Estimating Insurance Attrition Using Survival Analysis

Luyang Fu, Ph.D., FCAS

CAS Spring Meeting, May 2015,

The author’s affiliation with The Cincinnati Insurance Company is provided for identification purposes only and is not intended to convey or imply The Cincinnati Insurance Company’s concurrence with or support for the positions, opinions or viewpoints expressed.
Agenda

• Introduction
• Survival Analysis
• Cox Proportional Hazard Model
• A Case Study
• Q&A
Introduction

Attrition/retention is important to insurance companies

Growth (Top Line)

• “We were not successful in raising customer renewal rates, so that the new business success did not result in overall growth this quarter”
  All State 2010Q4 Earnings call

• Higher retention, less pressure of attracting new business

  To grow a book with 10,000 policies by 10%,
  – if retention is 90%, need to attract 2000 new accounts: 1000 to make up attrition, 1000 for the growth
  – if retention is 70%, need to write 4000 new accounts: 3000 to make up attrition, 1000 for the growth
Introduction

Attrition/retention is important to insurance companies

Profitability (Bottom Line)

• “Given our strong retentions as well as the new business and account growth we've achieved over the last few years, we have significant positive leverage to an improving environment.”
  Travelers 2010Q4 Earning call

• Ageing Phenomenon
  – D’Arcy and Doherty (1989; 1990): loss ratio improves with policy age
  – Wu and Lin (2009): renewal book on average has a loss ratio 13% better than new business by examining 8 lines of business, 25 books, $29 billion premium

• Price Optimization
  – Retention, conversion, price elasticity
  – Life-time value
Introduction

Two types of Attritions

• Mid-term cancellation
• End-term nonrenewal

Probability of Attrition: Cancellation vs. Nonrenewal
Two types of attritions behave differently

• Example 1: price elasticity
  – End-term nonrenewal is more sensitive to price change
  – Mid-term cancellation may be from non-pricing reasons
Introduction

Two types of attritions behave differently

- Example 2: policy size in commercial lines
  - Small policies may have a higher mid-term cancellation ratio than large policies
  - Large policies may have a higher end-term nonrenewal ratio

Attrition Ratios by Month: Large vs. Small Commercial Policies
Introduction

Traditional Retention Analysis

• Renewal ratio at expiration month
  – If 1,000 policies expire at May 2013, 920 of them are still with the company at 05/31/2013. The renewal ratio is 92%.
  – The evaluation lag may vary.
  – It ignores the attrition from mid-tern cancellation
  – It does not give an annual view of retention or attrition
Introduction

Traditional Retention Analysis

• Annual Retention: Snapshot comparison
  – If there were 10,000 inforced policies at 12/31/2011, 8,500 of them were still effective at 12/31/2012, the annual retention ratio is 85%.
  – Does not analyze the sources of attritions. 15% is the sum of mid-term cancellation and end-term nonrenewal
Introduction

Traditional Retention Analysis

• Logistics models
  – Data: snap-shot data
  – Variable of interest: yes or no
  – Do not model cancellation and nonrenewal separately (can be extended to model two ways of attritions independently).
  – Static view
Introduction

Why survival analysis?

• Estimate mid-term cancellation and end-term nonrenewal sequentially and simultaneously
  – Survival Analysis:
    • Reflect two ways of attritions through the seasonality within survival curve
    • Recognize the aging sequence of the same policy (panel data approach)
  – Logistics Regression:
    • Snap-shot data cannot separate mid-term and end-term attritions
    • Treat each record within the same policy panel independently
Introduction

Why survival analysis?

• Better estimation of life time value: not just whether a policy will leave, but when it will leave
  – Survival Analysis:
    • Target variable of interest: t (time to attrition)
    • If 10,000 policies are inforce at 12/31/2009, 8,500 of them were still effective after a year. Among 1,500 attritions, how many of them left by cancelation and non-renewal, and when they left?
  – Logistics Regression:
    • Target variable of interest: yes or no
    • Ignore the time of attrition
    • Do not predict the attrition for non-integer multiples of the evaluation horizon
Introduction

Why survival analysis?

