The Cause of China's Great Leap Famine Revisited Using Panel Data Approach for Program Evaluation

Ying Fang James Kung Yang Yang

July 9 2014, Japan

Content

- Introduction
- Methodology
  - Panel Data Approach for Program Evaluation
  - Illustrative Examples
- Empirical Results
  - The Political Impact Index (PII)
  - Mortality Variation across Provinces
  - PII versus Radical Policies
- Conclusion

Introduction

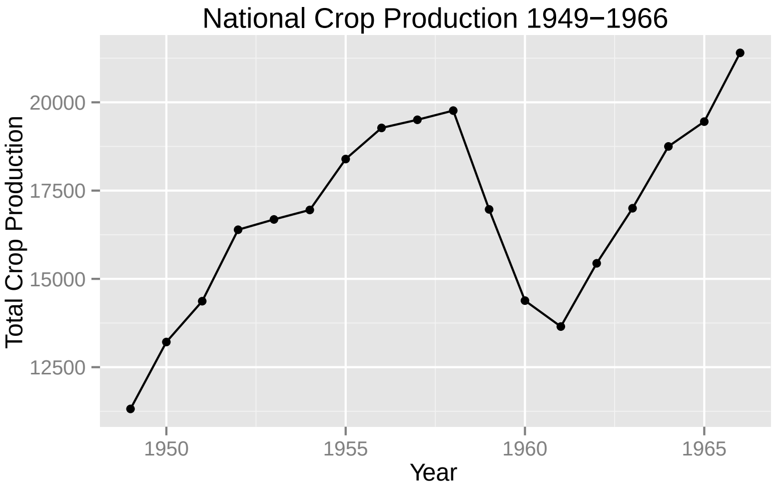
- The Great Leap Forward (GLF) refers to an economic and social campaign occurred in China from 1958 to 1961, which aimed to radically transforming the country from an agrarian economy into a communist society through rapid industrialization and collectivization.

- *People's Daily* : The GLF would propel China to surpass the Great Britain in industrial production in 15 years and the United States in 20 or 30 years (the New Year's Editorial of 1958).

The Consequences of GLF:

- a sharp collapse of crop production
- a catastrophic famine

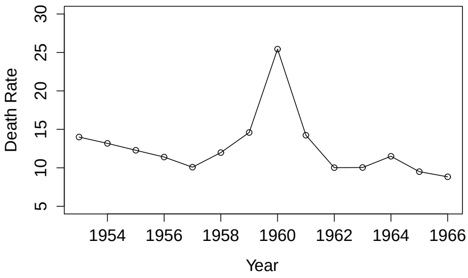
Introduction



Introduction

The estimates of premature deaths during the Great Leap Famine

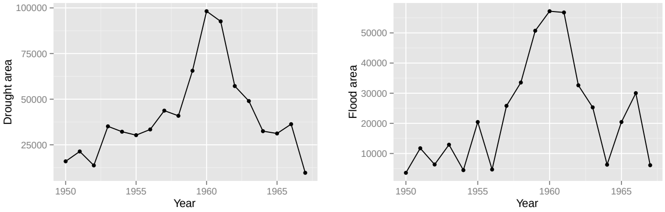
- Coale (1981): 16.5 million
- Yao (1999): 18.48 million
- Peng (1987): 23 million
- Ashton, et al. (1984): 29.5 million
- Banister (1987): 30 million



The Cause of the Famine

One usual explanation puts the blame mainly on the bad weather and the GLF period of 1959-61 is often called the "Three Years of Natural Calamities".

Figure: **The affected area by natural disasters in state level 1950∼1967**



Debates on the Cause of the Famine

- Kueh (1995): Bad weather was a contributing factor. But bad weather of similar magnitude in the past did not produce such a serious reduction in aggregate grain output.
- Lin (1990): Government may ... use the natural disaster as an excuse.
- Jin (1998): Weather condition of the three years is even better than the normal years.
- Li and Yang (2005): Government could change or falsify statistical data to cover the impact of radical policy.

The Cause of the Famine – Radical Policies

- Diversion of resources
- Political and institutional factors

Diversion of Resources

Diversion of Agricultural Resources	Papers
Incremental of Steel Production	Li and Yang(2005), Meng, Qian and Yared (2013
Gross Domestic Product (GDP) Per Capita	Kung and Chen(2011)
Reduction of Sown Areas	Li and Yang(2005), Meng, Qian and Yared (2013)
Proportion of Rural Population	Lin and Yang(2000), Kung and Lin(2003)
Proportion of Agricultural Income	Kung and Chen (2011)
Per Capita Grain Output in Rural Areas	Lin and Yang(2000), Kung and Lin(2003), Kung and Chen (2011), Meng, Qian and Yared (2013)

