
基于家庭收入分布的地区基尼系数的 测算及其城乡分解

段景辉 陈建宝*

内容提要 本文利用《中国城乡人民生活综合调查》(2004年)的有关家庭收入抽样调查数据,首次考察发现全国和各地区城乡家庭人均收入的对数服从由 Pareto分布、正态分布和指数分布构成的混合分布。与通常的利用分组数据计算基尼系数不同,本文综合了全部抽样家庭的人均收入信息,应用分布函数法对全国、各省(市、自治区)以及东、中、西部地区的城镇基尼系数、农村基尼系数和城乡混合基尼系数进行了测算,进一步对城乡混合基尼系数进行城乡分解,得到了城镇和农村内部收入差距以及城乡收入差距对混合基尼系数的贡献大小。研究结果对有关政府部门制定相关的居民收入分配政策提供了依据。

关键词 地区差距 收入分布 基尼系数

一 引言

自1978年中国实行改革开放以来,中国经济在取得快速发展的同时,居民收入差距也在不断扩大,社会不公平现象凸现,而这种不公平性尤其体现在城乡之间和区域之间。收入差距过大不仅影响经济效率,更重要的是会影响社会的稳定,带来一系列

* 段景辉:厦门大学经济学院计划统计系 361005 电子信箱: jhduan001@163.com;陈建宝:厦门大学宏观经济研究中心研究员。

本文获国家社科重大基金项目“扩大内需的宏观经济政策研究”(08&ZD034)、教育部人文社科重点研究基地基金项目“我国地区间收入分配差异与劳动力转移的经济增长效应分析”(07JJD790145)和福建省社会科学规划研究项目“我国城乡收入差异问题研究”(2009b051)的资助。同时,本文作者感谢匿名评审人提出的有益建议以及程永宏在计算方面提供的帮助。

世界经济 * 2010年第 期 · 100 ·

严峻的社会问题,有悖于中国建设和谐社会的目标。因此,居民收入差距保持在适当水平非常重要和紧迫,对其客观测量亦显得十分重要。

常见的地区收入差距分析方法主要有以下三种:一是采用加权变异系数法,对地区收入差距的产业或部门构成进行分解。如万广华(1998)、覃成林和李二玲(2002)、林毅夫和刘培林(2003)。二是采用泰尔系数和广义熵系数法,对地区收入差距的地理构成进行分解。如李实和赵人伟(1999)、孙靖和黄海滨(2007)。三是计算基尼系数并对其进行分解。如王祖祥(2006)利用分组数据对中国中部6省的基尼系数进行估算,发现中部6省农村和城镇内部的基尼系数都不大,但是城乡加总的基尼系数都很大;陈昌兵(2007)利用非等分数据计算了全国21个省的基尼系数,并利用非参数计量中的分布函数估计方法分析了各地区城乡军民基尼系数的变化特征;王云飞(2007)利用Dagum(1997)提出的分解方法,计算了中国东、中、西部以及三地区之间的基尼系数和贡献度,发现中国地区间的基尼系数贡献度占地区总基尼系数的比重达到80%,东部各省之间的差距有缩小趋势,而西部各省的差距正在不断扩大。综合来看,虽然已有的研究成果揭示了中国地区收入差距的规律性问题,但是也存在一些不足:(1)多数成果都是研究整个区域的差异,并没有区分区域内农村和城镇的差异情况,事实上,由于目前中国城乡居民收入差距悬殊,将二者混在一起计算的区域差异系数并不能反映区域间贫富的实际情况。(2)多数研究采用的数据都是分组数据,利用微观数据的不多,因此计算的区域差异系数并不精确,不能准确反映区域间和区域内的收入差距情况。针对以上问题,本文拟利用中国社会综合调查开放数据库(CGSS)中21个省、3个自治区和4个直辖市的2004年的微观数据对全国和这28个地区的城镇内部、农村内部和城乡混合基尼系数进行计算并分解,并进一步将这28个省(市、自治区)分为东、中、西三个地区,分别对这三个地区的城镇内部、农村内部和城乡混合基尼系数进行计算并分解。

本文采用分布函数法测算基尼系数。分布函数法是基于对指标的概率密度函数或概率分布函数的假设,估计其分布参数,然后利用洛仑兹曲线对基尼系数进行估计,如McDonald(1995)在总结已有的各类分布函数的基础上提出广义贝塔分布函数(GB)和幂函数形式的广义贝塔分布函数(EGB),并通过分析说明此函数可以用于描述居民收入分布和金融变量的分布等。成邦文(2005)通过对洛仑兹曲线和基尼系数的研究表明大量社会经济规模指标服从对数正态分布;王海港(2006)用帕累托分布拟合中国1988年和1995年两年的居民收入数据,研究中国居民收入分配的格局。国内不少学者曾分别引入过伽马分布函数、帕累托分布函数和正态分布函数等对实际数

据进行拟合,效果都不甚理想,主要是由收入分布两端拟合效果不佳引起的。据我们所知,目前还没有一种较理想的将中国居民家庭收入拟合得很好的分布。为此,本文在这方面做了有益的尝试。

基尼系数的分解可以按不同收入群体或不同收入来源的组成部分来进行。按收入群体分解可以了解社会各个群体的收入不平等是如何影响社会总体的,按收入来源分解可以探索各个收入来源组成部分的不平等是如何影响收入不平等的。如 Sundrum (1990)将收入群体分为“穷人”和“富人”两个群体,并对混合群体的基尼系数进行分解;Cowell(2000)指出混合基尼系数在不同群体间进行分解时,除了各群体内部差距外,还包括群体间差距和交叉项;董静和李子奈(2004)在假设“城镇和农村居民的收入服从正态分布”的条件下,对城乡混合基尼系数在“城镇居民”和“农村居民”两个群体下进行分解;程永宏(2007)在利用逻辑斯蒂函数估算城乡混合基尼系数后,对城乡两个群体进行分解混合基尼系数。由于中国是一个典型的二元经济社会,居民收入数据天然地分成了城镇家庭居民收入和农村家庭居民收入两套数据,因此,本文在计算全国和各地区城乡混合基尼系数的基础上,对“城镇家庭”和“农村家庭”两个收入群体进行分解。

本文尝试在前人研究的成果基础上,利用中国社会综合调查开放数据库(CGSS)中2004年的微观数据,将全国和各地区的城镇和农村家庭人均收入分为高、中、低三个收入层次,利用分布函数法寻找合适的函数对全国和各地区城乡家庭人均收入的每个收入层次进行拟合,构造更加准确的城乡居民收入分布函数,从而推导全国和各地区的城乡混合基尼系数并在城镇和农村两个群体分解,研究城镇和农村居民收入的不平等是如何影响收入差距的。

二 城镇和农村家庭收入的分布特征

本文数据来源于中国社会综合调查开放数据库(CGSS)中的《中国城乡居民生活综合调查》(2004年)的有关家庭收入抽样调查数据。该数据采用分层四阶段不等概率抽样方法,以家庭为样本单位进行调查。本文所使用的样本包括21个省、3个自治区和4个直辖市共9724个家庭样本单位。其中东部地区包括北京、上海、天津、广东、福建、浙江、山东、辽宁、江苏9个省(市);中部地区包括山西、陕西、江西、吉林、湖北、湖南、黑龙江、河南、河北、海南、安徽、重庆12个省(市);西部地区包括云南、贵州、甘肃、四川、内蒙古、广西、新疆7个省(自治区)。

