

量化历史与新史学*

——量化历史研究的步骤和作为新史学的价值

林 展 陈志武

【提要】 随着历史大数据时代的到来,大量历史数据库建成和向学界开放,历史研究也面临新的挑战,即如何有效利用大规模的史料。量化历史作为新史学的重要组成部分,是应对历史大数据挑战的重要方法之一。近年来,量化历史在国际学术界发展较快,形成了较为完善的分析方法,本文以基于《新教伦理与资本主义精神》展开的量化历史研究为例,详细说明其研究步骤,分析其在应对历史大数据的挑战、识别历史的长期影响、促进历史学与其他社会科学的交流与对话等方面可能发挥的作用。

【关键词】 量化历史 新史学 历史大数据

当代新史学的发展,引发了“史料之革命”,扩展了史料的范围,形成了多元的史料体系,进而也引发了历史资料的“大爆炸”。^① 这一发展的最近表现是历史数据库的大量出现。不少文章对数据库在历史研究中的价值进行了多角度的介绍,^②但是很少有学者讨论如何有效利用这些历史数据库,特别是如何基于结构化历史数据库开展研究,这不利于真正发挥历史数据库的价值,也不利于史料革命的进一步深化。量化历史在如何利用大规模数据库方面,已经形成了较为完善的分析方法,但史学理论中对这些方法和具体实施步骤的介绍还比较少。^③ 本文希望在量化历史研究步骤及其对历史研究的价值这两方面提出一些初步的想法,抛砖引玉,推动学界提供更多这方面的讨论。

一、量化历史的含义及与新史学的关系

量化历史研究是交叉学科,是用社会科学理论和量化分析方法来研究历史。量化历史研究目的是发现历史的规律,即人类行为和人类社会的规律。在量化历史研究中,称这些规律为因果关系,量化历史研究的过程,就是发现因果关系的过程。理解量化历史研究的含义,一般需要结合三个角度,

* 本文是国家社会科学基金重大项目“清末民国社会调查数据库建设”(项目编号:15ZDB041)的阶段性成果。

① 徐善伟《当代西方新史学与“史料之革命”——兼论中国新史学史料体系的重构》,《史学理论研究》2010年第2期。

② 相关介绍参见梁晨、董浩、李中清《量化数据库与历史研究》,《历史研究》2015年第2期;夏明方《大数据与生态史:中国灾害史料整理与数据库建设》,《清史研究》2015年第2期。

③ 对量化历史发展过程、价值和前景的介绍,参见陈志武《量化历史研究告诉我们什么》,《量化历史研究》2014年第1期;陈志武《量化历史研究的过去与未来》,《清史研究》2016年第4期;陈志武《量化历史研究与新知识革命:以财富差距与消费差距的历史研究为例》,《北京大学学报》2018年第4期。

即社会科学理论、量化分析方法、历史学。

理解量化历史的第一个角度,是其广泛借鉴社会科学的理论。社会科学包含经济学、金融学、管理学、政治学、法学、社会学、人类学、教育学及心理学等。就研究涉及的领域而言,量化历史包含对政治、经济、思想文化、环境等历史的方方面面的量化分析。上述不同的研究领域,一个共同的特征就是需要收集数据,通过量化分析的方法来回答特定的问题,进而来理解人类行为和人类社会。基于数据的研究通常也称为经验研究(empirical analysis)。

社会科学理论是关于人类社会某些特征的一组系统化的论断,或者是对于真实世界的系统化解释。这些理论有两个重要的特征:一是可以引申出可检验的假说或推测;二是随着支持这些假说或推测的经验证据越来越多,理论的可靠性程度也越高。^①

理论对于历史研究的价值,在于为分析、理解历史提供了一个基准和框架。这一框架能够帮助研究者从纷繁复杂的历史史料中快速梳理出一个研究的起点、参照系或者靶子,成为理解历史的重要工具。历史资料错综复杂、千头万绪,而人的认知能力有限;如果缺乏理论分析框架,既不利于认知历史规律,也不利于历史研究成为一门代际累进的学科。

理解量化历史的第二个角度是量化分析方法,这包括统计学、计量经济学、人工智能等领域的方法。量化方法不是一个静止的概念,而是处于不断发展中。随着相关学科知识的进步、计算机技术的发展,新的量化方法不断出现,能够处理的数据量越来越多、效率越来越高,发现的结论也会越来越可靠。起初,历史研究中的“量化方法”很简单,差不多就是“用数据说话”,加上图表和一些普通的统计指标,比如均值、方差和相关系数,^②但如今,不止如此。限于篇幅,本文仅选择三个角度来介绍量化方法,一是对历史现象的度量,二是寻找历史现象的相关性,三是发现历史对象之间的因果关系。为了发现历史规律,通常需要进行因果关系的分析,对历史现象的度量和寻找历史现象的相关性是寻找因果关系的步骤或起点,通常也可以作为一个单独的研究。这里的量化分析方法是大多数社会科学共同使用的方法,与自然科学实验方法的逻辑也是一致的。目前在社会科学中,经验研究方法正在从统计推断(statistical inference)向因果推断(causal inference)转变,这种转变被安格里斯特和皮施克称为经验研究的“可信性革命(credibility revolution)”。^③这场革命的核心是基于随机试验的思想来获得因果关系,即规律性认识。

需要强调的是,量化方法的本质是高效率的处理大规模信息,从中获得规律认识。其基于众多现实问题而产生,已发展成为一个内容丰富的学科,有些部分变得非常专业和高深,如何将不同的量化工具与历史研究有效结合起来,发挥量化方法的优势,需要不断尝试和探索。

理解量化历史的第三个角度是其与历史学的关系,这在陈志武之前的研究中已有讨论,两者的关系“体现在量化方法不是要取代传统历史研究方法,而是对后者的一种补充,是把科学研究方法的全过程带入历史领域。整理考证史料、注重文献是历史学研究的传统,量化史学同样注重对历史文献的考证、确认,这一点没有区别”。^④正如本文开篇提到的,由于新史学带来了史料革命,让史料规模爆炸式增长,这让定性方法面临挑战,而量化方法则可以较好地应对这一挑战。但仅有量化

① Janet Buttolph Johnson, H. T. Reynolds, Jason D. Mycof, *Political Science Research Methods*, CQ Press, 2016, p. 54.

