应用统计与数据科学

Applied Statistics and Data Science



ART AND TECHNOLOGY PRESS INC.

(517 666 0904)

263 S KENWOOD ST 560

CASPER,WY 82601

Copyright © 2026 by ART AND TECHNOLOGY PRESS INC. (United States)

Complimentary Copy

ART AND TECHNOLOGY PRESS INC
(United States)

编委会名单

主　　编



王国长（Guochang Wang），暨南大学（Jinan University），
邮箱：wanggc023@amss.ac.cn
研究兴趣：函数型数据分析、时间序列、充分性降维、统计与机器学习
functional data analysis, time series, sufficiency dimension reduction, statistics and machine learning

执行副主编



张兴发（Xingfa Zhang），广州大学（Guangzhou University），
邮箱：xingfazhang@gzhu.edu.cn
研究兴趣：金融统计、环境统计、机器学习
financial statistics, environmental statistics, machine learning

编　　委



胡志勇（Zhiyong Hu），广州大学（Guangzhou University）
邮箱：zhyhu65@163.com
研究兴趣：大数据，人工智能，财务与会计
big data, artificial intelligence, finance and accounting



雷田礼（Tianli Lei），深圳职业技术大学（Shen Zhen Polytechnic University）
邮箱：ltl@szpu.edu.cn
研究兴趣：数量经济、高职数学教育
auantitative economics，higher vocational mathematics education



舒连杰（Lianjie Shu），澳门大学（University of Macau），
邮箱：ljshu@um.edu.mo
研究兴趣：量化金融，统计学习，质量控制及管理
quantitative finance, statistical learning, quality control and management



王纬（Wei Wang），山东财经大学（Shandong University of Finance and Economics）
邮箱：wangwei_0115@sdufe.edu.cn
研究兴趣：统计机器学习、应用统计、计量经济
statistical machine learning，applied statistics, econometrics



杨凯（Kai Yang），长春工业大学（Changchun University of Technology）
邮箱：yangkai@ccut.edu.cn
研究兴趣：时间序列分析、高维数据分析、贝叶斯分析
time series analysis，high-dimensional data analysis，bayesian analysis



周彦（Yan Zhou），深圳大学（Shenzhen University）
邮箱：zhouy1016@163.com
研究兴趣：生物统计，数据科学
Biostatistics，data science



李永明（Yongming Li），上饶师范学院（Shangrao Normal University）
邮箱：lym1019@163.com
研究兴趣：非参数统计，金融统计，极限理论与统计大样本性质，小波方法
nonparametric statistics，financial statistics，limit theory and statistical large sample theory，wavelet method



刘照德（Zhaode Liu），广东财经大学（Guang dong University of Finance & Economics）
邮箱：lzhaode@163.com
研究兴趣：经济统计、多元统计、计量分析
economic statistics，multivariate statistics，econometric analysis



王小刚（Xiaogang Wang），北方民族大学（North Minzu University）
邮箱：wongone@163.com
研究兴趣：经济与社会统计
economic and social statistics



严晓东（Xiaodong Yan），西安交通大学（Xi' an Jiaotong University）
邮箱：yanxiaodong@xjtu.edu.cn
研究兴趣：统计机器学习、智能体统计学
statistical machine learning，agent statistics



朱柯（Ke Zhu），香港大学（The University of Hong Kong）
邮箱：mazhuke@hku.hk
研究兴趣：统计建模、金融时间序列分析、计量经济、金融大数据
statistical modeling, financial time series analysis, econometrics, financial big data

目录 CONTENTS

应用统计与数据科学

Applied Statistics and Data Science

第2卷 第10期　2026年10月刊

主办单位：ART AND TECHNOLOGY PRESS INC. (United States)

承办单位：广州统计师事务所

指导单位：广东省现场统计学会、广州市统计学会

主　　编：王国长

执行副主编：张兴发

编　　委：胡志勇、李永明、雷田礼、刘照德、舒连杰、王小刚、王　纬、严晓东、杨　凯、朱　柯、周　彦

地　址：263 S KENWOOD ST 560 CASPER,WY 82601

网　址：https://arttechpress.com

本刊说明：

凡向本刊所投稿件，全体作者需签署论文著作权转让声明书和论文发表承诺书，声明、承诺及相关事项如下：

1. 作者将论文的复制权、发行权、网络传播权、翻译权、汇编权、信息网络传播权、改编权等著作权在世界范围内免费转让给本刊。
2. 论文不侵犯他人著作权和其他权利，否则作者将承担由此产生的全部责任，并赔偿由此给出版单位造成的全部损失。
3. 论文署名作者享有该作品的完全著作权，署名作者的身份真实。
4. 论文未曾以任何形式公开发表过。
5. 作者所投本刊稿件，本刊编辑部拥有修改权。

统计与教育教学 | STATISTICS AND EDUCATION TEACHING

001　**OBE 导向多谱协同与 AI 赋能《数值分析》课程建设实践**　　**杨丹丹，华志强**
OBE-oriented Multi-spectrum Collaboration and AI-enabled Construction Practice of Numerical Analysis Course　　Yang Dandan, Hua Zhiqiang

数据科学技术、方法与应用 | DATA SCIENCE TECHNOLOGY, METHODS AND APPLICATIONS

004　**超越工具论：政务 AI 应用的认知误区、适配挑战与数据基础构建**　　**谷剑芳，张阳，王丹丹，袁云云，金玲，杨奎**
Beyond Instrumentalism: Cognitive Misconceptions, Adaptation Challenges, and Data Infrastructure Construction in Government AI Applications　　Gu Jianfang, Zhang Yang, Wang Dandan, Yuan Yunyun, Jin Ling, Yang Kui

012　**融合机器学习与结构方程模型的少数民族文化高质量发展指数（CHDI）构建与验证**　　**邱紫玥，刘付靖岚，倪一丹，张振华**
Development and Validation of the Cultural High-Quality Development Index（CHDI）for Ethnic Minorities Using a Combination of Machine Learning and Structural Equation Modeling　　Qiu Ziyue, Liufu Jinglan, Ni Yidan, Zhang Zhenhua

025　**地貌格局约束下秦直道布线规律与最优路径重构研究**　　**黄文乐，刘泽宜，马伊琳，邵涵，杨永霞**
Research on Layout Rules and Optimal Path Reconstruction of the Qin-Zhidao under Geomorphological Pattern Constraints　　Huang Wenle, Liu Zeyi, Ma Yilin, Shao Han, Yang Yongxia

031　**基于机器学习与 TPB 理论的青年非遗疗愈消费意愿及影响机制研究**　　**高坚国，朱金桂，詹富婕，王佩君，欧阳睿，周大镯，林培琪**
Young Consumers' Intention toward Intangible Cultural Heritage-Based Wellness Consumption and Its Influencing Mechanisms: A Study Based on Machine Learning and the Theory of Planned Behavior　　Gao Jianguo, Zhu Jingui, Zhan Fujie, Wang Peijun, Ouyang Rui, Zhou Dazhuo, Lin Peiqi

统计方法及应用 | STATISTICAL METHODS AND APPLICATIONS

038　**环境政策收紧与生产率动态——基于 OECD 国家的面板局部投影**　　**莫济瑜，张少谦，蔡梓萱，邝东阳**
Environmental Policy Tightening and Productivity Dynamics: Panel Local Projection Evidence from OECD Countries　　Mo Jiyu, Zhang Shaoqian, Cai Zixuan, Kuang Dongyang

046　**公共资金项目绩效评价中服务对象满意度调查体系设计与执行机制研究**　　**商黄鹤，谭怡**
Research on the Design and Implementation Mechanism of Service Recipient Satisfaction Survey System in Public Fund Project Performance Evaluation　　Shang Huanghe, Tan Yi

055　**一种稳健多变量自相关控制图设计方法及其在空气质量监控中的应用**　　**王志坚，张钰哲，杜玮**
A Design Method for Robust Multivariate Autocorrelation Control Charts and Its Application in Air Quality Monitoring　　Wang Zhijian, Zhang Yuzhe, Du Wei

062　**何为企业数字技术产出的创新鸿沟？——基于多种可解释模型的比较分析**　　**刘佳仪，陈丽宇，黄颖，程从华**
What Constitutes the Innovation Gap in Digital Technology Output of Enterprises? A Comparative Analysis Based on Multiple Interpretable Models　　Liu Jiayi, Chen Liyu, Huang Ying, Cheng Conghua

075 **考虑设备转运与购置决策的多车间协同整修调度研究** **魏陆炀**
Multiworkshop Maintenance Scheduling Considering Equipment Transfer and Acquisition Decisions Wei Luyang

080 **外贸风险预警的政企协同法治保障研究** **吴乔林，章津**
Research on the Legal Guarantee of Government-Enterprise Collaboration in Foreign Trade Risk Early Warning Wu Qiaolin, Zhang Jin

OBE 导向多谱协同与 AI 赋能《数值分析》课程建设实践

杨丹丹，华志强
内蒙古民族大学 数学科学学院，内蒙古 通辽 028000
DOI:10.61369/ASDS.2026100001

摘　　要： 《数值分析》为我校信息与计算科学专业核心基础课程，课程存在理论抽象、实践薄弱、分层教学难等问题。课程团队立足 OBE 理念，构建知识、问题、能力、思政多谱协同体系，借助 AI 技术辅助且以 Python 语言为工具引导学生结合现实热点问题完成算法实践，实现数值方法原理理解、代码实现、问题求解与价值塑造一体化培养。

关 键 词： OBE；知识图谱；AI 赋能；数值分析；课程建设

OBE-oriented Multi-spectrum Collaboration and AI-enabled Construction Practice of Numerical Analysis Course

Yang Dandan, Hua Zhiqiang
College of Mathematics Science, Inner Mongolia Minzu University, Tongliao, Inner Mongolia 028000

Abstract: Numerical Analysis is a core basic course for the major of Information and Computational Science in our university. This course is characterized by abstract theories, insufficient practical training, and difficulties in stratified teaching. Based on the Outcome-Based Education (OBE) concept, the course team constructs a multi-spectrum collaborative system integrating knowledge, problem-solving, ability cultivation, and ideological and political education. With the assistance of artificial intelligence technology and taking Python as the teaching tool, the course guides students to conduct algorithm practice combined with real-world hot issues, realizing the integrated cultivation of theoretical understanding of numerical methods, programming implementation, problem solving and value shaping.

Keywords: OBE; knowledge graph; AI empowerment; numerical analysis; curriculum construction

引言

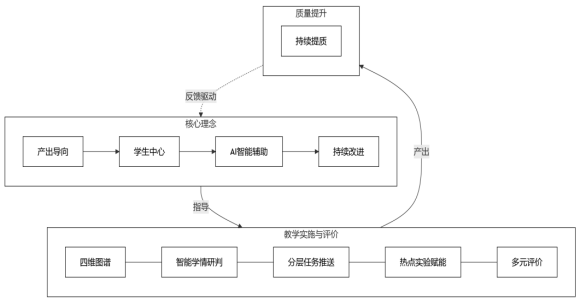
在教育数字化转型与新工科建设深入推进的背景下，利用人工智能、知识图谱等新一代信息技术赋能传统课程教学，已成为课程教学改革的重要方向。《数值分析》是我院信息与计算科学专业的核心基础课程，是后续课程学习的重要基石，总学时72学时（理论56学时，实验16学时），主要讲授插值、拟合、数值积分、线性方程组求解、非线性方程求根、常微分方程数值解法等内容，旨在培养学生运用数值方法近似求解数学问题的能力。然而，该课程的教学面临以下问题：理论抽象，课程内容涉及误差、收敛性、稳定性等抽象数学性质，同时包含复杂的算法构造逻辑。板书与常规课件仅能展示算法流程推演，误差推导、收敛证明等内容，但难以完整呈现知识间的内在关联，学生容易机械记忆公式步骤，难以把握算法本质；知识碎片化，各章节内容相对独立，知识点之间缺乏清晰关联，学生难以构建完整的课程认知框架；实验内容陈旧，缺乏时事热点与前沿技术融合；教学适配困难，学生数学基础差异大，传统教学采用统一内容、统一训练、统一考核的模式，授课内容、节奏与难度难以兼顾全体学生，导致学生的学习成绩差异越来越大，因材施教难以落地[1-3]。

基金项目：内蒙古自然科学基金面上项目（2025），“次线性期望空间下 WOD 随机变量序列的完全 f- 矩收敛性的研究”（项目编号2025MS01025）；内蒙古民族大学教育教学研究重点课题，“数学学科优质培育拔尖计划的研究”（项目编号 ZD2024001）；内蒙古民族大学博士启动基金项目（项目编号 BS650）；2026年自治区直属高校基本科研业务费项目，“面向不完备信息系统的三支多属性智能决策模型研究与应用”（项目编号 GXKY26Z036）。
作者简介：
杨丹丹(1984-)，女，内蒙古民族大学数学科学学院，讲师，主要从事数据挖掘、粒计算；
华志强(1982-)，男，内蒙古民族大学数学科学学院，副教授，主要研究方向概率极限理论和多元统计分析。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

一、《数值分析》课程建设实施路径

为解决传统教学难题、落实 OBE 成果导向育人要求，课程团队系统梳理课程核心内容、知识点之间的逻辑关联、教学重难点与考核要点，搭建“课程—章节—知识点”三级层级知识架构，并融合问题、能力、思政维度图谱，构建四维一体化知识体系，为课程数智化、可视化教学筑牢底层支撑 [4-6]。基于该体系，《数值分析》课程利用四维图谱搭建多谱协同教学框架，紧扣课程育人核心目标、聚焦学生学习产出，构建起“目标—活动—评价—改进”的数据驱动教学闭环（下图）。本次课程建设深度融合先进育人逻辑与 AI 大数据技术，通过四维图谱教学赋能、智能学情研判、分层任务推送、热点实验革新等数智化举措优化教学模式，下文将系统阐述本次教学改革的具体实施路径。



图：《数值分析》课程建设措施

课前阶段，为了丰富原有单一的课本预习形式，借助课程知识图谱搭建的预习体系，设置多元化自主预习任务，充分发挥知识图谱可视化、结构化、关联化的教学赋能作用。首先基于课程知识图谱发布前置知识点预习视频、导学课件、思维导图梳理或者基础题的测验等，系统采集学生预习完成度、答题情况等过程数据，通过数据能分析学生课前知识掌握水平，能精准定位共性薄弱点与个体学习差异。引导学生自主梳理预习知识点绘制思维导图搭建个人认知体系，并鼓励学生将预习过程中遇到的疑难问题及时与老师和同学沟通交流及时解决。同时将学生预习完成度、自测正确率等数据纳入过程性考核，进一步丰富考核维度与形式。

课中教学严格遵循 OBE 成果导向教育理念，围绕学生知识掌握、能力提升和素养培育的核心目标，以“知识图谱 + 思政图谱”双谱协同辅助课堂教学。知识图谱用于可视化呈现重难点和梳理知识逻辑关系等，帮助学生搭建起完整课程体系；思政图谱用于挖掘课程育人元素，将工匠精神、家国情怀、职业规范等有机融入教学，实现知识教学与价值引导同步推进。在此基础上，坚持以学生为主体，多种教学方式相结合，常常运用任务驱动、小组讨论、随堂测验、师生互动问答等方式教学，调动学习的积极性与主动性，引导学生独立思考、灵活运用所学知识。以复化梯形公式教学为例，教师以“如何提高梯形公式精度”为任务驱动，引导学生借助知识图谱自主梳理出“梯形公式—复化梯形公式—余项估计—收敛阶—步长”的知识脉络。学生通过小组合作推导复化梯形公式及其余项公式；在师生互动中，由余项公式分析得出：步长减半时，误差理论上约降为原来的四分之一。随后，学生列表计算几组数据加以验证，直观感受这一收敛规律。随后，围绕“步长是否越小越好”展开探究讨论，让学生进一步认识到计算量与舍入误差的互相制约。此时，在“步长与误差”节点引入“不积跬步，无以至千里；不积小流，无以成江海”思政图谱案例，说明复化梯形公式正是将整体积分化为若干小梯形逐步累加，体现积小成多、量变到质变；也启示学生脚踏实地、注重积累，同时把握适度、讲求效率。随后，教师借助学习通发布随堂测验，检测学生对复化梯形公式及收敛规律的理解。教学过程中，借助学习通平台实时收集学生答答题、互动、小组合作等数据，生成学情分析，据此调整教学进度与重点，针对薄弱内容补充讲解。课程实行全过程评价，将课堂参与、小组协作、随堂测试、思政素养等纳入过程性考核，以多元评价提升学生专业知识、综合能力与核心素养，达成 OBE 育人目标 [7-8]。

课后巩固环节，学习通 AI 助教 24 小时全天候智能答疑，助力学生精准纠错、深化理解。同时，通过作业智能查重等功能，规范课后学习行为，保障过程考核公平公正。通过学习通的任务引擎模块，以学生每次单元测验成绩为主要依据，为不同知识水平的学生推送初、中、高三个层次差异化的课后学习任务，实现因材施教目的 [9]。初阶任务面向基础薄弱学生，聚焦核心概念理解与基础算法应用，帮助学生夯实知识根基；中阶任务面向中等水平学生，侧重提升问题解决能力；高阶任务面向能力突出学生，旨在拓宽视野与创新思维培养。分层任务并非固定标签，而是随每次单元测验和学情数据动态调整，有效破解传统教学“一刀切”难题。课后学习形式丰富多样。课堂知识图谱梳理的知识脉络，为课后分层任务提供内容框架；思政图谱中的育人元素，为高阶任务中融入科学精神与责任意识提供引导。学生可在学习通讨论区围绕复化梯形公式的收敛性与计算效率等话题展开异步交流。平台还支持课后拓展阅读资料推送、小组协作任务提交等功能，满足不同学生的学习需求。教师借助平台数据统计和学情功能，了解各层学生的任务完成率、讨论参与度与质量，对任务难度及时调节，确保各层次学生均能获得适度挑战。课后巩固成为学生个性化成长的重要环节，切实提升专业知识、综合能力与核心素养。

实验环节使用 Python 语言进行编程，基于 OBE 理念反向设计实验体系，围绕学生岗位适配能力和前沿技术应用能力等核心产出目标，摒弃传统陈旧的验证性实验内容，引入问题图谱重构实验教学逻辑，借助问题图谱将真实场景中与数值分析相关的热点实际问题分层梳理，按章节知识点与具体解决方法归类，清晰呈现问题与方法的对应关系。实验内容围绕时事热点问题设计。比如气象数据拟合、环境监测数据分析等实验，学生拿到的不再是编好的例题，而是带缺失、带噪声的真实数据。从数据的预处理到建模再到求解和验证，大幅提升了实验的真实性与挑战性。学生需要独立完成实操流程，对学生来说是一个全新的挑战，但是这样设计的目的是倒逼学生融会贯通所学理论知识，实现综合性应用。有时为了最终解决问题除了应用数值分析方法还将人工智能经典算法（分类、聚类等）融入实验 [10,11]。例如在气象预测

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

实验中，采用插值、最小二乘拟合实现气象要素数值预估，结合随机森林分类算法完成天气类型判别。通过数值方法与智能算法开展实验，学生在完成实验的过程中既夯实了数值分析基础，又对接了人工智能前沿技术，大幅提升综合应用能力。这样岗位适配能力和前沿技术应用能力同步提升，综合应用能力也会明显增强，OBE 所强调的成果导向也就真正落到了实处。

二、结语

《数值分析》课程改革坚持以学生为中心，以 OBE 成果导向理念为核心，以四维知识图谱与人工智能赋能辅助教学，构建精准教学与持续改进闭环体系。改革遵循反向设计思路，搭建课前多元预习、课中探究教学、课后 AI 辅助、分层育人与智能算法融合实验的教学模式，利用 AI 技术优化教学实施，破解传统课堂同质化、重结果轻过程、实验内容陈旧等痛点。实践证明，“多谱协作 + AI 赋能”数智化模式能够激发学生学习主动性，强化算法实践能力，减负增效，全面提升课程育人质量。

参考文献

[1] 陈丽娟，王丽莎，张蕾，等 . 研究生数值分析课程建设实践 [J]. 高教学刊 ,2026,12(3):70-73.

[2] 傅钲淇，赵建昕，闫盼盼 . 智慧教学背景下研究生数值分析课程新形态教材建设论析 [J]. 大学教育 ,2025,(23):76-80.

[3] 周亚东，万明杰，康娟 . 以应用为导向的工科研究生数值分析课程教学改革与实践 [J]. 高教学刊 ,2026,12(25):146-149.

[4] 尹秀玲，赵春宇，李秋萍，等 .OBE 理念下的数值分析课程实践教学改革 [J]. 高师理科学刊 ,2025,45(10):57-63.

[5] 张波，陈维愿，蒋红，等 . 基于知识图谱的“结构力学”课程教学改革研究 [J]. 科技风 ,2026,(9):71-73.

[6] 郎迪，周晓晶，徐艳，等 . 思政元素精准融入视角下《数值分析》教学改革研究 [J]. 统计与咨询 ,2026,(3):33-36.

[7] 刘益林，董云桥，罗彪，等 .“数值计算方法”课程教学改革与实践研究的探索 [J]. 科技风 ,2026,(8):81-83.

[8] 伏文清，张丽丽，路畅 . 基于知识图谱的高等数学数字化教学探索与实践 [J]. 大学数学 ,2026,42(3):61-66.

[9] 付涛，陈久朋，王森，等 . 数值分析课程与人工智能技术深度融合的教学改革探索 [J]. 高教学刊 ,2025,11(30):156-161+166.

[10] 马维元，汤玉荣 . 大数据驱动下数值分析课程的教学改革探讨 [J]. 大学 ,2022,(26):193-196.

[11] 盛余洋，林轶 . 基于知识图谱与 AI 技术的 Python 程序设计课程教学新生态构建研究 [J]. 信息与电脑 ,2026,38(18):233-235.

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

超越工具论：政务 AI 应用的认知误区、适配挑战与数据基础构建

谷剑芳[1]，张阳[2*]，王丹丹[3]，袁云云[4]，金玲[5]，杨奎[1]
1. 河南省政务大数据中心，河南 郑州 450052
2. 河南省数据局，河南 郑州 450052
3. 河南科技大学商学院，河南 郑州 450052
4. 河南省遥感院，河南 郑州 450052
5. 河南省自然资源厅，河南 郑州 450052
DOI:10.61369/ASDS.2026100002

摘　　要： 2025年是一个明显的分水岭。DeepSeek 等国产模型在中文语境下的表现显著提升，私有化部署的门槛又降了下来，各地政务系统引入 AI 的节奏随之加快，智能客服、工单分派、公文辅助一类应用密集上线。问题也随之暴露：不少系统上线即闲置，答非所问，知识库长期不更新。笔者在参与省级政务数据治理与多地调研中感到，这类失败多半不在技术侧，而在决策者对大模型能力边界的判断上。本文据此梳理出三类偏差：最要紧的是把大模型当成查得准、不会错的政策知识库，忘了它的答案其实是按概率拼出来的，不是从库里查出来的；其次是立项时先看指标参数、后看业务痛不痛，又犯了只管建、不管用的老毛病；还有一条常被忽略——按“验收即完”的传统信息化方式管 AI，既没留数据更新的口子，也没留模型迭代的钱。为检验这三类偏差能解释多少结果，本文对比了四个样本：先看福田那批“AI 公务员”；再看漳州的“候喜”；杭州城市大脑是第三个；还有一个不在政务系统里，是沃尔玛。沃尔玛不在政务系统内，放进来是为了隔离一个变量——当数据不缺、预算不缺，工具本身碎片化会造成什么后果。案例分析沿“场景嵌得多深”与“平台统得多广”两条线索展开，由此归纳出需求、场景、数据、平台四个落地条件，并做成一张四维评分表（业务价值40%、数据基础25%、技术可行性20%、风险可控性 15%），供地方在立项前对候选场景打分排序。文章最后讨论支撑条件，认为标准、数据、平台三件事里，从数据供给侧到模型消费侧的那段加工链路最容易被跳过，也最值得先投。相比已有研究，本文想留下的不是又一个技术框架，而是一张能拿去开会用的打分表。

关 键 词： 人工智能；政务服务；大语言模型；认知误区；场景适配；高质量数据集；平台治理

Beyond Instrumentalism: Cognitive Misconceptions, Adaptation Challenges, and Data Infrastructure Construction in Government AI Applications

Gu Jianfang[1], Zhang Yang[2*], Wang Dandan[3], Yuan Yunyun[4], Jin Ling[5], Yang Kui[1]
1. Henan Provincial Government Big Data Center, Zhengzhou, Henan 450052
2. Henan Provincial Administration of Government Services and Data Management, Zhengzhou, Henan 450052
3.School of Business, Henan University of Science and Technology, Zhengzhou, Henan 450052
4. Henan Provincial Institute of Remote Sensing, Zhengzhou, Henan 450052
5. Department of Natural Resources of Henan Province, Zhengzhou, Henan 450052

Abstract: The arrival of DeepSeek and comparable Chinese-built large language models in 2025 changed how fast local governments in China began experimenting with AI. Deployment had become cheap enough, and private deployment secure enough, that smart customer service, work-order triage and drafting assistance moved from pilot to routine inside a single budget cycle. What has received far less attention is how unevenly these systems perform once they are in service. Drawing on the authors' own work in provincial government data governance and on field visits to Shenzhen, Zhangzhou, Hangzhou and Qingdao, this paper argues that the failures are not primarily algorithmic.

作者简介：
谷剑芳，本科，高级工程师，河南省政务大数据中心数据治理部副部长，研究方向为数据治理、数据开发利用、数据要素政策；
王丹丹，硕士研究生，教授，博导，河南科技大学商学院，研究方向为数据要素与数据治理，大数据管理与应用；
袁云云，本科，中级会计师，河南省遥感院财务科会计，研究方向数据要素与数据治理，数据资产管理与应用；
金玲，本科，河南省自然资源厅副处长，研究方向数据要素与数据治理，数据资产管理与应用；
杨奎，硕士，高级工程师，河南省政务大数据中心公共平台部，研究方向为信息系统工程，大数据工程和人工智能应用；
通讯作者：张阳，本科，正高级工程师，河南省数据局数字政府规划建设处处长，研究方向为数字政府规划建设、政务信息化项目建设管理。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

They are failures of judgment about what a language model can and cannot be asked to do. Three misjudgments recur. The first treats a chatbot as an authoritative policy database, when its answers are sampled continuations of text rather than look-ups against a verified record. The second ranks candidate projects by technical ambition rather than by the size of the administrative problem they would remove, reproducing a bias toward building over using that long predates AI. The third manages AI as a deliverable with an acceptance date, leaving no budget line for the data refresh and retraining a deployed model needs in order not to decay. To test how much these misjudgments actually explain, four cases are compared: Futian District's "AI civil servants", Zhangzhou's "Hou Xi" assistant, Hangzhou's City Brain, and – deliberately, from outside government – Walmart's consolidation of scattered AI agents into a single front end. Walmart is included because it isolates a variable the public-sector cases cannot: what goes wrong when an organisation already has the data and the budget, and still fails because its tools are fragmented. Two questions organise the comparison: how deeply a tool is embedded in the actual work process, and how much of the stack is shared. These are then turned into a four-part scoring sheet (business value 40%, data readiness 25%, technical feasibility 20%, risk controllability 15%) that a local authority can apply before committing funds. A closing section argues that of the three foundations usually cited – standards, data and platforms – the least developed in practice is the chain that turns raw administrative records into material a model can actually learn from, and that this is where marginal investment currently pays off most.

Keywords : **artificial intelligence; government services; large language models; cognitive misconceptions; contextual adaptation; High-Quality datasets; platform governance**

引言

（一）研究背景与现实问题

数字技术的快速演进正深刻重塑政府治理模式与公共服务形态。以大语言模型为代表的新一代人工智能技术，凭借其在自然语言理解、语义推理与文本生成方面的突出能力，为政务服务智能化升级、行政流程优化再造和辅助决策支持提供了前所未有的技术驱动力。2025年以来，以 DeepSeek 为代表的国产大语言模型凭借在中文语境下的显著性能提升、灵活可控的私有化部署方式及对数据安全的高度重视，迅速成为政务领域智能化探索的重要工具。深圳、杭州、福州、青岛等数字化治理基础较好的地区，已在智能客服、政策咨询、公文辅助、热线工单处理等场景中积累了一定实践经验。

然而，技术热潮之下，政务 AI 的深度落地并非一帆风顺。从实践观察来看，当前政务 AI 建设中存在三类较为普遍的结构性问题：一是认知层面，政务决策与执行主体对大语言模型的技术本质、能力边界和适用范围存在系统性误判，导致应用路径选择偏差；二是实施层面，技术导向优先于业务需求导向的路径依赖，使“重建设、轻应用”的历史顽疾在 AI 时代以新形式延续；三是机制层面，沿用传统软件项目的静态工程思维，忽视了政务 AI 系统持续运营、数据迭代与人机协同的内在复杂性。上述问题相互叠加，导致大量政务 AI 项目在建设规模扩张的同时，实际效能释放严重不足，甚至出现“上线即搁置”“形象工程”等现象，造成公共资源的显著浪费与公众信任的潜在损耗。

这一现实困境表明，推动政务 AI 从浅层试点走向深度落地，既需要对大语言模型技术本质形成准确认知，也需要在场景选择、数据治理与平台建设等基础能力上实现系统性突破。如何在正确认知 AI 能力边界的前提下，构建符合政务场景特性的应用框架与评估工具，成为当前政务信息化研究亟待回答的核心问题。

（二）相关研究综述

人工智能用在公共服务和政府治理上，这几年讨论得不少。但具体到选型、落地和运维，已有研究给出的答案还不够用。下面梳理的三点不足，就是本文要试着补上的地方。

关于 AI 进政府会遇到什么阻力，Wirtz 等人早在2019年就用技术接受模型加制度理论梳理过一遍，归结到三样东西：单位自己能不能接住、制度上给不给空间、公众买不买账[1]。这个判断今天看依然站得住，但它停在“要不要用”的层面，没往下回答“用在哪”。Sun 等人的一项跨国比较发现，丹麦和中国的项目卡住的地方几乎一模一样——数据底子薄，部门之间不通气[2]。不过要注意，他们研

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

究的是上一波以规则引擎和机器学习为主的 AI，大模型带来的不确定性并不在他们的问题域内，这也正是本文想接着往下做的地方。

国内的研究更贴近工程实践：孟庆国、鞠京芮关心的是平台型政府的技术框架与实施路径[3]，施敏、杨海军盯的是隐私保护这一侧，具体到大模型的数据脱敏与采集环节[4]。两篇合起来看，一个讲怎么搭，一个讲怎么防，但都默认数据已经就绪——这个前提在真实项目里往往不成立。

真正对本文有直接支撑的是关于大模型能力边界的讨论。Bender 那篇被广泛引用的“随机鹦鹉”[5]点破了要害：模型输出的是按概率拼接的语言形式，流畅不等于理解。Huang 等人的综述给出的数字也不乐观——主流模型回答事实性问题时，出错的比例仍然可观，更麻烦的是这些幻觉常以高置信度的语气出现，使用者很难当场识别[6]。Ji 等人划出的高风险区更麻烦：打官司、看病、解释政策，恰恰都是知识密集的事[7]。把这三篇的结论叠起来看，会得到一个对政务场景很不友好的推论：幻觉最高发的任务类型，恰好与政务 AI 最想做的事高度重合。这一点在已有文献里没有被认真对待过，多数讨论停在“这个技术有什么特点”，没有往下写“那么在政务里该怎么管”。

在政务 AI 场景适配与数据治理研究方面，现有研究已认识到数据质量对 AI 效能的决定性作用。Janssen[8] 等指出，政府数据的碎片化、异构性与敏感性构成了公共部门 AI 应用的独特挑战，呼吁建立政务数据治理的专门框架。国内学者从数字政府建设实践出发，对政务数据共享、数据标准化和数据安全合规进行了较为深入的探讨，但对如何将高质量数据集建设与具体 AI 应用场景相结合，尚缺乏系统性的操作框架。此外，现有研究对政务 AI 平台建设模式的讨论多集中于技术架构层面，对“统分结合”模式在实践中如何兼顾共性集约与个性适配的机制设计，尚存在明显研究空白。

综合来看，现有研究在以下三个方面存在不足：第一，缺乏针对大语言模型技术特性与政务场景特殊需求之间适配关系的系统分析；第二，缺乏面向政务 AI 场景甄选的可操作评估工具；第三，在政务 AI 可持续发展的基础能力建设路径上，理论框架与实操指导之间仍有较大距离。本研究旨在针对上述三项不足作出针对性回应。

（三）研究问题与研究方法

1. 研究问题。文献读完，有三个问题绕不开：

问题一：地方在引入大模型时反复踩的坑，究竟是“不懂”，还是“懂”得过于自信？这两种情况的表现和后果不一样，得分开说。

问题二：同样是大模型，为什么福田的“AI 公务员”能铺到 240 个场景，而不少地方的系统上线后无人使用？差别主要出在场景选择，还是出在平台支撑？

问题三：标准、数据、平台这三件事，如果财力只允许先做一件，该做哪一件？三者之间是否存在严格的先后次序，还是可以并行？

2. 研究方法。本文用文献梳理与案例对比两种办法。

案例方面，按“理论复制”的思路选了五个样本，其中四个来自政务系统，沃尔玛作为企业侧对照。在多案例对比分析方面，以“理论复制”逻辑为指导，选取深圳福田“AI 公务员”、漳州“候喜”、杭州城市大脑、广州政务 AI 平台及沃尔玛 AI 整合转型共五个案例，构成覆盖政务场景纵深嵌入、跨域平台整合、基础设施集约建设及跨行业对照的多维样本。案例选取遵循以下标准：一是具有较为明确的应用目标与建设路径；二是有相对可查的实施过程信息；三是在场景类型、建设规模或组织形态上存在显著差异，能够支撑比较分析。需要指出的是，受制于公开信息的可及性，案例分析所依据的数据主要来自政府官方发布、权威媒体报道及行业研究报告，部分案例的效果数据尚缺乏独立第三方评估支撑，相关结论应在此限制下加以理解。

在分析框架上，本研究以“认知层—路径层—机制层”三层结构为主轴，围绕认知误区识别、场景适配要素归纳与基础能力建设路径三个维度展开论证，最终形成“需求导向—场景适配—数据底座—平台治理”的政务 AI 落地分析框架。

（四）研究贡献与文章结构

1. 研究贡献。本文想做的事有三件。

第一件是把“认知误区”从口号变成可辨认的具体表现——我们给出了每类误区的判别特征和对应的失败案例，决策者可以拿去对照自查，这是已有研究没做的。

第二件是那张四维评分表，它不新在概念上，新在能直接打分：权重、判分口径、一票否决条件都写死了，可以直接用于立项评审。

第三件是路径建议，我们没有按“标准—数据—平台”平铺，而是给出了优先次序和理由，也说明了“统”的代价。需要说明的是，这套东西是在有限的试用范围内调出来的，换到数据基础差异大的地区，权重需要重新校准，不能照搬。

2. 文章结构。全文共分 5 个部分。第 2 节梳理中国政务 AI 整体应用现状，勾勒技术应用的宏观背景；第 3 节系统分析阻碍政务 AI 深度落地的三类核心认知误区，通过多案例比较归纳政务 AI 成功落地的核心准则，并提出场景评估矩阵；第 4 节探讨政务 AI 可持续发展的基础能力体系构建路径；第 5 节总结研究结论，说明研究局限并提出未来研究展望。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

一、政务 AI 整体应用现状

政务 AI 的推进节奏在2025年出现了一次明显变化。此前各地的尝试多是单点试验——一个智能问答框、一套人脸识别门禁、一个能查进度的语音热线，彼此不通气。2025年之后，DeepSeek 一类模型把本地部署的门槛压到县区级也能承受，加上私有化部署方案成型，数据安全这一关才算真正过得去。

从落地顺序看，最先跑通的是面向公众的高频咨询类场景。原因不难理解：问题相对标准、答案可以预先准备、答错了后果可控。沪、浙、粤三地已经把智能问答嵌进一网通办的办事环节，用户在填表环节就能拿到指引，这类应用目前运行得比较稳定。城市治理类场景铺得更广，但深浅不一。杭州城市大脑是少数真正打通了数据源的例子——它把交通、城管、应急的数据拉到一处，再按实时流量调信号灯，这套做法后来被不少城市照搬。更多的项目则只做到了可视化大屏，算法照旧跑在各部门自己的系统里，数据并没有真正合流。行政办公侧的普及速度超出预期。写材料、整纪要、翻文件这几件事，因为不直接对外产生效力，推进阻力最小，也是基层感受最明显的。法院走得更快：判决书可以先由模型出草稿，关键要素由它去抽取，这些在多地已经日常化。不过这里有个前提：法院的案件数据本身结构化程度高，换到数据基础差的部门，同样的工具往往推不动。

技术路线上，目前主流做法是“底下用通用模型，中间挂本部门的知识和规则，上面再按场景做微调”，人机分工也基本固定下来：模型出初稿，人来核。这个分工在可预见的时期内不会变。需要指出的是，覆盖面扩大不等于效能提升。跨部门数据拿不到，是最常被提到也最难解决的一条；其次是县乡一级没人维护知识库，省里训的模型到了基层容易失准；政务数据敏感度高，模型安全合规的要求比一般信息系统严，很多通用做法不能直接套用。这几条在后面两节会分别展开。

总体看，这几年政务 AI 的推进顺序是清楚的：先在外围的咨询、辅助类场景跑通，再往核心业务走。这个顺序不宜颠倒，也解释了为什么同样是上大模型，有的地方很快见到效果，有的则长期停在演示阶段。阻碍政务 AI 深度落地的核心症结在认知。DeepSeek 等大语言模型（LLMs）的突破性进展，点燃了2025年以来全社会各领域的 AI 应用热情，政务领域亦不例外。深圳龙岗政务服务大厅率先推出基于 DeepSeek 的“数字公务员”示范[9]，引发了各地政府的学习与效仿热潮，以期快速提升政务服务智能化水平。然而，在技术热潮涌动之际，一个深层次的挑战已然显现：政务主体对 AI 技术（尤其是 LLMs）的基本原理、核心能力边界、适配路径及长期运维逻辑普遍存在关键性认知偏差。这种系统性认知不足，源于政务决策与执行层面对 AI 相关通识知识的普遍欠缺，不仅导致其在应用路径选择上易陷于盲目性决策，更成为掣肘 AI 技术真正深度融入政务场景、实现可持续价值的核心瓶颈。

（一）能力认知偏差：过度泛化，误将 LLMs 视为“全域精准知识库”

GPT、DeepSeek、通义千问等生成式大语言模型（LLMs）的涌现，极大提升了社会对 AI 能力的预期。在政务领域，LLMs 的对话交互能力使其在办事查询、政策咨询等场景展现出显著便捷性，却也滋生了一种危险的认知偏差，即将 LLMs 等同于“全域精准、静态的知识库”。这种偏差直接导致对其可靠性的不合理预期。

现实风险已清晰显现：一是信息可信度危机。LMYS.org 2025年8月 Vicuna 评测显示，主流 LLMs 如 Gemini-2.0-Flash、OpenAI o3-mini-high、DeepSeek-V2.5均存在幻觉现象[10]。用户常遭遇 LLMs 以高度自信口吻输出虚构事实或错误信息。二是专业领域失准。美国律师因提交 ChatGPT 编造案例险酿司法事故[11]；DeepSeek-R1关于国产动画的回答充斥事实谬误[12]；更严峻的是，某医院 AI 处方系统忽略患者转氨酶异常，推荐可能加重肝损的降脂药瑞百安[13]，暴露其缺乏医学判断与风险规避能力。技术根源在于 LLMs 的运作本质。LLMs 的核心机制是基于海量语料进行分词、向量化，并利用注意力机制建模词汇关联与位置关系。其生成的“逻辑合理”文本，本质是统计概率驱动下的上下文拟合，而非基于事实核查或深度理解的理性推理。尽管模型持续优化，其“推理”过程仍具有自强化特性，极易在缺乏真实依据时生成“貌似合理”却偏离事实的“幻觉”内容，与人类基于认知与情感的判断存在根本差异。Emily M. Bender、Timnit Gebru 等研究者指出，LLMs 根据概率信息将训练数据中的语言形式片段拼接成序列，但并不具备人类的理解能力，“就像个随机鹦鹉”[14]。

LLMs 不等于知识库，其输出是概率产物，非精准数据库检索结果，不具备可靠的事实核查与专业判断能力。在现有技术框架下，统计模型驱动的自递归推理必然伴随一定程度的幻觉风险。在涉及法律、医疗、政策解读等高严谨性、高后果责任的政务场景中，过度依赖 LLMs 作为最终知识源或决策依据存在重大安全隐患。构建高质量政务语料库虽能降低错误概率，却无法解决 LLMs 基于概率生成“合理错误”的本质缺陷，需结合其他机制（如事实核查、规则约束、人类监督）构建安全边界。

（二）实施路径偏差：技术本位，忽视政务场景适配性的单向思维

“重建设、轻应用”的顽疾，长期困扰着政府信息系统建设，亦是催生数字形式主义的温床。尤其在综合类大型平台（如数据中台、监管平台）建设后，常陷入用户注册率高、实际活跃度低、功能长期闲置的窘境。这种脱离现实业务需求、盲目追求技术先进性的“为建而建”思路，导致技术与应用严重脱节，造成巨大的资源浪费与项目烂尾风险。

AI 项目的特殊性加剧了风险。相较于传统信息化项目，AI 项目（尤其是大模型应用）具有技术复杂度高、算法结果不确定性大、算力资源消耗巨大三大特征，其建设成本、实施难度与失败风险显著提升。比如，某地住房城乡建设厅投资1200万元建设的6个政务服务信息系统，片面追求硬件“高端大气”，却因缺乏实际业务支撑和用户需求牵引，部分系统建成即停用或长期闲置，沦为典型的“面子工程”与资源浪费[15]。政务服务的核心价值在于便民利民与行政效能提升。然而，脱离具体业务场景的 AI 大模

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

型建设，极易陷入追求技术复杂性的迷思。复杂、难用的 AI 交互设计，反而可能增加公众操作负担，背离“简化流程”初衷。“智能外壳”下若缺乏对流程优化、数据贯通等治理本质的支撑，非但无法提升效率，甚至可能因系统冗余、响应迟缓而降低服务效能。华而不实的 AI 项目不仅浪费公共财政，更会严重侵蚀公众对政府数字化能力的信任与满意度，损害政府公信力。

AI 立项必须以解决具体政务痛点、满足真实用户需求、提升核心业务效能为根本出发点，而非技术本身的先进性。坚决摒弃“为 AI 而 AI”的思维。地方政府需立足本地实际（业务成熟度、数据基础、财力人力），对 AI 项目的实用性、可行性、投入产出比进行严格论证。避免贪大求全、盲目跟风上马高风险项目。优先选择业务边界清晰、价值可衡量、技术风险可控的细分场景进行试点验证（如特定事项智能问答、表单自动填充），通过敏捷迭代、持续优化积累经验，再考虑规模化推广。避免一次性投入巨资建设“大而全”却无用的平台。AI 赋能政务服务的关键在于技术与治理的深度融合。必须同步推进业务流程再造、数据治理标准化、人员能力提升等治理工作，确保 AI 技术真正嵌入并服务于政务效能提升的目标，而非成为新的负担或“花瓶”。

（三）运维机制偏差：静态工程观，规避政务系统复杂性所需的动态闭环

当前受媒体渲染影响，将 AI 平台简单视作“即插即用”型工具，误以为部署即可输出稳定服务。这种认知严重低估了政务 AI 系统的复杂性与生命特性，本质上是将动态生命体错认为静态工程品。

政务 AI 平台（尤其基于大模型）是开放式系统，其核心能力（模型、高质量数据集、训练机制）需要持续运营而非一次性建设。AI 的精准性、适用性（如政策解读、问题解答）高度依赖持续注入的多维度、高质量政务数据流。这要求建立常态化的数据归集、清洗、标注、校核与训练机制。AI 功能价值的实现必须深度嵌入并适应动态变化的政务业务流程与规则，而非独立存在。比如，河北省石家庄市依托 DeepSeek 大模型对政务服务智能客服进行全面升级，通过百万级数据量政务知识库训练和本地化场景适配，提供 7×24 小时在线咨询解答 [16]。

在 AI 的应用实践中，真正有挑战的不是建立机器学习模型，而是在模型复杂而庞大的情况下，如何构建一个可以持续集成的机器学习系统，以及在生产环境中持续运行该系统 [17]。沿用传统软件项目“验收即结束”的思维，会导致后续数据更新停滞、模型迭代乏力、知识库陈旧，AI 性能迅速退化甚至失效；也会导致对持续性数据工程（DataOps）、模型运维（MLOps）所需资源（预算、编制、算力、专业团队）的长期规划和投入保障的缺失。AI 持续优化依赖技术方与业务方的高频深度协同（问题反馈收集、业务规则同步、场景适配调整）。静态思维下，协同机制往往流于形式或难以持续。非专业人士对 AI“易用性”和“智能水平”的认知混乱，叠加技术复杂性带来的“黑箱”效应，易导致决策者产生不切实际的预期或低估长期投入的必要性。这使得政策制定和项目规划脱离实际，最终在运维阶段暴露问题，造成项目“形同虚设”或“半途而废”。

所以与其说“AI 即服务”，不如把话说白：这类项目的钱，大头在验收之后。立项时就要把三件事写进方案，缺一件不建议开工。

一是数据的持续供给。模型上线只是起点，政策法规在改、办事指南在变，语料停更三个月，回答的准确率就会明显下滑。建议把数据归集、清洗、标注的责任落到具体处室，并明确更新频率——写进合同，而不是写在倡议里。石家庄的做法可以参考：依托 DeepSeek 升级智能客服时，同步建了百万级数据量的政务知识库，并配了本地化适配的维护流程，覆盖企业开办、执业登记、食品药品等200余项高频事项 [16]；真正难的其实不是建这个库，是让它一直新。

二是运维的责任主体。技术方撤场后谁来接，这个问题多数项目答不上来。比较可行的做法是设跨职能的“数字专员”，编制放在业务处室而非信息中心，负责收集问题、同步规则变更、协调模型再训练。同时给业务人员做一轮 AI 素养培训，目的不是让他们懂算法，是让他们知道什么样的错误必须反馈。

三是考核口径。建议把“模型更新频率、问题解决率、用户满意度”三项列为后评估的硬指标，与运维经费续拨挂钩。只考核“建成了几个”，结果一定是建完就停。

二、政务 AI 成功落地的核心准则与方法论

做信息化是为了办成事，不是为了有个系统。AI 也一样，先说清楚要解决谁的什么问题，再谈用什么模型。

（一）AI 应用要坚持需求导向

找场景不是列清单。调研中见过不少地方把“AI 应用场景”做成一张几十项的表，逐项比谁列得多，最后真正跑起来的不过三四项。问题往往出在一开始就问错了：不该问“AI 还能干什么”，而该问“哪些事现在办得最费劲”。

把两侧的能力摆在一起看会更清楚。大模型这边，真正成熟的本事集中在三处：读得懂用自然语言写的诉求，能从长文本里抽要点，能按模板生成格式固定的文稿。它明显不擅长的，是需要唯一正确答案、且答错了要担责的事。政务这边的特点恰好相反——规则硬、留痕严、一个事项常常牵扯好几个部门，多数环节不允许“大概齐”。两者一叠，能马上见效的范围其实很窄：咨询量大而答案相对固定的（政策问答、办事指引）；材料多而格式统一的（表单填报、证照识别）；以及需要在大量文本里找线索的（工单分派、投诉归类）。至于需要援引具体法条、直接影响相对人权利义务的环节，现阶段更适合让它写初稿、人来定稿，而不是直接对外输出。这条边界划清楚，后面讨论场景才有意义。

先说 AI 到底能干什么。把它和政务业务放在一起比对会发现，真正成熟的能力集中在三处：读得懂用自然语言写的诉求；能从长材料里抽出要点；能按固定模板生成文稿。把这三样放到具体岗位上看，最先受益的就是窗口咨询、热线应答、材料预审这类高频、重复、答案相对固定的环节——青岛12345把 AI 接进工单流转之后，话务诉求量同比下降了19.91%[18]，靠的就是这一

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

类替换。反过来，需要引用具体条款、直接影响相对人权利义务的事，现阶段更适合让它写初稿、人来定稿。

第二种用法是在跨部门的链条上找机会。企业开办、不动产过户这类“一件事”，群众要跑好几个部门、同一份材料反复交，是被抱怨最多的地方。模型能做的是把各部门的材料清单和政策条文通读一遍，把重复项和互相打架的条款标出来，由人决定砍哪一条。这一步在市县两级的效果差别很大：数据通的地方，几天就能出结果；数据不通的地方，模型只是把问题重新描述了一遍，事情照样卡在协调会上。

第三种用法，是让 AI 反过来治理数据本身。清洗、格式转换、跨库比对这类活儿，规则写清楚之后机器比人快，也不容易漏。我们试过的做法是：先让模型把明显矛盾的记录挑出来，附上出处，交给人去判。人的时间于是能花在判断上，而不是逐条翻台账。这一步的前提得说清楚——训练用的语料和更新机制必须先到位，否则机器校出来的东西同样不可信。

（二）AI 应用要适应场景需要

光讲道理不够，下面拿几个真实项目做对比，看跑起来的和趴窝的项目之间，究竟差在哪几个变量上。

案例一：福田那70个“AI 公务员”是怎么铺到240个场景的

先看深圳福田。福田在2025年年初上线了首批70个“AI 公务员”[19]。它们不是机器人，是挂在通用大模型上的一批小助手：每个助手只管一类事，比如公文初稿、民生诉求分拨，各自带着本部门的业务数据和业务规则。全区一共铺了240个场景，从公文处理到招商引资都有。数量看着惊人，做法其实很朴素——先把通用模型接进来，再灌业务数据，最后按具体事项切成一个个边界清楚的小体。

案例二：沃尔玛——把散落的 AI 工具收成一个入口

沃尔玛的教训在另一个方向。它很早就在用 AI：商品搜索一套、排班一套、供应商管理又一套。工具各自好用，放在一起却很别扭——入口不同、上下文不连通，员工要在几个界面之间来回跳，数据也各存各的。2025年7月，沃尔玛动手把这些工具重做成面向四类用户的“超级代理”，把搜索、推荐、客服收进同一个入口[20]。这次整合并没有换更强的模型，改的是架构。这一点对本文很重要：沃尔玛既不缺数据也不缺预算，它出问题纯粹是因为工具散着装——在政务案例里，这个变量和“数据不通”混在一起，很难单独剥出来。

两个案例摆在一起，能读出一条共同的规律：决定成败的不是模型好坏，是应用之间有没有被组织起来。福田的做法可以拆成三步——先接一个通用大模型，再灌入各部门自己的业务数据和规则，最后按具体事项做成独立的智能体。70个“AI 公务员”覆盖240个场景，看着是规模问题，实质是组织问题：每个智能体的边界划得很清楚，只做本部门那一类事，所以规则能写死、责任能落到人。反过来，如果一上来就做一个“什么都能答”的通用助手，规则写不进去，出错也找不到人负责。沃尔玛走的是相反的路：它先有了很多 AI 工具，问题恰恰出在“多”上——产品搜索、员工调度、供应商管理各有一套，接口不同、上下文不通，员工要在几个系统之间来回切。2025年7月的整合并不是换更好的模型，而是把入口合并成面向四类用户的“超级代理”[20]。值得注意的是，沃尔玛既不缺数据也不缺预算，它的失败完全来自架构碎片化——这个变量在政务案例里很难单独剥离出来，所以本文特意把它放进来做对照。由此得到的可迁移判据有两条：一是场景边界是否清楚到能写进规则，二是多个应用之间是否共用上下文。至于“技术先进度”，在这两个案例里都不是决定性因素，甚至没有出现在失败原因里。需要提醒的是，福田模式的隐含前提是部门数据已经打通，对数据基础薄弱的地区，照搬很难达到同样效果。

（三）政务 AI 场景甄选的价值引领与多维评估矩阵构建

为帮助决策者在立项前筛掉不成熟的想法，我们做了一张评分表（下称“评估矩阵”）。它的设计依据来自两份政策文件：《数字中国建设整体布局规划》强调应用实效，《生成式人工智能服务管理暂行办法》则划定了安全合规底线；前者对应业务价值，后者对应风险可控，中间的数据基础与技术可行性，是把已有研究的共识操作化之后补进去的。

先看业务价值，权重给到40%。理由很直接：一个场景如果做与不做的差别不大，后面三项再漂亮也没有立项必要。打分时建议只看三件事——这件事现在要花多少人天、出错率有多高、投诉集中在哪。三者有一项突出就往下走，三项都平平基本可以直接否掉。需要提醒的是，不少地方把“提升治理能力”“赋能科学决策”写进立项理由，这类表述无法证伪，建议在打分表里明确不计分。

数据基础给25%，排序在技术之前，是因为我们在实践中反复遇到同一件事：模型选得没错，场景也对，最后卡在数据上——要么拿不到，要么拿到的是三年前的。判断标准建议简化为两条：一是这批数据谁有权给、走什么流程、多久能到手（很多好想法死在“走流程”上）；二是数据有多旧、多久更新一次。至于完整性、准确性，可以用抽检代替全量评估：按批次随机抽样，错误率超过约定阈值的退回重标。

技术可行性给20%，是四项里权重最低的，也是刻意的。大模型应用的工程门槛这两年下降很快，真正因为“技术做不到”而失败的项目，比因为“数据不给”或“没人维护”失败的要少。这里只问三件事：现有算力够不够跑、团队有没有人能接手改、出问题能不能回滚。

风险可控性给15%，但它是“一票否决项”，不参与加权。涉及个人信息、可能影响相对人权利义务、或直接生成对外文书的，无论前三项分数多高，都得先把评估、备案这套合规流程走完，并且留一道人工审核。反过来，内部辅助、不对外、错了也无后果的场景，可以走简化流程。

表：AI 评估矩阵表

评估维度	指标解释和权重说明	参考权重
业务价值	看三件事：现有人天投入、当前出错率、投诉集中度。三者皆平则否掉。“提升治理能力”一类无法证伪的表述不计分。	40%
数据基础	看两条：数据能否拿到（权责与流程）、数据有多新（更新频率）。完整性、准确性用抽检代替全量评估，按批次随机抽样，超阈值退回重标。	25%

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

技术可行性	只问：算力够不够、有没有人接手改、出问题能否回滚。权重低于数据基础，因为实践中卡在技术上的项目少于卡在数据上的。	20%
风险可控性	不参与加权，一票否决。涉及个人信息、影响相对人权利义务、直接生成对外文书的，须过安全评估与算法备案，并保留人工审核环节。	15%

需要说明的是，这组权重是我们依据实际试用情况调出来的，不同省份数据基础差异较大，建议正式采用前先在本地试评两三个已知结果的项目，校准后再推开。评分表解决的只是“先做哪个”，做起来之后能不能持续，取决于后面一章讨论的那些支撑条件。

三、政务 AI 可持续发展的基石与体系化构建路径

前面回答了“选什么场景”，这一节回答“靠什么撑住”。数据、标准、平台是三件绕不开的事，但它们的紧迫程度并不相同，本节按这个顺序展开，并说明我们认为应该先做哪一件。

（一）标准规范先行，引领政务 AI 健康有序发展

标准先行这句话说了多年，真正卡住的不是没有标准，是标准落地时没人认。在省级数据共享平台的实际对接中，常见的情况是：接口规范发了，但各部门历史系统的改造排不上队，新标准只能在新建模块上生效，老数据依旧按老口径出。所以与其列一张涵盖架构、组件、协议、语料、流程、指标的全清单，不如先抓两件能验收的事。

一是模型与平台之间的调用接口，先统一到能在省内互认的程度；二是数据标注的质量要求，明确到“什么算标错、抽检比例多少、谁签字”这一层。其余的可以边做边补，一次把标准做全的项目我们没见过成功的。评价环节建议硬一点：明确每半年对在用模型评一次，评不过的就下线，而不是“优化或淘汰”这种两头都算的话。

（二）夯实数据根基，保障高质量政务数据供给

高质量数据集这个提法出自2024年底一份关于数据产业的指导意见，但落到政务口，难的不在“集”，在“炼”。政务数据的原始状态很差：省直部门的数据大多躺在各自业务库里，字段口径各定各的；市县一级还有大量扫描件、PDF 红头文件、Excel 台账。这些东西直接喂给模型，效果通常比不喂更糟——模型会把表格表头当正文读，把扫描件的水印和页码当关键词。因此，笼统地讲“全生命周期管理”没有意义，真正该单独立项的是供给侧到消费侧之间那段加工链路，可以拆成四步：先做面向训练的预处理工程，用自动化处理加人工抽检来消化扫描件和非结构化文本；再做标准化的接入通道，让共享平台能按模型需要的格式自动供数；第三是效果反馈，模型答错了要能回溯到是哪批语料的问题；第四是给标注成本单独算账——标注这件事靠业务骨干，不靠外包，得有长效机制留人。

（三）构建 AI 平台，强化共性支撑与集约发展

平台这块，争议一直不在“要不要统”，而在“统到哪一层”。杭州城市大脑和广州的政务大模型平台给出了两种答案。杭州从城市治理切入，拿 DeepSeek-R1 当底座，把十几个行业的模型放到同一套算力上来调度[21]；广州则先把算力、存储、网络、安全做成统一底座，再在上面长出工单分派、公文助手、政务服务“慧办小精灵”等十几类智能体[22]。两者的共同点在于底座统、应用分：算力、工具链、数据通道统一供应，行业模型各自训练。差别在起步点——杭州从场景倒推，广州从基础设施正推。对多数省份而言，后一条路更可复制：先把算力和数据通道统起来，让各部门不再各自买卡，再谈模型。需要一并说明的是，“统”也有代价——统一底座意味着单个部门的需求要排队，紧急需求响应变慢，这需要在立项时就把优先级规则写清楚，否则很容易在具体项目上反复扯皮。

四、结论与展望

（一）研究结论

本文从人工智能基本概念切入，首先明确 AI 在政务服务领域应用的正确认知，进而结合实际应用场景，总结了 AI 的创新案例，分析其面临的技术挑战。通过剖析典型案例，强调了技术适配场景、场景反哺技术以及充分激活数据价值的关键作用。研究认为，推动政务服务从数字化向智能化、人性化和精准化迈进，需要系统性谋划，包括挖掘核心应用场景、加工高质量数据集以筑牢数据基础、增强算力供给能力、采用统分结合的方式建设 AI 平台、完善运营服务体系等。

（二）研究展望

本研究虽构建了政务 AI 应用的系统性分析框架并提出实践建议，但选取的代表性案例多集中于特定地域与行业，且案例以成功实践为主，研究结论基于当前大语言模型与 AI 技术特征，部分结论可能受限于现阶段技术水平，对未来技术突破带来的新挑战与机遇预见不足。虽强调高质量数据集的重要性并设计 AI 评估矩阵，但对数据采集、清洗、标注等具体操作流程及工具的技术分析较为概括，未能充分展开实操层面的细节探讨。未来可重点从政务大模型的伦理规制体系构建、联邦学习、差分隐私等技术在政务数据处理中的适配性、主动感知技术的演进、以及人机对抗方向等方面进一步深化研究。

参考文献

[1]Wirtz ,Weyerer ,Geyer .Artificial Intelligence and the Public Sector—Applications and Challenges[J].International Journal of Public Administration,2019,42(7):596-615.

