# Estimate Attrition Using Survival Analysis

Hongyuan Wang, Ph.D.
Luyang Fu, Ph.D., FCAS, MAAA

March 2011

# Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

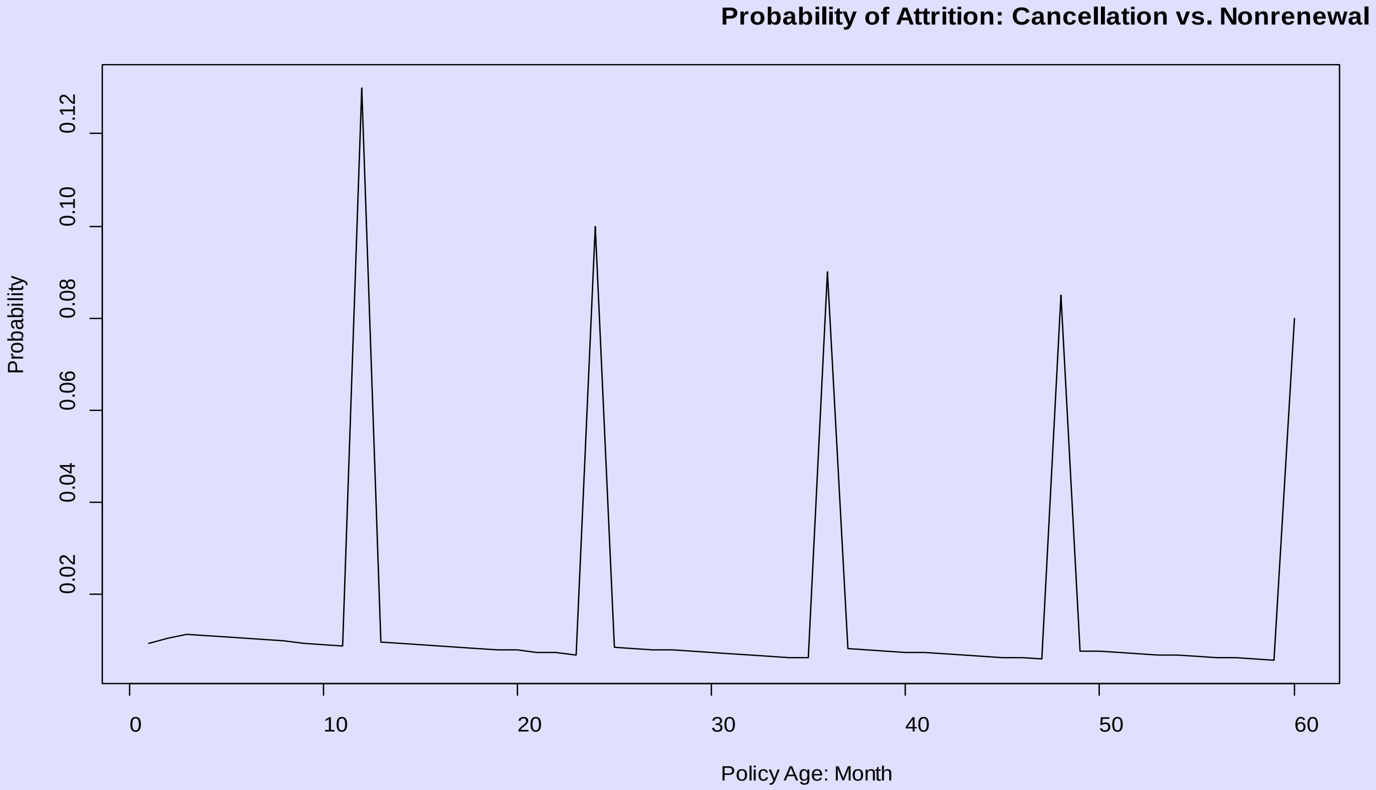
# Agenda

- Introduction
- Survival Analysis
- Cox Proportional Hazard Model
- A case study
- Q&A

# Introduction

# Two Ways of Attrition

- Mid-term cancellation
- End-of-term nonrenewal



# Snapshot View of Retention/Attrition

- If there were 10,000 inforced policies at 12/31/2009, how many of them were still with the company at 12/31/2010?
- Variable of interest: yes or no
- Do not separate cancellation and nonrenewal.
- Static view

# Dynamic View of Retention/Attrition

- If there were 10,000 inforced policies at 12/31/2009, how many of them left by cancelation and non-renewal, and when they left?
- Variable of interest: t (time of attrition)
- Cancellation and non-renewal occurs sequentially and dynamically.
- Time-varying variables (Unemployment, GDP change, Premium Change …) impact retention.

# Why Survival Analysis?

- Better estimation of life time value: not just whether a policy will leave, but when it will leave.
- Estimate cancellation and non-renewal sequentially and simultaneously.
- Measure the impacts of time-variant macroeconomic variables on attrition by incorporating monthly macroeconomic data in the regression.

