

Multinomial Logistic Model for Long-Term Value

Bruce Lund, Magnify Analytic Solutions, a Division of Marketing Associates, LLC

ABSTRACT

Customer Long-Term Value (LTV) is a concept that is readily explained at a high level to marketing management of a Company, but its analytic development is complex. This complexity involves the need to forecast customer behavior well into the future. This behavior includes the timing, frequency, and profitability of a customer's future purchases of products and services. This paper describes a method for computing LTV. First, a multinomial logistic regression provides probabilities for time-of-first-purchase, time-of-second-purchase, etc. for each customer. Then the profits for the first purchase, second purchase, etc. are forecast but only after adjustment for non-purchaser selection bias. Finally, these component models are combined in the LTV formula.

INTRODUCTION

LTV (Long-Term Value) is an assignment of a monetary value to a customer. The assignment is based on projecting future profits to be earned by the Company from the customer by selling, financing, and servicing the company's products. Since profits are calculated into the future, these profits are time-adjusted by a discount factor.^{1 2}

Generally, an LTV formula is a discounted flow of profits which are computed over discrete time periods (months, years) taking this form:

$$LTV = \sum_{t=1}^T \text{Profit}_t / (1 + d)^t \dots (A)$$

where Profit_t is the profit to be earned by the Company from the customer at time period t , d is a discount factor, and T is a fixed but distant time horizon.³

In this paper the focus is on applications where Profit_t will be zero for many of the future time periods for typical customers. For example, the customers of an automotive company typically buy a new-vehicle only every few years and many of these customers defect to other brands.

INDUSTRIES WHERE LTV CAN BE IMPLEMENTED

Common to these industries are: (1) large customer or donor databases which are built and continuously updated with customer transactional data and possibly overlaid with demographics, (2) clear identification of historical revenues and profits earned from the customer.⁴

PROFIT_t IN EQUATION (A)

In developing a customer-level LTV model the forecast of Profit_t from equation (A) depends on customer characteristics determined at time $t = 0$. In fitting the model, the time $t = 0$ will be an historical date and the customer characteristics will be determined as of this date.

In order to estimate Profit_t there will be a target variable that gives the customer's actual profit-to-the-company for the t^{th} period after the historical date of $t = 0$.

Since this paper considers applications where the actual Profit_t at time t for many of the customers is zero, it is easier to estimate Profit_t conditionally on the customer being a buyer during period t .⁵

$$E(\text{Profit}_t | \text{Buyer}_t) \text{ Prob}(\text{Buyer}_t)$$

¹ Long-Term Value, instead of Life-Time Value, is used because the time horizon for computing the customer's value is a fixed number of future periods.

² Marketing costs unique to the customer are ignored (but purchase rebates would be recognized in the purchase profit). It is likely these marketing costs are small, similar across most customers, and not planned by the company beyond one or two time periods.

³ An equation where T varies by customer is given by W. Potts (2005).

⁴ E. Malthouse (2013) provides a current list of references.

⁵ Alternative approaches to estimating profit directly, perhaps including finite mixture modeling, are not considered in this paper.

But a generalization of this expected value may be appropriate. For example, the profit from customers who buy for the first time in period t may be different than the profit from customers who have already made purchases during some of the time-periods prior to period t.

For example, a college alumnus has probabilities for year t of being a first-time annual donor, second-time annual donor, etc. Given there is a donation in year t, the amount may depend on whether the donation is first-time, second-time, etc. The expected donation amount for year t can be expressed by:

$$\sum_{J=1}^t E(\text{Profit}_{J,t} | \text{Donor}_{J,t}) \text{Prob}(\text{Donor}_{J,t})$$

where J gives the donation number (first-time, second-time, etc.).

These considerations lead to a new LTV formula given in the section below.

NEW LTV FORMULA

Notation:

- If a customer makes purchases during time period t, then $\delta_t = 1$. Otherwise, $\delta_t = 0$.
- The event $E_{J,t}$ occurs when $\sum_{i=1}^{t-1} \delta_i = J-1$ and $\delta_t = 1$

The notation $E_{J,t}$ is used in defining the new LTV formula. First, some examples.

Table 1 Examples of Occurrence / Non Occurrence of $E_{2,3}$

Customers	Time period			$E_{2,3}$
	1	2	3	
Customer 001	purchases	no purchase	purchases	YES
Customer 002	purchases	purchases	purchases	NO
Customer 003	no purchase	purchases	purchases	YES
Customer 004	purchases	purchases	no purchase	NO

The new LTV formula (B) generalizes the specification of Profit_t that appears in equation (A).

$$\text{LTV} = \sum_{t=1}^T (\sum_{J=1}^t \pi_{J,t} p_{J,t}) / (1 + d)^t \dots \text{(B)}$$

where:

Varies by customer (throughout the paper the indexing by customer is suppressed):

- $p_{J,t}$ gives the probability of event $E_{J,t}$... the “purchase probabilities”
- $\pi_{J,t}$ estimates profit of purchases in period t when event $E_{J,t}$ occurs ... the “profit forecasts”

For all customers:

- T: a fixed time horizon measured in equal-length time-periods such as years.
- d: the discount factor which is constant over the time horizon and does not depend on the customer.⁶

Example: $p_{2,3}$ gives the probability that a customer made purchases in period 3 and also made purchases in either period 1 or period 2 but not in both periods.

A practical consideration involves the upper index in the summation $\sum_{J=1}^t \pi_{J,t} p_{J,t}$. Specifically, the occurrence rate in the sample of $E_{t,t}, E_{t-1,t}$, etc. may be too small to allow modeling of these events. The modeler will need to perform exploratory data analysis and determine appropriate cut-offs for the J-index.

SAS® code is given in Appendix B for a simple simulation which includes demonstrations of the modeling techniques for fitting $p_{J,t}$ and $\pi_{J,t}$.

