2004

Decision Tree Analysis of Terminated Life Insurance Policies

Robert Keng Heong Lian
Nanyang Technological University, akhlian@ntu.edu.sg

Yuan Wu
Nanyang Technological University, aywu@ntu.edu.sg

Hian Chye Koh
Nanyang Business School, ahckoh@ntu.edu.sg

Journal of Actuarial Practice Vol. 11, 2004

Decision Tree Analysis of Terminated Life Insurance Policies

Robert Keng Heong Lian,* Yuan Wu,† and Hian Chye Koh‡

Abstract§

Statistical methods such as regression and survival analysis have traditionally been used to investigate the factors affecting the duration of terminated life insurance policies. This study explores a different approach: it uses a more recently developed data mining technique called decision trees. By sequentially partitioning the data to maximize differences in the dependent variable (duration in this study), the decision trees technique is good at identifying data segments with significant differences in the dependent variable. This identification can be useful when a company is trying to understand the factors driving or associated with the termination of life insurance policies. Decision trees also have an advantage over other techniques such as linear regression in their ability to detect nonlinear and other complex relationships that are more likely to exist in any practical data set.

Key words and phrases: *data mining, life insurance policies, termination, persistency, lapses*

---

*Robert Keng Heong Lian, M.Sc., A.S.A., is a associate professor emeritus at Nanyang Technological University. He was the general manager of British American Life Assurance in Singapore before joining the university. His research interests include health care financing and retirement planning.

Prof. Lian's address is: Nanyang Business School, Nanyang Technological University, Singapore, SINGAPORE 639798. E-mail: *akhlian@ntu.edu.sg*

†Yuan Wu, Ph.D., is an associate professor at Nanyang Technological University. His research interests include traffic accident analysis and application of statistics to business and finance.

Dr. Wu's address is: Nanyang Business School, Nanyang Technological University, Singapore, SINGAPORE 639798. E-mail: *aywu@ntu.edu.sg*

‡Hian Chye Koh is an associate professor and vice-dean of the Nanyang Business School. His research interests include data mining applications, statistical analysis and modeling in accountancy and business, and behavioral research.

Prof. Koh's address is: Nanyang Business School, Nanyang Technological University, Singapore, SINGAPORE 639798. E-mail: *ahckoh@ntu.edu.sg*

§The authors wish to thank the anonymous referee for helpful comments and suggestions that have improved this paper.

1 Introduction

From a management point of view, in order to contribute to the profitability of a business, quality sales are required. For life insurance, quality sales means policies sold that remain in force long enough to at least allow the life insurance companies to recoup their high acquisition costs. If a policy lapses (i.e., is terminated because of non-payment of premiums) too early, the cost can be high to all parties. For the policyholders, the premiums paid will be forfeited. For the company, on the other hand, the cost of acquisition will not be fully recovered. As a consequence, policyholders who persist (i.e., who keep their policies in force) must bear the unrecovered expenses, thus subsidizing those who terminate their policies. For agents, there will be loss of renewal commissions and, in extreme cases, maybe even a loss of their jobs. Therefore, it will be advantageous for life insurance companies to know the factors affecting the persistency of policies and work toward increasing persistency.

Motivated by the importance of persistency, the Life Insurance Association of Singapore (LIAS) has initiated a study on the termination of policies. In order to carry out the study, the LIAS has collected data on policies terminated because of non-payment of premiums. The data set includes individual factors such as policyholder age at purchase, gender, duration of policy at termination, type of policy, etc.

Several statistical methods such as regression analysis (Renshaw and Haberman, 1986, and Lian et al., 1993) and survival analysis (Lian, Wu, and Loi, 1998) have been used to analyze terminated life insurance policies. In the last decade, however, three new interrelated areas of data analysis have emerged that have emphasized obtaining more information from a data set: data warehousing, knowledge management, and data mining. Of particular interest to us is data mining, which aims to identify valid, novel, potentially useful, and understandable correlations and patterns in data (Chung and Gray, 1999). Data mining techniques such as neural networks and decision trees provide a different approach to predictive modeling. This study uses decision trees (also known as recursive partitioning algorithm) to analyze the LIAS data set.

2 Termination and Persistency of Policies

Although persistency is an important issue in insurance, few academic research studies have been done in this area. Instead, this issue has been discussed in trade journals where authors have emphasized the agency force and described how it can help to reduce terminations.

Willet (1986), for example, stressed the importance of good personal communication with clients and prospects in creating goodwill, faith, trust, and high persistency as compared to a telephone recorder answering system. Ingram (1987), in examining the relationship between effort and compensation, found that agents chose activities that maximize return of their time and effort. He suggested that new concepts of life insurance sales be developed to focus on the winning of new clients and the building of value when clients repeatedly pay premiums and purchase additional policies. Santos (1992) believed that life insurance agents could prevent higher termination rates by maintaining an attitude of service, being thorough in determining a client's needs, and by presenting a clear, systematic explanation of the proposed solution in the spirit of trust and friendship. Santos added that agents should also listen, interact, and promise to stay in touch to monitor their clients' needs and to adjust their insurance programs.

Hall (1990) and Berlin and Powell (1990) expressed their concern on the importance of market research and how it could help to improve persistency. Hall (1990) urged life insurance researchers to stay attuned to factors such as unemployment caused by mergers, downsizings, and closures, which contribute to sharp declines in income and can affect terminations. Berlin and Powell (1990) emphasized the importance of the evaluation of persistency to monitor the overall quality of the industry's marketing and service efforts. The direction and degree of persistency rates reflect how well a company and its agents excel in matching products and services with the insurance needs of its policyholders.

