Supporting Information

Xie and Zhou 10.1073/pnas.1403158111

SI TEXT

Data and the Calculation of Gini Coefficients. To measure the size of family-level income inequality in today's China, we calculated estimates of the Gini coefficient using data from seven recent nationally representative surveys. Besides CFPS 2010, which we introduced in detail in the main text, the other data sources are (*i*) the China 2005 1% Population Intercensus Survey (also called the 2005 minicensus, hereafter Mini-Census 2005), (*ii*) the 2010 and 2012 Chinese General Social Surveys (CGSS 2010 and CGSS 2012), (*iii*) the 2011 Chinese Household Finance Survey (CHFS 2011), (*iv*), the 2012 baseline wave of the China Labor Force Dynamic Survey (CLDS 2012), and (*v*) the 2012 wave of the China Family Panel Studies (CFPS 2012). Table 1 summarizes these data sources, with organization name, sample coverage, sample size, measure of family income, and estimated Gini coefficients. Note that all Gini coefficients were calculated for families with a positive annual income.

Mini-Census 2005, as part of the China census program, is an intercensus survey administered by the National Bureau of Statistics in China, which covers 973,159 families (about 2.6 million people) in all 31 provinces and equivalent administrative units of mainland China. It was conducted through a multistage stratified cluster sampling process with a household-based questionnaire collecting each household member's information pertaining to employment, occupation, and income, as well as basic demographic characteristics. For this data set, we define total family income as the sum of all family members' self-reported monthly incomes multiplied by 12. Among the 779,849 families who had a positive annual income, the Gini coefficient is 0.483. For the same sample, we also calculated the Gini coefficient for family income per capita, i.e., total family income divided by family size. In this calculation, we weighted the data by family size such that individuals, rather than families, were treated as the unit of analysis. The resulting estimate is 0.496, slightly higher than the family-level Gini coefficient. This difference may result from a negative correlation between family income per capita and family size in China. Because very poor families (in terms of family income per capita) are more likely to be larger families, relatively more weight is placed at the very low end of the income distribution in the latter calculation, thereby amplifying the measured level of inequality.

The CGSS is a nationwide, repeated, cross-sectional general survey project in China. The project was launched jointly by Renmin University of China and the Hong Kong University of Science and Technology in 2003 (1). We use two recent waves of the study, CGSS 2010 and CGSS 2012, to corroborate our estimates of the Gini coefficient measuring income inequality. Both CGSS 2010 and CGSS 2012 adopted a nationally representative sampling frame. Although the sample of CGSS 2010 covered all 31 provinces and equivalent administrative units of mainland China, the sample of CGSS 2012 did not include Tibet and Hainan. After excluding those families with missing or nonpositive income, the sample size is 10,260 for CGSS 2010 and 10,326 for CGSS 2012. We obtained Gini coefficients from these two samples at 0.545 and 0.539, respectively. Because CGSS 2012 also collected information on family size, we calculated the Gini coefficient for family income per capita. As in the case of Mini-Census 2005, the Gini coefficient at the individual level (0.563) is also higher than that at the family level (0.539).

The CFPS is a large-scale panel survey project conducted by the Institute of Social Science Survey at Peking University. The project was designed to study the long-term dynamics of social transition in China. With the household as the target of sampling, the survey collects comprehensive information on the household as a whole and all individual household members living in the sampled households. The CFPS adopts a nearly nationally representative sampling frame, which includes 25 provinces (excluding Inner Mongolia, Xinjiang, Tibet, Hainan, Ningxia, Qinghai, Hong Kong, Macau, and Taiwan) of mainland China, representing 95% of the Chinese population (2). We use data from the CFPS baseline survey, which was carried out in 2010 (CFPS 2010), and a follow-up survey in 2012 (CFPS 2012). Both CFPS 2010 and CFPS 2012 collected detailed information about households' incomes and expenditures in the previous year. We summed up the household incomes from wages and salaries, agricultural production, investment, and transfers. An earlier analysis compared income measures in CGSS 2010 and CFPS 2010 and found similar distributions of family income and expenditure from the two data sources (3). The baseline survey (CFPS 2010) interviewed 14,798 families, of which 13,837 reported a positive family income. Among these families, we calculated the Gini coefficient to be 0.530 for family income and 0.541 for family income per capita. The follow-up survey in 2012 reinterviewed about 90% of the original sample, including 13,316 families in total and 11,785 families with a positive annual income. From CFPS 2012 data, we obtained the Gini coefficient at 0.532 at the family level and 0.526 at the individual level.

