Entrepreneurship and the School of Hard Knocks: Evidence from China's Great Famine

Junhong Chu, I.P.L. Png, & Junjian Yi*

Abstract

Many entrepreneurs credit the School of Hard Knocks for their education. Such claims challenge the social value of entrepreneurship schooling and training. Here, we exploit geographical differences in the intensity of China's Great Famine as a quasi experiment to investigate the effect of hardship on entrepreneurship. We find robust evidence that, in counties that experienced greater hardship, the share of small, young privately-owned businesses is higher and individual residents are more likely to become entrepreneurs. The effect on individual entrepreneurship applies across all levels of hardship and is more pronounced among those aged between seven and twelve when the famine started. We show that the effect is due to conditioning towards greater risk tolerance.

Keywords: Entrepreneurship; Hardship; Conditioning; Risk Tolerance; China's Great Famine

JEL codes: L26, O15, O17

*National University of Singapore: junhong.chu@gmail.com, iplpng@gmail.com, junjian.yi@gmail.com. We thank Shangjin Wei, Chien-yi M. Png, Richard Lai, Sng Tuan Hwee and other NUS colleagues, participants at seminars at Cornell University, Syracuse University, the AIEA-NBER 2006 Conference for helpful advice, and Xiong Xi for outstanding research assistance. This research was funded by Academic Research Fund grant MOE2013-T2-2-045 (R-313-000-109-112).

1 Introduction

> "By definition, an entrepreneur is one who takes risk. It's an attitude and an appetite, one which may be hardwired into one's personality. ... My grandfather referred to me as having an MBA from the School of Hard Knocks", Stephen Greer (2010).

> "From the military school of life – What does not kill me, strengthens me", Nietzsche (1990: 8).

Given the importance of new enterprise for business growth and economic policy (Guzman and Stern, 2016), and the widespread belief that hardship develops entrepreneurs (witness the number of entrepreneurship training programs called "boot camps"), it seems natural to ask whether hardship indeed contributes to entrepreneurship, and if so, how. Does hardship contribute by winnowing out less suited people or does it contribute by conditioning behavior that fosters entrepreneurship?[1] Which dimensions of personality and attitudes do hardship select or cultivate?

Despite the importance of these questions, there appears to be no relevant research (Ahmetoglu et al., 2017). The challenge to any study is causal inference. Prior research into the effect of hardship on personality has relied on natural experiments such as economic shocks and natural disasters (Malmendier et al., 2011; Cameron and Shah, 2015; Bernile et al., 2017; Hanaoka et al., 2018). Adopting a similar approach, we investigate the effect of hardship on engagement in entrepreneurship by exploiting county-level differences in the intensity of China's Great Famine of 1959-61. The famine followed the Chinese government's headlong rush to collectivize agriculture and ramp up industrial production. To the extent that the famine exogenously imposed hardship to different degrees by geography, it presents a unique opportunity to identify the effect of previous life experience of hardship on personality and entrepreneurship.

[1] Interviewed by the *Harvard Business Review*, Stephen Greer, who built a multinational enterprise with annual sales of $200 million, first asserted that risk-taking is "hardwired", but then, referring to the "School of Hard Knocks", suggested that risk-taking and entrepreneurship are cultivated. Likewise, Nietzsche's characterization of the "military school of life" emphasizes both selection and cultivation.

This research presents three challenges: (i) How to measure the degree of hardship caused by the famine, (ii) How to account for possible endogeneity in the statistical relation between hardship and entrepreneurship, and (iii) How to distinguish whether the relation between hardship and entrepreneurship is due to selective mortality or conditioning.

During the period of study, the hukou (household registration) system severely restricted individual mobility within China. The restriction on mobility enables us to construct an intuitive county-level measure of hardship. Following Yao (1999), Meng and Qian (2009), and Meng et al. (2015), we calculate the population of each county by year of birth, and, fitting a county-specific trend to the cohorts born before or after the famine, project the counterfactual number of persons in the cohorts born during the famine. Then we calculate the rate of cohort loss as the difference between the projected and actually recorded number of people divided by the projected population. Our first construct, to represent the intensity of the famine and hardship in the county, is the relative cohort loss, defined as the difference in the rate of cohort loss between the famine and normal birth cohorts.

Motivated by research into climate change (Deschenes and Greenstone, 2007; Schlenker and Roberts, 2009), our second construct is thermal agricultural productivity, defined as the exposure to heat during the growing season. We validate the two constructs, showing that the rate of cohort loss is negatively correlated with the previous year's thermal agricultural productivity in normal periods, but positively correlated during famine periods. This contrast accords with prior findings that the famine was due to excessive procurement of grain from rural areas by officials projecting output from pre-famine years and officials being unwilling or unable to adjust procurement targets or redistribute food during the famine (Kung and Chen, 2011; Meng et al., 2015).

The various areas of China differed in the intensity of the Great Famine as well as subsequent entrepreneurship, measured either as the proportion of enterprises that are privately-owned and young, or as the likelihood that an individual is self-employed or owns a business. To account for possible endogeneity in the statistical relation between the famine and entrepreneurship, we apply an instrumental variables (IV) strategy. The essence of the IVs is to isolate the variation in the relative cohort loss induced by weather shocks, *abstracted from county fixed effects.*

We find an economically and statistically significant relation between the severity of the famine, as represented by the relative cohort loss, and the level of entrepreneurship in the county four to five decades later. If the relative cohort loss is one percent higher, the share of small, young privately-owned businesses is 0.5 percent higher and each individual resident is 0.9 percent more likely to be an entrepreneur. The individual effect is significant across all birth cohorts, and is more pronounced among those aged between seven and twelve years at the onset of the famine. Further, we also find an economically and statistically significant relation between the severity of the famine and the degree of entrepreneurial success, as measured by the entrepreneur's income.

The hardship imposed by the Great Famine ranged to extremes. We estimate the effects of less extreme hardship in two ways. One, following Meng and Qian (2009), applies quantile regression and shows that the positive effect of hardship on entrepreneurship does not vary significantly across quantiles. The other exploits the lower severity of the famine in urban areas. We find that the effect of hardship on entrepreneurship is slightly (but not statistically significantly) lower in urban areas than rural areas.

How did greater hardship during the Great Famine give rise to more entrepreneurship in later years? Was it because the famine selectively killed less enterprising people or because the famine conditioned individuals to be more enterprising? We show that, by county, *conditional on the population*, the number of entrepreneurs increased with the severity of the famine. Hence, the relation between hardship and entrepreneurship cannot be completely explained by differences in mortality, and so, must be due, at least in part, to operant conditioning of personality. Further, we find an economically and statistically significant relation between the rate of cohort loss and investments in shares. Based on previous research that characterizes risk aversion by investment in shares (Hvide and Panos, 2014), we interpret the estimates as showing that hardship cultivated more tolerance of risk.[2]

Our work contributes to three areas of research. Foremost, it adds to a better appreciation

[2]We rule out hardship increasing entrepreneurship by conditioning confidence, tenacity, resilience, opportunism, or skirting of authority. Also, our estimates are inconsistent with alternative explanations. One is the famine tilting the gender ratio towards females and inducing people with daughters to engage more in entrepreneurship to enhance marriage prospects (Wei and Zhang, 2011). Another is the famine leaving more people without sons to support them in old age, and so, causing them to engage more in entrepreneurship to be more self-sufficient in later years.

of the individual factors that lead people to choose careers in entrepreneurship. Economic research has emphasized personal skills. Lazear (2005) shows that individuals with balanced skills are more likely to engage in entrepreneurship. Empirically, Stanford MBA graduates with varied work experience and education were more likely to start businesses than those with focused backgrounds (see also Lazear (2004)). Older societies are less entrepreneurial as young people (who are more creative and risk-taking) are blocked from high-level positions that would teach them the skills necessary to start businesses (Liang et al., 2018).

Our contribution is to emphasize psychological factors in entrepreneurship. We show that hardship conditions individuals to become entrepreneurs. This finding also contributes to the long-standing debate over whether entrepreneurs are "born or bred". In a Swedish study comparing the influence of birth vis-à-vis adoptive parents on entrepreneurship among their children, Lindquist et al. (2015) found that the effect of pre-birth factors was half that of post-birth factors (notably role-modeling). We show that early life experience founds entrepreneurship in adulthood, and is consistent with entrepreneurship being nurtured.

Second, our work contributes to research into how early life experience shapes personality, and thereby, influences socio-economic outcomes in adulthood (Heckman, 2007). Generally, personality traits are not fixed, but rather, vary with experience and investment (Almlund et al., 2011). For instance, the Perry Preschool program reduced aggressive, antisocial, and rule-breaking behavior, which in turn, improved academic achievement, income, and employment, and reduced criminality when the individuals grew up (Heckman et al., 2013). Our contribution is to show that early life experience also affects career choices, specifically, hardship during the famine conditioned individual personality to foster entrepreneurship in later years.

Third, our work contributes to research into the long-term socio-economic consequences of famine. Most studies of China's Great Famine (Almond et al., 2010; Mu and Zhang, 2011; Tan et al., 2014) focus on the effect of in-utero experience of the famine, using post-famine birth cohorts as a control group for the cohort born during the famine. We cannot apply this empirical strategy as engagement in entrepreneurship is subject to strong family and peer effects (Lindquist et al., 2015; Lerner and Malmendier, 2013). Moreover, we find that the effect of hardship on entrepreneurship is more pronounced among those aged between seven and twelve when the famine began. Our study is closely related to Chen and de la

Rupelle (2016), who show that the severity of the famine is positively correlated with household savings. They interpret the results as the outcome of altruistic parents instilling lower time preference among their children. These findings are consistent with ours to the extent that greater patience and greater risk tolerance both lead people to increase savings and investment.

The remainder of this paper proceeds as follows. Section 2 describes the institutional setting of China's Great Famine, Section 3 presents the data, and Section 4 discusses our key constructs and validation. Section 5 explains our empirical strategy for causal identification. Section 6 reports our estimates and suggests possible mechanisms, and Section 7 considers alternative explanations of the relation between the famine and entrepreneurship. Section 8 concludes with discussion of policy and business implications, limitations of the analysis and findings, and directions for future work. An Appendix presents details of the data and supplementary estimates.

2 Institutional Setting

In 1958, the Communist government of China, led by Chairman Mao Zedong, launched the Great Leap Forward, aiming to surpass the industrial output of Britain within 15 years and the United States after a further 15 years. The government collectivized agriculture and forced people to eat in communal canteens, and redirected up to 100 million people from agriculture to collective works and industrial production, while vigorously enforcing increased procurement of food from rural areas for distribution to urban residents and export to foreign countries (Meng and Qian, 2009; Kasahara and Li, 2018). The result was a famine, starting in the winter of 1959 and lasting until 1961, when Chairman Mao reluctantly reversed his policies. The government blamed the famine on bad weather and is officially named as "three years of natural disasters" (*san nian zi ran zai hai*) (Meng and Qian, 2009).

The severity of the famine during the Great Leap Forward was not uniform across China. Figure 1 depicts the severity by county, as represented by our construct, the relative rate of cohort loss, defined as the difference in the rate of cohort loss between famine and normal periods, with darker colors representing more severe famine. Apparently, there was substantial geographical variation in the severity of the famine.

Meng et al. (2015) show that, nationally, China produced sufficient food to meet domestic consumption needs. Shortages of food in particular areas were due to the inflexibility of local government officials. They had set or agreed to targets for procurement, likely based on the previous year's harvest. In the face of shortages due to random fluctuations in weather, they could not or would not relax procurement or redistribute food geographically. Kung and Chen (2011) attribute the localized famines in part to the career concerns of provincial officials. Officials with lower rank in the Communist Party (hypothesized as aspiring to a higher rank) extracted about 3 percent more grain from their provinces than more senior officials.

Although the Chinese government supplied food to urban areas, people in the cities did not escape the famine. According to official Chinese government statistics, the ratio of the death rate during the peak of the famine in 1960 to the pre-famine death rate in 1957 was 1.6 and 2.6 in urban and rural areas respectively (Lin and Yang 2000: Table 2).

In 1958, at the start of the Great Leap Forward, China established the household registration (hukou) system to restrict domestic migration, particularly to prevent migration from rural to urban areas. Under the hukou system, individuals have legal status only in their registered hukou place of residence. The hukou system regulates almost every aspect of an individual's life, including birth, marriage, divorce, public housing, education, and employment. Importantly, local governments issued food rations only to people with the local hukou. Furthermore, the government permits changes of hukou registration only under very stringent conditions. Indeed, the hukou system persists and remains controversial, and the Chinese government is only now taking tentative steps to reform the system.

3 Data

This study uses data from multiple sources. Administratively, China is divided into provinces, which are further divided into prefectures, which are further divided into counties. The administrative division of China includes an overlapping category of "city", which might be at provincial, prefectural, or county level. We carry out the analysis mainly by county, which is the lowest administrative level at which government and survey data are available.

Our principal sources of information on demography, business, individual employment, investments, and personality are: (i) 1990 Population Census for information on population by county and year of birth, used to estimate cohort losses; (ii) 2004 Economic Census for information on private-sector enterprises; and (iii) 2000 Population Census, 2005 1% Population Census, China Family Panel Studies (CFPS) 2010, China Household Finance Survey (CHFS) 2011, and China General Social Survey (CGSS) 2008, 2010, 2011, 2012, and 2013 for information on individual employment, investments, and personality traits.

We use the 1990 Population Census rather than earlier or later censuses for various reasons. In matching the population, business, and individual data, we drop all counties which changed boundaries. The number of counties with changes in boundaries is over 800 between 1982 and 2005, and about 300 between 1990 and 2005. So, to minimize the loss of data, we prefer the 1990 to the 1982 Population Census. Further, we prefer the 1990 to the 2000 Population Census as migration was more limited at earlier times and the 1990 Census is considered to be more reliable (Lavely 2001; Zhang and Zhao 2006). We use the 2000 Census and the combination of the 1990 and 2000 Censuses in robustness tests.[3]

The 1990 Census reports only current residence and hukou registration. We limit the sample to people whose county of residence is the same as their hukou registration and have lived in the county for five or more years, and assume that their birthplace is in that county. The 2000 Census reports place of birth and county of residence. We limit the sample to people whose birthplace is in the county of residence.

We collect county-level data on private-sector enterprises from the 2004 Economic Census, which was the first comprehensive national record of business activity. We define a private enterprise as entrepreneurial if it is less than 10 years old and has fewer than 100 employees. The purpose of limiting the employment is to exclude large private businesses formed by spin off from the government or state-owned enterprises. In robustness checks, we define a private enterprise as entrepreneurial if it is less than 10 years old and has fewer than 500 employees. Our results are robust to this alternative measure.

Further, we collect individual data on employment and income from the 2005 1% Population Census, and define entrepreneurship in two ways – broadly as those who own an

[3]The 1995 mini-Census is not available at individual level or county level.

enterprise in the private sector or are self-employed, and narrowly as those who own a private-sector enterprise. We collect data on individual employment, investments, self-confidence, tenacity, resilience, and opportunism and skirting of authority from multiple household surveys, specifically, CFPS 2010, CHFS 2011, and CGSS 2008, 2010, 2011, 2012, and 2013.

From the China Meteorological Administration, we download daily instrumental weather records for 727 stations over the years 1951-70. We associate each county with the nearest weather station by the Euclidean distance from the county seat, and thus, on average, match three counties to one station. As explained below, we use the weather to construct instruments, not as the explanatory variable of interest, which is severity of the famine at county level.[4]

Table 1 presents summary statistics of the data. Table 1, panel (a), summarizes the 2004 Economic Census data for 2,194 counties. Entrepreneurial firms account for 47.1 percent (s.d. 16.0 percent) of all enterprises and 17.8 percent (s.d. 14.0 percent) of sales, and their median age is 37.7 months (s.d. 12.3 months), as compared with 47.4 months (s.d. 18.1 months) for all private enterprises. Entrepreneurial firms are more profitable. Even though they take up less than one-fifth of sales, they account for more than half of the profit. The share of entrepreneurial firms in employment is 20.5 percent, which slightly exceeds their share of sales.

Table 1, panel (b), summarizes the 2005 1% Population Census data. To focus on the effects of the famine, we limit the sample to people born before 1962. The Census covers 639,443 such persons in 2,212 counties, of whom 0.7 percent own private enterprises (1.1 percent and 0.3 percent among males and females, respectively), and 2.7 percent are self-employed (3.8 percent and 1.6 percent among males and females, respectively).[5]

Table 1, panel (c), reports summary statistics for the surveys, limited to people born before

[4]To the best of our knowledge, this is the most granular dataset of daily instrumental observations of weather available for China. Gridded datasets may seem to be more precise but they are based on interpolations of instrumental observations, and so, are not truly finer than the data that we use.

[5]We exclude counties whose boundaries changed between 1990 and 2005. The excluded counties are very similar to those included in our analysis. The percentages of individuals who owned private enterprises or were self-employed were respectively 0.8 and 3.0 in the excluded counties, as compared with 0.7 and 2.7 in the included counties (differences from Table 1(b) due to rounding). These differences are not statistically significant.

1962. These surveys cover 21,965 such individuals in 248 counties, of whom 2.1 percent own private enterprises, 4.3 percent have financial investments, and 2.7 percent participate in the stock market.

4 Constructs and Validation

To characterize the severity of the Great Famine in each county, we follow Yao (1999), Meng and Qian (2009), and Meng et al. (2015), and use the 1990 Population Census to construct the following measure. For each county and year of birth, the rate of cohort loss is the projected, counterfactual population who would have survived until 1990 less the actual population who did survive until 1990, divided by the projected population. Formally, for county c, and year y, our first key construct, the rate of cohort loss,

$$\mu_{cy} = \frac{\tilde{P}_{cy} - \hat{P}_{cy}}{\tilde{P}_{cy}} = 1 - \frac{\hat{P}_{cy}}{\tilde{P}_{cy}}, \tag{1}$$

where $\tilde{P}_{cy}$ and $\hat{P}_{cy}$ represent the projected and actual populations in 1990, respectively.

The famine affected individual health, education, employment, marriage, and other outcomes in the long term (Meng and Qian, 2009; Almond et al., 2010; Mu and Zhang, 2011). By taking a retrospective view from 1990, our construct for the severity of the famine, the rate of cohort loss, μ_{cy}, accounts for both immediate as well as long-term effects of the famine. We stress that the purpose of calculating the rate of cohort loss is not to measure excess mortality and depressed fertility during the famine as such. Rather, the objective is to represent differences in the geographical intensity of the famine in all its ramifications, including lower wages and worse jobs, up to 1990.

Referring to equation (1), we project the counterfactual population as follows. We use ordinary least squares (OLS) to regress the numbers of people born in the normal years, 1949-57 and 1963-70, as recorded in the 1990 Population Census on a linear year trend. For each county, we estimate the equation,

$$\hat{P}_{cy} = \alpha_{0c} + \alpha_{1c} y + \epsilon_{cy}, \tag{2}$$

where α_{0c} and α_{1c} are respectively the constant and coefficient of the year trend for county c, and ϵ_{cy} is a random error. We start the projection in 1949, the year in which Communist China was founded, as China was racked by war – first, the Japanese invasion, and then, the civil war between the Nationalists and Communists – in the previous two decades.