Goodwin (1993) objected to insurers coercing agents (against their code of ethics) to encourage clients to retain insurance policies that are against their clients' best interests. He suggested that if a penalty is to be imposed on an agent for a policy termination, there should be a scale of commission forfeitures keyed to the age of the terminated policy. Sweeney (1996) opined that adverse publicity over improper sales practice had worked against persistency and had affected the sales of new products. Zinseer (1996) suggested that improvement in the persistency of existing business could have a 35% impact on pre-tax earnings. He attributed the main reason for the delay in insurers exploiting their existing customer base to the lack of skills and the scale of field officers who were responsible for managing customers.

Lautzenheiser and Barks (1991) and Howard (1997) focused on how product design could encourage policyholders to keep their policies longer. Lautzenheiser and Barks (1991) advocated that the design of insurance products should provide features to cover life-cycle benefits and incentives to encourage policyholders to keep their policies to ensure that such benefits would be there when needed. Howard (1997) suggested that persistency could be improved by reinforcement visits, regular service and information, and termination could be minimized by product flexibility to meet changing policyholder insurance needs.

Purushotham (2001) presented the results of a LIMRA survey of company practices regarding the measurement of individual life persistency experience. Most of the responding companies have more than one tool in place for monitoring product persistency. Further, more than half of these companies conduct traditional actuarial studies on a regular basis. These traditional actuarial persistency studies involve the determination of lapse rates by policy count, face amount, and annualized premium level. Often lapse results are examined by product, policy year, issue age, premium payment frequency, and billing method.

In recent years, insurance companies have begun to recognize the impact of product persistency experience on company profitability. For example, Streiff (2001) expressed concern over shrinking annuity persistency and its adverse impact on profitability. He concluded that insurers need to recognize the problem and commit resources to solving it. Just as there is a team dedicated to new business (sales), there should also be a team dedicated to conservation. In a later paper, Streiff (2002) suggested that to increase persistency, insurers should carefully analyze past persistency patterns of their distributors in order to stop persistency problems before the product is sold. Further, in addition to meaningful contact and good service, it is important to make sure the product is performing as the policy owner expects.

While the papers cited above were not technical, there are a few papers that have used statistical techniques to analyze empirical data pertaining to the termination of policies. Renshaw and Haberman (1986) used lapse (or withdrawal) data with various policy characteristics to determine the reasons policyholders gave up their policies in the U.K. They applied a general linear model (GLM) to identify factors (such as type of policy, location of the office, and mode of payment) that could affect termination rates. They found that the main interaction involved policy duration and the type of policy.

Lian et al. (1993) applied multiple regression analysis and concluded that the variables age, gender, mode of payment, type of policy, and method of payment had a significant bearing on the duration of life insurance policies. In a later study, Lian, Wu, and Loi (1998) employed survival analysis to study the factors affecting the termination of policies and to evaluate the impact of each significant factor. The study provided important information such as the average life expectancy of terminated policies, high risk-of-termination time periods, and incidence of survival rates by gender, types of policy, modes of payment, method of payment, and service status.

To summarize, few academics have analyzed policy terminations, while several trade and professional publications have discussed persistency and the minimization of terminations. Many of the findings and suggestions contained in these papers are based on anecdotal evidence instead of empirical analysis. This paper attempts to fill this gap in the literature by analyzing a large database of terminated policies, deriving objective (statistical) evidence of the factors associated with lapses, and generating useful rules and visual results to help insurance companies and their agents to better understand and minimize policy terminations.

3 Data Mining

3.1 The Literature

More companies are now using data mining as the foundation for strategies that help them outsmart competitors, identify new customers, and lower costs (Davis, 1999). Data mining is now widely used in marketing, risk management, and fraud control (Kuykendall, 1999). For example, an insurer may data-mine customer information to develop competitive rates, a retailer may analyze data mined from transactional systems to refine direct mailings and advertising campaigns to improve sales, while an auditing system may use data mining to help identify what could be fraudulent insurance claims. In a typical lapse-related application, data mining can identify customers who are likely to be profitable and who are likely to leave or churn, thus helping the insurer to target the valuable customers for extra value-added customer services, special offers, and loyalty incentives (Peacock, 1998).

To date, data mining has been applied to a wide range of business applications including direct mailings and advertising (Storey, 2001), customer relationship management (Lesser, Mundel, and Wiecha, 2000), fraud detection and fraud control (Kuykendall, 1999), surveillance (Holliday, 2001), risk management, (Berger, 1999), customer acquisition (Peacock, 1998), lifetime valuation (Drew et al., 2001), and others. Although there is widespread application of data mining in the commercial world, this interest is not matched in the academic world.

