Environment and Development Economics (2024), **29**, 279–295
doi:10.1017/S1355770X24000111

RESEARCH ARTICLE

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

Weihua Yu, and Jingjing Hu*

Jinhe Center for Economic Research, Xi'an Jiaotong University, Xi'an, China
*Corresponding author: Jingjing Hu; Email: 787652794@qq.com

(Submitted 10 March 2023; revised 29 December 2023; accepted 03 February 2024; first published online 29 April 2024)

Abstract

This study exploits the correlation between Yellow River flooding and human capital accumulation using county-level data from Anhui, Hebei, Henan, Jiangsu, and Shandong provinces in China. Employing a spatial regression discontinuity approach, we compare the differences in human capital accumulation within and beyond the Yellow River floodplain areas (YRFA). Empirical results show human capital accumulation in the YRFA is at least 12.1 percent lower than outside the YRFA. Furthermore, our results demonstrate intergenerational transmission and gender differences in the negative impact of the Yellow River flooding on human capital accumulation. The baseline specifications of this study are not affected by drought and overall natural disasters. This paper documents historical human capital accumulation, clan power, and social trust, through which Yellow River flooding has a long-term impact.

Keywords: human capital accumulation; historical natural disasters; Yellow River flooding; Yellow River floodplain areas; regression discontinuity approach; long-term effect

JEL classification: I21; R11; N35

1. Introduction

Understanding how human capital accumulates is crucial for economic innovation and development (Mankiw *et al*., 1992). For example, religion and human capital in history have long-term effects on education (Becker and Woessmann, 2009; Chen *et al*., 2020; Xiong and Zhao, 2023). In addition, historical shocks affect regional human capital formation in the long term (Dell, 2010). In particular, the role of natural disasters in human capital cannot be ignored.

There is a debate about the net effect of natural disasters on human capital. A large body of evidence suggests that the effect may be largely negative. Empirical studies have documented that natural disasters such as typhoons, locust plagues, and droughts have a significant negative impact on children's educational attainment (Beegle *et al*., 2006;

de Vreyer *et al.*, 2015; Deuchert and Felfe, 2015). Specifically, the economic losses to government, society, and households brought on by natural disasters lead to the negative influence of natural disasters on human capital investment (Otero and Marti, 1995; Carter *et al.*, 2007; Baez *et al.*, 2010). However, some studies suggest that natural disasters are beneficial to human capital by generating Schumpeterian creative destruction. The introduction of technology and equipment in post-disaster reconstruction can lead to higher human capital investment and productivity (Hallegatte and Ghil, 2007; Teixeira and Fortuna, 2010). In addition, the risk of natural disasters alters the relative returns on capital. Specifically, disaster risk reduces the expected rate of return on physical capital and increases the relative rate of return on human capital (Skidmore and Toya, 2010; Zhang and Ruan, 2020). As a result, according to these studies, natural disasters contribute to an increase in human capital.

Overall, the impact of natural disasters on human capital is unclear, which means that research on the connection between different natural disasters and human capital is necessary. Therefore, this paper focuses on the long-term effects of the Yellow River flooding on human capital accumulation. The Yellow River has flooded many times in the past; when it overflows, some counties in Henan, Shandong, Hebei, Anhui and Jiangsu provinces are seriously threatened by flooding. The counties affected by the flooding of the Yellow River are mainly in the provinces of Hebei, Henan, Anhui, Jiangsu, and Shandong. The area historically flooded by the Yellow River is called the Yellow River floodplain area (YRFA). The five provinces of Henan, Shandong, Hebei, Anhui, and Jiangsu contain both YRFA and unaffected regions, which we call the non-Yellow River floodplain areas (non-YRFA). Taking advantage of this spatial discontinuity in the impact of natural disasters, we examine the lasting impact of the historic Yellow River flooding on human capital accumulation.

In this paper, we match the county-level data from the *2010 Sixth National Population Census* to Yellow River flooding data. We use average years of education as the explanatory variable, recorded in the census. Whether the county belongs to the YRFA is used as the central explanatory variable. Empirically, we refer to the semi-parametric regression discontinuity (RD) approach of Dell (2010) and control for the latitude and longitude polynomial of the county. First, we demonstrate the validity of the RD by using a balance test. The results show that the long-term flooding of the Yellow River has indeed reduced human capital accumulation in the YRFA. In addition, we use the parametric RD approach, different survey data, excluding bad controls, Conley standard errors, moving the boundary location, and deleting samples with different proportions near the boundary to conduct robustness tests. The results are still robust.

We also examine gender differences in the impact of natural disasters to better understand the effect of Yellow River flooding on the human capital of different groups. The results indicate that Yellow River flooding has a more severe negative impact on human capital accumulation (HCA) for women than for men. We also exclude the competitive hypothesis of drought and total natural disaster. We find that the lower level of HCA in the YRFA is not primarily due to drought or other natural disasters.

As for the mechanism, we believe that HCA is affected by Yellow River flooding mainly through three main channels: historical human capital, clan culture, and social trust. First, we find that the long-term flooding of the Yellow River had a detrimental effect on the quantity of jinshi (scholars who won the final level of the central government court examination in the ancient Chinese imperial examination system) in the dynasties of the Ming and Qing, and then had a long-term effect on the HCA through intergenerational transmission. Second, the long-term flooding of the Yellow River has damaged the environment in which clans live, affecting clan culture. Clan culture, which is based on blood ties, plays a very important role in post-disaster recovery. Third, social trust is conducive to HCA, but the flooding of the Yellow River has reduced the level of social trust.

Our research adds to the body of knowledge on the long-run effects of natural disasters on HCA in three ways. First, this is the first paper to assess the lasting effects of Yellow River flooding on HCA. While existing literature shows that natural disasters have an effect on the accumulation of human capital, the long-term economic effects of natural disasters are poorly understood (Cavallo and Noy, 2009). Previous literature has concentrated on the short-term effects of disasters. Especially, there is a lack of knowledge on the impact of historical natural disasters on the formation of human capital in the long term. Exploring the effect of historical natural disasters on long-term economic outcomes has important theoretical and practical significance for today's environmental climate research.

