

• 社会学研究 •

数据生产和数据造假: 基于社会学视角的分析

刘能, 马俊男

(北京大学社会学系, 北京 100871)

摘要: 本文从数据生产的过程入手, 追溯统计数据造假现象的缘由。首先, 误差扰动、接近值模拟和数据质量的资源投入水平, 解释了数据生产过程中的数据失真。其次, 从统计作为一项国家事业发展演进的角度来说, 国家能力和国家意图有时会影响统计数据的真实性和全面性, 并引发出制度诱导的造假行为。第三, 从个体层面的理性选择过程来看, 当违法收益显著高于违法成本时, 数据造假现象就成了经济领域中最常见的越轨行为之一。治理数据造假的潜力也和个体的动机和理性计算直接相关: 当我们把纯粹物质性的理性计算模型扩展到纳入社会成本和群体动力机制之后将会发现, 社会交往网络和同伴群体内部伦理声誉的受损, 将会比物质违法成本的提高更能约束数据造假行为。最后, 本文还讨论了数据造假问题的直接伦理损失和技术在治理数据造假过程中所扮演角色的两面性。

关键词: 数据生产; 数据造假; 国家统计; 伦理损失; 技术及其两面性

中图分类号: C915 **文献标识码:** A **文章编号:** 1009-8860(2019)03-0062-08

引言

进入现代以来, 随着统计科学的发展, 统计数据越来越成为人们决策的基础信息。然而, 与统计学学科日益专业性和科学性的发展相背离的是, 统计数据的真实性不断受到质疑, 数据造假和数据失真现象层出不穷。尤其是随着互联网的兴起, 数据失真现象已经超出以往经常讨论的官方统计数据的范畴, 变成了社会多个领域内不得不直面的重要问题。除了已被学界讨论过的政府数据造假、企业数据造假、海关数

据造假、经济数据造假、生育数据造假、收视率数据造假、直播数据造假之外, 还出现了水军刷榜、视频点击率刷榜、淘宝刷交易好评等社会上热议的数据造假现象。围绕这一现象, 以往学界的讨论更多地关注于统计数据的技术层面和统计主体的互动层面, 强调统计指标选择的价值有涉、利益博弈和统计机构的(非)独立性问题, 对于统计数据所引发的公共性议题的关注不够。实际上, 统计数字并不是对现实的一种价值中立式的反映, 而是与多样性价值意涵紧密联系在一起的社会产物。初看起来, 统计数据似乎是对经验现象的无偏测度和描述, 但

收稿日期: 2019-01-05

作者简介: 刘能(1970—), 男, 浙江嵊泗人, 北京大学社会学系副主任、教授、博士生导师, 北京大学中国社会与发展研究中心副主任, 主要研究方向为城市社会学、社会变迁和社会问题; 马俊男, 北京大学社会学系硕士研究生。

事实上,统计数据及其生产过程却与当代社会科学的一系列重要概念和议题——如社会建构、象征符号、认知方式、社会规范、权力控制、行动方式、社会叙事等——紧密勾连^[1]。原先人们所秉持的“让数字说话”的技术观念已显得落伍,因为统计数字早已超出了价值无涉的技术应用范畴而与众多社会公共议题相联系。与此同时,围绕数据造假的单一理论解释,如援用针对微观决策行为而言解释力度较强的博弈论模型所进行的讨论,虽然的确具有一定的启发意义,但这些讨论不但无法解释数据造假之风为何屡禁不止且进而向社会其他领域迅速蔓延,也无法解释整个社会对于数据失真现象所持的漠然态度。从社会学的学科视角来看,造成数据失真或数据造假现象的原因很多:除了微观层面个体行动者各异的行动动机之外,也还要考虑到中观组织和制度因素的影响,以及宏观文化心智层面的形塑和诱导。具体来说,表面上的数据失真和数据造假,其背后却是经过复杂考量的个体和组织选择,既有理性的计算和权衡,又有晋升预期的考虑,还有文化伦理上的认知架构。此外,数据生产的过程又是一个日益专业化的过程,包括技术因素和质量意识在内的方法论因素,都会对数据质量本身造成直接的影响。接下来,我们将依次从数据采集过程、国家权力与数据生产之间的关系,以及数据生产背后的文化价值因素等多个方面出发,对数据失真/造假现象展开具体讨论。

