University of Wisconsin CAE Research Grant Application

Title: Insurance Analytics: From Data Analysis to Decision Making

Summary: Insurance is a data-driven industry and analytics is a key to deriving information from data. As quantitative analysts in the insurance industry, actuaries use analytical tools to help make data-driven business decisions. In the era of big data, actuaries are facing new challenges from data analysis to decision making. This proposal describes a plan to promote research in insurance analytics in the actuarial society. We propose three broad themes: the first explores novel application of dependence models in insurance operations; the second focuses on predictive modeling in health care; the last concerns analytics to support insurance and risk management theory. We will explore both traditional and non-traditional areas where analytics add value to decision-making. The products of this project will expand actuarial knowledge in insurance analytics and will make significant impact to the actuarial profession.

Section 1. Description of Proposed Project

1.1 Impact of Analytics on Industry

Analytics is a key to deriving information from data, especially in a data-driven industry such as insurance. But what is analytics? Making data-driven business decisions has been described as "business analytics," "business intelligence," and "data science." These terms, among others, are sometimes used interchangeably and sometimes used separately, referring to distinct domains of applications. As an example of such distinctions, the term "business intelligence" suggests a focus on processes of collecting data, often through databases and data warehouses, whereas "business analytics" utilizes tools and methods for statistical analyses of data. In contrast to these two terms that emphasize business applications, the term "data science" can encompass broader applications in many scientific domains. For the purposes of this proposal, we use the term "analytics" to refer to the process of using data to make decisions. This process involves gathering data, understanding models of uncertainty, making general inferences, and communicating results. Analytics continues to enjoy increasing popularity among businesses. For example, the *Harvard Business Review* in 2012 published an article suggesting that the profession ``data science" was the ``sexiest job of the 21st century." (Reference URL: http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century).

As with all large firms, insurers have opportunities to employ analytics in several capacities, including sales and marketing, compensation analysis, productivity analysis, and financial forecasting, among others. Moreover, insurance companies have special opportunities for employing analytics that are not found in other industries, for instance, in underwriting and ratemaking, claims management, and reserving. Promoting analytics tools in the actuarial society will provide actuaries new opportunities in the data-driven industry.

1.2 Research Themes

Broad Theme 1. Dependence Models

Our goal is to produce novel applications of dependence models in the insurance operations performed by actuaries and quantitative analysts. Simply put, we seek to understand how the dependence among risks should influence analytics in the insurance industry. Dependence could be introduced in a variety of ways. For example, dependence could exist among multiple types of risk, either from different coverages within a contract or different business lines within a firm. Dependence could also be introduced through observations over time, space, or both aspects. In a more complex system, one could argue that all individual risks are related in subtle yet important ways due to latent factors.

To study this problem, we will study applications of predictive models in insurance and risk management that could benefit from dependence modeling. Below are some examples:

- *Consumer choice*: Studying policyholders' joint decisions between multiple products helps the insurer in sales and marketing to customize and bundle products to meet consumer's needs.
- *Mortality modeling*: Modeling lifetimes of dependent lives has historically been used in actuarial pricing of joint life products. It also has important implications in managing longevity risk and pricing mortality linked products.
- *Loss reserving*: Reserves represent a substantial liability for life, health and casualty insurers. Joint modeling of outstanding losses in multiple lines of business is critical in quantifying the variability of the liabilities and determining a reasonable reserve range for future obligations.
- *Portfolio management*: Insurers often are interested in managing a particular block of business when determining a risk management strategy, such as with facultative reinsurance arrangements. The dependence among individual risks is important in the prediction of the losses of the entire portfolio.
- *Cost management*: In a competitive market, insurers survive through a cost efficient management. Studying the relation between claims and expenses, and relations among different types of expenses helps facilitate insurer's cost control and identify cost effective practice.

In the work, we will investigate advanced dependence modeling techniques in various predictive applications, ranging from general multivariate regressions to specific methods for longitudinal data, spatial data, and time to event data. Copula regression models are of particular interest to us. The deliverables include introductions to copula modeling and other multivariate regression techniques (such as multi-state transition models), case studies in predictive analytics, as well as research papers for publication in top journals.