3.2 Data Mining Techniques and Decision Trees

Data mining can be defined as the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown patterns of data (SAS Institute, 1998). The SAS® data mining methodology comprises the following five stages: Sample, Explore, Modify, Model and Assess (also known as SEMMA):[1]

- Sampling is desirable if the data for analysis are too voluminous for reasonable processing time or if it is desirable to avoid problems of generalization by dividing the data into different sets for model construction and model validation.
- Exploration and modification refer, respectively, to the review of data to enhance understanding (e.g., by examining the summary measures) and the transformation of data (e.g., to induce a linear relationship or a normal distribution). It is noted that not every data-mining project needs sampling or modification of the data. Exploration is usually useful, however, and done as a form of preliminary analysis.
- The modeling stage is the actual data analysis. Most data mining software include traditional statistical methods (e.g., regression analysis and discriminant analysis) as well as non-traditional statistical analyses such as neural networks and decision trees.
- Finally, the assessment stage allows the comparison of models and results from any data mining model by using a common yardstick (e.g., lift charts or profit charts).

This study uses SAS® Enterprise Miner software for data analysis.[2]

Decision trees, one of the most commonly used data mining techniques, are often used to discover classification rules for a chosen attribute (i.e., target or dependent variable) by systematically subdividing

[1] Another data mining methodology is provided by CRISP-DM (Cross-Industry Standard Process for Data Mining—see <http://www.crisp-dm.org>). This framework proposes the following stages: (1) business understanding, (2) data understanding and data preparation, (3) modeling, (4) evaluation, and (5) deployment. Business understanding is critical, as it identifies the business objectives and hence the success criteria of data mining projects. Deployment refers to the operationalization and implementation of a data mining model in the organization. The data, modeling, and evaluation stages in CRISP-DM are similar to SEMMA.

[2] Other popular data mining software include Clementine (from SPSS, available at <http://www.spss.com>) and Affinium Model (from Unica, available at <http://www.unicacorp.com>).

information contained in the data set. Consequently, a decision tree-based approach gives clear decision rules for target variables such as who is likely to purchase a certain product, how much profit or loss one can expect by targeting a particular subgroup, which customer is likely to churn/turnover ... etc. The greatest benefit of decision trees is their ability to generate clear and understandable "if... then...' rules that can be readily applied in a business process. Related to this is the ability to visualize decision tree results. As with all decision-making methods, however, decision trees should be used in conjunction with common sense—decision trees are just one important part of a decision-making tool kit.

4 Model Development

4.1 The Singapore Life Insurance Industry

Singapore is a premier insurance center in Asia. There are currently 13 insurance companies competing in the life insurance sector (as listed in the ASEAN Insurance Industry Directory[3]). These include local insurance companies (e.g., NTUC Income and Asia Life), insurance companies associated with banks (e.g., UOB Life and Great Eastern Life) and foreign insurance companies (e.g., AIA and China Life).

The insurance industry in Singapore is regulated and supervised by the Monetary Authority of Singapore (MAS) (the central bank of Singapore) via a risk-based framework. In addition, parliamentary acts and MAS notices impact the life insurance industry. For example, the Financial Advisers Act implemented in October 2002 sets the minimum entry and examination requirements for financial advisers, while MAS Notice 318 sets the principles of appropriate market conduct for direct life insurers. Overall, MAS ensures that insurance companies in Singapore maintain a high standard of professionalism, financial management, and prudence.

Another unique feature in Singapore is the Central Provident Fund, which is a compulsory social security savings scheme. Employees can use their CPF funds to buy insurance policies such as term policies and health policies.

Based on data provided by MAS, the Singapore dollar amounts for new annual premiums for life insurance, new single premiums for life insurance, and new single premiums for annuity in 2002 are S$686.7

[3]Available at: <http://www.insurance.com.my/zone_industry/directory>.

million, S$5.9 billion and S$602.6 million, respectively. Investment-linked policies accounted for 37% of total new single premiums. A total of 1.42 million policies were sold in 2002 with a life insurance coverage of S$27.9 billion. Of these policies, approximately 14% were whole life policies, 22% were endowment policies, 62% were health policies, and 2.0% were term policies.

Total individual annual premiums in force amounted to S$5.2 billion in 2002. There were 5.89 million individual polices in force with a life insurance coverage of S$240 billion. About 90% of the Singapore population is estimated to be covered by life insurance. The average annual growth in life insurance in the years 1992 to 2002 is 12.4%.

Life insurance in Singapore is sold primarily through the traditional agency distribution channel. The importance of agents is evidenced by the formation of the Committee on Efficient Distribution of Life Insurance in March 2000, which has a mandate to make recommendations to enhance the sales advisory process, product disclosure, and professional training requirements. In the past several years, insurance companies have also moved into direct marketing through strategic alliance with banks (bancassurance). This channel is now becoming increasingly important in view of the emergence of financial advisers in banks. Some insurance companies also sell their policies through credit card companies and through the Internet. The latter is likely to become an effective and economical means of distribution given the high information technology penetration in Singapore.

In the event of surrender of life policies, the surrender value is calculated using the U.K. model, which that looks at the market conditions at the point of surrender. Also, orphan policies (i.e., policies without agents) are quickly assigned to other agents by group managers. In recent years, the declining mortgage rates and re-financing rates appear to have affected the persistency of endowment policies.

Growth in the life insurance industry has been sluggish in 2003 because of economic uncertainties, the war in Iraqi, and the effects of the SARS outbreak. These have contributed to the reluctance of the Singapore population to make long-term financial commitments. The outlook for 2004 is more positive, driven mainly by a more optimistic economic outlook.