Then, we use equation (2) to project the population for all years from 1949 to 1970, including 1958-62, the five years not used in the estimation. Figure 2 depicts the projected and actual populations for China as a whole. Evidently, the population in normal years fluctuated randomly around the time trend, and the difference between the counterfactual and actual populations increased sharply in the famine years. The famine was followed by a sharp baby boom in 1963. Figure 2 is quite similar to Meng et al.'s (2015) Figure 1(b).

The major concern with the construct, μ_{cy}, is non-classical measurement error, that is, error that is correlated with the severity of the famine. Classical measurement error is less of a concern because it causes the estimated coefficients to be attenuated, and thus, any statistical inference is conservative. Moreover, as explained below, we apply an IV method to estimate the effect of the famine on entrepreneurship. The method resolves any classical measurement error.

Accordingly, our key concern is with error that is correlated with the severity of the famine. One such source of error is migration between counties. Despite vigorous government enforcement, people did try to move away from famine-stricken areas (Thaxton 2008: 162-170; Yang 2012: 50). However, the 1990 Census enumerates persons by their hukou registration, and so, the impact of error due to migration is limited to people who were able to change their hukou registration. Migration and changes of hukou were severely restricted before the early 1990s (Lin and Yang 2000; Meng and Qian 2009: footnote 23).

The linear projection in equation (2) might give rise to errors, which may or may not be classical. Visually, Figure 2 suggests that the linear projection fits the aggregate pattern well during the normal period. Indeed, the average R^2 of the estimations is 0.490 across all counties, indicating strong predictive power for a very simple regression with just one explanatory variable. To check robustness, we omit the year 1963 from the basis of projection and also apply two non-linear projections, one quadratic and the other exponential. Our results below are robust to these alternative projections.

As an additional check of robustness, we also measure the intensity of the famine by the gender ratio. Mothers under stress produce more adrenal androgens, which increase spontaneous abortion of male fetuses, tilting the gender ratio of births towards females (James, 2015). The famine certainly increased stress on mothers. Based on the 1990 Census, the male/female gender ratio was 1.069 in the pre-famine years 1950-57, and fell to 1.054 in the famine years, 1959-61.[6]

We propose to measure the severity of the Great Famine by the rate of cohort loss, as computed from the 1990 Population Census. To validate this construct, we investigate its relation with predictors of agricultural productivity. Agronomic research shows that exposure to heat during the growing season affects crop yields in a nonlinear way. The temperature must exceed some threshold for plants to absorb heat. Yields increase with temperature above the threshold up to some ceiling of around 32 degrees Celsius, beyond which yields decrease (Ritchie and NeSmith, 1991; Deschenes and Greenstone, 2007).

Accordingly, we measure the potential agricultural productivity in each county and year by the sum of "degree-days", with a floor of 8 degrees Celsius and a ceiling of 32 degrees Celsius, between April 1 and September 30. For each county c and day d, the degree-day is computed according to

$$H_{cd} = \begin{cases} 0 \;\text{ if }\; T_{cd} < 8, \\ T_{cd} - 8 \;\text{ if }\; 8 \leq T_{cd} < 32 \\ 24 \;\text{ if }\; T_{cd} \geq 32, \end{cases} \tag{3}$$

where T_{cd} represents the average temperature in the county on that day. Exposure to extreme temperatures harms plant growth. In robustness checks, we account for the harmful effects through an alternative measure of degree days. Following Ritchie and NeSmith (1991), the alternative measure is $T_{cd} - 8$ if the daily temperature is between 8 and 33 degrees Celsius, $\frac{25}{8}[41 - T_{cd}]$ if the daily average temperature is between 33 and 41 degrees Celsius, and zero otherwise.

Then, we define our second key construct, the *thermal agricultural productivity* in county

[6]The gender ratio, as observed in the 1990 Census, might be subject to non-classical error as a measure of the intensity of the famine. During famines, Chinese parents discriminated in favor of sons, which raised their survival rate relative to daughters, and so, tilted the subsequent observed gender ratio towards males (Mu and Zhang, 2011).

c and year y, as the sum of the degree days during the growing season,

$$\text{Agri}_{cy} = \sum_{d=\text{April 1}}^{\text{Sept 30}} H_{cdy}. \tag{4}$$

This construct is relatively exogenous to manipulation. By contrast, direct measures of agricultural production may have been distorted by government officials during the Great Famine or incompletely adjusted by post-famine statisticians.[7] In the same spirit, Meng and Qian (2009) and Meng et al. (2015) also construct exogenous measures of agricultural productivity, which are based on soil characteristics and monthly weather.

Figure 3 illustrates the evolution of the average thermal agricultural productivity between 1951 and 1970 over all counties. Thermal agricultural productivity fluctuated and was only slightly below average in 1958-60, and actually exceeded the average in 1961. These patterns refute the official explanation that the famine was caused by bad weather.

Having constructed the measure of thermal agricultural productivity, we estimate the following regression to validate the rate of cohort loss as a measure of the severity of the famine,

$$\mu_{cy} = \beta_0 + \beta_1 \text{Agri}_{c,y-1} + \sum_{y=1953}^{1970} \lambda_y \text{Agri}_{c,y-1} D_y + \sum_{y=1953}^{1970} \eta_y D_y + \nu_c + \varepsilon_{cy}, \tag{5}$$

where D_y is a year indicator, and β_0, β_1, λ_y, and η_y are coefficients to be estimated, ν_c are county fixed effects, and ε_{cy} is a random error. This equation allows the effect of the previous year's thermal agricultural productivity to vary by year through the year-specific coefficients, λ_y. The equation also includes coefficients of year indicators, η_y, to capture any variation in the construct of the severity of the famine not explained by variation in the previous year's thermal agricultural productivity.[8]

Table 2 reports and Figure 4 illustrates the estimates. The coefficient of the average previous year's thermal agricultural productivity, $\beta_1 = -0.044$ (s.e. 0.024), is negative and

[7] During the Great Leap Forward, government chief statistician, Xue Muqiao, advised his staff, "Whatever statistical data the party leadership needs, we will provide it, and whatever direction the political campaigns and production campaigns take, statistical work will follow" (Xue (1958) quoted in Yang (2012: 257)).

[8] The weather data cover 1951-70. Since equation (5) includes a lagged term, $\text{Agri}_{c,y-1}$, the first observation in the regression is for the year 1952.

marginally significant, suggesting that, on average, a good harvest in the previous year was associated with lower cohort loss in the subsequent year. This is consistent with part of surplus agricultural production (in excess of government procurement) accruing to people in the county.

Importantly, as Figure 4 clearly illustrates, the coefficient of the previous year's thermal agricultural productivity is positive and significantly large in the famine years. Indeed, the coefficient is an order of magnitude larger in 1960, at the peak of the famine. The estimates imply that, during the famine years, *higher* agricultural productivity in the previous year was associated with *larger* cohort loss in the subsequent year. These results are consistent with previous research into the proximate causes of the famine: Higher food production led to higher grain procurement and aggravated the famine in the following year (Meng et al., 2015).

5 Empirical Strategy

Our thrust is to exploit county-level variation in the severity of the Great Famine to identify the effect of hardship on entrepreneurship. The proximate cause of the famine was official exaggeration of production, and unwillingness to revise procurement or redistribute food to mitigate shortfalls in production (Meng et al., 2015). The severity of the famine differed geographically according to the degree of career concerns of Communist Party officials (Kung and Chen, 2011). Differences in officials' career concerns during the famine period, however, are unlikely to be related to entrepreneurship in the 2000s.

The major empirical challenge to our research is unobservable heterogeneity between counties. For instance, the natural resources in a county may correlate with both the severity of the Great Famine and entrepreneurship in the 2000s, generating a spurious relationship between hardship and entrepreneurship. Previous research deals with such heterogeneity by using post-famine birth cohorts as a control group for the pre- and within-famine cohort (Almond et al., 2010; Mu and Zhang, 2011; Tan et al., 2014). However, in our context, this strategy is problematic as engaging in entrepreneurship is subject to strong family and peer effects (Lindquist et al., 2015; Lerner and Malmendier, 2013).

Our empirical strategy addresses unobservable cross-county heterogeneity by constructing two instruments for the severity of the famine that *purge county fixed effects*. Specifically, we first reformulate equation (5) into two periods, $p = f$ (famine) and $p = n$ (normal),

$$\mu_{cp} = \beta_0 + \beta_1 \mathrm{Agri}_{cp} + \eta D_p + \lambda \mathrm{Agri}_{cp} D_p + \nu_c + \varepsilon_{cp}, \tag{6}$$

where μ_{cp} and Agri_{cp} are the average rate of cohort loss and thermal agricultural productivity in county c in period p, and ν_c is a county fixed effect.[9] The indicator, $D_p = 1$ for famine years, $y = 1959, 1960$, and 1961, and $D_p = 0$ otherwise. Hence, the coefficient, η, represents the average effect of the famine on cohort loss across all counties other than through temperature.

The coefficient, β_1, measures the productivity effect of higher temperatures on the rate of cohort loss during normal years, conditional on the county fixed effect. Owing to localized administration or barriers to trade, greater production of food in a county raises local food consumption in the following year. Accordingly, we expect $\beta_1 < 0$.[10]

The coefficient, λ, measures the institutional effect of higher temperatures on the rate of cohort loss during the famine period, conditional on the county fixed effect. Generally, higher temperatures affect cohort losses during the famine in two conflicting ways. One way is through productivity, by which higher temperatures raise food production, and so, reduce cohort losses, which is captured by $\beta_1 < 0$. The other way is through institutions during the famine period. Higher food production was associated with higher grain procurement, which coupled with rigidity in procurement, led to more severe famine and greater cohort losses (Meng et al., 2015). Accordingly, we expect the institutional effect, $\lambda > 0$.

To estimate the effect of hardship during the famine on entrepreneurship, we represent the severity of the famine by the relative cohort loss between famine and normal periods,

$$\Delta\hat{\mu}_c = \hat{\mu}_{cf} - \hat{\mu}_{cn}, \tag{7}$$

where $\hat{\mu}_{cf}$ and $\hat{\mu}_{cn}$ are the rates of cohort loss during the famine and normal periods as

[9]Specifically, $\mu_{cf} = \frac{1}{3}\sum_{y=1959}^{1961}\mu_{cy}$, $\mu_{cn} = \frac{1}{15}(\sum_{y=1953}^{1958}\mu_{cy} + \sum_{y=1962}^{1970}\mu_{cy})$, $\mathrm{Agri}_{cf} = \frac{1}{3}\sum_{y=1959}^{1961}\mathrm{Agri}_{c,y-1}$, and $\mathrm{Agri}_{cn} = \frac{1}{15}(\sum_{y=1953}^{1958}\mathrm{Agri}_{c,y-1} + \sum_{y=1962}^{1970}\mathrm{Agri}_{c,y-1})$.

[10]In principle, the negative coefficient, $\beta_1 < 0$, might also represent an income effect as greater production of food raises overall income. However, prior to the Great Famine, the Chinese government had forcibly collectivized agriculture, and so, rural people did not gain by producing more.

predicted by the estimated coefficients of (6). Specifically, in normal periods, $D_p = 0$, and

$$\hat{\mu}_{cn} = \hat{\beta}_0 + \hat{\beta}_1 \text{Agri}_{cn} + \hat{\nu}_c, \tag{8}$$

while, during the famine, $D_p = 1$, and

$$\hat{\mu}_{cf} = \hat{\beta}_0 + \hat{\beta}_1 \text{Agri}_{cf} + \hat{\lambda} \text{Agri}_{cf} + \hat{\nu}_c. \tag{9}$$

Substituting from (9) and (8) in (7), the relative cohort loss,

$$\Delta\hat{\mu}_c \equiv \hat{\mu}_{cf} - \hat{\mu}_{cn} = \hat{\eta} + \hat{\lambda} \text{Agri}_{cf} + \hat{\beta}_1 \Delta \text{Agri}_c, \tag{10}$$

where $\Delta \text{Agri}_c \equiv \text{Agri}_{cf} - \text{Agri}_{cn}$ is difference in thermal agricultural productivity between famine and normal periods.

Importantly, both (8) and (9) include the county fixed effect, $\hat{\nu}_c$. Hence Agri_{cn} and Agri_{cf} represent the *random* shocks of weather on thermal agricultural productivity during normal and famine periods, *purged of any non-time-varying county characteristic* such as average agricultural productivity. Our claim is supported by estimates, reported in Appendix Table A1, which show that, after controlling for station fixed effects, thermal agricultural productivity is not serially correlated. So, in (6), the coefficients, β_1 and λ_y, are consistently estimated.

We then estimate the effect of hardship experienced during the famine on entrepreneurship, using the relative cohort loss to measure hardship,

$$Y_c = \omega_1 + \omega_2 \Delta\hat{\mu}_c + \upsilon_c, \tag{11}$$

where Y_c measures entrepreneurship and υ_c is an error term. Referring to (10), by construction, the relative cohort loss, $\Delta\hat{\mu}_c$, excludes the county fixed effect and the variation in relative cohort loss is due solely to random temperature shocks. Accordingly, the estimate of ω_2 renders a causal interpretation.

Our estimation procedure is essentially a two-stage least squares regression. The first stage, (10), estimates the relative cohort loss, $\Delta\mu_c$ as a function of two instruments – thermal agricultural productivity during the famine period, Agri_{cf}, and difference in thermal agricultural productivity between famine and normal periods, ΔAgri_c. The second stage, (11), estimates entrepreneurship as a function of the fitted value of the relative cohort loss rate, $\Delta\hat{\mu}_c$.

To check whether our empirical strategy successfully deals with cross-county heterogeneity, we also estimate (11) including county agricultural output in the early 1990s as an additional covariate. As Appendix Table A4 reports, this additional variable hardly affects the estimate. Moreover, our results are also robust to the inclusion of an extensive array of county-level socioeconomic controls, including ethnicity, gender, age, urbanization, and education. Such robustness is expected as, by construction, our identification strategy extracts the variation in cohort losses induced by weather shocks and abstracts from county fixed effects.

Referring to Figure 1, the severity of the famine varied considerably across counties. The mean of the relative cohort loss rate is 0.363 (s.d. 0.211) with a minimum of -0.716 and a maximum of 1.039. These are statistics for all counties, and so, differ slightly from Tables 1(a) and (b). The famine was most severe in the provinces of Sichuan, Anhui, Hunan, Guizhou, and Qinghai, and least severe in Heilongjiang, Inner Mongolia, and Tibet. There was also substantial variation within provinces. For instance, in Sichuan, the mean of the relative cohort loss rate is 0.592 (s.d. 0.151) with a minimum of -0.037 and a maximum of 0.789.

Table 3 reports estimates of the first stage regression, based on the 1990 Census, the 2000 Census, and the combination of both censuses. Consistent with our prediction, the institutional effect, λ, is positive, while the productivity effect, β_1, is negative. In the analysis below, we focus on the estimates based on the 1990 Census, and use the 2000 Census and combined censuses in robustness checks. Referring to Table 3, the joint F-statistic of the two instruments exceeds 21 across all specifications, indicating that the instruments satisfy the relevance condition. The F-statistic is 16.45 when accounting for spatial correlation in errors.

It is important to emphasize that excluding either of the two instruments from the first stage regression would result in specification error. Statistically, Agri_{cf} and ΔAgri_c are correlated because, by definition, $\Delta\text{Agri}_c = \text{Agri}_{cf} - \text{Agri}_{cn}$. Hence, using only one as an instrument would violate the exclusion restriction. Conceptually, referring to (10), during the famine, agricultural productivity, Agri_{cf}, affects the relative cohort loss rate, $\Delta\mu_c$, in two conflicting ways. Agricultural productivity, as represented by Agri_{cf}, reduces relative cohort loss. By contrast, institutions, as represented by ΔAgri_{cf} after controlling for Agri_{cf}, increase relative cohort loss. Thus, if one instrument is omitted, the relation between the other instrument and $\Delta\mu_c$ would be non-monotone, which violates the monotonicity assumption of instrumental variable estimation (Imbens and Angrist, 1994). Our empirical strategy distinguishes these two effects. Since the two instruments are correlated, the relevant diagnostic for the first-stage estimation is the joint F-statistic, rather than the individual t-statistics.

We cluster standard errors by weather station when measuring entrepreneurship by the 2004 Economic Census or 2005 1% Population Census. This procedure is conservative and accounts for possible correlation among counties due to multiple counties being associated with the same weather station. To investigate the mechanisms for the positive relation between famine severity and entrepreneurship, we use data from the various surveys. Then, we cluster standard errors by county because the counties in the surveys are not contiguous and are associated with different weather stations. In robustness checks, we account for spatial correlation in the errors, and also cluster standard errors by province and prefecture.

6 Estimates

Figure 5 illustrates our main result using the 2004 Economic Census. Counties that suffered more during the Great Famine were more entrepreneurial in subsequent years. The figure depicts the entrepreneurial share of all enterprises - the ratio of the number of private enterprises less than 10 years old and with fewer than 100 employees to the number of all enterprises - against the severity of the famine.

Following up, Table 4, column (a), presents an OLS estimate of equation (11) with the dependent variable being the entrepreneurial share of all enterprises. The coefficient of the relative cohort loss rate, which represents the severity of the famine, is positive and statistically significant.

Table 4, column (b), reports the first-stage regression of the relative cohort loss rate on the instruments. This estimate is similar to that reported in Table 3, column (a). The instruments are not weak (F-statistic is 20.29). Then, Table 4, column (c), reports the second-stage IV estimate of equation (11) with the dependent variable being the entrepreneurial share of all enterprises. The coefficient of the relative cohort loss rate, 0.658 (s.e. 0.119), is positive and statistically significant. The estimated elasticity of 0.511 implies that, if the famine is more severe by one percent, then the proportion of entrepreneurial enterprises would be higher by 0.511 percent.

Importantly, the IV estimate of the coefficient of the relative cohort loss rate is substantially larger than the OLS estimate. The Hausman test statistic is 30.66 ($p < 0.001$). The significant difference and Hausman test confirm that the severity of the famine is endogenous to entrepreneurship, and so, biasing the OLS estimate downward. One possible reason is that counties which suffered more during the Great Famine were more strictly controlled by the government, and the control persisted to the 2000s, and so, stifled entrepreneurship. Our robustness checks support this explanation. In an OLS estimate that controls for government expenditures (reported in Appendix Table A4), the coefficient of the relative cohort loss rate is positive, albeit not significant. Another potential reason is classical measurement error, which results in the estimated coefficients being attenuated.

Table 4, column (d), reports an IV estimate of equation (11) with the dependent variable being the entrepreneurial share of enterprise sales. The coefficient of the relative cohort loss rate, 0.264 (s.e. 0.086), is positive and statistically significant. Next, Table 4, column (e), reports an IV estimate of (11) with the dependent variable being the entrepreneurial share of enterprise profits. The coefficient of the relative cohort loss rate, 0.734 (s.e. 0.147), is positive and statistically significant.

Table 4, column (f), reports an IV estimate of (11) with the dependent variable being the median age of private enterprises. The coefficient of the relative cohort loss rate, −31.763 (s.e. 12.047), is negative and statistically significant. Finally, Table 4, column (g), reports an IV estimate of (11) with the dependent variable being the entrepreneurial share of employment. The coefficient of the relative cohort loss rate, 0.300 (s.e. 0.075), is positive and statistically significant. In four of the five estimates (columns (c) and (e)-(g)), the Hansen J-statistic is insignificant, which validates our use of the two instruments.

