

# 大数据环境下财政审计数据分析研究<sup>\*</sup>

裴文华 成维一

【摘 要】网络、移动通信把人类社会带入了一个大数据时代，利用大数据进行审计数据关联分析对提高审计效率和能力有非常重要的意义，但目前对大数据环境下财政审计数据分析的研究较少。基于此目的，以财政审计数据分析为着眼点，结合审计实践，首先，阐述了财政审计数据关联分析中的三种分析思路 and 模式，即以财政资金为核心、以企业为核心和以人员为核心探讨大数据背景下财政审计数据分析的思路和方法；其次，从数据共享、数据标准化、数据多样性和审计人员自身能力等方面探讨了当前数据分析中的难点；最后，要做好大数据环境下的财政审计，先要树立大数据环境下数据分析的意识，提前做好审计数据分析规划，集中解决财政审计中的难点问题，同时要不断提升审计人员数据分析能力，通过大数据的关联分析促进财政审计效率的提升。

【关键词】大数据 财政审计 关联分析

要建立健全与审计全覆盖相适应的工作机制，提高审计能力和效率，只有创新审计技术方法才能实现审计全覆盖。在大数据和审计全覆盖的背景下，如何做到财政审计方法创新，与时俱进，是值得我们深思的问题。本文结合财政审计的实践，就大数据环境下如何做好财政审计数据分析，谈一些思考。

## 一、大数据环境给财政审计数据分析带来的影响

### （一）大数据概述

“大数据”并非一个确切的概念，最初，这个概念是指需要处理的信息量过大，已经超出了一般电脑在处理数据时所能使用的内存量，因此必须改进处理数据的工具。这导致新处理技术的产生，如谷歌的 MapReduce 和 hadoop 平台，这些技术使得人们可以处理的数据量极大地增加。更重要的是，这些数据不需要像传统的数据库中以表格整齐地排列。2013 年，维克托·迈尔·舍恩伯格和肯尼斯·库克耶提出，大数据不用随机分析法（抽样调查）的捷径，而是采用所有数据进行分析处理，即“样本 = 总体”。普遍认同大数据有“4V”特点，即 Volume（大量）、Velocity（高速）、Variety（多样）、Value（价值）。后经学者发展完善，将“4V”理论中最后一个“V”改为 Veracity（真实性）。国内学者从不同角度对大数据给出了定义，如秦荣生（2014）认为大数据指的是所涉及的数据量规模大到无法利用现行主流软件工具，在一定的时间内实现收集、分析、处理或转化成为帮助决策者决策的可用信息。杨凯茜（2015）认为大数据是用传统的方法不能直接获取的来自于人、机、物的大量网络数据的集合，这些数据包括结构化数据、半结构化数据和非结构化数据，具有大量化、多样化和快速化的优点。

### （二）大数据给财政审计数据分析工作带来的转变

#### 1. 深入有效地利用结构化审计数据的必要性

一是提高现有数据的价值。大数据的价值不在单纯源于它的基本用途，而更多源于它的二次利用。随着审计数据定期报送制度化，数据集中力度将更高，加之历年审计项目中积累的各种数据，审计可用的数

\* 裴文华、成维一，审计署驻兰州特派员办事处，邮政编码：730030，电子信箱：peiwenhua@126.com。

据将越来越多,如何将现有多行业、多部门、多类型数据利用好,是当前财政审计数据分析的要务。二是应更加注重数据关联分析。数据的总和比部分更有价值,当多个数据集的总和重组在一起时,重组总和本身的价值也比单个总和更大。目前开展的财政审计数据分析已经实践了多方数据关联分析的有效性。只有加大财政业务数据与行业数据以及跨行业、跨领域数据的综合比对和关联分析力度,才能提高运用信息化技术查核问题、评价判断、宏观分析的水平。

### 2. 重视非结构化数据的利用

目前财政审计数据分析还主要停留在部门结构化数据的查询分析上,非结构化数据的利用还在初级阶段,但我们要认识到传统的数据分析方法在大数据环境下的变革,非结构化数据利用的重要性。如 2009 年,谷歌公司通过分析 5000 万条美国人最频繁检索的词条,并将其用于特定的数学模型,准确及时地判断了 H1N1 流感是从哪里来,预测结果与官方数据的相关性高达 97%,但比官方获得数据还快一两周。因此,大数据环境下,利用网络检索与审计相关的信息,建立分析模型,找出疑点线索,这给审计数据分析工作带来一个新的思路,即拥有“大数据”思维。