TERMINOLOGY

To avoid long cumbersome verbal descriptions, the phrase “Jth-time buyer during period t”, or simply “Jth-time buyer” when “t” is understood, will be used for customers who have experienced event $E_{J,t}$.

⁶ The discount rate might be set at the rate of high-risk corporate bonds with a maturity that matches T, the time horizon for LTV.

VALIDATION OF THE LTV MODEL

The component terms in the LTV formula should be validated on holdout samples. In particular, for each $t = 1$ to T the quantities $\sum_{j=1}^t \pi_{j,t} p_{j,t}$ are computed for each customer and compared with actual profit from purchases by these customers. These comparisons are performed by creating Lift Charts using these steps:

1. Compute expected profit (EP) for each customer in the holdout sample: $EP = \sum_{j=1}^t \pi_{j,t} p_{j,t}$
2. Rank the customers by EP
3. Compute actual profits (AP) obtained during period “t” for each customer
4. Compute $Lift(rank) = 100 * AP(rank) / AP(ALL)$.
e.g. $Lift(rank 0) = 100 * \$1180 / \$430 = 274$

The lift chart in Table 2 uses deciles but, in practice, 20 or even 100 ranks can be used if sample size is large. The criteria for success are subjective and include: predictive accuracy (near equality of EP and AP by rank), monotonic consistency in lift, high differentiation in lifts between the top and bottom ranks.

Table 2 Hypothetical Example of Lift Chart for period t

Rank	Average		Lift
	EP	AP	
ALL	\$420	\$430	
0	\$1,190	\$1,180	274
1	\$670	\$740	172
2	\$530	\$550	128
3	\$440	\$450	105
4	\$360	\$360	84
5	\$300	\$320	74
6	\$250	\$250	58
7	\$200	\$190	44
8	\$170	\$140	33
9	\$130	\$100	23

Finally, LTV values on a holdout sample are computed by equation (B)

$$LTV = \sum_{t=1}^T (\sum_{j=1}^t \pi_{j,t} p_{j,t}) / (1 + d)^t$$

A lift chart similar to Table 2 is computed. Here, input from marketing management is needed to help in judging success. For example, marketing can answer the question of whether the average actual LTV for the highest ranked customers is high enough to justify the development of CRM programs for this customer segment.

A VARIATION ON THE NEW LTV FORMULA

A variation on Equation (B) is discussed in Appendix A.

PREPARATION OF SAMPLES FOR MODELING OF $\pi_{j,t}$ $p_{j,t}$

Multinomial logistic modeling will be used to estimate the $p_{j,t}$. Ordinary least squares (OLS) will be used to estimate $\pi_{j,t}$ after consideration of selection bias issues. Before a discussion of modeling begins, it is important to discuss the preparation of the samples for the models that produce $p_{j,t}$ and $\pi_{j,t}$.

SAMPLE FOR $t = 3$

To give a concrete example we assume that “now” is sometime after 1/1/2015 and that we have purchase information for each customer by year for 2012, 2013, and 2014. In addition we assume we have the ability to assemble a customer profile (transactional history, demographics, etc.) for each customer at historical points in time.

Now suppose a sample is drawn of customers that existed in the database on midnight of 12/31/2011. Transaction history and demographics are assembled to create a customer profile as of 12/31/2011 (denoted by X_1 - X_k).

Next, the values of a target variable “TARGET” for multinomial logistic modeling are constructed. These values are given in the bottom row of [Table 3](#).

Values of TARGET ending in “5” indicate the customer made no purchase in 2014 but purchased during at least one of the years 2012 or 2013. In this case the values of TARGET are 10 times the count of the years-with-a-purchase plus 5. (E.g. purchases in 2012 but none in 2013 or 2014 makes TARGET = 15)

Otherwise, the values of TARGET are 10 times the count of the years-with-a-purchase. (E.g. purchases in 2014 but none in 2012 or 2013 makes TARGET = 10)

The connection between this coding scheme and the purchase probabilities $p_{J,t}$ is given later in the section called “Purchase Probabilities”.

Table 3 Pattern of Years-with-a-Purchase and Assignment of Values for TARGET for t=3
In Rows 2012, 2013, 2014 a “0” indicates no purchase and “1” indicates purchases.

2012	0	0	0	0	1	1	1	1
2013	0	0	1	1	0	0	1	1
2014	0	1	0	1	0	1	0	1
TARGET=	0	10	15	20	15	20	25	30

When TARGET = 10, then event $E_{1,3}$ has occurred. Likewise, TARGET = 20 corresponds to occurrence of $E_{2,3}$ and TARGET = 30 corresponds to occurrence of $E_{3,3}$.

SAMPLES FOR t = 1 AND t = 2

But before starting the discussion of modeling, it is helpful to compare the foregoing with the sampling for t = 1 and t = 2. The values of TARGET are displayed in [Table 4](#) and [Table 5](#). The rules for assigning values to TARGET follow the same pattern as for case of t = 3.

Table 4 Assignment of Values for TARGET for t=1

2014	0	1
TARGET=	0	10

Table 5 Assignment of Values for TARGET for t=2

2013	0	0	1	1
2014	0	1	0	1
TARGET=	0	10	15	20

DISCUSSION OF SAMPLING

The only requirement in creating these three samples is to have sufficient size for modeling. There is no requirement for constant sampling fractions or any other connection between the samples.⁷ The predictor variables for modeling do not need to be common across the three samples.

For each “t” the most recent purchase activity is utilized in fitting the multinomial logistic model and in all cases the year 2014 (the most recent year) is included in defining TARGET.⁸

⁷ To make the samples fully independent of one another a random sequence number could be applied to all customers. Then customers whose sequence was in range “1” would be taken for sample “1”, range “2” for sample “2”, etc.

⁸ This assumes that 2014 is believed to be a good indicator of the behavior of customers in future years.

PURCHASE PROBABILITIES

FINDING THE PROBABILITIES $p_{j,t}$

Consider the “Sample for $t = 3$ ” discussed earlier. The 3 periods are the years 2012, 2013, 2014.

