

# 车险费率厘定研究—基于驾驶行为视角综述与实证分析

张琳, 谢亚凤

(湖南大学 金融与统计学院, 湖南省、长沙市, 410079)

**摘要:** 本文从驾驶行为因素四大量化指标出发, 运用相关分析、主成分分析方法, 选取连接保单数据与驾驶行为数据的重要变量, 并寻找其替代变量, 引入车险费率厘定模型中。同时, 当存在车型因子、驾驶行为因素等多水平费率因子时, GLM 估计结果不可信, 通过构建 HGLM 解决此类问题。实证结果表明, HGLM 拟合效果优于 GLM, 加入驾驶行为因素可进一步提高 HGLM 的拟合效果。

**关键词:** 费率厘定; 驾驶行为因素; GLM; HGLM

**中图分类号:** F840.65

**文献标识码:** A

## 0 引言

全国机动车保有量创新高, 涉及上亿中国车主切身利益的关键性险种, 车险定价至关重要。截至 2016 年底, 全国机动车保有量达 2.9 亿辆, 机动车驾驶人 3.6 亿人。车险改革从未中断, 改革经验是活到最后的是大型保险公司, 如何在竞争中活下来甚至借助车险费改实现“弯道超车”, 是所有保险公司必须思考的问题。突破现有粗放式定价, 探索更公平、合理的定价机制, 实现精细化定价是解决问题的关键, 也是车险费率市场化的必然要求。

费率厘定研究包括费率因子研究与费率厘定模型研究。费率厘定因子方面。驾驶行为因素主要通过驾驶里程、驾驶时间、驾驶速度、道路状况等四大指标量化。Iqbal 和 Lim, (2006) 在《A Privacy Preserving GNSS Based Pay As You Drive Insurance Scheme》一文中指出设计 PAYD 保费计算公式, 可以表示为  $\frac{\text{Premium}}{\text{Second}} = \text{基准保费} \times \text{WCR} \times \text{LCR} \times \text{RHR} \times \text{SR} \times \text{RR}$ , 根据车龄、车型、保额、个人驾龄、索赔纪录、年龄、性别等计算出来的单位里程基准保费, 再根据 WCR (天气)、LCR (光线)、RHR (假期/非假期)、SR (道路状况)、SR (速度) 调整保费。Tselentis (2016) 归纳量化驾驶行为风险的指标有总驾驶里程数、道路类型、危险驾驶时间、旅行频率、超速行驶里程、急加速、急转弯、驾驶过程中手机使用情况等。同时市场上主要的测度驾驶行为的设备 (CarChipFleetPro、Sky-meter、OnStar、Freematics、MyRate Device), 测度的主要指标是驾驶里程、驾驶速度、驾驶时间、经纬度、发动机转速等。对驾驶行为因素量化后, 需要选择合适的方法将其引入费率厘定模型中。引入费率厘定中的思路有直接引入与压缩变量引入。压缩引入有助于简化模型。压缩驾驶行为变量, 主要有以下几种方法: 构造驾驶行为风险指数, 将指数当做变量引入模型 (Toledo, 2008); 对驾驶行为指标构建逻辑回归模型, 将模型结果作为新的费率因子引入定价模型 (Paefgen, 2013); 编写算法或专家评分 (Kantor, Stárek, 2014; 朱爽, 2015); 主成分降维 (刘征宇, 2015)。指标与方法汇总表见表 1。

费率厘定模型方面。GLM 在实务中较为常用, 但 GLM 存在两个缺陷, 一是 GLM 假设变量观测值之间相互独立, 与保险数据不符; 二是对数据量的要求较高, 只有每个类别的经验数据充足时, 才能保证 GLM 结果的可靠性, 但当多水平费率因子存在时, 该标准很难满足。HGLM 可隐形地克服 GLM 的缺陷。Lee 和 Nelder (1996) 年提出分层广义线性模型 (HGLM), 并针对

性的提出 h-似然估计法。Lars Rönnegård, Xia Shen (2010) 编写的 hglm 软件包使 HGLM 在 R 软件中较为容易实现。国内将 HGLM 应用于保险领域的研究较少。唐清霞 (2016) 利用 HGLM 进行未决赔款准备金评估; 李常剑 (2016) 构建 HGLM, 但并未对 HGLM 拟合效果与 GLM 效果进行比较, 实践应用中是否具有优势有待商榷。

