

DOI:10.3969/j.issn.1003-5060.2025.12.016

基于多模型联用的科技企业 R&D 评测方法与实证研究

南国君

(合肥工业大学 资源与环境工程学院, 安徽 合肥 230009)

摘要: 文章运用逐步最小二乘回归、分位数回归、灰色预测模型构成多模型联用方法, 系统评测企业研究与开发(research and development, R&D)经费投入的影响机制及未来趋势。研究表明: 基于多模型联用的评测方法, 可精准直观地评测科技企业 R&D 经费投入的现状; 地区生产水平、政府财政科技拨款、企业投入是驱动 R&D 经费增长的核心因素; 政府财政支持具有跨层级的持续有效性, 其回归系数稳定在 0.953~1.085 之间, 凸显政策端对创新投入的稳定撬动作用; 部分企业未来 5 年增速可能回调 3.2%~5.7%, 提示过度依赖单一主体的结构性风险。该文提出构建梯度化政策体系, 强化财政支持的普惠性、稳定性, 推动大中小企业协同创新等建议, 为区域创新政策制定和企业研发投入优化提供理论方法支撑。

关键词: 企业研究与开发(R&D)经费投入; 最小二乘回归; 分位数回归; 灰色预测模型; 多模型联用方法

中图分类号: F273.1

文献标志码: A

文章编号: 1003-5060(2025)12-1689-08

Empirical research on evaluation method for R&D in technology enterprises based on multi-model integration

NAN Guojun

(School of Resources and Environmental Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: This paper uses stepwise least squares regression, quantile regression, and grey prediction models to form a multi-model integrated method for evaluating the influencing mechanisms and future trends of enterprise research and development (R&D) investment. The results show that the multi-model integrated evaluation method can accurately and intuitively assess the current status of R&D investment in technology enterprises. Regional production levels, government fiscal allocations for science and technology, and enterprise contributions are the three core drivers of R&D funding growth. Government fiscal support demonstrates cross-level sustained effectiveness, with regression coefficients remaining stable in the range of 0.953-1.085, highlighting the stable leveraging role of policy-driven innovation investment. Some enterprises may experience a growth rate adjustment of 3.2%-5.7% over the next five years, signaling structural risks from overreliance on single entities. Based on these findings, the paper proposes the construction of a gradient policy system, enhancing the inclusivity and stability of fiscal support, and fostering collaborative innovation among small, medium, and large enterprises. This study provides theoretical and methodological support for regional innovation policymaking and enterprise R&D investment optimization.

Key words: enterprise research and development (R&D) investment; least squares regression; quantile

收稿日期: 2025-05-17; 修回日期: 2025-07-14

基金项目: 国家自然科学基金资助项目(62073113); 安徽省科技创新战略与软科学研究重点资助项目(202206f01050028)和 2025 年高校中青年教师培养行动资助项目(JNFX2025090)

作者简介: 南国君(1981—), 男, 甘肃会宁人, 博士, 合肥工业大学副研究员, 硕士生导师, 通信作者, E-mail: zzbngji@hfut.edu.cn.

regression; grey prediction model; multi-model integrated method

在全球新一轮科技革命与产业变革加速演进的背景下,科技创新已成为重塑区域竞争力的核心要素^[1]。安徽省作为长三角一体化发展的重要组成部分和中部地区创新发展的主体,近年来深入实施创新驱动发展战略,将“创新型省份”建设作为推动高质量发展的重要引擎^[2]。2024年安徽全社会研发投入1 264.7亿元,其中企业研发投入占比超过80%,企业在关键核心技术攻关、成果转化及产业升级中发挥显著的引领作用,推动全省区域创新能力连续多年稳居全国第1方阵。

研究与开发(research and development, R&D)作为科技创新的核心驱动力^[3],其投入规模与结构深刻影响国家和地区的技术竞争力与经济可持续发展能力^[4-5]。文献[6]通过对美国制造业企业的分析,证实R&D投入与全要素生产率存在显著正相关关系,奠定企业研发投入的经济价值理论基础;文献[7]研究R&D的投入、产出(专利数量和引用)及其相互作用与未来收益、经营现金流的水平及可变性之间的相关性;文献[8]基于2003—2018年的时间序列数据,研究高新技术产业产值与R&D投入、产业结构转型之间的关系,并进行实证分析;文献[9]采用多种计量模型证实样本考察期间政府直接R&D资助对企业R&D投入产生杠杆效应,即两者之间具有显著的正向线性关系而不存在倒U型关系;文献[10]选取2011—2020年A股非金融类上市企业为样本,研究表明地方政府补助是企业R&D投入的重要驱动因素;文献[11]利用中国23个省(自治区、直辖市)2007—2016年高新技术产业(high technology industry, HTI)面板数据,研究国外R&D对HTI创新绩效(innovation performance, IP)的影响及其空间溢出效应;文献[12]研究表明R&D经费投入可以改善科技成果转化政策实施效果,且这一激励作用呈现显著的区域异质性特征;文献[13]研究发现,R&D投入在数字化转型与企业创新绩效之间发挥重要中介作用。

