

数学模型分析对县域经济发展的信贷风险预测研究

杨卓

甘肃农业大学管理学院, 甘肃 兰州

收稿日期: 2024年2月6日; 录用日期: 2024年2月26日; 发布日期: 2024年4月17日

摘要

随着县域经济发展与金融创新的深入, 信贷风险管理日趋复杂化, 县域银行和金融机构面临着越来越多的挑战。基于此背景, 本文旨在通过数学模型对县域经济发展中的信贷风险进行预测研究, 为金融决策提供科学支撑, 本文首先说明了数学模型在县域经济发展信贷风险预测中的应用原则, 并以某县域经济发展基本情况为例, 说明了多元统计分析法数学模型的建立方法, 最后通过效果分析, 证明了本文所提出的数学模型的有效性和可行性, 实现了对县域经济发展中信贷风险的精确预测, 为相关部门提供了可靠的参考框架。

关键词

数学模型, 县域经济发展, 信贷风险预测

Research on Credit Risk Prediction of County Economic Development by Mathematical Model Analysis

Zhuo Yang

College of Management, Gansu Agricultural University, Lanzhou Gansu

Received: Feb. 6th, 2024; accepted: Feb. 26th, 2024; published: Apr. 17th, 2024

Abstract

With the development of county economy and the deepening of financial innovation, credit risk management is becoming more and more complicated, and county banks and financial institutions

are facing more and more challenges. Based on this background, this paper aims to forecast and study the credit risk in the development of county economy through mathematical models to provide scientific support for financial decision-making. This paper first explains the application principle of mathematical model in the prediction of credit risk in county economic development, and takes the basic situation of a county economic development as an example to illustrate the establishment method of mathematical model of multivariate statistical analysis. Finally, through the effect analysis, the validity and feasibility of the mathematical model proposed in this paper are proved, and the accurate prediction of the credit risk in the development of county economy is realized, which provides a reliable reference frame for the relevant departments.

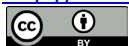
Keywords

Mathematical Model, County Economic Development, Credit Risk Forecast

Copyright © 2024 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. 前言

在全球化的经济环境下，金融市场激烈竞争，银行业面临着巨大的信贷风险，而县域经济，由于受地理位置、行业结构、人口素质等因素影响，其发展情况存在显著差异，信贷风险也各不相同。因此，针对县域经济开展信贷风险预测可帮助金融机构提高资金使用效率，避免坏账损失，但传统的信贷风险评估方法主要依赖单一或少数几个经济指标，这种方法过于简化了问题的复杂性，不能全面反映影响信贷风险的多元因素。随着大数据和计算技术的发展，应用数学模型进行信贷风险预测已成为可能，相关部门可收集大量相关数据，并运用多元统计分析等先进技术处理和分析这些数据，建立更精确、更全面的信贷风险评估模型。

本文旨在开发基于数学模型的县域经济发展的信贷风险预测系统，将收集和整理与县域经济发展相关的各类数据，利用数学模型中的多元统计分析法来评估这些数据中包含的信息，并根据主成分分析结果建立信贷风险预测模型，以此对县域经济中的信贷风险进行预测。

2. 县域经济发展信贷风险预测中数学模型分析的应用原则

2.1. 数据的全面性与代表性

为保证模型分析的有效性，应确保输入数据的全面性与代表性，即数学模型的建立过程应基于信贷历史数据、借款人的财务报告、信用评级以及宏观经济指标等各方面信息，尽可能提高输入数据的全面性。针对县域层面，模型建立还特别需要注意地方经济结构、产业特色和区域政策等因素的影响，比如如果某县以制造业为主导，则该县企业的信贷数据可能会显示出与服务业或农业主导地区不同的风险特征。其中全面性也意味着收集跨时间段的数据，以理解经济周期如何影响信贷风险水平，而且当地文化、法律和社会环境等软数据也应被考虑纳入建模过程中，只有当所用数据能够真实反映借款人在特定环境下的行为时，预测模型才具有更高可靠性[1]。

