Customer Lifetime Value and Churn Management

Data Science for Marketing Decision Making
Günter J. Hitsch
Chicago Booth

2017

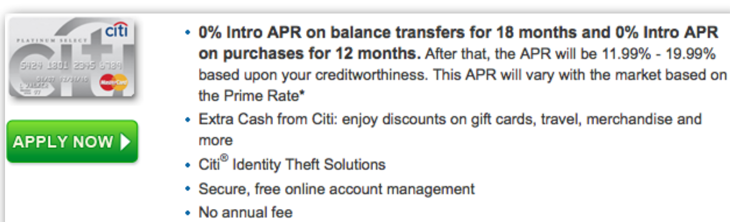
Overview

1. Customer lifetime value
2. Retaining customers — churn management

Customer lifetime value

Customer lifetime value (lifetime value of the customer)

- Citibank rents a list of names and addresses for a direct mail campaign to *acquire* new credit card customers and offers a 0% intro APR
- Citibank lowers the APR on an account to *retain* a customer



- Such direct marketing activities often show a negative return in the short run
  - Cost of sending many letters to acquire one customer
  - 0% or reduced APR reduces revenue on account
- But often benefits become apparent in the long run
  - Steady flow of revenues
  - Reduced churn rate

The lifetime value of the customer is a measure to evaluate customer profits over the long term and compare them to current marketing costs

$$\text{CLV} = \text{LTV} = \sum_{t=0}^{T} \frac{\mathbb{E}(\text{customer profit}_t)}{(1+r)^t}$$

- Customer lifetime value (CLV) and lifetime value of the customer (LTV) are synonymous
- r is the discount rate

revenues × %margin
- marketing costs

customer profit
× prob. active

discounted
expected profit

Year	0	1	2	3	4	5
Revenues	140	180	190	200	210	210
Margin	40%	40%	40%	40%	40%	40%
Marketing costs	20	15	11	11	10	10
Customer profit	36	57	65	69	74	74
Probability active	100%	89%	73%	64%	59%	55%
Expected profit	36	50.7	47.5	44.2	43.7	40.7
Present value of exp. profit	36	46.1	39.2	33.2	29.8	25.3
Lifetime value	**209.6**					

Example with constant customer profits and retention rate (90%):

Year	0	1	2	3	4	5
Revenues	188	188	188	188	188	188
Margin	40%	40%	40%	40%	40%	40%
Marketing costs	13	13	13	13	13	13
Customer profit	62.2	62.2	62.2	62.2	62.2	62.2
Probability active	100.0%	90.0%	81.0%	72.9%	65.6%	59.0%
Expected profit	62.2	56	50.4	45.3	40.8	36.7
Present value of exp. profit	62.2	50.9	41.6	34.1	27.9	22.8
Lifetime value	**239.5**					

(Discount rate: $r = 0.1$)

Retention rate

- The retention rate is the probability that a customer stays active between two periods
- If the retention rate α is constant through time, the probability that a customer is still active $t = 1, 2, \ldots$periods from now is α^t
- If the retention rate and customer profits are constant, the lifetime value formula simplifies to:

$$\text{LTV} = \sum_{t=0}^{T} \frac{\alpha^t \cdot \text{customer profit}}{(1+r)^t}$$

Calculating CLV/LTV: A simple approach

- A simple approach (based on house file data) to calculate LTV assuming constant retention and customer profits across time:
  1. Calculate the retention rate based on the fraction of customers who remained active between last year and now
  2. Calculate average customer profits last year
  3. Apply LTV formula on previous slide
- Technical note
  - This approach is only approximately correct, because the LTV evaluated at the average retention rate and customer profits is not the same as the average LTV across all customers
  - This problem is often ignored in practice

Calculating CLV/LTV: Cohort method

- The cohort method allows for time-varying retention rates and customer profits
- Approach
  1. Create a list of customers from your database, and record the profits derived from them at each age, i.e. time on file
  2. For each age, average over customers (including those who are no longer active) to derive the average customer profit
     - This average incorporates both retention and the average profit of those customers who are still active
- Technical note
  - The cohort method corrects the problem described on the previous slide

Example:

Year	Customer no.														% still active	Expected profit
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
0	92	102	56	90	104	67	209	89	102	77	102	34	104	104	100.0%	**95.14**
1	104	67	103	53	0	90	287	104	112	145	177	78	134	278	92.9%	**123.71**
2	134	0	209	103	0	23	189	209	159	201	219	0	278	78	78.6%	**128.71**
3	132	0	156	106	0	0	0	178	153	178	145	0	189	0	57.1%	**88.36**
4	119	0	176	134	0	0	0	213	172	205	0	0	204	0	50.0%	**87.36**
5	124	0	205	0	0	0	0	212	196	179	0	0	245	0	42.9%	**82.93**

average over all 14 original customers

Implied lifetime value (10% discount rate)

$$\text{LTV} = 95.14 + \frac{123.71}{1.1} + \frac{128.71}{1.1^2} + \frac{88.36}{1.1^3} + \frac{87.36}{1.1^4} + \frac{82.93}{1.1^5} = 491.5$$

- Assumption made in cohort method:
  - Retention rates and profits from customers in the past are predictive of future customer behavior
  - Unlikely to be true if your product, service level, pricing structure, competition, etc., have changed
  - Predicting retention and customer profits from last year's and this year's data might then give a more accurate prediction of LTV
- Extension of cohort method:
  - "Mix and match" customers who joined in different years
  - Year when customer joined $\rightarrow$ assign to age 0
  - Year t after customer joined $\rightarrow$ assign to age t
  - "Age 0" customers could have joined in 2008, 2009, 2010, etc.
  - Can be useful if there are only few customer observations in your database

LTV and customer heterogeneity

Goal: Allow for differences in customer behavior and lifetime value

Approach 1

- Segment database by demographics or other variables
- Calculate LTV for each customer segment using the cohort method

Approach 2

- Calculate realized LTV for each individual customer i
- Collect demographics and other variables for each customer
- Regress LTV of customer i on demographics and other variables:

$$\text{LTV}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + X_K X_{Ki} + \epsilon_i$$

Churn management

Retaining customers

Retaining customers—*churn management*—is an important task in many industries

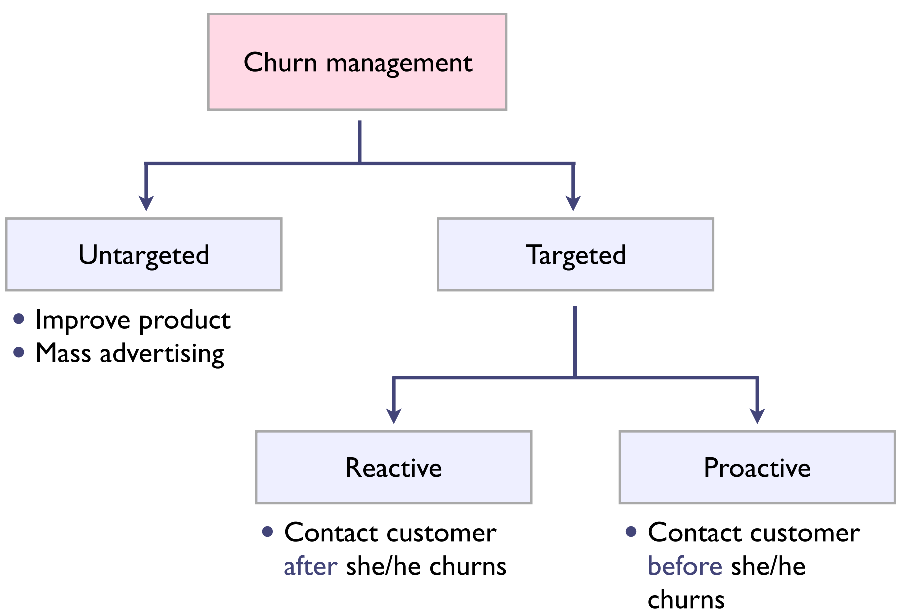
Industry	Year	Company	Churn rate (annual %)
Wireless phone	2006	Verizon	10
		Cingular/ AT&T	18
		Sprint/Nextel	22
		T-Mobile	22
Satellite radio	2005	XM Satellite	28
Financial services	2001	UK Industry Average	20-30

Source: Blattberg, Kim, and Neslin (2008)

Key concepts in churn management

- *Churn rate* = percentage of customers who terminate relationship with company between two periods
  - churn rate $= 1 -$ retention rate
- Expected lifetime of customer
  - expected lifetime $= 1/$churn rate
  - Example: churn rate = 25% per year ⇒ expected lifetime = 1/0.25 = 4 years
- Voluntary and involuntary churn
  - *Involuntary churn*: Company terminates relationship with customer (e.g. if customer is delinquent)
  - Management concern is mostly about *voluntary churn*

Overview of churn management



Basic economics of churn management

How does churn management affect customer profitability?

