The Great Leap Forward: Assessing Official Accounts and Long Term Effects

Andrew Tai

May 2019

Abstract

This paper investigates the Great Leap Forward, a period of famine in China from 1958-1962 caused mainly by central government policy. I assess the reliability of official government data on grain production by comparing against demographic and other agricultural data. Overall, I find little positive evidence of gross inaccuracies. I also adapt a framework for assessing the long term impacts of the Great Leap Forward on economic development, though data scarcity prevents implementation.

1 Introduction

In 1958, the Chinese government embarked upon the Great Leap Forward (GLF), a program of rapid Soviet-style agricultural and industrial reforms. These reforms were characterized by forced farm collectivization, massive labor movements from farms to industry, and draconian grain procurements from the countryside. The result was a disaster; widespread famine killed at least 22 million people, making it the worst in human history (Yang, 2010).

Though the Great Leap Forward was one of the most significant events of the 20th century, social scientists have only very partial understanding of the period. This understanding is largely inhibited by data quality and availability. In this paper, I attempt to investigate two aspects of the Great Leap Forward. First, I test the reliability of official Chinese statistics on grain production. These are the typical statistics used in literature investigating this period (see Peng, 1987 or Li & Yang, 2005). Local officials' incentives to misreport are well documented. Indeed, there appears to have been widespread inflation of production numbers in this period driven by political incentives (Becker, 1998). While official statistics from the period have since been revised, it is unclear whether Chinese authorities are willing or able to fully correct them. Second, I present a framework adapted from Davis & Weinstein (2002) for testing the persistence of the impact of the Great Leap Forward on subsequent economic development. While data restrictions prevent this paper from truly implementing this, I am hopeful that future data developments may offer a path forward.

The Great Leap Forward was the most severe non-war mass death in modern history; this on its own makes it an important object of study. However, data availability and trustworthiness have hampered social scientists' ability to study it. I hope to contribute to research in this vein by offering a description and assessment of the available data and a simple evaluation of the Great Leap Forward's long term impact. While the significance of the GLF makes it worthwhile to study in and of itself, this paper also fits into literature in economic geography investigating the spatial distribution of economic activity (like Davis & Weinstein (2002), from which the latter model is adapted). It also adds to literature studying catastrophic events, like Li & Yang (2005), which offers a structural growth-type analysis of the GLF, and Bai & Wei (2019), which studies the Cultural Revolution. It also contributes to literature studying China's miracle growth since the 1980s, which contrasts against the disastrous first three-quarters of the 20th century (see Zhu, 2012).

The paper proceeds as follows: section 2 offers a brief historical account of the GLF and presents a few facts relevant for this paper. Section 3 presents the tests and results for official data reliability. Section 4 presents the framework for testing long-term impact on development. Section 5 concludes.

2 Historical Setting

The Chinese Civil War ended in 1949 when Mao Zedong declared the People's Republic of China. At this time, as in much of Chinese history, the peasantry lived precariously on the edge of food sufficiency. Natural and political disasters had made mass starvations common through the Qing Dynasty, even into the early 1900s (see Li (1982) or Naughton (2007) for historical accounts).

In this context, the Communist Party of China began ambitious programs to industrialize the country. It began collectivizing agriculture and promoting industry almost immediately after the end of the Civil War, culminating in the announcement of the Great Leap Forward in 1957. Amid huge fanfare, the Communist Party announced its plan; among its goals was a 270% increase in gross value of agricultural output and a huge expansion of rural industry. In his optimism, Mao suggested that Chinese industry could surpass the United Kingdom's within 15 years.

