A NEW APPROACH TO ANALYZING PERSISTENCY OF INSURANCE POLICY

By

Loi Soh Loi, Wu Yuan, and Robert Lian Keng Heong

Nanyang Business School, Nanyang Technological University, Singapore

Abstract

Lapsation of life insurance policies has always been a worldwide concern. A study on the profile of the policyholder, policies and agent who produces better persistency based on the Singapore experience was made in 1993, using the regression model.

This paper uses the survival model to evaluate the impact of each factor (e. g. age, sex, type of policy, mode of payment, method of payment, service status etc.) on the survival curves over a period of time. The survival model not only tells us, for example, that policies with male policyholders (M) have a better chance of surviving than policies with female policyholders (F), it is also able to describe the magnitude as well as the incidence of such differential during the policy year. In other words, the survival model provides much more information to the management than what the regression model can offer.

A NEW APPROACH TO ANALYZING PERSISTENCY OF INSURANCE POLICY-[1]

1. Introduction

The persistency of life insurance has always been an important concern. If the premium payment is not persistent, the policy will either lapse or become surrendered or converted to reduced paid up or extended term insurance. Lapsation is extremely costly. The financial impact of lapsation is significant as it adversely affects the policyholder, the company, the agent and the industry in terms of forfeiture of premiums paid, cost of acquisition not fully recovered, loss of renewal commissions and wastage of scarce resources. To the policyholder, he will lose the premiums paid and the loss of insurance protection. To the company, the cost of acquisition may not be recovered. To the agent, he will lose his renewal commissions and job. To the industry, it may hamper the growth of the business. Robert Lian *et al* (1993) have studied an insight of the causes of lapsation and replacement based on the Singapore experience. The findings of the study will enable the companies to portray the profile of the policyholders who are likely to lapse or replace their policies.

Careful study of the causes of lapsation and replacement suggests that some of the analytical procedures can be improved by viewing these causes from a different approach. That is, to identify causes of persistency of the policies underlying the existing study and to introduce "survival analysis" as an enhancement to these problems. The approach outlined here will build on and improve existing policies in force. Survival analysis, when combined with traditional survey approaches, can improve the existing methodology.

[1] The research is supported by the Monetary Authority of Singapore (MAS), the Life Insurance Associate (LIA) and the Nanyang Technological University (NTU), Singapore. The authors are lecturer and senior lecturers at Nanyang Business School, NTU.

This paper consists of six sections. Section One introduces the topic while Section Two explains the methodology used in Robert Lian *et al*. (1993). Section Three introduces survival analysis, a technique we use in this paper to analyze the persistency of insurance policy. The section includes explanations of terminology, models and variables needed to develop a survival analysis study. In Section Four, we use a survival analysis approach to analyze the archival data and examine the effect of independent variables within the survival model. Finally, Section Five covers a discussion of implications of a survival analysis approach to the lapsation problems faced by companies and the industry as a whole. The analysis shows that the results are not only consistent with the results obtained in Robert Lian *et al*. (1993), but also enhance and supplement to the technique used. Section Sx gives the conclusion of the study.

In summary, this paper describes and demonstrates a procedure that allows the company, the agent and the industry to view the persistency of the policyholder from a different perspective.

2. Lapsation

The study of the causes of lapsation (Robert Lian *et al,* 1993), was divided into two parts. The first part identified and analyzed the factors that may have an effect on the persistency of the policies. These factors are listed in Table 1.

Table 1

Factors affecting persistency and their measures

Factor	**Measure of factor**	**Name**
Age at purchase	Age at purchase	AGE
Sex	M - Male F - Female	SEX
Marital status at purchase	M - Married S - Single D - Divorce W - Widowed	MS
Type of policy	WP - Whole life participating WN - Whole life non-participating EP - Endowment participating EN - Endowment non-participating TM - Term OR - Others	TP
Mode of payment	A - Annual S - Semi-annual Q - Quarterly M - Monthly	MP
Status of policy	E - Extended term insurance L - lapsed R - Reduced paid-up insurance S - Surrendered	SP
Method of payment	C - Cash / Cheque G - Giro / Autopay S - Salary deduction B - Bankers' order O - Others	MPT
Service status	S - Serviced O - Orphan	SS
Size of policy	Original sum assured in thousands for basic plan	SA
Duration of policy	Number of completed years from inception	DP

The primary data consisting of the input to the above factors were supplied by the Life Insurance Association (LIA) Singapore, based on all the individual policies terminated in year 1990 due to non-payment of premium. These include policies which have lapsed, surrendered, converted to reduced paid-up or extended term insurance but exclude policies terminated due to death, maturity, expiry or conversion to permanent plans. Fully paid-up policies and single premium policies are also excluded. Altogether there are 48,243 such policies issued by the eleven companies in Singapore.

Multiple regression analysis was used in the 1993 paper to measure the impact of the factors on persistency. A duration model is constructed by considering duration of policy DP as the dependent variable and the other factors the independent variables AGE, SEX, MS, TP, MP, SP, MPT, SS and SA.

The results showed that factors affecting persistency can be classified under the following categories:

- Policyholder - Related

Higher persistency is found among policyholders who are married, male, and older in age.

- Policy - Related

Higher persistency is found among the policies with the following characteristics:- term, non-life (such as medical and disability income types), smaller size, premium paid less frequently and paid by pre-authorized methods e.g. Giro.

- Agent - Related

Higher persistency is found on orphan policies.

In this paper, the factors affecting persistency will be analyzed by a new approach – the survival analysis. In the following section, an overview of survival analysis and the application of such approach to the 48,243 archival data will be given

3. Survival Analysis

An Overview

Survival analysis is a statistical technique which we will use to discuss the treatment of "mortality data". The direct meaning of "mortality data" is data that arise from recording times of death of individuals in a specified group. Individuals in the group may be humans, animals, machines, and so on. More generally we consider situations in which the replacement of "mortality" by the term "failure" is appropriate. A survival model is a probability distribution for a special kind of random variable -- T, the time to failure, for example, the time of death of a person, the time of failure of a manufactured item or the resignation of an employee. In our study the random variable is the number of years of duration of policy (dp). For simplicity, we use T to denote dp in the rest of the paper.

