779
7	3.7972386	2.6742871	3.1378772	2.9692742	2.7056807	3.0042329	3.1672809
8	4.9795783	2.4219743	3.2187693	3.0160751	1.9950378	2.9686381	3.2817686

9	5.4158333	2.9066667	3.3291667	3.0675000	2.6141667	3.1016667	3.2600000
10	6.5575000	3.7475000	4.0891667	3.9833333	3.5300000	3.8900000	4.3800000

Figure 39: Conditional Expectation Heterogeneous Population with Homogeneous Covariate Effect RMSE

One can note that while the model with covariate has a smaller error compared to the models without covariates in the case where actual purchases are small, they have bigger errors when the actuals purchases are higher. In other words, the high buying propensity individual's parameter was dragged down towards the population more in the case where the model has a covariate.

Heterogeneous Population with Heterogenous Covariate Effect

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	2.943017	2.167765	1.846482	1.868942	2.069687	1.923118	1.833032
Correlation	0.878269	0.9051894	0.9052252	0.9003293	0.9069125	0.9099661	0.900733

Figure 40: Conditional Expectation Heterogeneous Population with Heterogeneous Covariate Effect (All Periods Assumed to Have Covariates)

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	2.510003	1.635307	1.749004	1.798962	1.512732	1.823507	1.777798
Correlation	0.9014604	0.9467363	0.9494087	0.9423995	0.9509425	0.9507007	0.9460763

Figure 41: Conditional Expectation Heterogeneous Population with Heterogeneous Covariate Effect (All Periods Assumed to not Have Covariates)

One notes a similar case when the covariate effect is heterogenous. In the event where covariates were assumed to be present for all the next periods, the conditional expectation of the models with covariate is lower, whereas in the case where there were all covariates were set to zero, the model with covariate specification has a higher error. The outlier to this rule was the DP-G DP-B model that heavily mis-specified the dying propensity.

Actual Purchase	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
0	0.1519867	0.2124084	0.1967911	0.2072406	0.2119224	0.1847676	0.2025492
1	0.9618575	0.7810794	0.8179763	0.7305601	0.7179148	0.6829032	0.7376075
2	1.1842742	0.7994661	0.9412082	0.8931397	0.8730521	0.9972347	0.8457544
3	1.7744037	1.3441695	1.6142952	1.5019506	1.3410309	1.6725185	1.4936687
4	1.6490247	1.1544075	1.5289808	1.6089570	1.1599600	1.8939617	1.4952576
5	2.0602090	1.4465055	1.7508411	1.7325126	1.3962849	2.1927845	1.8031171
6	2.5855598	1.7091363	2.0457002	2.0658841	1.7679790	2.5141580	1.9462954
7	3.7301549	2.3787095	2.6928943	2.8052084	2.3254803	3.2326982	2.8047324
8	4.6530571	2.7493220	2.9238034	3.0100608	2.6144867	3.5391391	3.1396423
9	5.4876826	3.1395974	3.2660863	3.2998524	2.9035924	3.9804337	3.5243004
10	6.4400000	3.3161111	3.4261111	3.6066667	2.2600000	3.6055556	3.3744444

Figure 42: Conditional Expectation Heterogeneous Population with Heterogeneous Covariate Effect RMSE

The shrinkage pattern noted earlier was also present in this case, where the models with covariate effect again performed worse when actual purchase is larger.

Finite Mixture Population

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	1.950064	1.91723	1.902247	1.904884	1.886151	1.870595	1.836944
Correlation	0.8782898	0.8183846	0.8178485	0.7994399	0.7200887	0.7436967	0.7353538

Figure 43: Conditional Expectation Finite Mixture Population

In the finite mixture case, the DP-G DP-B model overperformed all other models in terms of RMSE but underperformed compared to all other models in terms of the out of sample correlation. This could be explained since the DP-G DP-B model shrinks the posterior distribution of an individuals to their closest mode thereby homogenizing groups of individuals propensities further. Such homogenization may lead to better prediction performance due to its shrinkage but could also lessen the variation across individuals thereby decreasing the correlation.

Finite Mixture Population with Heterogenous Covariate Effect

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate

RMSE	1.935816	1.815185	1.685702	1.727486	1.819086	1.720643	1.72891
Correlation	0.8580906	0.784386	0.8289517	0.7919509	0.7696435	0.759704	0.7573393

Figure 44: Conditional Expectation Finite Mixture Population with Heterogeneous Covariate Effect (All Periods Assumed to Have Covariates)

	Geometric Binomial Model	BGBB Model	BGBB Model with Homogenous Covariate Effect	BGBB Model with Heterogeneous Covariate Effect	DP-B DP-G Model	DP-B DP-G Model with Homogenous Covariate	DP-B DP-G Model with Heterogenous Covariate
RMSE	1.935816	1.815185	1.902253	1.946367	1.819086	1.914204	1.906634
Correlation	0.8580906	0.784386	0.8022488	0.7887746	0.7696435	0.7543559	0.7708036

