

社会经济统计研究中的若干处理方法

范正绮

随着社会主义市场经济的逐步完善,对社会经济的研究工作自然地要求抬高一个层次,要做到既定性又定量地研究问题。对社会经济的研究中,较为理想的手段之一就是利用数据建立经济模型,借助于现代化的计算机工具及先进的统计或其它软件,使得现实的社会经济的某些方面可以在计算机上再现。

然而在我们最近的研究中发现,有时候计算机统计模拟会出现较大的偏差。在理论上已经证明是科学的东西,在实用中却偏差惊人,这一点不能不引起我们的重视。经过认真分析,这里面的很多问题主要来源于原先的统计方法。统计方法的使用不当常会引起诸如数据缺陷或得到错误的信息以致令人误解等等。我们在下面对若干情况进行剖析。

1. 有关观察研究问题。

众所周知,定量研究经济首先要收集数据。社会经济方面的数据要进行不同情况的对照来获取并分析。比如,实行股份制是否对企业的经营有大幅度的促进这一课题的研究,通常需要对股份制企业与非股份制企业这两个组的各项经济指标进行观察、比较才能得出结论,或者通过某些企业的实施股份制前后的各项经济指标的对比来进行研究。这种对比的方法在自然科学、医药等各方面比较容易安排,常常由研究人员决定试验对象或单位分在对照双方的哪一方。例如想要试验某种新药的疗效,医药统计工作者可以将病人随机地分为两个组,一组用新药,一组

用老药,等等。但是,社会经济问题并非如此理想,对于对比的双方,我们常常只能观察而不能安排。举例来说,(假定)想调查假冒伪劣产品给自身带来的非法暴利以及给社会带来的危害,我们只能客观地观察并收集数据,无权也不可能故意安排一些单位去进行伪劣产品的炮制,从试验对象来说,也未必会去听从某些安排。倘若想研究酗酒的社会危害性,我们无法去安排对象进入酗酒的一方或者正常的一方,也不会有这样的人因为要奉承统计工作者的需要而故意去酗酒。

正因为社会经济现实中观察研究的对比双方不是由统计工作者决定的,因此在对比双方的形成过程中有时候会混进令人不易察觉的因素,从而使所得到的结果产生混淆。比如我们比较股份制企业与非股份制企业的效益,通常一个企业转为股份制的话,它本身必须具有一定的资产、人力资源,有一整套较先进的管理系统以及有一定的经济效益。由于上述条件在不同企业中是不一样的,因此对这两个方面作单纯的经济增长比较是不合理的。出于同样的理由,单纯地比较乡镇企业与国营大中型企业的经济增长并不能立即断定国营大中型企业远比乡镇企业来得差。乡镇企业起步晚底子薄,尤其是它在有些情况下得到了国营大中型企业无法得到的政策优惠,使得这两个方面在不少经济数字方面是不可比较的,要比较就有可能将人引入歧途。

既然观察研究中所得数据常有可能涉及到研究之外的其它因素,那么我们在将数据

输入计算机以建立数学模型前必须要对观察研究加以调整,否则数据一旦进入计算机就很难发现出错的原因。因为计算机里的数据仅仅是数据,而不会呈现出其它因素的影响,用这样的数据去建立社会经济模型当然不能反映实际情况。要消除其它因素对研究目的的影响,一个常用的办法是缩小对比双方的范围,将对比的两个组除了要比较的东西之外在其它因素方面做到尽可能地类似,甚至完全相同。这样观察到的有关数据才能在计算机上建立相应的能反映现实的模型。

2. 有关伪劣数据问题。

伪劣数据有多种的形式,有的是谎报的,有的是故意扩大的,有的是不想提供实情而胡乱填的,有的是笔误或通讯工具等原因造成的,有的是概念上的混淆而引起提供数据的失误,有的是错误的逻辑推理而形成的,等等。本文不准备去研究伪劣数据的形式以及它的各种来源。通过近年来一直在从事统计计算工作所获的经验,作者认为从统计计算的角度出发,我们至少得寻求一种手段来鉴别伪劣数据。在计算机上采用科学的推理,以做到“辨别香花,剔除毒草”。我们的这种想法已得到有关专家的重视。

除了认真贯彻执行有关统计法,尽可能地使数据保持真实之外,我们建议,在实际收集数据的过程中,统计工作者应当尽全力确保一部分数据的真实性。一般来说,通过细致科学的工作是可以做到这一点的。利用这些可靠的数据我们建立起社会经济的模型,在这个模型中,各个指标之间的相互关系,各个指标的基本趋向等情况应当在计算机上一目了然。这样的—个模型为我们鉴别其它数据的真伪提供了依据。—组数据一旦输入计算机,如果与我们已建立的模型拟合得很好,有理由认为它们是符合常规的,至少我们没有理由讲它们一定是伪劣。倘若与模型有很大差距,我们称之为“异常值”。对于这样的异常值必须作进一步调查与分析,它有可能虽异

常但却是真实的,这就促使我们去找出原因,从而达到对某些社会经济现象的阐述、解释与研究。若不然,我们就有理由怀疑其伪劣性,从而对伪劣的数据进行处理。

3. 有关删失数据。

统计工作中另一个存在的问题是数据的删失。譬如在抽样调查中有些敏感问题遭到了某些被调查者的拒绝回答,虽经再三做工作,这样的拒绝回答仍然存在(如果强迫被调查者回答,一来于情理上讲不通,二来有可能导致伪劣数据的出现,反而对统计工作不利)。

众所周知,计算机处理数据时是有规律的,利用计算机阅读调查表格若发生空白格现象,在紧接着的统计计算公式中会产生错误。对于这个问题,美国国家统计局采取的对策是将所有出现的空格全部填上人为的数据。我们知道这些数据不能也不应当过分地胡编乱造,否则就会出现上面所提到的伪劣数据。要使所填上的人为数据具有一定的依据,办法之一就是从小非空白的数据中进行随机地抽取。这样既使人为数据合理化又可使计算机顺利进行统计分析。但是这样的处理毕竟使我们统计分析产生了新的误差。当然这也是没有办法的办法。到底采用什么样的办法弥补空缺数字最好,看来需要统计工作者和社会经济研究人员进行很艰苦的努力,以寻找解决的方法。

4. 有关收集数据过少的问题。

在社会经济研究的调查中,有时会遇到某些数据只能少量地收集到。例如希望了解某项特殊经济政策(或措施)的作用。由于试点单位少以及时间方面的短缺,造成了有关数据只有少量被收集。数据的少量给统计计算及分析带来了较大的困难,有时由于误差较大有可能引起推断的失误。怎样充分利用所收集到的少量数据所提供的信息,又是当今统计工作者面临的一个棘手问题。