表 1 部分文献使用的方法与变量

作者	Litman	Iqbal 等	Boquete 等	Ferreira 等	Hultkranztz 等	Paefgen 等	Handel 等	Kantor 等	Ayub 等	陈星	王天梅
年份	2005 2008 2011	2006	2010	2010 2012	2012	2013 2014	2014	2014	2016	2016	2016
线性模型	√			√							
非线性模型		√	√	泊松回归	√	逻辑回归	算法	算法	Weibull	逻辑回归	泊松回归
驾驶里程	√		√	√		√		√		√	
夜间驾驶时间			√			√	√	√	√	√	√
早晚高峰时间		√				√	√	√		√	√
周末/非周末						√					
驾驶速度		√	√		√	√	√		√	√	√
急加速			√				√				√
急刹车							√				
急转弯							√				
道路类型		√	√			√	√	√	√		
能见度		√									
区域			√								
乘客数量			√								
天气		√									
经纬度							√				
使用手机			√								
能量耗损							√				
性别	√								√		
其他分类	√			√							

综上所述, 国内系统地研究量化驾驶行为因素及引入费率厘定模型的方法、HGLM 在保险领域的应用研究都较少。一方面, 笔者从量化驾驶行为因素指标出发, 并对引入方法进行

文献综述,选择驾驶行为因素的替代变量,为财险公司未来选择“驾驶行为因素”指标、实现基于“驾驶行为”车险费率厘定,提供理论参考;另一方面构建 HGLM 研究车险费率厘定的研究较少,笔者通过构建 GLM、HGLM、引入驾驶行为因素的 HGLM,并比较其拟合效果,为费改后车险费率厘定实践提供参考,57 个水平,宝骏是基准水平。

## 1 变量、数据与模型

### 1.1 变量选择与数据说明

本文采用数据是某地区某大型财险公司的 2011-2014 保单年的家庭自用车车辆损失保险数据,经整理有 547 970 条纪录可用。根据费改后现行家庭自用车主要费率因子是车辆种类、车型、车辆使用年限等。选择了 6 个费率因子(表 2),其中车型因子(car\_type)是多水平费率因子。

车辆实际价值。不同风险等级之间的索赔频率存在微小差异,索赔频率呈现逐步缓慢上升的趋势;案均赔款之间的差异较大,随着车辆实际价值的风险等级的提高,案均赔款的差异愈加明显,相邻两个风险等级的案均赔款比分别是 1.50、1.40、1.77、2.22。

车辆使用年限。索赔频率是车辆使用年限为 0-1 年最高;风险等级 1-2 年的案均赔款最高,最低的是 5 年以上的车辆。据索赔频率、案均赔款分布分析,1-2 年、2-5 年风险保费低于其他两个风险等级的主要原因是较低的索赔频率。

车型因子,对费率进行进一步调整。车型因子的索赔频率、案均赔款、纯保费范围分别是 0.33~0.85、680~9380 元、391~7628 元,即不同车型因子之间索赔频率、索赔强度、纯保费的差异较为明显。客户忠诚度。

承保的风险单位数在续保、新保、转保中分布较为均匀;索赔频率、案均赔款、纯保费三角度,新保客户都高于其他两个风险等级。

无赔款优待及上年赔款记录(NCD),主要是索赔频率上的差异,是费率厘定过程中需要重点考虑的因子。上年发生 4 次及以上赔款(NCD\_7)的索赔频率低于上年发生 3 次赔款(NCD\_6)的客户,出现这种反常情况原因是,被保险人存在隐瞒小额损失,而只报大额损失的现象,这一点反应在其案均赔款远高于其他类别无赔款优待记录投保个体上,该级别的案均赔款分别是 NCD\_1~NCD\_6 水平 1.459、1.413、1.382、1.384、1.558、1.515 倍。该行为背后的深层原因是行业的无赔款优待及上年赔款记录调整系数对各风险等级的调整系数设置不合理,过度地惩罚了上年出险次数多的客户。邸娜(2016)也指出行业无赔款优待系数的转移规则不合理、不公平,在市场化竞争的环境中容易引发逆向选择,相对保费存在对低风险客户的过度奖励和对高风险客户惩罚等。因此,将 NCD 作为费率因子直接引入模型中,而不是已知的调整系数。

年龄。根据原有保单库的年龄分类,年龄在 27 岁以下的客户,无论是索赔频率还是案均赔款,其风险程度都较高,27-33 岁、33-55 岁的客户,风险等级较为接近。

