Statistical Analysis of Life Insurance Policy Termination and Survivorship

Emiliano A. Valdez, PhD, FSA
Michigan State University

joint work with J. Vadiveloo and U. Dias

The School of Statistics
Renmin University of China
Beijing, CHINA, 7 January 2014

Preliminaries

- An individual has a future lifetime random variable T and is exposed to two possible reasons to fail: withdrawal (policy termination) or mortality (death).
- Denote the cause of failure by J with:
  - $J = w$ indicates failure due to withdrawal, and
  - $J = d$ indicates failure due to death.
- Convenient to introduce theoretical "net" lifetime random variables: T_w and T_d. Assume their respective distribution, survival and density functions exist: F_j, S_j and f_j, for $j = w, d$.
- *Competing Risk Models*: T_w and T_d are never observed simultaneously, but only (T, J) where $T = \min(T_w, T_d)$.
- Model identifiability is a common issue here: one approach is to specify the joint distribution or copula function associated with (T_w, T_d). See Tsiatis (1975).

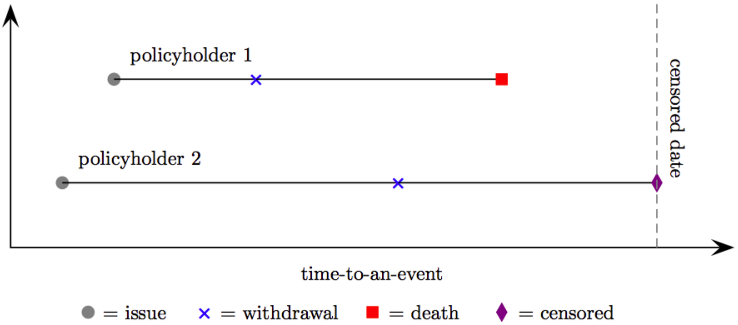
Competing risk models

- Competing risk models can be applied in several disciplines:
  - actuarial science: life insurance contracts
  - economics: duration till employment, cause of leaving employment
  - medical statistics: clinical trials
  - epidemiology: occurrence/recovery of diseases
  - engineering: time/cause of failure of a mechanical system
- In actuarial science, some of the literature:
  - Carriere (1994, 1998), Valdez (2000), Tsai, Kuo and Chen (2002)
  - Actuarial students study what is called "Multiple Decrement Models". Plenty of literature here.

Outline

Motivation for model constructions

- Data-driven. Our observables are best illustrated by the following figure:



- This diagram provides an illustration of the observed times until withdrawal and times until death.

Data source used in the calibration

- A sub-sample from a portfolio of life insurance contracts from a major insurer.
  - detailed information on the type of policies (e.g. PAR, TERM, UL, CONV) and additional characteristics
  - sub-sample consists of 65,435 terminated single-life insurance contracts with mortality dates tracked from the US Social Security System administration office
  - our data file recorded a 1918 as the year with the earliest policy issue date and the end of the observation period is 14 February 2008
- Our policy record indicates 61,901 of the total observations are censored, representing about 94.6% of the observation.
- For each contract observed, we have policy effective (issue) date, the termination date and the date of death, if applicable.

Policy characteristics and other observable information

Categorical variables	Description			
			Proportions	
PlanType	Type of insurance plan:	PlanTypeP	42.4%	
		PlanTypeT	28.0%	
		PlanTypeO	29.6%	
RiskClass	Insured's assigned risk class:	RiskClass = N	72.0%	
		RiskClass = Y	28.0%	
Sex	Insured's sex:	Male = 1	65.2%	
		Female = 0	34.8%	
Smoker	Smoker class:	Non-smoker = N	66.6%	
		Smoker = S	12.4%	
		Combined = C	21.0%	
Censor	Censoring indicator for death:	Censor = 1	94.6%	
		Censor = 0	5.4%	
Continuous variables		Minimum	Mean	Maximum
IssAge	The policyholder's issue age	0	37.70	89.65
Face Amount	The policy's insured amount	1	213,000	60,000,000
Temp FEAmt	Temporary flat extra amount (per 1000)	0.00	0.08	49.00
Perm FEAmt	Permanent flat extra amount (per 1000)	0.00	0.06	48.00
MEFact	Extra mortality factor	1.00	1.01	4.00
Dates				
IssDate	Policy effective or issue date			
BDate	Insured's date of birth			
WDate	Policy withdrawal or lapse date			
DDate	Insured's date of death, if applicable			

