

【统计应用研究】

中国城镇居民家庭收入差异的回归分解

文 晖¹, 徐 梅²

(1. 西安交通大学 经济与金融学院, 陕西 西安 710061; 2. 广西大学 商学院, 广西 南宁 530004)

摘要:用 CHNS 调查数据对中国城市居民的收入差异进行回归分解分析, 得出结论: 教育水平对收入差异的影响显著, 并且影响力逐年提高, 尤其对退休人群养老金收入差异的影响更大; 私营经济部门劳动者的工资收入要高于国有经济部门劳动者的工资收入, 但是私营经济部门退休人群的养老金收入则低于国有经济部门; 处于经济发达地区的人群, 无论是工作人群的劳动收入还是退休人群的养老金收入都高于中西部经济落后地区。

关键词:收入差距; 回归分解; 养老保险

中图分类号:F224 **文献标志码:**A **文章编号:**1007-3116(2009)04-0063-06

建立和谐、可持续发展的社会, 在追求经济增长和发展的同时不能忽视收入差距问题。严重的收入分配差距会引起社会动荡, 影响长期经济发展的数量和质量。改革开放 30 年来, 中国居民收入差距不断扩大已经成为无可争辩的事实, 引起社会各界的广泛关注。工作人群的工资收入和退休人群的养老金收入也存在明显的差异, 是什么因素造成了这些差异? 如何缩小收入差距, 尤其是老年人收入差距? 本文对中国城镇居民收入差异进行回归分解分析, 试图找到上述问题的答案, 并提出相关的建议。

一、引言

一般来说, 影响收入的变量也将决定收入的差异。因此, 经济理论和常识可以用来识别这些变量。换句话说, 人们可以较容易地找出一系列可能揭示收入差异的因素, 例如不同的资源禀赋和政策偏好。然而, 为了排列出政策的优先次序, 有必要从变量对整个差异影响程度的角度将变量排序并解析这一差异。收入差异的分解方法一般都遵循 Shorrocks^[1]的理论。在这些理论构架下, 人们可以通过按人口分组或按要素分组来分解收入差异——前者产生了所谓的“组内因素”和“组间因素”, 它用来调查例如

城乡收入差异、男女收入差异等问题, 但是这种分解方法并没有告诉人们哪些是这两组因素中导致收入差异的最基本的因素; 后者要求知道所有收入来源的完备信息, 要求用要素收入的总和来表示全体收入, 除了难以得到所需的数据外, 这种处理方法也不能起到用来量化收入差异的基本决定因素的作用。用回归分解方法通过量化收入函数中各人口特征对收入差异的贡献率, 分析收入差异的影响因素。回归分解方法之所以新颖, 是由于其可以根据任一种差异度量标准来排列这些决定因素。

二、数据和样本介绍

目前关于收入差异的研究大多从总体水平来考察。从全国来看, 省或地区经常被看作分析的对象。本文将使用家庭水平调查数据计算各种收入差异指标并分解收入差异。数据来源于美国北卡罗来纳大学和中国预防医学科学院联合对中国经济、人口、营养和健康进行的长达十几年的跟踪调查 (China Economic, Population, Nutrition and Health Survey, 以下简称 CHNS)。CHNS 自 1989-2007 年, 每隔三到四年都会进行一次调查, 并对调查得到的数据进行整理并予以公布。并且依据地理位置、经济

收稿日期: 2009-01-15

作者简介: 文 晖(1968-), 女, 湖南桃江人, 博士生, 高级讲师。研究方向: 货币政策与金融理论;

徐 梅(1977-), 女, 广西南宁人, 讲师, 研究方向: 数量经济学。

发展程度、公共资源的丰裕程度和健康指数覆盖了中国东、中和西部 8 个省,对辽宁、江苏、山东、河南、湖北、湖南、广西和贵州进行了调查(1997 年由黑龙江代替辽宁),每个省内的样本依据收入的差异进行选取。城镇方面,选取了省城和较低收入的城镇,以家庭为单位,对家庭中各成员的年龄、工作性质和教育程度等都有记录。Shi, Terry 和 Zhao^[2]简单介绍了样本,还把 1997 年 8 个省的样本与国家统计局的统计数据进行了对比,平均而言,CHNS 样本中农村和城镇居民家庭的收入低于国家统计局的调查,但差别不是很大,并认为该样本对中国居民家庭收入的调查是有代表性的,并可以作为数据进行分析。

