

灰色收入与国民收入分配

王小鲁

前言

2005~2006年，我们在全国各地几十个不同规模城市进行了城镇居民收入与消费调查。在此基础上，2007年作者发表了题为《我国的灰色收入与居民收入差距》的研究报告¹。在这个报告中，作者推算我国在2005年的城镇居民可支配收入中，有4.8万亿元没有反映在居民收入统计数据中的隐性收入，主要发生在高收入阶层。作者认为这些隐性收入的大部分属于“灰色收入”。作者通过对城镇居民分组收入数据进行校正，发现在包括了隐性收入后，城镇最高收入和最低收入各10%家庭之间的收入差距从9倍扩大到了31倍；全国居民最高收入和最低收入各10%家庭之间的收入差距从21倍扩大到了55倍。这意味着我国国民收入的分配失衡比过去所普遍了解的情况更加严重。

该报告还根据家用汽车拥有量、商品住宅销售量、私人出境旅游的人次数、以及银行存款的分布数据，从不同角度对居民收入（主要是高收入居民的收入）进行了估计，从而对上述推算结果进行了交叉

验证。同时也对灰色收入的来源进行了一些初步分析，认为这主要是由于体制上的缺陷造成的，因此迫切需要推进体制改革来解决国民收入分配失衡问题。

上述推算的方法和结果是否可信？经过几年之后，国民收入分配状况发生了哪些变化？为了回答这些问题，我们于2009年再次组织进行了城镇居民家庭收支调查（采集的数据为2008年全年数据），并在调查数据的基础上进行了新的研究分析。本报告就这次调查的范围、调查和分析方法、以及研究结果进行陈述，并在此基础上进一步分析国民收入分配失衡的现象和所需要的应对措施。

本报告的第一部分报告我们的城镇居民收入调查方法和样本分布情况。第二部分讲述分析方法，报告技术分析的结果。第三部分在调查数据和技术分析的基础上，推算我国城镇居民可支配收入的真实水平和真实的收入差距。第四部分是对灰色收入来源的进一步分析。第五部分分析灰色收入对国民收入分配格局的影响。第六部分是本报告的结论。

* 本文是中国经济体制改革研究会灰色收入课题研究报告，课题顾问：宋晓梧、石小敏。本课题在实地调查和研究过程中，得到了许多机构和个人的帮助；在报告讨论过程中，也有许多专家提出了意见和建议，有益于报告的修改完善；在此一并致谢。但报告中的错误由作者个人负责。

1 王小鲁，《比较》第31辑，2007年7月出版，中信出版社。这个课题是由中国经济体制改革研究会和中国改革基金会国民经济研究所资助，历时两年，于2007年完成。

一、城镇居民收入调查方法和样本分布

1. 如何获得真实的调查数据

作者认为，目前关于居民收入的统计调查数据存在重大失真。特别是关于城镇高收入居民的可支配收入，失真非常严重。这并不是指统计调查方法或计算的错误。目前的城镇和农村居民住户调查样本，是根据统计学的随机抽样方法确定的。这一方法本身并不存在系统性的错误。但必须注意到以下情况：

1. 现行的住户调查抽样是基于自愿原则确定的，但高收入居民中有相当大比例不愿意接受调查，导致在抽样过程中被迫更换样本，因此在样本更换过程中发生了对高收入居民的遗漏。

2. 包括在调查样本中的高收入居民中，有许多人不愿意提供他们的真实收入信息。在他们报告的收入中，工资性收入的真实程度较高，而其他收入可能偏低；特别是其中一些人有大量不愿暴露、来源不明的“灰色收入”。这部分收入基本上不可能反映在收入调查数据中。

以上原因使统计调查发生系统性偏差，无法真实地反映居民收入分配状况。这一问题，很难在现有的住户统计样本和现有的调查方法范围内得到彻底解决，需要探索另外的途径。基于这一原因，我们研究的首要目的是取得关于居民收入的真实数据。在2005~2006年调查中，我们借鉴了社会学调查方法，由各地的专业调查人员对他们熟悉的亲属、朋友、同事、邻里的家庭收支状况进行了调查。我们还采取了若干辅助措施保证调查数据的真实性。事实证明这一方法是可行的，调查取得的数据具有比较高的可信度。

在2009年调查中，我们使用了同样的调查方法，但采取了更严格的质量控制措施，并扩大了调查样本。必须说明，由于这一调查方法不同于随机抽样方法，因此我们无法直接用这个调查样本的数据来推算我国城镇居民收入的总体分布状况，而必须借助于其他方法进行推算。关于推算方法，将在本报告第二部分中详细介绍。本部分的主要目的是对调查方法和样本分布状况进行说明。

在调查之前，我们对各地调查人员进行了问卷和调查方法培训。为了消除受访者可能的疑虑，调查问卷采取无记名方式，在调查开始前向受访者提供了调查结果只用于研究的承诺和样本数据保密承诺。在调查方法上，也采取了若干降低调查敏感度和有利于获得真实数据的措施。例如，关于调查目的，主要强调研究消费结构而不是强调研究收入水平；在问卷设计上，先问消费问题后问收入问题，先问具体分项后问消费和收入总额。在收入来源方面，问卷只要求回答简单的收入分类（包括工资性收入、兼职和劳务收入、实体经营收入、金融投资收入、财产性收入、知识产权收入、各类转移收入、和未包括在以上各项的其他收入），不要求回答具体收入来源。在访问结束后，要求调查员填写他们与受访者的关系，以及他们对问卷调查结果可信程度（包括可能的偏差方向和偏差程度）的个人估计，作为问卷的参考信息。

调查完成后，我们对问卷进行了全面的质量检查。除了对信息完整性和调查地点正确性进行核对外，我们设计了一套检查程序，对各问题之间的逻辑关系，以及各项收入和消费数据之间的数量关系，进行合理性检查，对质量不符合要求的问卷（包括信息遗漏、涂改、数据反常、不同信

息之间存在逻辑错误而又无法鉴别正确信息等等问题的数量超过标准，以及调查对象不属于城镇居民）和信息真实性值得怀疑的问卷做了剔除。

2. 调查样本的分布状况

本次调查在全国 19 个省份（包括直辖市）的 64 个不同规模城市，以及 14 个县的县城和建制镇进行，

省份（括直辖市）包括北京、上海、山东、江苏、浙江、广东、山西、河南、湖北、安徽、江西、辽宁、黑龙江、四川、重庆、云南、陕西、甘肃、青海。这保证了东部、中部、西部和东北地区都有一定数量的样本分布，并照顾了南北方的分布。

城市包括北京、上海、济南、南京、杭州、广州、太原、郑州、武汉、合肥、南昌、沈阳、哈尔滨、成都、重庆、昆明、西安、兰州、西宁、深圳、青岛、苏州、大同、鞍山、抚顺、齐齐哈尔、大庆、徐州、扬州、阜阳、芜湖、六安、日照、襄樊、宜昌、东莞、中山、绵阳、忻州、开封、三门峡、驻马店、孝感、宜都、邳州、富阳、金华、绍兴、韶关、巢湖、滁州、赣州、吉安、景德镇、九江、丹东、铁岭、牡丹江、西昌、咸阳、白银、嘉峪关、天水、玉溪。其中直辖市、省会城市和“副省级”城市有 21 个，规模较小的地级和县级城市有 43 个，在不同规模城市之间保证了较为均衡的分布。

县城和建制镇所在的县份包括山西省繁峙县、江苏省沛县、浙江省象山县、山东省平原县、齐河县、河南省滑县、湖北省大悟县、重庆市垫江县、开县、忠县、陕西省咸阳市礼泉县、甘肃省皋兰县、涪川县、青海省民和县。这些县份的地理分布也是较为均衡的。

这次调查选取的城市数量较多，而样本在各城市的分布比较分散，这是基于两个考虑：首先，一个城市样本数量过多，就无法保证调查样本家庭都是专业调查人员所熟悉的家庭，与本次调查方案的初衷相悖。其次，样本中包括较多的城市，也保证了样本具有更好的代表性。

我们的调查方法也存在缺点。一个主要问题是，该调查是一次性进行的，关于受访者家庭收入和消费的数据都是由受访者根据记忆提供（但在选取受访者的过程中，已经排除了对家庭收入和消费状况不够了解的家庭成员）。与记账式的抽样调查相比，这会产生较大的数据误差。但记账式调查不仅比一次性调查成本高、耗时长、难度大，而且由于调查内容具有敏感性，更容易导致系统性偏差。而一次性调查因记忆不准确造成的数据误差，一般而言是随机分布的，而不是系统性的。在样本平均的意义上，随机性误差会因正负互相抵消而大大减少，而系统性偏差是无法自动抵消的。因此基于本课题的研究目的和研究条件，都必须采取一次性调查的方式。

本次调查总共包括样本家庭 4 909 个，经过严格的质量检验，剔除问卷质量不符合要求的 689 个样本，另有 25 个负收入样本未包括在分析中（因为分析表明，他们大部分在正常情况下不属于低收入家庭，负收入主要是临时性经营亏损造成的），实际分析采用有效样本 4 195 个。

表 1 列出了全部调查样本和有效样本的地区分布、按城市规模的分布、受访者年龄和户籍状况分布、样本家庭最高收入者的文化程度分布、受访家庭的人均可支配收入分布等情况。可以看到，样本在全国不同区域之间、不同规模的城市之间、以及受访者的年龄、文化程度的分布是比

表 1 按各种分组的样本分布状况

1. 按地区分布	样本总数	样本总数分布	有效样本数	有效样本分布
东部地区	1 863	37.95%	4 563	37.26%
中部和东北地区	1 848	37.65%	1 605	38.26%
西部地区	1 198	24.40%	1 027	24.48%
合计	4 909	100.00%	4 195	100.00%
2. 按城市规模分布	样本总数	样本总数分布	有效样本数	有效样本分布
200 万人以上城市	2 495	50.83%	2 083	49.65%
100~200 万人城市	915	18.64%	789	18.81%
100 万人以下城市	995	20.27%	889	21.19%
县城、建制镇	504	10.27%	434	10.35%
合计	4 909	100.00%	4 195	100.00%
3. 受访者年龄分布	样本总数	样本总数分布	有效样本数	有效样本分布
20~29	1 647	33.55%	1 411	33.64%
30~39	1 383	28.17%	1 196	28.51%
40~49	1 236	25.18%	1 062	25.32%
50~59	520	10.59%	425	10.13%
60 及以上	123	2.51%	101	2.41%
合计	4 909	100.00%	4 195	100.00%
4. 受访者户籍分布	样本总数	样本总数分布	有效样本数	有效样本分布
本市城镇	4 457	90.79%	3 808	90.77%
外地城镇	276	5.62%	234	5.58%
外地农村	156	3.18%	138	3.29%
漏答	20	0.41%	15	0.36%
合计	4 909	100.00%	4 195	100.00%
5. 家庭最高收入者文化程度	样本总数	样本总数分布	有效样本数	有效样本分布
小学或以下	165	3.36%	136	3.24%
初中	970	19.76%	832	19.83%
高中（包括同等学历）	1 833	37.34%	1 565	37.31%
大专、大学本科	1 822	37.12%	1 569	37.40%
硕士、博士	82	1.67%	74	1.76%
漏答或无法确定	37	0.75%	19	0.45%
合计	4 909	100.00%	4 195	100.00%
6. 家庭最高收入者职业	样本总数	样本总数分布	有效样本数	有效样本分布
一般专业技术人员	396	8.07%	353	8.41%
中高级专业技术人员	262	5.34%	227	5.41%
其他专业人员（科教文卫等）	339	6.91%	302	7.20%
党政军机关一般干部	193	3.93%	165	3.93%
党政军机关中高级干部	52	1.06%	47	1.12%
企事业单位、社团职员	561	11.43%	483	11.51%
企事业单位中层以上管理者	327	6.66%	268	6.39%
服务人员	317	6.46%	277	6.60%
工人	659	13.42%	562	13.40%
个体工商户、自由职业者	1 008	20.53%	853	20.33%
私企所有者、合伙人、股东	317	6.46%	277	6.60%
其他职业	73	1.49%	66	1.57%
学生、研究生	20	0.41%	17	0.41%
无职业（括退休退职人员）	349	7.11%	278	6.63%
漏答或无法确定	36	0.73%	20	0.48%
合计	4 909	100.00%	4 195	100.00%

