ADJUSTING LIFE TABLES TO INCORPORATE PERTINENT PERSONAL PROFILE INFORMATION

Patrick L. Brockett and Samuel Cox*

Section 1

Introduction and Summary

The purpose of this paper is to show how to use personal information such as medical history, family characteristics, insurance characteristics, etc. in a logical statistical fashion for the purpose of modifying an existing mortality table to reflect the collected underwriting information. The technique presented is applicable even when the relevant information has only been gathered on a relatively small group of individuals, as well as cases in which more extensive experience has been gathered. Consequently, some smaller scale medical studies can be made useful for actuarial calculations.

Additional benefits of the proposed technique are:

i) It uses information in all ages to gather information about the effect of the measured underwriting characteristics. In a sense it smooths the data over age as well as covariate experience groups.

ii) It modifies the entire standard mortality table in light of the pertinent information, and does not smooth the table by ad hoc ratio smoothing at specific ages.

iii) There is a firm statistical foundation for the technique. Accordingly, inference can be made about covariate experience effect upon mortality. This is impossible using the traditional actuarial methods which ignore the random aspects of the mortality curve adjustments, i.e., that the information gathered is sample information.

iv) Confidence intervals for the mortality table adjustments may be made, allowing risk protection in calculation of annuities and life

*Associate Professor of Finance and Actuarial Science at The University of Texas at Austin. Professor Brockett is also with Applied Research Labs, University of Texas. This work supported in part by the office of Naval Research N0014-81-K-0145.

insurance functions involving the experienceadjusted table.

v) The technique is easily programmed on a computer. This means that policy premiums can be calculated, and underwriting done on an individual by individual basis. Thus a policy can be constructed and priced tailored to each individual's personal history and expected mortality. Mortality tables are obsolete in this era of modern computers; however if an actuary desires, such a table complete with associated commutation functions may be printed out for each individual policy holder, or group of policy holders. The aggregation of such individual personal mortality characteristics for the purpose of reserve calculation, etc. poses no significant mathematical or statistical problems and shall be dealt with in a subsequent article.

Section 2

A method for lifetable adjustment in light of covariate information

Suppose we know certain information about an individual such as medical history, smoking or non-smoking, diabetes, mental illness, etc. How should the actuary take this information into account when underwriting? Actuaries traditionally use such covariate information as sex, age, years since purchase of insurance, etc. to obtain a lifetable, and then using this specific lifetable, they calculate the desired quantaties. For many covariates of interest for example in disability insurance or higher risk underwriting there is insufficient mortality data gathered within company on the specific covariates to directly obtain a mortality table. Often only small scale medical studies, and/or limited company experience can be brought to bare on the calculations. In this section we shall present a statistically based method for incorporating such information.

For a mathematical description of our procedure, let Z_1, Z_2, ..., Z_M denote the numerical values of $M \geq 1$ covariates for which we wish to adjust a standard mortality table. For example Z_1 might denote cholesterol level deviation from the normal, Z_2 might denote blood pressure deviation from normal, etc. Discrete variables may also be used, for example Z_3 might be 0 if no diabetes is present and one if diabetes is present, Z_4 might be one if a moderate smoker, zero otherwise, Z_5 might be one if a heavy smoker and zero otherwise.

Given the numerical measures Z_1, Z_2,..., Z_M , we assume that the effect of the covariate values is multiplicative upon the force of mortality. This is a quite natural assumption, since roughly it says that in a small interval of time (x, x+dx) the chance of death given the covariates Z_1, ..., Z_M is just a constant multiple of the same chance when $Z_1=0$, $Z_2=0$, ..., $Z_M=0$ [1]. Mathematically we write $\mu_t(Z) = \mu_t(0)\exp\{\beta_1 Z_1+..+\beta_M Z_M\}$ using the natural exponent to induce the multiplicative effect, and the symbol $\mu_t(Z)$ to denote the force of mortality at time t given the covariate vector $Z = (Z_1, Z_2.., Z_n)$.

