对大数据统计设计的思考*

赵彦云

内容提要: 本文认为大数据统计与三个问题有关: 大数据发展趋向极限无穷时,人类社会数据信息将发生什么变化? 大数据发展会不会产生危害社会进步的数据垃圾? 大数据即是一场革命,那么作为数据科学的统计学脱胎换骨地继承与发展的是什么? 本文对此的回答包括,提出了大数据发展的统计设计观点,从理论和实践上做出了论证分析,并联系我国实际,探讨了我国大数据发展中的统计设计理论和内容要点。

关键词: 大数据统计; 统计设计; 元统计; 降维与增维
中图分类号: C829.2　　文献标识码: A　　文章编号: 1002 - 4565(2015)06 - 0003 - 08

Reflection on Statistical Design in Big Data

Zhao Yanyun

Abstract: We find that big data statistics is related to three issues. Big data tends to be an infinite data information, what will happen to human society development for the role of big data in future? Will development of big data produce data garbage which does harm to social progress? Even if big data is a revolution, what is the inheritance and development of statistics as data science? This paper has made in-depth analysis and put forward the viewpoint about the development of big data from the point of statistics design. The scientific argument and analysis is made from the view of theory and practice, and theory and main content of China's statistical design in the context of big data are discussed based on the reality of China.

Key words: Big data statistics; Statistical Design; Element statistics; Statistical Dimension Reduction and Dimension Expanding

在计算机、互联网、云计算、大数据迅速发展的背景下,探索统计的科学作用至为关键。因为大数据以数据为主体,而统计是关于数据的科学,因此,统计科学应该在大数据发展中起主导作用。然而,现实当中,社会大众和业界人士似乎还没有发现统计科学的重要性。为此,本文提出大数据发展中统计设计这一主题,希望通过探索统计的科学思想和其发展的客观依据,引领大数据统计平台建设,进而更好地聚集能量,推动大数据为社会发展、技术革命和生产力水平提高做出贡献。

一、大数据统计的挑战

我们追踪大数据的发展历程,以及大数据头脑风暴的发散思维,从中发现大数据统计与三个问题有关: 第一个问题,假设大数据发展趋向极限无穷,人类社会数据信息将发生什么变化? 第二个问题,大数据发展会不会产生危害社会进步的数据垃圾? 第三个问题,大数据即是一场革命,那么作为数据科学的统计学脱胎换骨地继承与发展的是什么?

第一个问题,大数据发展,即数字化时代发展的最终目标仍应是使社会资源如何得到最优配置和利用,市场价格包括工资报酬、资本收益率、利率、技术价格、资源价格、产品价格、服务价格等的定价标准,进而保证市场的有效竞争,保证社会公平、公正。按照数据智能化发展的趋势①,货币政策、财政政策等宏观政策、区域政策都能内生系统化、智能自动化,等

* 本文获国家社科基金重大项目"经济社会公共数据的空间统计样本数据开发及应用研究"(11&ZD157)、全国统计科研规划项目重点课题"基于大数据政府统计改革发展的网络互联网公司统计报表制度设计研究"(2014)资助。《统计研究》许亦频副主编在本文修改中提出了宝贵意见,特此致谢。

① 知识增长的三个时代: 线性,大数据,智能数据,包括数据规模值的飞跃,跨设备用户识别,知识图谱,用户行为预测系统,深度学习等。

等,但现实判断的依据是: 虽然统计思想和统计工具还未能深入大数据并发挥主导作用,但是个性化推荐、社交化推荐是主流,即大数据发展仍处于计算机思维的主控期,即把所有数据都有序存储起来,并快速提取到,以及局部系统上的一些目标的关系挖掘,但是,更加科学深入的统计分析和实证研究却是空白。

第二个问题,大数据发展会不会产生危害社会进步的数据垃圾? 显然,目前的大数据着重于实时数据和短期历史数据,远期的历史数据是否还有用或部分有用,如果回答是否的话,那么,随时间的推移,大数据之外必然有不用或无用的数据,或在某一个时点之前的数据成为永远都不能用的数据,即使数据可能还有待发掘的价值,但随着时间趋向无穷,其数据使用价值趋向零,显然这些就是数据垃圾。那么大数据是否一定会产生大数据垃圾,如果不产生需要什么样的条件。现实看,大数据不能总是着眼眼前数据,这样的大数据会产生大数据垃圾。当前的大数据发展比较突出个性化和社交网络化,大数据价值的大小取决于个性化的个性单位最小化和社交网络数据上的大数据网络结点有效连接的最大化。然而,个性化与社交网络一体化的大数据,不能缺失时间上的连续性条件,也就是个性单位最小化和社交网络数据上的大数据网络结点有效连接的最大化与个性单位和社交网络的时间连续条件满足的话,那么,大数据发展可以避免大数据垃圾产生。

