The Institutional Causes of China's Great Famine, 1959-1961*

Xin Meng Nancy Qian Pierre Yared

March 3, 2015

Abstract

This paper studies the causes of China's Great Famine, during which 16.5 to 45 million individuals perished in rural areas. We document that average rural food retention during the famine was too high to generate a severe famine without rural inequality in food availability; that there was significant variance in famine mortality rates across rural regions; and that rural mortality rates were *positively* correlated with per capita food production, a surprising pattern that is unique to the famine years. We provide evidence that an inflexible and progressive government procurement policy (where procurement could not adjust to contemporaneous production and larger shares of expected production were procured from more productive regions) was necessary for generating this pattern and that this policy was a quantitatively important contributor to overall famine mortality.

Keywords: Famines, Modern Chinese History, Institutions, Central Planning

JEL Classification: P2, O43, N45

---

*We are indebted to Francesco Caselli, an anonymous co-editor and seven anonymous referees for invaluable suggestions. We also thank Daron Acemoglu, Alberto Alesina, Abhijit Banerjee, Robin Burgess, Claudia Goldin, Mikhail Golosov, Rick Hornbeck, Chang-tai Hsieh, Dean Karlan, Michael Kremer, James Kung, Naomi Lamoureaux, Christopher Leighton, Cormac O'Grada, Suresh Naidu, Emi Nakamura, Nathan Nunn, Gerard Padro-i-Miquel, Chris Udry, Dennis Yang and Fabrizio Zillibotti for their comments; the participants at the Berkeley Development Seminar, Kiel Institute of International Economics, UCLA Applied Economics Seminar, Stanford Economic History Seminar, University of Texas at Austin Applied Micro Seminar, University of Toronto Applied Micro Seminar, Tsinghua University Economics Seminar, the Applied Workshop at CCER PKU, the Economics Workshop at Hong Kong University, the Applied Micro Seminar at the Chinese University of Hong Kong, Yale University Macro Lunch and Harvard History Tea, and the participants at BREAD-CIPREE, CEPR-DE, the NBER Summer Institute for Economic Growth, the University of Houston Health and Development Conference, and the Conference on Fetal Origins and Famines at Princeton University for useful comments; and Louis Gilbert, Sara Lowes, Ang Sun, David Liu, Meredith Startz, Jaya Wen, Yiqing Xu and Kai Yan for excellent research assistance. We acknowledge the Harvard Academy Scholars program for financial support. Please send comments to any of the authors: xin.meng@anu.edu.au, nancy.qian@yale.edu, pyared@columbia.edu.

1 Introduction

During the twentieth century, approximately seventy million people perished from famine. This study investigates the causes of the Chinese Great Famine (1959-61), which killed more than any other famine in history: 16.5 to 45 million individuals, most of whom were living in rural areas, perished in just over three years.[1] The existing literature on the causes of the Great Famine has formed a consensus that a fall in aggregate food production in 1959 followed by high government procurement from rural areas were key contributors to the famine.[2] In this paper, we argue that these factors could not have caused the famine on their own because average rural food availability was too high to generate famine, even after accounting for procurement. Thus, any explanation of the famine must account for inequality in food availability and famine mortality across rural areas. Motivated by this reasoning, we analyze the spatial relationship between agricultural productivity and famine severity. Using a panel of provinces, we find a surprising *positive* correlation between mortality rates and productivity across rural areas during the famine. We provide evidence that a procurement policy that was *inflexible* (i.e., procurement could not adjust to contemporaneous production) and *progressive* (i.e., procurement was a larger share of expected production in productive regions) was necessary for generating this pattern and that this policy was a quantitatively important contributor to overall famine mortality.

The goal of this study is to make progress on understanding the root causes of the famine by providing evidence for the novel hypothesis that the inflexibility of the centrally planned procurement system was an important contributing factor to the famine. Our study proceeds in several steps. The first step is to document that after procurement, rural regions as a whole retained enough food to avert mass starvation during the famine. Since the entire rural population relied on rural food stores, we compare the food retained by rural regions after procurement to the food required by rural regions to prevent famine mortality. Using historical data on aggregate food production, government procurement and population (adjusted for the demographic composition), we find that average rural food availability for the entire rural population was almost three times as much as the level necessary to prevent high famine mortality. We reach these conclusions after constructing the estimates to bias against finding sufficient rural food availability. Our findings are consistent with Li and Yang's (2005) estimates of high rural food availability for rural workers and imply that the high level of famine mortality was

[1] See Sen (1981) and Ravallion (1997) for estimates of total famine casualties. For the Chinese famine, mortality estimates range from 16.5 million (Coale, 1981) to 30 million (Banister, 1987) to 45 million (Dikotter, 2010).

[2] For more details, see the discussion at the end of the Introduction and in Section 2.

accompanied by significant variation in famine severity within the rural population.[3] There must have been some factor that caused a large rise in the inequality of access to food across the rural population, which, in turn, resulted in significant variance in mortality outcomes across the rural population.

The second step is to verify this conjecture by documenting that the increase in mortality rates during the famine was accompanied by an increase in the variance of mortality across the rural population. Since the historical mortality data are limited to total mortality at the regional level, we focus on spatial variation in mortality. We find that mortality rates across provinces varied much more during the famine than in other years. To investigate whether this pattern remains true at a more disaggregated level, we also use the birth-cohort sizes of survivors observed in 1990 to proxy for famine severity at the county-level.[4] This is based on the logic that famine increases infant and early childhood mortality rates and lowers fertility rates such that a more severe famine results in smaller cohort sizes for those born shortly before or during the famine. The data show that there is much more variation in cross-county birth-cohort sizes for famine birth-cohorts relative to non-famine birth-cohorts. This is true both across and within provinces. These findings imply that an explanation of the famine will need to account for the increase in both the mean and the variation of rural mortality rates.

Third, we document the empirical relationship between mortality rates and productivity across rural areas. We find that across rural regions, famine severity is *positively* correlated with per capita food production, a pattern that is unique to the famine era. This surprising correlation holds at the province-level, where we use mortality rates to proxy for famine severity and *per capita* grain production data to measure productivity.

We acknowledge that these estimates could be biased by contemporaneous misreporting of the historical data for production and mortality. For example, the Chinese government during the Great Leap Forward era (GLF 1958-61) was known to have over-reported grain production. Hence, one may be concerned that the positive correlation between mortality and reported production reflects the over-reporting of production in the more famine-stricken regions. To address this, we impute production with data that were not known to have been manipulated, such as historical weather conditions and geo-climatic suitability, and data that would have been easy to correct after the GLF, such as total rural population and total land area. To minimize the possibility that our results are driven by misreporting, the main empirical analysis of the spatial patterns uses the constructed production measure. Moreover, at the county-level, we show that famine survivor cohort sizes are negatively correlated with

[3] We compare Li and Yang's (2005) estimates to ours at the end of Section 3.

[4] There are no county-level historical data on mortality rates or food production.

the predictors of high grain productivity (e.g., geo-climatic suitability for grain production, rainfall and temperature).

The fourth step is to examine whether our empirical findings can be explained by existing theories of the causes of the famine, which have primarily focused on the role that policies specific to the GLF played in causing the famine. We take the variables for GLF policies used in existing studies and show that regional GLF intensity cannot explain the patterns we observe in the data and a new explanation is needed. Moreover, we show that productivity explains much more of the variation in the mortality increase during the famine than GLF intensity. Thus, explaining the correlation between productivity and famine severity across rural areas can potentially explain a large fraction of overall famine mortality.

Finally, we provide an explanation for the empirical findings. We argue that the spatial patterns of famine severity were the result of an inflexible and progressive government procurement policy combined with a fall in per capita production that was typically larger in magnitude in more productive regions, but not so large that it changed productivity ranks across provinces. In the late 1950s, the central government procured as much grain as it could from rural areas while leaving rural workers with enough food to be productive laborers. It was thus *progressive* as it procured a higher percentage of total production from productive areas. It was also *inflexible* as the government set each region's procurement level in advance such that it could not be easily adjusted afterward. This was because weak state capacity, along with political tensions, made communication challenging and hampered the government's ability to respond quickly to a harvest that was below expectations. As a consequence, the level of procurement from a given rural region did not respond to the actual amount produced, but was instead based on an estimated production target established earlier, where this target was itself based on past production. After procurement took place, the food retained in a given region would be negatively correlated with the difference between target production and realized production, i.e. the "production gap". Since more productive regions experienced a larger absolute production drop while still remaining more productive relative to less productive regions, the procurement policy caused more productive regions to experience a larger per capita production gap, subjecting them to more over-procurement, which in turn, led to less per capita food retention, less per capita food consumption and higher mortality rates.

To examine the plausibility of our explanation, we test the prediction of our proposed mechanism that mortality rates should be positively correlated with the production gap and that this should be true in all years during the 1950s and 60s since they were subject to a similar procurement policy.

To estimate the production gap, we construct a measure of target production that is based on past production and past production growth. The results show that province-level mortality is positively associated with the production gap, a relationship that is robust to controlling for GLF intensity.

A back-of-the-envelope calculation shows that the inflexible and progressive procurement mechanism explains 32-43% of total famine mortality. Hence, our proposed mechanism is quantitatively important, and at the same time leaves room for other factors, such as GLF policies and the complex political environment of the time, to contribute to famine mortality.

In addition, we use historical province-level procurement data to provide evidence on the causal links of our hypothesis. For example, we document that during the famine, the positive correlation between regional procurement rates and productivity increases, and regional food retention, which is the difference between production and procurement, is decreasing in the size of the production gap.

The main challenges for our study are data availability and quality. We address these difficulties by using a large array of data from contemporaneous, archival, Chinese and international sources. As we discussed earlier, we address the possibility of systematic over-reporting of production by constructing a measure of production that does not use GLF-era production data. We are also able to address potential concerns that mortality rates are misreported by proxying for famine severity with survivor birth cohort size as observed in 1990. The fact that the results are qualitatively similar across all data sources suggests that the spatial patterns we detect between agricultural productivity and famine severity are not driven by measurement error.

This study makes progress on understanding the causes of the Chinese famine in several ways. First, it is the first to propose that an inflexible and progressive procurement policy was an important contributing factor of the famine, and the first to document the surprising *positive* association between famine mortality rates and agricultural productivity. Second, it is one of the few studies that examine the determinants of regional procurement (most existing studies treat procurement as an exogenous variable). Since government procurement is a key determinant of regional food retention, and we show that inequality in food retention across regions is necessary for causing a large famine, understanding the determinants of regional procurement levels (i.e., the inequality in procurement across regions) is critical for understanding the famine. Another study that examines the determinants of regional procurement levels is Kung and Chen (2011). They find that political radicalism increased regional procurement during the famine and explains approximately 16% of total famine mortality. As such, our mechanism complements theirs in explaining total famine mortality. Finally, to the best of our knowledge, we are also the first to address potential measurement error in the GLF-era grain production data. Past studies have taken the official production data as given.

In directly examining the determinants of mortality, our study is similar to Lin and Yang (2000), which finds that conditional on average grain production, lower grain production (and higher urban population share) were associated with higher mortality during the famine. That a decline in production relative to a region's average production leads to higher mortality is consistent with our inflexibility mechanism. In another study of famine mortality, Kung and Lin (2003) finds that conditional on procurement, higher production is associated with lower mortality, and that conditional on production, lower procurement is associated with lower mortality. Related to this, Li and Yang (2005) finds that during the famine, lower food retention (production minus procurement) and zealousness in pursuing GLF policies lead to lower food production in the following year. Our finding that food retention is negatively associated with famine is consistent with these earlier studies.

This study also adds to the larger literature on famines. Our results are broadly consistent with Sen's (1981) thesis that famines are mainly due to food distribution rather than aggregate food deficits. However, we document spatial patterns in famine severity and food production that are difficult to explain with market mechanisms.[5] Thus, together with other studies of the Chinese famine, we expand the literature on the causes of famine by studying the detailed mechanisms in a non-market context.[6] This is an important context since over sixty percent of all famine deaths in the twentieth century taken place in non-market economies (e.g., China's Great Famine in 1959-61, the Soviet Famine in 1932-33, and the North Korean Famine in 1992-95).[7] In the Conclusion, we discuss the similarities between the Chinese famine and famines in other centrally planned economies.

This paper is organized as follows. Section 2 provides a brief historical background. Section 3 estimates rural food availability during the famine. Section 4 documents spatial variation in famine severity. Section 5 estimates the correlation between famine severity and food productivity across rural regions. Section 6 explains the empirical patterns with the inflexibility and progressiveness of the historical food procurement policy. Section 7 offers concluding remarks.

[5]For example, Sen (1981) found that the Bengal Famine (1943) was partly due to the inability of rural wage workers in famine stricken regions to buy food from regions with surplus production. With such a mechanism, mortality rates would be higher in less productive regions.

[6]For recent studies on the causes of famines in market economies, see studies such as Burgess and Donaldson (2010), (Shiue, 2002, 2004, 2005) and OGrada (2007). Also, Dreze (1999) and OGrada (2007) provide overviews of this literature. See Section 2 for a discussion on studies of the famine in China besides the ones that we have already mentioned.

[7]Davies and Wheatcroft (2004) estimate that up to 6.5 million died across the Soviet Union during the 1932 famine. In North Korea, it is commonly believed that 2-3 million individuals, approximately 10% of the total population, died during this famine (e.g., see Haggard and Noland, 2005; and Demick, 2009). There are very few academic studies or reliable accounts of details related to this famine.

2 Background

In this section, we provide a brief discussion about the historical background of the famine. For more detailed discussions, please see our earlier working paper, Meng et al. (2010), and the references below.

2.1 Collective Agriculture

On the eve of the famine, the production, distribution and consumption of food in China were entirely controlled by the central government. This meant that the government was the sole insurer of food consumption in the event of a drop in production.[8] At the time, approximately 80% of the population worked in agriculture.[9] Land reforms that began in 1952 had resulted in full collectivization by the end of the decade. Private property rights to land and assets were erased, and markets for private transactions were banned (Fairbank, 1987: p. 281-5). Agricultural workers were forced to work under constant monitoring and were no longer rewarded for their marginal input into production (Johnson, 1998). By the end of the 1950s, there were no wages or cash rewards for effort.[10]

Grain was harvested and stored communally. Private stores of grain were banned, a rule that was sometimes enforced with virulent anti-hiding campaigns (Becker, 1996: p. 109). Grain was procured by the central government from communal depots after the fall harvest around November. Procured grain was fed to urban workers, exported to other countries in exchange for industrial equipment and expertise, and stored in reserves as insurance against natural disaster.[11]

The grain retained by rural regions was fed to peasants in communal kitchens, which were established so that the collective controlled food preparation and consumption. The government prevented peasants from migrating, and thus, they were mostly bound to consume the amount distributed in their collective (Thaxton, 2008: p. 166). When that was insufficient, famine occurred.

2.2 Famine Chronology

The Great Famine is officially defined by the Chinese government to be three years, 1959-1961, when mortality rates were the highest. Grain production grew nearly monotonically between 1949 and 1957. There are a few accounts of production falls in select regions in 1958 and many accounts of widespread production falls in 1959 and 1960. Famine became widespread when local stores of the 1959 harvest ran

[8] See the previous version of this paper, Meng et al. (2010), for a detailed discussion on how agricultural collectivization during the 1950s reduced rural households' ability to smooth consumption.

[9] We calculate this from data reported by the National Bureau of Statistics (NBS).

[10] See Walker (1965, p. 16-7) for a detailed description of collectivization.

[11] Historical central planning documents state that approximately 4-5 million tons per year were put into reserves as insurance against natural disasters (Sun, 1958). During the late 1950s, total grain exports were approximately 2% of total production (Walker, 1984: Table 52).

out during the early part of 1960 (Thaxton, 2008: p. 207-10). Between 16.5 and 45 million individuals died during the three years in total.[12] Mortality rates were the highest in the spring of 1960 (Becker, 1996: p. 94).[13] The official explanation provided by the government was bad weather. Recent studies have provided evidence that the fall in output was also partly due to bad government policies such as the diversion of resources away from agriculture to industrialization, as well as weakened worker incentives.[14]

Famine primarily struck the rural areas. Communal kitchens, which survivors recall as having served large quantities of food, suddenly ran out. Peasants scavenged for calories and ate green crops illegally from the field (*chi qing*) when they could (Thaxton, 2008: p. 202). Mortality rates were highest for the elderly and young children (Ashton et al., 1984: Tables 3 and A7; Spence, 1991: p. 583). Prime-age adults experienced relatively higher survival rates (Thaxton, 2008: p. 202-10).

Relative to other famines, infectious diseases did not play a major role in causing mortality. The low level of disease in rural areas during the famine has been attributed to limited population movements, the prevalent use of DDT prior to the famine, and public health measures taken by the government during earlier years (e.g., Becker, 1996; Dikotter, 2010: ch. 32; Fairbank, 1986: p. 279). "People really did die of starvation−in contrast to many other famines where disease loomed large on the horizon of death" (Dikotter, 2010: p. 285). This is an important point to keep in mind for calculating the caloric requirement for survival in the Chinese context.[15]

In recent years, a broad consensus has formed that the government over-procured grain from rural areas in the fall of 1959, and this exacerbated the production decline and caused the massive mortality in the spring months of 1960. There are many hypotheses for what led to over-procurement. The central government placed the blame on local leaders, accusing them of over-reporting production and consequently leading the central government to over-estimate true production (Thaxton, 2008: p. 293-9). Recent academic studies find that over-procurement was driven by multiple factors, including the

[12]For example, mortality estimates range from 16.5 million (Coale, 1981) to 30 million (Banister, 1987) to 45 million (Dikotter, 2010).

[13]See the Introduction for references to studies of total famine mortality. For example, "By the end of 1959, much of the peasantry was starving to death but the hardest time in the entire famine came in January and February when the greatest number perished" (Becker, 1996: p. 94).

[14]The policies include labor and acreage reductions in grain production (e.g., Peng, 1987; Yao, 1999), implementation of radical programs such as communal dining (e.g., Chang and Wen, 1997; Yang, 1998), reduced work incentives due to the formation of the people's communes (Perkins and Yusuf, 1984), and the denial of peasants' rights to exit from the commune (Lin, 1990). Li and Yang (2005) compile province-level panel data on grain production and attempt to quantify the impact of various potential factors. They find that in addition to weather, the relevant factors were over-procurement and the diversion of labor away from agriculture during the Great Leap Forward for projects such as rural industrialization.

[15]Consistent with the belief that infectious diseases were not an important feature of the Chinese Famine, the data show that famine mortality rates were low in densely populated places such as urban areas and were uncorrelated with latitude and elevation. Moreover, the results we show in the paper are robust to controlling for population density and its interaction with the famine dummy variable. These correlations are available upon request.

government's bias towards providing high levels of food to urban areas (Lin and Yang, 2000), the political zealousness and career concerns of provincial leaders (Yang, 1998; Kung and Chen, 2011) and an over-commitment by the central government to meeting export targets (Johnson, 1998). In addition, some have argued that mortality rates were exacerbated by food wastage in communal kitchens (Chang and Wen, 1997).

The Chinese government did not begin to systematically respond to the famine until the summer of 1960, after a large proportion of famine mortality had already taken place. The response came in several forms. The government returned workers who had been recently moved to urban areas to assist in industrialization back to their home villages. This was intended to replenish the greatly weakened and demoralized rural labor force in order to minimize further falls in production (Li and Yang, 2005; Thaxton, 2008: p. 169). Urban food rations were reduced, although typically not to below subsistence levels (Lin and Yang, 2000). The government also abandoned many of the more extreme policies of collectivization (Walker, 1965: p. 83, 86-92; Thaxton, 2008: ch. 6, p. 215-6). For example, households again stored and prepared their own food, peasants were again allowed to plant strips of sweet potatoes for their own consumption, and the government sometimes also turned a blind eye to the black market trading of food across regions and the illegal consumption of green crops; all this helped preserve lives until the next harvest (Thaxton, 2008: ch. 4).

These measures could not prevent another decline in production in 1960, this time caused by the diminished physical capacity of the rural labor force, the lack of organic inputs such as seeds and fertilizers, which had been consumed during the months of deprivation (e.g., Li and Yang, 2005), and the consumption by starving peasants of green crops from the field (Thaxton, 2008: p. 202). In 1961, the government finally ended the famine by sending large amounts of grain into rural areas. Thirty million tons of grain reserves were depleted (Walker, 1984: Ch. 5) and China switched from being a net exporter to a net importer of grain (Walker, 1984: Table 52). Grain production recovered gradually in subsequent years.

After the famine, procurement rates were kept at a much lower level than during the famine-era. Official statistics show that aggregate procurement rates gradually declined from a peak of almost 30% of total production in 1960 to approximately 10% by 1965, and remained between 6% and 19% for the next twenty years.[16] The procurement policy remained largely unchanged otherwise. Consistent with

[16]Note that the aggregate procurement rates discussed in the next section and displayed in Table 1 differ slightly from the national average. This is because our analysis utilizes a restricted sample of provinces. All of our results are unchanged if one uses the full sample (as we do in a previous version of this paper) or the restricted sample. Results using the full sample are available upon request.

the low procurement levels and the government's need to feed its growing urban population, China remained a net importer of grain for several decades. The government did not attempt to re-implement the extreme policies from the GLF that were abandoned during the initial reaction to the famine. China experienced several aggregate production drops of approximately 5%-10% in per capita terms, but never experienced another fall as large as that of 1959 (i.e., 15%). These factors, together, may explain why there were no subsequent famines in China (Walker, 1965: ch. 6; Thaxton, 2008: ch. 6).

Politically, the central government engaged in various public campaigns to preserve political support during the famine's aftermath. This was necessary since the famine had primarily affected the rural population which represented the support base of the communist regime. The government limited the reports of famine and minimized the mortality numbers; it initiated large-scale propaganda campaigns such as *yiku sitan* to convince the population that bad weather and corrupt bureaucrats were to blame for low production and over-procurement; and it initiated the *fan wufeng* movement to allow peasants to punish local leaders for famine crimes (Thaxton, 2008: p. 293-9).

Our study focuses on the three years of the highest mortality rates, 1959 to 1961. Since the mid-autumn harvest is used to support life for many months of the following year, we focus on grain production during 1958 to 1960. The chronology of production and mortality rates portray a consistent picture where roughly constant levels of production in 1958 led to above-trend mortality rates in 1959, and bigger falls in production in 1959 led to extremely high mortality rates in 1960. Per capita production was similar during 1960 and 1961, during which time mortality rates declined. Per capita production began to grow in 1962, at which point mortality rates returned to trend. We document these facts explicitly in the next sections.