Survival Function

Basically the problem in survival analysis is to estimate a distribution function F(t) of T by parametric methods (say, regression models) assuming F to belong to a particular class of distributions or by nonparametric methods (say, life table). The distribution function F may depend on several independent or risk variables X_i, $i = 1, 2, \ldots, n$. It is depressing to speak of death (failure) but more pleasant to speak of life. In analyzing survival data, the custom has grown not of using the cumulative probability of death but using an equivalent function which is called the *survivor function.* The survivor function is defined by

$$S(t) = P(T \geq t)$$

where S(t) depends on independent variables X_i, it is usually denoted by $S(t, x_1, \ldots, x_n)$.

Hazard Function

In the analysis of survival data, one is often interested in examining which periods have the highest or lowest risk of death. By risk of death, one has in mind the risk or probability among those alive at that time. For example, at a very old age there is a high risk of dying each year among those reaching that age. The probability of any individual dying, say, in the 100^{th} year is small because only a few individuals can live to 100 years old. This concept is made rigorous by the idea of the hazard function or hazard rate. For instance, if the random variable T is continuous, the hazard function is defined as $h(t) = f(t) / S(t)$ where $f(t) = F'(t)$ is the probability density function of T. Using $S(t) = 1 - F(t)$ gives

$$h(t) = -d\,(\ln S(t)) / dt$$

and conversely $S(t) = \exp\left\{-\int_0^t h(u)du\right\}$. Thus S(t) is derivable from h(t) and *vice versa*.

In other words, the hazard function is the probability of dying per unit time given survival to the time point in question. The hazard function is also called the *force of mortality, the age specific death rate*, *conditional failure rate* and *instantaneous death rate.*

Models For The Hazard Function

The simplest model for survival analysis assumes that the hazard function is constant across time, that is $h(t) = \lambda$, where λ is any constant greater than zero. This results in the exponential death density $f(t) = \lambda exp(-\lambda t)$, and the exponential survival function $S(t) = exp(-\lambda t)$. Graphically, the hazard function and the death density function are displayed in Figure 1. This model assumes that the fact of having survived up to a given point in time has no effect on the probability of dying in the next instant.

Figure 1 a: Hazard Function for Exponential Distribution with λ = 1

Figure 1 b: Death Density Function for Exponential Distribution with λ = 1

(Source of figures: A.A. Afifi, Virginia Clark, 1990)

If the hazard function is not constant, the Weibull distribution should be considered. For this distribution, the hazard function may be expressed as $h(t) = \alpha\lambda(\lambda t)^{\alpha-1}$. The expressions for the density function, the cumulative distribution function, and the survival function can be found in most of the research papers. Figure 2a and 2b show plots of the hazard and density functions for $\lambda = 1$ and $\alpha = 0.5$, 1.5, and 2.5. The value of α determines the shape of the distribution. Furthermore, as may be seen in Figure 2.a, the value of α determines whether the hazard function increases or decreases over time. Namely, when $\alpha < 1$ the hazard function is decreasing and when $\alpha > 1$ it is increasing. When $\alpha = 1$ the hazard is constant, and the Weibull and exponential distributions are identical. The Weibull distribution is used extensively in practice. In this paper we use a model in which this distribution is assumed.

Figure 2 a: Hazard Function for Weibull Distribution

with $\lambda = 1$ and $\alpha = 0.5$, 1.5, and 2.5

Figure 2 b: Death Density Function for Weibull Distribution

with λ = 1 and α = 0.5, 1.5, 2.5

(Source of figures: A.A. Afifi, Virginia Clark, 1990)

A good way of deciding whether of not the Weibull distribution fits a set of data is to obtain a plot of *log (-logS(t))* versus log time, and check whether the graph approximates a straight line. If it does, then the Weibull distribution is appropriate and the methods described in the following section can be used.

The Log-Linear Regression Model

In this section, we describe the use of multiple linear regression to study the relationship between survival time and a set of independent variables.

Suppose that T is survival time and $X_1, X_2 ...., X_p$ are the independent or risk variables.

Let $Y = log(T)$ be the dependent variable, where natural logarithms are used, then the model assumes a linear relationship between *log(T)* and the *X*'s. The model equation is

$$\log(T) = \alpha + \beta_1 X_1 + \beta_2 X_2 .....+ \beta_p X_p + e$$

where e is an error term.

This model is known as the log-linear regression model since the log of survival time is a linear function of the X's. If the distribution of *log(T)* is normal, it would be possible to use the regression methods to analyze the data. However, in most practical situations *log(T)* is usually not normally distributed (T is often assumed to have a Weibull distribution). For these reasons, the method of maximum likelihood is used to obtain estimates of β_i's and their standard errors.

4. Survival Analysis: An application

We apply both parametric and nonparametric methods to analyze the distribution of T (i.e. dp- the duration of policy) of the archival data which consist of 48,243 policy-holders from eleven life assurance companies in Singapore. For parametric method, we use the log-linear regression model to study the relationship between T and the factors that may have effect on the persistency of the policies (refer to Table 1). For nonparametric method, we use the life table to analyze the survival and hazard functions of T which are affected by different factors. The LIFEREG and LIFETEST procedures in SAS are used for both methods to analyze the data as follows:

The LIFEREG procedure of SAS is used for parametric methods. There are 47,529 out of 48,243 data without missing values. We analyzed the data without considering the marital status since most of the companies do not have the records of this variable. Since the plot of *log(-logS(t))* versus log time in our data set is a straight line, the Weibull distribution is appropriate for the survival curve. Therefore, we used the Weibull model to fit the data set.. Table 2 displays the resulting output. Shown are the maximum-likelihood estimates of the intercept (α) and regression (β_i) coefficients along with their estimated standard errors, and the probability of a larger chi-square value (Pr > Chi) testing whether the corresponding parameters are zero. The coefficients (or estimates) within each factor take the value of zero, positive or negative. Take for example, in the case of mode of payment (MP), the coefficient of semi-annual mode policy (S) with a certain duration of policy say dp(s) was set to zero before the test. After the test, the finding showed that annual mode policy (A) has a positive coefficient of 0.1034 and monthly mode policy (M) a negative coefficient of 0.3781. Stated in insurance language, this means that policy (A) will on the average last 0.1034 year longer than policy (S) while policy (M) 0.3781 year shorter than policy (S).

