Contents

A Data Appendix A–1

B Counterfactual Spacing B–1

C Size C–1

D Independent Shifts Distributions D–1

E Common Shifts Distributions E–1

F OLS Specification F–1

G Don't Subtract Counterfactual Means G–1

H Common Shifts H–1

I Counterfactuals at Least 5km From All Actual Factories I–1

J No Balance Restriction J–1

K Sample Includes Factories Near Modern Sugar Factories K–1

L Remove Factories Within 10 Km of Residency Capital L–1

M 10 Kilometer Sample M–1

N No Village Restrictions N–1

O No Factory Fixed Effects O–1

P Includes Catchment Indicator P–1

Q Leontief-Weighted Outcomes Q–1

R Education for All Cohorts R–1

S RD Plots S–1

T Additional Public Goods Outcomes T–1

U Crop Choice U–1

V Robustness to Bandwidth Plots V–1

W Single Linear Latitude-Longitude Polynomial W–1

X Quadratic Latitude-Longitude Polynomial X–1

Y One-Dimensional Linear RD Polynomial Y–1

Z 25 Km Boundary Segment Fixed Effects Z–1

AA No Geographic or Pre-Period Controls AA–1

AB Alternative Sample Restriction AB–1

AC Full Sample AC–1

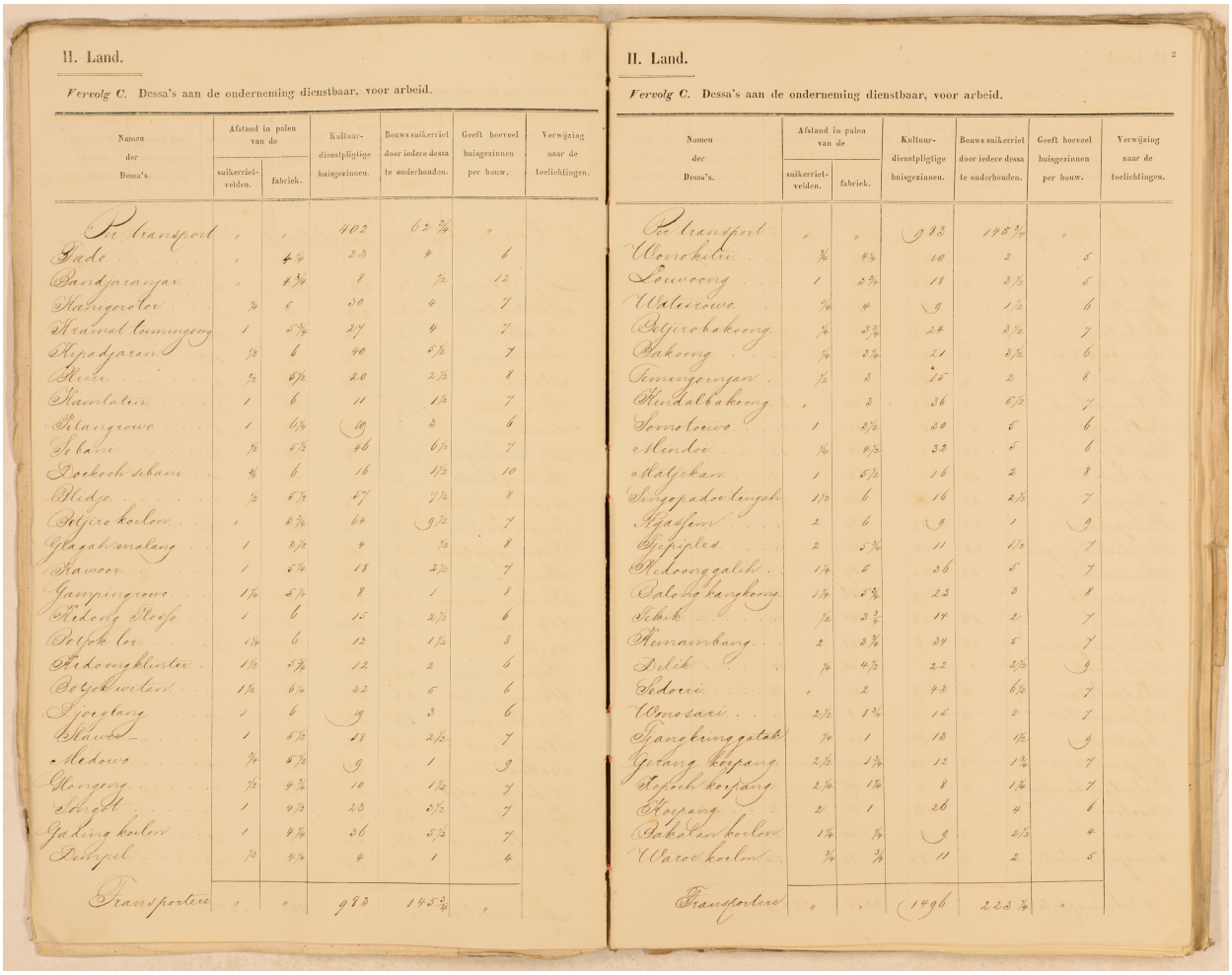
AD No Triangular Kernel Weighting AD–1

AE Instrumental Variables AE–1

AF Grid Cells AF–1

A Data Appendix

Figure A-1: Umbgrove Manuscript Example



II. Land.

Vervolg C. Dessa's aan de onderneming dienstbaar, voor arbeid.

Namen der Dessa's.	Afstand in palen van de suikerrietvelden.	Afstand in palen van de fabriek.	Kultuurdienstpligtige huisgezinnen.	Bouws suikerriet door iedere dessa te onderhouden.	Geeft hoeveel huisgezinnen per bouw.	Verwijzing naar de toelichtingen.
Per transport	"	"	402	62 3/4	"	
Gado	"	4 1/4	23	4	6	
Bandjaranjar	"	4 3/4	8	1/2	12	
Hanigorotor	1/4	5	30	4	7	
Kramat toemengoeng	1	5 3/4	27	4	7	
Hepadjaran	1/2	6	40	5 1/2	7	
Heere	1/2	5 1/2	20	2 1/2	8	
Kamlaten	1	6	11	1 1/2	7	
Pilangrowo	1	6 1/4	19	3	6	
Sebani	1/2	5 1/2	46	6 1/2	7	
Dockoch sebani	3/4	6	16	1 1/2	10	
Blidjo	1/2	5 1/2	57	7 1/2	8	
Betjiro koelon	"	3 3/4	64	9 1/2	7	
Glagah malang	1	3 1/2	4	1/2	8	
Kawoor	1	5 1/4	18	2 1/2	7	
Gampingrowo	1 1/4	5 1/4	8	1	8	
Kedong Flosso	1	6	15	2 1/2	6	
Botjok lor	1 1/4	6	12	1 1/2	8	
Kedoongklinter	1 1/2	5 3/4	12	2	6	
Botjok wetan	1 1/2	6 1/4	32	5	6	
Djoeglang	1	6	19	3	6	
Slawoe	1	5 1/2	18	2 1/2	7	
Medowo	3/4	5 1/2	9	1	9	
Glongong	1/2	4 3/4	10	1 1/2	7	
Songot	1	4 1/2	23	3 1/2	7	
Gading koelon	1	4 1/4	36	5 1/2	7	
Dempel	1/2	4 1/4	4	1	4	
Transportere	"	"	983	145 3/4	"	

II. Land. 2

Vervolg C. Dessa's aan de onderneming dienstbaar, voor arbeid.

Namen der Dessa's.	Afstand in palen van de suikerrietvelden.	Afstand in palen van de fabriek.	Kultuurdienstpligtige huisgezinnen.	Bouws suikerriet door iedere dessa te onderhouden.	Geeft hoeveel huisgezinnen per bouw.	Verwijzing naar de toelichtingen.
Per transport	"	"	983	145 3/4	"	
Wonokitri	3/4	4 1/4	10	2	5	
Loewoong	1	3 3/4	18	3 1/2	5	
Watesrowo	3/4	4	9	1 1/2	6	
Betjirobakoong	1/4	3 3/4	24	3 1/2	7	
Bakoong	1/4	3 1/4	21	3 1/2	6	
Temengoengan	1/2	3	15	2	8	
Kendalbakoong	"	3	36	5 1/2	7	
Somotoewo	1	3 1/2	30	5	6	
Mendoe	1/4	4 1/2	32	5	6	
Matjekan	1	5 1/2	16	2	8	
Singopadoe tengah	1 1/2	6	16	2 1/2	7	
Ngassem	2	6	9	1	9	
Tjepiples	2	5 3/4	11	1 1/2	7	
Kedoonggalih	1 1/4	5	36	5	7	
Balong kangkoong	1 1/4	5 3/4	23	3	8	
Tekik	1/2	3 3/4	14	2	7	
Kemambang	2	3 3/4	34	5	7	
Delik	1/4	4 1/2	22	2 1/2	9	
Sedoeri	"	2	43	6 1/2	7	
Wonosari	2 1/2	1 3/4	15	2	7	
Tjangkringgatak	1/4	1	13	1 1/2	9	
Gerang koepang	2 1/2	1 3/4	12	1 3/4	7	
Kopoch koepang	2 1/4	1 1/4	8	1 1/4	7	
Koepang	2	1	26	4	6	
Bakalan koelon	1 1/4	1/4	9	2 1/2	4	
Waroe koelon	3/4	3/4	11	2	5	
Transportere	"	"	1496	223 1/4	"	

Table A-1: Data Sources

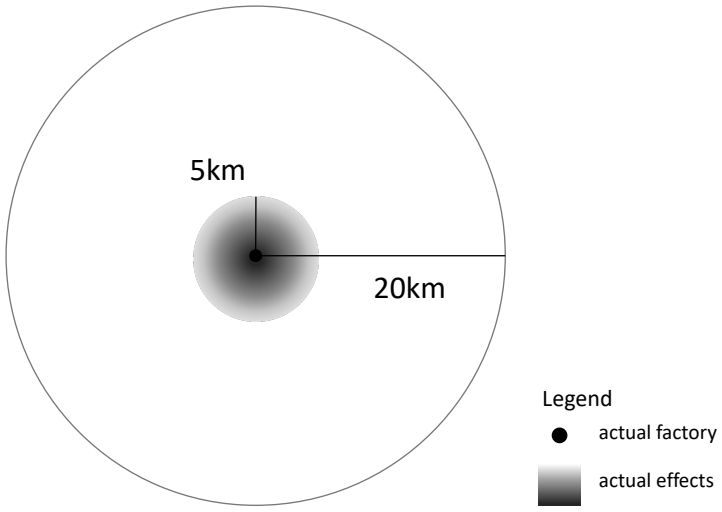
Data Source	**Publication Year(s)**	**Online Description**	**Variables**
1980 Population Census	1980	World Bank	Years of Education, Individual's Industry
2000 Population Census	2000	World Bank	Years of Education, Individual's Industry
PODES (Survey of Village Potential)	1980, 1996 2000, 2003 2005, 2008 2011	Duke Library	Village Head Years of Education, Number of High Schools; 1980 and 2003 Only: Village-Owned Land, Village Area; 1980 Only: Village-Owned Land, Village Area, Road Surface Type, Village has Electricity, Number of Teachers in Public Non-INPRES Primary Schools and INPRES Primary Schools, Number of School Buildings for Public Non-INPRES Primary, INPRES Primary, Junior High, and High; 2003 Only: Population, Tons of Sugar Cane Grown; 2011 Only: Distance to Nearest Sub-District Capital
Hydrosheds	2017	Hydrosheds	Elevation, Slope, Flow Accumulation
Author's Calculations	N/A	N/A	Distance to Coast
Digital Atlas of Southeast Asia	2017	USGS	Distance to Natural Harbor (Perennial lake or river within 10km of coast)
Indonesian Government Topographical Map	2017	Geospasial untuk Negeri	Distance to River
2006 Economic Census	2006	BPS	Employment by Village and Industry, Input-Output Table, Number of Firms, Value of Processed Sugar, Villages with Modern Sugar Factories
SUSENAS (National Socioeconomic Survey)	2001-2011	RAND	Individual's Industry, Household Consumption
Indonesian Government Road and Railroad Map	2007	Geospasial untuk Negeri	Road and Railroad Density
Agricultural Census 2003	2003	BPS	Hectares of Agricultural Land Used
Atlas van Nederlandsch-Indie	1856	Harvard Library	Great Post Road, 1830 Residency Capitals and 1856 Regency Capitals
Commissie Umbgrove	1858	N/A	Subjected Villages, Distance to Historical Factory
Author's Calculations (Based on Commissie Umbgrove)	N/A	N/A	Catchment Areas
Dutch Ships in Tropical Waters by Robert Parthesius	2012	OAPEN	VOC destinations

B Counterfactual Spacing

This appendix discusses in detail when the counterfactual factory method is likely to produce unbiased coefficients. Intuitively, if counterfactuals are placed too near the actual factories - relative to the rate at which the effects decline - the real effects will contaminate the counterfactual means. This will tend to bias the estimates downwards, making them too conservative. Importantly, as the following examples illustrate, downward bias will tend to be most pronounced for the bins that are closer to the factories, making the estimated effects too flat.

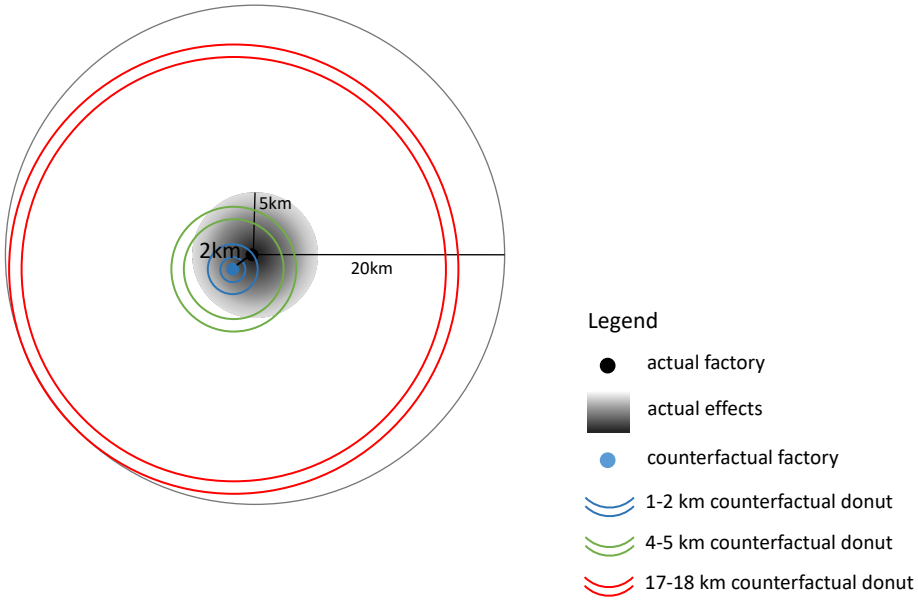
First consider the following example, which illustrates that when a counterfactual factory is placed too close to an actual factory, the counterfactual means tend to be most biased for the closer bins. Suppose that there is an actual factory, and the effect of proximity to that factory declines to zero over 5 kilometers (Figure B-1).

Figure B-1: Actual Factory



Suppose that a counterfactual factory is placed 2 kilometers from the actual factory (Figure B-2).

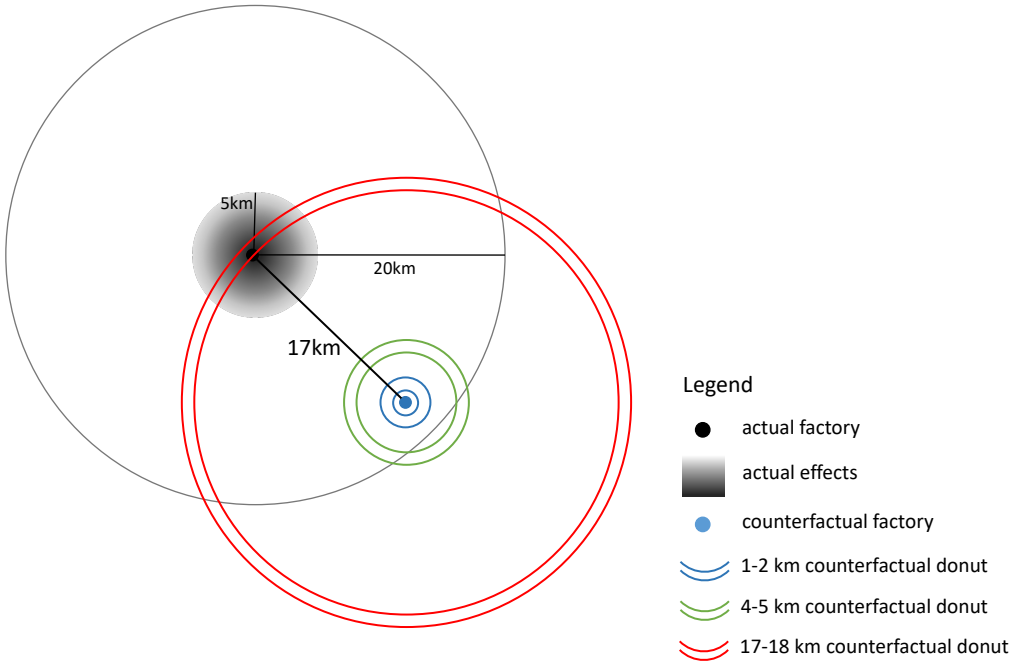
Figure B-2: Nearby Counterfactual



A high share of bins within 1-2 kilometers of that counterfactual will be contaminated by the actual factory effects (see the blue donut), whereas a lower share of villages in bins 4-5 kilometers away will be contaminated (see the green donut). The 1-2 kilometer blue donut is concentrated in a much smaller area, which is all near the actual factory (i.e. its circumference is 4π km versus the 10π km circumference of the 4-5 km green donut). By the basic principles of geometry, a larger donut cannot be entirely near a single actual factory. Beyond 7 kilometers from the counterfactual, none of the locations will be contaminated by the actual effects (see for example the red 17-18 km donut).

Now consider another example, which illustrates that for this bias to be large, the counterfatual needs to be sufficiently close to an actual factory, relative to the speed at which the effects decline to zero. Suppose there is an actual factory 17 kilometers away from a counterfactual (Figure B-3).

Figure B-3: Far Counterfactual



As long as the effects of the real factory do not decline too slowly, a much smaller share of the donut between 17-18 kilometers from the counterfactual will be contaminated contaminated than in the 1-2 kilometer donut discussed in the above example where the counterfactual is only 2 kilometers from the actual factory. The area of the 1-2 kilometer donut is $3\pi\ km^2$, spread over a circumference of $4\pi\ km$. The area of the 17-18 kilometer donut is $35\pi\ km^2$, spread over a circumference of $36\pi\ km$. While in theory the density of actual factories could be so high that the entire 17-18 kilometer donut was full of factories, in practice this would require a much larger number of factories than is present in our context.

The intuition behind this can be shown more systematically through a simulation exercise, which proceeds according to these steps:

1. We construct a grid of finely spaced points, one point to be the actual factory, and construct an outcome variable that declines to zero linearly when moving from 0 to k kilometers away from the actual factory. We vary k, which captures the degree to which the effect of being near a factory dissipates rapidly.

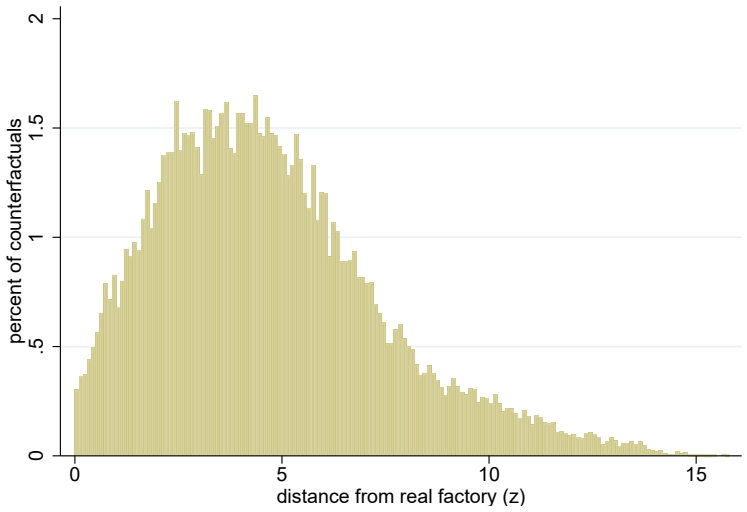
2. For each other "counterfactual" point on the grid, we compute the effect of being distance $0.1, 0.2, \ldots 20$ km that point

3. For each of the counterfactual points, we also compute its distance to the actual factory. We average the effects of being $0.1, 0.2, \ldots 20$ from a counterfactual factory, computed in step 2, within bins of distance $0.1, 0.2, \ldots$km from the actual factory

4. We then take a weighted average of the effects of being distance $0.1, 0.2, \ldots 20$ km from a counterfactual point computed in step 3), where the weights are the share of counterfactual factories in our baseline analysis that are distance x (i.e. $0 - 0.1, 0.1 - 0.2 \ldots$) from the nearest actual factory

Figure B-4 plots the distribution of straight-line distances between the counterfactual and nearest actual factories in our baseline analysis, which we use to construct the weights in step 4.

Figure B-4: Distribution of distances

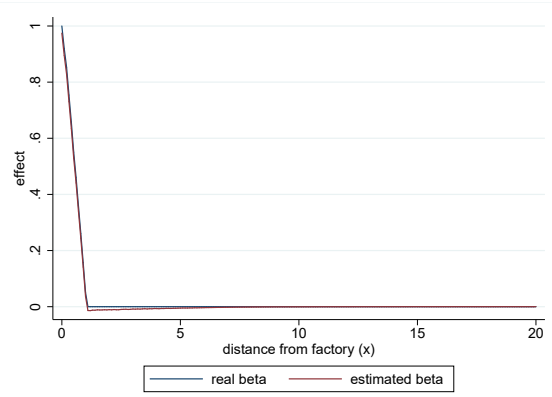


Notes: z is the distance from a counterfactual to the closest real factory. The counterfactuals are those used in the baseline analysis.

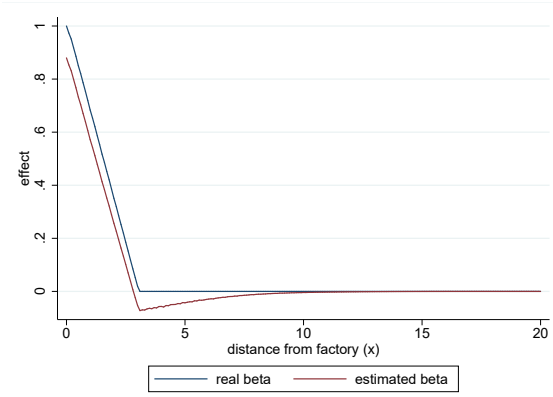
Figure B-5 shows the results from the simulation exercise. Panel a) assumes that the effects of the real factory decline linearly from 1 to 0 when moving from 0 to 1 kilometer from the actual factory, panel b) assumes they decline from 1 to 0 when moving from 0 to 3 kilometers from the actual factory, in panel c) they decline to 0 over 5 kilometers, in panel d) they decline to 0 over 10 kilometers, and in panel e) they decline to 0 over 20 kilometers. The blue line plots the actual effect and the red line plots the estimated effect that subtracts the counterfactual mean, with counterfactual means constructed according to the procedure outlined above.

Figure B-5: Real and estimated effects when actual effect dissipates at different rates

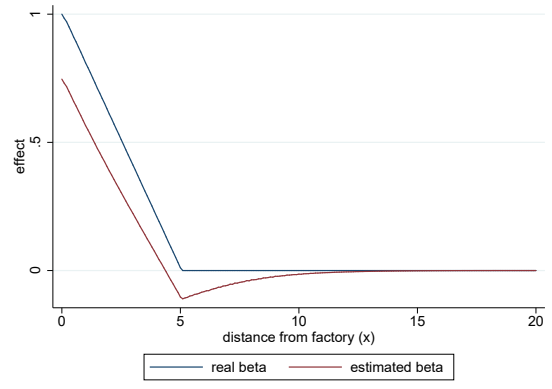
(a) $k = 1$



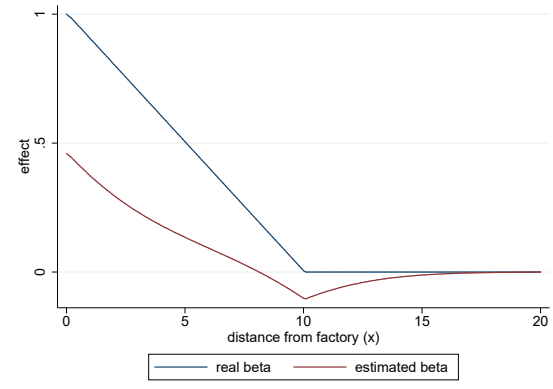
(b) $k = 3$



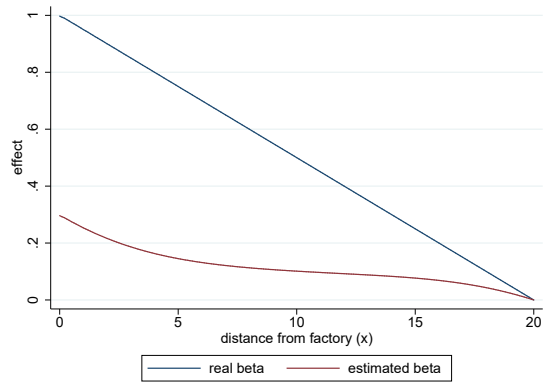
(c) $k = 5$



(d) $k = 10$



(e) $k = 20$



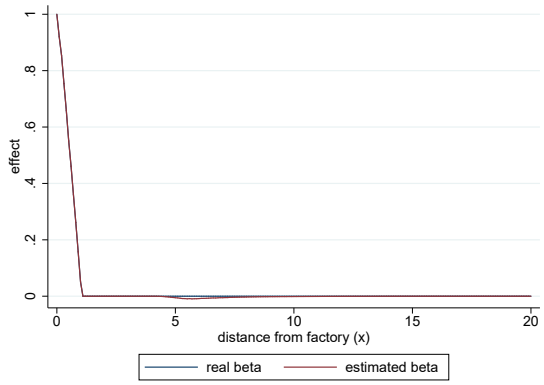
Notes: These plots are analogous to the distance-to-factory plots in the paper. Blue lines plot the real, underlying effect of distance to the nearest factory while red lines plot the estimated effect when counterfactual means have been subtracted. k is the maximum extent of the effect of proximity to the real factory—i.e., at distance k from the real factory, the outcome is zero.

When the effects decline quickly, the bias is quite small, because the radius in which the real effects are pronounced is too small to contaminate much of the counterfactual donuts. On the other hand, when effects decline slowly, the bias can be quite large, especially for the bins nearest the factory. This bias is typically towards zero, making the estimates that subtract the counterfactual means too flat. Appendix C will show that the estimates are correctly *sized* regardless; that is, under the null of no effect, we do not falsely over-reject the null hypothesis. The analysis in this section, however, suggests that when we detect an effect and when the true effect declines slowly with distance, then there may be a downward bias.

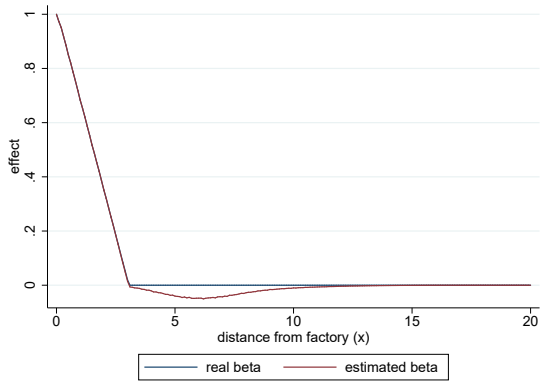
In our context, the fact that the counterfactual means are near zero in the baseline analysis suggests that we are in a world where effects decline quickly. To provide further evidence, Appendix I adds an additional restriction that all counterfactual factories must be at least 5 kilometers in straight line distance from the nearest actual factory, and we repeat the above simulation exercise where the weights are the share of counterfactual factories that are a given distance from the nearest actual factory under this restriction (Figure B-6):

Figure B-6: Real and estimated effects for different k, ensuring that all counterfactual factories are 5 km away from nearest actual factory

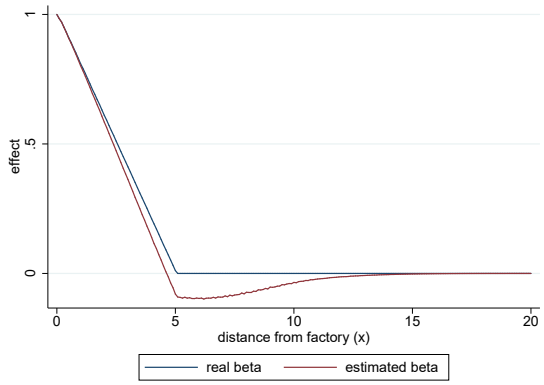
(a) $k = 1$



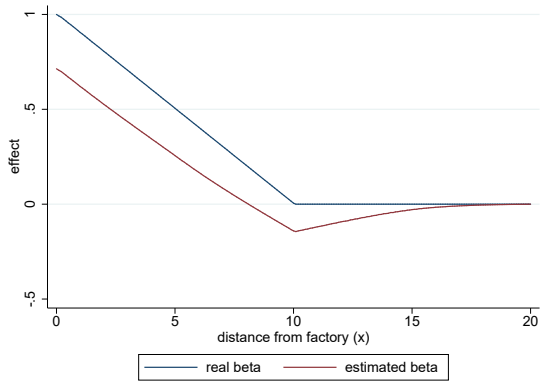
(b) $k = 3$



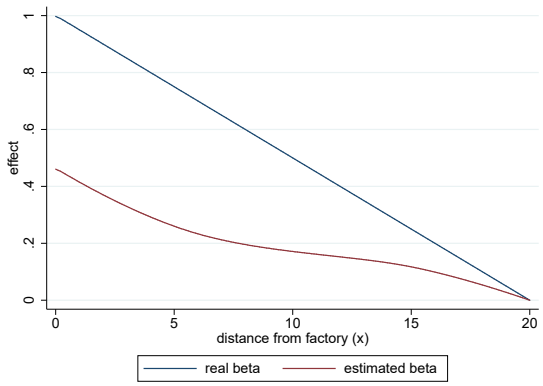
(c) $k = 5$



(d) $k = 10$



(e) $k = 20$



Notes: These plots are analogous to the distance-to-factory plots in the paper. Blue lines plot the real, underlying effect of distance to the nearest factory while red lines plot the estimated effect when counterfactual means have been subtracted. k is the maximum extent of the effect of proximity to the real factory—i.e., at distance k from the real factory, the outcome is zero.

Now, effects can decline more slowly without the bias becoming pronounced, since the spacing is on average greater. When we run our actual analysis using this spacing restriction - while it reduces the number of factories that have suitable nearby counterfactuals by nearly half - the estimates of the distance to factory effects are essentially unchanged (Appendix I). This provides further evidence that our context is one in which the effects decline quickly enough that the baseline spacing is unlikely to lead to significant bias.

C Size

To examine whether the counterfactual factory analysis is correctly sized under the null of no effect, we report the following simulation exercise. We create a grid and assign an outcome variable that consists of *iid* random noise to each point on the grid. The variance of the random noise is equal to the variance of the residuals from the distance to the factory analysis for the share of population working in agriculture, the outcome that we use in the paper to illustrate the methodology.

We randomly choose an "actual" factory location and 1,000 "counterfactual" locations, and compute where the actual distance to factory effects fall in the distributions of counterfactual distance to factory effects. If the absolute actual coefficient is greater than 95% of the absolute counterfactual coefficients, we reject the null with a threshold of $p < .05$. We repeat this entire exercise 5,000 times (i.e. generating outcome variables for each grid point, and choosing an "actual" and 1,000 "counterfactual" locations), and then compute the share of times we reject the null of no effect for each distance to factory bin. The below table shows that we reject the null about 5% of the time, as expected if the test is correctly sized.

Table C-1: IID Noise

Bin (1)	Share significant [$p < 0.05$] (2)
0-1	0.052
1-2	0.052
2-3	0.053
3-4	0.052
4-5	0.047
5-6	0.047
6-7	0.049
7-8	0.048
8-9	0.048
9-10	0.049
10-11	0.054
11-12	0.049
12-13	0.042
13-14	0.047
14-15	0.048
15-16	0.049
16-17	0.048
17-18	0.051
18-19	0.050
Mean	0.049

Notes: Column 2 shows the share of simulations in which the real estimate was larger in absolute value than 95% of the counterfactual estimates.

Conclusions are very similar if we use spatially correlated instead of iid noise, as documented in the below table. We group villages into 100 equally sized spatial clusters and generate random noise by combining two components, an independent component and a component that is equal within clusters and independent between clusters.

Table C-2: Spatially Correlated Noise

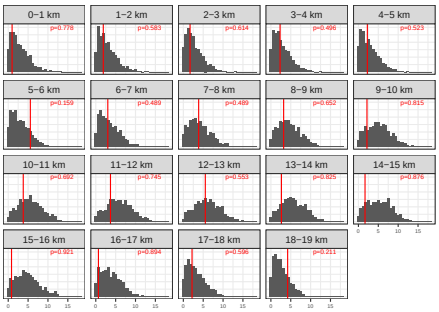
Bin (1)	Share significant [$p < 0.05$] (2)
0-1	0.045
1-2	0.047
2-3	0.049
3-4	0.048
4-5	0.051
5-6	0.052
6-7	0.052
7-8	0.046
8-9	0.049
9-10	0.050
10-11	0.046
11-12	0.052
12-13	0.052
13-14	0.045
14-15	0.047
15-16	0.053
16-17	0.047
17-18	0.047
18-19	0.049
Mean	0.049

Notes: Column 2 shows the share of simulations in which the real estimate was larger in absolute value than 95% of the counterfactual estimates.

