Online Appendices for "State Terror and Long-run Development": Online A-D, Dataverse ONLY E-I

Contents

A Identification Checks: Online Appendix **1**

B Robustness: Online Appendix **3**

C Alternative Explanations: Online Appendix **14**

D Human Capital Mechanism: Online Appendix **17**

E Data: Dataverse Only **25**
E.1 IDPoor . . . 25
E.2 Spatial Data . . . 26
E.3 Summary Statistics (Pretreatment Covariates) . . . 27
E.4 National Road 4: Google Maps Street View . . . 28
E.5 Excess Mortality Plot . . . 30
E.6 One-Dimensional RD Plots: Baseline Development . . . 32
E.7 One-Dimensional RD Plots: Human Capital . . . 33
E.8 One-Dimensional RD Plots: Wealth . . . 34
E.9 One-Dimensional RD Plots: Labor Markets . . . 35
E.10 One-Dimensional RD Plots: Health . . . 36

F Full Model Results **37**

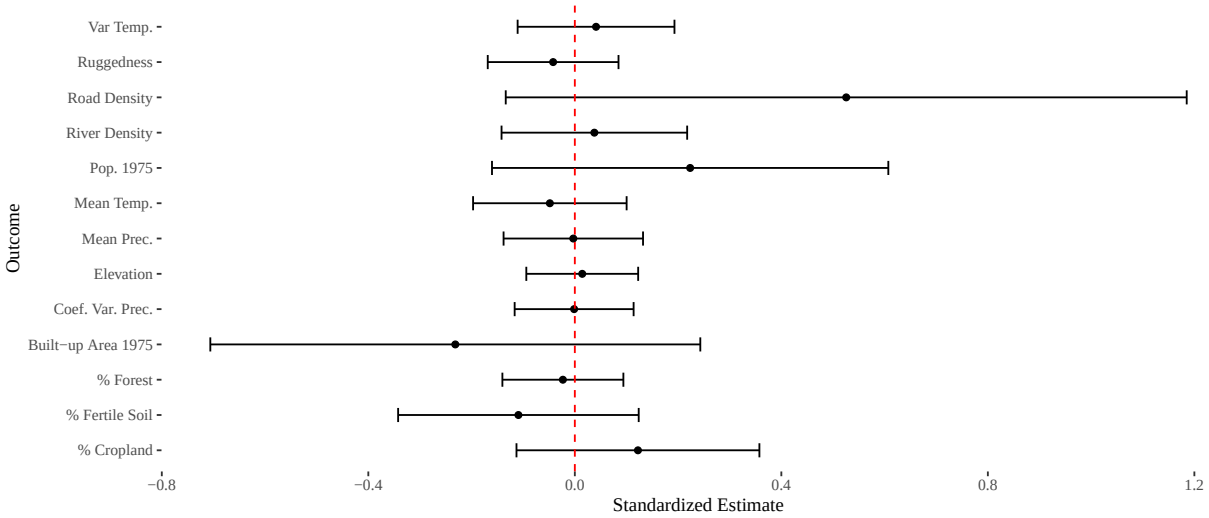
G Trimming Exercise Explanation: Dataverse Only **48**

H Background Information: Dataverse Only **52**
H.1 Zone Commanders . . . 52
H.2 Educational System Before, During, and After DK . . . 53

I Genocide Intensity: Dataverse Only **54**
I.1 Measuring Genocide Intensity . . . 54
I.2 Estimating Genocide Intensity . . . 59
I.3 Genocide Intensity Results . . . 60

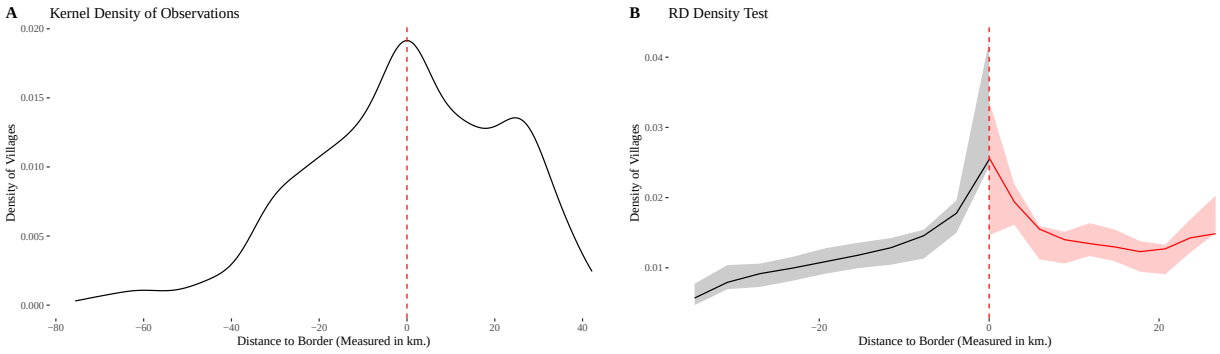
A Identification Checks: Online Appendix

Figure A.1: Balance Tests



Note: Outcomes standardized reported in horizontal axis, vertical axis refers to each respective outcome. Spatial heteroskedasticity and autocorrelation consistent standard errors used to construct equivalence confidence interval (ECI). Equivalence range selected using the sensitivity approach $\epsilon \pm .36\sigma$. Estimates using nonparametric RD within MSE optimal bandwidth, meaning no covariates to report in a tabular format.

Figure A.2: Density Test



Note: Kernel density of observations by running variable in Panel A. Test for discontinuity in density in Panel B showing smoothness of observations at the cutpoint.

B Robustness: Online Appendix

Table B.1: DHS Wealth

	(1) Levels	(2) Logs	(3) Categories
𝟙 SW	-0.80***	-0.42***	-1.23***
	(0.16)	(0.92)	(0.36)
Bandwidth	8.93	9.19	10.32
Effective N	3155	3155	3567

Outcome is the DHS wealth index constructed from the first principal component of household assets. Unit of analysis is the rural individual. Adjusting covariates include gender, age, age squared, and survey wave fixed effects. Column (1) reports the outcome measured in levels, Column (2) reports the natural log of the index, and Column (3) shows the outcome according to categories (quintiles).

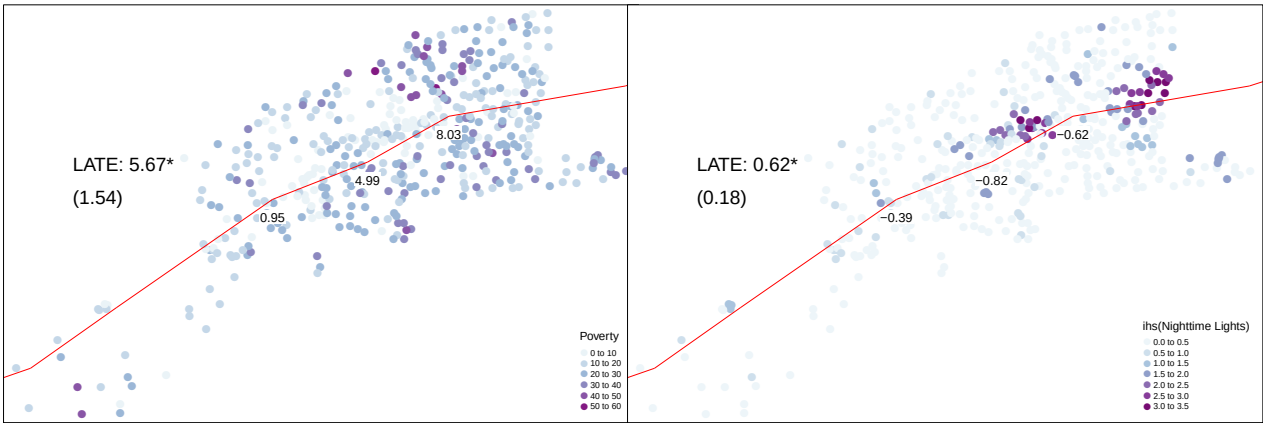
Table B.2: Baseline Results: Multidimensional Forcing Variable

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	%Poverty				IHS Luminosity			
𝟙 SW	5.45**	5.75**	6.37***	5.16***	−0.72***	−0.8***	−0.7***	−0.57***
	(1.8)	(1.85)	(1.43)	(1.45)	(0.12)	(0.13)	(0.11)	(0.1)
Effective N	334	324	502	484	422	389	452	568
Bandwidth	6.34	5.98	10.99	10.62	8.9	7.95	9.64	12.79
μ Control	20.95	20.95	20.95	20.95	0.43	0.43	0.43	0.43
σ DV	10.55	10.55	10.55	10.55	0.63	0.63	0.63	0.63

RD results using a polynomial in latitude-longitude space as the forcing variable. Linear forcing variable models latitude and longitude, denoted as x and y, as: $x + y + xy$. Squared model uses: $x + y + xy + x^2 + y^2 + xy^2 + yx^2 + y^2x^2$.

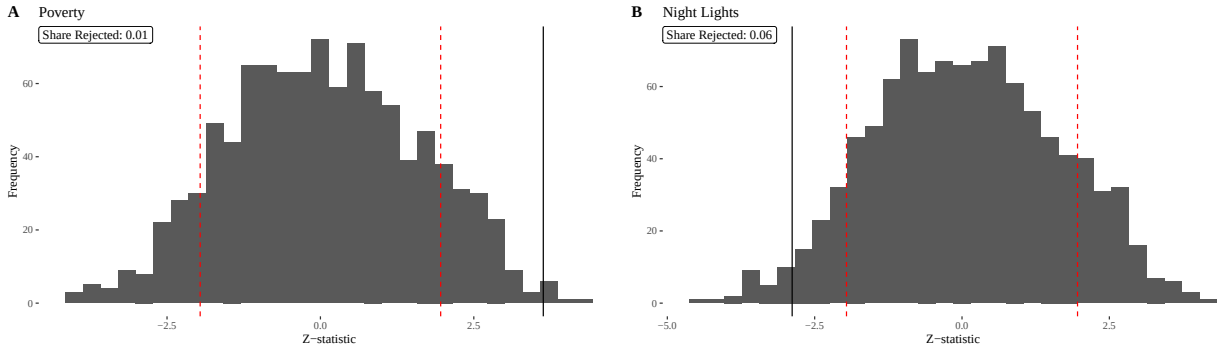
Figure B.1: Treatment Effect Curve

(a) Poverty (b) Luminosity



Notes: RD estimates along border points (reported at each particular point) and aggregated LATE (reported in upper left corner). Standard error computed via the bootstrap. Each dot represents a village with associated shading corresponding to level of poverty or luminosity respectively.

Figure B.2: Noise Simulations



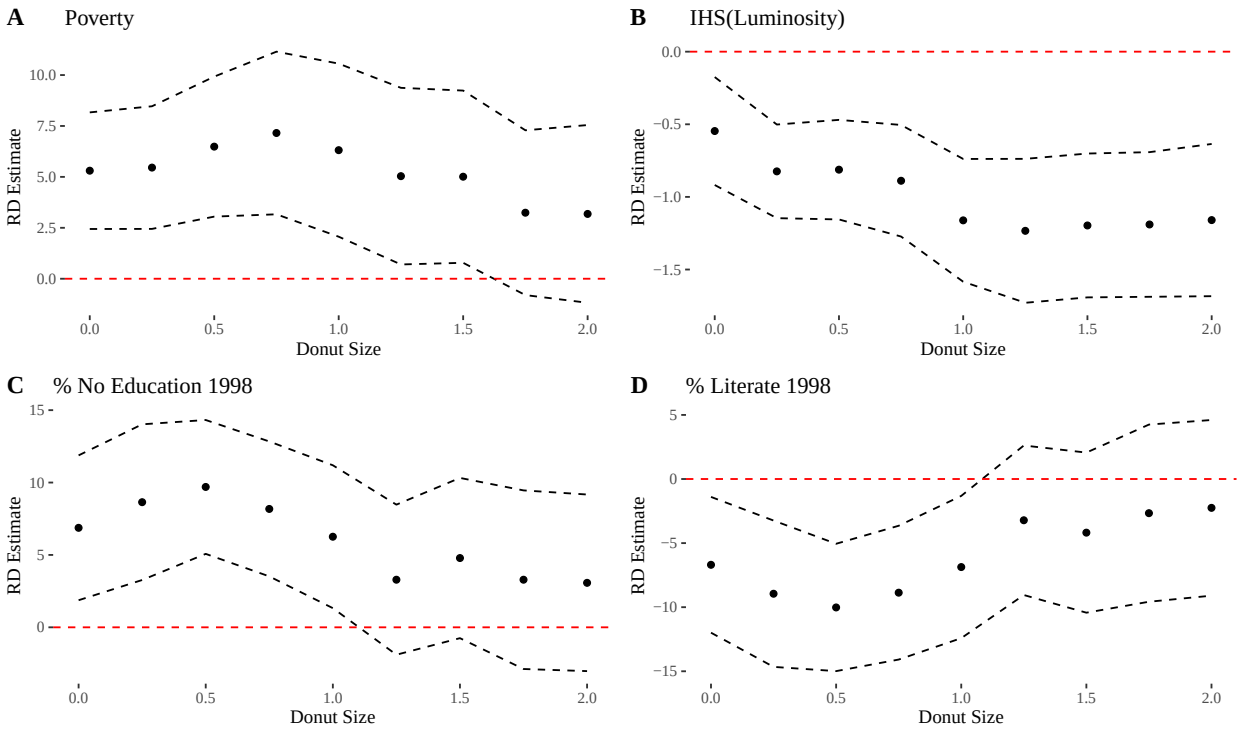
Note: Histograms of z-statistics from CCT robust standard errors. Outcome is simulated spatial noise for each respective outcome. Vertical red line is the z-statistic using the true data. Upper right corner is the proportion of z-statistics from simulations that are more extreme than the estimates from the true data.

Table B.3: Power Analysis:

		Power Against			
Kernel	H0: $\tau = 0$	$0.2^*\tau$	$0.5^*\tau$	$.8\ ^*\tau$	$\tau = \widehat{\tau}$
		Panel A: Poverty			
Uniform:	.05	.089	0.304	0 .638	0.824
Triangular:	.05	0.129	0.532	0.904	0.983
		Panel B: Luminosity			
Uniform:	.05	0.145	0.609	0 .947	0.994
Triangular:	.05	0.117	0.47	0 .854	0.965

Note: Power analysis of nonparametric robust bias-corrected regression discontinuity design for primary outcomes of interest (poverty and luminosity). Each column shows the power of the test against various null hypotheses based on the hypothesized effect size. The column to the furthest to the right reports the power against assuming the effect size detected in the study is the true value of τ, moving to the left the size of τ is decreasing. Power analysis includes border segment fixed effects.

