

【农村社会发展】

卜凯农户调查数据的整理与利用

陈德仙¹ 胡浩² 陈利根^{2*}

(1. 杭州师范大学 经济学院, 杭州 310000; 2. 南京农业大学 经济管理学院/公共管理学院, 南京 210095)

摘要:近代中国农村微观调查数据是农经史研究的珍贵资料。民国时期金陵大学农经系主任卜凯主持的“中国土地利用调查”极具代表性和影响力。本研究系统介绍该调查资料的抢救性整理过程,并解读数据的利用方向和潜在价值。已还原的卜凯农户调查数据涉及人口与劳动力、家畜、土地、肥料、作物产量及用途、借贷、特殊支出等22个方面,含有22省168地区16786个农场微观样本,能较为全面可靠地反映当时农村的生产和生活状况。卜凯微观及县级数据适用于地权分配、租佃关系、人口压力、家庭手工业、商品化、民间借贷、粮食价格、市场整合等近代中国农经史热点问题研究。卜凯数据的再利用将对经典农业经济理论提供新的证据或修正,也能够为中国式农业现代化总结历史经验和规律提供重要的数据支撑。

关键词:中国土地利用调查;卜凯数据;整理;利用

中图分类号:F329.0 **文献标志码:**A **文章编号:**1671-7465(2024)02-0086-13

一、引言

民国时期政治、经济和社会遽变。复杂的社会现实激荡出汹涌的学术思潮,各学派深入农村展开广泛的调查,收集农村、农业和农民的真实数据资料,进而提出各种理论思想和发展变革办法。如由学界主导的代表性的农村调查,包括毛泽东的湖南农民运动考察和寻乌调查,卜凯的中国农家经济调查和中国土地利用调查,陈翰笙的无锡、保定农村调查,费孝通的江村经济调查,李景汉的定县社会概况调查等^①。这些调查为理解民国时期农业农村发展的现象和规律提供了重要支撑。近年来,学界愈加重视这些珍贵的农村调查数据资料,并相继展开了数据挖掘、抢救性整理和数据库建设等工作。其中,基于卜凯中国土地利用调查整理的卜凯农户调查数据极具特色性、价值性和影响力,并被国内外学界广泛利用,驱动了农业经济理论的发展和农业农村改革路径的探讨。例如,卜凯的技术学派农业发展思想^[1]、黄宗智的农业内卷化理论^{[2]131-143}、马若孟的农民贫困理论^{[3]357-360}、张五常的佃农理论^[4]等。卜凯农户调查数据为研究民国时期的地权分配、租佃关系、农业商品化、民间借贷、农业生产力、农民贫困等重要问题提供了独特的微观数据来源。本文将系统介绍卜凯数据的挖掘整理过程,呈现目前已恢复的数据全貌,并从土地规模的角度评估数据的可靠性,以回应卜凯数据是富农样本的误解^{[2]32}。在此基础上,进一步探讨卜凯数据深度挖掘利用的方向。

收稿日期:2023-08-30

基金项目:国家社会科学基金重大项目“民国时期农村调查资料数据库构建与深度利用研究”(22&ZD233)

作者简介:陈德仙,男,杭州师范大学经济学院讲师;胡浩,男,南京农业大学经济管理学院教授;陈利根(通信作者),男,南京农业大学公共管理学院教授。

① 民国时期学界对农村展开了大量的调查。除了学界主导的农村调查,还包括国民政府部门的各类农村调查、中国共产党的农村调查、日本南满洲铁道株式会社(满铁)的中国农村调查等诸多调查。

二、卜凯的中国土地利用调查:22 省 168 地区 16786 农场

卜凯 (John Lossing Buck, 1890—1975), 毕业于美国康奈尔大学农学院, 1915 年以传教士身份来华, 1920 年进入金陵大学农学院工作, 1921 年组建中国近代第一个农业经济系, 并担任农经系主任。其后, 卜凯为收集民国农村资料及提出针对农业改良的建议, 先后主持了两次大规模的中国农村微观调查, 即 1921—1925 年的中国农家经济调查和 1929—1933 年的中国土地利用调查。其中, 中国土地利用调查共收集了 22 省 168 地区 16786 农场的数据和资料, 这构成了卜凯农户调查数据的核心来源。

(一) 中国土地利用调查问卷和内容

基于现代经济学的研究范式, 卜凯制定了契合中国农村社会场域的区域调查和农户调查问卷, 具体包括田场调查表、食物调查表、县调查表、地区调查表、人口调查表、生命统计调查表、农业概况调查表、作物表、自然区域调查表等, 其中的田场调查表、县调查表和地区调查表的调查内容显得尤为广泛和细致^①。

田场调查表主要针对农场进行微观调查。卜凯调查的农场具有如下特征: 农场由场主本人和其同居共食的家属及雇佣工人管理和耕种; 农场同时拥有耕地等生产用地和农舍等生活用地; 农场收入提供家庭费用占八成以上。田场调查表分为概要、农事、田场杂工、房屋衣服家具及农具四个子调查表。概要调查子表主要调查田场面积、坟地、灌溉、荒地面积、田地分布、租佃法、田场人工、牲畜、饲料之购进、作物、作物牲畜及畜产之用途、副产物、借款与负债、储蓄、特殊费用、作物面积之变迁等内容。农事调查子表主要调查田场作业、土壤情形等内容。田场杂工调查子表主要调查农场内外工作的种类与数量。房屋衣服家具及农具调查子表主要调查房屋、场主的服装、家具、农具等内容。

县调查表主要依据县属和官厅的记录以及有经验的绅商老者的回答进行填写。县调查表分为概要、农产贸易、食粮用途三个子调查表。概要调查子表主要记录县的面积、有生产的荒地面积与有生产的公有水面面积、土地所有权的分类、灾荒、过去特殊事项的影响、全县户口、田地的税额、地图等内容。农产贸易调查子表主要记录了输出农产品的数量及其占本市总贸易百分率的估计、农产贸易的运输费用及输往的市场、输入的农产品等内容。食粮用途调查子表主要记录了不同食粮用于人用食料、饲料、种子和工艺用品等各项用途的情况。

