Net Present Value Model Approach

By Jingjing Qu
Yumin Chen
Transamerica Life Insurance, Marketing Analytics

Contents

Introduction

The modeling approach in direct marketing channels is often debated. *Can an effective model predict beyond the initial response to an offer a sufficient degree that would positively impact revenue?* This white paper presents the business value as well as the methodology for an alternative approach to response modeling for traditional insurance direct marketing products. The paper covers the data requirements and statistical procedures, as well as the benefit of Net Present Value (NPV) models compared to a response model or a RIP (response/issue/paid) model. The statistical methods presented here are from a technical perspective and concentrate on developing a NPV model. The focus is to provide a means of helping statisticians build and evaluate an NPV model.

Problem Statement -- Benefit of NVP Model

A special feature of an insurance product is that its sale generates a sequence of premium payments up to an unknown future time. It is unknown because the buyer can stop paying the premium any time before the expiration of contract. Unlike the sale of consumption goods, the revenue of an insurance product at the time of sale has to include the assessment of the future potential premium payments. Currently, modeling practices focus either only on initial response or only on initial payment. The buyer's ability to make future payments is not assessed and explicitly considered. This not only leads to inaccurate computation of revenue, but also misdirects marketing resources simply because this approach can not distinguish loyal customers from transient customers.

The NPV model extends the current approach by explicitly modeling the attrition probability of the buyers. That is, in addition to predicting the possibility of initial payment, it also predicts the probability of premium payments over time. The expected revenue from the sale of the product is the net present value of the expected premium payments over time, which is calculated from the predicted probability.

Data Requirement

Before predicting a customer's intention to pay future premiums, or a customer's attrition probability, it is required to know a customer has the desire not only purchase but pay. A paid response model will serve the purpose of predicting probability to pay. Past campaign data with response and paid information are the sources for building response and paid models. To avoid bias, a random sample of the data is desirable. Bias reduction must be applied to the data if a random sample is not obtained.

Policy owner data contains the information necessary for modeling attrition probability. This data allows us to analyze the duration of policy owners.

Statistical Method

Logistic regression will be used to build response and paid models due to the binary nature of the possible outcomes: buy or not buy and pay or not pay. A decision tree model is another option. Both approaches are widely adopted and available in SAS and most existing statistical packages.

The statistical method of survival analysis is commonly adopted in analyzing the duration data. As the name suggests, it is the study of time between entry into observation and a subsequent event. Originally, the event of interest was death hence the term "survival analysis." The analysis consisted of following the subject until death. In the current application, it includes time until a buyer stops paying premiums.

Many procedures are available in SAS for doing survival analysis. We will choose Cox's proportional hazards regression to explain the effect of covariates on time until event. Reasons for adoption of this method are: the relative risk involved in this technique, no parametric assumptions are required, a partial likelihood function is utilized, and the creation of survival function estimates.

Relative Risk

The simplest interpretation given by the Cox model as "relative risk" type ratio is very desirable in explaining the risk of event for a certain covariate. For example, when we have a two level covariate with a value of 0 or 1, the hazard ratio becomes $\exp(\beta)$. If the value of the coefficient is $\beta=\ln(3)$, then it is simply saying that the subjects labeled with a 1 are three times more likely to experience an event than the subjects labeled with a 0. In this way we have a measure of difference between the exposure cohorts instead of simply knowing whether the two cohorts were statistically different.

No Parametric Assumptions

Another attractive feature of Cox regression is there is no need to choose the density function of a parametric distribution. This means that Cox's semi parametric modeling allows for no assumptions to be made about the parametric distribution of the survival times, making the method considerably more robust. Instead, the researcher has to validate the assumption that the hazards are proportional over time. The proportional hazards assumption refers to the fact that the hazard functions are multiplicatively related. That is, their ratio is assumed to be constant over the survival time.

