洛伦兹曲线模型研究综述和最新进展

颜节礼[1]，王祖祥[2]
(1.江南大学 商学院 214122；2.武汉大学 经济与管理学院，武汉 430072)

摘 要：洛伦兹曲线是收入分配分析的重要工具。在只有分组数据可用的条件下，估计洛伦兹曲线的方法主要有插值法、经验分布法和洛伦兹曲线模型法。传统洛伦兹曲线模型由于函数形式简单、参数太少，实际效果差强人意。Sarabia、Ogwang等学者相继提出组合模型思想，试图构造函数形式更为一般、参数更多的模型，期望提高精度和可靠性，但仍存在参数数量不够丰富、拟合精度不高等问题。在最新的研究中，Wang等提出了组合乘法加权与加法加权的方法，极大地丰富了洛伦兹曲线的函数形式和参数数量。

关键词：洛伦兹曲线；参数模型，；乘法加权；平衡估计法；收入不平等
中图分类号：F224 **文献标识码**：A **文章编号**：1002-6487(2014)01-0034-06

0 引言

收入分配统计分布是收入分配分析的重要工具，只要有了这种分布，就能对收入分配进行各种深入分析，例如可以测算或比较收入不平等、两极或多极分化、贫困与富裕等，可以进行深入的政策研究，亦可以直观地了解收入分配的动态变迁。国际上收入分配理论与应用研究中，一般依赖于收入分配的统计分布。

在收入分配抽样调查数据可用的条件下，可以使用核估计法构造收入分配的统计分布，理论上可以证明，这种方法依分布收敛于相应收入分配背后的统计分布。亦可以使用经验分布法，即使用一些经验统计分布对抽样数据进行拟合，并确定其中的参数。这两种方法各有不同的特点，究竟哪种更好，理论界还没有取得一致意见。

但实际工作中由于保密或保护隐私的原因，理论与应用分析人员不能获得收入分配抽样调查数据，能得到的可能是所谓的分组数据，这种数据中含有收入组的平均信息，即将人口依收入从低到高分成若干个组，组数一般不多，这种分组数据可以表示为：

$$\{(p_i, L_i)\}_{i=1}^{n} \tag{1}$$

$$\{(p_i, x_i/\mu)\}_{i=1}^{n} \tag{2}$$

其中p_i是第i个低收入端的累计人口比例，L_i是该人口组所拥有总收入的比例，$[0, x_i]$是该人口组所在的收入区间，μ是平均收入，n是组数。n一般不大，通常5到20组。

有时只有数据(1)可用，例如Solt (2009)整理了100多个国家若干年份的分组数据，其中没有数据(2)的信息。我国《统计年鉴》中从城镇可支配收入数据可以推算得到数据(1)，但没有x_i的信息，因此没有数据(2)的信息可用。我国《农村住户调查年鉴》提供了比较完整的分组数据，从中可以推算出数据(1)与数据(2)的相应信息，但遗憾的是该年鉴中高端收入人口的信息不足，这有可能影响分析的可靠性。

当只有数据(1)或数据(2)信息的条件下，一类重要的收入分配统计分配构造方法是洛伦兹曲线法，即利用洛伦兹曲线模型对数据(1)与数据(2)进行拟合，得到近似的洛伦兹曲线，然后利用洛伦兹曲线与相应统计分布的关系得到收入分配统计分布的估计。正是由于我国公开发布的收入分配数据是分组数据，因此洛伦兹曲线模型研究对我国具有特殊意义。实际上，利用分组数据可以大大减少有关部门提供收入分配数据的障碍。另外，分组数据条件下的洛伦兹曲线法也可用于数据有限条件下的其他科学技术领域。国外经济理论工作者对洛伦兹曲线模型进行了很多研究工作，本文对这一领域的研究进展进行综述性介绍。

1 洛伦兹曲线定义

洛伦兹曲线(Lorenz Curve)定义为低收入端人口累计比例p与该组人口拥有的总收入比例$L(p)$之间的函数关系，例如$L(0.1)=0.02$意味着低收入端10%的人口拥有的总收入的比例为2%。通过(0，0)和(1，1)的45°线段为$L(p)=p$，可以理解它为完全平等线，因为这时比例等于p的低收入端拥有总收入的比例等于p，因此对应的收入分配一定是完全平等的。一般，$L(p)$位于完全平等线下方，它离完全平等线越远，这时它与完全平均线所围成的面积越大，则收入不平等程度越大，反之越小。设收入分布的

基金项目：国家社会科学基金资助项目(10BJL015)；教育部社会科学规划基金资助项目(09YJA790152)；武汉大学自主社会科学基金资助项目
作者简介：颜节礼(1972-)，男，陕西咸阳人，硕士，副教授，研究方向：收入分配理论。
王祖祥(1953-)，男，湖北秭归人，博士，教授，研究方向：收入分配理论。

