

税收数据分析

方法与应用

S_HUISHOU
SHUJUFENXI
FANGFA YU YINGYONG
谭荣华 / 主编

 中国税务出版社

税收数据分析方法与应用

谭荣华 主编

中国税务出版社

图书在版编目 (CIP) 数据

税收数据分析方法与应用 / 谭荣华主编. --北京:
中国税务出版社, 2012. 5
ISBN 978-7-80235-804-1

I. ①税… II. ①谭… III. ①税收管理-数据-
分析 IV. ①F810.423

中国版本图书馆 CIP 数据核字 (2012) 第 094317 号

版权所有·侵权必究

书 名: 税收数据分析方法与应用

作 者: 谭荣华 主编

责任编辑: 黄 琳

责任校对: 于 玲

技术设计: 刘冬珂

出版发行: **中国税务出版社**

北京市西城区木樨地北里甲 11 号 (国宏大厦 B 座)

邮编: 100038

http: //www. taxation. cn

E-mail: swcb@taxation.cn

发行中心电话: (010) 63908889/90/91

邮购直销电话: (010) 63908837 传真: (010) 63908835

经 销: 各地新华书店

印 刷: 北京中科印刷有限公司

规 格: 880 × 1230 毫米 1/16

印 张: 31.5

字 数: 991000 字

版 次: 2012 年 7 月第 1 版 2012 年 7 月第 1 次印刷

书 号: ISBN 978-7-80235-804-1

定 价: 79.90 元

如有印装错误 本社负责调换

编辑委员会

主 任 李林军 谭荣华

副主任 李 伟 王 敏 谢波峰

主 编 谭荣华

副主编 谢波峰 王 敏

编 委 (以姓氏笔画为序)

王敬峰 冯 伟 司京民 朱庆锋 刘志安 刘若鸿

刘 超 刘尊涛 刘静华 安志伟 孙国光 孙晓明

孙 靖 李 硕 吴 强 张 茜 张 婉 陈远燕

陈 洧 周四新 庞瑞江 赵为民 赵芸淇 赵 涛

郝江波 袁有杰 夏日红 夏智灵 徐夫田 徐 乐

徐 康 栾春华 陶昭华 梁 季 彭启蕾 鲁钰锋

温 磊 楚文海 简春庆 谭 聪

序 言

经过近 30 余年的发展，我国税务系统的信息化建设已全面实现了从数据管理到信息管税的目标转变，开始进入以数据分析和深度利用为主要特征的新的发展阶段。这是新形势下税收核心竞争力的重要内涵。税收数据分析和深度利用的根本动因，是我国经济社会科学发展和构建和谐社会对建立科学的现代化税收管理体系的内在需要，是推进税制改革、加强税收征管、优化纳税服务、提高队伍素质的必经之途。

金税三期工程将要搭建全国大集中的税收信息平台，这将为税收数据分析和深度利用创造一个前所未有的崭新空间。在这个即将形成的平台上如何系统、有效地开展数据分析和深度利用工作，仍然是值得我们关注和思考的问题。例如从税收管理战略规划入手，建立宏观和微观层面的税收数据分析和深度利用的研究方法体系，就是税制改革和税收征管改革实践中所提出的重要问题。要解决这些问题，从长期看，需要高度重视分析主题的科学选择，高度重视分析数据的采集和质量评价，高度重视模型方法的合理运用，高度重视分析结果的多视角展现，高度重视应用系统的规划与实施，高度重视人才的培养。近期，则要重点考虑数据分析和深度利用的体制、机制和业务分析重点，建立业务数据模型，完善三期管理决策系统的路线图，从而初步形成统一规范的税收数据管理分析体系，并加快统分结合、层次分明的数据分析平台建设。

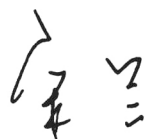
实现远期和近期目标，需要具备相关数据分析知识的人才来推动，因而加强整个税务干部队伍的数据分析能力建设，始终是一项非常重要的基础性工作。正是基于此，谭荣华教授以及他的研究团队编写了本书。

本书前三章从税收数据分析的概念、内涵、要素和基础理论入手，对税收数据分析的方法和必备要素——数据展开了全面阐述。后七章分别围绕税收收入预测、可计算一般均衡模型、税收

收入能力测算、税收分析指标体系、纳税评估、税收风险管理等分析主题做了深入探讨。

本书是谭教授的研究团队对我国税收管理和信息化工作长期跟踪研究的成果，是他们多年来对税务领域税收数据分析和深度利用研究的集中展现，也是各章作者多年来在税务系统实践税收数据分析和深度利用的切身体会，具有重要的理论和实践价值。

我希望此书的出版能够促进税务部门对税收数据分析方法和应用的研究，以利培养一大批掌握现代数据分析方法的高素质税务管理人才，进而推动我国的税收数据分析与利用工作，为我国税务管理水平的进一步提高贡献力量。