# Survival Analysis

# What is 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).

# What is Survival Time?

- Refers to a variable *t* which measures the time from a particular starting time (e.g., time initiated the treatment) to a particular endpoint of interest (e.g., attaining certain functional abilities).
- Examples:

  *Insurance Policy*: Started at Jan2005, terminated at Aug2008.

  *Products*: Bought at Dec2006, failed at Feb2007.

# 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 Functions

- Survival Function *S(t)* :

  $S(t)$ = Prob{T ≥ t}, here t ≥ 0 ;

- Lifetime Distribution Function *F(t)* :

  $F(t) = 1 - S(t)$ ;

- Event Density Function *f(t)* :

  Prob{t≤ T ≤ t+δt} = f(t)δt, $\frac{dF(t)}{dt} = f(t)$

- Hazard Function *h(t)* :

  $h(t) = f(t)/S(t)$

  or h(t)δt = Prob{t ≤ T≤ t+δt |T≥t};

# Survival Analysis Functions

All those functions are connected.

- Density function is the negative of the derivative of the survival function;
- Hazard function is the negative of the derivative of the log of the survival function.

$$f(t) = F'(t) = -S'(t)$$

$$h(t) = -\,{}^{d(\ln S(t))}\!/_{dt}$$

$$S(t) = \exp\left\{-\int_0^t h(s)ds\right\}$$

$$f(t) = h(t)\exp\left\{-\int_0^t h(s)ds\right\}$$

# Survival Analysis Functions

- The most popular distributions are exponential, Weibull, etc.
- Exponential: $S(t) = exp(-\lambda t) \quad \lambda > 0$ ;

  $f(t)= \lambda exp(-\lambda t);$

  $h(t) =\lambda$ ; ( so no ageing)
- Weibull; $S(t) =exp\ (-\beta t^{\alpha}) \qquad \alpha\ ,\ \beta > 0$ ;

  $f(t) = \alpha\beta t^{\alpha-1} (exp(-\beta t^{\alpha} ));$

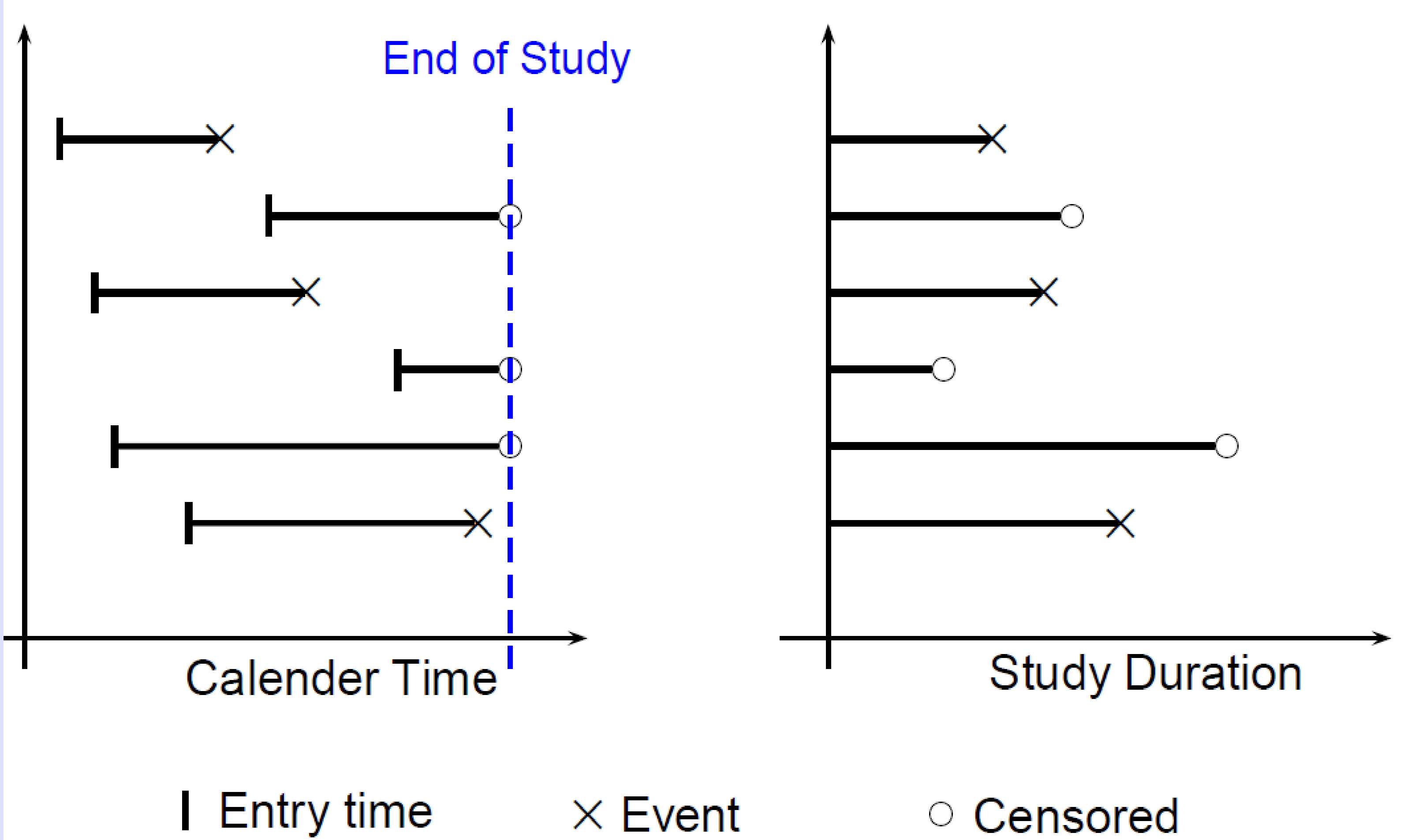
  $h(t) = \alpha\beta t^{\alpha-1}$ ;

