

参赛密码 \_\_\_\_\_  
(由组委会填写)

## 第十届华为杯全国研究生数学建模竞赛

学 校 国防科学技术大学

---

参赛队号 90002118

---

1.张志峰

---

队员姓名 2.王 吉

---

3.谭索怡

---

参赛密码 \_\_\_\_\_

(由组委会填写)



## 第十届华为杯全国研究生数学建模竞赛

### 题目 中等收入定位与人口度量模型研究

#### 摘 要：

监控收入分配格局的变化是经济社会发展的重要课题，其中中等收入人口比重是反映收入分配格局的重要指标。因此，对中等收入定位与人口度量方法的研究具有重要意义。本文针对中等收入定位与人口度量问题进行建模和求解；提出了满足洛伦兹曲线特征的新的洛伦兹曲线模型，基于两极分化指标改进了人口空间法，最后建立了一个新的中等收入人口界定模型。

首先，本文在经典洛伦兹曲线模型的基础上，构造了新的六参数洛伦兹曲线模型，并证明了其满足洛伦兹曲线特征，反映经济规律。在与 10 个已有模型的拟合精度比较中，其综合得分排在第 2 位。

接着，本文对传统的收入空间法与人口空间法进行了改进。对于收入空间法，本文提出了基于比例的界定方法与基于统计指标的界定方法，并分析了这两种改进方法的优缺点，进而提出了一种综合的收入空间界定方法。对于人口空间法，本文根据两极分化程度越严重则中等收入人口越小这一经济直观，分析了两极分化程度与中等收入人口规模之间的定量关系，提出了一种基于两极分化指标的改进人口空间法。

在前两步的基础上，本文对 A、B 两个地区前后两年的收入分配情况进行了定性、定量分析，发现 A 地区中等收入人口比重在减少，收入比重也在减少；而 B 地区的中等收入人口比重与收入比重均在增加。B 地区相较 A 地区的经济发展情况更为成熟、稳定。

最后，本文分析了大多数中等收入界定方法普遍存在着的“伪中等收入问题”，提出了一个新的考虑到个人生活质量的中等收入人口界定模型。该模型综

合考虑个人收入水平中等化评价指数  $Z$  和个人生活质量评价指数  $H$  对中等收入人口界定的影响，提出个人中等收入吻合度指数  $R$  来度量个人与中等收入人口特征的吻合程度，然后根据该指数的值来界定是否属于中等收入人口范畴。

**关键词：**洛伦兹曲线；两极分化指标；中等收入人口比重.

## 目录

一、问题重述.....	5
二、问题分析.....	5
2.1 洛伦兹曲线及其基尼系数计算.....	5
2.2 中等收入概念及界定方法研究.....	9
2.3 本文拟完成的工作.....	10
三、模型假设.....	11
四、符号说明.....	11
五、模型的建立与求解.....	12
5.1 洛伦兹曲线模型.....	12
5.2 中等收入人口界定方法.....	16
5.3 A B 两地区中等收入情况分析.....	21
5.4 一种新的中等收入模型.....	25
六、结果分析.....	28
七、模型评价与拓展.....	29
八、参考文献.....	29

## 一、 问题重述

我国处于经济转型期，收入分配格局处于重要的调整期，因此，监控收入分配格局的变化是经济社会发展的重要课题。自 2002 年 11 月党的十六次代表大会的报告中提出“扩大中等收入者比重”的战略举措以来，人们对中等收入的议论较多。到目前为止，“橄榄型”收入分配格局正处于形成阶段，理论界并没有形成统一的看法，官方也还没有给出权威性的解释。因而对中等收入的含义和界定中等收入的方法进行充分的讨论，对于理解扩大中等收入者比重的意义和落实扩大中等收入者比重的战略措施是很有必要的。

问题 1：构造满足洛伦兹曲线特征的新模型，使其能很好的拟合题中所给的分组数据、反映经济规律。

问题 2：（1）研究可否改进题目中提到的收入空间法，这时需要研究确定中等收入的范围、中等收入人口的范围的科学方法，以克服中等收入区间取法的任意性；（2）研究可否改进题目中提到的人口空间法，例如研究在各年中与取不同的值时，纵向比较各年中等收入人口与收入的变动的方法。

问题 3：利用题中表二~表五所附 A, B 两个地区前后两个不同年份的收入分配分组数据，研究：（1）对各地区、各年份的中等收入的数量（范围）、中等收入人口的数量或范围进行定量描述，说明中等收入人口的变化趋势；（2）比较两个地区的中等收入人口、收入等变化情况。

问题 4：除题中所述方法外，提出一个新的中等收入人口的定义、原理及经济学意义，并提出与之相应的中等收入人口的测算方法、模型或指数，说明其经济学意义。

## 二、 问题分析

### 2.1 洛伦兹曲线及基尼系数计算

关于收入差距的测度问题，是美国统计学家洛伦兹（Lorenz. M, 1905）<sup>[1]</sup>在其论文《财富集中度的测量方法》中最先提出的，即洛伦兹曲线。洛伦兹曲线把分析结果用一个简单明了的直角坐标图形表示出来，从曲线上可以直观地看出每个阶层的收入比重，从曲线的弯曲度可以观察到各个阶层的收入差别情况，通过对比不同的曲线了解不同国度总收入分配差别程度或同一国家不同时期的收入差别变动情况。离 45 度线越远、离 90 度线越近的曲线表示的收入差别程度越大。洛伦兹曲线作为一种分析收入分配不平等的方法，是国际上用来衡量一个国家或地区经济部门专门化（集中化）程度及国民收入分配不平等（贫富分化）程度的坐标图形，具有一定的趋势性。

收入分配的洛伦兹曲线  $L(p)$  在收入分配分析中具有重要地位,  $L(p)$  表示人口份额等于  $p$  的低收入端拥有的总收入份额, 因此  $L(p)$  是定义在区间  $(0, 1)$  上的函数, 按经济意义, 有效的洛伦兹曲线必须具有以下特征:

(1)  $L(0)=0, L(1)=1$

(2)  $L(p) \geq 0, L'(p) \geq 0, L''(p) \geq 0; p \in (0, 1)$

条件(1)是边界条件, 即 0%的人口拥有 0%的收入, 100%的人口拥有 100%的收入。

条件(2)表示洛伦兹曲线是非负的单调递增的凸函数。 $L(p) \geq 0$  因为收入不能是负额;  $L(p)$  是  $p$  的增函数, 即有  $L'(p) \geq 0$ , 因为所考虑的低收入端人口份额  $p$  越大, 该群体拥有的总收入份额应越大;  $L(p)$  是凸函数, 即满足  $L''(p) \geq 0$ 。因为  $p$  增加到  $p + \Delta p$  时, 人口份额所代表的是收入更高的群体, 因此  $p$  增加时,  $L(p)$  应以更大比例增加。如果对于任何  $p \in (0, 1)$  都有  $L(p) = p$ , 则此洛伦兹曲线是所谓平等收入线。对于任何洛伦兹曲线  $L(p)$ , 基尼系数定义为  $L(p)$  与平等收入线之间面积的 2 倍。

关于洛伦兹曲线与基尼系数的分析与测算, 目前方法主要有三种: 几何算法<sup>[2]</sup>、分布函数法<sup>[2]</sup> 和曲线拟合法。

### 几何算法

几何算法是根据分组数据描述洛伦兹曲线, 按几何图形分块计算基尼系数是根据分组数据刻画洛伦兹曲线与对角线之间的面积。

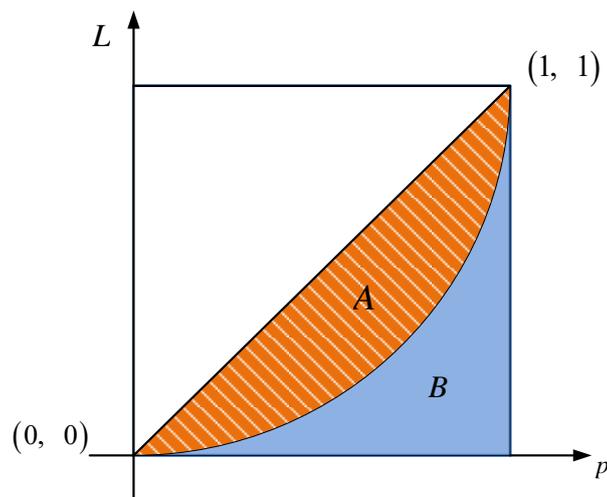


图 1 几何算法计算基尼系数

作为洛伦兹曲线拟合的“副产品”, 基尼系数可很容易地计算出来。按几何

图形分块计算基尼系数（见图 1），可以表示为：

$$G = \frac{A}{A+B} = \frac{A}{A+0.5-A} = 2A$$

然而，几何算法计算精度较差，是一种相对比较粗糙的算法。

### 分布函数法

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<sup>[2]</sup>在总结已有的各类分布函数的基础上，进行统一并提出了广义的贝塔分布函数（GB）和幂函数形式的广义贝塔分布函数（EGB），通过实证分析说明可以用于描述居民收入分布、金融有关变量的分布以及在回归分析方面的应用。成邦文<sup>[3]</sup>主要研究了在对数正态分布情况下洛伦兹曲线的性质、特征以及基尼系数的测算方法。黄恒君<sup>[4]</sup>通过收入分布函数序列为例，研究了位置—尺度分布族的近似构造。然而，由于分析的复杂性，到目前为止，很少见到根据指标的分布函数直接推导与分析洛伦兹曲线与基尼系数的研究与分析的论文。

