MANAGERIAL IMPACTS OF DIFFERENT COMPUTATION MODELS FOR CUSTOMER LIFETIME VALUE FOR AN E-COMMERCE COMPANY

Pavel Jašek – Lenka Vraná

Abstract

Long-term profitability of a customer is a really important subject for commercial organizations. The concept of Customer Lifetime Value helps companies to manage their acquisition and retention marketing activities based on appropriately discounted net contribution margin achieved per customer.

There are six main approaches for Customer Lifetime Value Modeling which are based on advanced statistical, econometric and data mining algorithms. Authors describe profoundly three of these approaches (Recency, Frequency and Monetary Value models, probability models and persistence models) on theoretical basis and compare their results on real-world data from an online fashion retailer.

Managerial implications of usage of different models are discussed. Companies need to decide what business goals are they trying to achieve with CLV analysis and predictions. Main decisions should be made on a level of detail (individual vs. aggregate) and a predicted variable (number of transactions, customer value or probability of being alive only). The article concludes specific use cases for various models: RFM serves great as an introductory tool to customer segmentation, Pareto/Negative Binominal Distribution predicts customer's probability of being alive and Vector Autoregressive Model interprets various relationships among customer acquisition metrics.

Key words: Customer Lifetime Value, Pareto/Negative Binominal Distribution, Econometrics Modeling

JEL Code: M21, C52, M31

Introduction

Customer Lifetime Value (CLV) is defined as the net present value of all the profits that a specific customer brings to the firm (Berger & Nasr, 1998). It can serve as an indicator of profitable individuals. Customer Equity (CE) is then sum of CLV of all the current and the future customers and can therefore serve as a tool how to measure the firm's performance.

(Gupta, Hanssens, Hardie, & Kahn, 2006) classify the CLV (or CE) modeling techniques into six branches: 1) *recency, frequency, monetary value (RFM) models* predict behavior of the customers in the next time period only, 2) *probability models* use probability distributions and their combinations (e.g., Pareto/NBD) to predict whether the customer will still be active in the future and if so, how will he act, 3) *econometrics models* combine models of customer acquisition, retention and expansion to estimate CLV, 4) *persistence models* also work with these elements, but treat them as a dynamic system and use methods of multivariate time series analysis, 5) *computer science models* consist of various data mining and machine learning algorithms and 6) *diffusion/growth models* focus on predicting customer equity.

We select three of these approaches (RFM models, probability models and persistence models), describe them in detail and compare their results.

1 Comparison of CLV Models Computations

Although we work only with three types of the algorithms described above, the comparison is very difficult. The RFM and probability models use detail data about customers and can predict CLV for each individual as well as CE. Persistence models analyze aggregated data and therefore can be used only to forecast CE. This leads to different managerial use cases.

We had to define metrics, which would enable us to evaluate the results, similarly to (Wübben & von Wangenheim, 2008). We decided to divide the dataset into two pairs of training and validation data. We divide the dataset into 50/50 and 90/10 time periods in order to compare long-term prediction based on short history and short-term prediction based on long history. We use the trained models to score one particular customer (for CLV estimation) and the whole customer set (to predict CE). Chapter 5 contains the actual comparison of the estimated and actual values.

1.1 Dataset Description

For the purpose of this article we have used real-world data from a Czech online retailer pseudo named MP. The business sells fashion primarily to mid-aged women and regularly twice a year changes a large portion of product catalogue in order to match summer and winter season. According to the customer base classification done by (Fader & Hardie, 2009, p. 63), this dataset has non-contractual relationship with customers and continuous opportunities for transactions.

This historical log contains 77 289 logged-in visits to the e-commerce website and 33 613 online purchases made by 29 589 different customers from the time period of September 1, 2011 to March 31, 2014 (134 weeks in total). Cut-off dates for 50/50 model validation would be December 15, 2012 and for short validation December 31, 2013.