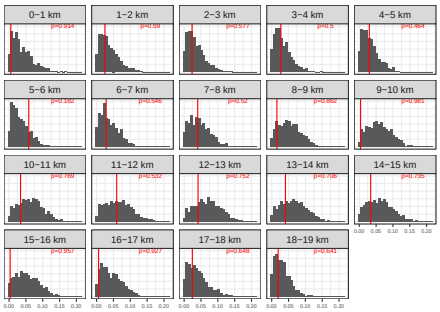
D Independent Shifts Distributions

Figure D-1: Geography

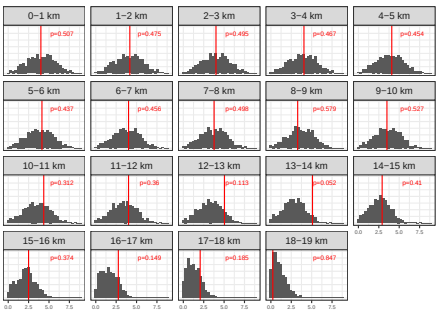
(a) Elevation



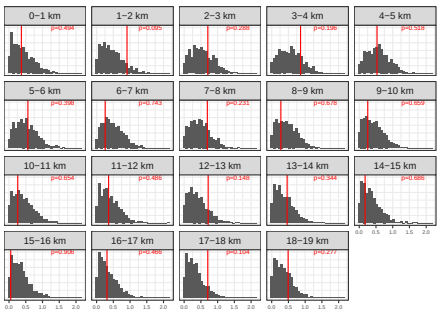
(b) Slope



(c) Distance to Coast

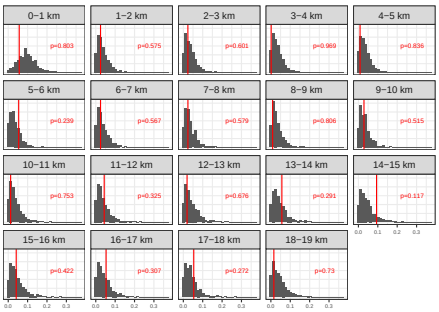


(d) Log Flow Accumulation

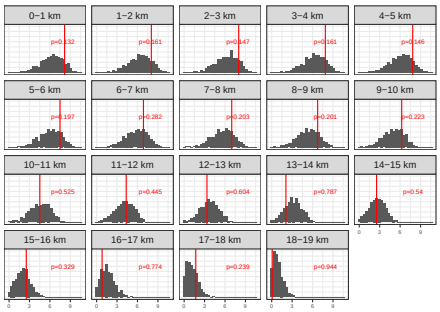


(e) Distance to River

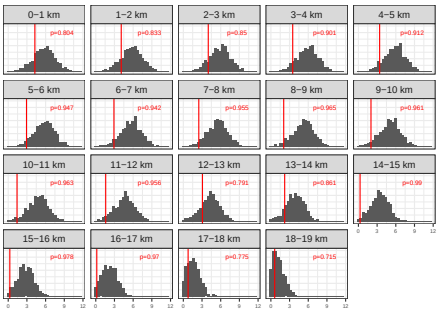
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



(h) Distance to VOC Port



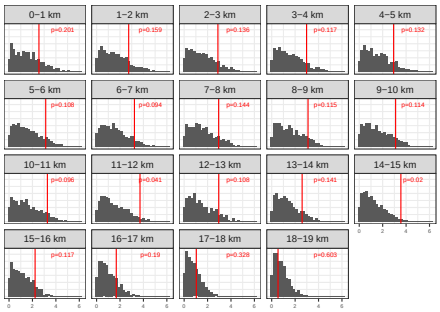
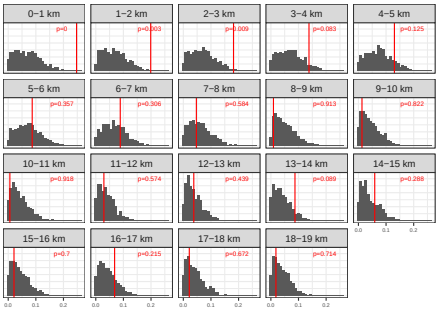
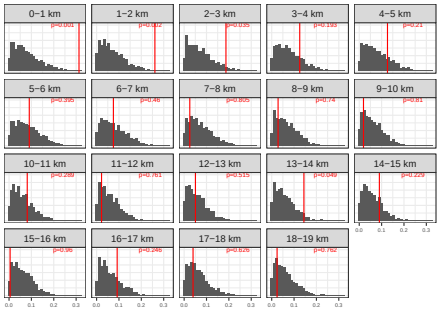


Figure D-2: Industry and Agglomeration

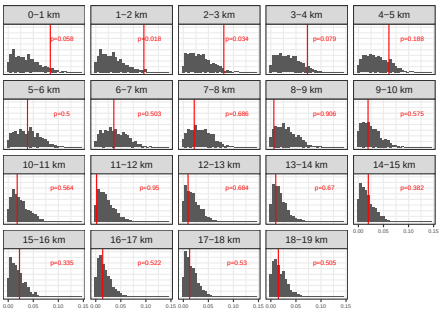
(a) Agriculture (Susenas 2001-11)



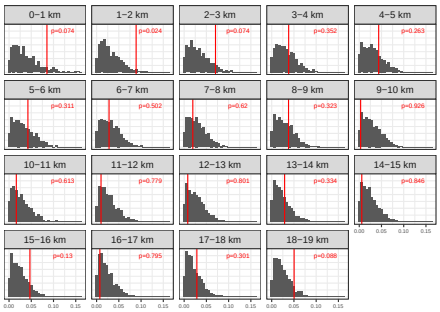
(b) Agriculture (Census 1980)



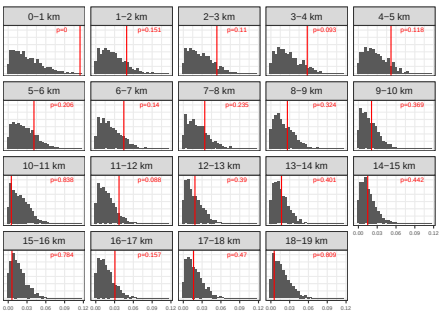
(c) Manufacturing (Susenas 2001-11)



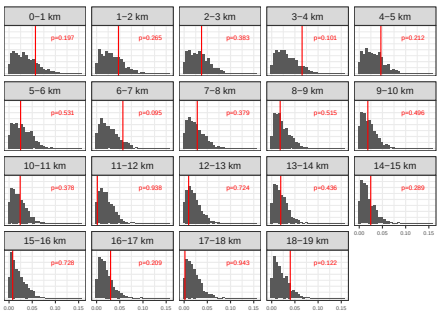
(d) Manufacturing (Census 1980)



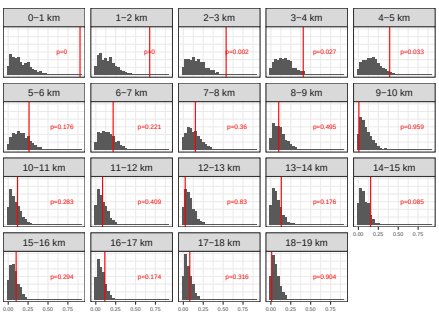
(e) Retail (Susenas 2001-11)



(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



(h) Log Pop. Density (PODES 1980)

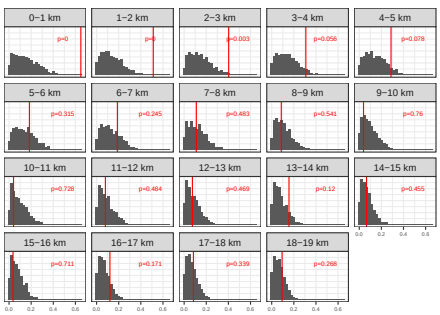
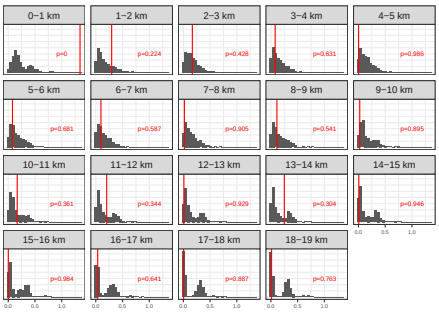
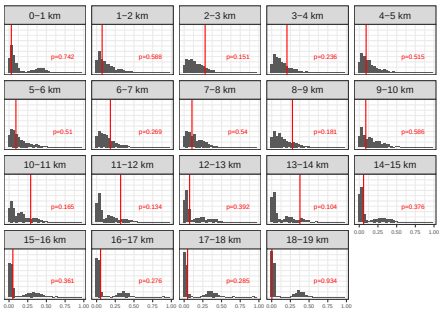


Figure D-3: Sugar and Linked Industries

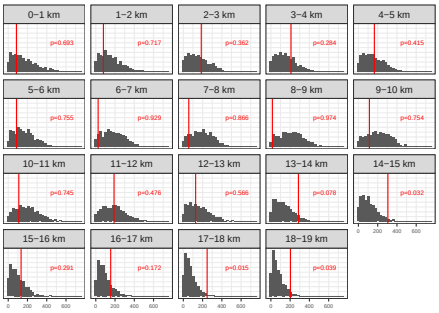
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



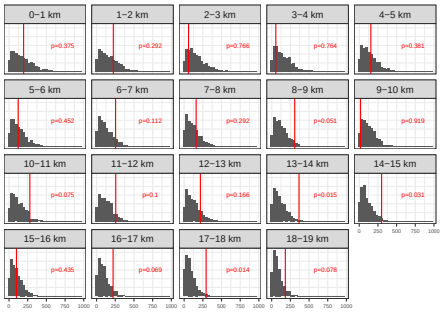
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



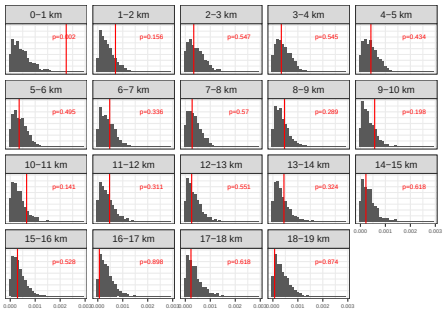
(c) Tons of Cane Grown (Full Sample, PODES 2003)



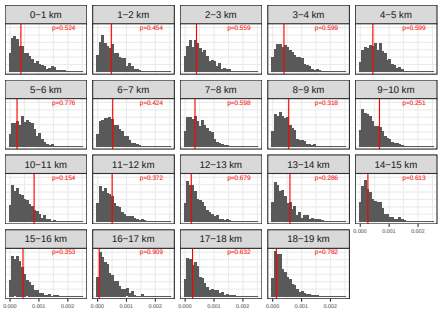
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



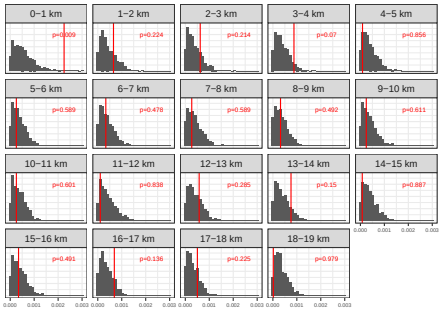
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)

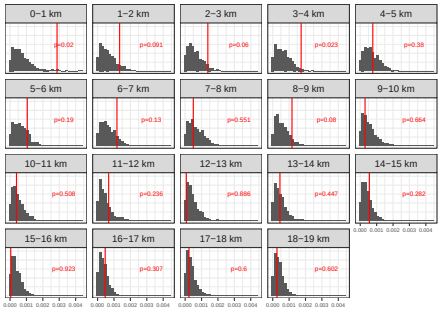
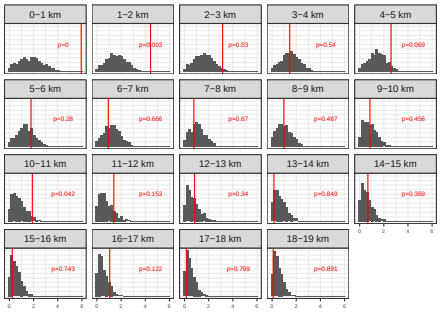
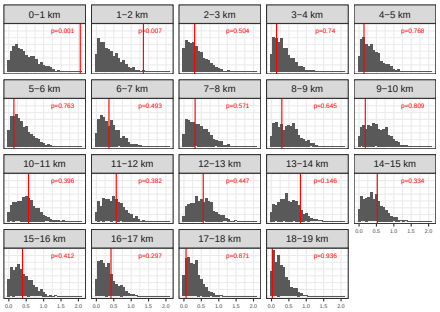


Figure D-4: Infrastructure

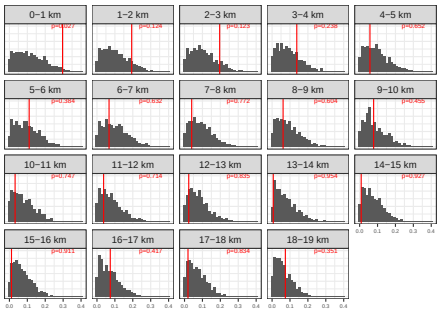
(a) Colonial Road Density (1900)



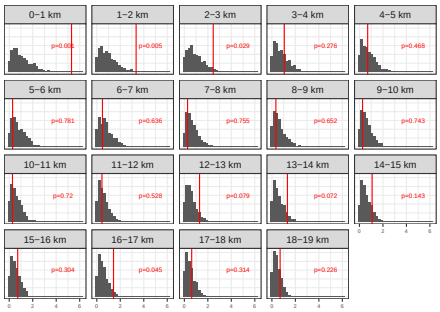
(b) Colonial Railroad Density (1900)



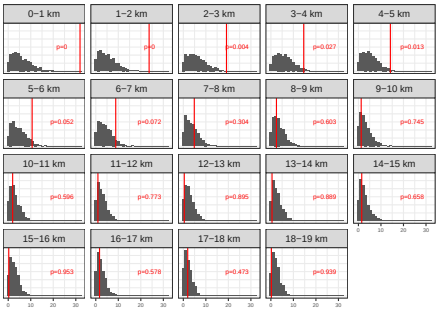
(c) Dirt Road (PODES 1980)



(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



(f) Railroad Density (2017)

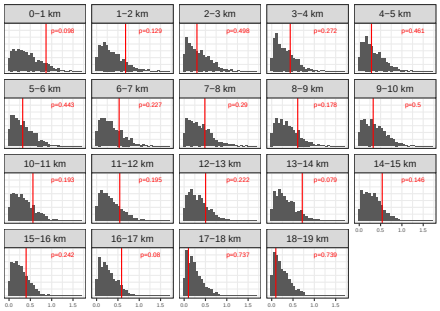
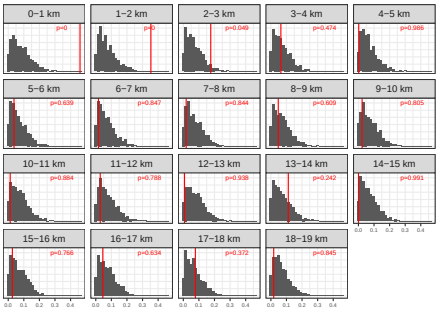
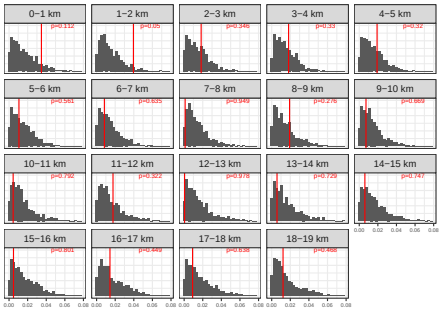


Figure D-5: Other Public Goods

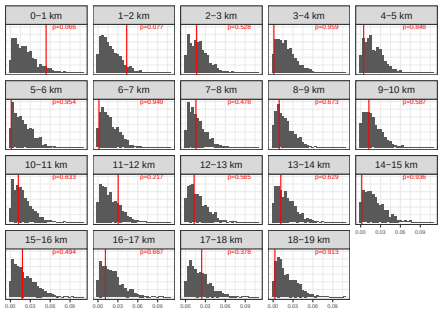
(a) Village Has Electricity (PODES 1980)



(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



(d) Distance to Subdistrict Capital (2011 PODES)

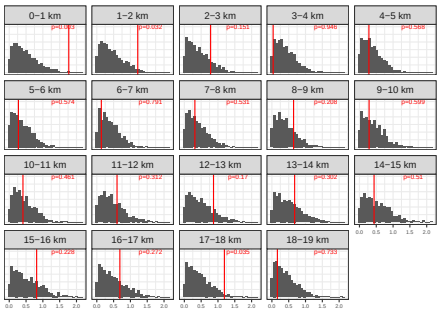
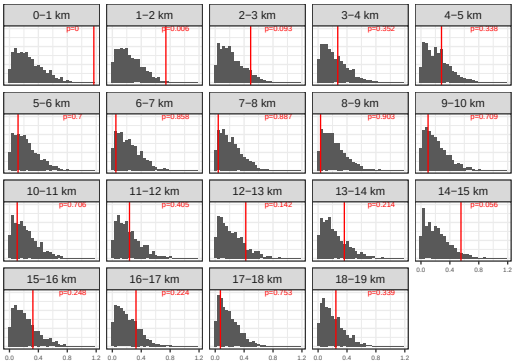
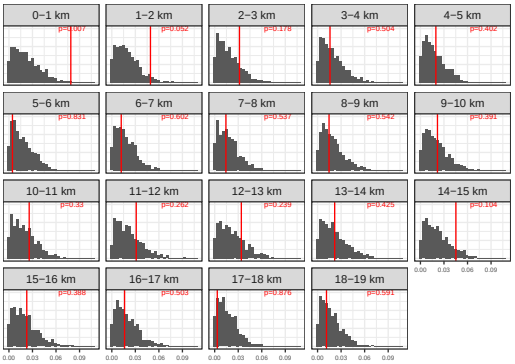


Figure D-6: Education

(a) Years Education



(b) Primary



(c) High School

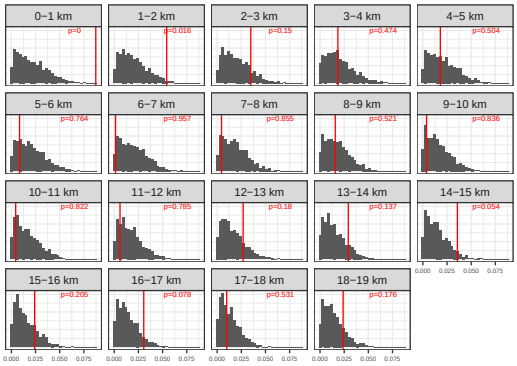
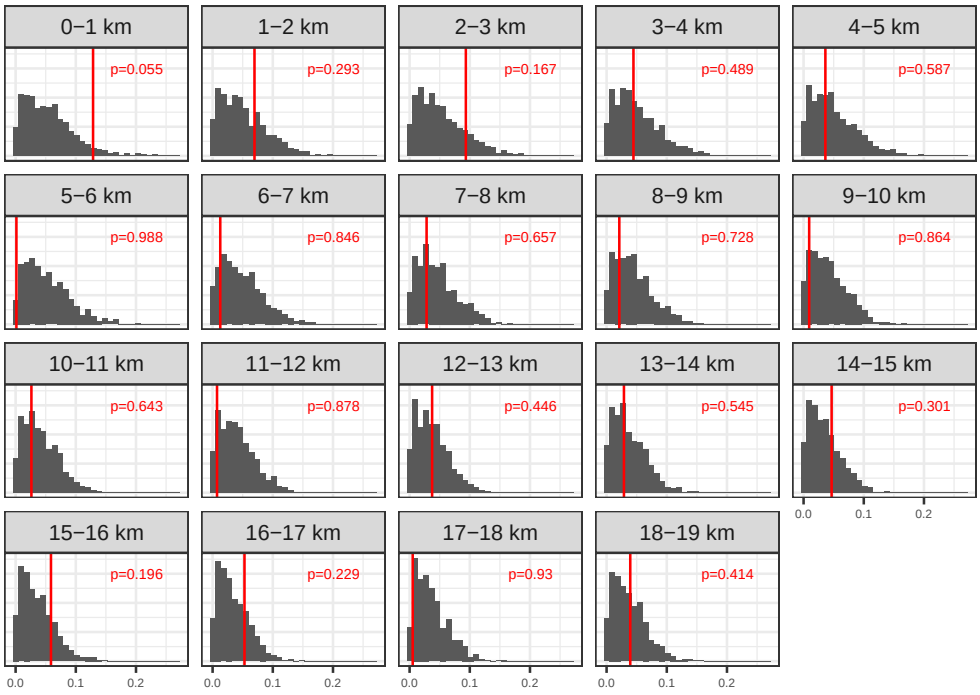


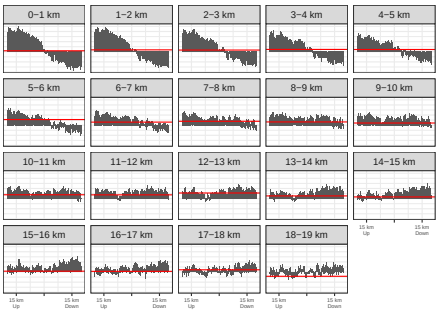
Figure D-7: Expenditure (2001-11)



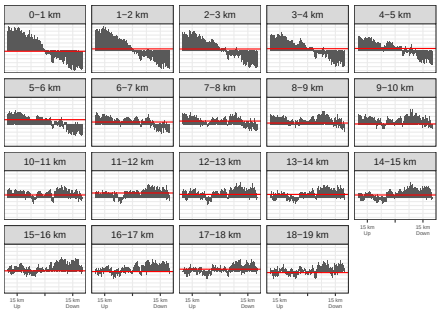
E Common Shifts Distributions

Figure E-1: Geography

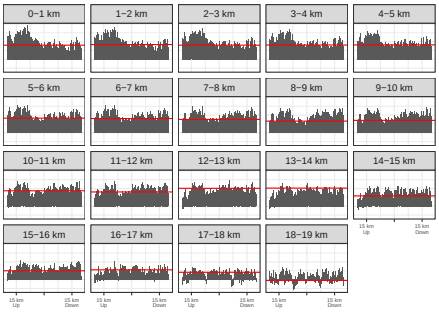
(a) Elevation



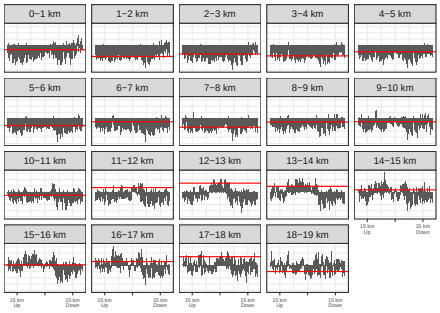
(b) Slope



(c) Distance to Coast

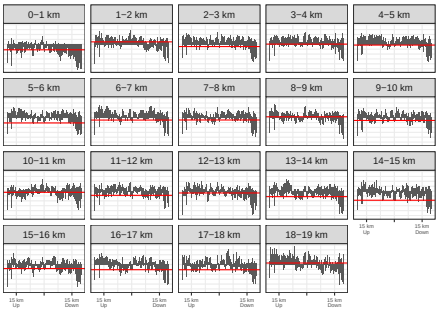


(d) Log Flow Accumulation

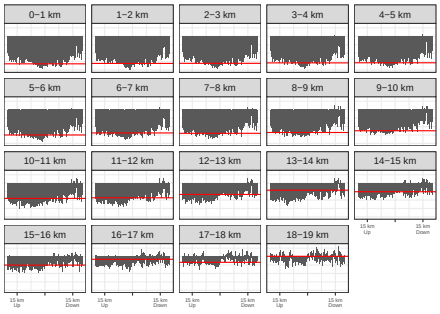


(e) Distance to River

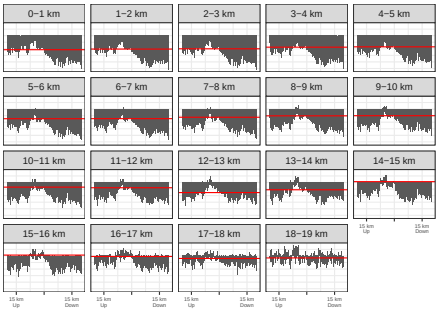
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



(h) Distance to VOC Port



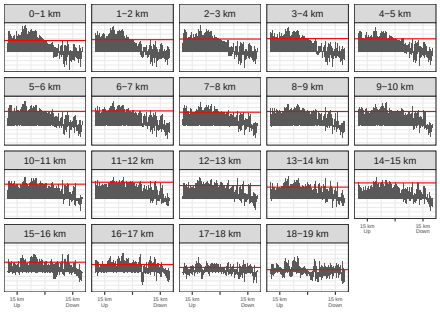
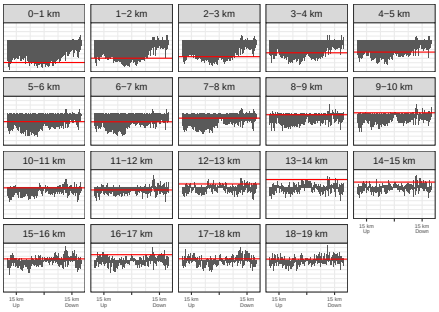
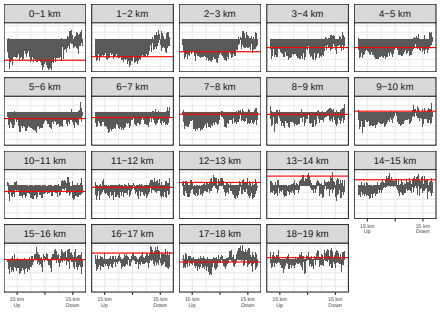


Figure E-2: Industry and Agglomeration

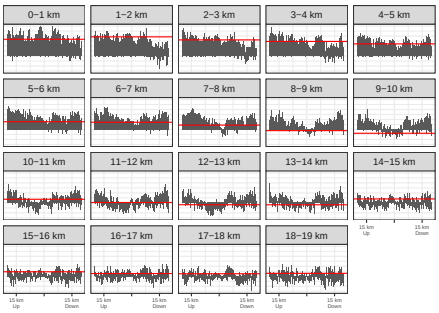
(a) Agriculture (Susenas 2001-11)



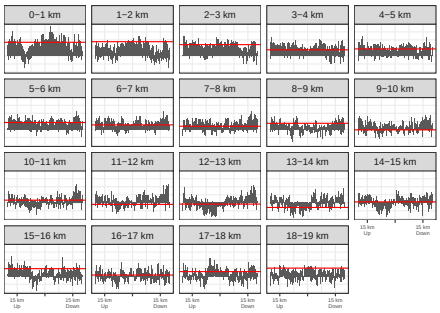
(b) Agriculture (Census 1980)



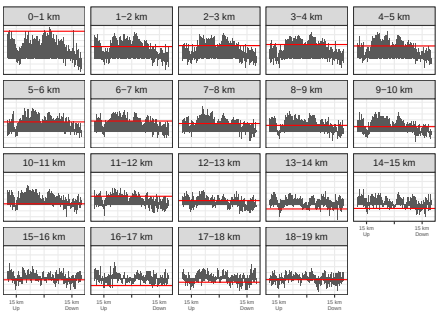
(c) Manufacturing (Susenas 2001-11)



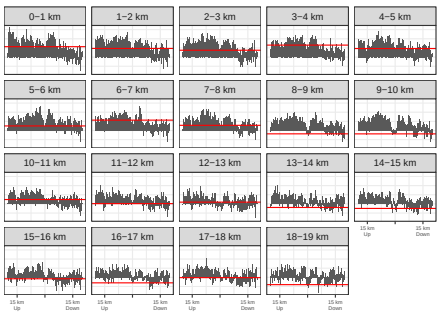
(d) Manufacturing (Census 1980)



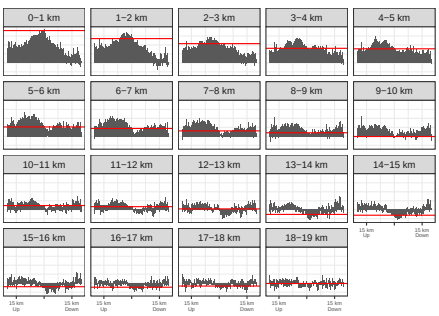
(e) Retail (Susenas 2001-11)



(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



(h) Log Pop. Density (PODES 1980)

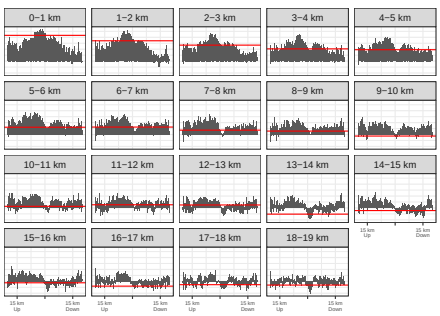
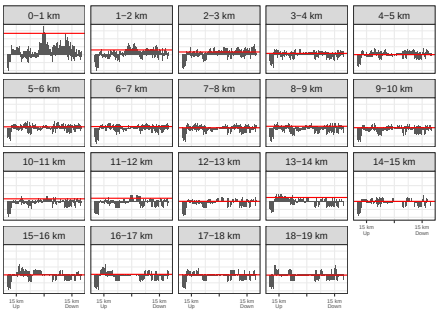
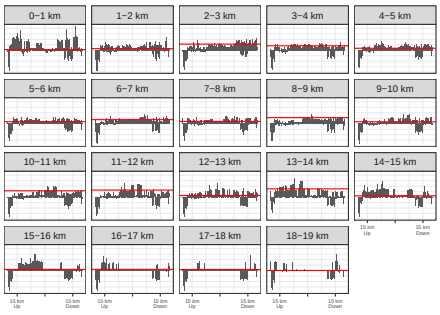


Figure E-3: Sugar and Linked Industries

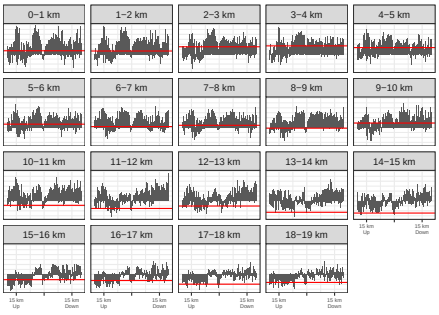
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



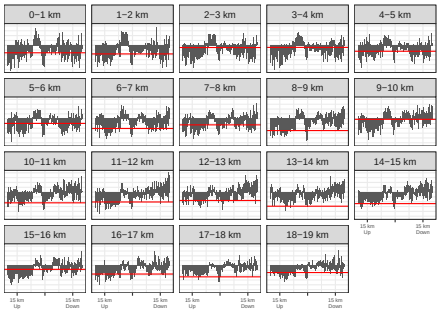
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



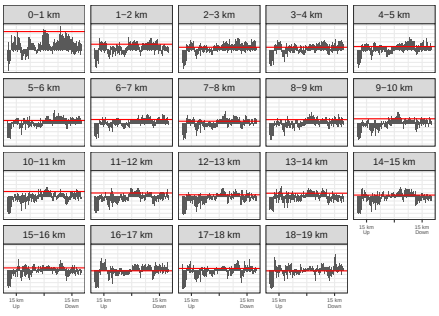
(c) Tons of Cane Grown (Full Sample, PODES 2003)



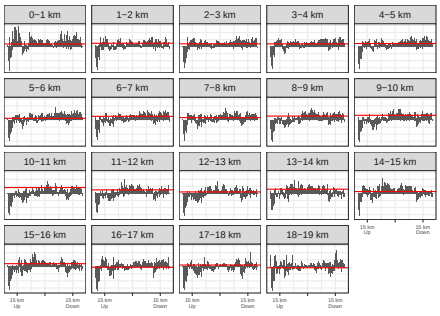
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



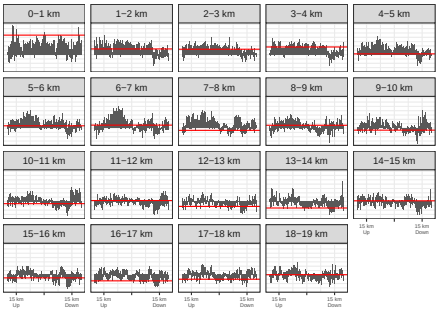
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)

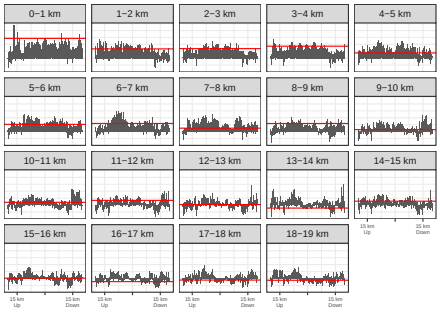
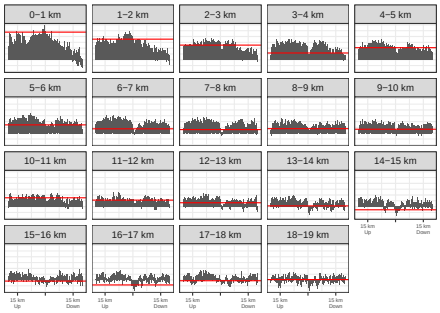
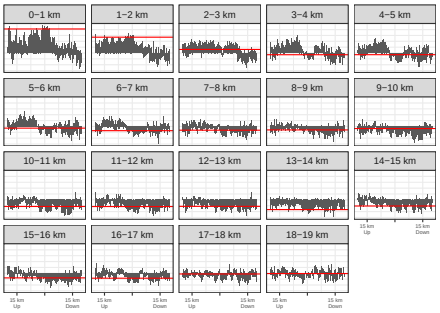


Figure E-4: Infrastructure

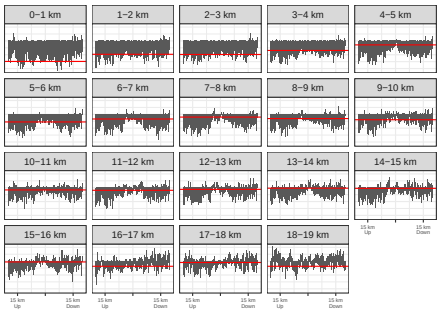
(a) Colonial Road Density (1900)



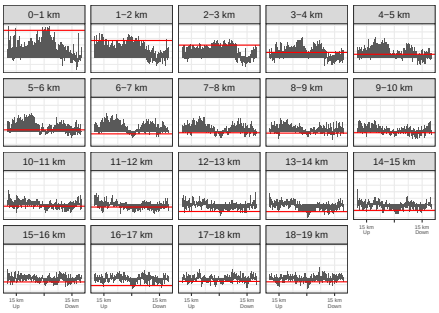
(b) Colonial Railroad Density (1900)



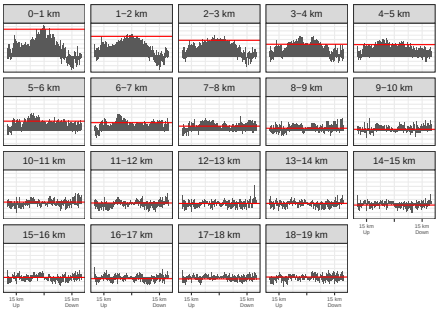
(c) Dirt Road (PODES 1980)



(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



(f) Railroad Density (2017)

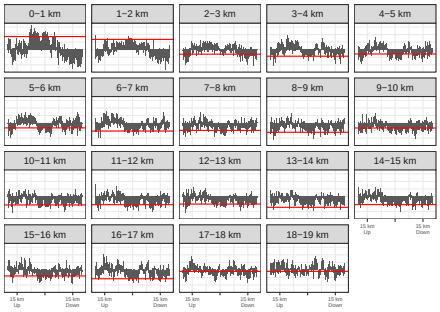
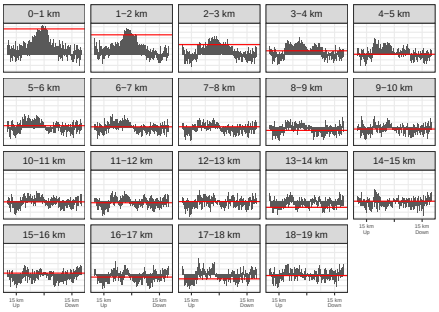
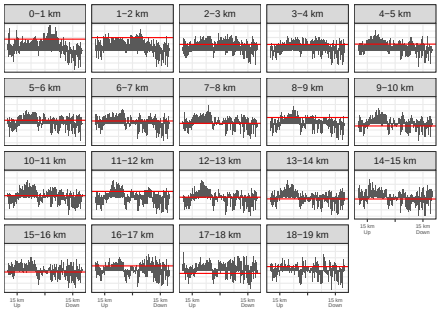


Figure E-5: Other Public Goods

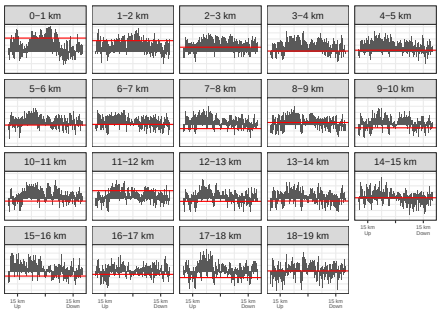
(a) Village Has Electricity (PODES 1980)



(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



(d) Distance to Subdistrict Capital (2011 PODES)

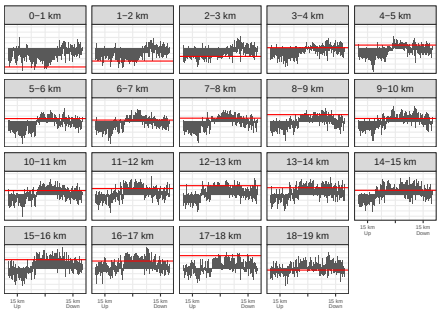
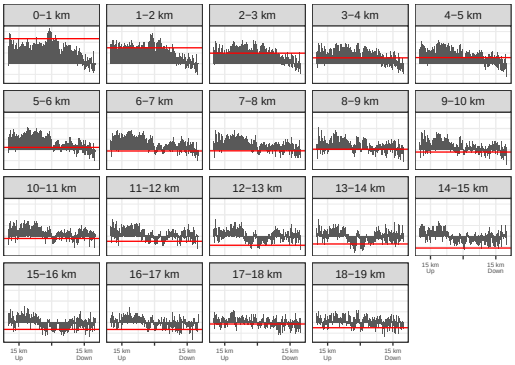
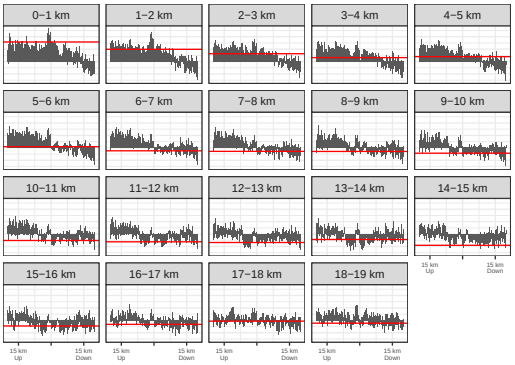


Figure E-6: Education

(a) Years Education



(b) Primary



(c) High School

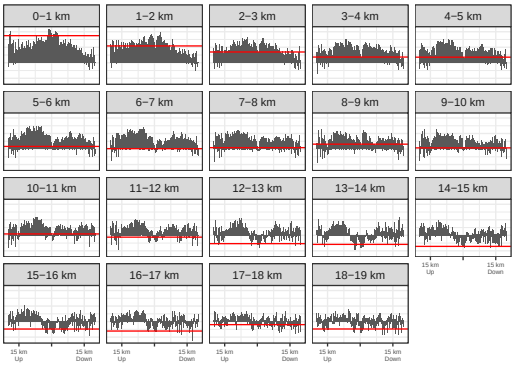
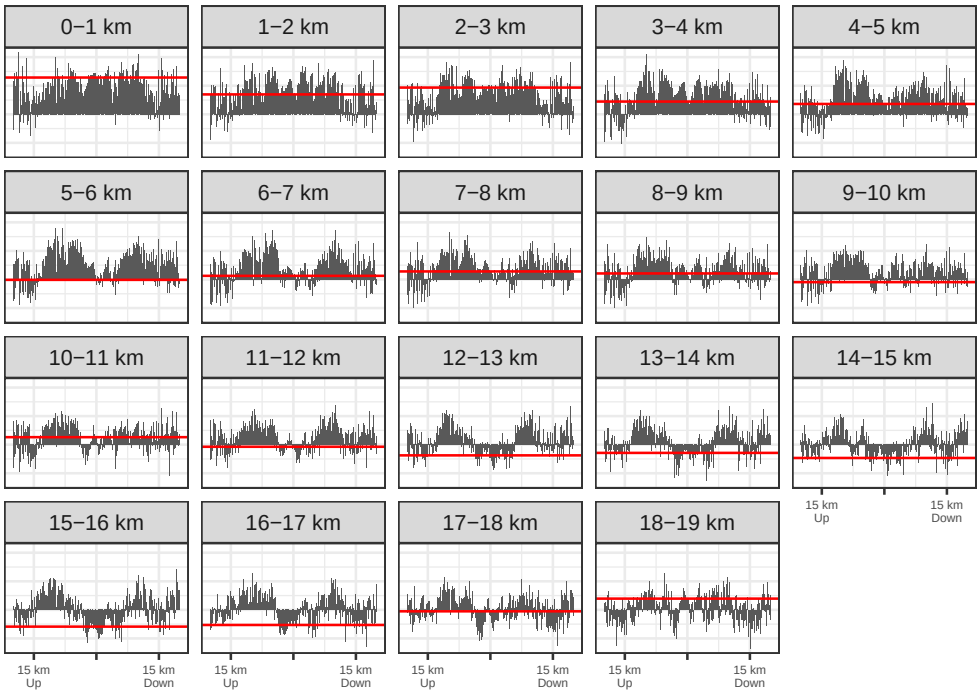


Figure E-7: Expenditure (2001-11)

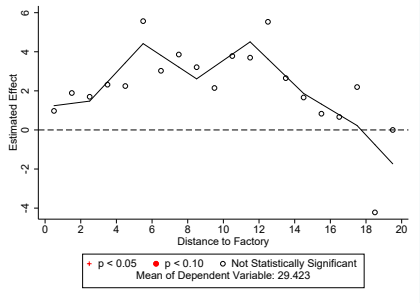


F OLS Specification

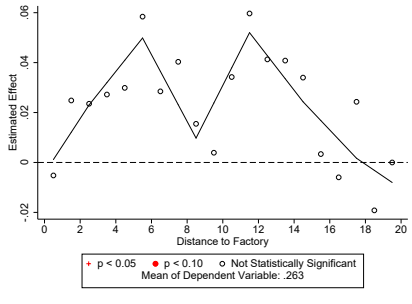
This appendix uses ordinary least squares analysis to estimate the impact of distance to the nearest factory, by 1 kilometer bins. Standard errors are clustered by 100 km^2 grid cells.