### 曲线拟合法

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。

目前有很多拟合方法（例如多项式、样条函数逼近、洛伦兹曲线模型）来确定洛伦兹曲线。Kakwani<sup>[5]</sup>早在 1980 年就曾利用广义二次函数来拟合洛伦兹曲线；庄康、张永光<sup>[6]</sup>采用多项式函数形式拟合洛伦兹曲线，并按照 20%、60%、20%收入群体比率，将基尼系数分解到各收入群体中；陈孝新<sup>[7]</sup>推导出洛伦兹曲线的三次分段多项式样条函数拟合的表达式，并通过样条回归确定系数。

实践证明使用洛伦兹曲线模型是比较理想的方法之一。

关于洛伦兹曲线模型，国外经济理论学界已经做了大量工作，发现了一系列洛伦兹曲线模型，用于拟合产生近似的洛伦兹曲线。这种模型都是含有一个或几个参数的函数关系式或经验公式，结合相关文献，我们在表 1 里归纳总结出 15 个模型。但有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，这不会影响基尼系数的估算精度，但在其他经济应用（如贫困评估）时往往会产生问题（如 Kakwani<sup>[5]</sup>1980 年提出的广义二次函数拟合洛伦兹曲线模型）。另外，Cheong<sup>[8]</sup>指出，有些洛伦兹模型还有一个缺点，即其估计值超出了 Gastwirth<sup>[9]</sup>用非参数方法给出的基于分组数据计算所得基尼指数的上下限，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，在模型的改进和优化方面，Ogwang 和 Rao<sup>[10]</sup> 以及我国的王祖祥教授<sup>[11]</sup>做了大量工作。

表 1 十五种洛伦兹曲线

模型	洛伦兹曲线	参数取值范围
Kakwani, Podder	$L(p) = p^\alpha \cdot e^{-\beta(1-p)}$	$1 < \alpha < 2, \beta > 0$
Rasche	$L(p) = [1 - (1-p)^\alpha]^\beta$	$0 < \alpha \leq 1, \beta \geq 1$
Chotikapanich	$L_\lambda(p) = \frac{e^{\lambda p} - 1}{e^\lambda - 1}$	$\lambda > 0$
Kakwani	$L(p) = p - ap^\alpha (1-p)^\beta$	$a > 0, 0 < \alpha \leq 1, 0 < \beta \leq 1$
Wang, Smyth	$L(p) = 1 - \left( \frac{e^{\lambda(1-p)} - 1}{e^\lambda - 1} \right)^\beta$	$\lambda \in (-\infty, 0) \cup (0, \ln \beta^{-1}]$ $0 < \beta \leq 1$
Gupta	$L(p) = p \cdot \alpha^{p-1}$	$\alpha > 0$
Sarabia	$L(p) = 1 - (1-p)^\beta$	$0 < \beta \leq 1$
Ortega	$L(p) = p^\alpha [1 - (1-p)^\beta]$	$\alpha \geq 0, 0 < \beta \leq 1$
Ogwang, Rao	$L(p) = \delta p^\alpha [1 - (1-p)^\beta] + (1-\delta) \frac{e^{\lambda p} - 1}{e^\lambda - 1}$	$\lambda > 0, 0 < \delta \leq 1$ $\alpha \geq 0, 0 < \beta \leq 1$
Lognormal	$L(p) = \Phi(\Phi^{-1}(p) - \sigma)$	$\sigma > 0$
Pakes	$L(p) = I_p(\delta, \eta)$	$\delta \geq 1, 0 < \eta \leq 1$
Basman	$L(p) = p^{ap+b} e^{-g(1-p^2) - h(1-p)}$	$a, b, g, h^{(1)}$
Schader, Schmid	$L(p) = p^\gamma - ap^\alpha (1-p)^\beta$	$a, \gamma, \alpha, \beta^{(1)}$
Sarabia	$L(p) = p^\alpha \left( \frac{e^{\lambda p} - 1}{e^\lambda - 1} \right)^\nu$	$\lambda > 0, \alpha > 0, \nu \geq 1$
Wang	$L(p) = \left[ \delta (1 - (1-p)^\beta) + (1-\delta) \frac{e^{\lambda p} - 1}{e^\lambda - 1} \right]^\eta$	$\lambda > 0, 0 \leq \delta \leq 1$ $\eta \geq 1, 0 < \beta \leq 1$

(1): 模型中参数的取值范围未确定

## 2.2 中等收入概念及界定方法研究

中等收入是一个抽象而又具体的概念。说它抽象，是因为每个人都可以对其给出自己的理解；说它具体，是因为一旦给出自己的理解，每个人都可以据此找到自己的位置。

关于中等收入阶层和人口的评定标准，国内外已开展了大量的研究。目前，大多数国家对中等收入阶层的界定尚停留在定性的阶段。一些发达国家如美国等，已经尝试采用定量的方法，包括选用一些数量指标，但总体上还处于摸索阶段，对于中等收入的定量界定虽然还没有一个十分成熟的方法。

题目中提到了“收入空间法”、“人口空间法”，但它们具有任意性和与经济直观不符合的缺点，除了这两种方法之外，通过阅读相关文献，我们还总结出另外四种界定中等收入阶层和人口的方法。

### 五分法或中收系数法

五分法即把全部家庭户或人口分为最低收入、次低收入、中等收入、较高收入和最高收入五个组，每组各占家庭户或人口总数的 20%。可以把中间三组的收入份额之和与整体收入份额之比称为中收系数，以此从一个侧面反映中等收入者比重现状，中收系数愈高，说明收入愈向中间集中，居民的收入差异较小，中等收入群体的比重愈大。这里需要说明的是，中收系数不等于中等收入者的实际比重，因为中等收入者本身就是一个相对的概念，目前还没有一个统一的标准，因此只有用按人口计算处于中间收入阶层的收入占总收入的比重来衡量。

五分法或中收系数法的一个最大好处是资料容易获得，可以很方便地揭示一个地区中等收入人群的成长过程。但一个致命的缺点是很难据以推算中等收入的确切水平和比重。

### 固定比例法

为了界定的方便，固定比例法也被经常用于中等收入界限的确定，即考虑收入落在中位收入  $m$  的一个范围内的人口为中等收入人口。如根据美国 1999 年人口普查结果，取全美家庭年收入中值 40800 美元的 80% 和 120% 为下限和上限。这样计算的结果把中产阶级的家庭年收入范围定在 33000 美元到 49000 美元这个收入区间之间。目前，西方国家普遍进入富裕国家行列之后，普遍改为按固定比例界定中等收入。

固定比例法的优点是简单，但是科学性不够，且具有很强的任意性。以一个固定的比例来确定中等收入群体的范围，难以反映中等收入群体不断成长的历程。

## 恩格尔系数法

隐性收入的大量存在是造成收入指标失真的主要原因,在这种情况下,选择职业地位和收入水平作为社会分层的主要指标,就有很大的局限性。据此,可以选择更容易准确测定的消费指标作为社会分层的依据,因为瞒报和虚报消费的可能性远比瞒报和虚报收入的可能性小得多。我们将国际上通行的衡量消费水平的恩格尔系数作为消费分层的划分依据(见表2),即可以确定一个合适的恩格尔系数区间所对应的收入水平作为中等收入的水平。恩格尔系数被更广泛地应用于测量一个家庭的贫困或富裕程度,这种方法的优点是可以忽略社会调查中人们对实际收入的瞒报和虚报行为。

表2 恩格尔系数与生活水平

恩格尔系数	生活水平
0.3 以下	最富裕阶段
0.3—0.4	富裕阶段
0.4—0.5	小康阶段
0.5—0.6	勉强度日阶段
0.6 以上	绝对贫困阶段

资料来源:何明升、唐魁玉《中国消费之谜》(1989)

## 主观评价法

就社会分层理论及方法而言,历来都有客观分层与主观分层之说。主观分层,是指对社会成员地位归属,采用主观自我评价和他人认定的方式所获得的分层结果,其主要依据是被调查者主观评价(包括自己的和他人的评价、认定)。在对中等收入者的界定和认识上,主观分层的方法更具有重要意义。因为是通过广泛的征求意见基础上形成的结果。