地区调查表是在不同的年份对某县或某县局部区域的若干村庄进行地区性的整体调查。地区调查表分为概要、食物、物价、土壤四个子调查表。概要调查子表主要调查了户口、土地、作物品种、作物荒歉原因、作物输栽法、农民产物销售、霜期、灌溉、肥料、农工工资、副业变迁、新作物引种、乡间谚语、度量衡制度、作物面积变迁、灌溉面积变迁等内容。食物调查子表主要调查了食物习惯、食物消用前的制造、食粮消用于各种状态的比例、食物商店数目的变迁等内容。物价调查子表主要调查了本地农民售出的农产物价、农民购买的物品价格、役畜的价格、长工全年的工资、农民对于耕地所负担的捐税、耕地价格。土壤调查子表主要采样了土壤标本并调查了不同土壤条件下农作物的通常产量。

(二) 田场调查的地区和分布

中国土地利用调查中的田场调查范围几乎覆盖全国, 田场调查的 22 省分别为河北、山

^① 本研究未介绍的其他调查表问卷内容详见《中国土地利用统计资料》附录。

西、绥远^①、辽宁、江苏、浙江、安徽、福建、江西、山东、河南、湖北、湖南、广东、广西、四川、贵州、云南、陕西、甘肃、青海、宁夏,一些边疆省份未做调查。由于当时中国某些地区受到灾害、战争、政治局势以及社会动荡的剧烈影响,出于安全等因素的考虑,调查地区、调查村庄及调查对象的选择均难以做到分层抽样或随机抽样,最终在每个地区若干村庄内选择了 100 个左右的家庭农场进行调查,共计调查了 168 个地区 16786 个农场。各省调查的地区数量、农场数量及其占比具体如下:江苏最多,包含 20 个地区,1905 个农场,农场占比 11.35%;宁夏最少,只有 1 个地区,100 个农场,农场占比 0.60%;其余省份介于二者之间(表 1)。

表 1 田场调查的省份、地区及其农场的分布情况

省份	地区(个)	农场(个)	占比(%)	省份	地区(个)	农场(个)	占比(%)
河北	11	1075	6.40	湖北	6	566	3.36
山西	12	1238	7.38	湖南	9	983	5.86
绥远	2	201	1.20	广东	7	701	4.17
辽宁	1	106	0.63	广西	2	204	1.22
江苏	20	1905	11.35	四川	7	702	4.17
浙江	13	1264	7.53	贵州	5	506	3.01
安徽	11	1088	6.47	云南	7	715	4.30
福建	5	508	3.03	陕西	8	797	4.75
江西	6	606	3.61	甘肃	5	500	2.97
山东	15	1515	9.03	青海	2	218	1.30
河南	13	1288	7.66	宁夏	1	100	0.60

资料来源:根据《中国土地利用统计资料》第 12 章中的表 1“调查材料来源之村别”整理。

(三) 调查资料和成果

中国土地利用调查的地区分布广泛,农场样本量大,调查内容系统翔实,在此基础上形成了具有里程碑意义的调查资料和学术成果。著名农业经济学家乔启明曾评价,中国土地利用调查就人地关系剖析中国土地利用实况是空前的。在中国土地利用调查的基础上,卜凯完成了《中国土地利用》《中国土地利用统计资料》《中国土地利用地图集》三部重要学术著作。《中国土地利用》主要按照不同农业区域进行了数据汇总统计,并对民国时期的自然因素、作物、家畜与土地肥力的保持、田场企业的大小、田场劳动力、物价与赋税、农产运销、人口、食物营养和生活程度展开了深入的分析和解释,成为中国农业经济学学科的第一代教科书。《中国土地利用统计资料》在县和地区层面展示了更精细的统计数据,数据资料结构具体分为气候、土地、食物营养、家畜与土地肥力、物价及税则、作物、田场企业大小、田场劳力、农产运销、生活程度、人口等 11 类 162 种统计数据表格^[5],提供了民国时期丰富的县级一手统计调查数据。《中国土地利用地图集》则对上述内容和项目分别进行了绘图,并特别拍摄了土地利用的航空相片,包括陕西、江苏、湖南、江西等省的梯田、灌溉田、稻田的真实影像图,填补了近代农村视觉影像资料的空白。

三、卜凯农户调查数据的整理过程及成果

(一) 中间统计表的保存、发现及抢救性整理

卜凯中国土地利用调查的相关资料得到了较为妥善的保存,1937—1951 年存于金陵大

① 中华民国国民政府时期的省级行政区,今为内蒙古自治区中部、南部地区。

学,1952—1971 年存于南京农学院,1971—1979 年存于江苏农学院,1979—1984 年存于南京农业大学图书馆,1984—2000 年存于南京农业大学农经系^[6]。2000 年,相关学者发现中国土地利用调查的中间统计表格,并发现大部分原稿保存良好,但部分资料出现残缺或字迹模糊。留存资料共有 118 个包装盒 24956 页。现存的中间统计表^①共计 83 个,其中 22 个是农户数据表,61 个是县级统计表。

对中国土地利用调查资料的抢救性整理工作由此展开。2002—2008 年,南京农业大学经济管理学院与日本东京国际大学展开两轮国际合作,分批次将这些珍贵资料进行电子化扫描及保护。2011 年,卜凯数据电子化和数据库建设项目正式启动,五年内系统地展开了以下四大项工作内容:(1)基础数据录入与核对。主要针对 22 个中间统计表下的农户数据表,将标题、地区、变量、数据等录成电子化数据并进行反复核对。(2)基础数据整合。将分散在不同中间统计表中的农户信息整合到同一标题内容下。(3)数据甄别与修复。对数值不清、反复修改和遗漏、英文单词拼写、中文方言表述等问题进行了推算和推测。(4)农户数据深度核对。核对不同标题下,相同地区、相同编号的农户信息是否存在逻辑对应关系,以及农户数据是否与《中国土地利用统计资料》中的统计数据整体一致。

(二)度量衡转换:面积、产量、距离、货币

民国度量衡纷繁复杂,各地度量标准不统一^[7]。针对度量衡问题,卜凯在《中国土地利用统计资料》中将地方单位转换为统一可比的公制单位,并在第 12 章中提供了一张各地区度量单位的转换率,主要对面积、产量、距离和货币四类数据进行了度量衡转换^[8]。通过查阅《中国土地利用统计资料》以及资料比较,便可以直接获得或推算出度量单位的转换率^②。