If churn management reduces then churn rate then:

- Lifetime of customer increases
- Hence, lifetime *value* of customer increases

Evaluating the profitability of a churn management program

- Compare customer LTV with churn management to baseline LTV
- Requires knowledge of the churn rate under the churn management program:

$$\text{churn rate with incentive} = \text{baseline churn rate} \cdot (1 - \gamma)$$

  - γ: *success rate* — probability that customer stays active for one period because of the incentive

Year	0	1	2	3	4	5
Revenues	200	200	200	200	200	200
Margin	40%	40%	40%	40%	40%	40%
Incentive cost	0	0	0	0	0	0
Customer profit	80	80	80	80	80	80
Probability active	100.0%	80.0%	64.0%	51.2%	41.0%	32.8%
Expected profit	80	64	51.2	41	32.8	26.2
Present value of exp. profit	80	58.2	42.3	30.8	22.4	16.3
Lifetime value	249.9					

baseline churn rate is 20%

incentive applied only in period 0

Year	0	1	2	3	4	5
Revenues	200	200	200	200	200	200
Margin	40%	40%	40%	40%	40%	40%
Incentive cost	C	0	0	0	0	0
Customer profit	80	80	80	80	80	80
Probability active	100.0%	90.0%	72.0%	57.6%	46.1%	36.9%
Expected profit	80	72	57.6	46.1	36.9	29.5
Present value of exp. profit	80	65.5	47.6	34.6	25.2	18.3
Lifetime value	271.2 $-C$					

all future expected profits increased due to permanently higher probability of being active

If incentive has 50% success probability of reducing baseline churn rate, the churn rate in year 1 is reduced from 20 to 10 percent

Value of incentive: $(271.2 - C) - 249.9 = 21.3 - C$

Predictive modeling for churn management

- Identify and select customers with high churn rates
  - Example
    - Baseline churn rate is 12%
    - Use model to identify customers who have 25% churn rates and higher
    - Target these customers for retention
- Determine *which factors are associated with high churn rates*
  - May be informative of why customers churn and how we can prevent churn
  - Example
    - Model reveals that customers who have experienced tech support wait times exceeding 10 minutes more than once during the last year are more likely to churn
    - Give priority to these customers when they make future service calls

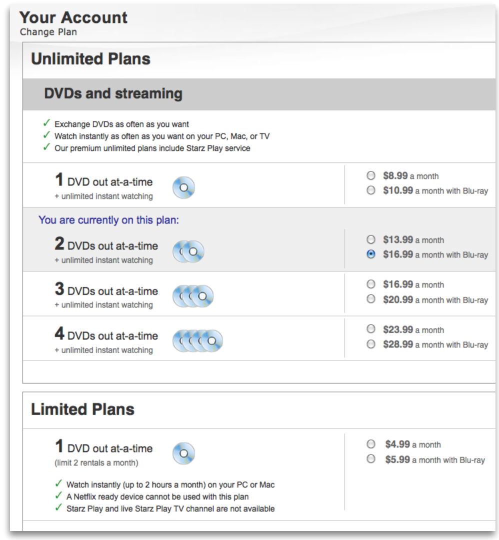
Churn management steps

1. Develop a model to predict customer churn
2. Use the model to identify customers who are most likely to churn (at least n percent more likely than average)
3. Use model to understand the main drivers of churn
4. Use insights from 3. to develop appropriate incentives/offers
5. Decide which offers to target to which customers
6. Evaluate results, ideally using A/B testing

Example: Netflix

- 81.5 million subscribers
- 100,000+ titles
- Revenue sharing with movie studios
  - \$1.5 per rental (high) — acquire DVD after initial time period
  - 50c shipping cost
  - Subscribers choose among various plans
  - Churn rate = 1.34% per month
    - 14.95% per year

$$= 1 - (1 - 0.0134)^{12}$$



Note: The numbers used in this example are hypothetical and not an accurate reflection of churn at Neflix

