Allameh Tabataba i University
ECO College of Insurance

Modeling Health Insurance Claims with Extreme Observations

The case study of Iran insurance company

Submitted in Partial Fulfillment of the Requirements of the Degree of Master of Science

In the subject of
Actuarial Science

By:

Mahsa Mir Maroufi Zibandeh

Supervisor: *Dr. Reza Ofoghi*

Advisor: *Dr. Ghadir Mahdavi*

Tehran- Iran
September -2012

To My Dearest Family

Acknowledgement

I am indebted to a great number of people who assisted me during my academic life as well as the completion of the present thesis. Although it is impossible to express my gratitude in words, and no word can show my heartfelt sincere to their encouragement and help, I feel a need to thank them with my simple words.

First and foremost, I want to appreciate my dearest professor, Dr. Reza Ofoghi, for his paramount help regarding the completion of this study by his great knowledge, patience, unending guidance and support, without whose help I could not do this research.

Impossible to pay their encouragement and assistance back, I want to thank my dearest professors, Dr. Ghadir Mahdavi , Dr. Atousa Ghoudarzi who helped me considerably during my MS study at the Eco College.

I appreciate my dear parents, whose existence was a source of considerable energy, and whose unconditional love and care, helped me during all stages of my life and my studies.

I thank all my friends whose moral help assisted me through the demanding stages of doing this research.

Finally, I am grateful to the kind participation of Iran Insurance Company in all phases of data collection for the current study.

Abstract

In modeling insurance claims, when there are extreme observations in the data, the commonly used loss distributions are often able to fit the bulk of the data well but fail to do so at the tail.

One approach to overcome this problem is to focus only on the extreme observations and model them with the generalized Pareto distribution as supported by extreme value theory. However, this approach discards useful information about the small and medium-sized claims which is important for many actuarial purposes. In this study we consider modeling large skewed data using a highly flexible distribution, the generalized lambda distribution, and the recently proposed semiparametric transformed kernel density estimation.

Considering the medical claim of Iran insurance company in 1389 and 1390, we have observed that the data is strongly skewed to the right. By applying our models for no threshold data, the transformed kernel and GPD model fit well to medical claims but GLD model is not good enough in modeling higher claims. For claims above 15,000,000 all models fit the empirical data well. Finally, Value at Risk estimation is given. We suggest using the transformed kernel density to estimate loss distribution based on the results. Consequently, losses can be estimated more accurately. Also the relevant premium can be charged and as a result of that, insurance company can witness a decline in loss ratio.

Key words: Extreme value theory; Kernel density; Value-at-risk; Generalized Pareto distribution; Generalized lambda distribution

Table of Contents

List of abbreviations and symbols

We have tried as much as possible to use uniquely defined abbreviations and symbols. In various cases, however, symbols can have different meanings in different sections. The lists below give the most typical usage. Commonly used mathematical symbols are not explained.

Abbreviation and symbols	**Explanation**
$\xrightarrow{d}$	convergence in distribution
$G_{\xi,\sigma}(X)$	*Generalized pareto distribution with shape parameter* ξ *and* scale parameter σ
EVT	*Extreme Value Theory*
GPD	*Generalized Pareto Distribution*
GLD	*Generalized Lambda Distribution*
ƒx	*Density Function*
ϕ	*Standard Normal Density Function*
MISE	*Mean Integrated Squared Error*
AMISE	*Asymptotic Mean Integrated Squared Error*
K(.)	*Kernel Function*
L	*Likelihood Function*
l	*Log Likelihood Function*
KMCE	*Transformation Kernel Density Estimator of* f *Based on Champernowne Distribution*
μ	*Location Parameter*
σ^2	*Scale Parameter*

Abbreviation and Symbols	Explanation
$\approx$	*Converge*
A	*Tale Index*
θ	*Location Parameter*
e(u)	*Mean Excess Function*
MLE	*Maximum Likelihood*
Π_q	*100q-th quantile of Distribution*
VaR	*Value-at-Risk*
Λ	logarithmic likelihood *function*
$\mathbb{R}$	*Real Line*
M	*Sample Maxima*

Chapter 1: Health insurance

1.1 Introduction

In practice, finding a good-fitting distribution for large data sets which contain some relatively large claim amounts such as insurance claims is not an easy task. In assessing distribution function for claims data, more accurate models must be used in order to be able to price the insurance product more precisely.

