Predicting e-Customer Behavior in B2C Relationships for CLV Model

Kaveh Ahmadi
Islamic Azad University, Islamshahr Branch
Tehran, Iran

Abstract

E-Commerce sales have demonstrated an amazing growth in the last few years. And it is thus clear that the web is becoming an increasingly important channel and companies should strive for a successful web site. In this completion knowing e-customer and predicting his behavior is very important. In this paper we describe e-customer behavior in B2C relationships and then according to this behavior a new model for evaluating e-customer in B2C e-commerce relationships will be described. The most important thing in our e-CLV (Electronic Customer Lifetime Value) model is considering market's risks that are affecting customer cash flow in future. A lot of CLV models are based on simple NPV (simple net present value). However simple NPV can assess a good value for CLV, but simple NPV ignores two important aspects of B2C e-relationship which are market risks and big amount of customer data in e-commerce context. Therefore, simple NPV isn't enough for assessing e-CLV in high risk B2C markets. Instead of NPV, real option analyses could lead us to a better estimation for future cash flow of customers. With real option analyses, we predict all the future states with probability of each of them. And then calculate the more accurate of future customer cash flow. In this paper after a brief history of CLV, we explain customer behavior in B2C markets especially for e-retailers. Then with using real option analyses, we introduce our CLV model. Two extended examples explain our model and introduce the steps in finding CLV of customer in a B2C relationship.

Keywords: Customer Lifetime Value (CLV), e-Commerce Relationships, Net Present Value (NPV), customer's behavior.

1. INTRODUCTION

CLV is an important parameter in B2B and B2C relationships. Managers could make better decisions to segment customers by CLV. The majority of contributions that investigate CLVs are based on simple net present value (NPV) considerations. Using simple NPVs to assess CLVs, the supplier discounts future cash flows from a specific customer to the present date, while deducting the investment expenditure associated with the customer. Despite of its broad acceptance, NPV isn't most appropriate approach with which to value customers in e-relationships. The most important reason for this inappropriate is regarding the environmental risks (such as fluctuations in demand, changing customer's needs, technological change, changing prices and production costs). Consequently, simple NPV analyses could assess CLV when the market has no environmental risks, and we can't use NPV-based models in real high-risk markets. In addition to this, e-commerce environments are very data-rich and traditional CLV models are unable to cope with this data-richness[1].

In this paper, at first we introduce the history of CLV models, and their problems in high risk markets. Then, we describe customer's behavior in B2C e-relationships, and define a period of time for customer's buying. Then, we estimate customer cash flows in future, by using history of customer's behavior. After that, real option analyses adapt the estimated future cash flows, to find accurate customer lifetime value in future. In this paper we complete our e-CLV model in three steps. In each step, our model will be improved and calculates more accurate value. And in third step, we introduce an accurate e-CLV model for B2C high risk markets by applying real option...
analyses. Then, we compare CLV models in B2C and B2B markets. Finally, we explain future works and limitation of our model to solve in future papers.

2. HISTORY OF CLV MODELS

Since Bursk’s article "View your customers as investments" in 1966, a number of scholars have adopted the idea of using NPV-based analyses to assess the value of customers in relationships[2]. In 1985, Jackson differentiated customers to lost-for-good and always-a-share. She proposed distinct approaches with which to assess industrial customers according to their buying behavior. Lost-for-good customers buying with high switching costs therefore they are reluctant to switch suppliers. Although these customers are committed to only one vendor, always-a-share customers may buy from more than one supplier. Switching costs are lower for the latter group than the lost-for-good customers. It means that the always-a-share customers can apportion their purchases among suppliers while maintaining low costs. Jackson suggested calculating different versions of NPVs to explore the value of the two customer groups [3]. Afterward, Dwyer extended Jackson’s analysis by refining and applying it to a direct marketing context [4].

