Irrigating the Killing Fields:

Economic Legacy of the Khmer Rouge Genocide

Rahul Mehrotra[†]

Job Market Paper

Click here to access latest version

Abstract

Radical nation-building under the Communist Party of Kampuchea resulted in approximately 2 million deaths, i.e. between 20-25% of the Cambodian population between 1975-79. How did this human capital destruction impact long-term economic development? Satellite data on the regime's failed irrigation projects provide a source of exogenous variation in mass graves. Spatial power relations between moderate and extremist administrative zones can be further exploited as a border discontinuity. My analysis finds persistent impacts of historical genocide on household wealth and village-level poverty, driven by lower human capital, limited structural transformation, restricted public goods provision and private investment flows. Directly impacted households and age-cohorts are significantly worse-off, while there are no differences in wealth for survivors with higher pre-genocide human capital. Political repression measured by proximity to Khmer Rouge prisons predicts significantly worse outcomes. There are no significant differences in internal migration rates among survivors. This paper contributes novel, country-level evidence on how the worst genocide since the Holocaust impacts economic development in Cambodia.

Keywords: political violence, genocide, nation-building, poverty, household wealth, human capital

JEL codes: D74, G5, I25, I3, O12, O53

[†]Address: Graduate Institute of International and Development Studies, Chemin Eugene-Rigot 2A, 1202 Genève, Switzerland. Email: rahul.mehrotra@graduateinstitute.ch. Website: rahul-mehrotra.com I am grateful to Tilman Brück, James Fenske, Ugo Panizza and Phillip Vervimp for helpful comments on this project. Thanks to participants at the North East Universities Development Consortium Annual Conference (Harvard University), Households in Conflict Network Annual Meeting (Warwick University); Annual Conference on Economic Growth and Development (Indian Statistical Institute, Delhi); Graduate Institute of International and Development Studies, Geneva; Centre for Study of African Economies (Oxford University); German Development Economics Conference (Technische Universität Dresden); Nordic Conference on Development Economics (University of Gothenburg), CAGE Research Centre (Warwick University) and Development Economics Network (DENS) Switzerland for their feedback. Any errors are my sole responsibility.

Agriculture is key both to nation-building and to national defence.
- Saloth Sar (alias Pol Pot), April 1975

1 Introduction

Genocide is the gravest crime against humanity.[1] Across the 20th century, multiple episodes of genocide and mass killings were implemented by extremist political regimes during periods of radical nation-building. Existing social and economic institutions were overthrown to build new states with a dominant cultural and economic identity. Repression against opposition groups was driven by competition over political control of the state and its economic resources, often exacerbated by threat of external conflict, adverse economic shocks and natural disasters (Alesina et al., 2020; Besley and Persson, 2011; Esteban et al., 2015). The resulting mass killings represent a significant shock to the human capital stock, however, existing analyses of the economic legacies of genocide find conflicting evidence. For instance, Yanagizawa-Drott and Rogall (2014) find positive impacts of the Rwandan genocide on household wealth and consumption, while Bai and Wu (2023) find a null effect of mass killings during the Chinese Cultural Revolution on industrialization in the long-term. Meanwhile, Acemoglu et al. (2011) find negative associations between the Holocaust in Russia and urban growth rates and per capita incomes. Overall, this literature remains limited and there is a paucity of evidence on the economic legacy of the largest genocidal episodes of the last century.

Radical nation-building under the Communist Party of Kampuchea, whose members were referred to as the Khmer Rouge, resulted in the largest recorded genocidal episode since the Holocaust.[2,3] The Khmer Rouge captured military control of Cambodia in 1975 after an extended period of institutional instability and civil war following independence from French colonial administration. Between 1975-79, the regime aimed to transform the socio-economic and environmental landscape of Cambodia to build a new nation-state called Democratic Kampuchea. Inspired by the radical Maoist vision of an egalitarian society with a prosperous agrarian economy, the Khmer Rouge implemented their own version of a Great Leap Forward economic plan. This involved abolition of all private property and existing currency, collectivization of land and evacuation of the urban population into agricultural labour camps. Members of previous governments, socio-economic elites

[1] The United Nations' Convention on the Prevention and Punishment of the Crime of Genocide (1948) defines genocide as any act committed with intent to destroy, in whole or in part, a national, ethnic, racial or religious group.

[2] Another comparable episode is the cultural revolution period in China (1966-71) when mortality equalled approximately 1.1 to 1.6 million deaths (Walder, 2014).

[3] Khmer Rouge is the popular, anglicized form of the original french term - *Khmers Rouges* referring to the communist Khmer people.

and ethnic minorities were targeted as a threat to internal security and were separated to prisons. Overall, this period resulted in approximately 2 million people, or between 20-25% of the national population, being killed due to political executions, famine and disease. Ethnic Vietnamese and Cham Muslim minority groups were especially targeted, while the ethnic majority Khmer were also killed in large numbers. The demographic effects of this period are well documented (Heuveline, 2015; Kiernan, 2003; de Walque, 2005). However, there is a lack of country-level analyses on the magnitude and mechanisms through which its economic legacy continues to impact Cambodia.

In this paper, I combine historical, geo-spatial information from before and during the Khmer Rouge period with individual, household and village-level data to address the research question: how does exposure to the genocide impact long-term economic development in Cambodia? Exposure is measured using proximity to mass graves locations identified across the country using local surveys and compiled by the Documentation Center of Cambodia (DC-CAM, 2022) and Genocide Studies Program (Yale University, 2022). Directly impacted households are identified using reported year of death of household members in Demographic and Health Surveys (DHS) Program surveys. This information is used to analyse the impacts on household wealth and village-level poverty.

Credible identification of causal impacts can be challenging in this context since the mass graves are non-randomly located. I identify and address three main challenges: first, mass grave locations can be potentially correlated with prevailing socio-economic, geographic and climactic characteristics. To address this concern, I digitize pre-1975 maps to construct historical measures of agriculture, industries, population density, road and water-based transportation routes that are combined with detailed geographic and climactic measurements. I also include geographic fixed effects in my analyses to control for significant, unobserved differences between rural and urban population clusters. Second, the Khmer Rouge forcibly evacuated the urban population into labour camps where many died and remaining survivors returned to their original locations after 1979. I investigate whether internal migration rates, especially the forced migration during 1975-79, is driven by proximity to mass graves since this could also explain any observed long-term effects. I do not find any statistically significant differences, including for survivors of the Khmer Rouge period. To further clarify the interpretation of mass graves as a measure of genocide, I analyse DHS household-level data to show that proximity is indeed a robust predictor of excess mortality during 1970s. Third, mass grave locations can be driven by unobserved differences in pre-1975 levels of regional economic development and these pre-trends would introduce omitted variable bias.[4] Since I can not measure historical household wealth and poverty, I analyse historical documentation produced by the Khmer Rouge regime to conceptualize two novel research designs for causal identification of long-run effects.

First, I exploit the failed agricultural infrastructure construction program of the Khmer Rouge to implement an instrumental variable (IV) strategy. The Communist Party of Kampuchea's economic plan published in 1976 identified increasing rice production and export as an economic priority to support domestic industrialization. Production targets set by the Party's Central Committee equalled a minimum 3 tonnes of rice per hectare per year across the country, increasing to 6 tonnes per hectare by doubling harvests per year. New irrigation infrastructure was identified as the critical input required to increase production from the previous level of approximately 1 tonne per hectare per year. In total, approximately 7,000 kilometres (km) of canals and hundreds of reservoirs were rapidly constructed across the country to meet the new targets (Chandler et al., 1988). However, this top-down, rapid expansion ignored natural hydrological systems, drainage patterns and soil suitability resulting in a catastrophic failure of the agricultural system (Bultmann, 2012; Himel, 2007). Conditional on the inclusion of a battery of geographic characteristics and pre-1975 socio-economic indicators, the locations of these abandoned and defunct irrigation reservoirs are therefore argued to be exogenous to long-run economic development. I digitize maps of large agricultural reservoirs that are geo-located by researchers comparing pre-1975 satellite images with post-1979 images taken soon after the regime's collapse (Tyner et al., 2018).

A battery of robustness tests are used to evaluate whether the IV is a robust predictor of mass grave locations, evaluate the consistency of estimations and ensure spatial-robust inference (Keane and Neal, 2022; Kelly, 2019; Young, 2022). As a conceptual test, I further demonstrate that the IV loses it's predictive power when using a zero first-stage sample, i.e. the IV only works in empirical settings consistent with its theoretical motivation (Lal et al., 2021; Rohner et al., 2013a). Plausibly exogenous IV estimators and causal machine learning methods are also deployed to re-confirm the validity of the IV strategy (Conley et al., 2012; Chernozhukov et al., 2015, 2018).

Second, I exploit the structure of spatial power relations within the Khmer Rouge administration to implement a border regression discontinuity design. Democratic Kampuchea followed a decen-

[4] I hypothesize that a naive ordinary least square (OLS) regression of present day development indicators on mass graves would yield an upward biased coefficient since omitted measures of historical economic development would be positively correlated with both the dependent variable and endogenous independent variable, respectively. Historically richer locations are expected to be richer today and the Khmer Rouge movement represented a rural uprising against urban elites, hence richer, more developed locations witnessed higher mortality.

tralized system of internal administration whereby the state was divided into distinct military-administrative zones. These zones emerged from local campaigns under regional commanders of the Khmer Rouge during the Cambodian civil war (1970-75). The respective military commanders transitioned to become Zone Secretaries after 1975 with a significant degree of autonomy over implementation of policies drafted by the party's Central Committee (Jackson, 1989a; Mertha, 2014; Vickery, 1986).[5] As a result, living conditions and level of political repression varied significantly across the different zones. I focus on the Northeast Zone originally under the command of a relatively moderate Central Committee member, Ney Saran (also known as Comrade Ya). As a party veteran, Ney Saran reportedly disagreed with other Khmer Rouge leaders over the rapid pace of land collectivization, evacuation of urban centres, and political killings of fellow Cambodians (Chandler et al., 1988; Human Rights Watch, 2007). Therefore, I argue that the border between the Northeast Zone and rest of Cambodia represents a geographic discontinuity between a moderate and extremist zone. A battery of proxy tests are used to test the validity of the research design relying upon no statistically significant differences in observable geographic, environmental, and pre-1975 economic activity across the border discontinuity (Calonico et al., 2019a,b).

My primary outcomes of interest include household wealth and village-level poverty rates. The DHS wealth index is provided by 3 waves of the Cambodia DHS surveys in 2005, 2010 and 2014. This index score measures each household's cumulative living standard measured using ownership of selected assets, housing characteristics and access to public goods and services. The DHS wealth index is a robust predictor of household welfare, especially health and education. Village-level poverty headcount rates are provided by the Cambodian IDPoor survey database constructed using two survey waves conducted in 2010 and 2011. This provides a measure of multidimensional poverty that accounts for levels of income, debt, asset ownership, health and education.

In terms of mechanisms, I focus on measures of human capital, structural transformation as well as public and private investment flows. These mechanisms are driven by the large-scale human capital destruction due to the genocide, including loss of lives, skills and experience of the victims, as well as foregone years of education of the survivors. Public investments in piped water supply, telephone network connectivity and provision of financial services are analysed to better illuminate the state's response to this historical genocide. Land concessions allocated for domestic and foreign investors

[5]Further sub-divisions within these zones were delineated after 1975, however the initial zone boundaries remain unchanged. These boundaries do not correspond with any pre-1975 or post-1979 geographic units, i.e. districts and provinces.

are also analysed as a measure of private investment flows to regions exposed to the genocide.

My findings can be summarized as follows: First, ordinary least squares (OLS) regressions of household wealth on exposure to the genocide (= number of mass graves in 10 kilometre buffer around DHS clusters) find a small negative coefficient that is not statistically significant. IV first-stage regression confirms that reservoirs (= total in a 10 kilometre buffer around DHS clusters) is a robust predictor of mass graves. The IV second-stage results find an economically significant decrease in relative wealth for households in close proximity to mass graves. One σ (standard deviation) increase in the number of mass graves (= 432 graves) has a negative impact equivalent to 0.29σ of relative wealth. A comparison of the OLS and IV coefficients indicates that the OLS coefficient is upward biased, as predicted by the hypothesized sign of the omitted variable bias driven by unobserved measures of historical economic development. In terms of the nationally-representative household wealth distribution, the mean household is 3.3 percentage points (p.p.) more likely to be in the poorest quintile (16.3 p.p. less likely to be in the richest quintile). Next, the geographic-RDD estimations show that village-level poverty rate is higher (11 p.p.) in villages on the radical area compared to the moderate Northeast Zone (sample mean = 28.1%).

Second, IV estimations to evaluate channels of persistence find significant impacts of proximity to mass graves on contemporary levels of human capital, structural transformation and investment flows. Results indicate that heads of households with the mean level of exposure to the genocide have 0.5 fewer years of education (1.8 fewer years for the median exposed head of household). The sample average for years of education equals only 4.5 years, hence these coefficients represent large, economically significant impacts. Splitting the total sample by birth cohorts confirms that the estimated impacts are driven by the cohorts born in the decades preceding the Khmer Rouge.

Third, I explore the impact of mass grave sites on the occupational choices of DHS survey participants in order to evaluate comparative patterns of structural transformation in the economy. The results indicate that adult males in mean household are 13 p.p. more likely to be self-employed in agricultural sector (20 p.p. more likely for adult females). Conversely, adult males are 6.5 p.p. less likely to be employed in the services sector while the corresponding estimates for adult females is 13 p.p. Comparing with sample means indicates these are economically significant effects.

