

居民收入差距的测量及其合理判断

高霞

北京师范大学

中国收入分配研究院

中国居民收入分配课题组

2013年5月20日



一、收入的含义

二、居民收入差距测量中的相关问题

1. 基础数据来源问题：样本点数据与分层（分组）数据
2. 收入指标的选取问题：家庭平均收入与家庭人均收入
3. 地区价格差异调整问题
4. 异常值的正确处理问题
5. 高收入人群样本的缺失问题

三、对中国基尼系数取值区间的合理判断



一、收入的含义

- 美国著名经济学家海利·西蒙斯（1938）：“个人收入可以定义为这样两项之和，一项是行使的消费权力的市场价值，另一项是资产存量价值的变化。”
- 收入指一定时期内获得的用于个人或家庭的现在与未来消费的资源的市场价值。具体到收入的实际计算上，与此定义最靠近的是家庭或个人的可支配收入。
- 按照国际惯例，在计算基尼系数的时候，一般使用的都是税后收入。
- 为了评估一个国家税收政策的再分配效应，也会计算税前收入的基尼系数来作为比较。2007年我国城镇居民的税前基尼系数为0.3459，税后基尼系数为0.3322，税后基尼系数比税前基尼系数低4%（岳希明、徐静，2012）。



二、居民收入差距测量中的相关问题

1. 基础数据来源问题：样本点数据与分层（分组）数据

■ 研究者在分析居民收入差距时所利用的基础数据，可能来自于样本点数据（**sample data**），也可能来自于分组数据（**grouped data**）。

■ 先分组后估算的基尼系数存在着估算上的误差问题，因为在不考虑各组内收入分布的情况下，利用分层（分组）资料所估算的收入差距忽视了组内差距，而只包含了组间差距，从而会造成估计上的一定误差；组内差距在总差距中所占的比重越大，这种估计上的误差就越大。

■ 分层越细，组别越多，组内差距占总差距的比重越小，从而基尼系数的低估误差越小。



表1 等分组中各组的平均收入与基尼系数

分层依据	组别	样本量	收入均值(元)	基尼系数
五等分组	1	5911	5416	0.153
	2	5910	9179	0.059
	3	5911	12572	0.048
	4	5910	17180	0.059
	5	5911	31239	0.185
十等分组	1	2956	4222	0.139
	2	2955	6611	0.051
	3	2955	8356	0.031
	4	2955	10003	0.027
	5	2956	11661	0.024
	6	2955	13484	0.025
	7	2955	15663	0.026
	8	2955	18697	0.032
	9	2955	23428	0.044
	10	2956	39047	0.165

数据来源：CHIP(2007)。



表2 利用分组资料模拟的基尼系数及其测算误差

	Gini2	(Gini2-Gini1)/Gini1 (%)	组内差距比重 (%)
五等分组资料	0.31565	-6.85	16.73
十等分组资料	0.33185	-2.07	8.30
二十等分组资料	0.33677	-0.62	4.26
一百等分组资料	0.33870	-0.05	1.06
地区分组资料	0.15568	-54.06	79.67

注：Gini1为利用样本点数据计算的基尼系数，Gini2为利用分组资料计算的基尼系数。

数据来源：CHIP(2007)。



二、居民收入差距测量中的相关问题

2. 收入指标的选取问题：家庭平均收入与家庭人均收入

表3 十等分组中各组的平均家庭规模

组别	平均家庭人口规模
1	3.72
2	3.54
3	3.42
4	3.22
5	3.21
6	3.11
7	3.08
8	2.97
9	2.94
10	2.88

■ 根据CHIP(2007)数据，如果选择家庭收入来计算居民的基尼系数，所得到的基尼系数为0.33553，这比用家庭人均收入所计算出的基尼系数（0.33887）要低1%。



表4 样本点数据的数值模拟

家庭代码	个人代码	家庭人均收入	家庭收入	实际个人收入
1	1	200	1000	200
1	2	200	1000	200
1	3	200	1000	200
1	4	200	1000	150
1	5	200	1000	250
2	1	300	1200	300
2	2	300	1200	300
2	3	300	1200	250
2	4	300	1200	350
3	1	450	1350	400
3	2	450	1350	400
3	3	450	1350	500
4	1	750	1500	800
4	2	750	1500	700
5	1	2000	2000	2000
基尼系数:		0.391	0.108	0.401



