

公积金提振家庭消费的微观机制与异质性研究

——基于可解释机器学习与因果森林算法

翟祥泽钰

北方工业大学理学院, 北京

收稿日期: 2026年7月6日; 录用日期: 2026年7月28日; 发布日期: 2026年8月7日

摘要

本文基于CFPS 2022微观家庭数据, 构建“基准预测 - 机制解析 - 因果效应估计”的递进建模框架, 系统分析公积金潜在流动性与家庭消费之间的关联机制, 并在双重机器学习识别假设下估计其平均处理效应。研究首先运用随机森林模型评估公积金在消费预测中的相对重要性, 继而通过SHAP可解释机器学习揭示缴存强度与消费的非线性关系及收入交互效应, 最后采用双重机器学习框架下的因果森林方法, 在控制收入、人口特征与家庭禀赋等混杂因素后, 测算公积金对消费的平均处理效应与个体异质性。研究发现: 第一, 公积金变量对家庭消费具有较强解释力, 但明显低于家庭纯收入, 表明公积金主要是依托收入渠道发挥作用的流动性补充手段。第二, SHAP分析显示公积金缴存强度与消费存在显著非线性关系, 月缴存额在1000~2000元区间对消费的边际拉动提升较快, 超过3000元后边际效应逐步放缓。第三, 因果森林估计表明, 在控制可观测混杂因素且识别假设成立的条件下, 模型估计显示公积金月缴存额增加与家庭消费之间存在显著正向关联, 对应平均处理效应约为13.84元。年化消费响应系数约为1.15。第四, 公积金消费效应存在显著异质性, 中高收入、城镇户口和年龄较高家庭的消费响应更强。第五, 公积金对文教娱乐等发展型消费的拉动强于食品等生存型消费, 有助于推动消费结构升级。

关键词

住房公积金, 可解释机器学习, SHAP, 因果森林, 异质性效应

A Study on the Micro-Level Mechanisms and Heterogeneity of How Housing Provident Funds Boost Household Consumption

—Based on Explainable Machine Learning and Causal Forest Algorithms

Xiangzeyu Zhai

College of Science, North China University of Technology, Beijing

文章引用: 翟祥泽钰. 公积金提振家庭消费的微观机制与异质性研究[J]. 统计学与应用, 2026, 15(8): 54-66.
DOI: 10.12677/sa.2026.158179

Abstract

Based on micro-household data from CFPS 2022, this paper constructs a progressive modeling framework of “benchmark forecasting-mechanism analysis-causal effect estimation” to systematically analyze the mechanism linking the potential liquidity of housing provident funds to household consumption, and estimates its average treatment effect under the dual machine learning identification hypothesis. The study first employs a random forest model to assess the relative importance of the housing provident fund in consumption forecasting. It then uses SHAP-based explainable machine learning to reveal the nonlinear relationship between contribution intensity and consumption, as well as the interaction effects with income. Finally, using the Causal Forest method within a dual machine learning framework, the study estimates the average treatment effect of the housing provident fund on consumption and individual heterogeneity after controlling for confounding factors such as income, demographic characteristics, and household endowments. The study found the following: First, the housing provident fund variable has strong explanatory power for household consumption, but it is significantly lower than that of household net income, indicating that the housing provident fund primarily functions as a liquidity supplement through the income channel. Second, SHAP analysis shows a significant nonlinear relationship between the intensity of housing provident fund contributions and consumption; the marginal increase in consumption rises rapidly when monthly contributions fall within the 1000~2000 yuan range, but the marginal effect gradually slows after exceeding 3000 yuan. Third, causal forest estimation indicates that, under conditions where observable confounding factors are controlled and identification assumptions hold, the model estimates a significant positive association between an increase in monthly housing provident fund contributions and household consumption, with a corresponding average treatment effect of approximately 13.84 yuan. The annualized consumption response coefficient is approximately 1.15. Fourth, the consumption effect of the housing provident fund exhibits significant heterogeneity, with stronger consumption responses observed among middle- and high-income households, those with urban household registration, and older households. Fifth, the housing provident fund exerts a stronger stimulatory effect on development-oriented consumption—such as culture, education, and entertainment—than on subsistence-oriented consumption—such as food—thereby helping to promote an upgrade in the consumption structure.

Keywords

Housing Provident Fund, Explainable Machine Learning, SHAP, Causal Forests, Heterogeneity Effects

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 绪论

1.1. 研究背景与问题提出

近年来,随着“租购并举”住房制度推进,公积金功能从购房支持扩展到租房保障、还贷减压和流动性支持等领域[1]。各地放宽提取条件、提高额度、简化流程,使沉淀资金具备更强可盘活属性,公积金逐渐成为缓解家庭预算约束、改善消费预期的重要制度性资金来源。

但公积金对消费的影响并非简单线性关系。高缴存家庭往往收入更高、就业更稳定，其消费增加可能源于收入优势而非公积金制度本身；不同收入、年龄、户籍家庭的住房压力和消费偏好差异显著，公积金能否转化为真实消费增量高度分化[2]。因此，本文核心问题不是判断“公积金是否与消费相关”，而是回答：公积金缴存强度能否有效解释消费差异？是否存在非线性响应边界？控制混杂因素后，边际效应在不同群体间如何分布？

基于以上问题，本文采用 CFPS 2022 数据，构建“随机森林 - SHAP - 因果森林”分析框架，从预测能力、非线性关系与群体差异三个维度，分析公积金潜在流动性与家庭消费之间的关联，并在控制可观测混杂变量基础上估计边际处理效应。需要说明的是，本文以月缴存额作为制度参与强度与潜在流动性的代理变量，检验其对消费的预测作用、非线性响应与异质性边际效应。