Fourth, regions with mean number of mass grave sites receive a lower provision of public goods and investment projects from the state. The mean household is less likely to have access to piped water-supply (13 p.p.), telephone connections (1 p.p.) and access to financial services (7 p.p.). Economic land concessions for domestic and foreign investment projects are 3.3 p.p. less likely to be allocated to regions with mean exposure to the genocide. Same as before, comparing these estimates with sample means indicate economically significant negative effects in the state's response.

Finally, my RDD analysis using DHS clusters is consistent with the previous IV results for channels of transmission. I find that households close to the border in the radical Khmer Rouge zone have lower wealth, restricted access to public goods, including piped water, telephone connectivity and financial services. Adult household members are less educated, more likely to be self-employed in agriculture and less likely to be employed in services sector. I do not observe robust differences in other covariates indicating the validity of the research design.

I add several extensions to my baseline results. I show that directly impacted households (measured by survey respondents' siblings' year of death occurring between 1970-79) are significantly worse off. Furthermore, households exposed to political repression in the prison system of the Khmer Rouge are also significantly poorer. Individuals with higher pre-genocide levels of human capital who survived the genocide do not appear richer and remain equally affected by the genocide. Finally, there are no statistically significant differences in migration rates driven by exposure to the genocide which allows me to rule out an alternative mechanism driving these findings. I also re-estimate the above results using alternative plausibly exogeneous IV estimators and lasso-IV machine learning methods. The results are consistent with my baseline analysis and presented as a robustness check.

This paper makes multiple contributions to the economic literature on political conflict and economic development. First, this research contributes to a robust growing literature on political violence and nation-building. Besley and Persson, 2011, Esteban et al., 2015 and Alesina et al., 2020 provide the foundational frameworks to analyze the drivers of political violence used to maintain or acquire power in a within-country conflict setting. This research highlights the role of shocks to income and aid, competition over political control of the state and its economic resources, and nation-building motivations in driving one-sided (i.e. repression) or two-sided violence (i.e. civil war). My findings provide supporting empirical evidence informed by these economic theories of political violence used to build new nation-states with a dominant cultural identity, at the cost of minorities and political opposition groups. There is a related empirical literature focused on forced migrations due to political violence focusing predominantly on Europe and Africa, resulting in a paucity of evidence on some of the largest episodes of mass killings from Asia (Becker et al., 2022; Justino, 2016).

Second, I provide novel, country-level evidence to highlight the magnitude and channels through which large-scale, one-sided political violence impacts comparative economic development. There is a limited literature on this question that finds conflicting results. In Rwanda, Yanagizawa-Drott and Rogall (2014) find positive impacts of Hutu-Tutsi genocide on household wealth and consumption which they argue to be consistent with the Malthusian hypothesis, i.e. existing productive assets are held by fewer individuals post-genocide. Meanwhile, Acemoglu et al. (2011) focus on the anti-Jewish holocaust in Russia and their analysis shows that mass killing areas are associated with slower urban growth rates and per capita incomes, which is explained by a change in the social structure. Finally, in a recent contribution Bai and Wu (2023) find short-term slowdown in industrialization in locations experiencing mass killings during the Chinese Cultural Revolution, however this effect dissipates leading to an long-term null effect. Therefore, the developmental impacts of large-scale genocide on affected states remains an open question in this literature.

Next, there is a growing focus on the Cambodian genocide in recent social science literature building upon the original demographic analysis from de Walque (2004). In contemporaneous work, Gangadharan et al. (2022) use lab experiments to study the prevalence of antisocial behaviour among individuals exposed to the genocide. In exhaustive research on the political and institutional legacy of the Khmer Rouge, Bühler and Madestam (2023) use rainfall shocks to proxy for local variation in mass killings and highlights the long-term impacts on political beliefs and voting behaviour. Finally, Grasse (2023) compares economic outcomes across the West and Southwest military zones to evaluate impacts on local poverty within Kampong Speu province. My paper uses novel identification strategies informed by historical research on the Khmer Rouge to provide country-level insights that are consistent with the findings reported in this literature.

Finally, there is a distinct literature on the impacts of violent conflict between states and within states, i.e. civil war. Unlike one-sided political violence studied in my paper, this literature studies impacts driven by the destruction of military and civilian infrastructure, displaced populations and disrupted education, as well as damage to social capital and political institutions. There is robust evidence of devastating short-term impacts, however, long-run analyses remains limited and often contradictory.[6] My paper contributes novel evidence on the conditions under which large-scale human capital destruction during political violence results in long-term poverty.

[6]Short-term analyses have focused on the impact of conflict on education (Akresh and de Walque, 2008; Shemyakina, 2011), child anthropometrics (Akresh et al., 2011; Akresh et al., 2012; Bundervoet et al., 2009) and social capital (Rohner et al., 2013b). Long-term analyses of conflict remain limited to Davis and Weinstein (2002) and Brakman et al. (2004) on urban growth; Ichino and Winter-Ebmer (2004), Akbulut-Yuksel (2014) and Merrouche (2011) on education attainment, and Miguel and Roland (2011), Riaño and Valencia Caicedo (2020) on poverty.

2 Historical Background: Radical Nation-building under the Khmer Rouge

The Cambodian communist movement emerged from the struggle against French colonization in the 1940s. After 9 decades as a French colonial protectorate, the Kingdom of Cambodia was established as an independent constitutional monarchy in 1953. However, communist factions sought to install a Marxist-Leninist government and established the Communist Party of Kampuchea. Party members were popularly referred to as the Khmer Rouge and they initiated a rural insurgency by 1968 that escalated into a civil war between 1970-75. The Khmer Rouge took control of the capital Phnom Penh in April 1975 and announced the establishment of the new state of Democratic Kampuchea. The state lasted less than 4 years before suffering economic collapse and military invasion by neighbouring Vietnam due to escalating border conflicts.

The Khmer Rouge regime's overall objectives were twofold: first, implement a great leap forward, i.e. Maoist China-style rapid socialist revolution to build a rural classless society, and, second, defend the new state against internal and external enemies. The urban, civilian population was forcibly evacuated into labour camps in rural areas to build the new agrarian economy. Significant numbers perished in these camps over the next four years due to forced labour, famine and disease. The mass graves discovered across Cambodia in subsequent years indicate the locations of these killing fields. Prison camp locations further delineate the centres of political violence where the regime's opponents, including members of previous government, socio-economic elites and minority ethnicities were held and executed (Chandler, 1999; Jackson, 1989b; Kiernan, 2002).

2.1 Rice Cultivation and Irrigation: Economic Plan 1976-79

The Communist Party of Kampuchea published a four year economic plan in 1976 that prioritized farmland collectivization and rapid expansion of rice cultivation. Earnings from rice exports were deemed critical to finance the new state's import substitution industrialization policy. The plan stated, *Objective: To produce rice for food to raise the standard of living of the people, and in order to export so as to obtain capital for the imports which we need. Thus paddy and milled rice are our capital base...For 100,000 tons of milled rice, we would get $20 million; if we had 500,000 tons we would get $100 million. We must increase rice production in order to obtain capital.* (Page 51, Chandler et al., 1988)

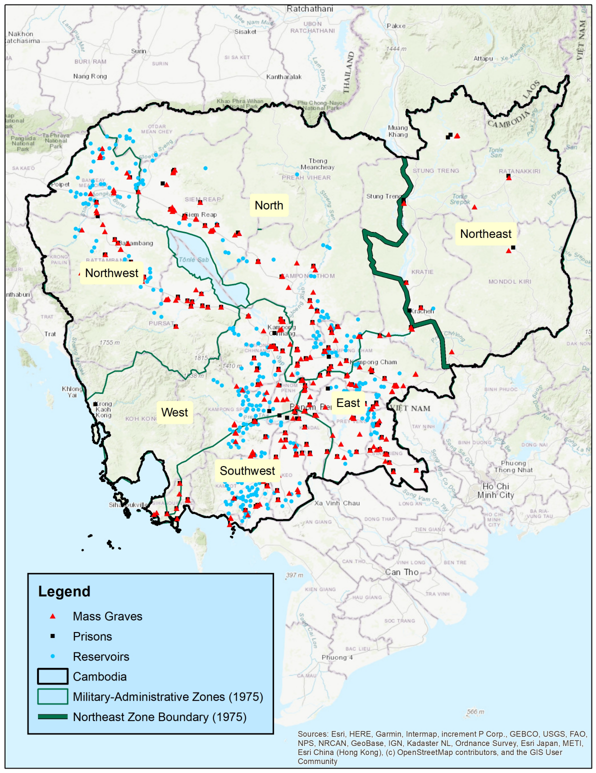


Figure 1: Democratic Kampuchea: Mass Graves, Prisons, Irrigation Reservoirs and Military Zones

Lack of adequate irrigation was identified as the primary obstacle for increasing rice production beyond limits imposed by traditional rain-fed irrigation. The economic plan set the following target, *Increase the degree of mastery over the water problem from one year to another until it reaches 100% by 1980 for first-class rice land and reaches 40 - 50% for ordinary rice land. In order to gain mastery over water there must be a network of dikes and canals as the basis. There must also be canals, reservoirs, and irrigation pumps stationed in accordance with our strategy.* (Page 89, Chandler et al., 1988). To achieve this economic plan, urban centres were evacuated and the entire civilian civilian population was set to work on rice cultivation, including clearing forests to expand rice fields and building vast networks of reservoirs linked by canals and dykes to provide additional irrigation. Ambitious rice production and export targets were set for the entire country and enforced by local commanders using forced labour (Twining, 1989; Chandler et al., 1988; Jackson, 1989a).

2.2 Democratic Kampuchea: Internal Administration

The Communist of Party of Kampuchea's new government in 1975 was assembled from previously fragmented local communist movements under different commanders who maintained considerable political and military power in the new state. Therefore, the Party divided Democratic Kampuchea into six *phumipheak* or administrative zones: North, Northeast, Northwest, East, Southwest, and West, plus the capital Phnom Penh (see internal boundaries demarcated in Figure 1).[7] These internal zones emerged from the regional military commands during the 1970-75 Cambodian civil war and the respective military commanders transitioned to become Zone Secretaries. The borders of these internal zones did not correspond with administrative boundaries for districts or provinces. As a result, this spatial administration system was temporary and exogenous to pre-1975 or post-1979 administrative features.

The zone borders are of particular interest since the implementation of policies drafted the Party's central committee were the responsibility of respective zone administrations. Consequently, Zone Secretaries were highly influential in their territories since the different commands were highly decentralized and a significant degree of autonomy was maintained across each vertical administrative unit. As a result, the living conditions in each zone varied significantly, especially with respect to the labour and security policies. Policies relating to farm collectivization, forced labour and political executions were applied with varying intensity driven by the degree of ideological alignment between the Central Committee in Phnom Penh and the respective Zone Secretaries and their administrations (Jackson, 1989a; Mertha, 2014; Vickery, 1986).

Overall, the Khmer Rouge's economic plan and it's administrative set-up was a catastrophic failure. Workers pressed into forced labour lacked agricultural experience and irrigation systems were hastily constructed under considerable pressure. Failure to meet rice production and export targets led to purges of local commanders and violence against workers. In several territories, the entire production had to be exported without maintaining any reserve to feed workers or seed the next harvest. As a result, the agricultural system completely collapsed by 1978 leading to wide-spread famine. The Khmer Rouge regime was finally deposed in 1979 after escalating border disputes led to a full-scale invasion by neighbouring Vietnam.

3 Data, Summary Statistics and Stylized Facts

This research compiles multiple historical and contemporary geo-spatial databases using two distinct units of analysis: first, I use DHS clusters from three survey waves (2005, 2010 and 2014) that

[7] Some zones were further sub-divided between 1976-78 to establish new sub-zones and autonomous sectors, often under the direct control of the Communist Party of Kampuchea's central organization.

are geo-located using their centroid coordinates. The sampling frame used for the Cambodian DHS survey is the complete list of all enumeration areas (EAs) of the General Population Census. Provinces are allocated into 19 domains which are further sub-divided into rural and urban strata. A two stage random-selection procedure is then implemented within each of the resulting 38 stratum to select EAs, followed by households within each EA. Survey respondents include individuals (heads of household and adult household members, especially females). To maintain anonymity, rural DHS clusters include survey respondents who are randomly displaced 10 kilometres (km) around the centroid, while urban clusters include respondents who are randomly displaced 5km around the centroid. For my baseline analyses, I build 10km buffer regions around each cluster radius to calculate my variables of interest, including mass graves and reservoirs. Second, I use village-level data where each village is identified by its centroid coordinates. The Cambodia IDPoor database provides the main variables of interest and I merge other geo-spatial variables using the nearest measurement to each village centroid. The data sources, variables of interest and descriptive statistics are discussed below:

Cambodia Demographic and Health Surveys: Information from 2 different DHS questionnaires are used in this analysis and reported in Table 1. The pooled sample includes 45,380 households from 1,765 DHS clusters spanning 3 waves of the survey, i.e. 2005, 2010 and 2014. First, the *household head questionnaire* is used to calculate the household wealth and access to public services (bank account, piped water and telephone) at the household-level, as well as the age, gender and education of the household head. As reported in Table 1: Panel A, the mean respondent is 46 years old and they have only 4.5 years of education. Approximately half the sample is born before 1975, i.e. the Khmer Rouge period, and a quarter of households are headed by female respondents. Next, the *female questionnaire* is used to record the occupation of adult females and their male spouse or partner. As reported in Table 1: Panel B, approximately 40% of the sample works in agricultural sector. A higher proportion of women are reportedly employed in the services sector (28%) compared to men (20%), while the reverse is observed for the manual work.