Figure 31: Conditional Expectation Finite Mixture Population with Heterogeneous Covariate Effect (All Periods Assumed to not Have Covariates)

Analyzing the finite mixture case with covariates reveals that even in a case where we presume there to be modalities in the distribution, the effect of covariates significantly outweighs the effect of a flexible prior in the context of individual level prediction. Like the earlier cases, the models with covariate (regardless of whether its BGBB or DP-G DP-B) performs much better than its vanilla counterpart when the covariate effect is presumed to exist, and worse when the covariate effect is absent (attributable to the shrinkage effect mentioned earlier).

Comparing the DP-G DP-B model to the BGBB model, it seems like the most parsimonious BGBB model with homogenous covariate offers superior performance than its counterpart. While more simulation-based analysis may need to be done to offer conclusive finding, this does shed light in how extremely flexible models may not necessarily to achieve good model performance.

Marketing Simulation

Going back to the functional form of how the covariates are included, the paper notes that the form used, being analogous to logistic regression, implies that the same covariate will vary in effectiveness for different individuals. Much like the sigmoid, the functional form dictates that the effect will be stronger for individual whose propensities hovers around 0.5 whereas it will basically be ineffective for those whose propensities are extremely low of extremely high. Principally, such effect makes sense as well, for individuals who are extremely likely to buy a product, the marketing campaign will be useless to them. The same also holds for the low propensity individuals.

Under this condition, this means that not all marketing efforts should be made equal. The paper assumes, as earlier, that there exists a homogenous covariate effect for all individuals. In the simulation, the paper introduces two covariates, both can be thought of as marketing campaigns rendered by a firm. The two campaigns vary in effectiveness, with the first having a coefficient of 0.6 on an individual's buying propensity and the second having a coefficient of 0.3. Due to the disparity in the effectiveness of the campaign, the two campaigns also have different costs. The first campaign cost $14 dollars per individual whereas the second campaign cost $6 dollars. Lets further assume whenever an individual converts, he/she brings in a fixed amount of $100. For the purpose of the simulation a discount rate of 0.1 per period will be used.

To establish some benchmarking, a couple of scenarios were first simulated.

- Scenario 1: Firm sends both marketing campaign to an individual if the probability that an individual is alive is greater than 0.5.
- Scenario 2: Firm does not send any marketing campaign at all
- Scenario 3: Firm sends only the $14 marketing campaign to an individual if the probability that an individual is alive is greater than 0.5.
- Scenario 4: Firm sends only the $6 marketing campaign to an individual if the probability that an individual is alive is greater than 0.5.

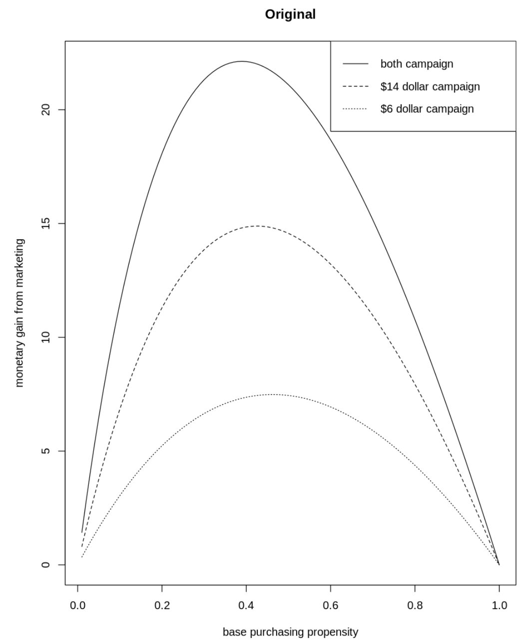
The parameters used to simulate the dying propensities are the posterior distribution obtained from the vanilla BGBB model.

The results can be found in the table below:

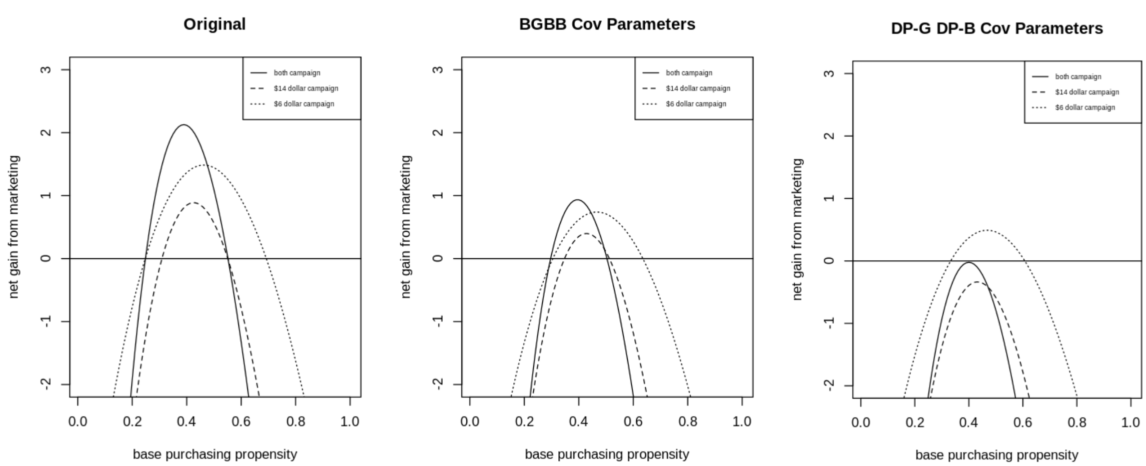
Scenarios	Quantiles	Net Present Value of an Individual's contribution (discounted conversion value less marketing expenses)
Scenario 1	**Mean**	**109.955541404643**
	0% Percentile	92.6653165368367
	25% Percentile	106.578594189838
	50% Percentile	110.726396780109
	75% Percentile	113.614003099488
	100% Percentile	121.456948207978
Scenario 2	**Mean**	**116.586431462874**
	0% Percentile	102.13460740616
	25% Percentile	113.427281826067

	50% Percentile	116.6062990329
	75% Percentile	119.745068037335
	100% Percentile	129.79965402895
Scenario 3	**Mean**	112.353258599838
	0% Percentile	95.7806332783563
	25% Percentile	109.411826560278
	50% Percentile	112.544822762004
	75% Percentile	115.329098084899
	100% Percentile	125.147052394333
Scenario 4	**Mean**	117.63622211443
	0% Percentile	101.25029595083
	25% Percentile	113.703745434894
	50% Percentile	117.640978529918
	75% Percentile	121.555979866937
	100% Percentile	132.237725843504

To reflect the point of how the ROI of a marketing campaign (even under a homogenous condition) is different depending on the innate propensity of an individual, the graphics below shows the true increase in purchasing propensity (using the 0.6 and 0.3 parameters) given the different form of marketing campaign ran scaled up by the fixed payment of 100.



One can obtain the net effect of such a marketing campaign by subtracting the cost of the campaign from the monetary gain. To gauge the efficacy of our model in this setting, the net effect graphics were also generated using the estimated covariate effect obtained from the BGBB with homogenous covariate effect model and the DP-G DP-B with homogenous covariate effect model.



Note that using the net effect graphics, a frontier emerged as to what the optimal marketing policy for individuals at the who has differing buying propensity. One can further note that while there exist discrepancies across models with regards to the exact monetary value of each policy, each model was able to identify the $6 dollar marketing campaign as an implementable policy with the BGBB with covariate model, further identifying a 2-campaign approach analogous to the simulated result from the original model.

Using the boundaries identified above, a simulation is ran with the targeting policy of both the BGBB with homogenous covariate model and the DP-G DP-B with homogenous covariate model. The posterior distribution used to gauge an individual's probability of being alive were also from these models respectively.

Scenarios	Quantile	Net Present Value of an Individual's contribution (discounted conversion value less marketing expenses)
BGBB With Cov	**Mean**	119.130883672808
	0% Percentile	105.973969392344
	25% Percentile	115.215392234813
	50% Percentile	119.55656589691
	75% Percentile	122.781467463989
	100% Percentile	136.069683103909
DP-G DP-B With Cov	**Mean**	119.536046163425
	0% Percentile	105.524181835381
	25% Percentile	115.911177039566

	50% Percentile	119.632257746352
	75% Percentile	123.503792216649
	100% Percentile	133.890119962238

As noted above, the targeting policies obtained from the model were both able to beat the best blanket strategies above by 2 dollars per individual. It is worth noting however, that the targeting strategy also greatly depends on the ability of the model to estimate whether an individual is alive, which may have been why DP-G DP-B model was able to obtain a slightly higher ROI despite having a worse targeting tactics.

DISCUSSION

Through the model and scenarios surveyed, one can note a couple of key takeaways. Firstly, the heavily parametrized models particularly the BGBB with covariate model, the DP-G DP-B model and the DP-G DP-B model with covariates, were all rather stable at recovering the parameters of rather simple scenarios. There were minimal signs of overfitting and the models were all rather desirable at predicting both the aggregated and individual level trend. Furthermore, in cases where a multi-modal distribution is presumed for individuals the DP-G DP-B model seems to extract conditional expectations with less error.

