Appendix for

Great Famine, Differential Fertility, and Income Inequality: Evidence from China

Chenxiao Shou,[1] Xuebo Wang,[2] Junsen Zhang[3]

[1] School of Business, Hunan Normal University. E-mail: scx@hunnu.edu.cn.
[2] School of Economics, Shanghai University of Finance and Economics. E-mail: wang.xuebo@mail.shufe.edu.cn.
[3] School of Economics, Zhejiang University. E-mail: jszhang@cuhk.edu.hk.

A1. Summary Statistics for Control Variables

Table A1 demonstrates the summary statistics for the control variables in the regressions (provincial and prefectural levels) of the main text.

We include 28 provinces and 291 prefectures in our analysis. We have no information on the rural mortality rate during the Famine for three provinces (Tibet, Chongqing, and Hainan). Thus, we exclude them from our analysis. We also exclude all the prefectures in Tibet from our prefectural-level analysis due to the following considerations: Tibet is different from other regions in China in many ways, the population size of most prefectures in Tibet is also very small, and the sample size of those prefectures in the 2005 mini census data is too small to generate a Gini coefficient. We do not exclude prefectures in Chongqing and Hainan Province from the prefectural-level analysis because we intend to include as many prefectures as possible in the regressions to make our analysis more representative.

In the regressions at the provincial level, we control the following variables: rural population share and agricultural productivity in 1958, post-Famine (1962–1985) rural share of women of childbearing age, and provincial characteristics in 2005, including income per capita, GDP, agricultural and industrial output shares in GDP, migrant population share in total population, population density, and unemployment insurance participation rate. Given that the data on rural population share and agricultural productivity in 1958 at the prefectural level are unavailable, we cannot control the corresponding pre-Famine characteristics in the prefectural-level regressions. However, our estimation results at the provincial level are not sensitive to the inclusion or exclusion of these two control variables. Therefore, not controlling them in the prefectural-level regressions may not be a big problem. Furthermore, we lack the information of the unemployment insurance participation rate in 2005 for each prefecture. Instead, we control the public expenditure per capita in 2005 in the prefectural-level regressions. Again, this practice may not be a problem because our estimation results are not sensitive to the inclusion or exclusion of a specific control variable.

The data on rural population share and agricultural productivity in 1958 (at the provincial level) are from Meng et al. (2015), and the data on provincial and prefectural characteristics in 2005 are from the National Bureau of Statistics. We obtain the information on post-Famine rural share of women of childbearing age (15–49 years old) for each year of 1962–1985 from the 1982 census data. We do not use the 1990 census data to obtain this variable because many childbearing-age women in 1962–1985 may had died in 1990. For example, childbearing-age women in 1962 are 15–49 years old and were already 43–77 years old in 1990; some may have died before. In the 1982 census data, these women were younger (35–69 years old) and more likely to survive and thus deemed a better sample. Be that

as it may, it is still possible that childbearing-age women in 1962–1985 have died in 1982. We will attempt to resolve this problem later.

Accurate information on respondents' hukou types (rural or urban) in the 1982 census data is unavailable. We identify respondents' hukou statuses on the basis of the industry which employed the head of the respondents' household. Specifically, if the industry that employed a household head is related to agriculture, that is, the variable "Industry" takes the values between 11 and 42, which refers to the five agricultural industries: farming, forestry, animal husbandry, side-line production, and fishery, then, we identify all the household members as rural residents. We identify other respondents as urban residents. In this way, we identify 78.59% of the total population as rural residents, which is very close to the officially announced rural share of total population in 1982 (i.e., 78.87%), which confirms the accuracy of our identification of respondents' hukou types.

