

China Family Panel Studies

CFPS

中国家庭动态跟踪调查

技术报告系列: CFPS-14

系列编辑: 谢宇 责任编辑: 胡婧炜

中国家庭动态跟踪调查 2010年农村家庭收入的调整办法

谢宇 张春泥 黄国英 许琪 徐宏伟

2012.12.20

1. 问题描述

CFPS 2010 共调查了 9,752 户农村家庭¹，占家户总样本的 65.9%，在具有全国代表性的再抽样样本中，农村家庭有 6,262 户，占再抽样家户样本的 64.8%。CFPS 2010 对农村家庭和城市家庭采用相同的问卷来测量每户家庭 2009 年的年收入。年收入的测量可概括为四部分：工资性收入、家庭经营收入、财产收入（包括地租、房租收入等）和转移性收入（包括离/退休金、社会保障金、低保等）。我们基于农村再抽样样本分别计算了这四部分收入的均值，并与国家统计局公布的 2009 年农村家庭人均收入情况进行了对比，结果见表 1：

表 1. 农村家庭 2009 年人均收入中各部分收入的均值（单位：元）

	国统局	CFPS 2010 ²
2009 年工资性收入	2,061.25	4,296.36
2009 年家庭经营收入	2,526.78	1,205.79
2009 年财产收入	167.20	169.46
2009 年转移性收入	397.95	287.99

国统局数据来源：《中国统计年鉴 2010》表“10-22 各地区按来源分农村居民家庭人均纯收入（2009 年）”，p. 365。

从表 1 中可以看到，与国统局的发布数据相比，CFPS 2010 调查得到的农村人均工资性收入较高，但以农业收入为主的人均家庭经营收入偏低。财产收入和转移性收入与国统局的数据基本吻合。在工资收入的测量上，CFPS 2010 与国统局数据的不一致主要是统计口径的不同所造成。国统局对“工资性收入”提供的指标解释为“农村住户成员受雇于单位或个人，靠出卖劳动获得的收入。”首先，该指标解释定义的“农村住户成员”仅为常住人员，而 CFPS 2010 定义的“家庭成员”是指所有同灶吃饭的成员，既包括住在家中的常住人口，也包括由于上学、打工、参军等原因暂时不住在家中的成员，因此 CFPS 2010 测量工资收入所涉及的家庭成员范围较广，尤其会包括进城打工的家庭成员，这部分家庭成员带来的工资收入可能较高。其次，国统局提供的指标解释未指出“出卖劳动获得的收入”具体包含哪些收入，而 CFPS 2010 将工资收入的内容定义为“含工资、奖金、补贴、分到个人名下的红利，但

¹ 农村家庭和城市家庭的划分主要根据村/居问卷中访员实地考察该社区是“居委会”还是“村委会”（变量 ca0）来判断。对 241 个没有村/居问卷的家庭，采用国家统计局资料的城乡分类变量（变量 urban）来判断。

² 此处计算 CFPS2010 的各项收入是基于具有全国代表性的再抽样样本。我们事先对各项收入的异常值进行了清理，异常值的判定标准详见沈艳、雷晓燕，2012，《中国家庭动态跟踪调查 2010 年收入支出数据整理（CFPS-15）》。

不包括离/退休金和其他政府补助”。因此，CFPS 2010 对工资性收入的测量不仅仅包括基本工资，还包括与非农工作相关的奖金、补贴、红利。这两点也许是造成 CFPS 2010 计算出来的农村家庭工资性收入偏高的原因。

CFPS 2010测量的家庭经营收入偏低的可能原因是CFPS 2010在测量农业收入时仅笼统询问了去年农村家庭农、林、牧、副、渔业的经营收入，而该收入仅是农村家庭出售农业产品的收入，并不能代表农村家庭从事农业生产的全部所得。中国农村家庭会将相当一部分农业生产产品用于自家消费。例如，在CFPS 2010中，有超过36%的从事农业生产的农村家庭填写的去年的农业经营纯收入在1,000元及以下（见图1），这类家庭的农业生产很可能用于自给自足，而非出售。在计算农业收入时，如果研究者只使用农业经营或销售的纯收入，则会低估农村家庭的收入，尤其会大大低估农产品市场化程度低的地区的农户收入，而其中又以贫困家庭为最甚。为了弥补调查中没有收集农村家庭自家消费农产品的价值所造成的损失，并对农村家庭收入做出较准确的估计，我们特别设计以下办法，在已有数据和信息的基础上，调整农村家庭农业生产收入的计算方式，并据此调整农村家庭的人均收入。