Figure B.3: Excluding Observations Near Threshold



Note: Estimation using CCT nonparametric approach and confidence intervals. Size of donut-hole expands at .25 kilometer increments starting with .5 kilometers. Each estimate drops additional data. Results estimated with nonparametic RD, as such there are no covariates to report.

Table B.4: Sensitivity Analysis

Treatment:	Est.	S.E.	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
			Outcome: *Poverty*			
treat	3.486	1.783	1.956	1.2%	10.2%	0%
df = 328		*Bound (4x Built Area 1975)*: $R^2_{Y\sim Z\|\mathbf{X},D}$ = 9.4%, $R^2_{D\sim Z\|\mathbf{X}}$ = 15.2%				
			Outcome: *Luminosity*			
𝟙 SW:	-0.574	0.14	-4.113	4.2%	18.9%	10.4%
df = 383		*Bound (4x Built Area 1975)*: $R^2_{Y\sim Z\|\mathbf{X},D}$ = 28%, $R^2_{D\sim Z\|\mathbf{X}}$ = 15.4%				
			Outcome: *Literacy Rate*			
𝟙 SW	-7.262	2.648	-2.743	2.4%	14.4%	4.3%
df = 311		*Bound (4x Built Area 1975)*: $R^2_{Y\sim Z\|\mathbf{X},D}$ = 0%, $R^2_{D\sim Z\|\mathbf{X}}$ = 4.2%				
			Outcome: *No Educ.*			
𝟙 SW	7.42	2.5	2.968	2.8%	15.7%	5.6%
df = 303		*Bound (4x Built Area 1975)*: $R^2_{Y\sim Z\|\mathbf{X},D}$ = 1.2%, $R^2_{D\sim Z\|\mathbf{X}}$ = 4.7%				

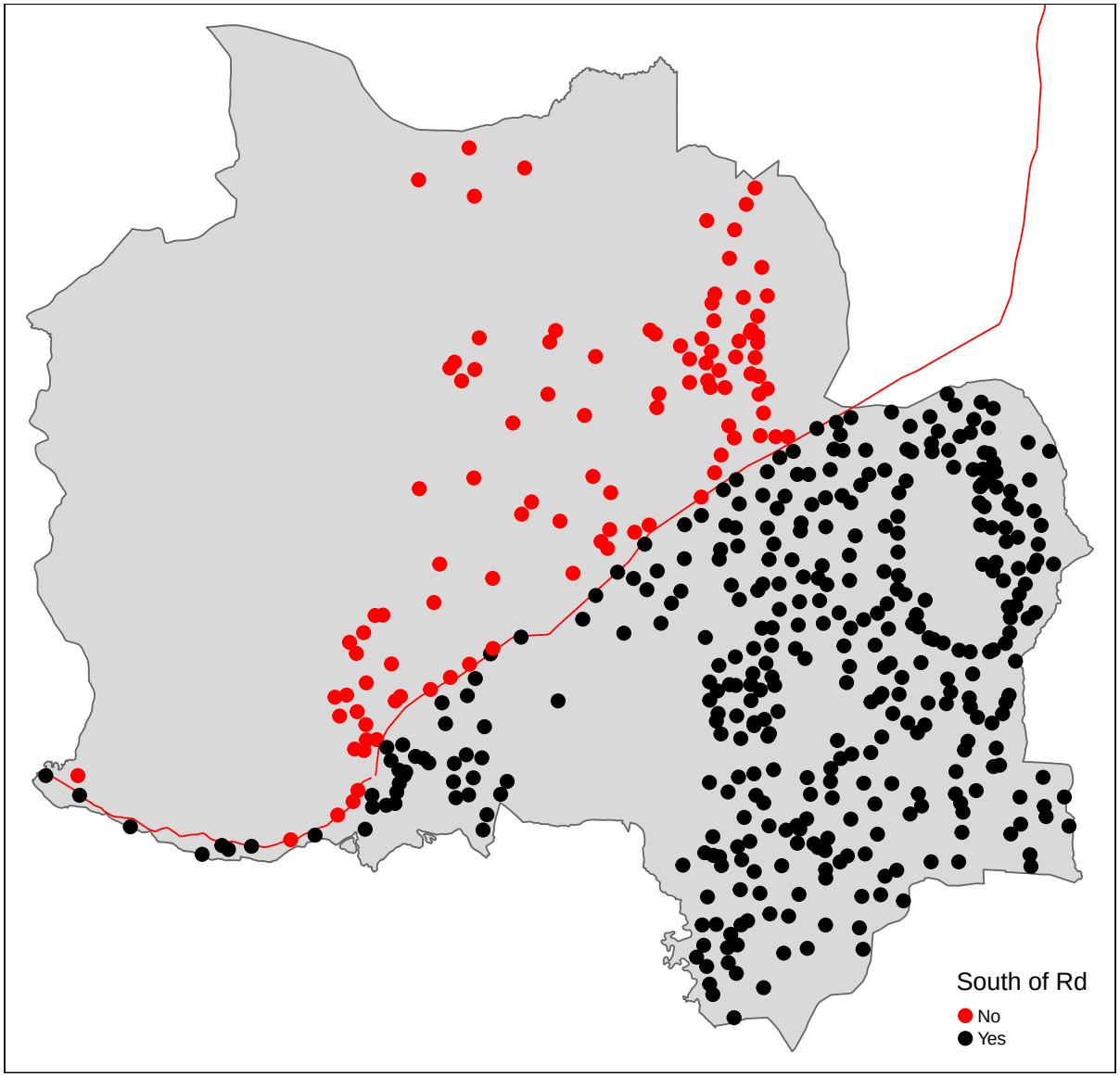
Sensitivity analysis results adjusting for road density and density of built up areas in 2 x 2 kilometer grids surrounding village centers. "Est." column is the estimate, "S.E." is the standard error, "t-value" is the t-statistic. $R^2_{Y\sim D|\mathbf{X}}$ reports how much residual variation in treatment exposure unobserved confounder would need to explain in order to erase the effect of treatment conditional on the unobserved confounder explaining all of the left out variance in the outcome of interest. $RV_{q=1}$ is the robustness value for bringing the estimate of Southwest to zero. Unobserved confounders that explain less than the robustness value's worth of both exposure to the Southwest zone and the outcome of interest are not sufficiently strong to explain away the observed effect.

Table B.5: National Road 3 Placebo: Kampot Province

Outcome	(1) Night Lights	(2) Poverty	(3) No Educ.	(4) Literacy
𝟙 South	-0.14	-1.15	-2.61	1.88
	(0.13)	(1.48)	(2.42)	(2.69)
Effective N	162	112	134	146
Bandwidth	7831.48	5009.04	6427	6974.87
μ Control	0.28	17.11	53.59	56.79
σ DV	0.5	7.38	12.57	14.2

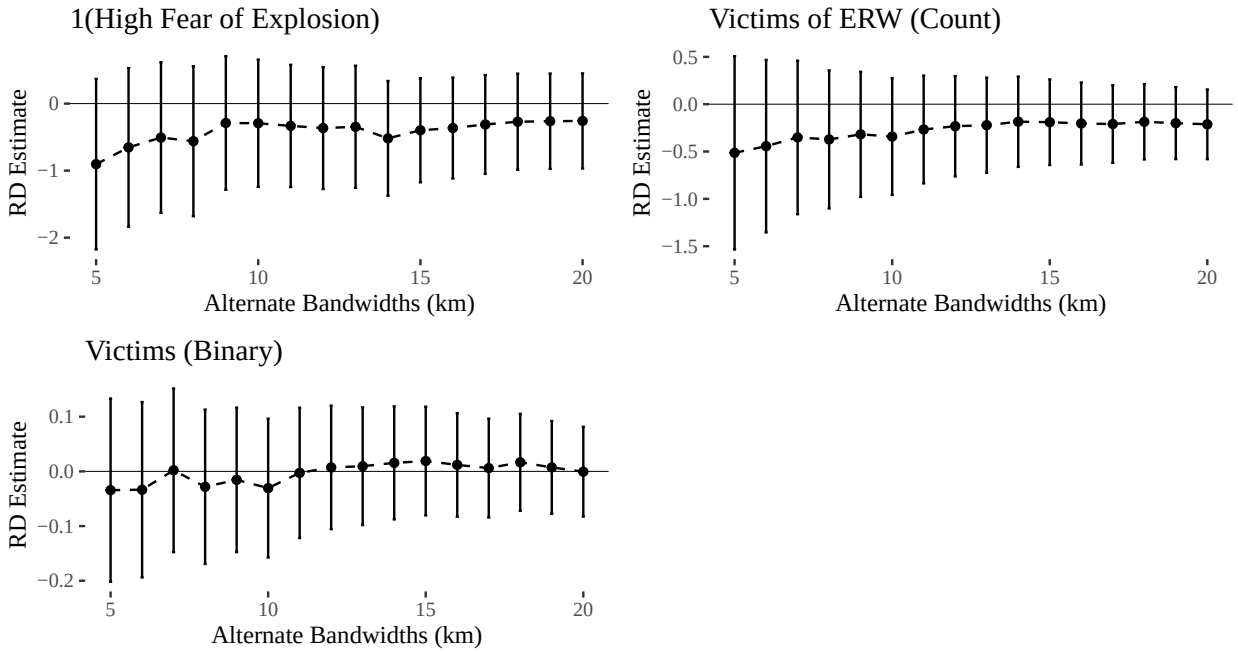
𝟙 South is a binary indicator for a village being South of National Road 3 within Kampot province (See Figure B.4 for reference). All villages within Kampot province, which was entirely in the Southwest Zone during the DK and civil war period (1970-1979).

Figure B.4: Road Placebo: National Highway 3



Note: Map showing the province and villages used for the National Highway 3 placebo test.

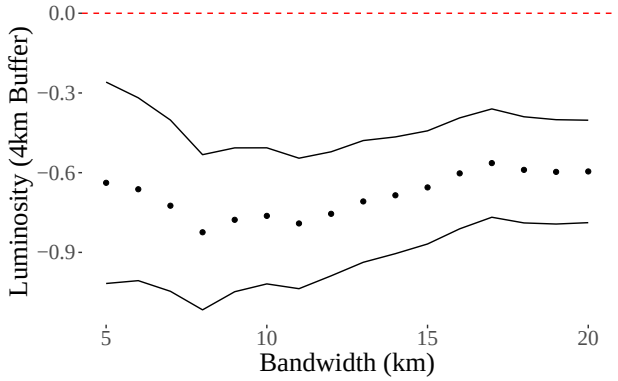
Figure B.5: Explosive Remnants of War (ERW) and Landmine Exposure: Post 2000



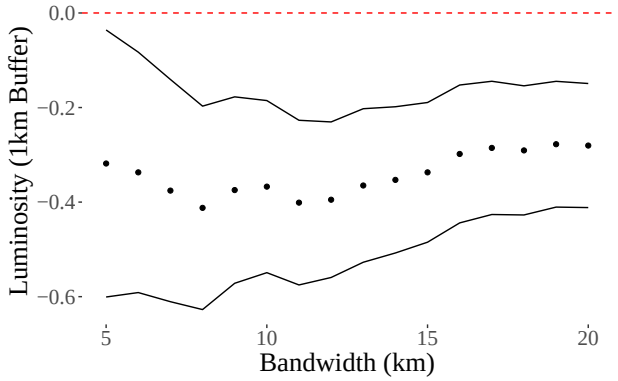
Note: 95% CCT robust confidence intervals shaded around estimates, uniform kernel, alternative bandwidths

Figure B.6: Luminosity: Other Aggregation Grids

(a) Nighttime Lights (4km)

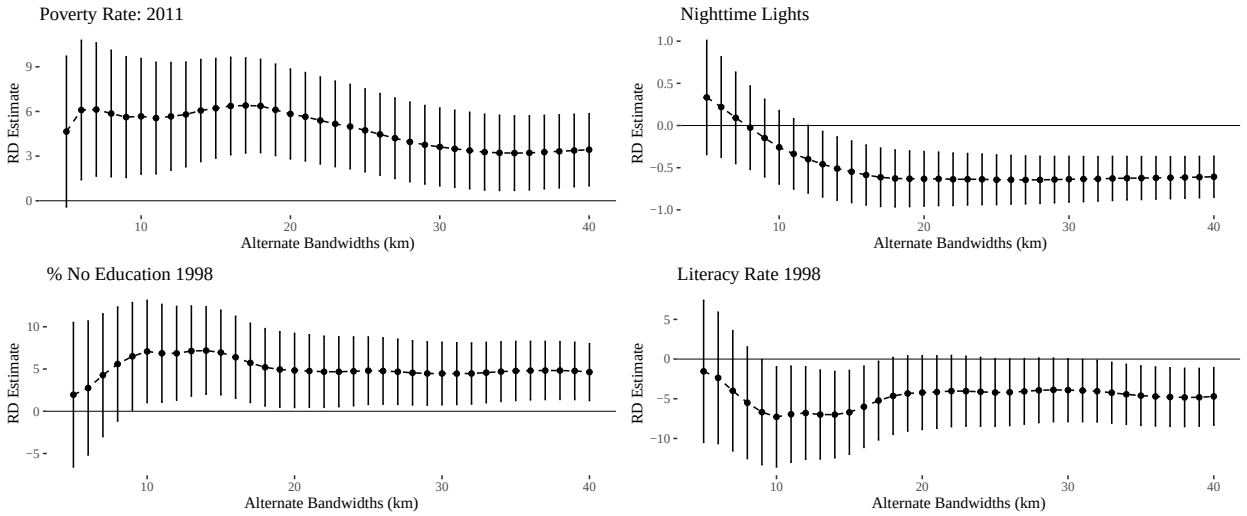


(b) Nighttime Lights (1km)