The DHS Wealth Index provides a useful proxy for expenditure and income-based measures of economic well being. International evidence indicates this measure is robust predictor of household welfare, especially in terms of health and education. The Index Score can be interpreted as an indicator of relative wealth calculated using information collected on household characteristics, consumer goods and assets owned, and access to public services. Each household asset is assigned a factor score generated through principal components analysis and the resulting asset scores are standardized with a mean of zero and a standard deviation of one. These scores are summed by household using which the sample can be divided into population quintiles (i.e., five groups with the same number of individuals in each). Wealth index scores can be used to evaluate relative well-being of households in close proximity to mass graves, in comparison to the mean household. The wealth quintiles can be used to evaluate where households in close proximity to mass graves fall in terms of their position in the nationally representative wealth distribution.

Genocide Data: Data on Democratic Kampuchea including the locations of mass graves, prisons and administrative zone borders are provided by the Cambodian Genocide Program's Interactive Geographic Database (Yale University, 2022). The database includes information on approximately 19,000 mass graves, 158 prison locations and administrative zone borders from the 1975-79 period. This information is merged with the DHS clusters using the procedure described previously, i.e. using a 10 km buffer around each rural and urban cluster. The summary statistics are reported in Table 1: Panel C and indicate that on average, 318 mass graves with 19,497 bodies are located in close proximity to a population cluster. The distribution is negatively skewed and the median DHS cluster has a significantly higher number of mass graves (1,258) and recovered bodies (79,375). The largest share of DHS cluster centroids (20%) are observed in the densely populated region that composed that erstwhile South-west military-administrative zone.[8]

Agricultural Irrigation Projects: Tyner et al. (2018) use remote sensing, archival records and field observation to systemically document the Khmer Rouge period irrigation schemes. Landsat satellite imagery is available from 1972 onward with sufficient spatial resolution to identify agricultural infrastructure. Pre (1973 - 75) and post-Khmer Rouge (1979 - 84) satellite images are analysed to identify locations of human-produced features, such as dams, canals, and reservoirs, followed by site visits to field-check the remote observations with surviving local population. Figure 1 presents a cartographic representation of the Khmer Rouge-era reservoirs used in this research.[9]

Geographic and Climate Data: Geographic characteristics including terrain ruggedness, elevation, slope are calculated for each DHS cluster using information provided by the EarthEnv

[8] Households within each cluster are randomly displaced to maintain their anonymity hence individual observations may fall outside the corresponding zone. This issue is addressed by applying doughnut regression discontinuity design when using DHS cluster-level data for the RDD analysis, i.e. dropping of rural clusters with centroids less than 10km from the Northeast Zone border and urban clusters with centroids less than 5km from the border.

[9] I use reservoirs since they represent spatially distinct and identifiable features that were crucial for water storage for irrigation.

database (Amatulli et al., 2018). Climactic characteristics including average annual rainfall and average temperature measurements are calculated for each cluster using data from the WorldClim 2.1 database (Fick and Hijmans, 2017).

Pre-1975 Economic and Demographic Data: I digitize historical maps provided by the Indochina Atlas (Central Intelligence Agency, 1970) to measure the locations of pre-1975 agricultural activity, industrial infrastructure, population density, roads, rivers and water-ways. I also use these historical maps to geo-locate the urban centres, including the national capital Phnom Penh and provincial capitals, that are used to calculate their distance from DHS clusters.

Village-level Poverty Data: I use the Identification of Poor Households Programme (IDPoor) database covering two rounds of village-level surveys from 2010 and 2011 implemented by the Cambodian Ministry of Planning. Since 2011, the IDPoor survey has been legally designated as the country's official poverty targeting mechanism and used nationwide by the public and private social sectors. According to this survey, villagers elect representatives amongst themselves to interview households using standardised, multi-dimensional poverty criteria including household assets, income, debt, education and health measurements. The village committees then discuss the survey results to finalise a list of poor households for each village. This procedure is run recurrently throughout the country to provide up-to-date poverty targeting data disaggregated at different administrative levels. The data is available via the Open Development Cambodia database and the implementation manual via the IDPoor official website.[10]

The survey sample includes information on 10,620 villages and the summary statistics are reported in Table 2. Each village can be geo-located using the geographic coordinates of their centroid. These coordinates allows me to calculate the distance to various features of interest, including mass grave sites, reservoirs constructed during 1975-79, and pre-1975 economic activities, including different types of agriculture and industries. The geographic and climactic characteristics (terrain ruggedness index, mean annual rainfall, mean annual temperature, slope, elevation) are calculated for each village centroid using the same data sources mentioned previously for DHS clusters.[11]

Stylized Facts: I use the assembled data to first derive two descriptive stylized facts to inform the following research designs. First, Figure 2 presents mortality rates by birth cohorts calculated

[10] Open Development Cambodia database can be accessed here: https://data.opendevelopmentcambodia.net (accessed: August 2022). IDPoor website is available here: https://idpoor.gov.kh/en/ (accessed: June 2023).

[11] Since the 2010 and 2011 survey rounds do not cover all provinces, further analysis will utilize alternate data sources for commune and district-level poverty rates

Table 1: Demographic and Health Survey Cluster Characteristics: Summary Statistics

	Obs.	Mean	S.D.	Min	Max
Panel A: Demographic and Health Surveys (2005, 2010, 2014): Household Head Questionnaire					
Wealth Index Score	45,380	-1,500	553,259	-934,030	5,277,000
Wealth: 5th Quintile (Poorest)	45,380	0.204	0.403	0	1
Wealth: 4th Quintile (Poorer)	45,380	0.198	0.399	0	1
Wealth: 3rd Quintile (Middle)	45,380	0.184	0.388	0	1
Wealth: 2nd Quintile (Middle)	45,380	0.189	0.392	0	1
Wealth: 1st Quintile (Richest)	45,380	0.225	0.417	0	1
Age of Household Head (years)	45,380	45.98	13.87	18	98
Year of Birth: 1990-99	45,380	0.0121	0.109	0	1
Year of Birth: 1980-89	45,380	0.142	0.349	0	1
Year of Birth: 1975-79	45,380	0.0923	0.289	0	1
Year of Birth: 1965-74	45,380	0.265	0.442	0	1
Year of Birth: 1955-64	45,380	0.235	0.424	0	1
Years of Education	45,127	4.555	3.884	0	20
Female Household Head	45,380	0.250	0.433	0	1
Piped water connection (dummy)	45,380	0.135	0.342	0	1
Owns bank account (dummy)	31,366	0.108	0.314	0	9
Own telephone (land-line)	31,368	0.0866	0.293	0	9
Panel B: Demographic and Health Surveys (2005, 2010, 2014): Female Questionnaire					
Occupation: Agriculture (Female)	46,156	0.398	0.489	0	1
Occupation: Services (Female)	46,156	0.281	0.449	0	1
Occupation: Manual Labour (Female)	46,156	0.117	0.322	0	1
Occupation: Agriculture (Male)	46,156	0.382	0.486	0	1
Occupation: Services (Male)	46,156	0.199	0.399	0	1
Occupation: Manual Labour (Male)	46,156	0.175	0.380	0	1
Panel C: Geographic Characteristics (10km buffer around DHS Cluster)					
Graves (#)	1,765	317.8	431.7	0	2,516
Bodies (#)	1,765	19,497	28,953	0	158,750
Reservoirs (#)	1,765	3.922	5.564	0	31
Military Zone: Central Phnom Penh	1,765	0.0918	0.289	0	1
Military Zone: East	1,765	0.148	0.355	0	1
Military Zone: North	1,765	0.136	0.343	0	1
Military Zone: North-east	1,765	0.1006	0.256	0	1
Military Zone: North-west	1,765	0.153	0.360	0	1
Military Zone: South-west	1,765	0.195	0.397	0	1
Military Zone: West	1,765	0.146	0.353	0	1
Distance to North-east Zone Boundary (kilometre)	1,765	165.67	84.78	0.06	352.55
Pre-1975 Rice Farming (Dummy)	1,765	0.744	0.437	0	1
Pre-1975 Inland Fish Farming (Dummy)	1,765	0.00510	0.0712	0	1
Pre-1975 Rubber Farming (Dummy)	1,765	0.0833	0.276	0	1
Pre-1975 Forestry (Dummy)	1,765	0.117	0.321	0	1
Pre-1975 Trucks & Vehicles Industry (Dummy)	1,765	0.0198	0.139	0	1
Pre-1975 Cement Industry (Dummy)	1,765	0.00793	0.0887	0	1
Pre-1975 Textiles Industry (Dummy)	1,765	0.0952	0.294	0	1
Pre-1975 Paper Industry (Dummy)	1,765	0.0119	0.108	0	1
Pre-1975 Tyres Industry (Dummy)	1,765	0.0283	0.166	0	1
Pre-1975 Sugar Industry (Dummy)	1,765	0.0147	0.121	0	1
Pre-1975 Population Density (Urban)	1,765	42,935	167,589	0	1,000,000
Pre-1975 Population Density (Rural)	1,765	130.8	126.3	1	700
Pre-1975 Roads (kilometre)	1,757	120.20	47.98	1.29	261.07
Pre-1975 Rivers and Waterways (kilometre)	1,762	144.32	91.47	0.06	518.29
Distance to Province Capital (kilometre)	1,765	23.89	19.18	0.05	90.84
Distance to National Capital (kilometre)	1,765	150.93	102.28	0.31	373.41
Economic Land Concession (dummy)	1,765	0.616	0.487	0	1
Terrain Ruggedness Index (Mean)	1,765	3.565	3.797	0.680	32.04
Slope (Mean)	1,765	0.882	1.023	0.120	8.938
Elevation (Mean)	1,765	52.15	75.07	3	694
Rainfall (millimetre, Annual Mean)	1,765	135.3	40.87	84.75	331
Temperature (Degrees Celsius, Annual Mean)	1,765	27.42	0.631	22.71	28.11

Notes: This table reports the summary statistics for the DHS cluster-level database containing 45,380 households across 1,765 clusters pooled from 3 waves of the Cambodia DHS surveys in 2005, 2010 and 2014. Panel A reports summary statistics for the DHS Household Head questionnaire focusing on household wealth, access to public goods and respondents' demographic characteristics. Panel B reports the summary statistics for variables sourced from the Female Questionnaire, focusing on occupational sectors. Panel C reports geo-spatial measurements calculated for a 10km buffer area around each DHS cluster. Measurements of the genocide are sourced from Cambodian Genocide Studies Database (Yale University, 2022). Information on irrigation reservoirs is from Tyner et al., 2018. Pre-1975 characteristics are sourced from the Indochina Atlas (Central Intelligence Agency, 1970). Climatic variables are sourced from the EarthEnv database (Amatulli et al., 2018) and geographic measurements are provided by the WorldClim 2.1 database (Fick and Hijmans, 2017).

Table 2: Village-level Characteristics: Summary Statistics

	Villages	Mean	S.D.	Min	Max
Poverty Rate (%)	10,620	0.281	0.144	0	3.850
Households (#)	10,620	53.59	44.44	0	624
Distance to Khmer Rouge Mass Grave (kilometre)	10,620	12.96	16.21	0.04	127.87
Distance to Khmer Rouge Reservoir (kilometre)	10,620	18.70	28.61	0.05	204.62
Military Zone: Central Phnom Penh	10,620	0.00621	0.0786	0	1
Military Zone: East	10,620	0.185	0.389	0	1
Military Zone: North	10,620	0.130	0.336	0	1
Military Zone: North-east	10,620	0.0455	0.176	0	1
Military Zone: North-west	10,620	0.189	0.391	0	1
Military Zone: South-west	10,620	0.274	0.446	0	1
Military Zone: West	10,620	0.165	0.371	0	1
Distance to North-east Zone Boundary (kilometre)	10,620	183.27	68.03	0.06	357.87
Pre-Khmer Rouge Rice Farming (Dummy)	10,620	0.474	0.499	0	1
Pre-Khmer Rouge Inland Fishing Area (Dummy)	10,620	0.0618	0.241	0	1
Pre-Khmer Rouge Rubber Farming (Dummy)	10,620	0.000565	0.0238	0	1
Pre-Khmer Rouge Forests (Dummy)	10,620	0.0305	0.172	0	1
Nearest Pre-Khmer Rouge Industry: Trucks (Dummy)	10,620	0.0303	0.171	0	1
Nearest Pre-Khmer Rouge Industry: Cement (Dummy)	10,620	0.119	0.324	0	1
Nearest Pre-Khmer Rouge Industry: Textile (Dummy)	10,620	0.309	0.462	0	1
Nearest Pre-Khmer Rouge Industry: Paper (Dummy)	10,620	0.232	0.422	0	1
Nearest Pre-Khmer Rouge Industry: Tyres (Dummy)	10,620	0.180	0.384	0	1
Distance to Nearest Pre-Khmer Rouge Industry (kilometre)	10,620	59.78	44.89	0.51	245.28
Terrain Ruggedness Index	10,620	2.184	3.367	0.156	75.81
Slope	10,620	0.490	0.925	0.0459	21.03
Elevation (meters)	10,620	36.00	61.52	1	1,250
Rain (Mean, Annual)	10,620	122.6	30.73	75.58	353.2
Temperature (Mean, Annual)	10,620	27.64	0.524	20.67	28.22

Notes: Data on village-level poverty rates and number of households is provided by the Identification of Poor Households Database (waves 4 and 5 conducted in 2010 and 2011, respectively). Information on the locations of mass graves and military-administrative zones is provided by the Cambodian Genocide Studies Database, while information on irrigation reservoirs is sourced from Tyner et al., 2018. Climatic variables, including rainfall and temperature, are sourced from the EarthEnv database (Amatulli et al., 2018). Geographic variables, including terrain ruggedness, slope and elevation, are provided by the WorldClim 2.1 database (Fick and Hijmans, 2017).