  $\alpha > 1$ (increasing hazard) , $\alpha < 1$ (decreasing hazard)

# 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 Data



# Examples

| Duration Times of Interest in Marketing | | |
|---|---|---|
| **Subdiscipline** | **Decision/Forecasting** | **Duration Time** |
| **Pricing/Promotion** | Timing of price chinages or promotions;<br>Measuring effect of promotion | Interpurchase duration; Timing of coupon redemption |
| **Salesforce Management** | Forecasting and managing salesforce turnover | Salesperson job duration |
| **New Product Development** | Forecasting trial, adoption, depth of repeat purchase | Duration time from new product introduction until initial trial;<br>Interpurchase times |
| **Marketing Research** | Forecasting response rates;<br>Forecasting size and composition of firm's customer base; | Time until survey response;<br>Time until customer becomes inactive or disaffected;<br>Time until cancellation of service contract; |

Sources: Kristiaan H. and D. C. Schmittlein, 1993, Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models; *Marketing Science*, Vol. 12, No. 4, page 396 .

# 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 Equation

Let $h(t \mid x_t)$ denote the resultant 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 $x_t = (x_{1t}, x_{2t}, \ldots\ldots, x_{kt})$ $\beta = (\beta_1, \beta_2, \ldots\ldots, \beta_k)$

$k$ is the total number of the covariates,

$\beta_j$ *is the constant Proportional effect of* $x_j$

The term *$h_0(t)$* is called the *baseline hazard*; it is the hazard for the respective individual when all independent variable values are equal to zero.

# Cox Proportional Hazard Model Equation

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:

$$\ln\{h(t \mid x_t)/h_0(t)\} = \beta' x_t$$

Taking partial derivative we have

$$\partial \ln h(t \mid x_t, \beta)/\partial x_{jt} = \beta_j$$

# Partial Likelihood Estimation of *β*

$$L(i \mid t, j_1, j_2, \ldots\ldots, j_{n(t)}) = \frac{h_i(t)}{\sum_{k=1}^{n(t)} h_{j_k}(t)} \qquad (1)$$

$$L(i \mid t, j_1, j_2, \ldots\ldots, j_{n(t)}) = \frac{h_0(t) e^{\beta' x_{it}}}{\sum_{k=1}^{n(t)} h_0(t) e^{\beta' x_{j_k t}}} \qquad (2)$$

$$L(i \mid t, j_1, j_2, \ldots\ldots, j_{n(t)}) = \frac{e^{\beta' x_{it}}}{\sum_{k=1}^{n(t)} e^{\beta' x_{j_k t}}} \qquad (3)$$

Estimation of ***β*** is obtained by Maximizing the Product of Expression (3) over all observed duration times.

# Literatures

- Kristiaan H. and D. C. Schmittlein, 1993, Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models; *Marketing Science*, Vol. 12, No. 4, pp. 395-414 .
- Graves S, D. Kletter, W. B. Hetzel, R. N. Bolton, 1998, A Dynamic Model of the Duration of the Customer’s Relationship with a Continuous Service Provider: The Role of Satisfaction, *Marketing Science*, Vol. 17, No. 1, pp. 45-65.
- Andreeva G., 2006, European Generic Scoring Models Using Survival Analysis, *Journal of the Operational Research Society*, Vol. 57, No. 10, pp. 1180-1187.
- Bellotti T. and J. Crook, 2009, Credit Scoring With Macroeconomic Variables Using Survival Analysis; *Journal of the Operational Research Society*, Vol. 60, pp. 1699–1707.

# A Case Study

# Case Study Data

- 6.5 years Commercial Line Policies.
- The Dependent Variable:

  Duration = The time until the policy cancellation
- If a policy is still alive at the end of study, it is right censored ( i.e. Censor = 1)
- Monthly policy data and economic data are stacked together to get the final model data.

# Annual Attrition Summary

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 24,570 | 156,478 | 16,907 | 197,955 | 12.41% | 79.05% | 8.54% |
| 200601 | 25,101 | 158,794 | 17,529 | 201,424 | 12.46% | 78.84% | 8.70% |
| 200701 | 24,756 | 159,079 | 18,057 | 201,892 | 12.26% | 78.79% | 8.94% |
| 200801 | 24,951 | 160,688 | 19,697 | 205,336 | 12.15% | 78.26% | 9.59% |
| 200901 | 27,398 | 162,875 | 20,787 | 211,061 | 12.98% | 77.17% | 9.85% |

The data is for illustration purpose.

