Online 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. Theoretical Discussions

Conceptually, whether and to what extent differential fertility across income classes affects income inequality of society largely depends on intergenerational income mobility. Specifically, if the intergenerational income correlation is one, that is, children have the same income levels with their parents, then differential fertility would have a direct and strong effect on the income distribution of the next generation, which is a pure population compositional effect. At the other extreme, if the intergenerational income correlation is zero, which means that children's income is uncorrelated with their parents' income, then differential fertility does not matter at all because children from different family backgrounds exactly have the same probability of falling into any income class. In reality, the intergenerational income correlation is generally between zero and one, and children of rich and poor parents have different probabilities of falling into any income class, which makes the issue very complicated.

Lam (1986) analyzes the dynamics of differential fertility across income classes and income inequality of society. In a model of differential fertility and intergenerational mobility based on a Markov process governing transitions across income classes, there are *1, 2, ……, n* income classes (the income level increases with the number) with different fertility rates, and intergenerational mobility is described by a matrix M, where element M_{ij} is the probability that a child of class j becomes a member of class i. The first step to examine the effect of a higher fertility of the poor on the income distribution of the society is to investigate how a higher fertility of the poor affects the proportion of the poor. Lam first analyzes the effect of a change in F_1, the fertility of the poorest class, on the proportion in that class in the next period. He concludes that if $M_{1i} \leq M_{11} \forall i,$ implying that parents of other classes are less likely to produce poor offspring than the poor themselves, then an increase in the fertility of the poor will always lead to an increase in the proportion of the poor in subsequent period. However, the effect of a change in F_1 on a potential income inequality index is much more complicated and cannot be predicted without knowing the actual magnitude of the change. For example, if we consider the effect of an increase in the proportion of the poor on the Gini coefficient, the Gini coefficient is likely to increase initially but must eventually decrease as the new entrants of the poor finally dominate the distribution and most people become poor. Lam further shows that two inequality measures, namely, variance of log income and coefficient of variation, move in opposite directions in both the steady sate and transition in response to the elimination of fertility differentials.

Such results raise serious concerns about the validity of income inequality measures and inequality comparisons when fertility rates differ across income classes.

The subsequent studies (Chu, 1987; Dietzenbacher, 1989) show that the issue is more complicated than expected and even Lam's simplest statement "if $M_{1i} \leq M_{11} \forall i,$ an increase of the fertility of the poor will lead to an increase in the proportion of the poor" does not necessarily hold and needs stronger conditions. For instance, if the children of the poorest class have a very high probability (say, 70%) of falling into the richest class, although they also have a slightly higher probability (say, 30%) of falling into the poorest class than the children of the richest class (say, 29%), then a higher fertility of the poor could lead to an increase in the proportion of the rich. Chu and Koo (1990) further systematically study the issue and conclude that under several more rigorous assumptions a higher fertility of the poorest class will lead to a higher proportion of the poor; moreover, the resulting income distribution will be conditionally first-degree stochastic dominated by the initial one (in a Markov branching process of income distribution dynamics). If an income distribution conditionally first-degree stochastic dominates (CFSD) the other one, it means it is better and more favorable than the other one in that it has a higher social welfare level. For example, if there are two initially identical income distributions and if we increase the income of the poorest class of the first one, then, it would dominate the other income distribution because its poorest class has a higher income level with other things being equal. Theoretically, the CFSD relation is an important concept and has stronger welfare implications than the income inequality measures. Specifically, a society with a lower income inequality and all people are poor does not necessarily dominate the other society with a higher income inequality. Although CFSD relation is a theoretically important concept, empirically testing such a relation is extremely difficult. Income equality is a practically more important concept and policy makers and the public care more about the income inequality measures such as the Gini coefficient. Although a lower income inequality level is not always the most desirable outcome, we still prefer a lower rather than a higher income inequality under most situations (except the situation under which income inequality is low and all people are poor). Thus, studying income inequality is practically important and also has strong policy implications.

Although we cannot reach a final conclusion about the effect of an increase in the fertility of the poor on different income inequality measures based on pure theoretical analyses, we can still make some reasonable inferences from these studies. According to Lam's analysis, an increase in the proportion of the poor is likely to increase the income inequality measures such as the Gini coefficient initially, but the Gini coefficient will eventually decrease as the proportion of the poor becomes very high and most people become poor. Therefore, before the income distribution reaches the extreme point at which most people are poor, an initial increase in the proportion of the poor will probably increase rather than decrease the income inequality of the society.

Our empirical investigations are based on the aforementioned theoretical literature. We intend to examine the effect of a plausible exogeneous increase of rural fertility in China on the income distribution of the region several decades later. According to Chu and Koo (1990), under three assumptions, an increase in rural fertility will induce a higher proportion of the poor and lead to a less favorable income distribution with a lower social welfare level. According to Lam (1986), an increase in the proportion of the poor is likely to increase the Gini coefficient (except the extreme situation under which the proportion of the poor becomes very high and most people are poor). Therefore, we discuss and empirically test the three assumptions in Chu and Koo (1990) and empirically identify the effect of an increase in rural fertility on the Gini coefficient of the country several decades later.

