An Improved BG/NBD Approach for Modeling Purchasing Behavior Using COM-Poisson Distribution

Mohamed Ben Mzoughia and Mohamed Limam

***Abstract*—The concept of Customer Lifetime Value (CLV) attempts to account for the anticipated future profitability of each customer during his lifetime with the firm. In non-contractual context, in which the firm does not observe customer defection, the measurement of the CLV metric presents the challenge of choosing the appropriate model that provides satisfactory prediction of the future purchasing behavior of customers. The most prevalent models in non-contractual setting are the Pareto/NBD and the BG/NBD which are based on statistical distributions and assume that the number of transactions follows a Poisson distribution. However, many applications have an empirical distribution that does not fit a Poisson model. In this paper we propose an improved BG/NBD approach for modeling purchasing behavior using COM-Poisson Distribution, which is a generalization of the Poisson distribution to a two-parameter distribution, offering more flexibility and fitting better real world discrete data. An empirical study based on customer credit card transactions shows that the proposed model has better forecasting performance than competing models.**

***Index Terms*—Com-Poisson, prediction, modeling, customer lifetime value.**

I. Introduction

Customer lifetime value is a customer level metric used to predict future customer behavior and to target those that are profitable. The most important challenge associate with the measurement of CLV is how to provide a satisfactory prediction of the individual customer profitability in different contexts. It is worth noting that research on CLV measurement has so far focused on specific contexts. The two types of context generally considered are non-contractual and contractual [1], [2]. The contractual context is one in which customer defections are observed, and longer customer lifetime implies higher CLV. The non-contractual context is one in which the firm does not observe customer defection, and the relationship between customer purchase behavior and customer lifetime is not certain [2], [3]. In the non-contractual context, the major task associated with the CLV measurement is the prediction of the number of transactions and the lifetime for each customer.

The most recognized models in non-contractual setting are the Pareto-NBD, proposed by Schmittlein [4], [5], and BG/NBG developed by Fader *et al*., [6]. These models assume that the number of transactions made by each customer follows a Poisson process with a heterogeneity in transaction rates across customers following a gamma distribution. These assumptions give us a Negative binomial distribution (NBD) for modeling the number of transactions made by the customer while he is alive.

Several extensions and alternatives of these models have been proposed in order to improve the quality of prediction, however, all these models assume that the number of transactions follows a Poisson distribution based on a single parameter, usually used for modeling count data. However, many real data violate the assumption of equi-dispersion that underlies the Poisson distribution.

To overcome this shortcoming, we propose an improvement to BG-NBD, by modeling the number of transactions using COM-Poisson (CMP) distribution which is a generalization of the Poisson distribution to a two-parameter distribution, providing more flexibility in modeling a wide range of over and under-dispersion and fitting better to discrete data [7]. The propose model called hereafter BG Gamma COM-Poisson (BG/GCP), will benefit from COM-Poisson properties that make it methodologically appealing and useful in practice.

The remainder of this paper is organized as follows. Section II presents the Pareto/NBD and the BG/NBD models. Section III deals with the proposed model and related expressions used both to estimates parameters and to predict future purchasing behavior. Section IV discusses an empirical study to evaluate the proposed model using real world data. Finally we summarize the merits of our modeling approach in Section V.

II. The Pareto/NBD and BG/NBD Models

The Pareto/NBD and BG/NBD models are based on the historical purchase behavior of each customer to forecast his future activity. Three past measures are required for every customer: the “cohort” T, which is the time from the entry of the customer of the company until now, the “frequency” x is the number of transactions that the customer has made after k time units and the “recency” t_x which is the time between the entry date and the last purchase date.

The Pareto/NBD model is based on six assumptions:

1) Customers go through two stages in their “lifetime” with a specific firm: they are “alive” for some period of time, and then become permanently inactive.
2) While alive, the number of transactions made by a customer follows a Poisson process with transaction rate

Manuscript received December 8, 2013; revised February 3, 2014.

Mohamed Ben Mzoughia is with the Department of Management at ISG of the University of Tunis, Tunisia (e-mail: mohamed.mzoughia@gmail.com).

