基于一族新洛伦兹曲线的收入不平等测度
——以湖北省城镇及农村的收入数据为例

王红霞 *

摘要: 本文阐述了一族新的洛伦兹曲线估计模型,并利用湖北省2000 - 2006年的城镇及农村收入数据对模型进行了检验。结果表明,采用此模型估计的基尼系数不会超出Gastwirth (1972)用非参数方法给出的基于分组数据计算所得的基尼指数的上下限,而且此模型的估计结果相当稳健。我们的分析也表明, 2000 - 2006年湖北省农村及城镇的收入不平等程度均不高,且呈小幅上涨趋势。

关键词: 洛伦兹曲线 基尼系数 收入不平等

一、问题的提出

收入分配是近来政府关注的重点问题之一。湖北省统计局的专题报告《构建和谐湖北进程中的居民收入差距问题研究》①指出,改革开放以来湖北省城乡居民收入差距呈现"W"形走势,且1998年以来一直呈递增态势,城镇居民内部收入分配的"马太效应"愈演愈烈。根据其计算结果,湖北省城镇基尼系数2000年为0.26, 2002年和2003年分别上升到0.34和0.32, 2004年、2005年略有下降,分别为0.267和0.275。城镇居民收入分配处于相对均等与相对合理的状态。农民内部收入差距持续扩大,不均衡性不断加剧。据国家统计局湖北调查总队的测算,湖北省农民收入分配的基尼系数自2000年以来快速增长, 2003年增大到0.298, 2004年为0.307, 2005年扩大到0.314。农民收入分配差距呈现出持续扩大的趋势。

以上是利用湖北省统计局的住户调查数据得到的结果,然而这种数据在一般研究中很难得到。相比于住户调查数据,分组数据则可从统计年鉴中得到,涵盖面也更广。然而,当我们使用分组数据进行研究时,则需要用洛伦兹模型估计真实的洛伦兹曲线。本文扩展了Sarabia等(1999)的洛伦兹曲线估计模型,并将其应用于2000 - 2006年湖北省城镇及农村的分组数据,估计了湖北城镇及农村的收入不平等,同时用可得的住户数据对比检验了本模型的估计误差与适用性。

本文结构如下:第二部分在Sarabia等(1999)模型的基础上阐述了一个新的洛伦兹曲线模型;第三部分以湖北省数据为例讨论了相关模型的应用问题,包括具体的估计方法、模型的适用性比较、模型估计的误差分析以及基于估计出的洛伦兹曲线进行的福利比较等;最后一部分是结论。

二、估算方法综述及新的估算模型

关于洛伦兹曲线与基尼系数的分析与测算,国内有学者提出的主要方法有几何计算法、分布函数法与曲线拟合法3种(成邦文, 2005)。几何计算法,是根据分组数据刻画洛伦兹曲线,按几何图形分块计算基尼系数,这是一种相对比较粗糙的算法。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,然后对洛伦兹曲线与基尼系数进行估计(McDonald and Xu, 1995;成邦文, 2005)。然而,由于分析的复杂性,这一方法使用较少。曲线拟合法首先假设收入分配服从某一特殊的统计分布函数形式,如对数正态分布、伽马分布、贝塔分布、帕累托分布、威布尔分布等,再据此为洛伦兹曲线选择适当的参数方程直接进行拟合、确定参数,由此估计出洛伦兹曲线与基尼系数。本文用到的即是曲线拟合法。

---

* 王红霞,武汉大学经济发展研究中心,邮政编码: 430072,电子信箱: moonsailorwhu@gmail.com。

本文为叶初升教授主持的教育部人文社会科学重点研究基地重大课题"发展经济学的微观基础研究"(07JJD790141)、国家自然科学基金资助项目"农村扶贫政策绩效评价及其动态瞄准机制设计"(70873088)的阶段性成果。感谢匿名评审人的有益建议,当然,文责自负。

①参见国研网(http://www.drcnet.com.cn/DRCNet.Common.Web/DocViewSummary.aspx?docid=1342554&leafid=12)。