表 2 费率因子的经验数据表

分类变量	变量水平及含义	风险单位%	索赔频率	案均赔款	纯保费
车辆实际价值 (car)	A5 万以下	15.36	0.62	847	523
	B5—10 万	46.14	0.65	1272	821
	C10—20 万	27.25	0.66	1775	1172
	D20—30 万	7.99	0.69	3136	2161

	E30 万+	3.26	0.72	6966	5029
车辆使用年限 (carage)	A0-1 年	24.99	0.70	1758	1229
	B1-2 年	25.96	0.61	1767	1083
	C2-5 年	41.61	0.64	1674	1070
	D5 年以上	7.44	0.69	1591	1092
客户忠诚度 (renew)	A 续保	42.40	0.61	1650	1001
	B 新保	31.13	0.72	1845	1334
	C 转保	26.47	0.64	1632	1041
无赔款优待 及上年赔款 记录 (NCD)	NCD_1 连续 3 年 没有发生赔款	5.64	0.34	1635	548
	NCD_2 连续 2 年 没有发生赔款	9.12	0.43	1689	721
	NCD_3 上年没有发生赔款	22.43	0.52	1726	898
	NCD_4 首年投保或新车 或上年发生一次赔款	48.25	0.70	1724	1212
	NCD_5 上年发生 2 次赔款	9.35	0.91	1531	1387
	NCD_6 上年发生 3 次赔款	2.66	1.05	1574	1650
	NCD_7 上年发生 4 次及以上赔款	2.55	0.96	2386	2298
年龄 (age)	A27-	14.43	0.74	1882	1397
	B27-33	20.88	0.68	1684	1149
	C33-55	60.59	0.62	1687	1046
	D55+	4.11	0.62	1552	969

## 1.2 构建模型

模型 1: 车型因子为固定效应的 GLM:

$$\begin{aligned}
 E(Y_{ijklmno}) &= \mu_{ijklmno} \quad \mu_{ijklmno} = g^{-1}(\eta_{ijklmno}) \\
 \eta_{ijklmno} &= \beta_0 + car_j \times \beta_{1j} + carage_k \times \beta_{2k} + renew_l \times \beta_{3l} \\
 &\quad + NCD_m \times \beta_{4m} + age_n \times \beta_{5n} + car\_type_o \times \beta_{6o}
 \end{aligned} \quad (1)$$

其中, car、carage、renew、NCD、age、car\_type 构成固定效应信息矩阵 X。假设索赔频率与案均赔款分别服从泊松分布、伽玛分布, 连接函数采用对数连接。

模型 2: 车型因子为随机效应的 HGLM:

$$\begin{aligned}
 E(Y_{ijlmno}|u_o) &= \mu_{ijlmno} \quad \mu_{ijklmno} = g^{-1}(\eta_{ijklmno}) \\
 \eta_{ijklmno} &= car_j \times \beta_{1j} + carage_k \times \beta_{2k} + renew_l \times \beta_{3l}
 \end{aligned}$$

$$+NCD_m \times \beta_{4m} + age_n \times \beta_{5n} + Zv(u_o)$$

$$v \sim f_v(v) \quad (2)$$

与 (1) 的区别是, car\_type 构成随机效应设计矩阵  $Z$ ;  $u_o$  是模型的随机效应部分, 一般通过某种单调转化, 转化成  $v(u_o)$ , 从而完成模型的参数估计过程。

模型 3: 引入驾驶行为因素的 HGLM。与 (2) 结构一样, 区别是随机效应设计矩阵  $Z$  由车型因子、驾驶行为因素构成, 即存在多个随机效应。

## 2. 实证结果与分析

### 2.1 车型因子作为固定效应实证分析

模型 1 参数估计结果见表 3 中。索赔频率, 前 5 个普通费率因子的索赔频率估计值, 与单因子分析结果比较一致, 且均具有统计显著意义。在案均赔款中转保、连续 2 年没有发生赔款统计意义并不显著。车型因子无论是索赔频率还是案均赔款模型, 大多数不显著, 同时车型系数之间的差距较大。说明可能由于车型分类较多, 导致部分分类经验数据不足, GLM 参数估计误差较大。