第三个问题,大数据统计即是一场革命,那么作为数据科学的统计学脱胎换骨地继承与发展的是什么? 谈起数据,人们就会与统计相连,因此大数据也应该如此,但是在计算机、数据库、分布计算新兴技术的大量普及等强势应用下,统计作用优势相对较弱,虽然数据挖掘、统计计算、统计模型与降维技术等都被认可,但是在大数据发展中的统计地位和作用仍需要努力开拓和发展。从思路上讲,统计学应紧跟大数据发展趋向,分析研究在大数据发展过程中,统计科学理论方法在哪些方面被弱化,哪些方面被追捧,重心、核心是什么,新体系如何变革演化等。事实上,对于统计科学理论方法,应用是本质特征,其中的学科发展基础和存在的问题是引发学科内在发展的关键。面向大数据挑战,统计学要继承与发展,当今统计必须充分考虑在强大的计算机网络及云计算等能力上的条件,继承样本总体、统计分布、统计描述、统计探索发现、统计推断、统计降维等理论方法,发展宏观与微观一体化、降维与增维并举、最小样本唯一码统计动态标准及智能自动化等新理论和新方法。

关于大数据应用,比较集中的一个观点是统计数据总体全面化,统计理论方法可能面临从样本推断总体的核心向外扩大发展,统计如何在整体上、过程上发挥作用,本文的观点是大力发展统计设计的理论方法,特别是从微观到宏观及增加时间因素的复杂系统的一体化统计设计,其中元统计基础的统计设计尤为关键和重要,实际上是探索人类社会定量化可持续的标准基础统计理论方法,以及在分析上的降维与增维的革命性思想和理论方法。

二、大数据统计发展的思路

互联网、云计算、大数据已发展成为当今的潮流,统计在大数据中具有怎样的地位和作用,是当前统计科学急需探索的重要问题。大数据统计应该按照什么样的思路探索? 本文给出公式化解析:

大数据 = (计算机 + 互联网) + 统计

(计算机 + 互联网) = 记录存储无限数据 + 最大社会网络

统计 = 可无限内部组合的最大统计总体 + 最小现实样本

探索统计大数据中的空间: 发展广义统计设计及其统计分析理论方法

我们应该深刻认识大数据是统计与计算机相结合发展的过程,即通过对经济社会活动与经济社会关系的定量、定性的观测与实验等过程,引入科学的技术和方法,达到对各种规律的有效把控、利用和管理的目的。在这个过程中,人们发展和利用各门科学知识,包括哲学、人文社会科学和自然科学、工程技术、计算机科学与计算机技术,以及统计学、经济学、管理学等,其中,计算机科学与计算机技术和统计学,成为当前大数据时代的核心科学推动力。

我们提出大数据统计发展的两个统计思想要点: 可无限内部组合的最大统计总体和大数据的最小现实样本。发展广义统计设计及其统计分析理论方法,目的是为了探索统计大数据。所谓广义统计设计是相对于一般统计的微观特点或专门性而言,例如,现实存在的统计实验设计、企业经营统计设计、产业统计设计、部门统计设计、国民经济综合统计设计,以及各种专题的统计设计,在这些统计设计

统一层面需要的是针对大数据统计发展的基础统计设计。因为在大数据情况下,统计数据的搜集整理是多方面、多形式的,利用互联网、移动互联网、物联网等发展的自然记录和有序进入数据库和云计算的海量数据,需要考虑统计设计与计算机设计相结合的发展模式,其中上述谈到的大数据统计发展下的两点统计思想是非常关键的。第一个,可无限内部组合的最大统计总体,是复杂大系统统计科学性的一个基本点,是对数据的系统一致性和可分解、可加总、可关联,能上能下地从微观到宏观的整体最优统计设计的要求; 第二个,大数据的最小现实样本,是统计设计的起点要求,或者是元数据的统计设计要求,最小样本是要求统计数据在一个样本上的数据现实客体的一致性,这是研究实际问题统计分析的重要前提。

三、大数据统计发展的基本问题

从大数据发展的现实趋势中寻找大数据统计发展灵感是非常必要的。大数据统计,属性上应用性质突出。中国应用统计相比发达国家而言有自己的特色,从学科上看经济社会统计与概率数理统计平行发展,形成竞争与合作的发展格局; 从应用上看,我们善于宏观经济系统与社会系统的统计设计,因此对统计在宏观问题上的应用比较重视,有利于统计应用发展解决复杂系统的能力建设。当然,也存在着相对忽视统计应用基础建设的问题。

