Paper BI-06-2014

Analyzing Collection Effectiveness using Incremental Response Modeling

Ryan Burton, CAPITAL Services, Sioux Falls, SD

ABSTRACT

Incremental response modeling (IRM) is commonly used to optimize direct marking campaigns. Traditionally marking campaigns are optimized by targeting customers most likely to respond without considering the incremental effect the campaign had. IRM targets those most likely to respond favorably to the campaign using a randomly split control and test group. SAS® Enterprise Miner™ now includes an IRM node in production status with version 12.1 or greater. This paper is divided into three sections including the applications and benefits of IRM, basic theory behind SAS's IRM node, and a case study analyzing collection effectiveness with IRM. Some applications of IRM are optimizing marketing actions, collection efforts, and personalized medicine. Generally IRM can be applied to maximize a desired response by applying a treatment to a scored population. By only applying the treatment to the most positively influential population, costs of the treatment can be minimized while returns maximized. An IRM application in collection effectiveness showed collection costs can be decreased without impacting revenue. By segmenting on independent variables like risk score and account age, a population was identified that was negatively influenced by early collection efforts. By focusing collection calls on customer segments positively influenced, collection effectiveness is maximized. The intended audience for this paper is anyone interested in maximizing the positive effect of a customer treatment that has a basic understanding of predictive modeling.

INTRODUCTION

Many companies have discovered the power of randomized testing. Few companies have extended this concept to leverage the complete value of their test and control data. The intuitive next step of calling a test a win or loss is to look at the effect of the test on subpopulations. The test may have won in one segment, but lost in another. With this information, a company could target those who were positively influenced by the test. IRM goes further to answer the question of who is positively or negatively influenced by a test by systematically modeling the impact. The increased flexibility of IRM results in better defined segments then if one were to manually search for differences in test effectiveness.

Traditionally campaign models target those most likely to respond to a campaign offer. This type of modeling does not accurately optimize the campaign efforts. The goal of any campaign is to make a positive difference by increasing responders by a campaign. IRM utilizes this goal more directly by categorizing those more likely to respond if they receive a campaign offer versus if they did not. Modeling the incremental response is intuitively superior but also adds complexity to the modeling process. Candidates for a campaign offer can be divided into three useful groups including those who have a higher response rate when in the campaign, those that are not influenced by the campaign, and those who have a lower response rate when in the campaign. IRM targets the first group, sometimes called the persuadables. Assuming there is some cost of a campaign, those not influenced and those negatively influenced decrease overall profit when included in the campaign. Therefore the power to classify candidates into these groups results in a more profitable strategy.

APPLICATIONS

IRM can be used to optimize marketing actions, personalize medicine, and increase collections effectiveness. Lee, Meng, and Ryan (2013) explain how IRM can be used to enhance direct marketing campaigns using SAS® Enterprise Miner™. In addition to the IRM, they explain how to make an incremental sales model. Kubiak (2012) showed lift in a direct marketing campaign with 1800 flowers by manually developing a probability decomposition IRM. They showed true lift in the response rate of direct mailing a floral marketing offer. Another example is the 2012 Obama campaign where voters likely to be positively influenced by a campaign were targeted (Siegel, 2013). They also modeled what message type and form of contact were best on an individual level. After collecting data by randomly testing marketing campaigns on swing states, they built IRM models and were able to persuade more voters than a traditional marketing campaign. There are many applications and benefits of IRM still being explored.

BASIC THEORY

IRM has three basic data requirements for the technique to be successful.

- Randomly selected treated and control groups
- A dependent variable, binary or continuous, that measures the outcome ("response", "dollar amount of response", etc.)

- Predictor variables that can differentiate the rate or amount of response between the treated and control groups.

Past campaign strategies have used models to target the probability of someone responding to an offer using only test data. This method would target those with the highest probability to respond

$$P(respond) = f(\mathbf{X}\beta)$$

The flaw in this method is that those who would have responded with the test treatment are still being targeted. IRM targets those with the highest incremental probability to respond given they're in the test versus control. A challenge of the IRM method is that the incremental probability cannot be measured directly because it is impossible to observe the actual incremental effect on a customer level. Instead the effect is estimated on a grouped customer level.

There are multiple techniques to estimate the incremental propensity for someone to be positively influenced by the treatment. SAS® Enterprise Miner™ uses the difference score technique. The difference score is defined as

$$Difference\ score = P_T(respond) - P_C(respond) = f_T(\mathbf{X}\beta) - f_C(\mathbf{X}\beta)$$

where p_T is the probability in the treated population and p_C is the probability in the control population. Targeting those with the highest difference score results in an optimal strategy.

