Allameh Tabataba'i University

E.C.O. College of Insurance

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

In

Actuarial Science

The Application of Multivariate Probit Models for Conditional Claim-Types (The Case Study of Iranian Car Insurance Industry)

By

Hojjat Shafiei Dizaj

Supervisor

Dr. Reza Ofoghi

Advisor

Dr. Reza Ghazal

September 2011

Abstract

This study considers statistical modeling of the types of claim in a portfolio of insurance policies. For some classes of insurance contracts, in a particular period, it is possible to have a record of i) whether or not there is a claim on the policy, ii) the types of claims made on the policy, and iii) the amount of claims arising from each type. A typical example is automobile insurance where in the event of a claim, the amounts that arise from say injury to oneself, damage to one's own property, damage to a third party's property, and injury to a third party can be observed.

Modeling the frequency and the severity components of the claims can be handled using traditional actuarial procedures. However, modeling the claim-type component is less known. In this study, I try to analyze the distribution of claim types using multivariate probit models, which can be viewed as latent variable threshold models for the analysis of multivariate binary data. Following Young et al (2008), I investigate the appropriateness of fitting a multivariate probit model to the conditional claim-type component in which the outcomes may in fact be correlated, with possible inclusion of important covariates. Then I compare the Multinomial Logit (MNL) model results with those from Multivariate Probit (MVP) model

Based on result of data analysis fitting a multivariate probit model to our data is superior to the MNL model when the outcomes are correlated. The flexible specification of the multivariate probit model allows an actuary to estimate the extent of correlation between different claim types. We run three separate univariate probit regression on policies for each claim type using covariates. In all three tests, the null hypothesis of homoskedasticity was strongly rejected in favour of heteroskedastic disturbances.

The estimated correlation coefficients, $\hat{\rho}_{sj}$, between each of the three claim-types are statistically significant. Surprisingly, the correlations between the disturbances of the "Injury" and "Own" equations and the "Own" and "Property" equations are negative. This suggests that the unobservable factors which increase the probability

of claiming with respect to “Injury”, for example, actually reduce the probability of claiming with respect to “Own”; a similar interpretation also applies to the negative correlation between “Own” and “Property”. The positive correlation between the “Injury” and “Property” equations is intuitive. Here, unobservable factors which increase the probability of claiming with respect to “Injury” also increase the probability of claiming with respect to “Property”. One can think when accident was occurred if third parties face with property damages then more possible that arise injury to third party. Furthermore, the likelihood ratio test for independence between the disturbances is strongly rejected, implying correlated binary responses between different claim-types.

Key Words: Actuarial science, Correlation; Insurance; Claim-Type Component; Multinomial Logit , Multivariate probit.

Table of Contents

ABSTRACT ____ I
TABLE OF CONTENTS ____ III
LIST OF TABLES ____ V
LIST OF FIGURES ____ VI
ACKNOWLEDGEMENTS ____ VII
DEDICATION ____ VIII

PART I ____ 1

CHAPTER1: A REVIEW ____ 2

1.1. Introduction ____ 3
1.2. Types of insurance: ____ 4
1.3. General Background ____ 7
1.4. Relevance and Importance of the subject ____ 10
1.5. Methodology ____ 15
1.6. Data requirements: ____ 17

CHAPTER 2: LITERATURE REVIEW ____ 20

2.1. Introduction ____ 21
2.2. Review of Literature ____ 22
2.3. Literature on claims frequency/severity ____ 25

CHAPTER 3: THEORETICAL FOUNDATIONS ____ 28

3.1. Introduction ____ 29
3.2. Statistical Model ____ 30
3.3. Multinomial Logit Model ____ 32
a. Multinomial Logit Model: Model Specification and Notation ____ 34
3.4. The Multivariate Probit Model ____ 38
a. Multivariate Probit Model: Model Specification and Notation ____ 40
b. The formulation of the multivariate probit model ____ 42
3.5. Research approach: ____ 44
a. Data ____ 44
b. Realm of Research: ____ 44
c. The period of investigation: ____ 44
d. Research site: ____ 44
e. Population (N): ____ 45
f. Sample (n): ____ 45
g. Sampling methods: ____ 45
h. Methods for analyze data and test hypotheses: ____ 45