3 Rural Food Availability

We argue in this section that rural areas retained enough food post-procurement to avoid mortality. In this endeavor, we consistently bias our exercise in the direction of overestimating the food shortage. The main analysis uses a sample of 19 of the 24 provinces in China during the famine era. For these provinces, we have data for all of the main variables used in the analysis: mortality, urban and rural population, production, procurement, weather, and geographical conditions. For consistency, we use the same sample for all of the estimates presented in the main paper. Our sample comprises 77% of China's total population in 1958 and approximately 87% of total famine mortality.[17]

[17] The provinces in our sample are Anhui, Beijing, Fujian, Guangdong, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Shaanxi, Shandong, Shanghai, Shanxi, Tianjin and Zhejiang. Famine mortality was calculated by the authors using mortality data reported by the National Bureau of Statistics. In this instance, it is

3.1 Rural Caloric Requirements

To calculate rural food availability when the famine began, we compare production to per capita caloric requirements. The main exercise uses official production data, but we later show that the main result is robust to using constructed measures of production. We focus on a seven-year window that includes the three official years of famine, 1959-1961. Mortality rates during 1959-1961 depend on the consumption of food produced during 1958-1960 since harvests are used to support life for a large part of the following calendar year and only a small part of the current calendar year.

We calculate two benchmarks for caloric requirements: the caloric needs for heavy adult labor and healthy child development and the caloric needs for staying alive. We use very generous caloric recommendations provided by the United States Department of Agriculture (USDA) for the first benchmark, and we calculate the lower benchmark to be 43% of the higher one.[18] To adjust for the demographic structure, we use the published tables extracted from the *1953 Population Census*. Thus, we assume that the per capita food requirement in 1959 is similar to that of 1953. In the Online Appendix, we describe how we apply demographic data to the USDA guidelines to calculate the average per capita caloric needs for the population. We estimate that daily average per capita requirements in 1959 were 1,871 calories for the first benchmark and 804 for the second benchmark. These estimates are lower than average caloric requirements for adult workers because they also account for young and elderly individuals in the population, who require fewer calories per person. It is important to note that our estimates are extremely conservative and constructed to obtain high levels of caloric requirement.

The main caveat in interpreting these estimates is that the demographic structure may have changed between 1953 and 1959 such that per capita requirements were higher in 1959. This is unlikely because the proportion of prime-age adults in the population decreased during 1953-59.[19] Since the very young

defined as the deaths during the famine that was in excess to the average level of mortality during the six years prior to and five years after the famine (1953-58, 1962-66). Please see the Online Appendix for a detailed discussion.

[18] There is little evidence for how many calories are needed to stay alive. This is partly because starvation is typically accompanied by other conditions such as disease, unsanitary conditions, war, etc. All of these other factors contribute to mortality, but they are less applicable in our context than for most other famines since rural China had a low level of endemic disease, had high sanitary standards, and did not experience war. The best evidence we have on calories and starvation is from the *Minnesota Starvation Experiment*, where the U.S. military systematically subjected military volunteers to high levels of physical exertion, chronic starvation and harsh conditions (e.g., high temperatures). In these experiments, the lowest level of caloric consumption was approximately 900 calories a day. This was far above the level that would cause mortality and the subjects were still required to exercise. Similarly, Dasgupta and Ray (1986), in their well-known paper on nutrition and work, chose the 900 calories per day benchmark as the amount required for an adult male to do some work. For these reasons, we also chose 900 calories as the lower benchmark and interpret it as the extreme upper bound of the amount of calories required to survive. Since 900 calories is 43% of the level needed for heavy adult labor for prime age men, we assume the same proportional needs for all other groups.

[19] Crude demographic projections suggest that between 1953 and 1959, youth dependency (the number of people aged 0 to 10 per 100 people aged 20 to 64) approximately increased from 90 to 100. The rate of increase was constant over time and shows no change during the GLF. Similarly, the dependency ratio of the elderly (the number of people over 64 per 100 people aged 20 to 64) was constant over this period. These statistics suggest that the proportion of prime-age

and elderly require fewer calories than prime age adults, this would most likely cause us to over-estimate *average* per capita caloric needs in 1959.

Applying the 1953 estimates to the years immediately after the famine is more problematic. Since the young and the elderly were more likely to have perished during the famine, the average per capita caloric needs during the post-famine years could have been much higher than in 1953. Hence, the caloric accounting for the years after the famine should be interpreted cautiously.

3.2 Rural Food Availability

Grain production, total population and urban population data are published by the National Bureau of Statistics (NBS). The three province-level municipalities were mostly urban, but still contained significant agricultural populations that engaged in grain production.[20]

The data on grain production is not disaggregated by types, which include rice, sorghum and wheat. Therefore, to convert data for the volume of retained grain into calories, we use the Chinese *Ministry of Health and Hygiene's (MHH)* estimate of calories contained in the typical mix of grains consumed by an average Chinese worker, and we assume that one kilogram of grain provides 3,587 calories. Moreover, we assume that individuals subsist solely on grain, which is a reasonable description of the diet of Chinese peasants in the 1950s (Walker, 1984).

The data for national grain production and total population are displayed in Table 1 columns (1) and (2). In column (5), we use these data to estimate per capita production, which is then converted in column (6) into per capita food availability in terms of calories per day using the MHH's estimate of calories per kilogram of grain. Aggregate production in column (1) and estimated per capita food availability in column (6) show that per capita production was roughly constant between 1957 and 1958. And although production and food availability declined from 1958 to 1959, national production provided approximately 2,421 calories per capita in 1959, which is approximately 300% of the 804 calories necessary for preventing mortality and 129% of the 1,871 calories required for heavy labor and healthy child development. In other words, national average food availability based only on production was about three times the level needed to avoid mortality. Similarly, in 1960, per capita production was 2,101 calories per day and far above the two caloric requirement benchmarks.

We examine the amount of rural grain retention by using the production data described above and national procurement data reported by the Chinese Ministry of Agriculture (1983).[21] The national

adults in the population was declining (Kinsella et al., 2009).

[20]During the 1950s, 34%, 30% and 47% of the populations of Beijing, Shanghai and Tianjin were rural.

[21]The data report gross procurement as well as the amount that is "sold back" to the countryside for each province.

procurement rate presented in column (4) shows that it was around 28% higher in 1959 than in 1958, which was in turn 24% higher than in 1957. In 1960, procurement declined back to 1958 levels, which is consistent with the government having begun to respond to the famine in 1960.

In columns (7) and (8), we deduct government procurement from production and divide it by the rural population to estimate average rural caloric availability after procurement. Columns (7) and (8) show that despite the increased procurement rates, average rural per capita retention in 1958 and 1959 was far above the two benchmarks for caloric requirements. For example, consider retention in 1959, which was associated with the year of highest famine mortality (1960). The rural population was left with roughly 2,329 calories per capita per day. This is approximately 290% of the 804 calories necessary for preventing mortality and 124% of the 1,871 calories required for heavy labor and healthy child development. In other words, we find that average rural food availability was almost three times as high as the level required for avoiding mortality at the peak of the famine. Similarly, in 1960, we find that average rural caloric availability was approximately 2,168, which is 270% of the requirement for avoiding mortality.[22]

The main caveat for interpreting these estimates is that official data may overstate production in order to minimize the appearance of failure of GLF agricultural policies, or to understate procurement to minimize government culpability. However, such misreporting is unlikely to overturn our result that aggregate rural retention exceeded aggregate rural nutritional needs for avoiding famine for several reasons. First, we have followed recent studies on the famine in using production and procurement data that was corrected and reported during the post-Mao reform era, when the government had no incentive to glorify or undermine the GLF. Second, we have made assumptions throughout the exercise to bias our calculations towards finding low rural food availability (and high food requirements). Third, for aggregate procurement to cause aggregate rural retention to be below the nutritional needs for avoiding famine mortality, procurement would have needed to be 73% of total production in 1959, 70% in 1960 and 65% in 1961. These are three to four times the reported procurement rates and seem unlikely

We take the difference between these two quantities to compute net procurement for each province. We then aggregate it across provinces to compute aggregate procurement. These data have recently been used by Kung and Chen (2011).

[22] Our estimates of per capita rural caloric availability are comparable, though slightly higher than Li and Yang's (2005) estimate of 2,063 calories per rural worker. The difference arises because our estimates are based on net procurement while their estimates are based on gross procurement, which does not take into account the grain "sold back" given back to provinces. However, both estimates show that rural food availability was too high to have caused a famine without the presence of inequality within the rural population. Note that Ashton et al.'s (1984) estimates of national per capita food availability in 1959 is 1,820 calories per day. This is lower than the estimates from our study because they assume that the grain remaining after gross aggregate procurement is used to feed the entire population (both urban and rural). In practice, most of the procured grain was used to feed the urban population and the post-procurement retention was used to feed only the rural population. However, even their estimates are near the level for heavy labor and healthy child development, and much higher than the level needed to avoid mortality.

as there are no documented accounts of such drastic under-reporting of procurement. Finally, in the Online Appendix Section G, we show that our finding changes little if we use constructed production measures, which we describe in Section 5.1, and reasonable upper-bound values for procurement.

4 Spatial Variation in Famine Mortality

4.1 Across Provinces

The fact that average rural food retention was too high to inevitably cause high famine mortality implies that food retention, and therefore famine severity must have varied significantly across the rural population. To examine this, we follow existing studies in using mortality rates, reported by the NBS, to proxy for famine severity. Figure 1a plots average mortality rates and the normalized variance in mortality rates over time (the cross-province standard deviation divided by the cross-province mean).[23] It shows that both the mean and the variance of mortality rates spike upwards during the famine years.

Note that the historical mortality data do not distinguish between urban and rural populations, but as most of the famine mortality occurred in rural areas, we interpret the figure to reflect patterns of rural mortality.[24] Next, we examine the variation across rural populations explicitly using another data set.

4.2 Across Counties

To examine spatial variation in famine severity at a finer geographic level, we use another proxy for famine severity: the birth-cohort size of survivors amongst the agricultural population of each county, constructed from the 1% sample of the *1990 China Population Census*.[25] Birth-cohort size is negatively correlated with famine severity as it captures the reduced fertility and increased mortality caused by the famine. We construct this famine severity index for the county-level, which is the lowest official

[23] The normalization of the standard deviation addresses the concern that the standard deviation can be mechanically positively correlated with the mean.

[24] If we assume that mortality only occurs within the rural population, we can address this issue by normalizing by the rural population (normalized mortality rates = total mortality rates $\times$ total population/rural population). Data for total and rural population are also reported by the NBS. This normalization does not change the observed pattern of a spike in the mean and normalized variance of mortality rates during the famine. These estimates are omitted for brevity and are available upon request.

[25] Agricultural populations are defined to be households officially registered as agricultural. These statuses were assigned in the early 1950s and there was very little mobility from being an officially identified agricultural household to a nonagricultural household between then and 1990. The main distinction for agricultural households is their obligation to deliver a grain tax to the central government, their right to farmland, and their lack of access to urban public goods such as health care, schooling and housing. For these reasons, there is an unwillingness on both the government's and the farmers' sides to switch official statuses. An alternative way to identify rural populations is to identify everyone living in a non-urban county as rural. This does not affect our results. Due to space constraints we do not report these results with the alternative data construction. They are available upon request.

administrative division in China.[26] In addition to allowing us to identify rural individuals, this proxy provides several advantages over the mortality rate data. First, it allows us to disaggregate our analysis to the county-level and examine whether the same spatial variations observed at the province-level also exist at this lower level of administrative division. Second, this measure of famine severity is not vulnerable to the misreporting caused by the government's desire to understate famine severity. Given the focus of our paper on rural inequality in famine severity, we focus our analysis on agricultural populations.[27] For consistency, we use the same provinces as in our province-level exercise.

The effect of famine can be observed in the birth-cohort size of survivors. Figure 1b plots the size of birth-cohorts in 1990 for all of China. The straight dotted line illustrates the positive trend in birth-cohort size over time, from approximately 5,000 individuals per county (i.e., our sample contains 1% of the whole population) to approximately 9,000 per county. This increase reflects the combined forces of increased fertility and reduced infant and child mortality. The comparison of the actual birth-cohort sizes and the projected linear trend shows that the former begins to deviate from the trend for birth-cohorts born as early as 1954, and sharply declines for individuals born during the famine (1959-61), when the average cohort size is only approximately 4,000 individuals. It returns to trend afterwards. The negative deviation from the trend suggests that individuals who were aged approximately five years and younger when the famine began (i.e. those born 1954-58) were more likely to perish than older children. The steep decline for individuals born during the famine captures the additional vulnerability of very young infants to famine and the reduction in fertility during the years of the famine. These patterns are consistent with the belief that adult famine victims are likely to stop bearing children (by choice or for biological reasons) before they starve to death, as well as qualitative accounts that very young children were more likely than adults to perish and that very few children were born during the famine. Note that the high survival rate of the child-bearing population is consistent with the observed rebounding of cohort sizes soon after the famine.

To adjust the cohort size in a way that is easily comparable to the mortality rate data shown in Figure 1a, we calculate a ratio of birth-cohort size in each year to the average county birth-cohort size over the period 1949-1966 (because there are no data on historical county population sizes for time of

[26]In the famine era, each county had approximately five communes (also known as collective farms). However, communes were not an official level of government. We know of no data that can be disaggregated to the commune level.

[27]Note that policies against labor migration caused there to be very little rural migration between when the famine occurred and when survivor cohort sizes are measured in our data (West and Zhao, 2000). To check that the birth-cohort size of survivors is a good proxy for famine mortality, we compare survivor cohort sizes and mortality rates at the province-level. We aggregate birth-cohort sizes to the province and year (birth year) level and regress the log of birth cohort size on the log of mortality while controlling for the log of total population and year and province fixed effects. The correlation is -0.28 and statistically significant at the 1% level.

the famine), and assume that the latter is highly correlated with historical county population size. As with the mortality rate data, we normalize the variance of this variable by its mean. These estimates are plotted in Figure 1c, which clearly shows a simultaneous drop in cohort size and an increase in its variance for the famine years.

Using birth cohort data at the county-level, we can also show that there is significant variation in famine severity across counties within provinces. See Online Appendix Section F and Table A.2.

5 Spatial Correlation between Famine Severity and Productivity

5.1 Measurement of Province-level Productivity

In this section, we document the empirical relationship between mortality rates and agricultural productivity across rural areas. We find a positive correlation between mortality and productivity that holds only during the famine years. This correlation holds both at the province-level and the county-level.

The NBS provides a measure of grain production at the province-level. Our main concern with this measure is the possibility that statisticians were unable to fully correct for government misreporting of production during the GLF. In that case, the association between productivity and mortality rates during the famine would be confounded by misreporting.

To address this concern, we construct a time-varying measure of province-level production that is unlikely to be affected by government misreporting by estimating a production function using data from non-GLF years (1949-57, 1962-82).[28] To estimate the production function, we regress production for province p in year t on the following production inputs: temperature and its squared term, rainfall and its squared term, grain suitability and its squared term, rural population and its squared term, total land area and its squared term, and all combinations of the double interactions of temperature, rainfall, suitability, rural population and total land area. The production function regression has an adjusted R-squared of 0.93, which means that the input variables explain 93% of the variation in production.[29] Then, we apply the input data from all years to create a measure of predicted production using the production function, which we will refer to as "constructed production" henceforth.

The only data from the GLF era that are used for the constructed production measure are the weather, total area and population variables that are inputs in the production function. Monthly mean temperature and rainfall are reported by scientific weather stations. Access to these data was restricted

[28]We restrict our attention to the years before 1982 so that our estimates can be easily comparable to the results in Section 6.3.2 which uses the procurement data.

[29]We do not report the production function coefficients because of the large number of regressors and the difficulty in interpreting each coefficient in the presence of interaction effects. They are available upon request.

to Chinese scientists until recently. The suitability measure is a time-invariant index of a region's suitability for the cultivation of the main procurement grain crops in China during the 1950s (rice, sorghum, wheat, buckwheat, barley). The index is produced by the GAEZ model developed by the Food and Agricultural Organization (FAO), and we assume the inputs are those that are similar to what was used in China during the 1950s.[30] The weather and suitability variables were never known to have been manipulated by the Mao-era government. The data for total rural population and total land area are reported by the post-Mao NBS. Since these data are difficult to manipulate and easy to correct retrospectively, their inclusion should not bias the constructed production measure.

Constructed production is similar to reported production. This can be seen in Online Appendix Figure A.1, which plots constructed grain production against reported grain production together with the 45-degree line for each of the four years of the GLF. The fact that constructed and reported production do not line up perfectly on the 45-degree line is consistent with the potential presence of misreporting. However, the two measures are highly correlated for all years. To be cautious, we use the constructed measure of production for the remainder of our analysis. The estimates using reported production data are similar.[31]

We note that an earlier study, Li and Yang (2005), used additional input variables to predict production: the total area sown for agriculture and for grain, irrigation, the amount of agriculture machine power, and fertilizer usage. Since these variables would have presumably been more difficult for Chinese statisticians to correct after the famine, we did not use them as inputs to construct our main production measure. However, we check that our results are robust to their inclusion.[32]

5.2 Province-Level Analysis

With these data, we explore the relationship between food production and famine mortality rates. To provide a precise estimate of the relationship between per capita food production and mortality rates, we pool all of the data together and estimate the following equation.

[30]These are based on fixed geo-climatic conditions and the technologies used by Chinese farmers in the late 1950s (e.g., low level of mechanization, organic fertilizers, rain-fed irrigation). See the Online Appendix for a detailed discussion of the weather and suitability data and the construction of province-level measures.

[31]See Online Appendix Table A.4 for the main results, as well as the previous version of the paper Meng et al. (2010).

[32]Online Appendix Figure A.2 plots our main constructed measure against the alternative measure, where we include these additional inputs for the famine years. The points are all close to the 45-degree line in the figure. Thus, the additional inputs make little difference. We also check that the positive association between grain productivity and famine mortality rates is robust to using this alternative measure of grain production. The results are available upon request.

Like Li and Yang (2005), we predict production with inputs. However, the two studies differ conceptually in that we use production to predict food availability and mortality, where as they use (lagged) food availability to predict inputs and do not examine mortality as a dependent variable.

$$m_{p,t+1} = \alpha P_{p,t} + \beta P_{p,t} \mathcal{I}_t^{Fam} + \mathbf{Z}'_{p,t}\gamma + \delta_t + \varepsilon_{p,t}, \tag{1}$$

where $m_{p,t+1}$ is the log number of deaths in province p during year $t+1$; $P_{p,t}$ is log constructed grain production; $P_{p,t}\mathcal{I}_t^{Fam}$ is the interaction of log constructed grain production and a dummy variable for whether it is a famine year, where $I_t^{Fam} = \{0, 1\}$ is a dummy variable that equals 1 if the observation is of year $t = 1958, 1959, 1960$; $\mathbf{Z}_{p,t}$ is a vector of province-year level covariates; δ_t is a vector of year fixed effects; and $\varepsilon_{p,t}$ is an error term. The vector of covariates in the baseline specification, $\mathbf{Z}_{p,t}$, includes the log total population, which normalizes our estimates so that we can interpret them in "per capita" terms. We also control for the log urban population to ensure that the estimates capture variation driven by rural areas. Year fixed effects control for all changes over time that affect regions similarly and they subsume the main effect for the famine year dummy. To address the presence of heteroskedasticity, all regressions estimate robust standard errors.[33] Equation (1) estimates the cross-sectional correlation between productivity and mortality rates for non-famine years as $\widehat{\alpha}$, and the correlation during the famine as $\widehat{\alpha + \beta}$.

In Table 2 column (1), we show the cross-sectional correlation between productivity and mortality rates during 1953 to 1982.[34] In column (2), we restrict the sample to five years before the famine began until five years after the famine (1953-1965) to show that our results are driven by the years close to the famine and not spurious correlations from long after the famine ended.[35]

In both columns (1) and (2), the coefficient for log constructed grain is negative and statistically insignificant. This means that during normal years, higher production per capita is uncorrelated with mortality rates. In contrast, the interaction terms are positive and statistically significant at the 1% level, which implies that the relationship between production and mortality is more positive during the famine years than the non-famine years. The sum of the the interaction coefficient and the coefficient for log grain, $\widehat{\alpha + \beta}$, is presented at the bottom of the table along with its p-value. They are positive and statistically significant at the 1% and 5% levels.

Columns (1) and (2) use the main nineteen province sample. Columns (3) and (4) use a larger sample that includes the four provinces with a large proportion of ethnic minorities (see Online Appendix Section B). In columns (5) and (6), we present the results using the reported historical grain production

[33] Our results are similar if we estimate wild bootstrapped standard errors that are clustered at the province-level and adjust for the small number of clusters (Cameron et al., 2008). These results are not presented due to space restrictions. The robustness of our estimates to different levels of clustering can also be seen later, when we present the county-level analysis that cluster standard errors at the province-level.

[34] Before 1953, mortality data were not available for all provinces.

[35] The estimates are similar if we use a sample with a longer time horizon (e.g., until 1998).

instead of our main measure of constructed grain production and return to the main nineteen province sample. The estimates are very similar. Due to space constraints, the remainder of the paper only reports results using the constructed grain production measures for the main sample of 19 provinces.

To examine whether our estimates are driven by outliers, we plot the residuals of the regressions in column (1) of Table 2. Figure 2 plots the residuals for the interaction of log constructed grain and the dummy variable for the three famine years. As seen in the regression, the relationship is positive and not driven by outliers.

The results thus far indicate that the positive relationship we observe between grain productivity and mortality rates during the famine period is very robust.

Next, we investigate whether grouping many non-famine years together causes us to capture spurious changes in the correlation between productivity and mortality rates over time. For example, the relationship may be very volatile over time and the interaction term may capture a spurious positive correlation, while the uninteracted grain productivity term averages the positive and negative correlations to produce a statistically zero coefficient.

To investigate this possibility, we estimate the yearly correlation between productivity and mortality rates in the following equation:

$$m_{p,t+1} = \sum_{\tau=0}^{T} \alpha_\tau P_{p,\tau} \mathcal{I}_t^\tau + \mathbf{Z}'_{p,t}\gamma + \delta_t + \varepsilon_{p,t}, \tag{2}$$

where $m_{p,t+1}$, $P_{p,t}$, $\mathbf{Z}_{p,t}$, δ_t, and $\varepsilon_{p,t}$ are the same as previously defined, and where $\mathcal{I}_t^\tau = \{0,1\}$ is a dummy variable that equals one if the observation year t is equal to the year τ, and we no longer control for the grain productivity main effect because we estimate interaction effects for each year. This estimate allows the coefficient on productivity to vary for each year. α_t is the cross-sectional correlation between mortality rates and productivity in year t.

Figure 3 plots the estimated coefficients of $\widehat{\alpha}_t$ and their 95% confidence intervals.[36] The figures show that the correlation between grain productivity and mortality rates is statistically similar to zero for most years (becoming slightly more negative over time), but spikes upwards for the first two of the three famine years. The point estimate for the interaction term of log constructed grain is statistically significant at the 1% level for the year of peak mortality, 1959 (i.e., 1959 production and 1960 mortality). The magnitude of the coefficient is also large for 1958. The results show that the positive association between productivity and mortality rates is unique to the first two years of the famine. Note that

[36]The coefficients and their standard errors are presented in Online Appendix Table A.3.

finding no association between productivity and mortality for the last famine year (1961 mortality and 1960 production) is consistent with the government having begun to respond to the famine in 1960, which would have weakened the relationship between food production and mortality rates.[37]

There are two important caveats for interpreting these results. First, as we show in greater detail in Section 6.2, more productive regions during the famine were also more productive historically. It is possible that historically more productive provinces pursued GLF policies more zealously. This could bias our constructed measure of production. If zealousness in pursuing GLF policies lowered the returns to climate, geography and total labor (the only GLF-era data used to impute production), then our constructed output measure will over-estimate output in productive regions. This, in turn, would cause our estimates of the association between constructed productivity and famine mortality to bias upward the true association between productivity and famine mortality. In Section 5.4, we will show that this is unlikely to be a problem in practice because historical productivity and political zealousness were not positively associated.