一、方法论层面的解释

方法论的层面看,数据采集过程本身要遭受多个来源的误差的侵袭,而对误差的控制不力,本身就会造成数据失真的结果^[2]。如从问卷调查的角度来看,效度问题是研究团队所能加以把控的,而信度问题则在很大程度上取决于回答者真实回答问题的认知和意愿。其次,数据质量的高低,既和数据生产单位的职业伦理氛围有关,又和数据生产的技术程序得到执行的水平,以及数据生产的资源投入水平有很大的关系。第三,数据失真也许本身就是一个无法克服的本体论问题:除了所谓的“测不准”

准则之外,一个更令人沮丧的事实是,“真值”是随时间而变动的,而数据的采集和利用过程却是截面的,故而我们只能去追求截面时点的最优接近值,将其作为事实上的“真值”看待。此外,流传在数据采集行业的一个最经常被提到的俗语便是,“没有免费的数据”;高质量的数据需要高密度的资源(人力、物力和时间)投入。因此,在一个不尊重专业伦理和专业程序,不尊重数据采集的质量原则和成本投入原则,短期化、功利化行为盛行,盲目追求表面绩效的社会,数据失真现象就屡见不鲜了。当然,本文的重点不在讨论这些潜藏于数据生产过程之中的、导致数据失真或数据造假的原因,而毋宁将它作为一个隐含的背景知识,更多地把焦点投注在导致中国社会中数据造假或失真现象的文化的、制度的、个体动机的和技术层面的多重缘由。

二、作为政治统治和技术治理手段的数据统计

体系化的统计活动起源于城邦、王国或帝国之类的政治主权行动者的诞生,出于征兵、收税和救灾等治理需要,有必要对国民活动的方方面面进行持续有效的计数和记录^[3]。统计学作为一种国家治理术,本身带有政治的烙印,由于其兼具公共性与技术性的特征,往往被当作一种别样的政治工具^{[4][5]}。现代国家对于统计数字的收集,使其有了相较于传统时代更理性的计算能力,一方面使其权力效能能够抵达最细小、最偏僻的地区,确保了现代权力细致入微的渗透力;另一方面,建立在统计信息基础上的反思和监控能力,也确保了现代国家能够无缝隙地采集到关于自身行政能力的反馈信息^{[6][7]}。借由统计数字带来的精确信息,现代国家相较于传统国家有了更强大的认识能力、资源整合能力和资源调取能力,国家权力也得以更进一步地深入到个人生活之中。在统计技术与国家权力紧密结合的情况下,国家治理职能的扩张、国力的兴衰,以及国家治理重心的变化,都有可能导致统计体系的变动,包括统计指标的变动、统计范围的增减,以及统计数据质量水平的波动。一般来说,某个朝代的政治越稳

定,吏治越好,其治期内的户籍登记和土地统计的质量就越高^[8]。日本学者迟田温也注意到了中国汉代至唐代时期国家对基层的控制力,随着统计能力的下降而下降的事实^[9]。黄仁宇提出的“数目字管理”概念,也隐晦地指向了国家能力与数字统计之间的关系^[10]。国内年轻学者张丽萍对当代中国统计体系是如何逐步采纳GDP统计框架的专门研究,也指出了GDP水分和中国省际竞争之间的关系^[11]。