Thoroughly in this article we would illustrate some calculations using one selected customer called Alice. Data sample in Table 1 gives a preview of sales data for such customer. In total Alice had 17 transactions of total revenue 14 162 CZK, dating from November 17, 2011 to December 12, 2013. In comparison with the average of 1.74 transactions and 2 086 CZK average sum of revenues per active customer it is evident Alice is an outlier, yet her example still serves great for explanative purposes.

Tab. 1: Dataset sample for Customer #22862 (Alice)

Row ID	**Customer #**	**Date**	**Sales (CZK)**
2712	22862	2011-11-17	1 540
3336	22862	2011-11-26	434
9257	22862	2012-05-05	1 120
9550	22862	2012-05-11	273

Source: Sample dataset for online retailer MP.

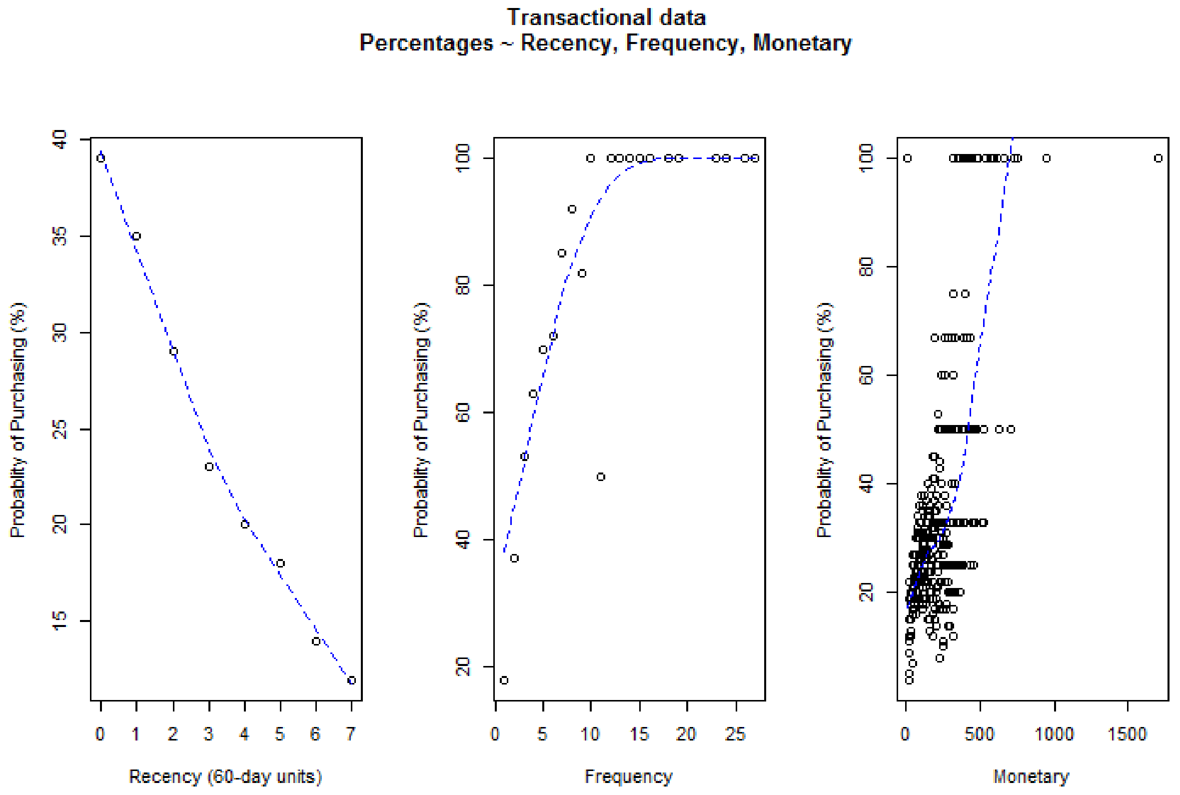
Unfortunately, the dataset does not include any cost data related to marketing expenses. One of ordinary assumptions for CLV modeling is its focus on profit and not only revenue. For the purposes of this article we will be consciously breaking this rule relying only on revenue metrics as we're not comparing different models of profitability computation. Once we started segmenting data by marketing channels, we would extremely need cost data in order to deliver valid results.