The difference score technique starts by modeling the probability to respond given the subject was tested by creating a model using the test data. Independently, the technique also models the probability to respond given the subject was not tested by creating a model using the control data. Next the control and test data are combined and each subject is scored with both scores. The difference of these scores measures the relative incremental probability for a subject to respond given they are tested versus not tested. The main advantage of this technique is its simplicity and intuitiveness. A con to the difference score technique is that the incremental probability is being estimated indirectly. SAS protects against this fault by providing good variable selection methods.

A standard credit scoring variable importance metric is the information value (*IV*) created using the weight of evidence (*WOE*) of binned attributes. Kullback (1959, p. 5) defines the *WOE* for group i as

$$WOE_i = \log\left(\frac{p(i \mid y = 1)}{p(i \mid y = 0)}\right)$$

where $p(i \mid \mathrm{y} = 1)$ is the probability[1] that an observation is in the i^{th} group given that the observation has a dependent variable equal to one. Similarly $p(i \mid \mathrm{y} = 0)$ is the probability that an observation is in the i^{th} group given that the observation has a dependent variable equal to zero. Therefore if WOE_i is positive, there is a higher probability of being in group i given the dependent variable is one compared to the probability of being in group i given the dependent variable is zero. WOE_i is negative if the complement is true. The information value is calculated for a complete independent variable estimating the value of the variable to predict the dependent variable. The information value is defined as

$$IV = \sum_{i=1}^{k}[(p(i \mid y = 1) - p(i \mid y = 0)) * WOE_i]$$

Thus an independent variable with a higher *IV* indicates the variable separates the target distribution better and shares more information with the dependent variable. Extending *WOE* and *IV* to a net lift application, one can use net weight of evidence (*NWOE*) and net information value (*NIV*), defined by Larsen (2010) as

$$NWOE = \log\left(\frac{p_T(i \mid y = 1)}{p_T(i \mid y = 0)} \div \frac{p_C(i \mid y = 1)}{p_C(i \mid y = 0)}\right)$$

and

$$NIV = \sum_{i=1}^{k}[(p_T(i \mid y = 1)p_C(i \mid y = 0) - p_T(i \mid y = 0)p_C(i \mid y = 1)) * NWOE_i]\,.$$

Therefore, the *NIV* can rank order variables according to their ability to predict incremental responders. There is also an option in SAS to compute a penalized net information value which considers the generalizability of the variable importance to a hold out data set. Another extension of this metric would be to consider the incremental importance of each variable considering multicollinearity between independent variables. Once the correct variables are selected,

[1] Precisely, $p(\mathrm{i} \mid \mathrm{y} = 1)$ is the percentage of "i" observations among all observations where y = 1.

the risk of overfitting the model decreases. Table 1 shows an example calculation of the *NWOE* and *NIV* for a hypothetical variable.

Table 1: Net information value example

Variable 1	Treatment Count	Treatment Response Rate	No Treatment Count	No Treatment Response Rate	Incremental Response Rate	NWOE	NIV
Attribute 1	500	20.0%	500	30.0%	-10.0%	-0.7903	0.0658
Attribute 2	500	30.0%	500	25.0%	5.0%	0.0000	0.0000
Attribute 3	500	40.0%	500	20.0%	20.0%	0.7295	0.0597
Total	1500	30.0%	1500	25.0%	5.0%	-	0.1255

SAS also has the capability to estimate the incremental outcome given response. This technique is similar to the difference score technique except a continuous variable is being estimated instead of a binary.

COLLECTIONS EFFECTIVENESS CASE STUDY

IRM can be directly applied to optimize collection calls. Delinquent accounts in collections can be divided into three key groups. Persuadables are more likely to make a payment when called. The neutral group is not influenced by collection calls. They will pay or not pay regardless if they receive a call. The final group is sometimes labeled do not disturb because they are less likely to make a payment if called. Since there is an associated cost with making a collection call, call centers ideally want to only target the persuadables.

The sample population for the case study includes credit card accounts that have recently gone delinquent. The specific data is from a test intended to decrease collection costs without impacting payment dollars. Test and control groups were randomly assigned to measure the effect of delaying collection efforts. The control group accounts were called the first day they went delinquent, and the test group accounts were called six days after they went delinquent. The performance metric was the amount collected in each group within the first month. The initial results showed that by delaying the collection calls by six days, the amount collected was significantly decreased, and the reduction of collection costs from not calling did not offset this amount. Thus, the control group was called the winner of the test. IRM dissects the test further to determine if there are segments of the population where the test group won.