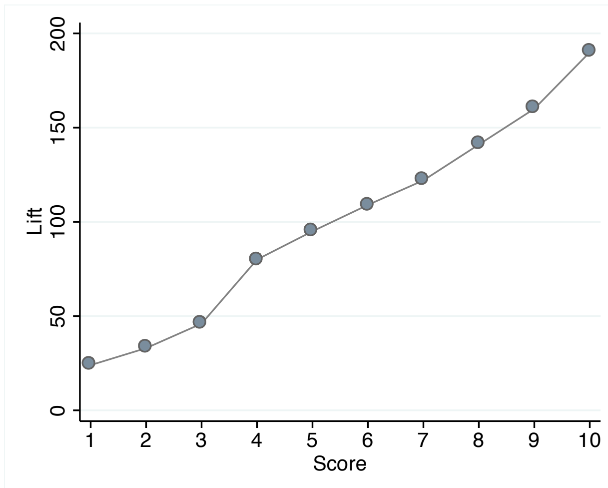
Step 1: Develop a model to predict customer churn

- Dependent variable
  - Did customer cancel Netflix in January 2009?
- Independent variables
  - July - December 2008, for 3-at-a-time plan

Predictor	Dummy	Mean	SD	Min	Max
Total no. of movies rented		33.87	21.94	0	141
Avg. monthly change in no. movies		0.2	0.78	-2.4	3.1
Avg. shipping delay		0.45	0.62	0	8
Lowest 10% in no. movies for plan	yes	0.1	0.3	0	1
No. reported scratched		0.6	0.84	0	5
No. reported missing		0.3	0.41	0	6
Moved in last 6 months?	yes	0.06	0.24	0	1
Avg. movie rating		3.63	0.83	1.7	5
Age 60+	yes	0.11	0.31	0	1
Resulted from referral	yes	0.23	0.42	0	1

Step 2: Identify customers who are most likely to churn

- Use predictive modeling to predict churn probability for each customer
  - Estimate the model
  - Validate the model to examine its predictive power



Step 3: Use model to understand main drivers of churn

- Which variables particularly influence the churn probability?
  - Examine precision/statistical significance

- What is the quantitative importance of these variables?
  - Predict how a change in one input, ΔX_k, affects the churn probability:

$$\Delta \Pr\{Y = 1|X\} \approx \Delta X_k \cdot \text{marginal effect}$$

  - Predict the marginal effects
  - If X_k is a dummy variable, then $\Delta x_k = 1$
  - Otherwise set Δx_k equal to standard deviation of X_k —allows for comparison of normalized (re-scaled) effect sizes

variable is statistically insignificant

variable has large effect on churn probability

Predictor	Coef.	z-stat	SD	Marginal effect	Change in prob.
Total # of movies rented	-0.011	0.943	21.94	-0.0008	-0.0174
Avg. monthly change in # movies	-0.892	7.34	0.78	-0.0641	-0.05
Avg. shipping delay	0.549	2.471	0.62	0.0395	0.0245
Lowest 10% in # movies for plan	1.196	13.458	0.3	0.086	0.086
# reported scratched	0.321	2.698	0.84	0.0231	0.0194
# reported missing	0.79	3.961	0.41	0.0568	0.0233
Moved in last 6 months?	0.004	0.572	0.24	0.0003	0.0003
Avg. movie rating	0.154	2.762	0.83	0.0111	0.0092
Age 60+	-0.188	3.866	0.31	-0.0135	-0.0135
Resulted from referral	-0.313	2.971	0.42	-0.0225	-0.0225

Step 4: Use insights to develop appropriate incentives

Predictor	Effect size	Effect sign	Actionable?	Where in customer Lifecycle?	Incentive/ offer
Avg. monthly change in no. movies	-0.05	-			
Avg. shipping delay	0.024	+			
Lowest 10% in no. movies for plan	0.086	+			
No. reported scratched	0.019	+			
No. reported missing	0.023	+			
Avg. movie rating	0.009	+			
Age 60+	-0.014	-			
Resulted from referral	-0.023	-			

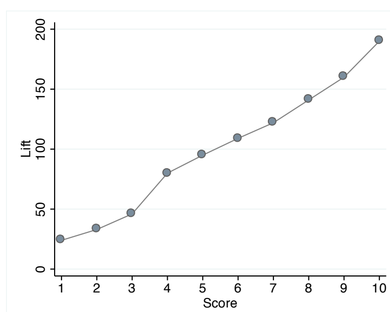
Predictor	Effect size	Effect sign	Actionable?	Where in customer lifecycle?	Incentive/offer
Avg. monthly change in no. movies	-0.05	-	yes	Retention	Send DVD on top of queue
Avg. shipping delay	0.024	+	yes	Retention	Give shipping priority for next 6 months
Lowest 10% in no. movies for plan	0.086	+	yes	Retention	Offer rebate equal to lower plan cost
No. reported scratched	0.019	+	yes	Retention	Offer rebate
No. reported missing	0.023	+	maybe	Retention	Offer rebate if not indicative of fraud
Avg. movie rating	0.009	+	no		
Age 60+	-0.014	-	yes	Acquisition	
Resulted from referral	-0.023	-	yes	Acquisition	

