

税制、收入与不平等

第一节 税制的设计

一、政府为什么要征税？

作为美国开国元勋的重要人物之一，本杰明·富兰克林（Benjamin Franklin）曾说过：“在这个世界上，没有什么是确定的，除了死亡和税收（In this world nothing is certain but death and taxes）。”这句话说明了税收在我们经济生活中的普遍性和重要性。那么，政府为什么要收税呢？

首先，根据曼昆的原理 7 或克鲁格曼等的原理 9，市场失灵时，政府可以进行干预从而提升社会福利。比如，提供市场生产水平严重低于最优水平的公共品（如社会秩序、国防）——注意，公共品是非竞争且非排他的物品，就是你、我、他均可以使用且你的使用不影响别人的使用。再比如，管理公共资源。公共资源虽然没有排他性，但是具有竞争性。你的使用可能会影响别人的使用，如清洁的河流。另外，就是之前提到的产权保护。而这些行为，都需要资金的支持，因此政府需要收入。

其次，税收本身也是一种经济激励手段，比如在管理公共资源时，可以将负外部性内部化。另外，税收本身可以影响人们的可支配收入，从而对整体宏观经济进行调控。当然，税收还有一些别的用途，比如扶贫、贸易保护等等。

二、中国的政府财政概览及与世界的对比

中国的政府财政是怎样的呢？1950 年，我国财政收入（窄口径，只计算全国一般公共预算收入）仅为 62.17 亿元。2023 年，这一数字跃升至 216795.43 亿元（约 3487 倍），人均贡献 1.54 万元（人口 140967 万）。

我国目前广义财政收入有一个“四本账”的概念，包括一般公共预算收入，政府性基金预算收入，国有资本经营预算收入、社会保险基金预算收入。2023 年，中国广义政府收入为 32.5 万亿元，可以看到，一般公共预算收入占广义收入的 66.7% 左右（大约三分之二），是政府主要的收入来源。

如图 2.1 所示，新中国成立之初，百废待兴，财力分散薄弱，建立了“统收统支”的财政体制——财政收入占 GDP 比重在 1960 年达 39% 左右的最高点。改革开放后，财政分配开始“放权让利”，中央与地方“分灶吃饭”，从 1978 年约 31% 下滑至 1993 年约 12%。1994 年的分税制改革是新中国成立以来范围最大、调整力度最

强、影响最为深远的一次财政体制改革：它处理了中央与地方的分配关系，也促进了国家财政收入的增长，调节了地区间的财力分配，加强了中央集权，维护了全国统一市场和企业平等竞争；财政收入占 GDP 比重在 1995 年触底（约 10%）反弹，逐步增长至 2015 年的 22% 左右。2013 年之后，为减轻企业负担，减税降费规模越来越大，财政收入占 GDP 比重有所下滑，2022、2023 年比重均不足 17%。

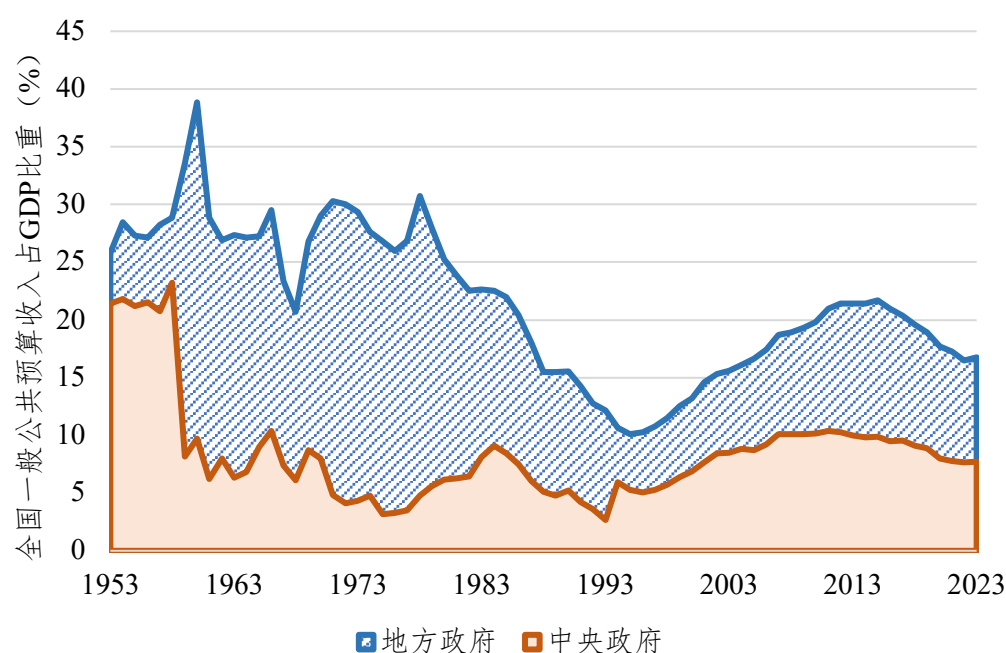


图 2.1：中国政府财政概览

注：数据来源于国家统计局，访问于 2025 年 2 月 13 日，笔者根据数据进行计算后绘制此图。

上图数据的获取流程是：

- 登录国家统计局网站，并点击数据模块 (<https://www.stats.gov.cn/sj/>)；
- 向下滑动至数据查询，并选择“年度数据”；
- 在右侧栏目中选择指标下的“财政”和“国民经济核算”；
- “财政”下的“中央和地方财政收入及比重”以及“国民经济核算”中的“国内生产总值”提供了计算全国一般公共预算收入占 GDP 比重所需的数据。

和中国类似，美国财政也有中央和地方之分，不过美国的中央政府被称为联邦政府（federal government），而地方政府还有州及（比州更小的）地方政府（state and local governments）之分。

相比我国中央财政收入略少于地方财政收入，美国的联邦政府收入占总政府收入的三分之二。另外，美国的中央政府也会向地方政府进行支付转移，主要目的是向欠发达地区进行政策倾斜，支持落后地区的发展，从而缩小地域差异。

关于财政支出的具体内容，我们留到以后阐述。

三、税收概览、结构与种类

政府收多少税呢？我国财政收入包括税收收入和非税收收入，而非税收收入中包括专项收入、行政事业性收费、罚没收入等等，比如一些明星被没收违法所得并处以罚款。我国国家税收收入占国家财政收入的比重如图 2.2 所示：

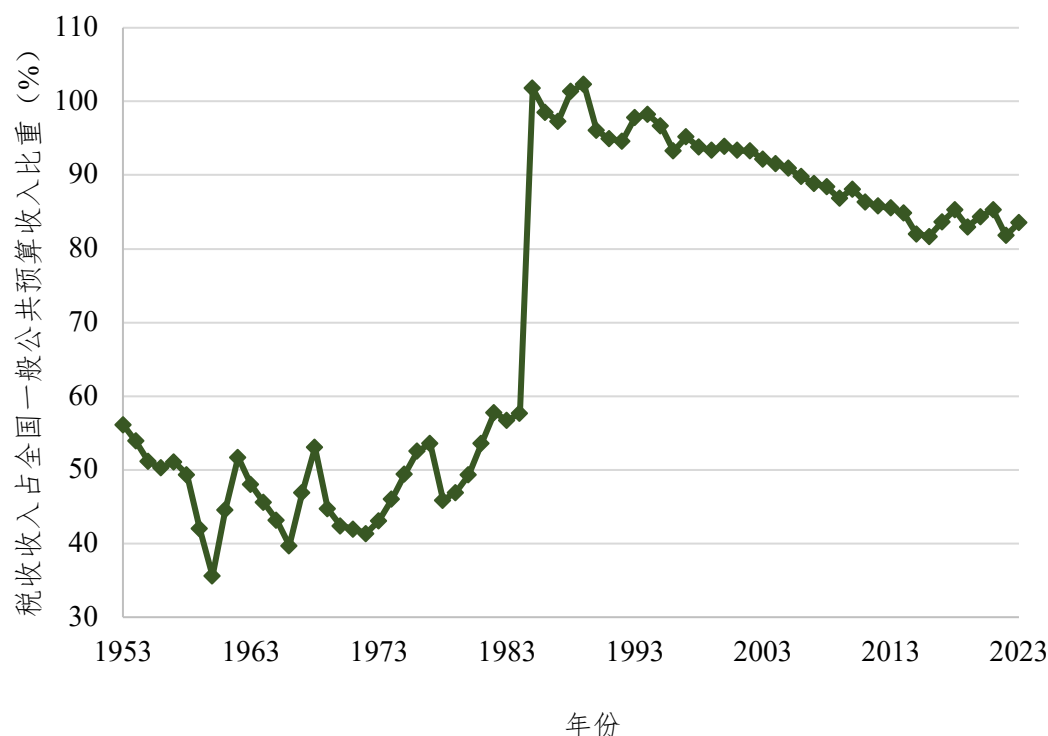


图 2.2：中国国家财政重要收入——税收

注：数据来源于国家统计局，访问于 2025 年 2 月 13 日，笔者根据数据进行计算后绘制此图。

其中，根据国家统计局，1985、1988、1989 三年的税收收入略大于财政收入，这是为什么呢？可能是因为“国有企业计划亏损补贴”冲减了财政收入所致（即非税收收入小于 0）。可以看到，1985 年起，国家税收收入成为我国财政的主要收入项目，比重超过 80%。其中，2023 年约为 83.6%。

我们再来了解一下其它国家的情况，看看税收收入占政府收入的比重是多少。根据 IMF（国际货币基金组织，International Monetary Fund）¹和 OECD（经济合作与发展组织，Organization for Economic Co-operation and Development）²数据，美国 2022 年的税收收入占 GDP 的 32.55%，财政收入占 GDP 的 27.6%，因此税收收入占财政收入的 85%左右。

在其它国家，政府也扮演着重要的角色。表 2.1 展示了税收收入占 GDP（也就是国内生产总值，具体见第三讲）的比重。可以看到，有的国家税收占 GDP 的比重

¹ 具体请访问 <https://www.imf.org/external/datamapper/rev@FPP/USA/>。

² 具体请访问 <https://www.oecd.org/tax/revenue-statistics-united-states.pdf>（访问于 2025 年 2 月 13 日）。

非常高——比如丹麦。丹麦的税收为什么这么高呢？事实上，这与这个国家的制度有关。丹麦是一个追求高福利的国家：在丹麦所有人从出生，到上幼儿园、小学、初中、高中乃至大学的所有学费都是由国家支出的；在丹麦除了那些上学的开销不用自己出之外，丹麦的教育者还会根据每个孩子不同的性格、习惯，给他们安排不一样的校园生活。例如，如果有人不喜欢在集体中学习，丹麦就会安排老师进行单独辅导。由此，丹麦常年被评为世界上最幸福的国家。根据世界银行的数据，他们的人均 GDP 在 2023 年达到 68453.9 美元。

表 2.1：各国政府 2023 年税收占 GDP 比重

国家	比重（%）	国家	比重（%）
法国	43.8	日本	34.4*
丹麦	43.4	加拿大	34.8
意大利	42.8	澳大利亚	29.4*
瑞典	41.4	韩国	28.9
希腊	39.8	美国	25.2
德国	38.1	土耳其	23.5
西班牙	37.3	智利	20.6
英国	35.3	墨西哥	17.7

注：数据来源于 OECD；表中标*数字使用 2022 年各国税收收入占 GDP 比重的数据，其余为 2023 年数据。

相比之下，美国 2023 年的税收占 GDP 的比重就不高了，只有 25.2%。而中国狭义税收占 GDP 比重仅有 14.0%左右，比大多数发达国家都低，体现我国税务负担（tax burden）较低。大部分欧洲国家的税务负担较高，因而能够支持更慷慨的社会保障网络（social safety net），包括对穷人和失业者更大的收入支持。

表 2.2：2022 年中美税收结构比较

税收种类	中国（万亿人民币）	美国（万亿美元）
个人所得税	1.5，比重 9.0%	3.1，比重 44.0%
公司所得税	4.4，比重 26.2%	0.5，比重 7.4%
社会保险税	0，比重 0%	1.6，比重 21.9%
增值税	4.9，比重 29.2%	0，比重 0%
产品和服务税	1.7，比重 10.0%	1.2，比重 16.3%
其它税种	4.3，比重 25.6%	0.7，比重 10.4%
总计	16.7，比重 100%	7.1，比重 100%

注：数据来源于国家统计局和 OECD，并由笔者计算。由于取整，每列的数字加总不一定等于总计。

接下来，我们进一步了解税收的结构。税收种类非常繁多，大致可以分为以下几类：个人所得税、公司所得税、社会保险税、增值税、产品和服务税以及其它税种。在美国，税收的主要成分个人所得税，而中国则是增值税；产品和服务税在我国只有消费税（或称国内货物税，实际上是一种特种消费行为税，即 **excise tax**），而在美国是销售税（**sales tax**）和特种消费行为税之和。

这里，我们以中美对比为主。表 2.2 展示了 2022 年的情况，其中一个非常明显的特征是：相比于美国，中国的个人所得税（**personal income tax**）比重小得多，而公司所得税（**corporate income tax**）占比大得多。

为了更好地了解我国个人所得税的规则，这里提供一道例题。

例 2.1

以某位北大老师为案例（数字有一定的虚构性）：一位北大老师税前年收入 40 万，个人养老金缴费 2.8 万，个人失业险 0.2 万，个人房基金 4 万，个人职业年金缴费 1 万，租房，独自养育一名 5 岁女儿并负担其教育费用，其父母均已过耳顺之年，作为独生子女的他是唯一赡养人。2020 年，她生了一场大病，而医保自付部分达 3 万元。基于这些信息，这位北大老师 2020 年大约交了多少税？

提示：（1）专项扣除包含养老保险、失业保险、医疗保险、住房公积金；（2）依法确定的其它扣除包括职业年金等；（3）部分专项附加扣除规则参见表 2.3。

表 2.3：部分专项附加扣除规则

项目	规则
住房租金	北京 1500 元/月
子女教育（3 周岁起）	1000 元/月/子女/家庭
赡养老人（60 岁起）	2000 元/月/赡养人
大病医疗	医保自付 1.5-8 万元部分据实扣除

注：本表格省略了继续教育、住房贷款以及部分细节。

表 2.4：2020 年中国居民综合所得税规则

应纳税个人所得额	税率（%）	速算扣除额
不超过 3.6 万元	3	0 元
3.6-14.4 万元	10	2520 元
14.4-30 万元	20	16,920 元
30-42 万元	25	31,920 元
42-66 万元	30	52,920 元
66-96 万元	35	85,920 元
96 万元以上	45	181,920 元

