

六普结果表明以往人口估计和预测严重失误*

郭 志 刚

【摘 要】文章以六普人口年龄结构为标准,模拟过去 20 年人口进程及其主要人口指标,然后与其他来源的人口指标进行了比较和分析。根据六普结果可以判断,1990~2010 年间的人口估计和预测存在的普遍问题是高估了出生人口数量,高估了生育水平,高估了人口增长,从而低估了人口老龄化程度。研究发现,由于高估出生和生育水平而导致的过分的统计调整,造成的偏差幅度甚至远远超过了原始调查统计的偏差,形成了严重的误导。这种状况反映出多年来人口统计上的迷茫,思想认识脱离实际。

【关键词】六普 模拟预测 人口指标 统计调整 偏差

【作 者】郭志刚 北京大学中国社会发展研究中心研究员、社会学系教授。

一、研究背景与研究目的

2010 年第六次全国人口普查(下称六普)公布的主要数据表明,中国大陆地区总人口为 13.3972 亿。其中,0~14 岁少儿人口占 16.60%,60 岁及以上老年人口占 13.26%,而 65 岁及以上人口占 8.87%。这些主要数据明确告诉人们,中国总人口的增长远低于以往人口预测和规划的水平,而在人口结构方面的少子化和老龄化则又明显超出了以往人口预测和规划。这些结果已经充分地揭示出,近 20 年来在人口预测与规划中普遍存在着对中国生育水平的严重高估,进而导致高估出生水平并低估人口老龄化。

图 1 提供了 2000~2010 年每个年度《国民经济和社会发展统计公报》的人口总数,同时提供了政府主管部门在制定“十五”和“十一五”人口规划时所使用的人口预测结果来进行比较。实际上,在近年制定 5 年人口规划时都将总和生育率(TFR)设定为 1.8 进行预测,然而对比历年统计公报的总人口数与最新公布的六普总人口数,已经足以表明以往的人口预测所使用的人口基数与总和生育率的假定严重脱离实际。

* 本研究为国家社科基金重点项目“我国人口发展与经济社会可持续发展战略研究”(项目批准号:08ARK001)的阶段性成果。

“十五”规划时人口预测所用的 2000 年人口基数要远高于五普公布数，而且其 2005 年预测人口数比 2005 年全国 1% 人口抽样调查公布的总人口数高出 2 000 多万。10 年前这个人口规划预测之所以产生如此大的偏差，一方面是由于当时高估了未来年份的生育率，另一方面是因为当时采用的人口基数过大，即高估了 20 世纪 90 年代的生育水平。

“十一五”规划时的人口预测没有再调整人口基数，直接采用了 2005 年全国 1% 人口抽样调查的人口数作为基数，然而总和生育率假定依然照旧，结果其预测总人口仅在 5 年后便又比六普总人口高了 2 000 万。产生这个偏差主要是由于高估了总和生育率。

人口预测在短短几年产生如此之大的偏差并不是人口预测方法的问题，而是因为预测采用的人口基数和人口变动参数假定严重脱离实际。笔者(2004)曾经根据多种人口调查数据的分析结果指出，在 20 世纪 90 年代后期，总和生育率(作为时期实际生育水平测量)已经处于 1.5 以下的可能性很大。即使考虑一定程度漏报而需要调整，仍用总和生育率 1.8 (甚至以上)作为当时的实际生育水平显然已不合适。然而，这种预警一直没有得到重视，近年来几乎所有人口预测和规划都忽略了实际生育水平更低的可能性。不过仅仅几年后，六普公报的总人口数便印证了以往人口预测和规划中存在的这种致命缺陷。可见，在 20 世纪 90 年代初中国总和生育率下降到更替水平以下之后，政府主管部门和人口学界反而失去了对总和生育率的正确把握，这种迷茫已经持续了 20 年。那么，以往 20 年间中国的总和生育率到底有多高？这个问题不解决，就无法回答以往人口预测和规划中到底对总和生育率高估多少的问题，也无法搞清楚人口出生的实际水平和高估幅度。

目前公布的六普主要数据并不能回答这些问题，而且即使正式出版的六普人口数据集提供的统计结果也不能直接回答上述两个问题。因为按照普查人口数据的标准表式，也只能提供人口普查前一年的生育率。然而，人口是一个系统化的发展进程，严格遵循人口变动的自身规律，因此可以应用人口预测技术来模拟再现出以往 20 年的人口进程，以获得对以往人口发展的总体把握，大体估计出各年相应的人口统计指标。通过这样的模拟估计，还可以大体推算出以往公布各种人口指标的偏差程度。本文的研究目的就是试图完成上述这种人口模拟预测工作，并进行多种主要指标的比较和分析。