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

密度函数为$f(x)$,分布函数为$F(x)$,平均收入为μ,又设$F(x)$的反函数$F^{-1}(p)$存在,那么洛伦兹曲线可以表示为:

$$L(p)=L\big(F(x)\big)=\frac{1}{\mu}\int_0^x tf(t)\mathrm{d}t=\frac{1}{\mu}\int_0^p F^{-1}(y)\mathrm{d}y \tag{3}$$

其中$p=F(x)$,$x\geqslant 0$是收入数量。可见,洛伦兹曲线满足:

$$L'(p)=\frac{dL(p)}{dx}\frac{dx}{dp}=\frac{x}{\mu} \tag{4}$$

$$L''(p)=\frac{1}{\mu}\frac{dx}{dp}=\frac{1}{\mu f(x)} \tag{5}$$

(3)是Gastwirth (1971)给出的洛伦兹曲线定义式。可见从收入分配的统计分布可以方便地构造相应的洛伦兹曲线,同时由(5)可见,由洛伦兹曲线也可以得到收入分配的密度函数。由洛伦兹曲线的经济意义与(4)、(5)可见$L(p)$应满足条件:

$$L(0)=0,L(1)=1,L'(p)\geqslant 0,L''(p)\geqslant 0 \tag{6}$$

从而$L(p)$为单调增加的凸曲线。

洛伦兹曲线$L(p)$的精确估计在收入分配分析中具有重要意义。例如,由于基尼系数定义为洛伦兹曲线与完全平均线之间面积的2倍,即$G=1-2\int_0^1 L(p)\mathrm{d}p$,可见$G$的估计精度依赖于$L(p)$的准确估计。再例如由(4)可以得到$x=\mu L'(p)$,解此方程可以得到任一收入数量$x$所对应的人口比例$p$,再由(5)可得收入分布密度函数$f(x)=\big[\mu L''(p)\big]^{-1}$。由此,就可以计算很多刻画收入分配侧面的指数,例如Sen贫困指数、Foster指数、Wolfson两极分化指数、Ducols极化指数等等。不仅如此,收入分布统计分布本身包含比这些指数更深入的信息,例如可以用它分析不同群体内的收入分配及其动态变化。

2 洛伦兹曲线估计方法

由于洛伦兹曲线的重要性,国外经济理论界考虑了若干种方法对其进行估计。分组数据条件下洛伦兹曲线的估计方法主要有三种,即插值法、分布函数法、模型法。

2.1 插值法

由于只有(1)或(2)形式的数据可用,即只知道洛伦兹曲线上的若干个点,插值法的基本思想是用适当的曲线段连接分组数据(1)中的相邻两个点,曲线段的选择标准是形成的整条洛伦兹曲线是增且凸的。线性插值法即用线性函数来近似表示每一区间上的洛伦兹曲线段,它忽略了各组内收入不平等程度,显然根据这一曲线估算的基尼系数将明显偏小,用其描述收入不平等时误差可能较大。

Gastwirth (1976)建议采用Hermite插值函数,即用m阶多项式逼近洛伦兹曲线。通过构造在插值点(即区间端点)处函数值相等、函数的一阶导数一直到m阶导数相等共$m+1$个插值条件决定插值函数。最常见的是二次样条(spline)插值和三次样条插值,这种插值整体逼近的效果很好(见Kakwani 1976,王祖祥 2001)。但是,这种插值法得到的曲线可能不满足洛伦兹曲线的凸性条件,由此计算密度函数时可能产生负的函数值,尤其是在收入分布的两端(Datt 1988)。

针对这一问题,Cowell (1982)从收入分布的直方图出发构造密度函数,除了考虑插值点以外,其插值条件还考虑(i) 在插值区间上密度估计$f(x)\geq 0$,(ii) 每个区间上的人口频率等于插值函数的积分。这一插值法需要数据(1)与数据(2)两者的信息。对于高收入组和低收入组,Cowell (1982)采用Pareto分布构造插值函数,而对于中间部分的收入区间仍采用多项式插值函数,然后连接相继的函数段,以此作为密度函数。但是,由于不同收入区间的函数形式不同,导致函数段的连接点选择得不同时,计算结果可能不同,估算过程也比较复杂(Datt 1988)。Ryu (1996)指出这种分段插值的方法只是孤立的考虑了各个区间的信息,并没有考虑到洛伦兹曲线的整体逼近效果。