为了回答该例题，我们还需要知道一个中国居民综合所得税的规则（表 2.4）。其中，

速算扣除额是为了方便快速计算税额而推出的一个数字，使我们在计算税额时只需要使用相应收入所对应的区间的单个税率。速算扣除额的原理就是，将使用高于实际平均税率计算出的税额导致的多交的税扣除，而由于分级的规则，这个多交的数是一个常数。

使用速算扣除额后，个税的计算公式就简单了：应纳个人所得税 = 应纳税个人所得额 \times 适用税率 - 速算扣除额。而这个公式汇总的应纳税个人所得额 = 年收入额 - 60000（起征点）- 专项扣除、专项附加扣除、依法确定的其它扣除。

由于税制的差异，在美国社保费纳入税收体系，而在我国社保费很长一段时间独立于税收体系。社保税（social insurance tax）在美国又称为薪酬税（payroll tax），是雇主从工资中扣除转交给政府的税，只是因为“payroll tax”大部分用于社保和老年医保，所以才称为社保税。

当前，中国也逐渐将社保费（social insurance charges）改称为社保税了，这就是所谓的“费改税”。2023 年，我国三项社会保险基金收入共 7.92 万亿元，若也作为税种之一，则其比重超过 30%。

增值税（value-added tax, VAT）是对商品生产、流通、劳务服务中多个环节的新增价值或商品的附加值征收的一种流转税。美国虽然没有增值税，但是前面提到的销售税（sales tax）与我国的增值税的本质是类似的，只是美国对消费者征税，而中国对生产者和销售者征收。根据税负转嫁的理论，税收归宿（tax incidence）并不取决于纳税人是谁，而只取决于需求和供给弹性：如果消费者的需求弹性很大，那么税收负担就低一些；否则，生产者的税收负担低。那么，为什么美国不采用增值税呢？可能是，增值税在实际操作中有种种问题，以及与美国的政治体制有一定的冲突。具体原因，可能大家需要选一门有关税制的课程来具体学习，这里我就不再展开。

无论是产品和服务税还是增值税，往往都是累退税（regressive tax）——也就收入越高、税率越低。收入税则相反，往往是累进的（progressive tax）。从公平的角度来说，累进税有利于缩小贫富差距。从美国的税收结构来看，由于个人所得税占比高，因此税收的总体效应可能是累进的。有研究表明，我国税制整体上是累退的，不利于改善居民的收入分配不均。另外，还有一个概念是比例税（proportional tax），也就是无论一个人的收入高低，他面对的税率都是一样的。

产品和服务税（goods and services tax, GST）包括一般的销售税和特种消费行为税（excise tax），后者在我国也简称消费税或货物税，注意与“consumption tax”（可以译为花费税）区分开。销售税针对大部分的产品和服务，而特种消费行为税则只针对一小部分产品和服务（比如奢侈品、烟、酒、美黑沙龙）。销售税往往是从价税（ad valorem tax），而货物税有不少是从量税（specific or per-unit tax）。从价税就是根据商品的价格来收税，比如 10% 的税率就意味着 100 元的商品要交 10 元的税，而 1000 元的商品要交 100 元的税。从量税就是根据商品的数量来收税，与商品的价格无关。比如，每一升油收取 1 元的税，每一支烟收取 0.5 元的税，每一克糖收取 0.1 元的税，等等。销售税主要的目标是修正人们不利于自身健康的行为。由

于个人的健康具有正外部性，或者个人的不健康具有负外部性，比如增大家庭、医疗体系的负担等等，政府引入了这种特别的税。

四、税收、效率与公平

首先，是税收与效率的关系。效率的损失主要来源于两个方面——无谓损失和行政负担。征税一定会带来无谓损失吗？前面也介绍过，市场可能发生失灵，比如存在负外部性，那么政府征收矫正税有利于减小无谓损失。当然，如果市场本来就是最优的，此时征税只会造成无谓损失。这里，我们需要进一步讨论分析中到底是使用平均还是边际税率。根据曼昆和克鲁格曼等的原理 3，由于理性人考虑边际量做出“多少”的决定，我们在评价税收对激励的扭曲时，也需要考虑边际税率。而当我们讨论纳税人的税务负担时，可能用平均税率更为合理。

与边际税率相关的一个概念是总额税——即对每个纳税人都征收相同的税额。这个税是最有效率的，为什么？因为它的边际税率为 0，所以不会扭曲人们的行为！而且，它的行政负担几乎为 0，所以几乎不需要计算，也很方便实施。然而，这种税是非常不公平的，因为它是非常累退的。

这里再简单说明一下行政负担，包括报税的时间成本、报税系统的建立和运营成本、为执行税法投入的资源。一些纳税人会雇佣律师和会计为他们报税，来实现合法避税（legal tax avoidance）的目的。合法避税有时是钻税法的漏洞（loopholes），这些漏洞有时候是行政机关的失误，有时候则是纳税人和政府谈判后达成的特殊待遇。这些漏洞让税法更复杂，增加了行政负担。当然，税法简化存在效率和公平的权衡取舍。

这里再与大家探讨一个问题：我们应该对收入还是花费征税？这里涉及到效率与公平的问题，从而让我们从效率的问题逐渐转移到对公平的思考上来。

美国的收入税（income tax）远高于花费税（注意区分 consumption tax 和 excise tax）。收入税激励人们减少储蓄，因为储蓄利息会被征税。一些经济学家主张对人们的花费进行收税，而不要对收入进行收税。美国的个人退休金账户（IRA）或 401K 计划就是例子：存在里面的钱和利息在取出来用以前都不收税。花费税激励人们增加储蓄，有利于资本积累，从而促进经济增长。然而，花费税可能是累退的——许多人认为这是不公平的。

接下来，让我们正式讨论税收与公平的关系。税收可以实现财富的再分配，因此能影响公平。税务负担该如何分配是政治辩论的一个重要议题。效率与公平的权衡取舍是一方面，什么是公平则是另一方面。如何分配才公平？以下介绍两个原则。

受益原则（benefits principle）：将纳税人从政府公共支出中受益的多少作为税收负担分配的标准。我们可以用受益原则来支持燃油税（gasoline tax）、累进的收入税等。

支付能力原则（ability-to-pay principle）：根据纳税人所能承受的负担来征税。支付能力原则引申出两个对公平的理解：纵向公平（vertical equity）——支付能力更强的人应纳更多的税；横向公平（horizontal equity）：有相似支付能力的人应纳等量税

收。这里存在两个争论点。第一，纵向公平中“更多”是多少？富人在绝对意义上纳更多税，可能是累进、累退或者比例的——到底哪种更公平并无显然的答案。纵向公平的现实情况是：如果将人群按收入进行五等分并比较不同收入等级的实际平均税率，我们就会发现在美国联邦政府征税是累进的，尤其是考虑了支付转移（可以被视为负的税或补贴）后更是如此。美国政府以个人所得税为主导税收项目，而我国政府则以增值税为主导项目，后者具有累退性。根据岳希明等在 2014 年的研究，我国税制整体上呈现累退性，不利于改善公民收入分配不均的情况。第二个争论点是，横向公平中的“相似”如何判断？在同一个收入等级（甚至有相同收入）的家庭未必有相似的支付能力。我们需要判断哪些因素决定支付能力，而哪些因素不相关。对于个人所得税，除了收入本身，我国还考虑：子女教育、继续教育、住房贷款、住房租金、赡养老人、大病医疗等。

税收归宿也是考量公平时不可忽略的问题——法律归宿（statutory incidence）和经济归宿（economic incidence）有差别。

经济学本身不能给出一个完美的答案，因为公平的定义并不统一，并且应该给效率和公平赋予怎样的权重因人而异。那么，政府的目标是什么？回答这个问题或许还需要政治哲学。经济学可以衡量税收归宿，为公平的评判提供基础。经济学还可以衡量经济效率，并避免没有改善公平但却牺牲经济效率的政策，并阐明在可能性边界（possibility frontier）上经济效率和公平的权衡取舍。

第二节 收入与歧视

收入分配理论自从古典经济学出现以来就一直是经济学中核心理论。随着凯恩斯《通论》的发表，收入分配研究开始与宏观经济更紧密地结合起来，比如将贫富差距与经济增长进行联系。起初，凯恩斯学派还未将宏观理论建立在理性个人的效用最大化之上，缺乏微观基础，因而研究方法相对局限，往往通过猜想构建理论并结合数据进行经验分析。库兹涅茨（Kuznets）于 1955 年提出了经济增长与收入不平等的“倒 U 型”关系假说（即著名的库兹涅茨曲线），引发众多学者的讨论与争议。

为了更好地理解收入不平等（相对贫困）与绝对贫困的问题，本小节将从收入的微观决定因素开始说起。

一、工资的决定、演变与差别

我们首先讨论工资的决定。工资（wage）其实是劳动的租金（rental price），而劳动则是一个生产要素（factor）。根据新古典分配理论（neoclassical theory of distribution），一个要素——如劳动、资本、土地——的租金取决于该要素的需求和供给。

从要素需求方的角度来看，租金就是一段时间内的使用价格，反映其支付意愿（willingness to pay）。对于一个竞争的、利润最大化的生产者来说，支付意愿就等于

边际产品价值（value of marginal product）。当需求方有垄断力量时，可能出现支付意愿小于边际产品价值的情况。那么，是否存在支付意愿大于边际产品价值的情况呢？在理性人和完美信息的假设下，是不存在的。为什么？因为理性决策者不愿意做亏本生意。值得一提的是，在完全竞争的劳动力市场中，支付意愿的下降也反映出边际产品的下降。对于一个竞争的、利润最大化的生产者来说，支付意愿就等于边际产品价值。那么，随着企业投入更多的劳动要素，边际劳动产品价值将越来越小，这反映一种规模收益递减（decreasing return to scale）的生产技术。

从要素供给方来看，租金则是一段时间内的收入，反映其出让意愿（willingness to sell）。对于竞争的供给方而言，卖价等于边际成本。³当供给方有垄断时，可能会出现卖价大于边际成本（即边际成本小于卖价）的情况。那么，是否会出现卖价小于边际成本的情况呢？回答同样是不会。为什么？因为这是亏本生意——理性决策者不会（不愿意）这么做。

从市场的角度来看，租金在一段时间内使得需求和供给双方达到平衡。因此，需求和供给双方共同决定租金（如图 2.3 所示）。

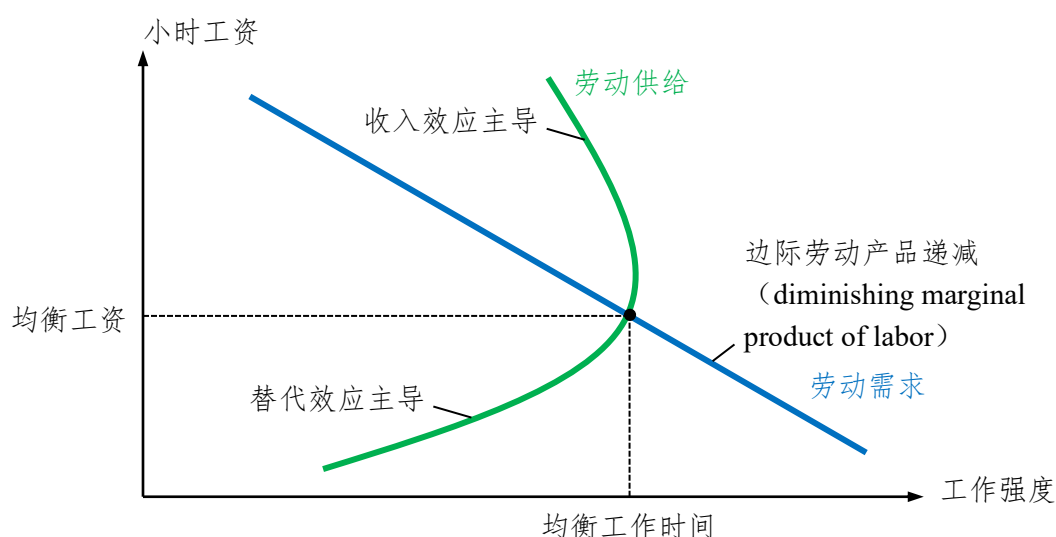


图 2.3：工资决定的供需模型

注意，图 2.3 的横轴是一段时间内的工作时间（工作强度），而纵轴是每多工作一小时的劳动报酬。需求和供给曲线的交点即为竞争性均衡点，决定了均衡的工资和工作时间。劳动需求是一个随着工资提升减少的曲线：工资提升时，雇主的购买力下降，如果劳动要素是一个正常投入品，需求会因收入效应下降；另一方面，若雇主有别的可替代劳动的要素（如机器人），这些替代要素的相对价格下降，需求也会因替代效应而从劳动向其它要素转移。工人的劳动供给比较有趣：先是随着工资的上升而上升，但当到达一个转折点时，会随着工资的上升而下降。这里，同样有收入和替代效应：工资提高时，替代效应意味着闲暇的机会成本上升，工人应分配

³ 注意，经济学中的成本是指机会成本。

更多时间给劳动而不是闲暇；收入效应则意味着收入（购买力）升高，如果闲暇是正常（消费）品的话，就应该更多地享受生活而不是工作。两个效应相互抵抗：在工资水平较低时替代效应占主导，而工资水平较高时收入效应占主导。

在我国，收入的主要来源是工资收入、资本收入和土地收入。根据国家统计局的数据，2022 年我国居民人均可支配收入工资性收入为 20590 元，占居民人均所有可支配收入的 56% 左右，占人均 GDP 的 24% 左右。资本收入有直接和间接两种，其中直接的资本收入来自居民储蓄（居民自己投资、企业通过银行或股市融资后投资），在我国 2022 年新增总量约为 7.3 万亿元，人均 5156 元；⁴间接收入是企业自留资金（45 万亿元⁵），这些收入能提高未来资本收入，从而间接地为居民带来收益。土地收入主要来自农村土地承包经营权流转等，主要针对农村居民而非城市居民。

我国工资的概念比较模糊：税前工资包括单位从个人工资中直接为其代扣或代缴的房费、水费、电费、住房公积金和社会保险基金个人缴纳部分等，不论是计入成本的还是不计入成本的，不论是以货币形式支付的还是以实物形式支付的，均列入工资的计算范围。在国外，工资特指小时工资，而按月或年发放的往往叫薪水（salary）或薪酬（compensation）。

