

Consumption and poverty of older Chinese: 2011-2020

Jinquan Gong

Gewei Wang

Yafeng Wang

Yaohui Zhao

Appendix A - Imputation of missing consumption values

With expenditures items totaling more than twenty, missing values become a major issue. In CHARLS, we first ask whether the household has an expenditure on a certain item. If the answer is no, then the expenditure is naturally 0. However, if a household reported yes to an item but could not remember the amount of money or refused to answer, such values should be missing, not 0. As a result, missing one item leads to missing for the whole household.

Missing values for expenditure items are common in household surveys. Although refusing the answer in expenditure surveys should theoretically be less than in income surveys, retrospectively recalling expenses may still be difficult, especially given cognitive decline among older people. In some surveys, households are asked to keep a diary of all expenses, but such a practice may usher in other biases, especially since many older people are illiterate.

In our data, many households had at least one missing value in the expenditure items, 17.4% in 2011, 37.6% in 2013, 35.9% in 2015, 17.8% in 2018, and 23.8% in 2020 (Table A1). Therefore, imputation of missing values is necessary to retain the representation of the sample. Fortunately, as revealed in Table A1, in most cases, the number of missing items is one or two, and those missing more than 12 items accounted for less than 10%. The missing data problem varied across years; the best was 2011, 2018 and 2020, when at least 92% of all households had less than three items missing; the worst was in 2013, when 9.3% of households skipped the whole expenditure module, followed by 2015 when 6.6% of households skipped it. The missing item problem significantly improved in more recent waves, reflecting better survey management. In the 2020 wave, only 0.77% of the households have twelve or more missing items.

We impute consumption values for households that had missing items based on information about the non-missing portion of the consumption and the imputed consumption share of the

missing portion. The basic idea for imputing the consumption share is to assign the community median to the households that need imputation.

Table A1. Household shares with missings by consumption item, 2011-2020

Number of missings	2011		2013		2015		2018		2020	
	%	cum. %	%	cum. %	%	cum. %	%	cum. %	%	cum. %
0	82.6%	82.6%	62.4%	62.4%	64.1%	64.1%	82.2%	82.2%	76.2%	76.2%
1	8.0%	90.6%	13.5%	76.0%	14.0%	78.1%	8.3%	90.5%	11.2%	87.4%
2	2.7%	93.3%	5.9%	81.9%	6.6%	84.8%	3.2%	93.6%	5.0%	92.4%
3	1.3%	94.6%	3.2%	85.1%	3.0%	87.8%	2.1%	95.7%	2.7%	95.2%
4	0.4%	95.0%	1.9%	87.0%	1.7%	89.4%	1.2%	97.0%	1.8%	97.0%
5	0.3%	95.3%	0.9%	87.8%	1.1%	90.6%	0.6%	97.5%	0.8%	97.8%
6	0.2%	95.5%	0.5%	88.3%	0.7%	91.3%	0.3%	97.9%	0.5%	98.3%
7	0.1%	95.7%	0.4%	88.8%	0.5%	91.8%	0.3%	98.1%	0.4%	98.7%
8	0.1%	95.8%	0.3%	89.0%	0.4%	92.2%	0.1%	98.2%	0.2%	98.9%
9	0.4%	96.2%	0.4%	89.4%	0.3%	92.5%	0.1%	98.3%	0.2%	99.1%
10	0.1%	96.3%	0.3%	89.7%	0.2%	92.7%	0.1%	98.3%	0.1%	99.2%
11	0.0%	96.3%	0.2%	89.9%	0.1%	92.8%	0.0%	98.4%	0.0%	99.2%
12	0.1%	96.4%	0.0%	89.9%	0.1%	92.9%	0.0%	98.4%	0.0%	99.3%
13	0.0%	96.4%	0.1%	90.0%	0.1%	93.0%	0.0%	98.4%	0.1%	99.4%
14	0.2%	96.6%	0.1%	90.1%	0.0%	93.0%	0.0%	98.5%	0.0%	99.4%
15	0.0%	96.6%	0.0%	90.2%	0.0%	93.0%	0.0%	98.5%	0.0%	99.4%
16	0.2%	96.8%	0.3%	90.5%	0.1%	93.2%	0.0%	98.5%	0.0%	99.4%
17	0.0%	96.9%	0.1%	90.6%	0.1%	93.3%	0.0%	98.5%	0.0%	99.4%
18	0.0%	96.9%	0.1%	90.7%	0.1%	93.3%	0.0%	98.5%	0.0%	99.5%
19	0.0%	96.9%	0.0%	90.7%	0.0%	93.4%	0.0%	98.5%	0.0%	99.5%
20	3.1%	100.0%	0.0%	90.7%	0.0%	93.4%	0.0%	98.5%	0.0%	99.5%
21			9.3%	100.0%	6.6%	100.0%	0.0%	98.5%	0.0%	99.5%
22							1.5%	100.0%	0.0%	99.5%
23									0.5%	100.0%
Total sample households	10,257		11,000		11,797		10,524		10,204	