表 3 GLM 索赔频率与案均赔款估计值

变量	索赔频率	案均赔款	变量	索赔频率	案均赔款
截距项	-1.211***	6.910***	NCDNCD_2	0.248***	-0.041
carB	0.061***	0.264***	NCDNCD_3	0.490***	-0.080*
carC	0.108***	0.588***	NCDNCD_4	0.746***	-0.125**
carD	0.137***	1.132***	NCDNCD_5	1.009***	-0.143***
carE	0.140***	1.797***	NCDNCD_6	1.149***	-0.134***
carageB1	-0.012	0.181***	NCDNCD_7	0.992***	-0.047
carageC2-5	0.120***	0.149***	ageB27-33	-0.062***	-0.134***
carageD5+	0.210***	-0.012	ageC33-55	-0.138***	-0.222***
renewB 新保	0.094***	0.240***	ageD55+	-0.129***	-0.325***
renewC 转保	0.053***	0.005	车型	宝骏是基准水平	

注: \*\*\*表示 p 值小于 0.001, \*\*表示 p 值小于 0.01, \*表示 p 值小于 0.10; 估计结果中, 每个分类的变量的第一个水平当成基准水平。

通过实际数据检验模型拟合效果。主要操作过程是, 利用参数估计值, 得到纯保费的估计值; 计算预期赔付率; 计算预期赔付率的 5%、10%、……100%分位数, 将原始数据分为 20 组 (表 4)。模型偏差是  $2.336 \times 10^{14}$ , 模型偏差较大其原因是, 低估左尾处的风险和高估右尾处的风险。即在多水平费率因子存在的情况下, 用 GLM 拟合的损失是较不稳定的。

表 4 GLM 结果拟合值与实际值对比表

组别	预期赔付率	实际赔付率	组别	预期赔付率	实际赔付率
1 组	0.283	0.338	11 组	0.619	0.655
2 组	0.357	0.406	12 组	0.650	0.640
3 组	0.396	0.438	13 组	0.684	0.631
4 组	0.428	0.480	14 组	0.721	0.691
5 组	0.456	0.490	15 组	0.763	0.682
6 组	0.484	0.517	16 组	0.814	0.727

7 组	0.510	0.498	17 组	0.876	0.736
8 组	0.536	0.530	18 组	0.959	0.814
9 组	0.562	0.533	19 组	1.087	0.938
10 组	0.590	0.562	20 组	1.505	1.159

## 2.2 车型因子作为随机效应的实证分析

随机效应函数选择伽玛分布、对数连接时，较正态分布、恒等连接时，总偏差略降低。故选择伽玛分布、对数连接，并对其进行参数估计。参数估计结果见表 5。无论是 GLM 还是 HGLM，案均赔款的 NCD\_7 的不具有统计显著性，与单因素分析过程发现 NCD\_7 类别下，案均赔款明显高于其他类别的情形明显不符。为与 GLM 对比，将宝骏调整为基准，系数调整为 1。索赔频率调整系数范围在 0.684~1.215 间波动，波动范围略小于 GLM 估计结果 0.581~1.274，案均赔款调整系数范围为 0.795~2.06，波动范围略小于 GLM 估计结果。对多水平费率因子的参数估计值相对较为稳定。

表 5 HGLM 索赔频率与案均赔款固定效应参数估计值

变量	索赔频率	案均赔款	变量	索赔频率	案均赔款
截距	-1.168***	6.984***	NCDNCD_2	0.248***	-0.032
carB	0.060***	0.274***	NCDNCD_3	0.491***	-0.067**
carC	0.106***	0.588***	NCDNCD_4	0.747***	-0.104***
carD	0.135***	1.131***	NCDNCD_5	1.010***	-0.136***
carE	0.144***	1.780***	NCDNCD_6	1.150***	-0.119***
carageB1	-0.012	0.155***	NCDNCD_7	0.993***	-0.019
carageC2-5	0.121***	0.118***	ageB27-33	-0.062***	-0.129***
carageD5+	0.211***	-0.050**	ageC33-55	-0.138***	-0.216***
renewB 新保	0.094***	0.192***	ageD55+	-0.129***	-0.309***
renewC 转保	0.053***	0.012			

根据预期赔付率分组，分组后损失拟合值、实际损失、预期赔付率、实际赔付率见表 6。车型因子作为随机效应之后，总偏差减少至  $1.236 \times 10^{14}$ ，仅为 GLM 偏差的 52.92%，偏差明显减少。说明与 GLM 相比，HGLM 显著减少估计误差，拟合值与实际值更为接近。即将多水平费率因子作为随机效应处理有助于提高模型拟合程度。

