基于区间型数据的城市家庭收入分布与基尼系数测算方法研究^①

张金宝¹ 廖理²

(1. 北京第二外国语学院财务管理系; 2. 清华大学五道口金融学院)

【摘要】本文提出在不完全信息条件下,估计家庭收入分布和测算基尼系数的方法。利用清华大学中国金融研究中心24个城市的家庭消费金融调查数据,估计我国地级以上城市家庭的收入分布。研究表明,复合概率分布更适合描述我国城市家庭的收入分布,我国城市居民家庭群体具有一定的复杂性。我国地级以上城市居民家庭收入的基尼系数在0.36267~0.38216之间,略高于现有文献对城镇居民收入的基尼系数估计,表明基于数值型数据或分组数据计算的基尼系数存在一定程度的低估。

关键词 区间型数据 收入分布 基尼系数

中图分类号 F064.1 **文献标识码** A

Research on the Method of Estimating the Income Distribution and Gini Coefficient of Urban Household Based on Interval Data

This paper proposes a method of estimating income distribution and Gini coefficient under incomplete information. Using the data derived from the household survey of 24 cities, we study the form of income distribution of urban household and find that composite distribution is more suitable to describe the income distribution of urban household in China, which testifies that urban household in China is made of different groups. The Gini coefficient of urban household is between 0.36267~0.38216, which is slightly higher than the published research results. It suggests that the existing calculation by using numeric data or group data maybe underestimated Gini coefficient to some extent.

Key words: Interval Data; Income Distribution; Gini coefficient

(续)

分布 (缩写)	参数	概率密度函数	累积分布函数
Singh-Maddala (Sm)	$\alpha_1, \alpha_2, \alpha_3$	$\frac{\alpha_1 \alpha_2 \alpha_3 x^{\alpha_2 - 1}}{(1 + \alpha_1 x^{\alpha_2})^{\alpha_3 + 1}}$	$1 - \frac{1}{(1 + \alpha_1 x^{\alpha_2})^{\alpha_3}}$
Weibull (Wb)	α, β $\alpha, \beta > 0$	$\frac{\alpha x^{\alpha - 1} e^{-(x/\beta)^\alpha}}{\beta^\alpha}$	$1 - e^{-(x/\beta)^\alpha}$

四、实证研究

1. 关于数据

本文选用清华大学中国金融研究中心在 2010 年对全国地级以上城市居民家庭进行消费金融调查所收集的数据。调研涉及家庭的基本信息、理财行为、资产负债情况、收入和支出、投资和融资、家庭的住房、医疗和养老保险、家庭的遗产规划等内容。此次调研根据我国地级以上城市规模、经济发展水平的差异,调查将这些城市分为 3 类。即经济发达的城市,经济较发达的城市和经济发展水平一般的城市(廖理和张金宝,2010)。考虑我国地区间经济发展的不平衡,本次调查将全国的城市(不含港澳台地区)在地理位置上按东北、华北、华东、华南、华中、西南、西北 7 个大区进行划分。调查家庭的样本数,由各区间根据人口数量(分别按东北地区每个类的城市 1 个,华北地区每个类的城市 2 个,华东地区每个类的城市 3 个,华南地区每个类的城市 4 个,华中地区每个类的城市 5 个,西南地区每个类的城市 6 个,西北地区每个类的城市 7 个)抽取 100 个市,随机抽样抽取 100 个家庭的样本。调研的数据,基本覆盖我国地级以上城市居民的经济、金融、消费、理财、投资、融资、住房、医疗、养老、保险、遗产规划等各个方面。此次调研共收集到 100 个家庭的收入、支出、理财、投资、融资、住房、医疗、养老、保险、遗产规划等数据。根据调研数据,将家庭收入(千元)的分布,划分为 10 个区间,区间划分的量表为

$[-10, 0], (0, 10], (10, 20], (20, 40], (40, 60], (60, 100], (100, 200], (200, 300], (300, 400], (400, 500], (500, 600], (600, 600]$

变量的 10 个分量(表 1)和 100 个家庭的家庭收入落入每个区间的频数(表 2)

收入区间	1	2	3	4	5	6	7	8	9	10
$[-10, 0]$	1	1	1	1	1	1	1	1	1	1
$(0, 10]$	1	1	1	1	1	1	1	1	1	1
$(10, 20]$	1	1	1	1	1	1	1	1	1	1
$(20, 40]$	1	1	1	1	1	1	1	1	1	1
$(40, 60]$	1	1	1	1	1	1	1	1	1	1
$(60, 100]$	1	1	1	1	1	1	1	1	1	1
$(100, 200]$	1	1	1	1	1	1	1	1	1	1
$(200, 300]$	1	1	1	1	1	1	1	1	1	1
$(300, 400]$	1	1	1	1	1	1	1	1	1	1
$(400, 500]$	1	1	1	1	1	1	1	1	1	1
$(500, 600]$	1	1	1	1	1	1	1	1	1	1
$(600, 600]$	1	1	1	1	1	1	1	1	1	1

注:1. 城市按大区分为:东北、华北、华东、华南、华中、西南、西北 7 个大区;2. 城市按经济水平分为:发达城市、较发达城市、一般城市。