

# 中国总体收入基尼系数的估计：1985—2008

胡志军 刘宗明 龚志民\*

**摘要** 给定收入分布的参数形式，我们使用农村、城镇的分组数据估计了分布的参数。我们考虑了多种常用于收入分布描述的分布函数，并比较了这些分布的拟合结果。结果表明，广义贝塔分布Ⅱ型比较适合于农村、城镇居民的收入分布拟合。在此基础上，我们计算了农村、城镇及总体的基尼系数。结果显示，1985—2008年间，总体基尼系数一直在上升，城镇具有上升趋势，农村1985—2003年间一直在上升，2003年以后具有缓慢下降趋势，而且城乡差距是我国总体差距的主要原因。

**关键词** 基尼系数，极大似然估计，广义矩估计

## 一、引言

基尼系数是一个国际上通用的描述收入不平等状况的指标。我国总体的基尼系数是多少也一直受到广泛关注。但国家统计局只是对外公布了城镇和农村分开的收入分组数据，并没有公布每年的居民总体基尼系数值。所以，在没有其他信息的情况下，如何根据这些分组数据来估计我国总体的基尼系数是值得研究的。

国内外有许多基于分组数据估计基尼系数的研究文献，一类是讨论基于特定分组数据形式（仅已知每组数据的均值、上下限，以及该组的人口份额）的离散的基尼系数估计的上下限（Gastwirth, 1972; Fuller, 1979; McDonald *et al.*, 1981; Gastwirth *et al.*, 1986; 等等）；一类是讨论基于分组数据如何估计出基尼系数，方法之一就是假定收入服从连续形式的某个特定分布，然后使用计量经济学的方法估计出分布参数，再根据收入分布函数计算出基尼系数（McDonald, 1984; Slotte, 1989; Chotikapanich *et al.*, 2007a; 等等）。

由于国家统计局公布的城镇和农村收入分组数据是分开的，没有城乡混合的分组数据，所以基于分组数据估计中国总体基尼系数的研究更侧重于如

\* 胡志军、刘宗明，上海财经大学经济学院；龚志民，湘潭大学商学院。通信作者及地址：胡志军，上海市三门路409号上海财经大学研究生公寓5-401, 200439；电话：13564271990；E-mail: hzhj0414@yahoo.com.cn。本研究受到上海财经大学研究生科研创新基金资助。作者感谢两位匿名审稿人对本文所提出的建议，当然，文责自负。

何构造全国的收入分组数据,或更进一步——如何构造出全国的收入分布。一类方法是采用离散形式的方法,陈宗胜(2002)使用一个“分层加权法”的方法计算全国基尼系数,该方法把城乡人口分组数据中各组人口比重的加权平均等同于全国人口分组数据中的各组人口比重,这是不合适的;胡祖光(2004)提出一个计算全国基尼系数的简易近似公式,在收入五分法中最高收入组与最低收入组各自所占的收入比重之差即为基尼系数值,但是如何构造出全国混合的收入五等分组是困难的。另一类方法是采用连续形式的方法,通过拟合城乡混合的收入分布来计算全国基尼系数,王祖祥(2006)使用二次样条函数逼近方法拟合农村收入分布,通过洛仑兹曲线的拟合得到城镇收入分布,从而得到拟合的城乡混合收入分布,并计算了中部六省的总体基尼系数值;程永宏(2006,2007)以广义logistic分布来拟合农村、城镇的收入分布,通过普通最小二乘方法(OLS)估计分布的参数,根据收入分布函数,给出一个二元经济中城乡混合基尼系数的计算方法,不足之处在于广义logistic分布的定义域为 $(-\infty, +\infty)$ ,而收入一般是正数,而且城镇分组数据的分布拟合难以使用OLS方法估计;Chotikapanich *et al.* (2007b)使用了韦布分布拟合农村收入分布、广义贝塔分布Ⅱ型拟合城镇收入分布,分别采用极大似然估计(MLE)方法估计农村收入分布参数,采用广义矩估计(GMM)方法估计城镇收入分布参数及城镇收入分组数据的上下限,然后得到总体分布函数并计算了总体基尼系数。

本文基于国家统计局公布的农村、城镇收入分组数据,采用拟合收入分布的方法来计算全国总体基尼系数。对于农村收入分布的分布参数估计,我们采用MLE;对于城镇收入分布的分布参数估计,则采用GMM。与Chotikapanich *et al.* (2007b)的研究相比,我们在农村、城镇收入分布的拟合过程中,考虑了常用于描述收入分布的三种分布:韦布分布、广义贝塔分布Ⅱ型和对数正态分布。而且在对城镇收入分布的参数进行估计时,我们只估计了分布参数,城镇收入分组数据的上下限则可以表示成分布参数的函数,即减少了估计参数的个数,提高了估计效率。

通过对1985—2008年我国农村、城镇分组数据的收入分布拟合,我们发现农村、城镇收入分布比较适合用广义贝塔分布Ⅱ型拟合。而且对我国1985—2008年的基尼系数估计结果表明:1985—2008年间,我国总体基尼系数一直在上升,由0.3168上升到0.4767;城镇基尼系数1985—2008年间具有上升趋势,由0.1665上升到0.3402;农村基尼系数1985—2003年间一直在上升,由0.2802上升到0.3576,2003年以后具有缓慢下降趋势。而且,城乡差距对我国总体差距的贡献率1985—2000年间由50.6%上升到60.7%,2001—2008年一直维持在60%左右;农村内部、城镇内部差距对总体差距的贡献率1985—2000年间由45.6%下降到35.5%,2001—2008年一直维持在35%左右;交叉项对总体差距的贡献率1985—2008年间稳定