18 世纪工业革命以前，劳动报酬的发放形式与现在的工资有较大的差异，或者称之为实物工资：食品、住宿和其它消费品。这种工资的一个好处是不受通胀影响，价值比较稳定。工业革命后⁶，市场经济兴起，为了方便交易，货币支付方式逐渐普及。随着社会的发展，主要进行脑力劳动的“白领”与从事体力劳动的“蓝领”不断分化，出现薪水的概念。而薪酬涵盖现代货币在工资的重要补充形式（如延期支付、带薪休假等），成为一种新的管理和激励手段。

随着后工业化、信息社会的发展，如今工作的本质差别在日益缩小，工资、薪水和薪酬概念的区分日益变得模糊和不再必要。现代经济学中，往往将雇员从雇主得到的各种货币收入统称为薪酬，或报酬（pay）。

我国各行各业的工资水平存在较大的差别。如表 2.5 所示，在我国，信息传输、计算机服务和软件业和金融业的工资水平较高，而农、林、牧、渔业和住宿和餐饮业的工资水平较低。为什么会有这么大的差异？为了回答这个问题，我们或许可以进一步思考：工资是否等于就业人员在劳动力市场上的收入？这些行业的劳动需求与供给有什么不同？

⁴ 该计算首先通过居民人均可支配收入和年末总人口计算了居民可支配收入总量，然后减去居民消费总量得到居民储蓄总量，然后除以年末总人口得到居民人均储蓄 5156 元；由于人口的单位为万，居民消费总量的单位为亿元，该计算可能因为取整、保留小数等原因产生误差。另外一个计算方式是直接用居民人均可支配收入减去居民消费水平：根据国家统计局，2022 年居民人均可支配收入为 36883 元，居民消费水平为 31718 元，差值为 5165 元，与上面的计算接近。

⁵ 该计算使用资本形成总额减去居民储蓄总量（见脚注 5 的计算）。

⁶ 我国的演化历程与西方有差异。事实上，东汉殇帝延平年间，我国就出现了非实物工资。先秦时期，工资主要是粮食，称为“禄”。在这以前，夏商周时期的“禄”按人口多少分配，相互之间差距不大。到了战国时期，“禄”的分配有悬殊。到汉代，“禄”改称为“俸”，以“石”或“斛”为计算单位发放粮食。东汉殇帝延平年间，改为半谷半钱，月俸改为“月钱”，有详细严格的按品位发放月钱的规定。唐代以后，薪俸才逐渐改为全部发现金。明代中叶，商品经济有了一定的发展，官俸改为薪金，当时称“月费”，继而改为“柴薪银”。

表 2.5：2022 年我国城镇非私营单位就业人员工资举例

部分行业	平均年工资（元）
农、林、牧、渔业	58,976
信息传输、计算机服务和软件业	220,418
金融业	174,341
科研、技术服务和地质勘查业	163,486
住宿和餐饮业	53,995
所有行业（包括以上和其它）	114,029

注：数据来源于国家统计局。

（一）补偿性工资差别

首先，需要注意的是，表 2.5 中的工资是税前工资，虽然这些工资包括丰富的形式，比如计时工资、计件工资、奖金、津贴和补贴、加班加点工资、特殊情况下支付的工资，但并非都是可支配工资。另外，由于接下来将提到的人力资本（human capital）的概念，即劳动者身上也蕴含一种资本，因此工人也分享资本利润，但这些资本利润并未计入上述工资之中。再次，工资中也未统计就业人员收到的非（难以）货币化的（隐性的）奖励或惩罚，这就要牵扯出补偿性工资差别（compensating differential）的概念：由不同工作的非货币特征引起的（货币）工资差距。

图 2.3 中，我们考虑了劳动强度的需求和供给，是一个集约边际（intensive margin）的概念。劳动强度的需求和供给被绘制在“价格-数量”平面上：当工资（价格）改变时，工作时间在相应的曲线上移动（movement along the curve）；当非货币特征发生改变时，曲线发生整体位移（shift of the curve），并导致均衡的改变。这个新古典分配理论模型是理解工资差别的起点。我们可以基于此继续分析：不同雇主、雇员的特征在劳动力市场中到底如何影响供求关系（从而影响均衡价格）。

图 2.4 中，我们从扩展边际（extensive margin）的概念出发，考虑劳动者数的需求和供给。此时，横轴是工人数，纵轴是相对工资。所谓相对工资，就是某一具体行业或职业相对于其它行业或职业的工资水平，是一个比例或倍数的概念，故没有单位。可以看到，当工作的非货币特征较差时，劳动供给（同一价格下愿意选择该工作的人）会减少，也就是为了吸引到同样数量的工人所需要支付的价格要更高（愿意接受的劳动卖价提高）；反之，当工作的非货币特征较好时，劳动供给会增加。这里，较好的特征补偿了货币工资较低的差异，而较差的特征需要由更高的货币工资来抵偿，所以我们将其称为补偿性工资差别。值得一提的是，图 2.4 中的工人数量是异质性工人数，即假设工人与工人之间有差异。如果工人与工人之间没有差异，那么劳动供给将是一条水平线——因为所有人愿意接受的劳动卖价是相同的（此时，较差的非货币特征将水平线上移）。还有一点，就是上一讲提到的辛普森悖论，在这里反映为：当我们将有转折点的个人供给曲线加总为市场供给曲线时，转折点可能会在可观察到的价格范围内消失。当然，这里横轴是工人数，并非工作时间的加总——而是将工作时间转化为“是否工作”的 0-1 变量后才加总，因此也可能会展现

出不同的供给曲线形状（如图 2.4）。

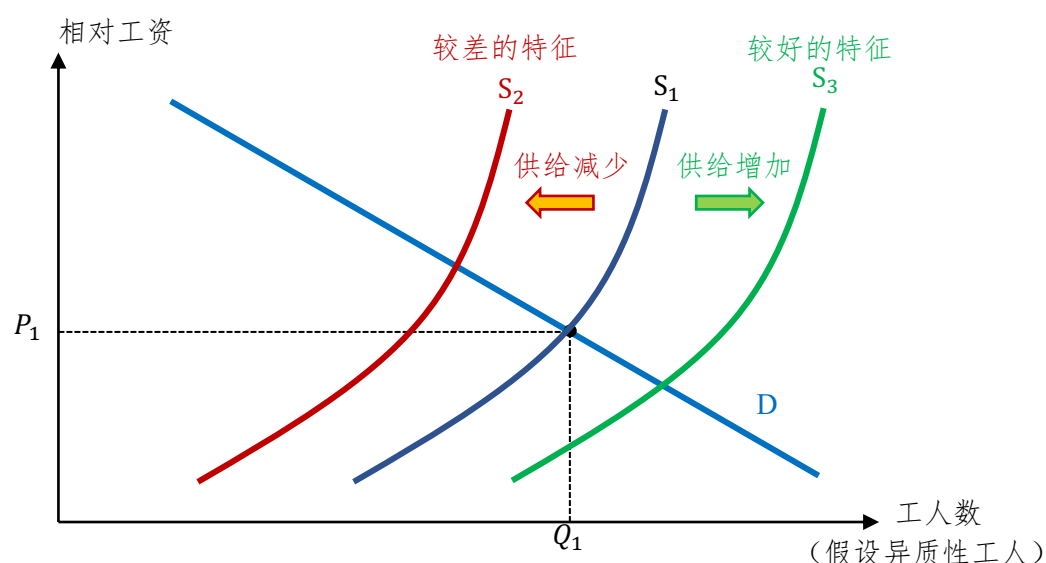


图 2.4：补偿性工资差别的供需模型

基于补偿性工资差别的模型，我们是否可以比较住宿、餐饮业和金融业之间的工资差异？更具体来说，为什么金融业的平均工资更高？我们或许可以为金融业更高的平均工资提出三点辩护：第一，金融知识和术语繁杂、对数学等技能的要求较高，对该学科内容本身真正感兴趣的人相对较少，因而需要更高的补偿性工资；第二，金融行业存在较多不确定性和风险，而大部分人不喜欢不确定性，存在风险厌恶（risk aversion）；第三，金融行业工作压力大、工作时间不规律，对身体和心理健康都存在不利影响。当然，这三点仅是基于笔者自身经验的定性分析，需要更严格的定量实证检验去验证这些说法的可靠性。我们或许能够找到一些理由支持金融业支付更高工资的做法，但两个行业之间巨大的差异似乎很难紧靠补偿性工资差异去解释——有人会说，住宿、餐饮业也有该行业的辛苦，并不比金融业轻松太多。所以，我们或许还需要继续寻找工资差别的原因。

在继续探索更多原因之前，我们需要明确补偿性工资理论的前提假设——这是我们在运用该理论时往往会忽略的问题：

- 工人追求效用最大化而不是工资最大化；
- 工人了解对他们重要的工作特征信息；
- 工人具有可流动性，有一系列可供选择的工作机会。

这些假设是相对现实的。第一点中，工人追求效用最大化的假设是比较符合直觉和常理的，也没有将只在乎钱的人排除——我们可以将只在乎钱的人的效用函数定义为只取决于钱的函数；唯一可能不太满足的情况是，人们或许不能够精确地计算出最大化效用的情况——或许计算出错，又抑或是嫌麻烦因而没有好好计算。第二点中，我们也允许信息不对称（information asymmetry）或信息摩擦（information

friction) 的存在，只要他们对重要的工作特征有稍微的了解即可，而不需要完全了解——只要有部分的了解，我们就有理由相信，他们在工作上的选择上是基于（部分）信息的。第三点，我们相信人们除了选择当前的工作以外，还是有别的选择的，因为失业也是一种选择——我们很难相信现实中有人会被强迫工作（为了生存而工作虽然可能有不情愿的成分，但在经济学中也被视为自愿——因为你是为了自己生存而工作）。在实证分析中，失业可以被定义为外部选择权（outside option）。

（二）人力资本

除了补偿性工资差别以外，另外一个解释金融业与住宿、餐饮业的工资差别的因素是人力资本（human capital）：这是一种非物质资本，是体现在劳动者身上的一种资本，如劳动者的教育文化水平、知识技能、健康状况等等。它同样是一种生产要素，对生产力的提高有着重要的作用。并且，它与人身自由紧密联系在一起。经济学中常常提到的人力资本投资，是对教育、卫生保健以及人才流动的投资。在现代人力资源管理理念中，雇员具有劳动和资本双重性，因而也要求分得劳动力市场价格和资本利润。我国人力资本总量是什么情况呢？表 2.6 用一组数据来展示。

表 2.6：2020 年我国不同受教育程度的 16-59 岁人口

受教育程度	16-59 岁人口（人）	比重（%）
未上过学	9,780,523	1.1
学前教育	588,876	0.1
小学	117,064,352	13.3
初中	366,582,059	41.8
高中	176,592,963	20.1
大学专科	105,169,983	12.0
大学本科	90,690,520	10.3
硕士研究生	9,337,884	1.1
博士研究生	1,246,184	0.1
合计	877,053,344	100.0

注：数据来源于《中国人口普查年鉴（2020）》。普查标准时点为 2020 年 11 月 1 日零时。数据包括肄业和在校生。

从表 2.6 中可以看出，我国目前人力资本存量不高：接受过高等教育的 16-59 岁人口只占到总数的 23.5%，而有研究生学历的只占到总数的 1% 左右。如果你本科毕业后保研、考研或者申请上了研究生，那么在教育人力资本上你将成为我国“百里挑一”的人。如果你是在北大读书，考虑到教育质量的差异，你可能就是“万里挑一”的人。

那么，人力资本和工资有什么关系呢？我们再看一组数据：表 2.7 的数据来自北京大学教育学院课题组进行的“2021 年全国高校毕业生就业状况调查”，涉及我

国 6 大行政区 34 所高校 2 万多名毕业生，涵盖了由专科、本科、硕士到博士的各学历层次毕业生的就业情况。其主要作者岳昌君教授担任中国教育与人力资源研究中心主任。可以看到，以受教育程度来衡量的人力资本越高，月起薪也就越高。回到新古典分配理论中，从需求的角度来看，我们可以认为，人力资本越高，边际劳动产品（生产率）也就越高，故劳动需求曲线也就越高；从供给的角度来看，人力资本越高，为其所付出（人力资本投资）的（机会）成本也就越高，需要获得更高的补偿才愿意提供劳动力。需求和供给都更高，那么工资自然也会更高。

表 2.7：2021 年我国高校毕业生起薪

受教育程度	算数平均值（元/月）	中位数（元/月）
大学专科	3910	3500
大学本科	5825	5000
硕士研究生	10,113	9000
博士研究生	14,823	15,000

注：数据来源于《全国高校毕业生就业调查报告（2021）》。

有趣的是，《全国高校毕业生就业调查报告（2021）》中还指出，工学、理学和交叉学科在平均起薪上具备优势，在各个学科门类中排名前三，分别为 8341、7909 和 7838 元/月，而在高考志愿中较受欢迎的经济学和管理学在起薪上并没有显示出优势。由此可见，除了以学历衡量的人力资本以外，还有别的因素（比如专业）影响着工具。另外，除了数量以外，教育质量也是重要的衡量人力资本的维度。该报告指出，一流大学建设高校、一流学科建设高校和普通本科高校的毕业生起薪分别为 10827、7346 和 6043 元/月。

近年来，有经济学家发现，人才越来越贵了。何以得此结论呢？从宏观数据中，研究者发现，本科毕业生与高中毕业生之间的工资差距在拉大。这里以美国为例：从表 2.8 中可以看到，与 30 年前相比，2022 年美国大学本科毕业生的平均收入与高中毕业生的绝对和相对差距都在扩大，并且这个趋势在男性和女性中均成立。由于男性和女性的比例相对稳定，可以预期辛普森悖论不会对总体结论造成太大的影响——将男性和女性混合在一起求平均，应该有相同的结论。

如果我们将大学本科毕业生称为熟练（skilled）工人，而高中毕业生称为非熟练（unskilled）工人，那么，为什么熟练和非熟练工人在一国内的收入差距在近些年变大了呢？经济学家对此给出两个猜想。第一，国际贸易（international trade）改变了市场对熟练和非熟练工人的相对需求：随着一国出口量的提升和出口产品技术含量的提升，国内市场对熟练工人的需求增多，对非熟练工人的需求减少。第二，科技的发展使得机器（如电脑）不断地替代非熟练工人的工作，并要求更多的熟练工人生产、操作和维护机器——这一现象也被称为技能偏向型技术进步（skill-biased technological change）。