现有研究多聚焦于单一模式或政策研究,但是方法创新方面的研究较少,因此难以全面捕捉R&D投入的复杂性与动态性^[14]。如普通最小二乘法(ordinary least squares, OLS)回归仅揭示平均效应,忽视条件异质性;分位数回归能剖析不同投入水平下的驱动差异但缺乏时序预测能力;灰

色预测运用有限样本预测趋势却依赖有效变量筛选。

本文创新性地构建“筛选—剖析—预测”递进式框架,首先运用逐步最小二乘回归从全局视角识别核心驱动因素;然后采用分位数回归揭示这些因素在不同R&D投入强度分位点上的异质性影响;最后聚焦关键主体,利用灰色预测模型预判其未来投入趋势。该方法整合了截面分析与实践序列预测,兼顾整体规律与结构性差异,为区域创新政策提供更系统的决策依据。

1 相关理论与方法

1.1 数据预处理

数据预处理是常用的数据统计分析方法,使研究数据更为稳定。为探究平减处理后数据的平稳性特征,综合采用ADF、PP检验开展单位根检验^[15]。作为时间序列分析中的经典检验方法,ADF、PP检验能够有效识别数据中是否存在单位根^[16],若存在单位根,则表明时间序列呈现非平稳特性,可能导致后续统计推断与模型估计出现偏差。这2种检验方法的综合运用,可以准确判断数据的平稳性状态^[17],为后续的计量分析与模型构建奠定基础。

1.2 逐步最小二乘回归法

普通最小二乘法(OLS)是一种在回归分析中广泛应用的参数估计方法,用于确定一组数据的最佳拟合直线或曲线^[18]。然而,OLS对数据中的异常值较为敏感,1个或几个极端值可能会对回归结果产生较大影响,导致估计的参数偏离真实值,降低模型的准确性和稳健性^[19]。

逐步最小二乘回归法作为变量选择的经典统计方法,在多元线性回归建模框架中,通过动态筛选自变量组合以获取最优解释变量子集^[20]。该方法以最小二乘法为参数估计基础,基于信息论准则或显著性检验标准^[21],采用前向逐步选择、后向逐步剔除或双向逐步筛选策略,在每一轮迭代中对候选自变量进行严格评估。通过不断更新回归方程结构,有效平衡模型复杂程度与拟合精度,规避因过度拟合导致的泛化能力下降问题,直至满足预设的终止条件,最终输出兼具统计显著性与稳健性的回归模型^[22]。

在多元线性回归模型中,假设因变量 y 与自变量 x_1, x_2, \dots, x_p 之间的关系为:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$ (1)
 其中: $\beta_1, \beta_2, \dots, \beta_p$ 为待估计的参数; ε 为误差项。

逐步最小二乘回归法就是通过不断调整自变量的组合估计出最优的回归系数 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$, 使得残差平方和 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 最小, 其中: y_i 为观测值; \hat{y}_i 为回归模型预测值。

1.3 分位数回归

分位数回归是另一种回归分析方法^[23]。作为 OLS 的扩展, 估计因变量的不同条件分位数与自变量之间的关系^[24], 能够更全面地描述数据的分布特征。与最小二乘回归关注因变量的条件均值不同, 分位数回归能够捕捉到因变量在不同分位数上的变化^[25], 因此在处理存在异方差性或极端值的数据时具有较强的鲁棒性^[26]。分位数回归广泛应用于经济学、金融学、医学等领域, 尤其适用对数据分布的不同部分进行细致分析。

分位数回归的一般形式如下:

$$Q_\tau(y_i | x_i) = x_i^T \beta_\tau \quad (2)$$

其中: x_i 为解释变量向量; $Q_\tau(y_i | x_i)$ 为在给定条件 x_i 下的分位数; τ 为分位数水平; β_τ 为与分位数 τ 对应的系数向量。

分位数回归通过估计一系列不同分位数下的回归系数, 为观察因变量分布的全貌提供系统分析框架^[27]。在面对具有偏斜分布特征或存在异方差性的数据时, 该方法优势显著, 无需对误差项的具体分布形式进行先验假设, 而是通过最小化特定损失函数, 直接刻画自变量对因变量不同分位点的作用机制^[28]。损失函数如下:

$$\sum_{i=1}^n \rho_\tau(y_i - x_i^T \beta_\tau) \quad (3)$$

其中, $\rho_\tau(u)$ 为分位数损失函数, 通常定义为 $u(\tau - I(u < 0))$, 损失函数考虑了分位数水平 τ , 并根据残差的正负进行调整。