2.2 经验分布法

这是一种重要的参数估计法。根据事先指定的经验分布来拟合收入分配数据,再从得到的统计分布构造洛伦兹曲线。对于收入分布形态的研究最早可以追溯到意大利经济学家Pareto (1897),他给出了收入分布的Pareto分布函数$N(x)=Ax^{-\alpha}(\alpha>1)$,这里$N(x)$表示收入大于$x$的人口比例,$\alpha$也被称作Pareto指数,也是反映收入平等程度的重要指标,Pareto讨论了α值,收入分配越平等α越大(Bronfenbrenner 1971)。之后Aitchison & Brown (1957)、Fish (1961)、Aiger & Goldberge (1970)、Singh & Maddala (1976)、Basmann (1984)等都利用不同的函数形式来研究收入分布形态,但实践证明,还没有发现一种特别的经验分布能够应对各种各样的收入分布。例如,Basmann (1990)指出在尾部拟合效果理想的函数,在收入的中间区间可能不尽人意,而另外一些函数形式则相反。因此,经济理论界还未找到在整个收入域都有良好逼近效果的经验分布(Slottje 1987)。

2.3 洛伦兹曲线模型法

从洛伦兹曲线条件出发,直接构造参数模型则是另一种方法。Kakwani & Podder (1973)在这一领域做了开创性的研究。先直接寻找满足洛伦兹条件(6)的参数模型,用其拟合数据(1),他们尝试的模型为:

$$L(p)=p^{\alpha}e^{-\beta(1-p)} \tag{7}$$

用(7)对澳大利亚1967–68消费者支出数据进行拟合(使用非线性最小二乘法),发现估计效果良好。Kakwani & Podder (1976)又提出了KP变换。在洛仑兹曲线坐标平面上,洛仑兹曲线的端点(0,0)保持不变,将坐标逆时针旋转45度,以完全平均线为横坐标,给出了新的模型,即$s=ar^{\alpha}\big(\sqrt{2}-r\big)^{\beta}$,其中$r=\big(p+L(p)\big)\big/\sqrt{2}$,$r\in[0,\sqrt{2}]$,坐标旋转的目的是克服洛伦兹曲线在接近点(1,1)时导数趋近于无穷大给积分带来的困难。Kakwani & Podder (1980)在研究贫困度量问题时再次采用了新的曲线模型:

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$L(p)=p-ap^{\alpha}(1-p)^{\beta} \tag{8}$$

发现这一模型的整体估计效果较之前出现的模型更优。这一模型也被后来S. Chen et al. (2000)在世界银行有关机构建立的“Program for Calculating Poverty Measures from Grouped Data” (POVCAL)软件中所应用，很多学者在收入不平等与贫困分析中应用了这一软件。

POVCAL选择的另一个模型是Villasenor & Arnold (1989)的二次式模型。Villasenor & Arnold通过要求二次型$aL^2+bLp+cp^2+dp+eL+f=0$的参数满足条件(6)来确定洛伦兹曲线模型，显然，这样确定的洛伦兹曲线模型可能是抛物线、椭圆、双曲线上的一段，因此称得到的模型为椭圆型洛伦兹曲线(Elliptical Lorenz curve)，最后得到的模型至多含三个参数，并导出了相应的密度函数表达式。该文最后的数值实证说明，对实际数据的逼近效果尚称满意。

Rasche et al. (1980)在评价Kakwani & Podder (1976)的模型时指出，该方法得到的曲线不满足洛伦兹曲线条件(6)，求导数后可以发现Kakwani & Podder (1980)的模型同样也不满足洛伦兹条件。尽管根据Kakwani & Podder (1980)的模型估计的基尼系数效果比较满意，但测算贫困时误差较大(王祖祥 2006)。Rasche et al. 建议重新回到洛伦兹曲线条件，从这一条件出发构造一般的洛伦兹曲线模型，文中给出了模型$L(p)=[1-(1-p)^{\alpha}]^{\beta}$，显然当$\alpha=\beta=1$时即为完全平均线，且$\alpha<1$，$\beta=1$时就是Pareto收入分布对应的洛伦兹曲线。

Gupta (1984)认为Kakwani & Podder (1980)、Rasche et al. (1980)等模型为非参数线性模型，其估计方法比较复杂，他给出了形式更为简单的模型$L(p)=pA^{p-1}$，通过对数变换可转变为线性模型，用普通最小二乘法(OLS)估计更为简单，但是这一模型其实只是Kakwani & Podder (1973)模型的特殊形式。

由于发现上述这些模型同样不能在整个收入区间内得到令人满意的拟合效果，Ortega et al. (1991)提出了模型：

$$L(p)=p^{a}\left(1-(1-p)^{\beta}\right)$$

并给出了相应基尼系数、K系数(Kakwani 1980)、I系数(Chakravarty 1988)的参数表达式，通过对21个国家和西班牙50个省的数据进行拟合，并与Pareto、K&P (1973)、Rasche (1980)、Gupta (1984)模型估计的洛伦兹曲线残差平方和(SSR)比较，指出这一模型在收入数据组数较少、组距较大时仍然有良好的拟合效果。Basmann et al. (1990)从Kakwani (1973)的模型出发，给出了更为一般的模型，其包含四个参数：

$$L(p)=p^{ap+b}e^{-g(1-p^2)-h(1-p)} \tag{9}$$

通过将反映洛伦兹曲线拟合效果的F-检验值以及该模型对应的基尼系数与之前模型估计结果进行比较，认为该模型效果更为理想。稍后Basmann et al. (1993)利用这一模型讨论了美国的收入分配，并形成了专著。实际上，(9)只是(7)的简单推广，拟合效果其实并不令人满意，在下文将进行进一步的讨论。