• Better utilization of time-varying macroeconomic variables
  – Survival Analysis:
    • Dynamic view of treasury yield, GDP change, and stock market return, etc.
    • Reflect interest rate, inflation, consumer confidence at the time of attrition
  – Logistics Regression:
    • Static view of those variables
    • If “yes” or “no” is constructed by comparing 2011 with 2012 year-end book, one summarized “unemployment rate” is used for all the records
    • Flinn and Heckman (1982): reliance on ad hoc procedures to cope with time-trended variables in logistic regression can produce very pathological estimates
Introduction

The disadvantages of survival analysis

• Model implementation is not as straightforward as binary model
  – Logistic
    • Probability of attrition is the direct output of model
  – Survival analysis
    • Develop baseline survival function
    • Derive hazard function for individual policies
    • Calculate the probability of attrition
Introduction

The disadvantages of survival analysis

• Time-varying macroeconomic variables are more difficult to predict than retention
  – How to capitalize the relationship between retention and time-varying macroeconomic variables
  – The models on interest rates and stock indexes are much more complex than retention models
  – Macroeconomic variables are more volatile than retention, and may introduce additional volatility into retention projection.
Introduction

Literatures on Marketing and Banking


Survival Analysis

• Another name for *time to event* analysis

• Statistical methods for analyzing survival data.

• Primarily developed in the medical and biological sciences (death or failure time analysis)

• Widely used in the social and economic sciences, as well as in Insurance (longevity, time to claim analysis).
Survival Analysis

Survival Time

• $t$ measures the time from a particular starting time (e.g., time initiated the treatment) to a particular endpoint of interest (e.g., patient died).

• Examples:
Survival Analysis

Censoring

• Occurs when the value of a measurement or observation is only partially known.

• Left Censoring:
  Example: Subject's lifetime is known to be less than a certain duration.

• Right Censoring:
  Example: Subjects still active when they are lost to follow-up or when the study ends.
Survival Analysis

Data

• Calendar time of whole study (Starting day, Ending day of the whole study period)
• Study Duration of each individual.
• Define the *censored observations*.
• Time measure units (Month, Year ... )
• Define the dependent variable and independent.
Survival Analysis

Beginning of Study

Calendar Time

End of Study

Duration Time

Origination Time ▶ Entry Time to study  ■ Attrition  ○ Censored
Cox Proportional Hazard Model

Advantages

• The dependent variable of interest (survival/failure time) is most likely not normally distributed.
• Censoring (especially right censoring) of the Data.
• Baseline hazard function is unknown.
• Whether and when the customer will leave.
• Dynamics covariates and duration
Cox Proportional Hazard Model

Hazard equations

\( h(t \mid x_t) \): hazard rate at time \( t \) for an individual have covariate value, \( x_t \)

\[
h(t \mid x_t) = h_0(t)e^{\beta'x_t}
\]

Here \( x_t = (x_{1t}, x_{2t}, \ldots, x_{kt}) \) \( \beta = (\beta_1, \beta_2, \ldots, \beta_k) \)

\( k \) is the total number of the covariates, \( x_j \)

\( \beta_j \) is the constant Proportional effect of

The term \( h_0(t) \) is called the baseline hazard; it is the hazard for the respective individual when there is no covariate impacts.
Cox Proportional Hazard Model

Hazard Equations

We can linearize this model by dividing both sides of the equation by \( h_0(t) \) and then taking the natural logarithm of both sides:

\[
\ln \left\{ \frac{h(t \mid x_t)}{h_0(t)} \right\} = \beta' x_t
\]

Taking partial derivative we have

\[
\frac{\partial \ln h(t \mid x_t, \beta)}{\partial x_{jt}} = \beta_j
\]
Cox Proportional Hazard Model

Literatures on Survival Analysis Theory


Case Study

Data

• Simulated commercial line data.
• Dependent variable:
  duration = the time until the policy leaves
• If a policy is still effective at the end of study, it is right censored (i.e. Censor = 1)
• External data (including macroeconomic data) are joined into policy data.
Case Study

Data

• Define rate changes, removing the impacts from
  — Exposure changes (add a building; cut a class)
  — Coverage changes (reduce limits; increase deductible)
  — Risk characteristics changes (have a violation/claim; add a youthful driver)

• Groupings/binnings can be arbitrary
  — Contractors vs. noncontractors
  — Size groups
  — Variable interactions:
    — Small, medium, large contractors
    — General, nongeneral with sub, artisan contractors
## Annual Attrition Summary