4.2 The Data

The data for the study were supplied by the Life Insurance Association of Singapore, based on all the individual policies that are terminated because of non-payment of premiums. A total of 48,243 such policies is included in the study, with an average duration of 2.58 years (standard deviation = 3.33 years). These include policies that have been terminated because of lapses (32,360 or 67.08%), surrender (11,705 or 24.26%), or conversion to reduced paid-up or extended term insurance (4,178 or 8.66%), but exclude policies terminated because of death, maturity, expiry, or conversion to permanent plans. For each policy, the following variables are provided in the LIAS data set: duration of policy, age, gender, type of policy, mode of payment, method of payment, status of service, and size of policy. The details are shown in Table 1.

Table 1
Variables Used to Study Duration of Terminated Policies

Variable	Measurement	Frequency (%)
Gender	Male	32,902 (68.2%)
	Female	15,341 (31.8%)
Type of Policy (TP)	Whole Life Par. (WLP)	16,211 (33.6%)
	Whole Life Non-Par. (WLN)	7,386 (15.3%)
	Endowment Par. (EP)	19,916 (41.3%)
	Endowment Non-Par. (EN)	183 (0.4%)
	Term (T)	3,916 (8.1%)
	Others (TO)	631 (1.3%)
Mode of Payment (MP)*	Annually	11,237 (23.3%)
	Semi-annually	3,187 (6.6%)
	Quarterly	3,150 (6.5%)
	Monthly	30,619 (63.6%)
Method of Payment (MPT)**	Cash/Check (CC)	25,370 (53.3%)
	GIRO	15,628 (32.9%)
	Salary Deduction (SD)	5,799 (12.2%)
	Banker's Order (BO)	593 (1.2%)
	Others (MO)	189 (0.4%)
Status of Service	Serviced Policy (agent)	44,387 (92.0%)
	Orphan Policy (no agent)	3,856 (8.0%)
Age at Purchase	In Years (Mean = 26.48, Std. Dev. = 9.95)	
Size of Policy (SP)	In S$1000s (Mean = 36.05, Std. Dev. = 52.58)	

Notes: *50 missing values; **664 missing values; Par. = Participating.

To ensure validity of the results, the data set is randomly partitioned into three parts, one part each for model construction, model validation, and model testing. Because the data set of 48,243 policies is large, a 70%, 20%, and 10% partitioning is used for model construction, validation, and testing, respectively. A significance level of 0.05 is used to guide the construction of the decision tree. For this study, the construction sample is used to construct the decision tree and the results are validated against the validation sample to avoid over-fitting the data. Finally, the reported p-values are computed with the test sample.

4.3 Decision Trees

The objective of decision trees is prediction and/or classification by dividing observations into mutually exclusive and exhaustive subgroups. The division is based on the levels of particular independent variables that have the strongest association with the dependent variable. In its basic form, the decision tree approach begins by searching for the independent variable that divides the sample in such a way that the difference with respect to the dependent variable is greatest among the divided subgroups. At the next stage, each subgroup is further divided into sub-subgroups by searching for the independent variable that divides each first-stage subgroup in such a way that the difference with respect to the dependent variable is greatest among the divided sub-subgroups. The independent variable selected need not be the same for each subgroup.

This process of division (or splitting in decision trees terminology) usually continues until either no further splitting can produce statistically significant differences in the dependent variable in the new subgroups or the subgroups are too small for any further meaningful division. The subgroups and sub-subgroups are usually referred to as nodes. Decision tree results can be converted into "if...then..." rules or graphically represented by a tree-like structure. These indicate the association of the independent variables with the dependent variable.

In the context of this paper, the decision tree shows the association between age, gender, type of policy, mode of payment, method of payment, status of service, and size of policy on one hand, and the average duration of terminated policies on the other. While univariate analysis can also be used to examine the relationship between duration and each independent variable to generate profiles that are more likely to experience policy termination at an aggregated level, decision trees lead to more powerful and richer results by:

1. Examining all independent variables at each split and selecting the particular variable that gives the best split;

2. Determining the threshold level to split interval variables such as age or size of policy;
3. Ordering the importance of the independent variables by their level or depth in the tree;
4. Segmenting the data into nodes that maximize the association between termination and the independent variables; and
5. Increasing usefulness of the results by generating rules and tree-like structures.

5 The Main Results

5.1 Decision Tree Structure

The structure of the decision tree, as summarized in Figure 1, consists of four levels. As mentioned earlier, a branch will not be split further if the difference in the mean duration of policy for all possible splits (based on all possible independent variables and all possible split thresholds) is not statistically significant at a 0.05 significance level. Two values are reported in the leading box of Figures 2 to 14. The first value refers to the mean duration of the particular branch and the second to the p-value for ANOVA. The abbreviations in Table 1 are used to describe the various branches.

At level 1 (see Figure 1), the overall sample is split by mode of payment into two branches: non-monthly and monthly. At level 2, the non-monthly branch is split by type of policy into four branches: **EP**, **EN or WLN**, **WLP**, and **T or TO**. The monthly branch is split by method of payment into three branches: **CC**, **SD or MO**, and **GIRO or BO**.

At level 3, for the non-monthly branch, three of the four branches at level 2 (i.e., **EP**, **EN or WLN**, and **T or TO**) are further split by size of policy whereas for the monthly branch, the three branches at level 2 are further split either by size of policy (**CC** and **SD or MO** branches) or type of policy (**GIRO or BO** branch). At level 4, only the monthly branch is further split either by type of policy (**CC → SP** branch at level 3) or age (**SD or MO → SP**, **GIRO or BO → WLN or TO**, and **GIRO or BO → EN or EP** branches at level 3).