Table 2 (Without considering the marital status)

Variables	Estimate	Standard Error	Pr >Chi
INTERCEPT	2.3117	0.0175	0.0001
AGE	0.0024	0.0003	0.0001
SEX			
F - Female	-0.0672	0.0058	0.0001
M - Male	0		
TP			
EN - Endowment non-participating	-0.0210	0.04326	0.6267
EP - Endowment participating	-0.0629	0.0067	0.0001
OR - Others	0.1619	0.0245	0.0001
TM - Term	0.6871	0.0117	0.0001
WN - Whole life non-participating	-0.0191	0.0085	0.0253
WP - Whole life participating	0		
MP			
A - Annual	0.1034	0.0122	0.0001
M - Monthly	-0.3781	0.0119	0.0001
Q - Quarterly	0.0036	0.0150	0.8100
S - Semi-annual	0		
SP			
E - Extended term insurance	-0.1728	0.00154	0.0001
L - lapsed	-1.7004	0.0068	0.0001
R - Reduced paid-up insurance	-0.0974	0.0127	0.0001
S - Surrendered	0		
MPT			
B - Bankers' order	0.1519	0.02492	0.0001
C - Cash / Cheque	-0.4327	0.0090	0.0001
G - Giro / Autopay	-0.1544	0.0089	0.0001
O - Others	0.01379	0.04267	0.7465
S - Salary deduction	0		
SS			
O- Orphan	0.2600	0.0099	0.0001
S- Serviced	0		
SA	0.0000 *	0.0001	0.6575

** value only be considered up to four decimals.*

Comparisons of the estimates of the coefficient of the significant variables on Table 2 with the estimates obtained in Robert Lian *et al*. (1993) shows that the results are consistent which each other: higher persistency is found among policyholders who are older in age ,male and having the policies with term coverage, smaller size, premium paid less frequently, paid by pre-authorized methods and not serviced by agent..

The log-linear regression equation can be used to estimate typical survival times for selected cases. For example, for a policyholder of age 30, the minimum case (shortest survival time) occurs when each explanatory variable has the minimum coefficient, i. e., the policyholder is a female (SEX = F) holding an endowment participating policy (TP = EP) with a monthly mode premium (MP =M) payable by cash (MPT = C) and was served by agent (SS = S) before lapsing her policy (SP = L). We then have log(t) = - 0.2526. Since this is a natural logarithm, the survival time will be t = exp.(-0.2526) = 0.7768 years. On the other hand, the maximum case (longest survival time) occurs when each explanatory variable has the maximum coefficient, i. e., policyholder is a male with the same age holding a term policy (TP = TM) with an annual mode premium (MP = A) payable by bankers' order (MPT = B) and was not serviced by agent (SS = O) before surrendering his policy (SP = S), we then have log(t) = 3.5861 and the survival time will be t = exp.(3.5861) = 36.093 years. Intuitively, it is more likely that the estimate will fall within these two extreme ends.

However, since term (TM) policy has no cash value and cannot be surrendered, it may be substituted by whole life participation (WP) policy which has also a greater coefficient than endowment participating (EP). Also, since GIRO (G) is much more popular than bankers' order (B) and also has a greater coefficient than cash payment (C), (B) may be substituted by (G). The probable maximum case will then have log(t) = 2.5927 or a survival time of t = 13.366 years. It should be noted that the above minimum and maximum survival time are applicable to age 30. For older policyholders, the survival times will be longer and vice versa.

The nonparametric survival analysis techniques can be derived from the life table analysis procedures, and are available in SAS - the LIFETEST procedure. Estimation of the survival distribution does not require that all sample policies start at the same point in time and follow over a fixed period in time interval, as would be true when following a specific cohort of new policies. Instead, survival analysis focuses on the length of duration, regardless of when the policies were bought. The life table (see Table 3) provides more specific information about the survival experience of the sample. Information provided in the life table includes:

- t_i, - the ith period (each is one year).
- n_i , - the total number of units in the study at the beginning of each t_i year.
- d_i , - the number of units that fail in the t_i year period, i.e. the number of policies lapsed.
- $q_i = d_i / n_i$ - the estimated probability that a unit will terminate a police in the t_i year.
- $S(t_i) = \Pi_{i=1}^{i}(1 - q_i)$ - the cumulative survival rate at the end of the t_i year.
- $f(t_{mi}) = S(t_{i-1})q_i / b_i$, - PDF at the middle of the ith period, where $b_i = t_i - t_{i-1.}$.
- $h(t_{mi}) = 2q_i / (b_i(1 + p_i))$ where $p_i = 1 - q_i$ hazard at the middle of the ith period.

Table 3 Life Table Survival Estimates For All Policy holders

t_i	n_i	d_i	q_i .	$S(t_i)$	$f(t_{mi})$	$h(t_{mi})$
0	47529	11254	0.2368	1.0000	0.2368	0.268579
1	36275	14349	0.3956	0.7632	0.3019	0.493084
2	21926	6608	0.3014	0.4613	0.1390	0.354849
3	15318	4201	0.2743	0.3223	0.0884	0.317836
4	11117	2647	0.2381	0.2339	0.0557	0.270281
5	8470	1713	0,2022	0.1782	0.0360	0.224995
6	6757	1410	0,2087	0.1422	0.0297	0.232981
7	5347	1292	0.2416	0.1125	0.0272	0.274835
8	4055	931	0.2296	0.0853	0.0196	0.259368
9	3124	617	0.1975	0.0657	0.0130	0.219144
10	2507	542	0.2162	0.0527	0.0114	0.242397
11	1965	481	0.2419	0.0413	0.0101	0.278921
12	1484	391	0.2448	0.0312	0.00823	0.303454
13	1093	279	0.2553	0.0230	0.00587	0.292606
14	.814	179	0.2199	0.0171	0.00377	0.247067
15	635	166	0.2614	0.0134	0.00349	0.300725
16	469	140	0.2985	0.00987	0.00295	0.350877
17	329	112	0.3404	0.00692	0.00236	0.410256
18	217	83	0.3825	0.00457	0.00175	0.472934
19	134	43	0.3209	0.00282	0.000905	0.382222
20	91	24	0.2637	0.00191	0.000505	0.303797

If the plots were requested, then printer plots are generated for the estimated survival distribution function (SDF) against the duration time (T), or the -log(SDF) against T or the log(-log(SDF)) against log(T).