② 陈志武《量化历史研究的过去与未来》,《清史研究》2016年第4期。

③ Joshua D. Angrist, Jörn-Steffen Pischke, “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics”, *Journal of Economic Perspectives*, Vol. 24, No. 2, 2010, pp. 3-30.

④ 陈志武《量化历史研究的过去与未来》,《清史研究》2016年第4期。

并不够,量化分析需要建立在扎实的定性研究基础之上,需要与历史学定性方法相结合,理解史料的历史背景,否则不管采用什么研究方法,得出的结论都不会可靠。只有与历史学的已有研究深度融合,量化历史才能更好地实现自身的价值。

一项合格的量化历史研究需要同时达到上述三个学科分支的要求。这也意味着,一项好的量化历史研究并不容易实现,需要细致的工作和艰辛的努力。

由量化历史的概念,可知其是新史学的重要组成部分,尽管新史学有不同的含义,也处在不断发展变化的过程中,但在对新史学的不同解释中,一般都强调尽可能结合人文社会科学的知识和方法来研究历史。随着人文社会科学研究方法的量化程度大大提高,量化分析已经在社会科学研究中唱主角,在人文学科中的影响也越来越大。

二、量化历史的研究步骤

《量化历史研究的过去和未来》一文介绍了科学研究的基本流程。这一流程大致分为五个步骤:第一是提出问题和假说;第二是寻找史料和数据;第三是对数据进行量化分析,寻找因果关系;第四是对发现的因果关系进行解释和寻找作用机制;第五是论文的写作。但该文没有对这些步骤的具体操作展开论述。^①在此,我们以对“韦伯假说”的相关量化分析为中心,结合其他成果,讨论量化历史研究过程中的主要步骤,限于篇幅,仅针对前四个步骤,论文写作在此不作介绍。

我们用来作为示范的例子是围绕《新教伦理与资本主义精神》展开的量化历史研究。这是马克斯·韦伯影响最大的著作之一,根据“谷歌学术”的统计,被引用超过三万次。在该书中,他认为新教地区有更为繁荣的经济,新教伦理对经济有重要的促进作用。虽然该书通常被认为是定性研究,但却建立在大量的统计数据之上,比如在书的开篇,韦伯就提出了下列基于统计数据的发现:

在一个各种宗教信仰混杂之处,只消一瞥其职业统计,往往便会发现一个屡见不鲜的现象,此一现象在天主教的报章和文献及德国的天主教会席上一再引发热烈的讨论,那就是:在近代企业里,资本家与企业经营者,连同熟练的上层劳动阶层,特别是在技术上或商业上受过较高教育训练者,全都带有非常浓重的基督新教的色彩。^②

由此开始,韦伯进一步提出了新教伦理与经济发展之间的关系,这被认为是关于文化与经济发展最重要的论断,即新教伦理越强的地方,经济可能越繁荣。^③在余下部分,为表述方便,我们暂时称之为“韦伯假说”。^④

韦伯假说隐含了新教伦理可能影响经济发展的途径,一是新教徒工作更为努力,二是新教徒储蓄更多,进而投资更多,从而提高了长期的生产率。上述两种途径使得新教伦理推动了资本主义

^① 陈志武《量化历史研究的过去和未来》,《清史研究》2016年第4期。

^② 马克斯·韦伯《新教伦理与资本主义精神》,康乐、简惠美译,上海三联书店2019年版,第8—9页。黑体字为原文所加。

^③ Daron Acemoglu, Simon Johnson, and James A. Robinson, “Institutions as a Fundamental Cause of Long-Run Growth”, in Philippe Aghion, Steven N. Durlauf, eds. *Handbook of Economic Growth*, Vol. 1a., Elsevier B. V., 2005, pp. 385–472.

^④ 当然,《新教伦理与资本主义精神》讨论的内容很多,涉及很多不同的观点,我们这里的讨论只集中于新教伦理是否推动了经济增长这一判断。

发展。

直到今天,韦伯的著作依然是不同学科的必读书,影响很大。但这只是解释资本主义经济发展的观点之一。对于这一观点,是否应该接受?如何来检验?如果基于史料,正如韦伯所发现的,新教徒数量越多的地方,经济发展更好,如此就能断定新教伦理推动了资本主义发展吗?对上述问题的回答,构成了一项典型的量化历史研究。

这一研究由两位经济学家贝克尔(Becker)和沃斯曼(Woessmann)完成。他们利用19世纪普鲁士^①452个县(郡)级政区的调查数据,发现新教徒占比越高的地区,经济发展确实越好(比如有更高的人均所得税收入、教师收入,更大的非农业部门规模),这与韦伯的发现一致。但是,当他们将各地平均识字率的差别和新教徒占比的差别放到一起时,发现后者的影响消失了。他们认为,新教之所以推动了经济增长,主要是由于马丁·路德呼吁所有人自己去阅读《圣经》并直接跟上帝对话,从而(意料之外地)提高了读写能力,推动了人力资本的提升。因此,是人力资本,而不是新教的思想伦理,推动了经济的增长。^②

下面,我们结合贝克尔和沃斯曼的研究和其他的量化历史研究,对量化历史研究的步骤进行详细说明。

(一) 提出问题与假说

量化历史研究的第一步是提出问题和假说,在介绍这些之前,需要先了解什么是变量和度量。

1. 变量与度量

变量是指可能变化的对象。比如每个地区的气温、降雨量,每个地区的人均收入水平等。这些变量与定性研究中的“影响因素”“结果”等概念接近,都是从对人类社会的观察中抽取出来的考察对象,为了表述方便,称之为变量。

在对韦伯假说的检验中,贝克尔和沃斯曼发现,普鲁士地区的教徒中大概有2/3是新教徒,1/3是天主教徒,每个地区新教徒数据之间存在较大的差别,因而可以用来检验韦伯假说。而在识字率方面,普鲁士的452个县级政区,差别也很大,比例跨度从37.4%到99.3%。

给变量赋予一个数字,就是变量的度量。有些同时还需要一个计量单位。比如,温度是多少摄氏度,身高是多少米。这里包含三个要素,即度量的对象(即变量本身),度量的数值和度量的单位。清晰的度量对于量化历史研究非常重要。

就度量的类型而言,可以分为直接度量和代理度量。直接度量是指度量的指标与要度量的变量之间是一致的,比如粮价,是多少两白银一石,这样的度量就是直接的,这要求变量本身的含义是确定的,能够找到一个单一的值来衡量。

代理度量是指在没有直接度量的情况下,对关心的变量进行测量。用来度量的值与实际关心的变量之间往往不完全等同。代理变量(proxy variable)是与希望分析但无法度量的变量高度相关的变量。比如什么是新教伦理,可以从不同的角度去讨论,但从实证研究的角度,需要有一个度量,无法度量则难以验证。由于没有一个对新教伦理直接度量的指标,解决的办法是引入代理变量的概念。

^① 普鲁士是新教发源地,是韦伯观察新教伦理与资本主义发展的地区,也是韦伯的家乡,基于这一地区的历史考察韦伯命题再合适不过了。

^② Sascha O. Becker and Ludger Woessmann, "Was Weber Wrong? A Human Capital Theory of Protestant Economic History", *The Quarterly Journal of Economics*, Vol. 124, No. 2, 2009, pp. 531 - 596.