主观评价法的优点是社会公认性强,受人质疑的可能性较小。但主观评价法的缺点是被调查者往往对中等收入者的内涵不了解,导致科学性降低。

综上,我们需要综合比较分析这些方法的优缺点,改进中等收入人口界定方法。

## 2.3 本文拟完成的工作

通过对问题的重述、题目要求的理解,我们阅读相关文献,进一步了解了相关问题的国内外研究现状并进行了比较分析,我们总结在本文我们需要完成的工作如下:

工作1:基于文献[11]的相关理论与方法,构造一个新的洛伦兹曲线模型,证明其满足洛伦兹曲线特征且符合经济规律。运用最小二乘法对其进行参数估

计，并将新模型与其他模型进行拟合精度比较。

工作 2：现有的大多数中等收入界定方法普遍存在确定范围随意、不能反映经济直观等缺陷，本文的工作立足于改进现有的中等收入界定方法，可以通过合理的分析与计算得到中等收入人口的范围，并反映出收入分配的特征。

工作 3：提出一种符合经济直观，易被社会群体主观接受的中等收入人口模型。

### 三、 模型假设

假设 1：收入分配曲线中仅考虑正的收入，且收入分配曲线是正偏的，即峰值点向左偏，右端拖一个长尾巴，意味着收入大于或等于平均收入的人口一定不到半数，因此是少数。

假设 2：收入分配的密度函数是连续的。

假设 3：入户统计调查个人相关数据时时，没有瞒报现象，也就是说可以获得较为准确的基础数据，从而确保模型运算结果的精度。

假设 4：忽略个人重大意外事故（例如重大疾病、重大交通事故等）导致其所属收入人口等级的改变。

假设 5：确定某地区群体中个人收入的最大值时要排除掉极端的少数人的过大收入。

### 四、 符号说明

符号	符号说明
$L(p)$	洛伦兹曲线
$G$	基尼系数
$\tau$	洛伦兹曲线模型中的参数变量
$x_l$	收入空间中中等收入人口的收入下限
$x_h$	收入空间中中等收入人口的收入上限
$m$	某地区收入的中位数
$f(m)$	收入分布的概率密度函数
$P$	两级分化严重程度指标
$W_{mid}$	中等收入人口的比重
$S$	中等收入人口状态
$Z$	个人收入水平中等化评价指数

$H$	个人生活质量评价指数
$K$	个人生活质量吻合度指数
$R$	个人中等收入吻合度指数

## 五、模型的建立与求解

### 5.1 洛伦兹曲线模型

正如我们在前面的问题分析中提到，目前国外经济理论文献中存在的模型大多是简单的函数关系式，含有的参数一般不超过三个，拟合效果不甚理想，有些甚至不满足洛伦兹曲线的条件。实际应用中，可以通过数据拟合的方法确定其中的参数来调整曲线的形状，使曲线尽量通过给定的各个数据点，即让拟合曲线与各数据点的误差尽可能小。显然，参数设置合理且参数个数适当的模型有可能得到更好的拟合效果。

为了方便后面的论述和证明，我们先给出我们在构造新的洛伦兹曲线模型时用到的王祖祥<sup>[11]</sup>总结的定理如下：

**定理** 假定  $f(p)$  和  $g(p)$  都是洛伦兹曲线，当满足  $\alpha \geq \frac{1}{2}$ ,  $\beta \geq \frac{1}{2}$ ，且对于任意  $p \in [0, 1]$ ,  $h''(p) \geq 0$  和  $g''(p) \geq 0$  时， $L(p) = h(p)^\alpha \cdot g(p)^\beta$  也是洛伦兹曲线。

通过对现有文献的大量阅读研究，首先，我们创新性的提出了一种新的洛伦兹曲线模型： $L(p) = 1 - (1 - p \cdot \mu^{p-1})^\alpha$ ，并证明当  $\mu \geq 1$ ,  $0 < \alpha \leq 1$  时，其满足洛伦兹曲线的条件。

然后，我们学习、运用了 Ogowang 和 Rao<sup>[10]</sup>提出的用于建立混合的洛伦兹模型的加权积法，并通过比较，引入王祖祥<sup>[11]</sup>提出的洛伦兹曲线模型作为我们新洛伦兹曲线模型的一个因子： $L(p) = 1 - L_\lambda(1-p)^\beta$ ，其参数的取值范围为：

$\lambda \in (-\infty, 0) \cup (0, \ln \beta^{-1}]$   $0 < \beta \leq 1$ 。根据前面提到的定理，我们进一步提出形式为  $L(p) = h(p)^\alpha \cdot g(p)^\beta$  的六参数洛伦兹曲线模型：

$$L(p) = \left[ 1 - (1 - p \cdot \mu^{p-1})^\alpha \right]^m \cdot \left[ 1 - L_\lambda(1-p)^\beta \right]^n,$$

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

参数的取值范围： $\mu \geq 1$ ,  $\lambda \in (-\infty, 0) \cup (0, \ln \beta^{-1}]$ ,  $0 < \alpha \leq 1$ ,  $0 < \beta \leq 1$ ,  
 $m, n \geq \frac{1}{2}$ 。

参数  $m, n$  的取值范围由前面提到的定理给出,  $\alpha, \beta, \lambda$  的取值范围参考了已有的经典洛伦兹模型的参数取值范围,  $\mu$  的取值范围是在证明其满足洛伦兹曲线特征时推导得出。

接下来我们证明我们的模型是满足洛伦兹曲线特征能反映经济规律。

**证明:**

记  $L_1(p) = 1 - (1 - p \cdot \mu^{p-1})^\alpha$ ,  $L_2(p) = 1 - L_\lambda(1-p)^\beta$ , 显然,  $L_1(0) = 0$ ,  $L_1(1) = 1$ ,  
 $L_2(0) = 0$ ,  $L_2(1) = 1$  成立。下面我们分别证明  $L_1'(p)$ 、 $L_1''(p)$ 、 $L_1'''(p)$  均大于零。

$$\begin{aligned} L_1'(p) &= -\alpha(1 - p \cdot \mu^{p-1})^{\alpha-1} \cdot (-\mu^{p-1} - p \cdot \mu^{p-1} \cdot \ln \mu) \\ &= \alpha(1 - p \cdot \mu^{p-1})^{\alpha-1} \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu) \geq 0 \end{aligned}$$

$$\begin{aligned} L_1''(p) &= \alpha \left\{ \left[ (1 - p \cdot \mu^{p-1})^{\alpha-1} \right]' \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu) + (1 - p \cdot \mu^{p-1})^{\alpha-1} \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu)' \right\} \\ &= \alpha \left[ (\alpha-1) \cdot (1 - p \cdot \mu^{p-1})^{\alpha-2} \cdot (-\mu^{p-1} - p \cdot \mu^{p-1} \cdot \ln \mu) \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu) \right. \\ &\quad \left. + (1 - p \cdot \mu^{p-1})^{\alpha-1} \cdot (\mu^{p-1} \cdot \ln \mu + \mu^{p-1} \cdot \ln \mu + p \cdot \mu^{p-1} \cdot \ln^2 \mu) \right] \end{aligned}$$

$$\text{记 } f(p) = 1 - p \cdot \mu^{p-1} \geq 0, \quad f'(p) = -\mu^{p-1} - p \cdot \mu^{p-1} \cdot \ln \mu \leq 0$$

$$\begin{aligned} L_1''(p) &= \alpha \left[ -(\alpha-1) \cdot f(p)^{\alpha-2} \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu)^2 \right. \\ &\quad \left. + f(p)^{\alpha-1} \cdot (2\mu^{p-1} \cdot \ln \mu + p \cdot \mu^{p-1} \cdot \ln^2 \mu) \right] \geq 0 \end{aligned}$$

$$\begin{aligned} L_1'''(p) &= \alpha \left[ -(\alpha-1) \cdot (\alpha-2) f(p)^{\alpha-3} \cdot f'(p) \cdot (\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu)^2 - (\alpha-1) \cdot f(p)^{\alpha-2} \cdot \right. \\ &\quad \left. 2(\mu^{p-1} + p \cdot \mu^{p-1} \cdot \ln \mu) \cdot (\mu^{p-1} \cdot \ln \mu + \mu^{p-1} \cdot \ln \mu + p \cdot \mu^{p-1} \cdot \ln^2 \mu) + (\alpha-1) \cdot \right. \\ &\quad \left. f(p)^{\alpha-2} \cdot f'(p) \cdot (2\mu^{p-1} \cdot \ln \mu + p \cdot \mu^{p-1} \cdot \ln^2 \mu) + f(p)^{\alpha-1} \cdot \right. \\ &\quad \left. (2\mu^{p-1} \cdot \ln^2 \mu + p \cdot \mu^{p-1} \cdot \ln^3 \mu) \right] \geq 0 \end{aligned}$$