Actuaries often find that standard models such as the lognormal, Weibull, and Pareto are able to fit the bulk of the data well but they fail to capture adequately atypical extreme observations. Since these extreme observations are disproportionately important for rating or reserving purposes, some actuaries have analyzed them separately from the bulk of the data.

Extreme value theory (EVT) provides a theoretical foundation for statistical methods that analyze extreme observations. The theory shows for a broad class of distributions that include the normal, gamma, lognormal, Weibull, and many other commonly used distributions conditional excess values over a sufficiently high threshold follow a generalized Pareto distribution (GPD).

This approach requires the determination of an appropriate threshold value. If a threshold is chosen too high, the number of observations above it could be too few to permit accurate estimates of upper quantile values. On the other hand, if the threshold is too low, the GPD may not apply to the moderate observations resulting in biased estimates. Since the moderate observations

usually form a large proportion of the sample, this bias can be a serious problem.

For rating purposes, it is important to utilize the information contained in all claims data. Ideally, one hopes to obtain a single density function that can fit the whole range of data well without the need to leave out any sample information.

When there are extreme observations in the data for modeling insurance claims, the commonly used loss distributions are often able to fit the bulk of the data well but fail to do so at the tail.

One approach to overcome this problem is to focus only on the extreme observations and model them with the generalized Pareto distribution as supported by extreme value theory. However, this approach discards useful information about the small and medium-sized claims that is important for many actuarial purposes. In this study we consider modeling large skewed data using a highly flexible distribution, generalized lambda distribution, and recently proposed semiparametric transformed kernel density estimation. We expect that both these approaches are credible options for the investigator in modeling insurance claims data that typically contain large extreme observations.

In this chapter, the importance of using relevant distribution function in modeling claims as well as the main risk factors in health insurance will be reviewed. In the second chapter, we will present literature review and the theoretical foundations of our approaches will be defined in details in the third chapter. In chapter four we will provide the simulation study and empirical

analysis for medical insurance claims. Finally the results and recommendations will be presented in chapter five.

1.2 Relevance and importance of the thesis

With a population of almost 70 million, Iran is one of the most populous countries in the Middle East. Total healthcare spending is expected to rise from $24.3 billion in 2008 to $50 billion by 2013 that reflects the increasing demand on medical services. Total health spending was equivalent to 4.2% of GDP in Iran in 2005. 73% of all Iranians have health care coverage.

On the other hand, according to declaration of the central insurance, loss rate of health insurance was 46.12% and 39.12% in 1388 and 1389 respectively. Central Insurance of Iran has also declared in its official website that sum of health insurance produced premium by Commercial insurance companies was 9850 billion Rials in 1389 and the compensation was about 9875 billion Rials which means loss ratio is 1.003 in 1389.Furthermore, sum of health insurance produced premium by commercial insurance companies was 15834 billion Rials in 1390 and the compensation was about 14661 billion Rials. Therefore, loss ratio is 0.926 in 1390. Loss ratio for health insurance in Iran insurance company was 0.94 and 0.9 in 1389 and 1390 respectively. In comparison to some insurance lines in our country, health insurance has higher loss ratio.

Insurance line	Paid loss		Produced premium		Loss rate (%)
	Number of paid loss in 1390	1390	Number of written policy in year 1390	1390	
Fire	22,648	1,016,024	2,076,356	4,648,555	21.6
Freight	6,565	297,700	264,232	1,099,782	27.1
Casualty	75,790	346,292	1,499,719	1,104,600	31.3
Driver casualty	10,943	677,601	10,759,881	2,478,808	27.3
Collision	474,850	4,004,473	2,430,424	6,176,089	64.8
Third party	1,327,214	24,827,761	14,858,276	39,415,971	63.0
Health	4,799,545	14,661,816	319,639	15,834,819	92.6
Ship	289	203,235	4,412	459,167	44.3
Airplane	469	204,589	476	645,662	31.7
Engineering	5,026	718,280	29,635	1,958,220	36.7
Liability	76,218	2,140,288	904,121	4,516,653	25.4
Other	193	19,609	5,270	155,056	383.5
Sum of nonlife	6,800,127	49,190,457	33,160,052	79,791,305	28.9
Life	88,952	2,441,642	2,068,890	7,529,903	34.5
Total	6,889,079	51,632,099	35,228,942	87,321,208	29.2