Notes: Calculated from CHARLS data. Each column reports the household shares by the number of missing consumption items. All items are from the expenditure submodule of the questionnaire except for rental payment. Rental payment before 2018 is from the housing module. Medical expenses are excluded. Cumulative percentages are shown in the second column for each wave.

A formal derivation is as follows. For an item i consumed by household j in community k , denote c_{ijk} the expenditure on item i . Define the total household consumption

$$c_{jk} = \sum_i c_{ijk}. \quad (1)$$

Denote the set of household j 's missing items by \mathcal{M}_{jk} . $c_{jk}^{NM} \stackrel{\text{def}}{=} \sum_{i \notin \mathcal{M}_{jk}} c_{ijk}$ is j 's total non-missing consumption that we actually observe. $c_{jk}^M \stackrel{\text{def}}{=} \sum_{i \in \mathcal{M}_{jk}} c_{ijk}$ is j 's total missing consumption that we don't observe. We have

$$c_{jk} = \sum_i c_{ijk} = c_{jk}^M + c_{jk}^{NM}. \quad (2)$$

Let the share of all missing items be $s_{jk}^M \stackrel{\text{def}}{=} c_{jk}^M / c_{jk}$ and the non-missing items be $s_{jk}^{NM} \stackrel{\text{def}}{=} c_{jk}^{NM} / c_{jk}$, and $s_{jk}^M + s_{jk}^{NM} = 1$, we have

$$c_{jk} = s_{jk}^M c_{jk} + s_{jk}^{NM} c_{jk} = s_{jk}^M c_{jk} + c_{jk}^{NM}. \quad (3)$$

The second term is known, so our goal is to estimate c_{jk} by finding a proxy for s_{jk}^M .

Our strategy is to assign the community (median) shares among households that do not have non-missing items, \hat{s}_{jk}^M , to the household. Denote s_{ik} the median of the community share for item i ,

$$\hat{s}_{jk}^M = \sum_{i \in \mathcal{M}_{jk}} s_{ik} \quad (4)$$

When only one item is missing, then the community median share for this item is used. In cases where more than one item is missing, we aggregate community median shares of all missing items and assign the total share to the set of all missing items \mathcal{M}_{jk} . Once this is done, we derive the total consumption c_{jk} for the household.

From Equation (3), we have

$$c_{jk} - \hat{s}_{jk}^M c_{jk} = c_{jk}^{NM} \quad (5)$$

thus,

$$\hat{c}_{jk} = \frac{c_{jk}^{NM}}{1 - \hat{s}_{jk}^M} \quad (6)$$

Our imputation method, i.e., assigning the community-median share to the household's missing items, may have consequences for poverty analysis. If the missing pattern is independent

of living standards, then imputation does not affect our analysis; otherwise, it may. There are reasons to believe that missing may not be random. For example, rich people may be more likely to hide their luxurious consumption for fear of revealing their wealth, or the poor may not remember their expenses due to worse cognitive health.

Although missing is non-random, we posit that the problem is not serious for average households because their consumption shares are likely conforming to the community averages and unlikely to cause major biases in the estimation. To understand biases for poor and rich households, we consider two cases in which either a necessary good is missing or a luxury good is missing. We are mainly concerned about whether the poor status is changed.

Scenario 1: Food consumption is missing.

For poor households. Due to Engel's Law, a poor household spends a larger share of consumption on food; thus, the community food share is likely an underestimation, and our imputed total consumption will be underestimated, making the household poorer. Therefore, our imputation is likely to cause some low-income households to fall below the poverty line, exaggerating the number of poor.

For rich households. A rich household spends a smaller share of consumption on food; thus, the community food share is likely an overestimation, and our imputed total consumption will be overestimated, making the household richer. Therefore, a rich household will stay rich; no impact on poverty is expected.

