

数字化时代的历史数据使用问题

——以中国近代“国民所得”人口数据为例

缪德刚

(中国社会科学院 经济研究所,北京 100836)

[内容提要]随着人文社会科学研究数字化发展,历史上的人口数据受到学界的关注和重视。与当今通过人口普查获取数据不同,直到20世纪三四十年代,中国尚未能开展覆盖全域的人口普查。由于人口数据缺失,中国“国民所得”估算者只得选取来源不一的人口数据。人口数据由此成为中国“国民所得”估算成果被诟病的因素之一,这同时折射出中国近代人口数量问题的复杂性。基于“国民所得”人口数据考察中国近代人口数量问题,既能够增进相关问题的社会历史分析深度,还可以为当前合理运用包括人口数据在内的历史数据提供参考。

[关键词]人口统计;中国近代;“国民所得”;历史数据

DOI:10.13469/j.cnki.zgshjjsyj.2024.02.002

[中图分类号]F092.6/F129

[文献标识码]A

[文章编号]1000-422x(2024)02-0073-10

引言

数字化不仅改变了日常生活方式,也实现了科研技术的突破。一些领域的人文社会科学工作者已经不满足于使用统计数据,而是将视野延展到史料中,去整理、挖掘历史数据,乃至构建大规模数据库开展经验分析。近十几年来,这种非传统的人文社会科学研究态势发展得如此之快,以致有人认为数字人文时代已经到来。^①目前,涉及历史数据整理与研究的成果相对丰富,其中以运用历史数据对相关经济理论予以经验分析的成果居多,^②此外还有历史档案的数字化、^③历史数据的补充与修订^④等方面的著述。从学术发展史上看,此前历史数据之所以未能在学术研究中规模化应用,除了技术原因外,历史数据与当前相关数据在获取途径上的差异也是不容忽视的制约因素。因此,在使用历史数据前,深入了解其收集、整理及早期应用无疑是重要的。

20世纪90年代中后期至今,提取自原始史料的人口数据大范围渗透到历史研究中,先后形成了加拿大巴尔扎克人口数据库、荷兰历史人口样本数据库、瑞典斯堪尼亚经济人口数据库等。中国方面的人口数据库也陆续建立并得到广泛运用。^⑤事实上,中国人口数据资料是十分丰富的,譬如中国近代人口数量长期以来是国内外学界共同关注的话题,该领域也由此形成了大量著述。综括来说,围绕中国近代人口数量问题的现有研究主要分为三类:一是关于中国近代人口数量问题的简

[基金项目]中国社会科学院学科建设“登峰战略”中国经济思想史特殊学科(编号:DF2023TS12)阶段性成果。

[收稿日期]2023-08-08

- ① 黄兴涛:《当代中国历史学的时代使命》,《历史研究》2019年第1期。
- ② 如杨虎涛、冯鹏程《技术—经济范式演进与资本有机构成变动——基于美国1944—2016年历史数据的分析》,《马克思主义研究》2019年第6期。
- ③ 如祁天娇《从历史档案到历史大数据:基于威尼斯时光机十年路径的探索》,《中国图书馆学报》2022年第5期。
- ④ 如许宪春《中国国内生产总值核算历史数据的重大补充和修订》,《经济研究》2021年第4期。
- ⑤ 梁晨、董浩、李中清:《量化数据库与历史研究》,《历史研究》2015年第2期。

要性介绍与评述;^①二是对中国近代史料中人口数据的系统化整理;^②三是利用已有资料对中国近代人口数进行估算。^③此外,也不乏基于中国近代人口数据探讨影响人口变动的因素、^④中国发展道路^⑤等问题的成果。从这些著述不难发现,先期的很多著述用到了中国近代人口数据。

从社会调查学术发展史来看,人口数据在不同历史时期的用途不尽一致。在该方面,中国人口数据也不例外。近代以前,对人口数量的掌握多是旨在赋税征敛。近代以后,人口数据的用途日趋多元,并与其他形式的社会调查数据合算为反映社会经济状况的各种指标,“国民所得”即为指标之一。中国“国民所得”估值主要是 20 世纪 40 年代做出的,迄今仍被中外学者称道。^⑥人口数据是此类成果重要的支撑资料,但这些成果采用的中国近代人口数据并不准确,以致这些数据的后续使用受到了限制。

近年来,对包括中国近代在内的历史国民经济总量的学术探索方兴未艾,相关估算数据的问世,为分析中国社会历史上的相关问题提供了更为直观、便捷的分析工具。人口数据不仅是估算中国历史上的国民经济总量的必要材料,也成为一些估值被诟病的因素。因此,中国近代人口数量问题关涉的不仅仅是历史问题,也关系到现阶段的学术探讨。审视中国近代人口数据问题,对于时下估算与使用相应的历史数据具有重要的参考价值。