Mohamed Limam is with Academic Affairs and Research at Dhofar University, Oman and Laboratory of Operations Research Decision and Control of Processes (LARODEC) at ISG of the University of Tunis, Tunisia (e-mail: Mohamed.limam@isg.rnu.tn).

λ;

3) Heterogeneity in transaction rates across customers follows a gamma distribution
4) A customer's unobserved "lifetime" of length τ is exponentially distributed with dropout rate μ;
5) Heterogeneity in dropout rates across customers follows a gamma distribution
6) The transaction rate λ and the dropout rate μ vary independently across customers.

The assumptions 2 and 3 give the Pareto Distribution and the assumptions 4 and 5 give the NBD model.

The Pareto/NBD is a well-known model, although it is difficult to implement due to computational challenges related to parameters estimation. To overcome this difficulty, Fader *et al.*, (2005) proposed an alternative of the Pareto/NBD model called BG/NBD [6]. The latter model holds the same assumptions as the Pareto / NBD model while modelling the number of transactions. On the other side, the BG/NBD model does not retain the assumptions 4 and 5, considering that the customer lifetime across customers follows a Beta-Geometric distribution and customers are assumed to defect immediately after purchasing unlike the Pareto/NBD model where it is assumed that customers can defect at any time.

As the two models give very similar results in a large variety of purchasing environments [8], the BG/NBD can be considered, in most applications, as an interesting alternative to the Pareto/NBD.

Section III deals with the proposed model, which is an improvement of the BG/NBG model based on COM-Poisson distribution, providing more flexibility and offering better forecasting performance.

III. The Proposed Model

A. Model Assumptions

The Pareto/NBD, BG/NBD and related models assume that the number of transactions made by each customer follows a Poisson process. However, many real data violate the assumption of equi-dispersion that underlies the Poisson distribution. The proposed model retains the same BG/NBD assumptions while modeling customer lifetime. However, it considers that the number of transactions follows a COM-Poisson distribution which is a two parameter generalization of the Poisson offering more flexibility in modeling discrete data.

Equation (1) displays the probability mass function of the COM-Poisson distribution:

$$P(X=x)=\frac{\lambda^x}{(x!)^\upsilon}\frac{1}{Z(\lambda,\upsilon)} \text{ , with } Z(\lambda,\upsilon)=\sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\upsilon} \quad (1)$$

Fig. 1 shows a graphic presentation of the probability mass function of COM-Poisson distribution for λ=2. When υ=1 and λ=2, the COM-Poisson distribution becomes the standard Poisson distribution. For the cases, υ<1 and υ>1, the COM-Poisson distribution describes an over or under dispersion of Poisson distribution with λ=2. The additional parameter μ offers more flexibility and better fit to discrete data.

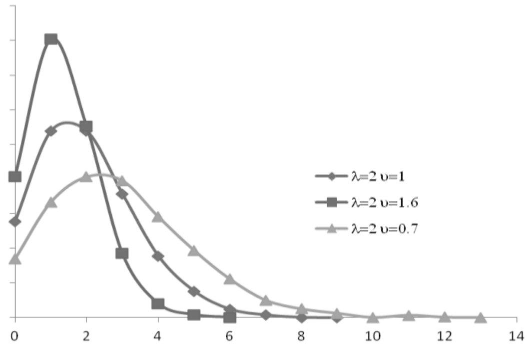

Fig. 1. Probability mass function of COM-Poisson distribution.

As presented in the Table I, the proposed model is an improved form of BG/NBD using COM-Poisson distribution to model number of transactions with heterogeneity in transaction rates across customers following a gamma distribution.

