Statistics and Application 统计学与应用, 2026, 15(9), 195-204
Published Online September 2026 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2026.159218

Hans汉斯

面向类别不平衡的家庭非房产债务压力风险预测

——基于代价敏感XGBoost的研究

黄炜坚

福建师范大学数学与统计学院，福建 福州

收稿日期：2026年8月18日；录用日期：2026年9月10日；发布日期：2026年9月21日

摘　要

家庭非房产债务往往对应消费信贷、经营周转及亲友借款等短期资金需求，其债务负担上升可能加剧家庭金融脆弱性。本文利用中国家庭追踪调查(CFPS) 2010~2022年数据构建家庭－年份非平衡面板，以下一调查期非房产债务收入比超过50%识别债务压力风险，并以当期家庭特征实施时间外推预测。针对高风险样本占比较低、漏识别成本较高的特点，本文在XGBoost框架中引入类别代价、风险强度动态权重和面向F2准则的分类阈值选择，同时以风险评分分组考察模型的筛查效能。结果表明：风险强度动态加权模型在测试集上的AUC为0.7426、KS为0.3650，召回率和F2值分别为0.6683和0.4512；风险评分最高10%家庭的实际风险发生率为33.76%，约为总体风险率的3.33倍。更换风险定义、滚动时间窗口以及剔除当期直接债务变量后，模型仍保持稳定的样本外区分能力。特征组消融结果进一步显示，资产缓冲和债务惯性是提高预测能力的主要信息来源。本文的贡献在于将家庭债务研究由当期相关性描述扩展为面向未来状态的预测问题，并给出适用于类别不平衡家庭调查数据的风险排序与筛查框架。

关键词

家庭非房产债务，债务压力风险，类别不平衡，代价敏感学习，XGBoost

Predicting Household Non-Housing Debt Pressure under Class Imbalance

—A Cost-Sensitive XGBoost Approach

Weijian Huang

School of Mathematics and Statistics, Fujian Normal University, Fuzhou Fujian

Received: August 18, 2026; accepted: September 10, 2026; published: September 21, 2026

文章引用: 黄炜坚. 面向类别不平衡的家庭非房产债务压力风险预测[J]. 统计学与应用, 2026, 15(9): 195-204.
DOI: 10.12677/sa.2026.159218

Abstract

Non-housing debt typically stems from short-term funding needs—such as consumer credit, business working capital, and loans from friends or family members—and a rising debt burden may exacerbate household financial vulnerability. This paper constructs an unbalanced household-year panel using data from the China Family Panel Studies (CFPS) spanning 2010 to 2022. It identifies debt pressure risk based on a non-housing debt-to-income ratio exceeding 50% in the subsequent survey wave and employs current household characteristics for out-of-time predictive modeling. Addressing the challenges posed by a low proportion of high-risk samples and the high cost of false negatives, the study incorporates class-specific costs, dynamic weighting based on risk intensity, and F2-oriented classification threshold selection into the XGBoost framework; it further evaluates the model's screening performance by segmenting households according to risk scores. Results indicate that the dynamic risk-intensity weighted model achieves an AUC of 0.7426 and a KS statistic of 0.3650 on the test set, with recall and F2 values of 0.6683 and 0.4512, respectively. Among the top 10% of households by risk score, the actual risk incidence rate is 33.76%, approximately 3.33 times the overall risk rate. The model maintains stable out-of-sample discriminative power even when the risk definition is altered, rolling time windows are applied, or current direct debt variables are excluded. Feature group ablation analysis further reveals that asset buffers and debt inertia are the primary sources of information for enhancing predictive capability. This study contributes to the literature by shifting the focus of household debt research from describing contemporaneous correlations to predicting future states, and by providing a risk ranking and screening framework tailored to class imbalance household survey data.

Keywords

Household Non-Housing Debt, Debt Pressure Risk, Class Imbalance, Cost-Sensitive Learning, XGBoost

1. 引言

居民部门债务扩张与家庭金融脆弱性是家庭金融研究的重要议题。债务本身并不必然意味着风险：合理借贷能够平滑消费、支持人力资本投资和经营活动；但当债务负担超过持续收入和资产缓冲能力时，家庭可能通过压缩消费、延迟支出或继续借款来缓解流动性约束，从而增加财务脆弱性。已有研究已从债务杠杆、偿债能力、数字金融和社会网络等角度讨论家庭债务行为及其福利后果[1]-[5]。相较于住房贷款，非房产债务通常缺乏稳定抵押物和长期还款安排，更易反映消费、经营周转和临时现金流缺口所形成的短期压力。