CHAPTER 4: EMPIRICAL MODEL ____ 47

4.1. Introduction: ____ 48
4.2. Market description and Market characteristic: ____ 49
a. Market description ____ 49

b. Markets characteristic 50
4.3. Data Description 50
4.4. Data gathering 53
4.5. Descriptive Statistics 54
a. Insurance companies: 54
b. Gender: 56
c. Damage locality: 57
d. Vehicle type: 59
e. Vehicle brand: 60
f. Usage type of vehicle: 62
g. Vehicle Age: 64
4.6. Empirical investigation 66
a. Estimation and discussion of results 70
CHAPTER 5: SUMMARY AND CONCLUDING REMARKS 79
BIBLIOGRAPHY 83
PART II : APPENDICES 86

List of Tables

Table 3. 1: Value of M, by Claim Type 31

Table 4 . 1: Description of Covariates varibles 53

Table 4. 2 : Frequency table of losses for insurance companies 54

Table 4. 3 : Frequency table of losses for gender 56

Table 4. 4 : Frequency table of damage locality 57

Table 4. 5 : Frequency table of damages for vehicle kind 59

Table 4. 6 : Frequency table of damages for vehicle System 60

Table 4. 7 : Frequency table of damages for usage type of vehicle.....62

Table 4 . 8 : Frequency table of damages for vehicle made year 64

Table 4 . 9 : Sample space of the conditional claim-type variable M 66

Table 4 . 10 : Summary statistics for Premium sorted by Claim-Type. (At 1000Rial).....67

Table 4 . 11 : Sample unconditional and conditional probabilities.....68

Table 4 . 12 : Results of fitting the probit model (for injury) to the policies 73

Table 4 . 13 : Results of fitting the probit model (for own) to the policies.....73

Table 4 . 14 : Results of fitting the probit model (for property) to the policies.....73

Table 4 . 15 : Results of fitting the TVP model to the policies.....74

Table 4 . 16 : The average marginal effect of Premium on the marginal claim-type probabilities. 75

Table 4 . 17 : Results of fitting the MNL model to the policies.....78

List of Figures

Figure 3.1 A graphical representation of the model in 3.6 under a full correlation structure. Observed nodes are shaded. 41

Figure 4. 1 : Column chart of losses for insurance companies..... 55

Figure 4 . 2 : Pie chart of losses for gender..... 56

Figure 4 . 3 : Column chart of damage locality 58

Figure 4 . 4 : Column charts of damages for vehicle kind 59

Figure 4. 5 : column charts of damages for vehicle System 61

Figure 4 . 6 : Column charts of damages for usage type of vehicle 63

Figure 4 . 7 : Column charts of damages for vehicle made year 65

Acknowledgements

Many people have provided me with support, kindness, and encouragement during my endeavor to complete my MSc degree. While there are too many to name, I would like to single out a few that I would especially like to thank.

First, I would like to thank my honorable supervisor and advisor, Dr. Reza Ofoghi and Dr. Reza Ghazal. They have provided adequate support, guidance and motions that made it possible for me to complete my thesis.

Second, I would like to thank my parents. Though they don’t be with me and both died. But they taught me that it is important to work hard for what I want, and that is the process of working for something that develops character. They supported me and gave me praise for my accomplishments. Thanks for your love and support.

Last, I would like to thank my family, my brothers and sisters, for their encouragement, patience, and inspiration. Without their love and support, my thesis would mean little. I would like to consider this thesis done for and by all of us. Thanks for your love, attention, support and guidance.

I thank God that helps me for this thesis. This is an example that through him all things are possible.

To My Family:

My Dear Parents

My Brothers

My sisters

این پایان نامه

با همکاری و حمایت پژوهشکده بیمه

انجام پذیرفته است.

Part I

Chapter1: A Review

1.1. Introduction

Due to population growth, urban expansion and development of roads, and multiplicity trips within urban and suburban and need to use much of the vehicles, cars play a special role in everyday life. Car is a tool that almost at late nineteenth century and into the epidemic form comes to human life in human societies and play important role in transporting passengers and goods transportation.

Although vehicle brings with itself welfare, comfort and speed but as well, causes growth of traffic accidents and incurs heavy bodily and financial damages to owners, passengers and third parties. The peril of collision and clash increased by most use of the vehicle and this was serious problems for users and losers in car accidents. The only way to compensate and assist to loser from this phenomenon is to eliminate or reduce the undesirable effects in loser's lives by use of insurance. In this context, providing variety of insurance coverages for owners of motor vehicles is inevitable.

