**说明文档**

**（一）文件说明**

本论文使用的数据代码文件包命名为“2024-00456+数据”，里面包含“2024-00456+程序代码、2024-00456+日志文件、2024-00456+说明文档”3个部分。

（1）2024-00456+程序代码。里面为全部的数据分析代码，其中标记了较为细致的注释，并根据文章正文及附录中相关表格与图形出现的顺序进行分析。本文在开展实证研究时，使用的计量软件为Stata17。

（2）2024-00456+日志文件。按要求提供了如下两部分：一是全部研究过程的stata-log文件（日志文件路径和名称.log），详细记录了全部实证分析过程；二是使用stata命令导出的原始实证结果（未对表格进行美化，保留了其原始状态），以及附录Ⅲ中的图A1和图A2。二者结果是一致的。

（3）“2024-00456+说明文档”，即本说明文档。

**（二）数据来源**

1.卜凯农户微观调查数据。该数据是1929～1933年期间，时任金陵大学农经系主任的卜凯教授（John L. Buck）组织的全国性的农村调查，是微观层面的农户数据。目前该数据的电子版本并未公开发行，由南京农业大学保管。但该套数据资料的一部分已经以纸质版形式由科学出版社出版，即《卜凯农户调查数据汇编（1929～1933）》，全书共13册。目前，相关学者也可以向南京农业大学申请在学校平台上使用。文中对数据来源以及该套资料的珍贵性进行了更为详细地介绍，此处不再赘述。

2.卜凯《中国土地利用（统计资料）》（1937）。该套资料由卜凯教授编著，利用微观农户调查数据以及其他调查资料从县级层面呈现了大量的数据与变量。本文中的县级田税、县级平均利率、市场距离、人口密度变量均从该套资料中整理获取。该套资料的纸质版是公开发行的，电子版也可以获取。

3.其他数据来源。为了获取县级典当数量、政府救济、社会动荡、现代银行四个变量数据，通过人工阅读、识别与整理各地区《地方志》资料获取。文章附录表Ⅰ2、表A1中人均耕地、图A1中收支有余者占比、图A2中借粮农户占比数据来自1930年代的《农情报告》与《全国土地调查报告纲要》。这些资料也都是公开的，可以获取。

**（三）数据处理说明**

本文以借贷市场为研究主题，仅关注在调查当年参与借贷交易的农户。结合研究主题与数据特征，在开展实证分析前对数据进行了如下处理：

第一，在总样本中筛选出借贷农户，提取利率、前五年借贷总额、家庭债务等数据，并依据附录Ⅱ中所述方法识别农户的债务是否上升。

第二，整理借贷农户的水稻、小米、玉米、黍子、高粱等农作物的当年产量与通常产量数据，将各作物的单位统一换算为“公斤/公顷”，并基于卜凯教授提供的折算系数对不同作物的产量进行加总（如球茎类及根菜类的折算系数为0.15，即球茎类及根菜类的公斤数乘以0.15，得到可用于加总的等谷物产量数值，详见《中国土地利用》第371～375页的分析与说明），以比较当年产量与通常产量的差异，作为产量是否向下波动的判断依据。

第三，整合土地财产、家庭规模、劳动力、役畜数量、非农就业等农户微观数据，并根据借贷农户所在县，将县平均利率、田税、市场距离、典当数量、政府救济等县级数据与农户微观数据进行匹配，初步形成数据集。

第四，针对利率、债务、作物产量等核心变量中存在的少量缺失值、标记为unknown（原始资料中即标记为unknown，无法明确其含义）的样本，以及借贷数据可能存在异常或不具代表性的样本（包括两户：一是四川省绵阳第74户，借贷金额为4153.85银元，利率为2.8分，但剔除该户后绵阳地区借贷金额平均值仅为47.07银元，而且借贷金额排第二位的农户仅借款238.46银元；二是湖北省枣阳第91户，借贷金额为6000银元，利率为2.5分，但剔除该户后枣阳地区借贷金额平均值为473.04银元，借贷金额排第二位的农户借款2600银元。这两户由于借贷金额明显较高，将其放在样本中可能不具有代表性，甚至会产生一定的估计偏差），本文予以剔除。

