

高净值人群缺失与居民财产差距

万海远 郁杨成

内容提要：高净值人群缺失和财产低报是全球各国面临的共同挑战。本文基于2018年中国家庭收入调查数据，使用新近发展的幂律分布技术校正了高净值人群缺失问题，分析由此带来的财产漏损和财产差距低估，并比较了过去约20年的变化趋势。结果发现，2018年我国居民财产存量的漏统计率较高，财产差距的基尼系数低估了7.43%，且低估比例随时间不断上升。考虑到财产漏损主要集中在高净值人群，财产差距迅速扩大的问题需要引起关注。

关键词：高净值人群 财产低估 财产差距 幂律分布

中图分类号：F222 **文献标识码：**A **文章编号：**1003-2878(2022)02-0044-17

DOI:10.19477/j.cnki.11-1077/f.2022.02.005

一、引言

近十年来，全球范围内爆发了大规模的群体抗争，其根源是长期存在的分配失衡，特别是资本强势回归、财产差距长期处于高位、社会流动性下降带来的群体分割问题（Bhattacharya 等，2015）。在全球化和技术进步的共同作用下，高净值人群的财产存量及其增速都明显上升，不过传统调查方式难以捕捉到财产最高1%的高净值人群，因此难以解释欧美国家中存在的“反对最高1%”现象，也难以解释资本强势回归后的财富分配固化问题。但在新的数据来源和估计技术下，就可以修补缺失的高净值人群，并反映各国财产分布格局的真实变动，从而解释财产分配背后的全球性经济社会问题。

在研究财产不平等问题时所面临的最大挑战，是住户调查数据普遍存在的高净值人群缺失与财产水平低报，由此造成财产分布的右尾端缺失，并导致国民财富漏统计和财产差距低估。一般认为，这种现象在发展中国家和转型国家尤其严重，特别是由于税务系统不够透明、政策监管力量薄弱导致的不合理或不合法财产的低估尤其明显（Meyer 等，2015）。基于历史间的比较发现，近年来因高净值群体缺失和财产低报带来的不平等低估问题正变得愈发严重，甚至会扭曲人们对真实财产差距变化趋势的认识。因此，在住户调查中如何校正高净值样本的缺失和财产低报，是财产分配领域的研究前沿，也是各国政策制定中的关键命题（Atkinson 等，2011）。

作者简介：万海远，北京师范大学经济与工商管理学院、北京师范大学中国收入分配研究院教授，电子邮箱：why842000@163.com。

郁杨成（通讯作者），上海财经大学经济学院博士研究生，电子邮箱：yyc961220@sina.com。

作者感谢匿名审稿专家所提宝贵建议，当然文责自负。

中国语境下的部分高净值人群与灰色或腐败财产相关,也与资产价格泡沫的明显上涨有一定关系,当然实体经济发展也塑造了部分高净值群体。人们普遍认为中国高净值人群的比重在以显著速度增加,而且越来越少地进入住户调查的范围,由此国民财富漏统计问题也越来越明显。近年来部分高净值财产的畸形积累,也带来了资本无序扩张、财富非正常外流等现象。目前关于高净值人群的财产分配研究存在一些不足,对于财产低估变化趋势的研究更是少见,这限制了我们对真实国民财富及其分配变化的认识。本文基于2018年全国住户调查数据,使用新近拓展的幂律方法,来校正财产分布右尾端的缺失问题,分析由此造成的居民财产漏统计及财产差距低估,并比较过去约20年的变化趋势。

二、文献评述

由于高净值群体数量有限,且具有较高的财产瞒报率,因此住户调查通常难以反映高净值群体财产的真实情况(Meyer等,2015)。当然,即便是国家层面调动上百万人员的大样本调查,也不可避免会遗漏部分高净值人群(王有捐,2010)。全球财产分配领域的学者在探讨如何补充缺失的高净值样本时,微观视角大致采用两类方法,即获得更高质量的住户数据和使用更先进的估计技术。实践中高质量数据是前提,而更好的技术会改善估计质量(Meyer等,2015)。

(一) 收集高质量数据

1. 住户调查中增加高净值样本。一是在分层抽样时,有针对性地增加高净值样本比例。中国家庭金融调查(CHFS)用非农人口比例排序来抽取居委会,在富裕群体中更多抽取富裕部分,人为提高了城镇高净值人群的抽样概率(李实和万海远,2013)。二是样本轮换时保证高净值人群。美国消费金融调查(SCF)严格执行高净值人群的轮换规则,使得中低财产组的样本流失率为5.2%,而最高财产组的样本流失率也仅为5.4%(Kennickell,2017)。三是在入户过程中留住高净值样本。王小鲁(2010)提出不能使用分层随机抽样法,而应借鉴社会学调查方法,由各地人员对亲属、朋友、同事、邻里进行滚雪球式调查,从而获得更多高净值样本。四是在访谈中合理引导高净值住户准确报告财产。住户调查一般要求访谈员客观记录住户信息,而不能给予明显诱导,但实践中若发现住户明显低报,访谈员能否及如何把握修正尺度仍有讨论空间(Korinek等,2006)。

2. 使用行政数据补充高净值样本。一是个人所得税数据。研究者可以将税务部门掌握的高收入纳税数据补充到住户调查中,但收入流量比财产存量更难核算,故一般是从财产来推测收入而不是相反(Atkinson等,2011)。二是遗产税数据。Piketty(2014)估算了20多个开征遗产税国家的财产差距,然而中国并没有遗产税制,通过遗产数据倒推的方法并不现实。三是资本性收入数据。Saez和Zucman(2016)认为高净值群体的重要财富来源是资本性收入,考虑到资本性收入的获取相对容易,不同人群的资本回报率也相对稳定,由此就能倒推财产水平。四是政府行政部门数据。由于政府部门的数据真实性和权威性较高,通过跨部门的信息联动平台,个人财产信息能够在行政部门间实现共享,使得高净值群体的财产估计变得可行(Meyer和Mittag,2019;万海远,2020)。

3. 专门收集高净值样本。一是媒体公布的富人榜。李实和罗楚亮(2011)提出利用富人榜信息对高净值样本加以修正,罗楚亮和陈国强(2021)根据富豪榜信息估计高净值群体的帕累托分布参数。然而,媒体公布的数据难以接受学术标准的审视,Li等(2020)就认为胡润榜和福布斯榜数据存在信息统计交叉、

个体与家庭数据重叠、收入与财产信息混淆等问题。二是近年来有学者建立了规范的高净值人群数据库，从行业代表性、领域重叠性和数据独立性角度提出了更高的要求。Li 等（2020）围绕高净值人群构建了多个子数据库，包括上市公司高管、私营企业主、金融从业人员、作家富豪、网络红人等。

4. 用行为反应数据来修正。一是避税行为数据。白重恩等（2015）基于家庭成本收益最优化推导了住户瞒报的反应函数，由此估计出各个分组的真实财产。然而，《中华人民共和国统计法》中明确规定住户调查是为了统计需要，微观家庭信息不会与税务部门有直接关联，因而从避税视角来解释居民瞒报的合理性仍然值得探讨。二是慈善捐赠数据。一些行为经济学研究从慈善捐赠数据倒推个体财产，如 Davies 等（2010）收集了公开的捐赠信息和捐赠者的财产，通过估计每一组的捐赠倾向来预测财产水平。三是消费行为数据。王小鲁（2010）提出财产水平越高，消费的占比就越低，故可以通过边际消费倾向倒推财产水平。但消费和财产可能没有稳定的线性关系，自雇佣群体外的工资性收入及消费本身也可能存在瞒报，且不同消费支出可能存在不同的瞒报原因，所以用消费来推断高净值群体的财产也可能存在问题（王有捐，2010）。

