

# 我国规模以下工业连续性抽样 调查方案设计研究<sup>\*</sup>

刘建平 李绍泰

**内容提要:** 本文在分析我国规模以下工业抽样调查现状和存在问题的基础上,提出改革建议,对工业企业和个体工业户分别设计了连续性抽样调查方案,对工业企业采用三维平衡三水平轮换模式,对个体工业户采用二维平衡单水平轮换模式,两种调查方案都分别设计了两种轮换模式,以适应各个地区的实际情况。

**关键词:** 规模以下工业企业; 个体工业户; 连续性抽样调查

中图分类号: C811 文献标识码: A 文章编号: 1002-4565(2014)01-0033-09

## The Design Research on Successive Sampling Survey Program for China's Industry under the Scale

Liu Jianping & Li Shaotai

**Abstract:** This paper analyzes the status and problems that based on China's industry under the scale, proposes some reform measures. We design successive sampling survey program separately on the industrial enterprises and individual industrial users, the three-dimensional balanced multi-level rotation for industrial enterprises and the two-way balanced one-level rotation for self-employed industrial household. Due to the difference of economic development between regions, two rotation sampling designs are provided for each of survey programs in order to meet the actual need of different areas.

**Key words:** Industrial Enterprises under the Scale; Individual Industrial Household; Successive Sampling Survey

我国早在新中国成立之初就开始建立工业统计体系。1950年3月,进行了第一次全国工矿企业普查,在此基础上,逐步建立和完善了月度、季度和年度工业生产、财务状况统计制度。1998年以前,我国的工业统计中,工业定期报表均按乡及乡以上法人工业企业统计来反映整个工业的情况。随着我国经济的发展,小企业和个体户快速增加,不仅数量庞大,而且经济总量也日益增大,1998年以后,将工业统计调查单位划分为规模以上工业(即年主营业务收入500万元及以上的工业法人企业,2011年提高到2000万元及以上)和规模以下工业。现行统计规模以上工业实行全面调查,每个月通过联网直报系统申报,规模以下则实施每季度一次的抽样调查<sup>[1]</sup>,以满足国民经济核算的需要。连续性抽样调查理论在国外得到充分实践,并取得了很好的实施效果<sup>[2]</sup>。本文从连续性抽样调查的角度出发,提出改革我国规模以下工业抽样调查的思路,并设计了

规模以下工业企业和个体工业户的连续性抽样调查方案。

### 一、现状分析

#### (一) 规模以下工业实施抽样调查的历程

我国规模以下工业抽样调查分为规模以下工业企业(法人企业)调查和个体经营工业单位(简称个体工业户)调查。我国工业抽样调查经历了以下三个阶段<sup>[3]</sup>: 1996—1997年为试点阶段,1996年底开始进行工业抽样调查试点,1997年在全国试行。范围为乡以下工业和乡以上年产值500万元以下的工业企业(单位),乡以上年产值500万元以下的工业企业(单位)采用目录抽样,乡以下工业采用二阶段

<sup>\*</sup> 本文为国家社科基金重点项目“我国政府统计调查体系研究”(12ATJ002)、教育部人文社科基金项目“改革和完善我国政府统计调查体系研究”(11YJA910006)的阶段性成果。

整群抽样; 1998—2002年为推广实施阶段,调查范围调整为年产品销售收入500万元以下的非国有工业企业和全部个体工业户,1999年和2000年规模以下工业抽样调查的调查频率都为两次,从2001年起改为四次。方法是对具有名录库的企业采用目录抽样,对城镇个体户采用二阶段目录抽样,对村以下企业和个体户采用二阶段或一阶段整群抽样。2003年以后进入正常运行阶段。在该阶段,统一了各省的抽样方法,确定每季度进行一次调查,目录企业采用目录抽样,非目录企业采用整群抽样,个体工业户采用一阶段整群抽样。

(二) 存在的问题

1. 实施固定样本调查,既不能反映总体结构变化,又易于引起样本老化。我国的工业抽样调查虽然进行过多次改革,但每次改革都是在一次性抽样设计的基础上进行,目前对规模以下工业企业采用的固定样本调查,就是一次性抽样的结果,第一年抽取的样本企业,连续调查五年,五年后进行一次性更新。这种传统的对同一调查对象经常进行的一次性抽样设计调查,不仅不能有效利用调查资源,也不利于形成科学的时间序列数据和面板数据。一次性抽样设计和固定样本调查的优势在于不用频繁地更换调查样本,减少组织者的工作负担和调查费用。但固定样本存在着弊端:第一,工业企业总体结构变化快。随着时间的变化,存在大量企业在规模上下转换的现象,企业由规下转为规上的可能性更大;规模以下企业,特别是小微企业,新生和消亡都很频繁。这些都会造成总体结构的变化,样本长期保持固定不变,会大大降低样本的代表性。第二,样本企业调查时间过长,容易引起被调查者的反感或不配合,从而降低调查质量,增大非抽样误差。

2. 抽样框更新维护不到位。抽样框的完整性影响着抽样的估计精度,但是目前我国各个地区企业名录库的更新维护时间和频率并不一致,有些地区更新频率快,即时或每个月都根据工商、税务、质监等部门的行政记录进行更新维护,但这样的地区不多,更多的地区是在经济普查时进行大规模的一次性更新。规模以下工业企业和个体工业户数量巨大,消亡和新生相当频繁,变化更新很快。抽样框更新维护不及时,即使采用科学的连续性抽样设计,也不能保证样本对总体的代表性,会使抽样估计结果

的准确性和可靠性大打折扣。

3. 统计调查机构设置不合理。我国目前的统计调查机构设置分为四级:国家—省(直辖市、自治区)—市(地区、州、盟)—一部分县(区、市)调查队。规模以下工业企业和个体户数量巨大,2009年个体工商户达到了6000万个<sup>[4]</sup>,分散在全国2000多个县(市)。对于抽中的企业,设置调查队的县(市)由调查队组织实施,没设调查队的由县(市)统计局负责实施。如果总是抽取设置了调查队的县(市)的企业或个体工业户进入样本,会降低样本的代表性,并挫伤调查人员和调查对象的积极性,使调查质量下降。如果同时选取了没设调查队的县(市)的企业或个体工业户进入样本,会额外增加当地统计局的负担,同样不能有效保证调查的质量。