In 1998, Berger and Nasr proposed a series of mathematical models for calculating CLV in different scenarios[5], whereas much of the earlier literature had been – in their words – ‘dedicated to extolling its use as a decision criterion’[1]. In 2000, these models were subsequently re-formulated and unified by casting them into a Markov chain framework by Pfeifer and Carraway[6]. Unlike cross sectional or basic longitudinal models for predicting CLV, Markov chains can be used to explicitly model the dynamics of how CLV develops over time for a given customer[1].

In 2001 Jacobs, Johnston and Kotchetova used a NPV based model in B2B context for calculating CLV[7]. Customer value also was evaluated in B2C context by Reinartz and Kumar in 2003. To calculate CLV, they used the present values of the customers’ estimated contribution margins[8]. Many recently publication about evaluating customer based in NPV and simple NPV. The difference between present values and net present values is that opposite to the present value, the net present value concept takes investment expenditures into account[9].

Consequently, most of these models calculate CLV by (simple) NPV. And there isn’t any attention to environmental risks in these models. If customers’ cash flows remained largely unaffected by risk, NPVs would be the correct assessment method. Since many markets are currently uncertain, simple NPV methods need to be extended to assess uncertain CLVs correctly. In 2004, Adams demonstrates a new model for CLV by using real options to assess customer equity in the financial services sector. In this model Adams show how the real options approach could be applied to assess an insurance firm’s customers equity[10]. In 2007, Ellen Roemer introduces a new CLV model based on real option analyses. Her model evaluates customers in buyer-seller relationships in B2B context. Her model suggests a typology of CLV models in accordance with the degree of environmental risk and the supplier’s flexibility[11].

In this paper we want to extend real option based models to B2C context. It means that our method uses real option analyses to evaluate customers in B2C high risk markets.

3. CUSTOMER’S BEHAVIOR IN B2C E-COMMERCE RELATIONSHIPS

First of all, for a CLV model in B2C context we need to illustrate customer’s behavior in B2C e-relationships. In this type of relationship, the customer is consumer and buys products or services for consuming then amount of each trade is less than a trade in B2B relationship, and customer’s purchases in B2C relationships often aren’t on a pre-defined contract. But on the other hand, in a retailer company number of customers in B2C relationships is more than B2B relationships. Then for a simplified model, we need to divide customers into different types. Depending on average period of time between two sequential purchases of a customer we propose different customers to different types.
As we said before, e-commerce context is very data-rich, then we have a lot of e-customer's data. These two factors give a better view of customer's behavior and are essential in Normal distribution. Let \( q_i \) be amount of customer's purchase in one period, then we have \( q_0 \) for current period and \( q_t \), \( t = -1, -2, -3, \ldots \) for previous periods down to the first period of customer's purchase. Equation 1 calculates \( \mu \) mean (average) of customer's purchase and equation 2 calculates \( \sigma^2 \) standard deviation of customer's purchase for \( n \) periods (from now down to the first period).

\[
\mu = E(q_t) = \frac{\sum_{t=0}^{\infty} q_t}{n}, \quad 1
\]

\[
\sigma^2 = Var(q_t) = E[(q_t - \mu)^2] = \frac{\sum_{t=0}^{\infty} (q_t - \mu)^2}{n}, \quad 2
\]

History of customer's behavior impacts on customer cash flows in the future, this is the base of most CLV models [1][11][12]. To assess the CLV, the customer's purchase should be estimated for different future states of the world, weighed with probabilities and discounted to the present date. Amount of customer's purchase in period \( t(q_i) \) is uncertain in future periods; we can assume that in future the customer's purchase is related to past periods. In this sense, amount of purchase follows a stochastic process in which the initial volume is known today, but future volume is unknown (stochastic)[11]. We have chosen normal distribution for customer's purchase in future as it is easy to use and many scientists like Schmittlein & Peterson and the others have modeled customer behavior by normal distribution[13][14]. If future cash flow is modeled as a normal stochastic process, probabilities for different future states can be mathematically derived from the stochastic processes. We use \( \mu \) and \( \sigma^2 \) that calculated in equations 1 and 2 as parameters for normal distribution. Then we can estimated \( q_t \), \( t = 1, 2, 3, \ldots \) for future periods, it shows in equation 3. By determining \( \mu \) and \( \sigma^2 \), we could generate normal distribution numbers. Excel, MATLAB, Minitab or other mathematical software can help us to generate normal distribution numbers.

\[
q_t = Norm(t; \mu, \sigma^2) \quad 3
\]

\(^1\)Series of \( q_i \) could be derived from customer's database.