2 RFM Modeling

RFM analysis is a marketing technique used for analyzing customer behavior such as how recently a customer has purchased (recency), how often the customer purchases (frequency), and how much the customer spends (monetary), as defined in (Bult & Wansbeek, 1995).

Traditionally, this discretization relies on quintiles. Implementation of RFM analysis done by (Han, 2013) discretized recency according to expected purchasing cycle, frequency according to number of purchases and monetary according to value intervals. Han followed this discretization by building and visualization of generalized linear model of quasibinominal family with logit link that we've used in this article as well. The graphical and statistical output can be seen in Figure 1 and Table 2, respectively.

Fig. 1: Scatter plot of RFM components for Probability of Purchasing



Source: calculated by the authors using R software and (Han, 2013) R script based on sample data.

Visualization provided in Figure 1 serves well for managerial discussions. The scatter plot suggests that there is an obvious linear or exponential fall relationship between the repurchase percentage and the Recency, and an obvious exponential rise relationship between the repurchasing percentage and the Frequency. However, there is no obvious relationship between the repurchasing percentage and the Monetary. Such results are completely aligned with CDNOW dataset examined by (Han, 2013).

Tab. 2: RFM Model estimated parameters for the sample dataset

Coefficient	Estimate	Standard Error
Intercept	-1.6768	0.0647 ***
Recency	-0.1500	0.0107 ***
Frequency	0.5915	0.0247 ***
Monetary Value	0.0009	0.0003 ***

Source: calculated by the authors using R software based on sample data. *** Significant at 1% level. Quasibinomal logit model for Buy ~ Recency + Frequency + Monetary.

Estimated coefficients for parameters in Table 2 indicates that historical purchases don't matter that much, but in contrary recent purchases and moreover the number of purchases is a really strong predictor of purchase behavior.

RFM Model in Table 2 can be used for individual predictions based on customer purchase information. In the 50/50 training data set, Alice made her last purchase on November 24, 2012, thus her Recency bucket is 0, Frequency is 7 transactions and Monetary value is 92 due to 10 CZK discretization. Predicted probability of Alice's repurchasing is 93.03 %. Expected CLV for validation period for Alice given 2 % discount rate is 580.87 CZK.

3 Probability Models

Pareto/NBD model implementation by (Fader, Hardie, & Lee, 2005) simplifies original Schmittlein's computation. This model serves for repeat-buying behavior in a non-contractual setting and is based on assumptions that the number of transactions made by a customer follows a Poisson process with transaction rate λ and that heterogeneities in transaction and dropout rates across customers follow a gamma distribution with shape parameters r and s, respectively, and scale parameters α and β, respectively.

Application in this article was inspired by 1) (Wadsworth, 2012) in his implementation of R's package BTYD (Buy 'Til You Die) by (Dziurzynski & Wadsworth, 2012), 2) (Baggott, 2013) and 3) package BTYD_plus that extend the basic package (Platzer, 2008).

Results of Pareto/NBD model calibration for sample dataset are shown in Table 3 and discussed below.

Tab. 3: Pareto/NBD Model estimated parameters for the sample dataset

Pareto/NBD model parameters	***r***	***α***	***r/α***	***s***	***β***	***s/β***	***Log-Likelihood***
Weekly aggregation, 50/50 training	*0.5699*	*33.4427*	*0.0170*	*0.2868*	*7.7247*	*0.0371*	*-27594.94*
Weekly aggregation, 90/10 training	*0.5446*	*39.0630*	*0.0139*	*0.4244*	*19.9702*	*0.0212*	*-62074.56*

