·管理科学与工程·

基于驾驶行为的UBI车险定价模型

□吴祥佑
[闽江学院　福州　435108]

[摘　要]　【目的/意义】随着5G时代的到来，更多地基于驾驶行为而非行驶里程定价将成为车险定价的关键。合理的UBI车险定价不仅有助于精算公平，更有利于UBI车险的普及推广。【设计/方法】基于UBI被保险人驾驶行为数据，构建了索赔次数影响因素的有序分类 logistic模型。【结论/发现】实证结果表明，高速行驶、急加速、违章次数和行驶里程数对出险索赔有显著的正向影响；急刹车和急转弯对出险索赔次数存在不显著的影响，其中急刹车将降低而急转弯则将提高出险次数。基于logistic模型所预测的被保险人在各索赔水平上的概率分布，可获得消除了运气扰动的期望索赔次数，以具体被保险人的期望索赔次数与全体被保险人平均索赔次数的比值为权重，可对车险实施较现行NCD定价更优的基于驾驶行为的定价。

[关键词]　UBI车险；驾驶行为；索赔次数；超常索赔比；定价
[中图分类号] F842　　[文献标识码] A　　[DOI] 10.14071/j.1008-8105(2019)-4014

UBI Auto Insurance Pricing Model Based on Driving Behavior

WU Xiang-you
(Minjiang University　Fuzhou　435108　China)

Abstract　[Purpose/Significance] With the advent of the 5G era, pricing based on driving behavior rather than mileage will become the key to auto insurance pricing. Reasonable UBI auto insurance pricing not only helps actuarial fairness, but also facilitates the popularization of UBI auto insurance. [Design/Methodology] Based on the driving behavior data of the UBI insured, an ordered logistic model of the influencing factors of the number of claims is constructed. [Findings/Conclusions] The empirical results show that high-speed driving, rapid acceleration, number of violations and mileage have a significant positive impact on insurance claims. Sudden braking and sharp turns have an insignificant effect on the number of insurance claims. Among them, sharp braking will decrease and sharp turns will increase the number of risks. Based on the probability distribution of the insured at each claim level predicted by the logistic model, the expected number of claims without luck disturbances can be obtained. Taking the ratio of the expected number of claims of the specific insured and the average number of claims of all the insured as the weight, it is possible to implement driving behavior-based pricing that is better than current NCD pricing for auto insurance.

Key words　UBI auto insurance; driving behavior; claim frequencies; over-level claim ratio; rate making

引言

随着经济的快速发展，我国已成为全球汽车增长速度最快的国家[1]。如图1所示，截至2018年底，我国汽车保有量达2.4亿辆，比2017年净增2 285万辆，增长了10.51%。虽然新车销售增速较上年下降了2.8%，但未来发展空间依然巨大。

[收稿日期]　2019－08－30
[基金项目]　福建省自然科学基金项目(2017J01799)；闽江学院课题（MYK17030、YS17035、MYK18041、MYS18048、MYS19014）；海西金融与财政发展研究中心（HXCJ1907）；闽江学院虚拟仿真实验教学项目（MJUS2019B004）.
[作者简介]　吴祥佑（1971－）男，博士，闽江学院新华都商学院副教授.

伴随着汽车保有量的增长，汽车保险稳步成为我国第一大财险险种。1988年我国车险保费首次超过企财险，以20亿元的规模、占财险38%的比重开启了高速增长的新纪元[2]。如图2所示，进入新世纪后，车险第一大财险的地位更趋稳固，保费由2000年的400亿元激增到2018年的8 183.38亿元，在财险中的占比接近8成。

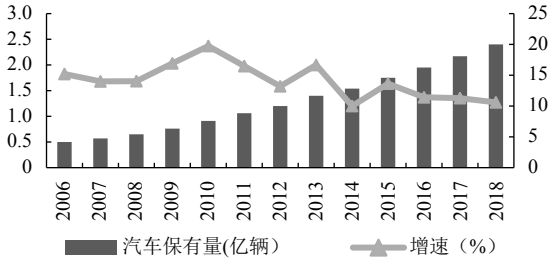

图 1 2006~2018年我国汽车保有量及增速

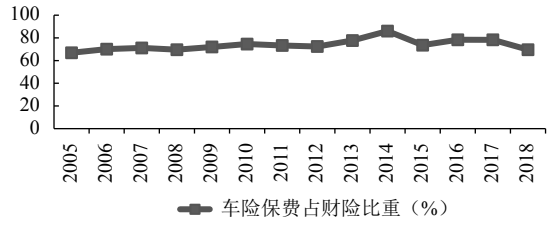

图 2 2005~2018车险保费占财险保费比重

汽车市场的成长和大数据时代的耦合催生了车联网的繁荣发展。车联网（Internet of Vehicles）以车内网、车际网和车载移动互联网为基础，实现了车与车、路、行人及互联网之间的无线通讯和信息交换，构建了智能化交通管理、智能动态信息服务和车辆智能化控制一体化的大系统网络[3]，使车辆行驶状态及其周边环境信息的采集、数据的传输与处理变得经济可行。车联网的发展为数字车险的发展提供了急需的数据，为车险市场效率低下、欺诈盛行、保费计算困难等老问题的解决提供了新的可能。由图2可见，我国车险保费日渐触顶，增速放缓已成常态。作为产险业最大的利润来源，2018年车险虽然实现了10.53亿元承保利润，但同比大幅减少63.36亿元，降幅达到85.75%，承保利润率仅0.14%，已逼近盈亏关口[4]。2018年车险对产险业的利润贡献不及保证保险和农业保险，已退居第三位，车险市场的发展亟需大数据技术的推广与应用。