In the equation

$$\mu_t(Z) = \mu_t(0) \exp\{\beta_1 Z_1 +...+\beta_M Z_M\} \tag{2.1}$$

the numbers $\beta_1,...,\beta_M$ represent unknown parameters expressing the actual effect of the covariates upon survival. When $\beta_j = 0$, the covariate j has no effect upon mortality, while if $\beta_j \neq 0$ there is an effect of

---

[1] The assumption that the effect of the covariates be multiplicative can easily be relaxed. If the covariates do not act multiplicitively, then one might stratify the population and assume $\mu_t(Z) = \mu_t^{(j)}(0) \exp\{\beta_1 Z_1+..+\beta_M Z_M.\}$ in the jth strata. See Kalbfleisch and Prentice (1980) for details.

covariate j upon survival. The model represented by equation (2.1) is called the proportional hazards model in the biostatistical literature. An excellent reference on the topic is the recent book by Kalbfleisch and Prentice (1980).

Our procedure for adjusting a standard mortality table to reflect the impact of the covariate vector $\underset{\sim}{z} = (z_1, z_2, \ldots, z_M)$ may now be briefly summarized as follows:

1) estimate the covariate effects $\underset{\sim}{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$ by $\hat{\underset{\sim}{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_M)$ using published medical data, or company data.

2) estimate the baseline average covariate value $\bar{\underset{\sim}{z}}^{(1)}$ corresponding to the sample study upon which the covariate effect assessment was made. This may not be the same average value as that for the population to which the adjusted table is to be applied. For example, if a company wishes to modify a standard mortality table of white male insured lives to obtain a table for white male insured lives with diabetes, they might refer to a study on mortality of diabetes patients published in a medical journal. The base line population for the medical journal study, is not the same as that in the insurance company, so this variation must be accounted for. In this step the medical study covariate average vector is represented by $\bar{\underset{\sim}{z}}^{(1)}$.

3) Using the results of 1) and 2) and the force of mortality function ν_t from the standard table we derive the <u>covariate adjusted standard</u> force of mortality via the equation

$$\nu_t(\underset{\sim}{z}) = \nu_t \exp\{\hat{\underset{\sim}{\beta}}'(\underset{\sim}{z} + \bar{\underset{\sim}{z}}^{(1)} - \bar{\underset{\sim}{z}}^{(2)})\} \tag{2.2}$$

where $\bar{\underset{\sim}{z}}^{(2)}$ is the average covariate value for the standard table upon which we are making the adjustment.

In this step the force of mortality has been adjusted to reflect the differences between the study group standard, and the company standard, as well as for the covariate effects.

4) The new adjusted mortality table $S_t(\underset{\sim}{z})$ reflecting the covariate information $\underset{\sim}{z}$ is now calculated in the usual manner from the force of mortality, namely the survival function is $S_t(\underset{\sim}{z})$ $\exp\{\int_0^t \nu_u(\underset{\sim}{z})du\}$

In the proportional hazards model this becomes

(2.3) $$S_t(\underset{\sim}{z}) = (S_t^{(std)})^{\exp\{\beta(\underset{\sim}{z} + \bar{\underset{\sim}{z}}^{(1)} - \bar{\underset{\sim}{z}}^{(2)}\}}$$

where $S_t^{(std)}$ is the standard survival curve upon which the adjustments are to be made. Thus the new table is obtained by raising the old table to a power, namely to the power $\beta_1(z_1 - \bar{z}_1^{(1)} + \bar{Z}^{(2)})$ $+ \ldots + \beta_M(\bar{z}_M^{(1)} - \bar{z}_M^{(2)} + z_M)$.

The only consideration still to be discussed is how to estimate the covariate effects $\underset{\sim}{\beta} = (\beta_1, \ldots, \beta_m)$ from the data in the medical study, or from the small experience company data. This will be provided in the next section where the innovative technique of partial likelihood due to D.R. Cox (1972) will be used.