<table>
<thead>
<tr>
<th>Year</th>
<th>Total</th>
<th>Renewed</th>
<th>Non Renewed</th>
<th>Midterm Cancellation</th>
<th>Non Renewal %</th>
<th>Midterm Cancellation %</th>
<th>Retention %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>197,954</td>
<td>156,477</td>
<td>24,570</td>
<td>16,907</td>
<td>12.41%</td>
<td>8.54%</td>
<td>79.05%</td>
</tr>
<tr>
<td>5</td>
<td>211,061</td>
<td>162,875</td>
<td>27,398</td>
<td>20,788</td>
<td>12.98%</td>
<td>9.85%</td>
<td>77.17%</td>
</tr>
</tbody>
</table>

* The data is for illustration purpose.
Case Study

Annual Attrition Summary

Annual attrition: end-term vs. mid-term
Case Study

Annual Attritions by Policy Age

Annual Attrition: NB vs old policies

- NB end term
- Old end term
- NB mid term
- Old mid term

Year 1 2 3 4 5

Annual attrition rates:
- 4.0%
- 6.0%
- 8.0%
- 10.0%
- 12.0%
- 14.0%
- 16.0%
- 18.0%
- 20.0%
Case Study

Annual Attritions by Policy Category

Annual Attrition: monoline vs package

Year
1 2 3 4 5

Mono end term
Package end term
Mono mid term
Package mid term
Case Study

### Monthly View: March Year1

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Attrition</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-term</td>
<td>16,939</td>
<td>2,086</td>
<td>12.32%</td>
</tr>
<tr>
<td>Others</td>
<td>182,160</td>
<td>1,609</td>
<td>0.88%</td>
</tr>
<tr>
<td>Total</td>
<td>199,099</td>
<td>3,695</td>
<td>1.86%</td>
</tr>
</tbody>
</table>

### Monthly View: End-term Nonrenewal Rate

- Year1Mar: 0.6%
- Year1Jun: 0.7%
- Year1Sep: 0.8%
- Year1Dec: 0.9%
- Year2Mar: 1.0%
- Year2Jun: 1.1%
- Year2Sep: 1.2%
- Year2Dec: 1.3%
- Year3Mar: 1.4%
- Year3Jun: 1.5%
- Year3Sep: 1.6%
- Year3Dec: 1.7%
- Year4Mar: 1.8%
- Year4Jun: 1.9%
- Year4Sep: 2.0%
- Year4Dec: 2.1%
- Year5Mar: 2.2%
- Year5Jun: 2.3%
- Year5Sep: 2.4%
- Year5Dec: 2.5%

### Monthly View: Mid-term Cancellation Rate

- Year1Mar: 1.0%
- Year1Jun: 1.1%
- Year1Sep: 1.2%
- Year1Dec: 1.3%
- Year2Mar: 1.4%
- Year2Jun: 1.5%
- Year2Sep: 1.6%
- Year2Dec: 1.7%
- Year3Mar: 1.8%
- Year3Jun: 1.9%
- Year3Sep: 2.0%
- Year3Dec: 2.1%
- Year4Mar: 2.2%
- Year4Jun: 2.3%
- Year4Sep: 2.4%
- Year4Dec: 2.5%
- Year5Mar: 2.6%
- Year5Jun: 2.7%
- Year5Sep: 2.8%
- Year5Dec: 2.9%
Case Study

Monthly Attritions by Policy Size

Monthly end-term attrition: by policy size

Monthly mid-term attrition: by policy size
Case Study

Monthly Attritions: Contractors vs. noncontractors

End-term attrition: contractors vs other

Mid-term attrition: contractors vs other
There are about 20 variables plus several interaction terms in the models. Only selected variables are reported.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Parameter Estimate</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Package Indicator</td>
<td>-0.12365</td>
<td>51.77775</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Rating Change</td>
<td>0.4847</td>
<td>9361.2017</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Policy Age</td>
<td>-0.00778</td>
<td>1838.8259</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.02942</td>
<td>58.5243</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### Case Study

Parameter Estimates from Logistic Regression

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Parameter Estimate</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Package Indicator</td>
<td>-0.1542</td>
<td>63.52335</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Rating Change</td>
<td>0.4167</td>
<td>899.4738</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Policy Age</td>
<td>-0.00691</td>
<td>3590.2861</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.0245</td>
<td>16.4331</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Case Study

Survival Curve for Policy Age
Case Study

Survival Curve for Policy Category
Case Study

Survival Curve for GDP Change (Percent)
Case Study

Survival Curve for Market Condition
## Validation of the Models (Table)