Political and Institutional Factors

Radicalism Chaos in Rural Area	Papers
Number of Mess Hall	Yang (1998)
Mess Hall Participation Rate	Chang and Wen (1997), Yang (1998),Kung and Lin (2003), Meng, Qian and Yared (2013)
Number of Agricultural Satellites	Kung and Chen(2011), Fan and Shi (2013)
Party Membership Density	Yang (1998), Kung and Lin(2003), Kung and Chen(2011)
Magnitude of the Anti-Right Purges	Kung and Chen(2011),Meng, Qian and Yared (2013)
Time of Liberation	Yang (1998), Kung and Chen(2011)
Net Procurement Ratio	Kung and Lin (2003)
Excess Procurement Ratio	Kung and Chen (2011)

Famine Theories

- The Food Availability Decline (FAD) Theory was firstly proposed by Adam Smith in *The Wealth of Nations*. It explains the cause of a famine as a decrease of the food supply. The FAD theory focuses on the factors that could affect agricultural production process such as weather conditions, incentive system, labor transfer and among others.

Famine Theories

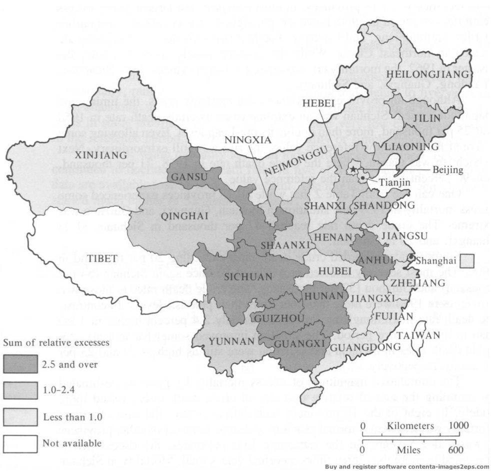
- The Entitlement Approach, proposed by Amartya Sen, considers the endowment vector of individuals, including the ownership of land, labor, physical condition and other property rights which could be used to produce or exchange for food. There comes a famine when both ways fail to work. Hence, a reduction of the endowment (a failure of rights to directly acquire food), an adverse change on food price (a failure of exchange rights to acquire food), and a distorted food allocation system could all cause a famine.

Unsolved problems in the existing literature:

- Although many scholars believe that policies play a more important role than bad weather in explaining the collapse of production, we do not know the real impact of radical policies on the collapse of crop production and the great famine.
- The existing literature is hard to explain the spatial variation of mortality across provinces during the GLF period.
- Many policy proxies are mentioned in the literature but lack of a comprehensive framework to assess their impacts and significance.

Unsolved problems in the existing literature:

Figure: Levels of excess mortality, China, 1958-62 by Peng(1987)



What we want to do?

- Using panel data program evaluation (Hsiao, et al., 2012), we quantitatively estimate the impact of policies on the collapse of crop production.
- We construct an aggregate measure of the extent of radical policies in each province during the GLF period and use it to effectively explain the cross-province mortality variation.
- We estimate the contribution of each policy proxy mentioned in the existing literature to the aggregate measure of the extent of radical policies.

Methodology

- The main difficulty lies in the fact that we cannot observe China's crop production without affected by radical policies during the GLF period.
- We estimate the counterfactual using the panel data approach for program evaluation proposed by Hsiao, et al. (2012).

Methodology

- Let Y_{it}^0 denote the crop production in region/country i at time t. Suppose that the crop production is generated by the following factor model:

$$Y_{it}^0 = \alpha_i + \beta_i^T f_t + \epsilon_{it}.$$

- Assume that a policy intervention occurred at time T_1, then the effect of the policy intervention is defined by $\Delta_{it} = Y_{it}^1 - Y_{it}^0$, where Y_{it}^1 and Y_{it}^0 denote the crop production with and without policy intervention.
- However, since we cannot observe Y_{it}^1 and Y_{it}^0 simultaneously, the observed data are in form $Y_{it} = d_{it} Y_{it}^1 + (1 - d_{it}) Y_{it}^0$, where $d_{it} = 1$ if unit i is under treatment at time t, and $d_{it} = 0$ otherwise.

Methodology

- Assume the first unit receives the policy intervention and other units not. Starting from period T_1, the crop production of the first unit is given by $Y_{it}^1 = \alpha_1 + \beta_1^T f_t + \Delta_{1t} + \epsilon_{1t}$.
- For other units and for all time periods t, we have $Y_{it}^0 = \alpha_i + \beta_i^T f_t + \epsilon_{it}$ for $i \neq 1$ and $t = 1, \cdots, T$.
- Hence, for $t \leq T_1$, we have $Y_t^0 = BF_t + \alpha + \epsilon_t$ where $Y_t^0 = (Y_{1t}^0, \cdots, Y_{Nt}^0)^T$, $\alpha = (\alpha_1, \cdots, \alpha_N)^T$, $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{Nt})^T$, and B is the loading matrix.