Some macroeconomics literature points out another possibility that income inequality may directly affect differential fertility. Kremer and Chen (2002) find that higher inequality levels tend to be associated with larger fertility differentials within a country. Croix and Doepke (2003) demonstrate that an increase in income inequality could increase the fertility differential between the rich and the poor, implying that additional weight is placed on families who provide little education; thus, an increase in inequality lowers average education and economic growth. Our empirical findings do not contradict such possibilities that income inequality could affect the current differential fertility across income classes. Furthermore, as we are studying the effect of the post-famine differential fertility on the income inequality in 2005, the two-way causality may not be a problem because the income inequality in 2005 is unlikely to affect the differential fertility in the 1960s. Nevertheless, we cannot exclude the possibility that the current income inequality could be correlated with previous differential fertility in other ways. Therefore, we need to find plausible exogeneous shocks on differential fertility to resolve the potential endogenous problems in our identification

A2. Summary Statistics for Control Variables

Table A1 demonstrates the summary statistics for the control variables in the regressions of the main text. We include 291 prefectures in our prefectural level analysis. We 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 prefectural level, we control for a set of prefectural characteristics in 2005, including income per capita, GDP, agricultural and industrial output shares in GDP, immigrant population share in total population, population density, and the fiscal expenditure per capita. The data on prefectural characteristics in 2005 are from the National Bureau of Statistics.

A3. Differential Fertility and Population Composition: Theoretical Clarification

We intend to investigate the effect of differential fertility of rural and urban households after the famine on the income inequality of the corresponding post-famine birth cohorts. However, in practice, we use the rural population share of the post-famine birth cohorts to substitute for the fertility differential between rural and urban households. Although differential fertility and population composition of the next generation are closely related, there is still subtle difference between them. We now prove that under certain scenarios, these two concepts could be equivalent.

Let α be the rural population share of a certain birth cohort; let N_1, N_2 be the number of women of childbearing age (15–49 years old) of the same year in rural and urban areas, respectively; let n_1, n_2 be rural and urban fertility in the same year, respectively. We can obtain the following:

$$\alpha = \frac{N_1 n_1}{N_1 n_1 + N_2 n_2} = \frac{1}{1 + \frac{N_2 n_2}{N_1 n_1}} \quad . \tag{1'}$$

Therefore, rural population share α is actually a function of the rural-urban fertility ratio $\frac{n_1}{n_2}$. Furthermore, if the ratio of rural and urban numbers of childbearing-age women ($\frac{N_1}{N_2}$) is given, there is a one-to-one correspondence between α and $\frac{n_1}{n_2}$. Thus, estimating the effects of the rural population share and the rural–urban fertility ratio on income inequality becomes equivalent when conditioned on the rural share of childbearing-age women.

Intuitively, differential fertility across income classes affects income inequality through changing the proportion of the offspring of the poor and rich in the total population and further affecting the income distribution of the next generation. For example, if rural fertility is much higher than the urban ones, then the population of the next generation would contain a larger share of rural children, and thus income inequality of this generation may increase accordingly. Considering that population composition is easier to measure and interpret and it is also the critical determinant of income inequality, we mainly investigate the effect of rural population share on the income inequality of the next generation in the empirical practice.

A4. Measuring Famine Severity at the Prefectural Level

We use famine severity to instrument for the rural population share of the post-famine birth cohorts, however, the data on the mortality rate during the famine at the prefectural level are unavailable. Meng et al. (2015) use the 1990 census data to calculate the birth cohort size of survivors during the famine among the agricultural population to proxy for famine severity at the county level. They show that birth cohort size during 1959–1961 is negatively correlated with famine severity because it captures the reduced fertility and increased mortality caused by the famine. In addition, they further argue that this famine severity index has several advantages over the mortality rate data.

Similar to the strategy of Meng et al. (2015), we first obtain the rural birth cohort size of each year over the period 1950–1970 for each prefecture from the 1990 census data and then fit a trend line of the prefecture rural birth cohort size during this period. We further calculate the gap between the actual and trend values of the rural birth cohort size during the famine (1959–1961) and finally obtain the ratio of this gap and the corresponding trend value to proxy for the famine severity at the prefectural level. Figure A1 plots the prefectural average rural birth cohort size of each year over the period 1950–1970 (the solid line) and the trend line of the birth cohort size during this period (the dotted line). Evidently, the birth cohort size decreased dramatically during the famine, resulting in a considerable gap between its actual and trend values. Thus, this variable, the average rural birth cohort size gap during the three years of the famine (1959–1961), measures to what extent rural birth cohort size shrinks during the famine and thus largely reflects famine severity. We also calculate the corresponding rural birth cohort size gap for each province and compare this indicator with the rural excess mortality rate during the famine, the alternative measure of famine severity. We determine that the correlation of these two variables is as high as 0.85 at the provincial level, which confirms the validity of the rural birth cohort size gap as an effective indicator of famine severity. Thereafter, we use this rural birth cohort size gap to instrument the post-famine rural population share for the 1962–1985 birth cohorts and obtain the corresponding estimates of the rural population share on the Gini coefficient of these birth cohorts in 2005 at the prefectural level.

A5. Discussions on Migrations across Regions

As discussed in the main text, we use the 1990 census data to calculate the rural population share of each prefecture, based on respondents' hukou status at the time of the census. To calculate the Gini coefficient, we use the 2005 census data, relying on respondents' place of residence and workplace.

All current labor force members working in each prefecture are included to derive the 2005 Gini coefficient for that prefecture. However, since individuals born in one prefecture may migrate to others over time, this approach might have limitations. Although migration was uncommon before 1990, there was a steady inflow of migrants from the western and central regions to the more developed eastern regions by 2005. Given that some migrant workers in each prefecture in 2005 may have originated from other areas, this could pose challenges to our strategy. To address this issue, we use information on respondents' migration status from the 2005 census data.