# Annual Attritions by Policy Type

**Line1**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 10,708 | 63,270 | 7,283 | 81,262 | 13.18% | 77.86% | 8.96% |
| 200601 | 11,292 | 65,190 | 7,924 | 84,407 | 13.38% | 77.23% | 9.39% |
| 200701 | 11,657 | 64,801 | 8,336 | 84,793 | 13.75% | 76.42% | 9.83% |
| 200801 | 11,525 | 64,178 | 9,539 | 85,242 | 13.52% | 75.29% | 11.19% |
| 200901 | 12,860 | 63,911 | 10,469 | 87,241 | 14.74% | 73.26% | 12.00% |

**Line2**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 9,630 | 67,757 | 7,101 | 84,488 | 11.40% | 80.20% | 8.40% |
| 200601 | 9,514 | 66,928 | 7,076 | 83,518 | 11.39% | 80.14% | 8.47% |
| 200701 | 8,666 | 66,705 | 6,799 | 82,170 | 10.55% | 81.18% | 8.27% |
| 200801 | 8,615 | 68,238 | 7,280 | 84,133 | 10.24% | 81.11% | 8.65% |
| 200901 | 9,611 | 70,428 | 7,516 | 87,555 | 10.98% | 80.44% | 8.58% |

**Line3**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 4,232 | 25,450 | 2,523 | 32,206 | 13.14% | 79.02% | 7.83% |
| 200601 | 4,295 | 26,676 | 2,529 | 33,500 | 12.82% | 79.63% | 7.55% |
| 200701 | 4,433 | 27,574 | 2,922 | 34,930 | 12.69% | 78.94% | 8.37% |
| 200801 | 4,810 | 28,272 | 2,878 | 35,960 | 13.38% | 78.62% | 8.00% |
| 200901 | 4,927 | 28,536 | 2,803 | 36,265 | 13.59% | 78.69% | 7.73% |

# Annual Attritions by Premium Change

**Annual Premium change < -x%**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 330 | 2,247 | 113 | 2,690 | 12.27% | 83.53% | 4.21% |
| 200601 | 3,657 | 22,867 | 880 | 27,405 | 13.35% | 83.44% | 3.21% |
| 200701 | 4,317 | 31,587 | 1,417 | 37,321 | 11.57% | 84.64% | 3.80% |
| 200801 | 5,103 | 37,126 | 1,602 | 43,831 | 11.64% | 84.70% | 3.66% |
| 200901 | 4,041 | 24,618 | 892 | 29,551 | 13.67% | 83.31% | 3.02% |

**Annual Premium change (-x%, 0%)**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 110 | 1,194 | 72 | 1,376 | 7.99% | 86.76% | 5.25% |
| 200601 | 1,514 | 12,866 | 676 | 15,056 | 10.06% | 85.45% | 4.49% |
| 200701 | 2,181 | 18,409 | 958 | 21,548 | 10.12% | 85.43% | 4.45% |
| 200801 | 2,306 | 18,315 | 829 | 21,450 | 10.75% | 85.38% | 3.87% |
| 200901 | 1,348 | 10,783 | 493 | 12,625 | 10.68% | 85.42% | 3.91% |

**Annual Premium change (0%, x%)**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 23,656 | 153,279 | 11,864 | 188,800 | 12.53% | 81.19% | 6.28% |
| 200601 | 13,448 | 85,676 | 7,538 | 106,661 | 12.61% | 80.33% | 7.07% |
| 200701 | 12,725 | 77,042 | 7,157 | 96,924 | 13.13% | 79.49% | 7.38% |
| 200801 | 13,844 | 84,623 | 9,115 | 107,582 | 12.87% | 78.66% | 8.47% |
| 200901 | 18,302 | 109,942 | 11,085 | 139,329 | 13.14% | 78.91% | 7.96% |

**Annual Premium change > x%**

| BaseMonth | nonRenewed | Renewed | Midterm_canceled | Total | nonRenewedPer | RenewedPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200501 | 474 | 4,232 | 478 | 5,184 | 9.15% | 81.64% | 9.21% |
| 200601 | 6,482 | 41,619 | 4,355 | 52,456 | 12.36% | 79.34% | 8.30% |
| 200701 | 5,533 | 36,853 | 3,928 | 46,313 | 11.95% | 79.57% | 8.48% |
| 200801 | 3,698 | 25,252 | 3,723 | 32,674 | 11.32% | 77.29% | 11.40% |
| 200901 | 3,708 | 21,809 | 4,235 | 29,752 | 12.46% | 73.30% | 14.24% |

