

# 统计数据准确性评估： 方法分类及适用性分析<sup>\*</sup>

王 华 金勇进

**内容提要:**对于统计数据准确性评估方法的归纳分类,可以从辅助资料信息的来源及表现形态、包含目标特征真值关键信息的参照标准的构造方法、实际统计数值与参照标准之间的比较逻辑等方面加以考察。针对总量统计数据的准确性评估,主要在纵向时间维度上开展,根据统计指标与其相关联指标变动趋势的偏离程度加以评判;针对个体或分类统计数据的准确性评估则主要在横向空间维度上开展,对数据的统计分布形态加以检验,或者利用重复调查或随机实验方法对事先假定的误差参数进行估计。不同方法具有不同的适用条件,在实际应用中,应基于所掌握辅助资料的翔实程度来选择评估方法,并对导致评估方法发生误判的因素加以细致分析和排除,以确保评估结论的说服力和可信度。

**关键词:**统计数据准确性;评估方法;适用性

**中图分类号:**C811      **文献标识码:**A      **文章编号:**1002-4565(2009)01-0032-08

## Statistical Data Accuracy Assessment : Methods Classification and Applicability

Wang Hua & Jing Yongjin

**Abstract:** For generalizing and classifying the assessment methods of statistical data accuracy, some factors should be reviewed such as source and form of the assistant information, conformation of contrast criterion containing key information of objective characters' real values, and logistic for comparing factual statistical data with contrast criterion. The assessment of gross statistical data accuracy has mostly been processed along the longitudinal way, the conclusion of which can be obtained according with actual statistics' trend departure from correlate indices. Meanwhile, the assessment of unit or section statistical data accuracy has mostly been processed along the landscape orientation, for which probability distribution of data would be tested, or error parameters be estimated based on re-interview or random experiment. In addition, there've been different applicability conditions for different assessment methods. In order to insure reliability of assessment conclusions, abundance degree of assistant information should be considered, and those factors inducing miscarriage of justice be checked and removed.

**Key words:** statistical data accuracy; assessment methods; applicability

### 一、引言

统计数据质量的概念内涵已经扩展到及时性、适用性、可比性、衔接性、可解释性、可取得性、有效性等特征,但准确性仍然是其核心特征。因此,作为统计数据质量控制的前提和依据的统计数据质量评估工作,也必然要以准确性评估作为其核心内容。事实上,现在大部分讨论“统计数据质量(评估)”的研究文献实质上仍然是意指“统计数据准确性(评估)”。

随着我国经济实力与国际影响力的不断增强,经济总量及其相关统计数据的准确性或可信性问题也日益引起世人关注。鉴于统计调查过程本身涉及面广,难以重复等特性,对其产出结果——统计数据的准确性进行“准确”评估存在诸多困难,导致在研究和实践中一直未能形成评估方法的公认体系和标

<sup>\*</sup>本文为国家自然科学基金项目“我国统计数据质量的评估系统研究”(项目编号:70671104)阶段性成果。

准,针对有关的评估结论更是颇多争议。这也成为国内外社会各界对我国统计数据质量的存疑长期难以消除的关键原因之一。

本文基于对统计数据准确性概念的分析,尝试对有关统计数据准确性的各类研究中所采用的评估方法进行归纳分类,对不同方法的特性及适用条件加以述评,以期促进对统计数据准确性评估研究与实践的规范化和体系化。

## 二、分类准则

由于宏观层面的总量统计数据是将基层个体单元的统计调查数据经由逐级或直接汇总得到,对于个体统计调查数据中存在的误差,最终也会在总量统计数据中有所体现。同时,如果对基层个体单元的调查误差服从具有一定系统偏差和方差的随机概率分布,则总量统计数值相对于唯一确定的目标特征真值,也会表现为具有系统偏差和统计不确定性(以方差表征)的随机变量。可见,统计数据准确性可以分别在个体数据与总量数据两个层次、系统偏差与方差两个方面加以定义。对于抽样调查数据资料及由此推估的总量数值,同样适用上述过程,不再赘述。

对统计数据的准确性进行评估的通行做法是,首先搜集与目标特征相关联的辅助资料信息,利用这些辅助资料,或者构造目标特征真值的估计量,或者构造包含目标特征真值关键信息的函数;再以此作为参照标准,与实际统计数值或是其函数进行比较,根据二者之间的差异状况对实际统计数据的准确程度加以评估判定。而辅助资料的搜集渠道与表现形态、参照标准的构造方法、实际统计数值与参照标准之间的比较逻辑等几项要素奠定了各种统计数据准确性评估方法的基本类型。