表 2.8：美国逐渐扩大的本科生溢价

受教育程度	1992 年平均收入	2022 年平均收入
男性		
高中毕业生	\$53,410	\$59,590
大学本科毕业生	\$87,870	\$111,700
差距	+64.5%	+87.4%
女性		
高中毕业生	\$37,540	\$45,650
大学本科毕业生	\$59,260	\$79,210
差距	+57.9%	+73.5%

注：数据来源于 U.S. Census Bureau 表 P-32，并结合笔者的计算。仅包括 18 岁以上的全职全年（工作 50 周及以上的）员工；收入的单位均调整为与 2022 年等价的美元。

（三）天赋、努力和机遇

第三类影响劳动力供求关系（进而影响工资）的因素包括天赋/能力（ability）、努力（effort）和机遇（chance）。这里将这三种因素打包为一类，是考虑到这些因素同样都是不太好测量的因素，故而在劳动经济学家的定量研究中往往就成为所谓的遗漏变量（omitted variables）。值得一提的是，在大多数研究中，可观测变量只能解释少于 50% 的工资变异性（variation）；由此推测，这些因素对工资的影响也是十分重要的，因此这一小节也将对它们进行讨论。

首先，天赋/能力有与生俱来的，也有后天产生的：其中一个比较有趣的维度就是外貌。外貌是否影响工资？如果影响的话，又是怎么影响？经济学家对此有讨论吗？

事实上，Daniel Hamermesh 和 Jeff Biddle 于 1994 年发表在《美国经济评论》（*American Economic Review*, 简称 AER）上的一篇文章⁷就发现外貌对工资有影响。为了衡量外貌吸引力（physical appearance/attractiveness）的高低，该研究团队请调查者给受访者打分（评级）：1-惊为天人的美貌，2-高于平均（好看的），3-平均，4-低于平均（相貌平平），5-不好看（homely）。他们的主要发现是：外貌高于平均的人比外貌平均的人收入多 5%，外貌平均的人比外貌低于平均的人收入多 5-10%。

为什么外貌会影响工资呢？首先，良好的外貌可能本身就是一种能提高生产率的先天能力，因而，从需求的角度来说，雇主愿意为更高的生产率支付更高的工资。其次，由调查员汇报/评价的外貌可能间接地反映了其它方面的（后天的）能力，如言谈举止等。第三，可能存在一种针对外貌的歧视——后文具体讨论。

第二，我们简单讨论一下“努力”这个因素。直觉上，在其它条件相同的情况下，工作努力程度越高的工人的生产率越高，因而，雇主愿意为其支付更高的工资。然而，努力程度往往是一个不好度量和监测的东西，那该怎么办呢？为此，薪酬设

⁷ Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *American Economic Review*, 84(5), 1174-1194.

计的概念应运而生——对员工劳动及其工作努力程度的报酬的一种制度安排。为了激励员工努力工作，在薪酬设计时，雇主往往会考虑责任底薪、提成工资、季度/年终奖、工作表现奖、成就奖励等等。第一讲提到，人们会对激励做出反应，而这里工资就是一种激励工具。

第三，机遇也影响工资。对于这一点，经济学家的研究很少，因为这是一个更难度量的东西。我们或许可以从某些侧面去讨论这一点：比如，我们可以观察高考填报志愿的趋势——近年来，逐渐以专业为主，“冷门”专业越来越不吃香。这间接地反映出，考生和他们的家长们也指导，机遇影响未来工资！

表 2.9 反映了我国“机遇”的变化：比如，服装设计专业曾作为艺术生出国留学的热门专业被考生所热捧，然而近年来，这个专业却逐渐在中国走上倒退的路线。服装与服饰设计专业近年来被 73 所高校撤销，体现出就业前景差、招生困难的局面。这里“机遇”到底发生了什么变化？这可能与我国制造业整体产业转型有关。如今，随着中国综合国力的提升，劳动力和土地成本的提高，“中国制造”（made in China）正在试图向“中国智造”（created in China）转型，大量国内服装工厂迁至东南亚等劳动力更廉价的国家或地区，从而减少了国内对相关人才的需求。

表 2.9：2017-2022 年我国撤销的部分本科专业数量排行榜

专业名称	分类	撤销数量（高校）
公共事业管理	管理类	106
信息管理与信息系统	管理类	104
服装与服饰设计	设计学类	73
产品设计	设计学类	72
信息与计算科学	理学类	71
教育技术学	教育类	53
电子信息科学与技术	电子信息类	51
广告学	新闻传播类	50
工业设计	机械类	48
生物技术	生物类	42

注：数据来源于教育部各年度普通高等学校本科专业备案和审批结果，结合笔者计算。

那么，为什么公共事业管理、信息管理与信息系统这样的专业为什么也会被撤销这么多呢？看起来，这些专业的就业面非常广泛，也顺应时代潮流。然而，问题在于，随着我国社会发展程度的提高，对专业化的要求也在不断提高。这些专业存在的问题在于定位过于“大而空”、宽泛、不够精细，从而使得它们培养的学生“高不成、低不就”，无法满足我国市场对人才的要求。由此可见，作为一种时代洪流的宏观“机遇”，也对个人的工资有很大的影响。

（四）信号

在第（二）小节中，我们对教育的人力资本理论进行了展开。然而，在经济学文献中对教育还有另外一个理解，那就是信号（**signaling**）。具体来说，就是有经济学家认为，获得高学历（比如大学本科学位）本身实际上并没有提高人们的生产率——应该是具有较高生产率的人（能够/愿意）选择上大学。这里，大学学历成为了一种信号的筛选器（**screening device**），将高能力的人过滤出来。

其基本原理在于，能力（生产效率）本身是一个难以直接观测和衡量的东西，因此需要有一个能够传递高能力信号的机制——高能力者往往更容易获得大学学历（因为成本更低），因此也更愿意通过完成大学学历来向雇主释放这一信号。这样一来，并不是教育本身带来了高能力（高生产率），而是这些人本来就具有高生产率，而教育只是让雇主确认了这一点，并成功将这些人选择出来进行招募。

可以看出，这两种理论都能解释我们现实中观察到的宏观数据：受教育程度更高的群体平均获得更高的工资。区别在于，人力资本理论认为，教育通过提高人们的生产率提高了人们的工资；信号理论认为，教育通过筛选出有高能力的人从而让雇主愿意为他们提供高工资——如果雇主没有像教育这样的筛选工具来区分高能力和低能力，就只能一视同仁支付一样的工资了。

这两种理论虽然都能解释我们观察到的数据，但却有非常不同的政策启示。根据人力资本理论，提高所有人的受教育程度能够提高所有人的生产率，从而能提高所有人的工资——因此，政策建议往往是支持政府投入更多的教育经费，使得全民教育提升。根据信号理论，提高所有人的受教育程度并不会带来生产率或者工资的改善，反而可能给雇主带来选择雇员的难题——如果大家都是博士了，那么谁才是更有效率的那一位呢？

那么，教育的真相到底是什么？或许，真相是介于这两种极端理论之间的一种情况。然而，我们能区分这两种理论到底谁占主导吗？通过定量研究，我们是否可以测量出两种理论对工资的相对贡献度有多少？准确地测量这两种理论在现实中的相对贡献度，对我们政府的决策有非常重要的影响。然而遗憾的是，Nick Huntington-Klein 于 2021 年在 *Empirical Economics* 上发表的一篇文章指出：经验证据可以拒绝单纯的人力资本或信号理论，但却无法严格区分两者。⁸ 这里笔者希望，读者在未来的研究中能对这一问题进行更深入的分析。

（五）超级明星现象

以上理论往往只能解释工资在平均意义上的差异，而这些差异往往是相对温和的。现实中，我们发现有的人有着天文数字般的工资，这就是所谓的超级明星现象（**superstar phenomenon**）。

以体育运动员为例。我国普通运动员的收入往往在百万以内，比如根据《2022-2023 赛季 CBA 联赛球员选秀、工资帽、聘用及交易管理规定》，CBA 新秀球员工资可低至 15 万人民币/年。然而，表 2.10 显示，有的运动员工资每年可达上亿人民币。

⁸ Huntington-Klein, N. (2021). Human capital versus signaling is empirically unresolvable. *Empirical Economics*, 60(5), 2499-2531.

事实上，收入最高的体育运动员并不一直都是表 2.10 中列出的运动员，比如琼·拉姆作为高尔夫球界新星在 2022 和 2023 年均未进入前十榜单。

表 2.10：2024 年体育运动员年收入排行榜前十名

运动员	领域	年收入	工资	场外
C 罗（Cristiano Ronaldo）	足球	2.6 亿	2 亿	6 千万
琼·拉姆（Jon Rahm）	高尔夫球	2.18 亿	1.98 亿	2 千万
梅西（Lionel Messi）	足球	1.35 亿	6.5 千万	7 千万
詹姆斯（LeBron James）	篮球	1.282 亿	4.82 千万	8 千万
字母哥（Giannis Antetokounmpo）	篮球	1.11 亿	4.6 千万	6.5 千万
姆巴佩（Kylian Mbappé）	足球	1.1 亿	9 千万	2 千万
内马尔（Neymar）	足球	1.08 万	8 千万	2.8 千万
本泽马（Karim Benzema）	足球	1.06 亿	1 亿	6 百万
库里（Stephen Curry）	篮球	1.02 亿	5.2 千万	5 千万
拉马尔·杰克逊（Lamar Jackson）	橄榄球	1.005 亿	9.85 千万	2 百万

注：数据来源于福布斯（Forbes），单位是美元。

根据前面讨论过的理论，同一个行业内不同人存在收入差距并不奇怪，但是为何有的行业的“顶薪”会如天文数字这般巨大？到底是什么决定了行业工资的上限？换句话说，到底是什么因素造就了超级明星？

经济学家指出，超级明星存在的市场需要具备两个特征：第一，市场中每个消费者都希望享受最佳生产者提供的产品。第二，制造这种产品的技术使得最佳生产者能够以较低成本将该产品提供给每一个顾客。以梅西为例，每个人都希望观看最佳足球运动员参加比赛，而电视、网络直播使得梅西可以为全世界人提供这项产品（服务）。2022 年 12 月 18 日，在 2022 卡塔尔世界杯足球赛决赛中，阿根廷队与法国队在常规时间和加时赛战成 3 比 3 平，通过点球大战，阿根廷队以总比分 7 比 5 获胜，夺得冠军。而当时，不少国内球迷即使感染了“新冠”，也要坚持着熬夜观看直播。

（六）高于均衡水平的工资

前面（一）至（五）小节讨论的因素主要影响的是均衡水平的工资。实际上，工资有时可能高于（完全竞争下的）均衡水平。本小节讨论三个因素：最低工资法（minimum-wage laws）、工会/工人联合会（labor union 或 trade union）、效率工资（efficiency wage）。

首先，关于最低工资法，我们要了解它的作用机制。最低工资，顾名思义，就是一种价格下限（price floor），即规定工资不得低于的价格。我国各个省、自治区、直辖市（甚至其辖区内不同地区）的最低工资标准存在差异——分为多个档位。北京采取单一档位的最低工资标准：根据人力资源和社会保障部，截至 2025 年 1 月 1

日，标准为 2420 元/月或 26.4 元/小时。

最低工资可能会造成劳动力过剩 (surplus)，也就是失业 (unemployment)，因为劳动供给会高于劳动需求。但是，最低工资也可能不会造成这样的问题，即它可能是不具有约束力的 (not binding) 的，比如最低工资远低于均衡工资时。另外，在有买方垄断 (monopsony) 时，最低工资可能不会加重失业。在现实的劳动力市场中，买方垄断是很少见的，但作为一种理论上可能发生的情况，让我们接下来试着进行分析。

假设在一个市场中唯有一个的公司在招聘。该公司的总花费： $E = w(L) \times L$ 。要注意的是，这里 $w(L)$ 是工人数 L 的函数，表示劳动供给曲线的具体表达式。该公司的边际花费 (marginal expenditure) 可以表示为： $ME = \frac{dE}{dL} = w(L) + \frac{dw}{dL} \times L$ 。可以看到，当劳动供给曲线斜率为正（一般是如此）时， $ME > w(L)$ 。由于公司是个边际决策者，该公司雇佣工人的策略就是一个一个地雇佣工人，直到雇佣新工人的边际花费（不断上升）等于边际劳动产品（不断下降）为止。试想，在此之前，公司每多雇佣一个工人，净利润都是正的；在此之后，公司每多雇佣一个工人，净利润都是负的。用图像模型来表示的话，就是图 2.5：

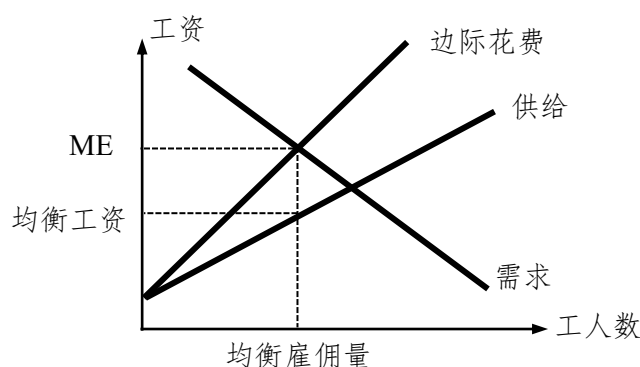


图 2.5：买方垄断的供需模型

此时，均衡雇佣量与供需曲线之间形成的三角形区域就是无谓损失 (deadweight loss)，因为从社会的角度上讲，即使再继续多雇佣一个工人生产，边际劳动产品仍大于工人的机会成本，因此会给社会带来净收益。为了改善这个情况，政府可以将最低工资设于供给与需求曲线的交点水平（见图 2.6）。

图 2.6 显示，由于最低工资的存在，该企业在需求曲线降至最低工资之前时都必须支付最低工资（而不是供给曲线上的工人愿意接受的工资——因为此前的供给曲线均低于最低工资）。由此，边际成本在此前都不随工人数 L 改变，是一条水平线。随后，由于工人愿意接受的工资大于最低工资，所以公司的边际花费将取决于工人数 L ，从而恢复到图 2.5 的情况。在这样的情况下，边际决策者（公司）选择 ME 等于边际劳动产品（需求）时，将会导致达到供给与需求的交点，即完全竞争情况下的均衡点，也是无谓损失最小（消失）的点。