The logistic model below outputs 6 probabilities for each customer to the data set named “Scored”. These probabilities are named by SAS as IP_0, IP_10, IP_15, IP_20, IP_25, IP_30. Here, IP_x is the probability that TARGET = x.

```
PROC LOGISTIC DATA = Work;
  MODEL Target(ref="0") = X1 - XK / LINK = glogit;
  OUTPUT OUT = Scored PREDPROB = I;
```

Three of these, IP_10, IP_20, IP_30, provide the required purchase probabilities $p_{j,t}$ for $J = 1$ to 3 for time period 3 of the LTV equation (B). The other three probabilities are not directly used in the LTV equation but their associated target values are required for the estimation of $p_{j,t}$.

CHOICES FOR THE MULTINOMIAL LOGISTIC MODEL

Values of TARGET can be meaningfully ordered by Count and Time. If the company can only sell once to the customer, it is better to sell earlier, therefore $10 < 15$. The ordering is: 0, 10, 15, 20, 25, 30.

With an ordering of the response variable there are, at least, three choices for the multinomial modeling approach: Proportional Odds (PO), Partial Proportional Odds (PPO), and the General (unordered) Logistic.⁹ This paper will focus on the general multinomial logistic model.

FOR GENERAL LOGISTIC MODEL THE $p_{j,t}$ ARE CORRECT, ON AVERAGE

The general logistic model has a desirable feature which is not entirely shared by the ordered logistic models. Specifically, means of probabilities for the target responses across the observations equals the empirical distribution of these responses. For example, the mean of IP_10 will equal the percentage of TARGET = 10 in the distribution for TARGET across its six values. See the Simulation in Appendix B for an example.

For ordered logistic models there is not a guarantee of equality. The advantage of the ordered logistic models is, importantly, far fewer parameters for estimation.

CONSOLIDATION AND EXPANSION OF TARGET RESPONSES

A TEST statement could be added to PROC LOGISTIC to test the consolidation of target responses. For example, the following statement conducts an hypothesis test of whether all coefficients for response “15” are equal to the corresponding coefficients for response “25”.

```
PROC LOGISTIC DATA = WORK;
  MODEL Target(ref="0") = X1 X2 X3 / LINK = glogit;
  TEST X1_15 = X1_25, X2_15 = X2_25, X3_15 = X3_25;
  OUTPUT OUT = Scored PREDPROB = I;
```

Consolidation of responses “15” and “25” into, say, response “90” would reduce the complexity of the logistic model but would adversely affect the fitting of the model since response “15” has a very different meaning than does response “25” (one purchase versus two). This consolidation also violates the ordering of target values and removes the opportunity of using an ordered logistic model.

A response such as “15” could be expanded into, say, “12” and “18” to distinguish the patterns of purchases of 1, 0, 0 and 0, 1, 0. The wide-spread expanding of such responses would significantly increase the number of target responses for large T. Yet the impact on LTV model validation (as illustrated by Table 2) would be small.

⁹ See Derr (2013) for a survey of PO and PPO.

Consolidation or expansion of target responses does not affect the “correct on average” property of the $p_{j,t}$ of the general multinomial logistic model. But the response probabilities for particular customers or segments of customers would be changed.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES AND MULTINOMIAL LOGISTIC REGRESSION

“Target responses” are also called “alternatives”. For multinomial logistic regression the log-odds of alternatives “j” and “k” for the i^{th} customer is given by

$$\log(P_{ij} / P_{ik}) = (\beta_j - \beta_k) \circ x_i \dots \text{ (IIA)}$$

The condition referenced by equation IIA is called the “Independence of Irrelevant Alternatives”.

In IIA the log-odds of alternatives “j” and “k” for the i^{th} customer involve the coefficients for alternatives “j” and “k” as well as the customer predictor values of x_i but involve no other alternatives. This restrictive condition may not be appropriate for some models.

Tests of the suitability of multinomial logistic regression, including violations of IIA, are performed by the Hausman’s Specification Test and the Small and Hsiao Likelihood Ratio Test. SAS implementations of these tests are discussed in SAS documentation of PROC MDC.

But findings reported by Cheng and Long (2006) show these tests to be unreliable for large-scale applications. A related short discussion of testing of the suitability of multinomial logistic regression is given by Paul Allison (2012b).

Alternatives to the multinomial logistic model for LTV include more generalized discrete choice models such as the heteroscedastic extreme value model and nested-logit model. Neither of these models is subject to the IIA condition. These and other discrete choice models can be fitted by PROC MDC.¹⁰

I have used the multinomial logistic model because the model is comparatively simple to fit and good validations can be obtained.

THE NUMBER OF LEVELS OF THE TARGET

The number of values for TARGET is given by the formula: NUMBER = 2 * T. For example, if T is 10, then NUMBER = 20. This is a fairly large number of response levels for PROC LOGISTIC but is feasible for the approach of this paper when the model is built on a very large sample.

FOR FIXED J THE $p_{j,t}$ ARE THE DENSITY FUNCTIONS FOR THE TIME TO EVENT J

For fixed J (becoming a J^{th} -time buyer) the $p_{j,t}$ are approximations to the values of the density function of the time to event J. By standard formulas the $p_{j,t}$ are related to the hazard functions $h_J(t)$.¹¹

LOGISTIC MODEL VALIDATION

One test to evaluate the fit of the logistic model is to measure the ability of a probability like IP_10 to “find” the response Target = 10. The SAS code below creates a “lift chart” for IP_10. Good evidence for validation of the model is given by finding a large column-percentage for Target = 10 in the top decile of rankings of IP_10 values. This test should be run on a holdout sample.

```
PROC RANK DATA = Scored OUT = rank_Scored GROUPS = 10 DESCENDING;
  VAR IP_10; RANKS rank_IP_10;
PROC FREQ DATA = rank_Scored; TABLES rank_IP_10 * Target
/ NOFREQ NOROW NOPERCENT;
```

This code is included in the simulation in Appendix B.