Use of Partial Likelihood Function

The Cox model has the flexibility to introduce time dependent explanatory variables and to handle censoring of survival times due to its use of the partial likelihood function. This was important to the study in that any temporal biases of the covariates over the time of study needed to be handled correctly.

Survival Function Estimates

With the SAS option BASELINE, a SAS dataset containing survival function estimates can be created and output. These estimates correspond to the means of the explanatory variables for each stratum.

Case Study—NPV Model vs. SPC Model

This case study used an example to show the NPV model approach and benefits. A sales per contact (SPC) model was also built using the same data to compare with the NPV model with respect to important marketing performance indicators. Sales here are defined as non paid responses to telemarketing contacts.

The results show that compared to the SPC model, the NPV model selects smaller number of the target variable (sales), but the selected targets have a higher persistency. As a result, the NPV model achieves higher revenue over time. The NPV model also has smaller number of sales who do not pay (i.e., no pay sales), so it saves cost by precisely targeting leads who are not only willing to purchase but willing to pay.

Data

- Paid, response, and SPC modeling file contains contacted names with completed paid information.
- Persistency modeling file contains years of paid policies. Paid policies were defined as policies where at least one month's premium was paid.
- Enhanced data was appended to the modeling file, plus age, gender and geographic region information were used as model predictors.

Statistical Approach

Two models were developed for NPV model:

Model 1- Paid response model: A multinomial logistic regression model which predicts the probabilities of not paying, paying individual coverage, and paying family coverage. The sum of the probabilities of paying individual coverage and paying family coverage is defined as the probability of paying. In SAS, a multinomial logistic regression model can be implemented by utilizing the proc logistic command with the glogit option.

Model 2 - Persistency model: A Cox proportional hazard model which utilizes paid policies to predict the probabilities of paying 1-36 months. In SAS, a Cox proportional hazard model can be implemented by proc phreg.

The NPV is the sum of

$$\frac{R_t}{(1+i)^t}$$

over t, where t is the time of the cash flow and i is the discount rate. 0.12 is used as the annual discount rate, while the equivalent monthly discount rate is 0.009489.

R_t is a product's normal monthly premium at time t, and is calculated as:

*R_t = normal monthly premium*probability of paying*probability of paying at time t*

Logistic regression was used to develop a SPC model to compare with the NPV model.

NPV Model vs. SPC Model Slice Reports

NPV Model Marginal

Decile	Sale	SPC	Paid/Sale	Paid/Contact	Paid Mons.	Paid Mon./Paid	Realized NPV	Total Expected 36m NPV
1	4739	9.1%	64.3%	5.9%	34516	11.3	449598	559902
2	4231	8.2%	63.6%	5.2%	30921	11.5	409123	469211
3	4157	8.0%	62.8%	5.0%	30034	11.5	401605	416262
4	3841	7.4%	61.0%	4.5%	25571	10.9	352589	369768
5	3698	7.1%	57.6%	4.1%	21496	10.1	302660	323730
6	3870	7.5%	53.3%	4.0%	19108	9.3	274954	277714
7	3564	6.9%	50.5%	3.5%	15080	8.4	221790	238268
8	2551	4.9%	49.3%	2.4%	9048	7.2	137341	197095
9	2052	4.0%	49.1%	1.9%	6273	6.2	99947	157818
10	1763	3.4%	47.7%	1.6%	5195	6.2	82038	122428

NPV Model Cumulative

Decile	Sale	SPC	Paid/Sale	Paid/Contact	Paid Mons.	Paid Mon./Paid	Realized NPV	Total Expected 36m NPV
1	4739	9.1%	64.3%	5.9%	34516	11.3	449598	559902
2	8970	8.6%	63.9%	5.5%	65437	11.4	858721	1029113
3	13127	8.4%	63.6%	5.4%	95471	11.4	1260326	1445375
4	16968	8.2%	63.0%	5.2%	121042	11.3	1612915	1815143
5	20666	8.0%	62.0%	4.9%	142538	11.1	1915575	2138873
6	24536	7.9%	60.6%	4.8%	161646	10.9	2190529	2416587
7	28100	7.7%	59.4%	4.6%	176726	10.6	2412319	2654855
8	30651	7.4%	58.5%	4.3%	185774	10.4	2549660	2851950
9	32703	7.0%	57.9%	4.1%	192047	10.1	2649607	3009768
10	34466	6.6%	57.4%	3.8%	197242	10.0	2731645	3132196