目前,已经有一些适用于分组数据的参数化洛伦兹模型(Cheong, 2002; Ogwang and Rao, 1996, 2000; Ortega et al, 1991; Sarabia et al, 1999, 2001, 2005; Schader and Schmid, 1994;欧阳植等, 1994;程永宏等, 1998),但这些模型用于逼近洛伦兹曲线时有一个共同的缺点,即它们不能完全满足洛伦兹曲线的定义(Kakwani and Podder, 1976; Kakwani, 1980)。另外,正如 Cheong(2002)指出的,有些洛伦兹模型还有一个缺点,即其估计值超出了 Gastwirth(1972)用非参数方法给出的基于分组数据计算所得的基尼指数的上下限,它们一般会低估或者高估基尼系数。最近的研究都关注于寻找同时满足理论与实际的新的洛伦兹模型,使得这些模型不仅能满足洛伦兹曲线的定义,而且应用于各种数据源时均呈现良好的性质(Ogwang and Rao, 2000; Ryn and Slottje, 1996)。

洛伦兹曲线 $L(p)$是一种函数关系,它刻画了所拥有的收入最少的百分之 p的人口所占的收入份额。假定 $L(p)$在 $[0,1]$上有定义,且是连续的,其二阶导为 $L''(p)$。如果对所有的 $p\in[0,1]$, $L(p)$满足 $L(0)=0$, $L(1)=1$, $L'(0^+)\geq 0$,且 $L''(p)\geq 0$,则 $L(p)$为洛伦兹曲线。

Sarabia等(1999)给出了基于经典帕累托分布的洛伦兹模型的基本形式:

$$S_0(p)=1-(1-p)^{\beta} \tag{1}$$

从这一基本形式出发,他们又介绍了一族广义帕累托族的洛伦兹曲线,都与(1)式密切相关,具体形式为:

$$S_1(p)=p^{\alpha}[1-(1-p)^{\beta}] \tag{2}$$

$$S_2(p)=[1-(1-p)^{\beta}]^{\eta} \tag{3}$$

$$S_3(p)=p^{\alpha}[1-(1-p)^{\beta}]^{\eta} \tag{4}$$

Ogwang和 Rao(2000)提出了用两种混合的方法建立洛伦兹模型:加权积(weighted product)及洛伦兹模型的凸组合(convex combination)。他们还发现,公式(2)和 Chotikapanich(1993)提出的洛伦兹模型:

$$L_C(p)=\frac{e^{\lambda p}-1}{e^{\lambda}-1},\ \lambda>0$$

的凸组合模型:

$$L_{OR}(p)=\delta p^{\alpha}[1-(1-p)^{\beta}]+(1-\delta)\frac{e^{\lambda p}-1}{e^{\lambda}-1},\ \delta\in[0,1]$$

符合洛伦兹模型的条件。

为了方便后面的论述,我们先给出如下定理:

定理　假定 $L(p)$为洛伦兹曲线,则对于任意的 $\alpha\geq 0$和 $\eta\geq 1$, $\tilde{L}(p)=p^{\alpha}L(p)^{\eta}$也是洛伦兹曲线。进而,如果对于所有的 $p\in[0,1]$都有 $L'''(p)\geq 0$,则当 $\alpha\geq 0$, $\eta\geq 1/2$且 $\alpha+\eta\geq 1$时, $\tilde{L}(p)$也是洛伦兹曲线(证明请参见附录)。

因为基于帕累托分布得到的 $S_0(p)$和 $L_C(p)$在$[0,1]$上都满足 $L'''(p)\geq 0$,故而,通过以上定理我们可以得到另一种洛伦兹曲线族模型:

$$L_{PC}(p)=p^{\alpha}\left\{\delta[1-(1-p)^{\beta}]+(1-\delta)\frac{e^{\lambda p}-1}{e^{\lambda}-1}\right\}^{\eta}$$

其参数范围为: $\alpha\geq 0$, $\beta\in(0,1]$, $\lambda>0$, $\delta\in[0,1]$, $\eta\geq 1/2$, $\alpha+\eta\geq 1$。

再考虑如下二元参数模型:

$$H_0(p)=1-(1-p)^{\beta}e^{-\gamma p} \tag{5}$$

$H_0(p)$是一个更一般化的经典洛伦兹曲线模型。可以看出,将其作为一个洛伦兹曲线的参数估计模型使用时,它不亚于公式(2)或公式(3)。而且,总的来说,它比 $S_0(p)$的性质更好。

因为 $H_0(p)$和 $L_C(p)$同时满足 $L'''(p)\geq 0$,我们可以给出另一种与前述 $L_{PC}(p)$相似的混合模型族,形式为:

$$L_{HC}(p)=p^{\alpha}\left\{\delta[1-(1-p)^{\beta}e^{-\gamma p}]+(1-\delta)\frac{e^{\lambda p}-1}{e^{\lambda}-1}\right\}^{\eta}$$

