

# 基于关联规则挖掘方法的结构 性减税实证研究

孙智勇, 刘 星

(重庆大学 经济与工商管理学院, 重庆 400044)

**摘要:**2009年以来,面对严重的国际金融危机,结构性减税已然成为中国税收政策的一大亮点,更成为刺激经济增长、扩大社会需求的有效财政工具。然而,尽管结构性减税政策适应了当前社会经济客观形势的需要,但也加大了财政收支压力,更对下一阶段如何运用结构性减税政策与完善税收制度相融合提出了新的挑战。文章通过关联规则的数据挖掘分析方法对本次结构性减税前后的各主要税种数据进行关联度的定量分析,从数据分析的角度探讨了减税政策对主要税种以及总的税收收入的影响,并通过分析主要税种与税收收入的相关度变化,对下一步在优化税制结构上的结构性减税方案提供了一些政策建议。

**关键词:**关联规则;结构性减税;税制改革

中图分类号:F813/817

文献标志码:A

文章编号:1008-5831(2010)01-0093-07

## 一、前言

2008年爆发的全球性金融危机,给世界各国的经济发展带来严重的影响。为了刺激中国经济走出困境,实现保增长、扩内需、调结构的目的,2009年国家出台了一系列“结构性减税政策”,为保持经济平稳较快发展起到了重要作用,并成为2009年中国税收政策的一大亮点。减税可以减轻企业负担,增加个人可支配收入,刺激投资与消费需求,更可以利用税收的乘数效应推动经济数倍于减税规模的增长。从2004年以来,中国实行了新一轮的税制改革,并采取了一系列减税政策,包括全面取消农业税、提高工资薪金个人所得税免征额、免征存储利息所得税、提高出口退税率、统一内外资企业所得税制、调整证券交易印花税等措施。可以说,结构性减税政策已经在中国稳步推进了相当长的一段时间。并且,还将继续成为贯彻实施积极财政政策、实现中国经济真正好转的重要一环。但是,随着经济形势的变化,积极财政政策的侧重点和方式也会发生些许变化。结构性减税应该在普遍减税的背景下,适当增加部分产业(行业)的税负,发挥税收调整结构的作用,促进经济结构优化和产业结构升级,增强中国经济发展的后劲,因此,需要根据经济社会的可持续性发展进行有增有减的税收制度改革<sup>[1]</sup>。进入2009年以来,国家据此又进行了一定的税制改革,通过推行增值税转型改革,鼓励企业技术改造,加快经济结构转型。实施新的消费税政策,实现扩大消费需求的宏观经济目标。同时,按照构建资源节约型和环境友好型社会的要求,积极推进资源税改革,积极研究开征社会保障税和环境保护税<sup>[2]</sup>。以上

收稿日期:2009-12-03

基金项目:国家自然科学基金项目(70772100)

作者简介:孙智勇(1977-),男,重庆人,重庆大学经济与工商管理学院博士研究生,主要从事税制改革研

种种步骤表明,国家在着眼当前严峻现实的同时,也已将调整结构这一长远经济发展目标纳入结构性减税的考虑范围,期望通过推进经济结构调整,寻找新的经济增长点促使中国尽快走出经济危机,为财政收入长期增长奠定基础。

在这样的背景下,笔者运用关联规则数据挖掘分析方法,针对1994-2008年的主要税种(国内增值税、企业所得税、国内消费税、个人所得税、营业税)的收入增长情况与2009年各月份的税种收入情况进行定量的关联规则挖掘分析。目前,各种数据挖掘算法已经开始在税务数据分析中得到应用,包括粗糙集、分类、聚类等数据挖掘技术在税务稽查中的应用等,它们能够给纳税评估、税源监控与预测、纳税信用等级评估等决策信息挖掘工作提供一些思路<sup>[3]</sup>。关联规则分析作为数据挖掘的一种重要方法也开始应用在税务系统中,例如,已有文献利用关联规则对历史的稽查数据中纳税人采用的主要违法违规手段之间的关联关系进行了数据挖掘,提高了稽查历史数据的管理水平,为税务稽查提供了有力的决策支持<sup>[4-5]</sup>。笔者则希望通过关联规则的数据挖掘分析方法对2009年结构性减税前后的各税种数据进行关联度的定量分析,以期得到几个主要税种对税收收入的影响因子,从而再对下一步在优化税制结构上的结构性减税措施进行分析评价,以适应调整税制结构、优化经济方式的客观要求。