在韦伯假说中,贝克尔和沃斯曼用来衡量新教伦理的代理变量是一个地方新教徒的比例。这里隐含的假定是新教徒越多的地方,平均而言,新教伦理的影响应该越大。衡量经济发展水平的是一个地区的收入税。之所以用收入税,是因为该税种基于一个人一年的收入水平来征收。因此,人均收入税越高的地区,平均而言,人均收入也越高。

代理度量与直接度量之间可能存在差别,所以需要通过一些办法来验证代理度量的可靠性。验证的办法是使用一些其他指标作为代理度量。比如,除了人均收入税的多少,贝克尔和沃斯曼还选择了男性小学教师的平均工资和当地非农就业的百分比作为经济发展的代理变量。如果不同的代理度量指标都指向同一个类似的结果,结论就更可信。

2. 问题和假说

在前面提到的韦伯假说中研究的问题是,新教伦理是否导致了资本主义的增长。相对于研究问题,假说则更进一步,认为新教伦理导致了资本主义增长。假说是对人类行为和社会规律的一个猜测(猜想)。这个猜测是对关心的变量关系的一个明确表述。比如儒家文化阻碍了创新,儒家文化减少了社会冲突,大运河导致商业革命,等等。由于是假说,其可能会被数据所证伪。

上面的假说,基本上都涉及两个变量,比如,韦伯假说中,涉及新教伦理和资本主义经济发展。与社会科学中经验研究一样,在量化历史研究中,把希望解释的对象叫做被解释变量、因变量或Y变量,把用来解释的对象叫做解释变量、自变量或X变量。下面为了表述方便,分别称为被解释变量和解释变量。韦伯假说中的被解释变量是资本主义经济发展,解释变量是新教伦理。影响资本主义发展的因素很多,这些因素和新教伦理共同影响了资本主义发展,为了发现新教伦理这一单一因素的作用,需要把其他影响因素的效果也揭示出来,排除出去。这些其他的影响因素,一般称之为控制变量。

假说被提出之后,研究的路线图也就形成了。一个假说需要证明解释变量如何影响被解释变量,因此首先需要对被解释变量和解释变量进行度量,然后通过量化方法寻求两者之间的关系,以及对这种关系进行解释。

一篇量化历史研究的论文,通常会关注一个被解释变量和一个解释变量,这样做的目的是为了论述更为集中,重点去探索两个变量之间的因果关系。^①这样做,并非是简化历史,而是分析复杂历史的可行方式。一个研究中的被解释变量,在其他研究中可能成为解释变量或控制变量。同样,一个解释变量可能成为被解释变量或控制变量。通过对同一个问题进行多个角度的量化分析,会有效增加对复杂历史的认识。这也是考虑到人的认知能力有限,为集中注意力而作的折中,这样做不是不考虑其他影响因素或解释变量,而是在分析中尽可能包括各种解释变量或控制变量,但在写作中有所侧重。

一个好的假说是研究质量的重要保障,好的假说需要含义明确、有一般性、可以被证明或证伪。含义明确包含两层含义,一是被解释变量和解释变量必须是具体的,能够找到一个指标来进行度量。比如什么是资本主义萌芽,就是一个模糊的概念,无法通过一个具体的指标进行度量。二是必须清晰说明二者的关系,即解释变量是如何影响被解释变量的,这也是下文将介绍的作用机制。

所谓一般性,是指假说不能只陈述个案。比如,旱灾导致了明朝的灭亡。由于明朝灭亡只发生

^① 对一些已有研究不足的探索性论文,通常会将不同的变量放在一起进行初步分析。

了一次,这个假说难以被验证,但是可以修改为一个一般化的假说,即旱灾发生频率越高的朝代,其存续的时间越短,或旱灾增加朝代崩溃的概率。一个假说可能来自于常识、来自于统计数据、来自于理论与其他研究或者来自于历史材料的描述。

即使在定性研究中,也存在大量的假说,只是由于研究者没有特意强调,不容易被人察觉。传统的历史研究,避免带着问题阅读史料,从而希望避免“以论带史”。那么量化历史研究先提出问题和假说,再去收集整理史料的好处是什么?

量化历史首先提出问题的的好处,一是可以方便跟已有文献的对话,确定自己的贡献在哪里。史料可能证实也可能证伪提出的问题或假说,只要史料的收集既全面又公允不偏,就不存在研究方法上的问题;如果证伪,已有的观点被推翻,更容易形成重要的发现。实际上,越是重要的假说,讨论的角度越多,一般也会有越多的竞争性假说。比如,朝代的灭亡,到底是因为旱灾、鼠疫,还是因为农民战争,等等。定性研究的特点在于,常常会有很多的假说在一起讨论,这帮助我们知道了更多的历史细节,但也不利于将复杂的历史条分缕析,将其中包含的规律一个个揭示出来。二是便于跟不同学科的交流,回答不同学科的重要问题。比如贝克尔和沃斯曼对韦伯假说的研究,虽然是讨论的普鲁士的情况,但其基本问题是文化如何影响经济发展,研究者可以在这一问题之下,讨论基督教在中国的发展对中国经济的影响,儒家文化对经济的影响,这不仅有利于在这些重大问题中提供中国经验、中国故事,也有利于知识的积累。三是可以基于要验证的假说,从被解释变量、解释变量和控制变量的角度,高效率地组织史料。史料如大海,如无问题视角,极易淹没其中。最后需要指出的是,量化历史在研究开始就提出问题和假说,但不是预设结论,而是猜测,这个猜测可能被史料证实,也可能被证伪。