2012年7月

前言

税收数据作为纳税人纳税义务、征税人征税依据及征纳过程与成果的数量描述，一直是将税收法律制度、税收执法主体以及纳税人三者联系在一起的纽带，也是税收工作的核心要素。对于税收工作而言，税收数据的价值不仅仅在于为税收执法提供了数量依据，更重要的是，由于税收数据——特别是在长期的税收工作中积累起来的海量数据——在一定程度上描述和记录了征纳双方在相应时期内的涉税活动，是深入研究、分析涉税活动发展变化规律并将其应用于税收工作实践的宝贵资源。

在社会主义市场经济不断发展的今天，经济主体结构逐渐多元化，经济活动形式日新月异，治税环境日趋复杂，税务部门征管手段、稽查手段、监控手段、内部监督机制、人员素质和税法宣传等方面都面临新的挑战。纳税人与税务部门、各级税务部门之间、税务部门与其他部门之间客观存在的信息不对称的现实与严格的依法治税之间的矛盾更加尖锐。在新形势下，税收科学化、精细化、专业化管理和“信息管税”的迫切要求，使得税收数据深度分析利用工作的重要性日益凸显。在金税三期工程实现税收数据的全国大集中后，深化税收数据分析与利用，实现由数据管理向信息管理，进而向知识管理、智能管理的跃升，有助于税务部门更好地履行职能，为科学发展和改善民生提供可靠的财力保障。

从建立健全符合科学发展、促进社会和谐的税收制度的要求出发，决策者不仅要考虑税制调整后税收收入的变化，还要考虑对纳税人的影响，对税源的影响，对产业结构调整的影响和对区域经济发展的影响；不仅要国内实际情况出发，还要基于经济全球化的视角，考虑国内企业参与国际竞争的问题和国际间税收竞争的问题；不仅要基于公平、效率和税收中性的原则对现有税种进行调整，还要考虑新税种的设计；不仅要分析历史数据，还要分析税收制度调整可能对经济发展、经济结构产生的影响。同时，为纳税人提供优

质、高效的纳税服务，构建和谐征纳关系是税务部门的重要使命。这就需要对纳税人的涉税行为进行分析，有针对性地为各类纳税人优化或重新设计纳税服务流程，提供不同的纳税服务手段，从而根本上提高服务水平，降低纳税成本，促进纳税遵从。可以说，所有这些都使税收数据分析利用逐步深化成为必然。

未来，在金税三期工程实现税收全国大集中后，将初步形成统一管理涉税数据的新格局。通过合理分析数据需求，简并纳税申报资料，规范数据采集清单和采集规范，优化数据采集渠道，将实现国税数据实时集中、地税数据及时集中、第三方数据适时交换的全国数据大集中，建成数据全面、存储灵活、综合服务的税务行业数据管理体系，为各项税收业务提供全方位、多层次的数据支持。在此基础上，一些诸如征管状况分析、查询统计报表、风险管理、政策评估、绩效管理、知识管理、收入核算、纳税人关系管理等以税收数据分析为核心内容的业务将逐渐发展、成熟。税收数据分析利用工作将充分借助现代信息技术和数据分析的理论、方法，基于全国大集中后相对完备的税收数据，实现对税收业务管理、决策的支撑，逐步构建横贯宏观、微观领域，满足各级税务机构、税务工作人员需要的综合分析体系。

税收数据分析和深度利用要用到多学科的理论、方法和模型，用到数据仓库、数据挖掘、联机分析系统等计算机方面的知识，还需要熟悉税收法律法规、征管规程和财务会计的内容，知识储备刻不容缓。正是基于上述考虑，中国人民大学金融与财税电子化研究所在国家税务总局的大力支持帮助下组织编写了本书。

本书分为以下内容：

第一章“导论”，从数据、信息、知识和智能的比较入手，引出了税收数据分析的基本概念，介绍了信息化发展理论与中国税收信息化的阶段划分，勾勒出了国内外相关理论基础和实践发展的脉络，并进一步指出了税收数据分析在我国税收信息化建设中的地位和作用，剖析了我国税收数据分析的现状和问题。本章旨在让读者对税收数据分析有一个初步的了解。

第二章“税收数据分析的过程和方法”，较为详细地介绍了税收数据分析过程中确定分析对象、提取分析数据、选择分析方法并构建指标体系、撰写分析报告、落实并反馈整改意见等5个日常环节的基本工作内容和原则，并据此提出了税收数据分析系统构建的思路和方案；简要介绍了当前国内外较为前沿和流行的数理经济方法、统计方法、计量经济方法等可以用于税收数据分析的方法和模型；介绍了税收数据分析的几个主要应用领域。

第三章“税收数据与数据整备”，本章从数据分析对数据管理的要求角度出发，通过对现代数据管理理论、技术及其应用沿革的介绍，结合涉税信息流理论模型，对面向主题的税收数据分析过程中可能涉及的数据来源、标准、抽取、组织等环节作了较为深入的分析；在此基础上，较为系统地介绍了服务于决策分析的税收数据整备的一般过程及其关键要素。