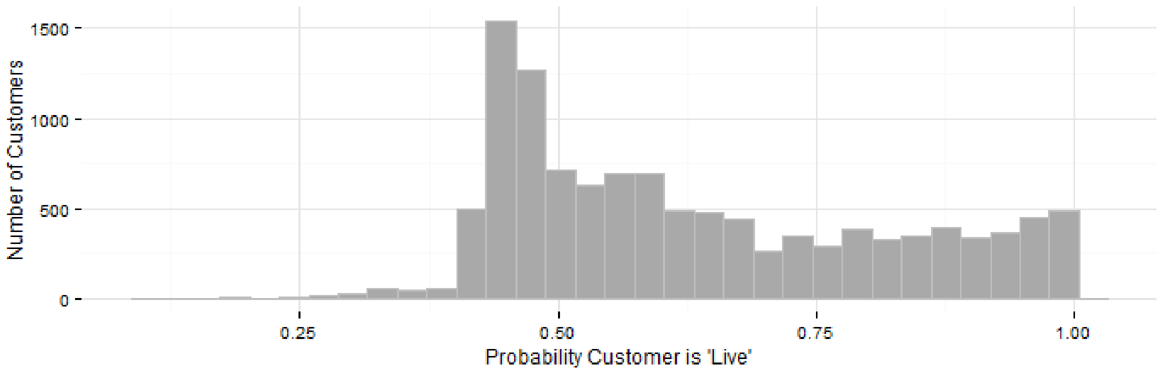
Source: calculated by the authors using R software based on sample data.

According to the interpretation of Pareto/NBD parameters described by (Wübben & von Wangenheim, 2008), r/α represents the number of purchases of an average customer in one time unit. In our dataset an average customer makes 0.02 transactions per week. The lifetime of an average customer is exponentially distributed with parameter s/β and has an expected value of 1/(s/β), where s/β represents the dropout rate of an average customer per time unit. An average customer remains active for 27 weeks, or 189 days respectively.

$$E[X(t) \mid r, \alpha, s, \beta] = \frac{r\beta}{\alpha(s-1)}\left[1-\left(\frac{\beta}{\beta+t}\right)^{s-1}\right] \tag{1}$$

Formula 1 proposed by (Fader, Hardie, & Lee, 2005) computes the expected number of repeat transactions in a period of length *t* for a random customer. For 52-week long period, this metric is 0.61. For Alice, expected number of transactions conditional on her past behavior (e.g. frequency 6 transactions, time of last interaction 53.29 weeks, 56.29 total time observed) is 3.42 in the same 52-week long period.

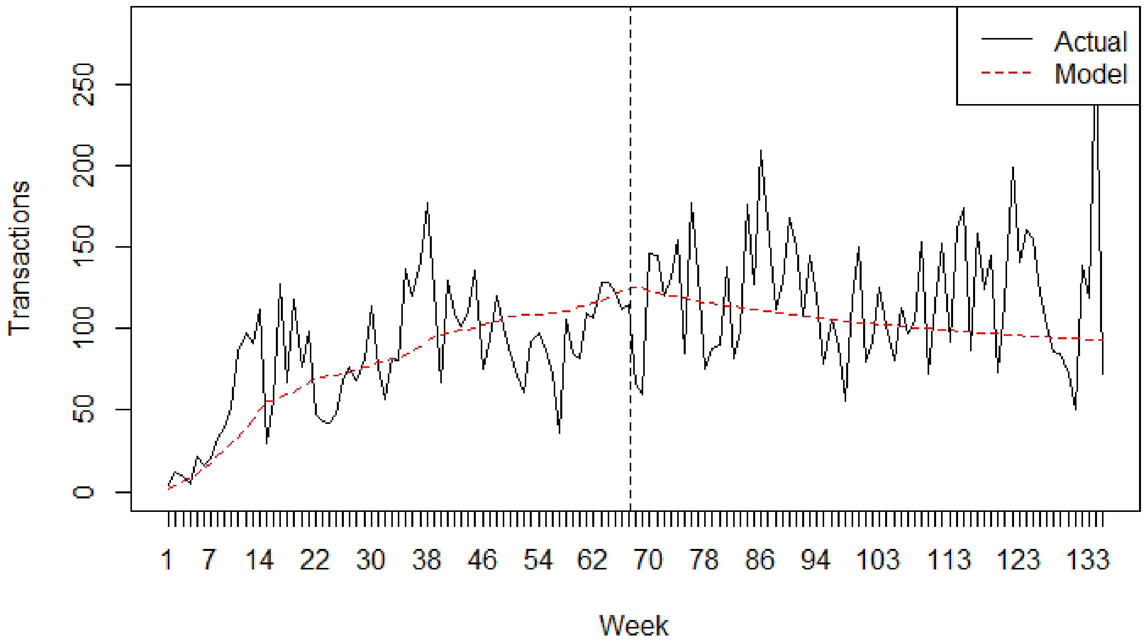
Fig. 2: Distribution of customers by probability of being alive



