May 2008

# Modeling Discrete-Time Transactions Using the BG/BB Model

Harvey Yang Zhang
*University of Pennsylvania*

# Modeling Discrete-Time Transactions Using the BG/BB Model

**Abstract**
For a business, it's important to know the repeat purchasing behavior of an existing customer. However, in a non-contractual setting, businesses cannot easily observe whether a customer is still active. The Beta-Geometric/Beta Binomial (BG/BB) model developed by Fader and Hardie in 2004 serves as a natural framework to model these types of behaviors. We further evaluate the BG/BB model by applying two different versions of the model to a dataset of 1975 customers over a period of 16 years to evaluate the predictive abilities of the BG/BB, and compares it with the Binary Logistic Regression model to determine whether the BG/BB model can give accurate forecasts over a long period of time.

**Keywords**
BG/BB, beta-geometric, beta-binomial, customer-base analysis, customer

**Modeling Discrete-Time Transactions Using the BG/BB Model**

Harvey Yang Zhang

Faculty Advisor: Professor Peter Fader

May 2008

For a business, it's important to know the repeat purchasing behavior of an existing customer. However, in a non-contractual setting, businesses cannot easily observe whether a customer is still active. The Beta-Geometric/Beta-Binomial (BG/BB) model developed by Fader and Hardie in 2004 serves as a natural framework to model these types of behaviors. We further evaluate the BG/BB model by applying two different versions of the model to a dataset of 1975 customers over a period of 16 years to evaluate the predictive abilities of the BG/BB, and compares it with the Binary Logistic Regression model to determine whether the BG/BB model can give accurate forecasts over a long period of time.

Key words: BG/BB, beta-geometric, beta-binomial, customer-base analysis, customer lifetime value, CLV, BG/NBD, Pareto/NBD

## 1. Introduction

The information on the number of repeat customers is important for event planners. The event planners make revenue based on the number of customers that attend the event, and customers who make more repeat attendances are more profitable. Therefore, predicting the number of repeat attendances of individual customers is important for the event planner to know who to market the event to, as well as budgeting its revenue and costs. The two most important managerial questions are predicting which customers are still active, and how many future transactions the customers will make.

These two questions are especially difficult to answer when the relationship between the manager and the customer is non-contractual; that is, the customer does not need to notify the manager when the customer becomes inactive. Many business relationships that involve customers making a purchase are non-contractual. For example, a customer does not need to notify a supermarket when the customer switches markets. A customer dying would be the ultimate example of becoming inactive, and surely the customer does not need to notify anyone in this case.

Many models have been developed in the past to model this non-contractual purchasing behavior. The Pareto/NBD model developed by Schmittlein, Morrison, and Colombo (Schmittlein et al, 1987) is used for continuous-time transactions, in which the customer can choose to make a transaction at any point in time. The Beta-Geometric/Beta-Binomial (BG/BB) model developed by Fader and Hardie, with a similar logic to the Pareto/NBD model, has been shown to be very good at modeling discrete-time transactions (Fader et al, 2004), while allowing for heterogeneity in both probability of purchase and probability of becoming inactive across the population. Thus, the BG/BB serve as the natural base models for this type of discrete-time transactions.

This paper applies the BG/BB model to a dataset that include attendance histories at a conference. We will first describe the BG/BB model, including its assumptions and underlying probability distributions. Then we will apply the model to the full dataset using two different approaches to provide an empirical analysis for the applicability of the model to event attendance. We will then test the predictive abilities of the model by forecasting the number of repeat attendances over four years using eleven years worth of data, and compare the forecasted results with the actual data. Finally, we will address the important managerial questions and discuss the limitations of both models.

**2. Model Development**

To fit the BG/BB model, the customer-level information has to be classified based on "recency" and "frequency" of each individual customer's purchases. Recency is defined as when the last transaction occurred, and frequency is defined as the total number of repeat transactions the customer made in the past. The notation used to represent this information is (x, m, n), where x is the total number of repeat transactions the customer made in the past, m is the index of the transaction opportunity of the most recent transaction, and n is the total number of repeat transaction opportunities. For example, a customer that had 6 transaction opportunities, made 3 purchases, and the last purchase was on the $5^{th}$ transaction opportunity would be classified as (3, 5, 6). Given these information about customer purchase behavior, we can fit the BG/BB model to describe their probability of still being active, as well as their expected number of purchases in the future conditioning on their (x, m, n) information.

The BG/BB model is based on the following underlying assumptions:

1. A customer is active when he joins the population by making the initial purchase. An active customer becomes inactive at the beginning of the next transaction opportunity with probability q, and the value of q stays constant throughout the customer's lifetime.
2. The values of q differ across the population, and follow a Beta distribution with parameters $\gamma$ and $\delta$.
3. If a customer is active at a transaction opportunity, the customer has probability p of making a purchase, and the value of p stays constant throughout the customer's lifetime.
4. The values of p differ across the population, and follow a Beta distribution with parameters $\alpha$ and $\beta$.
5. The values of p and q are independent across the population.

Assumptions 1 and 2 are the assumptions of a Beta-Geometric model, and assumptions 3 and 4 are the assumptions of a Beta-Binomial model. Therefore the model is called BG/BB. (For complete derivation of the BG/BB model as well as the proofs of the formulas below, see Appendix).