二、居民收入差距测量中的相关问题

3. 地区价格差异调整问题

- 如果要反映一个国家的实际居民收入差距，而不仅仅是名义居民收入差距，则需要考虑地区价格差异。
- 一般来讲，高收入地区的物价水平也相对较高，因此在进行了地区之间货币购买力的调整之后，所计算出的基尼系数要低于不考虑地区价格差异时的基尼系数。
- 如果直接利用江小涓和李辉（2005）所计算的各省会城市的价格指数来作为各省（直辖市）的价格指数，利用CHIP(2007)数据所得到的基尼系数将为0.31697，比没有经过地区PPP调整的Gini1低6.5%。



二、居民收入差距测量中的相关问题

4. 异常值的正确处理问题

- 异常值的出现可能来自于两种情况：一种异常值可能是总体固有的随机变异性的极端表现，这种异常值和样本中其余观测值属于同一总体；另一种异常值也可能是由于试验条件和试验方法的偶然偏离所产生的后果，或产生于观测、计算、记录中的失误，这种异常值和样本与其余观测值不属于同一总体。
- 当诊断出异常值，一般应首先选择进行样本清理，通过反复检查、核对、寻找出现异常的可确定的原因，以对其进行修正。若找不到出现异常的任何原因，则不能将这种数值随意删除。
- 异常值样本点是否需要剔除，取决于样本本身的质量，也反映了调查者对样本质量的信心。如果极端值样本点（即极高收入样本和极低收入样本）被简单粗暴地直接剔除，无疑将会带来基尼系数的低估。