本文使用 CHNS 中关于城镇居民家庭收入的数据,跨越 8 个地区,16 个城镇,共 2 658 个调查对象,涉及被调查者的收入水平、教育年限、年龄、工作类型、家庭负担水平和居住地区等,其中 1991 年整理了 586 个调查对象的资料、1993 年有 513 个、1997 年有 637 个、2000 年有 691 个、2004 年有 231 个。并且对各年的收入数据进行了通货膨胀调整,虽然不能说这组数据能代表全国,但能够涵盖不同的地理经济状况,而且比只依靠一年或一个省份数据的研究更有代表性。

三、回归分解方法和收入函数

回归分解方法是 Oxaca^[3]20 世纪 70 年代初提出的,但并没有引起人们足够的重视,直到最近情况才有所改变^[4],Wan^[5]曾提出了关于这种技术的具体描述。具体方法步骤如下:

(一)建立收入函数回归方程

对收入差异的回归分解方法是基于回归方程的结果来分析的,得到一个收入函数回归方程是进行回归分解的第一步。在确定中国城镇居民的收入函数时,有关人力资本理论认为,工资收入者依靠出卖劳动获得收入。按照该理论,这里应包括技能变量,例如教育、训练程度及经验(常用年龄代表)。为了和此类研究的普遍做法相一致,这里采用城镇居民的受教育水平及年龄。

即使人力资本是相同的,考虑那些可能改变收入的其他因素也是必要的。一个因素是城镇居民的工作类型。不同工作单位类型的劳动者工资收入和退休计划都会有所不同。显然,要表达来自不同工作类型的收入水平的差别,设定一系列虚拟变量是很必要的。另一个因素则是家庭中的负担比例。工作人数在家庭总人数中的比例影响着家庭收入,并

进一步影响着家庭生活水平。

地理位置是决定收入的一个很重要的因素。中国的地区经济发展和地理位置有着极强的对应关系,东部沿海地区是经济发达地区,中部地区是经济发展中等地区,而西部地区一般经济发展比较落后。应用地区虚拟变量能够反映地理和所在地的信息。

这样,收入函数所包含的变量有因变量和自变量。因变量是家庭人均年收入。自变量有:1. 教育:户主的受教育年限;2. 年龄:户主的年龄以及户主年龄的平方;3. 劳动者比例:家庭中劳动者占家庭总人数的比例;4. 虚拟变量:工作类型(政府机关和国有企事业单位、集体企业、三资企业和个体私营企业)和地理位置(东部地区、中部地区和西部地区)。

收入函数的公式为:

$$\text{收入} = f(\text{教育, 年龄, 劳动者比例, 工作类型虚拟变量, 地区虚拟变量}) \quad (1)$$

(二)回归分解方法介绍

当然,对收入差异的回归方程分解不仅需要确定收入函数,还需要选择合适的差异指标。由于本文使用的人口特征中包含很多定性变量,只能通过哑元变量的形式来纳入回归模型中,因此,本文基于回归方程的收入来源分解使用基尼系数作为收入差异的衡量指标。

基尼系数是 20 世纪初意大利经济学家 Gini 在洛伦兹曲线的基础上提出的一个用以衡量收入或财富分配不均等程度的测度指数,在实证研究和政策分析中得到了广泛应用^[6-7]。在实际运用中,根据不同的数据特点基尼系数有多种计算方法,由于本文使用的调查样本数据量较大,因此使用统计软件易于处理的协方差方法计算基尼系数,具体计算公式为:

$$G(y) = \frac{2}{ny} \text{cov}(y, r(y)) \quad (2)$$

其中 \bar{y} 表示 y 的均值, $r(y)$ 表示变量 y 的序数,即当有 n 个观测值时, y 值最小的被定义为 $r(y) = 1$,第二小的被定义为 $r(y) = 2$,最大的 y 值被定义为 $r(y) = n$ 。

在回归模型的选择上,由于计算基尼系数必须使用收入的原始观测值,即回归模型的因变量也必须是收入的原始观测值,因此本文假设收入决定函数 $y = f(x)$ 为一元线形回归函数,收入 y 可以表示为:

$$y = \alpha + \sum \beta_i x_i + \epsilon \quad (3)$$

其中 $\beta_i x_i$ 表示第 i 项个人特征产生的收入流。

按照费景汉—兰尼斯分解公式^[8],可以将基尼系数同时应用于式(3)两边,得到:

$$G(y) = \sum \phi_i R(y, x_i) G(x_i) \quad (4)$$

其中 $G(y)$ 表示个人收入基尼系数, $\phi_i = \frac{\beta_i x_i}{y}$ 表示个人特征收益权数,反映第 i 种个人特征产生的收入流在家庭收入中的比重; $R(y, x_i)$ 表示个人特征与收入之间的序数相关比,计算公式为: $R(y, x_i) = \frac{\text{cov}(x_i, r(y))}{\text{cov}(x_i, r(x_i))}$; $G(x_i)$ 表示各项个人特征的分布基尼系数。