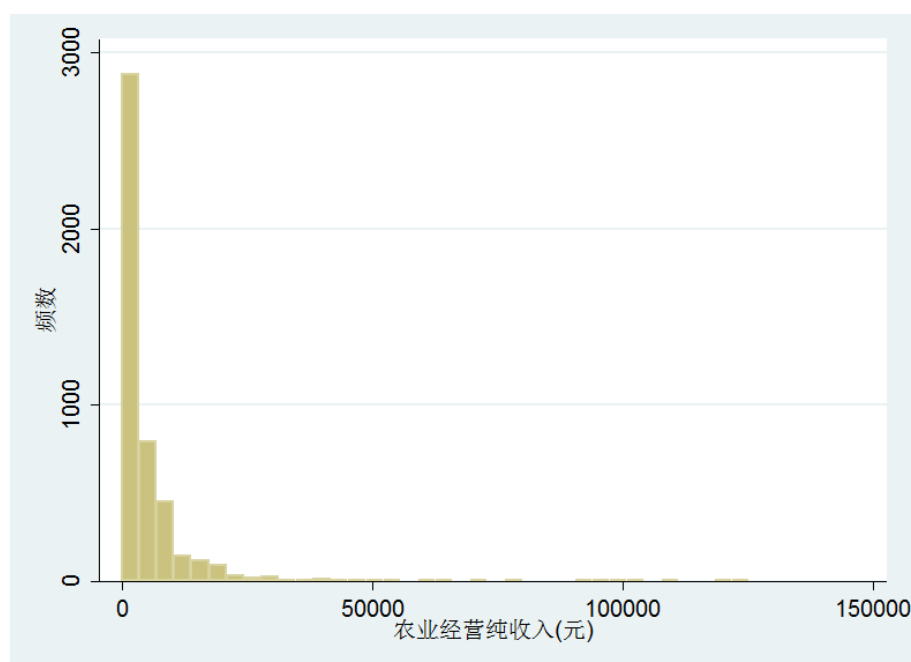


图 1. 从事农业生产的农村家庭的农业经营纯收入分布

2. 调整办法

尽管 CFPS 2010 对农业收入的笼统测量没有计入实物消费的部分，但所幸 CFPS 2010 调查了每户从事农业生产的家庭所生产的几种具体农产品的产量、销售量、销售收入和净收入。调查所涵盖的农产品种类囊括了农村家庭农业生产的绝大部分内容，具体包括水稻、小麦、玉米、棉花、林产品/林副产品 5 类农林作物，以及猪、牛、羊、鱼 4 类家畜及渔业产品。相关题目位于家庭问卷的 K6 和 K7 部分。研究者可以根据各项农产品的总产量与销售量的差值计算出农村家庭用于自家消费的农产品数量，并根据市场价格将这部分消费折算为收入，并与已知的出卖农产品所得的收入相加，得到家庭农业生产的总收入。具体步骤如下：

