

基于网络爬虫的农村金融与农民收入关系分析研究

刘朝旸

(河北经贸大学, 河北省石家庄市, 050061)

摘要: 乡村振兴战略的深入实施过程里, 农村金融慢慢变成推动农民收入增长的重要力量, 它的关键性越来越明显。过去十年间, 我国农村金融体系与乡村产业都获得了较快发展, 但金融服务水平怎样继续提高, 仍然是一个核心问题。^[1]传统研究在收集农村金融与农民收入相关数据时, 经常遇到数据更新不够快、覆盖区域有限的情形, 借助网络爬虫方法, 可以高效拿到来自不同渠道的农村金融信息及农民收入材料, 然后使用描述统计、关联性检验以及回归模型分析等办法, 探索农村金融与农民收入之间的内在关系。研究结果提到, 农村金融的深化有利于农民收入的提升, 同时农村金融体系内部表现出显著的协同发展趋势, 也提到了现有研究的不足之处, 给出了相关建议, 这样能为乡村振兴工作带来数据支撑与政策借鉴。

关键词: 网络爬虫; 农村金融; 农民收入; 相关性分析; 回归分析

中图分类号: F8 **文献标识码:** A

1. 引言

1.1 研究背景

2023 年《关于金融支持全面推进乡村振兴加快建设农业强国的指导意见》进一步强调金融支持对于推进乡村振兴的重要作用。这为以推动农村金融发展为路径, 促进农民创业就业增收、增进农民福祉、提高农民生活品质, 指明了方向。^[2]我国对农村经济发展给予高度重视, 农村金融作为农村经济的关键部分, 其发展程度直接关系到农民收入能否增加, 不过目前农村金融市场里存在着金融资源分配不够均衡、服务覆盖范围有限等情况, 这些因素限制了农民收入的提高。与此同时, 传统的数据收集方法不容易实时、全面地拿到农村金融与农民收入之间的动态数据, 这会影响到研究的及时性和准确度, 现阶段, 我国农村金融体系在财政政策方面还存在不少问题。积极发展农村金融不仅能够帮助积累资金, 也能够改善资金配置的局面, 从而有效应对农村资金短缺这一困难。^[3]

1.2 研究目的与意义

这项研究的目标是借助网络爬虫手段, 来探索农村金融和农民收入之间的关联。研究从描述统计、相关关系、弹性系数以及线性回归四个角度分别展开, 呈现分析得到的结果, 从而为乡村振兴战略给出理论层面的支撑。

在理论层面, 把网络爬虫技术带进农村金融与农民收入关系的研究中, 能够拓宽农村经济领域数据收集的途径, 充实了相关研究方法体系, 也为后来学者进行类似研究给出理论上的借鉴。

实践意义是可以及时、全面、充分地多渠道、灵活获取农村企业的各类信息, 也可以快速、便捷、科学地处理农村企业的各类信息, 为金融机构的决策提供有力的信息支撑, 从而有效降低了农村产业金融服务的信息成本。^[4]

1.3 研究解决的问题及方法

先梳理农村金融和农民收入涉及的理论以及网络爬虫技术的基本原理,再设计一套借助网络爬虫获取农村金融与农民收入数据的方案,并收集相关数据,采用统计分析等方法,探索农村金融与农民收入之间的关联及其作用机制,夯实基础,弥补不足,攻克难点,提升质量与效率,想方设法促进农业效益增加、农民收入提高、农村活力增强,推进乡村全面振兴和农业强国建设。
[5]

依托于对国内外的先行文献的深入挖掘和对农村金融与农民收入的关系的相关研究成果的充分梳理,结合网络爬虫技术的最新的应用现状对其所带来的机遇和挑战的深入的剖析为我们未来的研究奠定了坚实的理论基础。借助对政府部门官网、金融机构的平台等多个渠道爬取核心数据,将得到的数据进行描述性分析、相关性分析等,从而获取农村金融与农民的收入之间的关系。

2. 数据与方法

2.1 数据采集

2.1.1 数据来源

本研究利用网络爬虫技术获取的数据包括三大维度,取自于国家统计局年度统计公报的农村居民人均可支配收入(元)的数据,来源于国家金融监督管理总局等官方平台报告或通稿的农业保险覆盖农户数(万户)的数据,提取于中国人民银行《中国金融稳定报告》、中国银行业协会《农村中小银行发展报告》、国家金融监督管理总局《银行业金融机构法人名录》的农村商业银行数量(家)的数据。

2.1.2 核心指标定义

本研究采集了农村居民人均可支配收入(元)、农业保险覆盖农户数(万户)、农村商业银行数量(家)三方面维度的数据信息。为之后进行回归分析,本文把农村居民人均可支配收入(元)作为被解释变量 y ,用来反映农民收入增长的水平;解释变量 x_1 为农业保险覆盖农户数(万户),以反映农村保险服务的普惠程度;解释变量 x_2 是农村商业银行数量(家),以反映农村商业银行的服务网络密度。

2.1.3 数据集概况

最终构建的数据集包含 2019-2024 年共 6 条有效记录,核心指标说明如表 1 所示。

表 1 核心指标说明

指标名称	单位	数据范围
农村居民人均可支配收入	元	16021-23277
农业保险覆盖农户数	万户	18498-23985
农村商业银行数量	家	1478-1723

数据集无缺失值、异常值与重复记录，时间序列连续，指标口径一致，满足后续统计分析与建模需求。数据集统计概况如表 2 所示。

表 2 数据集统计概况

指标	观测值	均值	标准差	最小值	最大值	极差
农村居民人均可支配收入 (元)	6	19524.67	2732.99	16021.00	23277.00	7256.00
农业保险覆盖农户数(万户)	6	21008.17	2124.65	18498.00	23985.00	5487.00
农村商业银行数量(家)	6	1610.33	91.57	1478.00	1723.00	245.00