本文从社会调查学术发展史的视域出发,以“国民所得”所使用的人口数据为分析案例,对中国近代人口数量问题予以考释。文章首先阐释人口统计变革对社会历史分析指标化发展的推进作用,接着具体揭示人口数据在中国“国民所得”估算中的应用,藉此呈现中国近代人口数据在使用过程中的复杂性,最后结合“国民所得”人口数据探讨当前使用历史数据应有的认识。

一、人口统计在社会调查学术发展史中的作用渐变

“人口统计”的英文对应词为 Census,该词导源于公元前 6 世纪古罗马时期的拉丁文 Censcre,其大致含义是指古罗马公民须履行的户口及财产登记。^⑦除古罗马外,处于人类社会国家化发端期的古代中国、古巴比伦、古埃及也都存在过类似的社会调查统计活动。18 世纪前,这些调查统计活动通常是非定期的,主要目的是服务于政府管理,特别是了解辖区人口的赋税能力。在农业生产占主导的社会经济形态中,人口不仅是物质财富的依托,更被视为社会财富的一种。因此,在很长一段历史时期内,社会调查统计主要以人口统计的形式进行,一些能够揭示相应历史时期社会问题的官方、非官方材料由此得以累积。

① 王士达:《近代中国人口的估计(上)》,《社会科学杂志》(北平)1930 年第 1 卷第 3 期;《近代中国人口的估计(中)》,《社会科学杂志》(北平)1930 年第 1 卷第 4 期;《近代中国人口的估计(下)》,《社会科学杂志》(北平)1931 年第 2 卷第 1 期。

② 曹树基:《中国人口史》第 5 卷(清时期),上海:复旦大学出版社 2001 年版;侯杨方:《中国人口史》第 6 卷(1910—1953 年),上海:复旦大学出版社 2001 年版;严中平等编:《中国近代经济史统计资料选辑》,北京:科学出版社 1955 年版,第 362—374 页。

③ 陈启修:《中国人口的总数》,《国立北京大学社会科学季刊》(北京)1925 年第 3 卷第 4 号;Walter F. Willcox, “China’s Population—400,000,000 or 300,000,000?”, *The Chinese Students’ Monthly*, Vol. XXII, No.1 (November, 1926), pp.23-29.

④ 何炳棣著,葛剑雄译:《明初以降人口及其相关问题(1368—1953)》,北京:中华书局 2017 年版。

⑤ 葛剑雄、侯杨方、张根福:《人口与中国的现代化(1850 年以来)》,上海:学林出版社 1999 年版。

⑥ Paul B. Trescott, “The diffusion of western economics in China”, in Malcolm Warner, ed., *The diffusion of western economic ideas in east Asia*, London and New York: Routledge, 2007, pp.69-93.

⑦ 张延哲:《国势普查之意义及功用》,《国民经济建设》(镇江)1937 年第 2 卷第 5 期;梁方仲编著:《中国历代户口、田地、田赋统计》“总序”,上海人民出版社 1980 年版,第 2 页。

社会问题研究者对人口统计资料的整理、分析开启了调查统计学术化发展历程。^① 17 世纪左右,西欧地区的研究者在惯常采用的定性分析基础上融入包括人口统计数据在内的调查资料,藉此定量分析社会问题,统计资料分析逐渐发展成为一门学问,一些迄今仍被统计学界称道的著述陆续问世。约翰·格朗特(John Graunt)通过整理 17 世纪初教会编制的伦敦地区死亡人员报表撰写而成《根据死亡报表而作的自然和政治观察》(*Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality*),揭示了不同性别的出生率、死亡率差异等人口统计规律,并在此基础上编制了死亡率统计表。^② 威廉·配第(William Petty)则在《政治算术》(*Political Arithmetick*)中运用实际数据、推算数据、观察数据分析了英国、法国和荷兰三个国家的国力。威廉·配第因之被视为运用“算术的一般法则”论证“统治的事项,以及同君主的荣耀、人民的幸福和繁盛有极大关系的事项”的发明人。^③ 《政治算术》堪称运用数据进行国情比较研究的代表性著作,人口数据是支撑威廉·佩第论点的核心分析指标之一。此后,查尔斯·戴维南特(Charles Davenant)和格雷戈里·金(Gregory King)分别使用人口统计资料阐述了 17 世纪末英国部分地区家庭经济状况。然而,迨 18 世纪中叶,“政治算术”在英国渐呈颓势。