## 二、关联规则的描述及分析方法

### (一)关联规则定义

数据关联是数据库中存在的一类重要的可被发现的知识。若两个或多个变量的取值之间存在某种规律性,就称之为关联。关联可分为简单关联、时序关联、因果关联。关联分析的目的是要找出数据库中隐藏的关联网。关联规则可定义为<sup>[6-7]</sup>:设T为事务文件, $I = \{i_1, i_2, \dots, i_n\}$ 是一组数据项集。关联规则(associations rule, AR)定义为 $X \rightarrow Y$ ,其中X和Y均包含于I。它表示给定一个事务处理数据库T,在此数据库中找到形如 $X \rightarrow Y$ 的规则,要求X、Y之间存在一个依赖关系,即如果一个元素X的值可以推出另一个元素Y的值,则Y依赖于X,用 $X \rightarrow Y$ 表示,X和Y分别称为关联规则的先导(antecedent或left-hand-side, LHS)和后继(consequent或right-hand-side, RHS)。

在关联规则中,有以下几个重要的概念。

(1)关联规则在D中的支持度(support)是D中事务包含的百分比,即概率;支持度包括每个项集的支持度和每个规则的支持度。

$$D(X \rightarrow Y) = \text{Count}(XY) / \text{Count}(T)$$

(2)可信度(confidence)是包含X的事务中同时包含Y的百分比,即条件概率。

$$D(X \rightarrow Y) = \text{Count}(XY) / \text{Count}(Y)$$

(3)频繁项集(frequent itemsets):包含k个项的项集称为k\_项集。项集的出现频率是包含项集的记录数,简称为项集的频率、支持计数或计数。如果项集的出现频率大于或等于最小支持度与T中记录总数的乘积,项集满足最小支持度;如果项集的出现频率满足最小支持度,则称它为频繁项集。

如果满足最小支持度阈值和最小可信度阈值,我们就认为该关联规则是有趣的。这些阈值由用户或者专家设定。在这次实证数据的挖掘中,由于需要计算的是总税收收入与各主要税种之间的关联规则可信度,而显然各主要税种与总税收收入增长是强关联的,所以就将阈值设为0。关联规则挖掘过程主要包含两个阶段:第一阶段必须先从资料集中找出所有的高频项目组(Frequent Itemsets);第二阶段再由这些高频项目组中产生关联规则(Association Rules)。常用的算法有Apriori算法<sup>[8-9]</sup>、FP树频繁项集算法<sup>[10]</sup>等。

### (二)数值型关联规则

如果属性值是取连续的数值型,则是数值型关联规则。有关数值型关联规则的研究相对较晚,目前的处理方法一般是将属性值进行离散化,转化为对应布尔关联规则来发现强关联规则<sup>[11]</sup>。

记数据集为T,  $T = \{t_1, t_2, t_3, \dots, t_k, \dots, t_n\}$ ,  $t_k = \{i_1, i_2, i_3, \dots, i_m, \dots, i_p\}$ ,  $t_k (k = 1, 2, 3, \dots, n)$ 为事务, $i_m (m = 1, 2, \dots, p)$ 为项。设 $I = \{i_1, i_2, i_3, \dots, i_m\}$ 是T中全体项目组成的集合,I的任何子集X称为T中的项集, $|X| = k$ 称为集合X为k项集。如果 $X \in t_k$ ,称为 $t_k$ 包含项集X;如果是数值型关联,项的集合为 $I_x = I \times P \times P$ ,  $I_x = \{(x, l, u) | x \in I, l \in P, u \in P, l \leq x \leq u\}$ 。即属性x取值在l和u之间。

