

基于回归模型的消费行为统计研究

李林佳

山东建筑大学理学院, 山东济南, 250102;

摘要: 随着数字经济的快速发展, 消费者行为呈现出高度动态化与个性化特征。本文以回归模型为核心工具, 结合大规模真实消费数据, 系统分析影响个体消费决策的关键变量及其作用机制。区别于传统线性回归的单一视角, 本文引入分位数回归与面板数据固定效应模型, 揭示不同收入阶层、年龄群体及地域环境下消费弹性的异质性。研究发现: 收入对消费的影响在低收入群体中更为敏感; 数字支付习惯显著提升边际消费倾向; 而疫情后消费结构呈现“服务回补、商品降级”的双重趋势。本研究不仅丰富了消费行为理论的实证基础, 也为政策制定者优化消费刺激措施、企业精准营销提供科学依据。全文强调模型选择的适配性、数据的真实性与结论的可解释性, 力求实现学术价值与实践指导意义的统一。

关键词: 回归模型; 消费行为; 分位数回归; 边际消费倾向

DOI: 10.69979/3029-2700.26.03.050

引言

消费作为拉动经济增长的“三驾马车”之一, 其行为模式的准确刻画对宏观经济调控与微观市场策略具有重要意义。传统消费函数理论(如凯恩斯绝对收入假说、弗里德曼持久收入假说)虽奠定理论基础, 但在大数据时代面临解释力不足的挑战。当前消费行为受多重因素交织影响——收入波动、数字技术渗透、社会心理变化等, 亟需更精细的计量工具予以解析。回归模型因其良好的可解释性与灵活性, 成为消费行为实证研究的主流方法。然而, 现有研究多依赖普通最小二乘法(OLS), 忽视了消费分布尾部特征及群体异质性。本文旨在突破这一局限, 通过构建多维度回归框架, 深入挖掘消费行为的结构化差异, 并提出“弹性分层”新观点: 即不同群体对同一经济变量的反应强度存在系统性差异。本研究基于真实交易数据, 确保结论的科学性与实用性, 为理解当代中国消费图景提供新视角。

1 文献综述与理论框架

消费行为研究历经长期理论演进, 凯恩斯的绝对收入假说奠定了消费与当期收入正相关的基础逻辑, 弗里德曼提出的持久收入假说则强调消费者依据长期稳定收入进行支出决策, 莫迪利安尼的生命周期理论进一步将时间维度纳入分析框架, 指出个体在不同生命阶段平滑消费路径。这些经典理论为后续实证研究提供了重要支撑, 大量文献借助线性回归模型验证收入、财富与消费之间的关系。近年来, 随着数据规模扩大与算法进步,

部分学者尝试融合机器学习方法(如随机森林、梯度提升树)与传统计量模型, 以捕捉变量间的非线性交互效应, 提升预测精度。然而, 主流回归分析仍普遍依赖普通最小二乘法, 难以有效处理消费数据中广泛存在的异方差性、尾部偏态及群体响应差异。尤其在数字经济背景下, 技术使用习惯、心理预期变化等新型变量对消费决策的影响日益显著, 单一维度的收入驱动解释力明显不足。针对这一局限, 本文提出“收入-技术-心理”三维驱动理论框架, 将数字支付渗透率、消费信心指数等指标纳入回归体系, 主张回归模型的选择与设定必须契合消费行为的多维复杂性。该框架不仅拓展了传统消费函数的解释边界, 也为识别不同人群的边际消费倾向差异提供结构化分析路径, 使模型结果更具现实指向性与政策适配性。通过强化变量维度与模型灵活性的协同, 研究力求在保持计量严谨性的同时, 更真实地还原当代消费者的行为逻辑。

2 数据来源与变量设计

本研究采用某大型商业银行2020至2025年间匿名化处理的个人消费流水数据, 覆盖全国31个省、自治区和直辖市, 原始样本量超过200万条, 具有较强的全国代表性与时效性。被解释变量设定为个体月度人均消费支出, 并对其取自然对数以缓解右偏分布并符合经济变量的弹性解释惯例。核心解释变量包括经税务与社保信息校准的月度可支配收入、数字支付使用频率(以每月非现金交易笔数衡量)、年龄(以年为单位)、教育

程度（按初中及以下、高中/中专、大专、本科及以上学历四编类编码）、城市等级（划分为一线、新一线、二线、三线及以下四档）以及疫情前后虚拟变量（2020年1月至2022年12月赋值为1，其余为0）。控制变量引入家庭常住人口数量与职业类型（如公务员、企业职工、自由职业者、个体户等分类虚拟变量），以缓解遗漏变量偏误。所有连续变量均进行1%和99%分位的Winsorize缩尾处理，消除极端值干扰；分类变量统一采用哑变量编码，避免赋值主观性。描述性统计显示，样本月均消费支出中位数为2860元，数字支付平均使用频次达17.3次/月，本科及以上学历占比34.7%，三线及以下城市用户占样本总量的58.2%，整体结构贴近国家统计局公布的居民结构特征，保障了后续回归分析的外部有效性与现实基础。