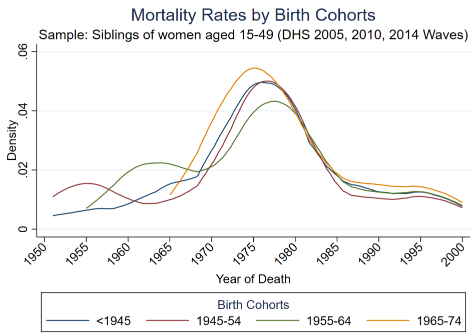


Figure 2: Mortality Rates, by Birth Cohorts

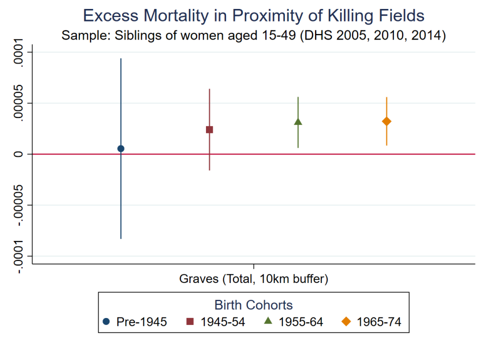


Figure 3: Excess Mortality in Proximity to Mass Grave Locations, by Birth Cohorts

using the sample of siblings born to female respondents in the Cambodian Demographic and Health Surveys (Waves 2004, 2010 and 2014). This survey represents the best available data source to illustrate birth-cohort-level trends in mortality rates for a representative sample of households. The results confirm that the mortality rates for all birth cohorts significantly increased during the Khmer Rouge period.

Second, I estimate excess mortality rates by regressing mortality by birth cohort on proximity to mass graves and include controls for age, gender and geographic location. The results are reported in Figure 3. The coefficients confirm that excess mortality is driven by proximity to mass graves, which can therefore be considered a valid measure of exposure to the genocide.

4 Research Design: Instrumental Variable Strategy

As discussed previously, the Khmer Rouge set a national target of 3 tonnes of rice per hectare and prioritized the rapid construction of irrigation infrastructure. However, this program was a catastrophic failure that contributed to mass famine and starvation deaths. Evidence from modern agricultural surveys of Cambodia confirms that this irrigation infrastructure did not account for local soil characteristics and hydrological systems, especially natural drainage patterns (Himel, 2007; Tyner and Will, 2015; Tyner, 2017). This contributed to failure to meet production targets in adequate quantities to meet strict export quotas and local consumption. After the downfall of the Khmer Rouge government in 1979, these agricultural projects were abandoned and ceased to function as the survivors returned to their places of origin. Recent agricultural surveys indicate the irrigation systems in Cambodia remain significantly underdeveloped and historic irrigation systems built in 1975-79 have become useless and non-functional (ADB, 2019; FAO, 2011; Himel, 2007).

I also exhaustively review the academic literature on the Khmer Rouge agricultural programme to identify potential predictors of reservoir locations. Land suitability for agriculture, proximity of urban centers and pre-1975 population density are measured using historical maps provided by the Indochina Atlas (Central Intelligence Agency, 1970). Additional geographic characteristics including slope, elevation and ruggedness, as well as climactic characteristics including average rainfall and temperature are measured. This battery of geographic characteristics and pre-1975 socio-economic indicators are included as covariates in the IV estimations. On this basis, I argue that the locations of these poorly-planned and since abandoned infrastructure projects are exogenous to long-term economic and political outcomes (Bultmann, 2012; Himel, 2007). Finally, I rigorously test the plausibility of this research design using a zero-first stage sample test, plausibly exogenous IV methods and lasso-IV models estimated using double-machine learning algorithms (Young, 2022; Conley et al., 2012; Chernozhukov et al., 2015, 2018).

To estimate the reduced-form relationship between mass graves and selected outcome variables, I use the following ordinary least squares (OLS) regression:

$$Y_{icpw} = \alpha_0 + \alpha_1 Graves_{cpw} + \alpha_2 X_{cpw} + Strata_p + Wave_w + \nu_{icwp} \tag{1}$$

where Y_{icwp} denotes the outcome variables of education, for example: household wealth, for household i, in DHS cluster c, in DHS strata p, from DHS wave w. $Graves_{cpw}$ denotes the endogeneous measurement of total mass grave sites in the 10km buffer around each DHS cluster c in strata p and from wave w. X_{cpw} denotes the vector of various geographic, climactic and pre-1975 characteristics of each DHS cluster c in strata p from wave w. I include DHS strata fixed effects $Strata_p$ and DHS wave fixed effects $Wave_w$. Standard errors ν_{icwp} are clustered at the level of DHS clusters.

As discussed previously, the reduced-form estimation are biased due to the endogeneity of mass grave locations. As a result, we instrument this endogenous variable with Khmer Rouge reservoirs. This structural relationship is then estimated using two-stage least squares (2SLS) model. The IV-first stage model for regressing the endogenous $Graves_{cpw}$ variable on our instrumental variable $Reservoirs_{cpw}$ is represented as:

$$Graves_{cpw} = \gamma_0 + \gamma_1 Reservoirs_{cpw} + \gamma_2 X_{cpw} + Strata_p + Wave_w + \epsilon_{cpw} \tag{2}$$

Finally, the second-stage estimation involves regressing our selected outcome variables of interest denoted by Y_{icp}, i.e. poverty and related mechanisms measured at the level of household i in DHS-cluster c and province p on the predicted values of the endogenous variables from the first-stage regression. The estimation model is represented by:

$$Y_{icpw} = \beta_0 + \beta_1 \widehat{Graves}_{cpw} + \beta_2 X_{cpw} + Strata_p + Wave_w + \nu_{icpw} \tag{3}$$

where $\widehat{Graves}_{cpw}$ denotes the fitted values from the first-stage model above, while the remaining variables are defined same as before. Standard errors ν_{icpw} are clustered at the level of DHS survey clusters. Spatial-robust standard errors calculated using a 3 decimal degree radius with bartlett linear decay and P-values from wild-bootstrap robust inference are also reported to address empirical concerns with the consistency of IV results, as per Kelly, 2019 and Young, 2022, respectively.[12] Robust weak IV test statistics indicating the first-stage F statistic that must be exceeded to minimize Negar bias are reported (Pflueger and Wang, 2015). P-values from an Anderson-Rubin weak IV robust test for the coefficient on the endogenous regressors is also reported for each IV estimation.

4.1 IV Results: Household Wealth

In this section, I present the results for the OLS and IV-2SLS analyses, alongside the associated robustness tests. Table 3 presents the results from the OLS and IV estimations using DHS household-level data. Overall, the results indicate that an increase mass grave sites within a 10km buffer around a DHS survey cluster corresponds to a decrease in household wealth, measured by the DHS wealth index based on the physical assets owned. The estimated coefficients indicate that a 1σ (standard deviation) increase in mass graves corresponds to an decrease equivalent to 0.29σ in the households wealth index score. Correspondingly, households with the mean number of mass graves in close proximity (=326 mass graves) are 3.3 percentage points (p.p.) more likely to be in the poorest quintile of the nationally-representative household wealth distribution (and 16.3 p.p. less like to be in richest quintile). The IV first-stage estimation confirms that total number of agricultural reservoirs constructed between 1975-79 is a strong predictor of mass grave sites (F-statistic 49.55 exceeds the minimum values proposed by the weak IV tests proposed by Pflueger and Wang, 2015). Anderson-Rubin test results further confirm that all estimated coefficients for mass graves endogenous variable are statistically significant.

The OLS estimate is also negative but significantly smaller in magnitude and not statistically

[12] 1 decimal degree equals 110km.

Table 3: Main Results: DHS Household Wealth

	Household Wealth Index			Household Wealth Distribution				
	OLS	IV-1^{st} Stage	IV-2^{nd} Stage	5^{th} Quintile	4^{th} Quintile	3^{rd} Quintile	2^{nd} Quintile	1^{st} Quintile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mass Graves	-24.34		-374.51	0.0001	0.0002	0.0001	0.0001	-0.0005
	(20.28)		(113.70)***	(0.0000)**	(0.0000)***	(0.0000)***	(0.0000)*	(0.0001)***
	{11.99}**		{119.15}***	{0.0000}**	{0.0001}**	{0.0000}***	{0.0000}*	{0.0001}***
	[0.24]		[0.00]***	[0.02]**	[0.00]***	[0.00]***	[0.05]**	[0.00]***
Reservoirs (IV)		25.24						
		(3.59)***						
		{6.29}***						
		[0.00]***						
Observations	45,095	45,095	45,095	45,095	45,095	45,095	45,095	45,095
R^2	0.18	0.47	0.14	0.10	0.03	0.02	0.04	0.22
F-statistic	-	49.55	49.55	49.55	49.55	49.55	49.55	49.55
Robust Weak-IV Test	-	37.41	37.41	37.41	37.41	37.41	37.41	37.41
A-R Test (p-value)	-	0.00	0.00	0.01	0.00	0.00	0.06	0.00
Mean LHS	-1153.28	325.55	-1153.28	0.20	0.20	0.18	0.19	0.23
Strata FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations in columns 2-8. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

significant. This is consistent with the prior expectation of the OLS coefficient to be upward biased due to omitted variable bias. Spatial persistence in development patterns should imply that any unobserved and omitted measures of pre-1975 economic development should be positively correlated with long-term development indicators, for example: household wealth. Furthermore, the Khmer Rouge represented a rural insurgency against urban elites who were pressed into forced labour in agricultural areas, therefore proximity to more developed areas and wealthier agricultural zones is expected to also predict the mass grave locations. Consequently, the sign of the omitted variable bias is expected to be positive leading to an upward bias in the OLS coefficient. This is consistent with the observed pattern of OLS and IV-second stage coefficients in Table 3.

5 Research Design: Border Regression Discontinuity Design (RDD)

For my geographic RDD, I focus on the Northeast Zone originally under the command of a moderate Central Committee member, Ney Saran (also known as Comrade Ya). As a Communist Party veteran, Ney Saran prominently disagreed with other Khmer Rouge movement leaders over the rapid pace of land collectivization, evacuation of urban centers, and politically mandated killings (Chandler et al., 1988; Human Rights Watch, 2007). Differences between Ney Saran and the rest of the leadership are prominently highlighted in an internal party document titled, Abbreviated Lesson on the History of the Kampuchean Revolutionary Movement Led by the Communist Party of Kampuchea, where the Khmer Rouge leadership condemns him by stating, "*We say that we made the National-Democratic Revolution because analysis showed that the feudal landlords were in antagonistic contradiction within our country, and outside (it) there were the imperialists. In our Party some friends said that this was not so.[...] Next there was the contemptible Ya (Ney Sarann) and various contemptible people. They resisted our analysis of the contradictions. They said it was necessary to live together in the world, i.e. with other Parties such as the Vietnamese one, and live together with Sihanouk inside the country. Therefore it was not just in analysis of the contradictions that we had to struggle, but also in declaring it.*" (Page 218, Chandler et al., 1988). Due to the influence of Zonal Secretaries over the implementation of policies in regions under their command, the border to Northeast Zone represents a discontinuity in terms of the implementation of Khmer Rouge's extremist policies, specifically relating to mass killings.[13]

In this section, I exploit the discontinuous change in exposure to repressive Khmer Rouge administration by comparing households in the Northeast military zone governed by a moderate zone administrator with households located in neighbouring military zones. The boundary between the Khmer Rouge military zones forms a multi-dimensional discontinuity in longitude-latitude space, however I collapse into a single dimension by calculating distance between observations and the zone boundary. The border is graphically illustrated in Figure 1 as the thick internal border-line located in eastern Cambodia.

The estimation equation for the geographic RDD takes the form:

$$Y_i = \omega_0 + \omega_1 Treatment_i + f(RD\,Polynomial)_i + \omega_2 Boundary\,Segment_i + \epsilon_i \tag{4}$$

where Y_i denotes the outcome variable for unit i, i.e. households, individuals or villages. $Treatment_i$ refers to the a dummy variable equal to 0 if the unit i is located in the Northeast military zone (moderate zone), else 0 (radical zone). $RD\,Polynomial$ refers to the polynomial term of the run-

[13]Similar to other zones, the Northeast Zone was further demarcated into two sub-zones and an autonomous sector between 1976-78. My research design only uses the original outer boundaries of the Zone as it existed in 1975. This outer boundary appears to have remained unchanged despite the creation of the internal sub-divisions (Yale University, 2022)

ning variable, i.e. distance to the Northeast Zone border. All RDD estimations further include $Boundary\,Segment_i$, i.e. 25 kilometre boundary segment fixed effects as recommended by Dell and Querubin, 2018. Robust standard errors are clustered at the district-level for village data. All estimations report p-values from the robust bias-corrected inference procedure and use the common-error robust (CER) optimal bandwidth as recommended by the latest advances in the RDD literature (Calonico et al., 2019a,b; Cattaneo et al., 2019).[14]

Finally, I also estimate the above RDD model using DHS cluster-level data for household wealth and associated mechanisms discussed in the previous section. In this case, I implement a doughnut RDD whereby the DHS clusters overlapping with the border discontinuity are dropped from the estimation, i.e. rural clusters within 10 kilometres and urban clusters within 5 kilometres are excluded. For all DHS-RDD estimations, I similarly cluster standard errors at the level of DHS clusters, report p-values from the robust bias-corrected inference procedures and use the CER optimal bandwidths, as recommended by Calonico et al., 2020.

