

保险精算中的预测模型实战考量

安永（中国）企业咨询有限公司 总监 姚佶
2017年12月

■ ■ ■
The better the question. The better the answer.
The better the world works.





目录

1

预测模型简介

2

保险精算中的应用实例

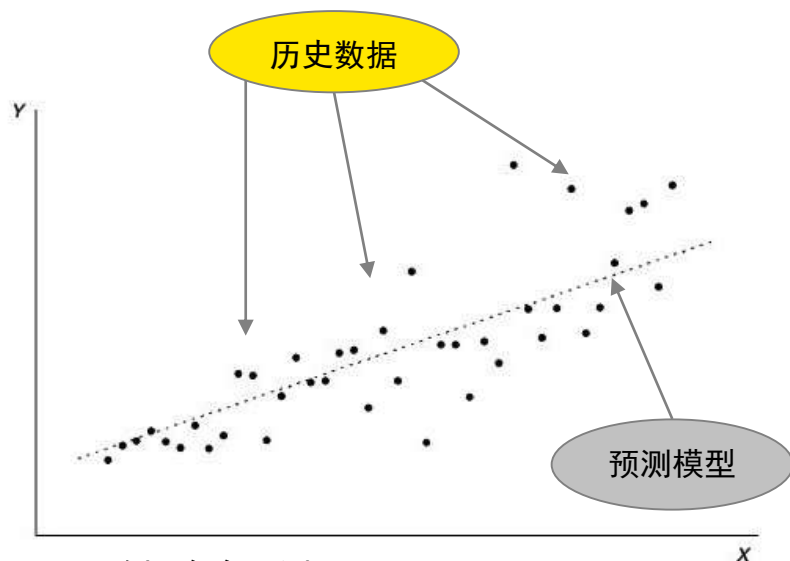
3

预测模型应用实践

4

小结和提问

预测模型简介



示例：气象预测



示例：路况预测



1

预测模型是什么

预测模型是一种基于历史数据并结合对未来趋势的判断而建立的模型。

预测模型可以预估未来某个指标的发展趋势，也可以基于某些已知的数据去建立计算模型，计算某些量变化后的结果。

2

预测模型有哪些类型

1、定性预测方法

根据人们对系统过去和现在的经验、判断和直觉进行预测，其中以人的逻辑判断为主，仅要求提供系统发展的方向、状态、形势等定性结果。该方法适用于缺乏历史统计数据的系统对象。例如常见的气象预测。

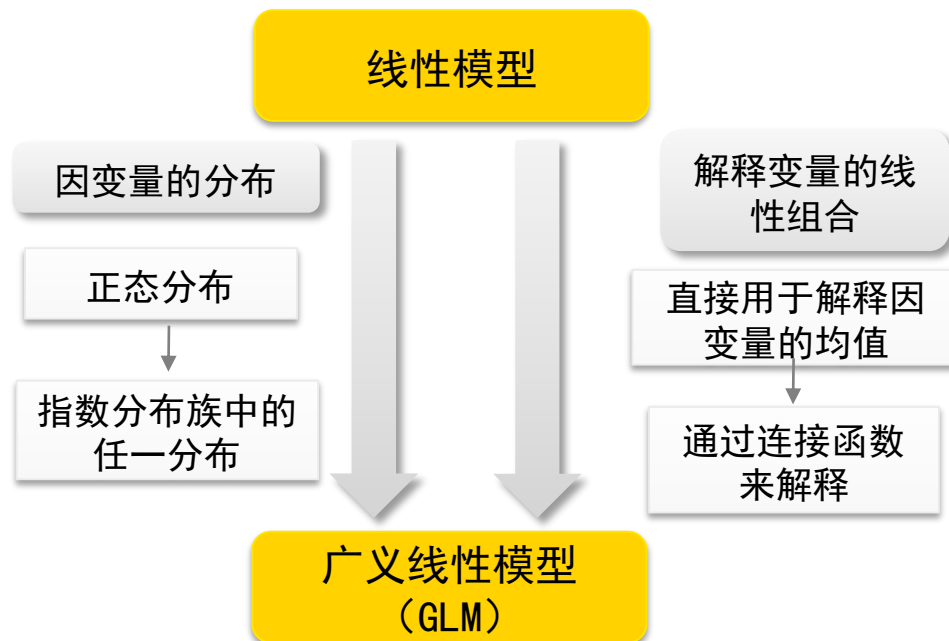
2、时间序列分析

根据系统对象随时间变化的历史资料，只考虑系统变量随时间的变化规律，对系统未来的表现时间进行定量预测。例如对路况拥堵情况的预测。

3、因果关系预测

系统变量之间存在某种前因后果关系，找出影响某种结果的几个因素，建立因与果之间的数学模型，根据因素变量的变化预测结果变量的变化，既预测系统发展的方向又确定具体的数值变化规律。

常见的预测模型及其功能评价

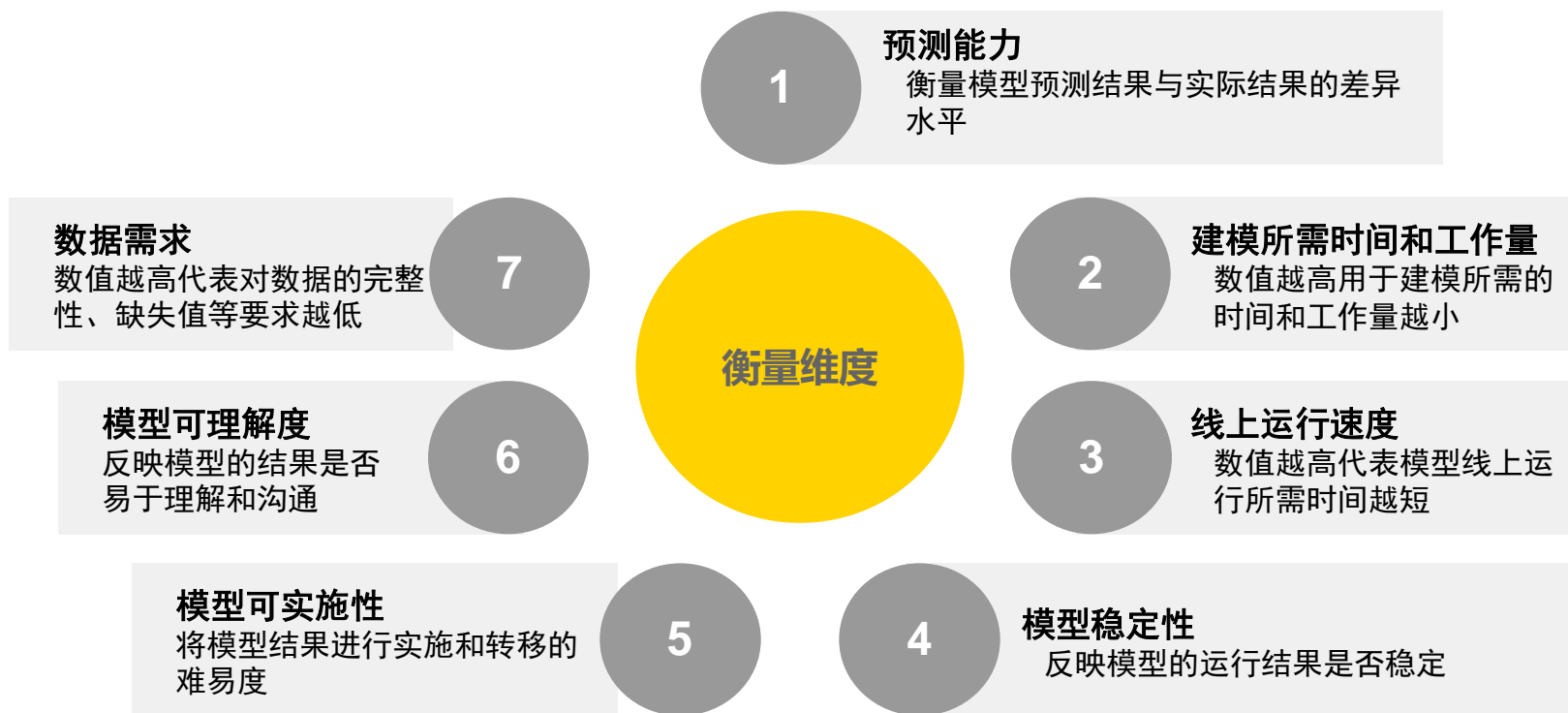


- 一般的预测函数形式为：

$$f(\underline{x}) = g^{-1}(\underline{x} \cdot \underline{\beta})$$

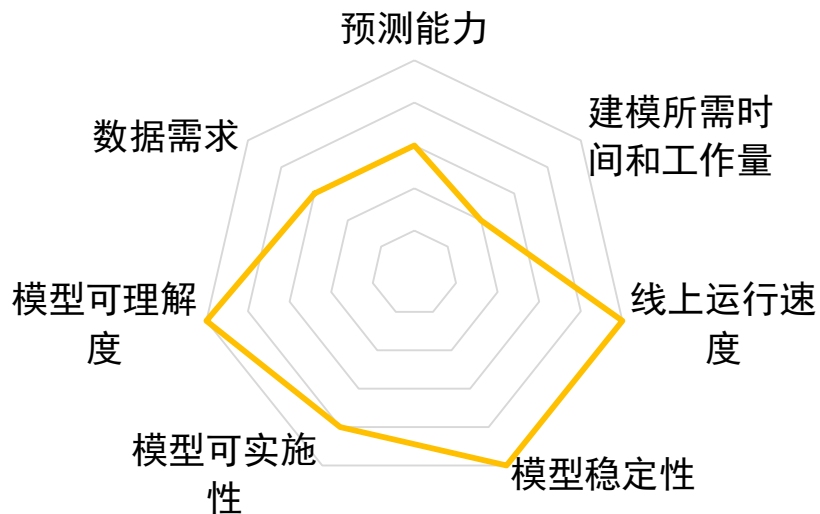
- 其中 β 是通过最小化损失函数 $L(\beta | X, y)$ 测算得到

我们可以从7个维度衡量模型的性能



GLM 模型的特点

GLM模型评价

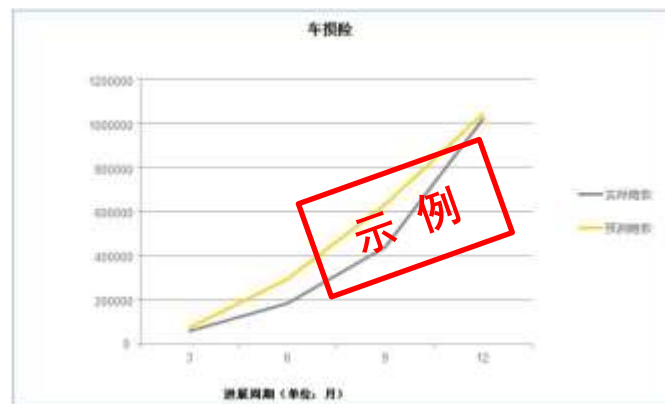


- ▶ GLM模型的预测能力低于GBM，但相对较容易理解，可实施性方面也更强。

梯度提升机（GBM）模型评价



公司实际赔付状况与广义线性模型预测结果对比示例





目录

1

预测模型简介

2

保险精算中的应用实例

3

预测模型应用实践

4

小结和提问

预测模型在保险精算领域的应用场景



风险保费模型根据对历史经验分析和建模，充分挖掘大数据的价值

历史赔案数据



驾驶员

- 年龄
- 性别
- 婚姻状况
- 子女数目
- 驾照年龄
- 驾照类别
- 居住区域
- 驾驶纪录
- 是否业主
- 疾病/残疾
- 职业
- 教育程度
- 是否吸烟
- 其他驾驶员