根据以上分析,为了对各种统计数据准确性评估方法进行有效归类,可作以下几点考察:第一,考察评估实施中辅助资料信息的来源及其表现形态,如独立调查的样本数据,国民经济核算体系中的其他统计指标,关于总体中各部门或基层单元的理论统计分布的先验知识,等等。第二,考察评估实施中构造参照标准的技术假定及方法范畴,如假定独立样本数据获得了调查单元的“真值”进而构造误差分解模型,假定同一体系中的其他统计指标与待评估统计指标间存在高度的相关关系并进行相关分析或回归分析,假定总体单元服从某种理论分布从而进

行分布一致性检验并设定表征误差界限的分位数,等等。第三,考察实际统计数据与参照标准的比较逻辑,如与样本重复调查数据相比较的偏差和方差估计,与相关统计指标比较的趋势偏离分析或回归系数变动分析,基于计量模型或分布检验的异常数值识别,等等。以上述几个方面的考察作为分类准则,可望实现对评估方法在实施主体、评估目标、操作细节等方面的有效区分。

区分宏观层面的总量统计数据与中微观层次的数据或个体统计数据两类资料,可以发现,前者仅存在时间序列,对其准确性的评估须在纵向时间维度展开;而后者则在某一时点存在同类部门或个体的数据,对其准确性的评估可在横向空间维度展开。本文即以此作为对评估方法进行分类的初始框架,对各类方法的一般技术特征和操作步骤、应用实践等事项加以评述。

## 三、针对总量数据的评估方法

针对总量统计数据准确性的评估研究和实践,既有统计数据的生产者,也有统计数据的使用者(社会经济领域的研究人员)。前者的评估目标是在统计数据的中间汇总环节,及时发现和纠正统计数据中可能存在的偏误,最大限度地保证最终公布统计数据的可靠性;后者则是为了考察统计数据的可用性,进而保障数量分析的可靠性。评估方法大致可分为逻辑规则检验、经验参数比对、相关指标比对、基于模型的异常数值识别与参数稳定性分析等,在此作逐一评析。

### (一)逻辑规则检验

在我国宏观统计体系当中,遵循特定统计核算规则产生的各项、各级同度量统计指标之间,在统计口径和范围方面往往存在一定的单向包含或相互平衡的内在逻辑关系,如期末存量应为期初存量与期间流量之和,就业人口应小于地区总人口,国民收入应小于社会总产值,GDP增长率应为三大产业增长率的加权平均值,三种核算方法下的GDP结果应相

---

按照统计调查领域对非抽样误差(尤其是调查计量误差)研究的技术假定,可将对基层个体单元的实际统计调查结果视为随机变量的一次实现。

总量与分类统计数据的区分标准是相对的:对于以部门或地区确立的分类统计数据,仅就某一部门或地区而言,即为总量统计数据;而在国际比较中,总量统计数据也只是以国家确立的分类统计数据中的一个单元而已。

等,等等。可以利用这些逻辑关系来衡量统计指标数值是否存在与客观规律不一致的现象,进而识别指标数据是否失真。

逻辑规则检验方法是以遵循逻辑关系进行对比的同度量统计指标彼此互为参照标准,为对方设立统计核算规则所允许的动态的取值范围。如果进行对比的统计指标间违背了特定的逻辑关系,表明其中必定有某个(或多个)统计指标的数值超出了正常的取值范围。在统计数据生成过程的各层环节中,逻辑规则检验是各级统计部门进行统计质量检查与控制的主要手段之一,这也是国际上通用的可解释性原则的应用(刘延年,2002)。

很明显,这一评估方法的结论是多重指向的——如果国内生产总值小于各地区生产总值之和,既可能国家级核算数值存在低估,也可能是地区级核算数值高估,或者是各方数值都存在误差。另一方面,由此揭示的信息也不能都归因于现行统计的质量(准确性)问题,因为统计制度方面的缺陷——如当统计核算资料来源因存在缺口和口径等方面的问题而需进行必要的估算和调整时(许宪春,1999)——同样可能会导致总量统计数据中出现逻辑不一致的现象。

