Case Final Report

"We have neither given nor received unauthorized aid on this deliverable."

- *Brian Sacco, Maggie Chen, Samantha Ripani, Caleb Deckard*

1. Consumption Variable

Consumption Variable: Consumption

We chose the Consumption variable as the consumption-like variable because it measures how many movies the customer orders each month. This variable clearly can help the company forecast its revenues. For example, if they see that a customer's consumption is declining and heading towards zero, the company can forecast that the customer may defect and that revenue stream will dry up. Also, the company can analyze its costs with this variable by seeing how much it costs to send that number of DVDs to the client each month. Descriptive statistics for the Consumption variable can be found in *Figure 1* of the Appendix.

Drivers of Consumption

After analyzing the correlation coefficients between the Consumption variable and the rest of the variables (see *Figure 2* in Appendix), we decided the following variables are drivers of Choice: month, priority, ucprice, and lead_op*.* Descriptive statistics for these variables can be found in *Figure 3* of the Appendix.

Marketing Insights

Month: Customers stay on average 5 months. The correlation between consumption and month is slightly negative, meaning that the longer the customer stays, the less DVDs they consume.

Priority: The average priority ranking of DVDs that customers consume is 3. If customers don't have access to their favorite movie, they may be less likely to stay. Fortunately, the modal value is 1, meaning that more customers than not are getting their number 1 choice of movie each month. Also, correlation between consumption is slightly positive, meaning that when customers have access to DVDs they are most interested in, they are more likely to continue consuming.

Ucprice: This is the pre-rental price, based on the customer's actual consumption (as opposed to the maximum consumption). The correlation is strongly negative, meaning that when customers take full advantage of their maximum consumption (quota), they consume less over their lifetime with the company.

Lead_op: This is the amount of the customer's over purchase in the previous period. The correlation is slightly negative, meaning that when customers over purchase in the previous period, their consumption in the current period decreases. Also, the average is 4.37, meaning that customers are indeed over purchasing routinely. Most importantly, the maximum value is 16, which shows that the company is not doing an effective job at marketing higher-quota plans to high-consuming customers.

2. Log-Linear (One Segment)

We have attached a screenshot of our log-linear regression model in the Appendix *(*see *Figure 4)*. We have also attached it to the submission - the file is named “LogLinear”.

<u>Coefficient Interpretation</u>
On a scale of 0-infinity, our MAE is 0.197. We believe that since this value is so close to 0, that we have a solid model. Our R^2 though is only .3178, so there is a lot of room for improvement as well.

Consumption .1573: This variable has a slight impact on the model. This makes sense since we are analyzing consumption and would expect the variable to have a large impact on the company’s revenues and costs, although this variable is lower than one may assume, it still affects the overall customer utility.

Month -.0097: This coefficient means that the n^{th} month of the customer lifetime with the company has a very slightly negative impact on consumption of DVDs. The longer the customer stays with the company, the less they consume. This is consistent with our findings when we analyzed the descriptive statistics for the month variable.

Priority .0068: Priority has very small, positive impact on the model. When customers obtain the DVDs that they prioritize, they are more likely to consume more.

Lead_op -.0129: The previous month’s level of over purchase has the smallest negative impact of the driver variables analyzed.

Uprice .0141: The price based on actual consumption also has a slight impact on the model. This is surprising because this variable can be looked at as a monthly variable cost for the customer.

Based .266: This variable is positive showing that this affects the overall customer utility. This variable represents the value of the product perceived by the customer. A higher value perceived by the customer would likely produce more customer utility. If the customer sees the product as valuable, they are more likely to purchase the product.

The company should target customers whose consumption is declining on a monthly basis. To regain the customer’s attention, they can advertise different genres or push movies similar to those on the customer’s priority list to the customer. The company should also engage in a discussion with the customer to understand why the customer’s consumption is declining. If the company successfully discovers why the customer’s consumption is declining, the company should tailor its solution specifically for that customer. As a result of this targeting, the company will increase retention rates, while also better understanding reasons that customers are defecting.

Since the correlation coefficients show that consumers are likely to upgrade to a better plan based on how many movies they watch, the company should make customers aware of the next plan upgrade available. The company should closely watch and evaluate the change in the consumption of DVDs per month. If the company better engages the consumer when consumption starts to decline or increase, they can propose a better plan to best fit the customer's needs. For example, if the customer currently uses the Standard plan, but is over purchasing, the company should let the customer know that they could upgrade to the Premium plan and be more economical.

3. Binary Customer Choice Variable (Pre-Step for 4)

Binary customer choice variable: Stay

We chose the variable Stay because it is a variable with a value of 0 if the customer leaves the company, and 1 if the customer stays. This is essentially the same as Retention. This has a large impact on the firm's revenues and profits because they can only earn revenue and profit if their customers stay with the firm. Once a customer defects, that revenue stream dries up. Descriptive statistics for this variable can be found in *Figure 5* of the Appendix.

Driver Variable: UPrice

UPrice: This variable is the per-rental price based on the maximum number of DVDs allowed by the plan. This is a marketing action variable because the firm can manipulate how much it charges for each DVD and each plan.

Descriptive statistics for the customer choice driver variables can be found in *Figure 6* of the Appendix.

Consumption variable: Consumption

Consumption: This variable shows the amount of DVDs the consumer has viewed in one time period (month).

Marketing Insights:

Uprice: The average pre-rental price is $3.28. The standard deviation is about $0.59. The correlation is slightly positive, meaning that when the pre-rental price increases, the likelihood of retention does too. This correlation is very little though, and likely has little impact.

