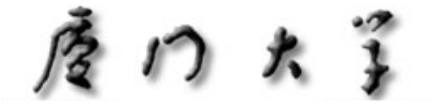
学校编码：10384
学号：15420111151911

分类号__密级_
UDC___



硕 士 学 位 论 文

# 有限混合广义线性模型在车辆保险理赔频率拟合中的应用

# Finite mixture of generalized regression models with application to vehicle insurance claim frequency modeling

唐荣

指导教师姓名：张志强教授
专 业 名 称：经济信息管理学
论文提交日期：2014 年 3 月
论文答辩时间：
学位授予日期：

答辩委员会主席：
评 阅 人：

2014 年 3 月

# 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

# 摘 要

机动车辆保险在非寿险业务中有着十分重要的地位。近年来，中国机动车辆保险保费总收入在财产保险保费总收入中的比例一直保持在60%以上，可见它的发展变化对整个非寿险市场的影响巨大。而机动车辆保险最重要的一环即是机动车辆保险定价，因而定价中机动车辆保险理赔频率的拟合则成为研究人员关注的焦点。

广义线性模型在车辆保险定价中应用广泛，尤其是在理赔次数或者频率拟合上。在保险数据中，过离散的现象，即方差大于均值，以及零膨胀的现象，即理赔数据为零的个体量庞大的现象，都非常普遍。这是因为现实保险条款中都存在免赔额，低于免赔额的损失保险公司不予赔偿，同时由于奖惩政策的存在，在高于免赔额不多的损失发生时，被保险人为了在下一年交付较低的保费，会考虑自己承担损失而不报损。为了解决过离散和零膨胀问题，负二项回归、零膨胀模型、两阶段回归等模型相继被提出，这些模型都是建立了在广义线性模型基础上的。

有限混合广义线性模型是在其他领域应用发展起来的一个包容性模型。但少见于机动车辆理赔频率拟合中。本文介绍了有限混合广义线性模型，并将其应用于机动车辆保险中保险理赔频率的拟合。鉴于车险数据的变量繁多，在变量选择上，本文对变量进行整合之后，应用对数似然函数贡献大小原则，择出关键变量，用于模型之中。与前人常用的变量选择相比，在模型拟合上取得了更优的结果。

本文首先介绍了机动车辆保险的基本知识，其次详述了有限混合广义线性模型的结构和特点。且基于某保险公司机动车辆理赔数据进行实证分析，基于Pearson 相关系数和变量整合，比较对数似然函数贡献，选取模型解释变量，应用于有限混合广义线性模型。最后，鉴于车辆保险理赔频率变量内部可能存在的非独立性，本文介绍了有限混合二维广义线性模型，希望今后能够用于日益复杂的保险关系数据的分析。

**关键词**：有限混合；理赔频率；零膨胀

# Abstract

Vehicle insurance occupies an important place in non-life insurance. In recent years，the total premium for vehicle insurance in china has always been 60% of that for non-life insurance, indicating a significant influence on the whole insurance market. One of the most important parts in vehicle insurance is ratemaking, that is why so many researchers focus on the topic of fitting the claim frequency for price making.

Generalized linear model (GLM) has been widely applied in the vehicle insurance pricing, especially in modeling the claim frequency or amount. The claim frequency data, with a big probability, is likely to be over-dispersed and zero-inflated, that is, variance is greater than the average and too many zeros in the distribution. This phenomenon is partly due to the franchise clause, in which loss below a certain limit would get no compensation from the insurer. At the same time, the bonus-Manus system (BMS) or the no claims discount system (NCD) tend to encourage the insured not to report small losses. In order to describe the overdispersion and zero-inflation, models like negative binominal regression, zero-inflated regression, and hurdle regression were put forward. The finite mixture of generalized linear model is an inclusive model developed in other fields.

The aim of this essay was to introduce the finite mixture of generalized linear models (FMM), with application in fitting the claim frequency in vehicle insurance. The way of choosing the explanatory variables featured this essay. Due to the large numbers of variables, the paper selected some key variables by the principle of more contribution to the log-likelihood function. When applied into the finite mixture of generalized linear model, we got a much better result than other simply chosen variables in former researches.

The paper was started by the basic knowledge of the vehicle insurance, and then describes the structure and characteristics of the finite mixture model in details. Then a case in real world was studied, the log-likelihood function accompanied by principal component analysis (PCA) and Pearson coefficient of correlation was put in

use to select variables for the model. The method was turned out to be excellent. In the end, considering the non-independence within the response variables, the paper introduced the bivariate situation of the model, hoping that it will be used in claim frequency modeling in the future.