（二）使用新技术手段

1. 分组估计技术。第一，金融部门会统计高净值人群不同财产分组的区间人数、平均财产水平、平均财产份额等。程永宏（2006）建立了分组混合基尼系数的算法，不依赖于两组分布不重叠的假定，并提出了度量总体差距的新指标。第二，Chotikapanich 等（2007）基于分组数据，放宽组内收入不变的假定，估计了东亚八个国家或地区 1993 年和 1998 年的差距低估情况。第三，Hajargasht 等（2012）基于更加细致的分组人口份额和平均收入，推导出通用矩条件和最优权重矩阵，从而以相对简单的形式表达附加高净值群体的目标函数。第四，有些国家的税务部门也会公布高净值样本分组数据，如 Piketty 和 Saez（2003）基于分段税收数据和住户数据，估计了高净值人群缺失带来的相对差距低估。总的来看，分组估计技术简单直接，但由于假定每组内部财产水平是一致的，一般会低估总体财产差距，因此可以视为真实财产差距的下限（Cowell，2011）。

2. 多元参数估计技术。第一，由于财产分布在不同分位上形状不同，使用对称的 Gaussian 曲线会扭曲中低财产样本的分布，使得高净值人群的分布过度拖尾（Benhabib 和 Bisin，2018）。帕累托分布对高净值人群的拟合效果更好，但在中低财产端的描述效果较差；Gamma 分布可以很好地模拟财产分布两端，但会夸大中间群体的扭曲程度。综合比较后，Arnold（2015）认为这些两参数模型都存在改善空间。第二，Dastrup 等（2007）提出应引入更多参数以描绘现实中的财产分布，如基于 Gamma 函数的 beta 技术、基于 lognormal 和 Weibull 的 GG 技术、基于 Singh-Maddala 和 Dagum 的技术等。第三，为适应现实中的财产分布，还可以进一步叠加更复杂的模型，然而 Schluter（2018）认为模型的复杂程度并不意味着估计的准确性，更有意义的方向是完善现有的估计方法。

3. 幂律分布技术。利用幂律分布技术估计高净值群体的财产分布具有方法学基础，同时也被学者广泛接受，然而在具体使用过程中存在一定差异。第一，已有文献大多直接假定幂律分布，很少严格检验数据集是否真正满足幂律分布、从何处开始满足幂律分布等问题。第二，以往关于幂律分布技术的应用研究从估计规模参数出发，直接对富豪榜数据使用 OLS 方法估计幂律分布的规模参数。然而 Clauset 等（2009）证明这些基于 OLS 的方法并不稳定，参数估计对分布尾部的波动比较敏感，可能会产生系统性误差。第三，传统方法通过富豪榜数据拟合顶端高净值人群的帕累托分布，并从右向左延伸至税法规定的自我申报门槛值（Diaz-Bazan，2015）或住户最高值（罗楚亮和陈国强，

2021)。Charpentier 和 Flachaire (2019) 认为从最高的富豪榜数据延伸至很低的门槛点, 可能导致样本分布两端过高, 因而总体基尼系数显著上升。Jenkins (2017) 改进了调查样本与模拟样本的混合拼接技术, 找到客观上最优的幂律分布门槛值, 在方法论上具有明显改进。

(三) 本文主要特色

本文利用跨期约 20 年的可比数据, 使用新近拓展的幂律分布技术, 分析了高净值人群缺失对财产分配差距的影响, 并比较了其长期变化趋势, 主要特色如下: 第一, 过去纠正高净值样本的研究普遍依赖于幂律分布的前提假定, 本文严格检验了多种备择分布的拟合优度, 避免对幂律分布的武断假设。第二, 在确定高净值样本分布时, 已有研究通常将频率分布的柱状图绘制在双对数轴上, 使用普通最小二乘法估计规模参数并通过图像判断门槛值, 往往会产生系统性误差。本文使用最大似然估计法并结合 KS 统计量, 能够得到渐进无偏的规模参数和更为客观的幂律分布门槛值 (Clauset 等, 2009)。第三, 传统方法主观采信媒体公布的富豪榜最大值, 从右向左拟合到某一门槛值, 而本文则先估计出幂律分布的门槛值, 再从向左最右端的高净值方向延伸。相比于过去拼接富豪数据与住户数据的方法, 本文幂律分布技术完全基于住户数据, 处理了极端值带来的分布影响, 避免了多源数据对财产定义不同造成的估计误差。

三、数据说明和变量定义

(一) 数据来源

本文数据来源于中国住户收入调查 (CHIP), 它由北京师范大学与国家统计局合作收集。该调查采用随机抽样方法收集了个体、家庭、社区 (村庄) 层面的微观数据, 包含农村居民、城市居民和农村进入城市的暂住户调查, 在不同年份覆盖全国 15—22 个省、直辖市和自治区, 包含了人口学特征、劳动就业、家庭收支和财产等信息。CHIP 团队负责问卷设计, 国家统计局负责调查实施, 其他数据清洗及权重指标等由 CHIP 团队完成。历年 CHIP 数据采用相同的调查方法、抽样程序和指标定义, 不同年份数据具有很好的可比性, 样本具有较好的随机性和全国代表性, 所以能用来分析全国层面的财产分布情况 (Sicular 等, 2020)。

CHIP 数据的优势之一是具有详细的收支和财产信息, 由此成为国内外收入分配等领域的重要数据来源。由于住户财产的主要构成是房产, 而中国住房市场化改革从 1998 年才开始, 在此之前的单位福利分房难以体现在市场价值中, 因此本文只使用 1998 年之后的数据。此外, 2007 年数据缺乏住房质量指标和农村居民房产的详细信息, 综合考虑后本文基于 2002 年、2013 年和 2018 年数据研究过去约 20 年间财产分配的动态变化, 在剔除极少数个体指标信息都缺失的观测值后, 各年对应的样本量分别为 63157 个、62101 个和 69411 个。^①

(二) 变量定义

本文使用家庭人均净财产的概念, 它是由净房产、金融资产、生产性固定资产、耐用消费品、其他资产、非住房负债和土地资产构成的。本文居民财产中不包括人力资本、社会资本及各种形式的自然资源等, 因此使用的是狭义的家庭财产定义, 并混合使用财产、资产、财富和资本等概念。

^① 与 Wan 和 Knight (2022) 不同的是, 本文保留了所有农民工样本。

对于净房产，它是在家庭自有住房评估价值的基础上进行核算的：若是与单位共有产权，则其房产价值计算为住房价值的 50%；若是个别住户的房屋产权属于私有，回答了房屋面积但未回答现住房的房产价值，则使用其余样本中的房产特征对房产价值作回归，^①代入回归系数后乘以该家庭的房屋面积得到其未回答的现住房价值；若是租房、由单位免费提供住宿或住在亲戚家的，则把住房价值归为 0。农村虽然没有正式的房产市场，但问卷中直接询问了家庭所有住房的私下交易价值或建房成本等。由于 2018 年数据样本中未回答其他住房价值，但回答了其他住房的租金收益和持有成本，我们把 2013 年的其他住房价值作为因变量，把租金收益和住房持有成本作为解释变量，并把城乡、地区、离城市中心距离、区县平均租售比、周边配套公共服务和家庭特征作为控制变量进行回归，代入 2013 年的回归系数至 2018 年，就可以插补得到 2018 年家庭的其他住房价值。^②最后在所有房产价值基础上，房产的长期贷款按其现期贷款的数额来确定，再扣除建房买房的贷款或借款，就得到家庭房产净价值。