To format the data for IRM, independent variables were appended, a binary treatment indicator of test versus control was created, and a binary paid versus not paid was created. The treatment was labeled a one if they were called on the first day and a zero if the call was delayed to the sixth day of delinquency. The independent variables included behavioral information of the credit card account like risk score, purchase history, payments history, and contact history. It is critical to be able to capture the non-linear effects the independent variables. The IRM node gains this flexibility by grouping the independent variables into binned categorical variables. This enables an appropriate risk score relationship to be modeled. For example those with extremely low risk scores are likely to not have the ability to pay and are not going to be influenced by collection calls. Those with high risk scores are delinquent most likely because they forgot to pay and will pay soon on their own. Those with medium risk scores have the highest propensity to be pay only when they are called and should be targeted with early collections. Table 2 shows an example of this relationship.

Table 2: Payment rate by risk score

	Payment Rate		
Risk Score Variable	**Early Collections**	**Delayed Collections**	**Incremental Payment Rate**
Low	82.3%	82.2%	0.1%
Medium	87.7%	85.1%	2.6%
High	90.3%	90.8%	-0.5%
Total	86.8%	86.0%	0.7%

By targeting just the medium group instead of the entire population with early collections, the incremental payment rate changes from 0.7% to 3.0%. The IRM will target the medium risk group by estimating a higher incremental response to the medium risk group.

Before training the model, the data was randomly partitioned into training and validation data sets using the SAS data partition node. Initial variable selection was done by calculating the NIV on all independent variables and excluding those past a rank percentage cutoff. Next, two logistic regression models are trained using the control data and the test data separately. Once the models were created, they were applied to the entire sample population so every account had a likelihood to pay estimation given they were called early versus if their call was delayed. Finally, the differences of the estimations were calculated to estimate the incremental expected likelihood to pay given they were collected on early. Figure 1 shows the SAS® Enterprise Miner™ project diagram used to make the IRM.

Figure 1: Project Diagram

The first node imports the data and assigns the correct variable roles. The second node randomly partitions the data into training and validation data sets. The incremental response node does the variable selection, and creates the IRM. The final SAS code node is used to export the model output for additional analysis. Figure 2 shows the properties used in the incremental response node.

Figure 2: Incremental response properties

Property	Value
General	
Node ID	IRM
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Prescreen Variables	Yes
Rank Percentage Cutoff	50
Model Selection	
Combined Model	No
Selection Method	Stepwise Selection
Selection Criterion	AIC
Significance Level for Entry	0.1
Stay Significance Level	0.1
Suppress Intercept	No
Two-Way Interaction	Yes
Revenue Calculation	
Use Constant Revenue	Yes
Revenue Per Response	10.0
Cost	0.0
Report	
Number of Bins	5
Status	
Create Time	6/8/14 9:19 AM
Run ID	f2364b9c-4436-45c2-a25d-95099f1a
Last Error	
Last Status	Complete
Last Run Time	6/8/14 9:31 AM
Run Duration	6 Hr. 49 Min. 57.64 Sec.
Grid Host	
User-Added Node	No

A rank percentage cutoff of 50 is used to select the best 50% of the variables according to *NIV*. Stepwise selection is used as a secondary variable selection with an Akaike information criterion (AIC) cutoff of 0.1 for entry and stay. A constant revenue of $10 was used. The number of bins in the report was set to five. Table 3 shows the likelihood to pay output of the IRM model for 10 sample accounts.

Table 3: Account level estimation of payment

dataobs	EM_P_TREATMENT_RESPONSE	EM_P_CONTROL_RESPONSE	EM_P_INCREMENT_RESPONSE
AccountID	P(Payment | Early Collections)	P(Payment | Delayed Collections)	Incremental Payment
1	0.9999	1.0000	-0.0001
2	0.5488	0.7753	-0.2266
3	0.9951	0.9380	0.0571
4	0.9865	0.9920	-0.0055
5	0.9956	0.9678	0.0278
6	0.9986	0.9800	0.0186
7	0.9478	0.9949	-0.0471
8	0.8997	0.9916	-0.0919
9	0.9982	0.9464	0.0519
10	0.9316	0.9859	-0.0542

The incremental likelihood to pay is calculated by subtracting the probability of payment given delayed collections from the probability of payment given early collections. Those with a negative incremental payment should not be targeted with early collections. Table 4 shows the incremental revenue output.