为了描述 2004 年中国城镇和农村家庭人均收入分布特征,将“家庭人均收入”这一指标做取对数处理后,经计算得到表 1 和表 2。

表 1 2004 年全国、各省(市、自治区)家庭人均对数收入分布特征

地区	城乡	家庭数目	均值	标准差	偏度	峰度	JB 正态性 检验 P 值
全国	城镇	5665	3.7022	0.3193	-0.3844	0.5826	0.0001
	农村	4059	3.1917	0.4042	-0.4529	0.2946	0.0001
北京	城镇	383	4.0038	0.3097	-0.1972	0.8336	0.0001
上海	城镇	382	4.1019	0.3239	0.5363	2.7703	0.0001
天津	城镇	400	3.8856	0.2705	0.7663	4.1311	0.0001
广东	城镇	328	3.8655	0.4197	0.6161	3.6601	0.0001
	农村	178	3.1629	0.4551	-0.2576	0.1319	0.0001
福建	城镇	200	3.9708	0.3231	0.4644	0.2241	0.0001
	农村	103	3.1378	0.4011	-0.3661	0.1393	0.0001
浙江	城镇	106	4.0186	0.3449	-0.0318	1.1074	0.0001
	农村	163	3.4011	0.5157	-0.2774	0.2269	0.0001
山东	城镇	317	3.8111	0.3407	-0.1376	0.1376	0.0001
	农村	329	3.3563	0.3308	-0.3137	1.5943	0.0001
辽宁	城镇	224	3.7125	0.3337	0.5266	1.6456	0.0001
	农村	188	3.2358	0.3382	-0.0641	0.4974	0.0001
江苏	城镇	282	3.8896	0.3898	-0.0173	0.5488	0.0001
	农村	328	3.3786	0.3844	-0.1556	0.5731	0.0001
山西	城镇	76	3.5315	0.2572	-0.2436	0.3368	0.0001
	农村	81	3.1342	0.3419	-0.7395	2.1806	0.0001
陕西	城镇	207	3.7654	0.2914	0.3517	2.2122	0.0001
	农村	124	3.0659	0.3135	-0.0256	0.1661	0.0001
江西	城镇	91	3.3962	0.3997	0.6389	0.5062	0.0001
	农村	104	2.9499	0.3907	-0.5708	1.6083	0.0001
吉林	城镇	129	3.4196	0.3671	0.0668	0.8679	0.0001
	农村	42	3.1894	0.2502	1.0398	0.8426	0.0001
湖北	城镇	241	3.6876	0.3404	-0.5059	0.9827	0.0001
	农村	247	3.2076	0.2769	-0.0509	0.4793	0.0001
湖南	城镇	263	3.7969	0.3671	0.3016	0.8613	0.0001
	农村	204	3.3537	0.2849	-0.1672	0.1586	0.0001

计算“基于家庭人均收入的基尼系数”更符合中国社会的文化背景,我们把“家庭”看作不可分的个体,以“家庭人均收入”作为该个体的收入,这样处理更具有现实意义。

基于家庭收入分布的地区基尼系数的测算及其城乡分解

地区	城乡	家庭数目	均值	标准差	偏度	峰度	JB 正态性 检验 P值
黑龙江	城镇	251	3.4741	0.3675	0.1169	0.1432	0.0001
	农村	60	3.3555	0.4283	0.1837	0.2501	0.0001
河南	城镇	261	3.5973	0.3618	-0.5825	4.2893	0.0001
	农村	328	2.9686	0.3588	-0.5051	0.7722	0.0001
河北	城镇	171	3.8259	0.3603	-0.4659	3.3312	0.0001
	农村	233	3.2599	0.3208	-0.6131	0.7945	0.0001
海南	城镇	52	3.4531	0.3725	-0.0142	0.7192	0.0001
	农村	12	3.1713	0.2205	-0.2156	0.5265	0.0001
安徽	城镇	259	3.7576	0.3899	0.2301	1.5554	0.0001
	农村	263	3.2779	0.3357	0.0616	0.1811	0.0001
重庆	城镇	42	3.5646	0.3132	0.4556	1.4482	0.0001
	农村	43	3.1665	0.3564	-0.3605	0.1365	0.0001
云南	城镇	150	3.5324	0.3939	-0.3516	0.0804	0.0001
	农村	158	3.1709	0.3198	0.0017	0.3285	0.0001
贵州	城镇	104	3.7243	0.2951	-0.3502	0.0629	0.0001
	农村	171	3.2716	0.2622	0.2427	0.2913	0.0001
甘肃	城镇	147	3.6506	0.3597	-0.8064	1.5732	0.0001
	农村	100	3.2444	0.3992	-0.2487	0.0279	0.0001
四川	城镇	258	3.6523	0.3815	0.0206	1.1923	0.0001
	农村	390	3.0391	0.3107	0.2581	0.7106	0.0001
内蒙古	城镇	85	3.6822	0.3427	0.4231	0.2102	0.0001
	农村	80	3.2206	0.2532	-0.1726	0.3043	0.0001
广西	城镇	180	3.6563	0.4146	0.5348	3.7609	0.0001
	农村	130	2.8807	0.3881	0.1474	0.1662	0.0001
新疆	城镇	76	3.2334	0.4677	0.3769	1.1062	0.0001

说明:缺少北京、天津、上海、新疆的农村家庭收入数据。

本文数据来源:根据中国社会综合调查开放数据库(CGSS)的数据计算所得,下同。

表 2 2004年中国东、中、西部地区家庭人均对数收入分布特征

地区	城乡	家庭数目	均值	标准差	偏度	峰度	JB 正态性 检验 P值
东部地区	城镇	2622	3.9203	0.3564	-0.2084	1.7429	0.0001
	农村	1289	3.3059	0.4069	-0.0237	0.6665	0.0001
中部地区	城镇	1794	3.6364	0.3911	-0.0355	1.0714	0.0001
	农村	1574	3.1856	0.3637	-0.3513	0.9305	0.0001
西部地区	城镇	1249	3.6366	0.3858	-0.1909	1.3563	0.0001
	农村	1196	3.1092	0.3369	-0.0991	0.2539	0.0001

如表 1 和表 2 所示,全国 28 个省(市、自治区)和东、中、西部三大地区的城镇家庭人均对数收入的均值都大于农村家庭人均对数收入的均值,说明城镇家庭人均收入要高于农村家庭人均收入;城镇和农村家庭人均对数收入的 JB 正态性检验的 P 值都小于 0.05,说明城镇和农村家庭的人均对数收入分布都没有通过正态性检验,不符合正态分布。偏度和峰度是两个无量纲的数值,常用来描述一组数据的分布形状。正态分布的 g_1 和 g_2 均为 0,若 $g_1 > 0$,则称 x 的分布是正偏(或右偏)的;若 $g_1 < 0$,则称 x 的分布是负偏(或左偏)的, $|g_1|$ 越大,说明分布偏斜得越厉害;若 $g_2 > 0$,则说明随机变量 x 分布的尾部比正态分布的尾部粗,并且 g_2 值越大,倾向认为尾部越粗;若 $g_2 < 0$,则说明 x 分布的尾部比正态分布的尾部细,且 $|g_2|$ 值越大,倾向认为尾部越细。2004 年全国以及东、中、西部三大地区的城镇和农村家庭人均对数收入分布的偏度 $g_1 < 0$,而峰度 $g_2 > 0$,说明全国以及三大地区的城镇和农村家庭人均对数收入的分布都是左偏厚尾的;中国 28 个省(市、自治区)中浙江、山东、江苏、山西、湖北、河南、河北和海南省的城镇和农村家庭人均对数收入分布都是左偏厚尾的,其他各省(市、自治区)大部分农村家庭的人均对数收入分布是左偏厚尾的,而城镇家庭人均对数收入分布是右偏厚尾的。所以,2004 年城镇和农村的家庭人均对数收入分布不可能用单一的分布函数来描述,我们将依据城镇和农村家庭人均对数收入分布的特点寻找适合低、中、高三收入层次的分布函数,以此来拟合整个城镇和农村家庭的人均收入分布函数。

在国外,研究收入分布规律已经有很长的历史,意大利经济学家 Pareto 是研究这一问题的第一人,他发现收入分布服从幂率分布。随后的大量研究结果表明高收入阶层的收入分布可以用幂率分布描述,而中低收入阶层的收入分布可以用对数正态分布或指数分布描述。本文拟采用这一思想,对全国和各地区城镇和农村家庭人均收入分布做一个拟合,经过反复验证高收入阶层可用帕累托分布拟合,中等收入分布可用正态分布拟合,低收入分布可用指数分布拟合。

帕累托分布是收入分配理论中一种重要的分布(Cowell, 2000)。帕累托分布的累积分布函数为:

$$F(x) = 1 - \left(\frac{a}{x}\right)^b, \quad x > 0, \quad a > 1$$

其中 a 是形状参数, b 是尺度参数(门限参数),累积分布函数度量收入小于等于

偏度定义: $g_1 = E[x - E(x)]^3 / [Var(x)]^{1/2}$ 。

峰度定义: $g_2 = E[x - E(x)]^4 / [Var(x)]^2 - 3$ 。

本文的“人均收入”即为“人均对数收入”的简称,下同。

的相对人口比例,我们将这种两参数帕累托分布记为 $Pareto(\alpha, \beta)$ 。因为帕累托分布的参数严格为正,仅有右尾,而且尾部比对数正态分布厚,能很好地拟合高收入阶层。在分析过程中,如果所得模型的平均方差(Average Variance,简称为 AV)远远超过均方误差(Mean Square Error,简称为 MSE),则数据的拟合效果好(彭昭英,2000)。