Count and face amount

Number of policies and average face amount by plan type, sex and issue age

Plan Type	Issue Age								Total
	Males				Females				
	≤ 30	30-50	50-70	> 70	≤ 30	30-50	50-70	> 70	
PlanTypeP									
Count	6,461	8,476	2,300	100	4,401	4,545	1,374	119	27,776
Face Amount	46,766	152,345	139,624	213,028	35,611	103,401	150,228	213,891	100,605
PlanTypeT									
Count	1,130	9,557	1,963	20	964	4,262	434	3	18,333
Face Amount	323,955	475,092	653,320	1,461,250	168,350	251,603	408,421	425,833	416,264
PlanTypeO									
Count	2,076	7,314	3,091	188	1,516	3,789	1,103	249	19,326
Face Amount	124,896	193,958	203,519	445,704	79,893	133,510	310,929	604,947	181,690

A class of duration models for time-until-withdrawal

Suppose we can write T_w as $T_w = \exp(\mu) T_0^\sigma$ for some non-negative rv T_0. With log-transformation,

$$\log(T_w) = \mu + \sigma \log(T_0) = \mu + \sigma\Lambda,$$

where $\Lambda = \log(T_0)$, μ and σ are location and scale parameter provided $\sigma \neq 0$ to avoid a degenerate distribution.

Because we can write the survival distribution function of T_w as

$$S_w(t) = \begin{cases} S_\Lambda\left(\dfrac{\log(t) - \mu}{\sigma}\right), & \sigma > 0 \\ 1 - S_\Lambda\left(\dfrac{\log(t) - \mu}{\sigma}\right), & \sigma < 0 \end{cases}$$

where S_Λ denotes the survival function of Λ, the distribution of T_w belongs to a log-location-scale family of distributions.

Covariates

Introduce covariates through the location parameter μ.

With $\mathbf{x}$ as a vector of covariates, such as policyholder characteristics, and β, the vector of linear coefficients.

Then replace $\mu = \mathbf{x}'\beta$.

We have $T_w = \exp(\mathbf{x}'\beta)T_0^{\sigma}$ and

$$\log(T_w) = \mathbf{x}'\beta + \sigma \log(T_0) = \mathbf{x}'\beta + \sigma\Lambda,$$

which generalizes the ordinary regression model.

This specification is a special case of the Accelerated Failure Time (AFT) model commonly studied in survival analysis.

Distribution of the time-until-withdrawal

Straightforward to find explicit form of the distribution of T_w in terms of the distribution of T_0.

- The survival function of T_w can be expressed as

$$S_w(t) = S_0\left((e^{-\mu}t)^{1/\sigma}\right).$$

- Its density can be expressed as

$$f_w(t) = \frac{1}{|\sigma|t}(e^{-\mu}t)^{1/\sigma} f_0\left((e^{-\mu}t)^{1/\sigma}\right),$$

  where S_0 and f_0 are respectively the survival and density functions of T_0.

- Within this class of models, oftentimes more straightforward to specify the distribution of T_0 rather than of its logarithm.

Class of distribution models considered

- Log-Normal Distribution: T_0 has a log-normal distribution with parameters 0 and 1.

$$f_w(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp\left[-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right].$$

- Generalized Gamma Distribution: T_0 is a standard Gamma with scale of 1, shape parameter m.

$$f_w(t) = \frac{1}{|\sigma| t}\frac{1}{\Gamma(m)}(e^{-\mu}t)^{m/\sigma} \exp\left[-(e^{-\mu}t)^{1/\sigma}\right].$$

- GB2 Distribution: T_0 has a Beta of the second kind (B2) density with parameters γ_1 and γ_2.

$$f_w(t) = \frac{1}{|\sigma| t}\frac{1}{B(\gamma_1,\gamma_2)}\frac{(e^{-\mu}t)^{\gamma_1/\sigma}}{\left[1+(e^{-\mu}t)^{1/\sigma}\right]^{\gamma_1+\gamma_2}}.$$

Survival models

Let the (fixed) issue age be z and X_d the age-at-death r.v. so that

$$X_d|z = z + T_w + (T_d - T_w) = z + T_w + T_{wd},$$

provided $T_{wd} > 0$.