The parameters of the BG/BB model can be estimated using the maximum likelihood procedure. The likelihood function for a randomly-chosen customer with purchase history (x, m, m) is shown to be:

$$L(\alpha,\beta,\gamma,\delta \mid x,n,m) = \frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)}\frac{B(\gamma,\delta+n)}{B(\gamma,\delta)} + \sum_{i=0}^{n-m-1}\frac{B(\alpha+x,\beta+m-x+i)}{B(\alpha,\beta)}\frac{B(\gamma+1,\delta+m+i)}{B(\gamma,\delta)}$$

where B(α,β) means the Beta function with parameters α and β.

For a sample of N customers, we can estimate the parameters by maximizing the sum of the log-likelihood function of each customer using standard numerical procedures in Excel or other programs.

The probability of a customer being alive given (x, m, n) is shown to be:

$$P(active \mid x,n,m,\alpha,\beta,\gamma,\delta) = \frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \frac{B(\gamma,\delta+n+1)}{B(\gamma,\delta)} \Big/ L(\alpha,\beta,\gamma,\delta \mid x,n,m)$$

The expected number of transaction a customer makes in the next $n^*$ periods given (x, m, n) is shown to be:

$$E(X^* \mid n^*,x,n,m,\alpha,\beta,\gamma,\delta) = \frac{B(\alpha+x+1,\beta+n-x)}{B(\alpha,\beta)} \times \left[\frac{B(\gamma-1,\delta+n+1)-B(\gamma-1,\delta+n+n^*+1)}{B(\gamma,\delta)}\right] \Big/ L(\alpha,\beta,\gamma,\delta \mid x,n,m)$$

The BG/BB with drift model is a more generalized BG/BB model in which the four parameters change over time according to a logistic equation. The mean and polarization index of a Beta distribution are defined as follows:

$$mean = \mu = \alpha/(\alpha+\beta),$$

$$polarization = \varphi = 1/(\alpha+\beta+1)$$

A logistic equation is used to model the drift in μ and φ separately for the BG and BB distributions. The equation is specified as follows:

$$\mu(t) = \frac{e^{a_1+b_1t+c_1\ln(t)}}{1+e^{a_1+b_1t+c_1\ln(t)}} \quad \varphi(t) = \frac{e^{a_2+b_2t+c_2\ln(t)}}{1+e^{a_2+b_2t+c_2\ln(t)}}$$

where t is the number of year since inception, and $a_1$, $b_1$, $c_1$, $a_2$, $b_2$, $c_2$ are constants.

This equation form is chosen because it is very versatile in modeling different types of drifts, since it has linear, log, and linear-log components. Using maximum likelihood procedures, the values of the constants for the equations can be estimated, and then the parameters α, β, γ and δ can be found as follows:

$$\alpha = \mu_{BB} \times (1 - \varphi_{BB}) / \varphi_{BB}$$
$$\beta = (1 - \mu_{BB})(1 - \varphi_{BB}) / \varphi_{BB}$$
$$\gamma = \mu_{BG} \times (1 - \varphi_{BG}) / \varphi_{BG}$$
$$\delta = (1 - \mu_{BG})(1 - \varphi_{BG}) / \varphi_{BG}$$

where $\mu_{BB}$ and $\varphi_{BB}$ are given by the drift process for the BB distribution, $\mu_{BG}$ and $\varphi_{BG}$ are given by the drift process for the BG distribution. (for proof, see Appendix)

The BG/BB with drift model can capture the drift of the parameters over time, which lends itself to modeling time series that span over many years and situations where the parameters are not likely to stay constant over time. (In some cases, when the drift over time in either the mean or polarization of either the BG or BB distribution is very small, that drift can be assumed to be zero over time.)

**3. Empirical Model Fitting**

To evaluate the applicability of the BG/BB model to real-life situations, we fit the model over a dataset taken from the ART marketing conference that is held every year. The dataset include the attendance history of 1975 unique customers from 1990 to 2005. These customers have made their first attendance between 1990 and 2005, so there are 16 cohorts of customers based on their years of initial attendance, and they had repeat transaction opportunities that range from 0 to 15.

Approach 1:

Two approaches were used to fit the data. First, we use four different parameters to fit a BG/BB distribution for each cohort of customers based on their years of initial purchase. This serves as a basic test to see if BG/BB can serve as a valid model to fit the data.

The model's validity is verified using a hold-out analysis, where the data from the years 2001 to 2005 are withheld from parameter estimation. To prevent over-fitting the data, we only fit the model over customers who had at least four repeat transaction opportunities in the past. Thus, we only used transaction histories of the customers that made their initial attendance between 1990 and 1996. This means that we are fitting 11 years worth of data over 7 cohorts of customers, and the transaction opportunities range from 4 to 11. For each cohort of customers based on their year of initial purchases, we use four different parameters to fit a BG/BB distribution using the maximum likelihood estimation. The results are shown below in Figure 1 below:

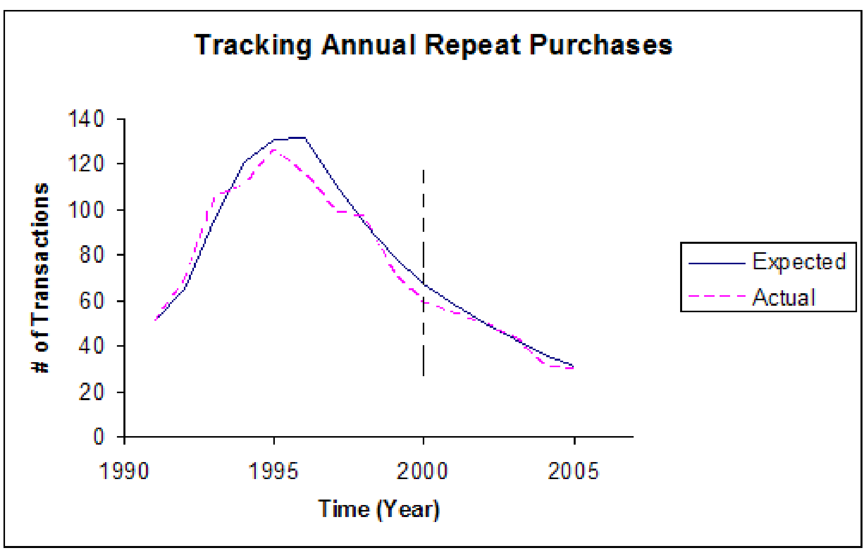


Figure 1: Actual vs expected annual repeat purchases over time, fitted over 11 years of data and extrapolated to 16 years.

The results of the hold-out analysis under approach 1 show that the BG/BB model fits the data very well. This justifies the using BG/BB to model discrete time purchasing behaviors. However, the approach of fitting four parameters to each cohort limits the predictive power of the model. First, we cannot predict the behaviors of cohorts with fewer than four years of data because of problems with data over-fitting. Second, we cannot predict the behaviors of any future cohorts because the model does not tell us any correlation between the parameters of cohorts. Thus, we further evaluated the BG/BB with drift model in the second approach.

Approach 2:

In the second approach, we assume that the four parameters of BG/BB follow certain type of drift over the years, and the drift can be modeled by constants. Then we use this BG/BB with drift model to fit BG/BB distributions over all eleven cohorts of customers. This way, we can test if there is any change of customer behavior across the years. This approach has more predictive value because this model can be used to predict the behaviors of cohorts that do not exist yet. From empirical observations, we noticed that the drift in the polarization in the BG distribution is negligible, thus that parameter was held constant over time. Thus, we used 10 parameters in this BG/BB with drift model.

This model's validity is again be verified using a hold-out analysis, where the data from the years 2001 to 2005 are withheld from parameter estimation. We used transaction histories of the customers that made their initial attendance between 1990 and 1999. This means that we are fitting 11 years worth of data over 10 cohorts of customers, and the transaction opportunities range from 1 to 11. The results are shown below in Figure X below:

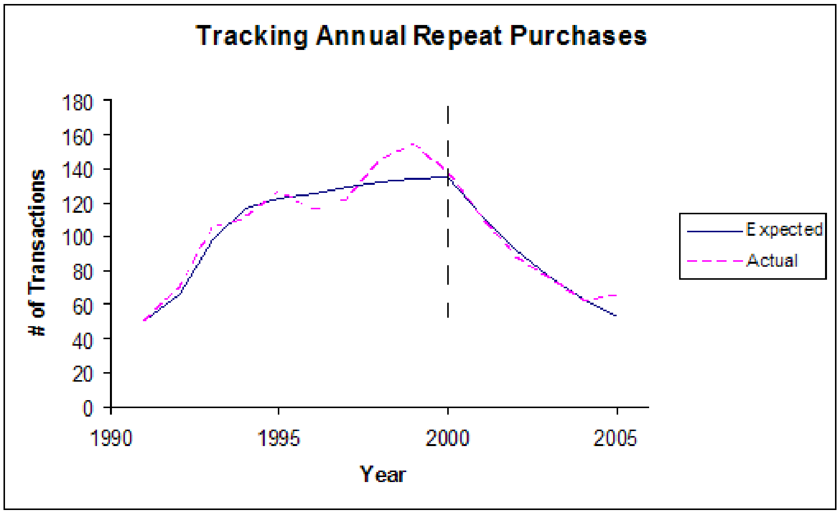


Figure 2: Actual vs expected annual repeat purchases over time, fitted over 11 years of data and extrapolated to 16 years.

Comparing Figure 1 with Figure 2 above, we can observe that holistically, the fit using the first approach is not visibly better than the second approach. Quantitatively, the Mean Absolute Percent Error (MAPE) of the year-by-year purchases in the holdout period is 6.24% using the first approach, and the MAPE is 6.42% using the second approach. This small difference in predictive ability further validates that the first approach is not significantly better than the second approach, even though the first approach uses 24 parameters, and the second approach uses only 10 parameters.

The most important use of the BG/BB model is to predict the number of future purchase based on a given past purchase history, or conditional expectations based on (x, m, n). Thus, we examined the accuracy of the conditional expectations using the second approach by calculating the expected number of future purchases in the holdout period for each customer. The aggregate results are shown in Figures 3 and 4 below that graphs conditional expectations based on x and m.