The preceding analysis investigates the effect of hardship on aggregate, county-level entrepreneurship. To probe more deeply, we draw on the 2005 1% Population Census to investigate at the individual level, limiting the sample to persons who were born before 1962. Using the Census record of individual employment, we define an individual to be an entrepreneur in two ways: if he/she reported being the owner of a private enterprise or self-employed (broad definition) or only if an owner (narrow definition). Then, we estimate linear probability models of equation (11), weighted by the Census weights, where the dependent variable is an indicator of being an entrepreneur.

Table 5, column (a), reports the OLS estimate of equation (11). The coefficient of the relative cohort loss rate, which represents the severity of the famine, is not statistically significant. Table 5, column (b), reports the first-stage regression of the relative cohort loss rate on the instruments. This estimate is similar to that reported in Table 3, column (a), with the coefficient being larger (0.094 as compared with 0.075) due to the difference in sample. The analysis of entrepreneurship at the individual level encompasses 2,212 counties, owing to the omission of counties that changed boundaries between 1990 and 2005. The instruments are not weak (F-statistic is 17.87).

Then, Table 5, column (c), reports the second-stage IV estimate of entrepreneurship. The coefficient of the relative cohort loss rate, 0.088 (s.e. 0.021), is positive and statistically significant. The estimated elasticity of 0.917 implies that, if the famine is more severe by one percent, then the probability of the individual being an entrepreneur would be higher by 0.917 percent.

Consistent with the aggregate analysis (Table 4), the IV estimated coefficient of the relative cohort loss rate is substantially larger than the OLS estimate. This result affirms that the severity of the famine is endogenous to entrepreneurship, and the endogeneity biases the OLS estimate downward. As discussed above, the downward bias may arise from unobserved heterogeneity in institutions. In particular, counties that suffered more during the Great Famine were those under stricter government control, which persisted and stifled entrepreneurship in later years. Our robustness checks support this explanation. In an OLS estimate that controls for government expenditures (reported in Appendix Table A4), the coefficient of the relative cohort loss rate is positive, albeit not significant.

To check the robustness of the findings, Table 5, column (d), reports an IV estimate with entrepreneurship defined narrowly as owning a private enterprise. The results are similar to those using the broader definition of entrepreneurship, reported in Table 5, column (c). The coefficient of the relative cohort loss rate, 0.013 (s.e. 0.006), is positive and statistically significant.

The Great Famine affected females and males differently (Mu and Zhang, 2011). Further, men are generally more likely to engage in entrepreneurship than women (Schiller and Crewson, 1997). We therefore check whether the effect of famine on entrepreneurship differed between men and women. Table 5, columns (e) and (f), report the IV estimates for the two different definitions of entrepreneurship. The coefficient of the interaction term between the relative cohort loss rate and male is positive and statistically significant.

It is also important to appreciate the effect of hardship on the intensive as well as the extensive margin of entrepreneurship. Table 5, columns (c)-(f), present estimates of engagement in entrepreneurship, which is the extensive margin. To study the effect on the intensive margin, which is income from entrepreneurship, we use data from the 2005 1% Population Census on the individual's monthly income. We observe the income of only those who become entrepreneurs. Hence, an estimate of the effect of hardship on income from entrepreneurship must account for selection into entrepreneurship as well as possible endogeneity of the measure of hardship.

Heckman (1979) suggests that observed samples of wages depend on household structure. Here, in the same spirit, we instrument for selection into the entrepreneur sample by the number of generations in the individual's household. Table 5, column (g), reports the IV estimate, applying a Heckman correction for selection into entrepreneurship (either owning a private enterprise or self-employed), and, in the second stage, using just one instrument – the thermal agricultural productivity during the famine.[11] The coefficient of the relative cohort loss rate is positive, statistically significant, and economically large. The implied elasticity is 0.315, which suggests that, if the relative cohort loss rate is one percent higher, the entrepreneurs' income would be 0.315 percent higher. The estimate in Table 5, column (h), shows that the result is robust to the narrow definition of entrepreneurship.

[11] The IV estimate with two instruments – the thermal agricultural productivity during the famine and the difference in thermal agricultural productivity between famine and normal periods – is over-identified. Accordingly, we use only the former as an instrument and control for the latter.

An obvious concern with these estimates is under-reporting of incomes. To the extent that mismeasurement of the dependent variable is classical, our estimates of the effect of hardship on entrepreneur incomes are conservative. Hence, the concern is that under-reporting of incomes by entrepreneurs in the 2005 1% Census is correlated with the severity of the famine as instrumented by the thermal agricultural productivity during the famine, which we consider to be unlikely.

Prior research shows that famines impair the long-term physical health and cognitive skills of those who survive (Meng and Qian, 2009; Almond et al., 2010; Mu and Zhang, 2011; Tan et al., 2014). To the extent that entrepreneurship depends on physical health and cognitive skills, survivors of the Great Famine would be less likely to engage in entrepreneurship. Accordingly, our estimates of the effect of hardship on entrepreneurship and income are conservative.

As reported in Appendix Table A4, our results are robust to various differences in measurement, specification, and estimation method: (i) excluding 1963 from the years used to project the counterfactual population, (ii) projecting the counterfactual population by a quadratic or exponential growth model, (iii) measuring the intensity of the famine by the gender ratio rather than cohort losses, (iv) measuring thermal agricultural productivity by a four-piece linear formula (Ritchie and NeSmith, 1991), (v) including controls for government expenditure, agricultural production, and individual demographics (ethnicity, gender, age, urbanization, and education),[12] (vi) clustering standard errors by province, prefecture, or county, (vii) accounting for spatial correlation in the errors, and (viii) estimating by probit regression rather than the linear probability model.

6.1 Heterogeneous Effects

We have found a positive relation between hardship during the Great Famine and engagement in entrepreneurship in later years. To better appreciate the policy implications of this relation, we next investigate two possible dimensions of heterogeneity. One is differences by the age at which the individual experienced the hardship, and the other is differences by the degree

[12] Gender, urbanization, and education are possibly endogenous, and so, the estimates including these variables should be interpreted with caution.

of the hardship.

To investigate possible differences in the effect of the famine by age, Table 6 presents IV estimates by three-year birth cohort. Figure 6 depicts the estimated elasticities with respect to the relative cohort loss rate and the corresponding 95 percent confidence intervals. The elasticity of entrepreneurship is large and statistically significant across all cohorts, with the effect larger among those who were 10 years or older at the beginning of the famine (born before 1950).

Note that the observed effect is the net outcome of negative and positive effects of hardship on entrepreneurship. We also caution that the apparent differences in Table 6 and Figure 6 should be interpreted with caution. Engagement in entrepreneurship might differ with age. Entrepreneurship is measured by the 2005 1% Population Census, in which the various birth cohorts differed in age. For example, in 2005, people born in the 1941-43 cohort were 62-64 years old, whereas those born in the 1959-61 cohort were 44-46 years old. So, the differences in Table 6 and Figure 6 might be due to differences in age at which the individuals were observed. Further, the differences are possibly confounded by the Cultural Revolution, which affected the education of people born between 1948 and 1961. Still, the estimated elasticities suggest that hardship increased entrepreneurship across all cohorts, and particularly among those aged 10 years or older at the onset of the famine.

The hardship imposed by the Great Famine ranged to fatal extremes. This presents the question of whether an analysis based on the famine generalizes to situations of less extreme hardship. We address this issue in two ways.

First, following the approach of Meng and Qian (2009), we use quantile regression to estimate the effects of different degrees of hardship. To sensibly apply quantile regression, the dependent variable must be continuous. Accordingly, we specify the unit of analysis as a county and the dependent variable as the percentage of the county population owning a private enterprise or self-employed. The instruments are thermal agricultural productivity in the famine period and the difference in thermal agricultural productivity between famine and normal periods. As Table 7, columns (a)-(c), report, the positive effect of hardship on entrepreneurship is not significantly different across the quartiles.

The second way to investigate the effect of less extreme hardship distinguishes rural and urban areas. The famine was substantially less severe in urban areas. The ratio of the death rate at the peak of the famine in 1960 to the pre-famine death rate in 1957 was 1.6 and 2.6 in urban and rural areas respectively (Lin and Yang 2000, Table 2). We use two-stage least squares method to estimate the effect of severity of the famine on entrepreneurship, just as in Table 5, column (c), but separately for rural and urban areas. As Table 7, columns (d) and (e), report, the effect of hardship is slightly smaller in urban areas, but the difference is not statistically significant (Hausman test statistic is 0.72 ($p = 0.395$)).

Overall, the analyses suggest that people who experienced more severe hardship during the Great Famine were subsequently more likely to engage in entrepreneurship. The effect was stronger among those aged 10 years or older, and did not differ by the severity of hardship.

6.2 Mechanism: Selective Mortality or Operant Conditioning?

How did greater hardship during the Great Famine induce more entrepreneurship? One possible mechanism is Darwinian – the famine selectively killed the fragile, leaving survivors who were more suited to the challenging environment. Another possibility is that the famine conditioned people to adapt their personality in ways that helped them cope and survive. The famine might have selectively killed and/or conditioned people on several dimensions of personality that support entrepreneurship – risk tolerance, confidence, tenacity, and resilience (Frese and Gielnik, 2014).

It is fairly intuitive that the famine may have selectively eliminated people and conditioned others on risk tolerance, confidence, tenacity, or resilience. During the Great Famine, strategies for survival include speculation and profiteering, dealing in the black market, and stealing food (Zhou 2012: Chapter 7). Villagers in Da Fo in northern Henan Province survived by eating raw or unripe crops in the fields, deliberately leaving crops in the fields during the communal harvest for later "gleaning", slacking during communal work, black-market dealing in salt earth and nitrates, and begging (Thaxton 2008: Chapters 5 and 6).

Table 6 and Figure 6 provide some indirect evidence on the mechanism by which hardship induced entrepreneurship – whether selective mortality or operant conditioning. Famines disproportionately kill infants and very old (Meng and Qian, 2009), and so, if the mechanism had been selective mortality, the relation between hardship and entrepreneurship should be stronger in the very young cohorts. This is contrary to Table 6 and Figure 6, which suggest that the relation was weaker among the very young.

To investigate the mechanism underlying the differences in entrepreneurship across counties at more depth, we construct and estimate a simple model. Suppose that the famine increased entrepreneurship only through selective mortality and not through any other mechanism. Denote the share of entrepreneurs in the population absent famine by e_c, the actual number of entrepreneurs by $\hat{E}_c$, and the share of entrepreneur deaths in cohort loss due to the famine by λ_c.

Accordingly, the actual number of entrepreneurs in county c,

$$\hat{E}_c = [e_c - \lambda_c \mu_c]\tilde{P}_c = [1 - \mu_c]\tilde{P}_c \cdot \frac{e_c - \lambda_c \mu_c}{1 - \mu_c}, \tag{12}$$

where, as defined around equation (1) above, μ_c is the rate of cohort loss, and $\tilde{P}_c$ the counterfactual projected population in the absence of famine. On the right-hand side of equation (12), $[1-\mu_c]\tilde{P}_c$ is the actual population as observed during the census, and $[e_c - \lambda_c \mu_c]/[1-\mu_c]$ is the share of entrepreneurs in the actual population.

Suppose that some entrepreneurs survived the famine, $e_c - \lambda_c \mu_c > 0$, so that $\hat{E}_c > 0$. Then, differentiating the logarithm of equation (12) with respect to μ_c, we have, conditional on $\tilde{P}_c$,

$$\frac{\partial \ln \hat{E}_c}{\partial \mu_c} = \frac{-\lambda_c}{e_c - \lambda_c \mu_c} \leq 0. \tag{13}$$

This states that, conditional on the counterfactual population absent the famine, the number of entrepreneurs should not increase with the severity of the famine.

Now, consider estimating the model,

$$\ln \hat{E}_c = \gamma_0 + \gamma_1 \mu_c + \gamma_2 \ln \tilde{P}_c + \epsilon_c, \tag{14}$$

where ϵ_c is an error term. By equation (13), if the famine increased entrepreneurship only through selective mortality, then $\gamma_1 \leq 0$. However, if, empirically, we find $\gamma_1 > 0$, then some other mechanism must have raised entrepreneurship so much as to outweigh the selective deaths.

The challenges in estimating equation (14) are that the rate of cohort loss, μ_c, and the counterfactual population absent the famine, $\tilde{P}_c$, are possibly endogenous. Accordingly, we operationalize μ_c by the relative cohort loss rate, $\Delta\mu_c$, and estimate equation (14) using two instruments – the thermal agricultural productivity during the famine period and the difference in thermal agricultural productivity between famine and normal periods – for the two endogenous variables, μ_c and $\ln \tilde{P}_c$.

As discussed above, famines disproportionately kill the young, but people are more easily conditioned at an early age. Hence, we are particularly interested in differences among cohorts in the extent to which the increase in entrepreneurship was not due to selective mortality. To broaden the analysis to include people who were teenagers during the famine, we extend the projection of cohort losses (equation (2) above) to include the years 1947 and 1948.

Table 8, columns (a)-(e), report IV estimates of equation (14) by three-year birth cohorts, specifically, 1947-49, 1950-52, 1953-55, 1956-58, and 1959-61, with entrepreneurs defined as owners of private enterprises or self-employed. To ensure comparability across estimates, the sample is limited to counties with a positive number of entrepreneurs in all cohorts.[13]

The coefficient of the relative cohort loss rate is positive and statistically significant in the two older cohorts (1947-49 and 1950-52), marginally significant in the 1953-55 cohort, but imprecisely estimated in the two younger cohorts (1956-58 and 1959-61). The relation between hardship and entrepreneurship is strongest, with an elasticity of 1.059, in the oldest cohort (1947-49), or those who were 10-12 years old at the onset of the famine. The relation is monotone decreasing with youth.

Referring to equation (14), the reduced-form estimate of γ_1 captures both selective mortality and operant conditioning. By equation (13), the selective mortality effect is negative. Hence, the positive estimates of γ_1 reported in Table 8 imply that the increase in entrepreneurship extended beyond selective mortality. In particular, the estimate of γ_1 provides a lower bound on the positive effect of hardship on entrepreneurship through operant conditioning.

Accordingly, the pattern of the estimates by birth cohort in Table 8 suggests significant operant conditioning among individuals who were between seven and twelve years old at the

[13] If $\hat{E}_c = 0$, the famine did not increase entrepreneurship, so it does not make sense to analyze whether the increase was due to selective mortality or conditioning.

onset of the famine (the 1947-49 and 1950-52 cohorts), with weaker evidence of conditioning among those aged four to six years (1953-55 cohort). Young children may have observed parents, siblings, other relatives and neighbors engaging in risky actions, such as taking raw crops from fields and going out at night to glean food. Older children might have directly participated in gathering food. Children learned behavior and developed personal traits that increased their chances of survival and stimulated entrepreneurship in later life.

It might seem interesting to compare the patterns of the estimates by birth cohort in Tables 6 and 8. However, the two sets of estimates are not directly comparable. Table 6 reports individual level analyses on a larger set of counties, while Table 8 reports county level analyses, controlling for cohort size, on a subset of the counties. Subject to that proviso, comparing the patterns by birth cohort in Tables 6 with 8, we infer that operant conditioning was stronger among the older children, while selective mortality was stronger among the very young and particularly those born during the famine.

In interpreting the differences between the cohorts, we caution that the differences across birth cohorts are possibly contaminated by age or cohort effects. Further, Table 8 presents the net outcome of two possibly conflicting effects. On the one hand, we argue that famine conditions personality traits that increase entrepreneurship. On the other hand, famine diminishes cognitive skills (Meng and Qian, 2009), which would reduce entrepreneurship. Accordingly, the estimates in Table 8 are biased downward as to the effect of operant conditioning of personality traits.

Development of both cognitive and non-cognitive skills (personality traits) is strongest in early childhood (Caspi and Silva, 1995; Heckman, 2007; Heckman et al., 2013). The empirical pattern of weaker effects on entrepreneurship in later cohorts is consistent with the famine having affected cognitive development relatively more than non-cognitive development among young children.

6.3 Conditioning and Personality Adaptation

As reviewed above, previous research points to risk tolerance, confidence (self-efficacy), tenacity, and resilience as being important ingredients in entrepreneurship. Through which of these personality traits did the Great Famine stimulate entrepreneurship? We investigate using individual data from the various surveys.

A conventional measure of risk tolerance is investment in shares (equities). Using data on 400,000 Norwegian individuals, Hvide and Panos (2014) validate investment in shares as a measure of risk tolerance and show that it is correlated with starting a new business. The CFPS 2010, CHFS 2011, and CGSS 2010, 2012, and 2013 include questions on investments. We combine responses to these surveys, limiting the sample to individuals born before 1962. Table 9, column (a), reports the first-stage regression of the relative cohort loss rate on the thermal agricultural productivity during the famine and the difference in thermal agricultural productivity between famine and normal periods. The instruments are not weak (F statistic is 13.96).

Table 9, column (b), reports the IV estimate of a linear probability model with the dependent variable being an indicator of investment in the share market. The coefficient of the relative cohort loss rate, 0.091 (s.e. 0.045), is positive and statistically significant. The result indicates that people who suffered more during the famine are more likely to invest in shares. This result is buttressed by the IV estimate of financial investment, reported in Table 9, column (c), in which the coefficient of the relative cohort loss rate, 0.127 (s.e. 0.062), is positive and significant. These results are consistent with the famine conditioning people to be more risk tolerant.[14]

Further, Table 9, column (d), reports an IV estimate of ownership of a private enterprise based on these survey data. The coefficient of the relative cohort loss rate is positive and statistically significant. People who suffered more during the famine are also more likely to own a private enterprise. This estimate provides further evidence in support of our main finding of an effect of hardship on entrepreneurship. In addition, these results tend to rule out the relation between hardship and entrepreneurship being due to conditioning of tenacity and resilience. Tenacity and resilience cannot account for the effect of hardship on investments in shares and other financial investments.[15]

[14]It might seem intuitive to control for income and wealth in these estimates. However, the famine caused people to adapt towards more risk-taking, and thus, it would have raised their incomes and wealth. Accordingly, income and wealth would be bad controls in a regression of risk tolerance on the severity of the famine.

[15]Although investment is a binary variable, we use linear regression to facilitate estimation and

Besides risk tolerance, the personality traits of confidence, tenacity, and resilience also contribute to entrepreneurship. However, as reported in the Appendix, we find no empirical evidence of any relation between severity of the famine and these other traits. We also investigate and find no evidence that the famine increased craftiness or opportunism, which behavior might have contributed to entrepreneurship.

7 Discussion

Exploiting geographical variation in the severity of China's Great Famine, we find robust evidence that sustained hardship is associated with more entrepreneurship, across different degrees of hardship. The effect is partly due selectively mortality. However, among those aged between seven and twelve at the time, the relation between hardship and entrepreneurship is at least partly the result of operant conditioning. We find evidence consistent with conditioning of risk tolerance.

We should qualify that our analysis is cross-sectional, not a life-cycle study. Even though we analyze the population by birth cohorts separately, the data do not allow us to follow the same persons over time to investigate how the famine affected their development at various stages of life. Our findings only address the long term effect of the famine on conditioning of personality and entrepreneurship.