作为对该方法的拓展,经济研究学者利用统计核算体系中的各项基础构成指标,遵循统计核算规则来重构总量统计指标,以此作为评判官方公布统计数据准确性的参照标准。这一技术思路在对经济总量(GDP)及其增长速度的准确性评估中尤为常见,如以实物产量增长速度的加权平均方法计算综合增长率的物量指数法(Wu, 2000; Madisson, 1998),利用各类价格指数对名义GDP进行缩减调整的价格指数法(Ren, 1997)等。这一方法的运用须以基础构成指标(如实物产量和价格指数)的准确无误作为前提,而由于统计指标的重构方法与实际统计核算方法不尽一致,评估结论只能提供一种方向性的启示(即统计指标被高估或低估),却无法给出确切的误差水平(即被高估或低估的程度)。

## (二) 经验参数比对

在一定客观条件下,反映社会经济现象的若干指标之间存在较为稳定的关联性或协同变动的规律性,通过对这些指标间联系的长期分析,可以得到指标间基本固定的经验数据标准。利用这些经验参数作为统计数据准确性评估的参照标准,如果统计指

标间的比率或比例关系值出现与经验参数的较大偏离,则可初步判定待评估统计指标的准确性存在问题。

在国民经济核算领域,诸如增加值率(增加值占总产值比重)、内部结构比重(规模以下工业产值或增加值占地区工业总产值或增加值比重)、三次产业占GDP的结构比重、财政收入占GDP比重、本级GDP占上一级GDP的比重、人均GDP(或人均消费水平)及在全国或全省的位次等,这些相对指标在一般情况下是相对稳定的(尹彦君,2004),可作为评估GDP核算数据准确性的经验参数标准。在人口统计领域,年龄偏好指数、惠普尔指数、迈耶尔指数、联合国年龄-性别指数、常规生育率等经验参数,也被普遍用于对人口统计或普查数据准确性的评估(张为民和崔红艳,2003;王广州,2004)。

通过与经验参数的比对分析,可以在一定程度上判断统计数据的准确性,但此方法的应用存在着许多理论和方法上的缺陷:一方面,用于评估的比率通常会涉及两个以上的统计指标,即使与经验参数没有显著差异,也不能排除构造比率的所有指标都存在失实的可能性;另一方面,经验参数毕竟是“经验之谈”,其适用范围仍有待验证,如果客观条件发生了变化,却仍采用惯常的参数水平作为参照标准,评估结论就难免出现误判。

## (三) 相关指标变动趋势比对

作为对经验参数比对方法在操作形式上的一种变通,相关指标变动趋势比对法要求寻找与待评估指标保持较高相关性的、已知正确的其他统计指标,作为评估参照标准,根据其变动趋势来检验待评估指标(及其变动趋势)的准确性。可见,两类评估方法的逻辑思路基本相同,但后者的操作形式更为灵活、适用范围更为广泛。

孟连和王小鲁(2000)选取价格指数、工业产品产量增长率、货物运输/周转总量和工业消费量/工业消费综合能源量增长率等理论上与经济总量(或工业增加值)增长速度应保持高度正相关的指标,发现1991-1997年间GDP和工业增加值增长速度与上述指标的变化趋势出现较大偏差,因而判定经济增长数据存在统计误差。Rawski(2001)则以

换言之,该方法同样存在结论的多重指向性,这也为人为干预统计数据和虚报行为提供了回旋余地。

1997 - 2000 年间中国经济高速增长的同时却伴随能源消耗相对下降、通货紧缩、就业人口增长缓慢等现象为由,质疑经济增长数据的可靠性。该文献结论旋即在国内外社会引起强烈反响,支持者甚众,但也不乏反对的声音。

Lardy(2002)以1997 - 2001年间中国财政税收与进口高达90%的增长速度作为证据,反对Rawski的“低增长猜想”。任若恩(2002)对Rawski(2001)及孟连和王小鲁(2000)在研究中所采用方法的合理性进行了驳斥,认为关于经济增长率与能源增长率应该大致相等的假定是不成立的,工业增加值与总产值(或工业产量)的增长率未必同步,与货运和能源消费增长间的相关关系也很难稳定。

可见,利用相关指标比对的结论来分析统计数据的准确程度,首先需要解决相关指标间的关联关系是否确实存在、是否长期稳定等问题,否则评估结论是难以令人信服的。不仅如此,相关参照指标的数据是否可靠,也是不容回避的一大问题。张新和蒋殿春(2002)即指出,中国官方公布的就业数字无法反映真实情况(如隐性失业),不足以作为评估GDP增长可信度的参照依据。Rawski(2002)在其答辩性的文章中也指出,Lardy(2002)的分析参照变量财政收入和进口都存在高估的倾向。