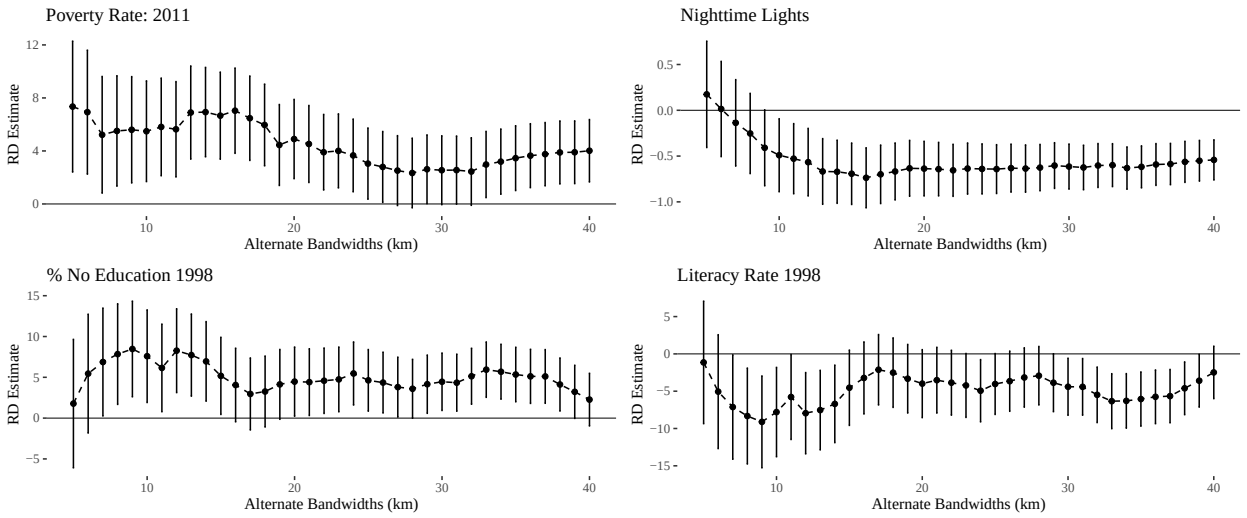
Notes: Semi-parametric RD estimates at alternative bandwidths. Panel A shows the results where a 4 km buffer is created around villages to compute luminosity. Panel B shows results where a narrower 1 km buffer is used to compute luminosity. SHAC standard errors used to construct 95% confidence bands. See Table F9 for full model results.

Figure B.7: Nonparametric RD: Alternative Bandwidths, Triangular Kernel



Note: Estimation using CCT nonparametric approach and confidence intervals at alternative bandwidths with triangular kernel.

Figure B.8: Nonparametric RD: Alternative Bandwidths, Uniform Kernel



Note: Estimation using CCT nonparametric approach and confidence intervals at alternative bandwidths with uniform kernel.

C Alternative Explanations: Online Appendix

Table C.1: Effect of Southwest on Village Development (Covariate Adjusted)

Outcome	(1)	(2)	(3)	(4)
	%Poverty		IHS Luminosity	
SW	$3.21^{\dagger}$	4.36^{*}	-0.53^{***}	-0.53^{***}
	(1.71)	(2.03)	(0.12)	(0.13)
Effective N	340	505	439	618
Bandwidth	6.58	11.12	9.31	14.37
μ Control	20.95	20.95	0.43	0.43
σ DV	10.55	10.55	0.63	0.63
Segment FE	✓	✓	✓	✓
Dist. Capital Covariate	✓	✓	✓	✓
Pre-DK covariates	✓	✓	✓	✓
Linear	✓	-	✓	-

Note: % Poverty is the count of level 1 and level 2 poverty divided by the number of households per village as measured by IDPoor in 2011. Nighttime lights are the inverse hyperbolic sine of the sum of estimated GDP from luminosity in a 2x2 kilometer grid cell surrounding the village centroid. Estimates include the following pre-DK covariates: distance to the provincial capital, the sum of built up area around the grid cell surrounding the village in 1975, road density in the grid cell surrounding the village. See Table F6 for full model results.

Table C.2: Public Goods Access

Outcome: Distance to:	Hospital (1)	School (2)	Commune Center (3)
𝟙 SW	0.21	-0.00	0.39
	(0.34)	(0.12)	(0.55)
N	297	378	365
BW	5.25	7.75	7.54

Notes: See Table 2. Outcomes are village distance to nearest public good (kilometers)

Table C.3: International Migration (Commune)

	(1)	(2)	(3)	(4)
𝟙	0.08	−0.03	0.10	0.07
	(0.09)	(0.24)	(0.08)	(0.23)
N.	87	87	87	87
Effective N.	87	87	87	87
District FE	-	-	7.46	7.46

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

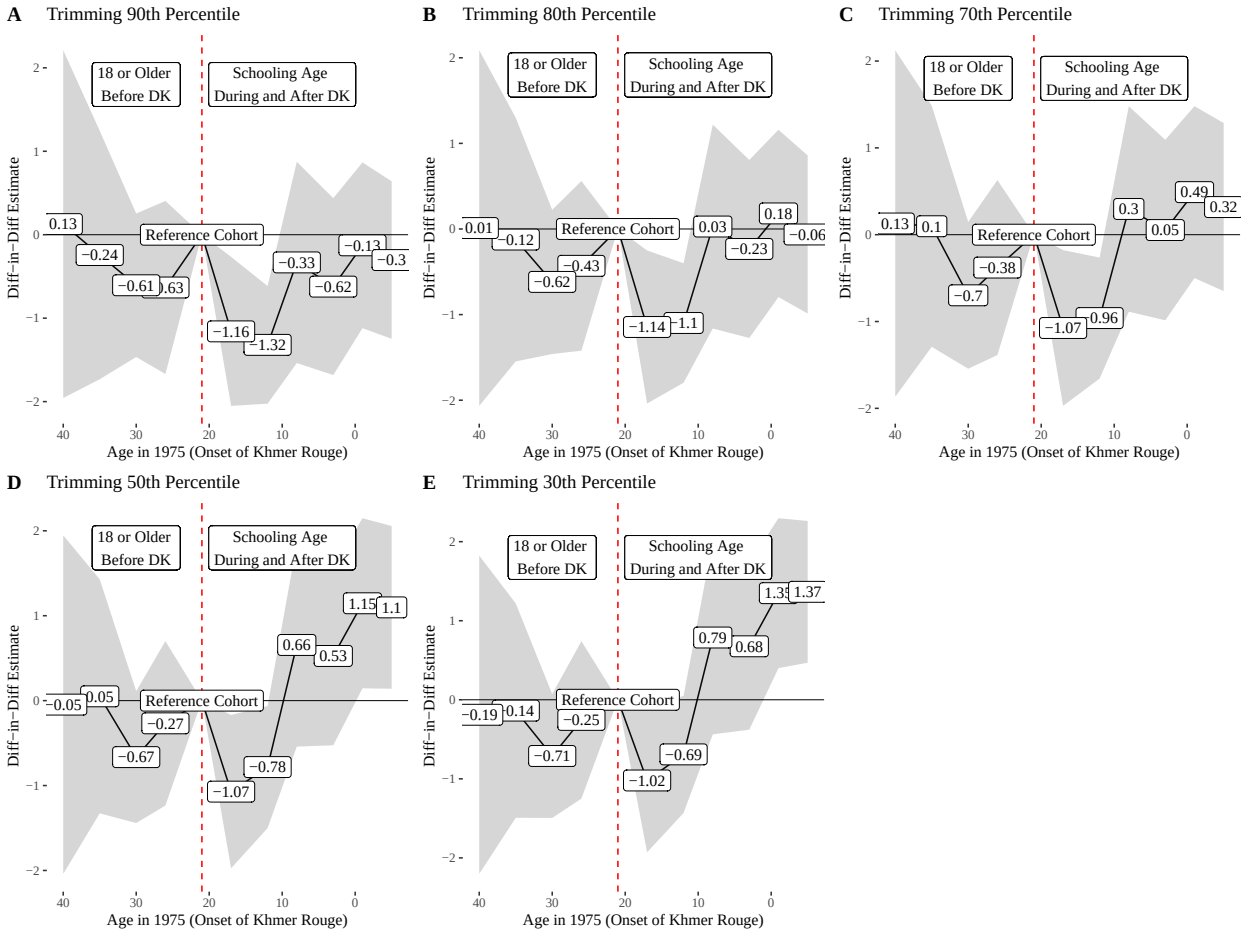
Robust standard errors reported in parentheses. Odd columns use a quadratic of the running variable

Table C.4: Results from Trimming: Wealth

	(1)	(2)	(3)	(4)	(5)	(6)
x%	95	90	85	80	75	70
𝟙SW	-0.77***	-0.71***	-0.62***	-0.50***	-0.37*	-0.21
2	(0.17)	(0.15)	(0.15)	(0.15)	(0.19)	(0.19)
Bandwidth	8996.16	9053.37	9190.09	9621.71	10270.48	10937.85
Total N	13100	13063	12936	12733	12505	12280
Effective N	3152	3138	3103	3148	3352	3388

Note: Outcome is DHS wealth data. Each column drops a percentile of top wealthiest persons in the West zone - e.g. Column (1) drops the top 5% wealthiest from the West zone and retains the bottom 95%, Column (2) drops the top 10% and retains the bottom 90%, ect.

Figure C.1: Event Studies Trimming Upper Education Percentiles



Note: Robust errors clustered at the village. Each panel drops top percentile of schooled persons from the West zone. See Table F11 for full model results.

D Human Capital Mechanism: Online Appendix

Table D.1: Human Capital in 1998: Years of Schooling and Attendance

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Yrs Educ.				Attendance Rate			
SW	-1.1^{**}	-0.58	-1.1^{*}	$-0.69^{\dagger}$	-6.02^{**}	-4.64^{*}	-7.79^{**}	-3.42
	(0.41)	(0.39)	(0.43)	(0.38)	(2.15)	(2.16)	(2.43)	(2.29)
Effective N	312	285	442	439	355	313	597	476
Bandwidth	5.78	4.85	9.87	9.66	7.13	5.85	14.27	10.78
μ Control	4.16	4.16	4.16	4.16	30.07	30.07	30.07	30.07
σ DV	1.9	1.9	1.9	1.9	12.8	12.8	12.8	12.8
Segment FE	-	✓	-	✓	-	✓	-	✓
Dist. Capital Covariate	-	✓	-	✓	-	✓	-	✓
Linear	✓	✓	-	-	✓	✓	-	-
Quadratic	-	-	✓	✓	-	-	✓	✓

Note: Yrs. Educ. is the average years of education in a village. Attendance Rate is the share of persons who are enrolled in school below 25 (i.e. schooling aged). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ See Table F7 for full model results.

Table D.2: Human Capital in 1998: Adjusting for Distance to Schools

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Literacy Rate		% No Educ.		Yrs. Educ.		Attendance Rate	
𝟙 SW	-4.94^{*}	-5.62^{*}	4.23^{*}	5.15^{*}	$-0.62^{\dagger}$	-0.76^{*}	-5.14^{*}	$-4.21^{\dagger}$
	(2.44)	(2.71)	(2.11)	(2.41)	(0.37)	(0.37)	(2.08)	(2.3)
Effective N	302	424	313	430	286	436	313	451
Bandwidth	5343.89	9249.23	5847.32	9352.39	4860.19	9592.27	5843.17	10170.76
μ Control	62.88	62.88	50.48	50.48	4.16	4.16	30.07	30.07
σ DV	17.03	17.03	14.85	14.85	1.9	1.9	12.8	12.8

Note: Literacy Rate is the percentage of persons over 15 who can read write. % No Educ. is the percentage of people who have no schooling. Yrs. Educ. is the average years of education in a village. Attendance Rate is the share of persons who are enrolled in school below 25 (i.e. schooling aged). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table D.3: School Outcomes

School Outcome	Staff/Student Ratio		Students Per Classroom	
SW	0.28	0.47	0.01	-46.26^{**}
	(0.20)	(1.08)	(3.13)	(16.47)
N.	495	495	495	495

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Robust standard errors reported in parentheses. Data collected at the school level. Staff/student ratio is the number of employees in the school divided by the number of students. Students per classroom is the number of students divided by the number of rooms in the school.

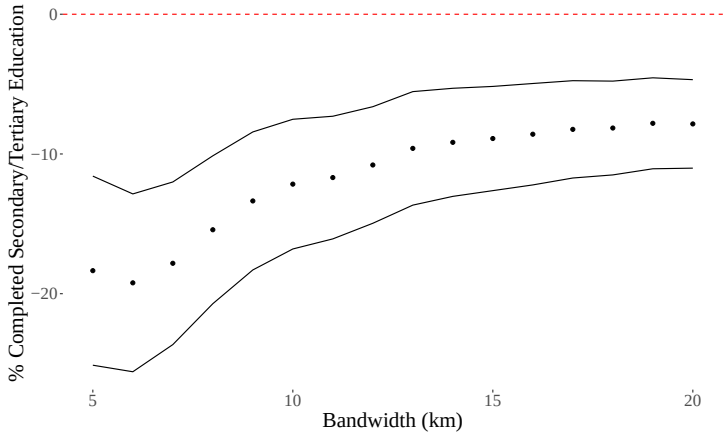
Table D.4: Schooling Persistence: 2008 Census

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Attendance Rate		Yrs. School		Literacy Rate		% No Educ.	
SW	-1.61	$-2.25^{\dagger}$	-0.13	-0.24	-1.29	-2.27	-0.82^{**}	-1.15^{*}
	(1.24)	(1.23)	(0.12)	(0.15)	(1.89)	(2.18)	(0.3)	(0.47)
Effective N	336	612	294	391	358	569	405	710
Bandwidth	6.48	14.11	5.03	8.04	7.17	12.8	8.49	17.1
μ Control	29.77	29.77	5.16	5.16	75.04	75.04	1.18	1.18
σ DV	6.46	6.46	0.65	0.65	14.07	14.07	2.13	2.13

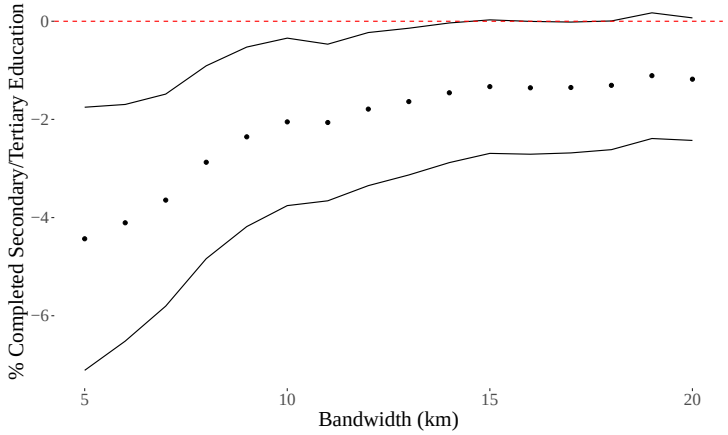
Note: Unit of analysis is the village. SHAC standard errors reported in parentheses.