2.2. 模型选择与优化

对于县域经济发展的信贷风险预测而言，数学模型的选择应基于数据类型、问题复杂度以及预期目

标, 结合县域经济发展实际, 对于大多数信贷风险预测问题, 可以选择线性回归、决策树、支持向量机(SVM)、神经网络等统计或机器学习方法。除了选择合适的模型外, 模型优化方面, 相关人员可调整模型参数、执行交叉验证和采取过拟合防止措施, 比如在神经网络训练过程中, 需要调整网络层数、每层节点数以及学习率等超参数, 以此提高模型的有效性。同时, 在面对县域经济发展中出现新趋势或突发事件时, 模型需要具备一定的灵活性以快速应对变化, 所以管理人员或技术人员还需要通过实时测试来验证模型在新数据上的表现, 并根据实际情况做出相应调整[2]。

2.3. 伦理与合规

相关部门和金融机构在使用数学模型分析进行信贷风险预测时, 必须遵循伦理和法律标准, 以保障客户数据安全和隐私作为基本前提, 所有收集和处理的的数据都需符合《个人信息保护法》, 而且在建立评分和决策系统时要避免歧视任何群体。例如, 不能基于种族、性别、年龄等敏感属性来设定贷款审批标准, 算法设计者需要确保模型不会放大现有不公平现象或创造新的偏见。另外, 在解释模型结果时应尽量透明明晰, 对于那些“黑箱”算法(像深度学习), 研究者和实践者需要开发出方法来解释其决策过程, 使之对用户和监管机构都可理解, 最大限度避免因数学模型应用造成的用户冲突。

3. 县域经济发展信贷风险预测中数学模型分析的应用策略

3.1. 县域经济概况

案例县城作为典型的制造业重镇, 2022年实现GDP增长6.2%, 工业总产值达到4000亿元(GDP约为800亿, 工业总产值与GDP比值约为5:1), 人均GDP超过12万人民币。然而, 在经济快速发展的背后, 该县域的金融支持体系尚显不足, 特别是中小企业融资难、融资贵的问题突出。根据最新统计, 该县中小企业贷款违约率为4.3%, 高于全国平均水平, 基于该指标, 县内银行和金融机构亟需精确预测信贷风险, 优化信贷资源配置。在此背景下, 该县相关部门采用多元统计分析法, 以该县中小企业的经营数据为样本, 选择了营业收入增长率、资产负债率、现金流量比率等关键财务指标作为模型输入, 并进行数据预处理, 包括缺失值处理、异常值剔除与数据标准化, 采用维度压缩技术如主成分分析(PCA)被运用来降低数据维度并提取主要信息, 最后建立逻辑回归模型来预测信贷违约概率, 并通过交叉验证方法评估模型的泛化能力。

3.2. 确定关键指标

针对案例县域, 其制造业是主要特征, 此显著特征给关键指标的选择提供了方向, 所以相关技术人员在确定信贷风险预测模型的关键指标时, 应从宏观和微观两方面入手, 宏观指标通常包括全县的经济增长率、工业总产值等, 反映整个地区经济发展水平的信息, 而微观指标则需要深入到中小企业层面, 涉及财务状况、运营效率和市场竞争力等。具体而言, 以制造业为核心的案例县域可以首先搜集中小企业过往几年内的详细财务报表, 比如利润表、资产负债表、现金流量表以及相关的比率分析报告, 并利用统计学中的相关性分析来识别那些与信贷违约率呈现显著正相关或负相关关系的财务指标, 以此帮助挑选出对信贷风险具有强预测能力的变量, 再通过研究历史上违约企业的财务特征, 案例县域可进一步验证初步筛选出的指标在违约情况下的表现是否符合预期[3]。

3.3. 数据预处理

确定关键指标后, 案例县域的数据预处理任务包括三个关键步骤: 缺失值处理、异常值剔除与数据标准化。其中缺失值处理方面, 由于案例县域的中小企业存在数据不全的情况, 对于缺失值, 使用均值