(一) 大数据统计发展的现实基础与融合

笔者认为,推进大数据统计发展的一个重要领域,应该在公共数据领域,因为这部分大数据直接关系经济发展与社会进步,特别是关系基础建设、社会保障、现代服务、宏观政策与市场经济繁荣等问题。就公共数据领域与非公共数据领域的大数据划分而言,前者是大河长江湖泊,后者是小溪支流,公共数据领域的统计设计做好了,可能是对大数据发展最积极的推动。

发展公共服务领域的大数据统计要从实际出发,其中重要的一点是要明确现在所有的公共统计数据内容及其背后的科学体系,能否为大数据发展提供统计的科学价值,实际的统计数据脉络又是如何。图 1 说明了公共统计数据当前的基本格局,这将成为大数据统计设计的起点。

大数据统计是在更大的系统范围内,提供统计

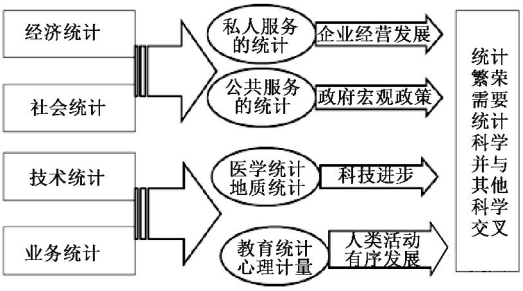

图 1　从现实出发的大数据统计设计发展基础

内涵多层次与统计内在一致性的数据体系。从现实统计直接的技术表现看,基本包括经济统计、社会统计、技术统计和业务统计,前两部分实际运用比较多,后两者有待在互联网和信息技术、物联网、APP 支持下增强。其中,业务统计贴近经济活动和社会活动,现实性强; 技术统计深入科学层面,对生产力及其发展有更有效的刻画。图 1 第二个层面,进一步解释上述统计服务的社会格局,经济统计和社会统计,主要为市场和非市场两大部分,所谓市场体现为个人、企业和市场竞争服务的统计,支持企业经营发展,保障消费需求和投资需求的有效实现; 非市场部分是为公共服务的统计,它是社会进步发展和保障市场有效竞争发展的重要内容,支持政府宏观政策,包括经济政策和社会政策等的科学制定、实施与效果评估。技术统计涉及各个科学领域,例如医学统计、地质统计、教育统计、心理计量等等。业务统计涉及所有的人类活动领域,体现为用统计手段描述人类社会活动的有序发展。针对现实存在和发展需求,我们提出从业务统计、技术统计、社会统计、经济统计四个方面发展大数据统计的科学规范,相对于过去仅从实物量统计和价值量统计二分法,可能在系统落实大数据统计研究上有新的进步。另外,上述四个层面的统计设计,更加追求大数据统计的生态自然过程,即从直接的活动业务属性、技术属性、社会属性、经济属性,系统测量测度人类社会活动的统计特征,累积统计数据,监测、引领、控制发展过程,为人类存在与发展造福。

(二) 统计交叉学科的重要作用

目前我国应用统计的交叉学科发展比较弱,各学科大多追求独立性的最大化,这种状况不利于大数据统计应用的科学发展。下面举例分析。

生产理论、生产函数是经济学的重要内容,也是数理经济学、计量经济学、经济统计学研究的重要方

面。从国民经济核算体系的设计基础看，它是严格按照经济学概念、理论体系的最简公约设计的，因此，生产、收入、分配、消费、储蓄、投资、国民财富等从流量到存量、从资产到生产及收入分配、从资产负债到金融的投融资等经济运行复杂系统的循环体系，生产理论、生产函数、经济增长等理论发挥着核心的作用。我们以生产函数应用的变化过程和经济增长核算为例来说明统计学内部交叉，统计学与经济学外部交叉的相互推动作用。

1. 生产函数模型及其应用的发展。

生产函数:

$$Y = AK^{\alpha}L^{\beta} \quad (1)$$

减少变量:

$$\frac{Y}{L} = A\left(\frac{K}{L}\right)^{\alpha} \quad (2)$$

假设 $\alpha+\beta=1$，即规模收益不变。

增加变量:

$$Y = A(K + K_i)^{\alpha}L^{\beta} \quad (3)$$

从生产函数模型式(1)、(2)、(3)变化中可以看到，这些变化融入了经济学与统计学及数学的交叉相互推动发展。柯布道格拉斯生产函数提出之后的应用发展，实际是与统计数据的发展紧密结合，发展探索对经济问题的研究，减少变量以突出核心问题，增加变量以进一步细化经济过程的细分因素，这些工作都必须要做到经济概念的可统计，统计数据质量可保证。但实际统计工作，如资本存量统计、劳动投入统计等方面，还有许多没有解决的问题，随着资本深化和金融创新等的发展，又有新的统计问题不断产生。所以，做好满足生产函数理论的统计模型分析，统计设计、统计调查、统计数据质量控制还需要更多的努力。而这个过程，需要统计学家与经济学家一起探索解决基于实际情况的具体问题。统计工作只是相对最优方案，在统计基础上，经济学家的思维也需要有所改变，即采取面对大山拦路、绕行开路还是隧道建设，应该联系成本优化考虑。事实上，经济发展，也有社会引领问题，统计是经济系统管理的一项基础工作，统计工作的主要手段是保证生产关系对生产力的实现，在现实制度上尤其最基本制度上，可以考虑统计与管理有效结合的长期发展设计。

随机前沿面生产函数:

$$\ln Y_i(t) = \alpha_0 + \alpha_1 + \alpha_L \ln L_i(t) + \alpha_k \ln K_i(t) + u_i(t) \quad (4)$$

超越对数生产函数:

$$\ln Y_i(t) = exp\left[\alpha + \alpha_k \ln K + \alpha_L \ln L + \frac{1}{2}\beta_{KK}\ln^2 K + \beta_{KL}(\ln K)(\ln L) + \frac{1}{2}\beta_{LL}\ln^2 L\right] \quad (5)$$

从生产函数进一步向式(4)、(5)的模型发展，也看出数学在经济学与统计学交叉合作中的积极作用，深入分析经济问题，可以通过数学工具实现简化的复杂分析，统计学的数理逻辑基础在一定程度上容易与数学模型方法相互衔接。图 2 利用数学几何表达对生产函数变动内涵的全要素生产率、生产效率和技术进步做出了逻辑严谨的定量分析，A 点代表生产无效率点，B 点和 C 点表示生产有效率点。无效率的程度可以用该点与前沿曲线的距离表示，距离越大越无效率。全要素生产率定义为从原点出发的射线的斜率。如果一个厂商从 A 点移到 B 点，斜率变大，意味着全要素生产率的提高，同时生产效率也得到了改进。从 B 点移到 C 点，生产效率不变，但是全要素生产率得到了提高。从生产前沿曲线 1 上移到生产前沿曲线 2 就是技术进步。

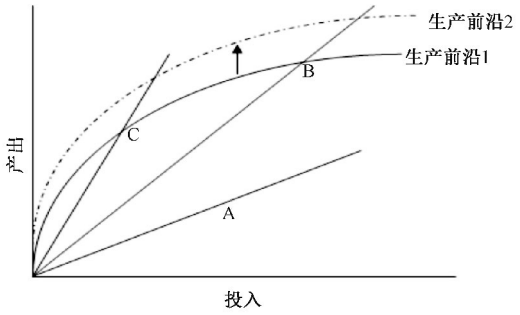

图 2 全要素生产率、生产效率和技术进步

2. 经济增长核算—从经济统计向相关学科的发展。

图 3 显示，生产函数模型引入经济增长核算，形成了经济学、统计学(经济统计、数理统计)、数学、计量经济学之间的交叉学科相互推动发展的过程解析。索洛增长方程，表面上是把非线性模型转化为线性模型，并有效与现实统计数据相结合的过程，进一步看，他在研究上引入了经济核算的思想和方式，余值变量引入了统计学思想和方式，即残差的进一步分解和统计假设条件的分析。

