

建档立卡贫困户的返贫风险 预测及返贫原因分析 ——基于 2019 年 25 省（区、市） 建档立卡实地监测调研数据

章文光 吴义熔 宫钰¹

【摘要】 基于 2019 年 25 省（区、市）建档立卡实地监测调研的贫困户数据，建立基于逻辑回归方法的贫困风险预测模型，评估返贫风险。运用优势比和交互信息，分析各致贫因素对返贫风险预测的影响，结果表明：预测模型的 ROC 曲线下面积达到 0.706，模型可用于预测返贫的风险。在解决相对贫困问题阶段，应针对返贫风险高的家庭加强贫困风险预测研究；创新发展基于“预警+救助+赋能”的扶贫策略；在实施保障救助的同时，加强致贫因素分析，采取有针对性的扶贫措施，防止返贫发生。

【关键词】 返贫风险 风险预测 脱贫攻坚

【中图分类号】 F323.8 **【文献标识码】** A **【文章编号】** 1003-7543（2020）12-0110-11

在精准扶贫精准脱贫方略指引下，我国脱贫攻坚工作取得了举世瞩目的成就，贫困人口和贫困县数量急剧下降。我国现行标准下农村贫困人口从 2012 年末的 9899 万人减少至 2019 年末的 551 万人，累计减少 9348 万人；贫困发生率从 2012 年的 10.2% 下降至 2019 年末的 0.6%。2020 年初，国务院扶贫办对 2019 年底未摘帽的 52 个贫困县实施挂牌督战，截至 11 月 23 日，52 个县全部脱贫摘帽，这也标志着全国 832 个贫困县全部出列，脱贫攻坚取得重大胜利，为实现第一个百年奋斗目标打下了坚实基础。与此同时，我国发展不平衡不充分的问题仍然突出，脱贫攻坚成果持续巩固拓展仍面临诸多困难和挑战：深度贫困地区兜底保障户的贫困问题仍然突出^[1-3]；绝对贫困问题得到解决后，相对贫困问题仍然长期存在^[4-5]；一些偏远的山区、边区和少数民族聚居地区受自然禀赋、历史因素和经济基础的影响，返贫风险仍然较高^[6-7]；在大规模解决区域性整体贫困的同时，非贫困地区的贫困问题逐渐显现^[1,5]；部分贫困户政策性收入占比过高，产业发展仍处于孵化期，脱贫攻坚政策在乡村振兴框架内如何衔接仍需路线图^[8-9]。

在脱贫攻坚战及成果巩固拓展过程中，防止返贫、降低返贫风险已成为所有挑战中最具现实性和紧迫性的问题之一。早在 2017 年，习近平总书记在参加十二届全国人大五次会议四川代表团审议时就指出“防止返贫和继续攻坚同样重要”。2020 年 3 月，习近平总书记在决战决胜脱贫攻坚座谈会上明确要求“加快建立防止返贫监测和帮扶机制”。当前是上承脱贫攻坚取得决定性成就、下启“十四五”时期巩固拓展脱贫攻坚成果的关键时期，建立返贫风险预测机制是确保脱贫攻坚成功的必然要求，是将事后帮扶变为事前预防、实现贫困人口动态清零的题中之义。我国已通过全国扶贫开发信息系统为全国 9899 万贫困户实现

¹**作者简介：**章文光，北京师范大学政府管理学院院长，教授、博士生导师，北京师范大学农村治理研究中心主任；吴义熔（通信作者），北京师范大学人文和社会科学高等研究院教授、博士生导师，北京师范大学农村治理研究中心研究员；宫钰，北京师范大学政府管理学院/农村治理研究中心博士研究生。

建档立卡，建档立卡指标体系包括户村县的基础信息库、业务管理系统、扶贫项目库等模块，拥有的海量数据为开展量化研究预测返贫风险提供了基础。进一步推进我国扶贫开发事业，一方面，有必要加强对精准扶贫精准脱贫方略的政策研究，为提高精准帮扶措施的靶向性、精确性提供参考和依据；另一方面，有必要拓宽思路，发挥学术研究在量化研究领域的优势，建立返贫风险预测机制，推出新的帮扶方法和策略。