Insurance companies with understanding the needs of human communities have offered different compensation plans. In car insurance, sometimes insurer obliges to compensate damage pertinent to insured's vehicle and sometimes compensate the damage pertinent to third party.

A primary attribute of the actuary has been the ability to successfully apply statistical techniques in the analysis and interpretation of data. In this thesis I analyze a highly complex data structure and demonstrate the use of modern statistical techniques in solving actuarial problems. Specifically, I focus on a portfolio of automobile insurance policies and, by analyzing the historical data

drawn from this portfolio, we are be able to revisit some of the classical problems faced by actuaries dealing with insurance data. This thesis applies statistical models in Iranian car insurance market using detailed- micro level records of automobile insurance policies.

This thesis will be structured as follows: Chapter 2 will provide a review of the literatures. In Chapter 3, the methodology used will be presented and discussed. In Chapter 4, we present the results of the empirical model. Chapter 5 concludes.

1.2. Types of insurance:

Often all adverse events from car accidents are mostly due to one of following three modes:

- The driver is at fault and financial and bodily loser is quite innocent.
- Loser is at fault and driver is completely innocent.
- Both the driver and loser are at fault.

Each of the above three cases diagnosed solely via traffic's officers. If the culprit and loser do not have agreement, concluded according to police officers hearing. Mention that to settle a heavy loss, insurance companies need reports from the police officer present at the scene.

Car insurance in the world has a special welcome and it is one of the biggest selling insurance. Generally car insurance companies compensate for losses caused by traffic accidents to the vehicle, its passengers and third parties.

Insurance companies offer different covers in the field of car insurance which each will be discussed briefly below.

- **Collision insurance**

In this type of insurance, the insurer plights that in case of damaging the car by any of the events covered under the insurance contract (fire, explosion, accident, reversal, falling and theft and additional risks) to pay for the damaged vehicle repairment or pay compensation in accordance with conditions of insurance policy.

Based on the conditions of contract, holder or owner of the vehicle insure own car against the dangers till in case of accidents and damages arise from accident to the vehicle, and insurance company compensate losses based on conditions of the policy.

- **Third party insurance**

According to the compulsory insurance of civil responsibility of land motor vehicles owners against third parties law, all owners of land motor vehicles and various trailers connected to the said devices and railway trains, whether rightful or legal persons are responsible for compensation physical and financial losses that incur to third parties in the events of said vehicles accident or them cargo and hence they must to insure themselves responsibilities into domestic institution.[1]

Third party insurance basically compensate for damages to third parties (who are endangered by the insured or the driver) which includes:

1 - central insurance of Iran

A) **Financial losses:**

generally financial damages payable under the insurance policy include all the financial losses due to civil liability of motor vehicle owners and drivers in traffic accidents. In case of insurance purchased by holders of the said vehicle, insurance companies act to compensation up to commitments contained in the text of the third-party insurance.

B) **Bodily damages:**

bodily damages of third party insurance include all bodily damages and indemnity such as death, injury, absolute permanent or relative disability as well as medical costs resulted from car accidents. It is compensated up to the purchased commitments and contained in the third party insurance.

Third party insurance is a one of liability insurance and despite the fact that mercantile insurance is generally optional, but in year 1347 in Iran in an effort to comply with supportive systems in the world, it became compulsory.

– **Passenger accident insurance**

The passenger is person that at the time of the incident was inside the insured car, whether the car is in motion or at stop. Public means of vehicle passenger is including the driver, helpmate driver and vehicle passenger.

Basically, in the event of an accident, two types of damage are observed: a) damages to the third parties that necessarily and by law, incident blamer required compensating it and, b) Other relevant damage.

First group of losses can be compensated from third party insurance. Second group of losses comprises of two parts: damage to property and death or injury. Damage to property can be compensated from collision insurance and death or

injury should be reimbursed from passenger accident insurance. However, if the passengers have been insured, the insurance company will compensate in case of death or injury.

This policy is offered as supplemental coverage along with third party insurance. Under this coverage, the driver and group of passengers in insured vehicle who are injured due to accidents and are not considered as third party can be compensated. Based on this policy, the medical costs and death and maim indemnity is compensated up to the ceiling specified in the insurance policy.