[2]Sun Q T ,Medaglia R .Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare[J].Government Information Quarterly,2019,36(2):368-383.

[3] 孟庆国、鞠京芮 . 人工智能支撑的平台型政府：技术框架与实践路径 [J]. 电子政务 ,2021(9):37-46.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

[4] 施敏，杨海军 . 大语言模型数据隐私保护的难点与探索 [J]. 大数据 ,2024(5):174-182.

[5] Bender E M, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots: can language models be too big?[C]//Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021: 610-623. DOI: 10.1145/3442188.3445922.

[6] Huang L, Yu W, Ma W, et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions[PP/OL]. arXiv(2023-11-09)[2026-09-09]. https://arxiv.org/abs/2311.05232. DOI: 10.48550/arXiv.2311.05232.

[7] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1-38. DOI: 10.1145/3571730.

[8]Janssen M ,Brous P ,Estevez E , et al.Data governance: Organizing data for trustworthy Artificial Intelligence[J].Government Information Quarterly,2020,37(3).

[9] 赵娟 , 孟天广 . 数字政府的纵向治理逻辑 : 分层体系与协同治理 [J]. 学海 ,2021(2):90-99.

[10] 刘泽垣 , 王鹏江 , 宋晓斌 , 张欣 , 江奔奔 . 大语言模型的幻觉问题研究综述 [J]. 软件学报 ,2025,36(3):1152-1185.

[11] 王秀平 . 人工智能司法应用法律风险及法治应对 [J]. 政法论丛 ,2024(2):151-160.

[12] 韩旭 , 张正彦 , 刘知远 . 知识指导的预训练语言模型 [J]. 中兴通讯技 ,2022,28(2):10-15.

[13] 李雪晴 , 王石 , 王朱君 , 朱俊武 . 自然语言生成研究综述 [J]. 计算机应用 ,2021(5):1227-1235.

[14]Bender E M, Gebru T, McMillan-Major A, et al. On the dangers of stochastic parrots: Can language models be too big?[C]Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,2021:610-623.

[15] 欧阳康 . 基层数字形式主义：表现、成因与对策 [J]. 国家治理 ,2024(13):10-14.

[16] 曾维和 .DeepSeek 嵌入政务智能生态的风险及治理 [J]. 理论探讨 ,2025(5):37-47.

[17] 林锋 . 持续训练在机器学习平台中的应用研究 [D]. 上海：上海交通大学电子信息与电气工程学院 ,2021.

[18] 王静 , 张红 . 人工智能在数字政府建设中的应用及其法治化路径 —— 基于 DeepSeek 在数字行政中应用场景的分析 [J]. 中外法学 ,2025,37(3):625-643.

[19] 霍旻含 .AI“公务员”，表现很亮眼 [N]. 人民日报海外版 ,2025-02-24(8).

[20] 毛伟杰 . 动态能力视角下跨国零售企业供应链数字化转型路径研究 ——以沃尔玛为例 [J]. 广西民族大学硕士电子期刊 ,2025-06-16—2025-07-15,2025(7).

[21] 赵路 , 屠友军 . 杭州全面启动城市大脑 3.0 建设 [N]. 浙江日报 ,2025-04-01(1).

[22] 王燕 , 方晴 . 广州：全国首个城市“三维数字底座”将这样支撑数字政府改革 [N]. 广州日报 ,2022-08-25(1).

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

融合机器学习与结构方程模型的少数民族文化高质量发展指数（CHDI）构建与验证

邱紫玥，刘付靖岚，倪一丹，张振华 *
广东外语外贸大学 数学与统计学院，广东 广州 510006

DOI:10.61369/ASDS.2026100003

摘　　要： 数字技术赋能与东西部协作背景下，少数民族文化传承正由传统保护模式向数字化、参与式和协同化发展转变，但现有研究对群众参与与数字赋能的协同作用及其内在心理路径关注不足，亦缺乏兼具理论解释力与量化评价功能的综合研究框架。本文以粤黔协作“山海同心”工程为实践背景，基于403份有效调查问卷，构建“群众参与—数字赋能—文化传承”分析框架，综合运用机器学习与结构方程模型，对少数民族文化高质量发展的核心驱动因素、作用路径及量化评价方法进行研究。

首先，采用随机森林回归（RF）、极端梯度提升（XGBoost）、神经网络回归（NNR）和支持向量回归（SVR）对影响文化传承的14个维度进行重要性评估，并依据多模型的动态加权后的特征重要性排序（R2为赋权依据），筛选得到7个核心影响因素。其次，基于选取的七个核心因子标准化得分与其机器学习权重相乘，进行标准化处理，构建少数民族文化高质量发展指数（CHDI），并通过信效度、相关性及测量模型检验验证其量化评价的可行性。进一步地，构建SOR–TPB融合理论检验传导路径，经检验，外部刺激变量对行为态度有显著正向作用，行为态度和知觉行为控制进一步通过文化传播行为意愿作用于文化传承行为，呈现“外部刺激—心理机制—行为意愿—传承行为”的链式中介关系。最后，采用RF、XGBoost、NNR和SVR构建CHDI指数，并通过多模型加权融合开展政策沙盘推演模拟。结果表明，融合模型测试集 R^2 达到0.9934，RMSE和MAE分别为0.0197和0.0152；进一步的恒定资源约束情景分析显示，围绕核心影响因素实施精准干预具有较好的资源配置效果。

研究发现：（1）群众参与是传承行为的核心驱动，数字赋能边际效益显著；（2）CHDI指数具良好测量学属性，可作为标准化量化评估工具；（3）外部刺激通过“心理机制→行为意愿”完全链式中介影响传承行为，情感体验与行为控制感是关键心理路径；（4）政府应以数字赋能优先、精准施策、避免资源错配实施文化政策。

关 键 词： 少数民族文化高质量发展；CHDI；数字赋能；群众参与；机器学习；结构方程模型

Development and Validation of the Cultural High-Quality Development Index（CHDI）for Ethnic Minorities Using a Combination of Machine Learning and Structural Equation Modeling

Qiu Ziyue, Liufu Jinglan, Ni Yidan, Zhang Zhenhua*
School of Mathematics and Statistics, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006

Abstract: Digital empowerment and East–West collaboration are reshaping how ethnic minority cultures are preserved, with the field shifting from traditional conservation models toward digital, participatory, and collaborative approaches. Existing research, however, has paid insufficient attention to how public participation and digital empowerment reinforce each other and through what psychological pathways; a comprehensive framework that is both theoretically explanatory and quantitatively evaluable remains absent. This paper takes the Guangdong–Guizhou Collaborative “Mountains and Seas, One Heart” Project as its empirical context and, using 403 valid survey responses, constructs a “Public Participation—Digital Empowerment—Cultural Heritage” analytical framework. Combining machine learning with structural equation modeling, it investigates the core drivers, pathways, and quantitative evaluation methods underlying the high-quality development of ethnic minority culture.

R2To rank the 14 dimensions influencing cultural preservation, we first ran four machine learning

作者简介：
邱紫玥（2005.10–），广东东莞人，广东外语外贸大学，本科在读，研究方向：社会发展多维计量评价，文化大数据与社会统计分析；
刘付靖岚（2006.02–），广东茂名人，广东外语外贸大学，本科在读，研究方向：统计机器学习混合建模与实证研究，人文社会数据科学计量研究；
倪一丹（2006.06–），浙江温州人，广东外语外贸大学，本科在读，研究方向：数字赋能文化传承研究，量化政策评估。
通讯作者：张振华，广东外语外贸大学，教授，研究方向：统计机器学习，因果推断与因果学习，科学计量学与信息计量学，大数据决策。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

models—Random Forest Regression (RF), Extreme Gradient Boosting (XGBoost), Neural Network Regression (NNR), and Support Vector Regression (SVR)—and used the dynamically weighted feature importance rankings across the models as the weighting criterion, which reduced the candidate set to seven core factors. We then multiplied the standardized scores of these seven factors by their machine-learning weights and normalized the results to construct the High-Quality Development Index for Ethnic Minority Culture (CHDI); tests of reliability, validity, correlation, and measurement models confirmed that the index is feasible for quantitative evaluation. Next, to examine the transmission pathways, we developed a framework integrating the Stimulus-Organism-Response theory (SOR) with the Theory of Planned Behavior (TPB). The results show that external stimulus variables significantly and positively affect behavioral attitudes, which, together with perceived behavioral control, drive cultural inheritance behavior through the intention to engage in cultural dissemination—a chain-like mediation running from external stimulus to psychological mechanisms to behavioral intention and finally to inheritance behavior. In the last step, we rebuilt the CHDI with RF, XGBoost, NNR, and SVR and conducted policy simulation through multi-model weighted fusion. The fusion model achieved an R^2 of 0.9934 on the test set, with RMSE and MAE of 0.0197 and 0.0152, respectively, and scenario analysis under constant resource constraints suggests that interventions targeted at the core influencing factors lead to better resource allocation.

Four findings emerge from this study. First, public participation is the primary driver of cultural preservation, and digital empowerment brings significant marginal benefits. Second, the CHDI has sound psychometric properties and can be used as a standardized quantitative assessment tool. Third, external stimuli shape cultural preservation through a complete mediation chain of "psychological mechanisms → behavioral intention," in which emotional experience and perceived behavioral control act as the key psychological pathways. Fourth, cultural policy should give priority to digital empowerment, adopt targeted measures, and avoid the misallocation of resources.

Keywords : high-quality development of ethnic minority culture; chdi; digital empowerment; public participation; machine learning; structural equation modeling

引言

近年来，在数字技术赋能与东西部协作协同推进的时代背景下，东西部协作与文化数字化国家战略深受我国重视。《“十四五”文化发展规划》（2022）对民族地区公共文化建设提出了明确要求，推动民族文化传播与传承发展。“十五五”规划（2026—2030）进一步强调了全面推进东西部协作，发挥跨区域重大工程支撑作用，强化文化遗产保护传承与活态传承。随着数字化时代，《关于推动文化高质量发展的若干经济政策》（2025）强调了探索文化与科技融合机制，以数字化、信息化赋能文化建设，运用数字技术推动少数民族文化活化传承已成为时代要求。

但政策上的重视与实际取得的传承效果之间仍然存在落差。文化自身时滞性和部分民族地区在资金、人才、渠道等方面条件有限，两方面的因素叠加使得一些少数民族文化长期处在“文化孤岛”的状态之中。以贵州省黔西南布依族苗族自治州晴隆县的苗族文化为例，当地对它的保护长期以静态方式为主，工作多停留在资料整理、实物保存和非遗名录申报等环节，但传播渠道也因此相对狭窄。由此保护的力度在增加，文化的活跃程度却未必同步提高，活态传承与有效传播两个目标都受到制约。

基于上述判断，本研究团队依托粤黔协作 “山海同心”工程，在广东省惠州市开展了实地调研，调研围绕晴隆民族文化在惠州的传播路径、影响机制与传承效果展开。在探索东西部协作背景下，数字赋能是驱动少数民族文化传承与发展的有效路径模式，其作为东西部协作的典型实践，有效推动了晴隆苗族、布依族等少数民族文化向惠州辐射传播，为民族文化跨区域传承提供了重要实践载体。

本研究在理论工作主要有三处推进。第一，把数字赋能与群众参与放进同一个分析框架中考察。已有研究往往把两者分开讨论，较

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

少追问二者在同一机制中如何相互作用，本研究试图补上这一环节，弥补现有研究对双维度孤立探讨的不足。第二，在理论模型的选择上，本文同时把 TPB（计划行为理论）与 TAM（技术接受模型）的核心维度嵌入其中。这样处理的好处在于，外部刺激、个体心理状态与行为反应之间的层次关系得以保留的同时特有的情感因素与技术接受因素也能同时进入模型，理论解释的边界因此有所扩展。第三，在方法上采用机器学习与结构方程模型的联合实证。前者负责因子重要性排序与发展水平预测，后者负责路径机制的检验与效应分解。二者的结论相互参照，既弥补了单一方法的不足，也让实证结果更为稳健。

在实践层面，本研究希望产出能够直接使用的工具与依据。其一，构建少数民族文化高质量发展指数（CHDI），使文化发展水平的量化评估有据可依，地方在比较不同区域、不同时段的工作成效时，不必再依赖主观印象。其二，借助机器学习识别群众参与与数字赋能两个维度中的关键因子和薄弱环节。哪些因素真正影响传承效果，哪些环节长期得分偏低，都可以得到较为明确的回答，这为“山海同心”工程的优化提供了具体方向，也有助于避免资源在各项任务上平均分配、政策在各类对象上泛化实施。其三，于机器学习的发展水平预测模型，实现文化政策干预前的效果模拟。即在资源有限的前提下，为政府决策提供“先预测、后施策”的技术支撑。

一、文献综述

（一）少数民族文化传承与传播研究

少数民族文化传承与传播研究是近年来学术关注的热点领域。邱婷等人结合“技术—组织—环境”理论框架，揭示了数字技术对非遗创新的驱动作用[1]。罗薇等人提出文化传承行为需强调政府－社区－群众多元协同关键作用[2]。王萍等人系统梳理了非物质文化遗产传承与创新的影响机制，指出国内对区域文化生态影响机制的研究仍较为薄弱[3]。

综上，现有研究多聚焦于非遗的数字化保护与区域文化生态等视角，缺乏对文化传承与传播背后的公众心理机制的量化实证。因此，本研究引入心理学框架，通过实证检验揭示数字赋能与群众参与影响文化传承的作用路径，弥补现有研究不足。

（二）数字赋能文化高质量发展研究

数字赋能背景下文化传承研究已成为学术前沿。刘明真等人从数字文化战略视域出发，做出数字赋能背景下文化传承的逻辑框架[4]。陈波、涂晓晗从数字赋能的视角，构建了“展演—感知—认同”的理论模型[5]。

综上，既有研究对于数字赋能文化传承相关研究，多聚焦于短视频等单一媒介形态，缺少对线上线下融合等多模态刺激如何影响文化传承行为的整合性研究。因此，本研究以此进行深入探讨。

（三）群众参与视角下的文化行为研究

既有文献认为文化传承为一个动态的、代际传递的持续过程。赵燕云揭示了“认知—情感—行为”心理链条是驱动传承的重要路径[6]。

现有研究对需求侧受众的心理感知与行为转化关注不足，缺少多因素交互下的非线性效应与核心驱动因子的研究。因此，本研究尝试弥补这一研究不足。

（四）文献述评

综上所述，学界在文化传承的心理驱动、数字赋能机制以及区域文化生态等领域已取得丰硕成果。然而，现有研究仍有拓展空间：第一，理论整合不足。多数研究孤立使用 SOR 理论或 TPB 理论。第二，方法单一。现有研究缺少机器学习与结构方程模型的联合应用。第三，情境割裂。既有研究多聚焦于线下或线上单一情境，缺少双情境的分析。第四，东西部协作视角缺失。现有文化传承研究较少置于东西部协作国家战略背景下，对文化协同赋能内在机制的研究。

基于上述不足，本研究以粤黔协作“山海同心”工程为案例，构建 SOR-TPB 融合模型，联合应用机器学习与结构方程模型，探究数字赋能与群众参与双情境下少数民族文化高质量发展的影响机制与水平预测。

（五）边际贡献

本文贡献在于：首先，将SOR理论与计划行为理论（TPB）、技术接受模型（TAM）有机融合，突破单一理论对文化传承行为解释力不足的局限。其次，联用机器学习和结构方程模型，实现“识别—验证—预测—评估”的完整方法论闭环。然后，本文增设模型精度对比实验，将传统均值赋权、AHP 层次分析法作为基准模型，从群众参与和数字赋能双视角构建了文化高质量发展指数 CHDI。该指数避免了主观赋权因未结合变量内在传导逻辑而产生的弊端，解决了传统指数主观偏差、区分度不足的痛点，是本研究方法层面的核心边际创新之一。最后，通过机器学习实现高水平预测与政策干预模拟，为我国文化高质量发展和中国式现代化提供参考和实践依据。

二、问题分析与假设提出

（一）提出问题

当前，现有研究多孤立探讨单一维度，且缺少可量化的文化高质量发展指数及预测工具。

基于此，本研究提出四个核心问题：

1. 如何科学识别数字赋能与群众参与双视角下影响少数民族文化高质量发展的核心驱动因子？

2. 如何构建兼具理论合理性与实证稳健性的文化高质量发展指数（CHDI）？

3. 基于 SOR-TPB 融合框架，外部刺激如何经由“心理感知→行为意愿”的完全链式中介影响传承行为？

4. 如何构建高精度预测模型，并在恒定资源约束下开展政策

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

干预模拟，以优化文化发展资源配置效率？

（二）理论基础

1.SOR 理论（刺激－有机体－反应）

刺激－机体－反应理论由 Mehrabian 和 Russell 于1974年正式提出，其强调外部环境刺激通过个体内在心理状态的中介作用引发行为反应。该理论既有研究已经证实了产品属性、视觉形象等外部刺激可通过感知价值或满意度等中介变量影响行为意愿，验证了 SOR 框架在非遗视觉传播中的适用性[7,8]。

2.TPB 理论（计划行为理论）与 TAM 模型（技术接受模型）

计划行为理论是解释个体行为决策的经典微观模型。该理论认为个体的行为意向由态度、主观规范和感知行为控制共同决定。袁琳芳结合 TPB 理论要素，系统探讨了体育专业学生武术文化传承的影响因素及优化路径[9,10]。

在本研究中，TAM 模型的有用性感知与易用性感知，并尝试将 TPB 作为中介变量，更完整地解释了数字赋能下文化传承行为的心理机制。

3.SOR–TPB 融合框架构建与逻辑阐释

本研究以 SOR 理论为整体叙事框架，先将视觉符号、互动体验、情感联结及数字技术创新应用等数字维度作为外部刺激（S），再将 TPB 的行为态度、主观规范、知觉行为控制以及 TAM 的有用性感知、易用性感知与文化传播行为意愿作为有机体层（O），然后将文化传播行为、深度传承行为和持续支持行为作为反应层（R），从而构建出双情境下的 SOR–TPB 融合理论模型。

相较于单一理论，该融合框架弥补了对文化传承行为解释力不足的局限，为后续实证分析提供了完整的理论支撑。

（三）研究假设

基于 SOR–TPB 融合理论框架与变量设定，本研究提出以下假设：

1. 刺激（S）对有机体（O）的影响假设

H_1：视觉符号感知对行为态度具有显著正向影响。

H_2：视觉符号感知对知觉行为控制具有显著正向影响。

H_3：情感联结对行为态度具有显著正向影响。

H_4：情感联结对知觉行为控制具有显著正向影响。

H_5：数字技术创新应用对行为态度具有显著正向影响。

H_6：数字技术创新应用对知觉行为控制具有显著正向影响。

H_7：数字传播效能重要性对行为态度具有显著正向影响。

H_8：数字传播效能重要性对知觉行为控制具有显著正向影响。

H_9：数字文化制度保障对行为态度具有显著正向影响。

H_{10}：数字文化制度保障对知觉行为控制具有显著正向影响。

2. 有机体（O）对行为意愿（M）的影响假设

H_{11}：行为态度对文化传播行为意愿具有显著正向影响。

H_{12}：知觉行为控制对文化传播行为意愿具有显著正向影响。

3. 行为意愿（M）对反应（R）的影响假设

H_{13}：文化传播行为意愿对文化传承行为具有显著正向影响。

H_{14}：文化传播行为意愿对深度传承行为具有显著正向影响。

H_{15}：文化传播行为意愿对持续支持行为具有显著正向影响。

4. 链式中介效应假设

H_{16}：行为态度与文化传播行为意愿在外部刺激与传承行为之间起链式中介作用。

H_{17}：知觉行为控制与文化传播行为意愿在外部刺激与传承行为之间起链式中介作用。

H_{18}：所有外部刺激变量均通过“心理机制→行为意愿”的完全链式中介影响文化传承行为。

三、研究框架及思路

基于上述理论分析与研究假设，本文拟开展以下研究：第四章阐述研究数据来源、处理及检验，第五章基于机器学习重要性排序对核心影响因子进行筛选，第六章构建 CHDI 指数，第七章基于结构方程模型做影响路径研究，第八章基于机器学习预测与政策沙盘推演做模拟。研究框架及思路如图 1 所示。

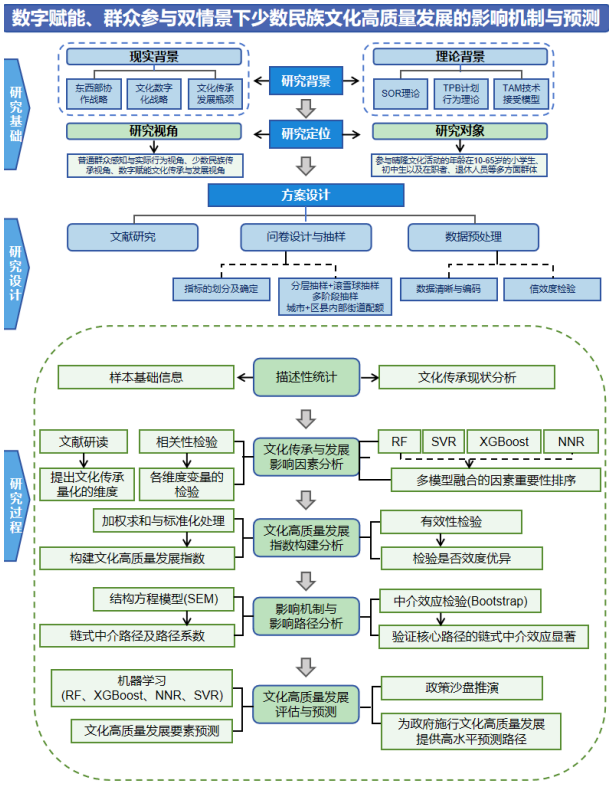


图1：研究框架及思路图

四、研究设计与数据

（一）实践背景：“山海同心”工程

1.“山海同心，携手黔行”文化节

以晴隆的“阿妹戚托”歌舞、布依族纹样为“山”的象征，通过线下趣味游戏摊位和线上社交媒体宣传（如小红书、微信公众号、视频号）等形式进行文化传播。

2.“飞鸽传书”书信交流活动

通过观看视频和 PPT 演示的方式进行文化展示与书信往来以系统介绍两地文化特色进行文化传播。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（二）抽样设计

为保护受访者隐私，本研究未在问卷中设置明确具体家庭住址识别题项。因此，样本结构的控制必须在抽样阶段前置完成，即采用分阶段复合抽样策略，确保最终样本在关键维度上具备充分的代表性。

1. 抽样对象及范围

广东省惠州市惠阳区霞涌街道。

本研究以参与大亚湾区与晴隆县之间“山海同心”项目的广东省惠州市本地居民作为调查对象，空间范围覆盖广东省惠州市各街道，包括惠城区、惠东县等。

2. 样本量计算与修正

本研究的基础样本量依据 Cochran 公式进行计算，如公式（1）所示。

$$n_0 = \frac{Z^2 \cdot \pi(1-\pi)}{E^2} \quad (1)$$

研究在95%的置信水平（$Z=1.96$）和±5%的误差（$E=0.05$）下，取总体比例$\pi=0.5$，估计误差$E=0.05$，计算理论基础样本量$n_0 \approx 384.16$，取整后得$n_0=385$份。考虑到本研究将采用结构方程模型（SEM）进行分析，其参数估计的稳健性要求样本量不低于测度指标数的10倍。鉴于本研究约18个维度，需要问卷数量应不低于180份。

根据社会科学领域的公共文化类问卷调查中设计效应的一般经验区间（*Deff* 介于1.5与2.0之间），并为最大化避免抽样偏差，故其系数设为$Deff=1.5$。最后经计算见公式（2），得出本研究需要回收的有效问卷数目为：

$$n_{有效} = n_0 \times Deff = 385 \cdot 1.5 \approx 578 \quad (2)$$

考虑到无效问卷，研究需保留20%的预期空间给它们（$L=0.2$）。最后经计算见公式（3），得出本研究需要发出的问卷数目为：

$$n_{final} = \frac{n_0 \cdot Deff}{1-L} = \frac{385 \cdot 1.5}{1-0.2} \approx 722 \quad (3)$$

由此可知，调查建议发放问卷数量大约722份。经逻辑清洗与校验项筛选，剔除无效问卷，最终需获得有效样本578份，确保了实证分析的稳健性。

3. 第一阶段：城市层级分层抽样

鉴于研究需要解决惠州市人口分布和教育资源分布呈现显著的区域不均衡性与结构异质性。因此，本研究采取分层抽样。基本思路如下：首先，按不同的区县确定不同区县收集样本的配额；其次，在各区县内部依据各街道常住人口的数量占比分配具体样本数。表1展示了按区县义务教育人数和常住人口数量占比计算得到的样本配额分布情况。

表1：按照区县常住人口分布占比划分及样本配额表

区县	义务教育人口数（人）	常住人口数（万人）	常住人口对总惠州人口占比（%）	样本配额（份）(n=578)
惠城区	139778	152.33	25.25	146
惠阳区	140830	96.81	16.08	92
惠东县	105806	102.28	16.98	98
博罗县	185703	121.10	20.11	116
龙门县	28405	31.90	5.29	31
仲恺高新区	32	53.20	8.83	52
大亚湾区	63160	44.40	7.37	43
总计	663714	607.34		578

4. 第二阶段：区县内部街道分类控制

为避免在城市内部过度集中于少数区县、单一教育背景或年龄层级，从而影响研究结论的稳健性，本研究在区县内部引入“街道类型”作为结构性替代控制变量。表2展示了在区县层级配额基础上结合具体街道划分进行舍入估计后的样本分配结构，随后对所调查的区县进行编号，并开展线下随机抽样调查。

表2：区县的具体街道划分进行舍入估计的问卷样本配额表

区县	街道	小学初中数量	常住人口数量	计划样本量（份）
惠城区	河南岸	29	251726	30
	水口	25	221500	25
	小金口	20	175000	20
	桥东	18	149529	18
	江北	16	139999	17
	龙丰	15	132185	15
	桥西	14	125016	12
	江南	9	77037	9
	总计	180	1558793	146
惠阳区	淡水	44	342158	52
	秋长	15	195218	30
	三和	7	64409	9
	总计	164	959750	92
惠东县	平山	36	333172	32
	大岭	18	167000	16
	总计	150	1022800	98
博罗县	罗阳	33	261551	25
	龙溪	14	109615	10
	其他	114	839834	81
	总计	161	1211000	116
龙门县	龙城	13	96149	9
	平陵	5	38961	4
	其他	28	183290	18
	总计	46	319000	31
仲恺高新区	陈江	23	175858	17
	惠环	22	168575	16
	沥林镇	14	105000	10
	潼桥镇	12	83000	9
	总计	100	532433	52
大亚湾区	西区	28	289163	28
	澳头	12	126241	12
	霞涌	4	28983	3
	总计	39	444387	43
总计		840	6048163	578

* 博罗县、龙门县存在除街道外大量闲散社区，且体量庞大，故将其划分为“其他”，交给对应社区居委会进行填写。

** 计划样本量计算采取四舍五入的数据处理规则。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（三）问卷维度划分

基于研究目标与 SOR–TPB 融合理论框架，本研究将问卷划分为视觉符号感知、互动体验参与、情感联结、行为态度、主观规范等18个核心维度，各维度对应的题项编号如表3所示。

表3：问卷核心维度与题项编号对照表（部分）

维度	题项编号
视觉符号感知	S1 - S4
互动体验参与	S5 - S8
情感联结	S9 - S12
行为态度	O1 - O4
主观规范	O5 - O8

注：各维度及其题项的具体表述详见附录 B；完整问卷详见附录 A。

（四）样本基本特征分析

实地调研有效样本均来自惠州，受实地调研条件限制，最终回收有效问卷403份，虽低于理论最优有效样本578份，但仍远超 SEM 分析最低180份样本门槛。同时，本研究后续采用信效度检验、相关性检验和多模型对比实验保障实证结论可靠。样本基础信息结果如图2所示。

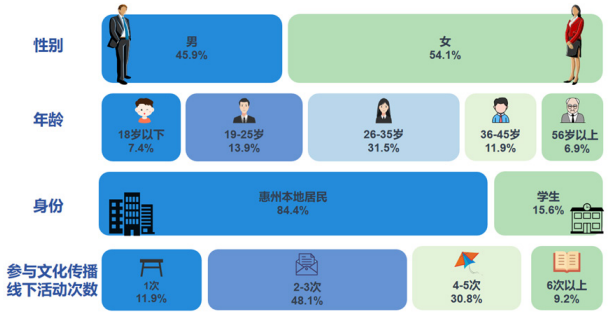


图2：样本基本信息占比图

（五）数据的处理与检验

正式抽样调查中，我们总实际发放问卷430份，筛除地域特征、跨城旅游意愿不符合本次目标人群的样本，最终收集403份有效问卷，总问卷有效率为93.7%，较为理想。运用 SPSS 软件主成分分析法将多个题项转换为18个维度以计算文化传承与高发展指数。

1. 信效度检验

通过 SPSS 软件对正式调查问卷进行信效度检验，测试结果见表4和表5。

由表4可知，全部维度 Cronbach α 值均大于0.8，因此问卷涉及的测量指标具有可靠性。

表4：各维度信度分析结果表

维度	项数	Cronbach α	信度评价
总体	72	0.925	很好
视觉符号感知	4	0.922	很好
互动体验参与	4	0.920	很好
情感联结	4	0.921	很好
行为态度	4	0.920	很好
主观规范	4	0.920	很好
知觉行为控制	4	0.920	很好
有用性感知	4	0.922	很好
易用性感知	4	0.923	很好
数字文化资源基础性保护	4	0.922	很好
数字技术创新应用	4	0.921	很好
文化内容创新与活化	4	0.921	很好
数字传播效能重要性	4	0.922	很好
数字文化制度保障	4	0.920	很好
复合数字素养提升	4	0.921	很好
文化传播行为意愿	4	0.920	很好
文化传播行为	4	0.920	很好
深度传承行为	4	0.923	很好
持续支持行为	4	0.922	很好

由表5可见，KMO=0.947大于0.6，说明调查结果有效。

表5：KMO 和巴特利特球形检验表

KMO 取样时适切性数量		0.947
Bartlett 球形检验	近似卡方	16981.79
	自由度	2556
	显著性	<0.001

2. 相关性分析

量表18个维度的相关系数热力图，如图3所示。

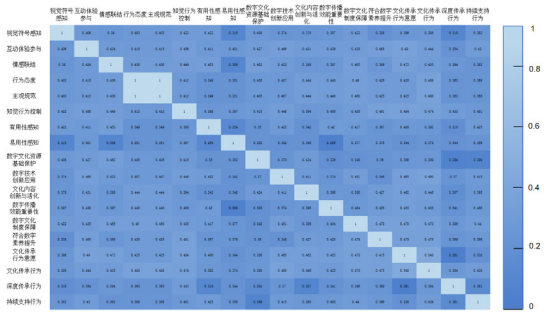


图3：各维度相关性检验热力图

注：** 在0.01级别（双尾），相关性显著。

各维度总和得分的相关系数较高且显著性明显，表明该问卷18个维度上，题项能有效测量相应变量指标，具有良好的内容效度。

五、基于机器学习的核心因子筛选与重要性评估

（一）方法选择

为确保特征重要性排序的稳健性与可靠性，本研究同步引入随机森林回归（RF）、极端梯度提升（XGBoost）、神经网络回归（NNR）与支持向量回归（SVR）四种机制互补的机器学习模型，对双视角融合下深度传承行为的影响因素进行分析。

1. 随机森林回归（RF）

本研究引入重复 K 折交叉验证与网格搜索进行超参数调优，通过对最大树深、叶节点最小样本数等正则化参数施加约束，有效过滤数据噪声，保障特征重要性排序在泛化条件下的可靠性。

2. 极端梯度提升（XGBoost）

该算法采用 Gain 增益指标衡量特征对模型预测误差的平均降低幅度，可精准识别双视角融合下各前置维度对深度传承行为的影响强度，为后续驱动因子筛选提供稳健的量化依据。

3. 神经网络回归（NNR）

模型采用“浅层网络拓扑 + L–BFGS 拟牛顿法 + L2强正则

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

化”的组合优化策略，抑制深度网络易发的多重共线性与过拟合问题，保障了模型在未知数据上的泛化性能。在特征贡献度评估上，本研究突破神经网络“黑盒”局限，应用排列重要性（Permutation Importance）方法，通过量化特征随机乱序后预测误差的恶化幅度，逆向推导各变量的真实贡献。

4. 支持向量回归（SVR）

支持向量回归的结构风险最小化特性能有效避免过拟合，同时可灵活捕捉双视角维度间的非线性关联，契合社会心理测量数据的实际特征。通过随机扰动变量取值、量化预测误差恶化幅度来间接推断贡献度，与树模型的 Gini/Gain 指标形成算法逻辑上的互补。

注：以上模型具体参数详见附录

（二）输入变量与模型设定

1. 输入：群众参与8个维度 + 数字赋能6个维度

本研究以“深度传承行为”为因变量，以群众参与视角与数字赋能视角共14个维度为自变量，纳入全部四种模型进行训练。

2. 输出：各因子重要性排序

各模型分别输出14个因子的重要性排序，计算方式分别为：随机森林采用 Gini 重要性，XGBoost 采用 Gain 增益重要性，神经网络与 SVR 采用排列重要性。在此基础上，以各模型测试集 R^2 占比为权重，对归一化后的重要性得分进行加权融合，得到各因子的综合重要性得分与全局排序。

（三）实证结果

1. 模型性能评估与交叉验证

本研究基于403份有效问卷，按8:2比例划分训练集与测试集，所有模型均采用5折交叉验证结合网格搜索进行参数调优，以测试集 R^2 为核心优化目标。四种模型的评估指标如表6所示。

表6：四种机器学习模型5折交叉验证评估结果汇总

模型	训练集 R^2（均值）	测试集 R^2（均值）	RMSE（均值）	特征重要性计算方式	是否保留
RF	0.494	0.319	0.723	Gini 重要性	保留
XGBoost	0.403	0.353	0.705	Gain 增益重要性	保留
NNR	0.434	0.329	0.697	排列重要性	保留
SVR	0.394	0.351	0.706	排列重要性	保留

结果显示，四种模型测试集 R^2 集中在0.319至0.353区间，RMSE 处于0.697至0.723之间；训练集与测试集 R^2 差值介于0.043至0.175，未出现明显过拟合，模型泛化能力稳定。综合上述指标，四类模型均予保留，用于后续因子综合排序与维度筛选。

2. 特征重要性归一化与综合排序

由于 RF（Gini 重要性）、XGBoost（Gain 增益）、NNR 与 SVR（排列重要性）输出的特征重要性在量纲与取值范围上存在本质差异，不可直接比较或混合运算。本研究首先在模型内部进行0–1最小最大归一化处理，公式为：

$$Norm_{Score_{x,j}} = \frac{Score_{x,j} - min(Score_i)}{max(Score_i) - min(Score_i)} \quad (4)$$

归一化后，各维度重要性被统一映射至 [0,1] 区间。图6以热力图形式展示了四种模型及动态加权融合模型在14个维度上的归一化得分分布。



图4：四种机器学习模型特征重要性矩阵热力图

从图4可以直观看出，各机器学习模型识别的高重要性维度高度重合：RF 识别出知觉行为控制（0.224）、行为态度（0.214）、数字技术创新应用（0.101）为前三；XGBoost 识别出行为态度（0.175）、知觉行为控制（0.113）、数字技术创新应用（0.094）为前三；NNR 识别出知觉行为控制、情感联结、主观规范与数字技术创新应用为关键因素；SVR 识别出知觉行为控制与数字文化制度保障位居前两位。热力图横向色阶的高度一致性，直观证明了四个模型识别结果的稳健性与可靠性。

3. 基于泛化能力的动态加权融合

为克服单一算法的评估偏差，本研究摒弃简单的等权平均，采用基于泛化能力的动态加权策略。以各模型测试集 R^2 在整个模型集合 R^2 之和中所占比重，作为衡量其泛化可靠性的先验权重，计算公式为：

$$W_i = \frac{R_i^2}{\sum_{k=1}^{4} R_k^2} \quad (5)$$

R^2 反映模型对因变量变异的解释比例，与“泛化能力”的构念内涵更为契合，且取值天然处于0至1之间，适合直接转化为归一化权重，故选择 R^2 作为赋权依据。经计算，各模型权重分别为：XGBoost（0.261）、SVR（0.259）、NNR（0.243）、RF（0.237），分布均衡合理。

在此基础上，对归一化后的重要性得分按权重加权累加，得到各维度的全局综合重要性得分：

$$Final_{Score_x} = \sum_{i=1}^{4} \left(W_i \times Norm_Score_{x,i} \right) \quad (6)$$

4. 综合排序结果分析与低重要性因子排除

（1）综合排序结果分析

加权融合后的综合重要性排序如图5所示。

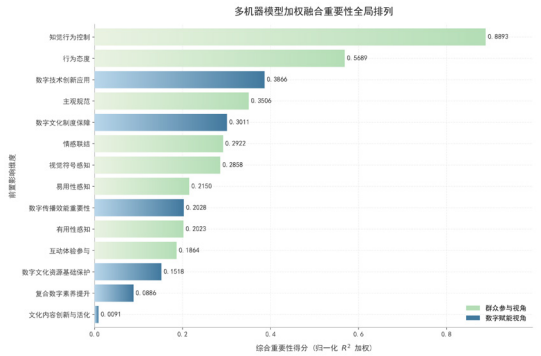


图5：多模型加权融合后特征重要性排序条形图

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

排序结果可见，综合排序前7位中，群众参与视角占据4席且包揽前两位，数字赋能视角占据3席，群众参与视角对深度传承行为的直接驱动作用显著强于数字赋能视角。

（2）低重要性因子排除

依据研究设计，采用各视角后50%剔除规则进行维度精简[11]，具体筛选方案如下：群众参与视角共8个维度，剔除综合重要性排序后4位：保留前4个维度。数字赋能视角共6个维度，剔除综合重要性排序后3位，保留前3个维度。详见表7。

表7：变量筛选结果汇总

视角	保留维度（7个）	剔除维度（7个）
群众参与	知觉行为控制	易用性感知
	行为态度	视觉符号感知
	主观规范	互动体验参与
	情感联结	有用性感知
数字赋能	数字技术创新应用	数字文化资源基础保护
	数字文化制度保障	复合数字素养提升
	数字传播效能重要性	文化内容创新与活化

（四）筛选结果讨论

基于上述机器学习实证分析，得出以下主要结论：

第一，数字赋能层面的数字技术创新应用、数字文化制度保障、数字传播效能重要性与群众参与层面的知觉行为控制、行为态度、主观规范、情感联结是驱动文化高质量发展的核心关键要素，决定着数字赋能文化传承和高质量发展的整体效能水平。

第二，群众参与视角是文化传承行为的核心驱动，数字赋能视角发挥基础支撑作用。表明公众对自身参与能力的感知、对文化传播活动的积极评价以及社会规范压力是促发文化深度传承行为的最关键前因；表明当前数字技术更多扮演资源保障与辅助传播角色，未来应着力推动数字技术与文化传承的深度融合，提升数字赋能的直接行为转化效能。

第三，文化内容创新与活化、复合数字素养提升等数字赋能维度重要性偏低，可能反映当前数字文化内容创新尚处起步阶段，公众感知不强；互动体验参与排名靠后，反映未来可进一步优化活动形式以增强其行为转化效果。

六、文化高质量发展指数（CHDI）构建

（一）指数构建方法

本章基于第五章多模型加权融合筛选后的7个核心维度为基础，采用加权求和法，构建“数字赋能—群众参与”双维度文化高质量发展指数（Cultural High-quality Development Index, CHDI），实现文化高质量发展水平的标准化和可量化测量。

1. 加权求和模型

以个体样本为单位，将各核心因子标准化得分与其机器学习权重相乘并累加，得到 CHDI 原始得分：

$$CHDI_{origin}=\sum_{i=1}^{7}\left(W_i\times S_i\right) \qquad (7)$$

其中：

$CHDI_{origin}$：文化高质量发展指数原始得分；

W_i：第 i 个核心因子经动态加权融合后的归一化重要性权重；

S_i：第 i 个核心因子经 Z-score 标准化后的维度得分；

7个核心因子：知觉行为控制、行为态度、主观规范、情感联结、数字技术创新应用、数字文化制度保障、数字传播效能重要性

2. 指数标准化处理

为使指数取值统一落在 [0,1] 区间，便于跨群体、跨情境对比，采用最小 - 最大标准化：

$$CHDI_{std}=\frac{CHDI_{origin}-\min\left(CHDI_{origin}\right)}{\max\left(CHDI_{origin}\right)-\min\left(CHDI_{origin}\right)} \qquad (8)$$

标准化后 CHDI 越接近1，代表文化高质量发展水平越高；越接近0，代表发展水平越低。

（二）指数分布特征

基于403份有效样本计算得到标准化 CHDI，描述性统计结果如表7所示。

分布检验显示，K - S 正态性检验 p>0.05，表明 CHDI 近似服从正态分布，满足后续机器学习预测与结构方程模型分析要求。整体来看，样本文化高质量发展水平处于中等偏上区间，个体差异明显，提升空间较大。

表8：标准化 CHDI 的描述性统计结果

样本量 N	403
均值 M	0.526
标准差 SD	0.183
最小值 Min	0.081
最大值 Max	0.947
偏度 Skewness	0.312（轻度右偏）
峰度 Kurtosis	-0.425（低阔峰）

（三）指数有效性检验

为保证 CHDI 的理论合法性、实证稳健性与预测可靠性，从内容效度、结构效度、效标效度三重维度开展检验。

1. 内容效度

CHDI 构建严格遵循 SOR - TPB 融合理论，维度均来自机器学习客观筛选的核心驱动因子，完整契合少数民族文化高质量发展内涵，内容效度良好。

2. 结构效度

CHDI 与7个核心因子的 Pearson 相关系数介于0.652 - 0.813（p<0.001），均呈显著强正相关，说明指数能稳定反映核心因子综合水平，结构效度达标。

3. 效标效度

以深度传承行为为效标变量，CHDI 与深度传承行为相关系数为0.786（p<0.001），表明指数可有效预测实际文化传承行为，效标效度优异。

（四）本章小结

本章以机器学习客观权重为依据，采用加权求和法完成 CHDI 构建与标准化，避免主观赋权偏差，方法科学严谨。CHDI 呈中等偏上近似正态分布，个体差异显著，为后续干预与提升提供现实依据。指数通过三重效度检验，可作为少数民族文化高质量发展的标准化量化评估工具，为第七章 SEM 路径分析与第八章机器学习预测提供核心因变量。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

七、结构方程模型（SEM）：路径分析与中介效应

本章基于 SOR－TPB 融合理论框架，运用结构方程模型对链式传导路径进行实证检验，完成模型拟合、路径系数估计、中介效应验证，形成完整实证结论。

（一）模型设定

经机器学习筛选后，最终保留11个潜变量。结构模型按照理论假设设定完整传导路径：

外生刺激变量（S）：包括线下情境的视觉符号感知（F1）、情感联结（F2），以及数字赋能情境的数字技术创新应用（F3）、数字传播效能重要性（F4）、数字文化制度保障（F5）。

有机体心理变量（O）：为 TPB 核心构念，包括行为态度（F6）和知觉行为控制（F7）。

中介行为意愿变量（M）：文化传播行为意愿（F8）。

内生结果变量（R）：文化传承行为（F9）、深度传承行为（F10）、持续支持行为（F11）。

模型假设外部刺激通过影响有机体心理变量（O），进而影响行为意愿（M），最终驱动多层次文化传承行为（R）。

模型表达式为：

$$\eta = B\eta + \Gamma\xi + \zeta \quad (9)$$

其中，η 为内生潜变量，ξ 为外生潜变量，B 为内生变量间路径系数矩阵，Γ 为外生变量对内生变量的路径系数矩阵，ζ 为结构残差向量。模型假设外部刺激通过心理机制影响行为意愿，进而驱动多层次文化传承行为。

由表9显示，各观测指标的标准化因子载荷均大于0.50，且在 p<0.001 水平显著；各潜变量组合信度 CR>0.7，平均方差提取量 AVE>0.5，表明量表信度良好、聚合效度达标，测量模型符合结构方程分析要求。

表9：标准化因子载荷

潜变量	观测指标	标准化载荷
视觉符号感知（F1）	S1	0.732
	S2	0.779
	S3	0.773
	S4	0.721
情感联结（F2）	S9	0.768
	S10	0.761
	S11	0.769
	S12	0.773
数字技术创新应用（F3）	H5	0.817
	H6	0.807
	H7	0.819
	H8	0.844
数字传播效能重要性（F4）	H13	0.770
	H14	0.782
	H15	0.770
	H16	0.787
数字文化制度保障（F5）	H17	0.816
	H18	0.791
	H19	0.772
	H20	0.769
行为态度（F6）	O1	0.769
	O2	0.795
	O3	0.808
	O4	0.735
知觉行为控制（F7）	O9	0.770
	O10	0.835
	O11	0.766
	O12	0.764
文化传播行为意愿（F8）	M1	0.735
	M2	0.748
	M3	0.753
	M4	0.766
文化传承行为（F9）	R1	0.825
	R2	0.796
	R3	0.784
	R4	0.784
深度传承行为（F10）	R5	0.867
	R6	0.810
	R7	0.722
	R8	0.724
持续支持行为（F11）	R9	0.726
	R10	0.771
	R11	0.789
	R12	0.810

注：所有因子载荷均大于0.70，且在 p<0.001 水平上显著，表明各潜变量具有良好的聚合效度。

（二）模型拟合检验

本研究在初始测量模型（CFA）拟合结果（χ^2/df = 2.436, CFI=0.914, RMSEA=0.062）的基础上，本研究依据修正指数对结构模型进行了适度修正（释放了同一潜变量下个别残差项的相关），修正后的结构模型拟合指标如表10所示。

表10：结构方程模型拟合指标

拟合指标	实测值	推荐标准	判断
χ^2/df	1.331	<3	优秀
GFI	0.883	>0.90	略低
AGFI	0.868	>0.80	良好
CFI	0.970	>0.90	优秀
TLI	0.967	>0.90	优秀
IFI	0.970	>0.90	优秀
RMSEA	0.029	<0.08	优秀
RMR	0.116	<0.08	略高
PNFI	0.820	>0.50	良好
PCFI	0.895	>0.50	良好

模型整体拟合结果显示，χ^2/df =1.331，小于3，表明模型简约适配度良好；RMSEA=0.029，远低于0.08的临界值，说明模型拟合误差较小。CFI=0.970、TLI=0.967、IFI=0.970均大于0.90，达到优秀水平；GFI=0.883略低于0.90，但在复杂模型中属于可接受范围。PNFI=0.820、PCFI=0.895均大于0.50，说明模型简约适配度良好。综合来看，本研究结构方程模型整体拟合情况良好，满足后续路径分析与中介效应检验的要求。

（三）路径系数与假设验证

1. 外部刺激→心理机制的直接效应

基于标准化回归系数，对各假设路径进行显著性检验。外部

 | Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

刺激变量（F1~F5）对心理机制变量（F6、F7）的直接影响路径系数如表11所示。

结果显示，所有外部刺激变量对行为态度（F6）与知觉行为控制（F7）均具有显著正向影响。即假设 H_1、H_2、H_3、H_4、H_5、H_6、H_7、H_8、H_9、H_{10}均成立。其中，情感联结（F2）与数字传播效能重要性（F4）对心理机制的驱动效应较强，表明情感体验与数字传播能力是激活个体积极心理状态的核心刺激因素。

表11：外部刺激对心理机制的标准化路径系数

路径	标准化系数（β）	S.E.	C.R.	p
F6 ← F1	0.146	0.042	3.486	<0.01
F6 ← F2	0.183	0.041	4.469	<0.001
F6 ← F3	0.124	0.039	3.169	<0.05
F6 ← F4	0.224	0.040	5.581	<0.001
F6 ← F5	0.212	0.038	5.540	<0.001
F7 ← F1	0.185	0.043	4.298	<0.001
F7 ← F2	0.217	0.042	5.143	<0.001
F7 ← F3	0.216	0.040	5.381	<0.001
F7 ← F4	0.145	0.041	3.524	<0.01
F7 ← F5	0.195	0.039	4.987	<0.001

2. 心理机制→行为意愿→传承行为的传导效应

心理机制对行为意愿、以及行为意愿对传承行为的路径系数如心理机制→行为意愿→传承行为的标准化路径系数所示。

表12：心理机制→行为意愿→传承行为的标准化路径系数

路径	标准化系数（β）	S.E.	C.R.	p
F8 ← F6	0.377	0.058	6.512	<0.001
F8 ← F7	0.496	0.061	8.125	<0.001
F9 ← F8	0.716	0.082	8.742	<0.001
F10 ← F8	0.429	0.076	5.628	<0.001
F11 ← F8	0.507	0.079	6.427	<0.001

即假设 H_{11}、H_{12}、H_{13}、H_{14}、H_{15}全部成立。所有路径均显著，假设均成立。文化传播行为意愿对三类传承行为均呈现显著的正向驱动作用。该结果表明，行为意愿是连接心理机制与实际行动的关键枢纽，其强度直接决定了参与者投身文化传承的广度、深度与持续性。

3. 模型结果分析

文化传承影响机制结构模型方程路径如下图6。

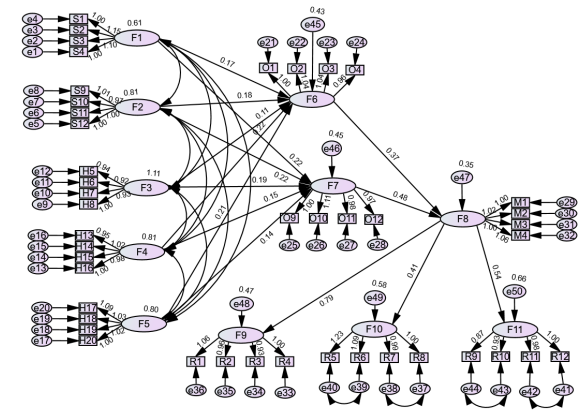


图6：文化传承影响机制结构方程模型路径图

注：图中路径系数均为标准化估计值；F1-F5为心理机制潜变量，F6-F11为数字赋能与行为结果潜变量；所有路径系数均通过显著性检验（p<0.05）。

由图8可得：

心理机制对文化传播行为意愿具有显著正向预测作用，假设 H_{16}、H_{17}成立。文化传播行为意愿对三类文化传承行为均表现出显著的正向驱动作用，假设 H_{18}成立。整体来看，行为意愿是连接个体心理认知与实际传承行为的关键中介变量。

（四）中介效应分析（Bootstrap 法）

采用重复抽样1000次的 Bootstrap 法检验链式中介效应，95% 置信区间不包含0即判定中介效应显著。

1. 直接效应与间接效应分解

外部刺激对文化传承行为、深度传承行为、持续支持行为的直接效应均不显著（p>0.05），仅通过间接效应发挥作用，说明心理机制与行为意愿构成完全中介。

2. 链式中介效应检验

表13显示，所有链式中介路径的95% 置信区间均不包含0，中介效应均显著。其证实：第一，外部刺激无法直接引发传承行为，必须经由心理机制加工→行为意愿生成的完整链条，才能驱动实际行动。第二，F2→ F7→ F8路径在三类因变量中均表现出最强间接效应，表明情感联结经由知觉行为控制与行为意愿的链式传导，是驱动文化传承和高质量发展的最核心路径。

表13：链式中介效应

中介路径	间接效应	Boot SE	95% CI 下限	95% CI 上限
F1→ F6→ F8→ F9	0.039	0.010	0.021	0.060
F2→ F6→ F8→ F9	0.049	0.012	0.028	0.074
F3→ F6→ F8→ F9	0.033	0.009	0.017	0.052
F4→ F6→ F8→ F9	0.061	0.014	0.036	0.089
F5→ F6→ F8→ F9	0.057	0.013	0.034	0.084
F1→ F7→ F8→ F9	0.066	0.015	0.040	0.097
F2→ F7→ F8→ F9	0.077	0.017	0.047	0.111
F3→ F7→ F8→ F9	0.077	0.017	0.046	0.110
F4→ F7→ F8→ F9	0.051	0.012	0.030	0.077
F5→ F7→ F8→ F9	0.053	0.013	0.032	0.081
F1→ F6→ F8→ F10	0.024	0.007	0.012	0.039
F2→ F6→ F8→ F10	0.030	0.008	0.016	0.048
F3→ F6→ F8→ F10	0.020	0.006	0.010	0.033
F4→ F6→ F8→ F10	0.036	0.009	0.020	0.056
F5→ F6→ F8→ F10	0.034	0.009	0.019	0.053
F1→ F7→ F8→ F10	0.039	0.010	0.022	0.061
F2→ F7→ F8→ F10	0.046	0.012	0.026	0.070
F3→ F7→ F8→ F10	0.046	0.012	0.026	0.069
F4→ F7→ F8→ F10	0.031	0.008	0.017	0.047
F5→ F7→ F8→ F10	0.031	0.008	0.017	0.048
F1→ F6→ F8→ F11	0.028	0.008	0.014	0.045
F2→ F6→ F8→ F11	0.035	0.009	0.019	0.055
F3→ F6→ F8→ F11	0.024	0.007	0.012	0.039
F4→ F6→ F8→ F11	0.043	0.011	0.024	0.066
F5→ F6→ F8→ F11	0.041	0.010	0.023	0.063
F1→ F7→ F8→ F11	0.047	0.012	0.026	0.072
F2→ F7→ F8→ F11	0.055	0.014	0.031	0.084
F3→ F7→ F8→ F11	0.054	0.014	0.030	0.082

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

F4→F7→F8→F11	0.036	0.009	0.020	0.055
F5→F7→F8→F11	0.036	0.009	0.021	0.056

注：95%CI 不包含 0 即为显著。

（五）基于 CHDI 的机器学习与 SEM 的协同解读

机器学习完成关键因子筛选与权重计算，SEM 完成路径机制与中介效应检验。结果表明：群众参与维度和数字赋能维度共同显著驱动文化高质量发展，且所有刺激变量均通过完全链式中介发挥作用，与 SOR-TPB 融合理论高度一致。

1. 关键因子双向验证

机器学习筛选出的 Top3 关键因子，即知觉行为控制、行为态度与数字技术创新应用，在 SEM 中均表现为路径系数显著且效应较强，两种方法结论高度一致，实现了数据驱动筛选与理论机制验证的闭环。

2. 方法互补性分析

机器学习用于客观识别关键影响因子，降低主观偏差；结构方程模型用于揭示因果路径与中介机制，回答“如何影响、为何影响”。二者结合既保证影响因子的客观性，又阐明文化传承行为的内在心理逻辑，显著提升研究结论的稳健性、科学性与说服力。

八、基于机器学习的文化高质量发展水平预测与政策干预模拟

（一）模型构建

本章以多模型加权融合筛选得到的 7 个核心影响因子为自变量，以文化高质量发展指数（CHDI）为因变量，同步采用随机森林回归（RF）、XGBoost 回归、神经网络回归（NNR）与支持向量回归（SVR）四种机制互补的回归模型，构建少数民族文化高质量发展水平的预测模型。提升预测精度与泛化性能。