一、相关研究综述

返贫风险预测及其精准帮扶的对象具有多元性。2014 年国务院扶贫办印发的《扶贫开发建档立卡工作方案》中，根据是否享受社会救助措施将贫困户分为特别困难供养户（特困户）、低收入保障贫困户（低保户）和一般贫困户。2015—2016 年在全国范围开展两次农村贫困人口建档立卡“回头看”之后，为保障脱贫稳定性，根据贫困户脱贫后是否继续享受政策分为脱贫不享受政策的稳定脱贫户和脱贫后仍享受各项帮扶政策的不稳定脱贫户。2018 年《中共中央国务院关于打赢脱贫攻坚战三年行动的指导意见》实施后，为减少扶贫的“悬崖效应”，非贫困群体中的低收入群体亦进入政策视野，被称为边缘易致贫户。2020 年受新冠肺炎疫情等影响的收入骤减户和支出骤增户也成为重点监测群体。从政策角度来看，上述群体均为返贫风险监测对象，但实践中为了克服监测覆盖面广、人员构成复杂、信息不对称等难题，各地普遍将特困户和低保户这两类“深度困难”群体^[10]作为监测返贫风险的主要对象。从理论上讲，一般农户或脱贫户重新成为深度困难群体时，意味着返贫风险的快速上升，此时帮扶措施能否及时兜住决定了返贫是否发生。只有当兜底措施和其他帮扶措施均失效时，才会通过“两公示一公告”或其他程序将该户认定为返贫户，即返贫过程通常经历从脱贫户，到深度困难群体，再到认定成为返贫户这三个阶段，因此脱贫后仍然深度困难的群体是返贫风险最高的群体，各地也通常称之为“返贫风险户”。精准扶贫和预防返贫成功的关键在于及时识别返贫风险户，找出致贫原因、摸清贫困特征，因地制宜、因贫施策，因人施策^[4]，阻断返贫发生。

返贫风险预测是防止因兜底政策失灵而返贫的重要措施之一。以特困供养和低保政策为主的社会救助制度是我国反贫困体系和社会保障体系的重要组成部分，在脱贫攻坚中发挥着兜底保障作用。然而，在社会救助特别是农村低保政策的实施过程中，兜底保障存在一些问题，包括瞄准偏差和遗漏^[11-12]、福利依赖^[13]、消极救助^[14]、“贫困陷阱”问题^[15]等。学者们提出了两种克服上述问题的策略：一是在发挥农村低保制度最低需求保障的同时，实施发展型社会救助项目^[16-17]，或以“工作福利”对抗“福利依赖”^[14]，或完善和建立“救助+赋能”并重的救助制度^[18]等；二是加强针对深度贫困或返贫的预测研究，明确最重要的致贫因素，实施有针对性的救助和扶贫措施，织牢兜底保障网，避免陷入返贫状态。

逻辑回归模型是返贫研究中被广泛运用的模型之一。作为返贫群体预测研究的一部分，农村低保制度研究备受学术界关注，包括低保对象的认定^[11-12, 18]、低保标准的制定^[19]及其减贫效果^[20-21]等。在对低保对象认定的研究中，邓大松、王增文基于 10 个省份 33 个县市农村地区的实地调研数据，利用逻辑回归模型，分析了各项指标对是否为农村低保户判定的影响^[18]。李辉利用逻辑回归模型，分析了各项指标对农村低保户或五保户贫困的影响^[22]。由此可见，逻辑回归模型虽已被用于返贫研究领域，但主要用于分析各项贫困指标与返贫之间的关系，研究还不够深入。因此，有必要进一步加强对各项贫困指标与返贫之间关系的研究，利用各个指标预测农户成为返贫风险户的概率。考虑到贫困的多维特性，有必要重点研究由各项贫困指标构建起来的模型整体上的风险预测性能。当前，学术界对于模型整体性能的研究相对有限，汪三贵等研究了逻辑回归模型对于贫困户识别的准确率，但未描述模型的训练和验证方式^[23]。针对我国县域农村贫困预测，冯娅娅等利用人工神经网络，将各自然、社会经济因子作为输入变量，预测了自然致贫指数和社会经济消费指数^[24]。

精准识别深度困难群体的主要致贫因素，优化贫困治理机制，对于切实防止深度困难群体脱贫后返贫具有重要的理论价值和现实意义^[1, 3]。由于在脱贫攻坚实践中的返贫户数量极少，且导致一般户成为深度困难群体并认定为贫困户的“致贫原因”，与导致脱贫户成为返贫风险户并认定为返贫户的“返贫原因”具有逻辑一致性，因而本文使用 2019 年 25 省（区、市）¹建档立卡实地监测调研得到的贫困户数据，建立基于逻辑回归方法的贫困风险评估模型，来预测成为返贫风险户的概率，并分析致贫原因，找出致贫因素，以阻断返贫发生。本文在数据来源上，采用的是 2019 年针对全国 25 省（区、市）建档立卡实地监测抽样调查数据，同时发展了针对返贫风险户的返贫风险预测模型，并探讨了模型的预测性能，为构建基于“预警+救助+赋能”的