1.2. 研究目的与现实意义

我国经济转向高质量发展阶段，扩大内需、拉动消费成为稳增长的关键抓手。居民消费受收入、住房压力、预防性储蓄与流动性约束等多重因素影响[2]。住房公积金兼具住房保障与强制储蓄属性，既能直接作用于住房支出，也能通过改善现金流预期带动一般性消费增长。

本文研究目的包括三方面：一是评估公积金变量在家庭消费预测中的解释能力；二是借助 SHAP 可解释机器学习识别公积金缴存强度与消费之间的非线性响应区间；三是利用因果森林估计不同家庭的条件平均处理效应，刻画公积金消费效应在收入、年龄和户籍维度上的异质性分布。现实意义在于：更准确理解公积金制度在扩大内需中的潜在作用；识别政策传导中的高敏感与弱传导群体；推动公积金制度从“总量调节”走向“分群体精准优化”。

1.3. 研究内容与技术路线

本文以住房公积金制度优化为背景，基于 CFPS 2022 年微观调查数据，构建涵盖家庭消费、收入结构、人口特征和公积金变量的指标体系[3]。研究内容主要包括：第一，描述性统计刻画公积金缴存分布及其与消费的相关关系；第二，利用随机森林和 SHAP 评估公积金对消费的预测贡献及非线性响应；第三，采用因果森林估计公积金对家庭消费的平均处理效应与异质性分布。研究技术路线图见图 1。

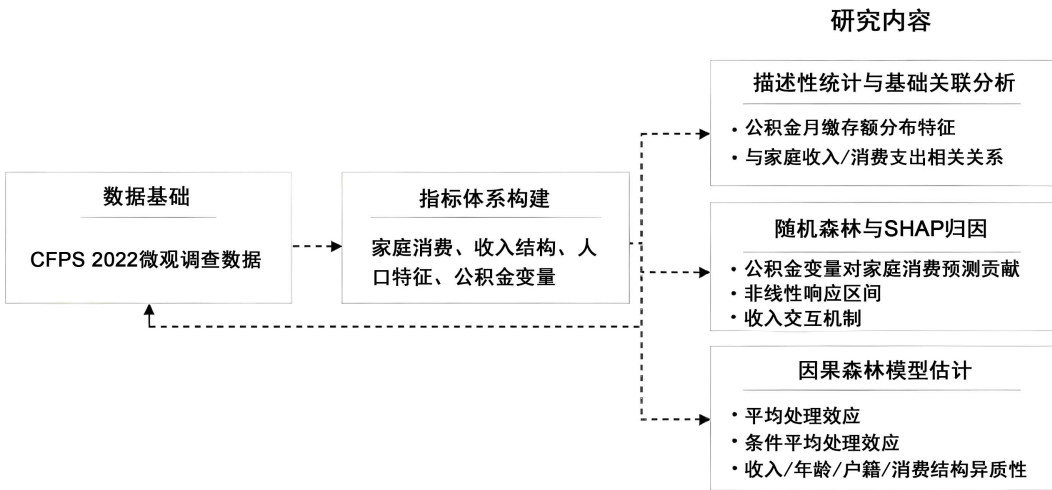


Figure 1. Research roadmap
图 1. 研究技术路线图

2. 住房公积金制度现状与理论基础

2.1. 各地提取优化政策演进与宏观导向

近年来，“租购并举”住房制度持续推进，各地优化公积金使用政策，重点提高提取额度、增加频次、简化流程，提升资金可及性与便捷性。北京、上海等城市结合市场租金动态调整上限，对新市民、青年等群体推出更宽松政策，推动公积金从传统住房储蓄工具转向兼具流动性支持功能的综合保障制度。各地普遍实现按月提取、线上办理和自动到账，使资金更及时进入家庭现金流，可能影响当期消费能力[4][5]。本文正是在这一背景下，从微观数据层面评估公积金缴存强度及潜在流动性对家庭消费的影响。

公积金存量资金地区分布不均衡，经济发达、人口流入地区缴存规模大，部分地区和群体覆盖不足。截至2024年末，全国累计缴存余额中近8万亿元用于个人住房贷款，在支持购房的同时约束了整体流动性。测算显示，可用于进一步支持消费的结余资金上限约1.4万亿元，且高度集中于北京、上海、广东、江苏等地，占比达57%，区域差距突出。这种结构性差异使公积金消费效应在区域、收入和户籍维度明显分化。

2.2. 理论机制分析

住房公积金影响家庭消费，主要可能通过以下三条机制实现：一是流动性约束缓解机制，公积金缴存和提取制度能够缓解家庭购房、租房和还贷压力，降低住房相关现金流压力后，可用于一般消费的预算空间可能增加；二是财富储备与预期稳定机制，较高的公积金缴存额和账户积累可能增强家庭对未来住房支出的安全感，降低预防性储蓄动机，从而增加弹性支出；三是收入协同与异质性机制，公积金消费效应并不独立于家庭收入存在，对于中高收入家庭，公积金更容易转化为消费信心，而对于低收入家庭，缴存可能短期内挤占可支配现金流。

基于上述机制，本文提出以下实证判断：住房公积金缴存强度对家庭消费可能具有正向影响，但该影响并非均匀分布，而是存在非线性边界和群体异质性。

3. 研究设计与模型构建

3.1. 数据来源与质量评估

本文使用中国家庭追踪调查(CFPS)2022年数据，通过匹配个人问卷与家庭经济问卷，构建家庭层面的消费、收入、人口特征和公积金变量数据库[6]。为保证建模稳健性，本文对连续变量进行上下1%缩尾处理，对缺失值采用中位数/众数填充，并剔除关键指标严重缺失的样本，最终保留2465个有效家庭样本。