车联网技术的兴起提高了保险业大数据的可得性，使其具备了对定价数据搜集、整理及分析的有效手段，催生了基于驾驶行为定价的新型车险UBI（Usage Based Insurance），有助于驾驶行为较好的被保险人获得适当的保费优惠。长期以来，由于我国保险公司对国家统一制定的条款费率的依赖过大，造成车险费率自由化改革后，行业缺乏车险定价所需的数据统计体系，造成定价风险过大、投保核保混乱等问题[5]。为更好地解决车险现存的问题，适应车险发展的趋势，保险公司必须改进现有车险的定价模式。在车联网及大数据兴起的时代背景下，以对UBI车险定价模式的改进为核心的车险市场化改革已成为必然趋势。

一、文献回顾

随着车联网的发展，基于驾驶行为定价势将成为车险的主流定价方法。传统车险以“人车”基本状况等先验因子为基准进行定价，高度依赖历史信息与既有的静态变量，不能将适时信息等动态因素纳入模型，会使部分“人车”状况并不理想，但谨慎驾驶者承担过高的费率，导致费率厘定的不公平。

学界很早就开始关注驾驶行为对车险定价的作用。Vickrey主张通过征收汽油税或根据平均轮胎寿命进行定价，以解决车险费率不能促使驾驶者改变驾驶习惯和减少驾驶的问题[6]。Butler认为行驶里程是车辆使用情况最重要的测度，对行驶里程不同的车辆不应适用相同的费率[7]。Litman指出，UBI保险根据行驶里程等驾驶行为定价，除更精准外，还有助于提高安全驾驶意识，降低出险率[8]。Litman进一步强调，在现有不同种类的里程保险中，基于车联网的里程保险更能反映车险的风险成本，更值得推行[9]。还有学者指出，现行的无赔款优待（NCD）完全取决于偶发性的索赔频次，忽视了驾驶行为改进的影响，也有悖精算公平[10]。

王和强调，信息技术一直是保险业存在的前提和支撑，更是其发展的引擎和决定力量，不断推动着保险科技的产业化[11]。随着车联网技术向车险的渗透，车险进入了移动互联网商业模式时代。在车联网和大数据技术的推动下，郁佳敏认为UBI车险将快速成长[12]。UBI车险将为财产保险业开辟新的成长空间，也将成为车联网商业模式成功的关键[13]。在车险费率市场化改革的大背景下，UBI车险使差异化、精细化、个性化的车险服务成为现实[14]。借助于大数据技术，UBI车险将显著改善车险市场效率，助推有车联网定价优势的保险公司胜出，提高车险定价的公平性[15]。

随着大数据技术的发展，越来越多的学者探索用数据挖掘技术，如随机森林、神经网络、支持向量机等机器学习算法为车险定价[16~18]。随着车联网

技术的发展，保险公司能够采集到被保险车辆较为详细的驾驶数据，这为促进汽车保险定价的合理性和公平性提供了可能[19~20]。行驶里程数是最重要的费率因子，但行驶里程数仅突显了车辆使用的风险暴露性，忽视了“三急”（急加速、急减速和急转弯）等车辆使用方式对风险的放大性[21~22]。对如何基于驾驶行为信息对车险定价，学界仍存在分歧。Weinder使用傅立叶分解法区分了不同的速度–加速度模式[23]；Wüthrich用聚类分析法对速度–加速度模式进行分类[24]；Gao & Wüthrich用主成分法和瓶颈神经网络提取速度–加速度模式[25]。高光远和孟生旺建立了预测索赔频率的泊松广义可加模型，发现驾驶行为因子是车险定价的重要费率因子，对索赔频率具有显著的非线性影响[26]。

在车联网、大数据技术日益成熟的背景下，车险应结合驾驶行为信息定价而非仅依赖行驶里程数定价。关于UBI车险定价，现有文献多试图从高频车联网数据中提取风险因子，很少从实际索赔数据中提取驾驶行为因子。对于UBI车险业务，我国保险实务界仍处于探索阶段，目前市场上尚无成熟的产品。学界虽然对此表现出了深厚的兴趣，但主要集中在理论探讨，仅提出了一些简单的定价模型。本文试图基于车联网大数据，从驾驶行为的视角构建车险索赔频率模型，并基于驾驶行为所导致的索赔频率在各水平间的概率分布对UBI车险实施动态定价，既贯彻NCD的思想，也间接影响、引导驾驶者改变驾驶行为，为车险的持续稳定发展提供支持。

二、数据与模型

（一）数据

数据由驾驶者基本信息和驾驶行为信息两部分构成，其中驾驶者基本信息来自某财产保险公司，主要包括性别、年龄、历史出险次数及车型、车龄等个人信息；驾驶行为信息由某车联网信息技术公司的车载OBD采集，主要内容包括行驶里程数、出行时间、超速行为、“三急”行为以及违章行为等[27]。

驾驶行为数据共包含10个特征变量和1个分类变量。10个特征变量分别是月总行驶里程、早晚高峰驾驶时间占比、夜间行车时间占比、周末行车时间占比、80~120 km/h行车时间占比、高于120 km/h行车时间占比、急加速次数、急减速次数、急转弯次数和违章次数。分类变量为被保险人的年出险索赔次数，取0~4共5种水平。样本规模为400个有效观测。表1给出了各变量的描述性统计。