Figure F-1: Balance

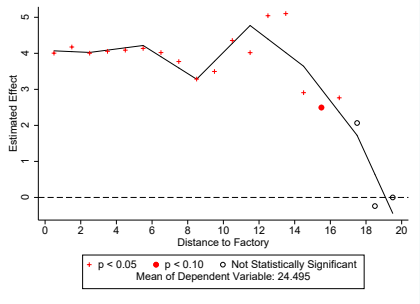
(a) Elevation



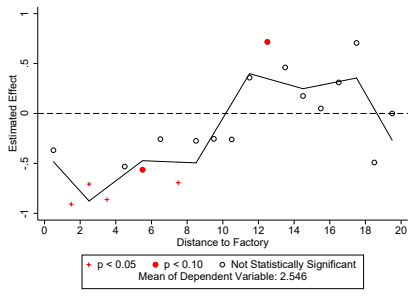
(b) Slope



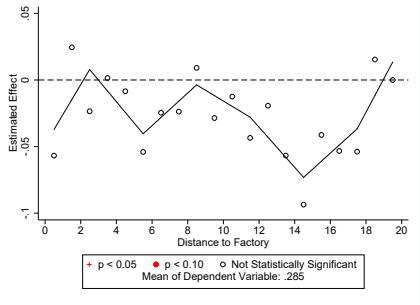
(c) Distance to Coast



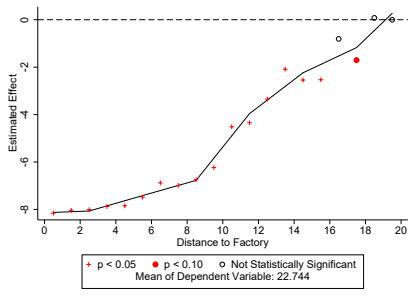
(d) Log Flow Accumulation



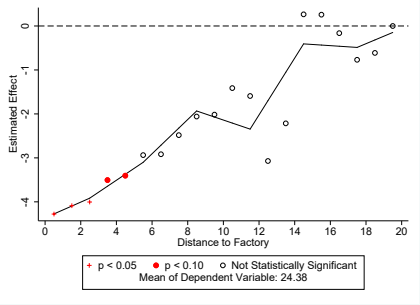
(e) Distance to River



(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



(h) Distance to VOC Port

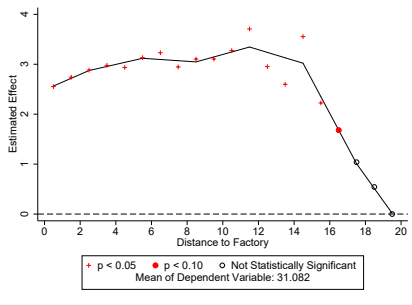
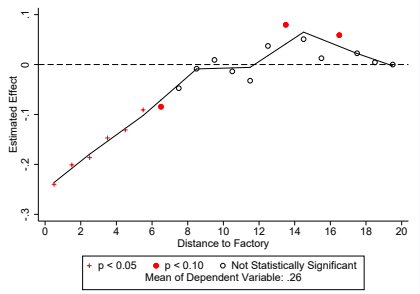
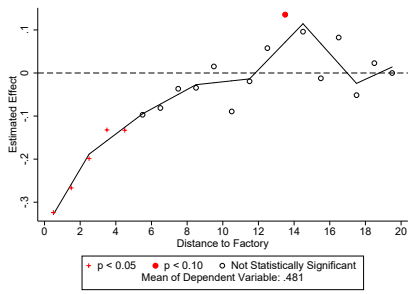


Figure F-2: Industry and Agglomeration

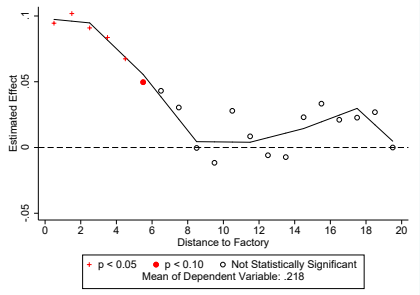
(a) Agriculture (Susenas 2001-11)



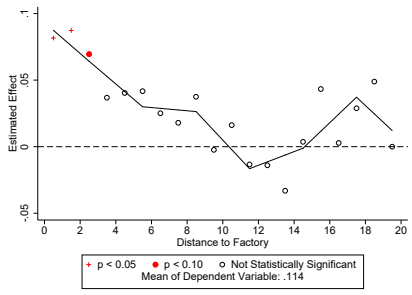
(b) Agriculture (Census 1980)



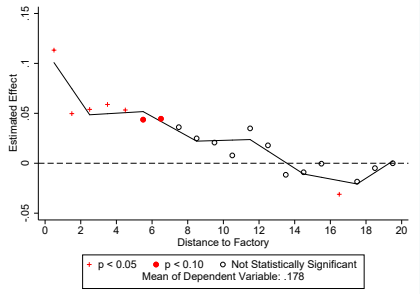
(c) Manufacturing (Susenas 2001-11)



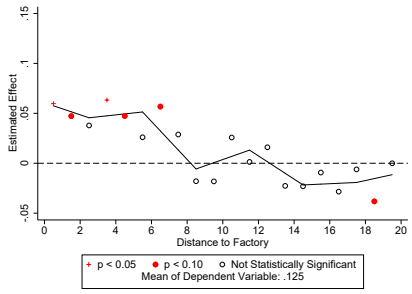
(d) Manufacturing (Census 1980)



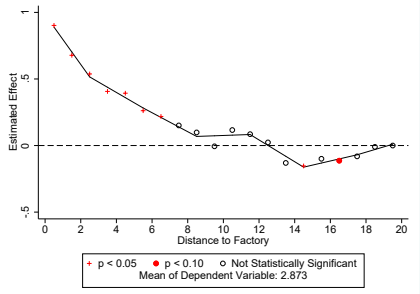
(e) Retail (Susenas 2001-11)



(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



(h) Log Pop. Density (PODES 1980)

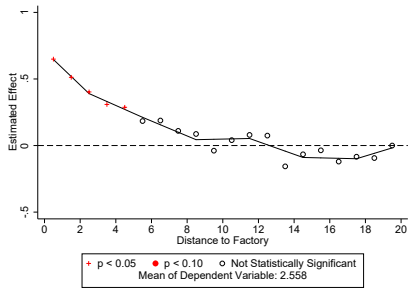
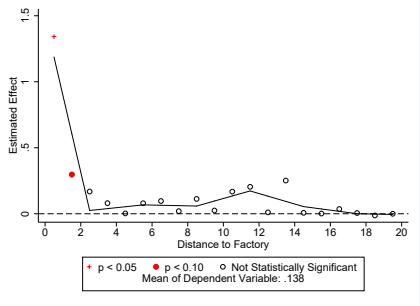
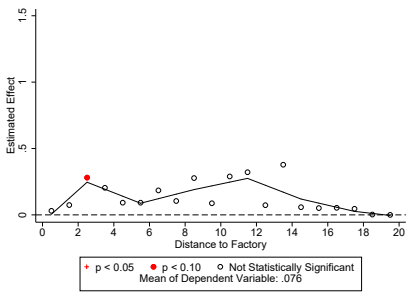


Figure F-3: Sugar and Linked Industries

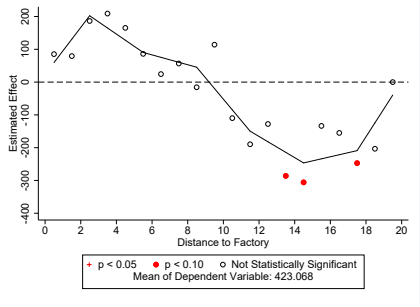
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



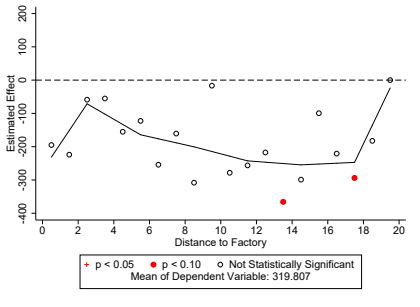
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



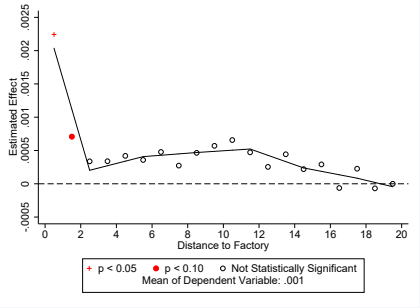
(c) Tons of Cane Grown (Full Sample, PODES 2003)



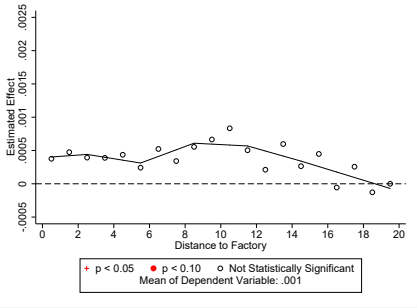
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



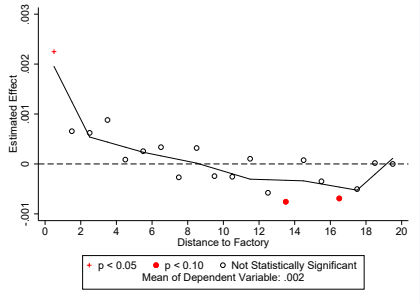
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)

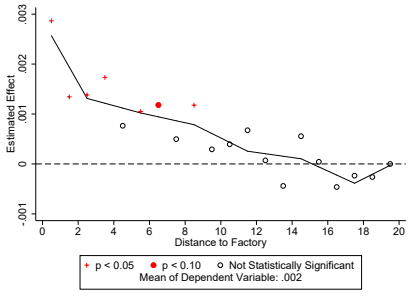
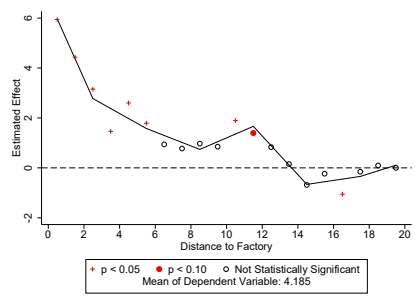
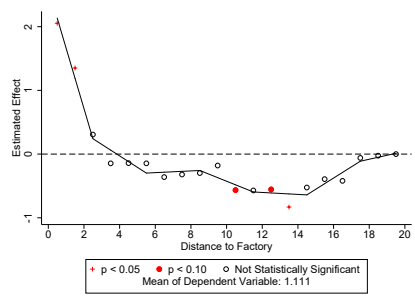


Figure F-4: Infrastructure

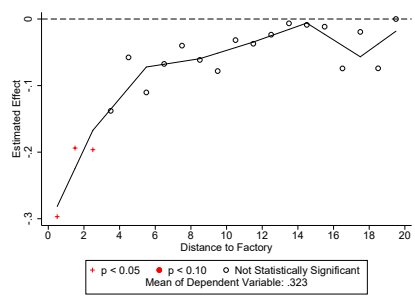
(a) Colonial Road Density (1900)



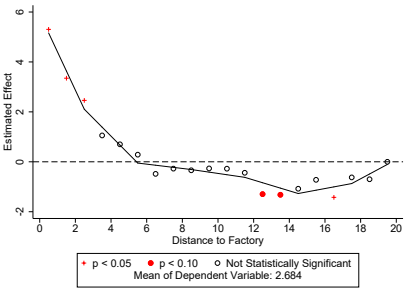
(b) Colonial Railroad Density (1900)



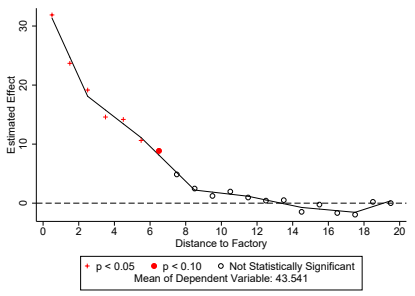
(c) Dirt Road (PODES 1980)



(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



(f) Railroad Density (2017)

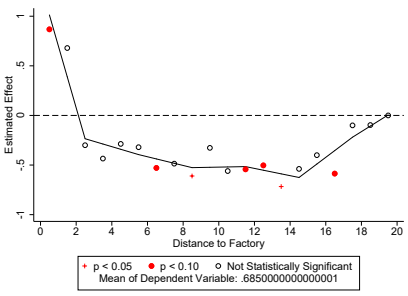
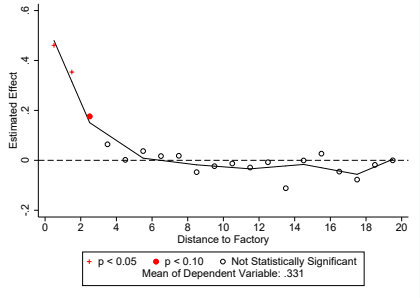
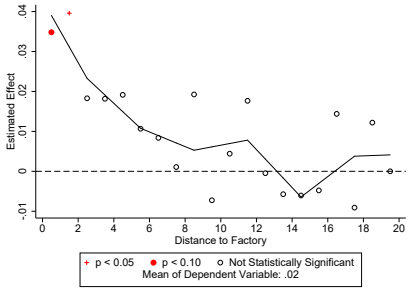


Figure F-5: Other Public Goods

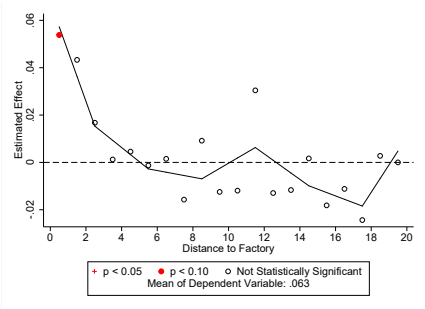
(a) Village Has Electricity (PODES 1980)



(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



(d) Distance to Subdistrict Capital (2011 PODES)

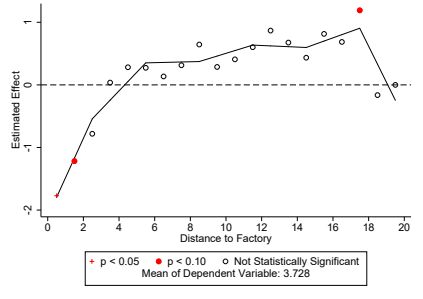
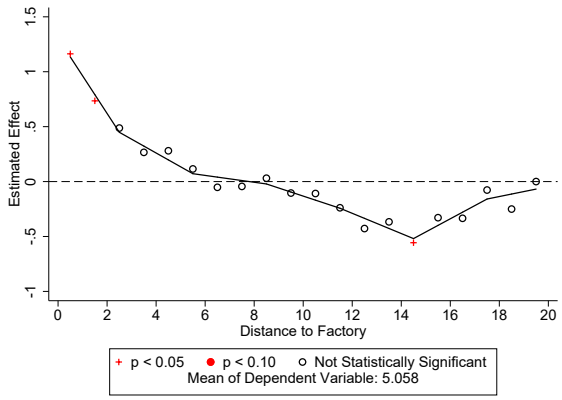
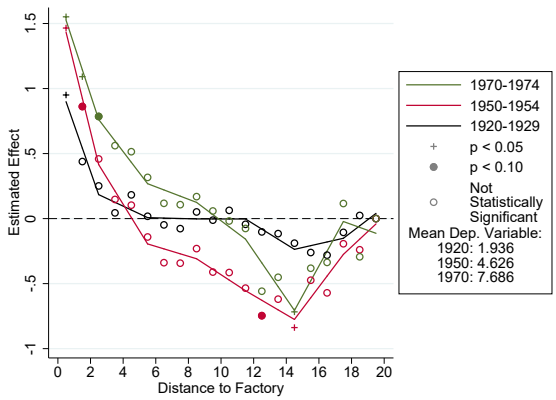


Figure F-6: Education

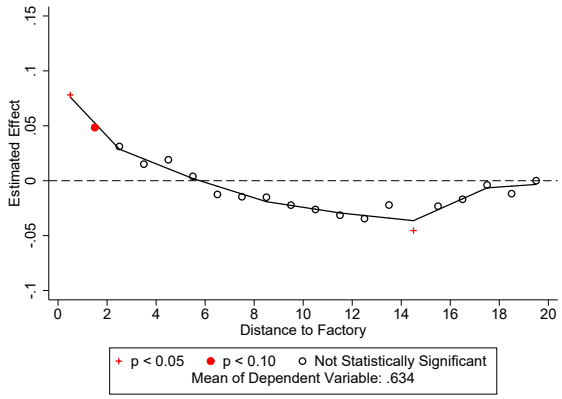
(a) Years Education (2000 Census)



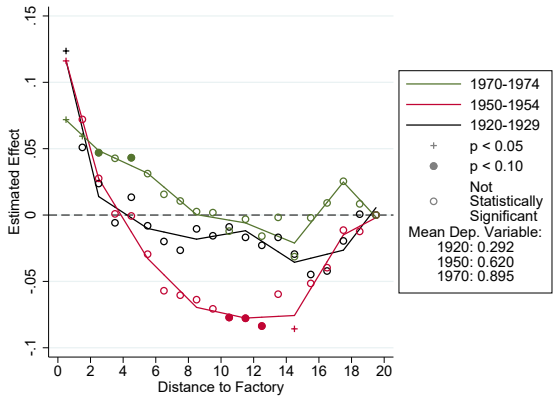
(b) Years Education by Cohort



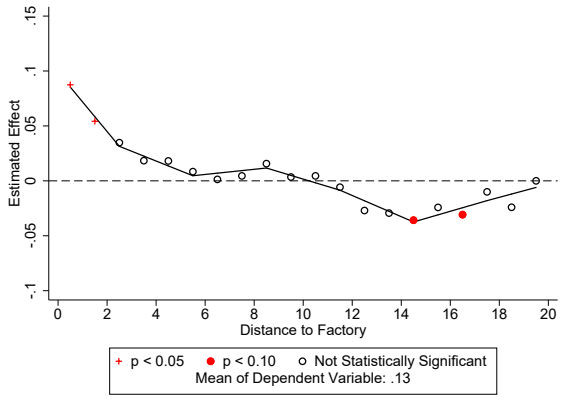
(c) Primary (2000 Census)



(d) Primary by Cohort



(e) High School (2000 Census)



(f) High School by Cohort

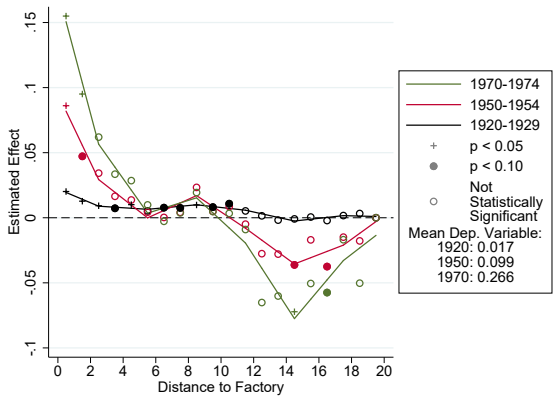
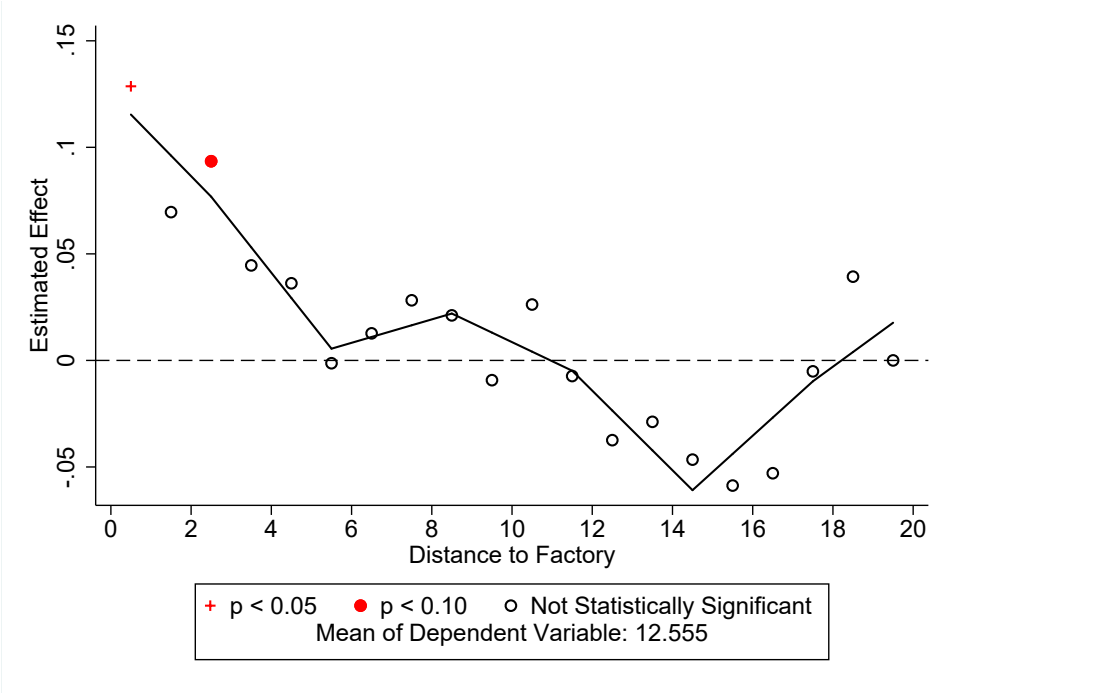


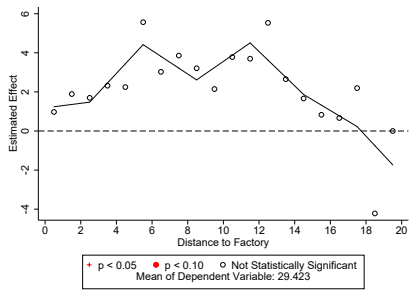
Figure F-7: Expenditure (2001-11)



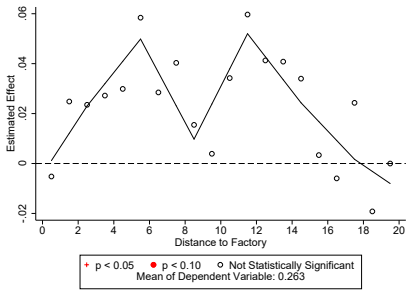
G Don't Subtract Counterfactual Means

Figure G-1: Geography

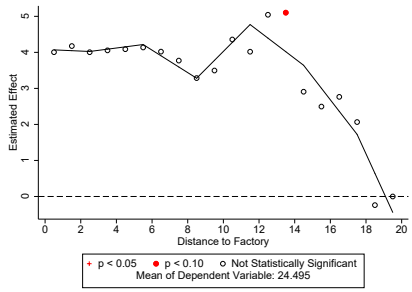
(a) Elevation



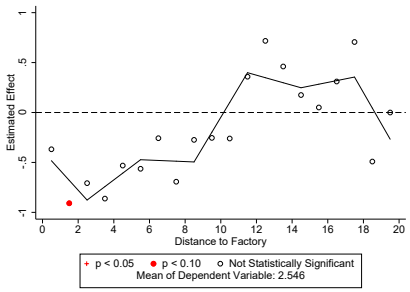
(b) Slope



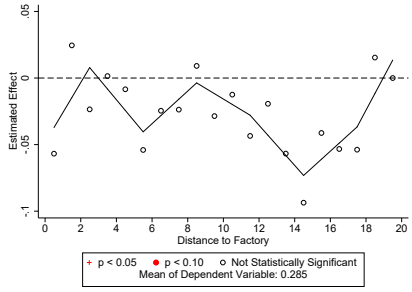
(c) Distance to Coast



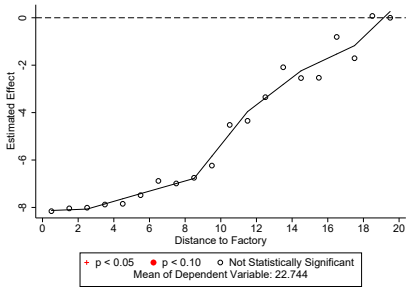
(d) Log Flow Accumulation



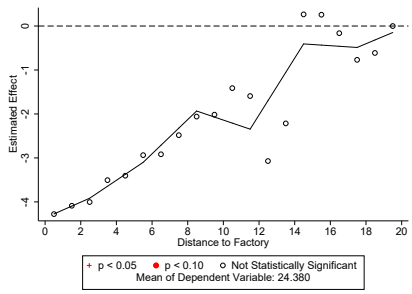
(e) Distance to River



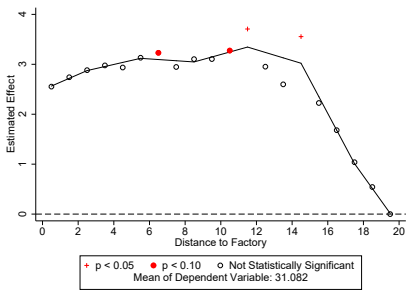
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



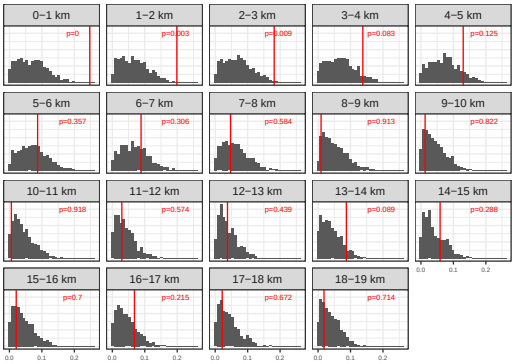
(h) Distance to VOC Port



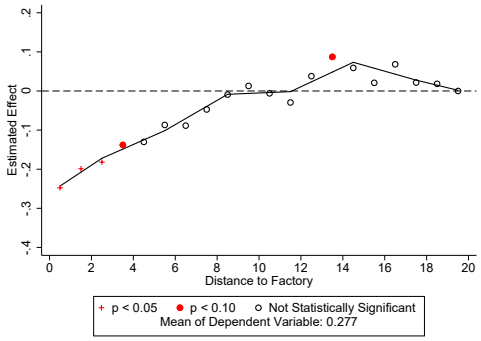
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The points are fit with a linear spline. p-values compare the effect proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-2: Share in Agriculture (2001-11): Illustration of Methodology

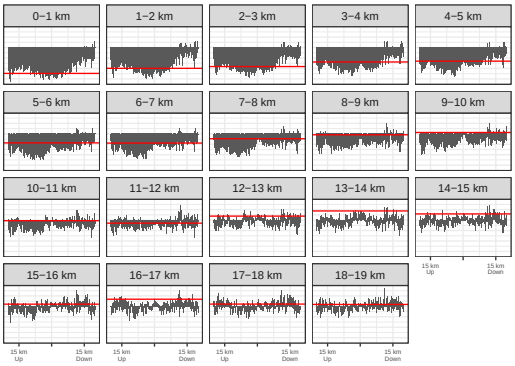
(a) Independent Shifts: Counterfactuals



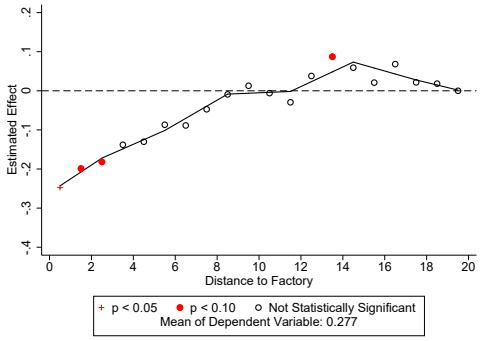
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



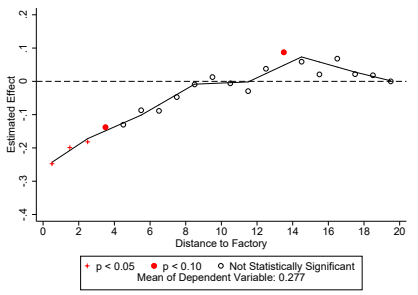
(d) Common Shifts: Plotted Coefficients



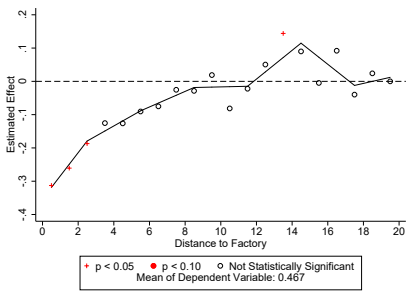
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots real coefficients for each bin, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots real coefficients for each bin, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure G-3: Industry and Agglomeration

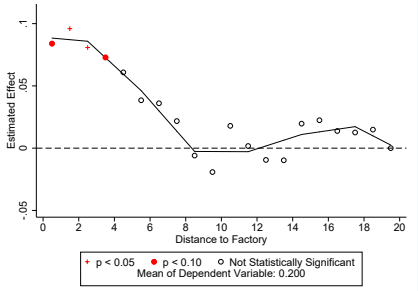
(a) Agriculture (Susenas 2001-11)



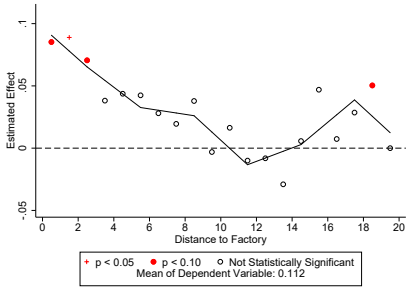
(b) Agriculture (Census 1980)



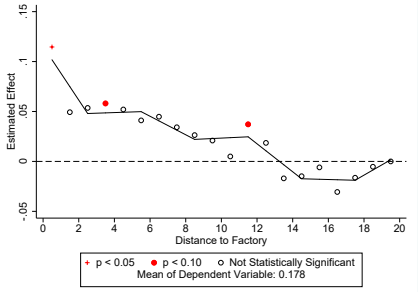
(c) Manufacturing (Susenas 2001-11)



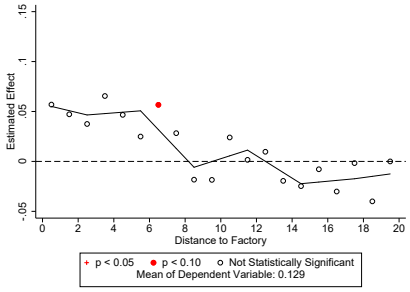
(d) Manufacturing (Census 1980)



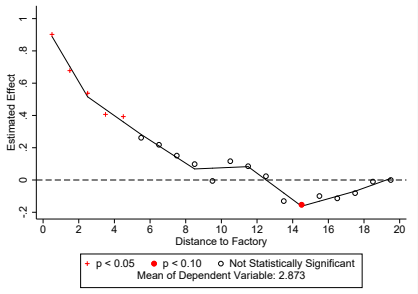
(e) Retail (Susenas 2001-11)



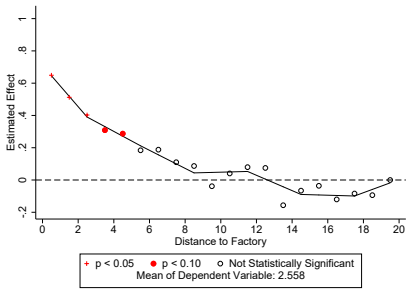
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



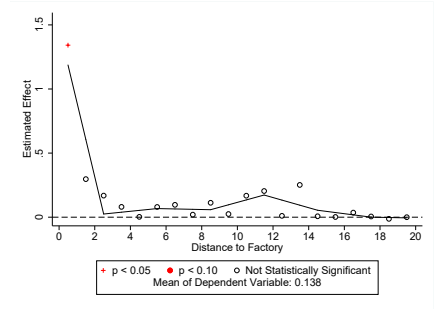
(h) Log Pop. Density (PODES 1980)



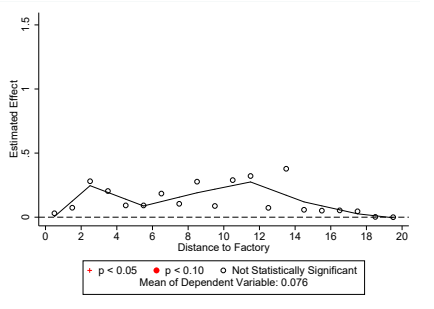
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearestcounterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-4: Sugar and Linked Industries

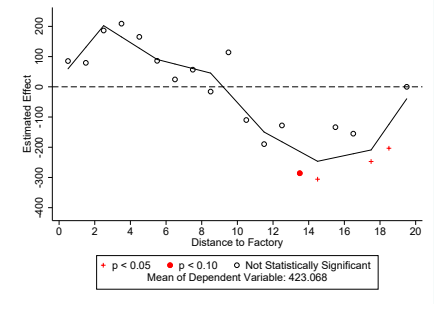
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



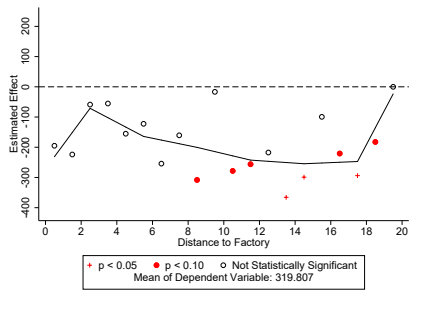
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



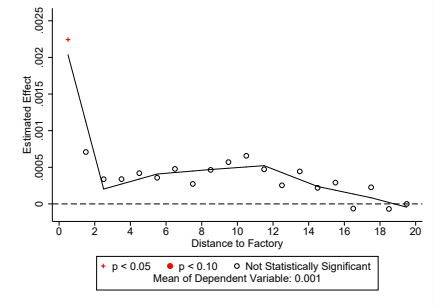
(c) Tons of Cane Grown (Full Sample, PODES 2003)



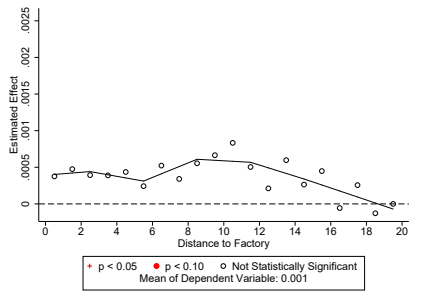
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



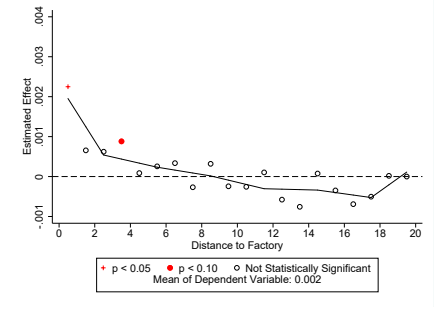
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



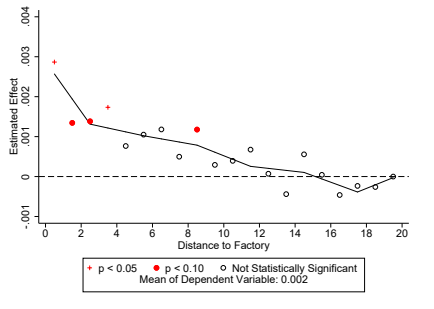
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



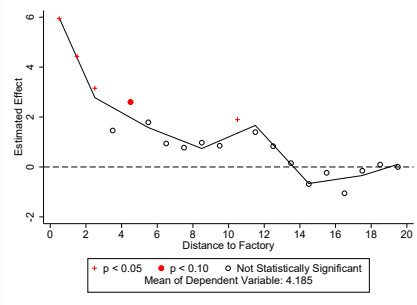
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



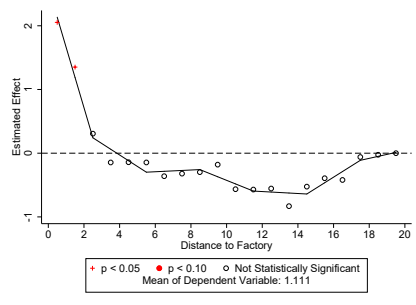
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-5: Infrastructure

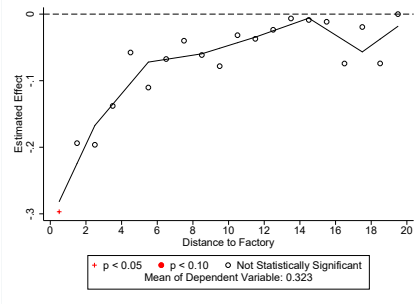
(a) Colonial Road Density (1900)



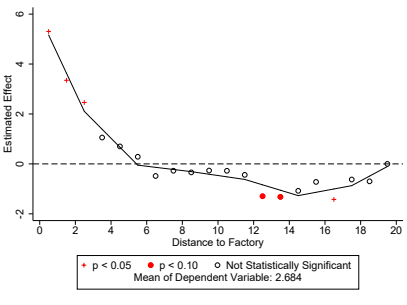
(b) Colonial Railroad Density (1900)



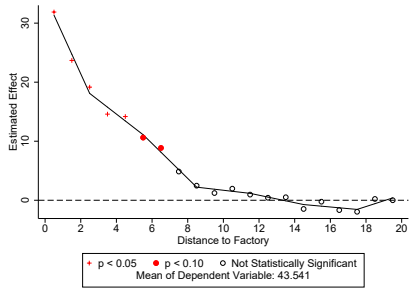
(c) Dirt Road (PODES 1980)



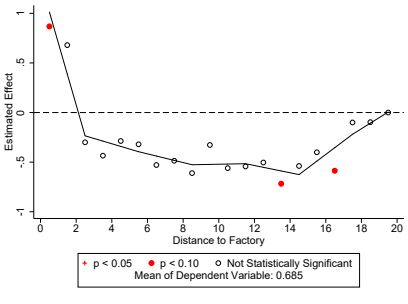
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



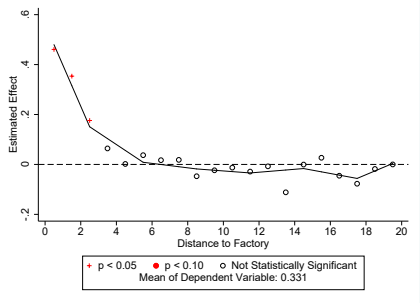
(f) Railroad Density (2017)



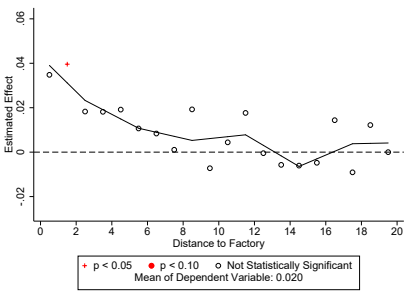
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-6: Other Public Goods

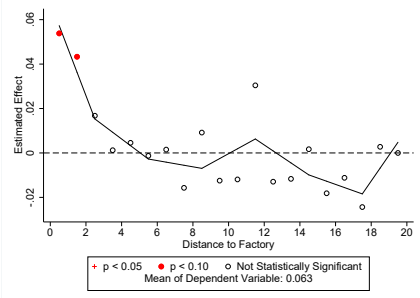
(a) Village Has Electricity (PODES 1980)



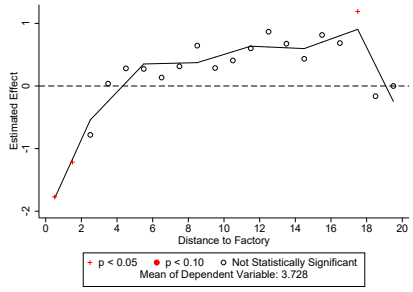
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



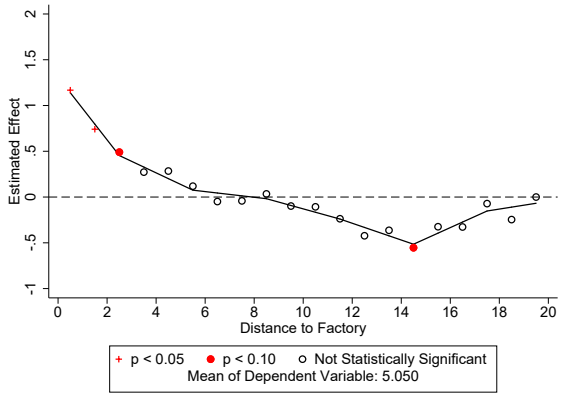
(d) Distance to Subdistrict Capital (2011 PODES)



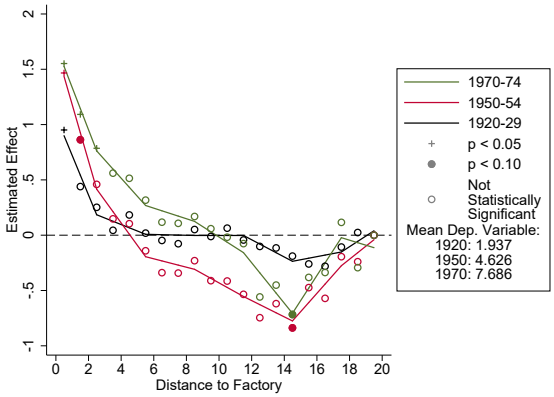
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-7: Education

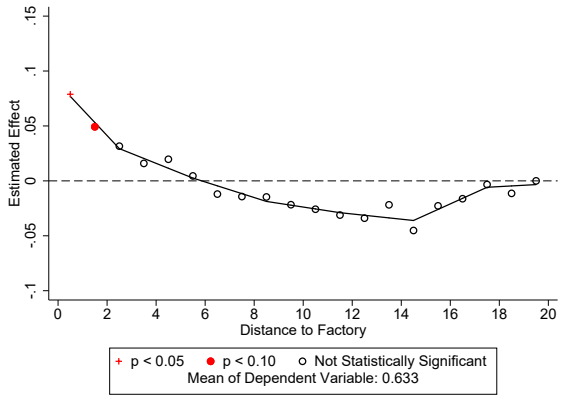
(a) Years Education (2000 Census)