Second, one may be concerned that productive regions may have experienced a larger rise in political radicalism during the famine, which increased mortality rates through channels other than grain production. We will address this concern in Section 5.4 by directly controlling for zealousness in our estimate of the relationship between productivity and mortality.[38]

5.3 County-Level Analysis

The province-level analysis faces a few additional limitations. First, we infer that the cross-province variation reflects differences in rural mortality rates because we control for urban population, but the mortality data do not explicitly distinguish between urban and rural areas. Second, the official mortality data, like the official production data, may be measured with error. Finally, there may be variation across rural areas within a province that we cannot observe with the province-level analysis.

We address these concerns by using county-level data that include only agricultural populations. This allows us to check that there is variation across rural areas and control for province fixed effects to

[37] The coefficients reflect the cross-sectional relationship between mortality rates and productivity. Finding no association means that mortality rates are similar across regions of different productivity. Thus, if the government responds to the famine by giving food or allowing black market "garden" agriculture (see the discussion in Section 2), then we will naturally see the link between productivity and mortality rates weaken by the end of the famine.

[38] Note that for brevity, the estimated standard errors presented in the paper do not correct for the fact that the main explanatory variables (log constructed grain production and its interaction with the famine period dummy variable) are estimates. Such a correction makes little difference because the inputs used to predict production explain 93% of the variation in production. For example, when we correct the standard errors presented in this section by instrumenting for log reported production and its interaction with the interactions of the famine period dummy variable and each of the production inputs used in Section 5.1, we obtain nearly identical standard errors and coefficients. Similarly, the F-statistic for the first stage equation in the 2SLS estimate is very large, 39. These estimates are available upon request.

examine whether there is variation within provinces as well as across provinces. This exercise also proxies for famine severity using data that was not potentially manipulated by the Chinese government in order to check that positive association between famine severity and food productivity at the province-level are not driven by government misreporting.

The main dependent variable is the alternative proxy for famine severity – birth cohort size, which is the size of each birth cohort observed in the 1% sample of the *1990 China Population Census.* The main drawback of using birth cohort size to proxy for famine severity is that it cannot capture the mortality rates of the elderly, who are also believed to have suffered high famine mortality rates. Another related issue is that while it is reasonable to assume that famine mortality is negatively correlated with survivor birth cohort size, we do not know the functional form of this relationship. Thus, we can only use the relationship between birth cohort size and productivity to test the robustness of the sign of the relationship between mortality and productivity.

The explanatory variables are measures of annual (time-varying) mean spring temperature, rainfall, and a (time-invariant) measure of the suitability for grain cultivation, which are the disaggregated measures of the variables used in the province-level production function in Section 5.1. We examine these directly (rather than a constructed production measure from the production function) to check that the earlier results are indeed driven by natural conditions. In this exercise, we interpret the measures of natural conditions as proxies of productivity.

For consistency, we restrict the sample to the same 19 provinces used in the province-level mortality rate analysis. The limited number of historical weather stations additionally restricts the final sample, which includes 511 counties. Our sample begins with 1950, the first year for which there is disaggregated weather data, and ends before 1966 to avoid potentially confounding effects of post-famine political events on fertility.[39]

We examine the relationship between survivor cohort size and proxies for agricultural productivity across rural counties. If the interpretation of our main results are correct, then we should find similar patterns over time between natural conditions that are advantageous for production and survivor birth cohort size – i.e., good conditions should have no relationship or a positive relationship with birth cohort size for cohorts not affected by the famine, but a negative relationship for cohorts affected by the famine. Due to space restrictions, we focus on the year-by-year estimates using equation (2).

[39]See the Online Appendix for a detailed description of the weather and suitability data. We exclude the post-1966 period because the Cultural Revolution (1966-76) caused political disruptions that may have delayed fertility. This is less of an issue for the province-level mortality estimates because this revolution was not associated with abnormal mortality rates. The results are similar if we extend the panel to include those born during or after 1966. They are not reported due to space constraints, but are available upon request.

This differs from earlier estimates in that observations are now at the county and year level, and we control for log average birth cohort size to normalize by the population. We also control for province fixed effects to examine whether the same patterns exist within provinces as across provinces. In our regression, we simultaneously include the interactions of year dummy variables with each of the three proxies of productivity: mean log spring temperature, mean log spring rainfall and the suitability for grain cultivation index. The standard errors are clustered at the county-level. The coefficients and standard errors are reported in Online Appendix Table A.6.

Figures 4a-4c plot the interaction coefficients and their 95% confidence intervals. The pattern is consistent with the province-level results. To interpret the correlation, recall that being born in a county that produces high levels of food per capita has two potentially offsetting effects for cohorts born prior to the famine. On the one hand, higher food availability may cause higher fertility and lower mortality rates, which increases the cohort sizes of survivors. On the other hand, these individuals are exposed to a more severe famine at very young ages, which reduces the cohort size of survivors. Similarly, for the cohort born during the famine, the more severe famine faced by women of child-bearing age may reduce fertility. Therefore, finding a negative correlation between cohort size and natural conditions that are good for production implies that the negative effects of famine exposure outweigh the positive effects.

Similarly, being born in a productive region has offsetting effects on survivor birth-cohort size for those born after the famine. On the one hand, survivors living in productive regions likely had access to more food after the famine relative to less productive regions.[40] This could speed the recovery from famine, increase fertility, reduce infant mortality, and thereby, increase survivor cohort sizes. On the other hand, famines of greater severity in productive regions mean that these regions suffer larger population losses, which result in a smaller population base for bearing and rearing children after the famine. The finding that the correlation between survivor birth-cohort sizes and natural conditions are zero or positive for those born after the famine implies that the positive effects outweigh or cancel the negative effects. This is not altogether surprising since individuals of child bearing age suffered the least from famine and are believed to have emerged relatively intact.

These results provide strong support that the insights from our main findings are not driven by measurement error in either historical mortality or constructed production data. Moreover, they also

[40] Recall from Section 2 that many of the GLF policies were lifted in response to the famine. If people can eat what they produce, then there will naturally be more food per capita in regions with good natural conditions. In addition, our hypothesis that the inflexible grain procurement policy caused the spatial patterns between productivity and famine severity predicts that those living in more productive regions have more food per capita during normal years. See Section 6 for details.

show that the same patterns that exist across provinces also exist across counties *within* provinces and motivate us to develop an explanation that is consistent with variation at high and low levels of government administration.[41]

5.4 Controlling for Political Factors

The three years of high famine mortality occurred during the GLF, a four-year period beginning in 1958 when many misguided policies were often carried out with extreme zealousness. The implementation of these policies is thought to be a contributing factor to both the production drop (e.g., Li and Yang, 2005) and the subsequent mortality rates (e.g., Kung and Lin, 2003) during the famine. For our study, it is important to understand whether more productive regions implemented GLF policies more zealously for the two reasons that are discussed earlier at the end of Section 5.2.

We investigate the possibility that GLF policies were pursued more zealously in historically productive regions, which could result us in biasing upward their constructed productivity during the famine. In this case, our constructed grain output measure for the famine-era is likely to over-state true production in the more productive regions because GLF policies reduced the returns to the inputs we use for our production function (climate, geography, rural population size, land area). For example, if GLF policies diverted land away from agriculture, then the returns to total land area would be lower during the famine years than prior years. Similarly, if there was more food wastage from higher participation rates in communal dining halls in more productive provinces, then our constructed grain output would overestimate food availability in rural areas. To examine the likelihood of this concern, we regress average pre-GLF constructed productivity during 1954-57 on proxies for GLF zealousness that past studies have found important for a cross-section of seventeen provinces. The GLF variables are not available for all of the provinces in our sample; for consistency, we restrict the sample to observations for which all of the GLF variables are reported.[42]

First, we follow Li and Yang (2005) and proxy for zealousness with a reduction in the areas sown for grain and an increase in steel production, since the GLF diverted resources from agriculture into manufacturing. The explanatory variables are the average annual change in these variables during 1958-1960. For ease of interpretation, we construct each variable so that an increase in the value corresponds

[41] The results are similar when we do not control for province fixed effects, or when we control for province-year fixed effects to control for time-varying changes across provinces such as political leadership. They are reported in an earlier version of the paper (Meng et al., 2010).

In the Online Appendix Table A.5, we also show that we obtain qualitatively consistent results when we use province-level survivor birth cohort size as the dependent variable in equation (1). Note that the sample size is smaller that the full sample province-level estimates because these estimates are restricted to 1950-1965 and because the county-level data do not include Shanghai (there were no rural counties in the Shanghai municipality which contained a weather station).

[42] Our results are similar if for each estimate, we use the maximum sample size. These results are available upon request.

to an increase in zealousness. Thus, the variable for sown area growth is actually the sown area growth multiplied by negative one.[43] Table 3 columns (1)-(2) show that the coefficients for the negative growth in per capita area sown and the positive growth in per capita steel production are both statistically insignificant.

Second, we proxy for zealousness with the participation rate in communal dining halls (Yang, 1998). By the eve of the famine, almost all workers ate in communal dining halls. Past studies, such as Yang (1998), have argued that communal dining participation rates during the mid-1950s, before the famine, can be used as a proxy for GLF zealousness. Thus, we use the data reported by Yang (1998) on GLF communal dining participation rates across provinces during the years before the famine to calculate the average participation rate. Column (3) shows that participation rates are negatively correlated with productivity, which means that more zealous provinces were less productive. The estimate is statistically significant at the 10% level.[44]

Third, we use Kung and Chen's (2011) proxy of zealousness, the magnitude of the Anti-Right purges that occurred in 1957, as an explanatory variable.[45] Column (4) shows that this variable is also negatively correlated with productivity and statistically significant at the 10% level. Again, this means that productive provinces were less zealous.

Fourth, we proxy for zealousness with a dummy variable for whether a province was "liberated" by the Communist Party after the official national liberation date of October 1, 1949. Studies such as Yang (1998) and Kung and Lin (2003) find that provinces that were liberated after the national liberation date were more likely to appoint politically zealous leaders. As with the other proxies of zealousness, the estimate in column (5) shows that being liberated after the national date is negatively correlated with productivity. The estimate is statistically insignificant.[46]

In column (6), we include all of these controls in one regression since they may be correlated. When we do this, the signs of the coefficients are similar, but none of the covariates are statistically significant. To address the possibility that the lack of statistical significance may be due to the large

[43]We use the same NBS data as in Li and Yang (2005).

[44]Yang (1998) does not report dining hall participation rates for the urban municipalities. To maximize sample size, we expand this sample by assuming that all workers in the three municipalities ate in communal dining halls, which is motivated by the facts that almost all non-agricultural workers ate in dining halls and that the municipalities were very urbanized.

[45]We use the measure of the percent of the total population that is purged (e.g., total purges divided by total population). We thank the authors of this paper for generously sharing their data with us. In their paper, they found that in the cross-section, the number of purges in 1957 is correlated with higher grain procurement during the famine.

[46]We expand the data sample of these previous studies to cover all provinces in our study by collecting liberation dates and months from a Chinese publication with the title that translates into the *Report on the Liberation of the Peoples Republic of China*. It is housed in the archives of the National Library in Beijing. The results are similar with other measures of liberation date (e.g., number of months before or after the national liberation, rank of liberation date). Chinese reference names are available upon request.

number of explanatory variables used in a relatively small sample, we alternatively control for the first principal component of the five GLF proxy variables instead of the individual GLF proxies in column (7). The coefficient for this parsimonious control is negative in sign and statistically significant at the 1% level. Taken together, the estimates in Table 3 show that if anything, GLF policies were pursued less zealously in more productive regions. If not accounting for GLF policies biases our constructed output measures, then the latter understates production in historically productive regions relative to historically unproductive regions; and our results on the positive association between productivity and mortality during the famine *understate* the steepness of the true positive relationship.

Next, we address the concern that the positive association between famine mortality and productivity is driven by omitted political variables which affect mortality through channels other than grain. Given the finding that less productive regions pursued GLF policies more zealously, one may be concerned that these regions received more favoritism from the central government and received earlier grain relief (a variable that we do not observe) and therefore suffered lower mortality. We have never seen evidence of this type of favoritism. Nevertheless, we can check that our estimates are robust to this and other types of omitted variable bias by directly controlling for GLF proxies in equation (1).

To maximize power, we first control for the principal component of the GLF factors. Note that since we do not have GLF variables for all of the provinces in our sample, the number of observations is reduced for this exercise. Column (1) of Table 4 presents the baseline estimates on this restricted sample. The estimates are very similar to the results with the full sample. Column (2) shows that adding the GLF principal component measure and its interaction with the famine dummy into the main specification does not change our main results. The interaction of constructed grain production and the famine dummy variable continues to be positively correlated with mortality. The interaction of the GLF variable and the famine dummy is statistically insignificant.

To evaluate the quantitative importance of the positive relationship between grain productivity and the rise in mortality during the famine, we present standardized coefficients in square brackets. In columns (1) and (2), the standardized coefficients for the interaction of productivity and the famine dummy show that a one standard deviation change in productivity is associated with an approximately 0.50 standard deviation rise in mortality rates during the famine. Thus, the positive association between productivity and famine morality rates is quantitatively large. In contrast, column (2) shows that a one standard deviation change in the interaction of the famine dummy variable and the GLF principal component indicator only corresponds to a 0.02 standard deviation in the dependent variable. Hence, even if the interaction between the GLF principal component and the famine dummy were statistically significant, we would still conclude that the positive association between food productivity and mortality rates is more quantitatively important for understanding mortality rates than the relationship between GLF factors and famine mortality rates.

In column (3), we check that the limited effect of the GLF in column (2) is not somehow an artifact of the way that we constructed the principal component measure by instead controlling for each component of the composite measure and its interaction with the famine dummy. The interaction of the famine dummy variable and the late liberation dummy variable is almost significant at the 10% level. However, the standardized coefficients show that the explanatory power of late liberation is quantitatively much smaller than that of grain productivity.

Moreover, a comparison of the standardized coefficients for the interactions of the the GLF variables and the famine dummy with the interaction of productivity and the famine dummy shows that all else equal, even if all of the estimates were statistically significant, the GLF factors would be quantitatively much less important than grain productivity for explaining famine mortality.

Since the separate controls do not change the main message of the results, we return to the more parsimonious component measure of GLF intensity for the rest of the analyses.[47]

In addition to the GLF factors, we also consider the role of urban-bias in driving food procurement, as proposed by Lin and Yang (2000). They argue that high government procurement was partly motivated by the government's commitment to feeding the urban population, even if it came at the cost of under-feeding the rural population. To examine this conjecture, we add a control for the interaction of log urban population and the famine period dummy variable to equation (1) (recall that the main effect of log urban population is already included in the main equation). Consistent with Lin and Yang (2000), the correlation between the interaction of urban population and the famine dummy variable in column (4) is positive in sign, but it is statistically insignificant. In column (5), we control for GLF intensity together with urbanization. As before, we find that our main results are very robust.

[47]In Online Appendix I, we show that migration and food wastage are unlikely to confound our results. Note that we also collect a large number of other potential proxies for political zealousness such as the percentage of the population in the late 1950s that were Communist Party members, party cadre density (number of cadres/number of total party members), age of party officials, the number of years that the provincial secretary and governor had been in office, provincial party secretary fixed effects, whether the provincial secretary or governor were purged during the Cultural Revolution later (many of those who were seen as to have resisted Mao during the GLF were purged during the 1966-76 Cultural Revolution), distance from Beijing, distance from the Long March route, distance from the coast, distance to a border, or the negative growth in agricultural population (relative to total population). The inclusion of these variables do not change any of our results, and most importantly, do not change the finding that productivity explains much more variation in famine than any proxy for GLF zealousness. Due to space constraints, we focus on the variables that have received the most emphasis in recent studies and do not present or discuss additional controls. The results are available upon request.

We conclude that the positive spatial relationship between productivity and famine mortality cannot be attributed to the political factors that have featured in existing studies of the causes of the Chinese Famine. Moreover, we document that conditional on political factors, the spatial variation in productivity is quantitatively important for explaining the variation in mortality.

6 The Inflexible Grain Procurement Policy

This section proposes an explanation for unequal rural food distribution that is consistent with the empirical facts presented in the previous sections. First, we document that the historical grain procurement policy was inflexible and progressive and could not easily respond to local aggregate shocks. Then, we use a stylized example to illustrate how such a procurement policy can cause the spatial patterns of famine severity and productivity observed in the data. Finally, we use the example to highlight the connection between mortality and the production gap, which is the difference between target (expected) production and actual production. This relationship allows us to quantify the approximate contribution of our proposed mechanism to overall famine mortality.

6.1 Grain Procurement System in the late 1950s

The grain procurement system in the late 1950s was progressive, in the sense that procurement rates were higher in more productive regions. This mechanically followed from the fact that the central government aimed to procure all grain produced above and beyond what was necessary for sustaining rural populations and future production (e.g., seed). This policy set each location's procurement level in advance according to a location-specific production target, and it could not easily adjust this procurement level following the drop in production. More specifically, production and procurement targets were set at the beginning of the year (Walker, 1984: ch. 2). Procurement was implemented top-down. The central government set procurement targets for the provinces and "the provincial party secretary divided the provincial target among the different prefectures" (Oi, 1991). Similarly, each prefecture would divide the target across counties.

The methods for determining procurement levels for each location are outlined in the *Three Fix Policy,* which based procurement targets on verified past production and estimates of subsistence needs.[48] Procurement occurred after harvest in October and November, after which the central government

[48]In 1956, this policy stipulated that to "fix" procurement levels for each collective, expected local production levels in 1956 should be based on past production, and subsistence levels of consumption and seed retention should be based on population and production needs. See Johnson (1998) for a discussion of the food procurement system. Historical grain policies are outlined in public government archives. See http://2006.panjin.gov.cn/site/gb/pj/pjjz/pjjz_detail.php?column_id=2382.

would measure the harvest to update production estimates. The most recent verified estimates were then used to determine procurement for the following year. Since it often took more than twelve months for the government to accurately account for production (Walker, 1965: p. 82), targets were, in practice, based on production from two or more years before. Collectives delivered the targeted amount of grain even if they were left with too little for their own consumption.[49]

The delays in information gathering imply that the system could not respond to shocks in an accurate or timely fashion. For example, reports of the government's response to regional production shortfalls in 1954 found that relief was made, but not in proportion to the regional decline in output (Walker, 1984, p. 48).[50] The long delays in gathering and responding to information are not surprising given the limited bureaucratic capacity of China in the late 1950s. At that time, virtually all decisions regarding procurement were made centrally by a Standing Committee of approximately seven individuals (Fairbank, 1987: p. 297-341; Spence, 1990: p. 542). Information on the effectiveness of policies was collected locally, aggregated by the regional government, and then eventually reported upward to the Standing Committee (Fairbank, 1987: p. 297-341; Spence, 1990: p. 542). Information collection and policy response proved challenging given China's massive geographic size and poor communication and transportation infrastructure, and this led to significant delays and inefficiencies.[51]

It is interesting to note that official procurement data show that aggregate procurement increased rapidly in the years leading up to the famine, both in levels and as a percentage of production. This pattern is consistent with the belief that the central government increased their procurement ambitions as it obtained more accurate information on the production capacity of each region. This would also result if local officials increased reported production figures in the years leading up to the famine.

These fundamental limitations in bureaucratic capacity were greatly exacerbated by various historical factors in the late 1950s. First, for budgetary reasons, the bureaucracy was understaffed (Eckstein, 1977: p. 186). Second, the Anti-Right purges in 1957 sent many able bureaucrats to re-education in 1958, reducing the number of statisticians and demographers available to project production figures in

[49]Contemporaneous and retrospective survivor accounts show that peasants and local leaders believed that the government, which rose to power with promises of ending famines, would replenish rural grain supplies once they were depleted. For example, Thaxton (2008: p. 109) provides examples of how peasants and village leaders recalled the government's commitment to prevent food crises and examples of food assistance prior to 1959. He also provides examples of how local leaders delivered grain based on a combination of faith in the new regime and its promise to deliver them from privations and a fear of punishment (Thaxton, 2008: p. 117).

[50]These delays also implied that procurement targets were often revised during the spring and summer months to adjust for updated estimates of production from the past year.

[51]Transportation networks were almost completely destroyed by decades of civil unrest and the war with Japan, and repairs had only recently begun (Fairbank, 1987: p. 278). Urban centers were relatively few and geographically concentrated, and it could take many weeks to reach an outlying collective. Moreover, rural areas were typically not connected by telecommunications infrastructure.

1959 (Spence, 1990: p. 580). With diminished statistical capacity, the central government had to wait longer than usual to know the true production figures.

Third, political tension during the GLF era led to an environment of suspicion and further reduced communication within the government. Mao and his followers were unwilling to believe reports of low production from local leaders (Dikotter, 2010; Thaxton, 2008; p. 193-8). This reduced the effectiveness of government, and was further exacerbated by Mao's attempts to solidify his power by reducing the frequency of Standing Committee meetings from twice a week to six times a year and by removing decision-making power from local governments (Fairbank, 1987: p. 303). The political resistance against believing that GLF policies could potentially reduce production is especially important for understanding why the famine became so severe and persisted for so long. The only high-level leader known to have voiced concerns about the possibility of famine was Peng Dehuai, China's defense minister at the time. If the other central leaders had believed and responded to the concerns he expressed at Lushan in July 1959, it could have reduced procurement in 1959 and mitigated famine mortality rates.[52] However, procurement in 1959 did not take Peng's reports into account. This was partly because many did not believe that there were production falls or increased mortality rates and argued that if these were true, they were due to lack of enforcement of the collective system rather than high procurement levels. It is possible that party leaders were genuinely over-optimistic that production would continue to grow.

Aside from limited bureaucratic capacity, one can argue that another reason for the inflexibility of the system is that it helped the central government limit problems of asymmetric information. Relative to the central government, local leaders had much better information on the level of production of their collective, but also had incentives to misreport production to the central government. Historically, local leaders often under-reported production in order to increase their support among local peasants, to retain the unreported grain for personal profit, and to suppress the central government's expectations of future harvests (Thaxton, 2008: p. 293-312).[53] Local leaders were also known to have over-reported production when they felt pressured to meet procurement targets or when promoting themselves politically by meeting or exceeding yield expectations. For the central government, both types of misreporting were problematic: under-reporting reduced government procurement while over-reporting risked government over-procurement and political instability in the countryside, which had

[52] See the Online Appendix for a more detailed discussion of Peng Dehuai.