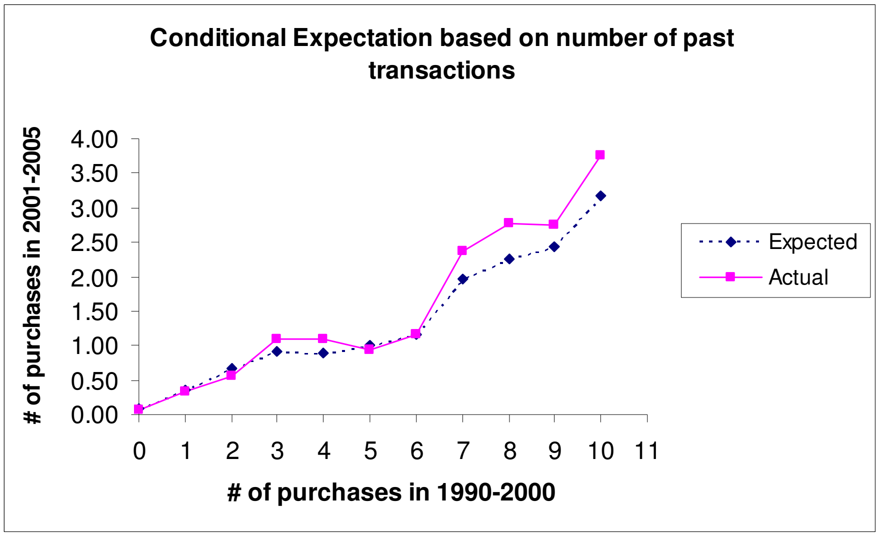


Figure 3: Expected vs actual number of purchases in the holdout period based for an individual based on x.

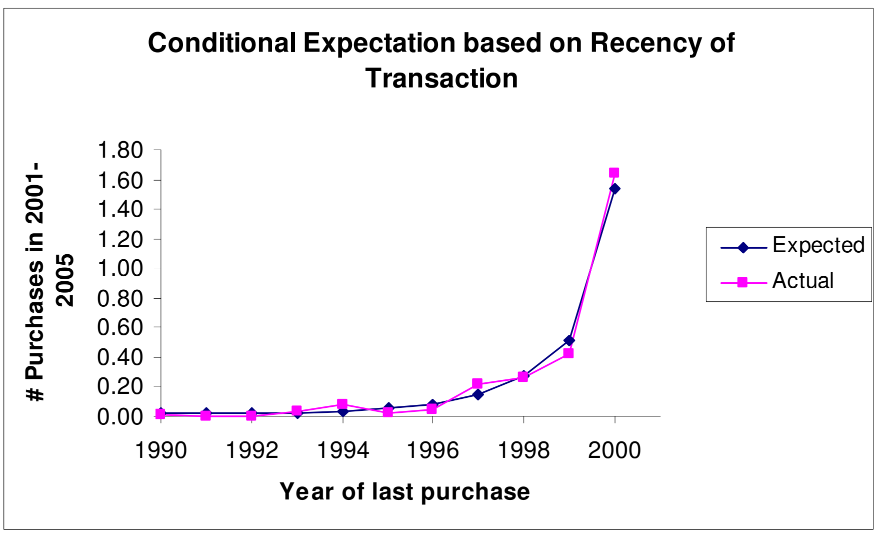


Figure 4: Expected vs actual number of purchases in the holdout period for an individual based on m.

To explain the above graphs with examples, Figure 3 above shows that if an individual made six transaction from 1990 to 2000 (x = 6), then that individual is expected to make 1.17 purchases over 2001-2005, and the actual average number of purchases for an individual with a x of 6 is 1.15. Figure 4 shows that if an individual's most recent transaction is in 1998 (m = 9), then that individual is expected to make 0.27 purchases over 2001-2005, and the actual average number of purchases for an individual with an m of 9 is 0.25.

From both graphs above, it can be observed that the BG/BB with drift model captures the trends of future number of purchase based on a customer's past purchase history under both x and m very well. The correlation between the expected and actual number of transactions is over 0.99 for both type of conditional expectations.

Another important use of the BG/BB with drift model is to predict the behavior of future cohorts. To test the predictive ability of the model further, we computed the expected number of annual repeat transactions of the cohorts in the hold-out period given that we know the size of each cohort, and compared that with the actual data. The result is shown in Figure X below:

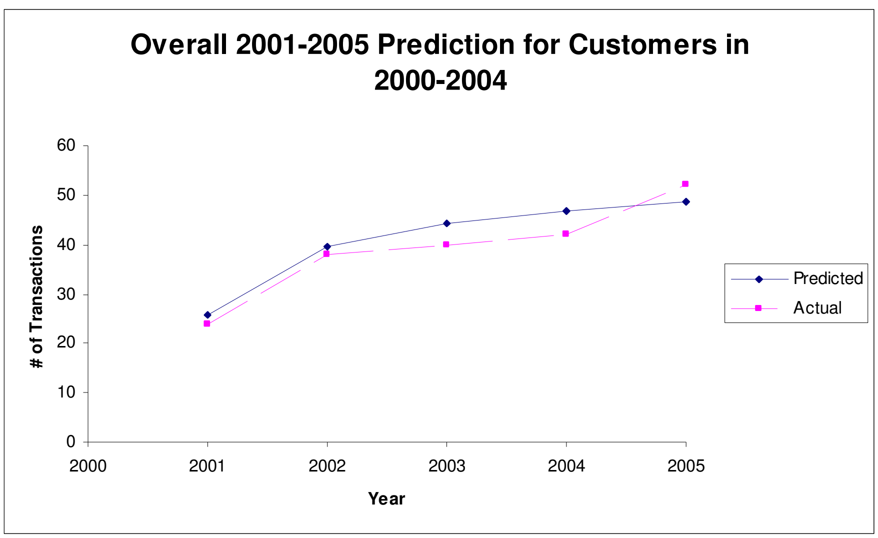


Figure 5: Actual vs expected annual repeat purchases over time, fitted using data up to 2000 and extrapolated to 2005.