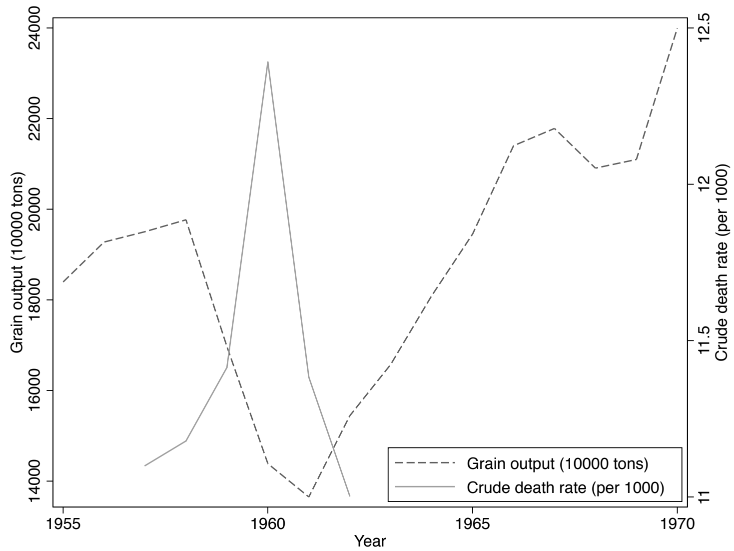
The basic agricultural reform was collectivization. In 1958, farmers were forced into collectives, in which land was pooled and worked in common. Management and pay were also determined at the collective level. Even communal food halls were established, where villagers would dine together, the cost paid for collectively. To encourage industrialization, the central government increased procurements of agricultural output, products which farmers were obligated to sell to the government at low prices. These procurements were diverted to urban areas to feed industrial workers or exported for foreign currency. The government also re-assigned millions of farmers to rural industries, which ended up producing poor quality pig iron.

While production began to rise at first, the conditions quickly deteriorated. Mis-incentivized local officials over-reported production, and the ecstatic central government raised production targets, increased procurements, and diverted more labor to industry. In actuality, poor economic incentives in the collectives and reduced labor inputs were causing agricultural production to fall dramatically. These factors were compounded by extreme weather in 1960. By 1959, starvation conditions prevailed in much of China's countryside. The Communist Party responded by intensifying the political atmosphere, launching the Anti-Rightist campaign; it was not until 1961 that reforms were rolled back and conditions began to recover. The crisis did not end until 1962. Figure 1 shows the official grain production series along with estimated death rates over this time; the catastrophic drop in production and spike in deaths are apparent. Even by conservative estimates such as Peng (1987), over 20 million people died as a result of the GLF famine, making it the worst in human history. Other estimates range as high as 50 million deaths.

2.1 Implications of historical facts for the analysis

During the GLF, there were wide urban-rural disparities and across provinces. This naturally led to internal migration, which gives rise to a difficulty for the analysis. If inter-provincial migration had general equilibrium effects and "exported" conditions between provinces, I am not able to disentangle

Figure 1: Grain output and crude death rates



Source: National Bureau of Statistics; Peng (1987)

these effects. The urban-rural divide also means I will generally study contemporaneous impacts of the GLF restricted to the rural population. If intra-provincial migration affected the province-level impacts of the GLF, then this can be considered part of the effect that I attempt to estimate. However, official policy was generally to forbid movement, so the effect of internal migration is plausibly low.

The GLF was followed by four relatively calm years before the start of the Cultural Revolution in 1966. The close succession of the Cultural Revolution makes it hard to disentangle the effects of the GLF and the Cultural Revolution. In practice, I appeal to the notion that the Cultural Revolution consisted of a very different set of reforms and crises, so that these effects will be independent. (See Bai & Wei (2019) for an account and analysis of the Cultural Revolution.)

In general, I appeal to the notion that the GLF was a unique and large event in Chinese history (indeed, all of history). Even if "clean" identification is ultimately impossible, I believe that the magnitude of the event allows for meaningful inference.

2.2 Definition of provinces

Modern China consists of 33 "provincial level" administrative regions. Two are Special Administrative Regions (Hong Kong and Macau); four are municipalities, and five are autonomous regions. The remainder are regularly administered provinces. This paper of course does not consider Hong Kong and Macau, which were returned to China in the 1990s. The municipalities are Beijing, Tianjin, Shanghai, and Chongqing (the last was established in 1997); as these are large, highly urban cities, and the Great Leap Forward was a largely rural event, I exclude them from the analysis. However, I attempt to show them in the descriptive data figures and tables where possible. Finally, I make the decision to include the autonomous regions with the exception of Tibet, which was newly annexed by China and experienced armed insurrection in 1959. The remainder of the autonomous provinces do not appear to be outliers in the data. Any other exclusions of provinces in the empirical work are due to data unavailability.

