

税务数据集中后的分析体系研究^①

Centralized Taxation Data Analysis System

叶志坚 (福建省国家税务局信息中心 福建 厦门 350003)

摘要: 当前全国基本实现了主要税收业务信息系统的省级集中,为税收分析提供了前提。本文从税收分析目标出发,分析税务数据源状况,归纳了税收分析方法属性和过程,指出当前税收分析的拓展面,提出税收分析的体系架构,最后,我们给出了几个应用举例。

关键词: 税收 分析 体系

1 引言

谢旭人先生指出:“加强税收分析,既要关注税收增长情况更要分析税收与相关经济指标关系,分析宏观税负、税收弹性等情况,并与企业纳税评估结合起来”。金税工程三期四个子系统中的“辅助决策分析系统”目标与之相适应,国家税务总局印发的《税收分析工作制度》规范了税收分析的工作层面;在税务决策支持系统研究方面,不少学者进行过探索。要加强税源管理首先需要掌握和了解,这就对税收分析工作提出了更高要求——税收分析不再是简单的收入计划执行情况分析,而应该着眼于税源进行税收与经济的相关分析,要为促进税收征管质量和效率的提高服务。当前全国开展了以税收分析、纳税评估、税务稽查为主要环节的“联动机制”工作,在各地陆续取得显著成效。国家税务总局印发了《税收分析工作制度》,规范了税收分析的工作流程等内容。不少学者进行过许多探索。宏观分析已经有过多年应用,但尚须拓展。微观税收分析方面,有学者已经总结出一套方法^[1],并为国家税务总局所推荐。实践方面,全国出现不少先例,近两年,山东省、云南省、河南省、福建省等国税开展大量的税收分析,取得初步成果。在发达国家,税收分析成为税务管理不可或缺的手段,例如,澳大利亚联邦税务局经过多年探索,已经形成一套税收分析的机制和方法。许多税务工作者,例如吴新联先生、焦瑞进先生,从各种角度出发,一直在研究具体的分析方法^[2]。有了对税收分析的初步认识,

有了许多税收分析的方法,有了税收分析实践经验积累,有了数据集中的前提,加上信息技术发展的支撑,可以探索出我国税收分析的体系架构。在信息技术基础上,到2006年底全国国税系统基本实现了主要税收业务数据的省级集中,另一方面,各地信息化建设已普遍取得突破性进展,这就为新形势下税收分析工作内容的拓展、税收分析水平的提高创造了必要前提。

必须指出,税收分析工作必须瞄准税务组织的目标来展开,根据大多数理论界的观点,税务组织就围绕三个目标开展工作^[3]:提高税收遵从度、降低税收成本、提高社会满意度。离开这个目标,税收分析工作可能陷入盲目状态。因此,在税收分析中,我们始终贯穿这个原则。

根据税务组织目标,税收分析工作可分为如下三类:一是税收预测,二是遵从度测算,三是税收政策评估。传统上,各级税务机关领导和税收分析主管部门主要从事第一类工作,征管部门更加关注第二类,税务部门 and 各级政府乃至公众关心第三类。综合以上,有了各面对税收分析的统一认识,有了税收分析方法的深入研究和更新发展,有了税收分析实践经验的不断积累,有了基础数据的大集中,加上信息技术发展的有力支撑,我们可以探索出构建新形势下税收分析体系的新思路。

2 数据集中后数据源的状况

数据集中后,可供进行税收分析的数据源有征管

^① 收稿时间:2008-11-10

数据、企业数据、社会统计数据、特别采集数据、其他数据。

2.1 综合征管软件数据

常称为 CTAIS, 可分为管理类、申报征收类、稽查类三大类数据。国税系统目前已经实现省级集中, 至少都有 2 年以上的数据, 多的可达 5 年以上。

2.2 增值税管理系统数据

增值税专用发票管理系统, 又称作金税二期, 记作 GT2。以增值税专用发票等为主的数据。2007 年起, 一般纳税人开具的普通发票也进入 GT2。经过数据整合, 2004 年起, 一般纳税人基本信息和发票领用存与 CTAIS 相一致。