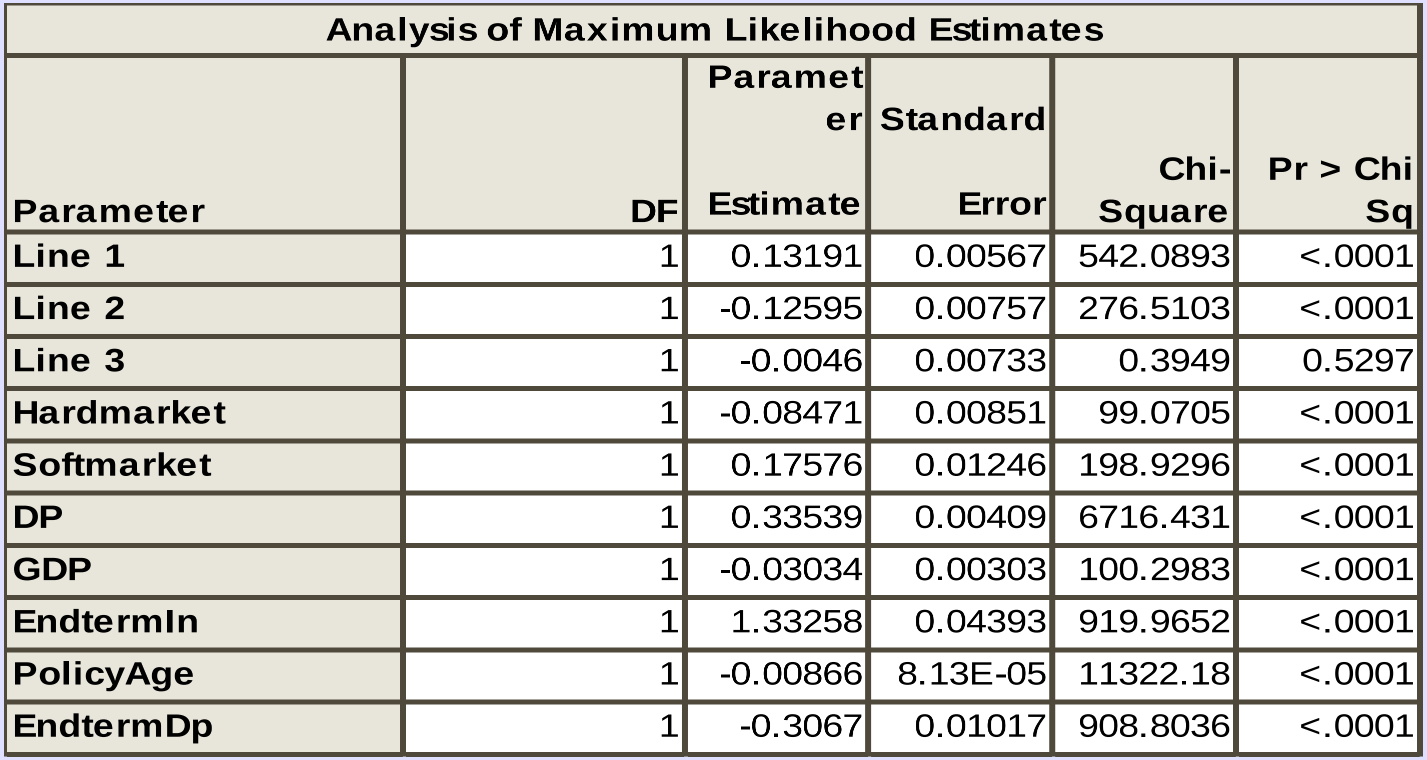
# Monthly View

**Monthly Snapshot**

| | Active | Withdraw | Percent |
|---|---|---|---|
| Endterm | 16,939 | 2,086 | 12.32% |
| Others | 182,161 | 1,609 | 0.88% |
| Total | 199,099 | 3,695 | 1.86% |

| BaseMonth | nonRenewed | Renewed | Midterm Canceled | Total | Mid-term Stayed | nonRenewPer | Midterm_cancelPer |
|---|---|---|---|---|---|---|---|
| 200503 | 2,086 | 14,852 | 1,609 | 199,099 | 180,552 | 12.32% | 0.88% |
| 200506 | 2,089 | 14,789 | 1,609 | 200,793 | 182,305 | 12.38% | 0.87% |
| 200509 | 1,750 | 12,879 | 1,502 | 201,314 | 185,183 | 11.96% | 0.80% |
| 200512 | 1,565 | 11,330 | 1,602 | 201,192 | 186,694 | 12.13% | 0.85% |
| 200603 | 2,228 | 15,292 | 1,775 | 201,657 | 182,362 | 12.72% | 0.96% |
| 200606 | 2,083 | 14,805 | 1,455 | 201,820 | 183,477 | 12.33% | 0.79% |
| 200609 | 1,797 | 13,096 | 1,684 | 201,698 | 185,120 | 12.07% | 0.90% |
| 200612 | 1,584 | 11,437 | 1,584 | 201,145 | 186,541 | 12.16% | 0.84% |
| 200703 | 2,284 | 15,597 | 1,634 | 202,562 | 183,047 | 12.77% | 0.88% |
| 200706 | 1,910 | 14,997 | 1,483 | 203,966 | 185,576 | 11.30% | 0.79% |
| 200709 | 1,725 | 13,237 | 1,690 | 204,830 | 188,178 | 11.53% | 0.89% |
| 200712 | 1,615 | 11,578 | 1,939 | 204,858 | 189,727 | 12.24% | 1.01% |
| 200803 | 2,174 | 15,955 | 1,763 | 206,118 | 186,226 | 11.99% | 0.94% |
| 200806 | 2,055 | 15,038 | 1,687 | 208,880 | 190,100 | 12.02% | 0.88% |
| 200809 | 1,895 | 13,291 | 1,750 | 210,140 | 193,205 | 12.48% | 0.90% |
| 200812 | 1,568 | 11,547 | 2,573 | 210,703 | 195,015 | 11.95% | 1.30% |
| 200903 | 2,328 | 16,087 | 2,111 | 212,861 | 192,334 | 12.64% | 1.09% |
| 200906 | 2,313 | 15,371 | 1,989 | 214,614 | 194,942 | 13.08% | 1.01% |
| 200909 | 2,168 | 13,759 | 1,910 | 214,595 | 196,758 | 13.61% | 0.96% |
| 200912 | 1,847 | 11,836 | 2,297 | 212,302 | 196,322 | 13.50% | 1.16% |