In the survey of the 2005 census, question "R6" is related to the respondents' hukou registration location. If the respondents' hukous were registered in where they were currently residing, then, we identify them as local residents. By contrast, if the respondents' hukous were registered in other places rather than where they were currently living, we identify them as migrant population and further take the registration location of their hukou as their hometown.

Such an identification of the respondents' migrant statuses may not be completely accurate. For example, if a respondent was born in one prefecture, later successfully admitted to college in another prefecture, obtained a hukou after graduation, and lived there permanently, then, we include her in the sample of the prefecture of her residence rather than her home prefecture in 2005. Theoretically, we should include all the respondents who were born during 1962–1985 in the prefecture in the sample of this prefecture in 2005, and such an inaccurate choice of the sample may induce problems. However, given Chinese government's extremely tight control on the hukou system over a long period, it is very difficult for most people to change their hukou statuses, particularly switch hukou from a region to another. Given that the most typical migration in China was that people left their hometown and worked in other cities without obtaining the hukou there, we can precisely identify the migration statuses of these people with the above method in the census data.

If we include those migrant workers in their home prefectures, then we should also consider the difference in living cost between their residences and their hometowns. For example, a migrant worker from Shaanxi worked in Shanghai and earned a 3000 Yuan monthly salary, while similar workers in his hometown can only earn 1500 Yuan per month. In this case, we cannot say that this worker's salary is twice as high as that of workers in his hometown because the cost of living in Shanghai is also substantially higher than that back home. Therefore, given the difference in the cost of living across prefectures, those migrant workers' income may be incomparable to that in their home prefectures.

We thus further adjust the price differences across prefectures to make the income earned in different regions comparable. Using the spatial price index for all prefectures in 2004, calculated by Brandt and Holz (2006), we adjust the nominal income in all prefectures accordingly. We then reestimate Equations (5) and (6) in the main text, incorporating these migration and price adjustments, and present the results in Table A2. The results remain consistent with the benchmark findings in Table 5 of the main text, confirming the robustness of our analysis.

A6. Extreme Weather Shocks during the Famine as an IV for Famine Severity

As discussed in the main text, we use extreme weather shocks as an instrument variable for famine severity. Specifically, we employ a dummy variable to indicate whether a rainfall shock occurred during the famine to measure extreme weather shocks. Following the methodology used in extensive literature (e.g., Shah and Steinberg, 2017; Addoum et al., 2020; Corno et al., 2020), we define a rainfall shock in a given prefecture is defined as annual rainfall below the 5th percentile or above the 95th percentile of the prefecture's long-term rainfall distribution observed during the period 1950–1966. The weather data is sourced from Meng et al. (2015). The largely random occurrence of a rainfall shock during the famine makes it a valid IV for famine severity. Specifically, a weather shock is unlikely to affect regional long-term income inequality through channels other than its impact on famine severity.

As established in the literature, the sharp decline in food production and high government procurement from rural areas were the primary causes of the famine (Thaxton, 2008; Meng et al., 2015). To explore how extreme weather shocks influenced famine severity, we examine their impacts on grain output and government grain procurement during the famine.

Given that data on grain output and grain procurement is only available at the provincial level, we employ a Difference-in-Differences (DID) framework with a panel spanning 1953–1961 at the provincial level for our analysis. Specifically, we compare the per capita grain output and grain procurement between the treatment group (provinces that experienced extreme weather shocks during the famine) and the control group (provinces that did not experience such shocks) before and during the famine. This allows us to identify the effects of extreme weather shocks on grain production and procurement during the famine period. In practice, we estimate the following equation:

$$lny_{pt} = a + \alpha_p + \gamma_t + b(GF_t \times RainShock_p) + \varepsilon_{pt} \qquad (2')$$

where lny_{pt} represents the logarithm of outcome variable (either per capita grain output or grain procurement) for province *p* in year *t*; GF_t is a dummy variable that equals 1 if year *t* falls within the famine period (1959–1961) and 0 otherwise; $RainShock_p$ is a dummy indicating whether province *p* experienced weather shocks during the famine; α_p and γ_t are province and time fixed effects, respectively; and ε_{pt} is an error term. *b* is the parameter of interest.

The results are presented in Table A3. As shown in Columns (1) and (2), the coefficients for both outcome variables are significantly negative. These results suggest that extreme weather during the famine led to a substantial reduction in per capita grain output, and the government accordingly reduced its grain procurement from these regions. As a consequence, regions that experienced a decline in grain output due to extreme weather shocks may have retained more of their grain, thereby suffering less from the famine. This explains why extreme weather shocks are negatively associated with famine severity at the prefectural level, as reported in Table 3 of the main text.

As mentioned in the main text, weather data is available for only 139 prefectures, representing less than half of all prefectures, raising concerns about potential sample selection bias in our analysis. To address this issue, we provide evidence that these 139 prefectures in the IV regression sample do not significantly differ from the full sample in most observable characteristics.