TABLE I: Models Comparison

Model	Number of transaction	Lifetime
Pareto-NBD	NBD (Poisson-Gamma)	Pareto
BG-NBD	NBD (Poisson-Gamma)	Beta Geometric
Proposed Model	COM-Poisson Gamma	Beta Geometric

The proposed CLV model is based on the following assumptions:

1) Customers go through two stages: "alive" for some period of time, and then become permanently inactive.
2) The number of transactions made by a customer follows a COM-Poisson process with parameters υ and λ. The Parameters υ is fixed across customers.
3) Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter r and scale parameter α

$$g(\lambda|r,\alpha)=\frac{\alpha^r\lambda^{r-1}e^{-\lambda\alpha}}{\Gamma(r)} \quad (2)$$

4) After any time period, a customer becomes inactive with probability p. The time period in which the customer "drops out" is distributed across transactions according to a (shifted) geometric distribution:

$$p(\text{inactive after jth time periode})=p\times(1-p)^{j-1} \quad (3)$$

5) Heterogeneity in p follows a beta distribution:

$$f(p|a,b)=\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} \quad (4)$$

where B (a, b) is the beta function, which can be expressed in terms of gamma functions:

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (5)$$

6) Parameters λ and p vary independently across customers.

The use of COM-Poisson requires a redefinition of the required data and a significant modification in the way in which we calculate model parameters.

B. Parameters Estimations

As we have already stated, the BG/NBD requires three historical measures for each customer: x, t_x and T.

The measure x is the total number of transactions made during the period of observation; in our case we need the number of transactions x_i in each time interval t_i. Indeed, we consider that the time intervals have the same duration u (eg week or month). The measures t_x and T are also assumed to be multiples of the time interval u, as displayed in Fig. 2.

For each customer, let us assume that the number of transactions xi which occurred during each time period t_i during the period (0, T] is known, illustrated as follows:

Fig. 2. Visual illustration of a typical data string.

The likelihood function is calculated as:

$$L(\lambda,\upsilon,p|T,t_x,x_1 \ldots x_x)=\frac{\lambda^{x_1}}{(x_1!)^{\upsilon}Z(\lambda,\upsilon)}\frac{\lambda^{x_2}(1-p)}{(x_2!)^{\upsilon}Z(\lambda,\upsilon)}\cdots \frac{\lambda^{x_x}(1-p)}{(x_x!)^{\upsilon}Z(\lambda,\upsilon)}\left[p+\frac{1-p}{Z(\lambda,\upsilon)}\right]^{T-tx}$$
$$=\frac{\lambda^{x}(1-p)^{t_x-1}}{(\pi_x)^{\upsilon}Z(\lambda,\upsilon)}\left[p+\frac{1-p}{Z(\lambda,\upsilon)}\right]^{T-tx} \qquad (6)$$

where $x=\sum x_i$ and $\pi_x=\prod(x_i!)$

Compared to BG/NBD, our model requires the additional variable π_x which is equal to the product of the factorial number of transactions x_i occurring during each time period t_i.

We remove the conditioning on λ and p by taking the expectation over the distributions of λ and μ to have

$$L(\upsilon,r,s,\alpha,\beta|T,t_x,x,\pi_x)=\int_0^{\infty}\int_0^{\infty}L(\lambda,\upsilon|T,t_x,x,\pi_x)g(\mu|s,\beta)g(\lambda|r,\alpha)\,d\mu\,d\lambda \qquad (7)$$

Using (5) and (6), the likelihood function is represented as:

$$L(\upsilon,\alpha,r,a,b|T,t_x,x,\pi_x)=\int_0^{\infty}\int_0^{1}\frac{\alpha^{r}}{\pi^{\upsilon}Z(\lambda,\upsilon)^{T}B(a,b)\Gamma(r)}\left[\lambda^{x+r-1}e^{-\lambda\alpha}p^{a-1}(1-p)^{t_x+b-2}\left(Z(\lambda,\upsilon)+1-p\right)^{T-t_x}\right]dp\,d\lambda \qquad (8)$$

Parameters υ, r, α, a and b can be calculated using the maximum likelihood, although it requires iterations and is computationally intensive.