Figure D.1: Tertiary and Secondary Education (All Data)

(a) Tertiary and Secondary Education



(b) Tertiary and Secondary Education (Dropping Outliers)



Notes: Parametric RD estimates at alternative bandwidths. Panel A shows results using all villages. Panel B shows results where outlying positive observations (highly educated villages) are dropped from the analysis. Horizontal axis reports different evaluation bandwidths. SHAC standard errors used to construct 95% confidence bands. See Table F11 for full model results.

Table D.5: Education Differences by Gender: 1998 Census

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Males	%No Educ. Males				Lit. Rate Males			
SW	6.83**	3.98†	4.31†	3.77	−5.49*	-3.23	−5.86*	−4.59†
	(2.23)	(2.06)	(2.39)	(2.38)	(2.25)	(2.13)	(2.66)	(2.49)
Effective N	320	311	624	452	317	321	528	435
Bandwidth	6157.66	5758.35	15246.56	10249.45	6003.24	6214.12	11983.37	9554.23
B: Females	%No Educ. Females				Lit. Rate Females			
SW	8.33**	3.81	8.35**	4.95†	−9.41**	-4.53	−10.82**	−6.4†
	(2.77)	(2.49)	(2.95)	(2.77)	(3.18)	(3.15)	(3.7)	(3.42)
Effective N	306	312	507	434	309	279	413	412
Bandwidth	5526.79	5808.96	11714.12	9432.27	5632.18	4690.57	8993.07	8951.34
C: Gap	%No Educ. Gender Gap				Lit. Rate Gender Gap			
SW	3.33†	1.64	3.8†	2.48	-1.38	-1.2	-1.52	-1.51
	(1.97)	(1.94)	(2.13)	(2.06)	(0.85)	(0.89)	(1)	(0.97)
Effective N	310	314	447	460	440	417	611	668
Bandwidth	5658.08	5893.41	10076.75	10401.42	9786.66	9125.82	14902.86	16266.69

RD estimates using education by gender as the outcomes of interest. Panel A studies the rates of no education and literacy by males, and Panel B by females. Panel C studies the gender gap in these outcomes, defined as the difference between human capital rates by group. Overall, I find little to no evidence of differential gender effects.

Table D.6: Placebo Tests: Cohort Analysis

	(1)	(2)	(3)	(4)
SW × 𝟙(DK age <= 35)	−0.49			
	(1.37)			
SW × 𝟙(DK age <= 30)		−0.58		
		(0.81)		
SW × 𝟙(DK age <= 25)			0.13	
			(0.66)	
SW × 𝟙(DK age <= 20)				0.42
				(1.08)
N.	537	537	537	537
SD DV	3.5	3.5	3.5	3.5
Village FE	✓	✓	✓	✓
Commune by Decade FE	✓	✓	✓	✓
Gender FE	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: Village clustered errors reported in parentheses. Outcome is the years of schooling. Sample is individuals who were over 18 years old in 1975, meaning they would have completed schooling before the DK regime began.

Table D.7: Child Health Between Zones by Maternal Education Level

Outcome	(1) Health Index	(2) Height/Age	(3) Weight/Age	(4) Weight/Height
	Panel A: Mothers with No Education			
𝟙 SW	-2.96^{**}	-1.70	-1.74^{**}	-3.31^{***}
	(0.93)	(1.65)	(0.53)	(0.38)
	[0.79]	[0.84]	[0.68]	[0.39]
N. Individuals	65	45	73	46
N. Clusters	25	17	27	18
	Panel B: Mothers with Education			
𝟙 SW	-0.30	0.14	-0.20	-0.84^{***}
	(0.40)	(0.22)	(0.29)	(0.20)
	[0.37]	[0.23]	[0.27]	[0.41]
N. Individuals	233	170	233	170
N. Clusters	38	26	38	26
Controls	✓	✓	✓	✓
SD DV	1.38	1.26	0.99	0.98

Note: Unit of analysis is the 3-5 year old individual from the 2000-2014 DHS survey waves - the children of the generation exposed to the Khmer Rouge. Health index (Column 1) is the first principal component of individual health measures. Height/Age is the standard deviations from the median of individual height for age (stunting), Weight/Age is standard deviations from the median of weight for age (wasting), Weight/Height is standard deviations from the median of weight for height (underweight). Analysis within rural households to maximize comparability. Controls include the age of the mother and its square and survey year fixed effects. Robust standard errors clustered at the survey area reported in parentheses. Panel A studies children with mothers without education. Panel B studies children of mothers with at least some education. Clustered standard errors, clustered by survey area, reported in parentheses. Wild cluster bootstrapped errors in brackets. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$ See Table F8 for full model results.

Table D.8: Schooling, Self Employment, and Income

	(1) Self Employment	(2) Income	(3) Income
Years of School	-0.03^{***}		
	(0.00)		
Age	4.44^{***}	-0.19	1.64
	(0.42)	(0.92)	(1.72)
Age^2	-2.75^{***}	1.67^{*}	0.69
	(0.41)	(0.82)	(1.13)
Rural	0.09^{**}	-0.29^{***}	-0.24^{***}
	(0.03)	(0.06)	(0.07)
Female	0.00	-0.02	-0.00
	(0.03)	(0.05)	(0.06)
Self Employed		-0.51^{***}	-0.86^{**}
		(0.06)	(0.28)
N.	975	975	975
Adj. R^2	0.23	0.11	0.09
Estimator	OLS	OLS	2SLS

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Labor Force Survey. Unit of analysis is employed working aged (11-59) individuals. Data from 2000-2001 Labor Force Survey. Pr(Self Employed) is scored 1 for persons who are own account workers. Income is individual wages, remuneration, earnings, tips reported from the last month in 10,000 riels, and productivity is riels divided by working hours.

E Data: Dataverse Only

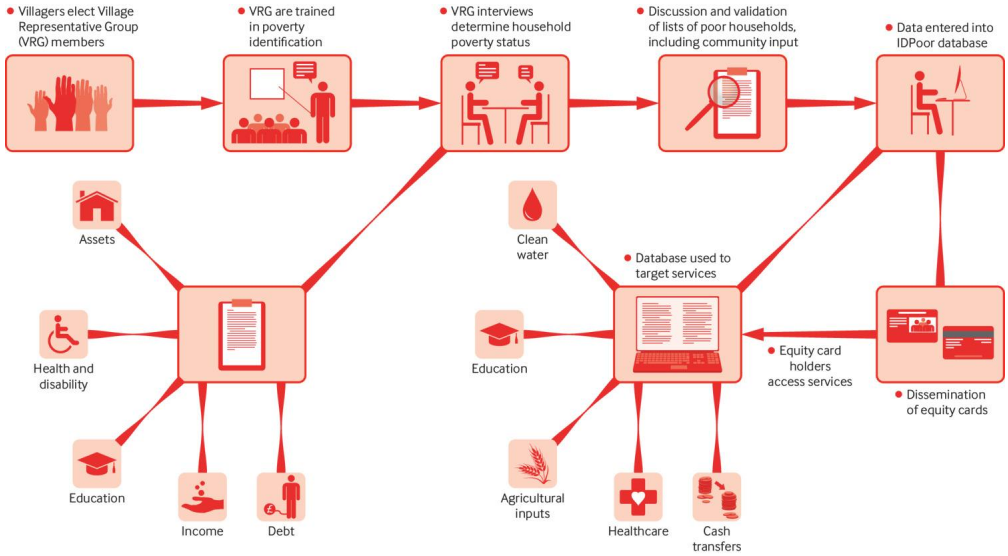
E.1 IDPoor

IDPoor provides a comprehensive measure of poverty in Cambodia. By combining household assets, education levels, and using community consultations, several dimensions of poverty are captured and validated (Kaba et al., 2018). For example, a measure of consumption alone may not account for individuals who consume the same amount but have different investments or assets.

IDPoor data collection process pictured in Figure E.1 from Kaba et al. (2018). Poor level 1 refers to the extreme poverty, with level 2 referring to poor. Those not in poverty are considered average by national standards or better off. By way of illustration, interviewers observe the building material of the roof of the household and score 8 it is made of soft materials (palm leaves or thatch), and 0 if its constructed with concrete. I measure the poverty rate of each village as the sum of households in category 1 or 2 divided by the total number of households to avoid capturing population differences.

During the process of data collection, the Ministry of Planning (MOP) monitors the implementation of the process. Once a list of poor households is drafted, the list is published and open to the community to enable validation of the list (Kaba et al., 2018). Indeed, "World Bank assessment determined that, on average, surveyed households rated the accuracy and implementation of the IDPoor process as high" (Kaba et al., 2018).

Figure E.1: IDPoor Data Collection Process



Note: Flow chart downloaded from Kaba et al. (2018).

E.2 Spatial Data

Mean Temperature: Average temperature per grid cell from 1970-2000 Fick and Hijmans (2017)

Seasonality Temperature: Standard deviation temperature between months times 100 Fick and Hijmans (2017)

Mean Precipitation: Average annual precipitation Fick and Hijmans (2017)

Coefficient of Variation: Average divided by standard deviation Fick and Hijmans (2017)

Elevation: Sea level elevation Farr et al. (2007)

Ruggedness: Standard deviation of elevation from Shaver, Carter and Shawa (2019)

% Crop Land: Percentage of grid cell classified as crop land Shaver, Carter and Shawa (2019)

Distance to Roads: Distance to nearest road

Distance to Rivers: Distance to nearest river

Built up Area: Global Human Settlement Pesaresi et al. (2016)

Population 1975: Global Human Settlement Pesaresi et al. (2016)

E.3 Summary Statistics (Pretreatment Covariates)

Table E.1: Summary Statistics (Spatial Data)

Zone		West			Southwest	
Variable	N	Mean	SD	N	Mean	SD
Temp. Mean	1427	272.273	4.049	1442	271.055	3.709
Temp. Var	1427	957.353	68.899	1442	903.258	61.999
Prec. Mean	1427	1458.876	180.991	1442	1442.967	183.759
Prec. Var	1427	67.214	1.884	1442	65.807	2.266
Forest Land	1427	0.741	0.427	1442	0.648	0.463
Crop Land	1427	0.067	0.239	1442	0.087	0.266
Soil Fertility	1427	0.432	0.46	1442	0.369	0.46
Ruggedness	1427	41.919	75.132	1442	75.949	111.381
Elevation	1427	73.708	72.288	1442	89.318	64.537
River Density	1427	214.766	618.628	1442	170.083	561.766
Road Density	1427	231.913	633.369	1442	353.749	779.167
Population (1975)	1427	68.14	365.967	1442	105.304	457.868
Built-up Area (1975)	1427	0.31	3.798	1442	0.136	0.989

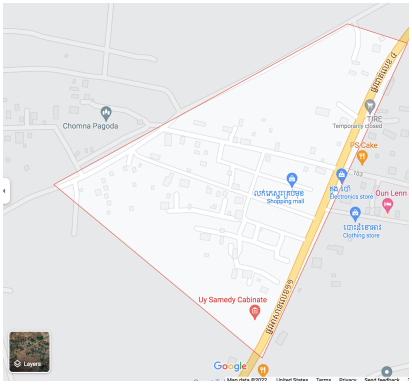
E.4 National Road 4: Google Maps Street View

Figure E.2: National Road 4 (Asian Highway 11) Street View in Kampong Speu

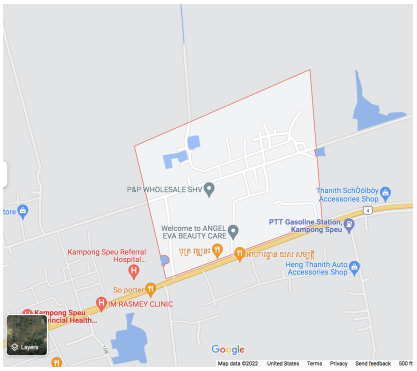
Note: Google Maps Street View of the highway (National Road 4, also known as Asian Highway 11) that separated the former West and Southwest zone. West zone shown on the left, Southwest zone shown on the right.

Figure E.3: Example Border Villages

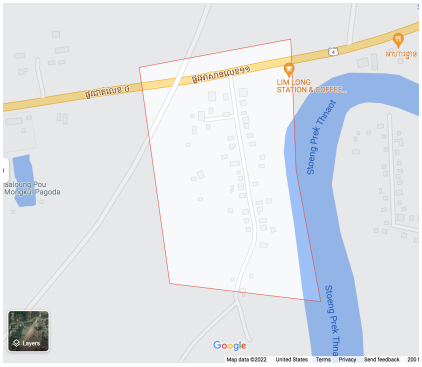
(a) Village A



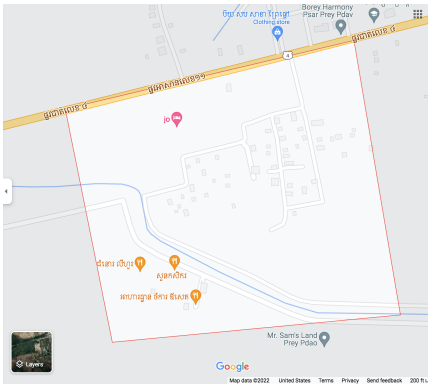
(b) Village B



(c) Village C



(d) Village D



Notes: Illustration of village polygons and their relationship to National Road 4 (the defunct border during the DK era). The maps show that in instances where village borders technically pass the road, villagers still tend to cluster towards the center, meaning classical measurement error of exposure to treatment is unlikely to significantly impact the main results.

