Data Mining Application of Marketing: A Bivariate Model of Customer Purchase Monetary and Interpurchase Time

Huang, Hui-Hsin[+]

Department of Business Administration, Aletheia University, Taiwan

Abstract. In the data mining application, customer database is often used to collect the information of customer consumption behavior which can predict the future purchase patterns. Many previous researches often demonstrated that there are stronger relations between customer purchase monetary and their interpurchase time. In this paper, we propose a bivariate stochastic model to describe the relationships between these two variables. We consider that the monetary volumes of customer consumption is followed normal distribution and interpurchase time length is followed exponetail distribution. Finally, this paper also conducts the empirical data of customer transactions from a credit card company A to compute the probability of customer behavior of interpurchase and monetary consumption.

Keywords: bivariate model, customer purchase behavior, marketing.

1. Introduction

When are customers going to purchase and how much are they going to spend on their purchase are two important questions that any retailer would like to answer [1]. Knowledge of a household's timing and amount of purchase decisions will help the online store to actively target these individual households and customize its offerings [2], either to induce them to shop more frequently or in some cases less frequently. The"when" and "how much'"decisions of the household also have important implications for the pricing structure for the delivery occasion and delivery volume. Typically these vary from a pay-per-use to fixed monthly amounts with a variety of combinations of these two.

A large number of sales models have been developed in the marketing literature. Such models have been utilized to set price, promotion and advertising policies [3] and much of their intended contribution has been to give practitioners better tools for understanding their markets [4]. Researchers in marketing also have been successful in developing models of purchase quantity [5], inter-purchase times [6-7], models of purchase quantities and timing [8], and scoring models that integrate these measures while allowing for their heterogeneous covariation.

Although there are a number of models of consumer behavior that have been developed, it needs a model that would jointly predict interpurchas and purchase monetary rather than simply a model of total sales. Almost all of the few existing models in the literature for the joint distribution of interpurchas time length and consumption monetary condition on known purchases elsewhere in the store [9-11] or treat the time length and monetary volume as independent [12-13]. Thus this research conducts a bivariate model to explore the relationships between interpurchase and monetary of customers and also calculate

2. Model

According Farlie-Gumbel-Morgenstern family of bivariate distributions [14-16], there are two independent univariate distributions and combine these two into a correlated bivariate distribution. The correlation can

---

[+] Tel.: 886-2-26212121-6208.

E-mail address: hoyasophia@gmail.com.

be positive or negative and it is induced using a parameter that empirically captures pontentail dependence across the two univariate distributions [17]. The function $F_{1,2}(X,Y)$ is the joint cumulative distribution function of the Farlie-Gumbel- Morgenster distribution which is defined as follows:

$$F_{1,2}(X,Y) = F_1(x)F_2(y)\{1 + \alpha[1 - F_1(x)][1 - F_2(y)]\}$$

(1)

Where the parameter α denotes the correlation in the distributions of purchase monetary, $F_1(x)$ and interpurchase times, $F_2(y)$. The bivariate density function associated with the previous distribution function is given by

$$f_{1,2}(x,y;\alpha) = f_1(x)f_2(y)\{1 + \alpha[1 - 2F_1(x)][1 - 2F_2(y)]\}$$

(2)

In this research, we consider x as the monetary volumes of customer consumption which is following normal density with the parameters, μ and σ^2

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

(3)

Its cdf, $F_1(x)$ is as following,

$$F_1(x) = \frac{1}{2}\left[1 + erf\left(\frac{(x-\mu)}{\sqrt{2\sigma^2}}\right)\right]$$

(4)

In equation (4), $erf(\cdot)$ is the error function which is denoted by

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-2t^2} \, dt$$

We consider y as the interpurchase time length of customer which is followed the exponential density with the parameter λ.

$$f_2(y) = \lambda \exp(-\lambda y)$$

(5)

Its cumulative distribution function is as following,

$$F_2(y) = 1 - \exp(-\lambda y)$$

(6)

Then, according Farlie-Gumbel-Morgenster distribution

$$F_{1,2}(X,Y) = \left[\frac{1}{2} + 2erf\left(\frac{(x-\mu)}{\sqrt{2\sigma^2}}\right)\right][1 - \exp(-\lambda y)]\left\{1 + \frac{\alpha}{2}\exp(-\lambda y)\left[1 - 4erf\left(\frac{(x-\mu)}{\sqrt{2\sigma^2}}\right)\right]\right\}$$

(7)