Second, our paper provides additional evidence on the channels through which historical natural disasters affect human capital investment. Analyzing such mechanisms is essential to understanding historical persistence (Cirone and Pepinsky, 2022). This study adds an analysis of the historical HCA, culture, and informal institutions through which historical natural disasters have affected human capital over time. By understanding these mechanisms, policymakers will be better able to design disaster-related policies and reduce the potential adverse impact caused by natural disasters. This is especially important given the natural disasters occurring with increasing frequency in a changing climate.

Third, the interaction of complex environmental factors makes it difficult to identify the causal relationship between natural disasters and HCA. Therefore, using similar samples to study the impact of natural disasters can minimize the influence of other factors on human capital investment. The YRFA and non-YRFA in Henan, Shandong, Hebei, Anhui, and Jiangsu provinces provide a quasi-natural experimental sample for our study. The RD approach helps us identify the causal effects of historical natural disasters on HCA.

The rest of the paper is structured as follows. Section 2 describes the historical background of the Yellow River flooding. We present the conceptual framework in section 3. We introduce the data in section 4. Section 5 demonstrates the empirical strategy and the pre-treatment balance. Section 6 describes the main results. Section 7 discusses the underlying mechanism and section 8 concludes the paper.

2. Historical background

The Yellow River basin is in northern China, with a temperate climate. Before modern flood control methods were available, there was no way for water to escape during the spring and autumn rainy season, so the river sometimes rose above the embankment, and the dikes burst. In addition, human factors such as the North-South conflict in 1128 contributed to the flooding of the Yellow River. According to the *History of Yellow River Water Resources* (Writing Group of History of Yellow River Water Resources, 2003), the flood history of the Yellow River spans from 168 BC to 1938 AD. In 2106 years, the Yellow River flooded 403 times. Henan, Shandong, Hebei, Anhui and Jiangsu were seriously affected by flooding when the river overflowed. About 250,000 square kilometers of the plains in the Yellow River's lower reaches were affected by the floods. The area historically affected by the Yellow River flooding is called the YRFA. In figure 1, we show the YRFA versus the non-YRFA.

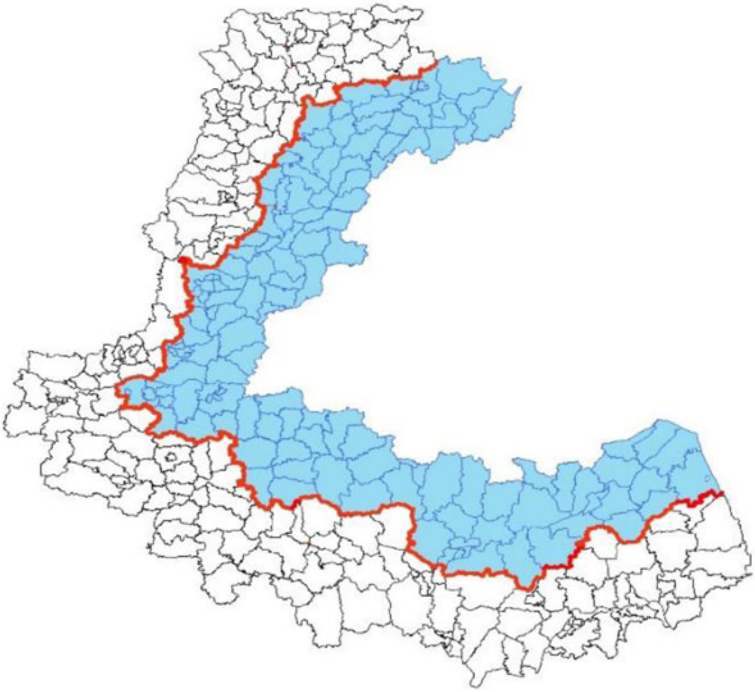

Figure 1. The spatial distribution of YRFA and non- YRFA.
Note: The red line represents the boundary of YRFA and non-YRFA. The blue area contains the counties located in YRFA, and the white area contains the counties located in non-YRFA. The map shows the counties within 100 km on either side of the boundary.

Historical living conditions in the YRFA have resulted in significant human and economic losses. For example, more than 10 million people were affected and more than 1.46 million homes destroyed when the Huayuankou Dam on the Yellow River burst in 1938, forcing 3.9 million people to flee their homes in Henan, Anhui, and northern Jiangsu provinces. Wherever the Yellow River flooded, there were crumbling buildings and starving people. The scene in Zhongmou County at the time was described in the Ta Kung Pao newspaper, published on 23 June 1938, when the population was reduced to eating bark, grass, and roots for more than 10 days. In addition to the significant economic and population losses caused by the flooding of the Yellow River, its high sand content has had a negative impact on salinization and desertification. As a result, people could only grow crops with low soil quality requirements. This has inevitably affected the profitability of agricultural production. The financial and human costs of managing the Yellow River ultimately fall on the shoulders of the general public. The Yellow River has not only caused temporary loss, but has also had a lasting impact on the YRFA.

3. Conceptual framework

How does the historical Yellow River flooding affect the accumulation of human capital today? From the perspective of theoretical mechanisms, path dependence provides an explanation. Path dependence proposes that small shocks change the historical process. Specifically, path dependence holds that the current and future states, actions, and decisions depend on the path of previous states, actions, and decisions (Page, 2006).

Existing research also shows that natural disasters have long-term effects on HCA. Tian *et al.* (2022) found that individuals who experienced high-intensity or low-intensity earthquakes in the womb had lower levels of education as adults. Xu (2011) took a long-term perspective and estimated the impact of earthquakes a quarter of a century later, finding that earthquakes had an adverse effect on educational outcomes. Deuchert and Felfe (2015) found that typhoons had negative and persistent effects on children's education, manifested as lower test scores, increasing retention rates, and declining overall levels of education. In addition to earthquakes and typhoons, the long-term effects of tropical storms and floods on education are also significant (Sotomayor, 2013; Dinkelman, 2017).