注：城市规模按市区常住人口计算。

较均衡的；但居住在较大规模城市、具有较高学历以及从事经营性活动和白领职业的人群所占比重，高于这些人群在全国城镇人口中的相应比重。这是因为根据作者2007年研究报告的结果，城镇居民收入水平的统计偏差主要发生在高收入居民。为了保证有足够数量的高收入居民样本以进行分析，在调查中有意识地增加了这部分人群的样本数量。我们使用的分析方法，将保证这种分布的差异不会影响到对全国城镇居民人均收入分布的推算结果。

二、分析方法和技术分析结果

1. 本报告使用的基本推算方法——恩格尔系数法

根据调查样本数据推算城镇居民可支配收入的方法可以概述如下：

首先，我们进行城镇居民收入调查的目的，并不是为了从调查样本直接推断城镇居民的总体收入分布状况，而是在真实可信的数据基础上，推算收入水平与若干消费特征参数之间的关系。其中一个关键的消费特征参数就是恩格尔系数（即居民家庭的食物消费支出占家庭消费支出总额的比例）。恩格尔系数是一个与收入水平相关的参数，其下降趋势能够反映收入水平的上升；这在经济学界是一个公认的事实。这是因为在满足了基本的温饱需求之后，居民会逐渐转向追求其他需求的满足，例如对出行和通信联络的需求、对奢侈品的需求、以及教育、文化娱乐等较高层次的需求。因此随着收入水平提高，居民用于食品的支出增量会递减，而用于某些较高层次消费的支出增量会递增，使它们在消费总量中的比例发生改变。

基于这个原理，我们可以基于一个比较可信、比较有代表性的调查样本，来计算居民家庭的恩格尔系数和人均可支配收入水平，并使用统计学或计量经济学方法，找出两者间的统计关系。依据得到的这些关系，我们可以对任意一组居民收入统计数据进行检验。也就是说，只要我们能够得到某一组统计样本的相对可靠的恩格尔系数，就可以依据该系数近似推算出该组居民的真实人均收入水平。因此，我们可以根据国家统计局的分组城镇住户的恩格尔系数，推算这些组别的平均收入水平，并将这些推算结果与公布的该组居民收入水平统计数据进行比较，以发现统计数据是否存在系统性的误差，以及这一误差有多大。我们称这一分析方法为“恩格尔系数法”。

当然，这样做的前提，是要求分组统计样本的恩格尔系数真实可信。一个自然会遇到的问题是，如果某一组居民的收入水平统计数据存在系统性偏差，他们的恩格尔系数数据会不会同样有系统性偏差呢？事实上，如果收入数据存在偏差（例如，被低估），那么消费和商品消费支出数据很可能也存在一定的偏差。但首先，只要消费支出和食品消费支出的偏差是同方向的，并在统计意义上大体上保持同比例，那么分组平均的恩格尔系数仍然是基本可信的。在这种情况下，我们仍然可以使用恩格尔系数来推算真实收入水平。其次，即使消费支出和食品消费支出的偏差不保持同比例，在计算恩格尔系数时，同方向的偏差仍然可以在很大程度上互相抵消，使恩格尔系数的偏差远远小于收入水平的偏差。因此仍然可以用来推算收入水平，只是推算结果的准确程度较低。

根据作者2007年的研究，发现在居民

收支统计数据中，高收入居民的收入水平数据偏差最大，明显低于他们的真实收入。他们的消费支出和食品消费支出也都存在一定程度的低估，但在程度上远远小于收入的偏差。其中食品支出的偏差比消费支出总额的偏差更小些。这也就是说，据此计算的恩格尔系数可能轻微偏高，因而根据这些恩格尔系数推算的收入水平有可能轻微偏低；但仍然可以在很大程度上校正原来的收入数据偏差。不过，我们也需要知道，我们所得到的收入水平校正结果，相比于真实收入，可能或多或少仍在一定程度上偏低。

还要注意到，使用这一方法对居民收入统计数据进行检验，并不能在数量上确定统计样本对高收入居民遗漏的情况，只能对现有统计样本收入数据的系统性偏差进行校正。因此校正之后的结果，仍然可能在一定程度上低估高收入居民的收入水平（由于样本遗漏的原因）。

以下，作者采取两种具体方法建立恩格尔系数与收入水平之间的关系，并对居民收入统计数据进行检验。这两种方法都属于恩格尔系数法，但分析手段和过程不同。在下面的叙述中，为简化起见，作者将“人均可支配收入”称为“人均收入”；将国家统计局的城镇居民住户调查样本称为“统计样本”，而将我们这次调查的样本称为“调查样本”；来自统计样本的人均收入将称为“统计收入”，而根据调查样本提供的参数及其与统计数据的比较结果推算的人均收入称为“推算收入”；以示区别。

2. 分组比较法

第一种具体推算方法可以称为分组比较法。在2007年研究报告中，作者采用的就是这个方法。其步骤如下：

第一步，分别计算全部调查样本的人均收入和恩格尔系数。

第二步，计算统计样本的分组恩格尔系数。国家统计局每年公布的城镇居民分组收入数据，是按人均收入，将全国城镇居民家庭分为七组。其中最低收入、（次）低收入、最高收入、（次）高收入这四组分别是按十等份划分的，即每组各占10%的城镇家庭。中间三组（中低收入、中等收入、中高收入组）是按五等份划分的，每组占20%的城镇家庭。这七组的平均恩格尔系数都是可计算的。根据本文前面的解释，我们假定统计样本的恩格尔系数是可信的²。

第三步，将全部有效调查样本按人均收入排序，从低到高进行分组。分组方法是，从最低收入开始，将样本逐个累加，直到该组的平均恩格尔系数与“统计样本最低收入组”的平均恩格尔系数相等为止。我们称这一组样本为“调查样本最低收入组”。这个分组过程不考虑样本数量。然后开始对“调查样本低收入组”进行分组，方法相同，也是使其平均恩格尔系数与“统计样本低收入组”（即第二个10%的城镇住户）相同。其余各组的分组方法类推。

第四步，分别计算调查样本各组的人均收入平均水平。根据前面解释的理由，我们假定一组居民的恩格尔系数与他们的收入水平具有唯一的对应关系。也就是说，给定某组居民一个恩格尔系数，则该组的人均收入就应当是我们计算得到的、对应于这个恩格尔系数的人均收入。

第五步，将各组调查样本与对应的统计样本人均收入进行对比，其差异就反映

² 国家统计局城镇住户样本的平均和分组数据见历年国家统计局：《中国统计年鉴》“人民生活”部分。

表 2 调查样本的分组分布及与统计样本分布的比较

分组	区间	调查样本		统计样本
		有效样本数	分布比例	分布比例
最低收入	1~7 000 元	365	8.7%	10%
低收入	7 001~10 000 元	622	14.8%	10%
中低收入	10 001~17 000 元	927	22.1%	20%
中等收入	17 001~26 500 元	650	15.5%	20%
中高收入	26 501~34 000 元	355	8.5%	20%
高收入	34 001~75 000 元	635	15.1%	10%
最高收入	75 001~400 000 元	565	13.5%	10%
剩余样本	>400 000 元	76	1.8%	0%
合计		4 195	100.0%	100.0%

注：统计样本总共包括城镇居民约 6.5 万户。

表 3 调查样本与统计样本的比较

分组	推算收入		统计收入		两样本比较	
	恩格尔系数	人均收入(元)	恩格尔系数	人均收入(元)	收入差额(元)	收入差率(%)
最低收入	0.4816	5 685	0.4814	4 754	931	19.6%
低收入	0.4595	8 646	0.4594	7 363	1 283	17.4%
中低收入	0.4297	13 392	0.4289	10 196	3 196	31.3%
中等收入	0.4065	20 941	0.4042	13 984	6 957	49.7%
中高收入	0.3790	29 910	0.3787	19 254	10 656	55.3%
高收入	0.3437	47 772	0.3403	26 250	21 500	82.0%
最高收入	0.2908	164 034	0.2918	43 614	120 420	276.1%
剩余样本	0.2241	658 811				

注 1：调查样本和统计样本对应各组恩格尔系数之间有很小的尾数误差，不影响分析，我们视同相等。

注 2：“收入差额”指调查收入高于统计收入的部分，“收入差率”是指收入差额相当于统计收入的百分比。

出统计样本的数据遗漏。表 2 是调查样本和统计样本的分组分布情况。可以看到在取了同样的恩格尔系数之后，调查样本的各组分布比例是不一样的。还可以看到，根据恩格尔系数分为七组之后，还有一部分人均收入更高的调查样本被留在了这七组之外（因为其恩格尔系数更低）。这一组的人均可支配年收入都超过 40 万元，最高为 176 万元。表 3 是调查样本和统计样本恩格尔系数和人均收入的对比结果。

从表 3 可以看到，在对应各组恩格尔系数一一相等的情况下，调查样本每一组的人均收入都高于对应的统计样本，但差额和差率都非常有规律地逐级扩大，特别是最高收入组的差额和差率最大，统计样本最高收入组的人均收入只有 4.3 万元，而

调查样本最高收入组为 16.4 万元，是前者的近 3.8 倍。其差额占了全部样本差额的 2/3。这种情况和作者在 2007 年研究报告中发现的情况基本一样，所不同的只是中、低收入各组的差额和差率在一定程度上大于 2007 年报告的推算。这样，也就基本上验证了 2007 年报告研究结果的可信性。但对于这一推算的可信度，我们还要在下一节中通过另一种推算方法来验证。

此外还需要说明，在作者 2007 年研究报告发表后，有少数读者误认为这一研究方法仍然等同于用调查样本来推算城镇居民总体收入分布状况，因而对研究结果的可信性提出质疑。因为无论本项研究的调查方法（非随机抽样）还是样本规模（规模偏小），都不适合用于直接推算城镇居民总

体收入分布。这种误解是由于对这项研究使用的恩格尔系数法，尤其是对分组比较的分析方法缺乏了解。实际上，即使不理解我们使用的分组分析法与从调查样本推算总体的方法有何实质区别，只要将两种方法得到的结果进行对比就清楚了。在表4中，作者使用本次调查的数据，把用分组分析法的结果与直接推算总体分布的结果进行比较。显然，两种方法不仅每组得到的恩格尔系数有差异，而且每组人均收入的差别更为显著。其中关于最高收入组的人均收入，分组分析法得到的结果是16.4万元，而推算总体的方法得到的是29.4万元。两者的区别是显而易见的。

3. 模型分析法

以分组比较法对收入水平进行推算，也存在缺点，即这种方法假定恩格尔系数只与收入水平相关。但实际上，恩格尔系数还可能受到其他因素的影响，例如消费品价格、不同地方居民的饮食习惯差异等等。因此某一个恩格尔系数是否真的只对应一个确定的收入水平？是可以怀疑的。

因此在这里，作者采用第二种具体推算方法，这可以称为模型分析法。这种方法以计量模型分析为基础，并能够把除收入水平以外还有可能影响恩格尔系数的其他变量作为控制变量包括在模型中进行检验，

并在计算恩格尔系数与收入水平的关系时把这些额外的影响因素排除在外。这种方法能够避免分组分析法的不足之处，具有明显的优点。其基本步骤可以概述如下：

第一步，确定控制变量。我们需要对调查样本的恩格尔系数和人均收入用计量经济学方法进行回归，找出人均收入对恩格尔系数的影响系数；而在进行这一步工作的同时，我们必须找到可能影响恩格尔系数的其他因素，作为控制变量包括在模型中，对其影响进行估计，才可能得到正确的收入影响系数。

首先，不同规模的城市，各类消费品的价格水平有很大差异。这可能会影响不同规模城市的恩格尔系数。例如，大城市由于远离农产品产地，农产品运输成本和损耗都较大，中间环节也较多，因此食品价格可能会显著高于中小城市，高出的幅度有可能大于其他消费品价格高于中小城市的幅度（这是因为蔬菜、肉类等农产品不耐保存，储藏成本和中间损耗较大）。因此在其他条件相同的情况下，大城市居民的恩格尔系数可能高于中小城市居民。因为无法得到物价绝对水平的数据，作者在模型中设定了一个表示城市规模的变量 city，其中对县镇、100万人以下城市（在这里作者称其为中小城市）、100~200万人之间的城市（这里称之为大城市）和200万人以