Scenario 2: Luxury good consumption is missing.

For poor households. We expect the community's luxurious goods share to be higher than poor households. Thus, our imputed total consumption for the poor household will be overestimated. If the poor have many luxury consumptions missing, this will artificially lift them out of poverty. Otherwise, it is unlikely to be an issue.

For rich households. A rich household spends a larger share of consumption on luxurious items; thus, the community share is likely an underestimation, and our imputed total consumption will be underestimated, making the household less rich. Despite this underestimation, a rich household is unlikely to fall below the poverty line because of the imputation.

Because of the above problems with missing imputations, we choose to limit the scope of imputation to households that have less than 12 items missing.¹To understand the extent of remaining biases, we examine in Table A2 whether rich or poor households tend to have missing items and whether they are more or less likely to miss consumption of food and luxury goods. These exercises are based on the subset of households for which consumption is not missing or successfully imputed, i.e., those with 12 or more missing items are excluded. We control for community fixed effects to consider within-community variations of item missing because we impute a share of missing items from the community median. Column (1) examines the missing patterns of food consumption, and Column (2) shows that for the beauty products or services which represent luxury goods. We find that the literate but not finishing elementary school had a little more propensity to have missing food items, but the association is not strong. Other than that, no education gradient can be found in missing food or missing beauty items. Based on these results, it appears that the imputation is unlikely to cause major biases in consumption or poverty estimations.

Figure A1 compares our imputed total consumption distributions against the distribution using only the sample with no missing items. There is barely any difference for waves 2011, 2018, and 2020. The imputations for 2013 and 2015 have shifted the distribution a little leftward, yielding a little higher poverty rates.

Some household expenditures cannot be imputed successfully either because all community households had missings ($n = 17$) or because the value of a household's non-missing consumption is zero ($n = 93$). In total, 4.13% of individuals are excluded from the imputation, mainly because their households had 12 or more missing items. Because these exclusions may cause our sample to lose national representation, we need to recalculate the weights. For this purpose, we run a probit model to predict sample inclusion. Details of the model are shown in Table A3. The inverse of the predicted inclusion probability is used to re-construct the sample weights.

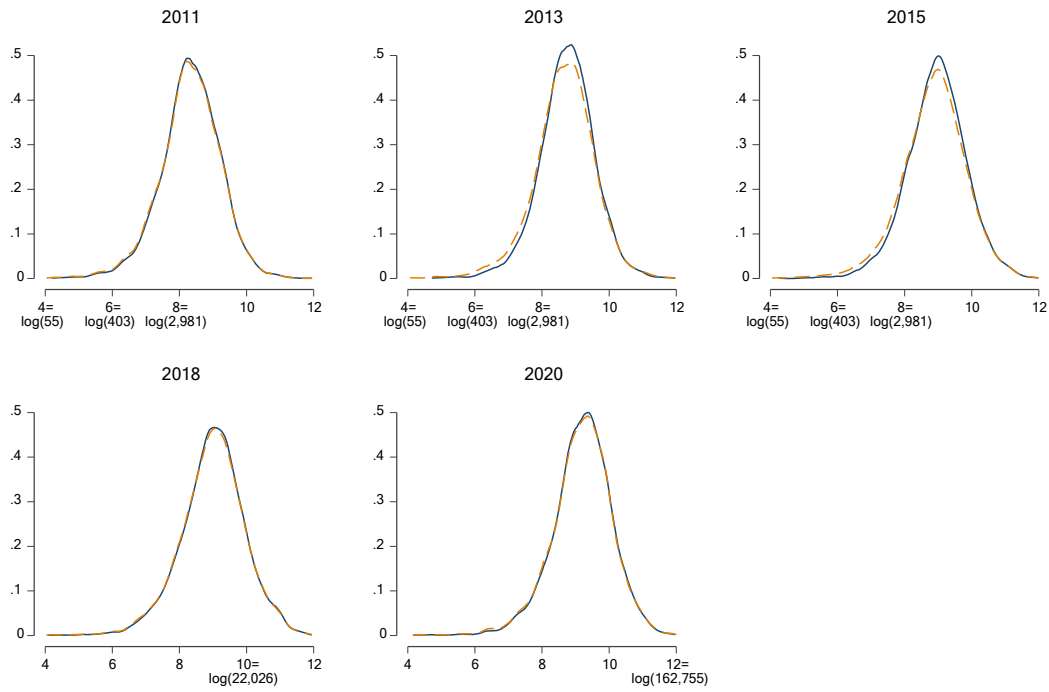
¹ We have repeated our imputations using other thresholds and the results remain similar.