数据集统计概况关键代码如下：

```
stats_summary = df[core_fields[1:]].describe().round(2)
print("[核心指标统计摘要]")
print(stats_summary.to_string())
```

2.2 研究方法

2.2.1 描述性分析

描述性分析是数据分析的基础，包括均值、标准差、最大值、最小值、极差、年均增长率等统计指标，描述三大主要指标的数量特征及在时间上的变化趋势，可以为之后的数据分析提供依据。其中，年均增长率的计算公式为：

$$\text{年均增长率} = \left(\frac{\text{期末值}}{\text{期初值}} \right)^{\frac{1}{n-1}} - 1 \quad (\text{公式 1})$$

年均增长率 = $\left(\frac{\text{期末值}}{\text{期初值}} \right)^{\frac{1}{n-1}} - 1$ (公式 1 中, n 为数据年份数 (本研究 n=6))

2.2.2 相关性分析

本文采用 Pearson 相关系数法，依据标准化数据计算三大指标间的相关系数矩阵，判断指标间的关联强度。相关系数的取值范围为 [-1,1]， $|r| \geq 0.8$ 为强相关， $0.5 \leq |r| < 0.8$ 为中等相关， $0.3 \leq |r| < 0.5$ 为弱相关， $|r| < 0.3$ 为无明显相关。相关性分析能够明确农业保险覆盖、农村商业银行数量与农民收入之间的关联特征。

2.2.3 量化影响分析

本研究采用弹性系数法，计算金融指标对收入的弹性系数，即金融指标每增长 1% 对应的收入增长百分比，反映指标间的相对变化关系。弹性系数大于 1，表明金融指标对收入增长具有“放大效应”；弹性系数等于 1，表明二者同步增长；弹性系数小于 1，表明金融指标对收入增长的推动作用相对有限。

线性回归分析以标准化后的农业保险覆盖数 x_1 、农村商业银行数量 x_2 为自变量，农村居民人均可支配收入 y 为因变量，构建多元线性回归模型 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ (β_0 为截距项， β_1 和 β_2 回归系数， ε 为随机扰动项)，量化金融指标对收入的贡献度，通过 R^2 值检验

验模型拟合效果。其中，回归系数 β_1 和 β_2 反映自变量每增长1个标准化单位，因变量的增长幅度，即金融指标对收入的标准化贡献度；决定系数 R^2 反映模型的拟合效果， R^2 越接近1，说明自变量对因变量的解释力越强；显著性检验（t检验、F检验）检验回归系数与模型的统计显著性，p值小于0.05表明结果具有统计意义。

3. 数据预处理

为确保数据分析的准确性与可靠性，本研究基于“按需处理，最小干预”原则，对采集的原始数据进行系统性预处理，具体流程与结果如下。

3.1 数据类型转换

原始数据采集后，存在数据类型不一致的问题，部分指标为字符串型（含逗号、单位等冗余字符），无法直接用于统计计算。因此，首先进行数据类型转换，使用Python pandas库的`dtypes`方法检查原始数据类型，结果显示为年份（year）：object（字符串型）；农村居民人均可支配收入（farmer_income）：object（字符串型，含逗号分隔符，如“16,021”）；农业保险覆盖农户数（agri_ins_coverage）：object（字符串型）；农村商业银行数量（rural_bank_num）：object（字符串型）。

为确保年份为整数，核心字段为数值类型，相关代码如下：

```
df = pd.DataFrame(integrated_data)
# 确保年份为整数，核心字段为数值类型
df[self.primary_key] = df[self.primary_key].astype(int)
for field in self.core_numeric_fields:
    df[field] = pd.to_numeric(df[field], errors="coerce") # 转换为数值类型，无效值设为
    NaN
logger.info(f"加载成功，数据量: {len(df)}条，年份范围: {df[self.primary_key].min()}-
{df[self.primary_key].max()}")
logger.info(f"核心字段数据类型: {df[self.core_numeric_fields].dtypes.to_dict()}")
return df
```

数据类型转换完成后，再次使用`dtypes`方法验证数据类型，结果显示`year: int64; farmer_income: float64; agri_ins_coverage: float64; rural_bank_num: int64`。数据类型转换成功，所有指标均为数值型，可直接用于后续统计分析与建模。

3.2 重复值处理

重复值会导致数据冗余，影响分析结果的准确性，因此本研究需要进行重复值检测与处理。结果显示数据无重复值，无需进行重复值处理。相关代码如下：

```
def handle_duplicates(self, df):
    """处理重复值：按年份（主键）去重，保留最新爬取的记录"""
    logger.info("\n开始处理重复值")
    initial_count = len(df)
    # 按年份去重，保留最后一条（假设最后一条是最新爬取的）
    df_cleaned = df.drop_duplicates(subset=[self.primary_key],
    keep="last").reset_index(drop=True)
    duplicate_count = initial_count - len(df_cleaned)
    if duplicate_count > 0:
```

```

    logger.info(f" 发现{duplicate_count}条重复记录 (按年份去重)，已保留最新记录")
else:
    logger.info(f"无重复值，无需处理")
return df_cleaned

```

3.3 缺失值处理

缺失值会导致样本量减少，影响分析结果的可靠性。统计结果显示为 farmer_income: 0, agri_ins_coverage: 0, rural_bank_num: 0，即三大维度数据均无缺失值。相关代码如下：

```

def handle_missing_values(self, df):
    """处理缺失值：仅校验核心数值字段，缺失则删除整行（时间序列数据不插值）"""
    logger.info("\n 开始处理缺失值")
    initial_count = len(df)
    # 统计核心字段的缺失情况
    missing_stats = df[self.core_numeric_fields].isnull().sum()
    logger.info(f"缺失值统计: \n{missing_stats.to_string()}")
    # 删除核心字段有缺失的记录（非核心元信息缺失不影响）
    df_cleaned = df.dropna(subset=self.core_numeric_fields).reset_index(drop=True)
    missing_count = initial_count - len(df_cleaned)
    if missing_count > 0:
        logger.info(f"删除{missing_count}条核心字段缺失的记录")
    else:
        logger.info(f"核心字段无缺失值，无需处理")
    return df_cleaned