表4 基于2008年调查样本的比较：分组分析法和推算总体方法

分组	分组分析法			推算总体的方法		
	恩格尔系数	人均收入(元)	样本分布比例	恩格尔系数	人均收入(元)	样本分布比例
最低收入	0.4816	5 685	8.7%	0.4794	5 884	10%
低收入	0.4595	8 646	14.8%	0.4654	8 362	10%
中低收入	0.4297	13 392	22.1%	0.4323	12 038	20%
中等收入	0.4065	20 941	15.5%	0.4146	19 285	20%
中高收入	0.3790	29 910	8.5%	0.3656	35 606	20%
高收入	0.3437	47 772	15.1%	0.3187	76 097	10%
最高收入	0.2908	164 034	13.5%	0.2645	293 769	10%
平均/合计		35 462	98.2%		51 771	100%

上城市（这里称之为特大城市）分别赋值为1、2、3、4。

其次，不同地区居民的消费习惯有差异。有些地区居民比其他地区居民有更高的美食偏好，因此可能在食品消费方面支出多于其他地区。通过对调查样本数据的分析，作者发现在其他条件相同的情况下，上海、江西、四川的恩格尔系数显著高于各省份平均水平。这三个省份用虚拟变量H1表示。北京、山东、湖北、广东、重庆、河南的恩格尔系数在一定程度上高于平均水平，这些省份用虚拟变量H2表示。而辽宁、山西的恩格尔系数低于平均水平，两者用L1表示。据此，在模型中包括了这几个虚拟变量。未包括在内的其他省份（有江苏、浙江、安徽、黑龙江、云南、陕西、甘肃、青海）的样本作为本底样本。

第三，家庭人口数对恩格尔系数可能有影响，因为人口较多的家庭在食品支出方面可能具有规模效应，能够节约食品支出。因此设定了一个代表家庭人口数的变量family。

第四，家庭成员的平均文化程度有可能对恩格尔系数有影响，因为文化程度较高的居民可能偏向于较多的精神需求，例如通信联络、教育、文化娱乐等，而文化程度较低居民则可能在这些方面需求较少，而在食品烟酒等消费方面支出较多。因此设定了一个代表成年家庭成员平均教育水平的变量edu18，由18岁及以上家庭成员的平均受教育年份表示。

第五，恩格尔系数还可能与家庭成员的就业面（就业的家庭成员占全部家庭成员的比例）有关。其原因比较复杂；一方面，家庭就业率较高，可能节约食品支出，因为从业者有可能在单位就餐，在某种程度上享受工作单位的食品补助。另一方面，

较高的就业率又有可能导致较多的外出就餐（因为在家里做饭更花费时间），因此导致较高的食品支出。究竟哪种因素占上风，还需要通过检验来证明。模型中设定了家庭就业面的变量emp。

第二步，设定模型的函数形式。从数据上我们可以直观地判断，恩格尔系数与人均收入之间的关系是非线性关系。因此作者分别选择了半对数函数、半对数二次函数、二次函数和三次函数模型进行估计。各函数均以恩格尔系数为被解释变量（以eng表示），半对数函数以对数人均收入lnY、以及控制变量city、family、edu18、emp、H1、H2、和L1为解释变量，称为函数（1）。半对数二次函数在函数（1）的基础上增加了lnY的平方项，见函数（2）。二次函数以人均收入及其二次项、以及各控制变量及其二次项作为解释变量，见函数（3）。三次函数在二次函数的基础上增加了各解释变量的三次项。函数（2）、（3）如下，函数（1）、函数（4）省略。

$$eng = C_1 + a_1 \ln Y + a_2 city + a_3 family + a_4 edu18 + a_5 emp + a_6 H_2 + a_7 H_1 + a_8 L_1 + a_9 (\ln Y)^2 \quad (2)$$

$$eng = C_2 + b_1 Y + b_2 city + b_3 family + b_4 edu18 + b_5 emp + b_6 H_2 + b_7 H_1 + b_8 L_1 + b_9 Y^2 + b_{10} city^2 + b_{11} family^2 + b_{12} edu18^2 + b_{13} emp^2 + b_{14} H_2^2 + b_{15} H_1^2 + b_{16} L_1^2 \quad (3)$$

第三步是对上述四个模型的估计。估计结果见表5。在初步回归分析中发现，其中二次和三次函数模型的某些变量二次项或三次项没有统计显著性，且t值很低。因此在表5的回归中已将这些项从模型中剔除。

从表5的回归结果看，尽管四个模型的调整R²不够高，但它们的绝大部分变量都有高的统计显著性，不仅证明恩格尔系数与人均收入水平之间存在十分显著的负

表 5 模型估计结果

变量	(1) 半对数函数		(2) 半对数二次函数		(3) 二次函数		(4) 三次函数	
	系数	t 值	系数	t 值	系数	t 值	系数	t 值
lnY	-0.05739	-28.66**	-0.12004	-4.63**				
lnY ²			0.00295	2.42*				
Y					-7.67E-07	-20.8**	-1.24E-06	-19.31**
Y ²					5.44E-13	13.88**	1.93E-12	12.15**
Y ³							-7.49E-19	-8.99**
city	-0.00664	-3.50**	-0.00677	-3.57**	-0.00385	-1.97*	-0.12508	-2.21*
city ²							0.05612	2.22*
city ³							-0.00774	-2.28*
edu18	-0.01116	-4.35**	-0.01066	-4.15**	-0.03194	-6.80**	-0.02741	-5.83**
edu18 ²					0.00117	2.84**	0.00098	2.39*
family	-0.01427	-6.41**	-0.01423	-6.40**	-0.01559	-6.78**	-0.01498	-6.54**
emp	-0.01585	-1.95*	-0.01350	-1.65*	-0.03781	-4.53**	-0.03164	-3.82**
H1	0.07106	11.47**	0.07078	11.43**	0.07601	11.89**	0.07543	11.89**
H2	0.02557	5.66**	0.02544	5.62**	0.02615	5.58**	0.02858	6.12**
L1	-0.03938	-6.06**	-0.03979	-6.13**	-0.03298	-4.93**	-0.03149	-4.74**
C	1.06077	49.76**	1.38627	10.19**	0.5790	37.80**	0.64580	16.57**
Adj.R ²	0.2463		0.2472		0.1973		0.2130	
Obser.	4195				4195		4195	

注：t 值标有 ' 号表示在 10% 水平显著，* 号表示在 5% 水平显著，** 表示在 1% 水平显著。

相关关系，同时也证明恩格尔系数还受家庭成员文化程度、家庭人口数、家庭成员就业面、城市规模、以及地域特征的影响。其中模型 (2) 的调整 R² 最高。计算表明，模型 (1) 和模型 (2) 的结果非常接近，而且在中、低收入区间的模拟结果与统计数据比较接近，而模型 (3) 和模型 (4) 的模拟结果在各个收入区间都与统计数据有较大差异，而且在收入水平很高的情况下不再保持恩格尔系数单调下降，这与事实相左。因此在下面的分析中将采用模型 (2) 的结果。

图 1 是用函数 (1) ~ 函数 (3) 模拟出的收入水平与恩格尔系数之间的关系曲线，显示了函数 (1) 和函数 (2) 之间很好的相似性。图中的纵轴表示恩格尔系数，横轴表示人均收入水平 (元)。

第四步，为了最后实现使用回归得到的各变量影响系数，求解与不同的恩格尔系数相对应的全国城镇居民收入水平，还

需要确定各影响变量在全国平均意义上的赋值。

根据 2007 年统计数据，城镇居民在特大城市、大城市、中小城市、县城和建制镇 (在模型中分别取值为 1、2、3、4) 的分布比例，大致为 21%、25%、33%、21%。加权平均取值为 2.5。但我们知道不同收入组别的人群在不同城市的分布是有区别的，高收入居民较多集中在特大城市和大城市，而低收入居民则更多集中在中小城市和小城镇。因此根据数据分析，作者把按人均收入从低到高排序的居民组合的城市规模取值，确定在 3.3~1.3 之间平滑变动。

关于城镇居民的人均文化程度 (18 周岁及以上)，模型中分别用从 1 到 5 的赋值来表示小学及以下、初中、高中和中专中职、大学专科和本科、硕士和博士学历。估算全国城镇平均取值在 3 左右。但文化程度在不同收入人群中的分布也是有差异的，作者把从最低收入到最高收入居民组合的

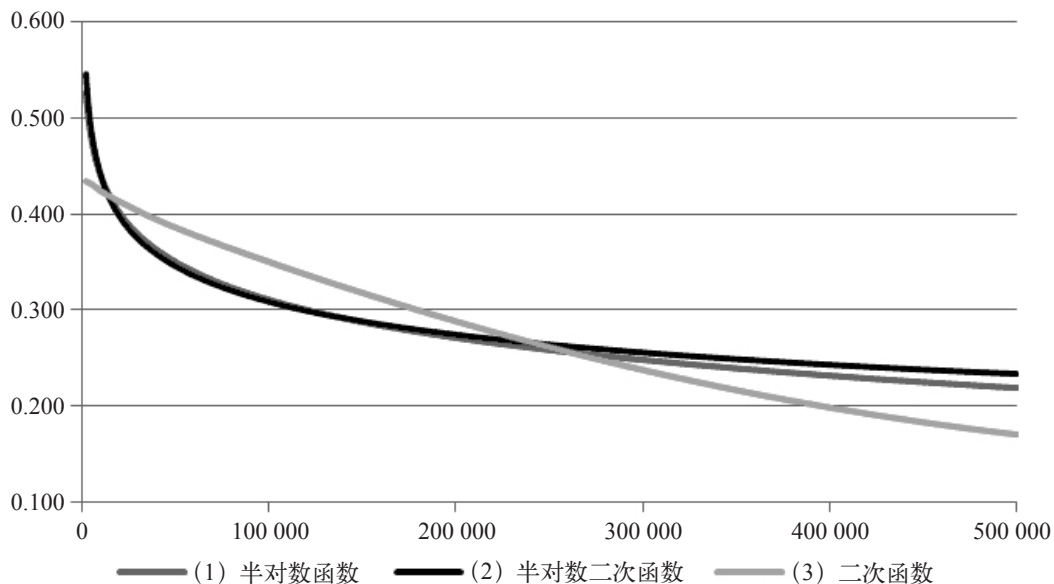


图1 函数(1)~函数(3)的模拟曲线

平均文化程度取值确定在 2.6~3.8 之间平滑变动。

关于城镇居民家庭成员的就业面,根据统计数据,全国平均大致为 0.5,但也存在不同收入分组的差异,从低到高取值在 0.38~0.62 之间变动。

关于城镇居民家庭人口数,统计显示全国平均为 2.9 人,但低收入居民家庭的平均规模相对较大,高收入家庭较小,变动范围在 3.3~2.6 之间。

最后,关于不同地区之间城镇居民的饮食习惯差异,模型中按在相同条件下恩格尔系数最高、较高、普通、较低的省份分为四组,它们的虚拟变量系数在正 0.071 到负 0.039 之间。全国平均按 0.01 取值。

在完成上述这些控制变量的赋值过程后,就可以根据模型回归得到的参数和各影响因素的取值,即在考虑其他影响因素对恩格尔系数影响的条件下,求解不同恩格尔系数所对应的城镇居民收入水平。这一结果在下部分报告。

三、求解城镇居民真实收入

1. 城镇居民分组收入推算结果

把上文模型(2)中各参数的估计结果(见表5)、统计样本中的分组恩格尔系数、以及各控制变量的全国平均赋值,带入模型(2),就可以倒推出对应于不同恩格尔系数的人均收入水平。表6将这些根据模型分析推算出的结果,与分组统计数据、以及使用分组分析法得到的结果同时列出,以进行对比。

从表6的数据看,用分组比较恩格尔系数和通过模型分析取得恩格尔系数与收入水平的关系这两种方法,得到的结果总体上是比较接近的。它们对最低收入、低收入和中低收入这三组收入水平的推算,虽然都不同程度上高于统计收入,但总体上差异不太大(其中模型分析的结果更为接近)。从中等收入组往上,推算收入与统计收入的差距明显扩大。差距最大的是最高收入组,其中分组比较法的结果是 16.4