正态分布的概率密度函数为:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

其中 μ 和 σ 分别是随机变量 X 的均值与标准差,我们记为 $N(\mu, \sigma^2)$ 。如果中等收入的城乡家庭数目小于 2000 户,则用 W 统计量检验中等收入阶层的对数收入分布是否符合正态分布。根据样本数据分布的形状和正态分布相比较,得出一个数值 P ($0 < P < 1$, 即实际的显著性水平)来描述“数据分布服从正态分布”的拟合程度。如果 W 统计量的 P 值大于 0.05,则样本分布通过正态性检验。

指数分布是一种连续概率分布,由于城乡家庭人均收入大于等于 0,因此它的累积分布函数为:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

其中 λ 是指数分布参数,我们记指数分布为 $Exp(\lambda)$ 。我们对所估计的参数 λ 进行 P 检验,如果参数估计的 P 值小于 0.05,则拒绝“参数 $\lambda = 0$ 的原假设,参数 λ 的估计值在 5%水平下显著。

三 基于分布函数的基尼系数的测算及分解方法

(一) 单一总体基尼系数的测算

下面以城镇家庭为例计算其基尼系数,农村家庭基尼系数的测算方法类似。

在城镇家庭这一总体中,家庭人均收入可以被近似的看作一个连续型变量 I ,因此,城镇家庭人均收入的分配状况可以用连续性随机变量 I 的概率分布函数表示,即:

$$P\{I < x\} = F_u(x) \quad (1)$$

表示城镇家庭人均收入不大于 x 的家庭数目占全部城镇家庭数目的比重是

在观察值的样本数 $n < 2000$ 时, Wilk - Shapiro 建议用 W 统计量: $W = \frac{[\sum_{i=1}^n (a_i - \bar{a})(x_i - \bar{x})]^2}{[\sum_{i=1}^n (a_i - \bar{a})^2][\sum_{i=1}^n (x_i - \bar{x})^2]}$ 在原假设 H_0 (数据分布服从正态分布)为真时, W 统计量标准化后渐近分布为标准正态分布,取值接近于 1。

$F_u(x)$ 。将城镇家庭人均收入分为低等收入、中等收入和高等收入三个层次,其中低、中、高三个收入层次的家庭数目占整个城镇家庭数目的比例分别为 $\frac{F_1}{W}$, $\frac{F_2}{W}$, $\frac{F_3}{W}$ 。则可设城镇家庭人均收入分布函数为 $F_u = F_1 + F_2 + F_3$ 其中, $\frac{F_1}{W} + \frac{F_2}{W} + \frac{F_3}{W} = 1$, F_1, F_2, F_3 分别表示城镇家庭低、中、高三个收入层次的家庭人均收入分布函数。

基尼系数是根据洛伦兹曲线构造的统计指数,从洛伦兹曲线的定义可知,洛伦兹曲线与城镇家庭人均收入分布函数间存在确定的关系,厉以宁和秦宛顺(1997)与程永宏和糜仲春(1998)曾以不同方式推导出这一关系:洛伦兹曲线上任意点 D 的横坐标 P 与纵坐标 L 为:

$$\begin{cases} P(D) = F_u(x) \\ L(D) = \frac{N}{W} \int_0^x f_u(t) dt \end{cases} \quad (2)$$

其中, N 表示城镇家庭总数目; W 表示城镇家庭人均总收入;横坐标 $P(D)$ 意味着城镇家庭人均收入不大于 x 的家庭数目(设为 n)占全部城镇家庭数目的比例;纵坐标 $L(D)$ 意味着这 n 个城镇家庭的累积收入占整个城镇家庭人均总收入的比重。

利用洛伦兹曲线的参数方程,可知洛伦兹曲线与 y 轴在 $y \in [0, 1]$ 上围成的面积 S 为:

$$S = \int_0^x F_u(x) dL(F_u(x))$$

与 $L(x) = 0$ 相对应的数量 x 为 0,与 $L(x) = 1$ 相对应的数量 x 就是城镇家庭人均收入中的最高收入者(用 X 表示)。利用(2)式,分部积分得到: $S =$

$$\frac{N}{W} \left(\frac{1}{2} X - \frac{1}{2} \int_0^X F_u^2(x) dx \right)$$

。则洛伦兹曲线与直线 $y = x$ 围成的面积 A 为: $A = S - \frac{1}{2} =$

$$\frac{N}{W} \left(\frac{1}{2} X - \frac{1}{2} \int_0^X F_u^2(x) dx \right) - \frac{1}{2}$$

由基尼系数的定义可知,城镇家庭基尼系数 G_u 可以用 A 表示为:

$$G_u = 2A = \frac{N}{W} \left(X - \int_0^X F_u^2(x) dx \right) - 1 \quad (3)$$

利用以上公式计算城镇家庭基尼系数需要解决一个问题:城镇家庭人均收入分布函数的具体形式。本文将在经验分析部分分别拟合各地区城镇和农村家庭低、中、高

在实际计算中,积分的下限为城镇家庭人均收入最低者。

三个层次的收入分布函数,从而得到城乡家庭混合人均收入分布函数。

(二)城乡家庭混合基尼系数的测算

根据单一总体基尼系数的测算方法以及城乡家庭人均收入调查数据,可以分别拟合城镇和农村家庭人均收入分布函数,从而获得城乡混合收入分布函数,进一步推算城乡混合基尼系数。

设城镇家庭数目为 N_u ,城镇家庭中的人均最高收入为 X_u ;农村家庭数目为 N_r ,农村家庭中的人均最高收入为 X_r ;某一地区居民家庭数目为 N_m ,显然这一地区居民家庭总数目为 $N_m = N_u + N_r$,这一地区居民家庭中的最高人均收入为 $X_m = \max(X_u, X_r)$ 。由(1)式可知,这一地区居民家庭人均收入可用随机变量 I_m 表示,相应的收入分布函数为 $F_m(x)$ 。则对于这一地区居民家庭人均收入来说,人均收入不大于给定 x 值的家庭绝对数目为 $M = N_u F_u(x) + N_r F_r(x)$,则相应的比重为: $M/N_m = [N_u F_u(x) + N_r F_r(x)]/N_m$ 。

由此得到这一地区城乡混合家庭人均收入分布函数 $F_m(x)$:

$$F_m(x) = P(I_m < x) = \frac{N_u F_u(x)}{N_m} + \frac{N_r F_r(x)}{N_m} \quad (4)$$

将这一分布函数(4)式代入(3)式,即可得到这一地区城乡家庭混合基尼系数 G_m :

$$G_m = \frac{N_m}{W_m} (X_m - \int_0^{X_m} F_m^2(x) dx) - 1 \quad (5)$$

进一步,因为这一地区居民家庭人均总收入 W_m 实际上就是累积收入在 $x = X_m$ 时的值,因此:

$$W_m = N_m \int_0^{X_m} x dF_m(x) = N_m (X_m - \int_0^{X_m} F_m(x) dx) \quad (6)$$

将(6)式代入(5)式得到:

$$G_m = \frac{X_m - \int_0^{X_m} \left(\frac{N_u F_u(x)}{N_m} + \frac{N_r F_r(x)}{N_m} \right)^2 dx}{X_m - \int_0^{X_m} \left(\frac{N_u F_u(x)}{N_m} + \frac{N_r F_r(x)}{N_m} \right) dx} - 1 \quad (7)$$

将(7)式进一步展开,并设 $X_u > X_r$,则 $X_m = \max(X_u, X_r) = X_u$,记:

在实际计算中,积分的下限为城乡家庭人均收入最低者。

$$\frac{N_u}{N_m} = \frac{N_r}{N_m} = 1 - \int_0^{X_u} F_u(x) dx = A_u, \int_0^{X_r} F_r(x) dx = A_r \quad (8)$$

$$\int_0^{X_u} F_u^2(x) dx = B_u, \int_0^{X_r} F_r^2(x) dx = B_r, \int_0^{X_u} F_u(x) F_r(x) dx = C$$