Source: calculated by the authors using R software based on sample data and script by (Baggott, 2013).

Figure 2 shows very positive results for the company: the majority of current customer base has probability of being alive higher than 0.5. Yet as we would see in Figure 3, the predictive ability is not ideal.

Fig. 3: Comparison of transactions by repeat buyers in Pareto/NBD model



Source: calculated by the authors using R software based on sample data of 13 661 transactions and 50/50 training/validation set. Transactions from customers with 1 purchase only were removed.

Figure 3 shows the output of Pareto/NBD model with actual number of transactions. It is visually understandable that overall performance of the model is good, yet it underestimates weekly deviations.

(Fader, Hardie, & Lee, 2005) also mention that a weakness of Pareto/NBD is its possibility of predicting number of transactions only.

4 Persistence Models

In 2008 Villanueva et al. described application of vector autoregressive model (VAR) to customer equity predictions. They researched impacts of customer acquisition on the company's performance. They examined the differences between customers gained by marketing activities and customers acquired spontaneously.

The model is designed as the classical VAR(p) model, where *p* stands for the number of lags. It captures dynamic relationships between three time series: number of customers acquired by marketing actions (*MKT*), number of customers acquired by word of mouth (*WOM*) and the firm's performance (*VALUE*):

$$\begin{pmatrix} MKT_t \\ WOM_t \\ VALUE_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \sum_{l=1}^{p} \begin{pmatrix} a_{11,l} & a_{12,l} & a_{13,l} \\ a_{21,l} & a_{22,l} & a_{23,l} \\ a_{31,l} & a_{32,l} & a_{33,l} \end{pmatrix} \begin{pmatrix} MKT_{t-l} \\ WOM_{t-l} \\ VALUE_{t-l} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{pmatrix}, \qquad (2)$$

where t stands for time, vector $(c_1, c_2, c_3)'$ contains the constant terms and vector $(e_{1,t}, e_{2,t}, e_{3,t})'$ contains the error terms with Gaussian white noise properties. Model in this form can describe the following relationships (Villanueva, Yoo, & Hanssens, 2008):

- *direct effects* of acquisition on the firm's performance (coefficients $a_{31,l}$ and $a_{32,l}$),
- *cross-effects* between two types of customer acquisition (coefficients $a_{12,l}$ and $a_{21,l}$),
- *feedback effects*, which states how the firm's performance affects the acquisition in the next time periods (coefficients $a_{13,l}$ and $a_{23,l}$),
- *reinforcement effects*, when value of series in time t affects its future values, e. g. customers acquired by word of mouth would spread the positive information about the firm which would lead to more acquisitions (coefficients $a_{11,l}$, $a_{22,l}$ and $a_{33,l}$).

Villanueva et al. discovered that customers gained by marketing promotions generate higher value in short term. However, customers acquired spontaneously had greater impact in long-term evaluation. We try to apply their approach on online retailer data and compare the results.

As we are missing more detailed data, we define newly acquired customers in time t as customers, who make their first purchase. Therefore we use first year data as a purchase history and we don't include them in the analysis. We work with weekly time series from September 1, 2012 to December 31, 2013. We keep 2014 data as the validation set.