现有家庭债务研究主要关注当期债务水平及其影响因素，而对家庭在未来调查期是否进入债务压力状态的识别相对有限。该差异并非单纯的技术问题：当研究目标由“解释当期债务”转为“识别未来风险”时，模型必须遵循时间顺序，并将评价重点由拟合优度转向样本外排序、风险覆盖和筛查效率。尤其在风险家庭占比较低时，准确率容易被多数类主导，默认 0.5 阈值也未必符合“尽量减少漏识别”的风险管理目标。

机器学习方法能够刻画收入、资产、支出、人口结构及历史债务之间的非线性关系和交互作用，在结构化表格数据中具有较强适用性[6]-[8]。但直接比较不同模型的 AUC 并不足以回应预警问题：AUC 衡量总体排序能力，不能决定应将多少家庭纳入重点筛查名单；阈值选择则对应具体的漏识别与误报权衡。基于此，本文以 XGBoost 为基础，将类别代价、风险强度动态加权和阈值选择纳入统一预测框架，并通过风险评分分组、时间滚动验证、替代标签和特征组消融检验结果的稳定性与信息来源。

本文有三点贡献。第一，将非房产债务收入比所刻画的家庭偿债压力转化为下一期风险状态的预测任务，强调时间外推而非当期拟合。第二，针对少数类风险识别，提出以风险严重程度调节正样本权重的训练方案，并将 F_2 准则用于决策阈值选择。第三，从风险排序、稳健性和特征组增量价值三个层面报告结果，避免将预测模型输出误解为变量的因果效应。

2. 数据、变量与预测任务

2.1. 数据来源与样本划分

本文使用中国家庭追踪调查(China Family Panel Studies, CFPS) 2010~2022 年家庭数据。CFPS 覆盖家庭收入、消费、资产、负债、健康、教育、就业及社区环境等信息，能够支持家庭层面的动态观察。本文首先将数据整理为家庭–年份非平衡面板；对同一家庭同一年出现的个人层面信息，按家庭特征聚合为均值、最高值或比例指标；对家庭总量变量保留其家庭口径。样本清理后，按照目标年份进行时间顺序划分，避免随机划分造成未来信息进入训练过程，具体样本划分情况及风险样本比例见表 1。

Table 1. Sample split and proportion of risk households
表 1. 样本划分与风险样本比例

样本	样本量	目标年份	风险样本比例
训练集	42,201	2012, 2014, 2016, 2018	0.1267
验证集	7732	2020	0.1094
测试集	6213	2022	0.1014

注：风险样本指下一调查期非房产债务收入比超过 50%的家庭。

2.2. 风险标签与特征体系

CFPS 并不提供家庭真实违约记录，因此本文研究的是“非房产债务压力风险”，而非严格意义上的贷款违约风险。则债务收入比定义为：

$$DTR_{it} = \frac{D_{it}^{\mathrm{NH}}}{Y_{it}} \tag{1}$$

若家庭在下一调查期的债务收入比超过 50%，则定义其进入债务压力风险状态。有：

$$R_{i,t+1} = \mathbf{1}\left(DTR_{i,t+1} > 0.50\right) \tag{2}$$

为降低标签阈值选择对结论的影响，后文进一步使用 30%阈值、训练期分位数阈值及“是否存在非房产负债”开展替代标签检验。特征变量不直接用于因果识别，而是服务于预测任务，主要包括基础人口与地区、偿债能力、支出冲击、资产缓冲、债务惯性以及宏观和社会经营六类信息，具体特征组及其经济含义见表 2。

Table 2. Feature groups and economic meanings
表 2. 特征组及其经济含义

特征组	主要信息	预测含义
基础人口与地区	户主年龄、家庭规模、教育、婚姻、城乡及地区	刻画家庭生命周期、信息条件与区域环境
偿债能力	家庭收入、支出、收支缺口、消费收入比	反映持续现金流与偿债能力
支出冲击	医疗、教育支出，不健康成员与老龄人口占比	反映刚性支出与突发支出压力
资产缓冲	净资产、金融资产、存款和资产收入比	反映短期流动性与风险承受能力
债务惯性	历史负债、历史风险比例与连续负债状态	反映债务状态的持续性
宏观与社会经营	地区经济、产业结构与家庭经营活动	补充外部条件和经营性现金流信息