表 1　各变量的描述性统计

变量	Obs	Mean	Std. Dev.	Min	Max	skew	curt	Jarque–Bera
category	400	1.347	1.466	0.000	4.000	0.577	1.884	8.048
over120ratio	400	1.724	1.801	0.000	6.580	1.224	3.503	19.507
suddenacceleration	400	8.200	6.036	0.000	29.000	0.747	3.322	7.295
suddenbrake	400	19.360	9.883	4.000	42.000	0.072	2.137	2.394
suddensteering	400	20.293	14.230	0.000	54.000	0.334	1.881	5.307
breaches	400	0.987	0.878	0.000	3.000	0.630	2.739	5.169
mileage	400	1264	562.35	217	2578	0.401	2.565	2.603
rushhoursratio	400	9.804	4.428	1.600	21.600	0.276	2.760	1.135
nighthoursratio	400	4.487	2.949	0.300	15.100	1.178	4.837	27.886
weekendhoursratio	400	12.344	4.314	3.000	20.500	–0.205	2.347	1.861
eighty120ratio	400	4.187	3.119	0.000	12.710	1.016	3.724	14.541

表1中，各变量的偏度几乎全部为正，反映了数据的右厚尾性，说明在“三急”和日常驾驶中，存在部分被保险人远超平均水平的情况。Jarque–Bera检验结果显示，仅高峰时段驾驶时间占比等少数几个变量的分布接近正态，其他变量均不符正态分布的假定。

基于驾驶行为进行保险定价，需了解各类驾驶行为是否对被保险人的出险索赔存在影响。表2给出了各驾驶行为变量是否对出险索赔存在影响的皮尔逊卡方检验，检验结果表明，仅“三急”、高速行驶和违章行为共5个变量对出险索赔次数存在显著影响。同时，表2中各变量的VIF值表明各变量间并不存在共线性问题。

（二）模型设定

在现有保险定价文献中，广义线性模型得到了广泛应用，但其对不同分布均有十分严格的假定。泊松分布要求被解释变量的均值与方差相等；负二项分布要求数据过离散，方差要大于均值[28]。在索

赔频次既定的条件下，各索赔频次间存在着有序的结构与分类，更适合运用有序分类的Logistic模型进行拟合。为此，本文建立以下多元有序分类的Logistic模型，

表 2 各特征变量对出险索赔次数影响的卡方检验

变量	皮尔逊卡方值	伴随概率	VIF
over120ratio	268.661	0.033	4.38
suddenacceleration	119.468	0.000	2.43
suddenbrake	168.510	0.001	3.59
suddensteering	184.749	0.003	4.47
breaches	52.349	0.000	1.72
mileage	272.711	0.476	2.72
rushhoursratio	194.997	0.844	1.58
nighthoursratio	218.984	0.170	1.61
weekendhoursratio	231.091	0.229	1.45
eighty120ratio	269.140	0.538	3.58

$$\begin{aligned} Y_i = & \\ & \beta_1 \text{over120ratio}_{i1} + \beta_2 \text{suddenacceleration}_{i2} + \\ & \beta_3 \text{suddenbrake}_{i3} + \beta_4 \text{suddensteering}_{i4} + \\ & \beta_5 \text{breaches}_{i5} + \beta_6 \text{mileage}_{i6} + \\ & \beta_7 \text{rushhoursratio}_{i7} + \beta_8 \text{nighthoursratio}_{i8} + \\ & \beta_9 \text{weekendhoursratio}_{i9} + \\ & \beta_{10} \text{eight120ratio}_{i10} \ldots + \mu_i \end{aligned} \tag{1}$$

Y_i是有序分类变量，分别取值0~4，代表年出险次数。over120ratio为速度超过120 km/h行驶时间占比；suddenacceleration为急加速次数；suddenbrake为急刹车（急减速）次数；suddensteering为急转弯次数；breaches为违章次数；mileage为行驶里程数；rushhoursratio为高峰时段行车时间占比；nighthoursratio为夜间行车时间占比；weekendhoursratio为周末行车时间占比；eight120ratio为速度在80~120 km/h行车时间占比。

由于因变量为0~4的有序分类变量，可以采用Logistic回归、决策树、随机森林、神经网络等方法进行分类。但有序多分类Logistic模型有助于准确捕捉因变量与自变量间的关系，识别对出险索赔有显著影响的驾驶行为，进而为UBI车险定价提供经验证据，为改变人们的驾驶行为提供正确的方向指引。在诸多分类方法中，Logistic模型有助于揭示自变量对因变量的影响程度及其显著性，既可指导剔除对因变量影响程度较小的自变量，又可基于AIC、BIC信息准则进行变量选择，得到精简高效的模型，为UBI车险实施更稳健的定价。同时，出险索赔次数有助于将被保险人分为不同的风险组别，而不只是简单地识别其在计数上的不同，因而分析驾驶行为对出险索赔次数的影响，更适宜用有序分类的Logistic模型。

三、实证结果与UBI车险定价

（一）模型回归结果

由于OBD采集了众多的驾驶行为数据，经筛选仍须保留10个解释变量。共线性检验表明各变量间并不存在共线性；皮尔逊卡方检验显示仅少数驾驶行为变量对出险索赔次数存在显著影响。基于UBI车险实践和现有文献资料，给出了UBI车险出险索赔次数的有序分类logistic模型回归结果（见表3）。