E.5 Excess Mortality Plot

Exact mortality estimates from the Khmer Rouge are impossible to obtain due to the absence of administrative records. Despite the fact exact mortality counts cannot be computed, I obtain estimates of estimated mortality to construct a relative estimate of excess mortality overtime between Southwestern and Western Kampong Speu. I follow De Walque (2006) and use the 2000 Demographic Health Survey questionnaire, which asks respondents about (1) their siblings demographic information and (2) the year their siblings died. Notably, the data I use will only record the sibling deaths of those who survived to the year 2000. As such, estimates cannot be taken as absolute levels. However, assuming women's survival probabilities were roughly equal between zones within the province - a reasonable assumption given the excessive targeting of men - one can make a conjecture about the relative intensive of violence overtime between places using the DHS data.

Since the data contain the year siblings died, one can plot deaths over time between spaces. However, death does not necessarily mean Khmer Rouge inflicted mortality, since some siblings may have been more likely to parish for demographic reasons irrespective of regime change. To account for this, I develop a model of mortality based on age, age squared, location of residence (urban/rural), and gender. The logistic regression is as follows.

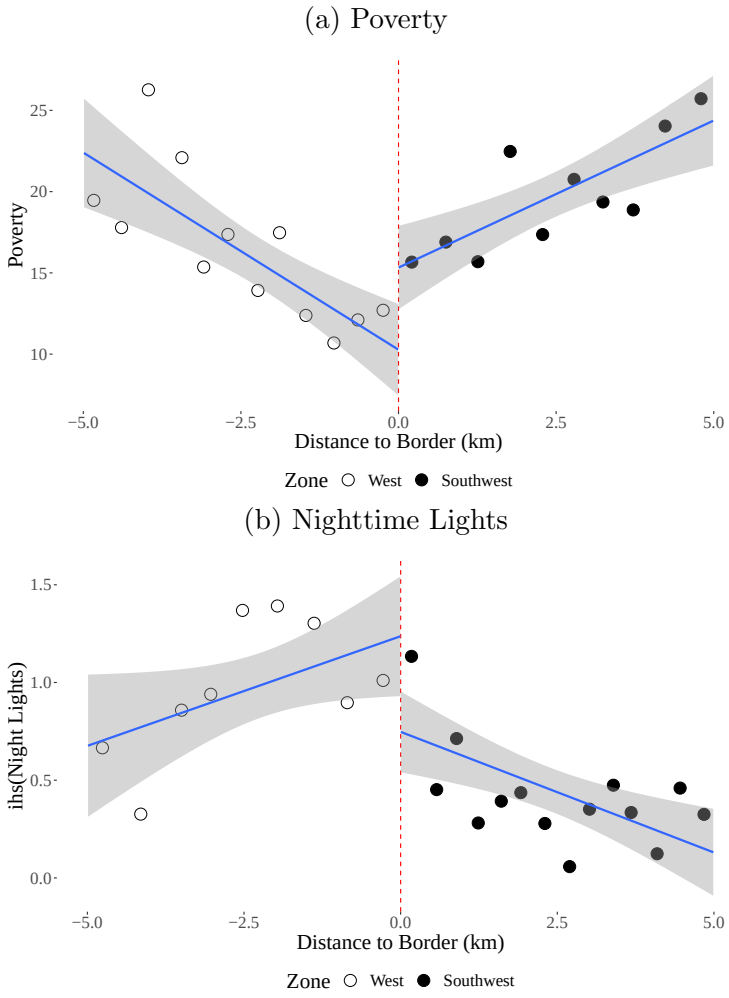
$$Pr(Y_{it} = Dead|X_{it}) = \frac{\exp(\alpha_0 + \alpha_1 Gender_i + \alpha_2 Age_{it} + \alpha_3 Age_{it}^2 + \alpha_4 \mathbb{1}Urban_i)}{1 + \exp(\alpha_0 + \alpha_1 Gender_i + \alpha_2 Age_{it} + \alpha_3 Age_{it}^2 + \alpha_4 \mathbb{1}Urban_i)}$$

I estimate the model on all siblings from the 2000 DHS survey round who were born at least before 1960 *outside* of Kampong Speu. I estimate the model outside of the province of interest to avoid overfitting. Then, I use the estimated coefficients to predict individual

mortality overtime in Kampong Speu using siblings values of the predictor variables. The difference between actual mortalities in a zone-year from predicted mortality in a zone-year represents excess mortality in the zone - that is, deaths which occurred which were not predicted by an individual's demographic traits.

E.6 One-Dimensional RD Plots: Baseline Development

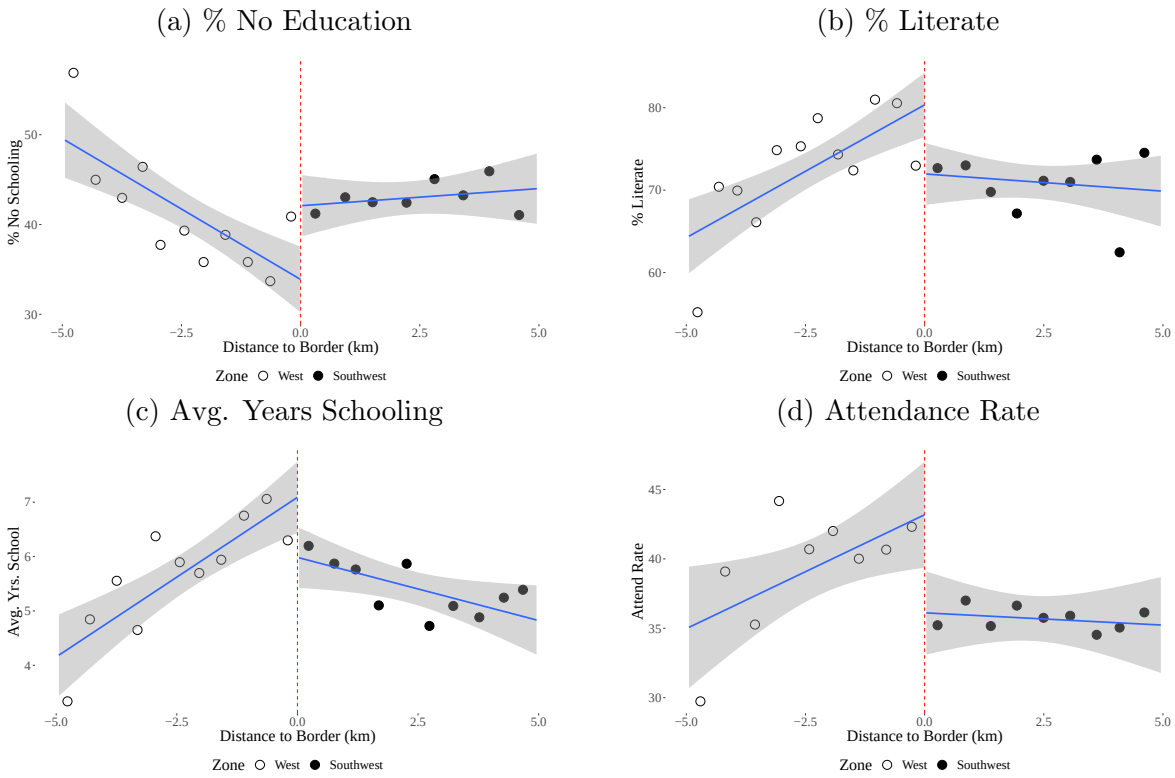
Figure E.4: One-Dimensional RD Plots: Baseline Development



Notes: RD plots illustrating local linear regressions within a 5 kilometer bandwidth. Vertical dashed line marks 0; observations to the right are in the Southwest zone and observations to the left of the line are in the West zone. Dots represent binned averages.

E.7 One-Dimensional RD Plots: Human Capital

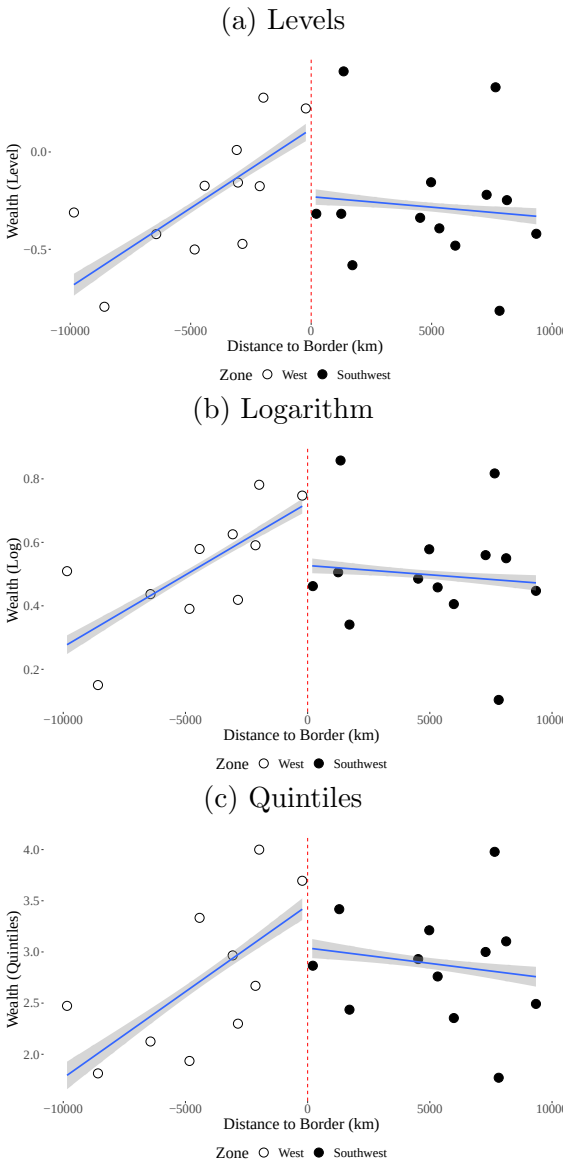
Figure E.5: Human Capital: 1998 Census



Notes: RD plots illustrating local linear regressions within a 5 kilometer bandwidth. Vertical dashed line marks 0; observations to the right are in the Southwest zone and observations to the left of the line are in the West zone. Dots represent binned averages.

E.8 One-Dimensional RD Plots: Wealth

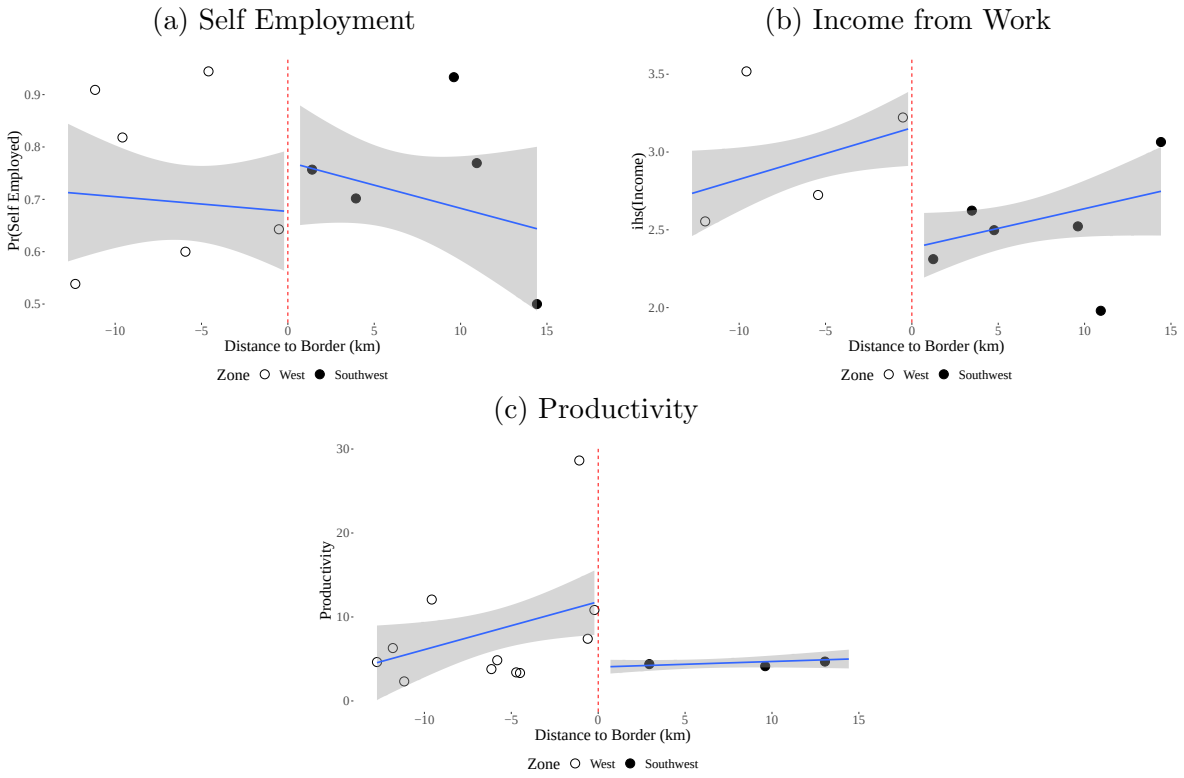
Figure E.6: One-Dimensional RD Plots: Rural DHS Wealth



Notes: RD plots illustrating local linear regressions within a 10 kilometer bandwidth. Vertical dashed line marks 0; observations to the right are in the Southwest zone and observations to the left of the line are in the West zone. Dots represent binned averages.