4. Binary-Logit (One Segment)

Binary Logit Model

We have attached a screenshot of our binary logit model to the appendix (see Figure 7). We have also attached the Excel file along with this submission. It is named “binarylogitonesegment”.

Coefficient Interpretation

Consumption 12.844:

This number means that if the consumer continues to consume the DVDs, there is an increase in 12.844 units in customer utility. This variable is extremely strong and is the strongest of our coefficients. This is likely because the more a DVDs a user watches, the more valuable they are to the company .

Uprice 0.205:

This variable is the marketing action variable, which means we could control this variable from the company side. “Uprice” is the per-rental price based on the maximum number of DVDs allowed by the plan. The coefficient we ran after the solver here means that if the company increases one unit of the “uprice”, then it will also increase 0.189 units of customer utility. This number is positive, meaning that by increasing behavior at the “uprice”, there would be a positive impact on customers’ value perception. This marketing action could be helpful for the company to increase the customer’s utility. Even though it is just a relatively small impact, the increase in the “uprice” still could help the company to increase its customer value perception for its products.

Base tendency stay 0.355:

This is the second strongest coefficient of the three measured. This means that when no marketing actions are taken, customers still relatively show more loyalty to the company. We could say that most of the customers would still choose to stay with the business. We could also say the customers are currently satisfied with the services offered.

Implications

Hit rate is the ratio we used to see the prediction of our estimation. Our hit rate is 89.88%, which is relatively high and a good sign. 89.88% here means that our prediction is around 90% correct, so we could use this model to estimate the marketing action of changing the “uprice” for the company whether would bring the negative or positive impacts for future estimation. We can use this information to our advantage when customers are about to defect. For example, if our model predicts a customer to leave after month 7, the firm should make sure to place that customer’s movie prioritizations as number one for that month.

While we changed this utility to a percentage, we could see that if the company doesn’t do anything, around 87% of the customers would still choose to stay with the business. We’re happy to see this ratio, because we realize that most of the customers are satisfied with the current services. Whether or not the company decides to make the marketing action "uprice", most of them would still stay with the company. Compared with the coefficient of the "uprice", which is 0.205, we realize the marketing action for the business could be helpful, but not necessary.

However, the month coefficient from the customer part is a negative impact. Then we could see that the longer the membership they have, the lower the value they would perceive in their minds. The firm can also send targeted promotions to customers based on what their top movie priorities are. For example, if Customer X prioritizes *Dodgeball, The Longest Yard,* and *Happy Gilmore* in his/her top three, the firm can tell that Customer X enjoys sport comedies the most. Armed with this information, the firm can push promotions for similar movies, such as *Talladega Nights* or *Caddyshack* to increase Customer X's priority and consumption rankings. This will help extend Customer X's life cycle, as well as increase revenues and profits for the firm. In conclusion, we can help customers filter or prioritize the movies they would prefer to watch in order to increase the product value in customers' minds for longer membership life.

5. Binary-Logit for "Retention" (Two segments)

We have included a screenshot of our 2-segment binary logit model in the Appendix (see *Figure 8*). We have also included the Excel file in this submission - the file is titled "binarylogit2segments".

<u>Hit Rate, LRI, and AIC</u>

Hit Rate: 89.88%

LRI: 0.327

AIC: 93.55

This information can be found in *Figure 9* of the Appendix.

<u>Coefficient Meanings</u>

Consumption: This coefficient indicates the consumption for the consumers in both segments. The b2 for seg 1: is 5.574 and the b2 for seg 2: 10.804, indicating that the consumer in the second segment consumers nearly double than those in the first segment. If a customer is in either of the segments they are consuming DVD's; however, it would be beneficial to try to focus marketing actions on those in seg 1 to increase their consumption and overall profitability for the company.

Uprice: This coefficient indicates the marketing action did by the company side. The b2 for seg1 is .317, which indicates that if the company chooses to increase one unit of the "upice", there would be .317 units increasing in customer utility. The b2 for seg2 is .205, and this means that if the company chooses to increase one unit of the "uprice", there would be .205 units increasing in customer mind.

From the overall looking, we could see that “increasing uprice” marketing action brings positive impact to customers from both of the segments. However, customers from the seg1 would benefit more from this marketing action.

Base tendency to stay: This coefficient indicates the based tendency to stay with the company from the customer side while company does not do any marketing actions. The attitudes for the loyalty from different segments are totally different. The b3 for seg1 is -.037, which means that customers in this segment are more likely to defect if the firm does not take any marketing actions. They are inherently less loyal than seg2 customers. The b3 for seg2 is .356, which means that customers in this segment are unlikely to defect if there are no marketing actions. This segment has customers that are inherently more loyal to the firm and depend less on marketing actions for retention.

Also, we have lambda this time. To change it into the percentage, we can read that around .0917% of probability that customers would be stay in segment 1.

This information can be found in *Figure 8* of the Appendix.

6. Comparing One-Segment and Two-Segment Model Performance

Comparison between One-segment model and Two-segment model performance

AIC: From the comparison of these two models, we could conclude that the AIC is lower in the “one-segment” model than the AIC in the “two-segment” model (85.505 < 93.555). This is better, because AIC is useful while we have two numbers to compare, and the lower one is the one which can be better used to analyze the overall dataset.