5.2 Decision Tree Results

The results are presented in a top-down order to give due consideration to the relative importance of the variables. Generally, variables

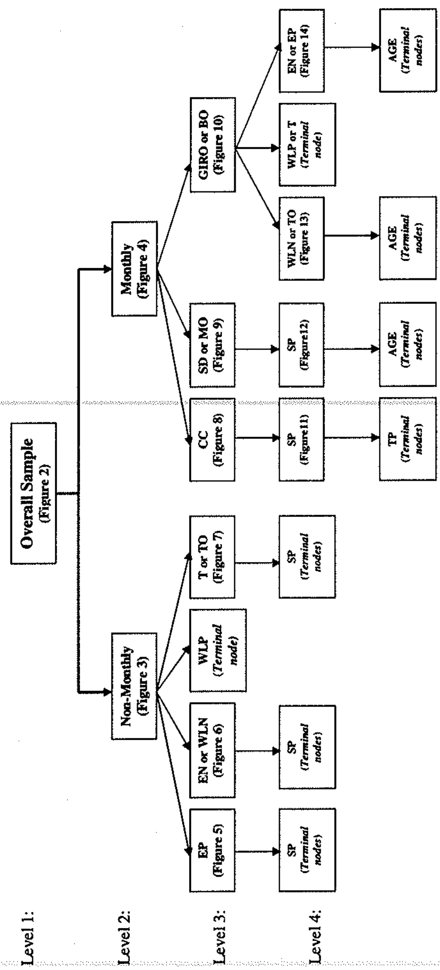

Figure 1: Structure of the Decision Tree

appearing at a higher level in a decision tree are more important in determining the average duration of terminated policies than variables appearing at lower levels. Also, variables not included in the decision tree are not statistically significant.

5.2.1 Levels 1 and 2

As shown in Figure 1, the mode of payment is the first variable in the decision tree. The overall sample is split into two branches, that is, non-monthly mode policies and monthly mode policies, as shown in Figure 2. The ANOVA p-value is 0.0001.

The results indicate that mode of payment is the most important factor in determining the average duration of the terminated policies. This finding is reasonable, as non-payment of premium would directly lead to termination. Intuitively, policyholders paying premiums by the monthly mode have to consider continuation of payment more frequently than those who pay premiums on a non-monthly mode. Therefore, it is not surprising that mode of payment is selected by the decision tree algorithm as the most important factor in determining the duration of terminated policies.

5.2.2 Levels 2 and 3

As the decision tree branches down from level 2 to level 3, type of policy is selected as the next most important variable to split non-monthly mode policies (Figure 3) and method of payment as the next most important variable to split monthly mode policies (Figure 4). The splits in Figures 3 and 4 are statistically significant since the p-values are both equal to 0.0001.

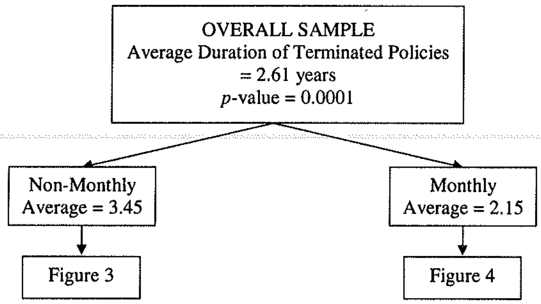

Figure 2: The Decision Tree: Levels 1 and 2

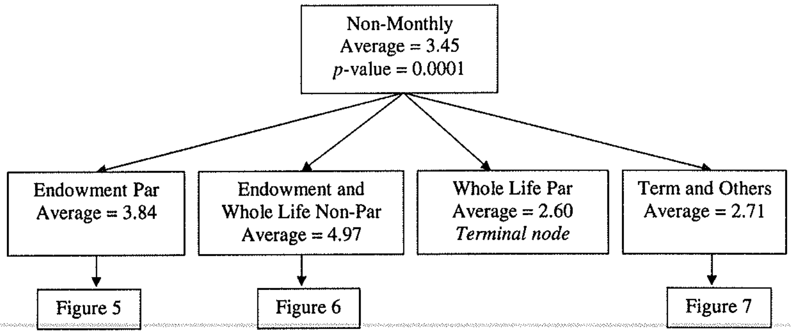

Figure 3: The Decision Tree: Levels 2 and 3, Non-Monthly Branch

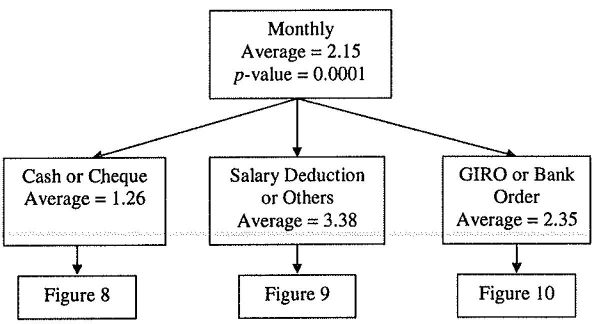

Figure 4: The Decision Tree: Levels 2 and 3, Monthly Branch

The results indicate that for non-monthly mode policies, different types of policies are associated with different average durations. One explanation for term and other policies (such as health insurance) being terminated earlier than all other types of policies (except whole life par) could be that they have no cash value to lose and are therefore more susceptible to competition and hence have a better chance of being replaced/terminated.