如果 $\forall (x, l, u) \in X, (x, v) \in T$ ,有 $l \leq x \leq u$ ,则称T支持X。

给定一个数据集T,数值型关联规则是挖掘出满足最小支持度(minsupport)和最小可信度(minconfidence)的规则。对于数值型关联规则的研究,一般是将其转化为布尔关联规则来进行挖掘,挖掘步骤一般如下:

(1)对于数值属性x,取值范围为 $[l, u]$ ,应用一定的算法划分成不同的区间;

(2)将划分后得到的 $\langle l_k, u_k \rangle$ 的数值,映射为布尔属性,所有映射来的布尔属性构成项目集;

(3)根据给定的最小支持度,找出频繁项目集;

(4)产生关联规则;

(5)确定感兴趣的强关联规则,并作为有用规则输出。

采用转化为布尔关联规则来进行挖掘的方法的焦点集中在如何对数值属性进行划分上。目前,对属性值划分的算法主要有等宽、等深、聚类等等。

### 三、利用关联规则方法分析主要税种数据

笔者以 1994 - 2008 年中国的税收收入以及 2009 年 1 - 11 月(因撰稿时 12 月份数据还未出,以税务统计年鉴,国家财政部税务司和国家税务总局网站的网站发布数据为来源)的税收收入作为关联规则的事务集,期望能够通过关联规则挖掘得到各主要税种与总税收收入之间的相关度,即相对应于关联规则中的可信度。显然,各个税种收入增长与总税收收入增长的关系是相当密切的。传统的贝叶斯相关性分析方法是假定各个税种之间独立的前提下,再进行相关度分析。但实际上各个税种之间的相关度也是相当大的。而关联规则分析方法则是将所有数据一起进行分析,甚至可以找出不同税种之间的变化关联度,因此,其准确性应该比传统的贝叶斯相关性分析方法更高。我们以总税收收入增长率、增值税增长率、消费税增长率、企业所得税增长率(2008 年前分为国内企业所得税与外资企业所得税,将其合并计算)、消费税增长率、个人所得税增长率、营业税增长率作为关联规则的事务数据集。

(一)进行数据预处理将数值型转换成布尔型数据集

数据预处理的过程就是需要将数值型的增长率数据进行离散化处理,如何离散化处理则是本计算模型的关键。

首先假设  $Y =$  总税收收入增长率;  $X_1 =$  增值税增长率;  $X_2 =$  企业所得税增长率;  $X_3 =$  消费税增长率;

$X_4 =$  个人所得税增长率;  $X_5 =$  营业税增长率。  $V_1 =$  增值税;  $V_2 =$  企业所得税;  $V_3 =$  消费税;  $V_4 =$  个人所得税;  $V_5 =$  营业税。目前对离散化的方法就是划分区间,笔者采用各税种收入的加权平均值作为分段标准,即有:  $\forall |X_i - Y| \leq |X_i * (\sum_{j=1}^5 V_j) - Y| - \epsilon$ ,就认为  $X$  落入关联区间,其中,  $\epsilon$  为可变参量。由于笔者只是利用关联规则计算各税种与总税收收入的关联度,而总税收收入等于各税种收入之和,因此,其关联度是肯定存在的。为了更好地体现各税种与总税收收入的关联度的差异,引入  $\epsilon$  能够使各税种划分区间更加弹性,更有利于后面的分析,可以考虑通过数据挖掘或者专家指定的方法来定义  $\epsilon$ 。在实际计算过程中,  $\epsilon$  取各增加率的平均值。

如果  $X_i$  落入关联区间,则在准备做关联规则挖掘的布尔型的事务数据表中添加  $X_i$ ,反之则填充值,最后生成一个布尔型的事务数据集。由于需要分析的数据量并不是很大,笔者将直接采用最简单又最明确的 Apriori 算法来进行关联规则的可信度计算。

