

# 《统计研究》的历史阶段性回顾与特征分析<sup>\*</sup>

朱建平 刘晓葳 欧阳汉

**内容提要:** 本文基于各个时代背景下我国统计学科的重要事件,联系过去 14 届中青年统计科学研讨会的会议主题,利用文本数据挖掘技术,对 1984 年第 1 期至 2014 年第 5 期《统计研究》刊登的 5192 篇文章的篇名、关键词和摘要等做深入分析。探讨了 30 年来我国统计学科的发展历程,剖析了《统计研究》期刊的整体特征,进一步明确了《统计研究》杂志对中国统计学科建设发展的重要贡献。

**关键词:** 统计研究; 期刊; 特征分析; 文本挖掘

中图分类号: C829.2 文献标识码: A 文章编号: 1002-4565(2014)09-0003-08

## The Historical Review and Feature Analysis for Journal of *Statistical Research*

Zhu Jianping Liu Xiaowei Ouyang Han

**Abstract:** Based on the important events in the statistics discipline under the background of all times in China, relating the themes of the past fourteen statistical science conferences for young and middle aged, and applying text data mining technology, this paper makes in-depth analysis of published 5192 articles in titles, key words and abstracts from 1984 to May, 2014, discusses the 30-years development course of statistics discipline in China, analyzes the overall characteristics of *Statistical Research*, further defines the important contributions of *Statistical Research* to the development and construction of statistical discipline.

**Key words:** *Statistical Research*; Periodical; Feature Analysis; Text Mining

### 一、引言

自 1984 年正式创刊到 2014 年,《统计研究》已走过 30 载春秋,刊登了诸多名家巨擘极富学术价值的重要文献,也记录着一批批优秀青年统计学者的思维轨迹。30 年来,《统计研究》始终坚持“交流科研成果、繁荣学术研究、创新理论知识、推动实际工作”的办刊宗旨,在学术性、创新性、前沿性、理论性方面不断突破,已成为国内统计学科最具权威性的学术刊物,也是统计学者、统计工作者们畅所欲言、交流思想、迸发火花的平台。30 年间,从丛刊到季刊,再到双月刊和月刊,刊物的发行形式发生了明显变化,刊登文章的数量持续增长,文章质量、学术影响力和社会影响力不断提升,而其承载的具体研究内容,充分反映了社会经济演进中统计学科的阶段性特点。

回顾 30 年间发表于《统计研究》杂志的文献,

既是对我国统计学科建设历程的清晰回溯,也是对改革开放巨变、市场经济发展脉络的深入剖析。在《统计研究》杂志创刊 30 周年之际,挖掘分析历年历期所刊登文献的内容,审视《统计研究》历史,总结文献的总体性及阶段性特点,对于把握中国统计学科建设的演变趋势,思考统计学未来发展路径,进一步开创统计科研工作新局面十分必要。为此,本文基于各个时代背景下我国统计学科的重要事件,联系过去 14 届中青年统计科学研讨会的会议主题,利用文本数据挖掘技术,对 1984 年第 1 期至 2014 年第 5 期《统计研究》刊登的 5192 篇文章的篇名、关键词和摘要等做深入分析,提取文章内容的特点、趋势与规律,并总结了《统计研究》杂志对中国统计学科建设发展的重要贡献。

<sup>\*</sup> 本文获国家社会科学基金重大项目(13&2D148)和国家统计局重点项目(2013516)资助。

## 二、分析思路及方法介绍

统计学文献往往体现出具体背景下的阶段性特征,其篇名、摘要及关键词中蕴含的主题与当时社会经济、统计学科发展状况紧密相关,围绕热点问题或前沿问题展开。而文献的研究成果又转而服务于社会经济和统计学科建设,从而推动统计学理论和应用的发展。这种文献内容与现实世界、学术总体水平的联系,在学科最前沿、高层次和最活跃的学术刊物上体现得最为显著。把握《统计研究》文献脉络,需要考察学科建设和经济发展历史,亦需对文本本身做精确的定量分析,文本挖掘(Text Mining)技术是将二者结合起来的有力工具。该技术是针对文本数据进行分析,并在分析过程中产生高质量信息的专门技术。典型的文本挖掘方法包括文本分类、文本聚类、观点分析等,随着大数据时代的到来,更多信息以文本形式储存,使得该技术快速发展并得到广泛应用。本文主要采用词频分析、主题分析和文本聚类等方式,综合研究《统计研究》刊登文献的总体和阶段性特征,以对其 30 年历史做回顾总结。