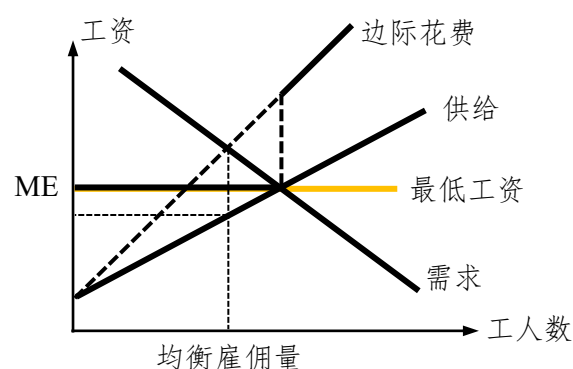


图 2.6：减少买方垄断带来的无谓损失的供需模型

这样看来，最低工资虽然会使得工资高于均衡水平，但有时候并非坏事——取决于原本的均衡到底是否是个好的均衡。这再次给政府干预提供了一个理由。

接下来，我们再讨论一下工会。所谓工会，就是职工为了争取更好的工作条件等共同目标而自愿联合起来的组织。中国和美国的工会存在差异，主要体现在独立性、手段和会员待遇这三个方面。中国工会的主要手段有协商、行政干预、法律诉讼，而美国工会可以与雇主进行集体谈判、平等协商、游行示威、联合抵制、罢工（strike）。比如，2022 年 11 月 14 日，来自美国加州大学 10 个校区的数万名学术员工停工，要求提高薪酬和福利。该工会代表约 4.8 万名助教、研究人员、博士后学者、导师和评分者（grader），他们要求大幅加薪和提高儿童保育补贴，以支付生活在洛杉矶、圣地亚哥和伯克利等房价飙升城市的日常费用；部分教师表示课程将无限期暂停。工会领导人称，这可能是加州大学系统有史以来最大规模的停工事件。关于会员待遇，中国工会会员与非会员之间工资差异不大，而美国工会会员比非会员工资高 10-20%。

最后，我们再讨论一下效率工资。企业管理者和经济学家认为，提高工资能提高员工的生产率，包括减少人员流动、提高努力程度、吸引更多优秀的人才等。代表性模型是 Carl Shapiro 和 Joseph Stiglitz 于 1984 年发表在 AER 上的怠工（shirking）模型。⁹在 Shapiro-Stiglitz 模型中，员工要么认真工作，要么偷懒；监督（monitoring）是高成本的或难以完美实现的；尽管如此，公司还是会进行一定程度的监督：一旦发现偷懒，就会开除偷懒者（由于无法完美监督，因此偷懒者只是面对一个被开除的概率）——虽然这样，提高工资也增大了被解雇的代价（即偷懒的机会成本）；由于人们会对激励做出反应，所以高工资能激励员工认真工作。模型中，公司提高工资是为了节省监督成本（只需要提供足够的开除威胁即可），只要提高工资的成本低于增加监督的成本即可。当大多数企业都这么做时，平均工资会高于均衡水平，从而导致非自愿失业（模型中还有一个外生因素导致的失业，这里就不对此进行展开了）。

⁹ Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, 74(3), 433-444.

这样决定工资的机制形成了一个不偷懒条件（no-shirking condition, NSC），它与需求的交点决定了效率工资。值得注意的是，为了简化分析，这里假设同质化（homogenous）的工人，如图 2.7 所示。

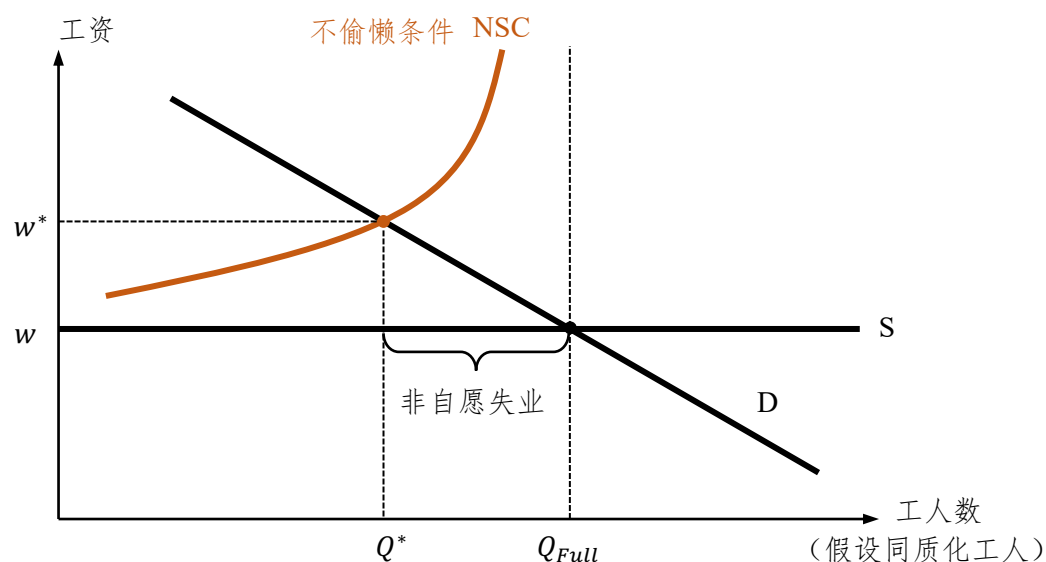


图 2.7: 一个导致非自愿失业的总工模型

（七）歧视

最后讨论一个工资差别的来源——歧视（discrimination）。什么是歧视？一般认为，歧视就是向除了种族/民族、性别、年龄、相貌等个人特征外相似的人提供不同的机会的情况。我们将重点讨论的问题是：歧视到底如何影响了工资？

首先，为了研究“如何影响”的问题，我们首先要能识别（identify）歧视的情况。下面我们先做一些简单的尝试，然后讨论这样的尝试是否是有效的。由于数据的可获得性问题，我们再次使用美国的数据。表 2.11 展示了不同性别中黑人与白人的收入差距。

可以看到，无论是对于哪个性别，黑人的中位数收入都低于白人；另外，无论是对于哪一个种族，女性的中位数收入都低于男性。此时，我们是否可以下一个结论：美国的劳动力市场存在对黑人和女性的歧视？或许，答案并没有那么简单，因为我们并没有控制中位数所对应的个人除了性别和种族以外是否相似。我们可以考察教育是否与这些个人特征有关（见表 2.12）。

可以看到，在黑人比白人工资低的同时，黑人中拥有大学本科及以上学历的比重也更少（而且男性中黑人比白人少的程度相较于女性中的情况来说更大，这也与工资的趋势一致）。或许，黑人与白人的收入差距可以在一定程度上由人力资本来解释，而不是歧视。然而，性别的差距无法被人力资本解释——美国女性中大学生占比更高，但收入更低。此时，我们可以说，美国的劳动力市场存在对女性的歧视了吗？需要注意的是，我们这里只考察了教育一个维度，而影响工资的因素有很多。

表 2.11：2022 年美国部分人群中位数收入

个人特征	白人	黑人	黑人较白人差距
男性	\$65,110	\$51,440	-21.0%
女性	\$53,120	\$48,970	-7.8%
女性较男性差距	-18.4%	-4.8%	

注：数据来源于 U.S. Census Bureau 表 P-38 并结合笔者的计算。含 15 岁及以上的全职全年（工作 50-52 周）员工；收入的单位均调整为与 2022 年等价的美元；白人指“White alone”，黑人指“Black alone or in combination”。

表 2.12：2022 年美国部分人群大学本科及以上学历的占比

个人特征	白人	黑人	黑人较白人差距
男性	33.5%	22.0%	-11.5%
女性	36.7%	27.7%	-9.0%
女性较男性差距	+3.2%	+5.7%	

注：数据来源于 U.S. Census Bureau¹⁰并结合笔者的计算。含 18 岁及以上的非公共机构居民；白人指“White alone”，黑人指“Black alone or in combination”。

或许，我们可以考虑工作经验（work experience）——可视为在工作中获得的人力资本。女性更可能因为生育、培养孩子而中断自己的事业，从而使得女性的工作经验比男性少。在美国 25-44 岁人口中，只有 75% 的女性加入劳动力，而男性则达 90%——这样一来，中年及以后，女性普遍比男性的工作经验要少。

我们还可以从补偿性工资差别来解释性别工资差异（gender wage gap）——这里我们需要考虑女性的择业行为。女性选择的职业可能更多是工资收入低，但其它非货币特征较好的职业。比如，女性更可能选择行政助理的职务，而男性更可能选择卡车司机的工作。行政助理和卡车司机的相对工资取决于这两类工作的工作环境、条件——事实上，在美国，卡车司机的工资较高。然而，由于难以对这些非货币特征定量衡量，我们不好说补偿性工资差别到底能对性别工资差距解释多少，因此也无法完全排除歧视的可能。

目前经济学文献的情况是怎样的呢？文献中有众多关于工资差别的研究，但他们也无法对歧视在劳动力市场中的普遍性达成一个清晰的结论或共识。大多数经济学家相信工资差别在一定程度上反映了歧视，但是他们无法回答具体是多大程度。

另外，即使我们将工资差异完全归结于人力资本的差异，如果我们要更追根溯源、刨根问底的话，我们甚至可以说人力资本的差异就来源于歧视！为什么？事实上，从历史上看，提供给女学生的课程大纲往往更粗糙，而黑人社区的教育资源也更差。美国经济的种种症状，或许来自政治的病根！

¹⁰ 请访问 <https://www.census.gov/data/tables/2022/demo/educational-attainment/cps-detailed-tables.html>。

拓展阅读 2.1

我们到底该如何通过经验研究（empirical research）验证歧视的存在呢？经济学家们采取了特别的策略——发送虚假简历！Marianne Bertrand 和 Sendhil Mullainathan 于 2004 年发表在 AER 上的论文做了实地实验（field experiment），¹¹并验证了美国劳动力市场存在歧视。他们向 1300 多个招聘广告投递约 5000 份假简历，随机让一半简历使用美国黑人常用名，另一半简历则使用白人常用名，其余特征相似。主要的发现是：使用白人常用名的简历获得的面试邀请比黑人常用名的简历要多 50%，并且这种情况出现在各行各业。

Philip Oreopoulos 于 2011 年发表在 *American Economic Journal: Economic Policy* 上的论文则验证了加拿大劳动力市场存在歧视。¹²他同样发送假简历，随机使用英语国家的常用名和非英语国家（如印度、巴基斯坦、中国和希腊）的常用名。主要的发现是：英语国家常用名简历获得的面试邀请比其它国常用名简历要多 39%，并且使用“英文名+中文姓”也同样获得较少的面试邀请。

2023 年诺贝尔经济学奖得主克劳迪娅·戈尔丁（Claudia Goldin）的某项研究¹³也揭露了美国的交响乐团在招聘中存在性别歧视。大多数交响乐团都是在候选人的性别已知的条件下进行招聘。某交响乐团招聘方式改为盲选时，通过拉幕布、铺地毯等方式使评委不能看到候选人，也无法根据脚步声判断性别。通过对比两种形式下的结果，克劳迪娅发现女性进入决赛成功率和最终录取率提高了。这一经验研究用数据检验了交响乐团在员工招聘中存在性别歧视的可能。

接下来，我们要探讨一个更深入的话题——到底谁才是歧视的罪魁祸首？是雇主吗？一个简单的回答是：答案没有那么简单。要问雇主是否采取歧视策略，就要问：歧视对雇主有什么好处？

让我们来做这样一个思想实验（thought experiment）：假设一个劳动力市场中只有两种工人，一种是蓝瞳孔工人，一种是黑瞳孔工人——他们具有同样的技能、经验和职业操守。然后，存在一个所谓的“歧视”，使得现在的雇主们不愿意去雇佣蓝瞳孔工人。由于对蓝瞳孔工人较低的需求，蓝瞳孔工人的工资远低于黑瞳孔工人。此时，对于一个希望进入市场的新企业来说，存在一个套利（arbitrage）的机会：若这个新的雇主雇佣蓝瞳孔工人，将能在不损害生产力的情况下降低生产成本，从而提高利润。假如市场是竞争的，那么这样的套利机会将不断吸引更多企业加入市场并组建由蓝瞳孔工人组成的企业，并不断抢夺市场份额。黑瞳孔工人企业面临激烈的竞争，市场份额将会不断减少。在这个过程中，市场对蓝瞳孔员工的需求将不断上升，而对黑瞳孔员工的需求将不断下降——最终，工资差别将逐渐消失，即所谓的“歧视”并不存在。这便是市场的自我修正（self-correcting）机制。

我们可以看到，在这个思想实验中，歧视对企业并没有什么好处，只会阻碍他

¹¹ Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991-1013.

¹² Oreopoulos, P. (2011). Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *American Economic Journal: Economic Policy*, 3(4), 148-171.

¹³ Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review*, 90(4), 715-741.