If T_w is known, then $(X_d|z, T_w = t_w) = z + t_w + T_{wd}$.

Thus, we have

$$\begin{aligned} P(T_{wd} > t_{wd}|z, T_w = t_w) &= P(T_d > T_w + t_{wd}|z, T_w = t_w) \\ &= \frac{P(X_d > z + t_w + t_{wd})}{P(X_d > z + t_w)} \\ &= \frac{S_d(z + t_w + t_{wd})}{S_d(z + t_w)}, \end{aligned}$$

where S_d is the survival function of X_d.

Survival models considered

- Gompertz Distribution: Survival function has the form

$$S_d(x) = \exp\left[e^{-m^*/\sigma^*}\left(1 - e^{x/\sigma^*}\right)\right],$$

where $m^* > 0$ is mode and $\sigma^* > 0$ is dispersion about this mode. See Carriere (1992). With $B = \frac{1}{\sigma^*}\exp(-m^*/\sigma^*)$ and $c = \exp(1/\sigma^*)$, it leads us to the hazard function

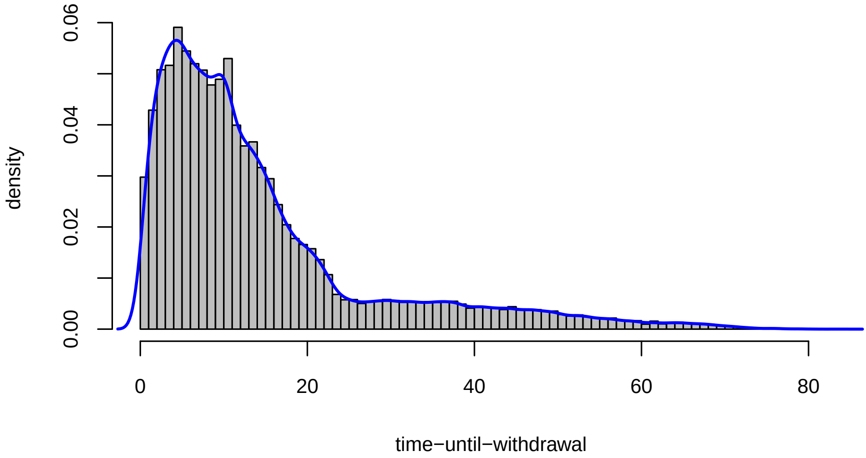
$$\mu_x = \frac{f_d(x)}{S_d(x)} = Bc^x.$$

- Weibull Distribution: Survival function has the form

$$S_d(x) = \exp\left[-(x/m^*)^{m^*/\sigma^*}\right],$$

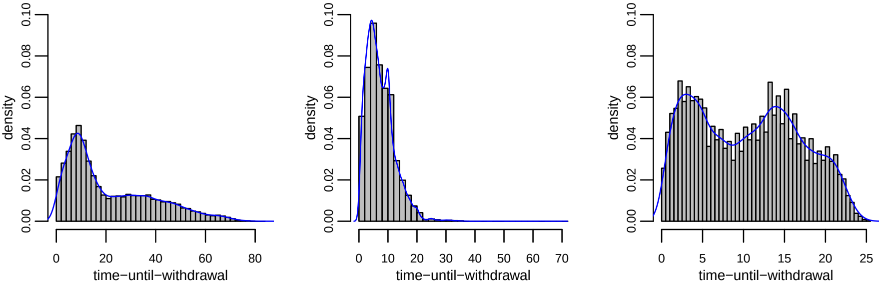
where $m^* > 0$ and $\sigma^* > 0$ are respectively location and dispersion parameters. See also Carriere (1992). Popularly known in survival analysis and reliability theory.