对于  $L_2(p) = 1 - L_\lambda(1-p)^\beta$ , 我们同样要证明  $L_2'(p)$ 、 $L_2''(p)$ 、 $L_2'''(p)$  均大于零。

$$L_2'(p) = -\beta L_\lambda(1-p)^{\beta-2} L_\lambda'(1-p) \geq 0$$

$$L_2''(p) = -\beta L_\lambda(1-p)^{\beta-2} L_\lambda'(1-p) \left[ (\beta-1)L_\lambda'(1-p) - \lambda L_\lambda(1-p) \right]$$

$$= -\beta L_\lambda(1-p)^{\beta-2} L_\lambda'(1-p) \frac{\lambda}{e^\lambda - 1} \left[ 1 - \beta e^{\lambda(1-p)} \right] \geq 0$$

$$L_2'''(p) = \left[ -\beta \cdot L_\lambda'(1-p) \cdot L_\lambda(1-p)^{\beta-3} \right] \cdot$$

$$\left\{ \left[ \lambda L_\lambda(1-p) + (1-\beta)L_\lambda'(1-p) \right]^2 - (1-\beta)L_\lambda'(1-p) \left[ -L_\lambda'(1-p) - \lambda L_\lambda(1-p) \right] \right\}$$

我们注意到，对任何的  $\lambda \neq 0$ ，都可以得到  $\lambda/(e^\lambda - 1) > 0$ ，且当  $0 < \beta \leq 1$ ，

$\lambda \in (-\infty, 0) \cup (0, \ln \beta^{-1}]$  时，对于任何的  $p \in [0, 1]$  都有  $1 - \beta \cdot e^{\lambda(1-p)} \geq 0$ ，

因此

$$(\beta-1)L_\lambda'(1-p) - \lambda L_\lambda(1-p) = \frac{\lambda}{e^\lambda - 1} \left[ 1 - \beta e^{\lambda(1-p)} \right] \geq 0$$

由于  $-L_\lambda'(1-p) \geq (\beta-1)L_\lambda'(1-p)$ ，

故  $-L_\lambda'(1-p) - \lambda L_\lambda(1-p) \geq (\beta-1)L_\lambda'(1-p) - \lambda L_\lambda(1-p) \geq 0$

由此可知，对于任何  $p \in [0, 1]$ ，都有  $L_2'''(p) \geq 0$ 。

最后，根据我们在前面提到的定理，

$$L(p) = L_1(p)^m \cdot L_2(p)^n$$

$$= \left[ 1 - (1-p \cdot \mu^{p-1})^\alpha \right]^m \cdot \left[ 1 - L_\lambda(1-p)^\beta \right]^n$$

在参数取值范围为： $\mu \geq 1$ ， $\lambda \in (-\infty, 0) \cup (0, \ln \beta^{-1}]$ ， $0 < \alpha \leq 1$ ， $0 < \beta \leq 1$ ，

$m, n \geq \frac{1}{2}$  时，满足洛伦兹曲线特征，能够反映经济规律。

**证毕。**

实际操作中，从分组数据出发，通常利用非线性最小二乘法极小化残差平方和：

$$\Delta = \sum_{i=1}^n (L(p_i, \tau) - L_i)^2$$

得到参数估计  $\hat{\tau} = (\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{m}, \hat{n})$ ，从而得到近似的洛伦兹曲线  $L(p, \tau)$ 。

应用某些参数变换，可以将约束非线性最小二乘问题化为无约束的，使得可以应用无约束最小二乘法进行参数  $\tau$  的估计。对于我们提出的模型，依次令：

$$\mu = 1 + x_1^2, \quad \alpha = \sin^2 x_2, \quad \beta = \sin^2 x_3, \quad \lambda = \ln \frac{1}{\sin^2 x_3} - x_4^2, \quad m = \frac{1}{2} + x_5^2, \quad n = \frac{1}{2} + x_6^2$$

将这些关系式代入模型，用无约束非线性最小二乘法求  $(x_1, x_2, x_3, x_4, x_5, x_6)$  的估计值，得到  $(\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5, \hat{x}_6)$  后再代回到这些关系式中，从而得到模型的参数估计：

$$\begin{aligned} \hat{\mu} &= 1.8839, \quad \hat{\alpha} = 0.9450, \quad \hat{\lambda} = 0.1961, \\ \hat{\beta} &= 0.6979, \quad \hat{m} = 0.7299, \quad \hat{n} = 0.8532 \end{aligned}$$

$$\hat{\tau} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{m}, \hat{n}) = (1.8839, 0.9450, 0.1961, 0.6979, 0.7299, 0.8532)$$

进一步，根据题目中给出的拟合精度好坏的三种标准（均方误差 MSE、平均绝对误差 MAE、最大绝对误差 MAXABS），我们从表 1 中找出 10 种经典的洛伦兹曲线模型，与我们提出的新模型进行拟合精度比较，其结果如表 3 所示：

表 3 洛伦兹曲线模型拟合精度比较

模型	$MSE \times 10^5$	MAE	MAXABS
Kakwani, Podder	26.1715	0.012307361	0.040884892
Rasche	1.82	0.003770065	0.007969035
Sarabia	305.1689	0.049608248	0.077400319
Chotikapanich	27.3831	0.014086406	0.035882638
Ogwang, Rao	0.847	0.002660626	0.005344326
Kakwani	0.397	0.001681175	0.004074109
Gupta	26.6009	0.01309733	0.039512905
Ortega	0.938	0.002741163	0.005523367
Wang	1.09	0.002923999	0.006295658
Wang, Smyth	25.1554	0.014405156	0.02606061
<b>我们的模型</b>	<b>0.354</b>	<b>0.00140883</b>	<b>0.005581551</b>

通过比较我们的模型和表中十个模型的拟合精度，我们观察到，以均方误差为评价标准，我们的模型排第一；以平均绝对误差为评价标准；我们的模型排第一；以最大绝对误差为标准，我们的模型排第三。

为了更直观、综合地反映模型的拟合精度好坏，我们在这里对其作一个简单的打分。记各项标准中，排第一位的模型得分为 10 分，第二位的得 9 分，依次递进，排最后一位的模型得 0 分。再把每个模型在 3 项评价标准中的得分相加，其总分即为该模型的拟合精度得分，得分越高，拟合精度越好。其结果如

表 4 所示:

表 4 模型拟合精度综合评定排名

模 型	得 分	拟合精度排名
Kakwani,Podder	8	8
Rasche	15	6
Sarabia	0	11
Chotikapanich	6	10
Ogwan, Rao	25	3
Kakwani	28	1
Gupta	7	9
Ortega	22	4
Wang	18	5
Wang, Smyth	9	7
<b>我们的模型</b>	<b>27</b>	<b>2</b>

可以发现, 在我们选出的 11 种模型中, 拟合精度最高的洛伦兹模型是 Kakwani<sup>[5]</sup>在 1986 年提出的三参数洛伦兹模型, 我们的模型排在第二位, 其拟合精度排在前列, 并不比之前的经典的洛伦兹模型差。

## 5.2 中等收入人口界定方法

在研究中等收入者的相关问题时, 首先要解决的问题就是界定中等收入者的范围。现今, 在经济理论学界, 通常使用两种方法——收入空间法与人口空间法来确定该范围。但简单地使用这两种方法会产生许多问题, 甚至会导致在分析数据时出现截然相反的结论, 本节我们的工作就立足于改进这两种方法, 提出一些合理的界定中等收入人群的方法。

### 5.2.1 改进收入空间法

收入空间法通过给出收入区间来确定中等收入人群。传统的收入空间法中, 收入上下界  $x_l$  与  $x_h$  的取法具有任意性, 不能说明经济发展与收入分布的实际情况, 为了克服这个不足, 我们从以下两个角度提出相关的改进方法。

#### 1. 基于比例的范围界定

采用传统的收入空间法时, 存在的一个重要不足是, 在固定上下界  $x_l$  与  $x_h$  的值后, 中等收入人口不能反映经济进步的影响。通常情况下, 所有人口的收入

会逐年变化，而使得全社会的收入区间发生移动，因此，直观来看中等收入区间也应随之移动。根据这一分析，我们通过下式来确定  $x_l$  与  $x_h$ ：

$$x_l = (1-a) \cdot m \quad x_h = (1+a) \cdot m,$$

其中， $a$  为一个给定的百分比值。由此，当全社会收入中位值发生变化时，中等收入的上下界也能随之变化，使中等收入区间始终位于社会收入中值附近的一个合理范围内，使纵向比较各年的中等收入人口时更为合理。

## 2. 基于统计指标的范围界定

基于百分比的范围界定方法解决了传统方法中  $x_l$  与  $x_h$  的取值不能反映经济发展、收入变化情况的不足，但采用这种方法时，没有考虑当年社会人口收入的总体分布情况，为此我们将收入分布的一些统计指标纳入计算之中。