1.4 灰色预测模型

灰色系统理论^[29]运用小样本、贫信息进行系统的建模与预测。GM(1, 1) 模型是其核心预测模型, 通过累加生成弱化原始序列的随机性, 挖掘隐含的系统演化规律, 模型方程为:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (4)$$

其白化方程为:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (5)$$

其中: a 为发展系数, 表示序列的衰减或增长趋势; b 为灰色作用量, 表示外部作用强度。该模型

特别适用于短期、趋势性明显但样本量有限的经济指标预测, 避免传统时间序列模型(如 ARIMA)对大样本的要求(如企业 R&D 投入强度)。

2 多模型联用的 R&D 经费投入分析

2.1 整体建模思想与框架

为了系统揭示 R&D 投入驱动机制, 构建多模型联用三阶段逻辑: 第 1 阶段为全局关键因素筛选; 第 2 阶段为异质性影响剖析; 第 3 阶段为趋势预测。多模型联用分析框架流程如图 1 所示。

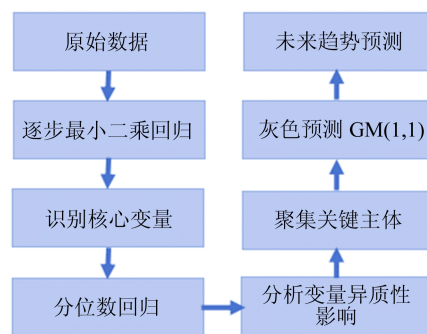


图 1 多模型联用分析框架流程

模型选择依据如下:

1) 逐步最小二乘回归。高效筛选对 R&D 投入具有显著全局影响的核心变量, 即以地区生产总值的比重表示该地区的生产水平 G_{DP} 、以相应的政府财政拨款表示政府科技投入的力度 C_Z 、以相应的科技投入金额表示企业科技投入的力度 C_Y , 避免多重共线性, 奠定了分析基础。

2) 分位数回归。深入研究核心变量在不同 R&D 投入强度条件下的差异化作用, 揭示条件异质性, 如边际效应递减、马太效应。

3) 灰色预测模型 GM(1, 1)。基于历史数据, 对分位数回归识别的关键主体(高投入企业)进行未来 3~5 年投入强度预测, 捕捉趋势性变化和潜在拐点。

其建模思想的主要创新点如下: ① 构建“筛选—分析—预测”闭环; ② 整合全局规律识别、结构差异挖掘和动态趋势预判。

2.2 构建模型方法

2.2.1 基于逐步最小二乘回归

从候选变量中识别对 R&D 经费投入(因变量 y)具有全局显著影响的核心驱动因素, 逐步最小二乘回归步骤如下。

1) 变量初始化。设候选自变量集 $X = \{x_1, x_2, \dots, x_p\}$ (包括 G_{DP} 、 C_Z 、 C_Y 等), 因变量为 y

(R&D 投入)。

2) 前向选择与后向剔除。前向选择从空模型开始, 逐轮添加使残差平方和下降为最显著的变量; 后向剔除为从全模型开始, 逐轮移除最不显著的变量; 终止条件为模型拟合优度 $R^2 < 0.01$ 或 AIC/BIC 准则不再优化; 参数估计的最终模型为:

$$y = \beta_0 + \beta_1 x_i + \beta_2 x_j + \dots + \varepsilon \quad (6)$$

输出全局显著变量子集 $X_{\text{core}} = \{G_{\text{DP}}, C_Z, C_Y\}$

和回归系数 $\hat{\beta}$ 。

2.2.2 基于分位数回归

基于筛选出的 X_{core} , 分析其对 y 在不同分位点 τ (投入强度水平) 的异质性影响。

模型设定为:

$$Q_y(\tau | X) = \beta_0(\tau) + \beta_{G_{\text{DP}}}(\tau)G_{\text{DP}} + \beta_{C_Z}(\tau)C_Z + \beta_{C_Y}(\tau)C_Y \quad (7)$$

其中, $Q_y(\tau | X)$ 为给定 X 下 y 的 τ 条件分位数。

损失函数最小化, 即

$$\min_{\beta(\tau)} \sum_{i=1}^n \rho_{\tau}[y_i - \mathbf{X}_i^T \boldsymbol{\beta}(\tau)],$$

$$\rho_{\tau}(u) = u(\tau - I(u < 0)) \quad (8)$$

输出各分位点 τ 的系数估计 $\hat{\boldsymbol{\beta}}(\tau)$ 及其显著性。

2.2.3 基于灰色预测模型 GM(1,1)

对分位数回归识别的高影响力主体进行 R&D 投入强度时序预测, 步骤如下。

1) 级比检验。原始序列为:

$$Y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n)) \quad (9)$$

级比为:

$$\lambda(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)}, \quad k = 2, 3, \dots, n \quad (10)$$