In Table A3: DID Estimates of the Effects of Extreme Weather Shocks on Per Capita Grain Output and Per Capita Grain Procurement

	(1) Per Capita Grain Output	(2) Per Capita Grain Procurement
GF×Rainfall Shock	-0.161**	-0.176*
	(0.072)	(0.098)
Prefecture FE	Y	Y
Year FE	Y	Y
Observations	261	252
R-squared	0.744	0.772

Notes: This table reports the IV estimates of the effects of extreme weather shocks on per capita grain output and per capita grain procurement during the famine, specifically presenting estimates of the coefficients from Equation (2′). The sample period is 1953-1961. The control variables include the prefecture and year fixed effects.
Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table , we compare the 139 prefectures and the full sample across various characteristics, including the Gini coefficient, rural birth population share, rural birth cohort size gap, GDP, income per capita, fiscal expenditure per capita, population density, immigrant population share, agricultural and industrial output share in GDP. For most variables, the mean differences between the IV regression sample and the full sample are not statistically significant. The only exceptions are population density and the agricultural output share in GDP, both of which show statistically significant differences at the 5% level. Overall, the evidence suggests that the selection bias in the IV regression sample is minimal.

A7. More DID Results

As shown in Figure 8 in the main text, the population compositions of the 1981–1985 birth cohorts were unaffected by the famine, making them a suitable choice for the control groups. Therefore, We perform additional DID regressions using the 1962–1980 and 1981–1985 birth cohorts—whose population composition was and was not affected by the famine, respectively—as the treatment and control groups. The results presented in Table A5 indicate that, compared to the 1981–1985 birth cohorts, the famine induced a much higher rural population share for the 1962–1980 birth cohorts, which, in turn, led to a higher Gini coefficient for them.

To further validate these findings, we conduct a placebo test, using the 1950–1956 and 1981–1985 birth cohorts—both of which had population compositions largely unaffected by the famine—as the treatment and control groups, respectively. Table presents the corresponding estimation results, including both the reduced-form estimates and the IV regression results. The results indicate that both the first- and second-stage coefficients are close to zero and statistically insignificant, and the reduced-form estimate is also not significant. These results further confirm that population composition could be an important mechanism through which the famine has affected long-term income inequality.

A8. Overidentification Test

As previously discussed, one concern is raised regarding the validity of the famine severity as the IV for the post-famine fertility or population structure as follows: the famine may affect the income inequality of affected regions through other channels rather than affecting the post-famine fertility structure. Undoubtedly, the famine could have comprehensive effects on affected regions, even in the long run. Specifically, the famine may affect regional institutions, physical and human capital investment, productivity, among others, all of which could have non-negligible effects on economic development and income inequality.

To further verify the validity of the famine as an IV for the post-famine fertility structure, we present another IV to perform an overidentification test. As shown in the literature, China's population control policy is more strictly implemented in urban areas than the rural ones, and such a two-tier population policy also induces a much higher rural fertility than the urban one (Zhang, 2017; Wang and Zhang, 2018). Therefore, we can use the implementation intensity of the population control policy as another IV for China's rural–urban fertility structure or population composition. The population control policy could be a valid IV because it directly affects fertility and is unlikely to affect the income inequality of affected cohorts by affecting other factors rather than fertility, such as institutions and regional productivity several decades later.

Given that China's one-child policy (OCP) was implemented after 1979 and the famine did not affect the population composition of the post-1980 birth cohorts, the OCP is not a feasible IV for the overidentification test. However, even before the implementation of the OCP, China has begun a voluntary yet strong family planning campaign with the slogan “Later, Longer, and Fewer” (LLF) in the early 1970s. This policy was very successful, and China’s overall fertility rate was halved during 1970 and 1978 (Zhang, 2017). Given that the famine also affects the rural–urban fertility structure of the 1970–1978 birth cohorts, we can use both the famine severity and implementation intensity of the LLF policy to instrument for the population composition of these birth cohorts to perform the overidentification test.

The LLF policy varies only at the provincial level, therefore, its implementation intensity was uniform across all prefectures within the same province. We can conduct the overidentification test at the prefectural level. Specifically, we estimate the following cross-section regression at the prefectural level:

$$Rshare_p = c_1 + \mathrm{b}FM_p + dL_p + \theta_1 X_p + \varepsilon_{p1}, \tag{3'}$$

$$Gini_p = c_2 + \beta\ Rshare_p + \theta_2 X_p + \varepsilon_{p2}. \tag{4'}$$

These regressions are similar to the Equations in section 4.1 of the main text; however, they differ from those in section 4.3, as we perform a cross-sectional regression here rather than a DID regression with panel data. L_p represents the implementation intensity of the LLF policy in prefecture *p*, and X_p are control variables, which are the same as those in Table 2.

In practice, we use the birth planning program timing to proxy for the implementation intensity of the LLF policy at the provincial level. As shown in Babiarz et al. (2018), the LLF policy timing shows significant variation across provinces. Table demonstrates the LLF implementation time for different provinces. Some provinces initiated the policy as early as in 1970, and some others launched the campaign after 1975. Given that we focus on the 1970–1978 birth cohorts, we can use the LLF policy timing to construct the variable of the policy implementation intensity for these cohorts. Specifically, if the policy was initiated in one province in 1970, then, all the 1970–1978 birth cohorts in this province were affected. Thus, we assign the number 1 as the policy implementation intensity to this province. Similarly, if a province initiated the policy in 1971, then 8 (1971–1978) out of 9 (1970–1978) birth cohorts were affected. Thus, we assign the number 8/9 as the policy implementation intensity to this province, and so on.

We estimate Equations $(2')$ and $(3')$ and report the results in Column (1) and (2) of Table . Column (1) shows the IV estimates without controls and the coefficient of rural population share is 0.107 and significant at the 5% level. Column (2) reports the IV estimates with controls and the corresponding coefficient is 0.140 and significant at the 1% level. The bottom Panel presents the results of overidentification test. The P values is much larger than 0.1 in Column (1) and nearly 0.10 in

Column (2), which indicates that we cannot reject the null hypothesis that the IVs satisfy the exclusion restrictions.