(1)面积单位。通过查阅《中国土地利用统计资料》,直接获得不同地区“Mow”(地方亩)与“公顷”的转换率。

(2)产量单位。原始资料中以“C”“O”“T”“P”等地方产量单位来记录作物的产量。“C”和“O”是重量单位,而“T”和“P”是体积单位。通过查阅《中国土地利用统计资料》,可以直接获得“C”与公斤的转换率。另外,通过将原始资料中主要作物的平均单位产量与《中国土地利用统计资料》进行对比,可进一步构建“T”“O”“P”与公斤的转换率公式。以 T 为例,转换公式具体如下:

$$\beta = \frac{X(\text{公担} / \text{公顷}) \times 100(\text{公斤} / \text{公担})}{\alpha(\text{种植面积} / \text{公顷}) \times Y(T / \text{种植面积})} (\text{单位:公斤} / T)$$

(3)距离单位。同样,通过两者资料的地区均值对比,将“Distance of the farthest plots”和“Average distance”等原始资料中记录的单位,推算出与“公里”的转换率。

(4)货币单位。同样,将原始资料中“D”(吊)和“\$”进行对比,推算出与“元”和“银圆”的转换率。

(三)卜凯微观农户调查数据的构建

在中国土地利用调查中间统计表抢救性整理的基础上,构建了一套卜凯微观农户调查数据,该数据主要于 2016 年基本复原。恢复的卜凯数据主要的调查时间是在 1929—1933 年,而江苏江都、宜兴、东台及湖北天门 4 个地区的调查则是在 1924—1925 年完成的。2017—2020 年陆续出版《卜凯农户调查数据汇编(1929—1933)》13 部,分别是《江苏篇》《浙江篇》

① 中间统计表不是每个农户的问卷,而是不同调查项目下的农户调查数据及合计等。卜凯农户调查数据正是基于这些中间统计表格进行的整理。

② 卜凯数据的度量衡转换率详见《卜凯农户调查数据汇编》。不过货币等转换率仍需进一步考证与谨慎使用。

《山东篇》《福建、广东、广西篇》《河北、青海、宁夏篇》《河南篇》《湖南、湖北篇》《江西篇》《山西、甘肃篇》《陕西篇》《云南、贵州篇》《四川篇》《安徽篇》,涵盖的内容及样本量见表 2。

表 2 基于中间统计表的卜凯农户调查数据整理状况

分类	标题及变量	样本量
劳动	1. 农户家庭规模:家庭人口数量	168 个地区、16786 个样本
	2. 健全男子一年内工作情况:工作、休闲及生病的人数、时间及其分布	103 个地区、10473 个样本
	3. 农场家庭劳动力和雇佣劳动力中男工、女工和童工分别从事农场工作和副业的工作时间:工作人数、时间,副业收入占比	105 个地区、10671 个样本
资本(家畜)	4. 各种家畜的数量:牛、马、骡、驴等役畜数量,猪、羊、鸡、鸭等用畜数量	117 个地区、11678 个样本
土地	5. 农场中不同用途的土地面积:农场总面积、耕地、农舍、道路、沟渠、池沼、坟墓、牧场、森林、生产水面、伐柴、可耕荒地等土地的面积	106 个地区、10776 个样本
	6. 农场内坟墓的数量及其所占的面积	99 个地区、10069 个样本
	7. 田地块数与丘数之大小距离及数量:田地、田丘数量,最远田块及所有田块与农舍的距离	66 个地区、6713 个样本
	8. 农场租用面积的百分比:是否属于分租、钱租、谷租等,农场的土地所有数量以及租佃面积	104 个地区、10574 个样本
	9. 不同农作物的耕地面积:春夏秋冬作物在沙地、泥地等以及灌溉和非灌溉条件下的面积	108 个地区、10971 个样本
资本(肥料)	10. 农场肥料生产数量:人粪、厩肥的数量及其总量	102 个地区、10371 个样本
	11. 每地方亩肥料的使用种类和数量:不同作物使用人粪、草灰、厩肥、堆肥、明矾、商品肥等数量	87 个地区、8855 个样本
	12. 农场施用肥料的变迁:施用较多肥料是否合算,施用肥料较十年前是否增加	106 个地区、10769 个样本
	13. 施用肥料种类的变迁:不同施用肥料是否增加,是否减少	70 个地区、7128 个样本
产量	14. 农作物的地方亩产量:不同农作物的当年总产量、通常产量、丰年产量、最高产量	103 个地区、10481 个样本
	15. 各种主要农作物的副产物每地方亩的通常产量	102 个地区、10372 个样本
	16. 按照土壤类型和灌溉地类型分组的主要作物通常产量	82 个地区、8374 个样本
产量分配	17. 各种农作物各项用途的数量:不同农作物用于出售、自用和剩余的数量	144 个地区、14512 个样本
	18. 各种作物各项用途的数量(农副产品):不同农副产品用于出售、自用和剩余的数量	141 个地区、14143 个样本
	19. 各种次要作物各项用途的数量	144 个地区、14512 个样本
资本(资金)	20. 储蓄:农作物和现款储藏的价值,借出的价值、利率、途径	106 个地区、10627 个样本
	21. 借贷和债务:生产和非生产用途的借贷金额、利率,借贷总额,借贷来源,近五年借款总额,负债总额	152 个地区、15318 个样本
	22. 特殊支出:结婚、嫁妆、祝寿、生子、丧葬的支出金额及其总量	145 个地区、14665 个样本

(四) 卜凯数据的可靠性及其检验

卜凯中国土地利用调查是一次具有代表性的民国农村调查,卜凯农户调查数据能够较为真实地反映当时中国农村的总体情况和基本特征,具有良好的可信度。例如,马若孟曾高度评价卜凯的调查及其数据,认为其是十分可信的,并广泛使用了卜凯数据^{[3]24-26}。然而,学界对卜凯数据也存在一种怀疑,认为卜凯中国土地利用调查主要是由学生完成的,当时的大学生主要来自富裕家庭,并据此认为调查的样本也是偏向于富裕样本。黄宗智曾对卜凯的中国土地利用调查提出过怀疑,认为该调查没有按照明确的标准选择农户,并指出卜凯在华北地区调查得出的人均耕地规模比土地委员会的《全国土地调查报告纲要》中的高出约 1 倍,据