车辆

- 厂牌/车型/年份
- 马力/重量
- 最高时速
- 排气量
- 新车购置价
- 使用性质
- 泊车地点
- 车龄
- 主驾驶员是否车主
- 是否为主要用车
- 家庭车辆数
- 年里程数
- 防盗设备
- 是否改装



其它

- 业务渠道
- 免赔额
- 责任限额
- 付款方式
- 无赔优待
- 防御性驾驶训练
- 多险别折扣
- 出单机构
- 多车折扣
- 是否拥有住宅
- 防抱死制动折扣
- 是否曾被拒保
- 保单期限
- 新单提前投保折扣

风险保费模型

目的：

清晰地展示业务品质发生变化的原因，对各因素影响程度的量化

预测指标：

• 满期赔付率、历年制赔付率、出险频度、案均赔款；

分析维度：

保单年月（到周，保费到日）、机构、险种、客户类别、交商同保、渠道

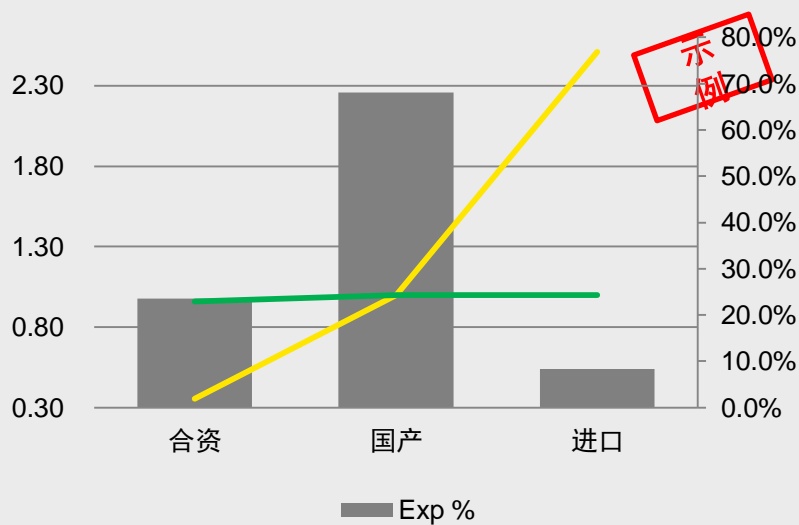
分析方法：

将分析指标设定为应变量，将维度设置为自变量，通过GLM或其他机器学习方法建立模型，对分析指标进行预测

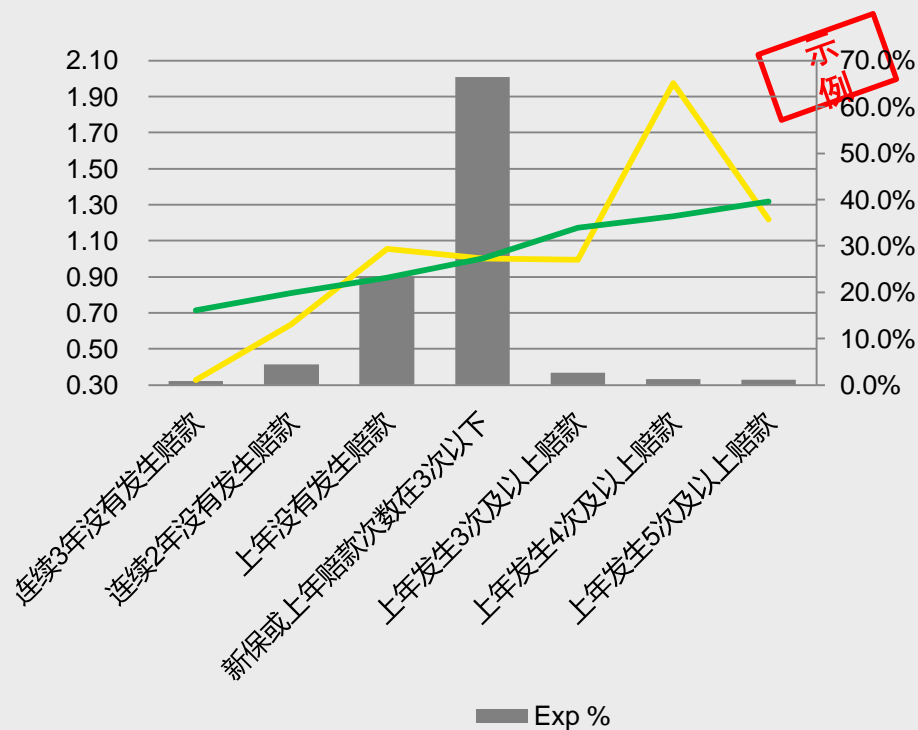
通过风险保费模型可以展示因子各类别之间的相对系数，从而帮助公司理解有重要影响的因子

(国内某公司) 模型结果展示 - GLM vs. One Way——以家庭自用车车损险为例

风险成本相对数 - 是否进口车

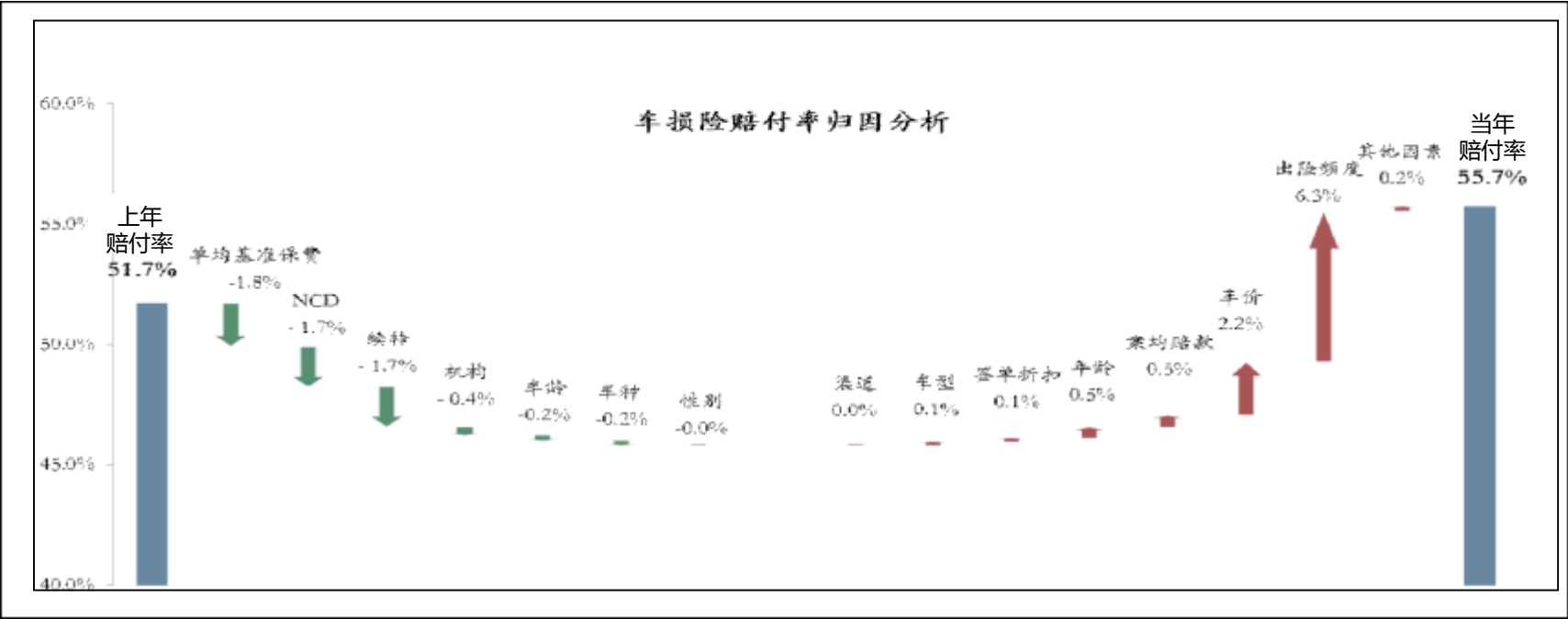


风险成本相对数 - 过去三年出险次数



基于预测模型，可以对业务指标变化进行归因分析，测算各因素对于预测指标的影响程度

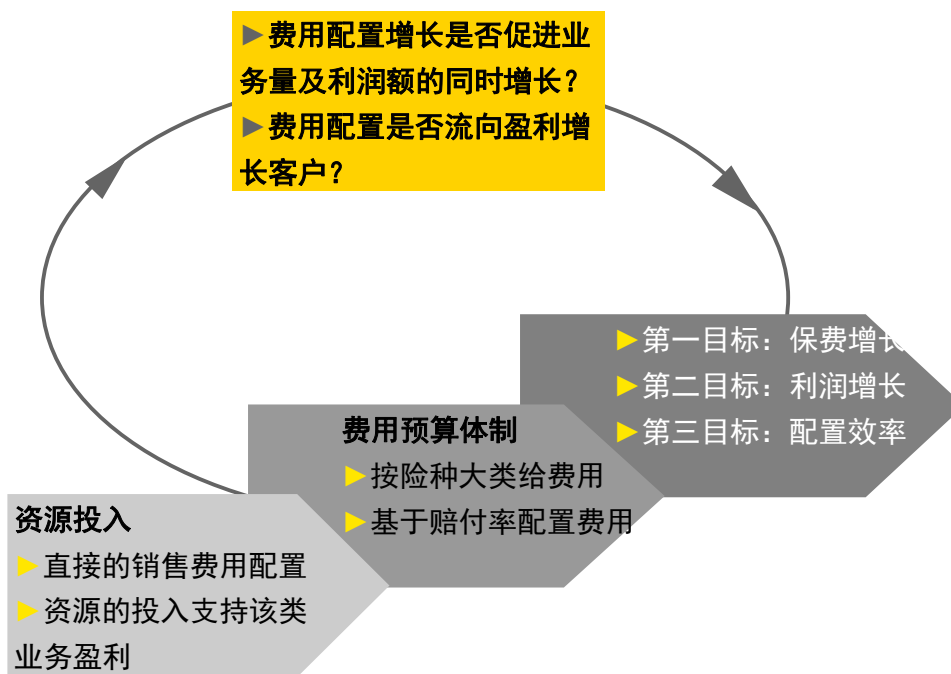
各因素的归因分析可通过“瀑布图”进行展示，如下图，可以看到出险频率对赔付率的增长影响最大