Resource: Central insurance of Iran website

This higher ratio may increase the potential loss of insurance company so that it is not an economical cost-effective activity for an insurance company. Regarding this high loss rate, the insurers have to estimate and forecast their loss distribution with more accurate modeling tools in order to compensate their insured and continue their activity. These estimations aid insurers to be able to set premiums fully align with their losses. Although detecting the extreme losses in this line occurs with lower probability, it has considerable importance and effects on estimation of future losses.

Recently, fat tailed distributions and extreme value theory are used by Iranian risk managements in modeling large claims for different lines such as fire insurance and catastrophe claims in reinsurance. These studies in general show importance of these models in estimating insurance claims.

1.3 The Hypothesis

Standard statistical methodology such as integrated error and likelihood does not weigh small and big losses differently in the evaluation of an estimator. Thus, these evaluation methods do not emphasize on important part of the error: the error in the tail.

Practitioners often decide to analyze large and small losses separately because no single, classical parametric model fits all claim sizes. This approach leaves some important challenges: choosing the appropriate parametric model, identifying the best way of estimating the parameters and determining the threshold level between large and small losses.

Our hypothesis is that the transformed kernel density estimation as a semiparametric approach has better performance and it can fit data better than GLD and GPD models. Therefore, it is more appropriate for heavy tailed distribution and yields more accurate estimation.

1.4 Risk Factors for Individual Health Care Insurance

Obtaining health coverage on the individual market can be more challenging than obtaining group coverage. With group coverage, the "law of large numbers" can help minimize risk because those who are healthy can often offset those who need to use insurance frequently. In the individual market, each applicant is assessed on a variety of risk factors, which determines eligibility and premium of prices from different insurance companies for very similar policies. Insurance companies assign different values to components in your risk profile. The information in your insurance application and your medical history comprise your risk profile.

Insurance companies use historic data and extensive product research to compile various factors and create a benchmark for risk profiles. The companies review your risk profile and the information that is included in your health insurance application. If they are against the company's benchmarks, they decide whether or not to provide you with insurance.

If an insurance company reviews your risk profile and determines that they want to provide you with health insurance, they calculate the insurance premiums based on the information in your individual application and provide you with a quote for the policy. Each company uses a different set of standards and assumptions to calculate that premium. Many different risk factors affect premium prices for most plans.

1.4.1 Physical and Medical Risk Factors

1. 4.1.1 Body Mass Index

On average, people with a high body mass index have significantly higher premiums than those with a normal BMI. People with a high BMI can suffer from, or develop diseases including diabetes, sleep apnea, and heart and joint problems. Additionally, for people with especially high BMIs, the treatment for normal occurrences such as pregnancy have elements of risk, and require specialized treatment or equipment. If you have a high BMI, insurance companies might charge you more for health insurance.

1.4.1.2 Tobacco Use

Insurance companies penalize applicants who use cigarettes, chewing tobacco, or snuff by charging higher insurance premiums or refusing to provide them with coverage. Insurance companies also penalize the people who have quit using tobacco within the last several years. It may take a very long time for someone to recover from a lifetime of tobacco use, and it takes time for his or her risk of disease to abate.

Some studies report that former tobacco users have always a higher risk of cancer and other ailments, even if they never use tobacco again. Many plans provide coverage for smoking-cessation products, including nicotine patches or nicotine gum provided that they are prescribed by a doctor. This coverage benefits both you and the health insurance company because when you quit using tobacco, your health risks begin to decrease immediately.

1.4.1.3 Gender

According to experts, women usually pay higher medical premiums than men for three reasons. Women may be more likely than men to go to the doctor regularly, take prescriptions, and be subject to certain chronic diseases. For insurance companies that offer maternity insurance, a delivery costs $13,000 on average excluding prenatal and postnatal care, which can also be a contributing factor of higher insurance costs.