3 Reliability of official grain output statistics

The first task is to test the reliability of official Chinese statistics of grain output. I seek to compare this against alternative measures of famine severity which I argue are more reliable, or at least can be cross-checked against each other. With these, I propose a non-structural and a structural test for reliability of official statistics.

3.1 Alternative measures

I present a few options for the alternative measures of famine severity. The first two are demographic: fertility rates and death rates. Peng (1987) constructs estimates of lost fertility from the GLF by province. The variation is large and shows expected patterns; there are significant urban-rural gaps and regional variation. Peng also estimates excess mortality by province. While Chinese demographic data have come under scrutiny, I argue that these statistics were harder to fudge, and that local officials had less incentive to do so. While the fertility data come from a 1982 survey, the One-Per-Thousand-Population Fertility Survey, the years considered predate the One Child Policy (1979) or any other large scale population control (1970). So it is unclear that there was incentive for respondents or local officials to misreport these numbers. Fudged mortality data would be even more obvious than fudged grain output data, so I feel comfortable offering Peng's excess mortality estimates as another alternative measure. More detailed discussion of the data follows in the next section.

Sen (1981) argues that death rates or fertility may not have a clear relationship with agricultural production in a famine; he notes that a lack of ownership of food is not the same as a lack of production of food. However, the GLF famine can be interpreted as the latter event, and Figure 1 shows an apparent relationship between death rates and grain production (even in perhaps understated data).

There is significant variation of grain output declines at the provincial level; for example, Jiangxi lost only 16.7% of grain production peak-to-trough, while Anhui lost 68.1%. Similarly, there is significant variation in impacts on total fertility rates. Table 1 shows peak-to-trough declines in grain production and estimated lost fertility.

The second category of alternative measures is non-headline agricultural output. Total grain production was the most salient series at the time, so I propose to compare it with other agricultural products, namely tubers and fiber output. I argue that since these statistics were less salient, officials had less incentive or attention to misreport these. Still, this may be problematic, given that central planning policy was changing the composition of agricultural output, so I take this less seriously than the demographic measures.

3.2 Data

Much of the work of this paper is in assembling the data, which come from disparate sources, some of them not public.

3.2.1 Agricultural output

The agricultural output series used are production of total grain, tubers, and fiber. The definition of grain includes cereal crops, potatoes, and beans. Note that tubers and grain have some overlap in potatoes. Fiber crops are typically those used in textiles, cordage, or paper; in provinces where total fiber output was not available, I used cotton output. Counts are the sum of all harvests in a calendar year; this is important over the GLF period in China, during which the central government pushed aggressively for multiple harvests a year, no matter the suitability of conditions. The data are obtained primarily from China's National Bureau of Statistics (NBS). As past literature has often relied on earlier versions of data, I checked these numbers against figures from a USDA-commissioned report by Colby, Crook, and Webb (1992) and did not find any major differences.

3.2.2 Fertility rates

The total fertility rate (TFR) in a year is defined as the average number of children a woman would have if she had contemporaneous age-specific fertility rates throughout her life, assuming she survives through her reproductive life.[1] So the TFR can be interpreted as the birth rate, controlling for the age structure of the overall population over time. TFRs are gathered from tables constructed by Coale & Li (1987), which is based on raw data from the 1982 One-Per-Thousand Sample Fertility Survey. Figure 2 shows the rural TFR and shows the large drop in TFR during the GLF followed by a minor "catch-up" in 1962-63. For reference, the figure also shows grain output over this period; both reach their nadir in 1961.

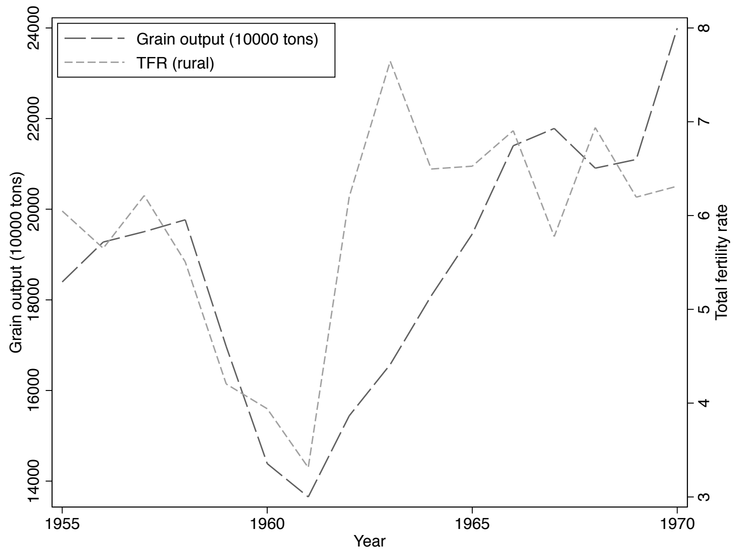
[1] This is potentially problematic given high death rates in the GLF period; for example, if less healthy (and less fertile) women died at higher rates, then TFR will be higher due to selection.

