

China Family Panel Studies



中国家庭追踪调查

技术报告系列: CFPS-29

系列编辑: 谢宇 责任编辑: 张聪

中国家庭追踪调查
2012 年和 2010 年财产数据技术报告

靳永爱 谢宇

2014-12-10

本报告主要记录了财产数据的清理、缺失值处理以及综合变量生成过程。

中国家庭追踪调查（以下简称 CFPS）2012 年和 2010 年在财产问题的设计上有一定差异，主要差别在于：（1）2012 年各项财产询问更细更准确，如耐用消费品在 2010 年被包含在家庭其他资产里面统一询问“您家的其他资产现值为___元”，而在 2012 年则是分项提问价值。2010 年财产中漏掉了政府债券、金融衍生品和其他金融产品。2010 年除了现住房外，其他房产是统一询问价值，2012 年则是分别询问其他房产中每套房产的价值。（2）2010 年财产相关问题的缺失值未进一步追问，而 2012 年则使用逼近法（unfolding brackets）的方式进行了追问。考虑到这些不同，两年数据在具体清理过程、缺失值插补以及综合变量生成上也存在一些差别，以下将分年份介绍数据处理过程。

一、 数据清理

本节介绍单位弄错的变量的处理过程。

（一）2012 年财产数据修正

对 2012 年数据清理和修正主要采取两个办法：一是回听所有大于 1000 万的值的录音，根据录音修正数据；二是因没有录音数据、录音听不清楚、地方方言等原因而无法依据录音修正的观测，结合家庭中的其他信息和 2010 年对应的信息进行判断与修正。

2012 年财产数据中，主要有以下变量存在不合理的值：①fq4a，您家现住房当前市价；②fr2a_a_1，离您家渐远的第 1 套房产市价；fr2a_a_2，离您家渐远的第 2 套房产；fr2a_a_3，离您家渐远的第 3 套房产市价；③系统自动生成的以“元”为单位的现住房和其他每套房产的市价，如 houseprice1、houseprice2_a_1、houseprice2_a_2、houseprice2_a_3、houseprice2_a_4；④fq501，现住房的待还房贷；fr301_a_2，其他房产中第 2 套待还房贷；⑤fq701，您家现住房建筑面积；⑥fm4_a_1，第一项个体私营总资产；fm4_a_2，第二项个体私营总资产；⑦fm7_a_1，第一项经营的净收入。下面将分别对这 7 种情形加以说明。

1.fq4a，现住房的市价，以万元为单位。这个变量存在两个方面的数据问题：一是 1000 万以上的房产可能是混淆了单位“万元”和“元”；二是在单位上多填或少填几个 0，比如 80 万填成了 800 万，这个问题与 2010 年房价对比可以发现。

报告的 fq4a 在 1000 万以上的共有 45 户，通过录音能够纠正其中的 11 户数据，剩下的 34 户中有 31 户可以根据现在家庭信息，如住房类型、住房面积、住房内设施（水、做饭燃料、厕所）、城乡性质、所在省份、家庭收入、各项支出以及 2010 年住房信息，判断其存在混淆了“万元”和“元”的问题。对于这些家庭户，其填答的 fq4a 值直接除以 10000 就可以得到修正值。其余 3 户虽然可以判断填答的值是弄错了单位，但不能确定是混淆了“元”和“万元”，因此将这三户赋为缺失值（fid12=312405、fid12=621784、fid=643364）。

2. fr2a_a_1，离家渐远的其他第一套房产市值，以万元为单位。值在 1000 万以上的有 8 户家庭，其中 3 户根据录音数据修正，其余 6 户可以判断是误将单位“万元”当作“元”填写，直接除以 10000 得修正值。这 6 户家庭为：fid12=140068、fid12=420330、fid12=430152、fid12=430182、fid12=140046、fid12=140058、fid12=430151。

fr2a_a_2，离家渐远的其他第二套房产市值，以万元为单位。有两户在 1000 万以上，通过录音数据纠正。

fr2a_a_3，离家渐远的其他第三套房产市值，以万元为单位。有两户在 1000 万以上，通过录音数据修正。

3.houseprice1 变量，数据库中系统根据变量 fq4a（万元）、fq4b（千元）、fq4c（百元）、fq4d（十元）、fq4e（元）综合生成以元为单位的现住房市价。对 fq4a 进行清理并处理极值后，houseprice 变量需要重新生成，在数据库中以 best 为后缀的 houseprice 变量表示。

houseprice2_a_1—houseprice2_a_4 是其他房产以元为单位的市价，一方面因为修改了以万为单位的极值，另一方面由于系统在生成以元为单位的房子市价时存在某些问题，有些计算不准确，所以这五个变量都根据新清理的值重新生成了以 best 为后缀的变量。

4.fq501，现住房待还房贷本息，以万元为单位。将房贷在 500 万以上的家庭挑出来，与该房子的市价、房子类型、住房内设施、城乡性质结合起来考虑，发现 500 万以上的家庭除了 1 家外，其他的都是误将单位“万元”当成了“元”填写。无法判断的这一户房贷填写的是 5000 万，但是房子市价“不知道”，所以将该户的房贷赋为缺失（fid12=282783）。

fr301_a_2，其他房产中第二套房产待偿贷款本息总额，以万元为单位。出现一户奇异值，待还房贷填写的是 10000 万元（fid12=230626），通过综合判断，认为是 10000 元。

5.fq701, 您家现住房建筑面积, 单位为平方米。有四户的面积出现了偏离总体分布过远的情况, 分别达到 5447 (fid12=220242)、5925 (fid12=130174)、8533 (fid12=211149)、5238 (fid12=230262) 平方米。经过与 2010 年面积对比, 发现 2012 年扩大了 100 倍, 而房屋类型并没有发生变化。由此判断可能是访员填答时漏掉了小数点, 故将这四户的面积在原值上除以 100 即得到修正值。