们提高利润。在竞争的市场中，企业的逐利行为是歧视天然的解药——对员工有歧视的雇主将在激烈的市场竞争中逐渐处于劣势，而更多不歧视员工的雇主将杀入市场取得商业站的胜利，从而淘汰歧视行为。

雇主往往对谋取利润更感兴趣，而不是对歧视本身。雇主产生歧视行为时，其根源可能不是雇主本身，而是其它因素导致了歧视与利润产生了挂钩。逐利的动机对歧视的修正作用往往受到这些因素的制约，从而展示出比较有限的力量。比如，在上面的思想实验中，考虑一个饮食行业，如果消费者更偏好黑瞳孔服务员提供的服务（即歧视蓝瞳孔服务员），那么为了保证消费者选择自己的服务，企业也只能更多地雇佣黑瞳孔服务员；政府也可能会颁布歧视性的法律，比如限制蓝瞳孔工人的工作范围。这些其它的力量，往往会使得歧视在经济体中持续下去。

拓展阅读 2.2

一篇 1988 年发表在 *Journal of Labor Economics* 上的文章¹⁴将指责歧视的矛头指向消费者。这篇文章研究了篮球比赛市场，因为篮球运动员的生产率或能力比较好衡量：比如得分、助攻、篮板、盖帽、抢断、失误等。该研究发现，在相似能力的情况下，黑人比白人运动员少挣 20%；与此同时，白人运动员比例更高的篮球比赛的上座人数更多。这就意味着，由于消费者偏好，黑人篮球运动员带给企业的盈利更低，从而使得企业给白人运动员更多机会（工资）。

另外，一篇 1990 年发表在 *Quarterly Journal of Economics*（简称 QJE）考察旧棒球卡（old baseball card, OBC），同样发现存在来自消费者的种族歧视。研究发现，在各方面特征相当的情况下，黑人击球员的卡片比白人击球者的卡片便宜 10%，而黑人投手的卡片比白人投手的卡片便宜 13%。¹⁵这两个案例都体现了来自消费者的歧视。

除了消费者和政府的影响，雇主展现出歧视行为的另一个原因是统计性歧视（statistical discrimination）。具体是指，一个不相关但可观测的个人特征与一个相关但不可观测的属性恰好呈现出相关性，从而引发雇主的歧视。比如，在某项工作中，女性的平均生产率低于男性，且方差更大，从而导致该工作向一个女性求职者提供的工资比一个能力相当的男性求职者更低。可以看出，性别这个可以观测的个人特征与能力这样一个不可观测的属性产生了相关性，从而导致雇主用性别来筛选能力。值得一提的是，统计性歧视的前提是信息不完全（incomplete information），并且或许信息是有成本的。统计性歧视的后果就是，雇主根据求职者的群体特征来推断个体特征，从而使得处于不利的群体遭受歧视。

统计性歧视可能是某项政策意料之外的结果（unintended consequence）。比如，某一项政策规定求职者不再需要向雇主提交犯罪记录（criminal records），这样一来，雇主无从知道雇员是否有犯罪历史（信息不完全），只能通过统计规律推断了。统计

¹⁴ Kahn, L. M., & Sherer, P. D. (1988). Racial differences in professional basketball players' compensation. *Journal of Labor Economics*, 6(1), 40-61.

¹⁵ Nardinelli, C., & Simon, C. (1990). Customer racial discrimination in the market for memorabilia: The case of baseball. *Quarterly Journal of Economics*, 105(3), 575-595.

规律显示，黑人比白人更有可能有犯罪历史。这样，雇主将会减少对黑人员工的雇佣，从而降低了他们的工资。这里，没有犯罪史的黑人处于不利的地位，遭受到了统计性歧视。可以看到，这项政策出于“防止歧视有犯罪记录的人”的“好意”，却造成了更多没有犯罪记录的人遭受歧视——实属是意料之外。

关于收入与歧视的内容介绍至此。不同工作的非货币特征会影响劳动供给，从而影响工资。受教育程度高的人平均工资更高，可能是因为人力资本得到了提升，也可能是向雇主释放了本身能力更高的信号。最低工资法、工会和效率工资会使工资高于“均衡”水平。歧视与其它因素混杂在一起，且不好量化。逐利动机可以修正歧视，但消费者偏好和政府力量可能限制了这种修正机制。

第三节 收入不平等与贫困

在阅读本节之前，请你尝试思考这些问题：如何衡量收入不平等？收入平等就公平吗？面对经济上的不平等，政府应该怎么做？是否有一些指导性的政治哲学？有哪些减少贫困的政策？

一、收入不平等的衡量

关于衡量收入不平等，我们主要想解决的四个问题是：我们的社会中收入分配到底有多么不均？我们的社会中有多少贫困人口？衡量不平等的水平时会遇到什么问题？人们在不同收入阶层的流动性如何？

让我们从五分位衡量（quintile measure）开始。根据《中华人民共和国 2023 年国民经济和社会发展统计公报》，我国居民五等份收入分组如表 2.13 所示。

表 2.13：2023 年中国居民五等份可支配收入分组均值

收入阶层	人均可支配收入（元）	份额（%）
低收入组	9215	4.4
中间偏下收入组	20442	9.9
中间收入组	32195	15.5
中间偏上收入组	50220	24.2
高收入组	95055	45.9

注：计算各阶层收入占全国总收入的比重时，笔者假定各收入阶层人数完全相等。由于取整，最后一列的数字加总不等于 100%。收入不包括应缴税款，是税后收入。

可以看到，我国高收入组（收入最高的 20%）的可支配收入是低收入组（最低 20%）的可支配收入的 10 倍左右。相比美国，五分位衡量所显示的收入不平等程度在中国更轻一些：表 2.14 显示，2023 年，美国高收入组的家庭年收入（注意收入数据来源不同）是低收入组的家庭年收入的 17 倍左右。

表 2.14：2023 年美国家庭五等份收入分组范围

收入阶层	家庭年收入范围（美元）	份额（%）
低收入组	33,000 及以下	3.1
中间偏下收入组	33,001 至 62,200	8.3
中间收入组	62,201 至 101,000	14.1
中间偏上收入组	101,001 至 165,300	22.6
高收入组	165,301 及以上	51.9
收入最高 5%	316,100 及以上	23.0

注：数据来源于 U.S. Census Bureau 表 H-1 和 H-2。收入包括应缴税款，即为税前收入。

美国公布的数据比中国多的地方在于，美国汇报了各个收入分组的分界线，而中国只有组内均值；另外，美国还汇报了收入最高 5% 组的范围和在经济中的份额。可以看到，美国收入最高 20% 的所有收入有 44% 被收入最高 5% 组拥有，展示出收入的厚尾（fat-tailed）特点。那么，收入最高 1% 是否有数据呢？

《21 世纪资本论》（*Capital in the Twenty-First Century*）的作者托马斯·皮凯蒂（Thomas Piketty）与其他经济学研究者联合执导的世界财富与收入数据库（World Inequality Database, WID），为最富有 1% 提供了估计，网址是：<https://wid.world/>。访问后，可以选择左上角的“World”，然后在新页面的左侧选中“Income Inequality”中的“Top 1% share”，接着就能看到右侧下方展示的历年（估计）数据。

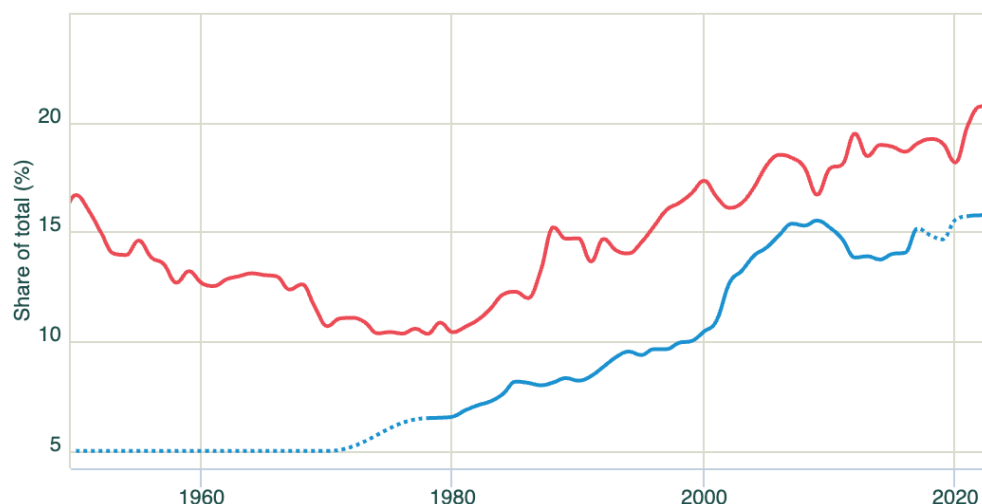


图 2.8：中国大陆与美国最富有 1% 人群的收入份额（1950-2023 年）

注：数据来源于 WID。纵坐标为份额，横坐标为年份。红色（较高）的曲线是美国数据，蓝色（较低）的曲线是中國大陸的数据，获取于 2025 年 2 月 13 日。

需要注意，图 2.8 与表 2.14 一样使用的是税前收入而不是可支配收入。中国的数据相对较不齐全，1978 年以前的数据点只有 1820、1850、1880、1900、1910、1920、1930、1940、1950、1960、1970 年（图中只展示了后三点），而美国在 1913 年后就

有逐年的数据。另外，图 2.8 与表 2.14 不同的地方是，图 2.8 使用的是个人而不是家庭作为计算单位，并且限定在年龄大于 20 岁的人。

根据托马斯·皮凯蒂等人的估计，收入最高 1% 组的收入份额衡量的收入不平等展示出美国的情况更严重一些。不过，自从 1978 年改革开放以来，中国的收入最高 1% 组人群的收入份额就在不断提升，从 6.5% 上升到 2023 年的 15.8%。美国则从 1978 年的 10.3% 上升到 2023 年的 20.7%。美国的收入不平等在 1935 至 1970 年期间逐渐减弱，但从 1970 至 2021 年却在逐渐强化。为什么改革开放以后，不平等不断上升呢？或许，国际贸易以及科技进步造成了这一结果。前文提到熟练和非熟练工人的工资差别在扩大，而这就是收入不平等的扩大。

除了纵向（跨期）比较，我们还可以考虑横向（跨国）比较，如图 2.9 所示。

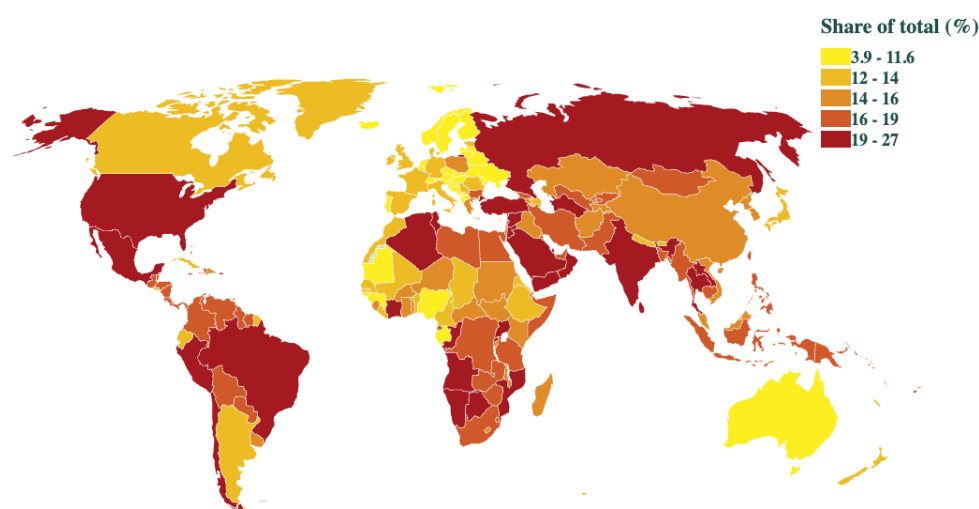


图 2.9：世界各国最富有 1% 人群的收入份额（2023 年）

注：数据来源于 WID。颜色越深代表份额越高，获取于 2025 年 2 月 13 日。

可以看到，收入最高 1% 组的收入份额为衡量的不平等展示出 2023 年我国的不平等程度在世界各国中不算高（中等），而美国的不平等程度则偏高。这里列出一些国家的数据：俄罗斯 23.8%，乌克兰为 9.3%，巴勒斯坦 18.2%，以色列 16.8%，墨西哥 21.6%。感兴趣的读者可以自行到 WID 网站上查询。

税前“收入最高 1% 人群的收入份额”这一指标可能是比较极端的衡量方式。当我们更换衡量方式时，国际比较的结论可能会改变。其它衡量方式有：收入最高 10% 组的收入份额、Palma 比值（收入最高 10% 组与收入最低 40% 组的收入份额比值）、五等份比值（高收入组和低收入组的收入份额比值）、基尼系数（Gini index）等等。

基尼系数是国际上较为通用的衡量一国或地区居民收入不平等的指标，其相关概念包括：

- 洛伦兹曲线（Lorenz curve）：一国或地区内，将人口从最穷到最富排列，累积人口百分比与累积收入份额之间的关系；
- 绝对平均线：当收入分配绝对平均时的洛伦兹曲线，此时呈一条与横轴形成

45 度夹角的直线。

若将任意洛伦兹曲线与绝对平均线之间的面积记为 A ，洛伦兹曲线与横轴之间的面积记为 B ，则基尼系数等于 $\frac{A}{A+B}$ （有的参考书会乘以 100，本讲义不乘）。因此，基尼系数的取值范围在 0 到 1 之间，数字越大代表不平等程度越大，反之则代表越平等，图 2.10 为示意图。

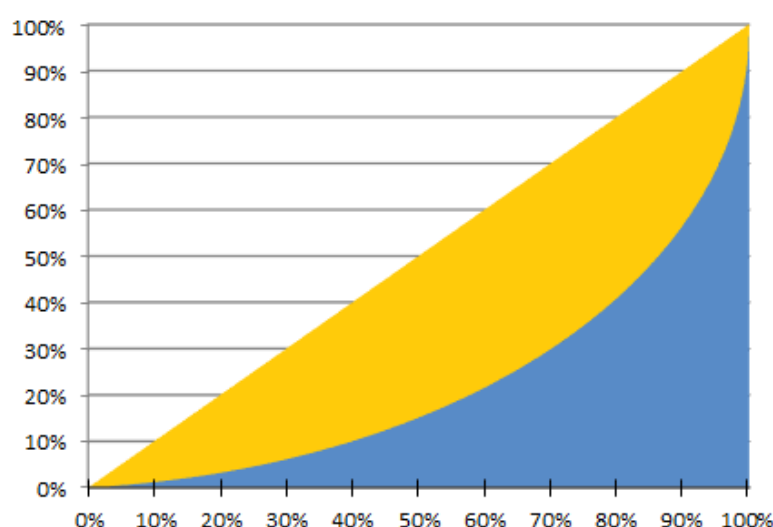


图 2.10：基尼系数的计算示意图

注：横轴为最贫穷人口的百分比，而纵轴是累积收入份额。黄色（45 度线与曲线之间）面积记为 A ，蓝色（曲线与横轴之间）面积记为 B ，则基尼系数是 $A/(A+B)$ 。曲线又称为洛伦兹曲线，而 45 度线又称为绝对平均线。

根据国家统计局的介绍，一般认为基尼系数小于 0.2 时收入分配过于平均，0.2-0.3 较平均，0.3-0.4 较合理，0.4-0.5 较不平等，大于 0.5 差距悬殊。根据国家统计局，2022 年中国居民人均可支配收入基尼系数为 0.467；美国人口普查局家庭结构调整（equivalence-adjusted）¹⁶ 的税前收入基尼系数在 2022 年也是 0.467，看起来中美不平等程度好像相当。但是，由于数据和计算方式差异，国际比较需要非常谨慎。