## 二、改革建议

(一) 建立规模以下工业连续性抽样调查体系<sup>[5]</sup> 实施轮换样本调查

我国的统计调查制度以周期性的普查为基础,经常性的抽样调查为主体,但经常性的抽样调查往往是独立的一次性调查,并不是连续的抽样调查。我国目前对规模以下工业企业采用的固定样本调查,就是一次性抽样的结果,虽然形式上也是连续性抽样调查,但是科学性不够。实施科学的连续性抽样调查,必须根据调查对象的变化特征有针对性地进行科学设计,设计科学的连续性抽样调查方案,选择科学的样本轮换模式,建立科学的连续性抽样调查体系。科学的样本轮换模式不仅保留了固定样本调查的优点,同时经过样本轮换,能减少样本老化和反映总体结构的发展变化,能得到更加准确的反映调查对象发展变化的时间序列数据和面板数据。根据这些数据可以更好地研究调查总体的变化发展规律。

(二) 及时更新维护抽样框

我国政府统计近两年实施的“四大工程”提出要建设真实完整、及时更新的基本单位名录库。各项统计调查使用统一的名录库,对基本单位名录库的维护更新有一定的促进作用。此外还要加强与工商、税务、质监等部门的合作,充分利用行政记录,更新维护新登记和注销的企业,及时增加由规模以上转入规模以下的企业,删除名录库中消亡或转为规

模以上的企业,真正做到名录库抽样框不重不漏、真实完整。

### (三) 取消县级调查队,充实地市级调查队<sup>[2]</sup>

取消县级调查队,充实加强地市级调查队,便于规模以下工业抽样调查按地市进行分层,定期实施样本轮换,不仅能满足市一级的核算需求,同时也能避免样本老化,提高样本的代表性和调查质量。

## 三、调查方案设计思路

本文对规模以下工业抽样调查按工业企业和个体工业户分别进行设计。

### (一) 工业企业连续性抽样设计

#### 1. 抽样框整理。

目前我国规模以下工业企业的调查区分两种情况选择抽样方式:第一,对有名录库的工业企业采用一阶段目录抽样,根据各省建立的企业名录库直接抽取样本企业。第二,对未包括在企业名录库中的非目录企业,采用一阶段整群抽样,在省一级直接抽取整群单位——村或居委会作为样本。

目录企业的抽样框,根据经济普查得到的省一级企业名录库建立本次调查的抽样框。抽样框内容应包括企业的基本属性指标和基本价值量指标。具体包括:企业详细名称、地址、组织机构代码、行业分类、企业登记注册类型等基本属性指标,以及工业总产值或营业收入等基本价值量指标。同时要结合工商行政登记部门、税务部门的资料整理和更新抽样框,如名录库中的一些企业在调查期间已经消亡或因经营效益好已经转变为规模以上企业,另外一些新增的工业企业也未包括在名录库中。因此在进行抽样前,要及时整理抽样框,对遗漏企业和新增企业通过相关部门的行政记录进行补充,对抽样框中重复的企业、已经消亡和不属于调查范围内的企业进行剔除。

非目录企业指不在省名录库中的企业。由于名录库的建立时间和调查时间的不同步,可能在建立企业名录库时有遗漏,也可能因为一些新生的企业没收入名录库中。因此非目录企业的抽样框采用区域框,利用个体工业一阶段整群抽样的样本,对整群样本范围内的规模以下工业企业进行核查,确认非目录企业样本。

对于规模以下工业企业,应该以目录企业抽样框为主体,以整群抽样框为补充的双重抽样框

设计,即样本企业的抽取从双重框中进行。非目录企业的调查是以整群单元——村或居委会为抽样框的,这为双重抽样框的使用提供了可能性,且抽样框覆盖的目标总体更全,从而提高抽样效率。在进行连续性抽样调查时,第一次调查确认的非目录企业,及时更新补充到企业名录库中,第二次调查时,将这些企业根据营业收入划分到对应的轮换组。

#### 2. 抽样实施步骤。

工业调查以省(直辖市、自治区)为总体,由工业企业所在的各地市(地区、州、盟)组织实施调查。对工业企业进行连续性调查时,对目录企业实施一阶段目录抽样。对非目录企业,因未包含在名录库中,实施第一次调查时,实施步骤与个体工业户调查相同,第一次调查完成后及时更新补充名录库;第二次调查时,则可按目录企业的抽样步骤实施,目录企业的抽样实施步骤分为以下几步:

第一步,对名录库中的目录企业进行分层。目录企业按企业所在地的地市进行分层——地域层;为了满足以全国分行业大类为总体抽样精度的要求,每个地域层中可以按企业的行业类型进一步分层——行业层;行业层的目录企业再按营业收入的高低进行分层——规模层。按地市和行业进行分层,是为了提高估计的精度,满足各地区推算总体和行业的需求。进一步分层的限制条件是,每个最终层内至少要包含4个以上企业。

第二步,从规模层中抽取样本。使用永久随机数法(PRN),永久随机数法要优于简单随机抽样。在连续性抽样调查中,一个企业被赋予一个随机数,则在以后的调查中都可以使用这个随机数,不仅能方便抽取层中样本,也方便在样本轮换时控制样本。在进行样本分配时,收入大的企业给予更高的抽样比,层与层之间抽样比差距不要太大,避免加大方差。

第三步,进行样本轮换。各地市按企业的营业收入划分成若干规模层,也可看作若干轮换组,每个轮换组包含一定数量的企业。在每月的调查中,从每个轮换组中抽取至少4个企业,组成最终的轮换样本进行调查。抽取至少4个企业的原因是为了防止由企业在调查期间消亡、由规模以下转为规模以上和无回答等情况造成该轮换组企业缺失。各地需要根据当地的实际情况选择不同的轮换模式进行后