扶贫策略奠定了基础。

二、数据来源和研究方法

（一）数据来源

在国务院扶贫开发领导小组办公室统筹领导下，北京师范大学政府管理学院/农村治理研究中心组织 500 余名调研人员在 25 省（区、市）开展了建档立卡实地监测调研。项目团队根据各省（区、市）贫困县总数量和分布情况，按比例随机抽取约 10% 的贫困县，总计 146 个县。其中，新疆、西藏由于地理条件局限，只抽取了 1 个县。针对每个县，项目团队根据贫困发生率、未脱贫人口等因素，抽取了 5 个贫困程度不同的村。针对每个村，项目团队随机抽取了 50 个农户开展实地入户核查，其中建档立卡贫困户 35 户、非贫困户 15 户。最终经过数据清洗后得到 23307 条实地核查过的建档立卡贫困户数据，包括行政村所在地、贫困人口数、生产生活条件、合作社经营、村属性等贫困村信息，以及户主姓名、家庭人口数、劳动力人口数、贫困户属性、致贫原因、扶贫措施、收入、交通情况、住房情况等信息。

（二）变量提取

为提取变量开展返贫风险户的致贫原因分析，我们从调研核查过的资料中选取了调查问卷编号、村属性、家庭人口数、劳动力人口数、致贫原因及贫困户属性等内容。村属性分为四类，包括非贫困村、脱贫村、贫困村和深度贫困村。依家庭人口数，将贫困户划分为四类家庭：1~2 人、3~4 人、5~6 人及大于 6 人的家庭。依劳动力人口数，将贫困户划分为五类家庭：0 人、1~2 人、3~4 人、大于 4 人及未知（劳动力人数）家庭。从致贫原因中，我们提取被记录的致贫因素，包括因丧、因婚、因学、因残、因灾、因病、缺土地、缺技术、缺水、缺资金、交通条件落后、自身发展动力不足和其他原因（共 13 种）。需要说明的是，致贫原因中的“缺劳力”由于和家庭劳动力人口信息重叠且后者信息更丰富，故未提取“缺劳力”。针对每户，我们计算了致贫因子总数目，其值为 1、2 或 3。贫困户属性包括特困供养户、低保贫困户、脱贫户和一般贫困户。总计有 5442 个返贫风险户（特困供养户或低保贫困户）和 17865 个普通贫困户（脱贫户或一般贫困户）。

（三）风险预测模型

我们利用逻辑回归方法建立贫困风险评估模型，用以预测家庭陷入深度困难成为返贫风险户的风险。逻辑回归是一种广义上的线性回归分析模型，常用于预测事件发生的概率和分析某一因素对事件发生的影响，因模型构建机理逻辑性和可解释性强，其被广泛地应用于经济社会活动预测等领域。在本文中，逻辑回归方法的模型如下所示：

$$\ln\left(\frac{p}{1-p}\right)=\alpha+\sum_{i=1}^K\beta_ix_i$$

其中， p 为家庭陷入深度困难成为返贫风险户的概率， α 为模型的截距， β_i 为贫困因子 x_i 的权重， K 为模型中贫困因子总数目。依靠收集到的数据，通过训练逻辑回归模型，可以得到 α 截距值和各个贫困因子的权重。当一个家庭的数据 x 被采集后，依靠截距和权重，可以得到其返贫的概率 p 。另外，通过衡量各贫困因子的权重，可以了解究竟哪些因素是家庭陷入深度困难成为返贫风险户的主要致贫因素。

为了保证模型预测性能的稳定性，我们采用 10-等份交叉验证（10-foldcross-validation）方法将所有贫困户样本随机分为 10 等份。每次选取其中的 9 等份样本用来训练预测模型，然后利用训练得到的模型去计算剩下的 1 等份样本。依次循环 10 遍实验后，得到 23307 个家庭陷入返贫风险户的风险评估值。

依靠风险评估值，我们可生成其受试者工作特征曲线（Receiver Operating Characteristic Curve-ROC 曲线）。在医学领域，ROC 曲线已被广泛用来描述一项医学测试的性能。真阳性率（TruePositiveRate, TPR）被定义为在实际为阳性的人群中，该医学测试判断为阳性的比例；真阴性率（True Negative Rate, TNR）被定义为在实际为阴性的人群中，该医学测试判断为阴性的比例。ROC 曲线用构图法揭示了 TPR 和 TNR 的相互关系，即通过设定多个不同的预测临界值，计算出一系列 TPR 和 TNR，再以 TPR 为纵坐标、（1-TNR）为横坐标绘制曲线。ROC 曲线下面积（Area Under ROC Curve-AUC）越大，该测试的准确性越高。AUC 取值在 0 至 1 之间，AUC 为 1 时，该测试完全正确；若 AUC 为 0，则该测试完全错误。一般认为，AUC 在 0.5 至 0.7 之间表示模型预测的性能较低；0.7 至 0.9 之间表示模型预测的性能中等；0.9 以上表示模型预测的性能较高^[25]。在本文中，AUC 被用来衡量贫困风险预测模型的准确性。

（四）返贫风险户的风险因子分析

在本文中，我们利用 OddsRatio^[26]和 Mutual Information^[27]确定返贫风险户的重要风险因子。对于家庭致贫的诸多因素，有些可能使家庭返贫，但也有些可能使其避免返贫。比如，因灾是家庭致贫的一种因素，由于只是相对短时间的消极影响，再加上政府和社会的及时关注救助，解决因灾致贫的可能性较大，因而这类家庭通常短时间内由返贫风险户恢复为普通户。