此认为卜凯数据是富农样本^{[2]32}。

需要澄清的是,中国土地利用调查并非由金陵大学的学生完成。中国土地利用调查主要是分区调查主任与地方调查员联合完成的,分区调查主任由金陵大学农经系教员担任^①,负责自然区域、地区或县等方面的调查内容,而地方调查员由分区调查主任聘请熟悉当地情况的本县或本村人员构成,并接受指导和培训,进而负责完成田场调查内容^[9]。

本研究在全国层面上,从土地使用面积、耕地使用面积和土地所有面积三个角度,将《卜凯农户调查数据汇编》与《全国土地调查报告纲要》作进一步对比分析(表 3)。

表 3 《卜凯农户调查数据汇编》与《全国土地调查报告纲要》的户均土地规模比较

省份	土地使用面积(亩)			耕地使用面积(亩)			土地所有面积(亩)		
	卜凯数据	土地纲要	差值	卜凯数据	土地纲要	差值	卜凯数据	土地纲要	差值
河北	40.15	20.77	19.38	29.44	22.75	6.69	31.10	21.03	10.07
山西	27.17	38.05	-10.88	24.51	34.52	-10.01	21.72	38.89	-17.17
江苏	25.25	15.20	10.05	23.60	14.33	9.27	15.78	15.42	0.36
浙江	14.37	10.39	3.98	12.44	14.35	-1.91	8.41	9.13	-0.72
安徽	27.26	16.80	10.46	24.14	15.98	8.16	12.99	15.82	-2.83
福建	6.41	9.01	-2.61	6.14	13.08	-6.94	1.95	8.02	-6.07
江西	12.64	10.73	1.91	11.81	19.85	-8.04	5.31	8.54	-3.23
山东	24.30	15.30	9.00	22.29	19.31	2.98	22.14	14.90	7.25
河南	21.47	18.22	3.25	19.65	23.16	-3.51	15.58	16.89	-1.31
湖北	10.93	11.89	-0.96	10.32	12.69	-2.37	6.42	10.97	-4.56
湖南	20.94	14.06	6.88	15.98	13.88	2.10	12.73	11.79	0.94
广东	18.29	5.96	12.33	17.36	8.22	9.14	6.11	1.40	4.71
广西	17.18	15.37	1.81	14.73	13.19	1.54	12.17	14.89	-2.72
四川	18.30			16.36	11.37	4.99	8.81		
贵州	8.59			7.25	39.50	-32.25	6.49		
云南	23.59			16.77	13.88	2.89	14.82		
陕西	31.00	22.71	8.29	26.56	28.51	-1.95	25.42	22.13	3.29
甘肃	24.67			20.92	28.84	-7.92	22.94		
青海	49.97			48.17			44.39		
宁夏	25.52			22.30			25.40		
中国	21.80	15.76	6.04	18.99	18.41	0.58	14.92	15.17	-0.26

资料来源:土地委员会编,《全国土地调查报告纲要》,第 24、27、31 页;
注:土地委员会的调查范围更广,具有明确的抽样原则,样本量更庞大。土地使用面积的数据源于《全国地方调查报告纲要》表 16《各省平均每户经营面积》,其包含了 163 个县、1534920 个农户;耕地使用面积的数据源于表 13《农民每户每人平均摊得耕地亩数》,其包含了 791 个县、32068541 个农户;土地所有面积的数据源于表 20《各省业主平均每户所有面积》,其包含了 163 个县、1295001 个农户。

(1) 在土地使用面积上,卜凯数据在大多数省份的户均结果比《全国土地调查报告纲要》要高,在全国层面高了 6.04 亩。这可能是两项调查包含的土地类型不同导致的^②。(2) 在耕地使用面积上,从总体上看,卜凯在全国调查的户均结果比土地委员会的高 0.58 亩,其

① 分区调查主任包括张心一、张履鸾、乔启明、李明良、刘润涛、潘鸿声、邵德馨、孙文郁、崔毓俊、尉迟秀藻、杨蔚、应廉耕 12 位。
② 卜凯调查的土地类型包括耕地、农舍、道路、沟渠、池沼、坟墓、牧场、森林、生产水面、伐柴、可耕荒地。土地委员会调查的土地类型包括水田、旱田、山林地、池塘地、无收益的荒地、宅地和晒场等。

中,卜凯的结果有一半省份相对较高,而另一半省份相对较低。在华北区域,卜凯在河北和山东的结果确实如黄宗智所讲的较高,但在山西和河南则较低。(3)在土地所有面积上,卜凯的全国户均结果反而比土地委员会的低了0.26亩,并在57.14%的省份上更低。由此可见,对户均耕地使用面积和土地所有面积而言,卜凯的结果与土地委员会的结果在总体上并不存在明显的系统性偏差,因此,在土地规模层面断言卜凯数据是富农样本缺乏充分的证据。此外,钟甫宁等^[10]从粮食产量角度否定了卜凯数据是富农样本。由此,卜凯数据在总体上是基本可靠的,当然也仍待更全面的学术检验,包括从不同地区、不同变量、不同度量等方面的检验。

四、卜凯数据在农经史热点问题中的利用

近代中国农经史热点问题主要包括地权分配、租佃关系、人口压力、家庭手工业、农产品商品化、民间借贷、粮价价格、市场整合、田税负担、灾害冲击等^①。卜凯微观及县级数据涉及的地区广泛、样本量大、信息丰富,并具有可靠性,其适用于上述热点问题的进一步探讨。

(一) 卜凯微观农户调查数据的学术应用领域

1. 利用土地数据分析地权分配及租佃制度

农场租用数据表主要记录了两块信息:所有权、租佃制度(表4)。(1)土地所有权方面,按照完全所有、混合所有(部分所有+部分租赁)、没有所有(全部租赁)三种不同的方式详细记录了土地的所有面积和租赁面积,这表示自耕农、半自耕农、佃农的土地权属结构以及农户的地权面积大小。自耕农占48.80%,户均地权为23.52亩;半自耕农占44.81%,户均地权为7.67亩,户均租佃面积为11.25亩;纯佃农^②占6.39%,户均租佃面积为28.86亩;总体户均地权面积是14.92亩(表4)。土地所有权记录的是农业生产经营者的数据,城镇地主的数据并不包括在内。(2)租佃制度方面,主要涉及半自耕农和佃农的租佃合约类型,分别包括分租、钱租、谷租、工佃分租、其他五种类型。谷租、钱租和分租是主要的租佃合约类型,分别有3668个、1260个、1033个(表4)。租佃制度方面缺乏租金数据,有土地流出的农户也被排除在外。