#### (四) 基于模型的异常数值识别与参数稳定性分析

在经验参数比对和相关指标变动趋势比对方法的运用中,难以保证待评估统计指标与辅助参照指标间存在确切而稳定的相关关系,作为对两类方法技术缺陷的一种弥补和完善,由此衍生出基于经济计量模型或时间序列模型的评估方法。此类方法的评估结论依赖于两种分析途径:一种是分析被解释变量实际统计值与模型拟合值之间的差异情况,从中识别出偏差较为显著的异常数值点;另一种是分析模型参数估计值的时期稳定性,判断模型所反映的经济运行机理是否明显有悖于社会经济常识。

异常数值是在数据集中明显远离其他大多数个体的数据点,因此有可能产生于完全不同的数据生成机制。基于模型的异常数值识别方法,其最初应用于统计诊断领域,目的在于发现对模型估计的可靠性产生不良影响的数值点(周建,2005)。而在评估统计数据准确性方面,该方法的基本操作思路是首先依据一定的经济理论,拟合以待评估指标作被解释变量的经济计量模型或时间序列模型;利用通

过各种检验的、拟合优度较高的模型来计算被解释变量在评估期的预测值,以此作为代表目标特征真值的参照标准;再分析计算实际统计值与预测值之间的差异程度,如果可从中识别出差异较为显著的异常数值点,则表明待评估指标存在数据准确性问题。在具体应用中,该方法又被称为趋势预测法或回归预测法(伍荣坤,1993)、事后预测及反常结果判断法(刘孝新等,1997)、趋势模拟评估法(刘洪和黄燕,2007)等。

在社会经济发展状况较为平稳的宏观背景下,用于反映经济运行机理的模型参数或由模型参数构造的函数也应该是比较稳定的。如果模型参数的估计值(或其函数)出现激增或异常跳动,而又无法找到适当的客观原因(如大规模的技术进步和体制变迁)加以解释,则可认定是被解释变量(待评估统计指标)存在准确性问题。该方法的应用最常见于以全要素生产率来评估经济增长速度的准确性,具体可在生产函数模型中引入用于划分时期的虚拟变量(孟连和王小鲁,2000),或者利用索洛余值法、数据包络分析法等来估算全要素生产率(杨冠琼,2006)。

虽然对经验参数比对与相关指标比对方法作了进一步深化,评估过程和标准更为科学严谨,但对于前述两种方法的限制因素,在基于模型的评估方法中同样存在:即要求用于拟合模型的历史统计数据(包括待评估指标本身)以及评估期的解释变量数据必须是准确可靠的,从而不会发生数据“污染”现象(周建,2005)。即便如此,对于被解释变量的异常数值,或者模型参数的异常变动,仍需结合异常数值产生的背景进行谨慎判断,以明确异常性是由客观原因(如外部冲击、体制变革等)造成的,还是源于统计数据的准确性问题。

## 四、针对个体或分类数据的评估方法

本部分讨论的统计数据包括统计调查基层单元的原始数据(可称之为个体数据)以及较低层级(如部门、地区)的统计汇总数值(可称之为分类数据),前者构成统计汇总数值的基础,后者在多级汇总数据结构中也表现出与前者同样的地位和特性。

个体数据和分类数据主要存在于统计调查的中间环节,成为统计部门评估数据准确性、改进统计工作制度的重要素材和依据,而统计数据的使用者(如社会经济领域研究人员)通常只能获得分类统计数据,因而对此类方法的应用较为受限。评估方法大

致可分为统计分布检验、基于重复调查的偏差估计与误差效应识别等。

### (一) 统计分布检验与基于统计分布的异常数值识别

如果掌握了统计指标在基层单元的个体数据或部门、地区的分类汇总数据,并且已知总体应服从特定的概率分布,则可以将个体(或分类)数据的具体分布形态与特定分布进行比较,若二者趋于一致,则表明统计数据是准确的;若二者不一致,则认为统计数据存在误差。特定的概率分布形态或者源于历史数据的经验估计,或者源于理论上的推断。如成邦文等(2000)证明,反映规模大小的统计量(如产量、产值、销售收入等),其来自总体的数据近似服从对数正态分布,因此可用对数正态分布的一致性检验来评估有关统计数据的准确性,进而可以设定表征误差界限的分位数(如95%分位数),用以识别个体或分类数据中远离按分布规律所对应位置的异常数值(成邦文和董丽娅,2002)。