在5%左右。

本文接下来的结构安排如下：第二部分介绍基于分组数据的分布拟合估计方法，并给出基于收入分布的基尼系数计算公式和组间分解公式；第三部分是基于我国1985—2008年农村、城镇分组数据的总体基尼系数的估计；第四部分是结论。

## 二、基于分组数据的收入分布拟合及基尼系数估计

### (一) 基于分组数据的收入分布拟合

假设居民收入  $x$  服从参数为  $\theta$  的某个分布，其分布函数为  $F(x; \theta)$ 。一般情况下，我们能够得到的样本是基于分组形式的数据。根据这些样本信息，可以估计出收入分布函数的参数  $\theta$ ，从而得到分布函数  $F(x; \hat{\theta})$  以及相应的基尼系数。

考虑到国家统计局公布的收入分组数据特点，我们考察两种分组数据形式的参数估计方法。

#### 1. 仅知各组收入上下限及人口份额（农村收入数据分组形式）

假设样本数据为  $(x_1, x_2, \dots, x_n)$ ， $x_i \in (a_{k-1}, a_k]$ ， $k=1, 2, \dots, K$ ， $p_k$  为落在第  $k$  个收入区间内的人口份额。其中， $a_k$ 、 $p_k$  已知， $F(a_0; \theta) = 0$ ， $F(a_K; \theta) = 1$ 。一般的，我们可以设定  $a_0 = 0$ ， $a_K = +\infty$ 。根据极大似然估计原理，可以得到  $\theta$  的一致估计量  $\hat{\theta}$ 。

我们给出一个不同于 Chotikapanich *et al.* (2007b) 的推导极大似然函数的方法：类似于一个多元选择模型，如果个体  $i$  的收入  $x_i \in (a_{k-1}, a_k]$ ，那么他报告  $y_i = k$ ，我们所知道的只有  $a_k$  和  $k$ 。那么  $\text{prob}(y_i = k) = \text{prob}(a_{k-1} < x_i \leq a_k) = F(a_k; \theta) - F(a_{k-1}; \theta)$ ，我们可以得到极大似然函数

$$L(\theta) = \prod_{i=1}^n \text{prob}(y_i = k) = \prod_{i=1}^n [F(a_k; \theta) - F(a_{k-1}; \theta)]^{1[y_i = k]}$$

和对数极大似然函数

$$\ln L(\theta) = \sum_{i=1}^n 1[y_i = k] \ln [F(a_k; \theta) - F(a_{k-1}; \theta)],$$

其中  $1[y_i = k]$  表示  $y_i = k$  时取值为 1， $y_i \neq k$  时取值为 0。所以

$$\ln L(\theta) = \sum_{k=1}^K p_k \ln [F(a_k; \theta) - F(a_{k-1}; \theta)].$$

$\theta$  的 MLE 估计量  $\hat{\theta}$  可以表示为

$$\hat{\theta} = \arg \max_{\theta} \sum_{k=1}^K p_k \ln [F(a_k; \theta) - F(a_{k-1}; \theta)].$$

2. 仅知各组收入均值及人口份额(城镇收入数据分组形式)

假设样本数据为  $(x_1 \leq x_2 \leq \dots \leq x_n)$ , 将样本按收入从小到大分成  $K$  组,  $\mu_k$  为第  $k$  组的收入均值,  $p_k$  为第  $k$  组内的人口份额。其中,  $\mu_k$ 、 $p_k$  已知。根据广义矩估计原理, 可以得到  $\theta$  的一致估计量  $\hat{\theta}$ 。

假设第  $k$  组的上下限分别为  $a_{k-1}$ 、 $a_k$ , Chotikapanich *et al.* (2007b) 将  $a_k$  ( $k=1, \dots, K-1$ ) 与  $\theta$  一起作为待估计的参数, 令  $\beta = (\theta, a_1, \dots, a_{K-1})$ , 则

$$\hat{\beta} = \arg \min_{\beta} \sum_{k=1}^K [e_k^2 + \epsilon_k^2].$$

其中

$$e_k = \left[ \frac{\int_{a_{k-1}}^{a_k} x dF(x; \theta)}{p_k} - \mu_k \right] / \mu_k, \quad \epsilon_k = \frac{\int_{a_{k-1}}^{a_k} dF(x; \theta) - p_k}{p_k}.$$

事实上, 我们可以仅将  $\theta$  作为待估计的参数, 因为  $a_k$  ( $k=1, \dots, K-1$ ) 是  $\theta$  的函数。具体来说, 我们可以通过以下联立方程组把  $a_k$  表示成  $\theta$  的函数  $a_k(\theta)$ 。

$$\begin{cases} F(a_1; \theta) - F(0; \theta) = p_1 \\ F(a_2; \theta) - F(a_1; \theta) = p_2 \\ \vdots \\ F(a_{K-1}; \theta) - F(a_{K-2}; \theta) = p_{K-1} \\ F(+\infty; \theta) - F(a_{K-1}; \theta) = p_K \end{cases},$$

