

基于银行流水数据的洗钱风险综合评估

张力芝¹, 赵胜利¹, 钟妤玥²

¹重庆理工大学理学院, 重庆

²重庆市渝高中学, 重庆

Email: zhaoshengli@cqut.edu.cn

收稿日期: 2020年12月30日; 录用日期: 2021年1月28日; 发布日期: 2021年2月4日

摘要

针对洗钱账户的行为特点, 在银行个人客户流水数据的基础上建立洗钱风险综合评价指标体系, 利用因子分析方法和熵值法建立了洗钱风险综合评价模型, 并利用广义极值分布对洗钱综合风险值的分布进行了拟合。采用重庆市某商业银行2018年个人流水数据进行了实证分析, 其结果表明, 该银行的绝大部分个人客户洗钱风险值都较低, 只有少数的客户可能有洗钱行为。根据拟合出的综合风险分布, 在0.05的显著水平之下, 有5239 (2.66%)位客户可能有洗钱风险, 需要进一步的鉴别, 这无疑将极大地提高银行的工作效率。

关键词

洗钱, 因子分析法, 熵值法, 综合评价模型, 广义极值分布

Comprehensive Evaluation of Money Laundering Risk Based on Bank Flow Data

Lizhi Zhang¹, Shengli Zhao¹, Yuyue Zhong²

¹School of Science, Chongqing University of Technology, Chongqing

²Chongqing Yugao Middle School, Chongqing

Email: zhaoshengli@cqut.edu.cn

Received: Dec. 30th, 2020; accepted: Jan. 28th, 2021; published: Feb. 4th, 2021

Abstract

According to the behavior characteristics of money-laundering, the comprehensive evaluation index system of money-laundering risk is established on the basis of the flow data of bank personal

customers. The method of factor analysis and entropy method are used to establish the comprehensive evaluation model of money-laundering risk, and the generalized extreme value (GEV) distribution is used to fit the distribution of comprehensive risk value of money-laundering. The empirical analysis is carried out by using the customers personal flow data of a commercial bank in Chongqing in 2018. The results show that the risk value of money-laundering of most bank individual customers is relatively low, and only a few customers may have money-laundering behavior. According to the fitted comprehensive risk distribution, at the significant level of 0.05, there are 5239 (2.66%) customers who may have money-laundering risk, and they need further identification, which will undoubtedly greatly improve the efficiency of the bank.

Keywords

Money-Laundering, Factor Analysis, Entropy Method, Comprehensive Evaluation Model, GEV

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

洗钱(Money laundering), 亦称洗黑钱、资金洗净, 指将通过犯罪或其他非法手段所获得的金钱、伪钞, 经过合法金融作业流程之类的方法, 以“洗净”为看似合法的资金。洗钱常与经济犯罪、毒品交易、恐怖活动及黑道等重大犯罪有所关联, 也常以跨国方式进行, 给国家的经济和社会安全带来了极大的伤害, 应该予以严厉的打击。随着科学技术的快速发展, 电子支付手段也越来越多, 在给人们的生活带来便利的同时, 也使洗钱行为呈现出手段专业化、主体组织化、资金来源多元化等特点, 更加难以监测。建立具有高度可操作性的洗钱风险评价指标体系, 对洗钱风险作出快速的评估, 对提高银行工作效率和打击洗钱活动具有重要的意义。

针对银行个人客户洗钱风险评估问题, 许多学者对其进行了研究。如陈云开[1]针对洗钱侦测问题, 通过给定的可疑交易判定模型, 提出了一个基于超图模型的高维聚类算法, 并运用该算法来对可疑交易进行判定。丁佳[2]通过建立第三方支付机构洗钱风险的评估指标体系, 采用层次分析法和熵权法对评价指标进行综合赋权, 将主观与客观、定性与定量指标结合, 实现了风险评估。肖琨[3]采用监督学习的方法建立了一个基于货币交易数据的洗钱风险检测和犯罪类别分类系统, 通过串联两种检测分类模型来提高整体性能, 验证了该系统的可行性。丁昱[4]从客户身份、客户正常交易数量以及客户行为这三个维度建立洗钱风险评估体系, 构建基于RBF神经网络、聚类算法和决策树模型预测结果的集成风险分类模型, 实现对洗钱风险的评估。这些研究都是通过建立合适的评价模型来分析客户洗钱风险程度, 为银行判断存在洗钱风险的客户提供了一定的参考作用。还有一些学者从理论上探讨了我国银行洗钱风险问题, 并提出相应的对策, 具体见文献[5] [6] [7]。