Preliminary investigation - histogram observed



By type of plan

Plan Type	Number	Min	Mean	Median	Max	Std Dev
PlanTypeP	27,776	0.08	21.46	14.80	83.75	17.24
PlanTypeT	18,333	0.01	7.34	6.42	70.15	4.83
PlanTypeO	19,326	0.08	10.51	10.62	25.01	6.36
Aggregate	65,435	0.01	14.27	10.01	83.75	13.57



MLEs for the various duration models

Parameter	Log-Normal		Generalized Gamma	GB2
Regression coefficients				
β_0 (intercept)	2.5534 (0.0263)		1.2138 (0.0419)	3.0034 (0.0238)
β_1 (PlanTypeP)	-0.4022 (0.0071)		-0.1604 (0.0061)	-0.1956 (0.0054)
β_2 (PlanTypeT)	-0.2808 (0.0068)		-0.1422 (0.0060)	-0.2805 (0.0055)
β_5 (RiskClassY)	-0.9787 (0.0063)		-0.6593 (0.0056)	-0.8199 (0.0060)
β_6 (Male)	0.0582 (0.0053)		0.0297 (0.0047)	0.0326 (0.0041)
β_7 (SmokerN)	0.2388 (0.0079)		0.3641 (0.0065)	0.1258 (0.0063)
β_8 (SmokerC)	1.6988 (0.0099)		1.7042 (0.0086)	1.2458 (0.0079)
β_{10} (Face Amount)	-0.0003 (0.0004)	*	-0.0027 (0.0003)	-0.0089 (0.0004)
β_{11} (Temp FEAmt)	0.0157 (0.0026)		0.0287 (0.0027)	-0.0258 (0.0020)
β_{12} (Perm FEAmt)	-0.0104 (0.0028)		-0.0167 (0.0023)	-0.0306 (0.0024)
β_{13} (MEFact)	-0.1168 (0.0240)		-0.6373 (0.0162)	-0.1553 (0.0216)
β_{14} (IssAge)	-0.0060 (0.0002)		-0.0092 (0.0002)	-0.0030 (0.0002)
Model specific parameters				
σ	0.6464 (0.0018)		1.2089 (0.0130)	0.2190 (0.0065)
m	-		4.5774 (0.0966)	-
γ_1	-		-	0.4303 (0.0168)
γ_2	-		-	1.2020 (0.0486)
Model fit statistics				
Number of observations	65,435		65,435	65,435
Log-likelihood	-209,054.1		-206,010.2	-201,199.5
Number of parameters	13		14	15
Akaike information criterion	418,134.19		412,048.47	402,428.96

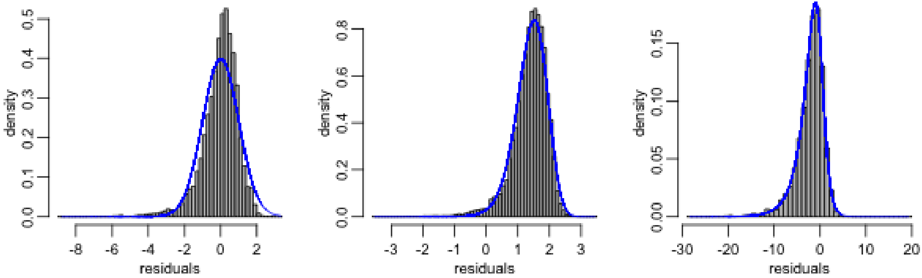
Notes:

a. Face amount is re-scaled in 100,000.

b. Standard errors are in parenthesis.

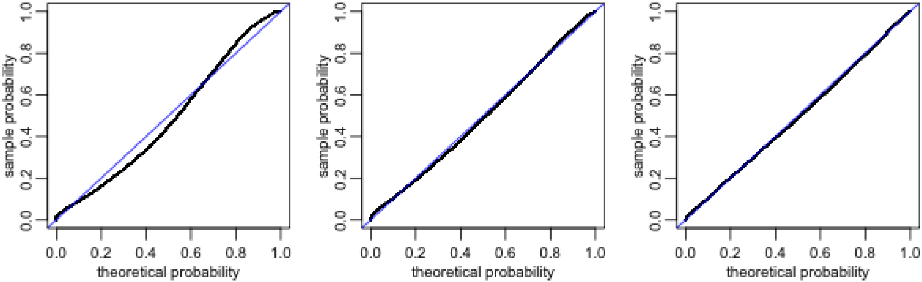
c. An asterisk * identifies ‘not significant’ at the 5% level.