规范的文本挖掘流程可表述为:文本预处理——文本挖掘与分析——可视化三个阶段。中文文本预处理内容主要包括分词和特征表示,其作用是将文本数据转化为可分析形式,文本信息源的非结构化性质使得数据预处理技术在挖掘中必不可少。文本数据经预处理过程转换为向(矩阵)量形式后,可进一步对生成的词条矩阵做挖掘分析(表 1)。

表 1 中文分词后生成的词条矩阵

词条 \ 文本	1	2	...	k
1	$n_{11}$	$n_{12}$	...	$n_{1k}$
2	$n_{21}$	$n_{22}$	...	$n_{2k}$
⋮	⋮	⋮	⋮	⋮
l	$n_{l1}$	$n_{l2}$	...	$n_{lk}$

表 1 的词条矩阵反映了各篇文档包含的文本信息,其中  $n_{ij}$  代表所分析第  $i$  篇文本中包含第  $j$  个词条的个数,若该文本中并无此词条,则  $n_{ij} = 0$ 。通过生成词条矩阵,原始文本的非结构化信息被转化为可分析的结构化数据,常用的数据挖掘思路和技术便可应用于信息挖掘,如文本结构分析、文本分类、文本聚类、文本相关分析、分布分析和趋势预测等。最后,将挖掘结果由数据可视化(Data

Visualization) 技术呈现出来。

具体研究中,本文以 R 语言为挖掘平台,采用基于 Java 的中文分词系统“rmmseg4j”<sup>①</sup>对《统计研究》5192 篇文章的篇名、摘要和关键词分别做分词,生成词条矩阵。对词条矩阵信息做阶段性分析,提取前 10 个高频词汇,并依据具体的历史背景详细挖掘若干重点词汇与其他词汇的相关性,总结代表性文献的内容,与统计学科发展相结合,以期把握《统计研究》期刊在各个历史阶段的特点与贡献。

需要指出的是,列举高频词汇时,将剔除对中心内容无明显指向作用的词语,包括文献惯用语(如“基础上”、“进一步”、“若干问题”等)或统计学科的常用词语(如“统计学”、“统计分析”等),但为保证文献原信息的完整性,亦结合具体历史背景对“统计学”等学科常用语做词汇相关性分析,以揭示特定阶段的学科研究主题。

## 三、《统计研究》历史阶段回顾与内容分析

创刊 30 年来,《统计研究》的成长伴随的是我国社会经济持续快速发展。以改革开放为起点,以改革创新为核心精神,从计划经济体制向市场经济体制的转变再到市场经济体制的日趋完善,从解决温饱到奔“小康”再到赶超中等发达国家生活水平,中国走的是一条不断解放社会生产力,提高人民生活水平的发展道路。中国统计事业在期间也历经了根本性变革,并产生了大量有益的理论与应用成果,这集中体现在峨眉会议引发的“一门两门”之争、学科建设问题的讨论、“大统计学”思想的提出、统计体制改革、复杂动态系统的统计方法的研究、医学统计、政府统计与公共卫生统计等分支学科的不断完善等重要历史事件中,围绕着以上焦点问题,统计理论的研究深度和统计方法的应用广度不断深化提升。作为统计学科最重要的学术交流平台,《统计研究》的内容形式历经演化更新,刊物内容质量、研究层面和前沿性日趋完善,在不同的社会历史背景下,其刊载文献的主题和成果不可避免地表现出了差异性特征,能够反映中国统计学科不同的发展阶段。

① HUANG Rongui: R interface to the Java Chinese word segmentation system of rmmseg4j, V0.2 - 0 2012.

结合统计学科的发展历史、全国中青年统计科学研讨会的会议主题及对《统计研究》文献内容的文本挖掘结果,可将《统计研究》的主题演化分为四个阶段:第一阶段为1984-1992年,以开展学科建设、落脚统计应用、适应改革开放新形势为主题;第二阶段为1993-2000年,即突破已有统计学科认识、深化统计理论、建设社会主义市场经济;第三阶段为2001-2010年,即加强统计创新、拓展研究方法、拓宽研究领域。第四阶段为2011-2014年,围绕着开展四大工程、加强一级学科、迎接大数据时代展开。经文本挖掘的词频分析,各阶段文本词频分析的结果如表2所示。