3.2. 变量体系设定

1) 被解释变量(消费支出指标)

本文被解释变量为家庭消费支出，重点关注家庭年度总支出，并以食品支出和文教娱乐支出分别作为生存型消费和发展型消费的代理变量，用于识别公积金影响消费结构的方向。表1列示了本文使用的主要消费支出变量。

需要说明的是，在后续核心建模部分，本文主要以家庭年度总支出作为整体消费水平指标，并以文教娱乐支出和食品支出作为消费结构分析的代表性变量。

2) 核心解释变量与控制变量

本文核心解释变量为公积金月缴存额，反映家庭公积金参与强度及未来可提取的住房储蓄规模。受

数据可得性限制，本文并非识别具体提取政策变动的冲击效应，而是考察公积金缴存强度及其潜在流动性与家庭消费的关系，结论更适合解释微观边际消费响应。

为缓解遗漏变量偏误，本文控制家庭收入结构(工资性、经营性、财产性、转移性、其他收入及家庭纯收入)和人口学特征(年龄、性别、学历、婚姻状态和户口类型) [7]。具体变量设定见表 2。

Table 1. Table of dependent variables
表 1. 被解释变量设定表

特征 (消费维度)	变量	指标处理/解释
生存型消费	食品支出	家庭食品相关消费支出，含食品、烟酒、饮食服务等基础生存消费
	衣着鞋帽支出	家庭衣着、鞋帽类消费支出，满足基本穿着需求
	居住支出	家庭居住相关消费，含租赁房租、住房维修及管理、水电燃料等
	家庭设备及日用品支出	家庭家具、家用器具、日杂用品、个人护理用品等日用品消费
	医疗保健支出	家庭医疗器具及药品、医疗服务等健康保障类消费
发展型消费	交通通讯支出	家庭交通出行、通信服务等发展相关消费
	文教娱乐支出	家庭教育、文化娱乐等发展性消费，含教育培训、文娱服务等
享受型消费	其他消费性支出	家庭除上述类别外的其他用品及服务消费
	福利性支出	家庭福利相关消费支出
专项支出	转移性支出	家庭转移性消费支出(如赡养、捐赠等)
	房贷支出	家庭住房贷款偿还支出
综合指标	家庭总支出	上述各类支出汇总，全面衡量家庭整体消费规模

Table 2. Specification of explanatory and control variables
表 2. 解释变量与控制变量设定表

变量类型	变量名称	指标说明
核心解释变量	公积金月缴存额	表征家庭公积金缴存强度及潜在流动性规模
辅助解释变量	公积金账户余额	反映家庭公积金存量积累水平
收入变量	工资性收入	家庭工资劳动收入
	经营性收入	家庭生产经营收入
	财产性收入	利息、租金、分红等财产收入
	转移性收入	养老金、补贴、赡养收入等
	其他收入	其他来源收入
	家庭纯收入	家庭各类收入汇总后的综合收入水平
人口变量	年龄	户主或家庭主要受访者年龄
	性别	男性赋值为 1，女性赋值为 0
	最高学历	按受教育程度分类赋值
	婚姻状态	未婚、已婚、离婚、丧偶等
制度变量	户口类型	反映城乡制度属性差异

3.3. 递进式机器学习框架构建

传统线性回归难以准确刻画公积金与消费之间的复杂非线性关系，且易受内生性干扰。为此，本文构建“基准预测 - 机制解释 - 因果推断”的机器学习分析框架。

1) 基准预测：随机森林回归模型

随机森林作为集成学习方法，通过多棵决策树提升预测稳定性，能较处理好变量间的非线性与交互关系[8]。本文采用该模型预测家庭消费，通过特征重要性比较公积金、收入与人口变量的解释力，并结合交叉验证保障模型稳健性。

2) 机制解释：SHAP 事后归因模型

传统特征重要性难以解释变量影响方向、非线性边界和交互机制。本文引入基于 Shapley value 的 SHAP 方法，将模型预测值分解为各特征的边际贡献，以识别公积金缴存强度与消费之间的非线性响应区间及收入交互效应[9]。特征 i 的 SHAP 值计算公式如下：

$$\phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (1)$$

其中， F 为全部特征集合， S 为不包含特征 i 的子集， $f_x(S)$ 为特征子集的预测值。借助该模型，本文能够识别公积金变量对家庭消费预测值的贡献方向和非线性响应区间。

3) 因果推断：因果森林模型(Causal Forest)

为进一步缓解收入等可观测混杂因素的干扰，识别公积金缴存强度对家庭消费的边际处理效应，本文引入基于双重机器学习框架的因果森林算法。该模型通过对处理变量(公积金缴纳额 T)与结果变量(家庭消费支出 Y)进行双重残差正交化处理，系统性控制高维控制变量 X 的干扰，核心目标函数如下：

$$Y - E[Y|X] = \theta(X) \cdot (T - E[T|X]) + \epsilon \quad (2)$$

其中， $E[Y|X]$ 和 $E[T|X]$ 分别为利用机器学习估算的条件期望， X 表示收入结构、人口特征和家庭禀赋等控制变量， ϵ 为随机误差项。方程求解出的 $\theta(X)$ 即为条件平均处理效应(CATE) [10]。

该模型的优势在于不预设处理效应同质，能够估计个体层面的条件平均处理效应。需要说明的是，因果森林模型的识别依赖于条件独立假设，即控制主要可观测混杂变量后，剩余未观测因素不会对公积金缴存强度与家庭消费之间的关系产生系统性偏误影响。本文通过引入丰富的家庭特征变量，并结合稳健性检验降低潜在遗漏变量偏误的影响。因此，本文估计的是在条件独立假设(Conditional Independence Assumption)成立前提下的平均处理效应，而非严格意义上的无条件因果效应。

