The long-term effects of natural disasters on human capital accumulation: a quasi-natural experiment based on the Yellow River floodplain area

Weihua Yu[1] and Jingjing Hu[1*]

[1]Jinhe Center for Economic Research, Xi'an Jiaotong University, Xi'an, China
[*]Corresponding author. E-mail: 787652794@qq.com

ONLINE APPENDIX

Table A1. Summary statistics

Variable	Definition	Observations	Mean	S.D.
Panel A. Historical variables				
Hf	Whether county located in the YRFA	449	0.389	0.488
Ljinshi	Logarithm of jinshi	360	3.056	1.037
Ldjinshi	The logarithm of the density of jinshi	334	5.172	1.194
Drought	Drought frequency from 1470 and 2000	449	0.598	0.089
Total natural disasters	Total natural disasters frequency from 1470 and 2000	449	1.193	0.717
Clan	Genealogical density	257	0.658	1.720
Panel B. Socioeconomic variables in 2010				
HCA	Average years of education	449	8.464	0.621
Credit-to-GDP ratio	Credit divided by Gross Domestic Product	449	0.420	0.278
GDP per capita	Gross Domestic Product divided by Total Population	449	2.881	2.310
Population	Population of the county	449	631727.2	302726.6
Share of ethnic minority population	The proportion of ethnic minority population to the total population	449	1.918	8.303
Area	Administrative area	449	1428.562	924.163
Gender ratio	Divide the number of males by the number of females	449	102.298	4.231
Eli in CGSS2010	Educational level of individual in CGSS2010	1522	6.527	4.476
Eli in CLDS2012	Educational level of individual in CLDS2012	1642	9.730	2.415
Fel in CLDS2012	Father's educational level in CLDS2012	3221	4.524	3.700
Mel in CLDS2012	Mother's educational level in CLDS2012	3257	2.813	2.993
Trust	Response to "Overall, do you believe that the majority of people can be trusted?"	2369	3.545	1.131
Panel C. Geographic variables				
Latitude	Latitude of the county	449	35.485	2.755
Longitude	Longitude of the county	449	116.405	2.280

Table A2. Balance test

	Mean difference			
	<150 km	<200 km	<250 km	Full sample
	(1)	(2)	(3)	(4)
Panel A. Historical variables				
County area in the Qing Dynasty	–0.005	–0.017	–0.025	–0.029
	(0.035)	(0.034)	(0.036)	(0.035)
Distance from the county to the prefecture during the Ming dynasty	13.413	13.644	9.167	8.144
	(12.415)	(11.990)	(12.803)	(12.737)
Panel B. Socioeconomic variables				
Credit-to-GDP ratio	0.017	0.015	0.003	0.002
	(0.037)	(0.038)	(0.041)	(0.041)
GDP per capita	–1.043	–0.935	–0.887	–0.927
	(0.548)	(0.515)	(0.501)	(0.493)
Log of population	0.091	0.125	0.169	0.184
	(0.134)	(0.153)	(0.174)	(0.177)
Share of ethnic minority population	0.623	–1.228	–1.319	–1.492
	(0.360)	(0.595)	(0.631)	(0.702)
Log of administrative area	0.032	–0.014	–0.066	–0.079
	(0.100)	(0.098)	(0.111)	(0.109)
Gender ratio	–1.254	–1.448	–1.593	–1.601
	(0.813)	(0.793)	(0.762)	(0.771)
Observations	396	419	439	449

Notes: Columns (1)–(4) show the differences in control variables between the treated and control counties under different bandwidths. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A3. Parameter estimation results

	Average years of education	
	(1)	(2)
Whether county located in the YRFA	–0.268	–0.325
	(0.076)	(0.108)
Polynomial	Linear	Quadratic
Observations	452	452
R-squared	0.132	0.132

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A4. Different survey data results