Chotikapanich (1993)给出了一种单参数模型：

$$L_{\lambda}(p)=\frac{e^{\lambda p}-1}{e^{\lambda}-1}$$

这一模型具备良好的数学性质，稍后将看到利用这一模型做为基本构件的更一般的模型。Ogwang (1996)认为文献中给出的洛伦兹曲线模型的基尼系数参数表达式都很复杂，他给出了一个洛伦兹曲线的隐函数形式，认为洛伦兹曲线可以看作圆上的一段弧，关键是找到合适的圆心和半径。通过洛伦兹曲线上的点与洛伦兹曲线的两个端点构成的夹角余弦值，采用普通最小二乘法就可以得到洛伦兹曲线，并给出相对简单的基尼系数参数表达式。但这种方法可以看做是前文Villasenor & Arnold (1989)椭圆型洛伦兹曲线的特例。

Gastwirth (1972)研究认为对于分组数据而言，无论收入分布形态如何，其基尼系数一定存在下限与上限，即Gastwirth区间，显然，若根据拟合的洛仑兹曲线计算的基尼系数高于Gastwirth上限或低于下限都是不合理的。Schader & Schmid (1994)对当时文献中出现的所有洛伦兹曲线参数模型进行了综述性介绍，并考虑了联邦德国1950–1988年间16年的数据，计算了洛伦兹曲线与基尼系数，并将基尼系数的计算结果与Gastwirth区间进行了比较，发现基尼系数可能低于下限或高于上限，因此认为这些模型的精确度和可靠性不能令人满意。Cheong (2002)也做了类似研究，通过对文献中洛伦兹曲线回归拟合优度R^2进行比较，认为只有Rasche et al. (1980)、Kakwani & Podder (1980)模型的结果接近1。而对于概率分布的估计，不同模型在收入高、低不同组的效果不同。对于模型估计的精确度和可靠性不能令人满意的原因，总结起来大致有两条：一是模型都比较简单，一般只有两三个参数；二是这其中大部分模型并不严格满足洛伦兹条件(6)，真正严格满足洛伦兹条件的只有Ortega (1991)、Rasche (1980)、Basmann (1990)、L_{λ}等四个模型。Rossi (1985)、Basmann et al. (1990)、Ryu & Slottje (1996)都曾进行类似研究，认为仅少数模型严格满足洛伦兹曲线的条件。

3 洛伦兹模型的组合思想和最新研究进展

3.1 洛伦兹模型的组合思想

从上述提到的模型可见，无论是从分布函数出发还是直接构造洛伦兹曲线模型，简单的函数形式一般难以满足收入分配计量的可靠性和精度要求。Schader & Schmid (1994)推广了之前的模型，提出四个参数的模型：$L=p^{\gamma}-ap^{\alpha}(1-p)^{\beta}$，显然，Otega (1991)和K&P (1980)是这一模型的特殊情形。Schader & Schmid (1994)用联邦德国的16组数据验证了根据这一模型所计算的全部基尼系数都通过了Gastwirth区间检验。这一推广说明，增加模型的参数或对模型进行组合能够提高模型的拟合精度和可靠性。

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

在之后的发展过程中，首先是Ryu & Slottje (1996)提出了利用多项式组合模型逼近的思想。他们采用两种组合函数方法：(ⅰ)通过指数多项式拟合收入密度函数，从收入分布的反函数得到洛伦兹曲线。(ⅱ)直接以Bernstein多项式拟合洛伦兹曲线。这种组合方法的优点一是得到的函数在整个收入域都有较好的拟合效果，二是产生的函数满足洛伦兹曲线的条件。Ryu认为以此方法得到的收入密度函数对于贫困计量的效果优于传统的简单函数。Sarabia (1999)继承了这一思想，并给出了洛伦兹曲线的更为一般的构造方法，从该文给出的两个定理可以推知：若$L(p)$满足洛伦兹曲线的条件，那么当参数满足条件$\alpha,\ \gamma\geqslant 1$ 或 $0\leqslant\alpha<1,\gamma\geqslant 1,L''(p)>0$ 时，组合模型$\tilde{L}(p)=p^{\alpha}L(p)^{\gamma}$也满足洛伦兹曲线的条件。但Sarabia并未给出定理的严格证明，对于参数满足的条件也没有进一步研究，可见参数的约束条件也比较复杂。