Section 3

The estimation of parameters

To implement the procedure outlined in the previous section, one might proceed in any of several ways: 1) assume a specific parametric model (e.g., Makeham or Gompertz) for the baseline survival function and use the principal of maximal likelihood to estimate $\hat{\beta}$. Gross and Clark (1975) is a good reference concerning this method and the inferential procedures associated with it. 2) Assuming the baseline survival function cannot be specified in advance, one may proceed using the partial likelihood method developed by D.R. Cox (1975). This is the method we shall pursue here. A very good account of partial likelihood and its inferential implication is to be found in Kalbfleisch and Prentice (1980).

Consider a study group or company sample containing n individuals. For individual i we record the data set (t_i, δ_i, z_i) where t_i is the time of death or event, or possibly censoring, δ_i is the indicator of failure with $\delta_i = 1$ if failure occurs, and $\delta_i = 0$ otherwise, and $\underset{\sim}{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iM})$ is the covariate vector for the ith individual. By censoring here we mean that the individual withdrew or was lost from the study before failure occured.

For convenience, let us order the failure times (i.e., those individuals with $\delta = 1$), so that $t_{(1)}, t_{(2)}, \ldots, t_{(k)}$ are the k reported failure times, and $\underset{\sim}{z}_{(1)}, \underset{\sim}{z}_{(2)}, \ldots, \underset{\sim}{z}_{(k)}$ are these people's covariate vectors. Additionally, let R(t) denote the set of individuals at risk of failure at time t, i.e., those people alive and uncensored at time t−0. Partial likelihood compares the covariate vector of the individual who failed at time $t_{(j)}$ with the covariate vector of those persons at risk at time $t_{(j)}$.

Thus, using the proportional force of mortality assumption (2.1) we have in essence

$$\Pr\{\text{individual(i) fails at } t_{(i)} \mid R(t_{(i)}), \text{ and one item fails at } t_{(i)}\}$$

$$= \frac{\mu_{t_{(i)}}(\underset{\sim}{z}_{(i)})}{\sum_{j \in R(t_{(i)})} \mu_{t_{(i)}}(\underset{\sim}{z}_j)}$$

$$= \frac{\exp\{\beta' \underset{\sim}{z}_{(i)}\}}{\sum_{j \in R(t_{(i)})} \exp\{\underset{\sim}{\beta}' \underset{\sim}{z}_j\}} .$$

Multiplying over all the failure times gives the partial likelihood.

$$(3.1) \qquad L(\underset{\sim}{\beta}) = \prod_{i=1}^{k} \frac{\exp\{\underset{\sim}{\beta}' \underset{\sim}{z}_{(i)}\}}{\sum_{j \in R(t_{(i)})} \exp\{\underset{\sim}{\beta}' \underset{\sim}{z}_j\}}$$

in the case of no tied failure times.

If $\hat{\beta}$ is chosen to maximize $L(\beta)$, then under suitable regularity conditions $\hat{\beta}$ behaves just like a maximum likelihood estimator, e.g., $\hat{\beta}$ is the unique solution of $\frac{\partial}{\partial \beta}\log L(\beta) = 0$, $\hat{\beta}$ is asymptotically normally distributed with mean β and approximate covariance matrix $I^{-1}(\hat{\beta})$ where

$$I(\beta) = \left(\frac{-\partial^2 \ln L(\beta)}{\partial \beta_i \partial \beta_j}\right).$$ For a proof of these results, see Tsiatis (1981). The beauty of this model is that no assumptions whatsoever need be made about the underlying study baseline force of mortality function, and yet still inference concerning the covariate effects may be using (2.1) and partial likelihood.

A fortran computer code to implement partial likelihood estimation may be found at the end of the book by Kalbfleisch and Prentice (1980). They additionally provide a program to implement the estimation when the covariates are time dependent (such as history of attacks of some ailments or disability).

Section 4

Other models and further extensions

In this paper we have shown how to incorporate the personal profile information in the simplist possible situation. This was done for ease of presentation, and to clarify the basic technique. In particular applications however there may be complications such as time varying covariates, non-proportional forces of mortality, tied failure times for the study group, or grouped data problems. Such complications can be handled by suitable extensions of the model. Kalbfleisch and Prentice discuss these problems and extensions in some detail, and the interested reader is referred to this book.