(二) 理解史料与数据

在提出问题和假说之后,我们需要寻找史料来建立数据库,即需要找到被解释变量、解释变量、控制变量等度量指标和数据。

在拿到相关的史料和数据之后,量化方法并非像很多批评所说的那样,对数据和史料拿来就用,而是先从不同的维度对它们的质量进行检验。这里的检验,主要是基于历史史料学的基本要求,由于这方面已经有众多教科书,此处不做介绍。下面主要讨论从量化历史的角度,如何理解史料,重点是史料的代表性问题。

在历史研究中,拿到的史料往往是关心的研究对象的一部分,只能够基于拿到的史料去开展研究。尽管随着历史大数据时代的到来,有时候可以拿到关于研究对象的全部史料,但这样的情况还比较少见。基于研究对象的部分信息对研究对象的特征进行分析,是统计学研究的核心之一。这里就涉及总体、样本和随机抽样的概念。

在统计分析时,把研究对象的全体称为总体,每一个研究对象称为个体,把从总体中抽取的一部分个体称为样本。比如,如果研究中国历史上的皇帝,那么,所有皇帝就是一个总体。如果只使用清朝的皇帝,那么就是一个样本。

什么对象是总体,取决于研究问题。如果希望知道清代中国的命案率,如果是以省为分析单位,那么每个省在每一年的命案率就是一个总体。如果只分析其中的一个或几个省,就是样本。如果只分析乾隆朝的命案,也是一个样本。前者是基于空间、后者是基于时间的抽样。

对于历史研究而言,基于空间和时间的抽样是非常常见的。一个典型的历史研究常常会限定地

域和时期,比如清代中后期江南经济研究。对时间和地区做这样的限定,是为了方便集中论述,避免将不同时期、不同地区的史料混用。抽样方法有其特有的优势,也是目前历史研究中常用的方法,但从实证研究、发现规律的角度看,这样的研究方法却可能会带来选择性偏差,研究结果不一定带有普遍性。

选择性偏差是指由于选择的样本不是随机抽取的(比如,专挑对研究结论有利的时期和地区),所以不能够基于这些样本的特征推断总体的特征。随机抽取是指保证总体中每个样本被抽中的概率都是一样的。例如,不能够基于江南的经济发展水平去推断整个中国的经济发展水平。这样的选择性偏差非常明显,也很容易被察觉,但历史研究中,还有大量不容易察觉到的选择性偏差问题。例如传统经济史研究中的“选精”“集粹”问题,其实质就是一个基于样本推断总体的问题。^①

具体而言,如果我们希望知道北宋整个朝代的年均亩产量,那么北宋所有土地上每一年的亩产量就是一个总体。由于不知道这个总体的规模有多大,我们不妨做一个假定,假设北宋耕地有7亿亩,平均每块耕地面积为70亩,那么耕地数量就有1千万块,再假设是一年一熟制,北宋总共存续了167年,那么理论上,北宋亩产量的观察值应该有16.7亿个,因此要精确地知道宋代的亩产量这个总体几乎不可能。如果研究者拿到了16700个北宋亩产量的历史记载,那么这1万多个亩产量占全体亩产量的比例约为十万分之一。我们自然要问,这十万分之一的样本能否推断总体呢?

这样的挑战不仅仅在经济史,在其他历史学分支同样会出现。比如,如果想知道民国时期中国人的民族观念,假设只考虑16岁以上的人,那么总体数量可能在两亿人以上。如果拿到了2000个当时中国民族观念的调查或者其他记载,能否基于这2000人来推断两亿中国人的民族观念?这里还需要假定一个人在一生中的民族观念变化不大。

由于研究者拿到的史料通常是样本,但希望了解的是总体,如何由样本推断总体就成为历史研究的挑战。这种挑战在定性和定量研究中都存在,尽管定性研究中不用这套术语,但不代表这个挑战不存在。定量研究的好处在于,可以明确地展示出这种挑战,同时利用已有的量化方法克服这种挑战。

基于总体和样本的理解,可以知道,历史大数据或增加数据量,并不必然会解决“选精”与“集粹”的问题。利用样本信息来推断总体的信息,是统计学中统计推断(statistical inference)的重要内容。应对选择性偏差,基于非随机抽样的样本来推断总体的特征,是量化分析方法特别是计量经济学分析的核心关注之一。

(三) 相关分析和因果分析

在提出问题和收集数据之后,第三步是量化分析。量化分析方法主要有三类。第一是对数据进行描述统计,主要包括数据的来源、处理过程,观察值也即样本量的多少,每个变量的统计特征,比如平均值、方差、最小值和最大值等。描述统计主要针对单个变量进行分析。这一方法比较简单,在历史学研究中较常见,不赘述。第二是相关性分析,主要分析方法包括画散点图、画地图、计算相关系数等。这一方法主要用于分析两个变量之间的关系。第三是因果关系分析,通常使用多元回归分析的方法,重点是处理内生性问题和发现作用机制。

1. 相关关系

当我们提出假说时,实际上就提出了一个相关关系,也就是两个变量的数值之间的变动关系。

^① 李伯重《“选精”、“集粹”与“宋代江南农业革命”——对传统经济史研究方法的检讨》,《中国社会科学》2000年第1期。

相关关系分为三种,分别指正相关、负相关和不相关。

正相关是指两个变量朝相同的方向变动,比如新教徒数量越多的地区,经济发展水平越高。负相关是指两个变量往相反的方向变动,比如儒家文化越发达的地方,发生暴力冲突的次数会更少。不相关是指两个变量的变动没有关联。韦伯在《新教伦理与资本主义精神》中开篇就提到他所发现的相关性。这种相关性构成了这一专著的起点和基石。展示相关关系的方式主要是画散点图、画地图和计算相关系数。散点图是用二维坐标展示同一个观察对象两种不同特征度量值关系的图。横轴标示一个变量的数值,纵轴标示另一个变量的数值。每一组数值对应散点图中的一个点,点的位置由两个变量的数值决定。通常,自变量标示在横轴,因变量标示在纵轴。如果两个变量之间不是解释与被解释的关系,可以将变量标记在任意坐标轴。画地图的方法,使得相关性在空间上的分布很直观地展示出来。从贝克尔和沃斯曼的文章所画的地图中,可以看到新教徒数量越多的地方,人均收入税确实越多。当然,这也从一个角度说明,如果只考虑某些特定的区域,比如普鲁士东北部地区,结果可能会是完全相反的。这正是考察大样本和考察个案所带来的差别。散点图和地图图示方法可以直观展示两个变量的相关关系,但并不精确,且也难以对不同组变量之间的相关性进行比较。解决办法是计算相关系数。相关系数描述两个数值变量之间线性相关关系的方向和强度。数值介于-1到1之间,符号的正负表示正相关和负相关,等于0时表示不相关。相关系数绝对值越大,表示相关性越强,反之,则越小。