As mentioned above, when we identify rural and urban childbearing-age women in each year of 1962–1985 in each province (or prefecture) from the 1982 census data, we should consider the possibility that some of them may have died before 1982 and were thus excluded in the sample. To resolve this problem, we combine the 1982 and 1990 census data to estimate the mortality rate of each birth cohort to adjust the rural share of childbearing-age women of each year. Specifically, if we obtain the number of rural and urban childbearing-age women in 1962 in a province from the 1982 census, who were 15–49 years old in 1962 and 35–69 years old in 1982 (who belong to the 1913–1947 birth cohorts), we can compare the sample size of these birth cohorts in the 1982 and 1990 census to estimate the mortality rate during the eight years. For example, if the sample size of the 1921 birth cohort (i.e., 61 years old in 1982) is 10,000 in a province in the 1982 census and decreases to 8000 in the 1990 census (i.e., 69 years old in 1990), then we conclude that the mortality rate from 61-year-old to 69-year-old women in this province is 20%. We thus divide the number of 69-year-old women in this province in the 1982 census by 0.8 to obtain the number of women of this cohort when they were 61 years old. Similarly, we can further obtain the number of women of this cohort when they were 53, 45,… years old, until an age at which the mortality rate is negligible. With this strategy, we can obtain the adjusted number of rural and urban women of each birth cohort in each province (or prefecture). Actually, the unadjusted and adjusted rural shares of childbearing-age women are highly correlated, and we obtain similar estimate results by controlling either of the two measures.

We obtain the rural share of childbearing-age women for 1962–1985 by the weighted average method, that is, we first obtain the rural share of childbearing-age women for each year of 1962–1985 and then calculate a weighted average value of the variable with the sample size in each year as the weight factor.

During 1982 and 1990, China experienced significant adjustment of the administrative division at the prefectural level and many counties were separated from their original prefectures and divided into other prefectures. No county-level information exists in the 1982 census data, thus preventing us from matching all the prefectures in 1982 to those in 1990 when combing the two census data sets. In fact, we can match a total of 230 in 291 prefectures in the two data sets. Therefore, if we control the variable of the rural share of childbearing-age women in the prefectural level regressions, we have to drop the 61 prefectures which cannot be matched in the two census data sets. We intend to include as many prefectures as possible to make our analysis representative. Thus, we do not control the rural share of childbearing-age women in the prefectural level regressions. However, it is reassuring that even when we include this control and run the regressions for the 230 prefectures, we still obtain similar results. In other words, our results are insensitive to the inclusion or exclusion of this control variable. Table A2 reports the DID estimates of the effect of the rural population share on the Gini coefficient for the 1962–1980 birth cohorts, which is similar to Table 9 in the main text. The only difference is that we control the rural share of childbearing-age women in all the regressions. Unsurprisingly, the results in Tables A2 and 9 are very similar, which confirm the robustness of our estimates.

A2. More Discussions about Migration

We use the 1990 census data to identify respondents' (the 1962–1985 birth cohorts) birth places and take their hukou registration location at the time of census as their birth places. We assume that respondents' hukou statuses did not change after they were born. In other words, we assume that there was no migration from rural to urban areas (within a province or prefecture) and across regions. As such migrations will induce that respondent's residences or hukou registration locations differ from their birth places and render the identification of their birth places impossible. Fortunately, as discussed in the main text, such migration is not common before 1990.

By contrast, when we study the Gini coefficient across province or prefecture in 2005, migration from rural to urban areas within province or prefecture is allowed and migration across regions will induce problems. For example, if a respondent was born in rural areas in a prefecture in 1962 and obtained an urban hukou in the same prefecture before 2005, then it would not be a problem when we compare the Gini coefficient across prefecture in 2005. Actually, intergenerational mobility is allowed in the aforementioned theoretical studies and our identification. Specifically, children of poor families can become rich after growing up and rural born children are allowed to obtain urban hukou later. We intend to examine the effect of a higher rural fertility on the income inequality of the next generation

at the provincial (or prefectural) level, and rural children can migrate to the urban areas within province (or prefecture) after growing up. By contrast, if many respondents were born in one prefecture in 1962 and obtained hukou in another prefecture before 2005, then the 1962–1985 birth cohorts in the prefecture does not correspond to the same group of people in this prefecture in 2005 of which we examined the Gini coefficient. Therefore, it would contaminate our prefectural-level estimates of the effect of the rural population share of the 1962–1985 birth cohorts on the income inequality of these cohorts in 2005.