Ogwang & Rao (2000)提出了加法模型和乘法模型(additive models and the multiplicative models)。加法模型即算术加权，即凸组合，乘法模型即乘法加权，类似于Cobb-Douglas生产函数。他们指出，若$L_1(p),L_2(p)$满足洛伦兹条件，则$\tilde{L}(p)=\delta L_1(p)+(1-\delta)L_2(p),(0\leqslant\delta\leqslant 1)$和$\tilde{L}(p)=L_1^{\gamma}(p)L_2^{\lambda}(p)(\lambda、\gamma\geqslant 1)$必然满足洛伦兹条件(2.4)。显然，两种组合法的参数个数都增加了，与传统的估计方法相同，也可以用非线性最小二乘法(NLS)进行参数估计。为了验证这种组合方法的有效性，Ogwang & Rao (2000)使用了Basmann et al. (1990)研究中引用的1977年美国收入数据，结果表明组合模型参数估计的标准差要小于分别采用两个分量函数估计的标准差，模型的回归估计标准误(SSE)也小于分别采用两个分量函数估计的结果，其它形式的组合模型都有类似的效果。

3.2 洛伦兹模型的最新研究进展

考察上述组合方法：一是构造的模型数量及参数仍比较有限，对参数应满足的约束条件研究也不充分；二是由于这些模型都比较简单，采用最小二乘估计时仅利用了分组数据(1)信息，而并没有利用数据(2)信息。王祖祥等(Wang et al. 2009，2011)在最近的两篇文章中对组合函数模型进行了研究，他们主要做了三方面的工作。

一是在函数模型构造方面。Wang et al. (2009)在研究中国农村居民收入不平等和贫困计量时，将Sarabia组合思想和Ogwang & Rao (2000)加法组合与乘法组合思想结合起来，利用Pareto模型和L_{λ}模型为分量函数构造了几个新的洛伦兹曲线模型，其中参数个数达到6个，大大增加了模型的灵活性，以此对中国农村居民1980~2006间部分年份的洛伦兹曲线和贫困指数进行了估计。在Wang et al. (2011)中，首先对Ogwang & Rao (2000)的乘法加权模型的参数与分量函数应满足的条件给予了证明，在此基础上，将两个函数的组合拓展到任意有限多个函数的乘法加权模型，即$\tilde{L}(p)=L_1(p)^{\alpha_1}L_2(p)^{\alpha_2}\cdots L_m(p)^{\alpha_m}$，并严格证明了其中分量函数$L_i(p)$与参数满足什么条件时$\tilde{L}(p)$将满足洛伦兹曲线条件(6)，指出$\tilde{L}(p)$中任何分量模型$L_i(p)$的二阶导数与一阶导数的比$L_i''/L_i'$单调递增时，$\tilde{L}(p)$具有最大的灵活性。该文还找到了一组所谓的广义帕累托(GP，generalized Pareto)模型，指出这组模型中的任何有限个模型的凸组合L满足L''/L'单调增条件，以此即得到了成千上万种洛伦兹曲线模型。

与此同时，在参数估计方面，传统估计方法只考虑使用分组数据(1)，使用非线性最小二乘法确定洛伦兹曲线模型的参数，即通过取

$$\sum_{i=1}^{n}\left(\hat{L}(p_i)-L_i\right)^2=\min \tag{10}$$

来确定$L(p)$中参数，这种最小化显然没有利用数据(1)信息。Wang et al. (2011)首次提出了平衡拟合(Balanced fit)的概念，即通过最小化

$$b\sum_{i=1}^{n}\left(\hat{L}(p_i)-L_i\right)^2+(1-b)\sum_{i=1}^{n}\left(\hat{F}(x_i)-p_i\right)^2 \tag{11}$$

来确定$L(p)$中参数，其中$b\in[0,1]$，$\hat{F}(x_i)$是方程$\mu L'(p_i)=x_i$的解，且是数据(1)与(2)对应的分布函数在$x=x_i$处的近似值。显然(11)充分利用了给定的数据信息，有可能得到更理想的洛伦兹曲线估计。当$b=1$时即传统的估计方法，这时标志着应用工作者仅对洛伦兹曲线的精度感兴趣；而$b=0$时对应一种新的确定洛伦兹曲线的估计方法，此时意味着应用人员仅在意分布函数的精度；而当$b\in(0,1)$时说明应用人员既在意洛伦兹曲线的精度，也关心收入分布函数的精度。调整b的大小还可以反映应用人员对两种精度不同的重视程度。

