

关于在我国推行绿色汽车保险的可行性分析 ——基于调研数据的Logistic模型研究

许可 黄若奕

(华东师范大学金融与统计学院, 上海, 200241)

摘要

自2009年绿色汽车保险的理念被引入我国以来, 如何合理设计这一新产品, 及新产品推行的可行性、必要性等就备受各界关注. 由于目前国内几乎所有财险公司的数据库中还没有行驶里程数的统计数据, 所以还没有出现直接利用汽车行驶里程数所做的研究. 本文通过抽样调查的渠道获取了行驶里程数的数据, 建立了一个Logistic模型, 并得出行驶里程数对索赔概率有着显著影响的结论, 从而论证了发展绿色汽车保险的可行性. 同时根据调研发现的问题, 首次在绿色车险中提出了相对行驶里程数的概念和对投保车辆进行分类的建议, 希望能对保险及相关行业有所启示.

关键词: 绿色汽车保险, 行驶里程数, Logistic模型.

学科分类号: O212.

§1. 引言

中国的节能减排政策正在追赶国际的脚步, 但是大多数政策集中在汽车产业, 后端服务方面的政策和措施较少, 尤其是金融保险服务这一块几乎是空白. 在这样的环境和契机下, 设计出绿色汽车保险、提高车主的节能环保意识、降低温室气体的排放等势在必行. 所谓绿色汽车保险, 是指在传统车险费率表中添加影响汽车尾气排放量的因素——汽车行驶里程数, 通过将其有效地与保费挂钩, 来间接鼓励人们尽量减少不必要的车程, 进而达到降低汽车二氧化碳排放量的目的.

在国外, 关于车险里程定价模式的研究开展较早, 研究认为PAYD (Pay-As-You-Drive) 车险产品的设计是合理的^[1], PAYD车险产品近几年已逐渐成为欧美发达国家市场上一款流行的产品. 在我国, 天平汽车保险公司于2009年11月起开始探索绿色车险产品, 但还没有成熟的产品进入市场. 同时由于具体国情不同, 汽车行驶里程数与保险索赔在我国是否存在一定的关系, 绿色汽车保险在我国推行是否可行等问题还需要进一步的探索. 由于目前国内几乎所有财险公司的数据库中还没有行驶里程数的统计数据, 难以对行驶里程数进行量化研究, 因此大多学者在这方面的研究都停留在理论上^[2, 4], 如段白鸽等在文献[2]中总结了国外车险里程定价的理论基础以及1998年以来在保险市场上的实践情况, 张连增等在文

本文2012年9月6日收到, 2012年12月12日收到修改稿.

献[3]中用公路里程数代替行驶里程数进行研究. 本文通过调研数据, 建立了一个汽车行驶里程数与发生索赔概率之间的量化模型.

§2. 含行驶里程因素的Logistic模型

本文利用Logistic模型建立行驶里程等影响因素与索赔发生概率之间的关系.

2.1 Logistic模型介绍

Logistic模型是一种非线性分类的统计方法, 属于广义线性模型, 主要用来处理因变量为分类变量的问题. 本文将建立因变量为二项分类变量的Logistic模型.

令因变量 Y 服从二项分布, 事件发生时取值为1、否则为0. 令 $Y = 1$ 的概率为 $\pi(Y = 1)$, 则 k 个解释变量为 X_1, X_2, \dots, X_k 所对应的Logistic回归模型为

$$\pi(Y = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (2.1)$$

或

$$\text{logit}[\pi(Y = 1)] = \ln \left[\frac{\pi(Y = 1)}{1 - \pi(Y = 1)} \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k, \quad (2.2)$$

其中, β_0 为截距(或称常数项), β_i 是 X_i ($i = 1, \dots, k$)所对应的偏回归系数(简称回归系数), $\ln[\pi(Y = 1)/(1 - \pi(Y = 1))]$ 称为 $\pi(Y = 1)$ 的logit变换值.

公式(2.1)与(2.2)相互等价. (2.1)通常被称为Logistic回归预测模型, 可以用来预测事件发生的概率. (2.2)与一般的回归模型更相似, 即左侧 $\pi(Y = 1)$ 的logit变换值与解释变量呈线性关系. Logistic回归模型实际上是对概率 $\pi(Y = 1)$ 进行logit变换后的线性回归模型, 所以通常也称Logistic回归模型为logit模型.