2.3 出口退税数据

出口退税系统相对独立, 传统上主要根据“两单两票”来设计的, 经过数据整合, 企业基本信息以 CTAIS 为准同步, 出口税票也从 GT2 中批量导入。

2.4 企业财务报表数据

税务局接收企业所得税申报时附表信息。这些附表反映企业财务情况。一般情况下, 这些表格与财务报表有一定的差异, 人们称作“财税差异”。

2.5 特别采集数据

如国家税务总局统一部署常见的有重点税源调查数据等, 加上各个省级的重点税源, 这部分数据越来越丰富。

2.6 社会统计数据

由国家统计局发布的数据, 各个行业组织发布的数据, 当前这些数据需要人工采集。

2.7 其他数据

其他数据包括稽查数据、个体定额调查产生的数据、行政系统相关的数据、历史相关分析数据、国际税收情报数据和国外其他参考数据等。

从完整性看, 与业务系统相联系的作业记录数据比较完整, 其他数据采集周期不稳定, 对象也不稳定, 因此不完整。税务部门拥有的数据及完整的数据应该进行日常分析、日常监控, 而来自外部门或不稳定的数据可以进行专题分析; 交叉部分可以进行相关分析。这些数据是信息化历史的积累, 目前已经发现了它的价值, 必须进行数据战略规划的研究, 国家税务总局已经把它作为税务信息化发展战略的一个组成部分, 也是把涉税数据转变为战略资源的一个重要方面。数据战略规划研究涉及的问题很多, 包括数据资源发展

的战略环境分析, 数据资源应用的战略目标校准, 数据资源应用发展的战略推进, 也就是实现战略目标的方法和途径, 也包括数据中心构架。

3 税收分析方法属性和过程

3.1 税收分析指标及其属性

概述中我们已经说明了按照税收分析目标进行的分类, 税收分析方法还有很有多种分类, 按照分析范围划分, 有宏观分析、微观分析; 按照分析工具划分, 可分为比较分析、数理分析; 按照工作性质划分, 可分为日常分析、专题分析; 按照技术划分, 有主题数据库技术、数据挖掘技术、数据仓库技术; 按照数据集中位置划分, 可分为总局、省级、市级、县(区)级等等。按照使用者用途划分, 可分为税源分析、收入预测等; 按照数据源划分, 有征管数据、企业数据、社会统计数据、特别采集数据、其他数据。这些划分都是从不同的角度出发, 各有用途及合理性, 但是不能反映某个方法全貌, 因此, 在下面的典型指标列举中, 如表 1 所示, 我们将这些划分方法都当作税收分析指标模型的属性看待。

表 1 税收分析指标及其属性

典型指标举例	相关税种	分析范围	工具	性质	技术	集中位置	用途	数据源
税收收入趋势	所有税种	宏观	时间序列法	月/季/年	线性/非线性模拟	全国/省/市/县/区	收入预测	CTAIS
增值税一般纳税人申报统计表(分行业、地区)	增值税	宏观	统计	日常	数据库	全国/省/市/县/区	税源分析	CTAIS/GT2
主营业务收入变动率	所得税	微观	比较/统计	日常	数据库/数据挖掘	县区	纳税评估	CTAIS/企业数据
产销成本、产销利润、产销率情况	消费税	宏观/微观	比较/统计	日常	数据库/数据挖掘	省/市	税源分析	CTAIS/企业数据
税收弹性系数跟踪表	所有税种	宏观	计量经济学	日常/专题	数据库/数据挖掘/数据仓库	全国/省/市	收入预测/税源分析	CTAIS/社会统计数据