[53] The government made extensive campaigns against under-reporting. The fact that this was a prevalent problem in the years leading up to the famine is consistent with under-reporting by local officials being included in a list of "ten evils" that was published in daily newspapers in the late 1950s (Walker, 1984: p. 65). For example, see *Jilinrijipao* [*Jilin Daily Newspaper* in Chinese], 1957 December 20th.

provided the core support for the Communist party.[54] As such, there was an advantage for the central government in preserving a system in which local procurement levels did not solely rely on local reports. Clearly, the incentives to misreport production would be minimized if the central government was committed to procuring a fixed target.[55]

6.2 A Stylized Example

We construct a stylized example to explain how an inflexible procurement policy can lead to a famine in which more productive regions suffer more. Before constructing the example, it is useful to describe the spatial pattern of the drop in production during the famine. We show that more productive regions typically experienced a larger absolute production drop while still remaining more productive relative to less productive regions. Figures 5a-5c plot the current year's per capita constructed production for 1958-1960 against the previous year's. The figures show that there is little change in production ranks over time. Thus, provinces that were the most productive in 1957 were still the most productive in 1958, 1959 and 1960. Figures 5d-5f plot the annual growth in per capita production against the previous year's per capita production for the three famine years. They show a negative correlation for each year, which means that the production growth was smaller for regions that produced more food per capita in the previous year.[56] Thus, the data show that productive provinces typically suffered larger falls in production during the famine, but not so much as to reduce their productivity rank.

Table 5 considers the implications of such a production drop together with an inflexible procurement policy. It provides an example of a country with three hypothetical regions: two rural regions (A and B) and a city. Each region has a similar population and therefore similar subsistence needs. For simplicity, we assume the latter to be 100 tons of food. However, agricultural endowments differ across regions. A is better endowed than B and hence produces more food per capita. The city produces no food. As was the case in China, the government restricts population movements so that individuals from low production regions cannot move to high production regions.

There are two states of the world. In the normal state, which occurs with 80% probability, A produces 250 tons and B produces 170 tons. The second state is caused by an aggregate shock which

[54]The government's concerns about over-reporting is shown by the continued political discussions that local leaders should not implement "commandism" during both the first and second Five Year Plans (1953-57, 1958-62) (Walker, 1965: p. 8). In a meeting in Changchou in 1957, Mao personally warned party members to not commit "Stalin's mistake" in terms of over-procurement, which he believed to have turned Soviet peasants against their government (Walker, 1984: p. 149). For a detailed example, see Thaxton (2008, p. 132).

[55]This idea can be easily formalized in a mechanism design framework in a setting where the planner tries to induce an agent with a preference for misreporting to truthfully reveal the amount produced. See an earlier version of this paper (Meng et al., 2010).

[56]The data exhibit similar patterns if we compare current production to a lagged moving average of production. Also, reported production data exhibit similar patterns. These figures are available upon request.

occurs with 20% probability, and production is 20% lower for each region, reducing production in A and B to 200 and 136 tons.[57] The government expects the aggregate shock to occur with 20% probability. The key feature of this production shock is that it is rank preserving, but at the same time, more productive regions experience larger falls in production. It follows that expected production of A and B are 240 ($250 \times .8 + 200 \times .2$) and 163 ($170 \times .8 + 136 \times .2$) tons, which sum to 403 tons of total production.

In accordance with the anecdotal evidence that the grain procurement policy was inflexible, the government consistently procures a fixed amount from each region in all states of the world. Since we do not know the precise historical objective function of the Chinese government, we assume for simplicity that the government chooses procurement to equalize expected consumption across its citizens, giving each of the three regions 134 tons of expected consumption. Thus, the government consistently procures the difference between expected production and expected consumption in all states of the world, taking away 106 (240-134) and 29 (163-134) tons from A and B and giving the city a subsidy of 134 tons. Actual consumption is not constant across states of the world since it equals the difference between actual production and procurement. Therefore, in the good state, A and B consume 144 (250-106) and 141 (170-29) tons of grain; in the bad state, they consume 94 (200-106) and 107 (136-29) tons. The city always consumes 134 tons.

This stylized example illustrates a procurement policy that is inflexible, since procurement does not respond to contemporaneous production, and progressive, since procurement is always a larger share of production in the more productive regions. The key insight is that when there is a fall in production that is larger in magnitude in more productive province (but not large enough to change productivity ranks across provinces), the inflexible and progressive procurement policy can cause high levels of famine mortality even when average rural retention is sufficient for avoiding mortality. Since procurement targets are set for each locality, our explanation is consistent with the observation that famine severity varies across counties within provinces as well as across provinces. More importantly, the example provides a clear illustration of how the government would have over-procured more in the more productive regions during the famine, even though this would not have been true during normal years. Mechanically, the absolute drop in production is larger for the more productive region, implying a larger gap between actual and expected production, resulting in over-procurement from the more productive region.

[57] We assume a proportional drop in production in this stylized example for simplicity since it produces a similar pattern to what we see in the data: more productive provinces would suffer a larger magnitude drop in production, but productivity ranks across provinces will not change.

There are several interesting points to keep in mind. First, the main insight from our example of how more productive regions have less food for consumption when there is an aggregate shock to production is not unique to our assumed objective function. For example, if the government's objective is to equalize consumption across rural regions in the normal state of the world, there will still be a negative correlation between productivity and consumption in the presence of a negative shock. A second and related point is that the correlation between productivity and consumption during normal years need not be strongly positive. For instance, in the previous example with the alternative objective function or in the case that the probability of a negative shock is near zero, there will be little or no correlation between productivity and consumption during normal years, which is what we observed in the data earlier.[58]

Finally, note that our hypothesis is consistent with accounts that some local bureaucrats historically exaggerated production, which could cause the government to over-estimate expected production if the government's post-harvest audits do not fully account for the exaggeration. This bias would exacerbate over-procurement and further reduce aggregate consumption. Moreover, if the level of exaggeration was positively correlated with actual production, then exaggerated production from previous years could also exacerbate the difference in consumption between productive and less productive regions. Thus, inflexibility and historical over-reporting are complementary explanations. The key point for our hypothesis is that inflexibility is still necessary for generating the observed spatial patterns because a flexible procurement system would have avoided over-procurement by adjusting procurement levels after the harvest was realized.[59]

6.3 Quantitative Evidence

6.3.1 Estimation of the Mechanism

A prediction of our stylized example is that mortality rates in rural regions are increasing in the difference between target (expected) production and actual production. More formally, let $P_{p,t}$ correspond to realized (actual) per capita production in region p in year t. From our description of the procurement system, food retention in region p in year t is realized production, $P_{p,t}$, minus procurement, and pro-

[58] For interpreting our empirical analysis, which use data for mortality rather than consumption, an additional reason for observing no correlation during non-famine years is that the relationship between food consumption and mortality is highly non-linear such that once a certain level of consumption is achieved, marginal consumption does not correlate strongly with survival probabilities.

[59] In a previous version of the paper, we present a formal version of the model and show that the constrained optimal policy of a central planner with utilitarian preferences will cause higher over-procurement in regions that are more productive when there is a fall in output that is broadly proportional across regions. We also evaluate the tradeoffs between a policy that targets quantities versus an alternative policy that targets prices (Meng et al., 2010). Due to space constraints, the formal model is excluded from this version.

curement equals the difference between expected production per capita, $\hat{P}_{p,t}$, and some baseline level of per capita food consumption common to all regions. Therefore, regional food consumption is negatively correlated with the regional per capita difference between target and realized production, $\hat{P}_{p,t} - P_{p,t}$, which we refer to as the *production gap*. In principle, this characterization of the correlation between per capita food consumption, and thereby, mortality rates, and the per capita production gap should apply to all years around the time of the famine that used a similar procurement policy.

In this framework, the case where the government could flexibly adjust procurement to give all rural regions the same baseline consumption is equivalent to the government procuring the difference between realized production (as opposed to expected production) and baseline food consumption. Since aggregate production was above consumption needs, a flexible procurement system is effectively equivalent to a production gap of zero, where all rural regions are allocated their baseline consumption.[60]

We construct a measure of expected production at date t. Based on the discussion earlier, we assume that the most recent audited production figures available to the government are from two years ago, $t-2$. To this production level, we assume that the government then applies the three-year moving average of the growth rates from two, three and four years ago for each province in order to estimate expected production $\hat{P}_{p,t}$.[61]

Our mechanism implies that when target per capita production exceeds actual per capita production (i.e., as the production gap increases), per capita retention declines and mortality rates increase. To examine whether this is true in the data, we use our time-varying observations of mortality and production at the province-level to estimate:[62]

$$m_{p,t+1} = F(\hat{P}_{p,t} - P_{p,t}) + \Gamma(urban_{p,t}) + \varepsilon_{p,t} \qquad , \tag{3}$$

where the log mortality rate in province p in year $t+1$, $m_{p,t+1}$, is a function, F, of the production gap $\hat{P}_{p,t} - P_{p,t}$; and a function, Γ, of the share of urban population $urban_{p,t}$. Since we wish to focus on a period that used a similar grain procurement policy as the famine era, we restrict the sample to the

[60] Because we do not have precise data on the distribution of procured food to urban centers, our framework cannot take into account that a zero production gap for all provinces means that there would be lower aggregate procurement, and consequently less food to deliver to urban centers. Nevertheless, our previous accounting exercise and the evidence from previous research suggests that as a whole, China had enough food to avoid mass starvation. This means that, theoretically, a famine would have been avoided under a zero production gap.

[61] More formally, $\hat{P}_{p,t} = \left(1 + \frac{1}{3}\sum_{\tau=2}^{4} g_{p,t-\tau}\right)^2 P_{p,t-2}$, where $P_{p,t-2}$ is the per capita production in province p from two years ago and $\frac{1}{3}\sum_{\tau=2}^{4} g_{p,t-\tau}$ is the average of growth rates over the three years prior to $t-2$. Our results are robust to alternative ways of projecting production, such as using national production growth, or using a four year moving average growth rate. These results are available upon request.

[62] Note that because of data limitations, this exercise does not take into account within-province production gaps and the redistribution of food within provinces.

years around the famine.[63] Since target production relies on five years of past production data, the first year that this variable is available for is 1954, four years prior to the start of the famine. For symmetry, we thus choose our sample to include four years each from the pre-famine and post-famine periods, such that the sample covers the years 1954 to 1964.[64] Note that the true structure of the function F is unknown. Thus, we allow F be a flexible step function defined over intervals of production gaps. We divide the data into nine groups of per capita production gaps.[65] For simplicity, we allow the function, Γ, to be linear.[66]

First, we test our main prediction that mortality was positively correlated with the production gap by regressing mortality rates on the nine production gap interval dummy variables. We estimate this regression with no constant. The coefficients and standard errors are shown in Table 6 column (1). The coefficients are also plotted in Figure 6a. The finding that the coefficients are positive and broadly similar in magnitude for the five lowest groups, where the gaps are negative or zero (e.g., realized production equals or exceeds target production), is consistent with the fact that death can occur for many reasons (e.g., natural causes). That the coefficients increase for the groups with positive gaps (e.g., realized production is below target production) is consistent with mortality increasing when the gap is positive. In other words, a positive production gap leads to "excess" mortality. The fact that the coefficients are roughly increasing with the magnitude of the production gaps when the gap is positive suggests that higher gaps lead to more mortality.

To verify that this pattern is not confounded by the influences of GLF intensity, we add the first principal component of the GLF as a control. Note that the sample size is reduced because we do not have GLF proxy variables for all provinces. However, the estimates shown in Table 6 column (2) are almost identical to those in column (1). In column (3), we further add year fixed effects to restrict the variation driving the coefficients to be cross-sectional. The relationship between mortality rates and the production gaps remain similar to before: they are roughly increasing in the production gap.

[63]This allows us to estimate the function F under the assumption that the baseline level of food consumption which applies to rural regions is not changing over this time period.

[64]Our results are not sensitive to changing the sample cutoff. Results using alternative samples are available upon request.

[65]Online Appendix Figure A.3 shows that we have very few values for the extreme values. Thus, each of the nine groups covers a production gap with a range of twenty kilograms per person, with the exception of the groups that contain the lowest and highest production gap values: the lowest group contains production gaps of -70 kg/person or less and the highest group contains gaps of more than 70 kg/person. Our results are robust to using more or fewer intervals as long as there are sufficient observations per interval. Our results are also similar if, for symmetry, we restrict the sample to exclude observations with production gaps of more than 100 kg/person or less than -100 kg/person. Similarly, they are unchanged if we use alternative specifications that replace the step function with a polynomial function. These results are not presented due to space restrictions.

[66]We could similarly allow the function, Γ, to be piece-wise linear. However, when we do this, we find that the relationship is roughly linear. Thus, we use a linear measure in the estimation.

Next, we examine the implication of the model that a production gap of zero, which would result from a fully flexible procurement policy, leads to much less famine mortality. This is obviously true when we apply the regression estimates from Table 6 column (1). The coefficient for the production gap group that contains the zero value (e.g., Group 5: -10 to 10 kg/person) is 13.6, the coefficient for the share of urban population is -8.23. The population-weighted urban population share of the provinces in our sample is 16.5%. Thus, the predicted mortality rate with zero production gap and the sample urban population share is 10.26 $(13.6 - 8.23 \times 0.165 = 10.26)$. This is similar to the 10.66 average mortality rate in the data for the three years right before the famine (see Figure 1a). The results are similar if we use the estimates from columns (2) and (3), where we control for GLF factors and year fixed effects. Thus, our empirical results are consistent with the prediction that a fully flexible procurement policy would not have generated excess mortality during the famine.

Finally, we attempt to provide a crude quantification of the aggregate impact of our mechanism for the famine. The *benchmark total mortality* in a given year is the mortality rate predicted by the econometric model when there is zero production gap in all provinces. This estimates the level of mortality under a flexible procurement system. *Predicted total mortality* from the model in a given year is the sum of the predicted mortality across provinces. For each province, this is the product of the predicted mortality rate under that province's production gap that year and the province's population that year. *Predicted excess mortality* is therefore the difference between the model predicted total mortality and the benchmark total mortality for the three years of famine 1959-1961. *Reported excess mortality* is the difference between the reported total mortality (the sum of all mortality across provinces in reported in the data) and the benchmark total mortality for 1959-1961. To assess the contribution of the hypothesized procurement policy to total excess mortality generated by the famine, we divide the predicted excess famine mortality by the reported excess mortality.

Based on the model presented in column (1), this is 42.6%. If we use the model in column (2) that controls for GLF intensity, the contribution of the inflexible procurement policy is 32.4%. If we use the model in column (3) that controls for GLF intensity and year fixed effects, then the contribution of the inflexible procurement policy increases further to 43.2%.

The estimated relationship between mortality and the production gap also illustrates the importance of progressiveness. Consider the estimated mortality in a counterfactual scenario in which all provinces are subject to the average production gap for the famine period, 14.9 kg/person.[67] Since procurement

[67] The calculation of the national average per capita production gap is similar to the procedure we described for calculating the per capita production gap for each province, except that we replace lagged province-specific production and growth rates with lags of aggregate production and growth rates.

was in practice proportional to the production gap, this counterfactual scenario results in slightly less procurement from the more productive regions and slightly more from the less productive regions, making the system less progressive. The regression estimates in column (1) show that the coefficient for the sixth production gap category, which covers the gaps that range from ten to thirty kg/person, is 12.6, which is very similar to the coefficient of the production gap category that includes zero production gap (the fifth production gap category 13.6). This means that the progressiveness of the procurement system is important for the large quantitative effect of the system's inflexibility on famine mortality.[68]

In the Online Appendix, we present the analogous results with reported production data. Using these data, we find that the procurement mechanism explains 40-43% of total excess famine mortality.

6.3.2 Regional Procurement

We have thus far assumed that procurement is driven by past production and that the positive association between the food production and mortality rates are driven by procurement. In this section, we provide evidence for both of these assumptions using regional procurement data, which are available for the years 1953-82. The procurement data report gross procurement and the amount that is "sold back" to provinces. The latter category includes food given back to rural areas for production the following year (e.g., organic fertilizer, seeds) as well as food given back in the event of a verified shock to production. We use these two measures to calculate net procurement. We then calculate the procurement rate as the ratio of net procurement to production.

As with government-reported production data, one may be concerned that the government underreported famine-era procurement in order to minimize its responsibility for the famine. In particular, to mitigate government culpability, the government would underreport procurement more for regions that were more productive and suffered higher mortality rates during the famine. Unlike the production data, we do not have a way for correcting for such measurement error. For these reasons, we cautiously interpret the findings in this section as supplementary evidence.

To examine the relationship between past production and current procurement, we examine both the two year lag of per capita production and a three year moving average of the two, three and four year lag of per capita production, always controlling for year fixed effects. Table 7 columns (1) and (2) show that reported procurement in year t increases by approximately 0.64 to 1.24 kg/person when production per capita two years prior increased by 1kg/person. The coefficients are statistically significant at the

[68] This exercise is equivalent to treating rural regions together as a single entity, which is in the spirit of the accounting exercise of Section 3. Our results here are consistent with those of the accounting exercise which suggest that inequality in food consumption across rural regions was necessary for the famine.

1% level. Columns (3) and (4) show that reported procurement in year t increases by approximately 0.68 to 1.25 kg/person when the moving average of production per capita two, three and four years prior increased by 1kg/person. The coefficients are statistically significant at the 1% level.

Columns (2) and (4) control for the principal component of GLF factors. The coefficients, taken literally, say that as past production increases by one kilogram per person, procurement in the current year will increase by 0.65 to 0.67 kilogram per person. That this association is positive and large in magnitude supports our claims. That this is the case after controlling for GLF factors goes against the suspicion that the relationship between past production and current procurement is entirely driven by a spurious correlation with political factors. At the same time, the fact that adding political controls reduces the magnitude of the coefficient is consistent with the belief that political factors mattered.

Next, we document that there is a higher increase in procurement rates in more productive regions during the famine. We estimate equation (2) with the percentage of production procured as the dependent variable. Figure 7 plots the coefficients and the 95% confidence intervals.[69] Consistent with the progressiveness of our hypothesized procurement policy, the estimated interaction coefficients are positive for all years. This implies that procurement rates are always higher in more productive regions.[70] More importantly, the interaction terms increase in magnitude for the famine years, which means that procurement rates during the famine increased more for productive regions. The estimates are very precise. The relative procurement rates across regions and their changes over time are both consistent with our hypothesis.

Third, we document that higher production gaps (as discussed in the previous section) lead to lower per capita food retention, which is defined as the difference between total constructed production and total reported procurement divided by the number of rural residents. We estimate equation (3) with per capita retention as the dependent variable. The estimates are shown in Table 6 column (4). The coefficients are plotted in Figure 6b. They show a clear negative relationship between rural food retention and the gap between government projected production and realized production. The more that government projections exceeded realized production, the less food was retained in rural areas. These results are qualitatively consistent with our mechanism that high production gaps increase mortality by reducing rural food retention. Column (5) additionally controls for GLF factors and column (6) further controls for year fixed effects. The pattern between per capita retention and the per capita production

[69] The coefficients and standard errors are presented in Online Appendix Table A.3.

[70] To see that procurement rates are always higher in the more productive regions under our hypothesis, see Table 5. Procurement rates in Region A and B during a the high production year are $106/250 = 0.42$ and $29/170 = 0.17$. During the low production year, they are $106/200 = 0.53$ and $29/136 = 0.21$. The difference between the two regions during the high production and low production years are thus $0.42 - 0.17 = 0.25$ and $0.53 - 0.21 = 0.31$.

gap is similar with additional controls.

Finally, using a simpler linear specification and the same sample as in Section 6.3.1 (1954-1964), we can also regress per capita retention on a linear measure of the production gap (while controlling for urban population share, GLF factors and year fixed effects). Table 8 column (1) shows that as government projected production exceeded realized production by one kilogram per person, per capita retention declined by approximately half a kilogram. In columns (2) and (3), we divide the data to non-famine and famine years. The estimates show that during non-famine years, a one kilogram increase in the production gap reduced retention by approximately 0.5 kilogram per person. However, during the famine, a one kilogram increase in the production gap reduced retention by approximately 0.7 kilogram per person. The increase in the magnitude of the reduction is consistent with the discussion in Section 6.1 that political factors during the GLF period exacerbated the inherent inflexibility of the centrally planned procurement system during the famine era. The fact that a one kilogram increase in the production gap caused a large reduction in retention also supports the assumption in our quantitative exercise that the famine-era government tried to procure the entire surplus production.

In column (4), we show the reduced form relationship between mortality rates and the linear measure of the production gap. It is positive and statistically significant. This means that as government projected production exceeds realized production, mortality rates increase. This is consistent with our hypothesis.

In column (5), we show the relationship between mortality rates and per capita retention, measured as the difference between constructed production and reported procurement. The estimate is negative and statistically significant. This means that as food retention declines, mortality rates increase. This is also consistent with our hypothesis.

Given that our theory of the cause of the famine focuses on the production gap, it is also interesting to examine the extent to which mortality is due to the component of food retention which is driven by the production gap. Note that this is a crude scaling exercise and we do not assume that the production gap only affects mortality rates only through its influence on food retention. In column (6), we instrument for retention with the production gap.[71] The 2SLS estimate is negative and statistically significant, indicating a negative relationship between mortality and food retention. This is again consistent with our hypothesis that as government projected production exceeds realized production, food retention declines and mortality increases.

We interpret the results in this section as evidence in support of our mechanism. At the same time,

[71] The first stage and reduced form estimates for this 2SLS estimate are shown in columns (1) and (4), respectively.

we conservatively interpret the quantitative magnitudes as illustrative. Note that since mortality is a non-linear function of consumption and there is significant variation in retention levels within a province, the coefficients cannot be interpreted as an elasticity of mortality with respect to consumption.

In the Online Appendix, we present analogous results using reported production data. The findings that retention is decreasing in the production gap and that higher retention reduces mortality rates are similar.

7 Conclusion

During the twentieth century, millions have perished from famine, and over sixty percent of total famine mortality has occurred in centrally-planned economies. The most deadly famine in history was the Chinese Great Famine, which in just a few years, killed up to 45 million individuals. This paper proposes that an inflexible and progressive government procurement policy is necessary for explaining the famine. Other explanations for the famine cannot be easily reconciled with the patterns of rural inequality in famine that we document, and the presence of rural inequality is necessary to generate such a massive famine given high average rural food availability. Our results show that the inflexible and progressive procurement policy contributed to 32-43% of total famine mortality. This means that our mechanism is quantitatively important. At the same time, it leaves much room for the contribution of other factors for famine mortality, such as the political factors that have been previously emphasized by famine scholars. Another way of interpreting our results is to say that absent these other factors, famine mortality could have been 57% to 68% lower.

There are several important points to keep in mind for understanding the role of inflexibility in causing the famine. The inability to aggregate information about true production would not have only caused the the government to over-procure, but it also delayed its ability to respond to famine by sending replenishments back to the affected regions. Given the data limitations, we are unable to distinguish the effect of over-procurement from the effect of delayed replenishments. Both effects are captured in our reduced form analysis and contribute to our back-of-the-envelope calculations of the quantitative contribution of inflexible procurement.

Our results do not mean that inflexible procurement must lead to famine. In the Chinese context, inflexible procurement caused a famine because the government miscalculated grain projections while aiming to procure an extremely large proportion of surplus production. After the famine, perhaps with the realization that some degree of government miscalculation is inevitable, the government lowered the procurement rate. As long as the government was committed to its inflexible procurement regime, lowering aggregate procurement goals was the only policy instrument available for avoiding another famine.