```

3.4 异常值处理

异常数据点指的是那些和整体数据分布明显不同的数值，这些点可能因为录入错误或者统计标准变动等原因出现，它们会对统计分析和回归模型带来显著影响。本篇研究是依据我国农村经济发展现实来开展异常值识别与应对工作，每个指标都给出了恰当区间（比如农村居民人均可支配收入设在 5000 到 50000 元之间），经过检查，全部数据都落在设定的合理区间里，并且没有发现异常值，因此不需要做删除或调整操作。相关代码如下：

```

# 筛选异常值
for field, (min_val, max_val) in outlier_thresholds.items():
    outliers = df_cleaned[(df_cleaned[field] < min_val) | (df_cleaned[field] > max_val)]
    if not outliers.empty:
        logger.warning(f"{field} 发现 {len(outliers)} 条异常值 (合理范围: {min_val}-{max_val}) : ")
        logger.warning(f"异常年份: {outliers[self.primary_key].tolist()}，异常值: {outliers[field].tolist()}")
        # 删除异常值
        df_cleaned = df_cleaned[(df_cleaned[field] >= min_val) & (df_cleaned[field] <= max_val)]
df_cleaned = df_cleaned.reset_index(drop=True)
outlier_count = initial_count - len(df_cleaned)
if outlier_count > 0:

```

```

    logger.info(f"共删除 {outlier_count} 条异常记录")
else:
    logger.info(f"无异常值, 无需处理")
return df_cleaned

```

3.5 标准化处理

因为三大数据的量纲不同（收入单位为“元”、保险覆盖数为“万户”、商行数量为“家”），本文为了保证相关性分析和回归分析不受影响，所以需要将原始数据进行标准化处理。本研究采用 Z-score 标准化方法，计算各指标的均值（ μ ）与标准差（ σ ），然后把生成新的标准化字段命名为原字段名+“_norm”。经过标准化后的数据，均值为 0，标准差为 1，这样一来就消除了量纲差异，方便用于相关性分析与线性回归建模。标准化后各指标的统计特征如表 3 所示。

表 3 标准化指标统计特征

标准化指标	均值	标准差	最小值	最大值
farmer_income_norm	0.000	1.000	-1.282	1.373
agri_ins_coverage_norm	0.000	1.000	-1.181	1.401
rural_bank_num_norm	0.000	1.000	-1.445	1.230

标准化处理关键代码如下：

```

def standardize_data(self, df):
    """标准化处理: Z-score 标准化 (均值=0, 标准差=1), 新增标准化字段 (保留原始字段) """
    logger.info("\n开始标准化处理 (Z-score) ")
    df_standardized = df.copy()
    for field in self.core_numeric_fields:
        mean_val = df_standardized[field].mean()
        std_val = df_standardized[field].std()
        if std_val == 0:
            logger.warning(f"{field} 所有值相同 (标准差=0), 无需标准化")
            df_standardized[f"{field}_norm"] = 0.0
        else:
            # Z-score 公式: 标准化值 = (原始值 - 均值) / 标准差
            df_standardized[f"{field}_norm"] = (df_standardized[field] - mean_val) / std_val
            logger.info(f"{field} 标准化完成: 均值={mean_val:.2f}, 标准差={std_val:.2f}")
    # 显示标准化后的数据预览
    norm_fields = [f'{field}_norm' for field in self.core_numeric_fields]
    logger.info("\n标准化字段预览 (前 3 条): ")
    logger.info(df_standardized[[self.primary_key] + norm_fields].head(3).to_string(index=False))
    return df_standardized

```

4. 结果与分析

4.1 分析结果

4.1.1 描述性统计结果

本研究通过描述性统计分析, 系统地呈现 2019-2024 年三大核心指标的数量特征与时间变化趋势, 结果如表 4 所示。

表 4 核心指标描述性统计结果 (原始数据)

指标	均值	标准差	最小值	最大值	年均增长率 (%)	累计增长率 (%)
农村居民 人均可支配收入 (元)	19524.67	2732.99	16021.00	23277.00	6.23	45.30
农业保险 覆盖农户数 (万户)	21008.17	2124.65	18498.00	23985.00	5.51	30.00
农村商业 银行数量 (家)	1610.33	91.57	1478.00	1723.00	2.71	16.50

乡村振兴战略在 2019 年到 2024 年这个阶段得到积极推动, 农村产业不断成长, 金融支持力度持续加大, 最终让农村居民人均可以支配的收入平均值来到 19524.67 元。这个收入从 2019 年的 16021.00 元起步, 慢慢增加到 2024 年的 23277.00 元, 总共增长金额有 7256.00 元, 每年平均增长率是 6.23%, 成为三个指标里增加速度最快的一个。农民收入的标准差达到 2732.99 元, 农民收入或许会受到农业生产周期变化、市场上下波动、自然灾害等各种因素带来的影响。

农业保险覆盖农户数在 2019 至 2024 年期间平均为 21008.17 万户, 标准差为 2124.65 元, 2019 至 2024 年整个时间段累计增长了 5487.00 万户, 年均增长率为 5.51%, 增长速度仅次于农村居民人均可支配收入。近些年来, 中央一号文件多次提到农业保险, 政府借助保费补贴、扩大投保面积及健全多层次保险体系等一系列政策措施, 促使农业保险覆盖面的不断扩大。

农村商业银行数量在 2019 到 2024 年这段时间里平均有 1610.33 家, 从 2019 年的 1478.00 家增加到 2024 年的 1723.00 家, 总共多出 245.00 家, 每年平均增长幅度是 2.71%, 因为农村商业银行的设立和转型要走完严格的审批程序, 并且一些农村信用社借助合并、重组这些方式升级变成农村商业银行, 所以增长步伐显得比较平缓。这组数据的标准差是 91.57 家, 在三个指标里面属于波动最不明显的那一个, 充分表明农村金融机构布局带有的稳定特征, 数量增长保持逐步向上, 能够为农村给出持续并且平稳的金融服务。