关于卜凯土地数据,于敏捷和胡浩已使用租佃制度数据,探讨了西北地区的租佃市场、合约特征及其影响因素^[11]。陈苏等同样基于租佃制度数据,使用C-D函数,比较了分成合约与固定合约的土地资源利用效率^[12]。陈苏使用Probit模型,证明了分成合约选择与风险规避有关,以及地主与佃农之间的博弈能力将决定合约形式^[13]。后续研究还可以在土地所有权及租佃制度等数据的基础上,通过描述统计方法、经济学评估指标、计量模型等更多元化的方法,测度近代中国地权分配的不平等程度,并结合灾害、税收、商业化、交通、副业、负债等数据,进一步分析地权分配的影响因素;掌握土地合约类型的结构及其地区差异,并结合肥料、役畜、农具、产出等数据,进一步分析合约类型对土地生产投资和生产效率的影响等相关

① 学界关于这些热点研究问题已经形成了丰硕的研究成果。例如在地权分配方面,包括胡英泽的《近代华北乡村地权分配再研究——基于晋冀鲁三省的分析》,龙登高、何国卿的《土改前夕地权分配的检验与解释》,赵冈的《地权分配的长期趋势》等。在商品化方面,包括黄宗智在《华北的小农经济与社会变迁》中对农业商品化加速的论述,曹幸穗在《旧中国苏南农家经济研究》中对农副产品及生活资料商品化的讨论等。在民间借贷方面,包括李金铮的《民国乡村借贷关系研究》,陈志武等的《高利贷与贫困陷阱:孰因孰果——反思民国时期农村借贷的利率问题》,林展的《社会资本与民国时期的农户借贷——基于1930年代北方农村调查》等。其他方面的研究不再赘述。

② 此处纯佃农的含义是指无任何土地产权(包括房屋)的农业生产者。因此其数值相较于其他调查资料偏低。实际上根据《中国土地利用统计资料》可以发现,卜凯根据不同的调查来源及不同的佃农含义提供了三个佃农占比值,分别为5.90%、16.90%、33.10%。

热点问题。

表 4 土地所有权和租佃制度的样本特征

	变量	样本量	平均值	中位数	标准差	最小值	最大值
土地所有权(亩)	自耕农的地权面积	5159	23.52	14.93	28.65	0.35	502.93
	半自耕农的地权面积	4737	7.67	3.65	12.89	0.02	212.95
	半自耕农的租佃面积	4737	11.25	7.16	13.48	0.16	206.14
	佃农的租佃面积	676	28.86	20.96	27.87	0.84	249.75
	地权面积	10572	14.92	7.97	23.42	0	502.93
租佃制度	是否分租	10572(1033)	0.10	0	0.30	0	1
	是否钱租	10572(1260)	0.12	0	0.32	0	1
	是否谷租	10572(3668)	0.35	0	0.48	0	1

资料来源:《卜凯农户调查数据汇编》中的表 8“农场租用面积的百分比”。
注:此处的样本量是清理后的样本,下同。括号内数字是属于该种租佃合约的样本量。

2. 利用劳动力数据分析人口压力与劳动就业

(1)家庭规模数据表记录了家庭人口数量。家庭人口仅有 1 人的有 129 户,占 0.78%; 2~4 人占 31.06%; 5~10 人的家庭最多,占 56.24%(表 5)。在《中国土地利用统计资料》中还有非常丰富的县级人口数据,包括各地区总户数与农民户数、家庭之大小、户之大小及人口之分配、人口密度、每公顷作物面积平均之人数、人口之变迁等。(2)劳动力分配数据表主要记录了三块信息:家庭副业工作类型、家庭副业工作时间、家庭副业工作收入。家庭副业工作类型种类繁多,不同地区记录的名称差异较大,卜凯的《中国土地利用统计资料》的数据显示,家庭手工业的占比是最高的,为 22.00%。没有家庭副业的占比达到 52.18%,副业平均收入占比为 14.66%(表 5)。

表 5 家庭人口与家庭副业的描述统计

家庭人口分组	频数	百分比	副业比例	频数	百分比
1	129	0.78	0	4460	52.18
[2, 5)	5114	31.06	(0, 25]	2044	23.91
[5, 10)	9258	56.24	(25, 50]	1485	17.37
[10, 20)	1874	11.39	(50, 75]	391	4.57
[20, 60)	88	0.53	(75, 100]	168	1.97
合计	16463	100.00	合计	8548	100.00

资料来源:《卜凯农户调查数据汇编》中的表 1“农户家庭规模”和表 3“农场家庭劳动力和雇佣劳动力中男工、女工和童工分别从事农场工作和副业的工作时间”。

关于卜凯劳动力数据,胡浩和郑微微已使用劳动力分配数据,基于古典经济学估算方法,测算了剩余劳动率并分析了劳动力剩余的原因^[14]。杨中卫使用 2SLS、Mincer 工资方程和分位数回归模型,研究了健康资本对农业劳动参与、副业参与、雇工收入的影响^[15]。后续研究还可以在家庭规模数据与劳动力分配等数据的基础上,通过描述统计方法、经济学评估指标、计量模型等更多样化的方法,评估近代中国家庭的人口压力,测算家庭手工业与副业的参与状况,进一步探讨人口压力增加与非农产业及就业市场发育有限的条件下,农业劳动力内卷化耕作及劳动力配置等相关热点问题;还可以结合生存生活资料、战争、饥荒、灾害等数据,与马尔萨斯陷阱进行对话。

3. 利用作物用途数据分析农产品销售与商品化

作物用途数据表记录了作物总产量、作物出售数量,进而可以得出农户层面以及不同地区

层面不同作物的农产品商品化率。例如,河北阜平的水稻和小麦商品化率分别为 75.49%和 67.49%;江苏昆山的水稻、小麦、油菜籽和棉花商品化率分别是 19.76%、75.62%、11.24%和 49.90%;湖南武冈的水稻、黄豆、油菜籽和棉花的商品化率分别是 17.21%、48.22%、32.03%和 10.83%(表 6)。作物用途数据表(包括作物产量数据)主要列出了当地一些重要作物的产量及销售 量,但并未呈现该地区所有作物的信息,实际上卜凯当时调查到作物类型是非常丰富的,可以进一步结合和参照《中国土地利用统计资料》第 6 章《作物》。