LRI: The ratio for the LRI from the “one-segment” model is higher than the first one (32.7% > 32.65%). The higher the LRI, the better we could conclude that the data fits our model. We could see that from the model construction part, the “one-segment” model is better than the “two-segments” model, because both two comparisons tell us the higher accuracy of the new model application.

Hit Rate: This tells us how accurate our prediction is. We could see that the hit rates from both of the two models are the same, (both are 89.88%), so we could say that the accuracy of our prediction from the model is not decreasing – this is good because we could conclude that our conclusion is still reliable. However, the second model would be the better choice for the manager because once the manager has a larger given dataset, this model could be more reliable to draw conclusions in the future.

This information can be found in *Figure 10* of the Appendix.

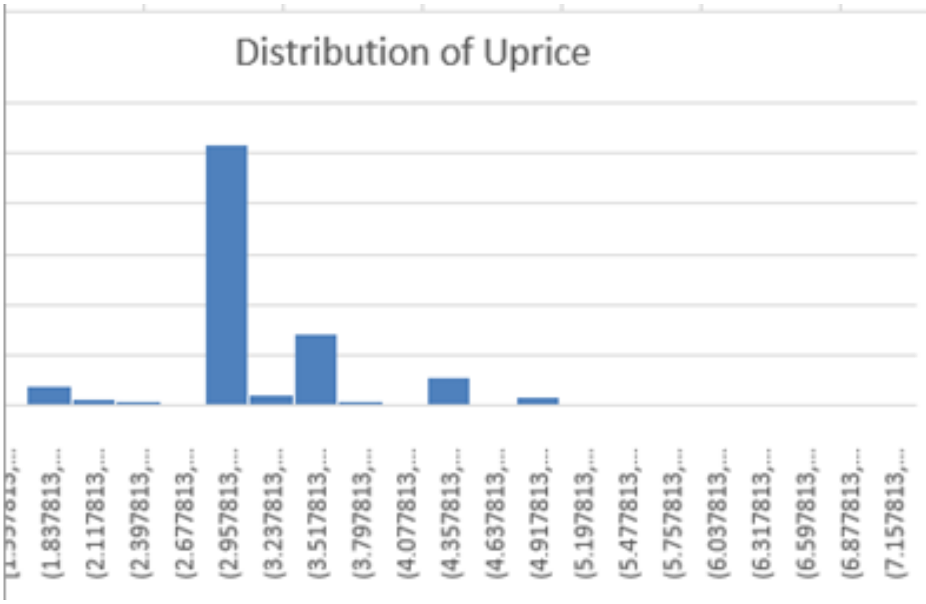
7. Dynamic Optimization

As determined in step 6, we are using a one-segment model to best determine marketing actions.

Based on the model we ran for our consumption variables, we take consumption, month, priority, lead_op, uprice as our considerable factors; based on the binary model, we chose uprice as the one to predict the probability of retention.

“Uprice” is the only marketing action variable from the business side – this is the only variable that could be controlled by the company side instead of consumer decisions. Then, in our case study, we would distribute our data into three sets based on the chosen level for “Uprice”: $1.84, $2.96 and $5.

As displayed from the histogram, we could see that most of the uprice the company would discharge to customers is concentrated on the lower-case level. Considering the minimum, maximum, and also the average value for this variable, our team takes these three uprice levels as marketing actions to predict the profit based on the dynamic programming.



min	1.557813
max	7.225
avg	3.282189

415

A-1	1.84
A-2	2.96
A-3	5

From our case, each consumption made by a customer would increase one unit of variable cost, which is $.45 for one-way delivery and an estimated $1.10 for overhead costs ($0.45 + $1.10 = $1.55). We still use 0.9 as our discount factor at calculation. Our revenue is collected from each payment made by a customer. Thus, based on the given data, our observed profit is calculated by the payment minus total cost (unit of variable cost * number of consumption) and also taken into account is the optimum profit, which can be obtained from next period if this customer chose to stay with us (discount factor * profit from next period).

discount factor				a (unit cost/consumption)				Uprice_1	1.84	
0.9				1.55				Uprice_2	2.96	
								Uprice_3	5	
										Block for t
customer	month	obs log(C)	obs Ret	obs Uprice	payment	consumptior	obs profit	priority	lead_op	log(C0) | A
1	0	0.301029996	1	3.570313	22.85	1	=F10-$E$6*G10+IF((D10=1)*AND(A10=A11),$A$6*H11,0)			
1	1	0.698970004	1	3.570313	22.85	4	87.5756205	2.8325	5.4	0.478174

After calculating the observed profit from the given data, we then begin to calculate the expected profit based on our log-linear model for consumption prediction and binary model for retention estimate.

Based on what we've run for our consumption and retention analysis, we should take those variables into our calculation. For a customer whose id is "1", we use the coefficient of "based tendency" as a direct result; starting at the second column for calculating the log consumption, we then use the "if" function for each column (This won't be misleading even at the following start of the new customer). Because we distribute our marketing actions into three Uprice levels – A=$1.84, A=$2.96 and A=$5, we need to multiply the coefficient of a4-uprice with each uprice level in the three columns, and then we have three results for our log consumption calculation.