注：具体变量均取自风险发生前的家庭信息；模型输出应理解为风险排序分数，而非结构参数或因果效应。

2.3. 预测任务与评价原则

本文以家庭 i 在 t 期的特征向量 X_{it} 预测其在下一调查期进入债务压力状态的条件风险，记为 $p_{i,t+1}=P\left(R_{i,t+1}=1\mid X_{it}\right)$。该设定使训练、验证和测试均遵循时间顺序。预测目标是将可能发生债务压力的家庭优先排列到高分组，而不是据此识别某一变量对风险的净因果影响。

$$p_{i,t+1}=P\left(R_{i,t+1}=1\mid X_{it}\right) \tag{3}$$

为保证风险排序与分类筛查的评价相互区分，本文同时报告 AUC、KS、Precision、Recall、F_2 以及前 10%风险提升倍数。AUC 刻画随机抽取一户风险家庭与一户非风险家庭时，前者获得更高预测分数的概率；KS 刻画不同阈值下两类样本分布差异的最大值，分别定义为：

$$\mathrm{AUC}=P\left(\hat{p}^{+}>\hat{p}^{-}\right) \tag{4}$$

$$\mathrm{KS}=\max_{c\in[0,1]}\left|\mathrm{TPR}\left(c\right)-\mathrm{FPR}\left(c\right)\right| \tag{5}$$

3. 代价敏感预测方法

3.1. 代价敏感 XGBoost 与风险强度权重

XGBoost 通过迭代提升树拟合条件风险函数，并在损失函数中加入复杂度惩罚。考虑到风险家庭为少数类，普通训练目标容易偏向非风险家庭。本文对风险样本引入额外权重，使模型在训练阶段更重视漏识别的风险家庭。设 $l(\cdot)$为二元对数损失，$\Omega(f)$为模型复杂度项，则加权目标可写为：

$$\min_{f}\sum_{i=1}^{n}w_i\ell_i+\Omega\left(f\right) \tag{6}$$

其中，训练样本中非风险家庭和风险家庭数量分别记为 N_0 和 N_1，基础正样本惩罚系数取二者之比；风险家庭的严重程度以超过阈值 τ 的债务收入比进行标准化，并由参数 α 调节其额外训练权重。非风险家庭权重为 1；对风险家庭，相应的基础惩罚系数、严重程度与样本权重可分别表示为：

$$\lambda=\frac{N_0}{N_1} \tag{7}$$

$$s_i=\frac{DTR_{i,t+1}-\tau}{\max\limits_{j:y_j=1}DTR_{j,t+1}-\tau} \tag{8}$$

$$w_i=\lambda\left(1+\alpha s_i\right),\ y_i=1 \tag{9}$$

参数 α 的确定仅使用训练与验证阶段信息，不接触测试集。由于本文的核心目标是时间外推预测，

为避免随机 K 折划分将未来年份信息混入训练过程，本文不采用随机交叉验证，而是在独立的 2020 年验证集上进行网格搜索。候选集合设为 $\alpha \in \{0.25, 0.50, 1.00, 1.50, 2.00, 3.00\}$；对每个候选 α 固定其余模型设定，并在验证集上同步选择使 F_2 最大的分类阈值，以验证集 F_2 作为主要选择准则，同时参考 Recall 与 Precision 之间的权衡。最终保留 $\alpha = 2$，并在 2022 年测试集上进行一次性样本外评价。

Table 3. Sensitivity of key metrics to the dynamic weight parameter α
表 3. 动态权重参数 α 的敏感性结果

α	AUC	KS	Precision	Recall	F_2	前 10%提升倍数
0.25	0.7424	0.3524	0.1919	0.6651	0.4455	3.4620
0.50	0.7446	0.3640	0.1927	0.6651	0.4463	3.4302
1.00	0.7428	0.3634	0.1915	0.6698	0.4467	3.4620
1.50	0.7448	0.3747	0.1799	0.7175	0.4490	3.3349
2.00	0.7426	0.3650	0.1963	0.6683	0.4512	3.3508
3.00	0.7420	0.3616	0.1779	0.7317	0.4509	3.2714

注：每个 α 对应的分类阈值均在 2020 年验证集上按 $F2$ 最大化原则确定，表中报告 2022 年测试集表现；参数搜索和阈值选择均未使用测试集信息。

从表 3 可以看出，α 变化主要改变 Precision 与 Recall 之间的权衡：随着 α 提高，模型总体上更倾向于覆盖风险家庭，Recall 上升，而 Precision 有所下降。在所考察区间内，F_2 仅在 0.4455~0.4512 之间波动，$\alpha = 2$ 时取得最高 F_2 (0.4512)，$\alpha = 3$ 时 F_2 为 0.4509，差异很小。这说明主结论并不依赖某一个极端的 α 取值，同时也表明继续增大风险样本权重并不会带来单调的综合性能改善。