其中, $\alpha\geq 0$, $\beta\in(0,1]$, $0\leq\beta+\gamma\leq\sqrt{\beta}$, $\lambda>0$, $\delta\in[0,1]$, $\eta\geq 1/2$,且 $\alpha+\eta\geq 1$。这也是本文进行洛伦兹曲线估计所使用的模型。

三、模型的应用与进一步讨论

我们研究收入分配的目的本质上是为了研究经济社会中人口福利随时间的变化,或在不同的经济地区

之间进行福利比较。

对于福利比较,Shorrocks(1983)提出了在实际中很常见的分布排序方法,也是我们进行福利比较的基础,即首先比较分布内部的不平等程度,再比较各分布的平均收入。假设 $L_A(p)$和 $L_B(p)$分别是收入分布 A 和 B 的洛伦兹曲线,且其相应的平均收入分别为 μ_A 和 μ_B,如果对任意的 $p\in[0,1]$有 $\mu_A\geq\mu_B$ 且 $L_A(p)\geq L_B(p)$,则对于一个偏好公平的政策制定者而言,A的社会福利高于 B。然而,这一标准却不能用于洛伦兹曲线相交时的情形。为了度量洛伦兹曲线相交时的社会福利,Shorrocks(1983)又引入了广义洛伦兹曲线(GLC)的概念,并将广义洛伦兹曲线定义为 $GL(p)=\mu L(p)$。用广义洛伦兹占优标准来排序的政策制定者同时偏好公平和效率(收入越高越好)。当且仅当对所有的 $p\in[0,1]$有 $\mu_A L_A(p)\geq\mu_B L_B(p)$时(暗含了两条广义洛伦兹曲线没有相交),他们认为 A优于 B。还有一些研究进一步讨论了当洛伦兹曲线或广义洛伦兹曲线相交有限次时的福利比较方法①,然而,在这些情形下排序的问题就变得相当复杂。因此,通常的处理方式是,在广义洛伦兹曲线相交时,用不平等指数(如基尼系数等)来进行福利比较。

本文的分析也遵从这样的思路,即先介绍模型估计的方法,再利用特定年份可得的不同数据类型比较估计结果以验证模型的适用性,并将其运用于更多年份的洛伦兹曲线估计,最后基于我们的估计结果计算基尼系数,并比较福利。

(一)估计方法

如果给定收入区间 $(x_{i-1},x_i]$, $i=1,2,\cdots$,那么收入不高于 x_i的人口份额为 p_i,由$\{(p_i,L_i)\}_{i=1}^{m}$给出,则$\hat{F}(x_i)$为 p_i的估计值。给定收入分配的密度函数 $f(x)$,累积分布函数的形式为 $F(x)=\int_0^x f(t)\,dt$,表示收入小于等于 x的人口所占的比例。这一分布的洛伦兹曲线是:

$$L(p)=\frac{1}{\mu}\int_0^y xf(x)\,dx,\ p=F(y)$$

其中 $p\in[0,1]$, μ是这一分布的平均收入。只需进行简单的求导运算就可得:

$$L'(p)=\frac{y}{\mu} \tag{9}$$

$$L''(p)=\frac{1}{\mu f(y)} \tag{10}$$

给定洛伦兹曲线 $L(p)$和收入 y,解公式(9)和公式(10)可得密度函数 $f(y)$,进而可得收入不高于 y的人口份额 $p=F(y)$。

如果有且仅有一组分组数据$\{(p_i,L_i)\}_{i=1}^{m}$可用于估计真实分布的洛伦兹曲线,则我们必须选择适合这一分组数据的洛伦兹模型。其中,p_i是按收入由低至高排序低端收入的人口比例,L_i是对应收入最低的 p_i的人口所占的总收入份额。我们选取的模型形式为 $L(p,\tau)$,令 τ为参数向量,即($\alpha,\beta,\gamma,\lambda,\eta,\delta$),取损失函数为残差平方和的形式,由最小残差平方和(RSS) $\sum_{i=1}^{m}(L_i-L(p_i,\tau))^2$的方式来估计。很显然,一阶条件为零即可得 $L(p_i,\tau)$的估计。