Consumption parameters								Retention parameters				
a0-consum	a1-month	a2-priority	a3-lead_op	a4-uprice	a5-based-tendency			b0-Uprice	b1-Consun	b2-Base tendency of Stay		
0.187263	-0.00841	0.006497607	-0.022129805	-0.064598859	0.595101			0.204856	12.84373	0.355802407		
discount factor				a (unit cost/consumption)				Uprice_1	1.84			
0.9				1.55				Uprice_2	2.96			
								Uprice_3	5			
										Block for time = 0 - no info on customer consumption		
customer	month	obs log(C)	obs Ret	obs Uprice	payment	consumptior	obs profit	priority	lead_op	log(C0) | A=1.84	log(C0) | A=2.96	log(C0) | A=5
1	0	0.301029996	1	3.570313	22.85	1	100.1180585	4.665	0	0.595101143	0.595101143	0.595101143
1	1	0.698970004	1	3.570313	22.85	4	87.5756205	2.8325	5.4	=IF($A11=$A10,$A$3*L10+$B$3*$B11+$C$3*$I11+		
1	2	0.477121255	1	3.570313	22.85	2	78.806245	5.5	2.4	$D$3*$J11+$J$5*$E$3+$F$3,$F$3+0+0)		

After that, we need to calculate the retention based on our coefficients from the binary model. Also, we need to calculate three retention utilities for the three uprice levels based on the result we have from the log consumption calculation.

Retention parameters								
b0-Uprice	b1-Consun	b2-Base tendency of Stay						
0.204856	12.84373	0.355802407						
Uprice_1	1.84							
Uprice_2	2.96							
Uprice_3	5							
		Block for time = 0 - no info on customer consumption and set past C = 0						
priority	lead_op	log(C0) | A=1.84	log(C0) | A=2.96	log(C0) | A=5		U | A=1.84	U | A=2.96	U | A=5
4.665	0	0.595101143	0.595101143	0.595101143		=$I$3*$J$5+$J$3*(10^L10-1)+$K$3		

Receiving the retention utilities from the three uprice levels, we need to use the exponential formula to transform into probability of retention. Still, we have three probabilities.

U | A=1.84	U | A=2.96	U | A=5		P(ret=1|A=1.84)	P(ret=1|A=2.96)	P(ret=1|A=5)
38.4473025	38.6767408	39.09465		=EXP(P10)/(1+EXP(P10))		1

Calculating the expected profit for each uprice level, we use "backward" solution, to define the last E[profit 0 | A=X] = "payment". For other obs, E[profit 0 | A=X] = payment + P[ret = 1 | A =X] * discount factor * best_profit from next period.

discount factor									
0.9									
		Block for time = 0 - no info on customer consumption and set past C = 0							
customer	payment	P(ret=1|A=1.84)	P(ret=1|A=2.96)	P(ret=1|A=5)	E[profit0|A=1.84]	E[profit0|A=2.96]	E[profit0|A=5]		best-profit
1	22.85	1	1	1	=F10+T10*$AS$6*$AG11		633.1613		633.161349
1	22.85	1	0.999999999	0.999996875	527.5462795	508.696279	592.2703		592.2703229

The next step is to find out which uprice level we should build for each customer at each time period based on the expected profits we have calculated. We first use the "max" function to find out the largest expected profit from the three uprice levels. Then, we use the "if" function to find out which uprice level the company should make to aim at a specific customer at a specific time period. Also, we use conditional formatting from Excel to highlight the uprice level, which should be taken. This visualization could be easier for a manager to adapt while using our model.

E[profit0|A=1.84]	E[profit0|A=2.96]	E[profit0|A=5]		Max	A=1.84?	A=2.96?	A=5?
555.8932906	534.0432906	633.161349		633.161349	=IF($AB10=X10,1,0)		
527.5462795	508.696279	592.2703229		592.2703229	IF(logical_test, [value_if_true],		
504.8180366	483.9680366	560.7736438		560.7736438	0	0	1
492.7860897	471.9360855	535.5200407		535.5200407	0	0	1

The final step is to find out the "best profit" based on the three marketing actions the company could make. This calculation should be the same result we calculated from the "max" profit chosen cell.

Max	A=1.84?	A=2.96?	A=5?		best-profit
633.161349	0	0	1		=AB10

8. Actual vs Dynamic CLV

After all these steps, we then use the “if” function to find out the expected profit for each customer at time = 0. We then could compare the difference between the obs profit (actual profits from given data) and expected profit (based on optimization through dynamic programming).

	at time =0
Total obs	**Total expected by optimization**
4763.184	32727.70401
	587.10%
obs profit	**exp profit (by DP) at time = 0**
100.1181	633.161349
0	0

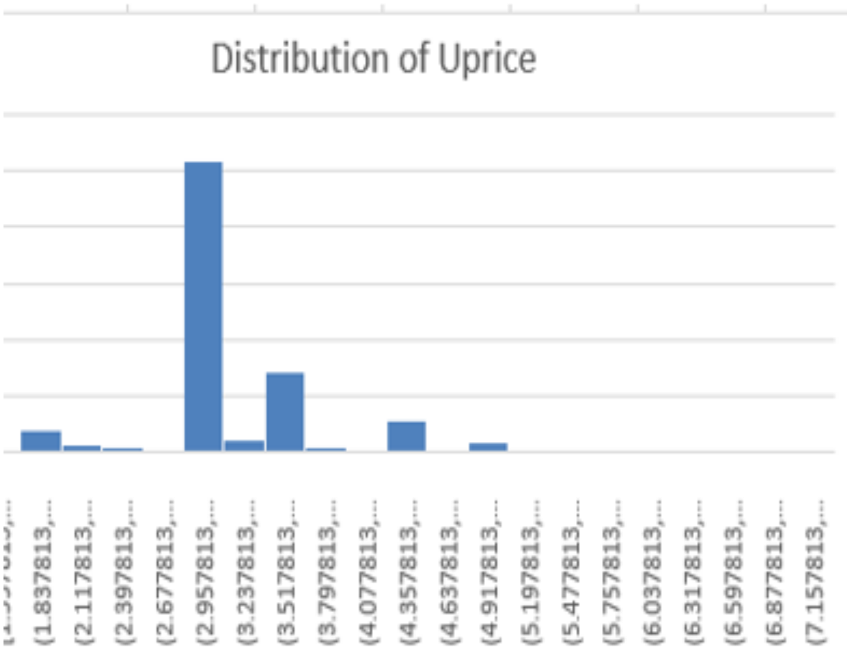
Our result here illustrates, that based on all the models we’ve run for choosing the proper consumption variables and retention estimation, our model could help managers find out the proper uprice level they could build for each customer at each period. Utilizing our model in estimation could help the company improve 587.10% in running.