(1) 用每种产品的总产量减去销售量得到该产品未销售的数量。

(2) 用该产品的价格乘以未销售的数量得到这部分产品的市场价值。

(3) 将各项农产品没有销售部分的市场价值相加得到该户生产出来但未销售的所有农产品的总价值。

(4) 将农业销售的毛收入（或纯收入³）与未销售部分的农产品总价值相加，得到农业生产总收入（或纯收入）。

调整公式如下所示：

$$Y_i = x_i + \sum_k p_k (T_{ik} - S_{ik})$$

公式中的下标 i 代表农户， k 代表农产品种类。 x_i 是 2009 年 i 农户出卖所有农产品所得的毛收入（或纯收入），在家庭问卷中对应的题目是 K3（或 K5）。 p_k 是第 k 种农产品的市场价格。 T_{ik} 是 i 农户 2009 年第 k 种农产品的总产量， S_{ik} 是 i 农户 2009 年第 k 种农产品的销售量；二者之差就是该农户 2009 年第 k 种农产品未销售的数量，将其乘以价格 p_k 折算为市场价值；再对农产品种类 k 求和就得到了 i 农户 2009 年所有未销售的农产品的市场价值；最后与 x_i 相加，当 x_i 是农产品销售的毛收入时， Y_i 是该户农业生产的总收入；但当 x_i 是农产品销售的纯收入时，我们得到的 Y_i 是该户农业生产的纯收入。

接下来，我们将分别介绍各步骤调整的结果、操作中遇到的问题和解决方案。

³ 问卷中询问的是农户经营农林牧副渔的总成本(K4)，所以严格来说，此处农产品销售部分的纯收入(K5)实际上已经扣除了自家消费部分的成本。

(1) 农产品未销售部分 $T_{ik} - S_{ik}$

我们对不同的农产品，分别计算了其未销售部分的数量，即用该产品的总产出减去该产品的总销售量。在计算之前，我们先对农产品的测量单位和填报的数量值进行了核查。

CFPS 2010 在调查时允许受访者在回答农产品产量和销售量时使用多种测量单位。其中，农林作物的可选单位包括公斤、斤和方，而家畜和渔业产品的可选单位包括公斤、斤和头。由于不同的农户填报同一农产品数量时选择的单位不同，以及少数农户在回答同一产品的产量和销售量时也选用了不同的单位，因此，我们事先对农产品的测量单位进行了统一。我们将水稻、小麦、玉米、棉花、林产品/林副产品的单位均统一为公斤，鱼的单位也为公斤，猪、牛、羊的单位均折算为头。头和公斤的折算方法是根据该农产品不同单位的市场价格之比近似估计出不同单位之间的换算率，比如一头牛的市场价格是 2,527.64 元，而一公斤牛肉的价格是 16.62 元，这样一头牛的重量大概是 152.08 公斤。

在核查数据值的过程中，我们发现了少部分农户的销售量比产量更高。对此，我们猜测可能有三种原因：第一，单位填错或者数值填错。第二，不排除有一些农产品是 2009 年之前生产但没卖出去的，在 2009 年给卖出去了，导致 2009 年销售的数量大于当年生产的数量。第三，养猪的农户出现销售量大于总产量的情况偏多，这种情况尤其出现在养殖户回答生产量 and 销售量单位不统一的时候。通常，市场上出售的活猪分两种：崽猪和肥猪，这两种猪的市场价相差较大，但 CFPS 2010 的问卷中并没有区分这两种猪。崽猪由于重量轻，每头的价格大大低于肥猪，根据样本计算出来每头猪的价格是混合了崽猪和肥猪之后的平均价格。这样，在计算公斤与头的转换率时，我们是用两种猪混合后的平均价格除以 1 公斤猪肉的价格得到的每头猪的平均重量，即平均了崽猪和肥猪以后的重量。计算得到一头猪大约重 63.82 公斤，这个重量高估了崽猪的平均重量，但也低估了肥猪的平均重量。其结果是，对于那些养殖肥猪的农户，当我们将他们以公斤计的销售量转换为头的时候，高估了销售的头数，导致销售量大于产量。对于以上各种销售量高于产量的情况，我们的处理办法是：假定该农户当年对该农产品销售没有剩余，不对其进行调整。

表 2 描述了问卷调查的各种具体农产品中未销售部分的平均数量和平均比例。从表中可以看到，农户生产的大部分水稻和小麦都未用于出售。在种植的作物中，仅棉花和林产品/林副产品的销售比例较高。在家畜及渔业养殖上，猪、牛、羊、鱼均有二成左右未出售，其中猪和牛的销售比例略高于羊和鱼。由此可见，农业生产中确有相当比例的产品并非用于出