### (1) 基于极差的计算方法

当整个社会中高收入与低收入群体的差距变大时，中等收入的范围也应有所扩大。据此，我们给出下式：

$$x_l = m - d \quad x_h = m + d, \quad d = (x_{\max} - x_{\min}) / n,$$

其中， $x_{\max}$  与  $x_{\min}$  分别为社会收入总体中的最大值与最小值， $n$  为整数参数。

### (2) 基于标准差的计算方法

当整个社会中收入的波动情况不同时，中等收入的范围也应有所变化。波动越大，直观上看中等收入的范围应该越大。据此，我们可以通过下式计算  $x_l$  与  $x_h$ ：

$$x_l = m - \Delta \quad x_h = m + \Delta,$$

其中， $\Delta$  为收入总体的标准差。该式的含义是，中等收入人口的收入应该在总体中值附近波动，而这个波动可以由总体的标准差来确定。

采用基于统计指标的界定方法反映了总体分布情况对中等收入人口范围的影响，但它也存在一些不足，例如当社会收入两极分化极其严重时（即  $d$  的取值非常大），采用基于极差的计算方法会使中等收入的范围变得非常大，这显然与现实情况不符合。为此，我们结合基于比例与统计指标的方法，提出一种综合的计算方法：

$$\begin{cases} x_l = \max \{ a' m, m - d \} \\ x_h = \max \{ a'' x_l, m + \Delta \} \end{cases}$$

其中， $a'$  与  $a''$  为两个比例参数，且  $a' < 1$ ， $a'' > 1$ ；其余参数的含义与上述相同。该计算方法的含义在于：当总体收入差距变大时，中等收入的下界可以适当放

宽，但不能低于中值的一定比例；中等收入的上界  $x_h$  是下界的  $x_l$  一定倍数，使中等收入区间内的人口收入差距不至于过大，但同时也要保证  $x_h$  大于总体的中值与标准差之和， $m + \Delta$  可以视作中值合理波动的上界。

### 5.2.2 改进人口空间法

另一种确定中等收入人口范围的方法是人口空间法，即选择中值  $m$  邻近的一个固定人口比例范围为中等收入人口，再用这个人口比例范围来确定中等收入人口的收入范围。这个方法简单易行，可以清晰描述在社会总人口中位于中间部分的人群的收入状况，特别适合经济发展成熟，社会结构稳定的发达国家，例如美国、日本等都采用该方法来确定中等收入范围。但该方法存在以下三个缺陷：

- 1、人口比重的确定比较随意，没有反映社会总体的收入情况；
- 2、中等收入人口占总人口的比重始终是固定的，无法反映经济发展、收入结构变化的情况，尤其不适当当下发展中国家不断发展、变革的实际；
- 3、如题目中所言，无法反映两极分化对中等收入人口所占比重的影响。

中等收入人口的多少与两极分化的程度有关，所谓两极分化，用密度函数表示就是严重右偏且厚尾，也即中间部分空洞化。因此直观上看，两极分化越严重，中等收入人口应越小，反之则意味着中等收入人口扩大。

为了克服人口空间法的缺陷，体现两极分化对中等收入人口范围的影响，一些学者研究使用基尼系数来确定中等收入人口的比重，但这种基于基尼系数的改进人口空间法存在一个问题：基尼系数描述了社会收入总体上的不平等程度，但并不能描述社会人口收入两极分化的情况。两极分化与收入不平等是两个不同的概念，收入不平等并不一定意味着两极分化。例如，当高收入人群与低收入人群内部的收入不平等十分严重时，总体的收入不平等程度也会加重，即基尼系数变大，但此时，这种内部的不平等使得位于收入分配两端的人群向中间部分靠拢，也就是说两极分化程度反而减轻了，中等收入人口数量有所提高。因此，我们需要使用准确的两极分化指标，重新测算两极分化与中等收入人口比重之间的关系。

文献[12]对两极分化的概念与现象进行了准确阐述，并给出了一种说明两极分化严重程度的指标  $P$ ，该指标值越大说明两极分化的程度越严重，其计算方式如下：

$$P = (T - G) \frac{\mu}{m},$$

其中， $T = (\mu^U - \mu^L) / \mu$ ， $\mu^U$  与  $\mu^L$  分别表示收入大于、小于中值  $m$  的人口的收

入均值， $\mu$  是整体的均值，由于  $(\mu^U - \mu)/\mu = (\mu - \mu^L)/\mu$ ，因此  $T$  反映了将每个在中值以下的人口收入提升到总体平均值所需要的收入，即相对中值的偏移程度； $G$  是基尼系数； $\mu/m$  表示了概率密度函数的右偏程度。

由文献[12]中的推导可得  $T = 1 - 2L(0.5)$ 。另，根据公式  $L'(p) = x/\mu$ ，可求得  $m = \mu L'(0.5)$ ，而  $G = 1 - 2\int_0^1 L(p) dp$ 。所以，根据人口收入的洛伦兹曲线，可求得  $P$  值如下：

$$\begin{aligned} P &= (T - G) \frac{\mu}{m} \\ &= \left(1 - 2L(0.5) - 1 + 2\int_0^1 L(p) dp\right) \frac{\mu}{\mu L'(0.5)} \\ &= \frac{2\left(\int_0^1 L(p) dp - L(0.5)\right)}{L'(0.5)}. \end{aligned}$$

为了获得两极分化指标与中等收入人口比重之间的定量关系，我们分析了世界银行公布的 42 个国家的五分法收入分配情况。不失一般性地，我们以中国的数据为例，说明具体分析计算步骤如下。

表 5 中国五分法收入分配表

	最低 20%	第二个 20%	第三个 20%	第四个 20%	最高 20%
收入比重(%)	5.9	10.2	15.1	22.2	46.6
收入累计比重(%)	5.9	16.1	31.2	53.4	100
人口累计比重(%)	20	40	60	80	100

### 步骤 1. 拟合收入分配的洛伦兹曲线

根据五分法收入分配表，可以得到人口累计比重与相应收入累计比重的五个数据点。由于数据点数量较少，不适宜采用参数较多的洛伦兹曲线模型进行拟合，因此，我们使用 Gupta 模型对这五个点进行拟合，可得到洛伦兹曲线模型如下：

$$L(p) = p \cdot 5.6762^{p-1}.$$

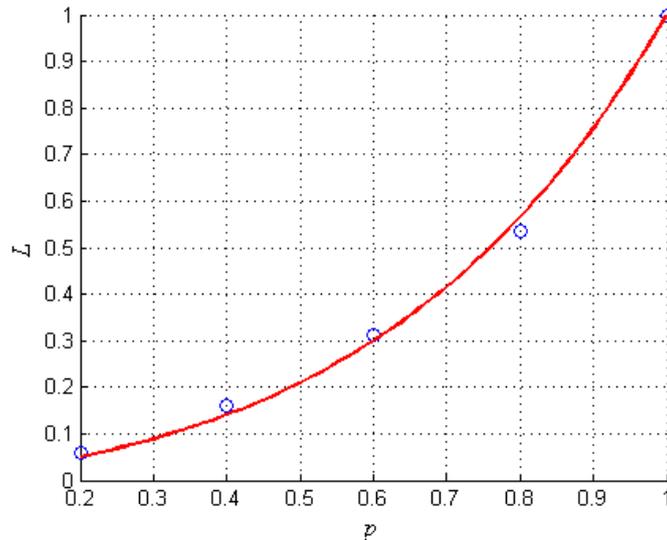


图2 中国收入分配的洛伦兹曲线

**步骤 2. 计算中等收入人口比重**

参照文献[13-14]的定义，以在中值附近，人群总收入占社会总收入 60% 的人群作为中等收入人口，从而可由下式求得中等收入人口的比重：

$$W_{mid} = L^{-1}(0.8) - L^{-1}(0.2).$$

对于中国而言， $W_{mid} = L^{-1}(0.8) - L^{-1}(0.2) = 0.92 - 0.49 = 0.43$ ，即中国的中等收入人口比重为 43%。