检验条件为:

$$\lambda(k) \in (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}}) \quad (11)$$

2) 累加生成。计算公式为:

$$Y^{(1)}(k) = \sum_{i=1}^k y^{(0)}(i), \quad k = 1, 2, \dots, n \quad (12)$$

3) 建立灰微分方程。方程如下:

$$y^{(0)}(k) + az^{(1)}(k) = b \quad (13)$$

其中, $z^{(1)}(k)$ 为背景值, $z^{(1)}(k) = 0.5[y^{(1)}(k) + y^{(1)}(k-1)]$ 。

4) 参数估计。计算公式为:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}, \quad \mathbf{Y} = \begin{bmatrix} y^{(0)}(1) \\ y^{(0)}(2) \\ y^{(0)}(3) \\ \vdots \\ y^{(0)}(n) \end{bmatrix} \quad (14)$$

5) 白化方程求解。计算公式为:

$$\hat{y}^{(1)}(t) = \left[y^{(0)}(1) - \frac{b}{a} \right] e^{-a(t-1)} + \frac{b}{a} \quad (15)$$

6) 还原预测值。计算公式为:

$$\hat{y}^{(0)}(k) = \hat{y}^{(1)}(k) - \hat{y}^{(1)}(k-1),$$

$$k = 2, 3, \dots, n \quad (16)$$

7) 模型检验。后验差比值计算公式为:

$$C = S_2 / S_1 \quad (17)$$

其中:

$$S_1 = \sqrt{\frac{1}{n} \sum_{k=1}^n (y^{(0)}(k) - \bar{y}^{(0)})^2};$$

$$S_2 = \sqrt{\frac{1}{n} \sum_{k=1}^n (\varepsilon(k) - \bar{\varepsilon})^2}.$$

精度等级如下: $C < 0.35$ 为优, $C < 0.5$ 为合格。

相对误差计算公式为:

$$\delta(k) = \left| \frac{y^{(0)}(k) - \hat{y}^{(0)}(k)}{y^{(0)}(k)} \right| \times 100\%,$$

$$\max \delta(k) < 20\% \quad (18)$$

3 实验与结果分析

3.1 实验数据

安徽省是全国农产品生产、能源、机械及原材料等领域的重要产业基地, 其区域创新能力连续多年位居全国前列。为了验证模型的科学精准性, 本文以安徽省典型科技创新型企业的 R&D 经费投入为实验对象。为保障样本的代表性和研究的聚焦性, 采用分层抽样方法, 依据如下: ① 在企业规模上, 覆盖龙头企业、央企、省级创新领军企业; ② 在行业分布上, 涵盖战略性新兴产业、高端装备制造企业、传统优势产业升级企业; ③ 在区位分布上, 既考虑省会合肥, 亦有芜湖、蚌埠区域; ④ 在数据完整性上, 采用连续 5 年 (2020—2024 年) 完整数据; ⑤ 在代表性方面, 所选 6 家企业覆盖安徽省三大主导产业, 且均为省级以上创新平台载体承担单位, 对其研究能有效反映全省 R&D 投入的核心机制与政策焦点。

根据研究工作实际需要, 样本数据主要来源于《安徽统计年鉴》中的 GDP, 安徽省科技经费投入统计公报中的 C_Z , 合肥、芜湖等地市财政决算报告以及相关科技企业公开年报中的 R&D 经费等。鉴于部分数据的敏感性, 本文对相关数据进行了脱敏处理, 企业 A、B、C、D、E、F 分别代表 6 家科技企业, 涵盖其 2020—2024 年政府财政拨款和企业研发投入经费数据, 见表 1 所列。

表 1 R&D 经费政府投入与企业投入经脱敏处理数据

年份	地区生产水平/(亿元)	政府投入						企业投入					
		A	B	C	D	E	F	A	B	C	D	E	F
2020	3 873.68	0.01	0.01	0.01	0.23	3.96	1.79	0.70	11.40	4.80	9.90	17.70	21.30
2021	4 299.57	0.30	0.02	0.39	0.69	4.39	1.07	1.90	14.90	6.40	8.50	21.40	16.00
2022	4 501.95	6.93	2.83	0.45	0.75	5.00	1.50	6.50	16.40	8.10	18.10	24.20	18.10
2023	4 701.48	4.19	1.80	0.50	0.80	5.50	1.60	13.20	18.70	11.60	45.10	29.40	17.90
2024	5 062.69	4.20	2.00	0.55	0.85	6.00	1.70	20.10	33.50	16.90	39.80	33.60	18.30

企业 R&D 经费投入强度见表 2 所列。

表 2 各企业 R&D 经费投入强度 %

年份	A	B	C	D	E	F
2020	0.05	1.35	4.65	1.22	22.39	4.25
2021	0.12	1.61	4.89	1.08	21.27	3.39
2022	0.37	1.65	4.18	2.22	18.55	4.23
2023	0.80	1.42	4.81	3.96	16.03	4.46
2024	1.52	2.75	4.20	3.90	17.83	5.02