4. 基准评估与特征事实：公积金对消费的全局预测

4.1. 描述性统计与基础关联

根据图 2 描述性统计结果显示，样本家庭公积金月缴存额平均值为 656.19 元，中位数为 480.00 元，标准差为 633.14 元，呈现明显右偏分布：0~1000 元占比 81.62%，1000~2000 元占 15.78%，2000 元以上仅 2.60%。收入与缴存额相关系数为 0.407，呈弱至中度正相关，且缴存额还受单位比例、基数、行业等因素影响。

4.2. 公积金与消费的关联关系分析

为厘清住房公积金变量与各类消费支出的基础关联特征，本文采用 Pearson 相关系数分析法，构建公积金相关指标与 11 类消费支出的相关性矩阵，并绘制相关性热力图见图 3。

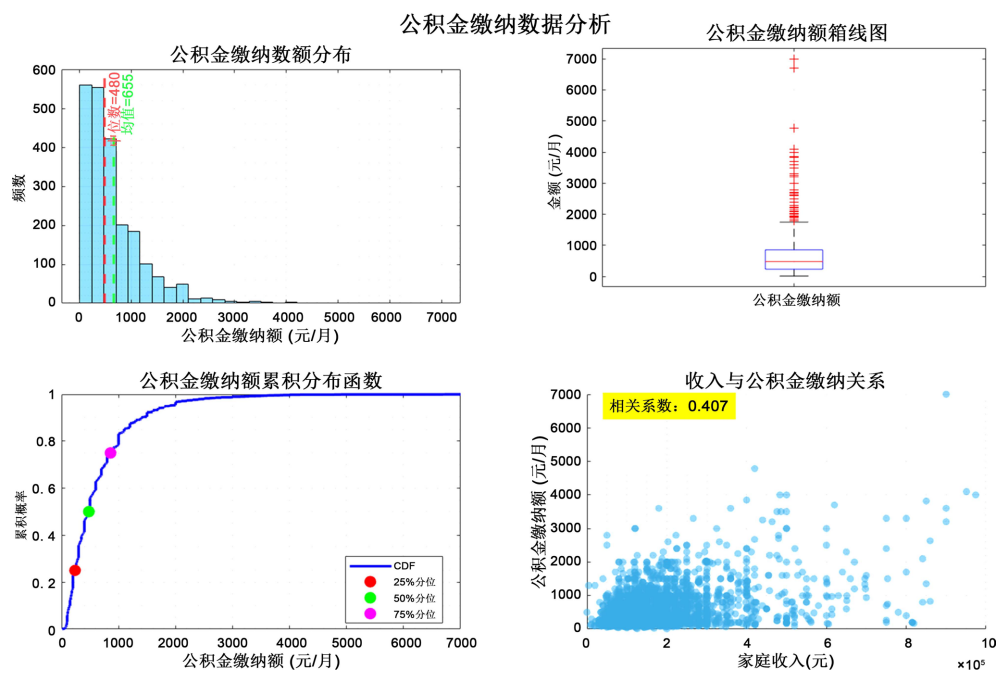


Figure 2. Distribution of monthly housing provident fund contributions and their relationship to household income

图 2. 公积金月缴存额分布及其与家庭收入的关系

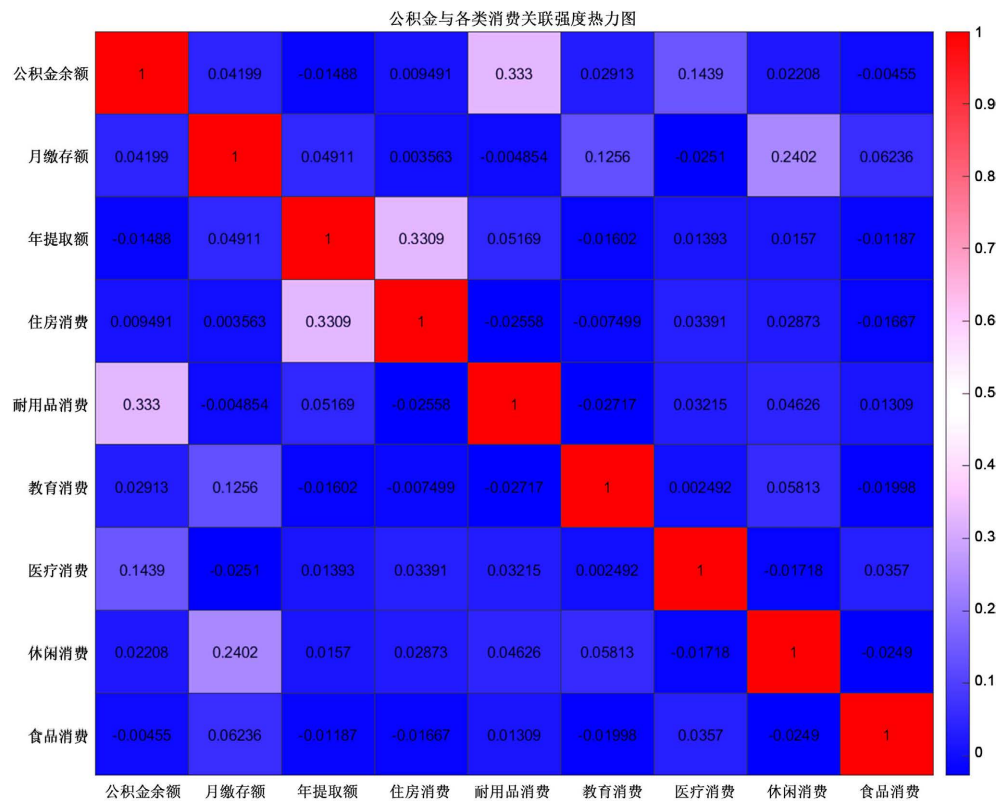


Figure 3. Heat map showing the correlation between housing provident fund variables and various types of consumer expenditures