Further, our analysis is not experimental and uses observational data, and so, the empirical relations are necessarily open to alternative explanations. In the Appendix, we consider several alternative explanations: (i) that the famine increased the proportion of females, and so, increased pressure on their parents to earn more through entrepreneurship and secure husbands for their children (Wei and Zhang, 2011); (ii) that the famine increased the number of couples without sons to support them in their old age, and so, induced them to earn more through entrepreneurship; and (iii) the government provided more help to areas that suffered relatively more during the famine, which fostered more entrepreneurship in later years. Our results are inconsistent with these alternative explanations.

---

interpretation. As shown in Appendix Table A6, the findings are robust to estimation by a probit model.

Another limitation is that the famine imposed hardship across the board rather than on particular individuals. Everyone in the county suffered the same deprivation, except to the extent that they could dodge the authorities and get more food. Our research context does not allow us to distinguish the direct personal effects of hardship from peer effects due to the hardship borne by family members and friends.

A further limitation is that we define entrepreneurship either in the aggregate as the share of small, young businesses or individually as owning a private enterprise or being self-employed. We do not claim that our analysis necessarily explains the development of high-growth businesses. On average, individual entrepreneurs in our dataset earned 11,659 Yuan (US$1424 at the 2005 exchange rate) annually, which is twice the average, 5,238 Yuan, of others. Our findings are directly relevant to the context of economic development (De Mel et al., 2008). Nevertheless, it is worth noting that Ren Zhengfei (Ren, 2001), founder of the telecommunications giant, Huawei, cited hardship during his formative years as being key to character and entrepreneurship.

Our research does contribute to discussions on whether entrepreneurs are born or nurtured (Lindquist et al., 2015), and more particularly, whether entrepreneurs can be trained. We show that people can be conditioned to become more entrepreneurial. In particular, we show that hardship, a factor much emphasized by institutions offering "boot camp" training, cultivates both more entrepreneurship and greater risk tolerance. This finding supports experiential learning – where students learn by facing and overcoming challenges in unstructured situations.

Although we find that hardship conditioned individuals aged between seven and twelve to become more entrepreneurial, our research design is limited in drawing inferences about the conditioning at earlier ages (six years old or younger). The effect for the younger children is also positive, albeit imprecise. The challenge with regard to younger children is that we only observe the net effect of the famine on entrepreneurship through cognitive and non-cognitive development. Research on child development (Heckman et al., 2013) shows that development of both cognitive and non-cognitive skills is fastest at early ages. Our inconclusive findings on the effect of hardship on entrepreneurship at early ages may be due to the conflict between negative effects on cognitive skills and positive effects on non-cognitive skills. An important direction for future research is to distinguish these two effects of hardship on entrepreneurship.

Famine kills people. Those who survive may be left with severely impaired health and cognitive skills (Meng and Qian, 2009; Almond et al., 2010; Mu and Zhang, 2011; Tan et al., 2014). To the extent that people who engage in business need better health and cognitive skills than wage earners, the famine would have diminished entrepreneurship. Accordingly, our empirical analysis under-estimates the positive effect of hardship on entrepreneurship.

References

Ahmetoglu, Gorkan, Tomas Chamorro-Premuzic, Bailey Klinger, and Tessa Karcisky, *The Wiley Handbook of Entrepreneurship*, John Wiley & Sons, Hoboken, NJ, 2017.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz, "Personality Psychology and Economics," in Eric A. Hanushek, Stephen Machin, and Ludger Woessman, eds., *Handbook of the Economics of Education*, Vol. 4, Elsevier, Amsterdam, 2011, pp. 1–181.

Almond, Douglas, Lena Edlund, Hongbin Li, and Junsen Zhang, "Long-term effects of early-life development: Evidence from the 1959 to 1961 china famine," in Takatoshi Ito and Andrew K. Rose, eds., *The Economic Consequences of Demographic Change in East Asia*, University of Chicago Press, Chicago, 2010, pp. 321–345.

Bernile, Gennaro, Vineet Bhagwat, and P Raghavendra Rau, "What Doesn't Kill You Will Only Make You More Risk-Loving: Early-Life Disasters and CEO Behavior," *Journal of Finance*, February 2017, *72* (1), 167–206.

Cameron, Lisa and Manisha Shah, "Risk-taking behavior in the wake of natural disasters," *Journal of Human Resources*, 2015, *50* (2), 484–515.

Caspi, Avshalom and Phil A. Silva, "Temperamental Qualities at Age Three Predict Personality Traits in Young Adulthood: Longitudinal Evidence from a Birth Cohort," *Child Development*, 1995, *66* (2), 486–498.

Chen, Heng and Maelys de la Rupelle, "The Great Famine and Savings Rate in China," University of Hong Kong 2016.

Deschenes, Olivier and Michael Greenstone, "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather," *American Economic Review*, 2007, *97* (1), 354–385.

Frese, Michael and Michael M Gielnik, "The Psychology of Entrepreneurship," *Annual Review of Organizational Psychology and Organizational Behavior*, 2014, *1*, 413–438.

Guzman, Jorge and Scott Stern, "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014," NBER Working Paper 22095 March 2016.

Hanaoka, Chie, Hitoshi Shigeoka, and Yasutora Watanabe, "Do Risk Preferences Change? Evidence from the Great East Japan Earthquake," *American Economic Journal: Applied Economics*, 2018, *10* (2), 298–330.

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

_ , "The Economics, Technology, and Neuroscience of Human Capability Formation," *Proceedings of the National Academy of Sciences of the United States of America*, 2007, *104* (33), 13250–13255.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, October 2013, *103* (6), 2052–86.

Hvide, Hans K. and Georgios A. Panos, "Risk Tolerance and Entrepreneurship," *Journal of Financial Economics*, 2014, *111* (1), 200–223.

Imbens, Guido W and Joshua D Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, March 1994, *62* (2), 467–475.

James, William H., "Proximate Causes of the Variation of the Human Sex Ratio at Birth," *Early Human Development*, 2015, *91* (12), 795–799.

Kasahara, Hiroyuki and Bingjing Li, "The Causes of China's Great Famine, 1959-1961: County-Level Evidence," CESifo Working Paper 6790 February 2018.

Kung, James Kai-sing and Shuo Chen, "The Tragedy of the Nomenklatura: Career Incentives and Political Radicalism during China's Great Leap Famine," *American Political Science Review*, 2011, *105* (1), 27–45.

Lazear, Edward P, "Balanced Skills and Entrepreneurship," *American Economic Review*, 2004, *94* (2), 208–211.

_ , "Entrepreneurship," *Journal of Labor Economics*, 2005, *23* (4), 649–680.

Lerner, Josh and Ulrike Malmendier, "With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship," *Review of Financial Studies*, 2013, *26* (10), 2411–2452.

Liang, James, Hui Wang, and Edward P Lazear, "Demographics and Entrepreneurship," *Journal of Political Economy*, 2018, *126* (S1), S140–S196.

Lindquist, Matthew J., Joeri Sol, and Mirjam Van Praag, "Why do Entrepreneurial Parents Have Entrepreneurial Children?," *Journal of Labor Economics*, 2015, *33* (2), 269–296.

Malmendier, Ulrike, Geoffrey Tate, and Jon Yan, "Overconfidence and Early-life Experiences: The Effect of Managerial Traits on Corporate Financial Policies," *Journal of Finance*, 2011, *66* (5), 1687–1733.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.

Meng, Xin and Nancy Qian, "The Long Term Consequences of Famine on Survivors: Evidence from a Unique Natural Experiment Using China's Great Famine," NBER 14917 April 2009.

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

Mu, Ren and Xiaobo Zhang, "Why does the Great Chinese Famine Affect the Male and Female Survivors Differently? Mortality Selection versus Son Preference," *Economics & Human Biology*, 2011, *9* (1), 92–105.

Ren, Zhengfei, "My Parents," *Huawei News (in Chinese)*, 2001, *114*, 4.

Ritchie, J.T. and D.S. NeSmith, "Temperature and Crop Development," in John Hanks and J.T. Ritchie, eds., *Modeling Plant and Soil Systems*, Crop Science Society of America, Inc., 1991, chapter 8, pp. 5–29.

Schiller, Bradley R. and Philip E. Crewson, "Entrepreneurial Origins: A Longitudinal Inquiry," *Economic Inquiry*, 1997, *35* (3), 523–531.

Schlenker, Wolfram and Michael J. Roberts, "Nonlinear temperature effects indicate severe damages to US crop yields under climate change," *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.

Tan, Chih Ming, Zhibo Tan, and Xiaobo Zhang, "Sins of the Fathers: The Intergenerational Legacy of the 1959-61 Great Chinese Famine on Children's Cognitive Development," IFPRI 01351 2014.

Wei, Shangjin and Xiaobo Zhang, "Sex Ratios, Entrepreneurship, and Economic Growth in the People's Republic of China," NBER Working paper 16800 2011.

Yao, Shujie, "A Note on the Causal Factors of China's Famine in 1959–1961," *Journal of Political Economy*, 1999, *107* (6), 1365–1369.

Table 1. Summary statistics

VARIABLES	Unit	Obs.	Mean	Std. dev.
Panel (a): Economic Census 2004				
Share of entrepreneurial firms	Ratio	2,194	0.471	0.16
Share of entrepreneurial firms in sales	Ratio	2,194	0.178	0.14
Share of entrepreneurial firms in profits	Ratio	2,194	0.526	0.19
Median age of entrepreneurial firms	Month	2,194	37.708	12.289
Median age of private enterprises	Month	2,194	47.448	18.126
Share of entrepreneurial firms in employees	Ratio	2,194	0.205	0.125
Share of high school educated employees	Ratio	2,194	0.17	0.11
Relative cohort loss rate		2,194	0.366	0.21
Thermal agricultural productivity during famine	1,000 degrees	2,194	2.41	0.623
Difference in thermal agricultural productivity between famine and normal years	1,000 degrees	2,194	-0.009	0.057
Panel (b): Population Census 2005				
Owner or self-employed	Indicator	639,443	0.034	0.18
Owner of private enterprise	Indicator	639,443	0.007	0.081
Self-employed	Indicator	639,443	0.027	0.162
Males: Owner of private enterprise	Indicator	322,933	0.011	0.103
Males: Self-employed	Indicator	322,933	0.038	0.19
Females: Owner of private enterprise	Indicator	316,510	0.003	0.051
Females: Self-employed	Indicator	316,510	0.016	0.125
Annual income: Owner or self employed	Yuan	21,409	11,659	16,317
Annual income: Owner of private enterprise	Yuan	4,267	20,125	31,157
Annual income: Self-employed	Yuan	17,142	9,552	8,285
Relative cohort loss rate		639,443	0.349	0.219
Thermal agricultural productivity during famine	1,000 degrees	639,443	2.527	0.593
Difference in thermal agricultural productivity between famine and normal years	1,000 degrees	639,443	-0.008	0.054
Panel (c): Combined survey sample				
Owner of private enterprise	Indicator	21,965	0.021	0.143
Any financial investment	Indicator	21,965	0.043	0.203
Stock market participation	Indicator	21,965	0.027	0.161
Relative cohort loss rate		21,965	0.343	0.212

Notes: Panel (c), sample: Respondents to CFPS2010, CHFS2011, CGSS2010, CGSS2012, or CGSS2013 born before 1962.

Table 2. Thermal agricultural productivity and cohort loss rate

VARIABLES	(a) β_1	(b) λ_y	(c) η_y
Thermal agricultural productivity in previous year	-0.044* (0.024)		
Year 1953		0.015 (0.018)	-0.024 (0.042)
Year 1954		-0.007 (0.020)	-0.026 (0.047)
Year 1955		0.029 (0.020)	-0.073 (0.049)
Year 1956		-0.002 (0.021)	0.088* (0.050)
Year 1957		-0.022 (0.021)	0.083 (0.050)
Year 1958		0.014 (0.021)	0.127** (0.050)
Year 1959		0.044** (0.021)	0.250*** (0.051)
Year 1960		0.127*** (0.022)	0.037 (0.054)
Year 1961		0.040** (0.020)	0.374*** (0.047)
Year 1962		-0.011 (0.020)	0.036 (0.048)
Year 1963		-0.037* (0.020)	-0.100** (0.049)
Year 1964		0.015 (0.018)	-0.030 (0.044)
Year 1965		-0.008 (0.018)	0.046 (0.043)
Year 1966		-0.004 (0.018)	0.049 (0.042)
Year 1967		-0.000 (0.017)	0.164*** (0.041)
Year 1968		0.004 (0.017)	-0.028 (0.041)
Year 1969		0.016 (0.018)	0.054 (0.042)
Year 1970		0.013 (0.017)	0.028 (0.041)

Notes: This table reports the OLS estimate of equation (5); Dependent variable: Cohort loss rate based on the 1990 census; with county fixed effects; 32,858 observations of 2,265 counties, covering years 1952-1970; R-squared: 0.368; F statistic: 31.81; Column (a): Coefficient of previous year's thermal agricultural productivity, β_1; Column (b): Coefficient of previous year's thermal agricultural productivity interacted with year, λ_y; Column (c): Coefficient of year indicator, η_y; Robust standard errors clustered by county in parentheses (* $p<0.1$; ** $p<0.05$; *** $p<0.01$).

Table 3. Relative cohort loss rate and thermal agricultural productivity

VARIABLES	(a) Based on Census 1990	(b) Based on Census 2000	(c) Based on Censuses 1990 and 2000	(d) Based on Census 1990: Spatial correlation in errors
Thermal agricultural productivity during famine, λ	0.075***	0.077***	0.079***	0.032***
	(0.013)	(0.013)	(0.013)	(0.009)
Difference in thermal agricultural productivity between famine and normal periods, β_1	-0.191	-0.192	-0.206	-0.105
	(0.147)	(0.140)	(0.140)	(0.081)
Observations	2,265	2,589	2,200	2,265
R-squared	0.058	0.062	0.069	
F statistic	21.71	21.89	22.02	16.45
Prob > F	< 0.001	< 0.001	< 0.001	<0.001

Notes: Sample: All counties; Estimated by ordinary least squares; Dependent variable: Relative cohort loss rate; Robust standard errors clustered by station in parentheses (* p<0.1; ** p<0.05; *** p<0.01). Column (a): Relative cohort loss rate estimated using the 1990 Population Census; Column (b): Relative cohort loss rate estimated using the 2000 Population Census; Column (c): Relative cohort loss rate estimated using the 1990 and 2000 Censuses; Column (d): Relative cohort loss rate estimated using the 1990 Population Census, allowing for spatial correlation in errors (estimated by Stata routine, spreg).

Table 4. Famine severity and entrepreneurship: County analysis

VARIABLES	(a) OLS: Entrepreneurial share of enterprises	(b) First stage	(c) IV: Entrepreneurial share of enterprises	(d) IV: Entrepreneurial share of enterprise sales	(e) IV: Entrepreneurial share of enterprise profits	(f) IV: Median age of private enterprises	(g) IV: Entrepreneurial share of employees
Relative cohort loss rate	0.161***		0.658***	0.264***	0.734***	-31.763***	0.300***
	(0.021)		(0.119)	(0.086)	(0.147)	(12.047)	(0.075)
Thermal agricultural productivity during famine		0.075*** (0.013)					
Difference in thermal agricultural productivity (famine – normal)		-0.191 (0.149)					
Counties	2,194	2,194	2,194	2,194	2,194	2,194	2,194
R-squared	0.04	0.06					
Hansen J statistic	.	.	0.56	10.85	0.39	1.96	1.08
p-value	.	.	0.45	0.00	0.53	0.16	0.30
Elasticity	0.125		0.511	0.544	0.511	-0.245	0.534
p-value	<0.001		<0.001	0.002	<0.001	0.008	<0.001

Notes: Sample: Enterprises covered by the 2004 Economic Census; Across columns (c) – (g), instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Columns (a) and (c): Dependent variable is entrepreneurial firms' share in all enterprises; Column (d): Dependent variable is entrepreneurial firms' share of all enterprise sales; Column (e): Dependent variable is entrepreneurial firms' share of all enterprise profits; Column (f): Dependent variable is median age in month of private enterprises; Column (g): Dependent variable is entrepreneurial firms' share of enterprise employment. Columns (c)-(g): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$), and first-stage F statistic = 20.29. Robust standard errors in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table 5. Famine severity and entrepreneurship: Individual analysis

VARIABLES	(a) OLS: Owner or self-employed	(b) First Stage	(c) IV: Owner or self-employed	(d) IV: Owner	(e) IV: Owner or self-employed	(f) IV: Owner	(g) Selection & IV: Income of owner or self-employed	(h) Selection & IV: Income of owner
Relative cohort loss rate	-0.003		0.088***	0.013**	0.046***	-0.001	0.887**	2.328*
	(0.005)		(0.021)	(0.006)	(0.014)	(0.002)	(0.399)	(1.399)
Thermal agricultural productivity during famine		0.094*** (0.017)						
Difference in thermal agricultural productivity (famine – normal)		-0.180 (0.169)					0.762* (0.431)	2.005** (0.919)
Male					-0.000	-0.002		
					(0.007)	(0.003)		
Relative cohort loss rate x male					0.081*** (0.020)	0.027*** (0.009)		
Observations	639,443	639,443	639,443	639,443	639,443	639,443	21,409	4,267
Counties	2,212	2,212	2,212	2,212	2,212	2,212	2,026	1,309
R-squared	0.000	0.060						
Hansen J statistic			1.300	1.890	6.966	2.148		
p-value			0.254	0.169	0.0307	0.342		
Elasticity			0.917	0.132			0.315	0.744
p-value			<0.001	0.021			0.026	0.096

Notes: Sample: Local residents born before 1962 covered by the 2005 Population Census; Columns (c)-(f) present estimates by two-stage least squares weighted by population, with instruments being thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Columns (g)-(h) present estimates by Heckman selection with number of generations in household as instrument for selection and two-stage least squares with thermal agricultural productivity in famine period as instrument and difference in thermal agricultural productivity between famine and normal years as control. Columns (a), (c), and (e): Dependent variable is indicator of owning private enterprise or self-employed; Columns (d) and (f): Dependent variable is indicator of owning private enterprise; Column (g): Sample limited to owners of private enterprises or self-employed, dependent variable is logarithm of monthly income; Column (h): Sample limited to owners of private enterprises, dependent variable is logarithm of monthly income; Columns (c)-(d): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$), and first-stage F statistic = 17.87. Robust standard errors clustered by station in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table 6. Famine severity and individual entrepreneurship: IV estimates by birth cohort