假定 τ的估计值是 $\hat{\tau}$,那么不仅能通过估计的洛伦兹曲线 $L(p,\hat{\tau})$来近似得到洛伦兹曲线 $L(p)$并计算各种不平等指数,而且可以从公式(9)、(10)中由 $L(p,\hat{\tau})$得到近似的密度函数 $\hat{f}(y)$。由于 $L(p,\hat{\tau})$是用最小化RSS的标准得到的最优估计,因此,$\hat{f}$是最好的可能估计。更进一步,对任意给定的 y,我们还可以通过解 $\hat{F}(y)=\int_0^y\hat{f}(x)\,dx$或 $\mu L'(p;\hat{\tau})=y$得到相应的累积分布函数 $\hat{F}(y)$。后面我们可以看到,比较用不同方法得到的密度函数(或累积分布函数)也可以看出模型的适用性。

最后,为了度量收入不平等进行福利比较,我们使用基尼系数:

$$G=2\int_0^1(p-L(p))\,dp=1-2\int_0^1 L(p)\,dp \tag{11}$$

(二)数据描述

我们所使用的是由湖北省统计局提供的 2000-2006年的湖北省农村及城镇的分组数据及 2006年的微观数据,所有数据均来自住户调查年报基层表。这些年份的数据包括组区间 $(x_{i-1},x_i]$, $i=1,2,\cdots$;这一收入

①例如,可参见 Davies和 Hoy(1995)。

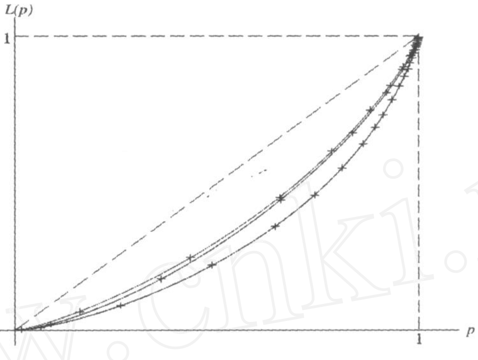
区间的人口数 n_i 和这一区间内的年人均收入 μ_i。

本文对洛伦兹曲线的估计使用的是上面给出的 L_{HC} 模型。通过这些数据我们可以得到洛伦兹曲线上的一系列点 $\{p_i, L(p_i)\}_{i=1}^{m}$,其中 $p_i = \frac{1}{n}\sum_{k \le i} n_k$,代表收入不高于 x_i 的人口占总人口的比例;$L(p_i) = \frac{1}{\theta}\sum_{k \le i} n_k \mu_k$,代表收入不高于 x_i 的人口拥有的收入占总收入的份额;$n = \sum_{k \le i} n_k$ 为总人口数量;$\theta = \sum_{k \le i} n_k \mu_k$ 为收入总额。

(三)估计比较:微观数据与分组数据

我们的逻辑是这样的,首先利用 2006年的微观数据和分组数据同时进行估计,比较所得到的结果,由此判定 $L_{HC}(p)$ 模型的拟合误差大小(适用度)。确认模型适用后,再估计 2000 - 2005年的分组数据,由此得到估计结果。

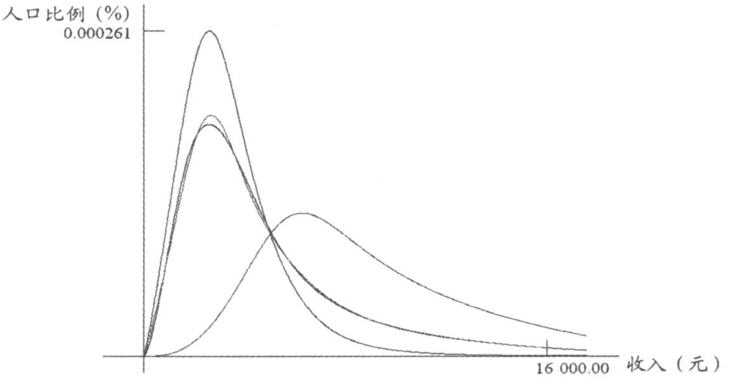
利用 2006年的分组数据 $\{p_i, L(p_i)\}_{i=1}^{m}$ ①我们精确地得到了洛伦兹曲线及其所应通过的各个点,每个点由十字星标记,见图 1;通过这些十字星的曲线则是我们估计所得的三条洛伦兹曲线。按从左到右的顺序,它们依次是湖北省城镇洛伦兹曲线、湖北省农村洛伦兹曲线、湖北省农村与城镇加总的洛伦兹曲线。