表 6 基于恩格尔系数法求解的城镇分组人均收入（元）

分组	恩格尔系数	统计收入	推算收入： (1) 分组比较法	推算收入： (2) 模型分析法
最低收入	0.481	4 754	5 685	5 350
低收入	0.459	7 363	8 646	7 430
中低收入	0.429	10 196	13 392	11 970
中等收入	0.404	13 984	20 941	17 900
中高收入	0.379	19 254	29 910	27 560
高收入	0.340	26 250	47 772	54 900
最高收入	0.292	43 614	164 034	139 000
全部城镇居民	0.379	16 885	35 462	32 154
剩余样本	0.224		658 811	625 000

注 1：全部城镇居民的平均统计收入 16 885 元是按各组统计收入加权平均计算的，而统计局公布的平均数是 15 781 元。

注 2：推算的全部城镇居民收入没有将剩余样本包括在内。

万元，模型分析法的结果是 13.9 万元，分别是统计收入的 3.76 倍和 3.19 倍。这两个结果所反映的收入分布基本格局是一致的，而且与 2007 年报告的结果（2005 年数据）相比，尽管数据来源不同，但具有较高的相似性。

表 6 显示，主要由于高收入和最高收入组的收入差异，按模型分析法推算得到的全国城镇人均收入平均值为 32 154 元，而不是按统计局的分组数据加权平均的 16 885 元，或统计局公布的 15 781 元。推算数与统计数几乎相差了一倍。

在表 7 中，我们也提供了 2005 年和 2008 年两次研究的推算收入（2008 年采用模型分析法的结果）与统计收入之比。我们发现，2008 年（次于最高收入组的）高收入组推算收入相对于统计收入差异明显变大，但总体而言差异最大的还是最高收入组。我们把这些没有包括在统计数据中的居民收入称为隐性收入，在表中最后一列计算了各组的隐性收入占隐性收入总量的比重。很明显，隐性收入主要集中在最高收入家庭。

以上两种推算结果中，分组比较法可能误差会大一些，这是由两个原因导致的。

第一，分组分析法只能通过计算各组的平均恩格尔系数和平均收入水平，得到与某一特定的恩格尔系数对应的人均收入。但无法排除收入以外的因素对恩格尔系数的影响。但这些因素对调查样本的影响和对统计样本的影响有可能是不同的，因此推算出的人均收入水平有可能不够准确。而模型分析法则控制了收入以外因素对恩格尔系数的影响，并计算了这些因素的全国平均影响程度。所得到的结果，是在这些影响之下，对应于某个恩格尔系数的人均收入水平，因此也应该更真实、更准确。

第二，分组分析法通过根据恩格尔系数重新分组的方法，在总体上避免了调查样本分布与统计样本分布不一致而对总体的收入分布产生错误影响。但这一方法对各组平均收入的计算，仍与组内的样本分布有关。而调查样本的组内分布有可能与统计样本的组内分布有所不同。尽管这对总体分布的影响不会很大，但不能排除产生一定误差的可能性。举例说，全国城镇居民最高收入组和最低收入组的组内收入水平很可能不是均等分布的，而是分别向低端和高端倾斜。但调查样本的组内分布，不一定符合这个特点。而模型分析法则

表 7 2005 和 2008 年推算收入与统计收入之比

分组	推算收入 / 统计收入 2005	推算收入 / 统计收入 2008	隐性收入的分布 2008
最低收入	99.1%	112.5%	0.4%
低收入	101.8%	100.9%	0.0%
中低收入	106.9%	117.4%	2.3%
中等收入	114.0%	128.0%	5.1%
中高收入	130.6%	143.1%	10.9%
高收入	138.7%	209.1%	18.8%
最高收入	337.6%	318.7%	62.5%
全部城镇居民	177.7%	194.4%	100.0%

注：推算收入与统计收入之比以各组的统计收入为 100%，隐性收入分布以 2008 年隐性收入总量为 100%。

根据收入和其他因素对恩格尔系数的影响系数，直接从恩格尔系数推算出对应的收入水平，准确程度应该高于分组比较法。

基于以上两点考虑，在以下的计算中，我们将根据模型分析法（模型 2）的估计结果进行。

图 2 更直观地给出了两种方法得到的推算收入与统计收入的比较。可以看到，两种方法具有一致性，但也有些差异。其中按模型分析法得到的高收入组人均收入明显高于按分组比较法得到的人均收入，而模型分析法得到的最高收入组人均收入则低于按分组比较法得到的收入。

此外，为了进行对比，我们把 2007 年研究报告提供的 2005 年推算数据与统计数据比较图（见图 3，依据分组比较法）也放在图 2 后面。这两张图很相似，显示了两次调查和研究结论的基本一致性。

还需要指出，在 2007 年报告中，作者判断根据调查样本推算的分组收入水平可能存在一定正方向误差，因此对推算结果做了一定程度的下调。从表 6 可见，下调后的 2005 年推算数与未进行调整处理的 2008 年推算数与统计数的比例比较一致，只有（次）高收入组例外。这看来说明，当时的下调是合理的，而且当时的误差可能主要来自分组比较法自身，因此这

次使用模型分析法进行推算，就消除了这一误差。

但要再次强调的是，这里的推算只解决了统计收入低报的问题，并没有解决统计数据遗漏高收入样本的问题，因此实际上这里对最高收入居民人均收入的推算仍然可能在一定程度上偏低。

这种情况也反映在我们的调查样本中。按分组比较法对调查样本分组之后，有少数高收入样本没有能够列入这 7 个收入组。这是因为这些样本家庭的恩格尔系数更低；加入这些样本后，最高收入组的平均恩格尔系数就会低于统计数据中最高收入组的恩格尔系数。在这次的有效样本中，有 76 个高收入样本被排除在外，他们的人均可支配收入都在 40 万元以上，平均收入为 66 万元。他们的平均恩格尔系数只有 0.224，而国家统计局城镇住户样本中最高收入组的平均恩格尔系数为 0.292。这似乎说明，在统计部门的城镇住户样本中，可能很少包括人均收入 40 万元以上的住户。因此严格地说，这里的最高收入组并不是收入最高的居民组，因为还有一些更高收入的居民没有包括在内。

但我们无法用这些调查样本来反映高收入居民的统计遗漏的情况，这是因为我们无法知道被遗漏的高收入人群在全部城

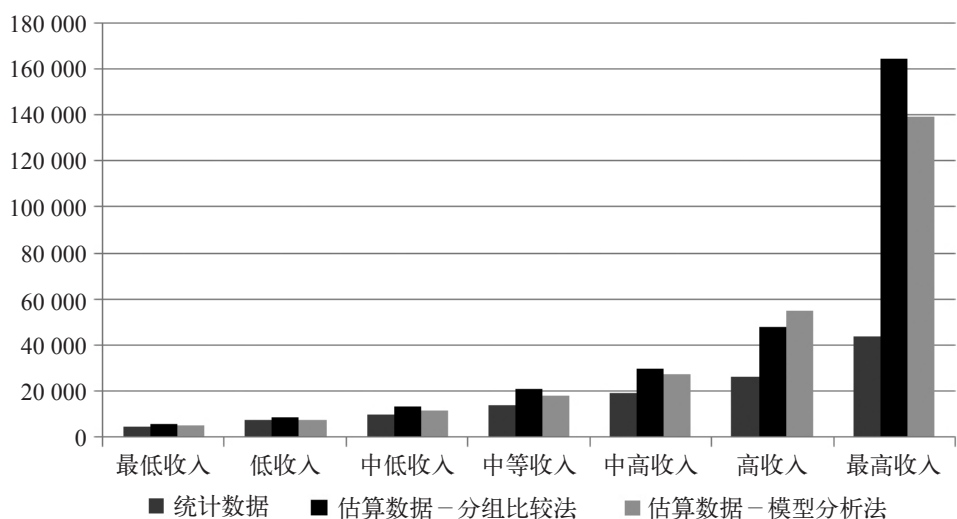


图 2 2008 城镇居民分组人均收入：推算数据与统计数据比较

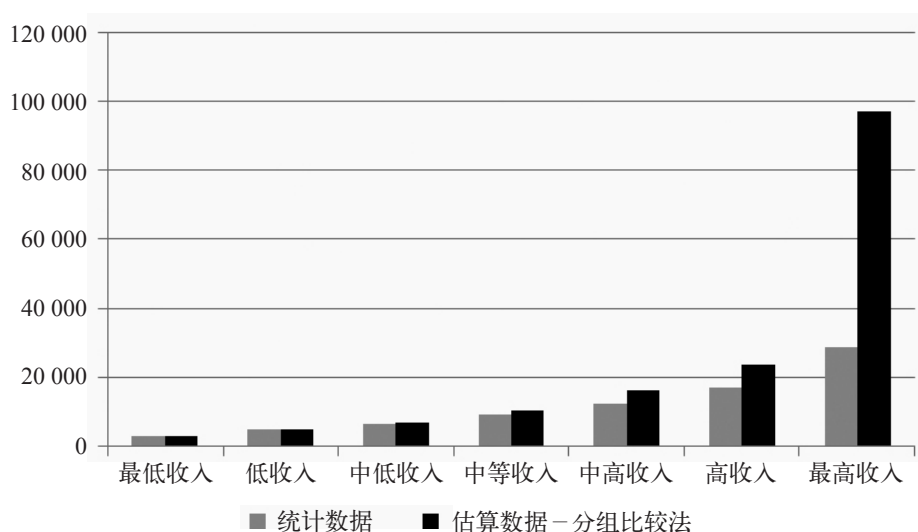


图 3 2005 城镇居民分组人均收入：推算数据与统计数据比较

镇居民家庭中占有多大比重，也不能用调查数据中这些没有包括在内的高收入样本的分布来代表全国被遗漏的高收入样本分布，因为调查样本的分布状况与全国更高收入居民的分布状况未必一致。而且，在我们的调查样本中也仍然存在遗漏。在这次调查样本中，收入最高家庭的人均可支配收入为 176 万元，比这一收入更高的样本还是没有包括在内。

这也从一个侧面说明，在统计调查中，更高收入的居民被遗漏了。而我们在这里所采取的两种恩格尔系数推算方法，都只能纠正已有的统计样本中对收入的低报，而无法纠正因遗漏高收入样本而导致的收入统计数据偏低问题。因此在某种程度上，我们对“最高收入组”人均收入的推算仍然是偏低的，不过其可靠程度还是远远高于目前的统计数据。

2. 隐性收入和居民收入差距有多大？

从表 6 和表 7 可以看到，2008 年占 10% 的城镇居民最高收入家庭，人均可支配收入应为 13.9 万元，而不是统计显示的 4.4 万元。两者相差 3.2 倍。最高收入家庭这些没有包括在统计数据中的隐性收入占城镇居民隐性收入总量的 63%。再加上高收入组，城镇 20% 高收入居民的隐性收入，占全部城镇居民隐性收入总量的 80% 以上。

据此推算城镇居民收入分配差距，按城镇居民家庭 10% 分组，2008 年城镇最高收入家庭与最低收入家庭的实际人均收入分别是 5 350 元和 13.9 万元，差距是 26 倍（上一份报告推算的 2005 年差距是 31 倍），而按统计数据计算只有 9 倍。（这次推算的收入差距小于上次的推算，主要是因为这次推算的最低收入组的人均收入比统计收入高了 12%。而这又与如下情况有关：在调查样本中有少数低收入样本没有包括在内，因为他们在正常情况下不属于低收入阶层，他们的低收入是暂时性的经营亏损造成的。但不包括这些样本，在一定程度上提高了最低收入组的推算收入。估计这种情况在统计样本中并没有剔除。这可能是最低收入组推算收入高于统计收入的一个主要原因。）

用城镇最高收入 20% 的家庭和农村最低收入 20% 的家庭来近似地代表全国最高和最低收入 10% 家庭³，那么全国最高 10% 家庭的人均收入是 9.7 万元，而最低 10% 家庭的人均收入是 1 500 元，两者在 2008 年相差 65 倍（2005 年的推算为 55 倍），而