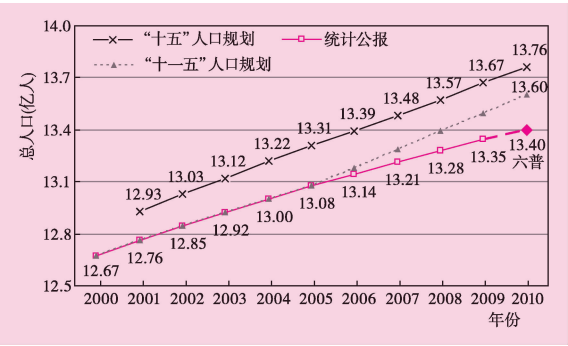


图 1 不同数据来源的总人口规模比较

注：图中“统计公报”为历年《国民经济和社会发展统计公报》公布的数据；“‘十五’人口规划”引用庄亚儿、张丽萍(2003)编：《1990 年以来常用人口数据集》；“‘十一五’人口规划”引用国家人口计生委(2006)：《全国“十一五”人口和计划生育事业发展规划》。

二、研究方法 with 数据

(一) 人口模拟预测方法

常规人口预测是在基期人口数量结构上按照人口变动参数发展变化,以取得相应的人口统计指标。其中的关键是两个假定:(1)假定所用基期人口数量结构数据准确;(2)假定人口变动参数合理。如果这两个假定都比较符合实际,那么人口预测结果便可以预示未来人口发展过程。即使这两个假定与实际情况存在一定差距,人口预测结果也可以大致反映未来人口进程的趋势。相对人口基数而言,未来人口参数的设定更难,因为这些参数是对未来的变化所做的预计。因此,人口预测对主要参数往往设置高中低的不同水平,实际上是给未来变化留出一定程度的余地。而以往对中国人口预测的普遍缺陷正是对生育水平很低的这种情况估计不足。

人口预测方法也可以反过来用,即反向预测。比如,用六普人口数量结构作为基础,然后按人口变动参数反向推算以前各年份的人口指标。这种反向预测要求作为预测基础的当前人口数据完整和准确,而人口变动参数则可以参考甚至直接采用以前年份的人口指标。本研究要做的则是介于上述两种预测之间。由于六普全面数据尚未公布,不能从当前人口状况反向预测以前年份。因此采用的操作步骤是:(1)进行1990~2010年的人口估计,与常规预测不同的是,这一阶段的实际人口进程已经有大量调查数据可以参考应用。(2)本研究的预测不是一次完成,而是反复对预测基数和参数进行试错调整,使2010年人口预测结果尽量趋近六普的结果。换句话说,本预测首先将六普人口作为一个特定的人口结果或目标,通过多次预测尝试和基数或参数的调整,将1990年的中国人口状况与六普结果之间建立起系统化联系,于是最后的模拟结果一方面提供了这20年间的人口指标,另一方面还要尽量保证这套人口进程指标与六普结果相一致。

本研究的预测模拟有双重探测的意义。一方面,假定六普结果正确,预测模拟以六普结果为靶标的性质可以探测出以往20年中国人口进程的大概状况。由于以往对实际人口态势迷茫,因而很多重要人口统计指标均做了大幅度的统计调整。而这套模拟结果则有助于挤掉以往人口统计指标中过分调整的“水分”,还原最可能的实际指标值(尤其是生育水平),并进一步了解以往高估总和生育率的偏差到底有多大。另一方面,也可以根据对预测模拟中在人口基数和参数上的调整幅度来判断其是否超过了合理的范围,反过来检验六普结果的质量和可接受性。在本研究中,笔者将侧重点放在对前一个方面的探讨上。

(二) 根据六普人口金字塔推计模拟预测的靶标

《2010年第六次全国人口普查主要数据》中提供了六普的单岁人口金字塔图形(见图2)。蔡泳根据此图用绘图软件测算了六普单岁年龄结构比例,然后结合六普公布的总人口数及其性别结构信息进一步推算出六普的单岁组性别年龄别人口数(蔡泳,2011)。尽管这套六普性别年龄别人口的推算数据会有一定的测绘误差,而且六普详细数据还有可能做进一步的清理和调整,但这些情况并不会导致显著的差异。蔡泳向本研究提供的这套推计数据

中的大量信息,在六普详细数据公布前便提供了本研究人口模拟的靶标,使后面根据六普人口结果探测 20 年来人口过程及其有关指标的研究得以实现。

（三）实际操作的简要报告

预测模拟先以 1990 年全国人口普查的分城乡的单岁性别年龄别人口数量为预测基数,其中已经加入了当次普查的军队人口,并且还相应调整了由 2000 年人口普查证实的低龄漏报人口。人口基数中的全部年龄队列将大体对应六普时的 20 岁及以上人口。对这些队列的模拟结果是否能与六普吻合的主要因素是:(1)模拟中的人口预期寿命参数;(2)四普人口数据质量;(3)六普人口数据质量;(4)模拟中对人口城镇化的处理是否得当,因为城乡人口的预期寿命水平存在着显著差异。