经济核算的思想和方式是什么？现实中就是会计核算、会计账户、国民经济核算账户的经济思想和方式，实质是对现实发生的经济活动计算价值，描述

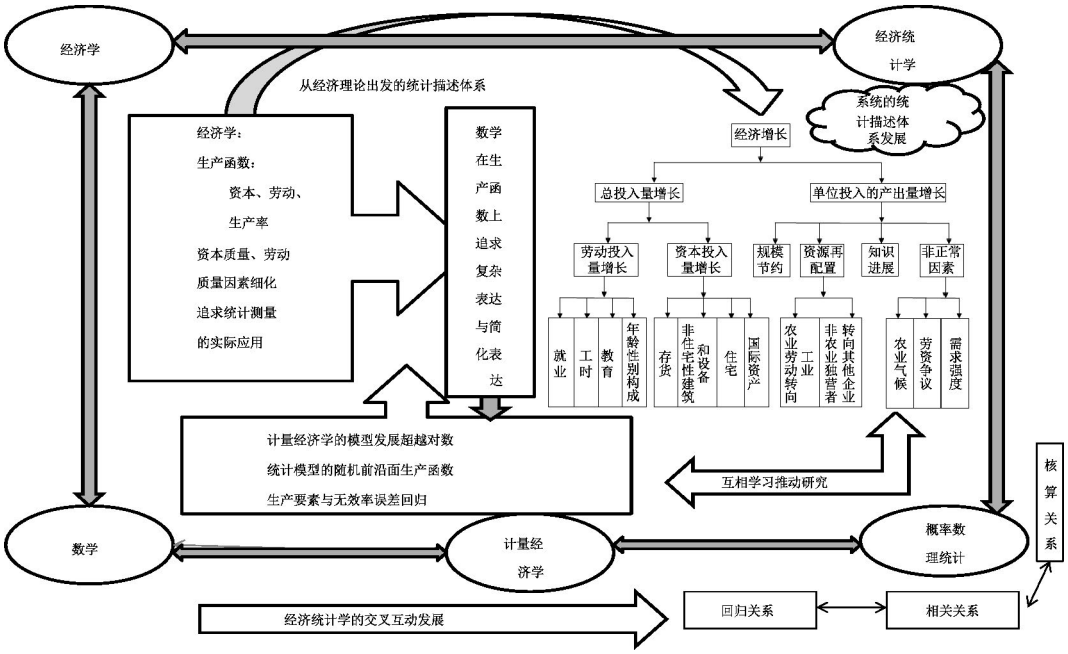

图 3　以经济增长分析为核心的交叉学科作用解析

相互之间的直接联系。就会计而言,企业或经济单位的每一笔经济活动成为最小单元计算价值,各种活动之间的价值直接联系完全是加减关系,由此逐级向上汇总就是国民经济核算账户,其中从会计向国民经济账户转化运用两个统计标准制度,一个是时间一致性的统计标准,一个是统计分类标准。时间标准是保证经济系统统计量化系统一致性,这是采用经济学分析原则的统计处理。统计分类标准,国民经济核算体系中最重要的是国民经济行业分类标准和国民经济机构部门分类标准,它们都是世界通用并统一编码的基础性统计标准,其科学依据仍然是经济学的生产理论、消费需求理论、投资理论等。

经济增长核算,从美国经济学家丹尼森和政府统计学家肯德里克等创建和发展以来,实际上是按照经济统计学的思想和方式方法发展的,突出特点是从经济统计现实出发思考经济学问题研究的因素,并追求影响因素的不断分解细化,研究他们的因素归属关系,还原到微观企业及经济活动的现实细节,以实现用从微观到宏观逐级汇总的统计数据进行描述和分析。经济增长核算首先分成总投入增长和单位投入的产出量增长两大类,前者是生产要素投入增长因素,后者是全要素生产率增长因素,这样划分完全是按照经济增长理论,也符合生产函数理论。进一步的经济增长核算,体现在总投入增长基础上,进一步按照劳动要素和资本要素的数量与质量及其不同类别进行划分核算因素,达到深入分析生产要素增长对经济增长的目的。明显看出,这些细化的生产要素因素,一方面遵循经济学理论,另一方面是遵循统计规则分解并保持核算的关系,因此可用于实证分析。同样在单位投入的产出增长即全要素生产率增长的进一步细分上,根据显示经济活动,考虑统计可行性,划分了规模节约、资源再配置、知识进展、非正常因素,并根据市场活动做了进一步的因素划分,这些做法同样考虑了经济统计可行和经济学分析的理论价值。经济增长核算,在引入统计核算思想和做法,并以经济学为指导,深入现实经济活动要点等方面非常突出,它为推动数理经济学和计量经济学的经济增长模型分析和数学与统计方法的运用产生了积极的作用。由此可见,统计学思想、经济统计学方法、数理统计逻辑,对经济学分析框架和因素体系,以及计量经济模型都产生重要的影响,今天我们深刻理解这种多学科的分工合作和"价值链"关系,对于未来的研究具有重要作用。

当然,统计学作为一门覆盖面比较宽的学科,其所包含的经济统计学、概率论、数理统计学之间,也应有交叉关系。但长期以来,经济统计学与经济学保持紧密关系,而经济统计学与概率论、数理统计学之间,在统计科学思想一致性与分工上,基本上没有什么交叉研究,因此造成连接地带空白的问题。当面对应用问题时,目前的统计分析多以相关关系、回

归关系为主，忽视了核算关系的基础作用，跨越经济统计学，必然会割裂了数理统计学与经济学、会计学，以及现实经济活动因素细节的关系，因此，我们也需要在统计学内部各分支学科上的交叉互动。