该方法要求个体或分类统计数据服从特定概率分布,但在很多情况下,数据的先验分布是未知的,或者无法用任何一种理论统计分布来拟合,因而应用范围受到限制。另一方面,即使统计数据确实符合特定分布,也不能排除在全部或大多数个体上都存在系统性数据偏差的可能性;而对于真值靠近分布中心位置的个体,即使出现较大的统计偏差,也未必会脱离多数数值点而表现为异常数值。因此,该方法虽然被认为“本身就具有抗虚假性,不真实的统计要通过检验是很困难的,未通过检验的统计数据也不可能通过不真实的数据修改而使之通过(成邦文等,2000)”,但实际评估操作的有效性却未必如此乐观。

为了克服总体理论分布未知的局限,傅德印(2001)提出了基于探索性数据分析及判别分析来识别原始(或分类)统计数据中异常数值的方法。茎叶图法、字母值法、箱线图法、编码表、悬浮式直方图及中位数平滑法等探索性数据分析方法,能够在不破坏原始数据中其他数据的前提下突出表现极端数据或无用的数据,为评估数据准确性提供了依据。但该方法仍然只能识别出远离数据总体的异常数值点,无法更深入地探索其他数据准确性问题。

### (二) 基于事后重复调查的偏差估计

事后重复调查(re-interview)是普查和大型抽样

调查中普遍采用的质量评估和控制手段,利用对随机子样本二次调查所获得的数据与初始调查数据进行比较,既可对数据准确性进行评估和纠正,也可实现对调查实施现场人员(主要是调查员)工作情况的监督和警示。

美国人口现状调查(the Current Population Survey, CPS)在各项抽样和调查技术的运用上已经成为很多国家实施大规模抽样调查的范本。CPS将重复调查程序视为限制非抽样误差发生的最主要工具之一。在为质量控制而实施的重复调查程序中,通过对调查员及其所调查住户的两阶段抽样,每月大约有2%的初始调查样本住户会被再次抽中接受重复调查,由有经验的调查员或项目督导借助计算机辅助电话调查的方式进行提问,对初始调查结果进行核查评估(U. S. Census Bureau and Bureau of Labor Statistics, 2002)。在美国10年一次的人口普查中,也通过事后抽样调查,采用基于捕获-再捕获模型(Capture-Recapture)的事后分层双系统估计方法,对普查数据的准确性进行评估,以此估计人口普查数字偏离总体“真实”人口数的程度(Hogan, 1993; 胡桂华, 2007)。

事后重复调查评估方法的运用,将系统性偏差的出现归因于调查设计的不合理与调查实施的不规范,因此要求对重复调查实施的调查条件加以控制,在更为理想和完善的调查条件下获得基层单元的真实结果。然而统计调查“真值”是否存在——尤其是在社会调查领域,重复调查结果是否可以代表真值,成为长久以来争论的一个话题;初始调查是否会对重复调查产生关联影响,也是必须考虑的问题。

### (三) 误差效应的测量及其模型构建

统计调查误差对于统计数据准确性的影响,不仅体现在系统偏差方面,更在于会增大统计结果的不确定性、减损统计推断的精度,即便对全面调查/普查亦是如此。结合各种调查误差来源的研究表明,各类调查误差的存在(如由调查员效应引起的相关计量误差)会极大增加调查统计量的方差(Groves, 2004)。美国普查局正是基于对普查员的相关误差影响的考虑而决定在1950年以后的普查中采用自填式普查设计,从而尽可能地杜绝家庭之间的相关

作为例外,捕获-再捕获估计则要求重复调查与初始调查的调查条件应保持一致,即两次调查的样本均应来自于同一总体。

计量误差 (Hansen 和 Marks, 1958)。

为了对统计调查误差效应进行测量评估,通常是采用重复调查或随机实验的方法。与旨在估计统计调查偏差的重复调查不同,为估计调查测量方差而实施的重复调查要求调查条件尽可能地与初始调查保持一致,是针对初始调查子样本的、与初始调查程序相独立的重复实施。随机实验又称为交叉子样本方法,是将相同设计、基本同质的随机子样本分配到与待测特定误差有关的不同处理(如不同的调查员、不同的问卷设计、不同的调查方式等)中去,对不同处理下的统计调查结果进行比较,采用方差分析等方法来估计具体的误差效应值。

卢依吉·法布里斯等(2002)利用中国第一次农业普查的事后质量抽查数据,对普查数据的可靠性进行了估计,通过对与农村住户及家庭成员有关项目的基本回答方差和平均总方差的计算,发现对调查估计最具打击性的影响来自调查员误差。杨清和吴伟霞(2000)主张重点判断原始资料中是否存在有系统性误差的资料,并基于资料分组利用方差分析与多重比较的方法来筛选出这类资料。屈耀辉和曾五一(2004)利用与杨清和吴伟霞(2000)相同的技术来识别农产量抽样调查中调查员的误差效应。