关于作物用途方面的卜凯数据,既有研究主要侧重于利用作物产量数据,探讨土地规模与农业生产效率之间的关系^[16-17]以及种植决策和结构问题^[18]。后续研究还可以在作物出售用途等数据的基础上,通过描述统计方法、经济学评估指标、计量模型等更多样化的方法,测算近代中国农产品商品化的水平,并结合市场及交通、工商资本、土地规模、家庭债务、税收等数据,研究农产品商品化经营的动力机制和商品化经济的市场发育;还可以结合农产品价格和家庭食物消费等数据,分析农产品商品化对农民的收入、生活水平和福利造成的影响等相关热点问题。

表 6 不同地区的农产品商品化率(7 个地区)

地区	样本量	水稻商品化率	小麦商品化率	黄豆商品化率	油菜籽商品化率	棉花商品化率
河北阜平	101	75.49(26)	67.49(100)			
江苏昆山	83	19.76(78)	75.62(78)		11.24(47)	49.90(80)
江西都昌	101	0.29(101)	0(82)	28.37(100)	24.23(54)	0(47)
湖南武冈	104	17.21(102)		48.22(52)	32.03(72)	10.83(38)
广西容县	106	11.45(106)	17.14(56)	20.52(105)		
贵州安顺	101	10.84(101)	92.69(43)	3.63(79)	99.23(37)	
陕西勉县	100	29.20(45)	19.91(96)			77.00(26)

资料来源:《卜凯农户调查数据汇编》中的表 17“各种作物各项用途的数量”。
注:括号内的数字是当地种植该作物的调查农户数。

4. 利用借贷和债务数据分析民间借贷与利率

借贷和债务数据表主要记录了三块信息:借贷金额、借贷利率、债务。(1)借贷金额方面,系统记录了当年的生产借贷和消费借贷的金额及其总额,以及近五年的借款总额。5680 个农户当年有借贷,户均借贷金额为 84.76 银圆。消费借贷比生产借贷更多,4796 个农户有消费借贷,户均消费借贷金额为 75.98 银圆;1696 个农户有生产借贷,户均生产借贷金额为 69.00 银圆(表 7)。(2)借贷利率方面,消费借贷利率比生产借贷利率整体更高,消费借贷年利率为 34.08%,生产借贷年利率为 30.34%,总体来看均处于很高的利率水平^①(表 7)。(3)债务方面,有 6971 个农户处于负债状态,即有 46.10%的农户无法偿清或来不及偿还其借贷,平均负债额达到 120.53 银圆(表 7)。

关于借贷和债务方面的卜凯数据,陈志武等使用相关数据和 2SLS,认为贫困和市场制度方面的缺陷是造成高利贷的原因^[19]。Turvey 和 Hu 采用 3SLS,从生产率的角度分析贷款的原因,以及贷款金额和贷款供给方的垄断特征等对借贷利率的影响^[20]。后续研究还可以在借贷与债务等数据的基础上,通过描述统计方法、经济学评估指标、计量模型等更多元化的方法,剖析近代中国民间借贷的基本运行机制,利率和债务的基本特征,并结合借贷渠道、借贷双方特征、灾害风险、地租、生病、特殊消费、物价等数据,更全面地分析高利贷的成因和家庭

① 当时的民间借贷利率往往很高,例如《全国土地调查报告纲要》也显示了类似情况,农村利率在 20%~40%的占 75.82%。

负债的形成机制等相关热点问题;还可以进一步分析农户为何无法摆脱债务陷阱,以及高利贷对农民生产和生活造成的影响等诸多重要内容。

表 7 借贷和债务数据的样本特征

变量	样本量	平均值	中位数	标准差	最小值	最大值
借贷金额(\$)						
生产借贷金额	1696	69.00	31.60	159.50	2	4153.85
消费借贷金额	4796	75.98	40	120.48	0.03	2000
借贷总额	5680	84.76	48	145.69	0.03	4153.85
近五年借款总额	3904	145.04	80	228.54	0.06	5200
借贷利率						
生产借贷月利率	1663	2.53	2.50	1.22	0	12
消费借贷月利率	4734	2.84	2.50	2.15	0	50
债务总额(\$)	6971	120.53	60	193.38	0.09	5200

资料来源:《卜凯农户调查数据汇编》中的表 21“借贷和债务”。

注: \$ 是中国土地利用调查中间统计表和《卜凯农户调查数据汇编》中使用的货币单位。根据《Land Utilization in China》及其中译本《中国土地利用》的比较,以及《卜凯农户调查数据汇编》与《中国土地利用统计资料》的比较, \$ 应是银圆。

(二) 中国土地利用统计资料成为农经史研究的数据源

《中国土地利用统计资料》中的县级和地区数据可以用于分析近代中国农经史的诸多重要内容。例如:(1)农产品售价数据表详细记录了中国 1906—1933 年 15 省 37 地区的农产品售价指数。该农产品售价指数是综合了米、早稻、晚稻、春小麦、冬小麦、大麦、玉蜀黍、黄豆、花生、油菜籽、棉花等 21 类主要农作物并按照不同权重得出的综合价格指数。农产品售价指数波动极其剧烈,在 1906—1929 年上升了 225.64%,在 1929—1933 年又狂跌 44.09%(图 1)。(2)粮食作物通常产量数据表详细记录了中国 22 省 161 地区壳类、豆类、油籽类、纤维类、球茎及根菜类、其他作物的通常产量信息。当时农业生产效率水平总体是非常低的,例如水稻亩产 223.30 公斤,小麦亩产 67.75 公斤(图 2)。(3)农人向政府缴纳的田税种类与数额数据表详细记录了中国 1907—1933 年 19 省 115 县的单位普通田地的正税、附加税及总税额。农户的田税负担逐渐加重,每亩田税额在 1907—1932 年上升了 64.67%(图 3)。(4)灾荒发生及损害比例数据表详细记录了中国 1904—1929 年 21 省 160 地区的各类自然灾害发生的平均次数、发生的通常时间以及作物受损害的平均比例(图 4)。当时处于灾害群发期,灾害频发,农户受灾严重。