In the above analysis, we assume that the birth planning policy affects the fertility immediately after its implementation. However, given the ten-month child-bearing period, the policy may have a time lag between its implementation and being effective to reduce fertility. For instance, many women may have already been pregnant when the policy was announced. Thus, the fertility will not decrease immediately in the future months. The policy is likely to reduce fertility effectively after 10 months or one year of its implementation. Therefore, we now assume that the policy affects the fertility after one year of its announcement and calculate the policy implementation intensity for the 1971–1978 birth cohorts and redo regressions (2′) and (3′).

Column (3) and (4) of Table report the estimate results. Compared to Column (1) and (2) of Table , the estimates in Column (3) and (4) show little change, which confirms the robustness of our identification.

A9. Famine Severity and the Implementation Intensity of the One-Child Policy (OCP)

Other factors, such as the population control policies, could also affect post-famine rural-urban fertility structure. However, if these factors are not systematically correlated with famine severity, they would not pose a significant problem. To address this, we now provide direct evidence on the correlation between famine severity and the implementation intensity of the population control policies. Figure A3 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 A4 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.

References

Addoum, J.M., Ng, D.T. and Ortiz-Bobea, A., 2020. "Temperature shocks and establishment sales," The Review of Financial Studies, 33(3): 1331-1366.

Babiarz, Kimberly, Paul, Ma, Grant, Miller and Shige, Song. 2018. "The Limits (and Human Costs) of Population Policy: Fertility Decline and Sex Selection in China under Mao," NBER Working Paper 25130.

Corno, L., Hildebrandt, N. and Voena, A., 2020. "Age of marriage, weather shocks, and the direction of marriage payments," Econometrica, 88(3): 879-915.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," American Economic Review 91(4): 795–813.

Meng, Xin, Nancy Qian, and Pierre Yared. 2015. "The Institutional Causes of China's Great Famine, 1959-1961," Review of Economic Studies 82(4): 1568–1611.

Shah, M. and Steinberg, B.M., 2017. "Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital," Journal of Political Economy, 125(2): 527-561.

Wang, Xuebo and Junsen Zhang. 2018. "Beyond the Quantity–Quality Tradeoff: Population Control Policy and Human Capital Investment," Journal of Development Economics 135:222–34.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and Its Effects on Family Outcomes," Journal of Economic Perspectives 31(1): 1–21.

Figure A1: Rural Birth Cohort Size Gap during the Great Famine (1959–1961) (Prefectural Level)

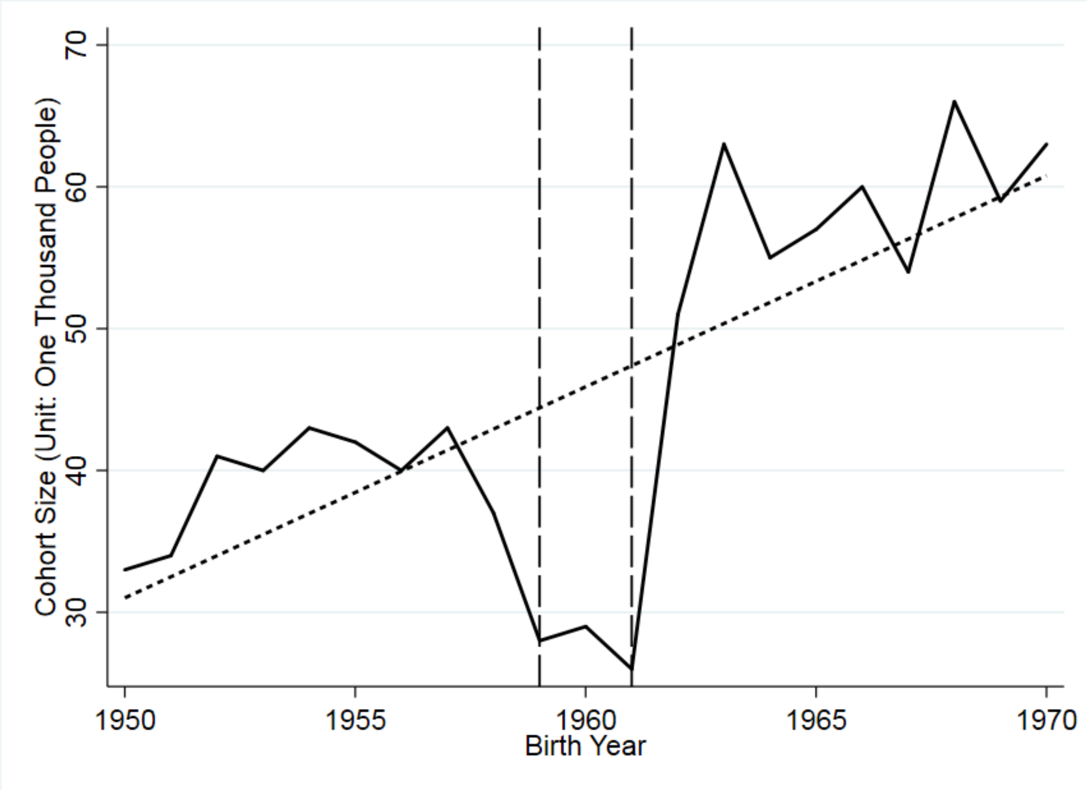

Figure A2: Coefficients of the Interactions Excess Mortality Rate×Birth Cohort (1962–1984) in Equation (5) (Prefectural Level)