3.2. 阈值选择与风险排序

模型训练后，预测分数必须经由阈值转化为“纳入重点关注”或“暂不纳入”的分类决策。默认阈值 0.5 隐含了对两类错误成本的特定假设，未必适合风险预警。本文在验证集中选择使 F_2 最大的阈值；F_2 相对 F_1 给予召回率更高权重，适合漏识别成本高于误报成本的前置筛查场景，其定义为：

$$F_2 = \frac{5PR}{4P + R}, \quad P = \text{Precision}, \quad R = \text{Recall} \tag{10}$$

在有限筛查资源下，本文进一步将风险评分最高 10%家庭的实际风险率(记为 $\bar{R}_{\text{Top }10\%}$)与总体实际风险率(记为 $\bar{R}$)之比定义为前 10%风险提升倍数，用以衡量高风险名单的集中筛查效能：

$$\text{Lift}_{10\%} = \frac{\bar{R}_{\text{Top }10\%}}{\bar{R}} \tag{11}$$

需要指出，阈值优化和前 10%名单划分改变的是分类操作点与筛查容量，并不改变模型本身的概率排序能力；因此，本文将其与加权训练产生的排序表现分开报告。

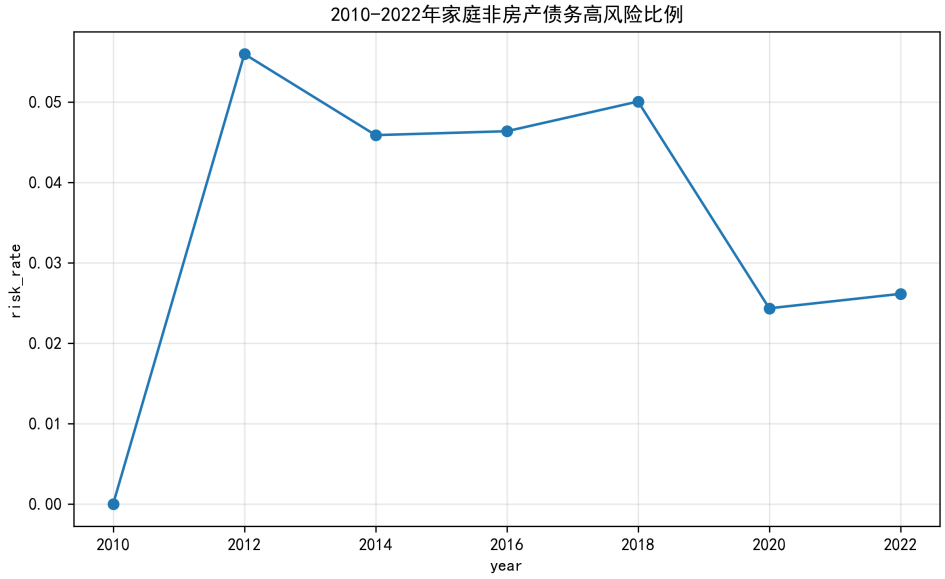
在模型融合部分，本文将代价敏感 XGBoost 得分与 Logit 得分进行线性组合，以检验风险排序的稳定性。融合不取代主模型，而作为对高分组风险集中度的补充检验。

为评价 XGBoost 相对于传统统计模型的增量表现，本文另设带类别权重的 Logit 模型作为基准。Logit 采用与主模型一致的特征集合和训练–验证–测试年份划分；数值变量按训练集尺度进行标准化，类别变量进行哑变量编码。类别权重按照训练集中类别频率自动平衡，使少数类风险样本在估计中获得更高权重；其分类阈值同样仅在 2020 年验证集上按 F_2 最大化原则选择，从而保证与 XGBoost 的比较口径一致。