Methodology

- Note that $Y_t = (Y_{1t}, \tilde{Y}_t)^T$ where $\tilde{Y}_t = (Y_{2t}, \cdots, Y_{Nt})^T$.
- Assume that $a = (1, -\tilde{a})$ satisfying $a^T B = 0$, then we can have
$$a^T Y_t \equiv Y_{1t} - \tilde{a}^T \tilde{Y}_t = a^T \alpha + a^T \epsilon_t.$$
- Then we can predict Y_{1t}^0 by the regression model $Y_{1t}^0 = \gamma_0 + \gamma^T \tilde{Y}_t$. In other words, in stead of estimating common factors, we can take advantage of the cross-sectional dependence in observed crop production of other units to predict the counterfactual crop production.

Methodology

- $E(\epsilon_{js} \mid d_{it}) = 0$ for $j \neq i$.
- heterogenous responses to the unobservable common factors.
- least data demand.
- only requiring reasonable T.

Methodology

- The ideal situation is to find units in the control group which share the unobservable common factor coming from the climate but not affected by radical policies during the period of 1959-1961.
- The units in the control group can response to the climatic shocks in a heterogenous way which means the units can have different systems and agricultural technology.
- We select American states in the same period based on the so called agro-climatic similitude ranges.

The Agro-climatic Similitude Range (Wei and Liu, 1994)

Agro-climatic Similitude Range is the method using a space range system to quantitatively, synthetically and sequentially measure the bioclimatic similitude between two or several regions. It takes climatic factors, such as temperature and rain falls, as different dimensions in the space and use 3-D sliding analogy to measure the similitude.

The Computation of Agro-climatic Similitude Range

$$d_{ij} = \sqrt{\frac{1}{m}\sum_{k=1}^{n}\left(X'_{ik} - X'_{jk}\right)}$$

where d_{ij} is the agro-climatic similitude range between regions i and j; $X^{'}_{ik}$ and $X^{'}_{jk}$ are the k-th standardized climatic factor for regions i and j, respectively; $\frac{1}{m}$ is a normalizing multiplier.

Methodology

- Using the panel data approach for program evaluation of Hsiao, et al. (2012), we estimate the counterfactual crop production for each province by selecting American states whose agro-climatic similitude range with that province is less than 1.

- The identification strategy is that these American states share the unobserved common factors of agro-climate but are not affected by all these radical policies implemented in China during the GLF period.

Methodology

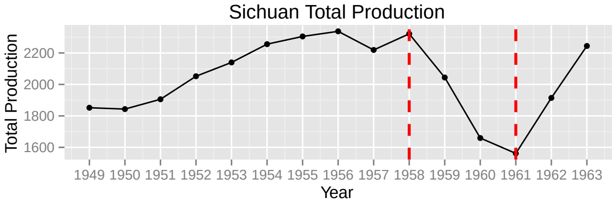
- We illustrate our methodology using the example of Sichuan province.
- When comparing crop production between China and the U.S., we only consider corn, rice, wheat, soybeans and potato, not including sorghum and buckwheat.
- We estimate the crop production per hectare rather than the total crop production in order to excluding the fluctuation in seeded areas across years and difference in seeded areas among regions.

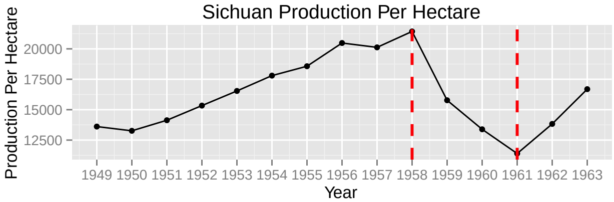
Methodology

Table: **Unit Conversion Table**

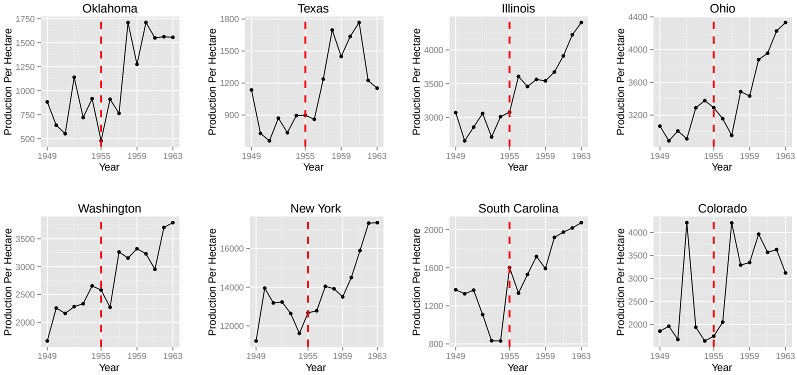
Crops	British Unit	Metric Unit
Corn Production	1 Bushel	25.40 kilogram
Rice Production	1 Bag	45.36 kilogram
Wheat Production	1 Bushel	27.22 kilogram
Soybean Production	1 Bushel	27.22 kilogram
Potato Production	1 Bushel	27.22 kilogram
Seeded area	1 Acre	0.4047 hectare