5.1 RDD Results: Village-Level Poverty

Table 4 reports the border-RDD estimation coefficients for village-level data on poverty rates, alongside a series of checks using geographic, climactic and pre-1975 characteristics. These results indicate that poverty is 11% higher in the radical Khmer Rouge territory, compared to the moderate Northeast Zone (i.e. approximately 40% of the sample mean). There are no statistically significant differences in other characteristics indicating that the identifying assumptions are met.

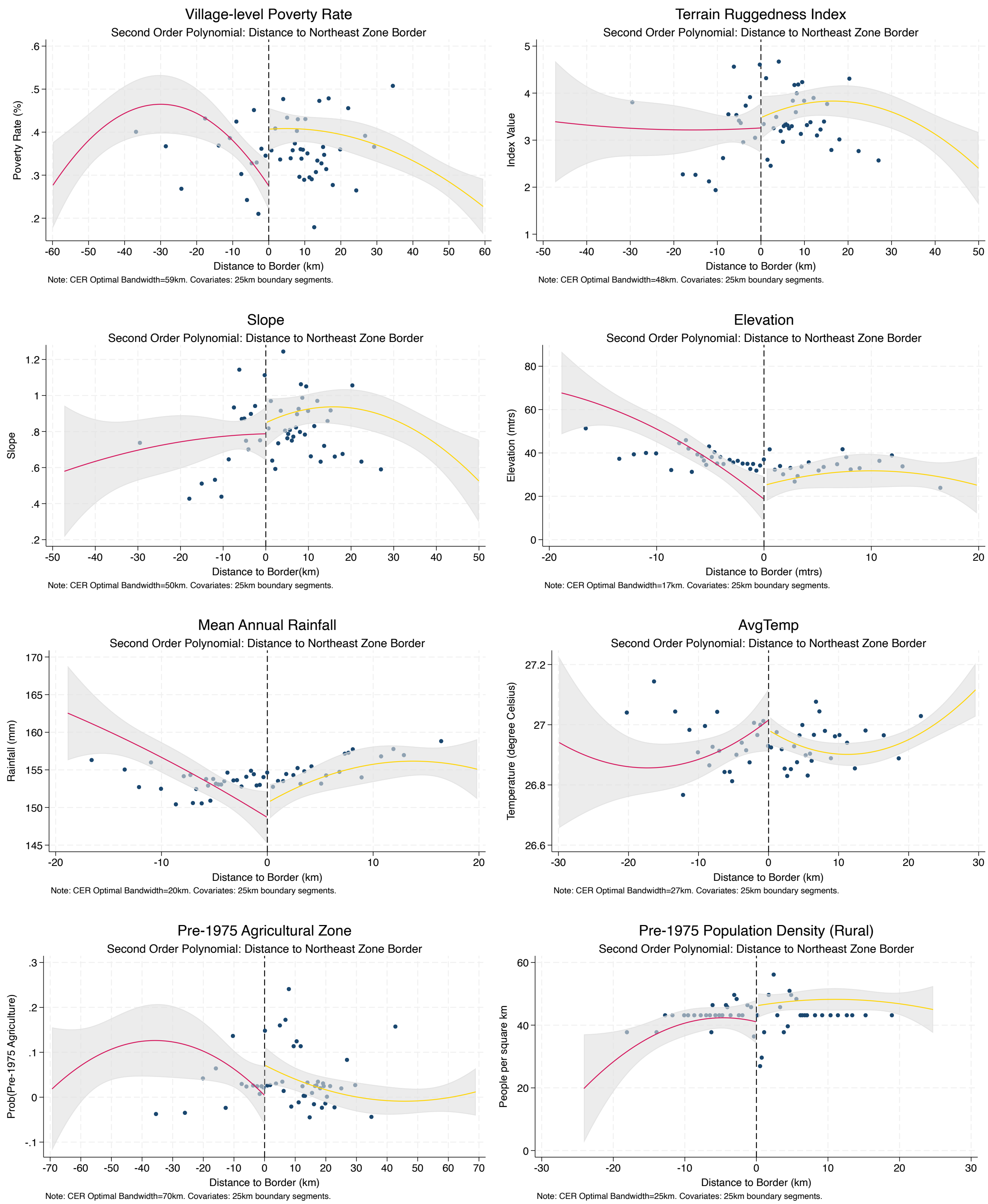
Table 4: Border Regression Discontinuity Results: 1st Polynomial Results

	Poverty Rate (1)	Terrain Ruggedness (2)	Slope (3)	Elevation (4)	Rainfall (5)	Temperature (6)	Pre-1975 Agri Zone (7)	Pre-1975 Pop Density (8)
RD Estimate	0.11	0.32	0.09	-0.10	0.62	0.01	0.05	3.04
	(0.02)***	(0.22)	(0.06)	(1.45)	(0.36)*	(0.03)	(0.04)	(1.80)*
	[0.00]***	[0.11]	[0.08]*	[0.88]	[0.53]*	[0.67]	[0.26]	[0.12]
Observations	10,620	10,620	10,620	10620	10,620	10,620	10620	10620
Bwidth (km)	58.97	48.14	49.80	17.39	20.26	26.80	69.95	25.40
Border Segment FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

[14] I use the rdrobust package provided by Calonico et al. (2017). Mass points in the geospatial data used for this analysis does not permit estimations using higher-order polynomial terms of the running variable.

Figure 4: Border Regression Discontinuity Design: 2nd Polynomial Results



As a graphical representation, Figure 4 presents a geographical representation of the RDD results when using the second polynomial of the running variables, i.e. distance to the Northeast Zone border. These results remain consistent with the estimations using the first polynomial reported in the previous table. There are statistically significant differences in poverty rates at the border, while the other proxy estimations do not find any statistically significant discontinuities.

Finally, I also report the results from a doughnut RDD using DHS clusters in Appendix A.3. Table 20 includes the results for household wealth and confirms the previous pattern of results using village-level data. Households in the radical Khmer Rouge territories are significantly more likely to fall in the poorest quintile and much less likely to be in the richest quintile of the wealth distribution.

6 Transmission Mechanisms

In this section, I discuss the results from IV-2SLS estimations to evaluate the channels of persistence, including human capital, structural transformation and investment flows. Corresponding results from the RDD estimations are also reported in Appendix A3.

6.1 Human Capital

Table 5: Channels of Persistence: Years of Education

	Years of Education (1)	Year of Birth: 1990-99 (2)	Year of Birth: 1980-89 (3)	Year of Birth: 1975-79 (4)	Year of Birth: 1965-74 (5)	Year of Birth: 1955-64 (6)
Mass Graves	-0.0014	-0.0027	-0.0009	-0.0010	-0.0016	-0.0021
	(0.0004)***	(0.0019)	(0.0006)	(0.0007)	(0.0006)***	(0.0005)***
	{0.0005}***	{0.0014}**	{0.0005}*	{0.0005}*	{0.0004}***	{0.0007}***
	[0.00]***	[0.12]	[0.16]	[0.12]	[0.00]***	[0.00]***
Observations	44,849	570	6,396	4,143	11,926	10,537
R^2	0.11	0.17	0.19	0.20	0.15	0.11
F-statistic	49.78	6.66	29.73	32.77	47.17	40.50
Robust Weak IV Test	37.42	37.42	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.06	0.13	0.14	0.00	0.00
Mean Dep Var	4.56	5.81	5.32	5.08	5.57	3.84
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Results in Table 5 indicate that heads of households with mean number of mass graves in close proximity receive 0.5 fewer years of education. This estimate increases to 1.8 fewer years of education on average for households with the median level of mass graves (=1258 mass graves) in close proximity. Splitting the total sample into birth cohorts confirms that this decrease is driven by the cohort born in the two decades preceding the Khmer Rouge period. Individuals born in 1965-74 received 0.5 fewer years of education on average, while individuals born in 1955-64 received 0.7 fewer year of education (median level of exposure predicts 2 and 2.6 fewer years of education for each preceding birth cohort respectively). There are no robust and statistically significant impacts observed in the cohorts born after the fall of the Khmer Rouge regime in 1979.

6.2 Structural Transformation

Next, Table 6 reports the impact of mass grave sites on the occupational sectors of DHS survey participants in order to evaluate comparative patterns of structural transformation in the economy.

Table 6: Channels of Persistence: Structural Transformation

	Agriculture: (Female) (1)	Agriculture: (Male) (2)	Services (Female) (3)	Services (Male) (4)	Manual Labour (Female) (5)	Manual Labour (Male) (6)
Mass Graves	0.0006	0.0004	-0.0004	-0.0002	-0.0000	-0.0000
	(0.0001)***	(0.0001)***	(0.0001)***	(0.0000)***	(0.0000)	(0.0000)
	{0.0002}***	{0.0002}**	{0.0001}***	{0.0001}**	{0.0000}	{0.0000}
	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.37]	[0.26]
Observations	45,889	45,889	45,889	45,889	45,889	45,889
R^2	0.10	0.09	0.06	0.04	0.09	0.03
F-statistic	43.21	43.21	43.21	43.21	43.21	43.21
Robust Weak IV Test	37.42	37.42	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.00	0.00	0.00	0.36	0.23
Mean Dep Var	0.40	0.38	0.28	0.19	0.12	0.17
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation are female DHS survey respondents who provide responses for both male and female members of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

The results indicate that adult males in households with the mean number of mass graves in their 10km proximity are 13 p.p. more likely to be self-employed in agricultural sector (20 p.p. more likely for adult females). Conversely, adult males are 6.5 p.p. less likely to be employed in the services sector. The corresponding estimates for adult females is 13 p.p. for services jobs, while there is no statistically significant difference in skilled or unskilled manual labour jobs. Overall, these results provide evidence consistent with limited structural transformation among households living in areas in close proximity to the genocide who remain stuck in low-value added primary sector, as opposed to higher value added jobs in the services and manufacturing sectors.

6.3 Public Goods Provision and Economic Concessions

Finally, Table 7 reports the impact of mass killings on contemporary household access to public infrastructure and proximity to investment projects.

Table 7: Channels of Persistence: Public Infrastructure

	Economic Land Concession	Piped Water	Bank Account	Telephone
	(1)	(2)	(3)	(4)
Mass Graves	-0.0001	-0.0004	-0.0002	-0.00003
	(0.0000)***	(0.0001)***	(0.0000)***	(0.00002)
	{0.0001}	{0.0001}***	{0.0000}***	{0.00000}***
	[0.00]***	[0.00]***	[0.00]***	[0.10]
Observations	45,095	45,095	31,208	31,208
R^2	0.39	0.17	0.06	0.01
F-statistic	49.55	49.55	40.41	40.42
Robust Weak IV Test	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.00	0.00	0.10
Mean LHS Var	0.95	0.14	0.11	0.09
Strata & Wave FEs	Yes	Yes	Yes	Yes

Notes: Unit of observation are households. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for $p<0.01$, ** for $p<0.05$ and * $p<0.1$.

DHS clusters with mean number of mass grave sites in close proximity have fewer households with piped water (13 p.p.), telephone connections (1 p.p.) and access to financial services (6.5 p.p.). This evidence is consistent with limited flow of public investments by the state in provision of public services in areas in proximity to the genocide.

Economic land concessions for domestic and foreign investments are 3.3 p.p. less likely to be allocated by the federal government to regions with mass grave sites. Therefore private investment flows to these areas is also restricted. My measures of land concessions comprise of spatial polygons, therefore it is not surprising that the estimated coefficients are not statistically significant when using spatial-robust inference methods.

7 Extensions

In this section, I extend the previous analysis by adding interaction terms to the previous discussed IV-2SLS estimation framework to test for heterogeneous impacts of exposure to the Khmer Rouge genocide. The augmented estimation model is now represented by the following equation:

$$Y_{icpw} = \beta_0 + \beta_1 Graves_{cpw} + \beta_2 Graves_{cpw} \times H_{icpw} + \beta_3 H_{icpw} + \beta_2 X_{cpw} + Strata_p + Wave_w + \nu_{icpw} \quad (5)$$

where H_{icpw} denotes the hetergeneous variable of interest (for example: pre-treatment human capital). This variables both directly and in an interaction term with the endogenous the explanatory variable ($Graves_{cpw} \times H_{icpw}$). I follow a similar instrumental variable strategy as before and instrument $Graves_{cpw}$ with $Reservoirs_{cpw}$. Meanwhile, the interaction term $Graves_{cpw} \times H_{icpw}$ is instrumented by $Reservoirs_{cpw} \times H_{icpw}$. All remaining variables are defined as before. Standard errors nu_{icpw} are clustered at the level of DHS survey clusters. Furthermore, I report spatial-robust standard errors using a 3 decimal degree radius with bartlett linear decay in curly brackets and p-values from wild-cluster robust inference tests, same as before.

7.1 Directly Impacted Households

First, I estimate heterogeneous effects of genocide exposure for directly impacted households. These households are identified by the reported year of death for any sibling ever born to the survey respondent, i.e. household head. Directly impacted households are defined as those which reported sibling mortality during the decade 1970-79 spanning the intense civil war that brought the Khmer Rouge to power (1970-75) and their period in government (1975-79).

The results are reported in Table 8 for the same outcome variables as before. The coefficients of the interaction-term indicate that directly impacted households in proximity to mass graves are significantly worse effected. On average, these households have lower levels of wealth, fewer years of education and are significant more likely to be employed in the agricultural sector, and less likely to be employed in the services sector, i.e. demonstrate lower levels of structural transformation.

