Data and Code for replication of "Early-life exposure to hardship increased risk tolerance and entrepreneurship in adulthood with gender differences"

Junjian Yi, Junhong Chu, and I.P.L. Png

February 2022

R1. Description of the replication package

Here, we describe the replication package, as deposited in the Harvard Dataverse (https://doi.org/10.7910/DVN/MC7B1H). Our analysis used county-level data from national censuses and representative surveys. As these sources restrict access to county-level data, the data that we can publish in the replication package are limited to that which would not reveal the county-level data.

The replication package includes the Stata code used to process the raw data for analysis (prep.do) and carry out the analysis as reported in the tables and figures (figtab.do) and measures of famine intensity by county (details in Section R2 below). In addition, we uploaded the daily instrumental weather records at 854 stations, 1950-2012, to a permanent repository: https://zenodo.org/record/6084180#.YgwYkt9BxEY.

To replicate our analyses, researchers should first apply to the respective sources (Section R3 describes the procedures) for access to the county-level data, and then run our code. Alternatively, to replicate the estimates using the CHFS, researchers may apply to the Southwestern University of Finance and Economics, Chengdu (chfs.swufe.edu.cn/), Ms Changyan Peng (pengchangyan@chfs.cn or contactus@chfs.cn), to access the replication file, Famine-code-data-log.zip.

R2. Replication files

Code:

1) Famine-intensity.do: Constructs the measure of famine intensity by county, using the rate of cohort loss and thermal agricultural productivity by county-year.
2) Census-sample.do: Processes the raw census data (1990 and 2000 Population Censuses, 2004 Economic Census, and 2005 mini-Census) to generate the cleaned data for empirical analysis.
3) Survey-sample.do: Processes the raw survey data (CFPS, CGSS, CHFS) to generate the cleaned data for empirical analysis.
4) Figtab.do: Carries out the analysis as reported in the tables and figures in the main text and Supplementary Information. The code follows the order of tables and figures in the paper with a few exceptions (Figure 2 is with Table S6; Figures 1, S1-S2 and Table S3 are in Famine-intensity.do).

Table R1 describes the preparatory files in more detail.

Data:

1. Station_GeoInfo.dta: Coordinates of 793 weather stations;
2. County_GeoInfo.dta: Coordinates of 2,870 counties;

3. China_coord.dta: Coordinates of 2,876 county boundaries;
4. match_00_ID.dta: Manually compiled match of the county IDs in 2000 and the IDs of the counties in China_coord.dta;
5. match_90_00.dta: Manually compiled match of the county IDs in 1990 and 2000;
6. match_00_04.dta: Manually compiled match of the county IDs in 2000 and 2004;
7. wthr-19502012.dta: Daily instrumental temperature records of 854 weather stations covering years 1950-2012, China Meteorological Administration;
8. county_controls.dta: County-level characteristics of 3,219 counties, including county government expenditure, GDP and population in 2000, and county agricultural productivity in 1993.
9. famine-intensity-estcy.dta: Predicted cohort loss rate based on thermal agricultural productivity by county-year for 2,727 counties covering years 1951-1970 (produced by Famine-intensity.do, lines 644-790).
10. famine-intensity-estc.dta: Famine intensity for 2,727 counties (produced by Famine-intensity.do, lines 792-834).

R3. Access to the raw county-level data: Contacts

1. China Population Census 1990: NBS-PKU Research Data Center, Peking University. To access the data, apply through the website (https://microdata.stats.gov.cn/), and click on the third link (国家统计局北京大学数据开发中心).

2. China Population Census 2000: Refer to Point 1.

3. China Population mini-Census 2005: Refer to Point 1.

4. China Economic Census 2004: Refer to Point 1.

5. China Household Finance Survey 2011 and 2013: Survey and Research Center for China Household Finance, Southwestern University of Finance and Economics, Chengdu (https://chfs.swufe.edu.cn/). To access the data, apply through the official platform (https://chfs.swufe.edu.cn/datacenter/apply.html).

6. China General Social Survey 2008, 2010, 2011, 2012, and 2013: National Survey Research Center, Renmin University of China, Beijing (http://cgss.ruc.edu.cn/English/Contact_Us.htm). To access the data, contact cgss@ruc.edu.cn.

7. China Family Panel Studies 2010: Institute of Social Science Survey, Peking University (http://www.isss.pku.edu.cn/cfps/). To access the restricted county-level data, apply through http://www.isss.pku.edu.cn/cfps/sjzx/xzsj/index.htm.

Table R1. Details of preparatory files

Code	Source data	Final replication sample
Famine-intensity.do	Station_GeoInfo.dta: Coordinates of 793 weather stations; County_GeoInfo.dta: Coordinates of 2,870 counties; China_coord.dta: Coordinates of 2,876 county boundaries, to plot Figure 1; match_00_ID.dta: Manually compiled match of the county IDs in 2000 and the county IDs in China_coord.dta; match_90_00.dta: Manually compiled match of the county IDs in 1990 and 2000; wthr-19502012.dta: daily instrumental temperature records of 854 weather stations covering years 1950-2012, China Meteorological Administration; Census00.dta: 2000 Population Census; Census90.dta: 1990 Population Census;	famine-intensity-estcy.dta; famine-intensity-estc.dta; Weather_sample_1.dta, Weather _sample_2.dta: To replicate Table S2.
Census-sample.do	Census90.dta: 1990 Population Census; Census00.dta: 2000 Population Census; Census04.dta: 2004 Economic Census; Census05.dta: 2005 Population mini-Census; match_90_00.dta: Manually compiled match of the county IDs in 1990 and 2000; match_00_04.dta: Manually compiled match of the county IDs in 2000 and 2004; county_controls.dta: County-level characteristics of 3,219 counties, including county government expenditure, GDP and population in 2000, and county agricultural productivity in 1993; famine-intensity-estc.dta.	Census05_sample.dta: To replicate Tables 1, S4-S5; Census00_sample.dta: To replicate Table S12 column (c); Op-cond-cor.dta: To replicate Table S6 and Figure 2.
Survey-sample.do	cfps2010family.dta: CFPS 2010 wave, family module; cfps2010adult.dta: CFPS 2010 wave, adult module. cgss2008.dta: CGSS 2008 wave; cgss2010.dta: CGSS 2010 wave; cgss2011.dta: CGSS 2011 wave; cgss2012.dta: CGSS 2012 wave; cgss2013.dta: CGSS 2013 wave. chfs2011_ind.dta: CHFS 2011 wave, individual module; chfs2011_hh.dta: CHFS 2011 wave, household module; chfs2011_city.dta: CHFS 2011 wave, city module; chfs2013_ind.dta: CHFS 2013 wave, individual module; chfs2013_hh.dta: CHFS 2013 wave, household module; chfs2013_city.dta: CHFS 2013 wave, city module. famine-intensity-estc.dta	cfps10_sample.dta: To replicate Tables S7, S12; cgss101213_sample.dta: To replicate Table S7; cgss08_sample.dta: To replicate Tables S11, S12; cgss11_sample.dta: To replicate Tables S10, S11; chfs11_sample.dta: To replicate Tables S7, S8; chfs13_sample.dta: To replicate Tables 2, S9.