第四章“税收收入预测”，介绍了税收收入预测方法的不同分类，提出了宏观、中观、微观税收收入预测框架体系，并分别针对宏观、中观、微观三个层面的税收收入预测框架，介绍了各自的数据要求、适用模型和底层分析工具。根据税收经济的实际数据，介绍了ARMA模型、协整理论及误差修正模型法和联立方程法建立税收经济预测模型等宏观预测方法，主要包括时间序列分析方法和联立方程组模型方法；从国民经济各部门的角度出发，介绍了可计算一般均衡模型（CGE）和投入产出两种中观分析和预测方法；以个人所

得税预测为例介绍了税收收入预测的微观方法。

第五章“税收可计算一般均衡(TCGE)建模及其在税制改革中的应用”,结合中国国情,从模型概念、理论基础、描述的经济系统循环、对税收与经济关系刻画、常用函数、与局部均衡模型比较、建模与税收政策模拟步骤等方面介绍了TCGE模型的基础理论、基本假设、宏观闭合规则、税制刻画以及经济描述;借鉴多个国家和区域TCGE模型建模经验,分析了中国经济运行特征和税制结构,构建了中国“标准”的TCGE模型;从税收社会核算矩阵(TSAM)理论入手,详细介绍中国TSAM构建的方法和步骤、主要数据来源和数据口径、TSAM数据采信原则、TSAM数据的计算过程;总结了国内关于增值税转型的研究成果,并据此设计模拟方案,调整模型方程,模拟并分析了增值税转型方案的结果,提出相应的政策建议。

第六章“税收收入能力测算”,首先简要界定了税收收入能力的口径,以及与相关概念的联系与区别,廓清研究范围。然后,对税收收入能力的衡量指标、估测原则、基本要求以及主要方法做了概括性的介绍。根据税收收入能力口径不同,丰富和完善了三种方法,即税柄法,代表性税制法以及分税种法。之后,对我国税收收入能力进行了实证分析。在系统介绍税柄法估测思路、模型的基础上,使用我国31个省、自治区、直辖市的1996~2005年的实际数据,对我国税收收入能力进行了估测。利用代表性税制法,对我国2000~2003年的省际间税收收入能力进行了比较;利用分税种法,对我国1999~2003年各主要税种的收入能力进行了估测,重点使用投入产出法估测了我国增值税的收入能力。

第七章“税收分析指标体系与纳税评估”,包括两个方面的内容:第一部分集中介绍税收数据分析所需的各类指标,包括宏观税收与经济关系类、微观税收与经济关系类等描述税收与经济关系的指标,以及税源监控、税收征收、税收管理、税收稽查、税收法制等税收业务类指标。第二部分重点阐述纳税评估相关内容,介绍了纳税评估的基本概念、指标分类、指标体系概念与构建基本原则、指标体系分类框架;提供了一般性的、流转税类、所得税类、资源税、财产税类、行为税类评估指标,分析了具体指标的含义及其应用;介绍分析了申报资料勾稽法、逻辑错误检索法、综合评分排序法、判别分析方法、Tobit模型估测方法、神经网络预测方法等纳税评估对象的选择方法、模型;阐述了如何运用数理原理、规则及其特定的、具体的计算来对评估对象中包含的量进行测定,从而对纳税人纳税申报的真实性进行判断分析的定量分析方法,以及综合运用税收政策、评估人员实践经验和逻辑推断,对纳税人纳税申报的真实性进行判断分析的定性分析方法等纳税评估分析方法、模型。

第八章“税收风险管理”,从风险的内涵和特征入手,介绍了税收风险管理的概念和相关理论,指出了税收风险管理与日常税务管理、纳税评估、税务稽查、纳税服务等其他征管措施的区别与联系;介绍了战略规划、风险识别、风险排序、风险应对、过程监控、绩效评估等税收风险管理流程中的主要环节和为满足风险管理体系建设需要的规划目标管理机制、风险识别管理机制、风险信息管理机制和风险应对管理机制四种主要运行机制;构建了适合我国的税收风险管理总体框架,介绍了税收风险管理的主题、方法以及国内外税收风险管理的实践经验。

第九章“税务系统绩效管理体系构建”,重点介绍了绩效管理的概念、税务系统构建绩效管理的必要性、税务系统构建绩效管理的目的和意义;在梳理国内外税务系统绩效管理的理论和实践基础上,介绍了国外税务系统绩效管理的成功经验及对我国的启示和借

鉴；介绍了税务系统绩效管理体系构建原则以及绩效管理的流程框架和绩效管理系统实现的相关内容。

第十章“税收数据分析模型支持系统的设计与实现”，在总结前九章的基础上，重点阐述系统的总体框架设计和关键技术。一方面，本部分确定了系统设计的指导思想和设计原则，按照数据整合和工具集成的思路，集数据库/数据仓库、模型库/方法库、知识库于一体，将数据与模型加以结合，为决策者提供分析问题、建立模型、模拟决策过程和方案的环境，有效解决税收决策中的半结构化和非结构化问题；明确了税收政策分析的数据标准规范、税收数据采集和分析数据库的建设、税收政策分析模型族的研制和各类模型工具的集成以及特定税制改革方案的分析模拟实验等具体目标和任务，并给出了系统的6层总体架构设计。另一方面，重点阐述了税收分析模型的概念和意义、模型分类、常用的模型软件工具以及这些工具的集成技术。此外，重点描述了税收分析模型运用的重要基础——数据质量问题，并进一步分析了数据质量问题产生的原因及对策。

