

卷烟异常流动事前预警模型的研究与应用

何朝阳 聂立佳 钟艺 王刚 乜堪雄

DOI:10.12238/ej.v7i10.1945

[摘要] 本文首先通过深度访谈、关键词云等方法进行数据挖掘,建立了兼具深度和广度的三级指标体系,再综合运用专家打分法、SPSS因子分析法对指标体系分析、优化、处理,构建“四维六因”预警模型。其次,以不同频次实时攫取近年来辖区外流涉案零售户在内管、专卖、营销、物流系统的数据,由Python语言进行建模和编译,通过数据清洗和数据特征选择,结合“四维六因”预警模型,确定辖区内各零售户的外流风险系数,根据零售户的风险系数进行风险评价,实时输出预警信息,实现对卷烟异常流动行为的高精度靶向监管。

[关键词] 卷烟异常流动; 事前; 指标体系; 预警模型

中图分类号: V271.4+7 **文献标识码:** A

Research and application of advance warning model of abnormal cigarette flow

Zhaoyang He Lijia Nie Yi Zhong Gang Wang Kanxiong Nie

[Abstract] Through in-depth interview, keyword cloud and other methods of data mining, this paper establishes a three-level index system with both depth and breadth, and then comprehensively uses the expert scoring method, SPSS factor analysis method to analyze, optimize and process the index system, to build the "four dimensional and six causes" early warning model. Second, with different frequency real-time grab jurisdictions outflow in recent years in retail, monopoly, marketing, logistics system data, by Python language modeling and compilation, through the data cleaning and data characteristics selection, combined with the "four dimensional six for" early warning model, determine the retail households within its jurisdiction of outflow risk coefficient, according to the retail risk assessment, real-time output early warning information, realize the high precision of cigarette abnormal flow targeted regulation.

[Key words] abnormal cigarette flow; advance; index system; early warning model

引言

“十四五”以来,国家烟草专卖局印发关于进一步加强真烟异常流动治理工作的通知,强调了促进重点单位加强工商协同、有针对性地调控,从根源上减轻、从渠道上制止、从案件上震慑,全面提升烟草异常流动治理水平。

目前,卷烟异常流动的现象屡禁不止,甚至在部分区域卷烟违法违规案件占比达到一半以上,其原因涉及多方面:利益驱使下的市场自发行为、卷烟投放策略与市场需求的 mismatch、地区间价格差异巨大、物流行业的高速发展等;卷烟异常流动严重扰乱了正常的卷烟销售秩序,损害了守法商户的利益,破坏了烟草行业的经营理念,对专卖制度的稳定有一定的危害。当前,如何有效地遏制卷烟异常流动现象,不仅是国家局对真烟异常流动加强治理的总体要求,也是地方各级专卖部门加强真烟外流预警及监管的迫切需要。

基于以上背景,本文旨在建立一种有效的卷烟异常流动事前预警模型,实现对卷烟异常流动行为的高精度靶向监管。

1 存在的问题及难点

1.1 数据样本量较少、数据有一定失真

数据样本问题体现在多方面,从零售户方面看,终端扫码数据作为卷烟市场动态的重要数据来源,辖区内的扫码客户仅有约30%,且扫码客户并不会每次销售时都使用终端扫码,同时,客户扫码的时间与实际销售时间有差异,出现集中时段大批量扫码、扫码商品不全等现象。

1.2 数据未得到有效地利用

在数据挖掘过程中,发现烟草内部各部门在工作中对数据的利用较少,专卖管理人员在办案途中主要依靠线索办案,线索的来源主要为:日常排查、蹲点、举报,未利用系统数据对办案提供实质性的帮助,所破获的案件中也没有直接与系统数据相关联的案件。

1.3 缺乏具有实践性、完整性的预警模型

现有的预警模型存在滞后性、缺乏科学性,无法有效地实现对卷烟异常流动问题的提前预警,难以达到防患于未然的理想效果。

卷烟异常流动监管分为事前、事中、事后三个阶段，目前系统中的预警行为主要为事后预警，经营者已经发生了异常经营行为，这种滞后的预警机制无法及时发现潜在的风险，难以有效地遏制卷烟异常流动问题的发生。

2 数据采集与分析

2.1 数据采集

在数据采集阶段，综合利用了主要和次要资源，包括深度访谈、咨询报告和小组焦点讨论，重点针对烟草各层级的一线工作者进行详尽访谈。在首次访谈中，我们选择12位具有代表性的烟草行业工作人员，为了保证访谈的科学性，确保访谈内容能够反映对应岗位群体的真实情况，我们每个岗位选取两人，将其分为两个深度访谈小组（如表1所示）。