Table 1: Peak-to-trough grain output decline and cumulative TFR loss

Province	Grain output decline	Peak year	Trough year	Cum. TFR loss
Anhui	-0.681	1957	1961	199.5
Beijing	-0.546	1958	1960	123
China	-0.370	1958	1961	109
Fujian	-0.320	1958	1961	77.30
Gansu	-0.756	1957	1960	119.9
Guangdong	-0.256	1957	1960	77.10
Guangxi	-0.237	1957	1963	67.20
Guizhou	-0.566	1958	1961	122
Hainan	-0.326	1957	1960	.
Hebei	-0.408	1957	1963	108.4
Heilongjiang	-0.614	1958	1961	31.10
Henan	-0.656	1958	1961	145.2
Hubei	-0.440	1957	1961	95.40
Hunan	-0.425	1958	1960	132.4
Inner Mongolia	-0.378	1958	1962	63.10
Jiangsu	-0.252	1957	1961	128.3
Jiangxi	-0.167	1957	1961	77.90
Jilin	-0.300	1959	1960	55.80
Liaoning	-0.665	1958	1960	122.3
Ningxia	-0.402	1958	1960	140.4
Qinghai	-0.604	1957	1961	155.9
Shaanxi	-0.354	1958	1961	75.40
Shandong	-0.454	1958	1960	134.1
Shanghai	-0.0207	1958	1961	177
Shanxi	-0.322	1958	1960	43.30
Sichuan	-0.640	1958	1961	186.5
Tianjin	-0.488	1957	1962	113
Xinjiang	-0.147	1959	1962	46.30
Yunnan	-0.256	1957	1960	105
Zhejiang	-0.193	1957	1961	99.90

Notes: Cumulative TFR loss is expressed as percent of a "typical year". Peak-to-trough grain output decline is expressed in log-difference. Source: NBS, Peng (1987).

Figure 2: Grain output and total fertility rates



Source: NBS, Coale & Li (1987)

There is some concern about the accuracy of the data given that the survey comes shortly after the One Child Policy's implementation in 1979. Indeed, survey responses for the late 70s are probably inaccurate. However, it is unclear that there is strong incentive for survey respondents or local officials to misreport data from earlier periods. There may have been incentive to inflate earlier numbers in order to emphasize the success of population control; as long as this only occurs sufficiently after the GLF period or is done consistently across years, this does not affect the results.

Lump-sum estimates of lost fertility from the GLF come from Peng (1987). Peng uses the same underlying data to construct his estimates, which are based on falls from pre-GLF fertility rates (see the reference for full details).

3.2.3 Other demographic data

Crude death rates (CDR) are defined as deaths per 1,000 in a year. Both the time series and lump-sum excess CDRs are from Peng (1987). The time series cover 1957-1963, though the data for 1957 are actually the averages of 1956-57.

3.3 Empirical tests

The non-structural test takes the form of OLS estimates of equations

$$\%\Delta GrainOutput_i = \beta \cdot \Delta A_i$$

where the subscript i refers to the province, $\%\Delta GrainOutput_i$ is peak-to-trough percentage decline in grain output, and A_i is an alternative measure of famine severity. In these tests, I reduce the time series to cross-sections; the ΔA_i's are estimates of cumulative lost fertility or excess mortality and peak-to-trough declines in other agricultural output. β is the parameter of interest; if I find that the variables are only weakly correlated, that is an indication that the official data are grossly wrong.