其中

$$F(0; \theta) = 0, \quad F(+\infty; \theta) = 1,$$

从而

$$\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^K [e_k(\theta)]^2,$$

其中

$$e_k = \left[ \frac{\int_{a_{k-1}(\theta)}^{a_k(\theta)} x dF(x; \theta)}{p_k} - \mu_k \right] / \mu_k,$$

可以看到, 我们利用了同样的信息, 但减少了  $K-1$  个待估计参数, 使得估计效率提高了。

## (二) 基于收入分布函数的基尼系数估计及组间分解

根据 Kendall and Stuart (1958) 和 Dorfman (1979) 的研究, 假设连续的收入变量  $x$  的分布函数为  $F(x)$ ,  $x \in [a, b]$ , 则收入的基尼系数

$$G = \frac{\int_a^b F(x)(1-F(x))dx - a}{\int_a^b (1-F(x))dx}.$$

一般情况下,  $a=0, b=+\infty$ , 则

$$G = \frac{\int_0^{\infty} F(x)(1-F(x))dx}{\int_0^{\infty} (1-F(x))dx}.$$

我们考虑总体基尼系数按两组分解的组间分解形式。假设总体被分成两部分, 其人口份额分别为  $\eta$  和  $1-\eta$ , 收入分布函数分别为  $F_1(x)$  和  $F_2(x)$ ,  $x>0$ 。为方便说明, 我们假设

$$\begin{aligned} \mu_1 &\equiv \int_0^{\infty} x dF_1(x) < \mu_2 \equiv \int_0^{\infty} x dF_2(x), \\ G_i &= \frac{\int_0^{\infty} F_i(x)(1-F_i(x))dx}{\int_0^{\infty} (1-F_i(x))dx}, \quad i=1,2. \end{aligned}$$

则总体的分布函数  $F(x)=\eta F_1(x)+(1-\eta)F_2(x)$ , 总体的收入均值  $\mu=\eta\mu_1+(1-\eta)\mu_2$ 。

如果代表组间差距的基尼系数定义为“各组内成员收入以其所属组的组内均值代替后的总体基尼系数”, 那么总体基尼系数不能表示成各组基尼系数的加权和(权数是各组平均收入与人口数量份额的函数)与代表组间差距的基尼系数之和。此时基尼系数不能按照组别进行完全的分解, 而是会产生一个分解剩余项。Cowell (2000)、徐宽 (2003) 等都对基尼系数的这一组间分解形式进行了总结。在连续情形下, 这一基尼系数的分解公式(以两组为例)为

$$\begin{aligned} G &= \frac{\mu_1 \eta^2}{\mu} G_1 + \frac{\mu_2 (1-\eta)^2}{\mu} G_2 + \frac{\eta(1-\eta)}{\mu} (\mu_2 - \mu_1) \\ &\quad + \frac{2\eta(1-\eta)}{\mu} \int_0^{\infty} [1-F_1(x)]F_2(x)dx, \end{aligned}$$

其中  $G_w \equiv \frac{\mu_1 \eta^2}{\mu} G_1 + \frac{\mu_2 (1-\eta)^2}{\mu} G_2$  为组内差距,  $G_B \equiv \frac{\eta(1-\eta)}{\mu} (\mu_2 - \mu_1)$  为组间差距,  $R \equiv \frac{2\eta(1-\eta)}{\mu} \int_0^{\infty} [1-F_1(x)]F_2(x)dx$  为分解剩余项(也称为交叉项)。

如果组间差距采用其他定义, 那么基尼系数的组间分解形式可能会不同

(程永宏, 2008; 洪兴建, 2008; 等等)。在本文的实证研究中, 我们采用了在基尼系数研究文献中使用较多的分解形式, 即  $G=G_w+G_B+R$ 。

### 三、基于分组数据的中国总体基尼系数 估计: 1985—2008

在本文拟合我国农村、城镇居民收入分布的实践中, 我们主要考虑几种常用于描述收入分布的函数: 对数正态分布、韦布分布和广义贝塔分布 II 型。这三种分布的一些特征详见表 1。

表 1 三种分布的特征

分布类型	密度函数	定义域	均值
韦布分布	$f(x \alpha, \beta) = \beta \alpha^{-\beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}$	$x > 0$	$E(x) = \alpha \Gamma\left(1 + \frac{1}{\beta}\right)$
广义贝塔分布 II 型	$f(x c, p, q) = \frac{x^{p-1}}{c^p B(p, q) \left(1 + \frac{x}{c}\right)^{p+q}}$	$x > 0$	$E(x) = \frac{cp}{q-1}$
对数正态分布	$f(x \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$	$x > 0$	$E(x) = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$

根据 McDonald (1984) 的研究, 这三种分布的基尼系数  $G$  分别可以表示为:  $G_{\lognorm} = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$ , 其中  $\Phi(\cdot)$  是标准正态分布;  $G_{weibull} = 1 - \frac{1}{2^{\frac{1}{\beta}}}$ ;

$G_{beta2} = \frac{2B(2p, 2q-1)}{p(B(p, q))^2}$ , 其中  $B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt$ 。

#### (一) 数据描述

本文中我们所用的数据均来自历年的《中国统计年鉴》, 农村居民的收入分组数据是农村居民按纯收入分组的户数占调查户比重的数据; 城镇居民的收入分组数据是城镇最低收入户 (10% 人口份额)、低收入户 (10%)、较低收入户 (20%)、中等收入户 (20%)、较高收入户 (20%)、高收入户 (10%) 和最高收入户 (10%) 家庭的平均每人可支配收入数据。