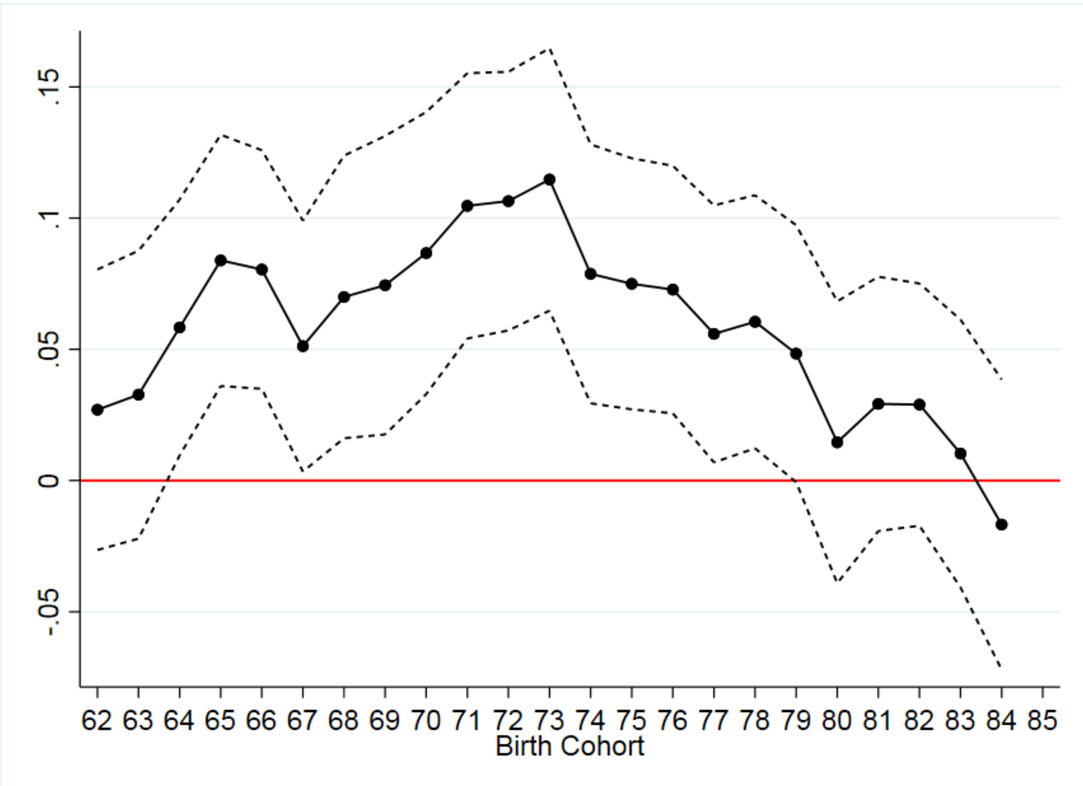

Figure A3: Fines for Excess Fertility in 1979 and the Average Rural Excess Mortality Rate in 1959–1961 (Provincial Level)

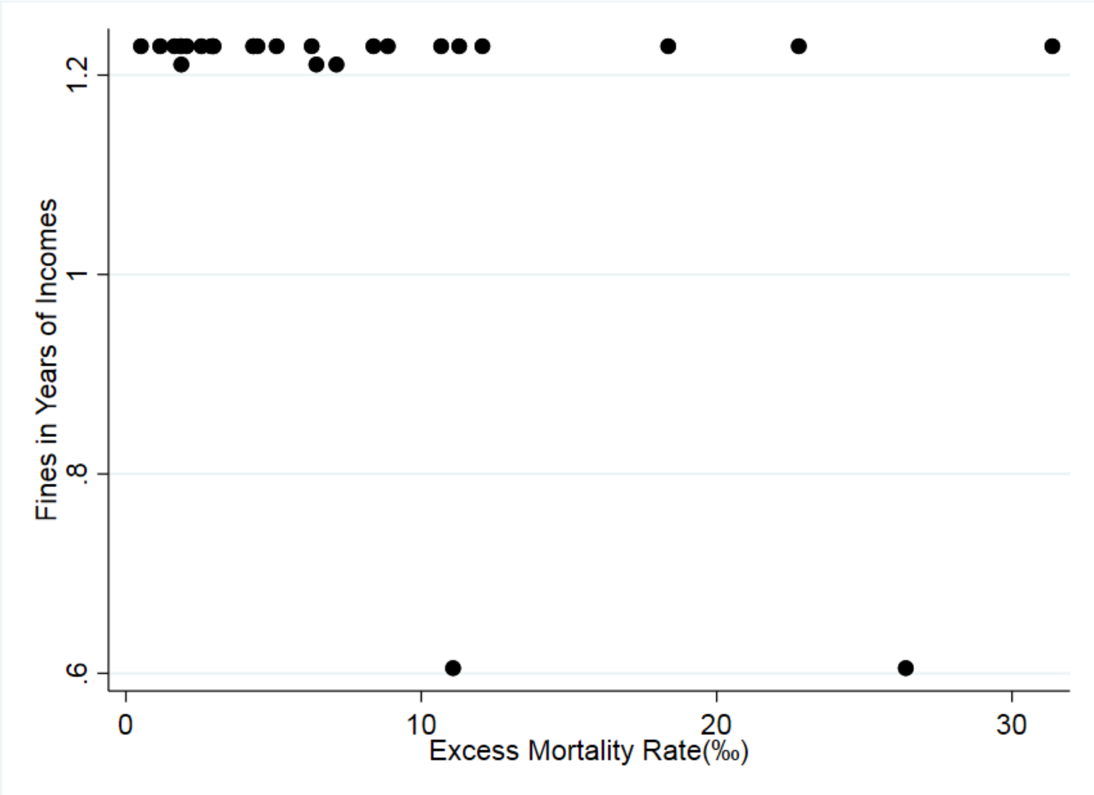

Figure A4: Fines for Excess Fertility in 1985 and Average Rural Excess Mortality Rate in 1959–1961 (Provincial Level)