填充法对连续变量进行处理, 技术人员可使用相关指标的平均数替代缺失值, 对于离散变量, 则采用众数填充法。如果某一指标的缺失值过多, 超过其总数据点的一定比率(如 30%), 则考虑删除该指标以避免引入过多噪声, 而且在特定情景下也可采用更为复杂的插补方法, 如使用相应行业或类似企业的属性进行回归填充[4]。

异常值剔除方面, 可通过箱形图探测法找出超出 1.5 个四分位距离的极端值, 并仔细检查这些值背后的原因, 如果异常值来源于录入错误或者是非典型事务(例如一次性大额交易), 将其从样本中剔除, 如果异常值反映了实际经济活动且属于正常波动范围内, 则保留, 以防损害数据真实性。

数据标准化方面, 为避免未经处理的不同规模和测量单位的原始数据导致模型收敛速度慢甚至无法收敛的问题, 案例县域可采取 Z 分数标准化方法, 按照每个指标的均值和标准差调整数据分布, 将每个观测值减去其所在指标的平均值并除以标准差, 使得各指标呈近似正态分布, 并且均值为 0, 方差为 1。

3.4. 维度压缩

案例县域在进行信贷风险预测时, 由于选取的多个关键财务指标之间可能存在高度相关性, 直接使用这些数据可能导致多重共线性问题, 为了解决这个问题并提高模型的效率和准确性, 案例县域采用主成分分析(PCA)等维度压缩方法来处理数据。

首先, 案例县域将营业收入增长率、资产负债率、现金流量比率等财务指标的原始数据集合做标准化处理, 即每个指标减去其均值并除以其标准差, 确保所有变量具有相同的尺度, 使它们在 PCA 中具有等权重。在此基础上构建协方差矩阵或相关系数矩阵来评估各个财务指标之间的线性关系强度, 再根据协方差矩阵, 应用特征值分解找出主成分, 选择那些具有最大特征值的成分(该特征值解释的方差数量较多), 以此种方法案例县域可以捕捉到最重要的信息, 并且舍弃那些影响较小的组成部分[5]。

其次, 在选择主成分数量时, 案例县域技术人员可采用累积贡献率来确定保留的主成分数目, 只保留累积贡献率达到 85% 以上的主成分便可足够代表原始数据集中绝大多数的信息。再应用所选的主成分, 生成新的低维度特征数据集, 该数据集包含与初始变量不同, 但是仍然反映了大部分信息的新变量集合, 如此便减少了后续逻辑回归模型计算复杂度, 并可能提高预测精度。最后, 在实施 PCA 并得到了压缩后的数据集之后, 技术人员可使用这些新生成的特征作为输入来构建逻辑回归模型来预测信贷违约概率。

3.5. 模型建立与验证

案例县域根据预处理和维度压缩后得到的数据集合, 构建了逻辑回归模型, 该模型适合于二分类问题, 在信贷风险预测领域内具有良好的表现, 技术人员可拟合逻辑函数来预测实际值为 0 或 1 的概率, 其公式可以表示为:

$$P(Y=1) = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

公式(1)中的 $P(Y=1)$ 指定事件发生的概率(例如企业违约), β_0 是截距项, 而 $\beta_1, \beta_2, \dots, \beta_n$ 是各个自变量(如营业收入增长率、资产负债率、现金流量比率等)对应的系数。验证方面, 案例县域可实施以下步骤以验证该逻辑回归模型: ① 将数据分割成训练集和测试集, 训练集用于训练模型, 而测试集则保留作为最终评估使用。② 利用训练集的数据拟合逻辑回归模型。这个过程涉及最大似然估计(MLE)来评估参数 β_i 的值。③ 采用交叉验证技术进一步评估模型性能, 比如使用 k-fold 交叉验证, 将训练集随机分割为 k 个大小相同的子样本, 轮流将其中一个子样本作为验证数据, 其余 $k-1$ 个子样本作为训练数据。④ 使用指标来评价模型性能, 比如准确率、召回率、F1 分数和 ROC 曲线下面积(AUC), 以此提供模型在不