表1 访谈安排及名单

访谈者	被访谈者	被访谈者职位	访谈者	被访谈者	被访谈者职位
七堪雄	陈军	打通区域稽查员	陈昱润	刘昊	稽查支队队员
	白太安	城郊区域稽查队长		杜娇	城区客户经理
	代鹏晋	城郊客户经理		张炜铭	打通区域稽查大队队长
	谢鑫	城区客户经理		杨泳	城区区域稽查大队队长
	王崇宇	城郊区域客户主任		敬凯	稽查支队队员
	冉文	城区区域客户主任		彭博	城区客户经理

本次访谈的主题围绕以下两点展开：一是了解所处职位人员负责处理烟草异动的相关内容，二是探讨烟草异动预警改进措施，以提升卷烟异常流动的市场靶向监管。

2.2 数据处理

表2 三级指标体系

一级指标	二级指标	三级指标
静态指标	客户经营状态	敏感品规、地区敏感品规、一人多证、客户档级、零售户基础资料名字或者电话变更
	专卖监管	大户控制小户、长期未开门标签、犯案次数、户籍信息
	内管数据	外流被查处+被我局查处
动态指标	客户经营状态	扫码时间规律性+结合区域性
	营销指标	可销天数、库存量、单品规订单满足率（敏感品规）、条销比、品规销售价格、节假日订货异常
	内管数据	物流送货不到位

2.2.1 三级指标体系建立

在分析词云后，我们召开了内部专题会议，以单独评估样本中关键词的层次，会议成员包含学术界的专家、研究生团队、烟草行业专家，利用层次分析法，对被访谈者所提到的关键词分层，得到了初代的指标体系（如表2所示）。

2.2.2 专家打分

第二轮的数据挖掘时，我们采用了专家打分法。首先，确定专家人选，所选取的专家应对烟草行业非常熟悉，有较高的权威性和代表性，选取的人数不宜太少，以减小误差，最终确定了县、市级专家116名；第二步设计卷烟异常流动预警指标调查问卷，详细解释了第一轮数据挖掘的30个预警指标的含义，根据指标的详细定义提出有关卷烟异常流动的问题，共71问（见附录II的调查问卷），将每个指标划分多个等级，并为各等级赋予定量数值，用于判断各指标对卷烟异常流动影响的程度，此次调查共收到116份调查问卷，其中相同打分问卷8份，未打分问卷1份，剩余有效问卷为107份。

3 “四维六因”模型构建

3.1 降维分析

本文利用SPSS软件对问卷量表进行严格的探索性因子分析，主要是为了找出影响观测变量的因子个数，以及各个因子和各个观测变量之间的相关程度，以试图揭示一套相对比较大的变量的内在结构。通过这一科学方法，我们能够准确评估问卷设计的有效性和可靠性，为后续模型建立提供坚实基础。

3.1.1 KMO与巴特利特球形检验

KMO 和巴特利特检验	
KMO 取样适切性量数	.908
巴特利特球形度检验	近似卡方 2707.505
	自由度 465
	显著性 .000

图1 KMO与巴特利特球形检验图

KMO是Kaiser-Meyer-Olkin的取样适当性量数，由图1可知，此处的KMO值为0.908，代表适合做因子分析。与此同时，Bartlett's球形检验的统计量值为2707.505，自由度为465，相应的概率P值为0，达到显著，表示样本的相关矩阵间有公因子存在，可以进行因子分析。

3.1.2 旋转后的成分矩阵

旋转成分矩阵可以分析出每个因子与变量的对应关系情况，当变量与因子对应的值小于0.5时，表示该变量与该因子关系不大；当变量与因子对应的值大于0.5时，就可以认为该变量与该因子有对应关系。然后将不合理的变量删除，直到所有变量与因子对应关系和预期基本一致。最终得到6个公因子，31个关键指标。

补充：（1）一个指标对应多个因子时，研究成员需结合烟草行业的专业知识进行综合判断归属于哪个因子；（2）当指标与因子在旋转成分矩阵中相关时，如果指标与因子无实际对应关系，

则删除该指标；(3)如果一个因子对应多个类型指标,此时可考虑增加1个因子；(4)如果一个因子对应指标单一,可根据其影响程度考虑删除该因子。

3.2 “四维六因”预警模型

我们根据各因子下指标的类型对因子进行命名,并将相同的因子放到同一维度内,得到“四维六因”预警模型,如图2所示:

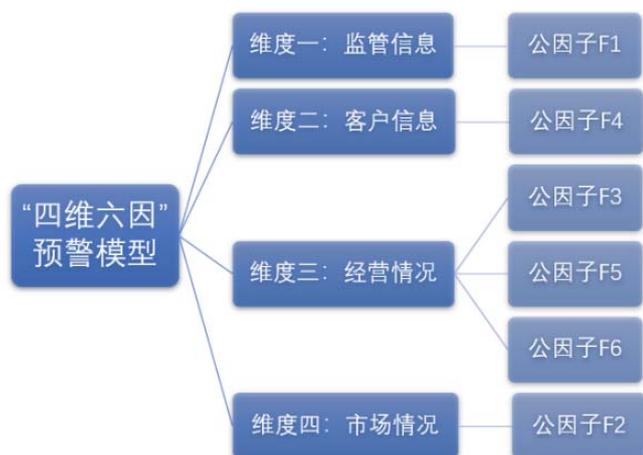


图2 “四维六因”预警模型

通过以上的研究,我们确定了“四维六因”预警模型,模型包含四个维度,每个维度的指标选取涉及专卖、内管、营销多个层面,并在研究过程中,综合利用SPSS分析软件、专题研讨会等方法剔除重复、无相关性、不准确、无数据价值的部分,最终得到了较为精准的预警模型。

3.3 预警模型工具开发

基于“四维六因”预警模型,抓取系统内的数据,由Python语言进行建模和编译,完成对零售户卷烟异常流动风险系数的计算;数据清洗和数据特征选择由Python脚本自动完成,数据清洗和特征选择后,将后端数据输入建好的数据模型,模型将自动测算零售户卷烟异常流动风险系数,根据零售户的风险系数进行风险评价,实时输出预警结果,预警结果直接推送至零售户所属区域监管人员,并进行可视化输出,形成数据看板。

3.3.1 模型算法确定

首先,抓取烟草系统内的数据,数据来源涉及专卖系统、内管系统、营销系统、物流系统,数据种类包含:持证人身份信息、基础资料、档级信息、监管记录、零售户基础信息维护记录、检查记录、订货时间;其次,根据“四维六因”预警模型中指标的详细描述,对数据进行清洗和编译。

例如对维度一监管信息风险系数的数据处理逻辑:第一步,

初始化后需要读取具体使用到的每一个工作表,目前使用的sheet表共有四个分别是:客户基础信息表、涉案关联户表、检查未开门表、监管次数记录表;第二步,根据零售户的身份证信息,筛选出一人多证的用户,对一人多证的零售户,给定权重,按持证数量进行叠加,同时,对一人多证的零售户持证档级累加;第三步,通过案件记录,找到案件关联户信息,对关联户给定权重,设置时间跨度选择窗口,选择特定时间内零售户案件的关联次数,以及涉案的数量,对关联户进行加权(对涉案数量加权是根据不同烟种类型进行计算;最后,对零售户的监管次数、未开门次数给定权重,并通过模型一监管信息的函数模型,完成对风险系数F1的计算。

3.3.2 模型结果分析

通过对数据清洗和特征选择,结合“四维六因”预警模型,最终得到了零售户的风险系数F1,如图15所示,Y轴为风险系数,X轴为零售户许可证号,共3052户,从左到右按风险系数由低到高排列。按风险系数划分四段风险等级:40-50分为一级预警,30-40分为二级预警,15-30分为三级预警,0-15为四级预警,预警结果为一级预警62户、二级预警106户、三级预警720户、四级预警2164户。

4 结语

卷烟异常流动指标体系构建与算法模型开发是一个复杂的系统工程,需要不断探索和改进。随着数字技术的发展和数据积累的不断丰富,相信未来能够构建更加完善的指标体系和更加高效的算法模型,为维护烟草市场秩序、打击非法经营活动提供有力支撑。

[参考文献]

- [1]程国亚,杨琳,张帅,等.卷烟规范经营预警模型研究与应用——以安庆市烟草专卖局为例[J].活力,2022,(19):127-129.
- [2]王金亮,刘金澎,纪鲁生,等.基于专销大数据的卷烟非法流通预警分析模型的构建与应用[J].现代商贸工业,2022,43(03).
- [3]刘赞.浅析真烟非法流通原因及对策[C]//广西壮族自治区烟草公司柳州市公司2021年学术论文汇编.柳州市城区烟草专卖局内管派驻组,2021:3.
- [4]潘申鑫,王雪芳,王曾.卷烟非法流通防控预警模型研究[C]//中国烟草学会.中国烟草学会学术年会优秀论文集.浙江舟山市烟草专卖局,2017:11.

作者简介:

何朝阳(1997--),男,汉族,四川省巴中市人,研究生,重庆交通大学,研究方向:供应链管理。