Kousky (2016) found that natural disasters damaged children's physical and mental health, destroyed homes and schools, and forced children to interrupt their education, thereby having long-term effects on their education. This suggests that Yellow River flooding may have affected early human capital, leading to long-term effects on HCA. This makes it seem like the investment in children's education by families during natural disasters is a temporary adjustment during difficult times, but often becomes a permanent change and has long-lasting effects.

Therefore, we propose *Hypothesis 1:* Yellow River flooding has had a long-term negative impact on HCA.

China's historical and current levels of human capital have a significant positive relationship (Chen *et al.*, 2020). In terms of human capital, the accumulation of human capital has the characteristic of intergenerational transmission (Becker and Tomes, 1986; Dong *et al.*, 2019). Parents' better education and attitudes towards education will affect the education level of the next generation. Therefore, ongoing damage to human capital is expected as a result of the historical Yellow River flooding if the historical Yellow River flooding has harmed historical human capital.

Therefore, we propose *Hypothesis 2*: Yellow River flooding affects contemporary HCA through historical HCA.

Studies have found that major historical shocks, such as natural disasters and wars, could shape culture (such as values and beliefs). Thus, historical natural disasters help clarify the origins of cultural and institutional differences. As Aristotle pointed out, different geographical environments give rise to different institutions and cultures. Cultures and institutions are particularly important in explaining economic growth and other economic outcomes (Dell, 2010; Guiso *et al.*, 2016). More importantly, culture and institutions have persistent effects on socioeconomic outcomes. It is because of the persistent effects of institutions and culture that contemporary development differs across countries (Acemoglu *et al.*, 2001; Tabellini, 2010). Therefore, we can see that natural disasters have long-run impacts on socioeconomic outcomes by affecting institutions and cultures (Nunn, 2009; Nunn and Wantchekon, 2011).

Clan culture, which is based on blood ties, serves as an essential complement to formal institutions. Clan resources can help families reduce or offset natural disasters' effects on consumption. Weak formal financial markets make it difficult for households to maintain normal consumption spending in times of crisis, which can have a significant and long-lasting impact on human capital. In the absence of formal financial markets, clan can be seen as a substitute or complement to physical capital, which can boost the labor productivity of low-income households and raise incomes. Ultimately, clan finance is likely to help disadvantaged households escape the cycle of poverty (Mogues and Carter, 2005; Chantarat and Barrett, 2012). Thus, after natural disasters, such informal social networks may play a role in risk sharing and smoothing consumption. Families may receive financial support from the clan to help them recover more quickly from natural disasters.

Additionally, the clan encourages its youth to pursue higher education by running schools and creating a system of student aid and scholarships (Bol, 1994). The emphasis on education in modern clan groups, and the old system of clan education and aid, both show that the clan values education. In short, clans support formal institutions by providing both financial aid and educational opportunities.

Therefore, we propose *Hypothesis 3*: Yellow River flooding affects contemporary HCA through clan culture.

Social trust is also an important part of informal institutions. There is evidence that trust promotes education (Yamamura, 2011). From a demand perspective, education is an important component of employment in a more trusting culture, which increases the returns to education (Knack and Keefer, 1997). From a supply perspective, in societies with higher levels of trust, people are more confident that their investment in education will be fairly rewarded. In addition, trust can ease market credit constraints, making it easier for individuals to obtain financing to invest in human capital. Therefore, education investment is less expensive in high-trust societies than in low-trust societies, resulting in faster educational progress in the former countries (Bjørnskov, 2009).

Therefore, we propose *Hypothesis 4*: Yellow River flooding affects contemporary HCA through social trust.

4. Data

This paper is based on a number of datasets that together provide detailed information on the socio-economic status and educational attainment at the county level. Table A1 in the online appendix contains summary statistics.

4.1 Data on human capital accumulation

The core data used in this study comes from the *2010 Sixth National Population Census*. This census is a survey of the population and households, conducted by the State Council of China. The survey covers natural persons residing in China. Census data are available at the county level. The samples in this paper include counties from Anhui, Hebei, Henan, Shandong, and Jiangsu provinces. We retain data for these five provinces in the *2010 Sixth National Population Census*.

The average years of education in the *2010 Sixth National Population Census* is used to measure HCA. We examine the robustness of the results by using education data from the *China General Social Survey 2010* (CGSS2010) and the *China Labor-force Dynamic Survey 2012* (CLDS2012). Specifically, we use the highest educational attainment of individuals in the CGSS2010 and the educational attainment of individuals and parents in the CLDS2012 to measure HCA. In order to construct a variable that reflects historical HCA, the jinshi data are drawn from Zhu and Xie's (1980) *Ming-Qing jinshi timing beilu suoyin*. This source contains the exact addresses of the jinshi. Therefore, we sort out the number of Ming and Qing jinshi in each county across the boundary.

4.2 Natural disaster data

We use the *Yellow River Basin Atlas* to get a map of the YRFA. In this paper, a county that is located in the YRFA is equal to 1, while a county that is located outside the YRFA is equal to 0.

Even counties within the YRFA differ in the extent to which they are affected by flooding. The number of Yellow River floods is used to measure this difference. In this paper, the flood frequency of the Yellow River from 168 BC to 1938 AD is manually sorted out using data from the *History of Yellow River Water Resources* (Writing Group of History of Yellow River Water Resources, 2003). The *Atlas of Drought and Flood Distribution in China in the Last 500 Years* (Modern Scientific Research Institute of the Central Meteorological Administration, 1981), records drought and flood data from 120 meteorological stations across the country from the years 1470 to 2000. We use the data from these 120 meteorological stations to obtain the average annual drought and total natural disaster frequency using the ordinary kriging method in ArcGIS. The drought and total natural disaster data are geo-georeferenced to the county.