We add data from Google Analytics to distinguish between *MKT* and *WOM* customers. Villanueva et al. used number of log-ins as *VALUE* series as they were working with data from internet firm that provided Web hosting. We tried to use income as firm's performance indicator, but there was no significant dependency of income on number of acquired customers, so we used number of purchases. Because the new customer is not identified until he makes her first purchase, the analysis focuses on any subsequent purchases.

All the series are tested for unit root by augmented Dickey-Fuller test and KPSS test and are recognized as stationary (their means and variances are time invariant). To minimize the Akaike information criterion we fitted VAR(1) model:

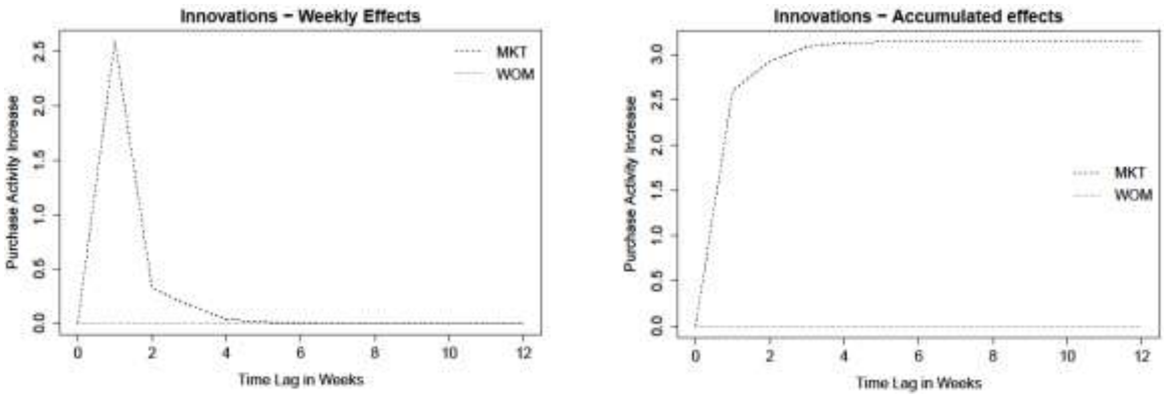
$$\begin{pmatrix} MKT_t \\ WOM_t \\ VALUE_t \end{pmatrix} = \begin{pmatrix} 68.51 \\ 40.82 \\ 237.96 \end{pmatrix} + \begin{pmatrix} 0.92 & 0.00 & -0.27 \\ 0.75 & 0.59 & -0.32 \\ 2.60 & 0.00 & -0.80 \end{pmatrix} \begin{pmatrix} MKT_{t-1} \\ WOM_{t-1} \\ VALUE_{t-1} \end{pmatrix}. \quad (3)$$

The coefficients a_{12} and a_{32} are insignificant and are set to zero, thus there is no direct effect of *WOM* customers on firm's performance and no cross effect of *WOM* customers on *MKT* acquisitions.

The actual number of purchases made during the validation period (from January 1, 2014 to March 31, 2014) is equal to 2 781. The model predicts 2 881 purchases in the corresponding period, however the predictions quickly revert to the mean.

Based on the fitted model we also create impulse response functions (Figure 4), that show the response of *VALUE* series to newly acquired customer via marketing promotion or word of mouth. The effect includes not only purchases made by the new customer, but also purchase activity of others which could have been encouraged by the newcomer (Villanueva et al., 2008).

Fig. 4: Direct effects of customer acquired through marketing promotions (MKT) and customer acquired spontaneously (WOM) on number of purchases (VALUE)



Source: calculated by the authors using R software based on sample data.

As the model doesn't find any direct effects of *WOM* on firm's *VALUE*, the impulse response function (weekly effects as well as accumulated) is constant and equal to zero. This means that the *WOM* customers usually make only one purchase (the one when they are identified as new customers) and no more. This zero effect of *WOM* is in contrast with the study done by (Smutný, Řezníček, & Pavlíček, 2013), where customers of a studied telecommunications company influenced their own interactions more than communications activities of the studied brand itself, thus impacting positively *WOM* channels.