¹⁰ Appendix C includes SAS code for running PROC MDC using TYPE = HEV (heteroscedastic extreme value) and TYPE = NLOGIT (nested logit) on the data set of the simulation of Appendix B. The structure of these models is the same as the multinomial logistic regression model except for the choice of HEV or NLOGIT. See Appendix C for discussion.

¹¹ $h_J(t_0) = p_{j,t_0}$ if $t_0 = 1$. $h_J(t_0) = p_{j,t_0} / (1 - \sum_{t=1}^{t_0-1} p_{j,t})$ if $t_0 > 1$

PROFIT FORECAST

EXAMPLE: ESTIMATING $\pi_{J,3}$

We continue to use the “**Sample for t = 3**” discussed earlier. Here, the 3 periods are the years 2012, 2013, 2014. Now 3 models are needed to estimate profit for t = 3:

- Model 1 for 1st-time buyers during period 3
- Model 2 for 2nd-time buyers during period 3
- Model 3 for 3rd-time buyers during period 3

Considering, for example, 2nd-time buyer during period 3, an ordinary least squares (OLS) model could be fit to this population of customers with their profit from period 3 (2014) as the target variable. But such a model might be affected by a selection bias. The bias arises when the model is applied to the entire population. The suspicion that the model would be biased comes by analogy to the classic case of selection bias that arises when estimating profits from buyers by OLS regression and then applying this model to buyers and non-buyers.¹²

CORRECTION FOR SELECTION BIAS FOR THE CASE OF BUYERS V. NON-BUYERS

The buyer v. non-buyer case of selection bias is addressed by the Heckman correction.^{13 14} The Heckman correction involves the following two steps:

1. A probit model is run of buyers v. non-buyers. For each customer the “xbeta” from the probit model output is saved in an output data set.
 - The Inverse Mills Ratio is defined as $M(x) = \varphi(x) / \Phi(x)$ where φ is the standard normal density and Φ is the standard normal cumulative distribution. For each customer, M is evaluated at xbeta.
2. An OLS regression is run on the subset of buyers with dependent variable PROFIT and with M being among the predictors. A good choice for a regression procedure is PROC GLMSELECT.¹⁵

Comments:

- PROC QLIM (SAS/ETS) can perform the Heckman two-step method in a single step.¹⁶
- The estimates by PROC GLMSELECT of the variances of the predictor coefficients will be incorrect in this setting.¹⁷ This problem does not adversely affect the STEPWISE variable selection by PROC GLMSELECT since the default selection of variables is by the Schwartz-Bayes criterion (SBC) and not the F-statistic of the predictors. Additional variable selection options are provided by GLMSELECT.
- In general, it is likely that some of predictors in the PROFIT regression model will also have appeared in the probit model. This leads to multicollinearity between M and these predictors.
- There is no assurance that M will be selected for the PROFIT regression model.

A PROPOSAL FOR ESTIMATING $\pi_{2,3}$ WITH CORRECTION FOR SELECTION BIAS

We continue with the population of 2nd-time buyers during period 3. A proposal is given for estimating $\pi_{2,3}$ while also addressing selection bias. It is a heuristic adaptation of the Heckman correction but without supporting theory. In a direct marketing application, such as LTV modeling, a measure of success of this proposal would be whether a strong validation of predictions can be attained on a holdout sample. There are two steps to the proposal.

Step 1: Pseudo Inverse Mill Ratios are created.

Two predictors, M_20L and M_20M, are introduced to take the place of the inverse Mills ratio of the “buyer v. non-buyer” classic case. The code to create M_20L and M_20M is given next.

¹² For a discussion of selection bias and related topics, see Breen (1996)

¹³ Heckman (1979)

¹⁴ Lee, et al. (2013, p. 5-6) briefly discuss selection bias for the buyer v. non-buyer case and present the Heckman correction.

¹⁵ See Cohen (2014) for a survey article about PROC GLMSELECT.

¹⁶ See discussion at http://support.sas.com/documentation/cdl/en/etsug/65545/HTML/default/viewer.htm#etsug_qlim_details17.htm

¹⁷ Ibid.


```
%LET P_20L = IP_0 + IP_10 + IP_15 + IP_20;
%LET P_20M = IP_20 + IP_25 + IP_30;
```

First, the conditional probabilities $IP_{20} / (&P_{20L})$ and $IP_{20} / (&P_{20M})$ are created. Next, the PROBIT function takes these conditional probabilities as arguments and returns the corresponding standard normal “z” values. Lastly, these “z” values are placed in the formula for the inverse Mills ratio to create M_20L and M_20M.

```
DATA Temp; SET Scored;
  Z_20L = PROBIT(IP_20 / (&P_20L));
  M_20L = PDF('NORMAL', Z_20L) / CDF('NORMAL', Z_20L);
  Z_20M = PROBIT(IP_20 / (&P_20M));
  M_20M = PDF('NORMAL', Z_20M) / CDF('NORMAL', Z_20M);
```

By employing the ordering of TARGET, the conditional probability $IP_{20} / (&P_{20L})$ is designed to mimic the probit model probability of the Heckman correction.

- $IP_{20} / (&P_{20L})$ is the probability of 20, given 20 or less
- $1 - IP_{20} / (&P_{20L})$ is the probability of not 20, given 20 or less.