VARIABLES	(a) Before 1940	(b) 1941-43	(c) 1944-46	(d) 1947-49	(e) 1950-52	(f) 1953-55	(g) 1956-58	(h) 1959-61
Relative cohort loss rate	0.026***	0.058***	0.097***	0.121***	0.116***	0.160***	0.149***	0.236***
	(0.006)	(0.015)	(0.024)	(0.027)	(0.029)	(0.035)	(0.038)	(0.051)
Observations	174,440	42,583	49,555	60,936	73,683	87,886	86,651	63,709
Counties	2,207	2,192	2,193	2,207	2,207	2,208	2,208	2,203
F statistic	15.09	19.65	17.36	20.06	23.12	19.76	19.17	17.33
Hansen J statistic	0.002	0.944	0.279	2.047	0.927	1.112	2.143	1.280
p-value	0.962	0.331	0.597	0.153	0.336	0.292	0.143	0.258
Elasticity	1.577	1.347	1.608	1.559	1.053	1.142	0.822	1.021
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Notes: Sample: Local residents born before 1962 covered by the 2005 Population Census; Dependent variable is indicator of owning private enterprise or self-employed; Estimated by two-stage least squares weighted by population; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Elasticity is calculated as the coefficient of relative cohort loss rate multiplied by the sample mean of relative cohort loss rate and divided by the sample mean of the dependent variable; Columns (c)-(d): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$); F statistic is for first stage. Robust standard errors clustered by station in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table 7. Heterogeneous effects by famine severity

Variables	(a) 25th percentile	(b) 50th percentile	(c) 75th percentile	(d) Rural region	(e) Urban region
Relative cohort	0.093***	0.084***	0.127***	0.060***	0.110***
loss rate	(0.020)	(0.018)	(0.023)	(0.015)	(0.038)
Observations	2,212	2,212	2,212	349,331	290,112
Counties	2,212	2,212	2,212	1,404	808
Hansen J statistic	.	.	.	2.471	6.003
p-value	.	.	.	0.116	0.014
Elasticity	1.536	1.160	1.492	0.862	0.766
p-value	<0.001	<0.001	<0.00	<0.001	0.004

Notes: Sample: Local residents born before 1962 covered by 2005 Population Census. Columns (a)-(c): Estimated by quantile regression using Stata routine, ivqreg, for 25th, 50th, and 75th percentiles of famine severity; Dependent variable is county-level percentage of population owning private enterprise or self-employed; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Columns (d) and (e): Estimated by two-stage least squares weighted by population; Dependent variable is indicator of owning private enterprise or self-employed; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Sample limited to rural (column (d)) and urban areas (column (e)); Robust standard errors clustered by weather station in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

Table 8. Operant conditioning: IV estimates by birth cohort

VARIABLES	(a) 1947-49	(b) 1950-52	(c) 1953-55	(d) 1956-58	(e) 1959-61
Relative cohort loss rate	2.993***	2.558**	2.784*	1.698	1.796
	(1.048)	(1.240)	(1.506)	(1.862)	(2.003)
Population (ln)	0.501	0.395	0.369	0.862	0.499
	(0.613)	(0.653)	(0.813)	(0.950)	(0.987)
Counties	477	477	477	477	477
Elasticity	1.059	0.905	0.985	0.601	0.635
p-value	0.004	0.039	0.065	0.362	0.370

Notes: Sample: Local residents born before 1962 covered by the 2005 Population Census; each column limited to counties with positive number of entrepreneurs (owners of private enterprises or self-employed) across all three-year birth cohorts between 1947-59 and 1959-61; Estimated by two-stage least squares weighted by population; Dependent variable is the logarithm of the number of entrepreneurs; Population is that of the respective birth cohort; Instruments are thermal agricultural productivity in famine period and the difference in thermal agricultural productivity between famine and normal periods; Elasticity is calculated as the coefficient of relative cohort loss rate multiplied by the mean of relative cohort loss rate; Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.05$), and F statistic = 12.87 and 4.16 for first stage regressions of relative cohort loss rate and logarithm of population respectively; Robust standard errors clustered by station in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table 9. Risk tolerance: Individual analysis

VARIABLES	(a) First stage	(b) IV: Share investment	(c) IV: Financial investment	(d) IV: Owner
Relative cohort loss rate		0.091**	0.127**	0.050**
		(0.045)	(0.062)	(0.025)
Thermal agricultural	0.123***			
productivity during famine	(0.024)			
Difference in thermal agricultural	0.069			
productivity (famine - normal)	(0.260)			
Survey fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes	Yes
Observations	21,965	21,965	21,965	21,965
R-squared	0.090			
Counties	248	248	248	248
Hansen J statistic		0.292	0.489	1.226
p-value		0.589	0.484	0.268

Note: Sample: Respondents to CFPS 2010, CHFS 2011, CGSS 2010, CGSS 2012, or CGSS 2013 born before 1962; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Column (b): Dependent variable is indicator of investment in share market; Column (c): Dependent variable is indicator of financial investment; Column (d): Dependent variable is indicator of owning private enterprise; Columns (b)-(d): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$), and first-stage F statistic = 13.96.; Robust standard errors clustered by county in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

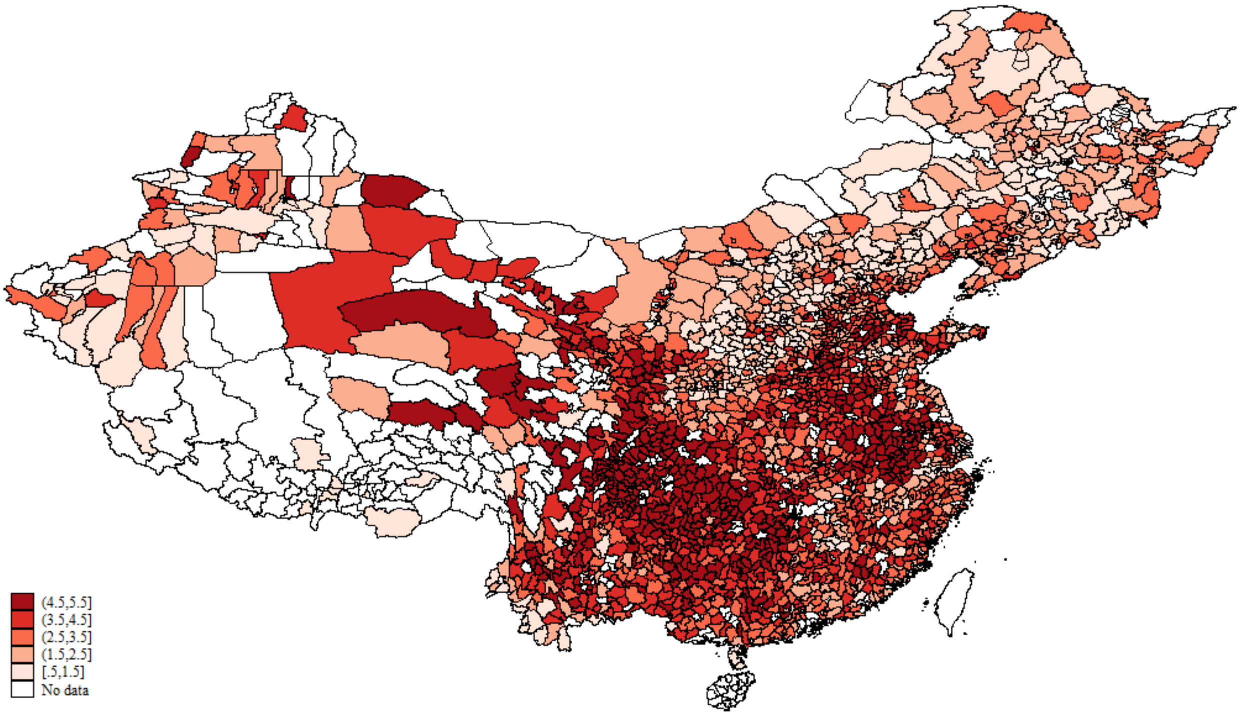


Figure 1. Famine severity: Variation by county

Notes: Quintiles of severity of famine as represented by the county relative cohort loss rate (difference between famine and normal periods) based on the 1990 Population Census; Darker color represents more severe famine.

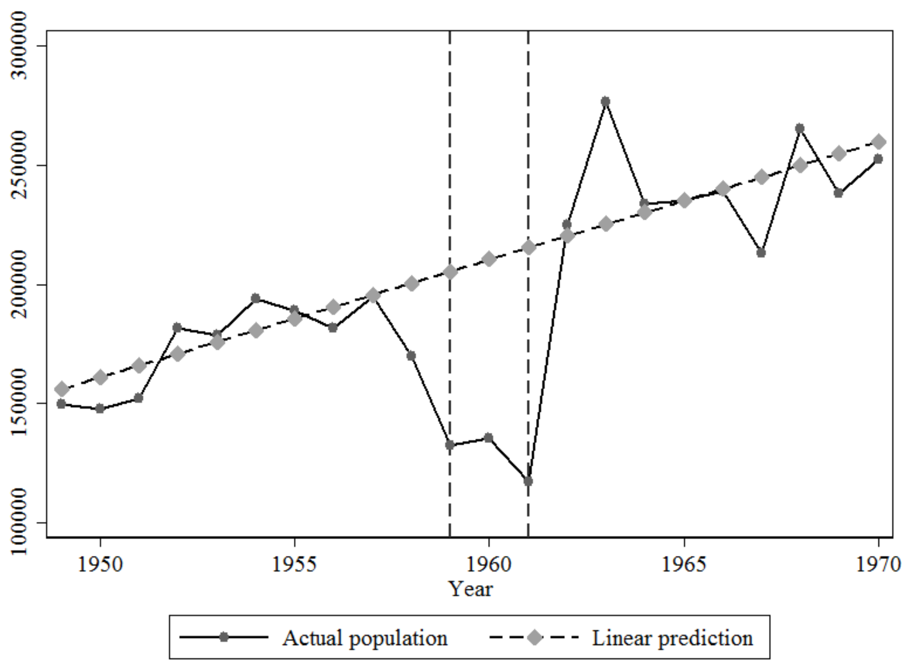


Figure 2. Population by birth cohort: Projected vis-à-vis observed

Note: Projected based on years 1949-57 and 1963-70 using the 1990 Population Census.

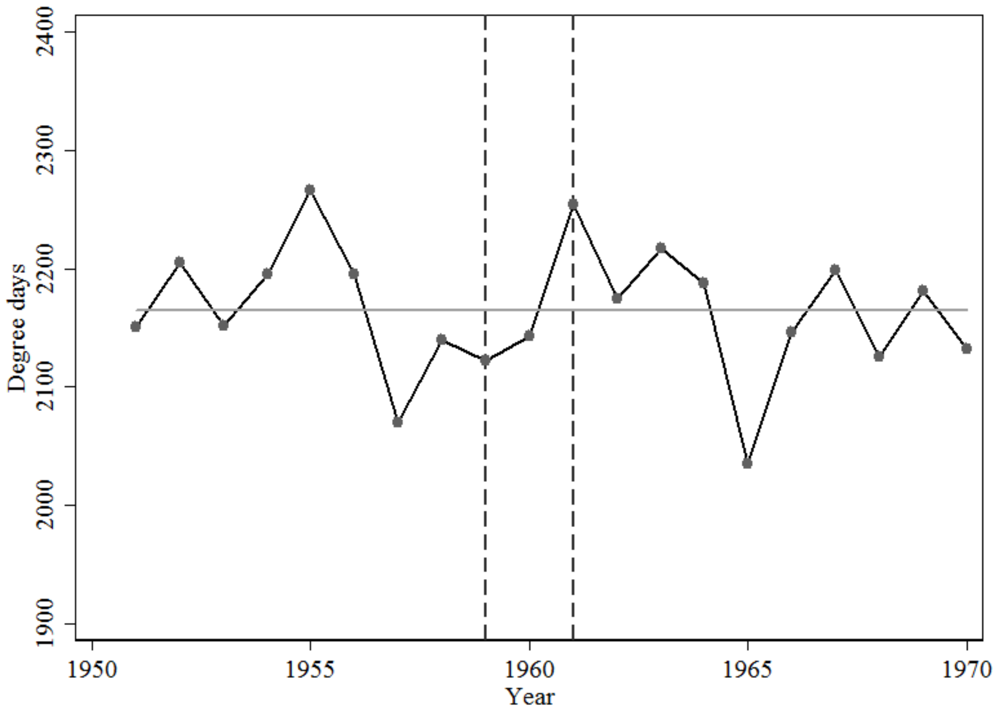


Figure 3. Thermal agricultural productivity

Notes: Graph depicts thermal agricultural productivity, or average sum of degree days between April 1 and September 30 (equation (4)) over all counties by year; the horizontal line depicts average thermal agricultural productivity over the period 1951-70.

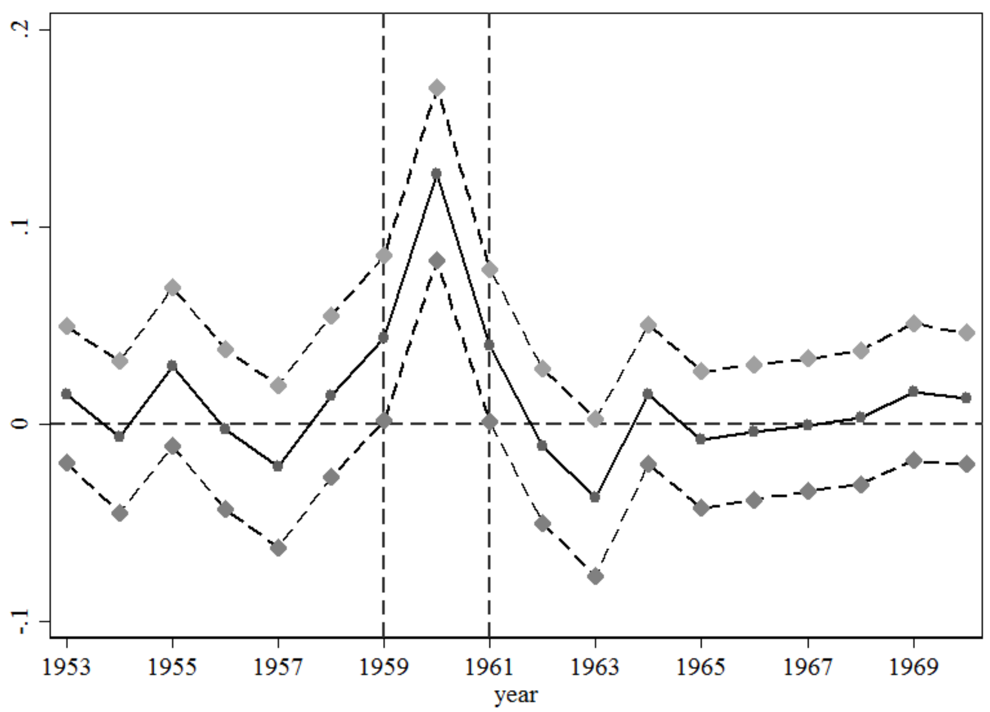


Figure 4. Cohort loss rate and thermal agricultural productivity

Notes: Graph depicts the estimated year-specific coefficients from regression of rate of cohort loss on previous year's thermal agricultural productivity (Table 2, column (b)) and 95% confidence intervals.

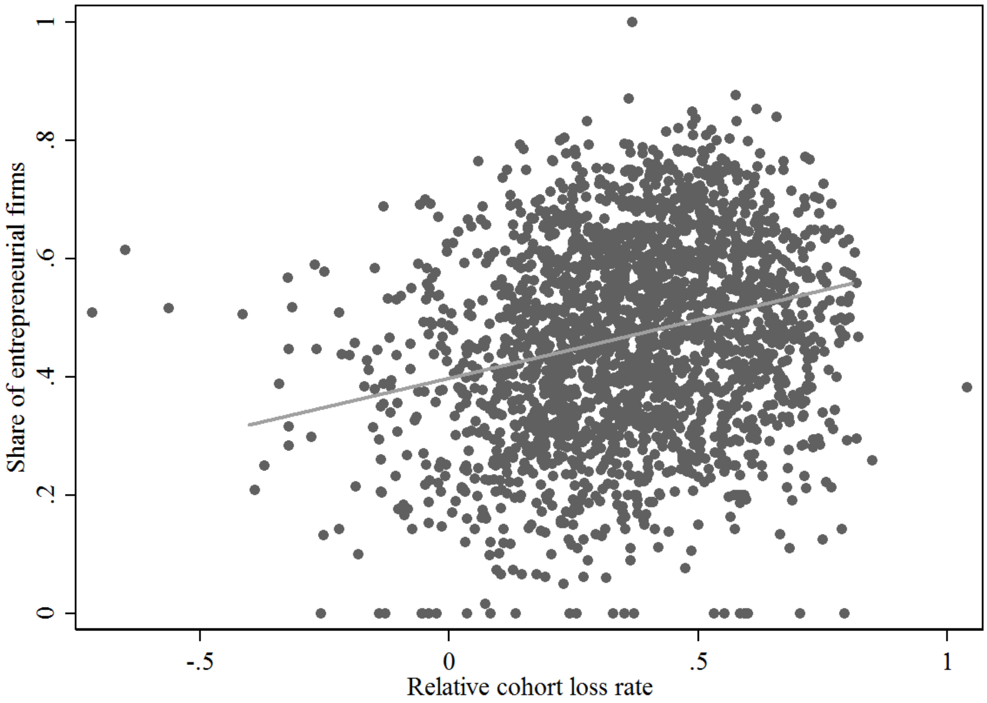


Figure 5. Famine severity and entrepreneurship
Notes: Based on 2004 Economic Census; Ratio of number of private enterprises with fewer than 100 employees and less than 10 years old to number of all enterprises; Each dot represents one county; the line depicts the linearly fitted values; Slope coefficient 0.147 (s.e. 0.020).

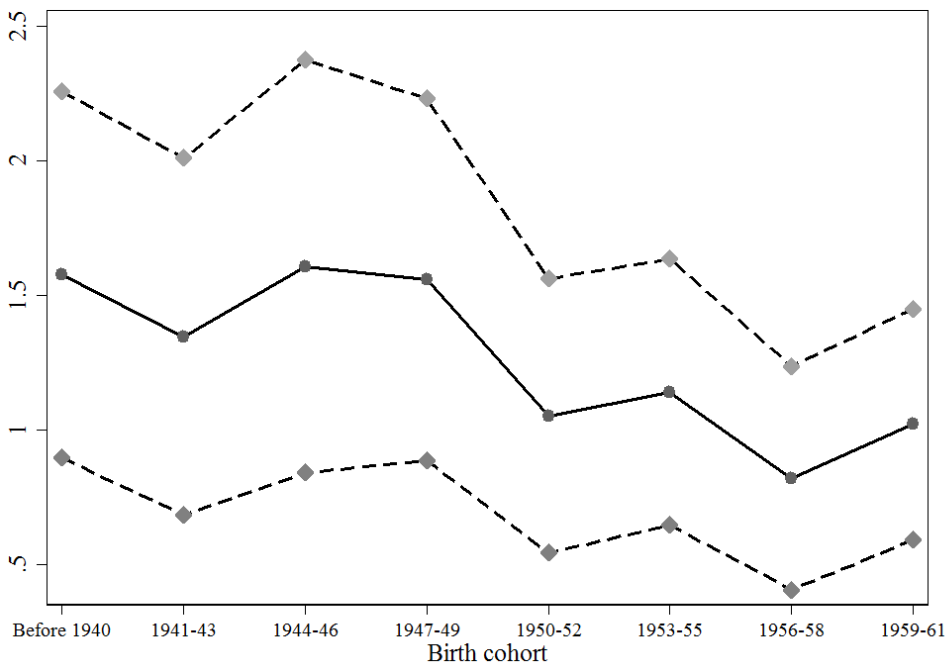


Figure 6. Famine severity and entrepreneurship: Individual analysis by cohort
Notes: Figure depicts estimated elasticities from Table 6, IV estimates of individual entrepreneurship (owner of private enterprise or self-employed) on relative cohort loss rate by birth cohort, and 95 percent confidence intervals.