基于AIC、BIC最小化原则，Brant比例优势检验和Omodel平行性检验，表3中模型2为最优模型。由表3中的模型2可知，over120ratio（速度超过120 km/h行驶时间占比）、suddenacceleration（急加速次数）、breaches（违章次数）和mileage（行驶里程数）4个变量对出险索赔次数存在显著的影响。Suddenbrake（急刹车次数）和suddensteering（急转弯次数）对出险索赔次数存在不显著的影响。

over120ratio（速度超过120km/h行驶时间占比）每上升1%，发生较多出险索赔次数的优势比将上升0.486 5（$e^{0.396}$–1=0.486 5）。高速行驶缩短了驾驶者的反应时间，将显著提高车祸发生的概率。suddenacceleration（急加速次数）每上升1次，发生较多出险索赔的优势将是发生较少出险索赔优势的1.292 3倍（$e^{0.256}$=1.292 3），表明急加速对车祸，尤其是对追尾类事故的发生存在着显著的正向影响。breaches（违章次数）每增加1次，发生更多出险索赔的优势将是发生较少出险索赔优势的3.129 4倍（$e^{1.141}$=3.129 4）。违章次数对出险次数的高解释能力表明，传统车险降低出险概率的做法依然十分有效，加强保险人与交警的合作既必要也高效。mileage（行驶里程数）每增长1公里，发生较多出险索赔次数的优势比将上升0.001 9（$e^{0.002}$–1=0.001 885），与传统广义线性模型以行驶里程为索赔次数重要解释变量的理论相符。suddensteering（急转弯次数）每增加1次，发生较多出险索赔次数的优势比将上升0.065 2（$e^{0.063}$–1=0.065 2），说明“一停二看三通过”，减少急转弯是避免车祸的有效手段。Suddenbrake（急刹车次数）每增加1次，发生较多出险索赔次数的优势比将下降0.015 6（$e^{-0.016}$–1=–0.015 6），说明关键时刻有效的刹车仍然是降低出险次数的有效手段。

表 3　有序分类logistic模型回归结果

变量	（1）	（2）	（3）	（4）	（5）	（6）
over120ratio	0.565***	0.396*	0.412*	0.412*	0.394*	0.791**
	（0.207）	（0.211）	（0.213）	（0.216）	（0.217）	（0.326）
suddenacceleration	0.262***	0.256***	0.251***	0.251***	0.245***	0.234***
	（0.075）	（0.080）	（0.079）	（0.079）	（0.080）	（0.078）
suddenbrake	–0.022	–0.016	–0.020	–0.020	–0.019	0.021
	（0.060）	（0.064）	（0.064）	（0.065）	（0.066）	（0.067）
suddensteering	0.077**	0.063	0.063	0.063	0.064	0.031
	（0.038）	（0.040）	（0.040）	（0.042）	（0.042）	（0.045）
breaches	1.165***	1.141***	1.142***	1.142***	1.153***	1.273***
	（0.419）	（0.424）	（0.424）	（0.429）	（0.427）	（0.448）
mileage		0.002**	0.002***	0.002**	0.002**	0.002***
		（0.001）	（0.001）	（0.001）	（0.001）	（0.001）
rushhoursratio			–0.050	–0.050	–0.026	0.014
			（0.078）	（0.078）	（0.086）	（0.091）
nighthoursratio				0.000	0.020	0.046
				（0.115）	（0.119）	（0.121）
weekendhoursratio					–0.057	–0.073
					（0.084）	（0.086）
eighty120ratio						–0.295*
						（0.176）
Constant cut1	4.339***	6.283***	5.861***	5.862***	5.449***	5.601***
	（0.864）	（1.292）	（1.407）	（1.439）	（1.542）	（1.571）
Constant cut2	5.801***	7.730***	7.319***	7.319***	6.931***	7.139***
	（1.004）	（1.390）	（1.494）	（1.526）	（1.612）	（1.654）
Constant cut3	7.890***	9.916***	9.514***	9.515***	9.195***	9.387***
	（1.202）	（1.586）	（1.670）	（1.694）	（1.746）	（1.784）
Constant cut4	10.470***	13.052***	12.689***	12.690***	12.344***	12.776***
	（1.537）	（2.086）	（2.135）	（2.143）	（2.178）	（2.292）
AIC	133.535 6	128.407 8	129.992 6	131.992 6	133.533 6	132.564 4
BIC	154.393	151.582 4	155.484 9	159.802 4	163.660 9	165.009 2
Brant test	0.999	1.000	0.999	1.000	0.984	1.000
Omodel test	0.191 9	0.613 7	0.700 5	0.237 1	0.172 5	0.168 6
Observations	400	400	400	400	400	400

Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1。

由表4可知，在其他变量均取其均值时，超速行使（over120ratio）在其均值处每增加1单位，将降低低出险索赔次数的概率，提升高出险次数的概率；在其他变量均取其均值时，急加速（suddenacceleration）在其均值处每上升1个单位，将降低低出险次数的可能性，提高高出险次数的可能性；在其他变量均取其均值时，急减速（suddenbrake）在其均值处每上升1个单位，将提升低出险次数的可能性而降低高出险次数的可能性；在其他变量均取其均值时，急转弯（suddensteering）在其均值处每上升1个单位，净降低低出险次数的可能性而提高高出险次数的可能性；在其他变量均取其均值时，违章次数（breaches）在其均值处每上升1个单位，净降低低出险次数的可能性而提高高出险次数的可能性；在其他变量均取其均值时，行驶里程数（mileage）在其均值处每上升1个单位，将降低最低出险次数的可能性而提高高出险次数的可能性，对高出险次数具有更高的显著性。