3 2008 年全国城镇人口略少于农村人口；但城镇居民每户平均人口数也少于农村居民，因此城乡户数相差不多。实际上城镇 20% 最高收入家庭占全国家庭数的比重为 10.7%，农村 20% 最低收入家庭占全国家庭数的比重为 9.3%，可以分别近似地代表 10% 的最高和最低收入家庭。

按统计数据计算只有 23 倍。

相信这种隐性收入绝大部分发生在城镇。近似假定农村居民收入统计数据是真实的，使用城镇居民人均推算收入数据来代替统计收入数据，与农村人均纯收入统计数据（分组加权平均结果）分别乘以城乡人口数，则可以近似得到 2008 年全国居民可支配收入总额为 23.2 万亿元⁴，但如果按城乡住户收入统计数据计算，2008 年全国居民收入总额只有 14.0 万亿元。两者相比，隐性收入高达 9.26 万亿元，比作者在 2007 年报告中推算的 2005 年隐性收入（4.85 万亿元）扩大了近一倍（上升 91%）。而同期，名义 GDP 只增长了 71.4%。这说明隐性收入仍在以快于 GDP 增长的速度迅速扩大。

在各自包括了隐性收入后，2008 年比 2005 年全国居民可支配收入总额增长了 69.3%，与名义 GDP 的增长幅度接近于同步。而按官方统计数据计算（不包括隐性收入），全国居民可支配收入 2008 年只比 2005 年上升了 57.4%，远远滞后于 GDP 的增长。因此它在 GDP 中所占比重从 2005 年的 48.4% 下降到 2008 年的 44.5%，下降了 3.9 个百分点。显然，隐性收入的增长

4 有几点需要说明：第一，农村“纯收入”与“可支配收入”在统计口径上有差异，我们在这里忽略这种差异，将其近似地视为可支配收入。第二，国家统计局根据住户调查公布的城、乡居民人均收入，与所公布的分组人均收入之间，似有不一致的地方。公布的城镇人均可支配收入比按分组人均可支配收入加权平均得到的结果低 5.5% 到 6.5%，公布的农村人均纯收入比分组加权平均结果低 8% 左右（见《中国统计年鉴》“人民生活”部分）。目前还不知道这种不一致是如何产生的。本报告的计算以分组加权平均计算结果为准。第三，依据住户统计调查的居民人均收入和全国城乡人口数计算得到的全国居民可支配收入（统计数据），与国家统计局在“资金流量表”中提供的全国居民可支配收入，有统计口径的差别。在本表中，该数据来自前一个来源。而在本报告下一部分关于国民收入分配格局的讨论中，所使用的居民可支配收入统计数据，则来自后一个来源。请注意这一区别。

表 8 2005~2008 年居民收入变动

	2005 年	2008 年	增长幅度
城镇居民人均可支配收入 (元, 统计数据) 1	11 100	16 885	52.1%
城镇居民人均可支配收入 (元, 推算数据)	19 730	32 154	63.0%
城镇人口 (亿人)	5.6212	6.0667	7.9%
农村人均纯收入 (元, 统计数据) 1	3 537	5 171	46.2%
农村人口 (亿人)	7.4544	7.2135	-3.2%
全国居民可支配收入 (亿元, 统计数据)	88 767	139739	57.4%
全国居民可支配收入 (亿元, 推算数据)	137 275	232 372	69.3%
推算隐性收入 (亿元)	48 508	92 633	91.3%
GDP 总量 (亿元)	183 217	314 045	71.4%
全国居民可支配收入 (统计数据) 占 GDP 比重	48.4%	44.5%	-3.9 (百分点)
全国居民可支配收入 (推算数据) 占 GDP 比重	74.9%	74.0%	-0.9 (百分点)
隐性收入占 GDP 比重	26.5%	29.5%	3.0 (百分点)

注 1: 城镇居民人均可支配收入和农村居民人均纯收入各有两组不同的统计数据。这里给出的是按统计局公布的分组收入统计进行加权平均计算得到的。而统计局公布的 2005 和 2008 年全部城镇居民人均收入分别为 10493 元和 15781 元, 比加权平均数分别低 5.5% 和 6.5%。公布的农村人均纯收入数比分组加权平均数低得更多 (8% 左右)。原因不明。

注 2: 全国居民可支配收入 (统计数据) 是根据统计公布的城镇人均可支配收入和农村人均纯收入 (表头以下第一行和第四行数字) 各自与城乡统计人口的乘积求和近似得到的。

注 3: 隐性收入占 GDP 的比重, 是在 GDP 不进行相应调整的情况下计算得出的。

速度远高于正常收入的增长速度。分别按统计数据和推算数据计算的城镇人均可支配收入、全国居民可支配收入总量及其占 GDP 比重在 2005~2008 年期间的增长幅度、以及名义 GDP 的增长幅度, 见表 8。

3. 从其他途径验证隐性收入

要检验以上推算是否合理, 除了方法问题需要探讨, 更重要的是要看推算结果与事实的吻合程度。长期以来, 在居民收入统计数据、宏观经济数据、住房、汽车和其他奢侈品销售方面, 存在许多明显的矛盾或者不可解释之处。而如果把上面推算的隐性收入包括到居民收入中来, 这些矛盾就基本上消失了。作者认为, 这应该是检验上述推算是否合理的一个主要方法。如果存在上文所估计的巨额隐性收入, 它也必然会在国民经济的宏观层面有一系列反映。下面我们试图通过几个不同途径来进行验证。

(1) 居民储蓄与消费总额

2008 年, 根据城乡居民收入统计数据和城乡人口数推算, 全国城乡居民储蓄总额为 3.55 万亿元 (这里的储蓄定义为可支配收入与消费之间的差额)。这是居民的真实储蓄规模吗? 居民储蓄以不同形式存在, 或直接、间接地用于各种投资。因此我们把宏观经济层面反映出的各种关于居民净投资的信息整理加总 (扣除借贷), 可以近似推算出居民总储蓄的规模。

这一年, 全国金融机构的城乡居民人民币储蓄存款余额增加了 4.54 万亿元 (我们知道金融机构的存款数据应该是较为准确的), 说明仅居民用于金融机构存款的储蓄部分就超过了按收入统计推算的储蓄总额 1 万亿元。

同一年, 商品住宅销售 2.12 万亿元 (不包括二手房), 扣除个人住房贷款增长额 3 000 亿元, 用去居民储蓄 1.82 万亿元。

同一年, 农村个人建房投资支出 3711 亿元。调查发现城镇居民个人建房也相当普遍, 在城镇居民自有住房中大约占到

15%~16%。虽然发生率低于农村，但单位投资额大幅度高于农村。估计总额不会小于农村自建房投资。而且这类投资使用贷款的情况很少，主要来自个人储蓄。据此估计，城乡居民自建房投资使用了个人储蓄7000亿元。

2008年，“规模以上”私营工业企业所有者权益净增加1.09万亿元。这些资金实际上是由企业所有者的收入与储蓄形成。服务业私营企业的个人出资的投资，估计不低于工业中的私人投资，应超过1万亿元。第二、三产业“规模以下”私营企业、股份制企业、和2874万户有证照的个体经营户个人出资的私人投资，以及建筑业的这类私人投资，粗略估计在5千亿~1万亿元之间。几项合计，第二、三产业来自个人储蓄的私人投资估计在2.5~3万亿元之间。

2008年，A股市场股票流通市值比2007年缩水50.9%，而上证综指和深证综指分别下跌65.4%和62.4%，跌幅均大于流通市值缩水幅度；两相抵消，估计当年仍然净吸纳投资1.35万亿元。国债和企业债券发行额共1.7万亿元。如果按这些投资的三分之一来自个人资金计算，股票（A股）和债券共吸纳私人储蓄1万亿元。

除此之外，居民对期货、黄金、外汇等金融产品的投资，加上居民手持现金和国外存款等项的增加，粗略估计5000亿元（也许低估了）。⁵

以上各项合计，2008年全国居民储蓄总额至少应为11万亿~11.5万亿元，而不

是按收入统计数据计算的3.5万亿元。这说明该年至少有7.5万亿~8万亿元的居民隐性收入没有包括在收入统计中。此外，根据住户消费统计计算的居民消费支出总额为9.46万亿元。尽管居民消费统计与居民收入统计相比遗漏较小，但调查发现还是有相当幅度的遗漏。如果按遗漏20%的保守估计，则意味着居民消费也被低估了近2万亿元。与被低估的储蓄相加，2008年实际的居民隐性收入总额至少应在9.5~10万亿元之间。这与上文根据模型分析法推算的9.3万亿元隐性收入是基本吻合的。

以上这种情况，并不仅仅发生在2008年。对近些年的宏观数据进行分析，都可以得到类似的结果。而且通过宏观数据分析所显示的隐性收入总量，有不断扩大的趋势。

（2）房价收入比

根据国外经验，通常商品住宅价格是居民家庭年收入的3~5倍，是可承受的价格，否则住房市场不会有良好的发展。而按收入统计计算，我国城镇居民的房价收入比一直在10倍左右，住房市场却持续火爆，这始终是国内外研究者面对的一个不解之谜。2008年，商品住宅销售2.1万亿元，2009年猛增到3.8万亿元。如果按统计收入算，扣除贷款，这也远远超过了城镇10%最高收入户的全部收入总额。这是一个无法用统计数据解释的现象。

根据我国2008年城镇居民收入统计，人均可支配收入1.578万元，平均家庭规模2.91人，合家庭年收入4.59万元，而同一年商品住宅销售2.12万亿元，共5.93亿平米，按每套平均110平米计，估算售出515万套，每套均价41.1万元，是统计显示的城镇居民平均家庭年收入的9倍。而按同样方法计算，2007和2009年都超过了10倍。

⁵ 各项推算依据的基础数据来自国家统计局关于城乡居民收支、人口、金融机构存贷款、私营工业企业所有者权益、各类固定资产投资、房地产市场、证券市场、股票交易等统计数据，见《中国统计年鉴2009》相关章节以及国家统计局《第二次全国经济普查主要数据公报》。个人住房贷款根据银行金融市场司副司长霍颖励在国新办新闻发布会上提供的情况（2009年1月，新华网）。

这远远超出了城镇居民的承受能力。还应当考虑到，一部分家庭是从炒房者手中购买的二手房，所以他们实际付出的价格超过原来的一手房价格，这没有包括在新房上市的价格中。但这没有妨碍住房市场的持续火爆。这些现象说明城镇居民平均收入可能被低估了一倍以上。

按照统计显示，占城镇家庭数 20% 的高收入家庭（最高收入 10% 和次高收入 10% 家庭合计）2008 年人均收入为 3.5 万元，平均家庭年收入 8.9 万元。按社会平均的房价计算，他们的房价收入比大致在 5 左右。这意味着即使高收入家庭，在 2008 年也仅仅是勉强达到有能力购买商品房的范围边缘。而事实说明这是不真实的。

1990~2009 年累计，商品住宅累计销售超过 4 600 万套，超过了城镇高收入家庭的总数（约 4 100 万户）。调查数据证明，实际上相当一部分高收入家庭不需要商品房（有公房、公司提供的住房、或“房改”房），而对已经购买商品房的高收入居民来说，他们的商品房房价标准远远高于社会平均的房价标准，而且大约有三分之一的高收入家庭拥有第二套、第三套或更多的住宅。这些情况说明，高收入居民的实际收入水平远不止被低估了一倍。

（3）家庭汽车与收入

根据我国的私人汽车拥有量统计，2008 年我国私人拥有的小型和微型汽车为 2814 万辆，如果按 90% 属于城镇居民计算，城镇家庭的家用汽车普及率已达 12.1%（每百户拥有家庭汽车 12.1 辆），说明占城镇住户 20% 的高收入家庭，除去一部分暂时没有买车意愿的家庭外，汽车已基本普及。实际上一些中等收入家庭也已经成为有车阶层。汽车拥有量统计有牌照发放和年检数为依据，应该是准确的。而根据城镇住户

抽样统计收入，每百户城镇家庭仅拥有 8.8 辆家用汽车。比前者少了近三分之一。这就从另一个角度说明住户抽样调查遗漏了相当数量的高收入家庭样本。而且更明显的矛盾是，有车家庭的统计收入水平根本不足以拥有汽车。