Table 8: Heterogeneous Effects: Directly Impacted Households

	Household Wealth (1)	Years of Education (2)	Agriculture (Female) (3)	Agriculture (Male) (4)	Service (Female) (5)	Service (Male) (6)	Manual (Female) (7)	Manual (Male) (8)
Mass Graves	-424.69	-0.0009	0.0006	0.0004	-0.0004	-0.0002	-0.0000	-0.0000
	(138.69)***	(0.0004)**	(0.0001)***	(0.0001)***	(0.0001)***	(0.0000)***	(0.0000)	(0.0000)
	{146.72}***	{0.0003}***	{0.0002}***	{0.0002}**	{0.0001}***	{0.0001}**	{0.0000}	{0.0000}
	[0.00]***	[0.02]**	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.59]	[0.40]
Sibling Died (1970-79)	20,693.53	-0.9215	0.0009	0.0105	0.0101	0.0751	-0.0111	0.0336
	(25,872.82)	(0.1005)***	(0.0165)	(0.0161)	(0.0129)	(0.0127)***	(0.0086)	(0.0104)***
	{24,709.14}	{0.1060}***	{0.0189}	{0.0128}	{0.0114}	{0.0154}***	{0.0096}	{0.0108}***
	[0.43]	[0.00]***	[0.96]	[0.51]	[0.45]	[0.00]***	[0.18]	[0.02]**
Mass Graves x Sibling Died (1970-79)	-148.88	0.0001	0.0001	0.0001	-0.0000	-0.0001	-0.0001	-0.0001
	(65.06)**	(0.0002)	(0.0000)***	(0.0000)***	(0.0000)	(0.0000)***	(0.0000)***	(0.0000)**
	{71.68}**	{0.0002}	{0.0001}**	{0.0000}***	{0.0000}	{0.0000}**	{0.0000}*	{0.0000}***
	[0.02]**	[0.65]	[0.01]***	[0.00]***	[0.79]	[0.00]***	[0.01]***	[0.00]***
Observations	45,889	45,889	45,889	45,889	45,889	45,889	45,889	45,889
R^2	0.16	0.19	0.10	0.09	0.06	0.04	0.09	0.04
F-statistic	21.58	21.58	21.58	21.58	21.58	21.58	21.58	21.58
A-R Test (p-value)	0.00	0.08	0.00	0.00	0.00	0.00	0.01	0.04
Mean Dep Var	49843.04	4.79	0.40	0.38	0.28	0.20	0.12	0.18
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation are heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistic and Anderon-Rubin test (p-value) are included in all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

7.2 Highly Educated Survivors

Next, I estimate heterogeneous effects of genocide exposure for households heads interacted with their years of education acquired before Khmer Rouge period, i.e. pre-1975. This allows me to test whether survivors with higher education levels became more wealthy due to increased returns for human capital after the massive loss of lives during the genocide. I split the sample of survivors into three sub-samples: born pre-1975, born pre-1965 (i.e. atleast 10 years old with some education before the genocide), and born pre-1955 (i.e. atleast 20 years old and completed basic education before the genocide). The results are reported in Table 9. In column (1), there is some evidence that highly educated survivors born before 1975 could be even worse-off as indicated by the negative coefficient for the interaction term. However, these results are not robust since levels of statistical significance vary when using difference inference methods. The coefficients lose statistical significant altogether when I restrict my sample to cohorts born before 1965 and 1955. Hence I conclude that there are no heterogeneous effects for exposure to the genocide by levels of pre-1975 education.

Table 9: Heterogeneous Effects: Highly Educated Households

Year of Birth	Pre-1975 (1)	Pre-1965 (2)	Pre-1955 (3)
Mass Graves	-300.01	-364.22	-383.47
	(124.51)**	(138.00)***	(149.21)**
	{106.60}***	{120.86}***	{137.37}***
	[0.01]***	[0.00]***	[0.01]***
Years of Education	37,048.49	41,178.33	35,391.47
	(3,918.43)***	(5,247.43)***	(5,987.25)***
	{3,185.27}***	{4,591.57}***	{5,456.36}***
	[0.00]***	[0.00]***	[0.00]***
Mass Graves x	-17.98	-14.87	-2.39
Years of Education	(9.20)*	(11.94)	(12.65)
	{6.03}***	{8.15}*	{9.39}
	[0.04]**	[0.20]	[0.87]
Observations	33,772	21,846	11,309
R^2	0.19	0.20	0.22
F-statistic	24.91	23.17	23.00
A-R Test (p-value)	0.00	0.00	0.01
Mean Dep Var	15,413.404	29,309.56	20,160.23
Strata & Wave FEs	Yes	Yes	Yes

Notes: Unit of observation are heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistic is included in all columns. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

7.3 Alternative Genocide Measures: Political Repression

Next, I test an alternative measure of genocide: proximity to Khmer Rouge prisons. This is measured as binary indicator equal to 1 if a prison is located within a 10 km buffer area of each DHS cluster, 0 otherwise. As discussed previously, the regime separated socio-economic elites, members of the previous political regime and demographic minorities from the labour camps into separate prison camps constructed for their interrogation and eventual executions. Therefore, I aim to use this alternative measure to attempt to distinguish between the effects due to political repression of opposition groups from the overall mortality during this period measured using the previously used proximity to mass graves. Finally, I propose to use the same IV strategy as before since the location of prisons since it simply represents an alternative measure of the genocide and their location choices follow the same logic as the previous explanatory variable, i.e. mass graves.

The results using DHS data on household wealth are reported in Table 10 below, while the results for the other dependent variables used to measure transmission mechanisms are reported in Appendix A1. The IV-first stage results confirm that reservoirs are a strong and robust predictor of proximity to Khmer Rouge prisons (Kleibergen Papp F-statistic = 51.51). The IV-second stage results confirm the same pattern of results as before. Households in close proximity to a prison are significantly less wealthy (1.47σ reduction in relative wealth), more likely to fall in the poorest quintiles of the wealth distribution. The results in Tables 14, 15 and 16 in Appendix A1 also confirm that these households have significantly lower human capital and demonstrate lower structural transformation. Public investment in services and private investments are also significantly less likely to flow to areas in close proximity to the erstwhile Khmer Rouge prisons.

7.4 Alternative Mechanism: Internal Migration in Cambodia

Finally, I use district-level survey data on internal migration from the Cambodian Consumption and Expenditure Surveys to analyse whether rates of internal migration vary by proximity to the genocide locations. Households located close to the killing fields and prison systems of Cambodia may choose to migrate to other parts of the country. This migration decision could vary by levels of household wealth, education and sector of employment. Therefore, international migration may be an alternative mechanism that explains the previous results.

In a modification of the variables used in the IV-2SLS model, I now utilize linear distance from the

Table 10: Alternative Genocide Measure: Khmer Rouge Prisons and Household Wealth

	Household Wealth Index			Household Wealth Distribution				
	OLS	IV-1st Stage	IV-2nd Stage	5th Quintile	4th Quintile	3rd Quintile	2nd Quintile	1st Quintile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prisons (Dummy)	-53,149.78**		-814,811.98***	0.23**	0.40***	0.29***	0.15*	-1.06***
	(25,516.81)**		(243,486.12)***	(0.10)**	(0.09)***	(0.07)***	(0.08)*	(0.18)***
	{33,786.78}		{233,894.91}***	{0.07}***	{0.10}***	{0.15}*	{0.10}	{0.35}***
	[0.04]**		[0.00]***	[0.02]**	[0.00]***	[0.00]***	[0.05]**	[0.00]***
Reservoirs (IV)		0.01						
		(0.00)***						
		{0.00}***						
		[0.00]***						
Observations	45,095	45,095	45,095	45,095	45,095	45,095	45,095	45,095
R^2	0.18	0.60	0.04	0.08	-0.03	-0.00	0.03	-0.12
F-statistic	-	51.51	51.51	51.51	51.51	51.51	51.51	51.51
Weak IV Test	-	37.42	37.42	37.42	37.42	37.42	37.42	37.42
A-R Test (p-value)	-	0.00	0.00	0.02	0.00	0.00	0.06	0.00
Mean Dep Var	-1153.28	0.75	-1555.24	0.20	0.20	0.18	0.19	0.23
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations in columns 2-8. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 11: Dependent Variable: Migration Outcomes

	Full Sample			Age 18+ in 1980		
	IV-1st Stage	No Migration	Last Migration 1979-80	IV-1st Stage	No Migration	Last Migration 1979-80
	(1)	(2)	(3)	(4)	(5)	(6)
Burial (km)		-0.02	0.01		-0.02	0.01
		(0.01)*	(0.07)		(0.14)	(0.04)
		{0.01}	{0.00}		{0.02}	{0.01}
		[0.11]	[0.35]		[0.13]	[0.37]
Reservoir (km)	0.15			0.16		
	(0.06)**			(0.07)**		
	{0.08}*			{0.09}*		
	[0.00]***			[0.00]***		
Observations	40,839	40,839	40,839	18,916	18,916	18,916
R^2	0.77	0.11	0.07	0.78	0.12	0.09
F Statistic	1,500.36	1,500.36	1,500.36	699.48	699.48	699.48
Mean Dep Var	0.43					
Province FEs	Yes	Yes	Yes	Yes	Yes	Yes
Survey Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network and length of rivers and pre-1975 waterways. All estimations also include province fixed effects and survey wave fixed effects. Bootstrapped standard errors clustered at district-level in parentheses. Spatial HAC robust standard errors in curly brackets. P-values from wild bootstrap test with 999 repetitions in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

district centroids to the mass graves as the endogenous explanatory variable which is instrumented using distance to the proximate reservoir. The analysis is conducted separately for the full sample of survey participants and a sub-sample of individuals who reached adulthood in 1980. This allows me to test whether there are any age-specific differences in internal migration decisions for households driven by their proximity to the genocide locations. Finally, I report results for two different outcome variables: $No\ Migration$ is a dummy variable equal to 1 if the respondents have never migrated previously, else 0. $Last\ Migration$ $(1979-80)$ is a dummy variable equal to 1 if they last reported migration occurred directly after the collapse of the Khmer Rouge regime. To be precise, I aim to test whether either measure of migration is driven by proximity to the genocide locations. The results are reported in Table 11. The IV-first stage coefficients indicate that distance to reservoirs is a strong and robust predictor of distance to mass graves (Kleibergen-Papp F-statistic = 1500). The IV second stage coefficients for both migration measures are not statistically significant for either sample. Therefore, I can rule out internal migration as an alternative mechanism that could explain the previously observed results.

8 Robustness Tests

8.1 Zero first-stage (ZFS) Test

In this section, I propose two tests for the validity of IV strategy. First, I use an placebo test to verify whether the IV (reservoirs constructed during 1975-79) meets the relevance assumption only under the expected conditions. Accordingly, I will use an auxiliary regression on a sub-sample where the underlying theoretical motivation indicates that the IV is not expected to influence treatment assignment, known as zero-first-stage (ZFS) test (Bound and Jaeger, 2000; Rohner et al., 2013a; Nunn, 2008; Nunn and Wantchekon, 2011). The underlying intuition is that since this ZFS sub-sample consists of *never takers*, the reduced form relationship between the IV and outcome should also be zero if the exogeneity assumption is satisfied.[15]

In the context of this research, the Northeast military zone of Cambodia under the Khmer Rouge was governed by the moderate zone secretary opposed to the extremist policies and persecution of the Central Committee. As a result, we expect our theoretical motivation for the IV to be

[15] As per Lal et al., 2021, ZFS placebo tests are particularly relevant in IV research designs used in historical political economy research, whereby specific geographic variables are proposed as valid IVs based on detailed domain knowledge. Therefore, these IVs can be argued as unlikely to be driving treatment assignment outside of a specific theoretical context.

inapplicable for units located in the Northeast zone and designate them as our ZFS sub-sample.[16] The following equation is then estimated for DHS clusters located in the ZFS Northeast zone and for comparison, also for the rest of Cambodia:

$$Y_{icpw} = \theta_0 + \theta_1 Reservoirs_{cp} + \theta_2 X_{cpw} + Strata_p + Wave_w + \xi_{icpw} \tag{6}$$

where all variables are defined as previously.

Table 12 reports the results from robustness tests for the validity of the proposed IV research design. In column 1, I report the coefficient from the regression of household wealth index score on the IV for the Northeast zone, i.e. the ZFS sub-sample. Consistent with the theoretical motivation, the estimated coefficient is not statistically significant for the Northeast zone. In column (2) I report the same regression for the sub-sample of DHS clusters located in the rest of Cambodia and find a statistically significant correlation. This correlation is expected since the IV impacts the outcome variable through its influence on the endogeneous variable, mass grave sites, in this sub-sample.

8.2 Plausibly Exogenous IV Estimations

For my second test, I apply a more formal plausibly exogenous methodology which tests for the validity of the IV results after relaxing the exogeneity assumption (Conley et al., 2012). Using this approach, we allow the IV $Reservoirs_{cp}$ to affect outcome variables Y_{icp} directly and the coefficient of $Graves_{cp}$ reveals how the 2SLS estimation is influenced if IV coefficient takes on different values.

Formally, we include the linear IV in the second stage and estimate the following equation:

$$Y_{icwp} = \zeta_0 + \zeta_1 Graves_{cp} + \zeta_2 Reservoirs_{cp} + \zeta_3 X_{cp} + Strata_p + Wave_w + \sigma_{icwp} \tag{7}$$

where all variables are defined as previously. In the conventional IV-2SLS methodology, I assume the coefficient ζ_2 on the IV $Reservoirs_{cp}$ to be zero based on theoretical arguments. In this approach, I relax this assumption under certain conditions by allowing the coefficient ζ_2 to vary and affect the outcomes Y_{icwp} directly, thus allowing the ζ_1 to reveal how the 2SLS estimation is influenced when the IV is only plausibly exogenous.