Choice of Cross-sectional Units for Sichuan Province: the total crop production and crop production per hectare of Sichuan province



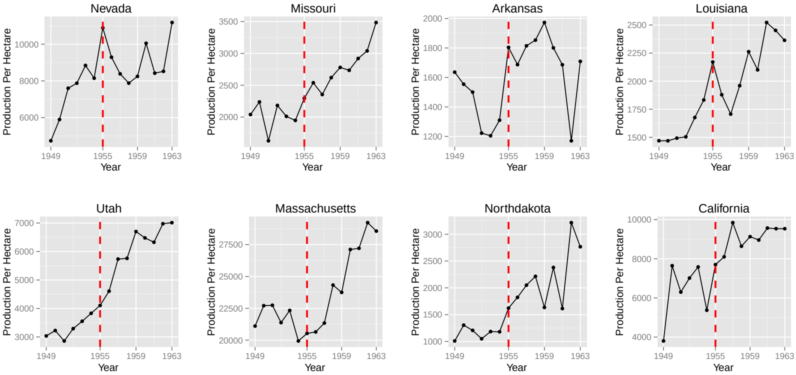


Choice of Cross-sectional Units for Sichuan Province: American states



Methodology

Choice of Cross-sectional Units for Sichuan Province: American states



The Crop Production per Hectare of Sichuan Province

Table: The Crop Production per Hectare of Sichuan Province and Its Control Group

Year	Sichuan	New York	Nevada
1952	2,051.895	13,234.070	7,875.870
1953	2,139.844	12,640.690	8,841.158
1954	2,255.906	11,610.260	8,147.118
1955	2,304.933	12,066.620	10,871.890
1956	2,337.516	11,563.180	9,252.979
1957	2,218.996	12,220.240	8,328.229
1958	2,321.550	11,485.200	7,803.826
1959	2,043.947	10,455.630	8,154.304
1960	1,659.203	10,848.890	9,939.173
1961	1,561.392	11,633.660	8,291.559

The red-colored data are de-trended. The unit is thousand kilogram per hectare.

Regression Result

	Dependent variable:
	Sichuan
New York	-0.148^{***}
	(0.016)
Nevada	0.028^{**}
	(0.009)
Constant	$3,775.756^{***}$
	(212.938)
Observations	7
R^2	0.964
Adjusted R^2	0.946
Residual Std. Error	$24.377 (df = 4)$
F statistic	$53.649^{***} (df = 2; 4)$
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Methodology

Compute the counterfactual total crop production

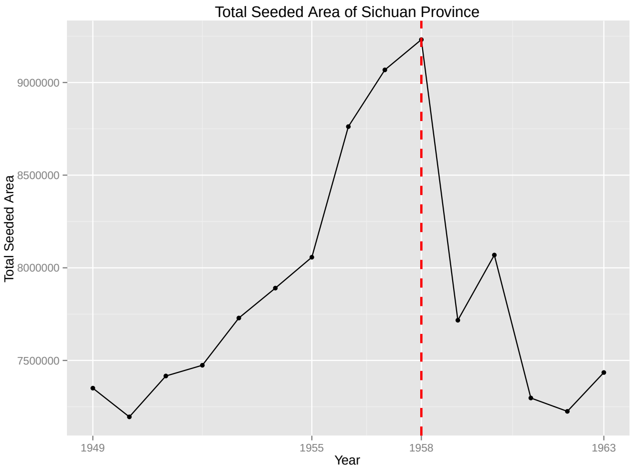
- We use the counterfactual crop production per hectare to compute the counterfactual total production:

$$Total\ Production_{Counterfactural} = Production\ per\ Hectare_{Counterfactural} \times \sum Seeded\ Area_{i,1958}$$

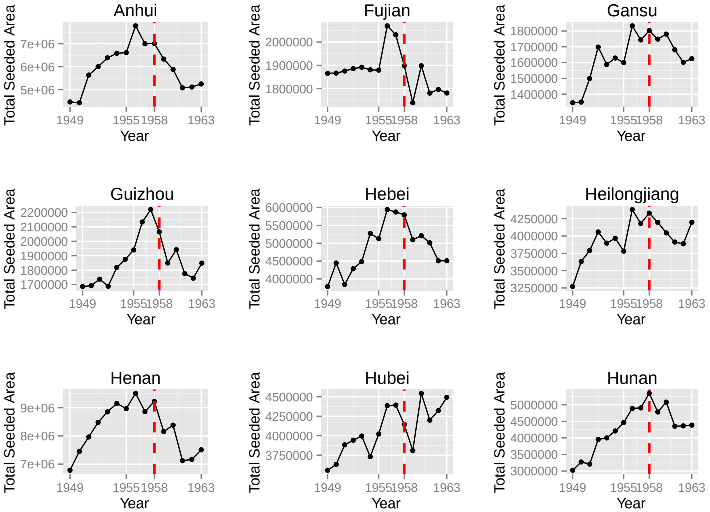
where i $\subset$ *Corn*, *Rice*, *Wheat*, *Soybean*, *Potato*

- Using the seeded areas in 1958 to eliminate the impact of the decrease of seeded areas due to radical policies during the GLF period.