C. Predicted Number of Purchases

Given that we don't know if a customer is alive at T, the expected number of purchases in the period (T, T + t] with purchase history x, t_x, π_x and T is measured as:

$$E(X(t)|\lambda,\upsilon,p,x,T,t_x,\pi_x)=E(X(t)|\lambda,\upsilon,p,alive\ \ at\ T)P(\tau>T|\lambda,\upsilon,p,x,T,t_x,\pi_x) \qquad (9)$$

As parameters λ and p are unobserved, we compute expected number of transactions by taking the expectation in (9) over the joint posterior distribution of λ and p:

$$E(X(t)|\upsilon,\alpha,r,a,b,x,T,t_x,\pi_x)=\int_0^{\infty}\int_0^{1}E(X(t)|\lambda,\upsilon,p,alive\ \ at\ T)P(\tau>T|\lambda,\upsilon,p,x,T,t_x,\pi_x)\,f(\lambda,\upsilon,p|r,\alpha,a,b,x,t_x,\pi_x,T)\,dp\,d\lambda \qquad (10)$$

Using Bayes' theorem, the joint posterior distribution of λ and p:

$$g(\lambda,\upsilon,p|r,\alpha,a,b,x,t_x,\pi_x,T)=\frac{L(\lambda,\upsilon,p|x,t_x,\pi_x,T)g(\lambda|r,\alpha)f(p|a,b)}{L(\upsilon,r,\alpha,a,b|x,t_x,\pi_x,T)} \qquad (11)$$

The expected number of purchases in the period (T, T +t] can be expressed as:

$$E(X(t)|\upsilon,\alpha,r,a,b,x,T,t_x,\pi_x)=\int_0^{\infty}\int_0^{1}\sum_{k=1}^{t}\bar{x}t(1-p)^{t}\,\frac{\lambda^{x}}{\pi_x{}^{\upsilon}Z(\lambda,\alpha)}(1-p)^{T-1}\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\frac{\alpha^{r}\lambda^{r-1}e^{-\lambda\alpha}}{\Gamma(r)}\,dp\,d\lambda \qquad (12)$$

where $\bar{x}=\sum_{j=0}^{\infty}\frac{j\lambda^{j}}{(j!)^{\upsilon}Z(\lambda,\upsilon)}$

Using parameters υ, α, r, a, b calculated using maximum likelihood and customer past purchase history x, t_x, π_x and T, the expected number of purchases can be measured of each customer for any future time period (T, T +t].

Thereby, the CLV can be computed as the discounted product of the predicted number of transactions and the expected profit per transaction.

IV. Empirical Analysis

We propose to evaluate the performance of the proposed model using real-world data model provided by a retail bank in North Africa. This evaluation is performed at two levels, an individual prediction evaluation and a segmentation performance evaluation.

A. Individual Prediction Performance

We explore a data set based on credit card transactions provided by an important retail bank in North Africa. The data set focuses on a single cohort of customers who made their first purchase in the first week of 2011. The dataset contains customers' card transactions data from January 2011 till

December 2012. The total number of transactions is 112,181 made by 1,857 customers.

As displayed in Fig. 3, the first 52 weeks data (Year 2011) are used to estimate model parameters. For the next 52 weeks (year 2012) data are used both to validate our model and to make a comparison with other models. ,

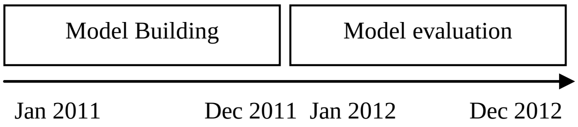

Fig. 3. Data analysis timelines.

We estimate the parameters of the Pareto/NBD, the BG/NBD and the proposed model, via MLE. Table II reports the estimated parameters and standard error of estimates.

TABLE II: PARAMETERS ESTIMATION

Channels	Pareto/NBD	BG/NBD	BG/GCP
υ			0.317
			(0.010)
r	0.259	1.555	2.301
	(0.000)	(0,003)	(0.078)
α	0.215	1.168	3.752
	(0.000)	(0,024)	(0.115)
s	0.623		
	(0.000)		
β	98.794		
	(0.171)		
a		0.420	0.993
		(0.020)	(0.041)
b		59.584	100.000
		(1,411)	(3.587)

The COM-Poisson parameter υ is 0.317, being smaller than 1 means that the number of transactions made by our 1,857 customers presents an under-dispersion case. This is a proof that the choice of Poisson distribution in modeling discrete number is not always recommended.

Table III reports the average weekly number of transactions, Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) statistics of the predicted number of transactions compared to the actual values.