(二)对离散化后的数据得到频繁项集并计算各税种与总税收收入的关联度

由于以税收收入增长率为基础,所以支持度均为 100%,  $X_i \rightarrow Y$  这种规则的可信度可以直接用  $X_i$  出现的数量除以总的数据集数来计算。我们这里只计算了与总税收收入有关的 2 阶频繁项集,表 1、表 2 分别为 1994 - 2008 年的各主要税种、总税收收入及其环比增长率,表 3、表 4 则为 2009 年 1 - 11 月的各主要税种、总税收收入及其环比增长率。表 5 与表 6 则是进行了离散化方法后产生的事务数据集。通过 Apriori<sup>[8-9]</sup> 算法计算出各条关联规则,其中与总税收收入相关 2 阶频繁项集的可信度则如表 7 所示。

表 1 1994 - 2008 年总税收及各税种的增长数据

单位:亿元

年份	总税收收入 (Z)	国内增值税 (V <sub>1</sub> )	企业所得税 (V <sub>2</sub> )	国内消费税 (V <sub>3</sub> )	个人所得税 (V <sub>4</sub> )	营业税 (V <sub>5</sub> )	其他
1994	5 126.88	2 308.34	708.49	516.00	72.67	670.02	851.36
1995	6 038.04	2 602.33	878.44	541.48	131.49	865.56	1 018.74
1996	6 909.82	2 962.81	968.48	620.23	193.19	1 052.57	1 112.54
1997	8 234.04	3 283.92	963.18	678.70	259.93	1 324.27	1 724.04
1998	9 262.80	3 628.46	925.54	814.93	338.64	1 575.08	1 980.15
1999	10 682.58	3 881.87	811.41	820.66	414.31	1 660.56	3 093.77
2000	12 581.51	4 553.17	999.63	858.29	660.37	1 868.78	3 641.27
2001	1 5301.38	5 357.13	2 630.87	929.99	995.99	2 064.09	3 323.31
2002	17 636.45	6 178.39	3 082.79	1 046.32	1 211.00	2 450.33	3 667.62
2003	20 017.31	7 236.54	2 919.51	1 182.26	1 418.00	2 844.45	4 416.55
2004	24 165.68	9 017.94	3 957.33	1 501.90	1 737.05	3 581.97	4 369.49
2005	28 778.54	10 792.11	5 343.92	1 633.81	1 728.94	4 232.46	5 047.30
2006	34 804.35	12 784.81	7 039.60	1 885.69	2 452.32	5 128.71	5 513.22
2007	45 621.99	15 470.11	8 769.47	2 206.9	3 185.54	6 581.99	9 407.98
2008	54 219.62	17 996.9	11 173.05	2 567.8	3 722.19	7 626.33	11 133.35

资料来源:1994 - 2008 年中国税务统计年鉴。  
 欢迎访问重庆大学期刊社 <http://qks.cqu.edu.cn>

表2 1995-2008的总税收收入、各主要税种与上年环比的增长率

年份	总税收收入 增长率( $Y$ )	国内增值税 增长率( $X_1$ )	企业所得税 增长率( $X_2$ )	国内消费税 增长率( $X_3$ )	个人所得税 增长率( $X_4$ )	营业税 增长率( $X_5$ )
1995	17.77%	12.74%	23.99%	4.94%	80.94%	29.18%
1996	14.44%	13.85%	10.25%	14.54%	46.92%	21.61%
1997	19.16%	10.84%	-0.55%	9.43%	34.55%	25.81%
1998	12.49%	10.49%	-3.91%	20.07%	30.28%	18.94%
1999	15.33%	6.98%	-12.33%	0.70%	22.35%	5.43%
2000	17.78%	17.29%	23.20%	4.59%	59.39%	12.54%
2001	21.62%	17.66%	163.18%	8.35%	50.82%	10.45%
2002	15.26%	15.33%	17.18%	12.51%	21.59%	18.71%
2003	13.50%	17.13%	-5.30%	12.99%	17.09%	16.08%
2004	20.72%	24.62%	35.55%	27.04%	22.50%	25.93%
2005	19.09%	19.67%	35.04%	8.78%	-0.47%	18.16%
2006	20.94%	18.46%	31.73%	15.42%	41.84%	21.18%
2007	31.08%	21.00%	24.57%	17.03%	29.90%	28.34%
2008	18.85%	16.33%	27.41%	16.35%	16.85%	15.87%