如果按家用汽车平均价格 10 万元、年相关费用 2 万元（燃油、维修保养、保险、年检、停车费、过路费等）计，通常至少需要 20 万元以上的家庭年收入来支持。根据统计数据，20% 高收入家庭的年可支配收入平均只有 8.9 万元，下限 6 万元，远远达不到普及家庭汽车的程度。而按本报告前面的推算结果，这 20% 城镇家庭的实际家庭年收入平均约为 24 万元，下限在 12 万元左右，大部分具备了购车能力。这再次说明高收入居民的统计收入被严重低估了，本报告前面对他们收入的推算则大体上与其购车能力相吻合。

（4）其他

除去以上列举的城镇居民住房、汽车等情况，城镇居民在其他许多方面的支出也是统计收入不能解释的。其中一个例子是国外消费。2008 年，全国因私出境人数超过 4 000 万人次，相当于 20% 的城镇高收入家庭每户一人次。如果假定 1/3 是商务原因而予扣除，其余为旅游、探亲、留学等，按人均花费 3 万元人民币计算，全国这方面的居民支出达到 8 000 亿元。这部分没有包括在国内社会消费品零售总额中的支出，同样在收入统计中无法找到来源。

此外据报道，中国每年有几百亿元私人资金存入国外银行；另据公安部官员估计，中国每年赌资外流至少在几百亿到上千亿元之间⁶。

6 中国广播网（2010 年 7 月 9 日）：“网络赌博久赌必输，我国每年因赌博资金外流数百亿”。

四、灰色收入及其来源

1. 巨额隐性收入说明什么？

在国民收入分配中，出现如此巨额的隐性收入，而且分布高度集中，显然是不能简单地用统计遗漏来解释的。根据本报告前一部分表7的推算，中、低收入居民也在一定程度上存在低报收入的倾向，幅度大致在0%到30%的范围内。这些仍然可以看做正常收入的统计遗漏，可能主要与调查对象的某些心理状态或社会环境有关。但中高收入阶层以上居民的隐性收入大幅度增加，特别是最高收入居民的实际收入高达统计收入的几倍，说明存在深刻的制度性原因。这些隐性收入中的主要部分，也不再是正常收入的统计遗漏，而只能用灰色收入来解释。

关于灰色收入的概念，近来引起了不少争议。有一种说法是：灰色收入就是非法收入，只能“取缔”，不能“规范”；所谓“规范”，有将非法收入合法化之嫌。另一种说法则强调这一概念本身“不科学”、没有明确定义，因此反对使用这个概念。但是，之所以产生“灰色收入”这个概念，就是因为现实生活中存在大量无法明确界定、不能清楚地判断其合法或非法的收入。即使“取缔”了灰色收入的概念，也决不等于取缔了灰色收入，实际上等于对如此重大的国民收入分配问题采取回避的态度，无益于问题的解决。这可能也说明，解决灰色收入问题会深深触及某些人的既得利益，必然会遇到很大的阻力。

我们使用灰色收入这个概念，通常是指两种情况。第一种情况，现实生活中有些收入，由于在制度上或法律上没有明确界定，因此处在合法与非法的中间地带。用

一个例子来说明，民间举行婚礼，收受亲朋好友的礼物礼金，既是民风民俗，也是法律所允许的。即便发生在党政官员的亲属、子女身上，如果在合理限度内，也未必能够简单禁止。但有些官员借子女、亲属婚礼的名义收受重金，聚敛财富，甚至以数十万、数百万计，实质上是索贿受贿。这就需要法律法规对官员亲属收受礼品礼金的情况，在情节和金额上划定一个合法的界限，并建立可行的监督办法。越过这个界限，必须严格禁止，以非法收入论处。但在法律法规没有明确规定的情况下，这类收入也只能算做灰色收入。

又如，企事业单位或政府机关在工资之外，给雇员提供某些奖励和福利待遇，一般而言是正常、合法的行为。但有些政府机关或具有垄断地位的国有企业，慷公共资金之慨，以种种名义给官员、企业管理层或职工提供远高于市场常规的福利或额外给付，可能还逃避了纳税，这实际上侵害了公众利益。这其中常常出现的一类情况，是违规违纪，但并不违法。在法律法规对此没有做出明确界定的情况下，这类收入也只有以灰色收入来形容。

当然，也不排除有些正当收入，由于法律没有明确赋予其合法地位，因而变成了灰色收入。

第二种情况，是某些怀疑为非法的收入，在来源不明而不能认定其非法的情况下，也只能作为灰色收入来看待。例如频繁出现的通过内幕交易、虚假拍卖获得的地产收益，通过内线消息、散布虚假信息、操纵市场而在股票市场、期货市场上获得的暴利，以及政府官员以权谋私，通过钱权交易获得的利益等等。但如果仅仅是公众或有关人士怀疑存在上述情节，而拿不到确凿证据，也无法作为非法收入来对待。

由于已经曝光的非法收入毕竟是少数情况，因此我们这里所讨论的灰色收入，除了包括对其合法性没有明确界定的收入外，也涵盖了大部分事实上的非法收入。

前一种情况，说明法律和制度不完善，使收入分配的许多环节存在模糊区间和漏洞；后一种情况说明在立法和执法、监督之间存在脱节，对一些经济活动的重要领域、对公务人员的行为缺乏有效监管；两者都导致国民收入分配发生扭曲。当一个社会出现大量灰色收入的时候，自然说明存在严重的制度不健全或漏洞。这是一个危险的信号，说明国民收入分配脱离了常轨。因此，要解决灰色收入和收入差距过大的问题，关键不在于如何给灰色收入下定义；缺乏针对性地空谈“取缔”也不解决实质性问题。根本出路，在于推进制度改革，理顺国民收入分配体制，建立和健全对收入分配、特别是涉及公共资金、公共资源及其收益的立法与监督机制，使国民收入分配得到有效的规范。

2. 灰色收入的来源

在市场经济体制下，国民收入分配主要依据生产要素的报酬进行，即区分为劳动、人力资本、资本、土地以及其他自然资源的回报。市场经济体制并不能自发保证收入分配公平合理，但一般而言有利于经济效率的发挥；因为在竞争条件下，回报率高的生产要素，自然得到较高的报酬，因而通过收入分配的激励促进了效率提高；而且生产要素的自由流动也保证了资源配置的优化。

但任何社会也都需要政府参与收入分配。因为社会必须以一部分资源或产出用于公共基础设施投资和维护、进行宏观调控、促进科技开发、保证国防和社会安全

等。由于市场经济不能保证收入分配的公平性，也需要通过政府来建立社会保障、公共服务和转移支付体系，以改善收入分配。这些都不是市场机制能够自发完成的。发达国家在整个 20 世纪，经历了这方面的一系列制度变革，减少了收入分配的不公平，也减少了社会冲突，从过去弱肉强食的传统资本主义演进到比较人性化的资本主义。这些改变增进了社会和谐与安定，并间接地保障了经济的持续发展。

这些情况，比经济学理论所描述的抽象的收入分配更复杂。20 世纪经济学的发展，尽管出现了有关经济外部性的公共经济学，有关宏观平衡和宏观管理的凯恩斯和后凯恩斯主义经济理论以及有关选举、公共政策与公众利益的公共选择理论等，各自涉及收入分配问题的某些方面，但总体而言这些理论进展仍然落后于社会实践，能够提供的指导是相对有限的。

其中一个重要领域，是如何保证公共资源及其收益的合理使用和分配。掌握这些资源的权力机构或人员，在利益的驱使下，很容易染指其使用或分配过程，侵占这些资源，或利用这些公共资源获取个人利益。在这种情况下，本来用于保障社会公平的公共资源，可能成为少数特权阶层的猎物，反而加剧收入分配的不公平。

广义而言，行政权力本身也是一种公共资源。因为掌握权力的人员可以利用行政权力来干预收入分配、资源配置和各种社会经济活动，通过滥用权力侵害其他社会成员的利益。

因此所谓灰色收入的一个重要来源，就是围绕权力产生的贪污、渎职、寻租等腐败行为。这就不仅需要公共资源的使用和权力的行使做出严格的法律界定，并建立一系列相应的规章制度，而且需要对

公共资源的整个使用和分配过程进行严格的监督。靠政府进行自我监督是远远不够的，关键的一点是使政府管理透明化，从而使社会公众能够对政府实行监督。

在当前的情况下，灰色收入往往与下列现象密切相关：

(1) 钱权交易，以权谋私

根据2006年一项涵盖全国四千家的调查，对于“贵企业去年用于政府和监管部门人员的非正式支付有多少？”这一问题，只有19.8%的企业负责人回答“没有”；回答“有一点”、“比较多”、和“非常多”的，总共占到了80.2%，其中回答“比较多”和“非常多”的占了18.1%。这一情况说明了权力机构的腐败问题已经严重到了何种程度。按行业分布来看，凡涉及资源、具有垄断性、和由于行业特点而涉及较多行政监管的行业，情况都更严重。其中回答非正式支付“比较多”和“非常多”的比例，采掘业占到了35.2%，电力和煤气生产供应业占24.3%，房地产业23%，化学原料和化学制品制造业24.2%，都明显高于平均水平⁷。

直接侵占公共资源的情况也相当普遍。根据国家审计署关于2009年度中央预算执行和其他财政收支的审计报告，抽查了56个中央部门已报销的29363张可疑发票，发现虚假发票5170张，列支资金1.42亿元。有一个情况可以为此做出注解，那就是卖假发票已经成了一个颇具规模的行业，公然兜售假发票的人员和广告不仅在城市街头随处可见，此类广告也充斥手机短信和电子邮件。从这些情况看，审计署已经查出的1.42亿元，不过是冰山的小小一角。公共资金正在以可观的规模，无声无息地流向与权力有关的人群。

⁷ 中国企业家调查系统，2006：《2006·企业经营问卷跟踪调查报告》，2006年11月发布。

另一个反映灰色收入膨胀趋势的“新兴产业”，是礼品收购业。城市街头正在越来越多地出现折价收购高档烟酒、贵重药材、金银饰品、以及礼券和购物卡的生意人。消费者大量地高价买进、低价卖出奢侈品或购物卡，当然是非理性行为。这种怪现象只有一个解释，即有很多人接受了大量不需要消费的贵重礼品，希望把它们转换成现金。而之所以有人要赠送如此多的贵重礼品，只是因为直接送钱不够方便或风险较大。送礼已经从一种维系友谊和社会交往的行为变成了钱权交易的商业行为。

还有一种特殊商品正在出现，就是政府官员的职位。这在有些地方，几乎到了明码标价的程度，卖官鬻爵成风，腐败程度令人震惊。

(2) 公共投资与腐败

公共投资领域是导致收入分配扭曲的一个重点领域。最近的一个例子是京沪高速铁路和西气东输二线工程西段，在审计中发现超进度计价和多计工程款8.15亿元，建设成本中列支其他费用和用虚假发票报销等问题17.94亿元。

借公共投资贪污公款并不仅仅发生在中央投资项目，在地方项目中更是频繁发生。作者在某地调查中采访的一位当地居民，对当地地方政府投资项目的说法非常形象：“上届政府刚建完，这届政府就翻修，要不然就重建。建了拆，拆了建；不拆不建从哪弄钱呢？投资1000万，他就敢把500万揣起来。”

一个时期以来，各地耗资巨大、大而无当、大而无用的形象工程、政绩工程层出不穷。深究起来，有些地方官员有如此强烈的投资冲动，恐怕不单纯是用追求“形象”和“政绩”能够解释的，而是背后有巨大的利益驱动。

在投资中直接“把钱揣起来”和用假发票报销毕竟是有风险的。更常见的做法，是将工程发包给“自己人”，再从承包方得到回报。据国家审计署发现，西气东输二线工程西段在截至2009年5月签订的36亿元施工合同中，有近80%未按规定招标，27亿元合同招标中存在人为拆分标段、违规确定中标人等问题。其实，更为隐蔽的做法也已经屡见不鲜；即工程招标过程照章进行，但通过招标条件设定和投标者资质筛选等过程中的“窍门”，保证事先安排好的“理想”投标人中标。