Odds Ratio(OR) 在医学领域是病例对照研究中的常用概念，俗称优势比。某种暴露条件下疾病发生的优势是该条件下病例组人数和非病例组人数的比值；非暴露条件下疾病发生的优势是该条件下病例组人数和非病例组人数的比值，两种优势之比即为优势比。优势比等于 1，表示该暴露因素对疾病的发生不起作用；优势比大于 1，表示该因素是危险因素；优势比小于 1，表示该因素是保护因素。优势比可由 Cornfield 的 2×2 表计算得到，也可由常用的回归模型的系数转化得到^[28]。尽管两者计算方法不一致，但其最终结果类似。由于回归模型具备诸多优势，因而得到了广泛应用。在扶贫研究领域，常常利用由回归模型的系数转化得到的优势比来识别重要的致贫因素^[18]。在本文中，优势比被用来判断一种贫困因子是否是导致返贫风险产生的危险因素，可大致衡量其对返贫风险户产生的影响程度。

Mutual Information 称为交互信息，是信息论中一种有效的信息度量，它可以看成一个随机变量中包含的关于另一个随机变量的信息量，或者说是一个随机变量由于已知另一个随机变量而减少的不确定性，不确定性越少，判断或预测越准确。在本文中，交互信息可表示为返贫风险户预测的不确定性由于知道一个贫困因子的信息而减少的数量。因此，交互信息可被用来衡量致贫因素对返贫风险户预测的重要性，交互信息值越大，该致贫因子对预测越重要。对于能取多个值的变量，交互信息可在变量层面指明其对预测的作用，而优势比在赋值层面上评价变量的重要性，因此在本文中我们同时应用这两种方法相互验证来确定重要风险因子。

表 1 家庭人口数的分布状况（单位：户数）

每户人数	1	2	3	4	5	6	7	8	9	10	>10
户数	1893	3919	3954	4818	3931	2902	1202	429	156	55	48

表 2 劳动力人口数的分布状况（单位：户数）

每户劳动力人口数	0	1	2	3	4	5	6	>6	未知
----------	---	---	---	---	---	---	---	----	----

户数	3168	4663	7980	4010	2292	676	178	41	299
----	------	------	------	------	------	-----	-----	----	-----

逻辑回归模型的训练及交互信息的计算均在 Weka(Waikato Environment for Knowledge Analysis) 软件平台上完成, 其是一款免费的、非商业化的, 基于 JAVA 环境下开源的机器学习以及数据挖掘软件, 是现今最完备的数据挖掘工具之一。

三、建模结果与检验

(一) 模型中的样本分布

本文取样涉及 23307 户贫困人口, 共 92299 人。贫困户家庭规模分布如表 1 所示, 其中 4 人户占比最大, 约为 20.67%; 家庭平均规模为每户 3.96 人。劳动力人口数分布如表 2 所示, 其中有 2 个劳动力的家庭占比最高, 约为 34.24%; 排除未知劳动力数量的家庭, 平均每户有 2.02 个劳动力。

(二) 模型整体预测性能

所提取变量的取值分布如表 3 所示。利用这些变量, 建立贫困风险评估模型, 用以预测家庭陷入深度困难成为返贫风险户的概率。该模型的 ROC 曲线如图 1 所示。其 AUC 为 0.706, 说明可以使用该模型预测家庭陷入返贫风险的概率。

(三) 各变量基于优势比对返贫风险户的预测情况

由训练逻辑回归模型得到的优势比如表 3 所示。各变量对家庭陷入深度困难成为返贫风险户预测的影响如下:

其一, 家庭人口数的影响。以 1~2 人家庭为基准, 当家庭人口数多于 2 个而少于等于 6 个时, 该类家庭不易成为返贫风险户。当家庭人口数多于 6 人时, 该类家庭容易成为返贫风险户, 其优势比为 1.228。

其二, 劳动力人口数的影响。若无劳动力, 该家庭容易成为返贫风险户。随着劳动力人口的增加, 几率将逐渐减少。当劳动力人口数多于 4 人时, 其优势比为 0.308, 这类家庭成为返贫风险户的优势约是无劳动力贫困家庭的 1/3, 这类家庭成为返贫风险户的几率很小。

其三, 因残致贫的家庭成为返贫风险户的概率高, 优势比是其他贫困家庭的 1.711 倍。

其四, 缺土地、因病致贫或自身发展动力不足的家庭成为返贫风险户的几率较高, 其优势比在 0.75 以上。

其五, 因缺技术、交通条件落后、因学、因丧或因其他因素而致贫的家庭成为返贫风险户的几率较小, 其优势比在 0.6 至 0.7 之间。其中因丧致贫的优势比, 其 p-value 为 0.457。由于该值大于常规的统计显著界限 0.05, 因而需要收集更多数据来进一步验证此结论。