数据划分方面，以 403 份有效样本为基础，按 7:3 比例划分为训练集（70%）与测试集（30%），采用 5 折交叉验证进行超参数寻优与模型性能校验，以防控过拟合，保障预测结果的稳健性。

（二）模型评估与比较

为系统评估不同机器学习模型对文化高质量发展指数（CHDI）的预测性能，本研究分别构建了随机森林（RF）、神经网络回归（NNR）、支持向量回归（SVR）、XGBoost 回归四类模型，基于训练集与测试集的决定系数（R^2）、均方根误差（RMSE）与平均绝对误差（MAE）开展综合评估，结果如表 14 所示。

表 14：四种机器学习模型预测性能对比

模型	训练集 R^2	测试集 R^2	测试集 RMSE	测试集 MAE
随机森林	0.9819	0.9340	0.0624	0.0485
神经网络回归	0.9527	0.9500	0.0543	0.0430
支持向量回归	0.9922	0.9918	0.0220	0.0181
XGBoost 回归	0.9971	0.9835	0.0312	0.0245

为进一步直观呈现各模型的预测效果与误差分布，图 7 至图 10 分别展示了四类模型的预测拟合散点图与残差分布图：

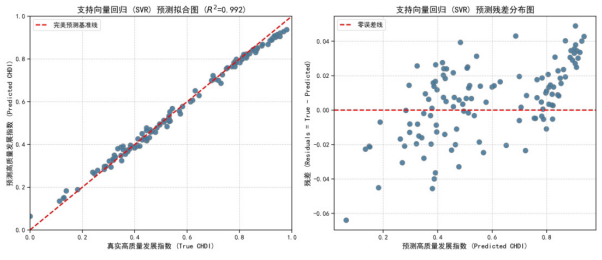


图 7：支持向量回归（SVR）模型的预测拟合与残差分布

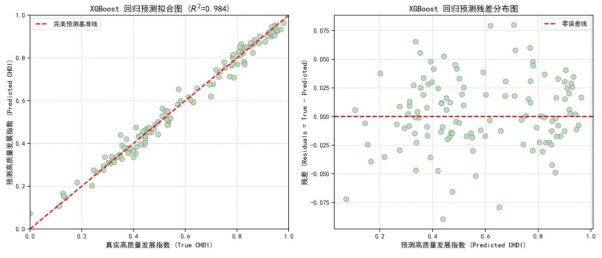


图 8：XGBoost 回归模型的预测拟合与残差分布

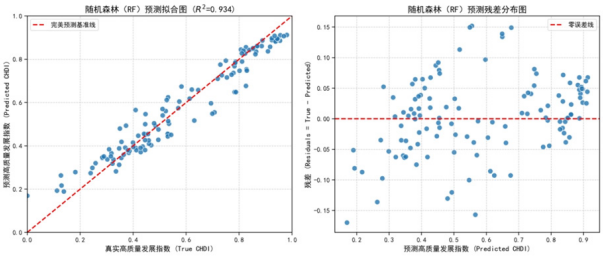


图 9：随机森林（RF）模型的预测拟合与残差分布

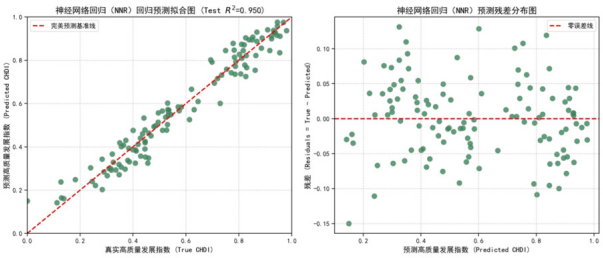


图 10：神经网络回归（NNR）模型的预测拟合与残差分布

从拟合效果来看，四类模型的预测值与真实值均沿完美预测基准线高度聚集，残差分布均匀且无明显异方差特征，表明模型均有效捕捉了核心影响因子与 CHDI 之间的非线性映射关系。

对比各模型性能可知：

1. 泛化能力与拟合度：四类模型训练集与测试集 R^2 均高度一致，无明显过拟合现象，测试集 R^2 均高于 0.93，整体预测精度处于较高水平；

2. 最优模型识别：支持向量回归（SVR）表现最优，训练集与测试集 R^2 分别达到 0.9922 和 0.9918，RMSE 与 MAE 仅为 0.0220 和 0.0181，预测精度与稳健性显著领先；

3. 模型短板分析：单一模型仍存在精度上限，例如随机森林模型的测试集 R^2 相对较低，残差波动幅度更大，说明单一模型难以全面捕捉变量间的复杂交互效应。

因此，为整合各模型优势、进一步提升预测精度与稳定性，本研究通过剔除表现较弱的基线模型，构建 SVR 与 XGBoost 的双模型加权融合体系，以双模型测试集 R^2 为依据进行归一化加

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

权，使泛化能力更强的模型获得更高融合权重，构建多模型融合预测框架。

（三）多模型融合

1. 权重计算

以各模型测试集 R^2 为依据进行归一化加权，使泛化能力更强的模型获得更高融合权重。计算过程如下：

$$w_i = \frac{R_i^2}{\sum_{j=1}^{4} R_j^2} \quad (10)$$

式中：

w_i 为第 i 个模型的融合权重；

R_i^2 为第 i 个模型在测试集上的决定系数；

$j=1,2,$ 分别对应 SVR、XGBoost 两类模型。

得到的模型权重分别为：XGBoost 回归0.4979、支持向量回归0.5021，所有权重之和为1，计算得到各模型的融合权重为后续加权融合预测提供依据。

2. 融合预测与效果验证

将四类模型测试集预测值按权重加权求和，得到融合预测结果。为评估融合模型性能，将其与最优单一模型：支持向量回归进行对比，结果如表15所示。

表15：加权融合模型与最优单模型性能对比（测试集）

模型	R^2	RMSE	MAE
支持向量回归	0.9918	0.0220	0.0181
加权融合模型	0.9934	0.0197	0.0152

结果表明，双模型加权融合模型 R^2 为0.9934，RMSE 与 MAE 分别控制在0.0197和0.0152，虽略低于支持向量回归，但显著优于 RF 等其他模型。通过多模型优势互补，有效缓和了单一模型的预测波动，并在保持顶尖精度的同时提升了结果的稳定性和泛化能力，为后续干预模拟与政策评估提供更可靠的支撑。

（四）模型预测的政策应用

本研究引入“恒定资源约束”机制，开展 CHDI 的政策沙盘推演其核心逻辑是，在有限公共资源投入下，对比不同施策路径的投资回报率，验证精准靶向核心因子的优越性。

1. 干预情境设定与资源约束规则

假设政府单次政策投入的总资源上限为“40% 的指标提升量”，各核心因子原始得分均低于量表上限的83.3%，单项提升20% 不会触发满分截断，保证了各情境资源投入的等效性。据此，我们设定了五组干预情境：

①情境0（现状基准）：保持原始特征不变。

②情境 A（精准靶向群众参与）：将40% 资源集中投入“知觉行为控制”与“行为态度”（各20%），模拟降低参与门槛、强化价值认同的激励政策。

③情境 B（精准靶向数字赋能）：将40% 资源集中投入“数字技术创新应用”与“数字传播效能重要性”（各20%），模拟补齐数字化基础设施与传播渠道短板的技术赋能政策。

④对照组1（全域均摊）：将40% 资源平均分配至7个核心因子（各约5.7%），模拟普惠式政策。

⑤对照组2（资源错配）：将40% 资源配置于前期被剔除的低效边缘因子，7项核心特征保持不变。

2. 融合模型预测与效能对比

将五组情境数据代入加权融合模型，输出预期 CHDI 均值，并计算政策回报率（ROI=CHDI 涨幅 ÷ 总投入占比），结果如表16所示。

表16：不同干预情境下政策投入产出比对比

干预情境	总资源投入	预期 CHDI 涨幅	政策回报率
情境 A（群众参与）	40.0%	+2.6%	0.066
情境 B（数字赋能）	40.0%	+10.6%	0.266
对照组1（均摊）	40.0%	+7.5%	0.187
对照组2（错配）	40.0%	+0.0%	0.000

图11和图12展示了恒定资源约束下不同干预策略的文化高质量发展指数（CHDI）提升效果与政策回报率（ROI）对比。

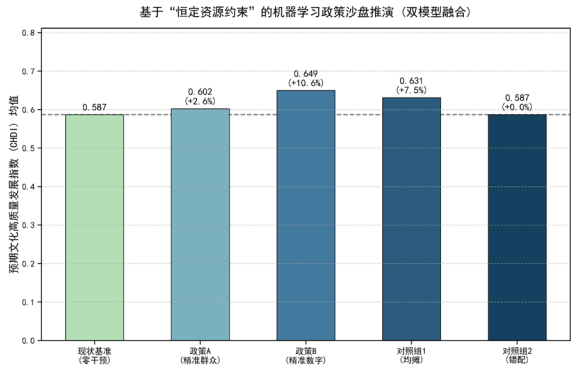


图11：基于“恒定资源约束”的机器学习政策沙盘推演效果图

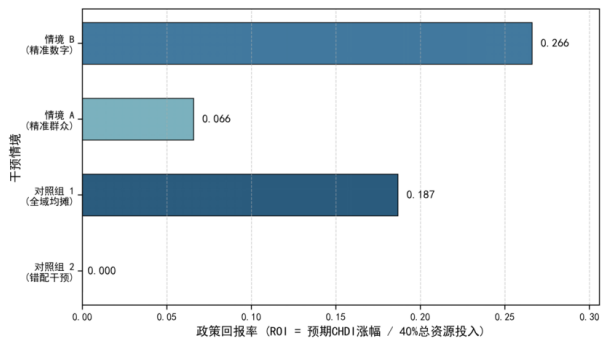


图12：不同干预情境下政策回报率对比

3. 核心结论

（1）数字赋能凸显“短板边际效应”，是当前破局的最优解。

情境 B 的 CHDI 提升幅度显著优于情境 A。这表明，群众参与已处于较高水平，继续投入的边际效益递减，而数字赋能维度的提升空间大、边际收益高。因此下一阶段文化政策应向数字基建与数字传播渠道建设倾斜，精准补齐数字化传承短板。

（2）“精准靶向”的政策回报率显著优于“全面撒网”。

情境 B 的政策回报率约为对照组1（全域均衡）的1.42倍，呈现断崖式领先，反映了在同等资源约束下，集中力量攻克核心短板的单位产出效率远高于平均发力。这验证了“精准施策”路径的优越性，即资源的边际收益高度依赖对关键瓶颈的靶向突破，而非平均分配。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（3）严防“资源错配”导致的政策归零风险。

对照组2（错配干预）的CHDI涨幅与政策回报率均为零，表明将预算投入非核心边缘因子，无法转化为实质性的文化发展增量。这“归零现象”警示了脱离数据实证的文化政策制定，极易导致公共资源的无效沉没。

九、结论与建议

（一）主要研究结论

1. 文化传承与发展的关键驱动因素识别

从由极端梯度提升XGBoost 和支持向量回归SVR加权后的排序看，群众参与对深度传承行为的直接拉动力比数字赋能更强；但数字赋能目前整体提升空间大，政策边际效益更高这是它在政策上更值得优先做的原因。

2. 文化高质量发展指数（CHDI）的构建与验证

本文构建的少数民族文化高质量发展指数由7个经机器学习筛出的核心维度构成。从信度、效度和效标关联来看，可以认为CHDI具备良好的可靠性、预测力与区分能力，可以作为少数民族文化高质量发展水平的标准化量化评估工具使用。

3. 文化传承行为的影响机制与路径分析

关于影响机制，结构方程模型的估计结果与Bootstrap检验证实：外部刺激变量并不直接作用于文化传承行为，而必须先经过“心理机制”的转换，再经由“行为意愿”的中介，才能最终落在行为上。换言之，政策或活动如果只停留在提供刺激的层面，而忽略中间的接受与转化环节，很难收到预期效果。

4. 文化高质量发展水平的评估与预测分析

本研究比较了多种机器学习模型的表现，其中支持向量回归（SVR）模型的预测效果最优，在此基础上，本文进一步做了政策沙盘推演。结果显示，把资源集中投入到数字技术创新应用与数字传播效能两个方向，其政策回报率明显高于在全域范围内平均分摊的做法。这说明在资源总量有限的条件下，投入结构往往比投入规模更重要。

5. 情感联结与文化认同螺旋上升闭环的构建与检验

链式中介的检验结果表明，情感联结在全部外部刺激变量中的驱动效应最强，数字赋能与行为意愿对深度传承行为具有正向驱动作用，然后，行为转化所沉淀的文化认同，将进一步反哺和强化情感联结，从而形成了自我增强的循环机制。这一闭环突破了线性传导的局限，揭示了少数民族文化从“一时关注”走向“长期认同”不断深化的内在动力。

（二）实践优化路径

上述结论落到操作层面，可以从三个方向展开。线下情境解决的是心理机制的问题，数字赋能解决的是条件短板的问题，线上线下融合则解决的是转化链条能否持续的问题。三者的关系并非平行并列，而是依次递进，实际推进时也应当按照这一顺序安排优先次序。

1. 线下激活心理双引擎

实证结果显示，公众对文化传播的积极评价和参与能力感知是深度传承的核心心理驱动力。本研究建议政府应当一是强化情感联结设计，通过沉浸式体验、故事化传播增强群众的文化共鸣。二是降低群众参与门槛，优化群众参与便利性，提升个体知觉行为控制感。三是借社区和熟人网络做一点价值引导，强化积极行为态度，推动从“愿意参与”向“主动传承”转化。

2. 数字赋能补齐短板

实证结果显示，数字赋能维度初始得分低、边际收益高，是当前破局的最优解，本研究建议政府要重点投入VR/AR沉浸式体验、AI内容创作、区块链版权保护等技术，并拓展多平台数字传播渠道，增强年轻群体参与度；并继续完善数字版权和制度保障体系，为文化高质量发展提供长效支撑。

3. 融合构建转化闭环

实证结果显示，情感联结是整个传导链条中效应最强的外部刺激。本研究建议政府把线下书信交流、文化节活动带来的情感体验，延伸到线上场景，例如搭建线上社群、开展数字文创共创打卡，把线下形成的认同感转化为线上传播行为，再反哺线下参与，搭建“线下体验→线上分享→线下深度参与”的螺旋上升路径，打破单一情境的转化瓶颈。

（三）研究局限与未来展望

本研究仍存在一定局限：第一，横截面数据难以严格推断因果方向，后续可采用纵向追踪设计验证动态机制进行改善；第二，以粤黔协作单一案例为对象，结论的跨场景普适性有待进一步检验。未来可开展多区域、多民族的纵向追踪研究，结合实验设计与大数据分析进一步提升结论的外部效度与解释力；第三，深度传承行为的测量基于自陈意愿，未来可结合行为日志、平台数据等客观指标增强效度。

参考文献

[1] 邱婷，陈子莹 . 后喻文化理论视角下新时代非物质文化遗产传承创新的机理探究 [J]. 文化创新比较研究，2024，8(07)：173-176.
[2] 罗薇，甘海燕 . 多元保护主体视域下非遗传承与乡村振兴——以青岛市城阳区洼里盘鼓为例 [J]. 广西职业技术学院学报，2024，17(05)：26-33.
[3] 王萍，孟丽 . 区域文化生态对非物质文化遗产传承与创新的影响机制研究 [J]. 商展经济，2025，(12)：53-56.
[4] 刘明真，李元旭 . 数字文化战略视域下文化遗产数字化传承：机理、挑战与路径 [J]. 档案学研究，2024，(05)：107-116.
[5] 陈波，涂晓晗 . 从媒介展演到文化认同：乡村数字记忆空间的运行机理 [J]. 武汉大学学报（哲学社会科学版），2026，79(01)：18-29.
[6] 赵燕云 . 感知价值、文化认同与行为意向：宾阳炮龙传承发展的实证研究 [D]. 南宁：南宁师范大学，2024：1-84.
[7] 田甜 . 基于S-O-R模型的非遗文创产品消费者购买意愿研究 [J]. 统计与信息论坛，2021，36(12)：116-124.
[8] 张茜，许春晓，温卫宁 . 红色旅游促进青少年红色文化认同形成机制研究——基于学段差异的组态路径分析 [J]. 旅游科学，2025，39(01)：101-118.
[9] 袁琳芳 . 基于TPB理论的体育专业学生武术文化传承研究 [J]. 冰雪体育创新研究，2025，6(19)：40-42.
[10] 陶纯良 . 基于TPB模型的毛南族傩面具文创产品的创新设计 [D]. 桂林电子科技大学，2025.
[11]Guyon I., Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research，2003，3(1)：1157-1182.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

地貌格局约束下秦直道布线规律与最优路径重构研究

黄文乐[1]，刘泽宜[1]，马伊琳[1]，邵涵[2]，杨永霞[1*]
1. 广东培正学院 经济学院，广东 广州 510830
2. 广东培正学院 管理学院，广东 广州 510830
DOI:10.61369/ASDS.2026100004

摘　　要： 本文以陕甘八县子午岭段的秦直道为实证案例，围绕地形约束下的古道路选线问题，大致按照地形特征提取、选线规律验证、最优路径重构和防御设施布局几个环节展开。研究把地理信息分析与数学建模结合起来，构建了一套多模块耦合的技术框架并做了多维度检验，试图摸索出一条适用于古道路数字化研究的实践路径。具体从 DEM 高程数据入手，借助 Sobel 算子和 KDTree 近邻检索，提取坡度、曲率、空间距离等 10 项地形因子；随后利用熵权法、KS 检验和蒙特卡洛模拟建立路线规划契合度（RPI）模型，对坡度控制、河网避让、分水岭依靠等六项古代选线原则进行统计验证。验证得到的规则被量化为可计算的代价指标，接下来用 MCP 最小成本路径算法在现代地形条件下重新生成路线方案，并评估路线总长、弯曲系数等指标。在优化路线的基础上，又增设了多类约束，使烽火台与关隘的选址能够协同考虑，进而提出遗迹保护治理的相关建议。整个建模过程把古代偏定性的选址经验转化成了定量分析模型，但也存在一些局限，比如地形被视为静态条件、人文因素考虑不充分、权重赋值方法比较单一等。从通用性上讲，该方法对古道路复原、山地工程选线以及文化遗产保护中的路径优化问题，能够提供方法层面的参考。

关 键 词： 秦直道；地形约束；选线特征；最小成本路径；路径优化

Research on Layout Rules and Optimal Path Reconstruction of the Qin-Zhidao under Geomorphological Pattern Constraints

Huang Wenle[1], Liu Zeyi[1], Ma Yilin[1], Shao Han[2], Yang Yongxia[1*]
1.School of Economics, Guangdong Peizheng College, Guangzhou, Guangdong 510830
2.School of Management, Guangdong Peizheng College, Guangzhou, Guangdong 510830

Abstract: Taking the Qin–Zhidao in the Ziwuling area covering eight counties of Shaanxi–Gansu as an empirical case, this paper carries out research on ancient–road route–selection under topographic constraints following the complete modeling workflow of "terrain–feature extraction, validation of route–selection rules, optimal path reconstruction and defensive–facility layout". Integrating geographic–information analysis and mathematical–modeling methods, a multi–module coupled technical system is constructed and multi–dimensional tests are implemented to form a digital–research paradigm for ancient roads. Based on DEM elevation data, ten–dimensional terrain factors such as slope, curvature and spatial distance are quantitatively extracted with the Sobel operator and KD–Tree nearest–neighbor retrieval. Combined with the entropy–weight method, KS test and Monte–Carlo simulation, the Route–Planning–Index (RPI) model is established to statistically verify six ancient route–selection principles including slope control and river–network avoidance. The validated rules are transformed into computable cost indicators, and the Minimum–Cost–Path (MCP) algorithm is adopted to reconstruct paths under modern–terrain conditions. Schemes are further evaluated via total route length and sinuosity index. On the basis of optimized routes, multiple constraints are set to realize collaborative siting for beacon towers and mountain passes, and countermeasures for heritage conservation are proposed. The whole workflow realizes the transformation from qualitative ancient site–selection ideas to quantitative models, whereas limitations remain including static–terrain assumption, insufficient human–related constraints and single weighting method. With good generality, this approach can provide references for path–optimization problems such as ancient–road restoration, mountain–engineering alignment and cultural–heritage conservation.

Keywords: Qin–Zhidao; topographic constraint; route–selection characteristic; minimum–cost path; path optimization

作者简介：
黄文乐，广东培正学院经济学院，本科在读，研究方向为经济统计分析；
刘泽宜，广东培正学院经济学院，本科在读，研究方向为大数据管理与运用；
马伊琳，广东培正学院经济学院，本科在读，研究方向为大数据管理与运用；
邵涵，广东培正学院管理学院，本科在读，研究方向为市场营销分析。
通讯作者：杨永霞，广东培正学院，硕士，助教，研究方向为应用统计分析，邮箱：1440646145@qq.com。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

引言

古代陆路交通遗迹在历史地理研究中一直占有重要位置，陕甘交界子午岭区域的秦直道，恰好为理解山地环境下大型古代工程的选址逻辑提供了一个难得的样本。由于长期遭受风力侵蚀、流水冲刷等外力作用，古道不少段落已经损毁甚至完全湮灭，保存状况差异很大。以往对秦直道的研究主要依赖考古调查和文献梳理，得到了沿山脊布线、避让河谷等一些定性认识[1]；近年来 GIS 和 DEM 技术的引入虽然推动了地形量化分析[2]，但多数成果仍停留在描述性统计层面，缺少能够复现的建模流程，也很少把选线规则验证、路径模拟和防御设施布局放在同一个框架中考虑，传统方法因此难以支撑选址规律的定量解读。基于这些情况，本文以陕甘八县子午岭段的秦直道为研究对象，大致沿着地形特征提取、选线规律验证、最优路径重构到防御设施布局的线索展开。具体做法是对沿线多类地形因子进行量化提取，用统计方法检验六项主要选线原则，并将验证后的规则用于现代地形条件下的最优路线重构，同时兼顾烽火台与关隘的协同选址。研究希望构建一套地形约束下路径优化的定量方法，为秦直道数字化保护以及同类山地线路规划提供理论依据和实践参考。

一、研究数据与研究方法

陕甘八县子午岭片区留存大量秦直道遗存，是探究地貌格局对古道布线约束的理想研究区[3]。为实现古道路数字化路径重构，本研究整合 DEM 高程栅格、秦直道历史矢量路线、分水岭、河网水系、烽火台与关隘遗迹多源空间数据，完成数据清洗、坐标系统一等预处理工作。综合运用 Sobel 地形梯度计算、KDTree 空间近邻检索、熵权赋权、统计显著性检验、蒙特卡洛模拟、最小成本路径算法，依次完成地形量化解析、选线原则校验、最优路径生成、防御设施选址整套研究流程。

（一）符号说明

符号	意义
Z_p	采样点的高程值
S	局部坡度
K_p	剖面曲率
K_{p1}	平面曲率
SI	路线顺直性指数
C	综合地形代价函数值
W_i	第 i 个地形因子的权重系数
L_{hist}	历史秦直道路线总长度
L_{opt}	优化后秦直道路线总长度

二、秦直道沿线地形特征量化分析

古代道路走向并非完全人为主观决定，很大程度上会被区域地形条件所限制。为客观揭示秦直道对周边地貌环境的适应模式，不能仅依靠肉眼观察图纸或者野外记录，需要借助栅格地形模型开展定量化测算。

本文基于经过预处理的 DEM 数据集构建地形特征参数模型[4]，定量测算坡度、坡向、剖面曲率、平面曲率、地形起伏度、地表粗糙度以及各类空间距离指标。依托矢量 – 栅格坐标映射方法以及 KD-Tree 近邻检索算法，完成全线采样点位与关键遗迹点位的地形信息批量提取，结合统计表格、空间可视化图表解析沿线地貌空间格局，为后面挖掘、验证秦直道选线内在规律提供标准化的数据集支撑。

（一）地形特征参数模型的建立与原理

1.DEM 数据预处理

对研究区 DEM 栅格开展无效像元识别与剔除，过滤掉研究区边界外以及数据缺损的栅格单元，只保留研究范围内有效的高程数据，保障后续各项地形参数计算结果不会被异常值干扰。本研究采用 Sobel 邻域差分算子求解地表高程梯度，在此基础上进一步换算得到坡度、坡向；通过高程二阶偏导数计算剖面曲率与平面曲率，分别刻画坡面纵向起伏形态和地表横向汇散特征；借助 3×3 滑动窗口统计窗口内部高程极差得到地形起伏度，并基于坡度完成地表粗糙度换算[5]。

在得到全部栅格地形图层之后，利用 KD-Tree 近邻检索算法快速计算每一个路线采样点到分水岭、河网水系的欧氏距离，将地形参数与距离参数组合，最终组装得到包含 10 个维度的地形特征向量，实现对古道全线地形环境的数字化刻画。

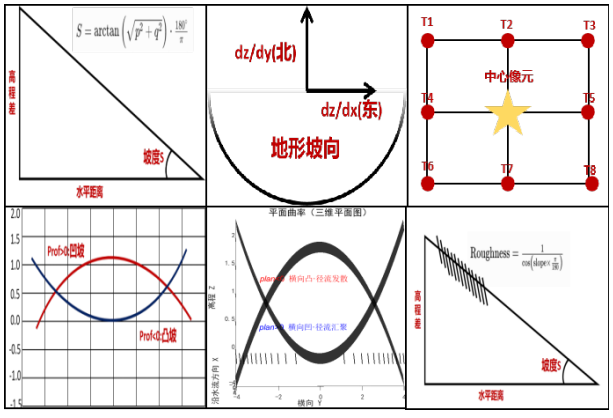


图 1：DEM 地形因子计算原理示意图

2. 核心地形指标计算公式

坡度反映地表单元倾斜程度，是古代山地道路选址时首要考量的地形指标，坡度过大会大幅提升施工难度，同时不利于部队行军与物资车马通行。

$$S = \arctan\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right) \cdot \frac{180^\circ}{\pi} \qquad (1)$$

地形起伏度表征局部窗口内部地势高差[5]，能够直观反映区域

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

地表破碎程度，起伏度越高代表地表沟壑越多，道路修建通行条件越差。

$$relief = \underset{3\times3}{max}(Z) - \underset{3\times3}{min}(Z) \quad (2)$$

（二）模型结果分析与可视化表达

完成 DEM 数据预处理与地形参数提取后，本研究结合历史文献、已有考古研究成果，借助可视化软件开展地貌特征挖掘。像《史记·秦始皇本纪》就提到秦直道修建时追求线形顺直、坡度可控，史念海先生在《秦直道遗迹的探索》中也注意到了这条古道常依托分水岭、同时避开河网沟谷的布线习惯；陕西考古研究院对遗址的调查则进一步说明，选线并不只看地形本身，高程是否平顺、遗迹之间能否协同布局，同样是实际需要面对的约束。这些来自不同角度的认识，为后续选线规则的提取和验证提供了基础。

在文献史料的基础上，借助 DEM 输出的多组可视化成果开展地形特征观察，图件包含研究区秦直道历史路线分布图、河网水系分布图、秦直道高程纵断面曲线、研究区坡度空间栅格图。一方面通过高程纵断面、坡度沿程变化曲线，解析整条古道行进过程中的高程起伏与坡度波动；另一方面通过全域空间栅格图，直观呈现古道与山地、水系、陡坡等地貌单元的空间位置关系。

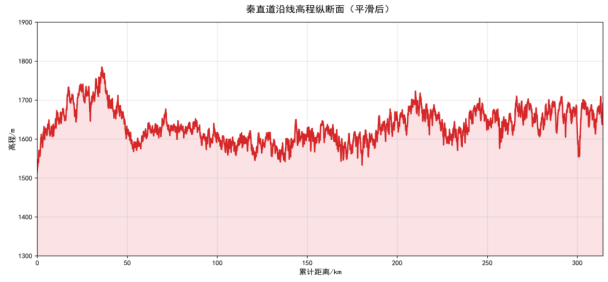


图2：秦直道高程纵断面

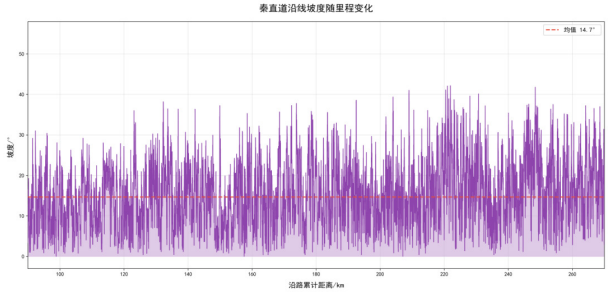


图3：秦直道坡度沿程变化图

结合剖面图表可以观察到，秦直道主线高程整体波动幅度有限，大部分通行路段坡度条件良好，仅穿越沟谷的局部区段出现短距离陡坡。

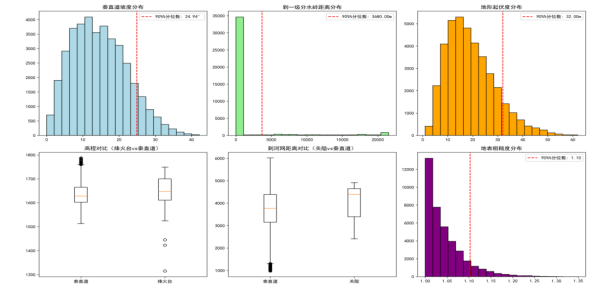


图4：全特征可视化

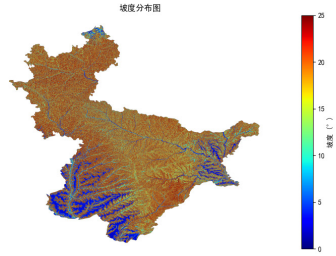


图5：坡度分布图

从全域空间格局来看，秦直道主线整体偏向山脊分水岭地带布设，主动避开大面积陡坡与河网洼地。综合文献记载与 DEM 可视化呈现的地貌现象，初步归纳出坡度控制、河网规避、分水岭依托、高程集中平顺、线形顺直、遗迹协同布局六大候选选线约束原则。

为进一步具象化地展示古道、烽火台、关隘等点位的地形条件，提高相关数据进行分析，以下是部分数据：

表1：10点地形特征采样结果（部分示例）

序号	类型	高程 (m)	坡度 (°)	到河网距离 (m)
1	秦直道	1617	29.01	4230.57
2	秦直道	1679	12.47	2153.89
3	秦直道	1593	35.02	3022.82
4	秦直道	1553	38.2	4277.97
5	秦直道	1571	11.26	4113.01
6	秦直道	1754	5.58	4376.69
7	烽火台	1702	3.59	3469.94
8	烽火台	1629	2.5	4486.1
9	关隘	1742	11.45	2404.4
10	关隘	1656	11.57	4912.9

采样统计结果进一步印证可视化图件反映出的地貌特征，上述六大候选选线约束将进入下一环节，开展各约束变量有效性的定量检验，为后续路径规划模型构建提供评价依据。

三、秦直道选线原则挖掘与定量验证

过去对秦直道选线规律的认识，更多是依靠考古调查和历史文献里那些定性的说法，缺乏统计学层面的量化检验，难以区分其选线特征是古人主动规划的结果，还是地形随机分布带来的偶然结果。

基于前文构建的全线多维地形数据集，再结合已有的考古研究成果，大致归纳出秦直道的六条选线准则。通过单因子统计建模，并对多个指标进行标准化处理，消除不同量纲的影响，在这个基础上构建路线规划契合度（RPI）综合评价模型，然后用 KS 检验和蒙特卡洛模拟来做统计验证，目的是把人为选址偏好和自然地形随机效应区分开来，进而还原这条古代军事通道背后完整的选址决策逻辑。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（一）路线规划原则与理论建模

1. 坡度控制原则（P1）：沿线坡度维持适中水平，主动避开长距离大陡坡，保障军队行军、车马物资转运的通行条件，同时降低开山修筑道路的工程成本。

2. 河网规避原则（P2）：路线主体与水系河道保持足够安全距离，减少汛期洪水冲刷、河道泛滥损毁路基的风险，降低跨河修建桥梁的工程投入。

3. 分水岭依托原则（P3）：线路尽量沿着山脊分水岭走向进行布设，山脊高地既能够获得开阔瞭望视野，又拥有相对稳定的路基条件。

4. 高程集中与平顺原则（P4）：路线高程不宜发生频繁剧烈升降，尽量减少大填大挖，降低工程开挖填筑的人力物力消耗。

5. 线形顺直与走向原则（P5）：线路整体以南北向作为主干走向，减少不必要的迂回绕行，契合秦直道作为快速军事通道的功能定位。

6. 遗迹协同布局原则（P6）：烽火台、关隘这类防御设施要和主线道路在空间上相互匹配，共同完成防御、预警、通讯整套军事功能。

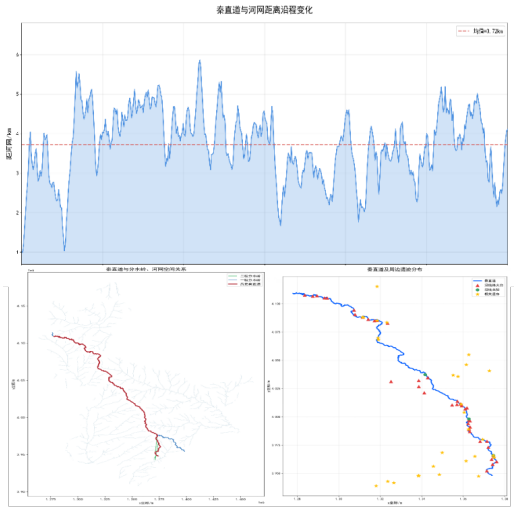


图6：秦直道沿程与河网距离变化及空间分布

各个原始评价指标量纲、数值区间各不相同，因此需要经过极差标准化消除量纲带来的干扰，之后采用熵权法客观确定各项选线原则的权重，得到路线规划契合度 RPI 综合评价公式。RPI 指数越高，代表该栅格单元越符合秦直道古代选线综合要求。

$$RPI_i = \sum_{j=1}^{6} w_j \cdot f'_{ij} \tag{3}$$

式中：w_j 为熵权法计算得到的第 j 项选线原则权重；f'_{ij} 为经过标准化处理之后的评价指标。

（二）模型算法流程与求解实现

完整计算流程分为四个步骤：第一步完成多源空间数据导入以及全部评价指标的标准化处理；第二步针对六大选线原则逐项开展定量测算；第三步依托熵权法求取各项指标权重，计算得到每一个栅格单元的 RPI 指数；第四步开展权重阈值扰动的敏感性测试，评判模型输出结果的稳定程度，同时根据 RPI 数值高低划分地形道路适配等级。

（三）模型结果与图表对应分析

经过模型测算可以得到一系列统计特征：秦直道沿线平均坡度14.66°，长距离陡坡片段占比很低；路线距离河网平均3.72km，基本不存在紧贴河道延伸的路段；到一级分水岭平均距离1470.60m；高程主要集中分布于1513–1787m 区间；线路整体呈南北走向，迂回程度有限；烽火台、关隘大多选址在地势高地，与主线道路空间匹配度较高。上述多项统计结果从定量层面证实六大选线原则客观存在，并非单纯来自文献主观推断。

四、基于综合代价栅格与 MCP 算法的路径重构

历史遗存秦直道受千年地貌变迁、人为损毁、遗迹断续等因素影响，现存线路存在局部迂回、断点缺失等问题，无法完全还原古代最优规划线路。本章将前文验证的六大选线原则转化为量化代价指标，构建综合地形代价模型，采用 MCP 最小成本路径算法实现现代地形条件下的最优路径重构 [6]，还原古人理想选线方案。

（一）构建模型思路

以秦直道军事通行、低耗建设、稳定存续的核心需求为导向，将坡度、水系、山脊、高程、地形平顺、遗迹适配六大约束转化为分级代价栅格，通过加权融合构建全域综合代价图层，搜寻累计代价最低的最优通行路径。

（二）综合地形代价栅格构建

综合地形代价由六项单因子代价加权构成，公式如下：

$$C_{total} = w_1 C_{slope} + w_2 C_{river} + w_3 C_{ridge} + w_4 C_{site} + w_5 C_{elev} + w_6 C_{terrain} \tag{4}$$

代入熵权计算所得权重：

$$C_{total} = 0.12 C_{slope} + 0.10 C_{river} + 0.30 C_{ridge} + 0.10 C_{site} + 0.16 C_{elev} + 0.22 C_{terrain} \tag{5}$$

由权重结果可知，山脊依托因子权重最高，是秦直道选线的核心控制条件，其次为地形平顺与高程约束，符合古代山地通道的建设逻辑。

（三）最小成本路径规划模型

以全域累计地形代价最小为目标 [7,8]，求解最优路径：

$$P^* = \arg\min_{P} \sum_{(i,j)\in P} C_{total}(i,j) \tag{6}$$

（四）路线评价指标

采用路线总长与弯曲系数两大指标量化评价线路优劣。路线总长公式：

$$L = \sum_{k=1}^{M-1} \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2} \tag{7}$$

弯曲系数用于表征线路顺直程度，数值越接近 1 线路越优：

$$SI = \frac{L}{\sqrt{(x_M - x_1)^2 + (y_M - y_1)^2}} \tag{8}$$

（五）新旧路线对比

重构路线与历史路线对比结果，重构线路总长更短、弯曲系

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

数更低，整体更加顺直高效，更契合古代军事快速通行需求。

表2：历史线路与新规划线路数据对比图

路线	总长度 /km	弯曲系数	点数
历史路线	301.07	1.6758	1800
新规划路线	251.08	1.339	2715

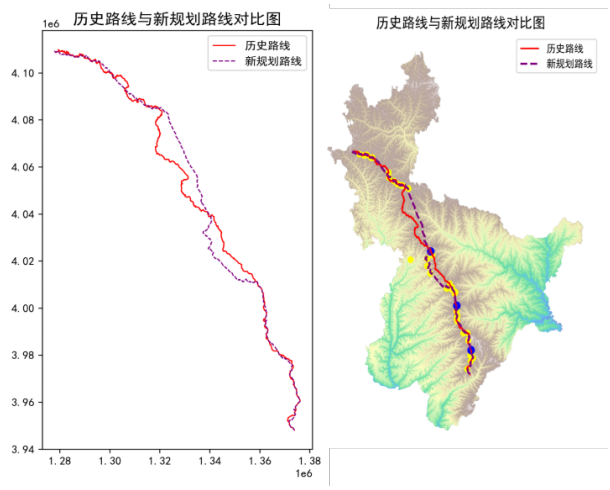


图7：秦直道历史路线与新规划路线空间对比

五、烽火台 – 关隘防御设施协同选址研究

秦直道并非单一交通线路，而是集通行、预警、防御于一体的复合型军事体系，烽火台与关隘的空间布局直接决定整体防御效能。本文依托重构最优路线，结合地形约束与古代军事布设经验，开展防御设施协同选址。

（一）选址约束条件

参考考古发现和同类古驿道的布设经验，约束条件并不是简单地按统一标准来设。烽火台和关隘的选址都限定在主线两侧300米以内，这样与主干道的联动才有保障。高程方面，百分位不低于70%，意味着要优先挑高地，瞭望需求才能满足。地面坡度控制在20° 以下，这主要是考虑到当时的修筑技术条件。至于间距，烽火台之间不小于5000米，关隘之间则不小于15000米，目的是让全线都能覆盖到，避免出现预警盲区。

（二）模型结果与分析

模型最终输出3处关隘、22座烽火台，点位整体位于山脊高地、地形平缓、通视条件优异，平均布设间距合理，通视检验覆盖率100%，能够完整实现古代分段预警、层层防御的军事功能。图6为重构路线与防御设施整体成果对比图，完整展示本研究最终优化成果。

六、模型有效性检验

为确保模型结果科学可信，本文从地形差异性、权重稳健性、路径显著性、设施布局合理性四个维度开展统计检验，排除算法随机误差与主观偏差。

（一）地形特征参数显著性检验

采用独立样本 t 检验对比古道点位与全域随机地形样本[9]，公式如下：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (9)$$

在0.05显著性水平下，坡度、高程、水系距离、山脊距离均显著拒绝原假设，证明秦直道选址存在明确地形偏好，并非随机分布。

（二）选线原则权重稳健性检验

对六大指标权重进行 ±10% 扰动测试，RPI 指数波动小于5%，无结果反转现象，模型稳健性良好；单因素方差分析显示各原则权重差异显著，赋权结果科学合理。

（三）重构路径蒙特卡洛模拟检验

通过随机路径对照模拟，重构路径 Z 统计量大于3，显著优于随机线路，证明路径重构结果具备统计可信度。

（四）防御设施布局统计检验

设施点位与随机点位对比检验结果显著，说明烽火台、关隘选址严格受高地、平缓地形约束，符合古代军事选址逻辑。

综合多维度检验可知，本文模型体系稳定可靠，选线规律、重构路径、设施布局结果均具备科学依据。

七、结果讨论

（一）主要研究发现

本文实现了秦直道传统定性选线思想的定量化、模型化表达，成功揭示古道六大核心选线规律，完成最优路径重构与防御设施协同选址，构建了一套可迁移、可复现的古道路数字化研究范式。

（二）研究存在局限

研究采用现代 DEM 数据推演，未充分考虑千年地貌变迁；仅采用熵权单一赋权方式，缺少主观经验修正；模型侧重地形约束，尚未充分纳入人文、军事、工程技术等多维因素[10]。

（三）后续研究展望

后续可结合古地貌数据还原秦代地形环境，采用主客观组合赋权，引入人文与工程约束，构建多维度综合选线模型，进一步完善古道路复原研究体系。

八、结论

本文以陕甘子午岭段秦直道为研究对象，基于 DEM 地形数据与考古遗迹资料，同时也参考了文献记载和地理建模方法，系统

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

开展了地形量化分析、选线原则验证、最优路径重构及防御设施布局等工作。研究发现秦直道沿线确实存在六条比较稳定的选线约束规律，并据此构建 $C_{total} = w_1C_{slope} + w_2C_{river} + w_3C_{ridge} + w_4C_{site} + w_5C_{elev} + w_6C_{terrain}$ 的最优路径方案。重构出来的路线跟历史遗存线路比起来，显得更顺直、更高效，和地形条件的匹配也比较好。防御设施的布局结果，也能较好地满足古代军事预警与防御方面的需求。这样一来，传统上偏定性的历史认知就被转成了可量化的模型成果，对秦直道遗产保护、古道路复原以及山地路径优化来说，都有一定的参考价值。

参考文献

[1] 史念海 . 秦始皇直道遗迹的探索 [J]. 文物 ,1975 (10):2837.

[2] 张健 , 王浩 . 基于 DEM 的秦直道沿线地形量化与空间分布分析 [J]. 地理科学进展 ,2022,41 (7):11451156.

[3] 孙洁 , 李婧 . 秦直道子午岭段地理环境与选线特征研究 [J]. 干旱区地理 ,2023,46 (5):891899.

[4]Wood J.The Geomorphological Characterization of Digital Elevation Models[J]. Geomorphology,1996,16 (1):1935.

[5] 王玲 , 吕新 . 基于 DEM 的新疆地势起伏度分析 [J]. 测绘科学 ,2009,34 (1):113116. [6] 刘学锋 , 孟令奎 . 基于栅格数据最优路径的公路选线分析 [J]. 地理空间信息 ,2011,9 (3):4547.

[7] 鲁敏 , 张金芳 . 栅格地形的最优路径分析 [J]. 武汉大学学报 (信息科学版),2010,35 (1):5963.

[8]Fisher T. LeastCost Path Modelling for Archaeological Landscape Reconstruction [J]. Journal of Archaeological Science,2011,38 (6):12561265.

[9] 姜启源 , 谢金星 . 数学模型 (第 5 版)[M]. 北京：高等教育出版社 ,2018.

[10] 王金铜 , 佳鑫 . 基于改进 A * 算法的施工道路智能选线方法研究 [J]. 土木建筑工程信息技术 ,2025,17 (3):4551.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

基于机器学习与 TPB 理论的青年非遗疗愈消费意愿及影响机制研究

高坚国，朱金桂，詹富婕，王佩君，欧阳睿，周大镯，林培琪
惠州学院 数学与统计学院，广东 惠州 516007

DOI:10.61369/ASDS.2026100005

摘　　要： 在青年情绪调节需求增长与非物质文化遗产活态传承的双重背景下，非遗手作、传统技艺和文化空间正逐渐成为兼具情绪价值与文化价值的新型体验消费载体。如何科学评估非遗疗愈的市场潜力、精准识别消费群体的异质性特征、系统揭示消费意愿的形成机制，是推动非遗活态传承与疗愈经济融合发展的关键问题。本研究以珠三角地区 15—44 岁青年群体为研究对象，基于 457 份有效问卷，综合运用描述性统计、K–means++ 聚类、随机森林、XGBoost 及结构方程模型，系统考察青年群体对非遗疗愈体验的市场认知、消费偏好、群体差异与意愿形成机制。研究发现，珠三角青年对非遗疗愈具有较为广泛的认知和参与基础，55.7% 的受访者知晓相关概念，71.2% 有过非遗体验经历，54.0% 愿意尝试非遗疗愈；消费偏好集中表现为“短时长、轻手工、低门槛、高性价比”，宣传触达不足、价格偏高和时间限制是主要障碍。K–means++ 聚类识别出中高消费家庭体验型、白领中消费高压型和青年低消费入门型三类细分群体，聚类后随机森林与 XGBoost 的预测效果均有所提升，表明市场细分能够增强购买意愿识别的准确性。结构方程模型（$\chi^2/df=1.016$，CFI=0.998，RMSEA=0.006）显示，情绪价值、文化价值和社会价值显著提升消费态度，态度、知觉行为控制与情绪价值进一步显著影响体验意愿，其中态度的直接作用最强（$\beta=0.525$）。基于此，本文提出构建分层产品矩阵、实施差异化运营、强化场景化传播并完善多主体协同机制。需要强调的是，本文所称“疗愈”主要指日常性的压力舒缓和情绪调适，不替代专业心理诊疗。

关 键 词： 非遗疗愈；消费意愿；扩展 TPB 模型；机器学习

Young Consumers' Intention toward Intangible Cultural Heritage-Based Wellness Consumption and Its Influencing Mechanisms: A Study Based on Machine Learning and the Theory of Planned Behavior

Gao Jianguo, Zhu Jingui, Zhan Fujie, Wang Peijun, Ouyang Rui, Zhou Dazhuo, Lin Peiqi
School of Mathematics and Statistics, Huizhou University, Huizhou, Guangdong 516007

Abstract: Against the dual background of rising emotional–regulation needs among young people and the living transmission of intangible cultural heritage (ICH), traditional crafts and cultural experiences are increasingly becoming experiential consumption carriers that combine emotional and cultural value. How to scientifically assess the market potential of ICH healing, accurately identify consumer heterogeneity, and systematically reveal the formation mechanism of consumption intention are key issues for promoting the integrated development of ICH living transmission and the healing economy. This study focuses on young people aged 15–44 in the Pearl River Delta, and based on 457 valid questionnaires, employs descriptive statistics, K–means++ clustering, random forest, XGBoost, and structural equation modeling to systematically examine market awareness, consumption preferences, consumer heterogeneity, and the formation mechanism of participation intention. The results show that 55.7% of respondents were aware of ICH healing, 71.2% had participated in ICH–related experiences, and 54.0% were willing to try such experiences. Preferences were characterized by short duration, light manual tasks, low entry barriers, and value for money, while insufficient publicity, high prices, and time constraints were the main obstacles. Three consumer segments were identified through K–means++

作者简介：
高坚国，惠州学院数学与统计学院，本科在读，研究方向为应用统计学；
朱金桂，惠州学院数学与统计学院，本科在读，研究方向为数学与应用数学；
詹富婕，惠州学院数学与统计学院，本科在读，研究方向为应用统计学；
王佩君，惠州学院数学与统计学院，本科在读，研究方向为数学与应用数学；
欧阳睿，惠州学院数学与统计学院，本科在读，研究方向为数据科学与大数据；
周大镯，惠州学院数学与统计学院，教授，硕士生导师，研究方向为数据挖掘；
林培琪，惠州学院数学与统计学院，辅导员，研究方向为思想政治。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

clustering: family-oriented high-spending experiencers, stressed white-collar consumers with medium spending, and price-sensitive young beginners. Prediction performance improved after clustering, indicating that market segmentation enhances the accuracy of purchase intention identification. The structural equation model (χ^2/df=1.016, CFI=0.998, RMSEA=0.006) indicates that emotional, cultural, and social value significantly enhance consumption attitudes, while attitudes, perceived behavioral control, and emotional value significantly influence participation intention, with attitudes exhibiting the strongest direct effect (β=0.525). Accordingly, this study proposes a tiered product portfolio, segment-specific operations, scenario-based communication, and a multi-stakeholder collaboration mechanism. "Healing" in this study refers to everyday stress relief and emotional adjustment rather than clinical treatment.