（二）基于驾驶行为的车险定价

根据模型对第i位驾驶者在0~4共5种出险索赔次数上的概率分布：

$$p_{ij}=\begin{cases} p_{i0} & \text{if claims}=0 \\ p_{i1} & \text{if claims}=1 \\ p_{i2} & \text{if claims}=2 \\ p_{i3} & \text{if claims}=3 \\ p_{i4} & \text{if claims}=4 \end{cases} \quad ,j=0,1,2,3,4$$

可得第i位驾驶者的期望索赔次数：$ex_claims_i=$

$\sum_{j=0}^{4}(p_{ij}*j)$。以该驾驶者的期望索赔次数与该公司全体车险被保险人的期望索赔次数的比值乘以基础费率，即可得到该被保险人的个人车险费率。以本文样本中前25位驾驶者为例，可得他们基于驾驶行为习惯的个人基础车险保费（见表5）。

由表5可知，对第6、9、18位驾驶者而言，因其期望出险索赔次数远低于全体被保险人索赔次数

表 4　各变量的边际效应分解

变量	（1）	（2）	（3）	（4）	（5）
over120ratio	–0.030 2*	–0.001 1	0.003 9	0.014 7**	0.012 6*
	（–1.82）	（–0.70）	（0.72）	（2.20）	（1.83）
suddenacceleration	–0.019 5***	–0.000 7	0.002 5	0.009 5**	0.008 2***
	（–3.79）	（–0.60）	（0.88）	（2.57）	（3.66）
suddendeceleration	0.001 2	0.000 04	–0.000 2	–0.000 6	–0.000 5
	（0.25）	（0.27）	（–0.26）	（–0.24）	（–0.25）
riskysteering	–0.004 8*	–0.000 2	0.000 6	0.002 3	0.002
	（–1.67）	（–0.63）	（1.12）	（1.29）	（1.63）
breaches	–0.086 2***	–0.003 1	0.011 2	0.042 4**	0.036 3***
	（–2.85）	（–0.64）	（0.90）	（2.22）	（2.80）
mileage	–0.000 01**	0.000 05	0.000 02	0.000 07**	0.000 06***
	（–2.55）	（–0.62）	（0.81）	（2.12）	（3.11）

Z statistics in parentheses*** p<0.01, ** p<0.05, * p<0.1。

表 5　基于驾驶行为定价的个人基础保费

序号	车辆品牌型号	初登日期	基础保费	期望出险次数	个人基础保费
1	路虎CJL6440LIAW5风尚版	2014	6 905.33	2.214	11 353.09
2	福克斯CAF7180B时尚型	2014	2 244.13	2.576	4 292.208
3	福特CAF7152BC4运动型	2013	1 939.73	3.873	5 578.193
4	福克斯CAF7102A51超能风尚型	2014	2 364.21	2.462	4 322.799
5	福克斯CAF7163B4	2008	1 514.07	1.621	1 822.461
6	福克斯CAF7102A38	2015	2 526.89	0.040	75.100
7	福特CAF7200a42	2015	2 382.69	2.253	3 986.672
8	雪佛兰SGM6470TAXD	2015	3 370.29	2.526	6 321.867
9	丰田GTM7160LVCE	2014	2 127.06	0.053	83.357
10	LVSHCFDB4DE242956	2011	2 071.55	0.057	87.919
11	LVSHCAAE9BF819215	2011	1 684.86	3.655	4 573.018
12	雪佛兰SGM7168ATB	2014	1 966.76	0.070	102.294
13	雪佛兰SGM6473ATA	2015	2 142.64	2.023	3 219.186
14	骐达2012	2013	1 961.63	1.687	2 457.534
15	福克斯2013	2012	1 930.12	0.299	427.827
16	途安2012	2012	1 951.16	0.188	273.095
17	本田歌诗图2012	2013	1 918.18	3.968	5 651.923
18	长城M42012	2015	2 405.86	0.033	58.558
19	一汽大众CC2012	2015	1 966.76	2.147	3 135.966
20	广本凌派2013	2013	2 234.49	2.214	3 674.113
21	吉普自由客PATRIOT2.4L	2012	2 076.18	0.298	458.692
22	大众FV7207FCDWG	2012	1 891.64	0.179	251.689
23	甲壳虫BEETLE2.0L	2010	1 609.50	1.775	2 121.284
24	长安SC7164B	2013	2 348.6	3.584	6 250.706
25	捷豹JAGUARXJL2.0T	2011	1 579.92	0.560	656.449

的均值（1.346 667），因而其期望出险次数与全体被保险人索赔次数均值的比值小于1，其个人基础保费也低于其所对应车型的基础保费。其他22位驾驶者，由于他们的期望出险次数远高于全体出险次数的均值，因而其个人基础保费也远高于对应车型的基础保费。这种基于有序分类logistic模型所获得的各位驾驶者在各种出险水平上的概率分布，通过引入具体驾驶者的期望出险次数与全体被保险人出险次数均值的比值的方式进行车险定价更好地体现了NCD的思想，实现了原则性与灵活性的有机结合，有利于车险费率的市场化改革的深入推进。

四、稳健性检验

（一）泊松分布

对出险索赔次数究竟应适用有序分类的logistic模型，还是应适用计数的泊松模型，学界一直存在争议。为检验本文所建模型的稳健性，特给出泊松模型回归结果（见表6）。