Table A2. Linear Regression of Consumption Non-Response

	1 if food consumption is missing (1)	1 if beauty consumption is missing (2)
Mean of the outcome variable	0.0522	0.0176
Female	0.00290* (0.00159)	-6.45e-05 (0.000951)
Marital Status (Base: Married)		
Widowed	0.0182*** (0.00358)	0.0125*** (0.00205)
Divorced	0.0138 (0.00979)	0.0175** (0.00733)
Never married	0.000853 (0.0117)	0.0107** (0.00534)
Hukou Rural	0.0115*** (0.00386)	0.00332 (0.00270)
Age Group (Base: 60-64)		
Age Group (65-69)	0.00932*** (0.00292)	-0.000674 (0.00178)
Age Group (70-74)	0.0138*** (0.00347)	0.00159 (0.00208)
Age Group (75+)	0.0300*** (0.00402)	0.000707 (0.00234)
Education (Base: finished high school or above)		
Illiterate	0.00593 (0.00454)	-0.00411 (0.00317)
Literate but did not finish elementary school	0.0106** (0.00456)	-0.00383 (0.00321)
Finished elementary school	0.00323 (0.00428)	-0.00257 (0.00307)
Finished middle School	-0.000336 (0.00417)	-0.000222 (0.00328)
Household size	0.00255*** (0.000946)	0.0105*** (0.000762)
Wave dummy (base=2011)		
Wave 2013	0.0544*** (0.00543)	0.00748*** (0.00269)
Wave 2015	0.0630*** (0.00536)	0.0117*** (0.00250)
Wave 2018	-0.0273*** (0.00383)	0.00528** (0.00231)
Wave 2020	-0.0511*** (0.00347)	0.0141*** (0.00257)
Community FE	Yes	Yes
Observations	45,203	45,203
R-squared	0.070	0.048

Notes: Estimated from CHARLS data by OLS. *, **, *** denote significance levels at 10, 5, and 1 percent. Standard errors in parenthesis are clustered at the household level.

Figure A1. Log(Consumption) distributions before and after imputations



Note: Blue lines are densities for log(consumption) among households with no missing items. Orange lines are densities for log(consumption) after imputations.

Table A3. Probit Regression of Household Inclusion

	(1) 2011	(2) 2013	(3) 2015	(4) 2018	(5) 2020
Female	-0.237*** (0.060)	-0.209*** (0.039)	-0.219*** (0.041)	-0.154** (0.061)	-0.227*** (0.083)
Age Group (Base: 60-64)					
Age Group 65-69	-0.113 (0.090)	-0.032 (0.067)	-0.137** (0.069)	-0.129 (0.104)	0.117 (0.172)
Age Group 70-74	-0.122 (0.100)	-0.371*** (0.070)	-0.325*** (0.073)	-0.232** (0.118)	0.504** (0.207)
Age Group 75-79	-0.557*** (0.089)	-1.082*** (0.065)	-1.168*** (0.061)	-0.618*** (0.099)	-0.327** (0.140)
Education (Base: finished high school or above)					
Illiterate	0.380*** (0.127)	-0.167 (0.102)	-0.450*** (0.108)	-0.043 (0.133)	-0.162 (0.230)
Literate but not having finished elementary school	0.339*** (0.126)	-0.006 (0.103)	-0.419*** (0.116)	0.134 (0.146)	-0.192 (0.245)
Finished elementary school	0.154 (0.121)	0.089 (0.100)	-0.100 (0.107)	0.126 (0.139)	-0.182 (0.223)
Finished middle School	0.095 (0.132)	0.118 (0.104)	-0.164 (0.114)	0.176 (0.152)	-0.117 (0.237)
Education Missing	-1.079*** (0.300)	-0.742 (0.729)	-0.559 (0.528)		
Constant	2.289*** (0.296)	2.223*** (0.507)	2.322*** (0.287)	2.403*** (0.233)	3.296*** (0.355)
Community FE#	Yes	Yes	Yes	Yes	Yes
Observations	7,412	8,433	9,734	10,413	11,246

Notes: Estimated from CHARLS data by probit separately for each wave. *, **, *** denote significance levels at 10, 5, and 1 percent. Standard errors in parenthesis are clustered at the household level. #: In some cases where no variations exist in a community, we used county fixed effects.