说明:按从左到右的顺序,它们依次是湖北省城镇洛伦兹曲线、湖北省农村洛伦兹曲线、湖北省农村与城镇加总的洛伦兹曲线。十字星代表点 $\{(p_i, L(p_i))\}_{i=1}^{m}$。

图 1 洛伦兹曲线:2006年分组数据估计所得

正如前面所描述的,通过估计所得的洛伦兹曲线我们还可以得到 2006年湖北省城镇、农村及加总的收入密度函数曲线 $\hat{f}(y)$,且我们将其与由微观数据得到的密度函数曲线进行了比较,如图 2,可以看出,它们之间的差别非常微小。



说明:按照峰度从高到低的顺序排列,四条曲线依次为:由分组数据估计出的农村密度函数曲线、由分组数据估计出的农村与城镇加总密度函数曲线、由微观数据估计的加总密度函数曲线、由分组数据估计的城镇密度函数曲线。同时由分组数据估计而得的三条密度曲线相交于同一点。

图 2 2006年湖北省密度函数曲线的比较

---

①因篇幅过长,数据未在本文中列出。2006年湖北省城镇数据共 11组,10个收入切分点;农村数据共 41组,40个收入切分点;加总数据有 21组,20个收入切分点。即 m 的取值分别为 10、40、20。若需要具体数据,可来信向作者索取。

进一步,我们还分别利用 2006年湖北省收入抽样调查的微观数据和分组数据计算了各种基尼系数①,见表 1。

表 1　基尼系数:分组数据与微观数据的比较

		农村	城镇	加总	Dagum分解
微观数据		0.307743	0.283625	0.398723	0.39872
分组数据	估计值	0.30802	0.28379	0.39881	0.398447
	Gastwirth上限	0.31564	0.31564	0.41302	-
	Gastwirth下限	0.29553	0.27596	0.39331	-

说明:表中加总基尼系数是指将农村和城镇数据混合起来,按照与估计农村基尼系数和城镇基尼系数同样的方法进行估计所得的基尼系数值。Dagum分解值是用 Dagum(1997)的方法计算得来的。Gastwirth(1972)给出了利用分组数据计算基尼系数的上下界,我们一并计算后在表中列出。

从表 1的结果中可以看出,利用我们提出的洛伦兹曲线模型估计而得的$\hat{L}_{HC}(p)$所计算出的基尼系数与用微观数据计算所得的基尼系数非常接近,且具有相当良好的性质。Gastwirth(1972)推导了由组数据得出的基尼系数的下界和上界。许多作者使用这一边界约束检验洛伦兹模型对分组数据的拟合度(Kakwani, 1980; Schader and Schmid, 1994)。Schader和 Schmid(1994)发现许多传统的洛伦兹模型所得出的基尼系数都超出了 Gastwirth边界。而表 1结果显示利用$\hat{L}_{HC}(p)$计算所得的基尼系数完全位于 Gastwirth(1972)给出的区间之内,也就是说,$L_{HC}(p)$模型的拟合度是非常好的。不仅如此,通过估计结果和样本值的比较可以发现$L_{HC}(p)$模型是相当令人满意的,对此我们在后面做了更为详细的分析。

(四)模型的应用:2000 - 2005年的洛伦兹曲线及福利比较

由 2006年数据估计结果的对比分析我们可以看出,L_{HC}模型有很强的适用性。因此,我们可以利用 L_{HC}模型和已有分组数据分别估计 2000 - 2005年湖北省的农村及城镇洛伦兹曲线$\hat{L}_{HC}(p)$。表 2和表 3分别包含了模型 L_{HC}的运用于湖北省城镇及农村所得的参数估计,在本文的第二部分我们已经证明过,其相应的估计曲线$\hat{L}_{HC}(p)$满足洛伦兹曲线的定义。