6.fm4_a_1, 第一项个体私营总资产, 单位万元。2012 年 CFPS 财产数据中的奇异值处理办法主要是通过录音纠正。2012 年家庭第一项个体私营总资产在 1000 万以上的共有 172 户。其中通过录音回听纠正了 115 户的数据。

将录音回听数据与原始数据对比可以看出, 绝大部分家户是错将单位“元”当成了“万元”。剩下的 57 户录音数据缺失, 根据 2012 年相应的股份比例、雇佣人数、净收入、城乡性质、家庭收入、家庭各项支出、住房类型、住房市价、住房设施(水、燃料、厕所)以及 2010 年对应的资产值修正。其中 53 户十分确定是错将单位“万元”当成了“元”的家庭, 将原资产值缩小到原值的 1/10000。其他无法确定的 4 户家庭, 直接赋值为缺失。

fm4_a_2, 第二项个体私营总资产, 单位万元。该变量值在 1000 万以上的共有 4 家, 其中 3 家通过录音进行了修正, 剩下的一家雇佣了 12 人, 净收入写的 30000 万, 原资产值填写的是 4 亿, 无法确定该值是否被扩大了 10000 倍, 所以赋为缺失。另外一家 fid12=430357 地处村委会, 填写的资产值为 500 万, 但是雇佣人数是 0, 净收入也是 0 元, 通过录音进行了纠正。

7.fm7_a_1, 第一项经营的净收入, 单位元。

通过比较 fm4_a_1 (单位: 万元) 和 fm7_a_1 (单位: 元), 发现有六户的 fm7_a_1 值有问题, 主要表现为前者与后者差距过大, 比如 fm4_a_1 为 130 万, 但 fm7_a_1 却仅为 30 元。有四户家庭 fid12=211393、fid12=411032、fid12=411568、fid12=510857 的第一项经营净收入填错了, 通过录音纠正了。其余两户 fid12=510803 和 fid12=530051 无法找到录音未改动。

fr2a_a_4、fr301_a_1、fr301_a_3、fr301_a_4、houseprice2_a_5、houseprice2_a_6 和 fm4_a_3、fm4_a_4、fm7_a_2、fm7_a_3、fm7_a_4 没有极值, 未做任何改动。

为了统一, 便于用户使用, 这些变量都重新生成了以 best 为后缀名的变量。

fm7_a_4_best 除了调整不合理的值外，还根据填报的最大值和最小值，采用取平均值的方法估计了回答不知道和拒绝回答的情况。

（二）2010 年财产数据修正

2010 年财产数据中，主要有四组变量存在极值问题：①fd4（上个月，您家现住房子的市值约为多少万元？）；②fd703（其他几处住房现在的总市值约为多少万元）；③fv4_a_1（经营第一家企业/公司的产业的总资产为多少万元）、fv4_a_2（经营第二家企业/公司的产业的总资产为多少万元）；④fv7_a_1、f7_a_2（经营的第一家、第二家公司去年的营业额为多少万元）；5.f8_a_1、f8_a_2（经营的第一家、第二家公司去年的税后净利润是多少万元）。

1.f4，上个月您家现住房子的市值约为多少万元，以万元为单位。

有 17 家的现住房市价在 1000 万以上，结合住房类型（单元房、平凡、小楼房、别墅等）、房内设施（厕所类型、做饭燃料、水）、住房面积、家庭年收入、各项支出以及城乡性质判断，确定有三户是混淆了单位“元”和“万元”，故将原值除以 10000（fid=320074、fid=350187、fid=530334）进行修正。其中有两户是上海和北京的住房，考虑到上海和北京房价高的问题，认为 1000 万以上的住房是合理的。

其余 12 户，能够确定这些家庭的住房不可能在 1000 万以上，但直接除以 10000 的值也不太符合家庭特征，因此，将这些家庭的住房市价赋值为缺失。这些家庭的户号（fid）为：

211308	311669	320193	320195	320198	320201
320202	330041	330045	330175	340363	350186

2. fd703，其他几处住房现在的总市值约为多少万元，以万元为单位。共有 77 户的其他房产在 1000 万以上。参照该家庭户现住房的市价、其他房产的面积以及其他在修正现住房市价 fd4 时考虑到的因素，判断是否为极值。同样的，上海的两户（fid=312050、fid=312391）fd703 不变动，另外有 14 户可以判定属于极值，但无法确定是否扩大了 10000 倍，所以赋值为缺失。

这些家庭的户号（fid）如下：

230109	370488	410775	411693	420054	430368	441455
441468	441829	510645	520008	520318	530120	620534

其余的家户均缩小为原值的 1/10000。

3.fv4_a_1, 家庭经营第一项资产, 以万元为单位。首先将资产在 1000 万及以上的家户全部挑出来, 结合该项资产的股份比例、雇佣人数、营业额(万元)、纯利润(万元), 2010 年家庭收入、各项支出、城乡性质、居住房子类型、市价及设施(如厕所类型、做饭燃料、水), 以及相对应的 2010 年的资产额, 综合这些信息判断 1000 万以上的资产是否是混淆了元和万元。

逐条判断后发现: 家户号为 fid=120015、fid=130608、fid=230503、fid=310172、fid=330169、fid=430088、fid=430613、fid=430639、fid=530342 的家庭资产在 1000 万以上是合理的, 不过需要注意的是其中除 fid=120015 和 fid=430088 两户外, 其他虽然填的资产是在 1000 万以上, 但是股份却是 0。