4.3 County socio-economic data

The level of economic and financial development also affects HCA. Therefore, we control for economic and financial indicators in the main estimations. Using credit scale, GDP, and population data from the County Statistical Yearbook, we calculate the credit-to-GDP ratio and GDP per capita, which measures the level of financial and economic development.

Clan genealogy is used to measure the strength of a county's clan culture. We matched county-level genealogical data from the *General Catalogue of Chinese Genealogy* (Wang, 2006) and population data from the *Land Population Survey of Cities and Counties in China* (Interior, 1935). To obtain the genealogical density, we divide the genealogical data by the population data. We use the genealogies per capita to measure the degree of clan culture in the county. This paper defines genealogy per capita by dividing the genealogy data by the total population.

This study uses social trust data from the CGSS2010. The question 'Overall, do you believe that the majority of people can be trusted?' in the CGSS2010 measures social trust. Responses range from strongly disagree to agree on a scale of 1–5, with a higher score indicating that the respondent trusts other people.

4.4 Geographical data

The Qing dynasty administrative area of each county and the distance from the county to the prefecture in the Ming dynasty are obtained from the CHGIS database.

5. Empirical design

5.1 Empirical approach

The identification strategy used in this paper is based on the assumption that, in the absence of the Yellow River flooding, the level of HCA would be similar across the YRFA and the non-YRFA. We use the map of the YRFA to identify the treatment and control groups.

We use the spatial RD approach to compare HCA in counties on either side of the YRFA boundary. The counties across the YRFA boundary are counterfactual because the counties on either side of the YRFA boundary had similar geographical and socioeconomic conditions. The ideal treatment and control counties are inside and outside the YRFA. This paper refers to the semi-parametric RD method of Dell (2010). The following estimation equation is proposed to investigate the lasting effect of Yellow River flooding:

$$HCA_i = \alpha_0 + \alpha_1 Hf_i + \beta X_i + f(\textit{geographic location}_i) + \varepsilon_i,$$

where HCA_i is the average years of education in county i. Hf_i is an indicator variable. It takes the value 1 for a YRFA county and 0 for a non-YRFA county. X_i is a vector of covariates, including output per capita, the share of credit capital in GDP, log of population, the gender ratio, the share of ethnic minority population, and the log administrative area. We refer to the empirical specification of Dell (2010), where $f(\textit{geographic location}_i)$ is an RD polynomial. $f(\textit{geographic location})$ is a two-dimensional polynomial that controls for the latitude and longitude of the county, which can absorb smoothing trends on either side of the boundary. α_1 is the coefficient on which we focus. It measures the influence of the Yellow River flooding on HCA. ε_i is the residual term. We adopt cluster standard errors at the province level, which is appropriate, since unobservables may be correlated at this level. A very small number of clusters ($n = 5$) can lead to downward biased standard errors. Therefore, we use wild cluster bootstrap standard errors.

We use local linear regression in the baseline specification. This controls for a linear latitude-longitude polynomial and uses a restricted bandwidth near the boundary. Since there is no well-accepted optimal bandwidth for RD, we restrict the samples to 150, 200 and 250 km on either side of the boundary (Dell and Olken, 2020). To confirm the reliability of the baseline specifications, we also present the results of estimation using parametric and a global polynomial approach.

5.2 Balance tests

RD estimation requires that the samples on either side of the boundary are similar, which means that all relevant factors affecting HCA must vary smoothly across the YRFA boundary. The YRFA includes five provinces, Hebei, Henan, Anhui, Jiangsu and Shandong, which are geographically close and have similar historical development paths. Moreover, the neighboring counties show a high degree of homogenization. To test this hypothesis, this paper compares characteristic variables of counties on both sides of the YRFA.

In online appendix table A2, we report the differences in control variables between the treated and control counties under different bandwidths. The historical variables both inside and outside the YRFA are described in panel A. Panel A demonstrates that the county areas during the Qing dynasty were statistically the same on both sides of the YRFA boundary. The ancient society was mainly based on agriculture, so land area can be used to measure the ability of the area to produce goods. This also suggests that there were similar initial levels of development inside and outside the YRFA. The counties that are closer to the prefecture have a relatively more developed economy and lower transport costs for students to take the exam to become jinshi. Panel A shows that the distance from the county to the prefecture in the Ming dynasty has no significant difference across the YRFA boundary.

Next, we test the balance of the county's major socioeconomic variables in panel B. Our study further compares the per capita output, the share of credit in GDP, the log of population, the gender ratio, the share of ethnic minority population, and the log of administrative area at the county level in 2010. The results in panel B demonstrate that these variables are balanced both inside and outside the YRFA. Socioeconomic variables at the county level are taken into account. The findings in panel B are encouraging because they show that the counties on either side of the boundary are comparable for these variables. These county-level socio-economic variables are controlled in our subsequent analysis.

6. Empirical results

6.1 Baseline results

We first use spatial RD to estimate the impact of Yellow River flooding on HCA. In the main empirical specification, we use the average years of education in 2010 as the dependent variable. Table 1 reports the main results. Columns (1)–(2) control for the linear polynomial in latitude and longitude. We control for the higher polynomial in latitude and longitude in columns (3)–(5). In the absence of an optimal bandwidth for RD that is widely accepted (Dell and Olken, 2020), table 1 presents results for bandwidths of 150, 200, and 250 km using a local approach, which are reported in panels A through C. This ensures the robustness of our estimates to specific bandwidth choices. Panel D is the result of estimation using the global polynomial approach. Across all specifications, the estimated coefficients are negative and statistically significant, which indicates that the estimates are stable.

The minimum estimated coefficient is −0.121 when using a 200 km bandwidth, indicating that the average years of education in the YRFA is at least 12.1 per cent less than that outside the YRFA. Thus, results illustrate that HCA is lower in the YRFA than outside the YRFA. Given that Yellow River flooding occurred repeatedly for more than 2000 years, the negative impact of Yellow River flooding on HCA is significant and long-lasting. Therefore, this study provides evidence that historical natural disasters inhibit HCA.

6.2 Robustness checks

We carry out four robustness checks to strengthen the conclusion that the effect of Yellow River flooding on HCA is lasting and negative.