Thus, our theory of the cause of the famine has different policy implications from the theories highlighted in previous studies, which have mostly focused on the zealous pursuit of misguided GLF policies. If one believes that the main contributor was bad GLF policies and political radicalism during the late 1950s, then China would have no more famines once GLF policies were abandoned. In contrast, if one believes that inflexibility was an important contributor, then the Chinese government would have had to take additional measures, or risk another famine (albeit one of smaller magnitude if political radicalism has subsided) the next time there was an unanticipated production shock. The fact that the government permanently reduced procurement rates after GLF policies were abandoned suggests that post-famine Chinese policymakers may have understood the risks inherent to an inflexible system.

It is important to note that the context of our study has many specific institutional features, and at the same time recognize that the inflexibility we describe stems from the fundamental problem of mistrust – farmers, who must give all surplus production to the government, do not have the correct incentives to truthfully report production and local bureaucrats may under- or over-report production depending on whether they curry the favor of their neighbors or the officials in the upper-levels of government. The problem we highlight is therefore a generic problem for any regime where the farmer is not the residual claimant of his production.

The most essential ingredients for our mechanism – the commitment to central planning, the incentives for regional bureaucrats and peasants to misreport production and to shirk, the inability to quickly aggregate and respond to new information – are common to several other centrally planned economies of the twentieth century. For example, we believe that our study provides potentially generalizable insights for understanding the causes of the Soviet Famine during 1932-33, which killed up to 6.5 million people in just one year (Davies and Wheatcroft, 2004). There is a consensus that high government procurement from rural areas caused the famine. Like the Chinese case, the most conservative estimates of rural retention in the most severely stricken Soviet state, the Ukraine, show that average rural food availability after procurement was deducted was approximately 170 kg/person during the worst year of the famine (Conquest, 1987). This provides a diet of approximately 1,671 calories per day. While it is not a rich diet, it is much more than the level needed to avoid the high famine mortality rates experienced by the Ukraine. Also, as in the case of China, the data suggest similar spatial patterns in mortality rates – they are higher in more productive regions.[72]

[72]Conquest (1987) discusses the extremely high mortality rates in the states that produced the most food such as the

These provocative similarities suggest that the role of the inflexible procurement policy in other historical famines like the Soviet Famine is a worthy topic of future research. Similarly, it is also important to develop a more generalized framework for understanding the conditions under which rigid food distribution mechanisms in centrally planned economies contribute to famine. The evidence offered in this study takes a first step in this agenda.

Ukraine and Kazakhstan. A recent study by Sharygin (2011) uses the RLMS data to construct regional famine mortality rates within Russia and finds that the agriculturally rich regions near the Volga suffered the highest mortality rates.

References

Ashton, Basil, Kenneth Hill, Alan Piazza, and Robin Zeitz, "Famine in China, 1958-1961," *Population and Development Review*, 1984, *10*, 613–645.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," Technical Report, National Bureau of Economic Research 2012.

Banister, Judith, *China's Changing Population*, Stanford University Press, 1987.

Becker, Jasper, *Hungry Ghosts: China's Secret Famine.*, John Murray (Publishers) Ltd, 1996.

Burgess, Robin and Dave Donaldson, "Can Openness Mitigate the Effects of Weather Fluctuations? Evidence from India's Famine Era," 2010.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, August 2008, *90* (3), 414–427.

Chang, Gene Hsin and Guanzhong James Wen, "Communal Dining and the Chinese Famine of 1958-1961," *Economic Development and Cultural Change*, 1997, *46* (1), 1–34.

Coale, Ansley J., "Population Trends, Population Policy, and Population Studies in China," *Population and Development Review*, 1981, pp. 85–97.

Conquest, R., *The Harvest of Sorrow: Soviet Collectivization and the Terror-Famine* Armenian Research Center Collection, Oxford University Press, Incorporated, 1987.

Dasgupta, Partha and Debraj Ray, "Inequality as a Determinant of Malnutrition and Unemployment: Theory," *The Economic Journal*, 1986, *96* (384), 1011–1034.

Davies, R.W. and S.G. Wheatcroft, *The Years of Hunger: Soviet Agriculture, 1931-1933* The Industrialization of Soviet Russia, Palgrave Macmillan, 2004.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, "Temperature Shocks and Economic Growth: Evidence from the Last Half Century," *American Economic Journal: Macroeconomics*, July 2012, *4* (3), 66–95.

Demick, Barbara, "The Good Cook: A Battle Against Famine in North Korea," *The New Yorker*, 2009, *November 2, 2009*, 58–64.

Dikotter, Frank, *Mao's Great Famine: The History of China's Most Devastating Catastrophe, 1958-1962*, Walker, 2010.

Dreze, Jean, *The Economics of Famine*, Vol. 101, Edward Elgar Publishing, 1999.

Eckstein, Alexander, *China's Economic Revolution*, Cambridge University Press, 1977.

Fairbank, John King, *Great Chinese Revolution 1800-1985* A Cornelia and Michael Bessie book Volume 1442 of Perennial library, HarperCollins, 1987.

Haggard, Stephan and Marcus Noland, *Hunger and Human Rights: the Politics of Famine in North Korea*, Committee for Human Rights in North Korea, 2005.

Jia, Ruixue, "Weather Shocks, Sweet Potatoes and Peasant Revolts in Historical China," *HiCNWorking Papers*, 2011.

Johnson, D. Gale, "China's Great Famine: Introductory Remarks," *China Economic Review*, 1998, *9* (2), 103–109.

Kinsella, K.G., H. Wan, and United States Bureau of the Census, *An Aging World: 2008* International Population Reports, U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, 2009.

Kung, James and Justin Yifu Lin, "The Causes of China's Great Leap Famine, 1959-1961," *Economic Development and Cultural Change*, October 2003, *52* (1), 51–73.

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

Li, Wei and Dennis Tao Yang, "The Great Leap Forward: Anatomy of a Central Planning Disaster," *Journal of Political Economy*, 2005, *113* (4), 840–877.

Lin, Justin Yifu, "Collectivization and China's Agricultural Crisis in 1959-1961," *Journal of Political Economy*, 1990, pp. 1228–1252.

_ **and Dennis Tao Yang**, "Food Availability, Entitlements and the Chinese Famine of 1959-61," *Economic Journal*, January 2000, *110* (460), 136–58.

Meng, Xin, Nancy Qian, and Pierre Yared, "The Institutional Causes of China- Great Famine, 1959-1961," *NBER Working Paper No. 16361*, 2010.

Nunn, Nathan and Nancy Qian, "The Impact of Potatoes on Old World Population and Urbanization," *Quarterly Journal of Economics*, forthcoming.

OGrada, Cormac, "Making Famine History," *Journal of Economic Literature*, 2007, *45* (1), 5–38.

Oi, J.C., *State and Peasant in Contemporary China: The Political Economy of Village Government* Center for Chinese Studies, UC Berkeley Series, University of California Presson Demand, 1991.

Peng, Xizhe, "Demographic Consequences of the Great Leap Forward in China's Provinces," *Population and Development Review*, 1987, pp. 639–670.

Perkins, Dwight H. and Shahid Yusuf, *Rural Development in China*, Baltimore: Johns Hopkins University, 1984.

Ravallion, Martin, "Famines and Economics," *Journal of Economic Literature*, September 1997, *35* (3), 1205–1242.

Sen, Amartya, *Poverty and Famines: An Essay on Entitlement and Deprivation* 1981.

Sharygin, Ethan, "Stunting and Selection Effects of Famine: A Case Study in Russia," *University of Pennsylvania Working Paper*, 2011.

Shiue, Carol H., "Transport Costs and the Geography of Arbitrage in Eighteenth-Century China," *American Economic Review*, 2002, pp. 1406–1419.

_ , "Local Granaries and Central Government Disaster Relief: Moral Hazard and Intergovernmental Finance in Eighteenth-and Nineteenth-Century China," *The Journal of Economic History*, 2004, *64* (01), 100–124.

_ , "The Political Economy of Famine Relief in China, 1740-1820," *Journal of Interdisciplinary History*, 2005, *36* (1), 33–55.

Spence, J.D., *The Search for Modern China*, W W Norton & Company Incorporated, 1990.

Thaxton, Ralph, *Catastrophe and Contention in Rural China: Mao's Great Leap Forward Famine and the Origins of Righteous Resistance in Da Fo Village*, New York: Cambridge University Press, 2008.

tzu Sun, Wei, "Principles for Organizing Grain Circulation Planning," *Jihuajingji [Planned Economy in Chinese]*, 1958, *29*, 24–7.

Walker, K.R., *Planning in Chinese Agriculture: Socialisation and the Private Sector,1956-1962*, Psychology Press, 1965.

_ , *Food Grain Procurement and Consumption in China* Contemporary China Institute Publications, Cambridge University Press, 1984.

West, Loraine A. and Yaohui Zhao, *Rural Labor Flows in China*, Vol. 42, Univ of California Inst of East, 2000.

Yang, Dali, *Calamity and Reform in China: State, Rural Society, and Institutional Change since the Great Leap Famine*, Stanford University Press, 1998.

Yang, D.L., *Calamity and Reform in China: State,rural Society and Institutional Change Since the Great Leap Famine* Political science. Asian studies, STANFORD University Press, 1998.

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

Table 1: Average Food Availability

	Grain Prod	Total Pop	Rural Pop	Nat'l Proc Rate	Per Capita Production		Per Capita Rural Retention	
	(Millions Tons)	(10000 Persons)	(10000 Persons)	(Proc/Prod)	(Kgs/Yr)	(Calories/Day)	(Kgs)	(Calories/Day)
Year	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1957	133	47699	39735	0.15	279	2745	285	2802
1958	142	48612	39955	0.19	292	2867	289	2842
1959	123	49933	39603	0.24	246	2421	237	2329
1960	106	49576	38937	0.19	214	2101	221	2168
1961	101	49739	40071	0.16	204	2005	212	2080
1962	115	50947	42092	0.15	226	2223	234	2301
1963	123	52366	43279	0.15	234	2303	242	2376

Notes: The sample contains 19 provinces: Anhui, Beijing, Fujian, Guangdong, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Shaanxi, Shandong, Shanghai, Shanxi, Tianjin and Zhejiang. Sources: CDSM50 (1999), CPIRC (2000), the Ministry of Agriculture (1983) and the authors' computations.

Table 2: The Correlation between Constructed (and Reported) Grain Productivity and Mortality Rates across Provinces

	Dependent Variable: Ln Deaths in Year t+1					
	A. Constructed Grain Production				B. Reported Grain Production	
	Main Sample: 19 Provinces		Larger Sample: 23 Provinces		Main Sample: 19 Provinces	
	1953-1982	1953-1965	1953-1982	1953-1965	1953-1982	1953-1965
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Grain x Famine Dummy	0.148	0.103	0.156	0.120	0.137	0.0905
	(0.0415)	(0.0413)	(0.0422)	(0.0424)	(0.0365)	(0.0373)
Ln Grain	-0.00678	-0.00513	-0.0409	-0.0257	-0.0240	-0.00109
	(0.0131)	(0.0122)	(0.0120)	(0.0133)	(0.0250)	(0.0361)
Observations	569	246	689	298	569	246
R-squared	0.954	0.945	0.953	0.940	0.954	0.945
Joint: Ln Grain x Dummy + Ln Grain	0.141	0.0977	0.115	0.0943	0.113	0.0894
p-value	0.00104	0.0255	0.00689	0.0339	0.00436	0.0456

Notes: The famine dummy is equal to one for the years 1958-1960. All regressions control for log total population, log urban population and year fixed effects. Robust standard errors are presented in parentheses. The samples comprise balanced panels of provinces with the exception that there is no mortality data for Beijing for 1953. Columns (1)-(2) and (5)-(6) comprise 19 provinces: Anhui, Beijing, Fujian, Guangdong, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Shaanxi, Shandong, Shanghai, Shanxi, Tianjin and Zhejiang. Columns (3) and (4) additionally include Qinghai, Gansu, Yunnan and Guizhou. Sources: CDSM50 (1999) and the authors' computations.

Table 3: The Correlation between Constructed Grain Productivity and GLF Intensity

	Dependent Variable: Average Ln Constructed Per Capita Grain Production 1954-1957						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(-) Per Capita Sown Area Growth (Avg. 1958-1960)	-754.0					-291.3	
	(532.9)					(862.9)	
(+) Per Capita Steel Production Growth (Avg. 1958-1960)		8.430				20.66	
		(58.84)				(66.87)	
% Participation in Communal Dining (Avg.1955-1958)			-2.000			-0.694	
			(1.079)			(1.480)	
% Purged during 1957 Anti-Right Movement				-16.34		-8.813	
				(8.381)		(14.65)	
Late Liberation Dummy					-94.45	-81.92	
					(59.36)	(77.39)	
Principal Component of Independent Variables in Columns (1)-(6)							-42.47
							(19.28)
Observations	17	17	17	17	17	17	17
R-squared	0.118	0.001	0.186	0.202	0.144	0.348	0.244

Notes: The sample comprises 17 provinces. In column (7), we control for the principal component the GLF variables used in columns (1)-(6). Robust standard errors are reported in parentheses. Sources: see text.

Table 4: The Correlation between Constructed Grain Productivity and Mortality Rates across Provinces – Robustness to Controls

	Dependent Variable: Ln Deaths in year t+1				
	(1)	(2)	(3)	(4)	(5)
Ln Constructed Grain Production x Famine Dummy	0.180	0.197	0.230	0.164	0.181
	(0.0543)	(0.0673)	(0.0813)	(0.0596)	(0.0684)
	[0.495]	[0.542]	[0.633]	[0.452]	[0.498]
Ln Constructed Grain Production	0.000062	-0.0140	-0.0521	0.00147	0.00217
	(0.0145)	(0.0187)	(0.0234)	(0.0143)	(0.0140)
	[0.000087]	[-0.0201]	[-0.0749]	[0.00211]	[0.00311]
GLF Principal Component x Famine Dummy		0.0362			0.0350
		(0.0498)			(0.0506)
		[0.0211]			[0.0204]
GLF Principal Component		-0.00895			-0.00919
		(0.00463)			(0.00461)
		[-0.0268]			[-0.0276]
Per Capita Sown Area Growth (x -1) x Famine Dummy			-0.347		
			(0.663)		
			[-0.0177]		
Per Capita Steel Production Growth x Famine Dummy			0.0418		
			(0.0811)		
			[0.0221]		
% Participation in Communal Dining x Famine Dummy			-0.00151		
			(0.00193)		
			[-0.0471]		
% Purged during 1957 Anti-Right Movement x Famine Dummy**			0.0258		
			(0.0188)		
			[0.0921]		
Late Liberation Dummy x Famine Dummy			0.130		
			(0.0712)		
			[0.0450]		
Ln Urban Population x Famine Dummy				0.126	0.121
				(0.133)	(0.131)
				[0.342]	[0.326]
Observations	483	483	483	483	483
R-squared	0.943	0.944	0.947	0.943	0.944
Joint: Ln Constructed Grain Prod x 1959 Dummy + Ln Constructed Grain Prod	0.180	0.183	0.178	0.166	0.168
p-value	0.00125	0.00886	0.0332	0.00627	0.0212

Notes: The famine dummy is equal to one for the years 1958-1960. All regressions control for log total population, log urban population and year fixed effects. In columns (2) and (5), we control for the principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61. Robust standard errors are reported in parentheses. In column (3), we also control for all of the uninteracted effects of the GLF intensity proxies. **In column (3), the coefficients for anti-right and anti-right x famine dummy are multiplied by 1,000 for presentation purposes. Standardized coefficients are reported in italics and brackets. The sample includes an unbalanced panel of 17 provinces for the years 1953-82 (the number of provinces vary across years due to missing values for the GLF variables, and because there is no mortality data for Beijing for 1953). Sources: See text.

Table 5: A Stylized Example of Grain Procurement

	Region A	Region B	City
Subsistence Needs	100	100	100
Production under High Shock (Probability 80%)	250	170	0
Production under Low Shock (Probability 20%)	200	136	0
Expected Production (0.8 x High + 0.2 x Low)	240	163	0
Expected Consumption	134	134	134
Procurement/Subsidy (Expected Production - Expected Consumption)	106	29	-134
Consumption under High Shock (High Production - Procurement)	144	141	134
Consumption under Low Shock (Low Production - Procurement)	94	107	134

Sources: Author's calculations.

Table 6: The Correlation between the Production Gap and Mortality Rates

	Dependent Variable					
	Number of Deaths per 1,000 in Year *t+1*			Grain Retention (Kg/person) in Year *t*		
	(1)	(2)	(3)	(4)	(5)	(6)
	1954-1964	1954-1964	1954-1964	1954-1964	1954-1964	1954-1964
Government Projected PC Prod - Realized PC Prod is						
Group1: <=-70	13.00	13.34	14.26	151.7	173.6	226.4
	(0.802)	(0.978)	(0.949)	(30.41)	(29.21)	(39.22)
Group2: (-70, -50]	11.85	11.93	12.20	164.2	182.7	230.2
	(0.542)	(0.532)	(1.117)	(34.42)	(36.81)	(42.85)
Group3: (-50,-30]	13.07	13.35	14.14	122.4	142.2	206.1
	(0.602)	(0.674)	(0.734)	(18.84)	(19.03)	(31.54)
Group4: (-30, -10]	12.18	12.40	11.96	84.70	105.7	175.7
	(0.691)	(0.746)	(1.031)	(26.31)	(23.67)	(35.02)
Group5: (-10, 10]	13.60	13.83	14.03	130.0	138.6	197.1
	(1.197)	(1.265)	(0.998)	(21.93)	(22.70)	(32.76)
Group6: (10, 30]	12.60	12.49	12.93	120.8	123.4	179.4
	(0.509)	(0.501)	(0.665)	(19.92)	(19.61)	(33.50)
Group7: (30, 50]	16.33	16.52	14.90	81.97	98.82	164.6
	(3.093)	(3.179)	(2.458)	(25.50)	(24.93)	(32.37)
Group8: (50, 70]	14.36	14.55	15.31	35.15	56.38	113.6
	(0.964)	(0.969)	(1.003)	(31.83)	(29.79)	(37.58)
Group9: >70	14.29	14.70	14.42	31.85	63.05	104.9
	(0.718)	(0.861)	(0.788)	(30.25)	(29.67)	(27.77)
Controls						
GLF PCA	N	Y	Y	N	Y	Y
Year FE	N	N	Y	N	N	Y
Observations	185	175	175	204	193	193
R-squared	0.831	0.828	0.876	0.953	0.950	0.955

Notes: Robust standard errors are in parentheses. All regressions control for urban population share. The control for GLF is the first principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61. The number of observations vary due to missing values for the GLF and mortality variables.
Sources: CDSM50 (1999) and the authors' computations.

Table 7: Procurement and Past Production

	Dependent Variable: Per Capita Procurement$_t$			
	(1)	(2)	(4)	(5)
Per Capita Grain Production$_{t-2}$	1.238	0.643		
	(0.106)	(0.0687)		
3 Year Moving Avg. Per Capita Grain Production$_{t-2}$			1.250	0.677
			(0.104)	(0.0704)
Controls				
GLF PCA, Urban Share	N	Y	N	Y
Observations	228	216	228	216
R-squared	0.396	0.835	0.408	0.838

Notes: Robust standard errors are in parentheses. All regressions control for year fixed effects. The control for GLF is the first principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61.The sample comprises of the years 1953-1964. The number of observations vary due to missing values for the GLF variables.

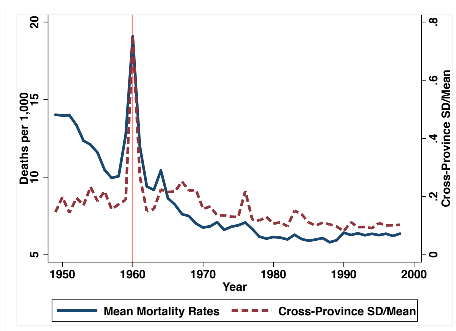
Table 8: Mortality, Grain Retention and the Production Gap

	Dependent Variable					
	Retention (Kg/Person)			Mortality in year t+1		
	(1)	(2)	(3)	(4)	(5)	(6)
	1954-1964	1954-1957, 1961-1964	1958 - 60	1954-1964	1954-1964	1954-1964, 2SLS
Production Gap (Kg/Person)	-0.546	-0.514	-0.720	0.0107		
	(0.150)	(0.168)	(0.244)	(0.00403)		
Retention (Kg/Person)					-0.00235	-0.0196
					(0.000946)	(0.00759)
Observations	193	140	53	175	198	175
R-squared	0.823	0.822	0.859	0.076	0.090	0.028

Notes: Robust standard errors are in parentheses. All regressions control for urban population share; the principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61; and year fixed effects. In column (6), retention is instrumented with production gap. The first stage equation for the 2SLS in column (6) is shown in column (1). The reduced form equation for column (6) is shown in column (4). The number of observations vary between columns (1) and (3) because of missing values in mortality data. There are fewer observations in columns (4) and (6) than column (5) because the production gap is based on government projected production, which is in turn based on data from *t-2*, *t-3* and *t-4* years. Sources: CDSM50 (1999) and the authors' computations.

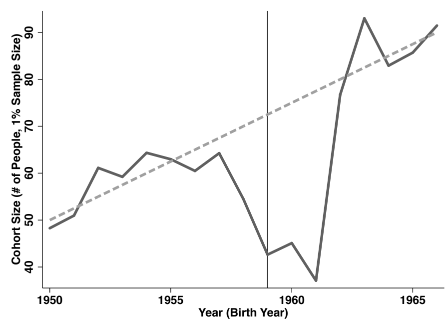
Figure 1: Average and Spatial Variation in Famine Severity

(a) Province-level Mortality Rates – Mean and Cross-Province Standard Deviation



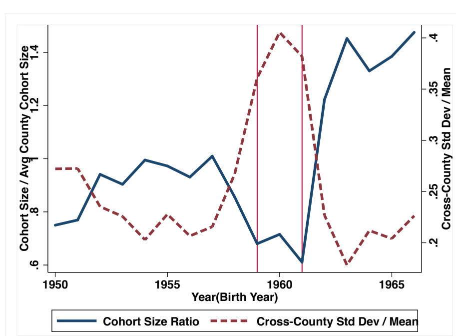
Notes: The solid line plots mean mortality rates, which are average mortality rates across provinces in each year. The dashed line is the standardized variance in mortality rates, which is the standard deviation in mortality rates across provinces in year t divided by the mean mortality rate in year t.