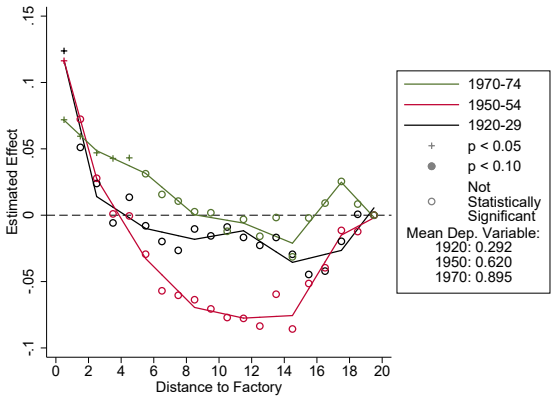
(b) Years Education by Cohort



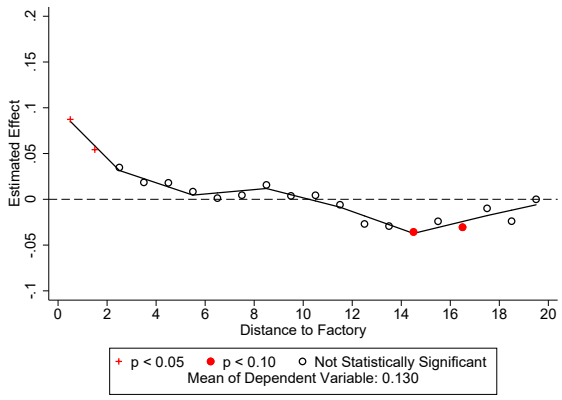
(c) Primary (2000 Census)



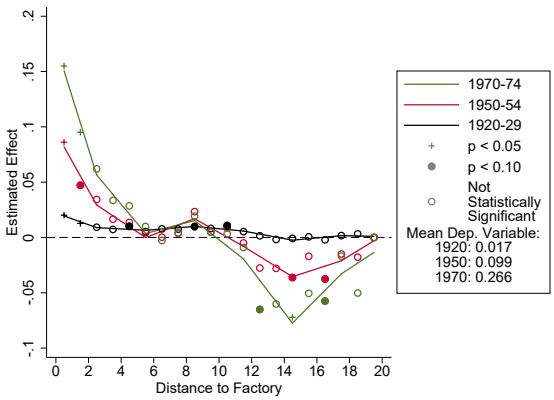
(d) Primary by Cohort



(e) High School (2000 Census)

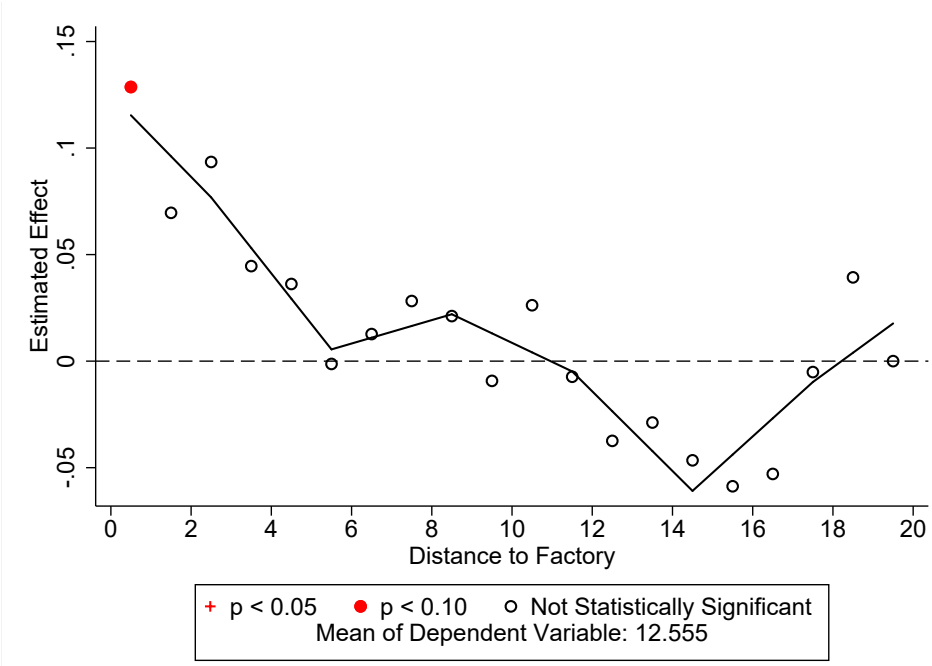


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure G-8: Expenditure (2001-11)

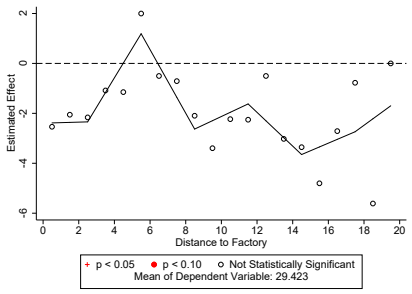


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

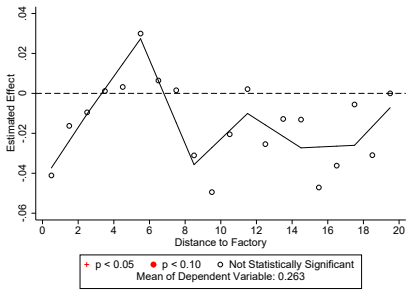
H Common Shifts

Figure H-1: Geography

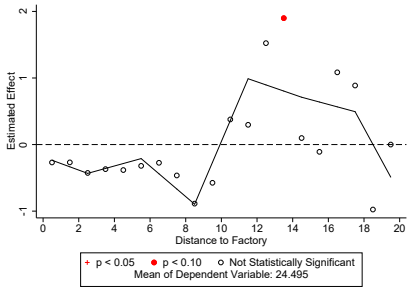
(a) Elevation



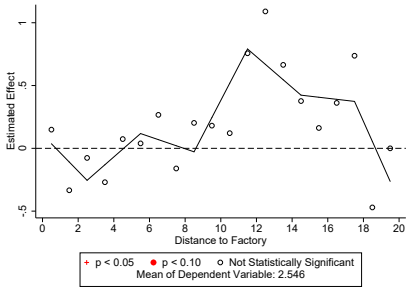
(b) Slope



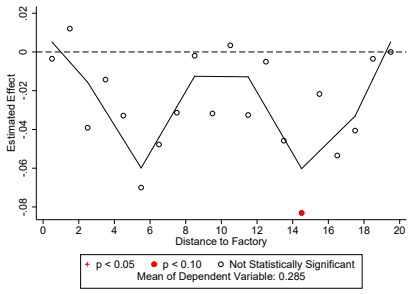
(c) Distance to Coast



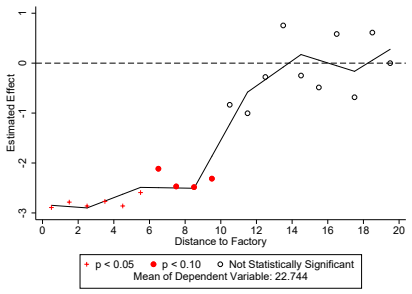
(d) Log Flow Accumulation



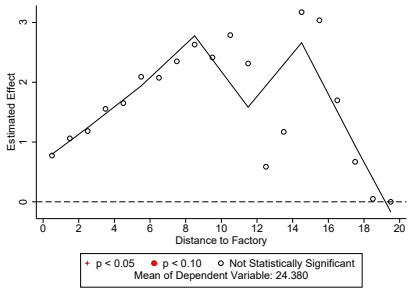
(e) Distance to River



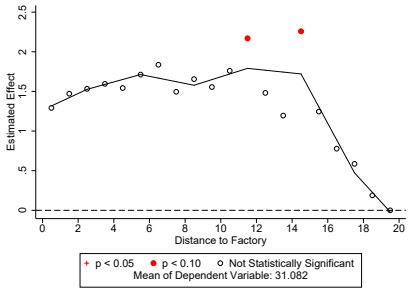
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



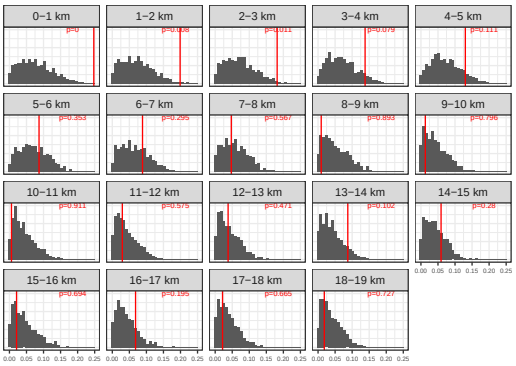
(h) Distance to VOC Port



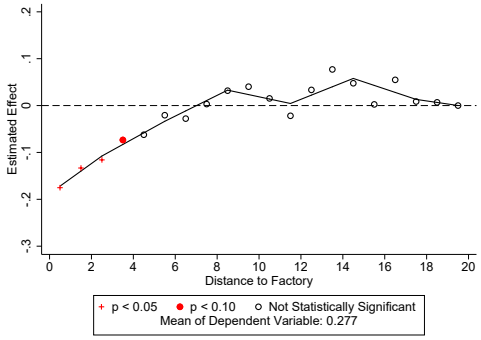
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-2: Share in Agriculture (2001-11): Illustration of Methodology

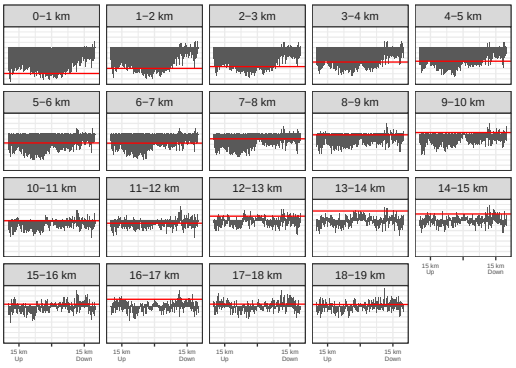
(a) Independent Shifts: Counterfactuals



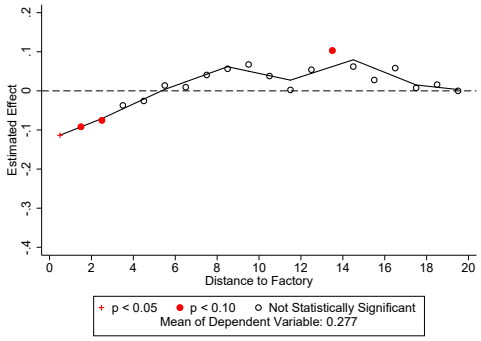
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



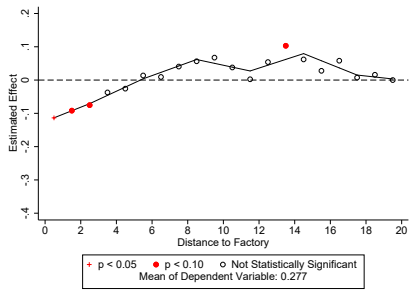
(d) Common Shifts: Plotted Coefficients



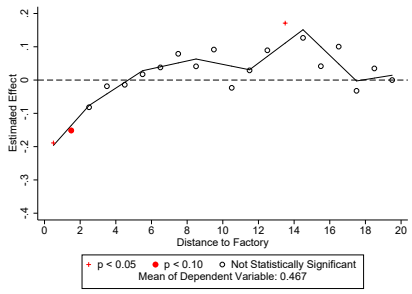
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure H-3: Industry and Agglomeration

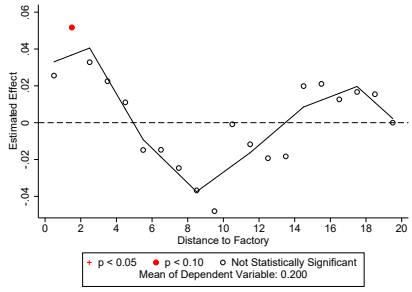
(a) Agriculture (Susenas 2001-11)



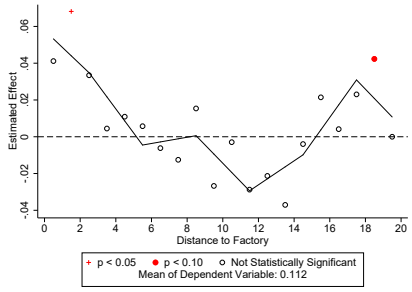
(b) Agriculture (Census 1980)



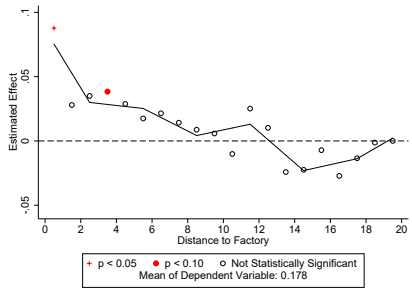
(c) Manufacturing (Susenas 2001-11)



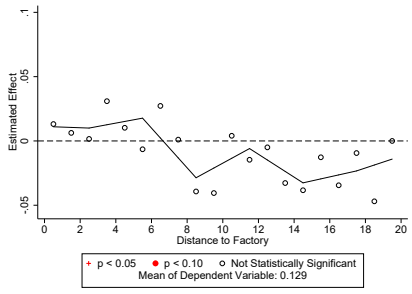
(d) Manufacturing (Census 1980)



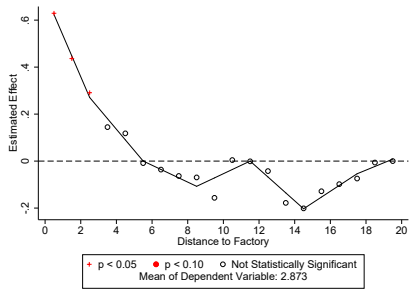
(e) Retail (Susenas 2001-11)



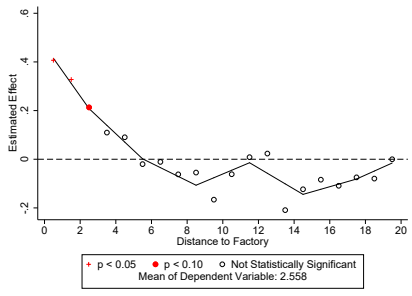
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



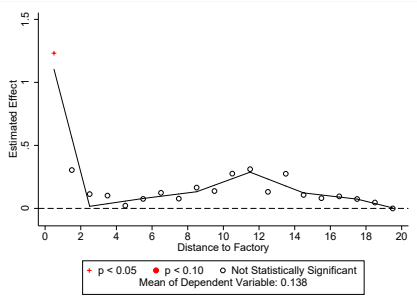
(h) Log Pop. Density (PODES 1980)



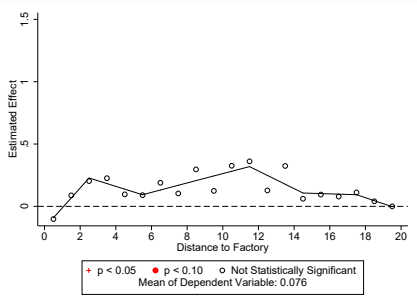
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearestcounterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-4: Sugar and Linked Industries

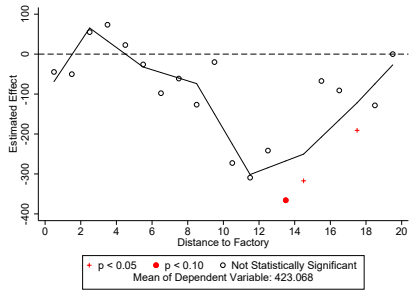
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



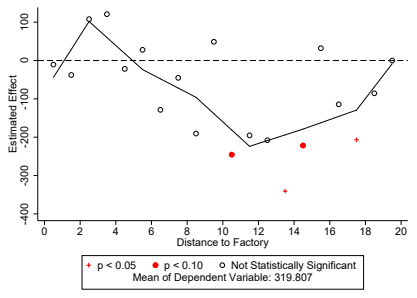
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



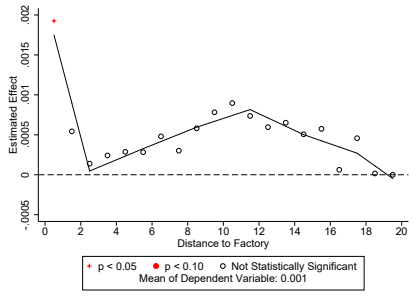
(c) Tons of Cane Grown (Full Sample, PODES 2003)



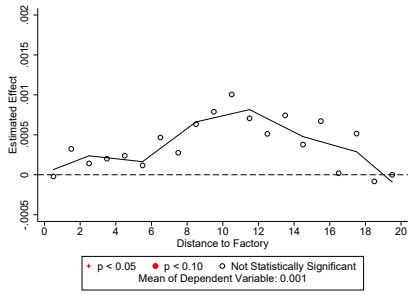
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



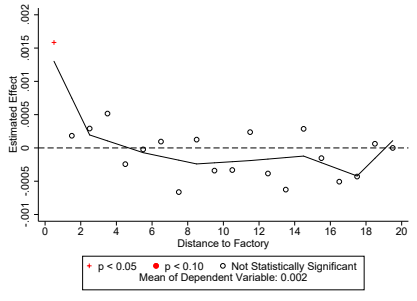
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



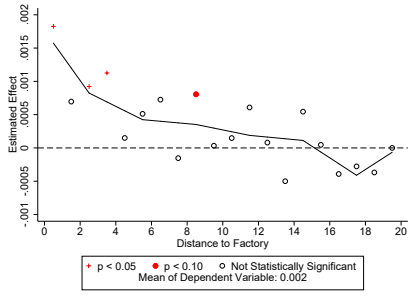
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



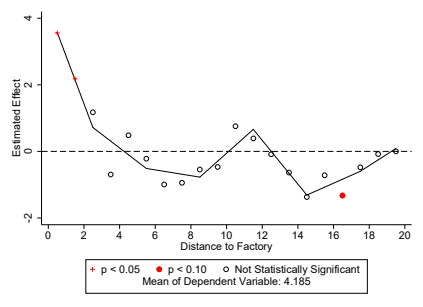
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



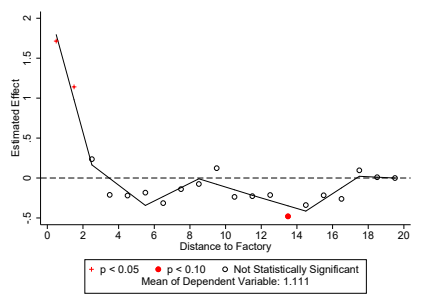
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-5: Infrastructure

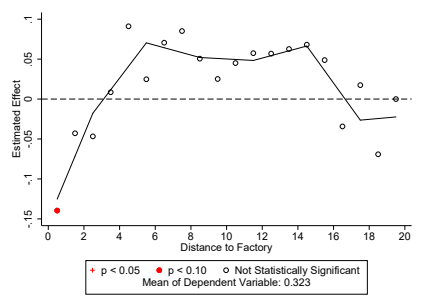
(a) Colonial Road Density (1900)



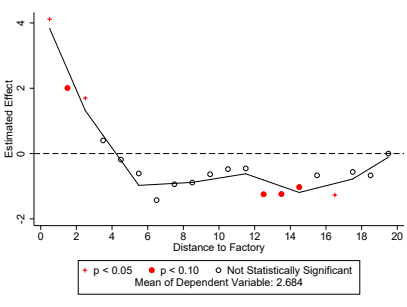
(b) Colonial Railroad Density (1900)



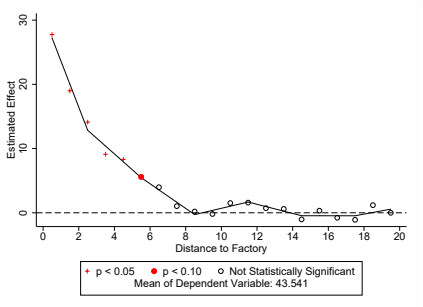
(c) Dirt Road (PODES 1980)



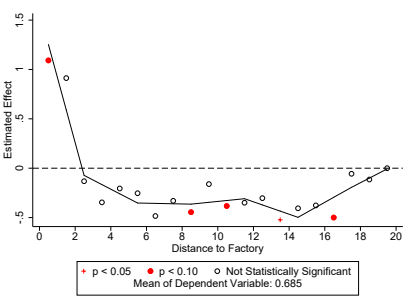
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



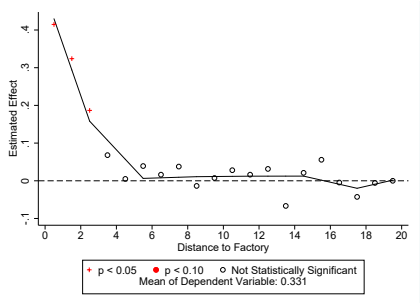
(f) Railroad Density (2017)



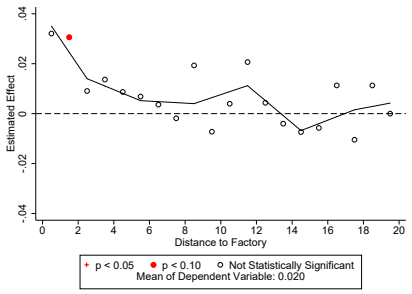
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-6: Other Public Goods

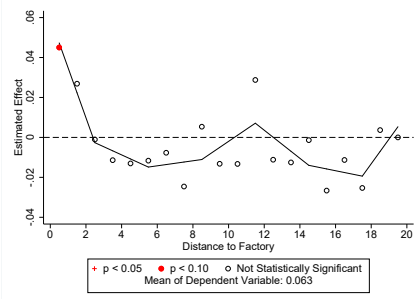
(a) Village Has Electricity (PODES 1980)



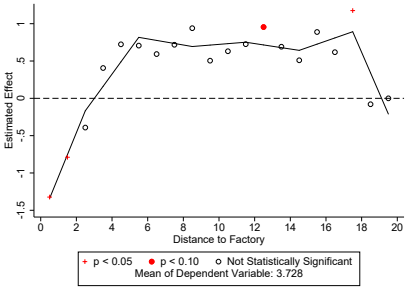
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



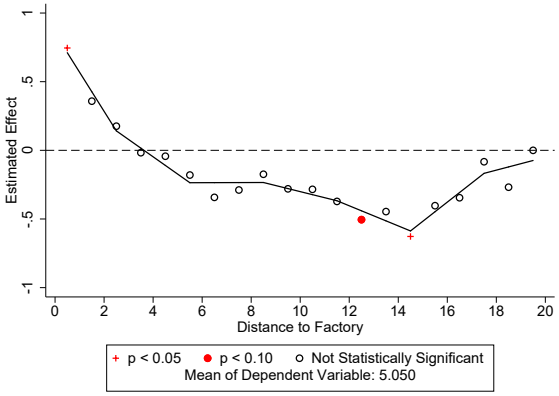
(d) Distance to Subdistrict Capital (2011 PODES)



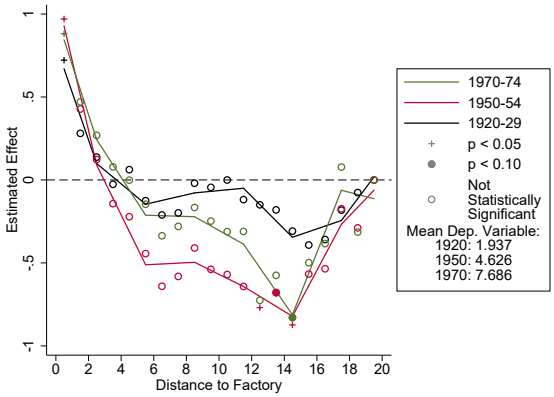
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-7: Education

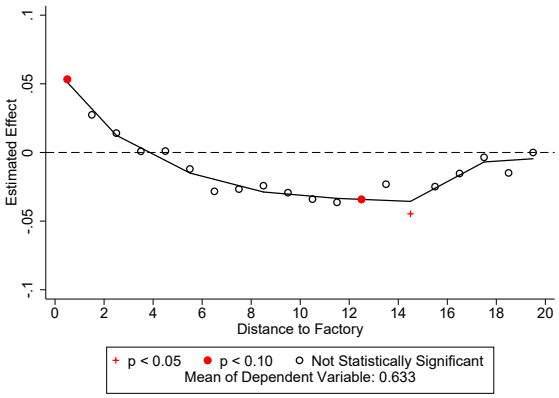
(a) Years Education (2000 Census)



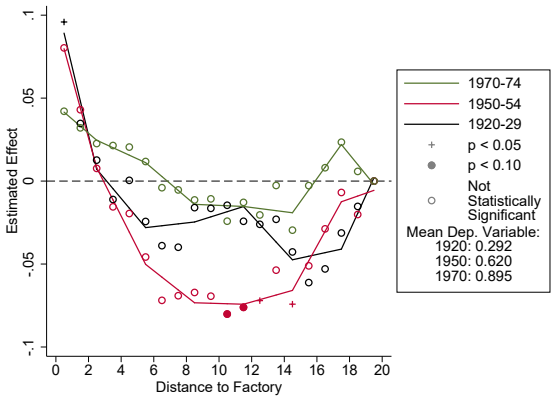
(b) Years Education by Cohort



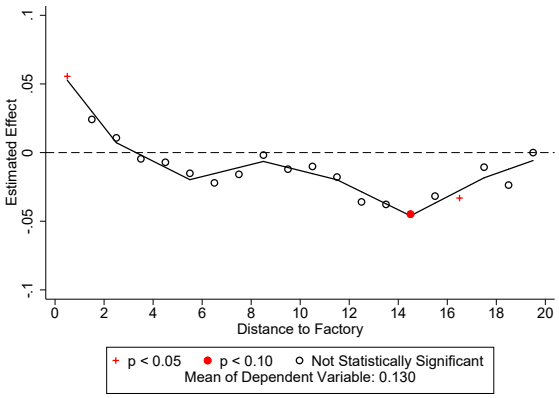
(c) Primary (2000 Census)



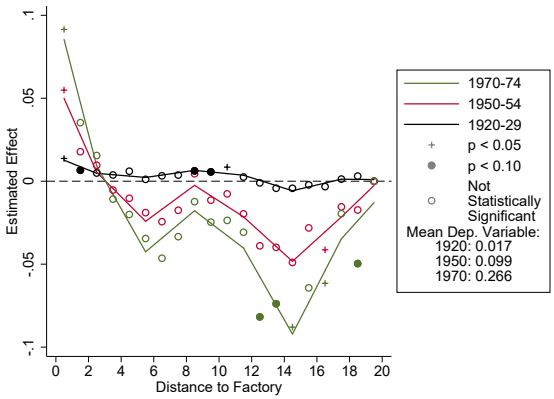
(d) Primary by Cohort



(e) High School (2000 Census)

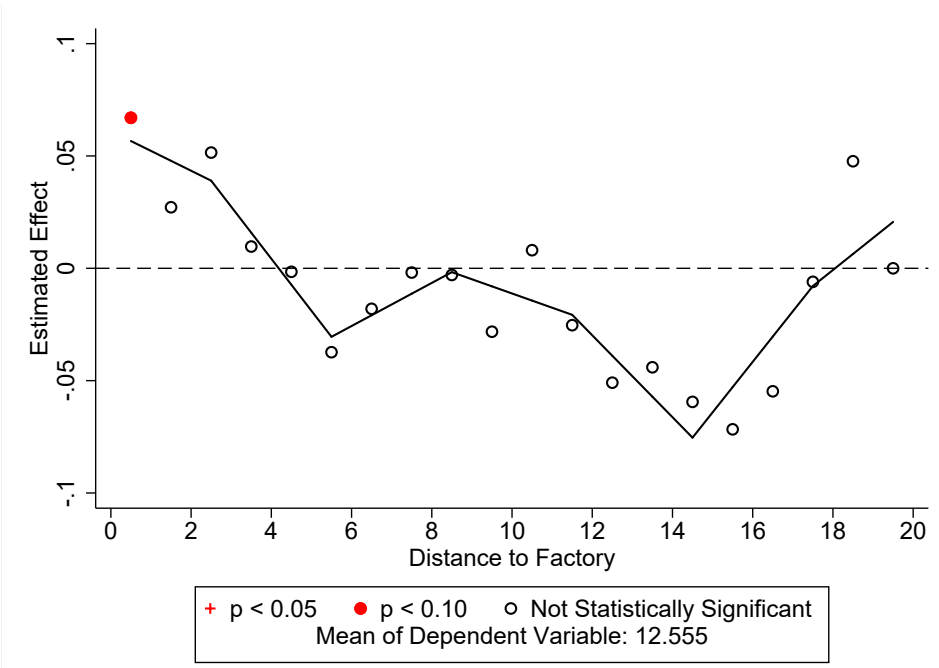


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

Figure H-8: Expenditure (2001-11)

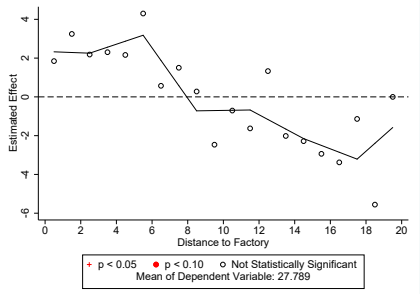


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 200 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 200 counterfactual factory configurations.

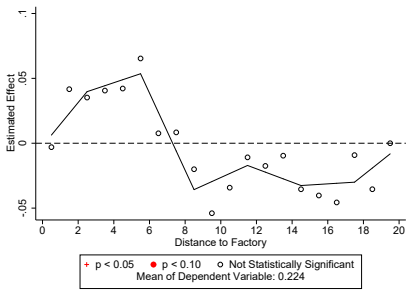
I Counterfactuals at Least 5km From All Actual Factories

Figure I-1: Geography

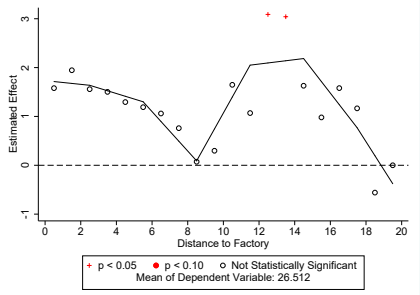
(a) Elevation



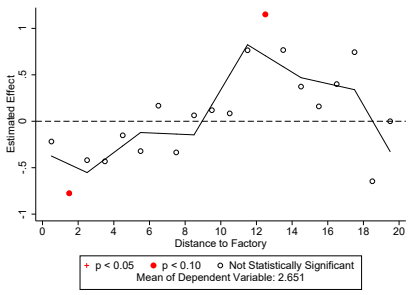
(b) Slope



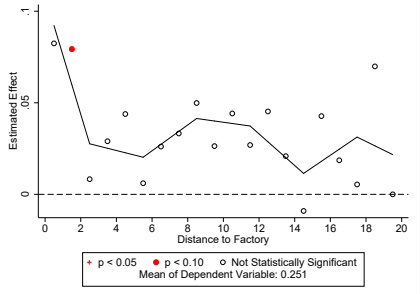
(c) Distance to Coast



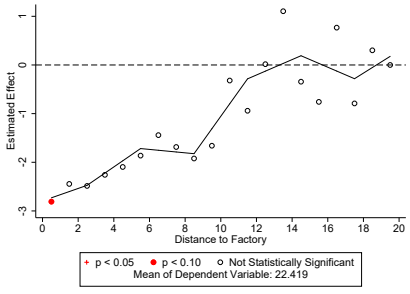
(d) Log Flow Accumulation



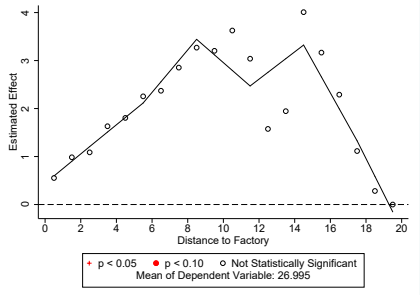
(e) Distance to River



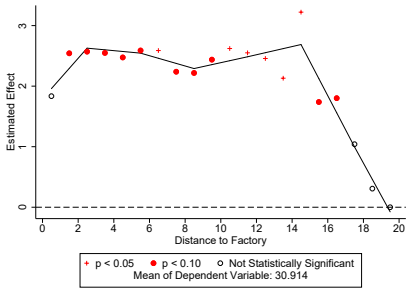
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



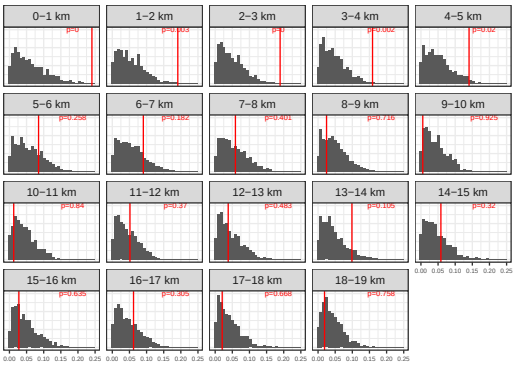
(h) Distance to VOC Port



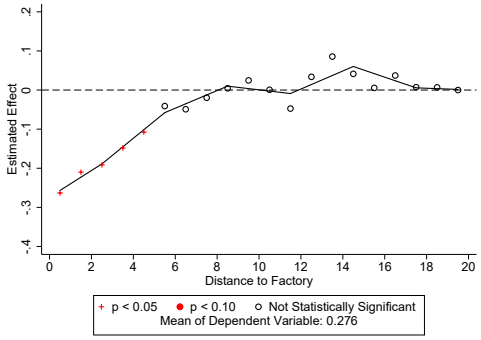
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-2: Share in Agriculture (2001-11): Illustration of Methodology

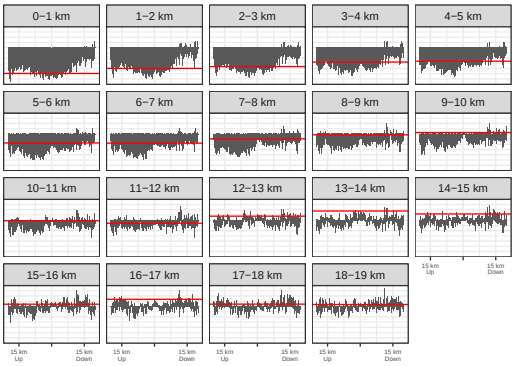
(a) Independent Shifts: Counterfactuals



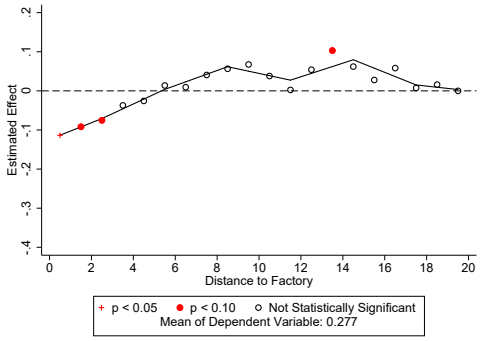
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



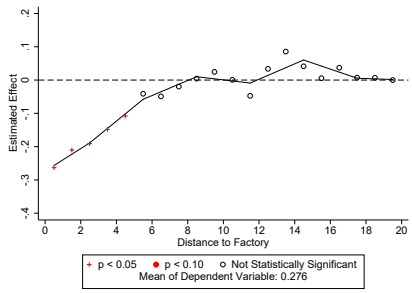
(d) Common Shifts: Plotted Coefficients



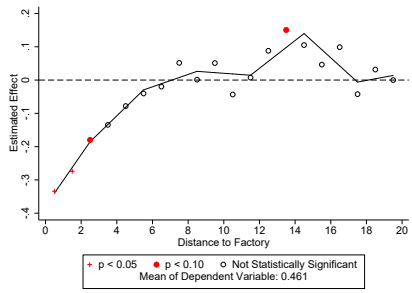
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure I-3: Industry and Agglomeration

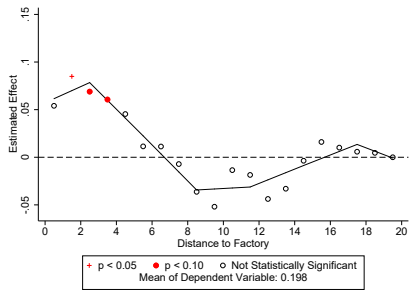
(a) Agriculture (Susenas 2001-11)



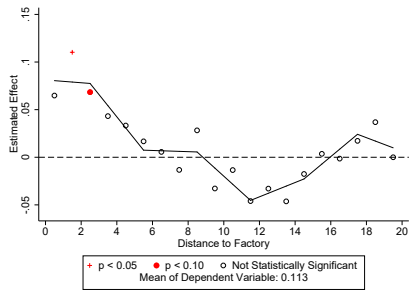
(b) Agriculture (Census 1980)



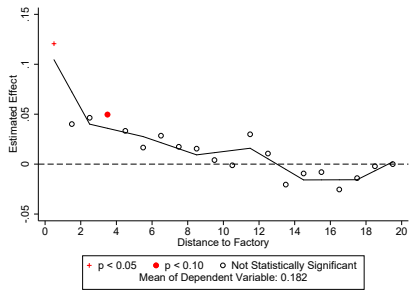
(c) Manufacturing (Susenas 2001-11)



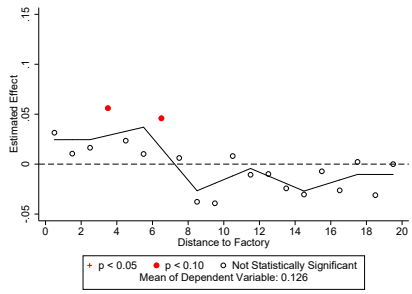
(d) Manufacturing (Census 1980)



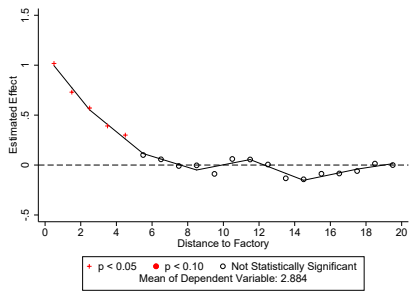
(e) Retail (Susenas 2001-11)



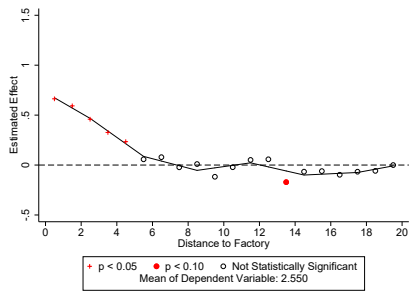
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



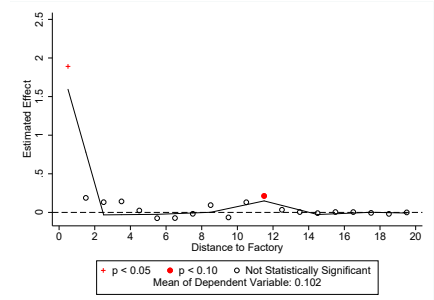
(h) Log Pop. Density (PODES 1980)



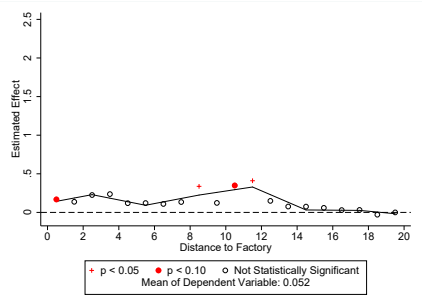
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-4: Sugar and Linked Industries

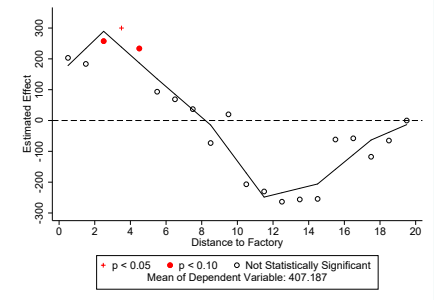
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



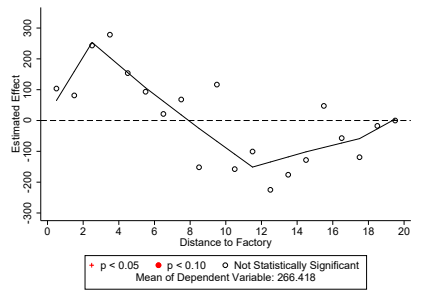
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