From Figure 5 above, it can be observed that the BG/BB with drift model predicts the behavior of future cohorts very well. The MAPE of annual repeat purchases during the hold-out period is only 4.80%, which is remarkable given that no observations are made yet on the purchasing behaviors of those cohorts. Based on the evidences in the predictive abilities of the model in the hold-out analysis, we conclude that the BG/BB with drift model provides a reasonable fit to the data.

## 4. Comparison with Econometric Models

From the hold-out analysis above, we conclude that the BG/BB with drift model provides a good fit to the data. We will also compare the BG/BB with drift model with the regular binary Logistic Regression model, which is commonly used in econometrics to model binary (Yes/No) data that the BG/BB model addresses. (See Appendix for the mathematical formulation of the Binary Logistic Regression)

We compared the Logistic Regression and the BG/BB with drift model in a hold-out analysis, where the data from the years 2001 to 2005 are withheld from parameter estimation. Logistic Regression uses two parameters to fit each cohort. To prevent over-fitting the data, we only fit the model over customers who had at least three repeat transaction opportunities in the past. Thus, we only used transaction histories of the customers that made their initial attendance between 1990 and 1997. The same calibration period was used for the BG/BB with drift model. The results are shown in Figure 6 below:

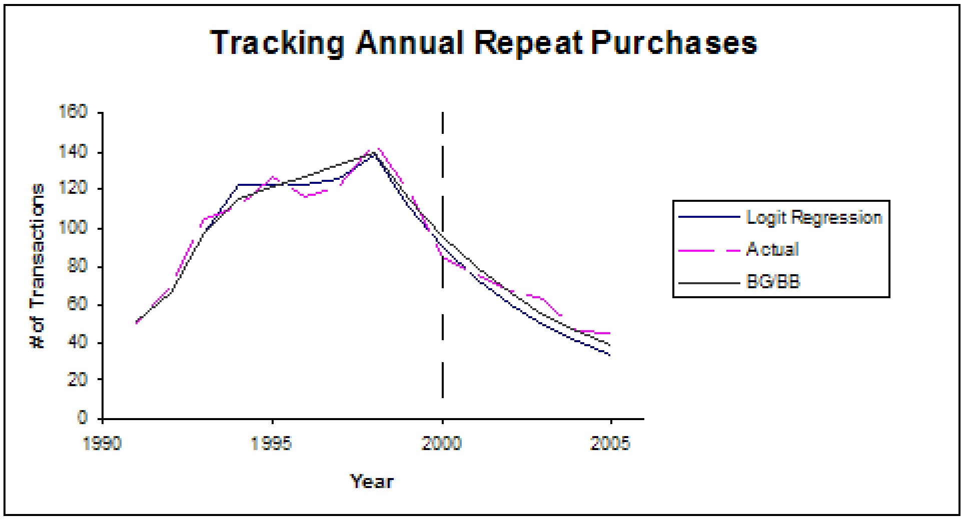


Figure 6: Actual vs expected annual repeat purchases over time, fitted using data up to 2000 and extrapolated to 2005.

From Figure 6 above, it can be observed that holistically, the fit using Logistic Regression is not visibly better than the BG/BB. Quantitatively, the MAPE of the year-by-year purchases in the holdout period is 17.88% using the Logistic Regression, and the MAPE is 5.43% using the BG/BB approach, even though the Logistic Regression uses six more parameters to fit the data. Also, the Logistic Regression model cannot be used to calculate conditional expectations based on an individual's past purchase behavior, as it can only give the cohort's average probability of making a purchase in each year. Thus, we conclude that the BG/BB with drift model is superior to the Logit Regression model in its predictive abilities.

## 5. Managerial Questions

We have concluded that the BG/BB model under the first approach is reasonable to describe the purchasing behavior of customers. We can now answer the two important managerial questions about the customer base: 1. given a customer, what is the probability that the customer is still active, and 2. given a customer, what is the expected number of purchases over the customer's lifetime in the future.

In Table 1 below, we use the cohort that made their initial purchase in 2000 as an example to illustrate the behavior of probability of being active in 2006, as the trends in probability being active are the same for all cohorts.

| | **Year of last purchase** | | | | | |
|---|---|---|---|---|---|---|
| **Number of purchases** | 2005 | 2004 | 2003 | 2002 | 2001 | 2000 |
| 5 | 0.681 | | | | | |
| 4 | 0.681 | 0.200 | | | | |
| 3 | 0.681 | 0.304 | 0.069 | | | |
| 2 | 0.681 | 0.371 | 0.152 | 0.040 | | |
| 1 | 0.681 | 0.418 | 0.229 | 0.108 | 0.039 | |
| 0 | | | | | | 0.063 |

Table 1: Probability that a customer is active in 2006 based on recency and frequency.