二、居民收入差距测量中的相关问题

5. 高收入人群样本的缺失问题

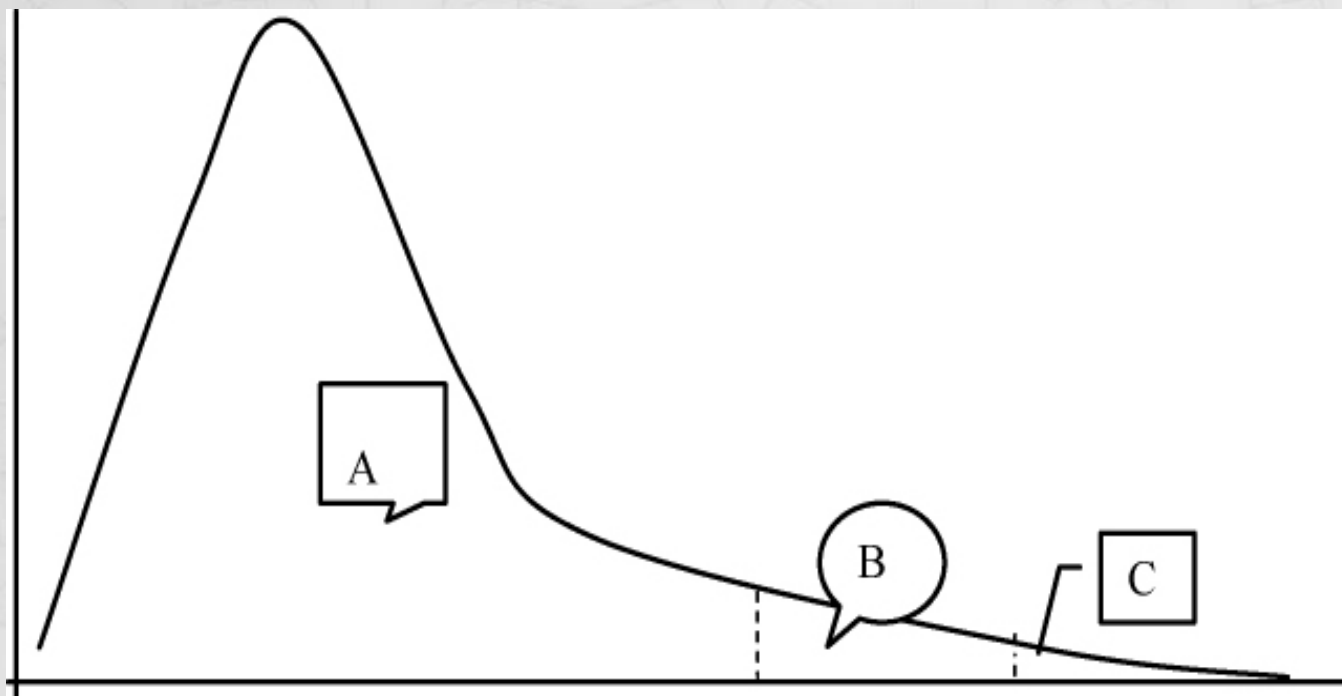


图1 高收入人群样本遗漏示意图

表5 对高收入人群收入及其分布的推算（2007年）

基于帕累托分布的推算	胡润榜+福布斯榜+2007年上市公司高管薪酬 (全国第二次经济普查法人单位数量的1/3加权)
高收入人群中的平均收入(元)	509105
收入水平高于12万元的人数	2804228
占城镇人口(59379万)的比重(%)	0.472
占全国人口(132129万)的比重(%)	0.212
高收入人群基尼系数理论值	0.619

资料来源：李实、罗楚亮(2011)。



表6 城镇和全国基尼系数的估计 (2007年)

收入定义		样本数据未 做任何处理	加权样本数据 (省份加权)	修正抽样误差(富人榜+上市公司数据)	
				1/3加权 (无地区权重)	地区加权 +PPP调整
1. 国家统计局 的收入定义	城镇	0.3364	0.3375	0.4199	0.4126
	全国 (不含流动人口)	0.4737	0.4810	0.5297	0.4822
	全国 (含流动人口)	0.4705	0.4784	0.5240	0.4792
2. 卡恩定义(包 括公有住房的房 租补贴和自有住 房的归算资金, 加实物收入)	城镇	0.3390	0.3371	0.4113	0.3987
	全国 (不含流动人口)	0.4905	0.4965	0.5392	0.4888
	全国 (含流动人口)	0.4831	0.4919	0.5307	0.4852

资料来源：李实、罗楚亮(2011)。



三、对中国基尼系数取值区间的合理判断

■ 2007年后我们课题组没有做全国的住户收入调查，不过从总体而言，我们认为无论是样本的代表性方面，还是收入指标的测量上，统计局住户调查数据质量比其他住户调查更高，尽管也可能存在一定的低估，但其所得到的基尼系数是基本可信的。

■ 当然，这并不意味着我国的基尼系数（不考虑非法收入的情况下）在2012年一定是0.474。数据收集工作的进一步完善和估计方法的不断改善都有可能使得现在公布的基尼系数出现上下2到3个百分点（即0.444~0.504）的波动。这属于正常情况，并不因此改变我们的基本判断。