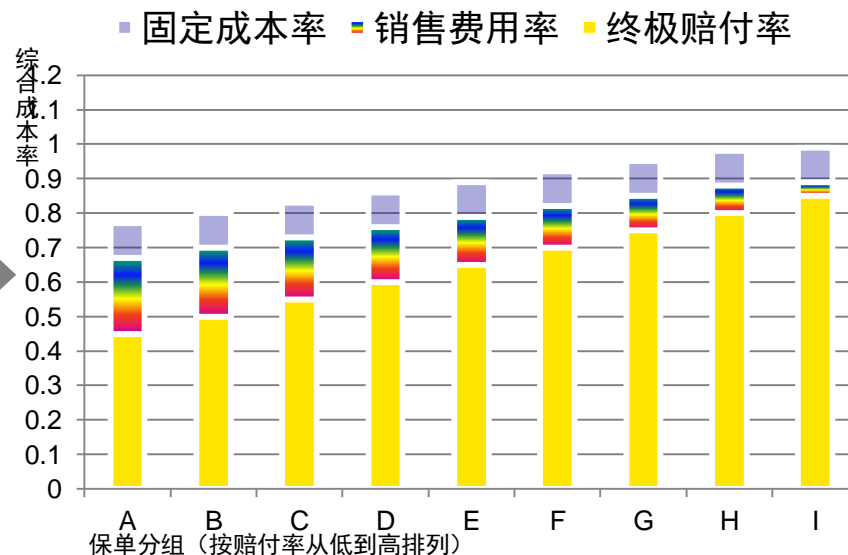
通过预测模型合理设置费用率，有效配置公司资源，促进保费和利润的增长

两率（赔付率与费用率）联动的特点和作用

销售费用的配置需要促进保费收入的达成以及配置效率最优



如下图，公司可以进行销售费用差异化配置，赔付率低的业务配置高于市场费用，赔付率适中的业务跟随市场费用，赔付率高的业务不超过最大可投放费用。



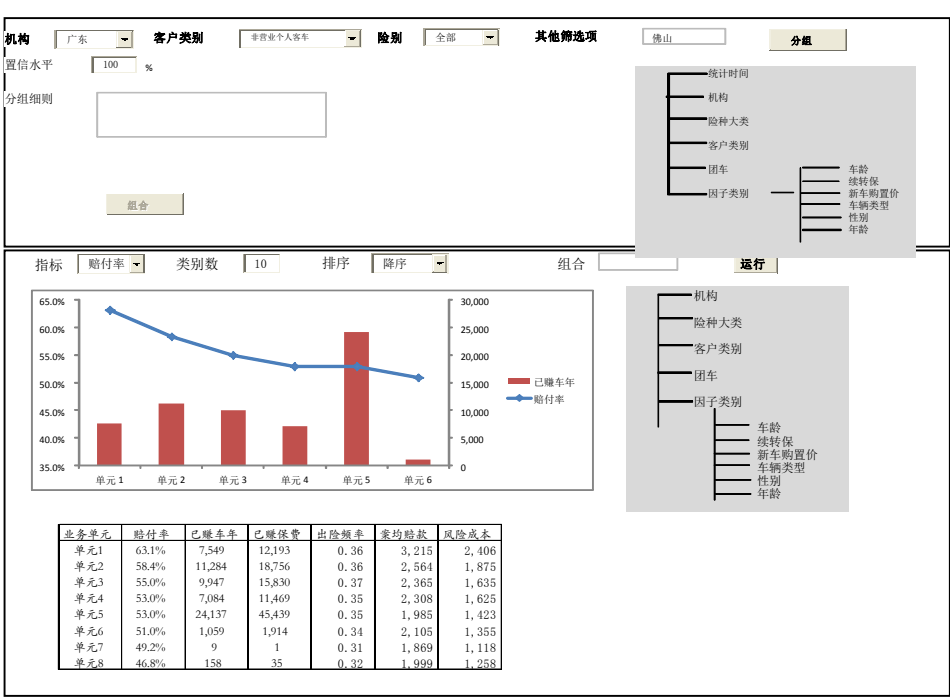
预测模型对于业务风险识别与筛选具有指导性作用，可以根据模型结果和战略要求设置相应的保费和核保规则

工作框架

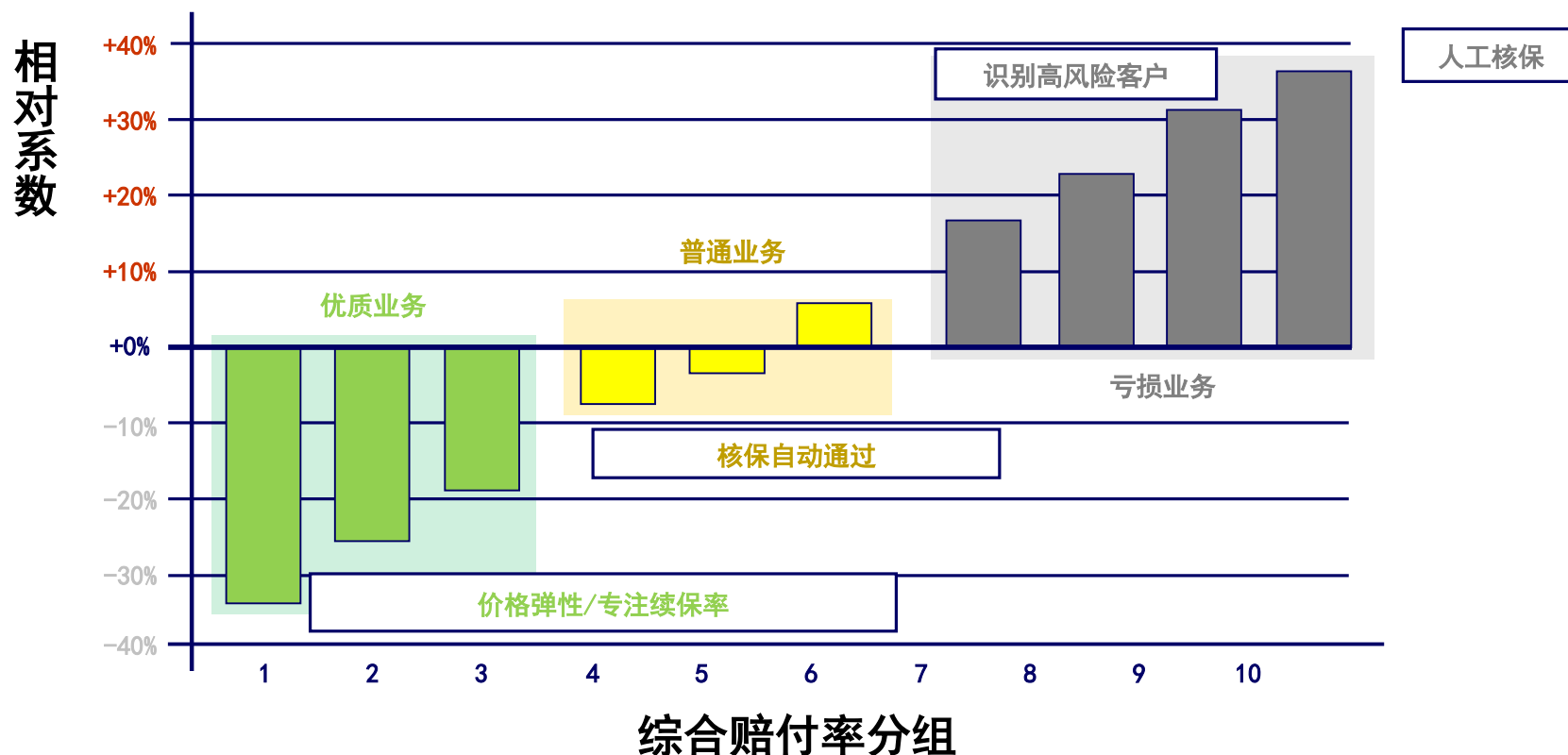
- 目的：
将业务细分，精确定位优质业务与亏损业务，并调整保费和核保规则
- 分析指标：
- 赔付率、出险频度、案均赔款、终极预测赔付率、风险成本；
- 分析维度：
- 统计时间（年度、月度、周）
 - 机构（包括三级机构）
 - 险种（交强、商业、车险）
 - 客户类别（非营业个人客车、非营业货车等）
 - 险别（车损险、三者险、盗抢、司机、乘客、不计免赔）
 - 团车（是、否）
 - 因子类别（车龄、年龄、性别、是否续保、新车购置价、车型、座位数、吨位数等定价因子）

- 分析方法：
1. 将某类业务按照一定规则，分解成业务单元
 2. 针对各个业务单元组的赔付率进行预测，进而调整价格以达到预期目标
 3. 也可以根据预期赔付率，确定核保规则