Appendix B - Additional tables and figures

Table B1. Expenditure items in CHARLS questionnaires, 2011-2020

	Defined items in the questionnaire (in Chinese)	Defined items in the questionnaire (in English)	2011	2013	2015	2018	2020
1	购买食品, 包括自家生产和消费的农产品, 不包括外出就餐、购买香烟、酒水等	Purchase of food, including agricultural products produced and consumed by the family, excluding restaurants, cigarettes, alcoholic beverages, etc.	Y				
	购买食品, 不包括外出就餐、购买香烟、酒水等	Purchase of food, excluding restaurants, cigarettes, alcoholic beverages, etc.		Y	Y		
	购买食品, 不包括摆酒、办酒席、外出就餐、购买香烟、酒水等	Purchase of food, excluding banquets, restaurants, cigarettes, alcoholic beverages, etc.				Y	Y
	消费的自家生产的农产品	Home-grown agricultural products		Y	Y	Y	Y
	外出就餐	Restaurants	Y	Y	Y		
	外出就餐, 不包括摆酒和办酒席	Restaurants, excluding banquets				Y	Y
	摆酒和办酒席的支出	Banquets				Y	Y
2	香烟、酒水	Cigarettes and alcoholic beverages	Y	Y	Y	Y	Y
3	邮电、通讯支出 (包括电话、手机、上网、邮寄等)	Expenditure on post and telecommunications (including telephone, cell phone, Internet, mail, etc.)	Y	Y	Y	Y	Y
4	水费、电费	Water and electricity expenses	Y	Y	Y	Y	Y
5	燃料费 (包括煤炭、煤制品、柴草、木炭、液化气等)	Fuel expenses (including coal, coal products, firewood, charcoal, liquefied petroleum gas, etc.)	Y				
	燃料费 (包括煤炭、煤制品、柴草、木炭、天然气、液化气等)	Fuel expenses (including coal, coal products, firewood, charcoal, natural gas, liquefied petroleum gas, etc.)		Y	Y	Y	Y
6	保姆、小时工、佣人等的支出	Expenses for domestic helpers, hourly helpers, maids, etc.	Y	Y	Y	Y	Y
7	在当地的交通费	Transportation expenses in the local area	Y	Y	Y	Y	Y
8	日用品包括美容化妆品 (如洗衣粉、香皂、肥皂、牙膏、牙刷、美容化妆品等)	Household essentials including beauty cosmetics (such as laundry detergent, soap, toothpaste, toothbrush, beauty cosmetics, etc.)	Y				
	日用品 (包括洗漱用品、家居用品、厨卫用品、装饰用品等)	Household essentials (including toiletries, household items, kitchen and bathroom items, decorative items, etc.)		Y	Y	Y	Y
	防疫支出, 包括购买口罩, 防护服和消毒液	Epidemic prevention expenditures, including the purchase of masks, protective clothing and disinfectant solution					Y
9	美容支出 (包括化妆品、美容护理、按摩等)	Beauty (including cosmetics, beauty care, massage, etc.)	Y	Y	Y	Y	Y
10	文化娱乐支出 (包括书报杂志、光盘、影剧票、歌舞厅和网吧等)	Cultural and entertainment expenses (including books, newspapers, magazines, CDs, movie and theater tickets, pubs and Internet cafes, etc.)	Y	Y	Y	Y	Y
11	衣着消费	Clothing and footwear	Y	Y	Y	Y	Y
12	家庭的旅游支出 (包括旅行时坐火车、汽车、飞机、轮船的费用)	Household's travel expenses (including the cost of train, bus, airplane, and ship)	Y				
	家庭的旅游支出	Household's travel expenses		Y	Y	Y	
	家庭的旅行支出, 包括在外地工作往返老家的费用, 旅游的车费和旅馆费等	Household travel expenses, including the cost of traveling between hometown and the current residence, transport, and hotels, etc.					Y
13	家庭的取暖费支出 (指集中供暖)	Household's heating expenses (refers to central heating)	Y	Y	Y	Y	Y
14	教育和培训支出 (包括学杂费、培训费等)	Education and training expenses (including tuition and fees, training fees, etc.)	Y	Y	Y	Y	Y
15	保健费用 (包括健身锻炼及产品器械、保健品等)	Health maintenance expenses (including gym, equipment, nutrition products, etc.)	Y	Y	Y	Y	Y