2.2 解释变量说明

在本文中, 假设因变量 $Y = 1$ 表示索赔发生、 $Y = 0$ 表示索赔未发生, $\pi(Y = 1)$ 表示保单发生索赔的概率. 本研究调查的样本数据包含7个解释变量: 性别、年龄(岁)、驾龄(年)、价格(万元)、排量(升)、平均每年耗油量(升)、平均每年行驶里程数(公里). 使用抽样调查的数据资料建立模型, 分析影响索赔的因素.

抽样调查所研究的各种因素的定义及说明见表1.

2.3 数据来源

本文使用的数据来源于分层随机抽样. 将全国各省市以南北分界, 将调查地点分成了两大层, 在南方地区中选取了江苏、安徽、上海, 在北方地区中选取了河北、山东, 由6位调查员在上述五省进行问卷调查. 问卷主要包括车主和汽车的客观基本信息, 及车主对绿色汽车保险产品的主观看法和建议. 本次调查共发放问卷352份, 收回341份, 回收率高达96.9%. 在收回的问卷中, 有效答卷329份, 有效率为96.5%.

表1 各因素说明表

因素	符号	含义(赋值)
性别	X_1	男(0), 女(1)
年龄(岁)	X_2	18 ~ 24(1), 25 ~ 30(2), 31 ~ 40(3), 40以上(4)
驾龄(年)	X_3	1 ~ 4(1), 5 ~ 8(2), 9 ~ 12(3), 13 ~ 16(4), 大于16(5)
价格(万元)	X_4	10以下(1), 10 ~ 20(2), 20 ~ 30(3), 30 ~ 40(4), 40 ~ 50(5), 50以上(6)
排量(升)	X_5	1以下(1), 1 ~ 1.6(2), 1.6 ~ 2.5(3), 2.5 ~ 4(4), 4以上(5)
平均年耗油量(升)	X_6	1000以下(1), 1000 ~ 1500(2), 1500 ~ 2000(3), 2000 ~ 2500(4), 2500以上(5)
平均年行驶里程数(公里)	X_7	0 ~ 15000(1), 15000 ~ 20000(2), 20000 ~ 30000(3), 30000 ~ 40000(4), 40000以上(5)

2.4 Logistic模型的建立

使用SAS软件建立Logistic模型.
第一步是只包含截距项而无解释变量时的结果.
第二步, 解释变量 X_7 进入模型, 得到表2.

表2 AIC、SC、 $-2\log L$ 指标值

Criterion	Intercept Only	Intercept and Covariates
AIC	118.880	113.152
SC	121.323	118.037
$-2\log L$	116.880	109.152

第三步对模型进行总的检验, 原假设是所有解释变量的参数为0, 检验结果为 p 值均小于0.05 (见表3), 故拒绝原假设, 即模型总体有意义, 解释变量的参数不全为0. 其中, $-2\log L$ 中的 L 就是Likelihood Ratio, 其 p 值是0.0054, Score的 p 值是0.0059, 都小于0.05, 故 X_7 是一个很显著的解释变量, 即验证了汽车行驶里程数是影响索赔发生与否的显著因素.

表3 模型的检验

Test	Chi-Square	DF	Pr > Chi-Square
Likelihood Ratio	7.7286	1	0.0054
Score	7.5738	1	0.0059
Wald	7.2295	1	0.0072

第四步, 当解释变量 X_2 进入后, 得到表4.

表4 添加 X_2 后的AIC、SC、 $-2\log L$ 指标值

Criterion	Intercept Only	Intercept and Covariates
AIC	118.880	109.347
SC	121.323	116.675
$-2\log L$	116.880	103.347

AIC(赤池)信息准则和SC(舒瓦茨)准则可以用来比较不同的模型, 它们的值越小, 模型就越好. 当解释变量 X_2 进入后, AIC、SC都变小了, 说明模型拟合的更好了. 同时 $-2\log L$ 和Score对应的 p 值也更小, 见表5.

表5 添加 X_2 后模型的检验

Test	Chi-Square	DF	Pr > Chi-Square
Likelihood Ratio	13.5332	2	0.0012
Score	12.5621	2	0.0019
Wald	11.1116	2	0.0039

最后得到极大似然估计的结果, 见表6.

