

## 附录

# 中国城镇家庭财富差距总体考察、 来源分解与形成机制

张 平 诸葛安东 田 也

## 附录一、基尼系数修正方式

常规基尼系数的表达式如下：

$$Gini_{it} = 2 \times \int_0^1 [p - L(p)] dp \quad (1)$$

其中,  $Gini_{it}$  表示财政年度  $t$  内聚类  $i$  的基尼系数,  $p$  和  $L(p)$  分别表示洛伦兹曲线中绝对均等和现实情况下家庭财富的百分比累积分布函数。

传统的基尼系数假定分布函数中的观测值均为非负数, 当出现负数时, 基尼系数的统计量将失去原有意义。本文的研究对象为家庭财富净值, 考虑到存在家庭财富净值为负, 即总负债大于总财富的情况, 本文采用 Diaz-Bazan(2014) 的基尼系数分解公式以及艾小青(2017)提出的方法对负净值财富基尼系数加以修正, 修正后的基尼系数计算公式如下：

$$Gini_{it} = P_1^2 \frac{|\mu_1|}{\mu} G_1 + P_2^2 \frac{\mu_2}{\mu} G_2 + P_1 P_2 \frac{|\mu_1 - \mu_2|}{\mu} \quad (2)$$

其中,  $P_1$  和  $P_2$  分别表示样本中家庭财富净值为负和为正的样本数量占全部样本数量的比重,  $G_1$  和  $G_2$  表示两个子样本单独测算得到的基尼系数,  $\mu$  为总样本的家庭财富均值,  $\mu_1$  和  $\mu_2$  为子样本的财富均值。修正后的负净值样本基尼系数在原有洛伦兹分布基础上取绝对值, 考虑到负净值样本的财富均值为负, 对  $\mu_1$  取绝对值以使其与修正后的基尼系数符号一致。

另外, 本文数据来源为不同年份的家户调查数据, 可能存在极高财富净值人群样本代表性不足的问题(李实、罗楚亮, 2011; 李家山等, 2021)。基尼系数对极高值的敏感度较高, 极高财富净值人群数据的缺失可能导致估计存在严重偏误。因此, 本文采用罗楚亮和陈国强(2021)的方法, 将相应年份的福布斯中国内地富豪榜数据和家户调查数据进行“拼接”。首先用帕累托分布拟合高财富净值人群的分布特征, 帕累托分布的对数线性表达式为：

$$\ln N = \ln K + \alpha \ln y \quad (3)$$

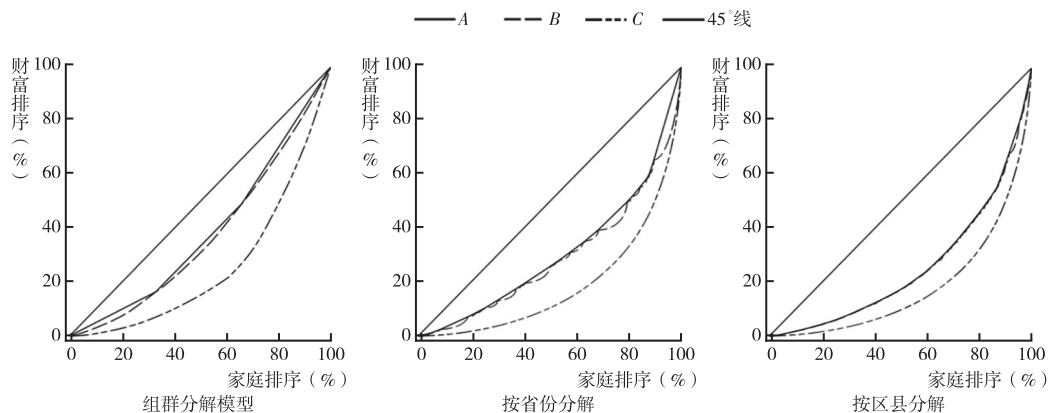
其中,  $N$  为财富个体  $y$  在样本中的位次,  $\ln K$  和  $\alpha$  为待估参数。式(3)建立了财富净值  $y$  和财富位次  $N$  的线性回归函数, 可用线性最小二乘法估计得出  $\alpha$ , 从而求出基尼系数。帕累托分布下基尼系数和参数  $\alpha$  存在以下关系：

$$Gini = \frac{1}{2\alpha - 1} \quad (4)$$

假设家户调查数据中包含了部分极高财富净值人群,但这部分样本缺乏代表性,现以绝对标准在原家户调查数据99%分位点“拼接”相应年份的福布斯中国内地富豪榜。首先,分别估计净财富位于99%分位点以下家庭样本的基尼系数和财富均值;其次,根据式(3)推算99%分位点实际的门槛值;最后,根据式(4),用99%分位门槛值和参数 $\alpha$ 估计极高财富净值家庭的财富均值和基尼系数,最终拟合得到相应年份的修正后基尼系数。