2. 因果关系

韦伯显然没有满足于只是发现新教徒数量与经济发展之间的相关性,而是继续追寻背后的因果关系。

因果关系的基本含义是改变一个变量的值,就可以使另一个变量的值改变。由于与相关关系类似,两个变量都有变动关系,人们常常将相关关系误读为因果关系。但相关关系不等于因果关系;相关关系中可能隐含有因果关系,也可能没有,一般需要进一步证明。在相关关系中,改变其中一个变量的数值不一定会引起另一个变量数值的改变。因果关系是指若一个事情发生,另外一个事情必然发生。当然,有一个重要的前提条件是给定其他条件不变。这一思想最早来自约翰·穆勒(John Stuart Mill)的差异法(Method of Difference),即通过“比较某现象出现的场合和不出现的场合,如果这两个场合除一点不同外,其他情况都相同,那么这个不同点就是这个现象的原因,两种情况的差异就是这个原因的因果效应”。^①

学界早已注意到,在定性研究中,本身就隐含了众多的因果推断,甚至几乎在每一页上都有一个隐含的函数关系,即我们这里讨论的因果关系。比如“春秋战国之际,随着铁农具和耕牛的使用日渐广泛,农业产量不断提高;每个王朝中后期,随着土地兼并现象日趋剧烈,农民阶级与地主阶级之间的矛盾日益尖锐”^②——尽管传统史学者不做基于大样本的统计分析、检验(也不用量化历史研究的术语),但这些典型的史学论断中,包含了许多因果关系结论,或者说,只要有这种论断,就是在下因果关系结论。由于通常情况下,历史学家没有明确意识到其中包含的某种函数关系,更难于想到在可能的条件下用量化方法检验。这使得这些论断,即使经过一定史实检验,陈述的时候也显得

^① Paul W Holland, "Statistics and Causal Inference", *Journal of the American statistical Association*, Vol. 81, No. 396, 1986, pp. 945 - 960. 这段话的翻译转引自赵西亮《也谈经济学经验研究的“可信性革命”》,《经济资料译丛》2017年第2期。

^② 庞卓恒主编《西方新史学述评》,高等教育出版社1992年版,第404页。

缺乏说服力,或者仅凭直觉或少量个案做出,经不住计量检验。^①

导致相关关系不等于因果关系的问题,一般也称为内生性(endogeneity)问题。造成内生性问题的主要原因,一种是遗漏变量,一种是反向因果。遗漏变量是指有一个变量在研究过程中没有被考虑到,一旦这个变量被考虑进来,就能够解释为什么之前观察的两个变量之间存在相关关系。比如在韦伯假说中,一旦考虑到人力资本这个变量,就会看到新教徒数量与经济发展水平之间没有了相关关系。这时候的人力资本就是一个遗漏变量(当然,因为新教鼓励教徒自读《圣经》,所以每个新教徒需要从小读书,因此人力资本更高)。反向因果是指因果关系的影响方向与预判的是相反的,或者存在互为因果的情况。比如,在韦伯假说中,就可能存在经济发展潜力越大的地区,越可能选择新教。这时候就不能得出结论说是新教徒越多的地方,经济越发展。

既然因果关系是研究的目的,因果推断在定性研究中也 very 常见(只是定性研究很难处理好因果关系的论断问题),那么,如何来进行因果推断呢?我们首先介绍自然科学和医学中进行因果推断的方法,接着介绍量化历史研究中使用的方法。

在自然科学中,可以通过重复试验来解决这一挑战。通过将影响实验结果的其他因素都控制住,考察一个因素变动带来的影响。在医学中,一般要通过大样本双盲对照试验发现因果关系。其背后的原理就是实验方法。以医学中对药物有效性的验证为例,在什么情况下,才能够说一种药物对于某种疾病有效?在新型冠状病毒肺炎治疗案例中,一种药物对某位美国患者有明显的疗效,是否就可以下判断说这种药物是有效的呢?这里面至少存在两个问题需要解决,首先是这个患者不仅仅只服用了这种药物,可能也服用了其他的药物(包括不同饮食),或者是使用了其他的治疗方式,到底是哪一种药物或治疗方式起作用,需要区分开,但是我们只能观察到多个因素共同起作用的结果。其次是某些疾病即使是不吃药,有些人也能自愈,因而就有可能遗漏掉患者会自愈这个影响因素。为了证明某种药物的作用,通常需要做大规模随机对照试验。

历史不能够做试验,但研究者发现,在某些情况下,历史可以很好地充当实验室。在量化历史中,一般称之为(准)自然试验,即依靠历史事件模拟随机对照试验,将历史对象分为对照组和处理组。贝克尔和沃斯曼对于韦伯假说的检验,就是利用准自然试验的方法,这也是目前量化历史研究中最常用的方法。^②

在经济学、政治学等社会科学领域,经验研究正在经历研究范式的转变,即从统计推断向因果推断转变。这也被认为是经济学经验研究的“可信性革命”。^③可信度革命是指在经验研究中,基于随机化试验的思想,通过良好的识别策略,发现因果关系。基于历史自然实验的思路,实证研究方法有工具变量法、差分方法、断点回归方法等。^④

3. 回归分析

上述发现因果关系方法的实现,都需要通过回归分析。回归分析是一种统计方法,通过建立统

① 庞卓恒主编《西方新史学述评》,第404—405页。

② 关于历史自然试验的介绍,参见 Davide Cantoni and Noam Yuchtman, “Historical Natural Experiments: Bridging Economics and Economic History”, *NBER Working Paper Series*, February 2020, 26754; Jared Diamond and James A. Robinson, eds., *Natural Experiments of History*, Harvard University Press, 2010.