表3 2009年1-11月总税收收入及各税种的增长数据

单位:亿元

月份	总税收收入	国内增值税	企业所得税	国内消费税	个人所得税	营业税	其他
1	5 984.96	1 690.08	1 495.54	256.08	432.25	1 013.99	1 097.02
2	3 956.92	1 475.52	267.89	430.74	442.56	509.23	830.98
3	4 122.02	1 426.06	404.56	322.90	347.32	561.37	1 059.81
4	5 672.89	1 409.64	1 533.80	375.03	318.73	840.30	1 195.39
5	5 915.20	1 458.48	2 086.66	379.49	290.32	601.19	1 099.06
6	5 742.76	1 838.36	1 122.84	429.23	310.24	766.78	1 275.31
7	6 168.10	1 417.83	1 786.79	434.33	306.26	919.07	1 303.82
8	4 923.97	1 491.63	830.42	453.70	292.36	681.48	1 174.38
9	5 009.90	1 655.92	476.41	467.79	309.85	719.71	1 380.22
10	6 386.60	1 595.64	1 833.78	495.52	297.12	911.23	1 253.31
11	4 646.92	1 612.69	340.23	400.85	278.54	681.21	1 333.40

资料来源:国家财政部税务司和国家税务总局计统司的网站发布数据。

表4 2009年1-11月总税收收入、各主要税种收入环比的增长率

月份	总税收收入环比	国内增值税环比	企业所得税环比	国内消费税环比	个人所得税环比	营业税环比
2	-33.89%	-12.70%	-82.09%	68.21%	2.39%	-49.78%
3	4.17%	-3.35%	51.02%	-25.04%	-21.52%	10.24%
4	37.62%	-1.15%	279.13%	16.14%	-8.23%	49.69%
5	4.27%	3.46%	36.05%	1.19%	-8.91%	-28.46%
6	-2.92%	26.05%	-46.19%	13.11%	6.86%	27.54%
7	7.41%	-22.88%	59.13%	1.19%	-1.28%	19.86%
8	-20.17%	5.21%	-53.52%	4.46%	-4.54%	-25.85%
9	1.75%	11.01%	-42.63%	3.11%	5.98%	5.61%
10	27.48%	-3.64%	284.92%	5.93%	-4.11%	26.61%
11	-27.24%	1.07%	-81.45%	-19.11%	-6.25%	-25.24%

表 5 预处理后的事务数据集

年份	国内增值税环比	企业所得税环比	国内消费税环比	个人所得税环比	营业税环比
1995	$X_1$	$X_2$			
1996	$X_1$	$X_2$	$X_3$		
1997			$X_3$		$X_5$
1998	$X_1$				
1999				$X_4$	$X_5$
2000	$X_1$	$X_2$			$X_5$
2001	$X_1$		$X_3$		
2002	$X_1$	$X_2$		$X_4$	
2003			$X_3$	$X_4$	$X_5$
2004	$X_1$			$X_4$	
2005	$X_1$	$X_2$	$X_3$		
2006	$X_1$	$X_2$	$X_3$		$X_5$
2007	$X_1$	$X_2$		$X_4$	
2008	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$X_i$ 可信度	0.785 7	0.571 4	0.500 0	0.428 6	0.428 6

表 6 预处理后的事务数据集

月份	国内增值税环比	企业所得税环比	国内消费税环比	个人所得税环比	营业税环比
2	$X_1$				$X_5$
3	$X_1$	$X_2$			
4			$X_3$		$X_5$
5	$X_1$	$X_2$	$X_3$	$X_4$	
6		$X_2$	$X_3$		
7	$X_1$	$X_2$	$X_3$		$X_5$
8		$X_2$	$X_3$	$X_4$	
9	$X_1$		$X_3$	$X_4$	$X_5$
10					$X_5$
11		$X_2$	$X_3$	$X_4$	$X_5$
$X_i$ 可信度	0.500 0	0.600 0	0.700 0	0.400 0	0.600 0