16	各种交通通讯工具的购买（如自行车、电动自行车和手机）、维修及配件费用	Transportation vehicles and communication tools (such as bicycles, electric bicycles and cell phones), maintenance and accessories costs	Y					
	各种交通工具（如自行车、电动自行车等，不包括汽车）、通讯工具（如电话、手机等）的购买、维修及配件费用	Transportation vehicles (such as bicycles, electric bicycles, etc., excluding automobiles), communication tools (such as telephones, cell phones, etc.), maintenance and accessories costs		Y	Y	Y	Y	
17	物业费（包括车位费）	HOA fees (including parking)	Y	Y	Y	Y	Y	
18	社会捐助支出（包括现金，食品、衣服等）	Donations (including cash, food, clothes, etc.)	Y	Y	Y	Y	Y	
19	房子或者床位的租金，包括家属成员在外的住宿费，如住校宿舍费，不包括旅游的宾馆费	House or bed rental, including the accommodation cost of household members away from home, such as dormitory fees, excluding hotel expenses when traveling					Y	Y
20	家具和耐用消费品的支出，包括电冰箱、洗衣机、电视和钢琴等高档乐器	Furniture and durable goods, including refrigerators, washing machines, TVs and high-end musical instruments such as pianos	Y					
	家具、耐用消费品及电器的支出（包括电冰箱、洗衣机、电视、电脑和高档乐器如钢琴等）	Furniture, durable goods and electric appliances (including refrigerators, washing machines, TVs, computers and high-end musical instruments such as pianos, etc.)		Y	Y	Y		
	家具、耐用消费品及电器的购买、维修及配件费用。家具、耐用消费品及电器包括电冰箱、洗衣机、电视、电脑和高档乐器如钢琴等	Purchase, repair, and accessories of furniture, durable goods and electric appliances (including refrigerators, washing machines, TVs, computers and high-end musical instruments such as pianos, etc.)						Y
	电器（包括笔记本电脑、台式电脑和配件等）	Electric appliances (including laptops, desktop computers and accessories, etc.)	Y					
21	购买汽车	Purchase of automobiles	Y	Y	Y	Y		
	汽车的购买、维修及配件费用	Purchase, maintenance and parts of automobiles						Y
22	上交给政府相关部门的税费和杂费（不包括所得税）	Taxes and miscellaneous fees paid to the government (excluding income tax)	Y	Y	Y	Y	Y	
23	医疗支出（包括直接或间接）	Medical expenses (including direct or indirect payment)	Y	Y	Y	Y	Y	

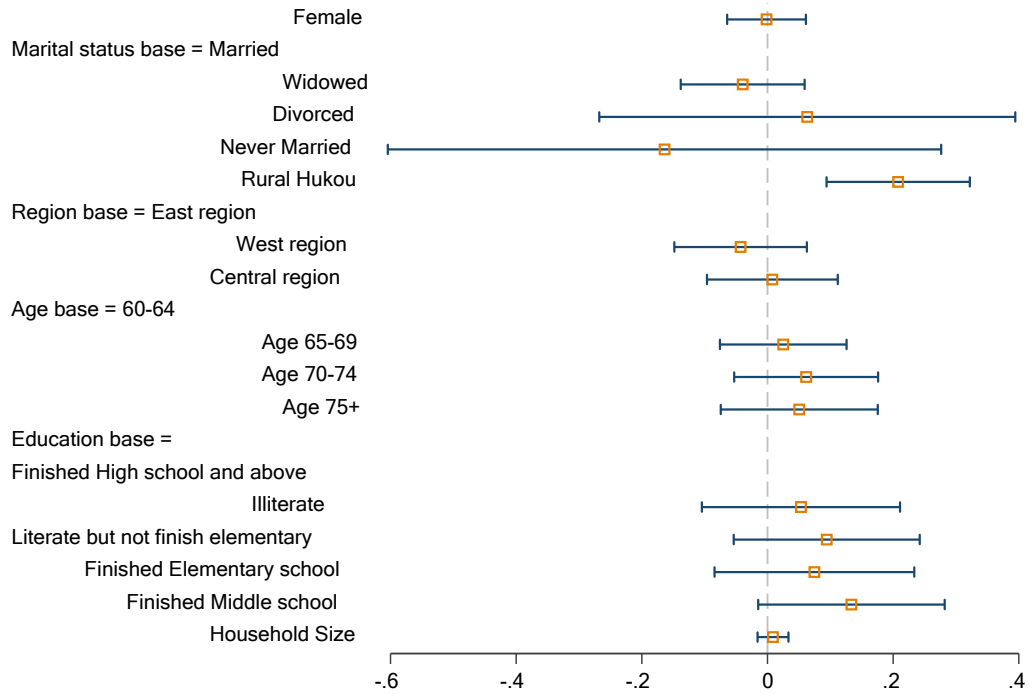
Notes: Definitions framed together are for the same item that changed at some wave. "Y" denotes the existence of an item in a wave's questionnaire.