These two probabilities are analogous to the “probability of buyer” v. the “probability of non-buyer” of the classic case. A similar discussion applies to $IP_{20} / (&P_{20M})$

Step 2: OLS regression is run with M_20L and M_20M among the predictors.

```
PROC GLMSELECT DATA = Temp;
  MODEL Profit = M_20L M_20M W1 - WL
  / SELECTION = STEPWISE(SELECT= SBC);
  WHERE Target in (20);
```

Comments:

- The estimates by PROC GLMSELECT of the variances of the predictor coefficients are probably incorrect (as was true for the buyer v. non-buyer case). This problem does not adversely affect the STEPWISE variable selection by PROC GLMSELECT since the default selection of variables is by the Schwartz-Bayes criterion (SBC) and not the F-statistic of the predictors.
- It is likely that the predictors M_20L and M_20M will be moderately to highly correlated.
- If some of the predictors $W_1 - W_L$ in the PROFIT regression model also appeared in the multinomial logistic model, it is likely there is multicollinearity between M_20L, M_20M and these predictors.
- There is no assurance that M_20L or M_20M will be selected for the PROFIT regression model.
- For 3rd-time buyers during period 3 the corresponding M_30M cannot be created because the conditional probability IP_{30} / P_{30M} is identically one (since $P_{30M} = IP_{30}$).
- If $t = 1$, the probabilities IP_{15} and IP_{25} do not exist and the formulas for pseudo inverse Mills ratio require adjustment.

HECKMAN CORRECTION EXTENSIONS

There are new methods for extending the Heckman correction to the case of multiple selection biases. A description of these methods is given by Bourguignon, Fournier, and Gurgand (2004).¹⁸ Additional work is required to determine if or how to apply these methods to the PROFIT equation.

MASS PRODUCTION OF MODELS

A seemingly formidable number of models are required to assemble the components of the LTV formula. A total of T models are required to produce $\pi_{j,t}$. Then another $T*(T+1)/2$ models are required to produce $\pi_{j,t}$. For T = 10 there would be 65 models. Mass-production of models is needed.

¹⁸ The authors discuss advanced methods which extend the Heckman correction to the case of multiple selection biases when fitting a regression equation. As a by-product of their investigations the authors developed a Stata® command that performs these methods. I am not aware of SAS code which performs these methods.

Step 1: One or two values of period “t” are selected and an exhaustive effort is conducted to fit the multinomial model to TARGET and to fit the OLS regression models to PROFIT. The goal is to prove-out predictor variables and achieve successful lifts on a holdout sample. Now the hope is these predictors will continue to perform well for other values of “t”.

Step 2: The component models for LTV are mass produced by these steps:

- First, a **master data set** is assembled of data that are needed to determine eligibility of a customer for inclusion in modeling for a given “t” and for the computation of the values of predictor and target variables. This master data set would be unique by customer and would be denormalized to include arrays of transactional data which are required to code the predictor and target values.
- A macro program with a macro parameter called “profile_date” is needed. Each profile date corresponds to a value of “t”. This macro will be run T times, for each of the $t = 1$ to T.
 - First, a DATA step is run to compute the values of the predictors in relationship to the “profile date”. For example, a predictor variable might be “the number of purchases in the 5 years prior to the profile date”. Using processing of an ARRAY the customer’s prior purchase-dates would be compared to the “profile date” to determine the value of this predictor.
 - A PROC LOGISTIC step follows. It runs a multinomial logistic to compute the $p_{J,t}$ on the training subset of the master data set. The scored output is used in the next DATA step to compute the pseudo inverse Mills ratios.
 - Finally, there are “t” runs of PROC GLMSELECT on the training dataset to compute $\pi_{J,t}$.
- LTV is computed for each customer using the $p_{J,t}, \pi_{J,t}$, and d (discount factor).

The component models for LTV are validated on holdout samples as part of the macro program. In particular, for each $t = 1$ to T the quantities $\sum_{J=1}^J \pi_{J,t} p_{J,t}$ are computed for each customer in the holdout sample, ranked, and compared with actual profit from purchases by these customers.

A full validation of the LTV score poses problems. It would be impractical to compute LTV scores for customers as of “today” and track customer purchases over the next T periods for comparison. Instead, one solution is to profile the characteristics of customers by their LTV rank and to judge the reasonableness of these profiles.

CONCLUSION

The LTV Equation (B) provides generality in the computation of customer value by incorporating frequency and recency into the targets which are predicted by the component models. But the development of the numerous component models may seem daunting. By using a pilot to develop a good model for a specific time period, the proven-out predictor variables from the pilot can then be used in a mass production of the component LTV models.

The sampling to create the T analytic data sets is straightforward with only the requirement to make the samples large enough for modeling. The profile dates for the observations in an analytic data set are all the same. This has the benefit of not mixing together customer characteristics from unlike market conditions. Target values are defined using the most recent data available.

A SIMULATION

SAS code for simulating the steps in fitting the component models of LTV is given in Appendix B.

A DATA STEP is run which creates data set “work”. “Work” is the analytic dataset for $t = 3$ with $J = 1$ to 3. Included in “Work” are these variables:

- TARGET (with values 0 10 15 20 25 30)
- PROFIT
- Predictors: X1, X2, X3, and V

PROC LOGISTIC modeling provides $p_{J,t}$
PROC GLMSELECT modeling estimates $\pi_{J,t}$

REFERENCES

- Allison, P. (2012a). *Logistic Regression Using SAS*, SAS Institute Inc.
- Allison, P. (2012b). "How Relevant is the Independence of Irrelevant Alternatives?", Oct 12, 2012, Statistical Horizons. Available at: <http://www.statisticalhorizons.com/iaa>.
- Bhat, C. (2000) "Chapter 5: Flexible Model Structures for Discrete Choice Analysis", in: D. Hensher and K. Button, editors, *Handbook for Transport Modelling, 2nd edition (2007)*, Emerald Group Publishing. Available at: http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Flexible_discrete_chapter.pdf
- Bourguignon, F., Fournier, M., Gurgand, M. (2004). "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons", National Center for Scientific Research (CNRS); National Institute of Statistics and Economic Studies (INSEE) - Center for Research in Economics and Statistics (CREST). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=555744
- Breen, R. (1996). *Regression Models, Censored, Sample Selected, or Truncated Data*, Sage Publications, Inc.
- Cheng, S. and Long, J. S. (2006). "Testing for IIA in the Multinomial Logit Model", *Sociological Methods & Research*: 35: 583-600.
- Cohen, R. (2006). "Introducing the GLMSELECT PROCEDURE for Model Selection", *Proceedings of the 31st Annual SAS Users Group International Conference SUGI 31, 2006*, Paper 207-31.
- Derr, B. (2013). "Ordinal Response Modeling with the LOGISTIC Procedure", *Proceedings of the SAS Global Forum 2013 Conference*, Paper 446-2013.
- Heckman, J. (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.
- Koppelman, F. and Bhat, C.(2006). A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models, prepared for U.S. Department of Transportation, Federal Transit Administration. Available at: http://www.caee.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf
- Lee T., Zhang R., Meng X., and Ryan L. (2013), "Incremental Response Modeling Using SAS® Enterprise Miner", *Proceedings of the SAS Global Forum 2013 Conference*, Paper 096-2013.
- Malthouse, E. (2013), *Segmentation And Lifetime Value Models Using SAS*, SAS Institute Inc.
- Potts, W. (2005). "Predicting Customer Value", *Proceedings of the 30th Annual SAS Users Group International Conference SUGI 30, 2005*, Paper 073-30.