Concerning alternative models for incorporating covariate information in adjusting standard tables, the first method which leaps to mind is the accelerated failure time model from engineering and biostatistics. In this model it is assumed that the effect of the covariates is to act multiplicative on time itself rather than upon the force of mortality. Thus the effect of the covariates is to speed up or slow down an individual's progression along the time axis, essentially ageing him at a rate possibly different than calendar time. See Kalbfleisch and Prentice (1980) for more details. Mathematically, in terms of the survival function we have

$$S(t|\underset{\sim}{z}) = S_o(te^{\beta' \underset{\sim}{z}})$$

where $S(t|\underset{\sim}{z})$ is the survival curve for someone with covariates and $\underset{\sim}{z}$ and $S_o(.)$ is the baseline survival function. The parameters $\underset{\sim}{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$ again represent covariate effect. If $S_o(\cdot)$ is known (e.g., Makeham or Gompertz) then parametric estimation procedures can be used to estimate $\underset{\sim}{\beta}$. Once $\hat{\underset{\sim}{\beta}}$ is found, the adjusted table is formed by taking

$$S(t|\underset{\sim}{z}) = S_o(te^{\underset{\sim}{\beta}' \underset{\sim}{z}}) \quad (4.1)$$

using S_o as the standard table.

As a practical matter, the Gompertz law fits mortality data quite well in most age ranges, and it is easily seen that the Gompertz curve simultaneously fits the proportional force of mortality model, and also accelerated failure time models. Hence the adjusted tables calculated using equation (4.1) and those calculated using equation (2.3) should be quite close. If the underlying survival function S_o is unknown, one can still use accelerated failure time modeling to estimate the $\tilde{\beta}$ parameters, either using the totally non-parametric method of Arnold and Brockett (1982), or of Buckley and James (1979), or else using the standard table to fill in expected survival times for the censored individuals and then using least squares on the log survival times. Of course S_o need not be known to use the proportional force of mortality technique with partial likelihood.

Concerning the computation of life insurance functions based upon the adjusted table we have for example the life insurance function

$$\bar{A}_x(\tilde{z}) = \int_0^\infty v^t \frac{S(t + x \mid \tilde{z})}{S(x \mid \tilde{z})} \mu_{x+t}(\tilde{z})dt$$

For the proportional force of mortality function assumption model this may be calculated on a computer. With the accelerated failure time model, we have the closed from solution

$\bar{A}_x(\tilde{z})$ at discount rate v, and covariates $\tilde{z}$
$= \bar{A}_y$ at discount $v^{\exp\{-\beta'\tilde{z}\}}$ using the standard table, where $y = xe^{\beta'\tilde{z}}$
is the "biologically equivalent age". Other life functions may similarly be related to the standard table.

Acknowledgements

The problem addressed in this paper was first brought to the author s' attention at the Actuarial Research Conference in Waterloo, Ontario Sept. 30-Oct. 2, 1982. Several speakers raised the issue in various guises. At this same conference, J. D. Kalbfleisch spoke on the proportional hazards model in biostatistics. Many of the ideas here can be traced to subsequent conversations with him. His talks also have impacted this paper.

References

Arnold, Barry C. and Patrick Brockett (1982). The Accelerated Failure Model with Random Right Censoring and Unknown Baseline Survival Function. Working Paper, Department of Finance, University of Texas at Austin.

Buckley, J., and I. James (1979. Linear Regression with Censored Data, Biometrika 66, 429-436.

Cox, D. R. (1972), "Regression Models and Life Tables (with discussion), Journal of the Royal Statistical Society, Series B, #34, 187-220.

______________ (1975), "Partial Likelihood," Biometrika 62, 269-276.

Gross, Alan J., and Virginia Clark (1975). Survival Distributions: Reliability. Application in the Biomedical Sciences, John Wiley, New York.

Kalbfleisch, J. D., and R. Prentice (1980) The Statistical Analysis of Failure Time Data, John Wiley, New York.