表 2　应用 L_{HC}模型对湖北省城镇 2000 - 2006年的参数估计值

	α	β	γ	λ	η	δ
2000	0.728432 (0.046057)	0.713538 (0.000198)	-0.246220 (0.049594)	13.232102 (0.894228)	0.515940 (0.052443)	0.915131 (0.013021)
2001	0.191009 (0.302391)	0.754418 (0.000577)	-0.042013 (0.089961)	13.672288 (3.885227)	1.081989 (0.319756)	0.976570 (0.021557)
2002	1.096916 (2.586658)	0.391307 (0.003684)	0.234238 (3.833136)	1.990190 (37.817108)	0.500009 (2.937611)	0.941731 (1.904116)
2003	1.030929 (1.139716)	0.446695 (0.001697)	0.221654 (2.309999)	6.480230 (39.052144)	0.500084 (1.507872)	0.937870 (0.515036)
2004	0.910000 (0.998162)	0.575116 (0.001023)	0.181458 (1.758942)	5.232855 (17.361422)	0.500000 (1.197925)	0.868708 (0.423010)
2005	0.584552 (1.798190)	0.657606 (0.000485)	0.153323 (1.454249)	3.209703 (13.777163)	0.794872 (1.947371)	0.872801 (0.777146)
2006	0.277595 (2.240302)	0.699033 (0.001841)	0.137049 (1.284744)	2.656201 (6.805046)	0.998767 (2.335806)	0.732291 (0.613113)

注:圆括号中为标准误。

表 4将这一模型用于湖北省 2000 - 2006年各年份的城镇及农村分组数据时所得的$\hat{L}_{HC}(p)$值与样本$L(p)$值进行了比较,列出了拟合残差平方和(由于这个值过小,表中列出的是扩大 100 000倍后的结果)、各估计值的最大拟合残差绝对值及残差绝对值之和。从表 4的统计结果中可以看出拟合的结果是相当好的。农村分组数据所得的各年份最大的拟合残差绝对值只有 0.00083,其中 2004年的拟合残差绝对值之和是最高的,也只有 0.01153;城镇数据相对粗糙,但估计结果也仍然令人满意,统计结果显示,2002年数据所得的

①对于分组数据,可用公式(11)计算基尼系数,其中 $L(p)$取$\hat{L}_{HC}(p)$;至于微观数据,则可采用离散数据情形下求基尼系数的公式来计算:$G = 1 + \frac{1}{N} - \frac{2\{x_N + 2x_{N-1} + 3x_{N-2} + \cdots + Nx_1\}}{N^2\mu}$,其中 N代表所观察的个体总数,$x_1 \le x_2 \le \cdots \le x_N$ 代表个体收入,μ为整体平均收入。也可用公式:$G = \sum_i \sum_j \frac{|x_i - x_j|}{2N^2\mu}$。这两个公式是等价的。

估计误差是最大的,其最大的拟合残差绝对值为 0.01079,拟合残差绝对值之和为 0.03006%。[①]

表 3　　应用 L_{HC}模型对湖北省农村 2000 - 2006年的参数估计值

	α	β	γ	λ	η	δ
2000	0.000000 (0.068752)	0.784170 (0.000032)	-0.116695 (0.022547)	7.807259 (1.897303)	1.219903 (0.074708)	0.988272 (0.002882)
2001	0.000000 (0.098485)	0.793794 (0.000048)	-0.081559 (0.042860)	9.197640 (1.879217)	1.243542 (0.110860)	0.965956 (0.005375)
2002	0.000000 (0.063987)	0.763856 (0.000049)	0.006064 (0.027151)	8.379918 (1.382187)	1.316953 (0.072796)	0.968954 (0.003853)
2003	0.422014 (0.028051)	0.699084 (0.000046)	-0.086292 (0.020024)	12.351521 (2.207097)	0.857054 (0.032807)	0.984766 (0.002356)
2004	0.000000 (0.314737)	0.745716 (0.000046)	0.117833 (0.133316)	3.367106 (3.535737)	1.429981 (0.343437)	0.952616 (0.084462)
2005	0.371173 (0.062699)	0.798274 (0.000028)	-0.138532 (0.025902)	25.153827 (0.917118)	1.005876 (0.068477)	0.972778 (0.003972)
2006	0.439044 (0.153146)	0.686692 (0.000016)	0.141977 (0.092278)	3.021918 (1.204781)	1.025866 (0.165770)	0.912704 (0.058463)

注:圆括号中为标准误。

表 4　　分组数据估计结果与样本值的比较 ($L_{HC}(p)$与$\hat{L}_{HC}(p)$)

	农村			城镇		
	拟合残差平方和 ×100000	最大拟合残差(绝对值)	残差绝对值之和	拟合残差平方和 ×100000	最大拟合残差(绝对值)	残差绝对值之和
2000	0.10492	0.00042	0.00399	0.06047	0.00038	0.00263
2001	0.35356	0.00066	0.00743	0.41161	0.00089	0.00698
2002	0.07214	0.00048	0.00284	16.35177	0.01079	0.03006
2003	0.15193	0.00064	0.00467	10.37903	0.00820	0.02217
2004	0.62694	0.00083	0.01153	2.22359	0.01070	0.00341
2005	0.19155	0.00079	0.00572	0.83339	0.00258	0.00598
2006	0.04579	0.00039	0.00212	0.21898	0.00095	0.00352