ACKNOWLEDGMENTS

Dave Brotherton of Magnify Analytics Solutions read drafts of this paper and gave helpful suggestions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bruce Lund
Marketing Associates, LLC
777 Woodward Ave, Suite 500
Detroit, MI, 48226
blund@marketingassociates.com or blund_data@mi.rr.com

All code in this paper is provided by Marketing Associates, LLC. "as is" without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Recipients acknowledge and agree that Marketing Associates shall not be liable for any damages whatsoever arising out of their use of this material. In addition, Marketing Associates will provide no support for the materials contained herein.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

APPENDIX A: VARIATION ON EQUATION (B)

The variation of the LTV formula is shown in equation (C). The meanings of some of the symbols are modified.

$$LTV = \sum_{t=1}^T (\sum_{j=1}^{J_{max}} \pi_{j,t} p_{j,t}) / (1 + d)^t \dots (C)$$

The phase “Jth purchase” now refers to the count of distinct purchases.

- $p_{j,t}$ is the probability that the Jth purchase occurs in period t
- $\pi_{j,t}$ is the estimated profit for purchases at t assuming Jth-purchase occurs at time t
- J_{max} is the maximum number of purchases to be considered (J_{max} can be greater than t). The modeler may choose to have J_{max} depend on t.

For example: If a customer purchases twice in period 1, zero times in period 2, and twice in period 3, then the value of J in period 3 equals 4 and the 4th purchase occurred in period 3.

CHANGES TO TARGET

One significant change is to the definition of TARGET. Considering the case of $T = 3$ and $J_{max} = 3$, the coding scheme for TARGET is shown in Table 6.

Table 6 Assignment of Values for TARGET

t = 1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	3
t = 2	0	0	0	0	1	1	1	2	2	3	0	0	0	1	1	2	0	0	1	0
t = 3	0	1	2	3	0	1	2	0	1	0	0	1	2	0	1	0	0	1	0	0
TARGET	0	10	20	30	15	20	30	25	30	35	15	20	30	25	30	35	25	30	35	35

Additionally, when developing the pseudo inverse Mills ratios for the case where $J = 2$, the probability for P_{20M} is changed by adding IP_{35} :

`%LET P_20L = IP_0 + IP_10 + IP_15 + IP_20;`

`%LET P_20M = IP_20 + IP_25 + IP_30 + IP_35;`

Similarly IP_{35} is added to P_{10M} and P_{30M} when $J = 1$ and $J = 3$.

Finally, it should be emphasized that the profits for all purchases in period t are used in defining the PROFIT target variable when estimating $\pi_{j,t}$. For example: If a customer purchases once in period 1, zero times in period 2, and twice in period 3, then the value of J in period 3 equals 3. The sum of the profits from the two purchases made during period 3 is used to define the value of PROFIT for modeling for this customer.

APPENDIX B: SIMULATION PROGRAM

A simulation generates a data set called “work” which is used in logistic modeling. Here is the code which runs PROC LOGISTIC to estimate $p_{j,t}$ and PROC GLMSELECT to estimate $\pi_{j,t}$ after consideration of selection bias issues:

```

DATA WORK;
do i = 1 to 25000;
/* Whether or not the customer makes a purchase in years 2012, 2013, 2014 */
r1 = ranuni(12345);
r2 = ranuni(12345);
r3 = ranuni(12345);
C2012 = (r1 > .70);
C2013 = (r2 > .75 & r1 < .92);
C2014 = (r3 > .80 & r1 < .95 & r2 < .92);
/* Defining TARGET only for t = 3 (2014) */
COUNT = sum(of C2012 - C2014);
If COUNT > 0 and C2014 = 0 then Target = 10*COUNT + 5;