家户号为 fid=330365、fid=340143、fid=340172、fid=340178、fid=360240、fid=441052、fid=620888 的家庭虽然能够判断资产值有问题, 但是不能确定是否单纯是单位出错, 所以将这几户家庭的资产值赋为缺失。

除此之外, 其他资产值在 1000 万以上的均除以 10000 进行修正。这样的处理也是根据 2012 年的规律做的, 通过回听 2012 年资产在 1000 万以上的家庭的录音发现, 绝大部分是混淆了“元”和“万元”单位。

fv4_a_2, 家庭经营第二项资产, 以万元为单位。以同样的方式判断, fid=370582、fid=430088 这两户家庭的资产在 1000 万以上是正常的。而 fid=410306、fid=420431、fid=510129 这三户家庭则是混淆了“元”和“万元”, 实际值应该除以 10000 进行修正。

4.fv7_a_1, 第一家公司/企业去年的营业额是多少万元, 根据对应的资产值进行修改。将营业额在 100 万以上的都挑出来, 逐条判断。主要有以下几种判断: (1) 原资产值在 1000 万以上, 但是经过逻辑检验, 认为是混淆了单位“元”和“万元”的, 相应的 fv7_a_1 也除以 10000 进行修正; (2) 原资产值在 1000 万以上, 经过逻辑检验认为是极值但是不能确定是将“元”当成“万元”, 因其可能是出现了“1000”“100”或者“10”的单位错误。这些样本的资产值赋为缺失, 相应的营业额也赋值为缺失; (3) 原资产值没有弄错单位, 与该资产值对比, 可以判定其营业额混淆了单位“元”和“万元”的家户, 将营业额原值缩小 100

倍进行修正。比如 fid=621253，其经营的资产值为 0.5 万元，但是营业额却填答了 2000 万元，故将该值修正为 2000 元，即 0.2 万元。

在原值上缩小了 10000 倍的家庭的家户编码 (fid) 为：

130636	130709	350057	360323	410138
410723	410930	411123	430320	440528
440598	441797	510172	510939	510948
530092	530343	620327	620536	621253

fv7_a_2, 第二家公司/企业去年的营业额是多少万元，仅有一家 fid=621662 的营业额可能有问题，原值为 8000 万元，但是该户家庭的企业资产值为 10 万，所以无法判断营业额的单位，将该户家庭的营业额赋值为缺失。

5. fv8_a_1, 第一家公司/企业去年的税后净利润是多少万元，判断方式同 fv7_a_1。

在原值上缩小了 10000 倍的家庭的家户编码 (fid) 为：

130630	130636	130709	350057	360248	360323
410138	410723	411047	411123	430320	440528
441797	450300	510172	510948	520427	530092
530343	620327	620536	621253		

fv8_a_2, 第二家公司/企业去年的税后净利润是多少万元，仅有一家 fid=621662 的税后净利润可能有问题，原值为 7000 万元，但是该户家庭的企业资产值为 10 万，所以无法判断单位，将该户家庭的税后净利润赋值为缺失。

剩下几个变量 fv4_a_3、fv4_a_4、fv7_a_3、fv7_a_4、fv8_a_3、fv8_a_4 无奇异值，没有进行任何改动。

所有调整过的变量均加上“best”后缀以示区别。为了用户使用的统一性，尽管企业资产第三项和第四项相应的资产值、营业额和净利润都没有变动，但也新生成了带“best”后缀的变量。

二、财产数据中缺失值及不合理值的处理

本节介绍缺失值及不合理值的处理过程。

（一）2012 年房产缺失值和不合理值的处理

1. 缺失值处理

2012 年,对于现住房市价回答“不知道”或“拒绝回答”的受访户,继续用逼近法(unfolding brackets)的方式进行了追问,这些受访户中有 74.4%的人回答了范围,对于两个数之间的房价,取中值:比如,回答的房价大于 10 万但小于 15 万,则该户的住房价值为 12.5 万。两端的值——小于 2.5 万元和大于 50 万的情形分别按一半和 2 倍计算,即前者为 1.25 万,后者为 100 万。

通过以上处理后,还有一部分家庭户的房价仍然是缺失状态,首先用社区同类型房子的单位面积房价和现住房面积估计,无法估计的采用社区同类型房子的平均值。经过以上处理后房价仍然未知的用同省同类型住房的单位面积房价和现住房面积估计;没有填答住房面积的,用同省同类住房均价代替。

2. 其他不合理房价值的处理

（1）房产为 0 值的处理

现住房市价填答为 0 元的情形,也跟“不知道”或“拒绝回答”一样进行估计,即先通过社区信息估计,无法用社区信息估计的,使用省份信息估计。

（2）与 2010 年相比不合理的值的处理

同时,因为 2010 年询问的房价是以“万元”为单位,而 2012 年以“__万__千__百__十__元”的形式询问,2012 年可能会出现几个单位填错位的情况,比如错把万位上的数字填到了千位上,因此需要与 2010 年房价对比进行对比纠正。用 2010 年房价除以 2012 年房价,商值特别大的表示 2010 年房子市价远远高于 2012 年,回听录音数据进行修正。

剩下没有录音数据的家庭户,作以下处理:从 2010 年到 2012 年住房类型没变或者变得更好,比如从平房变为小楼房、单元房,如果 2010 年房产值除以 2012 年房产值的商值在

8~80 之间（不包括 80）¹，则在 2012 年数值上乘以 10；如果商值在 80~800 之间（不包括 800），则在 2012 年数值上乘以 100；如果商值在 800~8000 之间，则在 2012 年数值上乘以 1000。对数据所做的此类修正最后只在综合变量之一——现住房市价中（resivalue）体现，并不在 fq4a、fq4a_best 和 houseprice1、houseprice1_best 中体现。