由于本文采用的数据是家庭人均收入数据, 所以我们计算的是基于家庭人均收入的基尼系数, 而不是基于个人收入的基尼系数。因此, 本文计算的中国总体基尼系数是以家庭为基本单位的, 我们没有计算以个人为单位的总体基尼系数, 即没有考虑家庭人口规模、家庭内部收入分配等因素对总体基尼系数的影响。

## (二) 估计结果

### 1. 农村居民收入分布的拟合

首先, 我们使用 MLE 对农村居民的收入分布函数进行拟合估计。根据本文第二部分第一节的估计方法, 我们使用 Matlab 软件, 对 1985—2008 年我国农村收入分布分别进行韦布分布、广义贝塔分布 II 型和对数正态分布的拟合估计。估计的详细结果见表 2。

表 2 农村收入分布拟合的 MLE 结果

时间	韦布分布			广义贝塔分布 II 型			对数正态分布			
	$\alpha$	$\beta$	极大似然值	$c$	$p$	$q$	极大似然值	$\mu$	$\sigma$	极大似然值
85	448.3972	1.9372	-2.0677	257.6843	9.8085	7.3636	-1.9699	5.8558	0.5032	-1.9722
90	740.6975	1.7782	-2.2443	510.1537	6.7011	6.1656	-2.1706	6.3244	0.5797	-2.1725
95	1848.1	1.6715	-2.7152	1963.7	4.7339	6.6074	-2.6544	7.2177	0.6329	-2.6607
00	2647.3	1.7281	-2.7043	6039.7	3.4484	9.6652	-2.6785	7.5810	0.6792	-2.6938
01	2759.4	1.7072	-2.7004	7259.1	3.2086	10.259	-2.6786	7.6204	0.7012	-2.6987
02	2875.8	1.7170	-2.6872	6544.7	3.3090	9.2243	-2.6647	7.6647	0.6999	-2.6815
03	3014.2	1.6937	-2.6782	8694.4	3.0017	10.422	-2.6612	7.7091	0.7268	-2.6814
04	3354.2	1.7889	-2.6050	15401	2.9874	15.979	-2.5933	7.8357	0.7110	-2.6152
05	3664.3	1.7927	-2.5503	2.9686 * E8	2.5416	2.2631 * E5	-2.5470	7.9327	0.7510	-2.5825
06	4009.8	1.8286	-2.4658	1.3952 * E6	2.5850	984.8872	-2.4626	8.0379	0.7525	-2.4914
07	4551.5	1.8850	-2.3182	3.6663 * E7	2.5731	2.2369 * E4	-2.3204	8.1951	0.7831	-2.3564
08	5165.4	1.8831	-2.1474	3.6925 * E10	2.4861	1.8923 * E7	-2.1518	8.3554	0.8354	-2.1890

从估计的极大似然值角度考虑, 在农村收入分布的拟合上, 广义贝塔分布 II 型和对数正态分布要优于韦布分布, 广义贝塔分布 II 型要微弱地优于对数正态分布。根据拟合的收入分布, 我们还可以计算出特定收入区间的人口份额拟合值。将拟合值与样本值进行比较, 具体比较情形详见图 1。

从每个特定收入区间的人口份额拟合值与样本值的比较角度考虑, 相对而言, 广义贝塔分布 II 型是最适合于农村收入分布的拟合。

所以, 从这些拟合结果判断, 我们认为, 农村居民收入分布适合用广义贝塔分布 II 型拟合。

### 2. 城镇居民收入分布的拟合

我们使用 GMM 对城镇居民的收入分布函数进行拟合估计, 根据本文第二部分第一节的估计方法, 对 1985—2008 年我国城镇居民收入分布分别进行韦布分布、广义贝塔分布 II 型和对数正态分布的拟合估计, 详细结果见表 3。根据拟合的收入分布, 我们可以计算出每个收入阶层的平均收入拟合值。将拟合值与样本值进行比较, 具体比较情形详见图 2。