Wang et al. (2011)中的模型看起来都非常复杂，但该文指出，这些模型都可以使用无约束非线性最小二乘法(UNLS)进行参数估计，而对这种优化问题已经存在非常有效的计算方法。其中关键是参数变换，例如对于模型

$$p^{\alpha}\left\{\delta_1\left[1-(1-p)^{\beta}\right]+\delta_2 L_{\lambda}(p)+\delta_3\left[1-(1-p)^{\gamma}\right]\right\}^{v} \tag{12}$$

其中参数$\delta_1、\delta_2、\delta_3\in[0,1]$且$\delta_1+\delta_2+\delta_3=1$，可使用参数变换

$$\delta_1=\sin^2\theta_1,\delta_2=\cos^2\theta_1\sin^2\theta_2,\delta_3=\cos^2\theta_1\cos^2\theta_2$$

这时θ_1与θ_2是无约束的，无论它们如何变化，有关$\delta_1,\ \delta_2,\ \delta_3$的条件都将得到满足。

3.3 组合模型的应用

为了检验上述组合模型和参数估计方法在应用中的有效性，Wang et al. (2011)根据GP模型构造了数个洛伦兹曲线模型，通过对两组数据拟合的结果与核估计法的结果进行比较，结果发现使用分组数据与组合模型的估计结果甚至优于使用抽样数据的核估计，对分布函数的估计误差也优于核估计。两组数据分别为美国收入分配数据(Current Population Survey，CPS)和2006年湖北省收入调查数据。其评价模型优劣的统计量选择了Sarabia et al. (1999，2001)使用的标准：即估计值$\hat{L}(p_i)$与观察值L_i之间的最大绝对偏差MAXABS、均方差MSE、平均绝对差MAE、以及

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

基尼系数的误差。

美国CPS收入数据(1977~1983)年被众多文献运用，比如 Basmann (1990，1991)、Ryu and Slottje (1996) 、Sarabia et al. (1999，2005)等。这一数据按照人口百分位数分组，形如$\{(p_i, L_i)\}_{i=1}^{99}$，每年数据对应着洛伦兹曲线上99各点。Wang et al. (2011)根据乘法加权组合方法构造了七个洛伦兹曲线模型，其中包括形如前文(12)的模型，还有如模型：

$$\left[1-L_{\lambda_1}(1-p)^{\beta_1}\right]^{\alpha}\left\{\delta p+(1-\delta)\left[1-\left(1-L_{\lambda_2}(p)\right)^{\beta_2}\right]\right\}^{\nu} \quad (13)$$

$$\left[\delta p+(1-\delta)L_{\lambda}(p)\right]^{\alpha}\left\{\delta_1\left[1-L_{\lambda_1}(1-p)^{\beta_1}\right]+(1-\delta_1)L_{\lambda_0}(p)\right\}^{\nu} \quad (14)$$

由于CPS数据不具备(2)的信息，因此参数估计时取$b=1$，对1977年的数据进行了洛伦兹曲线和基尼系数的估计，表1列出了该文七种不同模型估计精度值。

表1　　Wang et al. (2011)组合模型估计精度

Model	M1 (4.3)式	M2 (4.4)式	M3 (4.5)式	M4	M5	M6	M7
MSE × 10^6	0.0291	0.0298	0.0067	0.3835	0.0246	0.0308	0.3929
MAE	0.0001	0.0002	0.0001	0.0005	0.0001	0.0001	0.0005
MAXABS	0.0006	0.0005	0.0002	0.0028	0.0005	0.0007	0.0026
Gini	0.3683	0.3683	0.3683	0.3685	0.3683	0.3683	0.3681
估计标准差	0.0225	0.0210	0.0211	0.0216	0.0211	0.0229	0.0222

注：根据Wang et al. (2011)整理，Gini系数下面一行数字为Gini系数估计标准差。

Basmann (1990)针对1977年CPS数据，采用了(9)中的函数形式以及不同参数限制条件下的5种变形模型。在文中其评价模型优劣的标准采用了$\hat{L}(p_i)$的F统计量值和拟合优度R^2。为了将两者结果相比较，参阅 Basmann (1990)中给出的6种传统洛伦兹曲线模型估计值$\hat{L}(p_i)$，我们据此计算了6种不同模型的$\hat{L}(p_i)$估计精度值，如表2(前6个模型)。Sarabia et al. (1999)根据其提出的组合模型思想，以帕累托函数为基础组合了三个模型，同样对这一数据进行了估计，其估计精度见表2后三列。表格最后一行为根据相应模型计算出的基尼系数。

对比表1、表2发现，仅以MSE为例，Wang et al. (2001)组合模型的估计误差都在10^{-8} ~ 10^{-7}数量级，而Basmann (1990)传统模型对这一数据估计结果误差一般在10^{-2}数量级(这一计算结果也符合Sarabia et al. 2005 对传统模型估计精度的计算)，Ogwang & Rao (2000)根据其所给的组合模型对这一数据估计的MSE也在10^{-4} ~ 10^{-3}数量级之间，对比显示出乘法加权模型的估计精度不但优于传统模型，同时也较Ogwang & Rao组合模型精度要高。观察表1、表2中的MAE与MAXABS也能得到同样的结论。