## 附录二、基尼系数组群分解的几何含义

图A1展示了组群分解在洛伦兹曲线中的几何含义(Lambert和Aronson,1993)。洛伦兹曲线反映了家庭财富占比累积分布的情况,曲线C表示全样本的洛伦兹曲线,这一曲线越接近45°线,表示财富分布越均衡。如组群分解模型所示,假设将样本分为3组,曲线A表示组内财富全部相同情况下的财富累积分布,在此情况下组内差距为零,其与45°线所围成图形的面积表示全部的组间差距;曲线B表示在不同组群分布不重叠情况下引入组内差距后的财富累积分布,曲线B和曲线A所围成图形的面积体现了组内差距;当不同组群分布有所重叠时,真实的洛伦兹曲线如曲线C所示,曲线C和曲线B所围成图形的面积体现了交叠项对总基尼系数的贡献情况。图A1还展示了将2020年中国家庭追踪调查(CFPS)家户调查数据按省份和区县分解的情况,相较于按区县分解,按省份分解的组间差距贡献较小,组内差距贡献较大,这反映了省份之间的差距小于区县之间的差距,但省份内部的差距大于区县内部的差距。另外,按省份分解的交叠项贡献较大,反映了不同省份之间财富分布的重叠程度较高。通过组群分解,可以清晰地展示不同群体的组间差距、组内差距和交叠项变化的情况。



图A1 组群分解的洛伦兹曲线

附录三、微观数据地区分布

表 A1 各年份 CHIP 和 CFPS 抽样地区分布

省份	中国家庭收入调查(CHIP)				中国家庭追踪调查(CFPS)						合计
	1995 年	1999 年	2002 年	2008 年	2010 年	2012 年	2014 年	2016 年	2018 年	2020 年	
北京	500	599	484	0	98	64	124	113	115	97	2194
天津	0	0	0	0	55	50	64	67	62	58	356
河北	0	0	0	0	236	209	294	310	309	295	1653
山西	650	0	640	0	211	169	177	194	209	206	2456
内蒙古	0	0	0	0	0	0	4	6	8	5	23
辽宁	698	649	697	0	785	583	696	674	674	503	5959
吉林	0	0	0	0	169	145	144	159	128	103	848
黑龙江	0	0	0	0	386	304	331	292	312	225	1850
上海	0	0	0	500	1078	598	713	743	708	492	4832
江苏	800	647	729	600	162	162	189	207	187	163	3846
浙江	0	0	0	600	86	99	144	162	194	138	1423
安徽	499	0	492	550	139	132	167	175	172	129	2455
福建	0	0	0	0	68	41	60	71	77	59	376
江西	0	0	0	0	55	61	60	74	84	70	404
山东	0	0	0	0	239	225	236	275	310	264	1549
河南	600	698	680	650	569	556	653	611	634	508	6159
湖北	741	0	673	400	182	136	127	149	134	98	2640
湖南	0	0	0	0	209	162	245	292	264	193	1365
广东	540	0	544	699	696	551	651	736	778	600	5795
广西	0	0	0	0	52	59	106	115	107	88	527
海南	0	0	0	0	0	0	3	5	6	2	16
重庆	0	0	279	398	85	53	59	69	75	62	1080
四川	847	798	585	600	292	251	283	317	300	202	4475
贵州	0	0	0	0	105	115	108	135	117	96	676
云南	648	0	636	0	58	75	117	114	128	118	1894
西藏	0	0	0	0	0	0	0	1	2	2	5
陕西	0	0	0	0	120	121	148	160	146	128	823
甘肃	400	594	395	0	248	258	332	386	464	424	3501
青海	0	0	0	0	0	0	0	0	2	2	4
宁夏	0	0	0	0	0	0	2	3	4	5	14
新疆	0	0	0	0	0	0	5	12	29	31	77
合计	6923	3985	6834	4997	6383	5179	6242	6627	6739	5366	59275