	Educational level of individual in CGSS 2010	Educational level of individual in CLDS2012	Educational level of father in CLDS2012	Educational level of mother in CLDS2012
	(1)	(2)	(3)	(4)
Whether county located in the YRFA	–0.726	–0.583	–0.542	–0.591
	(0.239)	(0.279)	(0.233)	(0.260)
Observations	1,522	1,642	3,221	3,257
R-squared	0.011	0.018	0.020	0.034

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A5. Excluding bad controls

	Average years of education			
	(1)	(2)	(3)	(4)
Panel A: sample within < 150 km of bound				
Whether county located in the YRFA	–0.370	–0.183	–0.160	–0.160
	(0.032)	(0.067)	(0.085)	(0.085)
Observations	399	399	399	399
R-squared	0.314	0.480	0.515	0.515
Panel B: sample within < 200 km of bound				
Whether county located in the YRFA	–0.367	–0.188	–0.177	–0.182
	(0.035)	(0.064)	(0.078)	(0.081)
Observations	422	422	422	422
R-squared	0.325	0.488	0.525	0.527
Panel C: sample within < 250 km of bound				
Whether county located in the YRFA	–0.373	–0.212	–0.196	–0.193
	(0.034)	(0.050)	(0.064)	(0.068)
Observations	442	442	442	442
R-squared	0.320	0.479	0.523	0.525
Panel D: Full sample				
Whether county located in the YRFA	–0.379	–0.221	–0.207	–0.201
	(0.035)	(0.047)	(0.065)	(0.069)
Observations	452	452	452	452
R-squared	0.327	0.479	0.526	0.528
Polynomial	Linear	Quadratic	Cubic	Quartic
Covariates	Yes	Yes	Yes	Yes

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A6. Results using Conley standard errors

	Average years of education			
	(1)	(2)	(3)	(4)
Whether county located in the YRFA	–0.296	–0.269	–0.188	–0.177
Cutoff=50	(0.101)	(0.095)	(0.075)	(0.073)
Cutoff=150	(0.089)	(0.079)	(0.069)	(0.074)
Polynomial	Linear	Quadratic	Cubic	Quartic
Covariates	452	452	452	452
R-squared	0.374	0.463	0.533	0.539

Notes: Conley standard errors with different combinations of distance cutoffs (50 km, 150 km) are shown in parentheses.

Table A7. "Donut" RD result

	Average years of education				
	2×10 km	2×20 km	2×30 km	2×40 km	2×50 km
	(1)	(2)	(3)	(4)	(5)
Panel A: Sample within < 150 km of bound					
Whether county located in the YRFA	–0.220	–0.268	–0.281	–0.305	–0.321
	(0.071)	(0.079)	(0.090)	(0.099)	(0.093)
Observations	366	304	256	204	168
R-squared	0.484	0.508	0.497	0.529	0.547
Panel B: Sample within < 200 km of bound					
Whether county located in the YRFA	–0.217	–0.265	–0.282	–0.313	–0.340
	(0.073)	(0.082)	(0.095)	(0.102)	(0.098)
Observations	389	327	279	227	191
R-squared	0.494	0.515	0.505	0.538	0.554
Panel C: Sample within < 250 km of bound					
Whether county located in the YRFA	–0.220	–0.275	–0.306	–0.340	–0.380
	(0.075)	(0.088)	(0.103)	(0.115)	(0.118)
Observations	409	347	299	247	211
R-squared	0.477	0.489	0.477	0.506	0.511
Panel D: Full sample					
Whether county located in the YRFA	–0.228	–0.283	–0.315	–0.349	–0.392
	(0.075)	(0.088)	(0.104)	(0.115)	(0.117)
Observations	419	357	309	257	221
R-squared	0.476	0.489	0.479	0.509	0.516

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A8. Impact on different genders