However, comparing across the heavily specified model, the DP-G DP-B model class does appear to be less stable more often. As seen through the simulated example, the DP based models tend to have deviations in estimating the parameters of interest more frequently than the parametric models. This may partially be a result of the small sample (200 individuals) that were provided to the model but also the relatively short MCMC iteration. Nonetheless, the heavier parametrization of the model could have also contributed to this phenomenon.

In addition to that, in contrary to expectation, while the DP-G DP-B model can infer the presence of modalities in propensities, it offers rather limited improvement in terms of the aggregated level predictions and forecasting due to issues such as overfitting (particularly at the tail-end of the distributions). In the individual case, while the DP-G DP-B model produces results with less error, the prediction seems to correlate less with the actual result, which may be a cause of concern.

Secondly, model misspecification when a covariate effect is missing leads to a biased parameter estimates where in the presence of a positive covariate will lead to the buying propensity being overly inflated. In the case where such a covariate is present in future occurrences, this will cause the model to under-predict the number of conversions on both an individual and aggregated context.

Furthermore, even when the covariate is no longer present in the future, the upward bias of the parameter will still cause the model to underperform on aggregate. The model will overstate how many high buying propensity customers there is out there and overstate the initial adoption.

The covariate effects extracted can also be further used to perform consumer targeting. The paper experimented with the case where a homogenous covariate effect is assumed and found that by using the covariate effects in conjunction with the posterior distribution of an individual's propensities, firms can better target who to offer marketing activities to and boost their revenue per individual. Optimization done in this manner outperforms blanked based approach based on our simulation.

It is worth noting, however, that such a covariate-based model also comes with a trade-off of its own. When such covariate effects are included, the posterior distribution of individuals will be pulled more towards the mean since the effect of the covariate will account for some of the conversion. This means the posterior update process will be slower overall for individuals which may lead to the model performing worse when projecting out conditional expectation for high propensity individuals. That said, with enough observation, this issue should cease to exist.

Thirdly, while the combination of the DP prior and covariates offers a good theoretical model that can be extremely flexible in various cases, the simulation revealed that such model may offer rather little benefits in both aggregate and individual level predictions. Due to the presence of covariate effects the DP model often fails to identify the modalities in the buying propensities. This coupled with the fact that the DP model is already prone to overfitting issues causes the predictions to generally be about the same quality as the other models, while requiring more training time.

CONCLUSIONS AND NEXT STEPS

The models presented in this paper offers another dive at enriching the customer lifetime value modelling toolkits. However more work could be done to further explore the efficacy of this model. A natural extension for this paper would be to apply this model to an empirical dataset to test how useful such model is in a real-life setting.

In addition to added level of robustness check, more exploration could also be done at the point of incorporating correlations across model parameters. Past work (Fader, Hardie and Shang, 2009) have found that such correlation exists which can further boost the flexibility of the model and make it even more generalizable.

The covariate effect presented in this paper can also be further explored via marketing-based simulation. For instance, with the estimated covariate effect, one can devise a dynamic marketing strategy that takes advantage of the different level of customer sensitivity to marketing along with how effective the marketing is relative to an individual's innate parameters.

Lastly, covariate integration can also be attempted at the attrition level, as indicated earlier, certain marketing methods such as loyalty program is aimed at lowering the attrition propensity of individuals. Covariate integration at the attrition level could be more effective at accounting for these intricacies.

APPENDIX

Incorporating Covariates

To include the covariates in a non-binarized case where the magnitude of the covariates are considered one can imagine a diagram as follow:

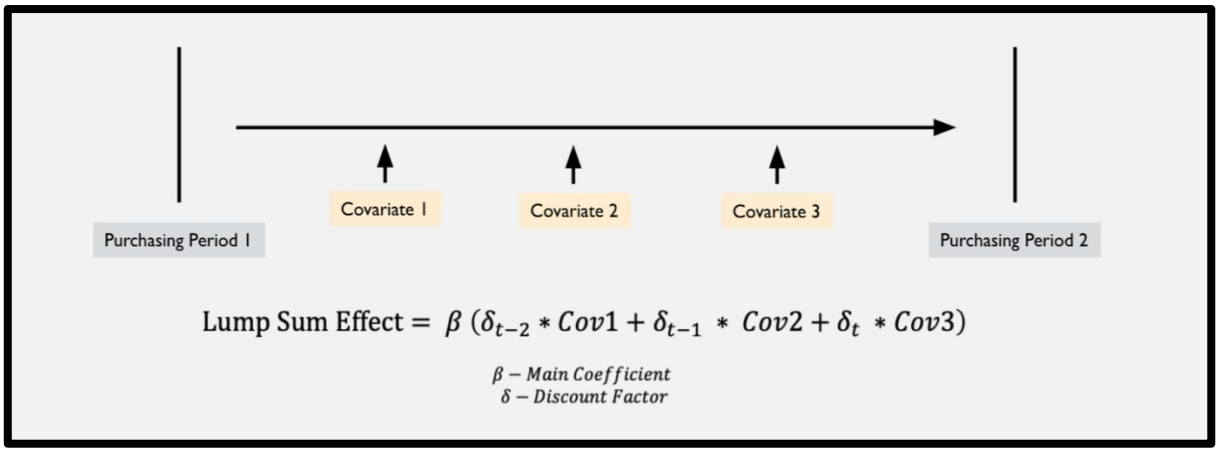


Figure 33: Visual Illustration of Covariate Effect

In such a continuous case however, the **X** vector used in the likelihood calculation would depend on the discount functions chosen. For instance, if straight-line discounting is used, then the formulation will be as follows:

$$X_{Cov} = \frac{t_1}{T} Cov_{t1} + \cdots + \frac{t_n}{T} Cov_{tn}$$

Where Cov_{t1} ... Cov_{tn}, are the different values of a covariate that were observed between the previous period and this, t_1 ... t_n represents the time when the covariates occurred and let **T** be the time between the previous period and the next.

Note that the discounting is necessary here as if one imagine a truly discrete buying process, one expects the recency effect to kick in closer to the period when the purchase can happen. However, if this model was applied to the discretized yearly retention case, one may reasonably think that discounting should not be used as covariate at every period should increase the customer's propensity to purchase in a specific period.

Model Parameters

	GB	BGBB	BGBB with Homogenous Cov Effect	BGBB with Heterogenous Cov Effect	DP-B DP-G Model (3 clusters)	DP-B DP-G Model with Homogenous Covariate (3 clusters)	DP-B DP-G Model with Heterogenous Covariate (3 clusters)
Homogenous Scenario	*p ~ 0.13* *q ~ 0.49* B_1 *~ 0* B_2 *~ 0*	*p ~ Beta* *(1.49,7.43)* *q ~ Beta* *(61, 64.55)* B_1 *~ 0* B_2 *~ 0*	*p ~ Beta* *(1.39,6.9)* *q ~ Beta* *(24.48, 24.07)* B_1 *~ -0.07* B_2 *~ -0.07*	*p ~ Beta* *(1.18, 5.52)* *q ~ Beta* *(23.8, 25.12)* B_1 *~ Normal* *(0.02, 0.28)* B_2 *~ Normal* *(-0.03, 0.11)*	*p | seg 1 ~ Beta* *(2.38, 2.81)* *p | seg 2 ~ Beta* *(0.55, 1.09)* *p | seg 3 ~Beta* *(1.42, 7.77)* *q | seg 1 ~ Beta* *(7.63, 5.5)* *q | seg 2 ~ Beta* *(0.59, 1.11)* *q | seg 3 ~ Beta* *(28.27, 28.76)* *P (seg 1) ~ 0.08* *P (seg 2) ~0.03* *P (seg 3) ~ 0.89* B_1 *~ 0* B_2 *~ 0*	*p | seg 1 ~ Beta* *(1.54, 8.17)* *p | seg 2 ~ Beta* *(1.94, 2.19)* *p | seg 3 ~ Beta* *(1.83, 1.47)* *q | seg 1 ~ Beta* *(36.44, 35.58)* *q | seg 2 ~ Beta* *(0.65, 1.46)* *q | seg 3 ~ Beta* *(0.44, 0.36)* *P (seg 1) ~ 0.96* *P (seg 2) ~0.02* *P (seg 3) ~ 0.02* B_1 *~ -0.06* B_2 *~ -0.06*	*p | seg 1 ~ Beta* *(1.66, 1.91)* *p | seg 2 ~ Beta* *(0.1, 0.21)* *p | seg 3 ~ Beta* *(1.83, 1.47)* *q | seg 1 ~ Beta* *(0.22, 0.41)* *q | seg 2 ~ Beta* *(1.27, 1.97)* *q | seg 3 ~ Beta* *(38.27, 37.51)* *P (seg 1) ~ 0.04* *P (seg 2) ~0.03* *P (seg 3) ~ 0.93* B_1 *~ Normal* *(-0.07, 0.32)* B_2 *~ Normal* *(-0.06, 0.46)*