### 3. 审计模式的转变

一是审计范围由“抽样审计”向“全量审计”转变,审计模式也将会转向全数据审计模式。当前财政审计数据分析由于受到人力、时间和分析条件的限制,无法做到对所有审计数据的收集和分析,所以在审计项目中一般采用依赖于审计重要性水平的审计抽样方法。而全数据审计模式可以精确定位问题所在,减少抽样审计带来的片面性和局部性,让风险控制更有针对性。二是审计数据分析呈现常态化,数据分析不只局限在审计项目开始时,而将在整个项目年度持续开展,一线作业与后台数据分析融合更加紧密,单点离散审计向多点联动审计转变,现场审计向现场审计与非现场审计相结合转变。审计人员将采用新技术,从海量数据中挖掘相关审计证据,提高审计能力、质量和效率,拓展审计监督的广度和深度。

## 二、财政审计三类数据关联分析的思路 and 模式

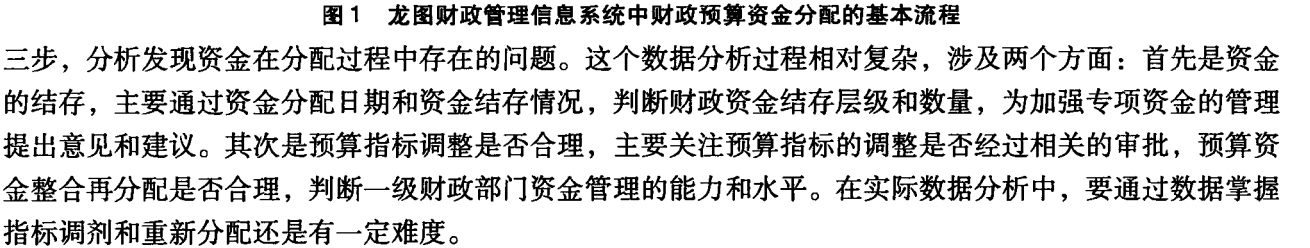
### (一) 以财政资金为核心的分析

#### 1. 数据分析思路和目标

以财政资金为核心的审计分析,要结合审计项目特点,摸清资金的总体规模、理清每笔资金的去向,判断其合规性和效益性等相关审计目标。沿着资金流向,加强对财政、国库、财政专户、预算部门等相关单位的审计数据分析,按照资金流向发现财政资金分配、管理和使用中存在的违规违法违纪和其他突出问题,从体制、机制和制度层面深入分析原因,提出改进和加强财政资金的管理、完善相关制度措施的意见和建议。

#### 2. 数据分析思路

在数据分析时,以被审计对象所在的同级财政部门的财政数据为源头,按照资金流逐层分析。目前财政部门使用较多的是龙图、用友和太极华青等财政管理信息系统。本文以龙图一体化财政管理信息系统为例,说明财政预算资金分配过程(其他财政管理信息系统原理与该系统原理一致),如图 1 所示。第一步,了解省级财政收到上级财政(即财政部)某一专项资金总额。龙图财政管理信息系统中,设计了一个控制表,包含 FromCtrlID 和 ToCtrlID 两个字段,通过这两个字段与其他表关联(在审计署 2014 版财政标准数据中,分别称为来源控制号(LYKZH)和目标控制号(MBKZH),含义和作用都一样)。通过这两个控制号将财政资金分配、支付、记账相关的预算指标表、用款计划表、支付申请表、资金支付表、总预算会计凭证表等关联起来。按照财政资金的分类,以来源控制号为 0、目标控制号不为 0 为筛选条件,可以掌握财政部下拨给各省区的专项资金总额。第二步,分析预算资金在省级财政部门的分配和下拨情况。财政部下拨的资金作为省级财政部门的总指标,分解后进入处室指标,通过处室指标分解进入部门指标和下级财政指标。分析仍然依靠上述的两个控制号,在不考虑指标增减变化等情况下,下一级资金的来源控制号对应上级部门的目标控制号。依据这个关系,设置查询条件就可以理出资金实际分配的情况和流向。同时要结合财政资金的性质,掌握是否需要省市资金配套(在数据中通过处室指标的分解可以看出)。第