三十年战争后,在欧洲大陆上形成中的民族国家之间竞争激烈。1660 年起,海尔曼·康令(Hermann Conring)在赫尔曼斯坦大学讲授囊括人口、土地、财政等内容的“国势学”(Notitia rerum public arum)课程。1748 年,高特弗瑞德·阿亨华尔(Gottfried Achenwall)在《欧洲帝国的政治学概要》一书的引言中将海尔曼·康令讲授的学问称为“统计学”(Statistics),^④其方法是通过研究既有资料来描述国家发展状况。^⑤ 此类“统计学”研究者不探究社会现象之间的因果关联,也轻视数字化表达,因此,他们又被称为记述学派(Discriptive School)。^⑥ 18 世纪中期,随着源自英国的“政治算术”在欧洲大陆的传播,偏重纯粹记述方法的“统计学”在德意志地区的大学课程中急剧减少。尽管如此,18 世纪中期后,以人口统计为中心的“国势调查”在欧美地区陆续拓展至领土、资源、行业产品、进出口额等其他事项,并先后被欧美地区之外的国家采纳。19 世纪中期,人口统计资料除了应用于分析社会问题,还被阿道夫·凯特勒(Adolphe Quetelet)、卡尔·克尼斯(Karl Knies)等人用以归纳社会发展规律、验证一般的经济学理论。在基于大量搜集统计资料对社会现象进行整体分析思想的影响下,一个代表国家整体社会经济水平的国民经济总量指标——“国富”以国势调查指标之一的形式出现。限于当时的技术条件,早期估算“国富”的基本数据主要依托人口调查统计收集。^⑦ 事实上,在当今部分国家的国势调查中,人口统计仍然居于首要位置。

① Lars Behrlich, “Statistics and Politics in the 18th Century,” *Historical Social Research*, Vol. 41, No. 2 (January, 2016), pp.238-257.

② Clarence J. Glacken, *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the End of the Eighteenth Century*, Berkeley: University of California Press, 1967, pp.398-400.

③ [英]威廉·配第著,陈冬野译:《政治算术》,北京:商务印书馆 2014 年版,第 1—2 页。

④ Gottfried Achenwall, “Vorbereitung zur Staatswissenschaft der heutigen fürnehmsten Europäischen Reiche und Staaten worinnen derselben eigentlichen Begriff und Umfang in einer bequemen Ordnung entwirft und seine Vorlesungen darüber ankündigt”, Göttingen: Vandenhoeck, 1748, p. 2, accessed December 3, 2022, <http://resolver.sub.uni-goettingen.de/purl?PPN637704622>.高特弗瑞德·阿亨华尔用德语“Statistic”称呼这门学科,由于“c”“k”在当时的德语中通用,所以后来文献中也将“Statistic”写作“Statistik”。“Statistic”是高特弗瑞德·阿亨华尔依据现代拉丁语“Statisticum”创造出来的,所以当前学界根据“Statistics”一词语源将统计学学科史追溯至古希腊时期的观点可能是值得商榷的。

⑤ 刘鸿万:《统计学史的命题》,《法学专刊》(北平)1936 年第 6 期;张一兵:《回到马克思——经济学语境中的哲学话语》(第四版),南京:江苏人民出版社 2020 年版,第 417 页。

⑥ 超超:《统计学考》,《周行》(广州)1936 年第 1 期。

⑦ 缪德刚:《国民经济核算思想在近代中国的传播——基于“国富”学说的考察》,《中国经济史研究》2019 年第 5 期。

19 世纪末、20 世纪初,新古典经济学、福利经济学的兴起先后影响了国民经济核算理念与方法的变革,经过德奥方法论之争,经济学理论与现实社会问题之间的嫌隙被不断弥补,经济学者越发重视发挥现实社会问题与经济学理论的交互作用。一战前后,财富和收入分配研究也被越来越多的经济学者视为缓解社会冲突的手段,能够反映个人福利水平的国民经济流量指标的测算愈发受到关注。在此背景下,布鲁金斯学会、美国国家经济研究局等致力于国民经济调查与研究的机构相继成立。20 世纪三四十年代,科林·克拉克(Colin Clark)出版了系列著作,对估算国民收入(National Income)在方法论上做出了重要建树,受当时经济科学主流理念的影响,科林·克拉克的研究一直秉持的是与个人福利相关的国民收入总量数值的估算。西蒙·库兹涅茨(Simon Kuznets)同样认为个人福利应该是国民经济核算关注的主要对象。20 世纪 40 年代中期,由国际联盟统计专家委员会国民收入统计小组委员会完成的《国民收入计量与社会账户构建》中,对国民收入的计量建立在个人收入统计基础之上。^① 这同样是国民经济统计学界针对国民收入的核算问题受到当时主流思想与方法影响的体现。不过,与科林·克拉克、西蒙·库兹涅茨等人不同的是,约翰·梅纳德·凯恩斯(John Maynard Keynes)认为,国民经济核算除了个人部分,还应当包括公共部门。在该主张的影响下,国民经济统计学界对国民经济流量指标核算的重心开始从国民收入转向国民生产总值(Gross National Product)。20 世纪 40 年代,“国民收入”和“国民生产总值”在绝大多数中文著述中被称为“国民所得”。由于规范化的国民账户体系尚未形成,加之统计机构尚不健全,几乎所有国家的“国民所得”数值系采用包括人口统计数据在内的既有资料合算得到。