### Out-of-sample Performance of Survival Analysis on the 1 year attrition

<table>
<thead>
<tr>
<th>Model Decile</th>
<th>Available Obs</th>
<th>Attrition Obs</th>
<th>Attrition Rate</th>
<th>Cumulative Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9,625</td>
<td>3,697</td>
<td>38.41%</td>
<td>9,625</td>
</tr>
<tr>
<td>2</td>
<td>9,627</td>
<td>2,714</td>
<td>28.19%</td>
<td>19,252</td>
</tr>
<tr>
<td>3</td>
<td>9,624</td>
<td>2,356</td>
<td>24.48%</td>
<td>28,876</td>
</tr>
<tr>
<td>4</td>
<td>9,628</td>
<td>2,116</td>
<td>21.98%</td>
<td>38,504</td>
</tr>
<tr>
<td>5</td>
<td>9,628</td>
<td>1,935</td>
<td>20.10%</td>
<td>48,132</td>
</tr>
<tr>
<td>6</td>
<td>9,626</td>
<td>1,722</td>
<td>17.89%</td>
<td>57,758</td>
</tr>
<tr>
<td>7</td>
<td>9,627</td>
<td>1,677</td>
<td>17.42%</td>
<td>67,385</td>
</tr>
<tr>
<td>8</td>
<td>9,625</td>
<td>1,498</td>
<td>15.56%</td>
<td>77,010</td>
</tr>
<tr>
<td>9</td>
<td>9,628</td>
<td>1,245</td>
<td>12.93%</td>
<td>86,638</td>
</tr>
<tr>
<td>10</td>
<td>9,626</td>
<td>1,054</td>
<td>10.95%</td>
<td>96,264</td>
</tr>
<tr>
<td>Total</td>
<td>96,264</td>
<td>20,014</td>
<td>20.79%</td>
<td>96,264</td>
</tr>
</tbody>
</table>

### Out-of-sample Performance of Logistic Regression on the 1 year attrition

<table>
<thead>
<tr>
<th>Model Decile</th>
<th>Available Obs</th>
<th>Attrition Obs</th>
<th>Attrition Rate</th>
<th>Cumulative Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9,622</td>
<td>3,567</td>
<td>37.07%</td>
<td>9,622</td>
</tr>
<tr>
<td>2</td>
<td>9,630</td>
<td>2,790</td>
<td>28.97%</td>
<td>19,252</td>
</tr>
<tr>
<td>3</td>
<td>9,627</td>
<td>2,303</td>
<td>23.92%</td>
<td>28,879</td>
</tr>
<tr>
<td>4</td>
<td>9,626</td>
<td>2,148</td>
<td>22.31%</td>
<td>38,505</td>
</tr>
<tr>
<td>5</td>
<td>9,628</td>
<td>1,929</td>
<td>20.04%</td>
<td>48,133</td>
</tr>
<tr>
<td>6</td>
<td>9,626</td>
<td>1,758</td>
<td>18.26%</td>
<td>57,759</td>
</tr>
<tr>
<td>7</td>
<td>9,626</td>
<td>1,641</td>
<td>17.05%</td>
<td>67,385</td>
</tr>
<tr>
<td>8</td>
<td>9,626</td>
<td>1,450</td>
<td>15.06%</td>
<td>77,011</td>
</tr>
<tr>
<td>9</td>
<td>9,627</td>
<td>1,310</td>
<td>13.61%</td>
<td>86,638</td>
</tr>
<tr>
<td>10</td>
<td>9,626</td>
<td>1,118</td>
<td>11.61%</td>
<td>96,264</td>
</tr>
<tr>
<td>Total</td>
<td>96,264</td>
<td>20,014</td>
<td>20.79%</td>
<td>96,264</td>
</tr>
</tbody>
</table>
Case Study

Validation of the Models (Lift)

Out-of-sample Lift: average attrition

Out-of-sample lift: relatively to average
Case Study

Validation of the Models (Gini Chart)

Out-of-sample Performance of Survival Analysis on the 1 year attrition

Out-of-sample Performance of Logistic Regression on the 1 year attrition
Conclusions

• Survival analysis addresses not only whether a policy will leave, but also when it will leave.
• Provide a dynamic insight by utilizing panel data and improve the static view derived from snapshot data.
• Analyze mid-term cancellation and end-term nonrenewal sequentially and simultaneously.
• Able to measure the impacts of time-variant macroeconomic variables on attrition.
• Empirical study does not show significant lift improvement over logistics regression