表2 《统计研究》高频词汇总

发展阶段	挖掘内容	出现频率最高的十个词汇
第一阶段(1984-1992)开展学科建设,落脚统计应用,适应改革开放新形势	篇名	抽样调查、国民经济、宏观经济、社会经济、统计工作、统计指标、投入产出、指标体系、统计工作、统计理论
	摘要	国民经济、抽样调查、国民收入、宏观经济、社会经济、统计工作、统计指标、投入产出、指标体系、统计理论
	关键词	国民经济、国民收入、社会经济、统计工作、统计指标、投入产出、指标体系、宏观经济、价格指数、统计理论
第二阶段(1993-2000)突破已有统计学科认识,深化统计理论,建设社会主义市场经济	篇名	抽样调查、国民经济、宏观经济、社会经济、统计工作、统计指标、投入产出、指标体系、SNA、国民生产总值
	摘要	SNA、统计调查、国民经济、社会经济、统计工作、社会主义市场经济、统计指标、投入产出、我国经济、国民生产总值
	关键词	SNA、国民经济、社会经济、生产总值、统计工作、统计数据、统计调查、统计指标、投入产出、指标体系
第三阶段(2001-2010)加强统计创新,拓展研究方法,拓宽研究领域	篇名	服务业、竞争力、劳动力、人力资本、统计数据、投入产出、增长率、制造业、CPI、FDI
	摘要	服务业、国民经济、竞争力、劳动力、人力资本、生产率、统计数据、投入产出、增长率、CPI
	关键词	GDP、服务业、国民经济、统计工作、统计数据、投入产出、指标体系、通货膨胀、宏观经济、FDI
第四阶段(2011-2014)开展四大工程,加强一级学科,迎接大数据时代	篇名	CPI、GDP、SNA、城镇居民、非线性、货币政策、通货膨胀、统计表、统计数据、国民经济
	摘要	SNA、非线性、货币政策、计量单位、价格指数、竞争力、人力资本、生产率、通货膨胀、统计表
	关键词	SNA、非线性、国民经济、计量单位、价格指数、竞争力、人力资本、生产率、统计数据、投入产出

第一阶段(1984-1992年)即开展学科建设,落脚统计应用,适应改革开放新形势。该阶段,我国原

有的统计制度尚能适应经济发展需要,但面对改革开放日新月异的社会环境,统计学理论和应用急需更深层次的变革。为促进统计实践的发展,使得统计学能够更好地服务于国民经济需要,学者们基于统计基本理论研究,更多的把目光聚焦在统计学方法的应用上,统计学的实用性和服务性被作为研究重点。因此,实证分析研究与统计方法应用研究,以及开展统计学科建设,成为第一阶段《统计研究》历史文献的焦点内容。

在此期间,文献普遍关注宏观经济的实证分析。从篇名、摘要及关键词的词频结构可以看出,“国民经济”、“宏观经济”、“社会经济”是该阶段文献的热点词汇,不少学者对建国以来的经济运行态势做了实证研究,总结经济发展的客观规律,并提出了一些统计指标作为衡量发展是否合理的标志,得出具有说服力的结论,“统计指标”和“指标体系”共同进入该阶段的高频词范畴。进一步的文本相关性分析发现,“指标体系”一词在文献中与“经济效益”一词关联性最为密切,相关系数为0.27,体现了学者们提倡建立科学合理的统计指标体系,并以此服务于经济发展,也侧面反映了改革开放初期,中国社会发展中以追求“经济效益”为主的基本导向。围绕第三次中青年统计科学研讨会提出的“我国人民生活水平的统计评价”、“解决国民收入的超分配问题”等议题,“国民收入”的有关探讨也成为该阶段文献的热点之一,作者们对我国当时国民收入的分配状况进行了实证分析,探讨国民收入超分配的根源,以“投入产出”等具体方法作为分析国民经济问题的主要工具。具有特殊时代意义的是,1979年12月邓小平首次提出的“小康社会”概念成为1984-1992年杂志刊文的讨论焦点之一,虽然“小康水平”、“小康生活”并未直接进入高频词汇列表,但“小康”一词,与“人均收入”、“国民收入”、“国民生产总值”、“人民生活”、“物质财富”、“消费水平”等描述和反映国民经济总体状况及人民生活水平的关键词相关性较强,相关系数分布于0.35~0.9之间,针对这一极富中国特色和时代特征的定义,作者们从统计意义、统计方法、统计指标、基本标准等各个方面做了综合分析,为建设社会主义现代化进程中这一重要标杆提供了深入的理论支持。

文献在统计分析方法应用方面也有所创新。在此期间,紧密联系三次中青年统计科学研讨会的重

要议题,期刊上刊登了诸多有关“通货膨胀的测定”、“指数理论完善”、“企业统计改革”领域的文献,产生了高水平的统计方法研究成果。围绕高频主题“价格指数”,作者们对西方国家价格指数的有关理论加以介绍,探讨价格指数构建问题,研究居民生活与价格指数之间的统计关系,检验价格指数的指示性,该主题词与“居民生活”、“西方国家”、“消费品”等词的相关性均在0.25以上。“抽样调查”也是统计方法讨论中的焦点,理论方面,作者们对抽样调查方法的随机性和局限性做了研究,同时与典型调查等其他方法做了比较;应用方面,重点以农产品产量为实证资料,对抽样调查的应用加以推广。对企业统计改革的贡献主要体现于企业经营管理中的应用研究,企业统计的相关文献有助于使企业充分认识到统计工作的价值,以达成企业统计改革的主要任务;同时,构建指标体系是指导企业运营的重要统计应用途径,在涉及“统计指标”、“指标体系”等主题的文献中,企业“经济效益”、“经营者”、“销售总额”等词汇也往往关联出现,建立完整的指标体系成为本阶段企业改革的重要内容与主要成果。