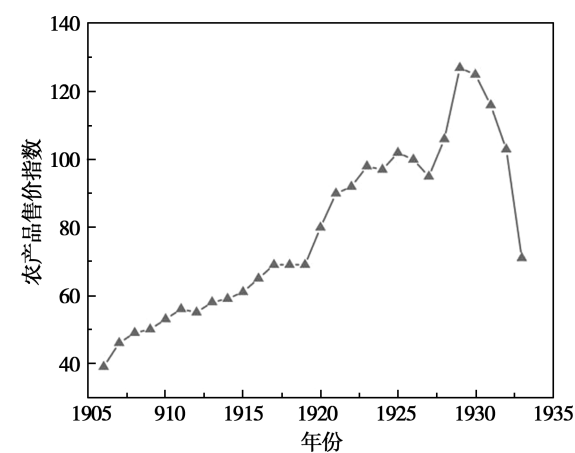


图 1 1906—1933 年中国农产品售价指数 (1926=100)

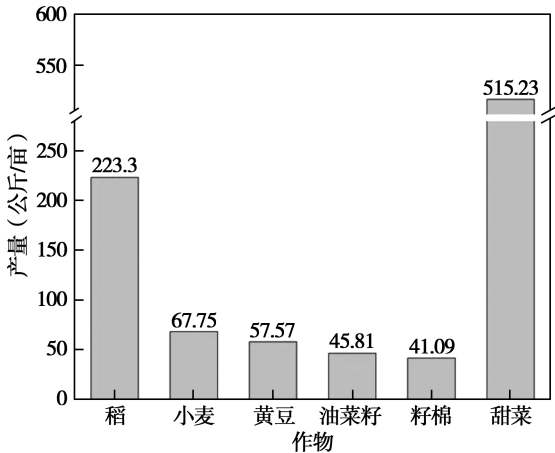


图 2 20 世纪 20—30 年代中国作物的通常产量

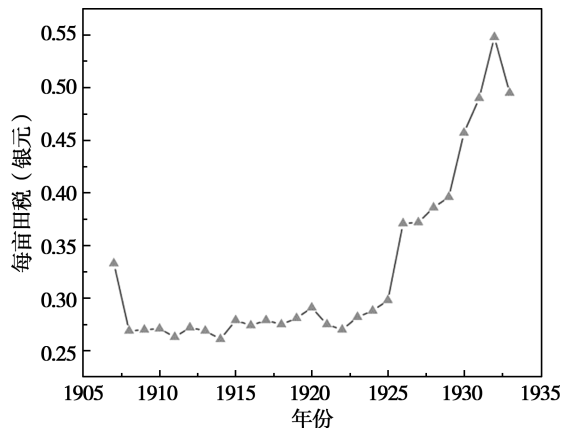


图 3 1907—1933 年中国每亩普通田地的税额

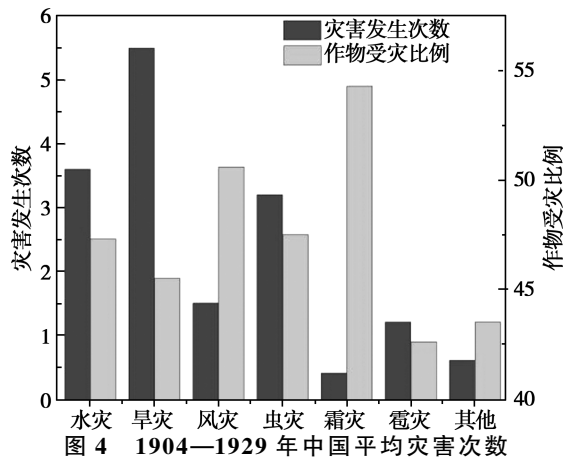


图 4 1904—1929 年中国平均灾害次数与作物受损

资料来源:《中国土地利用统计资料》中的第 5 章物价及税则的表 2“农家所得农产品售价之指数”、第 6 章作物的表 9“各种主要作物每公顷之通常产量”、第 5 章作物的表 8“每公顷普通田地农人向县政府所缴纳之税额”、第 1 章作物的表 9“1904—1929 年各种灾荒发生之次数及其受损害之百分比”。

在《中国土地利用统计资料》相关数据基础上,可以进一步讨论近代中国的粮食价格波动、农业生产力为何落后(地权分配、技术水平、战争与灾害、市场割裂等视角)、农产品市场整合与市场发育(自然地理、交通设施、灾害风险冲击、军阀割据等视角)、田税及灾害冲击对农户生产及生活造成的负担,以及对农村社会动荡的影响(村民逃荒离村、土匪滋生、农民与地主关系紧张等视角)等相关热点问题。

五、结论

卜凯主持的中国土地利用调查是民国时期最具代表性的农村经济调查之一。在中国 20 世纪 20—30 年代极其特殊的历史背景下,基于现代经济学的研究范式,应用现代化的调查与分析方法,对中国农村、农业进行了全景式的展示和剖析,最终形成了《中国土地利用》《中国土地利用统计资料》《中国土地利用地图集》等具有里程碑意义的调查资料和学术成果。卜凯中国土地利用调查资料覆盖地区广,涉及内容翔实,样本数量多,极具挖掘利用价值。

在中国土地利用调查的中间统计表被发现后,经过基础数据录入与核对、基础数据整合、数据甄别与修复、农户数据深度核对、度量衡转换率推算等五项系统化的数据整理工作,卜凯微观农户调查数据实现了电子化,并出版了 13 部《卜凯农户调查数据汇编》。通过《卜凯农户调查数据汇编》与《全国土地调查报告纲要》的土地规模对比可以发现,并不支持卜凯数据是富农样本,卜凯数据在总体上是可靠的。关于不同地区、不同变量等方面的卜凯数据,还有待更全面的检验。