可见,早期人口统计主要是一种政治管理辅助手段,其数据是治国理政的重要参考资料。至少从 17 世纪开始,人口统计除了的政治领域发挥过往的功能之外,人口统计数据的应用呈现出明显的学术化发展态势。18 世纪中期,国家之间社会经济比较、经验分析技术的提升、社会化生产的需求等因素推动了系列国民经济综合评价指标陆续出现,人们在对社会经济现象的分析中,愈发仰赖于此类指标,而人口统计数据是得到相关指标的基础数据之一。

二、人口资料在中国“国民所得”估算中的应用

20 世纪 40 年代,中外学者通过不同途径测算了中国货物与劳役的币值总额,先后问世了 10 余种估值。^② 在国民经济诸领域时间序列数据、截面数据不足以及缺乏规范的“国民所得”核算原则等多重困境下,估算者通过不同方法将各类调查、统计材料合算得到。在 10 余种中国“国民所得”估值中,来源不一的人口数据是大部分估值的重要依托材料。

较早用到人口数据对中国“国民所得”进行估算的是英国学者科林·克拉克。1940 年,他在《经济进步的条件》一书中披露了 1925—1934 年中国“国民所得”估值。科林·克拉克首先分别估算农业所得和非农业所得,然后将二者合计得到“国民所得”总额数值。^③ 在农业所得和非农业所得的估算中,科林·克拉克使用了理查德·托尼(Richard Tawney)《中国的土地与劳动》(*Land and Labour in China*)中的人口数据。《中国的土地与劳动》中的中国人口数据只是一个概数,理查德·托尼认为 20 世纪 30 年代初中国人口数量不少于 4.00 亿,不多于 4.50 亿,他进一步指出,农业人口占总人口的 3/4。^④ 科林·克拉克在其估算中参考了 20 世纪 30 年代中国总量人口数值和行业人口数值,《中国的土地与劳动》中的两种数值显然是十分粗略的。与理查德·托尼设想不同的是,

① Report of the Sub-committee on National Income Statistics of the League of Nations Committee of Statistical Experts, *Measurement of National Income and the Construction of Social Accounts*, Geneva: United Nations, 1947, pp.11-12.

② 缪德刚:《20 世纪 40 年代中国“国民所得”估算问题》,《清华大学学报(哲学社会科学版)》2022 年第 2 期。

③ Colin Clark, *The Conditions of Economic Progress*, London: Macmillan & Co. Ltd, 1940, pp.44-46.

④ R. H. Tawney, *Land and Labour in China*, London: George Allen & Unwin Ltd., 1932, pp.23-27.

科林·克拉克在估算中将农业、非农业人口比例取为 80%、20%。

$$\text{农业所得} = \frac{\text{每家平均收入}}{\text{每家人口}} \times \text{农业人口}$$

$$\text{非农业所得} = \text{大工厂工作人口净产额} + \text{其他行业人口总工资}$$

$$\text{“国民所得”} = \text{农业所得} + \text{非农业所得}$$

1942 年冬,刘大中在加拿大魁北克的蒙特布朗举办的太平洋国际学会第八次年会上提交会议论文《中国国民所得与战后工业化》(Some Preliminary Notes on the National Income and Post-war Industrialization of China),参考了科林·克拉克的成果估计了 1929—1934 年的中国“国民所得”。在这次估算中,刘大中采取了从宽(Liberal Estimate)的估计态度,所以他选用了理查德·托尼推测的中国人口总数最高值——4.50 亿,并将中国总人口数值的 80%作为由地主与农场经营主及雇佣劳工等组成的农业人口、20%作为工商业及政府官员等非农业人口,分别估算农业所得和非农业所得。^①

1946 年,刘大中完成并出版《1931—1936 年中国的国民所得:一个探索性研究》(*China's National Income 1931-36: An Exploratory Study*),首先估算了 22 个省份的农业所得和非农业所得,然后把二者相加并取均值后乘以 22 省份之外的其他省区的人口比例,计算出东北、热河、新疆、蒙古、西藏等其他省区的“国民所得”,最后通过加总 22 省“国民所得”和其他省区“国民所得”得到中国“国民所得”总额数值。^②

$$22 \text{ 省“国民所得”} = \text{农业所得} + \text{非农业所得}$$

$$\text{其他省区“国民所得”} = 22 \text{ 省“国民所得”均值} \times \text{其他省区人口比例}$$

$$\text{中国“国民所得”} = 22 \text{ 省“国民所得”} + \text{其他省区“国民所得”}$$

第二次估算过程中,刘大中秉持从严(Conservative Estimate)的估算态度,人口总数数值采用了 4.10 亿。该数值自 1940 年《中华民国统计提要》中记载的苏、皖、浙等 22 省的人口数额换算得到。《中华民国统计提要》载 22 省农业人数 2.93 亿,刘大中又根据《中国土地利用》中的农业、非农业人口比重 79%、21%,得到 22 省非农业人口 0.78 亿,故 22 省计 3.72 亿人,该数值约占全国人口总数的 90.7%。同时,依据《中华民国统计提要》的记载,22 省之外的各省区人口数占 9.3%,计约 0.38 亿人。故而,中国人口总额数值为 4.10 亿。^③ 后来,刘大中在与叶孔嘉合作的《中国大陆经济:国民收入与经济发展(1933—1959)》中结合沃尔特·威尔考克斯(Walter Willcox)1912 年的中国人口估值,并调低男性比例、增加瞒报儿童数,将 1933 年的中国人口总数估算为 5.00 亿。^④