From Table 1 above, we can observe two general trends. The first trend is that the probability of being active increases as the number of past purchase decreases. This can be attributed to the fact that customers who make fewer purchases tend to have a lower probability of making a purchase if they are active, so the fact that no purchase was made in 2005 can be attributed more to their low probability of making a purchase than to the probability of them being inactive. The second trend is that given the same number of past purchases, the probability of being active decreases as the year of most recent transaction decreases. This is because the longer the customer does not make a purchase, the more likely that the customer is inactive, since if a customer makes a purchase, the customer must be active at least in that year.

In Table 2 below, we use the cohort that made their initial purchase in 2000 as an example again to illustrate the behavior of Discounted Expected Transactions (DET) in a customer's remaining lifetime, as the trends in probability being active are the same for all cohorts.

| | Year of last purchase | | | | | |
|---|---|---|---|---|---|---|
| **Number of purchases** | 2005 | 2004 | 2003 | 2002 | 2001 | 2000 |
| 5 | 1.362 | | | | | |
| 4 | 1.110 | 0.326 | | | | |
| 3 | 0.857 | 0.383 | 0.087 | | | |
| 2 | 0.604 | 0.329 | 0.135 | 0.035 | | |
| 1 | 0.352 | 0.216 | 0.118 | 0.056 | 0.020 | |
| 0 | | | | | | 0.009 |

Table 2: Expected number of remaining transactions based on recency and frequency, discounted with an annual rate of 10%.

From Table 2 above, we can observe three general trends. The first trend is that if a customer made a transaction in the most recent year, then DET increases monotonically as number of past purchases increases. This is because if the probability of being active is the same, then more purchases in the past imply the customer has a higher probability of making a purchase in a year while active, which will lead to more transactions in the future. The second trend is that the DET does not increase monotonically as number of past purchase increases if a customer did not make a transaction in the most recent year. This can be attributed to the fact that customers who made more transactions in the past and did not make a transaction in the most recent year implies that their probability of being inactive is higher, as shown in Table 1. So the probability of being inactive may cause the customer to have fewer expected transactions even though the customer has a high probability of making a purchase while active. The third trend is that DET decreases monotonically as frequency decreases because the longer the customer does not make a purchase, the more likely that the customer is either inactive or has a low probability of making a purchase while active.

We can generate similar tables for each of the cohorts. Using these tables, we can compute the probability that any individual customer is inactive and the DET based on the customer's past transaction history.

**6. Discussion:**

In this study, we have developed a discrete-time model that explains the customer's behavior of making transactions. Using data from the attendance at a conference, we have shown that the BG/BB model can be applied to model not only purchases, but also event attendances. From the holdout analysis, the BG/BB has demonstrated very good predictive abilities; thus, we can use the BG/BB to predict whether a customer is active, as well as the customer's DET, which are the two important managerial questions facing any event planner.

We have considered two BG/BB models in this analysis. The first one fits four parameters over each cohort, and the second one fits 10 parameters over all cohorts. We showed that although both approaches produce very good fits to the data, the second approach is better than the first one in terms of having fewer parameters as number of cohorts grows large. Also, the second approach has much more predictive power since it is possible to predict the behaviors of future cohorts using past data even without any observation of the behaviors of those cohorts. Thus, by adding a drift over time to the parameters of BG/BB, we have shown that it is possible to pool different cohorts into one single model. This is an improvement over previously developed models for single cohorts, as noted in the discussions of previous models in previous papers. (Fader et al, 2005)  For similar reasons, we have also shown that the BG/BB model is superior to the Binary Logistic Regression model that is commonly used in economics.

The BG/BB model can be easily generalized to other datasets. For discrete-time transactions where a customer is not limited to purchase only one item, a mixing distribution (Normal or Gamma) can be added to the BB model to account for the multiple purchases at one time. (Schmittlein and Peterson 1994, Colombo and Jiang 1999). The BG/BB model can also serve as a benchmark for managers to measure the effect of marketing activities.

However, in predicting more long term purchase behaviors, several issues must be considered. First, the parameters of the BG/BB model may change over time at the individual level. Second, there may be covariates, such as marketing activity, that affect the behavior of transactions. Third, there may be specific segments of customers that come in cohorts, so the behaviors of customers are not independent. Nevertheless, even when the above issues are taken into consideration, the BG/BB should still be considered as an appropriate base model that could be extended to include the effect of the issues described above.

**References:**


Fader, Peter S., Bruce G.S. Hardie, and Paul D. Berger (2004), "Customer-Base Analysis with Discrete-Time Transaction Data." (Working paper)

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), “"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model,” Marketing Science, (Spring), 275-284.

Schmittlein, David C., Robert A. Peterson. 1994. Customer base analysis: An industrial purchase process application. Marketing Sci. 13 (Winter) 41–67.

Morrison, Donald G., Richard D. H. Chen, Sandra L. Karpis, and Kathryn E. A. Britney (1982), “Modelling Retail Customer Behavior at Merrill Lynch,” Marketing Science, 1 (Spring), 123–141.

Colombo, Richard, Weina Jiang (1999), A stochastic RFM model. J. Interactive Marketing 13 (Summer) 2–12.

**7. Appendix:**

Derivations of the BG/BB model equations are given by Fader and Hardie in their 2004 working paper, reproduced below with permission.