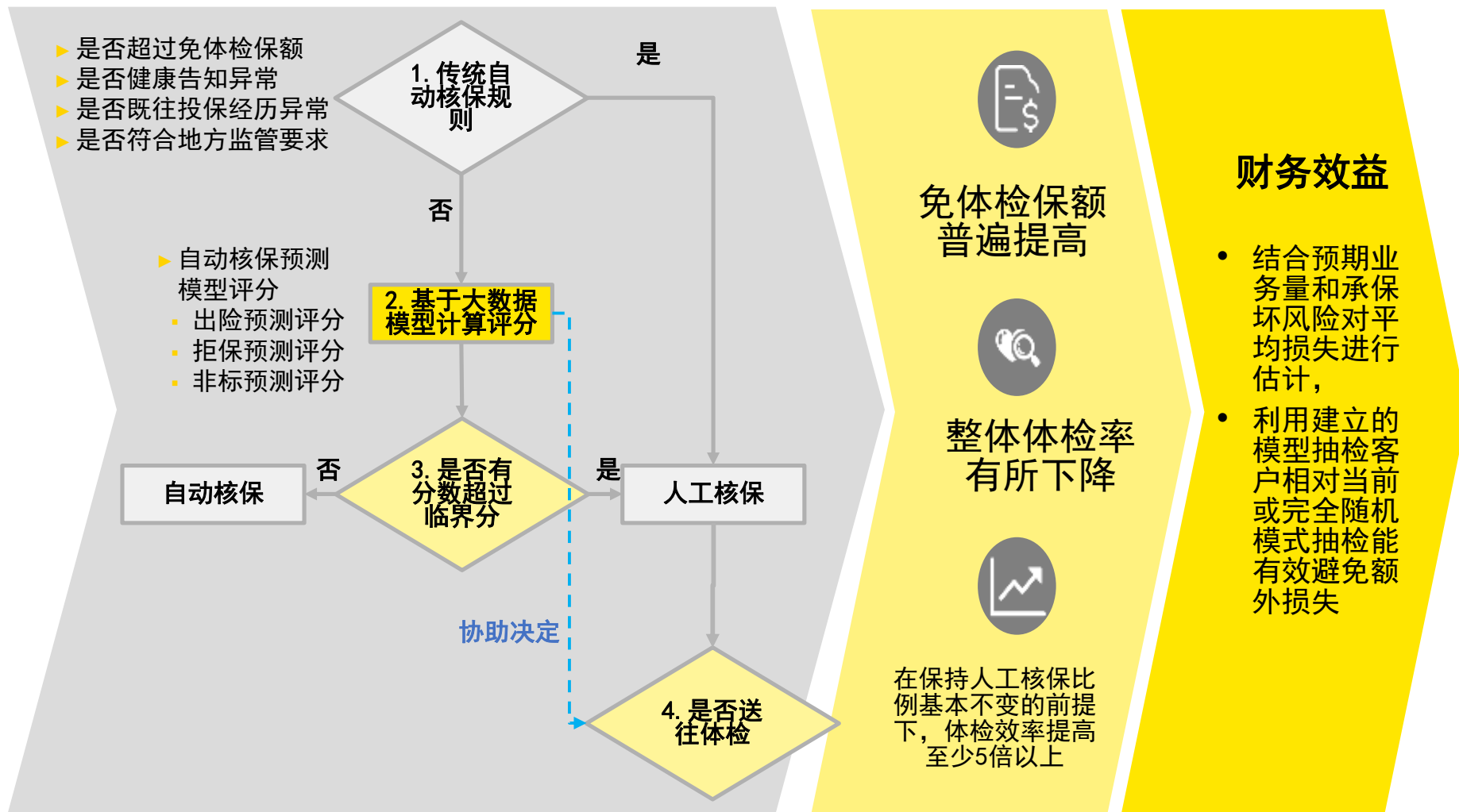
示例



公司可以根据预测模型的结果识别业务风险的高低，并对相应的业务组群采取合理的措施，为业务部门设置核保规则提供参考



根据经验数据建模，然后在寿险核保流程中嵌入模型，可以降低整体体检率，优化核保流程和效率。



基于前述的归因分析可筛选预测模型的自变量风险因素，并对未来的指标（例如赔付率）进行预测

工作框架

目的：
对当前业务在未来的赔付水平进行较为精准的预测

预测指标：
•终极赔付率、进展赔付率、核保人赔付率、边际贡献率、利润率；

分析维度：
•统计时间（年度、月度、周）
•机构（包括三级机构）
•险种（交强、商业、车险）
•客户类别（非营业个人客车、非营业货车等）
•险别（车损险、三者险、盗抢、司机、乘客、不计免赔）
•团车（是、否）
•因子类别（车龄、年龄、性别、是否续保、新车购置价、车型、座位数、吨位数等定价因子）

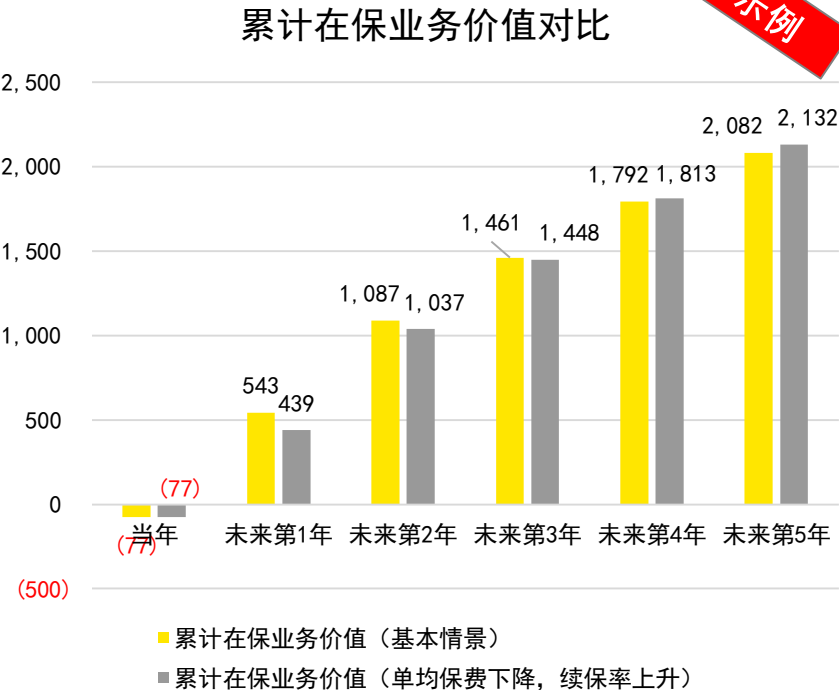
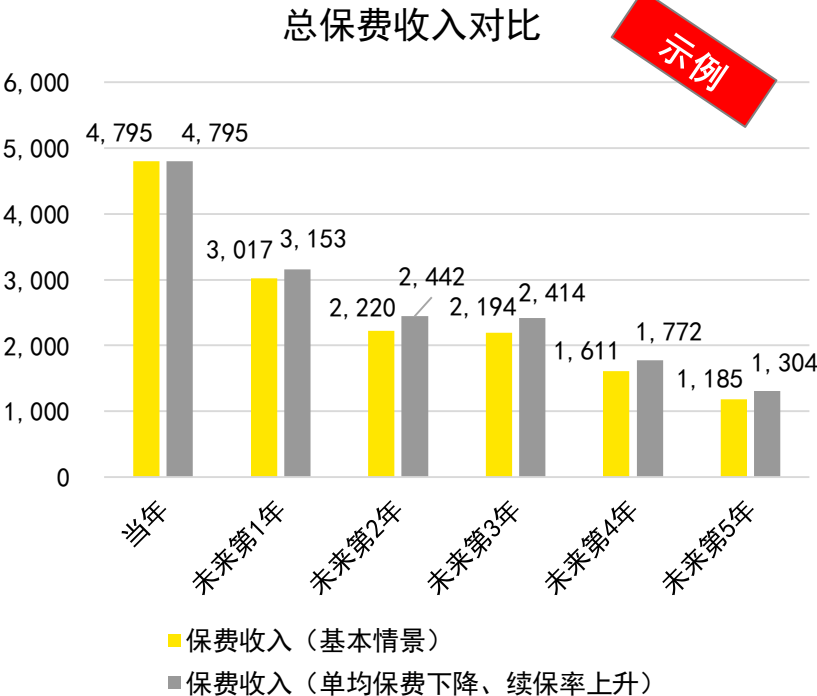
分析方法：
1、从多个维度，基于不同预测方法对赔付率等指标进行预测
2、可根据预测结果评估、调整核保人赔付率，指导核保规则
3、对多种预测方法预测出的赔付率结果进行比对
4、将进展赔付率的预测结果留存，以期与下一时点的实际赔付率进行比较，校验预测的准确性

示例

细分类别	已赚车年	已赚保费	实际赔付率	预测赔付率1	预测赔付率2	预测赔付率3	核保人赔付率
A0-1年	241,373	45,439	53.0%	54.6%	55.6%	52.4%	
B1-2年	112,838	18,756	58.4%	54.9%	61.3%	57.8%	
C2-3年	75,495	12,193	63.1%	67.6%	66.3%	62.5%	
D3-5年	99,472	15,830	55.0%	56.1%	57.8%	54.5%	
E5-8年	70,841	11,469	53.0%	54.6%	55.7%	52.5%	
F8-12年	10,590	1,914	51.0%	48.5%	53.6%	50.5%	
G12年以上	158	35	46.8%	47.2%	49.1%	46.3%	
H信息缺失	9	1	49.2%	49.7%	51.7%	48.7%	

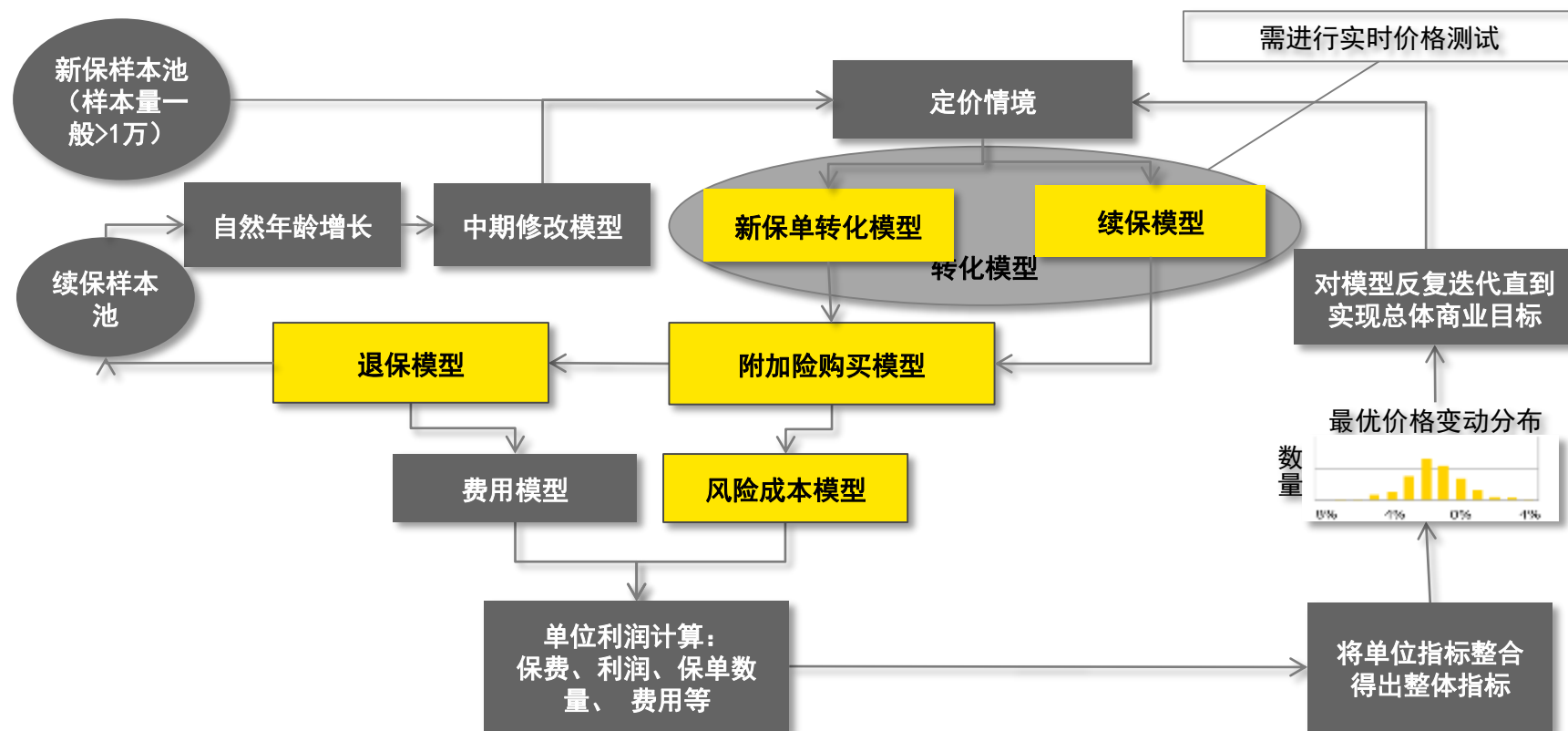
续保优化不仅可以提升当期续保率，而且可以增加公司业务的长期价值

- ▶ 在以下测算示例中，在客户生命周期的未来第1年虽然单均保费下降，但由于业务的续保率提高，总保费收入有所上升，尽管第1年利润有所下降，但考虑到后续业务续保率的上升，整体业务价值将有所提高
- ▶ 相比当期价值（利润），客户生命周期价值从理念上与公司长期价值更加契合。特别是在目前行业整体风险成本下降的趋势下，在展业时适当考虑客户生命周期价值，既能够为更多的消费者提供保障，也能够对行业规模和效益产生贡献，实现双赢。