The consumption variables and the retention estimation we built utilized into this dynamic programming, could help managers better realize which uprice level they should charge for most customers in order to earn a larger profit. Uprice is the only marketing action, which could be controlled and managed from the company side.

The most direct benefit we could realize why dynamic programming is better than given data estimation could be visualized from these screenshots.

Without utilizing dynamic programming, the company could just set up their uprice randomly and this could create problems with the company. The first chart is the distribution of uprice setup based on the raw data: the company would set their uprice mostly at the lower side; however, based on our model estimation, the conditional formatting at the second screenshot tells us the uprice level the company should build up in order to earn a larger profit. The company would earn a higher profit if they set the uprice level at $5, which is at the high end. Taking the time to figure out the best price to set the products at is valuable for this firm.

In conclusion, the company utilizing the dynamic programming would be more strategic for determining the price to set their products at. This estimation could help the company increase the response rate and increase the profit margin by taking marketing actions that are strategic and data driven.



>>

A=1.84	A=2.96	A=5?
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
1	0	0
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
1	0	0
0	0	1
0	0	1
0	0	1
0	0	1
1	0	0
1	0	0
0	0	1
1	0	0
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1

Appendix

Figure 1: Consumption variable descriptive statistics

consumption	
Mean	1.546987952
Standard Error	0.078206488
Median	1
Mode	0
Standard Deviation	1.593187289
Sample Variance	2.538245737
Kurtosis	0.965328224
Skewness	0.958343444
Range	9
Minimum	0
Maximum	9
Sum	642
Count	415

Figure 2: Correlation coefficients for consumption driver variables

	customer_id	*month*	*cmonth*	*tenure*	*choice*	*nswitch*	*censor_index*	*priority*	*rating*	*consumption*
customer_id	1									
month	-0.409608429	1								
cmonth	0.341023052	0.13288	1							
tenure	-0.582765443	0.70287	-0.17522	1						
choice	-0.372105644	0.34568	0.09204	0.2636	1					
nswitch	-0.725113034	0.26055	0.18209	0.3707	0.43654	1				
censor_index	0.000859312	-0.02427	-0.22882	0.01723	-0.05286	-0.07167	1			
priority	0.195881944	-0.1038	0.04209	-0.1078	-0.08098	-0.15666	-0.009315453	1		
rating	0.102624862	-0.24344	0.00631	-0.17863	-0.05876	-0.07533	-0.271888923	0.01229	1	
consumption	0.317397579	-0.31565	0.05363	-0.11594	-0.06167	-0.1233	0.088013107	0.18352	0.15034	1
totconsump	-0.169483791	0.55752	-0.02166	0.7932	0.00505	0.10775	0.112818523	0.04946	-0.15155	0.184752353
acconsump	-0.242845344	0.87793	0.14414	0.70048	0.18494	0.17438	0.04374049	0.00953	-0.20888	-0.031847494
quota	-0.050182619	0.01631	-0.10171	0.27897	-0.19071	0.05322	0.110471824	-0.02985	-0.01283	0.054872681
payment	0.055107633	-0.1312	-0.22623	0.10634	-0.4321	-0.12088	0.143201012	0.01784	0.02446	0.088406355
totpmt	-0.517640833	0.6703	-0.17376	0.95366	0.24691	0.30793	0.061681362	-0.09867	-0.1696	-0.092849671
accpmt	-0.404856063	0.96091	0.09833	0.71261	0.31604	0.25076	0.002596014	-0.10345	-0.22806	-0.286325322
uprice	-0.158668177	0.10005	-0.03312	-0.14349	0.49364	0.09375	-0.069324892	-0.00485	0.01057	-0.094521667
ucprice	-0.315489548	0.17295	-0.18966	0.27348	0.02392	0.18138	0.036869872	-0.10088	0.01292	-0.746603962
ngenre	-0.245771555	0.45357	0.03559	0.64532	0.12911	0.28879	-0.028836934	-0.12127	-0.10875	0.157672612
op	-0.210549585	0.17979	-0.11771	0.30651	-0.13568	0.1115	0.05118934	-0.12248	-0.09009	-0.475746928
lead_op	-0.301506579	0.37534	-0.04483	0.39334	0.03381	0.1882	0.045630999	-0.11631	-0.08774	-0.283726183
stay	0.047823556	-0.1127	-0.0105	0.0949	0.33552	0.00886	-0.004744333	0.06099	0.06229	0.326222721
defect	-0.750607543	0.27525	-0.40164	0.3916	0.14317	0.35931	0.062962725	-0.22354	-0.10562	-0.398751246
accept	-0.757281669	0.27794	-0.39765	0.39544	0.14852	0.36949	0.061102305	-0.2247	-0.10593	-0.399031327