将 fid12==450362 的家户的 resivalue 赋值为缺失，因为该家庭地处村委会，周围邻居的住房市价都远远低于 500 万，并且该户家庭 2010 年房价只有 3 万。

对于 2012 年房价远远高于 2010 年的，我们发现大部分变动都是因为该家庭的住房类型发生了很大的变动，比如从平房变为小楼房，所以重点校对了对商值小于 0.02 的观测，重要问题是 2010 年房价的问题，接下来对 2010 年房价清理说明时会详细介绍修改细节。

(3) 上海的特殊情况处理

上海情况比较特殊，如果直接按填答的数据计算，会发现 2012 年上海的平均房产价值低于 2010 年，这在房价持续上升的背景下是不合理的。经过检查发现，主要原因是一部分 2010 年有住房产权的家庭到 2012 年没有住房产权了，这样的家庭共有 96 户。其中，有 59 户家庭在 2010 年住房产权是完全自有，到 2012 年则填答的是“其他”。进一步结合其他信息，发现这些家庭基本都是住的“动迁房”，也就是说，这些家庭并不是在 2012 年失去了产权房，而是因为拆迁房子正在建设中。所以，对于这部分家庭，我们使用 2010 年住房市价和国家统计局公布的城市房价上涨比例估计 2012 年的住房市价²。

(二) 其他财产缺失值的处理

2012 年，多数财产（如房产、存款、债券、股票、基金等）都通过 unfolding 方式对“不知道”或“拒绝回答”的家庭户进行了追问。凡是回答了 unfolding 题目的家户，用 unfolding 答案估计，即采用最大值和最小值的平均值；两端开口处——比最小值还小的情形，采用最

¹ 以 8 为起点而不是直接以 10 为起点，主要是考虑到 2010—2012 年还有房价上涨问题，比如 2010 年房价为 80 万，但是 2012 年填的 10 万，考虑到城市房价上涨比例，2012 年很有可能是 100 万。

² 根据国家统计局公布的数据，2010-2011 年城市房价上涨比例为 6.8%，2011-2012 年城市房价上涨比例为 7.7%。（http://blog.sina.com.cn/s/blog_4373bd1f0101hysa.html）

小值的二分之一，比最大值还大的情形，采用最大值的 2 倍。接下来，如果还有缺失值，则用其他信息估计。

存款缺失值估计方法跟 2010 年一样，在用 unfolding 方式估计后，仍然有 100 多户家庭缺失，用同社区收入最相近的家庭存款额填补。

政府债券、股票、基金、金融衍生品和其他金融产品并不像存款那样普遍，而且其与收入的关系也很难界定，因此这些变量的缺失值我们没有进行填补。

家中耐用消费品共有 17 项，各项的价值如果缺失，则分别根据同社区人均家庭收入最接近的家庭对应的消费品价值填补。用这种方法估计后，还有缺失的，用社区平均值估计，最后还存在缺失的用所在省的平均值估计。

个体经营或企业资产有缺失的，用企业净利润代替。

因为农村中农用器械分布较为分散，可能某社区中仅一家拥有某种机械，因此难以根据同社区收入相近的农户进行估计。农村中农用器械的缺失值优先用社区平均值估计，无法用社区平均值估计的，用所在省平均值代替。

土地价值根据农业收入估计。农业收入包含了农林产品价值、禽畜和水产品总值、牲畜副产品总值以及出租土地的收入。计算农业收入时有缺失值的用社区平均值插补。对于有土地但是农业收入为 0 或者缺失的家庭，采用以下办法估计土地价值：用土地价值和土地面积都非缺失的家庭计算各家平均每亩的土地价值，然后在社区层面求平均值，得到社区层面每亩土地的平均价值，最后用家庭的土地面积乘以每亩土地价值得到估计值。如果无法通过社区平均值估计，则使用跟社区一样的办法用所在省平均值估计。

（三）2010 年房产缺失值处理

1. 现住房市价缺失值的处理

2010 年现住房市价“不知道”和“拒绝回答”的比例较高，占 25.34%，我们主要采取四种方法来填补这些缺失值：首先优先用 2012 年的值来估计 2010 年的值；如果无法通过这种方式估计值，则根据社区房价平均值估计；仍然存在缺失的情形，则用居委会填答的本社区商品房单位面积价格；通过以上方法都无法估计的，最后用所在省份平均值估计。

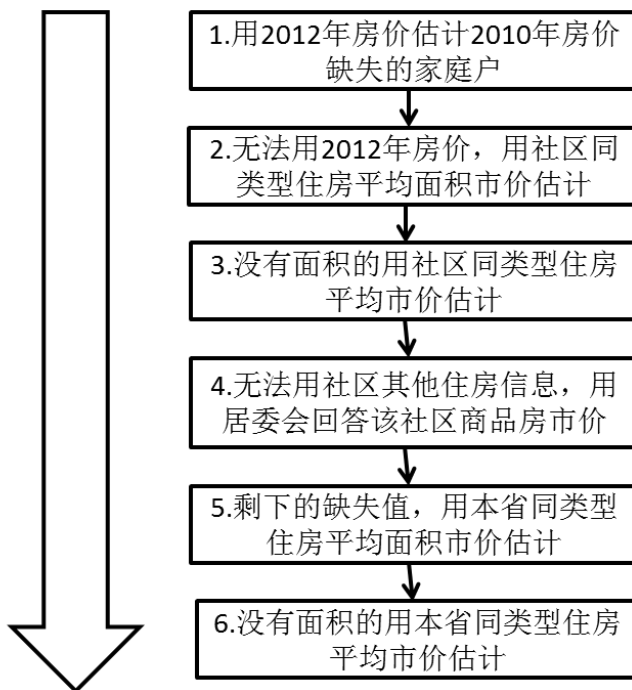
2012 年回答住房市价的比例更高，而且在 2012 年调查中，虽然也有一部分人回答“不知道”或“拒绝回答”，但通过 unfolding 的询问方式使得绝大部分家户的房价市值都能获得。因为城市有房价上涨问题，所以用 2012 年该家户的房价估计时将城乡分开进行。具体做法如下：2010 年、2012 年两年间该家户住房类型没有发生改变，而且该家户并没有分家现象，则使用 2012 年的房价值估计 2010 年。其中，村委会的住房用 2012 年房价除以 2010-2012 年两年的通货膨胀系数得到 2010 年房价估计值；居委会住房则用 2012 年份房价和国家统计局公布的 2010-2012 年房价上涨比例得到。