③ Joshua D. Angrist and Jörn-Steffen Pischke, “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics”, pp. 3–30.

④ 这些方法在常见的高级计量经济学教材中都有详细介绍,限于篇幅,在此不展开介绍。

计模型,用一个或多个解释变量来解释被解释变量。回归分析可帮助人们理解,当任意一个解释变量发生变化而其他变量保持固定时,被解释变量将如何变化。

回归分析是展示变量相关性的一种重要方法,同时也是因果推断的基础和最重要的工具之一。回归分析与相关分析不一样,回归需要选择解释变量与被解释变量,而相关系数不需要。回归分析可以处理两个以上变量的相互关系,一个被解释变量和多个解释变量,以及多个控制变量。

由于历史现象通常是多个因素共同作用的结果,也就是说,通常涉及两个以上的变量(即历史现象是一个多变量函数)。因此,在量化历史研究中,多元回归分析是最为常用的方法。多元回归分析能够帮助我们在保持其他影响因素不变的情况下,考察一个特别的因素对于历史结果的影响,即特定的自变量对因变量的影响。

还是以韦伯假说为例,我们能够观察到的是不同地区的经济发展水平,但影响经济发展水平的因素很多,基于目前经济增长理论的研究,直接原因有技术、物质资本、人力资本等,而根本原因则有制度、地理、文化、土质、产业结构、海外关系,等等。这么多的影响因素共同作用,影响了一个地区的经济发展。我们如何将新教伦理的因素与其他因素分离出来呢?如果我们观察到一个县(郡),新教徒的数量很多,但经济发展较慢,能否就说新教伦理对于经济增长起到了反向的作用,即新教徒数量越多,经济发展越差呢?或者没有影响呢?这里面可能遗漏掉另外的因素,比如存在一个因素,使得经济发展变慢,这时候新教伦理的作用就被掩盖掉了。具体来说,假如新教伦理平均可以让经济增长快1%,但另外一个因素可以让经济增长慢2%,最后,我们虽然看到一个地区新教徒数量很多,但是经济增长可能是负的,出现这样的问题,正是因为遗漏了重要的其他影响因素造成的。多元回归分析正是将重要的影响因素都尽可能考虑到并将不同因素的影响剥离开的方法。

吴承明很早就介绍了回归分析方法在经济史研究中的应用,^①但当时的回归分析基本是用手计算,差不多只能做单变量回归,样本数量不能太大,所以,回归分析的作用不大。随着计量经济学的兴起,尤其是计算机速度和成本的变化,今天的回归分析已经变得非常发达。回归分析成为社会科学中的主流分析方法,在人文学科中的使用也越来越多。

随着量化分析方法的不断改进,针对不同的数据类型,不同的研究设计,分别有对应的回归分析方法来处理。对这些方法的详细介绍,最好是参考计量经济学或统计学的专门书籍。

一项量化历史研究通常需要进行很多的回归分析,贝克尔和沃斯曼文章中的回归表格就有十个。这是因为需要将不同的影响因素考虑进来,需要考察是否有遗漏变量、反向因果的问题,以及对结果的可靠性程度,也要从不同的维度进行检验。

(四) 作用机制分析

通过上面步骤建立因果关系后,我们还需要知道为什么这种因果关系会发生,也就是解释变量是如何影响被解释变量的。

机制是指解释变量如何影响被解释变量,其产生影响的原因是什么。在对“韦伯假说”的论证中,贝克尔和沃斯曼发现新教徒数量与资本主义经济增长之间是正相关的关系,即新教徒数量越多

^① 吴承明《中国经济史研究的方法论问题》,《中国经济史研究》1992年第1期。

的地方 经济增长越快 其中起到作用的机制是新教徒的识字率更高 而不是新教伦理。

机制研究是量化分析的核心组成部分 其价值在于揭示发现的因果关系是如何起作用的 其原理在哪里 也是对因果关系的进一步支撑。

正如当我们知道某种药能够治疗某种疾病时 我们还需要知道这种药是如何治疗这种疾病的。比如青蒿素能够治疗疟疾 但是到底是如何起作用的。今天的药物研究 需要说明某种药物的有效成分是如何在分子层面起作用的 即具体是如何帮助治疗疾病的。以2020年新型肺炎为例 需要说明药物的有效成分作用在病毒的哪个成分上面 让病毒难以复制或死亡。

与此类似 我们对于在人类社会中发现的因果关系 也需要知道其作用机制是什么。作用机制的途径 通常也是来自于理论、其他学者的研究、常识与史料 在研究过程中 通常会出现多种可能的作用机制 因而需要对每一种可能的作用机制进行验证。验证的方法 仍然是依靠多元回归分析。

对于作用机制的分析 一般需要从三个方面进行。第一 起作用的机制变量。比如人力资本必须与解释变量高度相关;比如 新教伦理越高的地方 人力资本应该越高。第二 需要说明机制变量与被解释变量高度相关 即人力资本越多的地方 经济发展状况更好。第三 需要说明 当同时考虑作用机制变量和解释变量之后 解释变量的解释力变小或消失 即系数在统计上显著性下降或不显著;而机制变量有很强的解释力 即系数统计上显著。在回归分析中 就是显著性下降或消失。也就是说 当不同的解释出现的时候 最好的办法是使用“赛马模式” 将不同的解释变量放到同一个多元回归模型中进行分析。贝克尔和沃斯曼通过将识字率加入到回归分析中 发现新教徒数量的系数在统计上不再显著 表明新教徒数量对于经济发展的影响消失。而这时识字率的系数显著为正 表明新教徒占比对于经济发展的影响主要来自于识字率。

三、量化历史作为新史学的价值

量化历史是新史学的重要组成部分 陈志武在之前的研究中已经从新知识革命的角度介绍了量化历史对于知识创新的价值 通过不同案例说明量化历史不仅仅是验证已有的常识 也带来新知识、新认知。^① 本文所引用的案例也表明量化方法在知识创造、解决争议、纠正错误认识等方面的优势。接下来 笔者侧重从历史学研究面临的挑战出发 介绍量化历史作为新史学对于历史学研究的价值。