4.1.2 相关性分析结果

根据标准化数据并采用 Pearson 相关系数法, 计算三大指标间的相关系数矩阵, 三大

指标间均呈现强正相关关系, 详细结果如表 5 所示。

表 5 标准化指标相关性矩阵

指标	农村居民人均可支配收入	农业保险覆盖农户数	农村商业银行数量
农村居民人均可支配收入	1.000	0.994	0.992
农业保险覆盖农户数	0.994	1.000	0.981
农村商业银行数量	0.992	0.981	1.000

农业保险覆盖农户数与农村居民人均可支配收入的相关系数为 0.994, 呈极强正相关, 表明农业保险覆盖范围的扩大与农民收入增长高度同步。这一结果符合理论预期, 农业保险通过为农民提供风险保障, 降低自然灾害、市场波动等因素对农业生产的冲击, 减少农民因灾返贫、因灾减收的风险, 稳定农民的经营性收入; 同时, 农业保险的普及有助于农民扩大生产规模、尝试高附加值农产品种植, 进而提升收入水平。

农村商业银行数量与农村居民人均可支配收入的相关系数为 0.992, 同样呈极强正相关, 这说明了农村商业银行数量的增加对农民增收具有显著支撑作用。农村商业银行给予农户农业生产贷款和农村基础设施建设贷款, 为农民生产经营、农村产业发展提供资金支持。

农业保险覆盖农户数与农村商业银行数量为强正相关, 相关系数为 0.981, 反映了农村金融体系内部存在显著的协同发展特征。农村商业银行与农业保险机构实行“银保合作”模式, 农村商业银行利用其网点优势与客户资源, 营销农业保险产品, 农业保险机构为农村商业银行的信贷业务提供风险保障, 降低信贷违约风险。农业保险将促进农村信贷供给, 农村信贷将促进农业保险发展, 二者相互制约, 相互促进, [6]这种协同发展模式既扩大了农村金融服务的覆盖面, 又提升了服务效率, 对农民收入增长起了积极作用。

4.1.3 弹性系数分析

弹性系数代表金融指标每增长 1% 对应的收入增长百分比, 计算结果如表 6 所示。

表 6 金融指标对农民收入的弹性系数

金融指标	弹性系数	收入增长百分比
农业保险覆盖农户数	1.531	153.1%
农村商业银行数量	2.581	258.1%

农业保险覆盖农户数的弹性系数为 1.531, 即农业保险覆盖每增长 1%, 农村居民人均可支配收入增长 153.1%, 弹性系数大于 1, 表明农业保险对收入增长具有“放大效应”。这一结果与 Chen 等的研究结论一致, 农业保险通过风险保障稳定农民收入, 同时激励农民扩大生产规模、投入更多生产要素, 进而实现收入的超额增长。[6]

农村商业银行数量的弹性系数为 2.581, 即农村商业银行数量每增长 1%, 农村居民人均可支配收入增长 258.1%, 弹性系数也大于 1, 表明农村商业银行数量增长对收入增长具有“放大效应”。农村商业银行数量的增长意味着农民获取信贷、存款、理财等金融服务的便利性提升, 信贷可得性增强, 解决农民融资难问题, 进而推动农业生产经营规模扩大与农民收入增长。

4.1.4 线性回归分析

回归结果如图 1 所示。

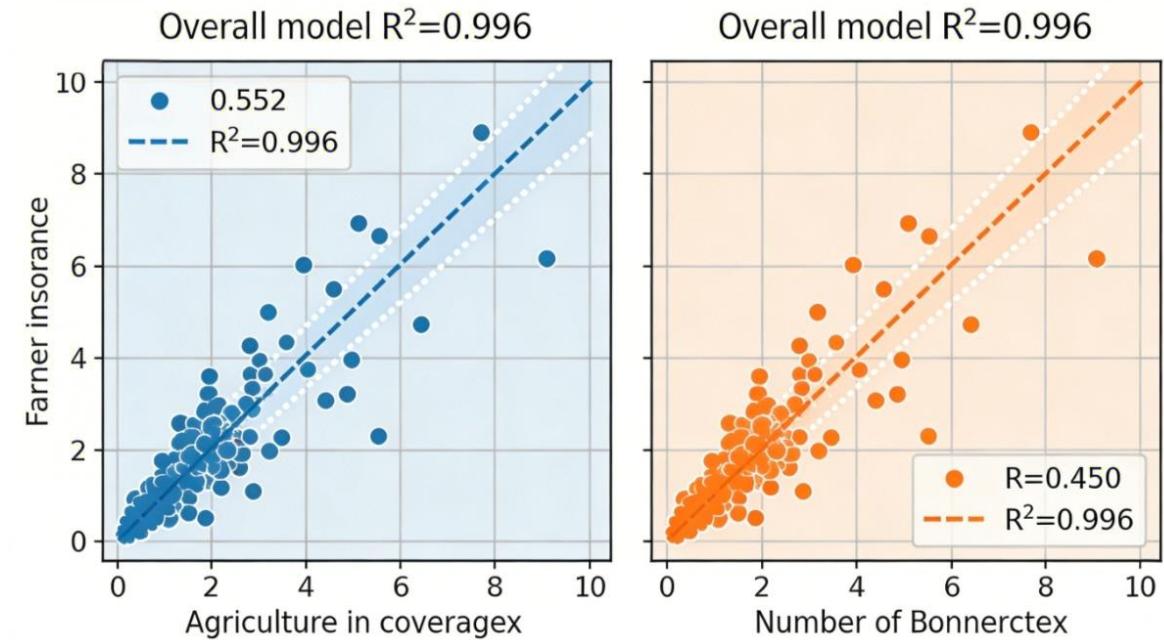


图 1 回归结果

农业保险覆盖标准化值的回归系数为 0.552，表明在控制农村商业银行数量的情况下，农业保险覆盖每增长 1 个标准化单位，农民收入增长 0.552 个标准化单位，即农业保险对收入的标准化贡献度为 0.552。