表 6 HGLM 结果拟合值与实际值对比表

组别	预期赔付率	实际赔付率	组别	预期赔付率	实际赔付率
1 组	0.272	0.338	11 组	0.589	0.663
2 组	0.341	0.403	12 组	0.619	0.646
3 组	0.378	0.454	13 组	0.650	0.659
4 组	0.408	0.452	14 组	0.685	0.669
5 组	0.435	0.501	15 组	0.726	0.688
6 组	0.461	0.520	16 组	0.772	0.714
7 组	0.485	0.497	17 组	0.830	0.749
8 组	0.510	0.525	18 组	0.907	0.798
9 组	0.536	0.544	19 组	1.025	0.946
10 组	0.562	0.552	20 组	1.399	1.178

### 2.3 驾驶行为因素实证分析

积极寻找驾驶行为因素的替代变量,对包含从车因素、从人因素、驾驶行为因素的 HGLM 进行实证分析。

#### 2.3.1 驾驶行为因素指标之间的关系

从驾驶里程、驾驶速度、驾驶时间、道路类型四方面入手,对驾驶行为因素变量进行量化,利用 1407 份驾驶行为样本数据,检验变量之间的相关性(部分相关系数见表 7)。发现夜间驾驶时间与夜间驾驶次数,除了速度 100-120 里程数与速度 120+里程数等与夜间驾驶时间相关性不强以外,与其他变量有较强的相关性。主成分分析过程中,根据特征值大于 1 的原则,选择 6 个主成分,其中前 6 个主成分分别提取总信息的 0.3725%、15.54%、10.73%、5.90%、5.24%、4.46%,累积提取全部变量的 79.12%的信息。根据 Hair 等人(1998)的观点,在社会科学相关领域中,由于其精确度不如自然科学,因而其萃取的共同因素累积解释变量能达到 50%以上,因素分析结果可被接受,60%以上表示共同因素是可靠的。其中在夜间驾驶次数在主成分 1、3、4 上具有较大的权重,分别为 0.279、-0.329、0.204,夜间驾驶时间在主成分 1 上具有较大的系数,系数为 0.289。

表 7 夜间驾驶次数、夜间驾驶时间与其他驾驶行为变量之间的相关关系

	夜间驾 驶次数	夜间驾 驶时间		夜间驾 驶次数	夜间驾 驶时间
总里程	0.384***	0.091***	开车超过 6 小时数	0.899***	0.465***
速度 0-40 里程	0.793***	0.659***	早高峰驾驶次数	0.733***	0.425***
速度 40-80 里程	0.828***	0.416***	早高峰驾驶时间	0.740***	0.275***
速度 80-100 里程	-0.421***	-0.026	晚高峰驾驶次数	0.786***	0.407***
速度 100-120 里程	-0.056***	-0.003	晚高峰驾驶时间	0.843***	0.383***
速度 120+里程	-0.063*	-0.028	急刹车次数	0.174***	0.141***
空挡滑行里程	0.559***	0.194***	长下坡次数	0.125***	0.084**
			夜间不开双闪次 数		
总驾驶时间	0.645***	0.417***		0.198***	0.113***

注:\*\*\*表示 Pearson 相关性检验的 p 值小于 0.001; \*表示 p 值小于 0.01; \*\*表示 p 值小于 0.1。

结合相关分析与主成分分析结果,夜间驾驶时间与夜间驾驶次数是反应驾驶行为因素的重要变量。故利用夜间驾驶时间与夜间驾驶次数建立驾驶行为数据与现有保单数据库的联系。

#### 2.3.2 驾驶行为因素替代变量分析

有夜间出险记录的被保险人,前提条件是存在夜间驾驶行为,存在夜间驾驶行为被保险人,具有更高的风险程度即更高的索赔频率与案均赔款。根据保单库中的数据分析,有夜间出险记录的索赔频率、案均赔款分别为 1.92、4784 元,而无夜间出险记录的索赔频率、案均赔款分别为 1.43、1660.95 元。

搜集某省各市 2011—2015 年公路里程数与交通事故情况,借鉴张连增研究 2006—2012 年的省公路里程数与事故发生率、案均赔款的关系的思路。检验发现 2011—2015 年各市的平均公路里程数与平均每起事故损失之间的相关系数 0.423, p 值为 0.1 左右,具有较强的正向相关性,即公路里程多的地区,平均每起损失更高,即公路里程数在一定程度上能反应不同地区的人的风险程度。出于数据保密性,按照各市的公路里程数从小到大的顺序将某省的 16