期调查的轮换。

3. 工业企业轮换模式的设计。

规模以下工业企业连续性抽样调查的轮换 根据连续性抽样理论和我国的实际情况,认为我国比较适合三维平衡三水平的轮换模式,原因主要有以下两个方面:

第一,三维平衡三水平的轮换模式本身优势。三水平轮换样本调查不仅吸收了固定样本的优点,同时经过轮换更新样本。在三维平衡模式下,能够有效地减少轮换组偏差和回忆的偏差。一次调查可以获得连续三个月的信息,可以有效得到月度总体总值(均值)和年度总体总值(均值)的估计量,可以估计月度间及年度间总体的变化情况。

第二,三维平衡三水平轮换模式的可行性。工业企业是依法成立的能进行独立核算的法人单位,跟个体工业户相比,它的核算基础更好,更符合三水平模式需要回忆两个时期的要求。我国目前规模以下工业企业抽样调查每季度调查一次,对于被调查的企业,每季度接受一次调查,这跟三水平模式相近。在三水平模式下,调查一次,得到调查企业连续三个月的信息,不会增加企业的调查负担,且根据企业前期调查资料,可利用回归估计提高精度。

本文提供了 4~7~3(3) 和 3~5~2(3) 两种三维平衡三水平轮换模式,供各地市根据本地区工业企业的实际情况选择不同的轮换模式。4~7~3(3) 三维平衡轮换模式的特点是轮换速度慢,轮换周期比较长,企业从第一次接受调查,经过七轮调查至完成退出调查需要 26 个月的时间,适合工业企业少且分散的地区。此外,我国从 2011 年起把规模以上工业企业的统计起点标准从年主营业务收入 500 万元提高到 2000 万元,规模起点标准提高后,纳入月度规模以上工业统计的企业数量减少较多,而需要抽样调查的规模以下工业企业数量增加较多。根据 2009 年工业年报数据测算,全国 2000 万元以上工业企业单位数比 500 万元以上企业数减少 40.6% [6]。4~7~3(3) 轮换模式下,有 7 个样本轮换组,适合地区内行业比较多、工业企业间营业收入差异大,需要多分组提高精度的地区。3~5~2(3) 的特点是轮换速度相对较快,轮换周期相对较短,从开始对企业进行调查到企业调查完成退出只需要

18 个月的时间,这种模式只有 5 个轮换组,适合企业间收入差距小,行业比较少的地区,同时也适合企业变化较快的行业和地区。

4~7~3(3) 三维平衡三水平轮换模式的实施步骤如下:各地市按照工业企业营业收入的高低分成 7 个轮换组,从每个轮换组中抽取至少 4 个企业形成轮换样本。在连续各期的调查中,轮换样本按照 4~7~3(3) 的轮换模式进行 2 轮的调查。在第 1 轮调查中,每个样本企业间隔 3 个月调查一次,连续调查 4 次,即在连续的 10 个月中连续调查 4 次,在随后的 9 个月中退出调查。之后进行第 2 轮调查,又连续调查 3 次,最后永久退出调查,同时从该轮换组中抽取新的轮换样本进行新一轮的调查。如表 1 中第 4 个轮换样本的第 5 个轮换组,在第  $t$  月时接受第 1 次调查,连续调查 4 次后暂时退出,在第  $t+19$  月时进行第 2 轮的调查,连续调查 3 次,在第  $t+25$  月时接受第 7 次调查,最后永久退出。表 1 中“①、②、③、④、⑤、⑥、⑦”表示第几次接受调查,“I”、“II”分别表示搜集轮换样本前一期和前两期的信息。

3~5~2(3) 三维平衡三水平轮换模式的实施步骤与 4~7~3(3) 的步骤类似:各地市按照工业企业营业收入的高低分成 5 个轮换组,从每个轮换组中抽取至少 4 个企业形成轮换样本。每个样本企业间隔 3 个月调查一次,连续调查 3 次,在随后的 7 个月中退出调查。之后又连续调查 3 次,最后永久退出调查,同时从该轮换组中抽取新的轮换样本进行新一轮的调查。

上述两种模式 4~7~3(3) 和 3~5~2(3) 属于  $r_{11} \sim r_{21} \sim r_{12}$  形式。在每个调查月份,调查样本都包括所有的  $r_{11} + r_{12}$  个轮换组,每月每个轮换组都包括三个不同的轮换样本,其中一个轮换样本在回忆时刻为“I”的调查内容,即搜集前一期的信息;另一个轮换样本在回忆时刻为“II”的调查内容,即搜集前两期的信息。从横向上看,调查样本都是由  $r_{11} + r_{12}$  个轮换样本构成,其中一个轮换样本在接受第 1 次调查,另一个样本在接受第 2 次调查,依此类推,直到最后一个样本在接受第  $r_{11} + r_{12}$  次调查,即轮换样本在横向上达到平衡;从纵向上看,对于每个轮换组中的轮换样本,在某月接受第 1 次调查,在另一个月接受第 2 次调查,依此类推,最后一个月接受最后一次调查,即轮换样本在纵向上达到平衡。