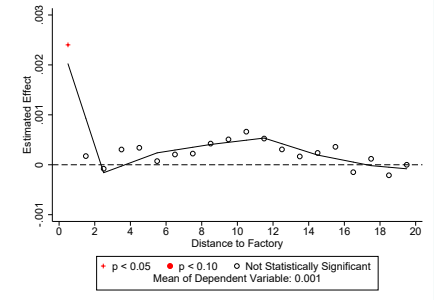
(c) Tons of Cane Grown (Full Sample, PODES 2003)



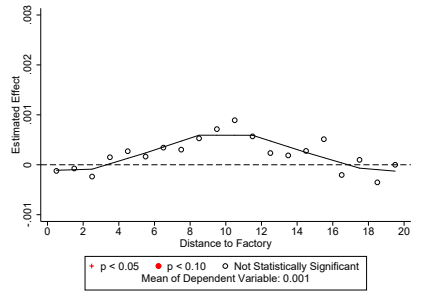
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



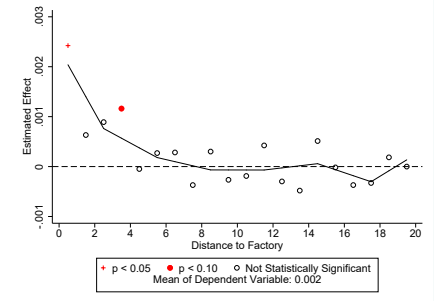
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



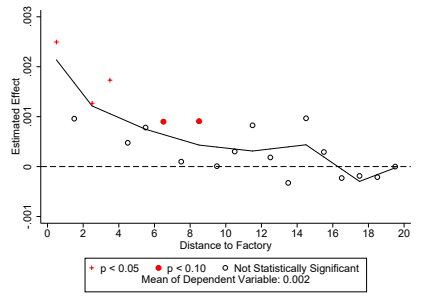
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



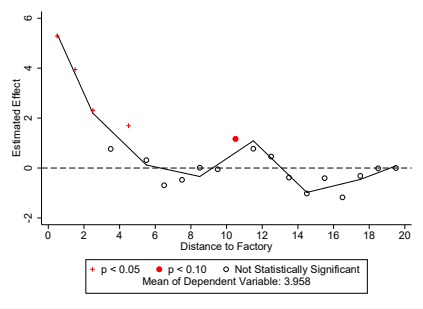
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



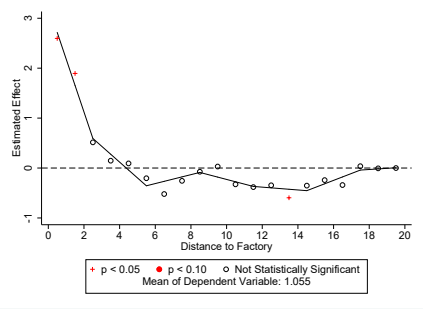
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-5: Infrastructure

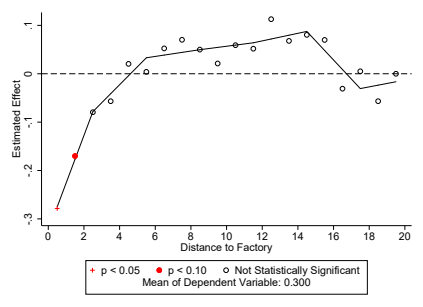
(a) Colonial Road Density (1900)



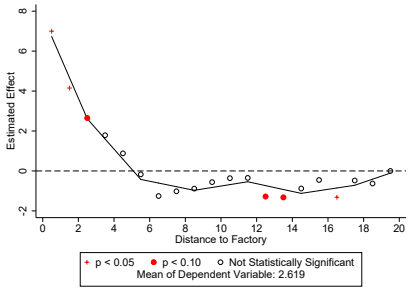
(b) Colonial Railroad Density (1900)



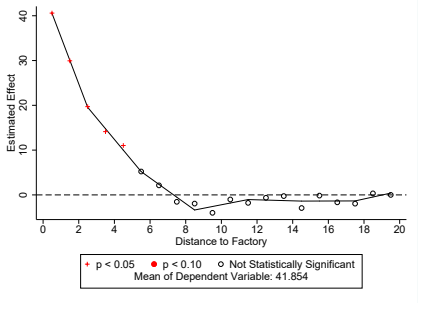
(c) Dirt Road (PODES 1980)



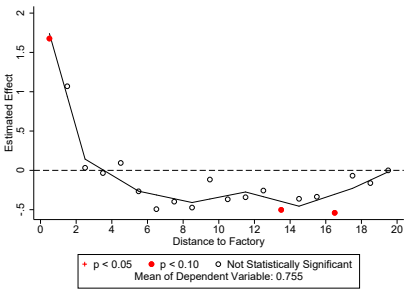
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



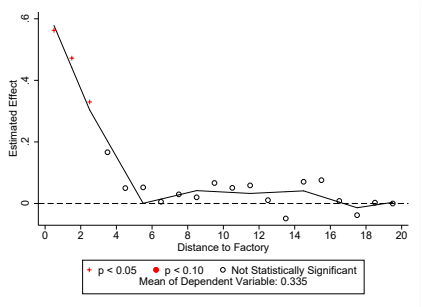
(f) Railroad Density (2017)



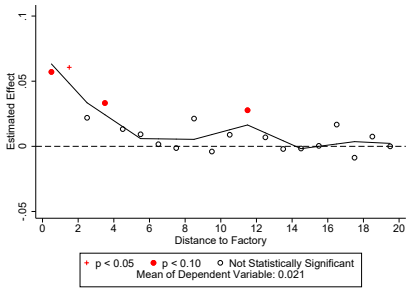
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-6: Other Public Goods

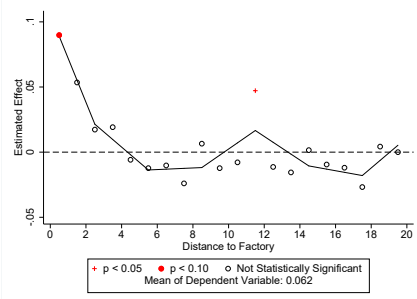
(a) Village Has Electricity (PODES 1980)



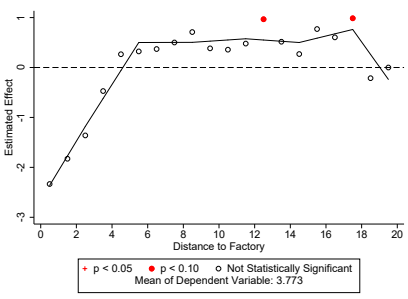
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



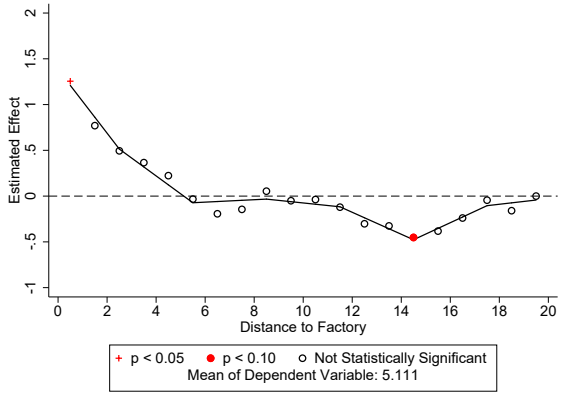
(d) Distance to Subdistrict Capital (2011 PODES)



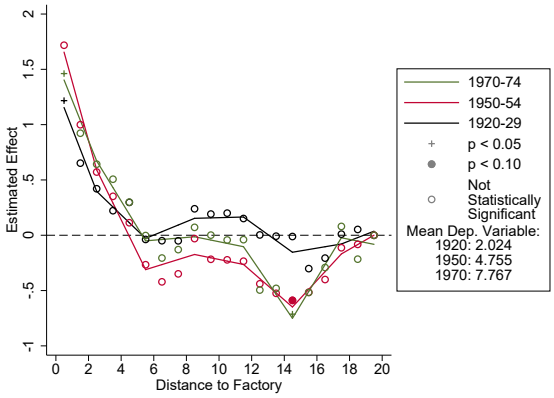
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-7: Education

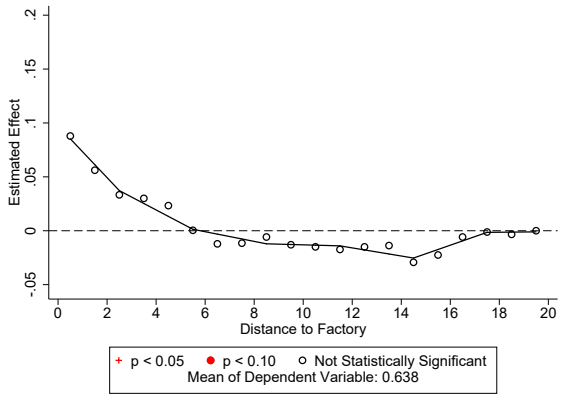
(a) Years Education (2000 Census)



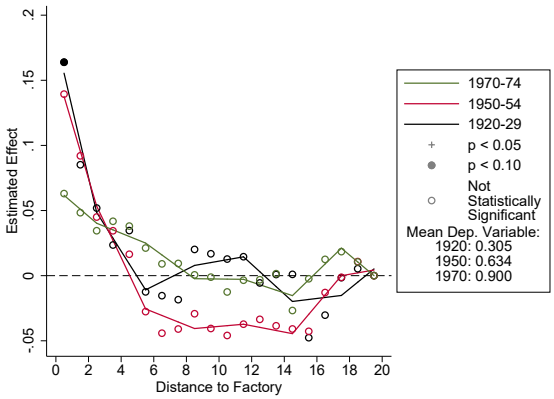
(b) Years Education by Cohort



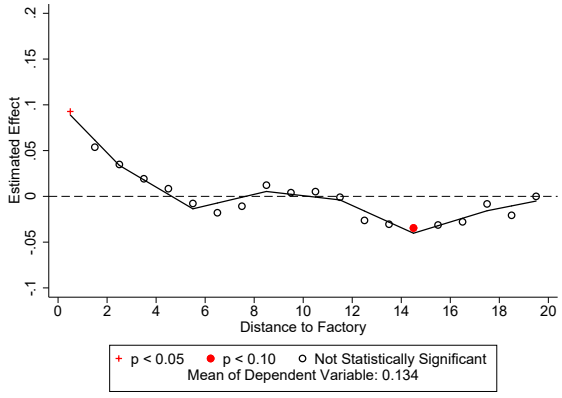
(c) Primary (2000 Census)



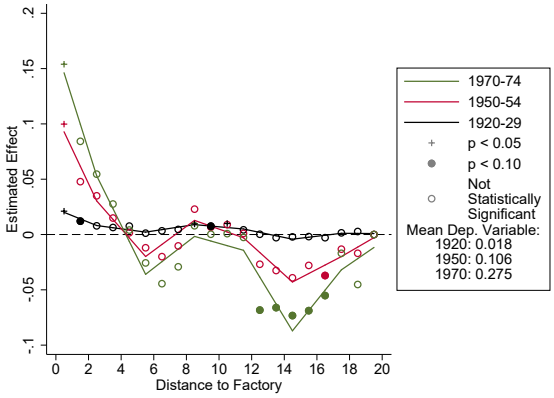
(d) Primary by Cohort



(e) High School (2000 Census)

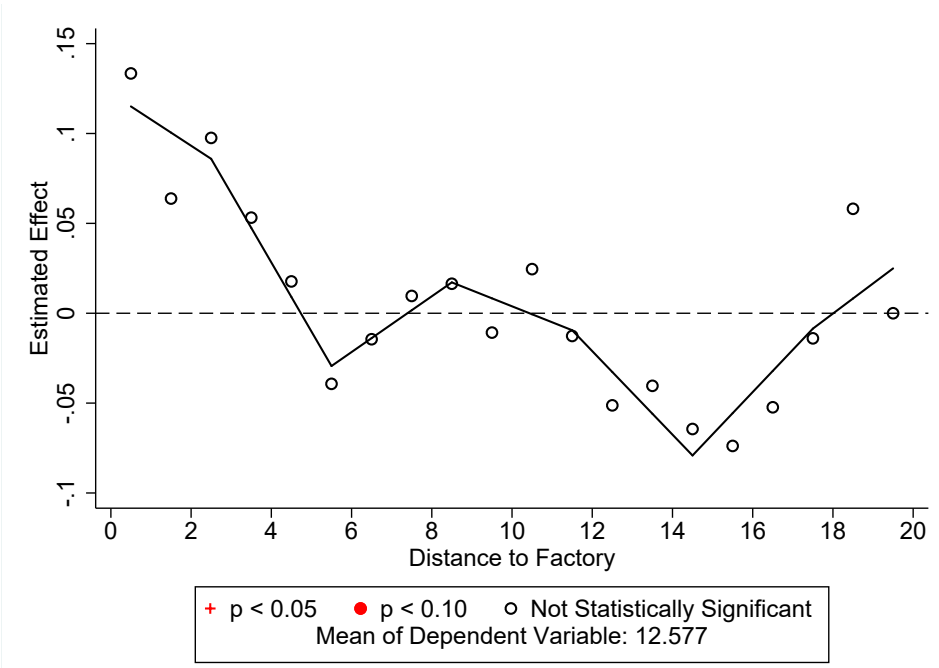


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure I-8: Expenditure (2001-11)

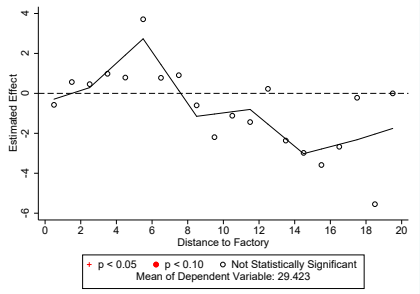


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

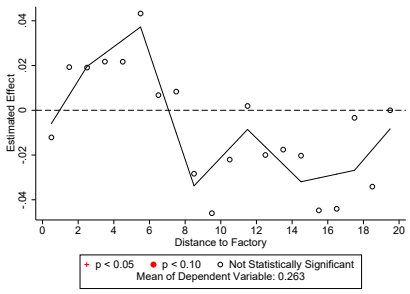
J No Balance Restriction

Figure J-1: Geography

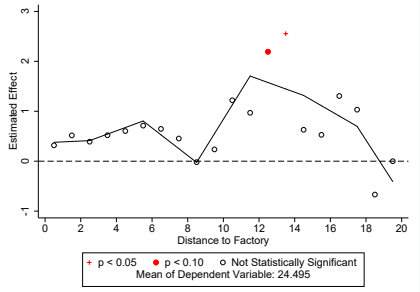
(a) Elevation



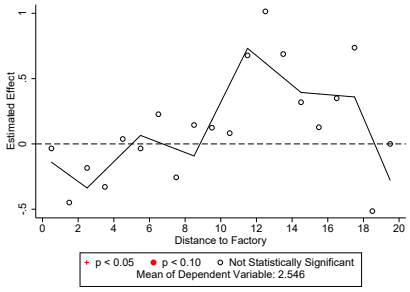
(b) Slope



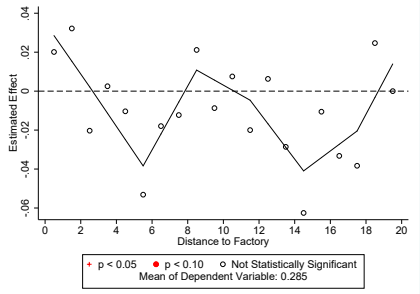
(c) Distance to Coast



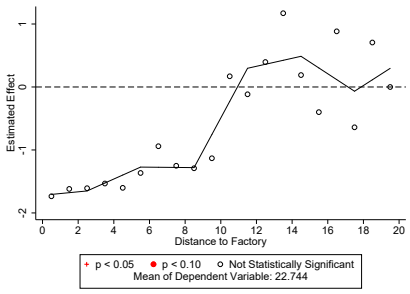
(d) Log Flow Accumulation



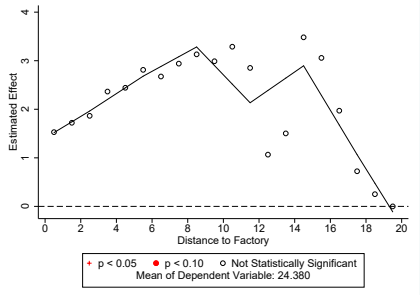
(e) Distance to River



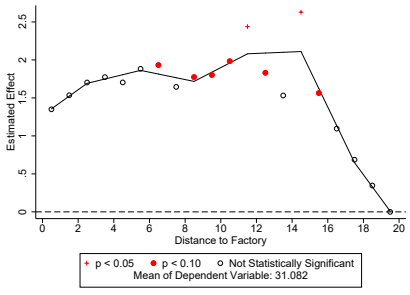
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



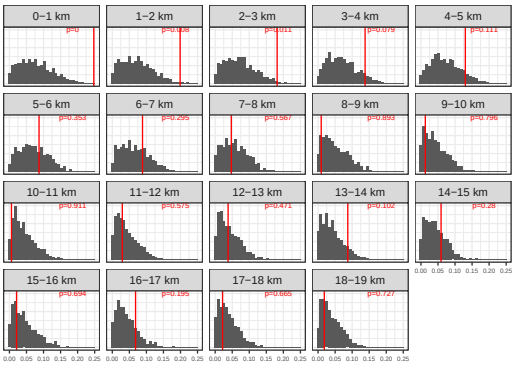
(h) Distance to VOC Port



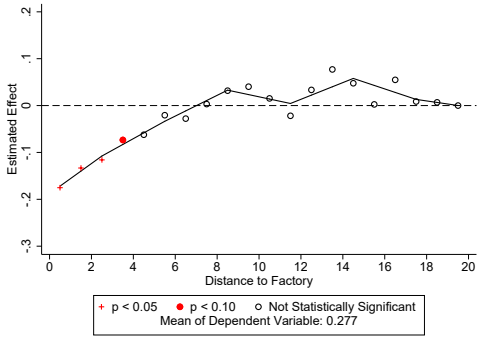
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-2: Share in Agriculture (2001-11): Illustration of Methodology

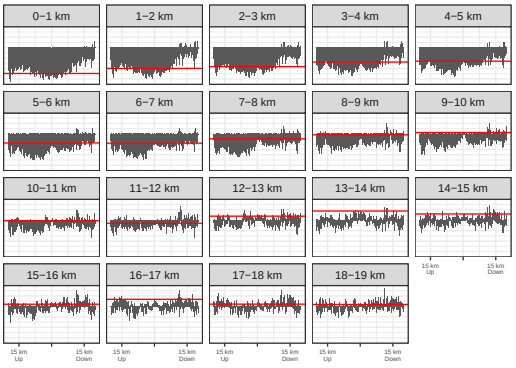
(a) Independent Shifts: Counterfactuals



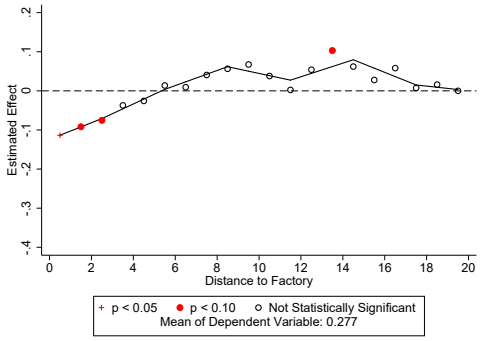
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



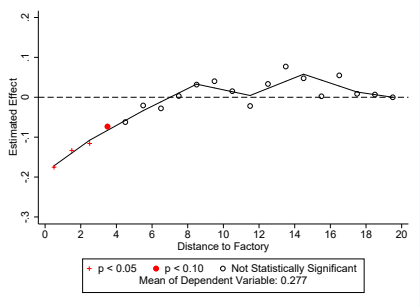
(d) Common Shifts: Plotted Coefficients



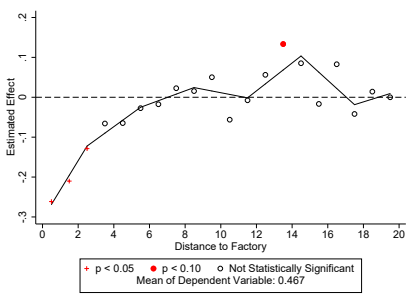
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure J-3: Industry and Agglomeration

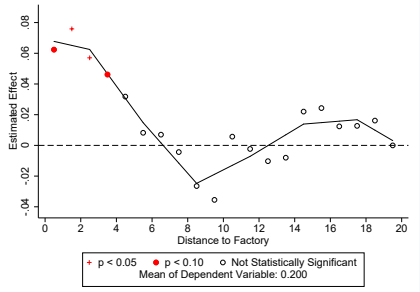
(a) Agriculture (Susenas 2001-11)



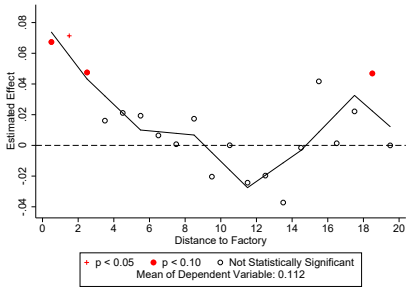
(b) Agriculture (Census 1980)



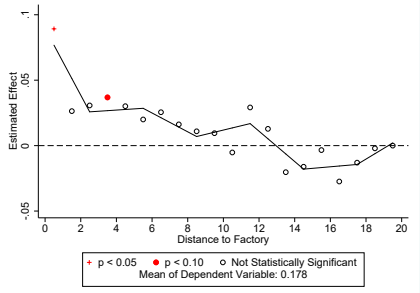
(c) Manufacturing (Susenas 2001-11)



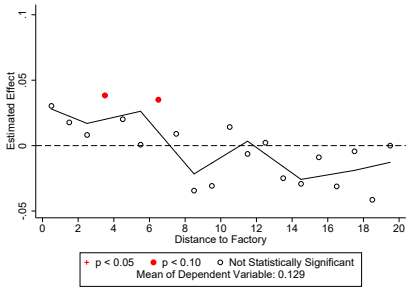
(d) Manufacturing (Census 1980)



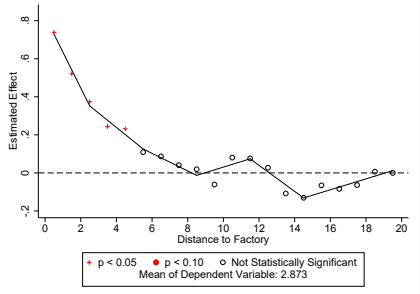
(e) Retail (Susenas 2001-11)



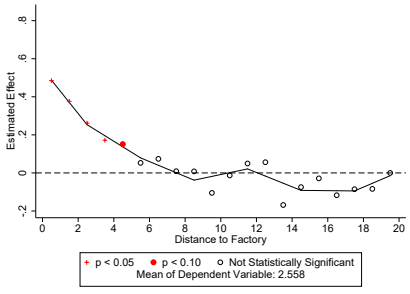
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



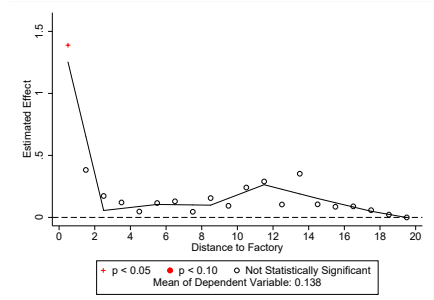
(h) Log Pop. Density (PODES 1980)



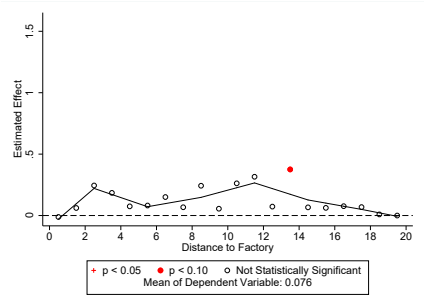
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-4: Sugar and Linked Industries

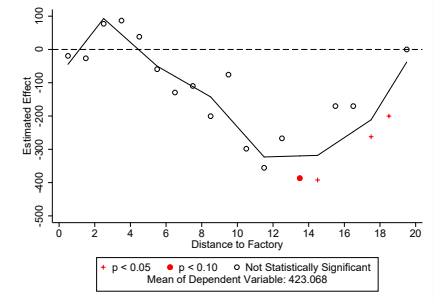
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



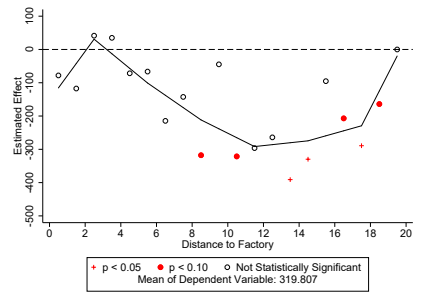
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



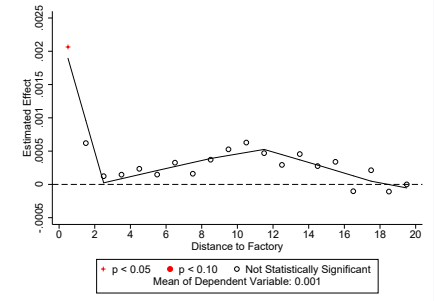
(c) Tons of Cane Grown (Full Sample, PODES 2003)



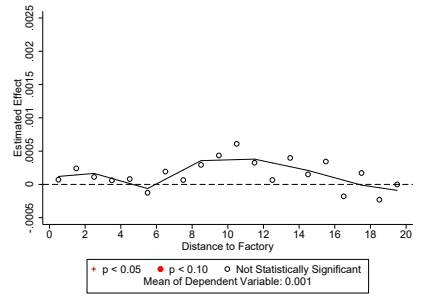
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



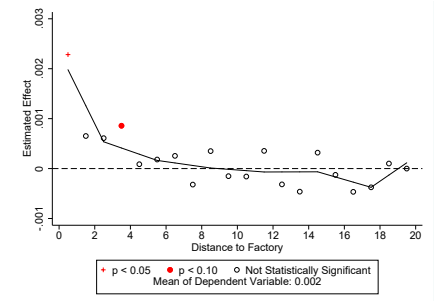
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



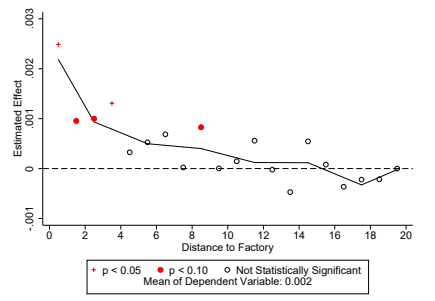
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



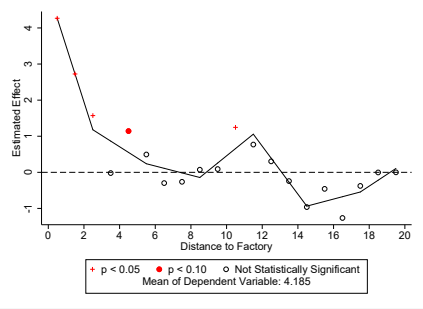
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



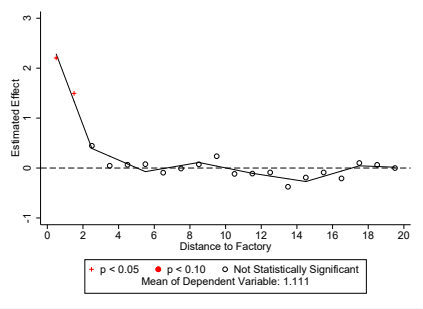
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-5: Infrastructure

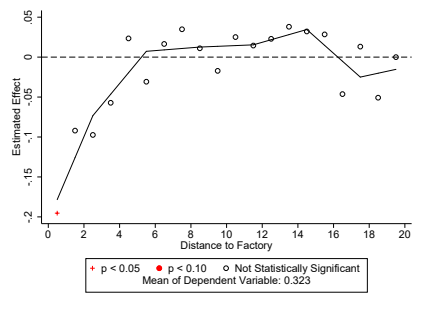
(a) Colonial Road Density (1900)



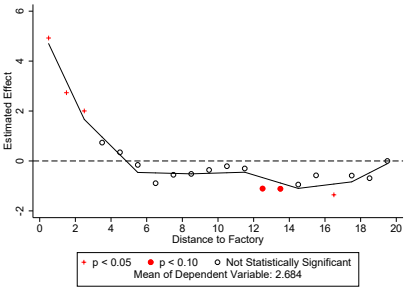
(b) Colonial Railroad Density (1900)



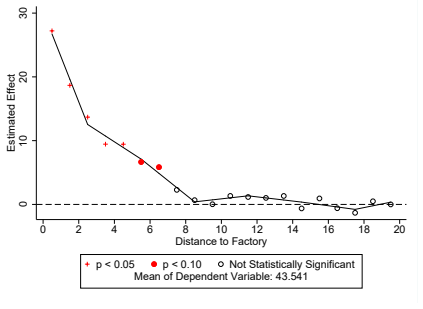
(c) Dirt Road (PODES 1980)



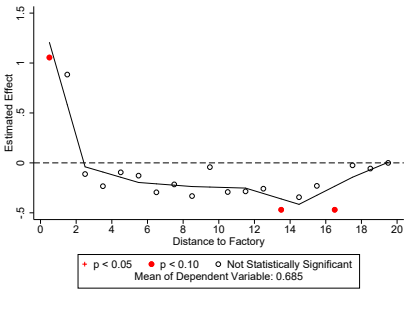
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



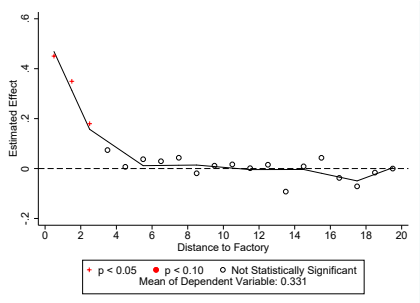
(f) Railroad Density (2017)



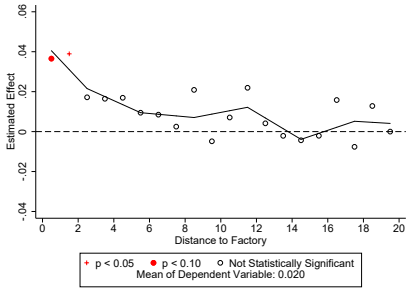
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-6: Other Public Goods

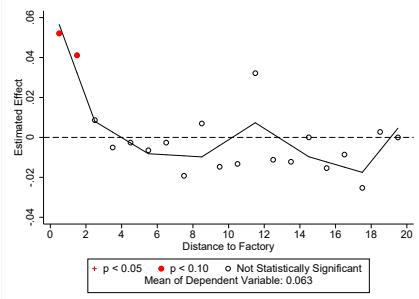
(a) Village Has Electricity (PODES 1980)



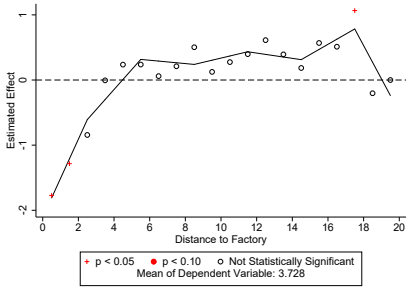
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



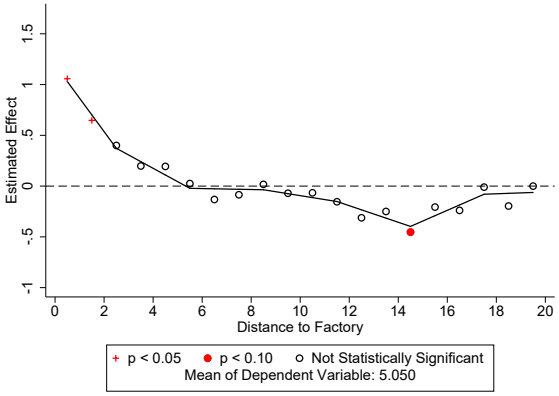
(d) Distance to Subdistrict Capital (2011 PODES)



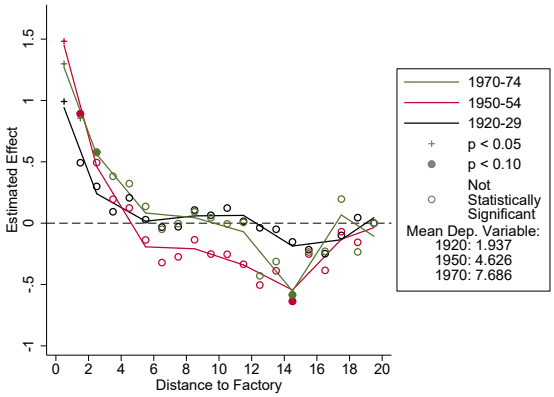
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-7: Education

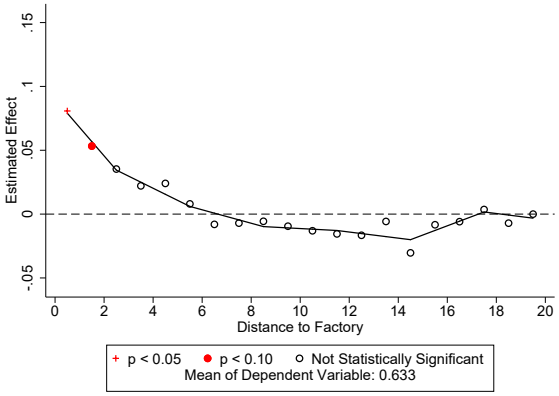
(a) Years Education (2000 Census)



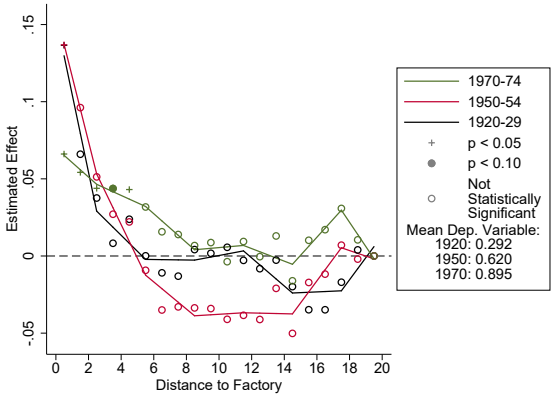
(b) Years Education by Cohort



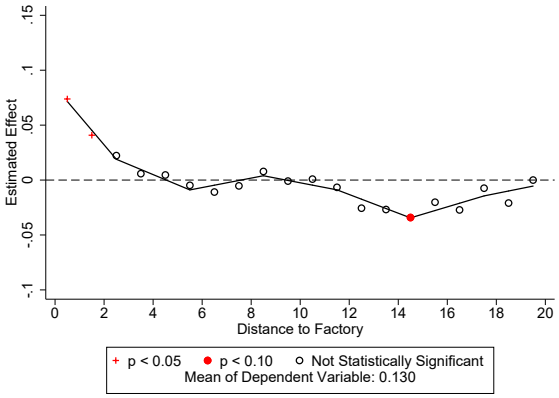
(c) Primary (2000 Census)



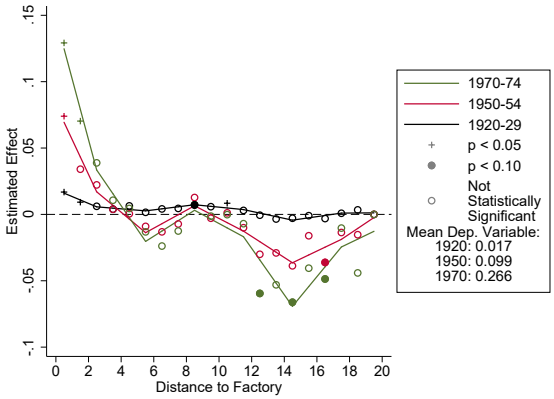
(d) Primary by Cohort



(e) High School (2000 Census)

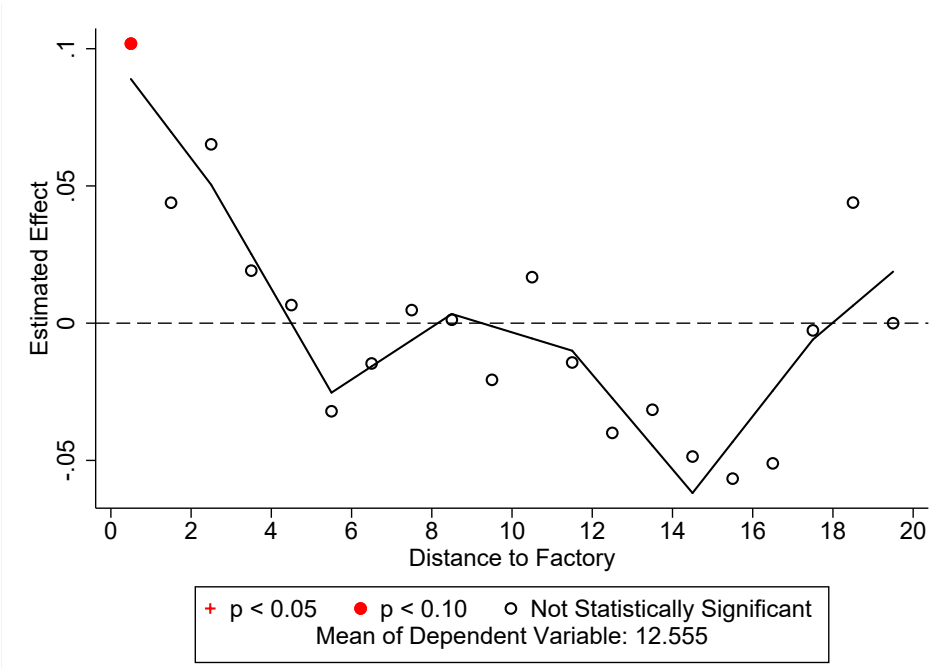


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure J-8: Expenditure (2001-11)

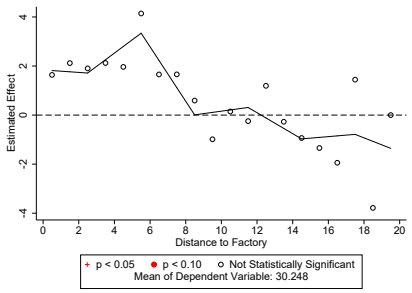


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

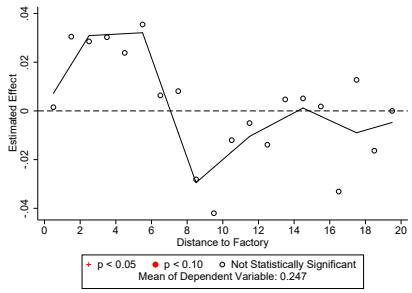
K Sample Includes Factories Near Modern Sugar Factories