中央设计局运用内政部人口统计调查等数据调整了刘大中在《中国国民所得与战后工业化》中的 1929—1934 年中国“国民所得”估值,得出 1932—1936 年中国“国民所得”数值。其中,农业、非农业人口比例分别取值为 80%、20%,热河、察哈尔、绥远、宁夏四个省份的农业人口占全国农民人口比例取值 1.6%,新疆、西藏、蒙古三省区人口占全国农民人口比例取值 3%。^⑤

中央研究院社会科学研究所巫宝三的主持下先后三次估算了中国“国民所得”,成果分别为

① 刘大中著,杨叔进译:《中国国民所得与战后工业化》,《大公报》(重庆)1944 年 5 月 21 日,第 3 版;《中国国民所得与战后工业化(续)》,《大公报》(重庆)1944 年 5 月 22 日,第 3 版。

② Ta-Chung Liu, *China's National Income 1931-36: An Exploratory Study*, Menasha: George Banta Publishing Company, 1946, p.10.

③ Ta-Chung Liu, *China's National Income 1931-36: An Exploratory Study*, pp.58-66.

④ Ta-Chung Liu, Kung-Chia Yeh, *The Economy of the Chinese Mainland: National income and Economic Development, 1933-59*, Santa Monica: The Rand Corporation, 1963, pp.254-265.

⑤ 陈光照:《巫宝三:国民所得概论》,《中农月刊》(南京)1945 年第 6 卷第 10 期;丁鹤:《中国国民所得的五种估计》,《中央银行月报》1946 年第 1 卷第 6 期。

《中国国民所得(一九三三年)》《“中国国民所得,一九三三”修正》《中国国民所得,1933、1936、1946》。《中国国民所得(一九三三年)》主要估算了1933年的中国“国民所得”,其总人口数值取自《中国年鉴(1936—1937)》(The China Year Book, 1936—1937)的4.29亿。^①南钟万、贝友林在用消费与投资法估计农户人口消费时,主要采取了张心一《中国农业概况估计》中的农业、非农业人口比重数据。其中,农户占比75.52%,非农户占比24.48%。^②巫宝三等人的估算还用到了矿业业工人人数、外籍工厂工人人数、商店店员数等行业人口数据。方锦认为,巫宝三等的这次估算中有些职业人口如律师只计算了全国律师公会会员数而非会员没有被包含在内,所以此类数字是有遗漏的。^③《“中国国民所得,一九三三”修正》是对《中国国民所得(一九三三年)》的修订,总人口数仍然取值为4.29亿。^④农业人口、非农业人口分别占总人口的75%、25%。其中,城市人口占全部非农业人口的20%。此次估算还用到了工厂工人人数、手工业常年工人人数、自由职业人数、家内仆佣所占的人口比例等。经过这次修正,中国人均所得增加了国币1元。《中国国民所得,1933、1936、1946》的主要目的是为了估算1946年的中国“国民所得”,这次估算的中国人口总数采用了1946年国民政府主计处统计的数值4.49亿。^⑤

1944年,刘大钧主持了军事委员会国民经济研究所的中国战前“国民所得”估算。这次估算是依据“所得法”将租金、利息、利润、薪金与工资四项相加得到,采用了中国境内的总人口数值4.50亿。其中,城市人口15%,农村人口80%,有5%的人口漏列。^⑥

表1 中国“国民所得”总量人口数据概况

	估算者	估算成果	估值年份	基础数据来源	总值(亿)	比例
1	科林·克拉克	《经济进步的条件》	1925—1934年	《中国的土地与劳动》	4.50	农业80%; 非农业20%
2	刘大中	《中国国民所得与战后工业化》	1929—1934年	《经济进步的条件》	4.50	农业80%; 非农业20%
3	刘大中	《1931—1936年中国的国民所得:一个探索性研究》	1931—1936年历年	《中华民国统计提要》 ¹	4.10	农业79%; 非农业21%
4	中央设计局	《中国国民所得之估定》	1932—1936年	内政部人口统计	4.67 ²	农业80%; 非农业20%
5	社会科学研究 所巫宝三等	《中国国民所得(一九三三年)》	1931—1936年历年	《中国年鉴》 《中国农业概况估计》等	4.28	农户75.52%; 非农户24.48% ³
6	巫宝三	《“中国国民所得,一九三三”修正》	1931—1936年历年	《中国年鉴》等	4.28	农户75%; 非农户25%
7	巫宝三	《中国国民所得,1933、1936、1946》	1946年	《中华民国统计提要》等	4.49 ⁴	—— ⁵