The structural version uses the panel data in a Chow test for a structural break in the relationship between

$$GrainOutput_{i,t} \sim \beta \cdot A_{i,t} + \mathbf{X}_i \gamma'$$

where $GrainOutput_{i,t}$ is grain output in province i at time t. To combine the different alternative measures into a single parsimonious specification, I use principal component analysis (PCA). $A_{i,t}$ is the first principal component, reduced from fiber output, tuber output, and rural TFR.[2] Since it is not clear that TFR and grain output should have a linear relationship, I also include a quadratic TFR term in the PCAs. $\mathbf{X}_i$ is a vector of control variables consisting of province fixed effects and time fixed effects. The time period considered is from 1953 to 1969; the end date is chosen as the last year before population control policies, and the start date is chosen to mirror the 5 years after the GLF end year. I test for a single break in either 1958 or 1962 and for both breaks.

3.4 Results

Table 2 shows the results of the non-structural tests. The peak-to-trough drop in grain output closely matches the alternative measures. Figure 3 also shows the equivalent scatterplots. With the exception of perhaps the CDR specification, the results do not hinge on single influential observations. The results provide essentially no positive evidence that the grain production data are inaccurate. The specifications strongly reject zero correlation between grain production and the alternative measures; even the R^2 are surprisingly high given the simplicity of the specifications.

The results of the Chow tests for structural breaks are more nuanced. When testing for breaks in either 1958 or 1962, the Chow test strongly rejects no break at the 99% level. The test for both breaks also rejects at the 99% level. The results suggest that there was a change in the relationship between the time series in grain output and the PCA-generated alternative measure during the GLF.

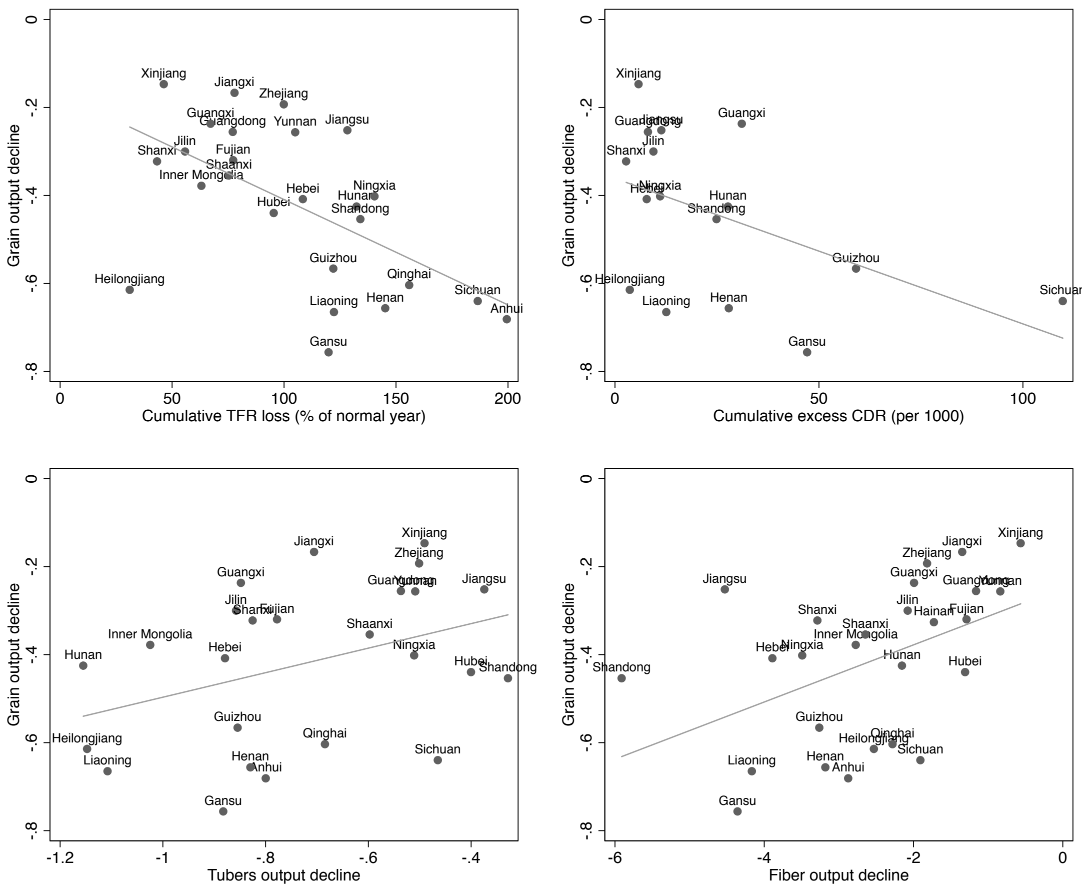
[2] Insufficient data history prevents me from using the crude death rate as well.