Figure K-1: Geography

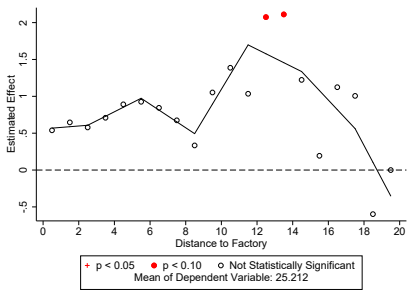
(a) Elevation



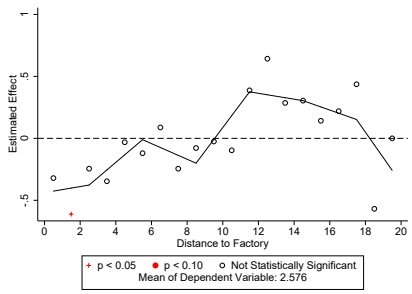
(b) Slope



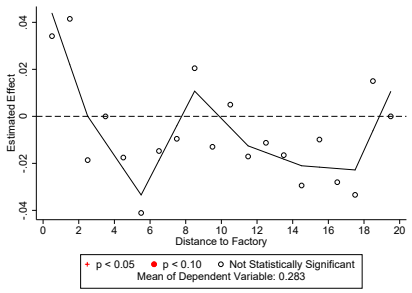
(c) Distance to Coast



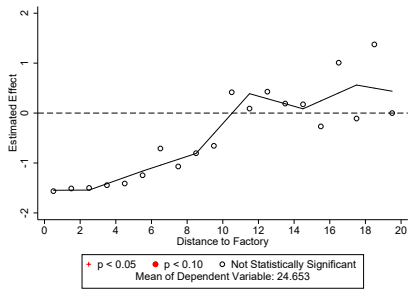
(d) Log Flow Accumulation



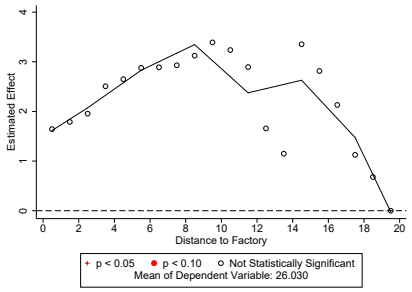
(e) Distance to River



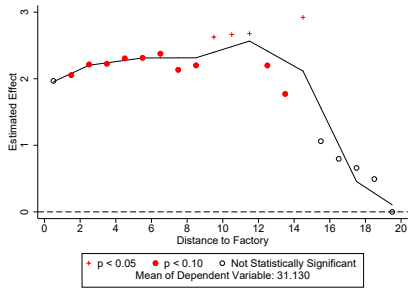
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



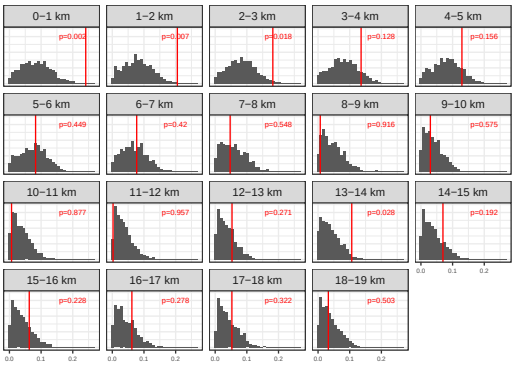
(h) Distance to VOC Port



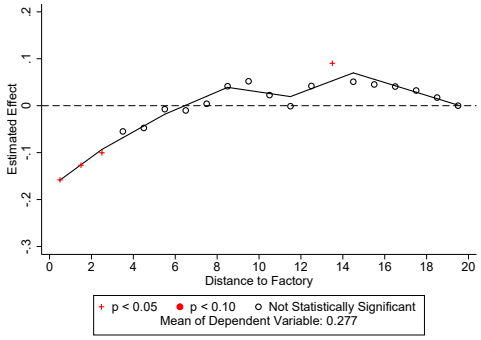
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-2: Share in Agriculture (2001-11): Illustration of Methodology

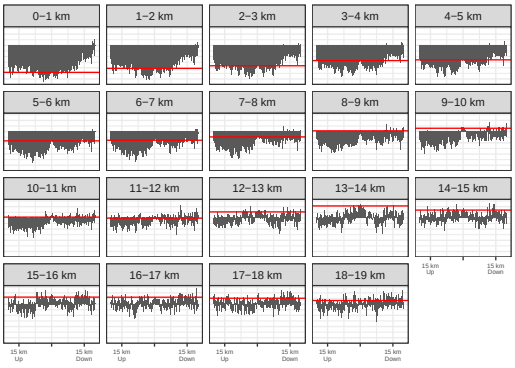
(a) Independent Shifts: Counterfactuals



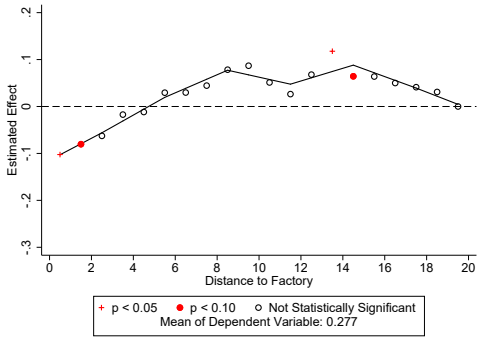
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



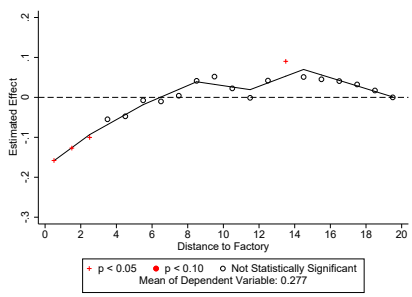
(d) Common Shifts: Plotted Coefficients



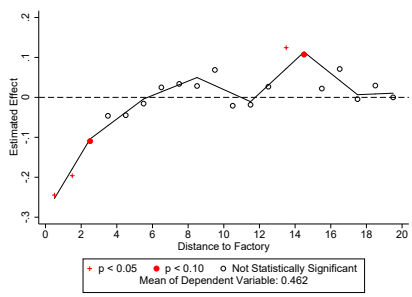
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure K-3: Industry and Agglomeration

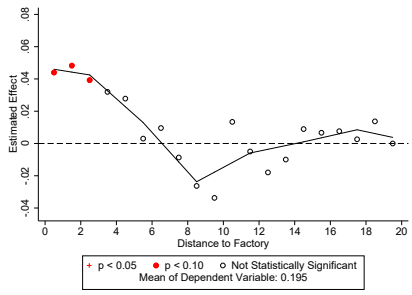
(a) Agriculture (Susenas 2001-11)



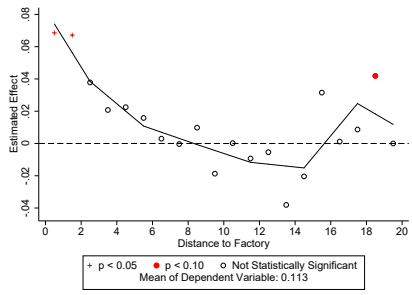
(b) Agriculture (Census 1980)



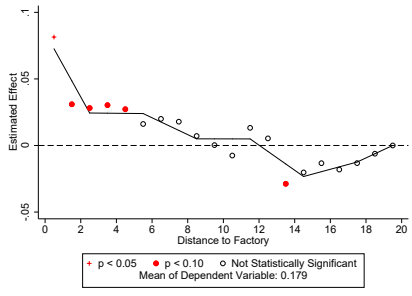
(c) Manufacturing (Susenas 2001-11)



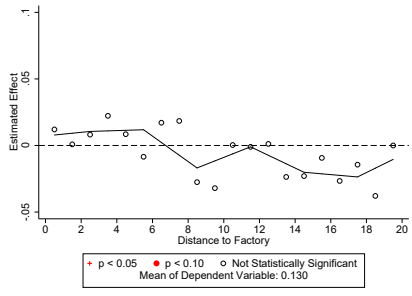
(d) Manufacturing (Census 1980)



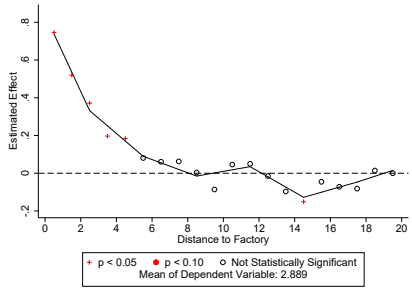
(e) Retail (Susenas 2001-11)



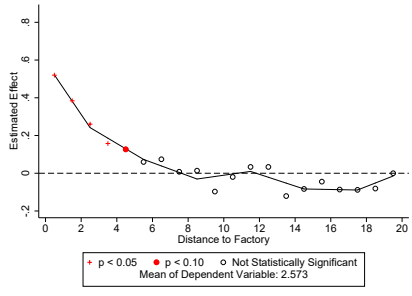
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



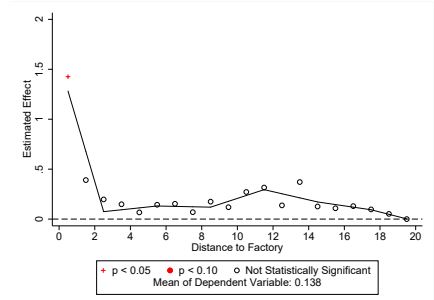
(h) Log Pop. Density (PODES 1980)



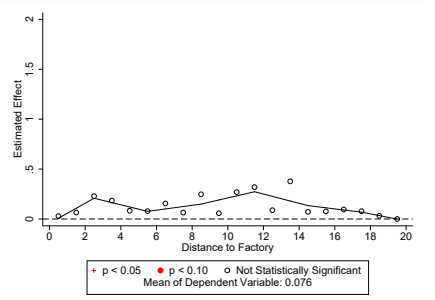
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-4: Sugar and Linked Industries

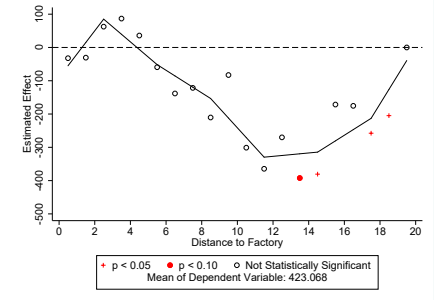
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



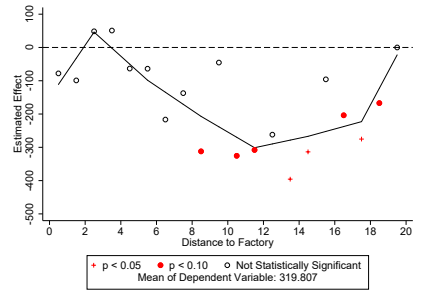
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



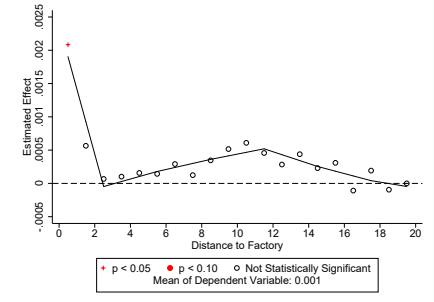
(c) Tons of Cane Grown (Full Sample, PODES 2003)



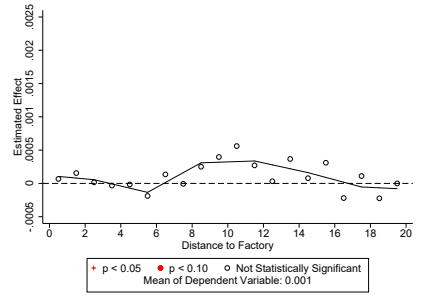
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



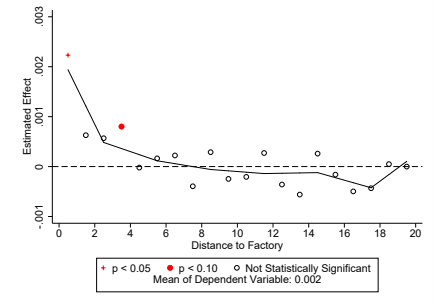
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



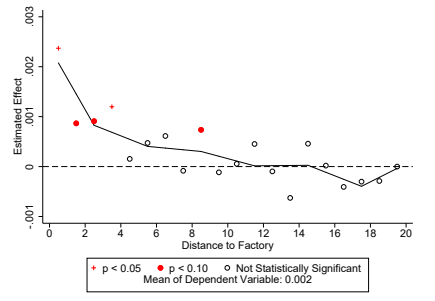
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



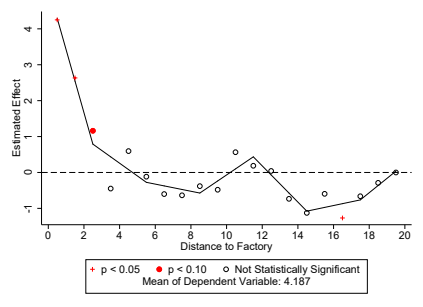
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



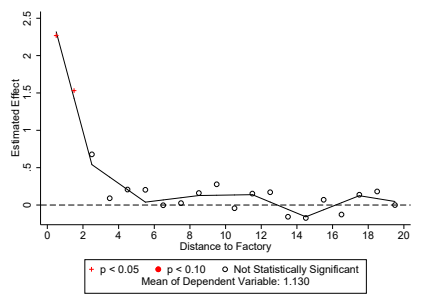
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-5: Infrastructure

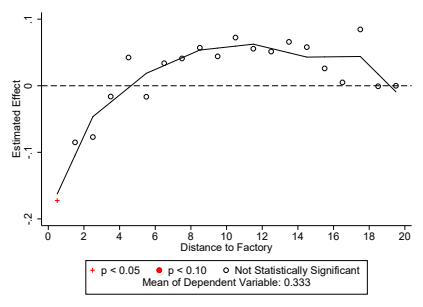
(a) Colonial Road Density (1900)



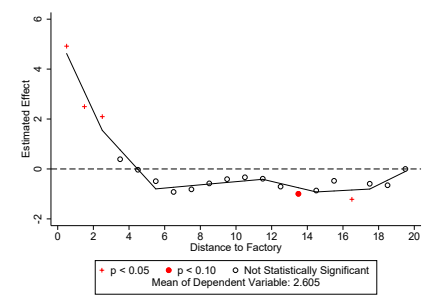
(b) Colonial Railroad Density (1900)



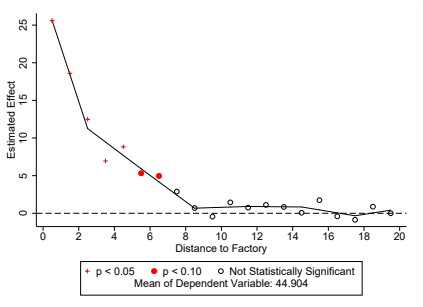
(c) Dirt Road (PODES 1980)



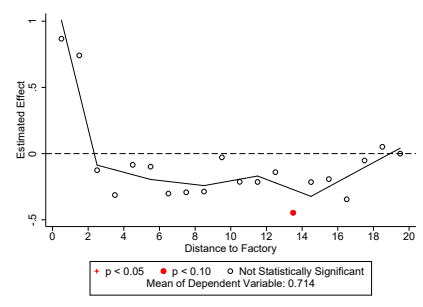
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



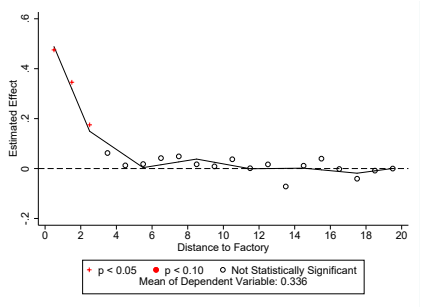
(f) Railroad Density (2017)



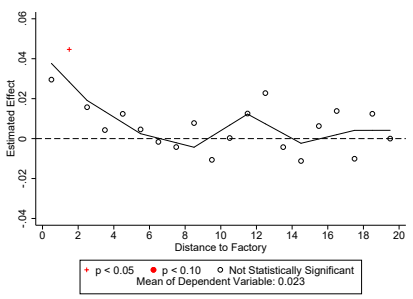
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-6: Other Public Goods

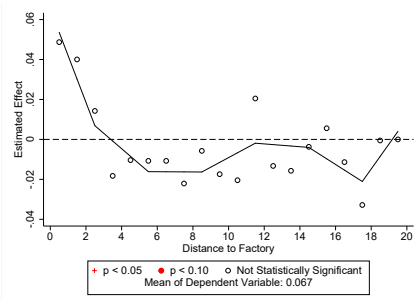
(a) Village Has Electricity (PODES 1980)



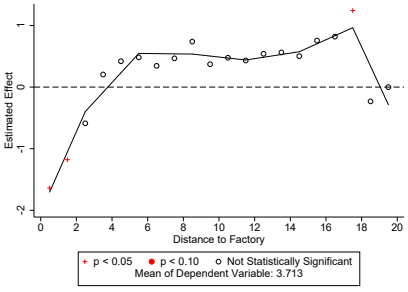
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



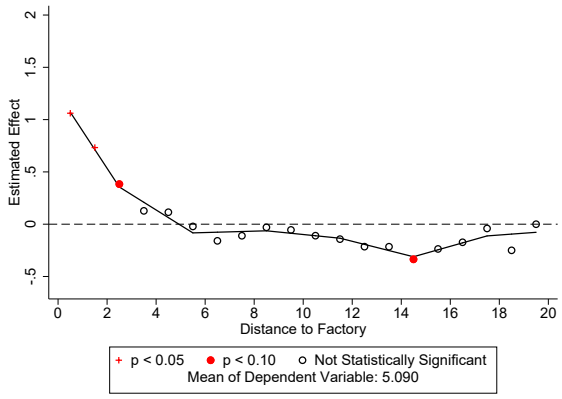
(d) Distance to Subdistrict Capital (2011 PODES)



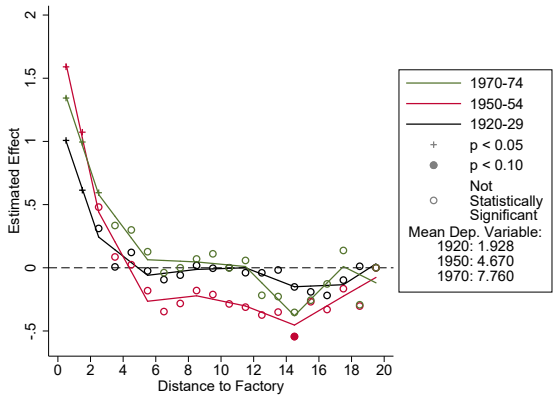
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-7: Education

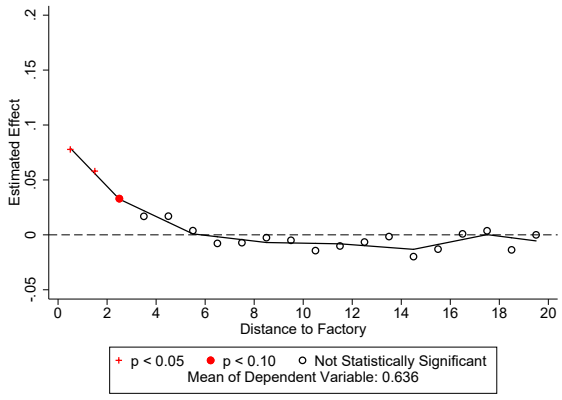
(a) Years Education (2000 Census)



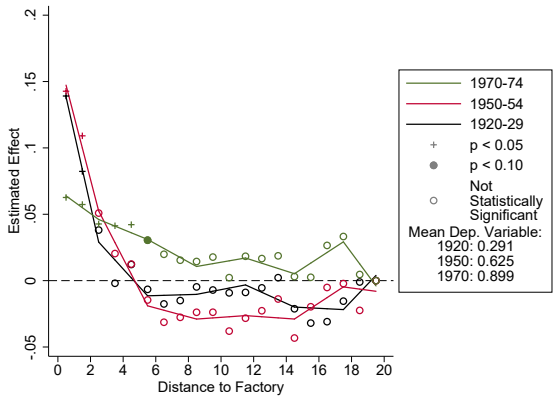
(b) Years Education by Cohort



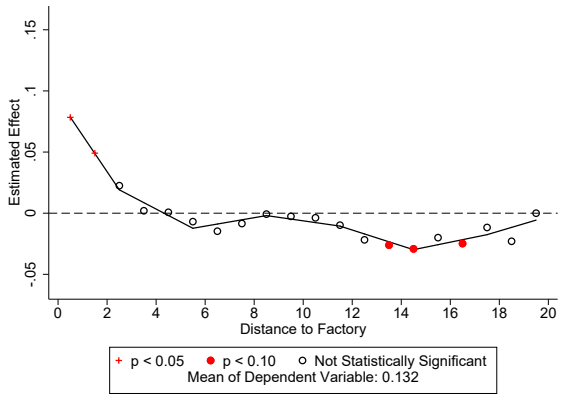
(c) Primary (2000 Census)



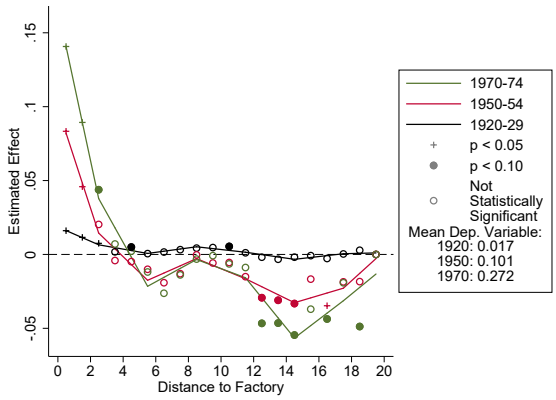
(d) Primary by Cohort



(e) High School (2000 Census)

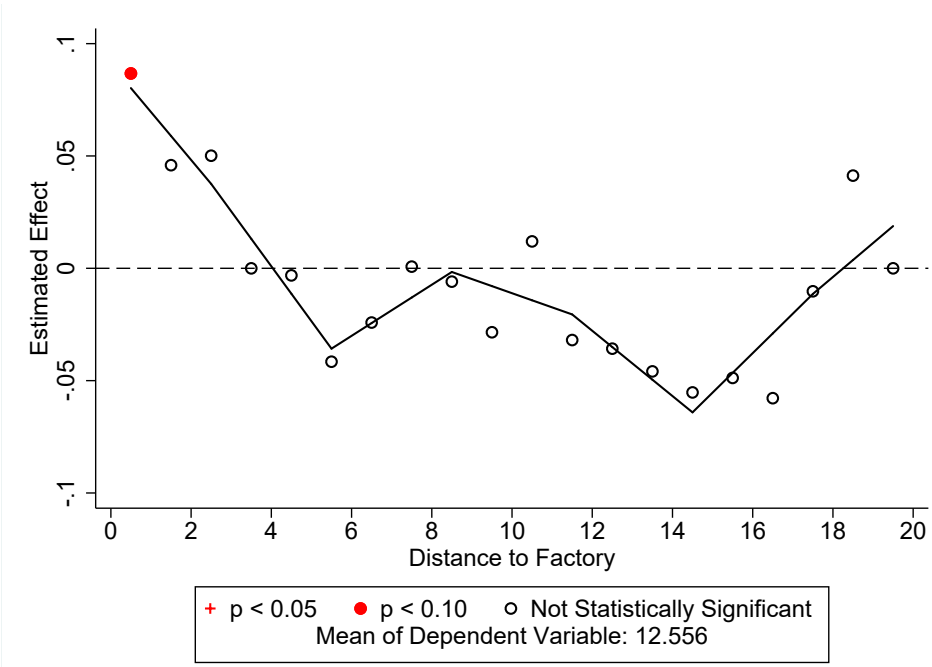


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure K-8: Expenditure (2001-11)

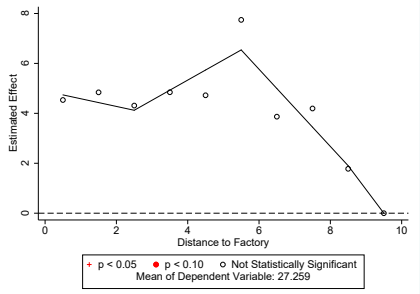


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

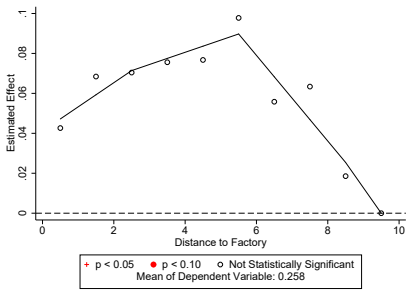
L Remove Factories Within 10 Km of Residency Capital

Figure L-1: Geography

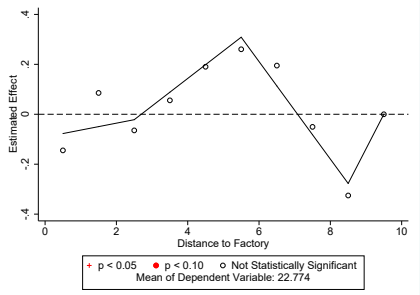
(a) Elevation



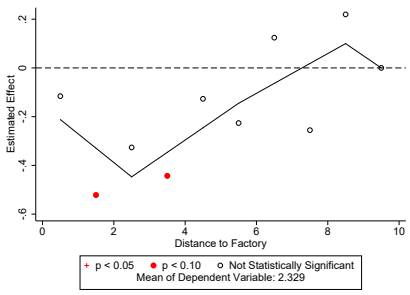
(b) Slope



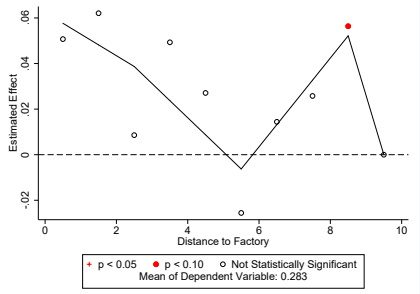
(c) Distance to Coast



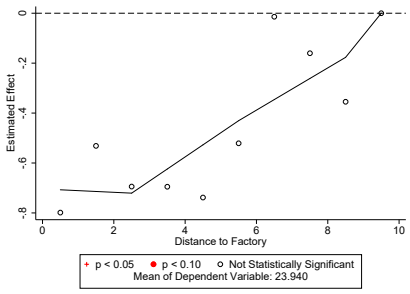
(d) Log Flow Accumulation



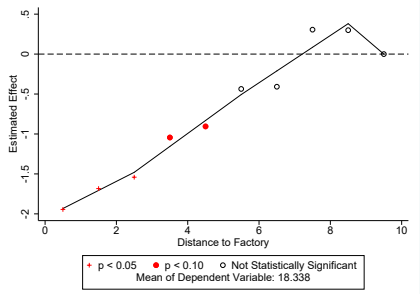
(e) Distance to River



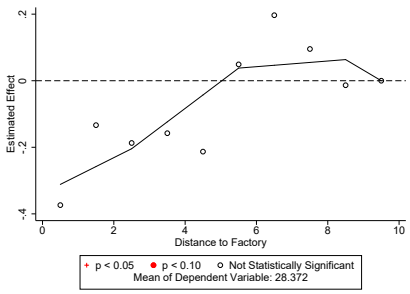
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



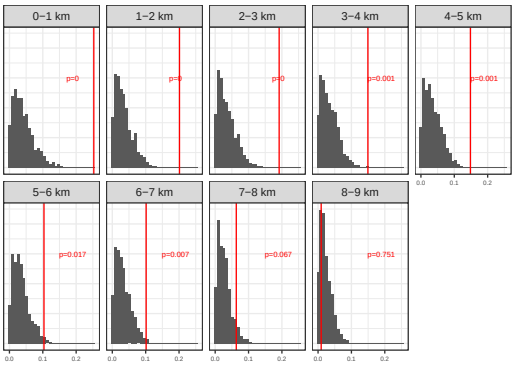
(h) Distance to VOC Port



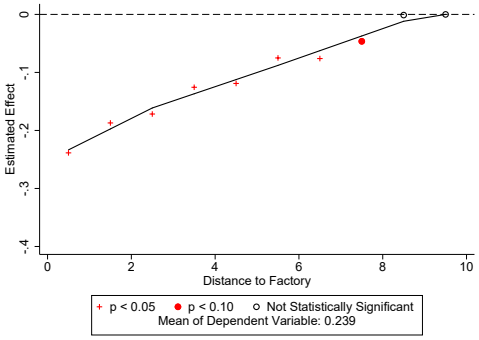
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure L-2: Share in Agriculture (2001-11): Illustration of Methodology

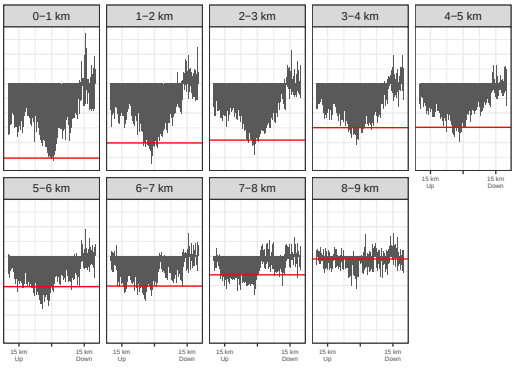
(a) Independent Shifts: Counterfactuals



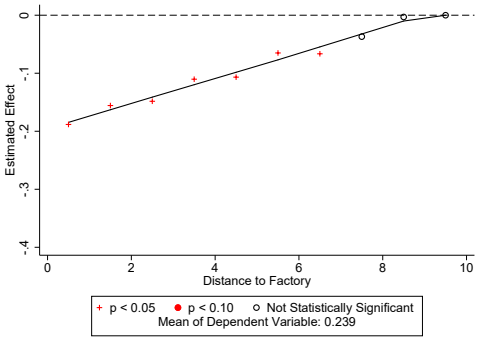
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



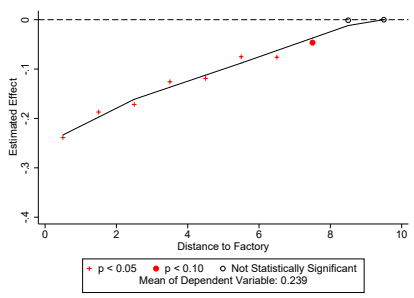
(d) Common Shifts: Plotted Coefficients



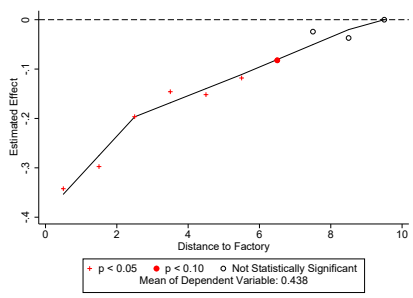
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure L-3: Industry and Agglomeration

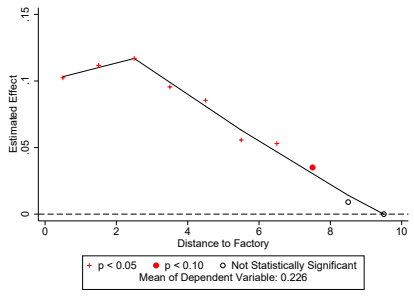
(a) Agriculture (Susenas 2001-11)



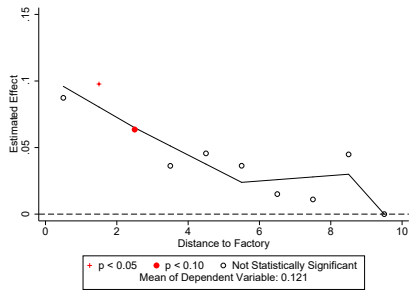
(b) Agriculture (Census 1980)



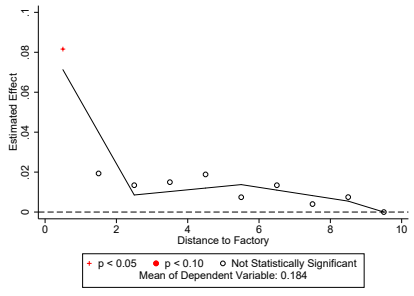
(c) Manufacturing (Susenas 2001-11)



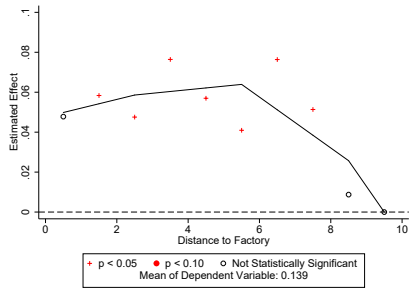
(d) Manufacturing (Census 1980)



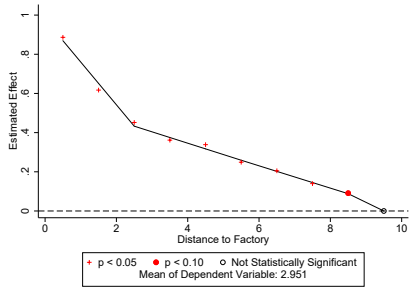
(e) Retail (Susenas 2001-11)



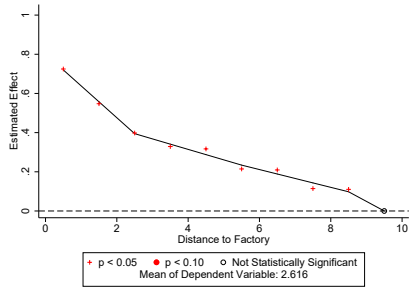
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



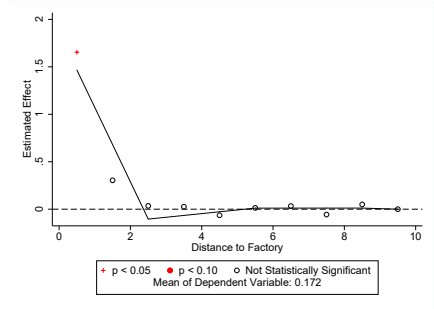
(h) Log Pop. Density (PODES 1980)



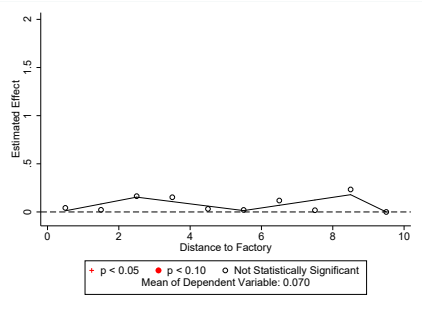
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure L-4: Sugar and Linked Industries

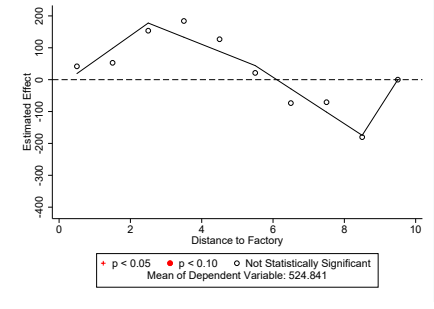
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



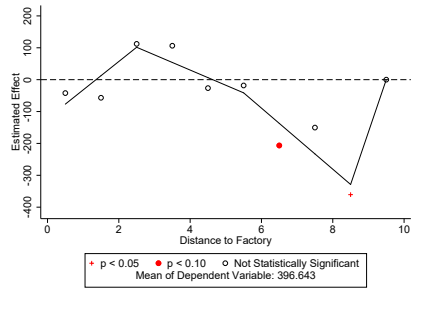
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



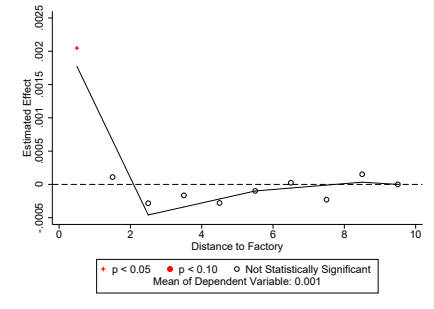
(c) Tons of Cane Grown (Full Sample, PODES 2003)



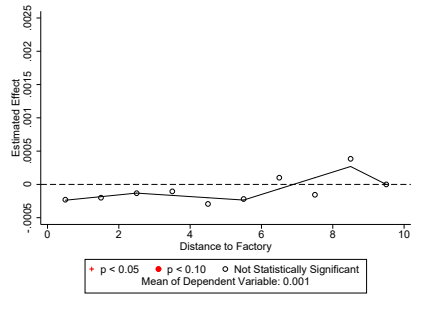
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



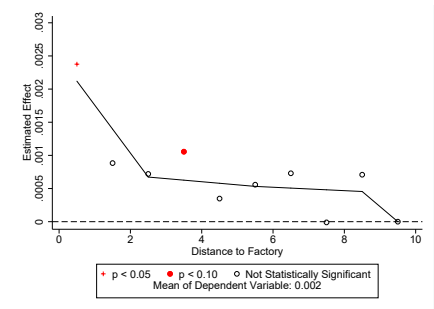
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



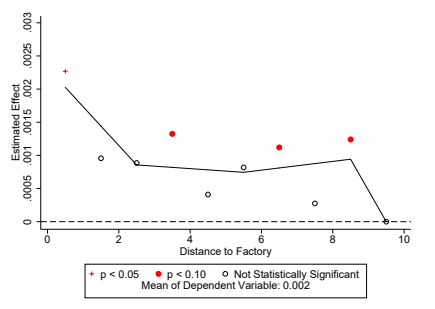
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



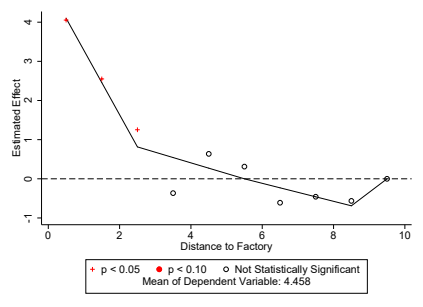
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



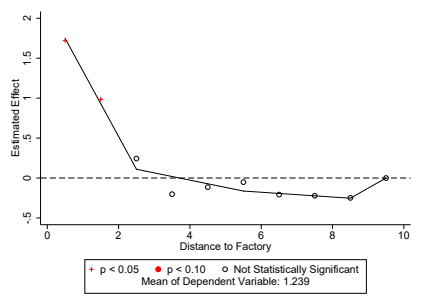
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure L-5: Infrastructure

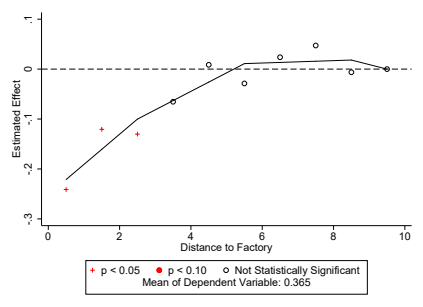
(a) Colonial Road Density (1900)



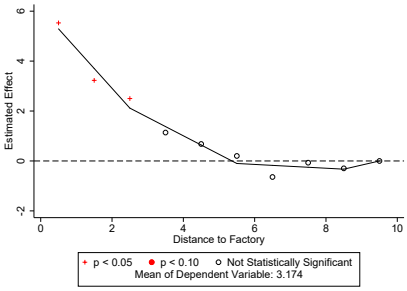
(b) Colonial Railroad Density (1900)



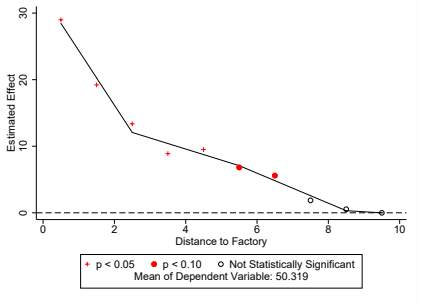
(c) Dirt Road (PODES 1980)



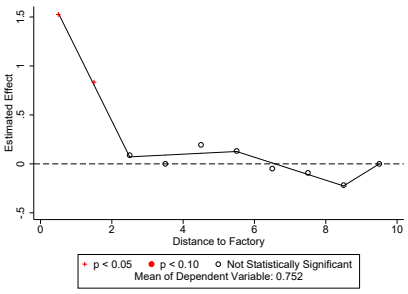
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



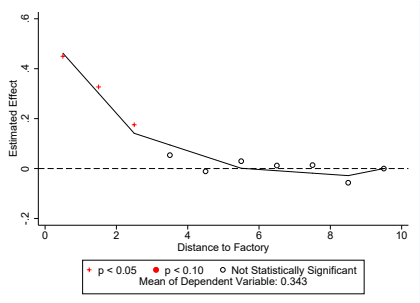
(f) Railroad Density (2017)



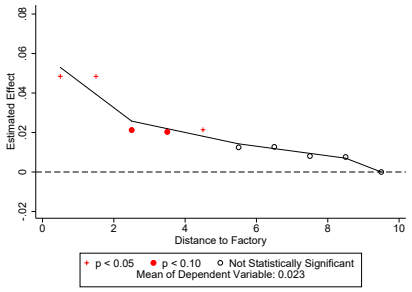
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations. z