考察Sarabia et al.的组合模型，Sarabia et al. (1999)利用组合模型对瑞典和巴西的估计、2001年、2002年先后用组合模型对19个国家洛伦兹曲线的估计，其估计结果均显示，$\hat{L}(p_i)$估计值的MSE一般落在10^{-4} ~ 10^{-2}数量级，这些研究案例都显示乘法加权组合模型对$\hat{L}(p_i)$的估计精度要优于之前文献中的模型；

进一步考察组合模型对基尼系数估计的精度，据Cheong (2002)估计，1977年美国基尼系数的经验值为0.3682，作者为了检验所给出的6种模型对基尼系数估计的精度，对美国1977年数据基尼系数进行了估计，表3列示了Cheong (2002)计算的基尼系数的结果。我们比较表1与表2、表3中基尼系数的计算结果，发现Wang et al. (2011)中组合模型结果更接近于0.3682。Wang文中还根据Efron & Tibshirani(1993)的Bootstrap法产生随机样本的方法(重复随机样本200次)，估计了每个模型对应的基尼系数标准差，标准差都在0.0210 ~ 0.0229之间，显示出该模型构造方法对基尼系数估计精度比之前模型理想。

表3　　Cheong (2002)计算的美国1977基尼系数

Model	Basmann et al. (1993)	Chotikapanich (1993)	Kakwani (1980)	Kakwani & Podder(1973)	Ortega et al. (1991)	Rasche et al. (1980)
Gini	0.3600	0.3611	0.3684	0.3681	0.3695	0.369

注：根据Kwang Soo Cheong (2002)文整理.

Wang et al. (2011)运用的第二组数据为2006湖北省城乡收入抽样数据。先将抽样数据按照收入等间隔划分成城乡各11个组，得到兼有数据(1)、(2)信息的分组数据，利用平衡估计法，b分别取1、0.5和0，该文使用了六种组合模型。表4仅以$b=0.5$为例，展示了六种模型$\hat{L}(p_i)$的估计精度。对于$\hat{L}(p_i)$的估计，几种模型的MSE都在10^{-5}~10^{-7}的数量级，MAE也在10^{-4}~10^{-3}的数量级($b=1$与$b=0$时对应的结果类似，详见Wang et al. 2011)，在之前文献出现的模型都没有达到过这样的估计精度。

表4　　组合模型对湖北2006城乡收入数据$\hat{L}(p_i)$估计精度

模型	MSE × 10^5	MAE	MAXABS
M1(文4.3)	0.0322	0.0005	0.0008
M2	4.0576	0.0050	0.0113
M3	0.0367	0.0005	0.0010
M4	0.0479	0.0006	0.0011
M5	2.9235	0.0043	0.0092
M6	0.0378	0.0005	0.0010

尤其是对于$\hat{F}(x)$的估计，该文运用平衡估计法对城乡收入分布函数进行了估计。表5比较了平衡估计法和核估计法得到的$\hat{F}_i$的精度，其组合模型采用前文(14)函数形式。结果显示，多数情况下，兼顾洛伦兹曲线和分布函数的平衡估计法比核估计的精度更好。该文分别列出了对于在城乡收入分组后每组的真实值L_i和F_i，同样的结果显示，相对于核估计而言，平衡估计法得到的无论是$\hat{L}_i$还是$\hat{F}_i$估计值，都更接近真值，在11个组中仅有一两组例外。

表2　　Basmann (1990)传统模型、Sarabia et al. (1999)组合模型估计精度

Model	H_1	H_2	H_3	H_4	H_5	H_6	S1	S2	S3
MSE	0.0107	0.0119	0.0105	0.0137	0.0231	0.0267	5.14×10^{-6}	1.5×10^{-6}	1.97×10^{-6}
MAE	0.0059	0.0066	0.0058	0.0074	0.0148	0.0212	0.00189	0.00078	0.0098
MAXABS	0.0300	0.0323	0.0296	0.0352	0.0488	0.0438	0.00494	0.00331	0.00379
Gini	0.36	0.36	0.36	0.36	0.34	0.40	0.3692	0.3693	0.3698

注：H_1 ~ H_6为Basmann使用的传统模型、S_1 ~ S_3为Sarabia使用的组合模型。

4　结语

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表5　　平衡估计和核估计对$\hat{F}_i$估计的比较

		Balanced Fit			
		b=1	b=0.5	b=0	Kernel
Urban	MSE × 10^5	0.9789	0.1673	0.1576	4.8196
	MAE	0.0024	0.0010	0.0009	0.0054
	MAXABS	0.0073	0.0027	0.0027	0.0125
Rural	MSE × 10^5	2.2934	0.0621	0.0290	1.4081
	MAE	0.0029	0.0007	0.0004	0.0030
	MAXABS	0.0129	0.0014	0.0010	0.0081

注：根据Wang et al. (2011)文整理.