金融资产包括银行存款、有价证券、股票、借出款、借入款和手存现金，而且把中国特色的住房公积金作为居民金融资产的一部分。近年来拆迁补偿的比例日益提高，一般被认为是土地资产的转化形式，实践中若转化成一笔现金补偿则归为金融资产，若转为住房补偿则直接归类到住房资产。耐用消费品包括汽车、摩托车、家用电器等；生产性固定资产包括机器设备、农用机械（如拖拉机等）、器具工具、生产性役畜等。其他资产是指以货币为计量单位的各种财产、资源、债权和其他权利，主要包括未收回的欠款及收藏品、古董和字画等。非住房负债包括用于教育、家庭成员治病、家庭日常开支和其他方面的借款或贷款。土地价值是先对土地面积按一亩水浇地等于两亩旱地进行调整，并假定农业经营毛收入有 25% 是由土地带来的且土地的收益率为 8%，从而倒推得到农村家庭的土地价值（McKinley, 1993）；过去城镇居民土地价值被设定为零，但 2012 年全国住户调查一体化之后，在城镇工作的农民工也被定义为城镇住户并包含在城镇样本中，从而城镇住户实际上也可能拥有农村土地财产。在中国特色养老保险制度下，居民养老金一般作为家庭收入的组成部分，过去作为工资性收入的一部分，但 2012 年国家统计局城乡一体化调查之后又划分为居民转移性收入，在进行国际比较时也可以单独作为一项收入类型（Sicular 等，2020）；但不管以何种角度看，中国居民财产类型都不应包括居民养老金，毕竟现实中无法把养老金作为市场中的金融投资产品，这一点与国外研究有一定区别。

在上述家庭净财产定义的基础上，除以家庭常住人口规模，就得到家庭人均净财产。进一步我们还对数据进行了以下调整。首先，国家统计局于 2012 年实施了全国住户一体化调查，调整了城乡居民定义，以常住地为标准把常年工作在城镇的农村户口居民也纳入城镇调查，我们使用这个标准调整了 2002 年数据的城乡定义方式。其次，为保持不同时期的可比性，采用国家统计局公布的

① 房产特征包括住房地理区位、住房配套设施、外部条件、内部结构和内部环境等方面，也包括教育、医疗、超市、交通等公共配套服务，如社区是否通公路、饮用水是否经过集中净化处理、是否集中供暖、垃圾是否集中处理、社区是否有绿化景观、社区是否有卫生站、社区幼儿园或学前班是否便利、周边上小学的便利程度、社区是否有专职安保人员等。

② 由于 2018 年数据仅包括家庭自有现住房的价值，而没有包括家庭所持有的其他房产价值，这与 Knight 等（2021）略有差异。这里也可以使用家庭所有其他房产的租金收入，除以当地的平均租售比推断得到其他房产的市场价值。使用这种方法估计出的财产差距基尼系数总体上是一致的。

CPI 指数对不同年份财产分项统一调整为 2018 年数据。^①再次,所有指标均使用地区与城乡层面的权重,从而加权后就可以得到具有全国代表性的数据(Sicular 等,2020)。最后,考虑到财产分布的个别畸高极端值可能会导致财产分配结果的扰动,所以我们慎重对待畸高值问题。美国的人口调查数据 CPS 在经过严格稳健性检验后建议做一些缩尾处理。^②我们检验了不同缩尾处理下幂律分布技术的处理结果,发现 2002 年、2013 年和 2018 年在分别缩尾 8 个、124 个和 138 个财产值最高的样本后,各个分布参数和财产分配基尼系数都趋于稳定,实践中与 Cowell 和 Flachaire (2007) 的估计结果也很接近,说明选择这个缩尾比例是比较稳健的。^③

四、幂律分布理论与应用

(一) 高净值样本拟合技术

在现实世界中,当居民财产超过特定门槛值后往往服从典型的幂律分布。为了数学上方便,研究中通常把居民财产视为连续态分布,用 x 代表财产,那么幂律分布的概率密度函数为:

$$p(x)dx = \Pr(x \leq X < x + dx) = Cx^{-\alpha} dx \quad (1)$$

其中 X 是观测值, C 是标准化常数。当 $x \rightarrow 0$ 时,概率密度函数是发散的,因此(1)式不适用于所有 $x \geq 0$,必须规定某个下限 x_{\min} 。假设 $\alpha > 1$,进行标准化后得到:

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha} \quad (2)$$

进一步计算幂律分布的互补累积分布函数为:

$$P(x) = \int_x^{\infty} p(x')dx' = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1} \quad (3)$$

接下来估计幂律分布的两个关键参数,包括门槛值 x_{\min} 和规模参数 α 。对于规模参数,最大似然法可以给出渐进无偏的估计,根据(3)式计算对数似然函数为:

$$\begin{aligned} L &= \ln p(x|\alpha) \\ &= \ln \prod_{i=1}^n \frac{\alpha - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}}\right)^{-\alpha} \\ &= \sum_{i=1}^n [\ln(\alpha - 1) - \ln x_{\min} - \alpha \ln \frac{x_i}{x_{\min}}] \\ &= n \ln(\alpha - 1) - n \ln x_{\min} - \alpha \sum_{i=1}^n \ln \frac{x_i}{x_{\min}} \end{aligned} \quad (4)$$

① 我们把所有年份数据都折现到 2018 年的价格,而 Knight 等(2021)把财产指标都调整为 2013 年。对于能否再用购买力平价指数(PPP)来调整财产存在争议。有观点认为居民财产大部分以房产形式存在,PPP 本身没有反映房价的变化,因而用 PPP 平减财产并不合适;也有研究考虑到居民财产流动性在增强,故利用 PPP 来对各项财产调整是必要的。Knight 等(2021)认为是否调整的差异较小,故本文不作调整。

② 可参见 Autor 等(2008)关于 top-coding 的讨论。

③ 本文也检验了对财产左侧缩尾的结果,发现左侧缩尾对总体基尼系数几乎没有任何影响。

设定 $\partial L/\partial \alpha=0$ ，得到规模参数 α 的解析解：

$$\hat{\alpha}=1+n\left[\sum_{i=1}^n\left(\ln\frac{x_i}{x_{\min}}\right)\right]^{-1} \quad (5)$$

下面先求解幂律分布的门槛值，再利用（5）式得到规模参数。具体步骤是：①任意给定门槛值 x_{\min} ，结合实际数据计算互补累积分布函数；②根据式（5）得到对应的规模参数 α ；③结合给定的门槛值和对应的规模参数，计算幂律分布的互补累积分布函数；④量化估计实际数据与幂律分布之间的“距离”；⑤求解使得两者最大距离最小化的门槛值，即为最终的门槛值 \hat{x}_{\min} ；⑥将 \hat{x}_{\min} 代入（5）式得到最终的规模参数 $\hat{\alpha}$ 。

值得说明的是，第4步量化两个概率分布“距离”的方法有很多，对于非正态数据最常见的方法是使用 Kolmogorov-Smirnov（KS）统计量，先计算实际数据互补累积分布函数 $(S(x))$ 与幂律分布互补累积分布函数 $(P(x))$ 之间的最大“距离”，再求解使最大“距离”最小化的门槛值。^①具体为：

$$\min D = \min \left\{ \max_{x \geq x_{\min}} |S(x) - P(x)| \right\} \quad (6)$$

（二）混合财产差距估计技术

在幂律分布中，一旦得到了门槛值和规模参数，那么幂律分布函数就唯一确定。通过幂律分布函数插补高净值人群后，就可以使用调查数据和模拟数据的混合样本来估计总体财产差距，^②计算公式为（Diaz-Bazan, 2015；艾小青和祁磊，2021）：

$$G = p_1^2 \times \frac{u_1}{u} \times G_1 + p_2^2 \times \frac{u_2}{u} \times G_2 + p_1 \times p_2 \times \left| \frac{u_2 - u_1}{u} \right| \quad (7)$$

其中， p_1 和 p_2 分别表示调查样本和高净值样本的人口份额； u_1 和 u_2 分别为两部分群体的平均财产； u 为两个群体的加权平均财产； G_1 和 G_2 分别表示两个群体的财产基尼系数，其中服从幂律分布的高净值人群的基尼系数 $G_2 = \frac{1}{2\alpha-3}$ （Alvaredo, 2011）。

财产大于门槛值以上任意 x_0 的群体人数 $num_{x \geq x_0}$ 可以通过互补累积分布函数求得：

$$\begin{aligned} num_{x \geq x_0} &= \frac{\Pr(x \geq x_0)}{\Pr(CHIP_{\max} \geq x \geq x_{\min})} \times num_{CHIP_{\max} \geq x \geq x_{\min}} \\ &= \frac{P(x_0)}{1 - P(CHIP_{\max})} \times num_{CHIP_{\max} \geq x \geq x_{\min}} \end{aligned} \quad (8)$$