Assessing the quality of the model fit



Log-Normal, Generalized Gamma and GB2, respectively

Assessing the quality of the model fit



PP plots of Log-Normal, Generalized Gamma and GB2, respectively

Observed deaths by issue age and sex

	Mortality status		
Issue Age	Survive	Death	Total
Males			
≤ 30	8,995	672	9,667
30-50	24,341	1,006	25,347
50-70	6,621	733	7,354
> 70	239	69	308
Total	40,196	2,480	42,676
Females			
≤ 30	6,532	349	6,881
30-50	12,202	394	12,596
50-70	2,653	258	2,911
> 70	306	65	371
Total	21,693	1,066	22,759

Maximum likelihood estimation technique

- Maximum likelihood techniques used.
- While we investigated several other parametric models, it boiled down to choosing between the Gompertz and Weibull models.
- Our observable data, $(z_i, t_{w,i}, t_{wd,i}, \delta_i)$, consists of the age at issue, the time of withdrawal, the time of death from withdrawal (if applicable), and a censoring variable.
- For an uncensored observation, the log-likelihood contribution is

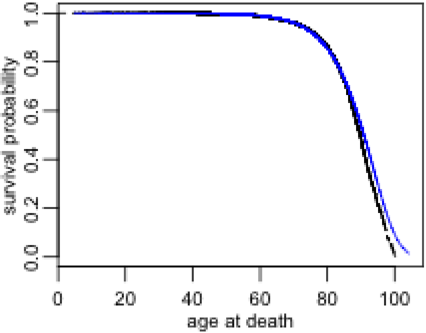
$$\log \frac{f_d(z_i + t_{w,i} + t_{wd,i})}{S_d(z_i + t_{w,i})}.$$

- For a censored observation, it is

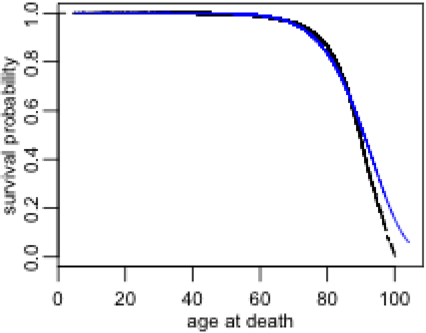
$$\log \frac{S_d(z_i + t_{w,i} + t_{wd,i})}{S_d(z_i + t_{w,i})}.$$

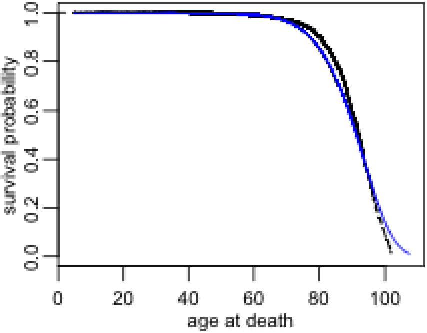
Maximum likelihood estimates

Parameter	Gompertz	Weibull
m^*	93.6031(0.1428)	94.2095 (0.1811)
σ^*	6.8420 (0.0975)	8.3039 (0.1337)
$\sigma^* \times$ Male	0.5206 (0.1161)	0.7507 (0.1481)
Model fit statistics		
Number of observations	65,435	65,435
Log-likelihood	-18,264.55	-18,433.82
Number of parameters	3	3
Akaike information criterion	36,535.11	36,873.63



Gompertz - Male and Gompertz - Female, respectively





Weibull - Male and Weibull -Female, respectively

What do all these results imply?

To understand the implications of results of our models, we examined two items:

- The presence of mortality antiselection: this refers to whether there is greater survival rate after termination of the insurance contract.
  - There is presence of antiselection at withdrawal in life insurance if
    $$S_{d|w}(t_d|t_w) > S_d(t_d), \quad \text{for every } t_d \geq t_w.$$
  - See Carriere (1998) and Valdez (2001).
- The financial cost of policy termination.