不同于仅从统计实践出发看待统计制度之分化演变的统计史研究,权力和数字之间存在着的双重建构关系,为研究者从社会科学视角出发看待数据造假问题提供了思路^[12]。一方面,统计技术为国家权力的下沉提供了技术上的保证:政治动员、资源分配、信息反馈等无不需要统计数据提供决策的基本依据。统计工作的技术性也使得各级政府可以掌握辖区内的各类重要信息,通过对数字的了解和把握而清晰地知晓辖区的社会经济发展状况。可以说,中国从前现代国家演进为现代国家的过程,也是统计技术及其应用不断提升和扩展的过程。另一方面,统计数据的政治性和公共性又使得它超越了单纯的技术性:在权力借助数字得以下渗时,数字也在运用自身的力量重构着权力。由数字所表征的发展成果,不仅是官员上升渠道的守护神,也是民众对官员政绩评价的重要依据。于是,统计数字便成了官员手中的一把双刃剑。“夸大自身的有利信息,减少自身的不利信息”可以使工作成绩更容易被上级认可;而不得不进行的数据公开则给希望监督官僚运作的公众提供了一个切实的监督渠道。社会科学界对当代中国GDP水分的解释,主要是制度层面的解释,将其归因为行政等级制中的“锦标赛体制”^{[13][14]}:在这个竞赛体制内,由GDP表现所衡量的政治能力和政治绩效,成为地方官员晋升的主要依据。由此,“数字出官,官出数字”,政绩考量带来的晋升动力促使一部分政府官员以不断注水的方式对待经济统计数字。

这也就不难解释为什么“大跃进”时期离奇的卫星数据被人们离奇地接受并且不断攀高,最终陷入一发不可收拾的局面。统计的专

业性和技术性特征,在面对政治性的制度环境时,不得不向权力屈服。国家作为政治锦标赛的目标制定者和规则提供者,在运用指标激发和考核地方政府的同时,也把数字建构的能力交给了地方政府。虽然之后的官方数据失真远不及之前的“卫星”离谱,但仍然延续了“大跃进”时期的竞赛模式。对于官员上报的统计数据而言,不仅在绝对值上要有所提高,在和同类竞争者的相对值的比较中,也要脱颖而出。“横向比较要先进,纵向比较要前进”,政治锦标赛对于指标和效率的追求,使数字腐败成了实现晋升目标的最方便渠道。在最初的数字造假带来晋升之后,政府间的数字攀比竞赛便开始了,基数虚假一旦开始,数字造假便再也无法停止。继任者即使自己不想加入到数字攀比的赛道之中,为了保全前任的政绩(和面子)也不得不继续注水。需要考量的,除了要造多大程度的假才能维护好前任与现任的政绩关系之外,更在乎的是要挤出多少水分才能被自己的上级所接受。因为在政治锦标赛开动的竞赛期间,上级领导也是竞赛中的一员,数字对于基层是重要的政绩,而汇总的基层数据也便成了中层领导的政绩。在这样的数字竞赛体制下,使统计数据返璞归真便成了牵一发而动全身的难题。这也就解释了为什么在不断查处数据造假人员,三令五申干净数据的重要意义的同时,数据造假依然大行其道。一方面,前人的注水数据让继任者骑虎难下,不得不继续注水;另一方面,数据造假带来的晋升预期使官员相信,上级领导为了自身的政绩在面临真正的检查时,会帮助自己掩盖数据失真的事实。这便是一种典型的政府间“共谋”行为^[15]。

因此,可以说,官方统计数据失真的背后,不仅仅是个别官员的错误政绩观,更是整个政治锦标赛体制所带来的结构性压力。政治信仰无法计算,社会经济发展成就却可以通过各项数据直观地反映出来。统计数字在给人们带来更直观认识的同时,也更加具有欺骗性和可篡改性。在面临升迁的考量时,巨大的政治利益往往推动着官员们铤而走险,而统计部门的分级管理 and 独立性不足,等于直接把数字制造的权力下放到了掌握决策权限的各级官员手

中^[16]。因此,可以说,围绕着当代中国 GDP 数据水分的研究,实际上指明了当代数据造假的一个主要的制度诱导和制度扩散的缘由:当数据造假的动机,从政界经由政策制定、接触和磋商,以及政商关系等中介机制传递和扩散到商业交易和日常生活的世界时,普遍的数据造假的现象也就屡见不鲜了(如淘宝卖家系统性地篡改交易记录和交易评价)。