```

```

    else Target = 10*COUNT;
/* x1 - x3 (and V below) characterize the customer and
predict the TARGET and PROFIT variable values */
/* Predictors */
X1 = r1 + 2*(ranuni(12345) - 0.5);
X2 = r2 + (ranuni(12345) - 0.5);
X3 = r1 + r2 + r3 + (ranuni(12345) - 0.5);
/* A predictor V is a measure of historic value */
V = 4*rannor(12345) + 40;
/* PROFIT in t=3 (2014). Depends on predictor V and r1 - r3 */
Profit = (C2014 > 0) * (V/2 + 2*r1 + 3*r2 + 4*r3 + 5*rannor(12345));
output;
end; /* END: do i = 1 to 25000 */
run;

/* FINDING THE PROBABILITY DENSITIES pJ,3 */
PROC LOGISTIC DATA = WORK;
MODEL Target(ref="0") = X1 X2 X3 / LINK = glogit;
/* TEST OF THE EQUALITY OF THE COEFFICIENTS FOR TARGET =15 AND TARGET =25 */
TEST X1_15 = X1_25, X2_15 = X2_25, X3_15 = X3_25;
/* INDIVIDUAL TARGET PROBABILITIES */
OUTPUT OUT = Scored PREDPROB = I;
TITLE1 "General Logistic Model";
run;

/* MEAN VALUES OF IP_10, IP_20, IP_30 EQUAL THE PERCENTS FROM PROC FREQ */
PROC MEANS DATA = Scored MEAN; VAR IP_10 IP_20 IP_30;
TITLE1 "Mean Values of IP_10 IP_20 IP_30";
PROC FREQ DATA = Scored; TABLES Target / LIST;
TITLE1 "Distribution of Values of Target";
run;

/* MEASURING THE MODEL FIT FOR IP_10 */
PROC RANK DATA = Scored OUT = rank_Scored GROUPS = 10 DESCENDING;
VAR IP_10; RANKS rank_IP_10;
PROC FREQ DATA = rank_Scored; TABLES rank_IP_10 * Target
/ NOFREQ NOROW NOPERCENT;
TITLE1 "Column Percents of Target Outcomes vs. Decile Ranks of IP_10";
run;

%LET P_10L = IP_0 + IP_10;
%LET P_10M = IP_10 + IP_15 + IP_20 + IP_25 + IP_30;

/* PROFIT EQUATION FOR FIRST-PERIOD BUYERS */
/* ADJUSTMENT FOR SELECTION BIAS FOR PROFIT FOR TARGET = 10 */
/* CREATING HECKMAN-LIKE CORRECTION USING LOGISTIC PROBABILITIES */
DATA Scored; SET Scored;
Z_10L = PROBIT(IP_10 /(&P_10L));
M_10L = PDF('NORMAL',Z_10L) / CDF('NORMAL',Z_10L);
Z_10M = PROBIT(IP_10 /(&P_10M));
M_10M = PDF('NORMAL',Z_10M) / CDF('NORMAL',Z_10M);
run;
PROC GLMSELECT DATA = Scored;
MODEL Profit= V M_10L M_10M / SELECTION= STEPWISE(SELECT=SBC);
WHERE Target in (10);
TITLE1 "Regression using Less and More pseudo inverse Mills ratio";
run;

```

```

/* CORRELATIONS BETWEEN PSEUDO INVERSE MILLS RATIOS */
PROC CORR DATA = Scored; VAR M_10L M_10M;
  WHERE Target in (10);
TITLE1 "Correlation between pseudo inverse Mills ratios";
run;

```

APPENDIX C: ALTERNATIVES TO MULTINOMIAL LOGISTIC REGRESSION

As discussed by Allison (2012a p. 202) the multinomial logistic model is a special case of the “conditional logit model”. Both PROC LOGISTIC and PROC MDC can fit conditional logit models.

In Appendix B the simulation generates a data set called “work” which is used in logistic modeling. The DATA step below transforms work into work2. The structure of work2 is designed so that PROC MDC can fit models that are directly comparable to the multinomial logistic model of Appendix B.

```

data work2; set work;
retain v1 - v6;
if _n_ = 1 then do;
  v1 = 0; v2 = 10; v3 = 15; v4 = 20; v5 = 25; v6 = 30;
end;
array targetx {6} target00 target10 target15 target20 target25 target30;
array c_value {6} v1 - v6;
array x1_ {6} x1_1 - x1_6;
array x2_ {6} x2_1 - x2_6;
array x3_ {6} x3_1 - x3_6;
do j = 1 to 6;
  targetx{j} = 0;
  x1_{j} = 0;
  x2_{j} = 0;
  x3_{j} = 0;
end;
do j = 1 to 6;
  choice = (target = c_value{j});
  targetx{j} = 1;
  x1_{j} = x1;
  x2_{j} = x2;
  x3_{j} = x3;
  output;
  targetx{j} = 0;
  x1_{j} = 0;
  x2_{j} = 0;
  x3_{j} = 0;
end;
run;

```

C_Logit: The SAS code below, running PROC MDC with TYPE = CLOGIT, will almost exactly reproduce the modeling results from the multinomial logistic model of Appendix B.

```

PROC MDC DATA = work2;
MODEL choice =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6 /
CHOICE = (j)
TYPE = CLOGIT; /* Conditional Logit */
ID i;

```

HEV: The heteroscedastic extreme value (HEV) logistic model generalizes the conditional logit model. In HEV the variances of the random terms associated with the utility functions of the alternatives are allowed to be unequal. As a result IIA is not forced by the HEV model. The difference between HEV logit and the conditional logit model is discussed by Allison (2012a, p. 210). A full explanation of the HEV model is given by Bhat (2000, section 2).

SAS code for the HEV version of the multinomial logistic model of Appendix B is shown below:

```
PROC MDC DATA = work2;
MODEL choice =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6 /
CHOICE = (j)
TYPE = HEV HEV=(UNITSCALE=1, INTEGRATE=LAGUERRE);
ID i;
```

The nested logistic (NLOGIT) model also generalizes the conditional logit model. In NLOGIT the alternatives are grouped by the modeler. Within these groups the variances of the random terms are correlated but are uncorrelated between groups. As a result IIA is not forced by NLOGIT. The difference between NLOGIT and the conditional logit model is discussed by Allison (2012a, p. 214). A full explanation of the nested logit model is given by Koppelman and Bhat (2006).

The decision of how to group the alternatives is guided by subject matter knowledge of the modeler and by experimentation. To systematically test all groupings of a choice set with 6 alternatives, with a limit of two levels of nesting, requires 31 runs. This complexity will limit the ability to run experiments and will emphasize subject-matter-based judgments.