在实证分析与方法应用研究取得一定成果的同时,作者们还就统计学科体系建设问题进行了探讨。分析文献高频主题“统计理论”及“统计工作”,关联度较高的主题为“现代化”(0.28)、“体制改革”(0.21)、“统计法”(0.17)、“分类”(0.15)、“统计监督”(0.13)等。一方面,作者们探讨了统计学科定位,认为社会经济统计学和数理统计学分属经济学、数学下属学科的状况不利于统计学的发展。另一方面,指出统计工作者应开阔视野,广泛地了解、深入地研究国际上统计学科的发展情况,并加以学习和借鉴,对我国统计学科的发展进行客观评价,找出问题和不足,明确发展方向,推动统计学科现代化建设。第三,也强调了统计工作中的法制规范和监督问题。这些有关统计学科建设的重要文献,较为全面地提出了统计学科发展的几大关键问题,即学科定位问题、现代化统计学科建设问题及统计工作的规范问题,引发了统计学者们广泛而持续的讨论,也为第二阶段进一步加强学科建设奠定了良好的理论基础。

第二阶段(1993-2000年)即突破已有统计学科认识,深化统计理论,建设社会主义市场经济。1992年党的十四大正式提出了建立有中国特色的

社会主义市场经济,我国开始进入了由市场经济体制逐步取代计划经济体制的转轨时期。统计学科建设和统计科研在许多方面面临双轨体制下经济形态变化的严峻挑战,中国新的发展阶段对统计学科、统计理论及应用提出了更高要求,与之相适应,统计形势与任务发生巨大变化。因此,加强统计学科建设,深入开展统计理论研究成为本阶段的核心内容。分析该阶段文献我们发现,承接上一阶段的讨论热点,“国民经济”、“社会经济”、“统计工作”、“统计指标”等词依旧是篇名、摘要及关键词中的高频主题,而讨论“社会主义市场经济”的有关文献的出现,符合该阶段鲜明的历史背景。

首先,“国民生产总值”、“社会主义市场经济”(或“市场经济”)等热点主题的出现,深刻反应了《统计研究》刊登文献内容的时代性特征。为响应十四大建设中国特色社会主义市场经济的号召,统计学者们针对该问题做了广泛探讨,在刊物上发表了有价值的学术成果。词频相关性分析显示,篇名中,与“市场经济”一词关系密切的为“我国社会主义”、“条件下”、“企业”、“体制改革”,揭示了市场经济的发展条件、背景以及发展市场经济的主要动力和基本途径;摘要及关键词中,“多元化”、“大中型企业”、“提高生产率”等词语与“市场经济”相关系数均超过0.30。而“国民生产总值”一词则多与“经济实力”、“世界排名”、“增长率”、“居民收入”同时出现,表示GDP成为厘定经济发展目标的主要标准。

其次,本阶段统计学科最突出的成就是统计认识上有了新的突破。社会主义市场经济的发展要求大胆吸取国外统计科学研究的最新成果,善于从丰富的统计改革实践中吸收营养,解放思想,转变观念,探索统计学发展的新思路。统计研究要超前于统计实践,也要服务于统计实践,深入研究,解决统计改革中出现的新问题。要加强统计基础理论的研究,以完善学科建设,丰富和充实学科体系。随着广大统计工作者的不懈努力,1992年11月1日,国家技术监督局发布了中华人民共和国国家标准《学科分类与代码》,将统计学列为一级学科予以公布,它极大地推动了我国统计科学的发展,成为中国统计学科建设的重大突破和历史里程碑。该时期《统计研究》各文献的摘要中,“学科分类”、“不同层次”、“各学科”三词均与“统计学”相关性较强,相关系数

分别为0.25、0.24和0.24;而关键词中,“设计理论”、“学科分类”是与“统计学”相关性较强的词汇,相关系数为0.23;除此之外,文献较多讨论了统计学课程设置问题,篇名中“课程内容”与“统计学”的相关系数达到0.30,“高等学校”、“标准化”等词语则与高频主题词“统计工作”强相关,可见该阶段文献围绕着统计学科建设问题主要进行两方面的工作:一是确定统计学的研究范围,构造统计学的二、三级学科体系;二是在具体工作中反映统计学的地位和作用,如统计教育改革问题和高等教育中关于统计学课程及专业设置问题。