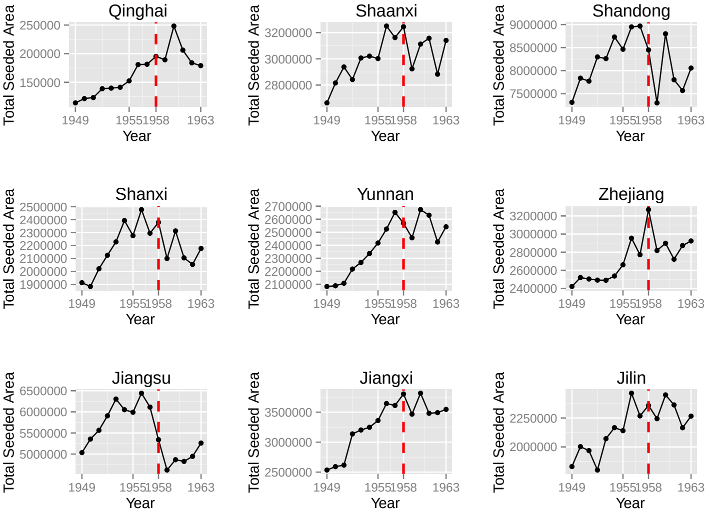
The Decrease of Seeded Areas in Sichuan province



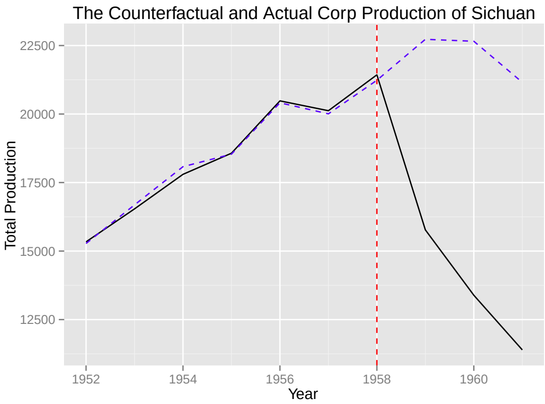
The Decrease of Seeded Areas During the GLF Period



The Decrease of Seeded Areas in China During the GLF Period



Counterfactual Crop Production versus Actual Crop Production



Methodology

We measure the extent of radical policies in each province by constructing a so called *Political Impact Index*$_t$:

The Political Impact Index

$$Impact\ Index_t = \frac{y_t^{counterfactual} - y_t^{actual}}{y_t^{counterfactual}}$$

The *Political Impact Index* of Sichuan Province,
$t = 1958, 1959, 1960, 1961, 1962$

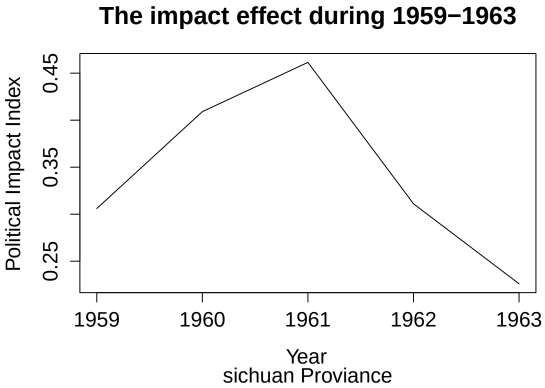
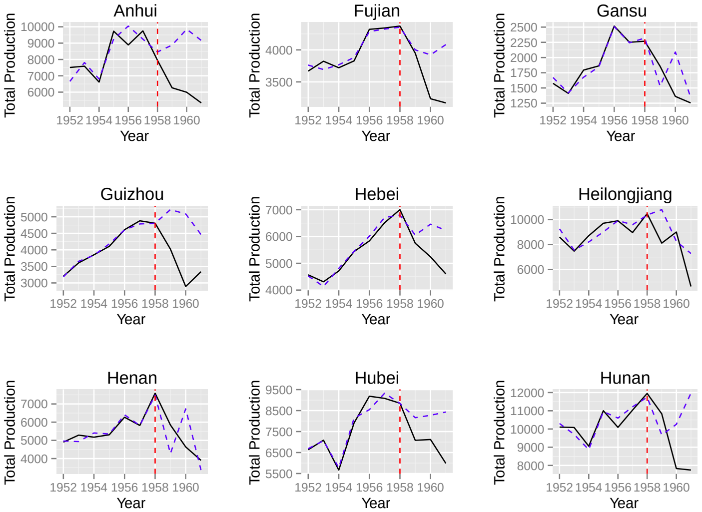