Note:

- A good predictor of churn does not mean we can also act on it
- A model of churn can also highlight actions that need to be taken at the acquisition or development stage

Step 5: Decide which offers to target to which customers

- Find customers with the highest churn rate
- Find factors that might lead to the highest success rate, i.e. are most likely to prevent churn
- For now: Assume (maybe based on data from prior plans) the success rate γ, i.e. the probability that the incentive prevents churn



Lowest 10% in no. movies for plan
Avg. monthly change in no. movies
Avg. shipping delay
No. reported missing
No. reported scratched

Given a 14.95% churn rate and a success rate of 50%, the churn rate drops to 7.48%. Hence, the retention rate increases from $\alpha = 1 - 0.1495 = 0.8505$ (85.05%) to $\alpha = 1 - 0.0748 = 0.9252$ (92.52%)

The customer lifetime values at the baseline retention rate and the increased retention rate under the incentive are:

$$\text{LT}V_{baseline} = \text{LTV}(\alpha = 0.8505)$$

$$\text{LT}V_{incentive} = \text{LTV}(\alpha = 0.9252)$$

Note: Calculate $\text{LT}V_{incentive}$ not including the cost of the incentive!

Using these results, we can calculate the profitability (value) and the ROI of the incentive:

$$\text{incentive value} = \text{LT}V_{incentive} - \text{LT}V_{baseline} - C$$

$$\text{ROI} = \frac{\text{LT}V_{incentive} - \text{LT}V_{baseline} - C}{C}$$

Design a precise incentive plan

	If customer is in lowest 10% in # movies for plan	Offer retrospective rebate equal to difference in current plan and lower plan for last 4 months (C = 3*4 = $12)
`else:`	If DVD rentals fall consistently over a 4 month period	Send DVD on top of queue 3 times (C = 2*3 = $6)
`else:`	If the average shipping delay exceeded 2 days	Put customer on top of priority scale during next 6 months if a movie is in short supply (C = ?)
`else:`	If reported 1 scratched DVD in last 3 months	Offer 1 additional at-a-time DVD for 1 month (cost depends on increased frequency*$2)
`or:`	If reported 2 scratched DVDs in last 3 month	Offer 3 additional at-a-time DVDs for 1 month (cost depends on increased frequency*$2)

Evaluate the profitability and ROI of these incentives using the approach on the previous slide

Step 6: Evaluate the results — plan testing

Value of incentive depends on the success rate γ which is not estimated! Hence, we cannot perfectly predict the effectiveness of the incentive plan purely based on the model estimates

Recommend approach to estimate the effectiveness of the incentive plan: A/B testing

- Focus on a specific incentive and all customers who qualify
- Random split into treatment and control group
  - No incentives given to the customers in the control group
  - Customers in treatment group are given the incentive

Example

- \$12 rebate for customers in "lowest 10% in no. of movies for plan" group
- Observe churn and revenues for 6 months (or other time period)

Group	N	Retained	LTV
Control	2,000	92.2%	**\$256**
Treatment	18,000	96.5%	**\$289**

- Churn rate reduced from 7.8% to 3.5% over 6 month horizon — 15% to 7% at annual frequency
- Value of incentive is 33 -12 = \$21

Appendix: Oversampling

Oversampling

Oversampling: Create a training sample from the whole data base that balances responses and non-responses.

- Old-fashioned, and typically applied when estimating a logistic regression if the original data set is very large

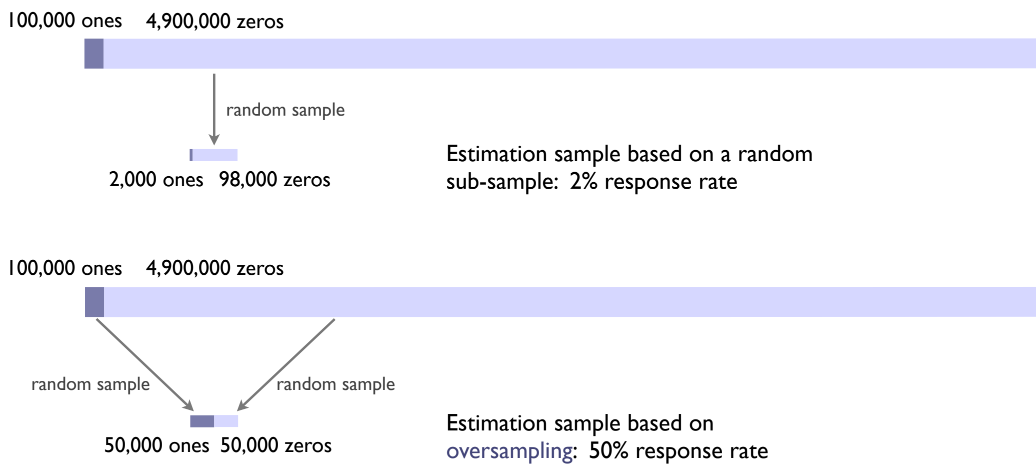
Approach:

- Choose target size of training sample, N
- Randomly sample $N/2$ data points among observations with $y = 0$
- Randomly sample $N/2$ data points among observations with $y = 1$

Rationale: In a balanced estimation data set a logistic regression model can more easily distinguish among the ones (responses) and zeros (non-responses) in the data, hence the precision of the estimates increases

Example

- Original data: 5,000,000 observations, response rate is 2%
- Goal: Create estimation sample with 100,000 observations



Oversampling in estimation sample, not validation sample

Note

- Oversampling is used to create the estimation sample, in order to increase the precision of the estimates
- Oversampling is not used to create the validation sample, because the validation sample should reflect the true response rate in the data

Oversampling and out-of-sample prediction

Fitted logistic regression model will predict the average response rate in the training sample — typically much larger than the true response rate

To be precise:

- Intercept overestimated
- Slope coefficients will be *unbiased*, however (only true in logistic regression model)

Over-prediction of response rates: Solution 1

Solution 1 applies if:

- We validate the predictive model in the validation sample (lift and gains table, etc.)
- Targeting is based on segments (scores), not individual-level targeting

In this case the over-prediction problem can be ignored, because over-sampling affects the predicted *response rates*, but not the predicted r*ank-order*. In particular, if p_i is the true response rate for observation i and p_i^o is the predicted response rate based on oversampling, then typically $p_i < p_i^o$. However, the rank-order is preserved: For any two observations i and k:

$$p_i < p_k \Longleftrightarrow p_i^o < p_k^o$$

Because we rank the observations correctly along predicted response rates, we create the same segments/scores that we would have created had we estimated the model without oversampling.

- Lifts and gain are correct, and we can use the observed response rates in the *validation sample* as a basis for targeting

Over-prediction of response rates: Solution II

Solution II directly corrects the predicted response rates

Approach

1. Calculate an offset variable:

$$\begin{aligned}\text{offset} &= \log\left(\frac{\alpha_e}{1-\alpha_e}\right) - \log\left(\frac{\alpha_s}{1-\alpha_s}\right) \\ &= (\log(\alpha_e) - \log(1-\alpha_e)) - (\log(\alpha_s) - \log(1-\alpha_s))\end{aligned}$$

   - α_s is the true response rate (and hence also response rate in validation sample), α_e is the response rate in the training sample)

2. Estimate the logistic regression model by supplying the offset variable to the estimation routine
   - The offset is a fixed number and will affect the estimated intercept only
3. Set the offset variable to 0 and then predict response probabilities and marginal effects

Why does this work? — Based on the logistic regression formulas for the response probabilities, we have:

$$\frac{\Pr\{Y=1|X\}}{\Pr\{Y=0|X\}} = \frac{\frac{\exp(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}{1+\exp(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}}{\frac{1}{1+\exp(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}}$$

Take the log on both sides of the equation, and note that $\Pr\{Y=0|X\} = 1 - \Pr\{Y=1|X\}$:

$$\log\left(\frac{\Pr\{Y=1|X\}}{1-\Pr\{Y=1|X\}}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Interpretation: The probability index predicts the log-odds ratio of the response rate

Setting offset $= 0$ is equivalent to

(i) Subtracting the incorrect log-odds ratio, $\log(\alpha_e) - \log(1-\alpha_e)$

(ii) Adding the correct log-odds ratio, $\log(\alpha_s) - \log(1-\alpha_s)$

Summary

- Predict customer lifetime value for any marketing decisions that have long-run profitability consequences
  - Retention and acquisition
- Churn management
  - Use predictive modeling to identify customers with high churn rates and to understand what factors cause churn
  - Target offers to prevent churn based on incremental customer lifetime value
  - Test churn management program