(b) County-level Survivor Birth Cohort Sizes – Mean



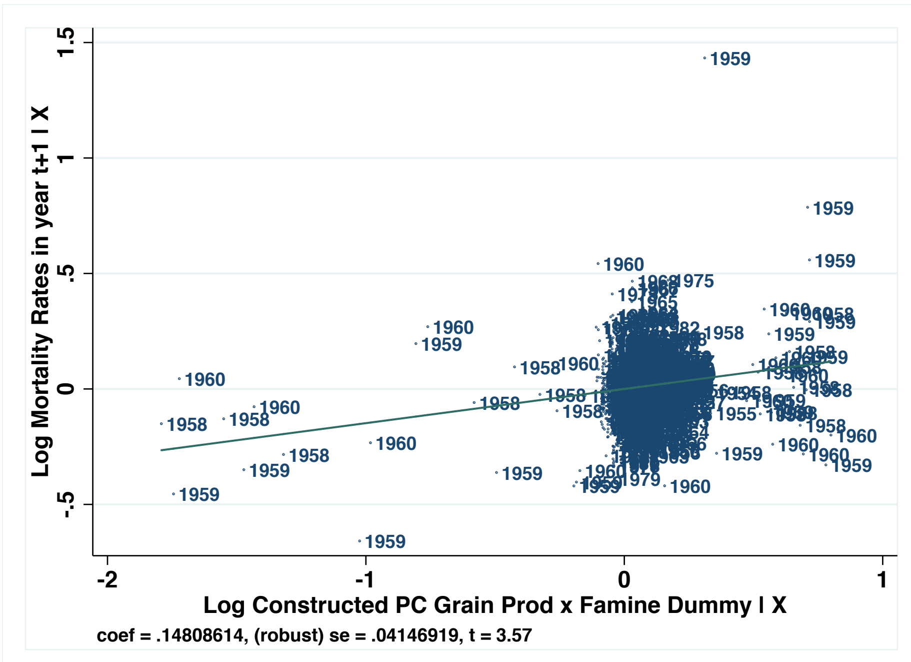
Notes: The solid line plots the 1% size of the birth cohort born in year t (e.g., 50 on the y-axis indicates 5,000 individuals). Individuals are observed in 1990. Cohort size reflects fertility and mortality. The dashed line is a linear trend.

(c) County-level Survivor Birth Cohort Sizes – De-trended Mean and Cross-County Standard Deviation



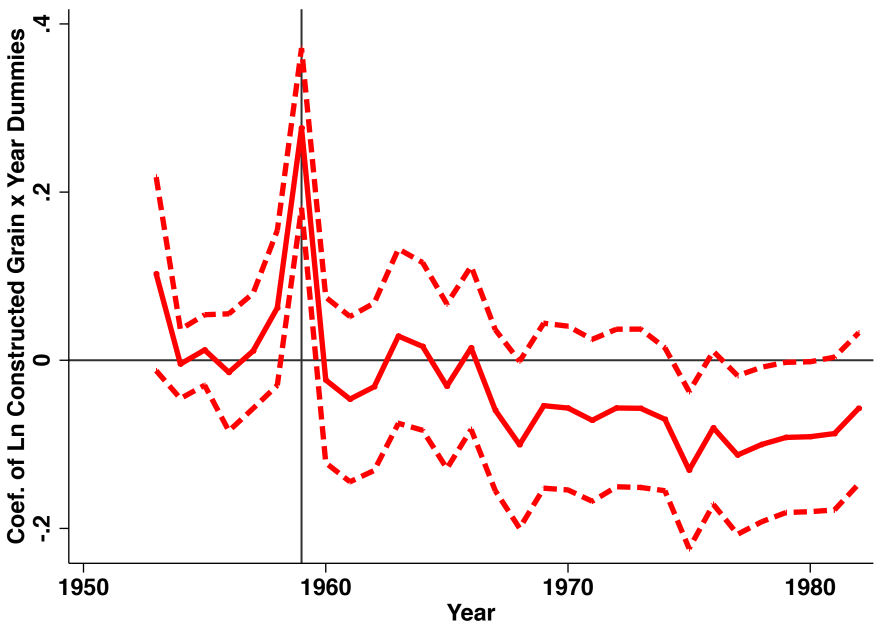
Notes: The solid line plots the de-trended 1% size of the birth cohort born in year t. The dashed line plots the normalized cross-county variance in birth cohort sizes. Individuals are observed in 1990.

Figure 2: Mortality Rates and Constructed Grain Productivity during the Famine – Residual Plot of $Ln\ Constructed\ Grain\ Production \times Famine\ Dummy\ Variable$



Notes: This figure plots the residuals and the regression line from regressing log mortality in year $t + 1$ on the interaction of log constructed grain production in year t and the famine dummy variable, while controlling for log constructed grain production, log total population, log urban population and year fixed effects (Table 2 column (1), $\hat{\beta}$ from Equation (1)). All of the explanatory variables are measured in year t. Constructed production is predicted by climate, geography, total land area, and total rural population as inputs.

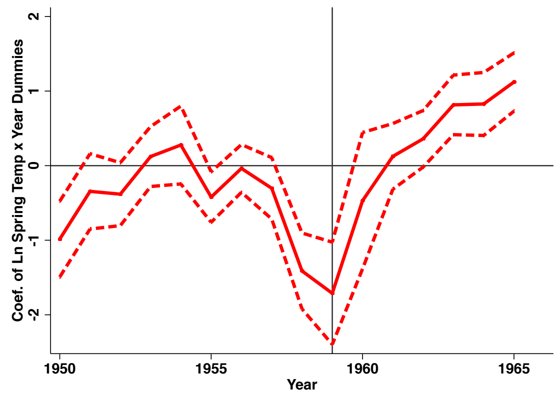
Figure 3: The Correlation between Constructed Productivity and Mortality Rates over Time – Estimated Coefficients of $Ln\ Constructed\ Grain\ Production \times Year\ Dummy\ Variables$ and 95% Confidence Intervals



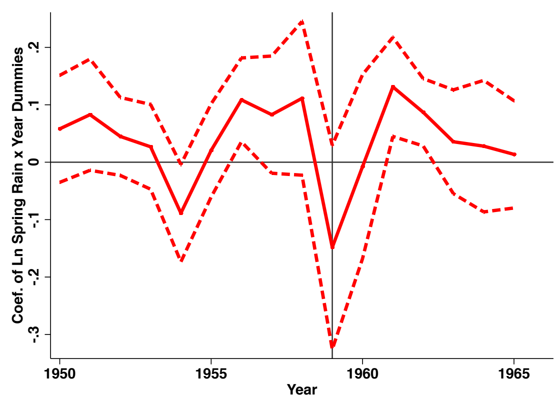
Notes: The solid line plots the coefficients of the interaction effects of log constructed grain production and dummy variables for each year, which are estimated by regressing log mortality in year $t+1$ on the interaction variables, while controlling for log total population, log urban population and year fixed effects. $\hat{\alpha}_{\tau}$ from equation (2). All of the explanatory variables are measured in year t. Constructed production is predicted by climate, geography, total land area, and total rural population as inputs. The coefficients and standard errors are presented in Online Appendix Table A.3.

Figure 4: The Correlation between Geo-Climatic Conditions and Survivor Birth Cohort Size over Time – Coefficients of $Natural\ Conditions \times Year\ Dummy\ Variables$ and their 95% Confidence Intervals

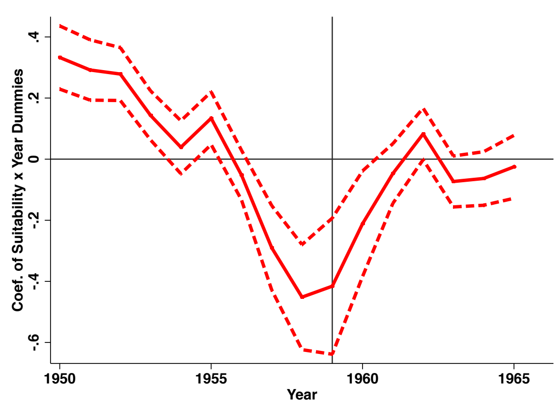
(a) Coefficients of $Ln\ Mean\ Spring\ Temperature \times Year\ Dummy\ Variables$



(b) Coefficients of $Ln\ Mean\ Spring\ Rainfall \times Year\ Dummy\ Variables$



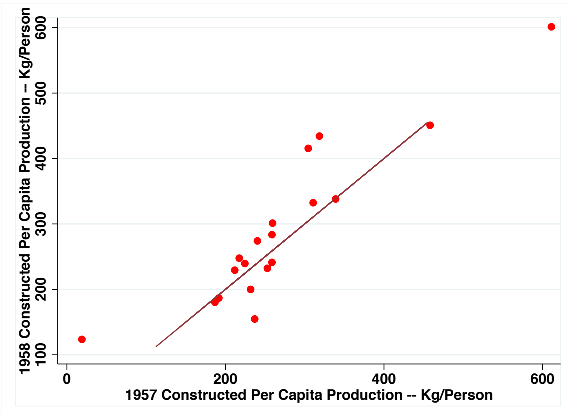
(c) Coefficients of $Suitability\ Index \times Year\ Dummy\ Variables$



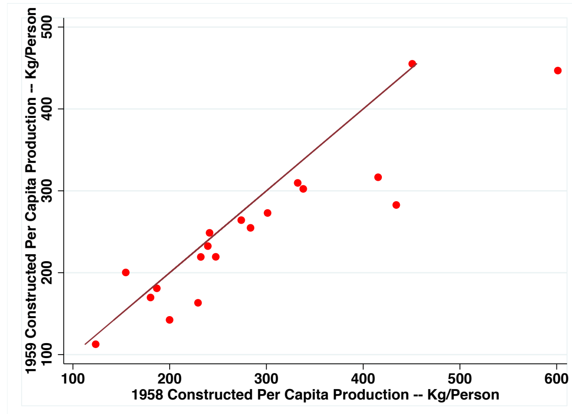
Notes: The interaction coefficients are estimated by regressing log birth cohort size in year $t+1$ on the interaction of the specified variable (log spring temperature in year t, log spring rainfall in year t and suitability for grain production, which is time-invariant) with birth year dummy variables. The coefficients for all three figures are estimated from one regression that controls for log average county birth cohort size, province and year fixed effects. The estimated coefficients and standard errors are shown in Online Appendix Table A.6.

Figure 5: The Correlation between Constructed Productivity and Productivity Growth versus Lagged Constructed Productivity

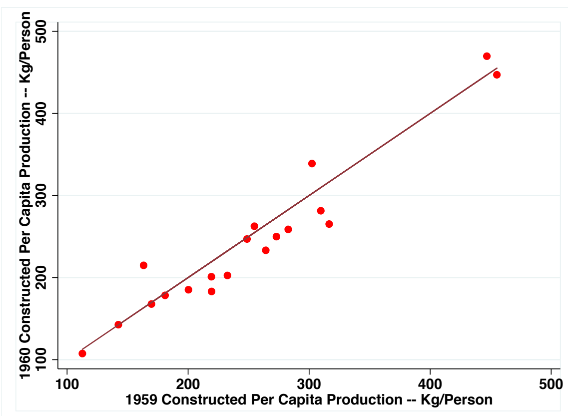
(a) 1958 vs. 1957 Productivity



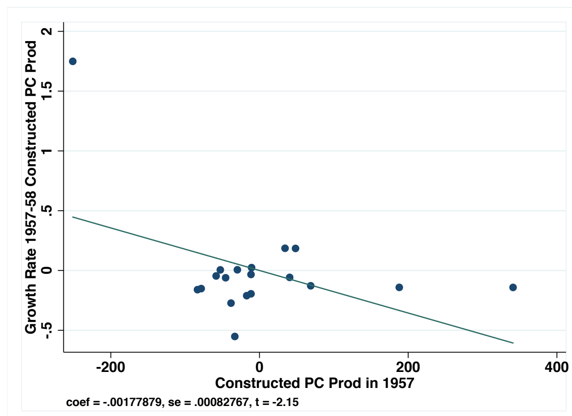
(b) 1959 vs. 1958 Productivity



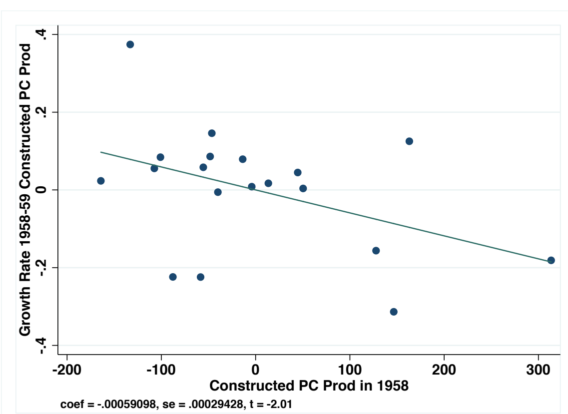
(c) 1960 vs. 1959 Productivity



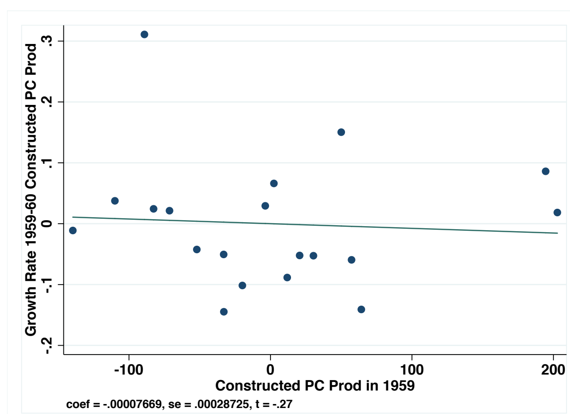
(d) 1958 Productivity Growth vs. 1957 Productivity



(e) 1959 Productivity Growth vs. 1958 Productivity



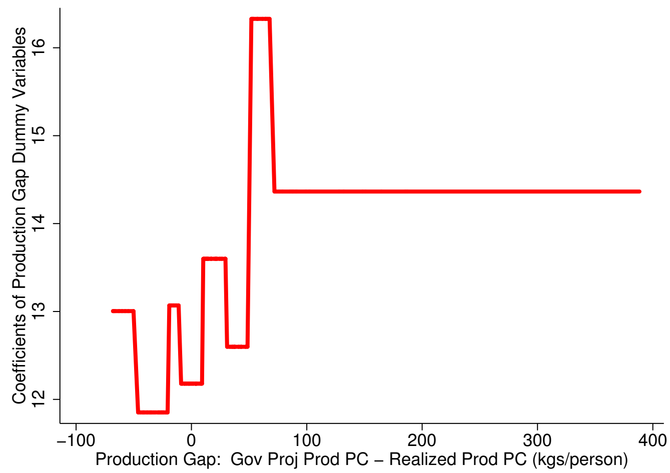
(f) 1960 Productivity Growth vs. 1959 Productivity



Notes: Constructed production is predicted by climate, geography, total land area, and total rural population as inputs. Constructed productivity = constructed province production/province population.

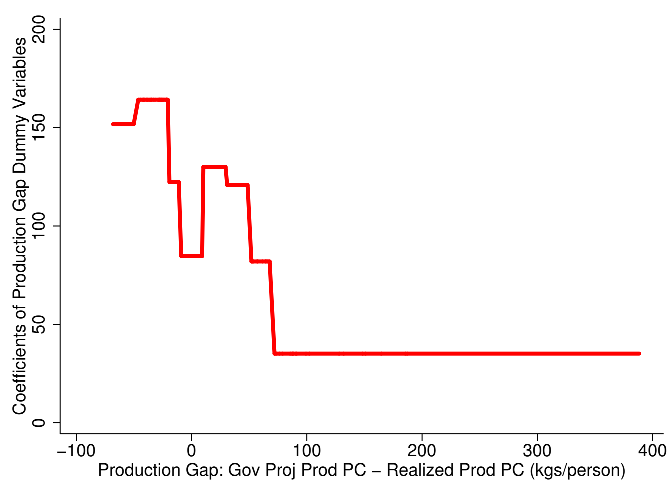
Figure 6: The Effect of the Production Gap (Government Projected Production - Realized Production) on Mortality and Food Retention, 1954 - 1964

(a) Mortality Rate (Deaths per 1,000)



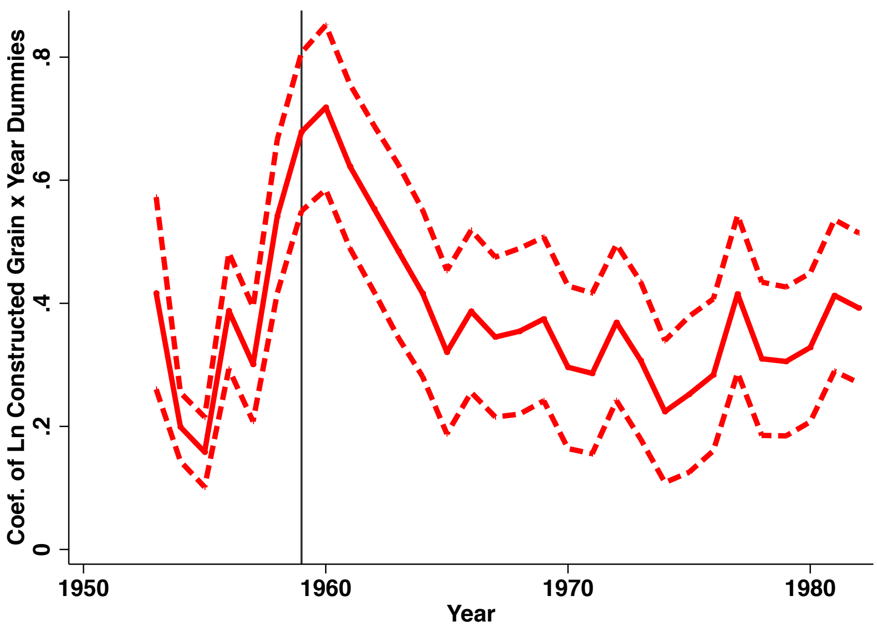
Notes: The coefficients of production gap dummy variables are estimated from regressing mortality rates in year $t+1$ on the gap between government projected production and realized production in year t, while controlling for the urban population share; we are estimating the function $F(\cdot)$ in equation (3), where the dependent variable is mortality rates in year $t+1$. The production gap is the difference between government projected production and realized production. The regression uses constructed production measures, which are predicted by climate, geography, total land area, and total rural population. The coefficients and standard errors are presented in Table 6 column (1).

(b) Per Capita Grain Retention (Kg/Person)



Notes: See previous note. We are estimating the function $F(\cdot)$ in equation (3), where the dependent variable is per capita retention. The coefficients and standard errors are presented in Table 6 column (4).

Figure 7: The Correlation between Constructed Productivity and Grain Procurement Rates – Estimated Coefficients of $Ln\ Constructed\ Grain\ Production \times Year\ Dummy\ Variables$ and their 95% Confidence Intervals



Notes: The interaction coefficients are estimated by regressing the of grain procurement on the interaction of log constructed grain with year dummy variables, while controlling for log urban population, log total population and year fixed effects. The plotted coefficients are $\hat{\alpha}_\tau$ from equation (2) where the dependent variable is log procurement. The estimated coefficients and standard errors are shown in Online Appendix Table A.3 columns (3)-(4).

ONLINE APPENDIX

A Production, Mortality and Procurement Data

Aggregate and provincial-level production, population, urban population and mortality data are reported by the *China Data and Statistical Materials for 50 Years, 1949-98* (CDSM50), which is published by the National Statistics Bureau (NBS). Historical province-level demographic breakdowns are published as tables from the *1953 Population Census*, and made available to us by the NBS. The typical concern for using production and mortality data reported by the Chinese government is that it has historically under-reported mortality and over-reported production to minimize the severity of the famine for political reasons. The data we use were published in the post-Mao reform era and have been carefully corrected by the NBS to address potential reporting errors from the Mao-years. The difficulty in such revisions results in missing data for some provinces and some years (e.g., there is no data for Tibet until 1978, data for Sichuan are missing for several years). Comparisons across data sources show that these demographic and production data have been adjusted from the Mao-era reports (Ashton et al., 1984). For example, the official figure for 1959 national grain production was 270 million tons according to 1959 reports, but revised downwards to 200 million tons in the official reports in 1980, which was similar to the amount reported by the United States Department of Agriculture and the U.S. Agriculture Attache in Hong Kong (Ashton et al., 1984: Table 6). The procurement data are reported by the Ministry of Agriculture in their publication*, Nongye Jingji Ziliao, 1949-1983*. The data we use are the best that are available to researchers today and have been used by all recent studies of the famine (see Section 2 of the paper for references).

B Main Analysis Sample Restriction

In the 1950s, there were officially 24 provinces in total.[73] These included three province-level municipalities that were mostly urban, but still contained large fractions of agricultural populations (Beijing, Shanghai and Tianjin). In addition, there were five regions that were given official political "autonomy" due to historical political reasons (Tibet, Xinjiang, Inner Mongolia, Guangxi and Ningxia). Our main sample excludes the five officially autonomous regions because they faced different political and economic policies. Amongst the 24 provinces, we exclude four provinces with large ethnic minority populations that live in autonomous communities (Qinghai, Gansu, Yunnan and Guizhou) because these

[73]The present day provinces of Chongqing and Hainan had not yet attained provincial status.

autonomous communities also faced different political and economic policies, and Sichuan because there is no mortality data for this province for 1958 and 1959. Thus, the main sample used in this paper uses 19 of the 24 provinces that existed during the famine era.

The main insights that rural retention was, on average, enough to avoid a famine and that famine severity was positively associated with food productivity are not unique to our chosen sample. In Section 4.1, we show that our main results are similar when we use all 23 provinces for which we have data. In an earlier version of the paper, we also show that our results are similar when we include the autonomous regions, though they are less precisely estimated (Meng et al., 2009). In addition, we also show that average rural food production in each province was sufficient for survival in all of the provinces in our sample, and was sufficient for heavy adult labor in all provinces except the municipalities, which suffered little famine mortality. The main difference for the excluded provinces is that the relationship between production and mortality is noisier, which is consistent with the fact that the excluded regions faced different procurement policies.

C Weather Data

The weather data is reported by the *Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1996)* data set. These data have been used by many recent studies, including Dell et al. (2012) and Nunn and Qian (2011). The Chinese weather data were never meant to be publicly released and therefore unlikely to have been manipulated by the famine-era government.[74] We use ArcGIS to calculate the monthly mean temperature and precipitation for each county for the years 1950-65.

For the weather variables used in estimating the production function using province-level data (Section 5.1), we calculate the sum of all rainfall and temperature in a province in each month. For each year and province, we have twelve rainfall variables and twelve temperature variables. We can alternatively use mean temperature or rainfall, or measure rainfall and temperature as deviations from their historic means (for each province). This does not change the constructed grain production estimates. Hence, we only report one specification for brevity. Results from other specifications are available upon request. Note that weather stations do not cover all areas in each province. This results in measurement error in independent variables. We can adjust for this by calculating the area covered by weather stations in each province and adjusting output proportionally (e.g., if weather stations cover 70% of a province,

[74]The data was made available only very recently for historical climatology studies. Detailed documentation is provided at http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts.html.

we adjust production and population downwards by multiplying each variable by 0.7). This does not change the results since we are interested in per capita production. These corrections are not reported for brevity.

For the weather variables used in the county-level estimates of the relationship between cohort size and natural conditions (Section 5.3), we calculate the sum of all rainfall and temperature in each county during spring months in a year. We construct one temperature and one rainfall variable per county-year observation so that the coefficients of this reduced form regression are easy to interpret. The focus on the spring months is motivated by studies of historical Chinese agriculture, which argue that warm spring temperatures and plentiful precipitation were typically the most important factors for a good grain harvest (Walker, 1965, 1984).

For the county-level analysis, we restrict our sample to the counties for that report historical weather data.