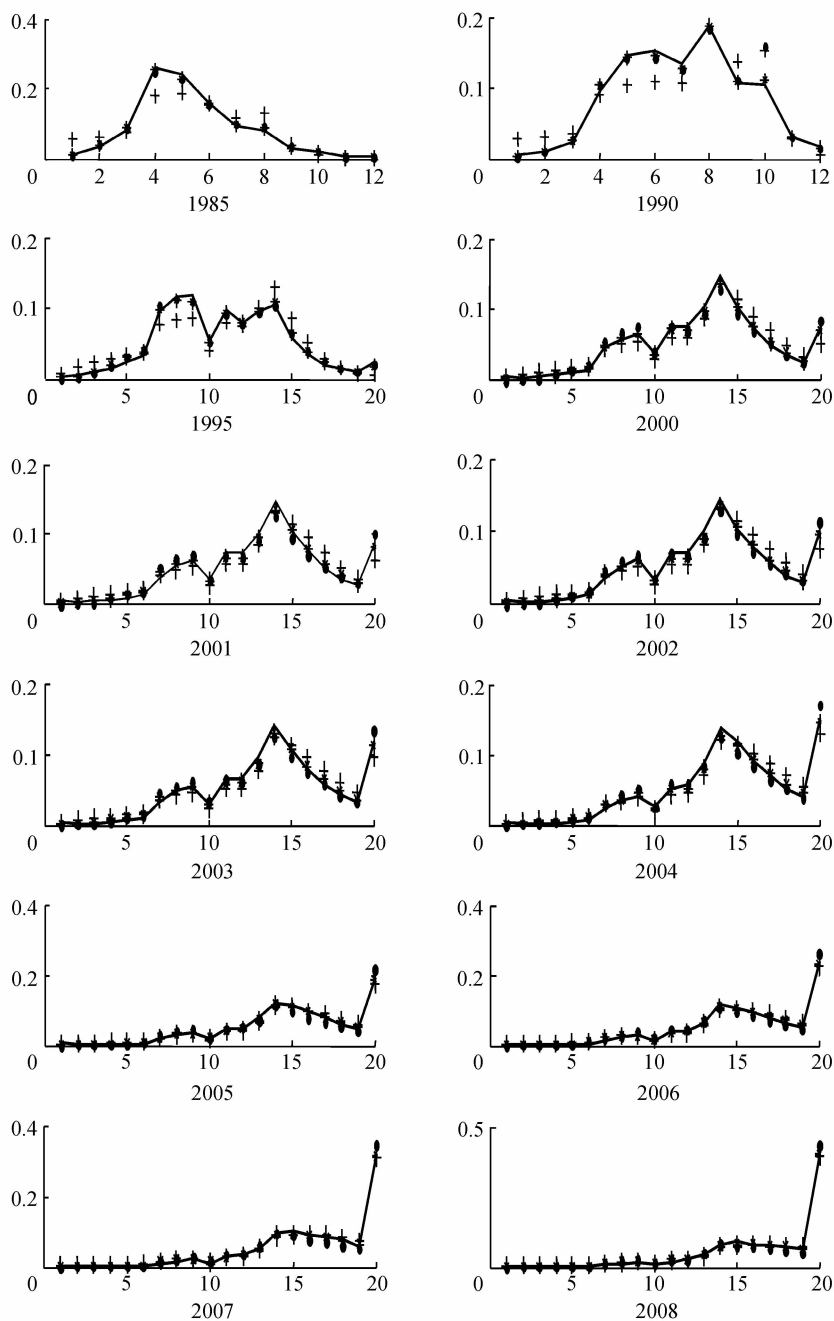


图1 历年农村居民收入分组数据的人口份额拟合值与样本值的比较

注：图中实线表示样本值，“+”表示韦布分布拟合值，“\*”表示广义贝塔分布Ⅱ型拟合值，“•”表示对数正态分布拟合值。



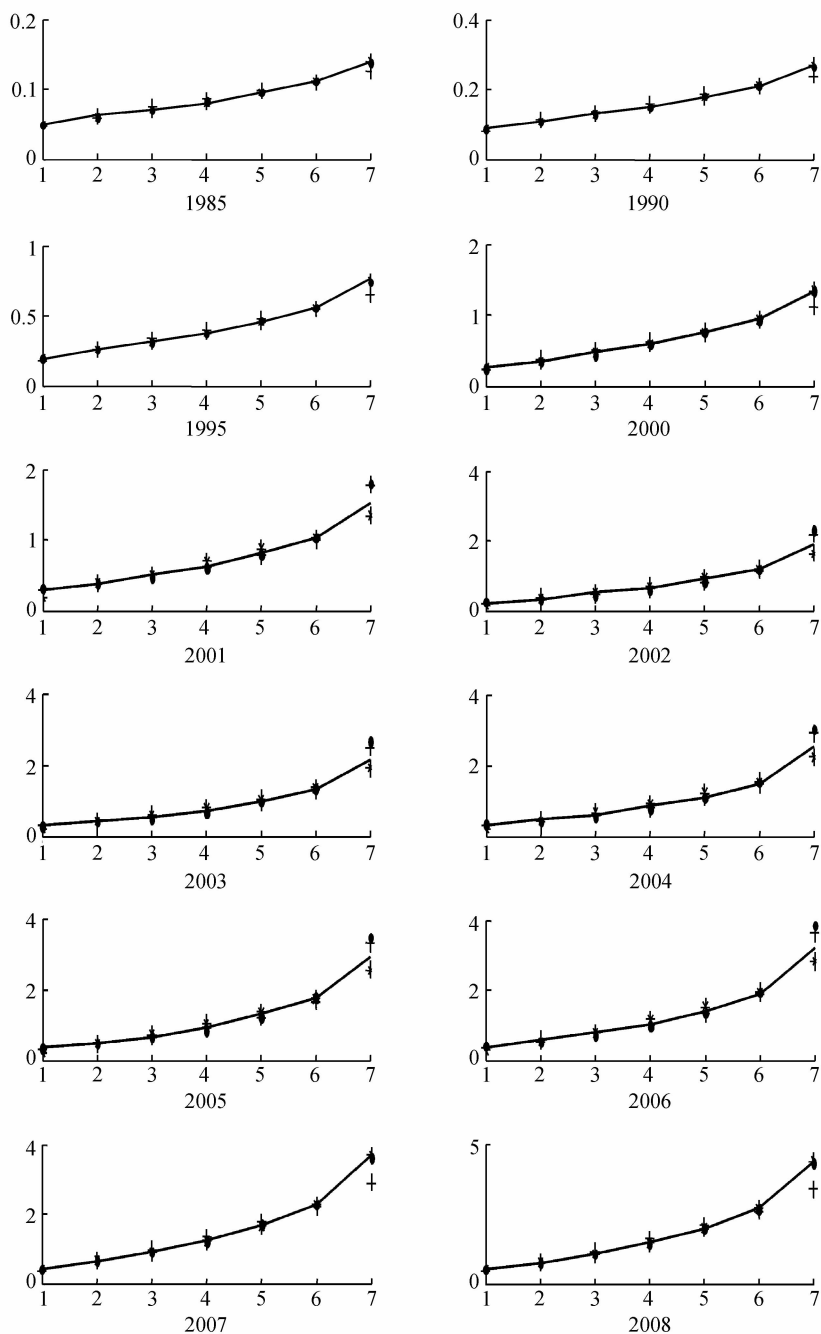


图2 历年城镇居民收入分组数据的组平均收入拟合值与样本值的比较

注：图中实线表示样本值，“+”表示韦布分布拟合值，“\*”表示广义贝塔分布Ⅱ型拟合值，“·”表示对数正态分布拟合值。

表3 城镇收入分布拟合的GMM结果

时间	韦布分布			广义贝塔分布II型				对数正态分布		
	$\alpha$	$\beta$	残差平方和	$c$	$p$	$q$	残差平方和	$\mu$	$\sigma$	残差平方和
85	933.1250	4.2669	0.0289	541.2951	28.8727	19.5306	0.0002	6.6943	0.2964	0.0004
90	1744.8	3.9504	0.0336	994.5778	24.9873	16.8194	0.0002	7.3092	0.3196	0.0004
95	4470.9	3.2771	0.0501	2283.6	18.2762	11.4410	0.0002	8.2207	0.3845	0.0007
00	7313.5	2.7206	0.0568	6782.6	9.3592	10.7605	0.0000	8.6750	0.4578	0.0001
01	7981.7	2.6145	0.0679	5750.4	9.7190	8.8640	0.0001	8.7553	0.4773	0.0002
02	8664.4	2.1576	0.0877	7029.5	6.2590	6.6629	0.0007	8.7905	0.5780	0.0017
03	9578.4	2.1061	0.1019	6541.7	6.4351	5.8770	0.0012	8.8863	0.5944	0.0027
04	10846	2.0644	0.1101	6899.3	6.3917	5.4961	0.0013	9.0063	0.6074	0.0034