附录四、家庭财富分类方式

表 A2 城镇家庭财富结构及其测量指标		
一级指标	二级指标	三级指标
资产	总房产	现住房市场价值
		其他房产市场价值
	金融资产	储蓄
		股票、基金、债券、金融衍生品、借给他人的款项、其他金融产品
	耐用消费品	汽车、电视、计算机、冰箱等家庭常见消费品
	生产性固定资产	农业使用机械、经营性企业资产等
负债	土地	—
	住房负债	连本带息尚未还清的建房、购房贷款或借款
	非住房负债	教育、医疗、购买耐用消费品等方面的债务
净资产	净房产	总房产-住房负债
	净金融资产	金融资产-非住房负债
	其他净资产	耐用消费品+生产性固定资产

附录五、基尼系数的交叉分解

表 A3 展示了不同财富构成按省份交叉分解的结果,即探讨不同财富构成在省份间的组间差距和组内差距的总体情况。

首先,本文按资产类型进行分类。结果表明,对于房产财富而言,组内差距的贡献始终较小,表明省份内部不同家庭的房产财富较为接近;组间差距的贡献呈现逐年增大的趋势,与此同时,交叠项系数则不断减小,这表明不同地区的房产价值差距持续扩大,城镇家庭房产财富的分布愈加不均衡。对于金融资产而言,组间差距和交叠项对财富差距的贡献较大,且两者的系数并未呈现随时间变化的明显趋势。对于其他资产而言,交叠项的贡献最大,这表明不同省份城镇家庭中其他资产的分布重叠程度较高,其他资产的地区间分布较为均衡。

其次,按自住房增值和自有收入积累分类的分解结果与按资产类型分类的结果相似。自住房增值反映房产价值增长的收益,在这部分财富中,对财富差距贡献最大的依然是组间差距,且这一比重常年接近 70%,表明地区间的房价增长存在较大差距。自有收入积累则是家庭财富中除去自住房增值外的其他部分,主要反映了来自工资性收入的积累和储蓄偏好。自 2010 年起,自有收入积累的组间差距贡献持续减小,交叠项系数上升,这表明全国范围内地区间的工资性收入积累差距逐渐缩小。

此外,本文还分别考察了有房家庭和无房家庭的基尼系数组群分解情况。结果表明,有房家庭之间的财富基尼系数小于无房家庭,这一结论与前文的分析一致,即家庭房产财富是家庭财富构成的主要部分,有房家庭的内部异质性小于无房家庭。从分解结构看,有房家庭的财富基尼系数主要由组间差距构成,这反映了房产财富在不同省份分布的差异性;而无房家庭的财富基尼系数主要由交叠项构成,且随着时间的推移,组间差距的影响不断减小,这反映了工资性收入在不同区域的均等化程度不断提高,各省份无房家庭通过自有收入积累的财富分布重叠程度不断提升。

