CITY UNIVERSITY OF HONG KONG
香港城市大學

Zero-inflated count data models for claim frequency in general insurance
多零數據模型於普通保險
索償頻數上之應用

Submitted to
Department of Management Sciences
管理科學系
in Partial Fulfillment of the Requirements
for the Degree of Master of Philosophy
哲學碩士學位

by

Yip Ching Han, Karen
葉靜嫺

March 2004
二零零四年三月
Abstract

One of the distinctive features of claim frequency data collected from general insurance is that it is often zero-inflated. Therefore, traditional application of Poisson and negative binomial distributions for model fitting may not be adequate due to the presence of excess zeros. The source of excess zeros may come from the unreported claims due to minor losses caused by the deductible agreement and/or under the no claim discount (NCD) system in motor insurance. Spurious dispersion appears as the number of observed zeros exceeding the number of expected zeros under the Poisson or even the negative binomial distribution assumptions.

The purpose of this study is to illustrate and discuss alternative methods of modeling claim frequency distribution in general insurance with the presence of many zero counts. Various zero-inflated count data models are considered. In addition, the use of quasi-likelihood approach is explored to address the over-dispersion problem. A motor insurance data set is used to demonstrate the application of various zero-inflated models. Performances of the models are evaluated by the log-likelihood and related statistics.

It is found that the zero-inflated models, especially the zero-inflated double Poisson regression model, provide a substantially better fit than traditional Poisson model and negative binomial model in predicting the insurance claim count. The quasi-likelihood modeling approach of zero-inflated Poisson model reduces to the negative binomial model, but with a different parameterization.

In conclusion, the zero-inflated count data model would be a scrumptious choice in modeling the claim count data in general insurance as it extends the Poisson model, incorporates the excess zeros and the extra dispersion in the Poisson part. In regression setting, with an appropriate set of risk factors, the level of risk of customers can be correctly evaluated in accordance to individual characteristics. The findings in this thesis provides insurance
practitioners a better and more precise method for the modeling of the claim frequency distribution, which in turn can help in refining their ratemaking and hence the loss reserving process.
# Contents

## Abstract

## Acknowledgements

### Chapter 1 Introduction
- 1.1 Background of study
- 1.2 Motivation of study
  - 1.2.1 The motor insurance data set
  - 1.2.1.1 Assumptions of the motor insurance data set
  - 1.2.2 A subset of the motor insurance dataset
- 1.3 Research objectives
- 1.4 Outline of this thesis

### Chapter 2 Literature review
- 2.1 Introduction
- 2.2 Unconditional versus conditional zero-inflated models
  - 2.2.1 Zero-inflated Poisson model
  - 2.2.2 Modified zero-inflated Poisson model
  - 2.2.2.1 Zero-inflated negative binomial model
  - 2.2.2.2 Zero-inflated generalized Poisson model
  - 2.2.2.3 Zero-inflated modified power series distribution
  - 2.2.2.4 Zero-inflated double Poisson model
- 2.3 Score test
- 2.4 Mixture of Poisson distributions
- 2.5 Quasi-likelihood model

### Chapter 3 The Unconditional Zero-inflated Count Data Models
- 3.1 Structure of the unconditional zero-inflated model
- 3.2 Zero-inflated Poisson model
- 3.3 Zero-inflated negative binomial model
- 3.4 Zero-inflated generalized Poisson model
- 3.5 Zero-Inflated double Poisson model
- 3.6 Statistical results of the zero-inflated models with unconditional approach
  - 3.6.1 Goodness-of-fit statistics and model selection criteria
  - 3.6.2 Unconditional distribution fits
  - 3.6.3 Unconditional regression model results

Technical notes for Chapter 3
Chapter 4  The Conditional Zero-inflated Count Data Models  35
4.1 Structure of the conditional zero-inflated model  35
4.2 Zero-inflated truncated Poisson model  36
4.3 Statistical results of the zero-inflated models with conditional approach
   4.3.1 Conditional distribution fits  37
   4.3.2 Conditional regression model results  38

Technical notes for Chapter 4  41

Chapter 5  The Quasi-likelihood Approach for Zero-inflated Poisson Models  43
5.1 Structure of the quasi-likelihood approach for ZIP models  43
   5.1.1 Features of the quasi-likelihood model  45
5.2 ZIP model with quasi-likelihood approach  45
   5.2.1 Relationship between QLZIP and NB models  46
5.3 Statistical results of the quasi-likelihood approach for zero-inflated Poisson models
   5.3.1 Distribution fits of the quasi-likelihood approach  47
   5.3.2 Regression fits of the QLZIP model  49
5.4 Further explorations about the QLZIP model  50
5.5 Discussion on the quasi-likelihood models  53

Technical notes for Chapter 5  53

Chapter 6  Concluding Remarks  57
6.1 Summary  57
6.2 Research objectives achieved  58
6.3 Discussion  59

References  61

Appendices  67
Appendix A  Fields in the motor insurance data set  67
Appendix B  S-plus programs of the ZIP models  69
Appendix C  GUASS programs of the ZIP models  79
Appendix D  GUASS programs of the ZINB models  83
Appendix E  S-plus program of the ZIGP distribution  87
Appendix F  GUASS programs of the ZIGP models  91
Appendix G  GUASS programs of the ZIDP models  96
Appendix H  Programs of the ZITP models  99
Appendix I  GUASS programs of the QLZIP models  102