预测模型计算最优价格，实现利润本年最优化

- 构建最优价格动态模型，通过机器学习不断模拟各项指标变化
- 之后加入预测模型，引入退保、保单修改和附加险收入等较为复杂的因素



预测模型计算最优价格，实现客户终身价值最大化

客户价值度量了客户在当前以及未来可能为公司带来的收益

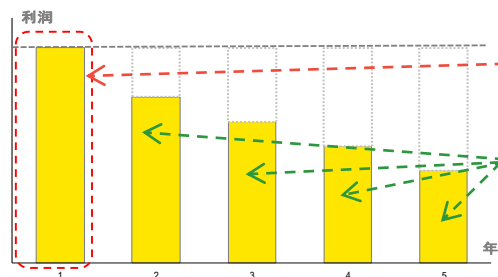
计量方法

客户价值 = 客户现值 + 附加险价值

$$\text{客户现值} = \sum_{j \in J} \int_t p_x^j W_t^j v^t dt$$

t: 时间 ρ: 续保率

W: 利润 v: 折现率



示例

▶ 每个现有客户在首个保单年的利润

▶ 考虑客户未来续保行为，将给公司带来利润

逐客户首年利润

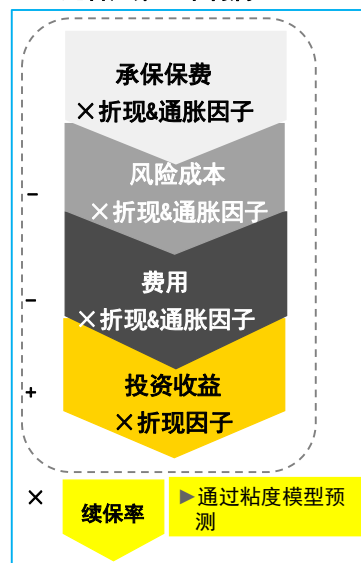
逐客户第二年利润

逐客户第三年利润

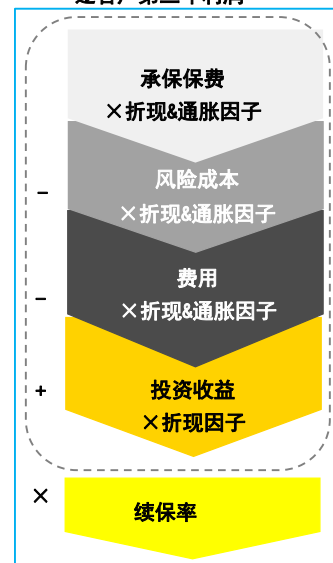
客户现值 =



+



+



+

...

理赔欺诈检测的方法



人工卷宗审核

- ▶ 根据金额或者频率等标准，对高风险赔案进行人工卷宗审核。
- ▶ 这种方式对识别欺诈非常有效，但所需资源投入过大，专业人才较为缺乏，且容易错过模式性欺诈。



基于规则的系统识别

- ▶ 在系统中部署已定的规则，将触犯规则的案件推送到人工审核。
- ▶ 这种方式可以有效查找已知的欺诈类型，但无法针对新的欺诈案件自动更新，而且非常容易被欺诈者得知后迅速失效。



替换零件回收确认

- ▶ 将替换下的零配件进行回收确认，用以减少虚假的零配件索赔。
- ▶ 这种方式对零件类的索赔会有一定的效果，但是成本高昂，不适合大范围应用。



人工实地审核

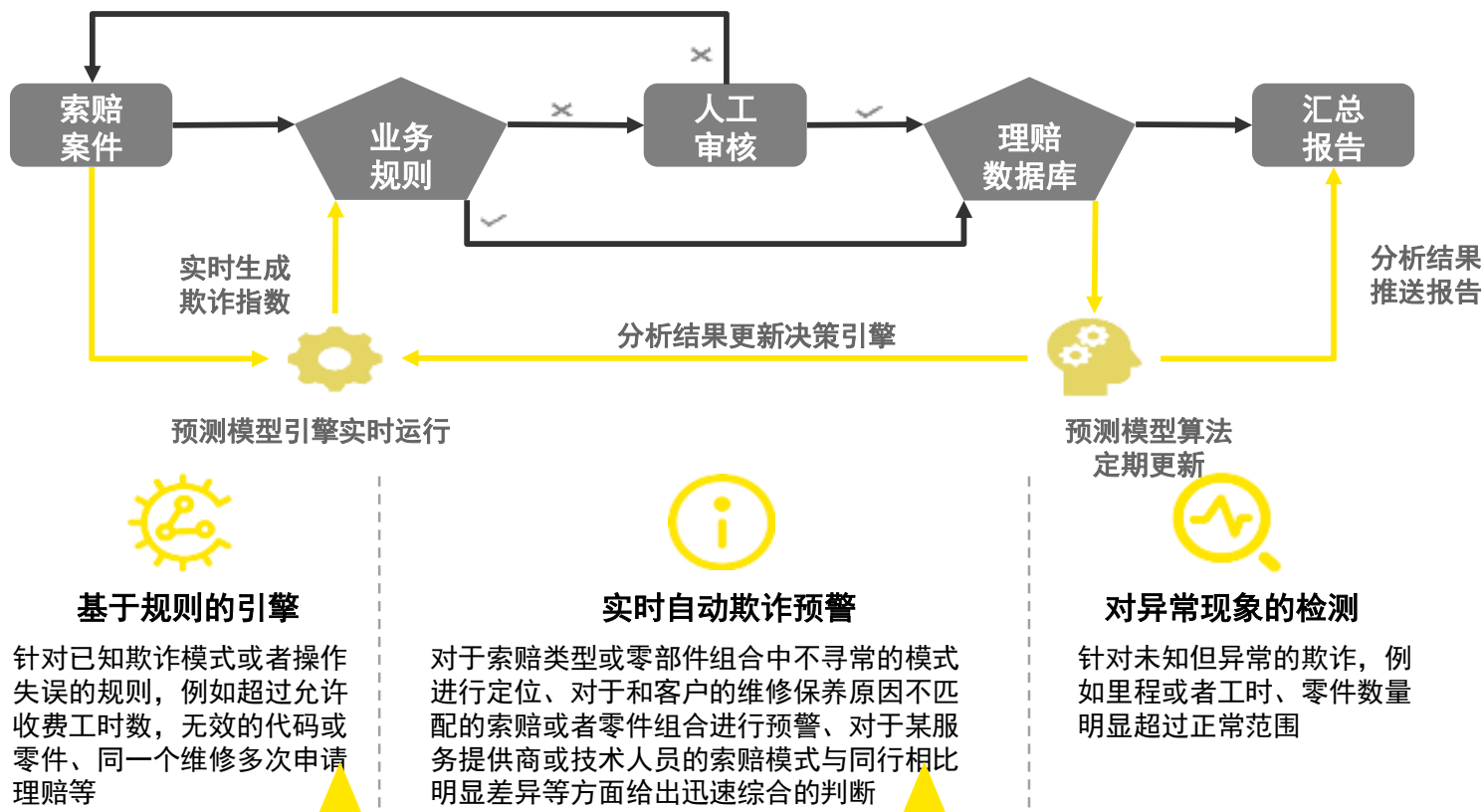
- ▶ 非常有效，但难以预先得知聚焦的服务商或案件。
- ▶ 应该采用什么作为人工审核的标准更加有效难以判断。



预测模型分析方法

- ▶ 利用数据分析和机器学习的方法，针对欺诈的可能性高低对每个索赔和服务提供商进行精准评分，从而合理分配审核员的工作重点。
- ▶ 随着时间推移，机器学习算法可以从服务提供商和审核人员的行为中学习，获得越来越多的欺诈判断标识，使判断更精准。