对于统计调查误差效应的识别估计,需要结合统计调查项目实施的背景进行细致分析,辨别出各种可能的误差来源,并针对有关误差的统计特性设置假定,进而利用误差分解模型进行估计和分析。

该类方法是各类统计数据准确性评估方法中仅有的一类关注统计调查方差的方法,虽早在半个世纪之前已经具有雏形,但由于涉及过强的误差参数假定和复杂的模型估计技术,同时社会调查领域实施独立重复调查也存在很多困难,因而离实际应用仍有很大差距。

### 五、各类评估方法的适用性比较

综上所述,七类统计数据准确性评估方法(分别以字母 A~G 表示)所针对的统计数据形态、评估实施主体、辅助资料类型、参照标准、方法应用的技术假定、比较逻辑以及导致评估结论发生误判的因素,可归纳如表 1 所示。

由表中各项特征决定,不同的评估方法的适用条件也不尽相同,可大致概括为可行性、有效性和可靠性三个方面。其中,评估方法的可行性是指评估所需辅助资料的可获取程度与方法操作的难易程度;评估方法的有效性是指评估结论对于影响统计数据准确性的各种因素的揭示程度,以及相关信息的实践指导价值;评估方法的可靠性则是指方法技术假定的合理性以及由此导致误判的可能性。

在可行性方面,针对总量统计数据准确性的 A~D 四类评估方法,都要求寻找与待评估指标具有逻辑相关性的统计指标作为辅助资料,这一类资料往往可以在与待评估指标同时公布的统计数据中获得;同时各类方法在参照标准构造与比较逻辑的灵活和复杂程度方面存在渐次递进的关系。针对个体

表 1 各类统计数据准确性评估方法的适用性比较

数据形态	评估方法	评估主体	辅助资料	参照标准
总量统计数据	A	数据生产者、使用者	同度量指标	同度量指标及其加减运算结果
	B	数据生产者、使用者	相关指标	经验比率、比例
	C	数据使用者	相关指标	相关指标的变动率
	D	数据使用者	相关指标	模型预测值或回归参数的时变性
个体统计数据	E	数据生产者、使用者	经验分布	经验或理论分布函数及其分位数
	F	数据生产者	重复调查数据	基于重复调查数据的总值估计
	G	数据生产者	重复调查数据、交叉子样本	两次调查结果的离散程度
数据形态	评估方法	技术假定	比较逻辑	误判因素
总量统计数据	A	逻辑包含、平衡关系	是否违背逻辑平衡关系	同度量指标含误差、统计制度缺陷
	B	比率、比例关系稳定	是否偏离经验比率	相关指标含误差、比率标准变化
	C	变动趋势相近	是否偏离相关指标变动趋势	相关指标含误差、相关关系变化
	D	变动趋势相近、稳定	是否偏离模型预测值	相关指标含误差、数值突变的客观背景
个体统计数据	E	服从经验统计分布	是否与经验分布不一致	经验分布假定不当、系统性偏移
	F	重复调查获得真值	是否偏离重复调查结果	重复调查含误差、两次调查结果相关
	G	调查条件不变、误差参数的统计特性	单元偏差的离散程度及其相关结构度量	误差参数假定不当、模型估计偏差

或分类统计数据准确性的统计分布检验(E)评估方法,必须事先掌握对于总体理论或经验分布的有关知识,这一点在很多情况下是难以做到的;而统计分布形态检验与异常数值识别方法本身则较易于实施操作。基于事后重复调查的偏差估计(F)与误差效应识别(G)有着共同的评判基础,即都要实施事后的重复调查,这需要有制度和大量人力物力的支持,因而只能适用于统计数据生产者(政府统计部门)方面的质量控制和评估实践;同时,这两类方法对于模型构造和参数估计也都有着较高的技术要求。