售，如果计算农业收入时不计入这部分实物的价值，则确实有可能导致对农业收入的低估。

表 2. 农产品未销售部分的数量及比例

	未销售量均值	单位	平均未销售比例	样本量（户）
水稻	982.41	公斤	0.80	2,395
小麦	777.65	公斤	0.69	3,091
玉米	575.73	公斤	0.49	4,710
棉花	20.73	公斤	0.20	214
林产品/林副产品	620.16	公斤	0.05	455
猪	1.31	头	0.18	1,284
牛	1.24	头	0.19	370
羊	4.34	头	0.27	333
鱼	193.91	公斤	0.24	69

(2) 市场价格 p_k

市场价格 p_k 的获取方式有两种：一种是直接采用已有统计资料中的各农产品市场价格；另一种是根据 CFPS 2010 的数据计算所有农户出售各种农产品的价格并在家户汇总的层次上取各产品售价的平均值。

我们先用后一种方法，在 CFPS 2010 数据中用各种农产品的销售收入除以销售量得到每户销售该农产品的单位市场价格，并在全国层次上取平均，计算得到各农产品的平均销售价格。但由于生产某些农产品的家户样本量有限，且存在填报不准确、经营亏损等情况，用 CFPS 2010 销售信息计算出来的水稻、小麦、玉米和棉花的价格明显偏高，以公斤为单位计算的牛的价格偏低，仅按头为单位计算的猪、牛、羊的价格接近市场的经验值。因此，我们最终使用的市场价格 p_k 的取值是以国统局“200 个主产县主要农产品价格变动情况”提供的价格标准为主，对于没有出现在国统局数据中的农产品，如林产品，以及存在多重测量单位的农产品，如猪、牛、羊，我们使用 CFPS 调查数据进行补充。表 3 描述了 CFPS 2010 调查中所涉及的所有农产品的市场价格 p_k 。

表 3. 全国各农产品单位产量销售价格⁴

农产品	单位	单位产量销售价格（元）	数据来源
水稻	公斤	2.21	国统局
小麦	公斤	1.97	国统局
玉米	公斤	1.87	国统局
棉花	公斤	6.09	国统局
林产品/林副产品	公斤	18.36	CFPS
	方	768.28	CFPS
猪	(肥猪) 公斤	10.87	国统局
	(仔猪) 公斤	15.81	国统局
	头	851.35	CFPS
牛	公斤	16.62	国统局
	头	2,527.64	CFPS
羊	公斤	16.73	国统局
	头	375.35	CFPS
鱼	公斤	9.36	CFPS

(3) 农业收入调整结果 Y_i

根据我们先前提出的公式, 各种农产品未销售部分的数量乘以其市场价格的加总即是农村家庭未出售的实物部分的折算收入。将这一收入与全部农产品销售毛收入（或纯收入⁵）相加, 即得到农户去年农业生产的总收入（或纯收入）。

表 4 比较了调整前后农业生产纯收入的均值、标准差、中位值的变化。计算的对象是 2009 年从事农业生产的家庭。可以发现, 经调整, 每户农业纯收入的均值提高了 2,469.7 元/年, 中位数提高了 2,275 元/年。

表 4. 调整前后 2009 年家庭农业生产纯收入的比较

	均值（元）	标准差	中位数	样本量（户）
调整前	4,889.0	12,796.6	2,105.0	7,586 ⁶
调整后	7,358.7	14,044.3	4,380.0	7,586

⁴ 国统局的数据来自 2009 年 6 月 1 日至 2010 年 5 月 31 日“200 个主产区主要农产品价格变动情况”各农产品价格 12 个月的平均值。其中, 水稻的价格是籼稻原粮和粳稻原粮的平均值。

⁵ CFPS 2010 分别询问了销售毛收入、总成本和销售纯收入。农业毛收入为农产品的销售收入, 农业纯收入为毛收入减去生产成本和税费。当出现填答的毛收入减总成本不等于填答的纯收入的情况时, 在计算纯收入（毛收入 $fk3$ -总成本 $fk4$ ）和填答纯收入（ $fk5$ ）之间取数值较大的一个作为最终的销售纯收入。纯收入为负数的情况作零值处理, 假定亏损等同于去年没有赚到钱。因此, 此处的销售纯收入是经过以上检查和更正以后的收入。