附录一“数据分析在其他领域的应用实践”和附录二“基于 Excel 的税收数据分析实例”属于扩充知识，其中附录一分别介绍了“风险管理”、“金融创新”和“客户关系管理”等银行业数据分析的应用实践，“分析上市公司财务报表的真实性”和“预测股票价格客户行为分析”等证券业数据分析的应用实践，“客户行为分析”、“精细化营销”、“优惠和折扣管理”等零售业数据分析的应用实践，“客户流失分析”、“欺诈行为分析”、“网络故障分析”等电信业数据分析的应用实践，“产品设计质量控制”和“产品客户关系管理”等制造业数据分析的应用实践，展示了不同行业数据分析的角度，以期为读者在税务系统数据分析运用中扩展思路。附录二以 Excel 作为一个门槛较低、方便实用有效的税收数据分析模型工具，以实例方式介绍了税收数据时间序列预测与回归分析、微观模拟分析以及税收数据多维分析的具体方法和步骤。具体介绍了使用 Excel 开展包括税收数据时间序列预测、税收数据随机扰动的平滑处理和税收数据回归分析等税收数据时间序列预测与回归分析的实例；使用 Excel 开展税收数据微观模拟分析的实例；使用 Excel 开展税收数据多维分析中税收数据的获取、分类汇总、透视表和数据透视图多维分析、基于 D 函数和模拟运算表的纳税户比例分析以及税收数据时间序列的形成等具体过程步骤；以使普通的税务研究和实务工作者认识到“税收数据分析的门槛并不高”，任何人都可以使用 Excel 这一基本办公软件很好地开展税收数据分析工作，税收数据分析马上就可以开始！

需要指出的是，为了方便读者阅读应用篇中的内容，更好地指导读者开展税收数据分析实践，同时也为了保证各章内容的相对完整与独立，应用篇中的七章内容都自成体系，也正因为如此，这些章节中的介绍经济指标和统计方法的部分内容可能有所重叠。

作者仅希望此书能对读者了解、掌握税收数据深度分析利用的知识有所帮助。受编写人员能力所限，本书尚有诸多不尽如人意之处，也难免内容疏漏与错误，热忱盼望各方批评指正。

曾华

2012 年 7 月

目 录

第一章 导论	1
第一节 税收数据分析的基本概念	1
第二节 税收数据分析的理论支持	3
第三节 税收数据分析的现状归纳	11
第二章 税收数据分析的过程和方法	18
第一节 税收数据分析的过程	18
第二节 税收数据分析的方法	23
第三节 税收数据分析的主要领域	34
第三章 税收数据与数据整備	39
第一节 税收数据的客观存在方式	39
第二节 基于分析主题的数据提取和审核	42
第三节 元数据和数据模型	48
第四节 面向应用主题的数据组织	55
第四章 税收收入预测	63
第一节 税收收入预测方法分类	63
第二节 税收收入预测的宏观方法研究	66
第三节 税收收入预测的中观方法研究	96
第四节 税收收入预测的微观方法研究 ——以个人所得税预测为例	120
第五章 税收可计算一般均衡（TCGE）建模及其在税制改革中的应用	134
第一节 税收可计算一般均衡（TCGE）模型理论框架	134
第二节 中国税收可计算一般均衡（TCGE）模型的建立	142
第三节 中国税收可计算一般均衡（TCGE）模型数据结构及模型校准	152
第四节 中国税收可计算一般均衡（TCGE）模型在税制改革中的应用：增值税转型分析	173
第六章 税收收入能力测算	190
第一节 税收收入能力测算的一般考察	190
第二节 税柄法	195
第三节 代表性税制法	203
第四节 分税种法	232

第七章 税收分析指标体系与纳税评估	252
第一节 税收与经济关系指标	253
第二节 税收业务指标	266
第三节 纳税评估概述	286
第四节 纳税评估指标体系概述	290
第五节 纳税评估指标体系：一般指标与应用分析	292
第六节 纳税评估指标体系：流转税类评估指标与应用分析	297
第七节 纳税评估指标体系：所得税类评估指标与应用分析	303
第八节 纳税评估指标体系：资源税、财产税、行为税评估指标与应用分析	313
第九节 纳税评估对象的选择方法、模型	316
第十节 纳税评估分析方法、模型	326
第八章 税收风险管理	347
第一节 税收风险管理的概念与理论	347
第二节 税收风险管理的机制	353
第三节 税收风险管理体系	357
第四节 税收风险管理的实践	366
第九章 税务系统绩效管理体系构建	374
第一节 税务绩效管理的理论概述	374
第二节 税务绩效管理的系统构建	393
第三节 税务绩效管理的技术实现	405
第四节 税务绩效管理的系统功能	417
第十章 税收数据分析模型支持系统的设计与实现	434
第一节 税收数据分析模型支持系统的指导思想和设计原则	434
第二节 税收数据分析模型支持系统的设计目标	436
第三节 税收数据分析模型支持系统的体系结构	437
第四节 税收数据分析模型支持系统的关键技术——模型与工具	441
第五节 税收数据分析模型支持系统的重要基础——数据质量	447
附录	451
1. 数据分析在其他领域的应用实践	451
2. 基于 Excel 的税收数据分析实例	458
参考文献	486
后记	489