最后,为了进行福利比较,我们根据估计的洛伦兹曲线$\hat{L}_{HC}(p)$和公式(11)计算了 2000 - 2006年的湖北省农村及城镇基尼系数,结果见表 5。

表 5　　湖北省 2000 - 2006年的农村及城镇基尼系数

	2000	2001	2002	2003	2004	2005	2006
农村	0.28278	0.28439	0.28759	0.29636	0.30238	0.30892	0.30802
城镇	0.27080	0.27720	0.35298	0.32984	0.27353	0.27655	0.28379

从表 5的结果中可以看出,湖北省 2000 - 2006年农村及城镇的收入不平等程度都不高。从同年比较来看,农村的收入不平等程度略高于城镇。2000 - 2005年以来农村的收入不平等程度逐年增加,2006年略有下降。城镇的收入不平等程度在 2002年急增,但 2003年迅速下降,自 2004年起回落至较低水平,同时 2004 - 2006年又呈缓慢上升趋势。该结果与湖北省统计局给出的结论也是相当接近的。

四、结论

本文引入了一族新的洛伦兹曲线模型,首先将其应用于湖北省 2006年的分组数据,并利用 2006年湖北省收入的微观统计数据对比研究了本模型的适用性问题。研究结果显示,本文提出的洛伦兹曲线模型呈现非常良好的性质,不仅满足洛伦兹曲线的定义,而且具有很高的拟合度。进而,我们将其应用于湖北省 2000 - 2005年的分组数据,通过估计值与样本值的比较,其结果进一步证实了本模型的高度拟合性。同时,我们还得到了以下有关收入不平等方面的结论。

第一个主要结论是,从湖北省 2000年以后的状况来看,无论是城镇还是农村,收入不平等程度均不高。在国内经济持续高速增长的背景下,这不是一个坏消息。第二个结论是,湖北省农村的不平等状况较城镇的不平等状况略差。这一点结论看起来似乎不太符合我们的直觉,我们认为有两种可能原因:一是农村人口迁移外出务工导致的收入差距拉大,二是由于城镇由传统沿袭下来的调查方式所产生的调查误差导致。因为

①限于篇幅,具体的估计结果不在文中列出,如有需要,可来信向作者索取。

城镇住户调查所选取的住户一般均是在国有企业或事业单位务工的职工家庭,忽略了在外资企业、私营企业工作或者自主创业的群体,因此,误差的存在是可能的。第三个结论是,尽管农村和城镇的收入不平等程度并不严重,但也呈现缓慢的上升趋势,尤其是城镇,在 2002年出现突然恶化的情形,这一点说明在经济增长的过程中不平等状况的急速恶化是可能的,因此,在制定和执行各种促进经济增长的政策时,务必兼顾考虑收入不平等的问题,这对于构建和谐社会至关重要。

另外,基于本文提出的方法可以找到更有效的模型,比如对参数向量 τ中的系数赋予更大的取值范围等。本文估计结果的局限之处在于,我们没有系统地讨论湖北省农村与城镇之间的收入不平等以及省内不同地区的收入不平等状况,也没有对其背后的微观机制给出合理的分析与解释。这是我们进一步研究的方向。

附录:定理的证明

对于第一个表述,我们只需要证明$\tilde{L}'(p) \geq 0$,且$\tilde{L}''(p) \geq 0$即可。

因为$\tilde{L}'(p) = \alpha p^{\alpha-1} L(p)^{\eta} + \eta p^{\alpha} L(p)^{\eta-1} L'(p)$,

$$\frac{\tilde{L}''(p)}{p^{\alpha-2} L(p)^{\eta-2}} = \alpha(\alpha-1) L(p)^2 + \alpha\eta p L(p) L'(p) + \eta p^2 [(\eta-1) L'(p)^2 + L(p) L''(p)] + \alpha\eta p L(p) L'(p) \quad (A1.1)$$