无法通过 2012 年房价估计 2010 年的家户则用社区平均值估计。具体做法为：用非缺失的现住房市价和现住房面积相除，得到单位面积的市价。接着，按社区和住房类型（平房、单元房、小楼房、别墅、联排别墅和其他）计算平均值，即得到每个社区每类住房的平均单位面积房价（2010 年共调查了 635 个社区）。然后，对于不知道或拒绝回答房价的家户，用其住房面积乘以估计的单位面积房价得到估计的住房价值。经过这一轮估计后，还存在几种情况导致房产价格无法被估计：一是该住房面积缺失，二是社区内某一类住房房价都缺失，导致该类住房无法通过社区平均值估计。

对于住房面积缺失的家户，进一步直接计算每个社区每类房子的平均价格，用社区同类型房子的均价作为估计值。对于社区内某类房子类型价格都缺失的情况，则使用社区居委会填答的该社区内商品房单位面积价格估计。对于商品房价格缺失的社区，计算出每个省份每一类住房的平均值，使用该平均值填补房价缺失的家庭户。

2010 年缺失值估计主要流程图如下所示：

对于房价缺失值的估计，从上到下，为依次优先使用的方法



2. 不合理值的处理

(1) 房产为 0 的值的处理

另外，2010 年有 105 户家庭有住房但住房市价填的是 0。对这部分家户，我们也使用以上估计“拒绝回答”或“不知道”的方法来估计这些家庭的住房价值。

(2) 与 2012 年对比不合理值的修改

同时，还与 2012 年对比修改了一些不合理的值，其仅体现在最后的综合变量——现住房市价 `resivalue` 上，而不改变 `fd4` 和 `fd4_best` 的值，具体方法用 `stata` 代码表示如下：

- `replace resivalue=. if fid12==130323` 【判断依据：居委会小楼房不可能是 150 元，但又不能确定具体房价】
- `replace resivalue=3000*100 if fid12==411333` 【判断依据：居委会的小楼房，根据 2012 年房价调整】
- `replace resivalue=200*1000 if fid12==620334` 【判断依据：居委会的平房不可能是 200 元，故根据 2012 年房价调整】

- `replace resivalue=. if fid12==130075` 【判断依据：该家庭户住房类型是别墅，不可能只值 200 元，但是无法确定具体价值】
- `replace resivalue=120*1000 if fid12==420323`
- `replace resivalue=500000 if fid12==440149` 【判断依据：与周围房子价格对比了下判断应该是 50 万】
- `replace resivalue=. if fid12==520572` 【判断依据：该家户住房类型是别墅，不可能只值 160 元，但是无法确定具体价值】

3. 其他房产缺失值的处理

在 2255 户有其他房产的家庭中，有 458 户不知道市价，7 户拒绝回答。对于其他房产的估计跟现住地的房子有一定差别，因为可能出现流动人口在外地打工租房住但是在老家有房产的情况。因此，对于其他房产的估计考虑了户口所在地和现居住地因素。

基本处理方法如下：因为可能存在家中成员户籍所在地不一致的情况，我们以家中主事者（变量 `tb7`）代表这一家的户口性质（农业户口和非农户口）和户口所在地。对于家中主事者变量缺失的家户，近似认为家庭成员中序号排位为 1 的人是主事者。首先，根据各家庭自报的房价（`fd4`）计算出各省分城乡的房产均价。接着，根据家中主事者的户口所在省份分为三类：（1）户口所在省份跟现居住省份一致并且居住地的城乡性质跟户口所在地的城乡性质一致的；（2）户口所在省份跟现居住地省份一致但是居住地的城乡性质跟户口所在地不一致的，比如省内发生了城乡流动的人；（3）户口所在省份跟现居住省份不一致的家户。

对于第（1）类，用现住房价值估计其他房产价值。填答了其他房产的面积，用现居住地社区内相应住房类型的平均值——单位面积的房产价格（`value_persq`）乘以面积；如果其他房产的面积未知，则直接用现住房的价格乘以其他房产的套数。

对于第（2）类，非农户口的家庭，用本省城市的房产均价乘以该家户其他房产套数。农村户口的家庭，用本省农村的房产均价乘以该家户其他房产套数。

对于第（3）类，用其所对应的户口性质和户籍所在地的房产均价代替。例如，调查中的一个家庭在北京居住，但是老家是湖北的，而且是农业人口，那么该家庭的其他房产就以湖北省农村平均房产代替，如果有多套，则乘以相应套数。

4. 其他类财产缺失值的处理

2010 年数据中，存款、股票、基金、收藏品价值以及其他资产都有缺失情况，我们用同一社区内家庭人均收入相似的家庭对应的值填补。缺失值最多的两个变量是存款和其他资产，其中存款有 780 个家庭户“拒绝回答”或“不知道”，其他资产有 767 户家庭“拒绝回答”或“不知道”。依据家庭人均收入信息，分别填补了 619 和 551 个值。股票和基金首先用该家户回答的去年本金总额填补，其余的缺失值根据家庭人均收入信息填补。