其六, 因灾或缺水而致贫的家庭不易成为返贫风险户, 其优势比小于 0.5。

其七, 因缺资金而致贫的家庭不易成为返贫风险户, 其优势比为 0.429。缺资金不是致贫根本原因, 而是其他致贫因素导致的结果。近年来, 由于全国整体经济发展水平提高和扶贫措施加强, 贫困家庭可以获得小额信贷支持, 这类家庭陷入深度困难

的几率较低。

表 3 各变量取值的分布和优势比

变量	变量取值	返贫风险户	普通贫困户	Odds Ratio	p-value
家庭人口数	1~2	3888	1924	—	
	3~4	7046	1726	0.816	<0.01
	5~6	5488	1345	0.898	0.053
	多于6人	1443	447	1.228	<0.01
劳动力人口数	9	1858	1310	—	
	1~2	9651	2992	0.549	<0.01
	3~4	5390	912	0.316	<0.01
	多于4人	748	147	0.308	<0.01
	未知	218	81	0.616	<0.01
因残	是	1737	1151	1.711	<0.01
	否	16128	4291	—	
缺土地	是	620	230	1.031	0.735
	否	17245	5212	—	
因病	是	5419	2139	0.908	0.06
	否	12446	3303	—	
自身发展动力不足	是	1828	503	0.79	<0.01
	否	16037	4939	—	
缺技术	是	8459	2039	0.694	<0.01
	否	9406	3403	—	
交通条件落后	是	1736	467	0.681	<0.01
	否	16129	4975	—	
因学	是	2551	601	0.665	<0.01
	否	15314	4841	—	

因丧	是	11	5	0.648	0.457
	否	17854	5437	—	
其他	是	122	30	0.633	0.034
	否	17743	5412	—	
因灾	是	451	99	0.499	<0.01
	否	17414	5343	—	
缺资金	是	6521	1203	0.429	<0.01
	否	11344	4239	—	
缺水	是	122	23	0.423	<0.01
	否	17743	5419	—	
因婚	是	36	4	0.336	0.043
	否	17829	5438	—	
村属性	非贫困村	3041	1194	—	
	脱贫村	4072	559	0.354	<0.01
	贫困村	10703	3685	1.153	<0.01
	深度贫困村	49	4	0.194	<0.01
致贫因素总数目	1	5263	1492	—	
	2	9970	3043	1.627	<0.01
	3	2632	907	2.758	<0.01

其八，因婚致贫的家庭不易成为返贫风险户，其优势比为 0.336。

其九，村属性的影响。在脱贫村，由于脱贫措施落实，整体贫困状况得以改善，家庭成为返贫风险户的概率较低，其优势比为 0.354。在非贫困村和贫困村，家庭成为返贫风险户的概率较高。在深度贫困村，家庭成为返贫风险户的风险低，一是可能由于样本量分布不够，属于深度贫困村的 53 个贫困户均来自湖南省邵阳县罗城乡保和村；二是可能深度贫困村获得了更多关注度和资源投入。具体原因还有待于进一步的考察研究。

其十，致贫因素总数目的影响。随着致贫因素总数目的增加，家庭成为返贫风险户的概率相应增加。有 2 个致贫因素的贫困户，其陷入深度困难的优势比为 1.627。有 3 个致贫因素的贫困户，其优势比为 2.758，这类家庭更易成为返贫风险户。

（四）各变量基于交互信息对返贫风险户的预测情况

利用交互信息对变量排序的结果如表 4 所示。对于返贫风险户的预测，最重要的变量是劳动力人口数，其次是村属性。因残、缺资金及家庭人口数也会严重影响家庭返贫风险预测结果。

四、结论与政策建议

基于 2019 年 25 省（区、市）建档立卡实地监测调研得到的贫困户数据，通过建模开展返贫风险预测及返贫原因分析，我们发现：基于逻辑回归的风险预测模型，其预测性能达到中等程度，可用于家庭成为返贫风险户的预警，在相对贫困治理阶段发挥应有作用；优势比和交互信息均可用来分析导致家庭陷入深度困难成为返贫风险户的因素，这为进一步开展此类研究提供了新思路。针对返贫风险户需要关注重要致贫因素，有的放矢加强相应保障力度，以提升防止返贫的成效。具体研究结论包括：