E.9 One-Dimensional RD Plots: Labor Markets

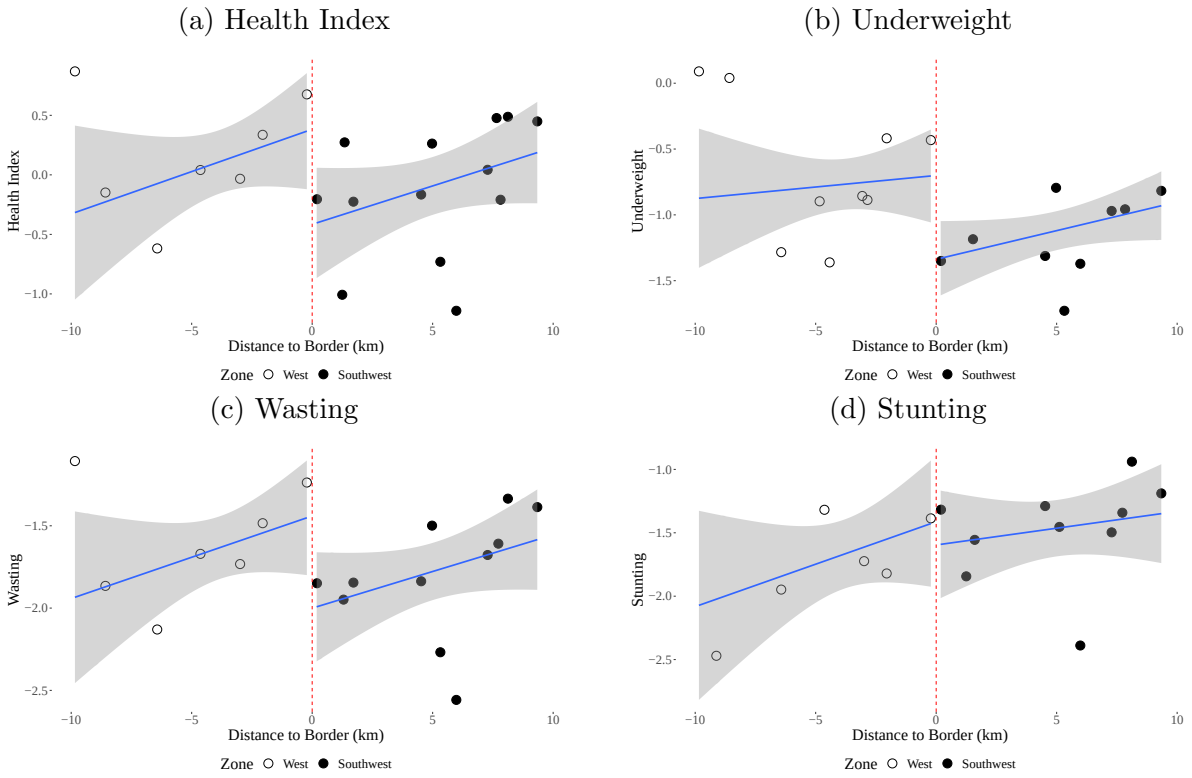
Figure E.7: Labor Force Survey (LFS) Labor Market Outcomes



Notes: RD plots illustrating local linear regressions within a 10 kilometer bandwidth. Vertical dashed line marks 0; observations to the right are in the Southwest zone and observations to the left of the line are in the West zone. Dots represent binned averages.

E.10 One-Dimensional RD Plots: Health

Figure E.8: DHS Rural Child Health



Notes: RD plots illustrating local linear regressions within a 10 kilometer bandwidth. Vertical dashed line marks 0; observations to the right are in the Southwest zone and observations to the left of the line are in the West zone. Dots represent binned averages.

F Full Model Results

Table F.1: Full Model Results: Baseline Development Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	%Poverty				IHS Luminosity			
$\mathbb{1}$ SW	4.53**	4.43**	5.81**	4.96**	−0.68***	−0.69***	−0.48*	−0.64***
	(1.66)	(1.68)	(1.91)	(1.88)	(0.16)	(0.16)	(0.23)	(0.19)
Border	−1.95***	−2.22***	−3.79***	−3.51***	0.14***	0.17***	0.16*	0.27***
	(0.38)	(0.39)	(0.70)	(0.76)	(0.02)	(0.02)	(0.07)	(0.04)
SW × Border	3.63***	3.77***	6.50***	6.09***	−0.19***	−0.23***	−0.35***	−0.41***
	(0.52)	(0.53)	(0.96)	(0.97)	(0.03)	(0.03)	(0.10)	(0.06)
Ln(Dist. Cap.)		1.27***		1.20***		−0.02		−0.02
		(0.28)		(0.26)		(0.03)		(0.02)
Border^2			−0.25***	−0.22**			0.00	0.01***
			(0.07)	(0.08)			(0.01)	(0.00)
SW × Border^2			0.05	0.01			0.01	−0.00
			(0.09)	(0.10)			(0.01)	(0.00)
(Intercept)	11.05***		8.99***		1.27***		1.28***	
	(1.11)		(1.29)		(0.12)		(0.16)	
N. Villages	334	324	502	484	422	389	452	568
Segment FE	-	✓	-	✓	-	✓	-	✓

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Results from Table 2 reproduced to show partial derivatives of adjusting covariates.

Table F.2: Full Model Results: Age Cohort Analysis

	Years of Schooling
Age	-61.87^{***}
	(13.57)
Age^2	-20.64^{*}
	(8.02)
Wave 2001	-1.05^{***}
	(0.13)
Female	-1.67^{***}
	(0.14)
SW × Age [-10, -5] in 1975	−0.17
	(0.53)
SW × Age [-5, 0] in 1975	−0.13
	(0.58)
SW × Age [0,4] in 1975	−0.50
	(0.54)
SW × Age [4,9] in 1975	−0.24
	(0.62)
SW × Age [10,14] in 1975	-1.32^{***}
	(0.36)
SW × Age [15,20] in 1975	-1.21^{**}
	(0.46)
SW × Age [21,25] in 1975	−0.64
	(0.61)
SW × Age [26, 30] in 1975	−0.38
	(0.51)
SW × Age [31,35] in 1975	−0.14
	(0.93)
SW × Age [36,40] in 1975	0.37
	(1.17)
N. Individuals	2285
N. Villages	71
Decade by Commune FE	Yes

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Results from Figure 5b reproduced to show partial derivatives of adjusting covariates.

Table F.3: Full Model Results: Human Capital: Education and Literacy in 1998

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	%No Educ.				Lit. Rate			
𝟙 SW	7.58**	3.49	7.39**	4.46	−7.89**	−4.85	−7.85*	−5.46
	(2.47)	(2.22)	(2.66)	(2.53)	(2.66)	(2.63)	(3.09)	(2.90)
Border	−2.67***	−2.63***	−3.27***	−4.18***	3.02***	3.34***	3.83**	4.69***
	(0.60)	(0.51)	(0.78)	(0.97)	(0.65)	(0.69)	(1.18)	(1.18)
𝟙 SW × Border	3.00***	3.38***	3.64**	4.83***	−3.56***	−4.23***	−4.13*	−5.16**
	(0.84)	(0.72)	(1.10)	(1.35)	(0.91)	(0.96)	(1.65)	(1.68)
Ln(Cap.)		1.63***		1.46***		−1.45**		−1.24**
		(0.42)		(0.38)		(0.51)		(0.42)
Segment 1		22.32***		24.69***		−20.69***		−25.20***
		(4.20)		(3.35)		(5.12)		(3.73)
Segment 2		7.53***		8.79***		−6.39***		−5.68***
		(1.56)		(1.37)		(1.82)		(1.57)
Segment 3		0.80		0.94		−0.78		0.46
		(1.63)		(1.36)		(1.89)		(1.54)
Border^2			−0.16**	−0.29**			0.24	0.35**
			(0.06)	(0.10)			(0.12)	(0.13)
SW × Border^2			0.22*	0.35*			−0.35*	−0.48*
			(0.09)	(0.14)			(0.17)	(0.19)
(Intercept)	34.60***	19.83***	34.44***	19.70***	80.04***	93.76***	80.15***	92.13***
	(1.77)	(4.04)	(1.91)	(3.84)	(1.90)	(4.86)	(2.20)	(4.30)
N. Villages	334	324	502	484	422	389	452	568
Segment FE	-	✓	-	✓	-	✓	-	✓

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Results from Table 3 reproduced to show partial derivatives of adjusting covariates.

Table F.4: Labor Market Effects of Repression (Full Model Results)

	(1) Pr(Self Employed)	(2) IHS(Income)	(3) Productivity
$\mathbb{1}$ SW	0.12	-0.68^{**}	-8.67^{*}
	(0.07)	(0.24)	(4.17)
Border	-0.04^{***}	0.07	0.59
	(0.01)	(0.04)	(0.33)
Age	5.43^{***}	-4.27	-13.22
	(0.77)	(2.24)	(8.59)
Age ^{2}	-2.54^{***}	1.45	15.34^{*}
	(0.69)	(1.51)	(7.01)
Female	0.22^{**}	-0.06	1.33
	(0.07)	(0.18)	(1.06)
2000 Wave	0.11^{**}	-0.40^{*}	1.75
	(0.04)	(0.18)	(2.71)
SW $\times$ Border	0.04	-0.07	-0.51
	(0.02)	(0.04)	(0.31)
(Intercept)	0.42^{***}	3.46^{***}	10.77^{***}
	(0.03)	(0.18)	(2.08)
Effective N	235	285	411
Covariates	✓	✓	✓
SD DV	0.45	0.86	8.81

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: Table 4 results from main text including the estimates and uncertainty for adjusting covariates.

Table F.5: Intergenerational Effects: Child Health Between Zones (Full Model Results)

Outcome	(1) Health Index	(2) Height/Age	(3) Weight/Age	(4) Weight/Height
$\mathbb{1}$ SW	-0.88^{***}	0.10	-0.59^{**}	-1.11^{***}
	(0.26)	(0.17)	(0.19)	(0.16)
Dist. Border	0.00^{*}	-0.00	0.00^{*}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
SW × Border	-0.00	-0.00	-0.00	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.14	-0.26	-0.11	0.07
	(0.17)	(0.16)	(0.12)	(0.11)
Age ^{2}	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
N. Individuals	243	298	243	195
N. Clusters	29	36	29	23
Bandwidth	11.05	11.55	11.21	9.34

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Note: Table 5 results from main text including the estimates and uncertainty for adjusting covariates.

Table F.6: Baseline with Additional Covariates: Full Model Results

Outcome	(1)	(2)	(3)	(4)
	%Poverty		IHS Luminosity	
Intercept	5.012	3.039	1.439	1.526
	(3.081)	(3.157)	(0.218)	(0.215)
SW	3.213	4.361	-0.525	-0.527
	(1.712)	(2.027)	(0.12)	(0.134)
Dist. Border	-1.558	-3.172	0.127	0.189
	(0.375)	(0.708)	(0.017)	(0.036)
Segment1	-2.084	0.226	-0.627	-0.533
	(3.493)	(2.857)	(0.266)	(0.201)
Segment2	-6.941	-8.222	-0.429	-0.425
	(2.78)	(2.298)	(0.194)	(0.137)
Segment3	-5.939	-6.262	-0.496	-0.455
	(1.189)	(1.046)	(0.089)	(0.072)
Segment4	-0.14	0.5	-0.332	-0.319
	(1.225)	(1.077)	(0.088)	(0.075)
Dist.Capital	1.17	1.143	-0.011	-0.015
	(0.297)	(0.29)	(0.023)	(0.021)
Road	-0.001	0	0	0
	(0.001)	(0)	(0)	(0)
Built Area	-0.15	-0.141	0.018	0.019
	(0.046)	(0.048)	(0.004)	(0.004)
SW x Dist.Border	2.804	-0.21	-0.179	0.008
	(0.522)	(0.064)	(0.025)	(0.003)
Dist.Border Sq.	5.36		-0.286	
	(0.981)		(0.05)	
SW x Dist.Border Sq.	0.037		-0.003	
	(0.086)		(0.004)	
Effective N	340	505	439	618
Bandwidth	6.58	11.12	9.31	14.37
μ Control	20.95	20.95	0.43	0.43
σ DV	10.55	10.55	0.63	0.63

Table F.7: Human Capital in 1998: Years of Schooling and Attendance

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Yrs Educ.				Attendance Rate			
Intercept	6.983	7.154	9.912	9.529	42.712	44.796	59.347	58.851
	(0.295)	(0.308)	(0.722)	(0.58)	(1.542)	(1.742)	(3.93)	(3.555)
SW	-1.105	-1.103	-0.581	-0.691	-6.023	-7.785	-4.639	-3.418
	(0.413)	(0.432)	(0.388)	(0.382)	(2.146)	(2.428)	(2.155)	(2.29)
Dist.Border	0.001	0.001	0.001	0.001	0.001	0.003	0.002	0.002
	(0)	(0)	(0)	(0)	(0)	(0.001)	(0.001)	(0.001)
Dist Border x SW	-0.001	0	-3.051	-3.06	-0.002	0	-19.47	-19.515
	(0)	(0)	(0.75)	(0.506)	(0.001)	(0)	(4.051)	(3.154)
Dist.Border^2		-0.001	-2.123	-1.907		-0.004	-7.551	-9.337
		(0)	(0.27)	(0.206)		(0.001)	(1.513)	(1.23)
Dist.Border^2 x SW		0	-0.916	-0.573		0	-2.792	-3.326
		(0)	(0.281)	(0.206)		(0)	(1.571)	(1.264)
Dist.Cap			-0.253	-0.207			-1.647	-1.559
			(0.075)	(0.057)			(0.405)	(0.357)
Dist. School			-0.001	0			-0.002	0
Effective N	312	285	442	439	355	313	597	476
Bandwidth	5.78	4.85	9.87	9.66	7.13	5.85	14.27	10.78
μ Control	4.16	4.16	4.16	4.16	30.07	30.07	30.07	30.07
σ DV	1.9	1.9	1.9	1.9	12.8	12.8	12.8	12.8

Table F.8: Full Model Results: Table D7

Outcome	(1) Health Index	(2) Height/Age	(3) Weight/Age	(4) Weight/Height
	Panel A: Mothers with No Education			
SW	-2.96^{**}	-1.70	-1.74^{**}	-3.31^{***}
	(0.93)	(1.65)	(0.53)	(0.38)
Dist. Border	0.00^{*}	0.00	0.00^{*}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
Dist. Border x SW	-0.00	-0.00	-0.00	-0.00^{**}
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.31	-0.46	-0.18	-0.16
	(0.39)	(0.50)	(0.24)	(0.25)
Age^2	0.00	0.01	0.00	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Num. obs.	65	45	73	46
N. Clusters	25	17	27	18
	Panel B: Mothers with Education			
SW	-0.30	0.14	-0.20	-0.84^{***}
	(0.40)	(0.22)	(0.29)	(0.20)
Dist. Border	0.00	-0.00	0.00	0.00^{**}
	(0.00)	(0.00)	(0.00)	(0.00)
Dist. Border x SW	-0.00	0.00	-0.00	-0.00^{*}
	(0.00)	(0.00)	(0.00)	(0.00)
Age	-0.14	-0.22	-0.10	0.03
	(0.16)	(0.19)	(0.11)	(0.12)
Age^2	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Num. obs.	233	170	233	170
N. Clusters	38	26	38	26

