

基于大数据的应用统计在金融风险预测中的实践探索

曲奕

中央民族大学理学院，北京 100081

DOI: 10.61369/SSSD.2025060002

摘 要： 大数据时代的来临，为金融行业带来海量数据资源，应用统计技术与大数据融合，在金融风险预测领域展现出巨大潜力。本文深入剖析基于大数据的应用统计在金融风险预测中的实践，阐述应用统计的优势，详细介绍信用风险、市场风险、操作风险预测中的具体应用，分析实践面临的挑战并提出应对策略，结合实际案例展示应用成效，展望未来发展趋势，旨在为金融机构利用大数据提升风险预测能力、增强风险管理水平提供有益参考。

关 键 词： 大数据；应用统计；金融风险预测；风险管理

Practical Exploration of Applied Statistics Based on Big Data in Financial Risk Prediction

Qu Yi

School of Sciences, Minzu University of China, Beijing 100081

Abstract： The advent of the big data era has brought massive data resources to the financial industry. The integration of applied statistical technology and big data has shown great potential in the field of financial risk prediction. This paper deeply analyzes the practice of applied statistics based on big data in financial risk prediction, expounds the advantages of applied statistics, introduces in detail its specific applications in the prediction of credit risk, market risk and operational risk, analyzes the challenges faced in practice and puts forward corresponding countermeasures, demonstrates the application effects with actual cases, and looks forward to the future development trend. It aims to provide useful references for financial institutions to improve risk prediction capabilities and enhance risk management levels by using big data.

Keywords： big data; applied statistics; financial risk prediction; risk management

引言

在数字化快速发展的当下，金融行业数据量呈指数级增长。据 IDC 预测，全球每年产生的数据量将从 2018 年的 33ZB 增长到 2025 年的 175ZB，金融数据作为其中重要部分，涵盖客户信息、交易记录、市场行情等多方面。传统金融风险预测方法依赖少量结构化数据，难以适应大数据环境，而大数据与应用统计结合，能从海量、复杂数据中挖掘有价值信息，为金融风险预测提供新途径。精准金融风险预测可帮助金融机构提前防范风险，减少损失，增强市场竞争力，维护金融市场稳定。因此，研究基于大数据的应用统计在金融风险预测中的实践意义重大。

一、大数据与金融风险预测相关理论基础

（一）大数据的特点与金融数据类型

大数据具有 Volume（大量）、Velocity（高速）、Variety（多样）、Value（价值）、Veracity（真实性）五个显著特点。在金融领域，数据类型丰富多样，如下。

1. 结构化数据如客户基本信息（年龄、性别、收入等）、交易流水数据（交易金额、时间、对手方等），存储于数据库中，格式规范，便于查询与分析^[1]。

2. 半结构化数据如 XML、JSON 格式的金融报表，有一定结

构但不规则。

3. 非结构化数据占比日益增大，像金融新闻文本、社交媒体上的金融讨论、客户服务语音记录等。这些不同类型数据从多维度反映金融活动，为风险预测提供丰富信息。

（二）金融风险的类型与传统预测方法局限性

金融风险主要包括信用风险（借款人违约可能性）、市场风险（因市场价格波动如利率、汇率、股价变动带来的风险）、操作风险（源于不完善或有问题的内部程序、人为失误、系统故障等的风险）。传统金融风险预测方法有信用评分模型（如 FICO 评分）、风险价值模型（VaR）等。

1. 信用评分模型主要依据客户财务报表、信用记录等少量数据评估信用风险，难以全面反映客户复杂信用状况，对新兴业务和缺乏信用记录客户评估准确性低。

2. VaR 模型假设市场因子波动服从特定分布，在极端市场条件下，实际波动与假设差异大，预测效果不佳。且传统方法处理大数据能力弱，难以应对数据量增长和多样化挑战。

（三）应用统计在金融风险预测中的优势

应用统计通过数据收集、整理、分析和推断，为金融风险预测提供关键支持。

1. 在数据处理阶段，运用数据抽样、清洗技术，从海量金融数据中获取准确、可用数据。统计分析方法如回归分析，能找出风险因素与风险指标间关系，预测风险趋势；聚类分析可对客户或金融产品分类，识别同类风险特征。

2. 相比传统方法，应用统计基于大数据，能综合考虑更多风险因素，挖掘数据深层关联，提高风险预测准确性和及时性^[2]。如在信用风险预测中，结合更多维度数据构建统计模型，更精准评估客户违约概率。