又因为 $x > X_r$ 时, $F_r(x) = 1$, 有 $\int_{X_r}^{X_u} F_r(x) dx = X_u - X_r$, $\int_{X_r}^{X_u} F_r^2(x) dx = X_u - X_r$, 所以在:

$$\begin{cases} \int_0^{X_u} F_r(x) dx = \int_0^{X_r} F_r(x) dx + \int_{X_r}^{X_u} F_r(x) dx = A_r + X_u - X_r \\ \int_0^{X_u} F_r^2(x) dx = \int_0^{X_r} F_r^2(x) dx + \int_{X_r}^{X_u} F_r^2(x) dx = B_r + X_u - X_r \end{cases} \quad (9)$$

把 (8) 和 (9) 式代入到 (7) 式的展开式中, 整理后得到城乡混合基尼系数的另一形式:

$$G_m = \frac{X_u - [\int_0^{X_u} B_u + (1 - \frac{N_r}{N_m})^2 B_r + 2(1 - \frac{N_r}{N_m})C + (1 - \frac{N_r}{N_m})^2 (X_u - X_r)]}{X_u - [\int_0^{X_u} A_u + (1 - \frac{N_r}{N_m})A_r + (1 - \frac{N_r}{N_m})(X_u - X_r)]} - 1 \quad (10)$$

以上公式涉及诸多积分计算, 形式复杂, 但是利用计算机软件 Matlab 强大的数值计算功能, 很好地解决了这一计算难题。更可贵的是, 上述公式 (10) 可以进一步分解成具有明确经济意义的简单形式, 为这一地区城乡家庭混合基尼系数的分解提供了可靠的理论依据。

(三) 城乡家庭混合基尼系数的分解

基尼系数的分解一直是相关文献研究的热点难点问题, 根据以上城乡家庭混合基尼系数的测算方法, 可以进一步对这一地区的城乡家庭混合基尼系数进行分解。

设这一地区城乡家庭人均收入的均值为 E_m , 根据式 (8) 可将式 (6) 表示为:

$$E_m = \frac{W_m}{N_m} \int_0^{X_u} x dF_m(x) = X_m - \int_0^{X_u} F_m(x) dx$$

$$= X_m - [\int_0^{X_u} A_u + (1 - \frac{N_r}{N_m})A_r + (1 - \frac{N_r}{N_m})(X_u - X_r)]$$

同理, 根据式 (3) 和式 (8), 城镇家庭人均收入的均值和农村家庭人均收入的均值分别为:

$$E_u = (A_u - B_u) / G_u, \quad E_r = (A_r - B_r) / G_r \quad (12)$$

将式 (11)、(12) 代入式 (10), 重新整理组合得到:

$$G_m = \frac{E_u}{E_m} G_u + \frac{(1 - \alpha) E_r}{E_m} G_r + \frac{(1 - \alpha) (B_u + B_r - 2C + X_u - X_r)}{E_m} \quad (13)$$

式 (13) 中 $B_u + B_r - 2C$ 可以表示为:

$$\begin{aligned} B_u + B_r - 2C &= \int_0^{x_m} [F_u^2(x) + F_r^2(x) - 2F_u(x)F_r(x)] dx - \int_{x_r}^{x_m} F_r^2(x) dx \\ &= \int_0^{x_m} [F_u(x) - F_r(x)]^2 dx - (X_u - X_r) \end{aligned}$$

我们记 $\int_0^{x_m} [F_u(x) - F_r(x)]^2 dx = D$, 则得到这一地区的城乡家庭混合基尼系数的分解式:

$$G_m = \frac{E_u}{E_m} G_u + \frac{(1 - \alpha) E_r}{E_m} G_r + \frac{(1 - \alpha)}{E_m} D \quad (14)$$

以上城乡家庭混合基尼系数分解式有明确的经济意义: D 是度量城乡差距的优良指标, 更全面地反映了城乡差距, 被定义为“绝对城乡差距指数”, 而 D 去除家庭人均收入这一“量纲”的影响后, 可以记为: $G_d = D/E_m$, 被定义为“相对城乡差距”(程永宏, 2006)。 G_u, G_r 的权数分别表示城镇和农村家庭人均收入占总收入的份额, 表明城乡内部基尼系数对混合基尼系数的影响取决于城乡家庭人均收入所占全国家庭人均总收入份额的大小。(14) 式左边可以看作城乡家庭人均收入因为总体不平等造成的福利损失; 右边第一项可以看作由于农村内部不平等造成的福利损失, 右边第二项可以看作由于城市内部不平等造成的福利损失, 右边第三项可以看作由于城乡差距造成的福利损失。因此本文的分解方法揭示了这一地区城乡内部以及城乡之间的收入差距。

三 经验分析

(一) 城镇和农村家庭人均收入的分布函数拟合

首先确定全国和各地区城镇和农村家庭低、中、高收入阶层的分界线。利用“交叉验证法”(赵桂芹和栗芳, 2008) 得到帕累托分布的门槛值 x_g , 即城镇(农村)高收入家庭的最小值, 根据门槛值 x_g 所处位置可以计算得到城镇(农村)高收入家庭数目占整个城镇(农村)家庭数目的比例; 从城镇(农村)概率分布图中得到门槛值 x_g 的概率 P_g 以及概率最大的值 x_0 后, 根据正态分布的对称性, 以 x_0 为对称中心, 在 x_0 的左端

找到概率同样为 P_g 的数值 x_d , 以 x_d 所处位置作为中等收入和低收入阶层的分界点, 由此可以分别计算出城镇 (农村) 中、低收入家庭数目占整个城镇 (农村) 家庭数目的比例和。

接下来, 分别利用帕累托分布函数、对数正态分布函数和指数分布函数对全国和各地区城镇和农村家庭人均收入的高、中、低收入阶层进行拟合。经验证我们发现每个帕累托模型的平均方差远远超过均方误差, 数据的拟合效果好, 全国和各地区城镇和农村中等收入阶层的家庭人均对数收入数据都通过正态性检验, 指数分布模型的参数估计结果均在显著性水平 5% 上显著, 如表 3 和表 4 所示。因此, 可以用以上模型得到全国和各地区城镇和农村家庭各收入阶层的分布函数, 见表 5 和表 6。

由城镇家庭人均收入分布函数公式 $F_u = F_1 + F_2 + F_3$ 可分别得到各地区的城镇家庭人均收入分布函数 F_{ui} (i 表示地区, 下同)。同样, 由农村家庭人均收入分布函数 $F_r = F_1 + F_2 + F_3$ 可分别得到各地区农村家庭人均收入分布函数 F_{ri} 。进一步, 可以分别得到 2004 年各地区城乡混合家庭人均收入分布函数 F_{mi} , $F_{mi} = i F_{ui} + (1 - i) F_{ri}$, i 表示第 i 地区城镇家庭数目占该地区家庭总数目的比例。

表 3 2004 年全国、各省 (市、自治区) 城镇和农村各收入阶层的分布检验

地区	城镇					农村				
	高收入		中等收入		低收入	高收入		中等收入		低收入
	AV	MSE	W	P	P	AV	MSE	W	P	P
全国	259.67	0.0429	0.6981	0.3039	0.0001	489.18	0.0419	0.7224	0.3022	0.0001
北京	455.43	0.0476	0.6844	0.2859	0.0001	-	-	-	-	-
上海	499.39	0.0352	0.9381	0.2074	0.0001	-	-	-	-	-
天津	691.09	0.0729	0.8419	0.2847	0.0001	-	-	-	-	-
广东	438.73	0.0169	0.5491	0.1308	0.0001	341.43	0.0193	0.9399	0.2739	0.0001
福建	372.06	0.0825	0.7649	0.1726	0.0001	734.56	0.0831	0.7519	0.2527	0.0001
浙江	294.87	0.0939	0.9201	0.2938	0.0001	492.81	0.0442	0.6429	0.1862	0.0001
山东	318.29	0.0136	0.9724	0.1744	0.0001	339.05	0.0401	0.9291	0.1582	0.0001
辽宁	294.71	0.0392	0.9173	0.0827	0.0001	450.21	0.0157	0.6011	0.2729	0.0001
江苏	593.99	0.0471	0.9742	0.1732	0.0001	290.74	0.0291	0.5472	0.1528	0.0001
山西	371.94	0.0962	0.5028	0.1143	0.0500	239.87	0.0247	0.9022	0.2632	0.0500
陕西	448.32	0.1003	0.8391	0.3192	0.0001	229.37	0.0591	0.4493	0.0837	0.0001
江西	397.12	0.0582	0.7392	0.1745	0.0500	279.37	0.0482	0.6938	0.0739	0.0001
吉林	377.91	0.0385	0.6391	0.0928	0.0001	307.37	0.0639	0.5983	0.0691	0.0500
湖北	492.91	0.1958	0.7329	0.0837	0.0001	401.32	0.1201	0.5911	0.0594	0.0001