（三）探索增维与降维的双向发展

从经济统计发展来看，投入产出表和投入产出模型的产生，为宏观经济定量分析引入了重要的统计思想，即对国民经济行业分类可综合与可细分，以及把国民经济账户用数学模型及矩阵化处理，大大推进了对经济与社会复杂系统关系及其影响因素的分析研究，被广泛运用到经济分析，国际贸易分析，金融流量分析，能源、水与环境等问题的分析研究中。今天看来，可以成为我们探索大数据统计增维与降维双向发展的一种思路。

目前的经济社会统计是一个复杂系统的数据体系，从威廉·配第的17 世纪中叶至今的 300 多年期间，这一体系不断改进统计设计，尤其在产品与服务分类标准、行业分类标准、机构部门分类标准，以及追求单位编码和身份证号码等统计基础建设上，有了重大发展，这一过程实际上是对复杂经济社会系统的简约化统计设计的过程，应该讲，这些也是大数据统计发展的基础，也为探索大数据统计增维与降维双向发展提供了有实践意义的统计思想。

（四）加强公共统计简约化标准建设

大数据向微观、宏观两个方向发展，需要新的大数据统计平台，包括新的统计思想和理论方法及新应用。但是，这个平台的科学内涵应该如何研究发现、构建和应用，笔者提出通过简约最小化要求的统计设计观点。图 4 是针对经济系统和社会系统及其关系的大数据系统，提出的一个基于最优简约最大化原则的研究思路，基本出发点是通过最基础的分类标准体系建设，实现大数据统计设计。图中的公共统计简约化标准说明的是探索全社会大数据统计的最基础编码标准，以及由此建立的统计分类体系。最底一级统计标准编码是身份证编码、单位编码和产品与服务分类编码，即建立人、单位唯一码的全社会统计标准对象。其中，单位编码与国民经济机构部门分类标准和国民经济行业分类关联，身份证号码与社会信用、社会保障分类标准关联。产品与服务分类编码是经济社会活动的最基础分类标准编码，它是国民经济机构部门分类标准和国民经济行业分类的基础。

统计设计是统计工作的首要阶段，包括实验设计、抽样设计、统计指标及指标体系设计等。那么统计设计面对大数据可以发挥什么样的作用，应该发挥什么样的作用。统计理论方法在大数据的降维方面，已经做了许多研究，已经在生物医学统计、基因信息分析研究等领域展开了应用，例如，用偏最小二乘方法在酵母菌细胞周期数据集同时进行数据降维和变量选择的研究（Chun 和 Keles 2010）、高维 ECG 数据（心电图）中使用主成分分析（PCA）进行数据降维（Johnstone 和 Lu 2009）、用张量型数据的主成分分析将不同试验条件下、不同时间点的基因表达量进行主成分降维（Omberg 等 2009）。但是统计的增维还没有涉及。大数据发展具备充分的系统数据信息，因此，大数据统计可以探索如同显微镜、望远镜、放大镜，可繁可简。因为统计具有总体完整、样本可比、指标范围可比、指标内涵可比、指标动态可比、指标平行归属单位样本可比的科学要求，统计数据库具有可组合的“魔方”应用价值，因此，在实际统计工作中，统计首先要在定量上有标准化的设计，才能发挥统计归纳、比较、描述分布、计算概率，进而开展估计、优化等精确测量和科学分析的作用。基于以上的思考，笔者认为我们可以面对大数据，研究统计设计的作用，从现实的各个领域的大数据出发，追寻统计的作用，进而研究全面、系统的统计设计。

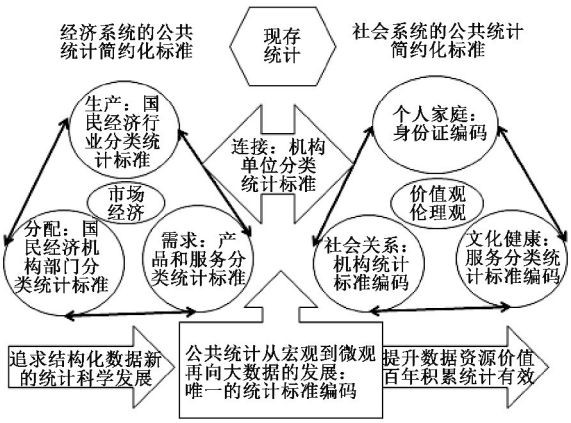

图 4 经济社会系统中的大数据简约化的统计标准

四、政府大数据统计设计

政府统计发展已经形成了新的格局，第一，政府统计范围日趋扩大和深入，从我国的实际情况看，国家统计局系统的综合统计、政府各机构的部门统计、非营利机构的公共统计共存，目前已呈分工协调有效合作的发展态势。第二，政府统计在根据统计法