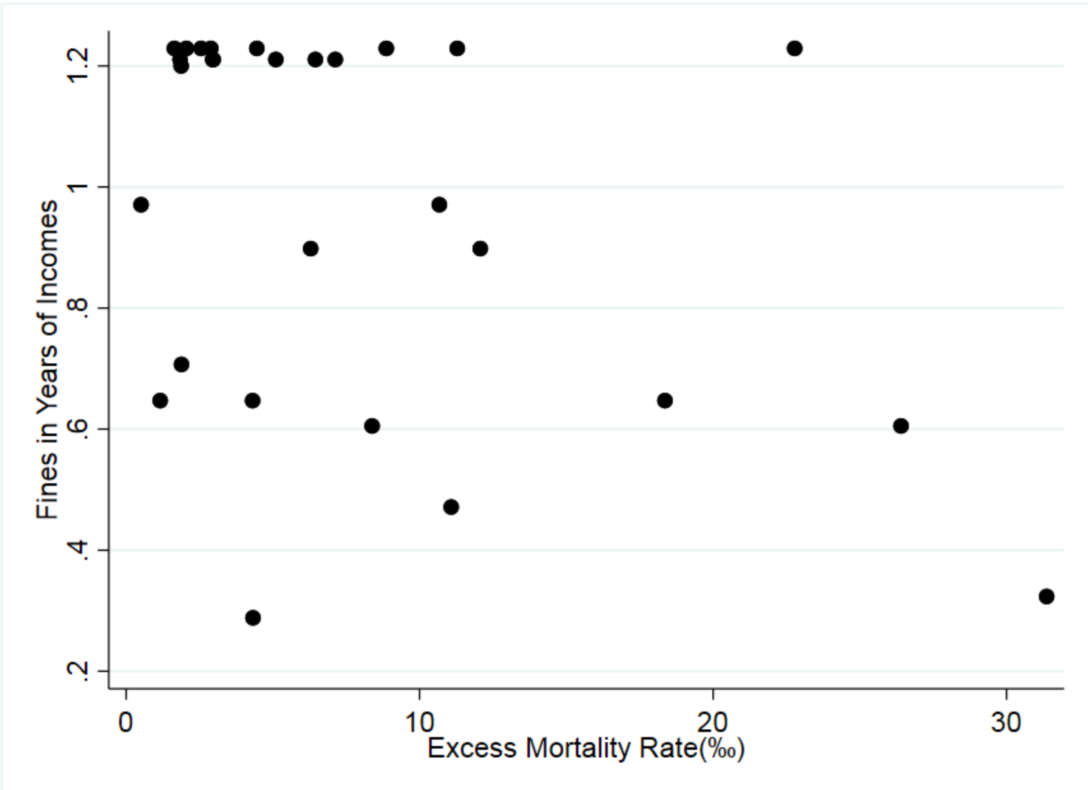

Table A1: Summary Statistics for Control Variables

Variables	Definition	Obs.	Mean	S.D.
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
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
Primary	Agricultural output share in GDP	291	0.186	0.117
Secondary	Industrial output share in GDP	291	0.444	0.126

Table A2: DID Estimates of the Effect of the Rural Population Share on the Gini Coefficient with Migration and Price Adjustments

	Dependent variable: Gini Coefficient			
	RF	IV	IV	IV
	(1)	(2)	(3)	(4)
Rshare		**1.037*****	**0.874*****	**0.881*****
		(0.289)	(0.261)	(0.281)
		First Stage		
Famine Severity× T_t	**0.038*****	**0.042*****	**0.046*****	**0.044*****
	(0.010)	(0.008)	(0.009)	(0.009)
GDP× T_t	-0.000		0.000	0.001
	(0.003)		(0.003)	(0.003)
Income× T_t	-0.031***		-0.030***	-0.031***
	(0.007)		(0.009)	(0.009)
Expenditure× T_t	0.000		-0.021***	-0.019**
	(0.005)		(0.005)	(0.008)
Density× T_t	-0.001		-0.010***	-0.006*
	(0.002)		(0.003)	(0.003)
Immigrant× T_t	0.019			0.088**
	(0.025)			(0.037)
Primary× T_t	0.086***			0.205***
	(0.028)			(0.050)
Secondary× T_t	-0.008			0.109**
	(0.018)			(0.045)
Prefecture FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Kleibergen-Paap F statistic		28.841	29.416	25.350
Observations	7306	7306	7306	7306

Notes: This table reports the DID estimates of the effect of the rural population share on the Gini coefficient for the 1962–1980 birth cohorts with migration and price adjustments, using the 1950–1956 birth cohorts as the control group and the 1962–1980 birth cohorts as the treatment group. The control variables are the same as those in Table 2 in the main text. The table includes both reduced form (RF) and IV estimation results.
Robust standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.10.

