

天气因素对航班延误险定价的影响分析

刘敬真 林荔圆

[摘要] 航班延误的概率通常会受到天气、航空管制、航空公司等多种因素的影响，而目前保险市场上航班延误险的价格通常是固定的，在不同天气状况下、不同航空公司之间并没有差异。这种定价方式容易导致投保人在提前获取天气信息之后采取逆向选择行为，保险公司可以通过采取分类费率来降低这种逆向选择行为带来的成本。本文首先通过分析航班延误时间数据和天气数据，发现天气因素对航班延误险的损失率有着重要的影响。其次，按影响程度对天气进行分类，对不同天气级别下的始发延误时间进行分布拟合，得出在不同天气级别下的始发延误时间分布。最后，建立模型对消费者的逆向选择行为进行分析，计算保险公司面临的逆向选择成本，比较发现收集数据确定分类费率的费用要低于节约的逆向选择成本，因此保险公司采用分类费率来节约逆向选择成本是可行的。

[关键词] 航班延误险；逆向选择行为；分类费率

[中图分类号] F840.69 **[文献标识码]** A

[基金项目] 本文受国家自然科学基金面上项目（11771466）、国家自然科学基金青年项目（11301559）资助。

[作者简介] 刘敬真，中央财经大学保险学院精算系副教授；林荔圆，中央财经大学保险学院硕士研究生。

一、引言

随着经济的发展，人们在旅行、工作出差时越来越多地选择飞机作为出行工具，航班延误问题也因此成为了旅客与航空公司之间爆发矛盾的一大焦点。根据2016年11月的全球机场准点率数据来看，我国的国际机场航班准点率排在世界百位之后，客流量较大的北京首都国际机场，上海虹桥国际机场和上海浦东国际机场航班准点率甚至位列倒数十位之内^①。航班延误影响了旅客的出行安排，作为承保航空公司准点起飞信用的航班延误险应运而生，以补偿旅客因航班延误产生的部分损失和安抚旅客。航班延误险可以作为主险销售，但常作为航空意外险的附加险进行销售。如今各大订票网站在机票订购环节就有是否购买航班延误险的选项。各大保险销售网站也出售航班延误险，如安联保险集团的航班延误险针对不同的延误时间有不同的赔付金额，平安财产保险公司将航班延误险作为航空意外险的附加险销售，承诺对延误4小时以上的航班赔偿300元，支付宝的保险服务中也有各种附加于航空意外险的航班延误险，这些保险在投保后的1~2日之后开始生效，不同的保险公司对于延误的定义都不尽相同，价格也不同。

目前的许多相关文献主要探讨航班延误险存在的延误定义不一、理赔困难、认知度不高等问题。本文主要针对由于天气因素给航班延误险带来的逆向选择风险和定价问题进行讨论，首先分析天气因素给航班延误险损失率带来的影响，再探讨航班延误险的分类费率定价模型和逆向选择成本。

二、文献综述

针对航班延误的原因分析，国外研究开始较早，Allan.S S等（2001）针对天气因素对航班起降、飞行过程产生的影响分析了航班延误问题。国内方面孟会芳（2015）在全面分析了造成航班延误的各种原因，并运用Cox比例风险模型

^① 数据来源：<http://www.variflight.com/>.2017.2.21.

对延误时间进行了建模。缪晨晖（2016）等用正态分布模拟了天气因素导致的航班延误的概率分布，建立了分季节排班计划的优化模型，以减少航班的波及延误。

针对航空公司的延误风险分级，国内的许多学者采取了不同的方法进行研究。石丽娜（2006）通过建立数学模型对航空公司的延误风险进行评定，将航空公司的延误风险按大小分为正常、低、中、高四种风险状态。刘雄（2012）等根据时间的多种因素，将航空公司的延误风险分为显著、中、较低、低四种风险状态。戚彦龙（2014）等运用层次分析法，以离差最大化为目标预测了每个航班的延误风险级别。在这些研究中，天气因素都被视为延误风险分级的一个关键因素。