1.4.1.4 Age

Premium prices are usually lower for younger individuals. Young people have fewer diagnosed and undiagnosed health conditions than older people, and are less likely to develop health problems. Young people are less likely to go to the doctor, although they are much more likely to get into accidents or have significant injuries.

1.4.1.5 Pre-existing Medical Conditions

Insurance companies know that a pre-existing medical condition can be costly. Even if the insurance doesn't cover pre-existing conditions, the policyholder may be more likely to have additional issues related to the pre-existing condition. Many plans don't sell insurance to people with specific pre-existing conditions.

1.4.1.6 Family History

For applicants whose families have a history of certain medical conditions or ailments, the premiums may be higher. Applicants with a family history of cancer, for example, may pay higher insurance premiums than applicants with no family history of cancer.

1.4.1.7 Denial of Coverage

In many states, insurers are permitted to deny coverage based on an applicant's health history if it includes certain conditions. Some of the more

common past conditions include heart disease, stroke, rheumatoid arthritis and kidney stones. This can make it difficult for a person with this type of history to obtain coverage anywhere on the private market without having to pay an exorbitant premium.

1.4.2 Lifestyle and Personal Risk Factors

Insurance companies have decades' worth of data that they analyze regularly to find relationships between risk and a multitude of different factors. The insurance company assumes that the people who share these factors have similar risks and price their policies accordingly.

1.4.2.1 Profession

The individuals exposed to hazardous chemicals or radiation, or the ones who work in jobs with high injury rates have much higher insurance premiums than those with less dangerous jobs. There is also a trend to charge higher insurance premiums to those who have extremely sedentary jobs, which require them to sit in one place all day or perform repetitive tasks or motions. People who work in sedentary professions may have an increased risk of cardiovascular disease. International business travel may also impact the cost of health insurance.

1.4.2.2 Zip Code

Whether due to a shared climate, a cultural aversion to exercise, or a lack of healthy food options, people who live in the same area tend to have similarities in their risk profiles. If your neighbors or coworkers have high health insurance premiums, you may also pay more for health insurance. This does not reflect upon you as an individual, but instead reflects the insurance company's statistics, which may show that people in your area aren't as healthy as people in other areas. Move to another area, and you might see a dramatic drop in premium prices.

Furthermore, insurance companies factor the cost of complying with local laws into the cost of your insurance premium. If the cost of doing business in your area is less expensive for the insurance company, that cost will be passed along to you.

1.4.2.3 Marital Status

For reasons that haven't been determined, married couples live longer and are generally healthier than single people. Getting married usually causes a drop in health insurance premiums, although the benefits are greater for men.

1.4.2.4 Previously Uninsured

If you've been without health insurance coverage, or if you are buying insurance for the first time, insurance companies may charge you a higher premium. The insurance company anticipates that previously uninsured policyholders have been waiting to get covered medical care for various

ailments. The insurance company also operates under the assumption that the previously uninsured will want to visit a dentist or physician, to begin receiving general health and wellness checkups. In addition, if you haven't received regular medical care including checkups, any health problems you have may be exacerbated. The insurance company takes all of this into account, and they price policies accordingly.

1.5 Health insurance

Health insurance is one of the insurance services. Humankind has always been in danger of lots of diseases, so people need to sponsor the charge of all this treatments. In order to help people in those situations, insurance companies represent various kinds of health insurance. Also, government employee or industrial and production units in most of the countries use the group health insurance.

Health insurance is one of the major fields of insurance in the world. This kind of insurance is offered as a group insurance in Iran.

1.5.1 Items that are covered by health insurance

In the insurance contracts, disease is defined as any adverse change and physical condition or any disruption in the normal and ordinary members of the body that is detectable by medical authorities and is not related with voluntary actions of insured.

Therefore, it is the insurer's obligations to repay all medical and hospital costs of each insured in accordance with contract terms and in compliance with the franchise agreement.

Franchise is a percentage of the cost health that should be paid by the insured. Franchise is inversely related to the premiums. If the amount of franchise is set at a lower level, the premium increases and vice versa. The franchise amount and the type of cover which are subject to franchise in the contract are determined by the insurer. Insurer may introduce different levels for franchise; hence insured can choose one of the levels.

In general, obligations of the insurer include the costs of doctor visits, surgical dressing, all kinds of medical tests, anesthesia costs, transportation