本预测采用国家统计局(2005:253)公布的 1990 和 2000 年人口预期寿命作为参数设置标准^①,然后分解到城乡,并制作出各类预期寿命均匀增长的相应系列参数。在对六普结果的模拟“打靶”中,主要问题是模拟的成年人口数(20 岁及以上)远低于六普结果。这种迹象表明模拟中采用的预期寿命参数过低,或者是四普存在人口漏报问题,而另一个可能原因则是六普的劳动年龄人口和老年人口存在一定的重报现象。本研究的主要关注点是以六普结果为标准再现以往人口历程,于是预测模拟首先尝试不断提高预期寿命参数的方法来提高成年人口的吻合度。继而,模拟研究又进一步发现男性和女性之间的差异,即男性寿命参数提高 1 岁左右便可以大致吻合六普结果,然而提高寿命参数对模拟女性成年人口却难以奏效,因为即使将女性寿命提高 9 岁(已远超过合理范围),20~45 岁妇女人数仍然预测不足,而老年段妇女人数却已经明显超出了。这种现象表明,六普人口与四普人口之间的不一致不是简单的寿命参数问题,要么是四普低龄女性漏报极多(只是四普时已经 10~25 岁的女性为什么漏这么多仍说不清),要么是六普 20~45 岁女性重报较多。

在现有数据条件下,研究六普数据质量问题还难以做到。因此,本研究将预期寿命参数提高到一个合理程度后,若模拟结果还不能与六普吻合,便将模拟的各年龄成年人口数与六普的差异作为四普相应队列的漏报或其他原因产生的误差,相应调整预测模拟所用的 1990 年人口基数。最后,预测模拟将国家统计局公布的 1990 年的男女预期寿命分别提高了 0.65 岁和 1.28 岁,将国家统计局公布的 2000 年的男女预期寿命分别提高了 1.20 岁和 1.80 岁。在此调整之后,2010 年的成年人口与六普结果之间的不一致将通过对 1990 年的预

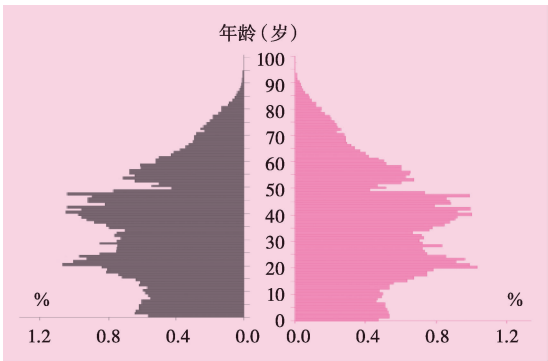


图 2 2010 年全国人口普查人口金字塔

资料来源:国务院第六次全国人口普查办公室、国家统计局人口和就业统计司编:《2010 年第六次全国人口普查主要数据》,中国统计出版社,2011 年。

① 1990 年全国平均预期寿命男性为 66.84 岁,女性为 70.47 岁;2000 年全国平均预期寿命男性为 69.63 岁,女性为 73.33 岁。



图3 1990年普查年龄别人口数及按六普模拟的调整数

表1 对1990年普查人数的调整幅度

年龄组(岁)	调整人数	年龄组(岁)	调整人数
0~4	9902157	35~39	-499026
5~9	1954901	40~44	864971
10~14	3879530	45~49	348600
15~19	2578436	50~54	821425
20~24	966962	55~59	605968
25~29	-2542056	60~64	1751843
30~34	3530641		

测基数调整来加以弥合。图3给出了1990年普查公布的年龄别人数(包含军队人口)与本研究模拟调整后的相应结果。调整幅度最大的是四普的0~4岁组,共调高了990万人,并且其中仅0岁组便调高503万人。其次为10~14岁组,调高387万人。位于第三位的是30~34岁组,调高353万人。本研究模拟时相对四普各5岁年龄组人数的调整幅度列在表1中。

在预测模拟中,1990~2010年的出生人口数及其性别构成按照六普0~20岁的人口数为基准,实际上已经考虑了出生后各队列人数所经历的死亡影响。预测模拟的人口城镇化进程设置参考了历年国民经济和社会发展统计公报及六普的城镇人口比重,而人口城镇化的性别年龄别模式与生育的年龄别模式等其他人口变动参数则参照了2000年人口普查、2005

年全国1%人口抽样调查和其他全国人口调查提供的信息。

三、模拟结果与其他来源统计的比较与分析

当前的人口总量及其年龄结构中镌刻着以往数十年的人口经历。本研究的预测模拟要达到的主要目标是:假定六普数据准确无误,模拟出1990~2010年各种人口指标;然后,用这套与六普性别年龄别人口数相一致的模拟结果与以往不同渠道公布的人口统计加以比较,探测以往人口统计能否经得起六普结果的检验。

(一) 前些年高报了出生水平

图4提供了模拟的以往年份年出生数与相应历年国民经济和社会发展统计公报的出生数的比较,从中可以看到,1990年统计公报的出生数显著偏低,1991~1993年公报出生数与六普结果较为吻合;1994~2003年公报人口出生数则显著偏高,尤其是1996~1999年严重偏高,达到400万左右。这些人在六普时已经10岁以上,不应该存在显著的漏报问题,死亡也不会对队列人数有明显的影响。

再来看以往出生的绝对数量,根据六普结果,1996年出生人口数已经降到1 658万,1999~2003年甚至降到1 500万以下。虽然最近几年出生人数略有回升,但仍保持在1 600万以下的水平。综合来看,尽管六普结果表明四普数据有严重的低龄人口漏报问题,但相