<table>
<thead>
<tr>
<th>Type</th>
<th>customer's purchases</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>These customers buy every week</td>
<td>1 week</td>
</tr>
<tr>
<td>2</td>
<td>These customers buy every two weeks</td>
<td>2 weeks</td>
</tr>
<tr>
<td>3</td>
<td>These customers buy every three weeks</td>
<td>3 weeks</td>
</tr>
<tr>
<td>4</td>
<td>These customers buy every four weeks</td>
<td>4 weeks</td>
</tr>
<tr>
<td>5</td>
<td>These customers buy sometimes</td>
<td>10 weeks</td>
</tr>
</tbody>
</table>

**TABLE 1:** Different types of customers

In table 1, different types of customers are shown and period shows the period of time between two sequential purchases of a customer, which we consider for each type. Customers in type 1 are very good and loyal customers and customers in type 5 aren't very important in B2C markets. Now, we consider \( q_i \) as amount of customer's purchase in one period, and don't pay any attention to the length of each period.
In the other side, there is a discount rate which discounts future cash flows or customer's purchase. Determining an appropriate discount rate is usually difficult. Customarily, companies expect rate of return for an equivalent investment in the capital market[11]. For simplicity, we assume that future margins will be discounted at the risk-free interest rate \( i \) (like most of other CLV models) \([15][16]\). To assess the CLV, the supplier considers \( n \) future periods. The CLV (step 1) can be calculated as a simple present value formula given in equation 4.

\[
CLV(\text{step 1}) = \sum_{t=1}^{n} \frac{q_t}{(1+i)^t},
\]

But step 1 has a big problem. It shows amount of customer's purchase not supplier's profit, we must subtract supplier's costs to determine accurate CLV. For this purpose, we assume \( x \) as a percentage of customer's purchase that is equal to supplier's profit (e.g. 10% or 15\(^2\)). Equation 5 shows CLV (step 2).

\[
CLV(\text{step 2}) = \sum_{t=1}^{n} \left( \frac{xq_t}{(1+i)^t} \right)
\]

A simple numerical example could demonstrate the CLV in step 2. Imagine a customer buys every two weeks, and average of his purchase is 40 $, and standard deviation of his purchase in previous periods is 10. We must estimate future amount of customer's purchase \((q_t, t=1,2,3,...)\). Equation 3 shows that \( q_t \) follows normal distribution, then \( q_t = \text{Norm}(t; 40, 10) \). For example in four periods \( q_1=22.03 \), \( q_2=49.36 \), \( q_3=26.13 \), \( q_4=34.16 \) (These numbers are generated by Microsoft Excel2007\(^2\)). In this example we set percentage of seller's profit \((x)=15\% \) and discount rate \((i)=10\%.\) Then according to equation 5 the CLV (step 2) is:

\[
CLV(\text{step 2}) = \sum_{t=1}^{4} \left( \frac{0.15 \times q_t}{(1+0.1)^t} \right)
\]

\[
= \frac{0.15 \times 22.03}{1.1^1} + \frac{0.15 \times 49.36}{1.1^2} + \frac{0.15 \times 26.13}{1.1^3} + \frac{0.15 \times 34.16}{1.1^4}
\]

\[
CLV(\text{step 2}) = 3.611 + 2.94 + 2.49 + 3.49 = 15.54
\]

These calculations show that the value of this customer is 15.54 $ for next four periods. In this example the length of each period is two weeks. Then four next periods are eight weeks. By this data the seller could decide to pay attention to this customer for a mid-term marketing plan or not. But yet, there is another problem in step 2. This is increasing or decreasing future customer's demand in high risk e-commerce relationships. Real option analyses could considers all the future states\([17][18]\) and solve this problem. In the next step, we use real option analyses to determine another factor \((F)\) and multiply \( F_t \) with amount of customer's purchase in each period \( t \), and then we find the accurate amount of customer's purchase in high risk markets.