综合上述研究, 其都是依据客户交易信息和个人信息两方面的数据进行的。然而, 客户的个人信息形式不一, 容易缺失或者真实性无法考证, 使得风险值的估计容易有偏差; 此外, 客户个人信息数据格式不统一, 难以进行批处理, 从而影响风险评估的效率。因此, 本文仅根据银行个人客户的流水数据, 建立洗钱风险综合评价模型, 对银行个人客户的洗钱风险进行分析, 并实证考查了该模型的准确性和精度。

2. 银行个人客户洗钱风险综合评价模型

银行个人客户洗钱风险评估的关键在于评价指标体系的选取,而当前的指标体系大都依赖于客户的个人信息和交易信息。本文从客户个人流水信息中提取出相应的评价指标,重新建立指标体系,进而构建洗钱风险综合评估模型。

2.1. 个人客户流水数据采集

银行流水是客户通过银行账户流入和流出的资金记录,是实时记录的客观数据,客户通过银行开立的账户进行各种资金往来,如转账、消费、用于理财等,都会被一一记录下来。银行流水基本包括有交易时间、交易金额、交易对象以及交易类型等,这些信息客观真实地反映了客户交易的基本情况,具备了定量计算所需数据以及定性分析所需信息。

2.2. 洗钱风险综合评价指标体系

当前有关银行客户洗钱的评价指标主要从客户的身份、正常交易的数量以及客户的行为这三个维度出发,以大额可疑异常交易的风险识别点和反洗钱的检查系统为标准,来进行筛选评估。丁昱[4]根据我国反洗钱的相关条例,从客户的身份以及客户正常交易数量与行为来设计了一套开放的、可根据后续需要进行调整的客户和账户风险分类指标体系。本文在该指标体系的基础上,结合银行个人流水数据的特点,重新构建洗钱风险综合评价指标体系。

2.2.1. 洗钱风险评价指标的选取原则

银行客户的洗钱风险评价指标体系是银行进行客户洗钱筛选的基础与前提,是判定结果是否具有客观性和真实性的标准,必须要遵循一定的指标选取原则[8]:

① 科学性原则。洗钱风险评估的各项指标应建立在科学实用的基础上,要有清晰明了的层次结构,并且能真实客观地反映银行客户的交易情况。

② 全面性原则。选取的评价指标不仅需要全面反映出与客户洗钱相关的各方面信息,还能够相互配合。

③ 可操作性原则。评价指标所需数据必须准确,易于采集。

④ 公正性原则。选取的洗钱风险评价指标必须客观反映出客户的交易情况,不能偏袒。

⑤ 可比性原则。评价指标的选取要具有概括性和普遍性,便于能够准确科学地进行纵向或横向比较。

2.2.2. 洗钱风险综合评价指标体系

本文在已有的指标体系的基础上,按照评价指标的选取原则,从银行客户的个人流水数据中选取交易次数、交易金额、应报大额交易数量以及疑似可疑交易这4个二级指标以及总交易次数、现金交易次数、总交易金额等11个三级指标来建立新的基于银行客户个人流水数据的洗钱风险综合评价指标体系,如表1所示。

对表1中的指标解释如下:

交易次数(Y_1),共包含2个三级指标,即总交易次数(X_1)和现金交易次数(X_2)。其中现金交易次数指的是存取款总次数,因为客户的取款与存款是属于现金交易的一种。通常情况下,犯罪分子会将洗钱的金额进行多次交易来降低被发现的风险,因此,交易频率会在一定程度上反映洗钱的可能性,说明交易次数越多,客户存在洗钱的可能性也就越大,洗钱风险越高。