① 巫宝三:《中国国民所得(一九三三年)》,北京:商务印书馆2011年版,第26页。

② 巫宝三:《中国国民所得(一九三三年)》,第202页。

③ 方锦:《巫宝三书评:中国国民所得(一九三三年)》,《经济评论》(上海)1948年第2卷第21期。

④ 发表于《社会科学杂志》(南京)1947年第9卷第2期的《“中国国民所得,一九三三年”修正》的1933年人口数值为429 494 138,而非后来2011年商务印书馆版的429 492 138(参见巫宝三《中国国民所得,一九三三》修正》,巫宝三:《中国国民所得(一九三三年)》,第721—722页)。

⑤ 巫宝三:《中国国民所得,1933、1936、1946》,《社会科学杂志》(南京)1947年第9卷第2期。

⑥ 丁鹤:《中国国民所得的五种估计》,《中央银行月报》(上海)1946年第1卷第6期。

续表

	估算者	估算成果	估值年份	基础数据来源	总值(亿)	比例
8	国民经济研究所刘大钧	《中国战前国民收入估计》	战前	——	4.50	农村 80%； 城市 15% ⁶

说明:人口数值为四舍五入后的值。

注:1.《中华民国统计提要(1940)》记载了全国人口数额 4.74 亿(不包括侨民),但是刘大中并没有采用该数值。结合刘大中后来的研究来看,该人口数据可能受到了沃尔特·威尔考克斯估计方法的影响,相关评述可参见邢慕寰《我国国民所得数字的商榷》,《资本市场》1948 年第 1 卷第 5 期。

2.具体数值为 466 85 856。

3.此人口比例数据系南钟万、贝友林自消费领域对中国“国民所得”进行估计时用到的。

4.具体数值为 449 255 736。

5.此次估计未涉及消费领域的估计,故只用到了全国人口总数。

6.国民经济研究所的人口比例遗漏了 5%。

除了上述估值之外,孙拯、程孝刚、方显廷、何廉等人也对中国“国民所得”进行了估算。在 20 世纪 40 年代的 10 余种中国“国民所得”估值中,仰赖于人口数据资料的主要是表 1 中所列示的 8 种。这 8 种估值除巫宝三等人的《中国国民所得,1933、1936、1946》外,基本上是估算者在 20 世纪 40 年代运用相关资料测算得到的 20 世纪 20、30 年代的中国“国民所得”数值。这些估算成果所使用的人口数据主要是人口总量数据、行业人口数据和地区人口数据。尤其是 8 种估值都使用到了人口总量数据,它们的人口总量数据范围在 4.10 亿至 4.67 亿之间。在缺乏充分资料的情况下,为了便于估算,大部分估算者按照产业或地域设置了农业和非农业人口比例。其中,从事农业生产活动的人口占中国总人口的 75%至 80%,从事非农业生产活动的人口占中国总人口的 15%至 25%,非农业部门的行业人口数据来源最为庞杂,关于这一点,巫宝三等《中国国民所得(一九三三年)》表现最为鲜明(见表 2)。利用到地区人口数据的分别是刘大中的《1931—1936 年中国的国民所得:一个探索性研究》、中央设计局的《中国国民所得之估定》。

表 2 《中国国民所得(一九三三年)》自由职业人口数据来源

职业名称	数据来源
私立学校教员	《中国年鉴》(1936—1937);《甘肃省统计季报》(1935);《广西年鉴》(1934);《福建统计年鉴》(1936);《湖北省年鉴》(1935);《四川省概况》(1935);《伪满洲国年报》(1933);《统计汇刊》(1929);《中国经济志》(南京、江宁、当涂、休宁,1933);《上海市统计》(1928);《杭州市经济调查》(1931);《中国经济志》(嘉兴,1931);《中国经济志》(长兴,1932);《中国经济志》(注县、宁国,1934);《芜乍路沿线经济调查》(1932);《渝柳段线川黔经济调查总报告》(1931);《湘滇线云贵段附近各县经济报告书》(1930)
医师	《内政统计季刊》(1933);《中国年鉴》(1931);《广西统计季刊》(1933、1937);《伪满洲国年鉴》(1933);《鄞县县政府统计特刊》(1933);《中国经济志》(吴兴,1933);《统计汇刊》(1929);《首都警察概况》(1934);《上海市统计》(1932);《湖北省年鉴》(1936);《浙江省政务概况》(1933);《福建统计年鉴》(1935);中央卫生设施实验处生命统计室(1933)
僧、道、喇嘛、教士	《申报年鉴》(1930);《中国年鉴》(1931);《蒙古年鉴》;《伪满洲国年鉴》;《蒙藏新志》
律师	《申报年鉴》(1933);《中国年鉴》(1931)
会计师	实业部登记数据
新闻记者、家内仆佣	《中国年鉴》(1931)