如前文所述，洛伦兹曲线和收入分布函数的计量与测度在收入分配问题研究中具有非常重要的作用。相对于收入分配问题在经济学领域和社会学领域的重要性而言，显然对于其测度方法再多的研究都是很有必要的。回顾对于洛伦兹曲线的探索，也经历了一个函数形式由简单到复杂、参数数量也逐渐增加的历程。但无论函数形式多么复杂，满足洛伦兹条件是其基本要求，这也是发展洛伦兹曲线模型必须遵循的准则。而就目前的最新进展而言，Wang et al. (2011)提出的这一洛伦兹曲线的组合思想不仅极大的丰富了函数形式，对于其参数限制也进行了论证，使其严格满足洛伦兹条件。其提出的平衡估计法思想使得研究能够同时兼顾洛伦兹曲线和收入分布函数。但是，平衡估计法是否具有普适性仍然需要用更多的不同空间和时间的数据进行检验，比较其与其它估计方法如核估计等的优劣。同时找出更多GP模型也是对函数组合思想的进一步充实和完善，毕竟，该文中给出的GP函数只是组合洛伦兹曲线的充分非必要条件，那么找出更多的分量函数，根据不同参数约束构造洛伦兹曲线仍有广阔的研究空间。

参考文献：

[1]Bronfenbrenner, M. Income Distribution Theory[M].方敏等译本.北京：华夏出版社，2001.

[2]Gastwirth, J.L. The Estimation of the Lorenz Curve and Gini Index[J]. Review of Economics and Statistics，1972，(54).

[3]Kakwani, N.C. , N. Podder . On the Estimation of Lorenz Curves from Grouped Observations[J]. International Economic Review，1973 ，(14).

[4]Kakwani, N.C. On a Class of Poverty Measures[J].Econometrica，1980，(48).

[5]Gupta, M.R. Functional form for Estimating the Lorenz Curve[J]. Econometica，1984，(52).

[6]Kakwani, N.C. , N. Podder .Efficient Estimation of Lorenz Curve and Associated Inequality Measures from Grouped Observatons[J]. Review of Economics Studies，1976，(43).

[7]Cowell, F.A. , F. Mehta .The Estimation and Interpolation of Inequality Measures[J]. Review of Economic Studies ，1972，(49).

[8]Chotikapanich, D. A Comparison of Alternative Functional Forms for the Lorenz Curve[J]. Economics Letters，1993，(3).

[9]Schader, F. Schmid .Fitting Parametric Lorenz Curves to Grouped Income Distrbutions-A Critical Note[J]. Empirical Economics，1994，(19).

[10]Datt, G. Computational Tools for Poverty Mesurement and Analysis [C]. International Food Policy Research Institute Food Consumption and Nutrition Division (FCND) Discussion Paper NO50，1998.

[11]Sarabia, J.M. et al. An Ordered Family of Lorenz Curves[J]. Journal of Econometrics，1999，(91).

[12]Sarabia, J.M. et al. An Exponential Family of Lorenz Cruves[J]. Southern Economic Joural，2001，(61).

[13]Sarabia, J.M. , M. Pascual A Class of Lorenz Curves Based on Linear Exponential Loss Fuctions[J]. Communications in Statistics – theory and Methods ，2002，(31).

[14]Sarabia, J.M. at el. Mixture Lorenz curves[J]. Economics Letters，2005，(89).

[15]Cheong, K.S. An Empirical Comparison of Alternative Functional Forms for the Lorenz Curve[J]. Applied Economics Letters，2002，(9).

[16]Ogwang, T , U.L. Gouranga Rao .Hybird Models of the Lorenz Curve [J]. Economics Letters,2000,(69).

[17]Shorrocks, A. , G. Wan .Ungrouping Income Distributions: Synthesising Samples for Inequality and Poverty Analysis[C]. World Institute for Development Economics Research (WIDER), Research Paper No16，2008.

[18]Wang, ZX. et al. A New Ordered Family of Lorenz Curves with an Application to Measuring Income Inequality and Poverty in Rural China[J]. China Economic Review,2009，(20).

[19]Wang, ZX. et al. A General Method to Create Lorenz Models, Monash University Business and Economics[C]. Discussion Paper, 2009.

[20]Wang, ZX. et al. A General Method for Creating Lorenz Models[J]. Reviews of Income and wealth,2011，(57).

（责任编辑/亦　民）

© 1994-2014 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net