同阈值下判断正类和负类的能力。⑤ 根据交叉验证结果调整模型参数，优化模型表现，结合实际改变正则化强度、调整学习速率等。

3.6. 结果解释与应用

在解释模型结果方面，案例县域可对逻辑回归模型中每个变量的系数进行分析，在公式(1)中输入测试企业的基本数据，将模型中得到的概率转换为分类结果，由管理人员设置阈值(例如 0.5)，当预测概率大于该阈值时，将企业划分为高风险群体，反之，则为低风险群体，以平衡风险识别的敏感性和特异性。在模型验证完成且性能合格后，案例县域可将此预测工具应用于日常信贷审批流程中，银行和金融机构可以利用该模型评估新申请贷款客户或现有债务人的违约风险，优化信贷资源配置。若某企业被预测为高风险，则可能需要进一步审查或要求额外担保。相反，对于那些被判断为低风险的企业，则可以提供更优惠的贷款条件。

3.7. 效果分析

在应用了逻辑回归模型之前，该县中小企业贷款违约率为 4.3%，明显高于全国平均水平，采用多元统计分析法后，关键财务指标被用来预测信贷风险并优化信贷资源配置，风险控制效果如表 1，由表可知在逻辑回归模型投入使用后一年内，贷款违约率下降了 1.2 个百分点。中小企业获得贷款，反映了银行和金融机构对信贷申请的审批能力有所提升，更多的合适借款人得到了资金支持。预警企业比例，表示逻辑回归模型计算得出的有可能违约的企业比例，应用该数据模型后，所申请的 15%的企业被标记为可能会违约。成功拦截高风险贷款数代表标记为高风险的企业最终未获得贷款，减少了潜在坏账。

Table 1. Effect analysis

表 1. 效果分析

年份	贷款违约率	中小企业获得贷款数(亿元)	预警企业比例(%)	成功拦截高风险贷款数(亿元)
2022	4.3%	100	未实施数学模型预测	未实施数学模型预测
2023	3.1%	120	15%	18

4. 结束语

综上所述，本文以某县域经济情况为例，说明了如何利用多元统计分析法构建基于数学模型的县域经济发展的信贷风险预测系统，为金融机构以及相关部门提供有效工具以改善其信贷决策过程。最后通过效果分析，证明了本文所构建的数学模型对县域经济发展中信贷风险预测有一定效果，相关部门和人员可运用该模型实现更客观、更精确地评估每个借款人或者每个企业的信贷风险水平，帮助金融机构制定适当的信贷策略以控制风险。但考虑到市场经济发展受多种因素影响，且部分因素为不可抗力，所以金融机构以及相关部门在使用本模型时，应结合实际经验和市场环境进行判断，并根据新收集的数据定期对模型进行更新和修正。未来，随着信息技术的发展，相关技术人员可尝试将信息技术应用到数学模型建立中，进一步提高信贷预测的准确性，加强其对于外部因素的鲁棒性，以此为相关部门或金融机构提供更为有效的风险预测数学模型，保障地区经济稳定发展。

参考文献

- [1] 张程浩. 数字经济发展对中部六省制造业升级的影响研究[D]: [硕士学位论文]. 郑州: 河南财经政法大学, 2023.
- [2] 肖引昌. 基于数学模型技术的农业大数据分析研究[J]. 大众标准化, 2023(7): 155-157+160.

- [3] 宫毓雯, 刘国晖, 王益成. 基于问卷调查的国际经济学中数学理论模型教学探析[J]. 长春师范大学学报, 2023, 42(2): 187-190+195.
- [4] 文江盼. 数字金融、区域创新与经济发展联动关系——基于广东省面板数据 PVAR 动态模型分析[J]. 中国商论, 2022(24): 120-123.
- [5] 郑丹丹. 基于混合模型的国家经济发展的分类比较[D]. [硕士学位论文]. 曲阜: 曲阜师范大学, 2022.