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table F.9: NTL at Alternative Bandwidths: Full Model Results

	4km	1km
Intercept	2.254	1.058
	(0.234)	(0.167)
SW	-0.763	-0.367
	(0.131)	(0.093)
Dist Border	0.184	0.091
	(0.017)	(0.012)
Segment 1	-0.996	-0.388
	(0.284)	(0.202)
Segment 2	-0.73	-0.211
	(0.213)	(0.152)
Segment 3	-0.775	-0.375
	(0.099)	(0.07)
Segment 4	-0.47	-0.238
	(0.098)	(0.07)
Dist. Cap	-0.018	-0.012
	(0.025)	(0.018)
SW X Dist.Border	-0.255	-0.136
	(0.025)	(0.018)

Table F.10: Full Model Results: Figure C1

	(1)	(2)	(3)	(4)	(5)
Age	-56.50^{***}	-47.95^{***}	-35.14^{**}	-23.37	-17.71
	(12.19)	(11.47)	(12.29)	(12.39)	(12.19)
Age^2	-10.71	-4.79	-5.34	-6.58	-7.56
	(6.54)	(6.41)	(6.33)	(5.67)	(5.55)
Female	-1.49^{***}	-1.40^{***}	-1.34^{***}	-1.28^{***}	-1.27^{***}
	(0.11)	(0.11)	(0.12)	(0.13)	(0.15)
Wave 2001	0.06	0.08	0.36^{*}	0.01	0.58^{***}
	(0.11)	(0.14)	(0.14)	(0.10)	(0.07)
SW x Age [-10,-5]	-0.30	-0.06	0.32	1.10^{*}	1.37^{**}
	(0.47)	(0.46)	(0.48)	(0.47)	(0.44)
SW x Age [-5,0]	-0.13	0.18	0.49	1.15^{*}	1.35^{**}
	(0.49)	(0.48)	(0.49)	(0.49)	(0.46)
SW x Age [0,4]	-0.62	-0.23	0.05	0.53	0.68
	(0.52)	(0.51)	(0.51)	(0.52)	(0.51)
SW x Age [4,9]	-0.33	0.03	0.30	0.66	0.79
	(0.60)	(0.59)	(0.58)	(0.59)	(0.60)
SW x Age [10,14]	-1.32^{***}	-1.10^{**}	-0.96^{**}	-0.78^{*}	-0.69
	(0.35)	(0.34)	(0.34)	(0.35)	(0.36)
SW x Age [14,20]	-1.16^{*}	-1.14^{*}	-1.07^{*}	-1.07^{*}	-1.02^{*}
	(0.44)	(0.44)	(0.44)	(0.44)	(0.44)
SW x Age [20,25]	-0.63	-0.43	-0.38	-0.27	-0.25
	(0.51)	(0.49)	(0.50)	(0.47)	(0.49)
SW x Age [26,30]	-0.61	-0.62	-0.70	-0.67	-0.71
	(0.43)	(0.42)	(0.42)	(0.38)	(0.38)
SW x Age [31,35]	-0.24	-0.12	0.10	0.05	-0.14
	(0.74)	(0.70)	(0.68)	(0.68)	(0.66)
SW x Age [36,40]	0.13	0.01	0.13	-0.05	-0.19
	(1.03)	(1.02)	(0.98)	(0.98)	(0.98)
Num. obs.	2174	2003	1898	1657	1528
N. Villages	71	71	71	71	71

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table F.11: Educational Persistence: Full Model Results

	Full Sample	Dropping Positive Outliers
SW	-12.16^{***}	-2.05^{*}
	(2.39)	(0.88)
Dist.Border	3.38^{***}	0.71^{***}
	(0.32)	(0.11)
Dist.Capital	-2.20^{***}	-0.68^{***}
	(0.47)	(0.18)
SW x Dist. Border	-4.68^{***}	-1.03^{***}
	(0.46)	(0.16)
N. Villages	462	366

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

G Trimming Exercise Explanation: Dataverse Only

Migration immediately after - and during - the Khmer Rouge is an important factor to consider. The Khmer Rouge forced urbanities to move across the country. Within rural areas, villagers were forced into mobile work teams, which traveled around sectors (ie, within the administrative area directly below the zone level.) Since I investigate the legacy of the regime among villages within two sectors, the rural areas I focus may have moved during the regime, but would have moved within the geographic area of their treatment assignment (within the West or Southwest zones).

That being said, some degree of internal migration could explain the result giving movement that occurred due to displacement or flight. The choice to move occurs post-treatment for many residents: after the Khmer Rouge administered repression and strictly controlled internal migration, some citizens may have sorted after the regime in ways correlated with their treatment status.

Since movement occurs post-treatment, to estimate the impact of the Southwest Zone among residents who never moved, I adopt a principal stratification framework. Consider migration $\mathcal{M} = \{0, 1\}$ where 1 means an individual has migrated (ie, lives in a place that was not their original place of birth) and 0 means the individual is living in the place they have always lived and Z $= \{0, 1\}$ where 1 denotes being exposed to the (former) Southwest Zone, which I refer to as "treatment" for brevity. Finally, a respondent's residence $R_i = \{0, 1\}$ is 1 when the person lives in the former Southwest Zone and 0 otherwise. I follow the method outlined in (Marbach, 2021) to estimate the effect of the SW zone net of migration.

As in Marbach (2021) we can define 4 principal strata based on these criteria using potential outcomes notation. Let $\mathcal{M}(0)$ denote migration status under control ($Z = 0$) and

$\mathcal{M}(1)$ denote migration status when assigned to the Southwest Zone ($Z = 1$).

Table G.1: Principal Strata

Types	Exposure and (Potential) Migration
Always Moves (AM)	$\mathcal{M}(0) = \mathcal{M}(1) = 1$
Never Moves (NM)	$\mathcal{M}(0) = \mathcal{M}(1) = 0$
Moves if Treated (MT)	$\mathcal{M}(0) = 0, \mathcal{M}(1) = 1$
Stay if Treated (ST)	$\mathcal{M}(0) = 1, \mathcal{M}(1) = 0$

The four strata ($S = \{AM, NM, MT, ST\}$) and labels for their types are presented in Table G.1. Qualitatively, these types are as follows:

- Always moves (AM): migrate regardless of treatment status; individuals who would have been mobile irrespective of the intensive margin of repression they were exposed to. These individuals may have preferences, income, or social ties that would have pushed them to move no matter what the Khmer Rouge did.

- Never Moves (NM): never migrate, whether repressed more or less. These individuals may have strong social ties to their place of residence, or may face very high transaction costs for moving (i.e., the entire immediate family would also need to move, lack of a car makes transportation of goods difficult, fixed assets that could be sold may be insufficient to cover for expenses, barriers to entry to new labor markets, ect).

- Moves if Treated (MT): People that migrate, but were only pushed because of Khmer Rouge repression. These persons may have gone from the Southwest zone to the nearby West zone after repression to place themselves in a more ideal labor market.

- Stays if Treated (ST): these are persons who had plans to move, or would have migrated, but did not because the Khmer Rouge shocked them into place. This could occur if income shocks were so large that assets that were going to go towards migrating had to allocated towards other expenses. Given low rural-rural migration rates and strong traditional ties to place, this subgroup is likely to be fairly small in practice.

From here, we can break down different movement types into several types displayed in Table G.2.

Table G.2: Individual Respondents Stratified by Exposure Status (Z) and Current Residency (R)

		Zone Historically	
		0	1
Residency	0	MT & NM	MT & AM
Today	1	ST & AM	ST & NM

Assumption 1: No "Stays if Treated" (Monotonicity) $\mathcal{M}_i(1) - \mathcal{M}_i(0) \geq 0 \ \forall \ i = \{1, ..., N\}$**.**

Assumption 1 says there is no type that only wants to live in the Southwest if they are exposed to mass repression. The assumption is substantively motivated and likely to be satisfied in this context. Those that wanted to leave Cambodia often did, as evidenced by the mass migration outflows from the country in the lead up to the Khmer Rouge, and the devastation wrought by violence is not likely to be a pull factor persuading otherwise mobile civilians to stay after treatment. Likewise, it makes little sense to think of a person who was satisfied staying in their home village, and was only promoted to move into the Southwest zone because they were assigned *less* repression. As such, I assume individuals will either

always move regardless of repression due to factors uncorrelated with treatment, will never move, again for reasons unaffected by treatment status, or will move if repressed.

Selective migration could explain the result of the following migration patterns occurred after the Khmer Rouge: suppose MT's left the Southwest Zone due to repression and poverty, and sought to relocate somewhere nearby that had impacted less by violence. They would likely pick the parts of Kampong Speu within the West Zone, since they have preexisting ties to the province. This could explain the result if the following assumption is true.

In support of Assumption 1, I report the rates of international migration by commune in both the West and Southwest. Table C.3 shows international migration rates are similar between zones. This evidence suggests that pull factors that may attract migrants are relatively equal by area, suggesting there is nothing about having been in the Southwest zone that would lead persons who would otherwise not move to relocate to the zone.

H Background Information: Dataverse Only

H.1 Zone Commanders

Mok (Chhit Choeun) was native to Takeo province in southwestern Cambodia, where he had helped lead an insurgency during the independence era, and at one point served under Sy before becoming a commander himself. Mok's rise to power placed him in direct competition with Sy, an older leftist who harbored less extreme views about the direction of the revolution. Mok promoted his family members - including his sons, daughters, and in-laws - in positions of power throughout the Southwest (Kiernan, 2008, p. 87). Mok was charged with purging party elites who were disloyal to the regime. He survived the regime, assisting in military leadership of the Khmer Rouge insurgency after Vietnam deposed the regime.

Sy (Chou Chet) had been a politically active leftist since the 1940s, participating in independence efforts against the French (Vickery, 1984, p. 129). Sy had been the secretary of the entire Southwest during the civil war (which included the Southwest and West during the DK regime) but was demoted in 1973 in favor of Mok (Kiernan, 1989).

Sy had pro-NVA opinions (Kiernan, 2008, p. 79) which had set him apart from more nationalist hardliners closer in Pol Pot's circle. During a CPK meeting in Kampong Speu, Sy's address stood out to audience members as he did not refer to the Vietnamese as enemies and spoke of providing housing for individual families (Kiernan, 2008, p. 390). Due to the Khmer Rouge's general hostility towards Vietnam, the remarks highlighted a contrast between the ideology of Chet and the rest of DK.

Chandler (2018) avers Sy revealed an earnest assessment of the Khmer Rouge's policy failures in the autobiography DK cadres forced him to concoct as a confession while jailed

in S-21.

> [I said that] the current regime was a highly dictatorial one, too rigid and severe, one that overshot the comprehension and consciousness of the people. Therefore a lot of people were muttering and moaning about how they were doing a lot of work and getting little back for it, how they couldn't get together with their families, couldn't rest, never had any fun, and so on.

Although the passage is presented as a dialogue that was almost certainly fictitious, as the party forced the biography to be a confession of crimes including collaboration with the CIA and the Vietnamese, the remark is consistent with what was reported about Sy: he was a more moderate leader who was skeptical of the regime's brutal approach.

H.2 Educational System Before, During, and After DK

Cambodia inherited its contemporary educational system from French colonization (Ayres, 2000). The Cambodian educational system was expanding prior to the onset of the civil war and the Khmer Rouge regime. The national government dedicated a large share of the budget to education, and in 1969 the country boasted "3202 primary schools, 163 secondary schools, and nine universities" (Ayres, 1999).

When the Khmer Rouge took power, traditional teachers were targeted for purges. 90% of schools were destroyed (Clayton, 1998). Whereas the regime attempted to provide education, schooling was subordinate to labor and occurred in irregular settings, such as fields and stables. Teachers were not "new people," the regime instead favored "base people" as educators, who had little to no experience. Students lacked supplies due to general resource shortages during the DK period (Ayres, 1999). Within a year of the Khmer Rouge's downfall, school reopening was attempted. Vietnam focused on quantity rather than quality, which

failed to address the dilapidated school infrastructure or limited teaching corps (Ayres, 2000, p. 148).

I Genocide Intensity: Dataverse Only

This section of the appendix presents indirect evidence that exposure to increased genocide intensity is correlated with contemporary development. I leverage data on the location of mass graves in Kampong Speu to construct a gravity-based measure of exposure to genocide. Then, I estimate the relationship between development (poverty and luminosity) and the intensity of genocide exposure instrumented by zone.

I.1 Measuring Genocide Intensity

The data on mass graves is from the Cambodia Genocide Program Geographic Database. The data was collected by a team of researchers who, through interviews with locals and archival documents, excavated mass graves during the DK era. The data include an estimated 1 million bodies and over 300 mass grave locations, with 16 of those locations residing in Kampong Speu. The Khmer Rouge era included deaths from other events, including a mass bombing campaign during the civil war. However, traditional Cambodian practices included cremation at the time, and many of those killed in bombings would have been incinerated. Cremation and other traditional practices were banned during the DK regime, meaning mass graves filled with bodies are most likely attributable to Khmer Rouge activity. Forensic evidence breaking down who in the graves were victims of execution versus other causes of death, such as starvation, is unavailable. Nonetheless, starvation was a DK tactic

to cheaply eliminate political rivals - to the extent the data include both sets of deaths, it may still capture the degree of DK brutality (Etcheson, 2000).

Mass grave sites were almost always constructed near security centers - indeed, most are within a kilometer of a security center. Security centers were managed at the zone level, meaning the head of the each respective zone would have retained authority over who to send to these centers and what to do with them, including torture, re-education, release, or execution (Etcheson, 2000).

One concern with attributing a village's geographic proximity to mass grave sites to genocide intensity is population transfers that occurred during the DK regime. One may reasonably object that a village near a mass grave with hundreds of bodies may have had few residents of that village buried in the grave, since people were forced to migrate internally during the regime to perform forced labor. As such, my measure may contain some error.

A key advantage of my approach is that I study a very localized instance of the genocide, which allows me to circumvent this empirical problem. The Southwest zone was subject to a mass population transfer, but it was an outflow of residents towards the Northwest zone, which was considered the breadbasket of the regime (Kiernan, 2008, p. 390). To the extent the mass graves may measure genocide exposure with error, it is likely an underrepresentation in the Southwest zone, since residents forced to move to the Northwest zone may have died there instead of the Southwest.