Keywords : intangible cultural heritage-based wellness; consumption intention; extended theory of planned behavior model; machine learning

引言

随着《健康中国2030规划纲要》深入实施与社会心理服务体系建设全面推进，国民心理健康需求日益凸显。珠三角地区18-34岁青年群体抑郁焦虑水平显著高于其他年龄段，对兼具文化深度与心理疗愈功能的体验式消费需求迫切。与此同时，非遗传承面临严峻代际断层危机（国家级非遗代表性传承人平均年龄63岁，40岁以下中青年传承人不足5%），亟需通过非遗+疗愈跨界融合实现双向赋能。基于SOR理论、正念疗法、艺术治疗、心流理论及感知价值理论，本文构建文化、心理、市场三维分析框架，选取珠三角15-44岁青年群体为调研对象，采用定性访谈与定量问卷相结合的混合研究方法，系统探究非遗疗愈市场潜力、青年消费偏好与意愿形成机制，旨在填补区域实证研究空白，为非遗场馆、文旅经营主体提供实证数据与经营决策参考。

一、文献综述

国内外学者已围绕非遗活态传承、艺术疗愈、文旅消费与青年心理干预开展大量研究，构建多类理论模型与实证分析框架，力求同步实现文化价值传递、文旅商户收益提升与青年消费者心理获益最大化。方思琳、杨立珩等通过对比多种艺术疗愈干预模式后证实，沉浸式实体手工创作可有效缓解青年焦虑、抑郁情绪，显著提升个体心理弹性[1]，是适配青年情绪调节的优质干预方式[2]。丁志峰等学者深入挖掘传统手工艺的情绪调节机制，针对剪纸、扎染、刺绣等岭南非遗构建三层递进疗愈模型，可根据青年压力状态动态调整体验内容，精准适配不同群体的情绪舒缓需求[3]。学者王浩阳、刘铁军（2022）则从具身认知理论视角切入，构建了信息循环与意识循环双重作用机制，为非遗疗愈提供理论支撑[4]。梁琨（2025）以戏剧疗愈为研究切入点，印证了传统文化元素在疗愈情境建构中的独特价值[5]。吴艳蓉（2024）立足信息时代背景，梳理了非遗教学与艺术疗愈融合的理论基础，并提出手工艺疗法、非遗故事疗法等五类落地实践方法[6]。学者杨硕（2023）深入研究手工陶艺的疗愈运行模式，提出适配陶艺非遗项目的三段式疗愈范式[7]。在兼顾文化认同感与消费价格感知的市场需求环境下，达维尔等将文化遗产作为心理健康干预媒介，诺兰证实古遗址具身体验的疗愈价值[8]。手工艺疗愈方面，TrujilloColmena等人（2025）经Meta分析证明具身性艺术干预可显著提升生活满意度，疗效因子为具身化、具象化与象征隐喻[9]。欧盟INT-ACT项目用扩展现实技术评估非遗体验对幸福感的影响，推动领域进入实证阶段[10]，但手工艺疗愈尤其是非遗技艺商业化转化的系统研究仍较薄弱。

现有研究不足主要有两点：一是针对珠三角这类非遗富集、青年密集、市场活跃的核心区域，缺乏系统的市场调研与需求画像；二是非遗手工艺疗愈研究薄弱，缺少理论整合，珠三角特色非遗技艺的疗愈机制与市场价值研究近乎空白。本文围绕市场潜力、人群细分、作用机制三大核心问题开展全链条实证分析。因此，本研究突破单一学科视角，转向市场调查与消费行为分析，以珠三角青年为对象开展实证调研，分析非遗疗愈认知、参与意愿、支付偏好及影响因素，将疗愈机制转化为市场潜力与需求预测，为非遗活态传承与情绪疗愈研究提供数据与决策参考，具有重要社会意义。

二、数据来源与研究方法

（一）数据来源与问卷设计

本研究数据包含问卷调查数据与网络文本数据两类，核心研究对象为珠三角地区15-44岁青年群体。

问卷调查采用配额抽样方法，按珠三角城市分布与青年群体类型设定样本配额，经抽样公式测算与有效率预留，共发放问卷493份，数据清洗剔除无效样本后，回收有效问卷457份，有效回收率为92.7%。样本覆盖广州、佛山等珠三角九市，包含在校大学生、

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

初入职场毕业生、高压职场白领、新中产育儿家庭四类核心群体，人口统计学特征与珠三角青年总体结构契合度较高，样本代表性良好。问卷设计基于SOR模型、计划行为理论（TPB）及感知价值理论，经过文献梳理、行业资料分析、前期访谈与预调查优化后形成正式量表，核心变量采用5点李克特计分，包含压力行为、认知态度、消费意愿、人口特征四大模块。样本基本特征如表1所示。

表1：样本基本特征

维度	类别	占比（%）
年龄	15—24岁	45.8
	25—34岁	28.3
	35—44岁	24.9
	其他	1.0
性别	男性	46.21
	女性	54.79
职业	高压职场白领	30.6
	初入职场毕业生	24.8
	大学生	23.8
	新中产育儿家庭	17.0

网络文本数据通过网络爬虫技术采集，覆盖抖音、B站、微博等主流平台的非遗手工相关高赞视频评论，经去重清洗后保留有效评论文本1.2万余条，用于挖掘公众对非遗体验的关注焦点与情感诉求，作为问卷数据的补充与交叉验证。

（二）数据质量检验

为保障实证分析的科学性与可靠性，本研究从信度、效度两方面对核心量表进行质量检验，核心结果如表2所示。

表2：量表信度与收敛效度检验结果

测量维度	题项数	Cronbach’s α	组合信度CR	平均方差提取量AVE
非遗疗愈认知与态度	9	0.889	0.892	0.568
非遗体验偏好	5	0.846	0.851	0.538
附加服务重视程度	3	0.801	0.805	0.518
总量表	17	0.938	–	–

结果显示，信度层面，各维度Cronbach’s α系数均高于0.8，组合信度CR均大于0.8，量表内部一致性良好。收敛效度方面，各维度AVE值均超过0.5的临界标准，收敛效度达标。结构效度上，验证性因子分析显示模型拟合优良（$\chi^2/df=1.780$，RMSEA=0.039，CFI=0.972）；区分效度检验显示各维度AVE平方根均大于维度间相关系数，维度区分度良好。

三、模型与方法

（一）K-means++聚类模型

为识别青年非遗疗愈消费群体的内部异质性，本文采用K-means++算法进行人群细分。该算法在初始聚类中心的选取中引入距离加权机制，可降低传统K-means对初始中心敏感的问题，并提高聚类结果的稳定性。

结合研究目的，实际参与聚类的变量包括月度解压支出、非遗认知水平、压力感知频率和价格接受度。职业属于名义尺度变量，直接纳入欧氏距离计算可能产生人为次序，因此仅用于聚类完成后的群体命名与画像解释。

设样本集合为$D=\{x_1,x_2,\cdots,x_n\}$，聚类数为K，第k个簇及其中心分别记为C_k和μ_k。模型通过最小化簇内平方误差和（SSE）获得类内紧凑的划分结果：

$$SSE=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2 \quad (1)$$

SSE越小，表明同一簇内样本差异越小。为避免仅依据SSE下降趋势主观确定聚类数，进一步采用Calinski-Harabasz（CH）指数评价不同K值下的类间分离度与类内紧凑度：

$$CH(k)=\frac{tr(B_k)/(k-1)}{tr(W_k)/(n-k)}=\frac{tr(B_k)(n-k)}{tr(W_k)(k-1)} \quad (2)$$

式中，B_k和W_k分别为类间离差矩阵与类内离差矩阵，n为样本量。CH值越大，说明聚类结构越清晰，据此综合确定最优聚类数。

（二）基于决策树的集成学习模型

为预测青年群体的非遗疗愈消费意向，并识别影响购买意向的关键因素，本文分别构建随机森林回归和XGBoost回归模型。两者均以决策树为基础，但在树的生成方式、误差修正机制及复杂度控制方面存在差异，可形成相互补充的预测结果。

1. 随机森林回归

随机森林回归（Random Forest Regression，RF）通过Bootstrap有放回抽样生成多个训练子集，并在每棵树的节点划分中随机抽取部分特征进行竞争，从而构建相互差异的回归树。样本随机化与特征随机化能够降低树之间的相关性，减弱单棵决策树的过拟合风险。

设共构建B棵回归树，第b棵树对样本x的预测结果为$T_b(x)$，随机森林的最终预测值为各树结果的平均：

$$\hat{y}(x)=\frac{1}{B}\sum_{b=1}^{B}T_b(x) \quad (3)$$

该平均集成方式能够降低预测方差，提高模型的稳健性和泛化能力。同时，可依据节点纯度下降或置换重要性衡量各解释变量对消费意向预测的相对贡献。

2.XGBoost回归

XGBoost（Extreme Gradient Boosting）采用串行梯度提升思想，通过逐轮加入回归树修正前一轮模型的预测误差。与随机森林的并行平均不同，XGBoost中新加入的树重点拟合已有模型尚未解释的残差，因此具有较强的非线性拟合与变量交互识别能力。

设模型包含K棵回归树，第k棵树为f_k，则样本x_i的预测值可表示为：

$$\hat{y}_i=\sum_{k=1}^{K}f_k(x_i),f_k\in F \quad (4)$$

为兼顾预测精度与模型复杂度，XGBoost将损失函数与正则化项共同纳入目标函数：

$$Obj=\sum_{i=1}^{n}l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k) \quad (5)$$

其中，$l(\ )$衡量真实值与预测值之间的误差，$\Omega(f_k)$用于惩罚

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

树结构的复杂度。常用的正则化形式为：

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \qquad (6)$$

式中，T 为叶节点数量，w_j 为叶节点权重，γ 和 λ 为正则化参数。模型在节点分裂时进一步依据目标函数增益选择最优切分点：

$$Gain=\frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda}+\frac{G_R^2}{H_R+\lambda}-\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right]-\gamma \qquad (7)$$

增益越大，说明该划分带来的误差下降越明显。由此，XGBoost 可在提升拟合能力的同时控制模型复杂度，适用于解释变量关系非线性、交互效应较强的消费意向预测任务。

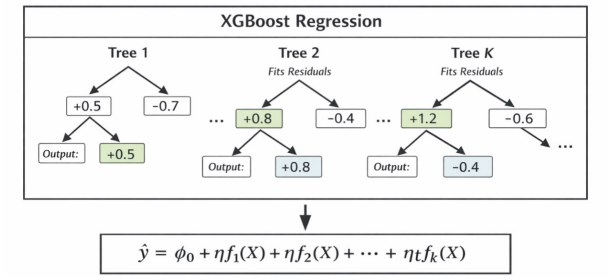


图1：XGBoost 原理图

（三）TPB 结构方程模型

为进一步解释非遗疗愈消费意向的形成机制，本文构建计划行为理论（Theory of Planned Behavior，TPB）与感知价值理论的整合结构方程模型。TPB 认为，行为意向主要受到态度（ATT）、主观规范（SN）和知觉行为控制（PBC）的共同影响；感知价值则划分为功能价值（FV）、情绪价值（EV）、社会价值（SV）和文化价值（CV）四个维度，用于刻画青年群体对非遗疗愈体验的综合评价。

在 SOR 理论逻辑下，产品信息、价格及体验设计等构成外部刺激（S），感知价值、态度、主观规范和知觉行为控制反映个体内部状态（O），消费意向则作为行为反应（R）。该整合框架既可检验各潜变量对消费意向的直接影响，也可分析感知价值通过 TPB 变量产生的间接作用。

实证分析依次开展描述性统计、量表信度与效度检验、测量模型检验及结构路径估计，并结合模型拟合指标和标准化路径系数判断理论假设是否成立。具体变量关系如图 2 所示。

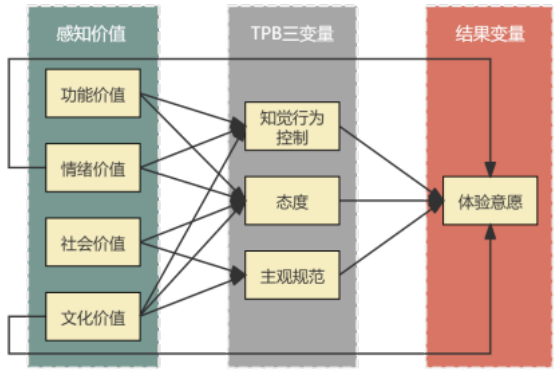


图2：理论模型图（TPB+ 感知价值整合框架）

四、数据探索性分析

（一）情绪疗愈市场背景与青年需求特征

随着社会生活节奏加快，青年群体的情绪调节需求不断上升。据中国科学院心理研究所出版《中国国民心理健康发展报告（2023-2024）》[11] 显示，抑郁与焦虑风险总体随年龄增长而下降，年轻群体的风险水平相对更高；同时，2023—2025 年短视频平台中“疗愈 + 美学”，“疗愈 + 旅修”，“疗愈 + 课程”等细分话题快速增长，表明疗愈需求正由泛化情绪安慰转向具体生活场景。

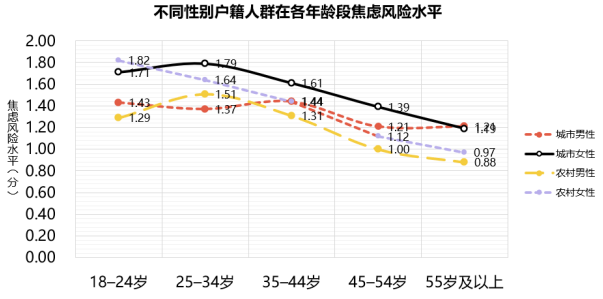


图3：不同性别户籍人群在各年龄段焦虑风险水平

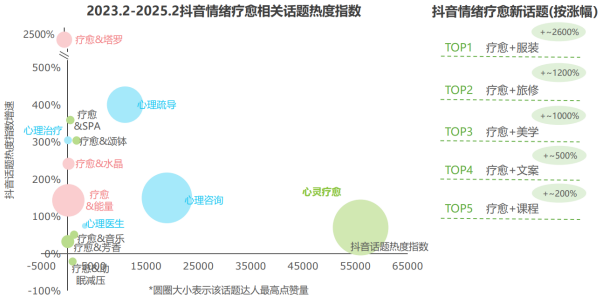


图4：2023—2025 年抖音情绪疗愈相关话题热度

（注：数据来源：巨量算数，统计时间范围：2023.2.1-2025.2.1，指数增幅对比分别为 2023.2.1-2024.2.1 与 2024.2.1-2025.2.1 两个时间段话题热度指数平均数对比 [12]）

本研究共获得 457 份调查样本。工作或学习压力是青年最主要的压力来源，其次为经济压力、家庭压力和自我成长焦虑。情绪问题出现后，动手制作类活动的选择人数最多，说明手工体验具有较好的青年接受基础。

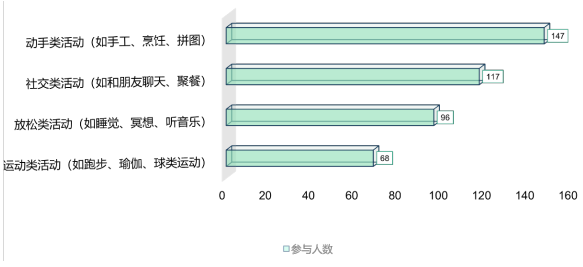


图5：青少年解决情绪问题方法调查

在信息获取方面，短视频平台和线下展览、文创市集是最主要的认知渠道，社交媒体、熟人推荐及学校或单位活动发挥补充作用。因此，非遗疗愈产品可采用“线上内容触达—线下体验转化—社交分享扩散”的推广路径。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

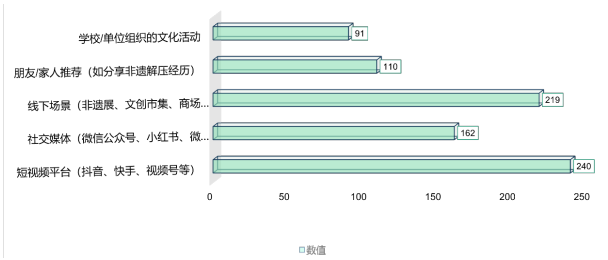


图6：用户对疗愈手段渠道

调查显示，青年对非遗已有一定认知基础：39.5% 的受访者将自身非遗认知水平评为 4 分，55.7% 的受访者知晓“非遗疗愈”概念。在参与行为上，71.2% 的受访者曾参加非遗体验，其中 44.3% 明确感受到身心放松；未参与者中仍有较多潜在兴趣人群。

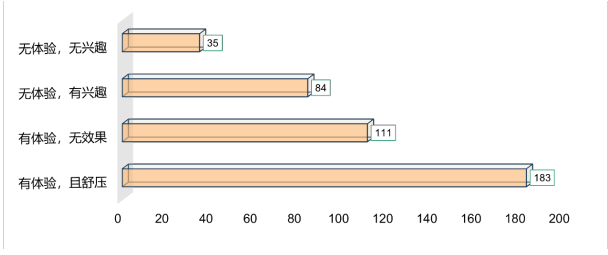


图 a：珠三角青年群体非遗参与现状

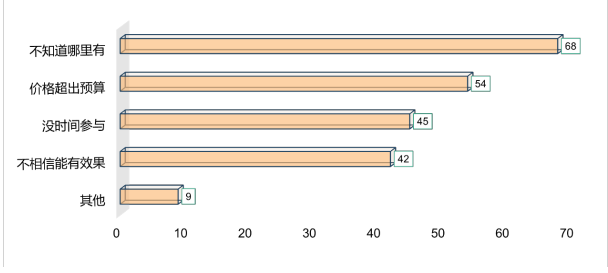


图 b：青年群体未参与非遗体验原因

图7：珠三角青年非遗疗愈体验市场认知现状与阻碍因素

宣传触达不足、价格偏高、时间受限和效果信任不足是影响参与的主要因素。总体来看，珠三角青年非遗疗愈市场已具备认知、参与和意愿基础，但仍处于由兴趣认同向实际消费转化的培育阶段。

（二）基于 K-means++ 的消费群体细分

为识别青年非遗疗愈消费需求的内部差异，本文采用 K-means++ 方法进行聚类。实际参与聚类的指标包括月度解压支出、非遗认知水平、压力感知频率和价格接受度，职业变量用于聚类后的群体解释与命名。各指标经 Z-score 标准化后，在 k=2-8 范围内比较 Calinski-Harabasz（CH）指数。

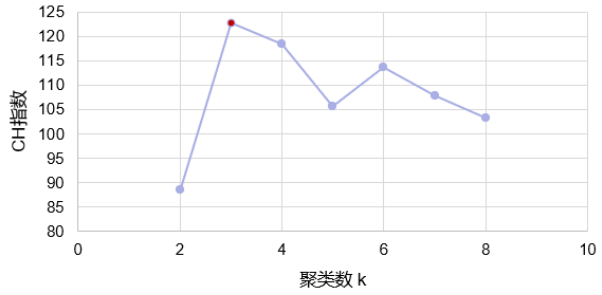


图8：不同聚类数下的 CH 值对比

结果显示，k=3 时 CH 指数最高，说明三类划分具有较好的簇内紧凑性和簇间差异。结合聚类中心、职业结构与消费特征，可将市场概括为中低频保守型群体，高消费核心群体，情绪敏感型群体。

表3：各聚类群体组内平方和

维度	全局均值	低频保守型群体	高消费核心群体	情绪敏感型群体
职业	3.11	2.75	4.40	2.44
每月可支配收入	3.01	2.62	3.71	2.76
购买频率	3.12	1.27	3.14	4.12
压力感知	3.14	2.96	3.24	3.17
溢价认同感	3.21	2.89	3.51	3.18

通过对最终聚类中心的分析可以看出，三个用户群体在人口学特征、认知成分和情感成分上均表现出较明显的差异特征。

五、非遗体验意愿预测与决策分析

本章首先运用集成学习模型识别不同细分群体的体验意愿。其次基于计划行为理论（TPB）与感知价值构建结构方程模型，进一步解释体验意愿的形成路径。

（一）基于集成学习的体验意愿预测

在 K-means++ 群体细分基础上，分别采用 XGBoost 与随机森林对整体样本及三类群体的体验意愿进行预测，并以决定系数（R^2）和均方误差（MSE）评价模型表现。

表4：集成学习模型预测结果对比

模型	无聚类		青年低消费入门型		中高消费家庭体验型		白领中消费高压型	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
XGBoost	0.3247	0.8292	0.4316	0.7152	0.3884	0.7421	0.4112	0.7284
Random Forest	0.3913	0.7475	0.4184	0.7318	0.4467	0.7013	0.4295	0.7196

结果表明，相较于未分群模型，三类群体在两种算法下均表现为 R^2 提高、MSE 降低。其中，随机森林对中高消费家庭体验型的预测效果最佳（R^2=0.4467，MSE=0.7013）；XGBoost 对青年低消费入门型的提升较为明显。由此说明，群体细分能够增强样本内部同质性，使模型更准确地识别不同群体的体验意愿规律。

（二）基于 TPB 结构方程模型的非遗体验影响路径分析

为揭示非遗体验意愿的心理形成机制，本文将计划行为理论与感知价值相结合，以态度（ATT）、主观规范（SN）和知觉行为控制（PBC）预测体验意愿（BI），并从功能价值（FV）、情绪价值（EV）、社会价值（SV）和文化价值（CV）四个维度解释其前因作用。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

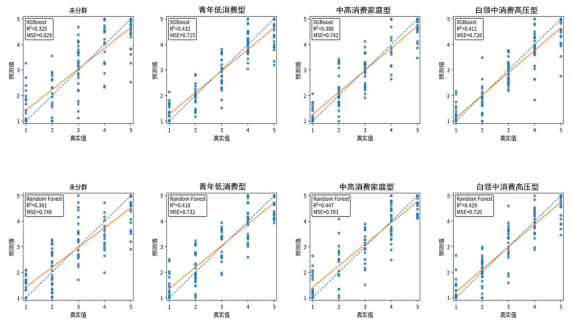


图9：群体划分前后预测效果

1. 模型拟合与路径检验

结构方程模型整体拟合情况较好。χ^2/df 为1.016，CFI 与 TLI 均为0.998，RMSEA 为0.006；除 AGFI 略低于0.90外，其余指标均达到或接近常用判别标准，表明模型能够较好地解释样本数据。

表5：结构方程模型拟合度指标（SEM）

指标	参考标准	数值
χ^2	p>0.05为良好	452.279
χ^2/df	<2为优秀，<3为良好	1.016
p 值	>0.05为良好，>0.10为优秀	0.396
CFI	>0.95为优秀，>0.90为良好	0.998
TLI	>0.95为优秀，>0.90为良好	0.998
RMSEA	<0.05为优秀，<0.08为良好	0.006
GFI	>0.95为优秀，>0.90为良好	0.908
AGFI	>0.95为优秀，>0.90为良好	0.897
NFI	>0.95为优秀，>0.90为良好	0.908

路径检验表明，态度（β=0.525）、知觉行为控制和情绪价值均显著促进体验意愿；情绪价值、文化价值和社会价值显著提升态度，社会价值增强主观规范，功能价值提高知觉行为控制，而情绪价值对知觉行为控制呈负向作用。

表6：路径关系假设检验结果表（β 为标准化系数）

假设	路径	β	p	显著性
H1a	EV → ATT	0.430	<0.001	显著
H1b	CV → ATT	0.351	<0.001	显著
H1c	SV → ATT	0.173	0.002	显著
H1d	FV → ATT	0.057	0.307	不显著
H2	SV → SN	0.630	<0.001	显著
H3a	FV → PBC	0.188	0.015	显著
H3b	EV → PBC	−0.165	0.032	显著（负向）
H3c	CV → PBC	−0.033	0.657	不显著
H4a	ATT → BI	0.525	<0.001	显著
H4b	SN → BI	0.098	0.053	不显著（边际）
H4c	PBC → BI	0.219	<0.001	显著
H4d	EV → BI	0.185	0.009	显著
H4e	CV → BI	−0.006	0.923	不显著

2. 影响机制与决策启示

综合来看，非遗体验意愿呈现态度主导、情绪直达、控制约束的作用机制。项目设计应突出情绪疗愈与文化审美价值，并通过清晰流程、分层定价和便捷渠道降低时间、价格与信息门槛。社会价值更适合作为口碑传播与社群触达的辅助工具。结合前述细分群体进一步实施差异化产品与传播策略，可提升体验意愿的识别与转化效率。

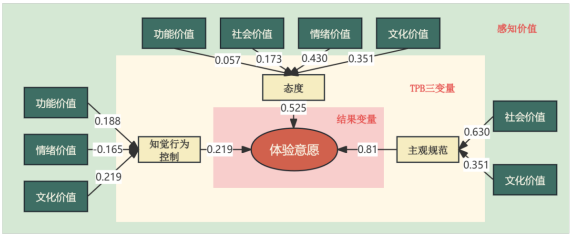


图10：非遗疗愈体验愿影响机制

六、结论与策略建议

（一）主要结论

第一，在市场认知维度方面，珠三角青年对非遗疗愈具有较为广泛的认知基础，但整体仍处于从兴趣认同向实际消费转化的培育阶段。55.7% 的受访者知晓“非遗疗愈”概念，71.2% 有过非遗体验经历，54.0% 愿意尝试非遗疗愈，核心消费群体集中在23–30岁职场青年。

第二，在消费行为维度方面，消费偏好集中呈现“短时长、轻手工、低门槛、高性价比”特征，超70% 的受访者接受1小时以内体验，60% 以上接受50元以下价格。宣传触达不足（31.2%）、价格偏高（24.8%）和时间限制（20.6%）是三大核心障碍。

第三，在市场细分维度方面，K–means++ 聚类（CH 指数在 k=3时取最优）识别出三类群体：①中高消费家庭体验型（新中产育儿家庭为主，月均解压消费301–500元占76.2%）；②白领中消费高压型（高压职场白领为主，月均解压消费101–300元占74.5%，压力频率每周1–2次占59.7%）；③青年低消费入门型（初入职场毕业生为主，月均解压消费100元以内占76.0%）。聚类后随机森林 R^2 从0.3913提升至最高0.4467，XGBoost 的 R^2 从0.3247提升至最高0.4316，证明细分能增强预测准确性。

第四，在作用机制维度：结构方程模型（χ^2/df=1.016，CFI=0.998，RMSEA=0.006）显示，态度对体验意愿影响最强（β=0.525），知觉行为控制（β=0.219）和情绪价值（β=0.185）同样显著促进意愿。情绪价值（β=0.430）和文化价值（β=0.351）是驱动态度的核心前因，呈“态度主导、情绪直达、控制约束”三重机制。

（二）发展策略建议

在产品策略方面，构建“引流—核心—定制”三级产品矩阵。引流型30–60分钟、50元以内，主打商圈快闪与校园活动；核心型1–2小时，融入文化讲解与传承人示范；定制型面向亲子家庭、企业团建，提供个性化方案。三级产品共享同一文化母题，形成从首次尝试到深度参与的升级路径。

在运营策略方面，中高消费家庭体验型突出亲子共创与家庭套票，提升客单价与推荐率；白领中消费高压型推行次卡、季卡，布局写字楼与商圈，聚焦复购率；青年低消费入门型降低首次成本，通过拼团、联名强化社交传播，关注获客成本。

 Copyright © This work is licensed under a Creative Commons Attribution–NonCommercial 4.0 International License

在传播策略方面，宣传应展示材料触感、手部操作与参与者真实反馈，减少不确定性，而非堆砌历史信息。“疗愈”表述采用“放松体验、压力舒缓、专注练习、文化陪伴”等审慎语言，避免夸大疗效，维护消费者合理预期。

在协同机制方面，建立政府引导、企业主导、传承人参与、高校支撑的多方联动机制。传承人负责工艺真实性，企业负责产品化运营，高校开展效果评估，政府完善行业规范。核心原则是以非遗为本体、以疗愈为功能，避免将非遗简单包装为疗愈噱头。

参考文献

[1] 方思琳 . “艺术疗愈”在高校大学生心理健康教育中的应用研究 [J]. 佳木斯职业学院学报 ,2024,40(03).

[2] 杨立珩 . 艺术疗愈理念下的美术课程价值重塑与路径探索 [J]. 美术教育研究 ,2026,(01).

[3] 丁志峰 , 木全熙 , 杨玉博 . 剪纸、扎染、刺绣等民间非遗技艺的体验消费与疗愈应用研究 [J]. 丝网印刷 ,2025,(23).

[4] 王浩阳 , 刘铁军 . 艺术疗愈视角下的情绪信息家具设计研究 [J]. 家具与室内装饰 ,2022,29(05).

[5] 梁琨 . 戏剧疗愈在大学心理健康教育中的应用：困境与突破 [J]. 扬州职业大学学报 ,2025,29(04).

[6] 吴艳蓉 . 信息时代基于非遗教学的艺术疗愈 [J]. 上海服饰 ,2025,(09).

[7] 杨硕 . 手工陶艺的艺术疗愈机制研究 [D]. 山东艺术学院 ,2023.

[8]Xie Y, Sun X. An overview of the integration and development trends between arts and psychotherapy[J].Frontiers in Psychology,2025,16:1617700.

[9]Fern á ndez-S á nchez J, Trujillo-Colmena D, Collado-Mateo D, Leon-Llamas J L, Lav í n-P é rez A M, Lorenzo-Gonz á lez M. Effects of art therapy interventions on satisfaction with life: A systematic review and meta-analysis[J].Advances in Integrative Medicine,2025,12(4):100532.

[10]N ú ñez Garc í a A, la Fuente Garc í a S, Lowemi E, Masoodian M, Vieira R, Rodrigues A, Luz S. A study protocol for assessing the effects of intangible cultural heritage experiences on human well-being[J].PLoS One,2025,20(11):e0336120.

[11] 孙向红 , 蒋毅 . 中国国民心理健康发展报告（2023-2024）[R]. 北京：社会科学文献出版社 ,2025.

[12] 艾瑞咨询研究院医疗组 . 中国情绪疗愈行业研究报告 [R]. 上海：艾瑞咨询 ,2025.

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

环境政策收紧与生产率动态——基于 OECD 国家的面板局部投影

莫济瑜[1]，张少谦[2]，蔡梓萱[1]，邝东阳[1*]
1. 中山大学 数学学院（珠海），广东 珠海 519082
2. 中山大学 软件工程学院，广东 珠海 519082
DOI:10.61369/ASDS.2026100006

摘　　要： 现有研究对环境政策与生产率关系尚未形成一致结论。受限于政策水平的静态估计以及对单个时期显著性的依赖，一般方法也较难同时识别政策调整的动态过程及其条件差异。本文以环境政策严格度的正向变化度量政策收紧，采用面板局部投影估计法，在控制国家与年份固定效应的基础上，考察政策收紧后0至5年内全要素生产率的累计响应。异质性分析从两个维度展开：其一为事前制度能力，其二为政策工具类型（市场型、非市场型及技术支持型）。为确保估计结果的稳健性，本文综合运用国家层面的聚类稳健标准误、受限野生聚类自助法、跨期联合显著性检验及多重假设检验校正，以系统评估证据强度。研究使用34个 OECD 国家2002—2020年的面板数据，其中环境政策、生产率、制度和宏观控制变量分别来自 OECD EPS 、Penn World Table 11.0 、World Bank WGI 和 World Development Indicators。结果表明，环境政策收紧后的平均生产率响应未形成稳定的动态路径：较高制度能力对应的点估计总体更趋向正向，不同政策工具也呈现一定的路径分化，但条件效应、跨期联合检验和多重推断均未确认稳定的制度或工具异质性。我们针对以上数据的研究说明，环境政策的生产率后果更适合从动态性、条件性和统计不确定性加以评价，政策实践中不宜依据单期估计或简单工具分类判断其生产率效应，而应同时关注政策实施环境、工具设计及动态调整过程。

关 键 词： 环境政策严格度；全要素生产率；局部投影；制度能力；政策工具；稳健推断

Environmental Policy Tightening and Productivity Dynamics: Panel Local Projection Evidence from OECD Countries

Mo Jiyu[1], Zhang Shaoqian[2], Cai Zixuan[1], Kuang Dongyang[1*]
1.School of Mathematics (Zhuhai), Sun Yat-sen University, Zhuhai, Guangdong 519082
2.School of Software Engineering, Sun Yat-sen University, Zhuhai, Guangdong 519082

Abstract: The existing literature has yet to reach a consensus on the relationship between environmental policy and productivity. Conventional approaches, constrained by static estimates of policy levels and reliance on statistical significance at individual horizons, often have difficulty identifying both the dynamic adjustment to policy changes and the conditions under which such effects may differ. This study measures environmental policy tightening as positive changes in environmental policy stringency and employs panel local projections, controlling for country and year fixed effects, to estimate the cumulative response of total factor productivity over horizons of zero to five years following a tightening. Heterogeneity is examined along two dimensions: ex ante institutional capacity and policy instrument type, including market-based, non-market-based, and technology-support instruments. To ensure the robustness of statistical inference, the analysis combines country-clustered robust standard errors, the restricted wild cluster bootstrap, cross-horizon joint significance tests, and multiple-testing corrections to systematically assess the strength of the evidence. The analysis uses panel data for 34 OECD countries over 2002–2020. Data on environmental policy, productivity, institutions, and macroeconomic controls are drawn from the OECD Environmental Policy Stringency database, Penn World Table 11.0, the World Bank's Worldwide Governance Indicators, and World Development Indicators, respectively. The results show no stable dynamic pattern in the average productivity response following environmental policy tightening. Point estimates tend to be more positive at higher

作者简介：
莫济瑜，中山大学数学学院（珠海），本科在读；
张少谦，中山大学软件工程学院，本科在读；
蔡梓萱，中山大学数学学院（珠海），本科在读。
通讯作者：邝东阳，中山大学数学学院（珠海），副教授，硕士生导师，研究方向为数据分析与挖掘，交叉科学中人工智能模型的开发应用，数学规范化及可解释。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

levels of institutional capacity, and the response paths also differ to some extent across policy instruments. However, neither the conditional effects nor the cross-horizon joint tests and multiple-testing adjustments provide statistically robust evidence of systematic heterogeneity by institutional capacity or policy instrument. Taken together, the evidence from this sample suggests that the productivity consequences of environmental policy are better assessed in terms of their dynamics, conditioning factors, and statistical uncertainty. Policy assessments should therefore avoid drawing conclusions from estimates at a single horizon or from simple classifications of policy instruments, and should instead consider the implementation context, instrument design, and the dynamic adjustment process jointly.

Keywords : **environmental policy stringency; total factor productivity; local projections; institutional capacity; policy instruments; robust inference**

引言

环境政策能否在改善环境质量的同时维持乃至提高生产效率，是环境经济学中一个长期存在而尚未获得统一答案的问题。传统观点强调环境规制的合规成本，更严格的排放标准、环境税费和技术要求可能占用资本、劳动与管理资源，从而在短期内降低生产率。Porter 和 van der Linde 则提出，设计合理且具有可预期性的环境规制可能通过诱导技术创新、流程改造和资源重新配置形成创新补偿，抵消部分甚至全部合规成本[1]；Palmer 等则指出，这种补偿并非无条件成立，其净效应仍需由经验事实判断[2]。此后围绕波特假说的研究逐渐从环境规制究竟促进还是抑制生产率，转向考察不同政策类型、时间尺度和制度环境下的异质性响应[3]。

环境政策的生产率后果可能具有明显的时间结构：政策收紧后，企业首先承担设备更新、污染治理和组织调整等成本，而技术采用、学习效应和创新回报通常需要更长时间才能实现。因此，同一项政策可能在不同年份产生方向不同的响应，仅考察政策严格度与当期生产率之间的平均关系，可能掩盖短期成本与中期适应的差异。国内研究也观察到类似复杂性：王杰、刘斌发现环境规制与企业全要素生产率并非简单线性关系[4]；黄庆华等指出环境规制与绿色全要素生产率之间具有明显时期依赖性[5]；刘伟江等进一步表明，环境规制与绿色全要素生产率、技术进步之间可能呈现非线性关系[6]。这些结果意味着，接近零的平均效应既可能反映政策影响有限，也可能来自不同调整阶段和经济主体反应的相互抵消。

近年来跨国研究进一步表明，环境政策的生产率效应具有明显异质性。Albrizio 等利用 OECD 环境政策严格度指数发现，环境政策收紧的生产率后果取决于企业和行业距离技术前沿的程度[7]；Koch 和 Themann 基于欧盟碳排放交易体系的企业证据同样发现，接近技术前沿的企业与落后企业可能呈现不同反应[8]；Benatti 等进一步用面板局部投影估计环境政策收紧后五年的动态响应，发现高污染企业受到的生产率负向冲击更明显，而且不同政策工具的影响存在差异[9]。因此，环境政策与生产率之间缺乏一致结论，更可能反映时间、制度、工具和主体能力等维度的异质性，而非研究结果彼此矛盾。

然而，现有研究仍存在三个值得进一步区分的问题。第一，已有研究虽越来越重视动态调整，但跨国宏观研究仍较多依赖环境政策严格度的水平值或固定滞后项。由于 OECD EPS 具有较强的时间持续性，政策水平所反映的长期国家差异与政策变化后的动态响应并非同一研究问题。第二，制度条件常被用于解释环境政策效果之间的差异，但制度能力既可能提高政策可信度、促进创新，也可能强化政策执行并增加短期合规成本，因此其调节方向并不确定。按照 North 的制度分析框架，制度通过塑造激励结构和交易成本影响经济主体行为[10]；相关研究也表明，治理结构和执行能力会通过政府约束、资源配置与创新行为影响绿色生产率[11]。第三，动态曲线上某一期的显著系数并不意味着整条响应路径得到统计确认；随着预测期和异质性维度增加，还需要同时考虑跨期联合推断与多重比较问题。

基于上述问题，本文考察环境政策收紧后的生产率动态调整，以及制度能力和政策工具是否形成可跨时期、跨推断方法确认的系统性差异。本文整合了多个权威数据源，构建34个 OECD 国家2002—2020年的面板数据。不同于直接使用高度持续的 EPS 水平，本文以环境政策严格度的正向变化刻画政策收紧，并采用面板局部投影估计政策发生后数年内的累计全要素生产率响应[12,13]。

本文的贡献主要体现在三个方面。第一，在研究对象上，从环境政策严格度的静态水平转向政策收紧及其动态响应，将短期合规成本与滞后适应过程置于统一时间框架。第二，在异质性分析上，不预设制度能力的调节方向，而将其作为政策作用的边界条件，并进一步区分市场型、非市场型和技术支持型三种工具。第三，在统计推断上，将国家层面聚类稳健推断、有限聚类条件下的受限野生聚类自助检验（Wild Cluster Restricted Bootstrap）、跨期联合检验和多重检验校正纳入统一证据体系，重点考察经济结论能否在不同合理推断方式下保持一致。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

一、理论机制与研究问题

（一）成本补偿与动态调整

环境政策收紧首先改变企业面临的技术约束和资源配置条件。从成本视角看，更严格的排放标准和环境要求会增加治理投资、资本替换和生产调整成本，因而可能在政策实施初期降低生产率[2]。另一方面，污染和能源使用成本上升也可能推动企业采用节能设备、开展清洁技术研发，通过技术创新和资源重配形成创新补偿[1,3]。

由于合规成本通常较早发生，而技术学习、资本更新和创新收益可能会滞后实现，生产率既可能呈现“短期下降、随后回升”，也可能持续为负，或因不同主体反应相互抵消而在宏观层面接近零。由此提出第一个研究问题：环境政策收紧以后，OECD国家的全要素生产率是否在0—5年内呈现稳定、可统计确认的动态响应?

（二）基于制度能力的边界条件

相同的环境政策在不同制度环境下可能对应不同的执行效果和企业调整行为。较强的制度能力一方面可提高政策连续性与可信度，促进长期投资和技术创新；另一方面也可能增强监管执行、减少规避空间，从而提高企业实际承担的短期合规成本[10,11]。

因此，制度能力对环境政策生产率效应的调节方向在理论上并不确定。为避免年度治理指标缓慢变化和测量误差干扰短期政策响应，本文采用政策期之前形成的事前制度能力。由此提出第二个研究问题：环境政策收紧后的动态生产率响应，是否会系统地随事前制度能力而变化，并形成能够通过跨期联合推断确认的制度异质性?

（三）政策工具差异

OECD EPS将环境政策划分为市场型、非市场型和技术支持型工具[14]。市场型政策主要通过价格信号改变污染和能源使用成本，非市场型政策通过标准和行政约束影响企业生产选择，技术支持政策则通过降低研发和技术采用成本发挥作用。由于三类工具在调整自由度和创新激励上存在差异，其动态生产率响应也可能不同[7-9]。由此提出第三个研究问题：不同类型环境政策收紧以后，是否呈现彼此不同且能够得到统计确认的动态生产率响应?

二、文献综述

（一）从约束与补偿之争到条件化波特假说

环境政策与生产率之间尚未形成稳定一致的经验结论。早期研究主要围绕合规成本与创新补偿展开，后续研究逐渐表明，环境规制的生产率效应并非简单的正向或负向关系，而会受到时间、主体特征和政策环境等因素影响[1-3]。王杰、刘斌发现环境规制与企业全要素生产率之间存在非线性关系[4]；Albrizio等表明政策影响会随企业或行业距离技术前沿的程度而变化[7]；黄庆华等和刘伟江等进一步从动态调整和技术进步角度说明环境规制影响具有明显条件性[5,6]。因此，相关研究的重点已逐渐由判断波特假说是否普遍成立，转向识别其动态过程和成立条件。

基于这一研究进展，本文将研究范围集中于政策收紧后的动态响应、事前制度能力以及政策工具差异三个方面。企业创新和污染暴露同样可能影响政策效果，但更适合利用微观数据识别，不属于本文34个OECD国家宏观面板能够充分回答的范围。这样的范围限定既回应已有文献中的主要争议，也避免在有限国家样本中引入过多异质性维度。

（二）从平均效应到能力环境工具

已有研究表明，仅考察平均政策效应可能掩盖不同经济主体和政策环境之间的重要差异。Koch和Themann发现，环境政策的生产率影响会随企业技术位置而变化[8]；范小敏和刘中强则表明，环境治理结构和执行能力能够影响政策约束向企业行为和生产率结果的传导[11]。相较于使用同期制度变化，本文采用政策分析期之前形成的制度能力，以更清晰地刻画各国既有政策执行和经济调整基础，并减少制度变化与政策响应之间的同期混合。

政策工具也是已有文献识别出的重要异质性来源。OECD更新后的EPS指标分别提供市场型、非市场型和技术支持型政策子指数[14]，Benatti等也发现不同环境政策工具可能对应不同的生产率调整过程[9]。因此，本文选择制度能力和政策工具作为两项主要异质性维度：前者考察政策实施环境是否改变生产率响应，后者考察不同政策设计是否产生不同动态结果。

（三）从静态相关到政策收紧的动态响应

OECD EPS具有较强的时间持续性，政策水平不仅反映当前环境规制强度，也可能包含收入水平、产业结构和治理体系等长期国家差异，因此EPS水平与生产率之间的静态关系并不能直接回答新的政策收紧发生后生产率如何调整。Albrizio等已将研究重点更多转向环境政策严格度的变化[7]，Benatti等进一步利用局部投影分析政策收紧后的动态生产率响应[9]。据此，本文以EPS正向变化刻画政策收紧，并将既有EPS水平作为控制条件。

在计量方法上，Jordà提出的局部投影能够直接估计不同预测期的响应，对完整动态系统施加的结构限制相对较少[12]；Jordà和Taylor进一步强调，动态分析应同时关注逐期估计、跨期联合推断和响应路径的不确定性[13]。这一方法与本文的研究目标一致：本文并不以某一个预测期是否显著作为主要判断依据，而是考察政策收紧后的生产率响应能否形成具有统计支持的稳定动态模式。

三、数据与研究设计

（一）样本与数据来源

本文以OECD环境政策严格度指数为政策数据基础。该指标覆盖1990—2020年40个国家和13类政策工具，并区分市场型、非市场型和技术支持三类子指数[14]。以34个OECD成员经济体作为样本。生产率及宏观投入变量来自Penn World Table 11.0；按照PWT的推荐引用，生产率数据口径参见Feenstra等[15]。制度变量来自世界银行全球治理指标，使用法治、监管质量、政府效能和腐败控制四个维度。贸易开放度与制造业增加值占比的数据来自世界发展指标。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

本文将动态分析期固定为2002—2020年。其中，h表示局部投影的预测期（horizon），取值为h=0,...,5；h=0表示政策收紧发生当期。h=0时有效样本为646个国家—年份观测，覆盖34个国家；随着前瞻期增加，末端年份自然退出，样本量依次下降为612、578、544、510和476，但各期均保留34个国家。

表1：主要变量与描述性统计

变量	数据来源	定义	观测数（N）	均值	标准差	最小值	最大值
TFP 增长率（%）	Penn World Table 11.0	相对TFP 的年度对数变化率	646	0.093	1.957	−7.073	5.083
环境政策严格度	OECD EPS	综合环境政策严格度指数	646	2.571	0.994	0.000	4.195
EPS 正向变动	OECD EPS	$max(\triangle EPS,0)$，用于刻画环境政策收紧	646	0.127	0.210	0.000	1.500
基准制度能力	World Bank WGI	1996、1998和2000 年制度PCA 的均值，并进行标准化处理	646	0.000	1.001	−2.329	1.187
市场型环境政策	OECD EPS	市场型环境政策严格度子指数	646	1.324	0.868	0.000	3.833
非市场型环境政策	OECD EPS	非市场型环境政策严格度子指数	646	4.361	1.493	0.000	5.750
技术支持型环境政策	OECD EPS	技术支持型环境政策子指数	646	2.031	1.390	0.000	5.500
资本深化（%）	Penn World Table 11.0	每名劳动者资本的增长率	646	1.527	1.908	−2.474	8.592
人力资本指数	Penn World Table 11.0	人力资本指数	646	3.267	0.369	2.092	3.767
人均GDP 对数	Penn World Table 11.0	实际人均GDP 的自然对数	646	10.679	0.418	9.589	11.494
贸易开放度（%）	World Development Indicators	进出口总额占GDP 的比重	646	94.002	54.785	22.453	310.281
制造业占比（%）	World Development Indicators	制造业增加值占GDP 的比重	646	14.695	5.068	5.002	28.522

（二）被解释变量与政策收紧

本文使用 PWT 相对全要素生产率水平（$Tf_{i,t}$）构造局部投影的累计响应变量：

$$Y_{i,t}^{(h)}=100\left[ln\left(Tf_{i,t+h}\right)-ln\left(Tf_{i,t-1}\right)\right],h=0,...,5. \quad (1)$$

其中，h=0 对应t−1至t的TFP 对数变化，h=5 对应t−1至t+5的累计 TFP 对数变化。因变量乘以100后以百分点表示，因此后文系数衡量的是政策收紧与累计 TFP 变化之间的条件动态关联。

环境政策收紧（$Tt_{i,t}$）定义为EPS（$E_{i,t}$）的一阶正向变化：

$$Tt_{i,t}=max\left(E_{i,t}-E_{i,t-1},0\right). \quad (2)$$

该定义保留政策严格度上升的幅度，并将政策放松年份记为0，使主估计聚焦于政策收紧，而不预设收紧与放松具有对称影响。由于 EPS 水平具有较强持续性，主模型同时控制$E_{i,t-1}$，以区分既有政策严格度与当期新增的政策收紧。

（三）事前制度能力

为避免把 WGI 年度缓慢变化和测量噪声直接解释为短期制度冲击，本文将制度条件固定在政策分析期之前。具体地，对每个国家计算1996、1998和2000年制度 PCA 得分$I_{i,t}$的均值，再在OECD 34个国家之间进行标准化：

$$I_i^{base}=z\left[\frac{I_{i,1996}+I_{i,1998}+I_{i,2000}}{3}\right]. \quad (3)$$

制度PCA 由法治、监管质量、政府效能和腐败控制四个维度构成。四项指标在进入 PCA 前进行标准化处理，并取第一主成分作为制度能力得分；主成分方向统一为数值越大表示制度能力越强。第一主成分解释总方差的93.21%。标准化后，I_i^{base} 对应样本国家的平均制度能力，正值和负值分别表示高于和低于样本平均水平。由于该指标在2002年以后对同一国家保持不变，其主效应被国家固定效应吸收，模型识别的是制度能力与政策收紧之间的交互差异。

（四）面板局部投影

平均动态响应模型为：

$$Y_{i,t}^{(h)}=\beta_h Tt_{i,t}+\rho_h E_{i,t-1}+\Gamma_h^{'}X_{i,t-1}+\alpha_i+\lambda_t+\varepsilon_{i,t}^{(h)},h=0,...,5 \quad (4)$$

其中，$X_{i,t-1}$ 包括资本深化、人力资本、人均GDP、贸易开放度和制造业增加值占比的一期滞后项；α_i 和 λ_t 分别表示国家固定效应和年份固定效应。β_h刻画在控制既有 EPS 水平、滞后宏观条件以及双向固定效应后，政策收紧与t−1至t+h累计TFP 变化之间的条件关联。

为考察制度条件下的响应差异，在式（4）中加入政策收紧与事前制度能力的交互项：

$$Y_{i,t}^{(h)}=\beta_h Tt_{i,t}+\theta_h Tt_{i,t-1}\times I_i^{base}+\rho_h E_{i,t-1}+\Gamma_h^{'}X_{i,t-1}+\alpha_i+\lambda_t+\varepsilon_{i,t}^{(h)},h=0,...,5 \quad (5)$$

在制度水平I下，政策收紧的条件响应为：

$$ME_h\left(I\right)=\beta_h+\theta_h I, \quad (6)$$

为便于比较不同制度条件下的动态路径，本文选取$I=-1,0,+1$三个代表性水平，分别对应低于样本均值1个标准差、样本平均水平和高于样本均值1个标准差的制度能力。

对于政策工具异质性，分别将市场型、非市场型和技术支持型政策子指数的正向变化定义为：

$$Tt_{i,t}^k=max\left(E_{i,t}^k-E_{i,t-1}^k,0\right),k\in\left\{M,N,T\right\}, \quad (7)$$

其中M、N和T分别表示市场型、非市场型和技术支持型政策。三类政策收紧变量同时进入同一局部投影：

$$Y_{i,t}^{(h)}=\beta_{M,h}\cdot Tt_{i,t}^M+\beta_{N,h}\cdot Tt_{i,t}^N+\beta_{T,h}\cdot Tt_{i,t}^T+\rho_h\cdot E_{i,t-1}+\Gamma_h^{'}\cdot X_{i,t-1}+\alpha_i+\lambda_t+\varepsilon_{i,t}^{(h)}. \quad (8)$$

因此，$\beta_{M,h}$、$\beta_{N,h}$和$\beta_{T,h}$分别表示在控制另外两类政策工具同期收紧、滞后总体EPS 和其他宏观条件后的条件动态关联，避免通过分开估计不同工具模型来选择性解释结果。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（五）分析与拟合方法

首先，在变量处理中对 TFP 、资本深化和人均 GDP 进行对数变换或对数差分，对 EPS 及其三类政策工具进行一阶差分并提取正向变化；制度变量通过四项 WGI 指标的主成分分析构造，并进一步形成事前标准化制度指标。交互效应通过政策收紧与制度能力的乘积项进入模型，低、中、高制度条件下的响应则由估计系数的线性组合得到。

其次，按照“平均动态响应—制度条件差异—政策工具差异”的逻辑逐步建立模型。对每一个预测期 h=0,…,5 分别估计一条面板局部投影方程，而不是由单一动态方程外推全部预测期。各模型均采用普通最小二乘法计算，并加入国家固定效应和年份固定效应；政策收紧使用当期值，既有政策水平和宏观控制变量使用一期滞后值。由于 $Y_{i,t}^{(h)}$ 需要 $t+h$ 期 TFP，预测期越长，样本末端可用观测越少，因此不同 h 下的样本量按数据可得性自然递减。

再次，在模型校验与推断中，正文以国家层面聚类稳健标准误为基础，以允许同一国家内部误差项存在任意序列相关和异方差 [16]。考虑到样本仅包含 34 个国家聚类，本文进一步采用 9999 次受限野生聚类自助检验（Wild Cluster Restricted Bootstrap, WCR），具体使用 WCR31 版本，对关键系数进行有限样本稳健检验；在国家层面采用 Rademacher 权重，并施加原假设约束 [16]。对于跨预测期整体推断，本文将六个预测期方程堆叠，并利用 FWL 重参数化得到与原固定效应模型等值的表示，在国家层面实施受限野生聚类联合自助检验。动态异质性不仅考察单个预测期的逐点显著性，还通过六个预测期的联合检验判断整条响应路径是否整体偏离零，并使用 Holm 校正控制同一检验族中的多重比较 [17]。双向聚类（国家和年份）和 Driscoll - Kraay 协方差估计作为误差结构敏感性分析，其中后者采用 Bartlett 核和带宽 3，用于检验结论是否依赖某一种标准误设定 [18]。

最后，模型评价不以某一预测期是否达到传统显著性阈值作为唯一标准，而主要比较系数方向、置信区间、有限聚类推断、跨期联合检验和多重校正后的证据是否一致。模型计算同时通过系数复现和重参数化等值性进行核验。

四、实证结果

本节围绕三个研究问题，依次考察环境政策收紧的平均生产率动态、制度条件差异和政策工具异质性，并结合国家层面聚类稳健检验、WCR31、跨期联合检验和多重检验校正，对逐点与整体证据进行综合评价。

（一）环境政策收紧后的平均生产率动态

平均动态模型首先回答环境政策收紧后是否存在稳定的宏观生产率调整路径。图 1 显示，h=0 至 h=5 的累计 TFP 响应始终在零附近波动，点估计在不同预测期之间多次改变符号，且所有逐点 95% 置信区间均覆盖零。这意味着样本中既没有出现持续的负向调整，也没有形成由负转正并逐渐增强的清晰动态轨迹。

表 2 进一步给出了逐期估计及其统计推断。在 5% 显著性水平下，六个预测期的国家层面聚类稳健检验均不能拒绝 $H_0:\beta_h=0$，且相应的 WCR31 检验均得到一致结论。其中，h=4 的点估计下降至 −0.505，但置信区间仍较宽；h=5 又回升至 0.220，说明中期负向点估计并未延续为稳定路径。由此，第一项研究问题得到较为明确的经验回答：在本文的 OECD 国家宏观样本中，没有观察到能够得到统计确认的平均生产率动态响应。

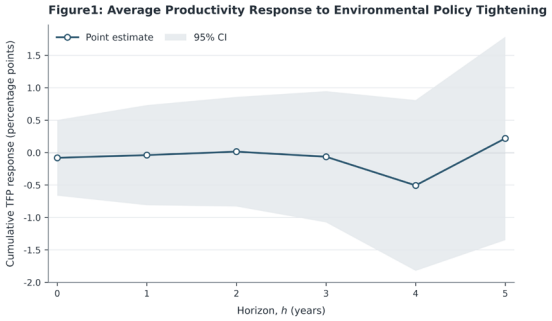


图 1：环境政策收紧后的累计全要素生产率响应

Figure 1: Cumulative TFP Response to Environmental Policy Tightening

这一结果既不支持稳定的“短期合规成本—中期创新补偿”路径 [1−3,9]，也不支持环境政策收紧持续降低生产率的单向解释。更准确地说，现有数据不能拒绝平均动态响应为零，而非证明环境政策的真实生产率效应严格等于零。这一区分也为后续异质性分析提供了基础：总体响应较弱并不排除不同制度条件或政策工具下存在方向相反的局部响应。

表 2：环境政策收紧的平均动态生产率响应

h	$\hat{\beta}_h$	95% CI	$p_{country}$	p_{WCR31}	N
0	-0.079	[-0.662, 0.505]	0.791	0.800	646
1	-0.038	[-0.809, 0.734]	0.924	0.922	612
2	0.016	[-0.829, 0.860]	0.971	0.972	578
3	-0.063	[-1.074, 0.948]	0.902	0.902	544
4	-0.505	[-1.821, 0.811]	0.449	0.450	510
5	0.220	[-1.347, 1.788]	0.781	0.785	476

注：因变量为式（1）的累计 TFP 变化；模型控制滞后 EPS、资本深化、人力资本、人均 GDP、贸易开放度、制造业占比及国家和年份固定效应。

（二）制度能力

第二项研究问题关注事前制度能力是否构成环境政策作用的边界条件。按照式（5），制度能力的作用首先体现为政策收紧与基准制度能力交互项 θ_h 的变化。表 3 面板 A 显示，h=0 时交互项估计为 0.603，国家聚类 $p=0.032$，WCR31 $p=0.046$，表明在政策收紧当期，制度能力越高，生产率响应的点估计越趋向正向。其余预测期的交互项多数仍为正，但估计精度明显下降，没有形成连续的逐期显著结果。

交互项的逐点显著并不等同于不同制度水平下的条件响应已经得到确认。图 2 和表 3 面板 B 进一步将交互模型转换为低制度（−1 SD）、平均制度和高制度（+1 SD）三种代表性条件下的边际响应。h=0 时，低制度条件下的点估计为 −0.634，高制度条件下为 0.572，呈现较为明显的方向分化；但相应 95% 置信区间均覆盖零。此后各预期的条件响应同样未在 5% 水平得到确认。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

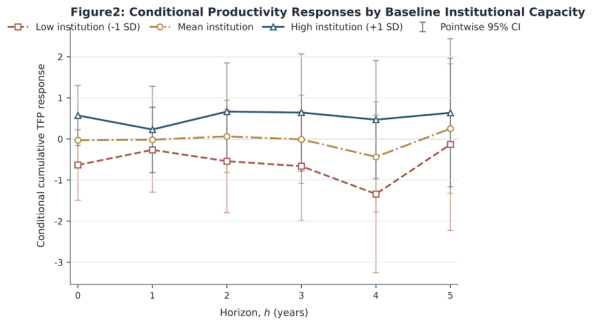


图2：不同事前制度能力下环境政策收紧的条件生产率响应

Figure 2: Conditional Productivity Responses by Baseline Institutional Capacity

进一步考虑整条动态路径后，制度异质性的统计证据明显收缩。六个预测期交互项的国家层面聚类稳健联合 Wald 检验 $p=0.160$，基于 FWL 重参数化的野生聚类自助联合检验 $p=0.405$；将六个预测期作为同一检验族实施 Holm 校正后，h=0 交互项的 p 值分别上升至0.193和0.275。同时，低、中、高制度条件响应的 Bonferroni 同时置信区间均覆盖零。

因此，第二项研究问题得到的是具有方向性但尚未得到整体推断确认的结果：事前制度能力改变了动态响应的点估计形状，尤其在政策收紧当期表现出较明显的正向制度梯度，但现不足以确认这一梯度能够稳定延伸至整个动态窗口。制度能力由此更适合被理解为具有提示性的条件因素，而不是已经得到稳健识别的单调调节效应。

表3：制度能力与政策收紧的动态异质性

面板 A：$T_t \times I^{base}$ 交互项						面板 C：跨期整体推断	
h	$\hat{\theta}_h$	95%CI	$p_{country}$	p_{WCR31}	N	检验	p 值
0	0.603	[0.051, 1.155]	0.032	0.046	646	国家层面聚类稳健联合 Wald 检验	0.160
1	0.248	[−0.416,0.912]	0.464	0.467	612	基于 FWL 重参数	0.405
2	0.604	[−0.240, 1.447]	0.159	0.136	578	基于 FWL 重参数化的联合野生聚类自助检验	0.193
3	0.653	[−0.213, 1.519]	0.138	0.111	544	Holm 校正后的 h=0 交互项（国家层面聚类稳健检验）	
4	0.904	[−0.133, 1.942]	0.086	0.072	510		
5	0.384	[−0.773, 1.541]	0.512	0.531	476	Holm 校正后的 h=0 交互项（WCR31 自主检验）	0.275

面板 B：不同制度水平的条件响应									
低制度（−1 SD）				平均制度			高制度（+1 SD）		
h	$ME_h(I)$	95%CI	$p_{country}$	$ME_h(I)$	95%CI	$p_{country}$	$ME_h(I)$	95%CI	$p_{country}$
0	−0.634	[−1.491,0.223]	0.147	−0.031	[−0.605,0.543]	0.916	0.572	[−0.158,1.303]	0.125
1	−0.264	[−1.296,0.767]	0.615	−0.017	[−0.817,0.784]	0.967	0.231	[−0.818,1.280]	0.665
2	−0.541	[−1.794,0.711]	0.396	0.062	[−0.817,0.942]	0.889	0.666	[−0.517,1.849]	0.269
3	−0.662	[−1.992,0.668]	0.328	−0.010	[−0.817,0.942]	0.986	0.643	[−0.783,2.069]	0.376
4	−1.339	[−3.255,0.578]	0.170	−0.435	[−1.774,0.904]	0.524	0.469	[−0.968,1.907]	0.521
5	−0.132	[−2.225,1.961]	0.902	0.253	[−1.321,1.826]	0.753	0.637	[−1.166,2.440]	0.488

注：面板 B 为式（6）的线性组合。Bonferroni simultaneous95% 区间亦全部跨零。图2展示的是逐点置信区间；整体推断以面板 C 为准。

（三）政策工具

第三项研究问题进一步考察政策设计本身是否对应不同的生产率动态。按照式（8），市场型、非市场型和技术支持型政策的正向变化同时进入同一局部投影，因此各工具系数是在控制另两类政策同期收紧及其他宏观条件后得到的条件动态关联。

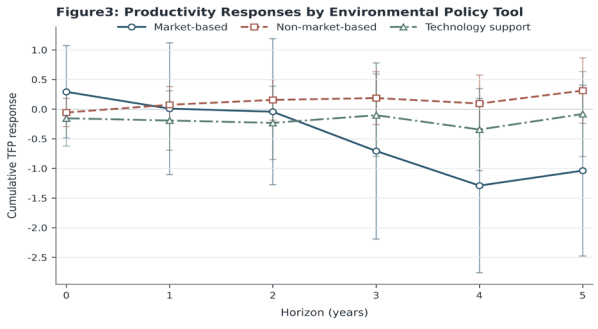


图3：不同环境政策工具收紧后的累计全要素生产率响应

Figure 3: Cumulative TFP Responses to Environmental Policy Tightening by Policy Instrument

图3显示，三类政策工具的点估计路径具有不同形态。市场型政策由 h=0 的0.293逐步转为负值，在 h=4 达到 −1.289；非市场型政策的点估计则由接近零逐渐转为正值，h=5 达到0.313；技术支持型政策在六个预测期内的点估计均为负。就路径形态而言，不同工具确实呈现出一定的动态分化。

但这些视觉上的路径差异并没有转化为相应的统计证据。表4显示，三类政策工具各预测期的95% 置信区间均覆盖零。市场型政策在 h=4 时最接近传统显著性阈值，但国家层面聚类稳健检验的 p 值为0.084，WCR31 自助检验的 p 值为0.110，在5% 显著性水平下均不能拒绝零响应的原假设。进一步从整条动态路径考察，市场型、非市场型和技术支持型政策的国家层面聚类稳健联合 Wald 检验 p 值分别为0.417、0.596和0.478；基于 FWL 重参数化的联合野生聚类自助检验 p 值分别为0.562、0.706和0.659。经工具内及全局 Holm 校正后，亦无结果达到5% 的统计显著性水平。

因此，第三项研究问题与制度能力分析呈现出相似的证据结构：政策工具之间存在经济上可辨识的点估计形态，但现有统计证据不足以将这种形态差异解释为稳定的工具异质性。换言之，本文能够描述不同政策工具对应的动态轨迹，却不能据此判断某一类工具具有系统性的生产率优势或劣势。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

表4：三类政策工具的动态响应与联合推断

h	市场型		非市场型		技术支持		N
	系数	95% CI	系数	95% CI	系数	95% CI	
0	0.293	[0.487,1.073]	−0.056	[0.293,0.181]	−0.153	[0.621,0.314]	646
1	0.009	[1.101,1.120]	0.075	[0.234,0.383]	−0.191	[0.694,0.312]	612
2	−0.043	[1.275,1.188]	0.157	[0.186,0.499]	−0.231	[0.849,0.388]	578
3	−0.705	[2.190,0.779]	0.188	[0.257,0.634]	−0.102	[0.796,0.593]	544
4	−1.289	[2.758,0.180]	0.096	[0.384,0.575]	−0.343	[1.034,0.348]	510
5	−1.036	[2.476,0.404]	0.313	[0.238,0.865]	−0.082	[0.797,0.634]	476
Country joint p		0.417		0.596		0.478	
Wild joint p		0.562		0.706		0.659	

注：三类政策收紧变量同时进入同一局部投影。逐点置信区间基于国家层面聚类稳健协方差估计；最后两行分别给出各政策工具六个预测期的国家层面聚类稳健联合检验和联合野生聚类自助检验。在各工具—预测期组合的逐点检验中，国家层面聚类稳健检验和 WCR31 检验均无结果达到 5% 显著性水平；经工具内及全局 Holm 校正后，亦无结果得到统计确认。

（四）推断敏感性与证据边界

综合三个模型可以看到，本文的主要结论并不取决于某一个预测期或某一种推断方法。平均动态模型中，国家聚类与 WCR31 在六个预测期均得到不能拒绝零响应的一致判断；制度模型中，h=0 交互项虽然在逐点推断下达到 5% 显著水平，但这一证据没有延伸到条件边际效应，并在跨期联合检验和 Holm 校正后明显减弱；政策工具模型中，三类工具的路径形态虽有差异，但逐点、跨期联合以及多重推断均未提供稳定支持。

即，平均动态响应没有形成稳定模式；制度能力呈现提示性的条件差异；政策工具呈现描述性的路径差异；后两类异质性均未通过整体推断得到稳健确认。这一结果与本文的推断策略相一致，即不能将单个预测期的显著性直接外推为整条动态路径，也不能将视觉上的曲线差异直接等同于得到统计确认的异质性。

从模型评价角度看，三个层次的分析并非通过寻找更复杂的设定来获得显著结果，而是逐步检验平均响应、制度条件和政策工具三个研究问题在不同推断标准下能否成立。本研究的推断标准由逐点检验逐步向有限聚类校验、跨期联合检验及多重比较控制收紧；在这一多维校验机制下，本文所得结论虽更为保守，却确立了更为严谨的统计推断边界。因此，本文实证结果所支持的核心判断不是环境政策收紧没有生产率效应，而是现有 OECD 宏观数据尚不足以确认一个稳定的平均动态路径，也不足以将制度和工具层面的点估计差异提升为稳健的动态异质性结论。

五、讨论

本文结果表明，环境政策收紧在国家平均层面未形成稳定的生产率响应，但制度能力和政策工具维度呈现出具有经济含义的路径差异。已有研究表明，环境政策的生产率影响会随企业技术水平、行业特征和调整能力而变化 [3,7,9]，因此较弱的宏观平均响应并不必然意味着政策影响不存在，也可能反映不同主体或政策环境下的正负响应在聚合过程中相互抵消。相较于静态政策水平，本文对政策收紧后动态路径的考察进一步表明，其生产率后果更适合从调整过程和条件差异加以理解。

（一）制度能力的条件差异与证据约束

制度能力分析提供了较为明确的局部证据。政策收紧当期，制度能力交互项为正，该数值结果在国家层面聚类稳健检验和 WCR31 检验下均达到 5% 显著水平，表明较高制度能力与更趋正向的当期生产率响应相关。不过，这一差异并未稳定延伸至后续预测期，跨期联合检验和多重检验校正后也未得到整体确认。一个可能的解释是，制度能力同时影响政策执行和经济主体调整：较强的治理能力可以增强政策可信度、促进资本和技术调整，但更有效的执行也可能提高短期合规约束 [10,11]。因此，本文结果更支持将制度能力理解为环境政策作用的条件因素，而非简单的单调调节规律。

（二）政策工具的动态分化与识别边界

三类政策工具呈现出不同的点估计路径，说明政策设计可能影响生产率调整的时间形态。市场型政策主要通过价格信号发挥作用，非市场型政策更多依赖标准和行政约束，技术支持则侧重降低技术采用和创新成本，因此其动态响应不必完全一致 [14]。尽管这些路径差异尚未获得逐点、跨期联合及多重检验的稳定统计支持，但其方向分化与已有研究关于政策工具效应依赖企业能力和行业特征的发现相一致 [7–9]。因此，本文更适合将其解释为政策工具异质性的动态线索，而非普遍的工具优劣排序。

（三）动态推断的证据层级

综合三组结果，本文最值得强调的是不同推断层次所对应的证据强度。平均动态响应没有形成稳定模式；制度能力和政策工具虽表现出一定的方向性差异，但当推断从单个预测期扩展到有限聚类稳健检验、跨期联合检验和多重检验校正后，这些异质性证据均有所减弱。

这说明，动态分析中的某一期显著、曲线存在差异和存在稳定的动态异质性并非同一层次的结论。本文因此不将个别时期的显著性外推为一般规律，而将研究结论限定为：环境政策收紧的生产率响应具有潜在的动态性和条件性，但现有 OECD 宏观样本不足以确认稳定的平均效应及系统性的制度或工具异质性。

六、结论

本文基于 34 个 OECD 国家 2002—2020 年的面板数据，将研究

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

对象从环境政策严格度水平转向政策收紧，采用面板局部投影估计 $h=0,\ldots,5$ 年的累计 TFP 响应，并进一步考察事前制度能力和不同政策工具的动态异质性。在推断上，本文结合国家聚类稳健推断、WCR31 受限野生聚类自助检验、跨期联合检验和多重检验校正，以避免依据单一预测期的显著性作出过强判断。

围绕三个研究问题，本文得到三点主要结论。第一，环境政策收紧后的平均生产率响应在 0—5 年内围绕零波动，未形成稳定的“短期成本—中期补偿”路径。第二，事前制度能力改变了动态响应的点估计形态，当期交互项呈现较明显的正向制度梯度，但条件效应、跨期联合检验和多重校正均未确认稳定的制度异质性。第三，市场型、非市场型和技术支持型政策呈现不同的动态轨迹，但这些工具差异同样未获得整体推断的稳健支持。因此，本文并非证明环境政策不存在生产率影响，而是表明在 OECD 国家宏观层面，政策收紧后的平均动态响应及其异质性仍具有较大的统计不确定性。

上述结论带来两方面启示。一方面，环境政策的生产率后果不宜仅依据当期系数或单一年度变化进行评价，而应同时考察动态调整过程及其统计不确定性。另一方面，制度能力和政策工具仍是理解政策异质性的重要维度，但现有宏观证据不足以支持简单的国家分类，也不足以判断某类政策工具具有普遍的生产率优势。本文由此将环境政策与生产率研究从静态平均关系进一步推进到政策收紧、制度条件和工具差异的动态比较，并通过多层次推断明确了相应的证据边界。

本文仍存在一定局限。首先，EPS 变化并非外生政策冲击，因此局部投影结果应解释为控制可观测因素后的动态关联，而非严格因果效应。其次，34 个国家的聚类数量限制了异质性和跨期联合检验的统计功效，尽管本文采用有限聚类稳健方法进行校验。最后，国家平均 TFP 难以直接揭示企业进入退出、行业重配和技术采用等微观机制。未来研究可进一步利用国家—行业或企业层面的细粒度数据，并结合具有明确政策时点的实验设计，检验本文所发现的制度政策工具差异及其具体传导机制。

参考文献

[1]PORTER M E, VAN DER LINDE C. Toward a new conception of the environment-competitiveness relationship[J]. Journal of Economic Perspectives, 1995, 9(4): 97-118.