Heterogenous Scenario	$p \sim 0.11$ $q \sim 0.54$ $B_1 \sim 0$ $B_2 \sim 0$	$p \sim Beta$ $(0.85, 6.97)$ $q \sim Beta$ $(1.21, 1.4)$ $B_1 \sim 0$ $B_2 \sim 0$	$p \sim Beta$ $(0.46, 2.83)$ $q \sim Beta$ $(1.54, 1.42)$ $B_1 \sim -0.39$ $B_2 \sim -0.15$	$p \sim Beta$ $(0.81, 6.68)$ $q \sim Beta$ $(1.26, 1.27)$ $B_1 \sim Normal$ $(-0.32, 0.17)$ $B_2 \sim Normal$ $(-0.14, 0.71)$	$p \mid seg\ 1 \sim Beta$ $(0.42, 3)$ $p \mid seg\ 2 \sim Beta$ $(0.61, 1.1)$ $p \mid seg\ 3 \sim Beta$ $(0.38, 0.35)$ $q \mid seg\ 1 \sim Beta$ $(1.47, 1.7)$ $q \mid seg\ 2 \sim Beta$ $(0.44, 0.33)$ $q \mid seg\ 3 \sim Beta$ $(0.76, 0.5)$ $P\ (seg\ 1) \sim 0.9$ $P\ (seg\ 2) \sim 0.05$ $P\ (seg\ 3) \sim 0.05$ $B_1 \sim 0$ $B_2 \sim 0$	$p \mid seg\ 1 \sim Beta$ $(0.51, 3.46)$ $p \mid seg\ 2 \sim Beta$ $(0.86, 0.92)$ $p \mid seg\ 3 \sim Beta$ $(0.09, 0.3)$ $q \mid seg\ 1 \sim Beta$ $(2.01, 1.83)$ $q \mid seg\ 2 \sim Beta$ $(0.4, 0.44)$ $q \mid seg\ 3 \sim Beta$ $(0.2, 0.2)$ $P\ (seg\ 1) \sim 0.9$ $P\ (seg\ 2) \sim 0.04$ $P\ (seg\ 3) \sim 0.62$ $B_1 \sim -0.38$ $B_2 \sim -0.14$	$p \mid seg\ 1 \sim Beta$ $(0.07, 0.24)$ $p \mid seg\ 2 \sim Beta$ $(1.06, 2.24)$ $p \mid seg\ 3 \sim Beta$ $(1.13, 8.77)$ $q \mid seg\ 1 \sim Beta$ $(1.41, 1.5)$ $q \mid seg\ 2 \sim Beta$ $(1.39, 1.24)$ $q \mid seg\ 3 \sim Beta$ $(2.14, 1.87)$ $P\ (seg\ 1) \sim 0.15$ $P\ (seg\ 2) \sim 0.22$ $P\ (seg\ 3) \sim 0.64$ $B_1 \sim Normal$ $(-0.36, 0.3)$ $B_2 \sim Normal$ $(-0.15, 0.41)$
Heterogenous Scenario with Homogenous Covariate Effect	$p \sim 0.11$ $q \sim 0.64$ $B_1 \sim 0$ $B_2 \sim 0$	$p \sim Beta$ $(0.65, 4.58)$ $q \sim Beta$ $(1.36, 1.0)$ $B_1 \sim 0$ $B_2 \sim 0$	$p \sim Beta$ $(0.72, 5.14)$ $q \sim Beta$ $(1.11, 1.07)$ $B_1 \sim 0.27$ $B_2 \sim 0.58$	$p \sim Beta$ $(0.69, 4.78)$ $q \sim Beta$ $(1.2, 1.1)$ $B_1 \sim Normal$ $(0.21, 0.41)$ $B_2 \sim Normal$ $(0.49, 0.19)$	$p \mid seg\ 1 \sim Beta$ $(2.38, 2.81)$ $p \mid seg\ 2 \sim Beta$ $(0.55, 1.09)$ $p \mid seg\ 3 \sim Beta$ $(1.42, 7.77)$ $q \mid seg\ 1 \sim Beta$ $(7.63, 5.5)$ $q \mid seg\ 2 \sim Beta$ $(0.59, 1.11)$ $q \mid seg\ 3 \sim Beta$ $(28.27, 28.76)$ $P\ (seg\ 1) \sim 0.14$ $P\ (seg\ 2) \sim 0.07$	$p \mid seg\ 1 \sim Beta$ $(0.73, 5.33)$ $p \mid seg\ 2 \sim Beta$ $(0.83, 5)$ $p \mid seg\ 3 \sim Beta$ $(0.12, 0.31)$ $q \mid seg\ 1 \sim Beta$ $(1.44, 1.35)$ $q \mid seg\ 2 \sim Beta$ $(0.16, 0.18)$ $q \mid seg\ 3 \sim Beta$ $(1.19, 0.95)$ $P\ (seg\ 1) \sim 0.91$ $P\ (seg\ 2) \sim 0.03$	$p \mid seg\ 1 \sim Beta$ $(0.88, 7.02)$ $p \mid seg\ 2 \sim Beta$ $(0.25, 0.41)$ $p \mid seg\ 3 \sim Beta$ $(0.1, 0.22)$ $q \mid seg\ 1 \sim Beta$ $(1.12, 1.02)$ $q \mid seg\ 2 \sim Beta$ $(12.62, 18.17)$ $q \mid seg\ 3 \sim Beta$ $(0.22, 0.41)$ $P\ (seg\ 1) \sim 0.92$ $P\ (seg\ 2) \sim 0.05$