保险理赔反欺诈---解决方案



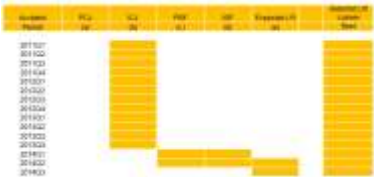
基于单个赔案的准备金计算方法

传统准备金模型

- 链梯法：应用于相对较早的事故季度



- 预期赔付率法：应用于较近事故季度

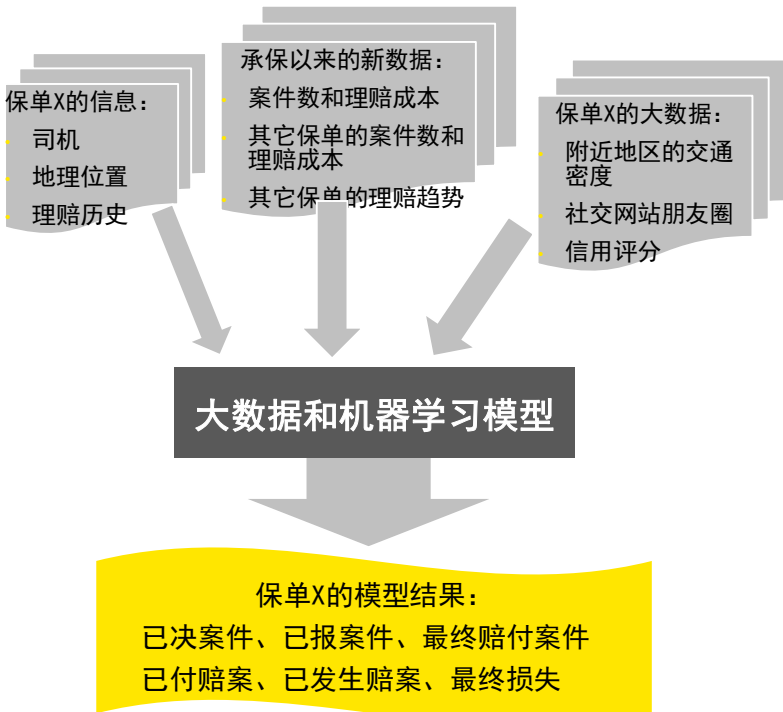


- B-F法：应用于较近事故季度

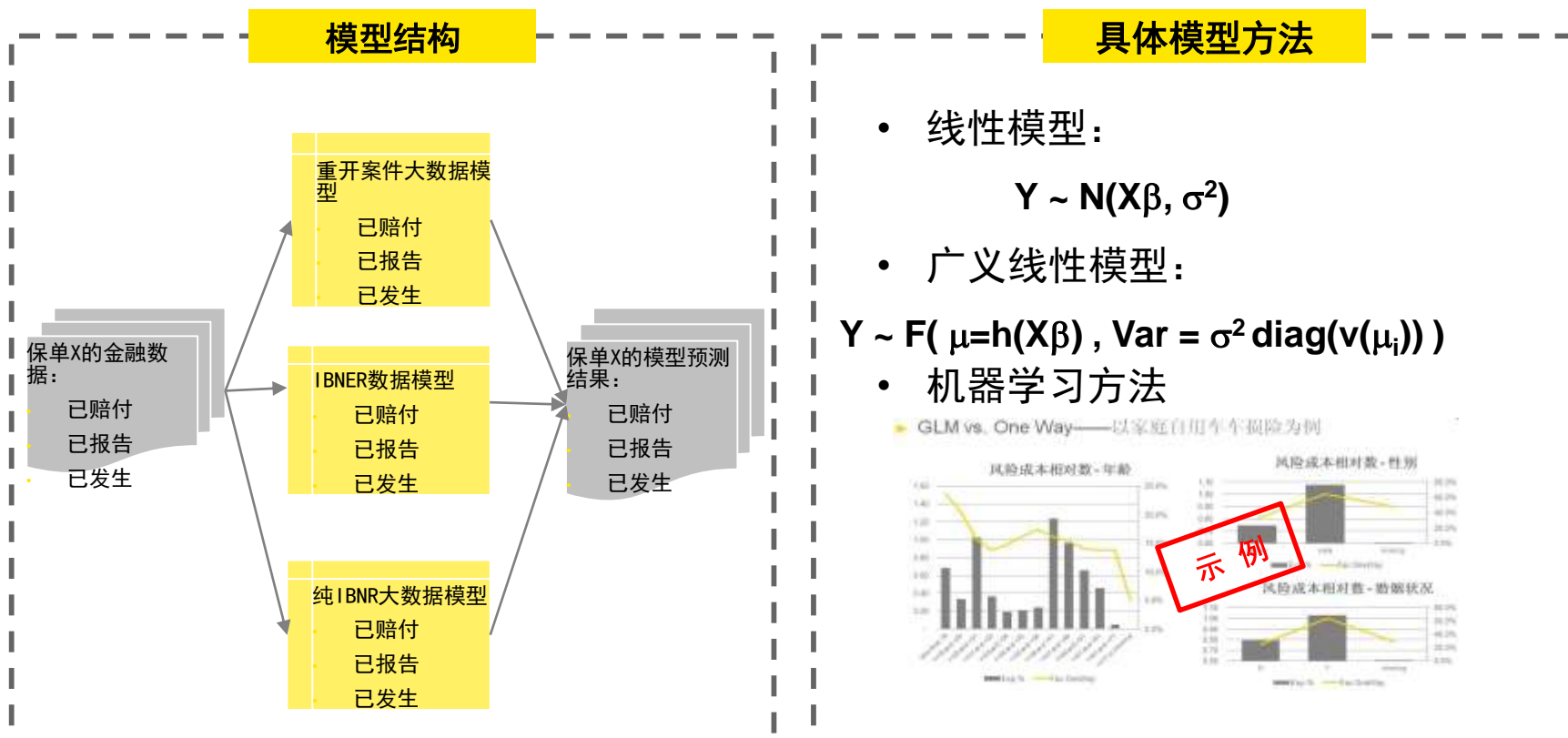
修正最终预期 = $x \cdot \text{链梯法最终预期} + (1-x) \cdot \text{预期赔付率法的最终预期}$

Accident Year	Expected Premium	Selected Loss Ratio	Expected Loss Ratio	Paid Ratio to UR	Expected Percent	Expected Loss	Paid to Date	Expected UR Loss	Expected UR Loss
2000	100	100%	100%	100%	100%	100	100	100	100
2001	100	100%	100%	100%	100%	100	100	100	100
2002	100	100%	100%	100%	100%	100	100	100	100
2003	100	100%	100%	100%	100%	100	100	100	100
2004	100	100%	100%	100%	100%	100	100	100	100
2005	100	100%	100%	100%	100%	100	100	100	100
2006	100	100%	100%	100%	100%	100	100	100	100
2007	100	100%	100%	100%	100%	100	100	100	100
2008	100	100%	100%	100%	100%	100	100	100	100
2009	100	100%	100%	100%	100%	100	100	100	100
2010	100	100%	100%	100%	100%	100	100	100	100
2011	100	100%	100%	100%	100%	100	100	100	100
2012	100	100%	100%	100%	100%	100	100	100	100
2013	100	100%	100%	100%	100%	100	100	100	100
2014	100	100%	100%	100%	100%	100	100	100	100
2015	100	100%	100%	100%	100%	100	100	100	100
2016	100	100%	100%	100%	100%	100	100	100	100
2017	100	100%	100%	100%	100%	100	100	100	100
2018	100	100%	100%	100%	100%	100	100	100	100
2019	100	100%	100%	100%	100%	100	100	100	100
2020	100	100%	100%	100%	100%	100	100	100	100
2021	100	100%	100%	100%	100%	100	100	100	100
2022	100	100%	100%	100%	100%	100	100	100	100
2023	100	100%	100%	100%	100%	100	100	100	100
2024	100	100%	100%	100%	100%	100	100	100	100
2025	100	100%	100%	100%	100%	100	100	100	100
2026	100	100%	100%	100%	100%	100	100	100	100
2027	100	100%	100%	100%	100%	100	100	100	100
2028	100	100%	100%	100%	100%	100	100	100	100
2029	100	100%	100%	100%	100%	100	100	100	100
2030	100	100%	100%	100%	100%	100	100	100	100
2031	100	100%	100%	100%	100%	100	100	100	100
2032	100	100%	100%	100%	100%	100	100	100	100
2033	100	100%	100%	100%	100%	100	100	100	100
2034	100	100%	100%	100%	100%	100	100	100	100
2035	100	100%	100%	100%	100%	100	100	100	100
2036	100	100%	100%	100%	100%	100	100	100	100
2037	100	100%	100%	100%	100%	100	100	100	100
2038	100	100%	100%	100%	100%	100	100	100	100
2039	100	100%	100%	100%	100%	100	100	100	100
2040	100	100%	100%	100%	100%	100	100	100	100
2041	100	100%	100%	100%	100%	100	100	100	100
2042	100	100%	100%	100%	100%	100	100	100	100
2043	100	100%	100%	100%	100%	100	100	100	100
2044	100	100%	100%	100%	100%	100	100	100	100
2045	100	100%	100%	100%	100%	100	100	100	100
2046	100	100%	100%	100%	100%	100	100	100	100
2047	100	100%	100%	100%	100%	100	100	100	100
2048	100	100%	100%	100%	100%	100	100	100	100
2049	100	100%	100%	100%	100%	100	100	100	100
2050	100	100%	100%	100%	100%	100	100	100	100
2051	100	100%	100%	100%	100%	100	100	100	100
2052	100	100%	100%	100%	100%	100	100	100	100
2053	100	100%	100%	100%	100%	100	100	100	100
2054	100	100%	100%	100%	100%	100	100	100	100
2055	100	100%	100%	100%	100%	100	100	100	100
2056	100	100%	100%	100%	100%	100	100	100	100
2057	100	100%	100%	100%	100%	100	100	100	100
2058	100	100%	100%	100%	100%	100	100	100	100
2059	100	100%	100%	100%	100%	100	100	100	100
2060	100	100%	100%	100%	100%	100	100	100	100
2061	100	100%	100%	100%	100%	100	100	100	100
2062	100	100%	100%	100%	100%	100	100	100	100
2063	100	100%	100%	100%	100%	100	100	100	100
2064	100	100%	100%	100%	100%	100	100	100	100
2065	100	100%	100%	100%	100%	100	100	100	100
2066	100	100%	100%	100%	100%	100	100	100	100
2067	100	100%	100%	100%	100%	100	100	100	100
2068	100	100%	100%	100%	100%	100	100	100	100
2069	100	100%	100%	100%	100%	100	100	100	100
2070	100	100%	100%	100%	100%	100	100	100	100
2071	100	100%	100%	100%	100%	100	100	100	100
2072	100	100%	100%	100%	100%	100	100	100	100
2073	100	100%	100%	100%	100%	100	100	100	100
2074	100	100%	100%	100%	100%	100	100	100	100
2075	100	100%	100%	100%	100%	100	100	100	100
2076	100	100%	100%	100%	100%	100	100	100	100
2077	100	100%	100%	100%	100%	100	100	100	100
2078	100	100%	100%	100%	100%	100	100	100	100
2079	100	100%	100%	100%	100%	100	100	100	100
2080	100	100%	100%	100%	100%	100	100	100	100
2081	100	100%	100%	100%	100%	100	100	100	100
2082	100	100%	100%	100%	100%	100	100	100	100
2083	100	100%	100%	100%	100%	100	100	100	100
2084	100	100%	100%	100%	100%	100	100	100	100
2085	100	100%	100%	100%	100%	100	100	100	100
2086	100	100%	100%	100%	100%	100	100	100	100
2087	100	100%	100%	100%	100%	100	100	100	100
2088	100	100%	100%	100%	100%	100	100	100	100
2089	100	100%	100%	100%	100%	100	100	100	100
2090	100	100%	100%	100%	100%	100	100	100	100
2091	100	100%	100%	100%	100%	100	100	100	100
2092	100	100%	100%	100%	100%	100	100	100	100
2093	100	100%	100%	100%	100%	100	100	100	100
2094	100	100%	100%	100%	100%	100	100	100	100
2095	100	100%	100%	100%	100%	100	100	100	100
2096	100	100%	100%	100%	100%	100	100	100	100
2097	100	100%	100%	100%	100%	100	100	100	100
2098	100	100%	100%	100%	100%	100	100	100	100
2099	100	100%	100%	100%	100%	100	100	100	100
2100	100	100%	100%	100%	100%	100	100	100	100