Figure 3: Descriptive statistics for consumption driver variables

month	
Mean	5.065060241
Standard Error	0.269204393
Median	3
Mode	0
Standard Deviation	5.484110419
Sample Variance	30.07546709
Kurtosis	3.366313606
Skewness	1.827809582
Range	28
Minimum	0
Maximum	28
Sum	2102
Count	415

priority	
Mean	2.992040605
Standard Error	0.207922593
Median	2.25
Mode	1
Standard Deviation	4.235705238
Sample Variance	17.94119886
Kurtosis	193.0663722
Skewness	12.19886242
Range	73.25
Minimum	0.5
Maximum	73.75
Sum	1241.696851
Count	415

ucprice	
Mean	10.44809736
Standard Error	0.37457582
Median	9.975
Mode	9.975
Standard Deviation	5.993213113
Sample Variance	35.91860342
Kurtosis	1.437271306
Skewness	1.383416098
Range	27.4575
Minimum	2.4925
Maximum	29.95
Sum	2674.712924
Count	256

lead_op	
Mean	4.375903615
Standard Error	0.16549719
Median	4.4
Mode	6.4
Standard Deviation	3.371434074
Sample Variance	11.36656771
Kurtosis	3.423983112
Skewness	1.447070905
Range	16
Minimum	0
Maximum	16
Sum	1816
Count	415

Figure 4: Log-normal regression

B5 =SUM(M8:M422)

	A	B	C	D	E	F	G	J	K	L	M	N	O	P	Q	R
1	consumption	month	priority	lead_op	uprice											
2	a0	a1	a2		a5											
3	0.330656936	-0.0043	0.01451486	0.0058323	0.020127272									Percentage		
4												Mean_of_Log_bal	0.319388	0.623583		
5	SSE	26.75501														
6																
7	customer_id	month	priority	lead_op	ucprice	consumption		log (C-cur+1)	log (C-prev+1)	prediction	diff^2	AE	MAE	SSR	SST	R^2
8	1	0	4.665	0	22.85	1		0.301029996	0	0.52762	0.051343	0.226590001	0.199165	0.043361	0.000337	0.524266
9	1	1	2.8325	5.4	5.7125	4		0.698970004	0.301029996	0.282824	0.173177	0.416145582		0.001337	0.144083	
10	1	2	5.5	2.4	11.425	2		0.477121255	0.698970004	0.546307	0.004787	0.069185285		0.051492	0.02488	
11	1	3	1.5	4.4	11.425	2		0.477121255	0.477121255	0.422258	0.00301	0.054863421		0.010582	0.02488	
12	1	4	1.5	4.4	9.975	2		0.477121255	0.477121255	0.388775	0.007805	0.088346005		0.004815	0.02488	
13	1	5	1.5	4.4		0		0	0.477121255	0.183708	0.033749	0.183707668		0.018409	0.102009	
14	1	6	1.5	6.4		0		0	0	0.033311	0.00111	0.033310777		0.08184	0.102009	
15	1	7	2	6.4	4.475	2		0.477121255	0	0.12634	0.123048	0.350781542		0.037268	0.02488	
16	1	8	1.5	0		0		0	0.477121255	0.145151	0.021069	0.145151432		0.030358	0.102009	
17	1	9	1.5	2		0		0	0	-0.00525	2.75E-05	0.005245459		0.105387	0.102009	
18	1	10	1.5	2		0		0	0	-0.00954	9.11E-05	0.009543498		0.108196	0.102009	
19	1	11	1.5	2		0		0	0	-0.01384	0.000192	0.013841537		0.111042	0.102009	
20	1	12	1.5	2		0		0	0	-0.01814	0.000329	0.018139576		0.113925	0.102009	
21	1	13	1.5	2		0		0	0	-0.02244	0.000503	0.022437615		0.116845	0.102009	
22	1	14	1.5	2		0		0	0	-0.02674	0.000715	0.026735654		0.119802	0.102009	
23	2	0	3.5	0	3.74375	8		0.954242509	0	0.126153	0.685731	0.828089023		0.03734	0.40304	
24	2	1	3	8	29.95	1		0.301029996	0.954242509	1.004244	0.494509	0.70321366		0.469027	0.000337	
25	2	2	3	15	9.983334	3		0.602059991	0.301029996	0.422908	0.032095	0.179152049		0.010716	0.079903	
26	2	3	1	13		0		0	0.602059991	0.276516	0.076461	0.276515957		0.001838	0.102009	
27	2	4	1	16		0		0	0	0.09064	0.008216	0.090639506		0.052326	0.102009	