基于家庭收入分布的地区基尼系数的测算及其城乡分解

地区	城镇					农村				
	高收入		中等收入		低收入	高收入		中等收入		低收入
	<i>AV</i>	<i>MSE</i>	<i>W</i>	<i>P</i>	<i>P</i>	<i>AV</i>	<i>MSE</i>	<i>W</i>	<i>P</i>	<i>P</i>
湖南	510.32	0.0991	0.5918	0.1193	0.0001	311.92	0.0773	0.6118	0.1148	0.0001
黑龙江	338.58	0.0639	0.7111	0.1148	0.0001	271.83	0.0492	0.7028	0.1039	0.0500
河南	291.28	0.0927	0.5822	0.0739	0.0001	407.25	0.0591	0.5091	0.0837	0.0001
河北	381.21	0.0762	0.6291	0.2248	0.0001	421.04	0.0397	0.6003	0.0581	0.0001
海南	229.34	0.1772	0.4999	0.1473	0.0001	-	-	-	-	-
安徽	308.65	0.0837	0.5118	0.0639	0.0001	229.72	0.0859	0.5933	0.0596	0.0001
重庆	112.96	0.0999	0.7291	0.1175	0.0500	308.67	0.0715	0.7301	0.1739	0.0500
云南	307.11	0.0629	0.6882	0.1928	0.0001	307.49	0.0482	0.6229	0.1236	0.0001
贵州	296.48	0.0582	0.5338	0.0593	0.0001	317.92	0.0695	0.5582	0.0982	0.0001
甘肃	229.74	0.0691	0.5915	0.0601	0.0001	288.49	0.0299	0.5118	0.0974	0.0001
四川	288.41	0.0883	0.4829	0.0739	0.0001	217.47	0.0682	0.4928	0.0636	0.0001
内蒙古	231.54	0.0119	0.3981	0.1791	0.0500	239.45	0.0194	0.6049	0.0791	0.0500
广西	387.69	0.1029	0.5598	0.1208	0.0001	291.58	0.0501	0.5821	0.0575	0.0001
新疆	285.48	0.0721	0.7101	0.0728	0.0500	-	-	-	-	-

说明:缺少北京、天津、上海、新疆的农村家庭收入数据;海南省农村家庭样本太少,无法检验。

数据来源:中国社会综合调查开放数据库(CGSS)的数据计算所得,下同。

表 4 2004年东、中、西部地区城镇和农村各收入阶层的分布检验

地区	城镇					农村				
	高收入		中等收入		低收入	高收入		中等收入		低收入
	<i>AV</i>	<i>MSE</i>	<i>W</i>	<i>P</i>	<i>P</i>	<i>AV</i>	<i>MSE</i>	<i>W</i>	<i>P</i>	<i>P</i>
东部地区	489.43	0.0147	0.8744	0.0737	0.0001	430.11	0.0457	0.8911	0.0629	0.0001
中部地区	317.49	0.0342	0.5281	0.0911	0.0001	220.14	0.0591	0.7972	0.0828	0.0001
西部地区	221.09	0.0829	0.7919	0.0816	0.0001	259.87	0.0741	0.7092	0.0532	0.0001

表 5 2004年全国、各省(市、自治区)城镇和农村家庭各收入阶层的分布函数

地区	城镇			农村		
全国	=0.21	$Exp(0.0272)$		=0.23	$Exp(0.1826)$	
	=0.68	$N(3.821, 0.0217)$		=0.63	$N(3.2363, 0.0225)$	
	=0.11	$Pareto(6085, 18, 1.2238)$		=0.14	$Pareto(5494, 17, 1.0591)$	
北京	=0.15	$Exp(0.0721)$		-	-	
	=0.64	$N(4.1538, 0.0397)$		-	-	
	=0.21	$Pareto(4.1956, 1.2975)$		-	-	

地区	城镇		农村
	=0.14	<i>Exp</i> (0.0829)	-
上海	=0.64	<i>N</i> (4.1587, 0.0847)	-
	=0.22	<i>Pareto</i> (4.2050, 1.3819)	-
	=0.16	<i>Exp</i> (0.0739)	-
天津	=0.66	<i>N</i> (4.0219, 0.0821)	-
	=0.18	<i>Pareto</i> (4.1807, 1.2482)	-
	=0.13	<i>Exp</i> (0.0391)	=0.20
广东	=0.68	<i>N</i> (3.9626, 0.0256)	=0.65
	=0.19	<i>Pareto</i> (4.0089, 1.3491)	=0.15
	=0.23	<i>Exp</i> (0.0587)	=0.26
福建	=0.60	<i>N</i> (3.9441, 0.0491)	=0.64
	=0.17	<i>Pareto</i> (4.0380, 1.1490)	=0.10
	=0.14	<i>Exp</i> (0.0399)	=0.20
浙江	=0.71	<i>N</i> (4.1849, 0.0831)	=0.65
	=0.15	<i>Pareto</i> (4.2162, 1.0273)	=0.15
	=0.15	<i>Exp</i> (0.0394)	=0.13
山东	=0.70	<i>N</i> (3.9267, 0.0217)	=0.75
	=0.15	<i>Pareto</i> (4.0150, 1.2411)	=0.11
	=0.14	<i>Exp</i> (0.0183)	=0.23
辽宁	=0.72	<i>N</i> (3.8212, 0.0421)	=0.63
	=0.14	<i>Pareto</i> (3.9907, 1.2911)	=0.14
	=0.16	<i>Exp</i> (0.0344)	=0.20
江苏	=0.72	<i>N</i> (3.9626, 0.0126)	=0.67
	=0.12	<i>Pareto</i> (4.0089, 1.3192)	=0.13
	=0.19	<i>Exp</i> (0.0517)	=0.18
山西	=0.68	<i>N</i> (3.6441, 0.0151)	=0.69
	=0.13	<i>Pareto</i> (3.8030, 1.1611)	=0.13
	=0.24	<i>Exp</i> (0.0285)	=0.16
陕西	=0.61	<i>N</i> (3.8849, 0.0753)	=0.72
	=0.15	<i>Pareto</i> (3.9029, 1.0222)	=0.12
	=0.14	<i>Exp</i> (0.0597)	=0.19
江西	=0.75	<i>N</i> (3.4267, 0.0239)	=0.68
	=0.11	<i>Pareto</i> (3.5980, 1.2210)	=0.13
	=0.14	<i>Exp</i> (0.0433)	=0.24
吉林	=0.76	<i>N</i> (3.5212, 0.0511)	=0.66
	=0.10	<i>Pareto</i> (3.6307, 1.2966)	=0.10