三、数据造假的个体动机考察

当我们把视线落在实际生活中每一个做出数据造假决定的微观社会行动者身上时,一个常规的社会科学解释模型——理性计算和理性选择理论——就出现了。理性选择模型最初在经济人类学中出现,用于解释看起来非理性交换行为背后的理性基因,如 Malinowski 对美拉尼西亚群岛践行的、名为“库拉圈”的地方性交换循环的解释^[17],以及 Mauss 对礼物含义的创新理解^[18];后来这一概念(尤其是扩展了的理性概念)被用来解释更多的人类活动,如生育、消费和农民的经济决策^{[19][20]},甚至犯罪和高风险活动(如社会运动)的参与等等^{[21][22]}。如果我们把数据造假活动看作是一种典型的越轨行为的话,那么理性选择理论的论证焦点——也即违法成本和违法收益之间的不平衡——便同样可以用来解释前者:从司法体系(代表中央权威)和地方执法机构(代表地方权威)处置经济犯罪的当代实践来看,短期经济繁荣或经济绩效(以 GDP 统计为表征)方面的考虑,显然胜过了对经济犯罪的长期负面后效的考量。由此造成了司法和执法过程中对经济越轨行为的惩罚相对不足,以及当事人动机的强化:当违法收益高于违法成本时,行为将持续下去。

沿着同一条理论路线,自博弈论成为实证社会科学的主流理论工具以来,统计学界依托博弈论的理论模型,对数据失真行为进行了系统分析。在建立造假/不造假、查处/不查处的博弈模型之后,研究者们更多地将数据造假的原因归结为造假成本、预期收益和预期风险之间的相对关系上^{[16][23][24][25][26][27][28]}。这些研究在分析统计数据失真的技术原因之外,更加

强调与统计技术有关的各方利益主体的动机和预期,并通过建立利益博弈的理论模型对各方主体在不同条件下的行动策略进行了预测。研究者们提出的改革政策也往往比较直接,如提高抽查概率、加大处罚力度^[28],或者改革统计部门的管理模式^[16],以及完善统计体制、强调统计机构的独立性和重构国家统计系统等^{[5][29]}等。具体来看,《中华人民共和国统计法》、《中华人民共和国统计法实施条例》等统计法规虽然都对虚报、瞒报、伪造、篡改数据等行为明确了法律责任,但是仍然停留在模糊的“处分”、“通报批评”和 5 万元以下罚款等较轻处罚的层次上。国家统计局的《统计执法监督检查办法》就更加力不从心,在法律责任部分只规定了通报和由任免机关或纪检监察机关给予处分,把处理权限更多地交给了纪律监察机关。在具体案件的处理中(主要是国家统计局部门组织的统计执法检查),违法者所承担的责任就显得更加轻微了。例如,在国家统计局 2018 年 12 月所通报的对于农业普查造假案件的处罚中,提到最多的仍然是警告和记过处分。在通报的七件数据造假案件中,只对甘肃省凉平市泾川县林业局局长执行了撤销党内职务和行政职务的处分,其余的均停留在记过和警告两个等级,部分普查负责人只受到诫勉谈话的处罚。

官方的统计数据是如此,在缺乏明确法规约束的社会中,情况便更加不容乐观。以最近兴起的网络直播平台为例,购买“粉丝”刷人数的行为,成了平台、主播和用户心照不宣的包装策略。直播平台和网红经纪公司甚至可以合作将主播的天价合约费用通过刷礼物的形式反馈给直播平台,如此一来,既抬高了主播的身价,包装了主播身份,制造了直播间火热的现象,也给直播平台带来一系列礼物数据的刷新记录,为直播平台在资本市场的数字角逐提供了方便^[30]。与此同时,无论是刷点击、刷粉,还是机器人账号和集体数据造假,都慢慢形成了一条利益分化的产业链条。数据造假成为直播平台日常运营的战略手段,而掌握数据制造技术的人则在淘宝上大行其道,贩卖数字。比如说,某个直播平台的造假数据离奇到称在同一个直播

间中同时观看的人数达到了 13 亿,其造假程度突破了人们可以理解的下限^[31]。尽管网络直播数据的造假侵犯了消费者的知情权、违反了广告法、违背了诚实信用原则甚至涉嫌不正当竞争行为,但是由于目前《反不正当竞争法》对于消费者上述权益保护的缺位,用法律武器规制直播平台的造假行为仍然难以实现^[32]。