随机准备金模型



基于单个赔案的准备金计算建模方法





目录

1

预测模型简介

2

保险精算中的应用实例

3

预测模型应用实践

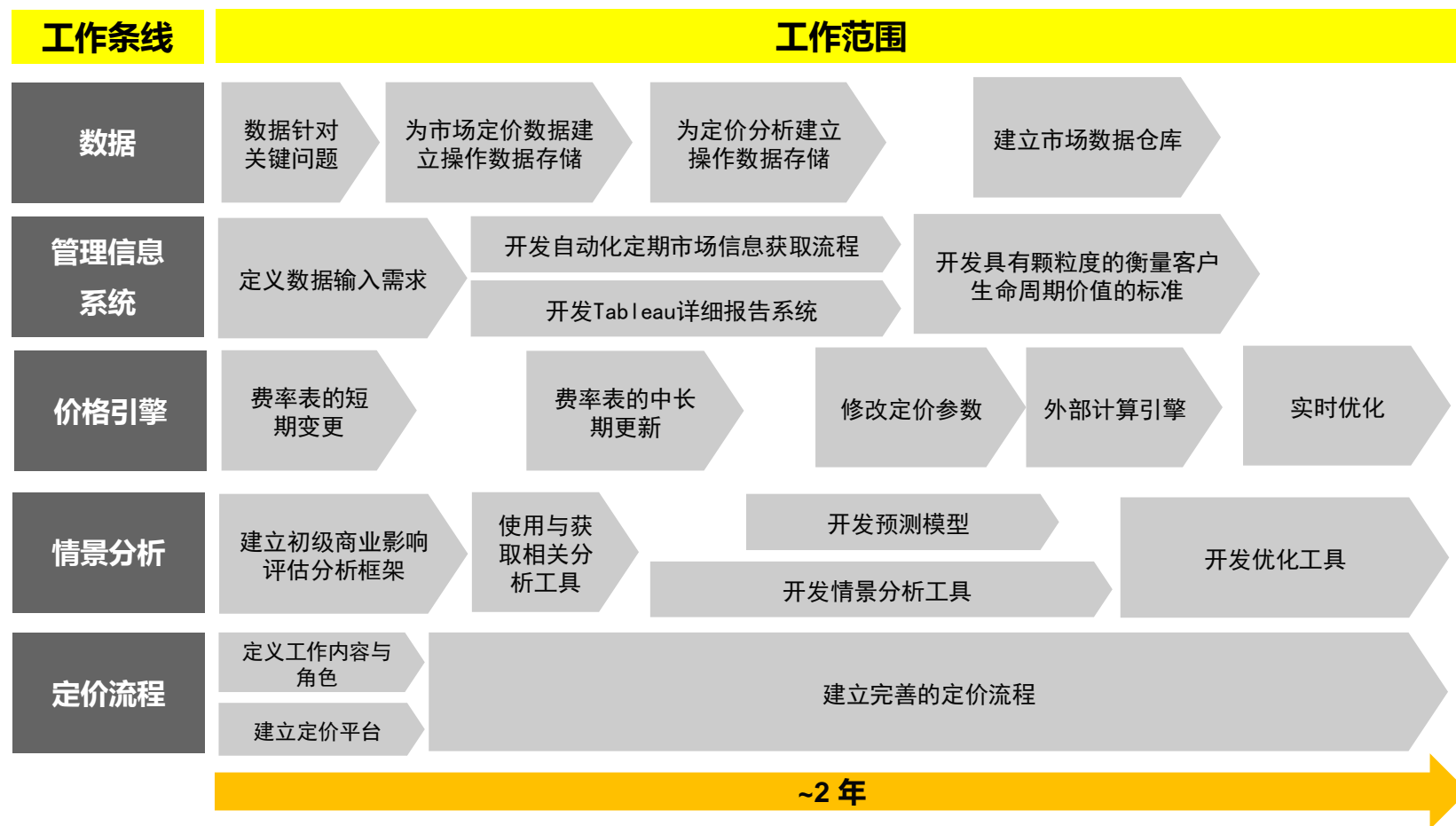
4

小结和提问

定价总体框架和流程

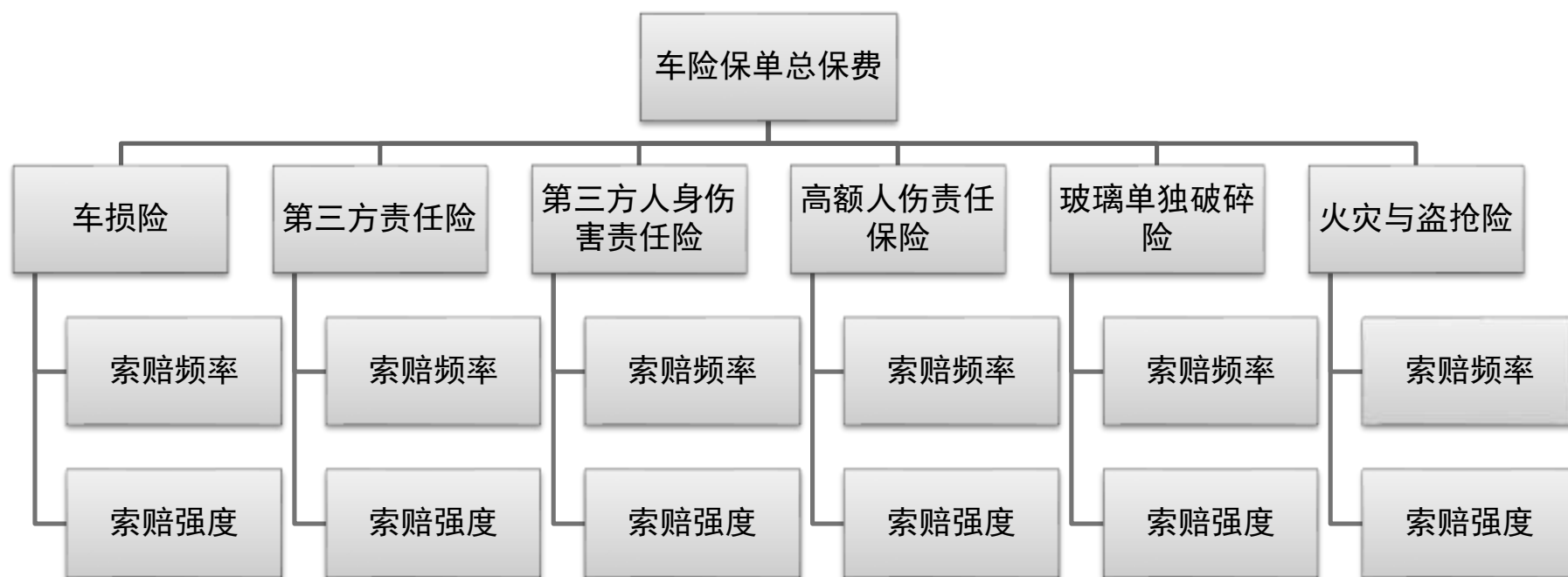


典型的定价核保职能的组建规划



典型的车险定价模型结构

- 英国常用车险模型的保费结构如下，车险保单中涵盖了各险别的纯风险保费，并需要考虑费用率和无赔款折扣系数进行保费计算



外部数据及相应的作用

▶ 外部数据源

行业数据



征信类数据



诚信上海

移动类数据



高德地图



谷歌地图



社交媒体类数据



法规类数据



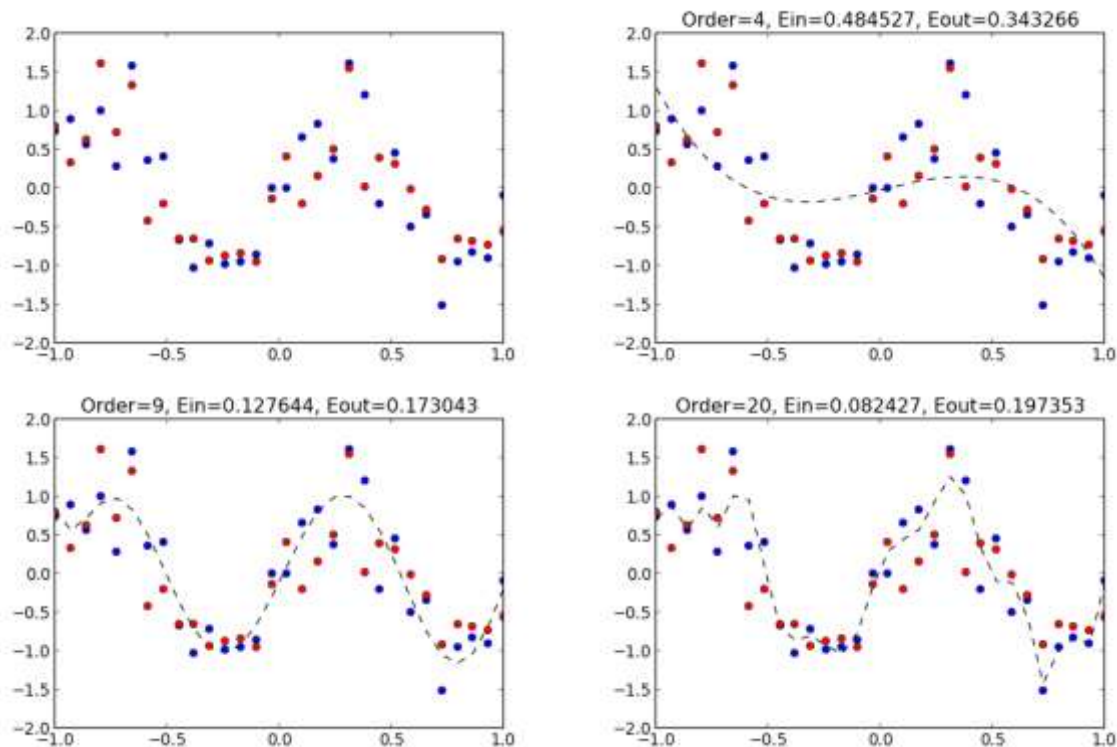
上海交警
app

- ▶ 模型结构的选择
 - ▶ 交叉项
- ▶ 参数的选择
 - ▶ P 值方法
 - ▶ 逐步回归法
 - ▶ 正则化线性回归
- ▶ 参数校验

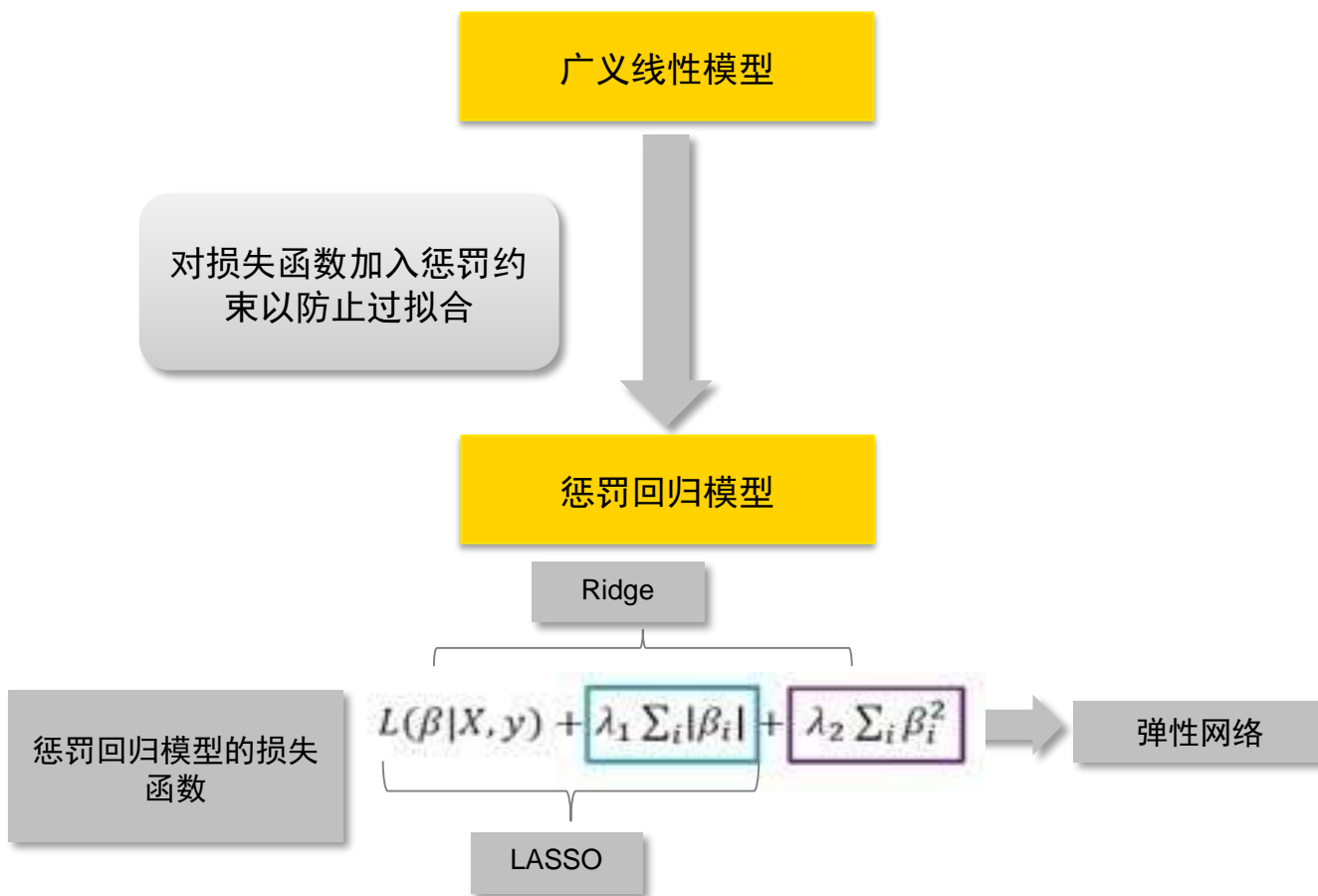
广义线性模型的过拟合问题

当模型过分拟合时，统计模型更多地预测了随机残差/噪声而不是真正的规律

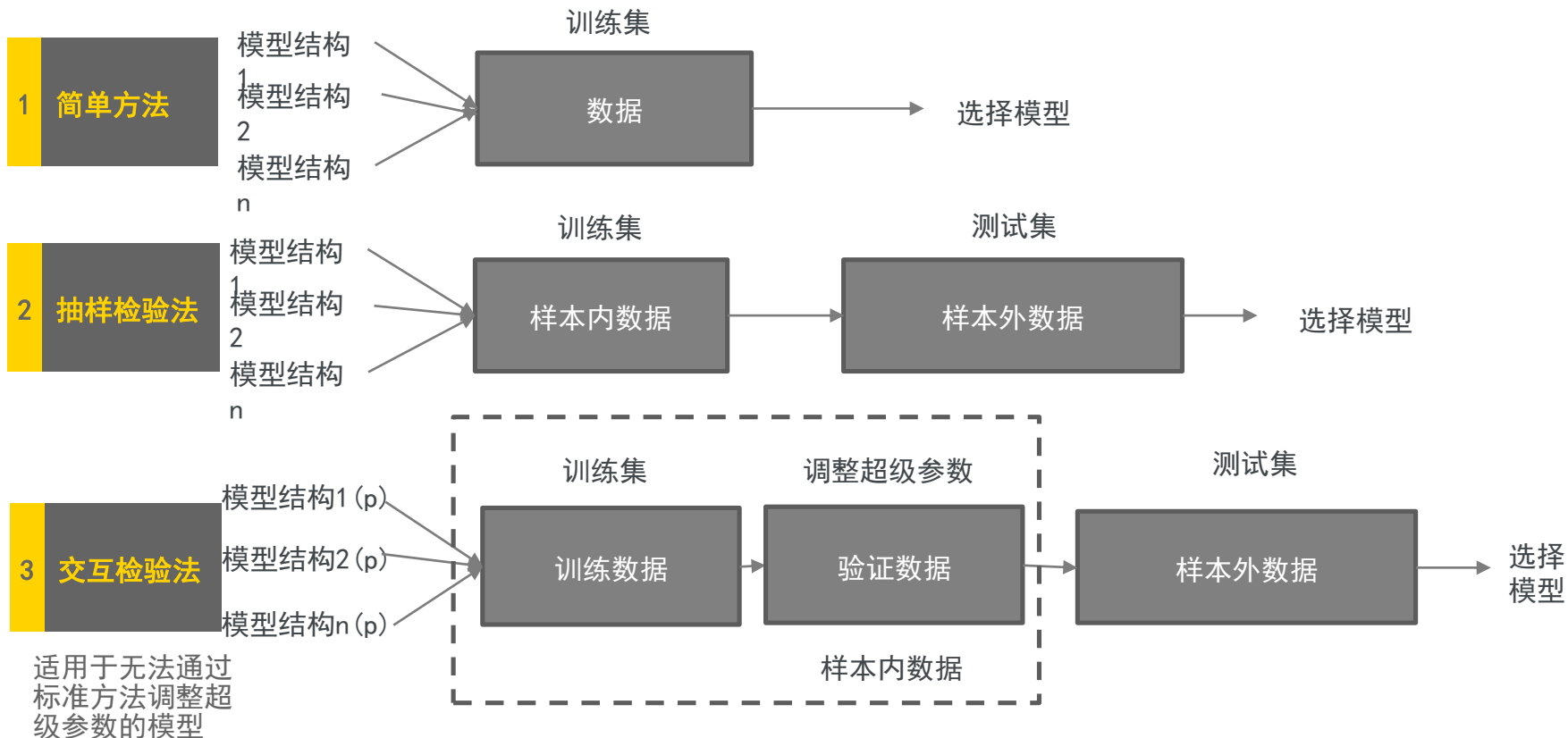
图表中的蓝色点是训练数据而红色点是测试数据



正则化回归模型：若使用得当预测能力较广义线性模型强，其他性能和GLM类似



模型中参数的选择方法



K折交叉验证进一步提高模型的稳定性

我们把整个建模数据集分成K份，每次选其中一份作为样本外（测试）数据，其余作为样本内数据，轮流K次。如果需要，样本内数据仍分成两部分，一部分用于训练，一部分用于验证，也可以使用K折方法。

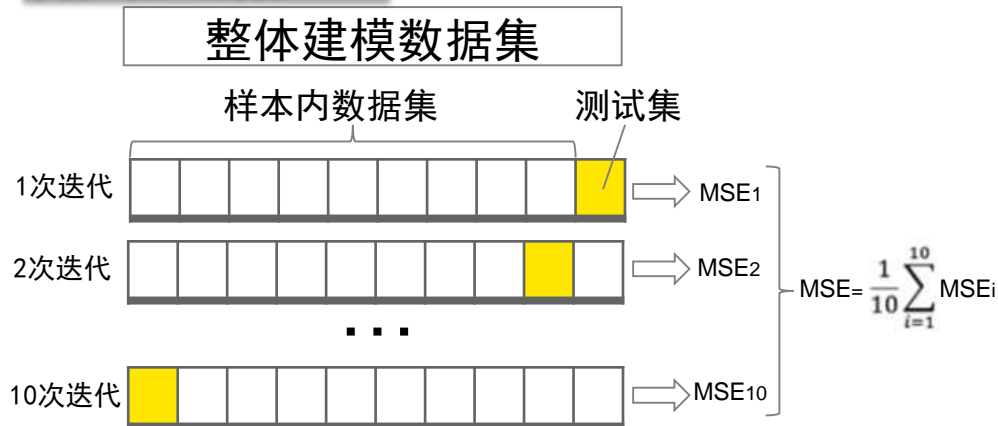
利用K折交叉验证的步骤是：

示例：以K=10为例

第一步：使用不重复抽样将原始建模数据随机分为K份

第二步：K-1份数据用于模型训练，剩下数据用于测试模型。然后重复K次，得到K个模型评估结果MSE_i

第三步：将K次的MSE_i取平均得到最后的MSE



- K的设定主要基于经验和可行性，一般采用5或者10。
- 如果数据集较小，则需要设定较大k值使训练集增大，因为K越大，训练集在整个原始训练集的占比就越多。
- 而K过大则会导致训练模型个数过多，导致高方差。



目录

1

预测模型简介

2

保险精算中的应用实例

3

预测模型应用实践

4

小结和提问

安永可以为客户提供以下服务



数据基础设施建设