					P (seg 3) ~ 0.79 B_1 ~ 0 B_2 ~ 0	P (seg 3) ~ 0.05 B_1 ~ 0.26 B_2 ~ 0.55	P (seg 3) ~ 0.03 B_1 ~ Normal (0.27, 0.61) B_2 ~ Normal (0.57, 0.32)
Heterogenous Scenario with Heterogeneous Covariate effect	p ~ 0.1 q ~ 0.62 B_1 ~ 0 B_2 ~ 0	p ~ Beta (1.01,7.91) q ~ Beta (2.71, 1.9) B_1 ~ 0 B_2 ~ 0	p ~ Beta (0.67,4.74) q ~ Beta (2.36, 2.17) B_1 ~ 0.17 B_2 ~ 0.55	p ~ Beta (0.69, 4.91) q ~ Beta (2.33, 2.13) B_1 ~ Normal (0.18, 0.61) B_2 ~ Normal (0.54, 0.22)	p | seg 1 ~ Beta (0.25, 2.35) p | seg 2 ~ Beta (0.7, 1.19) p | seg 3 ~Beta (1.45, 11.99) q | seg 1 ~ Beta (1.19, 0.58) q | seg 2 ~ Beta (2.98, 1.52) q | seg 3 ~ Beta (4.33,3.09) P (seg 1) ~ 0.14 P (seg 2) ~0.12 P (seg 3) ~ 0.74 B_1 ~ 0 B_2 ~ 0	p | seg 1 ~ Beta (4.72, 56.54) p | seg 2 ~ Beta (3.51, 2.61) p | seg 3 ~ Beta (0.95, 1.3) q | seg 1 ~ Beta (3.16, 3.38) q | seg 2 ~ Beta (1.9, 1.57) q | seg 3 ~ Beta (0.95, 1.3) P (seg 1) ~ 0.82 P (seg 2) ~0.05 P (seg 3) ~ 0.13 B_1 ~ 0.18 B_2 ~ 0.55	p | seg 1 ~ Beta (0.69, 5.21) p | seg 2 ~ Beta (0.13, 0.31) p | seg 3 ~ Beta (0.25, 0.72) q | seg 1 ~ Beta (2.83, 2.74) q | seg 2 ~ Beta (0.59, 0.59) q | seg 3 ~ Beta (0.6, 0.49) P (seg 1) ~ 0.91 P (seg 2) ~0.03 P (seg 3) ~ 0.96 B_1 ~ Normal (0.19, 0.71) B_2 ~ Normal (0.58, 0.46)
Finite Mixture Population	p ~ 0.16 q ~ 0.47 B_1 ~ 0 B_2 ~ 0	p ~ Beta (1.91,8.42) q ~ Beta (1.36,1.55) B_1 ~ 0 B_2 ~ 0	p ~ Beta (2.43,11.4) q ~ Beta (1.48, 1.74) B_1 ~ 0.08 B_2 ~ 0.02	p ~ Beta (0.69, 4.91) q ~ Beta (2.33, 2.13) B_1 ~ Normal (0.18, 0.61) B_2 ~ Normal (0.54, 0.22)	p | seg 1 ~ Beta (0.07, 0.35) p | seg 2 ~ Beta (9.37, 23.68) p | seg 3 ~Beta (0.61, 0.35) q | seg 1 ~ Beta (5.36, 11.09)	p | seg 1 ~ Beta (5.12, 11.77) p | seg 2 ~ Beta (0.44, 0.28) p | seg 3 ~ Beta (0.33, 3.43) q | seg 1 ~ Beta (2.54, 0.87)	p | seg 1 ~ Beta (0.26, 1.52) p | seg 2 ~ Beta (3.36, 7.17) p | seg 3 ~ Beta (0.79, 0.45) q | seg 1 ~ Beta (3.27,6.38)