05	12282	2.0102	0.1067	9066.6	5.6923	5.6494	0.0005	9.1220	0.6205	0.0017
06	13709	2.0411	0.1077	9488.1	6.0268	5.6161	0.0007	9.2368	0.6124	0.0022
07	16012	2.0598	0.1022	11798	5.9646	5.8738	0.0007	9.3937	0.6063	0.0019
08	18695	1.9998	0.1045	14514	5.5286	5.7532	0.0002	9.5402	0.6221	0.0011

与农村数据的拟合情形相似,无论是从残差平方和角度考虑,还是从每个收入阶层的平均收入拟合值与样本值的比较角度考虑,在城镇居民收入分布的拟合上,广义贝塔分布II型和对数正态分布要优于韦布分布,广义贝塔分布II型要微弱地优于对数正态分布。

所以,我们认为,城镇居民收入分布也最适合用广义贝塔分布II型来拟合。

### 3. 基尼系数的估计及城乡分解

在农村、城镇收入分布都采用广义贝塔分布II型拟合的基础上,我国总体收入分布就是一个混合的广义贝塔分布II型。1985—2008年间我国农村、城镇及城乡混合的收入密度函数详见图3。

从1985—2008年间我国农村、城镇及总体收入密度函数的变化可以看出,居民收入分布的峰度越来越小、尾部越来越厚。

根据本文第二部分第二节的步骤,我们可以估计出农村、城镇各自的基尼系数,也可以估计出城乡混合的基尼系数,具体结果详见表4。

为了方便考察基尼系数的变动趋势,我们给出农村、城镇及总体的基尼系数随时间变动的趋势图,见图4。

从表4和图4中,我们可以发现:1985—2008年间,我国总体基尼系数一直在上升,由1985年的0.3168上升到2008年的0.4767,上升了约0.16;城镇居民收入基尼系数在1985—2008年间具有上升趋势,由0.1665上升到0.3402;农村居民收入基尼系数在1985—2003年间一直在上升,由1985年的0.2802上升到2003年的0.3576,2003年以后具有缓慢下降趋势。