市编码为 A1-A16。分析 A1-A16 市有夜间出险记录赔付次数占该市赔付次数的比,发现在地区之间差异较大。其中有夜间总索赔次数最多的地区是 A11,是最少地区的 A8 地区的 6.29 倍;占比最高的地区是 A14,是占比最低地区 A2 的 5.79 倍;且有前 8 个地区夜间出险情况普遍低于后 8 个地区的情况(见表 8)。

表 8 各地区夜间出险记录与总出险记录比(%)

地区	夜间出险/总出险	地区	夜间出险/总出险
A1	1.28	A9	1.71
A2	0.69	A10	1.86
A3	1.87	A11	1.03
A4	1.93	A12	2.58
A5	1.77	A13	1.55
A6	1.21	A14	4.02
A7	1.70	A15	1.66
A8	1.14	A16	2.24

各地区夜间出险占比之间存在差异,说明部分地区驾驶员具有更强的夜间驾驶倾向。结合夜间驾驶与驾驶里程、驾驶速度之间的相关关系分析结果,发现大多是正向关系,因此部分地区的驾驶员具有更高的驾驶行为风险。同时,在量化驾驶行为因素过程中,除了驾驶里程、驾驶速度、驾驶时间外,常将道路类型、天气等纳入量化的指标体系,且各市之间的道路类型、公路里程之间存在差异。高速公路驾驶风险最低(AA Trust Report, 2006),高速公路占比越高的地区,其风险相对较低。分析各市 2011-2015 高速公路占公路里程的比发现,前 8 个地区高速公路占比普遍高于全省平均水平,而后 8 个地区低于全省平均水平,以 2015 年为例,占比最高 A1 是占比最低 A9 地区的 2.64 倍。因此,综合驾驶行为因素的四类量化指标,探索性地将经过排序的地区变量作为驾驶行为的替代变量,引入车险费率厘定模型,对于商业车险费率的车险定价进行补充。

### 2.3.3 引入驾驶行为因素的 HGLM 的实证分析

模型构建与公式(2)一样,区别是随机效应设计矩阵<sup>Z</sup>由车型因子、驾驶行为因素的替代变量地区因子 code 构成,即存在两个随机效应。固定参数估计结果见表 9。

表 9 HGLM 索赔频率与案均赔款固定效应参数估计值

变量	索赔频率	案均赔款	变量	索赔频率	案均赔款
(Intercept)	-1.064***	6.841***	NCDNCD_2	0.212***	-0.046
carB	0.060***	0.261***	NCDNCD_3	0.422***	-0.076***
carC	0.095***	0.586***	NCDNCD_4	0.641***	-0.098***
carD	0.119***	1.148***	NCDNCD_5	0.881***	-0.128***
carE	0.128***	1.821***	NCDNCD_6	1.020***	-0.180***
carageB1	-0.014	0.180***	NCDNCD_7	0.907***	-0.137***
carageC2-5	0.103***	0.160***	ageB27-33	-0.085***	-0.100***
carageD5+	0.179***	0.010	ageC33-55	-0.164***	-0.168***
renewB 新保	0.113***	0.187***	ageD55+	-0.182***	-0.251***
renewC 转保	0.036***	0.023*			

引入驾驶行为因素替代变量的 HGLM,参数估计结果更为合理,以 NCD 估计结果为例,

具有统计显著意义的是 NCD<sub>7</sub>，与实际情况较为相符。

地区因子索赔频率与案均赔款的参数结果见表 4.10。在 HGLM 中，不给出 p 值，只给出估计标准误，估计标准误大多数在 0.04 左右，估计误差较小，说明结果还是值得信赖的。地区的随机效应参数估计值，在纯保费一栏中，前 8 个地区随机效应调整系数小于 1，后 8 个地区随机效应调整系数大于 1 的规律，即与前 8 个地区夜间出险占比普遍低于后 8 个地区的现象相对应。特别地，A1 地区风险最低，调整系数为 0.817，A14 地区风险最高，调整系数为 1.336。地区因子估计值的之间差异，说明需要类似于地区因子的驾驶行为因素对费率进行进一步调整。