表 1 仅仅是为了说明问题而列举的一部分，实际运作时，指标应尽量全面，属性应更加明确，并且随着经济形势的变化不断补充。实际分析时，除了日常分析外，并非所有指标都进行。具体某个指标应用时，还得注意一些特殊数据的调整，吴新联先生曾提出税收分析时出现的一些问题^[4]，提醒口径和时间差方面的数据调整。

3.2 分析流程

分析流程如下：确定分析目标——选择方法与技术——选择分析模型——分析指标计算——调整并发布分析结果——验证分析结果(如图 1 所示)。

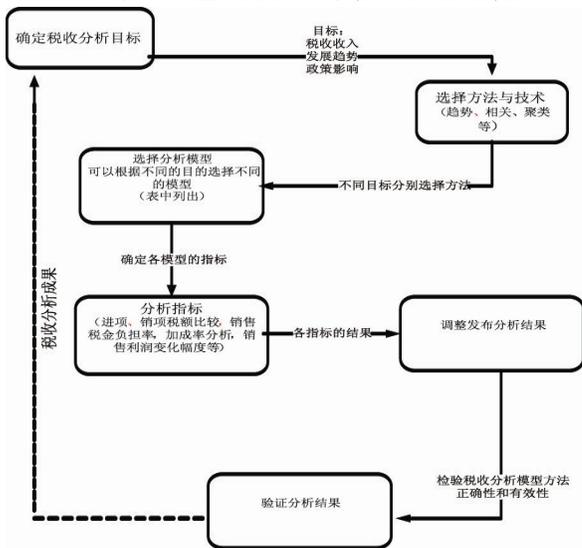


图 1 税收分析流程图

4 税收分析思路的拓展

上世纪 80 年代前，由于税收计划体制影响，我国税收分析主要侧重收入分析，一般对不超过 10 张的统计分析、不可比因素分析、上年同期数字比较，很少涉及比较复杂的分析方法。新税制改革后，特别是增值税管理带来的巨大挑战，促使税收分析有了很大发展，引入了许许多多分析方法。但是，不可否认的是计算机的广泛应用，数据的大量拥有，才使得这些新的分析方法得以实现。正像上面所述，税收分析方法已经有了一套比较多的指标体系，但是从理论和实践中，我们需要进行拓展。首先税收思路要更新，其次对新的税收分析领域要进行拓展并进行实践，还要探索补充新的分析方法。

4.1 理念的更新

一个传统的收入分析报告结构如下：

表 2 传统的收入分析报告结构

- 1、收入数、同比增长情况、完成计划进度。
- 2、宏观税负、弹性(与 GDP)。
- 3、分下级单位、分税种、分行业收入统计表。
- 4、本级重点税源情况补充说明(正、负)。
- 5、征管措施及其对收入的效果呈现(正)。
- 6、本期税收政策对收入的影响(正、负，正小于负)。
- 7、与全国(或全省)比较，上述 1-6 项指标的出超或差距。
- 8、预测下期或到本年底收入。方法为一般趋势(如年增长 12%)和上述 1-7 项因素影响结果

以上可以看出这么几点：一是这份分析报告比以前已经有了提高；二是这份报告还带有计划经济时代的成分；三是众多的税收分析方法没有充分利用；四是分析结果经常没有再验证。税收分析不再是收入计划分配的艺术，是围绕组织目标的工作行动，是联动机制的实践活动，这种理念应该转变过来。

4.2 新领域的实践

在许多国家，经常使用的一套方法有：时间序列法、计量经济学等式法、微观模拟法，还经常用宏观微观交叉印证法，分析报告中有篇幅的结果说明，结果定性判断及其修正说明。这些我们可以从一些发达国家税务部门公开的年度报告中查到。许多好的经验我们可以借鉴。但是我们也可以看到其中的许多局限性和困惑。也就是说，即使是国外，税收分析也有拓展空间。