- ▶ IT系统咨询与实施
- ▶ 内部数据库建设
- ▶ 外部数据资源整合



产品与服务创新

- ▶ UBI 产品
- ▶ 个性化产品推荐
- ▶ 车险快速理赔



市场调研与分析

- ▶ 目标客户群
- ▶ 竞争对手比对
- ▶ 行业数据分析



客户关系管理

- ▶ 客户分群
- ▶ 客户忠诚度管理
- ▶ 交叉及向上销售
- ▶ 客户生命周期管理



核保定价

- ▶ 现有定价模型评估
- ▶ 费率分配及两率联动
- ▶ 定价模型分析2.0
- ▶ 客户价格弹性定价
- ▶ 核保规则
- ▶ 价格最优化



核赔

- ▶ 理赔预防与缓解
- ▶ 理赔渗漏管理和反欺诈



绩效评估

- ▶ 业务品质监控
- ▶ 随机准备金评估

演讲者简介



姚佶 博士

电话: +86 173 1560 0803

Email: Jeff.Yao@cn.ey.com

- ▶ 现任安永精算与风险管理咨询服务总监，在保险精算行业超过十二年工作经验，还在国际知名高校精算系有三年的教学科研经验，致力于数据和模型在保险精算领域的理论研究和实践应用
- ▶ 联合编著《预测模型在精算科学中的应用》一书
- ▶ 在加入安永(中国)之前，他在英国开发机器学习和大数据挖掘的全新软件Tyche，是该软件的创始团队成员之一
- ▶ 除日常工作外，他现任世界多所大学和国际精算机构协会的考官，兼任上海交通大学中英精算项目学术主任，曾在英国肯特大学兼职讲授应用精算学硕士课程多年。他是英国高等教育学会会士(FHEA)
- ▶ 数学与统计专业博士，英国精算师(FIA)，特许企业风险分析师(CERA)

EY 安永 | Assurance 审计 | Tax 税务 | Transactions 财务交易 | Advisory 咨询

关于安永

安永是全球领先的审计、税务、财务交易和咨询服务机构之一。我们的深刻洞察和优质服务有助全球各地资本市场和经济体建立信任和信心。我们致力培养杰出领导人才，通过团队协作落实我们对所有利益关联方的坚定承诺。因此，我们在为员工、客户及社会各界建设更美好的商业世界的过程中担当重要角色。

安永是指 Ernst & Young Global Limited 的全球组织，也可指其一家或以上的成员机构，各成员机构都是独立的法人实体。Ernst & Young Global Limited 是英国一家担保有限公司，并不向客户提供服务。如欲进一步了解安永，请浏览 www.ey.com。

© 2017 安永，中国
版权所有。

APAC no.
ED MMY

本材料是为提供一般信息的用途编制，并非旨在成为可依赖的会计、税务或其他专业意见。请向您的顾问获取具体意见。

www.ey.com/china