First, we use a different RD approach to further illustrate the robustness of the baseline results. Lee and Lemieux (2010) suggest that the RD estimate should not depend on a specific model, and that parametric and nonparametric estimation cannot substitute for each other. To avoid bias caused by the choice of estimation method, this subsection employs parameter estimation to check robustness. In online appendix table A3, the first-order and second-order polynomials in the latitude and longitude estimates are shown in columns (1) and (2). The estimated coefficient in column (1) shows a 26.8 per cent decrease in HCA at the cutoff point. The outcomes are in line with the baseline results. This implies that despite the change in approach, the effect of Yellow River flooding on HCA has not changed.

Second, we eliminate the potential bias of using a single survey. The data from the CGSS2010 and the CLDS2012 are used to check the robustness of the baseline outcomes. In online appendix table A4, we use the highest educational attainment of an individual in the CGSS2010 as the dependent variable in column (1). We use the educational attainment of individuals and their parents from the CLDS2012 in columns (2) through (4). The results are in line with the baseline results.

Table 1. Baseline RD results

	Average years of education				
	(1)	(2)	(3)	(4)	(5)
Panel A: sample within <150 km of bound					
Whether county located in the YRFA	−0.361 (0.041)	−0.211 (0.070)	−0.121 (0.032)	−0.122 (0.049)	−0.124 (0.052)
Observations	399	396	396	396	396
R^2	0.195	0.479	0.580	0.583	0.583
Panel B: sample within <200 km of bound					
Whether county located in the YRFA	−0.327 (0.050)	−0.208 (0.073)	−0.127 (0.024)	−0.136 (0.041)	−0.141 (0.042)
Observations	422	419	419	419	419
R^2	0.195	0.489	0.586	0.588	0.589
Panel C: sample within <250 km of bound					
Whether county located in the YRFA	−0.301 (0.060)	−0.210 (0.073)	−0.161 (0.014)	−0.172 (0.035)	−0.165 (0.040)
Observations	442	439	439	439	439
R^2	0.159	0.475	0.560	0.573	0.578
Panel D: Full sample					
Whether county located in the YRFA	−0.296 (0.063)	−0.216 (0.072)	−0.174 (0.019)	−0.185 (0.039)	−0.174 (0.045)
Observations	452	449	449	449	449
R^2	0.156	0.474	0.554	0.570	0.576
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.

More importantly, the results further highlight the long-lasting impact of Yellow River flooding on the intergenerational transmission of human capital in columns (2)–(4). Education in a family can be transmitted to the next generation (Becker and Tomes, 1986; Dong *et al*., 2019). The educational level of the parents and their attitudes towards education influence the educational attainment of the next generation. This empirical finding also suggests that previous human capital continues to play a role in its intergenerational transmission.

Third, we consider the issue of poor controls. Climate variables may impact socioeconomic factors that are usually included as control variables, such as production or the level of financial development (Hsiang *et al*., 2013; Burke *et al*., 2015). This approach is referred to as bad controls, which may lead to mistaken conclusions (Angrist and Pischke, 2009). Therefore, we excluded per capita output and the level of financial development from our re-estimation to modify our estimates. The specific empirical results are presented in online appendix table A5; they demonstrate that our baseline conclusions are not influenced by bad controls.

Fourth, we re-estimated the results using Conley standard errors. Referring to Conley (1999) and Colella *et al.* (2020), we adopt Conley standard errors with cutoff windows of 50 km and 150 km for spatial correlation. The specific empirical results are shown in online appendix table A6. The negative impact of Yellow River flooding on HCA is significant, as indicated by the results.

6.3 Placebo tests

The smoothness of variables within and outside the YRFA has been tested above. However, we still cannot completely rule out any possible confounding factors – in particular, whether the differences in HCA between the two sides of the YRFA boundary are due to statistical coincidence. To test this, we move the YRFA boundaries 10, 20, and 30 km westward and eastward for the placebo tests. In the absence of other factors, no differences in HCA should be observed across the falsified boundary.

The results are shown graphically in online appendix figure A1 which shows the estimated coefficient plot of the falsified boundary. Panels A–C are the RD results for moving the boundary 10, 20, and 30 km to the west and east of the actual boundary, respectively. Figure A1 displays the coefficients with 95 and 99 per cent confidence intervals. As can be seen from figure A1, the coefficients of the falsified boundary formed by moving the boundary to either side are not significant. This indicates that there is no discernible difference in HCA on either side of the falsified boundary. The results show that, when the boundary of the YRFA is arbitrarily moved to a series of placebo locations, the differences in HCA found at the actual boundary do not occur.

6.4 Displacement effects

The counties on both sides of the boundary near the YRFA are taken as the treatment and control groups, since the relevant factors of the counties are smooth inside and outside the YRFA. However, the lower HCA in the YRFA might affect the counties outside the YRFA. HCA outside the YRFA may be affected by the migration of individuals or families within the YRFA. Although the RD estimate provides a causal explanation for the average treatment effect under this migration effect, the results also include unfavorable externalities of migration that cannot be explained by Yellow River flooding.

We eliminate the counties most likely to be affected by migration by removing the counties closest to the boundary. In this way, the 'donut' RD helps us to solve the problem of data accumulation across the YRFA boundary (Eggers *et al.*, 2015; Barreca *et al.*, 2016). Specifically, we exclude 10, 20, 30, 40, and 50 km samples on either side of the boundary to re-estimate the effect of Yellow River flooding.

Table A7 in the online appendix shows the results of the 'donut' RD estimates. Panels A–D in table A7 use different bandwidths. Columns (1)–(5) show the results of RD when excluding samples of 10, 20, 30, 40 and 50 km on either side of the boundary. We observe a statistically significant negative effect on HCA. This finding is consistent across bandwidths and sample exclusions. This demonstrates that the negative impact of Yellow River flooding on HCA persists even after excluding samples on both sides of the boundary.