Figure L-6: Other Public Goods

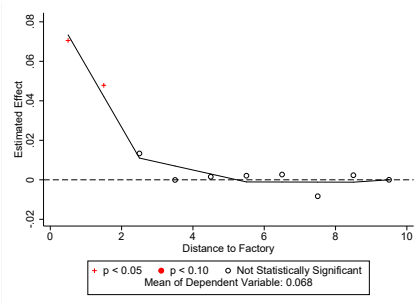
(a) Village Has Electricity (PODES 1980)



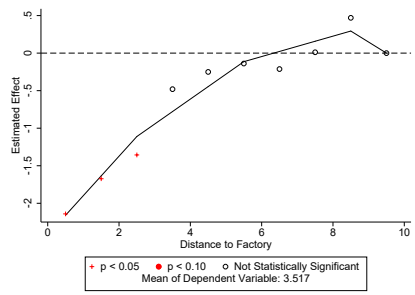
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



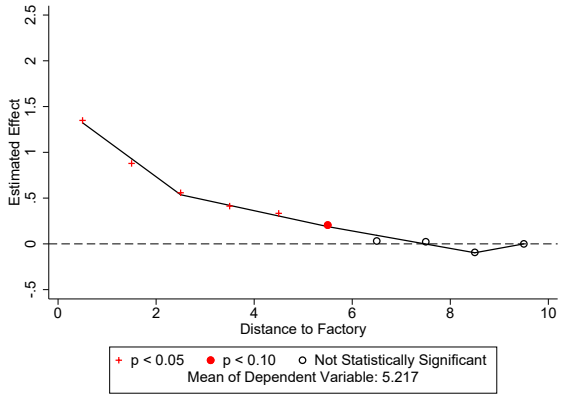
(d) Distance to Subdistrict Capital (2011 PODES)



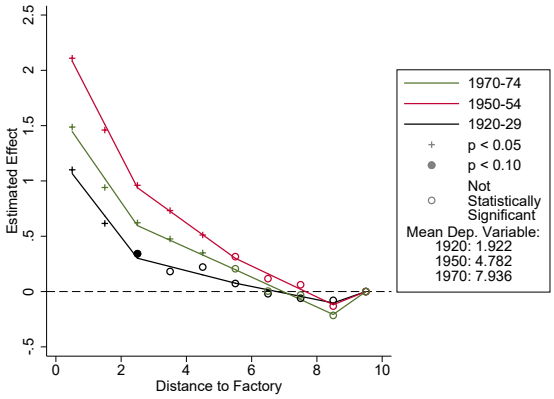
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure L-7: Education

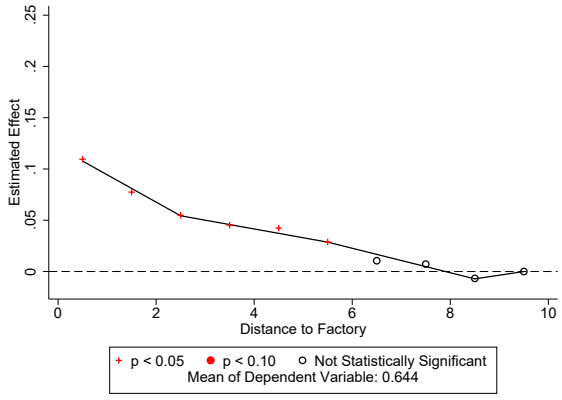
(a) Years Education (2000 Census)



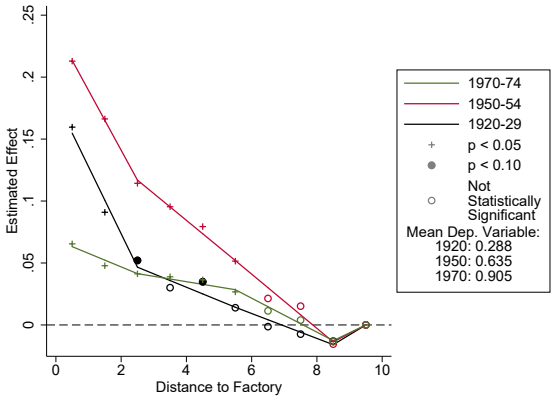
(b) Years Education by Cohort



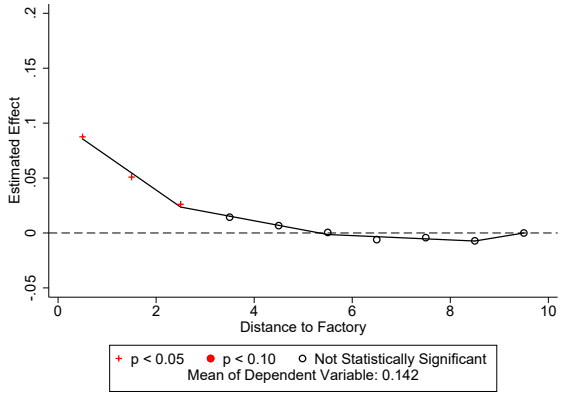
(c) Primary (2000 Census)



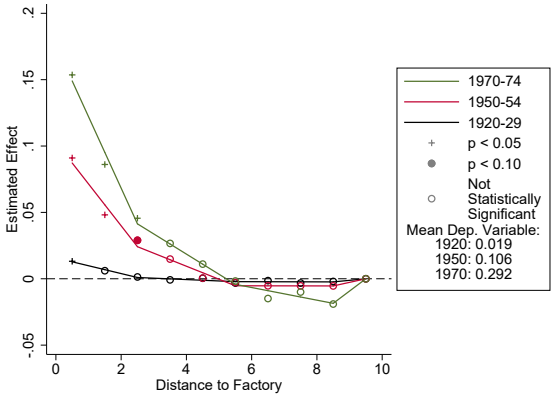
(d) Primary by Cohort



(e) High School (2000 Census)

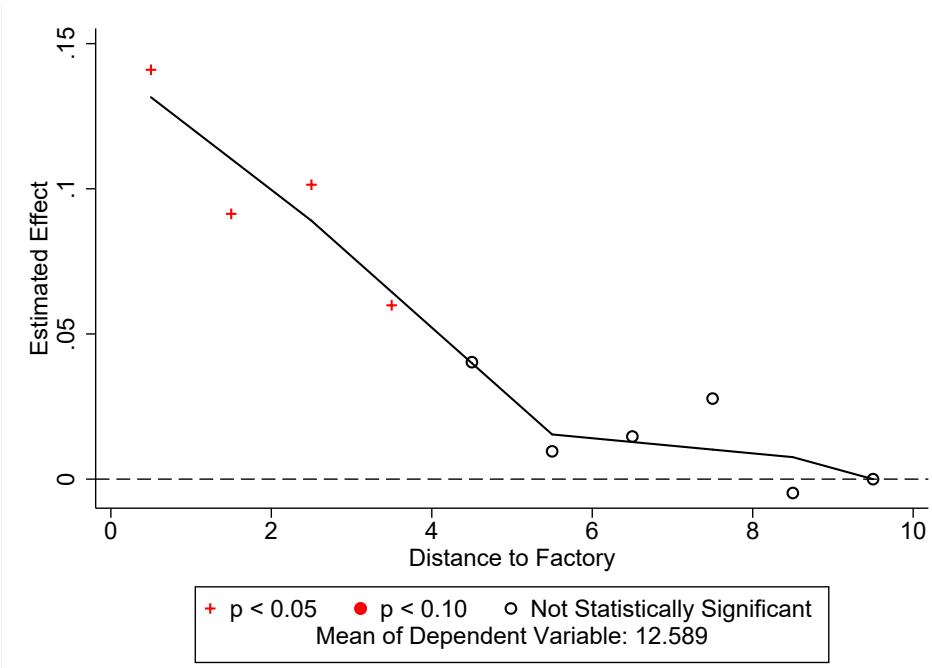


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure L-8: Expenditure (2001-11)

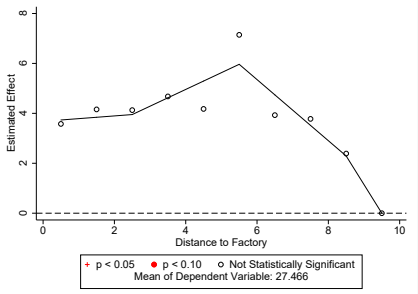


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

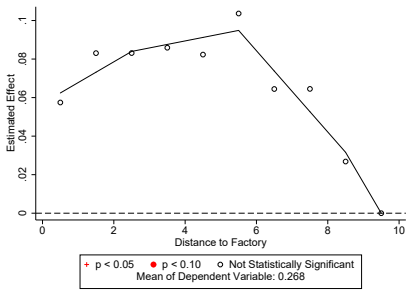
M 10 Kilometer Sample

Figure M-1: Geography

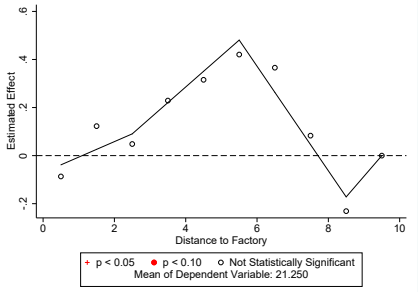
(a) Elevation



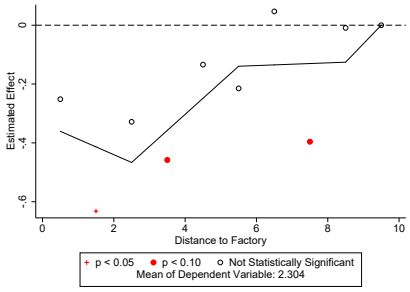
(b) Slope



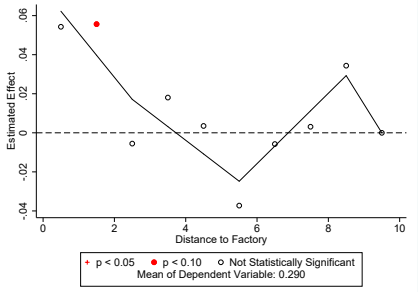
(c) Distance to Coast



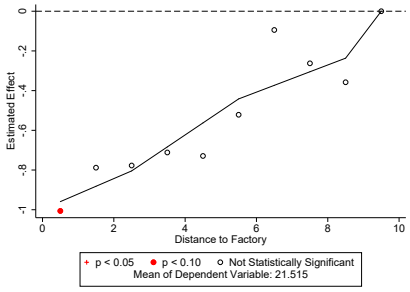
(d) Log Flow Accumulation



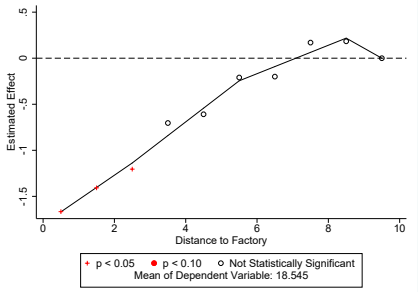
(e) Distance to River



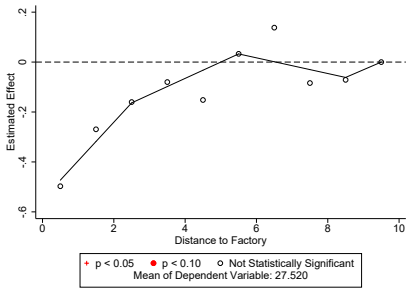
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



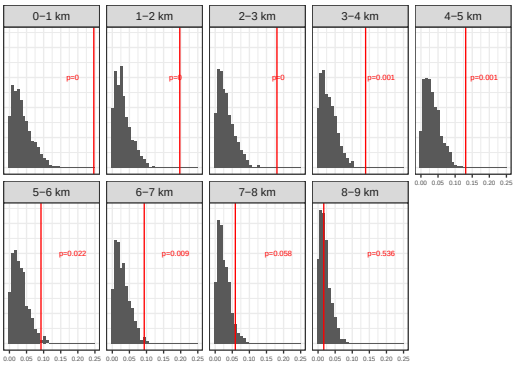
(h) Distance to VOC Port



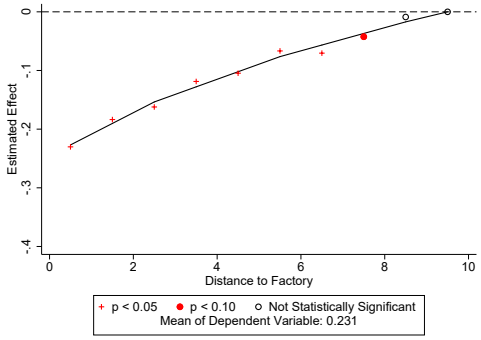
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-2: Share in Agriculture (2001-11): Illustration of Methodology

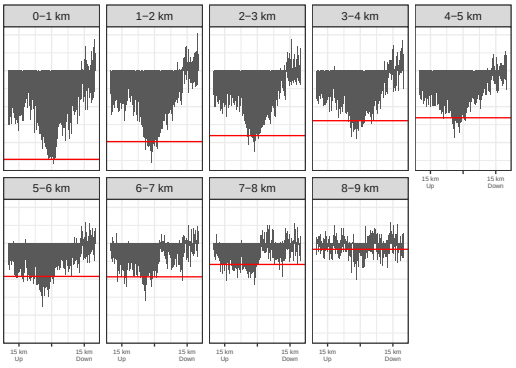
(a) Independent Shifts: Counterfactuals



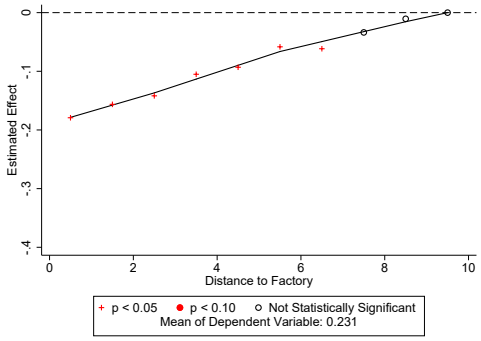
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



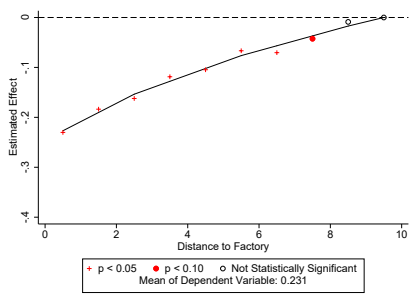
(d) Common Shifts: Plotted Coefficients



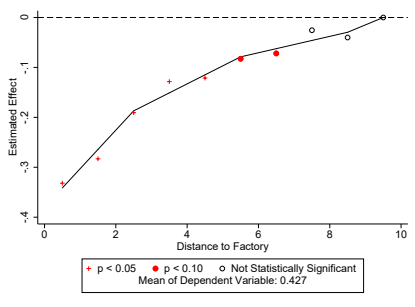
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure M-3: Industry and Agglomeration

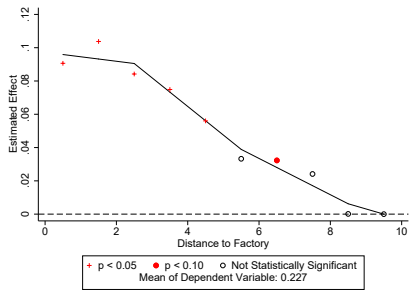
(a) Agriculture (Susenas 2001-11)



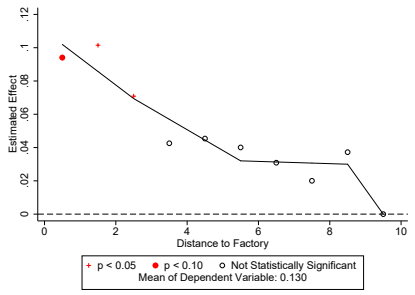
(b) Agriculture (Census 1980)



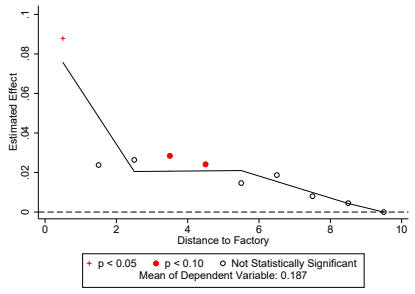
(c) Manufacturing (Susenas 2001-11)



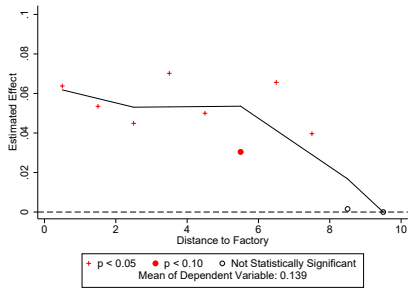
(d) Manufacturing (Census 1980)



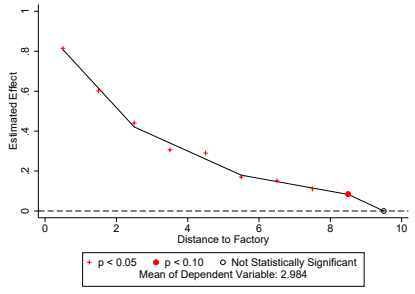
(e) Retail (Susenas 2001-11)



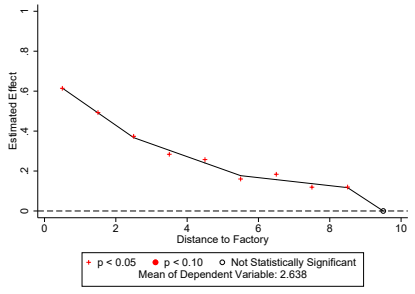
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



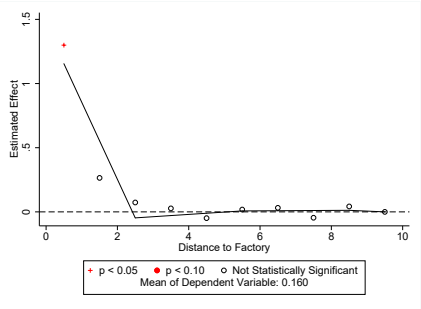
(h) Log Pop. Density (PODES 1980)



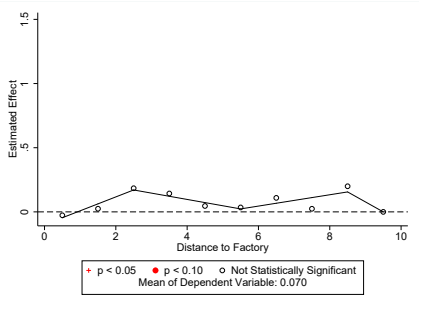
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-4: Sugar and Linked Industries

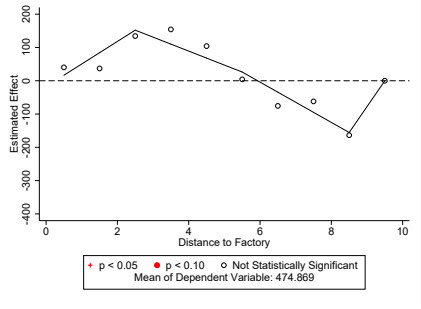
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



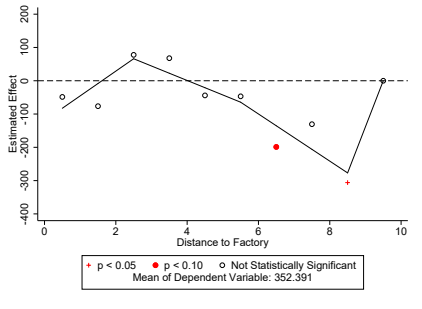
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



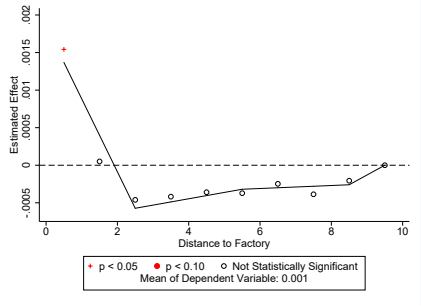
(c) Tons of Cane Grown (Full Sample, PODES 2003)



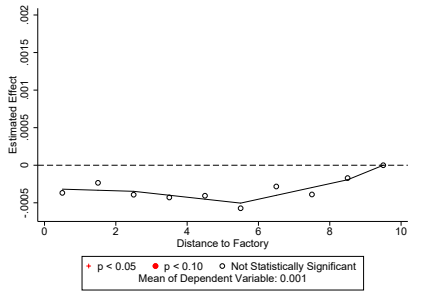
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



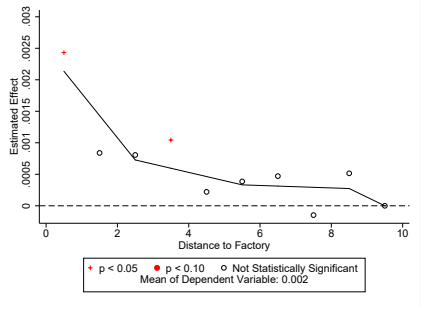
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



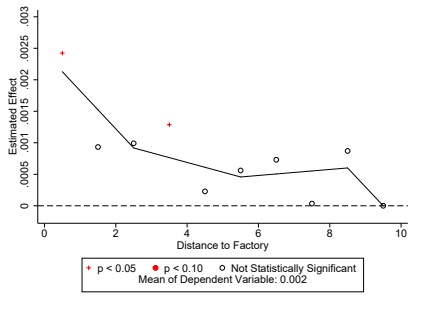
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



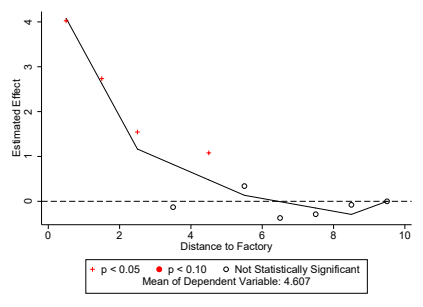
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



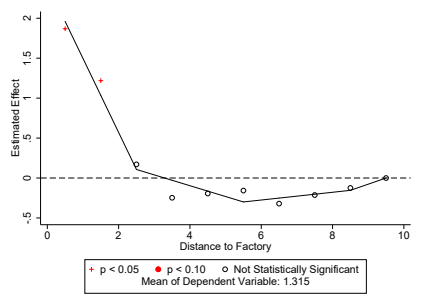
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-5: Infrastructure

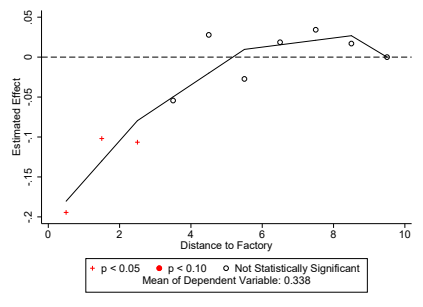
(a) Colonial Road Density (1900)



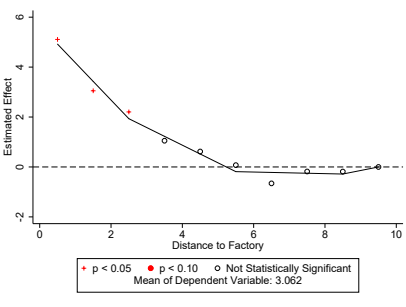
(b) Colonial Railroad Density (1900)



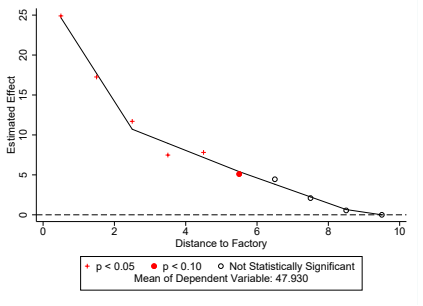
(c) Dirt Road (PODES 1980)



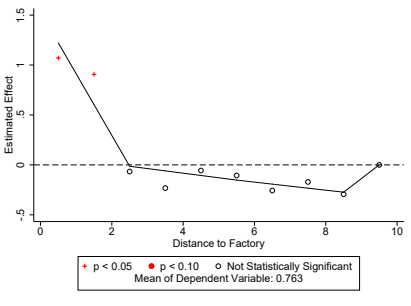
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



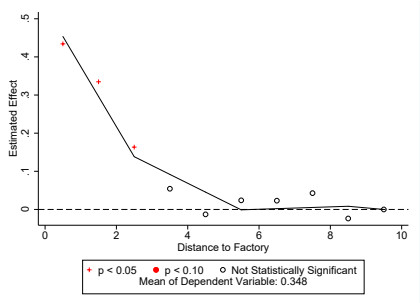
(f) Railroad Density (2017)



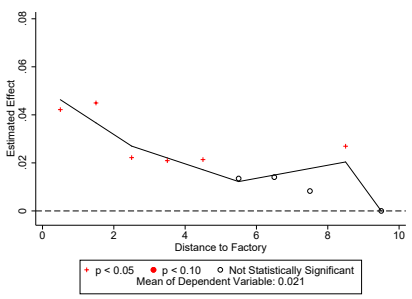
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-6: Other Public Goods

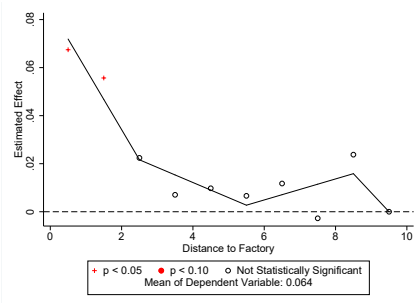
(a) Village Has Electricity (PODES 1980)



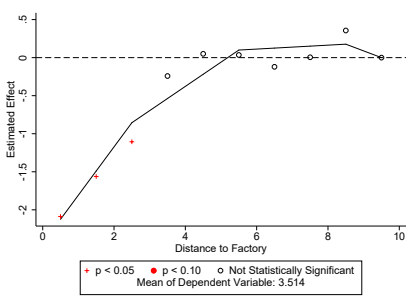
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



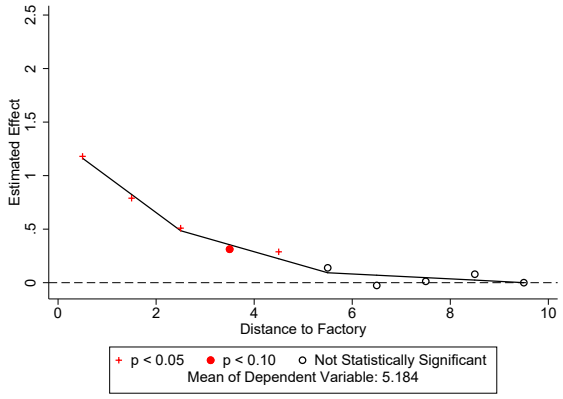
(d) Distance to Subdistrict Capital (2011 PODES)



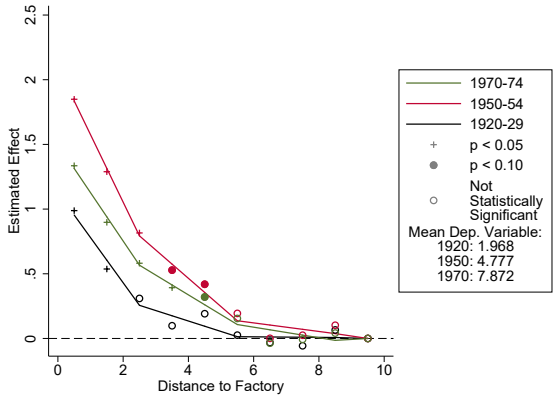
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-7: Education

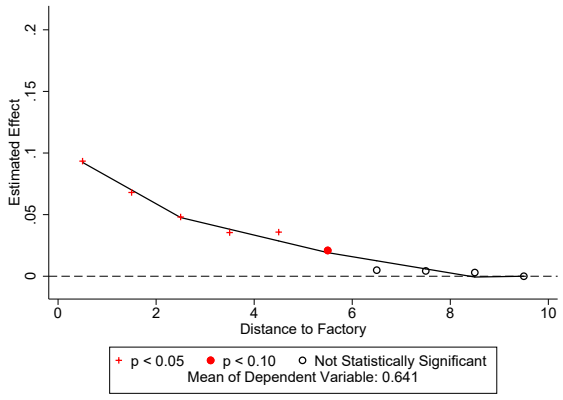
(a) Years Education (2000 Census)



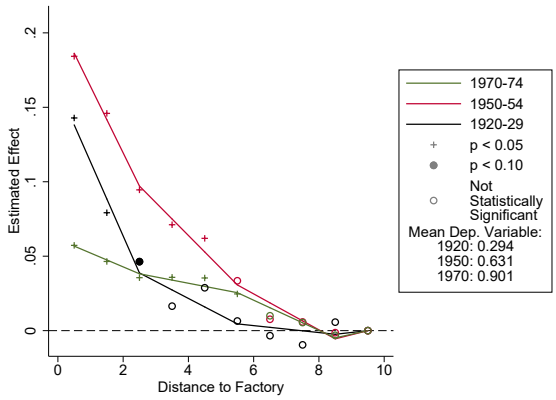
(b) Years Education by Cohort



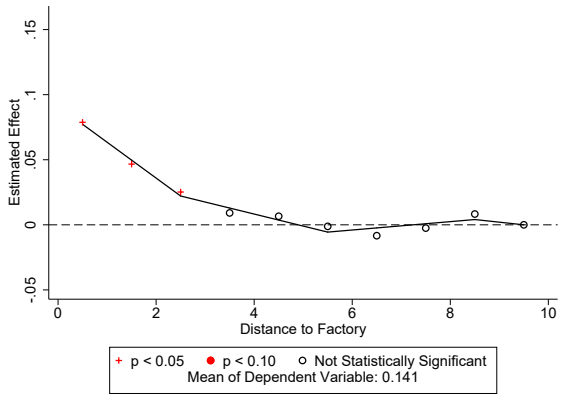
(c) Primary (2000 Census)



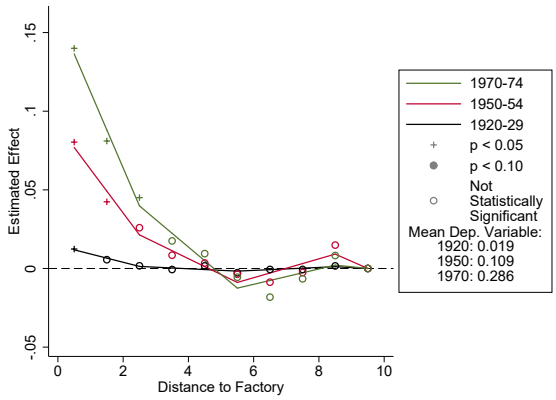
(d) Primary by Cohort



(e) High School (2000 Census)

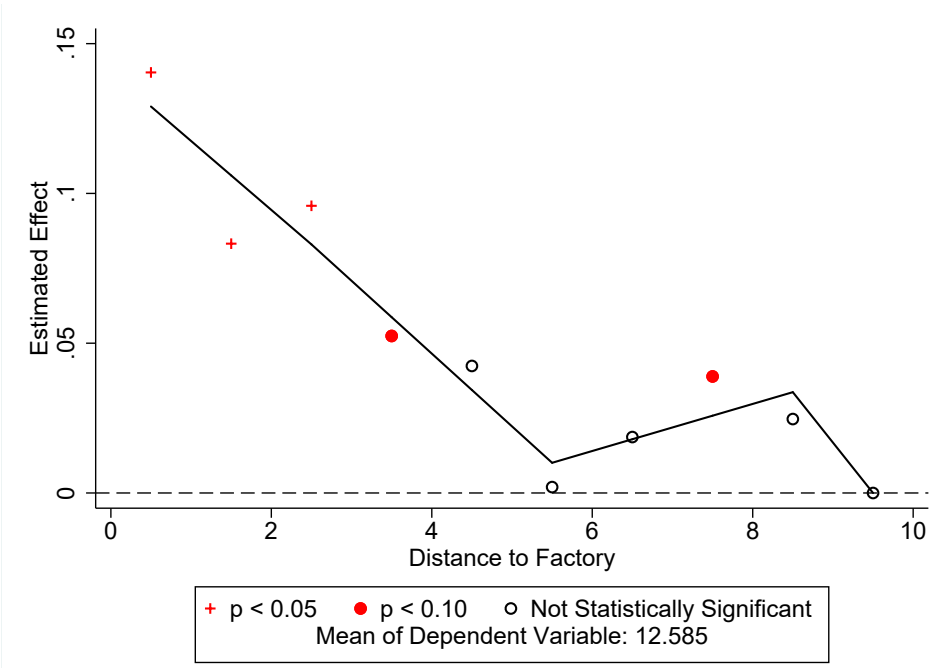


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure M-8: Expenditure (2001-11)

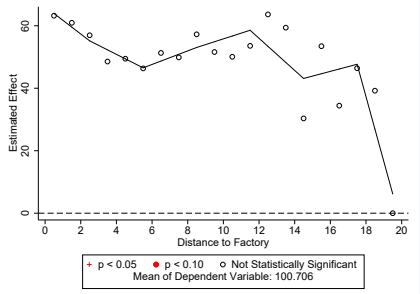


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

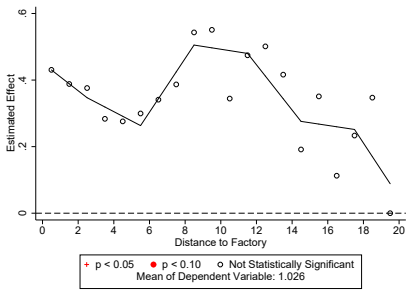
N No Village Restrictions

Figure N-1: Geography

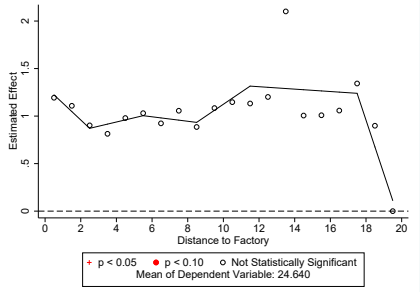
(a) Elevation



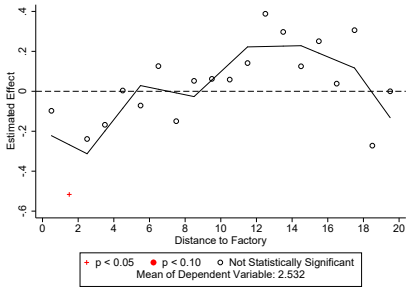
(b) Slope



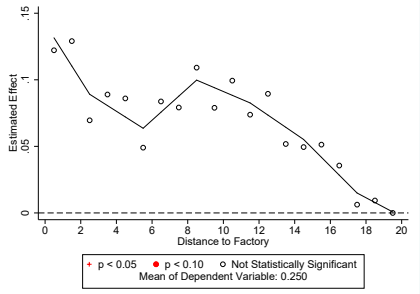
(c) Distance to Coast



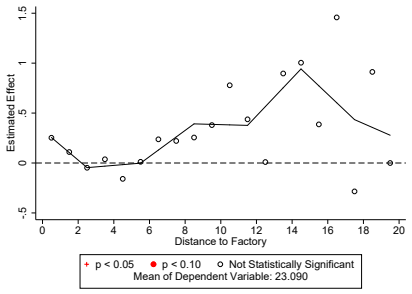
(d) Log Flow Accumulation



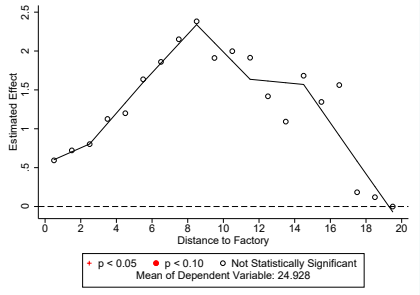
(e) Distance to River



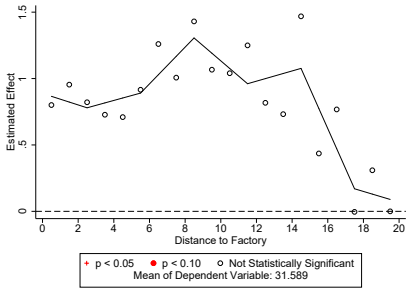
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



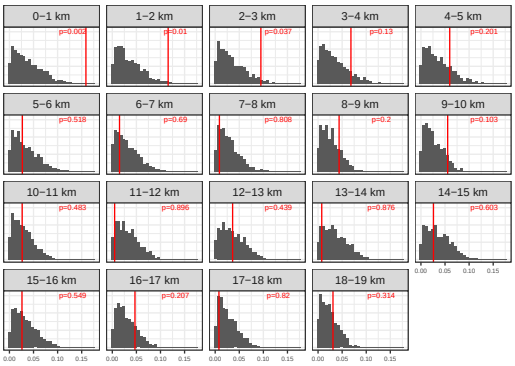
(h) Distance to VOC Port



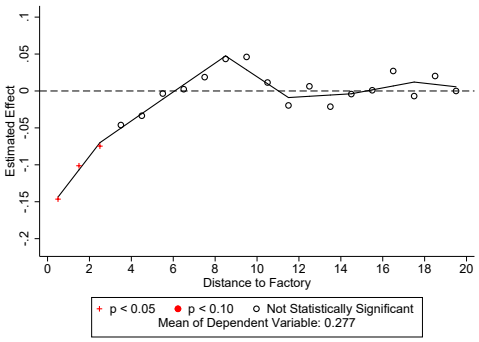
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-2: Share in Agriculture (2001-11): Illustration of Methodology

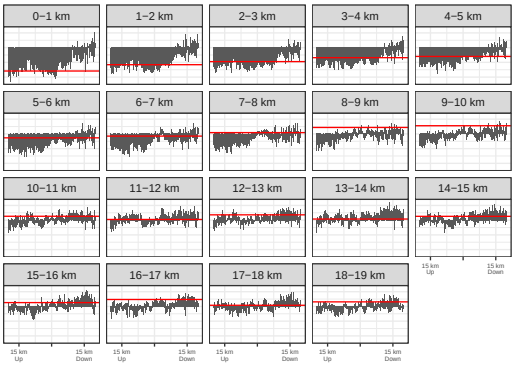
(a) Independent Shifts: Counterfactuals



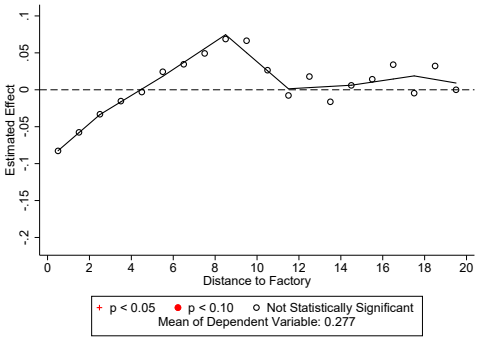
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



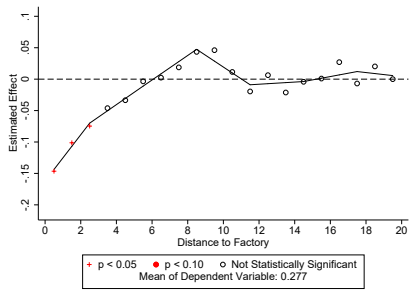
(d) Common Shifts: Plotted Coefficients



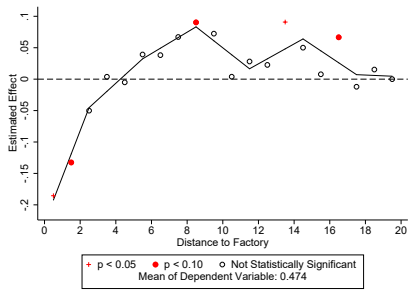
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure N-3: Industry and Agglomeration