图 3. 公积金变量与各类消费支出的相关性热力图

结果显示，公积金变量与不同消费类别存在选择性关联。公积金账户余额与家庭设备及日用品支出相关系数较高，说明账户积累水平较高的家庭更可能进行耐用品消费；与居住支出、房贷支出等住房相关支出也存在一定联系，符合公积金制度的住房保障属性。相反，公积金变量与食品等刚性消费的相关系数较低，说明基础生存型消费主要受人口规模和基本收入影响，对公积金变化不敏感。

该结果初步表明，公积金对消费的影响具有选择性，并非均匀作用于所有消费类别，可能主要通过缓解住房支出压力和增强流动性预期，影响耐用品和发展型消费。但相关系数仅反映线性关联，不能说明因果关系。为进一步刻画公积金变量在家庭消费预测中的重要性，本文继续采用随机森林模型进行全局预测分析。

4.3. SHAP 全局归因结果

根据图 4 SHAP 全局特征重要性结果显示，家庭纯收入位居家庭总支出预测模型的首位，工资性收入紧随其后，说明收入变量仍然是家庭消费最主要的决定因素。公积金月缴存额在变量重要性排序中处于较高位置，表明其对家庭消费预测具有较强补充解释力。从 SHAP 散点分布看，高收入家庭通常对应正向 SHAP 值，公积金月缴存额较高的样本也更多分布在正向 SHAP 区域，说明公积金缴存强度提高通常会增加家庭消费预测值。

这一结果表明，公积金并非仅具有住房保障功能，在微观家庭层面也体现出一定消费解释能力。但与收入变量相比，公积金的作用更像是一种依托收入基础发挥作用的流动性补充机制，而不是独立于家庭收入之外的消费决定因素。

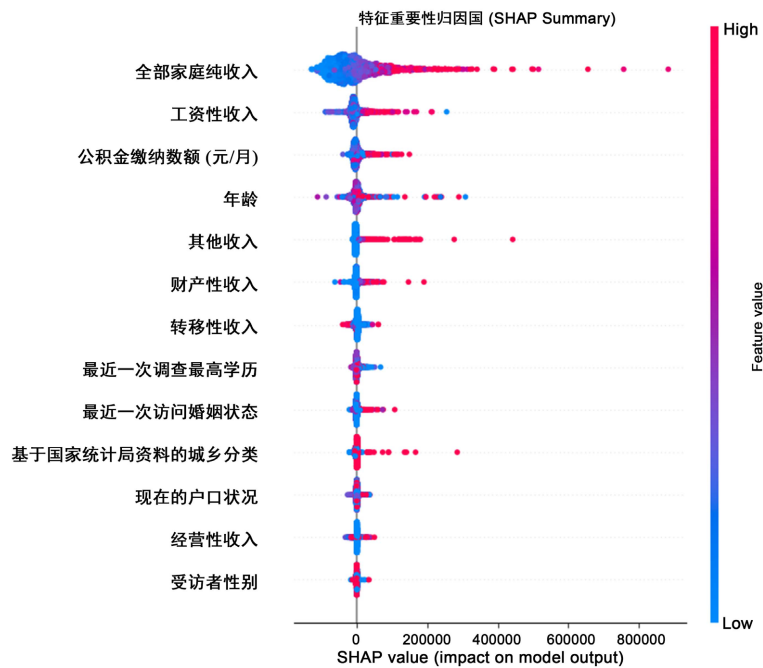


Figure 4. SHAP global feature importance for the household total expenditure prediction model

图 4. 家庭总支出预测模型的 SHAP 全局特征重要性

5. 机制透视：提振效应的非线性边界挖掘

根据图 5 SHAP 单变量依赖图显示，公积金月缴存额与家庭消费预测值之间并非简单线性关系，而

是呈现明显阶段性特征。

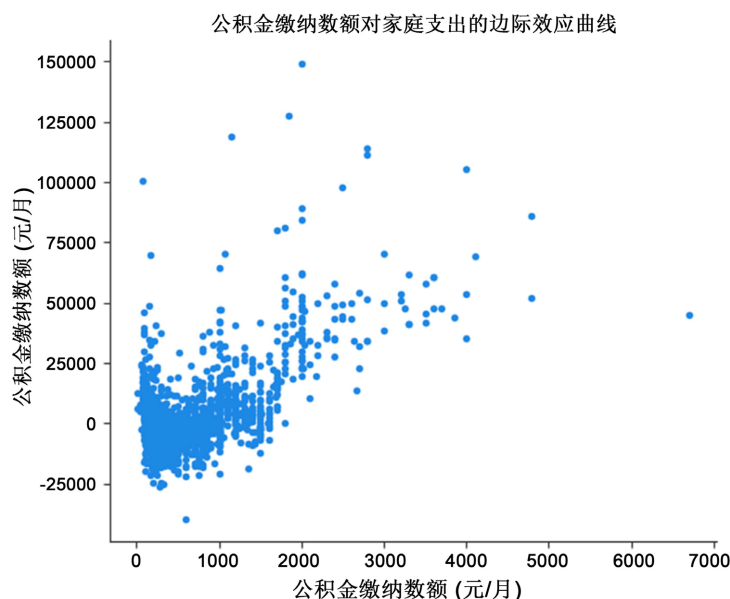


Figure 5. SHAP dependency plot of monthly housing provident fund contributions on projected total household expenditures