[16] As mentioned previously, the Northeast Zone was further sub-divided into sub-units including an autonomous zone before the collapse of the Khmer Rouge regime. In order to distinguish this section from the spatial border RDD, I use only those territories that remained under the Northeast zone throughout from 1975-79.

Table 12: Zero First-Stage and Plausibly Exogeneous IV Tests

	Northeast Zone (ZFS) (1)	Rest of Cambodia (2)	All Zones (3)
Reservoirs	302.14	-5,835.19	
	(92,325.74)	(2,020.10)	
	{0.99}	{0.00}***	
	[0.99]	[0.01]***	
Graves			-320.43
			(51.87)
			{0.00}***
			[0.00]***
Observations	2,788	41,371	45,095
R^2	0.32	0.16	
Strata & Wave FEs	Yes	Yes	Yes
Military Zone FEs	Yes	Yes	Yes

Notes: Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network and length of rivers and pre-1975 waterways. All estimations also include DHS strata fixed effects, DHS wave fixed effects and Khmer Rouge military zone fixed effects. Robust standard errors clustered at DHS survey cluster-level in parentheses and p-values in curly brackets{}. P-value from a wild bootstrap test in square brackets [] in columns (1) and (2). P-value from a cluster-bootstrap inference reported in square brackets [] in column (3). Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Column (3) in Table 12 reports the results from a plausibly exogeneous IV regression for the complete sample. The estimated coefficient is negative and statistically significant similar to our baseline results. Overall, these findings indicate that our proposed IV strategy is valid and robust.

8.3 Lasso-IV Estimations

Finally, I use an alternative causal machine learning procedure to re-estimate the IV-2SLS model presented in Equations 6 and 7. This estimation uses a Lasso shrinkage procedure implemented using the cross-fit, partialling out algorithm to select and include only the relevant covariates in the final estimations (Chernozhukov et al., 2018). The lasso procedure is most relevant in settings with a large number of covariates, where an unknown subset is most likely to be sufficient. Exclusion of noisy covariates can provide us with more precise estimation of the IV-first stage and serve as a useful robustness test for the baseline IV results (Cameron and Trivedi, 2022).

The Lasso-IV results for household wealth are presented in Table 13, while the results for the remaining dependent variables representing transmission mechanisms are reported in Appendix A2. Overall, these results are consistent with the baseline IV-2SLS results and magnitudes remain similar.

Table 13: Lasso-IV Estimation: DHS Household Wealth Index

	Household Wealth	Household Wealth Distribution				
	Household Wealth (1)	5^{th} Quintile (2)	4^{th} Quintile (3)	3^{rd} Quintile (4)	2^{nd} Quintile (5)	1^{st} Quintile (6)
Mass Graves	-367.0833***	0.0001***	0.0002***	0.0001***	0.0001**	-0.0005***
	(41.3040)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)**	(0.0000)***
Observations	45,095	45,095	45,095	45,095	45,095	45,095
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata and DHS wave fixed effects. Robust standard errors clustered at DHS cluster-level in parentheses. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

9 Conclusion

Multiple episodes of extreme political violence were implemented by radical regimes across the 20th century. In many cases, their objective was to build new national identity after destroying existing institutions and eliminating all political opposition. Theoretical literature on nation-building and strategic mass killings models political violence being driven by competition over political control of the state and its economic resources, often exacerbated by threat of external conflict, adverse economic shocks and natural disasters (Alesina et al., 2020; Besley and Persson, 2011; Esteban et al., 2015). The resulting mass killings represent a significant shock to the human capital stock, however, there is a paucity of existing analyses on the economic legacies of mass killings. Our common understanding on the magnitude and transmission mechanisms through which political violence impacts long-term economic development remains limited.

This paper aims to analyse the long-term economic impacts of mass killings under the Khmer Rouge regime (1975-79) on contemporary economic development indicators. I merge multiple historical and contemporary data sources, including geo-located information on the locations of mass graves. Two novel identification strategies are proposed to identify the causal impact of the genocide on individual, household and village-level economic indicators. The Khmer Rouge's failed agricultural irrigation projects are exploited as a source of exogenous variation in mass killings. Additionally, I exploit spatial power relations within the Khmer Rouge bureaucracy in the form of an internal administrative border within a spatial regression discontinuity design.

My analysis finds persistent impacts of historical genocide violence on contemporary household wealth and village-level poverty rates, driven by lower human capital, structural transformation and availability of public goods. Zero-first-stage placebo tests, plausibly exogenous instrument methods, and lasso-IV machine learning algorithms confirm the validity of the proposed IV research design. Robustness of the RDD design is established using a battery of proxy estimations that find no significant differences in observable demographic, geographic, climactic and pre-treatment characteristics along the border discontinuity.

In supplementary analyses, I show that directly impacted households identified by using the year of death of family members are significantly worse off. Furthermore, households exposed to political repression measured by proximity to Khmer Rouge prisons are also negatively impacted. Individuals with higher pre-genocide levels of human capital who survived the genocide do not appear to be wealthier in the long-run, indicating they are equally impacted by the genocide. Finally, there are no statistically significant differences in internal migration rates driven by exposure to the genocide which allows me to rule out an alternative mechanism driving the observed results.

This research aims to make multiple contributions to the economic literature on political conflict and economic development. First, I contribute to the growing literature on political violence and nation-building (Besley and Persson, 2011; Esteban et al., 2015; Alesina et al., 2020). My research is also closely linked to the related empirical literature on forced migration in response to political violence reviewed by Becker et al., 2022. Second, I provide novel, country-level evidence to highlight the magnitude and channels through which large-scale, one-sided political violence against civilian and political opponents impacts comparative economic development. There is a limited existing literature focusing on this question which finds conflicting results (Yanagizawa-Drott and Rogall, 2014; Acemoglu et al., 2011; Bai and Wu, 2023) Therefore, the developmental impacts of large-scale genocide on affected states remains an open question in this literature. Next, there is a growing focus on the Cambodian genocide in recent social science literature (de Walque, 2004; Gangadharan et al., 2022; Bühler and Madestam, 2023; Grasse, 2023). My paper proposes distinct identification strategies informed by historical research to provide country-level insights consistent with the overall findings reported in this literature. Finally, there is a distinct literature on the impacts of violent inter-state and civil conflict. There is robust evidence of devastating short-term impacts of violent conflict between states, however, analyses of persistent magnitude and channels of impact on long-term economic development remains limited and often contradictory. This paper contributes new insights on the mechanisms through which large-scale human capital destruction caused by one-sided political violence can result in long-term poverty.

References

Acemoglu, D., Hassan, T. A., and Robinson, J. A. (2011). Social Structure and Development: A Legacy of the Holocaust in Russia. *The Quarterly Journal of Economics*, 126(2):895–946.

ADB (2019). Kingdom of Cambodia: Preparing the Irrigated Agriculture Improvement Project. *Technical Assistance Consultant Report*, Project Number: 51159001 (TA 9349 CAM).

Akbulut-Yuksel, M. (2014). Children of war the long-run effects of large-scale physical destruction and warfare on children. *Journal of Human Resources*, 49(3):634–662.

Akresh, D. and de Walque, R. (2008). *Armed Conflict And Schooling: Evidence From The 1994 Rwandan Genocide*. Policy Research Working Papers. The World Bank.

Akresh, R., Lucchetti, L., and Thirumurthy, H. (2012). Wars and Child Health: Evidence from the Eritrean-Ethiopian Conflict. *Journal of Development Economics*, 99(2):330–340.

Akresh, R., Verwimp, P., and Bundervoet, T. (2011). Civil war, crop failure, and child stunting in rwanda. 59(4):777–810.

Alesina, A., Reich, B., and Riboni, A. (2020). Nation-building, nationalism, and wars. *Journal of Economic Growth*, 25(4):381–430.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, 5(1):180040. Number: 1 Publisher: Nature Publishing Group.

Bai, L. and Wu, L. (2023). Political conflict and development dynamics: Economic legacies of the cultural revolution. *Journal of Economic History*.

Becker, S. O., Mukand, S., and Yotzov, I. (2022). Persecution, pogroms and genocide: A conceptual framework and new evidence. *Explorations in Economic History*, 86:101471.

Besley, T. and Persson, T. (2011). The logic of political violence. *The Quarterly Journal of Economics*, 126(3):1411–1445.

Bound, J. and Jaeger, D. A. (2000). Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? In *Research in Labor Economics*, volume 19 of *Research in Labor Economics*, pages 83–108. Emerald Group Publishing Limited.

Brakman, S., Garretsen, H., and Schramm, M. (2004). The Strategic Bombing of German cities during World War II and its Impact on City Growth. *Journal of Economic Geography*, 4(2):201–218.

Bühler, M. and Madestam, A. (2023). State Repression, Exit, and Voice. Rationality and Competition Discussion Paper Series 408, CRC TRR 190 Rationality and Competition.

Bultmann, D. (2012). Irrigating a socialist utopia: Disciplinary space and population control under the khmer rouge, 1975-1979. *Transcience*, 3(1):40–52.

Bundervoet, T., Verwimp, P., and Akresh, R. (2009). Health and civil war in rural burundi. *Journal of Human Resources*, 44(2):536–563.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2019a). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019b). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, 101(3):442–451.

Cameron, A. C. and Trivedi, P. K. (2022). *Microeconometrics Using Stata: Volume II: Nonlinear Models and Causal Inference Methods*. Stata Press.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*. ISBN: 9781108684606 9781108710206 Publisher: Cambridge University Press.

Central Intelligence Agency (1970). *Indochina Atlas*. Directorate of Intelligence, Office of Basic and Geographic Intelligence, Central Intelligence Agency.

Chandler, D. P. (1999). *Brother number one: A political biography of Pol Pot*. Routledge.

Chandler, D. P., Kiernan, B., and Boua, C. (1988). *Pol Pot plans the future: confidential leadership documents from Democratic Kampuchea, 1976-1977*. Yale Center for International and Area Studies, New Haven, Conn. OCLC: 19789872.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015). Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments. *American Economic Review*, 105(5):486–490.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly Exogeneous. *The Review of Economics and Statistics*, 94(1):260–272. Publisher: The MIT Press.

Davis, D. R. and Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *The American Economic Review*, 92(5):1269–1289.

DC-CAM (2022). Documentation Center - Cambodia.

de Walque, D. (2004). The long-term legacy of the Khmer Rouge period in Cambodia. Technical Report 3446, The World Bank. Publication Title: Policy Research Working Paper Series.

de Walque, D. (2005). Selective Mortality During the Khmer Rouge Period in Cambodia. *Population and Development Review*, 31(2):351–368.

Dell, M. and Querubin, P. (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, 133(2):701–764.

Esteban, J., Morelli, M., and Rohner, D. (2015). Strategic Mass Killings. *Journal of Political Economy*, 123(5):1087–1132. Publisher: The University of Chicago Press.

FAO (2011). AQUASTAT Survey 2011 Irrigation in Southern and Eastern Asia. *FAO Land and Water Division*.

Fick, S. E. and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

Gangadharan, L., Islam, A., Ouch, C., and Wang, L. C. (2022). The long-term effects of genocide on antisocial preferences. *World Development*, 160:106068.

Grasse, D. (2023). State terror and long-run development: The persistence of the khmer rouge. *American Political Science Review*, pages 1–18.

Heuveline, P. (2015). The boundaries of genocide: Quantifying the uncertainty of the death toll during the Pol Pot regime in Cambodia (1975 - 79). *Population Studies*. Publisher: Routledge.

Himel, J. (2007). Khmer rouge irrigation development in cambodia. *Searching for the Truth*, pages 42–49.

Human Rights Watch (2007). Anatomy of an Interrogation: The Torture of Comrade Ya at S-21. *Human Rights Watch*.

Ichino, A. and Winter-Ebmer, R. (2004). The Long-Run Educational Cost of World War II. *Journal of Labor Economics*, 22(1):57–87.

Jackson, K. D., editor (1989a). *Cambodia, 1975-1978: Rendezvous with Death*. Princeton University Press.

Jackson, K. D. (1989b). *Cambodia, 1975-1978: Rendezvous with death*. Princeton University Press.

Justino, P. (2016). The microeconomic causes and consequences of genocides and mass atrocities. *Economic Aspects of Genocides, Other Mass Atrocities, and Their Preventions*, pages 211–29.

Keane, M. P. and Neal, T. (2022). A practical guide to weak instruments.

Kelly, M. (2019). The Standard Errors of Persistence. Working Papers 201913, School of Economics, University College Dublin.

Kiernan, B. (2002). *The Pol Pot regime: Race, power, and genocide in Cambodia under the Khmer Rouge, 1975-79*. Yale University Press.

Kiernan, B. (2003). The Demography of Genocide in Southeast Asia: The Death Tolls in Cambodia, 1975-79, and East Timor, 1975-80. *Critical Asian Studies*, 35(4):585–597. Publisher: Routledge.

Lal, A., Lockhart, M. W., Xu, Y., and Zu, Z. (2021). How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies.

Merrouche, O. (2011). The long term educational cost of war: Evidence from landmine contamination in cambodia. *The Journal of Development Studies*, 47(3):399–416.

Mertha, A. (2014). *The Khmer Rouge Bureaucracy*. Oxford University Press.

Miguel, E. and Roland, G. (2011). The long-run impact of bombing vietnam. *Journal of Development Economics*, 96(1):1–15.

Nunn, N. (2008). The Long-term Effects of Africas Slave Trades. *The Quarterly Journal of Economics*, 123(1):139–176.

Nunn, N. and Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, 101(7):3221–3252.