## （二）以企业为核心的分析

此类分析主要关注财政补贴资金使用的真实性和效益性，享受补贴企业是否符合条件，掌握企业接受财政补贴资金的原因和资金使用效果，通过关联分析，找出不符合规定企业领取补贴的情况。省级财政数据中对企业 and 个人的付款，通过财政数据中资金支付表中的收款人来判别。这种分析思路适合财政直接补贴企业的专项财政资金审计，如淘汰落后产能专项资金、新兴产业扶持资金等。

以 2013 年 A 省省长经济责任审计中淘汰落后产能中央财政奖励资金审计为例，说明以企业为中心的财政资金关联分析模式。通过分析近年淘汰落后产能企业名单，以关停企业为重点，按照大数据审计分析思路，关联多方数据，发现企业在申报淘汰落后产能奖补资金过程中，通过伪造虚假申报材料，共骗取中央财政淘汰落后产能奖励资金 935 万元的问题。

在数据分析时，首先检查享受补贴企业真实存在性。通过与工商、税务、电力数据关联，查看上述数据库中是否存在享受补贴的企业，如果不存在上述企业，应怀疑其属于编造，进入延伸核实名单。其次检

查企业享受财政补贴资金的合规性。已关停企业冒领补贴问题,通过企业的工商注销和年检信息,获取注销年度和最后一次年检年度,与应关停年度比较,判断是否以前年度已关停,虚报冒领骗取财政资金。应关未关企业问题,将获得奖补资金应关停企业作为核心,核实这些企业用电量信息、工商年检信息、税收数据,找出违规企业名录进入延伸核实名单。最后是检查其余应部分关停企业在关停年度及以后的纳税和用电数据变化情况,分析营业收入是否长期处于较低水平进入延伸核实名单。

### (三) 以人员为核心的分析

#### 1. 分析思路

以人员为核心的数据分析主要涉及两类:一是财政对个人的各类补贴资金的审计核实,如城乡最低生活保障金、惠农贷款和农资综合补贴等。这些业务无法在关联部门核实,只能通过与外部数据的关联比对,确定补贴资金发放的合规性。二是分析公务人员利用职务便利谋取非法利益问题。如公务人员开办企业,违规获得财政补贴、变相利益输送等重大违法违规问题。

#### 2. 数据分析模式

一是人员表的归类整理。涉农资金数据的质量差和不完整等问题比较突出,应先对需要分析的数据标准化整理,形成符合比对要求的汇总表。若数据质量高,可只进行简单的数据完整性验证后直接开展分析。二是与外部数据的关联。不论是个人领取补贴还是财政公职人员投资办企业,这两种分析思路基本一致,都需要以人员身份证号和姓名作为连接条件,与财政供养人员、车购税、工商、税收等外部数据关联,分析比对出须核实的人员名单。三是结果的筛选和核查。由于现实情况的多样性,通过重要性水平筛选出的分析结果,要认真核实,尤其是在大数据的背景下,数据分析结果与现实还存在一定的差距,对最终的结果要认真核实分析。如我们在 B 省粮食直补和农资综合补贴资金审计中,数据分析发现获得补贴人员身份证号码错误的情况,怀疑存在虚构人员骗钱财政资金的问题,选择了 2 个县区的 4 个乡做延伸时,经办人告知审计人员,确实编造了身份证号码,原因是多数村民外出打工,拿不到身份证号,为及时发放财政补贴资金,编造了人员身份信息,但业务真实有效。

### 三、大数据环境下财政审计数据关联分析存在的难点

#### (一) 种类繁多、数据孤立,数据共享难

从当前的财政审计数据分析工作看,数据信息共享在一定层面上仍然存在障碍和阻塞。一是各行业审计之间数据共享不够。不同行业数据管理呈现分散割据式的特点,数据的存放、保管基本是“谁审计哪个部门、谁负责收集数据”,数据使用的割据现象普遍存在。不同行业的数据关联软件也互不相关,无法将相关业务数据高效、完整地整合在一起。二是同一部门数据共享不够。以财政数据为例,全国财政管理信息系统各不相同,省市县三级财政数据管理平台也各不相同,一笔财政资金从中央财政到省市县乃至每个乡镇的资金使用从纵向无法直接实现关联,数据没有共享,导致无法通过计算机直接进行查询。