在有效性方面,统计数据的生产者可以利用A、B两类方法,在数据汇总的中间环节,依据统计核算规则、工作经验以及对社会经济运行状况的把握,及时发现和纠正各项统计数据之间存在的逻辑错误与虚/瞒报现象。统计数据的使用者(社会经济研究学者)还可利用C、D两类方法,基于社会经济理论对各项统计指标相关关系的论述,进行多方位的相关分析和模型分析,判断统计数据偏离“正常”水平的程度。总体上,A~D四类方法只能对影响统计数据准确性的偏差因素进行评估,而且这种评估通常只能给出方向性的提示,无助于进行具有公信力的统计修正。而基于事后重复调查的F、G两类方法,通过对统计数据的生成机制及其误差来源的全面把握,可以利用重复调查数据来对初始统计调查数据的各项误差参数(如偏差、简单计量方差、相关计量方差、调查员效应等)进行估计;经过科学设计的重复调查评估具有较高的精确性与可信度,评估结果对于统计部门改革统计工作制度、改善统计工作流程有很强的指导作用。

在可靠性方面,评估方法的技术假定如果不能满足,则评估结论可能导致两种误判倾向,一种是将实际可靠的统计数据判定为准确性差(可称之为“弃真”错误),另一种是将存在准确性问题的数据判定为可靠(可称之为“纳伪”错误)。针对总量统计数据准确性的A~D四类评估方法,由于不能保证逻辑相关指标的准确性,因而也难免结论的多重指向和误判倾向;总体上,四类方法犯“纳伪”错误的可能性更大一些,也即对统计数据中误差因素的识别能力不足,但也不排除存在将统计核算制度方面的缺陷,或社会变革导致的数据突变归咎于统计数据准确性问题的“弃真”可能。对于评估方法E,如果对经验分布的假定不当,会导致“弃真”错误的发生;而对于

所有(或大多数)原始单元统计调查的系统性偏差、真值靠近分布中心位置的个体误差,该类方法的识别能力不足,易于犯“纳伪”错误。基于事后重复调查的评估结论(尤其是关于调查统计量方差的估计),在很大程度上依赖于对误差机制(参数)所设置的假定与现实情况接近与否;模型估计本身也会带来一定的不确定性。

## 六、结论及应用展望

本文对用于统计数据准确性评估的各种方法加以归纳分类,对各类方法的技术细节及其适用条件进行了探讨和比较。针对总量统计数据的准确性评估,主要在纵向时间维度开展,根据统计指标与其相关联指标变动趋势的偏离程度加以评判,在此领域中统计数据的生产者和使用者都有参与。针对个体或分类统计数据的准确性评估,则主要在横向空间维度开展,对数据的统计分布形态加以一致性检验,或者利用重复调查或随机实验方法对事先假定的误差参数进行估计,主要由统计数据的生产者即统计部门实施。不同方法具有不同的适用条件,在具体对统计数据准确性进行评估时,必须基于实际可以掌握的辅助资料信息的翔实程度来选择评估方法,结合对各种资料来源可靠性和技术假定合理性的判断,对导致评估方法发生误判的因素加以分析和排除,以确保评估结论具有说服力和可信度。

为了在更深层次、更广范围内融入国际社会,使我国统计制度不断健全并与国际接轨,提高统计数据的透明度与可信度,关于统计数据准确性评估的研究和实践仍需不断深入进展。而在当前政府绩效考核与经济总量、能耗污染等统计相挂钩的宏观背景下,为了杜绝对统计数据的人为干预和造假行为,对于统计数据准确性进行适时、有效、可靠的评估更凸显其重要性和迫切性。加大这一方面的研究力度,开发相应的评估方法体系,推进统计数据质量评估的规范化与评估结果公布的程序化,才能为改革和完善统计管理体制与统计工作流程提供科学的参考依据、为进一步开展的统计质量控制指明方向,才能对统计造假者产生实质性的震慑作用,并增进社会公众对政府统计的信任。

在现阶段,鉴于各类方法在逻辑思路或技术方面都存在一定的缺陷,在讨论有关中国统计的问题时,应更多关注对整个经济数据的精确度、范围、覆盖面、统计方法以及其他技术问题(Rawski, 2002)。