根据本文第二部分第二节的基尼系数分解公式,我们还可以得到我国总体基尼系数的城乡分解结果,具体结果详见表5。

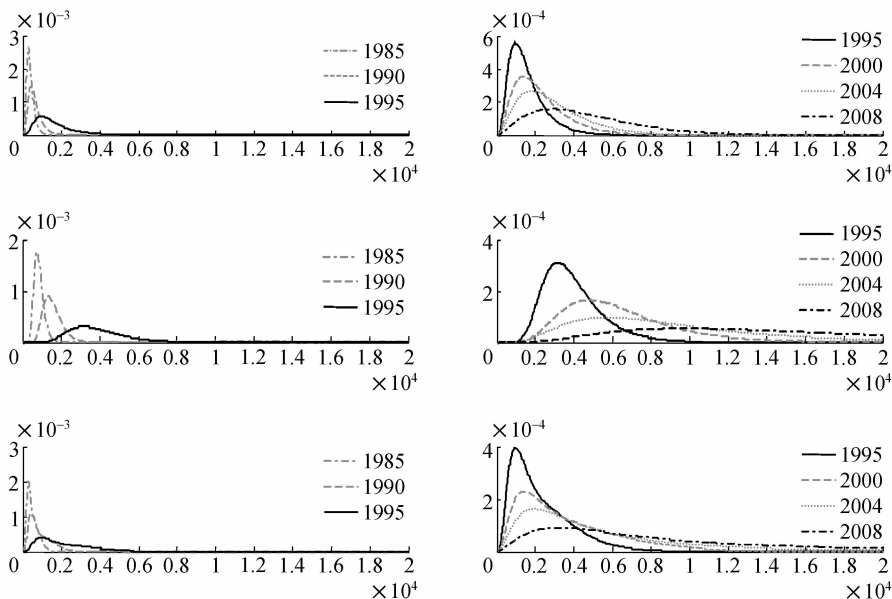


图3 历年农村、城镇及总体收入密度函数

注：左（上、中、下）分别为1985年、1990年、1995年的农村、城镇居民及城乡混合的收入密度函数；右（上、中、下）分别为1995年、2000年、2004年、2008年的农村、城镇居民及城乡混合的收入密度函数。

表4 农村、城镇及总体的基尼系数

时间	农村基尼系数	城镇基尼系数	交叉项	总体基尼系数
1985	0.2802	0.1665	0.0120	0.3168
1990	0.3198	0.1795	0.0136	0.3584
1995	0.3395	0.2157	0.0175	0.3761
2000	0.3451	0.2531	0.0153	0.4012
2001	0.3504	0.2648	0.0146	0.4121
2002	0.3522	0.3171	0.0179	0.4367
2003	0.3576	0.3274	0.0164	0.4493
2004	0.3424	0.3351	0.0142	0.4506
2005	0.3370	0.3400	0.0120	0.4565
2006	0.3349	0.3366	0.0109	0.4563
2007	0.3352	0.3326	0.0102	0.4559
2008	0.3404	0.3402	0.0252	0.4767

根据这些分解结果，我们认为：城乡差距是我国总体差距的主要原因。而且城乡差距对总体差距的贡献率呈现出以下特点：城乡差距对我国总体差距的贡献率1985—2000年间由50.6%上升到60.7%，2001年以后一直维持在60%左右；农村内部、城镇内部差距对总体差距的贡献率1985—2000年间由45.6%下降到35.5%，2001年以后一直维持在35%左右；交叉项对总体差距的贡献率1985—2008年间则稳定在5%左右。

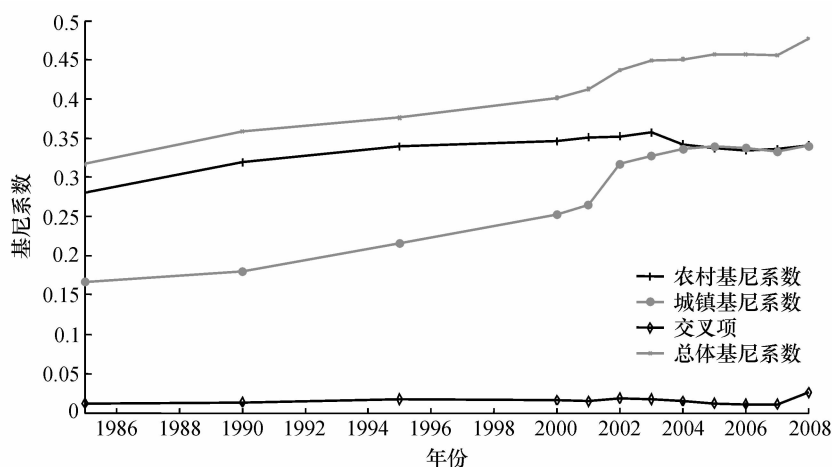


图4 基尼系数的趋势图

表5 总体基尼系数的城乡分解结果

时间	农村、城镇内部差距	农村、城镇内部差距 占总体差距的比重(%)	城乡差距	城乡差距占总体差距 的比重(%)
1985	0.1445	45.6100	0.1603	50.6135
1990	0.1489	41.5463	0.1959	54.6707
1995	0.1524	40.5128	0.2063	54.8390
2000	0.1423	35.4705	0.2436	60.7157
2001	0.1436	34.8396	0.2539	61.6186
2002	0.1553	35.5527	0.2635	60.3386
2003	0.1583	35.2312	0.2746	61.1131
2004	0.1580	35.0680	0.2784	61.7858
2005	0.1593	34.8851	0.2853	62.4844
2006	0.1588	34.7937	0.2866	62.8091
2007	0.1587	34.8112	0.2870	62.9544
2008	0.1629	34.1786	0.2885	60.5283

## 四、结 论

我们基于《中国统计年鉴》中的农村、城镇居民收入分组数据,估计了全国总体基尼系数。我们考虑了三种收入分布对农村、城镇居民收入分布的拟合情况,即韦布分布、广义贝塔分布Ⅱ型和对数正态分布,分别采用了MLE和GMM方法对分布函数的参数进行估计。与Chotikapanich *et al.* (2007b)的研究相比,在农村、城镇收入分布的拟合过程中,我们考虑了常用于描述收入分布的三种分布,而且在对农村收入分布进行MLE估计时,采用了一个多元选择概率模型来推导出极大似然函数,在对城镇收入分布进行GMM估计时,通过一组联立方程,我们可以只估计分布参数,即减少了估计参数的个数,提高了估计效率。