3 模型构建与方法选择

3.1 基准 OLS 模型设定与初步估计

为刻画收入对消费支出的平均影响，构建普通最小二乘（Ordinary Least Squares, OLS）回归作为基准模型。模型形式设定为：

$$\ln C_{it} = \alpha + \beta_1 Y_{it} + \beta_2 X_{it} + \varepsilon_{it}$$

其中 C_{it} 表示个体 i 在时期 t 的月度人均消费支出， Y_{it} 为可支配收入（取对数）， X_{it} 代表年龄、教育程度、城市等级等控制变量向量， ε_{it} 为随机扰动项。该模型假设所有个体对收入变化的反应具有一致性，适用于初步识别核心变量的符号方向与经济显著性。估计结果显示收入系数显著为正，符合凯恩斯消费理论的基本预期。然而，OLS 仅反映条件均值关系，无法捕捉消费分布不同位置的响应差异，尤其在存在异方差或尾部极端行为时可能产生偏误。模型残差图显示明显扇形扩散，提示误差方差随收入水平上升而增大，违背同方差假定，削弱了标准误的可靠性。尽管如此，该基准结果为后续更精细模型提供了参照系，并揭示了传统线性框架在解释当代复杂消费行为时的局限性。

3.2 固定效应面板模型控制个体异质性

为缓解由不可观测个体特征（如消费偏好、风险态度、家庭文化等）引发的内生性问题，引入个体固定效应（Fixed Effects, FE）面板模型。该模型通过组内变换消除不随时间变化的个体特异性成分，有效控制遗漏变量偏误。模型表达式为：

$$\ln C_{it} = \alpha_i + \beta_1 Y_{it} + \beta_2 X_{it} + \varepsilon_{it}$$

其中 α_i 为个体固定效应项。利用 2020 - 2025 年多期观测数据，模型能更准确识别收入变动对同一主体消费行为的动态影响。Hausman 检验结果支持固定效应优于随机效应，确认个体效应与解释变量相关。FE 估计下收入弹性略有下降，表明忽略个体异质性会高估收入的真实作用。该模型虽提升因果推断力度，但仍局限于条件均值分析，未能揭示低消费群体与高消费群体在边际反应上的结构性分化，需借助更灵活的分位方法进一步拓展。

3.3 分位数回归揭示消费弹性异质性

为突破均值回归的局限，采用 Koenker 与 Bassett 提出的分位数回归（Quantile Regression, QR）方法，分别在 10%、50% 和 90% 分位点估计收入对消费的影响。QR 模型直接建模条件分位函数，形式为：

$$Q_T(\ln C_{it} | Z_{it}) = \alpha(T) + \beta_1(T) Y_{it} + \beta_2(T) X_{it}$$

其中 T 表示分位水平。结果显示，10% 分位处的收入弹性为 0.84，显著高于中位数（0.63）与 90% 分位（0.41），证实低收入群体消费对收入波动更为敏感。这一发现支持“弹性分层”假说，说明统一政策刺激对不同阶层效果迥异。QR 还能稳健处理非正态误差与异常值，其系数估计不受极端高消费样本干扰。Bootstrap 法计算的标准误进一步验证了各分位系数差异的统计显著性。该方法有效揭示了消费行为背后的结构性不平等，为精准施策提供微观依据。

3.4 交互项模型与稳健性检验设计

为检验数字支付是否调节收入对消费的影响，构建包含交互项的扩展模型：

$$\ln C_{it} = \alpha + \beta_1 Y_{it} + \beta_2 D_{it} + \beta_3 (Y_{it} \times D_{it}) + \gamma X_{it} + \varepsilon_{it}$$

其中 D_{it} 为数字支付使用频率。交互项系数显著为正，表明高频数字支付用户具有更高边际消费倾向。模型诊断环节执行 VIF 检验，所有变量 VIF 值均低于 3.5，排除严重多重共线性；Breusch-Pagan 检验 p 值小于 0.01，确认异方差存在，故后续回归均采用聚类稳健标准误。稳健性检验包括：将数字支付频率替换为是否使用移动支付的二元变量；限定样本为 25 - 55 岁劳动人口；分东、中、西部子样本回归。各类替代设定下核心结论保持一致，证明研究结果具有较强可靠性与泛化能力。