And its bivariate density function is as following,

$$f_{1,2}(x,y;\alpha) = \left\{\frac{\lambda}{\sqrt{2\pi\sigma^2}}\exp-\left(\frac{(x-\mu)^2}{2\sigma^2} + \lambda y\right)\right\}\left\{1 + \alpha\left[-erf\left(\frac{(x-\mu)}{\sqrt{2\sigma^2}}\right)\right][2\exp(-\lambda y) - 1]\right\}$$

(8)

3. Empirical data analysis

We conduct the empirical data of customer transactions from a credit card company A. From 1, April 2010 to 30, April 2010, the total sample size is 2566 customers and the average monetary volumes of customer is from NT.349500 dollars to N.T.25 dollars. The transactions are from 6 (the largest) to 1(the smallest) and the largest time length of interpurchase is 29 days. The average interpurcahse time of total customer is 3.95 days.

We use these datas to compute $P(Y < y | X < x), E(X | Y = y)$ or $E(Y | X = x)$.

4. Conclusion

This paper demonstrates a bivariate stochastic model to describe the interpurchase time and monetary consumption of customers. This model can detect the relationships between these two variables and provides the some empirical reference to managers to make marketing strategy.

5. Acknowledgements

6. Reference

[1] Boatwright, P., Borle, S.and Kadane, B. J. A Model of the Joint Distribution of Purchase Quantity and Timing. In: *Journal of the American Statistical Association*. 2003, **98**(463): 564- 572.

[2] Rossi, P. E., McCulloch, R. E. and Allenby, G. M. The Value of Purchase History Data in Target Marketing. In: *Marketing Science*. 1996, **15**: 147-165.

[3] Jedidi, K., Mela, C. F. and Gupta, S. Managing Advertising and Promotion for Long-run Profitability. In: *Marketing Science*. 1999, **18**: 1-22.

[4] Bucklin, R. E. and Gupta, S. Use of UPC Scanner Data: Industry and Academic Perspectives. In: *Marketing Science*. 1999, **18**: 247-273.

[5] Schmittlein, David C. and Robert A. Peterson. Customer Base Analysis: An Industrial Purchase Process Application. In: *Marketing Science*. 1994, **13** (Winter): 41-67.

[6] Jen, Lichung, and Shih-Ju Wong. Incorporating Heterogeneity in Customer Valuation: An Empirical Study of Health Care Direct Marketing in Taiwan. In: *International Journal of Operations and Quantitative Management*. 1998, **4** (3): 217-228.

[7] Jen, Lichung, Chou Chien-Heng, and Greg Allenby . A Bayesian Approach to Modeling Purchase Frequency. In: *Marketing Letters*. 2003, **14** (1): 5-20.

[8] Boatwright, Peter, Sharad Borle, and Joseph B. Kadane. A Model of the Joint Distribution of Purchase Quantity and Timing. In: *Journal of the American Statistical Association*. 2003, **98** (463): 564-572.

[9] Bell, D. R. and Lattin, J. M. Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers prefer EDLP. In: *Marketing Science*. 1998, **17**: 66-88.

[10] Bucklin, R. E. and Lattin, J. M. A Two State Model of Purchase Incidence and Brand Choice. In: *Marketing Science*. 1991, **10**: 24-39.

[11] Gupta, S. Impact of Sales Promotions on When, What and How Much to Buy. In: *Journal of Marketing research*. 1988, **25**: 342-356.

[12] Van Praag, B. M. S. and Vermeulen, E. M. A Count-Amount Model with Endogenous Recording of Observations. In: *Journal of Applied Econometrics*. 1993, **8**: 383-395.

[13] Meghir, C. and Robin, J. –M. Frequency of purchase and the estimation of demand systems. In: *Journal of Econometrics*. 1992, **53**: 53-85.

[14] Farlie, D. J. G. The Performance of Some Correlation Co- efficients for a General Bivariate Distribution. In: *Biometrika*. 1960, **47**: 307-23.

[15] Johnson, N. L. and S. Kotz. On Some Generalized Farlie-Gumbel-Morgenster Distributions. In: *Communications in Statis- tics*. 1975, **A4** (4): 415-27.

[16] Johnson, N. L. and S. Kotz.: On Some Generalized Farlie-Gum- bel-Morgenster Distributions-II: Regression, Correlation and Further Generalizations. In: *Communications in Statistics*. 1977, **A6** (6): 485-96.

[17] Jen, L., Chou, C. H., and Allenby, M. G. The Importance of Modeling Temporal Dependence of Timing and Quantity in Direct Marketing. In: *Journal Marketing Research*. 2009, **46**(4): 482-493.