**步骤 3. 计算两极分化指标值**

依据中国收入分配的洛伦兹曲线，按照两极分化指标的计算公式，可得中国的两极分化指标值  $P = 0.2367$ 。

按照上述过程，依次计算其余 41 个国家的中等收入人口比重与两极分化指标，图 3 显示了这 42 个国家两极分化指标与中等收入人口比重之间的关系。

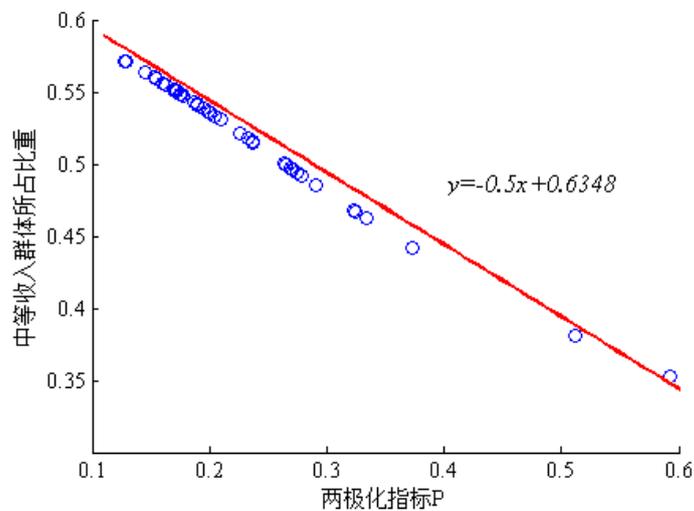


图3 两极分化指标与中等收入人口比重的关系

从图 3 中可以看出，两极分化指标与中等收入人口比重之间具有明显的线性关系，且两极分化指标越大，中等收入人口比重越小，符合经济直观要求。对上述散点做线性回归，可得到结果如下：

$$W_{mid} = -0.5P + 0.6348,$$

调整后  $R^2=0.9948$ ， $SSE=4.73 \times 10^{-4}$ 。由该结果可知两极分化指标与中等收入之间具有高度线性相关的关系。同时，由于计算过程采用的样本是各个处于不同发展阶段的 42 个国家，具有较好的代表性，所以上述计算分析结果较为可信，当得到两极分化指标后，可以通过上式计算得到中等收入人口比重。

这种基于两极分化指标的改进人口空间法，依据收入分配的洛伦兹曲线，通过合理的计算与分析，由两极分化指标得到中等收入人口比重。该方法克服了传统人口空间法中确定比重较为随意、无法反映收入分配结构与中等收入群体成长过程等缺陷，使分析结果更符合经济客观，更具有纵向与横向可比性。

### 5.3 A B 两地区中等收入情况分析

首先，我们对 A、B 两个地区前后两年的收入分配数据进行常用统计指标上的分析。图 4 显示了该四组数据的累积分布函数。

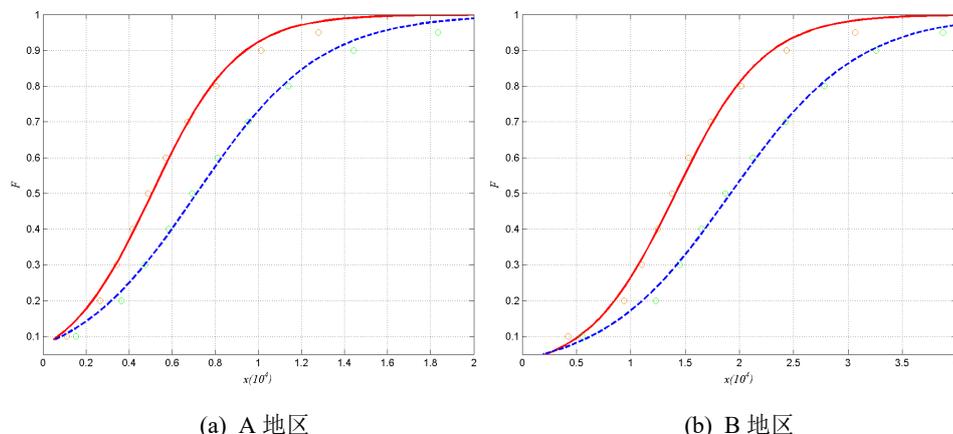


图 4 A、B 两地区各年份的累积分布函数

从图中可以看出 A、B 两个地区，后一年的累积分布函数曲线较前一年都出现明显的右移，这表明，随着社会经济的发展，A、B 两个地区所有人口的收入都提高了，A、B 两个地区中等收入区间也应随之右移。同时，仔细观察还可以看到，A 地区在高收入段的右移程度相较其他阶段更为显著，这说明 A 地区高收入群体的增加收入更多，社会财富进一步向高收入群体集中。表 6 说明了这四组数据在常用指标值上的取值情况。需要说明的是，为了便于计算，我们取各个收入分组的中值代表该分组的收入情况， $f(m)$  表示概率密度函数在总体中位数处的取值。

表 6 A、B 地区各年份统计指标值

地区	年份	均值 $\mu$	中位数 $m$	方差 $\Delta^2 (\times 10^7)$	$f(m) (\times 10^{-4})$
A	1	6281.34	4887.00	0.94	1.18
	2	8890.21	6872.78	1.95	0.82
B	1	16938.46	13797.87	4.45	0.55
	2	22228.53	18707.20	7.61	0.40

从上表可以看出，A、B 两个地区在前后两个年份的均值与中位数都有较大的增长，表明这两个地区的经济都有较快的发展。A 地区年份 2 的均值较年份 1 增长了 41.53%，而 B 地区均值的增长率为 31.23%，这说明 A、B 两地区的发展都比较大，但 A 地区的发展更为明显。比较中位数前后两年的变化可以发现，A 地区的中位数增长率为 40.63%，B 地区的为 35.58%，值得注意的是，A 地区中位数的增长率小于均值的增长率，而 B 地区恰恰相反。这个结果说明，A 地区的收入增长较快，但其中等收入人群的收入增长没能赶上平均增长率，中等收入人群收入所占社会总收入的比重在减小，结合图 4 中的分析，可以认为 A 地区的财富在向高收入群体集中。而 B 地区的收入增长主要发生在中等收入群体中，中等收入群体在社会总收入中的比重在增加。 $f(m)$  的大小体现了概率密度函数在中位数附近隆起程度的高低，即收入在中位数附近概率的大小，直观来看  $f(m)$  越大，则说明收入出现在中位数附近的人口越多，也就是中等收入人群越多。表中数据表明，A 地区后一年的  $f(m)$  值出现明显下降，减少了一个数量级，这说明 A 地区的中等收入人口规模在减小；而 B 地区的  $f(m)$  值变化不大，基本保持稳定，说明 B 地区的中等收入人口规模没有发生大的变化。

综合上述分析，我们可以得出一些定性的结论：

1. A、B 两个地区的人口收入都有明显的增长，且 A 地区的增长速度更快；
2. A 地区中，中等收入人群的收入占社会总收入的比重在减小，中等收入人群的规模也在减小，社会财富出现向高收入人群集中的现象；
3. B 地区中，中等收入人群的收入占社会总收入的比重在增加，且中等收入人群的规模变化不大，社会财富进一步向中间人群集中。

通过上述常用统计指标的分析，我们获得了一些定性结论，对 A、B 两个地区的发展情况有了概略的了解，但这些结论缺乏定量数据的支持，无法给出定量结论。

下面，我们利用前文所提出的模型与方法对 A、B 两个地区的收入分配情况、中等收入人群状态进行定量分析，以期获得相应的定量结果。

首先，我们利用章节 5.1 所提出的洛伦兹曲线模型对 A、B 地区前后两个年份的收入分配情况进行描述。对四组数据的拟合结果如下：

$$L(p) = \left[ 1 - (1 - p \cdot \mu^{p-1})^\alpha \right]^m \cdot \left[ 1 - L_\lambda(1-p)^\beta \right]^n, \quad \text{其中 } L_\lambda(1-p) = \frac{e^{\lambda(1-p)} - 1}{e^\lambda - 1}.$$

$$\text{A 地区年份 1: } \hat{\tau}_{A1} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{m}, \hat{n}) = (2.90, 0.72, -1.32, 0.97, 0.5, 0.53);$$

$$\text{A 地区年份 2: } \hat{\tau}_{A2} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{m}, \hat{n}) = (2.96, 0.74, -1.49, 0.95, 0.5, 0.51);$$

$$\text{B 地区年份 1: } \hat{\tau}_{B1} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{m}, \hat{n}) = (2.09, 0.65, -0.65, 0.99, 0.5, 0.5);$$

$$\text{B 地区年份 2: } \hat{\tau}_{B2} = (\hat{\mu}, \hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{m}, \hat{n}) = (2.47, 0.79, -0.85, 0.98, 0.5, 0.5).$$