[2]PALMER K, OATES W E, PORTNEY P R. Tightening environmental standards: The benefit-cost or the no-cost paradigm?[J]. Journal of Economic Perspectives, 1995, 9(4): 119-132..

[3]AMBEC S, COHEN M A, ELGIE S, et al. The Porter hypothesis at 20: Can environmental regulation enhance innovation and competitiveness?[J].Review of Environmental Economics and Policy, 2013, 7(1): 2-22.

[4] 王杰，刘斌 . 环境规制与企业全要素生产率——基于中国工业企业数据的经验分析 [J]. 中国工业经济，2014(3): 44-56.

[5] 黄庆华，胡江峰，陈习定 . 环境规制与绿色全要素生产率：两难还是双赢？ [J]. 中国人口·资源与环境，2018, 28(11): 140-149.

[6] 刘伟江，杜明泽，白玥 . 环境规制对绿色全要素生产率的影响——基于技术进步偏向视角的研究 [J]. 中国人口·资源与环境，2022, 32(3): 95-107.

[7]ALBRIZIO S, KOŹLUK T, ZIPPERER V. Environmental policies and productivity growth: Evidence across industries and firms[J]. Journal of Environmental Economics and Management, 2017, 81: 209-226.

[8]KOCH N, THEMANN M. Catching up and falling behind: Cross-country evidence on the impact of the EU ETS on firm productivity[J]. Resource and Energy Economics, 2022, 69: 101315.

[9]BENATTI N, GROISS M, KELLY P, et al. Environmental regulation and productivity growth in the euro area: Testing the Porter hypothesis[J]. Journal of Environmental Economics and Management, 2024, 126: 102995.

[10]NORTH D C. Institutions, institutional change and economic performance[M]. Cambridge: Cambridge University Press, 1990.

[11] 范小敏，刘中强 . 环保垂直管理改革如何提升企业绿色全要素生产率？ [J]. 上海财经大学学报，2026, 28(3): 90-105.

[12]JORDÀ Ò. Estimation and inference of impulse responses by local projections[J]. American Economic Review, 2005, 95(1): 161-182.

[13] JORDÀ Ò, TAYLOR A M. Local projections[J]. Journal of Economic Literature, 2025, 63(1): 59-110.

[14]KRUSE T, DECHEZLEPRÊTRE A, SAFFAR R, et al. Measuring environmental policy stringency in OECD countries: An update of the OECD composite EPS indicator[R]. Paris: OECD Publishing, 2022 (OECD Economics Department Working Papers, No. 1703).

[15]FEENSTRA R C, INKLAAR R, TIMMER M P. The next generation of the Penn World Table[J]. American Economic Review, 2015, 105(10): 3150-3182.

[16]MACKINNON J G, NIELSEN M Ø, WEBB M D. Fast and reliable jackknife and bootstrap methods for cluster-robust inference[J/OL]. Journal of Applied Econometrics, 2023, 38(5): 671-694.

[17]HOLM S. A simple sequentially rejective multiple test procedure[J]. Scandinavian Journal of Statistics, 1979, 6(2): 65-70.

[18]DRISCOLL J C, KRAAY A C. Consistent covariance matrix estimation with spatially dependent panel data[J/OL]. The Review of Economics and Statistics, 1998, 80(4): 549-560.

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

公共资金项目绩效评价中服务对象满意度调查体系设计与执行机制研究

商黄鹤，谭怡
深圳市维度数据科技股份有限公司，广东 深圳 518000
DOI:10.61369/ASDS.2026100007

摘　　要： 随着全面预算绩效管理改革深入推进，我国公共资金管理正由侧重投入和支出，向更加注重绩效结果与公共价值创造转变。服务对象满意度已成为衡量公共资金项目社会效益和公共价值的重要依据，但实践中仍存在调查对象界定不清、问卷设计同质化、样本代表性不足、调查执行不规范、质量控制薄弱和结果应用不充分等问题。本文采用文献分析法和案例分析法，围绕调查体系设计、执行机制、质量控制与结果应用，构建公共资金项目服务对象满意度调查的标准化技术路径。研究提出：建立涵盖项目认知、服务过程、项目质量、效益感知和总体评价的差异化指标体系，结合随机抽样、分层配额和多渠道数据采集提高样本代表性；构建覆盖前期规划、规范采集、数据核验、分层分析、反馈整改和资料留痕的全过程质量控制机制；将满意度结果与项目产出、资金执行和社会效益等客观资料交叉印证，形成“分析诊断—反馈整改—跟踪复核—综合应用”的绩效闭环。本研究可为规范服务对象满意度调查、提高财政资金配置效率和公共服务供给质量提供理论参考与实践指引。

关 键 词： 公共资金项目；绩效评价；服务对象满意度；调查体系；执行机制

Research on the Design and Implementation Mechanism of Service Recipient Satisfaction Survey System in Public Fund Project Performance Evaluation

Shang Huanghe, Tan Yi
Shenzhen Dimension Data Tech Co., Ltd., Shenzhen, Guangdong 518000

Abstract: As China’s comprehensive budget performance management reform advances, the management of public funds is shifting from a primary focus on inputs and expenditures toward greater emphasis on performance results and public value creation. Satisfaction among service recipients has become an important basis for assessing the social benefits and public value of publicly funded projects. In practice, however, such surveys still face problems including unclear respondent identification, homogenized questionnaire design, inadequate sample representativeness, nonstandard implementation, weak quality control, and insufficient use of findings. Using literature review and case analysis, this study develops a standardized technical framework for service recipient satisfaction surveys in publicly funded projects, covering survey system design, implementation mechanisms, quality control, and the use of findings. It proposes a differentiated indicator system encompassing project awareness, service processes, project quality, perceived benefits, and overall evaluation, supported by random sampling, stratified quota sampling, and multi-channel data collection to improve sample representativeness. It also establishes an end-to-end quality control mechanism covering advance planning, standardized data collection, data verification, stratified analysis, feedback and rectification, and recordkeeping. Survey findings are further cross-validated against objective evidence on project outputs, budget execution, and social benefits to create a performance feedback loop of “analysis and diagnosis-feedback and rectification-follow-up review-integrated application.” This study offers theoretical insights and practical guidance for standardizing service recipient satisfaction surveys, improving the efficiency of public fund allocation, and enhancing the quality of public service delivery.

Keywords: public fund projects; performance evaluation; service recipients satisfaction; survey system; implementation mechanism

作者简介：
商黄鹤，深圳市维度数据科技股份有限公司，总经理，主要研究方向为养老产业统计、智慧养老产业研究、公共绩效评价；
谭怡，深圳市维度数据科技股份有限公司，高级研究员，主要研究方向为政府及公共部门的第三方绩效评估、公众满意度调查、服务质量监测。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

引言

（一）研究背景与问题提出

随着我国全面实施预算绩效管理改革，公共财政资金管理模式　　正在发生深刻的范式转移：由传统的“重投入、重支出”的合规性监管，逐步转向以绩效目标实现、公共价值创造和公众实际受益为导向的全过程管理。《关于全面实施预算绩效管理的意见》与《项目支出绩效评价管理办法》等政策明确提出，要加快建成全方位、全过程、全覆盖的预算绩效管理体系，实现由“花钱必问效、无效必问责”向更加注重结果导向和价值创造转变。这意味着，公共资金项目绩效评价的关注重点也由单纯考察资金使用是否规范、项目任务是否完成，逐渐拓展到项目实施是否有效满足公共需求、公共服务质量是否提升以及服务对象是否真正受益等多维目标。

公共资金项目具有较强的财政属性与公共属性，其绩效既体现为活动场次、参与人数和成果数量等客观产出，也体现为服务对象的实际体验、需求满足和公共价值实现。在目前的公共财政实践中，大量专项资金是以“项目制”的形式运作的。项目制运作具有周期明确、阶段性强、受众流动性高、场景多元化等特征。传统的以财务指标、数量指标和时效指标为核心的评价方式，在一定程度上能够反映财政资金的合规性和基础产出，但对于服务对象实际体验、多元化需求满足程度以及公共价值实现程度等结果性信息反映不足。因此，引入服务对象满意度调查，将公众作为公共服务绩效评价的重要参与主体和价值评判主体，已经成为完善全面预算绩效评价体系、优化财政资源配置的关键方向。

满意度调查源于西方新公共管理运动对服务质量和用户体验的关注，其核心理念在于通过收集服务对象对项目实施过程、服务供给质量和最终效果的主观评价，弥补传统绩效评价中“重投入、重产出、轻感受”的不足。相关研究指出，公众满意度是现代政府治理中“公民—政府”反馈路径的关键一环[1]。满意度调查能够将公众由公共服务的“被动接受者”转变为“绩效评价的重要参与主体”，从而推动财政资金配置从“政府投入视角”向“公众受益视角”的转变。

当前，满意度调查已在公共资金项目绩效评价中得到广泛应用，但部分调查仍停留在“末端打分”层面，存在调查对象遴选不准确、样本量不达标、问卷设计模板化、原始资料留存缺失、调查结果难以核验等问题。这表明，尽管满意度调查在现实绩效管理实践中已得到广泛运用，但其结果往往难以转化为可归因、可比较、可追溯，可支撑预算安排与政策调整的绩效评价证据，尤其缺乏对“服务对象感受—实施过程—产出质量—实际效益”内在关联的系统揭示[2-4]。基于此，本节提出的核心研究问题是：在预算绩效管理提质增效、财政审计监督突出效益性、人民性与证据性的制度背景下，如何紧扣公共资金项目的财政属性与项目制特征，构建涵盖“体系设计—执行机制—质量控制—多维分析—闭环反馈”的标准化技术路径？这将推动满意度调查从分散化、经验化的辅助工具，转变为能够有效测度项目社会效益与公共价值、精准诊断项目管理薄弱环节、为财政预算绩效管理与政策优化提供科学参考的重要手段。

（二）研究意义

1. 理论意义

目前，满意度调查相关研究主要集中于公共服务质量评价、政府绩效评价和顾客满意度测评等领域，而针对公共资金项目（尤其是项目制运作下的专项资金项目）绩效评价中的系统研究仍显不足。在财政绩效评价场景下，满意度往往仅作为社会效益或服务对象满意度三级指标的简单测量工具，针对其调查对象识别、指标设置、样本选取、调查实施、数据分析及结果应用等环节，尚未形成系统性研究框架。

本研究将满意度调查置于公共资金项目绩效评价的完整链条中进行分析，重点研究服务对象主观感受与项目投入、实施过程、项目产出及实际效益之间的关系，有助于丰富公共资金绩效评价中满意度指标的研究视角，推动满意度调查从一般公共服务评价向财政绩效评价场景延伸。同时，通过探索满意度调查结果与项目绩效目标、社会效益指标及项目管理问题之间的关联，可进一步完善公共资金项目社会效益和服务质量的测量方法，为构建兼顾客观绩效与服务对象感受的综合评价体系提供理论参考。

此外，本研究强调满意度调查不仅用于反映服务对象“是否满意”，还应发挥识别需求差异、分析问题原因和解释绩效结果的作用，有助于拓展满意度数据在公共项目绩效评价中的功能定位，增强绩效评价理论对公众感受、获得感和公共价值等内容的关注。

2. 实践意义

本研究围绕体系设计、调查实施、质量控制、数据分析和结果应用环节构建标准化技术路径，可为财政部门、业务主管部门、项目实施单位及第三方评价机构开展相关工作提供操作参考。

一是规范满意度调查实施流程。通过明确调查对象、指标设置、问卷设计、抽样方式、数据采集与处理要求，减少调查对象识别不准确、问卷内容与项目不匹配、样本结构失衡、调查程序不规范等问题。在统一核心指标和基本调查口径的基础上，还可提升同类项目跨地区、跨年度调查结果的可比性。

二是提高调查结果的可靠性与解释力。通过加强调查过程监测、异常问卷识别和数据复核，结合总体评价、分项评价、群体差异和开放意见开展分析，避免仅以平均满意度得分判定项目成效，可更加准确地反映服务对象的实际体验、主要诉求和具体问题。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

三是强化满意度调查对绩效评价的支撑作用。将满意度调查结果与项目绩效目标、实施过程、产出质量和社会效益等资料相互印证，可为绩效评分、问题诊断和形成评价结论提供补充依据，降低仅凭主观判断或单一得分作出评价结论的风险。

四是推动评价结果转化应用。满意度调查既能够反映服务对象对项目的总体评价，也可以发现项目宣传、服务流程、内容供给、资源配置和组织管理等方面存在的不足。相关调查结果可为项目整改、政策优化和后续预算安排提供参考，推动绩效评价工作从侧重结果打分向问题诊断和管理改进延伸。

（三）研究内容与方法

1. 研究内容

本研究以公共资金项目服务对象满意度调查为研究对象，在梳理相关理论与实践经验基础上，结合公共资金项目的财政属性、实施特点和公共服务目标，分析调查对象识别、指标设置、组织实施、质量控制和结果应用环节存在的问题，进而提出与之适配的调查体系与执行方法。

首先，系统梳理公共资金项目绩效评价、公共服务满意度、公众参与和公共价值等相关理论，明确满意度调查在财政绩效场景下的功能定位。其次，结合公共资金项目绩效评价实践，深度解剖调查对象界定模糊、指标高度同质、抽样便利化严重、过程控制失效、数据真实性难验、结果分析浅显等核心困境。在此基础上，根据不同项目的服务内容（如公共文化活动、体育赛事、文艺创作等）和受益群体特征，系统设计涵盖“项目认知、服务过程、项目质量、效益感知、总体评价”五位一体的结构化适配指标体系，并确立基于统计学最小样本量计算和分层随机抽样的多渠道采集路径。建立涵盖“前置规划—规范采集—数据核验—结果反馈—全程留痕”的全流程质量控制机制。综合运用总体分析、分项分析、群体差异分析、交叉分析和原因分析等方法，研究调查结果与项目绩效指标评分、绩效等级确定、问题诊断、项目整改和预算安排之间的衔接机制，推动满意度调查结果由单纯的得分展示向绩效证据、问题诊断和管理改进转化。

2. 研究方法

本研究坚持理论研究与实践分析相结合、规范研究与实证研究相结合，主要采用以下研究方法。

（1）文献分析法

系统梳理预算绩效管理、公共价值、公共服务满意度及调查质量控制方面的政策文件、学术文献和技术标准，分析绩效目标、项目产出、实际效益与服务对象满意度之间的关系，为研究框架和技术要求提供依据。

（2）案例分析法

以公共文化活动、体育赛事、文艺创作、文化交流及宣传推广等公共资金项目为案例来源，比较不同项目在服务对象、实施方式、调查题项、样本获取和结果应用方面的差异，总结满意度调查中的共性问题和可操作经验。

一、文献综述

（一）公共服务满意度研究

满意度研究最早应用于市场营销与服务质量评价领域，主要用于考察服务对象对产品或服务的实际感受，后续逐步拓展至政府治理、公共服务、政府绩效评价等领域。在公共服务研究中，满意度一般界定为服务对象结合自身原有期望，对实际服务体验作出的主观评价，评价结果不仅与服务质量、服务效果直接相关，同时也受个人期望、个体特征、信息获取、参与经历等多方面因素影响[5]。

在满意度测量模型方面，美国顾客满意度指数模型（ACSI）和服务质量差距模型（SERVQUAL）应用较为广泛。ACSI 模型通过分析顾客期望、感知质量、感知价值、总体满意度、顾客抱怨和顾客忠诚之间的关系，解释满意度的形成过程。SERVQUAL 模型则从有形性、可靠性、响应性、保证性和移情性五个方面评价服务质量，重点比较服务对象的期望与实际感知之间是否存在差距[6]。国内学者主要结合政府绩效评价和公共服务供给实践，对满意度的概念、测量方法和应用方式进行了研究。贾奇凡等从行为公共管理学视角梳理了政府满意度的概念、测量方式及影响因素，认为公众满意度并不是政府客观绩效的直接反映，而是服务表现、公众期望和个体认知共同作用的结果[1]。季丹等将第三方机构引入公共服务质量评价，讨论了评价独立性和专业性问题[7]。郑方辉、雷比璐提出在地方政府绩效评价中重视公众满意度，使评价结果更能体现公共需求和公众感受[8]。

总体来看，满意度调查可弥补行政统计、项目台账难以反映服务对象实际体验的不足，但满意度属于主观评价范畴，不能直接等同于公共服务的客观绩效。分析调查结果时，需结合服务供给、项目产出和实际效益等相关资料，研判满意或不满意评价产生的具体原因。

（二）预算绩效评价与满意度指标研究

预算绩效评价是预算绩效管理的重要组成部分，其核心是依据绩效目标，对财政资金投入、组织管理、项目产出和实施效益进行综合测量、分析和判断。郑方辉等认为，财政绩效评价不仅是一种财政管理工具，还具有强化公共责任、提高财政资金使用效益和改善公共治理的重要价值，其指标体系应覆盖资金投入、过程管理、目标实现和社会满意等方面[9]。郑方辉、廖逸儿在对财

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

政专项资金绩效评价的研究中进一步指出，专项资金绩效评价需要明确评价目标、评价主体、评价内容和结果应用机制，避免将绩效评价简单等同于财务检查或项目验收[10]。

从功能定位看，满意度指标是连接项目客观产出与服务对象主观感受的重要补充指标。活动场次、参与人数、作品数量、任务完成率等指标能够反映项目是否完成计划任务，但不能充分说明项目内容是否符合公众需求、项目组织是否规范、参与过程是否便利以及项目是否形成预期的社会效果。通过满意度调查，可以从服务对象角度了解项目实施质量、服务体验和实际获得感，为判断项目绩效目标实现程度提供补充证据。

（三）调查方法与质量控制研究

伴随我国公共治理精细化的发展，公共服务满意度调查的技术规范与国家标准体系正逐步走向健全。由国家市场监督管理总局、国家标准化管理委员会发布的GB/T37273—2018《公共服务效果测评通则》与GB/T37229—2018《公共服务效果测评模型和方法指南》，为公共服务测评的基本原则、结构模型、问卷编制和统计方法提供了坚实的国家标准化依据。而于2026年4月1日起正式实施的GB/T46916—2025《基本公共服务社会满意度测评通则》，更是针对新时期基本公共服务的均衡性、均等化以及社会公众的满意度测评，制定了更为系统、严密的执行规范与抽样标准。这充分预示着我国满意度测评工作正从粗放的经验式调查向法治化、标准化、规范化的技术范式演进[11-13]。

综上，国内外关于满意度理论、公共服务质量、预算绩效评价和调查方法的研究已经形成了较为丰富的成果，为公共资金项目服务对象满意度调查提供了理论、制度和技术基础。但是，现有研究仍存在一定不足。一是公共服务满意度研究多集中于政务服务、医疗、教育和社会保障等持续性公共服务，对文化活动、体育赛事和文艺创作等项目制、活动型公共资金项目关注不足；二是预算绩效评价中的满意度指标与项目绩效目标、实施过程、项目产出和实际效益之间的对应关系尚不充分，满意度调查容易停留在通用问卷和总体得分层面；三是一般社会调查和公共服务测评方法尚未完全转化为适用于公共资金项目的标准化执行机制，特别是在调查对象识别、现场样本控制、异常问卷处理和调查资料留痕等方面仍需进一步细化；四是满意度调查结果与绩效评分、问题诊断、项目整改和预算安排之间的衔接机制尚不完善。

因此，本研究结合公共文化服务类公共资金项目特点，围绕调查对象、指标体系、执行机制、质量控制和结果应用构建标准化技术路径。

二、公共资金项目满意度调查的现状与瓶颈分析

（一）公共资金项目满意度调查的实践现状

在全面预算绩效管理向纵深推进的背景下，“服务对象满意度”已成为公共资金项目绩效目标设置与绩效评价环节的通用指标。特别是公共文化、体育赛事、民生微实事等直接面向公众提供服务的项目，一般均通过问卷、访谈等方式征集服务对象评价，并将调查结果纳入单位自评或第三方绩效评价工作。

当前，满意度调查主要由项目实施单位、预算主管部门、财政部门或第三方评价机构组织开展。部分实施单位在活动现场自行发放、回收问卷，第三方机构多采用现场访问、电话回访或线上调查等方式。不同调查主体在人员配置、专业能力和独立性方面存在差异，调查实施与质量控制要求也不尽一致，会对调查结果的可靠性产生影响。

从问卷内容设计来看，多数项目采用“总体评价 + 分项评价”的基本结构。总体评价主要用于掌握服务对象对项目的整体满意程度，分项评价则覆盖项目内容、组织安排、服务态度、参与便利性、现场环境和设施保障等具体维度。该设计可基本反映服务对象的参与体验，但部分问卷仍以通用题目为主，与项目绩效目标和实际服务内容结合不够紧密。

调查方式通常结合项目实施场景和服务对象名录情况确定。体育赛事、文化演出等现场参与人数较多的项目，主要采用现场访问或扫码填写问卷的方式开展；存有报名记录、服务台账或受众名单的项目，可通过定向发送线上问卷或开展电话回访进行调查。线上调查便于快速归集数据，但如果问卷公开传播、缺少身份筛选和重复填写管控，样本是否为实际服务对象往往难以核实。

现阶段，满意度调查结果主要用于核算绩效评价得分，并作为项目获得公众认可的佐证材料。部分评价报告会进一步整理开放性意见，提出项目整改建议，但调查结果与后续预算安排、政策调整和项目管理之间的关联仍然不够紧密。部分报告侧重呈现满意度高低，对不同群体评价差异及不满意原因分析不够深入，调查的诊断作用尚未得到充分发挥。

总体来看，满意度调查已逐步从辅助性工作转变为公共资金项目绩效评价的常规内容，应用范围持续扩大。但在实际执行中，仍有部分调查偏重问卷回收与结果计分，对样本来源、调查过程和数据质量关注不足。部分项目满意度长期维持较高水平，不同项目、不同年度之间差异较小，难以依托该结果识别具体问题。同时，部分调查缺少抽样记录、原始问卷、回访记录和数据审核资料，无法充分说明样本代表性与调查过程规范性，也限制了满意度结果在绩效评价中的证明效能。

（二）公共资金项目满意度调查的现实困境分析

首先，调查设计与项目实际场景的适配性不足。不同公共资金项目在服务内容、实施方式、受益对象和绩效目标等方面差异较大，但部分项目仍采用较为通用的满意度指标和问卷模板，未能充分体现文化活动、体育赛事、文艺创作和宣传类等项目在内容质量、参与体验、组织实施和社会效益等方面的特点。同时，抽样设计也较为粗放，样本常集中于特定场次、时段或参与渠道，难以保证调查结果的代表性。此外，部分项目过度依赖现场问卷或线上调查，数据采集渠道较为单一，容易产生样本覆盖不足、重复填写和调查对象身份难以核实等问题。

其次，调查执行过程的规范性不足。部分项目在调查前缺少完整的调查方案，对调查对象、样本数量、调查时间、抽样方式和人员分工等事项规定不明确，问卷未经预测试便直接使用，调

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

查人员培训也不够充分。在现场调查过程中，样本选择随意、调查口径不统一、项目实施单位直接参与问卷发放等情况较为常见，可能对服务对象的回答产生影响。调查结束后，部分项目未对填写时长、答案逻辑、重复样本和异常数据进行系统审核，结果分析主要停留在满意度比例和平均得分层面。

然后，调查质量控制体系尚不健全。当前部分项目的质量控制主要集中于问卷回收后的形式审核，尚未形成覆盖调查方案审查、问卷测试、人员培训、现场监督、数据核验和资料归档的全过程控制机制。同时，公共资金项目满意度调查在调查对象识别、样本分配、无效问卷判定、满意度计算和数据复核等方面缺少统一、细化的操作标准，不同评价主体往往依赖经验开展调查，影响调查结果的规范性和可比性。此外，部分调查由项目实施单位自行组织，即使引入第三方机构，也可能缺少独立抽样、过程监督和结果验证，调查方案、原始问卷、访问记录和数据清洗资料保存不完整，导致调查结果的真实性和可追溯性难以得到充分保障。

最后，调查结果的分析深度和绩效应用程度不足。现有分析往往停留在总体满意度、满意率和平均得分等单一结果呈现上，例如仅以“满意率达到95%”说明项目实施效果，缺少对不同年龄、性别、居住区域、参与渠道、参与频次和受益程度等群体的差异分析，也较少深入解释低分指标、“一般”及不满意评价产生的具体原因。更为突出的是，满意度这一主观感知数据经常被孤立使用，未能与项目资金执行率、活动场次、实际参与人数、服务台账、成果验收和传播覆盖等客观绩效资料进行交叉印证，导致满意度得分与项目实际产出、实施过程和社会效益之间缺乏清晰联系。由此，满意度调查容易沦为证明项目获得认可的形式化工具，难以真正发挥绩效评价支撑、问题诊断和项目改进作用。

三、公共资金项目服务对象满意度调查体系设计

（一）核心设计原则

公共资金项目服务对象满意度调查，应当坚持以公共价值创造和财政绩效评价需求为导向，统筹兼顾科学性、客观性与可操作性。

表1：公共资金项目服务对象满意度调查体系核心设计原则

1. 公共价值导向原则：围绕项目绩效目标，关注服务对象的实际体验、受益程度和公共需求满足情况。
2. 分类适配原则：保留必要的通用指标，同时根据项目类型、实施场景和服务对象差异化设计调查内容及抽样方式。
3. 全过程控制原则：将调查方案设计、问卷测试、数据采集、审核清洗和结果反馈纳入统一流程。
4. 客观独立性原则：调查样本抽取、问卷发放和数据回收应尽量减少项目实施单位的直接干预，避免选择性发放和引导作答。
5. 实操可行性原则：在满足调查质量要求的前提下，合理控制问卷长度、样本规模和实施成本，保证调查方法能够实际执行。

（二）调查对象识别

公共资金项目满意度调查对象应根据项目绩效目标、服务内容、实施方式和受益范围进行识别，主要包括直接受益对象、潜在受众以及项目实施与管理相关方。

1. 直接受益对象：指实际参与项目活动、接受项目服务或使用项目成果的个人和组织，如文化演出的现场观众、体育赛事参赛人员、公益培训学员和文体场馆使用者等。直接受益对象能够对项目内容、服务过程和实际效果作出直接评价，是满意度测算的主要对象。

2. 潜在受众群体：指符合项目服务定位、处于项目辐射范围内，但尚未实际参与项目的公众。对此类群体开展调查，应当重点了解项目知晓度、未参与原因、潜在需求和参与意愿。

3. 项目实施与管理相关方：包括项目承办单位工作人员、行业专家、志愿服务组织、预算主管部门管理人员等。调查内容主要围绕组织管理、资源配置、部门协同和政策执行情况开展，用于补充分析项目实施过程中遇到的困难。

（三）指标体系设计

公共文化服务类项目的调查指标，应当按照“绩效目标—调查维度—具体指标—问卷题项”的逻辑框架设置，在保持基本框架统一的基础上，结合不同项目的服务内容和预期效益作出差异化调整。调查内容主要涵盖项目认知、服务过程、项目质量、效益感知和总体评价五个部分以及背景信息题。

表2：调查指标内容设计

调查维度	指标类型	核心指标	具体评价内容	结果用途
（一）项目认知	辅助分析	1. 项目知晓度	• 是否在活动前知晓该项目 • 获取信息的途径（微信公众号 / 社区海报 / 口口相传等） • 宣传信息（时间、地点、门票 / 参与规则）是否明确	分析项目传播范围、信息触达效果及知晓度不足的原因
		2. 信息获取渠道		
		3. 宣传清晰度		
（二）服务过程	核心计分	1. 参与便利性	• 报名 / 预约 / 入场等环节是否便捷、顺畅 • 现场引导、秩序管理与工作人员服务态度 • 现场环境卫生的舒适度与安全保障措施	判断服务流程和现场组织中存在的问题
		2. 流程顺畅度		
		3. 现场秩序与服务		
		4. 安全与应急保障		
（三）项目质量（分类适配）	核心计分	1. 文化活动类：内容丰富度、艺术 / 专业水平、现场呈现效果	• 内容或活动安排是否具有吸引力和专业水准 • 场馆设备、音响灯光或竞赛设施是否完备 • 项目供给与服务对象需求的匹配程度	反映项目实施质量，为同类项目比较和改进提供依据
		2. 体育赛事类：竞赛组织专业性、裁判执法、场地设施与医疗保障		
		3. 文艺创作及宣传类：作品思想 / 艺术品质、展示传播效果、受众共鸣度		

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

调查维度	指标类型	核心指标	具体评价内容	结果用途
（四）效益感知	核心计分	1. 需求满足度	• 是否丰富了精神文化生活或提升了身心健康 • 对提升区域文化氛围或社区凝聚力的作用 • 公众对该项目未来继续获得财政资金支持的态度	判断项目实际效益及服务对象的主观获得感
		2. 文化获得感 / 实际受益		
		3. 社会影响力		
		4. 项目持续开展必要性		
（五）总体评价	综合 / 辅助	1. 总体满意度（量表打分）	• 对项目的综合体验进行整体评价 • 未来是否愿意再次参加或推荐给他人（用于校验数据真实性） • 具体的改进建议与不满原因（梳理“问题清单”依据）	总体满意度作为直接评价结果，再次参与和推荐意愿用于辅助校验
		2. 再次参与意愿		
		3. 推荐意愿		
		4. 开放性意见 / 改进建议		
背景信息（人口学特征）	辅助分析	1. 受访者基本特征	• 年龄、性别、居住区域、职业等人口学特征 • 是否首次参与、往期参与频次等	不计分（用于交叉分析与群体差异诊断）
		2. 参与频次与深度		

（四）问卷设计

公共资金项目服务对象满意度调查问卷应以项目绩效目标和指标体系为基础，主要包括卷首说明、筛选题、满意度量表题、分类信息题和开放建议题，题项应做到简明、具体、中性，避免使用专业术语、双重问题和诱导性表述。

问卷应保留总体满意度等通用题目，以便同类项目之间进行比较，同时根据项目特点设置差异化内容。文化活动类项目可重点评价活动内容、专业水平、组织安排和文化体验；体育赛事类项目可重点评价赛事组织、参与流程和安全保障；文艺创作及宣传推广类项目可重点评价作品质量、传播渠道、受众认同和社会影响。

不同调查对象应使用不同版本的问卷。直接受益对象重点评价服务过程、项目质量和实际受益情况；潜在受众重点调查知晓渠道、未参与原因和服务需求；项目实施与管理相关方重点反映组织管理和政策执行情况。各版本可保留少量共同题目，但不宜使用同一套题目评价不同群体。

正式调查前应进行预测试，重点检查题目是否容易理解、选项是否完整、填写时间是否合理，以及题目之间是否存在重复或逻辑冲突，并根据测试结果进行修改。

（五）计分方式设计

满意度调查通常采用五级李克特量表，将“非常满意、比较满意、一般、不太满意、很不满意”分别赋予相应分值，并根据评价需要计算总体满意度、分项满意度、满意率和综合得分。

指标权重应结合项目绩效目标、服务重点和专家意见确定，同一评价周期内的同类项目宜保持基本一致。对于普惠性项目，可适当提高服务便利性和实际受益情况的权重；对于专业性较强的项目，可适当提高项目质量的权重。如采用德尔菲法或层次分析法，应说明专家构成、判断过程和权重计算依据。

总体满意度得分计算公式如下：

$$S = \sum_{i=1}^{n} W_i \times \bar{X}_i$$

其中，S 为项目满意度综合得分，W_i 为第 i 个评价维度的权重系数，$\bar{X}_i$ 为该维度下所有样本评分的算术平均值。

结果分析除呈现总体满意度、综合得分和分项得分外，还应比较不同场次、参与方式和服务对象的评价差异，重点归纳“一般”及“不满意”评价的具体原因。样本量不足、结构明显失衡或异常问卷较多时，应说明数据局限，审慎将调查结果用于绩效评分。

（六）抽样设计与数据采集路径

公共资金项目满意度调查的抽样设计，应当以服务对象识别为基础，结合项目受众规模、服务范围、活动场次和实施方式，合理确定样本来源、抽样方法及样本量。对于具备完整参与名单的项目，可采用随机抽样或分层抽样方法；对于文化活动、体育赛事等难以获取完整受众名单的项目，应当按照活动场次、举办地点、参与时段、参与方式和受众类型合理分配样本，避免样本过度集中于单一场次或特定群体；对于文艺创作及宣传推广类项目，应当根据作品传播渠道和目标受众，分别覆盖现场观众、线上受众及重点传播群体，确保调查样本与项目实际服务范围相匹配。

为有效消除样本偏差，可建立“分层配额抽样模型”。一是根据项目实施的总受众规模或历史台账数据，依据统计学公式计算满足置信度（通常为95%）和抽样误差（控制在 ±5% 以内）的最小样本量（N）：

$$N = \frac{Z^2 \times P \times (1-P)}{E^2}$$

二是在确定总样本量后，结合项目制特征开展多阶段分层抽样。第一阶段：将资金支持的活动总场次，按举办区域（如中心城区与边缘地区）、活动类型（如演出、讲座、赛事）、举办时间（如工作日与周末、白天与夜间）开展分层配额；第二阶段：在选定的典型场次中，采用“等距拦截法”或“随机数表法”抽取具体服务对象，保障各子群体的样本比例与实际受众结构高度契合。

在数据采集路径方面，应当构建“多源采集、线上线下协同、异常拦截”的混合采集路径，结合项目实施特点灵活选取现场调查、线上问卷、电话回访等采集渠道。现场调查适用于具备明确活动场景的项目，线上调查适用于传播范围较广、受众较为分散的项目，同时应当采取身份筛选、填写次数限制、重复样本识别等管控措施。针对老年人、不熟悉线上操作的群体，应当适当增加线下访问或辅助填写环节。对于不同渠道调查结果差异较大、满意度异常偏高或偏低，以及调查结果与项目客观绩效不一致的情形，可通过开放性意见征集、电话回访或补充访谈核实原因，最终形成“分层抽样、多渠道采集、异常核验、定量与定性

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

相互印证”的数据采集路径。

四、公共资金项目满意度调查的闭环执行与质量控制机制设计

（一）调查技术方案规划与前置准备

满意度调查实施前，评价主体应紧扣项目绩效目标、服务内容和实际实施场景，编制调查技术方案，明确调查目的、对象范围、样本规模、抽样方式、数据采集渠道、无效问卷判定规则、结果用途及应急处理要求。针对文化活动、体育赛事、文艺创作及宣传推广等不同项目，应结合活动场次、举办地点、传播渠道和受众构成设计适配性问卷及样本分配方案；具有完整服务对象名录的，可采用随机或分层抽样，难以取得完整名录的，可按场次、时段、地点和受众类型进行配额抽样或现场拦截。正式调查前应组织专家评议和小范围预测试，重点检验题项理解、选项设置、填写时长及初步信度，并根据测试结果调整问卷；同时对调查人员开展职业操守、中立访问话术、操作流程和突发情况处置培训，明确保密及利益冲突回避要求，从源头减少调查执行中的人为偏差。

（二）数据采集与过程管控

数据采集应根据项目服务方式和服务对象分布，综合采用现场调查、线上问卷、电话回访和补充访谈等方式，并重点保障调查过程的独立性和样本真实性。文化活动、体育赛事等现场项目应由调查人员按照既定抽样规则独立选取受访者，合理覆盖不同场次、时段、地点和受众类型，项目承办单位原则上不得参与问卷发放、代填或引导作答；文艺创作及宣传推广类项目则应结合展示平台、传播渠道和目标受众确定样本来源。线上调查应通过对象筛选、填写次数限制、答题时长监测、重复样本识别和异常时段预警等措施，防止非目标对象填报、重复提交和集中代填，必要时可在依法合规和保护个人信息的前提下，利用账号、设备或区域信息辅助核验。电话调查应完整记录接通、拒访、无应答和无效样本等情况，并对样本量异常增长、渠道结果差异明显等情况及时开展人工复核；对于老年人、低频参与者和不熟悉数字化操作的群体，应适当增加线下访问或辅助调查，提高样本覆盖的完整性。

（三）数据核验、清洗与分层统计分析

调查数据回收后，应按照预先设定的统一规则，对问卷完整性、填写时长、答案逻辑、重复提交和异常作答进行核验，对关键评价题缺失、作答时间明显低于预测试合理区间、前后答案矛盾、非目标对象填报以及多项特征显示机械作答的样本予以剔除。连续选择同一选项或全部给予高分不宜单独作为无效判定依据，应结合填写时长、开放题内容和作答逻辑综合判断，联系方式等非必要信息缺失也不应直接认定为无效。对清洗后的有效样本，除计算总体满意度、分项满意度和满意率外，还应根据项目类型、活动场次、参与渠道、年龄、区域和参与频次等开展交叉比较；在样本量和统计条件允许的情况下，可采用卡方检验、独立样本 T 检验或方差分析判断群体差异，并运用重要性—满意度分析识别优先改进事项。同时，应将调查结果与活动场次、实际参与人数、上座率、传播覆盖、业务台账、成果验收和资金执行情况等资料进行综合比对，避免仅凭满意度高低判断项目整体成效。

（四）调查结果应用与绩效闭环反馈

经核验分析的满意度调查结果应纳入公共资金项目绩效评价报告，作为判断公共服务质量、社会效益和服务对象获得感的重要参考。对分项得分较低、群体差异明显以及开放性意见集中反映的问题，应从项目设计、组织执行、资源配置和服务保障等方面分析原因，形成问题清单和整改清单，明确责任主体、整改措施和完成时限，并反馈至项目实施单位、主管部门和财政部门。后续可通过绩效监控、整改回访或下一周期满意度调查跟踪改进效果，并将满意度表现、整改成效与项目产出、资金效益和历年绩效综合分析，为下一年度预算安排、项目调整和政策优化提供参考。对于满意度持续偏低、目标群体覆盖不足、实际效益不明显且整改不到位的项目，可综合采取调整项目内容、压减预算、暂停新增支持或重新论证必要性等措施；对于公众认可度较高、社会效益较好且具有示范价值的项目，可总结推广实施经验，形成“分析诊断—反馈整改—跟踪复核—综合应用”的绩效闭环。

（五）全流程质量控制机制

公共资金项目满意度调查应建立覆盖方案设计、调查执行、数据处理、独立评估和资料归档的全过程质量控制机制。方案质量控制重点审核调查对象识别、指标设置、问卷内容、样本分配和调查渠道是否与项目实际相匹配；执行质量控制主要通过调查人员培训、统一操作标准、现场督导、后台监测和抽查回访，减少调查过程中的人为偏差；数据质量控制应建立明确的无效问卷判定和异常数据处理规则，识别重复填写、逻辑矛盾、异常作答及样本结构失衡等问题；独立性控制应尽量减少项目实施单位对调查对象选择、问卷发放和结果统计的直接干预，必要时可由第三方机构独立或协助实施；档案留痕控制应完整保存调查方案、问卷版本、抽样记录、原始数据、数据清洗记录、回访复核材料和整改反馈资料，形成完整、可追溯、可复核的调查证据链。

五、满意度评价结果的多维分析与政策应用

本研究从指标设计、调查实施和质量控制等方面，对公共资金项目满意度调查的基本方法作出说明。获取调查数据后，分析工作不能仅停留在计算总体得分层面，还需进一步阐释不同群体、不同评价维度之间的差异，验证满意度结果与项目实际实施情况是否相符，并将发现的问题转化为具体改进举措。由于满意度反映的是服务对象的主观感受，开展分析时应当结合项目产出、资金执行和社会效益等客观资料综合研判。本章主要阐释满意度结果在绩效评价、问题诊断和后续管理中的具体应用。

（一）支撑绩效评价结论

满意度调查结果可作为公共资金项目绩效评价的一项重要依据，但不能单独用于判定项目整体绩效。总体满意度反映服务对象对项目的综合评价，分项结果用于分析信息获取、参与流程、

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

组织实施、内容质量、服务保障和实际受益等方面的具体情况。

不同类型项目的绩效目标和服务方式存在差异，分析内容应当各有侧重。文化活动类项目可重点聚焦内容质量、组织安排和文化体验；体育赛事类项目可重点分析赛事组织、安全保障和参与体验；文艺创作及宣传推广类项目应当结合作品质量、传播覆盖和受众反馈开展评价。

在样本量符合要求的情况下，还可按照年龄、居住区域、受众类型、参与场次、参与渠道和参与频次等开展分组比较，摸清不同群体间是否存在明显评价差异，为研判项目服务覆盖和资源配置情况提供参考。

满意度结果还需与活动场次、参与人数、传播覆盖、成果验收、业务台账和资金执行情况等资料结合分析。若主观评价与项目实际完成情况存在明显差异，应当进一步核查样本结构、调查过程和数据来源。例如，满意度较高但目标群体覆盖不足，可能说明调查样本主要集中于已参与项目的人群，不能以此判定项目整体绩效良好。只有在主客观资料相互印证的基础上，满意度结果才能为绩效评价结论提供较为可靠的支撑。

（二）服务项目问题诊断与舆情补充

满意度评价应重点发挥问题识别和原因分析作用。针对分项得分较低以及选择“一般”或不满意的样本，评价主体需结合关联题项、开放性意见和必要的电话回访展开深度剖析，以此判定问题究竟是源于宣传覆盖不足、参与渠道不畅、内容供给与需求不匹配、现场组织不到位，还是项目定位和资源配置不合理；在此基础上，还需清晰地界定个别体验问题、具体执行问题与系统性项目机制问题。

同时，规范的抽样设计赋予了满意度调查广泛的代表性，能够有效获取普通参与者、低频参与者和弱覆盖群体的真实意见，从而有力弥补网络评论和投诉热线中负面情绪集中、普通受众反馈不足的制度性偏差。在实际应用中，调查结果应与热线工单、投诉建议、网络评论、现场记录和访谈资料进行跨源交叉比对：对多个渠道反复涌现的问题实施重点核查，对单一渠道的意见则结合发生频率和影响范围审慎研判其代表性。需要明确的是，满意度调查主要发挥补充识别与原因分析的辅助作用，并不能直接替代专门的舆情监测与投诉处理工作。

（三）推动结果应用闭环

为避免满意度调查仅停留在结果得分和报告展示层面，应当建立“结果分析—问题反馈—整改落实—跟踪复核—管理应用”的工作机制。在将满意度结果用于绩效评分和项目研判前，应当首先核查样本数量与结构、调查实施过程和数据质量。针对样本量不足、结构明显失衡或调查过程存在缺陷的，可优先组织开展补充调查；无法开展补充调查的，应当说明数据局限性，并按照事前明确的评价规则降低权重或不纳入计分，避免临时调整影响评价工作的公平公正。

在调查结果可靠的基础上，可结合总体满意度、分项评价、群体差异、不满意原因和开放性意见梳理形成问题清单。问题清单应当明确具体问题、判断依据、涉及环节和可能成因，并据此建立整改台账，明确整改措施、责任单位和完成时限。针对低分项目、群体评价差异较大或主客观数据不一致的项目，还应当进一步核查项目实施过程，避免仅依据问卷结果提出笼统整改建议。

项目实施单位应当依据调查结果优化项目内容、宣传方式、参与流程和服务保障；主管部门负责跟踪整改进度，并在后续检查或下一轮评价中复核改进情况；财政部门可将整改成效与项目产出、资金使用效益和历年绩效表现结合开展分析，将其作为后续预算安排、支持规模调整和项目管理的重要参考。通过持续反馈与复核，推动满意度调查真正服务于项目改进和预算绩效管理。

六、结论与展望

（一）研究结论

公共资金项目满意度调查不仅要反映服务对象是否满意，还应说明各环节具体存在的问题，并与项目产出和实际效益相互印证。针对公共资金项目满意度调查中常见的对象界定不清、问卷内容同质化、样本代表性不足、执行不规范以及结果应用不充分等问题，本文以公共文化服务类项目为主要研究对象，从指标设计、样本抽取、调查执行、质量控制和结果应用等方面提出了较为完整的技术路径。主要结论如下。

1. 差异化设计指标体系，兼顾基本共性与项目差异。满意度调查可围绕项目认知、服务过程、项目质量、效益感知和总体评价设置基本指标，同时根据文化活动、体育赛事、文艺创作、宣传推广等项目的服务方式和成果特点调整具体内容。这样既能保持同类项目之间的基本可比性，也能避免“一套问卷用于所有项目”，使调查结果更贴近项目实际。

2. 调查对象和样本结构是影响结果可靠性的关键。直接受益对象应作为满意度测算的主要对象，潜在受众、项目实施单位和管理人员的意见可作为补充信息，原则上不宜与直接受益对象的数据合并计算。对于服务对象流动性较强、缺少完整名录的项目，可根据实施场景采用随机抽样、分层配额和现场拦截等方式，并对调查时间、地点和人群结构作出合理安排，提高样本与项目实际服务范围的匹配程度。

3. 强化全过程质量控制。质量审核不应仅集中于数据回收环节，应当覆盖调查前、调查实施、调查收尾全流程，涵盖问卷预测试、调查业务培训、现场督导、后台监测、抽查回访、数据清洗等各工作环节。对调查过程、异常问卷处理和原始资料进行必要留痕，有利于及时发现集中代填、异常重复和逻辑矛盾等问题，也便于后续复核。

4. 推动满意度结果与客观资料结合运用。服务对象的主观评价不能替代项目台账、成果验收和实际参与数据。项目知晓度可与传播覆盖情况对照校验，参与体验可结合现场组织情况核查核验，效益感知可依托上座率、参与人数、社会反响等资料开展分析。在此基础上，将分项得分、群体差异和开放意见归纳为具体问题，形成整改清单，为实施单位改进服务、主管部门完善管理、财政部门安排后续预算提供参考。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（二）研究展望

本文提出的调查方法主要面向公共文化服务类项目，其适用性还需要在更多项目类型和地区实践中进一步检验。后续研究可重点关注以下三个方面。

1. 拓展满意度数据采集方式。在依法合规、保护个人信息的前提下，可将问卷调查与报名、票务、预约、热线工单、网络评论等数据结合，探索在服务完成后开展简短即时评价，减少回忆偏差与填报负担。同时，严格控制重复填报，关注样本结构，避免少数高频网络用户意见被过度放大，为老年人等群体保留线下调查和电话回访渠道。

2. 逐步建立跨项目、跨年度的满意度数据库。在统一核心指标和数据口径的基础上，可通过纵向对比跟踪项目整改成效，通过同类项目横向对比分析实施差异。开展对比分析时，应当充分考量项目目标、服务对象和实施条件差异，不得简单按照满意度得分进行排名。待数据积累到一定量级后，可进一步分析财政投入、实施方式和公众感受之间的关联，为识别满意度持续下降、主客观评价差异较大的项目提供参考。

3. 完善数字化评价的数据治理。进一步明确数据采集范围、授权方式、保存期限和使用边界，建立算法分析结果人工复核与异常纠正机制。同时，加强对第三方调查机构的数据安全管理和职业规范约束，结合不同类型项目实际应用情况，持续检验调整调查指标、抽样方法及质量控制标准。

参考文献

[1] 贾奇凡，尹泽轩，周洁 . 行为公共管理学视角下公众的政府满意度：概念、测量及影响因素 [J]. 公共行政评论，2018,11(1):62-82+220.

[2] 马亮，于文轩 . 第三方公共服务绩效评价的评价：一项比较案例研究 [J]. 南京社会科学，2013,(5):55-63.

[3] 深圳市统计局 .2022 年度政府绩效满意度调查项目绩效评价报告 [OL].2023.

[4] 北京市丰台区人民政府 . 关于丰台区预算绩效管理情况的说明 [OL].2023.

[5]VANRYZINGG.Expectations，“performance,and citizen satisfaction with urban services”[J].JournalofPolicyAnalysisandManagement,2004.

[6]PARASURAMANA,ZEITHAMLVA,BERRYLL.SERVQUAL，“Amultiple-item scale for measuring consumer perceptions of service quality”[J].JournalofRetailing,1988.

[7] 季丹，郭政，胡品洁 . 公共服务质量第三方评价研究——基于华东地区的试点应用 [J]. 中国行政管理，2016,(1):41-44.

[8] 郑方辉，雷比璐 . 基于公众满意度导向的地方政府绩效评价 [J]. 中国特色社会主义研究，2007,(3):47-52.

[9] 郑方辉，廖逸儿，卢扬帆 . 财政绩效评价：理念、体系与实践 [J]. 中国社会科学，2017,(4):84-108+207-208.

[10] 郑方辉，廖逸儿 . 财政专项资金绩效评价的基本问题 [J]. 中国行政管理，2015,(6):46-52.

[11] 国家市场监督管理总局，国家标准化管理委员会 . 公共服务效果测评通则，GB/T37273—2018[S]. 中国标准出版社，2018.

[12] 国家市场监督管理总局，国家标准化管理委员会 . 公共服务效果测评模型和方法指南，GB/T37229—2018[S]. 中国标准出版社，2018.

[13] 国家市场监督管理总局，国家标准化管理委员会 . 基本公共服务社会满意度测评通则：GB/T46916—2025[S]. 中国标准出版社，2025.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

一种稳健多变量自相关控制图设计方法及其在空气质量监控中的应用

王志坚[1,2]，张钰哲[1]，杜玮[1]
1. 广东财经大学 统计与数据科学学院，广东 广州 510320
2. 新疆政法学院 经济管理学院，新疆 图木舒克 843900
DOI:10.61369/ASDS.2026100008

摘　　要： 传统 VAR 控制图面对含异常值的多变量时间序列时，往往容易被个别异常点带偏，且往往无法准确指出所有发生异常的变量。监控结果一旦与真实情况错位，决策者就很难及时锁定并调整异常变量。针对这一短板，本文把稳健估计量与加权方法 Zt 结合起来，构造了一种稳健 VAR 控制图，并通过不同污染率、不同维度下的模拟实验检验其表现。结果显示，与原有方法相比，新控制图在异常识别的准确率和灵敏度上都有改善，异常值比例较高时优势更明显。这项工作为多变量时间序列监控增加了一种可选工具，也为数据驱动决策中的稳健性分析提供了参考。

关 键 词： 向量自回归控制图；高维数据处理；异常值检测；过程控制

A Design Method for Robust Multivariate Autocorrelation Control Charts and Its Application in Air Quality Monitoring

Wang Zhijian[1,2], Zhang Yuzhe[1], Du Wei[1]
1.School of Statistics and Data Science , Guangdong University of Finance & Economics, Guangzhou, Guangdong 510320
2.School of Economics and Management, Xinjiang University of Political Science and Law, Tumxuk, Xinjiang 843900

Abstract: Traditional VAR control charts tend to be pulled off course by individual outliers when monitoring multivariate time series containing anomalies, and they often fail to clarify which variables actually went wrong. Once monitoring results deviate from the real situation, decision-makers struggle to pinpoint and adjust anomalous variables in time. To address this weakness, this paper combines robust estimators with the weighted Ztmethod and develops a robust VAR control chart. Simulation experiments under different contamination rates and dimensions are used to test its performance. The results show that, compared with the original method, the new chart improves both accuracy and sensitivity in outlier detection, and the advantage is clearer when the proportion of outliers is high. This work provides an additional tool for multivariate time series monitoring and a reference for robustness analysis in data-driven decision-making.