从航班延误的原因来看，天气因素是一项影响出险概率的重要因素，并且随着科学技术的发展，人们对于天气的预测能力提高，对于航班延误的概率有了更加准确的预判。因此，对于不同天气状况下的航班延误险制定相同价格，就会产生逆向选择风险，从而给承保公司带来损失。而针对如何将天气因素引入航班延误险的定价模型，进行费率分类的文献却少之又少。因此，本文主要探讨如何将天气因素引入航班延误险的定价模型，从而进一步分析进行费率分类的必要性和可行性。

三、天气因素与损失率的相关性检验

（一）数据收集

中国民用航空局的数据显示，2016年影响航班正常率的因素中，天气因素占到了56%^①。因此，首先采集历史天气数据^②和航班起飞和降落的具体时间数据^③。在考虑了数据量、航班、机场等因素产生的交叉影响之后，本文选取了西安咸阳国际机场、三亚凤凰国际机场、哈尔滨太平国际机场、广州白云国际机

① 数据来源：<http://www.chinanews.com/cj/2017/03-03/8165032.shtml>。

② 数据来源：<http://www.tianqihoubao.com.2016.12.31>。

③ 数据来源：<http://zh.flightaware.com.2016.11.7>。

场、长沙黄花国际机场、厦门高崎国际机场、北京首都国际机场、上海虹桥国际机场、上海浦东国际机场9个机场之间12条航线上的61个航班的7136条航班数据进行分析。数据来自Flightaware网站的航班的始发地、目的地、出发时间、到达时间、飞行时间等数据。根据各大订票网站上显示的计划时间来确定飞机原定出发时间，从而再根据实际出发时间来计算航班延误时间。

文献显示，雨季对于航班延误的影响要远远大于冬季下雪对于航班延误的影响，同时再对2016年11月至2017年3月上述航班的数据和是否下雪进行相关性分析，得出的 p 值为0.6，则可以认为航班延误与下雪天气没有太大的相关性。因此本文取天气因素中是否下雨这一项和航班延误的相关性因素进行了分析。同时，考虑到天气因素具有季节性，因此本文选取了2016年7月8日至2016年11月7日这一雨季的相关数据。

（二）数据分析

本文取中国平安财产保险股份有限公司平安旅行附加个人航班延误保险进行分析，条款表明，延误时间指：（1）自飞机原定出发时间起至飞机实际起飞时间，或至航空公司安排的替代性交通工具的出发时间为止；（2）自飞机原计划到达目的地时间起至飞机实际到达目的地时间，或至航空公司安排的替代性交通工具到达目的地时间为止。保险公司对航班延误时间超过4小时以上的旅客一次性赔偿300元^①。因此，本文将飞机原定出发时间起和飞机实际起飞时间之差与飞机原计划到达目的地时间起和飞机实际到达目的地时间之差的最大值超过4小时的航程以及被取消了的航班都定义为延误，进而对所有的数据进行判断，统计出延误的航班数、遇到雨天的航班数、遇到雨天且延误的航班数，从而计算出不考虑天气因素时总延误率和雨天延误率，进行对比。

通过统计，在7136条数据中，有2973条为雨天，其中，共有314次航班延误，雨天延误有115次，雨天延误率比总延误率高0.813%。

进行卡方检验，判断雨天与航班是否延误是否具有相关性。

^① 中国平安财产保险股份有限公司平安旅行附加个人航班延误保险条款，[2015]附157号。

表1 卡方相关性检验

单位：次

	延误	不延误	合计
有雨	155	2818	2973
无雨	159	4004	4163
合计	314	6822	7136

$$T_{11} = \frac{(a+c)(a+b)}{a+b+c+d} = \frac{314 \times 2973}{7136} = 130.8187$$