农村商业银行数量标准化值的回归系数为 0.450，表明在控制农业保险覆盖的情况下，农村商业银行数量每增长 1 个标准化单位，农民收入增长 0.450 个标准化单位，即农村商业银行对收入的标准化贡献度为 0.450。

模型的决定系数 $R^2=0.996$ ，拟合效果极佳，99.6%的农民收入变化可由农业保险覆盖与农村商业银行数量两大金融指标共同解释。

4.2 数据可视化

这项研究进行多角度数据可视化工作，会拿图表样子展示农村金融跟农民收入的变化走向、相互联系还有增长速度比较情形，这么做可以提升研究结论的可读度和说服力。

2019 到 2024 年三个核心指标的年度变化走向在图 2 里显示，左侧的 y 轴代表农村居民人均可支配收入，右侧的 y 轴代表农业保险覆盖农户数目与农村商业银行数目。农村居民人均可支配收入表现出一直往上走的势头并且曲线倾斜度最大，说明农民收入一直增加而且增加速度比较快，农业保险覆盖农户数目同样表现出一直往上走的走向，其增加曲线倾斜度排在收入曲线后面，农村商业银行数目表现出缓慢往上走的走向，增加曲线倾斜度最小，跟“年均增长比例 2.71%”的结论能够对应上。三个指标的同时增加，验证了农村金融进展与农民收入增加之间具备正面关联。

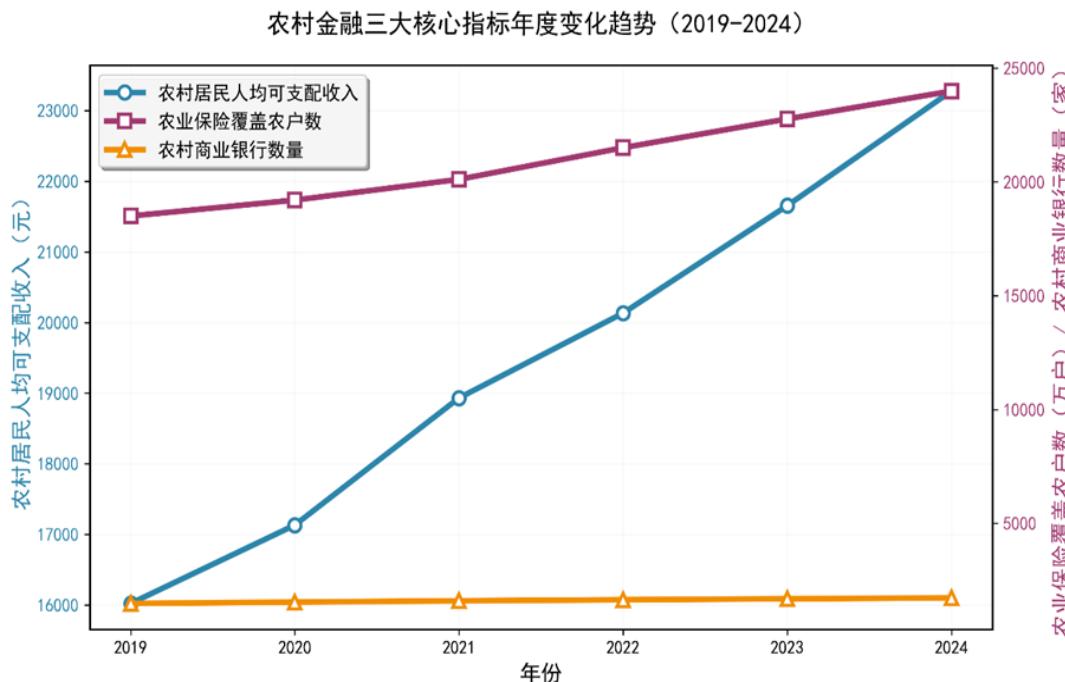


图 2 农村金融三大核心指标年度变化趋势（2019-2024）

图 3 为农村居民人均可支配收入与农业保险覆盖农户数的散点图及线性趋势线，散点为圆形，趋势线为直线，图中标注了 Pearson 相关系数（0.994）。各散点的分布明显是呈线性的，无明显偏离点，这说明了农业保险覆盖与农民收入的线性关联强度极高。趋势线斜率为正，随着农业保险覆盖农户数的增加，农民收入稳定增加。图中标注的相关系数为 0.994，二者具有极强的正相关关系，与相关性分析结果一致，增强了研究结论的说服力。