# Parameter Estimates Using PHREG



Analysis of Maximum Likelihood Estimates

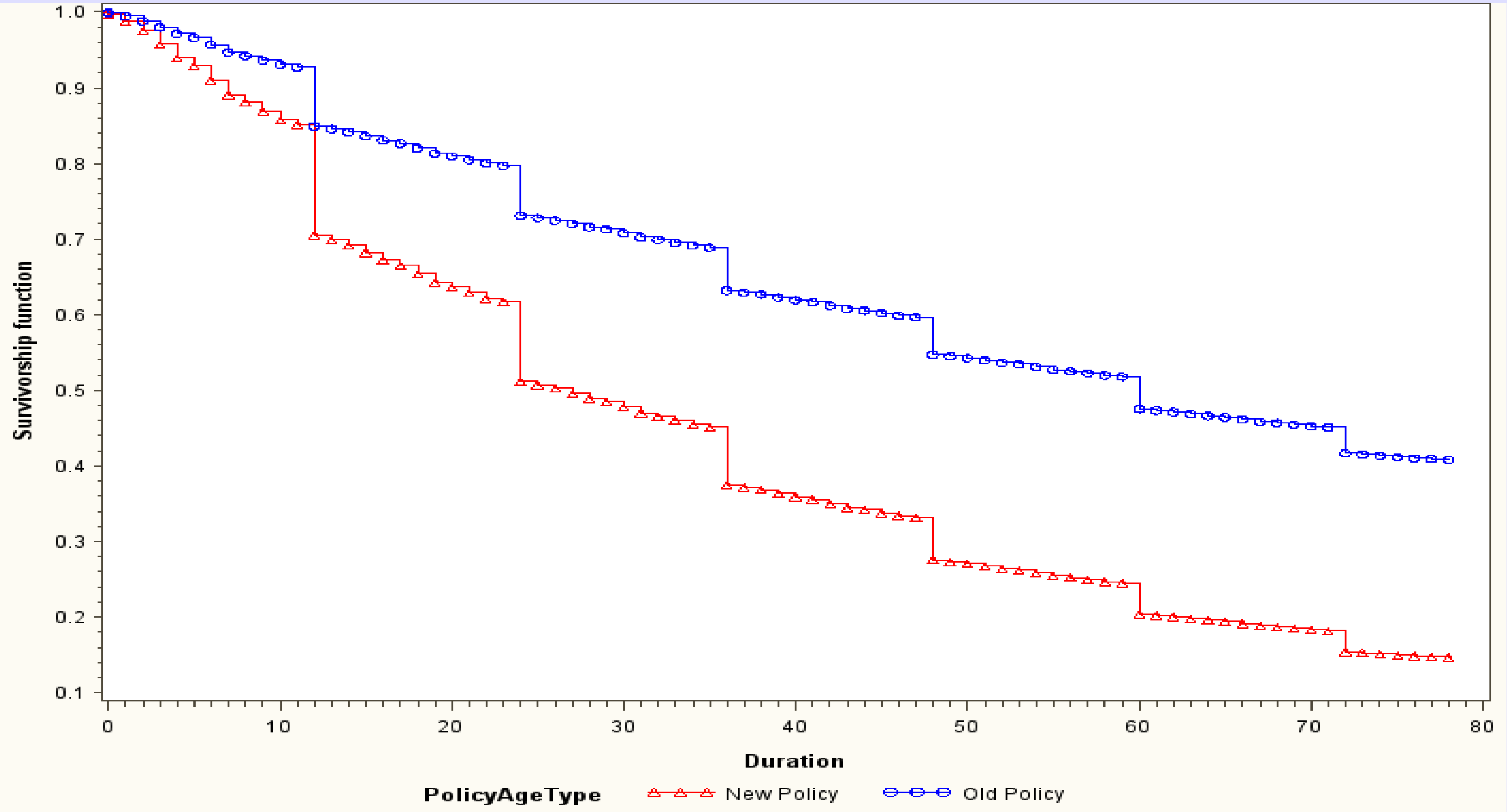
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| Line 1 | 1 | 0.13191 | 0.00567 | 542.0893 | <.0001 |
| Line 2 | 1 | -0.12595 | 0.00757 | 276.5103 | <.0001 |
| Line 3 | 1 | -0.0046 | 0.00733 | 0.3949 | 0.5297 |
| Hardmarket | 1 | -0.08471 | 0.00851 | 99.0705 | <.0001 |
| Softmarket | 1 | 0.17576 | 0.01246 | 198.9296 | <.0001 |
| DP | 1 | 0.33539 | 0.00409 | 6716.431 | <.0001 |
| GDP | 1 | -0.03034 | 0.00303 | 100.2983 | <.0001 |
| EndtermIn | 1 | 1.33258 | 0.04393 | 919.9652 | <.0001 |
| PolicyAge | 1 | -0.00866 | 8.13E-05 | 11322.18 | <.0001 |
| EndtermDp | 1 | -0.3067 | 0.01017 | 908.8036 | <.0001 |