A survival function rate $S(t_i)$ reflects the cumulative survival probabilities across time, the graph of a survival function (SDF) against T can provide important statistical information as well. For instance, Figure 3 portrays the survival function of 47,529 policies over a 20-year period. One can easily determine from the shape of the survival function which time periods present a high risk and lower survival rates (steep slope) and which have a relatively low risk (flat area in the curve).

Another statistic commonly used in survival analysis for each specific time period is the 'hazard rate' The hazard function rate $h(t_i)$ is a conditional probability that policies will terminate in a specific time period given that they have survived to the beginning of that time period. Hazard function can be plotted in the same manner as survival functions. Figure 3 portrays the hazard rates for the 47,529 policy holders. A constant hazard rate would reflect a random pattern to termination. Hazard rate plots which have one or more peaks identify time periods with a high risk of termination. It is clearly showed in Figure 3 that there are a few high risk of termination time periods, say, 1^{st} , 18^{th} years. It will be sufficient to consider the time period after including these two high peaks at the tail part of the hazard function because the numbers of the survival are getting lesser at a later duration and the effect is minimal. Therefore, a duration up to 20 years is selected.

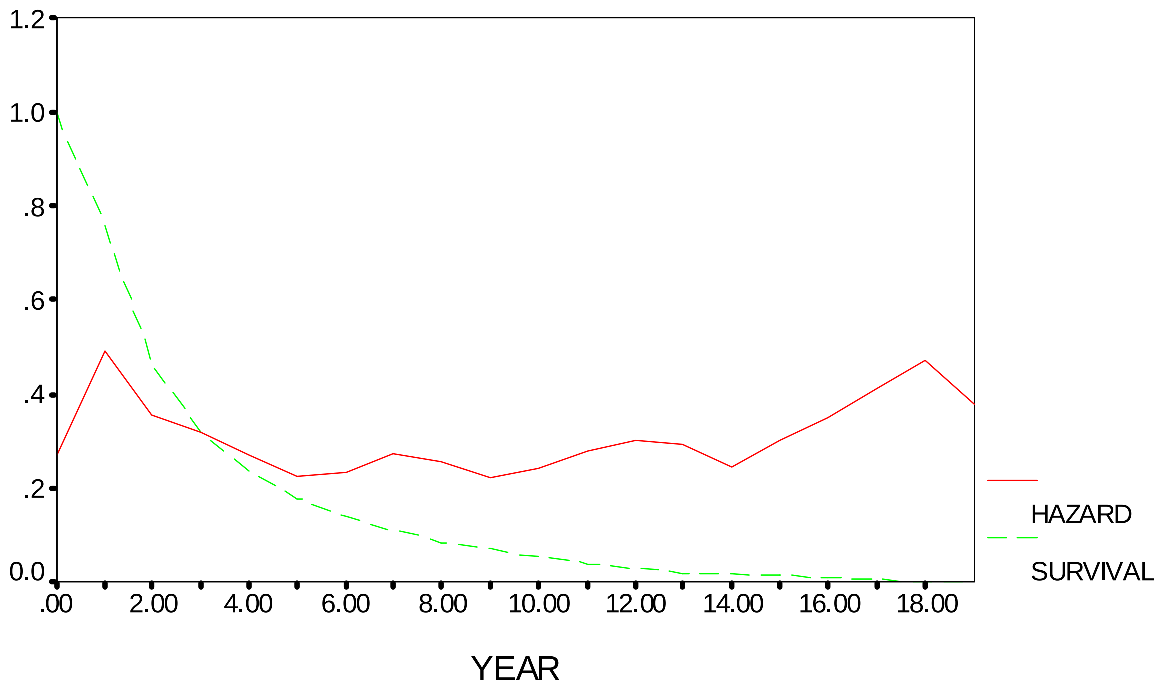

Figure 3: Survival and Hazard Functions for all policy holders

Survival analysis can also assess the impact of independent variables on the survival distribution. This is done by creating subgroups of policies based on these independent variables. Different factors for different samples have different survival rates. This allows an evaluation of different factors (independent variables), such as whether the gender, type of policy, mode of payment, status of policy, method of payment or service status would affect the survival rate. The following sections show how the survival rates are affected by different groups of factors. The life tables survival estimates for different subgroups are listed in Appendix at the end of the paper.

- **Comparing the survival rates between the gender (SEX).**

From the life tables for Female (F) or Male (M) policy holders as shown in Table 4a and 4b, the results indicate that there is a difference in survival rates between male and female. This can be seen from the plot of the two survival rates in Figure 4. The survival rate for the male is slightly higher than for the female throughout the duration. This result may be expected as the male tends to stay on to the job longer than the female and therefore has a more regular income to pay the premium.

Table 4:a Life Table Survival Estimates for Female (F)

t_i	n_i	d_i	q_i	$S(t_i)$	$f(t_{mi})$	$h(t_{mi})$
0	15108	3821	0.2529	1.0000	0.2529	0.289525
1	11287	4643	0.4114	0.7471	0.3073	0.517874
2	6644	2171	0.3268	0.4398	0.1437	0.390573
3	4473	1455	0.3253	0.2961	0.0963	0.388466
4	3018	787	0.2608	0.1998	0.0521	0.299867
5	2231	489	0.2192	0.1477	0.0324	0.246162
6	1742	397	0.2279	0.1153	0.0263	0.257208
7	1345	380	0.2825	0.0890	0.0252	0.329004
8	965	256	0.2653	0.0639	0.0169	0.305854
9	709	162	0.2285	0.0469	0.0107	0.257962
10	547	128	0.2340	0.0362	0.00847	0.26501
11	419	116	0.2768	0.0277	0.00768	0.32133
12	303	75	0.2475	0.0201	0.00496	0.282486
13	228	51	0.2237	0.0151	0.00338	0.251852
14	177	44	0.2486	0.0117	0.00291	0.283871
15	133	37	0.2782	0.00880	0.00245	0.323144
16	96	34	0.3542	0.00635	0.00225	0.43038
17	62	29	0.4677	0.00410	0.00192	0.610526
18	33	12	0.3636	0.00218	0.000794	0.444444
19	21	11	0.5238	0.00139	0.000728	0.709677
20	10	1	0.10000	0.000662	0.000066	0.105263