交易金额(Y_2),共包含4个三级指标,即总交易金额(X_3)、总现金交易金额(X_4)、平均交易金额(X_5)和平均现金交易金额(X_6)。其中平均交易金额是指总的交易金额除以总的交易次数,而平均现金交易金额则

是现金交易金额除以现金交易次数。交易金额的大小很大程度上会反映出客户是否存在洗钱的可能，即交易金额越大，则存在洗钱风险的可能性也就越大。

应报大额交易(Y_3)，共包含 2 个三级指标，即大额现金交易次数(X_7)和大额转账交易次数(X_8)。其中大额现金交易次数指的是存款与取款的总金额不低于 20 万元的交易次数，而大额转账交易次数指的是转入和转出的总金额不低于 50 万元的交易次数。洗钱的金额是相对比较庞大的，大额的交易金额次数会反映出客户是否有洗钱的可能，且存在大额交易的客户，其交易次数越多，洗钱风险较高。

疑似可疑交易(Y_4)，共包含 3 个三级指标，即个人短期交易金额或现金交易金额巨大的交易次数(X_9)、短期分散转入集中转出或集中转入分散转出的比例(X_{10})和个人短期大量金额为整数的交易次数(X_{11})。其中个人短期交易金额或现金交易金额巨大的交易次数指的是 5 天时间内总交易金额大于 100 万元或 10 天时间内现金交易金额大于 100 万元所发生交易的次数，而短期分散转入集中转出或集中转入分散转出的比例指的是 10 天时间内转入转出比或转出转入比不小于 5 的比例值，个人短期大量金额为整数的交易次数指的是 10 天时间内至少发生 5 次大笔整数交易，且交易总金额大于 40 万元的次数。犯罪分子在洗钱过程中可能为了方便，直接进行整数交易，而大多数普通客户在交易时多多少少都会有小数的产生，因此，疑似可疑交易的存在也会使得客户的洗钱风险增大。

Table 1. Integrated evaluation index system for money laundering risk

表 1. 洗钱风险综合评价指标体系

一级指标	二级指标	三级指标
综合评价指标	交易次数(Y_1)	总交易次数(X_1)
		现金交易次数(X_2)
	交易金额(Y_2)	总金额(X_3)
		总现金交易金额(X_4)
		平均交易金额(X_5)
		平均现金交易金额(X_6)
	应报大额交易(Y_3)	大额现金交易次数(X_7)
		大额转账交易次数(X_8)
	疑似可疑交易(Y_4)	个人短期交易金额或现金交易金额巨大的交易次数(X_9)
		短期分散转入集中转出或集中转入分散转出的比例(X_{10})
		个人短期大量金额为整数的交易次数(X_{11})

2.3. 洗钱风险综合评价模型

从对表 1 中评价指标的解释可以看出，指标之间存在一定的相关性，为了消除指标之间的相关性并对数据进行降维，本文采用因子分析的方法提取公因子，利用公因子来建立综合评价模型。

2.3.1. 综合评价指标的降维优化

首先，对所有的指标进行归一化处理，消除各个评价指标之间的量纲关系，使得数据具有可比性，提高评价结果的准确率。然后，计算该评价指标体系的 KMO 值并进行巴特利特球度检验，得到 KMO 值为 0.657，大于 0.5，且巴特利特球度检验统计量的观测值为 1,345,454.048，相应的 p 值为 0.000，则说明可以进行因子分析。最后，通过累计贡献率高于 75% 的标准来提取公因子，根据运行结果的总方差解释来看，前五个公因子共解释了原有变量总方差的 79.521%，总体上，原有变量的信息丢失较少，因子分析效果尚可，如表 2 所示。

Table 2. Factor total variance explanation**表 2.** 因子总方差解释

Initial Eigenvalues			Extraction Sums of Squared Loadings			Cumulative %
Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	
3.133	28.479	28.479	3.133	28.479	28.479	28.479
1.987	18.066	46.545	1.987	18.066	46.545	46.545
1.649	14.995	61.540	1.649	14.995	61.540	61.540
1.001	9.097	70.638	1.001	9.097	70.638	70.638
0.977	8.883	79.521	0.977	8.883	79.521	79.521
0.846	7.693	87.214				
0.747	6.791	94.005				
0.307	2.787	96.791				
0.210	1.907	98.698				
0.119	1.081	99.779				
0.024	0.221	100.000				

