The Value of Logistic Regression in Mortality Table Construction

SCOR regularly reviews its mortality pricing tables to ensure that they are the most up-to-date, accurate and client-specific based on available proprietary data. Our pricing team, comprised of actuaries, product developers, statisticians and underwriters, reviewed SCOR’s tables last year and introduced several changes that reflect both emerging experience and our best estimates for future projection.

We found that logistic regression tools can be useful in creating or updating mortality experience tables. The models’ results have been instructive in helping us not only understand our inforce business but also project how business may perform under different scenarios. This allows us to price reinsurance rates more accurately on a client-by-client basis.

However, constructing a useful model is not a simple task. Perhaps most importantly, the modeling requires a large amount of data on existing business to produce credible results. I outline some of the features, benefits and challenges below.

Logistic Regression Model

Logistic regression can be used as a predictive model to estimate mortality for an insured population. This model has the general form of

\[ q_x = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots)}} \]

Where

- \( q_x \) is the predicted mortality rate (dependent variable)
- \( x_1, x_2, \ldots \) represent the risk drivers used in the study (independent variables)
- \( \alpha, \beta_1, \beta_2, \ldots \) represent model coefficients derived from the experience data

In other words, the linear regression common to most business school graduates only allows for a dependent variable value of 0 or 1 – in the case of mortality, for example, determining whether a subject is dead or alive. In contrast, logistic regression allows for non-binary values. In the case of mortality experience, we expect the value of the dependent variable \( q_x \) to range from zero to one, allowing an analyst to determine the probability of death given particular risk drivers.

Companies can determine their own set of risk drivers (x) based on the data they have available. The more credible and specific the data, the more credible results the model will produce. Figure 1 provides examples of such drivers, grouped into distinct categories.
Issue age, duration and study year may be treated as continuous for three reasons:

- To estimate smoothed relationships between q and these variables
- To allow the coefficients of these variables to be transformed as mortality slopes
- To enable model-based mortality projections for older ages and longer durations where data may be sparse or unavailable

Adjusting the Experience

After constructing base experience tables, the pricing team will typically need to make adjustments to create tables suitable for pricing new business. A couple of examples of adjustments that may be needed include negating the effect of selective lapsation and accounting for mortality improvement.

Negating the effect of selective lapsation

Much of the experience for later duration mortality comes from issue year eras that experienced very high lapsation. For example, during the 1980s average lapse rates for term insurance ranged from 15-20 percent in durations 1-10 to 9 percent for durations 11 and after. In comparison, today’s lapse rates range from around 6-7 percent for those earlier durations, grading down to 3 percent. If the cohorts issued insurance in the 1980s experienced some level of selective lapsation (i.e., better risks lapsed leaving poorer risks inforce), then the experience we measure today for those groups is higher than what we should expect from a newly issued cohort going forward. To negate some of this anti-selective mortality, a company can perform successive Dukes-MacDonald calculations to back out the effects. (For a description of the Dukes-MacDonald theory, see my November 2011 *Messenger* article “2001 VBT – Caution: Steep Hill Ahead.”) Removing the effects of selective lapsation will result in lower mortality in later durations.

Accounting for mortality improvement

A company may develop their own annual mortality improvement rates based on historical US population data (for example, from the Human Mortality Database, available at www.mortality.org). Annual improvement rates by gender and attained age can help in developing factors applied to both select and ultimate experience mortality to reflect the impact of secular trends. Experience tables should be generationally improved from the mid-point of the exposure period to the current pricing era (e.g., 2013). Then, separate durational factors can be added to reflect future improvement.

Although we have seen continual improvement in mortality throughout the 20th and early 21st centuries, this does not imply that it will continue forever into the future. A direct writer’s actuaries will need to determine the most appropriate length of time to incorporate durational improvement, depending on their view of future trends.

Logistic Regression’s Benefits, Challenges

Logistic regression can be a useful tool in reviewing and updating pricing mortality tables. Its particular strengths include:

- The ability to control and analyze multiple explanatory variables directly related to an insured population or block of business
- Insights into relative mortality relationships among the risk drivers
- Less stringent theoretical assumptions
- Relative ease in implementing with commonly available software systems

However, the benefits do not come without challenges. Critical to the model is the requirement for copious amounts of policy-level data on a large heterogeneous block of business, in a useable format. While industry data may suffice as a proxy (alone or in tandem with some degree of internal company data), a carrier’s pricing team must remain aware of the unique features of its own business. Heavy reliance on industry data may require at least some adjustment to credibility expectations.

Additionally, even with available data and systems, the modeling process is a laborious task that should be reviewed by multiple layers and departments within the insurer. SCOR’s pricing team includes pricing and marketing actuaries, statisticians, sales staff, underwriters and risk managers. Recruiting this team and gaining their commitment can be an overlooked but critical challenge.
Your Reinsurer Can Help

SCOR’s current pricing mortality tables, along with deal-specific adjustments based upon client experience, represent our best estimate view of future expected mortality for a block of fully underwritten life reinsurance issued today. Using a combination of our own and industry experience, our pricing team has developed robust statistical models to create best-fit mortality rates for male and female / smoker and nonsmoker. Using actuarial expertise, we have adjusted the experience to produce tables appropriate for pricing new business. These tables are the current expected basis in our experience study system with which we can monitor emerging results against this new set of mortality expectations.

As an aggregator of record-level data, SCOR has invested considerable resources in examining company data and the resulting “fit” to the client’s business strategies (see next article for a discussion on real-world benefits we have offered to clients). Our sales and pricing staff look forward to assisting our clients in developing their own effective reviews of existing and proposed pricing strategies. ∞