The average weekly number of transactions predicted using the proposed model is the closest to the actual values.

The MAD statistic is used to evaluate various predictions. Referring to this statistic, we show that our proposed model outperforms the Pareto/NBD and the BG/NBD models with a MAD lower than 22, against values greater than 22 for competing models.

The MAPE metric shows that the prediction error from the proposed model represents about 54.23% of the average future number of transactions, against 55.47% and 54,88% for BG/NBD and Pareto/NBD models, respectively.

This analysis demonstrates the high degree of accuracy of the proposed model compared to the BG/NBD and the Pareto/NBD models while forecasting individual purchasing behavior.

The main objective of predicting CLV at the customer level is to be able to target future profitable customers. The next sub-section deals with the segmentation performance of our proposed model compared to the Pareto/NBD, the BG/NBD models.

TABLE III: MODEL COMPARISON

Model	Average weekly number of transactions	MAD	MAPE
Actual	0.836		
Pareto/NBD	1.056	25.665	54.88%
BG/NBD	0.908	22.005	55.47%
Proposed	0.838*	21.787*	54.23%*

B. *Segmentation Performance*

Customer segmentation is an important approach offering to managers the ability to better fit customers' needs and to optimize the firm's marketing resources. Recent research suggests CLV as a new base to customer segmentation [9].

To validate the ability of the proposed model to offer an improved segmentation performance in real world case, we propose to compare clusters generated using real data to those generated using predicted number of transactions. The prediction is made through the Pareto/NBD, the BG/NBD and the proposed model.

In our analysis, we segment customers using k-means algorithm, which is a partition-based method frequently used in data mining. The number of clusters is obtained by optimizing the Bayesian Information Criterion (BIC). This method has been applied successfully to decide which among two or more partitions closely matches the data. Using real data, the optimal number of clusters is 6, as displayed in Fig. 4.

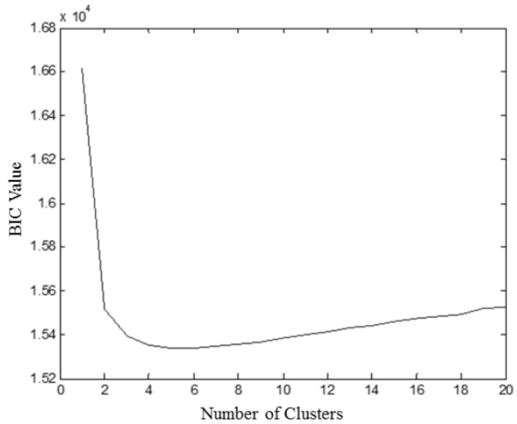

Fig. 4. Optimal bayesian information criterion.

To compare segments generated using predicted data to real segments, we use a set of well-known statistical measures. The sensitivity metric measures the proportions of customers which real segments are correctly identified using predicted number of transactions. Table IV shows that the proposed model allows identifying correctly more than 74% of customer's segments. The Pareto/NBD and the BG/NBD presents smaller sensitivity, about 59% and 73% respectively. The proposed model gives also a good precision level about 59%. The precision statistic stands for the fraction of retrieved instances that are relevant. The challenged models give lower value of the precision, 54% and 49% for BG/NBD and Pareto/NBD models, respectively.

TABLE IV: SEGMENTATION PERFORMANCE OF MODELS

Statistics	Pareto/NBD	BG/NBD	BG/GCP
Accuracy	0.50	0.56	0.62*
Sensitivity	0.59	0.73	0.74*
Specificity	0.41	0.39	0.52*
Precision	0.49	0.54	0.59*
Recall	0.59	0.74*	0.74*
F-Measure	0.53	0.62	0.66*

F-measure statistic is the harmonic mean of precision and recall which measures the effectiveness of retrieval with respect to a user who attaches more importance to recall as precision [10]. The proposed model gives an F-measure value of 0.66 which is higher than those given by the Pareto/NBD and the BG/NBD models.

This analysis demonstrates the high degree of accuracy of the proposed model compared to the BG/NBD and the Pareto/NBD models, as well its performance in predicting customer's future purchasing behavior.