表 1 4~7~3(3) 三维平衡三水平轮换模式的轮换过程

轮换样本 $\alpha$	1							2							3							4														
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7								
第 $t$ 月	⑦	I	II	⑥	I	II	⑤	I	II								④	I	II	③	I	II	②	I	II	①	I	II								
$t+1$		⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I	II	②	I	II	①	I									
$t+2$			⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I	II	②	I	II	①									
$t+3$				⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I	II	②	I	II									
$t+4$					⑦	I	I	I	II	⑤	I	II								④	I	II	③	I	II	②	I									
$t+5$						⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I	II	②									
$t+6$							⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I	II									
$t+7$								⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③	I									
$t+8$									⑦	I	II	⑥	I	II	⑤	I	II							④	I	II	③									
$t+9$										⑦	I	II	⑥	I	II	⑤	I	II							④	I	II									
$t+10$											⑦	I	II	⑥	I	II	⑤	I	II							④	I									
$t+11$												⑦	I	II	⑥	I	II	⑤	I	II							④									
$t+12$													⑦	I	II	⑥	I	II	⑤	I	II															
$t+13$														⑦	I	II	⑥	I	II	⑤	I	II														
$t+14$															⑦	I	II	⑥	I	II	⑤	I	II													
$t+15$																⑦	I	II	⑥	I	II	⑤	I	II												
$t+16$																	⑦	I	II	⑥	I	II	⑤	I	II											
$t+17$																		⑦	I	II	⑥	I	II	⑤	I	II										
$t+18$																			⑦	I	II	⑥	I	II	⑤	I	II									
$t+19$																				⑦	I	II	⑥	I	II	⑤	I	II								
$t+20$																					⑦	I	II	⑥	I	II	⑤	I	II							
$t+21$																						⑦	I	II	⑥	I	II	⑤	I	II						
$t+22$																							⑦	I	II	⑥	I	II	⑤	I	II					
$t+23$																								⑦	I	II	⑥	I	II	⑤	I	II				
$t+24$																									⑦	I	II	⑥	I	II	⑤	I	II			
$t+25$																										⑦	I	II	⑥	I	II	⑤	I	II		
$t+26$																											⑦	I	II	⑥	I	II	⑤	I	II	
$t+27$																												⑦	I	II	⑥	I	II	⑤	I	II

表 2 3~5~2(3) 三维平衡三水平轮换模式的轮换过程

轮换样本 $\alpha$	1					2					3					4					5																		
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2																	
第 $t$ 月	⑤	I	II	④	I	II							③	I	II	②	I	II	①	I	II																		
$t+1$		⑤	I	II	④	I	II							③	I	II	②	I	II	①	I	II																	
$t+2$			⑤	I	II	④	I	II							③	I	I	II	②	I	II	①	I	II															
$t+3$				⑤	I	II	④	I	II							③	I	II	②	I	II	①	I	II															
$t+4$					⑤	I	II	④	I	II							③	I	II	②	I	II	①	I	II														
$t+5$						⑤	I	II	④	I	II					II		③	I	II	②	I	II	①	I	II													
$t+6$							⑤	I	II	④	I	II				I	II		③	I	II	②	I	II	①	I	II												
$t+7$								⑤	I	II	④	I	II			④	I	II		③	I	II	②	I	II	①	I	II											
$t+8$									⑤	I	II	④	I	II		II	④	I	II		③	I	II	②	I	II	①	I	II										
$t+9$										⑤	I	II	④	I	II	I	II				③	I	II	②	I	II	①	I	II										
$t+10$											⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II									
$t+11$												⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II								
$t+12$													⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II							
$t+13$														⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II						
$t+14$															⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II					
$t+15$																⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II				
$t+16$																	⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II			
$t+17$																		⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II		
$t+18$																			⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II	
$t+19$																				⑤	I	II	④	I	II	II					③	I	II	②	I	II	①	I	II

因此 4 ~ 7 ~ 3 (3) 和 3 ~ 5 ~ 2 (3) 轮换模式都属于三维平衡三水平轮换模式<sup>[7]</sup>。

#### 4. 工业企业指标的估计。

第一,总体指标的估计。上文提出的 4 ~ 7 ~ 3 (3) 和 3 ~ 5 ~ 2 (3) 属于三维平衡三水平轮换模式  $r_{11} \sim r_{21} \sim r_{12} (3)$ ,  $x_{ij}$  是轮换样本在第  $t$  月接受第  $i$  次调查时第  $j$  个回忆水平下的估计量,调查次数  $i = 1, 2, \dots, G, G = r_{11} + r_{12}$ , 回忆水平  $j = 0, 1, 2$ 。参考 Y. S. Park et al. (2007) 的研究<sup>[8]</sup>, 总体的广义组合估计量的公式可表示成:

$$y_t = \sum_{i=1}^G \sum_{j=0}^{l-1} a_{ij} x_{ij} - \omega \sum_{i=1}^G \sum_{j=0}^{l-1} b_{ij} x_{t-1, i, j} + \omega y_{t-1} = a' x_t - \omega b' x_{t-1} + \omega y_{t-1} \quad (1)$$

其中:  $0 \leq \omega < 1$ ,  $a, b$  表示未知的系数向量,  $a = (a_{1,0}, \dots, a_{G,0}, \dots, a_{G,2})'$ ,  $b = (b_{1,0}, \dots, b_{G,0}, \dots, b_{G,2})'$ , 且  $1'a = 1'b = 1$ ,  $x_t = (x_{t10}, \dots, x_{tG0}, \dots, x_{t12}, \dots, x_{tG2})'$ 。通过最小化广义组合估计量的均方误差, 可得出最优系数。

第二,方差的估计。广义组合估计量  $y_t$  的方差公式为:

$$(1 - \omega^2) V(y_t) = a' Q_1 a + \omega^2 b' Q_1 b - 2\omega b' Q_2 a \quad (2)$$

其中:  $Q_1 = V_0 + 2\omega B_{1,0}$ ,  $Q_2 = B_{1,0} + B'_{1,1} B_{1,0} = \sum_{k=0}^{\infty} \omega^k V_{1+k} B_{1,1} = \sum_{k=0}^{\infty} \omega^{k+1} V_k V_{t_1} = Cov(x_t, x_{t+t_1})$ ,  $t_1 \geq 0$ 。