我们要大力拓展税源领域分析，这方面，理论界和不少税务机关已在投入产出分析、纳税能力估算等方面进行大量研究，但受各方面条件尤其是信息技术约束，其成效尚不显著，也很难大范围铺开，亟需信息资源采集及其真实性验证问题；在微观方面，同业税负分析思路和方法在互动机制中的税收分析环节大量应用，即同一行业的企业工艺、进料、能耗、销售费用、财务费用，都有很强的同类性，利润空间、增值率应该差不多，税收政策基本相同，税收和税负具有很强的可比性，当税负异常时，可作为一个重要的预警指标。应该说，上述两种微观分析方法是目前开展数据应用分析非常重要也非常有前景的不可或缺的重要方法。

4.3 分析方法的拓展

数理统计方法非常丰富，许多方法可以用到税收

分析中来。例如聚类分析，“2002 年地区税收与 GDP 增长关系聚类分析”一文，就 2002 年税收增长、经济增长、税收弹性情况进行聚类分析，得出税收增长与经济发展和都市化程度联系密切的结论^[5]，因此居民收入水平与税收收入预测可以相联系。又如相关分析，“关于重点税源监控企业税收与税源相关分析的分析”一文^[6]。这些分析方法从理论和实践上都很有意义，但是这种分析太少得到应用。然而，是不是没有别的方法了？显然不是，数理分布统计技术中许多方法可以应用。税收来自经济税源，纳税人经济运行受内部和周围环境影响，换句话说，如果周围环境趋于一致，该纳税人的经济运行也趋于一致，但是内外环境毕竟不同，这就造成税收差异，正象图形学中鉴定一幅数码图形是自然拍摄还是有经过编辑的方法类似，这种一致性与差异性符合一定的分布规律^[7]，凡是不符合这种分布的纳税人，可列入异常，进而纳税评估对象。这需要我们进一步探索。即使是发达国家，税收分析也在不断探讨，例如，税收预测更加许欢用简单线性趋势拟合方法，这属于无条件分析，而由于复杂数学技巧的原因，复杂自回归方法却极少使用，这也是方法拓展点。

5 税收分析架构建立

我们已经认识到了税收分析的形势，分析了数据集中后数据源的状况，确立了税收分析方法属性和过程，税收分析思路也进行了拓展，现在可以建立我们的架构了。根据业务驱动原则，这个架构由四部分组成：先建立需求架构，对信息资源进行规划，然后搭建技术基础平台，最后建立税收分析队伍。

5.1 需求架构

根据国家税务总局金税工程三期设计，共设计四个应用系统，其中第四个系统是“辅助决策分析”，税收分析属于“辅助决策分析”，五个模块中的“分析”、“预警”和“预测”三个模块。

5.2 信息资源规划

前一节我们描述了数据源状况，认为已经得到可利用的信息，但是这一步对税收分析来说还远远不够，仅仅使用现有的数据资源进行税收分析还是会遇到很大困难。首先是数据源属主部门化，数据取得的困难；其次是数据源的真实性难以信任；再次是数据源异构性带来实现性麻烦；最后就是现有数据面向事务处理