比之下,随后年份出生人口数高估、高报的问题更为突出。计算各年公报出生数减去六普模拟出生数的偏差,然后从 1990 年开始逐步累计出生偏差数就可以看出,20 世纪 90 年代前期的统计公报偏低结果在 1996 年时便已经被后来高报的出生人口数所完全抵消,2000 年时出生累计偏差为 1 600 万,到 2005 年时已高达 2 400 万。由于近年来统计公报的偏差很小,因此最近 5 年中上述出生累计偏差仅增加 74 万。

图 4 中还提供了翟振武、陈卫(2007)用教育统计的小学生数据估计的 20 世纪 90 年代的中方案出生人口数曲线(“战略中方案”),这套出生人口数量估计曾被用于国家人口发展战略研究的人口预测来调整低龄人口基数。一些研究曾经指出,很多证据表明教育统计数据与其他统计一样也存在着本身的问题,并不见得那样纯净和准确,并且研究估计在逻辑和方法上也有值得商榷之处(蔡泳,2009;郭志刚,2010b)。从图 4 可以看出,这套出生人口数量估计在前期略高于统计公报出生数,在后期略低于统计公报出生人口数。然而,如果以六普结果为标准来检验,这套出生估计的主要问题也同样是严重偏高。

实际上中国统计部门早在 21 世纪初就已发现,教育统计的小学在校学生数甚至已经高于人口统计中的学龄人口数(教育部教育管理信息中心,2006)。但当时人们只是怀疑人口统计中的漏报问题,并不怀疑教育统计。然而,看到六普结果后再重新审视,其实教育统计的重报高报问题本身也是上述矛盾的原因之一。由于存在小学在校人数高于人口统计中的相应年龄人口数的矛盾,中国统计部门回避向联合国教科文组织提交小学适龄人口数。于是联合国统计机构使用中国小学生人数作为分子,用 1990 年中国人口普查数据辅以 1.8 的总和生育率推算出小学适龄人口来作为分母,将中国的小学入学率计算得极低,只有 91.7%。中国教育部当即提出异议,并自己用 2000 年中国人口普查数据辅以 1.6~1.7 的总和生育率算出相应分母,计算的小学入学率为 97.6%(教育部教育管理信息中心,2006)。

这种计算结果的差异,暴露出中国教育统计与人口统计之间的矛盾,继而又国内人口统计根据小学生数反推出生人数并估计漏报水平的研究。3 种估计都涉及了教育统计的小学生数、小学入学率、人口统计的小学适龄人口数及其背后体现的生育率这几种数据。需要指出的是,尽管当时已知小学生数超过了五普的相应年龄人口数,但所有这些推导估计都未对小学生数的统计质量产生任何怀疑。联合国相信四普数据和较高的生育水平,结果把中国的小学入学率算得很低,而教育部门相信五普结果和较低的生育水平,推算出较少的小学适龄人口数,并得到较高的小学入学率。而在人口统计研究中采用相反的论证逻辑,用较低的小学入学率从小学生数反推出较高的低龄人口数,来证明出生漏报严重和 1.8 的总和生育率估计,而这样一来也就几乎否定了五普的低龄人口数据质量。

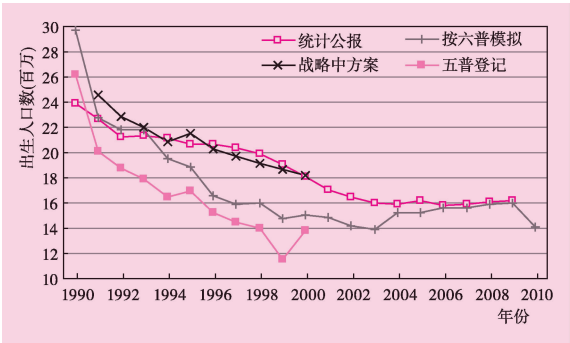


图 4 1990~2010 年出生人口数的比较

图4中还提供了2000年人口普查相应低龄队列的人口数,并没有考虑死亡影响来回推队列出生时的人数(实际上要更多)。与六普结果比较,五普的低龄组人口数的确存在显著漏报。但与以往年龄越小漏报越严重的经验相反,五普时5岁以上的年龄组(即1990~1994年出生)的偏差更大,而更幼小的年龄组人数反而相对更接近六普结果。更重要的是,若以六普人口为标准,1995~2000年,统计公报出生数和从教育统计估算的出生人口数反而要比五普低龄人口的偏差幅度更大。这就表明这些年份出生人口估计中包含的漏报调整过大,基本上被六普结果所否定。

1992年国家计生委的38万人调查初次显示,总和生育率远低于更替水平后,人们一时难于理解和解释,因而对出生漏报的指责和压力越来越大,于是后来进行统计调整的幅度也就越来越大。直到2000年第五次全国人口普查及2005年1%人口抽样调查以后,人们才逐渐意识和接受生育水平很低的现实,因而对出生统计也不再像以前那样过分调整,反而开始对以前高报的出生水平加以修正。然而,如上所述,以往出生人口统计高报的累计偏差已经非常大,很难一下调整下去。如果以人口出生率指标来比较,1996~1999年这4年统计公报的人口出生率均比模拟结果高出3个千分点以上。