6.5 Male versus female

The impact of natural disasters varies by group and the negative impact of natural disasters on women's education is more severe than that of men (Baez *et al.*, 2010; Paudel and

Ryu, 2018). Further, there is a more obvious preference for boys over girls in the YRFA (Liang, 2022). To delve into the differential effect of Yellow River flooding on different social groups, we use gender as a proxy to capture socio-economic inequality.

Table A8 in the online appendix shows the impact of Yellow River flooding on HCA of different genders under different bandwidths. Overall, Yellow River flooding has a significant negative impact on HCA of different genders. Moreover, female education is more negatively affected by Yellow River flooding. The *p*-value is used to test the significance of the difference in the coefficient of whether a county is located in the YRFA between groups. From the *p*-value, it can be seen that in panels A, C and D, women are more affected by Yellow River flooding than men in terms of education, and this gender difference is significant. Although historical natural disasters affect both male and female HCA in the long run, the impact is more pronounced for women (Foster and Gehrke, 2017). Prior literature has attributed this effect to the socio-economic status of woman, which means that the higher women's socioeconomic status is, the less likely they are to be affected by natural disasters (Maccini and Yang, 2009). It appears that the long-term effects of the Yellow River floods are associated with the cultural characteristics of patriarchal preference, in which women devote less time to school and more to family activities. Our findings are consistent with prior literature that women are more severely affected by natural disasters due to gender bias (Nakamura *et al*., 2017; Sawada and Takasaki, 2017).

6.6 Ruling out competing hypotheses

Although the results are consistent across different bandwidths, local and global approaches, and semi-parametric and parametric RD, there may still be factors that cause bias. Differences in HCA within and outside the YRFA may be the result of other natural disasters. Contrary to the results of our study, Zhang and Ruan (2020) found that long-term natural disasters in China benefit HCA. This raises concerns that the difference in human capital between the two sides of the boundary may be the result of other natural disasters or total natural disasters, rather than just Yellow River flooding. China has always been a country with frequent natural disasters, with droughts and floods accounting for 57 and 30 per cent of such events, respectively. Therefore, we need to clarify whether the lower HCA in the YRFA is driven by drought or total natural disasters.

We use historical data on drought and total natural disasters from 120 meteorological stations in China between 1470 and 2000 to address this concern statistically. In table A9 (online appendix), we added the natural disasters frequency from 1470 to 2000 in columns (1)–(4), and the droughts frequency from 1470 to 2000 in columns (5)–(8). Based on the results in table A9, we can rule out the impact of the total natural disasters and droughts from 1470 to 2000 on the conclusion of this paper.

6.7 Is there a migration effect?

Population migration is an unavoidable issue in the long-term historical formation process of the YRFA. As Liang (2022) pointed out, if the net migration is caused by the population outside the YRFR with the aforementioned cultural characteristics, it will result in an upward bias in the estimation results, thus overestimating the treatment effect. Therefore, we excluded the top 10 per cent of counties in the migration rate ranking in the YRFR based on the population migration rate of counties in the YRFR to reduce the impact of migration on the baseline conclusion. We show the specific

Table 2. The effect of Yellow River flooding on Ming-Qing jinshi

	Logarithm of jinshi				Logarithm of the density of jinshi			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Whether county located in the YRFA	−0.164	−0.250	−0.301	−0.276	−0.003	−0.298	−0.384	−0.352
	(0.105)	(0.125)	(0.121)	(0.121)	(0.130)	(0.159)	(0.139)	(0.140)
Observations	360	360	360	360	334	334	334	334
R^2	0.134	0.149	0.247	0.254	0.024	0.125	0.373	0.384
Polynomial	Linear	Quadratic	Cubic	Quartic	Linear	Quadratic	Cubic	Quartic

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

empirical results in online appendix table A10. We still see a significant negative impact of the Yellow River flooding on contemporary HCA. We cannot obtain historical migration data, but this conclusion at least indicates that contemporary migration has not challenged the baseline conclusion of this paper.

7. Underlying mechanisms

In this section, we analyze the underlying mechanisms. Specifically, we focus on three channels: historical human capital, clan culture, and social trust. The results document that Yellow River flooding inhibits the formation of historical human capital. In addition, Yellow River flooding has an impact on social culture. These further have impacts on HCA in the present day.

7.1 Ming-Qing jinshi

To verify whether the effect of Yellow River flooding operates through its impact on historical HCA, our study uses jinshi data from the Ming and Qing dynasties to measure historical HCA. Specifically, table 2 use the logarithm of jinshi and the logarithm of density of jinshi (the number of jinshi divided by the area of the county during the Qing dynasty) as the dependent variables for RD estimate. Table 2 presents the empirical outcomes. The coefficients in columns (1)–(8) show that there were fewer jinshi in the YRFA during the Ming and Qing dynasties than in the non-YRFA. In this way, China's historical civil examination system continues to influence human capital outcomes. This may explain why HCA is lower in the YRFA today.

7.2 Clan culture

To estimate the effect of Yellow River flooding on clan culture, our study uses genealogical density to measure clan culture. Genealogical density is used as an explanatory variable in the RD estimate. In table 3, we can see that the historical Yellow River flooding has had a negative impact on clan culture. The population loss and migration caused by Yellow River flooding undermined the clan model of 'living together'. In the context of a weaker clan culture, disasters will lead to larger and more permanent changes in the expenditure of credit-constrained households. This is similar to how asset losses reduce

Table 3. The effect of Yellow River flooding on clan culture

	Genealogical density			
	(1)	(2)	(3)	(4)
Whether county located in the YRFA	−0.344 (0.076)	−0.208 (0.088)	−0.197 (0.092)	−0.196 (0.091)
Observations	257	257	257	257
R^2	0.216	0.275	0.348	0.348
Polynomial	Linear	Quadratic	Cubic	Quartic

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 4. The effect of Yellow River flooding on social trust

	Do you believe that the majority of people can be trusted			
	(1)	(2)	(3)	(4)
Whether county located in the YRFA	−0.022 (0.151)	−0.057 (0.024)	−0.258 (0.093)	−0.256 (0.093)
Observations	2,369	2,369	2,369	2,369
R^2	0.043	0.045	0.061	0.061
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.

investment in education. In many developing countries, this income effect appears to be very significant (Santos, 2007). Weakened clans are also unable to provide educational opportunities through schools. Long-term natural disasters weaken clan culture, making it harder for families to recover from natural disasters as quickly as possible through mutual aid. Therefore, long-term human capital development may be hampered as a result. This indicates that clan culture is a key mechanism for the long-term impact of Yellow River flooding on human capital.