表 7 2009 年结构性减税调整前后各主要税种与总税收收入的可信度对比

	国内增值税	企业所得税	国内消费税	个人所得税	营业税
1994 - 2008 年(调整前)	0.785 7	0.571 4	0.500 0	0.428 6	0.428 6
2009 年调整后	0.500 0	0.600 0	0.700 0	0.400 0	0.600 0

#### 四、分析与展望

通过关联规则分析方法,我们得到了各主要税种对税收收入的支持度。通过比较发现,经过 2009 年的结构性减税调整之后,中国在全国所有地区、所有行业全面实施增值税转型改革,增值税转型改革

和降低小规模纳税人征收率影响了国内增值税增收。2009 年减税后,2009 年 1 - 11 月份国内增值税直接收入(不含免抵调库)完成 1 485.957 88 亿元,同比下降 1.6%,国内增值税与总税收收入的相关度在 2009 年减税前后分别为 0.785 7、0.500 0,下降变

化明显。但是,2009年1-11月国内增值税收入占税收总收入的比重为29.2%,2008年为33.2%,通过对比可以发现增值税在总税收收入的所占比重变化并不是很大,依然是所有税种中与税收增长相关性比较高的一个税种。这意味着作为中国第一大税种,其未来增值税减税空间依然很大。

企业所得税对税收增长的相关度则略有变化但影响不大,1-11月份企业所得税收入完成9875.96亿元,同比增长1.1%。分月看,企业所得税收入从6月份开始出现大幅度增长,6、7、8、9、10、11月等6个月分别比去年同期增长10.1%、17.2%、239.2%、91.1%、14.8%和90.2%。由此可见,上半年企业所得税收入大幅减收的主要原因是来自于受金融危机影响,企业利润普遍下滑;企业所得税2008年“两法合并”后,某些优惠政策实施带来的减收效应等因素影响。但经过半年的磨合,减税政策的效应逐步显现,随着企业负担的减轻,在内外资税负公平的激励下,国内企业大力推进产业结构优化升级,更加注重提高自主创新能力,经营状况逐步回升。

国内消费税在今年减税过程中增幅明显,对税收增长的相关度变化最大,从0.5直接攀升到减税后的0.7。消费税同比增速由1月份的13.2%提高到11月份的119.3%。前11月共计完成4445.6583亿元,同比增长86.8%。应当指出,国家采取的调整成品油消费税率、提高烟产品消费税率、加强白酒消费税征收管理、完善小汽车行业消费税政策等措施在增收消费税的同时,也减轻了汽车、摩托车等受金融危机影响较大行业的税负,更体现了减少某些行业税负的同时,增加某些行业税负的增减结合原则。

个人所得税在减税措施进行后对税收增长的相关度变化不大,但其税种收入增速回落明显,说明在屡次增加工资薪金所得扣除标准和暂免征收存款利息所得税后,个人税后收入是有所增加的。另据统计,从1994年到2008年,个人所得税收入占GDP的比重由0.15%上升至1.24%,占税收收入的比重由1.4%上升至6.4%。以上分析表明,个人所得税具备向主体税收进发的潜力。同时,如何在全面调节收入分配差距的前提下求得中低收入者个人所得税税收负担的减轻,让高收入者比低收入者多纳税并以此调节居民之间的收入分配差距,应是我们未来结构性减税过程中考虑的目标。

营业税收入依旧保持较快增长,其与总税收收入的相关性从0.4268增加到了0.6,与国内消费税收入一道成为流转税制中抵消国内增值税税收收入减少影响的中流砥柱。这说明在实行这次结构性减