					q | seg 2 ~ Beta (4.84 1.82) *q | seg 3 ~ Beta (0.22,0.34)* *P (seg 1) ~ 0.36* *P (seg 2) ~0.57* *P (seg 3) ~ 0.07* B_1 *~ 0* B_2 *~ 0*	*q | seg 2 ~ Beta (0.12,0.37)* *q | seg 3 ~ Beta (5.18, 10.29)* *P (seg 1) ~ 0.5* *P (seg 2) ~0.06* *P (seg 3) ~ 0.44* B_1 *~ 0.09* B_2 *~ 0.02*	*q | seg 2 ~ Beta (1.25,0.54)* *q | seg 3 ~ Beta (1.07,1.17)* *P (seg 1) ~ 0.45* *P (seg 2) ~0.49* *P (seg 3) ~ 0.06* B_1 *~ Normal (0.12, 0.47)* B_2 *~ Normal (-0.04, 0.91)*
Finite Mixture Population with Heterogeneous Covariate effect	*p ~ 0.18* *q ~ 0.61* B_1 *~ 0* B_2 *~ 0*	*p ~ Beta (1.68,6.68)* *q ~ Beta (1.32,0.99)* B_1 *~ 0* B_2 *~ 0*	*p ~ Beta (2.45,10.8)* *q ~ Beta (0.84, 0.93)* B_1 *~ 0.34* B_2 *~ 0.67*	*p ~ Beta (0.69, 4.91)* *q ~ Beta (2.33, 2.13)* B_1 *~ Normal (0.18, 0.61)* B_2 *~ Normal (0.54, 0.22)*	*p | seg 1 ~ Beta (0.47, 0.44)* *p | seg 2 ~ Beta (0.17, 0.5)* *p | seg 3 ~Beta (3, 9.66)* *q | seg 1 ~ Beta (1.47, 0.89)* *q | seg 2 ~ Beta (1.39, 1.93)* *q | seg 3 ~ Beta (2.45,1.07)* *P (seg 1) ~ 0.11* *P (seg 2) ~0.25* *P (seg 3) ~ 0.64* B_1 *~ 0* B_2 *~ 0*	*p | seg 1 ~ Beta (0.18,0.75)* *p | seg 2 ~ Beta (1.49, 2.01)* *p | seg 3 ~ Beta (2.09, 7.03)* *q | seg 1 ~ Beta (1.05, 2.05)* *q | seg 2 ~ Beta (0.81, 0.4)* *q | seg 3 ~ Beta (1.21, 0.96)* *P (seg 1) ~ 0.27* *P (seg 2) ~0.14* *P (seg 3) ~ 0.6* B_1 *~ 0.34* B_2 *~ 0.64*	*p | seg 1 ~ Beta (0.89, 3.83)* *p | seg 2 ~ Beta (1.27, 0.29)* *p | seg 3 ~ Beta (6.7, 17.49)* *q | seg 1 ~ Beta (1.39, 1.64)* *q | seg 2 ~ Beta (0.83, 1.16)* *q | seg 3 ~ Beta (0.59, 0.28)* *P (seg 1) ~ 0.72* *P (seg 2) ~0.05* *P (seg 3) ~ 0.23* B_1 *~ Normal (0.25, 0.4)* B_2 *~ Normal (0.66, 0.91)*

Figure 3: Actual Parameter Values Generated by the Model when Fitted onto the Datasets

Heterogenous Marketing Effect

REFERENCES

Blackwell, D. and J. MacQueen (1973). Ferguson distributions via Polya urn schemes,The Annals of Statistics 1: 353-355.

Chou, P., Chuang, H., Chou, Y., Liang, T. (2020). Predictive Analytics for Customer Retention: Low-Dimensional Probability Modeling or igh-Dimensional Machine Learning?

Dew, Ryan and A. Ansari. “Bayesian Nonparametric Customer Base Analysis with Model-Based Visualizations.” Mark. Sci. 37 (2018): 216-235.

Fader, P. S., Hardie, B. G. S., and Lee, K. L. (2005). “counting your customers” the easy way: An alternative to the pareto/nbd model. Marketing Science, 24(2):275–284.

Fader, P., Hardie, B., and Shang, J. (2009). Customer-base analysis in a discrete-time noncontractual setting. SSRN Electronic Journal.

Gelman A, Carlin JB, Stern HS, Rubin DB (2014). Bayesian Data Analysis, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gupta, Sunil (1991), “Stochastic Models of Interpurchase Time with Time Dependent Covariates,” Journal of Marketing Research, 28 (1), 1-15.

Jerath, K., Fader, P., and Hardie, B. (2011). New perspectives on customer death using a generalization of the pareto/nbd model. SSRN Electronic Journal.

Kim, J. G., Menzefricke, U., and Feinberg, F. M. (2004). Assessing hetero- geneity in discrete choice models using a dirichlet process prior. Review ofMarketing Science, 2(1).

Ma, S.-H. and Liu, J.-L. (2007). The mcmc approach for solving the pareto/nbd model and possible extensions. Third International Conference on Natural Computation (ICNC 2007).

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models, Journal of Computational and Graphical Statistics 9: 249-265.

Quintana, F. A. and Marshall, P. (2014). A bayesian non-parametric pareto/nbd model: Individual and cluster analysis.

Schweidel, D. and Knox, G. (2013). Incorporating Direct Marketing Activity into Latent Attrition Models.

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

Singh, Siddarth, Sharad Borle and Dipak Jain (2009), "A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes Are Not Observed," Quantitative

Marketing and Economics, 7 (2), 181-205.

Xie, S., Huang, C. (2020). Systematic comparisons of customer base prediction accuracy: Pareto/NBD versus neural network