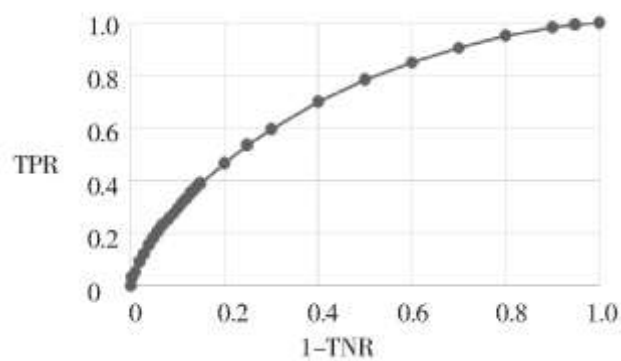


图 1 返贫风险预测模型的 ROC 曲线

第一，家庭人口数是影响家庭成为返贫风险户的重要因素。当家庭人口数多于 2 个而少于等于 6 个时，该类家庭通常家庭人员结构合理，不易成为返贫风险户。当家庭人口数太少或太多时则容易陷入深度困难成为返贫风险户。为此，需要在预测中对不同类型进行区分。调研发现，家庭人口数太少的贫困户主要包括自然因素和突发因素两种情形，自然因素如未婚育夫妻、独居老人等对社会保障政策较为依赖，需要持续稳定的社保兜底；突发因素如因灾祸导致家庭成员亡故、家庭突发重大变故等，其对社会保障、临时救助的依赖性较强。而家庭人口数太多的贫困户体现为超生严重，其对义务教育、住房保障等帮扶措施的稳定性较为依赖。

第二，劳动力人口数是影响家庭成为返贫风险户的重要因素。目前全国扶贫开发信息系统中对劳动力的数据采集包括全劳力、半劳力、部分劳力和无劳力四种类型，数据在每年动态管理中进行更新并定期关联民政和公安部门数据进行比对。若家庭无劳动力或缺劳动力，则容易成为返贫风险户。随着劳动力人口的增加，几率将逐渐减少，这是由于在农村地区较多的劳动力意味着可以获得稳定持续的家庭收入，尤其是随着移风易俗、智志双扶工作的推进，劳动力生产效率不断提升。而缺劳动力或无劳动力家庭帮扶措施通常包括民政一揽子兜底保障措施和部分社会救济，其对帮扶针对性和政策稳定性较为依赖。

表 4 各变量的交互信息值及重要性排序

交互信息值	重要性排序	变量名
0.0260108	1	劳动力人口数

0.0149508	2	村属性
0.0139923	3	因残
0.0127047	4	缺资金
0.0125976	5	家庭人口数
0.0051454	6	缺技术
0.0046409	7	因病
0.0012021	8	因学
0.0004999	9	致贫因素总数目
0.0002953	10	因灾
0.0002031	11	缺土地
0.0001986	12	交通条件落后
0.0001542	13	缺水
0.000148	14	因婚
0.0001427	15	自身发展动力不足
0.0000359	16	其他
0.0000161	17	因丧

第三，村属性对返贫风险预测有较大影响。除去数据本身的局限性，在实践中发现脱贫村农户成为返贫风险户的风险低，而在非贫困村和贫困村，农户成为风险贫困户的风险较高。村属性会产生影响首先印证了在脱贫攻坚中提出解决区域性整体贫困问题是必要和科学的；其次，村属性作为预测模型的重要参数是基于建档立卡村信息表中直接关联的数据，相关内容包括村产业发展情况、集体经济情况、公共设施水平、经济禀赋情况等，是对经济发展水平的整体性描述。2019 年显示的脱贫村多为贫困程度相对较低、脱贫难度相对较小的村，在同等政策力度帮扶下脱贫成效相对明显。而未脱贫村由于贫困程度深、脱贫难度大，非贫困村因长期未享受到脱贫攻坚政策红利，返贫风险均较大，需要在脱贫攻坚任务结束后从城乡统筹视角下实现政策有序退出和补齐民生保障短板。

第四，因残致贫家庭成为返贫风险户的概率较高，缺资金、因婚、因灾、缺水致贫家庭成为返贫风险户的概率较低。有残疾成员的家庭尤其是家庭主要劳动力残疾的，不仅自身无法工作还需家庭成员看护，再加上治疗费用等多种原因共同作用，增加了该类家庭成为返贫风险户的概率。风险预测模型结论告诉我们，在相对贫困阶段，帮扶措施需关注并对不同残疾类型开展分类分层的“精准滴灌”，提高政策资源的针对性。缺资金致贫主要是具备产业发展所需的人力资本，但缺乏产业发展资金和物料，在以提升“造血”能力为目标的帮扶措施下基本能够实现稳定脱贫。因灾、缺水致贫家庭成为返贫风险户更容易受到国家和社会关注，能够通过获得资助而避免成为返贫风险户。因婚致贫多是受地区风俗影响超越家庭承受能力举办婚事，但长期来看对家庭增收也会产生助力。

为助力相对贫困阶段预防返贫和巩固脱贫攻坚成果，我们发展了基于逻辑回归的风险预测模型，用来预测家庭成为返贫风险户的概率，并基于“预警+救助+赋能”的帮扶策略，提出对返贫风险较高家庭应采取有重点的帮扶措施，做到早发现、早干预、早预防，变事后帮扶为事前预防。我们将优势比和交互信息结合起来，识别了返贫风险户的重要风险因素。在此基础上，这里提出如下政策建议：