Let p be the probability that an active customer makes a transaction in any given opportunity. Heterogeneity in p follows a beta distribution with pdf:

$$f(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < p < 1$$

Let q be the probability that an active customer becomes inactive at the beginning of the next transaction, Heterogeneity in p follows a beta distribution with pdf:

$$f(q \mid \gamma, \delta) = \frac{q^{\gamma-1}(1-q)^{\delta-1}}{B(\gamma, \delta)}, \quad 0 < q < 1$$

Then in general, for a customer with purchase history (x, m, n), the likelihood function for p and q given (x, m, n) is:

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

Then for a randomly-chosen customer with purchase history (x, n, m), the likelihood function for the parameters α, β, γ, δ can be obtained by taking the expectation of the equation above over the two beta mixing distributions for p and q:

$$L(\alpha, \beta, \gamma, \delta \mid x, n, m) = \int_0^1 \int_0^1 L(p, q \mid x, n, m) f(p \mid \alpha, \beta) f(q \mid \gamma, \delta) dp dq$$

which, after simplification, becomes:

$$L(\alpha, \beta, \gamma, \delta \mid x, n, m) = \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)} + \sum_{i=0}^{n-m-1} \frac{B(\alpha+x, \beta+m-x+i)}{B(\alpha, \beta)} \frac{B(\gamma+1, \delta+m+i)}{B(\gamma, \delta)}$$

**Derivation of P(active | x, n, m):**

For a customer with purchase history (x, m, n), the probability of the customer being active in period n+1 is:

$$P(active \mid x,n,m,p,q) = \frac{p^x(1-p)^{n-x}(1-q)^{n+1}}{L(p,q \mid x,n,m)}$$

Because p and q are unobserved, P(active | x, n, m) is the expectation of the equation above over the mixing distributions of p and q with parameters α, β, γ, δ:

$$P(active \mid x,n,m,\alpha,\beta,\gamma,\delta) = \int_0^1\int_0^1 P(active \mid x,n,m,p,q) f(p,q \mid x,n,m,\alpha,\beta,\gamma,\delta)\,dpdq$$

By Bayes' theorem, the joint posterior distribution of p and q is given by:

$$f(p,q \mid x,n,m,\alpha,\beta,\gamma,\delta) = \frac{L(p,q \mid x,n,m) f(p \mid \alpha,\beta) f(q \mid \gamma,\delta)}{L(\alpha,\beta,\gamma,\delta \mid x,n,m)}$$

After simplification, we obtain:

$$P(active \mid x,n,m,\alpha,\beta,\gamma,\delta) = \frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \frac{B(\gamma,\delta+n+1)}{B(\gamma,\delta)} \Big/ L(\alpha,\beta,\gamma,\delta \mid x,n,m)$$

**Derivation of** $E(X^* \mid n^*, x, n, m, \alpha, \beta, \gamma, \delta)$**:**

Let $E(X^* \mid n^*, x, n, m, \alpha, \beta, \gamma, \delta)$ denote the expected number of purchases over the next $n^*$ periods by a customer with purchase history (x, m, n). The probability that the customer remains active in any period follows a geometric distribution with parameter q, and the probability that the customer makes a purchase while active is a constant p. Thus, together, the expected number of transactions in an interval is a geometric series.

In general, given p and q, the expected number of transactions in the interval (n, $n^*$) is:

$$E(X^* \mid n^*, p, q) = \frac{p}{q} - \frac{p}{q}(1-q)^{n^*}$$

Multiplying the above equation by the probability that a customer with purchase history (x, m, n) is still active, and taking the expectation over the joint posterior distribution of p and q, we obtain:

$$E(X^* \mid n^*, x, n, m, \alpha, \beta, \gamma, \delta) = \frac{B(\alpha + x + 1, \beta + n - x)}{B(\alpha, \beta)} \times \left[ \frac{B(\gamma - 1, \delta + n + 1) - B(\gamma - 1, \delta + n + n^* + 1)}{B(\gamma, \delta)} \right] \Big/ L(\alpha, \beta, \gamma, \delta \mid x, n, m)$$

**Binary Logistic Regression formulation:**

Binary Logistic regression assumes that given a cohort, whether a customer in the cohort makes a transaction in a given period is binomially distributed. For the $k^{th}$ year, the probability that a customer makes a purchase is given by $p_k$. Thus, the total observed number of purchases in year k from a cohort with size n follows a binomial distribution with parameters n and $p_k$.

Binary logistic regression also assumes that $p_k$ is given by the following equation:

$p_k = \frac{1}{1 + e^{-(a + bX_k)}}$, where $X_k$ is the observed number of transactions in year k, and a,b are constant parameters.

Then given data on $X_k$, the values of the parameters a, b can be estimated by maximum likelihood procedures.