3.2 逐步最小二乘回归法筛选关键因素

3.2.1 影响因素的选取

充分考虑是否可以获取相关统计数据,本文选择影响 R&D 投入的变量为 G_{DP} 、 C_Z 、 C_Y 。

3.2.2 数据预处理

为确保统计分析的有效性和可靠性,本文对所有变量进行对数转换,并利用出口价格指数等对 G_{DP} 、 C_Z 、 C_Y 数据进行平减处理,以消除价格波动的影响,使数据更稳定、更具可比性。此外,为避免伪回归现象,对各变量进行单位根检验(ADF、PP 检验)。结果表明原序列均为非平稳序列,存在单位根。但经过一阶差分处理后,所有变量均变为平稳序列, $\ln G_{DP}$ 在 10% 显著性水平下平稳,其他变量在 5% 显著性水平下平稳,表明这些变量均为一阶单整序列。

3.2.3 基本回归模型的构建

考虑外商直接投资和人力资本的积累对安徽省企业 R&D 经费投入的力度影响并不显著,因此将这 2 个变量从模型中剔除,最终确定的企业 R&D 经费投入影响因素回归模型为:

$$\ln R_{R\&D} = \beta_0 + \beta_1 \ln G_{DP} + \beta_2 \ln C_Z + \beta_3 \ln C_Y + \mu_t \quad (19)$$

回归分析结果显示了模型的基本系数,由于模型的 Durbin-waston 统计量 D_w 值较小,判断可能存在自相关现象。为减小其影响,本文在模型中加入一阶自回归模型 $A_R(1)$,最终得到企业 R&D 经费投入影响因素回归模型,即

$$\ln R_{R\&D} = -2.575 + 0.372 \ln G_{DP} + 1.504 \ln C_Z + 1.028 \ln C_Y + A_R(1) \quad (20)$$

该回归模型的 $R^2 = 0.997$, F 统计量为 1 298.438,上述模型表明在企业 R&D 经费投入的影响因素中,地区生产水平、政府对科技创新的财政支持力度和典型科技创新性企业自身的投入力度影响显著,因此这 3 个因素是 R&D 活动的主要影响因素。

3.3 分位数回归实证模型的有效性

根据模型式(20),运用 Stata 软件进行分位数回归处理得到各分位点上的分位数回归结果,见表 3 所列。

表 3 分位数回归系数估计值和检验统计量结果

影响因素	0.1		0.2		0.5		0.7		0.9	
	估计值	统计量	估计值	统计量	估计值	统计量	估计值	统计量	估计值	统计量
地区生产水平	0.452***	2.898	0.545***	4.937	0.362***	4.643	0.346***	3.352	0.295***	2.823
政府投入	1.033***	7.840	1.027***	10.178	1.021***	15.004	1.075***	12.392	1.085***	12.214
企业投入	1.087*	1.672	0.968*	1.871	1.182**	2.565	1.410***	2.873	1.685***	2.929
R^2	0.910	0.918	0.943	0.950	0.951					

注: * 表示在 10% 的显著性水平下显著; ** 表示在 5% 的显著性水平下显著; *** 表示在 1% 的显著性水平下显著。

由回归结果可以得到如下结论:

1) 在所有点位上,地区生产水平、政府财政拨款、典型和重点企业的投入力度 3 个影响因素对企业 R&D 投入的影响是显著的。

2) 地区生产水平在 0.2~0.9 点位间下降,回归系数从最高点处的 0.545 降至 0.295。由此可知,在 R&D 经费投入不断增加的同时,地区生产水平对其影响在不断衰减。

3) 政府财政拨款相较另外 2 个因素表现得更加稳定,在 0.1~0.9 点位间,回归系数始终在 0.953~1.085 区间波动,并且位置较高,说明政府财政科技拨款在推动 R&D 投入增长方面发挥重要的积极作用。

4) 企业投入力度的回归系数从最初 0.2 点位的 0.968 攀升至 0.9 点位的 1.685,其对 R&D 投入强度的影响呈现明显的上升趋势,并且在 3 个因素中影响最大,说明企业投入力度的加强对整个 R&D 经费投入起到至关重要的作用。

3.4 基于灰色预测模型的 R&D 投入强度预测

3.4.1 数据检验与处理

本文运用累加生成,通过数列各时刻数据的依次累加得到新的数据和数列,为了保证 GM(1,1)建模方法的可行性和可靠性,需要对已知数据进行必要的检验处理。

假设原始数列为:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (21)$$

依次累加生成的数列为:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \quad (22)$$

其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k=1, 2, \dots, n$ 。

定义 $x^{(1)}$ 的灰导数为:

$$d(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1) \quad (23)$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的邻值生成数列,即

$$z^{(1)}(k) = ax^{(1)}(k) + (1-a)x^{(1)}(k-1),$$

得到 GM(1,1) 的灰微分方程模型为:

$$d(k) + az^{(1)}(k) = b \quad (24)$$

3.4.2 GM(1,1) 模型的建立

原始数列 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ 满足级比检验,建立 GM(1,1) 模型 $x^{(0)}(k) + az^{(1)}(k) = b$,采用回归分析求得 a, b 的估计量,得到相应的白化模型为:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \quad (25)$$