保护个人和单位隐私前提下,通过科学手段逐步扩大微观数据的开放,公共统计数据的层次和容量不断扩大。第三,政府统计工作越来越在财务数据、金融数据、生态环境监测、行政记录等交叉工作中完成。第四,政府统计从事后统计,逐步向事中和事前统计发展,统计方法从调查事后的硬统计,向包括问卷调查等主观测度潜在变量和定性内容定量化的软统计和硬统计的新统计体系快速发展。

政府统计大数据工作,目前的主要思路是在原来统计工作基础上加以补充的方式进行,例如,针对网上购物,采取增加网上购物的统计数据。相对而言这是一种比较简单直接的做法,但是其他方面,如通过互联网信息和数据,替代原有统计的做法则要复杂得多,需要更加严谨、科学的统计设计。事实上,面对大数据,如何做好统计设计和统计工作设计,上述就事论事的补充的方式,可能遇到特别复杂的问题,而且难以科学解决,随之,还可能引起原来政府统计工作质量下降、新的互联网统计质量难以提高的尴尬局面。因此,我们面对互联网、云计算、大数据的迅猛发展,应该全面系统地分析研究,针对不同发展阶段,提出全面改进、完善统计设计和统计工作设计的方式方法,通过基础设计、核心设计、流程设计等方式,迎接各种发展的挑战。

(一) 大数据政府统计的基础设计

大数据政府统计的基础设计主要体现在国民经济产品和服务分类标准、国民经济行业分类标准、国民经济单位机构编码标准的建设上。目前,产品和服务分类标准已经到 10 位码,但是实际使用却非常综合,而且许多服务分类达不到 10 位码的科学划分层次,不能及时发展更新标准可能是统计基础的最大问题。国民经济行业分类目前只使用到 4 位码,相对产品和服务分类标准的 10 位码,显然国民经济行业部门分类标准综合性更强,恐不能适应大数据统计发展的要求。国民经济单位机构编码工作目前还是分割状态,工商局系统负责企业、事业单位的登记和管理,中央编办①系统负责各级政府机构单位的登记和管理,民政部系统负责非营利机构单位的登记和管理。其中,机构编码是一项技术性基础工作,原则上由国家质检总局负责,但是,这项工作还没有发展到科学制定和科学管理的程度。面对互联网、云计算、大数据,这些标准的科学性和发展都是首要的统计设计工作。应该认识到,此处谈到的三个基础标准的统计设计是为大数据的全社会科学管理而设计的标准,它可以保障全社会公共统计大数据完整有序,实现对全社会科学服务价值最大化。身份证号码,已经被认为是全社会人口有效管理的标准设计,具有唯一性和号码内含信息科学性,并具有国际规范化,为国家社会保障、社会人口管理与互联网相结合发挥了巨大的作用。然而,相比之下,机构编码的工作,远远没有达到像身份证号码那样的应用效果,全社会企事业单位、政府机构单位、非营利机构单位的管理,只在局部使用,缺乏统一的统计标准设计,严重阻碍与互联网结合的大数据应用。国民经济产品和服务分类标准、国民经济行业分类标准是有内在联系的统计标准,尽管已经有了非常广泛的应用,但是动态的科学发展,特别是在与互联网的结合和为互联网创造更大的发展应用空间方面,还存在许多问题。面对物感网和穿戴设备的移动互联网,我们应该用更加发展的视角来把这两项国民经济统计标准设计好,产品和服务,无论在生产技术特征和满足需求的服务价值方面,都在日益发展,我国的统计标准设计恰恰忽视了这个动态发展的问题,因此,导致政府统计数据使用价值严重滞后于客观的实际发展。相关部门在上述统计标准设计上,应该发挥全面政府职能,积极做好公共服务中的基础工作,其中统计基础标准(包括编码的统计设计和实施管理)是最重要的基础工作。

(二) 大数据政府统计的核心设计

我们可以把中国的政府统计划分为三大类,一是计划经济下的政府统计,二是市场经济下的政府统计,三是现代互联网信息技术下的政府统计,依据这三大类来讨论政府统计的核心统计设计就比较清楚了,这说明政府统计是根据内外部条件决定其发展及统计设计的,核心统计设计只是进一步突出核心统计内容的设计特点。

计划经济下的政府统计的主要特征是对应行政命令的计划管理工作,政府统计核心设计是全部国有企业、集体企业和政府部门,强调统计台账,基层统计与宏观综合统计高度一致,使用统计数据主要

① 中央机构编制委员会办公室是中央机构编制委员会的常设办事机构,在中央机构编制委员会领导下负责全国行政管理体制和机构改革以及机构编制的日常管理工作,既是党中央的机构,又是国务院的机构。