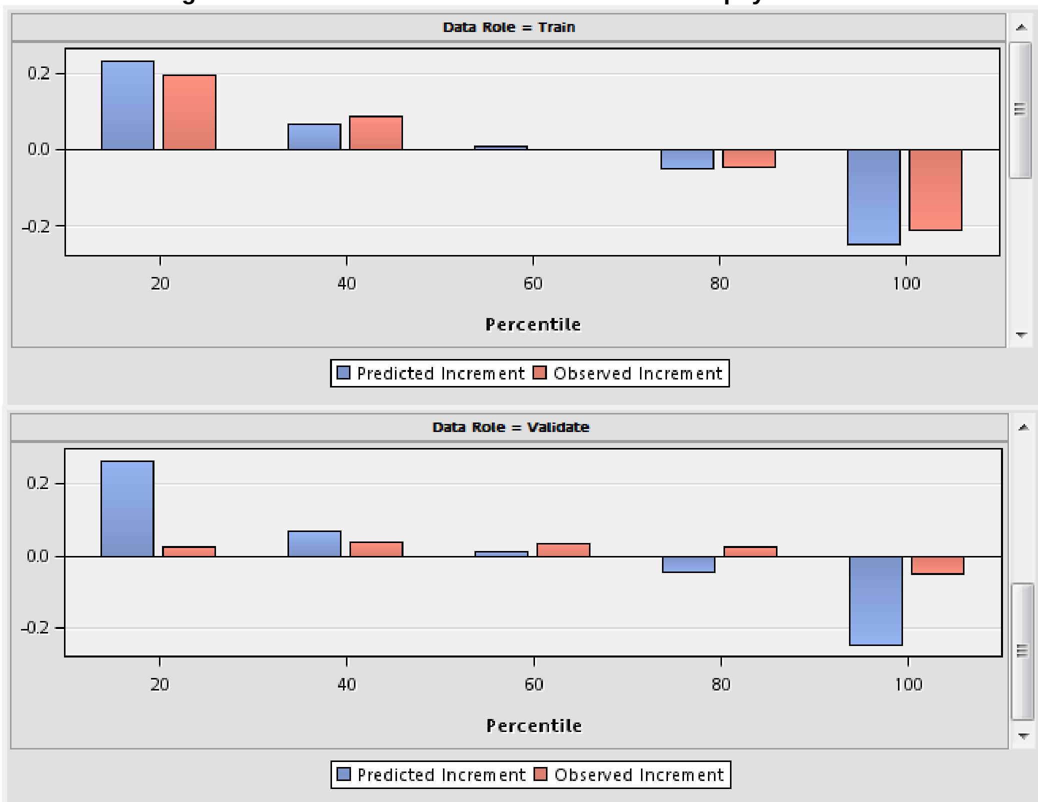
Table 4: Account level estimation of payment

dataobs	EM_REV_TREATMENT	EM_REV_CONTROL	EM_REV_INCREMENT
AccountID	Revenue | Early Collections	Revenue | Dealyed Collections	Incremental Revenue
1	9.9986	10.0000	-0.0014
2	5.4876	7.7534	-2.2658
3	9.9507	9.3800	0.5707
4	9.8651	9.9203	-0.0552
5	9.9560	9.6776	0.2784
6	9.9857	9.7996	0.1861
7	9.4777	9.9489	-0.4712
8	8.9969	9.9156	-0.9187
9	9.9823	9.4637	0.5186
10	9.3164	9.8587	-0.5423

The incremental revenue is calculated analogously assuming revenue given a payment of $10.

Figure 3 shows the predicted versus observed incremental payment rates within the training and validation datasets grouped into five evenly split groups ranked using the IRM by likelihood to be positively influenced by early collections.

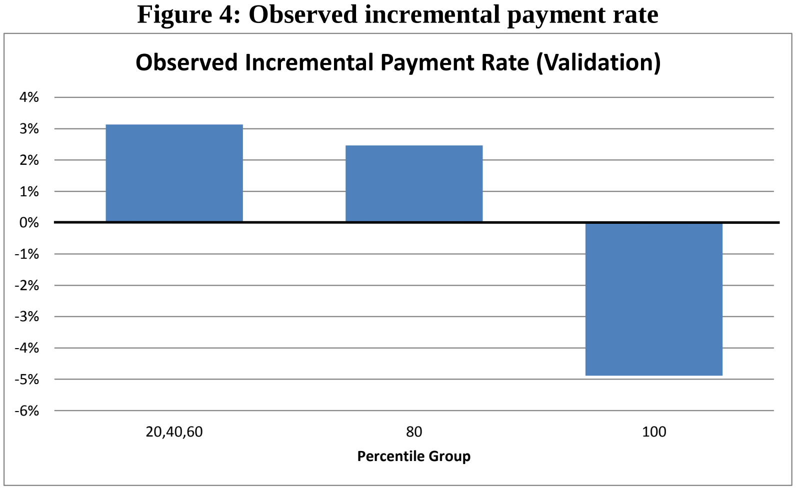
Figure 3: Predicted versus observed incremental payment rates