4 实证结果与异质性分析

4.1 收入弹性分层：从均值幻觉到结构真实

普通最小二乘回归估计的收入弹性为0.62，意味着可支配收入每提升1%，平均消费支出增长0.62%。该数值虽符合宏观经验判断，却掩盖了群体内部的巨大差异。分位数回归结果清晰揭示出“弹性分层”现象：在消费分布10%分位处，收入弹性高达0.85；中位数水平为0.63；而90%分位群体仅0.41。低收入群体因缺乏储蓄缓冲与信贷渠道，收入变动直接转化为消费调整，边际倾向强烈；高收入者则更多将增量收入用于资产配置或预防性储蓄，消费反应迟钝。这一发现挑战了传统消费函数中“同质响应”的隐含假设，凸显政策设计需考虑受益对象的结构特征。例如，向低收入家庭发放现金补贴的乘数效应远高于对高收入群体减税。回归系数的95%置信区间在各分位点无重叠，差异具有统计与经济双重显著性。图1展示的分位数弹性曲线呈明显递减趋势，直观印证了消费行为的非线性机制。

4.2 数字支付对边际消费倾向的激活效应

数字支付使用频率与边际消费倾向之间存在显著正向关联。交互项模型显示，在控制收入与其他协变量后，高频数字支付用户（月均交易 ≥ 15 笔）的收入弹性比低频用户高出约0.07，相当于边际消费倾向提升12%。该效应在18-35岁年轻群体中尤为突出，其弹性增幅达0.11，可能源于移动支付降低交易摩擦、促进即时消费冲动及嵌入场景化营销（如直播带货、拼团优惠）。相比之下，55岁以上群体未呈现显著调节作用，反映技术采纳存在代际鸿沟。值得注意的是，数字支付本身并不直接增加收入，而是通过改变支付心理（“无感支付”弱化支出痛感）与提升交易便利性，放大既有收入的消费转化效率。图2中交互项系数的置信区间远离零点，且在子样本回归中保持稳健。这一结果为推动数字金融基础设施下沉提供了实证支持，表明技术普及不仅是效率工具，更是激活内需的结构性杠杆。

5 政策启示与商业应用

实证结果为政策制定与商业实践提供明确指引。财政刺激措施若聚焦低收入群体，可显著提升公共资金的消费乘数效应，因其边际消费倾向接近0.85，远高于高收入阶层。在县域及农村地区加快数字支付终端覆盖与使用引导，有助于降低交易成本、释放潜在需求，尤其对年轻消费者具有强激活作用。企业可依据用户收入层级、支付行为与地域特征进行精细化分群，针对高弹性

群体推出小额高频促销，对高净值客户则侧重服务升级与体验优化，实现产品组合与定价策略的动态适配。大型电商平台或零售企业可将本文构建的回归模型嵌入智能决策系统，结合实时交易数据滚动更新个体消费预测，精准调控库存周转与广告投放节奏。模型输出的分位弹性参数与交互效应系数具备直接工程化价值，能支撑从宏观政策模拟到微观用户运营的全链条应用。研究不仅验证了消费行为的结构差异，更展示了计量模型在真实商业场景中的落地能力，凸显其科学性与实用性的统一。

6 结束语

本文通过构建多层次回归模型，对当代中国居民消费行为进行了系统性实证分析，突破了传统OLS模型对“平均效应”的过度依赖，揭示了消费响应机制中存在的显著群体异质性。研究提出的“弹性分层”观点，不仅深化了对消费函数微观基础的理解，也为精准施策提供了量化依据。在方法论上，本文强调模型选择应服务于研究问题本质——分位数回归有效捕捉了低收入群体的高敏感性，面板固定效应控制了个体特质偏差，体现了计量工具与现实问题的有机统一。数据的真实性和处理的严谨性保障了结论的科学性与可复制性。未来研究可进一步融合文本数据（如社交媒体情绪）与高频交易流，构建动态贝叶斯回归或因果推断模型，以捕捉消费行为的实时演化机制。总体而言，回归模型在消费行为研究中仍具强大生命力，关键在于根据研究目标灵活组合模型、审慎解读参数、紧密联系现实。唯有如此，学术研究才能真正服务于经济高质量发展与人民美好生活需要。

参考文献

- [1] 袁晓惠,王慧雪,向雪飞,等.基于阈值惩罚自回归模型的消费者信心指数的统计研究[J].吉林师范大学学报:自然科学版,2020,41(3):7. DOI:10.16862/j.cnki.issn1674-3873.2020.03.008.
- [2] 周先波,白虎杰.家庭社会互动与教育支出溢出效应研究——基于CFPS的经验证据[J].计量经济学报,2025(3):842-874.
- [3] 刘汉中.我国GDP周期波动的非对称性研究方法 with 统计检验[J].湖南工程学院学报:社会科学版,2008(A02):5. DOI:CNKI:SUN:GCHS.0.2008-02-003.