\(^2\)This percentage depends to market's type or other parameters.

\(^2\)In Microsoft Excel2007 formula \( \text{NORMINV}((\text{RAND()});40;10) \) could generate numbers that follow normal distribution.
4. REAL OPTIONS IN FUTURE

In a real high risk market, the future demand can go either up or down. Customer cash flows are affected by environmental risks in many guises. Risk can appear as operational risk due to the nature of a firm’s business activities, country risks, risks from competitors’ actions, technological risk and demand-side risks [19], in addition to these in e-commerce, context switching cost is very low, and customer could switch its vendor by only one click. All types of risk result in fluctuations in demand, price and/or costs and thus have an immediate impact on the customer’s cash flows. Then we focus on demand risk that affects customer cash flows and thus follow other papers recommendation to treat risks separately [11][17]. Many markets are affected by demand risk, which complicates the projection of future cash flows. Demand for products that require high investments in production can fluctuate with a country’s economic situation. Moreover, firms developing and launching new products are confronted with high demand risk because future demand is difficult to forecast [11].

To assess the CLV, the demand should be estimated for different future states of the world, weighed with probabilities and discounted to the present date. We can assume the future demand can go up or down. As we said before factor $F_i$ is the key for this purpose. $F_i$ shows the probability of increase or decrease in each period. Then equation 6 shows the CLV model in step 3. In this step each period multiply with $F_i$ and this equation considers increasing or decreasing in future demands in a real high risk market.

$$CLV(step \ 3) = \sum_{i=1}^{n} \left( \frac{F_i}{(1+i)^t} \right)$$

For $F_i$ we have chosen a binomial approach for increasing or decreasing future demand as it is easy to use and can span a large range of applications (this corresponds to [11][18][19]). If future demand is modeled as a binomial stochastic process, probabilities for different future states can be mathematically derived from the stochastic processes. For binomial stochastic we need three factor $u$ (increasing factor in each period), $d$ (decreasing factor in each period) and $p$ (probability of happening increased state in the next period).

Demand is known today but uncertain in future and follows a multiplicative binomial process in which demand can either improve by factor $u>1$ in future, or decrease by $d<1$. Therefore, $F_i$ at the first period $(t=1)$ is either $u$ or $d$. The probability $p$ of an increase in demand in the next period can also be derived from mathematical process. It is clear that if $p$ is probability of increase in next period then $(1-p)$ is probability of decrease in next period. Then at the first period $(t=1)$ the customer purchase must be multiplied with $F_1 = (p.u+(1-p).d)$ and in the other periods the customer’s purchase must be multiplied with $F_t = (p.u+(1-p).d)^t$. This binomial approach is used in many papers that estimated future demands with real option analyses [11][17][18]. All of the factors $u,d$ and $p$ can be calculated by historical data from customer’s database. Equations 7, 8, 9, 10 show $F_i$ in four periods $(t=1, 2, 3, 4)$.

$$F_1 = p.u+(1-p).d$$
$$F_2 = (p.u+(1-p).d)^2 = p^2.u^2+2.p.u.(1-p).d+(1-p)^2.d^2$$

Equation 11 shows our final CLV model (step 3).
As it can be seen in above equation, in this model different future potential demands are calculated (e.g., $u^3q_3$), multiplied with the supplier’s profit $x$, weighed with probabilities $p^3$ and discounted to the present date. If upward and downward factors converged towards each other so that $u=d=1$, demand would remain constant and CLV (step 2) would overcome. Similarly, if $p$ converged towards 1, the future upward state would become certain and step 2 would again overcome.