四、数据造假的扩散社会后果： 伦理秩序的长期损失

数据的欺骗性呈现,作为一种高端策略和战术,在中国历史上的多个叙事文本中得到了展演。比如说,在传说中的鬼谷子的两个徒弟孙臆与庞涓的争斗中,前者就分别使用了减灶法和升灶法(此处,灶的数量和兵力的数量有一个基本的映射关系)等计谋。在文化名著《三国演义》中,主角诸葛亮最善于使用的也是以迷惑为本的计谋,无论是草船借箭,还是空城计等,其实质都是操纵数据的表面来迷惑对手,以期取得战争或战役的胜利;连带着该作品中的其他主要人物,如周瑜、曹操等,也都是策略博弈造假迷惑的高手。由于这些历史叙事的巨大影响力和对民间博弈行为的系统影响,无论是在商业交易当中,还是在商业竞争当中,通过数据的欺骗性呈现来获得竞争优势的行为,实际上成为中华商业文化中的一个重要心智因素;因此,从文化工具箱(cultural tool kit)的角度来看,如果说数据造假在当代商业竞争领域能够如此盛行,部分地与这种来自历史叙事的文化赞赏和文化容忍不无关系。

成王败寇的历史逻辑使人们更多关注交往的结果而非交往的过程,对于迷惑性和欺骗性的手段则更多抱持容忍甚至赞赏的态度。但是这种容忍和赞赏所伤害的,却是人们交往过程中所共同遵守的伦理秩序,是正义、信任和效率等价值元素。就数据造假行为来说,对于眼前利益的追求与强调,实际上放弃的是更加长远的伦理秩序,是交往中的公平、正义等程序观念。当前中国社会中流行的工具主义价值观,赋予金钱和盈利能力以很高的价值地位,并由此主导了某些社会人群的伦理秩序,因此,很容

易在这些人群中形成鼓励或宽容数据造假现象的集体认知氛围。但是,随着社会总体舆论中对数据造假行为的谴责水平的日益提高,以及小圈子内对商业道德要求的提高,伴随着声誉监控和扩散技术的提升(如微信群内的及时通报和实时新闻事件后效的扩散效应),伦理秩序存在着从接纳数据造假,向拒斥数据造假演变的可能性。消费者权益保护运动的兴起,便是其中一个很好的例子^[33]。如果说人们都是生活在一个相互不接触、相互不交往的环境之中的话,则他人的意见对自己的价值就相对较少。相反,当社会行动者处于如下这样一个环境——即个人的地位和声誉取决于周围其他人的评价,而这一地位和声誉又很大程度上决定了他的一般社会福利——中时,他必须对周围观众的意见和看法施以相当的关注,有时甚至需要付出社会服从的姿态,或改变自己的行为以求社会接纳。因此,一个人所处的伦理语境对数据造假问题的批评态度和社会控制力度,构成了个体实施数据造假的一个重大社会成本。与此相反,当人们所处的伦理语境从更高等级的价值(如生存伦理)出发,包容它的存在,或者当小圈子中的社会行动者形成了“造假是必需的”的集体共识的话(如积累原始资金以求经济自立,或抵制大型垄断厂商的压制和剥削),那么数据造假现象就会以道义经济或反叛经济^[34]的面貌出现。当前社会整体舆论对于数据造假行为的声讨,为我们扭转工具主义对待统计数据的行为伦理提供了一个契机。通过在社会上宣传统计数据造假的不良社会影响、在统计从业者行业宣扬统计人员的职业道德观念、在统计培养学校宣扬数据真实的神圣地位,可以在一定程度上重建统计行业从业者的道德感和职业伦理,进而在职业群体内部形成职业压力,在社会上形成舆论压力,使得从业者之后在面临造假选择时,除了面对正式的法律规制和处罚之外,还不得不考虑到巨大的社会道德压力和职业群体内部的伦理压力,从而放弃对即时经济利益的考虑而重新回归到提供真实数据的道路上来。