Entrepreneurship and the School of Hard Knocks:
Evidence from China's Great Famine
Appendix (not for publication)

A.1 Famine severity and agricultural productivity

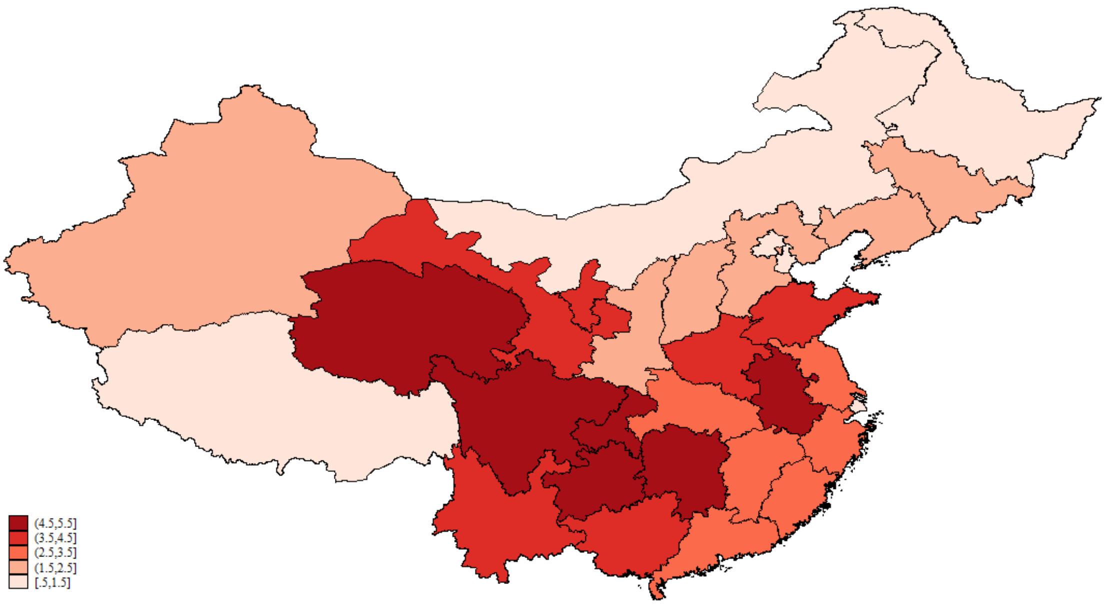
A.1.1 Severity of Great Famine

We use the 1990 Population Census, 1% sample, to construct the measure of the severity of the Great Famine. This census covers 2,600 counties. We retain those who have *hukou* registration in their county of residence and have lived in the county for five or more years, which amount to 93.95% of the population. For each county and year of birth, we calculate the population and use 1949-1957 and 1963-1970 as basis years to project the counterfactual population. We drop any county with zero persons in any cohort in the period 1949-1970, thus, reducing the number of counties in the sample to 2,545. Then, for each county and year of birth, we construct the cohort loss as the difference between the projected and actually recorded number of people. The rate of cohort loss (cohort loss divided by the projected population) in the famine birth cohorts represents the intensity of and hardship during the famine.

Similarly, we use the 2000 1% Population Census to construct alternative measures of the severity of the famine to use in a robustness check. This census covers 2,869 counties, among which we drop 186 that have zero persons in some cohorts, leaving 2,683 counties.

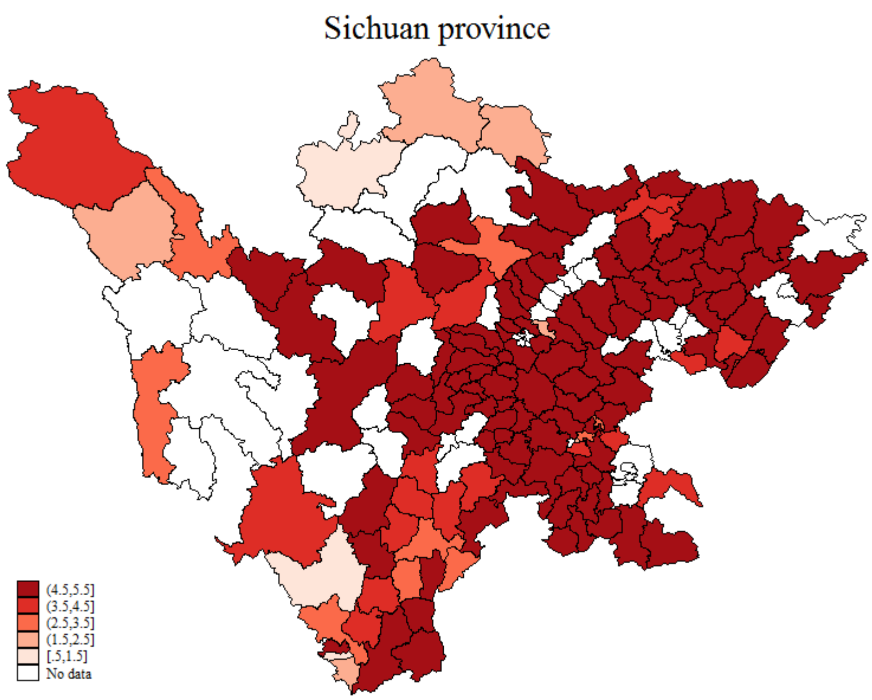
The severity of the famine during the Great Leap Forward differs across provinces and within provinces. Figure A1 depicts the severity by province, defined as the average rate of relative cohort loss in all counties in the province. Figure A2 depicts the severity by county in Sichuan province, one of the provinces that suffered most during the Great Famine. Apparently, there was substantial geographical variation in the severity of the famine.

Figure A1. Famine: Variation by province



Notes: Based on the quintiles of severity of famine as represented by the average of the county relative cohort loss rate (difference between famine and normal periods) in the province; Darker color represents more severe famine.

Figure A2. Famine in Sichuan: Variation by county



Notes: Based on the quintiles of severity of famine as represented by the county relative cohort loss rate (difference between famine and normal periods); Darker color represents more severe famine.

A.1.2 Agricultural thermal productivity

From the China Meteorological Administration, we download daily instrumental records for 727 weather stations over the years 1951-70. We measure the thermal agricultural productivity at each station and year by the sum of "degree days" according to equation (3), during the growing season, defined as April 1 to September 30. We exclude any station-year with incomplete coverage during the growing season. Table A1 reports estimates for the serial correlation in thermal agricultural productivity during 1951-1970. Columns (c) and (d) shows that, after controlling for station fixed effects, our construct of thermal agricultural productivity is a random walk.

Table A1. Thermal agricultural productivity: Serial correlation

	(a) Previous year	(b) Previous two years	(c) Previous year	(d) Previous two years
Thermal agricultural	0.987***	0.496***	0.001	-0.030
productivity (t-1)	(0.001)	(0.007)	(0.019)	(0.020)
Thermal agricultural		0.496***		0.013
productivity (t-2)		(0.007)		(0.017)
Station fixed effects	No	No	Yes	Yes
Stations	727	726	727	726
Observations	9,520	8,522	9,520	8,522
R-squared	0.977	0.983	0.000	0.001

Notes: Sample: All weather stations, years 1951-70; Estimated by ordinary least squares; Dependent variable: Agricultural thermal productivity; Robust standard errors clustered by station in parentheses (*** p<0.01, ** p<0.05, * p<0.1).

In a robustness check to account for the harmful effects of extremely high temperatures, we follow Richie and NeSmith (1991) and compute degree-days by the following four-part piece-wise linear function:

$$H_{cd} = \begin{cases} 0, & T_{cd} < 0 \\ T_{cd} - 8, & 8 < T_{cd} < 33 \\ \frac{25}{8}[41 - T_{CD}], & 33 \le T_{cd} < 41 \\ 0, & T_{cd} \ge 41 \end{cases} \qquad \text{(Eq. A1)}$$

Then, using equation (4), we compute the thermal agricultural productivity for this alternative measure. The average thermal agricultural productivity for the counties in the 2005 Population Census is 2527.49, which is similar to the average, 2527.43, using our baseline formula for degree-days.

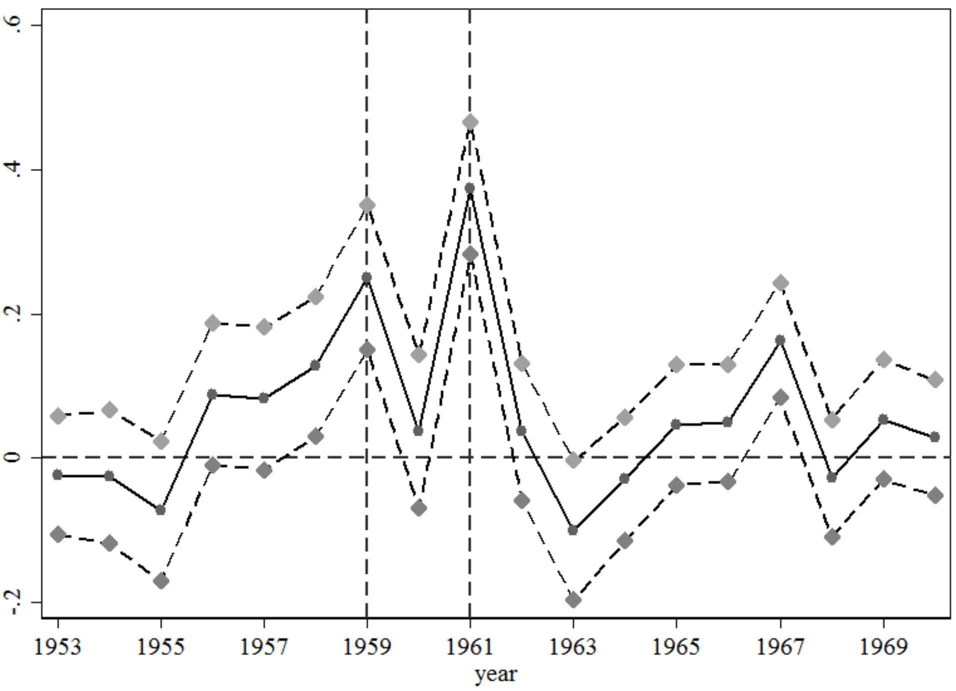
A.1.3 Thermal agricultural productivity and cohort loss rate

To associate the weather stations with the counties, we collect the longitude and latitude of the centroid of each county in the year 2000 from *Michigan China Data Center*, and then match each county with the nearest weather station by the Euclidean distance from the county centroid. On average, the distance between the weather station and county centroid is 29.51 km. To avoid measurement error, we limit our analysis to counties for which the distance from the centroid to the nearest station is less than 100km.

The county-level cohort loss rate constructed from the 1990 Population Census is then linked to thermal agricultural productivity by the county identifier in the year 2000. We match to county in 1990 and 2000 based on the record in China's Administrative Division History.[1] After dropping counties that changed borders in 1990 or 2000 (and counties with incomplete weather information), the 1990 Population Census covers 2,265 counties, the 2000 Census covers 2,589 counties, and the combined 1990 and 2000 Censuses cover 2,200 counties.

Figure A3 illustrates the coefficient of year indicators, η_y, in Table 2. The thermal agricultural productivity in the preceding year explains the variation in missing persons, except for the unusual years 1958, 1959, 1961, and 1967.

Figure A3. Coefficients of year indicators, η_y



Notes: Graph depicts the estimated year-specific coefficients, η_y, in regression of rate of cohort loss on year dummies (Table 2, column (c)).

[1] www.gov.cn/test/2007-03/23/content_559267.htm contains detailed information with which to track changes in administrative divisions at the county level between 1949-2006.

To further validate our construct for the severity of famine, Table A2 presents estimates of equation (5) with the rate of cohort loss based on the 2000 Census and combined 1990 and 2000 Censuses.

Table A2. Thermal agricultural productivity and rate of cohort loss

	Estimates for 2000 Census			Estimates for combined census		
	(a)	(b)	(c)	(d)	(e)	(f)
VARIABLES	β_1	λ_y	η_y	β_1	λ_y	η_y
Thermal agri prody	-0.059**			-0.045**		
previous year	(0.023)			(0.020)		
Year 1953		-0.015	0.068		-0.001	0.027
		(0.018)	(0.044)		(0.013)	(0.031)
Year 1954		-0.028	0.044		-0.015	0.009
		(0.019)	(0.047)		(0.016)	(0.039)
Year 1955		0.016	-0.029		0.023	-0.048
		(0.019)	(0.047)		(0.017)	(0.041)
Year 1956		0.003	0.068		0.002	0.078*
		(0.019)	(0.046)		(0.017)	(0.042)
Year 1957		-0.035*	0.122**		-0.025	0.099**
		(0.019)	(0.048)		(0.018)	(0.043)
Year 1958		-0.001	0.137***		0.008	0.140***
		(0.020)	(0.049)		(0.018)	(0.043)
Year 1959		0.042**	0.249***		0.040**	0.263***
		(0.020)	(0.050)		(0.019)	(0.046)
Year 1960		0.130***	-0.010		0.132***	0.015
		(0.021)	(0.052)		(0.020)	(0.048)
Year 1961		0.048**	0.343***		0.035**	0.393***
		(0.019)	(0.048)		(0.017)	(0.041)
Year 1962		-0.010	0.003		-0.019	0.053
		(0.019)	(0.048)		(0.017)	(0.040)
Year 1963		-0.022	-0.157***		-0.038**	-0.095**
		(0.019)	(0.047)		(0.017)	(0.041)
Year 1964		0.003	-0.021		-0.000	0.007
		(0.017)	(0.043)		(0.015)	(0.036)
Year 1965		0.006	-0.011		-0.014	0.059*
		(0.017)	(0.043)		(0.015)	(0.035)
Year 1966		-0.002	0.062		-0.011	0.078**
		(0.016)	(0.040)		(0.014)	(0.032)
Year 1967		0.009	0.152***		-0.008	0.191***
		(0.016)	(0.040)		(0.014)	(0.033)
Year 1968		0.005	-0.036		-0.003	-0.008
		(0.016)	(0.040)		(0.014)	(0.033)
Year 1969		0.018	0.081**		0.007	0.100***
		(0.016)	(0.040)		(0.014)	(0.034)
Year 1970		0.041**	-0.045		0.021	0.014
		(0.016)	(0.040)		(0.014)	(0.033)

Counties	2,589	2,200
R-squared	0.370	0.480

Notes: This table reports the OLS estimate of equation (5); Dependent variable: Cohort loss rate; with county fixed effects; Column (a) and (d): Coefficient of previous year's thermal agricultural productivity, β_1; Column (b) and (e): Coefficient of previous year's thermal agricultural productivity interacted with year, λ_y; Column (c) and (f): Coefficient of year indicator, η_y; Robust standard errors clustered by county in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

A.2 Effect of famine on entrepreneurship: Robustness checks

A.2.1 Aggregate analysis

The 2004 Economics Census was the first National Economic Census covering all economic sectors. This census contains detailed financial information for 1,375,148 business entities in 2,860 counties. Among private enterprises, 719,610 are less than 10 years old and have fewer than 100 employees, which we classify as entrepreneurial. Using this definition, we calculate the entrepreneurial firms' share in all enterprises, entrepreneurial firms' share of all enterprise sales, entrepreneurial firms' share of all enterprise profit, median age of private enterprises, and entrepreneurial firms' share of all enterprise employment for each county. After dropping counties with incomplete information for some of the above measures of entrepreneurship and those that changed boundaries between 1990 and 2004, our sample for county-level analysis comprises 2,194 counties.

Table A3 reports checks of robustness of the county-level analysis of the relation between famine severity and entrepreneurship presented in Table 4. Table A3, columns (a)-(f) present IV estimates of equation (8) with the dependent variable being the entrepreneurial share of all enterprises. Table A3, column (a), is the preferred estimate from Table 4, column (c). Table A3, column (b), includes controls for county-level gender ratio, Han ethnicity, migration rate, urbanization, average years of education, and relative size of each age cohort (based on the 2000 Population Census). Table A3, column (c), includes control for government expenditure.[2] This variable accounts for the possibility that the government's policy towards entrepreneurship is endogenously related to the Great Famine. Table A3, column (d), includes control for county-level agricultural productivity in 1993, measured as agricultural output per capita. This variable accounts for the possibility that agricultural productivity is related to the severity of the famine and also related to the opportunity for entrepreneurship. Table A3, column (e), reports an IV estimate accounting for spatial correlation in errors. In Table A3, column (f), enterprises are defined as entrepreneurial if they are less than 10 years old and have fewer than 500

[2] The source of the data is the Ministry of Finance's Statistical Reports of All Prefectures, Cities, and Counties (*Quanguo Dishixian Caizheng Tongji Ziliao*) for the year.

employees. Table A3, column (g), reports an IV estimate of equation (8) with the dependent variable specified as the entrepreneurial firm share of employment of persons with at least high school education in the private sector.

Table A3. Famine severity and county-level entrepreneurship: Robustness checks

VARIABLES	(a) Preferred: Entre. share of private enterprises	(b) Entre. share of private enterprises: Incl. county controls	(c) Entre. share of private enterprises: Incl. govt exp	(d) Entre. share of private enterprises: Incl. agri prody	(e) Entre. share of private enterprises: Spatial correlation	(f) Entre. share of private enterprises with < 500 employees	(g) Entre. share of private enterprise high school educated employees
Relative cohort	0.658***	0.713***	0.608***	0.677***	0.723***	0.698***	0.220***
loss rate	(0.119)	(0.231)	(0.110)	(0.121)	(0.161)	(0.127)	(0.065)
Counties	2,194	2,194	2,172	2,086	2,194	2,194	2,194
Hansen J stats	0.56	0.01	0.05	0.67		0.35	0.72
p-value	0.45	0.93	0.83	0.41		0.56	0.40
Elasticity	0.511	0.553	0.536	0.525	0.561	0.506	0.476
p-value	<0.001	0.002	<0.001	<0.001	<0.001	<0.001	0.001

Notes: Sample: Enterprises covered by 2004 Economic Census; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Dependent variable is entrepreneurial firm share in all private enterprises, except column (f); Column (a): Baseline estimate from Table 4, column (c); Column (b): Including county-level controls; Column (c): Including county government expenditure; Column (d): Including county agricultural productivity; Column (e): Allowing for spatial correlation in errors; Column (f): Entrepreneurial firm is defined as less than 10 years old with fewer than 500 employees; Column (g): Dependent variable is entrepreneurial firm share of employment of persons with at least high school education in the private sector; Robust standard errors clustered by weather station in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

A.2.2 Individual analysis

Table A4 reports robustness checks of the individual-level analysis presented in Table 5. Table A4, column (a), presents the OLS estimates controlling for county-level government expenditure, column (c). Table A4, columns (b)-(f), report estimates that respectively omit the year 1963 from the base years used to project the counterfactual population, project the counterfactual population by a quadratic model, project by an exponential growth model, and measure the intensity of the famine by the gender ratio rather than cohort loss, and project the counterfactual population using the 2000 Census.