1.3. General Background

Vital role of car in routine human life and being tangible dangers caused by traffic accidents, cause particular welcomed the purchase automobile insurance in the world and it is considered one of the biggest selling insurance. In Iran, share of the premium in automobile insurance from market production is about 61 percent[2]. Car insurance redress losses and damages and injuries resulting from accidents that incur to third parties, passengers or the car owner by use of vehicles.

In the motor vehicle insurance industry, a large number of insurance companies offer very similar products. The utilization of a set of criteria and actions to insure a vehicle is known as risk selection. This means that the conditions of the insurance policies allow the expected claims ratio to come as close as possible to the real claims ratio. The rules of contract hold great

2 - Bimeh Markazi Iran (Central Insurance of Iran)

importance for proper conduct in the insurance business and a technical and commercial balance must be sought.

Insurance is based on community and risk and each insured transfers his risk to this community. If all the insured have not equal risk, just that every member of coffer put funds according to their risk. So the main work for actuary is to determine tariffs based on classification insured to homogeneous groups so that each person pays the fairest premium.

Insurance companies sell protection to policyholders against many types of risks: property damage or loss, health and casualty, financial losses, etc. In return for this risk protection, insurance companies receive a premium from the policyholder that is used to cover expenses and the expected risk. For longer-term risk protections, part of the premiums is invested to get higher yields.

Although insurance companies compete for the same business, economic forces dictate that the loss experience of insurers can differ. During the sales process, insurers use different underwriting standards and pricing structures to attract different mixes of business. After an accident, insurers differ in their procedures to settle a claim, including legal, and the calculations of claims adjustments, thereby realizing different loss experience across companies. Moreover, there are issues of moral hazard, a term used to refer to the tendency of the insured to alter its behavior in the presence of an insurance coverage. Thus, it is possible that an insured with a policy from a particular company may have a different claims experience than if the insured were contracted with another company. For example, some insurers establish premium rating systems

that encourage policyholders to avoid reporting minor losses, even if they are contractually covered under the insurance policy.

To determine the premium, expected amount of damage claims for each contract should be estimated first. This conditional mathematical expectation on the information of each insured and own contract (that they are called input variables) is base to calculate gross premiums for the insured. Also by estimation damage claims for different types of damages can be has a better determination for gross premium.

Insurers usually group individual risks such that the risks within each group are as homogeneous as possible in terms of certain observable risk characteristics. A common premium for the group, also known as the manual premium, is then calculated and charged. The grouping is made primarily to reach a fair and equitable premium across all individuals. Such grouping also helps isolate a large group of independent and identically distributed risks so that the law of large numbers can be invoked during the claims prediction process. This minimizes variability in the claims experience precise, thereby causing the risks within each group to be not entirely homogeneous. An unknown number of unobservable traits will always contribute to the possible presence of heterogeneity among the individuals.

In premium calculation, which also requires prediction of claims, historical claims experience provides an invaluable insight into the unobservable characteristics of the individual risks. Furthermore, it is a common practice to allow for past claims experience of the insured individual in claims prediction and in premium calculation. This exercise is known as experience rating and is

generally made for the purpose of reaching a fair and equitable insurance premium rate.

1.4. Relevance and Importance of the subject

One of the most important aspects of a typical insurance company is predicting claims as accurately as possible. Actuaries require accurate predictions for pricing, for estimating future company liabilities, and for understanding the implications of these claims to the solvency of the company.

Actuaries and other financial analysts that work with short term coverage such as automobile insurance and healthcare expenditures typically have massive amounts of in-company data. Actuarial applications can be performed when modern statistical methods are applied to detailed, micro-level automobile insurance records. With modern computing equipment, analysts can readily access data at the individual policyholder level that termed "micro-level." Actuaries use statistical models to summarize micro-level data that subsequently need to be interpreted properly for financial decision-making. For example, automobile insurers typically differentiate premium rates based on policyholder characteristics such as age, gender and driving history. Gourieroux and Jasiak (2007) have dubbed this emerging field the "micro econometrics of individual risk."

General modeling method for operating loss is considering the frequency and severity of damage. If the insurance company able to classify damages in accordance with different types, correlations between the types of damage might appear. The claim-type methods able insurance companies to find influence of one type of risk on resulting damages from other types. Expansion cumulative