SPC Model Marginal

Decile	Sale	SPC	Paid/Sale	Paid/Contact	Paid Mons.	Paid Mon./Paid	Realized NPV	Total Expected 36m NPV
1	5302	10.2%	59.6%	6.1%	31317	9.9	421784	438359
2	4626	8.9%	59.1%	5.3%	28516	10.4	390560	414552
3	4220	8.1%	60.1%	4.9%	27117	10.7	370818	390992
4	3960	7.6%	58.8%	4.5%	25569	11.0	351167	370212
5	3777	7.3%	57.5%	4.2%	23398	10.8	319211	351830
6	3378	6.5%	59.7%	3.9%	21266	10.5	295749	334659
7	3129	6.0%	56.8%	3.4%	18429	10.4	254867	306421
8	2606	5.0%	53.3%	2.7%	10578	7.6	154594	216329
9	1916	3.7%	49.1%	1.8%	6191	6.6	96596	165376
10	1552	3.0%	47.5%	1.4%	4861	6.6	76299	143464

SPC Model Cumulative

Decile	Sale	SPC	Paid/Sale	Paid/Contact	Paid Mons.	Paid Mon./Paid	Realized NPV	Total Expected 36m NPV
1	5302	10.2%	59.6%	6.1%	31317	9.9	421784	438359
2	9928	9.6%	59.4%	5.7%	59833	10.2	812344	852911
3	14148	9.1%	59.6%	5.4%	86950	10.3	1183162	1243903
4	18108	8.7%	59.4%	5.2%	112519	10.5	1534329	1614115
5	21885	8.4%	59.1%	5.0%	135917	10.5	1853540	1965945
6	25263	8.1%	59.2%	4.8%	157183	10.5	2149289	2300604
7	28392	7.8%	58.9%	4.6%	175612	10.5	2404156	2607025
8	30998	7.5%	58.4%	4.4%	186190	10.3	2558750	2823354
9	32914	7.0%	57.9%	4.1%	192381	10.1	2655346	2988730
10	34466	6.6%	57.4%	3.8%	197242	10.0	2731645	3132194

Scenario Selecting Top 40% of File

The table to the right shows that the selection in the NPV model has:

- Lower SPC rate but higher paid rate;
- Lower number of sales and paids, but also a lower number of no pay sales. It saves cost by minimizing no pay sales;
- Higher total paid months, higher persistency months per paying customer;
- 12% more total expected 36 months' NPV than SPC model selection.

Conclusion

The NPV model extends the capability of the paid response model to include the ability to predict the attrition probability of buyers. As a result, the NPV model allows for targeting the future value of buyers. To achieve this, the NPV model requires more data and analytical tools than the response model.

NPV and SPC Model Compare at Slice 4

Model	Sale	SPC	Paid/Sale	Paid/Contact	Paid Mons.	Paid Mon./Paid	Realized NPV	Total Expected 36m NPV
NPV	16,968	8.2%	63.0%	5.2%	121,042	11.3	1,612,915	1,815,143
SPC	18,108	8.7%	59.4%	5.2%	112,519	10.5	1,534,329	1,614,115
%Difference		94%	106%	100%		108%	105%	112%

Reference

Cox, D. R. and Oakes, D. (1984), *Analysis of Survival Data*, New York: Chapman and Hall

Horowitz, J. L. (1998), *Semiparametric Methods in Econometrics*, New York: Springer-Verlag

Klein, J. P. and Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and truncated data*, New York: Springer Verlag

Appendix: SPC and NPV Lift Charts—NPV Model vs. SPC Model