表 A3		基尼系数构成按省份交叉分解									
财富构成	组群分解	1995 年	1999 年	2002 年	2008 年	2010 年	2012 年	2014 年	2016 年	2018 年	2020 年
房产	基尼系数	0.826	0.651	0.541	0.457	0.687	0.645	0.613	0.730	0.629	0.688
	组内差距 (%)	7.5	14.3	7.0	9.2	8.3	6.5	6.0	6.1	5.9	5.6
	组间差距 (%)	70.0	37.8	55.3	59.2	69.8	58.3	62.6	67.5	65.4	65.5
	交叠项 (%)	22.5	48.0	37.7	31.6	21.9	35.2	31.3	26.4	28.8	28.9
金融资产	基尼系数	0.612	0.635	0.615	0.623	0.805	0.775	0.718	0.787	0.759	0.751
	组内差距 (%)	8.6	15.7	8.4	11.0	9.9	7.3	7.4	7.0	6.8	6.2
	组间差距 (%)	44.6	43.9	39.5	47.0	55.8	48.4	54.6	48.2	46.8	47.9
	交叠项 (%)	46.8	40.4	52.1	42.1	34.3	44.3	38.1	44.8	46.4	45.9
其他资产	基尼系数	0.498	0.565	0.617	0.635	0.808	0.752	0.660	0.711	0.662	0.627
	组内差距 (%)	8.8	16.2	8.4	11.8	7.8	6.4	6.5	5.8	6.1	6.1
	组间差距 (%)	39.8	24.9	29.6	37.6	38.2	26.9	27.5	35.2	28.3	31.4
	交叠项 (%)	51.4	59.0	62.0	50.7	54.1	66.6	66.0	58.9	65.5	62.6
自住房增值	基尼系数	NA	NA	NA	NA	0.659	0.654	0.696	0.747	0.736	0.754
	组内差距 (%)	NA	NA	NA	NA	9.2	8.5	5.8	5.9	6.0	6.0
	组间差距 (%)	NA	NA	NA	NA	78.8	73.7	69.1	73.0	71.4	73.0
	交叠项 (%)	NA	NA	NA	NA	12.0	17.8	25.1	21.1	22.6	21.0
自有收入积累	基尼系数	NA	NA	NA	NA	0.670	0.617	0.582	0.584	0.592	0.600
	组内差距 (%)	NA	NA	NA	NA	10.7	9.6	6.8	6.3	6.4	6.6
	组间差距 (%)	NA	NA	NA	NA	53.5	51.8	42.3	38.0	34.4	39.2
	交叠项 (%)	NA	NA	NA	NA	35.8	38.5	50.9	55.7	59.2	54.2
有房家庭净财富	基尼系数	0.494	0.454	0.440	0.438	0.641	0.604	0.618	0.656	0.639	0.651
	组内差距 (%)	8.5	14.8	7.2	9.2	7.2	6.3	6.0	6.1	5.8	5.7
	组间差距 (%)	57.4	49.3	54.8	60.7	68.2	53.7	63.3	63.3	63.4	65.2
	交叠项 (%)	34.1	35.9	38.0	30.2	24.6	40.0	30.7	30.7	30.7	29.2
无房家庭净财富	基尼系数	0.503	0.615	0.580	0.633	0.864	0.801	0.753	0.827	0.606	0.737
	组内差距 (%)	9.3	18.1	9.9	49.9	27.4	25.1	14.3	8.4	9.0	11.6
	组间差距 (%)	45.2	37.6	48.4	9.5	56.8	55.2	33.0	57.6	57.6	18.3
	交叠项 (%)	45.6	44.3	41.7	40.6	15.8	19.7	52.7	34.1	33.4	70.2

附录六、回归分析控制变量描述性统计

表 A4		回归分析控制变量描述性统计						
层面	变量符号	变量名称	数据来源	样本量	均值	标准差	最小值	最大值
家庭	<i>fri</i>	家庭总收入(万元)	CFPS、CHIP	57578	5.282	8.972	-0.535	603.360
	<i>edu</i>	户主受教育年限(年)	CFPS、CHIP	57912	9.679	4.277	0	35
	<i>health</i>	户主健康状况	CFPS、CHIP	52235	0.219	0.414	0	1
	<i>age</i>	户主年龄	CFPS、CHIP	59273	48.761	13.966	19	97
	<i>fsize</i>	家庭规模(人)	CFPS、CHIP	59275	3.379	1.454	1	17
省份	<i>gdp</i>	省份 GDP(万亿元)	《中国统计年鉴》	189	2.367	2.026	0.055	11.076
	<i>pop</i>	省份总人口(亿人)	《中国统计年鉴》	189	0.526	0.259	0.125	1.260
	<i>gdpr2</i>	第二产业增加值占比(%)	《中国统计年鉴》	189	44.334	7.914	15.8	57.3
	<i>gdpr3</i>	第三产业增加值占比(%)	《中国统计年鉴》	189	45.155	10.408	27.788	83.900

注:户主健康状况为受访者对自身健康状况的自我评估值,为了使数据具有纵向可比性,本文将其处理为虚拟变量,1 表示更好的健康状况。

## 附录七、无条件分位数回归

考虑到净财富的偏态分布情况,本文采用无条件分位数回归(Unconditional Quantile Regression, UQR)深入探究不同财富水平城镇家庭的财富积累机制。相较于均值回归, UQR 可以放松对扰动项同方差和正态分布的假定,同时探究自变量对不同分布样本的差异性影响(Firpo等, 2009)。对于不同分位数水平的城镇家庭来说,其净财富的影响因素不尽相同,采用UQR可以分析家庭净财富影响因素的全面信息。本文选择5个具有代表性的分位数10%、25%、50%、75%、90%估计结果。回归结果如表A5所示,人均财政支出对各分位数的家庭净财富均具有显著的正向作用,同时这一效应对家庭净财富分位数更高的家庭影响更大,表明地区差异是导致家庭净财富差异的主要因素。自住房持有年限对家庭净财富的影响基本随着分位数的提高而逐渐增大,但对净财富极高人群(90%分位数)影响不显著。