(3) 土地收益的分配

长期以来，土地转让产生的巨大收益已经成为某些利益集团的主要目标。这方面的制度不健全为此提供了极大的方便。由于现行财政体制下地方政府常规财源不足，不能满足基础设施、城市管理、公共服务等需要，土地出让金的支配权被交给了地方政府，已经成为一些地方政府不可缺少的主要财源。但在这个过程中，政府实际上预支了未来70年的土地价值，造成了短期的过度繁荣。更严重的问题，是管理这笔巨大收益的制度漏洞。

据“中国指数研究院”推算，2009年全国土地出让金总额已达到1.5万亿元（新华网）。这相当于全国6.8万亿元财政预算的20%以上（但未包括在预算内）。在一些地方，这部分收益管理混乱，成了当地官员任意开支的钱口袋。而地方官员对土地出让金的豁免权，还可能给他们带来更大的利益。据审计署披露，2009年11个省区有684亿元土地专项资金未纳入预算管理，还有381亿元土地出让金应征未征。不难想象，数百亿元应征未征的出让金，会给有关官员带来多大的回报。

土地收益的分配问题不仅限于土地出

让金。由于土地资源的稀缺性，政府掌握了批地、征地、卖地的权力，使得土地供应和相关的房地产开发都具有某种垄断性，从而围绕着与土地相关的权力，产生了一系列寻租行为。根据一项估算，2009年全国城镇商品住房各项成本、税费合计为每平方米2500元，按销售面积和销售额计算的平均销售价格约4500元，房地产业出售商品住宅的利润高达1万多亿元⁸。而根据国家统计局更新的2009年商品住房销售面积和销售额数据，按同样方法计算，实际上应为1.7万亿元。这相当于2008年全国工业利润总额的一半以上。而按就业人数计，房地产业只相当于工业就业人数的1.3%；按资产总计算，也不过相当于工业的6%（均为2008年数据）。因此如果上述估算基本准确，房地产业可以说是不折不扣的暴利行业。可以作为佐证的是，在2010年福布斯世界富豪榜的中国前20名和胡润富豪榜的前10名中，房地产商都占了6名，是所有产业中富豪最集中的产业。

不过，1.7万亿元可以说是房地产业的利润空间，却不会是房地产商实际得到的利润。因为房地产商要拿地，往往需要给握有批地或其他相关权力的人大量“进贡”。这可能已经通过各种方式打入了房地产成本，而账面利润自然会低得多。因此，1.7万亿元巨额利润，实际上是在资本和权力之间进行分配的。这些利润最终是由千千万万购买商品住房用于自己居住的普通中产阶级支付的。因此在实际上，通过土地资源的垄断性和缺乏透明度的分配，实现了财富在中产阶级和超级富豪之间的逆向财富再分配。

由于土地的垄断和投机哄抬导致的高

⁸ 据全国人大代表、重庆市政协副主席陈万志估算；《东方早报》2010年3月4日，见于新华网。

地价，也转化为商业用房产的高房租，大幅度推高了城市商业区的商品和服务价格，从而成为城市居民必须缴纳的一项税款。这进一步加剧了财富的聚集。

可以预计，在房产税改革和相关的土地制度改革、财税体制改革迟迟不出台的情况下，高地价——高房价——逆向收入再分配这样一个怪圈，是不可能根本打破的。

(4) 其他垄断收益的分配

根据2008、2009年的工资统计，具有垄断性的行业（石油、烟草、电力、电信、银行、证券等等）平均工资与全国平均工资之比，除了证券业接近6倍，其他都仅在两倍左右或者不到两倍。但这些数据没有反映出实际的行业间收入差距。

第一，垄断性行业的实际人均收入水平远远高于其工资水平。据权威人士估计，这些行业与其他行业实际人均收入的差距平均在5~10倍之间⁹。

第二，垄断性行业内部分配极度不均。在一些行业内，普通员工和管理层（特别是高管层）之间收入差距极其悬殊。

垄断性和竞争性行业之间收入相差悬殊，是因为它们的利润相差悬殊。而前者的高额利润或者来自它们所占有的稀缺资源所产生的收益，或者来自垄断性价格产生的收益，而不能看做正常的经营性收益。

资源或市场的垄断会导致垄断价格和超额利润，使国民收入的分配向垄断部门倾斜。一个经济体中，由于自然资源的稀缺性和规模经济的排他性导致某些部门的天然垄断是难以避免的。对这些领域的问题，需要靠税收调节、反垄断立法、价格听证、公众监督等手段来解决。这对我国的资源性行业和具有垄断地位的国有企业

⁹ 劳动和社会保障部副部长步正发在2006年一次会议上的讲话，见《中国经济周刊》2006年5月22日。

而言尤为关键。经过多年的酝酿，国有企业红利上缴和资源税改革试点已相继出台（红利上缴5%~10%，资源税按5%从价计征）。但这些改革是否已经到位，仍然很值得探讨。

还有些本属于竞争性领域的部门，也可能由于不适当的经济政策而导致垄断。某些不适当的行政性价格干预，也是因素之一。垄断性领域越多，或对垄断的制约机制越不健全，市场竞争的范围就越小，收入分配就越可能失衡。因此，限制垄断、促进竞争、调节垄断收益、防止垄断部门的既得利益集团侵犯公众利益并影响经济政策，就成为解决因垄断导致的收入分配扭曲的关键。

五、重新审视国民收入分配

根据国家统计局城乡住户调查数据和人口数据近似计算，2008年全国居民可支配收入总额为14.0万亿元，占国民可支配收入总额的44.5%。但在国家统计局的“资金流量表”核算中，住户部门可支配收入总额为17.9万亿元，比按照城乡住户收入统计数据计算的可支配收入总额高出3.9万亿元。但与本报告推算的23.2万亿元居民可支配收入，仍然少5.4万亿元。资金流量表和住户统计数据不一致，主要是由于数据来源的不同；前者主要依据的是经济普查资料，数据主要来自企业，而不是统计局的住户调查资料。这说明经济普查比统计住户调查对居民收入的遗漏要小。然而这仍然无法解决灰色收入不能进入统计视野的问题。这里，作者把推算收入与资金流量表居民收入的5.4万亿元差额，视为灰色收入。

按照“资金流量表”的数据，2008年

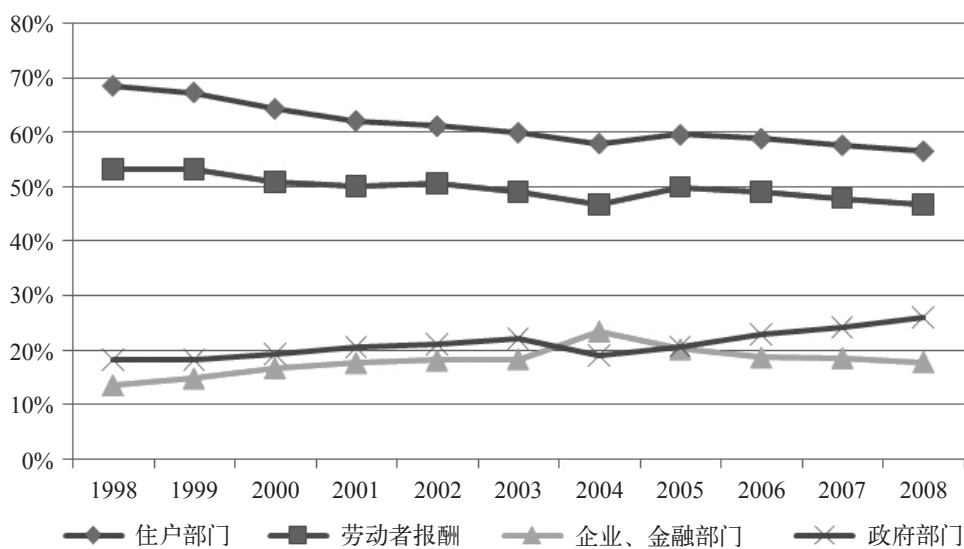


图 4 资金流量表所反映的国民收入分配结构变化

资料来源：国家统计局：《资金流量表历史资料 1992~2004》；《中国统计年鉴》2008、2009。

的居民可支配总收入为 17.87 万亿元，占 GDP 的份额为 56.4%¹⁰。其中劳动者报酬 14.75 万亿元，占 GDP 的 46.6%。这也就是说，居民的非劳动收入大约占 GDP 的 10%。另外企业（包括金融部门）和政府的可支配收入分别占 17.7% 和 25.9%。

图 4 是根据统计局历年“资金流量表”数据做出的国民收入分配结构变化图。该图显示，2008 年与 1998 年相比，居民可支配收入（住户部门）占国民总收入的比重下降了 12 个百分点，劳动者报酬的比重下降了近 7 个百分点。居民可支配收入与劳动者报酬这两条曲线之间的距离还有所收窄，显示出非劳动收入所占比重也下降了 5.5 个百分点。而企业和政府收入分别上升了 4 个和近 8 个百分点（企业比重从 2005

年以后有所下降）。

2008 年与三年前（2005 年）相比，居民可支配收入的比重下降了 3 个百分点，其中下降的主要是劳动者报酬。

但是如果用本报告前面推算的、包括隐性收入在内的居民可支配收入来代替资金流量表中的居民可支配收入，情况还是会发生很大的改变。因为灰色收入基本上不可能是劳动报酬（劳动报酬没有必要隐瞒，也很难隐瞒），这里作者把推算居民收入与资金流量表居民收入的差额（2008 年 5.37 万亿元，2005 年 2.67 万亿元；后者是根据作者 2007 年报告的推算）作为灰色收入分配到居民非劳动收入中。

在将居民可支配收入做了这一调整后，如果不相应调整国民可支配总收入（以及国民总收入和 GDP），那么居民可支配收入占后者的比重会大幅度上升，而企业和政府收入的总额和所占比重都要大幅度下调。居民收入是国民总收入的一个主要部分，我国的国民经济核算虽然主要采用生产法，

10 国民可支配总收入与国内生产总值（GDP）和国民总收入（GNI）口径接近。2007 年，国民可支配总收入只比 GDP 高 1.9%，比 GNI 高 1.1%。国家统计局尚未公布 2008 年的国民可支配总收入及其各个分项；这里使用的数字是根据国家统计局“资金流量表”中 2005~2007 年的数据进行线性外推并根据价格变动情况进行调整后得出的近似数据。

但收入法和支出法仍然是重要的参照和补充，居民收入的遗漏不可能不影响到国民总收入和 GDP 的核算，而且生产法的核算可能仍然存在遗漏。

例如，不少企业将内部人员的灰色收入，以及企业支付的回扣和行贿支出，以种种方式打入物耗成本或者计入差旅、交通、会议等支出，虚增了生产成本和管理费用，低报了企业实际增值的部分。这种情况在房地产业可能很普遍，在其他产业中也不是个别的做法。这也部分解释了假发票买卖为何如此盛行。这些情况自然会导导致 GDP 总量核算偏低。

灰色收入的另外一部分来自公共资金、公共财产、他人收入及财产的转移，是一种暗中进行的再分配。这包括公共投资资金以及其他各种政府财政和非财政资金流失到私人手中的部分、国有资产的流失、土地开发转让中的价值流失，也包括收受贿赂、房地产投机等造成的个人财产和收入的转移。这一类灰色收入并不来自 GDP 的漏算，但实际上减少了政府和企业收入，增加了少数人的个人收入；或者减少了一部分人的收入而增加了另一部分人的收入。这后一种情况，通常会减少普通居民的收入，而增加部分高收入居民的收入；或者是将少数人从普通公务人员和处于某种垄断地位的人员造就成高收入者。医疗、教育界相当普遍的私下收受红包，也是其中的一部分。

基于以上原因，作者在调整居民收入的同时，在表 9 中把 2008 年国民可支配收入以及国民总收入做一定程度的上调，调整幅度相当于当年 5.4 万亿元灰色收入的 60%。其余的 40%，需要相应地将企业和政府收入做一定程度的下调。对 2005 年的国民总收入，也采用同样方法进行调整。需

要说明的是，由于国民经济核算是一个非常复杂、工作量浩大、需要大量基础数据支撑的工作，本项课题不可能对之进行重新核算，更不能以估计数据来替代原有的核算数据。在下表的数据中，作者只是依据本报告前面部分的推算，对居民可支配收入尝试进行了修正。其他部分乃至国民总收入的数据调整，是根据一定的假设条件做出的。仅仅为研究界和各界读者提供一个参考。