2.3.2. 因子旋转与公因子命名

通过旋转后的因子载荷矩阵, 见表 3, 可以看出每个原始指标所归于哪一个公因子中。第一个公因子(记为 F_1')包含有大额转账交易次数、个人短期交易金额或现金交易金额巨大的交易次数和总交易次数, 都是关于客户进行交易的次数, 可将其命名为交易次数因子; 第二个公因子(记为 F_2')包含有现金交易次数和大额现金交易次数, 都是关于现金交易的次数, 可命名为现金交易次数因子; 第三个公因子(记为 F_3')包含有交易总金额、平均现金交易金额、平均交易金额和现金交易金额, 是关于银行客户进行各种交易的金额, 可命名为交易金额因子; 第四个公因子(记为 F_4')为个人短期大量金额为整数的交易次数, 可命名为整数交易因子; 第五个公因子(记为 F_5')为短期分散转入集中转出或集中转入分散转出的比例, 可命名为转入转出因子。这五个公因子的名称以及构成见表 4。

Table 3. Factor load matrix after rotation**表 3.** 旋转后的因子载荷矩阵

Rotated Component Matrix	Component				
	1	2	3	4	5
Zscore (大额转账交易次数)	0.940	-0.010	0.062	0.016	0.041
Zscore (短期内大额交易次数)	0.918	-0.069	-0.004	0.018	0.039
Zscore (总交易次数)	0.797	0.519	0.049	0.008	0.022
Zscore (现金交易次数)	0.046	0.977	0.025	-0.002	-0.003
Zscore (大额现金交易次数)	0.049	0.975	0.041	-0.001	-0.002
Zscore (交易总金额)	0.345	0.160	0.699	0.032	-0.025
Zscore (平均现金交易金额)	-0.074	-0.028	0.693	-0.015	0.035
Zscore (平均交易金额)	-0.017	-0.118	0.664	0.000	-0.013
Zscore (现金交易金额)	0.017	0.430	0.655	-0.001	-0.001
Zscore (个人短期大量金额为整数的交易次数)	0.028	-0.002	0.001	0.999	-0.001
Zscore (短期分散转入集中转出或集中转入分散转出比例)	0.069	-0.001	0.005	-0.001	0.996

Table 4. Common factor composition of risk assessment index system for money laundering
表 4. 银行个人客户洗钱风险评价指标体系公因子构成表

公因子名称	公因子构成指标
交易次数因子(F'_1)	个人短期交易金额或现金交易金额巨大的交易次数、大额转账交易次数、总交易次数
现金交易次数因子(F'_2)	现金交易次数、大额现金交易次数
交易金额因子(F'_3)	交易总金额、平均现金交易金额、平均交易金额、现金交易金额
整数交易因子(F'_4)	个人短期大量金额为整数的交易次数
转入转出因子(F'_5)	短期分散转入集中转出或集中转入分散转出的比例

2.3.3. 基于熵值法的洗钱风险综合评估模型

本文利用熵值法对表 4 中的公因子计算权重, 构建银行个人客户洗钱风险综合评估模型。熵值法能科学客观地计算各公因子的权重, 基本步骤如下:

① 对第 j 项指标进行等分分段, 统计每段范围内所包含的客户个数, 并计算每段范围的概率:
 $P_{ij} = X_{ij} / \sum_{i=1}^k X_{ij}, i = 1, 2, \dots, k; j = 1, 2, \dots, m$, 这里的表示将第 j 项指标等分为 k 段, X_{ij} 表示第 j 项指标第 i 段范围内所包含的客户个数。

② 计算第 j 项指标的熵值: $e_j = -a \times \sum_{i=1}^k P_{ij} \ln(P_{ij})$, 其中 $a = 1/\ln(k) > 0, 0 \leq e_j \leq 1$ 。