The function of weekly effects shows that each unexpected acquisition made through the marketing channel generates 2.60 additional purchases during the first week and then the effect fades. The new *MKT* customer causes 3.14 additional purchases during her whole lifetime.

For the case of Alice, who is customer gained by the marketing activity, VAR model expects her to generate 4.14 purchases during her whole lifetime. This value underestimates the actual number as Alice is an outlier customer.

The results of our analysis are opposite to the results of Villanueva et al. Their company's value is affected mostly by *WOM* customers; our model suggests that the *WOM* customers don't have any significant impact on the firm's performance after their first purchase.

5 Discussion and Managerial Impacts

As mentioned in chapter 1, the comparison is not easy. In this article we researched three different types of models that can be used in CLV computations, but each of them yields diverse results and is based on inconsistent assumptions even though sharing many e.g. convenience in non-contractual settings.

5.1 Model Comparison

Primary implementations of models used in this article could not compete with entire comparisons of stochastic models in the paper of (Wübben & von Wangenheim, 2008), in which many important metrics were proposed. Table 4 shows some of comparable results. Pareto/NBD results for number of transactions is significantly lower than actual data. Meanwhile, VAR model's estimation of 2 881 transactions in validation period compared with 2 781 transactions in actual data is a great result (moreover, counting aggregate-level only).

Tab. 4: Comparison of CLV computation models

Statistic for validation period	**Actual data**	**RFM**	**Pareto/NBD**	**VAR**
Transactions in 50/50 test method	8 180	N/A	7 137	N/A
Transactions in 90/10 test method	2 781	N/A	1 742	2 881
Transactions for Alice in 50/50 test method	10	N/A	0.18	N/A
Purchase value for Alice in 50/50 test method	7 749	5 809	N/A	N/A
Transactions for Alice in 90/10 test method	0	345	N/A	N/A

Source: calculated by the authors using R software based on sample data. VAR model training period started on September 1, 2012.

Pareto/NBD implementations used in this article understood Recency as a difference of last and first transaction date. In contrary, RFM Model captures Recency as the difference between the studied end date and last transaction date.

5.2 Discussion on Managerial Implications

From such incomparable results it is noticeable that the company needs to decide what business goals it is trying to achieve with CLV analysis and predictions. Main questions should lead to decision on level of detail (individual vs. aggregate) and predicted variable (number of transactions, customer value or probability of being alive only).

Each model serves its purpose and can be adapted for specific business goals. RFM Model exhibited clear visual explanation of its factor strength, Pareto/NBD worked well with probabilities of customers being alive and vector autoregressive model indicated possible relations between variables and various time shifts.

RFM Model demonstrated the prediction strength of Recency and Frequency and prediction weakness of Monetary value. High-value historical purchases doesn't implicate future transactions.

On the other hand, high probability of repurchasing by very recent customers with high number of transactions can lead to managerial decisions of increased marketing activities on this segment of customers. For future research, additional information about marketing activities should be incorporated.

According to (Wübben & von Wangenheim, 2008) the performance of heuristic estimation of customer value and segmentation can be as good as stochastic models. The paper also states that one third of top 20 customers gets predicted badly. Managerial implications are very clear – another customers should deserve being part of top 20 privileges, but the model does not nominate them. Such customers could spread negative feedback about the company.

Another criticism mentioned by Wübben et al. is the reliability of customer-centric models when training and predictions results only for several transactions in long time period only.

VAR models in CLV analyses can characterize behavior of the average customer gained by marketing campaigns or by word of mouth. In case of online retailer, we find out that *WOM* customers aren't loyal and that they don't affect the performance of the firm after their first purchase. On the other hand, customers acquired by marketing promotion generate 4.14 purchases during their whole lifetime.