D Suitability Data

The measure of suitability for grain cultivation is constructed from a model and data provided by the *Food and Agriculture Organization's GAEZ* (2002) database. It is based purely on the biophysical environment of a region and is not influenced by which crops are actually adopted in an area. Factors that are easily affected by human actions, such as soil pH, are not parameters in this model. The suitability measure at the county-level is the fraction of grids within a county that is suitable for the production of all types of grains that were produced in China during the 1950s, a measure we use for the sake of computational ease. The suitability measure at the province-level is the fraction of grids within a province that is suitable for the same set of crops. These data and similar measures have recently been used for the historical Chinese context in Nunn and Qian (2011a, 2014) and Jia (2011). See Nunn and Qian (2011a) for a detailed discussion of the underlying parameters of the GAEZ model.

E Per Capita Caloric Requirements in 1953

The first benchmark of the caloric requirement for heavy adult labor and healthy child development is constructed by combining data on caloric requirements by age and sex, as recommended by the United States Department of Agriculture (USDA), with data on population, age and sex distribution in China from the *1953 Population Census*.[75] The USDA recommends caloric requirements according to the age and sex of an individual, and the degree of activity. For this calculation, we use the caloric guidelines

[75] USDA guidelines are published at http://www.cnpp.usda.gov/Publications/USDAFoodPatterns/EstimatedCalorieNeedsPerDayTable.pdf

for the highest level of activity. Since the Census divides individuals into broader age categories than the USDA guidelines, we calculate the average caloric requirement for each sex-age group as the average recommended caloric requirements across all of the ages in each group. It is important to note that the USDA recommendations are extremely generous and approximately 20% higher than the average caloric consumption of East Asian countries during the mid-1960s.[76]

The second benchmark of the caloric requirement for avoiding mortality is calculated to be 43% of the first benchmark due to the scant medical evidence on the amount of calories needed to stay alive. Therefore, we follow Dasgupta and Ray's (1986) assumption that 900 calories are required for an adult male to do some work, and conservatively assume that a similar amount is needed to stay alive. Since 900 calories is 43% of the USDA recommendations for heavy labor by an adult male, we construct the benchmark calories for survival by assuming that the requirement is 43% of USDA recommendations for all age and sex groups.[77]

The estimation for average caloric requirements once the demographic breakdown is taken into account is shown in Online Appendix Table A.1. We estimate that for China, as a whole, 1,871 calories were needed per person per day on average for heavy labor and normal child development, and 804 calories were needed per person per day on average to stay alive in 1953. Our accounting exercise assumes that the per capita requirements during the famine are similar to 1953.

F Within-Province Variation in Famine Intensity

We construct a proxy for famine severity that is each county's birth-cohort size when it was lowest during the famine (1959-1960) relative to its non-famine cohort size (1954-1957). This proxy takes a smaller value when the famine is more severe. For example, it is less than one if the birth-cohort size of individuals born during the famine is smaller than that of individuals who were age two to five years when the famine began. A value that is greater than one would suggest that a county was relatively unaffected by the famine. Online Appendix Table A.2 columns (1)-(4) present the estimated mean, normalized variance (standard deviation divided by the mean), and the minimum and maximum values of this famine severity proxy for China as a whole and for each province. These descriptive statistics

[76] Average per capita food consumption is reported for different time periods and regions by the Food and Agriculture Organization (2002).

[77] Note that our calculated thresholds are generally extremely generous and more than sufficiently take into account additional calories needed for cold winter conditions. See an earlier version of this paper, Meng et al. (2010) for a discussion of the medical evidence. When comparing the caloric needs between our two benchmarks, one should also note that the relationship between calories consumed and work capacity is potentially highly non-linear, as suggested by the nutritional poverty trap theory (Das Gupta and Ray, 1986). This does not play an important role for our study since production in 1959 was in excess of both benchmarks.

show that for China as a whole, the famine birth-cohort size is approximately 70% of the pre-famine birth-cohort size. However, there is enormous spatial variation. For the hardest hit county, famine birth-cohort size was less than 14% of the pre-famine cohort. Column (5) shows the fraction of counties within a province that suffered relatively little famine, which we measure as the fraction of counties that have famine birth-cohort sizes that are equal to or greater than pre-famine cohort sizes. Approximately 11% of counties experienced little famine according to this measure. Note that the birth-cohort sizes of those who were born during 1954-57 were also reduced by famine exposure. Thus, our proxy is likely to understate the fraction of counties that suffered famine. The province-specific estimates show that there is both substantial cross-province and within-province variation in famine severity. The variation across provinces can be observed by comparing the famine birth-cohort size in the hardest hit provinces of Anhui and Sichuan, which are nearly half as large as the pre-famine birth-cohort size, to that of Beijing, where the famine and pre-famine cohort size are similar. The variation within provinces can be observed from the large standard deviations of within-province famine severity shown in column (2). We also investigate whether the spike in variance is driven by outliers by plotting histograms of the county-level famine-to-non-famine birth-cohort size ratio. They show that famine severity was distributed roughly normally across counties in China, and across counties within each province. Thus, the observed variance in cross-county and cross-province famine severity are not driven by outliers. The histograms are not presented for brevity. Please see the previous version of the paper for these results.

G Caloric Accounting Exercise with Corrected Production and Procurement Data

In Section 3 of the paper, we document that reported aggregate rural retention (reported aggregate production minus reported aggregate procurement) was almost three times as high as the level required for avoiding mortality at the peak of the famine. However, one may be concerned that this result is driven by misreporting of the production or procurement data. For example, the government may have exaggerated production to emphasize the success of GLF policies. It may also have overstated or understated procurement, depending on whether it wished to emphasize the successes of the GLF policies for production or to minimize the role of government procurement in causing the famine. We can check whether such misreporting is likely to overturn the aggregate rural food availability estimates by first comparing predicted aggregate production (see Section 5.1 of the paper) to reported aggregate production. These numbers are very similar: the difference between the constructed and reported measures (constructed - reported) during 1958 to 1960 show that aggregate reported production at most over reported production by two million tons (in 1958). Thus, over-reporting of production is unlikely to overturn our finding that aggregate retention exceeded the aggregate rural nutritional needs for avoiding famine.

Next, we turn to the procurement data. This is harder to correct. One reasonable number to use as a benchmark for the upper bound aggregate procurement rate is to take the highest procurement rate of any province each year and apply it to the entire country for that year. The maximum procurement rates in our sample are 23% in 1958, 27% in 1959 and 15% in 1960. When we substitute these procurement rates for those shown in Table 1 of the paper, we find that rural retention declines relative to those shown in Table 1. In terms of calories per day, they become 2,686 in 1958, 2,229 in 1959 and 2,274 in 1960, which are still all far above the 804 calories required for avoiding famine mortality, as well as the higher benchmark of 1,871 calories.

H Reports of 1958 Production Falls

The political climate in 1959 was extremely tense and most likely caused leaders to follow rules, even those that were likely to prove problematic later. The GLF had been received with cynicism from many party officials from the very beginning, and its failures and successes were crucial to Mao's political leadership. In December 1958, at a meeting of the Central Committee of the CCP in Wuhan, party leaders refused to fully endorse GLF policies. Following this meeting, Liu Shaoqi replaced Mao, who remained Party Chairman, as the Head of State in early Spring of 1959 (Spence, 1990, pp.581). Many historians view this as an unwilling step down by Mao. It is therefore not surprising that further challenges of the GLF resulted in a strong response from Mao. In July 1959, Mao famously purged Peng Dehuai, a field marshal of exceptional political standing, for criticizing collectivization and other GLF policies and expressing forebodings of famine. These problems of the collective system mandated by the GLF were a source of contention between communist party moderates and hard-liners who backed Mao. However, with the exception of Peng Dehuai who did a tour of the countryside during the spring of 1959, there is no evidence that any top leader ever obtained an accurate picture of the problems of collectivization and the danger of famine. Peng reported these problems to Communist Party Chairman Mao Zedong in a personal letter. The problems he mentioned included reduced incentives to work, a diversion of labor away from agriculture, and over-procurement of grain by mid-level party leaders who were under-pressure to fulfill grain target quotas that had been set too high. Fearing a political revolt against his leadership based on perceived failures of the GLF, Mao used the contents of this letter to purge Peng as a rightist at the historic Lushan conference in July of 1959. At this conference, the top party leaders made clear that the first year of the GLF was a success and that collectivization was increasing grain harvest more than ever (Becker, 1996, p. 87-92).

I Food Waste and Migration

Food wastage is often mentioned as a feature of the centrally planned agricultural economy. We discuss here how it is difficult for food wastage to explain the spatial patterns we document. For example, Yang (1998b) and Chang and Wen (1997) find that collective kitchens created substantial inefficiencies. However, Tables 3 and 4 show that communal kitchen participation rates were uncorrelated with the productivity and that the positive relationship between famine mortality and productivity is very robust to controlling for communal kitchen participation rates.

Another source of food wastage occurs during collection. Communes were told the window during which procurers would come to collect the grain. Grain would be lined up in waiting on the sides of the road. If there was bad weather, then an entire harvest could be lost. This was a problem for China as late as the 1980s (Oi, 1991). Given that the root cause of this loss is in part due to the logistic difficulty of transporting food and the uncertainty of when the procurer would arrive, it follows that there would be fewer such losses in places that are closer to major transportation networks used for transporting procured grains such as railroads or to urban areas. However, the results in Online Appendix Table A.9 suggests that for two counties of equal productivity, famine survival was higher in the county further away from railroads and cities (i.e. the triple interaction of suitability for grain cultivation, the famine dummy, and distance from the city is positive and significant at the 20% level; the triple interaction of log spring temperature, the famine dummy and distance from the city is positive and significant at the 1% level; the triple interaction of log spring temperature, the famine dummy and distance from the railroad is positive and significant at the 1% level). This is inconsistent with the idea that famine is more severe in regions where food rotted away as it waited to be transported. Instead, it is consistent with the idea that all else equal, famine mortality was higher in places where the cost of procurement was lower.

Another potentially confounding factor for our interpretation is if productive regions are also more "distant" regions (e.g., far from cities) such that the relationship between productivity and famine mortality is partly driven by the fact that people living in less productive regions were able to escape more easily. The results in Online Appendix Table A.9 suggests that this is unlikely. Productive regions closer to the cities and railways experienced higher famine mortality than productive regions further from urban areas, even though urban areas had more food during the famine than rural areas (i.e., the interactions of suitability and the famine dummy and of log spring temperature and the famine dummy are negative and significant at the 1% level). These results are again consistent with the notion that rural areas closer to cities faced heavier government procurement, and are inconsistent with the alternative hypothesis that the spatial patterns are driven by differential migration costs for escaping the famine.[78]

J Mortality with Reported Production Data

In Appendix Table A.7, we present the analogous results as Table 6, except that we use reported production data instead of the production that we predict using natural conditions. Note that the sample is slightly smaller than the main results that use predicted production because there are several observations which do not report production, but for which we have input data to construct our measure of predicted production. Figure A.4 is a histogram of the production gaps using the reported production data.[79] As in the main exercise, we allow F be a flexible step function defined over intervals of production gaps. We divide the data into seven groups of per capita production gaps. Please note that the standard deviation for the production gaps from reported data, shown in Appendix Figure A.4, are twice as large as that from the constructed data. We therefore use wider intervals to define production gap groups. As before, we try to keep the intervals similar across groups while preserving a sufficient number of observations within each group.

Columns (1)-(3) show the relationship between mortality rates and the production gaps. Figure A.5a plots the coefficients from column (1). It shows that mortality rates are increasing in the production gaps. Column (2) additionally controls for GLF factors. Column (3) further controls for year fixed effects. The regressions are estimated with no constants.

The quantification of the aggregate impact of our mechanism, computed as in the main exercise, shows that our mechanism is quantitatively important. Based on the model presented in column (1), this is 40%. If we use the model in column (2) that controls for GLF intensity, the contribution of the inflexible procurement policy is 29%. If we use the model in column (3) that controls for GLF intensity and year fixed effects, then the contribution of the inflexible procurement policy is 32%.

[78] The data for distance variables measure the nearest distance from the center of each county to major railways and cities. They are taken from Banerjee et al. (2012).

[79] The main difference between the constructed and reported production data is that there are more negative production gaps (i.e., production surpluses) in the latter. Since the production gap is essentially the difference between current production and past production, this is consistent with the belief that provinces were under pressure to report increasingly higher production levels.

K Retention and Mortality with Reported Production Data

In Appendix Table A.7 columns (4)-(6), we present the analogous results as Table 6 (4)-(6), except that we use reported production data instead of the production that we predict using natural conditions. Columns (1)-(3) show the relationship between mortality rates and the production gaps. As in the main exercise, column (1) only controls for the share of urban population. Figure A.5b plots the coefficients. It shows that per capita retention is decreasing in the production gaps. The pattern is similar to when we use the constructed production data. Column (2) additionally controls for GLF factors. Column (3) further controls for year fixed effects. The regressions are estimated with no constant.

Appendix Table A.8 presents the relationship between the production gap and retention and mortality. This is analogous to Table 8 in the main paper, except that it uses reported production data instead of constructed production data. The signs of all of the estimates are similar to those from using constructed production data. Thus, our qualitative results are not sensitive to the choice of data. The magnitudes of the estimates are also mostly similar between the estimates in Appendix Table A.8 and Table 8 in the main paper. One exception is the relationship between retention and the production gap during the famine era in column (3). This is consistent with the belief that the reported production data during the famine is measured with error.

Table A.1: Caloric Requirements Accounting

Age Bracket	Population (100)	Daily Caloric Needs	Population Daily Caloric Need	Average Daily Caloric Need
(1)	(2)	(3)	(4)	(5)
	A. 1954 Caloric Needs for Heavy Agricultural Labor (or Healthy Child Development)			
Female				
0-5	495,641	1,300	64,433,330,000	
6-10	335,192	1,800	60,334,560,000	
11-15	294,474	2,200	64,784,280,000	
16-20	298,419	2,200	65,652,180,000	
21-50	1,055,377	1,800	189,967,860,000	
51-100	432,744	1,300	56,256,720,000	
Male				
0-5	542,455	1,300	70,519,150,000	
6-10	373,404	1,800	67,212,720,000	
11-15	347,053	2,500	86,763,250,000	
16-20	343,704	3,000	103,111,200,000	
21-50	1,165,685	2,100	244,793,850,000	
51-100	387,607	1,600	62,017,120,000	
Total	**6,071,755.00**		**1,135,846,220,000**	**1,870.70**
	B. 1954 Caloric Needs for Avoiding Mortality			
Female				
0-5	495,641	559	27,706,331,900	
6-10	335,192	774	25,943,860,800	
11-15	294,474	946	27,857,240,400	
16-20	298,419	946	28,230,437,400	
21-50	1,055,377	774	81,686,179,800	
51-100	432,744	559	24,190,389,600	
Male				
0-5	542,455	559	30,323,234,500	
6-10	373,404	774	28,901,469,600	
11-15	347,053	1,075	37,308,197,500	
16-20	343,704	1,290	44,337,816,000	
21-50	1,165,685	903	105,261,355,500	
51-100	387,607	688	26,667,361,600	
Total	**6,071,755.00**		**488,413,874,600**	**804.40**

Notes: Columns (1) and (2) are reported in Coale's (1981) corrected version of the *1953 Population Census*. Column (3) in Panel A is calculated from guidelines provided by the United States Department of Agriculture. For children and adolescents, reference height and weight vary. For adults, the reference man is 5 feet 10 inches tall and weighs 154 pounds. The reference woman is 5 feet 4 inches tall and weighs 126 pounds. Column (3) in Panel B are estimated to be 43% of those in panel A. This is projected from the observation that an adult male labor need approximately 900 calories to engage in some labor, which is approximately 43% of the requirement for heavy physical labor. Column (4) is the product of column (2) x (3) x 100.

Table A.2: Within-Province Variation in Rural Famine Intensity

	Famine Birth Cohort Size (59-60)/ Pre-famine Birth Cohort Size (1954-57)				
Region	Mean (1)	Normalized Standard Deviation (Standard Deviation/Mean) (2)	Min (3)	Max (4)	>1 (5)
China	**0.706**	**0.354**	**0.137**	**1.987**	**0.111**
Beijing	0.966	0.079	0.877	1.082	0.400
Tianjin	0.775	0.249	0.497	1.109	0.200
Hebei	0.708	0.251	0.226	1.105	0.095
Shanxi	0.834	0.304	0.399	1.697	0.216
Neimeng	0.897	0.243	0.517	1.733	0.246
Liaoning	0.725	0.181	0.410	1.162	0.020
Jilin	0.858	0.176	0.424	1.206	0.160
Heilongjiang	0.862	0.254	0.519	1.641	0.278
Shanghai	0.730	0.298	0.407	1.018	0.200
Jiangsu	0.591	0.263	0.243	0.917	0.000
Zhejiang	0.691	0.171	0.463	0.993	0.000
Anhui	0.448	0.452	0.169	0.861	0.000
Fujian	0.720	0.205	0.400	1.042	0.028
Jiangxi	0.742	0.228	0.553	1.071	0.158
Shandong	0.633	0.259	0.211	1.095	0.034
Henan	0.564	0.227	0.320	0.908	0.000
Hubei	0.666	0.251	0.373	1.125	0.043
Hunan	0.517	0.253	0.258	0.769	0.000
Guangdong	0.789	0.286	0.513	1.787	0.068
Guangxi	0.669	0.254	0.313	1.284	0.035
Hainan	0.888	0.215	0.613	1.143	0.300
Sichuan	0.514	0.442	0.162	1.368	0.048
Guizhou	0.547	0.363	0.137	1.171	0.034
Yunnan	0.751	0.379	0.267	1.886	0.153
Tibet	1.371	0.152	1.071	1.765	1.000
Shaanxi	0.872	0.182	0.652	1.409	0.220
Gansu	0.719	0.396	0.213	1.987	0.137
Qinghai	0.623	0.340	0.213	0.949	0.000
Ningxia	0.619	0.226	0.389	0.882	0.000
Xinjiang	0.948	0.322	0.333	1.973	0.392

Notes: The statistics refer to counties within each province. Source: Authors' computations using the *1990 Population Census.*

Table A.3: Year-by-Year Correlations between Ln Constructed Grain Productivity and Mortality Rates, and Ln Constructed Grain Productivity and Procurement Rates

	Dependent Variables			
	Ln Death in Year t+1		Procurement Rate	
	(1)	(2)	(3)	(4)
	coef	se	coef	se
Ln Constructed Grain x Year Dummy				
1953	0.103	(0.0587)	0.416	(0.0797)
1954	-0.00442	(0.0210)	0.199	(0.0285)
1955	0.0122	(0.0214)	0.158	(0.0290)
1956	-0.0144	(0.0355)	0.388	(0.0482)
1957	0.0111	(0.0349)	0.301	(0.0473)
1958	0.0624	(0.0470)	0.542	(0.0637)
1959	0.276	(0.0485)	0.679	(0.0656)
1960	-0.0235	(0.0503)	0.718	(0.0681)
1961	-0.0462	(0.0501)	0.623	(0.0680)
1962	-0.0315	(0.0508)	0.553	(0.0689)
1963	0.0287	(0.0529)	0.485	(0.0718)
1964	0.0164	(0.0509)	0.416	(0.0690)
1965	-0.0309	(0.0500)	0.321	(0.0678)
1966	0.0148	(0.0495)	0.387	(0.0672)
1967	-0.0595	(0.0489)	0.345	(0.0663)
1968	-0.100	(0.0506)	0.355	(0.0687)
1969	-0.0540	(0.0500)	0.375	(0.0679)
1970	-0.0568	(0.0497)	0.296	(0.0674)
1971	-0.0714	(0.0491)	0.286	(0.0666)
1972	-0.0568	(0.0478)	0.369	(0.0648)
1973	-0.0571	(0.0480)	0.307	(0.0651)
1974	-0.0701	(0.0434)	0.224	(0.0588)
1975	-0.131	(0.0476)	0.252	(0.0646)
1976	-0.0804	(0.0464)	0.284	(0.0630)
1977	-0.113	(0.0481)	0.415	(0.0652)
1978	-0.100	(0.0469)	0.310	(0.0635)
1979	-0.0918	(0.0456)	0.306	(0.0617)
1980	-0.0909	(0.0455)	0.328	(0.0617)
1981	-0.0872	(0.0464)	0.413	(0.0628)
1982	-0.0572	(0.0459)	0.392	(0.0621)
Observations	569		569	
R-squared	0.959		0.722	

Notes: All regressions control for log total population, log urban population and year fixed effects. Robust standard errors are presented in the parentheses.These estimates use a balanced province-level panel that includes 19 provinces and the years 1953-1982, with the exception that there is no mortality data for Beijing in 1953. Sources: CDSM50 (1999), the Ministry of Agriculture (1983) and the authors' computations.

Table A.4: The Correlation between *Reported* Grain Productivity and Mortality Rates across Provinces – Robustness to Controls

	Dependent Variable: Ln Deaths in year t+1				
	(1)	(2)	(3)	(4)	(5)
Ln Grain Production x Famine Dummy	0.184 (0.0515) [0.504]	0.219 (0.0729) [0.598]	0.261 (0.0921) [0.713]	0.170 (0.0567) [0.464]	0.204 (0.0754) [0.557]
Ln Grain Production	-0.0312 (0.0292) [-0.0368]	-0.0615 (0.0300) [-0.0725]	-0.0986 (0.0323) [-0.116]	-0.0258 (0.0290) [-0.0305]	-0.0566 (0.0292) [-0.0667]
GLF Principal Component x Famine Dummy		0.0491 (0.0510) [0.0286]			0.0480 (0.0515) [0.0279]
GLF Principal Component		-0.0116 (0.00465) [-0.0348]			-0.0117 (0.00459) [-0.0350]
Per Capita Sown Area Growth (x -1) x Famine Dummy			-0.287 (0.629) [-0.0147]		
Per Capita Steel Production Growth x Famine Dummy			0.0383 (0.0813) [0.0202]		
% Participation in Communal Dining x Famine Dummy			-0.000831 (0.00203) [-0.0260]		
% Purged during 1957 Anti-Right Movement x Famine Dummy**			0.0271 (0.0197) [0.0966]		
Late Liberation Dummy x Famine Dummy			0.103 (0.0728) [0.0357]		
Ln Urban Population x Famine Dummy				0.147 (0.129) [0.399]	0.144 (0.125) [0.390]
Observations	483	483	483	483	483
R-squared	0.943	0.944	0.947	0.943	0.944
Joint: Ln Grain Production x 1959 Dummy + Ln Grain Production	0.153	0.157	0.162	0.144	0.147
p-value	0.00508	0.0420	0.0905	0.0151	0.0631

Notes: The famine dummy is equal to one for the years 1958-1960. All regressions control for log total population, log urban population and year fixed effects. In columns (2) and (5), we control for the principal component of GLF indicators: the late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61. Robust standard errors are reported in parentheses. In column (3), we also control for all of the uninteracted effects of the GLF intensity proxies. **In column (3), the coefficients for anti-right and anti-right x famine dummy are multiplied by 1,000 for presentation purposes. Standardized coefficients are reported in italics and brackets. The sample includes an unbalanced panel of 17 provinces for the years 1953-82 (the number of provinces vary across years due to missing values for the GLF variables, and because there is no mortality data for Beijing for 1953). Sources: See text.