残差计算公式为:

$$e(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1, 2, \dots, n \quad (26)$$

设原始序列 $x^{(0)}$ 及残差序列 e 的方差分别为 S_1^2, S_2^2 , 则有:

$$\begin{aligned} S_1^2 &= \frac{1}{n} \sum_{k=1}^n (x^{(0)}(k) - \bar{x})^2, \\ S_2^2 &= \frac{1}{n} \sum_{k=1}^n (e(k) - \bar{e})^2 \end{aligned} \quad (27)$$

其中: $\bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k); \bar{e} = \frac{1}{n} \sum_{k=1}^n e(k)$ 。

后验差比计算公式同式(17)。

根据检验结果和数值要求分析得出企业 A、B、C、D 近 5 年经费投入值的数据拟合精度高,企业 E、F 近 5 年资金投入值的数据拟合精度为合格或基本合格。因此,案例企业均可有效进行灰色模型的构建。

由 GM(1,1) 模型估测的 a, b 值,可以分别求出各企业的灰色预测模型,即

$$\begin{aligned} \hat{x}^{(1)}(k+1)_1 &= [x^{(0)}(1) + 8.466] \times \\ &\quad e^{0.176k} - 8.466, \\ \hat{x}^{(1)}(k+1)_2 &= [x^{(0)}(1) + 24.076] \times \\ &\quad e^{0.118k} - 24.076, \\ \hat{x}^{(1)}(k+1)_3 &= [x^{(0)}(1) + 29.597] \times \\ &\quad e^{0.144k} - 29.597, \\ \hat{x}^{(1)}(k+1)_4 &= [x^{(0)}(1) - 313.288] \times \\ &\quad e^{-0.073k} + 313.288, \\ \hat{x}^{(1)}(k+1)_5 &= [x^{(0)}(1) + 56.681] \times \\ &\quad e^{0.069k} - 56.681, \\ \hat{x}^{(1)}(k+1)_6 &= [x^{(0)}(1) - 155.125] \times \\ &\quad e^{-0.032k} + 155.125 \end{aligned} \quad (28)$$

3.4.3 GM(1,1) 模型的求解与检验

已知白化模型的解为:

$$x^{(1)}(t) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-a(t-1)} + \frac{b}{a} \quad (29)$$

于是可以得到原始值,即

$$\begin{aligned} \hat{x}^{(1)}(k+1) &= \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a}, \\ &\quad k = 1, 2, \dots, n-1 \end{aligned} \quad (30)$$

预测值为:

$$\begin{aligned} \hat{x}^{(0)}(k+1) &= \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \\ &\quad k = 1, 2, \dots, n-1 \end{aligned} \quad (31)$$

相对残差计算公式为:

$$\begin{aligned} \epsilon(k) &= \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, \\ &\quad k = 1, 2, \dots, n \end{aligned} \quad (32)$$

相对误差计算公式为:

$$\rho(k) = 1 - \frac{1-0.5a_\lambda}{1+0.5a_\lambda}(k) \quad (33)$$

根据模型可以得到拟合后的原始值、预测值,并计算残差、相对误差,结果见表 4 所列。

根据相对误差值和检验标准分析可得,GM(1,1) 模型对案例企业经费原数据的拟合都满足要求,更有部分数据的拟合效果非常理想,高度符合模型特征。借助 SPSS 软件,根据 2021—

2024 年典型科技创新性企业 R&D 经费投入强度构建的灰色预测模型,计算可得 2025—2029 年

各典型企业和重点企业未来 5 年的 R&D 经费投入强度分布趋势,如图 1 所示。

表 4 模型拟合后残差、相对误差

年份	企业 A		企业 B		企业 C		企业 D		企业 E		企业 F	
	残差	相对误差	残差	相对误差	残差	相对误差	残差	相对误差	残差	相对误差	残差	相对误差
2021	0.10	19.55	-0.16	4.64	-0.32	19.14	0.80	3.74	0.25	15.43	0.15	3.09
2022	-0.05	12.18	0.24	5.65	-0.01	0.53	-0.48	2.59	-0.03	1.48	-0.41	9.80
2023	-0.08	10.11	-0.03	0.69	0.76	19.22	-1.66	10.35	-0.59	19.52	0.36	7.59
2024	0.08	5.49	-0.03	0.66	-0.42	10.67	1.39	7.80	0.38	13.86	-0.10	2.50