对于公司资产，如果回答“不知道”或“拒绝回答”，但是填答了公司的税后纯利润，则以该值代替。

土地价值用农业收入估计，对于有土地但是农业收入为 0 或者缺失的家庭，采用以下办法估计土地价值：用土地价值和土地面积都非缺失的家庭计算各家平均每亩的土地价值，然后在社区层面求平均值，得到社区层面每亩土地的平均价值，最后用家庭的土地面积乘以每亩土地价值得到估计值。如果无法通过社区平均值估计，则使用跟社区一样的办法用所在省平均值估计。

三、综合变量

家庭财产主要由以下几类组成：土地、房产、金融资产、生产性固定资产、耐用消费品。家庭负债包括：住房负债和非住房负债。

土地是否应该纳入家庭资产是一个有争议的话题，争论的焦点在于土地是农民从国家或集体承包过来的，不能随意买卖（陈宗胜，2000），而有研究者强调，土地的收益权和剩余索取权归农民所有，至少部分产权是属于农民的（李实，2005）。土地是大部分农民赖以维生的主要生产资料，不纳入财产计算则难以反映农村真实的财产分配状况，因此本文采用后者观点，在财产计算中包括了土地价值。土地价值的估算使用 McKinney（1993）年提

出的办法，即假定家庭农业总收入的 25% 来源于土地，而土地的收益率为 8%，从而测算出土地价值。

房产包括现住房和其他房产总价值。现住房中有全部产权和部分产权之分，部分产权是指允许家庭使用、出租和继承房产，或者允许将房产用作抵押品。与全部产权的区别在于：部分产权房在五年内不能卖给他人，出售房屋时，售房的收入按个人与单位或政府各自所占的产权比例进行分配（Walder & He, 2014）。因为无法确定部分产权的比例而且家庭具有永久的使用权，所以部分产权房价值按全部产权计算。

金融资产则包括存款、股票、基金、债券、金融衍生品、其他金融产品及借款。生产性固定资产则包括经营性企业资产、农业使用机械等。耐用消费品包含汽车、电视、电脑、冰箱等一些家庭常见的消费品。非住房负债指的是来源于教育、医疗等方面的债务。住房负债则是调查时点受访者自报的“连本带息尚未还清的房贷”。这些变量最后被纳入了数据库，为方便研究者使用，具体变量名称及说明如下：

表 1 2012 年财产类综合变量说明

变量名	内容	备注
land_asset	土地资产（元）	根据农业收入估算
houseasset_gross	家庭总房产（元）	resivalue+otherhousevalue
houseasset_net	家庭净房产（元）	resivalue+otherhousevalue-house_debts
resivalue	现住房价值（元）	超过 1000 万的，录音纠正；超过 1000 万的，逻辑修改；插补了缺失值；房价为 0 元的修改；跟 2010 年对比修改不合理的值
houseprice1_best	现住房价值（元）	以 fq4a_best, fq4a, fq4b, fq4c, fq4d, fq4e 为基础生成的以元为单位的现住房市价
otherhousevalue	其他房产价值（元）	超过 1000 万的，通过回听录音和用其他信息对比修改；插补了缺失值；房价为 0 元的修改
houseprice2_a_1_best~houseprice2_a_6_best	离您家渐远的第 N 套房产（元）	在 fr2a_a_1_best~fr2a_a_6_best 以及 fr2b_a_1~fr2b_a_6.....fr2e_a_1~fr2e_a_6 基础上生成的以元为单位的房子市价
house_debts	总房贷（元）	house1_debts+ houseother_debts
house1_debts	现住房房贷（元）	对不合理值进行了修正
houseother_debts	其他房产贷款总额（元）	对不合理值进行了修正
fixed_asset	生产性固定资产（元）	company+agrimachine

company	经营资产（元）	超过 1000 万的，通过回听录音和用其他信息对比修改；插补了缺失值
agrimachine	农用器械价值（元）	插补了缺失值
finance_asset	家庭总金融资产（元）	Savings+ govbond+ stock+ funds+ derivative+ otherfinance+ debit_other
savings	现金和存款总值（元）	插补了缺失值
govbond	政府债券（元）	插补了缺失值
stock	股票（元）	插补了缺失值
funds	基金（元）	插补了缺失值
derivative	金融衍生品（元）	
otherfinance	其他金融产品（元）	
debit_other	别人欠自己家的钱（元）	
nonhousing_debts	非房贷负债（元）	bank_debts+ind_debts
bank_debts	银行贷款（元）	
ind_debts	非金融机构贷款额（元）	
durables_asset	耐用消费品价值（元）	
total_asset	家庭净资产（元）	所有资产加总

表 2 2010 年财产类综合变量说明

变量名	内容	备注
land_asset	土地资产（元）	根据农业收入估算
houseasset_gross	家庭总房产（元）	resivalue+otherhousevalue
houseasset_net	家庭净房产（元）	resivalue+otherhousevalue-house_debts
resivalue	现住房价值（元）	超过 1000 万的，录音纠正；超过 1000 万的，逻辑修改；插补了缺失值；房价为 0 元的修改；跟 2010 年对比修改不合理的值
otherhousevalue	其他房产价值（元）	超过 1000 万的，用家中其他信息对比进行修改；插补了缺失值
house_debts	总房贷（元）	对应 fh203_a_1
company	经营资产（元）	超过 1000 万的，通过回听录音和用其他信息对比修改；插补了缺失值
valuable	家中收藏品（元）	插补了缺失值
otherasset	其他资产现值（元）	插补了缺失值
financeasset_gross	家庭总金融资产（元）	Savings+ govbond+ stock+ funds+ derivative+ otherfinance+ debit_other
savings	现金和存款总值（元）	插补了缺失值
stock	股票（元）	插补了缺失值
funds	基金（元）	插补了缺失值
debit_other	别人欠自己家的钱（元）	
nonhousing_debts	非房贷负债（元）	edu_debts+consu_debts+medi_debts+li

		fe_debts+other_debts
edu_debts	教育借贷（元）	对应 fh203_a_2
consu_debts	贷款用于购买消费品（元）	对应 fh203_a_3
medi_debts	贷款用于家庭成员治病（元）	对应 fh203_a_4
life_debts	贷款用于家庭日常生活开支	对应 fh203_a_5
other_debts	贷款用于其他方面	对应 fh203_a_6
total_asset	家庭净资产（元）	所有资产加总