Figure 5: Customer choice variable descriptive statistics and correlation matrix

stay	
Mean	0.898795181
Standard Error	0.014822819
Median	1
Mode	1
Standard Deviation	0.301963772
Sample Variance	0.09118212
Kurtosis	5.068840697
Skewness	-2.654135555
Range	1
Minimum	0
Maximum	1
Sum	373
Count	415

	customer_id	month	cmonth	tenure	choice	nswitch	censor_index	priority	rating	consumption	totconsump	acconsump	quota	payment	totpmt	accpmt	uprice	ucprice	ngenre	op	lead_op	stay	defect	accept
customer_id	1																							
month	-0.409608429	1																						
cmonth	0.341023052	0.132883	1																					
tenure	-0.582765443	0.70287	-0.175217505	1																				
choice	-0.372105644	0.345684	0.092042508	0.263603313	1																			
nswitch	-0.725113034	0.260555	0.182093558	0.370700829	0.436537	1																		
censor_index	0.000859312	-0.02427	-0.228816764	0.017234798	-0.052864	-0.07167	1																	
priority	0.195881944	-0.103805	0.04208946	-0.10780453	-0.080977	-0.156658	-0.009315453	1																
rating	0.102624862	-0.243435	0.006313687	-0.178628472	-0.058758	-0.075331	-0.271888923	0.012291	1															
consumption	0.317397579	-0.315649	0.053631912	-0.115939886	-0.06167	-0.1233	0.088013107	0.18352	0.15034	1														
totconsump	-0.169483791	0.557517	-0.021658943	0.793200751	0.005055	0.107752	0.112818523	0.049459	-0.151554	0.184752	1													
acconsump	-0.242845344	0.877933	0.144135873	0.700481592	0.184944	0.174381	0.04374049	0.009528	-0.208885	-0.031847	0.778716	1												
quota	-0.050182619	0.016308	-0.101712684	0.278967593	-0.190711	0.053218	0.110471824	-0.02985	-0.012832	0.054873	0.241177	0.132088	1											
payment	0.055107633	-0.131198	-0.226230085	0.106344617	-0.432103	-0.120881	0.143201012	0.017837	0.024463	0.088406	0.161293	0.014329	0.886523	1										
totpmt	-0.517640833	0.670297	-0.173755092	0.953657032	0.246914	0.307935	0.061681362	-0.09867	-0.169597	-0.09285	0.759967	0.67282	0.486487	0.299798	1									
accpmt	-0.404856063	0.96091	0.098327632	0.712613458	0.316037	0.250764	0.002596014	-0.10345	-0.228059	-0.286325	0.55141	0.839922	0.196341	0.05298	0.746621	1								
uprice	-0.158668177	0.100054	-0.033117809	-0.143494498	0.493642	0.093747	-0.069324892	-0.004847	0.010565	-0.094522	-0.235942	-0.079158	-0.814541	-0.688342	-0.287281	-0.026468	1							
ucprice	-0.315489548	0.172952	-0.189658803	0.273484511	0.023922	0.181382	0.036869872	-0.100877	0.012922	-0.746604	0.029956	0.054521	0.320294	0.27945	0.319855	0.228331	-0.177043	1						
ngenre	-0.245771555	0.453574	0.035588496	0.645316605	0.129114	0.288795	-0.028836934	-0.121266	-0.108748	0.157673	0.76178	0.625402	0.220208	0.080983	0.620329	0.461952	-0.176215	0.10313	1					
op	-0.210549585	0.179793	-0.117707331	0.306507512	-0.135679	0.1115	0.05118934	-0.122475	-0.090095	-0.475747	0.115629	0.133048	0.852151	0.734614	0.477212	0.323018	-0.668	0.674904	0.111349	1				
lead_op	-0.301506579	0.375338	-0.044828194	0.393342431	0.033811	0.188196	0.045630999	-0.116306	-0.087742	-0.283726	0.180809	0.298451	0.650467	0.509152	0.527899	0.493846	-0.436143	0.205731	0.181625	0.721699	1			
stay	0.047823556	-0.112703	-0.010496951	0.094904927	0.335524	0.00886	-0.004744333	0.060994	0.062293	0.326223	0.131404	0.027132	-0.00841	-0.032907	0.09315	-0.09423	0.028249	#DIV/0!	0.129922	-0.178377	-0.229225	1		
defect	-0.750607543	0.275245	-0.401641849	0.391602104	0.143174	0.359313	0.062962725	-0.223542	-0.10562	-0.398751	0.037954	0.105152	0.035526	0.027559	0.344424	0.263349	0.103874	0.259172	0.002524	0.240275	0.257837	-0.152447	1	
accept	-0.757281669	0.277941	-0.39765399	0.395436981	0.14852	0.369485	0.061102305	-0.224696	-0.105928	-0.399031	0.039143	0.107002	0.035634	0.025143	0.347638	0.266007	0.105048	0.26055	0.006988	0.240516	0.259419	-0.151443	0.999926	1

Figure 6: Customer choice driver variables descriptive statistics

month	
Mean	5.065060241
Standard Error	0.269204393
Median	3
Mode	0
Standard Deviation	5.484110419
Sample Variance	30.07546709
Kurtosis	3.366313606
Skewness	1.827809582
Range	28
Minimum	0
Maximum	28
Sum	2102
Count	415

uprice	
Mean	3.282189089
Standard Error	0.028948917
Median	3.117188
Mode	3.117188
Standard Deviation	0.589734271
Sample Variance	0.34778651
Kurtosis	6.239994389
Skewness	1.035708114
Range	5.667187
Minimum	1.557813
Maximum	7.225
Sum	1362.108472
Count	415