新形势下国民经济核算的理论和方法是该阶段的另一热点议题。从长期采用MPS的框架组织国民经济核算到向SNA全面接轨,是20世纪90年代国民经济核算的重要特征,这种转变适应社会主义市场经济发展的需要,具有重要理论意义和现实意义,《统计研究》该阶段也刊登了大量以SNA为主要议题的论文。一方面,全面接轨期,诸多全新问题值得探讨,另一方面,作者们以不断发展的视角研究SNA相关理论。“SNA”成为文献中的高频词汇,而与其相关性较强的词汇则为“MPS”、“区别”、“新方法”、“账户”、“经济核算”、“生产者”、“总供需”、“投入产出”等,具体文献涉及SNA与MPS的比较、生产劳动与生产范围问题、社会总供需的测算与分析、投入产出核算等,这一时期,在《统计研究》这一优秀的学术平台上,统计学者和工作者们贡献了大量关于SNA研究的理论成果,为我国进一步适应市场经济需要和早日实现与国际SNA接轨做出巨大贡献。

第三阶段(2001-2010年)即加强统计创新,拓展研究方法,拓宽研究领域。新世纪我国统计发展的战略核心取向在于统计创新,包括理论创新、体制创新、制度创新、方法创新等。在此期间,统计学科在统计理论和统计学科研究方面取得了令人瞩目的成果,在实践中也取得了一定的成效,拓宽了统计学的研究领域。相对于前两阶段,这一时期《统计研究》刊登的文献集中体现出创新性、广泛性的特点,现代统计分析中的新方法被引入和扩展,统计学的实证应用面更宽,通过对该阶段文献的文本挖掘,可以发现一些更为具体的实证领域或方法词汇成为高频的主题词汇。

首先,刊物继续讨论统计学思想、统计学科性质

及分类问题,并取得了新的认识。“统计工作”仍然是该阶段的高频词汇,值得注意的是在文章篇名中,“迫在眉睫”一词与“统计工作”一词相关性达到0.5,表明作者们意识到新世纪继续大力发展统计工作的重要性。部分文献从哲学的高度来认识统计,认为统计是一种思想,统计的地位甚至可以与哲学相提并论,只不过各有分工而已。在统计学性质的认识上,对大英百科全书的观点则得到更多的认可,认为统计学虽有数理统计学和经济统计学之分,但总体上均属方法论的科学。数理统计学是研究随机性数据的方法论科学,经济统计学是研究确定性数据方法论的科学。有的学者谈到,应把统计学独立出来,发展统计方法。统计学不是纯数学,也不是经济学,应当在应用中发展统计学。可以说,对于统计学性质、学科分类和统计核心思想的讨论,促进了学科进一步发展,为统计学在新世纪注入了新的活力,使其能够更好地服务于社会经济的长足进步。

其次,统计理论研究较前两阶段呈现出新特点。“宏观经济”、“国民经济”、“投入产出”、“通货膨胀”、“CPI”仍然是这一阶段理论研究的热点,但具体研究方法有所更新。以“通货膨胀”为例,这一时期刊登的文献中,与该热点词相关性超过0.1的包括“SVAR”(结构向量自回归)、“VAR”(向量自回归)、“BSVAR”(贝叶斯结构向量自回归)等大量方法指向性词汇,反映该阶段作者们较多运用一系列前沿的统计量化分析方法解决具体问题;同时“可容忍”、“国际石油”、“能源安全”、“中长期”、“变异性”等词语与通货膨胀问题的相关性亦较高,显示该领域的理论研究无论是思想还是研究范围都有所延伸。同时,“非线性”一词更多地出现在该阶段统计方法的研究中,其思维也被用于研究“房地产价格”、“通货膨胀”、“通货紧缩”、“物价水平”等经济生活的方方面面,代表性模型“STAR”(平滑转移自回归)模型被引入、扩展和应用,说明经济的非线性现象较之前一阶段更为受到国内学者们的重视,相关方法也有了补充和拓展。另外“数据挖掘”(或“data mining”)亦是本阶段出现频次较高的方法指向词,在篇名中、关键词、摘要中分别出现15次、20次和12次,文章有对数据挖掘技术的综述及对国外先进方法的介绍引入,有对数据挖掘和统计学关系的探讨,也有新方法的提出及结合金融市场等具体经济现实的应用,从理论与实践角度介绍了数据挖