表 A5 城镇家庭净财富主导因素的分位数回归结果

变量	(1)	(2)	(3)	(4)	(5)
	家庭净财富				
	10%	25%	50%	75%	90%
人均财政支出	2.786*** (0.615)	9.587*** (1.368)	49.096*** (2.859)	118.015*** (5.199)	211.184*** (9.187)
自住房持有年限	0.134*** (0.006)	0.150*** (0.007)	0.116*** (0.009)	0.105*** (0.027)	0.058 (0.063)
省份固定效应	是	是	是	是	是
年份固定效应	是	是	是	是	是
控制变量	是	是	是	是	是
样本量	47997	47997	47997	47997	47997

注:括号内为经过Robust调整的标准误,经过了家庭层面的聚类修正;\*、\*\*和\*\*\*分别表示在10%、5%和1%的水平下显著。下同。控制变量与正文中表4一致。

## 附录八、工具变量回归

本文采用清代驿站数量作为交通基础设施的工具变量。古代驿站设置与现代交通基础设施具有一定的相关性,一方面,铺设驿站的规律与现代基础设施建设相似;另一方面,驿站设立在客观上促进了区域经济发展(欧阳艳艳、张光南,2016)。因此,省份内各地区清代驿站数量的差距能够反映地区间交通基础设施的差距,且清代驿站数量作为前定变量,与模型中其他未控制变量无关,具有较好的外生性。

如表A6所示,本文采用两阶段最小二乘法(2SLS)估计。根据第一阶段的估计结果,清代驿站数量基尼系数与交通基础设施基尼系数显著正相关,验证了前文关于工具变量相关性的假设。同时,从弱工具变量检验结果看,列(2)的Kleibergen-Paap rk Wald F统计量低于临界值(Andrews和Stock, 2005),可能存在弱工具变量的风险。目前学界普遍采用有限信息最大似然法(LIML)解决

弱工具变量问题,如果存在弱工具变量问题,则 LIML 法的小样本性质优于 2SLS 法(袁歌骋、潘敏, 2021)。本文采用 LIML 法估计后的回归结果与 2SLS 法估计结果没有明显区别,表明弱工具变量对估计结果的影响较小。第二阶段的估计结果与正文中的表 6 相似,用于衡量省份内地区间组间差距的交通基础设施基尼系数与净财富基尼系数显著正相关,进一步验证了上述实证结果的稳健性。

表 A6 各省份财富基尼系数与省份内不同维度差距(工具变量估计)

变量	(1)	(2)
	净财富基尼系数	
交通基础设施基尼系数	1.270*** (0.459)	1.717** (0.697)
第一阶段因变量	交通基础设施基尼系数	
清代驿站数量基尼系数	0.246*** (0.073)	0.194** (0.076)
地区固定效应	是	是
年份固定效应	是	是
省级控制变量	否	是
样本量	116	116
Kleibergen-Paap rk Wald F 统计量	13.056	7.646

参考文献:

1. 艾小青:《财富不平等的度量:存在负值时基尼系数的应用》,《统计与决策》2017 年第 4 期。  
2. 李家山、易行健、何启志:《中国居民财富不平等的测算修正、异质性与驱动机制》,《财政研究》2021 年第 12 期。  
3. 李实、罗楚亮:《中国收入差距究竟有多大?——对修正样本结构偏差的尝试》,《经济研究》2011 年第 4 期。  
4. 罗楚亮、陈国强:《富豪榜与居民财产不平等估算修正》,《经济学(季刊)》2021 年第 1 期。  
5. 欧阳艳艳、张光南:《基础设施供给与效率对“中国制造”的影响研究》,《管理世界》2016 年第 8 期。  
6. 袁歌骋、潘敏:《家庭部门杠杆变动对经济增长的影响分析》,《财贸经济》2021 年第 2 期。  
7. Andrews, D. W., & Stock, J. H., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 2005.  
8. Diaz-Bazan, T., *Measuring Inequality from Top to Bottom*. SSRN Working Paper, No.2506219, 2014.  
9. Firpo, S., Fortin, N. M., & Lemieux, T., *Unconditional Quantile Regressions*. *Econometrica*, Vol.77, No.3, 2009, pp.953-973.  
10. Lambert, P. J., & Aronson, J. R., *Inequality Decomposition Analysis and the Gini Coefficient Revisited*. *The Economic Journal*, Vol.103, No.420, 1993, pp.1221-1227.