总的来说,数据造假的社会后果极其严重,尤其是在它对社会价值体系的潜在冲击和伤害

方面。以官方统计数据造假为例,其最为恶劣的结果,不是中央对于地方实际情况掌握的失真和误导,而是官员间恶性竞争模式导致数字腐败的逆向选择,更是人们对于政府公信力的质疑。数据公信力无疑是政府公信力的重要组成部分,当政府发布的数据与人们的切身体验相去甚远时(比如某些地方政府对于空气质量和水质数据的离谱造假),人们便对政府数据不再采信,即使之后通过整改恢复了对数据的真实反映,重新建立公信力也是很漫长的过程。在商业世界中,数据造假同样也带来深远的负面后效。以淘宝卖家系统篡改交易规模数据和交易评价数据来说,造假行为将极大损害买家对之后同类交易活动中所遇到的卖家的一般信任水平,提高他们甄别同类风险的交易成本;而对于同属淘宝卖家的良心商户而言,一部分黑心商家的系统数据造假行为,扰乱了他们所处的商业环境,使得他们直接面对了不公平的竞争,并因此遭受业绩上和利润上的损失。对于整个淘宝交易系统来说,有意识的数据造假行为,同时也对整个交易体系的效率水平带来了巨大的伤害。因此,商业世界的系统数据造假行为,至少在正义、信任和效率这三大价值要素上,对特定社会的伦理秩序造成了直接的冲击。有时候需要一代,甚至两代人的时间,或者发起一场重新定义的文化战争,才能抵消这些行为的长期负面影响^[35]。可以说,普遍的数据造假行为,不仅仅是一种简单的、孤立的越轨经济行为,而是反映了整个社会伦理秩序的失范、社会控制的松弛和初级人际影响力的削弱,以及社会的日益工具化和次级化。

五、技术的两面性

技术的两面性,已经成为社会科学的一个重要话题。举例来说,在未来社会人们享受人工智能提供高效服务,满足特殊人群特定需求的同时,人们还心存疑虑,担心人工智能不会转而对整个社会实施技术控制。人们对待技术的矛盾态度,在统计技术的应用上也有体现,而数据造假则更多体现了人们对于统计技术的忧虑。一方面,技术的进步使得数据造假

活动的欺骗性和隐蔽性大大提高;另一方面,技术上的进展也使得我们获得了治理数据造假现象的有力工具:具体来说,当数据的生产、使用和呈现脱离了人类行动者的主观干预之后,数据造假的机会窗口将显著缩小(比如说,联网传感器系统对相关数据的自动化呈现、记录和展示,在某种程度上就遏制了人为造假的数据呈现);而当行为数据、信用数据和评级数据成为企业行动者和国家联合打造的数据治理体系的一部分时,数据造假行为的可侦测性和遭受潜在惩罚的可能性也大大提高。国家正在大力推进的全社会征信体系建设,实际上对抑制特定人群的系统化造假行为,起着巨大的作用,因为它其实是前面我们所说的声誉监控和声誉扩散体系的一个重要组成部分。可以说,在大数据技术不断被提倡的现在,提升数据的技术性特征无疑成了对抗数据造假一个有效手段。除了行为数据的自动收集之外,新型数据的深度利用也为摆脱数据造假提供了可能。以卫星遥感获得的夜间灯光数据为例,它为学者们研究城市化进程和城市空间变迁提供了新的视角。具体来看,学者们以灯光数据为基础,对贫困县认定^[36]、省级交界区空间^[37]、城市群发展特征^{[38][39]}和城镇化建设^{[40][41]}等方面进行了探讨。与以往经济数据需要大规模的人为操作不同,灯光数据由卫星遥感实时监测汇报,最大限度地减少了人为因素的干扰,其数据对于现实经济的反映状况也受到了国内外经济学家们的认可,成为GDP数据之外的一个替代选择^[42]。这为我们通过技术手段解决统计数据造假问题提供了一个思路,即强化统计数据收集、上报、汇总中的专业性、技术性和科学性,借以与统计收集过程中的公共性和政治性相抗衡。

结 语

统计数据造假现象层出不穷并愈演愈烈的背后,有着一系列的社会根源。从数据的本体论和方法论的角度来说,误差扰动、接近值模拟和数据质量的资源投入在一定程度上解释了数据失真的客观因素。但在从数据失真到数据造