4. 实证结果

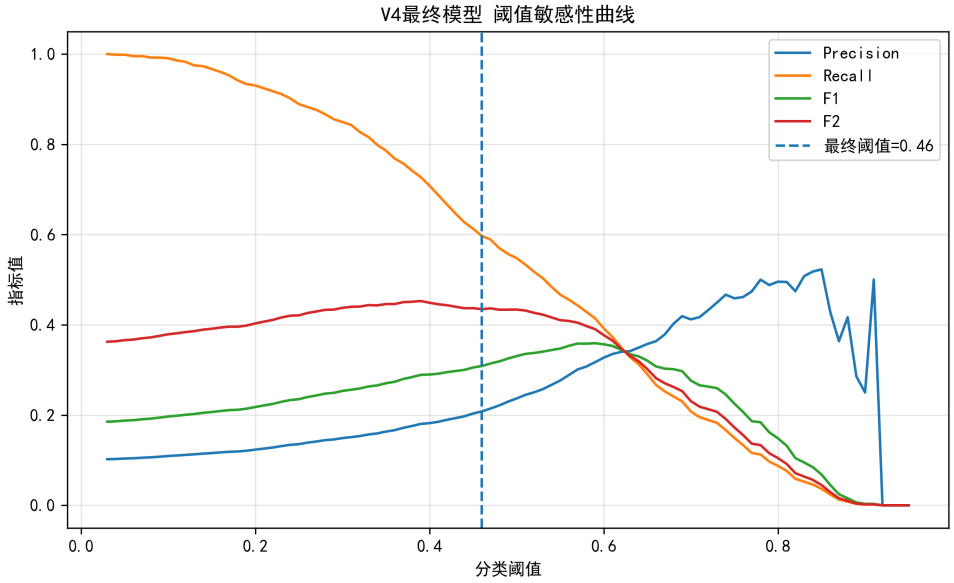
4.1. 风险发生特征与阈值选择

测试集风险家庭比例为 10.14%，表明即使使用 50%债务收入比阈值，风险样本仍明显少于非风险样本。图 1 显示，样本期内债务压力风险比例随时间波动，说明家庭债务压力具有动态性；这也支持使用历史信息预测后续状态，而不是将家庭债务视为静态结果。



注：资料来源：根据 CFPS 数据计算。

Figure 1. Trend in household non-housing debt pressure risk rate
图 1. 家庭非房产债务压力风险比例趋势



注：资料来源：根据验证集模型输出计算。

Figure 2. Sensitivity analysis of classification threshold
图 2. 分类阈值的敏感性分析

图 2 报告验证集上的阈值敏感性。随着阈值下降，模型能够覆盖更多高风险家庭，但同时会增加被识别为风险的非风险家庭。F_2 最大的阈值被作为主报告阈值，其目的在于明确预警场景中风险覆盖与筛查成本之间的取舍。

4.2. 不同优化策略的样本外表现

表 4 比较类别权重 Logit 与各 XGBoost 优化方案。各方法 AUC 均在 0.74 左右，说明模型复杂度对总体排序上限的影响有限；因此本文重点比较 Recall、Precision、F_2 和前 10%提升倍数，考察不同方法在风险覆盖、误报与高风险集中度之间的权衡。

Table 4. Out-of-sample performance of different optimization strategies
表 4. 不同优化策略的测试集表现

方法	AUC	KS	Precision	Recall	F_2	前 10%提升倍数
类别权重 Logit	0.7382	0.3631	0.1704	0.7582	0.4486	3.1098
基础 XGBoost	0.7441	0.3557	0.1955	0.6397	0.4399	3.3985
代价敏感 XGBoost	0.7430	0.3483	0.2046	0.6048	0.4347	3.4461
动态加权 XGBoost ($\alpha = 2$)	0.7426	0.3650	0.1963	0.6683	0.4512	3.3508
难样本再加权 XGBoost	0.7420	0.3570	0.2055	0.6190	0.4414	3.3985
CS-XGB + Logit 融合	0.7449	0.3562	0.2065	0.6079	0.4377	3.5096

注：类别权重 Logit 采用训练集类别频率平衡权重，其分类阈值按验证集 F2 选择；动态加权模型的 α 和分类阈值均基于验证集选择；融合模型按风险评分前 10%提升倍数选择。前 10%提升倍数为最高十分位风险率与总体风险率之比。

带类别权重的 Logit 在测试集上的 AUC 为 0.7382、Recall 为 0.7582、F_2 为 0.4486，说明传统线性基准在类别加权和阈值优化后同样具有较强的风险覆盖能力。风险强度动态加权 XGBoost 的 AUC 为 0.7426、Precision 为 0.1963、F_2 为 0.4512，分别高于 Logit 的 0.7382、0.1704 和 0.4486，其前 10%风险提升倍数也由 3.1098 提高至 3.3508，但 Recall 低于 Logit。由此可见，XGBoost 带来的增益并非体现在所有指标上均显著领先，而主要表现为非线性排序、高分组风险集中度以及 Precision-Recall 综合权衡的改善。融合模型的前 10%风险提升倍数为 3.5096，进一步表明在筛查容量严格受限时，可将排序表现更强的融合分数作为辅助工具；若首要目标是尽可能少漏掉风险家庭，则仍需结合 Recall 和筛查成本选择操作点。