(二) 死亡水平略有高报

图5提供了1990~2010年死亡人数的比较,从中可以看到,1991~1998年统计公报的死亡人数与六普结果的差异逐渐扩大,1998年两者的差距约为32万人。1999~2004年,二者之间的差距又逐渐缩小。2004和2005年二者已经很接近,然而在此之后,二者之间的差距又骤然扩大,达到59万。应该指出,最近4年统计公报的死亡水平迅速攀升是个奇怪的结果,由于六普模拟结果已经将人口结构变化和期望寿命提升的影响包括在内了,因此这样的死亡水平攀升显得不成比例。笔者发现,近年统计公报死亡水平的攀升实际上越来越接近联合国2008年对中国人口预测的水平。然而,联合国2008年预测的中国死亡率水平在整个20年的预测模拟期间一直远远高于按六普人口模拟的死亡率水平。于是可以推测,近年统计公报的死亡人数偏高可能是为了抵消部分人口增长,来追求全国总人口数及其增长方面的平衡所进行的调整。若从死亡率指标来比较,1993~2000年的多数年份二者之间的

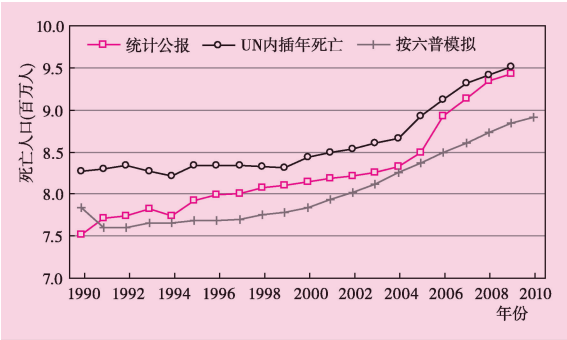


图5 1990~2010年死亡人数的比较

差距都在0.3个千分点以上,2004和2005年二者比较接近,但最近几年二者之间差距又增大为0.4‰左右。

(三) 高报了人口增长数

图6提供了每年人口增长数之间的比较,从中可以看出,1991~1993年统计公报的人口增长数与六普模拟结果之间的差异很小,反映出在20世纪90年代前期,统计部门对人口指标的调整并不太

大。然而,1994~2003 年的统计公报人口增长数与六普结果之间的差异则是先增大、再缩小。差距最大时处于 1996~1999 年,统计公报人口增长数偏高均在 350 万以上。但是,2004 年后统计公报的人口增长数却已经十分接近于六普结果。因此,六普结果揭示出以往统计公报明显高估了人口增长。由于人口增长是由出生和死亡两个部分的人口变动决定的,而以上比较揭示出以往的出生偏差要以百万计,死亡偏差则是以 10 万计。以往出生和死亡人数的高估反映到人口增长上时并不是二者叠加,而是相互抵消,所以图 6 中的人口增长偏差完全是高估出生人数被高估死亡人数抵消后的剩余影响。

图 6 中还比较了根据“十一五”人口规划预测各年人口数计算的年增长人数。尽管该人口规划声称是用 1.8 的总和生育率做的预测,其实 2001~2005 年人口增长数直接采用了统计公报结果,2006~2010 年规划期间的每年人口增长数则是一个常数(1 048.8 万),且数值远高于统计公报和六普模拟的相应结果。从人口自然增长率指标来比较,在差距最大的 1996~1999 年这 4 年中,统计公报中的人口自然增长率的偏高约为 3 个千分点。在 2004 年以后,两者的差距很小。

（四）以往的人口预测低估了老龄化水平

笔者(2010a)曾指出,高估出生水平和生育水平必然会导致严重低估人口老龄化,并且发现近年来统计公报公布的老年人口比例已经明显高于用总和生育率 1.8 预测的结果。这已经表明以往人口预测中的偏向现在就已不能忽视,那么这种偏向对未来的人口老龄化的低估将更严重。一些学者的观点正好相反,认为由于出生漏报和生育水平被严重低估,以往人口预测反而高估了未来人口老龄化和劳动力数量减少的风险。

图 7 提供了六普数据所反映的以往 65 岁及以上老年人口比例的动态变化,来对比历年统计公报公布的数字^①,以及 21 世纪初国家人口计生委按总和生育率 1.8 进行预测的老年人口比例^②。从中可以看出,相对于从六普结果回推的老年人口统计,不仅国家计生委 10 年前预测的老年人口比例明显偏低,而且以往历年统计公报公布的老年人口比例也都是偏低的;以往统计公报结果与六普结果之间的差距也是先扩大再缩小的。尤其是从统计公报的 2009 年底统计到六普开展调查只有不到 1 年的时间,但老年人口比例提高幅度却明显大于以往年份提高幅度。这并不能理解为是 2010 年有规模非常大的队列步入 65 岁以上所致,因为从六普结果模拟的老年比例中已经包含了这个因素。更合理的解释是,以往高估出