$$T_{12} = \frac{(b+d)(a+b)}{a+b+c+d} = \frac{6822 \times 2973}{7136} = 2842.181$$

$$T_{21} = \frac{(a+c)(c+d)}{a+b+c+d} = \frac{314 \times 4163}{7136} = 183.1813$$

$$T_{22} = \frac{(b+d)(c+d)}{a+b+c+d} = \frac{6822 \times 4163}{7136} = 3979.8$$

$$\chi^2 = \sum \frac{(T-A)^2}{T} = 8.014609$$

得 $p=0.00464 < 0.05$ ，所以我们有理由相信航班延误与是否下雨有极大的相关性。

四、逆向选择成本分析

非寿险产品在进行费率厘定时，为了消除逆向选择行为的影响，通常要考虑许多不同的风险分级变量，用于获得相对同质的风险子集，使定出的保险费率可以反映出真实的风险状况。在选择风险分级变量的时候要遵循精算标准、经营标准、社会标准、法律标准四大标准（孟生旺，2004）。总而言之，选择的分级变量要便于区分，为大众所接受，同时要保证分组后每组内都要有足够的保单数据以符合大数定律，并且使每组的费率有所区别。

从上文的分析来看，对于航班延误险的定价，天气因素是一个符合四大标准的风险分级变量。首先，天气因素是可以客观判断的，并且国家有着对天气

状况分类的标准，也是大众容易获得的信息。其次，影响航班延误的天气状况并不像地震等巨灾那样发生概率极低，针对天气状况进行分类后，各组依然会有充足的数据。最后，通过上文的分析，天气因素确实是航班延误险损失率的重要影响因素。因此本文将把天气因素作为航班延误险定价的一个风险分级变量，进而讨论逆向选择行为带来的成本。

(一) 始发延误时间分布

Muller (2002)曾就美国十大机场的数据得出航班的始发延误时间符合指数分布。由于所得数据有限，本文使用北京首都国际机场9个航班的2016年7月至11月的始发延误时间和对应时间段的北京的历史天气进行分析。根据不同天气因素对航班延误的影响不同，将这一时间段所有的天气状况分为三个等级，第一级为晴、多云、阴；第二级为小雨、小到中雨、中雨、阵雨；第三级为雷阵雨、大雨、暴雨、霾。同一天中有两种天气状况的，以较高级别的天气状况来划分当日的天气级别，因为投保人更加关心天气会有多恶劣。根据天气级别，将所得的始发延误时间的数据分为三类，分别进行个体分布识别。三组数据的直方图见图1、图2和图3。