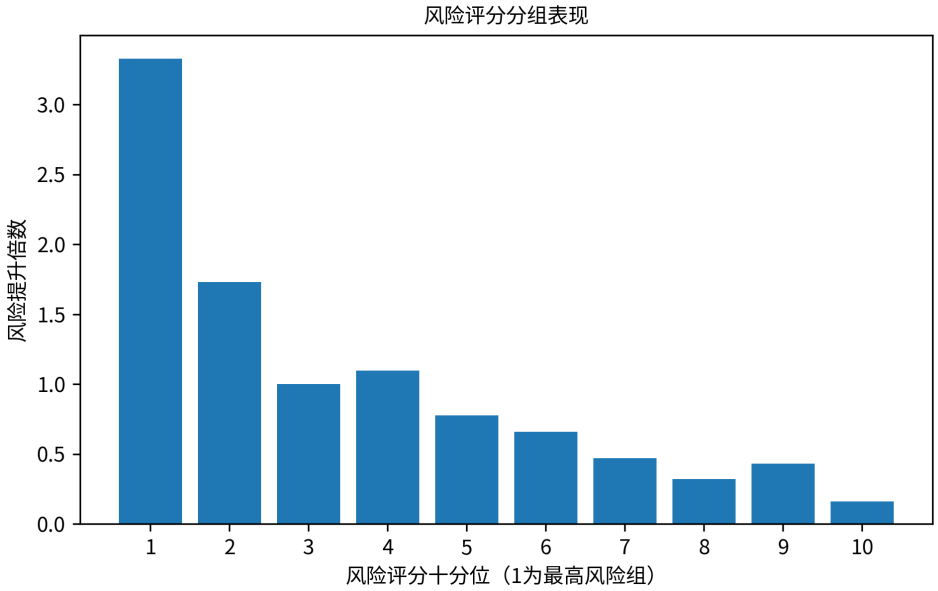
4.3. 风险评分分组与筛查价值

对于实际风险管理而言，模型通常用于生成分层名单，而不是直接替代人工判断。图 3 给出风险评分分组的实际发生率。最高十分位家庭的风险率为 33.76%，约为总体风险率的 3.33 倍；第二十分位的风险率为 17.55%，此后风险率明显下降。该结果说明模型能够将相对更多的真实风险家庭集中于高分组，具有优先筛查的应用价值。

5. 稳健性与信息来源分析

5.1. 替代标签、时间滚动与信息泄露检验

为检验主要结论是否依赖于单一风险阈值，本文依次使用债务收入比 30%、训练期 Q_{75}、训练期 Q_{80} 和“是否存在非房产负债”构造替代标签。表 5 显示，在不同标签下模型均保持正向的样本外区分能力，前 10%风险提升倍数介于 2.7898 和 4.2498 之间。风险事件越稀少时，Precision 和 F_2 会更易受基准发生率影响，因此不宜仅依据单一指标进行横向比较。



注：资料来源：根据测试集预测分数和实际风险标签计算。

Figure 3. Realized risk rate by risk score group
图 3. 风险评分分组的实际风险发生率

Table 5. Robustness results under alternative labels and rolling windows
表 5. 替代标签与时间滚动的稳健性结果