第一章

导 论

当前,围绕税收数据分析的理论研究和实践应用已经成为我国税收工作中一项令人瞩目的内容。特别是当税收信息化发展到一定阶段,科学化、规范化、专业化的管理与服务成为衡量税收征管水平的核心标准时,税收数据分析的重要作用便愈发彰显。一方面,在经济发展改革进程中,方方面面都对税收管理、税源监控、税收决策以及税收优化服务提出了更高的要求,而税收数据分析正是满足这些要求的必要手段之一;另一方面,以数据大集中为标志的税收信息化建设取得了阶段性成果,数据集中和系统整合为大规模的、有组织的、有体系的和深层次的税收数据分析提供了数据基础和系统平台。这表明我国税收数据分析的全面应用时代已经来临。

第一节 税收数据分析的基本概念

一、数据、信息、知识和智能

认识是人的头脑对客观世界的反映。在人们长期总结、改进和验证认识内容和方法的过程中,认识被逐步分成了四个范畴,即数据(Data)、信息(Information)、知识(Knowledge)和智能(Intelligence)。

数据由原始事实组成,代表着真实世界的客观事物,是未经组织的数字、文本、声音和图像等,例如:国土面积、人口数量、申报纳税的原始资料。一般来说,数据除了它自身之外没有多余的价值,但是经过加工整理后的数据(即信息)却具有了新的价值。数据是认识的起点,数据的获取和生成是为了增值利用,从这个意义上说,数据采集应保持数据的真实性、准确性和完整性,数据的生成亦必须遵循一定的格式,以便为后续检索、加工和处理奠定基础。

信息是以一定目的为前提,以有意义的组织形式加以排列和处理的数据。美国管理大师和知识经济理论的创始人彼得·F·德鲁克(Peter F. Drucker)将信息定义为“有目的性和关联性的数据”^①。与数据相比较,信息具有两方面的特性:首先,信息已经在各种数据之间建立了相互联系并进行了整合;其次,信息的生成具有一定的目的性。因此,无须进一步加工和处理,信息就是有意义的,可以直接回答一些相对简单的、贴近客观实际的问题,例如是什么(What)、是谁(Who)、什么时间(When)和什么地点(Where),而数据并不具备这一特性。在信息系统中,信息可以通过简单的索引功能或查询模块生成。

知识是建立在数据和信息之上的,但并不是数据和信息的简单叠加,而是在信息与数据、信息与信息之间建立了有意义的联系,可以用于指导实践和解决问题。迈克尔·J·马奎特(Michael J. Marquardt)将知识定义为“体现了信息的本质、原则和经验,能够积极地指导任务的执行和管理,能够用于决策和解决问题”^②的认知。知识可以区分为显性知识和隐性知识,显性知识是已经或可以文本

^① 彼得·F·德鲁克,等. 知识管理[M]. 杨开峰,译. 北京:中国人民大学出版社,1999.

^② 迈克尔·J·马奎特. 创建学习型组织五要素[M]. 邱昭良,译. 北京:机械工业出版社,2003.

化的知识，易于进行传播；隐性知识是存在于个人头脑中的经验或知识，需要进行大量的分析、总结和展现才能转化为显性知识。与信息相比，知识也可以直接回答一些问题，但是这些问题更加复杂、深入和抽象，例如怎么样（How）和为什么（Why）。在普通的信息系统中，知识通常是无法直接获取的，只有当智能或者更高层次的知识被转化为流程或算法并编写成系统中的高级分析模块时，知识才能够生成。而且，只有当知识被呈交给能深入理解这一过程的人后，它才有可能真正用于支持决策和解决问题。

智能的概念和创新息息相关，是人类所表现出来的一种独有的能力，是人类对为达到某种目标而应该使用哪些知识以及应该如何使用这些知识的理解，是在把知识用于改造世界并且能够产生新知识的动态过程中所表现的能力。智能主要表现为收集、加工、应用和传播信息与知识的能力，以及对事物发展的前瞻性的看法。由于智能的概念在某种程度上等同于人类独有的创新能力，因此，信息系统本身无法实现真正的智能，但二者并非毫无关系：一方面，作为信息技术领域前沿之一的人工智能（Artificial Intelligence，AI）所研究的核心内容正是如何让“机器”学会像人类一样思考；另一方面，智能可以用来指导构建复杂的信息系统，从而让信息系统本身可以直接生成高级的信息和知识，例如，商务智能（Business Intelligence，BI）是基于信息技术构建的智能化管理工具，它可以实时地对 ERP、CRM、SCM 等管理工具生成的企业数据进行分析，并将企业的各种数据及时地转换为企业管理者感兴趣的信息或知识。