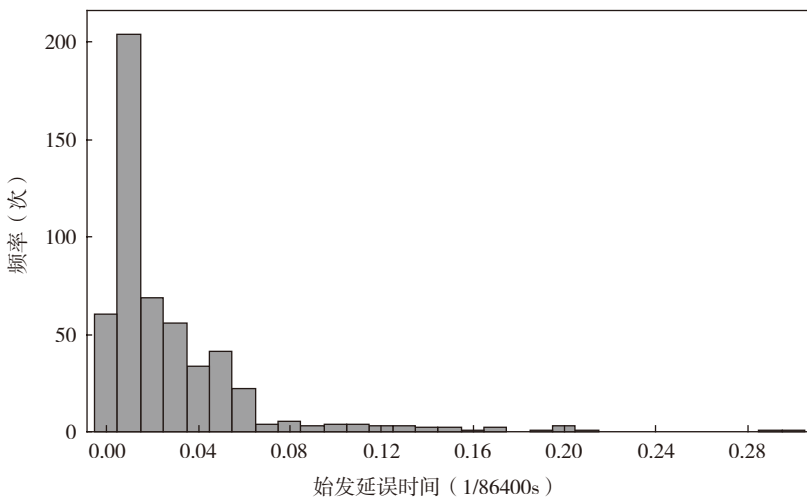


图1 一级始发延误时间直方图

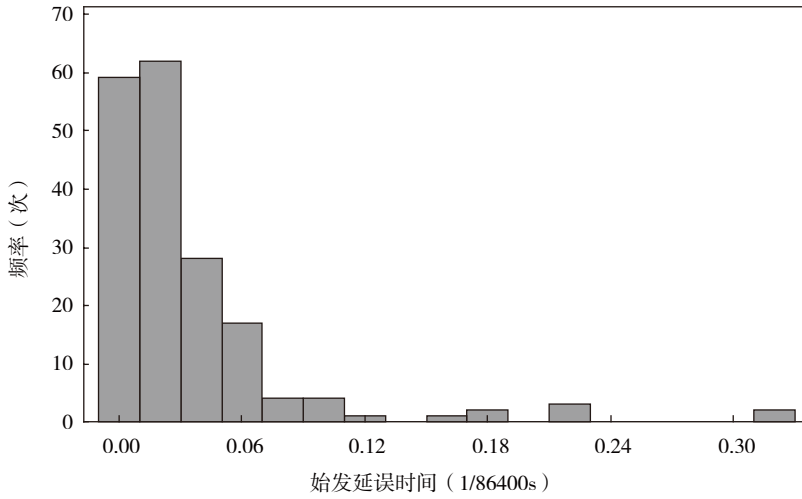


图2 二级始发延误时间直方图

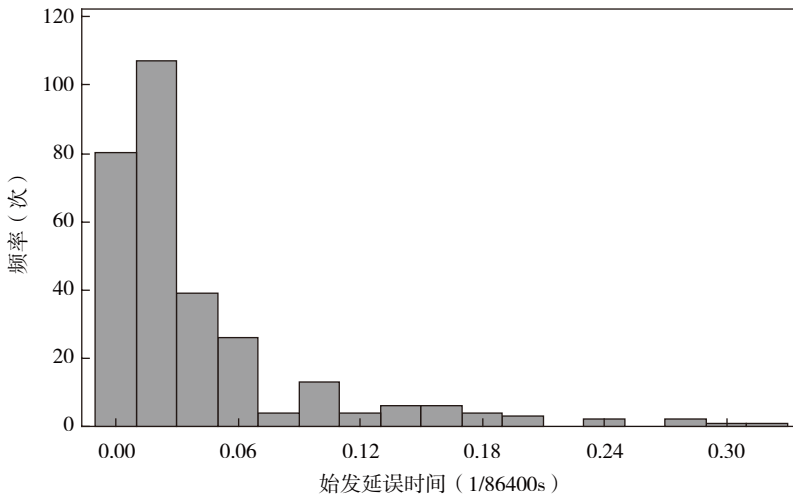


图3 三级始发延误时间直方图

通过直方图，猜测始发延误时间可能服从对数正态分布，为了消除数据中存在的零对分布拟合的限制，需要剔除零后对三组数据采用不同的分布进行拟合。三组数据中零出现的概率都约为0.05。

表2 一级始发延误时间拟合优度检验

分布	<i>AD</i>	<i>P</i>	极大似然比 <i>P</i>
正态	50.982	<0.005	
对数正态	1.349	<0.005	
三参数对数正态	1.361	*	0.955
指数	8.433	<0.003	
三参数对数Logistic	1.688	*	0.052

表3 二级始发延误时间拟合优度检验

分布	<i>AD</i>	<i>P</i>	极大似然比 <i>P</i>
正态	22.471	<0.005	
对数正态	0.224		0.822
三参数对数正态	0.221	*	0.852
指数	3.758	<0.003	
三参数对数Logistic	0.269	*	0.222

表4 三级始发延误时间拟合优度检验

分布	<i>AD</i>	<i>P</i>	极大似然比 <i>P</i>
正态	31.977	<0.005	
对数正态	0.502		0.204
三参数对数正态	0.513	*	0.772
指数	7.44	<0.003	
三参数对数Logistic	0.501	*	0.142