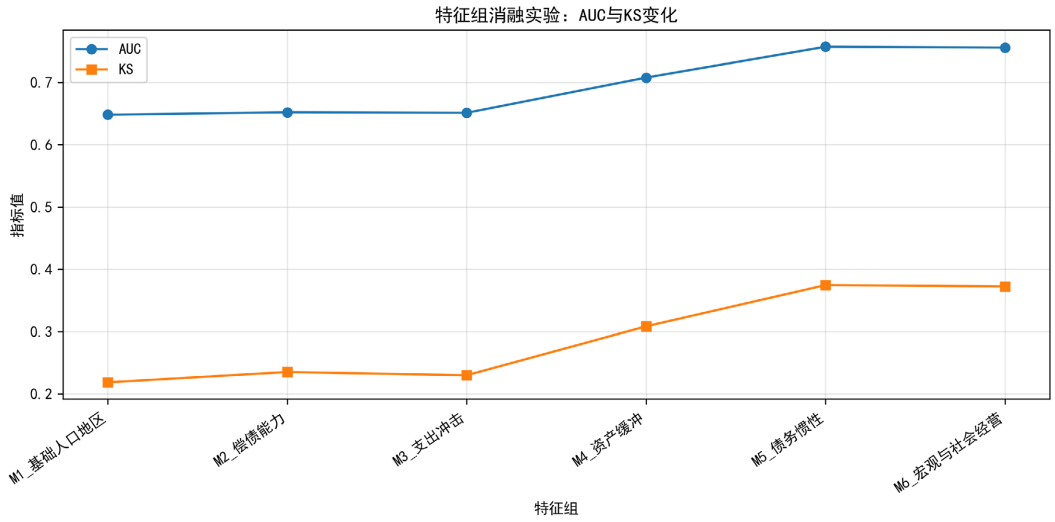
检验类型	AUC	KS	Recall	F_2	前 10%提升倍数
风险标签：债务收入比 ＞30%	0.7495	0.3575	0.8396	0.4947	3.3134
风险标签：训练期 Q_{75}	0.7668	0.4664	0.2919	0.2318	4.0392
风险标签：训练期 Q_{80}	0.7616	0.4542	0.2743	0.1902	4.2498
风险标签：是否存在非房产负债	0.7401	0.3517	0.9019	0.5558	2.7898
滚动窗口：2012~2018 训练、2022 测试	0.7557	0.3725	0.7316	0.4577	3.6438
剔除当期直接债务变量	0.7505	0.3674	0.7630	0.4578	3.3611

注：替代标签结果和滚动窗口结果均来自独立时间外测试；“剔除当期直接债务变量”不包含当期非房产债务金额、债务收入比和是否存在非房产负债等变量。

此外，使用 2012~2018 年样本训练、2020 年验证、2022 年测试的滚动窗口设置下，AUC 仍为 0.7557，前 10%风险提升倍数为 3.6438。剔除当期直接债务变量后，模型的 AUC 为 0.7505、F_2 为 0.4578，表明收入、资产、支出结构和历史状态本身包含可用于风险识别的信息。该结果并不意味着债务变量不重要，而是说明模型并非完全依赖“当期债务预测下一期债务”的机械延续。

5.2. 特征组消融与异质性

特征组消融实验按基础人口地区、偿债能力、支出冲击、资产缓冲、债务惯性和宏观社会经营信息逐步扩展特征集合。具体消融结果见表 6。图 4 显示，加入资产缓冲后 AUC 和 KS 明显改善；在加入债务惯性特征后，模型的排序和筛查能力进一步提升。这一结果与家庭金融脆弱性的理论预期一致：资产缓冲决定家庭应对短期冲击的能力，而历史债务状态包含借贷约束和现金流压力的持续信息。



注：资料来源：根据特征组消融实验计算。

Figure 4. AUC and KS changes after sequential addition of feature groups
图 4. 特征组逐步加入后的 AUC 与 KS 变化

Table 6. Feature group ablation results
表 6. 特征组消融结果

特征组模型	特征数	AUC	KS	Recall	F_2	前 10%提升倍数
M_1 基础人口地区	36	0.6480	0.2186	0.8006	0.3862	2.0104
M_2 + 偿债能力	60	0.6520	0.2351	0.8226	0.3842	2.0418
M_3 + 支出冲击	84	0.6511	0.2299	0.8619	0.3811	2.0261
M_4 + 资产缓冲	130	0.7077	0.3084	0.7363	0.4312	2.6386
M_5 + 债务惯性	165	0.7571	0.3746	0.7802	0.4615	3.6124
M_6 + 宏观与社会经营	186	0.7557	0.3725	0.7316	0.4577	3.6438

注：“+”表示在前一组特征基础上新增该类信息；不同模型的分类阈值均按验证集 $F2$ 确定。

异质性结果同样支持上述解释。在乡村家庭、低资产家庭、有不健康成员家庭和高医疗支出家庭中，模型的 Recall 或 F_2 相对更高。例如，高医疗支出家庭的 AUC 为 0.7713、F_2 为 0.5125；有不健康成员家庭的 AUC 为 0.7600、F_2 为 0.5077。这些分组并不构成因果机制检验，但提示健康冲击、刚性支出与资产缓冲不足是模型识别债务压力的重要信息维度。