表 6　泊松分布的回归结果

变量	（1）	（2）	（3）	（4）	（5）	（6）
over120ratio	0.105***	0.069*	0.069	0.075*	0.070*	0.073
	（0.040）	（0.040）	（0.043）	（0.043）	（0.039）	（0.061）
suddenacceleration	0.044**	0.045***	0.045***	0.043***	0.042**	0.042**
	（0.017）	（0.016）	（0.016）	（0.016）	（0.017）	（0.017）
suddendeceleration	0.004	0.004	0.004	0.003	0.002	0.002
	（0.018）	（0.018）	（0.018）	（0.018）	（0.018）	（0.022）
suddensteering	0.024*	0.020	0.020	0.021	0.021*	0.021
	（0.013）	（0.013）	（0.013）	（0.014）	（0.013）	（0.016）
breaches	0.332***	0.285***	0.285***	0.309***	0.281***	0.282***
	（0.093）	（0.087）	（0.087）	（0.100）	（0.100）	（0.101）
mileage		0.000*	0.000*	0.000*	0.000**	0.000**
		（0.000）	（0.000）	（0.000）	（0.000）	（0.000）
rushhoursratio			0.001	–0.001	0.006	0.006
			（0.020）	（0.020）	（0.021）	（0.021）
nighthoursratio				–0.016	–0.003	-0.002
				（0.026）	（0.027）	（0.027）
weekendhoursratio					–0.034	–0.034
					（0.025）	（0.025）
eighty120ratio						–0.003
						（0.047）
Constant	–1.564***	–1.800***	–1.807***	–1.757***	–1.497***	–1.496***
	（0.249）	（0.269）	（0.324）	（0.331）	（0.378）	（0.378）
AIC	179.118	179.449	181.448	183.288	184.292	186.291
BIC	193.023	195.671	199.988	204.145	207.467	211.783
Observations	400	400	400	400	400	400

Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1。

表6显示，尽管基于AIC、BIC最小化原则，较之模型1，模型2并不是最优的，但在全部6个模型中仍属次优，因而本文依然以模型2为基准展开分析。表6中的模型2显示，over120ratio、suddenacceleration、breaches和mileage依然对出险索赔次数存在显著的影响。suddenbrake和suddensteering对出险索赔次数存在不显著的影响。

这与表3中模型2所显示内容完全相符，说明本文针对驾驶行为对出险索赔次数影响所构建的有序分类logistic模型与泊松模型具有高度的一致性，反映本文所构建的模型具有较高的稳健性。

（二）负二项分布

在探索对出险索赔次数影响因素建立模型的过程中，负二项分布由于在处理过分散、零膨胀数据方面的优势而得到了广泛的应用。本文尝试构建驾驶行为影响出险索赔次数的模型，也给出了对应的负二项分布模型的回归结果（见表7），以进行比较分析。

同样地，在AIC、BIC最小化原则之下，表7中模型2劣于模型1，但优于其他4个模型。不影响一

致性，仍以模型2为基准展开分析。表7中模型2同样显示，over120ratio、suddenacceleration、breaches和mileage对出险索赔次数存在显著的影响，而suddenbrake和suddensteering则存在不显著的影响。表7所揭示的内容与表3、表6并不存在显著的差异，既说明这4种驾驶行为是影响出险索赔次数的关键变量，也显示本文所构建的模型具有很强的稳健性。

表 7　负二项分布的回归结果

变量	（1）	（2）	（3）	（4）	（5）	（6）
over120ratio	0.105***	0.069*	0.069	0.075*	0.070*	0.073
	（0.040）	（0.040）	（0.043）	（0.043）	（0.039）	（0.061）
suddenacceleration	0.044**	0.045***	0.045***	0.043***	0.042**	0.042**
	（0.017）	（0.016）	（0.016）	（0.016）	（0.017）	（0.017）
suddendeceleration	0.004	0.004	0.004	0.003	0.002	0.002
	（0.018）	（0.018）	（0.018）	（0.018）	（0.018）	（0.022）
suddensteering	0.024*	0.020	0.020	0.021	0.021*	0.021
	（0.013）	（0.013）	（0.013）	（0.014）	（0.013）	（0.016）
breaches	0.332***	0.285***	0.285***	0.309***	0.281***	0.282***
	（0.093）	（0.087）	（0.087）	（0.100）	（0.100）	（0.101）
mileage		0.000*	0.000*	0.000*	0.000**	0.000**
		（0.000）	（0.000）	（0.000）	（0.000）	（0.000）
rushhoursratio			0.001	–0.001	0.006	0.006
			（0.020）	（0.020）	（0.021）	（0.021）
nighthoursratio				–0.016	–0.003	–0.002
				（0.026）	（0.027）	（0.027）
weekendhoursratio					–0.034	–0.034
					（0.025）	（0.025）
eighty120ratio						–0.003
						（0.047）
Constant	–1.564***	–1.800***	–1.807***	–1.757***	–1.497***	–1.496***
	（0.249）	（0.269）	（0.324）	（0.331）	（0.378）	（0.378）
AIC	181.118	181.449	183.448	185.288	186.292	188.291
BIC	197.340	199.989	203.305	208.463	211.785	261.100
Observations	400	400	400	400	400	400

Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1。

（三）基于SVM的稳健性检验

如前所述，机器学习算法在数据分类上具有较高的效率与精度。支持向量机（Support Vector Machine,简称SVM）算法等方法近年来在车险出险索赔预测上得到广泛的运用，对比分析有序分类logistic模型和支持向量机（SVM）的预测结果显得尤为必要[29]。表8给出了样本前25个观测值的实际出险索赔次数和logistic模型、泊松模型、负二项模型及支持向量机预测的出险索赔次数。