For better understanding, we extend previous example to step 3. In the previous example we have a customer that buys every two weeks. Mean of his purchase’s amount is 40$ and standard deviation of his purchases is 10. We found that if there isn’t any market’s risk, he approximately gives 15.54 $ profit to the seller, in next four periods. In real e-commerce markets we have many risks; these risks make some fluctuations in customer’s demand. As we said before, for these fluctuations we have $u, d$ and $p$ to model customer’s purchase by binominal distribution. These factors can be estimated by historical data and other information about market’s risks. In this example the seller expects demand to either grow by factor $u=1.4$, or go down by factor $d=0.8$ in future. The probability that demand will increase in the next period is $p=0.5$, and the probability that demand will decrease is $(1-p) =0.5$. Real option analyses according to equations 8, 9, 10 generate an event tree for customer’s cash flow in four next periods.

Figure 1 shows this event tree; in each node of this tree we show one state of future period. As we said in previous example, in four periods $q_1=22.03$, $q_2=49.36$, $q_3=26.13$, $q_4=34.16$ (These numbers are generated by Microsoft Excel2007).
This tree shows many facts, for example in third period ($t_3$) estimated amount of purchase $q_3$ was equal to 26.13 $. But the event tree shows that in a real relationship market's risk may affect this amount. And amount of customer's purchase in $t_3$ is a variable amount from 13.37 $ to 71.70 $. In this example, we suppose the probability of increase and decrease equal to 0.5 ($p= (1-p) =0.5$), then probability of appearance of each state in one period is equal. Now the final amount of each period could be calculated by multiplying the probability of each state to the amount of its, the below calculations show the final CLV. In this example we set percentage of seller's profit ($x$)=15% and discount rate ($i$)=10%.

$$CLV_{(stages)} = \frac{0.15 \times [0.5 \times 30.82 + 0.5 \times 17.62]}{1.1} + \frac{0.15 \times [0.5^2 \times 96.74 + 2 \times 0.5^2 \times 55.28 + 0.5^2 \times 39.48]}{1.1^2}$$
The final CLV (step3) is $20.52$, then total profit that is gained from this customer in next four periods (eight weeks) is $20.52$. In this example we supposed increase factor \( u=1.4 \) and decrease factor \( d=0.8 \), then in this market total trend of customer's demand is increasing. Comparing CLV in step3 (20.52) and step2 (15.54) shows this fact.

5. COMPARING CLV MODELS IN B2C AND B2B MARKETS

This paper is the third paper of our series of papers in developing a CLV valuation model based on environmental risk. In our first paper in this sequence, which is used in ref. 12, we describe our model for B2B high-risk relationships. On that paper, we divided B2B relationships into four different types, and introduce a CLV evaluating model for each type. That paper used real option analyses to find accurate customer value in B2B markets. That model had some limitation that makes it suitable only for B2B markets. In this paper we change our approach and develop a model based on real option analyses for B2C markets. The aim of this paper is to develop a CLV model by considering environmental risks in B2C e-commerce markets. Comparing these CLV models in B2B and B2C relationships lead us to two differences:

1. Usually, in B2B markets buying are done on long term or mid term contracts, then amount of buying and number of customers approximately are constant and pre-defined in contracts. Then estimating future demand is easier than B2C markets. We used only binomial distribution to determine how environmental risks make future demands to grow up or down in B2B markets. But B2C relationships almost haven't any contract and customer's demand fluctuates very fast in future. Then at first, we must estimate future demand by using normal distribution (mean and standard deviation calculated by history data of the customer) and then adapted it with a binomial distribution to show how environmental risks make demands to grow up or down in B2C high risk markets.