图 5. 公积金月缴存额对家庭总支出预测值的 SHAP 依赖图

公积金月缴存额与家庭消费预测值之间并非简单线性关系，而是呈现明显阶段性特征。公积金月缴存额低于 1000 元时，SHAP 值接近零，对消费的边际作用有限；缴存额在 1000~2000 元时，SHAP 值快速上升，对消费的正向拉动明显增强，这一区间是消费响应由弱转强的关键；超过 3000 元后，正向作用依然存在，但边际增幅逐步放缓，高缴存家庭的额外提升空间有限。

需要说明的是，SHAP 结果仅反映预测层面的边际变化，不能直接作为政策因果阈值，主要为后续异质性分析提供依据。

6. 因果识别：消费刺激的异质性与靶向画像

6.1. 因果森林模型设定与解释边界

在 SHAP 揭示非线性关系和交互机制后，本文进一步使用因果森林模型估计公积金月缴存额对家庭消费的边际处理效应。该方法在控制高维可观测混杂因素基础上，估计平均处理效应和个体条件平均处理效应。

本文将公积金月缴存额作为处理变量，家庭年度总支出作为结果变量，收入结构、年龄、性别、学历、婚姻状态和户口类型等作为控制变量[11]。由于使用截面数据且处理变量为公积金月缴存额而非外生政策冲击，结果应解释为“在可观测混杂因素可控条件下的边际处理效应”，有助于缓解收入等混杂因素带来的偏误，但不能完全排除不可观测内生性。

6.2. 平均处理效应估计

根据图 6 因果森林模型估计结果显示，在控制家庭收入结构、人口特征和户籍类型等可观测混杂因素后，公积金月缴存额对家庭年度总支出具有正向平均处理效应。模型估计的 ATE 为 13.84，在控制收入、人口特征等可观测变量后，模型估计得到平均处理效应(ATE)为 13.84。若折算为年度缴存口径，年

化消费响应系数约为 1.15:

$$\frac{13.84}{12}=1.15$$

(3)

首先，从永久收入理论和生命周期理论来看，家庭消费不仅取决于当期收入，还受到未来收入预期和生命周期资源配置的影响。较高的公积金缴存额通常意味着家庭收入更加稳定、未来住房保障水平更高，有助于改善长期收入预期，降低预防性储蓄动机，从而促进家庭提前释放消费需求[12]。因此，消费增长可能不仅来自公积金缴存本身，还受到未来收入预期改善的影响。其次，从财富效应和流动性约束缓解角度来看，住房公积金账户积累增强了家庭对未来住房保障的信心，降低了住房支出的不确定性，使家庭能够将更多资金配置于教育、文化娱乐等发展型消费。

本文估计得到的消费响应系数并不意味着每新增 1 元公积金缴存必然带来 13.84 元消费增长，而是在双重机器学习和因果森林框架下，控制可观测混杂变量并满足条件独立假设时估计得到的平均处理效应。该效应反映的是收入预期改善、财富效应、流动性约束缓解等多种机制共同作用的综合结果，而非公积金缴存本身的直接消费乘数。

从个体条件平均处理效应分布看，CATE 在不同家庭之间存在较大差异。部分家庭的处理效应接近零甚至为负，另有部分家庭的处理效应明显高于均值，说明公积金消费效应并非均匀分布在所有家庭中。

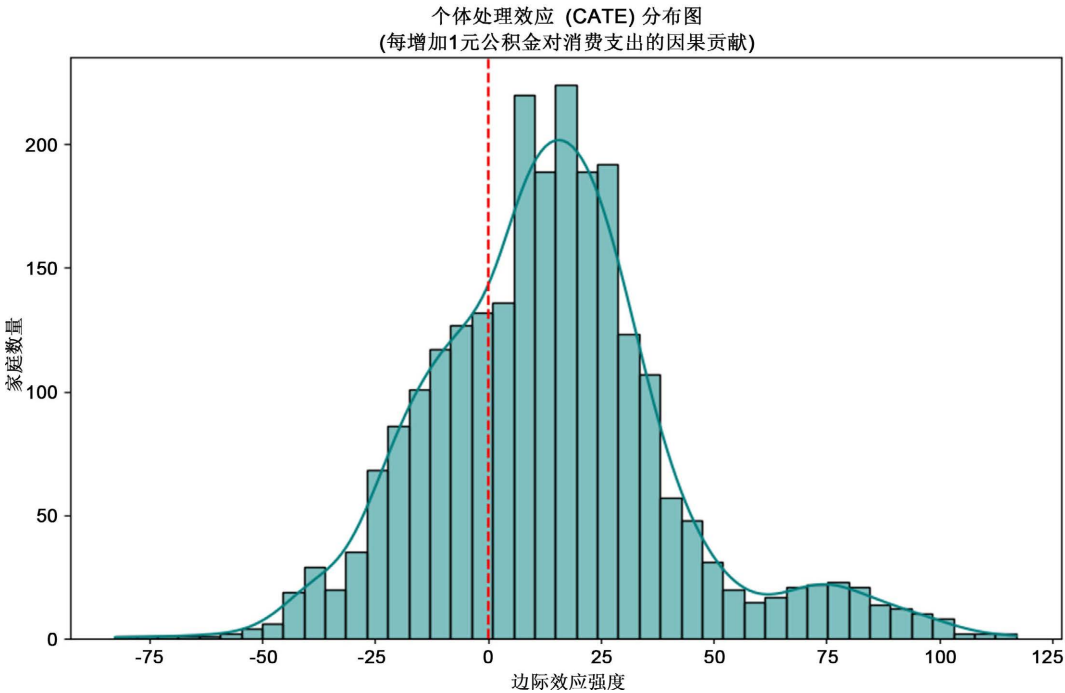


Figure 6. Distribution of individual processing effects on the consumption effect of housing provident funds
图 6. 公积金消费效应的个体处理效应分布