与“收入最高 1% 人群的收入份额”一样，基尼系数也是有缺陷的指标。洛伦兹曲线的构建对样本可靠性较依赖或敏感；基尼系数无法区分不同类型/结构的收入不平等，故也难以用于分析收入分配的结构转型；另外，我们还不能通过简单叠加或加权平均得到某一区域或整个世界基尼系数。

各国/地区数据的可获得性不同。即使在可获得数据的国家和地区中，数据搜集和统计的方式也有差异。比如，有的地方统计个人收入，有的统计家庭收入（正如

¹⁶ 家庭结构调整考虑了不同大小的家庭在资源共享和规模经济（economies of scale）上的差异，主要考虑的因素有三个：（1）儿童比成人消耗少；（2）家庭规模的扩大不会造成花费的同比例上升（往往花费的上升更缓慢）；（3）新生儿给单亲家庭带来的花费增长要大于给双亲家庭带来的花费增长。若不进行调整，基尼系数为 0.488。具体调整方式见：<https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/equivalence.html>。

表 2.13、表 2.14 和图 2.8 之间的差异那样)。根据 WID，2022 年中国税前收入基尼系数为 0.57；美国略高，为 0.59。使用 WID 的数据，我们会得出中国不平等情况比国家统计局的数据所显示的情况更严重的结论（超过了 0.5）。后文具体讨论这些不平等衡量的潜在问题。

一个与收入分配息息相关的概念是贫困率（poverty rate），能够从另外一个侧面反映收入不平等。贫困率是一国或地区收入低于贫困线（poverty line）的人口比例，而贫困线就是一个具体的收入值——不同结构的家庭的值可能不同，不同国家或地区政府也有不同的标准，并且随时间发生变化。这个收入值可以是绝对（absolute）的，也可以是相对（relative）的，还可以是弱相对（weakly-relative）的。

世界银行多次上调极度贫困线：2015 年 10 月将每人每天生活支出 1.25 美元上调至 1.9 美元，2022 年 9 月又上调至 2.15 美元（按 2017 年购买力平价）。按购买力平价的意思就是通过调整各国通货膨胀的程度以及生活成本的高低，使得各国之间的数字可比，也即能够买到的商品和服务（数量和质量）是相当的。

图 2.11 是根据世界银行的数据绘制的中国和美国极度贫困率自 2012 年以来的变化趋势。可以看到，中国的极度贫困率极度下降，在 2019 年低至 0.14%，而美国常年保持在 1% 左右。

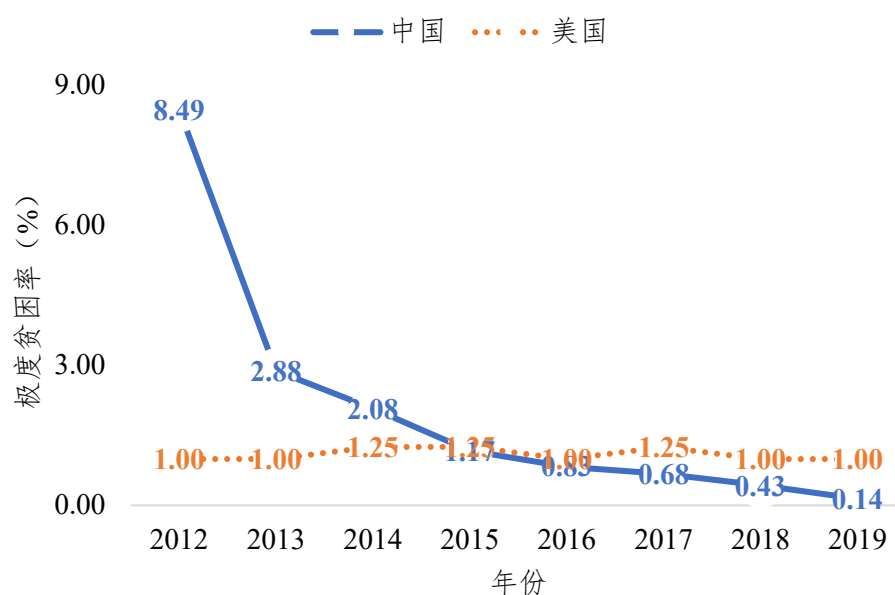


图 2.11：极度贫困率的中美比较（2012-2019）

注：数据来源于世界银行 Poverty and Inequality Platform，采取的极度贫困标准是\$2.15 国际美元（按 2017 年购买力平价）。

尽管中国的绝对贫困问题在近年来得到很大的缓解，相对贫困问题仍然需要受到重视。图 2.12 展示出，如果使用收入或消费中位数的 40% 作为（相对）贫困线，则中国的贫困率在 2019 年仍然有 6.38%，而美国则在 12% 左右。

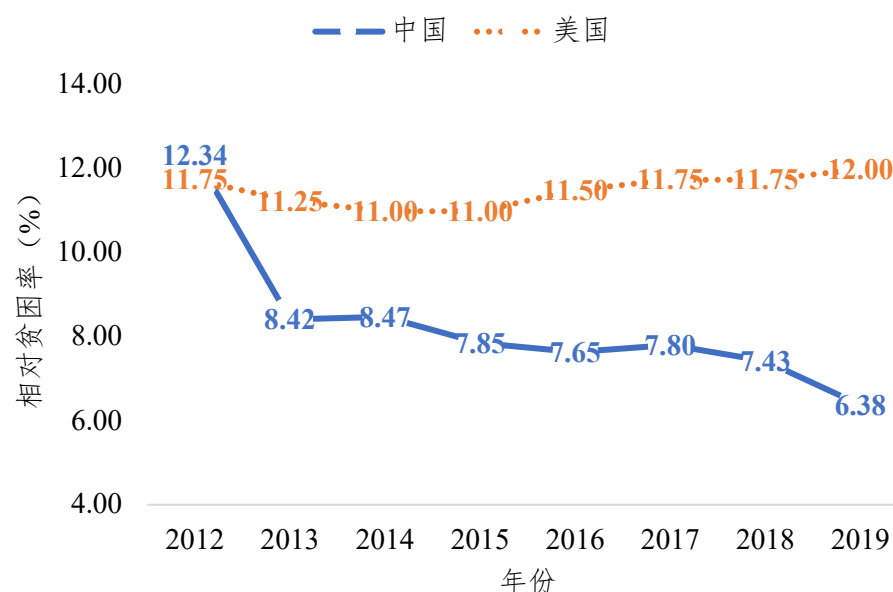


图 2.12：相对贫困率的中美比较（2012-2019）

注：数据来源于世界银行 Poverty and Inequality Platform，采取的极度贫困标准是\$2.15 国际美元（按 2017 年购买力平价）。

需要注意，世界银行数据存在一些问题降低了国际比较的可信度。首先，虽然图 2.11 和 2.12 已经按 2017 年的 1 美元在美国的购买力换算为各年各国货币的实际价值，但我们依然可以质疑这个换算的可靠性。它的标准是什么？数据来源是什么？其次，像美国这样的高收入国家的贫困率往往是根据收入调查数据计算的，而其它发展中国家（比如中国）往往采用消费调查来计算。采用消费和收入衡量不平等，得到的结论应该是不同的。而且，根据消费调查数据衡量的不平等程度往往低估了实际的不平等程度。为什么？

简单的解释是， $\text{收入} \approx \text{消费} + \text{储蓄} - \text{借贷}$ （向他人或自己）。对于高收入者，往往储蓄较高，借贷较低（甚至是负数，即出借给他人），因此消费远远小于收入；对于穷人来说，往往储蓄较低，借贷较高（比如花呗、借呗），其消费占收入很大比重，甚至大于收入！这样一来，消费随着收入的增长速度是递减的，故缩小了消费分布的范围，从而弱化了对贫富分化程度的衡量。

我国二十大报告指出，“全国八百三十二个贫困县全部摘帽，近一亿农村贫困人口实现脱贫，九百六十多万贫困人口实现易地搬迁，历史性地解决了绝对贫困问题”。可以承认的是，我国为全球减贫事业做出了重大的贡献。在接下来的发展过程中，防止返贫复贫，继续缩小相对贫困，将成为我国减贫工作的重心。

除了前面提到的衡量不平等时需注意的问题（货币价值调整、数据类型）外，还需注意以下问题（如数据来源和统计时间长短）。

第一，很多不平等和贫困的衡量采用的是年度税前货币收入，无法完整地刻画不平等。税前货币收入不仅忽略了实物转移支付（in-kind transfers）等非货币收入，也忽略了针对低收入人群的税款抵免（tax credits）。

第二，大多数不平等的衡量区间是一年，因而对长期收入（permanent income）、生命周期（life cycle）、暂时性变异（transitory variation）等因素都没有考量——这些因素往往会导致不平等程度被高估，即将年度间的自然变异性算入不平等。下面是三个可能的生命周期：（1）脚踏实地型，中规中矩，符合大多数人的规律——年轻时收入较低，随着年龄、阅历、地位的增长，收入不断提高，在 50 岁左右达到巅峰，随后快速下降至退休年龄；（2）少年得志型，年轻时快速积累一大笔财富，而中年到晚年主要靠吃老本；（3）大器晚成型，年轻时不得志，晚年突然事业起飞，走向人生巅峰。

图 2.13 显示，这三种类型的人或许终身收入差不多（曲线的积分，也即曲线以下、横轴以上的面积），但是他们在不同时间却显示出不同的不平等程度。当社会中存在各种各样人生阶段的个体时，我们如果只用某个时点的收入来衡量不平等程度，可能会将这种与不平等无关的“变异性”也纳入了不平等的计算之中，从而高估了不平等程度。

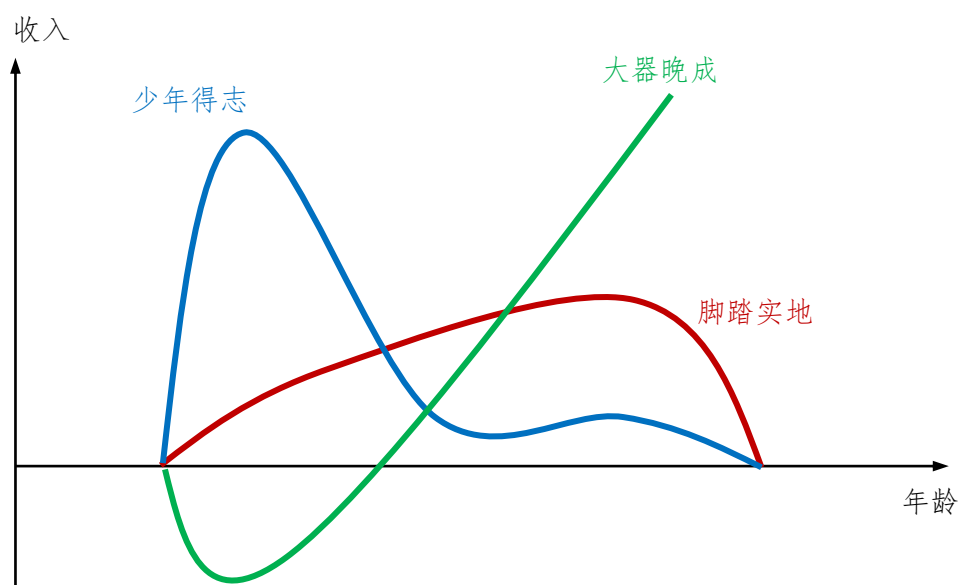


图 2.13：不同的收入-年龄曲线

最后，介绍一个与不平等衡量相关的概念：经济流动性（economic mobility），即不同的收入阶层不会永远被同一伙人占有的情况，即人们在不同收入阶层中流动的现象。

从图 2.13 可以看出，临时性（短暂）的贫困往往比一生（长期）的贫困要普遍。在美国，大约四分之一的家庭每 10 年中就有至少 1 年的收入是低于贫困线的，然而只有 3% 的家庭连续 8 年以上每年收入都低于贫困线。因此，为了将贫困衡量的尺度拓展到更长期，我们考虑使用代际流动性（intergenerational mobility）的概念，即多辈人之间是否存在贫困的延续。

研究表明，代际流动性在不同国家差异较大，且与不平等存在负相关关系。这

种负相关关系被一条“了不起的盖茨比曲线”（The Great Gatsby Curve）刻画出来（如图 2.14）。根据《Global Social Mobility Index 2020》，我国的代际收入弹性在各国中处于中间水平，而基尼系数也是中间水平。如果将各个国家的代际收入弹性与基尼系数绘制在一个平面上，我们可以发现各个国家的坐标散落在一条斜率为正的直线周围。基尼系数越高，代表不平等程度越大；代际收入弹性越高，则代表流动性越小。