7.3 Social trust

Finally, this section examines the lasting effect of Yellow River flooding on social trust. This study uses the 2010 CGSS: 'Overall, do you believe that the majority of people can be trusted?' This question's response assesses social trust. The higher the rating, the more the respondent trusts others. As this is a multivariate ordered variable ranging from 1 to 5, the ordinal logit model is used for estimation. The effects of Yellow River flooding on social trust are shown in table 4. According to the empirical results, social trust decreases in areas where the Yellow River has flooded more frequently. In conclusion, the historical flooding of the Yellow River has reduced social trust in the YRFA. This confirms our hypothesis 4, that Yellow River flooding hinders HCA by reducing social trust.

8. Conclusions

Our research looks at how natural disasters affect HCA in the long term using data from the Yellow River flood area. To conduct this study, Yellow River flood data are used and an RD design is adopted. Specifically, this paper compares the differences in HCA within and outside the YRFA. The empirical results show that HCA in the YRFA is lower than in the non-flood area. We find that HCA in the YRFA is at least 12.1 per cent lower than non-YRFA. The results further demonstrate that the influence of Yellow River flooding on HCA has obvious characteristics of intergenerational transmission. Moreover, women's HCA is more negatively affected by flooding. We exclude the possible effects of drought and total natural disasters on the difference in HCA between the two sides of the flood boundary. In the mechanism analysis, we find that the historical Yellow River flooding has a lasting impact on HCA through three channels: historical human capital, clan power, and social trust. Thus, this paper once again confirms the enduring influence of historical natural disasters.

The findings of this paper are important because our paper is the first comprehensive empirical assessment of how historical Yellow River flooding affects HCA. Our study provides evidence of the persistent adverse impact of the historical environment. More importantly, with regard to frequently-occurring natural disasters, our study highlights additional perspectives on how to respond to long-term future impacts. In particular, responses to natural disasters should not only consider the immediate impact, but also focus on the long-term effects.

This paper concludes that even if history cannot be changed, the government can make up for the shortcomings of the formal institution by building a good social trust environment and guiding the development of clan organizations, to create better social conditions for the accumulation of human capital in the long term. In addition, the problem of natural disasters affecting women's educational achievement more than men's requires a broad range of comprehensive practices in terms of social attitudes and the construction of legal systems to guarantee women's right to education.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355770X24000111.

Acknowledgements. We appreciate the valuable comments provided by the Co-Editor of Environment and Development Economics and two anonymous reviewers of this journal, which have significantly contributed to improving our paper. All errors are our own.

Competing interests. The authors declare none.

References

Acemoglu D, Johnson S and Robinson JA (2001) The colonial origins of comparative development: an empirical investigation. *American Economic Review* **91**, 1369–1401.

Angrist J-D and Pischke J-S (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Baez JE, Fuente AD and Santos I (2010) Do natural disasters affect human capital? An assessment based on existing empirical evidence. Discussion Papers 5164, Institute of Labor Economics (IZA), Bonn.

Barreca AI, Lindo JM and Waddell GR (2016) Heaping-induced bias in regression discontinuity designs. *Economic Inquiry* **54**, 268–293.

Becker GS and Tomes N (1986) Human capital and the rise and fall of families. *Journal of Labor Economics* **4**, S1–S39.

Becker SO and Woessmann L (2009) Was weber wrong? A human capital theory of protestant economic history. *The Quarterly Journal of Economics* **124**, 531–596.

Beegle K, De Weerdt J and Dercon S (2006) Poverty and Wealth Dynamics in Tanzania: Evidence from A Tracking Survey. Washington, DC: World Bank.

Bjørnskov C (2009) Social trust and the growth of schooling. *Economics of Education Review* **28**, 249–257.

Bol PK (1994). *'This Culture of Ours': Intellectual Transitions in T'ang and Sung China*. California: Stanford University Press.

Burke M, Hsiang SM and Miguel E (2015) Climate and conflict. *Annual Review of Economics* **7**, 577–617.

Carter MR, Little PD, Mogues T and Negatu W (2007) Poverty traps and natural disasters in Ethiopia and Honduras. *World Development* **35**, 835–856.

Cavallo EA and Noy I (2009) The economics of natural disasters: a survey. RES Working Papers 4649, Inter-American Development Bank, Research Department.

Chantarat S and Barrett CB (2012) Social network capital, economic mobility and poverty traps. *The Journal of Economic Inequality* **10**, 299–342.

Chen T, Kung JK and Ma C (2020) Long live Keju! The persistent effects of China's civil examination system. *The Economic Journal* **130**, 2030–2064.

Cirone A and Pepinsky TB (2022) Historical persistence. *Annual Review of Political Science* **25**, 241–259.

Colella F, Lalive R, Sakalli S and Thoenig M (2020) ACREG: Stata module to perform arbitrary correlation regression. Available at https://EconPapers.repec.org/RePEc:boc:bocode:s458889

Conley TG (1999) GMM Estimation with cross sectional dependence. *Journal of Econometrics* **92**, 1–45.

De Vreyer P, Guilbert N and Mesple-Somps S (2015) Impact of natural disasters on education outcomes: evidence from the 1987–89 locust plague in Mali. *Journal of African Economies* **24**, 57–100.

Dell M (2010) The persistent effects of Peru's mining mita. *Econometrica* **78**, 1863–1903.

Dell M and Olken BA (2020) The development effects of the extractive colonial economy: the Dutch cultivation system in Java. *The Review of Economic Studies* **87**, 164–203.