其中， $CHIP_{\max}$ 代表 CHIP 数据中的最高财产值， $num_{CHIP_{\max} \geq x \geq x_{\min}}$ 代表 CHIP 数据中财产值在门槛值与最大值之间的人数。

此外，财产大于 x_0 群体的平均财产，可以通过以下积分函数获得：

$$u_2 = \int_{x_0}^{\infty} xp(x)dx = \frac{\alpha-1}{\alpha-2}x_0 \quad (9)$$

① 除 KS 统计量外，也存在其他统计量，如使得两者最小距离最大化的 Kuiper 统计量或 Anderson-Darling 统计量，Cowell（2011）认为他们的标准误相对较大，尤其是在厚尾分布且样本量较小时，这种估计误差会显著增加。

② 这里不能用幂律函数直接生成高净值样本的微观个体数据，原因在于幂律函数究竟要覆盖多少规模个体还不确定，这取决于全国总人口和住户调查所代表的总人口差异，因此这里只能结合两段样本来共同推算总体财产差距。

五、幂律参数估计与检验

（一）门槛值与规模参数估计

根据幂律模型，图 1 描述了规模参数与门槛值的对应关系，发现当门槛值上升时，规模参数估计值总体是递增的，但局部也可能下降。观察式（5）发现，门槛值上升会对规模参数产生正反两方面的影响：一方面高于门槛值的样本量减小，会降低规模参数的估计值；另一方面求和项中的每一子项都减小，会导致规模参数估计值上升。后者在多数情况下起主导作用，因此规模参数估计值总体呈上升趋势。

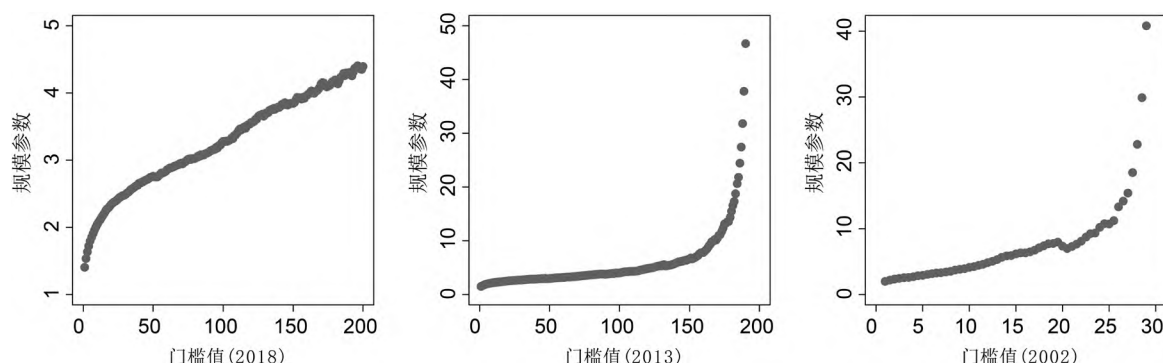


图 1 门槛值与规模参数的对应关系

注：2018 年和 2013 年数据的横坐标间隔 1 万元绘制散点图，2002 年间隔 0.5 万元绘制散点图。

图 1 中究竟哪个门槛值是最优的？这需要比较不同的门槛值，先通过最大似然估计法确定对应的幂律分布，然后基于（6）式计算使得实际数据与幂律分布最大“距离”最小化的门槛值。图 2 显示了最大“距离”随门槛值的变化情况。可以发现 2018 年使最大“距离”最小化的门槛值是 37 万元，而 2013 年和 2002 年分别是 33 万元和 14 万元。换句话说，2002 至 2018 年间的门槛值都高于 12 万元，说明传统方法主观选择 12 万元作为门槛值带来的偏差会越来越明显。进一步从（5）式计算得到规模参数，其在 2018 年、2013 年和 2002 年分别为 2.6066、2.7546 和 5.8316。值得注意的是，幂律分布技术依赖于数据对高净值样本的覆盖程度，这里 2013 年和 2018 年的规模参数在 2 至 3 之间，而 2002 年则超过 5，可能是由于早期的高净值人群规模较低，财产分布的尾部更薄，因此规模参数也更高。

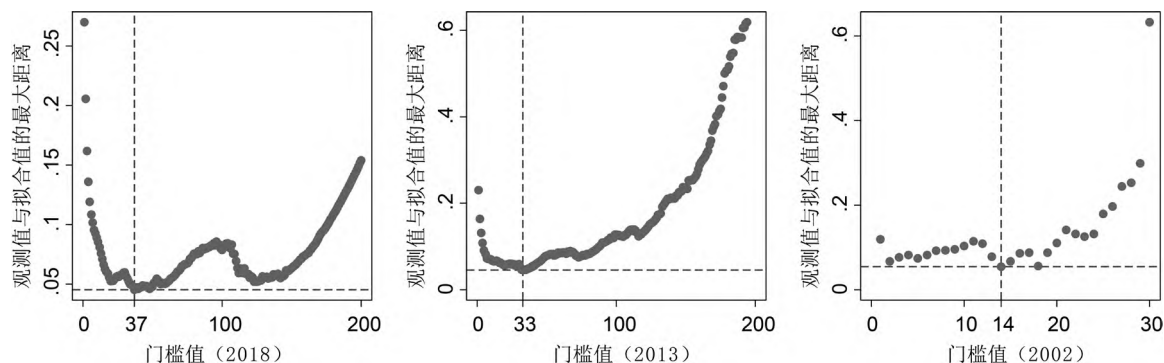


图 2 观测值与拟合值的最大距离变化

注：根据 CHIP 住户调查数据计算，横坐标单位是万元。

前文成功估计了幂律分布的两个主要参数，结果是否准确、与观测数据匹配是否良好，可以先通过图像直观地观察，再进行统计检验。图 3 显示了幂律分布与实际数据的拟合效果，散点是 CHIP 数据的观测值，曲线是用于拟合的幂律分布。从图形上看，发现幂律模型在直观上是吻合的。^①我们还将各年财产的互补累计分布函数绘制在双对数图像上，发现使用幂律分布技术得到的规模参数估计值在 0.01% 的水平下仍然统计显著。

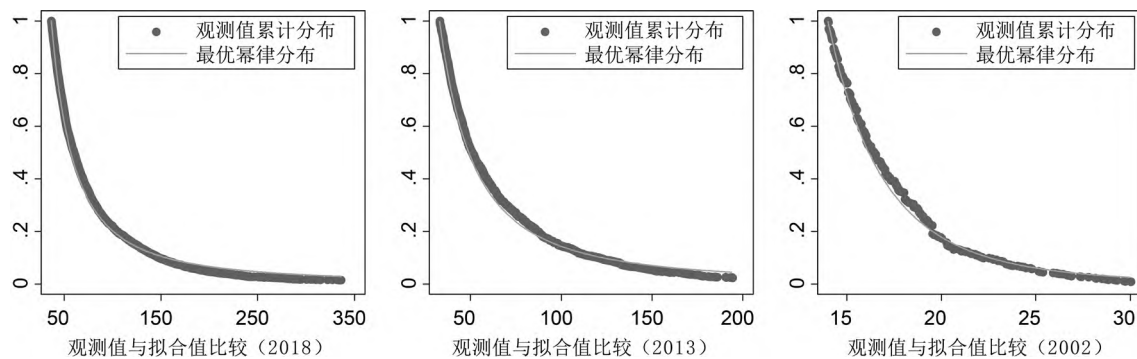


图 3 观测值和拟合值比较

注：根据 CHIP 住户调查数据计算，横坐标单位是万元。

（二）参数估计方法讨论

前文直接使用最大似然估计法（MLE）估算了幂律分布的规模参数，下面将基于 MLE 的估计值与基于最小二乘回归（OLS）的估计方案比较，以检验本文结果的稳健性。

1. 固定门槛值。基于 OLS 方法的思路是，首先将观测数据按区间分组，在双对数图上绘制频率分布直方图；其次对式（3）两边取对数，当分布函数在双对数图上接近于直线时则认为数据遵循幂律分布，并将 OLS 回归得到的直线斜率视为规模参数。在实际使用 OLS 方案时，存在 4 种不同的处理方式，其不同点有两处（见表 1）。一是绘制直方图所基于的函数不同，方案 2 和方案 4 基于概率密度函数，而方案 3 和方案 5 基于累积分布函数；二是区间划分不同，方案 2 和方案 3 将财产划分为等宽区间，方案 4 的区间宽度由高到低呈指数下降，而方案 5 按照财产高低排序，计算每个财产值出现的频率。