掘这一新的数据分析技术,为大数据时代的到来构筑了必要的技术基础。

第三 统计学应用领域有极大扩展,作者们在文献中更多地运用统计理论和技术方法解决实际问题,使人们真正感受到统计的重要性,这无疑充分发挥了统计学的应用价值,提高了统计的地位,进一步扩大了学科影响力。“服务业”、“劳动力”、“人力资本”、“竞争力”、“生产率”、“增长率”、“FDI”、“制造业”等词汇在摘要和关键词中成为热词,这些应用领域指向的词汇的出现,代表统计理论和方法的应用也正在向社会经济各个领域渗透,具体文献内容开始广泛涉及收入分配、金融证券、宏观经济分析、国际资金循环、外商直接投资、统计综合评价、城市可持续发展、产业发展、市场研究、最低工资、投资者信心指数、进出口问题、人力资本、城乡居民消费状况、民间投资与股价动态、社会保障支出与扩大内需、医疗卫生服务产出、工业污染物排放影响因素等。

第四 对政府统计有了全新看法,提出政府统计工作的改革问题。“统计数据”成为在篇名、摘要和关键词中均高频出现的主题,围绕2009年夏季举办的第十二次中青年统计科学研讨会中关于“政府统计的全新认识与公信力提高的问题”、“统计能力、统计数据质量的提高问题”等重要议题,诸多文献从统计数据质量评估、统计制度和统计监督及全面提升统计数据质量、提升政府公信力等角度讨论了政府统计工作改革问题,这一时期,与“统计数据”相关性大于0.1的词汇包含“标准化”(0.17)、“努力提高”(0.17)、“准确性”(0.17)、“可靠性”(0.17)、“理论研究”(0.13)、“质量问题”(0.13)等。针对当前的政府统计,作者们认为统计工作应该由独白转换为对话,应该建立调查对象视角的数据采集系统,建立受众视角的传播体系,提高政府统计的公信力及统计数据的质量问题。同时也提出统计的变革和发展需要更多的“传教士”式的统计教育者,这样可以帮助学生更加积极地看待统计学在日常生活中的价值。

第四阶段(2011-2014年)即开展四大工程,加强一级学科,迎接大数据时代。“十二五”期间,是加快统计事业科学发展的关键时期,是深化统计改革、推进统计建设的攻坚阶段,这一时期,统计学科也面临大数据时代带来的新变化。抓住统计事业发

展的重要战略机遇期,在奋力提高统计能力、提高统计数据质量和提高统计公信力的同时,加快“四大工程”建设,即建设基本单位名录库、企业一套表制度、数据采集处理软件系统和联网直报系统。这一阶段的文献,统计学方法和应用的前沿性继续提升,政府统计改革与发展有关问题继续深化,几篇国内最早系统讨论大数据与统计学关系问题的文章也在期刊上刊载,可以说,新时期《统计研究》的文献为统计学一级学科的建设及大数据时代统计学科的发展方向奠定了基础。

努力推进“三个提高”加快“四大工程”建设方面,一些作者从宏观角度探讨了对当前统计工作的体会和认识;从政府统计的发展目标、机遇和挑战等方面进行了分析;从统计法制建设、统计调查制度、统计标准的建立、统计调查手段改进、统计服务建设等方面总结了中国政府统计做出的改革创新;依据“三个提高”和“四大工程”建设的构想,探索了政府统计改革与发展的思路和方法。2011年第7期、第11期以及2012年第7期,分别刊载了以“四大工程”为篇名中心词的文章,重点探讨法制建设、信息化建设及统计系统建设等各项内容。除此之外,还有详细讨论“建设基本单位名录库”的文章3篇,“企业一套表制度”相关文章5篇,以及“联网直报系统”文章1篇。

统计学一级学科地位的进一步确立是该时期统计学界的重大事件,2010年国务院学位委员会办公室、教育部研究生司启动了改革开放后的第四次研究生专业目录调整工作,最终审议通过将原列为经济学一级学科门类“应用经济学”下的“统计学”二级学科,与原列为类一级学科“数学”下的“概率论与数理统计”二级学科合并为“统计学”,并确立其一级学科地位,可授予理学或经济学学位。自此,统计学学科正式在研究生层面上成为一级学科,并继1992成为科研科技统计专业目录上的一级学科之后,又成为教育专业目录上的一级学科,这无疑将对我国统计教育改革、统计学科发展和统计制度建设形成极大的促进作用。在此基础上,以《统计研究》为交流平台,作者们对加强一级学科建设提出了各种宝贵意见,在篇名、关键词和摘要中,“一级学科”出现12次,“统计学科”出现7次,“学科建设”出现2次,其他与统计学科建设有关的词汇(或短语),如“统计教材建设”等,也频繁出现。其中有代表性的文章如2011年第11期