Table 2: Non-structural tests against alternative measures

	Grain output decline			
Cumulative TFR loss (% of normal year)	-0.00239***			
	(0.000723)			
Cumulative excess CDR (per 1000)		-0.00331**		
		(0.00114)		
Tubers output decline			0.279**	
			(0.122)	
Fiber output decline				0.0649**
				(0.0272)
Constant	-0.170*	-0.361***	-0.218**	-0.248***
	(0.0905)	(0.0552)	(0.0955)	(0.0675)
Observations	25	16	25	25
R-squared	0.332	0.245	0.141	0.218

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 3: Non-structural tests against alternative measures



Note: Figure shows scatterplots of peak-to-trough grain output decline against alternative measures of famine severity. See Section 3.1 for details.

However, the Chow tests for a single break actually reject at the 95% level against a break in *any* year in the sample, though the test statistic is highest in 1958. This suggests that my specification misses a fundamental part of the relationship between grain output and the alternative measures, though it does capture some important part of the relationship, since the strongest break date is the ex ante expectation of the beginning of the GLF. To help visualize this test, Figure 4 shows the scatterplots of grain output and the alternative measure grouped into the relevant time periods by color. It appears that the grouped data indeed have different relationships, though this is far from obvious.

Given the null result in the non-structural test and the ambiguity of the Chow tests, I suggest the official statistics be taken as given. Further research could pursue refinements of the structural test. The PCA approach could be easily applied to richer data, which might better capture the true relationship.

4 Long-term impacts of the GLF

The second task of the paper is to evaluate the long-term impacts of the GLF. I interpret the GLF as a shock on province economic development, and I adapt the model from Davis & Weinstein (2002) to measure its impact. Let $y_{i,t}$ be province economic development; I will call it industrialization. Let each province have an initial industrialization level Γ_i, and let province-specific shocks be $\varepsilon_{i,t}$. Then write industrialization at time t as

$$y_{i,t} = \Gamma_i + \varepsilon_{i,t} \tag{1}$$

I model persistence in the shocks as

$$\varepsilon_{i,t+1} = \rho\varepsilon_{i,t} + \nu_{i,t+1} \tag{2}$$

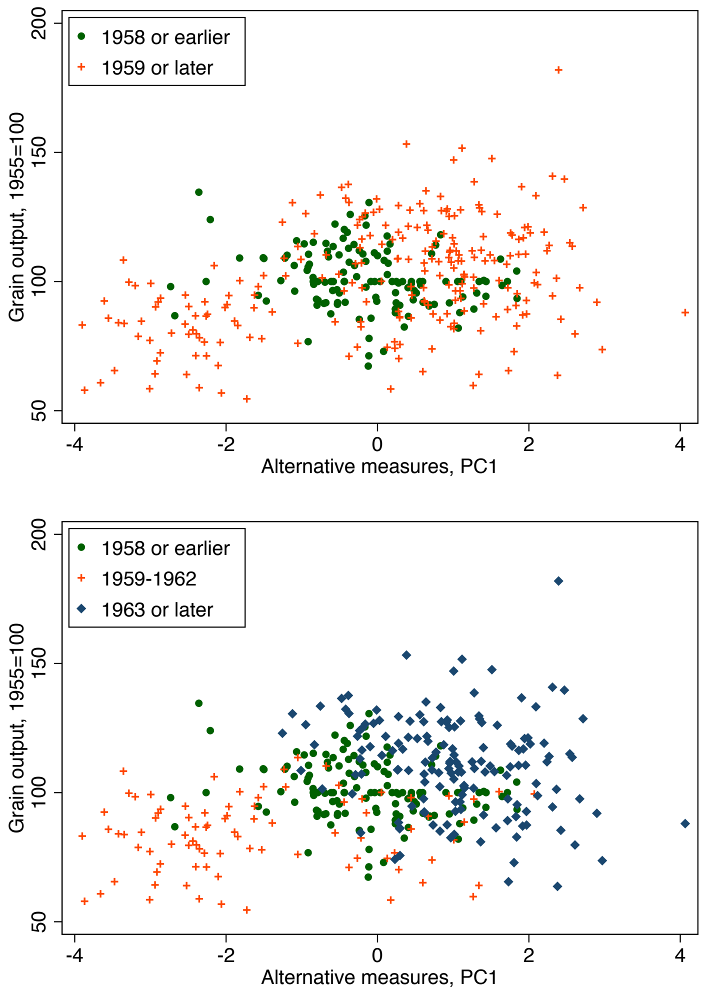
where the “innovations”, $\nu_{i,t+1}$, are normally distributed error terms.