表6 极大似然估计(MLE)结果

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
Intercept	1	-4.3355	1.4473	8.9735	0.0027
X_2	1	0.8141	0.3609	5.0888	0.0241
X_7	1	0.4469	0.1535	8.4748	0.0036

表6给出了模型参数的估计, 据此可以写出该回归方程

$$\text{logit}[\pi(Y=1)] = \ln \left[\frac{\pi(Y=1)}{1-\pi(Y=1)} \right] = -4.3355 + 0.8141X_2 + 0.4469X_7.$$

该回归方程保留了解释变量 X_7 (平均年行驶里程数), 验证了行驶里程数对索赔发生概率有一定影响. 可以推断, 行驶里程数在一定程度上决定了该车与其他车辆相遇及发生危险的可能性大小, 从而影响了索赔发生. 与此同时, 解释变量 X_2 (年龄)也被保留, 因为年龄在一定程度上决定车主反应的快慢及其驾驶态度, 从而影响索赔发生. 该回归方程剔除了其他解释变量, 比如性别、汽车价格、汽车排量等. 由于性别仅影响驾驶员开车习惯等, 汽车价格只是一定程度上决定汽车好坏, 汽车排量主要影响的是百公里油耗, 这些都与驾驶员驾车熟练程度等其它可能影响到事故发生的因素之间没有联系, 因此不会影响事故发生的概率, 从而不是影响索赔发生的关键因素.

2.5 模型评价

评价Logistic回归模型的主要指标有模型卡方统计和预测准确性.

模型卡方统计检验的是模型中所包含的解释变量对因变量有显著的解釋能力, 即所设模型优于零假设模型(即只包含常数项的模型). 在Logistic中用似然比检验, 在SAS输出结果中就是Likelihood Ratio对应的值. 上文中已说明Likelihood Ratio对应的 p 值小于0.05, 即所选解释变量对因变量有显著的解釋能力.

在实际应用中, 人们往往更关心预测准确性, 即观测值和模型预测的条件概率的关联强度, 表7通过预测数和观测数的关联性分析度量每个个体是否接近观测值.

表7 预测数和观测数的关联性分析

Percent Concordant	68.0
Percent Discordant	22.8

表7中, 一致性比率Percent Concordant为68.0%, 不一致性比率Percent Discordant为22.8%, 说明预测值与观测值在现有水平上有较强的关联性, 回归模型有很强的预测能力. 同时, 通过残差分析, 可以得出学生化残差大于2.000的观测(共一个), 如表8所示.

表8 学生化残差大于2.000的观测列表

case	Selected Status	Observed Y	Predicted	Predicted Group	Temporary Variable	
					Resid	ZResid
50	S	1**	0.114	0	0.886	2.792

综上, 模型检验结果比较理想, 模型拟合较好.

§3. 对绿色汽车保险产品的改进

3.1 相对里程数的引入

本次调查中有一项是车主对于推出绿色汽车保险产品的基本态度, 结果有182人支持, 超过一半的支持率(55.3%)也进一步肯定了推出该产品的实际可行性. 另外还有97人持观望态度, 50人反对. 通过进一步调研发现, 在持反对态度的车主中, 有86%的车主的年行驶里程数较高. 为了鼓励年行驶里程数较高的车主也积极参与到绿色汽车保险行列中来, 本文首次引入“相对行驶里程数”这一概念. 只要今年相对去年的里程数减少(即使行驶里程的基数较大), 年末仍然可以得到一定的保费返还, 返还保费 = 本保险年度保费 \times B1. 其中B1为奖励系数, 由投保车辆本保险年度行驶里程数与往年平均行驶里程数之比决定, 在新产品推行后可根据获得的历史数据进行调整. 初始拟定的奖励系数是通过在调研中与车主交流并咨询专家得到, 见表9. 通过这种奖励措施, 无论绝对行驶里程数处于哪个水平的车主都有机会获得经济利益, 支持该产品的比例一定会大大提高.

表9 绿色车险年末奖励系数表

系数因子	符号	系数值	说明
投保车辆在	B1a	0.2-0.3	$B1 < 0.6$
本保险年度中	B1b	0.14	$0.6 \leq B1 < 0.7$
行驶里程数与	B1c	0.1	$0.7 \leq B1 < 0.8$
往年平均行驶	B1d	0.07	$0.8 \leq B1 < 0.9$
里程数之比B1	B1e	0.05	$0.9 \leq B1 < 1.0$

3.2 新保险产品适用的汽车类型

新产品推出前应认真分析其适用范围. 之前调查主要针对私家车车主, 后又对安徽省池州市59台长途客运汽车的年行驶里程数和索赔金额进行了调查. 由于长途客运汽车的行驶路线固定, 所以其行驶里程数也相对固定, 可以获得准确的相关数据. 通过以索赔额为因变量, 年行驶里程数为解释变量, 进行了各种曲线估计, 方程包括线性、对数、二次、三次、幂、指数、Logistic等. 使用SPSS统计软件进行拟合, 模型汇总结果见表10, 其中Sig.是显著性水平, 应当越小越好, 同时必须在0.05以下, 但拟合结果中所有Sig.均大于0.05, 可见这些方程中没有一个能够较好的拟合长途客运汽车的年行驶里程数和索赔金额之间的关系, 初步判定两者之间没有一定的相关关系. 又从散点图(图1)中再次验证两者之间没有其他显著的相关关系. 据此推测, 由于长途客运汽车功能的特殊性, 以及其行驶里程数变化的弹性很小等原因, 其年行驶里程数和索赔金额之间不存在正相关关系, 因此, 新产品不适合长途客运汽车. 可见新产品推出前要对适用车辆进行分类处理.