的刊文,在回顾中国统计学科百年历史的基础上讨论学科现状,并提出要抓住机遇,在做大统计学科的基础上进一步加强学科建设。

在国民经济核算研究方面,“SNA”依旧是热词,这一阶段对我国国民经济核算体系的探讨主要集中在改进建议等方面。一些文献从我国实际出发探讨全球化对中国国民经济核算体系的影响;提出了我国产业分类体系的改进建议和讨论了核算中的数据衔接问题。与“SNA”相关程度较高的词分别为“国民经济”(0.28)、“社会保险”(0.23)、“新面貌”(0.23)、“中央银行”(0.23)。

针对统计理论与方法应用的研究,在统计方法应用面继续不断扩展的同时,前沿性方法在文献中的出现频次有所增加。如“非线性”首次成为词频最高的前10个词汇之一,而文献中非线性模型的应用也较上一阶段有了明显扩展,非线性现象的研究,从时间序列领域延伸到面板数据领域,“PSTR”(面板平滑转换回归)这一研究面板非线性的前沿模型被介绍到国内统计学界,而“FDI”、“通货膨胀率”、“金融危机”等指示应用领域的词汇和该热词联系紧密。

值得注意的是,大数据时代的到来,引起了政府、业界的极大关注,也将给统计学科带来长远影响。2013年10月召开的第十七次全国统计科学讨论会正是以“大数据背景下的统计”为主题,第十五次全国中青年统计科学研讨会也将大数据有关内容当作理论部分的重点,可见统计学界对大数据问题的广泛重视。2014年《统计研究》新年献词中,“大数据”被解读为即将到来的新时代,强调了其对统计学可能带来的深远影响,同时第1、2期刊物也连续刊载了4篇系统探讨大数据与统计学思维、大数据时代统计学科面临的机遇与挑战的文献,可以说在这一时期,《统计研究》同样成为了国内统计学者、工作者最先对大数据发声的平台,走在了大数据时代统计学科的最前沿。

#### 四、整体特征分析与总结

通过对《统计研究》历史内容的阶段性回顾,发现《统计研究》与经济社会发展阶段联系十分密切,作为学者们交流思想的平台,刊载的文献适应社会经济发展需要,贡献了大量有理论价值和现实意义的学术成果。《统计研究》始终是我国统计科研前沿阵地,是我国统计学科建设、统计方法研究及统计

领域改革的导向性刊物,也促使统计实践工作在社会各领域发挥应有作用。总结30年发展历程,《统计研究》有以下几个重要的整体性特点。

(一) 研究领域广泛、热点主题突出、理论与应用相结合

在5192篇文章中,理论研究论文约2050篇,占比39.48%;应用研究论文3142篇,占比60.52%,体现了期刊在选文时坚持理论与应用相结合。所刊文献内容覆盖面广,涉及“中国知网”划定的40个学科分类,大类有数学(1375<sup>①</sup>)、社会学及统计学(1103)、宏观经济管理与可持续发展(1014)、经济统计(815)、经济体制改革(501)等;按照应用领域,则包含金融(463)、投资(290)、贸易经济(171)、企业经济(145)、工业经济(130)、证券(125)、农业经济(112)、财政与税收(89)、市场研究与信息(62)、保险(62)、人才学与劳动科学(56)等。若以统计学重要主题分类,文献对基本理论问题、方法应用问题、核算问题、经济分析与统计分析问题都有重点关注,内容完整、主题鲜明。期刊牢牢把握统计学科发展的主旋律,以词云分析<sup>②</sup>(图1)为可视化方法研究文献题目,可以发现统计指标、统计数据、统计调查、通货膨胀、GDP、统计分析、社会经济、核算等统计学科重点问题受到作者们的广泛关注,使得期刊能准确记录、反映统计学科核心问题的研究脉络,为任一时期的研究者们提供有效参考。



图1 文献题目的词云分析

(二) 学术影响力大、研究成果传播面广

《统计研究》杂志的发展,伴随的是其影响力的不断扩大。以CSSCI(2014-2015)收录的期刊为研究范围,2010-2013年间,杂志影响因子连年提升,

① 篇数,在该分类方法下,一篇文章可能从属多类。

② Tagxedo Creator <http://www.tagxedo.com>。

从 1.959 升至到 2.567, 四年平均值 2.111, 高于其他三种统计学期刊(图 2); 同时考虑经济学、管理学、统计学、社会科学综合四类期刊,《统计研究》在 226 种期刊中四年平均影响因子排名第 30 位。这里需要提及的是,2013《统计研究》再次入选“中国最具国际影响力学术期刊(人文社会科学类)”,排名由 2012 年的第 29 位上升至第 24 位,具有较为广泛的影响力,为传播统计学研究成果提供了强有力的渠道保障。另外,文献一直保持着较高的引用量和阅读量,以被引次数最高的 600 篇文章为例,单篇平均引用数达 62.26 次,平均下载量达 842.22 次,进一步发掘引用频次较高的文章,发现主要涉及金融市场、工业经济、企业经济、能源经济、综合评价、统计基本理论等研究维度,充分反映《统计研究》促进统计学方法与应用相结合,有效推动了统计实践在社会经济运行中发挥其重要作用。