基于家庭收入分布的地区基尼系数的测算及其城乡分解

地区	城镇	农村
湖北	=0.16 <i>Exp</i> (0.0321)	=0.14 <i>Exp</i> (0.0359)
	=0.72 <i>N</i> (3.7626, 0.0184)	=0.75 <i>N</i> (3.3849, 0.0742)
	=0.12 <i>Pareto</i> (3.8955, 1.3752)	=0.11 <i>Pareto</i> (3.4629, 1.0661)
湖南	=0.19 <i>Exp</i> (0.0517)	=0.14 <i>Exp</i> (0.0519)
	=0.68 <i>N</i> (3.8441, 0.0169)	=0.76 <i>N</i> (3.4422, 0.0153)
	=0.13 <i>Pareto</i> (3.5530, 1.1740)	=0.10 <i>Pareto</i> (3.5530, 1.1440)
黑龙江	=0.24 <i>Exp</i> (0.0299)	=0.16 <i>Exp</i> (0.0249)
	=0.62 <i>N</i> (3.4849, 0.0497)	=0.72 <i>N</i> (3.3899, 0.0543)
	=0.14 <i>Pareto</i> (3.5829, 1.0239)	=0.12 <i>Pareto</i> (3.4029, 1.0200)
河南	=0.14 <i>Exp</i> (0.0313)	=0.19 <i>Exp</i> (0.0579)
	=0.75 <i>N</i> (3.6267, 0.0262)	=0.68 <i>N</i> (3.0451, 0.0149)
	=0.11 <i>Pareto</i> (3.7750, 1.2190)	=0.13 <i>Pareto</i> (3.1130, 1.1620)
河北	=0.14 <i>Exp</i> (0.0133)	=0.24 <i>Exp</i> (0.0311)
	=0.76 <i>N</i> (3.9212, 0.0321)	=0.66 <i>N</i> (3.3849, 0.0110)
	=0.10 <i>Pareto</i> (4.0007, 1.2185)	=0.10 <i>Pareto</i> (3.3329, 1.0218)
海南	=0.16 <i>Exp</i> (0.0490)	-
	=0.70 <i>N</i> (3.5626, 0.0129)	-
	=0.14 <i>Pareto</i> (3.6989, 1.3332)	-
安徽	=0.19 <i>Exp</i> (0.0807)	=0.14 <i>Exp</i> (0.0511)
	=0.68 <i>N</i> (3.8445, 0.0111)	=0.76 <i>N</i> (3.3466, 0.0199)
	=0.13 <i>Pareto</i> (3.9130, 1.1220)	=0.10 <i>Pareto</i> (3.4430, 1.1610)
重庆	=0.14 <i>Exp</i> (0.0301)	=0.26 <i>Exp</i> (0.0311)
	=0.69 <i>N</i> (3.6809, 0.0550)	=0.62 <i>N</i> (3.2849, 0.1060)
	=0.17 <i>Pareto</i> (3.7129, 1.0251)	=0.12 <i>Pareto</i> (3.3429, 1.0261)
云南	=0.14 <i>Exp</i> (0.0314)	=0.17 <i>Exp</i> (0.0527)
	=0.75 <i>N</i> (3.6247, 0.0267)	=0.75 <i>N</i> (3.2441, 0.0081)
	=0.11 <i>Pareto</i> (3.7750, 1.2050)	=0.08 <i>Pareto</i> (3.3000, 1.1090)
贵州	=0.24 <i>Exp</i> (0.0133)	=0.34 <i>Exp</i> (0.0307)
	=0.66 <i>N</i> (3.8212, 0.0031)	=0.59 <i>N</i> (3.3849, 0.0460)
	=0.10 <i>Pareto</i> (3.9107, 1.2285)	=0.07 <i>Pareto</i> (3.4629, 1.0291)
甘肃	=0.16 <i>Exp</i> (0.0280)	=0.14 <i>Exp</i> (0.0307)
	=0.72 <i>N</i> (3.7626, 0.0106)	=0.75 <i>N</i> (3.3461, 0.0281)
	=0.12 <i>Pareto</i> (3.8900, 1.3152)	=0.11 <i>Pareto</i> (3.4430, 1.1090)
四川	=0.19 <i>Exp</i> (0.0217)	=0.14 <i>Exp</i> (0.0355)
	=0.68 <i>N</i> (3.7281, 0.0431)	=0.76 <i>N</i> (3.1859, 0.0390)
	=0.13 <i>Pareto</i> (3.8030, 1.1490)	=0.10 <i>Pareto</i> (3.2911, 1.0611)

地区	城镇	农村
内蒙古	=0.14 $Exp(0.0330)$	=0.24 $Exp(0.0417)$
	=0.69 $N(3.7809, 0.0160)$	=0.65 $N(3.3411, 0.0211)$
	=0.17 $Pareto(3.8429, 1.0291)$	=0.11 $Pareto(3.4930, 1.1630)$
广西	=0.14 $Exp(0.0269)$	=0.14 $Exp(0.0377)$
	=0.75 $N(3.7401, 0.0481)$	=0.67 $N(2.9843, 0.0290)$
	=0.11 $Pareto(3.8630, 1.0390)$	=0.19 $Pareto(3.0029, 1.0921)$
新疆	=0.14 $Exp(0.0258)$	-
	=0.76 $N(3.3848, 0.0380)$	-
	=0.10 $Pareto(3.5429, 1.0202)$	-

表 6 2004年东、中、西部地区城镇和农村家庭各收入阶层的分布函数

地区	城镇	农村
东部地区	=0.19 $Exp(0.0539)$	=0.13 $Exp(0.0716)$
	=0.66 $N(4.0987, 0.0328)$	=0.73 $N(3.5113, 0.0119)$
	=0.15 $Pareto(4.256, 1.2838)$	=0.14 $Pareto(3.7871, 1.0556)$
中部地区	=0.14 $Exp(0.0354)$	=0.20 $Exp(0.0587)$
	=0.75 $N(3.8229, 0.0249)$	=0.77 $N(3.3814, 0.0246)$
	=0.11 $Pareto(3.9150, 1.2351)$	=0.13 $Pareto(3.5246, 1.1346)$
西部地区	=0.16 $Exp(0.0163)$	=0.13 $Exp(0.0211)$
	=0.74 $N(3.6012, 0.0491)$	=0.79 $N(3.1565, 0.0348)$
	=0.10 $Pareto(3.7707, 1.2459)$	=0.08 $Pareto(3.2947, 1.0364)$

(二)城镇和农村基尼系数及其混合基尼系数分解

根据拟合所得的家庭人均收入分布函数,利用公式(10)可以计算得到全国和各地区中国城镇基尼系数、农村基尼系数、城乡差异基尼系数和城乡混合基尼系数。表7和表8给出了相应的结果。按照联合国有关组织规定,基尼系数若低于0.2表示收入绝对平均,0.2~0.3表示比较平均,0.3~0.4表示相对合理,0.4~0.5表示收入差距较大,0.6表示收入差距悬殊,而国际上通常以0.4作为收入分配差距的“警戒线”。

从表7可以看出,2004年全国基尼系数为0.4554,同样各省(市、自治区)的城乡混合基尼系数不容乐观,除9个省份以外,多数省份的城乡混合基尼系数已经超过收入分配差距的“警戒线”,尤其处于西部地区的云南、贵州、甘肃、四川这四个省份的城乡混合基尼系数达到0.46以上,说明中国城乡差距贫富差距拉大,收入分配不公平。

2004年全国城镇基尼系数为0.3703,多数省(市、自治区)的城镇基尼系数都在

0.3~0.4之间,说明这些省份的城镇内部居民收入分配存在一定差距,而尤以江苏省的城镇基尼系数最高,达到0.3622。

2004年全国农村基尼系数为0.3589,小于全国城镇基尼系数。而绝大多数省份的农村基尼系数要高于该省的城镇基尼系数,说明农村内部的收入分配相对于城镇来说更不公平;除了江西、吉林、湖南、湖北、河南、安徽、甘肃、四川这8个省份外,其他各省(市、自治区)的农村基尼系数都在0.3~0.4之间,且内蒙古的农村基尼系数达到0.3641,说明农村内部收入分配也存在一定差距,成为不可忽视的问题。

2004年全国城乡差距基尼系数为0.4286,说明城乡收入存在巨大的差距。绝大多数省份的城乡差距基尼系数都在0.1左右,说明城镇和农村收入分配差距问题的存在已经成为不争的事实,四川省的城乡差距基尼系数最大,为0.2049,而江苏省的城乡差距基尼系数最小,为0.0393,可以看出收入越欠发达的地区,城乡收入差距相对越大,收入分配也相对越不公平。

如表8所示,从横向来看,2004年东部地区的城镇基尼系数大于农村基尼系数和城乡差距基尼系数,说明东部地区城镇内部的收入分配不公平现象相对于农村内部严重;中部地区的城镇基尼系数略高于农村基尼系数,说明中部地区也存在城镇内部收入差距相对较大的问题;而西部地区的农村基尼系数高于城镇基尼系数,说明西部地区的农村内部收入差距相对城镇内部较大,存在收入分配不公平、不合理的问题。

从纵向来看,2004年西部地区的城乡混合基尼系数最大,已经超过收入分配差距的“警戒线”,其次为东部地区,最小的为中部地区,说明西部地区的城乡收入差距最大,存在收入分配不公平的现象;东部地区的城乡混合基尼系数也到达0.3905,城乡收入分配略显不公平;中部地区的城乡混合系数为0.3118,城乡收入差距相对较小。2004年城镇基尼系数以东部地区最大,西部地区次之,中部地区最小,说明东部地区的城镇内部收入差距相对中、西部地区较大。而2004年农村基尼系数以西部地区最大,东部地区次之,中部地区最小,说明对于农村地区来说,西部地区的收入差距相对较大,不公平现象较严重。2004年西部地区的城乡差距基尼系数最大,为0.1453,远远超过东部地区的0.0575和中部地区的0.0274,说明西部地区的城乡收入差距问题非常严重,已经阻碍了西部地区城乡均衡发展。