Table: The Choice of Cross Sectional Units

Province	R^2	Adjusted R^2	Control Group 1	Control Group 2
Jiangxi	0.970	0.955	Louisiana	New York
Guizhou	0.967	0.950	New York	Minnesota
Sichuan	0.964	0.946	New York	Nevada
Gansu	0.953	0.930	North Dakota	Illinois
Henan	0.945	0.917	Oklahoma	North Dakota
Qinghai	0.944	0.916	Massachusetts	Ohio
Jiangsu	0.940	0.910	Arkansas	Minnesota
Hebei	0.921	0.881	Colorado	Georgia
Shanxi	0.915	0.872	Minnesota	Oklahoma
Hubei	0.914	0.871	Georgia	California
Fujian	0.896	0.844	Georgia	Nevada
Hunan	0.878	0.817	Tennessee	New York
Yunnan	0.875	0.813	Colorado	Massachusetts
Shandong	0.833	0.750	Missouri	New York
Shaanxi	0.808	0.712	Texas	Florida
Zhejiang	0.793	0.690	Washington	Texas
Jilin	0.786	0.743	New York	——
Anhui	0.733	0.599	South Carolina	New York
Heilongjiang	0.630	0.445	Washington	Missouri

Empirical Results

Counterfactual Corp Production versus Actual Corp Production



Counterfactual Corp Production versus Actual Corp Production

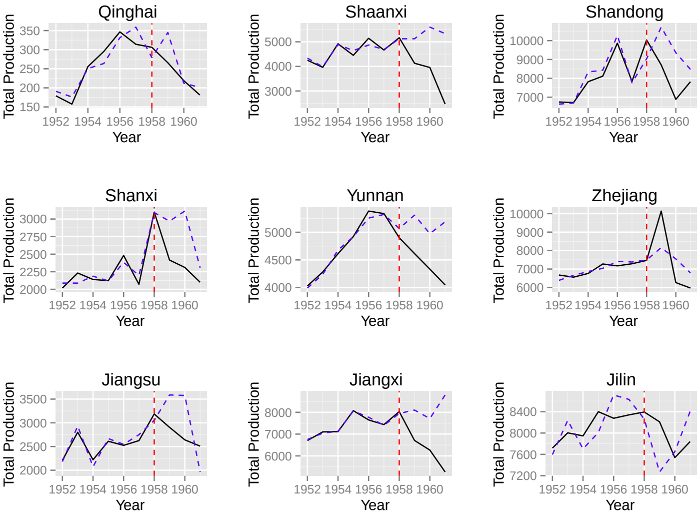


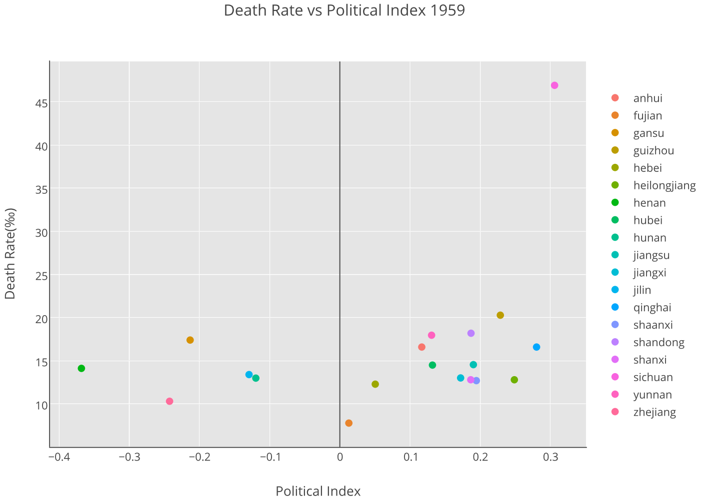
Table: The Provincial-level Political Impact Index

Province	Index 1959	Index 1960	Index 1961	Three Year Average Index
Jiangxi	0.172	0.187	0.402	0.254
Guizhou	0.229	0.431	0.252	0.304
Sichuan	0.306	0.409	0.461	0.392
Gansu	−0.213	0.349	0.067	0.068
Henan	−0.368	0.311	−0.158	−0.072
Qinghai	0.583	0.449	0.488	0.507
Jiangsu	0.190	0.262	−0.278	0.058
Hebei	0.051	0.189	0.261	0.167
Shanxi	0.187	0.258	0.090	0.178
Hubei	0.132	0.140	0.290	0.187
Fujian	0.013	0.175	0.223	0.137
Hunan	−0.120	0.237	0.353	0.156
Yunnan	0.131	0.129	0.220	0.160
Shandong	0.187	0.264	0.077	0.176
Shaanxi	0.195	0.293	0.538	0.342
Zhejiang	−0.243	0.170	0.121	0.016
Jilin	−0.129	0.014	0.068	−0.016
Anhui	0.117	0.277	0.401	0.265
Heilongjiang	0.249	−0.090	0.363	0.174