There are about 20 variables plus several interaction terms in the models. Only selected variables are reported.

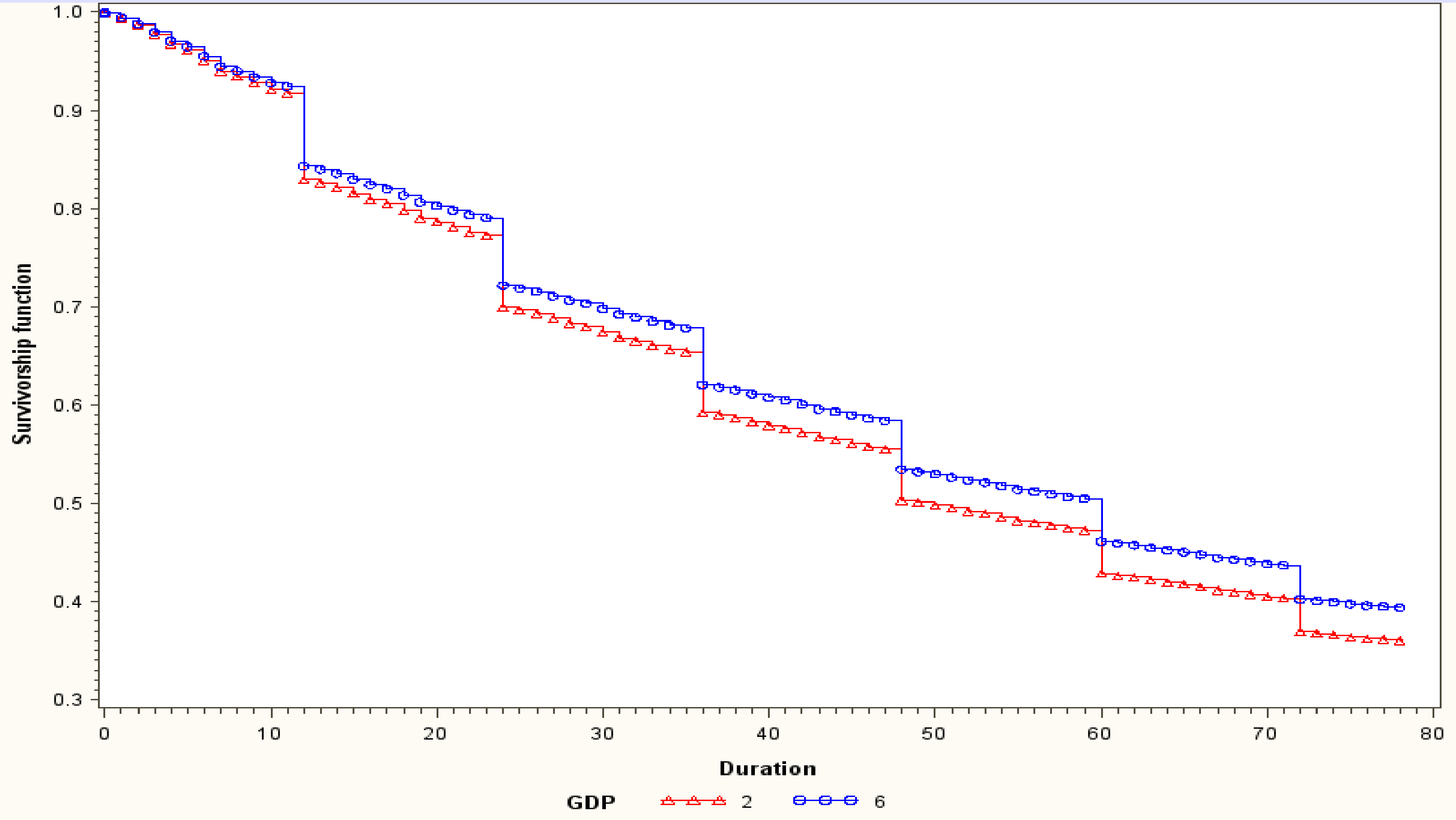
# Parameter Estimates Using Logistic

| Analysis of Maximum Likelihood Estimates | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > Chi Sq** |
| **Intercept** | 1 | -1.6028 | 0.0804 | 397.1769 | <.0001 |
| **Line 1** | 1 | 0.0816 | 0.0145 | 31.513 | <.0001 |
| **Line 2** | 1 | -0.1732 | 0.0143 | 146.9164 | <.0001 |
| **Line 3** | 1 | -0.0539 | 0.0164 | 10.7724 | 0.001 |