Differencing equation (1) and substituting in (2) yields

$$\begin{aligned} y_{i,t+1} - y_{i,t} &= \varepsilon_{i,t+1} - \varepsilon_{i,t} \\ &= (\rho - 1)\nu_{i,t} + [\nu_{i,t+1} + \rho(1-\rho)\varepsilon_{i,t-1}] \end{aligned} \tag{3}$$

ρ is the parameter of interest. If $\rho = 1$, then shocks are permanent and should have no relationship with future growth rates in industrialization. If $\rho < 1$, then there should be negative correlation between the shock and future growth rates in industrialization. I interpret $\nu_{i,t}$ as the innovation of the GLF shock, so $\nu_{i,t+1} = 0$. The most direct way of measuring this is to use the change in industrialization over the GLF period. However, autocorrelation in industrialization growth rates

Figure 4: Grouped scatterplots of grain output and the PCA alternative measure



would mean that this number includes information on past industrialization, and may bias the estimate in an unknown direction. So I use cumulative lost TFR as a 2SLS instrument for change in industrialization over the GLF period. This does require the (perhaps heroic) assumption that fertility is not correlated in the same way with past industrialization.

In the actual attempted estimations, $y_{i,t}$ is industrialization in 1964, and $y_{i,t+1}$ is industrialization in 1982, 1990, and 2000. Change in industrialization over the GLF is from 1953-1964. These dates are governed by data restrictions, and are obviously problematic. The measured GLF period is much wider than the actual event, which is from 1958-62. The post-GLF industrialization data points are much later, and may capture events such as the Cultural Revolution. The whole exercise admittedly has an air of delicacy. The actual estimations are of the form

$$y_{i,post} - y_{i,1964} = (\rho - 1) \cdot (y_{i,1964} - y_{i,1953}) + \rho(1 - \rho)\varepsilon_{i,1953}$$

where $y_{i,1964} - y_{i,1953}$ is instrumented by $\Delta_{cumulative} TFR_{1958-1962}$ in the first stage.

4.1 Industrialization data

Industrialization is measured as the share of the population employed in non-agricultural jobs. The data are from Bai & Wei (2019), who use Chinese censuses from 1964, 1982, 1990, and 2000, supplemented with local gazettes from 1953. Using this definition of industrialization is of course problematic; industrial production would be a more natural variable to use. However, such series are not available at province levels in China before the 1980s. Bai and Wei's data are at the county level; I aggregate to province level by taking population-weighted averages. Due to data restrictions, I use county populations in 2010. This comes with an obvious concern - more industrialized counties probably grew faster. This measure also endogenizes the fact that central government policy shifted labor toward industry, no matter the poor productivity of it. Future refinements of this paper can use more precisely constructed versions of these series if the data become publicly available (or can be re-assembled from the underlying censuses).

4.2 Results

Unfortunately, a weak instrument problem prevents the actual implementation of this design. This is perhaps not surprising, given the (non-)richness of the industrialization data. In fact, most provinces show almost zero change in the percent of employment in non-agriculture from 1953 to 1964, so it is unsurprising that cumulative fertility loss is a very weak instrument. Figure 5 shows the industrialization series; the display is rather clumsy, but adequately shows little movement from 1953 to 1964, with large increases after. Table 3 shows the results of the first stage; it can be seen here that the coefficient is close to 0 with large a standard error.