概括而言，数据是基础和原料，信息是有组织的、有特定目的的数据的集合，知识是有价值和有效用的信息，而智能则是建立在前三者基础之上的更高层次的知识运用和创新。表 1-1 列举了一些数据、信息、知识和智能在税务信息系统中可能的表现。

表 1-1 税务信息系统中的数据、信息、知识和智能

名称	可能的表现
数据	文本型字段：纳税人全称、法定代表人姓名。数据型字段：应税所得额、应纳所得税。逻辑型字段：是否一般纳税人、是否已经申报纳税。日期型字段：纳税企业成立日期、纳税申报日期。
信息	某一纳税人特定期内某一税种的缴纳总金额。
知识	具备某一特征的一类纳税人税负总水平的高低、在某一政策实施后税负总水平的变化。
智能	无法通过税务信息系统直接生成，但是税务智能（Taxation Intelligence，TI）可以指导税务信息系统的构建，从而可以更加有效、直接、科学地生成有用的信息和知识。

二、税收数据分析

在认识和理解数据、信息、知识与智能等概念的基础上，我们可以将数据分析定义为：基于对某个领域的认知，将原始的客观事实（即数据）经过筛选、分类、组织、加工和研究，进而上升到直接可用的信息和知识的过程（见图 1-1）。

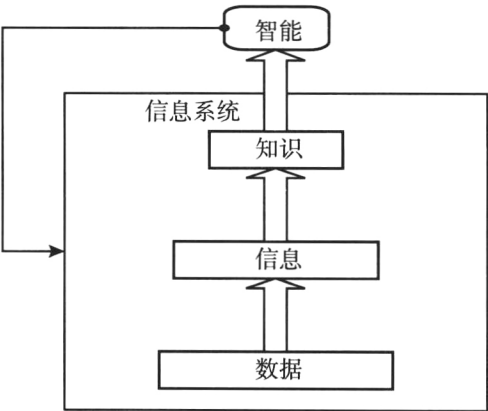


图 1-1 基于信息系统的税务数据分析过程

数据分析包括由数据到信息和由信息到知识两个环节。通常情况下,由数据到信息是相对简单的数据分析,基于一些常规的理论 and 简单的方法,即使在脱离信息系统的手工操作阶段,借助人和简单计算工具也可以完成这一层次的数据分析工作,信息系统的出现只不过实现了这一操作的流程化和自动化,并大大提高了其效率和准确度;而由信息到知识则是相对复杂的数据分析,基于一些高深的理论和复杂的方法,需大量使用统计学分析方法和计量经济模型,涉及大量的复杂计算,因此在完全脱离信息系统的情况下几乎是无法完成的。

智能的生成取决于人的创新能力,并不能在信息系统中直接生成,所以传统意义的信息系统框架只能覆盖由数据到信息、由信息到知识的分析过程,而对于知识到智能的上升环节,只能提供部分的支撑和协助,因此人的创新能力在这一环节必须发挥关键性的作用。

但是,智能并非与信息系统无关,智能的发展和优化决定了信息系统框架下数据分析的边界。也就是说,当信息和知识积累到一定程度时,通过人的创新能力实现由量变到质变的转换,从而使智能得以出现;反过来,智能可以用来指导由数据到信息、由信息到知识的数据分析过程,并且通过一定方式体现在信息系统的功能扩展、性能提升和结果优化等方面。

结合税收领域,我们可以将税收数据分析定义为:在税务系统及全社会信息化建设成果的基础上,利用经济、税收及相关学科的原理、方法和模型,针对税务部门数据集中、系统整合及外部数据交换收集整理的大量准确涉税数据(包括税务系统内部数据,也包括其他政府部门、企业、居民等外部数据),进行深入研究,探索其中蕴涵的规律和知识,帮助税务部门做出科学决策的分析活动。^①

第二节 税收数据分析的理论支持

一、信息化发展理论与中国税收信息化的阶段划分

信息化发展理论和中国税收信息化的实践表明,税收数据分析是税收信息化发展到一定阶段的必然结果。

1. 信息化发展的理论分析

20 世纪 70~90 年代,反映信息系统发展进程的阶段理论构成了国际上社会技术系统学派最杰出的代表性研究成果,这些理论主要包括诺兰模型、西诺特模型和米歇模型。^②

(1) 诺兰模型。美国著名的信息系统专家、哈佛大学教授理查德·诺兰(Richard Nolan)于 20 世纪 70 年代末提出了著名的企业信息化建设的阶段划分理论,通称“诺兰模型”,他将企业信息化建设分为六个阶段:

初始阶段。在这一阶段计算机刚进入企业,仅作为办公设备使用,应用范围也非常狭窄,通常只用来完成一些报表统计工作,甚至大多数时候被当作打字机使用;企业对计算机基本不了解,更不清楚信息技术可以为企业带来哪些好处,解决哪些问题;IT 的需求只被作为简单的改善办公设施的需求来对待,采购量少,只有少数人使用,在企业内部没有得到普及。

扩展阶段。在这一阶段企业对计算机有了一定了解,想利用计算机解决工作中的问题,例如,想通过进行更多的数据处理,给管理工作和业务带来便利。于是,应用需求开始增加,企业对信息技术应用开始产生兴趣,并对开发软件热情高涨,投入也开始大幅度增加。但是,在这个阶段也很容易出现盲目采购、盲目定制开发软件的现象,缺少计划和规划,因而应用水平不高,IT 的整体效用无法凸显。

控制阶段。在前一阶段盲目采购、盲目定制开发软件之后,企业管理者开始意识到计算机的使用已超出控制,虽然 IT 投资增长快,但取得的效益并不理想,于是开始从整体上控制计算机信息系统的发

^① 谭荣华,彭启蕾.论税收数据分析的五项核心要素[J].税务研究,2007(5).

^② ERP 信息网。

展,在客观上要求组织协调,解决数据共享问题。此时,企业 IT 建设更加务实,对 IT 的利用有了更明确的认识和目标。在这一阶段,一些职能部门内部实现了财务系统、人事系统、库存系统等应用,但各软件系统之间还存在“部门壁垒”和“信息孤岛”。信息系统呈现单点、分散的特点,系统和资源利用率不高。

统一阶段。在控制的基础上,企业开始重新进行规划设计,建立了基础数据库,并建成了统一的信息管理系统。企业的 IT 建设开始由分散和单点发展到成体系。此时,企业 IT 主管开始把企业内部不同的 IT 机构和部门统一到一个系统中进行管理,使人、财、物等资源信息能够在企业集成共享,从而更有效地利用现有的 IT 系统和资源。但是,这样的集成所花费的成本很高,耗费的时间很长,而且系统亦不够稳定。

数据管理阶段。在这一阶段,企业高层意识到信息战略的重要性,信息成为企业的重要资源,企业的信息化建设也真正进入到数据处理阶段。企业开始选定统一的数据库平台、数据管理体系和信息管理平台,统一数据的管理和使用,各部门、各系统基本实现了资源整合、信息共享。IT 系统的规划及资源利用更加高效。

成熟阶段。这一阶段企业信息化进入了最高层次,信息系统已经可以满足企业各个层次的需求,从简单的事务处理到支持高效管理的决策。企业真正把 IT 同管理过程结合起来,将组织内部、外部的资源充分整合和利用,从而提升了企业的竞争力和发展潜力。

(2) 西诺特模型。1988 年西诺特(W·R·Synnott)参照诺兰模型提出了一个新的模型——西诺特模型。这是一个过渡性的理论,主要考虑到了信息随时代变迁的变量。他用四个阶段的推移来描述计算机所处理的信息。从计算机处理原始数据的“数据”阶段开始,逐步过渡到用计算机加工数据并将它们存储到数据库的“信息”阶段;接着,经过诺兰所说的“技术性断点”,到达把信息当作经营资源的“信息资源”阶段;最后到达将信息作为带来组织竞争优势的武器,即“信息武器”阶段。

西诺特还提倡,随着计算机处理的信息机器作用的变化,作为信息资源管理者的高级信息主管或称为首席信息官(CIO)的重要性应当受到重视。当前,发达国家都接受了西诺特对诺兰模型的改善,将信息资源管理作为企业的头等大事来抓。

(3) 米歇模型。诺兰模型和西诺特模型均把系统整合(集成)和数据管理分割为前后两个阶段,似乎可以先实现信息系统的整合,然后实施数据管理,但后来的大量事实表明这是行不通的。美国的信息化专家米歇(Mische)于 20 世纪 90 年代初对此作了进一步修正,揭示了信息系统整合与数据管理密不可分,系统整合期的重要特征就是搞好数据组织,或者说信息系统整合的实质就是数据整合或集成。此前的研究仅仅集中在数据处理组织机构的管理和行为的侧面,而没有更多地研究各种信息技术的整合集成,忽视了将信息技术作为企业的发展要素而与经营管理相融合的策略。米歇的信息系统发展阶段论研究成果可以概括为:具有“四阶段、五特征”的企业综合信息技术应用连续发展的米歇模型。

米歇将综合信息技术应用的连续发展划分为四个阶段:起步阶段(20 世纪 60~70 年代)、增长阶段(20 世纪 80 年代)、成熟阶段(20 世纪 80~90 年代)和更新阶段(20 世纪 90 年代中期至 21 世纪初期)。其划分的特征不仅是数据处理工作的增长和管理标准化的建设,而且涉及知识、理念、信息技术的综合水平及其在企业经营管理中的作用和地位,以及信息技术服务机构在提供成本效益和及时性方面都令人满意的解决方案的能力。