Broad Theme 2. Health Care Analytics

With the proliferation of health care and related data, the role of actuaries has changed over the years with health care analytics becoming a hot area for practicing actuaries. Actuaries are interested in predictive models to help in the analysis of their business, both in premium setting and in the design and evaluation of disease management programs. Data are available in the traditional actuarial forms of claims data with actuaries now interested in collecting and using new types of data, such as risk assessment data, data from the electronic medical (or health) record, or data from public sources such as that collected for use in national health surveys. The traditional tools of setting claims reserves has been supplemented with the use of more advanced statistical techniques to analyze the data.

The availability of additional data and the knowledge of the availability of statistical techniques, like regression modeling, provide a new opportunity for actuaries to become knowledgeable in not only other available tools, but also in how to better use the data that are becoming available. Actuaries would also benefit from the knowledge of what others are doing in similar fields, such as those in health services research or in the statistical literature that are similar to actuarial interests. For example, having access to millions of claims records does not necessarily yield more information than several thousand records, if the big data are not representative of the entire population under study. Another possible area of interest is assessing the consequence of excluding data from a study because they are not complete, such as data with only a partial year of exposure.

The end product of this work will be a set of applied studies that introduce novel statistical methods in situations routinely encountered by practicing actuaries. In addition, one or more of these studies will be submitted for publication to the *North American Actuarial Journal* and similar quality actuarial outlets, including the *Annals of Actuarial Science*. Material will be made available for public use, whether for SOA actuarial study notes, continuing education for professionals, or for use in actuarial programs at universities.

The applied studies will combine the topic and the tool. These studies will utilize data from the Medical Expenditure Panel Study (MEPS), a publically available dataset that are continually updated. The MEPS data are a longitudinal study that offers a rich source of information of the non-institutional US population, and with proper adjustment, are representative of the US expenditures. The annual survey is co-sponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics. Data collection for MEPS started in 1996 and provides data on health care expenditures and other detailed information at a person-level (or household-level) for use by researchers and policymakers. Practicing actuaries are beginning to use MEPS for their work to supplement their own data. These applied studies would provide practicing actuaries more examples of using the MEPS data with more advanced statistical tools.

A topic/tool combination could include a study of defining population health with population health metrics applied to the MEPS data. Other topics could include wellness issues, health care

disparities, risk adjustment, issues related to the Affordable Care Act and the uninsured, Medicare/Medicaid, issues related to mental health, issues related to drugs, end of life care, or analysis of high-cost utilizers. Tools related to analyzing the impact of missing data, big data issues, different predictive modeling approaches, Markov models, decision trees, propensity scores, working with data from a complex survey, and causal inference could be explored. Integral to these tools will be a review of regression modeling and its extensions, such as logistic models and generalized linear models, as well as exploration of data mining techniques.

Broad Theme 3. Analytics to support Insurance and Risk Management Theory

The goal is to broaden the reach of actuarial ideas and theories to influence the development of risk management and insurance theory. In addition to having the grant support core actuarial faculty and their research, we support the broader efforts of our department that encompass applications in micro-insurance, behavioral economics, and corporate finance. Please note that the mechanism is primarily through our doctoral students who serve as the bridge between actuarial faculty and risk management and insurance faculty. Below are some examples of the research topics under this theme:

- *Consumer protection and firm risk-taking*: Risk sensitive customer demand represents an important restraint on risk-taking by financial firms. Consumer protection schemes protect consumers in the event of a financial firm's default, and thus reduce consumer incentives to monitor their financial institutions, leading firms to take excessive risk. This project examines whether consumer protection schemes dull customer sensitivity to firm risk.
- *Micro insurance*: The product has gained popularity in large part due to government support and subsidies, but also because of concerns among the poor about longer life expectancy and decreased ability to rely on children for support. Research interests include demand for micro-pensions and "social impact" factors for micro-insurance.
- *Reputation risk*: We will study insurance against loss from a damaged reputation as well as methods to manage reputation risk, which is included in Solvency II as part of the sources of variability that insurers need to include in their capital models.
- *Asymmetric information*: Information asymmetry is an important aspect of insurance economics and is crucial for optimal contract design and insurance ratemaking. The goal is to examine the adverse selection and moral hazard in various insurance markets and its severity.

The final products of this work will be research articles that are to be submitted to top insurance journals including the *Journal of Risk and Insurance*, as well as journals in business and applied economics.