表 10 HGLM 索赔频率与案均赔款随机效应参数估计值

地区	索赔频率	案均赔款	纯保费	地区	索赔频率	案均赔款	纯保费
codeA1	0.787	1.074	0.846	codeA9	1.129	0.901	1.018
codeA2	1.133	0.760	0.861	codeA10	1.161	0.825	0.958
codeA3	0.711	1.263	0.899	codeA11	1.254	0.787	0.986
codeA4	0.991	0.895	0.886	codeA12	0.975	1.045	1.018
codeA5	0.895	1.127	1.009	codeA13	1.056	0.985	1.040
codeA6	1.153	0.738	0.851	codeA14	0.941	1.405	1.322
codeA7	1.179	0.745	0.878	codeA15	1.123	0.793	0.890
codeA8	0.756	1.167	0.882	codeA16	0.757	1.491	1.129

地区因子、车型因子作为随机效应之后，地区因子、车型因子作为随机效应之后，总偏差减少至  $1.059 \times 10^{14}$ ，是不加地区因子作为随机效应的 HGLM 偏差的 85.67%，仅为最初 GLM 偏差的 45.33%，加入地区因子，使偏差进一步减少。即加入地区因子的 HGLM 提高了总体拟合效果，更接近实际值。对比表 6 与表 11，分段来看两个模型都有在低赔付率低估风险的趋势，但加入地区因子的 HGLM 在 7-16 组具有拟合优势，拟合值几乎与实际值之间的差异比较小。

表 11 加入地区因子 HGLM 结果拟合值与实际值对比表

组别	预期赔付率	实际赔付率	组别	预期赔付率	实际赔付率
1 组	0.267	0.335	11 组	0.587	0.601
2 组	0.337	0.390	12 组	0.617	0.690
3 组	0.373	0.419	13 组	0.650	0.644
4 组	0.403	0.492	14 组	0.686	0.676
5 组	0.431	0.443	15 组	0.727	0.714
6 组	0.457	0.522	16 组	0.774	0.741
7 组	0.482	0.493	17 组	0.833	0.755
8 组	0.507	0.507	18 组	0.911	0.855
9 组	0.533	0.535	19 组	1.031	0.922
10 组	0.559	0.561	20 组	1.411	1.216

### 3 结语

结合实证分析结果，与目前国内普遍应用的车险费率厘定因子与费率厘定方法，主要提出以下建议：

一是，改变费率因子的分级，充分挖掘风险信息。以机动车中占比最大的家庭自用车

为例，目前主要是从车型、车辆使用年限、车辆座位数位分类变量维度进行分级。一方面，应该加入从人因素中的年龄、驾龄等维度。实证结果中，三个模型驾驶人的年龄因素对索赔频率与案均赔款均具有显著影响，在 GLM 中，TYPE3 检验，显示年龄因素应该引入费率模型中，在 HGLM 中，年龄因素在索赔频率中的统计显著性 p 值远小于 0.0001。另一方面，对现有从车因素分级进行进一步细分或者寻找相应的替代变量。如车辆使用年限行业分类是 1 年以下、1-2 年、2-6 年、6 年以上，根据现实数据分析，家庭自用车，车辆使用年限 92.6% 集中在 5 年及以下，建议对家庭自用车的车辆使用年限进一步细化，以充分体现不同车辆使用年限之间的风险。对分级制度和内部分级的改变，有利于充分挖掘现有变量的信息，识别风险，从而建立更加公平的费率体系，实现更公平、更自主、更合理的定价机制。

二是，加强驾驶行为的数据搜集，车联网数据与保险数据的对接，为实现 UBI 定价奠定数据基础。公路里程数更多的地区，其夜间出险比例也相应较高，索赔频率、案均赔款也相应更高即具有更高的风险等级。同时，结合国内外已有 UBI 研究，驾驶行为因素是影响赔付支出的重要因素。

三是，加强技术培训，提高人才素质。目前，驾驶行为数据未能引进费率厘定模型主要原因是，驾驶行为数据较难获取，无法实现驾驶行为数据与保险公司已有数据的对接，针对此，建议财险公司加强技术人员培训、提高其风险识别能力，根据赔案库中的出险原因、出险时间、出险地区分布，对被保险人的习惯进行一定的识别。驾驶行为数据收集后，如何对其进行数据整理、引入定价模型，对财险公司定价人员提出了更高的要求。