三、对中国基尼系数取值区间的合理判断

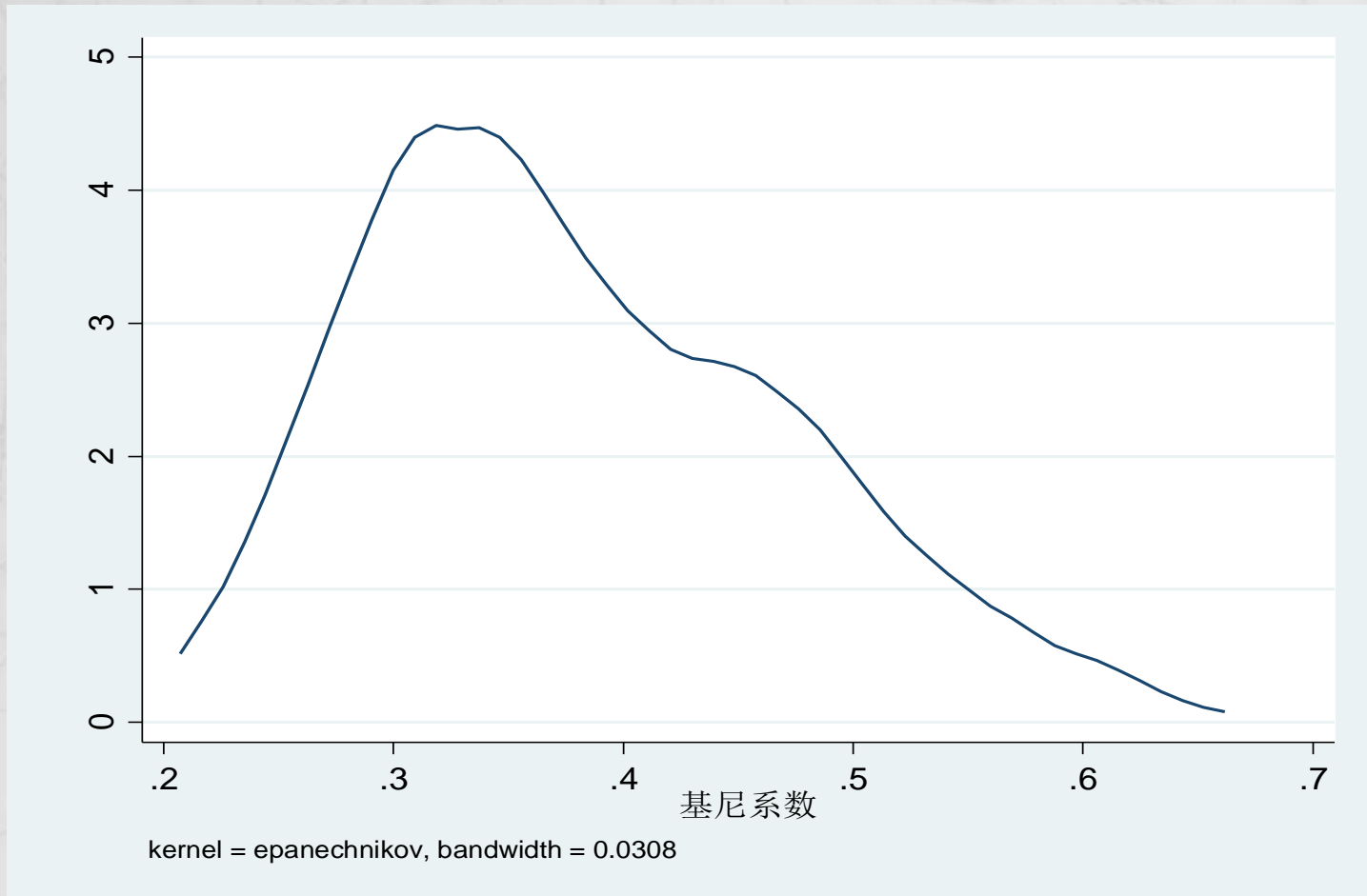


图2 119个国家基尼系数的核密度分布



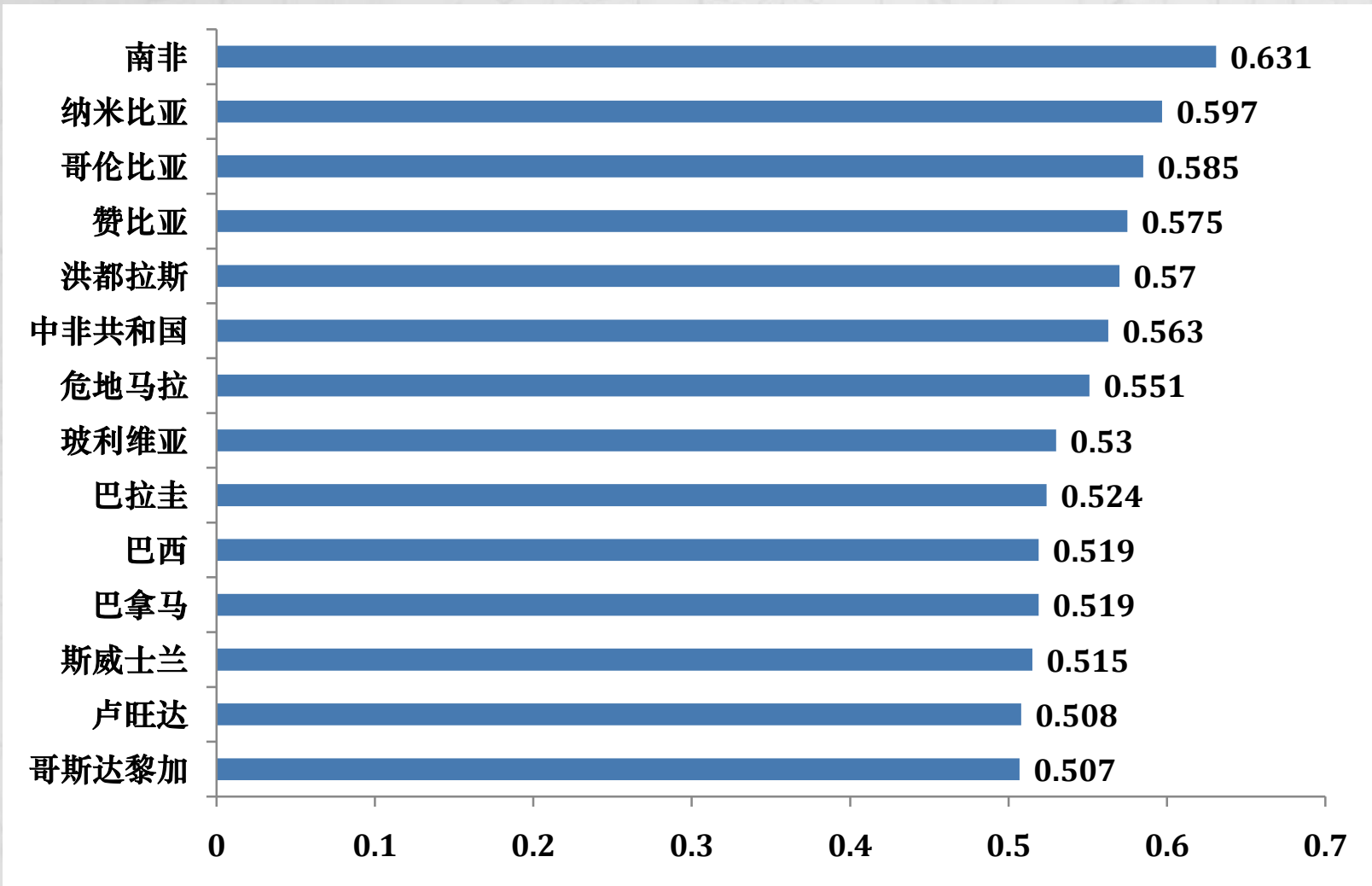


图3 基尼系数在0.5以上的国家



表6 与中国具有可比性的主要国家的基尼系数

国家	基尼系数	国家	基尼系数	国家	基尼系数
美国	0.378	印度尼西亚	0.356	新加坡	0.459
中国	0.474	巴基斯坦	0.3	乌拉圭	0.453
日本	0.329	尼日利亚	0.397	哈萨克斯坦	0.289
德国	0.295	孟加拉国	0.321	阿根廷	0.445
法国	0.293	菲律宾	0.43	巴拿马	0.519
巴西	0.519	越南	0.356	阿塞拜疆	0.337
英国	0.345	埃塞俄比亚	0.336	秘鲁	0.481
意大利	0.319	埃及	0.308	白俄罗斯	0.265
俄罗斯	0.417	土耳其	0.4	多米尼加	0.472
印度	0.339	泰国	0.394	安哥拉	0.427
加拿大*	0.324	南非	0.631	约旦	0.354
西班牙*	0.317			厄瓜多尔	0.493
澳大利亚*	0.336			阿尔巴尼亚	0.345
墨西哥	0.472				
韩国*	0.311				



■ 需要指出的内容：

- (1) 由于各国的统计口径和测算方法并不完全一致，进行基尼系数的严格国际比较需要审慎对待。
- (2) 一国的基尼系数与该国的经济发展阶段、经济结构、教育水平、人口结构、再分配政策等实际情况有关。很难判断对于一个国家而言，究竟是0.4还是0.5才是该国基尼系数的合理取值。
- (3) 基尼系数是一个反映总体收入相对差距的客观数字指标，该数字背后的过程是极其复杂的，需要在充分的专业的调查数据的基础上，通过严格的数据处理来完成其测算过程。
- (4) 随着近年来我国要素流动自由度的提高、农村劳动力人口向城镇的大量转移、不同地区之间市场竞争程度的加强、城镇户籍在大部分城市更加易得，我国的基尼系数近年来基本保持稳定甚至有所下降的结果是可解释也是基本可信的。
- (5) 基尼系数作为衡量收入差距的数字指标之一，并不能衡量收入分配的不公之处，收入差距和收入分配不公应该结合起来考虑。



谢谢各位!