通过对1985—2008年我国农村、城镇分组数据的一个实证研究，我们得到以下一些结论：

(1) 我们认为，在本文所考虑的三种分布中，相对而言，广义贝塔分布Ⅱ型适合用于我国农村居民收入分布的拟合，也适合用于我国城镇居民收入分布的拟合；而且，在该分布的前提下，我国农村、城镇及总体的收入分布呈现出峰度越来越小、尾部越来越厚的特点。

(2) 我国城乡混合的总体基尼系数一直在上升，由1985年的0.3168上升到2008年的0.4767，城镇基尼系数1985—2008年间具有上升趋势，由0.1665上升到0.3402；农村基尼系数1985—2003年间一直在上升，由0.2802上升到0.3576，2003年以后具有缓慢下降趋势。

(3) 城乡差距是我国总体差距的主要原因。具体而言，城乡差距对我国总体差距的贡献率1985—2000年间由50.6%上升到60.7%，2001—2008年一直维持在60%左右；农村内部、城镇内部差距对总体差距的贡献率1985—2000年间由45.6%下降到35.5%，2001—2008年一直维持在35%左右；交叉项对总体差距的贡献率1985—2008年间稳定在5%左右。

## 参 考 文 献

- [1] 程永宏，“二元经济中城乡混合基尼系数的计算与分解”，《经济研究》，2006年第1期，第109—120页。
- [2] 程永宏，“改革以来全国总体基尼系数的演变及其城乡分解”，《中国社会科学》，2007年第4期，第45—60页。
- [3] 程永宏，“基尼系数组群分解新方法研究：从城乡二亚组到多亚组”，《经济研究》，2008年第8期，第124—144页。
- [4] 陈宗胜、周云波，《再论改革与发展中的收入分配》。北京：经济科学出版社，2002年。
- [5] Chotikapanich, D., W. Griffiths, and D. Rao, “Estimating and Combining National Income Distributions Using Limited Data”, *Journal of Business and Economic Statistics*, 2007, 25(January), 97—109.
- [6] Chotikapanich, D., D. Rao, and K. Tang, “Estimating Income Inequality in China Using Grouped Data and the Generalized Beta Distribution”, *Review of Income and Wealth*, 2007, 53(1), 127—147.
- [7] Cowell, F., “Measurement of Inequality”, in Atkinson, A., and F. Bourguignon (eds.), *Handbook of Income Distribution*. Amsterdam: North Holland, 2000.
- [8] Dorfman, R., “A Formula for The Gini Coefficient”, *Review of Economics and Statistics*, 1979, 61(1), 146—149.
- [9] Fuller, M., “The Estimation of Gini Coefficients from Grouped Data: Upper and Lower Bounds”, *Economics Letters*, 1979, 3(2), 187—192.
- [10] Gastwirth, J., “The Estimation of the Lorenz Curve and Gini Index”, *Review of Economics and Statistics*, 1972, 54(3), 306—316.
- [11] Gastwirth, J., T. Nayak, and A. Krieger, “Large Sample Theory for the Bounds on the Gini and Related Indices of Inequality Estimated from Grouped Data”, *Journal of Business & Economic Statistics*, 1986, 4(2), 269—273.

- [12] 洪兴建,“一个新的基尼系数子群分解公式——兼论中国总体基尼系数的城乡分解”,《经济学(季刊)》,2008年第8卷第1期,第307—324页。
- [13] 胡祖光,“基尼系数理论最佳值及其简易计算公式研究”,《经济研究》,2004年第9期,第60—69页。
- [14] Kendall, M., and A. Stuart, *The Advanced Theory of Statistics, Vol. 1, Distribution Theory*, 1st ed. New York: Hafner Publishing Company, 1958.
- [15] McDonald, J., and M. Ransom, “An Analysis of the Bounds for the Gini Coefficient”, *Journal of Econometrics*, 1981, 17(2), 177—188.
- [16] McDonald, J., “Some Generalized Functions for the Size Distribution of Income”, *Econometrica*, 1984, 52(3), 647—663.
- [17] Slottje, D., *The Structure of Earnings and the Measurement of Income Inequality in the U. S.* Amsterdam: North-Holland, 1989.
- [18] 王祖祥,“中部六省基尼系数的估算研究”,《中国社会科学》,2006年第4期,第77—87页。
- [19] 徐宽,“基尼系数的研究文献在过去八十年是如何拓展的”,《经济学(季刊)》,2003年第2卷第4期,第757—778页。

## Estimation of the Gini Coefficient in China: 1985-2008

ZHIJUN HU ZONGMING LIU

(*Shanghai University of Finance and Economics*)

ZHIMIN GONG

(*Xiangtan University*)

**Abstract** We estimate China's income distribution using group data from the urban and rural areas. We consider some familiar distributions in the literature, compare the results of those fitted distributions, and find that the generalized beta II distribution is the most suitable distribution for the group data. We then use this distribution calculate the Gini coefficient. From 1985 to 2008, the total Gini coefficient and the urban Gini coefficient rise throughout the whole period, and the rural Gini coefficient rises until 2003 but declines slightly after 2003.

**JEL Classification** C13, C46, D31