Table A.5: Constructed Grain Production and Province-Level Survivor Birth Cohort Size

Dependent Variable: Province Birth Cohort Size in year *t+1*		
	Birth Cohort Size	Ln Birth Cohort Size
	(1)	(2)
Grain x Famine Dummy	-0.704	
	(0.160)	
Grain	0.657	
	(0.144)	
Ln Grain x Famine Dummy		-0.0911
		(0.0631)
Ln Grain		0.0786
		(0.0234)
Observations	306	306
R-squared	0.550	0.681
Joint Statistic	-0.0474	0.853
p-value	0.790	-0.0125

Notes: The famine dummy is equal to one for the years 1958-1960. In column (1), the regression control for total population and urban population. In column (2), the regression controls for the logarithms of these values.The sample comprises a balanced panel of eighteen provinces: Anhui, Beijing, Fujian, Guangdong, Hebei, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Shaanxi, Shandong, Shanxi, Tianjin and Zhejiang. Both regressions control for year fixed effects. Robust standard errors are presented in parentheses. Sources: CDSM50 (1999) and the authors' computations.

Table A.6: Year-by-Year Correlations between Survivor Birth Cohort Sizes and Natural Conditions

	Dependent Variables: Ln Birth Cohort Size in year *t+1*							
	(1)	(2)		(3)	(4)		(5)	(6)
	Coefficient	Std. Err.		Coefficient	Std. Err.		Coefficient	Std. Err.
Suitability x			Ln Spring Temperature x			Ln Spring Rainfall x		
1950	0.333	(0.0528)	1950	-0.985	(0.259)	1950	0.0581	(0.0475)
1951	0.292	(0.0503)	1951	-0.345	(0.258)	1951	0.0828	(0.0494)
1952	0.278	(0.0442)	1952	-0.382	(0.215)	1952	0.0449	(0.0346)
1953	0.143	(0.0409)	1953	0.122	(0.205)	1953	0.0268	(0.0377)
1954	0.0389	(0.0443)	1954	0.277	(0.267)	1954	-0.0887	(0.0433)
1955	0.134	(0.0440)	1955	-0.421	(0.173)	1955	0.0206	(0.0414)
1956	-0.0517	(0.0414)	1956	-0.0372	(0.163)	1956	0.109	(0.0372)
1957	-0.290	(0.0704)	1957	-0.304	(0.210)	1957	0.0828	(0.0521)
1958	-0.452	(0.0877)	1958	-1.412	(0.260)	1958	0.111	(0.0683)
1959	-0.416	(0.114)	1959	-1.712	(0.350)	1959	-0.148	(0.0906)
1960	-0.211	(0.0882)	1960	-0.470	(0.466)	1960	-0.00755	(0.0817)
1961	-0.0471	(0.0498)	1961	0.124	(0.225)	1961	0.131	(0.0440)
1962	0.0825	(0.0430)	1962	0.360	(0.193)	1962	0.0870	(0.0299)
1963	-0.0731	(0.0426)	1963	0.815	(0.204)	1963	0.0356	(0.0461)
1964	-0.0633	(0.0447)	1964	0.827	(0.216)	1964	0.0280	(0.0584)
1965	-0.0252	(0.0526)	1965	1.121	(0.199)	1965	0.0137	(0.0476)
Observations				8176				
R-squared				0.918				

The regression controls for log average county birth cohort size, year and province fixed effects. The standard errors are clustered at the county level.
Notes: The estimates use a county-level panel of birth cohorts that includes 511 counties for the years 1950-65 constructed from the *1990 Population Census*, *the Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series* and the FAO *GAEZ* database.

Table A.7: The Correlation between the Production Gap and Mortality Rates – Using Reported Production Data

	Dependent Variable					
	Number of Deaths per 1,000 in Year *t+1*			Grain Retention (Kg/person) in Year *t*		
	(1)	(2)	(3)	(4)	(5)	(6)
	1954-1964	1954-1964	1954-1964	1954-1964	1954-1964	1954-1964
Government Projected PC Prod - Realized PC Prod is						
Group1: <=-165	12.45	12.62	8.842	363.0	354.1	258.4
	(0.742)	(1.302)	(0.777)	(26.67)	(29.09)	(31.08)
Group2: (-165, -90]	12.48	12.43	10.65	226.0	215.1	165.2
	(0.681)	(0.820)	(0.841)	(22.91)	(25.65)	(28.60)
Group3: (-90,-60]	13.04	13.08	11.77	206.8	203.4	172.1
	(0.549)	(0.642)	(0.726)	(16.34)	(16.35)	(19.80)
Group4: (-60, -10]	13.20	13.15	11.79	198.6	196.7	172.2
	(0.671)	(0.744)	(0.740)	(12.84)	(12.96)	(17.44)
Group5: (-10, 5]	13.61	13.79	13.18	183.5	197.4	205.6
	(0.995)	(1.199)	(0.784)	(22.56)	(21.39)	(21.61)
Group6: (5, 35]	16.46	16.96	15.67	164.0	170.7	176.2
	(3.507)	(3.999)	(3.028)	(15.54)	(15.96)	(18.65)
Group7: (35,300)	14.28	14.24	14.10	166.7	157.9	170.6
	(0.910)	(0.950)	(0.741)	(13.74)	(15.34)	(16.02)
Controls						
GLF PCA	N	Y	Y	N	Y	Y
Year FE	N	N	Y	N	N	Y
Observations	166	157	157	175	175	175
R-squared	0.820	0.818	0.867	0.963	0.975	0.972

Notes: Robust standard errors are in parentheses. The control for GLF is the first principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61. The number of observations vary due to missing values for the GLF and mortality variables. Sources: CDSM50 (1999) and the authors' computations.

Table A.8: Mortality, Grain Retention and the Production Gap – Using Reported Production Data

	Dependent Variable					
	Retention (Kg/Person)			Mortality in year t+1		
	(1)	(2)	(3)	(4)	(5)	(6)
	1954-1964	1954-1957, 1961-1964	1958 - 60	1954-1964	1954-1964	1954-1964, 2SLS
Production Gap (Kg/Person)	-0.418	-0.501	-0.371	0.00818		
	(0.0803)	(0.109)	(0.116)	(0.00422)		
Retention (Kg/Person)					-0.00728	-0.0203
					(0.00423)	(0.0108)
Observations	193	140	53	175	175	175
R-squared	0.876	0.873	0.894	0.349	0.348	0.323

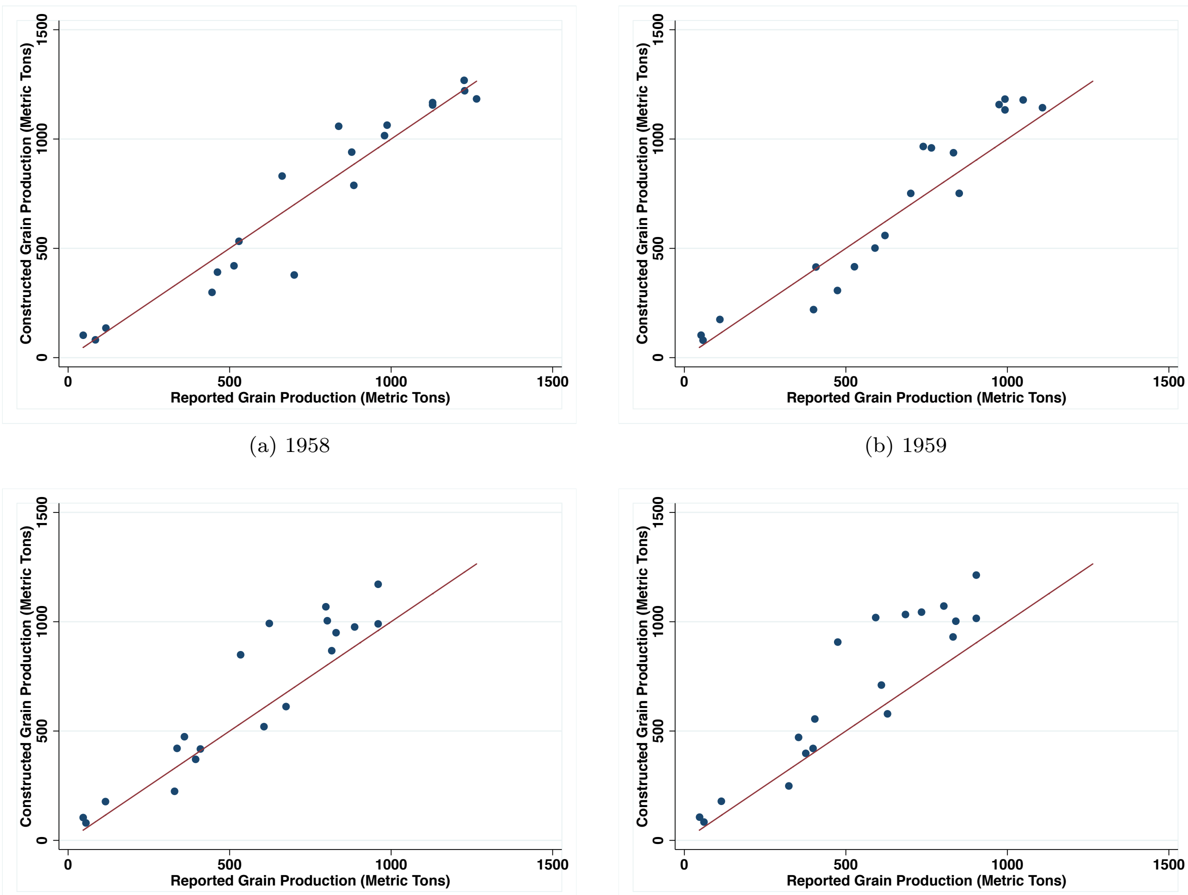
Notes: Robust standard errors are in parentheses. All regressions control for urban population share; the principal component of GLF indicators: late liberation dummy, the % of the population that was purged during the 1959 Anti-Right movement, communal kitchen participation rates during 1958-59, the growth in steel production and the negative growth in area sown for agriculture during 1958-61; and year fixed effects. In column (6), retention is instrumented with production gap. The first stage equation for the 2SLS in column (6) is shown in column (1). The reduced form equation for column (6) is shown in column (4). The number of observations vary between columns (1) and (3) because of missing values in mortality data. There are fewer observations in columns (4) and (6) than column (5) because the production gap is based on government projected production, which is in turn based on data from *t-2*, *t-3* and *t-4* years. The production gap is constructed using reported production data. Sources: CDSM50 (1999) and the authors' computations.

Table A.9: The Effect of Distance from Transportation Networks and Urban Areas on the Relationship between Geo-Climatic Conditions and Survivor Birth Cohort Size

Dependent Variable: Ln Birth Cohort Size in year *t+1*		
	(1)	(2)
Suitability x Famine Dummy	-0.387	-2.289
	(0.0827)	(1.132)
x Ln Dist to Railways		-0.0161
		(0.0556)
x Ln Distance to City		0.175
		(0.119)
Ln Spring Temp x Famine Dummy	-1.403	-20.61
	(0.292)	(4.077)
x Ln Dist to Railways		0.576
		(0.162)
x Ln Distance to City		1.130
		(0.363)
Ln Spring Rain x Famine Dummy	-0.0341	2.601
	(0.0614)	(1.911)
x Ln Dist to Railways		-0.00218
		(0.0461)
x Ln Distance to City		-0.218
		(0.194)
Observations	8,176	8,176
R-squared	0.916	0.917

Notes: The regressions control for all of the double interactions and main effects, log avg birth cohort size and province fixed effects. The standard errors are clustered at the county level. The sample comprises 511 counties for the years 1950-1965.

Figure A.1: Constructed Production versus Reported Production for the four GLF Years 1958-1961



(a) 1958

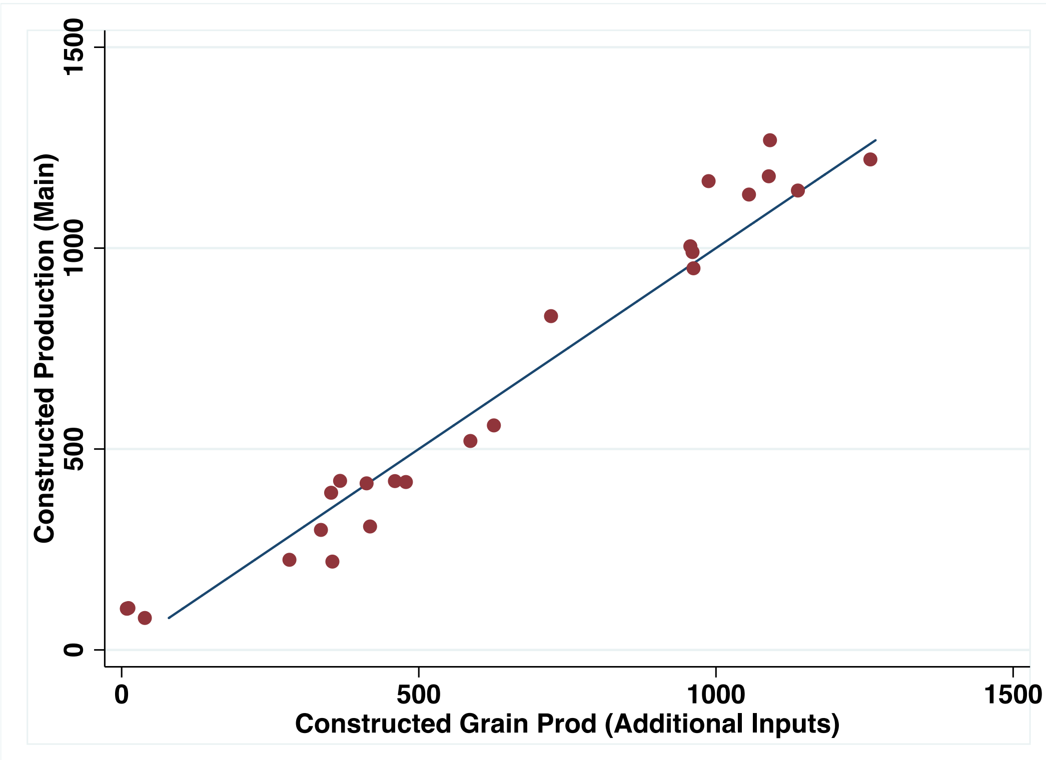
(b) 1959

(c) 1960

(d) 1961

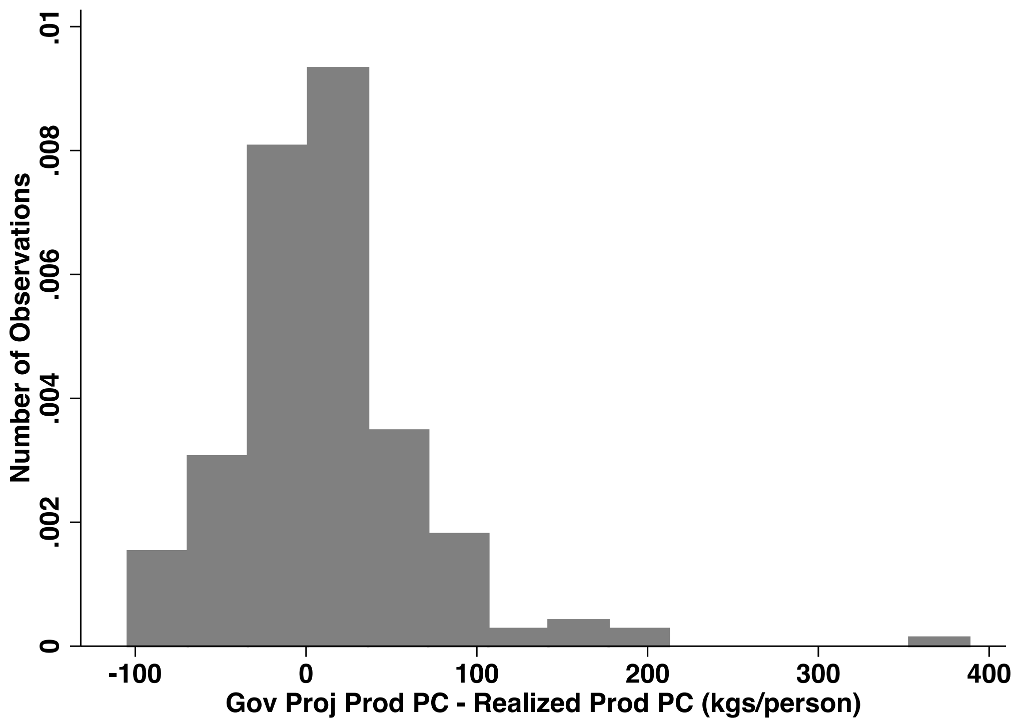
Notes: The y-axes plot constructed production (which is predicted from time-varying weather conditions, time-invariant agro-climatic suitability for grain production, total land area and total rural population). The x-axes plot reported grain production. The diagonal line is the 45-degree line.

Figure A.2: Alternative Methods of Prediction Production



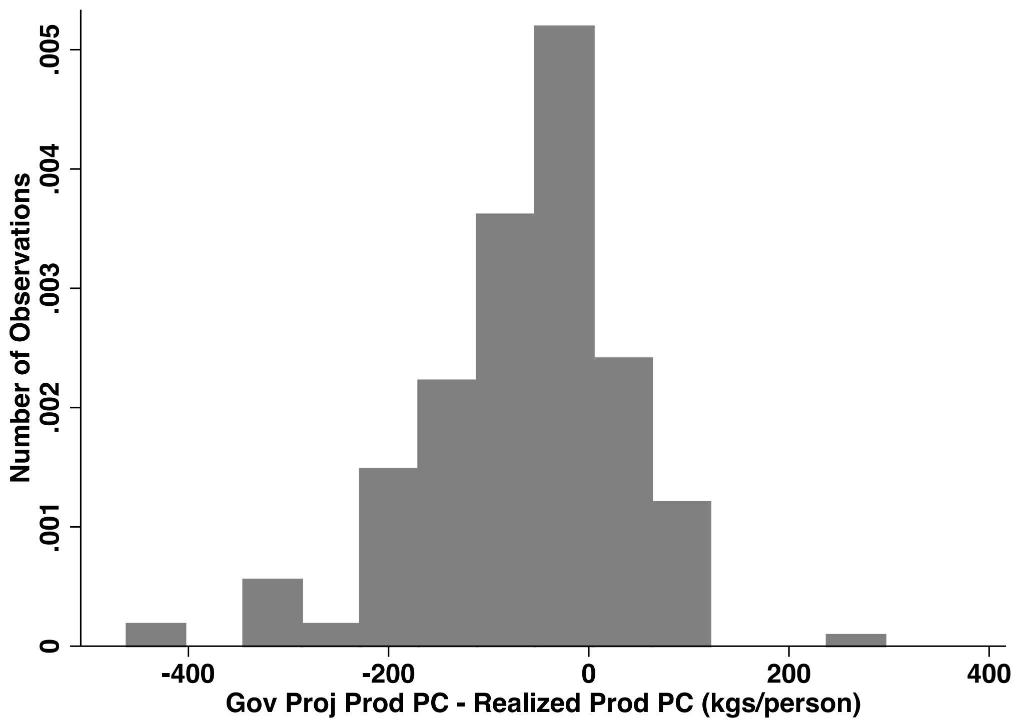
Notes: The y-axis plots constructed production, which is predicted by climate, geography, total land area, and total rural population. The x-axis plots an alternative measure of constructed production, which in addition takes into account the total area sown for agriculture, the total area sown for grain, the total amount of agricultural machine power, the % of irrigated land and total chemical fertilizer consumption as inputs. The observations are at the province and year level. The sample includes the years 1958-60. The diagonal line is the 45-degree line.

Figure A.3: Histogram of Production Gaps, 1954 - 1964



Notes: The production gap is the difference between government projected production (which is projected using realized production from two years ago and the average growth rate from two to four years ago) and realized production. These are calculated using constructed production, which is predicted by climate, geography, total land area, and total rural population.

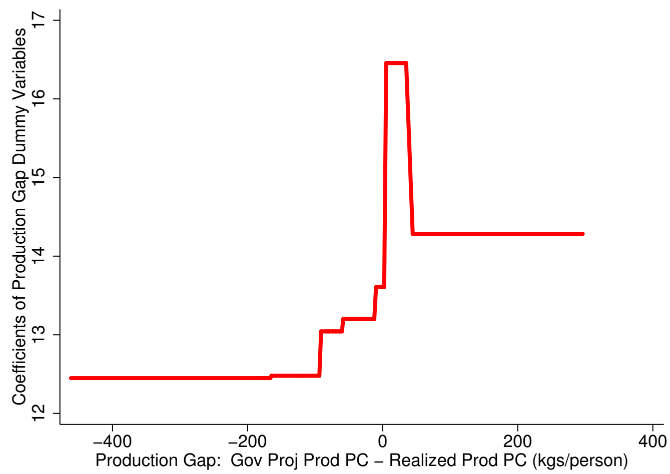
Figure A.4: Histogram of Production Gaps, 1954 - 1964 – Using Reported Production Data



Notes: The production gap is the difference between government projected production (which is projected using realized production from two years ago and the average growth rate from two to four years ago) and realized production. These are calculated using reported production data.

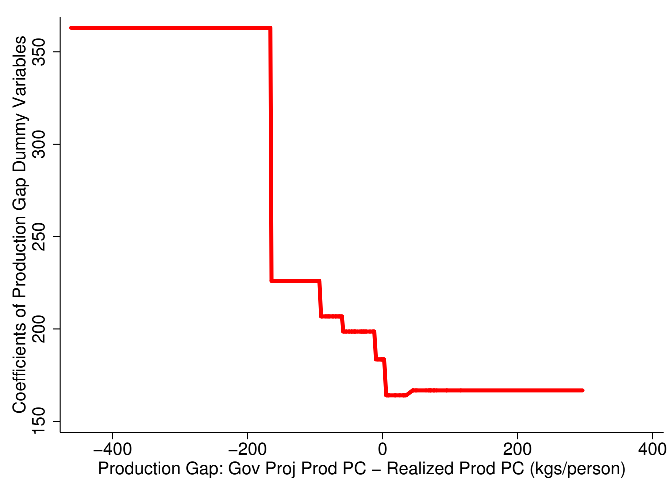
Figure A.5: The Effect of the Production Gap (Government Projected Production - Realized Production) on Mortality and Food Retention, 1954 - 1964 – Using Reported Production Data

(a) Mortality Rate (Deaths per 1,000)



Notes: The coefficients of production gap dummy variables are estimated from regressing mortality rates in year *t+1* on the gap between government projected production and realized production in year t, while controlling for the urban population share; we are estimating the function $F(\cdot)$ in equation (3), where the dependent variable is mortality rates in year $t+1$. The production gap is the difference between government projected production and realized production. The regression uses reported production data. The coefficients and standard errors are presented in Appendix Table A.7 column (1).

(b) Per Capita Grain Retention (Kg/Person)



Notes: See previous note. We are estimating the function $F(\cdot)$ in equation (3), where the dependent variable is per capita retention. The coefficients and standard errors are presented in Appendix Table A.7 column (4).