这表明，公积金消费效应并非均匀分布在所有家庭中。平均处理效应只能反映总体方向，真正具有政策意义的是识别哪些家庭能够有效释放消费，哪些家庭存在政策传导不足。

6.3. 收入异质性分析

见图 7 CATE 分布显示，家庭纯收入越高，公积金的消费拉动效应越强。低收入家庭 CATE 波动大，

部分甚至为负，说明收入偏低时公积金缴存可能挤占当期现金流，消费提振效果不稳定。中高收入家庭 CATE 更高且更稳定，现金流充足、预算压力小，能把公积金的住房保障与流动性优势更快转化为实际消费。

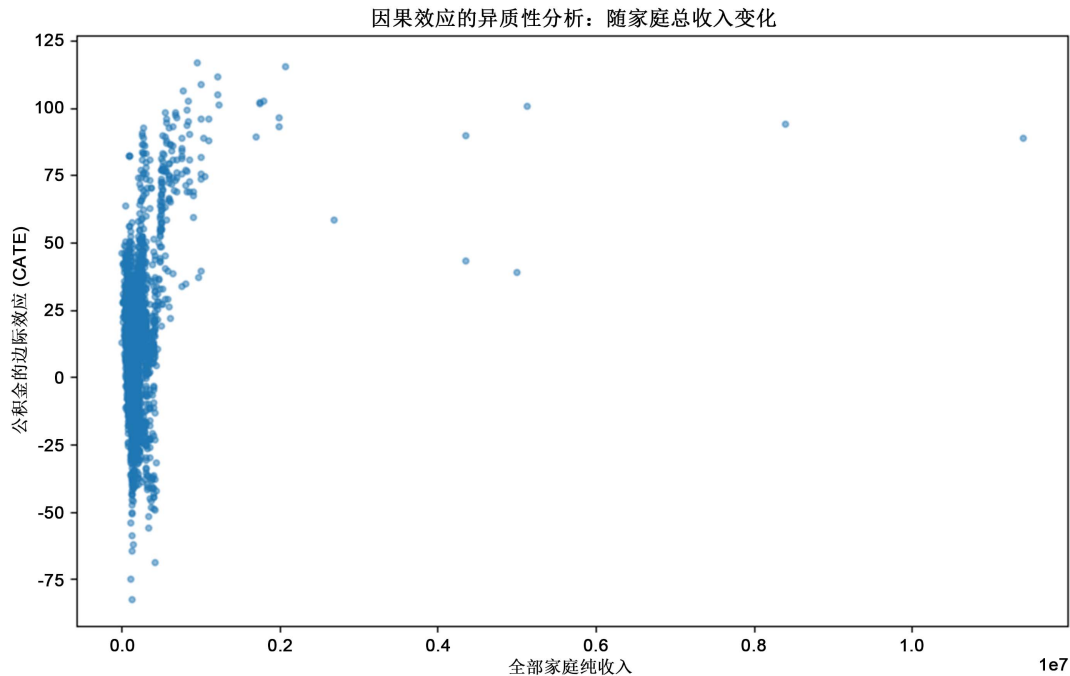


Figure 7. CATE distribution of the consumption effect of housing provident fund contributions as a function of household net income
图 7. 公积金消费效应随家庭纯收入变化的 CATE 分布

6.4. 模型可信度说明

为保证实证结果可靠，本文从三方面开展稳健性检验：一是替换消费统计口径，区分基础食品消费与文教娱乐发展型消费，核心结论不受指标选取影响；二是对连续变量做前后 1%缩尾，规避极端值干扰[13]；三是严格区分模型功能，通过随机森林判定变量重要性、SHAP 拆解非线性作用、因果森林测算异质性效应，分层递进分析[14]。综合检验可见，本文实证结论前后逻辑统一、结果稳健。

7. 结论与政策建议

7.1. 主要研究结论

研究得出四点核心结论：

第一，公积金能有效解释家庭消费变化，但作用依附于收入水平。收入仍是决定消费的首要因素，公积金影响力虽不及收入，却远超常规人口特征变量，早已不只是住房保障工具，还能通过流动性改变家庭消费选择。第二，公积金缴存额度和消费拉动效果并非线性变化。月缴存额千元以内拉动作用有限，1000 至 2000 元区间效果大幅提升，突破 3000 元后增长势头放缓，呈现先升后平的变化规律。第三，在控制可观测混杂变量后，模型估计结果表明公积金缴存与家庭消费之间存在显著正向关联。月缴存额每增加 1 元，家庭年消费约增加 13.84 元，年化响应系数为 1.15。该结论仅适用于微观家庭样本，不能直接套用为宏观政策效应。第四，公积金消费带动效果存在明显群体差异。高收入、城镇户籍、中老年家庭

消费响应更突出，青年和农村家庭相对偏弱，效果更多集中在收入稳定、住房负担小、能便捷享受政策的群体。

综上，公积金对不同家庭的消费刺激效果并不均等，受缴存水平、收入条件、家庭年龄结构及政策可及性共同影响。政策完善无需一味放宽缴存额度，重点应精准识别消费释放潜力群体，疏通政策传导受阻环节，实行差异化、场景化的制度调整。