分解公式(4)可以扩展到回归模型中自变量与因变量之间存在非线性关系的分解分析上,这时需要计算出某项个人特征产生的收入流,然后再考察该收入流对总的收入差异的影响。例如,假设年龄因素产生的收入流为:

$$y_{age} = \beta_1 \times age + \beta_2 \times age^2$$

相应地年龄因素对总的收入差异的贡献为:

$$CO_{age} = \frac{\bar{y}_{age} R(y, y_{age}) G(y_{age})}{\bar{y}}$$

从定义上来看,残差项 ϵ 是“白噪声”。在计量经济学分析中通常忽略其对因变量的影响,但由于残差项代表观测不到的因素或不包括在回归模型中的因素也决定了个人收入 y 的大小,进而对居民收入差异产生影响,因此 Wan^[9] 在使用回归方程分解收入差异时提出考虑残差项的影响。类似的分析也被用于对截距项的讨论,当将一个正的常数加到每个人身上时,基尼系数必然会下降,因此在分解基尼系数时也需要考虑截距项的影响。根据这一思路,笔者按照 Shorrocks 提出的分解的自然法则对费景汉—兰尼斯分解公式进行了一定的修正,计算回归模型中截距项和残差项对总体收入差异程度的贡献。

首先考虑残差项 ϵ 的影响,假设 $y(\epsilon = 0) = \alpha + \sum \beta_i x_i = y'$, 则 $G(y | \epsilon = 0) = G(y')$, 那么, ϵ 对总收入差异的贡献可以定义为:

$$CO_\epsilon = G(y) - G(y') \quad (5)$$

在式(4)中,截距项 α 对总的收入差异程度没有影响,但随着 α 的变化,个人收入均值 \bar{y} 会随之发生变化,进而对个人特征收益权数 $\phi_i = \frac{\beta_i x_i}{\bar{y}}$ 产生影响。例如, α 值的增加会带来 \bar{y} 的增加, ϕ_i 进而减小。因此,可以认为截距项对收入差异的贡献被分配到了式(5)右边的各项中。

设 $y'(\alpha = 0) = \sum \beta_i x_i = \bar{y}$, 则 $G(y' | \alpha = 0)$

$= G(\bar{y})$, α 对总收入差异的贡献为:

$$CO_\alpha = \frac{\beta_i \bar{x}_i R(\bar{y}, x_i) G(x_i)}{\bar{y}} \quad (6)$$

四、回归分解结果和讨论

本文首先以家庭人均收入作为被解释变量,对调查中所有城镇居民家庭人均收入进行回归,得到结果见表1。教育水平在各年间一直对家庭人均收入有显著的影响,并且教育程度对家庭人均收入的正向影响力在加强;家庭劳动者比例越高,家庭人均收入也越大,其影响也是显著的;工作单位类型中,私营程度越高,家庭人均收入越低,其中1991和1993年影响显著;被调查者居住地经济程度越发达,其家庭人均收入越高,其中2000年和2004年影响是显著的。值得注意的一点,无论是年龄还是年龄的平方项,对家庭人均收入均无显著影响,这令人迷惑不解。

根据生命周期理论,个人的收入将随着其所处生命周期的不同而呈现不同的起伏,因此,本文进一步将人群分为工作人群和退休人群,建立回归方程进行分析,只是工作人群的被解释变量换成了劳动收入,退休人群的被解释变量换成了养老金收入(见表1)。通过分析发现,教育水平对退休人群的养老金收入影响程度要更甚于工作人群的劳动收入。工作单位类型对劳动收入和养老金收入的影响正好相反:私营程度越高,劳动收入越高,但是退休后的养老金收入却越低。如果居民处在东部经济发达地区,无论是工作人群的劳动收入,还是退休人群的养老金收入,都比中西部地区有明显的提高。

中国目前教育水平较高的退休人群大多集中于政府部门和国有企事业单位,这些部门是养老金体系相对比较完善的领域,所以教育水平凸现了养老金的差距。随着改革开放的深化,中国私营经济逐渐深入发展,处于该经济体的工作人群收入提高幅度快,其中许多人的工资甚至超过了处于国有经济部门内的劳动者,但是由于中国养老保险体系的滞后,覆盖面还只是在国家机关、国有企事业单位和部分集体企业,所以退休人群养老金的变化和工作人群劳动收入的变化恰恰相反,处于国有经济部门的退休者比处于私营经济部门的退休者能得到更多的养老保障。经济发达地区本身的收入水平就比较高,并且社会养老保险也比较完善,所以能获得更多的劳动收入和养老金收入。