决定这些阶段的特征有五个方面:技术状况,代表性应用和集成程度,数据库和存取能力,信息技术融入企业文化,全员素质、态度和信息技术视野。其实,每个阶段的具体属性还有很多,总括起来有 100 多个不同的属性。这些特征和属性可用来帮助一个企业确定自己在综合信息技术应用的连续发展中所处的位置。

米歇模型可以帮助企业和开发机构把握自身当前的发展水平,了解自己的 IT 综合应用在现代信息系统的发展阶段中所处的位置,是研究一个企业的信息体系结构和制定变革途径的认识基础,从而找准这个企业建设现代信息网络及其应用的发展目标。

2. 我国税收信息化的主要阶段

如果从时间上划分,税收电子化和税收信息化建设经历了四个主要阶段。^①

(1) 应用起步期(1983~1989年)。在应用起步期税收电子化的基本思路是:以个人计算机(PC)作为应用平台,以计划、会计、统计等税务内部的信息处理作为应用切入点,以基层为重点,提出“抓基层、抓应用、抓效果”,使广大税务人员从繁杂的手工劳动中解脱出来,并着力培养计算机应用意识。由于当时的税收征管模式是“一员到户、各税统管、征管查一人负责、上门收(催)税”,与此相适应,计算机应用便主要集中在会计、统计报表处理、税收月报处理等方面,涉及征管等其他方面的极少。实际上,这时期会计统的计算机应用还可分三步:第一步是从报表到报表,原始报表由人工计算出来,再输入到报表软件中,然后层层汇总;第二步是从税票到报表,先将每一张税票数据录入计算机,然后分类产生原始报表;第三步是从税票到会计记账,再到报表。因此计算机主要用来输入、产生向上报送的报表。全国各单位应用水平及程度普遍不高,设备配备也十分有限,到1989年年底,整个税务系统PC机实际装备数量仅有5300多台。

在应用起步期,数据的基本形态主要有两种,即手工处理的会计账簿和会计应用系统中的基础数据,其中电子化的数据主要集中在省、市级税收管理部门的报表汇总和报送机构。在这个阶段,数据基本上是零散的和缺乏联系的,所发挥的作用也只是记录和形成一些报表。

(2) 初步应用期(1990~1993年)。1990年税务总局决定全面进行征管改革,在全国范围内推广“征收、管理、检查三分离”或“征管、检查两分离”和纳税人主动申报纳税的征管新模式。征管模式的变化客观上为税收信息化工作的开展创造了有利条件。1990年4月,在广东省召开的全国税务系统第一次计算机应用工作会议上,税收征管电子化作为整个税收电子化的重要组成部分被提上议事日程。随后,总局提出要将征管改革、会计改革和开发应用电子计算机结合起来(即通常所讲的“两同步、三结合”),这标志着中国税收电子化信息化工作进入了一个新的时期。在这个时期,计算机的应用从税务系统内部引申到外部,能够建立简单的纳税人档案和纳税申报资料。为了正确引导全国各地广泛使用计算机,总局制定了《税收业务分类代码》、《税收征管软件业务规范》、《税务系统计算机应用软件评测标准》等一系列规范。

通过“两同步、三结合”试点工作,税收电子化建设有了进一步发展。首先,业务人员对计算机的认识有了较程度的提高,计算机在他们的心目中,已经从可有可无逐步变为能够很好地辅助进行税收征管的工具,有些应用得较好的基层单位,甚至感到离不开计算机;其次,计算机应用已从省市一级逐步转移到了征管第一线,对基层的税收征管工作发挥了重要作用;最后,应用的深度有所提高,如征管软件中征管数据和会统账已经建立了有机联系和监控关系。

在初步应用期,随着“微机通用税收征管信息处理系统”和“税收管理网络系统”等税收征管软件的推广,以及通过开发“出口退税计算机管理系统”实现了对出口退税全过程的电子化管理,电子化征管手段开始逐步取代手工操作,原来账簿形态的数据也基本上转化为电子化形态,这意味着电子化数据真正开始在税收基层征管部门逐步积累。在这一阶段,纳税人档案和纳税人申报资料等数据开始充实税收征管系统的基础数据,而且在各类数据之间已经建立了一些联系,但是从数据所发挥的作用上看,初步应用期和应用起步期相比并没有发生根本的改变。

总体看来,我们可以把应用起步期和初步应用期归纳为税收管理电子化阶段。以税收管理信息系统的普遍应用为标志,税收机关在这个阶段为实现日常工作与税收征管操作、管理环节的电子化,在征管业务、出口退税、办公自动化等各类业务层面建立了一些计算机应用管理系统,使税收业务的管理实现了手工向计算机的过渡。这个阶段为税收管理效率的提高奠定了良好的基础。

(3) 应用发展期(1994~2000年)。1994年,我国进行了较大规模的财税体制改革,建立了以增值税为主体的流转税体系,实行了分税制,税收征管业务也随之发生了较大变化。征管的信息已从

^① 蔡金荣. 推进中国税收电子化信息化建设[J]. 信息系统工程, 2003(12).