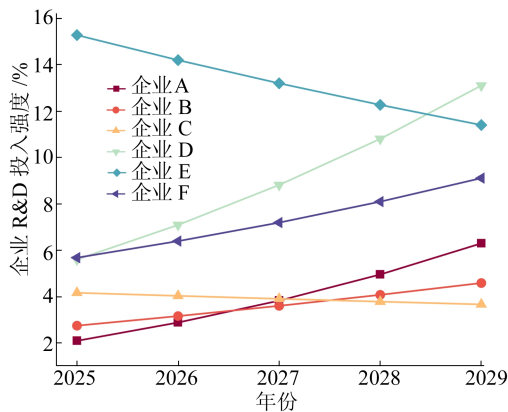


图 1 2025—2029 年企业 R&D 经费投入强度

3.5 结果分析

本文通过多模型联用系统揭示了企业 R&D 经费投入的影响机理和动态演变规律。

1) 基于逐步最小二乘回归法模型的全局分析表明,地区生产水平、政府财政科技拨款及典型和重点企业的投入力度是驱动 R&D 经费投入的三大核心因素,验证了经济基础、政策支持和企业主体行为对创新投入的协同作用。

2) 采用分位数回归模型细化不同投入强度下的差异化影响,可得如下结论:① 地区生产水平的回归系数从 0.2 分位点的 0.545 降至 0.9 分位点的 0.295,表明其边际效应随 R&D 投入规模扩大而递减,源于经济总量对创新驱动的阶段性饱和;② 政府财政拨款回归系数稳定在 0.953~1.085 之间,凸显其跨投入层级的强支撑性,是政策端持续发力的关键抓手;③ 典型和重点企业投入力度回归系数从 0.968 跃升至 1.685,反映头部企业对创新资源的虹吸效应和规模化研发的边际收益递增规律,其影响力随投入强度提升占据主导地位。

3) 采用灰色预测模型验证了未来趋势的分化,多数企业 R&D 投入持续增长,但部分头部企业呈现阶段性回调,提示需要警惕过度依赖单一

主体的风险,通过政策引导优化投入结构。

综上所述,三大模型逐层递进,逐步最小二乘回归锁定关键变量,分位数回归揭示动态异质性,灰色预测预判结构矛盾,表明安徽省 R&D 投入已从“经济总量驱动”转向“政策-企业双轮驱动”阶段。未来需巩固财政支持稳定性,强化重点企业与中小企业的协同创新网络,并建立动态监测机制以平衡投入结构的可持续性。

4 结 论

1) 企业研发投入呈现多主体协同驱动特征。区域创新体系建设需兼顾经济基础夯实、政策杠杆撬动与市场主体培育的协同发展。其中地区生产总值作为经济基础要素,其边际效应随投入规模扩大呈递减趋势,因而经济发展到一定阶段后需要通过结构性调整突破创新瓶颈。

2) 政策支持具有跨层级的持续有效性。财政支持对各类规模企业的研发投入均产生了显著的正向激励。这种政策效力的非衰减性特征,凸显了政府科技投入在创新生态中的基础性支撑作用。

3) 企业主体呈现动态演进特征。揭示研发投入存在显著的“马太效应”,头部企业通过创新资源集聚形成规模效应,但其虹吸效应可能导致区域创新体系的结构失衡。灰色预测模型进一步表明,未来 3 年重点企业研发投入增速,提示过度依赖单一创新主体的潜在风险。

基于上述结论,建议构建“梯度化”政策支持体系,对中小微企业强化普惠性财政支持,缓解创新资源获取壁垒;对重点企业实施“激励-约束”双向调控,通过税收优惠与研发绩效考核引导其发挥创新引领作用;建立跨区域研发资源共享平台,促进创新要素的均衡配置。后续将进一步优化多模型系统方法,研究数字经济背景下新型研发组织模式对传统投入机制的影响。

[参 考 文 献]