第三,月度变化和年度变化的估计量。为了反映工业指标月度和年度的变化特征,如计算同比、环比等政府部门需要的指标,构建月度和年度变化的广义估计量:

$$\Delta = y_t - y_{t-t_0}, t_0 = 1, \dots, t-1 \quad (3)$$

当  $t_0 = 1$ , 式(3) 可估计指标月度环比的变化; 当  $t_0 = 12$ , 则可估计指标年度同比的变化。

### (二) 个体工业户连续性抽样设计

#### 1. 抽样框整理。

个体工业经营单位(个体工业户)一阶段整群抽样的抽样框包括全省范围内所有的村或居委会,既包括有个体工业户的村,也包括没有个体工业户的村。抽样框的内容包括:村名称、地址代码、村内个体工业户数据、营业收入、总资产等信息。抽样框包括没有个体工业户的村,是考虑到建立抽样框时没有个体工业户,调查时可能会有,只有包括了没有个体工业户的村,抽样框才是完整的。一般是将没

有个体工业户的村单独作为一层处理。

以省为总体,对个体工业户采用一阶段整群抽样。首先,对个体户数量多的村(居委会)进行划分,划分为几个虚拟村,划分标准根据各市的实际情况及调查员所能承担的工作量确定<sup>[9]</sup>。总体由虚拟村和没划分虚拟村的行政村(居委会)构成,每个村(包括虚拟村和没划分虚拟村的行政村)作为一个独立的初级抽样单元。其次,对抽中的初级抽样单元包含的所有个体工业户进行调查。采用一阶段整群抽样设计主要是针对村或居委会作为抽样单元变化相对比较稳定,而村或居委会中的个体工业户的变化是经常发生的这一特点。在连续性调查期间,各个村或居委会会发生个体工业户数量的变化,这需要分两种情况对待:①划分了虚拟村的情况。当个体工业户增加时,要把增加的工业户分解到各个虚拟村;个体户减少时则不需要处理。②没划分虚拟村的情况,无论增减都不需要处理。

#### 2. 抽样实施步骤。

我国个体工业户抽样调查适合以各省(自治区、直辖市)为总体,各地市组织实施一阶整群抽样抽取调查样本。实施步骤如下:

第一步,对各省(自治区、直辖市)内包含的初级抽样单元进行分层。先按初级抽样单元的所在地市进行分层——地域层;各地市根据每个初级抽样单元的营业收入(个体户销售或营业收入之和)作为辅助信息进行分层——规模层,分层的原则是层内差异小,层间差异大。为了提高估计量的精度,最终层的抽样单元最少要达到 4 个。

第二步,在规模层中抽取初级抽样单元进入轮换组。样本的抽取使用永久随机数法(PRN),在规模层中,某些营业收入大的初级抽样单元作为必选的抽样单元;进行样本分配时,给营业收入大的初级抽样单元更高的抽样比,各层之间的抽样比差距不要太大,避免加大方差。

第三步,进行样本轮换。各地市按初级抽样单元的营业收入的高低划分成若干层,可看作若干轮换组,每个轮换组也可看作一个子名录库。以初级抽样单元来进行轮换,是因为初级抽样单元相对比较固定。在每月的调查中,从每个轮换组中抽取初级抽样单元,组成最终的轮换样本进行调查,并根据当地的实际情况选择不同的轮换模式进行后期调查的轮换。

3. 个体工业户轮换模式的设计。

个体工业户的轮换模式与工业企业不同,本文采用二维平衡单水平轮换模式,原因如下:

第一,个体工业户的核算能力较弱。工业企业能独立进行会计核算,可以通过会计核算资料和相关记录查询前两期的信息,而个体工业户会计核算制度不健全,有些甚至没有书面记录,若调查时仅仅依靠回忆得到前几期的有关调查项目的信息,得到的调查资料容易出错且不可靠,根据回忆估计的结果会加大误差。因此不宜采用多水平轮换模式。

第二,二维平衡单水平模式的优势。在单水平轮换模式下,不仅能够得到每月总值以及年总值的估计量,而且能反映月度间及年度间估计量的变化情况。可以根据前期资料做回归估计,提高估计的精确度。二维平衡模式能降低轮换组偏差,并能测量出轮换样本间的两类相关,即同一轮换样本在不同调查时期间的相关和同一轮换组内不同的轮换样本间的相关。

第三,二维平衡单水平模式适用于个体工业户调查。我国目前个体工业户的调查采用每季度调查一次。本文设计的两种二维平衡单水平的轮换模式  $3^3-3^2$  和  $6^2-12^1$  采取每月调查,  $3^3-3^2$  模式连续调查 3 个月,接着退出 3 个月,再接着调查 3 个月,再退出 3 个月,最后再调查 3 个月后永久退出。  $6^2-12^1$  模式先连续调查 6 个月,接着连续 12 个月退出调查,再连续调查 6 个月后永久退出。这两种模式实行每轮调查 3 个月或者 6 个月,适合我国工业个体户的实际情况和政府统计的习惯,且在调查期间,能在两个年度对相同月份数据进行比较。

对于个体工业户的抽样调查,本文提出了  $3^3-3^2$  和  $6^2-12^1$  两种二维平衡单水平轮换模式。  $3^3-3^2$  模式调查 3 个季度,退出 2 个季度,企业从接受第一次调查到完成最后一次调查然后永久退出只要 15 个月,轮换速度快,轮换周期较短,适合个体工业户变化快的地区。相邻两个月之间有 6 个样本重叠,样本重叠率达到了 66.7%;连续两年同一月份有 3 个样本重叠,样本重叠率达到了 33.3%。  $6^2-12^1$  模式是调查 4 个季度,退出 4 个季度,企业从开始调查到完成退出需要 24 个月即两年时间,轮换速度相对较慢,适合个体工业户变化较慢的地区。相邻两个月之间有 10 个样本重叠,样本重叠率达到了 83.3%。如果调查是为了反映月度、年度间的变