6. 结论与讨论

本文基于 CFPS 2010~2022 年家庭数据，研究了非房产债务压力风险的时间外推预测问题。主要结论如下：第一，使用风险发生前的家庭特征可以对下一期债务压力状态进行一定程度的样本外识别。第二，α 敏感性检验显示，在 0.25~3.00 的候选区间内 F_2 仅在 0.4455~0.4512 之间波动，$\alpha = 2$ 时综合表现较优，说明风险强度加权结论对参数设定具有一定稳健性。第三，与带类别权重的 Logit 相比，动态加权 XGBoost 在 AUC、Precision、F_2 和前 10%风险提升倍数上略有提高，但 Recall 并不占优，表明其增量价值主要体现在非线性排序和高风险集中度，而非所有指标的全面提升。第四，风险评分最高 10%家庭的实际风险发生率约为总体的 3.33 倍，模型具有将风险对象集中到有限筛查名单中的能力；替代标签、滚动时间窗

口和剔除当期直接债务变量的结果亦支持主要结论，资产缓冲和债务惯性是预测信息增益的主要来源。

本文的实际含义在于：家庭债务风险监测不宜仅依赖是否负债或平均债务规模，而可综合收入、资产、刚性支出和历史债务状态形成分层关注名单。模型输出应作为资源配置、金融教育或债务咨询的辅助信息，而不应直接替代人工核验或作为排斥性决策依据。

模型公平性需要在实际部署中单独审视。第 5.2 节的异质性结果显示，模型在乡村、低资产、有不健康成员和高医疗支出家庭中的 Recall 或 F_2 相对更高。这既可能反映这些群体的风险信号更突出，也意味着同一模型和统一阈值在不同群体中可能产生不同的漏识别率与误报率。调查样本代表性、变量测量误差、追踪损耗、缺失模式以及历史金融可得性差异，还可能将既有结构性差异带入模型。因此，不应将某一群体的较高预测分数直接解释为其内在“信用质量更差”，更不能据此形成自动化排斥。若模型进入实际监测场景，应按城乡、收入与资产水平、健康冲击等关键群组定期报告 Recall、Precision、误报率、漏报率和概率校准差异，并设置群体差异预警；当差异持续扩大时，可通过补充代表性样本、数据质量审查、样本再加权或分群校准、阈值复核，以及保留人工复核与申诉路径等方式进行缓解。模型更适合作为辅助筛查和资源优先级安排工具，不应作为拒贷、福利排除等高影响决策的唯一依据。

仍需注意三点局限。其一，债务收入比是债务压力的代理指标，不能等同于真实违约；其二，调查数据存在回忆误差、追踪损耗和变量测量误差，限制了可达到的预测上限；其三，本文虽然报告了群体异质性并提出公平性监测原则，但尚未系统检验不同群体的真阳性率、假阳性率与概率校准差异，也未评估真实干预效果。未来可在获得更高频交易、征信或政策干预数据后，进一步开展概率校准和群体公平性审计，并评估模型在实际风险治理中的增量价值与潜在分配影响。

参考文献

[1] 李波, 朱太辉. 债务杠杆、财务脆弱性与家庭异质性消费行为[J]. 金融研究, 2022(3): 20-40.

[2] 徐佳, 李冠华, 齐天翔. 中国家庭偿债能力: 衡量与影响因素[J]. 金融研究, 2022(11): 98-116.

[3] 姚健, 臧旭恒, 周博文. 社会网络与中国家庭金融脆弱性[J]. 金融研究, 2024(4): 151-168.

[4] 周利, 张浩. 互联网普及如何影响中国家庭债务杠杆率[J]. 南方经济, 2021, 50(3): 1-18.

[5] Funke, M., Li, X. and Zhong, D. (2023) Household Indebtedness, Financial Frictions and the Transmission of Monetary Policy to Consumption: Evidence from China. *Emerging Markets Review*, **55**, Article 100974. https://doi.org/10.1016/j.ememar.2022.100974

[6] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. https://doi.org/10.1145/2939672.2939785

[7] Liu, W., Fan, H., Xia, M. and Xia, M. (2022) A Focal-Aware Cost-Sensitive Boosted Tree for Imbalanced Credit Scoring. *Expert Systems with Applications*, **208**, Article 118158. https://doi.org/10.1016/j.eswa.2022.118158

[8] Grinsztajn, L., Oyallon, E. and Varoquaux, G. (2022) Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data? *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 507-520. https://doi.org/10.52202/068431-0037