Nested 1: SAS code for running the nested logit with TARGET = 0 in one group and all other values of TARGET in a second group is shown below. Specifically, this formulation assumes no correlation between pure non-buyers and the second group where all have at least one purchase.

```
PROC MDC DATA = work2;
MODEL choice =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6 /
CHOICE = (j)
TYPE = NLOGIT;
UTILITY U(1, ) =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6
;
NEST LEVEL(1) = (1 @1, 2 3 4 5 6@2),
LEVEL(2) = (1 2 @1);
ID i;
```

Nested 2: SAS code is given for running the nested logit with TARGET = 0 in one group, TARGET in (2, 3) in another group, and TARGET in (4, 5, 6) in the final group. This nested logit converged and to a solution that was very different from the conditional logit.

```
PROC MDC DATA = work2;
MODEL choice =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
```

```

x3_2 x3_3 x3_4 x3_5 x3_6 /
CHOICE = (j)
TYPE = NLOGIT;
UTILITY U(1,) =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6
;
NEST LEVEL(1) = (1 @1, 2 3 @2, 4 5 6 @3),
LEVEL(2) = (1 2 3 @1);
ID i;

```

Nested 2 with starting values from C_Logit: The same SAS code was used for running “Nested 2” but with the starting values taken from the conditional logit solution and, in addition, starting the variance parameters at 1, 0.5, 0.5. This nested logit converged to a solution that is somewhat similar to the “C_Logit” solution

```

PROC MDC DATA = work2;
MODEL choice =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6 /
CHOICE = (j)
TYPE = NLOGIT
START = (-4.2947 -4.2141 -9.8179 -11.3469 -18.0641 -0.2486 0.6476 0.4253
1.0081 0.7281 -0.7885 1.0478 -0.1035 2.509 1.6381 2.398 2.2817 4.6991 4.212
6.7102 1 .5 .5);
UTILITY U(1,) =
target10 target15 target20 target25 target30
x1_2 x1_3 x1_4 x1_5 x1_6
x2_2 x2_3 x2_4 x2_5 x2_6
x3_2 x3_3 x3_4 x3_5 x3_6
;
NEST LEVEL(1) = (1 @1, 2 3 @2, 4 5 6 @3),
LEVEL(2) = (1 2 3 @1);
ID i;

```

Comments and Conclusions:

- The conditional logit (equal to the multinomial logistic model) provided a fit at least as good as any of the other models as measured by AIC and the lift charts for the probability IP_10 (probability of TARGET = 10). See [Table 7](#) and [Table 8](#).
- The run of HEV produced the solution in [Table 7](#) but the optimization algorithm failed to converge. Convergence may be more likely by using the solution found via TYPE=CLOGIT as starting values. Also, experimentation with HEV options INTORDER and INTEGRATE=HARDY may help. Setting UNITSCALE=1 2 3 4 5 6 will approximately reproduce the “C_Logit” solution so experiments with UNITSCALE settings may be useful.
- As shown in [Table 7](#) for “Nested 2” and “Nested 2 start = C_Logit” the solution of the nested logit model may be significantly dependent on the starting values.
- Model fitting for an LTV model would be more complicated if using PROC MDC with TYPE= HEV or NLOGIT in comparison with the general logistic regression by PROC LOGISTIC.
 - Run times will be long for models on large samples with many variables and target values.
 - Convergence failures may be a problem if using HEV.
 - In the case of NLOGIT the testing of nesting configurations adds complexity and NLOGIT may

- have significantly distinct solutions dependent of the starting values.
- MDC does not provide a variable selection tool such as stepwise selection.

Table 7 Coefficient Estimates for Five Models applied to the Data and Variables of Appendix B
All Models Converged Except HEV

	C_Logit	Hev	Nested 1	Nested 2	Nested 2 start=C_Logit
Parameter	Coefficients				
target10	-4.295	-9.740	-3.758	-34.480	-4.026
target15	-4.214	-3.897	-3.652	-34.428	-3.936
target20	-9.818	-5.684	-9.254	-121.861	-20.625
target25	-11.347	-5.873	-10.782	-123.827	-22.421
target30	-18.064	-9.826	-17.506	-131.326	-29.557
x1_2	-0.249	-1.435	-0.300	3.190	-0.272
x1_3	0.648	0.613	0.581	4.130	0.612
x1_4	0.425	0.503	0.361	7.528	1.167
x1_5	1.008	0.617	0.937	8.155	1.796
x1_6	0.728	0.621	0.659	7.887	1.515
x2_2	-0.789	-3.504	-0.875	4.028	-0.821
x2_3	1.048	0.960	0.960	5.797	0.994
x2_4	-0.104	0.593	-0.200	10.012	0.977
x2_5	2.509	1.100	2.425	12.668	3.690
x2_6	1.638	1.045	1.547	11.908	2.852
x3_2	2.398	4.126	2.039	21.817	2.220
x3_3	2.282	2.171	1.912	21.727	2.108
x3_4	4.699	3.060	4.330	57.658	9.765
x3_5	4.212	2.961	3.841	57.344	9.352
x3_6	6.710	4.157	6.343	60.131	12.018
INC_L2G1C1			1	1	1
INC_L2G1C2			1.172	0.106	1.082
INC_L2G1C3				0.080	0.472
Log Likelihood	-26077	-26123	-26076	-26084	-26073
AIC	52194	52296	52197	52214	52191
run-time ¹⁹	4 seconds	48 min	4 min	11 min	4 min

Table 8 Lift Charts for Target=10 v. Deciles of IP_10

Deciles for IP_10	Column Percentage of Target = 10 by Decile				
	C_Logit	Hev	Nested 1	Nested 2	Nested 2 start=C_Logit
0	24.74	24.02	24.63	24.82	24.67
1	16.80	16.72	17.07	16.42	16.91
2	13.21	11.72	13.29	13.78	13.25
3	10.92	11.30	10.58	10.61	10.77
4	9.66	9.35	9.55	9.62	9.70
5	8.71	8.17	8.59	8.63	8.74
6	6.53	6.91	6.99	7.03	6.45
7	5.12	6.34	5.00	4.89	5.19
8	3.36	4.24	3.40	3.25	3.40
9	0.95	1.22	0.92	0.95	0.92

¹⁹ SAS 9.4 running on windows 7 PC with Intel® Core™ i7-3540M CPU @ 3.00 GHz, RAM 16 GB, 64-bit