⁶ CFPS 2010 家庭总样本中共有 7,798 户家庭回答去年从事过农业生产。此处的 7,586 户是家庭农业生产纯收入调整前后均无缺失值的样本, 其中包含从事农业生产但居住在城市社区的家庭。

3. 调整前后农村家庭收入的比对

表 5 比较了调整农业生产收入前后 CFPS 2010 农村家庭 2009 年人均收入的变化。由于加入了农业生产收入中实物消费的部分，人均年收入的均值上调了 10%。调整后的家庭经营收入更接近国统局的 2,526.78 元/人年的标准。余下的差异可能来自于无法调整的没有包含在 CFPS 问卷中的具体农产品的收入，比如，CFPS 中没有询问鸡的养殖，以及大豆、油料作物、瓜果的种植，而这部分农产品会有相当比例用于农户自家消费⁷。农业收入的调整对估计农村贫困家庭的收入有较大的影响，由于调整前没有计入农业产品中自家消费的部分，因此低估了从事自给自足农业生产的贫困家庭的收入。除此之外，致使家庭经营收入偏低的原因还可能包括非农家庭经营利润的漏报或瞒报。样本中有 7.8% 的农村家庭回答参与了非农经营，但仅有 1.6% 的家庭填写了非农经营的税后纯利润。由于 CFPS 对家庭非农经营的调查主要是提问公司/企业的经营、获利情况，并以万元为单位，从事小规模生意的家庭有可能漏报或不报，而拥有较大规模生意的家庭有可能瞒报。

表 5. 调整前后农村家庭 2009 年人均收入及收入各组成部分的均值⁸（单位：元）

	CFPS（调整前）	CFPS（调整后）
2009 年人均收入 ⁹	5,959.59	6,597.97
2009 年工资性收入	4,296.36	4,296.36
2009 年家庭经营收入	1,205.79	1,683.38
2009 年财产收入	169.46	169.46
2009 年转移性收入	287.99	287.99

调整后的 CFPS 农村家庭人均年收入仍与国家统计局农调队的数据有一定的出入，尤其是在工资性收入上。但研究者必须注意到 CFPS 与国统局在收入数据的收集和测量上的确采

⁷ 根据《中国农村统计年鉴 2010》表“10-1 全国种植业产品成本与收益”(p.259)，大豆的商品率为 91.3%，两种油料（花生和油菜籽）的平均商品率为 74%。由此可见，此类种植产品有相当比例用于自家消费或其他非出售目的，但在 CFPS2010 问卷中没有具体询问此类种植产品的生产和销售情况，故无法调整生产此类种植产品农户的农业纯收入。

⁸ 在收入的计算过程中，我们发现有部分家庭的工资性收入、家庭经营收入、转移性收入和财产性收入均为 0。对此，我们做了谨慎的处理和调整。我们比对了这部分家庭的消费数据。如果零收入家庭的消费亦为零或为缺失，则视该户家庭为放弃作答，整户家庭的收入作缺失处理。如果该家庭的消费不为零，则将该家庭当年的消费作为该家庭当年的收入。最终用于计算农村家庭人均纯收入和各项收入的样本量为 5,958 户。由于对一部分零收入的家庭做了额外的调整，人均纯收入的均值会略大于四个收入子部分均值的加总。

⁹ CFPS 2010 家庭问卷中还询问了其他收入，其他收入包括亲友馈赠和受访家庭汇报的其他收入。由于其他收入难以归入国家统计局定义的四项收入中，为了与统计局数据进行比较，此报告中的人均收入未计入这些“其他收入”。

用了不同的统计口径、指标界定和收集方法。严格意义上说，两个数据并非完全可比。此处的比较只是对 CFPS 数据稳健性的检测。

为了进一步说明 CFPS 数据的可靠性，我们将 CFPS 2010 测量的农村家庭年收入与测量的农村家庭年消费做比较。通常，收入和消费是相对应的两方面，尤其是对低收入的家庭来说，其收入的绝大部分即用于消费。因此，如果收入的数据合理，则其与消费的差异应该是随收入的增高而增高。我们计算了在不同年收入百分位上农村家庭的年人均收入与年人均消费之差。如图 2 所示，我们可以看到，年收入与消费之差与年收入的确存在正向的关系。在低收入的群体中，这一关系主要体现为入不敷出到收支相抵，在高收入群体中，则会有较高的收入累积。