设计的，直接分析性能很低。需要进行处理和转换。信息资源规划是解决这些问题的，根据金税工程三期的规划，其过程如图 2 所示：

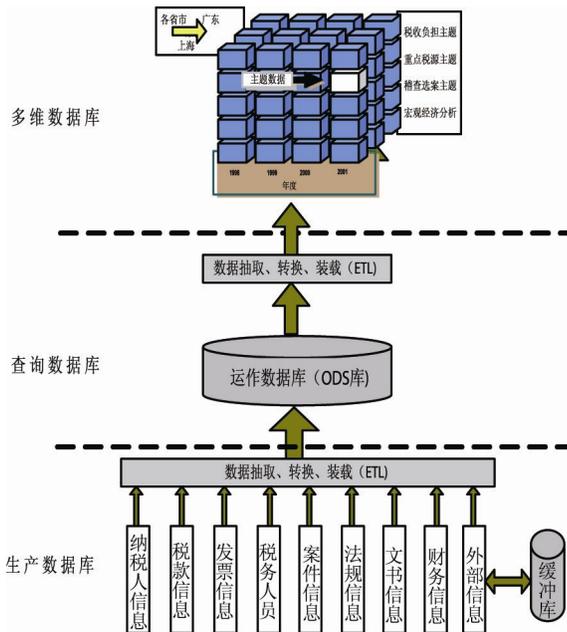


图 2 信息资源规划图

5.3 基础平台建设

规划了信息资源，就付诸实现，即基础平台，包括网络、主机、软件，对应需求和数据源规划，其架构如图 3：

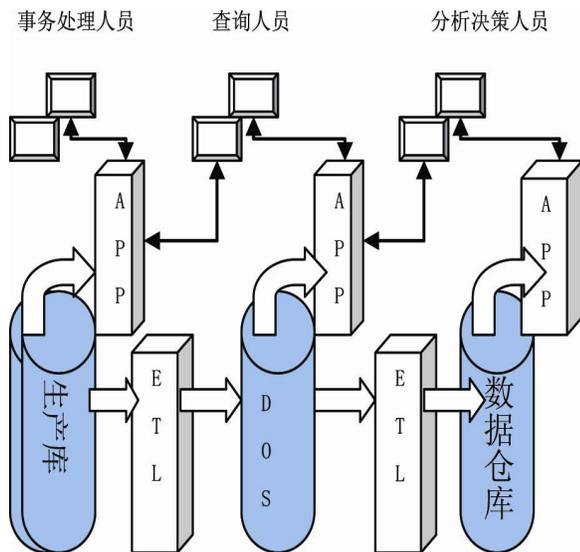


图 3 基础平台架构图

上面每个设备采购需要一定投入，设备上运行的系统软件需要费用，ETL 规则编写和 APP 开发需要人

力资源,更重要的是查询人员、决策分析人员真正用起来,对税务组织目标的更好实现才具有价值。

5.4 分析体系的组织机构和人员

最后我们要谈到人员及其组织。从长远来看,设立专门的机构和人员长期从事税收分析工作是提高税收分析质量和绩效的可行方法。我们可以借鉴其他国家的成功经验,如:澳大利亚的税务局设有与信息中心平行的专门的收入分析部从事税收分析预测工作,在2005-2006财政年度,联邦分析部门配备员工90人,为税收分析预测工作的顺利开展提供了强有力的组织保障,当期联邦税收收入2080.85亿澳元[8]。

6 应用

根据上述体系研究,并且进行项目开发,我们实际应用中内容比较丰富,这里不一一列举。我们仅看几个实际应用中的例子?注?:

确认数据源:我们选择CTAIS2.0中的dj_nsrxx、dm_swjg、dm_hy、dm_hydl、dm_djzclx和sb_zzs_2003_ybnsr表,这些表的选择是针对行业大类和注册登记类型分析的,为了分析经过清洗和装载,得到税负表lzs_sf_mx,再分成三类,得到lzs_sf_hy、lzs_sf_hydl和lzs_sf_djzclx。清洗需要许多判断过滤脚本,例如过滤或纠正纳税人识别号异常、申报税款所属期异常的等等(脚本略),这些脚本通过存储过程实现,存储过程还清除了与分析无关的历史数据、确定抽取时间周期、和作其他一些特殊情况的处理。然后对事实表LZS_SF_MX中的数据按机关代码、行业代码、登记注册类型、时间进行数据转换后插入到相关表中。

数据清洗与装载:以XX省国家税务局2007年3月份数据为例,该省一般纳税人户数是24000多户,sb_zzs_2003_ybnsr表中98000多,经过上述数据的确认、清洗、装载后,lzs_sf_mx中记录数为24000多,再经过数据汇总转换后,lzs_sf_hy事实表中记录数为2300多,lzs_sf_hydl事实表中记录数为1400多,lzs_sf_djzclx事实表中记录数为1400多,可见经过数据抽取处理后,提高了数据分析效率。