显然,当 $\alpha \geq 1$时,$\tilde{L}(p)$为洛伦兹曲线。假定 $\alpha \in [0,1)$,$\tilde{L}(p)$为洛伦兹曲线的充分条件是方程 (A1.1)右端的两项之和非负或者:

$$(\alpha-1) L(p) + \eta p L'(p) \geq 0 \quad (A1.2)$$

这一不等式显然成立,因为左边的式子在 $p=0$时为零,且在 $[0,1]$区间上是递增的。

对于第二个表述,注意在当前假设下,式 (A1.2)仍然成立,因此只要式 (A1.1)的右边两项之和非负即可。将式 (A1.1)右边两项除以 ηp后记作 $g(p)$,我们只需证明对所有的 $p \in [0,1]$:

$$g(p) = p[(\eta-1) L'(p)^2 + L(p) L''(p)] + \alpha L(p) L'(p) \geq 0 \quad (A1.3)$$

对任意的 $\eta \in [1/2, 1)$,且满足 $\alpha + \eta \geq 1$的 $\alpha \geq 0$均成立即可。在给定假设下这一等式是成立的,因为 $g(p)$在 $[0,1]$区间上递增,且 $g(0)=0$。

参考文献:

1. 成邦文:《基于对数正态分布的洛伦兹曲线与基尼系数》,载《数量经济技术经济研究》,2005(2)。
2. 程永宏 等:《利用个人收入分配函数确定基尼系数的新方法》,载《华东经济管理》,1998(1)。
3. 欧阳植 等:《分组数据的收入分布拟合及洛伦兹曲线与基尼系数》,载《数量经济技术经济研究》,1994(6)。
4. Cheong, K. S., 2002. "An Empirical Comparison of Alternative Functional Forms for the Lorenz Curve." Applied Economics Letters, Vol 9, pp. 171 - 176.
5. Chotikapanich, D., 1993. "A Comparison of Alternative Functional Forms for the Lorenz Curve." Economics Letters, Vol 3, pp. 187 - 192.
6. Dagum, C., 1997. "A New Approach to the Decomposition of the Gini Income Inequality Ratio." Empirical Economics, Vol 22(4), pp. 515 - 531.
7. Davies, J. and Hoy, M., 1995. "Making Inequality Comparisons When Lorenz Curves Intersect." The American Economic Review, Vol 85(4), pp. 980 - 986.
8. Gastwirth, J. L., 1972. "The Estimation of the Lorenz Curve and Gini Index." Review of Economics and Statistics, Vol 54, pp. 306 - 316.
9. Kakwani, N. C., 1980. "On a Class of Poverty Measures." Econometrica, Vol 48, pp. 437 - 446.
10. Kakwani, N. C. and Podder, N., 1976. "Efficient Estimation of Lorenz Curve and Associated Inequality Measures from Grouped Observations." Econometrica, Vol 44, pp. 137 - 148.
11. McDonald, J. B. and Xu, Y. J., 1995. "A Generalization of the Beta Distribution with Applications." Journal of Econometrics, Vol 69(2), pp. 427 - 428.
12. Ogwang, T. and Rao, U. L. G., 1996. "A New Functional Form for Approximating the Lorenz Curve." Economics Letters, Vol 52, pp. 21 - 29.
13. Ogwang, T. and Rao, U. L. G., 2000. "Hybrid Models of the Lorenz Curve." Economics Letters, Vol 69, pp. 39 - 44.
14. Ortega, P.; Martin, G.; Fernandez, A.; Ladoux, M. and Garcia, A., 1991. "A New Functional Form for Estimating Lorenz Curves." Review of Income and Wealth, Vol 37, pp. 447 - 452.
15. Ryn, H. K and Slottje, D. J., 1996. "Two Flexible Functional Form Approaches for Approximating the Lorenz Curve." Journal of Econometrics, Vol 72, pp. 251 - 274.
16. Sarabia, J.; Castillo, E. and Slottje, D. J., 1999. "An Ordered Family of Lorenz Curves." Journal of Econometrics, Vol 91, pp. 43 - 60.
17. Sarabia, J.; Castillo, E. and Slottje, D. J., 2001. "An Exponential Family of Lorenz Curves." Southern Economic Journal, Vol 67, pp. 748 - 756.
18. Sarabia, J.; Castillo, E.; Pascual, M. and Sarabia, M., 2005. "Mixture Lorenz Curves." Economics Letters, Vol 89, pp. 89 - 94.
19. Schader, M. and Schmid, F., 1994. "Fitting Parametric Lorenz Curves to Grouped Income Distribution − A Critical Note." Empirical Economics, Vol 19, pp. 361 - 370.
20. Shorrocks, A. F., 1983. "Ranking Income Distributions." Economica, Vol 50, pp. 3 - 17.

(责任编辑:孙永平)