四、我国家庭财产状况

以下用 2012 年和 2010 年两年数据简要分析我国家庭财产状况、分布、结构及两年内的变化。

在使用 2012 年数据估计我国家庭财产存量和分布时，为了调整顶端极富人群的分布，使用外部数据《2012 雅居乐海南清水湾胡润百富榜》³对顶端极富人群进行了调整。具体做法如下：首先，利用《2012 雅居乐海南清水湾胡润百富榜》数据拟合帕累托函数，根据该函数估计顶端家庭的财产，对最顶端的家庭财产则使用排行榜的实际值，二者结合得到顶端 0.1%家庭的财产；其次，对于其余 99.9%的家庭，使用调查数据，根据抽样权数将调查样本扩充为总体 99.9%的家庭；最后，将经过权数调整后的调查数据与帕累托函数拟合的数据以及“胡润百富榜”中最顶端家庭的实际财产数据合并，形成全国家庭的总体数据。

表 3 展现了全国家庭财产的存量。该表同时保留了仅使用 CFPS 调查数据直接计算的未调整结果和使用外部数据调整后的结果。在使用外部数据调整后，我们得到 2012 年全国家庭净资产均值为 42.15 万元，中位值为 16.79 万元。在各个分位数的分布上，2012 年全国有 25%的家庭其财产小于 6.33 万元，75%的家庭小于 33.07 万元，顶端 10%的家庭其财产高于 69.17 万元，顶端 5%的家庭高于 112.78 万元。

表 3 2012 年全国家庭净资产 (单位：万元)

	平均值	25%	50% (中位值)	75%	90%	95%
未调整	32.19	6.33	15.75	32.97	68.66	111.49
调整后	42.15	6.33	15.79	33.07	69.17	112.78

³ 数据来源：《2012 雅居乐海南清水湾胡润百富榜》(<http://www.hurun.net/CN/HuList.aspx?nid=14>)。

表 4 显示，我国财富不平等程度很高，大量的财富集中在极少数人的手中。将家庭财产按从低到高排序，排名在 25% 以下的家庭的财产总量仅占全国财产总量的 1.23%，排名在 50% 以下的家庭的财产总量仅占全国财产总量的 7.54%。然而，排名在顶端 25% 的家庭拥有全国 78.79% 的财产，排名在顶端 10% 的家庭拥有全国 62.10% 的财产，排名在顶端 5% 的家庭拥有全国一半以上的财产，排名在顶端 1% 的精英阶层拥有全国三分之一以上的财产。由此可见，中国的财产分布呈现出明显的两极分化。

90/10 比率体现的是财产两极的差距。2012 年全国财产的 90/10 比率高达 32.94，说明排名在 90% 分位数上的家庭财产是排名在 10% 分位数上的家庭财产的近 33 倍，而同期收入的 90/10 比率是 13.10（谢宇等，2013），可见，财产分配呈现出更加严峻的贫富两极分化，比收入分配更不平等。

表 4 2012 年全国的财产分布

	0—25% (%)	0—50% (%)	75—100% (%)	90—100% (%)	95—100% (%)	99—100% (%)	基尼 系数	90/10 比率
未调整	1.61	9.85	72.29	50.50	37.04	16.13	0.64	32.69
调整后	1.23	7.54	78.79	62.10	51.71	35.30	0.73	32.94

在分析财产存量和分布时，同时展示了未调整和调整后的数据，并用到了全国家庭层面的权数。接下来的分析因为无法获得外部数据调整，所以直接用 CFPS 调查数据分析。考虑到辽宁、上海、河南、甘肃、广东五个省份是过度抽样，这里使用具有全国代表性的样本分析。

在比较城乡差异时，财产水平和分布未进行极端值的调整，这会导致分析结果尤其是城镇的平均财产水平和分布出现偏差。但我们关注的是城乡差异和特征，使用没有调整的数据并不影响结论。实际上，调整后的结果更有利于加强结论，因为绝大多数财产极大值都分布在城镇。表 5 显示，城镇的财产存量远远高于农村。同时，城镇的财产不平等程度也高于农村。从分位数上看，农村底端家庭财产总量占有所有农村家庭财产总量的比例要高于相应的城镇家庭，比如农村财产总量排名 50% 以下的家庭占全国农村家庭财产总量的比例为 13.15%，高出城镇 2.91 个百分点。而农村顶端家庭财产总量占全国农村家庭财产总量的比例要低于相应的城镇家庭，比如财产总量排名在顶端 10% 的农村家庭拥有全国农村家庭 43.47% 的财

产，比城镇低了 5.48 个百分点。另外，农村的 90/10 比率也远远低于城镇，农村为 19.05，而城镇高达 44.68。

表 5 2012 年分城乡的家庭净财产及分布

	平均净财产 (万元)	0—25% (%)	0—50% (%)	75—100% (%)	90—100% (%)	95—100% (%)	90/10 比率
城镇	44.04	1.54	10.24	71.62	48.95	35.45	44.68
农村	18.65	2.69	13.15	65.90	43.47	31.30	19.05