Table B2. China's official poverty lines

	2011	2013	2015	2018	2020
Rural poverty line	2,536	2,736	2,855	2,995	3,315
Urban poverty line (by approximation)	3,581	3,863	4,031	4,229	4,681

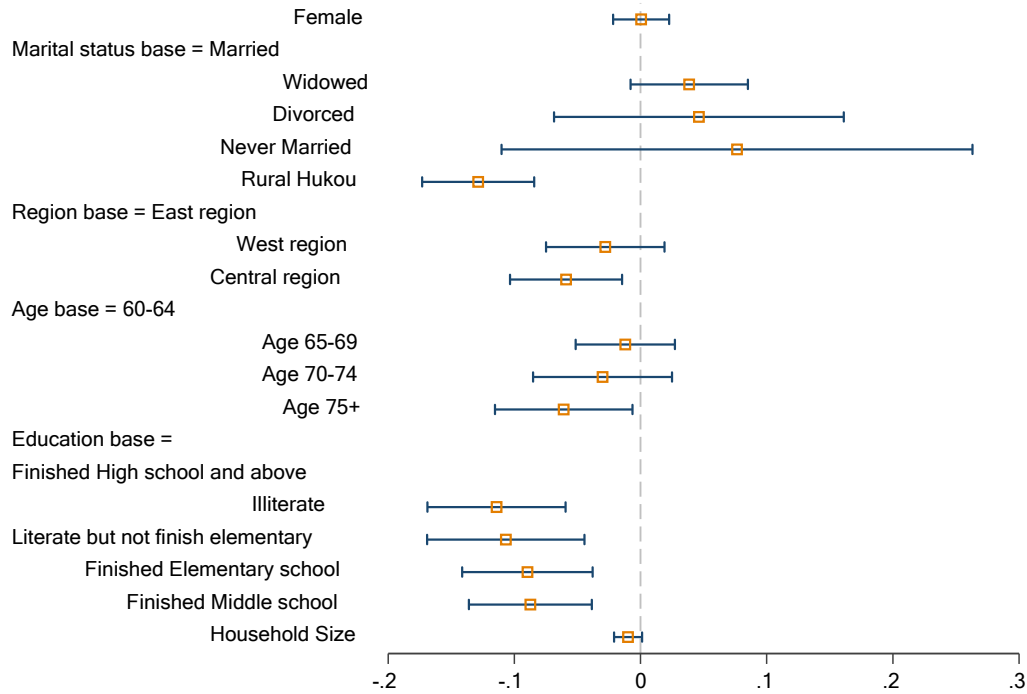
Notes: All poverty lines are in current prices. Rural poverty lines are from the National Bureau of Statistics. Urban poverty lines are rural poverty lines multiplied with a constant of 1.412.

Figure B1. Coefficient plot for demographic variables in the consumption regression, 2011 vs. 2020



Notes: These coefficients are estimated from a weighted pooled regression based on the sample of CHARLS 2011–2020, aged 60 and over. Each point estimate corresponds to the coefficient of a demographic variable interacted with the 2020 wave dummy and thereby measures the change from 2011 to 2020. The segment of each line represents a confidence interval at the level of 95 percent, while the yellow square in the middle indicates the coefficient estimate. Standard errors are clustered at the household level.

Figure B2. Coefficient plot for demographic variables in the poverty regression, 2011 vs. 2020



These coefficients are estimated from a weighted pooled regression based on the sample of CHARLS 2011–2020, aged 60 and over. Each point estimate corresponds to the coefficient of a demographic variable interacted with the 2020 wave dummy and thereby measures the change from 2011 to 2020. The segment of each line represents a confidence interval at the level of 95 percent, while the yellow square in the middle indicates the coefficient estimate. Standard errors are clustered at the household level.