As for monthly mode policies, method of payment affects average duration significantly. The findings shown in Figure 4 are not surprising because salary deduction and GIRO (a popular type of electronic fund transfer service used in Singapore) can enhance the probability of premium payments, thus increasing the expected persistency.

The results presented above show that decision trees work on the strength of the effect that a current factor has on the preceding factor. Therefore, it is not surprising to see that different variables emerge for policies with different modes of payments, i.e., method of payment for monthly mode policies and type of policies for non-monthly monthly mode policies.

5.2.3 Levels 3 and 4

As the decision tree branches down from level 3 to level 4, for the non-monthly payment-mode branch, three of the four branches at level 3 are further split by size of policy (Figures 5 to 7); whereas for the monthly payment-mode branch, the three branches at level 3 are further split either by size of policy or type of policy (Figures 8 to 10). Again, the splits in Figures 5 to 10 are statistically significant, as all the p-values are 0.0001.

As shown in Figures 5 to 7 for non-monthly mode policies, a positive relationship exists between policy size and average duration for all types of policies except endowment par. Generally, policyholders with larger policies enjoy greater protection and have more reasons to hold on to them (see Figures 6 and 7). For endowment par, however, there is a negative relationship between the duration and size of policies (Figure 5). It is probable that because endowment par is the most expensive plan, it would be more difficult to keep a bigger policy in force.

As for monthly mode policies, size of policy is the next most important variable for cash/check and salary deduction and other methods of payment as shown in Figures 8 and 9. For GIRO and banker's order, type of policy is the next most important variable as shown in Figure 10.

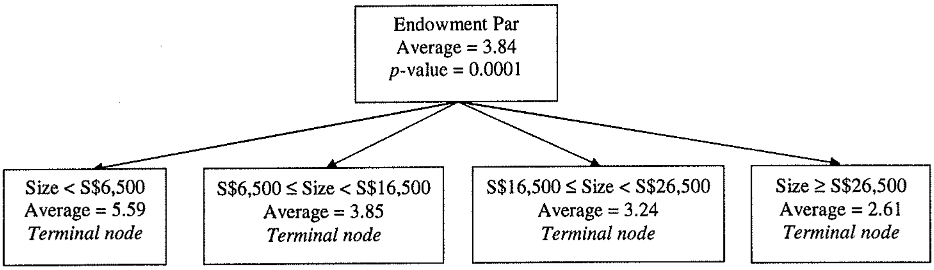

Figure 5: The Decision Tree: Levels 3 and 4, Endowment Par (EP) Branch

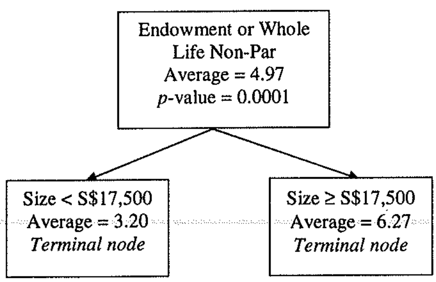

Figure 6: The Decision Tree: Levels 3 and 4, Endowment or Whole Life Non-Par (EN or WLN) Branch

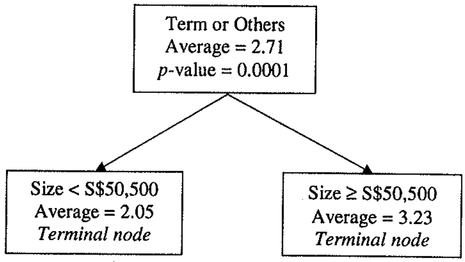

Figure 7: The Decision Tree: Levels 3 and 4, Term or Others (T or TO) Branch

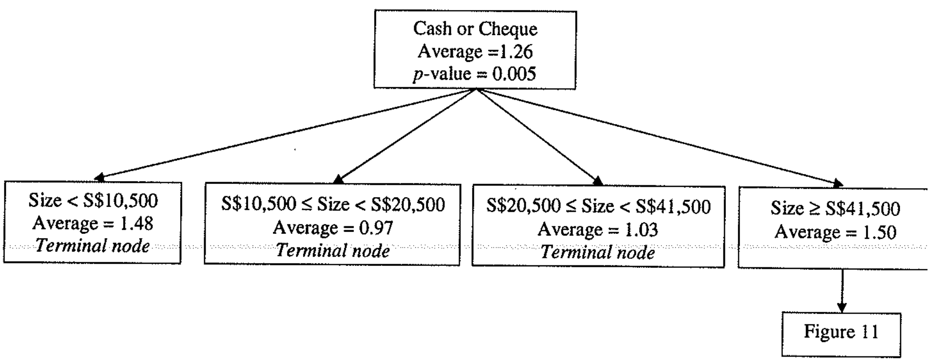

Figure 8: The Decision Tree: Levels 3 and 4, Cash or Check (CC) Branch

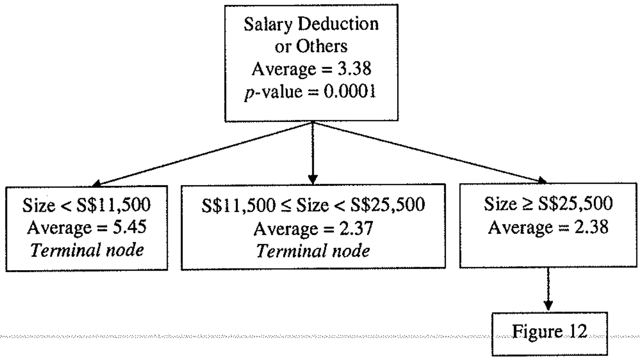

Figure 9: The Decision Tree: Levels 3 and 4, Salary Deduction or Others (SD or MO) Branch