第一，推进返贫风险监测机制的模型化智能化。随着人工智能技术的飞跃发展，各类定量研究方法不断涌现，给诸多应用领域的发展带来了突破和创新。一些学者将人工智能模型引入扶贫研究，但大多数工作偏向于依靠模型的系数来研究单个致贫因子的重要性，较少探讨各致贫因子影响的模型预测功能。我们发展了基于逻辑回归的风险预测模型，用来预测家庭成为返贫风险户的概率。利用 10-等份交叉验证的方式训练模型，得到模型的 AUC 值达到 0.706。风险预测模型以实地监测调研得到的贫困户致贫因素为输入变量，得到中等程度的深度困难预测性能，表明了建立模型预测家庭陷入深度困难成为返贫风险户的可行性。在大数据和人工智能快速发展背景下，通过多种算法模型对返贫风险开展监测成为可能，本文分析所选取的指标只占全国扶贫开发信息系统全部指标的小部分，系统在与民政、公安、住建、教育等部门进行数据比对过程中也会从行业部门引入部分信息，基于大数据通过不断迭代的神经网络模型、决策树等智能算法可实现返贫风险预测的智能化。

第二，构建“预警+救助+赋能”帮扶机制。当前，对于陷入深度困难状态的社会群体所提倡的扶贫措施，可以概括为救助和赋能并重的模式，一方面提供以最低生活保障、残疾人帮扶、特困供养等为主的社会救助，另一方面通过各项扶持政策帮助贫困户发展扶贫产业、提高就业技能，通过赋能增加其收入水平。运用风险预测模型的预测结果可加强救助和赋能政策执行的针对性和有效性，节约政策资源集中于返贫风险群体，形成基于“预警+救助+赋能”新的帮扶机制。全面建成小康社会后需要快速补齐建立预警监测机制并与现有帮扶机制实现对接，在监测对象认定、监测群体分类、监测程序制定、监测机制建立等方面持续完善。在中央统筹、省负总责、市县抓落实的工作架构下，探索中央预警监测机制与地方监测机制的协调互动，将完善的预警信息纳入地方决策系统，从根本上推动扶贫开发实现从事后反馈向事前监管转变。

第三，在政策过渡期内积极探索人员分类管理和政策有序退出。习近平总书记在 2020 年 3 月决战决胜脱贫攻坚座谈会上明确要求，“要保持现有帮扶政策总体稳定，扶上马送一程，可以考虑设个过渡期”。过渡期设置是落实“四个不摘”要求、实现从脱贫攻坚向乡村振兴转变的重要过程，其中一个重要内容是扶贫政策如何有序退出问题，而返贫风险预测对解决这一问题至关重要。本文的贫困风险预测模型能针对每一个贫困家庭计算出其成为返贫风险户的概率，通过采用合适的概率切割点^[23]，模型可以用来辅助识别风险群体，并按照概率大小划分风险程度，实现返贫风险户的分类管理。在实地调研中，大部分地区认为过渡期以 2~3 年为宜，深度贫困地区则倾向于设置 3 年以上的过渡期以实现脱贫稳定，需要运用一套评价指标对政策退出年限因地制宜进行设置。风险预测模型在当前针对家庭进行预测的基础上值得引入更多关联规则算法，构建起区域性返贫风险预测模型，为政策制定和完善提供有力支撑。

第四，在城乡统筹框架下完善帮扶措施工具包。在 2017 年 12 月召开的中央农村工作会议上，习近平总书记明确指出，“2020 年全面建成小康社会之后，我们将消除绝对贫困，但相对贫困仍将长期存在。现在针对绝对贫困的脱贫攻坚举措要逐步调整为针对相对贫困的日常性帮扶措施，并纳入乡村振兴战略架构下统筹安排。”党的十九届四中全会会议公报明确提出“建立解决相对贫困的长效机制”。相对贫困治理问题是后扶贫时代扶贫开发的主线，在坚持农业农村优先发展总方针指导下，相对贫困治理离不开城乡融合发展的时代背景。本文所采用的优势比和交互信息分析方法可以用来精准识别导致家庭陷入深度困难的重要风险因素，从而有利于制定完善的帮扶措施工具包以开展针对性帮扶，巩固和提升脱贫攻坚工作质量，但从解决相对贫困的角度来看，还需要在城乡统筹框架下丰富政策帮扶工具包内容。例如，在农村，因残导致家庭成为返贫风险户的概率较高，这类家庭在城市也会存在，因而要统筹考量相关帮扶政策，因地制宜采取针对性措施；又如，缺资金的家庭不容易成为返贫风险户，是在贫困群体可以获得小额信贷支持下得出的结论。为了解决融资难的问题，城乡融合过程中需要创新开发多样化的金融帮扶产品。