The sixth data set analyzed for this study is the 2011 baseline of the China Household Finance Survey (hereafter CHFS) (website of the project is at www.chfsdata.org/). The CHFS is a survey project aimed at understanding household finance in China. The project is directed by the Survey and Research Center for the China Household Finance Survey at the Southwestern University of Finance and Economics. The baseline survey was carried out in 2011. It collected information about household assets, income, expenditures, and social and commercial insurance in 2010. The baseline survey interviewed 8,438 households, sampled in 80 districts/counties in 25 provinces. We included this data set in our study not only because of its unusual detail in collecting household financial information but also because of the amount of publicity it received. The shockingly high level of the Gini coefficient based on this survey (i.e., the 0.61 estimate) aroused both media attention and an academic debate as soon as it was released (4, 5). The survey and the Gini coefficient based on this survey were prominently featured in a *Science* article on issues concerning data accuracy in China (6). The survey also aroused suspicion of the Chinese government's official statistics, which are believed to either conceal or underestimate the worsening income inequality in China. As shown in Table 1, among the 8,092 families with a positive annual income in the CHFS, the Gini coefficient is 0.611 at the family level and 0.633 at the individual level.

Finally, we use data from the baseline wave of the China Labor Force Dynamic Survey (i.e., CLDS) to corroborate our findings based on other data sets. Launched by the Department of Sociology at Sun Yat-sen University, this survey project was designed especially to monitor labor market dynamics among Chinese adults. The baseline survey (CLDS 2012) interviewed 10,612 households in 28 provinces of mainland China (excluding Tibet, Chongqing, and Hainan). As in CGSS 2012, a single question was asked concerning the total amount of family income in 2011. Among the 9,735 families who reported a positive income, the Gini coefficient

is 0.536, which closely resembles the estimates from CGSS and CFPS.

For all of the data sets except Mini-Census 2005, we also calculated Gini coefficients with purchasing power parity (PPP) adjustment that control for differences in cost of living between rural and urban areas and among provinces. To construct regional price differences, we updated the spatial price deflators in 2000 reported by Brandt and Holz (7) to the year of 2010 using provincial rural and urban consumer price indices published by the National Bureau of Statistics of China (NBS) (8). As expected, the PPP adjustment slightly reduces the measured level of inequality for all of the data sets.*

Table S1 shows income and price differences by province and rural/urban status. All entries are ratios relative to urban families in Beijing. Raw income data are given in the first two columns. Price data are given in the next two columns. Income data adjusted for price variations are given in the last two columns. To compare variability in the three series across the province-rural/urban combinations, we also present the Gini coefficients for the series in the last row of Table S1. The results show that the variation in price by province and rural/urban status (the second series) is much less than that in income, either raw or adjusted (the first or the last series).

Regression Analyses. To compare the structure of income inequality between China and the United States, we use data for regression analyses from CFPS 2010 and the March supplement of the Current Population Survey in 2010 (March CPS 2010). The Current Population Survey (CPS), sponsored jointly by the US Census Bureau and the US Bureau of Labor Statistics, is a monthly survey that provides up-to-date labor force statistics for the US population. Although the CPS is used primarily to provide information on employment and unemployment, it also collects demographic information, including age, sex, race, marital status, educational attainment, and family structure. The March supplement of the CPS, also known as the Annual Social and Economic Supplement, includes supplemental questions on work experience, income, noncash benefits, and migration. A respondent's total income is defined as the sum of his/her money wages and salaries, net income from self-employment, and income other than earnings received in the previous calendar year. The total family income is defined as the sum of the amounts received by all income recipients in the family. The 2010 data for March CPS contain 88,957 family records. In this study, we restricted the sample to those families who had a positive total income in 2009, yielding 85,564 family records.

In CFPS 2010, the total income of a family is defined in the same way as in March CPS 2010. As introduced in the previous section, among the 14,798 families who were interviewed, 13,837 had positive income. To conduct regression analyses, we also excluded those observations with missing data on any of the covariates (i.e., region, area type, education, race/ethnicity, and family structure), a procedure that resulted in 12,523 family records. In fact, most of the sample reduction in this step is due to failure in matching family records with individual records of family heads. This failure occurred because many family heads were not present at the time of the interview. There are no missing values on the covariates in March CPS 2010.