V. CONCLUSION

Customer lifetime value is a relevant metric for any business activity. The difficulty faced by firms in measuring CLV is the choice of the proper model which offers a satisfactory prediction of customer purchasing behavior.

In a non-contractual relationship, the BG/NBD model is a powerful model to predict customer's purchasing behavior. This model uses the Poisson distribution to model the number of transactions, yet this distribution based on a single parameter presents a shortcoming of not being able to model real discrete data in the case of over-dispersion or under-dispersed of data. To overcome this limitation, we propose an improvement to BG/NBD model based on COM-Poisson distribution offering more flexibility and accuracy to predict future customer's transactions over time.

The empirical analysis confirm the ability of the proposed model to fit better to real data and to offer a improved segmentation performance compared to competing models. This performance is verified on a single cohort of data which presents an under-dispersion case. The segmentation performance is tested using a partition-based clustering method (i.e., the K-means method).

Further investigation is needed to evaluate the performance of our proposed model in over-dispersion case and using other clustering algorithms as the hierarchical or the spectral methods.

REFERENCES

[1] W. J. Reinartz and V. Kumar, "On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing," *Journal of Marketing*, vol. 64, no. 4, 17-35, October 2000.

[2] W. J. Reinartz and V. Kumar, "The impact of customer relationship characteristics on profitable lifetime duration," *Journal of Marketing*, vol. 67, no. 1, pp. 77-99, January 2003.

[3] P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using iso-value curves for customer base analysis," *Journal of Marketing Research*, vol. 42, no. 4, pp. 415-430, November 2005.

[4] D. C. Schmittlein, D. G. Morrison, and R. Colombo, "Counting your customers: Who are they and what will they do next?" *Management Science*, vol. 33, pp. 1-24, January 1987.

[5] D. C. Schmittlein and R. A. Peterson, "Customer base analysis: An industrial purchase process application," *Marketing Science*, vol. 13, pp. 41-67, 1994.

[6] P. S. Fader, B. G. S. Hardie, and Ka Lok Lee, "Counting your customers the easy way: An alternative to the Pareto/NBD model," *Marketing Science*, vol. 24, no. 2, 275-284, 2005.

[7] G. Shmueli, T. Minka, J. B. Kadane, S. Borle, and P. B. Boatwright, "A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 1, pp. 127–142, 2005.

[8] K. Jerath, P. Fader, B. Hardie, "New perspectives on customer 'Death' using a generalization of the pareto/NBD model." *Marketing Science*, vol. 30, no. 5, pp. 866–880, September–October 2011.

[9] S. Y. Kim, T. S. Jung, E. H. Suh, and H. S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert System with Applications*, vol. 31, pp. 101–107, July 2006.

[10] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Butterworth-Heinemann Newton, MA, USA, 1979.

Mohamed Ben Mzoughia holds an engineering degree in industrial engineering from National School of Engineering of Tunis (ENIT), Tunisia, in 2002, and he received his master degree in management of production systems, from the Same University in 2004.

Mohamed is a Ph.D. student in the Department of Management at ISG of the University of Tunis, Tunisia and a member of the laboratory of operations research decision and control of processes (LARODEC). He already had 11 years of professional experience in Banking and finance as corporate controller. His research interests include customer profitability and precisely customer lifetime value measurement and customer segmentation based on profitability.

Mohamed Limam holds a M.Sc. in statistics from Oregon State University (OSU), USA, in 1981, and a Ph.D. in statistics from the same university in 1985.

Mohamed started his career as an assistant professor at OSU, then at King Saud University, Saudi Arabia. In 1994, he moved to Tunisia as an associate professor of Statistics at the University of Tunis. Currently, he is deputy vice chancellor for academic affairs and research at Dhofar University, Oman. His research interests include the area of applied statistics, computational statistics, data mining, statistical quality control and quality management. Prof. Limam is the author of many research publications and the advisor of many Ph.D. students. He is a member of the editorial board of the International Journal of Quality and Standards. He was the head of the laboratory of operations research decision and control of processes (LARODEC). He is the founder and president of the Tunisian association of statistics and its applications from 2008 to 2011 and co-founder of the Tunisian Management Science Society.