Table 4.b: Life Table Survival Estimates for Male (M)

t_i	n_i	d_i	q_i	$S(t_i)$	$f(t_{mi})$	$h(t_{mi})$
0	32421	7433	0.2293	1.0000	0.2293	0.2558949
1	24988	9706	0.3884	0.7707	0.2994	0.482046
2	15282	4437	0.2903	0.4714	0.1369	0.339649
3	10845	2746	0.2532	0.3345	0.0847	0.289907
4	8099	1860	0.2297	0.2498	0.0574	0.25945
5	6239	1224	0.1962	0.1924	0.0378	0.217523
6	5015	1013	0.2020	0.1547	0.0312	0.224687
7	4002	912	0.2279	0.1234	0.0281	0.257191
8	3090	675	0.2184	0.0953	0.0208	0.245232
9	2415	455	0.1884	0.0745	0.0140	0.208
10	1960	414	0.2112	0.0605	0.0128	0.236167
11	1546	365	0.2361	0.0477	0.0113	0.267693
12	1181	316	0.2676	0.0364	0.0975	0.308895
13	865	228	0.2636	0.0267	0.00703	0.303595
14	637	135	0.2119	0.0196	0.00416	0.23705
15	502	129	0.2570	0.0155	0.00398	0.294857
16	373	106	0.2842	0.0115	0.00327	0.33125
17	267	83	0.3109	0.00824	0.00256	0.368071
18	184	71	0.3895	0.00568	0.00219	0.478114
19	113	32	0.2832	0.00349	0.000987	0.329897
20	81	23	0.2840	0.00250	0.000709	0.330935

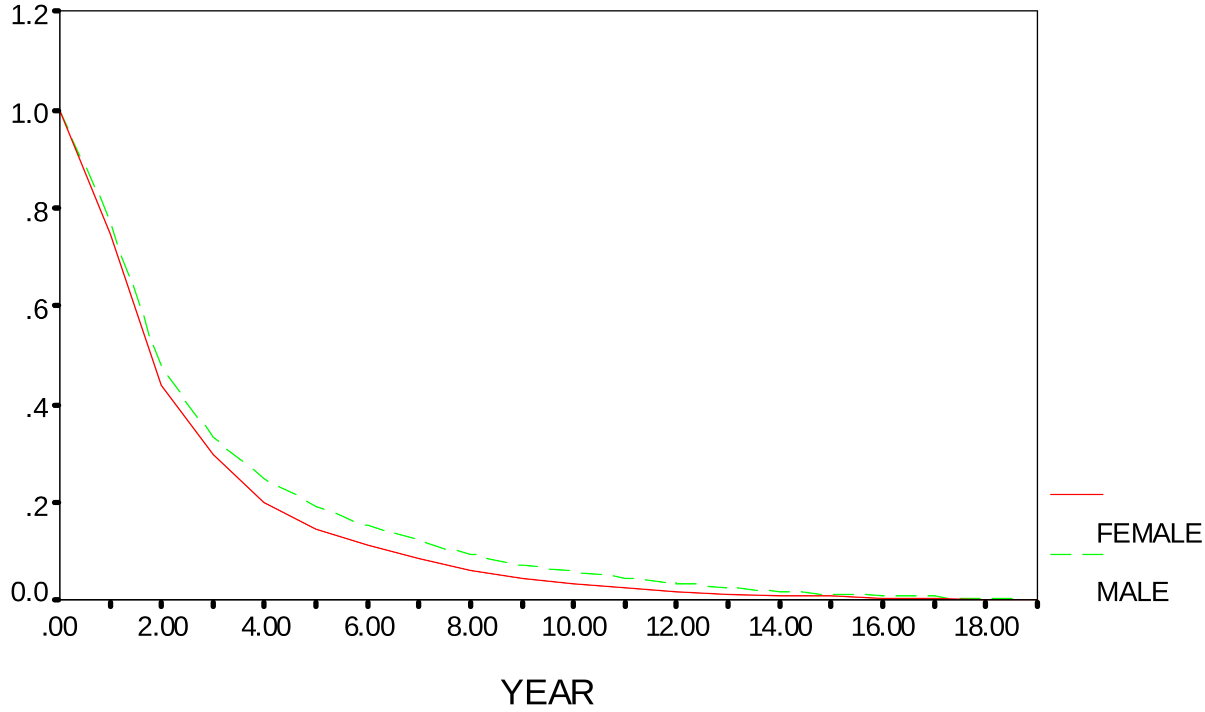

Figure 4: Comparing the Survival Rates by Sex of Policyholder (SEX)

- **Comparing the survival rates between the type of policy (TP).**

The life tables for different types of policies which include EN, EP, OR, TM, WN and WP show that there are differences in survival rates among the types of policies. This can be seen from the plot of the six survival rates in Figure 5. The survival rates for EP is higher than that of WP within the duration of three to thirteen years. This result may be expected as EP is purchased mainly for the purpose of savings with a target amount to save up and an incentive (e. g. maturity bonus) to do so and therefore has a better persistency