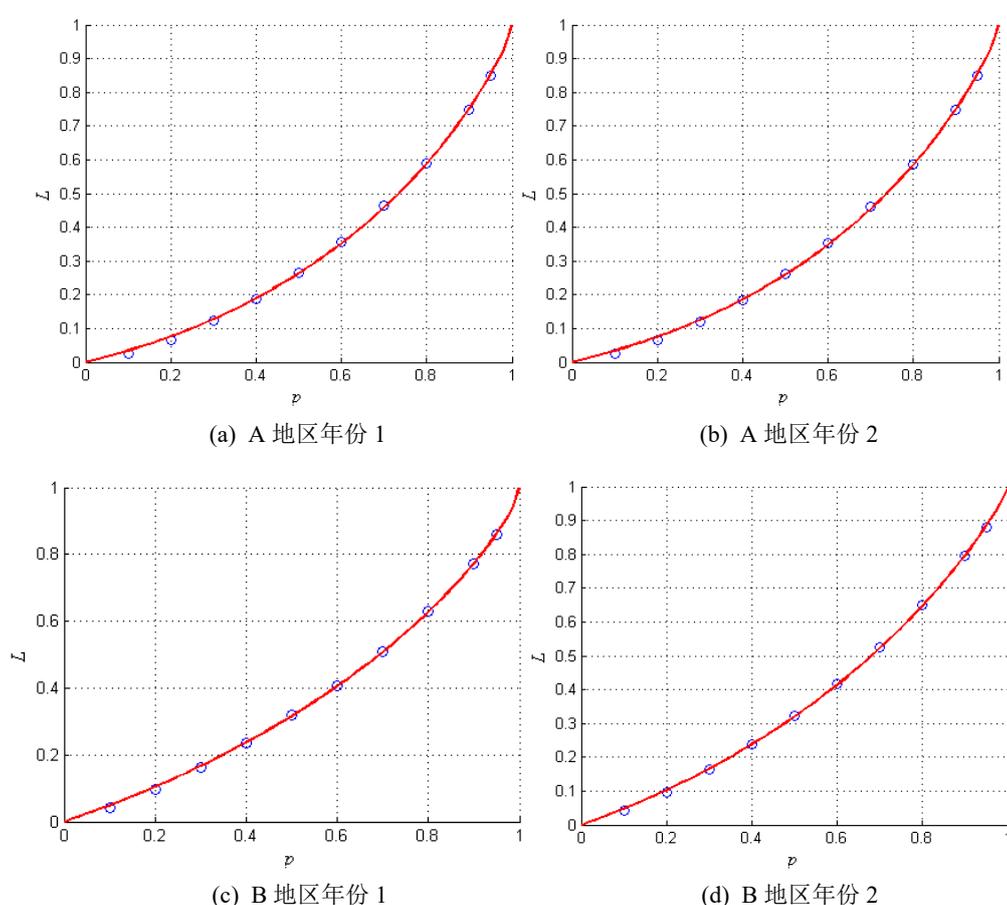


图 5 A、B 地区各年份的洛伦兹曲线

根据各组数据的洛伦兹曲线，按照章节 5.2 中提出的基于两极分化指标的改进人口空间法，我们可以计算得到各组数据的两极分化指标、中等收入人口比重。进而，我们可以获得中等收入范围的上下界与中等收入人口的状态，这两个值得计算过程较为简单，且已在题目中阐述，这里不再做介绍。我们直接将相关结果的数据列在下表中。

表 7 A、B 地区各年份收入分配情况

地区	年份	基尼系数 $G$	两极分化指标 $P$	中等收入人口比重 $W_{mid}(\%)$	收入下界 $x_l$	收入上界 $x_h$	中等收入人口状态 $S(\%)$
A	1	0.338	0.192	53.88	2847	9433	47.01
	2	0.344	0.196	53.69	3967	13383	46.88
B	1	0.268	0.145	56.25	9342	24348	49.24
	2	0.254	0.135	56.71	12739	31556	52.44

从上表的结果数据可以发现，A 地区的基尼系数、两极分化指标均变大，表明 A 地区收入分配的不平等程度与两极分化程度都在加剧，这直接导致中等收入人口比重减少了 0.19%，中等收入人口状态，即中等收入人口的总收入占社会总收入的比例减少了 0.13%。这个结果定量地说明，A 地区中等收入人口规模在减少，且占有的社会总收入量也在减少，结合上文的定性分析，我们可以认为该部分减少的量在向高收入群体转移。观察 B 地区的数据可以发现，B 地区的基尼系数与两极分化指标都在减小，说明 B 地区的收入分配变得更为平等，两极分化程度在减弱，这在中等收入人口情况上的表现就为中等收入人口比重增加了 0.46%，占有的收入总量增加了 3.20%，该结果定量地表明 B 地区中等收入人口规模在扩大，同时占有的社会总收入也在增加，社会财富在向中间阶层集中。

综合上述定性与定量分析，我们可以得到如下结论：

1. A 地区的中等收入人口比重与收入状态均在减小，表明 A 地区的中等收入人口规模与占有的财富总量均在缩水；
2. B 地区的中等收入人口比重与收入状态均在增加，说明 B 地区的中等收入人口规模与占有的财富总量均在扩大；
3. A 地区的收入分配结构有呈到三角的趋势，社会财富在向高收入人群集中；B 地区的收入分配结构已形成稳定的橄榄型（中等收入人口的规模与收入占有量均超过 50%）；
4. 可以认为 A 地区处于经济发展阶段，而 B 地区经济比较发达，已进入稳定成熟的阶段。

## 5.4 一种新的中等收入模型

对中等收入人口的界定及研究，是理论经济学的重要研究内容和研究方向，也是应用经济学新的研究领域和重要的研究方向。

对中等收入人口的界定，依赖于社会经济发展的现状，依赖于社会居民的结构，还依赖于人们的认知。如何对中等收入人口进行科学的界定，以及如何对中等收入人口进行科学合理的测算一直是国内外学者普遍关注的一个问题。

中等收入是一个相对而不是绝对的标准，它具有如下三个特征：

1. 中等收入是相对于一个国家或一个地区的高收入和低收入而言的，并没有一个固定的标准；
2. 中等收入是一个动态标准而不是静态标准；
3. 中等收入是区间数量而不是固定数量，即中等收入是一个大体的区间，不是一个固定不变的量。

通过阅读文献，我们发现，目前中等收入人口界定所面临的一个很大的问题是“伪中等收入问题”。据《2010年北京社会建设分析报告》公布的数据，北京中等收入者人数为540万，占北京市户籍人口的40%以上，然而，被认定人群中，68.7%的人不认可自己的“中等收入”标签，社会一片哗然。究其原因，工薪阶层作为中等收入人群的主体，面临子女教育、购房、赡养老人等长期较大支出，如果在支付这些固定项目后用于日常消费的支出寥寥无几，就不能算是中等收入，这就是“伪中等收入问题”。

因此，我们在本文中提出一个新的考虑到个人生活质量的中等收入人口界定模型，着力解决中等收入问题中的“伪中等收入问题”。

**定义：**中等收入人口是指收入水平处于某一地区全体居民某一时期内中位收入水平上下一定范围内并且生活质量较好的人口数量。

**原理：**(1) 该定义是以收入水平、生活质量作为确定中等收入者的条件，具有经济规定性。

- (2) 我们定义的中等收入水平不是一个绝对数，而是一个收入水平区间，且这一区间还是以一定地区、一定时期为特定条件的，具有社会规定性。

基于以上中等收入人口定义，我们进一步提出与之对应的中等收入人口测算模型。

在模型中，我们提出一个反映个人与中等收入人口特征的吻合程度的**个人中等收入吻合度指数  $R$** 。我们认为，该指数  $R$  受到个人收入水平和生活质量两方面的影响，这里我们定义**个人收入水平中等化评价指数  $Z$**  以及**个人生活质量评价指数  $H$**  来定量刻画其对  $R$  的影响。其中，个人收入水平中等化评价指数  $Z$  反映了某居民个人收入与中等收入人口收入水平的接近程度；个人生活质量评

价指数  $H$  反映了某居民现有生活质量的好坏程度。

我们现将这两个指数的生成原理及经济学意义分别介绍如下：

### 1. 个人收入水平中等化评价指数 $Z$

记第  $i$  个人的个人收入为  $x_i$ ，该地区个人收入的最大值和最小值分别为  $x_{\max}$ 、 $x_{\min}$ ， $m$  为收入分配密度函数曲线的中位数点对应的个人收入，则我们定义第  $i$  个人的个人收入水平中等化评价指数  $Z_i$  为：

$$Z_i = 1 - \frac{|x_i - m|}{\max(m - x_{\min}, x_{\max} - m)}$$

显然  $0 \leq Z_i \leq 1$ ， $Z_i$  越大，说明居民个人收入  $x_i$  距离中位收入点  $m$  的距离越近，表示该居民在收入上越接近于中等收入。该指数同时也与经济直观相符，因为随着群体的两极分化扩大， $Z_i$  的值是逐渐减小的，通过对群体中每个人的  $Z_i$  值进行统计就会发现， $Z_i$  取值较小的人口占总人口的比重就会增大，即群体中偏离中等收入的人口增大，中等收入人口下降。

### 2. 个人生活质量评价指数 $H$

我们定义，可自由支配收入是除去子女教育、购房、赡养老人等必须支付的固定收入支出后居民用于提高其生活质量的收入，我们这里认为其主要由娱乐支出和个人储蓄两部分组成。显然，可自由支配收入占总收入的比例越高，表示居民的精神文明生活越丰富，手头闲钱越多，自然生活质量也越高。