鉴于国民总收入是一个读者更为熟悉的概念，以下调整前后的计算均以国民总收入为 100%，而不是以国民可支配收入为 100%。各项可支配收入占国民总收入的百分比，与占国民可支配总收入的比例相比，只有微小的差别，合计略超过 100%，基本上可以忽略。

国民总收入各构成项目调整前后的数额见表 9，它们占国民总收入的百分比见表 10。

从表 10 可见，调整前，2008 年居民（住户部门）收入只占国民总收入的 56.5%，调整后提高到 66.7%，比重上升了 10 个百分点。按调整前的数据，居民收入在 2005~2008 年期间下降了 3.6 个百分点，而按调整后的数据在此期间下降了 1.9 个百分点。这说明在加上被遗漏的灰色收入后，实际的居民收入占 GDP 比重并不像原来我们认为的那样低，其比重下降的幅度也不像原来认为的那样快。

但这并不是一个值得乐观的结果。表 10 显示，劳动报酬所占比重比原来更低了，下降的速度也更快了。2008 年从原来的 46.7% 调整到 42.3%，下调了 4.4 个百分点。与 2005 年相比，下降了 4.1 个百分点。而非劳动收入的比重则大幅度上升，2008 年为 24.4%，比原来的比例提高了 14.5 个百

表 9 国民总收入结构的调整：2005~2008 年单位：万亿元)

	调整前		调整后	
	2005 年	2008 年	2005 年	2008 年
住户部门	11.06	17.87	13.73	23.24
劳动报酬	9.28	14.75	9.28	14.75
居民非劳动收入	1.78	3.12	4.45	8.49
企业、金融部门	3.73	5.61	3.20	4.74
企业部门	3.60	5.20	3.09	4.39
金融部门	0.13	0.41	0.11	0.35
政府部门	3.83	8.20	3.29	6.92
国民总收入	18.41	31.62	20.01	34.84

注：2008 年调整前的国民收入结构数据是根据 2005 ~ 2007 年国家统计局“资金流量表”数据进行线性外推并做了某些价格调整得到的估计数，在比例上与 2007 年公布数相差不多。调整后的数据是在此基础上将估算的隐性收入进行了分配后得到的数据。

资料来源：1. 国家统计局，《中国统计年鉴》2008、2009，“资金流量表”；2. 作者推算结果。

表 10 国民总收入结构的调整：2005~2008 年（当年国民总收入 = 100%）

	调整前		调整后	
	2005 年	2008 年	2005 年	2008 年
住户部门	60.1%	56.5%	68.6%	66.7%
劳动报酬	50.4%	46.7%	46.4%	42.3%
居民非劳动收入	9.7%	9.9%	22.2%	24.4%
企业、金融部门	20.3%	17.7%	16.0%	13.6%
企业部门	19.6%	16.4%	15.5%	12.6%
金融部门	0.7%	1.3%	0.5%	1.0%
政府部门	20.8%	25.9%	16.4%	19.9%
合计	101.1%	100.2%	101.0%	100.2%

资料来源：同表 9。

分点。尽管 2005 年的该比例也有大幅度上调，但与 2005 年相比，2008 年还是上升了 2.2 个百分点。根据调整后的数据，2008 年相对于 2005 年的变动情况，可以从图 5 和图 6 的比较中更直观地看到。

调整后的居民收入比重虽然显著高于原来的统计，但高出的部分是非正常收入。正常收入的比重并没有提高。正常收入在国民收入中比重过低的推断仍然是成立的。特别是劳动报酬比重更低且进一步下降，非劳动报酬比重更高且进一步上升，尤其是最高收入居民的收入大幅度增加，这意味着收入差距进一步扩大，收入分配的公平性进一步下降。必须注意到，灰色收入也不同于正常的资本回报，对鼓励竞争、改善效率不会有任何积极作用。相反，其相

当部分可能来自企业收入和政府收入的漏失，或者是对正常的居民收入及财产的侵蚀和掠夺。这种状况不仅不利于公平，同时也损害了经济效率，并成为导致社会冲突、加剧不安定的主要因素。

与世界多数国家、特别是与发达国家相比，我国的劳动报酬占国民总收入的份额都远远低于一般水平，居民的非劳动收入份额则高于一般水平。调整后的数据更加强了这种情况。如果按调整后的估计数字来看，政府收入所占份额虽然低于原来的统计数据，但这并不说明政府集中的财力太少，而是说明有一部分政府收入由于管理不善而漏失了。

此外，尽管我国的政府收入比重低于发达国家的一般水平，但在政府收入的使

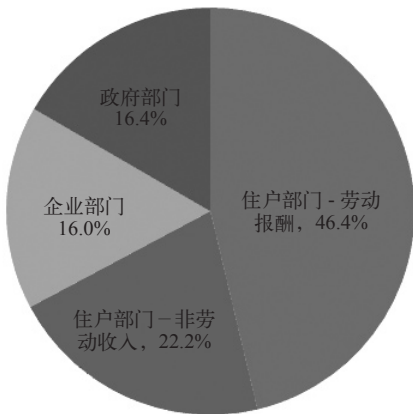


图 5 2005 年国民总收入分配状况 (调整后)

注 1: 各部分收入百分比按可支配收入占国民总收入的比例计算, 合计略高于 100%。
 注 2: 企业部门包括金融部门。
 资料来源: 国家统计局“资金流量表”、王小鲁《灰色收入与居民收入分配》(2007)、本报告计算。

用方向上与发达国家之间有重要区别。因为在大部分发达国家, 政府收入的最主要用途是社会保障, 以及提供教育、医疗、住房等方面的服务; 建立了一整套覆盖全体公民、人人可以分享的社会保障和公共服务体系。在这些方面, 我国虽然近年来发生了不少改变, 但社保体系不仅保障水平较低, 而且还远远没有实现覆盖十三亿人的目标。公共服务的提供不均等, 很大部分人群还很少能够享受到公共服务。其中尤其以保障性住房的提供严重不足。城市居民中, 能够获得廉租房的居民大约只占 1%~2%。其他保障性住房存在严重的分配不均问题, 其中公务员获得保障性住房的比例远高于社会平均水平。

另一方面, 我国的政府收入用于政府自身消费和各种投资活动的比例畸高, 在各种政绩工程、形象工程方面浪费巨大, 更严重的是大量公共资金通过各种渠道流失, 而转化为少数人非正常收入。这说明我国的公共资金和公共资源管理体制亟待改革; 在这些方面制度不规范、操作不透明、严

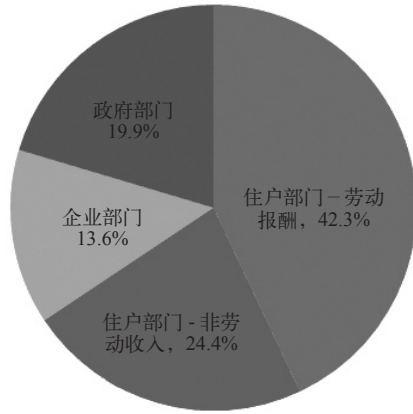


图 6 2008 年国民总收入分配状况 (调整后)

注: 计算口径和资料来源同图 5。

重缺乏公众监督等问题已经成为导致收入分配扭曲的关键因素。

六、结论

本报告基于 2009 年对全国 19 省份 64 城市四千多户城市居民家庭收支情况的调查数据, 分别采用分组比较法和模型分析法, 取得了恩格尔系数和人均收入水平对应关系的参数, 并用来对全国城镇居民分组收入统计数据进行了推算和检验。推算结果显示, 占全国城镇居民家庭 10% 的最高收入家庭, 2008 年人均可支配收入至少应为 13.9 万元, 而不是统计数据显示的 4.4 万元。两者相差 3.2 倍。其次 10% 的高收入家庭, 其人均可支配收入应为 5.5 万元, 而不是 2.6 万元, 两者相差 2 倍。没有体现在收入统计中的隐性收入, 80% 以上集中在 20% 的城镇最高收入和高收入家庭。其中 10% 的城镇最高收入家庭, 至少占有隐性收入的三分之二。

这样就使全国城镇居民人均可支配收入

入达到 3.2 万元，比原来统计的人均收入高了差不多一倍。推算 2008 年全国居民可支配收入总额为 23.2 万亿元，比按国家统计局城乡住户收入统计计算的 14 万亿元高出 9.3 万亿元（文中将其称为隐性收入），比国家统计局资金流量表基于经济普查资料计算的住户可支配收入总额 17.9 万亿元高出 5.4 万亿元。作者将这部分经济普查也无法查到的收入称之为灰色收入。以上结果，与根据居民储蓄和投资方面的宏观数据推算的结果非常接近。但由于可获得信息的局限，不排除以上估计仍然偏低的可能性。

与作者前一个报告推算的 2005 年居民收入相比，2008 年无论是 9.3 万亿元隐性收入，还是 5.4 万亿元灰色收入，都比 2005 年增加了差不多一倍，增长速度明显快于 2005~2008 年期间名义 GDP 的增长速度。由于灰色收入的存在，GDP 和国民总收入可能也在一定程度上被低估。作者粗略估计，2008 年灰色收入大约占到国民总收入的 15%。而这一比例在 2005 年约在 13% 左右。

隐性收入的存在显著扩大了收入差距。以城镇居民最高收入和最低收入各 10% 的家庭来衡量，其人均收入差距应从统计数据显示的 9 倍调整到 26 倍。以全国居民最高收入和最低收入各 10% 的家庭来衡量，其人均收入差距应从统计数据显示的 23 倍，调整到 65 倍。将这一因素考虑在内，全国居民收入分配的基尼系数，会显著高于近年来国内外有关专家计算的 0.47~0.50 的水平。

由于隐性收入的分布高度集中，说明其主要部分不是一般的统计遗漏，而是由于制度不健全或缺陷导致的灰色收入。这既包括法律法规没有明确界定其合法与否

的收入部分，也包括不能认定其来源、但实质属于非法收入的部分。大量现象说明，灰色收入主要是围绕权力产生的，是与腐败密切相关的，往往是来自凭借权力实现的聚敛财富、来自公共资金的流失、来自缺乏健全制度和管理的公共资源，或者来自对市场、对资源的垄断所产生的收益。

灰色收入的大量存在严重扭曲了国民收入分配，说明我国在经历了过去三十年经济体制改革后，在政治体制改革方面已经严重滞后。经济体制改革以市场经济体制代替了过去由自上而下的权力主导的、缺乏效率和活力的计划经济体制，同时也带来了经济利益对权力的侵蚀。

在社会经济活动全面转向以经济利益为驱动力，而不再以行政命令为驱动力的情况下，过去那种靠自上而下管理、缺乏社会公众监督的权力结构，自然也会加入追逐经济利益的潮流，使用其手中的权力谋取自身利益，原本应该为社会服务的权力就可能蜕变为某些人用来掠夺社会的工具。权力一旦与资本相结合，就将逐步取代自由竞争的市场经济，演变为垄断性的权贵资本主义，导致越来越不公平的收入和财富分配、越来越低的经济效率、越来越尖锐的社会冲突。经济体制改革带来的那种平等竞争、积极进取、全社会奋发向上、经济迅速发展的局面将不复存在。这种与资本结合的权力结构，最终会将社会引向发展停滞和严重的危机。

一个在经济上保持竞争活力和发展动力的市场经济社会，同时也必须是一个政治上公开透明、民主和谐，行政管理有序高效的社会。因此与经济体制改革相适应，必须推进政治体制和政府管理体制的改革。任何掩盖社会矛盾、无限期推迟改革的作法，只能获得表面的、暂时的稳定，却埋

藏下越来越严重的内在不稳定和社会危机的种子。

靠行政权力本身进行自我约束和自我清理，即使在理想的情况下也只能延缓这种腐化趋势，而不可能根本改变这一趋势。要制止这种趋势，唯一的出路是靠健全制度、靠社会公众的监督保持政权的自身健康。这就需要推进政府管理体制改革，建立一套严格、合理、公开、透明的制度及管理方式，形成一个在阳光下运作的政府，也才能建设一个和谐、健康、公平、公正、持续发展的社会。