表 1 不同估计技术的规模参数比较

	方案名称	参数备注	规模参数 $\hat{\alpha}_{2018}$	规模参数 $\hat{\alpha}_{2013}$	规模参数 $\hat{\alpha}_{2002}$
1	MLE+KS	最大似然	2.6066	2.7546	5.8316
2	OLS+PDF	等宽区间	2.4787	2.6057	5.2733
3	OLS+CDF	等宽区间	2.8027	2.9477	6.4072
4	OLS+PDF	对数区间	0.1179	-0.0394	1.3598
5	OLS+CDF	秩频	2.7949	2.9419	6.3297

注：根据 CHIP 住户调查数据计算。2018 年和 2013 年区间宽度为 1 万元，2002 年为 0.5 万元。

^① 我们使用 CHIP 数据本身，绘制了数据在幂律分布门槛值两端的分布形态，左侧大致是对数正态，右侧则不太可能是对数正态。限于篇幅，未在正文中列示，留存备案。

表 1 显示了 5 种不同方案下的规模参数估计值。与 MLE 估计值相比, OLS 方案都有所偏差。方案 2、方案 3 和方案 5 下的估计值偏差较小, 其中方案 2 的估计值低于 MLE 估计值, 而方案 3 和方案 5 的各年结果都显示出高估的共性特点。方案 4 下的估计值偏差较大, 甚至出现了负的结果。Clauset 等 (2009) 在理论和实践中证明了, MLE 方法在幂律模型中可以提供渐近无偏的规模参数估计, 虽然在有限样本下也会存在偏差, 然而这种偏差在大多数情况下可以忽略, 至少比 OLS 方法的误差小得多。

2. 变化门槛值。为进一步检验估计方案的稳健性, 下面讨论不同门槛值下 5 种方案规模参数的变化情况。图 4 发现, 基于累积分布函数的 OLS 方案 (方案 3 和方案 5) 的规模参数, 在任何门槛值下都几乎相等。事实上, 方案 5 可以视为方案 3 的极限情况, 当方案 3 的区间划分越来越细时, 估计的规模参数也会越来越接近方案 5。在 KS 统计量确定的门槛值附近, 方案 2、方案 3 和方案 5 的规模参数与 MLE 结果相对接近, 而方案 4 的估计值明显偏低。

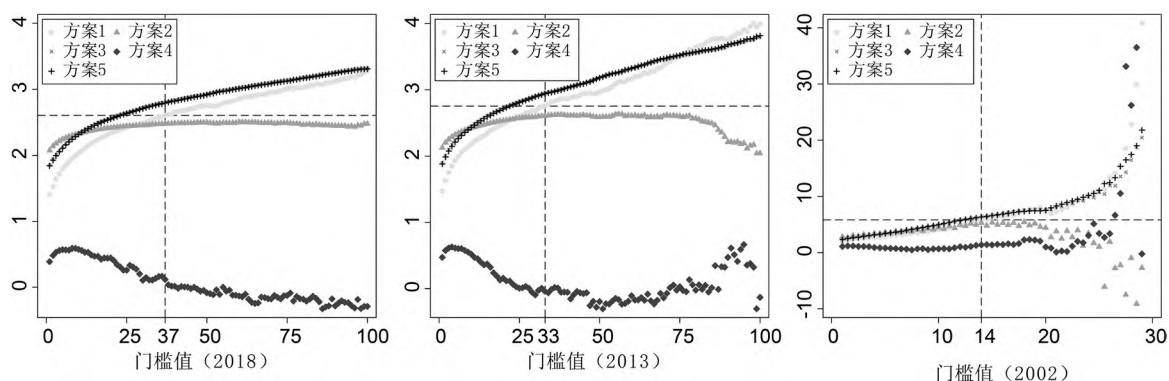


图 4 不同估计技术下的参数估算结果比较

注: 根据 CHIP 住户调查数据计算, 横坐标单位是万元。

从图 4 的结果来看, 门槛值的选择对于规模参数估计至关重要。当选定的门槛值略高于实际门槛值时, 虽然幂律模型覆盖的有效样本量会减小, 但规模参数的最大似然估计值仍接近于准确值。然而, 当选定的门槛值低于实际门槛值时, 规模参数的估计值会因观测数据偏离幂律分布而迅速产生偏差。综合来看, 传统 OLS 方法无论是否建立在精确的门槛值基础上, 都可能造成规模参数的估计偏差, 从而使财产差距的估计出现偏差 (Clauset 等, 2009)。

(三) 幂律分布统计检验

前文讨论了如何估计幂律分布的参数, 但前提问题是不同年份的 CHIP 数据是否以及从哪里开始符合幂律分布? 对于任何给定的数据集, 即便它不是幂律分布, 使用前面的方法也可以得到一个相对接近的幂律分布参数。^①特别是由于高净值观测值非常少, 不同函数形式都能在一定程度上吻合实际数据, 因此任何观察法或前置假定法, 都无法排除其他分布函数的存在 (Benhabib 和 Bisin, 2018)。即使从

^① 事实上, 无论数据是从哪个真实分布中产生的, 总是可以找到最优的幂律分布参数, 因此需要一种方案来检验数据是否符合幂律分布。

幂律分布产生的人工数据集，由于抽样的随机性，生成的一小部分数据也可能不遵循幂律分布。目前关于高净值人群的研究中，普遍使用了幂律分布的前置假定，由此带来前提条件被否定的风险，因此确认数据集是否符合幂律分布的严格统计检验是非常必要的。

下面先假设观测数据是从前面确定的幂律分布中产生的，然后通过构造一个 p 值来验证假设是否可信。具体来说：①先使用前述方案确定的幂律模型拟合观测数据，记录 KS 方法得到的统计量 D ；②根据确定的幂律分布，生成与观测数据样本量相同的人工数据集；③重复上述步骤②，将人工数据集拟合到幂律分布中，并计算每个数据集的统计量 D_i ；④重复生成 2500 次人工数据集，计算 D_i 大于 D 的概率，将其记为 p 。需要说明的是， p 值代表“人工数据集与幂律模型的距离”大于“观测数据与幂律模型的距离”的概率。如果 p 值较大，则观测数据与幂律模型的差异仅归结于统计偏差；如果 p 值很小，则幂律模型不是实际观测数据的合理拟合（Clauset 等，2009）。

表 2 幂律分布的优度检验

数据集	CHIP 样本量	幂律分布 门槛值	服从幂律 分布的 总样本量	缺失的 高净值 样本量	小于幂律分布 门槛的 样本量	模拟 次数	幂律 分布概率	指数 分布概率	广延 分布概率	截止 分布概率
	n	x_{\min}	n_{tail}	n_{top}	n_{low}	t	p	p	p	p
2018	69411	37	10204.27	294.74	59501.47	2500	0.87	0.01	0.04	0.00
2013	62101	33	6533.98	291.14	55858.16	2500	0.90	0.00	0.05	0.01
2002	63157	14	1496.08	37.38	61698.30	2500	0.57	0.08	0.02	0.01