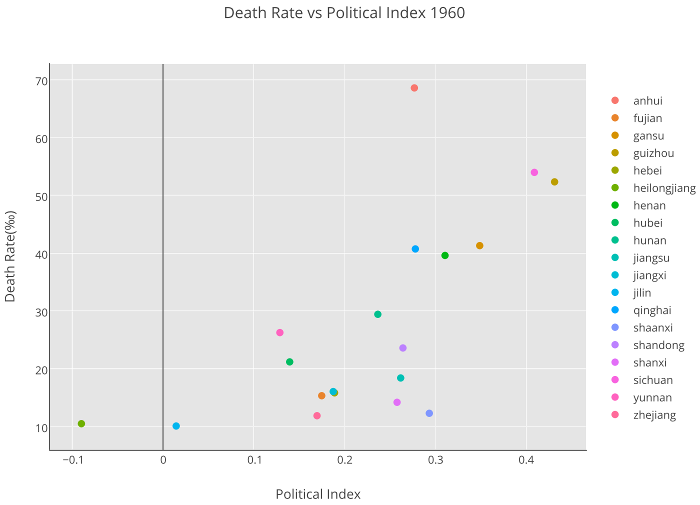
Table: The National Political Impact Index

	1959	1960	1961	Three-year Average Index
Index	0.1129	0.2408	0.2815	0.2117

The Scatter Plot: Death Rates *v.s.* Political Impact Index 1959



The Scatter Plot: Death Rates *v.s.* Political Impact Index 1960



The Scatter Plot: Death Rates *v.s.* Political Impact Index 1961

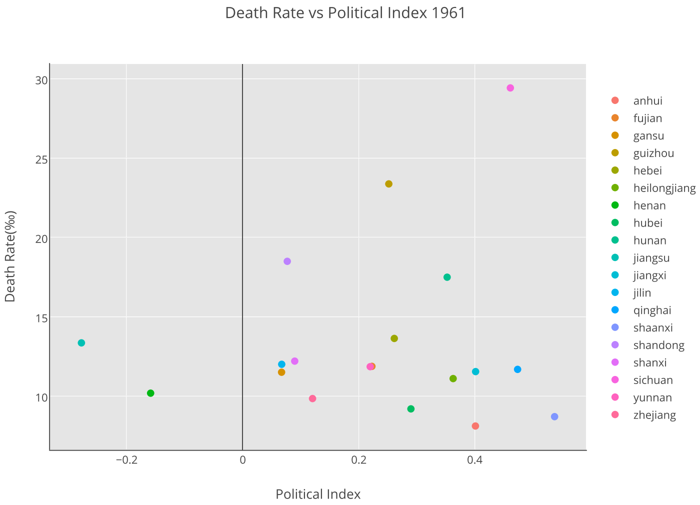


Table: Means and Standard Deviations of Death Rates

Statistic	N	Mean	St. Dev.	Min	Max
Year 1959					
Death Rate	19	16.062	8.037	7.780	46.910
Political Index	19	0.072	0.195	-0.368	0.306
Year 1960					
Death Rate	19	27.453	17.188	10.100	68.580
Political Index	19	0.225	0.125	-0.090	0.431
Year 1961					
Death Rate	19	13.448	5.338	8.110	29.420
Political Index	19	0.222	0.213	-0.278	0.538

Table: Means and Standard Deviations of Death Rates (Excluding Sichuan)

Statistic	N	Mean	St. Dev.	Min	Max
Year 1959					
Death Rate	18	14.348	3.051	7.780	20.280
Political Index	18	0.059	0.192	-0.368	0.280
Year 1960					
Death Rate	18	25.980	16.406	10.100	68.580
Political Index	18	0.215	0.120	-0.090	0.431
Year 1961					
Death Rate	18	12.561	3.786	8.110	23.370
Politicl Index	18	0.209	0.211	-0.278	0.538

Table: Linear Regression Results: 1959

	Dependent variable:				
	Death Rate				
	(1)	(2)	(3)	(4)	(5)
PII	15.224		15.409	16.635	14.000
	(0.185)		(0.182)	(0.176)	(0.223)
Net.Procurement.Ratio		0.056	−0.020	0.050	0.314
		(0.696)	(0.882)	(0.837)	(0.387)
Ln.Grain.Production				3.468	3.529
				(0.396)	(0.373)
PMD					0.518
					(0.872)
MHPR1959					0.119
					(0.291)
TOL					0.412
					(0.391)
Constant	14.969***	14.626**	15.464**	−2.917	−25.613
	(0.000)	(0.000)	(0.000)	(0.898)	(0.463)
R^2	0.136	0.003	0.137	0.228	0.429
Adjusted R^2	0.085	−0.055	0.029	0.073	0.144
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table: Linear Regression Results: 1960