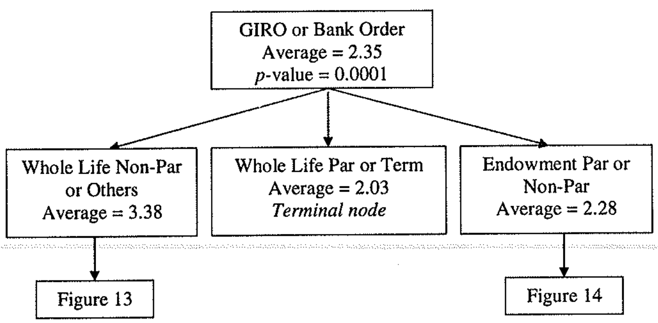

Figure 10: The Decision Tree: Levels 3 and 4, GIRO or Bank Order (GIRO or BO) Branch

Figures 8 and 9 indicate that the relationship between size of policy and average duration is non-monotonic (U-shaped) for cash/check method of payment and for salary deduction and others. For GIRO and banker's order payment, average duration varies by type of policy. These findings differ markedly from those for non-monthly mode policies. This suggests that simple rules of thumb cannot be applied to all situations.

5.2.4 Levels 4 and 5

At level 4 of the decision tree, only the monthly branch is further split either by type of policy (Figure 11) or age (Figures 12 to 14). It is noted that the p-values in Figures 11 to 14 are 0.0020, 0.0001, 0.0020, and 0.0001, respectively; therefore the differences of the average durations among various branches are statistically significant.

Figure 11 shows that whole life non-par or other policies tend to last longer than whole life par or endowment par policies. This is probably because whole life non-par or other policies are cheaper and hence have a greater chance of keeping them in force.

Figures 12 to 14 indicate that generally policies payable on a monthly mode tend to have longer duration as the age of purchase of the policyholders increases. Apparently, older policyholders (at time of purchase) tend to value insurance policies more and keep them longer as they are more security conscious. (This finding does not apply to non-monthly mode policies and those of certain sizes in any significant manner.)

6 Closing Comments

The decision tree results can help insurance companies and their agents isolate and focus on the key variables and combinations of variables that increase persistency. On the flip side, they can avoid those undesirable combinations that decrease persistency. For example, the decision tree in Figure 1 shows that the most persistent policies (those with longest average duration) have an average of 6.27 years for policies with the following combination (interaction) of variables: non-monthly payment-mode policies (Figure 3), endowment non-par or whole life non-par (Figure 6), and size of policy exceeding S$17,500. Such policies are expected to be more profitable. On the other hand, the shortest average duration is 0.97 years for policies with the following attributes: monthly payment-mode (Figure 4), cash/check payment (Figure 8), and size of policy between S$10,500 and S$20,500.

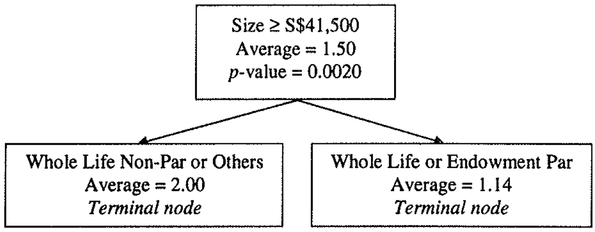

Figure 11: The Decision Tree: Levels 4 and 5, Size > S$41,500 Branch

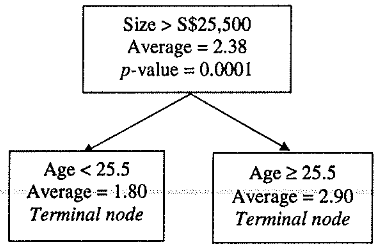

Figure 12: The Decision Tree: Levels 4 and 5, Size > S$25,500 Branch

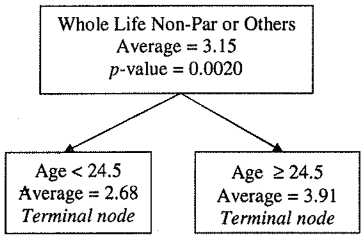

Figure 13: The Decision Tree: Levels 4 and 5, Whole Life Non-Par or Others (WLN or TO) Branch

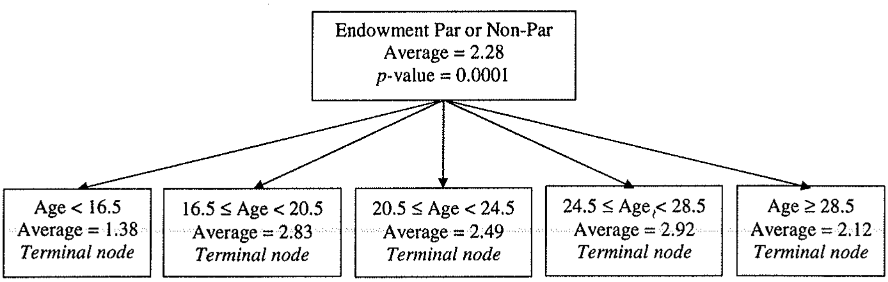

Figure 14: The Decision Tree: Levels 4 and 5, Endowment Par or Non-Par (EP or EN) Branch