基于前文提出的量化历史研究步骤和方法 我们认为量化方法对历史学的价值主要体现在如下方面。

(一) 应对历史大数据的挑战 帮助分析复杂的历史

历史学建立在史料的基础上 随着大规模史料的出版、公开、数据库化 研究者已经注意到 历史大数据时代和“数字人文时代已经到来”。^②

历史大数据包含两层含义。第一是接近全样本 也就是可以掌握研究对象的总体。比如关于清代官员的信息 由于有缙绅录和地方志 清代中后期绝大多数官员的信息都能够得到还原。根据前述样本与总体关系 随着历史学家掌握总体 统计推断带来的挑战会下降 但因果推断的挑战依然存

^① 陈志武《量化历史研究与新知识革命:以财富差距与消费差距的历史研究为例》,《北京大学学报》2018年第4期。

^② 黄兴涛《当代中国历史学的时代使命》,《历史研究》2019年第1期。

在,需要应用最新的量化方法来应对。

第二是数据量巨大,这又表现在两个方面,首先是由于引入定性变量和代理变量,可以将海量的历史资料转为数据。比如,仅清代刑科题本档案就有近65万本。不仅仅是清代,即使是很早以前的历史,也蕴藏着海量的数据。以考古墓葬为例,仅仅是香港大学“中国考古数据库”所收录的先秦已经发掘的遗址数量,就接近6万个。

其次是数据之间的关联。数据库的潜能不仅仅是在每一个单独的数据库本身,更为重要的是不同数据库之间的关联和匹配。举例来说,可以将缙绅录中近500万个官员的记录与《清实录》、清代档案中的官员行为记录匹配起来。缙绅录数据本身就是历史大数据,如果与官员的行为记录匹配起来,其数据量将达到数亿级别。在相关数据可得的情况下,依靠计算机,这样的匹配可以在很短的时间内就完成。当越来越多的信息匹配起来之后,我们有机会建立起关于复杂历史的更为真实的图景。

历史大数据出现之后,使用定性方法显然难以将这么多的数据做一个整体分析,从中得出经验性的认识,但量化分析方法却可以较好地应对这种挑战。因为数据量的加大很大程度上只是对计算机计算能力提出了更高的要求,而量化分析的基本原理没有发生大的变化。量化分析中的多元回归分析可以有效应对数据库匹配之后带来的多变量分析的挑战。实际上,本文介绍的量化历史研究,基本都是在对多个历史数据进行匹配的基础上开展的。

量化方法在应对历史大数据挑战的同时,也有助于分析复杂的历史。历史的复杂性体现在很多方面,梁启超早就注意到“然因果关系至复贖而难理,一果或出数因,一因或产数果,或潜伏而易代乃显,或反动而别证始明,故史家以为难焉。”^①

量化方法是处理复杂性的有效方法,其办法主要是将历史现象拆分成不同的因果关系。比如,新教改革无疑是非常复杂的历史现象。量化历史对新教改革的研究,大体可以分为两组研究,一是新教改革的原因,二是新教改革的后果,包括短期和长期的后果等。本文介绍的贝克尔和沃斯曼是研究新教改革后果的量化历史研究之一,贝克尔等学者对新教改革研究提供了一个详细的综述,涉及量化历史研究达50多项。^②

上述研究的特点均是使用历史大数据,分析因果关系。看起来每篇文章的结论都可以用一两句话进行概括,但其结论经过了上述四个量化分析步骤,可靠性程度大大提升,也大大增加了我们对新教改革这一复杂历史的认识。

(二) 识别历史的长期影响,形成贯通性认识

历史的长期影响,^③是人们关心历史的重要原因。今天很多重要的社会结构性特征,都与历史有关。比如,儒家文化对今天中国人的影响。那如何识别出历史的影响呢?实际上,在量化历史研究中,这是很大一批文献,一般称之为“遗产(legacy)”研究。

一个例子是科举制度的长期影响。科举制度在中国持续了1300多年,1905年被废除,这样的制度对于今天的中国人是否产生了持续的影响,如果有影响,是通过什么渠道产生的?陈婷(Chen Ting)等

① 梁启超《中国历史研究法 中国历史研究法补编》四川人民出版社2018年版,“自序”第7页。

② Sascha O. Becker, Steven Pfaff, and Jared Rubin, “Causes and Consequences of the Protestant Reformation”, *Explorations in Economic History*, Vol. 62, 2016, pp. 1–25. 中文介绍参见贺崑崑《宗教改革的前因后果》,“量化历史研究”微信公众号第358篇推送。
<https://mp.weixin.qq.com/s/fYkKeMvINiDpsJWluBL3w> [2019-09-27]

③ Nathan Nunn, “The Importance of History for Economic Development” *Annual Review of Economics*, Vol. 1, No. 1, 2009, pp. 65–92.

结合历史上的科举数据、今天的统计数据 and 大量微观调查数据 对此进行了分析 他们发现 科举制度即使废除了 那些历史上出进士较多的地区 今天平均的受教育年限更长 对教育的重视程度更高。具体来说 在排除其他影响因素的情况下 明清时期一个府每万人中每多增加一个进士 到 2010 年时人均受教育年限要多增加 0.7 年。之所以产生这样的影响 主要是由于对教育观念的重视。^①

量化方法对历史长期影响的分析 也有助于形成贯通性的认识。历史研究强调“通古今之变 成一家之言” 通史的训练是历史学中非常重要和基础的部分。历史学最重要的特征之一是历史的时间性 即考虑人类社会长时段的规律。

尽管强调时间性和打通断代是历史学研究的重要诉求 但中国悠久的历史留存下来的史料汗牛充栋 要从中对某些历史特征进行长时段的描述 具有挑战性。这种挑战 梁启超有过生动的表述 他说“中国历史可读耶？二十四史、两《通鉴》、九通、五纪事本末 乃至其他别史、杂史等 都计不下数万卷 幼童习焉 白首而不能殫 在昔犹苦之 况于百学待治之今日 学子精力能有几者？”^②

这种困难 通过简单的量化分析 就可以得到缓解。基于时间序列数据和面板数据^③ 不仅可以快速地对历史时期非常重要的历史事件进行长时段的描述 与世界其他地区已有的研究进行对比 更重要的是 这有助于改善知识创造和积累的方式 以后的学者 将可以在此基础上做进一步的推进 而不需要从头开始阅读史料。这方面的早期探索包括竺可桢对五千年气候变化的研究^④ 《历代战争年表》^⑤ 等。在此基础上 如何建立起气候变化与战争之间的因果关系 则是量化方法擅长的事情。这正是龚启圣和白营的工作 他们研究了游牧民族在什么情况下更有可能攻打中原。^⑥