卜凯农户调查数据为我们把握民国农村社会的实际情况和问题提供了独特的数据来源,为验证农业经济经典理论及结合中国实际情况修正经典假说提供了支撑,对近代中国农业经济史热点研究问题具有很高的利用价值。未来可以利用农场租用面积数据、家庭规模数据与劳动力分配数据、作物用途数据、借贷和债务数据、农产品售价指数数据、主要粮食作物通常产量、农人向政府缴纳的田税种类与数额、灾荒发生次数及受损害比例等卜凯微观和县级数据,深度研究地权分配、租佃关系、人口压力、家庭手工业及副业发展、农业商品化、乡村民间借贷、粮价波动与市场整合、田税负担及灾害冲击的影响等近代中国农业经济史热点研究问题。当然,卜凯数据资料的应用并不局限于此。

卜凯数据的整理与利用并联合更多的资料,将有利于学界更真实地探索近代中国的农村社会问题,完善传统农业经济理论和思想,通过纵向比较进而发现中国农村的演变逻辑和规律,乃至为现代农村制度改革、农业经济增长、农业农村现代化等方面提供重要的历史指导和现实借鉴。

致谢:感谢中国农业博物馆曹幸穗研究员、南京农业大学经济管理学院钟甫宁教授、虞祎副教授以及中国经济史学会 2023 年年会各专家学者、审稿人对本文提供的宝贵修改建议。

参考文献:

[1] 卜凯. 中国土地利用[M]. 台北:台湾学生书局,1971:1-20.

[2] 黄宗智. 华北的小农经济与社会变迁[M]. 北京:法律出版社,2014.

[3] 马若孟. 中国农民经济:河北和山东的农民发展(1890—1949)[M]. 南京:江苏人民出版社,2013.

[4] 张五常. 佃农理论[M]. 2 版. 北京:中信出版社,2017:195-199.

[5] 卜凯. 中国土地利用统计资料[M]. 上海:上海商务印书馆,1937:444-463.

[6] Hu H, Suga M, Zheng W W, et al. The Discovery and Restoration of Buck's Original Data[M]//Hu H, Zhong F N, Turvey C G. Chinese Agriculture in the 1930s. Switzerland: Palgrave Macmillan, 2019:85-112.

[7] 陈翰笙,等. 亩的差异:无锡 22 村稻田的 173 种大小不同的亩[M]. 南京:国立中央研究院社会科学研究 所,1929:2.

[8] 胡浩,钟甫宁,周应恒. 卜凯农户调查数据汇编:浙江篇(1929—1933)[M]. 北京:科学出版社,2017: 8-9.

[9] 孙文郁. 中国土地利用调查之经过及进行概况[J]. 地政月刊,1933(3):305-316.

[10] Zhong F N, Hu H, Su Q. Reliability of John Lossing Buck's Land Utilization Survey Data: A Preliminary Test of Grain Yields[M]//Hu H, Zhong F N, Turvey C G. Chinese Agriculture in the 1930s. Switzerland: Palgrave Macmillan, 2019:113-120.

[11] 于敏捷,胡浩. 近代西北地区租佃问题初探——基于卜凯原始农户数据的分析[J]. 中国农史,2016, 35(1):85-93.

[12] 陈苏,胡浩,张利国. 马歇尔分成租佃制真的导致土地资源利用“无效率”——基于 20 世纪 30 年代卜凯 农家调查资料的检验[J]. 农业技术经济,2020(1):17-32.

[13] 陈苏. 产量预期、博弈能力与土地合约选择研究——基于 20 世纪三十年代卜凯农家调查数据[D]. 南 京:南京农业大学,2018:142.

[14] 胡浩,郑微微. 民国中期农业剩余劳动力的估算及区域差异研究——基于卜凯的中国农家调查数据 [J]. 中国农史,2014,33(3):59-66.

[15] 杨中卫. 健康人力资本对农业雇工收入影响研究——基于卜凯调研数据的分析[D]. 南京:南京农业大 学,2016:87-91,94-109.

[16] 胡浩,于敏捷. 中国 20 世纪早期农户耕地面积与土地生产率关系研究——基于卜凯农村社会调查[J]. 中国经济史研究,2015(5):68-77.

[17] Hoken H, Su Q. An Analysis on the Inverse Relationship between Yield and Farm Size in Rural China in the 1930s[M]//Hu H, Zhong F N, Turvey C G. Chinese Agriculture in the 1930s. Switzerland: Palgrave Macmillan, 2019: 171-192.

[18] 郑微微. 家庭粮食安全约束对传统农户种植决策与劳动力利用的影响——基于卜凯的农家调查数据 [D]. 南京:南京农业大学,2015:61-75.

[19] 陈志武,彭凯翔,袁为鹏. 高利贷与贫困陷阱:孰因孰果——反思民国时期农村借贷的利率问题[J]. 量 化历史研究,2017(Z1):1-32.

[20] Turvey C G, Hu H. Farm Credit Demand and Supply in 1930s China[M]//Hu H, Zhong F N, Turvey C G.

Chinese Agriculture in the 1930s. Switzerland: Palgrave Macmillan, 2019: 225–258.

(责任编辑:刘浩)

The Collation and Utilization of Peasant Household Survey Data by Buck

CHEN Dexian HU Hao CHEN Ligen

Abstract: Data on modern Chinese rural areas is limited, making it an invaluable resource for researchers studying the history of modern agricultural economics. The “Land Utilization Survey in China,” conducted by J. B. Buck, the director of the Department of Agricultural Economics at Jinling University during the Republic of China era, was a highly representative and influential study. This paper provides a comprehensive introduction to the process of rescuing and preserving the survey data, as well as an interpretation of its utilization direction and potential value. The restored data from Buck’s peasant household surveys covers 22 aspects, including population and labor force, livestock, land, fertilizer, crop yield and usage, borrowing and special expenses, and encompasses 16,786 farm-level microsamples from 168 regions across 22 provinces. This comprehensive and reliable data set offers a detailed reflection of rural production and living conditions of the time. The microdata and county-level information compiled by Buck are highly relevant to current hot topics in the history of modern Chinese agricultural economics, such as land ownership distribution, tenancy relations, population pressures, the cottage industry, commercialization, private lending, grain prices, and market integration. The reuse of Buck’s data set can provide fresh evidence to challenge or confirm classical agricultural economics theories, and it allows for the summarization of historical experiences and the identification of developmental patterns that can inform China’s ongoing agricultural modernization efforts.

Keywords: Land Utilization Survey in China; Buck Data; Collation; Utilization