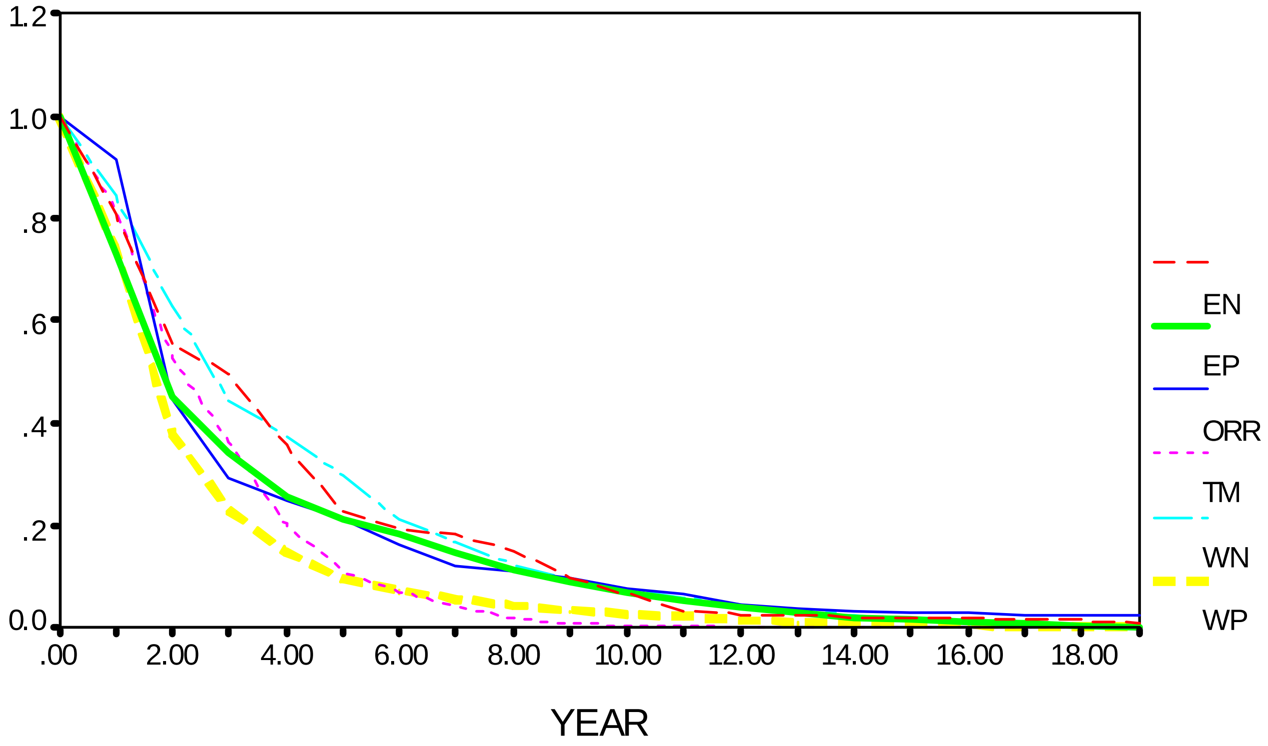

Figure 5: Comparing the Survival Rates by Type of Policy (TP)

- **Comparing the survival rates between the mode of payment (MP).**

The life tables for different modes of policy which include A, M, Q and S indicate that policies paid monthly have the lowest survival rate while those paid annually have the highest survival rate among four different modes of payment. This can be seen from the plot of the four survival rates in Figure 6. This result is again expected as the policyholder paying monthly has twelve times the opportunity to stop paying as compared to those paying annually.

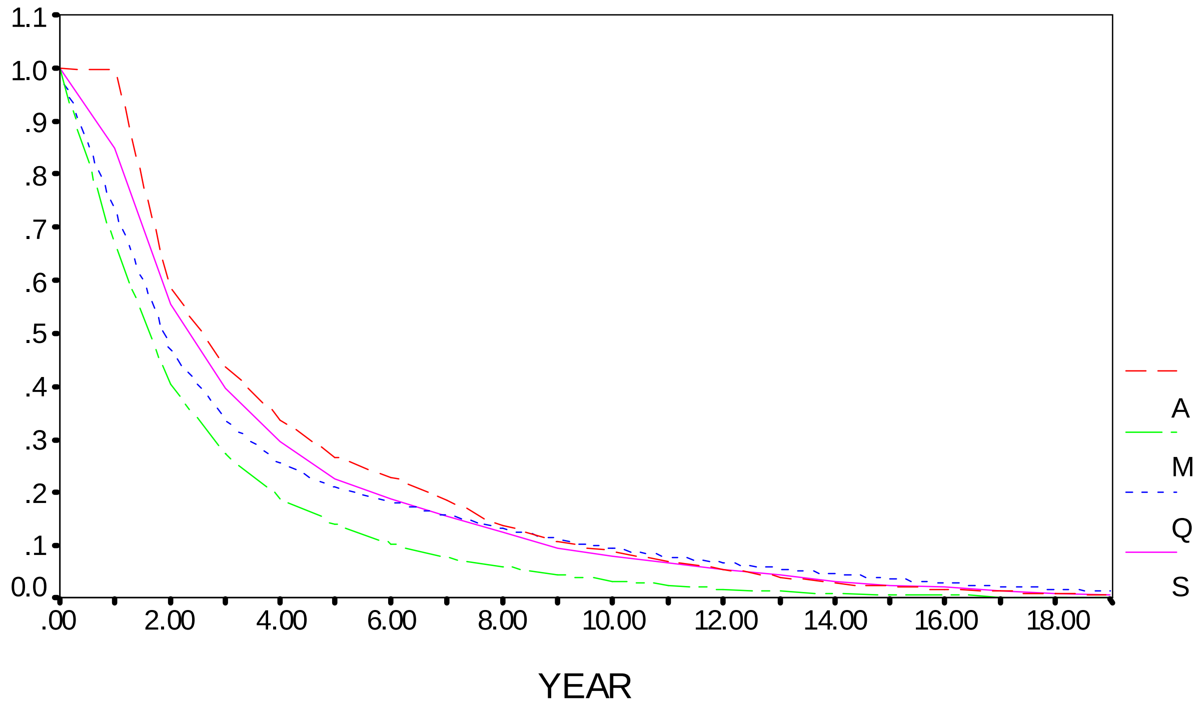

Figure 6: Comparing the Survival Rates by Mode of Policy (MP)

- **Comparing the survival rates between the status of policy (SP).**

The life tables for status of policy which include E, L, R and S indicate that the lapsed policy (L) has a survival rate much lower than the others. This can be seen from the plot of the four survival rates in Figure 7. This has to be true because a policy is defined as lapsed when it has not acquired cash value. Since policies with cash value must acquire cash value after they are kept in force for three years as required by the insurance law in Singapore; they can only lapse within the first three years. This explains why the survival rate for lapsed policies drops to almost zero by end of year three. The reason why it has not actually dropped to zero is because some policies with term or medical coverage have no cash value and can lapse anytime even after three years. The other finding that survival rates for E, R and S are fairly close to each other is also expected because by definition, extended term insurance (E) and reduced paid up (R) are the other non-forfeiture options the policyholder can choose if he/she does not want to surrender. They therefore belong to the same family and expected to exhibit similar characteristic.