Table 3: Long run impact 2SLS, first stage

VARIABLES	Change in % non-agricultural, 1953-1964
Cumulative TFR loss (% of normal year)	0.000320
	(0.00611)
Constant	0.293
	(0.774)
Observations	24
R-squared	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

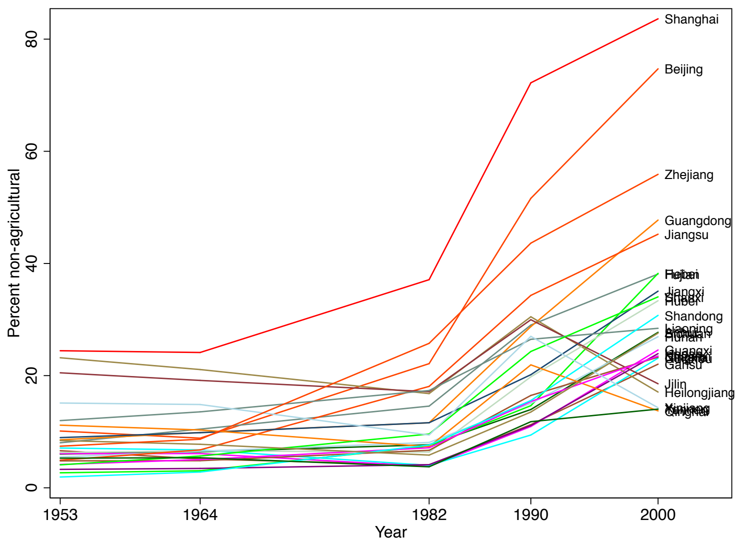
In future research, this model could be applied to better industrialization data. Recall that the series are aggregated using 2010 population weights, and that the dates are governed by the dates of the censuses. More sensibly constructed data and dates that more closely align with the actual GLF period would likely yield better results.

5 Conclusion

In this paper, collect and investigate data on the Great Leap Forward and make a partial attempt at investigating its long-term impacts on development. Data quality is an important factor in social science research on this period. While it is unclear whether Chinese authorities have fully corrected the historical data on agricultural production, much research takes it as given for lack of clear alternatives. I assess grain production against non-headline agricultural production and demographic data via both simple cross-sectional regressions and Chow tests applied to the panel data. I do not find much positive evidence that the grain production series are wrong, though the Chow tests indicate that there may be irregularities during the Great Leap Forward. The specification for the Chow tests misses some important part of the relationships (as it suggests breaks in every year), though it does suggest that 1958 and 1962 (the GLF bounds) are special dates.

I also present a framework to assess long-term impacts of the Great Leap Forward on industrialization, adapted from Davis & Weinstein (2002). Data quality prevents implementation of this framework, though I am hopeful that richer data in the future would enable implementation.

Figure 5: Industrialization, as percent of non-agricultural population



References

Bai, Liang and Lingwei Wu. 2019. "Political Conflict and Development Dynamics: Economic Legacies of the Cultural Revolution."

Becker, Jasper. 1998. *Hungry Ghosts: Mao's Secret Famine.* New York: Holt.

Coale, Ansley J. and Chen Sheng Li. 1987. "Basic Data on Fertility in the Provinces of China, 1940-1982." Papers of the East-West Population Institute, 104.

Colby, W. Hunter, Frederick W. Cook, and Shwu-Eng H. Webb. 1992. *Agricultural Statistics of the People's Republic of China, 1949-90.* U.S. Department of Agriculture.

Davis, Donald, R., and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." American Economic Review, 92 (5): 1269-1289.

Li, Lillian M. 1982. "Food, Famine, and the Chinese State." The Journal of Asian Studies, 41:4, 687-707

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

Naughton, Barry. 2007. *The Chinese Economy: Transitions and Growth.* Cambridge: MIT Press.

Peng, Xizhe. 1987. "Demographic Consequences of the Great Leap Forward in China's Provinces." Population and Development Review, 13:4, 639-670.

Sen, Amartya. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation.* New York: Oxford University Press.

Yang, Jisheng. 2010. "The Fatal Politics of the PRC's Great Leap Famine: the preface to Tombstone." Journal of Contemporary China, 19:66, 755-776.

Zhu, Xiaodong. 2012. "Understanding China's Growth: Past, Present, and Future." Journal of Economic Perspectives, 26:4, 103-24.