可以看出，三参数对数正态分布是最好的用于拟合三组数据的分布。

表5 始发延误时间参数估计

延误时间	位置	尺度	阈值
一级始发延误时间	-4.08452	1.03192	-0.00001
二级始发延误时间	-4.03629	1.11467	0.00008
三级始发延误时间	-3.84101	1.17351	-0.00011

因此始发延误时间的分布为零和三参数对数正态分布的混合，以0.05的概率取值为零。得到了不同天气级别下的始发延误时间的分布后，就可以根据保险公司的赔付规则进行统一定价或费率分类，再根据投保人的投保行为计算出在统一定价的情况下保险公司要面临的逆向选择行为带来的成本。

(二) 逆向选择成本分析

从上文的分析来看，天气因素对于航班延误险的出现概率的确有影响。但是目前市场上的航班延误险的价格并没有根据天气情况进行改变的机制，即使天气预报报道了航班当日存在恶劣天气，也没有提高保费的相应措施。而投保人又很容易通过天气预报得知乘机当天的天气状况，这样就很容易产生逆向选择行为。投保人若提前得知当天天气晴好，购买航班延误险的意愿就比较低；若提前得知出行当日有不利于航班起降的天气状况，就会有较大的可能购买航班延误险。这使在天气状况不好的时候，保险人会获得较多高风险的保单，而在天气状况良好的时候保险人却只能获得少量风险较低的保单，从而提高保险人额外的成本。

本文假设天气预报是准确的，且乘坐航班的每名乘客是否投保的概率是相同的。

航班延误时间在第*i*级天气状况下的分布为 $f(t|\xi=i)$ ， $P(\xi=i)=q_i$ 。

对于一个航班上的任意一名乘客而言，设 $X = \begin{cases} 1 & \text{投保} \\ 0 & \text{不投保} \end{cases}$ ， $P(X=1|\xi=i)=p_i$ ，

则有 $E(X) = \sum_{i=1}^3 p_i q_i$ 。

保险公司的赔付为 $h(t)$ ，若保险公司在不同的天气状况下收取统一净保费 Q ，则有 $Q = E[h(t)] = \sum_{i=1}^3 E[h(t)|\xi=i]P(\xi=i) = \sum_{i=1}^3 Q_i q_i$ ，其中 $Q_i = E[h(t)|\xi=i]$ 。

则对于乘坐航班的每名乘客而言，保险公司收取的保费为 QX ，赔付为 $h(t)X$ ，保险公司收益为 $QX-h(t)X$ ，故有 $E[h(t)X] = \sum_{i=1}^3 Q_i q_i p_i$ 。

消费者存在逆向选择行为，所以若 $E[h(t)|\xi=i] \geq E[h(t)|\xi=j]$ ，即 $Q_i \geq Q_j$ ，则有 $p_i \geq p_j$ ，从而 $E[QX-h(t)X] = (Q_2 - Q_1)(p_1 - p_2)q_1q_2 + (Q_3 - Q_1)(p_1 - p_3)q_1q_3 + (Q_3 - Q_2)(p_2 - p_3)q_2q_3 \leq 0$

因此当乘客存在逆向选择时，保险公司面临平均收益小于零的状况，这就是保险公司因为乘客逆向选择产生的成本，如果保险公司对不同的天气状况收取不同的保费，可以降低甚至消除这一成本。在给出 $h(t)$ 后，通过数据可以统计出 Q_i, q_i 。取 $h(t)=300, t>0.1667$ ，即如果始发延误时间大于4小时，保险公司一次赔付300元。由此计算 $|(Q_2 - Q_1)(p_1 - p_2)q_1q_2 + (Q_3 - Q_1)(p_1 - p_3)q_1q_3 + (Q_3 - Q_2)(p_2 - p_3)q_2q_3|$ （见表6）。