	< 150km		< 200 km		< 250 km		Full sample	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Linear Polynomial								
Whether county located	–0.129	–0.284	–0.123	–0.284	–0.122	–0.288	–0.127	–0.297
in the YRFA	(0.067)	(0.079)	(0.068)	(0.084)	(0.068)	(0.084)	(0.068)	(0.083)
Observations	396	396	419	419	439	439	449	449
R-squared	0.403	0.509	0.407	0.521	0.401	0.506	0.397	0.506
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P value	0.000		0.000		0.000		0.000	
Panel B: Quadratic Polynomial								
Whether county located	–0.124	–0.110	–0.122	–0.123	–0.139	–0.174	–0.148	–0.191
in the YRFA	(0.026)	(0.047)	(0.021)	(0.038)	(0.022)	(0.015)	(0.025)	(0.021)
Observations	396	396	419	419	439	439	449	449
R-squared	0.480	0.628	0.478	0.635	0.468	0.603	0.460	0.598
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P value	0.611		0.965		0.238		0.142	
Panel C: Cubic Polynomial								
Whether county located	–0.108	–0.129	–0.104	–0.160	–0.130	–0.206	–0.148	–0.225
in the YRFA	(0.036)	(0.067)	(0.036)	(0.056)	(0.034)	(0.043)	(0.025)	(0.049)
Observations	396	396	419	419	439	439	449	449
R-squared	0.490	0.632	0.485	0.640	0.477	0.624	0.460	0.623
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P value	0.509		0.090		0.019		0.006	
Panel D: Quartic Polynomial								
Whether county located	–0.100	–0.142	–0.105	–0.170	–0.130	–0.195	–0.133	–0.208
in the YRFA	(0.035)	(0.070)	(0.034)	(0.058)	(0.034)	(0.053)	(0.037)	(0.059)
Observations	396	396	419	419	439	439	449	449

R-squared	0.491	0.633	0.486	0.642	0.477	0.632	0.474	0.632
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
P value	0.203		0.049		0.028		0.014	

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors. The P-value is used to test the significance of the difference in the coefficient of whether county located in the YRFA between groups.

Table A9. Other historical natural disasters

	Average years of education				Average years of education			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: sample within < 150 km of bound								
Whether county located in the YRFA	–0.199	–0.122	–0.115	–0.122	–0.144	–0.133	–0.135	–0.143
	(0.071)	(0.031)	(0.043)	(0.052)	(0.082)	(0.049)	(0.055)	(0.063)
Observations	396	396	396	396	396	396	396	396
R-squared	0.483	0.580	0.584	0.585	0.494	0.580	0.584	0.585
Panel B: sample within < 200 km of bound								
Whether county located in the YRFA	–0.197	–0.128	–0.131	–0.136	–0.145	–0.134	–0.146	–0.156
	(0.075)	(0.024)	(0.037)	(0.039)	(0.086)	(0.041)	(0.044)	(0.048)
Observations	419	419	419	419	419	419	419	419
R-squared	0.492	0.586	0.589	0.591	0.501	0.586	0.589	0.591
Panel C: sample within < 250 km of bound								
Whether county located in the YRFA	–0.205	–0.162	–0.171	–0.160	–0.174	–0.159	–0.177	–0.179
	(0.076)	(0.015)	(0.034)	(0.038)	(0.087)	(0.032)	(0.035)	(0.043)
Observations	439	439	439	439	439	439	439	439
R-squared	0.476	0.560	0.573	0.580	0.479	0.560	0.573	0.579
Panel D: Full sample								
Whether county located in the YRFA	–0.213	–0.175	–0.185	–0.170	–0.191	–0.175	–0.192	–0.191
	(0.075)	(0.020)	(0.039)	(0.044)	(0.088)	(0.032)	(0.034)	(0.040)
Observations	449	449	449	449	449	449	449	449
R-squared	0.474	0.555	0.570	0.577	0.476	0.554	0.571	0.577
Polynomial	Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