表10 模型汇总

方程	R^2	F统计量	df1	df2	Sig.
线性	0.002	0.088	1	52	0.768
对数	0.001	0.072	1	52	0.789
二次	0.003	0.068	2	51	0.935
三次	0.003	0.056	3	50	0.982
幂	0.001	0.053	1	52	0.819
指数	0.000	0.017	1	52	0.896
Logistic	0.000	0.017	1	52	0.896

§4. 结 论

传统车险基于时间收取固定保费, 不考虑行驶里程数, 然而本文研究表明年行驶里程数是影响索赔发生的重要因素. 对于私家车来说, 行驶里程数与索赔发生之间存在着正相

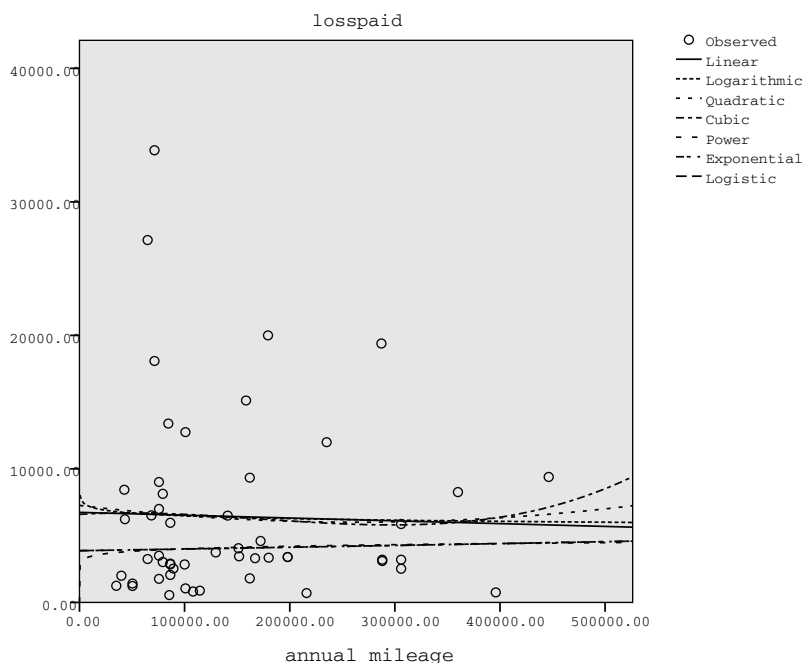


图1 行车里程数与索赔额的散点图

关的关系,绿色汽车保险将行驶里程数作为费率因子添加到车险费率表中的方案是可行的.同时,设计绿色汽车保险产品时应关注适用车型,还可以考虑添加相对行驶里程数等一类易被忽略的但一定程度上会影响车主投保积极性的费率因子.这种新产品的推行一方面可以达到保护环境的目的,另一方面也会为保险公司减少理赔的数量和金额,节约成本.总之,绿色汽车保险是添加了行驶里程数这一重要因素的新型车险,在我国推行具有可行性,也具有积极的社会效应,保险公司应不断探索和完善其公平精确的费率模式.

本文为行驶里程数与索赔发生之间建立的Logistic模型,可以为绿色汽车保险推广的可行性提供一个参考.但由于本文所用数据的局限性,参数估计结果的准确性等还有待进一步的改进.

参 考 文 献

- [1] Ferreira, J., Jr. and Minikel, E., *Pay-As-You-Drive Auto Insurance in Massachusetts: A Risk Assessment and Report on Consumer, Industry and Environmental Benefits*, the Department of Urban Studies and Planning, Massachusetts Institute of Technology, 2010.
- [2] 段白鸽, 余东发, 张连增, 国外车险里程定价理论与实践, *保险研究*, 2(2012), 72-79.
- [3] 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, *保险研究*, 6(2012), 29-38.
- [4] 张媛媛, 在中国推行基于行车里程定价的车险(PAYD)研究, *群文天地*, 5(2011), 208-209.