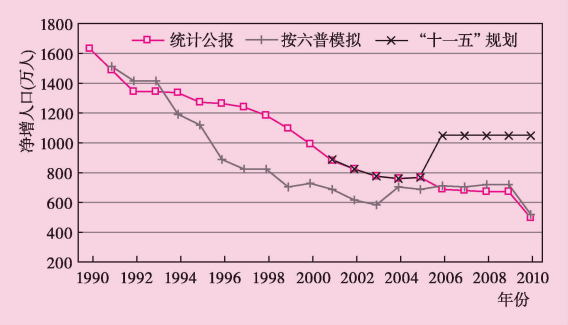


图 6 1990~2010 年人口增长数的比较

① 在统计公报的人口老龄化曲线中,2000 和 2010 年的统计实际上来自于人口普查结果。

② “十一五”人口规划文本中没有涉及人口老龄化方面,也没有找到相应人口预测数据。

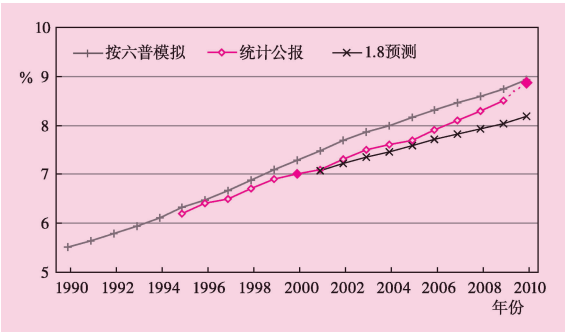


图7 1990~2010年65岁及以上老年人口比例

总之,尽管六普结果表明以往调查存在漏报,但如图4所示,以往对出生和低龄人口的漏报问题被严重夸大才是以往老年人口比例被低估的最主要原因。以往对人口预期寿命水平估计偏低,而且在人口预测中对预期寿命的提高速度又估计不足,则是次要原因。

(五) 其他年龄结构指标的问题

值得注意的是,以往年份统计公报中的老年人口比例变化比较均匀,但其他年龄结构比例则不是这样。图8、图9分别提供了统计公报的少儿人口比例和劳动年龄人口比例,并与按六普模拟的相应指标进行比较。从中可以看出,统计公报中的这两种人口结构统计

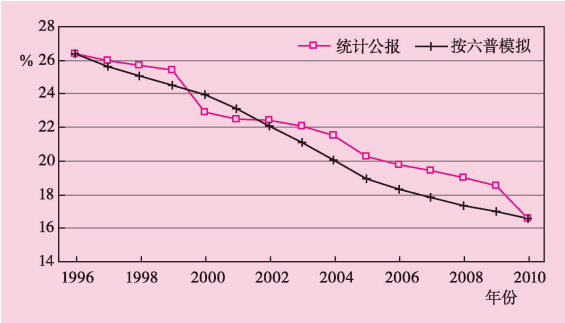


图8 1996~2010年0~14岁少儿人口比例

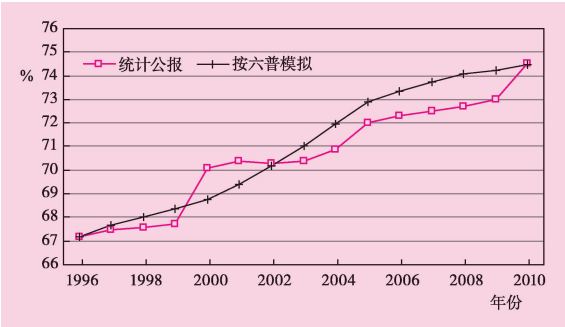


图9 1996~2010年15~64岁劳动年龄人口比例

生及其相应统计调整低估和掩盖了人口老龄化的实际水平。换句话说,以往人口统计的问题是在用猜测存在但实际并不存在的出生人口在统计上“缓解”了人口老龄化。这两条曲线之间的差异从2005年起开始缩小,其实反映的是国家统计局对上述人口统计偏向有所认识,并且开始有意识地进行反向调整,以便克服以往的偏差。

指标的变化都呈现出极为明显的阶梯状,并且跳跃性变化均发生在实施全国人口普查或全国1%人口抽样调查的年份。此外,这两个指标的跳跃性变化呈反相对应,且幅度相当。

这种夸张的变化特征并不是人口进程的实际表现,而是反映出作为每年统计公报的数据基础的人口变动调查的年龄结构特征与人口普查或1%抽样调查的年龄结构特征总是差异很大。但更可能是人口普查或1%人口抽样结果才足以反映到统计公报的年龄结构统计中去,而其他年份则是结合当年人口变动调查结果与以往统计的趋势外推两者来估计,少儿比例的外推便很容易受到高估出生人口数偏向的影响。几年高估的出生人口数便累积很多,然后再根据新的更大规模调查结果做一次调低。于是,使这两个指标均产

生了跳跃性的变化。根据六普模拟结果和统计公报这两种年龄结构指标的变化特点可以看出,以往平常年份统计公报的少儿比例倾向于偏高,而劳动年龄人口比例则倾向于偏低。这种情况表明,人口变动调查的主要漏报特点并不像通常想像的那样,主要漏报的其实并不是低龄人口,而是劳动年龄人口,很可能是青壮年流动人口。