Table A10. Effect of population migration

	Average years of education				
	(1)	(2)	(3)	(4)	(5)
Panel A: sample within < 150 km of bound					
Whether county located in the YRFA	–0.408	–0.231	–0.112	–0.122	–0.121
	(0.051)	(0.066)	(0.047)	(0.056)	(0.055)
Observations	381	378	378	378	
R-squared	0.227	0.505	0.590	0.593	0.593
Panel B: sample within < 200 km of bound					
Whether county located in the YRFA	–0.377	–0.229	–0.120	–0.137	–0.141
	(0.057)	(0.068)	(0.045)	(0.051)	(0.051)
Observations	404	401	401	401	
R-squared	0.227	0.512	0.595	0.598	0.599
Panel C: sample within < 250 km of bound					
Whether county located in the YRFA	–0.348	–0.232	–0.163	–0.175	–0.165
	(0.066)	(0.069)	(0.037)	(0.047)	(0.054)
Observations	424	421	421	421	
R-squared	0.186	0.493	0.568	0.582	0.587
Panel D: Full sample					
Whether county located in the YRFA	–0.342	–0.240	–0.184	–0.192	–0.179
	(0.068)	(0.068)	(0.039)	(0.052)	(0.059)
Observations	434	431	431	431	
R-squared	0.181	0.491	0.561	0.579	0.584
Polynomial	Linear	Linear	Quadratic	Cubic	Quartic
Covariates	No	Yes	Yes	Yes	Yes

Notes: Two-dimensional geographic controls are used in all regression analyses. Standard errors in parentheses are clustered at the province level using wild cluster bootstrap standard errors.

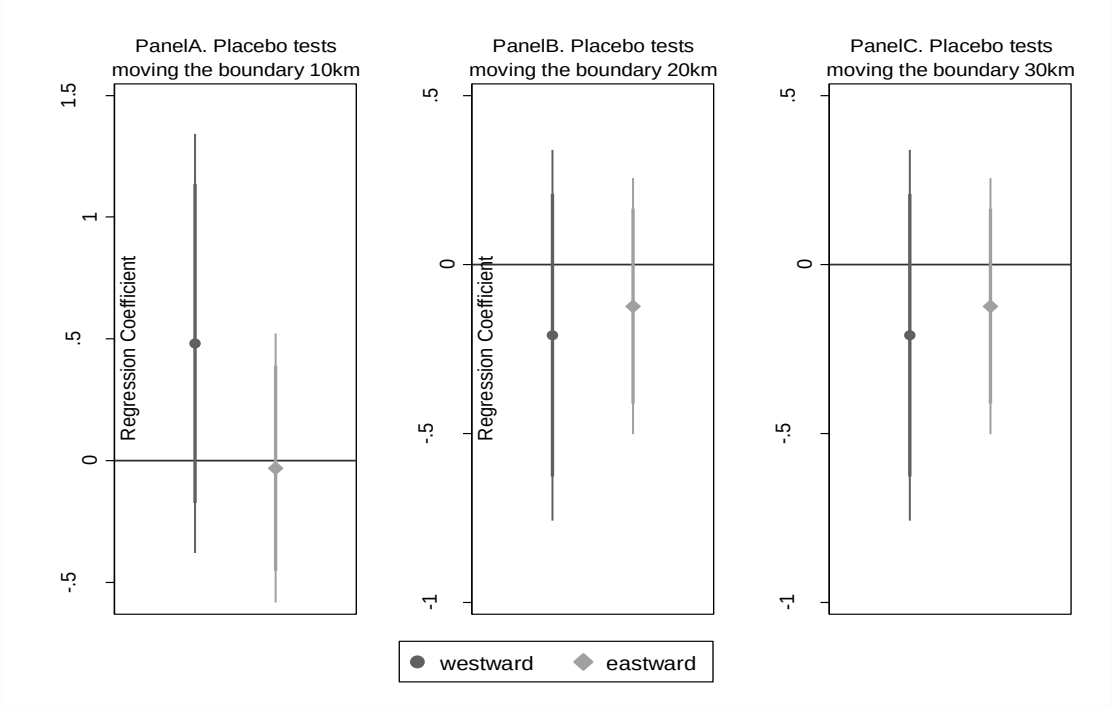

Figure A1. Estimated coefficient plot of falsified boundary

Notes: Both 95% and 99% confidence intervals are used for each estimate. The estimated sample is within 200km of bandwidth. Panel A–C are estimates when the boundary is moved 10, 20 and 30 km westward and eastward, respectively.