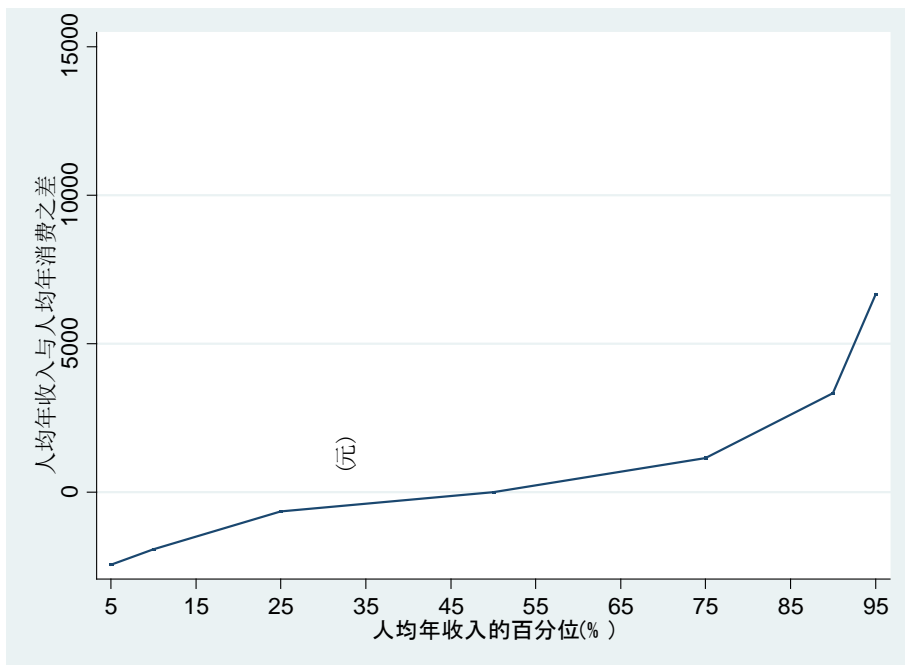


图 2. 农村家庭不同年收入百分位上的收入与消费之差

4. 生成农业收入及家庭收入的综合变量

在本报告中，我们介绍了如何利用农产品生产及销售的信息调整农业收入的方法，并对调整后的农业收入和农村家庭收入与国统局的数据进行了比较。为了方便研究者使用数据，我们将调整前后的农业收入、调整前后的家庭收入、其他各部分的收入均生成了变量并添加到家庭问卷数据库中。以下是这些变量的名称及标签：

	变量名	变量标签
1	faminc_old	未调整的家庭总收入
2	faminc_net_old	未调整的家庭纯收入
3	faminc	调整后的家庭总收入
4	faminc_net	调整后的家庭纯收入
5	indinc	调整后的家庭人均总收入
6	indinc_net	调整后的家庭人均纯收入
7	inc_agri	调整后的农业生产总收入
8	net_agri	调整后的农业生产纯收入
9	foperate	调整后的经营性总收入（包括农业生产和非农经营）
10	foperate_net	调整后的经营性纯收入（包括农业生产和非农经营）
11	firm	非农经营收入
12	finc	工资性收入
13	fproperty	财产性收入
14	welfare	转移性收入
15	felse	其他收入

在以上变量中，总收入是指没有扣除农业生产成本的收入，纯收入是指扣除农业生产成本后的收入。人均总收入和人均纯收入是用总收入或纯收入除以家庭规模得到的平均收入，计算所使用到的家庭规模是 T1 表同住人员的人数。家庭总（纯）收入是五个部分收入的加总，即工资性收入、经营性总（纯）收入、财产性收入、转移性收入和其他收入。其中，经营性收入包括农业生产收入和非农经营收入。研究者可以直接使用这些新增变量，或者根据数据中的原始变量自行生成相应的收入变量。