资料来源:巫宝三:《中国国民所得(一九三三年)》,第 167—179 页。

三、使用近代人口数据应有的认识

20世纪前叶,不同的估算者根据各异的方法测算了中国人口数据。有观点指出,仅1911年至1949年间,关于中国人口各类估值有100余种。^①整体上看,1912年前由不同估算者推测的同类数值之间差异幅度较大,1912年后相关数据的差异幅度明显降低。从表3列示的得到其他学者较多介绍或引用的全国人口数据来看,其中一些数据在20世纪40年代已经发布,它们应该是可资“国民所得”估算使用的。然而,不同“国民所得”估算者所使用的人口数据不尽相同。造成“国民所得”人口数据使用差异的原因包括:其一,人口数据可获得性不足,如《经济进步的条件》一书用到了理查德·托尼估计的中国人口概数;其二,估算方法不同,如《中国国民所得(一九三三年)》中,依据所得法的估算比消费与投资法使用到的人口数据更为详备;其三,估算态度有别,如尽管《中国国民所得与战后工业化》与《1931—1936年中国的国民所得:一个探索性研究》都出自刘大中的估算,但前者“从宽”、后者“从严”,刘氏因此选取了不同的人口数据;等等。

表3 20世纪前叶中国人口数量估计

估算者	年份	人口数量(亿)
柔克义	1904	2.75
丁瑞门	1910	3.29
沃尔特·威尔考克斯	1910	3.40
清政府民政部	1910	3.43
沃尔特·威尔考克斯	1912(调整前)	3.93
沃尔特·威尔考克斯	1912(调整后)	4.12
柔克义	1912	3.11
刘大中、叶孔嘉	1912	4.34
北洋政府内务部	1912	4.06
陈华寅	1912	3.93
海关调查	1918	4.39
中华续行委办会	1918—1919	4.41
邮政局	1920	4.28
海关报告	1923	4.45
陈长蘅	1923	4.43
海关报告	1925	4.48
陈启修	1925	5.47
海关调查	1927	4.58
国民政府内政部	1928	4.42
陈长蘅	1928	4.70
Harry P. Howard	1928	4.97
邮政局	1928	4.86

^① 葛剑雄:《中国人口发展史》,福州:福建人民出版社1991年版,第66页。

续表

估算者	年份	人口数量(亿)
沃尔特·威尔考克斯	1929	4.40
陈华寅	1929	4.45
陈长蘅	1930	4.56
国民政府内政部	1931	4.75
刘大中、叶孔嘉	1933	5.00
国民政府主计处统计局	1933	4.44
胡焕庸	1935	4.59
国民政府内政部	1936	4.71 ¹
国民政府内政部	1937	4.79
国民政府主计处统计局	1944	4.64
国民政府主计处统计局	1946	4.56
国民政府内政部人口局	1947	4.61

资料来源:1.1904年、1912年柔克义的数据来自徐履诚《论我国之人口统计》,《东方杂志》(上海)1944年第40卷第23号。

2.1910年丁瑞门的数据来自 Raymond P. Tenney, "Census Returns of China," *Daily Consular and Trade Reports*, Washington, July 13, 1911, pp.184-187.

3.1910年、1912年(调整前)、1929年沃尔特·威尔考克斯的数据来自《国际统计会议中中国人口之论争》,《银行周报》(上海)1930年第14卷第37号;《国际统计会议中中国人口之论争》,《东方杂志》(上海)1930年第27卷第18号。

4.1912年(调整后)沃尔特·威尔考克斯的数据来自 Ta-Chung Liu, Kung-Chia Yeh, *The Economic of the Chinese Mainland: National income and Economic Development, 1933-59*, pp.254-265.

5.1910年清政府民政部、1918年海关调查数据来自陈启修《中国人口的总数》,《国立北京大学社会科学季刊》(北京)1925年第3卷第4号。

6.1920年邮政局数据来自王士达《近代中国人口的估计(下)》,《社会科学杂志》(北平)1931年第2卷第1期。

7.1912年北洋政府内务部、1928年国民政府内政部、1931年国民政府内政部、1930年陈长蘅、1935年胡焕庸、1936年国民政府内政部、1947年国民政府内政部人口局数据来自葛剑雄《中国人口发展史》,第257—258页。

8.1912年、1933年刘大中及叶孔嘉的数据来自 Ta-Chung Liu, Kung-Chia Yeh, *The Economy of the Chinese Mainland: National income and Economic Development, 1933-59*, pp.254-265.