参考文献：

-
- [1]李春根, 陈文美, 邹亚东. 深度贫困地区的深度贫困: 致贫机理与治理路径[J]. 山东社会科学, 2019(4):69-73.
- [2]左停, 徐加玉, 李卓. 摆脱贫困之“困”: 深度贫困地区基本公共服务减贫路径[J]. 南京农业大学学报(社会科学版), 2018(2):35-44.
- [3]李俊杰, 耿新. 民族地区深度贫困现状及治理路径研究——以“三区三州”为例[J]. 民族研究, 2018(1):47-57.
- [4]燕继荣. 反贫困与国家治理——中国“脱贫攻坚”的创新意义[J]. 管理世界, 2020(4):209-220.
- [5]万兰芳, 向德平. 中国减贫的范式演变与未来走向: 从发展主义到福利治理[J]. 河海大学学报(哲学社会科学版), 2018(2):32-38.
- [6]贾林瑞, 刘彦随, 刘继来, 等. 中国集中连片特困地区贫困户致贫原因诊断及帮扶需求分析[J]. 人文地理, 2018(1):85-93.
- [7]殷丽梅, 杨紫锐. 公共价值导向的精准扶贫绩效评价研究——基于恩施州少数民族贫困地区的实证[J]. 重庆工商大学学报(社会科学版), 2018(4):45-51.
- [8]祖俊涛, 董黎明. “后脱贫攻坚时代”背景下农村低保对象精准识别的实现困境和路径优化——以安徽省 S 县为例[J]. 北京化工大学学报(社会科学版), 2019(4):11-21.
- [9]张松彪, 曾世宏, 袁旭宏. 精准扶贫视阈下城乡居民低保资源配置差异及瞄准效果比较分析——基于 CHIP2013 数据的实证[J]. 农村经济, 2017(12):37-43.
- [10]张明皓, 豆书龙. 深度贫困的再生产逻辑及综合性治理[J]. 中国行政管理, 2018(4):44-50.
- [11]易红梅, 张林秀. 农村最低生活保障政策在实施过程中的瞄准分析[J]. 中国人口·资源与环境, 2011(6):67-73.
- [12]朱梦冰, 李实. 精准扶贫重在精准识别贫困人口——农村低保政策的瞄准效果分析[J]. 中国社会科学, 2017(9):90-112.
- [13]肖云, 吴国举, 刘慧. 农村最低生活保障退出机制构建研究[J]. 西北人口, 2009(4):43-47.
- [14]徐丽敏. 反福利依赖视角下的低保制度审视——兼析《社会救助暂行办法》[J]. 华东理工大学学报(社会科学版), 2014(6):109-114.
- [15]刘丽娟. 我国城乡低保家庭基本状况分析——基于 2016 年“中国城乡困难家庭社会政策支持系统建设项目”的调查[J]. 中国民政, 2017(21):49-51.
- [16]左停, 贺莉, 赵梦媛. 脱贫攻坚战略中低保兜底保障问题研究[J]. 南京农业大学学报(社会科学版), 2017(4):28-36.
- [17]白永秀, 刘盼. 全面建成小康社会后我国城乡反贫困的特点、难点与重点[J]. 改革, 2019(5):29-37.
- [18]邓大松, 王增文. “硬制度”与“软环境”下的农村低保对象的识别[J]. 中国人口科学, 2008(5):8-25.

-
- [19]童星,王增文.农村低保标准及其配套政策研究[J].天津社会科学,2010(2):49-51.
- [20]陈文琼,刘建平.论农村低保救助扩大化及其执行困境[J].中国行政管理,2017(2):85-90.
- [21]韩华为,高琴.中国农村低保制度的保护效果研究——来自中国家庭追踪调查(CFPS)的经验证据[J].公共管理学报,2017(2):81-96.
- [22]李辉.基于 Logistic 模型的深度贫困地区贫困人口致贫因素分析[J].西北民族研究,2018(4):51-58.
- [23]汪三贵,王姮,王萍萍.中国农村贫困家庭的识别[J].农业技术经济,2007(1):20-31.
- [24]冯娅娅,潘竟虎,杨亮洁.中国县域农村贫困的空间模拟分析[J].地球信息科学学报,2018(3):321-331.
- [25]SWETS J A.Measuring the accuracy of diagnostic systems[J].Science,1988,240(4857):1285-1293.
- [26]CORNFIELD J.A method of estimating comparative rates from clinical data:applications to cancer of the lung,breast,and cervix[J].Journal of Chronic Diseases,1951,32(6):139-154.
- [27]SHANNON C E.The mathematical theory of communication[J].Bell Labs Technical Journal,1950,3(9):31-32.
- [28]HAILPERN S M,VISINTAINER P F.Odds ratios and logistic regression:further examples of their use and interpretation[J].The Stata Journal,2003,3(3):213-225.

注释:

1 不包括北京、天津、上海、江苏、浙江、广东 6 个省市。