Table S2 provides a detailed description of the covariates that are used in our regression analyses, as well as the associated numbers of degrees of freedom. Specifically, region corresponds to the 25 province-level administrative units in China covered by CFPS 2010 and 51 state-level administrative units in the United States. This variable is used to examine the role of regional variation in generating the overall inequality in China vs. that in the United States. Area type is defined as a dichotomous variable for rural and urban in China and a trichotomous variable for metropolitan, non-metropolitan, and not identified in the United States. Education denotes the educational attainment of the family head. We use the same six-category classification for both China and the United States except that the fifth category corresponds to vocational (3-y) college in China and some college in the United States. Race/ethnicity is also defined as that of the family head. Although 55 ethnic minority groups are recognized in China, only 22 of them are represented in CFPS 2010. Thus, we have 23 ethnic categories in our analytic sample. The CPS questionnaire uses a 21-category classification of race and a separate question asking if the respondent is Spanish, Hispanic, or Latino. To be conservative, we cross-tabulated these two variables and treated each nonempty cross-classification as a separate group, a procedure that yielded 38 categories of race/ethnicity for the US sample. Finally, for both countries, we adopt the same five-category classification to characterize the family structure: (*i*) primary-individual family, (*ii*) single-parent family, (*iii*) married couple with no children, (*iv*) married couple with child(ren), and (*v*) extended family.

In columns 1–3 of Tables S3 and S4, we present the sample percentages and average family income per capita by different population subgroups in China and the United States. Sample percentages are weighted by family size such that the numbers reflect the proportion of the Chinese/US population in different groups. It is notable that the population shares of different types of families differ significantly between China and the United States; whereas more than 80% of the Chinese population live in nuclear families [i.e., married couple with child (ren)] or extended families, family structure is far more diverse in the United States. In particular, nearly 18% of the US population lives in primary-individual families, whereas the corresponding figure is around 2% in China. Average family per capita is defined as the geometric mean of family income per capita weighted by family size. We chose to report the geometric means because they are less sensitive than arithmetic means to extreme values of income. It is apparent that the average income varies greatly across provinces in China but not across states in the United States. For instance, the average incomes in Beijing and Shanghai are several times higher than those in Sichuan and Guizhou. Similarly, we can see a huge urban-rural gap in China (3,674 vs. 7,805 Yuan) but not in the United States. On the other hand, different types of families vary greatly in economic status in the United States but not much in China. For example, average income in families consisting of a married couple with no children is more than four times as large as in single-parent families in the United States, whereas the corresponding ratio in China is 1.79 (6,748/3,771 = 1.79). These descriptive statistics echo our findings reported in Fig. 3.

In the last two columns of Tables S3 and S4, we present the estimated coefficients in simple and multiple regressions, which are denoted by $\beta_{\text{bivariate}}$ and β_{full}, respectively. We also report the *F*-test results corresponding to each set of predictors in both types of regressions. These *F*-test statistics measure the statistical significance of these five sets of predictors in explaining income inequality. They also bear a one-to-one correspondence with the bivariate and partial R^2 that are reported in Fig. 3. We observe that all of the *P* values are <0.001, indicating that all bivariate and partial R^2 plotted in Fig. 3 are significantly different from zero.

*In this study, the definition of family income does not include agricultural products that are self-consumed in rural areas. Considering that rural families may keep a substantial fraction of their agricultural products for self-consumption, ignoring this part of income may overestimate the rural-urban income gap and nationwide inequality. For CFPS 2010 and 2012, we also calculated Gini coefficients that take into account the imputed value of this type of income. The resulting estimates are 0.512 for CFPS 2010 and 0.498 for CFPS 2012, lower than those reported in Table 1.

1. Bian Y, Li L (2012) The Chinese General Social Survey (2003-8): Sample designs and data evaluation. *Chin Sociol Rev* 45(1):70–97.
2. Xie Y (2012) *User's Manual of China Family Panel Studies Data 2010.* Available at www.isss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/a16476476881bc46c90086fa1508880b.pdf. Accessed September 1, 2013.
3. Xu Q, Zhang C, Zhou X, Xie Y (2012) *Comparison of Income Distribution: CFPS, CGSS, CHIP, and CHFS*. Available at www.isss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/a29e71f54de2f5dcd71c6284e4eed017.pdf. Accessed September 1, 2013.
4. Anonymous (2012) To each, not according to his needs: A new survey illuminates the extent of Chinese income inequality. Available at www.economist.com/news/finance-and-economics/21568423-new-survey-illuminates-extent-chinese-income-inequality-each-not. Accessed September 1, 2013.
5. Kao E (2012) China wealth gap continues to widen, survey finds. *South China Morning Post.* Available at www.scmp.com/news/china/article/1102047/china-wealth-gap-continues-widen-survey-finds. Accessed September 1, 2013.
6. Hvistendahl M (2013) The numbers game. *Science* 340(80):1037–1039.
7. Brandt L, Holz CA (2006) Spatial price differences in China: Estimates and implications. *Econ Dev Cult Change* 55(1):43–86.
8. National Bureau of Statistics of China (2011) *China Statistical Yearbook 2011* (China Statistics Press, Beijing).

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)
Table S4 (DOCX)