| **Softmarket** | 1 | 0.1281 | 0.0267 | 22.9947 | <.0001 |
| **DP** | 1 | 0.4227 | 0.0118 | 1290.378 | <.0001 |
| **GDP** | 1 | -0.0311 | 0.00486 | 40.8976 | <.0001 |
| **PolicyAge** | 1 | -0.00717 | 0.000094 | 5866.395 | <.0001 |

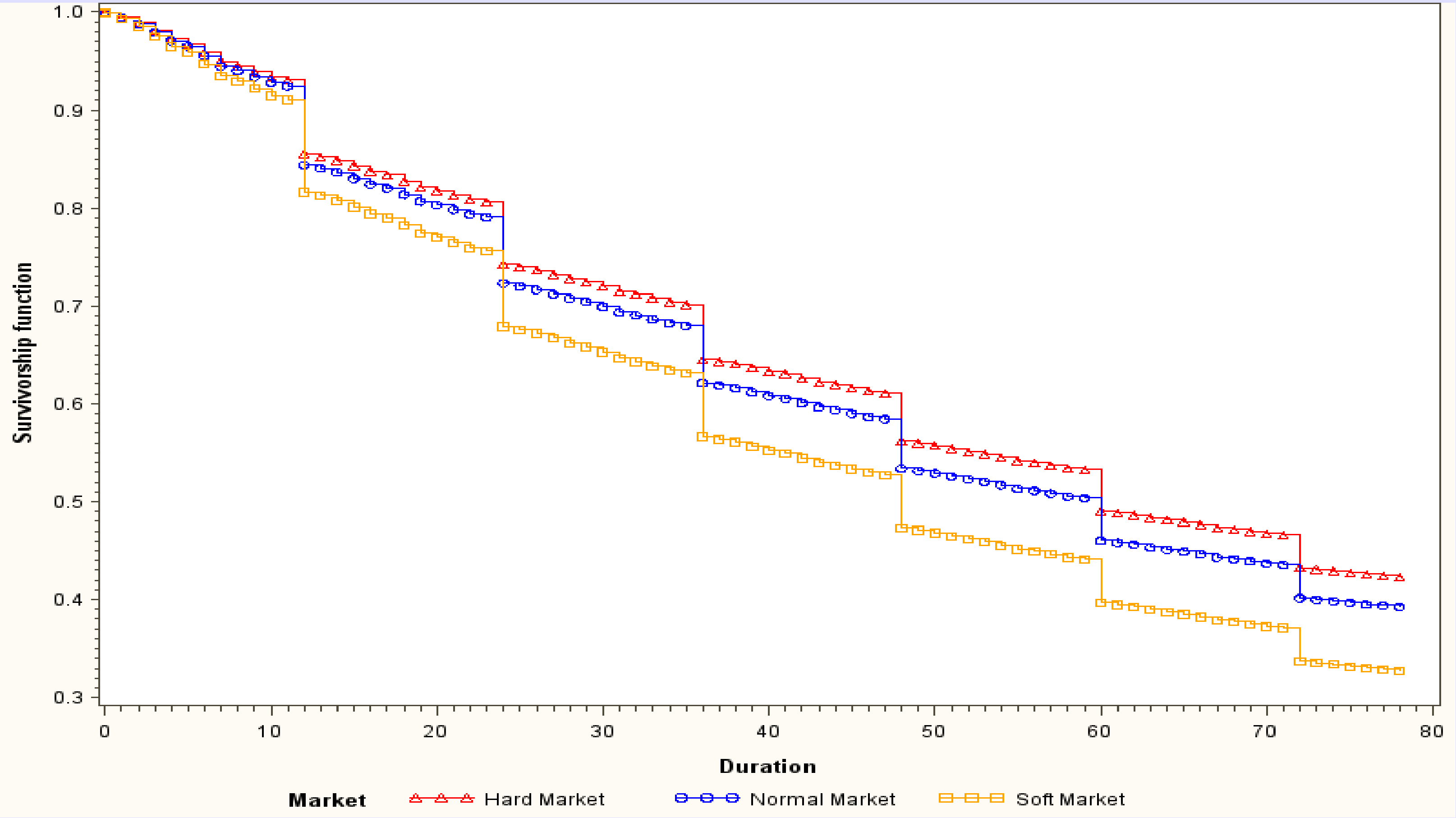
# Survival Curve for Policy Age



# Survival Curve for GDP Change (Percent)



# Survival Curve for Market Condition



# 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.

# Q & A

STATE AUTO®
Insurance Companies