③ 计算第 j 项指标的差异系数: $g_j = 1 - e_j$

④ 求第 j 项指标的权重: $w_j = g_j / \sum_{j=1}^m g_j$

由上述熵值法确定各项评价指标权重的基本步骤, 可根据各个公因子得分来计算相应评价指标的熵值与权重, 得到每项评价指标的熵值与权重, 如表 5 所示。

Table 5. Entropy value and weight of evaluation index
表 5. 评价指标的熵值与权重

评价指标	熵值	权重
交易次数因子(F'_1)	0.7672	0.1699
现金交易次数因子(F'_2)	0.7446	0.1864
交易金额因子(F'_3)	0.8293	0.1246
整数交易因子(F'_4)	0.6266	0.2725
转入转出因子(F'_5)	0.6619	0.2467

根据表 5 中所得的权重, 将各个因子评价指标进行加权求和, 得到银行客户洗钱风险的综合评价, 即

$$F_i = \sum_{j=1}^6 F'_{ij} \times w_j \quad (i = 1, 2, \dots, n) \tag{1}$$

3. 实证研究

本文根据重庆市某商业银行 2018 年的客户个人流水数据, 利用上述的洗钱风险综合评价模型, 对个人客户的洗钱风险进行了实证分析。

3.1. 重庆市某商业银行个人客户洗钱综合风险计算

将客户个人流水数据代入模型，计算出了每一位银行客户的洗钱风险值。所有客户洗钱风险综合评价值的最高值约为 99.037，最低值几乎趋近于 0，并且绝大部分银行客户的综合评价价值都分布在[0, 0.3]范围内。这说明绝大部分的银行个人客户的洗钱风险都很小，只有少数的账户存在显著的洗钱风险，这与实际情况相符。由于客户数太大(约 20 万)，仅给出了排名前十名客户的综合风险值，如表 6 所示。

Table 6. Comprehensive evaluation of money laundering risk for top ten customers

表 6. 前十名客户的洗钱风险综合评价得分

客户编号	综合评价价值	排名	客户编号	综合评价价值	排名
60349	99.03712101	1	162278	35.90974416	6
171775	70.32162761	2	73568	30.9320882	7
68555	58.8867813	3	197451	28.50623648	8
75174	52.53483518	4	172424	26.94520984	9
70447	37.14977976	5	150112	22.64177028	10

风险值排名第一的 60349 号客户在短时间内频繁交易，且每次交易都是大额资金交易，都是多个账户向该客户转账，转账金额高达 100 万，因此具有最高的洗钱风险。排名第二的 171775 号客户在短时间内的交易频繁，并且也是其他账户向该客户转账，转账金额高达 124 万，同时转账金额为整数，这些都说明了洗钱风险值高的银行客户均进行了大额资金交易，且短时间内交易频繁，符合洗钱行为特征。

3.2. 洗钱风险分布拟合

为了对该银行所有客户的洗钱风险情况进行整体的评估，需要对洗钱风险进行分布拟合。通过分析洗钱风险综合值的直方图发现，洗钱风险具有类似于广义极值分布的密度，因此采用广义极值分布对洗钱风险分布进行拟合。广义极值分布的分布函数为

$$F(x) = \exp \left\{ - \left[1 + k \left(\frac{x - \mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\} \quad (2)$$

其中 $1 + k \left(\frac{x - \mu}{\sigma} \right) > 0$ ， μ 表示位置参数， σ 和 k 分别表示尺度参数和形状参数。选取所有小于 1 的洗钱风险综合值，用广义极值分布的密度函数进行拟合，如图 1 所示。

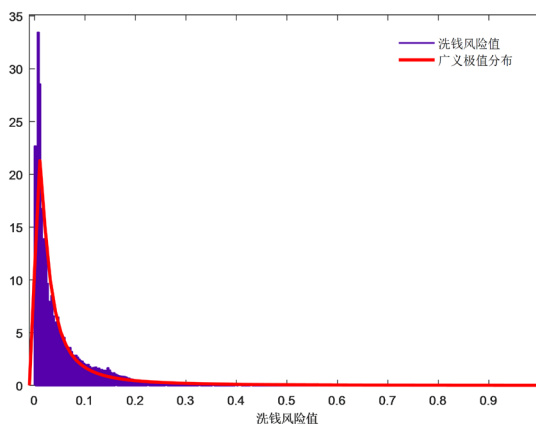


Figure 1. Matched chart of integrated value density function of money laundering risk
图 1. 洗钱风险综合值密度函数拟合图