Figure 7: One-segment Binary Logit Model

Uprice	Consumption	Base tendency of Stay									
b0	b1	b2	% of base tendency								
0.204856	12.84373456	0.355802407	0.588023937								
				Prop 1's	0.89879518		Likelihood Ratio Index			# of parameters	3
MLE:	-39.75247551			Rest. MLE	-59.0660957		LRI	0.326983		AIC	85.50495
U	p[stay=0]	p[stay=1]	actual likelihood	log(likelihood)	pred_retain	correct pre	Hit Rate				
13.93094	8.90986E-07	0.999999109	0.999999109	4.30404E-05	1	1	0.898795		One Segment Model		
52.46214	1.6443E-23	1	1	4.34273E-05	1	1					
26.77467	2.35455E-12	1	1	4.34273E-05	1	1					
26.77467	2.35455E-12	1	1	4.34273E-05	1	1					
26.68185	2.58358E-12	1	1	4.34273E-05	1	1					
0.994376	0.270048591	0.729951409	0.729951409	-0.136646557	1	1					
0.994376	0.270048591	0.729951409	0.270048591	-0.568397293	1	0					
26.96	1.95623E-12	1	1	4.34273E-05	1	1					
1.272532	0.2188242	0.7811758	0.7811758	-0.107195628	1	1					
1.272532	0.2188242	0.7811758	0.7811758	-0.107195628	1	1					

Figure 8: Two-segment Binary Logit Model

	Uprice	consumption	base tendency to stay														
util	b1	b2	b3		lambda	q1=P(Seg1)		Prop 1's	0.898795		LRI			# of Parameters:	7		
seg1:	0.316651556	5.574419422	-0.03657		6.993389	0.000917		Rest. MLE	-59.0661		0.326558			AIC	93.55513		
seg2:	0.204798976	10.80435752	0.355777														
MLE:	-39.77756722																
pred-U1	p1	p1-cum	p1-final	pred-U2	p2	p2-cum	p2-final	p-actual	log(p-actua	Revised P(	Revised P(	Pred :	p(retain)	p(leave)	pred_retair	correct_pre	Hit Rate
6.668393	0.998731173	0.998731173	1	11.89133	0.999993	0.999993	1	1	0	0.000916	0.999084	2	0.999992	8.00537E-06	1	1	0.89879
23.39165	1	0.998731173	1	44.3044	1	0.999993	1	1	0	0.000915	0.999085	2	1	6.35048E-14	1	1	
12.24281	0.99999518	0.99872636	1	22.69569	1	0.999993	1	1	0	0.000914	0.999086	2	1	4.54228E-09	1	1	
12.24281	0.99999518	0.998721546	1	22.69569	1	0.999993	1	1	0	0.000912	0.999088	2	1	4.53669E-09	1	1	
12.09933	0.999994437	0.99871599	1	22.60289	1	0.999993	1	1	0	0.000911	0.999089	2	1	5.22223E-09	1	1	
0.950491	0.721213893	0.720287848	1	0.994173	0.729911	0.729906	1	1	0	0.000899	0.999101	2	0.729904	0.270096348	1	1	
0.950491	0.721213893	0.200806245	1	0.994173	0.729911	0.197139	1	1	0	0.000916	0.999084	2	0.729904	0.270096493	1	0	
12.52928	0.999996381	0.200805518	1	22.88097	1	0.197139	1	1	0	0.000933	0.999067	2	1	3.4922E-09	1	1	
1.380444	0.79906233	0.160456125	1	1.272252	0.781128	0.153991	1	1	0	0.000972	0.999028	2	0.781145	0.218854558	1	1	
1.380444	0.79906233	0.128214445	1	1.272252	0.781128	0.120287	1	1	0	0.001036	0.998964	2	0.781147	0.21885341	1	1	
1.380444	0.79906233	0.102451333	1	1.272252	0.781128	0.093959	1	1	0	0.00113	0.99887	2	0.781148	0.218851733	1	1	
1.380444	0.79906233	0.081865001	1	1.272252	0.781128	0.073394	1	1	0	0.00126	0.99874	2	0.781151	0.218849397	1	1	
1.380444	0.79906233	0.065415239	1	1.272252	0.781128	0.05733	1	1	0	0.001437	0.998563	2	0.781154	0.218846215	1	1	
1.380444	0.79906233	0.052270853	1	1.272252	0.781128	0.044782	1	1	0	0.001677	0.998323	2	0.781158	0.218841912	1	1	
1.380444	0.79906233	0.010503183	0.010503	1.272252	0.781128	0.009802	0.009802	0.009802	-2.00867	0.001797	0.998203	2	0.78116	0.218839763	1	0	
45.15152	1	1	1	87.17399	1	1	1	1	0	0.000917	0.999083	2	1	0	1	1	

Figure 9: Hit Rate, LRI, AIC of Two-Segment Binary Logit Model

	LRI			# of Parameters:	7			
	0.326558			AIC	93.55513			
P(	Revised P(	Pred :	p(retain)	p(leave)	pred_retair	correct_pre	Hit Rate	
L6	0.999084	2	0.999992	8.00537E-06	1	1	0.898795	
L5	0.999085	2	1	6.35048E-14	1	1		
L4	0.999086	2	1	4.54228E-09	1	1		
L2	0.999088	2	1	4.53669E-09	1	1		
L1	0.999089	2	1	5.22223E-09	1	1		

Figure 10: One-segment and Two-Segment model performance comparison

Likelihood Ratio Index			# of parameters	3
LRI	0.326983		AIC	85.50495
Hit Rate				
0.898795		One Segment Model		

LRI			# of Parameters:	7			
0.326558			AIC	93.55513			
			Two Segment Model				
Revised P(	Pred :	p(retain)	p(leave)	pred_retair	correct_pre	Hit Rate	
0.999084	2	0.999992	8.00537E-06	1	1	0.898795	