平衡预算应用例:日常各级政府或税务机关为了平衡预算,需要随时知道税收收入进度,用表格展现很简单,按照税种分页卡、表头按照税务机关、预算、收入、进度、占预算百分比即可;然后根据税务机关

逐层展开。

行业税负分析例:分析的例子经常选用“税负”指标。如图4所示,对重点税源,经过两层钻取如图4所示,第一层钻取可知,本地区“环境管理业”税负最高,而“非金属矿物制品业”税负低得多,决策者们可能用于产业结构调整时做政策效果评估或提出进一步措施的参考建议。

行业	应纳税额	销售收入	税负
环境管理业	16,073.71	16,939.96	94.09
建筑业	201.26	625.42	43.63
房地产业	905.82	328.36	17.00
畜牧业	061.81	380.23	15.41
餐饮业	322.99	692.09	15.01
交通业	869.20	583.24	12.84
制造业	189.29	331.49	12.60
开采业	705.09	338.99	11.89
采矿业	108.99	490.61	11.51
服务业	221.39	905.12	10.93
制造业	944.05	941.67	9.48
制造业	335.41	856.83	8.45
制造业	173.28	658.84	8.19
采矿业	982.85	683.63	8.03
采矿业	316.57	658.05	7.98
电力、热	223.21	514.33	7.21
非金属矿物制品业	161,451,733.99	2,288,852,393.96	7.05

图4 税收负担率分行业统计

在上述应用例子中,系统硬件设施是一台IBM570/8CPU,连接EMC2存储,应用服务器用2台PC服务器,前端为一般PC机及其IE,系统软件有Oracle9i, Weblogic等,展现和挖掘暂时通过编程实现,其他有工具包的使用更加多彩,各级税务机关正在受益中。

7 展望

然而,事物总是发展着的,税收分析也不例外,影响税收分析的重要因素至少有如下三个方面:首先是金税工程三期,它投资可能超过100亿元[9],将实现三个目标,本文阐述的税收分析的体系,其目标与上述目标的一致性尚待调较。其次,信息技术的不断发展,征管模式将进一步发生变化。在金税工程三期征管需求编写十条原则中有“流程导向原则”。税收分析一定还要创新,也就是说,在技术业务双重驱动下,税收分析必然与时俱进。第三,有些领域我们还未深入探讨。就拿税收成本分析来说,税收成本包含效率成本、管理成本和遵从成本[10]三个部分,目前见到的税收成本分析报告极少,而且仅仅涉及第一部分管理成本,林雪女士就一直在做这方面的研究[11]。

(下转第84页)

参考文献

- 1 余铁军,张鸿.税务决策支持系统研究与探索.信息技术应用论文集,北京:中国税务出版社,2007.
- 2 吴新年,焦瑞进,刘新利等.微观税收分析指标体系及方法.北京:中国税务出版社,2006.
- 3 李林木.税收遵从的理论分析与政策选择.北京:中国税务出版社,2005.
- 4 吴新联,董丽红.关于近几年增值税与增加值弹性和税负情况的分析.税收经济调研,2004,83(11).
- 5 国家税务总局.2002年地区税收与GDP增长关系聚类分析.税收经济调研,2003,15(4).
- 6 国家税务总局.关于重点税源监控企业税收与税源相关分析的分析.税收经济调研,2004,17(5).
- 7 吴贛昌.概率论与数理统计.北京:中国人民大学出版社,2006.
- 8 国家税务总局.赴澳大利亚税收分析预测方法考察报告.税收经济调研,2006,50(10).
- 9 胡雨.金税三期盼什么.每周电脑报,10(15).
- 10 Tran-Nam. National Tax Journal, 2000,53(2):229-252.
- 11 林雪.中国税务系统征收成本分析.上海:上海财经大学,2006.