#### (二) 良莠不齐、结构复杂,数据标准难

当前财政审计数据关联分析要依赖标准化数据,只有数据可用,才能使用计算机技术实现自动化程度较高的分析。数据标准化受审计人员自身能力和行业数据本身两方面因素制约,目前高度信息化与低度信息化单位并存。信息化程度高的部门,业务处理复杂,审计人员对数据和信息系统的理解和掌握有个过程,通过数据流还原业务流难度较大,因而数据关联分析的难开展。同时,还有很多行业部门自身信息化程度低,数据环境差,所提供数据不完整或数据碎片化严重,给数据使用带来困难。如财政资金支付数据中收款企业名称,有简称、有全称,形式多样,与工商、税务、电力数据系统中企业名称不一致,无法通过企业名称直接关联,这都影响了财政审计数据分析的效率和质量。

#### (三) 纷繁多样、优劣掺杂,数据利用难

互联网时代,网络、移动通信、存储的快速发展把人类社会带入了一个以 PB 为单位的结构与非结构数据的新时代。互联网数据获取更加容易、便捷,输入检索字段,通过搜索引擎可以获得上亿条数据;数



据结构更加多样,尤其是多元异构跨领域的大数据使得审计人员面对的数据分析范围发生了扩增,从原先局限的被审计单位内部数据扩展到与其相关的所有外部数据,这就带来另一个问题,审计人员如何把隐没在海量异构、杂乱无章的电子数据信息集中、萃取和提炼出来,揭示其内在的规律,从中分析挖掘找到审计人员所需的数据,也是步履维艰。

#### (四) 数据量大、查询频繁,分析环境建设难

大数据背景下,对财政审计数据分析环境提出了新要求,传统的数据分析平台已不能承载海量数据的分析工作。按照目前国家审计数据分析网设计思路,成立审计署审计数据中心,各特派办经授权后通过数据专网登陆到署审计数据中心查询,这种方式对服务器等硬件设备提出了更高的要求,如果没有基于大数据下技术和方法的应用,服务器负担太重,查询效率可能比较低。因此要在大数据环境下做好财政审计分析工作,从硬件、软件和网络等各方面都需要做好准备。

#### (五) 认识不足、能力有限,数据分析难

在近些年的财政资金审计中,审计人员多数采用抽样分析的思路,如抽审某地区财政资金中的某些科目进行审计,以判断财政资金的使用情况。按照维克托·迈尔·舍恩伯格观点,大数据时代“样本=总体”,数据分析应该是全数据模式的分析。当前审计模式还局限在抽样、局部等传统的财政审计方式中,主要原因是审计人员对大数据的认识不足,运用大数据开展财政审计数据分析的能力不强。抽样审计的财政数据分析,都受人员和时间限制,开展全地区全面财政资金审计更是物力维艰。因此,在大数据背景下财政审计数据分析中,审计人员的综合能力不足和大数据下财政数据分析要求高之间的矛盾,是当前财政审计数据分析中的难点之一。

### 四、做好大数据环境下财政审计数据分析的思考

#### (一) 变革思维,树立大数据环境下数据分析的意识

目前财政审计数据分析已经初见效果,但应用还不够广泛,与“五个关联”还有很大的差距,要做好财政数据分析工作,需要三个转变。一是转变工作思路。审计人员要深刻认识到大数据所带来的变革和经济新常态下开展审计工作的思维转变,学会熟练运用信息化技术评价判断、核查问题,提高审计工作能力、质量和效率。二是转变审计模式。要改变从会议纪要、举报信中寻找线索的审计手段,要改变将电子数据只是视为账册的另一种形式的观念,要深刻认识到一线作业与后台数据分析融合的重要性,单点离散审计向多点联动审计的转变,现场审计向现场审计与非现场审计相结合转变,适应新的审计模式。三是转变对大数据分析的认识。大数据的时代背景决定应当将数据分析工作常态化,彻底改变数据分析“慢热”的现状,使数据分析贯穿项目整个阶段,对制定审计实施方案、确定审计重点和延伸审计对象起到引领作用。

#### (二) 加强共享,加强财政审计的关联分析

随着信息化的飞速发展,数据的大集中、大关联、大计算将是未来信息化环境下审计工作的主要模式,今后的财政审计要更加注重多维数据关联分析,构建大数据审计工作模式,通过搭建综合数据分析平台,集中海量化、共享融合化多部门数据资源,开展关联分析,从多个维度分析相关环节的疑点线索,挖掘有价值的内容,利用大数据分析拓展审计的深度和广度,提高审计“集中分析,精准打击”的效率,也减少对查询银行账户、证券账户等手段的依赖。因此,打破思想壁垒,支持数据共享,把数据共享规范化制度化,真正形成想共享、能共享的局面。