税政策中,营业税的调整达到了中央扩大内需、刺激经济增长的目标。

综上所述,2009年所采取的结构性减税政策效应已经初步显现。虽然上半年受到金融危机的影响,减税政策中增值税转型改革影响了国内增值税增收;企业汇算清缴适用所得税税率降低导致企业所得税下降较多;取消或降低多种出口商品的出口关税等政策调整在很大程度上影响了关税收入;多次提高出口退税率也造成税收总收入减少。但是,在积极财政政策和适度宽松货币政策刺激下,各项宏观调控措施全面发力,中国经济在二季度触底回升,一些重要的宏观经济指标出现积极变化,使部分税种实现较快增长,从6月份开始,单月税收收入已经扭转了上半年连续5个月减收的局面,并且有加快增长的趋势,到11月份累计税收收入同比小幅正增长。特别是消费税税收政策调整,2009年以来基础设施建设的投资力度加大导致营业税收入实现大幅增长等因素,增加了一部分税收收入,部分抵消了结构性减税的减收效应。因此,从长远来看,我们必须采取一种“增减结合、调整结构、实质下降”的税制改革方案,既要在保障和改善民生的税种政策方面减税,又要对某些特定群体和项目进行增税。例如,对高污染企业开征环境税,规范并扩大消费税课税范围(尤其是对高能耗企业、高档消费行为等开征或提高税率),推进资源税重大改革,实行综合与分类相结合的个人所得税制等。更重要的是,在以上基础上最终实现结构性减税的出发点和归宿:让纳税人实质税负水平下降。这样就可以真正实现中央优化税制结构、服务于经济增长和经济发展方式转变的长远目标。

#### 参考文献:

- [1] 贾康,张学诞,施文波. 深化中国税制改革应选择有增有减的结构性调整[J]. 税务研究,2009(9):3-8.
- [2] 丛明,朱乃肖. 积极稳妥地推进税制改革[J]. 税务研究,2007(9):3-5.
- [3] 蒋丽华. 数据挖掘技术在税务稽查中的应用[J]. 税务研究,2007(5):84-86.
- [4] 王敏,刘希玉. Apriori算法在税务系统中的应用[J]. 计算机技术与发展,2009,11(19):175-177.
- [5] 崔贤岳,李际军. 数据挖掘技术在税务系统中的应用[J]. 计算机工程,2007,33(14):283-284.
- [6] HAN JIA-WEI, KAMBER M. 数据挖掘[M]. 范明,孟小峰,译. 北京:机械工业出版社,2005.
- [7] ZHU Hong-lei, XU Zhi-gang. An effective algorithm for mining positive and negative association rules [C]. //Computer

- Science and Software Engineering, 2008 International Conference. 2008, 4:455-458.
- [8] 朱孝宇,王理冬,汪光阳. 一种改进的 Apriori 挖掘关联规则算法[J]. 计算机技术与发展,2006,16(12):89-90.
- [9] 马盈仓. 挖掘关联规则中 Apriori 算法的改进[J]. 计算机应用与软件,2004, 21(11): 82-83.
- [10] HAN Jia-wei, JIAN Pei. Mining frequent patterns without candidate generation: A frequent-pattern tree approach [J]. Data Mining and Knowledge Discovery, 2004, 8(1): 53-87.
- [11] RASTOGI R, KYUSEOK S. Mining optimized association rules with categorical and numeric attributes [C]//Knowledge and Data Engineering. IEEE Transactions, 2002, 14(1): 29-50.

## An Empirical Study on Structural Tax Reduction Based on Association Rules Mining

SUN Zhi-yong, LIU Xing

(College of Economics and Business Administration, Chongqing University, Chongqing 400044, P. R. China)

**Abstract:** Since 2009, facing intensified international financial crisis, structural tax reduction has already become not only a major highlight of China's tax policies but also an effective financial tool to stimulate the economic growth and social needs. Although the policy of structural tax reduction adapted to the needs of the current socio-economic objective situation, it has increased the pressure on revenues and expenditures and posed a new challenge, involving how to integrate the use of the policy of structural tax reduction with the improvement of the tax system in the next phase. In this paper, association rules data mining analytical method is introduced to make the quantitative data analysis of some main categories of taxes before and after the current tax reduction. And the impact of the tax reduction policy on some main categories of taxes and the total tax revenues is analyzed. Additionally, some suggestions on the following goals of structural tax reduction that feature the basis of the adjustment of the tax system are proposed through the analysis of main taxes associated with the degree of change in tax revenue.

**Keywords:** association rules; structural tax reduction; tax reform

(责任编辑 傅旭东)