另外,我们还利用Sen(1973)提出的基尼系数公式进行了计算,结果见表7和表8。结果显示:用此公式计算出的城乡混合基尼系数普遍小于用分布函数法计算出的城乡混合基尼系数,其差值幅度在0.0002~0.035之间。由于选取的样本一致,因此我们认为这种差异的成因主要来自于计算方法。

表 7 2004年全国、各省(市、自治区)城乡家庭人均收入基尼系数

地区	城乡混合基尼系数	城镇基尼系数	农村基尼系数	城乡差距基尼系数	城乡混合基尼系数*
全国	0.4554	0.3703	0.3589	0.4286	0.4506
北京	-	0.2303	-	-	-
上海	-	0.3122	-	-	-
天津	-	0.3191	-	-	-
广东	0.3694	0.3137	0.3398	0.0578	0.3411
福建	0.3915	0.2911	0.3264	0.1091	0.3752
浙江	0.3877	0.3182	0.3121	0.0838	0.3820
山东	0.3816	0.3083	0.3299	0.0705	0.3707
辽宁	0.3847	0.2883	0.3194	0.0961	0.3729
江苏	0.3908	0.3622	0.3488	0.0393	0.3906
山西	0.4307	0.2996	0.306	0.1541	0.4283
陕西	0.4539	0.2929	0.3627	0.1683	0.4527
江西	0.3984	0.2967	0.2671	0.1482	0.3883
吉林	0.4007	0.2916	0.2708	0.1829	0.3899
湖北	0.4132	0.2997	0.2686	0.1481	0.4118
湖南	0.4313	0.3328	0.2786	0.1359	0.4201
黑龙江	0.3573	0.3214	0.3407	0.0601	0.3492
河南	0.4476	0.3382	0.2917	0.1661	0.4291
河北	0.4506	0.3437	0.3791	0.1029	0.4385
海南	-	0.3546	-	-	-
安徽	0.4433	0.3476	0.2779	0.1461	0.4087
重庆	0.4431	0.2638	0.3003	0.1926	0.4215
云南	0.4808	0.3088	0.3711	0.1695	0.4799
贵州	0.4705	0.3064	0.3315	0.1844	0.4607
甘肃	0.4703	0.3318	0.2682	0.1934	0.4539
四川	0.4626	0.3161	0.2896	0.2049	0.4499
内蒙古	0.3702	0.303	0.3641	0.0459	0.3583
广西	0.4507	0.2633	0.3631	0.1864	0.4471
新疆	-	0.2977	-	-	-

说明:缺少北京、天津、上海、新疆的农村家庭收入数据;海南省农村家庭样本太少,无法计算。
 标注*号的城乡混合基尼系数是按照 Sen(1973)提出的计算公式所得。Sen(1973)提出的计算公式为: $G = (n+1)/n - 2 \sum_{i=1}^n [(n-i+1)y_i]/(n^2\mu)$,其中 y_i 表示按家庭人均收入排序的第 i 个家庭的收入, n 为所有的家庭数目, μ 为平均收入。

数据来源:中国社会综合调查开放数据库(CGSS)、《中国统计年鉴2005》及各省(市、自治区)的统计年鉴相关数据计算所得,下同。

基于家庭收入分布的地区基尼系数的测算及其城乡分解

表 8 2004年东、中、西部地区城乡家庭人均收入基尼系数

地区	城乡混合基尼系数	城镇基尼系数	农村基尼系数	城乡差距基尼系数	城乡混合基尼系数*
东部地区	0.3905	0.3499	0.3274	0.0575	0.3724
中部地区	0.3118	0.2957	0.2791	0.0274	0.3002
西部地区	0.4522	0.3093	0.3542	0.1453	0.4516

对城乡混合基尼系数进行分解一直是收入分配研究的重要课题。本文在计算城镇基尼系数、农村基尼系数和城乡差距基尼系数的基础上,对城乡混合基尼系数进行分解,并利用公式(14)分别计算了各自的贡献率,如表9和表10所示。

表 9 2004年全国、各省(市、自治区)城乡混合基尼系数分解后的贡献率 %

地区	城镇贡献率	农村贡献率	相对城乡差距贡献率
全国	43.0563	35.9582	21.9855
广东	58.8081	28.2860	12.9059
福建	52.8479	24.1152	23.0369
浙江	35.6618	45.5222	18.8160
山东	42.2107	41.2837	16.5056
辽宁	43.2808	35.0761	21.6431
江苏	46.1031	44.8554	9.0415
山西	35.7478	34.5357	29.7165
陕西	43.3739	26.1972	30.4288
江西	37.3734	33.3983	29.2283
吉林	55.8216	15.7419	28.4365
湖北	39.2337	29.8425	30.9239
湖南	45.7901	26.2627	27.9472
黑龙江	73.0796	17.8861	9.0344
河南	37.0918	33.1779	29.7304
河北	35.2967	45.2001	19.5032
安徽	41.5799	29.4465	28.9736
重庆	31.1789	32.2797	36.5415
云南	33.0121	37.5116	29.4763
贵州	26.6417	41.6328	31.7255
甘肃	43.9684	21.4870	34.5446
四川	30.2640	34.8758	34.8602
内蒙古	44.8925	44.4071	10.7003
广西	37.2336	29.2171	33.5493

从表 9 可以看出,2004 年全国城镇基尼系数的贡献率大于农村、城乡差距的贡献率,说明城镇内部收入差距是造成全国收入差距的最主要的原因,其次为农村内部收入差距和城乡相对收入差距。除了浙江、河北、重庆、云南、贵州和四川这 6 个省份外,其他省份的城镇基尼系数的贡献率都高于农村基尼系数的贡献率和城乡差距基尼系数的贡献率,说明城镇收入差距是造成各省(市、自治区)收入差距的主要原因。从数值比较来看,黑龙江省的城镇基尼系数贡献率最大,达到 73% 以上,而四川省的城镇基尼系数贡献率最小,仅 30.26%;河北省的农村基尼系数贡献率最大,为 45.2%,而吉林省的农村基尼系数贡献率最小,为 15.75%;四川省和甘肃省的相对城乡差距基尼系数的贡献率都较大,达到 34% 以上,说明这两省份的城乡收入差距非常不公平,而江苏省和黑龙江省的相对城乡差距基尼系数的贡献率较小,仅 9% 多,说明这两省份的城乡收入差距相对较小。

从表 10 可以看出,2004 年东、中、西部地区的城镇基尼系数贡献率都高于农村基尼系数贡献率和相对城乡差距贡献率,说明在这三大地区中城镇收入差距是造成这三个地区城乡收入差距的主要原因。从城镇基尼系数贡献率来看,东部地区最大,中部地区次之,西部地区最小,说明东部地区的城镇收入差距相对最大,收入分配不公平。从农村基尼系数贡献率来看,中部地区农村基尼系数贡献率均大于东、西部地区农村基尼系数贡献率,说明中部地区农村收入差距相对较大,而东部地区农村收入差距相对较小。从相对城乡差距基尼系数的贡献率来看,西部地区最大,达到 27.13%,而东部地区和中部地区仅为 12.09% 和 7.48%,说明西部地区城镇和农村的相对差距是导致西部地区收入差距不可忽视的重要原因。

表 10 2004 年东、中、西部地区城乡混合基尼系数分解后的贡献率 %

地区	城镇贡献率	农村贡献率	相对城乡差距贡献率
东部地区	63.3433	24.5712	12.0855
中部地区	53.6222	38.9005	7.4772
西部地区	37.6088	35.2598	27.1314

另外,我们同时也用泰尔指数进行了城乡分解,结果见表 11 和表 12,其结果显示收入差距主要来自城乡差距,而城镇和农村各自内部差距并不大,这一分解结果与基尼系数的城乡分解结果存在差异。程永宏(2007)对这一问题进行了解释,他指出:“对泰尔指数进行城乡分解的方法并不完善。万广华(2005)早已注意到:经验证据表明,对泰尔指数进行分解时,组间差距贡献率依赖于分组数目;当分组数目较大时,组