表8显示，从离差平方和（Sum of Squares of Deviations）来看，有序分类logistic模型的预测结果仅劣于支持向量机的预测，但优于泊松分布和负二项分布模型的预测结果，且后两者的预测则完全相同。支持向量机的整数预测结果少于小数预测结果，这与孟生旺等（2018）所强调的可加性观点相符[30]。有序分类logistic模型虽劣于机器学习算法，但优于泊松分布和负二项分布模型，说明本文所构建的模型较为稳健。支持向量机的预测结果对有序分类logistic模型预测结果的支持，说明本文所构建的模型在能准确识别出险索赔影响因素的同时，具有较高的稳健性。

表 8　各种模型所预测的出险索赔次数

序号	实际索赔次数	逻辑模型预测加权索赔次数	泊松模型预测索赔次数	负二项模型预测索赔次数	支持向量机径向预测整数索赔次数	支持向量机径向预测非整索赔次数
1	2	2.214	1.310	1.310	2	1.912
2	3	2.576	1.621	1.621	3	3.147
3	4	3.873	4.909	4.909	4	3.854
4	3	2.462	1.576	1.576	2	3.089
5	1	1.621	0.952	0.952	2	1.146
6	0	0.040	0.307	0.307	0	–0.131
7	3	2.253	1.631	1.631	3	2.826
8	3	2.526	1.709	1.709	3	2.854
9	0	0.053	0.333	0.333	0	0.147
10	0	0.057	0.309	0.309	0	0.100
11	4	3.655	3.432	3.432	4	3.854
12	0	0.070	0.362	0.362	0	0.074
13	3	2.023	1.379	1.379	3	2.854
14	2	1.687	1.064	1.064	2	1.944
15	0	0.299	0.470	0.470	0	0.030
16	0	0.188	0.440	0.440	0	–0.091
17	4	3.968	6.124	6.124	4	3.848
18	0	0.033	0.286	0.286	0	0.146
19	2	2.147	1.519	1.519	2	2.146
20	2	2.214	1.383	1.383	2	2.146
21	0	0.298	0.527	0.527	0	0.146
22	0	0.179	0.437	0.437	0	0.147
23	2	1.775	1.282	1.282	2	1.882
24	4	3.584	3.445	3.445	4	3.853
25	2	0.560	0.782	0.782	2	1.294
SSD		5.496	21.428	21.428	2.000	0.905

五、结论与建议

基于中国大陆某财产保险公司UBI车险所收集的被保险人驾驶行为数据，构建了出险索赔次数影响因素的有序分类logistic模型，实证结果表明，高速行驶、急加速、违章次数和行驶里程将显著提高车险出险索赔次数；急刹车和急转弯对出险索赔次数存在影响，但均不显著。其中，急刹车有助于降低，而急转弯将提高出险次数。全部6个解释变量中，交通违章次数对出险次数有极强的解释能力，说明传统车险所重视的变量对UBI车险定价仍具有极高的指导意义。“三急”中急加速和急转弯将显著提高索赔次数，但与现有文献中的结论相悖，本文发现急刹车（急减速）能降低而非提高出险索赔次数。传统车险定价的核心变量，行驶里程数确实对出险索赔次数有显著的影响，但其显著性远低于急加速和违章次数等驾驶行为变量，显示UBI车险将成为未来车险发展的大方向。

基于被保险人驾驶行为，以有序分类logistic模型预测其出险索赔次数，再以预测的出险索赔次数及具体车辆的基准费率进行UBI车险定价显著优于现行的NCD车险定价方法。不可否认，部分被保险人的实际出险次数可能低于或高于基于其驾驶行为预测的出险索赔次数，但这种高于或低于预测值的情况可能仅仅是由于运气等随机扰动造成的。因而，基于有序分类logistic模型的预测值而非NCD进行车险定价更有理论依据，更贴近驾驶行为，也更能反映真实的驾驶风险。泊松分布模型、负二项分布模型及支持向量机等方法均证明有序分类logistic模型在基于驾驶行为预测出险索赔次数上具有更高的准确率与稳定性。

违章次数和行驶里程数是传统车险定价的核心变量，也是UBI车险定价的关键变量，但与其说这两个变量是驾驶行为变量，还不如说它们是反映驾驶行为结果的变量。为通过UBI车险定价改变人们的驾驶行为，降低出险索赔次数，交通管理部门在驾驶证照考试中应强化违章教育，提高人们对交通法律法规的敬畏与尊重。保险公司可基于实证研究，提高违章次数在定价中的仅重，以价格这只无形的手来引导、规范人们的驾驶行为。同时，提高行驶里程数在UBI车险定价中的权重可诱导人们改变驾驶行为，减少开车出行频次，既有助于降低出险索赔次数，也对减炭减排、保护环境有积极的意义。

“三急”中急加速对车险出险索赔次数存在显著的正向影响；急转弯对车险出险索赔次数存在不显著的正向影响；急刹车（或急减速）对出险索赔

次数存在不显著的负向影响。因此，交通管理部门应强化现行的路口加速限制；对转弯弧度设置相应的规定；鼓励紧急情况下的急刹车，并在车险定价中设置相应的激励机制，引导人们改变驾驶行为。

超速行驶将显著增加出险索赔次数，是UBI车险定价的关键变量。交通管理部门应强化时速限制，加强行人安全教育，寻求交通效率与交通安全的平衡，在UBI车险定价中设置相应的激励机制，引导人们尽可能多地常速驾驶。通过在UBI车险定价中设置相应的激励机制，以有序、渐近的办法引导人们改变驾驶行为，逐步形成全社会健康驾驶、文明驾驶的氛围，将有助于支持车险业的长期健康发展。

参考文献

[1] 申相彬. 汽车后市场发展前景及人才需求分析[J]. 经济研究导刊, 2013(17): 70-72.