（六）长期以来总和生育率一直很低

种种迹象表明,中国生育率的发布好像并不是由国家统计局负责,而是由国家人口计生委负责。长期以来,国家人口计生委将中国的总和生育率固定在 1.8 的口径上,成了一道多年不变的风景线。但是,以往多次全国人口调查数据直接统计的生育水平都远远低于总和生育率 1.8 的水平,基本处于 1.3~1.5 之间。这些人口调查既包括国家统计局进行的历年人口变动调查^①,也包括国家人口计生委自己开展的多次人口调查。

六普主要数据公布以后,一些人口学者曾根据六普主要数据做出经验判断或生育率匡算,已经得出中国生育率确实很低的结论。而本研究用模拟预测方法则能将比较分析的视野扩展到以往 20 年。图 10 提供了不同来源总和生育率的比较。其中,按六普结果模拟的总和生育率十几年都处于 1.5 以下的低水平,不少年份甚至连 1.4 都不到。这就表明,六普结果确认了以往多次人口调查已经揭示过的极低生育水平。相比之下,政府主管部门长期坚持的总和生育率 1.8 的口径则极为脱离实际。

至于图 10 中 2006 年全国人口和计划生育调查表明 2005 年前后的总和生育率大幅度回升,既没有得到同期或以后的其他人口调查的印证,而且又一次被六普结果所否定。笔者(2009)曾经专门撰文指出,2006 年的调查结果之所以异常是因为

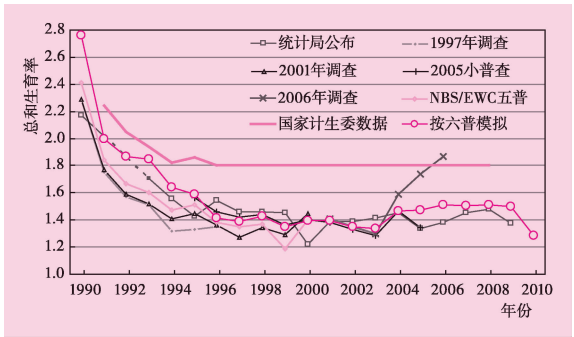


图 10 1990~2010 年总和生育率的比较

注:(1)统计局公布:根据历年人口变动调查的年龄别生育率(国家统计局人口和就业统计司,1990~2010)计算。其中逢全国人口普查或全国 1%人口抽样调查的年份为普查或小普查的结果。1991、1992、1993 年人口变动调查生育率未公开,由两端统计内插推算。(2)1997 年调查:郭志刚(2000)根据 1997 年全国人口与生殖健康调查数据计算。(3)2001 年调查:丁峻峰(2003)根据 2001 年全国人口与生殖健康调查数据计算。(4)2005 年小普查:郭志刚(2008)根据 2005 年全国 1%人口抽样调查样本数据估计的调整值。(5)2006 年调查:引自张维庆等(2008):《2006 年全国人口和计划生育调查数据集》。其中 2006 年数字引自该调查主要数据公报(国家人口计生委法规司,2008)。(6)NBS/EWC 五普:中国国家统计局、美国东西方中心(2007)根据中国 2000 年人口普查数据的估计。(7)国家人口计生委:引自国家人口计生委发展规划与信息司、中国人口与发展研究中心(2009)。(8)按六普模拟:本研究在 1990 年人口基数上模拟推算至六普公布人口“金字塔”的推计人数。

① 除 1991、1992 和 1993 年外,每年的《中国人口(和就业)统计年鉴》都公布人口变动调查的年龄别生育率,很可能是由于生育率骤降而感到没有把握。将各年龄生育率加总便能计算出相应的总和生育率。

其调查样本偏差极大所致。依托这样一个样本严重有偏的调查结果,所谓“生育反弹、形势严峻”之风曾经一度劲吹,以至党中央国务院发出的决定中都提出要“千方百计稳定低生育水平”。但后来事态发展却表明并没有“生育反弹”,而人口老龄化却先期而至了。然而这股风在当时的确大大助长了对中国生育水平的错误认识,对判断人口形势和人口决策产生了误导,危害很大。令人遗憾的是,这个问题一直被回避,并未得到应有的正本清源。

此外,笔者还想强调图 10 中所示以往总和生育率变化的动态特点。本来这些特点已经体现在以往的调查结果中了,但按六普模拟的结果则表现得更为清晰:20 年来,中国总和生育率的变化动态并不是一个徐徐下降的线性过程。在 20 世纪 90 年代初,总和生育率急剧下降。1996~2003 年最低,仅为 1.4 左右。最近几年略有回升,也不过只有 1.5 的水平。这个下凹的非线性特征在以往研究中也往往被忽视,结果是采用两点确定的简单线性假设也会自然高估生育率和出生水平。

总之,六普结果再次表明,中国已经确凿无疑地进入了低生育水平的时代。中国的总和生育率不仅远低于更替水平,而且生育率之低远超以往的认识和想象。

四、小 结

本研究根据六普人口金字塔测算的年龄别人口数,采用预测模拟方法推算了以往 20 年中国人口进程及其相应指标。本研究虽然是探测性的,而且在比较分析中以六普结果为标准,并未对六普数据本身的质量加以研讨,但基本结论仍然具有重要的启示和参考价值。