假的过程中,我们也看到了统计的技术性特征逐步褪色,而政治性特征逐步显现。从统计作为国家事业的角度来说,数字与权力的双重建构关系解释了“数字出官、官出数字”的理论基础,而政治锦标赛体制又不断强化已经存在的数据造假现象。从社会行动者个体的视角来看,由于目前存在的“造假收益高,违法风险低”的执法现状,促成了越来越多的围绕数据造假的越轨行动。从文化认知的角度来讲,在具体社会实践中操纵数据成了人们的一个日常策略,获得了一定的文化宽容。数据造假的负面影响更是显著而持久的:数据造假不但对内容容易导致劣币驱逐良币的逆向选择,对外则更容易陷入塔西佗陷阱,造成政府和商业行动者的公信力尽失。除此之外,数据造假的大行其道也冲击了人们传统的交往规则和信任机制,造成伦理秩序的混乱。相对应的,要解决数据造假问题,在官方层面应当认识到政治锦标赛体制导致了地方政府内部对数据指标的盲目追求,从而逐步改革过度依靠单个指标的评价机制;在法律层面,应当确认数据真实的长远意义,提高违法成本,保护守法者;而在文化层面,应当重建合理的数据观念,恢复人们的价值认同。除此之外,还要充分发挥统计数字的技术性特征,找寻运用技术手段解决数字造假问题的有效途径。

参考文献:

- [1]韩志明. 统计数字的多维解读及其理论诠释[J]. 统计与决策 2007 (20).
- [2]Groves Robert M., Floyd J. Fowler. Survey methodology [M]. Wiley 2009.
- [3]刘奇. 从统计史的角度看现代统计学[J]. 财经理论与实践 1999 (6).
- [4]邱东. 别样的政治工具[J]. 中国统计 2004 (10).
- [5]李金昌. 浅论政府统计的本质[J]. 中国统计 2005 (12).
- [6][法]米歇尔·福柯. 安全、领土与人口: 法兰西学院演讲系列: 1977—1978 [M]. 钱翰, 杨薇薇译. 上海: 上海人民出版社 2010.
- [7][英]安东尼·吉登斯. 民族—国家与暴力[M]. 胡宗泽, 赵力涛, 王铭铭译. 北京: 生活·读书·新知三联书店, 1998.
- [8]蔡光柏, 刘英杰. 历史人口的统计范围和资料处理[J]. 西北人口 1983 (3).
- [9][日]池田温. 中国古代籍账研究[M]. 龚泽铎译. 北京: 中华书局, 1984.
- [10]黄仁宇. 万历十五年[M]. 上海: 三联书店, 1997.
- [11]张丽萍. GDP: 中国实践中的双重建构[D]. 北京: 北京大学社会学系博士论文, 2018.
- [12]赵胜忠. 数字与权力[D]. 南京: 南京大学政治学博士论文, 2012.
- [13]周黎安. 中国地方官员的晋升锦标赛模式研究[J]. 经济研究 2007 (7).
- [14]周飞舟. 锦标赛体制[J]. 社会学研究 2009 (3).
- [15]周雪光. 基层政府间的“共谋现象”——一个政府行为的制度逻辑[J]. 开放时代 2009 (12).
- [16]李军林, 张威, 刘思璇. 统计造假治理与改革路径研究——基于监督博弈模型的分析[J]. 区域经济评论 2015 (6).
- [17]Malinowski, Bronislaw. Argonauts of the Western Pacific [M]. Waveland, 1988.
- [18]Mauss, Marcel. The Gift: The Form and Reason for Exchange in Archaic Societies [M]. W. W. Norton 2000.
- [19]Becker, Gary S. The Economic Approach to Human Behavior [M]. Chicago: The Chicago University Press, 1978.
- [20]Popkin, Samuel L. The Rational Peasant: The Political Economy of Rural Society in Vietnam [M]. Berkeley: University of California Press, 1979.
- [21]Beccaria, Cesare. An Essay on Crime and Punishment [M]. Forgotten Books 2017.
- [22]Olson, Mancur. The Logic of Collective Action: Public Goods and the Theory of Groups [M]. Harvard University, 1971.
- [23]雷宏. 对统计数据造假行为的博弈理论分析[J]. 集美大学学报(哲学社会科学版) 2005 (2).
- [24]雷宏, 雷春. 我国统计数据造假的成因及治理对策探讨[J]. 求实 2005 (2).
- [25]叶青, 李丽媛. 政府统计的博弈研究[J]. 统计与决策 2014 (8).
- [26]朱新玲, 黎鹏. 非对称信息下统计造假行为的博弈分析[J]. 统计与决策 2007 (15).
- [27]李卫东, 刘似臣. 统计数据质量的博弈分析[J]. 中国统计 2005 (4).
- [28]朱秀君. 政府统计中造假与治假的博弈分析[J]. 数量经济技术经济研究 2000 (7).
- [29]李金昌. 完善统计体制——从统计职能与政府统计本质说起[J]. 中国统计 2018 (1).