- [1] ZHOU M, WEI P, DENG L. Research on the factorial effect of science and technology innovation (STI) policy mix using multifactor analysis of variance (ANOVA)[J]. *Journal of Innovation & Knowledge*, 2022, 7(4):100249.
- [2] 赵狄娜. 安徽省聚力高水平创新型省份建设[J]. *小康*, 2024(7):33.
- [3] XIE Y, ZHOU L K. Assessing the green technology innovation efficiency in Yangtze river delta region under dual carbon background[J]. *Processes*, 2023, 11(4):1082.
- [4] SU H, QU X, SHUO T, et al. Artificial intelligence empowerment: the impact of research and development investment on green radical innovation in high-tech enterprises[J]. *Systems Research and Behavioral Science*, 2022, 39(2):489-502.
- [5] 杨武, 任禹佳. 中国大中型规模以上工业企业科技创新景气指数研究[J]. *中国科技论坛*, 2025(4):75-84.
- [6] GRILICHES Z. Productivity, R&D, and basic research at the firm level in the 1970's[J]. *The American Economic Review*, 1986, 76(1):141-154.
- [7] PANDIT S, WASLEY C, ZACH T. The effect of research and development (R&D) inputs and outputs on the relation between the uncertainty of future operating performance and R&D expenditures[J]. *Journal of Accounting, Auditing & Finance*, 2011, 26(1):121-144.
- [8] ZHONG J, TANG Y. Empirical analysis of R&D investment, industrial structure transformation and development of high-tech industry in guangdong province; based on time series data from 2003-2018[J]. *IOP Conference Series Earth and Environmental Science*, 2019, 358(3):32-48.
- [9] 屈天佑, 李健, 李娜. 政府直接 R&D 资助和企业 R&D 投入: 基于中国省际面板数据的实证分析[J]. *技术经济与管理研究*, 2021(6):20-24.
- [10] 郭珂, 郭雪萌, 周煜皓. 地方政府补助双重性与企业 R&D 投入: 基于 A 股上市企业的经验证据[J]. *北京交通大学学报(社会科学版)*, 2022, 21(4):113-123.
- [11] HAN X, FENG H. The impact of foreign R&D on the innovation performance of China's high-tech industry and its spatial spillover effect[J]. *PLoS ONE*, 2023, 18(3):1371.
- [12] 薛阳, 王健康, 胡丽娜. R&D 经费投入、创新动机与科技成果转化政策实施效果[J]. *统计与决策*, 2023, 39(3):170-174.
- [13] 光辉, 刘颖. 数字化转型 R&D 投入与企业创新绩效[J]. *现代管理科学*, 2024(3):187-195.
- [14] 黄姝婷. 政府研发补贴对企业科技创新能力的影响研究[D]. 南昌:江西财经大学, 2024.
- [15] 洪志生, 周源, 薛澜. 从东部走向中西部 从中国走向世界: “十三五”期间战略性新兴产业发展阶段性特征分析[J]. *中国战略新兴产业*, 2015, 3(7):18-20.
- [16] 靳庭良. ADF 检验与 PP 检验的可靠性比较[J]. *石家庄经济学院学报*, 2007, 30(1):71-76.
- [17] 陈昭. 时序非平稳性 ADF 检验法的理论与应用[J]. *广州大学学报(自然科学版)*, 2008, 7(5):5-10.
- [18] CHEN D Y, ZHANG J H, ZHANG Z X, et al. Analyzing the effect of light on lettuce F_v/F_m and growth by machine learning[J]. *Scientia Horticulturae* 2022, 306:111444.
- [19] 蒋红卫, 夏结来. 偏最小二乘回归及其应用[J]. *第四军医大学学报*, 2003, 24(3):280-283.
- [20] 袁芳. 基于多元回归模型的北京市 $PM_{2.5}$ 影响因素研究[J]. *现代营销(下)*, 2023(1):75-77.
- [21] 王乐洋, 赵英文, 陈晓勇, 等. 多元总体最小二乘问题的牛顿解法[J]. *测绘学报*, 2016, 45(4):411-424.
- [22] QIAN X, ZHANG Y, LIU L, et al. Exploring the potential of leaf reflectance spectra for retrieving the leaf maximum carboxylation rate [J]. *International Journal of Remote Sensing*, 2019, 40(14):5411-5428.
- [23] KOENKER R, BASSRTT G. Regression quantiles [J]. *Econometrica*, 1978, 46(1):33-50.
- [24] 谢兰云, 王维国. 基于分位数回归的我国 R&D 经费投入影响因素的动态研究[J]. *数学的实践与认识*, 2012, 42(2):43-52.
- [25] 陈实. 中国企业 R&D 经费投入特点与统计数据问题解析[J]. *中国科技论坛*, 2015(11):100-106.
- [26] SHENG D S, WANG D H. Quantile regression estimation for poisson autoregressive models[J/OL]. *Journal of Time Series Analysis*, 2025:1-36[2024-12-05]. <https://onlinelibrary.wiley.com/doi/epdf/10.1111/jtsa.12811>.
- [27] FRUMENTO P, BOTTAI M, FERNANDEZ-VAL I. Parametric modeling of quantile regression coefficient functions with longitudinal data[J]. *Journal of the American Statistical Association*, 2021, 30:1237-1258.
- [28] DUCCI N, GRILLI L, PITTAVINO M. Comparing flexible modelling approaches: the varying-thresholds model versus quantile regression[J]. *Advances in Data Analysis and Classification*, 2025, 19:493-514.
- [29] ZENG B, TUO Y B. Multivariate grey prediction model with fractional time-lag parameter and its application[J]. *Applied Mathematical Modelling*, 2025, 141:115940.

(责任编辑 张 镅)