四是，多种定价方法比较，保证其定价的合理性。保险公司在费率厘定过程中，在使用传统模型的同时，应勇于创新，逐步探索更适合公司经营的费率厘定模型。

创新与不足。HGLM 应用于车险费率厘定中，可以解决多水平费率因子参数估计结果过于复杂、估计误差较大等问题；探索性将地区因子作为驾驶行为因素的替代变量，引入费率厘定过程中，具有一定的创新性。利用地区因素对驾驶行为因素进行替代，可提高模型拟合效果，但替代程度无法量化，且低估了低赔付率处的风险，可能存在更好的替代变量。

#### 参考文献

[1] Greaves S P, Fifer S, Ellison R. Exploring Behavioral Responses of Motorists to Risk-Based Charging Mechanisms[J].

[1] Greaves S P, Fifer S, Ellison R. Exploring Behavioral Responses of Motorists to Risk-Based Charging Mechanisms[J].

[2] Transportation Research Record Journal of the Transportation Research Board, 2013, 2386(2386):52-61. Tselentis D I, Yannis G, Vlahogianni E I. Innovative motor insurance schemes: A review of current practices and emerging challenges[J]. Accident; analysis and prevention, 2016, 98:139.

[3] Ferreira Jr J, Minikel E. Measuring per mile risk for pay-as-you-drive automobile insurance[J]. Transportation Research Record: Journal of the Transportation Research Board, 2012 (2297): 97-103.

[3] Boucher J P, Marín A M P, Santolino M. Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident[J]. Anales Del Instituto De Actuarios Españoles, 2013:135-154.

[5] Paefgen J, Staake T, Thiesse F. Evaluation and aggregation of pay-as-you-drive insurance rate

factors: A classification analysis approach[J]. Decision Support Systems, 2013, 56: 192-201.

[6] Paefgen J, Staake T, Fleisch E. Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data[J]. Transportation Research Part A: Policy and Practice, 2014, 61: 27-40.

[7] Kaplan S, Guvensan M A, Yavuz A G, et al. Driver Behavior Analysis for Safe Driving: A Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(6):1-16.

[8] Charity Mkajuma Wamwea, Benjamin Kyalo Muema, Joseph Kyalo Mung'atu. Modelling a Pay-As-You-Drive Insurance Pricing Structure Using a Generalized Linear Model: Case Study of a Company in Kiambu. American Journal of Theoretical and Applied Statistics. 2015,4(6): 527-533

[9] Kantor S, Stárek T. Design of algorithms for payment telematics systems evaluating driver's driving style[J]. Transactions on Transport Sciences, 2014, 7(1):9-16.

ETSC, 2010. Road Safety Target in Sight: Making up for Lost Time. 4th Road Safety PIN report.

[10] 刘征宇,夏伟,刘宁波,张利. 基于主成分分析法的车险定价因子研究[J]. 合肥工业大学学报(自然科学版),20 15,(03):419-423.

[11] 王天梅,胡伊,肖永慧. 关于我国车险费率厘定方法创新研究——基于驾驶行为的车险费率厘定实证分析[J]. 价格理论与实践,2016,(11):122-125.

## Research on Automobile Insurance Ratemaking Based on Driver Behavior Perspective Review and Empirical Analysis

ZHANG Lin , XIE Yafeng

(Finance and Statistics Hunan University, Changsha, 410079)

**Abstract:** This article embarks from the driver behaviors of four quantitative indicators, using correlation analysis and principal component analysis method to find important variables in behavior factor of driver and connect them with data of policies, and looking for alternative variable and introducing into automobile insurance ratemaking. At the same time, The HGLM is built to overcome the Lack of credibility due to the vehicle model and driver behavior. Result of study reflects HGLM is better than GLM and if we add replacement variables of driver behavior o in this model, the fitting effect of model will increase.

**Keywords:** Ratemaking; Driver Behavior; GLM; HGLM

### 作者简介(可选):

张琳(1963.3至今),女,湖南长沙人,教授,管理学博士,湖南大学金融与统计学院风险管理与保险学系主任,中国保险学会理事、中国精算师协会正会员,研究方向:保险精算。湖南省长沙市岳麓区石佳冲109号;邮编:410079 手机:18684716118 邮箱:lindazhang0203@126.com

谢亚凤(1992.11至今),女,湖南永州人,湖南大学金融与统计学院硕士研究生,研究方向:保险精算。手机:18074681576 邮箱:1466574570@qq.com