- [30]张子雨. 网络直播平台陷入烧钱战 数据造假存隐忧[J]. 广西质量监督导报 2016 (8) .
- [31]艾小雅. 我国网络直播平台的发展现状与前瞻研究[J]. 新媒体研究 2018 (4) .
- [32]陈全森. 网络播放平台数据造假行为的法律规制[D]. 南昌: 南昌大学法学硕士论文 2018.
- [33] Brobeck ,Stephen ,Robert N. Mayer. Watchdogs and Whistleblowers: A Reference Guide to Consumer Activism[M]. Greenwood 2015.
- [34]Scott ,James. The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia [M]. Yale University ,1977.
- [35]郑也夫. 信任论 [M]. 北京: 中国广播电视出版社 2006.
- [36]叶宸辰 ,唐斌斌 ,刘林平. 中国贫困县划分精度商榷——基于夜间灯光数据的检验[J]. 中国农业大学学报(社会科学版) 2018 (5) .
- [37]曾冰 ,邱志萍. 省际交界区经济网络空间结构研究——以湘鄂赣的灯光数据为实证[J]. 财经科学 2018 (11) .
- [38]赵晓伟 ,孟治国 ,郑兴明 ,范俊甫 ,王庆帅. 夜间灯光数据的东三省城市群发展特征研究[J]. 测绘科学 2018.
- [39]钟洋 ,林爱文. 长江中游城市群空间格局演变及优化研究——基于 DMSP/OLS 夜间灯光数据的方法[J]. 水土保持研究 2018 (6) .
- [40]周翼 ,谢保鹏 ,陈英 ,赵鸿雁 ,裴婷婷. 基于灯光数据的中国县域城镇建设用地产出效率时空演变特征[J]. 地球信息科学学报 2018 (12) .
- [41]张滔 ,唐宏. 基于 Google Earth Engine 的京津冀 2001—2015 年植被覆盖变化与城镇扩张研究[J]. 遥感技术与应用 2018 (4) .
- [42]徐康宁 ,陈丰龙 ,刘修岩. 中国经济增长的真实性: 基于全球夜间灯光数据的检验[J]. 经济研究 , 2015 (9) .

责任编辑 黄 杰

Data Production and Data Fraud: An Analysis Based on Sociological Perspective

LIU Neng & MA Jun-nan

(Department of Sociology , Peking University , Beijing 100870 , China)

Abstract: This paper starts with the process of data production and traces the causes of statistical data fraud. First , the level of resource in error perturbations , proximity simulations , and data quality explains the distortion of data during data production. Second , from the perspective of statistics as a country's cause development , national capabilities and state intentions sometimes may affect the authenticity and comprehensiveness of statistical data , and lead to institutional-induced fraud. Third , from the perspective of the rational selection process at the individual level , when the illegal income is significantly higher than the illegal cost , the data fraud phenomenon becomes one of the most common deviant behaviors in the economic field. The potential of falsifying data is also directly related to individual motivation and rational calculation: when we extend the purely material rational computing model to include social costs and group dynamics , we will find that social networks and peer ethics can cause damage to internal ethical reputations , and this damage can be a greater constraining force to prevent data fraud than the increase in material illegal costs. Finally , the paper also discusses the direct ethical loss of data fraud as well as the duality of the role of technology in the governance of data fraud.

Key Words: data production; data fraud; national statistics; ethical losses; technology and its two duality