#### (三) 统筹规划,集中解决财政审计中的难点问题

“五个关联”其核心是以财政数据为中心的多方数据关联的审计思路。当前多数省市的财政数据中,数据管理信息系统不统一,中央财政到省市县的资金使用均无法直观掌握。要做好关联分析,首先要做好财政数据标准化和贯通问题,实现财政数据的省市县三级时间、空间和业务维度的贯通。审计署应全盘考虑,统筹规划,把财政数据的贯通问题作为重点课题,集中研究,解决财政分析中的瓶颈。同时,应加强财政数据分析成果总结和提升,对常规问题可设计计算机审计分析软件,通过固化的方式形成可以直接应

用的分析模型,提高财政审计效率。

#### (四) 高瞻远瞩,做好审计数据分析规划

要从国家审计层面认真研究,建立科学的数据规划。从长远规划、近期规划和当前过渡期三个层面,认真研究,深入了解各行业、系统自身的数据环境,分类制定数据环境建设规划。把建设国家审计数据系统和大数据数字化审计平台作为远景规划,在建设过程中应通过明确近期规划和过渡期规划,出台明确务实的数据管理使用制度,指导当前的财政审计数据分析工作。

#### (五) 未雨绸缪,提高审计人员的财政数据分析能力

大数据背景下,要实现财政审计数据分析思路明晰、工具高效、结果精准等目标,对审计人员的能力提出了更高的要求。目前业务精、计算机水平高的人才稀缺,普通审计干部即使通过计算机中级培训,大多也只能使用简单的查询语句进行对比分析,无法适应大数据背景对数据分析人员的能力要求。因此,要培养审计人员掌握大数据处理、多维分析、数据挖掘等数据处理和分析的能力,学习在大数据环境中 No-SQL 技术、可视化分析等审计分析新方法的应用,满足大数据环境下财政数据分析能力要求。同时,要积极发挥数据分析团队作用,通过数据分析团队引领数据分析工作,加大技术创新,适应当前“两办”提出的加强审计全覆盖、促使财政审计数据分析向大数据迈进的要求。

#### 主要参考文献:

- 顾洪菲. 2015. 大数据环境下审计数据分析技术方法初探[J]. 中国管理信息化(3):45-46.  
李强,谢汶莉. 2016. 大数据审计中的可视分析[J]. 中国内部审计(2):79-86.  
马志娟,梁思源. 2015. 大数据背景下政府环境责任审计监督全覆盖的路径研究[J]. 审计研究(5):28-29.  
秦荣生. 2014. 大数据、云计算技术对审计的影响研究[J]. 审计研究(6):23-28.  
维克托·迈尔·舍恩伯格,肯尼斯·库克耶(盛杨燕,周涛译). 2013. 大数据时代生活工作与思维的大变革[M],杭州:浙江人民出版社.  
杨凯茜. 浅谈大数据审计的特点及实现——以审计署对 2012 年中石油的审计结果为例[J]. 财经界(12):315-316.  
张兆信. 2015. 大数据环境下的审计问题研究[J]. 内蒙古科技与经济(5):40-41.

## A Study on Public Finance Audit Data Analysis in Big Data Environment

Pei Wenhua Cheng Weiyi

**Abstract:** Internet and mobile communication has contributed greatly to the booming of big data. It is critical to use big data analytics, such as correlation analysis to help improve the efficiency and effectiveness of the audit work. However, few studies focus on the public finance audit data analysis using big data analytics. Based on audit experiences, this paper first proposes three methodologies of how to conduct correlation analyses in public finance audit data centering on financial funds, corporate and personnel respectively. Second, it discusses several difficulties and practical problems in the current data analysis, like data sharing, standard data, heterogeneous data and the ability of auditors. Finally it puts forward practical solutions to the key challenges in audit data analytics, i. e. to raise awareness of big data analysis, make data planning, and solve difficult problems in public finance audit. Meanwhile, we should improve the data analysis ability of auditors, and promote the efficiency of public finance audit through correlation analysis.

**Key words:** big data, public finance audit, correlation analysis

(责任编辑:王志伟)