化,样本重叠率低更好<sup>[10]</sup>。

表 3  $3^3-3^2$  二维平衡单水平轮换模式的轮换过程

轮换样本 $\alpha$	1									2												
轮换组 $g$	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9				
第 $t$ 月	⑨	⑧	⑦				⑥	⑤	④				③	②	①							
$t+1$		⑨	⑧	⑦			⑥	⑤	④				③	②	①							
$t+2$			⑨	⑧	⑦		⑥	⑤	④		⑤	④			③	②	①					
$t+3$				⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④			③	②	①			
$t+4$					⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④			③	②			
$t+5$						⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④			③			
$t+6$							⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④					
$t+7$								⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④				
$t+8$									⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④			
$t+9$										⑨	⑧	⑦		⑥	⑤	④		⑥	⑤	④		
$t+10$											⑨	⑧	⑦		⑥	⑤		⑥	⑤			
$t+11$												⑨	⑧	⑦		⑥		⑥				
$t+12$													⑨	⑧	⑦							
$t+13$														⑨	⑧	⑦						
$t+14$															⑨	⑧	⑦					
$t+15$																⑨	⑧	⑦				

$3^3-3^2$  模式的实施步骤如下:各地市按初级抽样单元营业收入的高低划分成 9 个轮换组,每个轮换组抽出至少 1 个初级抽样单元组成最终的轮换样本。每个轮换样本都要接受 3 轮的调查。第 1 轮,轮换样本连续接受 3 个月的调查,接着退出 3 个月;然后进入第 2 轮调查,再进行 3 个月的调查,之后退出;最后进入第 3 轮的调查,再调查 3 个月最后永久退出。具体轮换过程见表 3。从横向看,每一期的调查样本都包括 9 个轮换样本,且在每期调查中,都有一个轮换样本接受第 1 次调查,另一个接受第 2 次调查,依此类推,最后一个样本接受第 9 次调查,达到横向平衡。从纵向看,任何一个轮换组,都存在某期接受第 1 次调查,另外一期接受第 2 次调查,依此类推,最后一期接受最后一次即第 9 次调查,从而达到了纵向上的平衡<sup>[11]</sup>。

$6^2-12^1$  模式的实施步骤与  $3^3-3^2$  相似,各地市按初级抽样单元营业收入的高低划分成 12 个轮换组,每个轮换组抽出至少一个初级抽样单元组成最终的轮换样本。每个轮换样本连续接受 6 个月的调查,接着退出 12 个月;再进行 6 个月的调查,最后永久退出。具体轮换过程见表 4。从横向看,每一期的调查样本都包括 12 个轮换样本,且在每期调查中,都有一个轮换样本接受第 1 次调查,另一个接受第 2 次调查,依此类推,最后一个样本接受第 12 次

表 4 表 6<sup>2</sup> ~ 12<sup>1</sup> 二维平衡单水平轮换模式的轮换过程

轮换样本 $\alpha$	1												2																						
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12											
第 $t$ 月	⑫	⑪	⑩	⑨	⑧	⑦													⑥	⑤	④	③	②	①											
$t+1$		⑫	⑪	⑩	⑨	⑧	⑦												⑥	⑤	④	③	②	①											
$t+2$			⑫	⑪	⑩	⑨	⑧	⑦											⑥	⑤	④	③	②	①											
$t+3$				⑫	⑪	⑩	⑨	⑧	⑦										⑥	⑤	④	③	②	①											
$t+4$					⑫	⑪	⑩	⑨	⑧	⑦									⑥	⑤	④	③	②	①											
$t+5$						⑫	⑪	⑩	⑨	⑧	⑦								⑥	⑤	④	③	②	①											
$t+6$							⑫	⑪	⑩	⑨	⑧	⑦							⑥	⑤	④	③	②	①											
$t+7$								⑫	⑪	⑩	⑨	⑧	⑦						⑥	⑤	④	③	②	①											
$t+8$									⑫	⑪	⑩	⑨	⑧	⑦					⑥	⑤	④	③	②	①											
$t+9$										⑫	⑪	⑩	⑨	⑧	⑦				⑥	⑤	④	③	②	①											
$t+10$											⑫	⑪	⑩	⑨	⑧	⑦			⑥	⑤	④	③	②	①											
$t+11$												⑫	⑪	⑩	⑨	⑧	⑦		⑥	⑤	④	③	②	①											
$t+12$													⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①											
$t+13$														⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①										
$t+14$															⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①									
$t+15$																⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①								
$t+16$																	⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①							
$t+17$																		⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①						
$t+18$																			⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①					
$t+19$																				⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①				
$t+20$																					⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①			
$t+21$																						⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①		
$t+22$																							⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①	
$t+23$																								⑫	⑪	⑩	⑨	⑧	⑦	⑥	⑤	④	③	②	①

调查,达到横向平衡。从纵向看,任何一个轮换组,都存在某期接受第 1 次调查,另外一期接受第 2 次调查,依此类推,最后一期接受最后第 12 次调查,从而达到了纵向上的平衡<sup>[11]</sup>。

4. 指标的估计。

第一,总体指标的估计。上文提出的 6<sup>2</sup> ~ 12<sup>1</sup> 和 3<sup>3</sup> ~ 3<sup>2</sup> 都属于二维平衡单水平轮换模式  $r_1^m \sim r_2^{m-1}$  的形式,  $x_{ti}$  是轮换样本在第  $t$  月接受第  $i$  次调查时得到的估计量,调查次数  $i = 1, 2, \dots, mr_1$ 。参考 Y. S. Park et al. (2001)<sup>[12]</sup> 的研究,总体的广义组合估计量的公式可表示为:

$$y_t = \sum_{i=1}^{mr_1} a_i x_{t,i} - \omega \sum_{i=1}^{mr_1} b_i x_{t-1,i} + \omega y_{t-1} = a'x_t - \omega b'x_{t-1} + \omega y_{t-1} \quad (4)$$

其中:  $x_t = (x_{t,1}, \dots, x_{t,2r_1}, \dots, x_{t,mr_1})'$ ,  $a, b$  表示未知的系数向量  $a = (a_1, \dots, a_{2r_1}, \dots, a_{mr_1})'$ ,  $b' = (b_1, b_2, \dots, b_{mr_1})$ , 且向量满足约束条件:  $1'a = 1'b = 1, 0 \leq \omega < 1$ 。通过最小化广义组合估计量的均方误差,可得出最优系数。