[2] 陈敏, 王华. 车辆保险欺诈的信号传递博弈模型[J]. 江汉大学学报: 自然科学版, 2012, 40(1): 44-46.

[3] 朱相全. CPS中数据处理相关技术的应用研究[D]. 南京: 南京邮电大学, 2014.

[4] 王选鹤, 孟生旺, 王雅实. 基于厚尾损失分布的汽车保险定价模型及其应用[J]. 保险研究, 2017(4): 67-78.

[5] 张永霞, 孟生旺. 我国商业车险奖惩系统研究[J]. 保险研究, 2016(10): 3-15.

[6] VICKREY W. Automobile Accidents, Tort Law, Externalities, and Insurance: An Economist's Critique[J]. Law & Contemporary Problems, 1968, 33(3): 464-487.

[7] BUTLER P, BUTLER T. Driver Record: A Political Red Herring That Reveals the Basic Flaw in Automobile Insurance Pricing[J]. Journal of Insurance Regulation, 1989, 8(2): 200-234.

[8] LITMAN T. Distance-Based Vehicle Insurance-A Practical Strategy for More Optimal Vehicle Pricing[EB/OL]. [2019-02-24]. https://link.springer.com/chapter/10.1007/978-3-642-57669-0_8.

[9] LITMAN T. Distance-Based Vehicle Insurance Feasibility, Costs and Benefits[J]. Victoria Transport Policy Institute, 2011(10): 1-90.

[10] LEMAIRE J. Bonus-Malus Systems in Automobile Insurance[M]. Dordrecht: Springer, 1995.

[11] 王和. 信息技术: 继往开来的力量——保险信息技术发展40年巡礼[J]. 保险研究, 2018(12): 118-122.

[12] 郁佳敏. 车联网大数据时代汽车保险业的机遇和挑战[J]. 南方金融, 2013, 4(48): 89-95.

[13] 杨磊. 车联网时代保险公司开发UBI车险产品策略探析[J]. 保险研究, 2014(D7): 33-35.

[14] 乔木. 大数据语境下UBI发展现状及趋势研究[J]. 现代商业, 2015(1): 53-54.

[15] 蒋寅, 王洁. 车联网业务与保险业务的融合创新[J]. 电信科学, 2012, 28(6): 22-24.

[16] LEO G. Gradient Boosting Trees for Auto Insurance Loss Cost Modeling and Prediction[J]. Expert Systems with Applications, 2012, 39(3): 3659-3667.

[17] LIU Y, WANG J B, LV S G. Using Multi-class Adaboost Tree for Prediction Frequency of Auto Insurance[J]. Journal of Applied Finance and Banking, 2014, 4(5): 45-53.

[18] HERE, 瑞士再保险. 未来的机动车保险——车联网和先进驾驶辅助系统将如何影响车险市场[J]. 中国经济周刊, 2016(39): 70-73.

[19] 王选鹤, 孟生旺, 杨默. 车险索赔次数预测模型的扩展与应用[J]. 保险研究, 2018(11): 82-92.

[20] 孟生旺, 黄一凡. 驾驶行为保险的风险预测模型研究[J]. 保险研究, 2018(8): 21-34.

[21] 张连增, 段白鸽. 行驶里程数对车险净保费的影响研究——基于公路里程对交通事故损失的影响视角[J]. 保险研究, 2012(6): 29-38.

[22] AYUSO M, GUILLEN M, PÉREZ-MARÍN AM. Telematics and Gender Discrimination: Some Usage-based Evidence on Whether Men's Risk of Accidents Differs from Women's[J]. Risk, 2016, 4(2): 10.

[23] WEIDNER W, TRANSCHEL F W G, WEIDNER R. Classification of Scale-sensitive Telematic Observables for Risk Individual Pricing[J]. European Actuarial Journal, 2016, 6(1): 3-24.

[24] WÜTHRICH M V. Covariate Selection from Telematics Car Driving Data[J]. European Actuarial Journal, 2017, 7(2): 89-108.

[25] GAO G, WÜTHRICH M V. Feature Extraction from Telematics Car Driving Heatmaps[J]. European Actuarial Journal, 2018, 8(2): 383-406.

[26] 高光远, 孟生旺. 基于车联网大数据的车险费率因子分析[J]. 保险研究, 2018(1): 90-100.

[27] 高岩. 基于驾驶行为分类的UBI费率厘定模型[D]. 北京: 北京交通大学, 2017.

[28] 李政宵, 孟生旺. 相依风险条件下的汽车保险定价模型[J]. 保险研究, 2016(7): 68-77.

[29] 孟生旺, 李天博, 高光远. 基于机器学习算法的车险索赔概率与累积赔款预测[J]. 保险研究, 2017(10): 42-53.

[30] 韩家群, 刘南杰, 黄波, 等. 基于车联网大数据的UBI系统研究[J]. 计算机技术与发展, 2016(12): 26-29.

编 辑 邓婧