中国人口进入低生育水平的 20 年来,人口统计研究陷入一种迷茫状态,思想认识严重落后于实际,失去了对人口进程的整体把握。本文通过对比六普结果与以往人口统计之间的差异,旨在说明人口统计迷茫的根本原因既不是数据质量问题,也不是技术方法问题,而是研究背后的思想认识出了问题。回顾人口统计研究的经验教训,以往对人口调查及其统计结果没有给予应有的充分尊重,对统计数据的调整过于主观随意,结果陷入了一个人口统计怪圈:一方面不断开展人口调查,而另一方面又不断用过分的统计调整来否定这些调查结论。本研究的比较和分析表明,多年来对统计数据所做的缺乏科学根据的过分调整甚至比不调整的问题更大,而且这种过分调整导致更大的思想混乱,更加误导人口形势的判断和应对决策的制定。

历史的经验值得注意。2000 年五普以后,人口统计的质量几乎遭到彻底否定,很多研究都依据以往年份统计公报的出生数来论证五普数据质量极差。其实,不论五普数据质量有什么问题,毕竟它来自于实际调查,用人为大幅度调整过的以往公报统计来证明直接调查结果有误首先在逻辑和方法上就是荒谬的。更有一些人口统计研究居然援引“人口学多数”或者“联合国估计”来作为自己的论据,更是研究上的奇谈怪论。

六普以后,人口学研究是否会再次步入那个人口统计怪圈,仍然值得警惕。俗话说,“吃一堑长一智”,真正可怕的并不是犯错,而是长期屡犯又坚持不改。本文并不是主张六普质

量没有问题,而是主张人口学研究应该充分、深刻地总结和汲取以往研究实践中的经验教训,坚持实事求是的原则,充分尊重调查结果,因为它是社会科学中体现实践检验的主要途径。对于统计调整要极为慎重。即使需要统计调整,也应当首先公布原始汇总统计,然后再谈统计调整。对人口统计数据的调整必须严格遵循科学原则,统计调整的根据必须经过严格的论证,统计调整的过程要公开化和透明化。

有人说,统计就是数字化的历史。人口统计本身既是在书写人口发展的历史,也是在书写人口研究的历史。因此,我们更应当以史为鉴,不唯上,不唯书,只唯实,将人口统计研究工作兢兢业业地做好,使人口统计经得起实践和历史的检验。

参考文献:

1. 蔡泳(2009):《教育统计真的是估计生育水平的黄金标准吗?》,载于《人口与发展论坛:再论生育水平》,《人口研究》,第4期。
2. 蔡泳(2011):《六普分年龄性别人数》,美国北卡罗来纳大学网(<http://www.unc.edu/~caiyong/papers/sex.age.2010.pdf>),10月18日。
3. 丁峻峰(2003):《浅析中国1991~2000年生育模式变化对生育水平的影响》,《人口研究》,第2期。
4. 国家人口和计划生育委员会(2006):《全国“十一五”人口和计划生育事业发展规划》,中国人口网,12月28日。
5. 国家人口和计划生育委员会发展规划司(2007):《2006年全国人口和计划生育抽样调查主要数据公报》,中国人口网。
6. 国家人口计生委发展规划与信息司、中国人口与发展研究中心编(2010):《人口和计划生育常用数据手册(2009)》,中国人口出版社。
7. 国务院第六次全国人口普查办公室、国家统计局人口和就业统计司编(2011):《2010年第六次全国人口普查主要数据》,中国统计出版社。
8. 国家统计局人口(和就业)统计司(1990~2010):《中国人口(和就业)统计年鉴》,中国统计出版社,相应年份。
9. 郭志刚(2000):《从近年来的时期生育行为看终身生育水平》,《人口研究》,第1期。
10. 郭志刚(2004):《中国1990年代生育水平的研究与讨论》,《人口研究》,第2期。
11. 郭志刚(2008):《2005年全国1%人口抽样调查的生育数据分析》,载于国家统计局人口就业统计司编:《2005年全国1%人口抽样调查课题论文集》,中国统计出版社。
12. 郭志刚(2009):《近年生育率显著“回升”的由来——对2006年全国人口和计划生育调查的评价研究》,《中国人口科学》,第2期。
13. 郭志刚(2010a):《中国的低生育率与被忽略的人口风险》,《国际经济评论》,第6期。
14. 郭志刚(2010b):《中国的低生育水平及相关人口研究问题》,《学海》,第1期。
15. 教育部教育管理信息中心(2006):《戴井岗同志在2006年教育统计工作会议上的讲话》(<http://emic.moe.edu.cn/edoas2/website18/level3.jsp?id=1170053429486064>),12月12日。
16. 张维庆主编(2008):《2006年全国人口和计划生育调查数据集》,中国人口出版社。
17. 中国国家统计局、美国东西方中心编(2007):《中国各省生育率估计:1975~2000》,中国统计出版社。
18. 庄亚儿、张丽萍编(2003):《1990年以来中国常用人口数据集》,中国人口出版社。

(责任编辑:朱犁)