对于评估争论的最终解决办法,要依赖于国家统计局所使用方法的改进和在日常工作中对各种数据来源的改善(任若恩,2002)。

### 参考文献

- [1] Rawski, T. G. 近年来中国 GDP 增长核算:目前的状态[J]. 经济学季刊, 2002, 2(1).
- [2] 成邦文, 董丽娅, 杨峻. 研究与开发机构统计数据质量与异常点的对数正态分布检验与识别[J]. 统计研究, 2000(1).
- [3] 成邦文, 董丽娅. 社会经济统计数据的误差分析[J]. 统计研究, 2002(11).
- [4] 傅德印. 利用探索性数据分析法对统计汇总数据进行质量控制的尝试[J]. 数理统计与管理, 2001(1).
- [5] 胡桂华. 美国 2000 年人口普查的准确性和数字评估调查[J]. 西北人口, 2007(4).
- [6] 刘洪, 黄燕. 我国统计数据质量的评估方法研究——趋势模拟评估法及其应用[J]. 统计研究, 2007(8).
- [7] 刘孝新, 朱慧明, 胡先红. 事后预测及反常结果判断法在统计数据质量评估中的应用[J]. 统计研究, 1997(2).
- [8] 刘延年. 如何评价统计数据的质量与可靠性[J]. 统计研究, 2002(8).
- [9] 卢依吉, 法布里斯, 徐志全, 戴宏国. 中国农业普查事后质量抽查[J]. 统计研究, 2002(8).
- [10] 孟连, 王小鲁. 对中国经济增长统计数据可信度的估计[J]. 经济研究, 2000(10).
- [11] 任若恩. 中国 GDP 统计水分有多大——评两个估计中国 GDP 数据研究的若干方法问题[J]. 经济学季刊, 2002, 2(1).
- [12] 屈耀辉, 曾五一. 农产量抽样调查中计量误差来源分析及其事后甄别[J]. 财经理论与实践, 2004(2).
- [13] 王广州. 对第五次人口普查年龄结构数据的评估与调整[J]. 人口与经济, 2004(5).
- [14] 伍荣坤. 定期统计报表数据质量组合评估方法初探[J]. 统计研究, 1993(1).
- [15] 许宪春. 中国国民经济核算体系改革与发展(修订版)[M]. 北京:经济科学出版社, 1999.
- [16] 杨冠琼. 中国经济增长数据可信度检验研究:理论、模型与实证检验[M]. 北京:经济管理出版社, 2006.
- [17] 杨清, 吴伟霞. 对原始资料偏误的监控技术初探——论提高统计数据质量的重要步骤[J]. 统计研究, 2000(6).
- [18] 尹彦君. 地区 GDP 数据质量的评估控制方法研究[A]. 国家统计局国民经济核算司. 地区 GDP 核算论文集[C]. 北京:中国统计出版社, 2004.
- [19] 张为民, 崔红艳. 对中国 2000 年人口普查准确性的估计[J]. 人口研究, 2003(4).
- [20] 张新, 蒋殿春. 中国经济的增长——GDP 数据的可信度以及增长的微观基础[J]. 经济学季刊, 2002, 2(1).
- [21] 周建. 宏观经济统计数据诊断:理论、方法及其应用[M]. 北京:清华大学出版社, 2005.
- [22] Groves, R. M. *Survey Errors and Survey Costs*. New York: John Wiley & Sons, 2004.
- [23] Hansen, M. H., and E. S. Marks. "Influence of the Interviewer on the Accuracy of Survey Results." *Journal of the American Statistical Association*, 1958, 53: 635 - 655.
- [24] Hogan, Howard. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association*, 1993, 88: 1047 - 1060.
- [25] Lardy, Nicholas. "China Will Keep Growing." *Wall Street Journal online*, 14 June, 2002.
- [26] Maddison, Angus. *Chinese Economic Performance in the Long Run*. OECD Development Centre, Paris, 1998.
- [27] Rawski, T. G. "What's Happening to China's GDP Statistics?" *China Economic Review*, 2001(4): 347 - 354.
- [28] Ren, Ruoen. *China's Economic Performance in An International Perspective*. OECD Development Centre, Paris, 1997.
- [29] U.S. Census Bureau, and Bureau of Labor Statistics. *Current Population Survey: Design and Methodology*. Technical Paper 63RV, March, 2002. Available at <http://www.census.gov/prod/2002pubs/tp63rv.pdf>.
- [30] Wu, Harry X. "An Alternative Estimation of the Post-War Chinese Industrial Production and Growth." Discussion Paper, D99-10, 1-51, Institute of Economic Research, Hitotsubashi University, 2000.

### 作者简介

王华,男,1978年生,山西翼城人,2006年毕业于中国人民大学统计学院,获经济学博士学位,现任厦门大学台湾研究院助理教授、硕士生导师,国家“985工程”二期台湾研究哲社创新基地成员,研究方向为统计数据质量评估、区域经济统计分析。

金勇进,男,1953年生,北京市人,中国人民大学经济学博士,现任中国人民大学统计学院院长、教授、博士生导师,中国人民大学应用统计科学研究中心研究人员,研究方向为抽样调查、应用统计。

(责任编辑:赵曾琪)