二、基于大数据的应用统计在金融风险预测中的具体应用

（一）信用风险预测

1. 数据收集与预处理

金融机构收集多源数据用于信用风险预测^[3]。

（1）内部数据有客户基本信息、账户流水、贷款还款记录等；

（2）外部数据包括第三方征信报告、公共信用信息、社交媒体数据（部分金融机构探索使用）。收集到的数据需预处理，数据清洗去除重复、错误、缺失值，如通过分析交易流水数据逻辑关系纠正错误交易金额；

（3）数据标准化将不同量纲数据统一尺度，如将客户收入、资产等数据标准化，使数据在模型中权重合理，提升模型性能。

2. 构建信用风险评估模型

逻辑回归模型是经典信用风险评估模型，通过构建线性函数，综合客户多个特征变量预测违约概率，模型简单可解释性强^[4]。

决策树和随机森林模型能处理非线性关系，决策树通过一系列条件判断对客户分类，随机森林集成多个决策树，提高模型稳定性和准确性^[5]。以某银行信用卡客户信用风险评估为例，构建随机森林模型，选取客户年龄、收入、信用历史时长、信用卡使用频率等特征，经训练模型在测试集上准确率达85%，有效识别高风险客户。

深度学习模型如多层感知机（MLP）、循环神经网络（RNN）及其变体长短期记忆网络（LSTM）也应用于信用风险预测^[6]。LSTM 能处理时间序列数据，捕捉客户信用状况随时间变化特征，在预测长期信用风险上有优势。

3. 模型评估与监控

采用准确率、召回率、F1 值、受试者工作特征曲线（ROC）

下面积等指标评估信用风险模型性能。

准确率衡量模型预测正确样本比例，召回率反映模型正确识别正样本能力，F1 值综合两者。

ROC 曲线展示模型在不同阈值下真正率和假正率关系，曲线下面积越大模型性能越好。

模型监控实时跟踪模型预测结果和性能指标，如定期计算模型在新数据上准确率。当模型性能下降，分析是数据漂移（数据分布变化）还是模型过拟合等原因，若是数据漂移，更新数据重新训练模型；若是过拟合，调整模型结构或参数，确保模型持续准确预测信用风险。

（二）市场风险预测

1. 市场数据的获取与分析

金融市场数据来源广泛，证券交易所提供股票价格、成交量等数据，外汇交易市场提供汇率数据，金融资讯平台提供宏观经济数据、行业动态等。

获取数据后进行分析，时间序列分析常用，如自回归移动平均模型（ARMA）及其扩展自回归条件异方差模型（ARCH）、广义自回归条件异方差模型（GARCH）。

GARCH 模型能刻画金融资产收益率波动聚集性和时变性，预测市场风险波动。

对股票市场收益率数据用 GARCH（1,1）模型分析，结果显示模型有效捕捉收益率波动特征，为投资者制定投资策略提供参考。

2. 风险价值（VaR）与条件风险价值（CVaR）模型的应用

VaR 模型衡量在一定置信水平和持有期内，投资组合可能遭受的最大损失。计算方法有历史模拟法、蒙特卡罗模拟法、方差-协方差法。

历史模拟法基于历史市场数据模拟投资组合价值变化计算 VaR；蒙特卡罗模拟法通过随机模拟市场因子变化计算 VaR。

CVaR 模型在 VaR 基础上，考虑超过 VaR 值的损失均值，更全面反映极端风险^[7]。在投资组合管理中，投资机构用 VaR 和 CVaR 模型评估风险，根据不同置信水平下 VaR 和 CVaR 值调整投资组合资产配置，降低市场风险。

3. 压力测试与情景分析

（1）压力测试模拟极端市场情景（如金融危机、利率大幅波动）对金融机构资产组合影响，评估机构承受极端风险能力^[8]。

（2）情景分析设定多种可能市场情景，分析不同情景下金融风险变化。某银行对其贷款组合进行压力测试，假设房价下跌30%、利率上升200个基点极端情景，评估贷款违约率和损失情况，结果显示在该情景下银行不良贷款率上升10个百分点，需提前准备应对措施。通过情景分析，银行制定不同市场情景下风险管理策略，提高应对不确定性能力。

（三）操作风险预测

1. 操作风险数据的整理与特征提取

操作风险数据收集困难且分散，包括内部损失事件报告、审计记录、业务流程监控数据等。整理数据后提取特征，如内部人员违规操作次数、系统故障时长、业务流程复杂度等，这些数据