基于家庭收入分布的地区基尼系数的测算及其城乡分解

表 11 2004年全国、各省(市、自治区)的泰尔指数及其城乡分解贡献率

地区	城乡混合 泰尔系数	城镇内部 泰尔系数	农村内部 泰尔系数	城乡之间 泰尔系数	城镇内部 贡献率(%)	农村内部 贡献率(%)	城乡差距 贡献率(%)
全国	0.1034	0.0381	0.0225	0.0428	36.85	21.76	41.39
北京	-	0.0293	-	-	-	-	-
上海	-	0.0281	-	-	-	-	-
天津	-	0.0247	-	-	-	-	-
广东	0.1215	0.0383	0.034	0.0492	31.52	27.98	40.49
福建	0.1103	0.0318	0.0299	0.0486	28.83	27.11	44.06
浙江	0.1024	0.0324	0.0261	0.0439	31.64	25.49	42.87
山东	0.0975	0.0338	0.0232	0.0405	34.67	23.79	41.54
辽宁	0.1147	0.0418	0.023	0.0499	36.44	20.05	43.50
江苏	0.1015	0.033	0.0233	0.0452	32.51	22.96	44.53
山西	0.0934	0.0318	0.0233	0.0383	34.05	24.95	41.01
陕西	0.0904	0.0284	0.0243	0.0377	31.42	26.88	41.70
江西	0.1097	0.042	0.0239	0.0438	38.29	21.79	39.93
吉林	0.0986	0.0355	0.0235	0.0396	36.00	23.83	40.16
湖北	0.1094	0.0361	0.0235	0.0498	33.00	21.48	45.52
湖南	0.108	0.0375	0.0207	0.0498	34.72	19.17	46.11
黑龙江	0.0772	0.0221	0.0286	0.0265	28.63	37.05	34.33
河南	0.097	0.0385	0.0219	0.0366	39.69	22.58	37.73
河北	0.0901	0.032	0.0225	0.0356	35.52	24.97	39.51
海南	-	0.0326	-	-	-	-	-
安徽	0.1077	0.034	0.0281	0.0456	31.57	26.09	42.34
重庆	0.1172	0.042	0.0243	0.0509	35.84	20.73	43.43
云南	0.111	0.0322	0.0213	0.0575	29.01	19.19	51.80
贵州	0.1009	0.0212	0.0228	0.0569	21.01	22.60	56.39
甘肃	0.1079	0.0226	0.0231	0.0622	20.95	21.41	57.65
四川	0.1106	0.0322	0.0231	0.0553	29.11	20.89	50.00
内蒙古	0.1045	0.0257	0.0241	0.0547	24.59	23.06	52.34
广西	0.1101	0.0263	0.0237	0.0601	23.89	21.53	54.59
新疆	-	0.0377	-	-	-	-	-

数据来源:中国社会综合调查开放数据库(CGSS)、《中国统计年鉴2005》及各省(市、自治区)的统计年鉴相关数据计算所得,下同。

表 12 2004年东、中、西部地区的泰尔指数及其城乡分解贡献率

地区	城乡混合 泰尔系数	城镇内部 泰尔系数	农村内部 泰尔系数	城乡之间 泰尔系数	城镇内部 贡献率 (%)	农村内部 贡献率 (%)	城乡差距 贡献率 (%)
东部地区	0.1076	0.0432	0.0193	0.0451	40.15	17.94	41.91
中部地区	0.0675	0.0249	0.0169	0.0257	36.89	25.04	38.07
西部地区	0.1044	0.0254	0.0204	0.0586	24.33	19.54	56.13

间差距影响较小;当用于城乡分解时,由于只有两个分组,组间差距显得尤为重要。本文计算的城乡差距基尼系数尽管很大,但其贡献率并不是最高的。所以,城乡差距贡献率的大小随分解方法的不同而不同,泰尔指数的城乡分解结果夸大了城乡差距贡献率。

四 结论

理论上,从使用数据的角度看,与通常的通过分组数据计算基尼系数相比较,我们所采用的方法计算出来的结果显然应该更准确,原因在于我们考虑了所有抽样家庭的人均收入分布,而分组数据将所有抽样家庭按人均收入的多少又分成几个组,每一组计算其收入平均,这肯定会导致大量关于家庭收入分布的信息损失,无法得到基尼系数的准确值。由于我们首次得到了中国和各地区居民家庭收入的较准确的拟合分布,因此,基于这种拟合分布得到的基尼系数的估算值至少从分布函数法的角度看应该是目前为止最好的。事实上,国内外学者曾分别引入过伽马分布函数、帕累托分布函数和正态分布函数等对实际家庭收入数据进行拟合,效果都不甚理想,主要是因为收入在高、中、低不同阶层的分布形状不一样,有关文献见 McDonald(1995)、成邦文(2005)和王海港(2006)等。另外,利用 Sen(1973)提出的离散公式所计算的基尼系数比分布函数法计算的基尼系数偏小,但相差不大;泰尔指数城乡分解与基尼系数城乡分解的结果存在差异,这一差异主要是由于分解方法不同造成的,泰尔指数的城乡分解结果往往夸大了城乡差距贡献率。

本文利用分布函数法对全国和各地区城镇、农村和城乡混合基尼系数进行测算,并对城乡混合基尼系数进行城乡分解,发现以下结论:(1)中国2004年全国和各地区的城乡低、中、高收入阶层的家庭人均收入分布的对数可以分别用指数分布、对数正态分布和帕累托分布来描述;(2)2004年全国和大部分省份的城乡混合基尼系数已经超过收入分配差距的“警戒线”,城乡贫富差距拉大,收入分配不公平。(3)2004年中国

西部地区的城乡收入差距最大,东部地区次之,中部地区最小。(4)城镇收入差距是造成全国和大部分省(自治区、市)城乡收入差距的主要原因。(5)城镇收入差距也是造成东、中、西部地区城乡收入差距的主要原因。(6)城镇和农村的相对差距是导致城乡收入差距扩大不可忽视的影响。关于如何制定相关政策以便缩小中国各地区收入差异,它是一项系统工程,涉及方方面面,不是三两句话可以解决问题的,但本文的研究成果对于有关政府部门制定相关的政策提供了理论依据。

参考文献:

- 成邦文(2005):《基于对数正态分布的洛仑兹曲线与基尼系数》,《数量经济技术研究》第2期。
- 程永宏(2006):《二元经济中城乡混合基尼系数的计算与分解》,《经济研究》第1期。
- 程永宏(2007):《改革以来全国基尼系数的演变及其城乡分解》,《中国社会科学》第4期。
- 程永宏和糜仲春(1998):《利用个人收入分配函数确定基尼系数的新方法》,《华东经济管理》第1期。
- 陈昌兵(2007):《各地区居民收入基尼系数计算及非参数计量模型分析》,《数量经济技术经济研究》第1期。
- 董静和李子奈(2004):《修正城乡加权法及其应用 - 由农村和城镇基尼系数推算全国基尼系数》,《数量经济技术经济研究》第5期。
- 厉以宁和秦宛顺(1997):《现代西方经济学概论》,北京大学出版社。
- 李实和赵人伟(1999):《中国居民收入分配再研究》,《经济研究》第4期。
- 林毅夫和刘培林(2003):《中国的经济发展战略与地区收入差距》,《经济研究》第3期。
- 彭昭英(2000):《SAS应用开发指南》,北京希望电子出版社。
- 孙靖和黄海滨(2007):《泰尔指数在东、中、西部地区收入差距分析中的应用》,《商业研究》第4期。
- 覃成林和李二玲(2002):《中国南北区域经济差异研究》,《地理学与国土研究》第4期。
- 万广华(1998):《中国农村区域间居民收入差异及其变化的实证分析》,《经济研究》第5期。
- 万广华(2005):《收入差距的地区分解》,《世界经济文汇》第3期。
- 王祖祥(2006):《中部六省基尼系数的估算研究》,《中国社会科学》第4期。
- 王海港(2006):《我国居民收入分配的格局 - 帕累托分布法》,《南方经济》第5期。
- 王云飞(2007):《我国地区收入差距变化趋势 - 基于基尼系数分解的分析》,《山西财经大学学报》第8期。
- 赵桂芹和粟芳(2008):《汽车交通事故损失分布的尾部估计》,《数理统计与管理》第6期。
- Cowell, F. "Measurement of Inequality." Handbook of Income Distribution, North Holland 2000.
- Dagum, C. "A New Approach to the Decomposition of the Gini Income Inequality Ratio." *Empirical Economics* 1997(4), pp. 515 - 531.
- McDonald, J. B. "A Generalization of the Beta Distribution with Application." *Journal of Econometrics*, 1995, 66, pp. 133 - 152.
- Sen, Amartya K. "On Economic Inequality." Oxford: Clarendon Press, 1973, pp. 139 - 148.
- Sundrum. "Inequality Decomposition." *Bulletin of Economic Research*, 1990, 40, pp. 309 - 311.

(截稿:2009年10月 实习编辑:王徽)