注：根据 CHIP 住户调查数据和模拟数据计算，其中 p 值在不同随机状态下也会不同。

表 2 第 2—3 列显示了 CHIP 数据的样本量和幂律门槛值，第 4—6 列是推算的服从幂律分布的总样本量（包括缺失的高净值样本）、缺失的高净值样本量、小于幂律分布门槛值的样本量。第 7—8 列是模拟次数和幂律拟合优度检验的 p 值。一般认为，若 $p \leq 0.1$ 则幂律分布假说就被排除。^③换句话说，如果从幂律分布中生成的人工数据集与幂律分布之间的“距离”大于观测数据与幂律分布之间的“距离”的概率小于 10%，则观测数据不服从幂律分布。从表 2 的结果来看，2018 年、2013 年和 2002 年门槛值以上的样本符合幂律分布，从而验证了以往研究假定“高净值群体服从幂率分布”的合理性。进一步，第 9—11 列用同样的方法检验了指数分布、广延分布和截止分布的拟合优度，发现 p 值都很低，说明门槛值以上的数据不符合这些分布形式。

① 假设观测数据总共有 n_{tail} 个观测值满足 $x \geq x_{\min}$ 和 n_{low} 个数据满足 $x < x_{\min}$ 。首先根据幂律模型生成 n_{tail} 个随机数，然后与实际调查数据中 $x < x_{\min}$ 的 n_{low} 个要素拼接，生成一个样本量为 $n_{\text{low}} + n_{\text{tail}}$ 的完整数据集，它和观测数据具有相同样本量，在 x_{\min} 以上遵循与观测数据相同的幂律分布，并且在 x_{\min} 以下与观测数据也具有相同的非幂律分布。

② 为了将统计偏差和由于数据来自非幂律分布而产生的偏差区分开来，需要从真实的幂律分布中抽样出许多人工数据集，进而测量它们与幂律形式的偏离程度。如果观测数据比人工数据离幂律形式远得多，那么幂律就不是对观测数据的合理拟合。观测数据也可能产生于非幂律分布，但由于抽样波动导致观测数据接近幂律。在这种情况下，模拟测试将给出错误结论，但这种情况发生的概率随着模拟次数的增加而减小。

③ 当然也可使用更宽松的 $p \leq 0.05$ ，但这可能会使一些实际不遵循幂律分布的数据被误认为遵循幂律分布。

六、混合样本拼接与稳健性检验

（一）混合样本拼接

前文虽然估计了分布参数、检验确定了幂律分布形式，但混合使用调查样本和模拟样本来估计总体财产差距时，又会遇到两个样本“拼接点”选择的问题。从表 3 观察到 CHIP 数据最大值都大于幂律分布门槛值，因此方案 1 选取 CHIP 数据最大值作为拼接点，而方案 2 选取幂律分布门槛值为拼接点，拼接点以下使用调查数据，拼接点以上使用模拟数据。方案 3 同样以 CHIP 最大值为拼接点，不同点是产生模拟数据的过程是从幂律分布门槛值开始的。

表 3 实际财产差距估计

年份	CHIP 标准差	CHIP 最大值	门槛值	规模参数	财产差距 (原始)	财产差距(处理后)		
						方案 1	方案 2	方案 3
2018	33.62	336.00	37	2.6066	0.6244	0.6707	0.6610	0.6706
2013	22.10	194.33	33	2.7546	0.6022	0.6426	0.6299	0.6423
2002	3.41	30.04	14	5.8316	0.4950	0.4976	0.4963	0.4976

注：根据 CHIP 住户调查数据和模拟数据计算。第 2—3 列是样本缩尾后的标准差与最高值，财产差距用基尼系数表示，财产单位是万元。第 6 列是使用原始数据估计的财产差距。

我们先使用规模参数计算服从幂律分布样本的基尼系数，然后利用式(7)得到总体财产差距。从表 3 看，3 种方案的估计值非常接近，使用方案 1 计算的基尼系数略高于方案 2，方案 3 和方案 1 的基尼系数几乎一致。与 Han 和 Cheng(2019)类似，本文使用的幂律分布技术将缩尾后的调查数据视为“帕累托分布的截尾形式”，这意味着缩尾后的调查数据在其所覆盖到的右尾端具有良好代表性，因此估算财产差距应尽可能使用调查数据。相比于方案 1，方案 2 模拟了幂律分布门槛值到住户调查最高值之间的样本，势必产生更大的统计误差；方案 3 在方案 2 基础上进行了优化调整，结果与方案 1 非常接近，但实践中技术也更为复杂。综合考虑，方案 1 包含了调查数据的所有样本信息，数学上也更为便利，因此作为基准方案。当然是否要选择 CHIP 数据最高值作为拼接点，也取决于住户调查中高净值人群的数据质量。虽然幂律分布技术可以解决最顶端的高净值样本缺失问题，但我们也担心住户调查数据中的次顶端样本同样存在缺失的情况。通过比较宏观资产负债表数据与缩尾后的住户调查数据的家庭人均财产，我们发现两者的结果非常接近，说明缩尾后住户调查数据中的高净值人群在其所覆盖的区间内代表性良好，由此进一步说明选择方案 1 来拼接是可以接受的。

（二）幂律参数敏感性检验

从式(7)来看，高净值样本对总体财产差距的影响取决于组内差距和组间差距，但本质上都由幂律分布的两个参数决定。基于表 3 中的方案 1，控制其中一个参数不变，可以测试财产差距对幂律参数值的敏感程度。虽然利用幂律分布技术得到的两个参数估计值始终同时变化，但这种控制变量法的讨论仍有必要。

表 4 财产差距对幂律参数的敏感性检验

门槛值变动						规模参数变动					
$\hat{x}_{\min 2018}$	G_{2018}	$\hat{x}_{\min 2013}$	G_{2013}	$\hat{x}_{\min 2002}$	G_{2002}	\hat{a}_{2018}	G_{2018}	\hat{a}_{2013}	G_{2013}	\hat{a}_{2002}	G_{2002}
37	0.6707	33	0.6426	14	0.4976	3.2066	0.6285	3.3546	0.6096	6.4316	0.4966
87	0.6785	83	0.6565	64	0.4949	3.0066	0.6356	3.1546	0.6159	6.2316	0.4969
137	0.6815	133	0.6666	114	0.4949	2.8066	0.6480	2.9546	0.6260	6.0316	0.4972
187	0.6797	183	0.7450	164	0.4949	2.6066	0.6707	2.7546	0.6426	5.8316	0.4976
237	0.6847	233	0.5977	214	0.4949	2.4066	0.7137	2.5546	0.6712	5.6316	0.4982
287	0.7128	283	0.5977	264	0.4949	2.2066	0.8001	2.3546	0.7232	5.4316	0.4988
337	0.6181	333	0.5977	314	0.4949	2.0066	0.9904	2.1546	0.8283	5.2316	0.4995

注：根据 CHIP 住户调查数据和模拟数据计算。

表 4 左侧显示了门槛值变动对总体财产基尼系数的影响，使用幂律分布技术得到的门槛值和总体基尼系数以加粗显示。当 2018 年门槛值从实际值逐渐上升时，总体基尼系数变化不大。值得注意的是，由于 2002 年的最高财产值仅为 30.04 万元，当门槛值达到这一数值后，基尼系数不再随门槛值变化。表 4 右侧显示了规模参数变动对基尼系数的影响，使用幂律分布技术得到的规模参数和基尼系数以加粗显示，发现基尼系数对规模参数的变化更加敏感，因此规模参数的准确估计至关重要。

七、中国居民财产差距估计

（一）漏损的财产存量估算

前文确定了观测数据的幂律函数，将分布从门槛值向右尾端延伸，从而模拟出调查数据中未收集到的高净值样本。基于调查样本和模拟样本，我们使用表 3 方案 1 的拼接方式，计算了调整前后的财产漏统计和财产差距低估情况。表 5 显示，在补充高净值样本后 2018 年家庭人均财产水平由 21.12 万元增加到 24.45 万元，居民财产的漏损统计率达到 15.76%。^①这是比较可观的国民财富低估，但既低于王小鲁(2010)使用社会学滚雪球调查数据估计出 2008 年 29.5% 的收入漏统计率，也低于白重恩等(2015)利用统计局城镇住户调查数据 UHS 估计出 2009 年 24.3% 的收入漏损率。不过本文财产是存量累积数据，而收入是流量新增数据，因此也需要看到它们之间的明显差异。