2. In our first paper we discussed about supplier's flexibility that is very important in B2B relationships. If supplier is flexible when customer's demand grows up, it will respond to the increasing demand. But if supplier's investment is limited, it won't respond to the increasing demand[11]. On that model we subtracted capital expenditure from customer value to calculate accurate CLV. Then our CLV model in B2B markets suggests to increase or decrease investment for a customer. Therefore, CLV model is a good way to make investment strategy in B2B relationships. But in B2C markets we face to a big majority of customers, and our CLV model in B2C market divides customers into appropriate segments. In this way, company could make different marketing plans for different segment of customers. Therefore, CLV model in B2C markets could make good marketing strategies instead of investment strategy.

6. CRITICAL DISCUSSION

As we said in the first of paper, most of CLV models were based on history of customers. It means that many of them calculate cash flow of each customer in previous periods and determine a value for each customer. This is a good way for determining customer value, but there are many environmental risks in real markets that cause customer's cash flow to decrease or increase in future periods.

In proposed model, we used real option analyses to predict future periods. Equation 11 and the following example in figure 1 show usage of real option analyses in this model. Figure (1) shows...
that all the future periods are calculated in this model and then the formula multiplies the value with their occurrence probably and finally these weighed values add to each others.

In addition, another difference in the proposed model shows in equation 3 and 4. In this model we didn't consider \( q \) (the customer's cash flow) a constant value. Instead of that, we consider customer's cash flow in future periods following a normal distribution. This assumption helps us to achieve a solution for B2C context.

The proposed e-CLV model has important implications for e-marketing theory and practice. The presented model according to the environmental risk is essential for the development of the CLV construct. Specifically, real option analysis represents a fundamentally new way of analyzing CLVs in e-relationships. To improve relationship management, continuous monitoring of CLVs in relationships is necessary [11]. Customer relationship management (CRM) systems may facilitate the assessment of CLVs by using data from marketing, market research and management accounting. CRM systems are especially useful for planning marketing strategy. Nowadays, websites compete with each other by using better and more accurate CRM systems.

Our model uses some mathematical functions to calculate CLV. These functions could be developed by software. Then CRM software could use our CLV model to determine customer's value automatically, and suggest different marketing campaign for different customers.

6.1 FUTURE WORKS AND STUDIES

Our model evaluates e-customer in high risk B2C markets. It works very accurately because it uses real option analyses to determine customer's future cash flows in each period of time. But there are two limitations in our models, that could be solved in future works.

The first limitation is that input parameters have to be estimated for the valuation metrics and are thus subject to an estimation error. Various estimated data are necessary to assess the CLV in buyer–seller relationships. Real option model is highly sensitive to the underlying input parameters such as probability of increase or decrease demand in the future \( (p) \), increase or decrease factor for demand in next period \( (u,d) \) and discount rate \( (i) \). In the case of financial options, these parameters could be determined from historical market data. In the context of real options, these data are sometimes difficult to obtain, especially if the asset is not traded and market data are not available for the asset. Consequently, the necessary data have to be subjectively estimated and could be a source of error[20]. For future works it is a good idea to develop a model to estimate these parameters, accurately.

Secondly, Our CLV analyses were based on economic parameters. Soft factors such as trust, social bonds or closeness that contribute to CLVs in relationships are difficult to include, as they are generally difficult to quantify[11]. However, they may be introduced as moderating variables on a qualitative basis.

6.2 CONCLUSION

We develop a new CLV model to evaluate e-customer value in B2C high risk markets. Real options analyses help us to determine exact future cash flow of a customer and calculate more accurate CLV than NPV-based models. In addition, e-commerce context is very data-rich[1] and our CLV model by mathematical functions could use all of these data to find better CLV. Real option analyses lead us to binomial distribution that estimate future increased or decreased demands. One of the most important consequences of evaluating customers in our model is better decision in marketing management. This model is a good way to segment customers. By suitable segmenting, we can use better marketing strategy for each segment. In addition, if customer relationship management (CRM) systems use CLV model to evaluate customers, those could suggest better ideas for managing relationships between companies and their customers[21]. This paper's CLV model could be used in many CRM systems. And this model lead CRMs to better decision in high risk and real markets.
7. REFERENCES