Nonetheless, population movements occurred within zones overtime. The Khmer Rouge formed mobile work teams who would perform tasks within their zone to construct irrigation canals. A resident may therefore be buried within their zone, but could have been executed near a village that they were not from, and very far from their village of birth.

To address this issue, I adopt a gravity based measure of genocide exposure. The intuition behind the measure is that villages that are closer to more large mass graves were likely more exposed to mass killing themselves. The assumption is reasonable, since a villager living near a security center who was accused of some sort of wrongdoing could be sent to a security center to be executed at a lower cost compared to a village who was far from a prison.

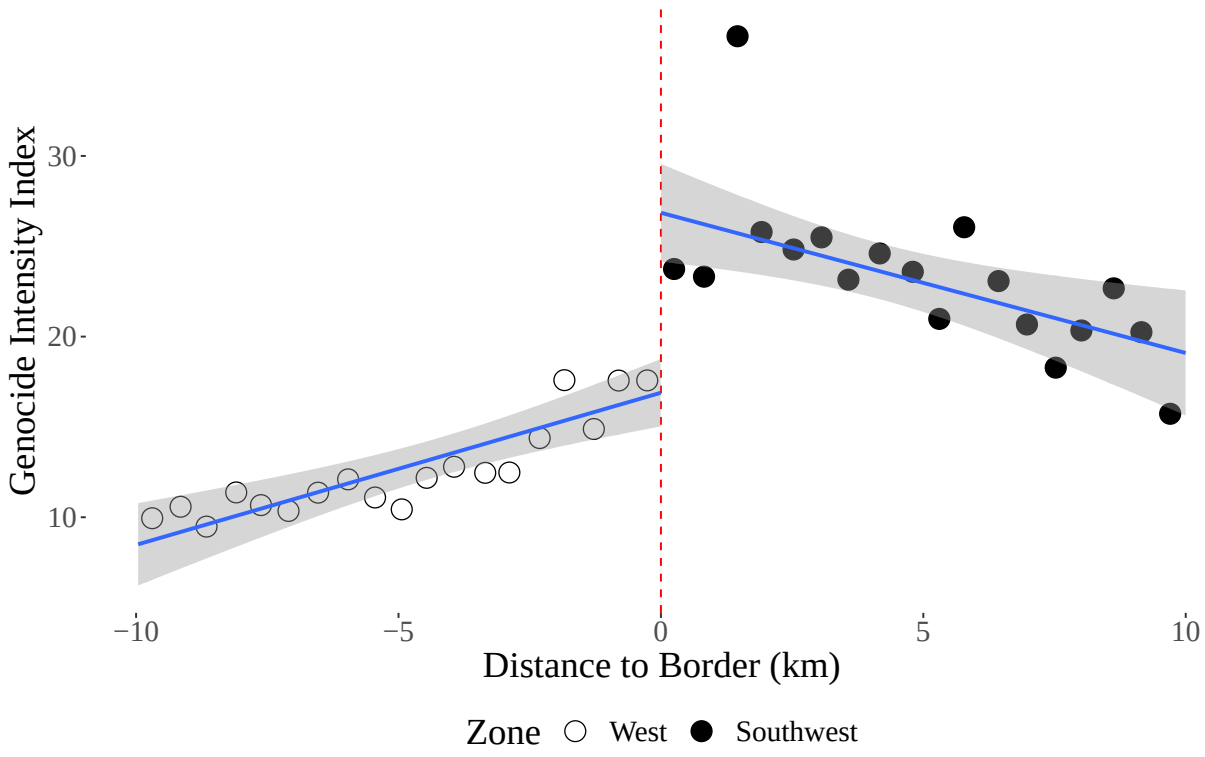
$$\text{Genocide Intensity}_v = \sum_{j \in J_z} \left(1 + distance(\text{Village}_v, \text{Mass Grave}_j)^{-1} \times \frac{\text{Bodies}_j}{\sum_{j \in J_z} \text{Bodies}_j} \right)$$

First, I compute the distance between a village v and each mass grave j within the province/zone z. I add 1 to prevent very proximate mass graves from having excessive influence, and take the inverse of the quantity, so larger values represent being closer to a grave and smaller values representing more distance. Second, for each grave, I multiple this quantity by a weight which equals the proportion of total bodies in zone grave j holds. Third, I sum these values together for each mass grave to obtain a final measure of genocide intensity for each village. Finally, to ease interpretation, I scale the variable from 0-100 as follows.

$$\left(\frac{\text{Genocide Intensity}_v - min(\text{Genocide Intensity})}{max(\text{Genocide Intensity}) - min(\text{Genocide Intensity})} \right) \times 100$$

The measure represents a village's exposure to genocidal violence at the intensive rather than the extensive margin - we cannot gleam from the measure whether a village had executions or not, but it does capture the intensity of mass killing surrounding the village. While imprecise, population transfers outside of the Southwest suggest the measure will be biased against the hypothesized positive effect of the Southwest zone on genocide intensity.

Figure I.1: Genocide Intensity in Kampong Speu Province

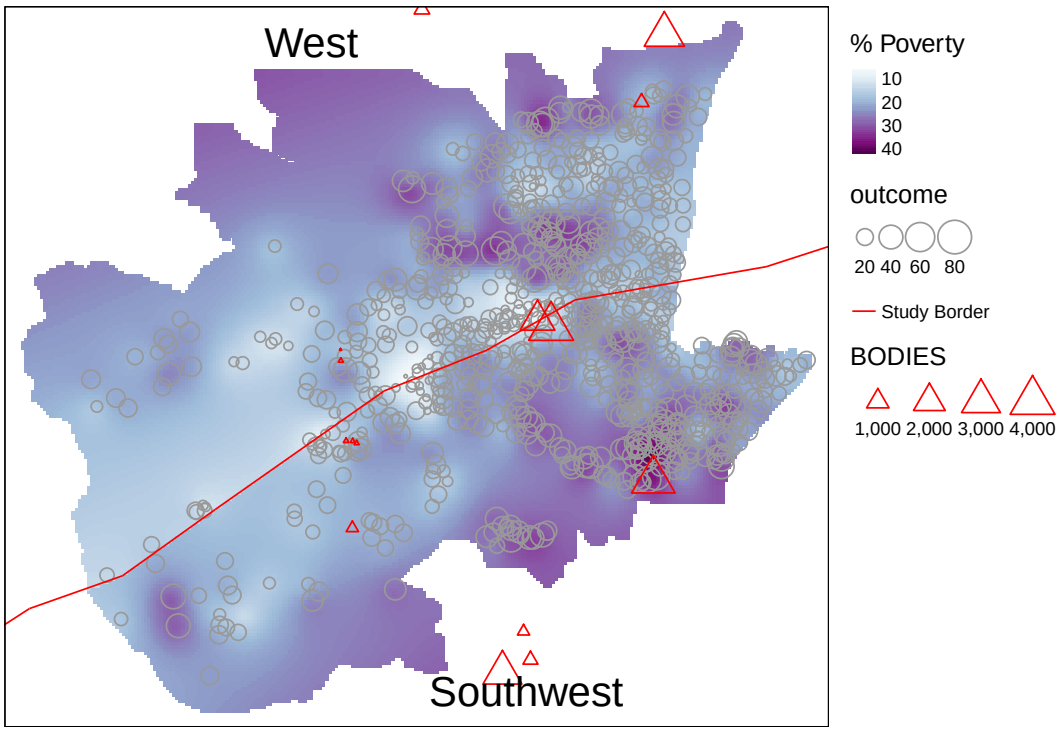


Another bias concern may be that the location of mass graves covary with distance to the capital. To partial out the distance to the capital, I adjust for this covariate in my estimates.

Figure I.1 plots the discrete increase in genocide intensity moving from the West to the Southwest zone.

Figure I.2 displays predicted poverty across space, actual poverty rates in 2011 at the village level, and the location and size of mass grave sites.

Figure I.2: Poverty and Mass Repression in Kampong Speu Province



Note: Dots are village locations, shaded regions are predicted values using ordinary kriging, Red line shows the border dividing the Southwest and West. Boxes with triangles are mass graves, sized to represent the estimated number of bodies found within them.

I.2 Estimating Genocide Intensity

One approach to unpacking the effect of genocide intensity and development would be regressing development directly on the measure of exposure. However, this will likely yield biased estimates. The location of mass graves was not randomly assigned; the Khmer Rouge disproportionately executed educated persons and former officials. This suggests that proximity to mass violence would be positively correlated with past development. If it was the case that genocide has a negative effect on development, and that places which were exposed more to genocide were more developed beforehand, one may recover a positive or zero coefficient from regressing development on repression due to the omission of unobservable, pre-DK factors.

To estimate the first order effect of zone assignment on genocide intensity, along with the influence of genocide intensity on development, I use the following two-staged least squares (instrumental variable) approach.

$$\text{Genocide Intensity}_{cv} = \alpha_c + \delta \mathbb{1}(\text{Southwest Zone}_{cv)} + \sum_{k=1}^{K} \zeta^k X_v^k + \varepsilon_{cv}$$

$$out_{cv} = \alpha_c + \beta \widehat{\text{Genocide Intensity}}_{cv} + \sum_{k=1}^{K} \zeta^k X_v^k + \eta_{cv}$$

The first stage estimates the impact of zone assignment on genocide intensity. The parameter δ captures the average difference between genocide intensity between the Southwest and the West, net of the K covariates X (distance to Chbar Mon and Phnom Penh). I include fixed effects for commune (α_c) which absorbs differences between administrative units and compares villages in a similar environment.

The second stage uses the predicted increase in genocide intensity to estimate the relationship between increased concentration of violence around a village and development. For both equations, standard errors are clustered at the commune, of which there are 87, to account for serial correlation induced by the fixed effects.

The key identifying assumption is zone assignment does not effect development through any channel outside of its effect on genocide intensity, after conditioning on covariates. While this assumption cannot be proven, it is reasonable in this context. As discussed in the main text, there is a dearth of qualitative evidence suggesting the zones differed along non-repression based dimensions.

I.3 Genocide Intensity Results

Results reported in Table I.1 show villages in the Southwest zone experience a standard deviation increase in genocide intensity. The result corresponds with qualitative accounts that the Southwest zone was the "toughest sanctuary of the Khmer Rouge movement."

The 2SLS results show a similar pattern. Remarkably, a standard deviation increase in genocide intensity corresponds with a 3.9% increase in poverty, which is fairly close to the RD benchmark of 4.53% in the main text. The luminosity results trend in the same direction, albeit at a smaller magnitude than the RD results. Overall, the evidence suggests villages exposed more to the genocide by virtue of being in the Southwest zone are less developed today.

The results provide evidence consistent with the account that genocide exposure from the Southwest zone reduced contemporary economic development. If it was the case that the Southwest zone indicator was not correlated with genocide intensity, and that instru-

Table I.1: Genocide Intensity Results

	Genocide Intensity	Household Poverty	Village Luminosity
$\mathbb{1}$(Southwest)	12.84***		
	(2.39)		
$\widehat{\text{Genocide}}$		0.33**	−0.04***
		(0.12)	(0.01)
ln(Distance to Chbar Mon + 1)	0.79	−1.62	−0.01
	(2.24)	(1.39)	(0.14)
ln(Distance Phnom Penh)	11.84	−0.84	−0.22
	(6.18)	(3.64)	(0.28)
N. Villages	1359	1359	1359
N. Clusters	87	87	87
Estimator	First Stage	2SLS	2SLS
SD DV	12.04	10.55	0.76
Commune FE	✓	✓	✓
Covariates	✓	✓	✓

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Models include commune fixed effects and covariatates adjusting for distance to Phnom Penh and Chbar Mon. Robust standard errors clustered at the commune reported in parentheses.

mented genocide intensity did not predict development, the argument in the paper would be implausible. Nonetheless, I caution that the result does not establish with certainty that repression directly causes poverty.

References

Ayres, David. 1999. "The Khmer Rouge and education: beyond the discourse of destruction." *History of Education* 28(2):205–218.

Ayres, David M. 2000. *Anatomy of a crisis: Education, development, and the state in Cambodia, 1953-1998*. University of Hawaii Press.

Chandler, David P. 2018. *Brother number one: A political biography of Pol Pot*. Routledge.

Clayton, Thomas. 1998. "Building the new Cambodia: Educational destruction and construction under the Khmer Rouge, 1975-1979." *History of Education Quarterly* 38(1):1–16.

De Walque, Damien. 2006. "The socio-demographic legacy of the Khmer Rouge period in Cambodia." *Population studies* 60(2):223–231.

Etcheson, Craig. 2000. "Mapping Project 1999: The Analysis." the Number": Quantifying Crimes Against Humanity in Cambodia.".

Farr, Tom G, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth et al. 2007. "The shuttle radar topography mission." *Reviews of geophysics* 45(2).

Fick, Stephen E and Robert J Hijmans. 2017. "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas." *International journal of climatology* 37(12):4302–4315.

Kaba, Mary White, Klaus Baesel, Bunnak Poch, Samnang Bun, Sabine Cerceau, Louise Bury, Bastian Schwarz, Ouly Keo, Rathavy Tung, Kannitha Cheang et al. 2018. "IDPoor: a poverty identification programme that enables collaboration across sectors for maternal and child health in Cambodia." *bmj* 363.

Kiernan, Ben. 1989. "The American Bombardment of Kampuchea, 1969-1973." *Vietnam Generation* 1(1):3.

Kiernan, Ben. 2008. *The Pol Pot Regime: Race, Power, and Genocide in Cambodia Under the Khmer Rouge, 1975-79*. Yale University Press.

Marbach, Moritz. 2021. "Causal Effects, Migration and Legacy Studies.".

Pesaresi, Martino, Daniele Ehrlich, Stefano Ferri, Aneta Florczyk, Sergio Freire, Matina Halkia, Andreea Julea, Thomas Kemper, Pierre Soille, Vasileios Syrris et al. 2016. "Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014." *Publications Office of the European Union* pp. 1–62.

Shaver, Andrew, David B Carter and Tsering Wangyal Shawa. 2019. "Terrain ruggedness and land cover: Improved data for most research designs." *Conflict Management and Peace Science* 36(2):191–218.

Vickery, Michael. 1984. *Cambodia: 1975-1982*. South End Press Boston.