It is useful to know how age, gender, type of policy, mode of payment, method of payment, status of service, and size of policy are associated with the termination of policies—these associations were described earlier and illustrated in Figures 1 to 14. To gain full value from the decision tree results, however, the practical application of the results to an analytical model should be considered. In particular, life insurance companies and agents can focus on factors that are associated with the most significant improvement on the average duration of policies. Factors that are applicable to all policies such as mode of payment (i.e., non-monthly) and method of payment (i.e., non-cash or non-check) should be given greater emphasis. Measures such as premium discount, bonuses, and other incentives could be offered to induce policyholders to adopt a non-monthly mode and a non-cash (or non-check) payment method. These can have a positive effect on persistency. As for larger policies, which are already receiving size discounts (due to their lower unit cost of administration), there is an additional justification to do so: their better persistency.

While persistency of policies is critical to the profitability of life insurance companies, it is also the social responsibility of the companies to ensure that the policyholders are receiving value for their money. It is therefore important that insurance companies always keep track of the persistency of their policies and be guided by findings to take measures in ensuring that policyholders keep their policies in force for as long as possible. Longer average duration will not only yield greater profitability to the companies but also provide greater protection for the policyholders.

It should be noted that the data set used in this study is confined to terminated policies. The analysis is therefore restricted to the duration of terminated policies. If in-force policies are also included in the data set, a more comprehensive set of conclusions could be made.

Finally, it is hoped that this paper can make a contribution to the literature as well as the practice of life insurance companies and their agents.

References

Berger, C. "Data Mining to Reduce Churn." *Target Marketing* 22, no. 8 (1999): 26–28.

Berlin, J.R. and Powell, J.M. "Persistency—One Indicator of Customer Satisfaction." *Limra's Marketfacts* 9, no. 1 (1990): 15–18.

Chung, H.M. and Gray, P. "Data Mining." *Journal of Management Information Systems* 16, no. 1 (1999): 11-13.

Davis, B. "Data Mining Transformed." *Informationweek* 751 (1999): 86-88.

Drew, J.H., Mani, D.R., Betz, A.I., and Datta, P. "Targeting Customers with Statistical and Data-Mining Techniques." *Journal of Service Research* 3, no. 3 (2001): 205-219.

Goodwin, D. "The Replacement Dilemma." *Best's Review* 93, no. 12 (1993): 31-14.

Hall, E.R. "Incomes Are Central to Life Insurance Sales." *Limra's Marketfacts* 9, no. 4 (1990): 30-32, 48-49.

Holliday, A. "When Crimes Pay." *Director* 54, no. 8 (2001): 50-55.

Howard, L.S. "U.K. Regulators Watch Persistency." *National Underwriter* 101, no. 26 (1997): 9-12.

Ingram, D.N. "Compensation for Value." *Best's Review* 88, no. 3 (1987): 14-20.

Kuykendall, L. "The Data Mining Toolbox." *Credit Card Management* 12, no. 6 (1999): 30-40.

Lautzenheiser, B. and Barks, B. "Why Punish Persistency?" *Best's Review* 95, no. 2 (1991): 18-21, 116.

Lesser, E., Mundel, D. and Wiecha, C. "Managing Customer Knowledge." *The Journal of Business Strategy* 21, no. 6, (2000): 35-37.

Lian, K.H., Loi, S.L., Wong, Y.W.A. and Hsu, W.C. "Lapsation and Replacement—What Are the Causes? The Singapore Experience." *The Institute of Actuaries of Australia, Quarterly Journal* (September, 1993): 22-48.

Lian, K.H., Wu, Y. and Loi, S.L. "Survival Analysis of Terminated Life Insurance Policies." *Singapore International Insurance and Actuarial Journal* 2, no. 1, (1998): 101-119.

Peacock, P.R. "Data Mining in Marketing: Part I." *Marketing Management* 6, no. 4 (1998): 8-18.

Purushotham, M.C. "Individual Life Persistency Measures." *Limra's Marketfacts* 20, no. 4 (2001): 70-73

Renshaw, A.E. and Haberman, S. "Statistical Analysis of Life Assurance Lapses." *Journal of the Institute of Actuaries* 11 (1986): 451-469.

Santos, M.A. "Achieve High Persistency with a Service Attitude." *Life Association News* 87 no. 11 (1992): 152, 154.

SAS Institute. *From Data to Business Advantages: Data Mining. The SEMMA Methodology and SAS Software.* Cary, NC: SAS Institute, 1998.

Storey, C.K. "Don't Get Stuck." *Credit Union Management* 24, no. 7 (2001): 38–41.

Streiff, T.F. "Tackling the Problem of Annuity Persistency." *National Underwriter* 105, no. 50 (2001): 42–43.

Streiff, T.F. "How to Maximize Annuity Persistency." *National Underwriter* 106, no. 5 (2002): 50

Sweeney, P.M. "Lapse Ratio Turns after 20-Year Low." *Best's Review* 97, no. 8 (1996): 36–37.

Willett, L.F. "Let's Get Personal ...." *Insurance Sales* 129, no. 10 (1986): 36–37.

Zinsser, A. "Present Policyholders Are the Best Resource." *Best's Review* 97, no. 8 (1996): 74–77.