**Key words：** finite mixture model; claim frequency; zero-inflation.

# 目录

# Table of Content

# 第一章 导论

## 1.1 选题背景与研究意义

汽车保险，属于运输工具保险，是车辆由于遭受自然灾害或意外事故造成的损失或民事赔偿责任的综合性财产保险。1983 年，我国将汽车保险改称为机动车辆保险。

第十九世纪末期，汽车在欧洲流行，汽车交通事故导致的意外伤害和财产损失不断增加，汽车保险随之出现。汽车保险的保费收入在发达国家通常要占财产保险总保费的 50%左右。而随着中国的经济持续快速健康地发展，群众的消费水平不断提高，生活质量日渐改善，机动车辆的消费需求逐渐增加，中国加入世界贸易组织后，为了迎合世界发展新形势，中国保险监督管理委员会已经将车险费率的厘定开放为市场化。2006 年 7 月 1 日，国务院出台的《机动车交通事故责任强制保险条例》正式施行；2013 年 3 月 1 日，国务院发布《国务院关于修改〈机动车交通事故责任强制保险条例〉的决定》施行，机动车辆保险在产险市场中的核心地位日渐凸显，成为财产保险的重要组成部分。 根据中国保险监督管理委员会的统计数据，2013 年，我国产险业务原保险保费收入 6212.26 亿元，同比增长 16.53%，而其中，产险业务中，交强险的原保险保费收入达到 1258.86 亿元，同比增长 12.99%。可见机动车辆保险的发展变化对整个非寿险市场的影响巨大。

机动车辆保险，具有较高的风险性，因为其标的物是活动的，经常处于运动状态。机动车辆保险拥有了越来越多的消费者，每个人的投保动机、所处环境以及个体之间的特点不尽相同。准确地分析人们的投保倾向，分析机动车辆保险市场的风险因素，合理地厘定机动车辆保险产品的价格，把握人们对机动车辆保险消费的有效需求，对于保险公司立足保险市场，保证公司稳定健康发展，并促进社会的稳定，有着极其重要的意义。

机动车辆保险产品开发中，最重要的一环便是根据风险因素影响情况，针对个体，确定机动车辆保险产品的价格。车辆保险定价中，车辆保险理赔数据的拟合是前驱步骤。因此，本文以机动车辆保险理赔数据为基础，针对索赔频率分析，

研究有限混合广义线性模型在理赔频率数据拟合中的应用，探究模型中解释变量的选择方法，旨在为机动车辆保险定价提供有用的参考信息。

## 1.2 文献综述

### 1.2.1 国外研究文献综述

国外机动车辆保险起步早，发展成熟，研究成果也极其丰富，参考价值大。

广义线性模型,是由 Nelder 和 Wedderbum (1972)提出的，他们制定了统一的处理正态数据的模型框架，包括线性回归 、方差分析、logistic 回归和对数线性模型 。 Steven Haberman 和 Arthur E.Renshaw（1996）回顾了广义线性模型在 20 世纪 80 年代以来在精算问题中的应用，如死亡率、多状态模型、失效、保险费率厘定和准备金等。 Piet DE Jong，Gilltan Z. Heller（2008）从基础知识出发，系统详细的介绍了广义线性模型及其推广模型在分析保险数据中的应用。

在机动车辆保险领域，针对机动车辆保险数据中存在的过离散、零膨胀现象，Karen C.H. Yip 和 Kelvin K.W. Yau（2005）比较了零膨胀泊松回归模型、零膨胀负二项回归模型、零膨胀广义泊松回归模型、零膨胀双泊松回归模型，关于解决零膨胀的机动车辆保险理赔数据的效果，得到了零膨胀双泊松回归模型拟合效果最好的结论。在变量选择上，直接选取了对泊松似然函数贡献最大的前五个变量。Mathew Flynn，Louise A（2009）介绍了两种更灵活的广义线性模型：零膨胀模型和两阶段模型。并提出了用于合并风险等级的基于数据挖掘的 CHAID 树模型。模型直接借鉴应用了 Yip 和 Yau 选取的五个变量。Morata(2009)针对机动车辆保险数据的非独立性问题，采用了二维泊松回归及其推广模型进行拟合并比较，充分证明了二维泊松回归模型的优越性； Morata(2012)应用了二维泊松回归模型的有限组合来解决了保险数据中的过度离散化问题。Morata 文中的变量都延用 Pinquet(2001)中使用的经验变量。

有限混合广义线性模型，已经在很多领域得到广泛的应用，比如生物学、基因学、药学、市场营销学等等。Igor Vladimir Cadez（2002）探讨了有限混合广义线性模型在高维序列和转换数据集中的应用。重在研究模型拟合算法，对于