表6 参数计算结果

i	Q_i	q_i
1	3.746325	0.522642
2	6.283395	0.180189
3	11.49519	0.29717

在极端情况下，假设 $p_1 = p_2 = 0, p_3 = 1$ ，那么有

$$|(Q_2 - Q_1)(p_1 - p_2)q_1q_2 + (Q_3 - Q_1)(p_1 - p_3)q_1q_3 + (Q_3 - Q_2)(p_2 - p_3)q_2q_3| = 1.72$$

即乘客只有在天气状况处于第三级时才会购买保险，且一定会购买保险。这是一种十分极端的逆向选择行为，在这种状况下，保险公司采用统一费率将会面临每名乘客产生的1.72元的逆向选择成本，这也说明在上述假设下，逆向选择成本最高不会超过1.72元。但相比航班延误险的低价位和高销量，这一成本也十分可观。为了消除这一逆向选择成本，保险公司需要收集数据对 Q_i, q_i, p_i 进行分析进而制定分类费率，其中 Q_i, p_i 可以通过公司业务积累的数据进行分析，而 q_i 可以通过分析天气数据得到，这些数据都是简易可得的，分析模型也相对简单，因此相较于所节省的逆向选择成本而言，这些数据的收集和分析费用相对较低。

五、结论

本文在参考了前人大量关于航班延误险和分类费率的资料的基础上，对航班延误险与天气状况这一因素进行了更深入的分析。通过收集大量的航班起降数据进行整理，用卡方列联表检验了天气状况与航班延误是否具有相关性。结

果表明,是否下雨和航班是否出现延误有极大的相关性,基于此,将天气因素引入定价模型。按影响程度不同,将天气状况分为三个等级,对始发延误时间用三参数对数正态分布拟合,得到个体分布后即可根据保险条款上的理赔规则计算纯保费。由于在不同的天气状况下,航班延误险存在不同的出险概率,旅客会有逆向选择行为,在天气状况差的时候会以较高的概率购买航班延误险,在天气状况良好的时候购买航班延误险的概率较低,给承保人带来了隐性成本。通过模型分析可以发现对于一份延误4小时赔付300元的保单,这一成本达到每单1.72元,而保险公司可以通过实行分类费率来降低这一成本。制定分类费率所需的数据可以通过以往销售的保单来积累,所耗费用要远小于节约的成本。因此保险公司可以通过制定分类费率来降低航班延误险的逆向选择成本。

参考文献

- [1] 冯杰,吴薇薇,刘钟佳文等.基于Logit模型的航班延误险购买行为研究[J].华东交通大学学报,2015,32(5):44-50.
- [2] 刘雄,邵维亮,张硕.基于灰色多层次方法的航班延误风险评价[J].西安航空技术高等专科学校学报,2012,30(3):27-29.
- [3] 孟会芳.基于统计分析的航班延误等级划分研究[D].南京:南京航空航天大学硕士学位论文,2015.
- [4] 缪晨晖,朱星辉,高强.天气因素导致的航班延误问题研究[J].航空计算技术,2016,46(3):17-19.
- [5] 孟生旺.保险定价:经验估费系统研究[D].中国人民大学硕士学位论文,1998.
- [6] 戚彦龙,王琨,朱星辉,郑松林.基于离差最大化原理的航班延误风险分级研究[J].武汉理工大学学报,2014,38(1):162-166.
- [7] 石丽娜.多等级模糊评价方法在航班延误中的应用[J].上海工程技术大学学报,2006,20(3):276-279.
- [8] Allan S S, et al.. Analysis of Delay Causality at Newark International Airport[R].Santa Fe:4th USA/Europe Air Traffic Management R&D Seminar,2001.
- [9] Mueller, et al.. Analysis of Aircraft Arrival and Departure Delay Characteristics:AIAA 2002-5866[R].Los Angeles,2002.