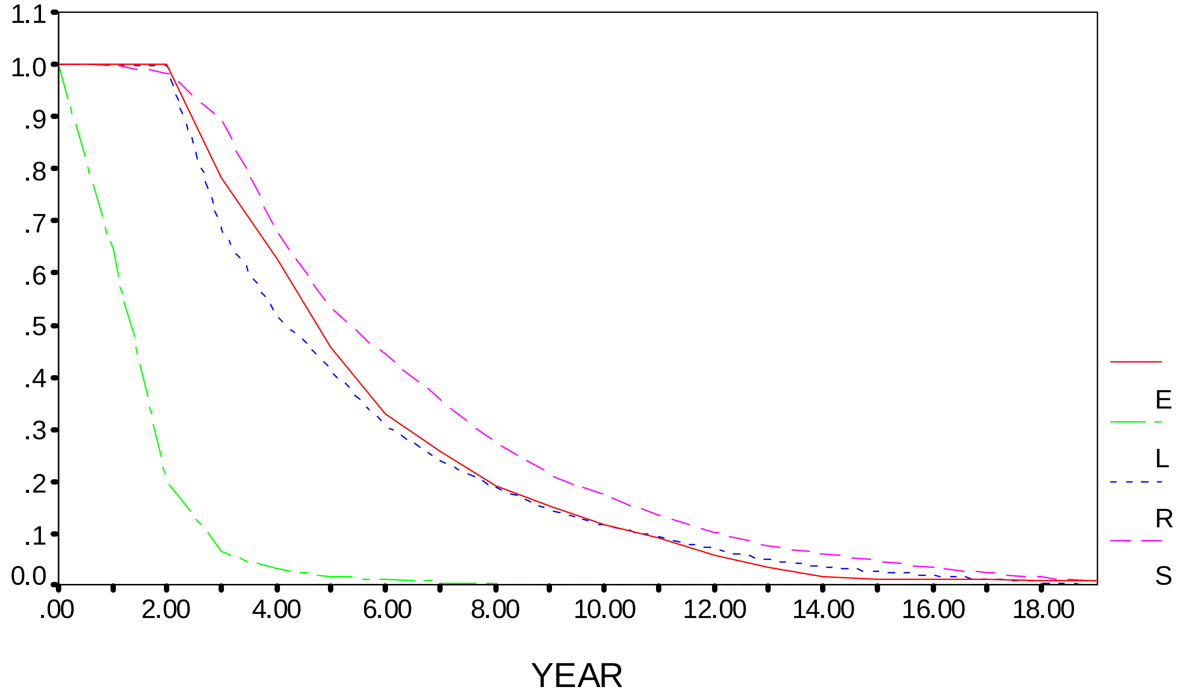

Figure 7: Comparing the Survival Rates by Status of Policy (SP)

- **Comparing the survival rates between the method of payment (MPT).**

The life tables for method of payment which include B, C, G, O or S show that the survival rate for those policy holders who make payment through pre-authorized arrangement such as G or S are higher than those who pay by cheque or cash i.e. C, within the first few years. This can be seen from the plot of the five survival rates in Figure 8. This may be expected because the policyholder has to make an effort to revoke the pre-authorized arrangements in order to lapse his/her policy. Due to automative deduction and the human inertia to change, the company can expect better persistency from this group of policyholders. Though inertia and the advantage of pre-authorized arrangement losses their effect after five years for G and fifty years for S, they are long enough to help the company to recover its first year high acquisition costs if the survival rate differential of G and S is high enough compared to C.

Note that though the survival rate for other payment methods O i.e. credit card or NETS, is noticeably higher than the others, in terms of volume it is too small to be reliable or to have any impact.

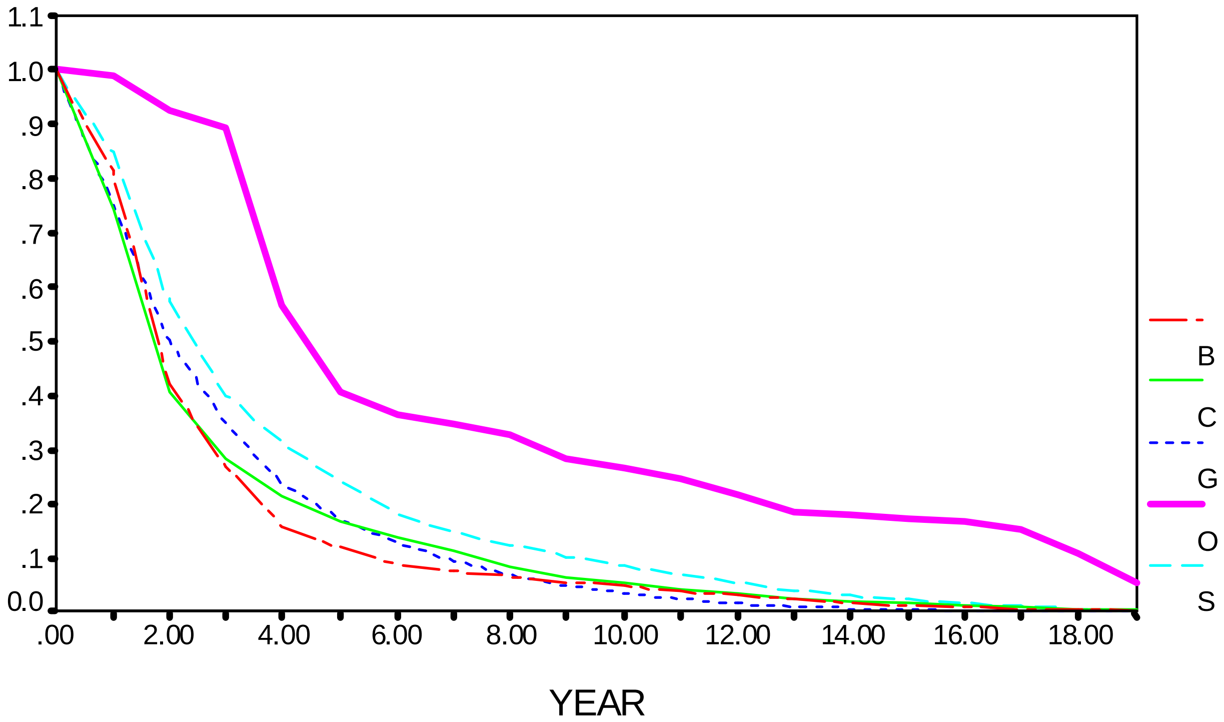

Figure 8: Comparing the Survival Rates by Payment of Policy (MPT)

- **Comparing the survival rates between the serviced and orphan status (SS).**

The life tables for the policy of serviced versus the orphan status give results that indicate a significant difference in survival rates between serviced and orphan status. This can be seen clearly from the plot of the two survival rates in Figure 9. The survival rate for the orphan status is higher than that for the serviced status for all durations. Intuitively, a company would expect the duration of policy for serviced status to last longer than the orphan status policy, but from the “statistically” analysis we obtain the reverse results. Further “mental” analysis will tell us that the orphan policyholder without the agent serving does not have to face the risk of the agent who instead of providing the service on the existing policy, would advise him/her to lapse/surrender and replace it with a new policy. This is certain a “good” source for replacement of policy. Control for such undesirable practice will have to come from highly emphasized agent training to refrain them from policy replacement, besides the restriction of first year commission within one year before or after the lapsation /surrender /ETI /RPU of existing policy.