	Dependent variable:					
	Death Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
PII	90.978***		115.727***	115.667***	116.199**	93.487**
	(0.000)		(0.000)	(0.001)	(0.001)	(0.010)
Lag(PII)				0.129	0.220	17.557
				(0.990)	(0.984)	(0.472)
Net.Procurement.Ratio		−0.604***	0.419	0.418	0.426	0.236
		(0.001)	(0.174)	(0.203)	(0.228)	(0.478)
Ln.Grain.Production					0.321	−6.021*
					(0.911)	(0.095)
PMD						−14.621
						(0.129)
MHPR1959						0.201
						(0.273)
TOL						−1.232
						(0.294)
Constant	6.949	39.466***	−6.964	−6.944	−8.691	55.562
	(0.142)	(0.000)	(0.548)	(0.573)	(0.678)	(0.194)
R^2	0.439	0.129	0.468	0.468	0.468	0.650
Adjusted R^2	0.406	0.078	0.402	0.362	0.317	0.428
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table: Linear Regression Results: 1961

	Dependent variable:					
	Death Rate					
	(1)	(2)	(3)	(4)	(5)	(6)
PII	2.942		3.056	2.738	4.418	4.226
	(0.623)		(0.627)	(0.623)	(0.461)	(0.553)
Lg(PII)				34.916^{*}	33.507^{**}	28.175
				(0.062)	(0.049)	(0.139)
Net.Procurement.Ratio		−0.082	−0.085	0.285	0.241	0.160
		(0.274)	(0.279)	(0.169)	(0.198)	(0.484)
Ln.Grain.Production					1.984	3.173
					(0.179)	(0.146)
PMD						2.324
						(0.267)
MHPR1959						−0.024
						(0.678)
TOL						0.335
						(0.183)
Constant	12.794^{***}	14.777^{***}	14.139^{***}	0.353	−7.989	−18.142
	(0.000)	(0.000)	(0.000)	(0.962)	(0.449)	(0.179)
R^2	0.014	0.017	0.032	0.353	0.422	0.487
Adjusted R^2	−0.044	−0.041	−0.089	0.224	0.257	0.160
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table: Panel Data Results: 1959-1961

	Dependent variable:						
	Death Rate						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PII	15.762*		$-$22.229	$-$4.226	$-$3.986	20.561**	20.379**
	(0.069)		(0.243)	(0.806)	(0.813)	(0.048)	(0.037)
Lag(PII)			$-$56.007***	$-$41.448**	$-$41.096**		
			(0.003)	(0.022)	(0.020)		
Net.Procurement.Ratio		0.060		1.219***	1.161**	0.455	0.072
		(0.723)		(0.004)	(0.026)	(0.126)	(0.818)
Ln.Grain.Production						$-$6.796	$-$11.786
						(0.646)	(0.437)
GDP.per.capita					0.010		0.136*
					(0.806)		(0.091)
Observations	57	57	38	38	38	57	57
Note:					*p<0.1; **p<0.05; ***p<0.01		

Table: Panel Data Results of Policy Proxies

	Dependent variable:					
	PII					
	(1)	(2)	(3)	(4)	(5)	(6)
Net.Procurement.Ratio	-0.013^{***}					-0.013^{*}
	(0.000)					(0.083)
LN.Grain.Production		-0.526^{***}				-0.314
		(0.015)				(0.411)
GDP.per.capita			-0.002^{*}			-0.0002
			(0.017)			(0.805)
Commerical.Crops				-0.022		0.015
				(0.105)		(0.355)
Irrigation.Cultivating					0.015	0.028^{***}
					(0.193)	(0.002)
Observations	57	57	57	57	57	57
R^2	0.208	0.180	0.085	0.053	0.044	0.367
Adjusted R^2	0.135	0.117	0.055	0.035	0.029	0.213
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table: PCA Summary

	Importance of Components:			
	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.6880620	1.3424679	1.0848549	0.9287501
Proportion of Variance	0.3561941	0.2252775	0.1471138	0.1078221
Cumulative Proportion	0.3561941	0.5814717	0.7285854	0.8364075

Table: PCA Panel Data

	Dependent variable:			
	PII			
	(1)	(2)	(3)	(4)
Comp1	-0.117^{**}	-0.163^{**}	-0.161^{*}	-0.179^{**}
	(0.015)	(0.019)	(0.063)	(0.046)
Comp2		0.185	0.190	0.354^{*}
		(0.292)	(0.300)	(0.060)
Comp3			0.007	-0.029
			(0.966)	(0.878)
Comp4				-0.173
				(0.294)
Observations	57	57	57	57
R^2	0.121	0.142	0.142	0.181
Adjusted R^2	0.079	0.090	0.087	0.108
Note:		$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Conclusion

- Using panel data approach for program evaluation, we quantitatively estimate the impact of radical policies on the collapse of crop production, which is about 0.21 on average.
- The Political Impact Index which measures the extent of radicalism among Chinese provinces can significantly explain the spatial distribution of mortality rates.
- We quantitatively evaluate the relative importance of proxies used in the literature for various radical policies by regressing them on PII.

Thank You!