是制定计划和检查计划完成情况，以及宏观上的总体总结宣传。

市场经济下的政府统计，主要特征是在统计法约束下公共统计数据的生产和使用，理论上应该为产品及服务市场和要素市场有效运行提供以价格为中心的公共统计数据，保护市场参与者利益，推动公平竞争，促进技术创新和管理及组织等创新，引领要素合理流动和资源最优配置，追逐高效率，不断为国民增加财富。显然，我国目前正处于这阶段，按照国际惯例建立与国际接轨并反映中国特色的国民经济核算体系是核心统计设计。然而，我国在市场经济下的政府统计设计，对市场经济这一复杂系统中以价格为中心的统计建设不足，市场供求双方没有形成以统计为客观分析依据的习惯和氛围，基础产品如水、药、能源等价格的管理也缺乏充分的统计基础和相应的市场原则性分析，人为主观导致许多错误的市场信号，造成巨大经济损失，产生许多“被统计”的社会不良影响。另一个问题表现在基础统计与宏观统计系统脱节，部门统计各行其是，分割状态下的政府统计造成统计数据的作用迅速衰减，直接影响我国全面深化改革和四个全面发展的进程，如果，没有运用大量系统统计数据对我国存在的现实问题做出科学有效的分析研究，改革发展措施的客观可行性就没有保障，显然政府统计基础工作是科学发展的充要条件。

现代互联网信息技术下的政府统计的核心设计应体现统计法和技术基础设施支持下的政府统计数据的生产和使用，因为有了物联网和穿戴设备移动互联网、云计算、大数据等，政府统计的核心设计可以在现行政府统计内容基础上，更加细化升级，提升对市场复杂系统的统计设计功效，在更大范围即全国、地区、行业，以及企业、金融、政府、社会非营利组织等大系统下形成可分级、可分层、可组合、可动态、可中心化的强大政府统计功能，为企业、产业、政府、社会提供公共统计数据信息服务。

（三）大数据政府统计的流程设计

大数据政府统计的流程设计是把基础统计设计和核心统计设计连接起来的具有可操作性的政府统计设计。计划经济下的政府统计流程设计是统计报表体系，即所有单位都要完成统计报表中的统计内容。除了统计报表体系之外，还有普查（人口普查、经济普查、农业普查）、抽样调查（农产量、1%人口、住户收支、消费价格），正在扩大发展的有政府部门数据共享、行政记录、互联网数据等。政府统计大数据的流程设计，本质上是全面利用现代互联网信息技术，进一步扩大服务目标。具体流程设计要点是：①政府统计流程设计，首要的是市场经济与社会活动主体的一体化和唯一编码的设计应用，包括个人、家庭、企业、事业单位、政府机构、社会非营利组织机构的各类市场经济参与者。②以生产活动统计为起点，按照国民经济产品和服务标准分类编码推行细化的起点统计，并逐步发展像汽车发动机唯一号码的标准信息运用。③对生产经营、金融交易、社会保障等发票编号建立全国统一标准制度，开发研究个人身份证号码和单位机构编码与发票使用码相关联的反映经济交易的唯一号码生成机制，为复杂的市场交易过程的政府统计自动化和智能化服务。④开发研究市场经济活动的统计词库，包括资源、人口、劳动力、资产、负债、生产、收入分配、消费、投资、金融、出口、进口等各类实际应用的统计词库，并将编码数据化，成为单位之间交易活动的具体统计内容，形成数字化系统自动统计方式。⑤开发研究市场产品、服务、生产要素等价格统计词库和计量单位统计词库，并将编码数据化，成为市场活动价格和数量单位的具体统计内容，形成数字化系统自动统计方式。⑥以会计报表为基础，开发研究单位内容与外部联系的内部核算规范流程设计，并将编码数据化，成为单位内部的具体统计内容，形成数字化系统自动统计方式。

参考文献

[1] Steve Lohr. The Age of Big Data [N]. The New York Times, 2012-02-11.
[2] 张小彦. 大数据与社会管理 [A]. 2012 年清华大学大数据论坛.
[3] McKinsey Global Institute. China's digital transformation: The Internet's impact on productivity and growth [J]. 2014(7).
[4] 陈辉. 智能数据时代的技术准备 [J]. 阿里商业评论，2014(3).
[5] 中国互联网络信息中心. 中国互联网络发展状况统计报告，2014.
[6] McKinsey Global Institute. Open data: Unlocking innovation and performance with liquid information.

作者简介

赵彦云，男，1957 年生，天津武清人，中国人民大学应用统计科学研究中心教授、博士生导师，统计学院院长，中国统计学会副会长，中国统计教育学会副会长。研究方向为经济统计和分析，竞争力与创新指数。

（责任编辑：许亦频）