Pflueger, C. and Wang, S. (2015). A robust test for weak instruments in stata. *The Stata Journal*, 15(1):216–225.

Riaño, J. F. and Valencia Caicedo, F. (2020). Collateral damage: The legacy of the secret war in laos.

Rohner, D., Thoenig, M., and Zilibotti, F. (2013a). Seeds of distrust: conflict in Uganda. *Journal of Economic Growth*, 18(3):217–252. Publisher: Springer.

Rohner, D., Thoenig, M., and Zilibotti, F. (2013b). Seeds of distrust: conflict in Uganda. *Journal of Economic Growth*, 18(3):217–252.

Shemyakina, O. (2011). The effect of armed conflict on accumulation of schooling: Results from tajikistan. *Journal of Development Economics*, 95(2):186–200.

Twining, C. H. (1989). The economy. In Jackson, K. D., editor, *Cambodia, 1975-1978: Rendezvous with death*, chapter 5, pages 109–150. Princeton University Press, New Haven.

Tyner, J. A. (2017). *From Rice Fields to Killing Fields: Nature, Life, and Labor under the Khmer Rouge*. Syracuse University Press.

Tyner, J. A., Munro-Stasiuk, M., Coakley, C., Kimsroy, S., and Rice, S. (2018). Khmer rouge irrigation schemes during the cambodian genocide. *Genocide Studies International*, 12(1):103–119.

Tyner, J. A. and Will, R. (2015). Nature and post-conflict violence: Water Management under the Communist Party of Kampuchea, 1975-1979. *Transactions of the Institute of British Geographers*, 40(3):362–374.

Vickery, M. (1986). *Kampuchea: politics, economics, and society*. Marxist regimes series. F. Pinter ; L. Rienner Publishers, London : Boulder.

Walder, A. G. (2014). Rebellion and repression in china, 1966–1971. *Social Science History*, 38(3-4):513–539.

Yale University (2022). Cambodian Genocide Program.

Yanagizawa-Drott, D. and Rogall, T. (2014). The legacy of political mass killings: Evidence from the rwandan genocide. *mimeo*.

Young, A. (2022). Consistency without Inference: Instrumental Variables in Practical Application. *European Economic Review*, 147:104112.

Appendix

A1. Alternative Genocide Measures: Khmer Rouge Prisons

This section reports the results from estimating the IV-2SLS model indicated in Equations 6 and 7 using proximity to Khmer Rouge prisons as an alternative measure of political repression under the Khmer Rouge. The estimates reported in Tables 14, 15 and 16 indicate that households exposed to the political repression component of the genocide are significantly worse affected, in terms of fewer years of education, more limited levels of structural transformation and reduced access to improved public infrastructure. These results indicate that the baseline results are robust to using an alternative measure of political repression from the Khmer Rouge period (1975-79).

Table 14: Alternative Genocide Measure: Khmer Rouge Prisons and Years of Education

	Years of Education (1)	Year of Birth: 1990-99 (2)	Year of Birth: 1980-89 (3)	Year of Birth: 1975-79 (4)	Year of Birth: 1965-69 (5)	Year of Birth: 1955-64 (6)
Prisons (Dummy, 10km buffer)	-3.0262	-7.7917	-2.1922	-2.3134	-3.1446	-4.3203
	(0.8809)***	(5.4895)	(1.5490)	(1.6455)	(1.1987)***	(1.1112)***
	{0.9751}***	{3.9371}**	{1.5115}	{1.0325}**	{1.1317}***	{1.2846}***
	[0.00]***	[0.08]*	[0.17]	[0.14]	[0.01]***	[0.00]***
Observations	44,849	570	6,396	4,143	11,926	10,537
R^2	0.08	-0.10	0.16	0.17	0.11	0.04
F-statistic	51.69	4.21	22.95	33.43	59.32	46.28
Weak IV Test	37.42	37.42	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.06	0.13	0.14	0.00	0.00
Mean Dep Var	4.56	5.81	5.31	5.07	5.56	3.83
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 15: Alternative Genocide Measure: Khmer Rouge Prisons and Structural Transformation

	Agriculture: (Female) (1)	Agriculture: (Male) (2)	Services (Female) (3)	Services (Male) (4)	Manual Labour (Female) (5)	Manual Labour (Male) (6)
Prisons (Dummy, 10km buffer)	1.2093	0.9080	-0.7608	-0.3968	-0.0659	-0.0894
	(0.2216)***	(0.1801)***	(0.1523)***	(0.0943)***	(0.0731)	(0.0758)
	{0.3206}***	{0.2068}***	{0.2246}***	{0.0942}***	{0.0535}	{0.0571}
	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.37]	[0.25]
Observations	45,889	45,889	45,889	45,889	45,889	45,889
R^2	-0.22	-0.09	-0.10	-0.01	0.09	0.03
F-statistic	47.83	47.83	47.83	47.83	47.83	47.83
Weak IV Test	37.42	37.42	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.00	0.00	0.00	0.37	0.23
Mean Dep Var	0.40	0.38	0.28	0.20	0.12	0.18
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation are female DHS survey respondents who provide responses for both male and female members of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 16: Alternative Genocide Measure: Khmer Rouge Prisons and Investments

	Economic Land Concession (1)	Piped Water (2)	Bank Account (3)	Telephone (4)
Prisons	-0.3130***	-0.7838***	-0.3901***	-0.0720
	(0.0885)***	(0.1440)***	(0.0826)***	(0.0450)
	{0.2600}	{0.3497}**	{0.1515}**	{0.0361}**
	[0.00]***	[0.00]***	[0.00]***	[0.10]
Observations	45,095	45,095	31,208	31,208
R^2	0.44	-0.09	0.03	0.03
F-statistic	51.51	51.51	40.84	40.84
Weak IV Test	37.42	37.42	37.42	37.42
A-R Test (p-value)	0.00	0.00	0.00	0.10
Mean LHS Var	0.95	0.14	0.11	0.09
Strata & Wave FEs	Yes	Yes	Yes	Yes

Notes: Unit of observation are households. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Kleibergen-Paap F-statistics, Montiel-Pflueger robust weak instrument test statistics and Anderson-Rubin test (p-values) are reported for all IV estimations. Robust standard errors clustered at DHS survey cluster-level in parentheses. Spatial-robust standard errors estimated for 3 decimal degrees with bartlett linear decay are reported in curly brackets. P-value from a wild bootstrap test in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

A2. Robustness: Lasso-IV Results

Below I report the results from implementing the Lasso-IV methodology discussed in Section 8.3 using DHS data. The estimates reported in Tables 17, 18 and 19 finds that households exposed to the genocide have significantly fewer years of education, demonstrate more limited levels of structural transformation and have reduced access to improved public infrastructure. These results indicate that the baseline results are robust to using an alternative Lasso shrinkage estimator to select relevant covariates from the initial vector of pre-1975, geographic and climactic covariates.

Table 17: Lasso-IV Estimation: Years of Education

	Years of Education (1)	Year of Birth: 1980-89 (2)	Year of Birth: 1975-79 (3)	Year of Birth: 1965-74 (4)	Year of Birth: 1955-64 (5)
Mass Graves	-0.0013	-0.0005	-0.0005	-0.0018	-0.0019
	(0.0002)***	(0.0005)	(0.0007)	(0.0004)***	(0.0004)***
Observations	44,849	6,396	4,143	11,926	10,537
Mean Dep Var	4.56	5.32	5.08	5.57	3.84
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation is individual heads of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Robust standard errors clustered at DHS survey cluster-level in parentheses. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 18: Lasso-IV Estimation: Structural Transformation

	Agriculture: (Female) (1)	Agriculture: (Male) (2)	Services (Female) (3)	Services (Male) (4)	Manual Labour (Female) (5)	Manual Labour (Male) (6)
Mass Graves	0.0006	0.0004	-0.0004	-0.0002	-0.0000	-0.0000
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)	(0.0000)*
Observations	45,889	45,889	45,889	45,889	45,889	45,889
Mean Dep Var	0.40	0.38	0.28	0.19	0.12	0.17
Strata & Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Unit of observation are female DHS survey respondents who provide responses for both male and female members of household. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Robust standard errors clustered at DHS survey cluster-level in parentheses. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 19: Lasso-IV Estimation: Public and Private Investments

	Economic Land Concession (1)	Piped Water (2)	Bank Account (3)	Telephone (4)
Mass Graves	-0.0004	-0.0003	-0.0002	-0.0000
	(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)**
Observations	45,095	45,095	31,208	31,208
Mean Dep Var	0.95	0.14	0.11	0.09
Strata & Wave FEs	Yes	Yes	Yes	Yes

Notes: Unit of observation are households. Covariates include terrain ruggedness, slope, elevation, temperature, rainfall, distance to Province capital, distance to Phnom Penh (national capital), pre-1975 agriculture indicators, pre-1975 industry indicators, length of pre-1975 road network, and length of rivers or waterways. All estimations include DHS strata fixed effects and DHS wave fixed effects. Robust standard errors clustered at DHS survey cluster-level in parentheses. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

A3. Doughnut RDD Results: DHS Clusters

Below I report the results from estimating the doughnut RDD indicated in Equation 4 using DHS cluster-level data. The estimates reported in Tables 20, 21, 22 and 23 finds that households close to the border discontinuity in the radical Khmer Rouge territories are less wealthy, have fewer years of education, demonstrate more limited levels of structural transformation and have reduced access to improved public infrastructure when compared to clusters in the moderate Northeast Zone. Overall, I do not observe statistically significant differences in other household-level, geographic and pre-1975 characteristics in Table 24, indicating that the previous RDD estimates are reliable.

Table 20: Doughnut Regression Discontinuity: DHS Household Wealth

		Household Wealth Distribution				
	Household Wealth (1)	5^{th} Quintile (2)	4^{th} Quintile (3)	3^{rd} Quintile (4)	2^{nd} Quintile (5)	1^{st} Quintile (6)
RD Estimate	-112,424.01	0.23	0.10	0.01	-0.05	-0.31
	(68,507.07)	(0.06)***	(0.03)***	(0.03)	(0.04)	(0.05)***
	[0.09]*	[0.00]***	[0.00]***	[0.75]	[0.26]	[0.00]***
Observations	44,432	44,432	44,432	44,432	44,432	44,432
Border Segment FEs	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	80.83	82.49	69.76	76.05	87.00	82.52

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 21: Doughnut Regression Discontinuity: DHS Years of Education, by Birth Cohort

	Years of Education Education (1)	Year of Birth 1990-99 (2)	Year of Birth 1980-89 (3)	Year of Birth 1975-79 (4)	Year of Birth 1965-74 (5)	Year of Birth 1955-64 (6)
RD Estimate	-2.09	-2.51	-2.52	-2.89	-2.51	-2.39
	(0.37)***	(1.42)*	(0.72)***	(0.72)***	(0.57)***	(0.38)***
	[0.00]***	[0.10]*	[0.00]***	[0.00]***	[0.00]***	[0.00]***
Observations	44,183	561	6,326	4069	11730	10388
Border Segment FEs	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	72.76	106.82	71.13	83.70	74.72	62.51

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 22: Doughnut Regression Discontinuity: Structural Transformation

	Agriculture: (Female) (1)	Agriculture: (Male) (2)	Services (Female) (3)	Services (Male) (4)	Manual Labour (Female) (5)	Manual Labour (Male) (6)
RD Estimate	0.38	0.31	-0.26	-0.16	-0.03	-0.09
	(0.07)***	(0.06)***	(0.04)***	(0.03)***	(0.02)*	(0.02)***
	[0.00]***	[0.00]***	[0.00]***	[0.00]***	[0.15]	[0.00]***
Observations	51,616	51,616	51,616	51,616	51616	51,616
Border Segment FEs	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	77.76	81.23	74.33	78.75	60.69	98.46

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 23: Doughtnut Regression Discontinuity: Investment

	Bank Account (1)	Piped Water (2)	Telephone Connection (3)
RD Estimate	-0.13	-0.57	-0.08
	(0.03)***	(0.19)***	(0.03)***
	[0.00]***	[0.02]**	[0.00]***
Observations	30,785	44,432	30,785
Border Segment FEs	Yes	Yes	Yes
Bandwidth (km)	83.11	21.00	73.30

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for p<0.01, ** for p<0.05 and * p<0.1.

Table 24: DHS Doughnut Regression Discontinuity: Robustness Tests

	Household Size	Female Head	Age	Slope	Elevation	Rainfall	Temperature	Pop Density	Unexploded Ordnance	Bombing (Weight)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RD_Estimate	0.13	-0.03	0.54	0.04	8.39	0.11	-0.12	-11.48	-1.42	414.77
	(0.11)	(0.02)	(0.72)	(0.04)	(1.67)***	(0.45)	(0.02)***	(3.56)***	(1.34)	(369.50)
	[0.33]	[0.10]*	[0.50]	[0.39]	[0.01]**	[0.97]	[0.00]***	[0.15]	[0.21]	[0.33]
Observations	45416	45,416	45,416	1765	1,765	1765	1,765	1,765	1,765	1,765
Border Segment FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	64.87	88.68	81.66	46.26	24.03	26.15	29.61	47.62	63.96	67.52

Notes: All estimations use the first polynomial of the running variable, i.e. distance to Northeast Zone border, and include 25km border segment fixed effects. Heteroskedasticity-robust standard errors clustered using nearest neighbour observations are reported in parentheses. P-values from the robust bias-corrected inference procedure proposed by Calonico et al. (2019a) reported in square brackets. Levels of statistical significance indicated by: *** for $p<0.01$, ** for $p<0.05$ and * $p<0.1$.