VAR models use aggregated data and do not allow analysis of data with finer granularity. They could be used to forecast the future firm performance, but the predictions (in our case) are quickly reverting to the mean and don't capture the actual fluctuations well.

The advantage of VAR models is their scalability: different indicators can be used as *VALUE* series and therefore we can examine different sets of relationships. One of the possible extensions could be to include more than one *MKT* series to assess the impacts of several marketing campaigns. The multivariate time series analysis is well described topic nowadays and so the possibilities of its application are wide.

Conclusion

This paper presented three main approaches to computations used in CLV analysis and modeling. RFM, Pareto/NBD and VAR models were compared on real-world data from an online fashion retailer based in the Czech Republic. Extensive dataset helped with apt visualizations of model results and plenty of validation possibilities.

During the research, many important questions and possible research extensions arose. Estimation of profit data in case of Pareto/NBD and VAR models and transactions in case of RFM predictions are two of the main fragilities of such modeling.

Various managerial impacts were discussed. RFM served great as an introductory tool to customer segmentation, Pareto/NBD predicted customer's probability of being alive and VAR model interpreted various relationships among metrics.

Acknowledgement

This paper was written with the financial support of IGA grant F4/18/2014.

References

Baggott, M. (2013). *Predicting Customer Behavior with R: Part 1*. Retrieved from Slideshare: http://www.slideshare.net/mattbagg/baggott-predict-customerinrpart1

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: marketing models and applications. *Journal of interactive marketing, 12(1)*, 17–30.

Birant, D. (2011). Data Mining Using RFM Analysis. *Knowledge-Oriented Applications in Data Mining*. InTech.

Bult, J. R., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. *Marketing Science* (pages 378-394). Catonsville: Informs.

Dziurzynski, L., & Wadsworth, E. (2012). *BTYD: Implementing Buy 'Til You Die*. Načteno z R package version 1.0: http://cran.r-project.org/web/packages/BTYD/BTYD.pdf

Fader, P. S., & Hardie, B. G. (2009). Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing*, pages 61-69.

Fader, P., Hardie, B., & Lee, K. (2005). *A Note on Implementing the Pareto/NBD Model in MATLAB*. Retrieved from BruceHardie.com: http://www.brucehardie.com/notes/008/pareto_nbd_MATLAB.pdf

Gupta, S., Hanssens, D., Hardie, B., & Kahn, W. (2006). Modeling Customer Lifetime Value. *Journal of Service Research : JSR*, pages 139-155.

Han, J. (2013). *Calculating Customer Lifetime Value with Recency, Frequency, and Monetary (RFM).* Retrieved from Data Apple: http://www.dataapple.net/?p=133

Platzer, M. (2008). *Stochastic Models of Noncontractual Customer Relationships.* Vienna University of Economics and Business Administration.

Smutný, Z., Řezníček, V., & Pavlíček, A. (2013). Measuring the Effects of Using Social Media in the Marketing Communications of the Company: Presentation of Research Results. *IDIMT-2013 Information Technology Human Values, Innovation and Economy. Prague* (pages 175-178). Linz: Trauner Verlag.

Tsay, R. (2013). *Multivariate Time Series Analysis: With R and Financial Applications.* John Wiley & Sons.

Villanueva, J., Yoo, S., & Hanssens, D. M. (February 2008). The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth. *Journal of Marketing Research*, stránky 48-59.

Wadsworth, E. (2012). *Buy 'Til You Die - A Walkthrough.* Retrieved from The R Project for Statistical Computing: http://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf

Wübben, M., & von Wangenheim, F. (2008). Instant Customer Base Analysis: Managerial Heuristics Often “Get It Right”. *Journal of Marketing*, pages 82-93.

Contact

Ing. Pavel Jašek
University of Economics, Prague
W. Churchill Sq. 4, Prague 3, Czech Republic
pavel.jasek@vse.cz

Ing. Lenka Vraná
University of Economics, Prague
W. Churchill Sq. 4, Prague 3, Czech Republic
lenka.vrana@vse.cz