9.1912年、1929年陈华寅数据来自陈华寅《民国十八年中国人口总数之推测》,《统计月报》(南京)1930年第2卷第9期。

10.1918—1919年中华续行委办会的数据不包括外蒙古,该数据来自中华续行委办会调查特委会编、蔡咏春等译《1901—1920年中国基督教调查资料》(上卷),北京:中国社会科学出版社2007年版,第28页。

11.1923年、1925年、1927年海关报告数据来自陈华寅《最近中国之人口统计》,《统计月报》(南京)1929年第1卷第1期。

12.1923年陈长蘅数据来自陈长蘅《中国近百八十年来人口增加之徐速及今后之调剂方法》,《东方杂志》1927年第24卷第18号。

13.1925年陈启修的数据来自陈启修《中国人口的总数》,《国立北京大学社会科学季刊》(北京)1925年第3卷第4号。

14.1928年 Harry P. Howard 数据来自王士达《最近十年的中国人口估计》,《社会科学杂志》(北平)1931年第2卷第2期。

15.1928年陈长蘅、1928年邮政局数据来自陈长蘅《中国人口总数的新推测》,《统计月报》(南京)1929年第1卷第3期。

16.1933年国民政府主计处统计局、1937年国民政府内政部、1946年国民政府主计处统计局数据分别来自国民政府主计处统计局编《中华民国统计提要(廿四年辑)》,上海:商务印书馆1936年版,第227页;Directorate of Statistics National Government, *Statistical Abstract of the Republic of China 1940*, Chuanking, p.24;民国三十六年国民政府主计处统计局编《中华民国统计提要》,南京:国民政府主计处统计局1947年版,第2页。

17.1944年国民政府主计处统计局数据来自郑成林选编《民国时期国情统计资料汇编》(第一册),北京:国家图书馆出版社2016年版,第43页。

说明:人口数据经四舍五入得到。

注:1.该数值不包括0.08亿的海外侨民。

数字化时代的社会历史分析中,不可避免地使用到历史数据。相较于当前通过统计方法得到的梳理,历史数据通常存在或多或少的“缺陷”。从中国近代人口数量问题一度引起的学术争论及其在“国民所得”估算中的应用来看,包括人口数据在内的历史数据不仅是学术问题的载体,也为利用、剖析相关历史数据提供了路径。

综括来看,中国近代人口数据大致可以分为直接数据和间接数据两种。其中,直接数据是通过调研得到的数据,如1910年清政府人口调查中县一级造册数据、1910年邮政局在地方通过问询得到的数据、北洋政府内务部制定的各类人口调查表中的数据、中华续行委办会收到的县级人口报告数据、地方志中的数据等,此类资料包含的数据量庞大,但未必切实,使用此类数据需要对调查方式有全面的了解。间接数据可以分为汇总数据和再估算数据。汇总数据是由地方分类统计数据合计而成的国家级、区域级、省级人口数据,如全国人口数据、22省人口数据、东北人口数据等,不同时期辖区范围有无变化是当前使用此类数据需要重点关注的。再估算数据是在同类或非同类数据基础上得到的估值,《1931—1936年中国的国民所得:一个探索性研究》中东北、热河、新疆、蒙古、西藏等边疆民族地区的人口数量是根据这些地区占全国人口数量比重折算得到的,该数据属于由同类数据换算而来的观察值,而北洋政府时期海关利用生活物品的消费量推算的人口数据是使用非同类数据计算的推测值。一般而言,再估算数据较之基础数据的客观度稍次。

同样应该认识到,已有著述中的中国近代人口数据是有时空局限的。20世纪中期前,接触到中国人口数据的基本上是三类人员:第一类是曾经到访过中国、对中国情况相对熟悉的外国研究者,他们往往对统计技术缺乏必要的了解,却较早地整理了中国人口数据;第二类是具有统计知识储备的外国研究者,其中不乏国际知名的统计学者,但他们几乎没有到访过中国,对于中国的具体情况并不了解,为文化差异所限,他们一般难以直接使用中国人口材料,因此第一类人员整理的数据成为他们的分析材料;第三类主要由掌握了一定统计技术的中国学者构成,他们能够发现中国人口数据使用过程中的相关问题,但较之第一类、第二类人员,他们在某些领域的学术影响力不足,有关成果难以获得国际性的认可。近代以前,第一类人员的成果在欧美地区最具影响,而到了近代早期,第二类人员的著述提升了国际学界对中国人口问题的关注度,20世纪三四十年代中国学者成为该领域不可小觑的研究力量。以上三类人员曾面临的时空局限在他们的著述中皆有一定的体现,这是参考他们成果时应当注意的。比如,第一类人员在对中国人口规模进行估计时,由于受资料限制,着重于估算汉族聚集区的人口数量,而这种学术传统在一定程度上影响了第二类人员对中国近代人口问题的探讨。相较而言,中国本土学者对人口数量的估算从地理区域上更贴近中国当前的幅员,但中国近代边疆民族地区人口数量相较于汉族聚集区域的人口数量准确性依然不足,直到1953年中华人民共和国中央人民政府着手全国人口普查,该问题才得以解决。