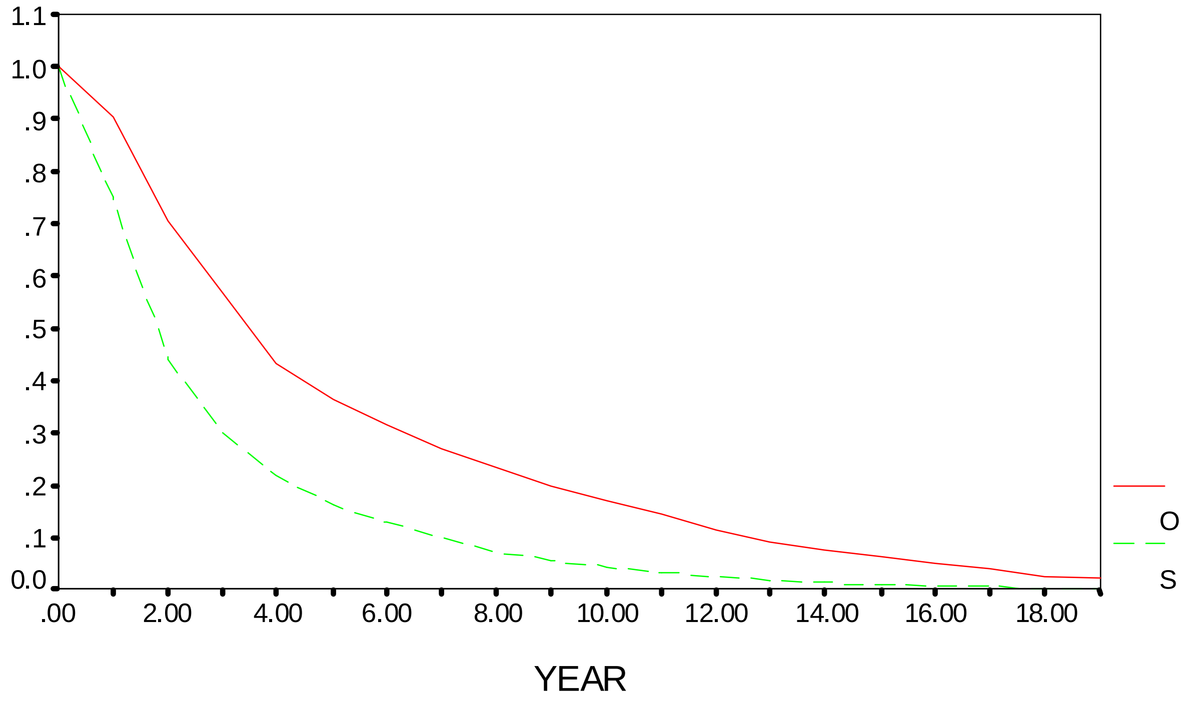

Figure 9: Comparing the Survival Rates by Service Status (SS)

From the results given above, certainly the persistency of the policy is an important goal in the insurance industry. Survival analysis allows the industry, agents and management to examine directly the effect of different independent variables on persistency of life policies.

- **Advantage of Survival Analysis**

In addition to providing the above information about duration of holding a policy, survival analysis provides a new perspective on the persistency problem. Survival analysis provides management with relevant information about aggregate policies rather than individual policies. Agents can thus focus upon variables related to the persistency of aggregate policies within their units. By analyzing the life table and the plot of the survival and hazard functions, lots of information can be gleaned to profile the policyholder, product and agent, that produce better persistency for the industry or company

5. Discussion

Lapsation of policy has always been a major concern in life insurance. Lapsation can be very disruptive in the industry or agencies. Not only does a lapsation affect customer-agent relationships, it also leaves the industry open to criticism.

Survival analysis has the ability to examine the effect of time varying variables. Different variables are thought to exert different pressures at different points in duration. For example, annual payment mode is a key variable in keeping the policy in force longer than the other modes of payment.

Studying persistency of life policy using the survival analysis model involves a conceptually different way of examining the problem of lapsation. Survival analysis not only identifies the key variables that can significantly affect the survival of the policy, it can also measure the magnitude and incidence if the policy would have a better or poorer chance of survival under the influence of a particular variable than other variables during the policy years.

The lapsation indicates how often policy would lapse. The agent's benefit of persisting the policy comes from discovering ways to improve the persistency or duration rate of policy. While survival analysis approach does not provide all the answers to persistency/lapsation problems, it does provide a methodology that may provide better managerial answers for company or industry. Survival analysis is not a new or controversial approach for most disciplines. However, the infrequent use of survival analysis in the field of insurance provides an opportunity for new research on insurance.

We also take note that the survival analysis approach is not without its problems. Survival analysis requires a data set with extensive data over a long period of time. However, many companies have the necessary data available in archival form. The typical personnel data may include date of buying life policy, current marital status, lapsation dates etc.. In addition, the validity of any longitudinal analysis increases with the length of the study period. The problem of obtaining a large and comprehensive data set is common to all studies of policy force, and does not represent a unique disadvantage of survival analysis.

6. Conclusion

To date the arena of life policy research has seen a great deal of activity. Unfortunately, this activity has left most researchers and agents wanting more guidance and understanding. The absence of an appropriate methodological tool to study the unique occurrence of policy lapsation contributes to this problem.

A research technique developed in the biomedical life sciences possesses strengths that can allow insurance researchers to move into a new and more productive era.

References

A.A. Afifi, Virginia Clark "Computer-Aided Multivariate Analysis" 2nd edition,1990, Vann Nostrand Reinhold.

Robert Lian Keng Heong, Loi Soh Loi, Adrian Wong Yoke Wai & Hsu Wei Ching "Lapsation and Replacement - What are the causes? The Singapore Experience", *The Institute of Actuaries of Australia, Quarterly Journal*, Sept. 1993, pp 22 - 48.

SAS User's Guide: Statistics, Version 6 Edition.