Keywords: vector autoregression control charts; high-dimensional data processing; outlier detection; process control

引言

统计过程控制（Statistical Process Control, SPC）一直是质量管理的基础工具，在自动化程度高、、制造环节复杂的生产现场尤为常见。多变量统计过程控制（Multivariate Statistical Process Control, MSPC）能同时盯住多个质量特性，因此在复杂生产过程中很难被替代。不过，工业数据的复杂度和维度都在上升，传统多变量控制图用起来越来越吃力。

问题出在两个地方。其一，许多工业过程的数据里，当前观测值和过去观测值并不独立，也就是存在自相关。传统控制图偏偏假设观测独立，这个假设在实际数据里根本不成立，控制图就容易被过程偏差带偏，虚警和漏报都会冒出来。其二，工业数据中还常混有异常值，它们可能来自测量误差、设备故障或突发事故。这些异常点一旦掺进来，均值和协方差的估计就会被拉动，报警限随之失真，整个质量监控系统的可靠性也跟着打折扣。

基金项目：本文获得广东省基础与应用基础研究基金项目“多变量自相关过程稳健控制设计及评价方法研究”（项目编号：2026A1515011461）资助。
作者简介：
王志坚（1982-），男，江西余干人，博士，副教授，研究方向：统计过程控制、时间序列分析；
张钰哲（2000-），男，湖南长沙人，硕士，研究方向：时间序列分析；
杜玮（2005—），女，河南驻马店人，硕士研究生在读，研究方向：时间序列分析。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

为了克服这些问题，近年来，稳健统计方法在多变量控制图中的应用得到了广泛关注，取得了较好的研究成果。理论层，Centofanti 等构建的功能回归控制图突破传统有限维监控边界 [1]，Elisa 等的收缩估计方法解决高维小样本协方差矩阵估计奇异痛点 [2]，Bayati 等的非参数深度控制图、Pulido 等的谨慎学习机制，从分布假设、受控性能两个维度完成方法底层升级 [3,4]，王志坚等将稳健估计量应用于Shewhart控制图取得了良好的效果，为稳健控制图的应用提供基础支撑 [5]。方法优化层，AliH T 等拓展变异性监控能力 [6]，S V 等完善理性子组下的小样本监控体系 [7]，薛丽等首次将自相关时序特征纳入多变量控制图优化框架 [8]。应用层，Capezza 等、Will 等分别完成环境排放、生物测序场景的跨域验证 [9,10]，雷彬文等在工业生产场景完成实证落地 [11]。但现有研究均未兼顾空气质量数据极端污染值、多污染物耦合与时序自相关的叠加干扰，面向该场景的专用稳健控制图的研究仍亟待完善。

为了解决上述问题，本文提出了一种基于稳健统计和时间序列模型的稳健多变量自相关控制图设计方法。该方法结合了重加权最小协方差行列式（RMCD）和重加权最小体积椭球体（RMVE）等稳健估计量，能够有效减少异常值对均值和协方差估计的影响；同时，采用一阶向量自回归模型（VAR(1)）描述多变量数据的时间依赖性，从而提高了对过程异常变化的检测灵敏度和准确性。通过蒙特卡洛仿真实验，验证了所提出方法在高维度、强自相关及高异常值率的场景下的优越性能，尤其在早期异常检测和过程偏移诊断方面，显著提高了过程监控的效率和稳定性。

本文的工作为工业过程质量监控提供了一条新思路。借助稳健多变量自相关控制图，复杂工业数据中的异常值和自相关问题有望得到更好的处理，过程控制的可靠性和实时性也随之改善。就目前的结果看，这项工作在理论和应用层面都还有继续挖掘的空间。

一、传统向量自回归控制图的原理及其不稳健性

（一）传统向量自回归控制图的原理

要看清稳健 VAR 控制图为何有必要，得先弄明白传统 VAR 控制图是怎么运作的。

设有 p 个变量的时间序列数据，每个变量有 n 个观测值，写成向量 $X_t=\left(x_{1t},x_{2t},\cdots,x_{pt}\right)$，其中 $t=1,2,\cdots,n$，x_{it} 表示第 i 个变量在时间 t 的取值。一阶 p 维的 VAR 模型 (VAR(1)) 可以写成：

$$X_t=\mu_t+\Phi\left(X_{t-1}-\mu_t\right)+\varepsilon_t$$

其中 μ_t 为 t 时刻的均值向量，μ_t 为均值向量为零的独立正态随机变量向量，协方差矩阵 Σ，Φ 为自相关参数 $p\times p$ 矩阵。进一步假设 X_t 是平稳的。在此假设下，μ_t 随时间的变化为常数，因此可以写成：

$$X_t=\mu+\Phi\left(X_{t-1}-\mu\right)+\varepsilon_t$$

根据 A. A. Kalgonda and S. R. Kulkarni[12] 提出的 VAR 控制图构建的建模步骤可知，假定在时间点 t 观测到的 p 维随机变量 X_t 严格遵循第一阶向量自回归过程 VAR(1)，即 $X_t\sim N_p(\mu,\Gamma(0))$，其中 $\Gamma(0)$ 是交叉协方差矩阵。当过程处于控制状态时，X_t 服从均值向量为 $\mu_0=\left(\mu_{10},\mu_{20},\cdots,\mu_{p0}\right)$ 的多元正态分布，且 μ_0 是目标均值向量。

当过程失去控制时，通常可能是由于变量子集的变化，而不是所有变量同时发生变化 (Hawkins，1991[13])。因此，在这种情况下，如果至少有一种假设不是真的，则认为过程失控更为合适：

$$H_{i0}:\mu_i=\mu_{i0},i=1,2,\cdots p$$

如果上式对所有的 i 成立，则认为该过程受控，即为，

$$H_0:\bigcap H_{i0}$$

设置控制限：下控制限 (LCL) 设为 0，上控制限 (UCL) 设为 $C_{\rho(0),\alpha}$，其中 $C_{\rho(0),\alpha}$ 是基于置信水平和自相关结构确定的常数。

$$Z_t=Max_{1\le i\le p}\left[\left|Z_{it}\right|\right]$$

其中 Z_{it} 是标准化后的观测值，$\sqrt{\gamma_{ii}(0)}$ 是 $\Gamma(0)$ 对角线上的元素。

$$Z_{it}=\frac{x_{it}-\mu_{i0}}{\sqrt{\gamma_{ii}(0)}},i=1,2,\cdots,p$$

当 H_{i0} 为真时，可知 Z_{it} 服从标准正态分布。显然 Z_{it} 不是独立的，因此，当过程处于受控状态时，可知，

$$Pr_{H0}\left[\left|Z_{it}\right|\le C_{\rho(0),\alpha},i=1,2,\cdots,p\right]=1-\alpha$$

根据上式在取定置信水平 a 时，$C_{\rho(0),\alpha}$ 的估计为样本 Z_t 的第 $(1-\alpha)$ 百分位点。

到此，已经完成了 VAR 控制图的各要素计算，接下来绘制 Z_t 的散点图，判断控制状态：如果 Z_t 值在控制限 (LCL 和 UCL) 范围内，则认为过程处于控制状态；如果超出范围，则认为过程失控。一旦确定过程失控，根据 Z_t 对应的 Z_{it} 可以直接确定发生失控的变量，从而进行针对性的调整或干预。

（二）传统向量自回归控制图的不稳健性和局限性

先看参数估计这一环。传统 VAR 控制图大多借助最小二乘法或极大似然法来估计模型参数，而这两种方法都绕不开样本均值和协方差矩阵。问题在于，这两个统计量恰恰对异常值很敏感。过程数据里只要混进少量异常观测，VAR 模型的参数估计就会明显跑偏，残差序列也随之不再满足独立同分布假设，后续搭起来的控制统计量便失去了理论根基。结果就是，过程明明还稳定，控制图却可能频繁报警，虚警概率一路往上升。

异常值和自相关一旦同时出现，传统 VAR 控制图的下滑会更明显。自相关结构会充当放大器，把异常值对参数估计的干扰成倍放大，异常观测借着时间依赖关系向相邻样本扩散；与此同时，异常值又会削弱模型刻画真实自相关结构的能力，造成拟合不足或过度拟合。两股作用叠加在一起，控制限就很难如实反映过程的真实波动水平，控制图对过程状态变化的判别能力也随之

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

明显下降。

为验证这一点，本文设计了一组数值模拟实验。先构建一阶向量自回归模型，生成二维多变量时间序列样本，用来刻画变量之间随时间演化的动态相关结构，其中当前观测由前一时刻状态和随机扰动共同决定。在此基础上随机生成100个观测值作为基准数据集，并从中抽取约10%的样本施加人为扰动，模拟异常值。通过分析污染数据条件下VAR控制图的监控表现，观察异常值对其稳定性和报警性能的影响。结果见表1。

表1：含异常值模型的参数以及检测效果

	系数阵	均值向量	协方差阵	UCL	报警次数
对照组	$\Phi=\begin{bmatrix}0.26 & 0.18\\ 0.24 & 0.36\end{bmatrix}$	$\mu=\begin{bmatrix}0.03\\ 0.01\end{bmatrix}$	$\Sigma=\begin{bmatrix}1.09 & 0.58\\ 0.58 & 1.08\end{bmatrix}$	2.22	4(0)
实验组	$\Phi=\begin{bmatrix}0.38 & -0.17\\ 0.53 & -0.24\end{bmatrix}$	$\mu=\begin{bmatrix}0.39\\ 0.58\end{bmatrix}$	$\Sigma=\begin{bmatrix}2.49 & 2.62\\ 2.62 & 4.38\end{bmatrix}$	3.23	5(10)

由表1可以看出，当时间序列中掺入10%的异常观测后，传统基于样本均值向量和协方差矩阵的估计结果出现明显偏移。由于这类统计量对异常值高度敏感，参数估计误差在VAR(1)模型拟合过程中被进一步放大，导致模型系数发生显著变化。受此影响，基于残差构造的监控统计量Z_t以及相应的控制上限均发生系统性偏移，从而削弱了控制图对过程异常的识别能力。

再比较污染前后的控制图参数，异常值一进来，上控制限（UCL）就明显抬高，传统VAR控制图在异常数据环境下的稳健性不足，从这里就能看出来。控制限被抬高，等于把报警门槛往上挪，控制图整体跟着变迟钝，过程真的发生偏移时往往不能及时报警，过程失控却未被识别的风险也就上去了。与此同时，异常观测还会干扰对过程结构的判断，把潜在的变化模式盖住，控制图在过程诊断和问题定位中的辅助作用也跟着打折扣。

污染率较高时，还有另一个问题：同一时刻常常有多个变量同时受到异常干扰。但传统VAR控制图采用

$$Z_t=\max_{1\le i\le p}|\ Z_{it}|$$

作为综合监控统计量，报警机制只指向最显著的那个失控变量，多变量同时异常的情况很难被完整反映出来。比如在本次模拟实验中，第35期观测点发生异常时，多个变量都受到了污染影响，可控制图只对其中一个变量给出报警，其余变量的同步失控风险被放过去了。放在实际工业决策中，这就可能造成信息利用不充分，甚至误判，如表2所示。

表2：检测结果（部分）

No.obs	X_1	X_2	$\lvert Z_{it}\rvert$	$\lvert z_{2t}\rvert$	Z_t
34	−0.2971	−0.2537	0.3280	0.3969	0.3969
35	2.6230 (−0.40025)	7.0736* (1.66285)	1.4141	4.1130*	4.1130*
36	0.6955	−0.9947	0.1463	0.7511	0.7511
37	1.4739	0.2576	0.6866	0.2022	0.6866

在多变量过程监控中，若控制图每次报警后均需对所有变量进行逐一排查，将显著增加时间与人工成本，尤其在高维情形下更为突出。因此，传统仅提供整体失控信号的控制图难以满足实际决策需求，亟需一种兼具稳健性且能够准确指示失控变量的多变量控制图方法。

二、稳健向量自回归控制图的构建

（一）分位数加权的计算

在通过考虑每个Z_{it}在其变量分布中的相对重要性（使用分位数作为权重）来更精细地整合多变量的信息。这种方法的核心思想是根据变量值的统计重要性赋予不同的权重，从而可能提高异常检测的灵敏度和特异性。

首先，对于观测序列矩阵的每个数据点，计算其标准化分数Z：

$$Z_{it}=\frac{X_{it}-\mu_i}{\sigma_i}$$

其中，X_{it}是第i个时刻第t个变量的观测值，μ_i是对应变量的均值，σ_i是对应变量的标准差。接着计算其在所有$|Z_{it}|$中的分位数：

$$q_{it}=F_{Z_i}\left(|Z_{it}|\right)$$

其中，F_{Z_i}是第i个变量的Z值的分布函数，由于在实际生产过程中异常值的在总过程中的占比不会超过50%，这里仅考虑变量的异常程度足够大的时候才考虑其对整体的影响，进一步减少噪声的干扰，使控制图对过程中真正的变化更为敏感。根据分位数q_{it}计算权重：

$$\omega_{it}=\begin{cases}0, q_{it}<0.5\\ q_{it}, q_{it}\ge 0.5\end{cases}$$

对于每个时刻的Z_{it}，把它的权重做归一化处理，让每个时刻的总权重恰好为1：

$$\omega_{it}^{'}=\frac{\omega_{it}}{\sum_{t=1}^{p}\omega_{it}}$$

有一种情况需要单独处理：如果某时刻所有ω_{it}都是0（也就是所有q_{it}都小于0.5），那就让每个$\omega_{it}^{'}$直接取$\frac{1}{p}$。

接着算每个时刻的加权Z_t值，权重就用归一化后：

$$Z_t=\sum_{t=1}^{p}\omega_{it}^{'}|Z_{it}|$$

再把优化后的Z_t值的$(1-\alpha)$分位数拿来当上控制限：

$$UCL=F_{Z_t}^{-1}(1-\alpha)$$

其中a是置信水平，$F_{Z_t}^{-1}$是Z_t分布的逆函数。

（二）稳健打点值的选择

设计多变量自相关控制图时，传统做法一般靠样本均值和协方差矩阵来估计过程的中心和散布。本文换一个思路，从仿射不变性和高崩溃点两个角度来挑稳健估计量。仿射不变性确保的是估计结果别受变量线性变换和尺度差异的影响；高崩溃点确保的是在存在一定比例异常值的情况下，估计量还能不能稳住。基于上述准则，本文选了RMCD和RMVE估计量，用于后面稳健控制图的搭建。

RMCD估计量的思路是：先挑出一个协方差行列式最小的子集，再拿这个子集去估计位置和散布。简单来说，RMCD估计量

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

旨在寻找一个具有最小行列式的子样本集，从而最小化协方差矩阵的估计误差，让估计量对异常值不那么敏感。

给定样本数据集$\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n\}$，RMCD 的估计分三步走：

1. 选择子集：从样本数据里挑出一个大小为 h 的子集，其中 $h \le n$，选择的标准是这个子集的协方差矩阵行列式最小。

2. 估计位置：把子集的均值拿来当稳健位置估计，记为：

$$\hat{\mu}_{\text{RMCD}} = \frac{1}{h}\sum_{i=1}^{h}\mathbf{x}_i$$

3. 估计散布：把子集的协方差矩阵拿来当稳健散布估计，记为：

$$\hat{\Sigma}_{\text{RMCD}} = \frac{1}{h-1}\sum_{i=1}^{h}(\mathbf{x}_i-\hat{\mu}_{\text{RMCD}})(\mathbf{x}_i-\hat{\mu}_{\text{RMCD}})^T$$

此外，RMVE 估计量通过最小化覆盖数据的椭球体体积来进行散布的估计。该方法选择一个适当的子集，其余估计步骤和 RMCD 估计量一致，在此不再赘述。

通过计算 RMCD 和 RMVE 的稳健估计量，可得出下面是基于 RMCD 和 RMVE 对应的 Z_t^R：

$$Z_{t-RMCD}^R = \sum_{t=1}^{p}\omega_{it}^{'}Z_{it-RMCD}^R \quad (1)$$

$$Z_{t-RMVE}^R = \sum_{t=1}^{p}\omega_{it}^{'}Z_{it-RMVE}^R \quad (2)$$

以上为稳健控制图中稳健打点值部分的优化步骤。

（三）控制限的优化计算

本文通过蒙特卡洛模拟以获得不同参数下对应的控制限 UCL。从标准多变量正态分布 $MVN(0,I_p)$ 中生成了 n=2000 个大小为 m 的样本，维度为 p。由于 Z_{t-RMCD}^R 和 Z_{t-RMVE}^R 统计量具有不变性，这些限值适用于任何 μ 和 Σ 的值。本文使用崩溃值为 $\gamma=0.50$ 的加权 MCD/MVE 估计量（$\bar{x}_{RMCD}, S_{RMCD}, \bar{x}_{RMVE}$ 和 S_{RMVE}），按照公式（1）（2）计算观测样本数据集中每个观测值的 Z_{t-RMCD}^R 和 Z_{t-RMVE}^R 统计量，并记录每个大小为 m 的数据集中达到的最大值。通过计算 Z_{t-RMCD}^R 和 Z_{t-RMVE}^R 的逆函数的最大值的经验分布，确定了 $(1-\alpha)100\%$ 的分位数。本文使用了由 Torodov 编写的 R 函数“CovMcd()”在“rrcov”包中计算 RMCD/RMVE 估计量。

本文构建了在 γ=0.50 的条件下，样本量 (m=[30(1)50,55(5.100, 110(10)200])，维度 (p=(2,3…10))，以及 (a=(0.05,0.01,0.005)) 时的 Z_{t-RMCD}^R 和 Z_{t-RMVE}^R 的经验分布。

$$f_{p,\alpha,\gamma,m} = a_{1,(p,\alpha,\gamma)} + \frac{a_{2,(p,\alpha,\gamma)}}{m^{a_{3,(p,\alpha,\gamma)}}} \quad (3)$$

表3为不同参数对应下 UCL 控制限拟合的参数估计，为节约空间仅展示常用参数下的估计值。

表3：Z_{t-RMCD}^R 和 Z_{t-RMVE}^R 控制图的模型参数 $a_{1,(p,\alpha,\gamma)}$，$a_{2,(p,\alpha,\gamma)}$ 和 $a_{3,(p,\alpha,\gamma)}$ 的估计

p	a=0.05				a=0.01			
	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	R^2
				Z_{t-RMCD}^R				
2	2.067	15.934	1.128	0.979	2.582	12.543	1.091	0.958
4	2.007	157.032	1.580	0.987	2.413	168.087	1.566	0.978
6	1.901	571.627	1.816	0.990	2.222	656.758	1.806	0.988
8	1.781	533.681	1.700	0.988	2.022	622.341	1.683	0.987
10	1.680	459.048	1.591	0.992	1.880	616.460	1.611	0.990
				Z_{t-RMVE}^R				
2	2.179	34.903	1.290	0.992	2.710	20.057	1.146	0.984
4	2.084	226.930	1.786	0.997	2.519	530.152	2.020	0.992
6	1.971	162.440	1.697	0.997	2.312	223.008	1.759	0.995
8	1.881	162.676	1.618	0.997	2.153	173.112	1.596	0.997
10	1.795	115.033	1.441	0.999	2.028	156.129	1.478	0.999

（四）稳健向量自回归控制图的优化步骤

构建稳健多变量自相关控制图的具体步骤如下：

阶段 1

1. 确定样本量 n、变量数 p 以及置信水平 $1-\alpha$。

2. 以明确的周期间隔收集阶段 I 数据 $\{x_1,\cdots,x_n\}$。

3. 根据样本规模参数选调合适的稳健统计量（下以 RMCD 为例），使用阶段 I 数据计算 RMCD 的估计值 x_{RMCD} 和 S_{RMCD}，并选择崩溃点 $1-\gamma=0.5$。

4. 根据期望的 a 和 p 值，从表格 3 中选择最小二乘法估计参数 $a_{1,(p,\alpha,\gamma)}$，$a_{2,(p,\alpha,\gamma)}$ 和 $a_{3,(p,\alpha,\gamma)}$ 然后使用公式 (3) 计算控制限。

阶段 2

5. 根据公式（3-14，15）计算每个新观测值的 Z_t^R，并将其绘制在已在阶段 I 中（步骤 4）确定的控制限的控制图上。

6. 对控制图进行动态监测，及时查找失控点确定农长变量。如有必要，对过程进行诊断。

通过以上步骤，即可有效地实施稳健多变量自相关控制图来监控过程中的多元数据。

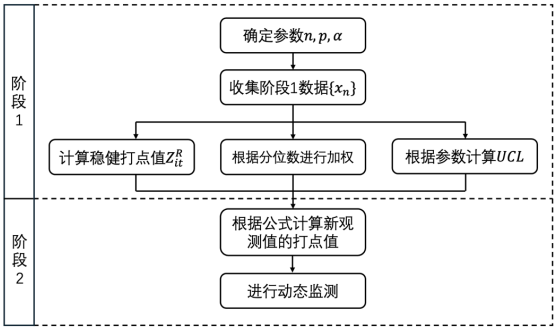


图1：稳健多变量自相关控制图建模流程图

三、模拟与实证分析

（一）模拟分析

为了说明稳健 VAR 控制图在实际生产过程中有良好的检测效力，本文将模拟一个含污染值的 VAR(1) 过程，构建含污染样本的目的是模拟现实世界数据中可能遇到的异常情况，这些异常值可能由不同来源引起，如测量误差、数据录入错误或极端事件等。通过在模拟的 VAR 过程中人为添加污染值，我们可以评估控制图

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

在实际应用中面对异常值时的表现和稳健性。

模拟实验这样安排。先定义基本的 VAR(1) 模型，其中 X_t 表示在时间 t 的多变量观测值向量，Φ 为系数矩阵，ε_t 为误差项：

$$X_t = \Phi X_{t-1} + \varepsilon_t$$

这里假定 ε_t 为具有协方差矩阵 Σ 的多元正态分布，即 $\varepsilon_t \sim N(0,\Sigma)$。

污染比例设为 a，一般取 $0<\alpha<1$，对于发生异常值的时刻 t，样本值为：

$$X_t = X_t + \delta_t$$

其中，δ_t 是叠加到观测值 Y_t 上的污染向量，用来模拟异常情况下的数据点。实际操作中，δ_t 可以从均值（μ_δ）和协方差矩阵（Σ_δ）都与原分布明显不同的多元正态分布中抽取，以体现异常值的特征，即 $\delta_t \sim N(\mu_\delta,\Sigma_\delta)$。这样一来，污染值会让被选中的观测值明显偏离原有数据分布，现实中可能遇到的异常情况也就被还原出来了。

下面展示的模拟演示，均基于样本量 $n=500$，VAR(1) 模型系数 $\Sigma=\begin{bmatrix}0.5 & 0.2\\ 0.1 & 0.4\end{bmatrix}$，均值向量 $\mu=\begin{bmatrix}0\\0\end{bmatrix}$ 的条件进行，结果如图2和表3所示。

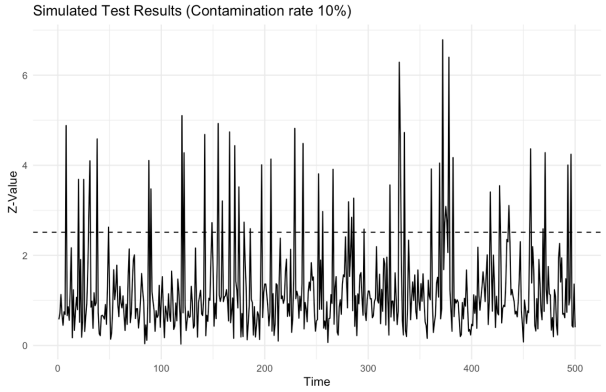


图2：在污染率为10% 条件下的模拟检测效果

表4展示了不同控制图在不同污染率下的检测效果。监控多变量时间序列数据中的异常活动时，3Sigma 控制图靠设定统计阈值来识别偏差，残差 VAR 控制图则检查 VAR 模型的残差，这两种都是生产过程中常见的控制图。下面把它们与本文提出的的稳健 VAR 控制图放在一起对比，看看在不同程度的数据污染下，谁识别异常的能力更强，进而确定复杂数据环境下更管用的监测工具。

表4：不同控制图在不同污染率下的检测效果

污染率	$\alpha=5\%$			$\alpha=15\%$			$\alpha=30\%$		
	3Sigma	残差 VAR	稳健 VAR	3Sigma	残差 VAR	稳健 VAR	3Sigma	残差 VAR	稳健 VAR
报警个数	190	9	39	178	11	72	234	2	112
准确率	67.00%	96.80%	96.80%	78.20%	87.20%	96.60%	70.00%	70.40%	90.80%
查准率	13.16%	100.00%	61.54%	40.44%	100.00%	90.28%	50.00%	100.00%	96.43%
查全率	100.00%	36.00%	96.00%	96.00%	14.67%	86.67%	78.00%	1.33%	72.00%
F1分数	0.23	0.53	0.75	0.57	0.26	0.88	0.61	0.03	0.82

把不同污染率下三种控制图的表现摆在一起看，残差 VAR 控制图对异常值的检测偏保守，查全率因此远低于稳健 VAR 控制图，到了高污染率（$\alpha=30\%$）时，直接检测失效。

传统的3Sigma 在低污染率下能把真实异常值全部揪出来，可误报也跟着上来了，查准率并不好看。污染率一往上走，它的准确率和查准率还在往下掉，这一点也就更明显。查全率是保住了，但实际用起来，频繁误报既浪费资源，也容易让人对报警麻木。

稳健 VAR 控制图的表现就不一样了，污染率怎么变它都比较稳。污染率升上去，它照样能把异常值认出来，报警不多，漏报也少，准确率和查准率因此一直在线。更值得一提的是，污染率往上走，稳健 VAR 控制图的性能并没有明显往下掉，崩溃点和稳健性都撑得住。它的 F1分数在所有污染条件下都保持在高位，说明它在查准和查全之间找到了平衡，这一点在实际应用中很要紧。

接下来，研究把目光转向维度。在中等污染率（α=15%）下逐步加维度，看看控制图面对更复杂的数据结构还稳不稳，见表5。

表5：污染率 $\alpha=15\%$ 下不同维度的检测效果

	p=2	p=3	p=4	p=5	p=6
报警个数	72	75	60	62	66
准确率	96.60%	99.20%	96.60%	97.00%	97.00%
查准率	90.28%	97.33%	98.33%	98.39%	95.45%
查全率	86.67%	97.33%	78.66%	81.33%	84.00%
F1分数	0.88	0.97	0.87	0.89	0.89

表5显示，在中等污染率（$\alpha=15\%$）下，稳健 VAR 控制图的性能随维度增加基本能稳住。报警个数在不同维度 (p=2至 p=6) 之间有些起伏，但整体还在相对稳定的范围内，没有出现维度越高报警越多或越少的趋势。控制图在标识潜在异常时的敏感度，并没有因为维度增加而被明显拉低。

更值得一提的是，各维度下准确率都超过了95%，这个数字说明控制图能把正常点和异常点分得比较清楚，即便数据里掺了污染，也还能保持可靠。查准率和查全率同样稳当，没有随着维度增加而明显下滑，说明控制图在不牺牲太多敏感性的前提下，仍能维持较高精确度。所有维度中 F1分数都在0.87以上，这些结果说明控制图在各个维度上都能保持住自己的检测性能。

为了进一步看稳健 VAR 控制图在不同维度下的检测情况，下面用 F1分数来综合刻画它的检测效力，对不同维度做多次模拟实验，并画出小提琴图，见图3：

从图3可以看出，维度增加后，稳健 VAR 控制图的 F1分数中位数在大多数维度上都维持在较高水平，分布也大多比较集中，说明高维情况下该控制图的表现相对稳定。不过，$p=5,6$ 时出现了少量的离群点，F1分数的整体分布范围也更宽。这一点值得留意：维度再往上走，模型复杂度跟着涨，稳健估计量本身也会开始吃力，性能波动因此被放大。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

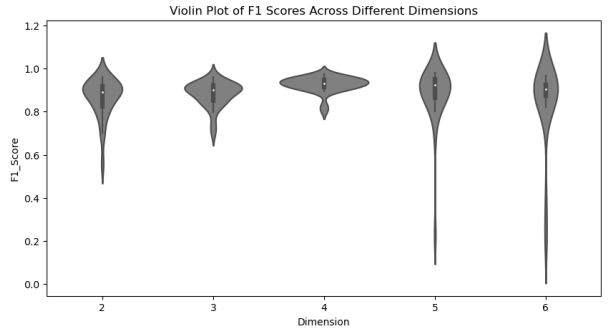


图3：不同维度下 F1 值的小提琴图

综上，不同维度和污染率下的实验结果表明，稳健 VAR 控制图面对复杂数据时稳得住。即便污染率高达30%，准确率仍保持在90% 以上，报警个数也只是适度增加，说明它对异常的识别既敏感又准确。F1 分数在不同维度上的稳定性说明，该控制图能在查准率和查全率之间取得平衡，检测性能也因此较为可靠。

（二）实证分析

模拟完成，接下来把方法放到真实数据测试一下。

本研究使用了广州市2022年的每日气象数据，重点分析空气质量相关的六个污染物指标，包括 PM2.5、PM10、SO2、NO2、O3（8小时均值）和 CO。数据集共365条记录，每条代表一天的污染物浓度和相应 AQI 级别。AQI 级别是评估空气质量的标准，也是衡量空气污染程度和对公众健康影响的关键指标。在该数据集中，AQI 级别大于等于3意味着空气质量较差，需要做异常值分析，找出可能由环境或人为因素导致的空气质量异常。

应用稳健 VAR 控制图方法后，我们在广州市2022年的气象数据中识别出若干异常值。这些异常值对应的日子主要集中在特定时间内，可能与季节变化、人为活动增加或特殊环境事件有关，见图4。

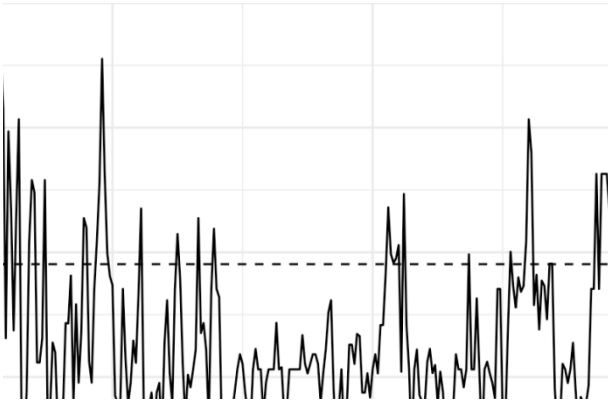


图4：广州市2022年空气污染物稳健控制图

为验证异常值检测的准确性，我们把检测到的异常值与 AQI 级别做了对比。当检测到的异常值对应的 AQI 级别大于等于3（轻度污染）时，就认为这次异常值检测是准确的。

图5是本实证结果的前100期可视化的对比图，其中浅灰色色块代表正常值，未检测出异常情况，黑色为发生异常的变量，颜色的深浅代表着影响程度的强弱。不难发现，在处理多变量问题时，由于数据的复杂度增加，传统的 VAR 控制图已经不能准确的识别失控过程，使得检测结果可信度降低，反观稳健 VAR 控制图，在处理高维数据时依然能表现稳定的检测效力；更值得注意的是，稳健 VAR 控制图还提供了不同变量在异常时刻的影响程度可视化，根据各时刻加权 Z_t 的系数可以绘制以上影响程度色块图，这里不但指出了异常时刻中影响程度最大的变量，还一并将发生变异的其余变量通过不同深浅的颜色标注出来，为决策者提供了准确的异常变量信息，以便决策者后续的判断和调整。

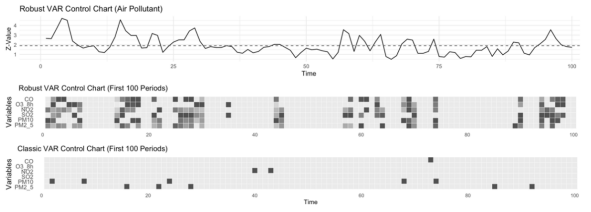


图5：传统 VAR 控制图和稳健 VAR 控制图检测结果（部分）的可视化

根据广州市2022年的大气数据，基于本文提出的稳健 VAR 控制图检测出异常值检测的准确率为83.01%。这表明，稳健 VAR 控制图方法在广州市2022年的气象数据中表现出了较高的准确性，能够有效地识别空气质量的异常变化。

四、结束语

在本研究中，我们对面临数据污染时多变量时间序列监控的挑战进行了系统的探索，详细评估了稳健 VAR 控制图的性能。模拟实验展示了在不同污染率下，稳健 VAR 控制图相较于传统 3Sigma 和残差 VAR 控制图，无论在低还是高污染水平，均能有效地识别异常值。

特别是在中等至高污染率下，稳健 VAR 控制图展现了其优越的稳定性和较高的崩溃点，整体准确率高达90% 以上。其 F1 分数普遍高于0.8，彰显了即使在数据复杂性升高的情形下，该方法依然能够提供精准的警报，最小化误判，并且在高维数据上表现出稳定的检测效力。并且可以为决策者提供发生异常时刻各影响变量的程度可视化，能够帮助决策者找到并分析异常值的相关影响变量，提升监测精度的同时提高了检测效率避免了不必要的浪费。

综上，通过在多种污染场景下的强表现，本研究提出的稳健 VAR 控制图证明了自己在多变量时间序列数据监控领域的实用性和可靠性。其稳健性和高准确率使之成为适用于当今数据密集型产业的理想工具，尤其是在存在数据质量问题的实际应用中。未来的研究将建立在这一基础之上，进一步扩展控制图在多元数据环境中的应用，以不断增强其在过程监控中的效率和效果。

参考文献

[1]Centofanti, Fabio, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo, and Simone Vantini. Functional Regression Control Chart[J]. Technometrics,2021,63(3):281 - 94.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

[2]Elisa C,E. R L. Robust multivariate control chart based on shrinkage for individual observations[J].Journal of Quality Technology,2022,54(4):415-440.

[3]Bayati M ,Nasrollahzadeh S ,Moghadam B M . Robust Monitoring of Multivariate Location Using a New Nonparametric Depth - Based Control Chart[J].Journal of Applied Mathematics,2026,2026(1):1-13.

[4]Pulido D J A ,Franco C E A ,G ó mez T G V . Nonparametric multivariate control chart with guaranteed in-control performance and cautious learning[J].Computers &Industrial Engineering,2026,218:112124.

[5] 王志坚 , 郑健松 . 稳健 Shewhart 控制图的构建及应用 [J]. 统计与决策 ,2018,34 (24): 39-42.

[6]AliH T ,SedeeqS B ,SalehM D , et al. Robust multivariate quality control charts for enhanced variability monitoring[J].Quality and Reliability Engineering Internation al,2023,40(3):1369-1381.

[7]S V ,Sebastian N ,TA S . Robust Bivariate Control Chart for Process Monitoring with Rational Subgroups[J].Quality and Reliability Engineering Internation al,2025,41(7):3232-3243.

[8] 薛丽 , 李聪凯 , 贾元忠 , 等 .VSI 多变量自相关过程 VAR 控制图经济设计 [J]. 数理统计与管理 ,2023,42(04):626-636.

[9]Capezza C, Centofanti F, Lepore A, etal. Functional regression control chart for monitoring ship CO2_2 emissions[J].Quality and Reliability Engineering Internation al,2022,38(3):1519-1537.

[10]Will M ,Mark R . SampleQC: robust multivariate, multi-cell type, multi-sample quality control for single-cell data.[J].Genome biology,2023,24(1):1-22.

[11] 雷彬文 , 龚立雄 . 多变量控制图及过程能力指数在生产中的应用 [J]. 湖北工业大学学报 ,2023,38 (01):5-10.

[12]Kalgonda, A. A., and S. R. Kulkarni. Multivariate Quality Control Chart for Autocorrelated Processes[J].Journal of Applied Statistics, 2004,31 (3):317 - 327.

[13]Hawkins, Douglas M. Multivariate Quality Control Based on Regression-Adiusted Variables[J].Technometrics,1991,33(1):61 - 75.

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

何为企业数字技术产出的创新鸿沟？——基于多种可解释模型的比较分析

刘佳仪[1]，陈丽宇[2]，黄颖[1]，程从华[1*]
1. 肇庆学院 数学与统计学院，广东 肇庆 526000
2. 肇庆学院 经济与管理学院，广东 肇庆 526000
DOI:10.61369/ASDS.2026100009

摘　　要： 中国企业数字技术创新产出在数字经济不断推进的大环境下呈现出很明显的“零膨胀”和“长尾”双重分布状态，也就是很多企业长期没有专利产出，少数企业却占据了大部分创新成果，公司治理结构在这里面也起到一定作用，本文把这种“创新鸿沟”拆成“门槛鸿沟”和“差异鸿沟”两个前后递进的层次，使用了2000到2024年中国A股上市公司45848条年度数据，搭了一个四层递进式模型框架，第一层用可解释模型打底，第二层用黑箱集成模型做主分析，第三层用SHAP做事后解释，第四层用滚动窗口完成动态验证，本文系统考察了线性回归、Lasso、梯度提升、随机森林、XGBoost、LightGBM这六类模型在样本内外的拟合好坏和误差指标，还通过10折交叉验证确认了结果是否稳健，之后本文结合EBM分类模型的形函数和SHAP依赖图，捕捉变量之间的非线性影响和交互效应，最后本文以5年滚动窗口画出2011到2021年变量重要性变化图，结果显示：（1）推动力量已经从“政策输血”全面转向“内部人力造血”，政府补助的重要性排名从第2位下降到第14位，研发人员数量则上升到首位；（2）监管层总人数有明显的门槛作用，超过临界点后，它会和研发投入形成正向的“放大器”协同；（3）机构投资持股比例呈倒U型，和研发人员数量之间有一个“最优配合区间”；（4）第一股东持股比例、高管货币激励、两职合一、IT背景标签这些传统治理变量持续失效。本文的建模框架为处理具有门槛作用和企业之间差异很大的创新数据提供了可以重复使用的方法范例，实证结论也为优化公司治理结构和精准设计科技政策提供了数据驱动的决策参考。

关 键 词： 数字技术创新；创新鸿沟；可解释模型；多模型比较

What Constitutes the Innovation Gap in Digital Technology Output of Enterprises? A Comparative Analysis Based on Multiple Interpretable Models

Liu Jiayi[1], Chen Liyu[2], Huang Ying[1], Cheng Conghua[1*]
1.School of Mathematics and Statistics, Zhaoqing University, Zhaoqing, Guangdong 526000
2.School of Economics and Management, Zhaoqing University, Zhaoqing, Guangdong 526000

Abstract: The innovative output of Chinese enterprises in the digital technology sector exhibits a distinct dual distribution pattern of "zero expansion" and "long-tail" within the broader context of the rapidly advancing digital economy—namely, many enterprises have historically produced no patent outputs, while a small number of enterprises account for the majority of innovative achievements; corporate governance structures also play a role in this phenomenon. This paper decomposes this "innovation gap" into two sequentially progressive layers: the "threshold gap" and the "differential gap." Utilizing 45,848 annual data points from Chinese A-share listed companies spanning the period 2000–2024, the study constructs a four-tiered progressive model framework. The first layer employs an interpretable model as the foundation; the second layer uses a black-box integration model for primary analysis; the third layer applies SHAP for post-hoc interpretation; and the fourth layer employs a rolling window approach for dynamic validation. This paper systematically examines the fit quality and error metrics of six types of models—linear regression, Lasso, gradient boosting, random forest, XGBoost, and LightGBM—both within and outside the sample, and confirms the robustness of the

基金项目：本研究得到广东省自然科学基金项目（编号：2024A1515010983）；广东省研究生教育创新计划项目（编号:2025JGXM_181）；肇庆学院专业学位研究生课程案例库项目（No.YBANLK2025_09）；肇庆学院第四批创新科研团队项目（编号:TD202410）的资助。
作者简介：
刘佳仪，肇庆学院数学与统计学院，硕士研究生在读，研究方向为大数据分析；
陈丽宇，肇庆学院经济与管理学院，硕士研究生在读，研究方向为数字贸易；
黄颖，肇庆学院数学与统计学院，本科在读；
通讯作者：程从华，肇庆学院数学与统计学院，硕士生导师，研究方向为生存分析。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

findings through 10-fold cross-validation. Subsequently, by leveraging the shape function of the EBM classification model and the SHAP dependency graph, the study captures nonlinear influences and interaction effects among variables. Finally, using a five-year rolling window, the paper plots the evolution of variable importance from 2011 to 2021. The results indicate that: (1) The driving force has fully shifted from "policy-driven external support" to "internal human capital development"; the importance ranking of government subsidies has declined from second place to 14th place, while the number of R&D personnel has risen to the top position; (2) The total number of regulatory personnel exerts a distinct threshold effect; once this threshold is exceeded, it interacts in a positive "amplification" manner with R&D investment; (3) The institutional investment shareholding ratio follows an inverted U-shaped curve, with an "optimal coordination range" existing between this ratio and the number of R&D personnel; (4) Traditional corporate governance variables—such as the shareholding ratio of the largest shareholder, monetary incentives for senior executives, the integration of dual roles, and IT background indicators—continue to prove ineffective. This paper's modeling framework provides a reusable methodological example for analyzing innovation-related data characterized by threshold effects and significant heterogeneity across firms; its empirical findings also offer a data-driven decision-making reference for optimizing corporate governance structures and precisely designing science and technology policies.

Keywords : **digital technology innovation; innovation gap; interpretable models; multi-model comparison**

引言

在数字经济时代下，数字技术创新能力已经成为企业核心竞争力的一个重要方面，国家“十五五”规划纲要明确要加快数字中国建设，推动数字技术和实体经济深度融合，这样一来数字技术产出的重要性就被提升到了国家战略层面，但是在实际经济活动里，不同企业的数字技术产出分化很明显，有的企业依靠持续的研发投入和较好的治理机制实现了技术上的跃升，还有大量企业却一直处在创新产出低水平徘徊甚至完全没有产出的状态，这种企业间的“创新鸿沟”现象，已经成为制约数字经济高质量发展的一个主要障碍，本文将创新鸿沟分成两个前后递进的层次，第一个层次是“门槛鸿沟”，也就是企业能不能实现“从0到1”的数字技术突破，第二个层次是“差异鸿沟”，也就是企业跨过门槛之后，在创新产出的数量和质量上继续出现分化，这两个层次共同构成了企业数字技术产出差异的结构性根源，所以有必要从公司内部治理的角度去深入分析它的形成原因和破解路径。

针对技术创新驱动因素，现有文献从人力资本、公司治理、政策支持等多个角度进行了讨论，大多数研究采用的是传统线性回归模型，对于变量之间复杂的非线性关系缺乏有效描述，股权集中程度、研发人员占比、薪酬差距等因素的非线性作用形式关注不足，同时不同所有制企业创新驱动路径的系统比较和动态演变趋势同样缺少考察；因此本文选取2000—2024年中国A股上市公司作为研究对象，综合使用内部解释模型和外部解释方法SHAP，对公司结构因素驱动发明专利产出的机制以及其中的非线性特征进行系统考察，并揭示企业数字技术产出差异背后的结构性原因，并为优化公司治理结构、完善科技政策设计提供理论依据和实证参考。

已有研究指出，企业开展数字化转型能够明显促进技术创新产出[1]，并且可以通过企业创新渠道推动高质量发展[2]；数字技术应用可以优化资源配置效率、降低信息不对称程度，还能增强技术吸收和知识整合能力，从而提升创新的数量和质量[3]；进一步来看，数字技术应用在技术溢出影响技术创新产出的过程中发挥调节作用[4]。此外，数字化转型可能通过多个中介机制影响企业财务绩效，这种影响逻辑和创新产出效应需要进行区分[5]，政府补助在这一过程中也发挥重要的调节作用[6]。但是现有研究大多关注数字技术的平均效应，对公司治理机制如何影响数字技术创新的效率和方向缺少深入讨论。

“创新鸿沟”这一概念是从数字鸿沟理论延伸出来的，是指企业在创新产出和能力方面出现的结构性分化[7]；按照“接入—应用—转化”的分析框架，企业在数字技术获取、使用以及价值转化这些环节上存在鸿沟，这些鸿沟会造成创新能力的非连续性差距[8]，而组织惯性、路径依赖以及动态能力方面的不同则是导致这一鸿沟的重要原因[9]；现有研究大多集中在数字技术接入层面的不平等问题上，对于企业内部结构性因素如何塑造创新鸿沟缺少系统检验。

数字技术产出和创新鸿沟之间存在的双向关联，数字金融发展能够通过缓解融资约束来缩小企业间的创新差距，但是这个过程会受

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

到数字鸿沟的削弱效应影响[10]；科技金融政策和数字并购则可以帮助企业跨越创新能力门槛，进而实现技术赶超[11,12]。

已有文献从数字经济政策、公共数据开放以及数字技术应用等视角进行了研究，结果发现数字经济政策可以促进传统企业进行跨界数字创新[13]，公共数据开放能够带来技术创新效应[14]，数字技术应用则有助于企业开展合作创新[15]；总体来看，现有研究仍然大多关注平均水平效应，对于非线性作用、门槛效应以及不同所有制情况下的差异化表现还缺少足够的关注。

基于上述文献综述，可以发现现有研究存在以下不足：一是多依赖线性模型，难以刻画变量间的非线性关系；二是对股权集中度、研发人员占比等因素的非线性作用形态关注不足；三是缺乏对不同所有制企业创新驱动路径的系统比较。本文拟采用多模型对比与可解释框架弥补上述研究缺口，并参考吴学福等人对零膨胀计数型数据进行的 Hurdle 模型的分析框架[16]，对当前问题分两阶段考察创新鸿沟：第一阶段聚焦“门槛鸿沟”，采用 EBM 内在可解释模型识别企业跨越数字技术创新门槛的关键结构条件；第二阶段聚焦“差异鸿沟”，综合选取线性模型、集成模型、树模型三类模型，各取两个代表模型进行评估指标的比较，选取最佳模型，结合 SHAP 事后解释方法比较各方法来揭示跨越门槛后企业间创新产出差异的驱动机制。

基于上述研究设计与实证分析，本文的创新点主要体现在以下三个方面；首先，在研究视角上，本文突破了传统线性回归模型只关注平均效应的局限，针对企业创新数据普遍存在的“零膨胀”与“长尾分布”特征，构建了分阶段建模框架；围绕创新鸿沟的两个核心问题，一个是企业如何跨越数字技术创新门槛，另一个是企业间创新产出差异的驱动因素，本文分别考察了从“无专利”到“有专利”和从“低产出”到“高产出”这两个阶段的驱动逻辑，揭示了其中明显的异质性。

其次，在模型构建上，本文构建了“内在可解释模型基准参照—黑箱集成模型主分析—SHAP 事后解释机制挖掘”的递进式建模管线；该管线先利用 EBM 二分类功能来识别跨越创新门槛的关键条件，再通过三类六种模型的纵向比较确定最优预测模型，接着借助 SHAP 依赖图捕捉变量之间的非线性形态和二阶交互效应，最后采用时间滚动窗口来检验模型结论的动态稳健性；这一设计从数据结构特征出发逆向推导模型需求，为处理具有门槛效应和强异质性的企业创新数据提供了可以重复使用的分析范式。

再次，在动态分析上，本文首次把时间滚动窗口和 SHAP 重要性排序结合起来，绘制了2011—2021年变量重要性动态演变图谱，从时间顺序角度考察了各个解释变量解释力的变化；这种方法不仅丰富了创新驱动因素的动态证据，也为统计建模中模型结论的时间稳定性问题提供了一个直接的实证检验方案。

一、变量选取与建模数据基础

（一）数据来源

本文数据来源于国泰安（CSMAR）数据库，包含2000-2024年中国 A 股上市公司面板数据和高管数据。合并清晰过程中以“股票代码—年度”为匹配对象。样本筛选过程中，剔除了金融类上市公司、ST 及 *ST 类企业以及主要变量存在严重缺失的观测样本，最终用于本文建模分析的数据集包含4887个企业共45848条年度观测值。以上数据皆公开，数据来源真实、可验证，能够有效支撑本文的实证分析。

（二）特征选取

为系统考察公司结构背景对数字技术产出的影响机制，本文分别从以下维度选取变量：企业特征、财务状况、成长能力、创新投入、行业环境、股权结构、公司治理、管理层特征、创新人才、政策扶持及数字化转型。选取的变量分别作为控制变量与解释变量。

1. 被解释变量

被解释变量采用上市公司‘发明专利授权量’，其记录了企业当年发明专利授权量的个数，能够体现企业数字技术产出量，因此使用该指标作为被解释变量用于研究企业之间产生差异的鸿沟何在。‘专利产出’是通过‘发明专利授权量’构造的指标，是一个二分类指标，0表示企业在当年无产出，1则有，这一指标体现了数字技术产出的门槛，即企业如何从无到有。

2. 解释变量

解释变量从以下五个方面选取：第一，股权结构方面选取第一股东持股比衡量股权集中度、股权性质、机构投资持股比，以考察股权结构在公司结构背景系统下的作用；第二，公司治理方面选取董事会规模、独立董事人数、监事会规模、监管层总人数及持股数量、两职合一，以考察治理结构对创新决策的影响；第三，管理层特征方面选取高管薪酬总额与高管前三薪酬总额、高管教育背景与信息技术背景，这个方面用于考察管理团队的知识资本与技术经验；第四，创新人才方面选取研发人员数量衡量绝对投入规模、研发人员占比考察团队结构配置的边际影响；第五，政策扶持方面选取政府补助金额衡量政策支持力度。

3. 控制变量

为缓解遗漏变量偏误，选取企业年龄与企业规模控制企业特征，选取总资产净利润率、资产负债率、固定资产占比控制财务状况，选取托宾 Q 值与营业收入增长率控制成长能力，选取研发投入总额与强度控制创新投入水平，引入行业集中度控制行业竞争格局。

二、数据清洗

为保障后续实证回归结果的稳健性与结论解释力，本文基于原始企业面板数据，遵循”变量统一—缺失修复—异常管控—尺度变换”的逻辑框架，开展多源数据系统性清洗与标准化预处

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

理。本文原始数据取自企业财务报表、公司治理数据库及创新投入专题数据库，以“股票代码—年度”为匹配标识完成多表对接与样本整合。

（一）数据清洗

1. 多源数据整合与主表构建

本文研究涉及财务、公司治理、创新特征多维度数据表，为统一样本统计口径，以企业年度财务数据为主表，选取“股票代码、年份”作为联合主键，对公司治理表、创新特征表依次采用“左连接（Left Join）”方式合并。该方式可最大限度保留主表企业年度观测样本，同步补充治理、创新维度变量信息，规避内连接、交叉连接造成的样本大量损耗问题。

2. 主营业务文本分类编码

原始主营业务为非结构化文本数据，存在行业表述繁杂、口径不统一、描述冗余等问题。本文采用“关键词匹配 + 文本筛选”方法对主营业务进行标准化分类编码，聚焦软件开发、云计算、大数据、人工智能、信息服务、电子商务、集成电路等数字技术业务关键词，构建企业数字业务虚拟变量：企业主营业务涉足数字技术领域赋值为1，未涉及则赋值为0，实现文本信息可量化建模。

3. 股权性质变量统一赋值

赋值编码规则如下：国有企业、央企、地方国资控股等公有制企业赋值为1；民营企业、外资企业及混合所有制非国资控股企业等非公有制企业赋值为0，统一变量内涵与实证解释口径。

4. 缺失值分层初步处理

通过数据信息核验与变量缺失率统计，分层开展缺失值处理：专利数量、研发投入、政府补贴、技术投入等投入计数类变量，遵循“未披露即未发生”的经济逻辑，缺失值统一赋值为0；股权持股比例、持股金额等核心治理变量，缺失将严重影响实证有效性，直接剔除对应年度观测样本，剔除后样本量仍满足计量建模基本要求。

5. 重复值剔除与格式标准化

统一股票代码、年份字段数据格式与字段类型；以“股票代码 + 年份”为唯一识别维度开展重复值去重，保留字段信息最完整的观测记录，杜绝重复样本干扰回归系数估计结果。

（二）数据预处理

建模前先完成变量定义与描述性统计，刻画变量分布形态、离散程度及潜在异常特征，在此基础上开展精细化预处理。

1. 缺失值精细插补修复

本文采用“逻辑回归插补法”进行缺失值拟合填补，这能有效捕捉到变量间统计关联关系，从而更精准地估算缺失数值，相比传统均值填补更能保留原始数据分布结构。

2. 异常值识别与缩尾处理

本文利用箱线图同时结合“四分位距”法则，识别与判断异常值，并发现多数连续型变量存在极端的异常值。为避免异常值对参数估计的影响，本文对所有连续变量进行1%–99% 缩尾处理。

3. 连续变量 Z–score 标准化

资产规模、盈利能力、研发强度等连续变量量纲差异显著，为消除量纲干扰、提升模型收敛速度与系数可比性，对全部连续变量进行 Z–score 标准化：

$$Z = \frac{X - \mu}{\sigma}$$

式中：μ 为变量均值，σ 为变量标准差。标准化后变量满足均值为0、标准差为1。

4. 计数型变量对数变换

专利申请量、政府补贴次数、投资事件数等计数变量呈明显右偏长尾分布，本文对非零观测值进行自然对数变换：通过对数变换压缩长尾、弱化极端大值干扰，改善变量正态性；取值为0的样本保留原值，坚守变量经济含义不变。

经过上述数据预处理操作后，最终得到口径统一、结构完整、统计特征稳定的企业面板数据集。预处理显著优化了数据的完整性、一致性与分布合理性，为后续基准回归、模型比较、稳健性检验及拓展分析提供坚实的数据支撑。

（三）变量说明及描述性统计

表1与表2分别报告了控制变量、被解释变量与解释变量的描述性统计结果。被解释变量‘发明专利授权量’的均值为13.90，但中位数为0，且标准差高达153.05，表明大量企业都没有发明专利产出，且有产出的企业之间差距较大，呈现严重的“零膨胀”与“长尾”现象。若此时采用传统回归方法可能出现严重的过拟合，因此进行两阶段建模与多模型比较是有必要的。另外发明专利申请量的均值为89.12、标准差为786.22，进一步验证了创新产出的极端不均衡性。

在解释变量方面，研发人员数量的均值为310.65，但中位数为0，说明大多企业尚未配备专职研发人员，与发明专利授权量呈现相似的现象。这一现象预示着企业在结构治理上跨越创新门槛很可能收到研发人员规模的影响。监管层总人数的均值为16.49、中位数为16、标准差为3.96，分布相对集中，但其对创新的影响是否呈现非线性形态有待模型验证。机构投资持股比的均值为45.56%、标准差为24.52%，覆盖从极低到极高持股的完整区间。政府补助的均值与中位数差异悬殊，表明补助资金高度集中于少数企业。

表1：解释变量与被解释变量的定义和描述性统计

变量名	定义	样本量	均值	中位数	标准差
解释变量					
第一股东持股比	第一大股东持股数量 / 总股数，衡量股权集中度与大股东控制程度	43157	34.9487	32.55	15.3895
股权性质	虚拟变量，0表示非公有制，1表示公有制，控制产权性质差异股权性质。	45848	0.3785	0	0.485
董事会规模	董事会成员总人数，衡量董事会规模与治理结构	45848	8.7308	9	1.9057

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

独立董事人数	独立董事人数，衡量董事会独立性与监督效率	45848	3.1667	3	0.7284
研发人员数量	研发人员总人数，衡量创新人力资源投入	45848	310.6479	0	1753.455
研发人员占比	研发人员 / 员工总数，衡量创新人才结构	45848	7.9729	0	12.7944
监事会规模	监事会成员人数，衡量内部监督机制强度	45848	3.6574	3	1.2427
高管薪酬总额	高管年度薪酬总额，衡量管理层激励水平	45848	3936863.21	2520750	5626588.61
政府补助	当年获得的政府补助金额，衡量政策支持力度	45848	23408090.33	1091207.56	291331341
机构投资持股比	机构投资者持股比例，反映机构投资者对公司的关注和认可程度	45384	45.5592	47.7233	24.5245
监管层总人数	公司监管层的总人数，反映公司治理结构的规模	45848	16.485	16	3.9593
监管层持股数量	公司监管人持有的公司股份总数	44199	49836993.82	689931	127085381.6
高管前三薪酬总额	公司年度内薪酬最高的三位高级管理人员的薪酬总和（包括工资、奖金、福利等）	45848	2221881.13	1541000	2858113.65
两职合一	董事长与总经理是否为同一人；1= 是，0= 否，衡量治理结构	45848	0.0822	0	0.2747
高管教育背景	董监高是否具有研究生教育背景；1= 是，0= 否	45848	0.2914	0	0.4544
主营业务	主营业务数字化转型程度，是否有数字技术类业务，0表示不是，1表示是	45848	0.1919	0	0.3938
高管信息技术背景	虚拟变量，CEO 具有技术从业背景时为1，否则为0	45848	0.0384	0	0.19
被解释变量					
发明专利授权量	发明专利授权数量，衡量高质量创新产出	45848	13.9036	0	153.0518
专利产出	企业是否有专利产出，0表示不是，1表示是	45848	0.3364	0	0.4725