为操作风险预测提供基础。

2. 基于统计模型的操作风险预测方法

(1) 泊松回归模型常用于操作风险损失频率预测，假设损失事件发生服从泊松分布，通过分析历史数据确定模型参数，预测未来损失事件发生次数^[9]。

(2) 负二项回归模型在泊松回归基础上考虑过度分散问题，使预测更准确。在操作风险损失程度预测上，常用广义线性模型（GLM），根据损失数据分布特点选择合适链接函数和误差分布，构建损失程度预测模型。银行用负二项回归模型预测操作风险损失频率，结合 GLM 模型预测损失程度，经实际验证，模型能较好预测操作风险，为银行制定操作风险准备金提供依据。

(3) 机器学习在操作风险预测中的创新应用

机器学习算法在操作风险预测中创新应用，支持向量机（SVM）能有效处理小样本、非线性问题，在操作风险分类（如区分不同类型操作风险事件）上有优势；

神经网络模型如深度信念网络（DBN）可自动学习数据深层次特征，对复杂操作风险模式识别能力强。金融科技公司利用 DBN 模型对操作风险数据进行分析，识别出传统方法难以发现的操作风险关联模式，提前预警潜在操作风险，降低风险损失。

三、基于大数据的应用统计在金融风险预测中的实践探索——以 H 金融科技公司为例

(一) 系统设计思路

该金融科技公司为金融机构提供市场风险预警服务，系统设计基于大数据实时分析技术。实时收集股票、债券、外汇等市场数据，运用机器学习算法和统计模型分析数据，捕捉市场风险变

化，及时发出预警信号。

(二) 数据来源与实时分析技术应用

数据来源于各大金融交易所、数据供应商和新闻媒体。采用 Apache Flink 流计算框架实时处理数据流，对市场数据实时清洗、分析和建模。如用 LSTM 模型实时预测股票价格走势，当价格波动超过设定阈值，系统自动发出风险预警。

(三) 风险预警模型的构建与验证

构建基于深度学习和统计分析的风险预警模型，综合考虑市场价格、成交量、宏观经济指标等因素^[10]。用历史数据训练和验证模型，通过回测评估模型在不同市场条件下表现。模型在历史数据回测中，准确预警多次市场大幅波动，预警准确率达 85% 以上。

(四) 实际应用成效与客户反馈

系统应用后，帮助金融机构及时调整投资组合，降低市场风险损失。客户反馈系统预警及时性和准确性高，为投资决策提供有力支持。

综上所述，在全球经济一体化和金融市场互联互通背景下，金融风险预测重要性日益凸显。未来基于大数据的应用统计在金融风险预测中的发展，将进一步提升金融机构的服务效率与质量。金融机构可借助更精准的风险预测模型，为不同风险等级的客户提供差异化、个性化的金融产品与服务。投资者能获取更全面、及时的风险信息，了解不同投资产品的潜在风险。总之，基于大数据的应用统计在金融风险预测中前景广阔，虽面临挑战，但随着技术创新、数据资源深度开发、行业协作加强以及监管体系完善，将为金融行业风险管理提供更坚实保障，推动金融行业稳健发展，同时为金融机构和投资者带来实实在在的益处。

参考文献

- [1] 张晓燕, 李楚, 龙亮. 数字金融对系统性金融风险的影响——基于金融监管的调节效应分析 [J]. 财经科学, 2025, (02): 1-15.
- [2] 刘兆轩. 大数据与人工智能在金融风险中的应用 [J]. 今日财富, 2025, (05): 37-39.
- [3] 王田. 大数据在金融风险预测与管理中的应用的相关思考 [J]. 老字号品牌营销, 2024, (21): 96-98.
- [4] 木强, 秦琳琳. 金融风险中的统计模型与方法 [J]. 商讯, 2024, (13): 88-91.
- [5] 郭长冬. 多模多维的大数据驱动型金融风险感知与评估系统. 北京市, 度小满科技(北京)有限公司, 2023-12-01.
- [6] 徐永. 基于机器学习的中国系统性金融风险测算及影响因素研究 [J]. 金融发展评论, 2023, (10): 44-61.
- [7] 陈耀辉, 马凌云. 基于 Copula-GARCH 模型的互联网金融市场风险测度 [J]. 南京财经大学学报, 2021, (01): 22-33.
- [8] 谭中明, 王道, 杨素敏. 基于压力指数的系统性金融风险跨市场传导研究 [J]. 经济论坛, 2023, (07): 139-152.
- [9] 梁丽波, 张小婉, 张瑞恒, 等. 双向开放与银行业风险: 基于泊松回归的“U”型关系再检验 [J]. 西部金融, 2025, (02): 16-21.
- [10] 曾青云. 资本市场量化投资的风险及风险控制策略研究 [J]. 投资与合作, 2022, (06): 4-6.