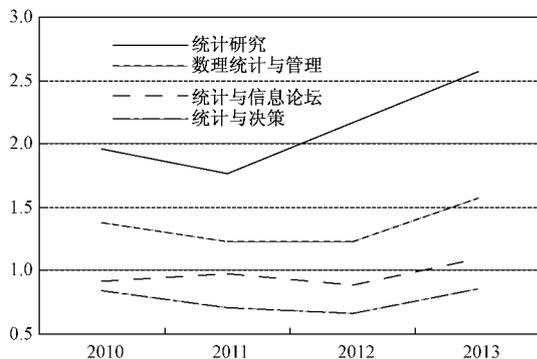


图 2 2010-2013 年统计学期刊影响因子对比

(三) 为高校学者提供良好平台,有效推动学科发展和统计教育改革

一方面,《统计研究》为高校学者们提供了公正、广阔、高质量的学术平台,促进高校统计学科建设,是分享、输送统计学知识,为社会经济发展服务的重要渠道与载体,按照 30 年间期刊刊文的第一作者隶属单位统计,高校作者论文 2722 篇,占比 52.4%。另一方面,发表于《统计研究》上的文献与各届中青年统计科学研讨会议题契合度高,期刊吸收了大量有见地、有研究热情和学术潜力的中青年统计学家的有益观点,不失为培养统计学高端人才的沃土。第三,对文献内容的阶段性回顾揭示,期刊刊载了大量探讨统计教育发展改革的文章,涉及内容、方法、教材和体制等方面,“统计教育”一词在文献题目、关键词及摘要中分别出现 11 次、107 次和

31 次,这些对统计教育的有益探讨,无疑对推动学科发展和教育改革起到了重要作用。

以上回顾总结尽可能从文献本身出发,力求客观准确地反映《统计研究》30 年足迹,把握统计学科历史脉络,但限于篇幅,文中所论难免挂一漏万,在运用文本挖掘技术时,基于不同的挖掘角度,也可能得到不尽相同的信息。但毫无疑问的是,30 年来《统计研究》不断进步、完善,为我国统计学者、业界工作者们提供了良好的学术平台,在学科的建设中也起到了愈加重要的作用。展望大数据时代,统计学科在理论与实践方面均面临着更高层次的机遇与挑战,《统计研究》将一如既往地担当学术交流、知识输送的重任,助力中国统计研究事业更上新高。

参考文献

[1] Michael W. Berry, Malu Castellanos. Survey of Text Mining: Clustering, Classification, and Retrieval [M]. Springer 2007.

[2] Meyer D, Hornik K, Feinerer I. Text Mining Infrastructure in R [J]. Journal of Statistical Software, 2008, (25).

[3] Gupta Vishal, Gurpreet S. Lehal. A Survey of Text Mining Techniques and Applications [J]. Journal of Emerging Technologies in Web Intelligence 2009(1).

[4] 朱建平著. 世纪之交中国统计学科的回顾与思考[M]. 中国经济出版社, 1999, 12.

[5] 韩明. 《统计研究》论文作者数量的统计分析[J]. 统计研究, 2002(2).

[6] 肖红叶. 中国经济统计学科建设 30 年回顾与评论——基于三大事件框架的研究[J]. 统计研究, 2010(2).

[7] 邢志强. 《统计研究》论文作者、合著及引文分析[J]. 统计研究, 2001(6).

[8] 袁卫. 机遇与挑战——写在统计学成为一级学科之际[J]. 统计研究, 2011(11).

[9] 朱建平、戴颖. 全国中青年统计科学研讨会的回顾与思考[J]. 统计研究, 2010(1).

作者简介

朱建平,男,2003 年获南开大学理学博士学位,现为厦门大学两岸关系和平发展协同创新中心和经济学院教授,博士生导师,厦门大学数据挖掘研究中心主任。中国统计学会副会长、教育部高等学校统计学类专业教学指导委员会秘书长、中国统计教育学会常务理事。研究方向为数理统计、数据挖掘。

刘晓葳,男,现为厦门大学经济学院统计系博士研究生。研究方向为数据挖掘、计量经济模型。

欧阳汉,男,现为厦门大学经济学院统计系博士研究生。研究方向为数据挖掘、计量经济模型。

(责任编辑:程 晔)