控制变量大多呈现非正态分布特征，总资产净利润率的均值为0.56%，标准差109.86%、托宾 Q 值 D 与营业收入增长率的均值与中位数之间同样存在较大偏离，同样暗示我们样本内存在极端盈利差异。综合数据特征，本研究在模型选取上需要能够处理“有 / 无”二分类门槛问题的分类模型、能够捕捉非线性关系与高阶交互作用的回归模型、具备可解释性以揭示变量作用方向的参考模型。

表2：控制变量的定义和描述性统计

变量名	定义	样本量	均值	中位数	标准差
企业年龄	企业成立至观测年度的年限，反映企业生命周期与经营成熟度	45843	16.51283	16	6.279152499
总资产	期末总资产账面价值（元），衡量企业整体资产规模	45848	54401166807	2931758243	746476150138
企业规模	总资产的自然对数，消除规模异质性，标准控制变量	45848	22.0193	21.7989	1.5191
总资产净利润率	净利润 / 平均总资产，衡量资产利用效率与整体盈利能力	45848	0.5584	0.0364	109.8558
托宾 Q 值 D	资产重置成本法托宾 Q，衡量企业成长性与投资机会	44133	4.2629	2.1289	243.5202
营业收入增长率	营业收入同比增长率，衡量企业成长能力	44916	-3.7961	0.1011	800.4029
资产负债率	负债总额 / 资产总额，衡量企业长期偿债能力与财务杠杆水平	45848	0.155	0	1.1891
固定资产占比	固定资产净额 / 资产总额，衡量企业资产结构与重资产程度	45848	0.0686	0	0.1388
行业集中度 D’	赫芬达尔指数 D，基于营业收入衡量行业竞争程度	45848	0.0334	0	0.0878
研发投入总额	年度研发投入资金总额，衡量创新投入规模	45848	135552502.5	20741293.02	748212118

三、多模型构建与性能比较

（一）模型说明

1. 模型评估指标与选取

为全面评价各模型的预测性能，本案例采用样本内拟合优度、样本外拟合优度、均方误差、平均绝对误差四个评估指标。样本内拟合优度 R_{in}^2 和样本外拟合优度 R_{out}^2 用于衡量模型对训练数据和测试数据的泛化能力，指标计算方法详见公式（2）和（3）。

$$R_{in}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

$$R_{out}^2 = 1 - \frac{\sum_{i \in Test}(y_i - \hat{y}_i)^2}{\sum_{i \in Test}(y_i - \bar{y}_{Train})^2} \tag{3}$$

均方误差 MSE 可以反映预测值与真实值之间偏差的平方的期望，然而均方误差对异常值较为敏感，为模型预测提供更多稳健性指标参考，本研究还使用了可以反映预测值与真实值之间绝对偏差平均水平的平均绝对误差 MAE。指标计算方法详见公式（4）和（5）。

$$MSE = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{5}$$

另外，本文研究聚焦于公司结构背景数据对数字技术创新产出的影响，需要在保证模型预测效果的同时，清晰识别关键影响变量，因此采用可解释机器学习思路进行模型选取。可解释机器学习（Explainable Machine Learning, XML）的目标是在保证模型预测性能的同时，提高模型决策过程的透明度，根据解释方式不同可分为内在解释和事后解释 [16]。

内在可解释（Intreinsic Interpretability）指模型自身结构简单，使用者能够直观看到模型内部结构。常见内在可解释模型

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

有：线性回归、逻辑回归、决策树等。

事后可解释（Post-hoc Interpretability）指模型训练之后，使用一定方法增强模型的可解释性，挖掘模型学习到的信息。常用事后可解释的方法有：LIME 事后可解释方法、SHAP 事后可解释方法[17]。

图1所示的是可解释机器学习中模型精度和模型可解释性的关系，由香港大学张爱军教授提出。横轴代表模型的可解释性，越往正方向模型可解释性越高；纵轴代表模型的精度，越往正方向模型精度越高。可以看到内在可解释的一些模型精度相对低，但是模型可解释性高，需要使用到事后可解释的模型精度高，可解释性偏弱。

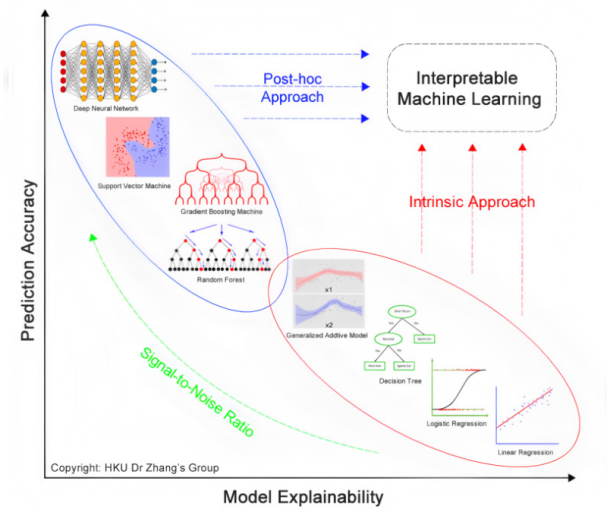


图1：可解释机器学习：模型精度和模型可解释性的关系

鉴于此，本案例首先选取适用于二分类变量预测的内解释模型 EBM 模型，来探究企业跨越数字技术创新门槛的因素。再通过多模型比较的方式确定进行回归预测与事后可解释的模型选取，并以此来探究企业数字技术创新的差异何在。

2.EBM 模型

EBM（可解释提升机，Explainable Boosting Machine），是一种内在可解释机器学习模型，其初衷是在保证预测性能接近复杂黑箱模型的同时，提供全局与局部的透明解释[18]。EBM 结合了广义加性模型和梯度提升树，能够提供特征的单一变量与二阶交互作用的变量重要性排序，适用于二分类目标的预测。

其原理是在给定一个分类任务数据，训练样本为 $(D_i=(x_i,y_i), i=1,\cdots,l)$，其中 (x_i) 为输入样本，(y_i) 为样本的类别值，基础形式为：

$$g\left(E\left(Y|X\right)\right)=\sum_{i=1}^{N}f_i\left(X_i\right) \quad (6)$$

为进一步捕捉特征间的非线性交互效应，EBM 引入二阶交互项，将模型扩展为：

$$g\left(E\left(Y|X\right)\right)=\sum_{i=1}^{N}f_i\left(X_i\right)+\sum_{i,j}f_{ij}\left(X_i,X_j\right) \quad (7)$$

其中 $g(\cdot)$ 为链接函数（回归任务为恒等函数，分类任务为 logit 函数），$f_i(X_i)$ 为特征 X_i 的主效应形函数，由梯度提升树拟合得到；$f_{ij}(X_i,X_j)$ 为特征 X_i 与 X_j 的二阶交互效应形函数。

EBM 的训练先通过梯度提升后向拟合训练所有特征的主效应形函数，再基于模型残差，利用 FAST 算法快速筛选并加入 Top-K 二阶交互项。特征重要性以形函数的标准差衡量：$\mathrm{std}(f_u)=\sqrt{E(f_u^2)}$。

EBM 主要的优缺点见下表。

表3：可解释提升机（EMB）主要优缺点

优点	缺点
1. 精度接近复杂黑箱模型，且具备天然可解释性	1. 仅支持二阶交互，难捕高阶特征
2. 支持全局与局部双重解释	2. 高维特征下计算复杂度上升
3. 自动识别筛选二阶特征交互项，降低计算成本	3. 形函数受限，非线性拟合弱于深度模型
4. 数据预处理成本低，无需复杂归一化分箱操作	4. 相比单一决策树，训练耗时略高

3.Linear 与 lasoo

线性回归（Linear Regression）是最基础的参数化回归模型，其核心原理是假设因变量与自变量之间服从线性关系，通过最小化残差平方和来估计模型参数。其数学表达式为：

$$y=\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta pxp+\varepsilon$$

其目标函数为：

$$\min_{\beta}\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

该模型的优点在于结构简单、可解释性极强、计算效率高，能够直接量化各变量对因变量的边际影响；但缺点同样明显——对多重共线性高度敏感，当特征数量超过样本量时无法求解，且极易过拟合，拟合能力局限于线性关系，无法捕捉变量间的非线性影响过程。

套索回归（Lasso Regression）则加入了 L1 正则化项，通过压缩部分回归系数至零来实现特征选择与过拟合抑制[19]。其目标函数为：

$$\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\}$$

其中，λ 为正则化强度参数，$\sum_{j=1}^{p}|\hat{a}_i|$ 为 L1 惩罚项。Lasso 能够自动筛选不相关变量，提取关键的变量，极大提升了模型的效率；但同样存在弊端，即当预测变量高度相关时，Lasso 倾向于仅选择其中一个而忽略其它变量，导致系数估计存在一定偏误，因此需要谨慎调参。相较于线性回归，Lasso 能够显著提高模型的稳健性与特征筛选能力，但两者均属于线性框架，无法从根本上解决非线性关系的建模问题。

4. 随机森林与 LightGBM

随机森林（Random Forest）是一种基于装袋法的集成学习模型，其通过自助采样法生成多个训练子集，并随机选取特征子集来构建大量相互独立的决策树，最终通过平均或多数投票的方式输出预测结果。其回归预测表达式为：

$$\hat{y}=\frac{1}{B}\sum_{b=1}^{B}f_b(x)$$

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

其中，B 为决策树的总棵数，$f_b(x)$ 为第 b 棵树的预测值。该模型能够自动处理缺失值与异常值，并且敏锐捕捉到非线性关系和特征交互；再通过随机特征选择有效降低了单棵树的方差，使得模型泛化性能优且不易过拟合。但由于构建多个深度较大的树，模型的内存占用大、预测耗时高、可解释性较差，且在噪声较大的数据上容易产生过拟合风险。

轻量级梯度提升机（LightGBM）则是一种基于梯度提升决策树框架的高效集成模型，采用叶节点按叶子生长策略，结合单边梯度采样和互斥特征捆绑技术，在保持高精度的同时大幅提升训练速度。其通用梯度提升目标函数为：

$$L^{(t)}=\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)}+f_t(x_i)\right)+\Omega(f_t)$$

其中，$l(\cdot)$ 为损失函数，$\hat{y}_i^{(t-1)}$ 为前一轮的预测值，$f_t(x)$ 为当前轮次拟合的决策树，$\Omega(f_t)$ 为正则化项。相比随机森林，LightGBM 的训练速度快、内存消耗低，能够高效处理大规模高维数据、更快地收敛至更高精度。但对小样本数据比较敏感，噪声数据拟合效果差，需要精细调整学习率与树深度等超参数。

5. 梯度提升 Boosting 与 XGBoost

极端梯度提升（XGBoost）是梯度提升框架下的集大成者，其在传统提升算法的基础上引入了二阶泰勒展开、L1 与 L2 混合正则化、列采样以及并行化节点分裂等创新技术 [20]。其目标函数在每轮迭代中的通用形式为：

$$L^{(t)}=\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)}+f_t(x_i)\right)+\sum_{k=1}^{t}\Omega(f_k)$$

其中正则化项定义为：

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2+\alpha\sum_{j=1}^{T}|w_j|$$

此处，T 为树的叶子节点个数，w_j 为叶子节点的权重，γ、λ 和 α 分别为控制树复杂度、L2 正则化与 L1 正则化的超参数。在每次分裂时，XGBoost 利用损失函数的一阶导数与二阶导数（即海森矩阵）进行节点划分，其增益计算公式为：

$$\text{Gain}=\frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda}+\frac{G_R^2}{H_R+\lambda}-\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right]-\gamma$$

其中，G 和 H 分别为一阶和二阶导数的累加和。XGBoost 正则化项的引入能够有效抑制过拟合风险，其二阶导数信息能够让树的生长方向更精准，训练效率高的同时能够自动处理缺失值。但 XGBoost 的缺点在于超参数空间庞大，调参过程较为复杂，对硬件资源的消耗也相对较高。

传统梯度提升机（Boosting）同样采用序列化的前向分步算法，每轮迭代拟合上一轮模型的负梯度方向以降低偏差，其迭代公式为：

$$F_m(x)=F_{m-1}(x)+\nu\cdot\gamma_m\cdot h_m(x)$$

其中，ν 为学习率，γ_m 为当前弱学习器的最优步长，$h_m(x)$ 为拟合残差或负梯度的基学习器。传统 GBM 的优点在于模型灵活、可解释性尚可，且通过逐步提升能够显著提高精度，但其训练过程是串行的形式，而不是并行化处理，这就导致训练时间极长。相较于 XGBoost，传统 GBM 在训练速度、泛化性能和工程实现层面均已全面落后，XGBoost 通过二阶近似与结构化正则化超越了经典 GBM 的性能。

以上六类模型的选取既顺应了数据特征，也结合了复杂度、精度、可解释性的比较标准。第一组作为线性框架的上下界，Linear 和 Lasso 作为线性模型类共同检验线性假设在此数据上是否足够。随机森林和 LightGBM 模型则检验集成策略的差异是否显著影响预测精度。此外 Boosting 和 XGBoost 模型用于检验算法迭代是否带来了可观测的性能跃升。三组对照从线性到树模型到梯度提升递进，保证每一组控制变量一致，通过改变和比较核心算法，从而使得模型比较结论具有清晰的归因路径。

（二）模型训练与评估

1.EBM 模型对‘专利产出’的预测

在前文我们将企业数字技术产出的‘创新鸿沟’划分为‘门槛鸿沟’和‘差异鸿沟’，第一步我们针对‘门槛鸿沟’进行研究，即为什么有些企业始终无法跨越门槛实现数字技术产出？为此，本案例在第一阶段采用 EBM 可解释提升机，针对二分类变量“专利产出”建立预测模型。以“专利产出”为被解释变量，采用与第二阶段完全一致的特征集。模型训练采用 EBM 默认参数，按年份分层抽样划分 90% 训练集与 10% 测试集。

EBM 模型评估结果如表 4 所示。在类别不平衡的情况下，我们更加关注的是分类模型的 PR 曲线下方面积，结果显示 PR_AUC 达到 0.7034，说明在这个情况下模型仍然能够有效预测和区分企业数字技术是否有产出。另外，EBM 模型的准确率为 0.7626，查准率为 0.6541，查全率为 0.5959 均达到标准，验证这个模型在当前数据集下有很好的区分能力。

表 4：EBM 模型评估指标

准确率	查准率	查全率	pr 曲线下面积
0.7626	0.6541	0.5959	0.7034

2. 三类模型对‘发明专利授权量’的回归比较

第二步，我们对‘差异鸿沟’进行研究。由于‘发明专利授权量’是一个计数型变量，对他进行回归预测非常考验模型的性能，因此评价标准也更加苛刻，从而选用更多的模型进行比较和稳健性建议。在此我们采用线性回归、集成模型与树模型对‘发明专利授权量’这一被解释变量进行回归。然后筛选出含有专利产出的企业，与 EBM 模型训练一致，均按照年份划分训练基于测试集，每年抽取 10% 作为测试集，90% 作为训练集进行模型训练，并通过网格搜索的方式进行超参数调优。

（1）解释变量适配度

表 5 给出了三种模型在仅含控制变量的基准模型与加入公司治理解释变量的完整模型下的四个评估指标。通过比较三类基准模型与完整模型，判断加入解释变量在不同模型下之后是否能显著提升模型的拟合优度、降低误差。如果可以则说明我们的解释变量对被解释变量是存在显著影响的。

可以看到相比基准模型，随机森林的样本外 R 方从 0.3766 大

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

幅提升至0.5234，是三个模型中表现最佳的模型。梯度提升的完整模型同样优于基准，但表现不及随机森林模型。线性回归的完整模型与基准模型之间差异极小，表明线性模型几乎无法利用公司治理变量提升预测能力。综合以上结果，随机森林在样本外泛化能力上显著领先，是本研究最合适的预测模型，而线性模型的低 Out_R² 验证了其内在可解释性高但精度低的缺陷。这一结论也说明加入的公司结构解释变量能够提升模型拟合能力，公司结构变量对企业数字技术创新存在显著的影响。

表5：模型评估指标表

Model	In_R²	Out_R²	MSE	MAE
Random Forest—基准	0.9149	0.3766	1.1452	0.7532
Random Forest	0.9327	0.5234	0.8756	0.6641
Boosting—基准	0.3782	0.3336	1.2242	0.7846
Boosting	0.4199	0.3767	1.145	0.7633
Linear Regression—基准	0.2203	0.2136	1.4446	0.907
Linear Regression	0.2625	0.2594	1.3605	0.857

（2）模型比较

由于不同模型适应不同种类的数据，为避免模型差异带来结果的不稳定性，选择三类模型进行横向比较。此外为能够使用上最佳的模型进行预测与可解释，我们选取三类共6个模型：Lasso 对比 Linear、XGBoost 对比 Boosting、LightGBM 对比 Random Forest 进行同类模型之间的纵向比较。通过设置调参空间，选取了六个模型中的最佳模型。表6显示各最佳模型中，XGBoost 相比 Boosting 样本内外拟合优度有显著提升，且 MSE 和 MAE 均有所下降，所以选取 XGBoost 进行后续的可解释；随机森林样本内拟合优度虽然高达0.9327，但样本外拟合优度仅有0.5234，说明随机森林模型在这个数据下很可能存在过拟合现象，反之 LightGBM 虽然拟合优度不如随机森林，但也具有一定的拟合效果，且样本内外拟合优度相差较小，因此综合考虑选择 LightGBM 模型进行后续实证；最后是线性模型，这里采用了 lasso 回归和线性回归进行比较，结果显示，二者的评估指标相差不大，Linear 相比 lasso 具有极小的优势。这是因为线性模型比较难以捕捉各变量之间的交互关系，因此线性模型在高维数据下，相比集成模型和树模型效果大大降级，但线性模型可以呈现一些变量与因变量之间的正负关系，也具有一定的参考价值。

表6：检验模型评估指标

Model	In_R²	Out_R²	MSE	MAE
XGBoost	0.7087	0.4907	0.9355	0.6798
Boosting	0.4199	0.3767	1.145	0.7633
LightGBM	0.5213	0.4288	1.0494	0.7288
Random Forest	0.9327	0.5234	0.8756	0.6641
lasso	0.2621	0.2598	1.3597	0.8569
Linear Regression	0.2625	0.2594	1.3605	0.857

图2横轴表示发明专利授权量的真实值，纵轴表示预测值。以散点图形式展示了六个模型在完整模型设定下对测试集样本的预测值与真实值之间的拟合程度。当数据点分布越靠近对角线，表明拟合程度越好。随机森林与 XGBoost 模型的样本店最为紧密地围绕对角线分布，拟合效果显著优于其余模型；梯度提升与 LightGBM 模型次之，数据点离散程度略高；线性模型的数据点分布的最为分散，且存在大量偏离对角线的样本点。这体现了集成学习方法在处理高维、非线性关系的数据具有明显优势。

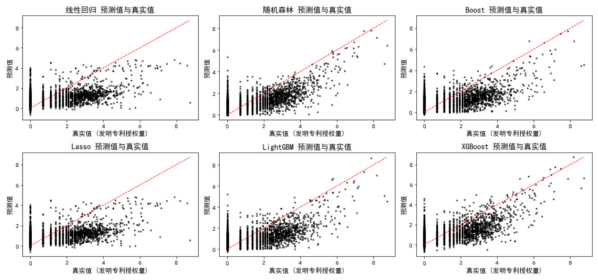


图2：预测值与真实值比较图

（3）10折交叉验证

通过上述模型选取，最终选择模型 Linear、LightGBM、XGBoost 三个不同类别的模型进行后续研究。但可能存在数据集划分不合理导致模型性能不足，为避免这类情况的发生，采用10折交叉验证的方式，进行模型的稳健性检验。

10折交叉验证：将数据随机划分为10组，依次使用每一组作为测试集，其余9组作为训练集，共训练10个模型，得到10个模型的评估指标，并对其取平均值，再与原始划分方法得到的模型评估指标进行比较。

在交叉验证下的结果显示，三个模型的样本内外拟合优度和 MSE、MAE 差距几乎可以忽略，并且文章中使用到的数据划分方式略优于交叉验证时的随机划分。

表7：交叉验证模型评估指标

Model	Sample_In_R2	Sample_Out_R2	MSE	MAE
LightGBM—10fold	0.5207	0.4291	1.0589	0.7351
LightGBM	0.5213	0.4288	1.0494	0.7288
Linear—10fold	0.2624	0.2608	1.3708	0.8650
Linear	0.2625	0.2594	1.3605	0.8570
XGBoost—10fold	0.7062	0.4832	0.9587	0.6874
XGBoost	0.7087	0.4907	0.9355	0.6798

四、构建模型可解释

（一）内在可解释模型

1.EBM 模型探究如何跨越‘门槛’

通过 EBM 模型输出的特征重要性排序显示，影响企业能否跨越数字技术产出门槛的因素众多，本文重点聚焦于四个与公司治理结构密切相关的核心变量：机构投资持股比、监管层总人数、研发人员数量、高管前三薪酬总额。这四个变量在全局重要性排序中名列前茅，说明在控制研发投入等基础因素后，治理结构与人才激励深刻影响着企业能否跨越门槛。其中，机构投资持股比反映了企业外部的资本对企业创新决策的干预；监管层总人数是内部治理架构的体现；研发人员数量代表了创新人力资本的绝对规模；而高管前三薪酬总额则揭示了核心管理层激励的集中度与公平性。这四个变量共同构成了跨越创新门槛的重要因素。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

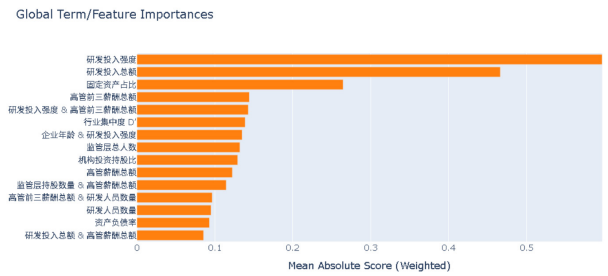


图3：特征重要性排序图

图3展示了上述四个关键变量的 EBM 形函数图，横轴为变量标准化取值，纵轴为对有“专利产出”概率的贡献。这些形状揭示了企业跨越创新门槛的具体临界条件与风险区间。

机构投资持股比的形函数图呈现出先平稳后抑制再促进的三段形态。当机构持股比例处于较低水平时，对专利产出的贡献接近于零。当持股比例上升至某个较高区间后，机构持股比对数字技术产出的影响出现了明显的负向抑制，这一发现提示，机构持股对跨越创新门槛的影响不是线性的，中等偏高的机构持股比对企业数字技术产出存在抑制现象。

监管层总人数的形函数图呈现出平稳抑制到缓慢转向促进再到达到峰值后促进效率降低的渐变形态。在监管层人数较少的阶段，对专利产出的贡献为负或接近零，说明规模过小的监管层难以形成有效的治理与决策支持；随着人数逐渐增加，贡献由负转正，且上升斜率逐步加大；当人数超过某一临界值后开始趋于平稳甚至略有下降。这表明，跨越门槛存在一个最优的监管层规模区间，过小则无力推动，过大则可能因协调成本上升而削弱决策效率。

从研发人员数量来看，形函数呈现出一段先平稳促进、达到一定数量后严重抑制的倒 V 形形态。在研发人员数量较低时，贡献微弱甚至略为负向，说明零散的研发人员难以促进企业创新发展。当研发人员数量进入中等规模后，迅速提升至强烈的正向作用，然而，一旦研发人员数量超过某个阈值，其贡献开始急剧下降甚至转为负作用，且抑制程度随人数增加而加深。这一发现说明：在跨越门槛阶段，研发人员并非越多越好。过大的研究团队可能因管理协调成本激增、研究方向分散等原因，反而降低了企业产出专利的概率。

最后关注到高管前三薪酬总额，这一变量的形函数图呈现出一条平滑的 S 形曲线。在薪酬总额较低的阶段，呈现负向影响。进入中等薪酬区间后，影响则由负转正并快速上升。当薪酬总额继续升高时，正向影响趋于平稳，不再随薪酬增加而提升，表明出现此时高管前三薪酬总额在其中的作用达到饱和。这警示企业适当的高管高薪有助于推动企业跨越专利门槛，但超出合理范围后并不会带来额外收益。



图4：单变量特征形函数图

2. 线性回归探究差异来源

探究完企业跨越门槛的因素，紧接着我们采用线性回归、随机森林与梯度提升模型，结合 SHAP 事后可解释方法，探究具有数字技术产出的企业又是通过怎样的企业结构治理来拉开创新差异。首先是线性回归的内解释作为基础比较。

表8展示了线性回归模型在基准模型与完整模型下的回归系数、z 值及显著性水平。企业结构变量中，监管层持股数量、高管薪酬总额、主营业务、政府补助、研发人员数量、监管层总人数以及高管教育背景对企业数字技术产出呈现显著的正向作用。其中监管层总人数的系数绝对值最大，为0.5151，这表明当监管层每增加一人，数字技术产出预期提升约51.5%，说明治理层人力规模在推动创新中起到了重要作用。另外监事会规模、高管前三薪酬总额、第一股东持股比、研发人员数量占比、机构投资持股比、董事会规模以及高管信息技术背景则对企业数字技术产出具有显著的负向作用。

其中存在一个矛盾的现象，即研发人员数量为正而研发人员占比为负，暗示线性模型可能无法准确刻画比例的边际效应；高管薪酬总额为正而高管前三薪酬总额为负，则预示着高管团队内部薪酬差距过大可能抑制协同创新。

表8：线性回归系数表

指标	系数	z 值	显著性	系数	z 值	显著性
控制变量						
企业年龄	0.2313	17.0497	***	0.0670	4.5785	***
总资产净利润率 ROA	0.0330	5.0950	***	0.0251	3.8732	***
总资产	−0.0900	−11.6696	***	−0.0489	−6.1944	***
企业规模	0.1443	15.6995	***	0.0114	1.0508	
营业收入增长率	0.0119	1.8655		−0.0048	−0.7710	
托宾 Q 值 D	−0.0131	−1.9368		−0.0321	−4.7481	***
资产负债率	0.0054	0.6204		−0.0233	−2.1680	*
固定资产占比	0.0922	11.2793	***	0.0586	6.8227	***
行业集中度 D’	−0.0873	−11.5097	***	−0.0914	−11.8318	***
研发投入总额	0.4243	60.1467	***	0.3258	44.3123	***
研发投入强度	0.3004	45.9443	***	0.1830	21.5187	***

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

解释变量			
监事会规模	−0.0796	−2.8659	**
监管层持股数量	0.0179	2.5111	*
高管薪酬总额	0.0655	3.1721	**
高管前三薪酬总额	−0.0507	−2.6490	**
第一股东持股比	−0.0169	−2.4302	*
主营业务	0.0816	4.9352	***
股权性质	0.0179	1.1964	
政府补助	0.0864	13.4233	***
研发人员数量	0.1304	38.2884	***
研发人员数量占比	−0.0777	−7.6554	***
监管层总人数	0.5151	9.6021	***
机构投资持股比	−0.0309	−3.6653	***
董事会规模	−0.1131	−2.5739	*
独立董事人数	0.0665	1.9006	
两职合一	0.0173	0.7234	
高管教育背景	0.1643	6.1375	***
高管信息技术背景	−0.0804	−2.9502	**

注：*、**、*** 分别代表0.05、0.01、0.001显著性水平。

（二）SHAP 事后可解释方法

SHAP 可解释机器学习（SHAPley Additive exPlanations）是一种基于合作博弈论中 SHAP 值的模型解释框架，旨在提升机器学习模型的可解释性。相较于特征重要性指标，可解释 SHAP 机器学习具有以下优势：第一，SHAP 值可以反映特征指标的贡献大小和正负向作用，为检验研究假设提供直接证据。第二，具备局部解释能力，可为局部样本生成独立的贡献值分布，从而捕捉特征在不同样本之间的异质性影响。第三，具有模型无关性，不依赖模型具体结构，可适用于各类机器学习算法。SHAP 值的计算流程如下：首先，利用加性特征归因方法将预测值 $\hat{f}(x)$ 分解为每个特征指标的贡献值之和，其中 ϕ_0 为所有样本预测均值，ϕ_i 为第 i 特征指标的 SHAP 值：

$$\hat{f}(x)=\phi_0+\sum_{i=1}^{m}\phi_i \tag{8}$$

其次，计算一个特征在所有可能的特征组合（子集）中的平均边际贡献。第 j 个因素的 SHAP 值 ϕ_i^j 计算公式如下：

$$\phi_i^j=\sum_{S\subseteq F\setminus\{i\}}\frac{|S|!(|F|-|S|-1)!}{|F|!}(\gamma_*(S\cup\{j\})-\gamma_*(S)) \tag{9}$$

其中 S 是不包含第 j 个因素的集合，共可构造出 $\sum_{S\subseteq F\setminus\{i\}}\frac{|S|!(|F|-|S|-1)!}{|F|!}$ 个不包含 j 因素的子集，$\frac{|S|!(|F|-|S|-1)!}{|F|!}$ 可以视为所有可能边际贡献的权重。$(\gamma_*(S\cup\{j\})-\gamma_*(S))_j$ 表示在变量集合 S 基础上，第 j 个因素对预测值的期望影响大小。

本案例使用到的 SHAP 可解释体现方式有以下几种：

（1）SHAP 可解释蜂群图。蜂群图能够绘制每个变量的每个样本对该模型的解释力度，从上至下按照变量总解释力度由高到低排列（即 SHAP 绝对值高低）；最右边色条表示变量数值的高低，数据点越红表示数值越高，越蓝表示数值越低；样本点分布在中轴线右边表示该样本的这个数据对结果预测是正向的影响，如果在中轴线左边则表示负向影响。

（2）SHAP 依赖图。SHAP 依赖图通过散点图的形式体现单一变量的值与 SHAP 值的关系。SHAP 蜂群图可能会有一些变量存在样本分布冗杂的情况，无法直观看出正负向影响，或者要探究其中非线性关系时，可以绘制 SHAP 依赖图。

1. 全局可解释提取关键指标

图5分别是 XGBoost 和 LighGBM 模型在 SHAP 事后可解释的蜂群图，其中研发人员数量和监管层总人数在两个模型下的排名都名列前茅，研发人数和监管层总人数较多的样本多分布在右侧，正向效应显著，说明研发人力资本以及优质的企业监管是驱动数字技术产出最核心的因素。政府补助和高管薪酬总额排名居中，政府补助在两个模型中高值样本均偏右，高管薪酬总额在 LightGBM 模型中高值样本偏右，表明外部资金支持与高管薪酬对创新产出具有强力促进作用。然而高管薪酬总额在 XGBoost 中呈现出左右均出现高值样本，低值样本集中在中间区域，两个模型呈现出不同的结果可能是受不同模型的影响，也可能是存在一定的交互作用。

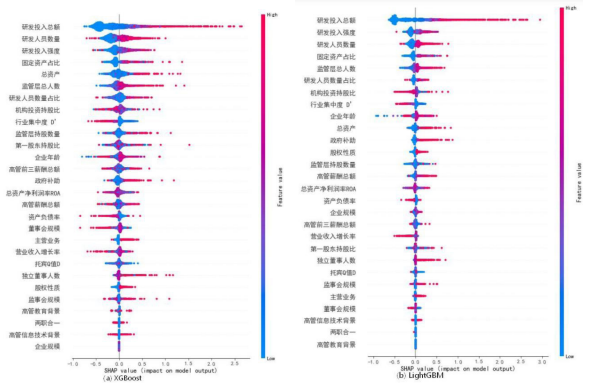


图5：SHAP 可解释蜂群图

值得特别关注的是，研发人员数量占比、机构投资持股比在整个排名中也位居前列，但这几个解释变量的 SHAP 值分布呈现明显的双向性：高值与低值样本交杂在一起，无法清晰捕捉其中正负向影响。此外，高管教育背景的影响力排名中等，但其SHAP

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

值多为正值，说明高学历管理团队普遍有益于创新。整体而言，数字技术创新的“差异鸿沟”首先源于研发人力与资金投入的差距，其次是不当的股权集中与监管层规模对创新活力的抑制。

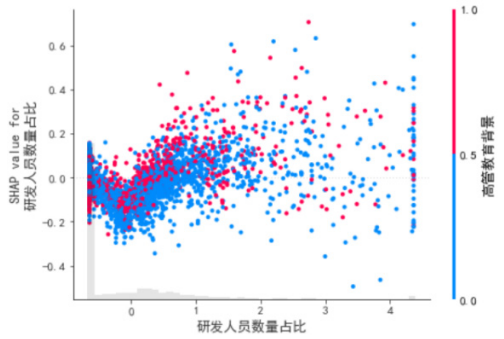
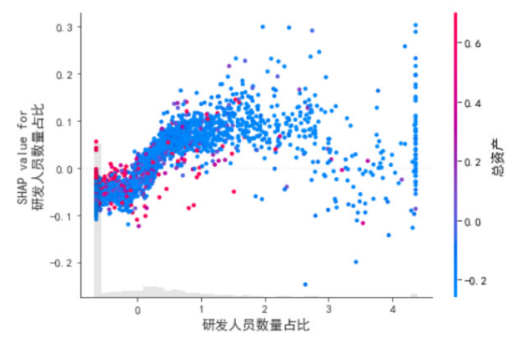
2. 单一可解释寻找非线性与交互关系

源于全局可解释下数据难以观察到非线性关系以及交互作用，于是对单个变量进行 SHAP 依赖图解释，发现以下三个指标与数字技术创新具有较为特殊的关系，且存在一定的交互项。图中数据点的颜色代表当下样本对应右侧变量的值的高低，红色点表示值较高，蓝色则相反。

（1）研发人员数量占比

表9揭示了研发人员数量占比这一变量对数字技术产出的影响，以及该变量与高管教育背景和总资产的交互作用。可以发现与 EBM 模型的形函数图存在相似的情况，即研发人员数量并非越多越好，而是存在一定峰值。并且重要的是交互作用显示较高的高管教育背景可以提升整个研发团队的水平。

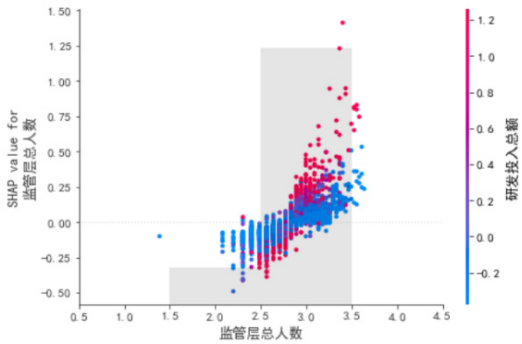
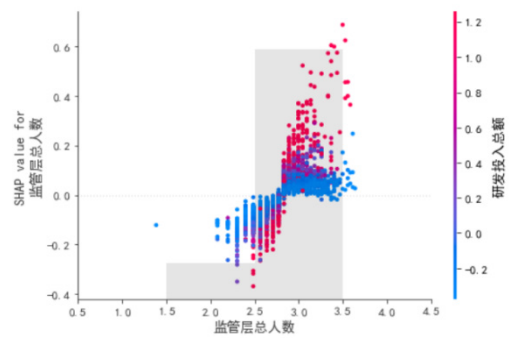
表9：研发人员数量占比分析

模型	XGBoost	LightGBM
单变量分析	随着研发人员数量占比升高，SHAP 值先是呈现‘U’式下降后升高，当占比达到一定数量时，SHAP 值开始离散并缓慢上升直至平缓。对企业数字技术创新先是抑制后促进。	与 XGBoost 模型不同的是，研发人员数量占比较少时 SHAP 值并未呈现‘U’形，而是直接上升，但后段同样变得平缓甚至出现下降，对企业技术创新产生抑制作用
交互项分析	研发人员数量占比 × 高管教育背景 高管教育背景与研发人员数量占比的交互作用下分布比较均匀，红色点的 SHAP 值分布相比蓝色点整体更高，说明高管的受教育水平较高时能够提升研发人员数量占比对企业技术创新的正向作用。	研发人员数量占比 × 总资产 总资产高值样本相比低值样本少很多，在整个依赖图在较为分散，主要集中在研发人员数量占比较少的企业，一小部分高资产企业在研发人员数量占比较少时能够降低对企业技术创新的抑制作用。
SHAP 依赖图		

（2）监管层总人数

表10呈现了监管层总人数的单一指标影响以及和研发犹如综合的交互作用。其中 XGBoost 与 LightGBM 模型的 SHAP 依赖图呈现高度相似。通过依赖图的分析可以得到以下：监管层总人数是企业数字技术创新进程中一个具有“开关”意义的治理变量，其影响力呈现指数型增长，而高研发投入总额能够放大监管层总人数的影响效应。

表10：监管层总人数分析

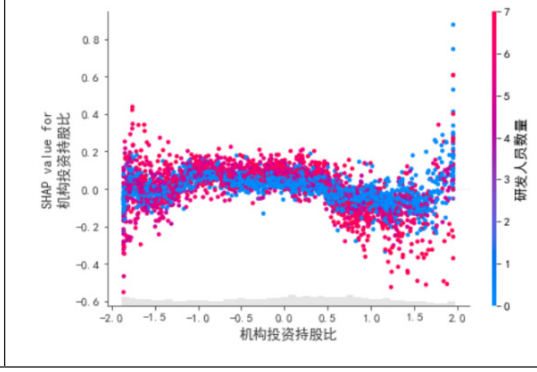
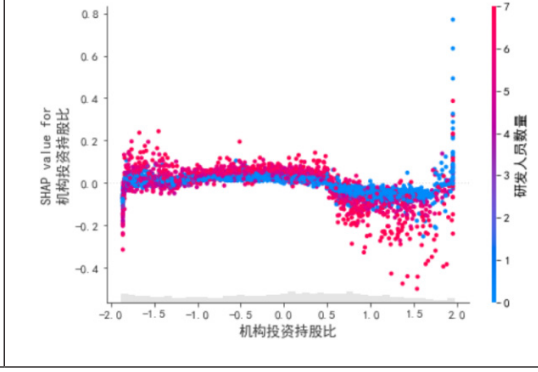
模型	XGBoost	LightGBM
单变量分析	图中显示随着监管层总人数的增加，SHAP 值呈现指数型增长，监管层总人数少的时候先是抑制数字技术创新产出，紧接着当监管层总人数增加则迅速从抑制转为促进。	
交互项分析	监管层总人数 × 研发投入总额 图中显示高研发投入在监管层影响企业技术创新过程中呈现两极分化：当监管层总人数较低，对企业数字技术创新呈现抑制效果时，高研发投入的企业数据点集中在下缘，说明此时高投入反而使得抑制效果更加明显；反之，当监管层总人数较高，对企业数字技术创新呈现促进作用是，搞研发投入的企业数据点集中在上缘，此刻高投入使得促进效果更加明显。	
SHAP 依赖图		

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

（3）机构投资持股比

表11展示的是两个模型下机构持股比的单一影响过程和其与研发人员数量的交互影响过程，且两个模型的SHAP蜂群图在该变量下也呈现出一致性。结合表内分析发现，机构持股比过高会抑制企业数字技术产出。交互作用展现初大量的研发人员数量也会放大机构持股比在其中的作用。机构持股比与研发人员数量之间存在一个“最优配合区间”，只有在机构持股适中的范围内，研发人员的规模优势才能与外部治理力量形成正向协同。

表11：机构投资持股比分析

模型	XGBoost	LightGBM
单变量分析	当机构持股比升高时，SHAP值为正，说明此时机构持股比对企业数字技术创新呈现轻微的促进作用，但当上升到一定峰值开始，SHAP值开始降到0值以下，说明在这个时候机构持股比对企业技术创新转为抑制作用。	
交互项分析	机构持股比 × 研发人员数量 研发人员数量在机构持股比对企业数字技术创新中的影响也呈现两极分化，当机构投资持股比较少或居中时，研发人员数量越多对其促进作用越有效；反之当机构持股比较高时，大量投入研发人员数量反而使得抑制效果更明显。	
SHAP依赖图		

3. 时间滚动窗口期

将时间窗口缩短至数字技术创新及企业数字化转型的发展和热门阶段，即2011–2021年期间，固定窗口期为5年，依次绘制2011–2015、2012–2016、2013–2017・・・以此类推的5年内指标特征解释性排序，形成时间滚动窗口期。

将不同周期下各解释变量的SHAP绝对值排序与完整周期下的变量SHAP绝对值排名比较，探究随着时间的推移，哪些指标始终是影响数字技术产出的关键因素，哪些指标是具有时代效应的。

综合考虑Linear、LightGBM、XGBoost模型，三者评估指标中，XGBoost的拟合优度最高，误差最小。而SHAP可解释下的交互作用LightGBM于XGBoost相差不大。因此在进行时间滚动窗口期时以XGBoost模型为代表。表12是不同时间周期内，各解释变量在所有变量中SHAP绝对值的排名。

表12：解释变量SHAP绝对值排名

变量	11–15	12–16	13–17	14–18	15–19	16–20	17–21	完整周期
第一股东持股比	4	5	5	8	8	8	11	11
股权性质	23	21	21	23	22	16	20	23
董事会规模	20	20	19	20	20	19	21	18
独立董事人数	22	24	23	21	21	23	22	22
研发人员数量	16	4	3	2	3	3	4	2
研发人员数量占比	15	16	15	13	5	9	8	7
监事会规模	21	23	22	22	23	22	23	24
高管薪酬总额	12	11	16	18	14	17	10	16
政府补助	2	2	10	12	16	18	16	14
机构投资持股比	9	14	7	6	7	6	6	8
监管层总人数	18	19	14	7	10	5	5	6
监管层持股数量	5	9	11	10	11	10	9	10
高管前三薪酬总额	8	6	13	9	13	12	12	13
两职合一	26	26	26	25	26	26	26	26
高管教育背景	24	22	27	26	27	27	27	25
主营业务	25	25	24	24	24	24	24	19
高管信息技术背景	27	27	25	27	25	25	25	27

基于时间滚动窗口与完整周期的解释变量重要性排名演变比对发现，政府补助的重要性从2011至2015年的第2位跌至后续窗口的第18位，完整周期仅列第14位，而研发人员绝对数量则从第16位跃升至第4位并长期稳居高位，这一轨迹说明政府补助和研发人员具有时代效应，并非固定效应。且企业创新动力已从政策补贴转向人力资本，表明人力结构优化正逐渐受到重视。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

外部治理变量中，机构投资持股比稳定在第6至8位，监管层总人数从早期的第18位跃升至第6位，两者排名均保持前列或显著上升，反映出外部监督力量在创新解释体系中的地位日益强化。而传统治理变量逐渐淡出或持续影响力不足，例如第一股东持股比从早期的第4位持续下滑至第11位，股权集中的解释力明显衰退。高管薪酬总额、高管前三薪酬总额长期徘徊于第10至16位且无上升趋势。高管信息技术背景、高管教育背景、两职合一等变量重要性始终处于末端，这说明货币激励与表层治理手段并非关键解释因子，常规治理架构设置同样不是决定数字创新水平的核心。

五、研究结论

本研究使用了2000到2024年中国A股上市公司45848条年度数据，考虑到企业数字技术创新产出具有“零膨胀”和“长尾分布”同时存在的特点，把企业之间的创新差距分成了“门槛鸿沟”和“差异鸿沟”两个前后递进的层次，并搭建了一个分阶段的分析框架，先用分类模型来识别门槛，再用回归模型来解释差异。主要结论有下面几点：

第一，多种模型组合起来的方法在预测非线性很强的创新产出时，比线性模型要准确很多，而且模型评价结果在内部比较稳定；随机森林和XGBoost在预测新数据时的准确程度明显高于线性回归和Lasso，其中XGBoost在预测准确度和结果稳定性之间平衡得更好，线性模型只适合用来辅助判断系数方向；把数据分成十份反复验证后发现，各个模型评价指标在不同分法下变化很小，所以结论不会受到特定数据划分方式的影响。

第二，同一个变量在跨越门槛和扩大差异这两个阶段，发挥作用的方式有很大不同；中等规模的研发团队最容易帮助企业实现从无到有的突破，而研发人员的总人数在突破之后的差异阶段会一直起到正向作用，所以呈现出“起步阶段靠精干团队、扩大差距阶段靠人员规模”的差异特点；监管层人数规模存在明显的门槛作用，超过某个临界点之后，它的促进作用会明显增强，这一点被EBM分类模型和SHAP回归分析一起证实。

第三，用SHAP做事后解释发现了三组线性模型抓不到的非线性交互作用；监管层规模和研发投入之间存在一种“先抑制后放大”的门槛协同关系，机构持股比例和研发人员数量之间有一个最优的配合范围，研发人员占比和高管教育背景之间是正向互补的；这些交互作用说明，治理结构、股权结构和人力资本之间是否匹配，是影响数字技术创新表现的关键因素。

第四，通过时间滚动窗口分析可以看出，2011到2021年企业数字技术创新的驱动因素发生了结构性的变化；政府补助的重要性明显下降，研发人员数量和监管层规模的重要性不断上升，它们在时间上的变化轨迹像一个剪刀差，这说明创新驱动正在从依靠政策扶持转向依靠人力资本和治理能力；第一大股东持股比例和高管信息技术背景这些变量的重要性一直处于较低位置并且比较稳定，说明它们的解释能力存在结构性限制。

参考文献

[1] 钟祖昌，赵芊，余佩璇，等．企业数字化转型的技术创新效应——基于沪深A股上市公司的实证分析[J]. 科技管理研究，45(23):144-157,2025.

[2] 宁靓，王培伟．技术溢出影响企业的技术创新产出吗？——基于数字技术应用与市场竞争的联合调节效应[J]. 技术经济，44(11):110-122,2025.

[3] 白福萍，刘东慧，董凯云．数字化转型如何影响企业财务绩效——基于结构方程的多重中介效应分析[J]. 华东经济管理，36(09):75-87,2022.

[4] 丛昊，张春雨．数字技术与企业高质量创新[J]. 中南财经政法大学学报，(04):29-40,2022.

[5] 肖土盛，吴雨珊，亓文韬．数字化的翅膀能否助力企业高质量发展——来自企业创新的经验证据[J]. 经济管理，44(05):41-62,2022.

[6] 温湖炜，王圣云．数字技术应用对企业创新的影响研究[J]. 科研管理，2022,43(04):66-74.

[7] 谢婼青，方顺超，朱平芳．中小企业数字化转型的鸿沟：内在机制与弥合路径[J]. 学术月刊，58(02):71-85,2026.

[8] 孙晓华，马雪娇，孔一杰，等．中国制造业的数字鸿沟——“接入—应用—转化”的分析框架[J]. 管理世界，41(04):38-58,2025.

[9] 李柏洲，李宜函，曾经纬，等．企业如何跨越数字创新中的非连续性鸿沟？——基于博世（中国）的案例研究[J]. 科学学研究，42(06):1334-1344,2024.

[10] 徐毅，杜凤龙．数字金融、数字鸿沟与企业技术创新[J]. 哈尔滨商业大学学报（社会科学版），(06):60-71,2024.

[11] 周闯，曲植，曾思棚，等．科技金融与关键数字技术突破：来自“专精特新”企业的证据[J]. 世界经济，48(11):3-32,2025.

[12] 谢红军，齐翔．数字技术与企业价值——基于数字并购场景的研究[J]. 管理世界，41(10):146-175,2025.

[13] 陶锋，翟少轩，王峤．数字经济政策与传统企业跨界数字创新[J]. 中国工业经济，(02):118-136,2025.

[14] 沈坤荣，林剑威．链“岛”成“陆”：公共数据开放的技术创新效应研究[J]. 管理世界，41(02):83-104,2025.

[15][15] 杨鹏，尹志锋，孙宝文，等．数字技术应用与企业合作创新[J]. 经济管理，47(01):108-127,2025.

[16] 吴学福，刘振球，吴明山，等．零膨胀计数资料几种模型方法的比较研究[J]. 中国卫生统计，2020,37(3):331-334.

[17]James W M，Chandan S，Karl K，et al. Definitions，methods，and applications in interpretable machine learning.[J].Proceedings of the National Academy of Sciences of the United States of America,2019,116(44):22071-22080.

[18] 石井 一夫．5分で分かる⁉ 有名論文ナナメ読み：Marco T. Ribeiro et al.：“Why Should I Trust You?”：Explaining the Predictions of Any Classifier[J]. 情報処理，2021,62(10):568-570.

[19]Sinha K R . Book review: Christoph Molnar. 2020. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable[J].Metamorphosis，2024,23(1):92-93.

[20]Robert T . Regression Shrinkage and Selection Via the Lasso[J].Journal of the Royal Statistical Society: Series B (Methodological),2018,58(1):267-288.

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

考虑设备转运与购置决策的多车间协同整修调度研究

魏陆炀

众言科技股份有限公司，广东 广州 510000

DOI:10.61369/ASDS.2026100011

摘　　要： 针对多车间整修中工序衔接、移动设备竞争与设备增购的耦合问题，建立以最大完工时间最小化为目标的资源约束调度模型。将五车间重复工艺展开为27道工序，描述同类设备并行、异类设备协同、分项释放及跨车间转运。在同步开工和设备能力线性叠加假设下，分别采用粒子群、遗传和蚁群算法求解单班组、双班组与预算购置情景。三种方案工期分别为159574、82180和54360 s，后两者较前一情景缩短48.50%和33.85%。购置方案在50万元预算内增购9台设备，支出48万元。抛光与检测设备是基准情景的紧缺资源，增购后末完工车间由C转为E。明细核验支持排程可行性，但尚不能证明全局最优或算法优越性。研究为可分工程量和移动设备共享条件下的整修规划提供了分析方法。

关 键 词： 多车间整修；资源约束调度；设备转运；设备购置；启发式算法

Multiworkshop Maintenance Scheduling Considering Equipment Transfer and Acquisition Decisions

Wei Luyang

Zhongyan Technology Co., Ltd. Guangzhou, Guangdong 510000

Abstract: This study examines the interaction between operation precedence, mobile equipment sharing, and equipment acquisition in multiworkshop maintenance. A resource-constrained model is developed to minimize makespan. A five-workshop case is expanded into 27 operations, incorporating parallel processing by same-type equipment, joint completion by different equipment types, separate equipment release, and interworkshop transfer. Under synchronized starts and linearly additive capacities, particle swarm optimization, a genetic algorithm, and ant colony optimization are applied to single-team, two-team, and acquisition-enabled scenarios, respectively. The resulting makespans are 159574, 82180, and 54360 s, with successive reductions of 48.50% and 33.85%. Nine machines are purchased for CNY 480000 within a CNY 500000 budget. Polishing and sensing equipment are scarce in the baseline scenario, and the last-finishing workshop shifts from C to E after acquisition. Record-level checks support schedule feasibility but do not establish global optimality or comparative algorithm superiority. The approach supports maintenance planning with divisible workloads and shared mobile equipment.