农村居民人均可支配收入与农业保险覆盖关联分析

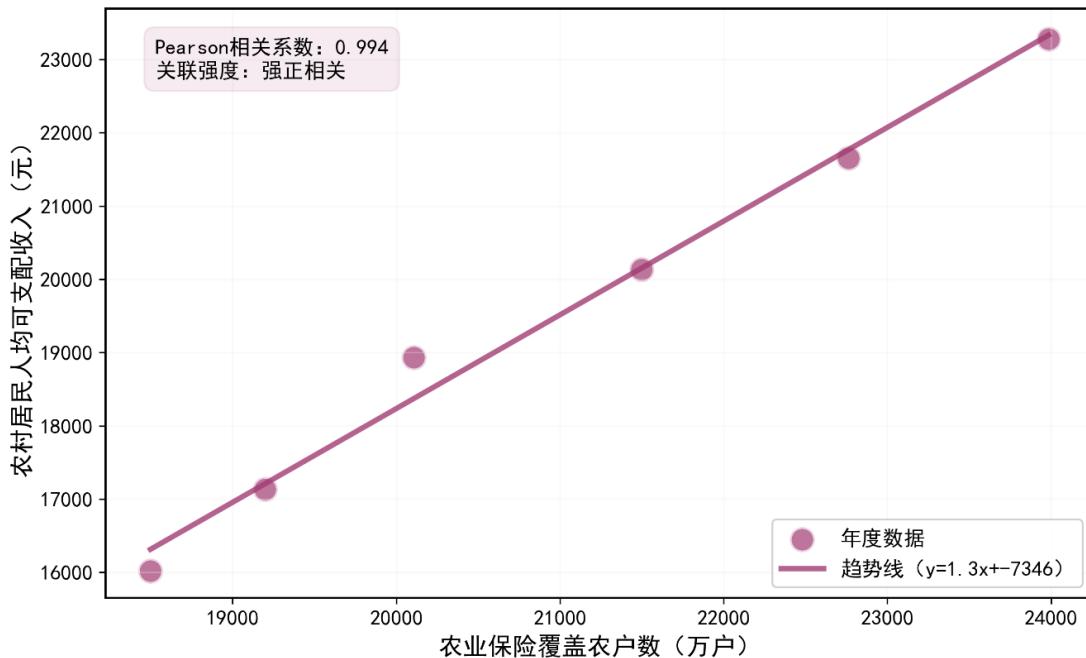


图 3 农村居民人均可支配收入与农业保险覆盖关联分析

农村居民人均可支配收入与农村商业银行数量的散点图及线性趋势线如图 4 所示，散点为圆形，趋势线为直线，图中标注了 Pearson 相关系数（0.992）。各个散点的分布是显

著的正相关，趋势线斜率为正，表明随着农村商业银行数量的增加，农民收入会稳定增加。图中标注的相关系数为 0.992，二者具有极强的正相关关系，与相关性分析结果一致，进一步验证了农业保险覆盖范围对农民增收的支撑作用。

农村居民人均可支配收入与农村商行数量关联分析

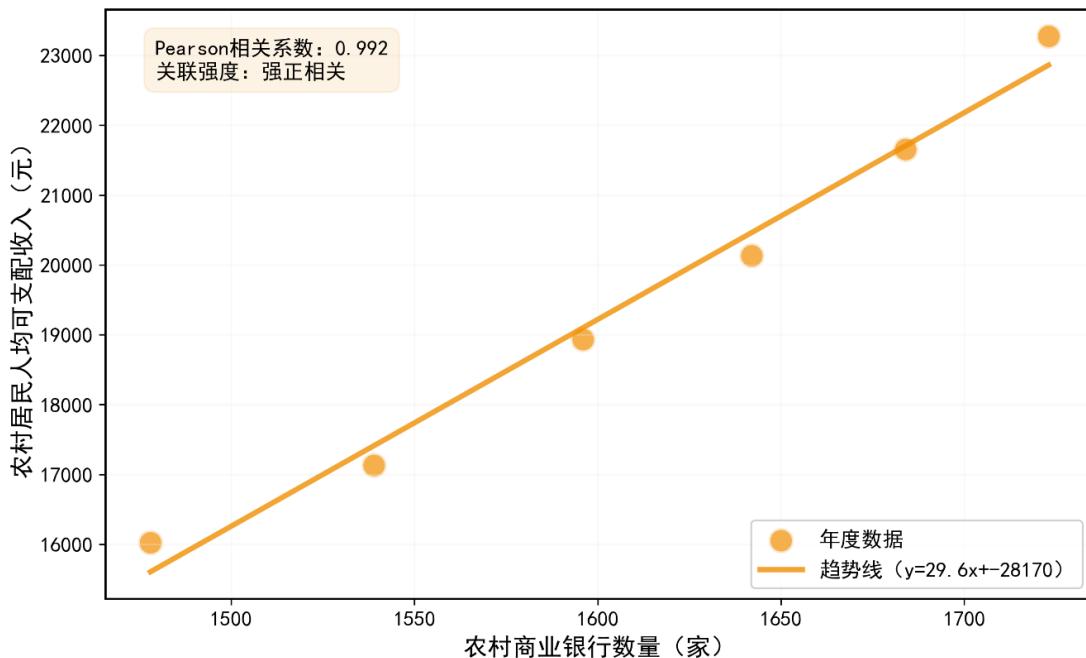


图 4 农村居民人均可支配收入与农村商行数量关联分析

2020 到 2024 年三个关键指标的年度增长情况对比柱状图在图 5 里给出，横轴是年份，纵轴是年度增长率，不同色彩的柱状图分别对应农村居民人均可支配收入、农业保险覆盖农户数、农村商业银行数量各自的增长幅度。农村居民人均可支配收入增长幅度比较大，起伏也比较突出，2021 年农村产业融合进程加快加上农产品价格走高带来这一年增长最快，2023 年受整体经济形势和自然气候灾害的干扰造成 2023 年增长最慢，农业保险覆盖增长整体表现出先涨后跌的走势，2022 年冲到最高点，这和中央一号文件等政策安排存在联系，因为覆盖基数变大、政策拉动作用逐步减弱等原因，2024 年落到 3.6%，从图里能够看出，农村商业银行发展具备平稳性，其数量增长幅度最为平稳，一直保持在 2% 到 4% 这个范围里，2020 年增长最快，2023 年增长最慢。

图 5 里面展示 2020 到 2024 这五年三个关键指标的年度增长情况对比柱状图，横轴表示年份，纵轴给出年度增长率，不同颜色柱状图分别对应农村居民人均可支配收入、农业保险覆盖农户数跟农村商业银行数量各自的增长幅度，农村居民人均可支配收入增长幅度相对较高，波动也较为明显，2021 年农村产业融合进程加快加上农产品价格走高，使得这一年增长速度最快，2023 年因为整体经济形势跟自然气候灾害带来影响，增长最为缓慢。农业保险覆盖增长整体先升后降，2022 年冲到最高点，这和中央一号文件等政策安排有关，由于基数变大和政策拉动效果变弱，2024 年落到 3.6%，从图中可以看出，农村商业银行发展保持平稳性，其数量增长幅度最为稳定，始终处在 2% 到 4% 这个范围里，2020 年增长最快，2023 年增长最慢。