记第  $i$  个人的收入为  $x_i$ ，娱乐休闲支出为  $a_{1i}$ ，个人储存部分为  $a_{2i}$ ，则第  $i$  个人的个人生活质量评价指数  $H_i$  为：

$$H_i = \frac{a_{1i} + a_{2i}}{x_i}$$

显然， $0 \leq H_i \leq 1$ ， $H_i$  越大表示该居民的生活状态越好。一般情况下，高收入者对应的  $H$  较大，低收入者对应的  $H$  较小。随着收入的增加， $H$  也在增加，只不过在收入处于不同阶段时， $H$  增加的快慢程度不同，但总体上来说  $H$  是单调增的。随着  $H$  在超过一定范围后继续增大时，该居民就会向高收入者行列迈进。由此可见， $H$  只有在一定的范围内才能有效界定中等收入人口。我们定义

函数  $K = f(H)$ ,  $0 \leq K \leq 1$ , 定量刻画出某居民在生活质量方面与中等收入人口的接近程度, 记群体的个人生活质量评价指标的中位数为  $H_m$ , 则应满足  $f(H_m) = 1$ ,  $f(0) = 0$ ,  $f(1) = 0$ 。这三个约束显然满足经济直观要求, 当  $H$  的取值为中值  $H_m$  时, 显然与中等收入人口的生活质量最为接近, 即  $f(H_m) = 1$ ; 当  $H$  为最大或最小值时, 应当与中等收入人口的生活质量相差最大, 即  $f(0) = 0$ ,  $f(1) = 0$ 。

同时,  $f(H)$  还应满足约束: 当  $0 < H < H_m$  时,  $f(H)' > 0$  且  $f(H)'' > 0$ , 也就是说此时随着  $H$  的增大,  $K$  也在增大且增大的越来越快; 当  $H_m < H < 1$  时,  $f(H)' < 0$  且  $f(H)'' > 0$ , 也就是说此时随着  $H$  的增大,  $K$  在减小且减小的越来越慢。这显然也是符合经济直观要求的, 当  $H$  的取值远离  $H_m$  时, 人群对生活质量的感受就不敏感, 相应的  $K$  值变化也会变小。例如, 当  $H_m = 0.3$  时,  $H = 0.98$  与  $H = 0.99$  人群对生活质量的差别感应该不大, 则  $K$  值的变化也应较小。图 6 显示了  $f(H)$  曲线的一种合理形式。

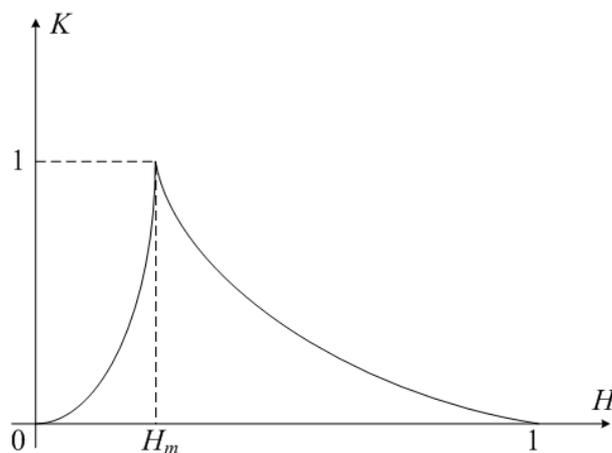


图 6  $H$ - $K$  曲线示意图

### 3. 个人中等收入吻合度指数 $R$

综合上面提出的个人收入水平中等化评价指标  $Z$  和个人生活质量吻合度指数  $K$ , 我们可定义某地区第  $i$  个人与中等收入人口特征的吻合程度指数的表达式为:

$$R_i = Z_i^\alpha \cdot K_i^\beta, \text{ 其中 } \alpha > 0, \beta > 0, \alpha + \beta = 1,$$

易知， $0 \leq R_i \leq 1$ ，且  $R_i$  越大，第  $i$  个人与中等收入人口特征的吻合程度就越好。取一个合理的  $\theta$  值 ( $0 < \theta < 1$ )，如果  $\theta \leq R_i \leq 1$ ，就可判定第  $i$  个人属于中等收入人口。

现记群体的总人数为  $N$ ，则中等收入人口的比例  $W_{mid}$  为：

$$W_{mid} = \frac{\sum_{i=1}^N M_i}{N},$$

$$\text{其中 } M_i = \begin{cases} 1, & \theta \leq R_i \leq 1 \\ 0, & 0 \leq R_i < \theta \end{cases}$$

上述所提模型综合考虑了个人收入与生活质量对界定中等收入人口范围的影响。不妨举一例做说明，在某地的个人收入情况基本不变的情况下，如果实行了全面的医保，则会降低社会人口的医疗支出，由此生活质量评价指数提高，有更多的人可以达到中等收入吻合度阈值，中等收入人口比重扩大；反之，若没有实行医保，则显然会使中等收入人口比重缩小。这种模型符合人们生活的直观感受，能在确保满足经济直观的同时，使社会群体更易于接受中等收入的界定，避免出现“伪中等收入问题”。

## 六、 结果分析

本文将十个已有的洛伦兹曲线模型与我们新提出的模型进行拟合精度比较，以均方误差为评价标准，我们的模型排第 1；以平均绝对误差为评价标准；我们的模型排第 1；以最大绝对误差为标准，我们的模型排第 3。进而通过对 3 个评价指标进行综合打分处理，我们的模型排在第 2 位。

在第二问中，对传统的收入空间法与人口空间法进行了改进。提出了一种基于比例的界定方法与统计指标的界定方法的综合收入空间界定方法；基于两极分化程度越严重则中等收入人口越小这一经济直观，提出了一种具有经济学意义的基于两极分化指标的改进人口空间法。

在第三问中，针对题中所给 A、B 两个地区前后两年的收入分配情况，我们运用我们提出的六参数洛伦兹模型拟合洛伦兹曲线，运用改进的方法界定中等收入人口比重，得出 B 地区相较 A 地区的经济发展情况更为成熟、稳定。

最后，提出了一个与经济直观相符的，解决目前普遍存在的“伪中等收入问题”的新的个人中等收入吻合度指数  $R$  来度量个人与中等收入人口特征的吻合程度，然后根据该指数的值来界定是否属于中等收入人口范畴。

## 七、 模型评价与拓展

1. 本文提出的洛伦兹曲线模型虽然拟合精度很高，但参数较多，模型较为复杂；可以考虑综合使用多项式拟合、样条拟合等其他拟合方法，以此来进一步提高精度并降低模型复杂度；
2. 本文对收入空间法的改进还有不足，相关参数的定义比较模糊，确定比较随意，可以做进一步优化；
3. 本文对人口空间法进行改进时，仅考虑了两极分化程度的影响，没有考虑其他因素，这显然是不够的。下一步应挖掘更多的经济直观，深入改进人口空间法；
4. 本文新提出的中等收入人口界定模型没有充分地考虑两极分化程度的影响，且未给出具体的函数表达形式，实际操作存在困难。下一步要研究各指数的函数表达形式，使指数的求值过程能简单、直接，并具有深刻的经济意义。

## 八、 参考文献

- [1] Lorenz M, Methods of Measuring the Concentration of Wealth [J], Journal of the American Statistical Association Vol. 9, pp.: 209-219, 1905
- [2] McDonald J.B, Xu Y.J, A generalization of the beta distribution with application [J], Journal of Econometrics Vol. 66, pp.:133-152, 1995
- [3] 成邦文, 基于对数正态分布的洛伦兹曲线与基尼系数[J], 数量经济技术经济研究, 第2期, 2005
- [4] 黄恒君, 位置一尺度分布族的近似构造——以收入分布函数序列为例[J], 管理世界, 第3期, 2003
- [5] Kakwani, Nanak C, On a class of poverty measures [J], Econometrics, Vol. 48, pp.: 761-766, 1980
- [6] 庄健, 张永光, 基尼系数和中等收入群体比重的关联性分析[J], 数量经济技术经济研究, 第4期, 2007
- [7] 陈孝新, 基于多项式样条回归的基尼系数的算法[J], 科技广场, 第2期, 2005
- [8] Cheong K.S, An empirical comparison of a iterative functional forms for the Lorenz curve [J], Econometrics Letters, Vol.9, pp.:171-176

- [9] Gastwirth J.L, The estimation of the Lorenz curve and Gini index [J], Review of Economics and Statistics, Vol. 154, pp.: 306-316, 1972
- [10] Ogwang T and Rao U.L.G, Hybrid models of the Lorenz curve [J], Economics Letters, Vol. 69, pp.:39- 44, 2000
- [11] Wang, Z.X., Y-K Ng, and R. Smyth, A general method for creating Lorenz curves. The Review of Income and Wealth Vol. 57, 561-582, 2011
- [12] Foster, J.E, Wolfson, M.C, Polarization and the decline of the middle class: Canada and the U.S [J], Journal of economic inequality, Vol. 8, pp.: 247-273, 2010
- [13] 胡荣华, 夏德智等, 南京城市居民中等收入界定及分析, 地方经济社会发展研究[J], Vol. 1, 2006
- [14] 庄健, 张永光, 基尼系数和中等收入群体比重的关联性分析[J], 数量经济技术经济研究, 第 4 期, 2007