第五，对于其他控制变量中的少量缺失值，为了避免信息损失，本文采用同一县内其他农户的平均值进行填补。

最终筛选出2539位借贷农户作为研究样本。

**（四）Dofile里面使用的指标说明（do文件中也添加了说明）**

1. ‘核心变量设定.dta’中的变量说明：

pinterest：生产借贷利率

cinterest：消费借贷利率

productive：生产借贷金额

cproductive：消费借贷金额

total：借贷总额

winterest：利率（加权月利率）

pfd：产量向下波动

borrow5：前五年借贷金额

debt：总债务

gddebt：深陷债务

povertyt：贫困陷阱

1. ‘实证分析数据.dta’中的变量说明：

povertyt：贫困陷阱（处于贫困陷阱中=1；没有=0）

winterest：利率（加权月利率）

relativec：是否亲属借贷（亲属借贷来源=1；其他借贷来源=0）

pland：人均耕地面积

lnpland：人均耕地面积的对数值：ln（1+人均耕地面积）

size：家庭人口

alabor：农业劳动力

livestock：役畜数量

nonae：非农就业

network：社会网络（婚丧嫁娶等特殊支出费用）

lnnetwork：社会网络对数值：ln（1+婚丧嫁娶等特殊支出费用）

save：储蓄金额

lnsave：储蓄金额的对数值：ln（1+储蓄金额）

rice：水稻种植占比

tax：田税

interestc：县平均利率

pawn：典当数量

location：市场距离

relief：政府救济（有=1；无=0）

flood：水灾次数

drought：旱灾次数

region：八大农区

winterest3：月利率3分阈值

pinterest15：年利率15%阈值

health：重要男性劳动力是否生病（是=1；否=0）

socialins：农民与地主冲突次数

newfina：是否有现代银行（有=1；无=0）

poppre：人口密度

poutput：人均粮食产量

apovertyp：主成分分析方法测度的资产指数

apovertyp\_D：主成分分析方法测度的资产贫困陷阱二元0-1变量

apovertyr：生计回归分析方法测度的资产指数

apovertyr\_D：生计回归分析方法测度的资产贫困陷阱二元0-1变量

1. ‘资产指数测度数据.dta’中的变量说明：

land：家庭总耕地面积

lnland：家庭总耕地面积的对数值：ln（1+家庭总耕地面积）

save：储蓄金额

lnsave：储蓄金额的对数值：ln（1+储蓄金额）

livestock：役畜数量

lnlivestock：役畜数量的对数值：ln（1+役畜数量）

farmh：农舍面积

lnfarmh：农舍面积的对数值：ln（1+农舍面积）

man：健全男子数量（人力资本）

lnman：健全男子数量的对数值：ln（1+健全男子数量）

social：社会资本（人均特殊支出费用）

lnsocial：社会资本的对数值：ln（1+人均特殊支出费用）

apovertyp：主成分分析方法测度的资产贫困指数

lnlandq：耕地面积的平方项

lnsaveq：储蓄的平方项

lnlivestockq：役畜数量的平方项

lnfarmhq：农舍面积的平方项

lnmanq：健全男子数量的平方项

lnsocialq：社会资本的平方项

lnlandsave：lnland与lnsave的交互项

lnlandlive：lnland与lnlivestock的交互项

lnlandfarm：lnland与lnfarmh的交互项

lnlandman：lnland与lnman的交互项

lnlandsoci：lnland与lnsocial的交互项

lnsavelive：lnsave与lnlivestock的交互项

lnsavefarm：lnsave与lnfarmh的交互项

lnsaveman：lnsave与lnman的交互项

lnsavesoci：lnsave与lnsocial的交互项

lnlivefarm：lnlivestock与lnfarmh的交互项

lnliveman：lnlivestock与lnman的交互项

lnlivesoci：lnlivestock与lnsocial的交互项

lnfarmman：lnfarmh与lnman的交互项

lnfarmsoci：lnfarmh与lnsocial的交互项

lnmansoci：lnman与lnsocial的交互项

poutput：人均粮食产量

apovertyr：生计回归方法测度的资产贫困指数

1. ‘省份对比数据.dta’中的变量说明：

povertytp：省级贫困陷阱农户占比

surplusp：省级收支有余农户占比

graincp：省级粮食借贷农户占比

**注：**不同数据集中存在一些相同的变量，但在各个数据集中均说明了变量含义。