Deuchert E and Felfe C (2015) The tempest: short-and long-term consequences of a natural disaster for children's development. *European Economic Review* **80**, 280–294.

Dinkelman T (2017) Long-run health repercussions of drought shocks: evidence from South African homelands. *The Economic Journal* **127**, 1906–1939.

Dong Y, Luo R, Zhang L, Liu C and Bai Y (2019) Intergenerational transmission of education: the case of rural China. *China Economic Review* **53**, 311–323.

Eggers AC, Fowler A, Hainmueller J, Hall AB and Snyder Jr JM (2015) On the validity of the regression discontinuity design for estimating electoral effects: new evidence from over 40,000 close races. *American Journal of Political Science* **59**, 259–274.

Foster AD and Gehrke E (2017) Consumption risk and human capital accumulation in India. Working Paper S-89212-INC-1, National Bureau of Economic Research.

Guiso L, Sapienza P and Zingales L (2016) Long-term persistence. *Journal of the European Economic Association* **14**, 1401–1436.

Hallegatte S and Ghil M (2007) Endogenous business cycles and the economic response to exogenous shocks. Working Paper No. 20.2007, Fondazione Eni Enrico Mattei (FEEM), Milano.

Hsiang SM, Burke M and Miguel E (2013) Quantifying the influence of climate on human conflict. *Science (New York, N.Y.)* **341**, 1235367.

Interior (1935) *Land Population Survey of Cities and Counties in China*. Nanjing, China: Department of Statistics, Ministry of the Interior.

Knack S and Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics* **112**, 1251–1288.

Kousky C (2016) Impacts of natural disasters on children. *The Future of Children* **26**, 73–92.

Lee DS and Lemieux T (2010) Regression discontinuity designs in economics. *Journal of Economic Literature* **48**, 281–355.

Liang R (2022) Natural calamity and cultural formation: a study on Yellow River flooding region. *China Economic Quarterly International* **2**, 15–28.

Maccini S and Yang D (2009) Under the weather: health, schooling, and economic consequences of early-life rainfall. *American Economic Review* **99**, 1006–1026.

Mankiw NG, Romer D and Weil DN (1992) A contribution to the empirics of economic growth. *The Quarterly Journal of Economics* **107**, 407–437.

Modern Scientific Research Institute of the Central Meteorological Administration (1981) *Atlas of Drought and Flood Distribution in China in the Last 500 Years*. Beijing, China: Map Press.

Mogues T and Carter MR (2005) Social capital and the reproduction of economic inequality in polarized societies. *The Journal of Economic Inequality* **3**, 193–219.
Nakamura H, Dorjjadamba R and Sodnomdarjaa D (2017) The impact of a disaster on asset dynamics in the Gobi region of Mongolia: an analysis of livestock changes. *The Journal of Development Studies* **53**, 1944–1961.
Nunn N (2009) The importance of history for economic development. *Annual Review of Economics* **1**, 65–92.
Nunn N and Wantchekon L (2011) The slave trade and the origins of mistrust in Africa. *American Economic Review* **101**, 3221–3252.
Otero RC and Marti RZ (1995) The impacts of natural disasters on developing economies: Implications for the international development and disaster community. Report from the Yokohama World Conference on Natural Disaster Reduction, 23–27 May 1994, World Bank, Washington, DC.
Page SE (2006) Path dependence. *Quarterly Journal of Political Science* **1**, 87–115.
Paudel J and Ryu H (2018) Natural disasters and human capital: the case of Nepal's earthquake. *World Development* **111**, 1–12.
Santos I (2007) *Disentangling the effects of natural disasters on children: 2001 earthquakes in El Salvador* (Ph.D. thesis). Harvard University, Cambridge, Mass.
Sawada Y and Takasaki Y (2017) Natural disaster, poverty, and development: an introduction. *World Development* **94**, 2–15.
Skidmore M and Toya H (2010) Do natural disasters promote long-run growth? *Economic Inquiry* **40**, 664–687.
Sotomayor O (2013) Fetal and infant origins of diabetes and ill health: evidence from Puerto Rico's 1928 and 1932 hurricanes. *Economics & Human Biology* **11**, 281–293.
Tabellini G (2010) Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association* **8**, 677–716.
Teixeira AA and Fortuna N (2010) Human capital, R&D, trade, and long-run productivity. Testing the technological absorption hypothesis for the Portuguese economy, 1960–2001. *Research Policy* **39**, 335–350.
Tian X, Gong J and Zhai Z (2022) Natural disasters and human capital accumulation: evidence from the 1976 Tangshan earthquake. *Economics of Education Review* **90**, 102304.
Wang HL (2006) General Catalogue of Chinese Genealogy. *Library Journal* **1**, 73–75.
Writing Group of History of Yellow River Water Resources (2003) *History of Yellow River Water Resources*. Xi'an, Shaanxi: Yellow River Water Conservancy Press.
Xiong H and Zhao Y (2023) Sectarian competition and the market provision of human capital. *The Journal of Economic History* **83**, 1–44.
Xu G (2011) Long-Run Consequences of natural disasters: evidence from Tangshan. Discussion Paper No. 1117, German Institute for Economic Research.
Yamamura E (2011) The role of social trust in reducing long-term truancy and forming human capital in Japan. *Economics of Education Review* **30**, 380–389.
Zhang Z and Ruan J (2020) Do long-run disasters promote human capital in China? – The impact of 500 years of natural disasters on county-level human-capital accumulation. *International Journal of Environmental Research and Public Health* **17**, 7422.
Zhu BJ and Xie PL (1980) *Ming-Qing jinshi timing beilu suoyin*. Shanghai, China: Shanghai Classics Publishing House (in Chinese).

Cite this article: Yu W, Hu J (2024). The long-term effects of natural disasters on human capital accumulation: a quasi-natural experiment based on the Yellow River floodplain area. *Environment and Development Economics* **29**, 279–295. https://doi.org/10.1017/S1355770X24000111