第二,方差的估计。计算二维平衡单水平的方差,要先知道二维平衡的协方差。为了简化方差公式,先定义三个矩阵。定义第一个矩阵  $(mr_1 + (m-1)r_2) \times (mr_1 + (m-1)r_2)$  维  $L_s$ , 该矩阵元素满足:

$$(L_s)_{ij} = \begin{cases} 1, & \text{当 } j - i = 1 \\ 0, & \text{其他} \end{cases} \quad (5)$$

此外,还有如下性质  $L_s^{t'} = L_s^{t'-1} \cdot L_s, t' \geq 1$ , 且  $L_s^0 = I$ 。

再定义第二个矩阵  $mr_1 \times mr_1$  维  $L$ , 该矩阵元素满足如下关系:

$$(L)_{ij} = \begin{cases} 1 & \text{当 } i = r_1, 2r_1, \dots, pr_1 \text{ 时 } j = (\text{mod}_m\{m - m^* + i/r_1\}r_1 + 1) \\ 1 & \text{当 } i \neq r_1, 2r_1, \dots, pr_1 \text{ 时 } j = i + 1 \\ 0 & \text{其他} \end{cases} \quad (6)$$

$L^{t'} = L^{t'-1} \cdot L$ , 这里  $t' \geq 1$ , 且  $L^0 = I$ 。

最后,定义第三个矩阵  $mr_1 \times mr_1$  维矩阵  $L^{t'}_1$ , 这个矩阵是通过移除矩阵  $L^{t'}_s$  的第  $i$  行和第  $i$  列得到

的  $j = kr_1 + (k - 1)r_2 + 1, kr_1 + (k - 1)r_2 + 2, \dots, k(r_1 + r_2)$  其中  $k = 1, 2, \dots, m - 1$ 。

根据定义的矩阵  $L_1$  和  $L$  得到协方差:

$$\text{Cov}(x_t, x_{t+t'}) = \rho_{1t'} V L_1' V + \rho_{2t'} V L_2' V \quad (7)$$

其中:  $t' \geq 0, V = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{mr_1}), L_1' = L_1' - L_1',$  且  $L_1^0 = I, L_2^0 = 0$ 。

可以得到估计量的方差公式:

$$(1 - \omega^2) V(y_t) = a' V(I + 2\omega Q_{1, \rho}) Va + \omega^2 b' V(I + 2\omega Q_{1, \rho}) Vb - 2\omega b' V(Q_{1, \rho} + Q_{1, \rho}') Va \quad (8)$$

其中:  $Q_{1, \rho} = \sum_{j=0}^{T-t_0+k} \omega^{j+k} \rho_{1, t_0+j-k} L_1^{t_0+j-k} + \sum_{j=0}^{\infty} \omega^{j+k} \rho_{2, t_0+j-k} L_2^{t_0+j-k},$   $t_0 \geq k, T \equiv mr_1 + (m - 1)r_2 - 1$  且当  $T - t_0 + k < 0$  时, 第一项为 0。

第三, 月度变化和年度变化的估计量。为了反映指标月度和年度的变化特征, 如计算同比、环比等指标, 构建月度和年度变化的广义估计量:

$$\Delta = y_t - y_{t-t_0}, t_0 = 1, \dots, t - 1 \quad (9)$$

当  $t_0 = 1$ , 式(9)可估计指标月度环比的变化; 当  $t_0 = 12$ , 则可估计指标年度同比的变化。

### (三) 其他估计方法

除了广义组合估计量外, 美国在现时人口调查中用了广义方差函数 (Generalized Variance Function) 估计方法来估计人口总量<sup>[13]</sup>。复杂样本方差的估计方法还有很多, 下面主要介绍刀切法 (Jackknife Method)<sup>[14]</sup>, 刀切法是通过子样本估计量间的变异来估计全样本估计量的方差, 使用刀切法的关键是随机组的划分。

令  $\hat{y}$  表示由全样本得到的指标  $y$  的估计量。把全部样本分成  $k$  组, 令  $\hat{y}_1$  表示去掉第一个随机组后, 利用剩下的  $k - 1$  组的样本构造得到的  $y$  的估计量。令  $\hat{y}_2$  表示去掉第二个随机组后, 利用剩下的  $k - 1$  组的样本构造得到的  $y$  的估计量, 依此类推, 依次去掉  $k$  个随机组, 得到  $k$  个  $y$  的估计量。令  $\hat{y}_t$  是去掉第  $t$  组样本后, 用与  $\hat{y}$  相同的方式构造的  $y$  的估计量, 记

$$\hat{y}_T = k\hat{y} - (k - 1)\hat{y}_t \quad (10)$$

其中  $T = 1, 2, \dots, k$  则  $y$  的估计量为:

$$\hat{Y} = \frac{1}{k} \sum_k \hat{y}_T \quad (11)$$

$\hat{Y}$  的方差估计为:

$$v(\hat{Y}) = \frac{1}{k(k - 1)} \sum_{T=1}^k (\hat{y}_T - \hat{Y})^2 \quad (12)$$

得到  $\hat{y}$  的方差估计为:

$$v(\hat{y}) = \frac{1}{k(k - 1)} \sum_{T=1}^k (\hat{y}_T - \hat{y})^2 \quad (13)$$

## 四、结束语

本文分析了我国规模以下工业抽样调查的现状及存在的问题, 从抽样调查方案的设计, 到抽样框的更新维护和估计量的估计方法, 提出了改革的建议, 对工业企业和个体工业户分别设计了连续性抽样调查方案, 对工业企业采用三维平衡三水平的轮换模式, 对个体工业户采用二维平衡单水平的轮换模式, 每种调查方案都设计了两种轮换模式, 以适应各个地区的实际情况。

本文对我国规模以下工业抽样调查设计的方案属于方法体系的改革内容, 在实践方面, 还需要相关制度改革、组织机构、人员经费等的跟进和配套。从理论研究逐步转向应用实践, 从一次性的调查向科学的连续性抽样调查转变, 形成科学的连续性抽样调查体系, 可以产生更加可靠准确的时间序列数据和面板数据, 为科学的分析和预测决策服务。