(三) 推动历史学与社会科学的交流与对话

史学被认为是一切社会科学的基础 应该成为社会科学理论创新的源泉 但从各个社会科学的实际情况看 史学还没有起到应有的作用。

量化历史直接从问题和假说出发开始研究 这些问题和假说也是不同社会科学关注的问题。如果假说得到证实 由于历史提供的实验室 使得理论被接受的可能性和可靠性得到增强 基于中国历史的例子也可以很好地融入社会科学理论之中。如果假说和理论被证伪 则会更好地推动理论的修改 形成重要的理论创新 这正是社会科学真实的发展过程。当然 历史的作用在这里 不仅仅是社会科学理论的实验室 对于理论的证实和证伪 可以从不同的角度增进对历史的认识。比如 诺贝尔经济学奖得主米尔顿·弗里德曼等人通过对 1867—1960 年美国货币史的量化历史研究 推导出了著名的货币层次理论及货币供应理论 是对经济学理论的重要贡献。^⑦

① Ting Chen, James Kai-sing Kung, and Chicheng Ma, “Long Live Keju! The Persistent Effects of China’s Civil Examination System”, *The Economic Journal*, Vol. 130, No. 631, 2020, pp. 2030–2064.

② 梁启超《中国历史研究法 中国历史研究法补编》,“自序”第 7 页。

③ 时间序列数据由一个或多个变量在不同时间点的观测值构成 比如历史上每 50 年的人口数量。一组观察对象在不同时间的观测值汇集一起 就构成了面板数据 比如历史上每个府每月的粮价。

④ 竺可桢《中国近五千年来气候变迁的初步研究》,《考古学报》1972 年第 1 期。关于这一历史时期气候变化数据的重建 参见葛全胜、方修琦、郑景云《中国历史时期温度变化特征的新认识——纪念竺可桢〈中国过去五千年温度变化初步研究〉发表 30 周年》,《地理科学进展》2002 年第 4 期。

⑤ 中国军事史编写组《中国历代战争年表》,中国人民解放军出版社 2003 年版。

⑥ Ying Bai and James Kai-sing Kung, “Climate Shocks and Sino-nomadic Conflict”, *Review of Economics and Statistics*, Vol. 93, No. 3, 2011, pp. 970–981.

⑦ 陈争平《大数据时代与经济史计量研究》,《中国经济史研究》2016 年第 6 期。

结 语

随着历史大数据时代的到来,如何高效率地处理大规模史料并从中获得规律性认识,是当代历史学面临的新挑战。量化方法经过数十年的发展完善,已经在应对大规模数据库、发现因果关系方面走在了前面。将量化分析方法和历史大数据结合起来,是新史学的重要内容,也是一种必然趋势。本文对典型量化历史研究的步骤进行了详细说明,并介绍了其在应对历史学挑战方面的价值。除此之外,量化方法的类型还很多,留待将来再做介绍。

强调量化历史研究的优势,并非意味着这些优势能够自动实现、或者很快就能够实现,一项好的量化历史研究需要很多条件的配合,也需要大量坚实的工作。而量化历史研究作为一个新兴的领域,仍然处于不断完善的过程之中。

在使用量化历史研究方法的过程中,也需要注意其适用的条件,任何一种方法都有其适用的范围和局限,一项研究的发展也需要学术共同体的监督和批评。量化方法作为“史无定法”中方法的一种,在历史大数据时代,其作用将越来越大。

(作者林展,中国人民大学清史研究所讲师;邮编:100872;陈志武,香港大学经济管理学院与亚洲环球研究所教授;邮编:999077)

(责任编辑:尹媛萍)

(责任校对:敖 凯)

《近代以来中外史学交流史》

张广智主编,复旦大学出版社2020年版。全书分为上、下两编。上编为域外史学在中国,主要阐述近代以来域外史学(主要为西方史学)传入中国及其产生的影响;下编为中国史学在域外,主要叙述中国史学输出域外及其对该国或该地区史学所产生的影响。全书史料宏富,论证翔实,清晰、具体地展示了近代以来中外史学交流的历史进程,对于了解中外史学交流的脉络、特点与趋向,乃至中外文明之间的对话与互鉴大有裨益。该书的出版,为中外史学史研究,尤其为构建中外史学交流史的学科体系、学术体系与话语体系做出了开创性的贡献。

SUMMARY OF ARTICLES

In the Spirit of Science in Dialogue with History: An Introduction to President Xi Jinping's Important Discourse on Historical Science // Yu Huamin

President Xi Jinping's important discourse on historical science points out an organic method to ground classical theories of historical materialism within China's history and culture through the active engagement of building the socialism with Chinese characteristics. It offers guiding theories and practical methods as well as profound reflection and insightful illumination by which to position historical science as a discipline to scientifically discover wisdom from history, creatively generating a series of new ideas, new hypotheses, and new points of view. President Xi's important discourse on historical science has broadened the scope of historical understanding and constantly generated new ideas. It is an embodiment of the true power of Marxism as well as of the contemporary Chinese Communists' ambitious global aspirations and their uncompromising sense of duty for history. It has a guiding significance in developing the studies of historical science in the new era in China. To study President Xi Jinping's important discourse on historical science, we need to carefully grasp its internal logic, historical reasoning, and scientific spirit, thus to discover the principles of history and to learn from the past facing the future.

Quantitative History as New History: Its Methods and Value // Lin Zhan, Chen Zhiwu

In the era of big data, historical research faces a new challenge, that is, how to make better use of the large-scale historical materials given in a large number of historical databases which have been established and become accessible to the public. Quantitative history, as a critical branch of the new history, is one of the important strategies by which to deal with this challenge. In recent years, as a burgeoning field in international academia, quantitative history has developed sophisticated analytical methods. Taking the study of *The Protestant Ethic and the Spirit of Capitalism* as an example, this article details the research process of quantitative history and analyzes its potential positive role in coping with the challenge of historical big data. It identifies long-term impact on historical research, and promotes exchange and dialogue between the history discipline and other subjects in social sciences.

On the Paradigm of Early Chinese Historiography: Meng Wentong as a Case Study // Chen Yong, Gong Chen

In the first half of the twentieth century, Chinese historiography transformed from tradition to modernity. One of its central themes is how to reorient Chinese historiography. This question gradually