**Derivation of estimating α, β, γ and δ from $\mu_{BB}$, $\varphi_{BB}$ $\mu_{BG}$, and $\varphi_{BG}$:**

$$\begin{aligned}\mu_{BB}\times(1-\varphi_{BB})/\varphi_{BB} &= \frac{\alpha}{\alpha+\beta}\frac{1-1/(\alpha+\beta+1)}{1/(\alpha+\beta+1)} \\ &= \frac{\alpha}{\alpha+\beta}\frac{(\alpha+\beta+1)(\alpha+\beta)}{\alpha+\beta+1} \\ &= \alpha\end{aligned}$$

$$\begin{aligned}(1-\mu_{BB})\times(1-\varphi_{BB})/\varphi_{BB} &= (1-\frac{\alpha}{\alpha+\beta})(\frac{1-1/(\alpha+\beta+1)}{1/(\alpha+\beta+1)}) \\ &= \frac{\beta}{\alpha+\beta}\frac{(\alpha+\beta+1)(\alpha+\beta)}{\alpha+\beta+1} \\ &= \beta\end{aligned}$$

Derivations of γ and δ are similar to the derivations of α and β above.

| | | | | | |
|---|---|---|---|---|---|
| alpha | 0.26 | | B(alpha,beta) | 3.8651 | |
| beta | 1.00 | | | | |
| gamma | 284.89 | | B(gamma,delta) | 4.6E-283 | |
| delta | 881.19 | | | | |
| n* | 1 | | | | |
| | | | | | |
| LL | -201.66 | | Maximize LL | | |
| | | | | | |
| | | | | | |
| | | | | | |
| x | m | n | # cust. | | L(.\|X=x,t_x,n) |
| 5 | 5 | 5 | 2 | -8.83 | 0.01210 |
| 4 | 5 | 5 | 2 | -11.72 | 0.00285 |
| 3 | 5 | 5 | 2 | -12.70 | 0.00175 |
| 2 | 5 | 5 | 0 | 0.00 | 0.00233 |
| 1 | 5 | 5 | 1 | -4.91 | 0.00739 |
| 4 | 4 | 5 | 1 | -4.87 | 0.00766 |
| 3 | 4 | 5 | 1 | -5.74 | 0.00323 |
| 2 | 4 | 5 | 2 | -11.23 | 0.00364 |
| 1 | 4 | 5 | 2 | -9.11 | 0.01052 |
| 3 | 3 | 5 | 3 | -13.38 | 0.01155 |
| 2 | 3 | 5 | 0 | 0.00 | 0.00733 |
| 1 | 3 | 5 | 6 | -24.66 | 0.01640 |
| 2 | 2 | 5 | 4 | -15.05 | 0.02324 |
| 1 | 2 | 5 | 5 | -17.69 | 0.02906 |
| 1 | 1 | 5 | 10 | -27.05 | 0.06689 |
| 0 | 0 | 5 | 120 | -34.72 | 0.74874 |

Figure 7: A sample fitting of the BG/BB in excel over the cohort of 2000.

| | | | |
|---|---|---|---|
| alpha | 0.267517 | B(alpha,beta) | 3.7004 |
| beta | 1.027819 | | |
| gamma | 151.0433 | B(gamma,delta) | 6.1E-184 |
| delta | 823.9281 | | |
| n* | 4 | | |
| | | | |
| LL | -199.06 | Maximize LL | |

| x | m | n | # cust. | | L(.\|X=x,t_x,r | Number of Repeats | ue Probability | E(X*\|n*,X=x,t_ | Expected | Actual |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 5 | 3 | -11.62 | 0.02077 | 1 | 0.020775 | 2.24207 | 7 | 5 |
| 4 | 5 | 5 | 3 | -15.89 | 0.00500 | 4 | 0.020014 | 1.81643 | 5 | 1 |
| 3 | 5 | 5 | 0 | 0.00 | 0.00311 | 6 | 0.018631 | 1.39079 | 0 | 0 |
| 2 | 5 | 5 | 3 | -16.46 | 0.00415 | 4 | 0.016585 | 0.96515 | 3 | 1 |
| 1 | 5 | 5 | 0 | 0.00 | 0.01318 | 1 | 0.013176 | 0.53951 | 0 | 0 |
| 4 | 4 | 5 | 1 | -4.63 | 0.00971 | 1 | 0.009706 | 0.93635 | 1 | 1 |
| 3 | 4 | 5 | 3 | -16.16 | 0.00458 | 3 | 0.013753 | 0.94201 | 3 | 6 |
| 2 | 4 | 5 | 2 | -10.42 | 0.00547 | 3 | 0.016408 | 0.73169 | 1 | 2 |
| 1 | 4 | 5 | 4 | -16.46 | 0.01634 | 1 | 0.016336 | 0.43514 | 2 | 0 |
| 3 | 3 | 5 | 2 | -8.86 | 0.01190 | 1 | 0.011903 | 0.36281 | 1 | 1 |
| 2 | 3 | 5 | 2 | -9.47 | 0.00879 | 2 | 0.017573 | 0.45544 | 1 | 0 |
| 1 | 3 | 5 | 3 | -11.50 | 0.02164 | 1 | 0.021643 | 0.32844 | 1 | 1 |
| 2 | 2 | 5 | 2 | -7.69 | 0.02138 | 1 | 0.021381 | 0.18717 | 0 | 0 |
| 1 | 2 | 5 | 5 | -17.23 | 0.03186 | 1 | 0.031856 | 0.22314 | 1 | 0 |
| 1 | 1 | 5 | 7 | -19.83 | 0.05887 | 1 | 0.058866 | 0.12076 | 1 | 3 |
| 0 | 0 | 5 | 89 | -32.85 | 0.69139 | 1 | 0.691392 | 0.04078 | 4 | 2 |
| | | | | | | | | | Total | Total |
| | | | | | 0.92813 | | 1.00000 | | 31 | 23 |

Figure 8: A sample fitting of the BG/BB in excel over the cohort of 1996, and calculating the conditional expectations over the next four years.