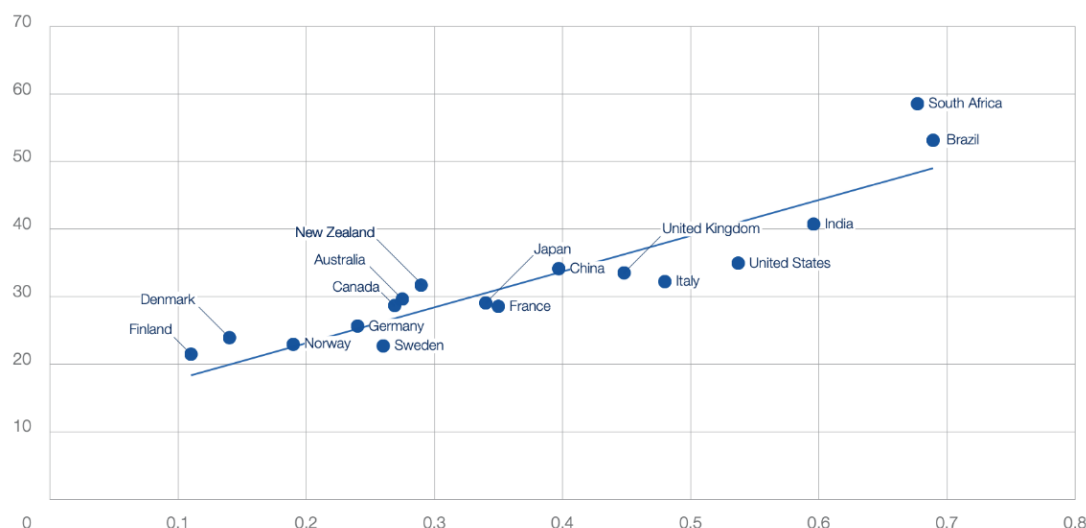


图 2.14：一条了不起的盖茨比曲线

注：横轴为代际收入弹性，纵轴是基尼系数（乘以 100）。图片来源于《Global Social Mobility Index 2020》。

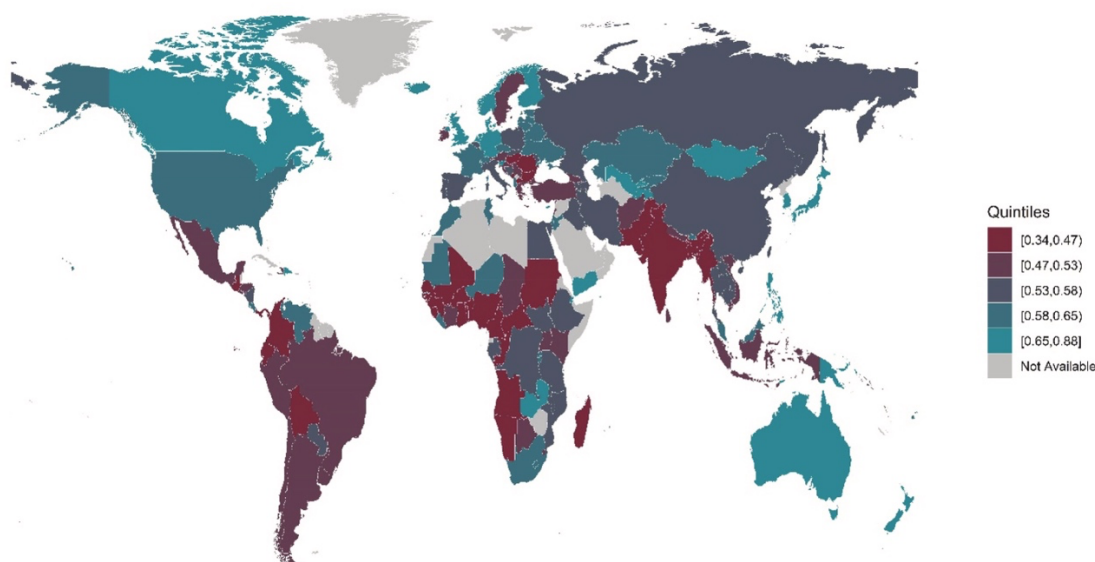


图 2.15：教育流动性在世界各国的分布

注：图片来源于 van der Weide 等（2024）的论文，图例是 1 减去代际受教育年限相关系数。

代际流动性除了用代际间收入的相关性衡量，还可以用健康、教育等指标衡量。比如，van der Weide 等人于 2024 年 1 月发表在 *Journal of Development Economics* 上的文章¹⁷就研究了世界各国“80 后”的教育流动性。根据这篇论文，中国大陆有 57% 的“80 后”的受教育程度比他们的父母都要高；父母最高受教育年限每增加 1 年，其受教育年限增加 0.48 年（不一定是因果）。¹⁸图 2.15 反映了中国的教育流动性在世界各国之中居于中间位置。

二、收入平等与公平的关系

关于该问题，推荐读者阅读 2020 年 Kris-Stella Trump 发表在 *Current Opinion in Behavioral Sciences* 题为“*When and why is economic inequality seen as fair*”的论文。简而言之，收入平等与公平之间不能划等号，不平等与不公平之间也不能。两者的关系比较复杂，且涉及到政治哲学，本讲义不试图进行具体讨论。

三、收入再分配的政治哲学

虽然经济学家们更倾向做实证分析（positive analysis）而非规范分析（normative analysis），但政府在收入再分配（income redistribution）中的角色引起诸多有关经济政策的辩论，我们或许也应该了解一些非实证经济学方面的相关内容。我们主要将讨论三类政治哲学：功利主义（utilitarianism）、自由主义（liberalism）以及自由至上主义（libertarianism）。

功利主义的代表人物：杰里米·边沁（Jeremy Bentham, 1748-1832）和约翰·斯图亚特·穆勒（John Stuart Mill, 1806-1873）。主要观点是政府应选择最大化社会中所有人的总效用（total utility）的政策。具体的应用是根据边际效用递减法则，政府应将富人的收入再分配给穷人，使得收入更平等。然而，由于人们会对激励（如收入）做出反应（如工作的努力程度），政府还需要在完全平等带来的效用提升和效率下降带来的总收入下降之间寻找平衡点。因此，根据功利主义，绝对平等并非社会的最优选择。

自由主义的代表人物是约翰·罗尔斯（John Bordley Rawls, 1921-2002），其代表作为《正义论》。主要观点是政府应选择“在无知之幕后”（veil of ignorance）背后的公正旁观者所认为的公平的政策。主要应用是，对于一个不知道自己收入将会落入哪个阶层的“无知者”来说，可能更担心落入最低收入阶层的福利，故提出最大最小原则（maximin criterion），即政府应该最大限度改善社会中状况最糟的人的福利。这样的政治哲学同样不会导致完全收入平等，否则总收入下降，也会使得最糟糕的情况更糟。自由主义与功利主义相比，如果最大最小原则就是“无知之幕后”背后的公正旁观者所认为的公平的政策，那么将对向穷人的再分配提出更强的要求。

¹⁷ van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A., & Gupta, R. (2024). Intergenerational mobility around the world: A new database. *Journal of Development Economics*, 166, 103167.

¹⁸ 数据来自 <https://datacatalog.worldbank.org/search/dataset/0050771/Global-Database-on-Intergenerational-Mobility>。其中，变量 cohort 选择 1980，变量 parent 选择 max，变量 child 选择 all。变量 MIX 和 BETA 即为讨论依据。

自由至上主义的代表人物是罗伯特·诺齐克（Robert Nozick, 1938-2002）。主要观点是政府应惩罚犯罪，执行自愿协议，而不应该再分配收入。因为，只要收入分配的过程是自由公平的，那么收入分配的结果也是公平的——不管它有多么不均。该政治哲学强调，“无知之幕”背后真正的公正旁观者更应关注过程公平，保证“能者多得”；一旦规则建立且无人破坏，政府无需再干预。

四、减贫政策

尽管哲学家们对政府应该做什么存在争议，但大部分人相信，政府应为社会提供安全网（safety net），帮助最需要帮助的人。我们主要考虑四类政策：最低工资法——前面已讨论过；福利项目（welfare programs）；负所得税（negative income tax）；实物转移支付。

福利项目的一个经典例子是美国的贫困家庭临时援助（Temporary Assistance for Needy Families, 即 TANF）项目，1996 年以前的前身是抚养未成年子女家庭援助（Aid to Families with Dependent Children, 即 AFDC）计划。经济学家对此的批评主要是认为该福利项目鼓励家庭破裂或私生子，从而激化了这个项目原本想要解决的问题。为此，经济学家提出对策，使得 TANF 在 1996 年改革后设定了援助年限。在我国，还没有特别针对单亲家庭的补助政策。

关于负所得税，有经济学家主张，应该给穷人补贴（subsidy）。比如：考虑一个应缴所得税 = $0.2 \times \text{收入} - 10000$ 。这样一来，年收入低于 50000 的人将缴纳负的所得税，也就是获得政府补贴。从而，没有人的收入会低于 10000 元。对于这样的主张，经济学家的批评是：它不仅资助了不幸的人，也资助了懒人。那么有什么对策呢？美国的工作收入所得税抵免（Earned Income Tax Credit, EITC）政策只针对工作者。对此，中国还有待试点。

实物转移支付是指以物品和劳务形式而不是现金形式向穷人转移支付。物品可以是食物、衣服、避难所、玩具等；劳务可以是医疗服务（如美国的 Medicaid）。那么问题是，到底应该给实物还是给现金？经济学界有争论。一些实施实物转移支付的慈善机构的例子：中国乡村发展基金会、马来西亚的 MyKasih 以及美国的 SNAP（Supplemental Nutrition Assistance Program），等等。

减贫政策带来的“意料之外”的效应可能是“躺平的穷人”。人们会对有效边际税率（effective marginal tax rate）这项激励做出反应。假如超过 50 万的全部额外收入均被政府征收，那么对于收入恰好在 50 万的人来说，有效边际税率为 100%，那他们将不会有动力做更多的工作。类似地，对于减贫政策的受惠者而言，收入提高可能意味着失去所有的补助，而这些补助的价值可能远远高于他们收入的提高——有效边际税率超过 100%，于是他们缺乏摆脱贫困的动力，形成一种“躺平文化”。

针对这个效应，我们可以考虑三个对策：第一，我们可以渐进地减少补助，使得有效边际税率低于 100%；但这样会扩大受惠者范围，从而增加了政府的负担。第二，考虑工作福利制（workfare），即领取补助的人同时还要接受一个政府提供的工作。第三，考虑限制受惠年限；比如，TANF 限制一人一生中只能享受五年的援助。

总之，研究贫困与不平等绝非易事。我们不仅要能够准确地定义、衡量贫困和不平等，还要弄清楚其影响因素，同时要对其（以及针对它们的政策）后果有很好地了解。

课后思考题

- 1、假设未来的某一天你去美国 X 州工作和生活。你在 X 州要交 4% 的个人收入给州政府，并根据下表向联邦政府上缴部分的个人收入，且联邦政府规定薪酬税为 15.3%。

应纳税个人所得额	税率 (%)
不超过 9325 美元	10
9326-37950 美元	15
37951-91900 美元	25
91901-191650 美元	28
191651-416700 美元	33
416701-418400 美元	35
418401 美元及以上	39.6

请问：

- 假如你的税前年收入是 3 万美元，那么你的各类税费是多少？
 - 考虑所有类型的税，你的平均税率是多少？边际税率是多少？
 - 如果你的税前年收入是 6 万美元，那么你对 a 和 b 的回答将有何不同？
- 2、以下各种情况符合受益原则还是支付能力原则？为什么？
- 北京许多公园对游客收取门票。
 - 北京征收房产税，并利用该税收收入资助小学和中学的建设。
- 3、假设海南打算提高公共教育的预算，并考虑以下四种不同的税收方式来支持这项计划：
- 对食品征收销售税（或产品和服务税）
 - 对有学龄儿童的家庭额外收税
 - 对度假型住房征收房产税
 - 对珠宝征收销售税（或产品和服务税）
- 假设这些方式最终都会带来相同的税收收入。请依次与同学们讨论（必要时请一些同学回答）以下问题：
- 请用受益原则分析这些征税方案。
 - 请用横向公平原则分析这些征税方案。
 - 它们更可能是累进、比例还是累退的？
 - 你更倾向于建议海南政府采用哪种税？为什么？
- 4、如何看待部分经济学家取消对食品和衣物征收消费税（或产品和服务税）的建议？
- 5、假设我们将目前中国的经济系统替换为一个所有人的工资都相同的体系——每个人都获得 GDP/N ，这里 GDP 是国内生产总值， N 是人口。

- a. 你个人是否会倾向于生活在这样一个经济系统？为什么？
 - b. 这样一个系统里可能会存在什么问题？
 - c. 可以通过哪些机制克服这些问题？
 - d. 谁是这样一个经济系统的受益者？
 - e. 在这样一个经济系统中，什么样的工作缺少应聘者？
- 6、当前，很多北大同学会选择在暑假实习，有的同学甚至在刚过去的寒假里就已经做了两份实习。然而，很多实习工作给同学们支付的报酬很少甚至没有。
- a. 做实习的机会成本是什么？
 - b. 为什么同学们会愿意做这样的实习？
 - c. 比较暑期参加无薪实习的同学以及打零工赚钱的同学，你认为他们未来的收入会有什么区别？
- 7、在过去的四十年中，中国最富有 10% 的收入发生了什么变化？为什么？
- 8、假设一个经济体里只有 10 人。请从绝对贫困（即一般意义上的贫困）、相对贫困（即不平等）和政治哲学的角度，对以下 4 种不同的年收入的分配方式进行比较和点评：
- a. 5 人收入 10 万，5 人收入 2 万
 - b. 5 人收入 25 万，5 人收入 2.5 万
 - c. 9 人收入 6 万，1 人收入 2 万
 - d. 10 人收入 5 万

作业题

- 1、请使用美国 2000 年人口普查的公用微观数据样本（见“data_pums_2000.csv”文档）绘制洛伦兹曲线，并计算基尼系数。
 - a. 请使用 Excel、Python、R、Stata 或 Matlab 完成该题，详细介绍你的操作过程，并提供任何用来实现计算或生成图片的代码。
 - b. 请比较你计算出的数字与任意官方的美国 2000 年基尼系数公告，简要谈谈异同与原因。
 - c. 基于你的计算，你认为中国 2003 年（可以在国家统计局查找）的不平等程度更高还是美国 2000 年的不平等程度更高？支持你的结论的理由是什么？有什么因素可能会导致你的结论受到经济学家的质疑？
- 2、某北京程序员 2020 年税前年收入 50 万，个人养老金缴费 3 万，个人失业险 1 万，个人房基金 5 万，个人职业年金缴费 2 万，离婚，无房也不打算买房，租房，独自养育一名 2016 年出生的女儿并负担其教育费用，其父母均已过古稀之年，作为独生子女的他唯一赡养人。这一年，他生了一场大病，医保自付部分达 5 万元。基于这些信息，回答以下问题。
 - a. 这位程序员 2020 年的平均税率是多少？
 - b. 这位程序员 2020 年的边际税率是多少？

提示 1：中国居民综合所得税规则

应纳税个人所得税额	税率（%）	速算扣除额
不超过 3.6 万元	3	0 元
3.6-14.4 万元	10	2520 元
14.4-30 万元	20	16920 元
30-42 万元	25	31920 元
42-66 万元	30	52920 元
66-96 万元	35	85920 元
96 万元以上	45	181920 元

提示 2：下面是一些可以用得上的公式。

（1）应纳个人所得税=应纳税个人所得税额×适用税率-速算扣除额

（2）应纳税个人所得税额=年收入额-起征点-专项扣除、专项附加扣除以及依法确定的其它扣除

（3）起征点=60000

提示 3：

（1）专项扣除包含养老保险、失业保险、医疗保险、住房公积金

（2）北京专项附加扣除条件及标准：<http://bj.bendibao.com/news/201914/256634.shtml>

（3）依法确定的其它扣除包括职业年金等