农村金融三大核心指标年度增长率对比（2020-2024）

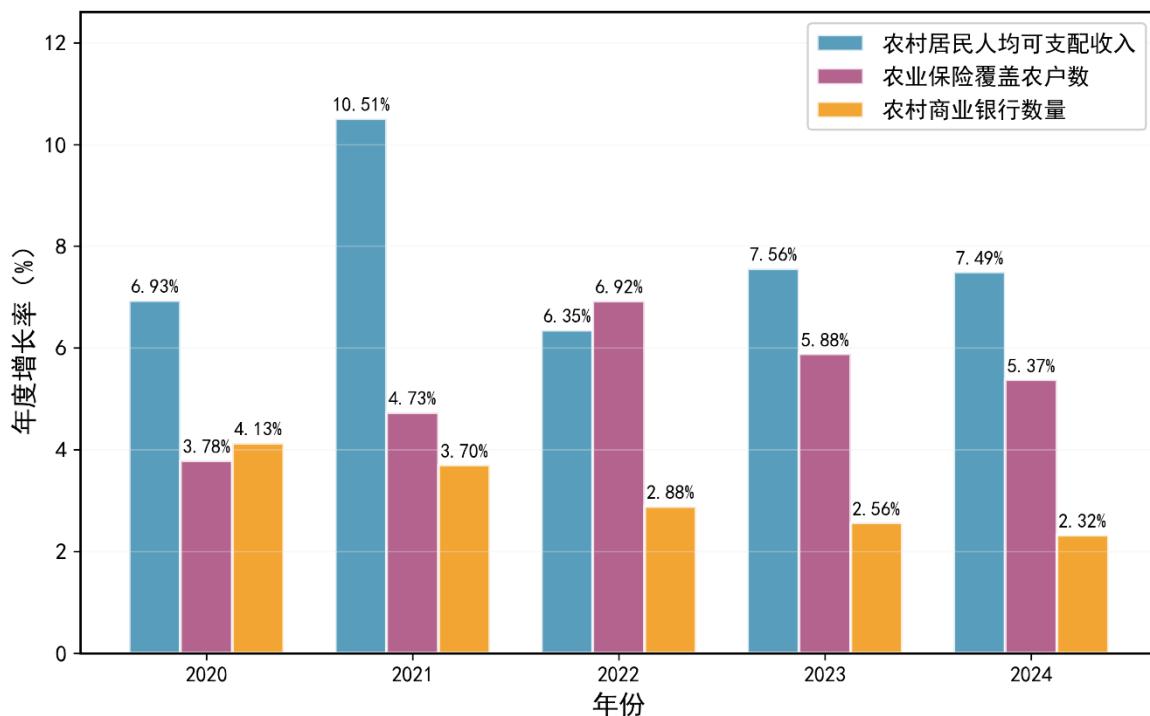


图 5 农村金融三大核心指标年度增长率对比（2020-2024）

图 6 是三大核心指标的相关性热力图，横坐标和纵坐标都是指标名称，右边的颜色条显示了相关系数和颜色的关系。颜色深说明相关系数大，颜色浅说明相关系数小。热力图用颜色深浅显示了三个指标的相关程度，这样读者可以很快看到指标间的关系。图中显示这三个核心指标之间的关系很强。

农村金融三大核心指标相关性热力图（标准化数据）

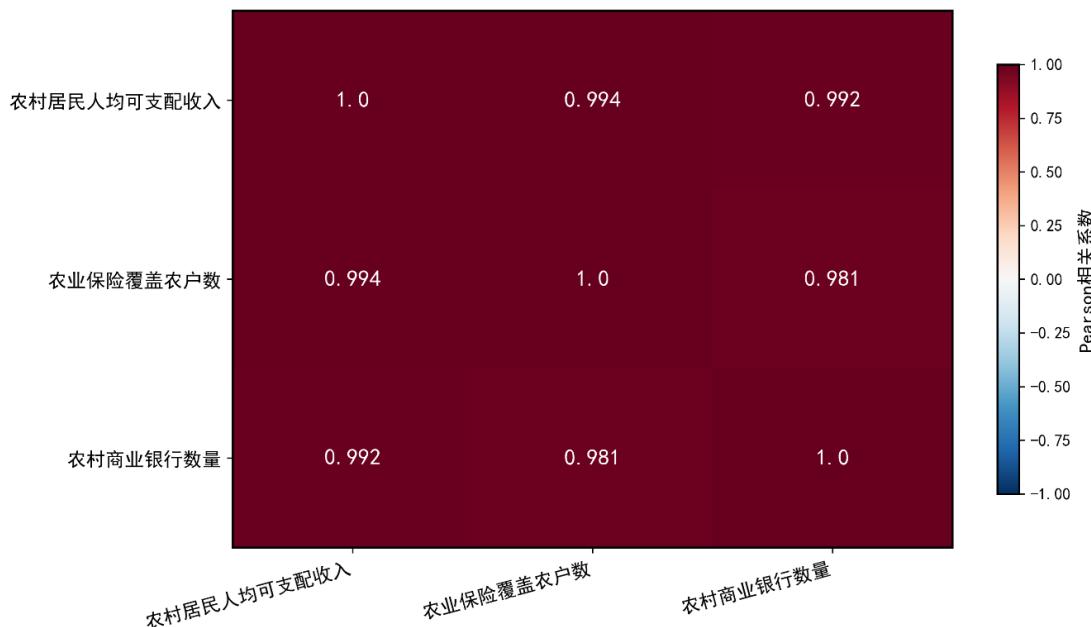


图 6 农村金融三大核心指标相关性热力图（标准化数据）

5. 结论与建议

5.1 结论

网络爬虫技术能高效又全面地进行农村金融与农民收入相关数据的采集工作，解决了传统数据获取方式存在的滞后以及覆盖面狭窄的问题，给农村经济研究提供了新的数据采集途径。农村金融深化依靠农村商业银行服务的优化、金融效率的提高以及农业保险覆盖率的增大，发挥出促进农民增收的效应。农村金融体系内部机构得实现协同发展，才能推动农民收入增长。展开来说，农村金融深化借助农村金融规模的扩大、金融效率的提升以及金融结构的优化，对农民增收起积极作用。^[8]农村地区GDP、农业机械化水平等因素对于农民收入增长也是有关键影响的，农村金融要和农村经济发展、农业现代化建设相协同，才可起到推动农民收入增长的作用。