Table A3: DID Estimates of the Effects of Extreme Weather Shocks on Per Capita Grain Output and Per Capita Grain Procurement

	(1) Per Capita Grain Output	(2) Per Capita Grain Procurement
GF×Rainfall Shock	-0.161**	-0.176*
	(0.072)	(0.098)
Prefecture FE	Y	Y
Year FE	Y	Y
Observations	261	252
R-squared	0.744	0.772

Notes: This table reports the IV estimates of the effects of extreme weather shocks on per capita grain output and per capita grain procurement during the famine, specifically presenting estimates of the coefficients from Equation (2′). The sample period is 1953-1961. The control variables include the prefecture and year fixed effects.
Robust standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A4: Balance Test between the Full Sample and IV Sample of Prefectures

Variables	Full Sample		IV Sample		Difference	P Value
	Obs.	Mean	Obs.	Mean		
Gini	291	0.396	139	0.395	0.001	0.886
Rshare	291	0.772	139	0.771	0.001	0.954
Rural Birth Cohort Size Gap	291	0.405	139	0.397	0.008	0.632
GDP	291	3.538	139	3.723	-0.185	0.103
Income	291	6.165	139	6.183	-0.018	0.623
Expenditure	291	7.225	139	7.234	-0.009	0.874
Density	291	5.320	139	5.593	-0.273	0.030
Immigrant	291	-0.007	139	-0.009	0.003	0.778
Primary	291	0.186	139	0.163	0.023	0.046
Secondary	291	0.444	139	0.460	-0.015	0.226

Table A5: Alternative DID Estimates of the Effect of the Rural Population Share on the Gini Coefficient for the 1962–1980 Birth Cohorts

	Dependent variable: Gini Coefficient			
	RF	IV	IV	IV
	(1)	(2)	(3)	(4)
Rshare		0.485**	0.754***	0.862***
		(0.191)	(0.200)	(0.227)
			First Stage	
Famine Severity× T_t	0.049***	**0.055*****	**0.061*****	**0.057*****
	(0.010)	(0.008)	(0.009)	(0.009)
GDP× T_t	0.001		-0.004	-0.002
	(0.003)		(0.003)	(0.004)
Income× T_t	-0.008		-0.008	-0.024**
	(0.008)		(0.009)	(0.011)
Expenditure× T_t	0.026***		0.032***	0.013*
	(0.006)		(0.005)	(0.007)
Density× T_t	0.008***		0.007**	0.004
	(0.002)		(0.003)	(0.003)
Float× T_t	0.097***			0.146***
	(0.020)			(0.031)
Primary× T_t	0.019			-0.038
	(0.043)			(0.043)
Secondary× T_t	-0.018			0.004
	(0.024)			(0.029)
Prefecture FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Kleibergen-Paap F statistic		43.536	43.915	37.104
Observations	6648	6648	6648	6648

Notes: This table reports the DID estimates of the effect of the rural population share on the Gini coefficient for the 1962–1980 birth cohorts, using the 1981–1985 birth cohorts as the control group and the 1962–1980 birth cohorts as the treatment group. The control variables are the same as those in Table 2. The table includes both reduced form (RF) and IV estimation results.

Robust standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.10

Table A6: Placebo Estimates of the Effect of the Rural Population Share on the Gini Coefficient

	Dependent variable: Gini Coefficient	
	RF	IV
		(3)
Rshare		-0.120
		(0.958)
		1st Stage
Famine Severity× T_t	0.002	-0.014
	(0.013)	(0.011)
Controls	Y	Y
Prefecture FE	Y	Y
Year FE	Y	Y
Kleibergen-Paap F statistic		1.743
Observations	3,312	3,312
R-squared	0.589	0.112

Notes: The treatment and control groups are the 1981–1985 and 1950–1956 birth cohorts, respectively.
Robust standard errors in parentheses: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A7: LLF Birth Planning Program Timing

Year	Num. of Provinces	Remarks
1970	3	Jiangsu, Guangdong, Hainan
1971	7	Liaoning, Jilin, Guangxi, Chongqing, Sichuan, Guizhou, Gansu
1972	8	Tianjin, Hebei, Heilongjiang, Jiangxi, Shandong, Hubei, Yunnan, Qinghai
1973	5	Shanxi, Shanghai, Fujian, Shaanxi, Ningxia,
1974	3	Anhui, Henan, Hunan
1975	1	Xinjiang
1979	1	Inner Mongolia

Table A8: Overidentification Test

	Current Period: 1970-1978		One Period Lag: 1971-1978	
	Without Controls	With Controls	Without Controls	With Controls
	(1)	(2)	(3)	(4)
Rshare	0.107^{**}	0.140^{***}	0.108^{**}	0.143^{***}
	(0.046)	(0.048)	(0.047)	(0.048)
	First Stage		First Stage	
Famine Severity	0.383^{***}	0.360^{***}	0.385^{***}	0.361^{***}
	(0.060)	(0.069)	(0.060)	(0.069)
Birth Control Intensity	0.053	0.089	0.050	0.089^{*}
	(0.055)	(0.056)	(0.051)	(0.053)
Controls		Y		Y
Prefecture FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Kleibergen-Paap F statistic	20.411	15.466	20.480	15.518
Observations	288	223	288	223
Over-identification Test				
Hansen J statistic	0.130	2.834	0.085	2.479
P Value	0.719	0.092	0.770	0.115

Notes: The sample period is 1970-1978 for Columns (1) and (2) and 1971-1978 for Columns (3) and (4). The control variables are the same as those in 2 in the main text. The table includes IV estimation results with and without controls.
Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.