Thus, we need to identify those migrants across regions (province or prefecture) and put them back to the sample of their hometown to resolve the problems induced by such migrations. Fortunately, the most common migration across regions is that migrant workers leave their hometown and work in another city, and we can precisely identify them in the 2005 census data.

A3. Famine Severity and the Implementation Intensity of the One-Child Policy (OCP)

In the main text, we argue that the Famine severity and the implementation intensity of the population control policy are not correlated. We now provide some direct evidence on this statement. Figure A1 plots the implementation intensity of the OCP as measured by fines for excess fertility in 1979 and the Famine severity as measured by the average rural excess mortality rate in 1959–1961 for all provinces. Evidently, the fines for excess fertility in most provinces are the same in 1979 when the OCP was initially implemented and was thus uncorrelated with the Famine severity. Figure A2 plots the fines for excess fertility in 1985 and the Famine severity for all provinces. In 1985 when the OCP had been implemented for six years, the fines for excess fertility across province show considerable variations. However, as shown in Figure A2, the points are extremely scattered and the two variables still seem uncorrelated at all.

In sum, we have no reason to expect that the Famine severity and the implementation intensity of the population control policy are correlated in any way, and the above figures confirm that the two variables are not systematically correlated.

Table A1: Summary Statistics for Control Variables

Variables	Definition	Obs.	Mean	S.D.
	A. Provincial Level			
Provincial Characteristics in 1958				
Rshare58	Rural population share	28	0.773	0.153
Grainpc58	Logarithm of grain output per capita (unit: kg.)	28	5.904	0.290
Provincial Characteristics in 2005				
GDP	Logarithm of Gross Domestic Product (unit: one billion yuan)	28	6.232	0.878
Income per capita	Logarithm of monthly income per capita (unit: yuan)	28	6.283	0.375
Primary	Agricultural output share in GDP	28	0.130	0.059
Secondary	Industrial output share in GDP	28	0.470	0.066
Unemp insurance rate	Unemployment insurance participation rate for urban population	28	0.192	0.048
Population density	Logarithm of population density (unit: people per square kilometer)	28	5.367	1.312
Migrant population share	Proportion of net inflow of population in total population	28	0.026	0.082
Rural share of childbearing-age women	Post-Famine (1962–1985) rural share of women of childbearing age	28	0.701	0.174
	B. Prefectural Level			
Prefectural Characteristics in 2005				
GDP	Logarithm of Gross Domestic Product (unit: one billion yuan)	291	3.538	1.105
Income per capita	Logarithm of monthly income per capita (unit: yuan)	291	6.165	0.355
Primary	Agricultural output share in GDP	291	0.186	0.117
Secondary	Industrial output share in GDP	291	0.444	0.126
Fiscal Expenditure per capita	Logarithm of fiscal expenditure per capita (unit: yuan)	291	7.225	0.589
Population density	Logarithm of population density (unit: people per square kilometer)	291	5.320	1.321
Migrant population share	Proportion of net inflow of population in total population	291	-0.007	0.108

Table A2: DID Estimates of the Effect of the Rural Population Share on the Gini Coefficient for the 1962–1980 Birth Cohorts (Prefectural Level)

	Dependent variable: Gini coefficient					
	A. Benchmark regression		B. Migration adjustment		C. M&P adjustment	
	RF	IV	RF	IV	RF	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Rshare		**0.913*** **		**1.046*** **		**1.060*** **
		(0.193)		(0.228)		(0.223)
Excess Mortality Rates× T_t	0.053***		0.057***		0.057***	
	(0.011)		(0.012)		(0.012)	
		1st Stage		1st Stage		1st Stage
Excess Mortality Rates× T_t		0.058***		0.055***		0.054***
		(0.011)		(0.011)		(0.011)
Prefecture FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
K-P F statistic		63.37		63.24		63.37
Observations	5,256	5,256	5,304	5,304	5,304	5,304
Adj R-squared	0.648	0. 648	0.535	0.535	0.580	0.580

Notes: Control variables include rural share of women of childbearing age in each year of 1962–1985, and a set of prefectural level characteristics in 2005, including income per capita, GDP, agricultural and industrial output share in GDP, migrant population share in total population, population density, and fiscal expenditure per capita. Panel A reports the benchmark estimation results, and Panels B and C list the estimation results with migration and price adjustment, respectively.
Robust standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1.