从图中可以看出, 分布的拟合情况良好, 密度曲线能准确的反应出洗钱风险值的分布情况。最终得到分布函数的拟合结果为: $\mu = 0.018$, $\sigma = 0.022$, $k = 0.860$, 即

$$F(x) = \exp \left\{ - \left[1 + 0.86 \left(\frac{x - 0.018}{0.022} \right)^{0.86} \right]^{-\frac{1}{0.86}} \right\} \quad (3)$$

3.3. 银行个人客户洗钱风险总体评估

利用拟合出来的洗钱风险分布的分布函数, 计算不同显著性水平下客户有洗钱行为的可能性、阈值以及客户人数, 见表 7。

Table 7. Possibility of money laundering at different significant levels and number of clients

表 7. 不同显著水平下洗钱可能性以及客户数

显著性水平	洗钱可能性	阈值	客户数
0.1	0.90	0.172	11,205
0.05	0.95	0.326	5239
0.01	0.99	1.347	681

由表 7 可知, 当选取显著性水平为 0.05 时, 有 5239 位银行客户的洗钱综合风险值大于 0.326, 这说明有 95% 以上的把握认为这些客户有洗钱行为。因为洗钱综合风险值的分布只是近似分布, 这里的洗钱可能性并不等同于客户洗钱的概率, 银行可以根据自己的需求选择恰当的水平 and 阈值。比如, 当阈值选取为 1.347 时, 有 681 位客户可能有洗钱行为, 至于究竟这 681 位客户到底有没有洗钱行为, 还需要银行进一步的分析和确认。

4. 结论

本文从银行客户的个人流水数据出发, 研究了重庆市某商业银行客户的洗钱风险综合评价指标体系, 并通过因子分析法和熵值法最终建立了银行个人客户洗钱风险综合评估模型。利用重庆市某商业银行 2018 年个人客户流水数据, 对客户洗钱风险进行了实证分析, 最后运用广义极值分布对银行个人客户洗钱综合风险值进行分布拟合, 并对该银行个人客户的洗钱风险进行了整体评估。在 0.05 的显著性水平之下, 有 5239 位客户可能存在洗钱行为, 需要银行的进一步确认。也就是说, 相对于总计近 20 万的银行客户, 只有大约 2.66% 的客户需要仔细鉴别是否真正有洗钱行为, 能极大地提高银行的工作效率。

基金项目

重庆市社会科学规划项目(2018BS56)。

参考文献

- [1] 陈云开, 卢正鼎, 刘芳, 等. 一种高维聚类算法及在洗钱侦测中的应用[J]. 计算机科学, 2007(6): 195-197+217.
- [2] 丁佳, 封思贤. 第三方支付机构洗钱风险的评估方法研究[J]. 金融发展研究, 2018, 444(12): 51-58.
- [3] 丁昱. 商业银行个人客户洗钱静态风险分类方法研究[J]. 金融理论与实践, 2017(8): 59-63.
- [4] 肖琨, 王云, 张桂刚. 基于识别和多重分类的反洗钱系统[J]. 小型微型计算机系统, 2019(10): 2046-2051.
- [5] 金大卫. 我国反洗钱监控机制的完善及创新[J]. 经济管理, 2010(4): 28-34.

- [6] 童文俊. 金融机构洗钱风险评估监管体系构建研究[J]. 金融与经济, 2013(2): 73-76.
- [7] 王宝运, 王怡靓. 我国商业银行反洗钱有效性影响因素分析[J]. 金融理论与实践, 2017(7): 99-103.
- [8] 胡爱平, 邱世芳, 伍度志. 重庆市城镇化综合评价指标体系应用研究[J]. 重庆理工大学学报(自然科学), 2016, 30(03): 140-148.