Table A4, column (g), includes control for rebound of the population in 1962. Table A4, column (h), reports an estimate using agricultural thermal productivity measured by equation (A1) to account for the negative effect of extremely high temperatures. Table A4, columns (i)-(k), report estimates with estimated standard errors clustered by province, prefecture, and county respectively. Table A4, columns (l)-(n), include controls for government expenditure, agricultural productivity and ethnicity, gender, age, urbanization, and education, respectively. Table A4, columns (o)-(p), respectively account for spatial correlation in errors and report an estimate using probit rather than the linear probability model in the instrumental variables regression.

Table A4. Famine severity and individual entrepreneurship: Robustness checks

VARIABLES	(a) OLS: Government expenditure	(b) Exclude 1963 from base years	(c) Quadratic projection	(d) Expon-ential projection	(e) Gender ratio	(g) Projection using 2000 census	(f) Rebound	(h) Extreme temp. degree day
Relative cohort	0.001	0.087***	0.102***	0.087***	0.307	0.092***	0.088***	0.088***
loss rate	(0.005)	(0.020)	(0.023)	(0.020)	(0.191)	(0.021)	(0.018)	(0.020)
Observations	639,443	639,443	639,443	639,443	635,256	725,214	639,443	639,443
Counties	2,212	2,212	2,212	2,212	2,150	2,527	2,212	2,212
Hansen J stats		1.150	0.481	1.179	3.413	1.144	0.723	1.205
p-value		0.283	0.488	0.278	0.0647	0.285	0.395	0.272
Elasticity	0.0130	0.928	1.162	0.900	-0.0990	0.897	0.922	0.918
p-value	0.818	< 0.001	< 0.001	< 0.001	0.108	< 0.001	< 0.001	< 0.001

Table A4 cont'd

VARIABLES	(i) Govern-ment expenditure	(j) Agricu-ltural produc-tion	(k) Individual controls	(l) Std error clustered by province	(m) Std error clustered by prefecture	(n) Std error clustered by county	(o) Spatial correlation	(p) Probit
Relative cohort	0.081***	0.090***	0.134***	0.088	0.088***	0.088***	0.092***	1.843***
loss rate	(0.018)	(0.021)	(0.024)	(0.058)	(0.024)	(0.013)	(0.0331)	(0.281)
Observations	639,443	639,443	639,443	639,443	639,443	639,443	2,212	639,443
Counties	2,212	2,212	2,212	2,212	2,212	2,212	2,212	2,212
Hansen J stats	0.0964	1.018	0.465	0.575	1.583	3.819		
p-value	0.756	0.313	0.495	0.448	0.208	0.051		
Elasticity	0.988	0.943	1.440	0.917	0.917	0.917	1.020	1.843
p-value	< 0.001	< 0.001	< 0.001	0.129	< 0.001	< 0.001	0.006	< 0.001

Notes: Sample: Local residents born before 1962 covered by 2005 Population Census; Estimated by two-stage least squares weighted by population except columns (o) and (p); Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Dependent variable is indicator of owning private enterprise or self-employed; Robust standard errors clustered by weather station in parentheses, except columns (i)-(k) (*** p<0.01, ** p<0.05, * p<0.1). Column (a): OLS estimate controlling for government expenditure; Column (b): Exclude 1963 from base years for projection of counterfactual population; Column (c): Project counterfactual population by quadratic model; Column (d): Project counterfactual population by exponential growth model; Column (e): Represent intensity of famine by gender ratio rather than cohort loss rate; Column (f): Project counterfactual population using 2000 Census; Column (g): Control for rebound of population size in 1962; Column (h): Represent thermal agricultural productivity by sum of degree days according to equation (A1); Column (i): Control for county-level government expenditure; Column (j): Control for county-level agricultural productivity in 1993; Column (k): Control for ethnicity, gender, age, urbanization, and education; Column (l): Cluster estimated standard errors by province; Column (m): Cluster estimated standard errors by prefecture; Column (n): Cluster estimated standard errors by county; Column (o): Account for spatial correlation in errors (estimated by Stata routine, spreg); Column (p): Estimate by probit with instrumental variables (Stata routine, ivprobit).

Figure A4 depicts the estimated elasticities of the number of entrepreneurs (owners of private enterprise or self-employed) with respect to the relative cohort loss rate by birth cohort from Table 8, and 95 percent confidence intervals.

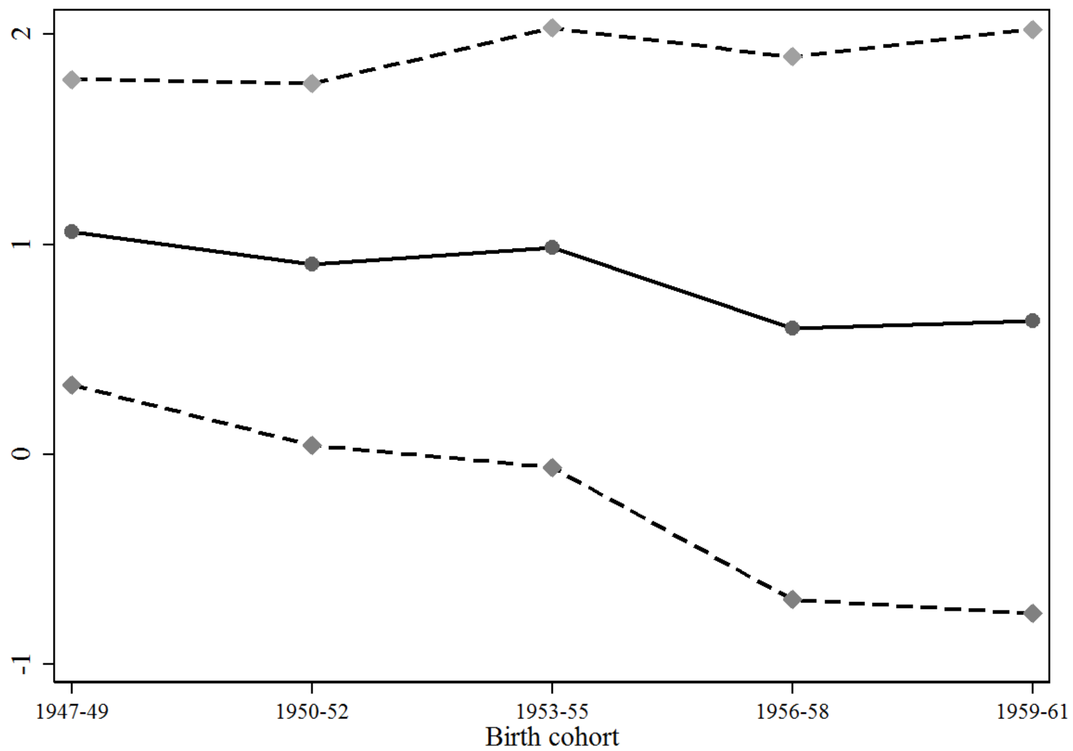


Figure A4. Operant conditioning: IV estimates by birth cohort

Notes: Figure depicts estimated elasticities from Table 8, IV estimates of logarithm of the number of entrepreneurs (owner of private enterprise or self-employed) on relative cohort loss rate by birth cohort, and 95 percent confidence intervals.

Table A5 reports robustness checks of the individual-level analysis of the effect of the famine on entrepreneurship beyond selective mortality presented in Table 8. Table A5, column (a), presents the IV estimate of the logarithm of the number of entrepreneurs on the relative cohort loss rate and logarithm of the population without famine. The sample is all persons born between 1947 and 1961. The estimation method is two-stage least squares regression weighted by population, with the instruments being the thermal agricultural productivity in the famine period and the difference in thermal agricultural productivity between famine and normal periods. Table A5, columns (b)-(n) check robustness to the method of projecting the counterfactual population, estimation of standard errors, and controls for local and individual characteristics.

Table A5. Increase in entrepreneurship beyond selective mortality: Robustness checks

VARIABLES	(a) IV: Basic estimate	(b) Exclude 1963 from base years	(c) Quadratic projection	(d) Expon-ential projection	(e) Gender ratio	(g) Projection using 2000 census	(f) 1962 rebound	(h) Extreme temp. degree day
Relative cohort	2.286*	2.295*	2.824*	2.314*	-6.225	2.458	2.633**	2.273*
loss rate	(1.322)	(1.327)	(1.593)	(1.344)	(6.628)	(1.583)	(1.252)	(1.325)
Population (ln)	0.572	0.571	0.451	0.554	2.000**	0.793	0.327	0.582
	(0.704)	(0.703)	(0.746)	(0.717)	(0.915)	(0.682)	(0.644)	(0.707)
Counties	477	477	477	477	477	477	477	477
Elasticity	0.809	0.833	1.108	0.815	0.0840	0.837	0.931	0.804
p-value	0.084	0.084	0.076	0.085	0.348	0.120	0.036	0.086

Table A5 cont'd

VARIABLES	(i) Govern-ment expend-iture	(j) Agri-cultural produc-tivity	(k) Std error clustered by province	(l) Std error clustered by prefecture	(m) Std error clustered by county	(n) Individual controls	(o) Spatial correlation
Relative cohort	2.626**	2.616**	2.286	2.286*	2.286**	3.836**	3.836**
loss rate	(1.251)	(1.212)	(2.540)	(1.201)	(0.948)	(1.539)	(1.539)
Population (ln)	0.341	0.415	0.572	0.572	0.572	0.389	0.389
	(0.558)	(0.552)	(0.785)	(0.568)	(0.530)	(0.928)	(0.928)
Counties	443	477	477	477	461	477	477
Elasticity	0.936	0.926	0.809	0.809	0.809	1.357	1.650
p-value	0.036	0.031	0.368	0.057	0.016	0.013	0.035

Notes: Sample: Local residents born before 1962 covered by 2005 Population Census; limited to counties with positive number of entrepreneurs (owners of private enterprises or self-employed) across all three-year birth cohorts between 1947-59 and 1959-61; Estimated by two-stage least squares weighted by population except column (o); Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Dependent variable is the logarithm of the number of entrepreneurs; Robust standard errors clustered by weather station in parentheses, except columns (i)-(k) (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Column (a): Basic estimate; Column (b): Exclude 1963 from base years for projection of counterfactual population; Column (c): Project counterfactual population by quadratic model; Column (d): Project counterfactual population by exponential growth model; Column (e): Represent intensity of famine by gender ratio rather than cohort loss rate; Column (f): Project counterfactual population using 2000 Census; Column (g): Control for rebound of population size in 1962; Column (h): Represent thermal agricultural productivity by sum of degree days according to equation (A1); Column (i): Control for county-level government expenditure; Column (j): Control for county-level agricultural productivity in 1993; Column (k): Control for ethnicity, gender, age, urbanization, and education; Column (l): Cluster estimated standard errors by province; Column (m): Cluster estimated standard errors by prefecture;

Column (n): Cluster estimated standard errors by county; Column (o): Account for spatial correlation in errors (estimated by Stata routine, spreg).

A.3 Individual psychology: Risk tolerance and other personality traits

A.3.1 Risk tolerance

We use the China General Social Survey (CGSS) 2010, 2012, and 2013, China Family Panel Studies (CFPS) 2010, and the China Household Finance Survey (CHFS) 2011 to derive measures of risk tolerance.

The CGSS is the first comprehensive national wide social survey project in China and is administered by Renmin University of China and the Hong Kong University of Science and Technology. It covers 125 counties from 31 provinces. The survey has two phases, during the periods, 2003-2008 and 2010-2019, with five waves in each phase cycle. The questions asked differ by wave. Only the CGSS 2010, 2012, and 2013 contain the question on financial investment. We collect and compile the measure of investment from the answers to the following questions:

Are you currently engaging in the following investment activities?				
1.Stock	2. Fund	3. Bond	4. Futures	5. Options
6. Real estate	7. Foreign currency	8. Others	9. None of the above	

The CFPS is a large-scale, almost nationally representative panel survey conducted by the Institute of Social Science Survey at Peking University. The baseline wave started in 2010 and covered 157 counties from 25 provinces of China. We collect and compile measures of investment from the answers to the following question:

At the end of last year, did you hold any of the following financial products?			
1. Stock	2. Fund	3. Bond	4. None of above

The CHFS, administered by the Southwestern University of Finance and Economics, is the only nationally representative survey in China that elicits detailed financial information about household finance and assets, including housing, business assets, financial assets, and other household assets. The first wave of the survey was conducted in summer 2011 with a sample size of 8,438 households and 29,500 individuals. We collect and compile the measure of investment from the answers to the following questions:

Does your family have any stock accounts?
1. Yes 2. No

Which of the following assets does your family own? (Can choose multiple answers)
1.Bonds 2.Mutual Funds 3.Derivatives 4. Wealth Management Products 5.None

We combine responses to these surveys, limiting to individuals born before 1962. After dropping the counties which changed boundaries between 1990 and the time of the survey, our sample comprises 21,965 individuals residing in 248 counties. Of these, 9,596 are from the CGSS, 9,560 are from the CFPS, and 2,809 are from the CHFS.

As investment is a binary variable, Table A6 presents the robustness checks for risk tolerance using a probit model.

Table A6. Risk tolerance: Robustness checks

VARIABLES	(a) ivprobit: Share investment	(b) ivprobit: Financial investment	(c) ivprobit: Owner
Relative cohort loss rate	1.141**	1.065*	0.993*
	(0.578)	(0.551)	(0.510)
Survey fixed effects	Yes	Yes	Yes
Wave fixed effects	Yes	Yes	Yes
Observations	21,965	21,965	21,965
Counties	248	248	248

Note: Sample: Respondents to CFPS 2010, CHFS 2011, CGSS 2010, CGSS 2012, or CGSS 2013 born before 1962; Estimated using Stata routine, ivprobit; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years (* $p<0.1$; ** $p<0.05$; *** $p<0.01$).

A.3.2 Self-confidence

Self-efficacy, a psychological concept which we interpret as self-confidence, has been related to entrepreneurship (Koellinger et al. 2007: Landier and Thesmar 2009: Hayward et al. 2010: Bullough et al. 2014). The CGSS 2011 includes the following five questions that gauge self-confidence. We recode all measures in the positive direction, so that a larger value represents a higher degree of self-confidence.

1. On a basis 1 to 10, where 10 represents highest level of power, please rate your power.
2. Have you ever lost confidence in the past four weeks?
   1. Very frequently 2. Frequently 3. Sometimes 4. Rarely 5. Never
3. Have you encountered any difficulties that you cannot overcome in the past four weeks?
   1. Very frequently 2. Frequently 3. Sometimes 4. Rarely 5. Never
4. In the past four weeks, have you have the ever had the following feeling: "I cannot control my life"?
   1. Very frequently 2. Frequently 3. Sometimes 4. Rarely 5. Never
5. In the past four weeks, have you have the ever had the following feeling: "I am confident that I can handle problems in my life"?
   1. Never 2. Rarely 3. Sometimes 4. Frequently 5. Very frequently

Table A7 reports IV regressions of self-confidence on the severity of the famine, as represented by the relative cohort loss rate. The coefficients of the relative cohort loss rate for self-assessed power and ability to overcome difficulties are positive. However, perhaps due to the small sample size, the coefficients are not statistically significant. The coefficients of the relative cohort loss rate are negative in three other measures of self-confidence, which is not consistent with hardship inducing greater confidence. Accordingly, we infer that the data do not support an inference that the famine induced greater self-confidence.

Table A7. Self-confidence: Individual analysis – IV estimates

VARIABLES	(a) Power	(b) Confidence	(c) Overcome difficulties	(d) Control own life	(e) Confidence
Relative cohort loss rate	1.136	-0.295	0.113	-0.143	-0.200
	(1.211)	(0.528)	(0.586)	(0.446)	(0.602)
Observations	1,671	1,628	1,652	1,683	1,680
Counties	47	47	47	47	47

F statistic	3.85	3.83	3.86	3.82	3.83
Hansen J statistic	2.177	0.977	0.933	2.577	0.124
p-value	0.140	0.323	0.334	0.108	0.724

Notes: Sample: Respondents to CGSS 2011 born before 1962; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Column (a): Dependent variable is self-assessed power on 10-point Likert scale with higher score representing greater power; Column (b): Dependent variable is response to question, “Have you ever lost confidence in the past four weeks?” on 5-point Likert scale with lower score representing stronger agreement; Column (c) Dependent variable is response to question, “Have you encountered any difficulties that you cannot overcome in the past four weeks?” on 5-point Likert scale with lower score representing stronger agreement; Column (d): Dependent variable is agreement with statement, “I cannot control my life.” on 5-point Likert scale with lower score representing stronger agreement; Column (e): Dependent variable is agreement with statement, “I am confident that I can handle problems in my life.” on 5-point Likert scale with higher score representing stronger agreement; Columns (a)-(e): Kleibergen-Paap rk LM statistics reject under-identification (p < 0.05); F statistic is for first stage. Robust standard errors clustered by county in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

As self-confidence is measured on integer Likert scales, Table A8 presents the robustness checks for self-confidence using an ordered probit model.

Table A8. Self-confidence – Ordered probit

VARIABLES	(a) Power	(b) Confidence	(c) Overcome difficulties	(d) Control own life	(e) Confidence
Relative cohort loss rate	0.957**	-0.434	0.035	-0.386	-0.205
	(0.387)	(0.403)	(0.395)	(0.397)	(0.377)
Observations	1,671	1,628	1,652	1,683	1,680
Counties	47	47	47	47	47

Notes: Sample: Respondents to CGSS 2011 born before 1962; Estimated using Stata routine, cmp; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years (* p<0.1; ** p<0.05; *** p<0.01).

A.3.2 Tenacity and resilience

Another personality trait that promotes entrepreneurship is determination, persistence, or tenacity (Baum and Locke 2004; Markman et al. 2005). The CGSS 2008 and 2011 include two questions that elicit this trait. One asks the respondent the degree to which he/she agrees with the statement, “I will exert effort to finish the task even when not feeling well”, and the other asks the degree to which he/she agrees with the statement, “I will exert effort to finish the task even it takes a long time to finish”.

Table A9 presents IV estimates of tenacity. The coefficients of the relative cohort loss rate are negative, which is not consistent with hardship raising tenacity. We infer that the data do not support an inference that the famine induced greater tenacity.

Table A9. Tenacity: Individual analysis – IV estimates

VARIABLES	(a) Even if not well	(b) Even if long time
Relative cohort loss rate	-0.483	-0.346
	(0.308)	(0.267)
Wave fixed effects	Yes	Yes
Observations	3,343	3,289
Counties	116	116
F statistic	8.79	8.72
Hansen J statistic	0.017	0.001
p-value	0.895	0.979

Notes: Sample: Respondents to CGSS 2008 or CGSS 2011 born before 1962; Instruments are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Column (a): Dependent variable is agreement with statement, "I will exert effort to finish the task even when not feeling well", on 4-point Likert scale with higher score representing stronger agreement; Column (b): Dependent variable is agreement with statement, "I will exert effort to finish the task even it takes a long time to finish", on 4-point Likert scale with higher score representing stronger agreement; Columns (a)-(b): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$); F statistic is for first stage. Robust standard errors clustered by county in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Yet another personality trait that promotes entrepreneurship is resilience (Bullough et al. 2014). Did hardship during the Great Famine underpin later entrepreneurship by selecting more resilient people or conditioning people to be more resilient? To gauge resilience, we exploit Chairman Mao Zedong's policy during the Cultural Revolution to move graduates of junior and senior high school from urban areas to live in the countryside. In a double randomization, we study the contingent effect of the famine and having been "sent-down" on physical health and tenacity.[3]

The CGSS 2008 includes information on whether the respondent reported being sent down during the Cultural Revolution, and the respondent's health and tenacity. Table A10, columns (a)-(c), report IV estimates of the respondent's self-reported health. In Table A10, column (c), the coefficient of the interaction between the relative cohort loss

[3] Gong et al. (2014) exploit the "send-down" movement to analyze the effect of adversity during adolescent years on non-cognitive skills.

rate and having been sent down, -11.196 (s.e. 9.772), is negative, which is not consistent with a more severe famine increasing resilience.