Keywords: multiworkshop maintenance; resource-constrained scheduling; equipment transfer; equipment acquisition; metaheuristic algorithms

引言

集中整修中，清洗、灌装、抛光和检测设备需在多个车间之间共享。局部工序提前完成，未必能够缩短总体工期，后续作业仍可能等待紧缺设备或受转运限制。因此，整修计划需要同时确定工序起止时间、设备指派与释放时刻，以及新增设备的配置方向。

资源约束项目调度关注工序先后与有限资源分配。Hartmann 和 Briskorn[1]、Ding 等[2]归纳了该类问题的扩展，多项目调度进一步讨论多条任务链的资源竞争[3]。柔性作业车间调度则重点协调工序排序与机器选择[4]。本文以工序链和共享资源为建模基础，借鉴相邻调度研究处理设备选择与移动约束。

运输约束是模型落地时不可省略的环节。Yan 等[5]研究了有限运输条件下的调度与解码，Berterotti è re 等[6]将运输资源显式纳入柔性作业车间模型。张国辉等[7]讨论了有限 AGV 资源配置与调度之间的联系。上述研究主要涉及工件或物料在机器之间流转，本文则以作业对象位置固定、设备跨车间移动为研究对象，并允许多台同类设备共同处理可分工程量。二者均需协调加工与移动，但资源流向

作者简介：魏陆炀，众言科技股份有限公司高级大数据分析师，研究方向为智能决策与数字化管理。主持多项企业自研项目，相关成果包括软件著作权7项、专利5项，发表学术论文3篇。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

和持续时间形成机制不同。

Fan 等 [8] 针对有限辅助模块下的可重构机床调度，设计了专门的编码、解码与局部搜索策略。Chen 等 [9] 进一步研究有限运输资源下的动态重调度。本文聚焦确定性计划阶段，通过设备容量、驻地和预算变化构造资源情景，分析资源配置与排程的耦合，并以作业明细和工期下界检验结果。

一、数据来源与作业机制

（一）算例与工序展开

本文采用包含五个车间和两个班组配置的确定性整修调度算例。输入由工序流程、班组配置和车间距离三类数据构成，包括工程量、设备效率、数量、移动速度、购置单价及位置间距离。算例用于检验调度模型与配置方案，不将计算结果直接解释为企业实际生产绩效。

A、B、D、E 车间分别含3、4、6、3道工序；C 车间在C1、C2之后重复三轮 C3、C4、C5，展开为11道工序。合计27道工序、41组“工序—设备类型”需求。各车间内部顺序固定，车间之间无预设工序依赖。

（二）设备资源与时间口径

两个班组的初始设备数量相同，每组16台，配置见表1。

表1：初始设备配置与购置单价

Table 1: Initial equipment capacities and unit prices

设备类型	班组1/台	班组2/台	单价/万元
自动化输送臂	4	4	5.0
工业清洗机	5	5	4.0
精密灌装机	5	5	3.5
自动传感多功能机	1	1	8.0
高速抛光机	1	1	7.5

设备移动速度均为2 m/s。班组1、2距 A 车间分别为400、500 m，首次到达需200、250 s。作业后设备位置更新为当前车间，据此计算下一次转运，同车间转移时间为零。

工程量与单台效率分别以 m³ 和 m³/h 计量，持续时间换算为秒后向上取整，超过24 h 的工期按累计小时表示。两个班组原有配置均含高速抛光机，各情景将其计为既有设备。

（三）模型假设与适用边界

工程量和设备效率确定，同类设备均分工程量、能力线性叠加，适用于作业面可分割且相互干扰较弱的情况。异类设备各自完成工序规定的全部工程量，不能将不同类型效率相加。

选定设备在同一工序内同步开工，分项作业结束后立即释放；后继工序须等待前序全部需求类型完成。同步开工是解码策略的附加限制，未覆盖异类设备异步开工的全部排程。

作业不可抢占，单台设备不能同时执行多项作业或边移动边加工。设备可在自身作业结束后提前转运。不考虑故障、道路容量、人员班次和安装准备；新增设备在任务开始前已于所属班组驻地可用，购置费用仅含设备单价。

二、设备配置与调度模型

（一）变量与工序持续时间

设 I、W、K 分别为工序、车间和设备类型集合。工序 i 的需求类型、工程量与类型 k 单台效率分别为 K_i、Q_i、v_{ik}；E_k 为可用设备集合，n_{ik} 为调用数量，取1至可用数量之间的整数。S_i、C_i 分别为工序开始与整体完成时刻，F_{ik} 为类型 k 设备的释放时刻。

$$\mathrm{T_{ik}}=\mathrm{ceil}\left(\frac{3600\mathrm{Q_i}}{\mathrm{n_{ik}v_{ik}}}\right),\qquad \mathrm{F_{ik}}=\mathrm{S_i}+\mathrm{T_{ik}} \tag{1}$$

$$\mathrm{C_i}=\max_{\mathrm{k}\in\mathrm{K_i}}\mathrm{F_{ik}},\qquad \mathrm{minC_{max}}=\min\max_{\mathrm{i}\in\mathrm{I}}\mathrm{C_i} \tag{2}$$

其中，ceil 表示向上取整。设备数量决定加工持续时间，工序在各需求类型全部完成后结束。设备实际占用区间为 S_i 至 F_{ik}，先完成的类型可转向其他车间，无需等待至 C_i。

（二）优先关系与设备路径约束

记 P_i 为工序 i 的直接前序集合，固定流程要求：

$$S_i \ge C_j,\quad j\in P_i \tag{3}$$

每台设备形成按执行次序排列的作业路径。设设备 e 从位置 a 至 b 的距离为 d_{ab}、速度为 u_e、耗时为 τ_{abe}。路径中相邻工序 i、j 须满足：

$$\tau_{abe}=ceil\left(\frac{d_{ab}}{u_e}\right),\quad S_j\ge F_{i,k(e)}+\tau_{w_iw_je} \tag{4}$$

其中，k(e) 为设备类型，w_i、w_j 为工序所属车间。首项作业以班组驻地和零时刻替代上一位置及释放时刻。每台设备在同一工序内只分配一次，类型须匹配，参与设备数须等于 n_{ik}，由此保证数量、初始可达性和作业运输不重叠。

（三）预算约束下的资源扩展

设 b 为班组，N_{kb}、z_{kb} 为既有和新增数量，p_k 为单价，B 为预算。新增设备沿用既有设备的效率与移动规则；固定资源情景令 z_{kb} 为零，购置情景联合选择数量及路径：

$$\sum_{k\in K}\sum_{b}p_kz_{kb}\le B,\quad z_{kb}\in Z_{\ge0} \tag{5}$$

$$N_k(z)=\sum_b(N_{kb}+z_{kb}),\quad 1\le n_{ik}\le N_k(z) \tag{6}$$

取 B=500000元。模型仅最小化总工期，支出作为约束；工期较短的预算可行方案不等于全寿命成本最低或投资回报最佳。

（四）用于解释解质量的下界

给定配置下，类型 k 的理想累计设备工时记为 H_k。忽略运输、等待和取整增加量，得到资源负荷下界：

$$H_k=\sum_{i:k\in K_i}\frac{3600Q_i}{v_{ik}},\quad L_R=\max_k ceil\left(\frac{H_k}{N_k}\right) \tag{7}$$

另将全部同类设备分别供各车间独占，并忽略运输与外部竞

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

争，按固定工序链累加最短加工时间，得到链长下界 L_P。取 L=max{ L_R， L_P }，则任何可行方案工期 C 均不小于 L。本文用（C − L）/L 表示相对下界的偏离程度。该值用于刻画可证明下界与当前方案之间的距离，不是已知真实最优解时的实际最优差距。购置情景的下界只对其选定配置成立，不能用于证明该购置组合在全部预算可行组合中最优。

三、求解策略与计算设置

（一）可行排程的串行生成

固定资源情景采用27个工序优先级基因和41个调用数量基因，总维度68。数量基因 g 位于 [0，1]，映射为：

$$n_{ik} = min\left\{N_k, 1 + floor\left(g_{ik} N_k\right)\right\} \tag{8}$$

其中，floor 表示向下取整。先从前序已排定的工序中选优先级最高者，再逐类型选取所需数量的最早可达设备。设 F_e、l_e 为设备释放时刻和当前位置，E_i 为所选集合，则：

$$A_{ie} = F_e + \tau_{l_e w_j e},\quad S_i = max\left\{\underset{j\in P_i}{max}\, C_j, \underset{e\in E_i}{max}\, A_{ie}\right\} \tag{9}$$

无前序时，式（9）的前序完成项取零。排定后更新设备释放时刻、位置及工序完成时刻。“已排定”指完成时刻已确定，不要求先排定作业在时间上先完成，车间仍可并行。设备任务按现有路径尾部追加，通过时间递推保证可行，但未搜索全部空闲插入区间。

（二）搜索方法与参数

单班组采用粒子群优化 [10]，以个体和群体历史最优更新优先级及调用数量，并设置递减惯性权重、速度截断和随机重置。双班组采用遗传算法，通过锦标赛选择、混合或均匀交叉、随机扰动与精英保留搜索排程。

购置情景采用蚁群搜索，分别构造设备购置、工序顺序与设备调用数量。购置阶段仅从剩余预算允许的候选项中选择；工序选择同时参考信息素和剩余工序链的启发信息，随后由同一时间递推规则完成设备指派。按方案工期给予信息素反馈并进行挥发，使预算配置与调度效果共同影响后续构造。信息素协同搜索的基本机制见文献 [11]。

参数见表2，主要随机生成器种子均为2026。各情景采用已有单次运行，未进行多随机种子或同资源条件的算法对照，不能推断均值、方差及统计显著性。

表2：算法计算设置

Table 2: Computational settings

情景与算法	规模与轮数	主要参数
单班组 PSO	90个粒子；260次	惯性权重0.90→0.40；学习因子均1.80；速度上限0.25；重置概率0.01
双班组 GA	140个个体；320代	精英6个；交叉概率0.88；变异概率0.055；扰动尺度0.12
购置 ACO	90只蚂蚁；220次	信息素指数1.2；启发指数2.5；挥发系数0.18；精英权重2.0

（三）结果核验规则

从设备作业明细逐条检查起止差与记录时长是否一致，按工序与设备类型重新聚合调用数量，并检查工程量是否完成；按车间检查后继工序不早于前序整体完成；按设备路径检查前一作业结束、转运与下一作业开始的时间衔接；最后核对新增设备、购置支出及最大完工时间。收敛日志仅用于识别单次运行首次达到最终值的轮次，曲线后段平稳不作为全局最优或跨运行稳定性的证明。

四、算例结果与分析

（一）单车间时间核算

班组1独立承担 A 车间时，设备到达需200 s。A1投入5台灌装机和4台输送臂，处理300 m³，两类设备均需1080 s；A2投入1台抛光机和5台清洗机，处理500 m³，分别需18000、1440 s；A3由1台传感设备处理500 m³，需18000 s。总工期为200 + 1080 + 18000 + 18000=37280 s，即10:21:20。

在严格串行、线性能力和无故障假设下，各工序投入全部可用同类设备，达到加工时间下界，故该单车间结果可证明为条件最优，但不能与五车间情景直接比较降幅。

（二）多车间资源情景比较

表3中，单班组16台设备的工期为159574 s（44:19:34）；双班组32台设备时为82180 s（22:49:40），缩短48.50%；增购后41台设备时为54360 s（15:06:00），较双班组缩短33.85%，较单班组累计缩短65.93%。

表3：五车间情景的工期与下界

Table 3: Makespans and lower bounds across scenarios

情景	工期 /s	下界 /s	相对下界偏离 /%	较前情景缩短 /%
单班组	159574	135000	18.20	—
双班组	82180	67500	21.75	48.50
双班组增购	54360	39120	38.96	33.85

三情景链长下界为93734、46867、26675 s，均低于资源负荷下界。当前工期与下界存在间隔，不能称为全局最短。不同情景的配置与搜索对象同时变化，偏离幅度不宜用于比较算法优劣。

情景差异反映资源条件与求解方案的联合结果，说明本算例存在通过共享和增购压缩工期的可行途径，但不能单独识别资源扩容的净效应。

（三）设备负荷与车间完工差异

单班组抛光设备累计作业135000 s，利用率84.60%；传感设备累计作业117360 s，利用率73.55%。利用率按加工时间与总工期之比计算，不含转运。两类设备各仅1台且被多个车间反复调用，其负荷与占用可解释部分等待，其他设备空闲不能替代其能力。

表4显示，C 车间在前两情景均最后完成，重复工序链多次竞争抛光和检测设备。A 车间在单班组多车间方案中于140899 s结束，显著晚于独占设备时的37280 s，反映跨车间共享引起的等待。

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

表4：各车间完成时刻

Table 4: Workshop completion times in seconds

车间	单班组 /s	双班组 /s	双班组增购 /s
A	140899	60607	41405
B	64664	55592	52680
C	159574	82180	46730
D	132914	67752	52280
E	31604	75957	54360

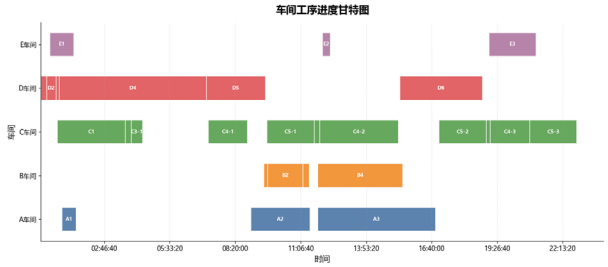


图1 双班组条件下的车间工序进度

Figure 1 Workshop operation schedules with two teams

E 车间在单班组时最早完成，双班组时却延后至75957 s。总体工期缩短不保证每个车间都提前，因为目标仅约束最大完工时间。图1呈现双班组工序进度，空档受设备可用性与运输衔接共同影响。

（四）设备购置方案与瓶颈变化

表5的方案向班组1、2分别增购7、2台设备，共9台、48万元，使用预算的96%。余款2万元低于最低单价3.5万元，既定组合无法继续增购，但不代表已达到成本与工期的最佳权衡。

表5：新增设备的数量与支出

Table 5: Equipment acquisition quantities and expenditure

设备类型	班组1/ 台	班组2/ 台	合计 / 台	支出 / 万元
自动化输送臂	2	0	2	10.0
工业清洗机	1	1	2	8.0
精密灌装机	2	0	2	7.0
自动传感多功能机	1	0	1	8.0
高速抛光机	1	1	2	15.0
合计	7	2	9	48.0

抛光设备从2台增至4台，传感设备从2台增至3台，其余类型各增2台。C 车间提前至46730s，较双班组减少35450s；E 车间以54360s 成为末完工车间。图2显示 C 车间重复工艺跨度缩小，E 车间末端检测决定当前方案结束时刻。该变化是排程观察，不能替代增购边际效应试验。

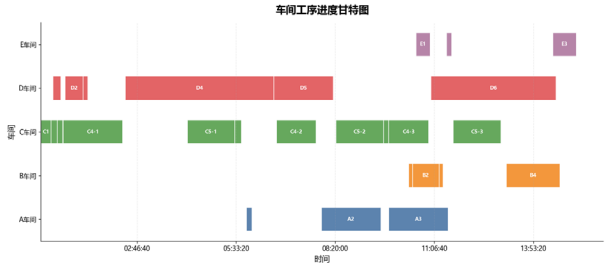


图2：增购设备后的车间工序进度

Figure 2: Workshop operation schedules after acquisition

原有与新增设备累计作业约101.8、39.4 h。双班组增购后，设备累计运输时间由66200 s 增至99105 s，但总工期下降。多台设备可并行移动，累计运输时间不能直接从工期扣除；实施时仍需评估相应运输能力。

（五）可行性核验与收敛记录

三方案分别含96、182、316条设备作业记录，均覆盖27道工序，未发现设备重叠、转运时间不足或后继提前开工。按单价与数量复算购置费为480000元。

复算发现购置情景 D4 由3台抛光机处理1500 m^3，单台效率120 m^3/h，理论耗时15000 s，记录为15001 s，源于浮点计算后向上取整。该保守误差不影响工程量完成，保留原作业区间及54360 s 工期，其余分项时长与式（1）一致。

粒子群、遗传和蚁群分别在第73次、第75代和第48次首次达到最终报告值，此后在设定轮数内未改善。这说明单次搜索趋于平台，尚不足以评价初始化敏感性或跨运行稳定性。

五、讨论

设备增配应结合工艺位置与负荷判断。抛光和检测被多车间反复调用，增加普通设备未必能消除排队。采购前应识别紧缺类型与可替代性，再用完整排程评价候选组合。增购后末完工车间从 C 转为 E，提示管理重点需随配置调整；若车间恢复生产期限不同，应加入完工上限或延期惩罚。

线性能力、同步开工和确定性运输假设限制了结果外推。现场存在作业面拥挤、协同损耗或道路容量约束时，设备翻倍未必使工时减半。应用前应由作业记录校准数量与效率关系，并考虑准备时间和安全裕量，不能直接套用本算例的工期降幅。

研究尚缺少同一资源情景下的多算法等计算预算比较，也未开展多随机种子试验、预算梯度与效率扰动分析。因此，不能报告统计显著优势或普遍适用的最优购置比例。后续可采用混合整数规划或约束规划验证小规模实例，在固定资源条件下重复比较算法，再分析资源变化的净效应。对于设备故障和订单取消，可借鉴动态重调度研究[9]；人员因素、能源消耗与运行韧性亦应纳入扩展评价[12]。

六、结论

本文将工序先后、同类设备并行、异类协同、分项释放和跨车间转运纳入统一调度模型，以设备集合变化描述班组共享及预算购置，得到可落实至设备起止时刻的计划。

在给定假设与单次搜索下，单班组、双班组和增购方案工期为159574、82180、54360 s，后两者较前情景缩短48.50%、33.85%，增购支出48万元满足预算。基准情景的抛光和检测设备需重点关注，增购后末完工车间由 C 转为 E。明细核验支持可行性，下界比较则表明尚缺全局最优证明。

该方法可用于计划阶段比较设备配置方案。实际采购仍需结合现场约束、重复试验和同情景基准比较，补充解质量与长期价值的证据。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

参考文献

[1]HARTMANN S, BRISKORN D. An updated survey of variants and extensions of the resource-constrained project scheduling problem[J]. European Journal of Operational Research, 2022, 297(1): 1-14.

[2]DING H, ZHUANG C, LIU J. Extensions of the resource-constrained project scheduling problem[J]. Automation in Construction, 2023, 153: 104958.

[3]GÓMEZ SÁNCHEZ M, LALLA-RUIZ E, FERNÁNDEZ GIL A, et al. Resource-constrained multi-project scheduling problem: A survey[J]. European Journal of Operational Research, 2023, 309(3): 958-976.

[4]DAUZÈRE-PÉRÈS S, DING J, SHEN L, et al. The flexible job shop scheduling problem: A review[J]. European Journal of Operational Research, 2024, 314(2): 409-432.

[5]YAN J, LIU Z, ZHANG C, et al. Research on flexible job shop scheduling under finite transportation conditions for digital twin workshop[J]. Robotics and Computer-Integrated Manufacturing, 2021, 72: 102198.

[6]BERTEROTTIÈRE L, DAUZÈRE-PÉRÈS S, YUGMA C. Flexible job-shop scheduling with transportation resources[J]. European Journal of Operational Research, 2024, 312(3): 890-909.

[7] 张国辉 , 蔡翌豪 , 李志霄 , 等 . 考虑有限 AGV 运输资源的柔性作业车间调度研究 [J]. 中国机械工程 ,2025,36(8):1811-1823.

[8]FAN J, ZHANG C, LIU Q, et al. An improved genetic algorithm for flexible job shop scheduling problem considering reconfigurable machine tools with limited auxiliary modules[J].Journal of Manufacturing Systems, 2022, 62: 650-667.

[9]CHEN R, WU B, WANG H, et al. A Q-Learning based NSGA-II for dynamic flexible job shop scheduling with limited transportation resources[J]. Swarm and Evolutionary Computation, 2024, 90: 101658.

[10]KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of ICNN'95 - International Conference on Neural Networks. Perth, WA, Australia: IEEE, 1995, 4: 1942-1948.

[11]DORIGO M, GAMBARDELLA L M. Ant colony system: A cooperative learning approach to the traveling salesman problem[J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1): 53-66.

[12]DESTOUET C, TLAHIG H, BETTAYEB B, et al. Flexible job shop scheduling problem under Industry 5.0: A survey on human reintegration, environmental consideration and resilience improvement[J]. Journal of Manufacturing Systems, 2023, 67: 155-173.

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

外贸风险预警的政企协同法治保障研究

吴乔林，章津
广州大学 法学院，广东 广州 510000
DOI:10.61369/ASDS.2026100010

摘　　要： 当前全球贸易格局深度调整，外贸风险已从传统的关税博弈演变为涵盖合规、标准、环保、数据跨境等多维度的系统性挑战。本文以政企协同为切入点，探讨如何通过法治化路径完善外贸风险预警机制。研究发现，我国外贸风险预警的政企协同已形成一批可借鉴的地方实践——义乌的合规指引、安吉的分级监测、绍兴的反馈处置等，均验证了政企联动的实际价值。但整体而言，现有机制仍存在突出短板：政府帮扶重资金补贴、轻信息服务，供需结构性错配明显；预警信息碎片化、滞后化，缺乏统一闭环管理；中小微企业的信息弱势地位缺乏制度性保障；前端预警与事后救济相互割裂，尚未形成全周期风险管理体系。上述问题的核心根源，在于政企协同的权责边界、运行规则与责任约束尚未上升至法治层面。基于此，本文提出以法治路径完善现有机制：一是以法定形式划定政企权责边界，推动政府履职重心从“事后补贴”向“事前预警＋事后救济并重”前移；二是健全多元主体的责任约束与权利救济体系，推动预警机制与出口信用保险、商事纠纷化解等事后救济渠道深度衔接，形成风险防控的完整闭环。

关 键 词： 外贸风险预警；政企协同；法治保障；中小外贸企业；问卷调查；地方实践

Research on the Legal Guarantee of Government-Enterprise Collaboration in Foreign Trade Risk Early Warning

Wu Qiaolin, Zhang Jin
School of Law, Guangzhou University, Guangzhou, Guangdong 510000

Abstract: The current global trade landscape is undergoing profound adjustments, and foreign trade risks have evolved from traditional tariff competition into systemic challenges covering multiple dimensions such as compliance, standards, environmental protection, and cross-border data flows. This paper takes government-enterprise collaboration as the entry point to explore how to improve the foreign trade risk early warning mechanism through a legal pathway. The study finds that a number of local practices worth learning from have emerged in China's government-enterprise collaborative foreign trade risk early warning system — such as Yiwu's compliance guidelines, Anji's tiered monitoring, and Shaoxing's feedback handling mechanisms — all of which verify the practical value of government-enterprise linkage. However, overall, the existing mechanism still suffers from prominent shortcomings: government assistance focuses heavily on financial subsidies while neglecting information services, resulting in a clear structural mismatch between supply and demand; early warning information is fragmented and delayed, lacking unified closed-loop management; the information disadvantage of small and micro enterprises lacks institutional safeguards; and front-end early warning is disconnected from post-event relief, failing to form a full-cycle risk management system. The root cause of these problems lies in the fact that the boundaries of rights and responsibilities, operational rules, and accountability constraints of government-enterprise collaboration have not yet been elevated to the level of the rule of law. Based on this, this paper proposes to improve the existing mechanism through a legal pathway: first, delineate the rights and responsibilities of government and enterprises by statutory means, shifting the focus of government performance from "post-event subsidies" to "pre-event early warning combined with post-event relief"; second, establish a sound system of responsibility constraints and rights remedies for multiple stakeholders, and promote the deep integration of the early warning mechanism with post-event relief channels such as export credit insurance and commercial dispute resolution, thereby forming a complete closed loop for risk prevention and control.

Keywords: foreign trade risk early warning; government-enterprise collaboration; legal guarantee; small and medium-sized foreign trade enterprises; questionnaire survey; local practices

基金项目：本文为第二十届“挑战杯全国大学生课外学术科技作品竞赛”广州大学校赛参赛作品《国际外贸政策变化对我国外贸型企业的影响与应对》的最终成果之一。
作者简介：
吴乔林，广州大学法学院，本科在读；
章津，广州大学法学院，本科在读。

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

引言

2024年以来，全球贸易环境持续承压。关税壁垒反复加高、非关税措施快速扩张、地缘政治冲突向经贸领域传导，外贸企业面临的风险已经从单一的价格竞争升级为涵盖合规、标准、环保、数据跨境等多维度的系统性挑战。联合国贸发会议《2025贸易与发展报告》指出，全球价值链加速重构，地缘政治已成为国际贸易首要风险源，非关税措施的贸易成本已显著高于关税本身，中小企业遭受巨大冲击[1]。世界经济论坛《2026全球风险报告》将地缘经济对抗列为年度第一大风险，50%的受访专家预判未来两年全球形势将处于“动荡”或“风暴”状态[2]。经合组织（OECD）2025年首份《供应链韧性评估》进一步揭示，全球进口集中度较上世纪90年代大幅攀升，单纯将生产回迁国内将导致全球实际GDP损失超5%[3]。

在这种背景下，外贸风险预警的重要性不言而喻——提前获取政策变动信息、准确研判风险等级、及时调整经营策略，已经成为企业稳定经营的刚需。然而，当前政企之间的风险预警协同仍存在明显短板：政府侧的帮扶多以出口补贴、融资扶持等“硬性资源补给”为主，前置性的信息预警与法律服务供给严重不足；企业侧则因信息渠道有限、政策解读能力薄弱，面对快速变化的国际贸易规则往往反应滞后。2026年3月国务院公布的《关于产业链供应链安全的规定》明确将“风险监测预警”列为法定制度，标志着国家层面已经认识到这一问题的紧迫性。

立足这一现实问题，本文以政企协同为切入点，探讨如何通过法治化路径完善外贸风险预警机制。本文通过问卷调查的方式对20家中小外贸企业开展预调研，梳理浙江义乌、湖州安吉、江山、绍兴、宁波等地的地方实践经验，并综合国内外相关文献，试图回答一个核心问题：现有预警机制并非空白，但为何仍不足以有效保护中小外贸企业？法治保障能在其中发挥什么作用？

一、外贸风险预警政企协同的现状分析

（一）外贸企业面临的风险环境：从“关税战”到“规则战”

当前外贸企业面临的外部环境呈现“复杂”“多变”“高频”三大特征。从政策层级来看，关税调整、出口退税属于基础高频政策，影响范围覆盖几乎所有外贸主体；进口配额、外汇管理、出口信保、市场准入合规则属于常规运营政策，涉及多数企业；而反倾销调查、政治制裁、长臂管辖、电商税等特殊政策虽非普遍适用，但一旦触发，往往对企业造成严重冲击。

问卷调研结果显示，85%的受访企业表示受到国际贸易政策调整的显著影响，其中45%的企业承受了成本上升、订单流失、合规压力增大、利润下滑等负面冲击，仅15%的企业表示未受实质影响。这一结果与宏观层面的判断基本一致。傅娟、杨皓森对2024年我国外贸形势的分析指出，全球经济疫后恢复仍面临重重挑战，芯片等关键领域的“卡脖子”问题依然突出，产业断链断供风险持续处于高位[4]。OECD在《经济安全与变革中的世界》报告中进一步指出，大宗商品和农业等领域的供应链高度集中于少数国家和企业，“关键节点”被扼制的风险日益突出[5]。WTO发布的《2025世界贸易报告》也提供了佐证：针对AI相关产品的贸易限制措施从2012年的130项增至2024年的近500项，低收入经济体获取关键技术的门槛不降反升[6]。

（二）现有政企协同预警的实践探索

面对日趋严峻的外部环境，部分地方政府和行业协会已开展了一系列探索，形成了一批可借鉴的实践经验。

在信息采集与合规指引方面，浙江义乌的做法较为典型。2025年4月，义乌市档案局印发《义乌外贸企业合规治理文档控制清单指引》（义档〔2025〕2号）。该指引由档案部门联合司法局、海关、商务局、商城集团等多部门共同参与制定，系统梳理了国际贸易法律法规、行业准则及国际条款。这一实践的价值在于将分散的政策信息整合为企业可直接使用的合规文档，有效降低了中小企业的信息获取成本[7]。

在风险监测与分级预警方面，湖州安吉县的“订单+清单”预警监测系统值得关注。作为浙江省首批五个试点县之一，安吉县自2018年起全面推进以家具行业为重点的监测预警系统建设。截至2019年9月底，该系统已上线外贸企业595家，累计填报订单23620笔，涉及金额6.81亿美元。系统运行中，安吉县商务局监测到嘉瑞福（浙江）家具有限公司出口美国市场销量比重超70%的集中风险后，立即上门核实并召开专题会议研究对策，协助企业通过产品升级、参加国际展会、拓展跨境电商等途径开拓新兴市场，最终使出口回归稳定[8]。

在事后应对与反馈方面，绍兴县的经验具有参考意义。2012年，当地连续遭遇4次贸易摩擦，涉案企业48家、涉案金额达1.81亿美元。为此，绍兴县商务局采取多项举措帮扶出口企业：从2012年2月起在全县开展外贸风险预警论坛活动，已举办5期、培训企业300多家；组织400多家企业参加各级贸易摩擦应对专题培训；依托网络开辟“预警平台”，将最新对华贸易壁垒情况第一时间通报企业；8月初出台《关于进一步加大外贸出口扶持力度的补充意见》，升级外贸支持力度。这些举措有效增强了企业应对国际贸易壁垒和规避贸易风险的能力[9]。

（三）现有机制的不足与完善必要性

上述实践表明，外贸风险预警的政企协同已有不少地方经验和技术工具。但与此同时，现有机制的有效性和覆盖面仍存在明显不足，机制完善具有迫切的现实必要性。

第一，政府帮扶“给资金多、给信息少”，结构性错配突出。问卷显示，70%的受访企业仅享受出口补贴、融资扶持等基础支持，只有10%的企业叠加了海外政策预警推送，仅1家企业（5%）获得了法务咨询与纠纷维权服务。而企业端的核心诉求恰恰相反：75%的企业最需要常态化外贸新政预警机制，70%的企业需要分国家、分品类的定制化合规提示。供给与需求之间的错位

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

说明，政府的履职重心仍停留在“事后补贴”，尚未有效前移到“事前预警”。

第二，预警信息呈现“碎片化”与“滞后化”特征，缺乏统一的闭环管理。现有地方实践多为自发探索，尚未形成全国统一的信息采集、过滤、发布与反馈机制。不同地区、不同行业的预警标准参差不齐，信息更新频率各异，企业难以获得稳定、可预期的预警服务。其后果是：分散的信息难以转化为有效的行动指引，预警的“信号”无法及时到达需要的人手中。正如王文中等学者在综述中所指出的，中国外贸损失率是发达国家的20倍，年损失超过万亿元，而现有研究在风险识别、评估与防范的系统性方面仍存在明显不足[10]——这既说明风险的严重性，也说明预警机制的系统性缺陷是长期存在的结构性问题。

第三，中小企业的信息弱势地位缺乏制度性保障。大型企业与政府交往密切、配套服务齐全，能够较早获取政策信息并作出调整；而中小微企业既缺乏信息渠道，也缺乏政策解读能力，往往在风险已经显现后才被动应对。问卷显示，多数企业的应对措施仍停留在开拓新市场、补充合规认证等表层，未能对深层模式和产品做出调整。OECD在供应链韧性评估中特别强调，中小企业因资源有限，在供应链中断中恢复最慢、受损最重，需要公共政策给予针对性的制度支撑[3]。

第四，“事后救济链条”与前端预警相互割裂，未能形成全周期管理。预警的价值在于防患于未然，但外贸风险不可能完全消除。胡涛基于义乌平台12个典型案例的研究表明，中小外贸企业信用风险传染问题突出，一家企业的违约可能通过供应链赊销、担保链等渠道迅速波及其他企业[11]。这意味着风险一旦突破预警防线，如果没有及时的事后救济机制衔接，就可能引发连锁反应。新冠疫情对全球供应链的冲击同样揭示了类似教训——Hung、Chuang与Suresh（2026）在涵盖多国供应链中断案例的系统综述中指出，拥有更强公共治理基础和预警协调机制的经济体，供应链恢复速度明显更快——韩国、德国等国家和中国台湾地区凭借高效的检测追踪能力和早期预警机制，有效缩短了供应链中断时间[12]。这说明，预警体系强则恢复快，预警与救济脱节则损失放大。前端预警与事后救济必须统一设计、一体推进，才能真正形成风险管理的完整闭环。

综上，现有外贸风险预警机制的真正问题在于：职责边界不清、信息流转不畅、中小企业保障不足、预警与救济脱节。地方实践经验已经证明了政企协同预警的价值，但要将这些分散的、自发的探索上升为稳定、可预期、可追责的制度安排，就必须通过法治化路径明确政府的预警职责、企业的配合义务、行业协会的过滤权限，以及信息迟报漏报的问责机制和企业的救济渠道。这正是下一部分机制完善方案所要回应的核心问题。

二、机制完善方案——以法治补齐现有预警体系短板

（一）明确政企权责：把“谁该做什么”固定为法定义务

现有预警机制并非“空白”，许多外贸企业都在有意识地建立健全预警机制、防范风险，但是在此过程中大部分政府部门只起到了呼吁的作用，并未能够提供更多有效预警支持，政企职责模糊、激励错位的问题时有发生。我们团队的问卷调研显示，有30%的企业未享受任何政府或协会帮扶，70%的企业仅享受出口补贴、融资扶持等基础支持，只有10%的企业有叠加海外政策预警推送、并且仅1家企业（5%）获得法务咨询与纠纷维权服务。这一结构说明，政府支持明显偏重“资金补贴”这一硬性资源补给，前置性、保障性的涉外风控与法律服务严重不足，中小微企业覆盖面存在缺位。由此，完善外贸风险预警政企协同法治保障机制的首要任务，是要把政府和企业的权利、义务与边界用法固定下来。

第一，政府应当从“事后补贴”转向“事前预警 + 事后救济并重”。依据新修订《对外贸易法》关于建立健全货物、技术、服务贸易预警体系与贸易调整援助的规定，应将“统一发布预警、审核信息风险等级、承担企业报送数据保密义务”列为商务主管部门的法定职责[13]。在具体履职层面，政府部门应当主动作为，打通信息供给环节的全链条。第一点，就是要统一信息采集。针对关税调整、绿色壁垒、反倾销调查等各类纷繁复杂的信息，政府要统筹收集整理，我们可以参考义乌《外贸企业合规治理文档控制清单指引》建立标准化合规信息档案[13]，为外贸企业，特别是中小微企业提供可靠的信息来源，破解单个企业信息收集能力不足的问题。其次，第二点是要做好信息过滤工作，广泛招纳行业协会、涉外法律人才参与风险政策解读，对于外贸企业关注度较高的各类政策做好合规翻译，在此基础上，灵活运用量化工具，如易靖韬、严欢构建的“小波分解 + ARIMA-GRU混合模型”，以贸易竞争力指数刻画外贸风险，为外贸风险管理提供更加科学的依据[14]。最后则是要规范预警的发布，以法律形式明确风险分级标准与推送时限，要求政策变动后按照风险等级在24-48小时内通过邮件、短信、小程序等方式推送给外贸企业，及时告知以便让他们做好充足准备，未雨绸缪，合理应对。在这个环节，我们可以参考安吉县的“订单 + 清单”预警监测系统[8]和江山“导航中心”模式[15]的两个成功经验，他们都是通过政府的合理干预，成功规避了外贸企业运营的各种国际风险。

第二，企业从“被动接收”转为“双向互动”。企业一方面享有免费获取预警信息、对信息准确性申请复核的权利；另一方面需要承担如实报送外贸经营数据、配合行业风险调研的义务。浙江义乌的成功实践已经证明了企业配合政府进行数据收集整理的重要性：义乌市档案局联合当地企业，深入调研外贸风险防控需求，系统梳理国际贸易法律法规、行业准则及国际条款，发布了《义乌市外贸企业合规治理文档控制清单指引》，这一文件成为外贸企业合规经营的“操作指南”和“风险预警手册”，助力全市外贸企业防范经营风险，强化合规治理文档管理的规范性与有效性[16]。同时企业应当承担反馈的义务，针对预警信息偏差，国际外贸运营中出现的新型风险，应当及时向相应主管部门反映，推动政企联动的外贸预警体系不断迭代升级。绍兴县和县商务局合作应对贸易摩擦的经验[9]更是进一步表明，企业主动反馈风险线索，能够帮助主管部门快速研判预警体系缺陷导致的个案问题，将个案处置经验沉淀为可广泛适用的群体经验，提升全行业的抗风险能力，在下一次问题来临时更好应对和处理，甚至是未雨绸

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

缪，把风险扼杀在摇篮里。

（二）完善责任与救济：让机制有锋芒

权责法定的落地必须配套以合理的责任约束与权利救济机制，没有它的兜底，制度要求极易流于纸面，成为空中楼阁，也难以持续激发政企双方协同的动力，对外贸行业发展的促进也只会成为泡影。针对各类主体，应当在划定权责边界的同时，搭建多元化的救济体系，形成权责统一、追责有据、救济有力的闭环。

第一，政府责任。未依法发布、迟报漏报风险预警信息致企业重大损失的，应纳入行政问责；泄露企业重要数据的，应严格依照数据安全相关法律法规追责。宁波中院对外贸企业常见法律风险进行提示、强调要“未雨绸缪做好防范”[17]，这个先例可作为政府预警发布义务的具体参照；相应地，对政府信息发布不及时、信息发布错误的情况应配套救济，明确权责须一致，用权受监督，侵权需赔偿 / 救济。

第二，行业协会责任。行业协会承担信息辅助研判、传导行业诉求的协同职能，对于篡改、截留预警信息，利用信息优势谋取不当利益的行为，应依照社团管理规定及反垄断规则予以处理。考虑到风险研判本身存在客观不确定性，为打消协会履职顾虑，我们仅在协会存在故意或重大过失并直接造成企业经济损失时，才要求其承担赔偿责任，兼顾制度约束与行业参与积极性。

第三，企业责任。企业虚报瞒报经营数据致行业预警失真的，要记入外贸主体信用档案；滥用预警信息从事投机活动，扰乱外贸秩序的，依照对外贸易管理相关规定处理。同时兼顾中小微企业现实处境，建立报送环节的商业秘密保护机制与特殊情形免责条款，消除企业“不敢报、不愿报”的心理障碍，保障预警数据的真实性与全面性。

三、总结

当前全球贸易格局深度调整，外贸风险已从传统关税博弈转向多维度规则竞争，中小外贸企业面临的合规压力与经营风险持续攀升，构建高效风险预警机制已成为稳外贸的重要抓手。

本文结合问卷调研与地方实践案例分析发现，我国外贸风险预警政企协同已形成一批可借鉴的地方探索成果，义乌合规指引、安吉分级监测、绍兴反馈处置等实践，均验证了政企联动在风险防控中的实际价值。但整体而言，现有机制仍存在突出短板：政府帮扶重资金补贴、轻信息服务，供需结构性错配明显；预警信息碎片化、滞后化，缺乏统一闭环管理；中小微企业信息弱势地位缺乏制度支撑；前端预警与事后救济相互割裂，尚未形成全周期风险管理体系。这些问题的核心根源，在于政企协同的权责边界、运行规则与责任约束尚未上升至法治层面，导致制度稳定性与可预期性不足。

针对上述短板，本文提出以法治路径完善现有机制的核心思路：一是以法定形式划定政企权责边界，推动政府履职重心从“事后补贴”向“事前预警 + 事后救济并重”前移，统筹信息全链条公共服务供给，引导企业从被动接收转向双向互动，夯实协同的数据基础与反馈机制；二是健全多元主体的责任约束与权利救济体系，区分政府、行业协会、企业的责任边界与归责标准，兼顾制度刚性与参与活力，推动预警机制与出口信用保险、商事纠纷化解等事后救济渠道深度衔接，形成风险防控完整闭环。

外贸风险治理是长期动态迭代的过程。以法治固化政企协同的制度成果，既能提升预警机制的稳定性与可追责性，也能为中小外贸企业提供稳定制度预期，最终为我国外贸高质量发展筑牢系统性风险防线。

附件1

国际外贸政策变化对我国外贸型企业的影响与应对调查问卷

第一套：企业版调查问卷（共15题）

卷首语

尊敬的企业负责人 / 外贸经理：

您好！我们正在开展《外贸企业风险预警与政企联动》学术调研，旨在了解外贸企业面对国际贸易摩擦、海外政策变动的现实困境，调研数据仅用于学术研究，所有信息严格保密。问卷大概耗时5–8分钟，感谢您的支持与配合。

1. 贵公司获取海外外贸政策信息主要使用哪些渠道？（可多选）

A. 贸促会平台 B. 贸法通 C. 驻外使领馆公开信息 D. 行业商会 / 协会

E. 同行交流 F. 公众号、自媒体 G. 企业海外办事处 H. 其他 ____

2. 您使用贸法通这类官方政策平台的频率？（单选）

A. 几乎每天 B. 每周数次 C. 每月偶尔查阅 D. 很少使用 E. 从未使用

3. 若较少使用官方政策平台，主要原因是？（可多选）

A. 信息量大筛选困难 B. 专业术语多理解门槛高 C. 信息与自身业务不匹配

D. 信息更新时效性不足 E. 缺少专职人员解读 F. 其他 ____

4. 贵司风控与外贸法律事务配置情况？（单选）

A. 设置专职外贸法务岗位 B. 无专职法务，长期外聘外部法律顾问

C. 仅内部人员兼职处理 D. 没有任何法务 / 法律顾问相关配置

5. 企业内部是否有专人负责监测、解读海外关税、合规政策变化，并输出业务执行指引？（单选）

A. 有固定专人，定期输出业务指引 B. 有兼职人员，不定期整理信息

C. 有收集信息，但不会转化为业务行动指南 D. 完全没有相关工作

6. 贵企业向政府反馈外贸相关问题的主要渠道？（可多选）

A. 政府组织企业座谈会 B. 线上政务平台留言反馈 C. 通过行业协会代为转达

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

D. 直接电话 / 线下对接业务主管部门 E. 几乎没有渠道，未曾反馈 F. 其他 ____

7. 您如何评价现有政企沟通渠道的实际效果？（量表，1完全无效；2效果较差；3一般；4效果较好；5效果很好）

□1 □2 □3 □4 □5

8. 贵企业是否通过行业协会 / 企业联盟获取合规支持、联合风险应对经验？（单选）

A. 经常，获得较多支持 B. 偶尔，有少量帮助 C. 有参与，但几乎得不到支持

D. 从未通过协会获取相关支持

9. 对比政府、行业协会、外部律师三类外部支持，您最信任哪类主体提供的外贸风险信息？（单选）

A. 政府部门（信息权威全面） B. 行业协会（贴合行业实际） C. 外部律师（专业法律视角）

D. 三者差异不大 E. 都不太信任

10. 对于政府推出分行业、分国别“避坑式”分类合规指引，您的期待程度？（量表：1完全不期待；2较低期待；3一般；4较高期待；5非常期待）

□1 □2 □3 □4 □5

11. 如果政府 / 第三方提供付费订阅式分类合规指引服务，您企业是否愿意付费？（单选）

A. 愿意付费 B. 看价格再决定 C. 不愿意付费

12. 遭遇境外突发风险（海关扣货、客户拒付、合规调查、海外新规突变），企业第一优先求助的外部对象？（可多选）

A. 行业协会 B. 贸促会 C. 政府主管部门 D. 涉外律师 E. 同行企业 F. 尽量不向外求助

13. 发生外贸纠纷后，贵司过往处理方式有哪些？（可多选）

A. 双方协商和解 B. 商事仲裁 C. 涉外诉讼 D. 未实际遭遇过外贸纠纷

14. 发生外贸风险事件之后，企业是否会开展复盘，调整风控、市场开拓策略？（单选）

A. 每次都会复盘调整 B. 重大风险才复盘调整 C. 简单总结，很少调整制度策略

D. 基本不复盘

15.【问答题（仅1题）】结合企业实际，您对政府外贸风险预警服务有哪些建议？（如推送方式、内容、精准度等）

问卷结束，感谢您填写！

第二套：行业协会版调查问卷（共 15 题）

卷首语

尊敬的协会会长 / 秘书长 / 负责人：

您好！本问卷来自广州大学法学院课题调研，主题为外贸企业风险预警与政企协作机制，调研聚焦行业协会在企业风险预警、政企桥梁中的实际运行情况。问卷内容仅用于学术研究，严格脱敏保密，感谢您拨冗填写。

1. 在政企分工视角下，您认为行业协会的核心职能与政府外贸管理部门核心职责的区分度如何？（量表：1完全无区分；2区分较小；3一般；4区分较清晰；5边界十分清晰）

□1 □2 □3 □4 □5

2. 政府向贵协会传递外贸风险预警、海外政策红利信息的主要方式？（可多选）

A. 正式文件下发 B. 工作会议口头通报 C. 政务系统线上推送 D. 几乎不会推送相关信息

E. 其他 ____

3. 贵协会将政府预警信息过滤、筛选后精准推送给对口行业企业的落实情况？（单选）

A. 常态化完成信息筛选，定向推送企业 B. 简单转发原始信息，不做筛选加工

C. 仅选择性转发部分重要信息 D. 基本不转发政府预警信息

4. 协会开展信息筛选、精准推送企业工作，遇到的主要困难？（可多选）

A. 缺少专业解读人员 B. 企业数量多，分类匹配难度大

C. 信息时效性不足 D. 企业接收意愿参差不齐 E. 暂无明显困难 F. 其他 ____

5. 贵协会招聘外贸风控、合规相关工作人员，重点看重哪些能力？（可多选）

A. 外贸业务实操经验 B. 涉外法律专业背景 C. 多语种语言能力

D. 熟悉海外各国贸易政策 E. 行业产业知识储备 F. 其他 ____

6. 贵协会与外部律所 / 法律机构开展外贸合规、风险处置合作现状？（单选）

A. 建立长期稳定合作机制 B. 有零散临时项目合作 C. 有接触，但几乎无实质合作

D. 完全没有开展相关合作

7. 协会与律所开展合作主要存在哪些壁垒？（可多选）

A. 经费资金限制 B. 双方工作目标不一致 C. 专业能力匹配度不足

D. 企业付费意愿不足 E. 不存在明显壁垒 F. 其他 ____

8. 贵协会与政府部门在外贸风险预警业务上协作机制成熟度？（量表：1很不成熟；2较不成熟；3一般；4比较成熟；5非常成熟）

□1 □2 □3 □4 □5

9. 您认为外贸企业主动向协会反馈海外外贸动态、风险线索的意愿如何？（量表：1完全不愿意；2意愿低；3一般；4意愿较高；5意愿很高）

□1 □2 □3 □4 □5

10. 从客观条件看，贵协会承接企业外贸风险信息收集、上报反馈职能的能力如何？（单选）

A. 完全可以承接该职能 B. 基本可以承接，但存在部分短板

C. 勉强承接，条件缺口较大 D. 尚不具备承接条件

11. 贵协会向企业传递政策、风险预警信息常用渠道？（可多选）

A. 协会公众号 / 微信群 B. 行业会议、线下讲座培训 C. 定向

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

邮件推送

D. 企业会员一对一通知 E. 很少主动推送信息

12. 在企业遭遇涉外贸易纠纷时，协会能够提供哪些支持？（可多选）

A. 政策信息解读 B. 对接法律资源 C. 组织企业集体协商应对

D. 向政府代为反映企业诉求 E. 基本无法提供实质支持

13. 当前协会开展外贸风险预警服务，最大短板是？（可多选）

A. 专业人才短缺 B. 运营经费不足 C. 获取海外一手政策信息渠道有限

D. 企业会员参与度不足 E. 和政府联动衔接不畅 F. 其他 ____

14. 您认为企业最希望协会新增哪一类外贸相关服务？（单选）

A. 国别合规风险专题解读 B. 纠纷调解与法律资源对接

C. 行业集体对外维权组织 D. 政策信息定期推送 E. 其他 ____

15.【问答题（仅1题）】为完善外贸风险预警、助力企业出海，您认为行业协会还可以从哪些方向优化改进？

问卷结束，感谢您填写！

第三套：政府部门版调查问卷（共15题）

卷首语

尊敬的政府部门工作人员：

您好！本问卷为广州大学法学院学术课题调研，研究主题为外贸企业风险预警与政企协作机制，调研旨在了解外贸风险信息传导、企业反馈渠道建设现状。问卷所有内容仅用于学术研究，不对外公开，感谢您的支持。

1. 当前面向外贸企业，是否设立专门的外贸政策咨询、风险问题反馈渠道？（单选）

A. 已设立多条常态化专门渠道 B. 有通用政务渠道，无专门外贸专项窗口

C. 仅有临时性对接方式 D. 暂未设置相关反馈渠道

2. 外贸风险预警信息主要触达哪类企业群体？（可多选）

A. 大型头部外贸企业 B. 中型外贸企业 C. 中小微型外贸企业 D. 各类企业基本均等覆盖

3. 目前预警信息触达中小企业的效果评价（量表：1效果很差；2效果较差；3一般；4效果较好；5效果很好）

□1 □2 □3 □4 □5

4. 企业遇到海外政策、贸易纠纷问题，向政府部门反馈诉求的主要路径？（可多选）

A. 政务平台线上留言 B. 行业协会收集汇总后转报 C. 企业座谈会现场反馈

D. 电话、线下业务对接 E. 缺少清晰反馈路径

5. 企业自下而上反馈风险线索的信息传导链条主要痛点？（可多选）

A. 企业反馈积极性不足 B. 信息层级流转耗时长，时效性弱

C. 信息碎片化，难以整合加工 D. 没有明显痛点 E. 其他 ____

6. 您是否支持建立企业外贸风险信息反馈激励机制？（单选）

A. 非常支持，很有必要 B. 可以尝试试点探索 C. 必要性一般，观望即可

D. 不建议设立该类机制

7. 如果推行企业风险信息反馈激励机制，您认为可行激励方式？（可多选）

A. 政策咨询优先服务 B. 荣誉表彰 C. 补贴 / 经费激励 D. 优先推送定制化风险预警

E. 不适合物质类激励 F. 其他 ____

8. 当前政府输出的外贸风险、合规指引材料最大现实问题？（可多选）

A. 内容偏宏观，缺少细分行业、商品的实操指引 B. 专业术语多中小企业理解困难

C. 更新迭代时效性有待提升 D. 传播触达渠道有限 E. 现有材料整体质量较好，无突出问题

9. 在政企协作链条中，行业协会在风险预警工作上发挥的实际作用评价（量表：1几乎无作用；2作用较小；3一般；4作用较大；5作用突出）

□1 □2 □3 □4 □5

10. 政府部门向行业协会下发风险预警信息，希望协会完成哪些工作？（可多选）

A. 转发告知会员企业 B. 结合行业做二次解读加工 C. 收集企业风险诉求反馈回政府

D. 组织企业风险培训宣讲 E. 仅做简单传递即可

11. 当前政府、协会、企业三方风险预警协同存在哪些堵点？（可多选）

A. 信息双向流转不通畅 B. 协会专业人员力量不足 C. 中小企业接收信息主动性弱

D. 部门间内部协同不足 E. 无明显堵点 F. 其他 ____

12. 对于给中小企业提供分国别、分行业的“避坑式”合规指引，当前推进主要难点？（可多选）

A. 国别政策变化快，维护成本高 B. 细分行业、商品数量庞大工作量大

C. 专业解读人才缺口 D. 企业需求差异大，难以做到全部适配 E. 暂无明显难点

13. 在风险预警落地层面，未来更优先的工作方向？（单选）

A. 扩大预警信息覆盖广度，下沉中小企业 B. 提升内容精细化，增强实操指导性

C. 完善企业风险线索反馈闭环机制 D. 强化与行业协会协同联动

14. 对于外贸突发风险事件，现有预警响应速度是否满足企业现实需要？（量表：1完全不满足；2较不满足；3基本满足；4较好满足；5充分满足）

□1 □2 □3 □4 □5

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

15.【问答题（仅1题）】进一步完善外贸风险预警体系、助力企业稳健出海，您有哪些工作建议?

问卷结束，感谢您填写!

附件二

问卷结果分析

企业

1、政策帮扶层面

类型	具体内容	企业数	占比
未享受任何帮扶	无政策、协会服务支持	6	30%
享受帮扶	出口补贴 + 融资扶持	14	70%
	叠加海外政策预警推送	2	10%
	叠加法务咨询、纠纷维权服务	1	5%

大部分企业享受过基础资金支持，但仍有不少企业没有受到任何帮扶，受海外政策预警推送、法务咨询、纠纷维权等涉外法务保障企业少。由此可见，政府和行业的帮扶存在缺位情况，中小微企业覆盖面不足，且扶持仅停留在传统资金补贴这一硬性资源补给上，缺乏涉外风控与法律服务保障这一软性支持，前置性、保障性、赋能性服务未能很好落实。

2、企业核心诉求层面

类型	企业数	比例
常态化外贸新政预警机制	15	75%
分国家 / 品类定制化合规提示	14	70%
行业抱团共建供应链风控联盟	11	55%
产品绿色壁垒、关税新规专项辅导	11	55%
政企专项对接答疑通道	9	45%

常态化外贸新政预警机制、分国家 / 品类定制化合规提示这两个服务需求量大，是企业的核心诉求；需要行业抱团共建供应链风控联盟、产品绿色壁垒、关税新规专项辅导这两个服务的企业数超过一半，体现企业对协同和专项辅导的诉求；政企专项对接答疑通道需求量不低，体现企业渴望与政府进行对话。由此可见，当前海外政策更新变化快，企业不一定能及时获取这类信息，存在信息差。不同国家规则不同，不同品类规则不一，企业完全查收这类信息难度大，解读相关规则能力不足。单一企业难以独立应对复杂的国家外贸环境，行业协同，相互配合成为企业普遍共识。企业对政企联动需求强烈，希望在政企对话中赋能企业发展。

3、外部贸易环境层面

政策层级	具体政策	提及频次
基础高频政策	关税调整、出口退税	最高
常规运营政策	进口配额、外汇政策、出口信保、市场准入合规	较高
特殊政策	反倾销、政治制裁、长臂管辖、电商税、区域专项出口政策	中等

调整类别	具体表现	覆盖企业范围
核心调整	关税上调、市场准入合规门槛提升	全覆盖
新增管控	碳排放、ESG 环保要求、反倾销调查增加	多数企业
区域政策	欧美：高关税、海关查验、贸易制裁 欧亚：关税同盟、清关政策变动	部分定向行业 / 企业
衍生影响	海运费大幅上涨	全覆盖

贸易规则升级，传统关税壁垒持续影响企业，非传统关税快速普及，外贸行业的竞争从价格竞争转向综合实力竞争，对合规、标准、环保等方面越来越重视。贸易保护主义盛行，趋向常态化，国家市场普遍上调关税、收紧市场准入。关税、物流等多重因素影响，企业运营成本增加。

4、企业受影响层面

主要影响	影响程度	企业数	占比
定价承压、被迫调整目标市场、产品适配合规、打乱生产排期、调整海外备货模式	受明显影响	17	85%
企业布局稳定，未收海外政策波及	无实质影响	3	15%

国际贸易政策的变化对绝大部分企业造成影响，且对企业全产业链均产生影响，政策变动深刻影响企业的发展，提前预测、有效应对是企业稳定发展的重要因素。企业面向市场从单一到多元，有转向亚太等准入门槛更低的区域的趋势。

核心特征	判定类型	企业数量	占比
成本抬升、订单流失、合规受阻、利润下滑	负面冲击	9	45%
行业出清，头部企业抢占市场，轻资产 / 差异化模式优势凸显	转型升级机遇为主	8	40%
短期承压，长期具备转型空间	冲击与机遇并存	3	15%

负面影响与正面影响占比不相上下，负面略胜一筹，企业受困境的几率大，而部分企业看到政策变化带来的转型机遇。不断抬高的市场门槛对中小微企业不友好，生存空间被大型企业挤压，订单、资源也流向头部企业。市场多元化布局、企业差异化经营具备竞争优势。

5、企业应对层面

调整方向	具体措施	成效反馈	企业占比
市场布局	市场多元化、开拓新区域、设立海外仓 / 子公司	风险分散，效果较好	多数企业
产品合规	获取准入认证、选用合规原材料、适配环保标准	合规能力提升，效果中等	多数企业
供应链 / 贸易模式	转口贸易、调整产地、优化结算方式、动态定价	小幅减缓压力，效果一般	部分企业
内部管理	规范报关、票据、汇率风险	小幅降险，效果有限	少数企业
未调整	无应对		少数企业

多数企业采取开拓海外市场、补充合规认证等手段，仅停留在基础、表层，未对深层模式、产品等方面做出调整。政企深度协作、行业联合风控等顶层配套机制未完善，难以抵御系统性外部风险，故企业成效有限。

补充：在表格横向对比中，我们发现极少数企业受政策变化的影响不大，可以看出大型企业与政府交往密切，配套服务齐

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

全，受外部环境干扰少。而中小企业绝大多数受影响大。

专家

1、海外贸易政策调整对我国外贸企业的影响

类型	具体内容
成本层面	进出口关税上涨，企业经营成本显著增加
市场层面	产品准入门槛提高，海外市场拓展受阻，海外订单、合作项目缩减
运营层面	合规审核、海关监管流程复杂化，企业运营效率下降
发展层面	外部政策倒逼企业优化产品结构、加大研发投入，提升综合竞争力

2、海外贸易管控主要发力方向

类型	具体内容
关税壁垒	多国大范围加征各类进口关税
配额与准入壁垒	设置贸易配额、市场准入限制等显性壁垒
标准壁垒	抬高产品质量、安全、环保、劳工相关准入标准
合规管控壁垒	强化供应链、跨境数据、海关全流程监管审查
技术管制壁垒	限制高端技术、精密设备、芯片等核心品类出口

3、国际外贸政策频繁调整的原因

类型	具体内容
经济因素	各国优先保护本土产业，国际市场竞争加剧，全球经贸格局重构
地缘政治因素	国际政治博弈、地区冲突频发，贸易阵营化特征明显
规则因素	全球环保、合规、质量等通用贸易标准持续升级
外部环境因素	疫情暴露全球供应链短板，全球经济复苏乏力且不均衡

4、海外典型贸易政策

类型	具体内容
关税类	美国对华加征关税（CBAM）、各国保护性进口关税
产品准入类	产品环保、质量、安全、劳工标准，区域市场准入条款
合规管控类	跨境数据监管、投资并购审查、转口贸易限制、供应链核查
技术管制类	美国实体清单、高端芯片 / 设备出口管制、境外资本投资限制
通用规则类	WTO 多边贸易协定、国际贸易术语、区域自由贸易协定

5、政策变动对企业的实际影响及转型案例

类型	具体内容
制造型企业	欧盟碳关税倒逼企业升级生产工艺，投入低碳技术，调整出口布局
高科技企业	技术封锁造成核心设备、芯片供给不足，加速产业链国产替代与供应链多元化
跨境电商企业	小额出口政策、关税变动冲击原有模式，企业转向布局海外仓，运营成本上升
外贸商贸企业	欧美传统市场受阻，企业转向开拓东南亚等新兴市场，开展转口贸易
配套服务企业	涉外法律等外贸配套业务萎缩，相关从业者被迫转型

6、多方主体应对策略

类型	具体内容
政府层面	落实稳外贸、减税降费政策；加大金融、海外仓基建支持；优化通关服务，开拓多元国际市场；对接国际经贸规则
行业协会层面	搭建国内外资源对接平台；实时跟踪并预警海外新政；提供合规咨询服务；推动内外贸融合发展
企业层面	加大技术、低碳领域研发；搭建多元化市场与供应链体系；强化风险管控；布局跨境电商、海外仓业态

7、疫情后外贸政策整体趋势及成因

类型	具体内容
整体趋势	贸易保护主义抬头，隐形贸易壁垒增多；贸易走向区域化、近岸友岸化；贸易逻辑从效率优先转为安全优先；绿色、技术类新型壁垒常态化；政策不确定性大幅提升
成因	疫情凸显全球供应链脆弱性，各国推动供应链本土化、短链化；大国科技与产业博弈加剧；全球经济复苏动力不足；全球绿色低碳转型催生新规则；经济民族主义兴起，多边贸易体系影响力弱化

以上数据反映出四个信息

1. 传统关税壁垒仍对大多数企业产生影响，标准、合规、环保、技术等非关税壁垒逐步叠加，新型壁垒对中小企业冲击大，中小企业继续提升自身实力的同时寻找外部帮助，以此应对外贸政策的变化。各国实行贸易保护政策，优先保障本土产业，传统跨区域长链条贸易持续收缩，企业更倾向于临近的、友好的区域。

2. 国际贸易政策的变化对企业全链条全环节均产生影响，倒逼企业转型升级，拓展海外市场，高耗能企业绿色低碳转型，企业加快自主研发，外部压力与限制下低端、低附加值的发展模式逐渐淘汰。不同行业受不同政策变化影响，行业风险发布不均。

3. 国家贸易政策变动受多种因素影响，多边贸易体系弱化，短期内外贸行业不稳定趋势不变。

4. 政府、行业协会、企业三方协作是应对国际外贸政策变动的重要因素，企业需提升核心竞争力，重视技术研发、提升合规能力、供应链多元化、进行风险管控是必要的。

参考文献

[1]UNCTAD. Trade and Development Report 2025: On the Brink – Trade, Finance and the Reshaping of the Global Economy[R]. Geneva: United Nations Conference on Trade and Development, 2025.

[2]WEF. Global Risks Report 2026[R]. Geneva: World Economic Forum, 2026.

[3]OECD. Supply Chain Resilience Review: Navigating Risks[R]. Paris: Organisation for Economic Co-operation and Development, 2025.

[4] 傅娟 , 杨皓森 .2024 年我国外贸形势分析与风险预警 [J]. 可持续发展经济导刊 ,2024.

[5]OECD. Economic Security in a Changing World[R]. Paris: Organisation for Economic Co-operation and Development, 2025.

[6]WTO. World Trade Report 2025: How Trade and AI Can Contribute to Inclusive Growth[R]. Geneva: World Trade Organization, 2025.

[7] 义乌市档案局 . 义乌市档案局关于印发《义乌外贸企业合规治理文档控制清单指引》的通知（义档〔2025〕2 号）[EB/OL]. (2025-04-03).

[8] 陈颖 . 监测预警上前 防风险稳外贸 [N]. 湖州日报 .

[9] 肖莹 . 绍兴县多管齐下应对贸易摩擦 [N]. 中国纺织报 .

[10] 王文中，汤若阳，邓旭虹. 企业外贸风险研究综述 [J]. 江苏商论 .

Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

[11] 胡涛 . 中小外贸企业的信用风险传染与防范对策研究——基于 " 义乌国际贸易综合服务及经济案事件预警平台 " 的典型案例分析 [J]. 中国商贸 .

[12]Hung T, Chuang C H, Suresh N. A systematic review and analysis of COVID-19 disruptions to supply chains[J]. International Journal of Engineering Business Management, 2026, 18.

[13] 中华人民共和国商务部. 中华人民共和国对外贸易法（2025修订）专题 [EB/OL].2026-03-01施行 .

[14] 易靖韬，严欢 . 基于小波分解和 ARIMA-GRU 混合模型的外贸风险预测预警研究 [J]. 中国管理科学 .

[15] 毛广绘，姜小武，汪礼国 . 蜜蜂跟着 " 导航指令 " 飞 [N]. 浙江日报 .

[16] 陈宇，邓英楠 . 浙江义乌 为外贸企业合规经营提供 " 风险预警手册 "[N]. 中国档案报 .

[17] 余晓辰 . 风险预警 " 护航 " 甬城外贸 [N]. 宁波日报 .

 Copyright © This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License