Figure A1: Fines for Excess Fertility in 1979 and the Average Rural Excess Mortality Rate in 1959–1961 (Provincial Level)

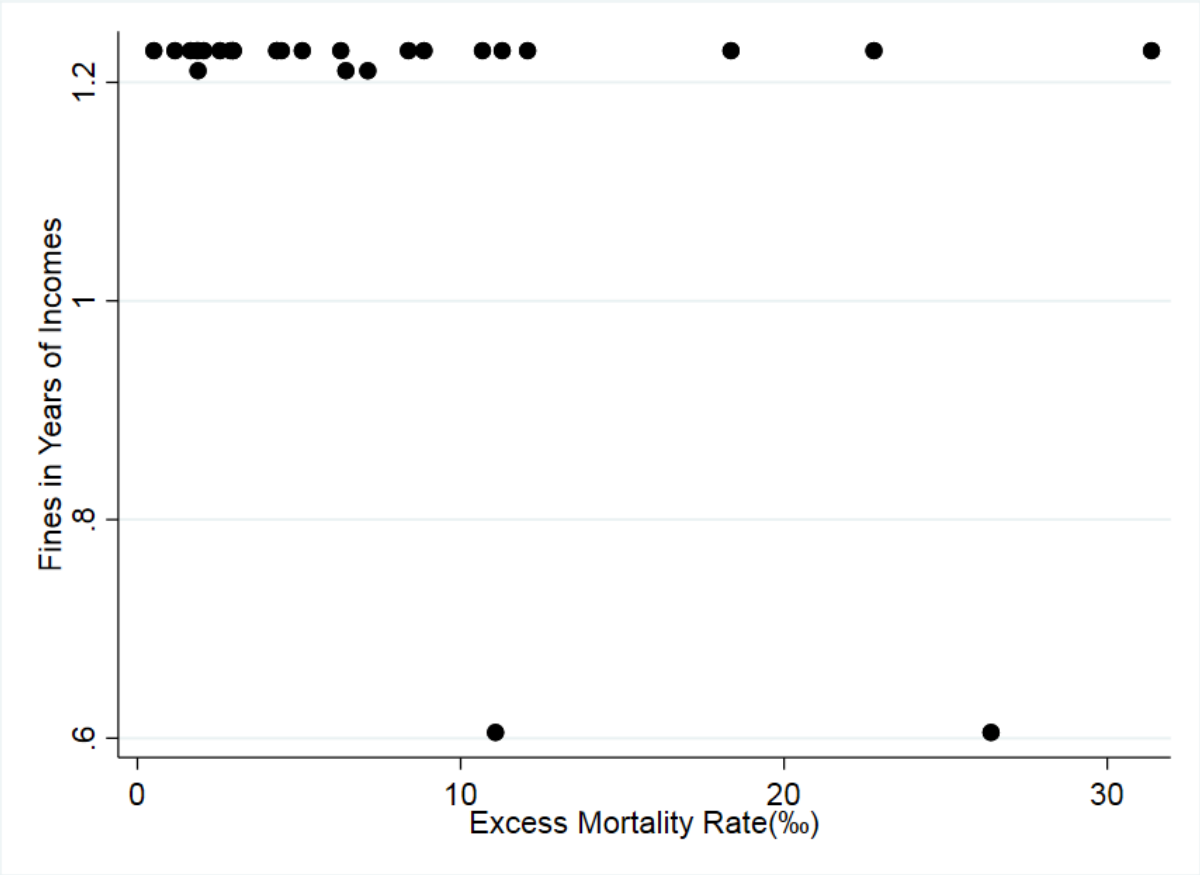

Figure A2: Fines for Excess Fertility in 1985 and Average Rural Excess Mortality Rate in 1959–1961 (Provincial Level)