7.2. 政策建议

基于异质性研究结果，本文提出以下政策建议：

1) 从“一刀切放宽”转向“分群体精准优化”

公积金消费效应存在明显群体差异，中高收入、城镇及中老年家庭响应更强，低收入、青年和农村家庭带动效果偏弱。对消费响应较好的群体，应持续简化流程、完善按月提取和租房直付等服务；对低收入群体则重点降低使用门槛，避免缴存占用日常流动资金。

2) 针对青年家庭缓解住房绑定，释放一般消费空间

青年家庭公积金多被租房、还贷占用，难以转化为日常消费^[15]。应完善青年与新市民租房提取规则，实现按月自动到账；在把控风险前提下，适度拓宽使用范围，支持职业培训、育儿托育及耐用品购置，助力提升消费与长远发展能力。

3) 对低收入家庭建立缴存与提取的弹性机制

低收入家庭缴存行为易挤占可支配收入，压制短期消费。可适当下调最低缴存比例，允许在失业、大病、高额房租等特殊情形下优先提取；同时联动住房补贴、公租房保障等政策，防止制度拉大收入差距。

参考文献

- [1] 万广华, 张茵, 牛建高. 流动性约束、不确定性与中国居民消费[J]. 经济研究, 2001, 36(11): 35-44+94.
- [2] 孙凤. 预防性储蓄理论与中国居民消费行为[J]. 南开经济研究, 2001(1): 54-58.
- [3] 周绍杰, 张俊森, 李宏彬. 中国城市居民的家庭收入、消费和储蓄行为: 一个基于组群的实证研究[J]. 经济学(季刊), 2009, 8(4): 1197-1220.
- [4] 尹志超, 郭沛瑶. 精准扶贫政策效果评估——家庭消费视角下的实证研究[J]. 管理世界, 2021, 37(4): 64-83.
- [5] 康书隆, 余海跃, 刘越飞. 住房公积金、购房信贷与家庭消费——基于中国家庭追踪调查数据的实证研究[J]. 金融研究, 2017(8): 67-82.
- [6] 雷心悦, 杜丽群. 放宽公积金提取条件对家庭消费的影响——基于 CHFS 数据的实证研究[J]. 河南大学学报(社会科学版), 2024, 64(5): 8-16+152.
- [7] 李伟军, 邓国营. “城市梦”: 住房公积金与流动人口居留意愿[J]. 浙江工商大学学报, 2024(1): 110-121.
- [8] 詹鹏, 万海远, 李实. 住房公积金与居民收入分配——基于可计算一般均衡模型的研究[J]. 数量经济技术经济研究, 2018, 35(9): 22-40.
- [9] 刘景江, 郑畅然, 洪永淼. 机器学习如何赋能管理学研究?——国内外前沿综述和未来展望[J]. 管理世界, 2023, 39(9): 191-216.
- [10] 陈燕君. 线上消费者购买行为的决策机制与可解释性研究[D]: [硕士学位论文]. 广州: 广东工业大学, 2024.
- [11] 王元彬, 张尧, 李计广. 数字金融与碳排放: 基于微观数据和机器学习模型的研究[J]. 中国人口·资源与环境, 2022, 32(6): 1-11.
- [12] 田坤, 黄坤, 行伟波. 数字经济、市场潜能与乡村振兴——基于双重机器学习的因果推断[J]. 山西财经大学学报, 2023, 45(11): 73-85.
- [13] 李伟军, 吴义东. 住房公积金、金融知识与新市民住房租购决策——基于 CHFS 的证据[J]. 中南财经政法大学学报, 2019(4): 139-148.

-
- [14] 张寒, 刘璨, 刘浩. 林地调整对农户营林积极性的因果效应分析——基于异质性视角的倾向值匹配估计[J]. 农业技术经济, 2017(1): 37-51.
- [15] 王耀中, 黄选爱, 胡尊国. 绿色创新对企业环境绩效的影响研究——基于因果森林算法的分析[J]. 财经理论与实践, 2024, 45(3): 125-130.

附录 数据清洗过程、样本筛选标准及描述性统计

1) 数据来源与样本构建

本文数据来源于中国家庭追踪调查 2022 年调查数据。本文研究对象为家庭层面的消费行为，因此主要使用 CFPS 2022 年家庭经济数据和个人成人数据，通过家庭编号(Family ID)进行匹配，构建包含家庭消费、公积金缴存、收入结构及人口特征等信息的微观家庭数据库。

2) 数据清洗过程

首先，根据家庭编号(Family ID)对家庭经济数据库与个人成人数据库进行匹配，将家庭消费、收入、公积金缴存情况以及家庭成员人口特征等信息整合至家庭层面，构建用于实证分析的家庭微观数据库。在变量筛选过程中，保留家庭年度总支出、食品支出、文教娱乐支出等消费指标，选取住房公积金月缴存额作为核心解释变量，并纳入家庭纯收入、工资性收入、经营性收入、财产性收入、转移性收入以及年龄、性别、教育程度、婚姻状态和户口类型等控制变量，以降低家庭经济状况和人口特征差异带来的估计偏误。

在数据预处理阶段，首先删除家庭消费、公积金缴存额等核心变量缺失的样本；对于控制变量中存在的少量缺失值，连续型变量采用中位数填补，分类变量采用众数填补，以尽可能保留有效样本信息。考虑到家庭收入、消费支出以及公积金缴存额等变量通常存在右偏分布，且极端值可能影响机器学习模型训练效果，本文对连续变量进行上下 1%的缩尾处理，降低异常观测对模型估计结果的干扰。

经过上述数据匹配、变量筛选、缺失值处理及异常值调整后，最终获得 2465 个有效家庭样本。