本研究存在某些不足，农村金融同农民收入之间关系还受到不少外部因素作用，像农村产业结构变化、农业现代化程度、财政支农政策强度、农村人力资本状况等这些方面。农村金融核心内容不光有机构数量跟保险覆盖范围，还牵扯到信贷投放规模、金融服务可获取性这些层面，可是本研究只是选了两类关键指标，对农村金融深化程度衡量不够完整，或许漏掉一些重要作用因素。

5.2 建议

目前农村金融遇到供给结构错位、供给效率不够、供给主体能力不匹配这些现实挑战。^[9]一方面，要持续把粮食主产区的奖补激励制度完善起来，加大对产粮大县的支持力度，同时推行粮食产销区省际横向利益补偿机制，创新区域协同补偿模式，形成主产区供粮，主销区反哺的共赢局面，逐渐改变产粮大县财政薄弱的状况。另一方面，要持续推动农产品内外贸一体化，促进内贸与外贸，进口与出口协调发展，还要发挥期货市场的作用，帮助农业强国建设，借助价格发现、风险对冲、资源配置这些方式，提高农业抗风险能力，稳定农民收入。^[10]

后续研究可以扩展数据采集的时间和地域范围，同时可以参照 Morduch (2019) 在《Journal of Development Economics》里提出的金融包容性与收入不平等的分析框架，进一步探讨农村金融对不同收入群体的影响，这有助于提升研究的理论深度和政策针对性。^[11]需要留意的是，《Journal of Development Economics》近年来发表的中国特稿提到，我国农村存在数字鸿沟代际差异，即 60 岁以上的农户，数字金融使用率不到 20%，这与 Morduch 强调的金融教育配套相呼应。所以，未来研究可以在爬虫数据里增加农户金融素养这个指标，构建“金融工具供给-能力匹配-收入分化”的链式分析，为政策制定给出更精准的理论支撑

参考文献

- [1] 冉光和. 探寻农村金融服务乡村产业振兴的系统路径——《农村金融服务乡村产业振兴的机制与路径研究》书评[J]. 湖南农业大学学报(社会科学版), 2025, 26(06): 2-19.
- [2] 翁可. 农村金融深化与农民增收[D]. 武汉轻工大学, 2025. 000175.
- [3] 卞晓伟, 张玲睿, 王帅. 吉林省农村金融对农民福祉提升的影响机制及支持路径研究[J]. 福祉研究, 2025, 5(00): 72-76.
- [4] 刘永智, 刘霖瑶. 数字金融赋能农村产业高质量发展的路径探究[J]. 农业经济, 2025, (11): 117-119.
- [5] 布仁吉日嘎拉. 农村金融发展对农村经济增长的影响研究[J]. 山西农经, 2025, (18): 220-222.
- [6] 温铁军, 姜柏林. 把合作金融还给农民——重构“服务三农的农村金融体系”的建议 [J]. 农村金融研究, 2007 (1).
- [7] Chen, K., Ye, M. H., & Wang, S. B. (2024). Agricultural Insurance, Rural Credit and Stable Income Growth of Farmers: An Empirical Analysis Based on Simultaneous Equation Model. *Lanzhou Academic Journal*, (1), 124-138.
- [8] 涂爽, 徐玖平, 徐芳. 农村金融发展对农民收入的影响: 基于收入结构的视角 [J]. 农村金融研究, 2022.
- [9] 张海燕, 王小慧. 农村金融供给侧结构性改革路径探析[J]. 村委主任, 2025, (17): 244-246.
- [10] 孔祥智, 孙倩楠. “十四五”时期农民收入变动及“十五五”时期发展趋势[J]. 河北学刊, 2025, 45(05): 1-12.
- [11] Goutte S, Sanin E M. Main challenges regarding development and sustainability in economics and finance[J]. *Development and Sustainability in Economics and Finance*, 2024.

Research on the Relationship Between Rural Finance and Farmers' Income Based on Web Crawlers

Liu Zhaoyang

(Hebei University of Economics and Business, Shijiazhuang City, Hebei Province, 050061)

Abstract: In the process of the in-depth implementation of the rural revitalization strategy, rural finance has gradually become an important force in promoting the growth of farmers' incomes, and its significance has become increasingly evident. Over the past decade, China's rural financial system and rural industries have both developed rapidly, but how to further improve the level of financial services remains a core issue.^[1] Traditional research often encounters problems such as data updates being too slow or coverage being limited when collecting data related to rural finance and farmers' incomes. By utilizing web scraping methods, it is possible to efficiently obtain rural financial information and materials on farmers' incomes from different channels, and then use descriptive statistics, correlation tests, and regression model analysis to explore the intrinsic relationship between rural finance and farmers' incomes. The research findings indicate that the deepening of rural finance is beneficial to the increase of farmers' incomes. At the same time, the rural financial system shows a significant trend of coordinated development internally. The study also points out the shortcomings of existing research and provides relevant recommendations, which can offer data support and policy references for the work of rural revitalization.

Keywords: Web crawler; rural finance; farmers' income; correlation analysis; regression analysis