Table A10. Resilience: Individual analysis – IV estimates

VARIABLES	(a) Health	(b) Health	(c) Health	(d) Tenacity	(e) Tenacity	(f) Tenacity
Relative cohort loss rate	0.787		1.634**	0.207		0.131
	(0.613)		(0.664)	(0.303)		(0.295)
Send-down		0.774	3.519		0.423	-0.071
		(0.815)	(2.358)		(0.512)	(0.815)
Relative cohort loss rate			-11.196			3.328
x send-down			(9.772)			(3.689)
Observations	1,544	1,544	1,544	1,530	1,530	1,530
Counties	73	73	73	73	73	73
F statistic	5.35	28.65	5.96	5.29	29.24	5.91
Hansen J statistic	1.985		2.509	1.347		1.860
p-value	0.159		0.285	0.246		0.394

Notes: Sample: Respondents to CGSS 2008 born before 1962; Estimated by two stage least squares; Instruments for relative cohort loss rate are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Instrument for send-down is birth cohort times *hukou* registration type; "Send-down" as reported by survey respondent; Columns (a)-(c): Dependent variable is self-reported health condition on 5-point Likert scale with higher score representing better health; Columns (d)-(f): Dependent variable is agreement with statement, "I will exert effort to finish the task even when not feeling well", on 4-point Likert scale with higher score representing stronger agreement; Columns (a)-(f): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.05$); F statistic is for first stage regression of relative cohort loss rate. Robust standard errors clustered by county in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table A10, columns (d)-(f), report IV estimates of the respondent's tenacity, as represented by the degree of agreement with statement, "I will exert effort to finish the task even when not feeling well". (To address the endogeneity of being sent-down, we use the birth cohort and hukou registration type at birth as instrumental variables.) In Table A10, column (f), the coefficient of the interaction between the relative cohort loss rate and having been sent down, 3.328 (s.e. 3.689), is positive, which is consistent with people experiencing more severe famine becoming more resilient. However, perhaps owing to the small sample, the coefficient is not precisely estimated. Accordingly, we infer that the data do not support an inference that the famine induced greater resilience.

Tenacity and resilience are measured on integer Likert scales, Table A11 shows that the findings with respect to tenacity and resilience are robust to estimates using ordered probit and Poisson regressions.

Table A11. Tenacity and Resilience: Robustness checks

VARIABLES	(a) Even if not well	(b) Even if long time	(c) Health	(d) Health	(e) Health	(f) Tenacity	(g) Tenacity	(h) Tenacity
Relative cohort loss rate	-0.933***	-0.734**	0.803**		0.502***	0.333		0.039
	(0.281)	(0.288)	(0.368)		(0.142)	(0.404)		(0.098)
Send-down				0.754	0.737***		0.795	-0.016
				(0.755)	(0.270)		(0.813)	(0.261)
Relative cohort loss rate					-2.359			0.826
x send-down					(1.725)			(0.967)
Observations	3,343	3,289	1,544	1,544	1,544	1,530	1,530	1,530
Counties	116	116	73	73	73	73	73	73

Notes: Sample: Columns (a) and (b): Respondents to CGSS 2008 or CGSS 2011 born before 1962; Columns (c)-(h): Respondents to CGSS 2008 born before 1962; Instruments for relative cohort loss rate are thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal years; Instrument for send-down is birth cohort times *hukou* registration type; Columns (a)-(d) and (f)-(g) are estimated using Stata routine, cmp; Columns (e) and (h) are estimated using Stata routine ivpoisson as Stata cannot implement IV-oprobit with multiple endogenous variables; (* p<0.1; ** p<0.05; *** p<0.01).

A.3.4 Craftiness and opportunism

Risk tolerance, self-confidence, tenacity, and resilience are positive traits. In addition to these, particularly in societies with weak institutions, entrepreneurship is also associated with opportunism, craftiness, and skirting authority. An old Chinese saying is *wu shang bu jian, wu jian bu shang*, which means “no businessman is not crafty, without craftiness there is no business”. It is plausible that crafty or opportunistic people were more likely to survive the famine, or the famine conditioned people to become more crafty or opportunistic, which led to more entrepreneurship in later life.

The CGSS 2008 includes six questions, listed below, to elicit respondents' views on the use of *guanxi* (personal connections) to gain advantage. We use these to gauge the individual’s degree of opportunism or craftiness. We apply principal components analysis to the six questions, and code the principal component such that a higher score represents stronger agreement with the use of *guanxi*.

It is very common to use personal connections to gain advantage, do you agree with the following statements based on your own experience:					
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
a. Everybody uses *guanxi* to gain advantage	1	2	3	4	5
b. Using *guanxi* to gain advantage is a Chinese tradition	1	2	3	4	5
c. People use *guanxi* to gain advantage only if they have limited power	1	2	3	4	5
d. Using *guanxi* to gain advantage does not violate the rule of fairness	1	2	3	4	5
e. Using *guanxi* to gain advantage, the earlier the better	1	2	3	4	5
f. Using *guanxi* to gain advantage, the closer the relationship, the better	1	2	3	4	5

Table A12, column (a), reports an IV estimate of the respondent's agreement with using *guanxi* to gain advantage. The coefficient of the relative cohort loss rate is negative, which is not consistent with more opportunistic or crafty people being more likely to survive the famine or famine conditioning people to become more opportunistic or crafty.

Table A12. Craftiness, opportunism, obedience to authority:
Individual analysis – IV estimates

VARIABLES	(a) Use of *guanxi*	(b) Use of bribery	(c) Police (ln)
Relative cohort loss rate	-0.468	0.223	-5.494***
	(1.102)	(0.463)	(1.684)
Population (ln)			1.663***
			(0.324)
Observations	1,506	5,634	2,180
Counties	73	83	2,180
F statistic	6.12	7.25	21.06
Hansen J statistic	2.696	3.644	
p-value	0.101	0.056	

Notes: Estimated by two-stage least squares with instruments being thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Column (a): Sample comprises respondents to CGSS 2008 born before 1962; Dependent variable: The principal component of agreement with six statements regarding use of *guanxi* to gain advantage, where higher score represents stronger agreement; Column (b): Sample comprises respondents to CFPS 2010 born before 1962; Dependent variable: Agreement with using bribery to achieve success on 5-point Likert scale with higher score representing stronger agreement; Column (c): Sample comprises counties with any person reporting police as occupation in 2000 Population Census; Dependent variable is logarithm of number of police; Relative cohort loss rate and population specified as endogenous; Columns (a)-(c): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.05$); F statistic is for first stage. Columns (a)-(b): Robust standard errors clustered by county; Column (c): Robust standard errors clustered by weather station (* p<0.1; ** p<0.05; *** p<0.01).

Another test of opportunism or craftiness uses the CFPS 2010, which asks respondents to indicate the degree to which they agree that “To achieve success in today’s society, it is impossible to avoid bribery”, on a 5-point Likert scale with higher score representing stronger agreement. Table A12, column (b), reports the IV estimate of the respondent's agreement with using bribery. The coefficient of the relative cohort loss rate, 0.223 (s.e. 0.463), is positive, which is consistent with people experiencing more severe famine being more opportunistic or crafty. However, perhaps due to the small sample, the coefficient is imprecisely estimated.

Yet another trait through which the famine might have stimulated entrepreneurship is skirting or outright disobedience of authority. Absent better data, we investigate this aspect through the number of police in the county, with a control for the population. Counties whose residents are less law-abiding might employ more police. We use information on occupation in the 2000 Population Census to enumerate the police. Table A12, column (c), reports a county-level IV estimate of the logarithm of the number of police officers. The coefficient of the relative cohort loss rate, -5.494 (s.e. 1.684), is negative and statistically significant. Apparently, counties that experienced more severe famine had fewer police, which suggests that their residents are more law-abiding. This empirical finding is inconsistent with the famine inducing people to skirt or disobey authority.

A.4 Alternative explanations

Our analysis of the effect of hardship on entrepreneurship is not experimental and uses observational data, and so, the empirical relations are necessarily open to alternative explanations. We next consider several alternative explanations based on the effect of the famine on the gender ratio, old-age support, and local government and institutions.

A.4.1 Gender ratio

The Great Famine killed male fetuses at a higher rate than females, thereby raising the ratio of girls to boys (Mu and Zhang 2011). When these children reach the age of marriage, the imbalance in gender would challenge the girls to get attractive marriage partners. This might induce parents of girls to work harder and take more risk, and specifically, engage in entrepreneurship to earn more, and so to better compete to win marriage partners for their daughters (Wei and Zhang 2011).

To check, we investigate whether the famine induced people with daughters to engage in entrepreneurship relatively more. The 2005 1% Population Census asks only women for the gender of their children, and so, our analysis is limited to women. Table A13, column (a), reports an IV estimate of entrepreneurship (defined as owner or being self-employed) among women. The coefficient of the relative cohort loss rate, 0.071 (s.e. 0.019), is slightly smaller than the corresponding coefficient for men and women, 0.088 (s.e. 0.021) (reported in Table 5, column (c)), which is consistent with women being less likely to engage in entrepreneurship.

Table A13. Alternative explanations: Children and state capacity

VARIABLES	(a)	(b)	(c)	(d)	(e)	(f)
	IV:	IV:	IV:	OLS:	IV:	IV:
	Entrepre-	Entrepre-	Entrepre-	Road	Entrepre-	Entrepre-

	neurship	neurship	neurship	density	neurship	neurship
Relative cohort	0.071***	0.057**	0.076***	-1.777***	0.080***	0.056***
loss rate	(0.019)	(0.023)	(0.019)	(0.425)	(0.019)	(0.018)
Daughter		-0.018***				
		(0.006)				
Relative cohort loss		0.022				
rate x daughter		(0.018)				
No son			0.016*			
			(0.008)			
Relative cohort loss			-0.013			
rate x no son			(0.025)			
Gradient				-0.359***		
				(0.056)		
Road density					-0.011	-0.006
					(0.007)	(0.022)
Relative cohort loss						-0.004
rate x road density						(0.062)
Observations	228,535	228,535	228,535	2,171	629,805	629,805
Counties	2,211	2,211	2,211	2,171	2,171	2,171
F statistic	19.39	14.96	20.31	37.85	18.70	14.34
Hansen J statistic	0.026	0.102	0.131		0.953	12.57
p-value	0.872	0.950	0.937		0.329	0.002

Notes: All columns except (d): Estimated by two-stage least squares with instruments being thermal agricultural productivity in famine period and difference in thermal agricultural productivity between famine and normal periods; Dependent variable is indicator of owning private enterprise or self-employed. Columns (a)-(c): Sample: Local female residents born after 1940 and before 1962 covered by 2005 Population Census; Column (b): Additional instrument -- thermal agricultural productivity in famine period interacted with daughter; Column (c): Additional instrument -- thermal agricultural productivity in famine period interacted with no son; Column (d): Estimated by ordinary least squares; Dependent variable is road density; Columns (e)-(f): Sample: Local residents born before 1962 covered by 2005 Population Census; Column (e): Additional instrument – county average gradient; Column (f): Additional instruments – county average gradient and thermal agricultural productivity in famine period interacted with gradient; Columns (a) and (e): Kleibergen-Paap rk LM statistics reject under-identification ($p < 0.001$); F statistic is for first stage regression of relative cohort loss rate. Robust standard errors clustered by station in parentheses (* p<0.1; ** p<0.05; *** p<0.01).

Table A13, column (b), reports an IV estimate of entrepreneurship among women, distinguishing those with daughters. The coefficient of having a daughter is negative and significant, suggesting that women with daughters were less likely to engage in entrepreneurship. This is consistent with the Chinese practice of the bridegroom's family bearing the expenses of marriage. The coefficient of the relative cohort loss rate interacted with having a daughter is positive, 0.022 (s.e. 0.018), which is consistent with women with daughters and who suffered relatively more during the famine being more likely to engage in entrepreneurship. However, the estimate is not statistically significant.

The 2005 1% Population Census does not report the children of men, so we cannot carry out the corresponding analysis for men. Intuitively, given Chinese men's strong preference for sons (Tan et al. 2014), we expect any relation among men with daughters to be weaker than the relation among women with daughters. The estimate of the effect of famine on women with daughters is imprecise. Accordingly, we are inclined to rule out the hypothesis that the famine induced people to engage in entrepreneurship so as to earn more to attract spouses for their daughters.

A.4.2 Old-age support

Another alternative explanation emphasizes the effect of the famine on old-age support. The Great Famine killed many children and reduced fertility, even leaving some women permanently unable to bear children. Traditionally, Chinese people look to their sons for support in old age. Without any sons, a couple must plan to provide for themselves in later years. One way is to work harder and take more risk by engaging in entrepreneurship to build up savings and even to continue economic activity after the normal retirement age.

Table A13, column (c), reports an IV estimate of entrepreneurship among women, distinguishing those with no sons. The coefficient of having no son is positive and marginally significant, consistent with the traditional notion that people without sons must worker harder and take more risks to support themselves in old age. The coefficient of the relative cohort loss rate interacted with not having any son is negative, -0.013 (s.e. 0.025), which is not consistent with women with no sons and who suffered relatively more during the famine being more likely to engage in entrepreneurship.

A.4.3 Government and institutions

The relation between entrepreneurship and the famine might be due to differences in government. Differences in the intensity of the famine might have been due to differences in state capacity (Lu et al. 2016). In areas with greater state capacity, officials were better able to enforce procurement, and so, caused more suffering. Or, in areas with weaker state capacity, officials were less able to mobilize production, and thus caused more suffering. Differences in state capacity would affect the opportunity for private-sector entrepreneurship, depending on whether state action is a complement to or a substitute for entrepreneurship (Jia and Lan 2014). However, as argued in the main text, representing the severity of the famine by the difference in the rate of cohort loss abstracts from any non-time varying county-level differences. Indeed, Table A4 shows that our findings are robust to controls for demographic structure, urbanization, and education.

A more serious concern is that government policies responded endogenously to the severity of the famine. Specifically, the government may have provided more help to areas that suffered relatively more during the famine, which fostered more entrepreneurship in later years. Table A4, reports an estimate including control for government expenditure to represent government assistance. The coefficient of the relative cohort loss rate is similar to that in the estimate without control (reported in Table 5).

Information on county-level government expenditure is coarse, and, in particular, government expenditure in urban areas is only available by district, which is one level above the county. To represent government assistance more precisely, we use the density of roads (kilometers of roads per square kilometer area) in the year 2000. Since this variable might be endogenous to entrepreneurship, we instrument for road density by the average gradient of the terrain. The gradient would affect the cost of building roads, but would not directly affect entrepreneurship.

Accordingly, we extract the length of roads in each county in the year 2000 from the China Data Center at the University of Michigan, and compute the density as the ratio of the length of roads in kilometers to the area of the county in square kilometers. We obtain information on terrain from the China Historical Geographic Information System (CHGIS), Harvard Yenching Institute, and calculate the average gradient as the difference in altitude between the highest and lowest points in the county, divided by the area of the county.

We first regress, at county level, road density on the severity of the famine, as represented by the relative cohort loss rate. As Table A13, column (d), reports, the coefficient of the relative cohort loss rate, -1.777 (s.e. 0.425), is negative and statistically significant. This suggests that, in counties where the famine was more severe, the government in recent times has been less effective and state capacity is lower.

Next, we regress entrepreneurship on the relative cohort loss rate and road density. Referring to Table A13, column (e), the coefficient of road density is not statistically or economically significant. The coefficient of the relative cohort loss rate, 0.080 (s.e. 0.019), is quite similar to the estimate without controlling for road density (Table 5, column (c)).

Finally, we add the interaction between the relative cohort loss rate and road density, with road density specified relative to the sample average. As Table A13, column (f), reports, the coefficient of the relative cohort loss rate, 0.056 (s.e. 0.018), is positive and statistically significant, and the coefficient of road density, -0.006 (s.e. 0.022), is negative and insignificant. Both coefficients are much smaller than the estimates without controlling for the interaction between the relative cohort loss rate and road density (Table A13, column (e)). The coefficient of the relative cohort loss rate interacted with road density, -0.004 (s.e. 0.062), is almost zero, suggesting that hardship and state assistance are relatively independent in stimulating entrepreneurship.

References

Baum, Robert J. and Edwin A. Locke, “The Relationship of Entrepreneurial Traits, Skill, and Motivation to Subsequent Venture Growth”, *Journal of Applied Psychology*, 2004, 89 (4), 587-598.

Bullough, Amanda, Maija Renko, and Tamara Myatt, “Danger Zone Entrepreneurs: The Importance of Resilience and Self-Efficacy for Entrepreneurial Intentions”, *Entrepreneurship Theory and Practice*, 2014, 38 (3), 473-499.

Gong, Jie, Yi Lu, and Huihua Xie, “Adolescent Environment and Noncognitive Skills,” NUS Business School, November 2017.

Hayward, Mathew, William Forster, Saras Sarasvathy, and Barbara Fredrickson, “Beyond Hubris: How Highly Confident Entrepreneurs Rebound to Venture Again”, *Journal of Business Venturing*, 2010, 25 (6), 569-578.

Jia, Ruixue and Xiaohuan Lan, “Capitalism for the Children: Entrepreneurs with Cadre Parents Under Big Government”, University of California San Diego, December 2014.

Koellinger, Philipp, Maria Minniti, and Christian Schade, “I think I can, I think I can”: Overconfidence and Entrepreneurial Behavior, *Journal of Economic Psychology*, 2007, 28 (4), 502-527.

Landier, Augustin and David Thesmar, “Financial Contracting with Optimistic Entrepreneurs”, *Review of Financial Studies*, 2009, 22 (1), 117-150.

Lu, Yi, Mona Luan, and Tuan-Hwee Sng, “The Effect of State Capacity Under Different Economic Systems”, SSRN Working paper 2728605, January 2016.

Markman, Gideon D., Robert A. Baron, and David B. Balkin, “Are Perseverance and Self-Efficacy Costless? Assessing Entrepreneurs’ Regretful Thinking”, *Journal of Organizational Behavior*, 2005, 26 (1), 1-19.

Mu, Ren and Xiaobo Zhang, “Why does the Great Chinese Famine Affect the Male and Female Survivors Differently? Mortality Selection versus Son Preference”, *Economics & Human Biology*, 2011, 9 (1), 92-105.

Ritchie, J.T. and D.S. NeSmith, “Temperature and Crop Development,” in John Hanks and J.T. Ritchie, eds., *Modeling Plant and Soil Systems*, Crop Science Society of America, Inc., 1991, chapter 8, pp. 5–29.

Wei, Shangjin and Xiaobo Zhang, “Sex Ratios, Entrepreneurship, and Economic Growth in the People's Republic of China”, NBER Working paper 16800, 2011.