表1 中国城镇居民收入的回归结果表

因变量	自变量	1991年	1993年	1997年	2000年	2004年
家庭人均收入(全部样本)	教育水平	20.68***	26.57***	30.92***	55.40***	131.54***
	年龄	20.23	54.53	33.34	41.37	100.02
	年龄平方	-0.29	-0.40	-0.23	-0.11	-0.21
	家庭中劳动者比例	403.21**	569.46**	1207.99***	409.73**	3809.24***
	工作类型 ^a	122.88***	102.27**	48.77	43.69	-58.61
	地区	4.00	58.27	-48.48	154.05*	518.54**
	F值	13.84***	6.72***	20.03***	16.31***	26.79***
	可决系数 Adj. R ²	0.1381	0.0766	0.1895	0.1416	0.2347
	样本量 n	482	415	523	559	531
	劳动收入(工作人群)	教育水平	8.56	67.46*	222.78***	490.62***
家庭中劳动者比例		4526.74***	5864.12***	10904.05***	10997.50***	30523.37***
工作类型		45.92	807.37***	478.90*	201.07	311.07
地区		-613.69***	-590.55	519.27	1822.41	3994.82***
F值		25.58***	10.31***	31.33***	8.84***	42.50***
可决系数 Adj. R ²		0.2039	0.1045	0.2368	0.0752	0.3400
样本量 n		385	320	379	387	335
养老金(退休人群)		教育水平	32.61***	41.79*	75.77***	286.51***
	家庭中劳动者比例	-587.92	-1444.85	2635.78**	-3515.97	-10364.55***
	工作类型	18.98	-51.60	-102.8	-925.86**	-1069.04
	地区	547.41***	86.55	-372.62	1414.86*	2327.58**
	F值	6.53***	1.59	3.17**	13.97***	35.712***
	可决系数 Adj. R ²	0.1873	0.0245	0.063	0.2338	0.4165
	样本量 n	97	95	145	172	196

注：*表示在10%水平上显著；**表示在5%水平上显著；***表示在1%水平上显著。工作类型虚拟变量的设置：政府机关和国有企事业单位=0，集体企业=1，三资企业=2，个体私营企业=3；地区虚拟变量的设置：西部=0，中部=1，东部=2。

对收入函数回归结果作了初步分析后，现采用回归方程的收入来源分解方法计算出中国城镇居民家庭人均收入差异的贡献如表2所示。通过分析发现，本文列出的人口特征对中国城镇居民家庭人均收入差距的解释能力越来越低，说明影响家庭人均收入差异的因素越来越多，比如家庭投资就是目前一个很重要的家庭收入来源。从表面上看，生命周期理论在实际家庭人均收入分布格局上并不具有解释力，但是如果将人群分为工作人群和退休人群之后，对不同人群主要收入的差异进行分解可以进一步发现这些人口特征的影响作用。

教育水平、年龄结构分布、家庭中劳动者的比例、工作类型和地区等特征对养老金收入的解释能力比较高，基本上能够说明退休人群养老金收入的分配差异。教育水平对养老金收入差异的贡献率最大，曾一度达到了60%以上。地区间经济差异对养老金分布差异的贡献率明显大于对家庭人均收入和工作人群的劳动所得差异的贡献率，由此判断，地区间的社会养老保险体系的完善程度并不一致，社会养老保险也需要经济发展给予资金支持。由于工作人群收入主要是通过劳动取得，所以家庭中劳动者人数比例对其劳动收入分布差异的影响更大一些。