在我国家庭财产的构成中，房产占绝对主导地位。如表 6 所示，房产在全国家庭平均财产中占了 74.29%。金融资产占家庭财产的 10% 左右。另外，生产性固定资产占家庭财产的 8% 左右，耐用消费品占 6% 左右。非住房负债比例略高于住房负债。但总的来说，除房产外，其他各项财产占家庭财产的比例都较小。城镇和农村家庭的财产结构有所不同。城镇家庭的房产占家庭总财产的比例远高于农村家庭，约为 80%，比农村家庭高了 17 个百分点左右。土地是农村家庭一项非常重要的资产类型，占家庭财产比例的 20.04%。⁴城镇家庭金融资产所占的比例略高于农村家庭，而生产性固定资产和耐用消费品所占的比例小于农村家庭。在负债上，城镇家庭的住房负债和非住房负债比例相当，而农村家庭的债务主要体现在非住房负债上。城镇的住房负债大于农村，但农村整体负债比例要高于城镇。

表 6 2012 年全国及分城乡的家庭财产构成 (单位：%)

财产结构	全国	城镇 ⁵	农村
土地	8.19	2.74	20.04
房产	74.29	79.47	62.54
金融资产	10.11	10.52	9.31
生产性固定资产	7.84	6.97	9.85
耐用消费品	5.69	5.58	5.83
住房负债	-2.22	-2.34	-1.76
非住房负债	-3.90	-2.95	-5.82

⁴ 土地资产的估算方法是用 McKinley 和 Griffin (1993) 提出的办法，假定家庭农业总收入的 25% 来源于土地，而土地的收益率为 8%，从而估算出土地价值。

⁵ 城镇之所以有土地资产，是因为城镇中也有一部分从农村进入城镇的流动人口，或者部分地区虽然在行政区划上被划归为城镇，但居民实际从事的仍是农业劳动。

表 7 对比了不同分位数下家庭平均净财产从 2010 到 2012 年的变化情况。全国家庭平均净财产呈明显上升趋势。家庭财产水平越低的家庭增长幅度越高，如家庭财产在全国排序 25% 以下的家庭，其平均财产增长幅度高达 62%，而排序在 50—75% 的家庭平均财产增长幅度为 25%。各个分位数上城镇和农村家庭的财产都出现了增长，但增长模式有所不同。对于底端 0—25% 的家庭，农村家庭财产的增长速度明显慢于城镇。这样的增长模式会使农村底端家庭处于更加不利的地位，拉大了农村家庭与城镇家庭的整体差距。

表 7 各个分位数上的家庭平均净财产 (单位: 万元)

	全国			城镇			农村		
	2010	2012	增长 (%)	2010	2012	增长 (%)	2010	2012	增长 (%)
0—25%	1.3	2.1	62.0	1.6	2.7	68.2	1.2	2.0	58.9
25—50%	7.8	10.2	30.7	12.0	15.3	27.8	5.8	7.8	33.3
50—75%	17.6	22.0	25.1	27.1	31.9	17.6	11.9	15.6	30.9
75—100%	77.2	88.8	14.9	108.4	126.4	16.6	41.6	49.1	18.1
合计	26.0	30.7	18.4	37.3	44.1	18.2	15.1	18.6	22.9

表 8 将 2010 年和 2012 年的家庭按财产从低到高排序，分为 0—25%、25—50%、50—75% 和 75—100% 四组，分析从 2010 年到 2012 年家庭相对财产的变化。对角线上加粗的数字表示两年间家庭财产的相对位置未发生变化，对角线以上表示向上流动，对角线以下表示向下流动。

可以看到，2010 年家庭财产排序处于 25% 以下、25—50%、50—75%、75—100% 四个水平上的家庭在 2012 年分别有 57.88%、41.50%、43.27% 和 64.97% 保留相对位置不变。底端 25% 的家庭和顶端 25% 的家庭在 2012 年保持相对位置不变的比例更高，反映出穷人阶层和富人阶层的流动性更小。

分城乡看，城镇和农村表现出同样的规律，但城镇家庭能够保持相对位置不变的比例高于农村，而且财产水平越高，城乡之间保持相对位置不变的家庭所占的比例差距越大。2010 年城镇家庭财产排在 25% 以下的家庭有 60.14% 在 2012 年仍保持在这个水平，农村则为 55.42%，二者相差 4.72 个百分点；而在 75—100% 财产水平上，城镇为 66.14%，农村为 56.33%，二者相差达 9.81 个百分点。所以，城镇的家庭财产更为稳定，而且富人阶层固化更严重。

表 8 2010—2012 年家庭相对财产流动

(单位: %)

家庭财产分位数	2012 年家庭平均净财产分组			
	0—25%	25—50%	50—75%	75—100%
全国				
0—25%	57.88	25.12	11.13	5.88
25—50%	25.85	41.50	23.61	9.05
50—75%	10.71	25.9	43.27	20.12
75—100%	5.56	7.48	21.99	64.97
2010 城镇				
年家 0—25%	60.14	23.76	10.36	5.74
庭平 25—50%	24.97	44.77	21.82	8.44
均净 50—75%	9.01	24.89	46.40	19.71
财产 75—100%	5.85	6.64	21.37	66.14
分组 农村				
0—25%	55.42	25.92	12.77	5.90
25—50%	26.28	37.68	24.06	11.98
50—75%	11.40	24.73	38.07	25.80
75—100%	6.86	11.69	25.12	56.33

参考文献

McKinley, Terry and Keith Griffin. 1993. "The Distribution of Land in Rural China." *The Journal of Peasant Studies* 21(1): 71—84.

Walder, Andrew G. and Xiaobin He. 2014. "Public Housing into Private Assets: Wealth Creation in Urban China." *Social Science Research* 46: 85—99.