For example, the 100th percentile group has a 5% lower payment rate if they are called right away according to the validation dataset. On average calling this group will result in fewer payments. All other groups are positively influenced by calling on the first day of delinquency according to the validation dataset. The observed increments of the validation data set are not as dramatic as the training data set. More data would help stabilize these estimations. Because of the indirect modeling of incremental response, one must be cautious when generalizing.

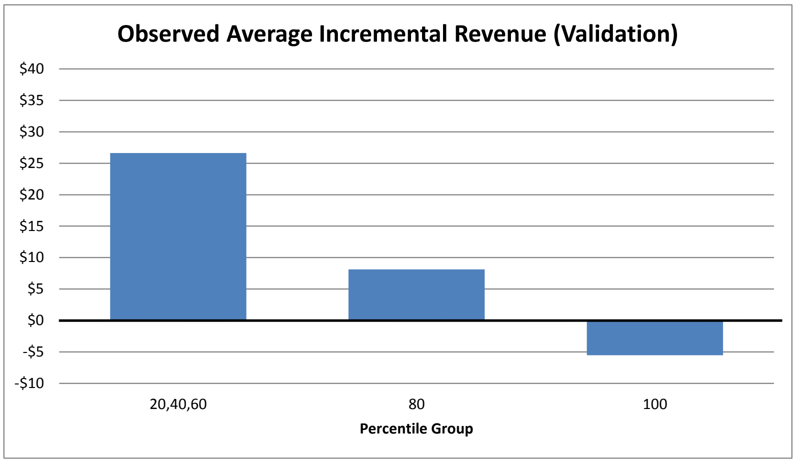
To stabilize the results, percentile 20, 40 and 60 were grouped together. Figure 4 shows the binned observed incremental payment rate of the grouped percentiles for the validation data set. With more volume in each group, the incremental payment rate rank orders as expected.

Figure 4: Observed incremental payment rate



Using the validation data set, the observed average incremental payment dollars in each group is shown in Figure 5.

Figure 5: Observed average incremental revenue



The cost to collect early per account was estimated to be around two dollars. Therefore, a positive profit is gained by calling the first two groups and around seven dollars per account is lost when considering lost revenue and calling expenses by calling the third group. The third group responds negatively to collection calls within the first 6 days of delinquency. By discontinuing targeting this group, collection costs will decrease and total payments will increase.

CONCLUSION

Many companies have mastered random testing but are not fully leveraging their valuable data. IRM is a good tool to systematically split the population into groups from most likely to least likely to be positively influenced by an action. Once the population is ranked, the action can be optimized. Applying IRM to collections, the case study showed value by delaying collection efforts on a group likely to be negatively influenced. The application of IRM results in decreased collection expenses and an increase in profit. As IRM techniques are advanced and more applications are discovered, IRM will continue to grow as a popular modeling application.

REFERENCES

- Kubiak, R. (2012). “Net Lift Model for Effective Direct Marketing Campaigns at 1800flowers.com”, SAS Global Forum 2012, Paper 108-2012.
- Kullback, S. (1959). Information Theory and Statistics. New York: John Wiley & Sons.
- Larsen, K. (2010). “Net Lift Models: Optimizing the Impact of Your Marketing Efforts.” SAS Course Notes. Cary, NC: SAS Institute Inc.
- Lee, T., Zhang, R., Meng, X., & Ryan, L. (2013). “Incremental Response Modeling Using SAS® Enterprise Miner™”, SAS Global Forum 2013, Paper 096-2013.
- Siegel, E. (2013). “The Real Story Behind Obama’s Election Victory”, The Fiscal Times.

CONTACT INFORMATION

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

Name: Ryan Burton
Enterprise: CAPITAL Services
Address: 500 E. 60th Street North
City, State ZIP: Sioux Falls, SD 57104
Work Phone: 605-782-3381
E-mail: ryan.burton@capitalsvcs.com
Web: www.capitalsvcs.com