### 参考文献

- [1] 中华人民共和国国家统计局. 工业统计报表制度 [EB/OL]. [http://www.stats.gov.cn/tjzd/gjtjzd/t20110506\\_402724057.htm](http://www.stats.gov.cn/tjzd/gjtjzd/t20110506_402724057.htm). 2011-05-06.
- [2] 陈光慧, 刘建平. 我国农村住户连续性抽样调查方案设计 [J]. 统计研究, 2013, 30(2)
- [3] 田秀华. 企业抽样调查实践 [M]. 北京: 中国统计出版社, 2005: 20-21, 31-60.
- [4] 中华人民共和国国家统计局. 奋进中的中国统计: 走向开放透明的中国统计 [EB/OL]. [http://www.stats.gov.cn/tjdt/tjr/d2jtjr/gfjy/t20100920\\_402752846.htm](http://www.stats.gov.cn/tjdt/tjr/d2jtjr/gfjy/t20100920_402752846.htm). 2010-09-20.
- [5] 陈光慧, 刘建平. 我国经常性抽样调查体系改革研究 [J]. 统计研究, 2010, 27(10): 3-8.
- [6] 中国统计信息网. 国家统计局有关负责人就提高工业和投资统计起点标准答记者问 [EB/OL]. [http://www.stats.gov.cn/tjdt/gjjjdt/t20110308\\_402709132.htm](http://www.stats.gov.cn/tjdt/gjjjdt/t20110308_402709132.htm). 2011-03-08.
- [7] 陈光慧, 刘建平. 三维平衡多水平轮换设计及其连续性估计方法研究 [J]. 统计研究, 2009, 26(11): 2-6.
- [8] Y. S. Park, J. W. Choi, K. W. Kim. A balanced multi-level rotation sampling design and its efficient composite estimators [J]. Journal of Statistical Planning and Inference, 2007, 137: 594-610.
- [9] 王文颖. 企业抽样调查方法及其应用 [M]. 北京: 中国统计出版社, 2007: 95-109.
- [10] Y. S. Park, K. W. Kim, B. Choi. Multi-level composite estimators and the evaluation of the three-way balanced multi-level rotation designs. Journal of the Korean Statistical Society, 2011, 40: 90-108.

# 核心通货膨胀扣除食品的可行性检验<sup>\*</sup>

谭本艳 黄双超

**内容提要:** 对于广泛使用的通过扣除食品和能源来度量核心通货膨胀的方法,中国学者对其在中国的适用性提出了质疑。本文在分析核心通货膨胀的含义及其度量方法的基础上,根据 2001 年 1 月至 2013 年 5 月中国 CPI 消费价格分类指数数据,应用 Vahid 和 Engle(1993)提出的基于共同趋势和共同周期模型的时间序列长期成分和短期成分分解技术,对核心通货膨胀扣除食品的可行性进行了检验。检验结果表明,食品价格既是中国整体 CPI 的短期主导力,又是长期主导力。因此,通过扣除食品度量中国的核心通货膨胀并不符合核心通货膨胀的本质内涵。进一步,本文提出了关注 CPI 结构的建议。

**关键词:** 核心通货膨胀; 食品; 共同趋势; 共同周期; 长短期主导力

**中图分类号:** C812      **文献标识码:** A      **文章编号:** 1002-4565(2014)01-0042-07

## Feasibility Test of Core Inflation Excluding Food

Tan Benyan & Huang Shuangchao

**Abstract:** Some Chinese scholars oppugn the scientificness for applying the widely used measure of core inflation by excluding food and energy to measure China's core inflation. Based on analyzing the definitions and measures of core inflation, this paper apply the technique to test permanent and transitory domain force in time series based on common trend and common cycle model proposed by Vahid and Engle (1993), making analysis on the feasibility for measure of core inflation by excluding food according to the monthly data of CPI by Category from 2001.1 to 2013.5 in China. The study turns out that the food price is not only the permanent domain force of CPI, but also the transitory domain force, so the measure of core inflation by excluding food don't accord with the connotation of core inflation.

**Key words:** Core Inflation; Food; Common Trend; Common Cycle; Permanent and Transitory Domain Force

### 一、引言

众所周知,《中国人民银行法》规定的中国货币

政策目标是“保持货币币值稳定,并以此促进经济增长”其中“保持货币币值稳定”的直观表述就是保持通货膨胀水平的稳定。那么,如何科学地看待

<sup>\*</sup> 本文获国家自然科学基金项目(11CJY009)、教育部人文社会科学项目(10YJC790230)、三峡大学人才科研启动基金项目(KJ2011B045)的资助。

① 国家统计局原总经济师姚景源 2007 年 9 月 11 日接受中央电视台经济半小时栏目采访时的谈话。

[11]陈光慧,刘建平. 二维平衡单水平轮换设计及其广义组合估计方法研究[J]. 数理统计与管理, 2011, 30(4): 696-704.

[12]Y. S. Park, K. W. Kim, J. W. Choi. One-level rotation design balanced on time in monthly sample and in rotation group [J]. Journal of the American Statistical Association, 2001, 96: 1483-1496.

[13]U. S. Department of Labor, U. S. Department of Commerce. The Current Population Survey: Design and Methodology, Technical Paper 66 [M]. Washington: Government Printing Office, 2006.

[14]P. Duchesne. A Note on Jackknife Variance Estimation for the

General Regression Estimator [J]. Journal of Official Statistics, 2000, 16(2): 133-138.

#### 作者简介

刘建平 男, 62 岁, 山西怀仁人, 暨南大学经济学院教授, 博士生导师。研究方向为统计调查与计量分析。

李绍泰 男, 25 岁, 浙江临海人, 暨南大学经济学院统计学专业在读硕士研究生。研究方向为统计调查与数据分析。

(责任编辑: 程 晔)