Mortality antiselection

To interpret the previous definition:

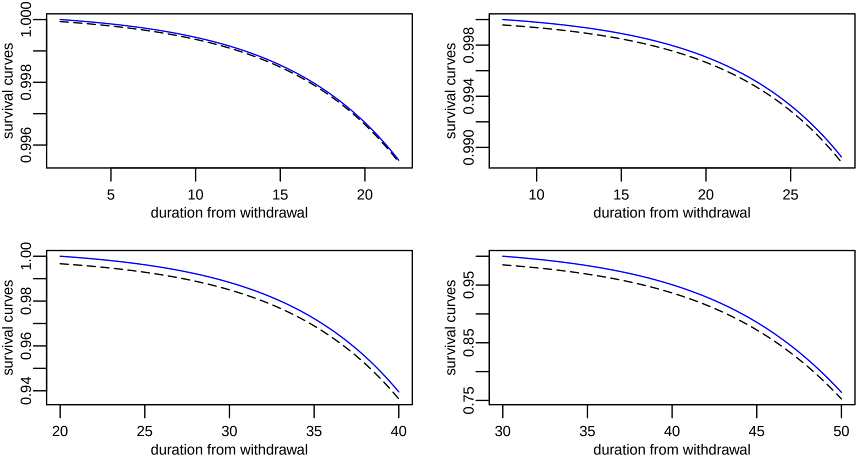
- Antiselection is evidently present when survival of those terminated policies, conditional on all periods of termination, have generally better unconditional survival.

Now, to look for evidence in our data, we consider a specific type of a policyholder with the following characteristics:

- issue age 35, permanent whole life, a non-smoker, male, face amount of 250,000, and not-so-risky with no flat extra charges.

Then, we compare the conditional and unconditional survivorship curves for this policyholder for terminating in different years from issue: withdrawals for years 2, 4, 6, 8, 10, 15, 20 and 30.

Survival curves after policy termination for (35)



For various policy terminations: years 2, 8, 20 and 30.

The financial cost of policy termination

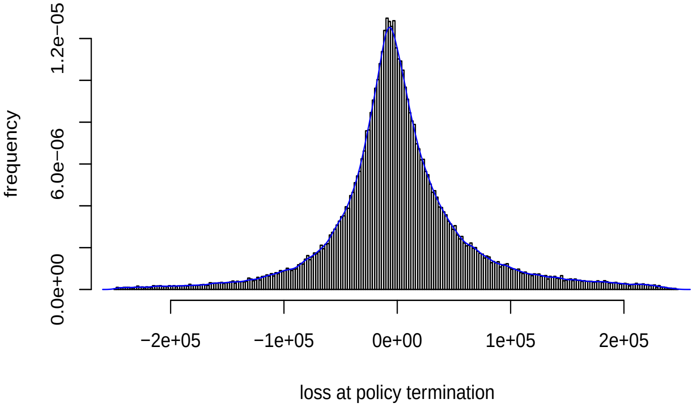
We considered the following case for illustration:

- Issue age 35, male, non-smoker, permanent whole life policy, death benefit of 250,000
- Two types of expenses - assumptions based on Segal (2002, NAAJ):
  - acquisition cost: 80 plus 4.5 per 1,000 of death benefit
  - maintenance expense: 60 plus 3.5 per 1,000 of death benefit
- Interest rate is 5%

Time-until-withdrawal were simulated based on Generalized Gamma. Age-at-death were simulated based on Gompertz.

The financial impact is the loss incurred when policy terminates: accumulated values of all past expenses incurred, plus policy reserves, reduced by the accumulated value of all past premiums paid.

Distribution of the loss at policy termination



Summary statistics of loss at policy termination

Number	Min	Mean	Median	Max	Std Dev
100,000	-249,500	1,223	-3,128	248,000	19,065

Concluding remarks

- We examined and modeled life insurance policy termination and survivorship:
  - time-until-withdrawal - duration models
  - age-at-death - survival models
- Our modeling aspect was driven by the observable data in our dataset. We find that:
  - several policy characteristics do affect policy termination, but not survivorship after policy termination.
- The modeling results can be used for:
  - understanding the presence of mortality selection of policy withdrawal, and
  - predictive modeling of loss upon policy termination.

Acknowledgement

The authors wish to acknowledge financial support from the

- Actuarial Science program at Michigan State University, and
- Janet and Mark L. Goldenson Actuarial Research Center of the University of Connecticut.

Special thanks also to Dr. Yugu Xiao for inviting us to visit your university.