表 2 基尼系数的回归分解结果表

	1991年		1993年		1997年		2000年		2004年	
	$G(x)$	贡献率	$G(x)$	贡献率	$G(x)$	贡献率	$G(x)$	贡献率	$G(x)$	贡献率
全 家庭人均收入	0.268 7		0.316 4		0.288 5		0.336 1		0.410 5	
部 教育水平	0.004 1	40.72%	0.339 3	14.38%	0.269 4	20.14%	0.252 0	21.2%	0.212 6	12.69%
样 年龄	0.345 5	22.46%	0.055 3	-0.42%	0.069 0	-5.55%	0.126 0	1.19%	0.121 6	-1.96%
本 家庭中劳动者比例	0.278 8	16.55%	0.337 1	6.89%	0.400 3	29.71%	0.529 6	2.07%	0.693 8	11.92%
工作类型	0.482 6	9.04%	0.539 6	2.82%	—	—	0.986 7	-0.04%	—	—
地区	0.258 0	0.06%	0.270 0	0.55%	0.227 6	-0.29%	0.197 1	1.91%	0.189 5	1.65%
所有个人特征		88.83%		24.22%		44.51%		26.33%		24.31%
工 劳动所得	0.295 5	0.401 1		0.372 5		0.397 8		0.607 5		
作 教育水平	0.294 9	2.81%	0.260 8	8.76%	0.180 6	13.83%	0.145 8	11.51%	0.130 1	3.00%
人 家庭中劳动者比例	0.222 9	96.51%	0.263 9	33.34%	0.261 2	34.12%	0.300 3	21.28%	0.430 8	35.00%
群 工作类型	0.470 8	0.66%	0.520 3	9.75%	0.655 3	3.15%	0.730 2	0.46%	0.785 4	1.26%
地区	0.226 7	10.30%	0.232 1	-0.43%	0.201 8	0.80%	0.174 1	0.56%	0.175 1	1.95%
所有个人特征		110.28%		51.42%		51.90%		33.81%		41.21%
退 养老金收入	0.856 9		0.651 9		0.611 7		0.508 4		0.458 4	
休 教育水平	0.781 4	10.80%	0.615 2	49.89%	0.593 8	61.65%	0.417 6	53.95%	0.368 5	62.32%
人 家庭中劳动者比例	0.494 3	5.19%	0.587 7	40.50%	0.955 5	28.03%	—	—	—	—
群 工作类型	0.509 6	-0.06%	0.639 6	0.91%	—	—	—	—	—	—
地区	0.473 2	40.39%	0.496	0.84%	0.341 7	7.47%	0.195 9	8.79%	0.217 9	7.26%
所有个人特征		56.32%		92.14%		97.15%		62.74%		69.58%

五、结束语

本文采用回归分解技术分析中国城镇居民收入差异问题,使用家庭收入水平数据是对现有研究的补充。通过研究发现,教育程度是导致收入差异最主要的因素。私营经济部门的劳动者收入要更高,但退休人群养老金收入却更低;而经济发达地区的

工作人群和退休人群收入比经济落后地区高。

以上分析说明中国实行社会养老保险制度是有一定成效的,缓解了不同工作类型工资收入的差异。今后应该打破“双轨制”,提高社会养老保险的覆盖面,并逐步实现全国统筹。

教育对收入差异的影响很大,因此,应加强城市低收入群体的教育培训,进而促进其收入的提高。

参考文献:

- [1] Shorrocks A F. Inequality decomposition by factor components[J]. *Econometrica*, 1984, 52(6): 1369-1385.
- [2] Shi X Zh, Terry, Zhao Y H. Analyzing urban - rural income inequality in China[C]. Discussion Draft, CCER, Peking University.
- [3] Oxaca R. Male - female wage differences in urban labor markets[J]. *International Economic Review*, 1973, 14(3): 693-709.
- [4] Juhn C, Kevin M M, Brooks P. Wage inequality and the rise in returns to skill[J]. *Journal of Political Economy*, 1993, 101(3): 410-412.
- [5] Wan G H. Regression - based inequality decomposition: pitfalls and a solution procedure[C]. World Institute for Development Economics Research Discussion Paper, 2002.
- [6] 徐宽. 基尼系数的研究文献在过去八十年是如何扩展的[J]. *经济学季刊*, 2003(4): 757-778.
- [7] 李虎. 关于基尼系数分解分析的讨论[J]. *数量经济技术经济研究*, 2005, 22(3): 127-135.
- [8] Fei J, Rainis G, Kuo S. Growth and the family distribution of income by factor components[J]. *Quarterly Journal of Economics*. 1979(1):17-53.

(下转第 78 页)

(上接第 67 页)

[9] Wan G H. Accounting for income inequality in rural China: a regression - based approach[J]. Journal of Comparative Economics, 2004, 32(2): 348 - 363.

(责任编辑:马 慧)

Regression - based Household Income Inequity Decomposition of Chinese Urban Residents

WEN Hui¹, XU Mei²

(1. School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, China;

2. School of Business, Guangxi University, Nanning 530004, China)

Abstract: Using the CHNS survey data, this paper decomposed the household income inequity based on regression of Chinese Urban residents. We had some conclusions: firstly, the educational level influences the income inequity significantly year by year; secondly, the average wage of private sector more than that of state sector, but the average pension of private sector less than that of state sector; finally, whether wage of the work force or pension of the retired in economically developed areas more than that of economically undeveloped areas.

Key words: income inequity; regression - based decomposition; social pension