从时间趋势上看，国民财富漏统率在持续攀升，由 2002 年的 0.64%，上升到 2013 年的 12.97%，并继续提高到 2018 年的 15.76%。综合表 5 的结果看，虽然住户调查的财产水平在增长，但没有纳入的漏统计财产数量及占比也在明显增长。在修补高净值样本后，发现我国居民财产的增长速度很快，不仅高于发达国家的平均增长速度，也高于我国居民收入的实际增长速度，显示了国民福利的加速增长态势。

^① 我们还使用幂律分布技术，计算了 CHIP2018、CHFS2017 和 CFPS2018 高净值群体的财产漏统率，发现计算结果比较接近，说明财产漏损是所有住户调查面临的共同问题。另外也要看到不同数据来源、不同数据年份、不同估计方法所带来的结果的差异性，因此难以简单横向比较。

表 5 漏损的财产存量

	2018 年		2013 年		2002 年		2018 年	2013 年	2002 年
	A	B	C	D	E	F	$G = (B - A) / A$	$H = (D - C) / C$	$I = (F - E) / E$
	原始	处理后	原始	处理后	原始	处理后	漏统率	漏统率	漏统率
财产水平	21.12	24.45	14.49	16.37	3.11	3.13	15.76%	12.97%	0.64%
财产差距	0.6244	0.6707	0.6022	0.6426	0.4950	0.4976	7.43%	6.72%	0.53%

注：根据 CHIP 住户调查数据和模拟数据计算。

（二）高净值人群缺失的财产差距影响

由于被漏损的财产主要集中在高净值人群，由此带来居民财产差距的低估也比较明显。表 5 显示住户调查中 2018 年财产差距基尼系数为 0.6244，而在修补漏统计的高净值样本后，财产差距基尼系数达到 0.6707，比调整前高出 7.43%。谢宇和靳永爱（2014）使用 2012 年家庭追踪调查数据并基于 OLS 的帕累托方法，发现修补高净值群体后财产差距基尼系数会从 0.63 增加到 0.73。Sicular 等（2020）使用 CHIP 2013 年数据并基于扩大最富裕人群权重的方法，发现高净值样本调整后财产差距基尼系数会从 0.61 增加到 0.69 左右。从过去 20 年的变化趋势来看，调整高净值群体样本会使 2002 年和 2013 年财产差距基尼系数分别提升 0.53% 和 6.72%，在 2018 年则显著提升 7.43%，由此显示高净值群体在财产差距变化中的重要作用。

前文发现高净值群体越来越难以进入住户调查，财产水平的漏统计程度明显增加，由此带来财产差距低估的问题也会更加严重。表 6 继续观察高净值群体修补前后的财产份额，发现处理后的高净值群体财产份额无论在哪一年都显著增加；而且越是在财产分布顶端，其财产份额的增加就越多，最高 0.1% 群体的财产份额增长幅度要显著大于其他高净值群体。这说明高净值人群内部存在明显的低估效应差异，因此用传统调查数据来估算总体财产差距的挑战在持续增加（Meyer 等，2015）。

表 6 高净值人群的财产份额

单位：%

	2018		2013		2002		Alvaredo 等（2017）					Alvaredo 等 （2018）	Bharti （2018）	Saez 和 Zucman （2016）
	原始	处理后	原始	处理后	原始	处理后	中国 2015	中国 2013	法国 2013	俄罗斯 2015	韩国 2013	英国 2013	印度 2012	美国 2012
0.1%	2.95	8.83	2.41	6.51	0.87	1.33	—	—	—	—	—	—	18.35	22.04
1%	13.20	25.25	12.05	22.49	6.61	7.19	29.63	27.25	22.90	42.58	24.98	16.2	30.69	41.83
5%	33.75	42.91	32.59	40.60	23.37	23.76	—	—	—	—	—	33.8	50.46	—
10%	47.76	54.97	46.68	53.02	36.95	37.19	67.41	66.56	54.85	71.32	65.67	47.3	62.77	77.24
20%	64.87	69.75	64.40	68.27	54.28	54.94	—	—	—	—	—	—	78.35	—
50%	89.51	90.95	88.23	89.63	82.60	82.61	93.56	93.39	93.59	96.51	98.20	—	94.29	98.29

注：根据 CHIP 住户调查数据和模拟数据计算。

表 6 显示，无论是否纠正高净值样本，我国居民财产差距并不太高。使用幂律分布技术后，我国财产差距水平低于 Alvaredo 等（2017）的测算结果，如本文 2013 年最高 1% 群体所占财产份额为 22.49%，低于其估计的 2013 年 27.25%，本文最高 1% 群体在 2018 年 25.25% 的财产份额也小于其 2015 年 29.63% 的结果。从国际比较看，Hundenborn 等（2019）发现南非 2014 年财产基尼系数会从 0.742

上升到0.791,远高于中国2018年调整后的0.6707。Saez和Zucman(2016)发现调整后美国2012年最高1%人群占有的社会财产份额为41.83%,而中国2018年仅为25.25%。综合来看,在全球范围内中国高净值人群所占有的财产份额及其带来的分化程度并不太高,无论是最高1%还是最高10%群体的财产份额都没有明显超出同类国家。但考虑到这种结构变化仅用了不到二十年时间,说明我国居民财产积累和财产分化的速度很快,这需要引起充分重视。

八、简要结论

高净值人群缺失和财产低报问题在全球各国普遍存在,不仅容易掩盖资本外逃和国民财富低估问题,更可能扭曲对实际财产分布情况的认识,导致经济社会政策制定的偏差。近年来我国出现对高净值人群财富管理、使用和继承问题的担忧,也在局部范围内出现资本无序流动与侵蚀蔓延等乱象,亟需对财产存量及其分配状况作出准确研判。本文使用2018年中国住户收入调查数据,结合新近拓展的幂律分布技术来修补高净值样本,讨论幂律分布在调查数据中的适用性,检验不同的估计方法和分布形式,研究其对国民财富漏统计及财产差距的影响,并比较过去约20年的变化趋势。

结果显示,高净值人群缺失带来的财产漏统计问题比较明显,2018年漏损财产占居民财产总量的15.78%。由于漏损财产集中在财产分布的顶端,修补高净值样本后,2018年的财产基尼系数达到0.6707,比原始调查数据提高了7.43%。从时间趋势来看,高净值人群缺失造成的财产差距低估还在持续上升,由此说明高净值人群在整个财产分布中愈发重要。从国际比较来看,在修补高净值样本后,居民财产的增长速度很快,还远高于发达国家的平均增速,也高于居民收入的实际增长速度,显示出国民福利的加速提升态势。

高净值人群脱离于住户调查范围之外,这在扭曲国民财富增长的同时,也会导致财产差距出现低估。与其他国家相比,当前中国的财产差距水平并不高,但财产差距迅速扩大已经成为经济社会的潜在问题,需要引起充分重视。尤其要警惕少数群体利用政策不完善获得不合理财产,少数资本在关键产业中进行无序扩张,防范部分财富的过度空转和低效传递,并关注由这些漏统计财产造成的财产差距扩大现象。建议政府部门与科研机构合作,在保障个体隐私的前提下,加大对高净值人群信息的开放力度,提升住户调查数据对高净值样本的有效覆盖,加强对综合行政大数据的分析使用,从而有利于对高净值人群财产存量及其分布的相关宏观决策。

参考文献

- [1] 艾小青,祁磊.信息不完全下收入或财富基尼系数的估算[J].数量经济技术经济研究,2021,38(6):146-165.
- [2] 白重恩,唐燕华,张琼.中国隐性收入规模估计——基于扩展消费支出模型及数据的解读[J].经济研究,2015(6):4-18.
- [3] 程永宏.二元经济中城乡混合基尼系数的计算与分解[J].经济研究,2006(1):109-120.
- [4] 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[J].经济研究,2011(4):68-79.
- [5] 李实,万海远.提高我国基尼系数估算的可信度——与<中国家庭金融调查报告>作者商榷[J].经济动态,2013(2):43-49.
- [6] 罗楚亮,陈国强.富豪榜与居民财产不平等估算修正[J].经济学(季刊),2021,21(1):201-222.
- [7] 万海远.高净值人群修补技术新进展[J].中国人口科学,2020(5):91-105.
- [8] 王小鲁.灰色收入与国民收入分配[J].比较,2010(48):53.
- [9] 王有捐.也谈居民收入的统计与调查方法[EB/OL].国家统计局网站,2010年8月24日,<http://www.stats.gov.cn/zjtj/ztfx/>

grdd/201008/t20100824_59068.html.

- [10] 谢宇, 靳永爱. 家庭财产 [A]. 谢宇, 张晓波, 李建新. 中国民生发展报告 [C]. 北京: 北京大学出版社, 2014.
- [11] Alvaredo, F. A Note on the Relationship between Top Income Shares and the Gini Coefficient. *Economics Letters*, 2011, 110 (3) : 274-281.
- [12] Alvaredo, F., Chancel, L., Piketty, T., et al.. Global Inequality Dynamics: New Findings from WID.WORLD. NBER Working Paper, No.23119, 2017.
- [13] Alvaredo, F., Atkinson, A. B., & Morelli, S.. Top Income Shares in the UK over More Than A Century. *Journal of Public Economics*, 2018, 162 (2018) : 26-47.
- [14] Arnold, B. C.. Pareto Distributions. New York: Chapman and Hall, 2015.
- [15] Atkinson, A. B., Piketty, T., & Saez, E.. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 2011, 49 (1) : 53-71.
- [16] Autor, D., Katz, L., & Kearney, M.. Trends in U.S. Wage Inequality: Revising the Revisionists, *Review of Economics and Statistics*, 2008, 90 (2) : 300-323.
- [17] Benhabib, J., & Bisin, A.. Skewed Wealth Distributions: Theory and Empirics. *Journal of Economic Literature*, 2018, 56 (4) : 1261-1291.
- [18] Bharti, N. K.. Wealth Inequality, Class and Caste in India, 1961-2012. WID Working Paper, No.2018/14, 2018.
- [19] Bhattacharya, S., Deb, J., & Kundu, T.. Mobility and Conflict. *American Economic Journal: Microeconomics*, 2015, 7 (1) : 281-319.
- [20] Charpentier, A., & Flachaire, E.. Pareto Models for Top Incomes. Working Paper, No.02145024, 2019.
- [21] Chotikapanich, D., Griffiths, W. E., & Rao, D. S. P.. Estimating and Combining National Income Distributions Using Limited Data. *Journal of Business and Economic Statistics*, 2007, 25 (1) : 97-109.
- [22] Clauset, A., Shalizi, C. R., & Newman, M. E. J.. Power-Law Distributions in Empirical Data. *SIAM Review*, 2009, 51 (4) : 661-703.
- [23] Cowell, F. A.. Measuring Inequality. New York: Oxford University Press, 2011.
- [24] Cowell, F. A., & Flachaire, E.. Income Distribution and Inequality Measurement: the Problem of Extreme Values. *Journal of Econometrics*, 2007, 141 (2) : 1044-1072.
- [25] Dastrup, S. R., Hartshorn, R., & McDonald, J. B.. The Impact of Taxes and Transfer Payments on the Distribution of Income: A Parametric Comparison. *Journal of Economic Inequality*, 2007, 5 (1) : 353-369.
- [26] Davies, J., Sandström, S., & Shorrocks, A. B.. The Level and Distribution of Global Household Wealth. *Economic Journal*, 2010, 121 (551) : 223-254.
- [27] Diaz-Bazan, T. V.. Measuring Inequality from Top to Bottom. World Bank Policy Research Working Paper, No.7237, 2015.
- [28] Hajargasht, G., Griffiths, W. E., Brice, J.. Inference for Income Distributions Using Grouped Data. *Journal of Business and Economic Statistics*, 2012, 30 (1) : 563-575.
- [29] Han, X., & Cheng, Y.. Does the ‘Missing’ High-Income Matter? -Income Distribution and Inequality Revisited with Truncated Distribution. *China Economic Review*, 2019 (57) : 101-337.
- [30] Hundenborn, J., Woolard, I., & Jellema, J.. The Effect of Top Incomes on Inequality in South Africa. *International Tax and Public Finance*, 2019, 6 (1) : 28-57.
- [31] Jenkins, S. P.. Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. *Economica*, 2017, 84 (334) : 261-289.
- [32] Kennickell, A. B.. Getting to the Top: Reaching Wealthy Respondents in the SCF. *Statistical Journal of the Iaos Journal of the International Association for Official Statistics*, 2017, 33 (1) : 113-123.
- [33] Knight, J., Li, S., & Wan, H.. Why has China’s Inequality of Wealth Increased Rapidly in the Twenty-First Century?. *Review of Income and Wealth*, 2021, 31 (2) : 1-30.
- [34] Korinek, A., Mistiaen, J. A., & Ravallion, M.. Survey Non-Response and the Distribution of Income. *Journal of Economic Inequality*, 2006, 4 (1) : 33-55.
- [35] Li, Q., Li, S., & Wan, H.. Top Incomes in China: Data Collection and the Impact on Income Inequality. *China Economic*

Review, 2020 (62) : 101495.

- [36] McKinley, T.. The Distribution of Wealth in Rural China. London: Macmillan Press, 1993.
- [37] Meyer, B. D., Wallace, K. C. M., & Sullivan, J. X.. Household Surveys in Crisis. Journal of Economic Perspectives, 2015, 29 (4) : 199-226.
- [38] Meyer, B. D., Wallace, K. C. M., & Sullivan, J. X.. The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences. NBER Working Paper, No.15181, 2015.
- [39] Meyer, B. D., & Mittag, N.. Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net. American Economic Journal: Applied Economics, 2019, 11 (2) : 176-204.
- [40] Piketty, T.. Capital in the Twenty-First Century. Cambridge MA: Harvard University Press, 2014.
- [41] Piketty, T., & Saez, E.. Income Inequality in the United States, 1913-1998. Quarterly Journal of Economics, 2003, 118 (1) : 1-39.
- [42] Piketty, T., Yang, L., & Zucman, G.. Capital Accumulation, Private Property and Rising Inequality in China, 1978-2015. American Economic Review, 2019, 109 (7) : 2469-2496.
- [43] Saez, E., & Zucman, G.. Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. Quarterly Journal of Economics, 2016, 131 (2) : 519-578.
- [44] Schluter, C.. Top Incomes, Heavy Tails, and Rank-Size Regressions. Econometrics, 2018, 6 (1) : 1-16.
- [45] Sicular, T., Li, S., Yue, X., et al. Changing Trends in China's Inequality: Evidence, Analysis, and Prospects. New York: Oxford University Press, 2020.
- [46] Wan, H., & Knight, J.. China's Growing but Slowing Inequality of Household Wealth, 2013-2018: A Challenge to 'Common Prosperity'. Oxford University Working Paper, No.389, 2022.

The Missing Top Wealthy and Wealth Inequality

Wan Haiyuan Yu Yangcheng

Abstract: A global challenge with household surveys is the wealth level and inequality underestimation caused by the missing and underreported top wealthy in household surveys. This paper applied a newly developed power-law distribution technique to correct for the missing samples, analyzed the wealth leakage and inequality underestimation, and compared their changes in the past 20 years based on the latest China Household Income Project data in 2018. We found that the wealth per capita in 2018 was underestimated to a large extent, while the Gini coefficient of wealth inequality was understated by 7.4% and continued to rise over time. Since the wealth leakage was mainly concentrated in the top wealthy, it is necessary to pay attention to the widening gap.

Keywords: The Top Wealthy; Wealth Underestimation; Wealth Inequality; Power Law Distribution

(责任编辑 : 黄 蕾)