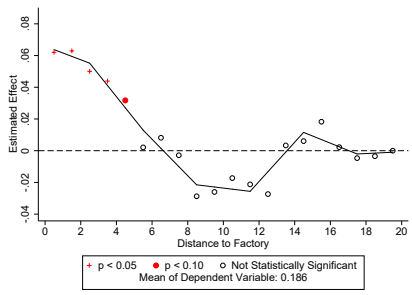
(a) Agriculture (Susenas 2001-11)



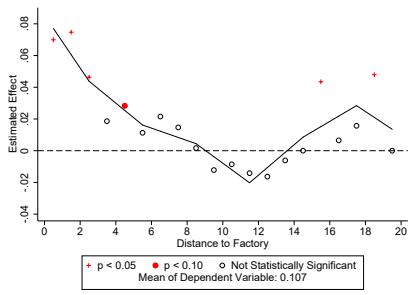
(b) Agriculture (Census 1980)



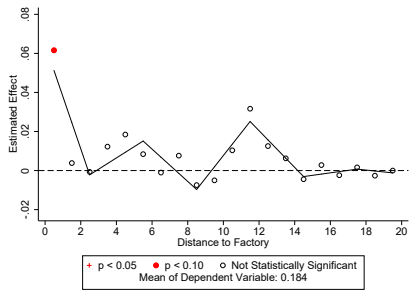
(c) Manufacturing (Susenas 2001-11)



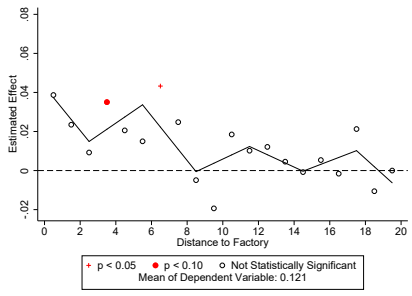
(d) Manufacturing (Census 1980)



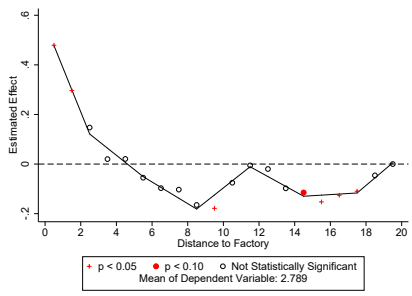
(e) Retail (Susenas 2001-11)



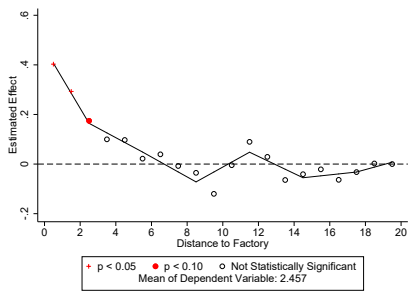
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



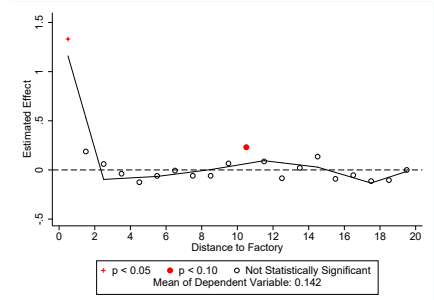
(h) Log Pop. Density (PODES 1980)



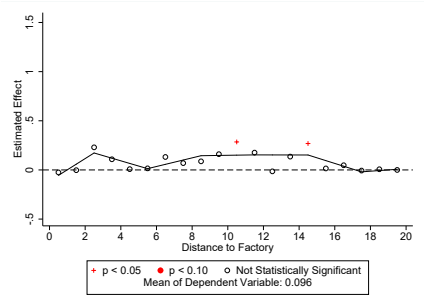
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-4: Sugar and Linked Industries

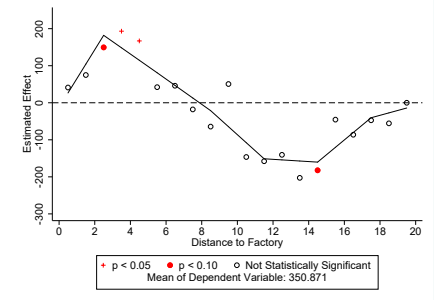
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



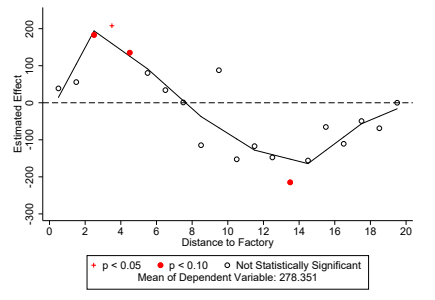
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



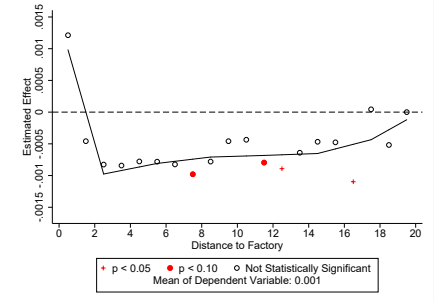
(c) Tons of Cane Grown (Full Sample, PODES 2003)



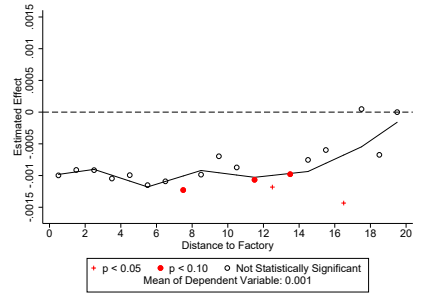
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



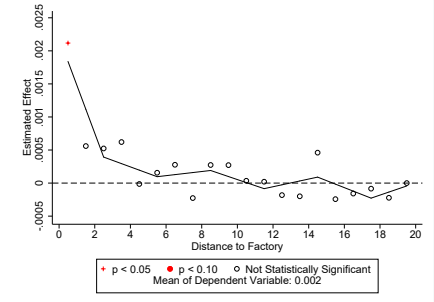
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



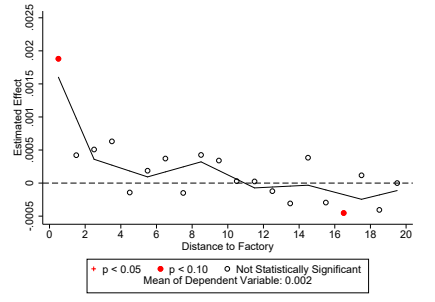
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



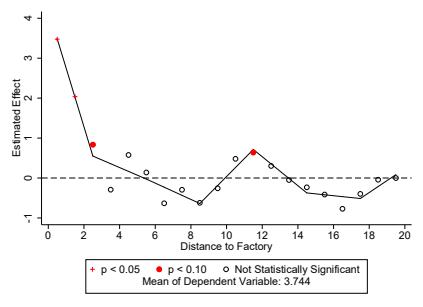
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



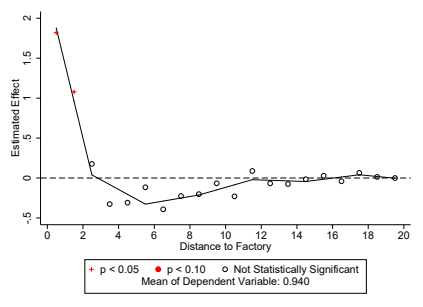
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-5: Infrastructure

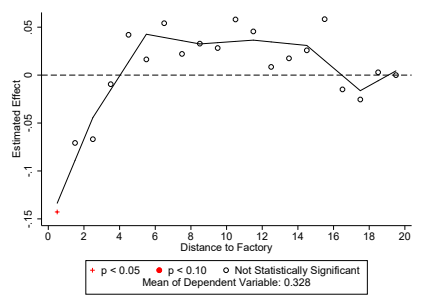
(a) Colonial Road Density (1900)



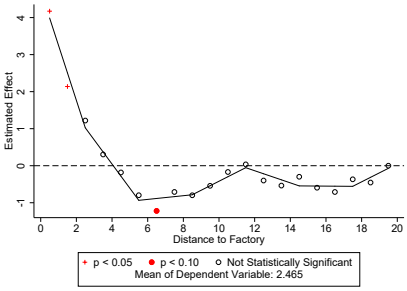
(b) Colonial Railroad Density (1900)



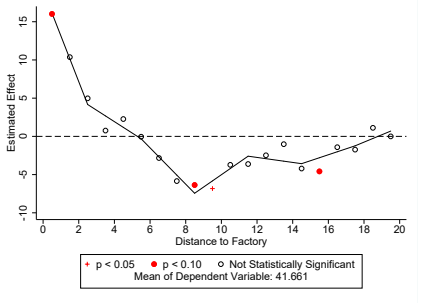
(c) Dirt Road (PODES 1980)



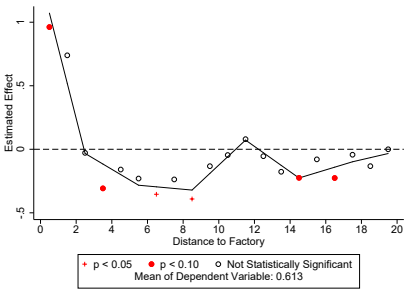
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



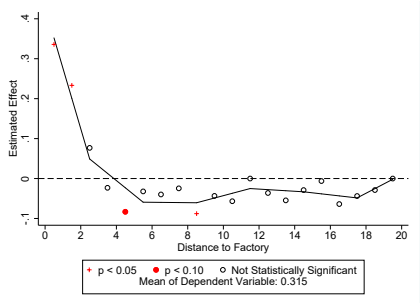
(f) Railroad Density (2017)



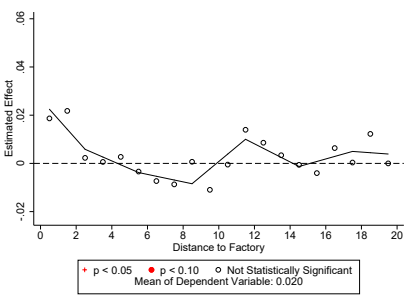
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-6: Other Public Goods

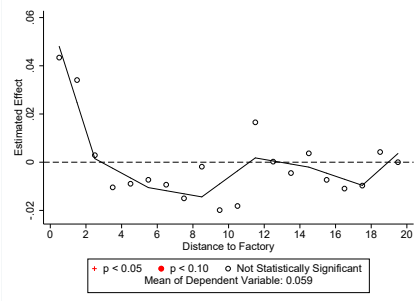
(a) Village Has Electricity (PODES 1980)



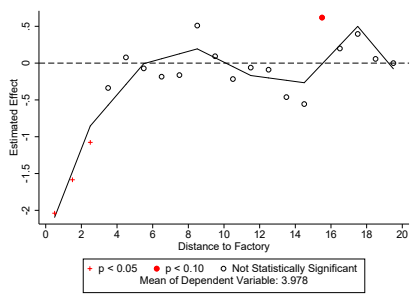
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



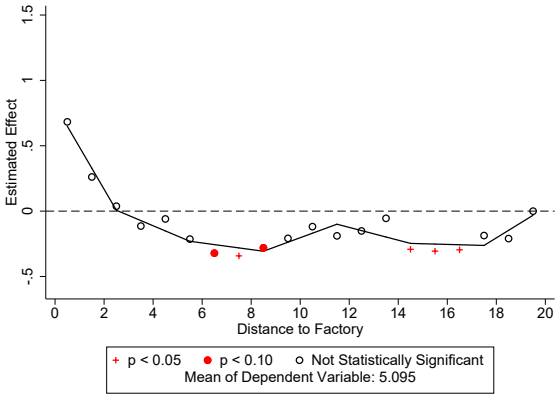
(d) Distance to Subdistrict Capital (2011 PODES)



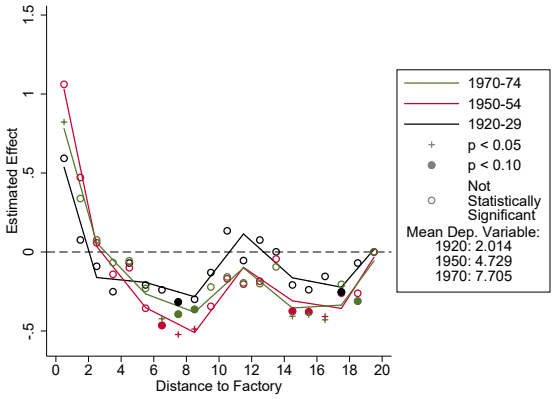
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-7: Education

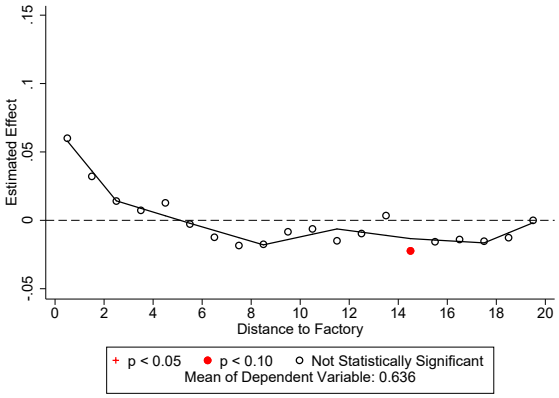
(a) Years Education (2000 Census)



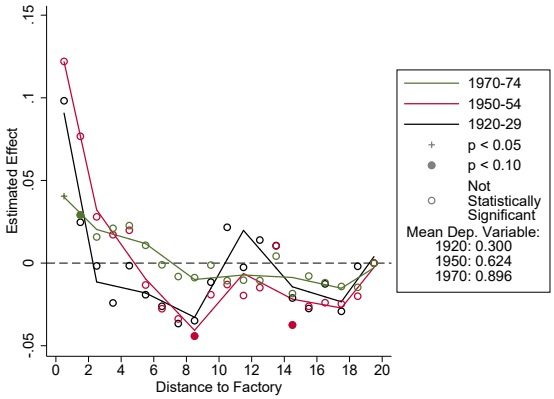
(b) Years Education by Cohort



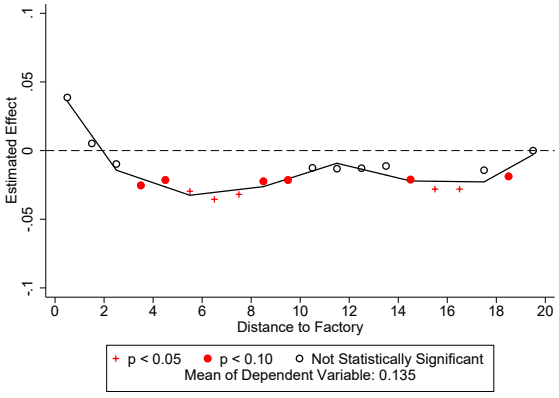
(c) Primary (2000 Census)



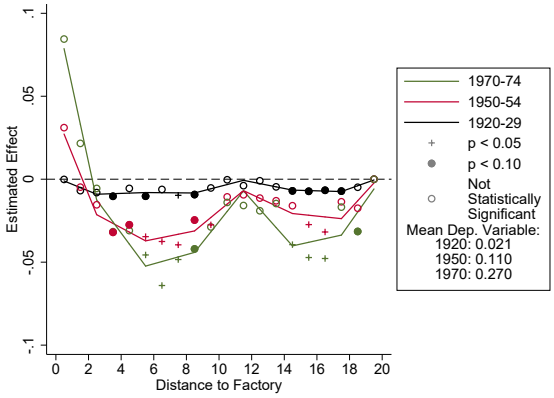
(d) Primary by Cohort



(e) High School (2000 Census)

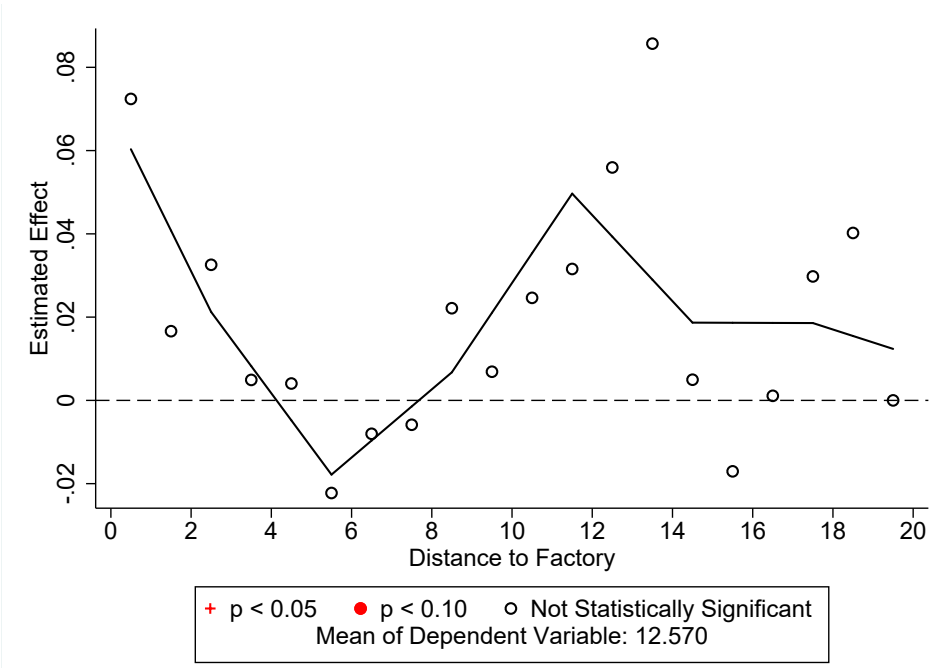


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure N-8: Expenditure (2001-11)

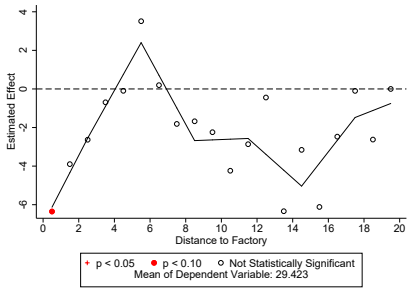


Notes: This figure plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

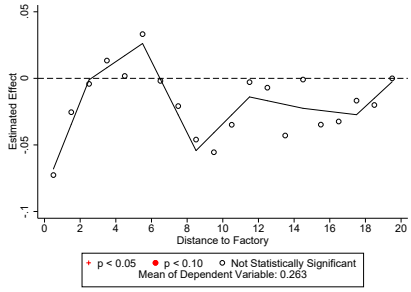
O No Factory Fixed Effects

Figure O-1: Geography

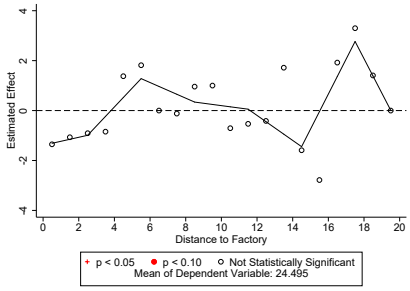
(a) Elevation



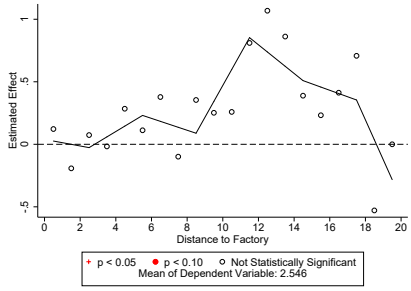
(b) Slope



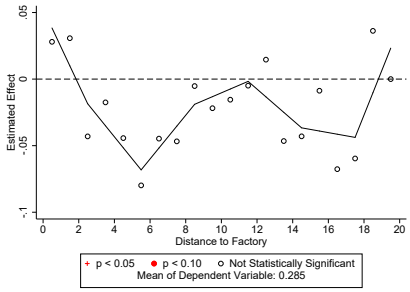
(c) Distance to Coast



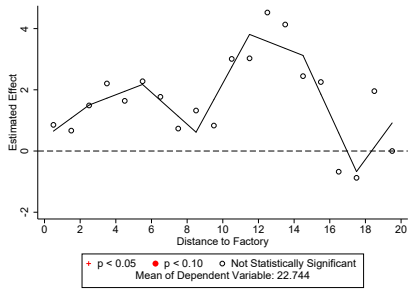
(d) Log Flow Accumulation



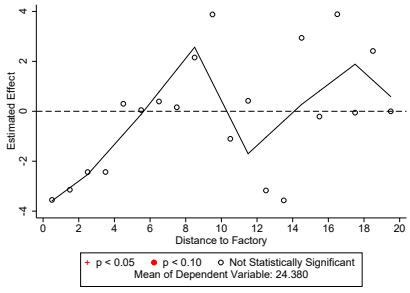
(e) Distance to River



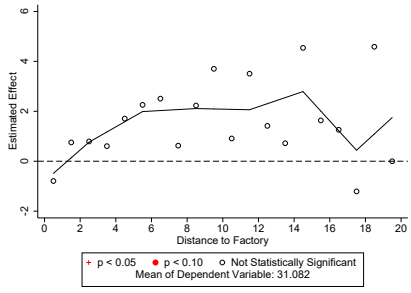
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



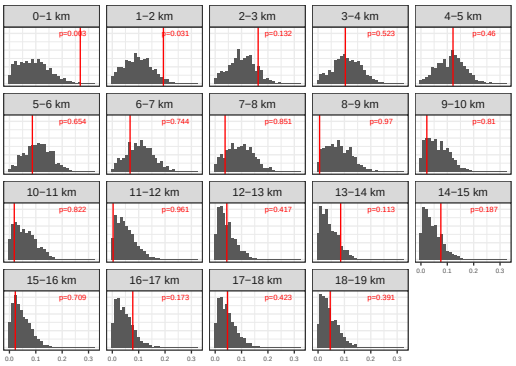
(h) Distance to VOC Port



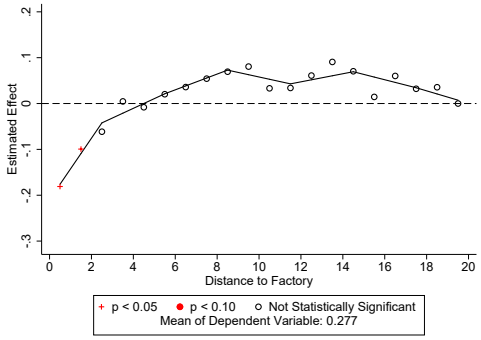
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-2: Share in Agriculture (2001-11): Illustration of Methodology

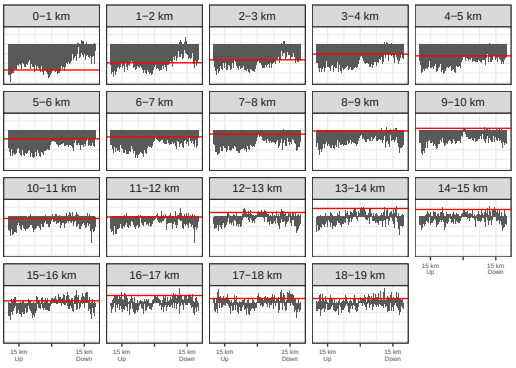
(a) Independent Shifts: Counterfactuals



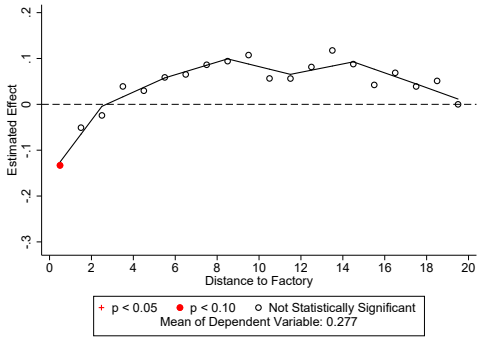
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



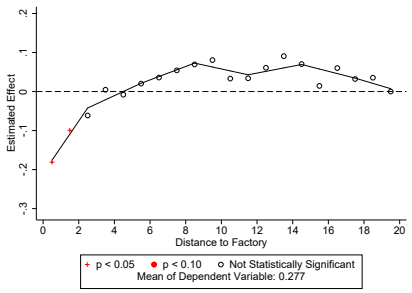
(d) Common Shifts: Plotted Coefficients



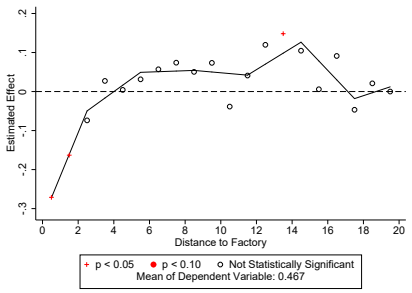
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for geographic and pre-period characteristics and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure O-3: Industry and Agglomeration

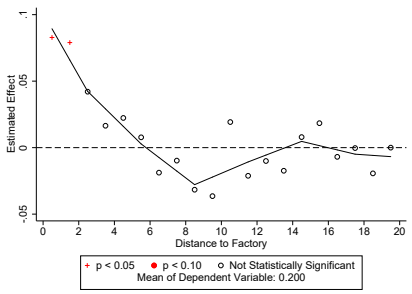
(a) Agriculture (Susenas 2001-11)



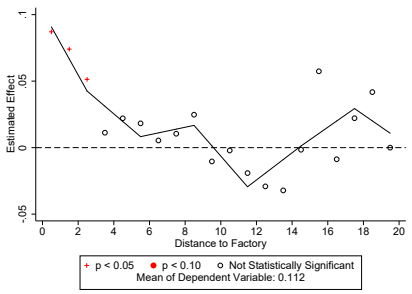
(b) Agriculture (Census 1980)



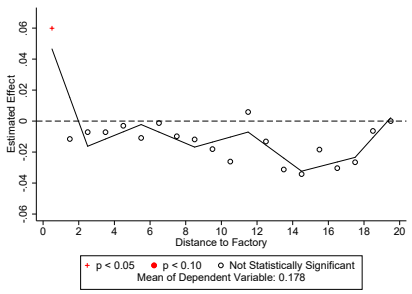
(c) Manufacturing (Susenas 2001-11)



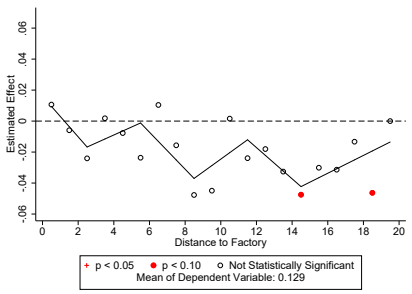
(d) Manufacturing (Census 1980)



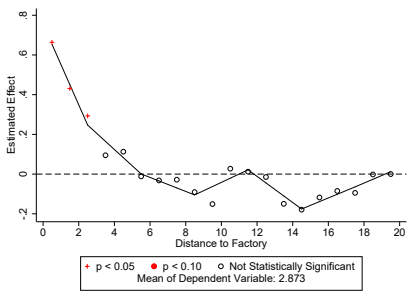
(e) Retail (Susenas 2001-11)



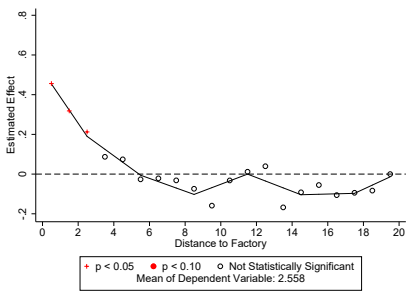
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



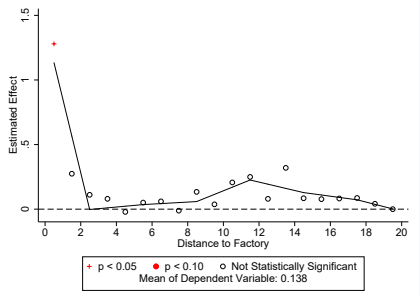
(h) Log Pop. Density (PODES 1980)



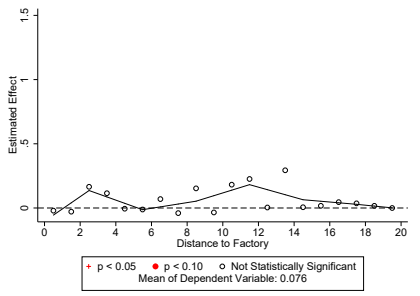
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearestcounterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-4: Sugar and Linked Industries

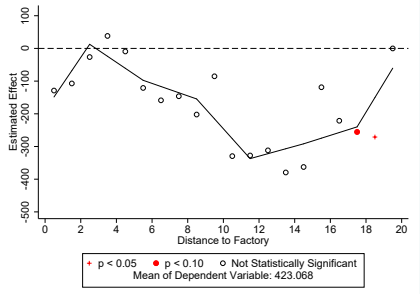
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



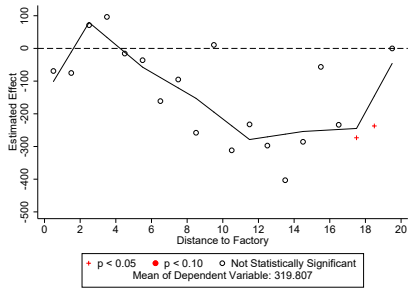
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



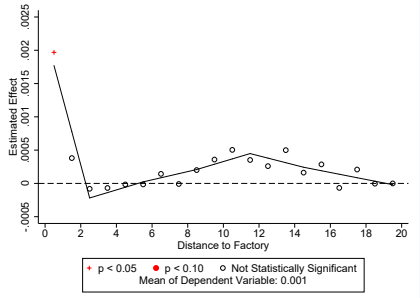
(c) Tons of Cane Grown (Full Sample, PODES 2003)



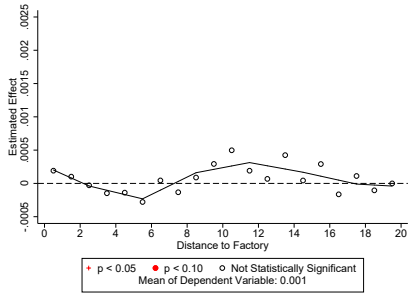
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



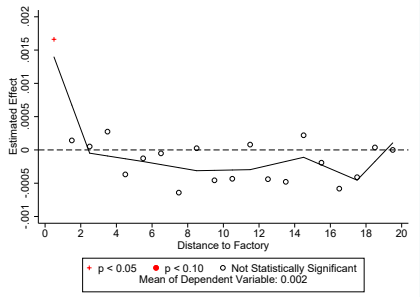
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



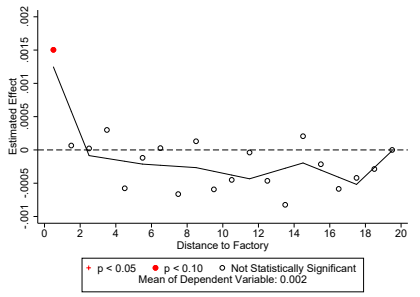
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



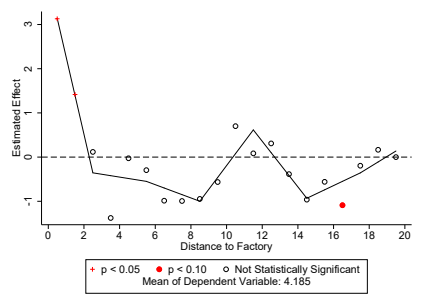
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



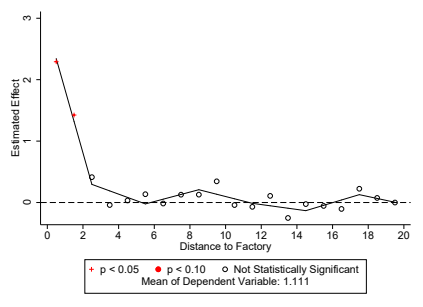
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-5: Infrastructure

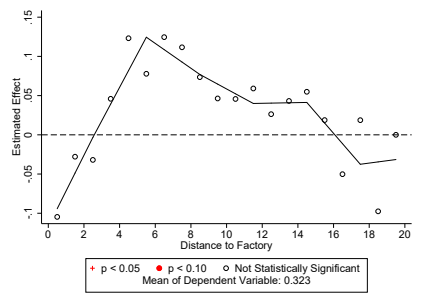
(a) Colonial Road Density (1900)



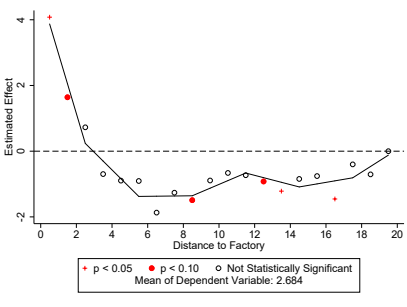
(b) Colonial Railroad Density (1900)



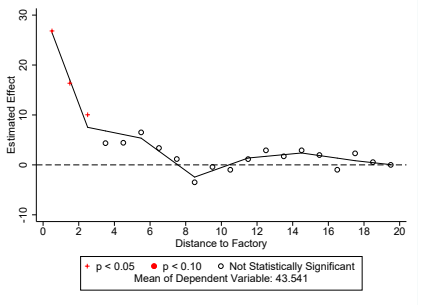
(c) Dirt Road (PODES 1980)



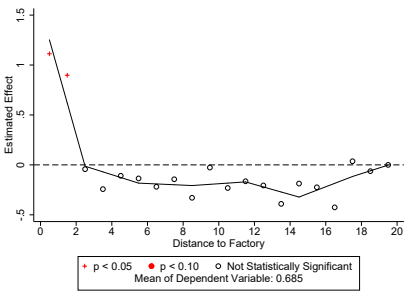
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



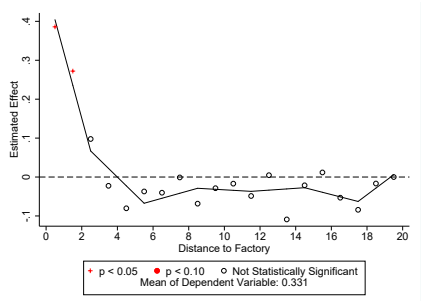
(f) Railroad Density (2017)



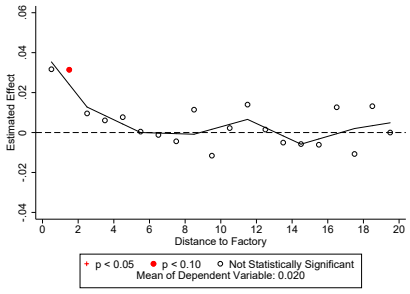
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-6: Other Public Goods

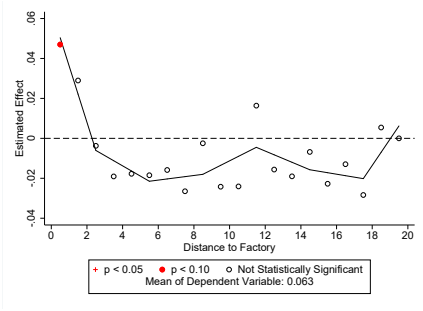
(a) Village Has Electricity (PODES 1980)



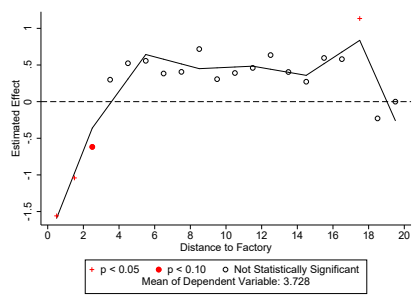
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



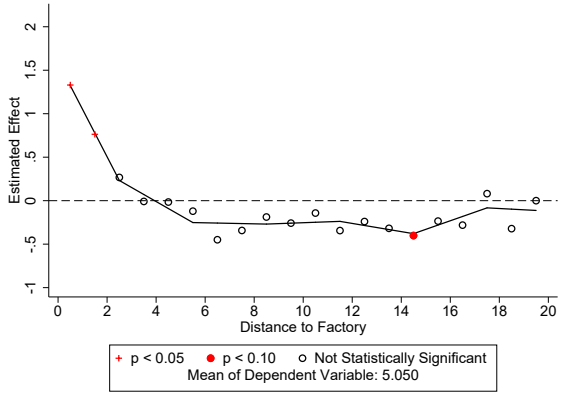
(d) Distance to Subdistrict Capital (2011 PODES)



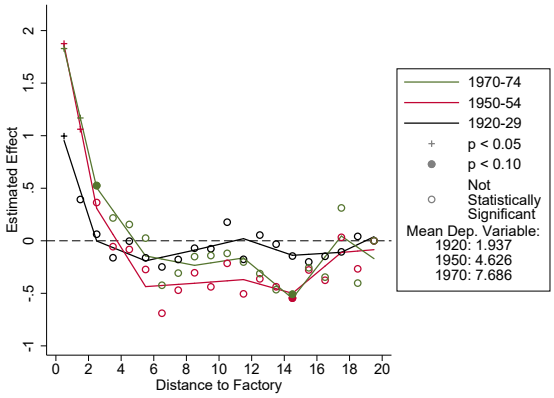
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-7: Education

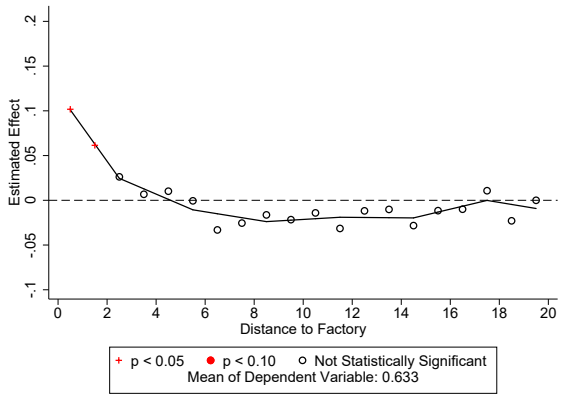
(a) Years Education (2000 Census)



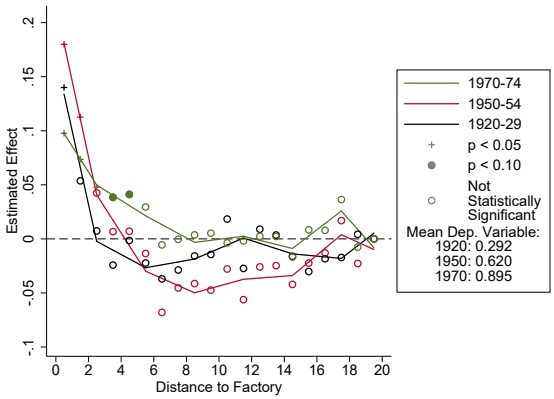
(b) Years Education by Cohort



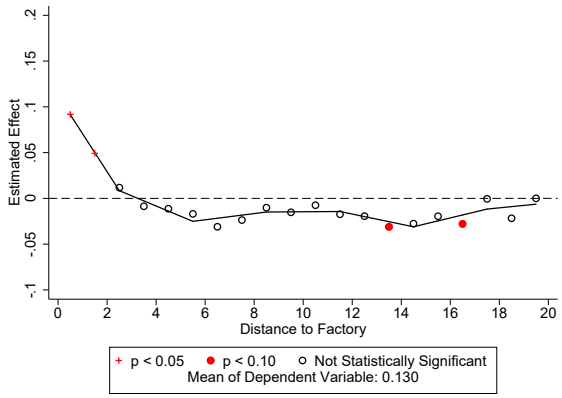
(c) Primary (2000 Census)



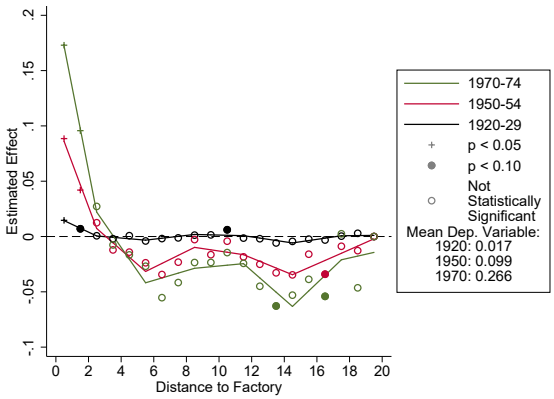
(d) Primary by Cohort



(e) High School (2000 Census)

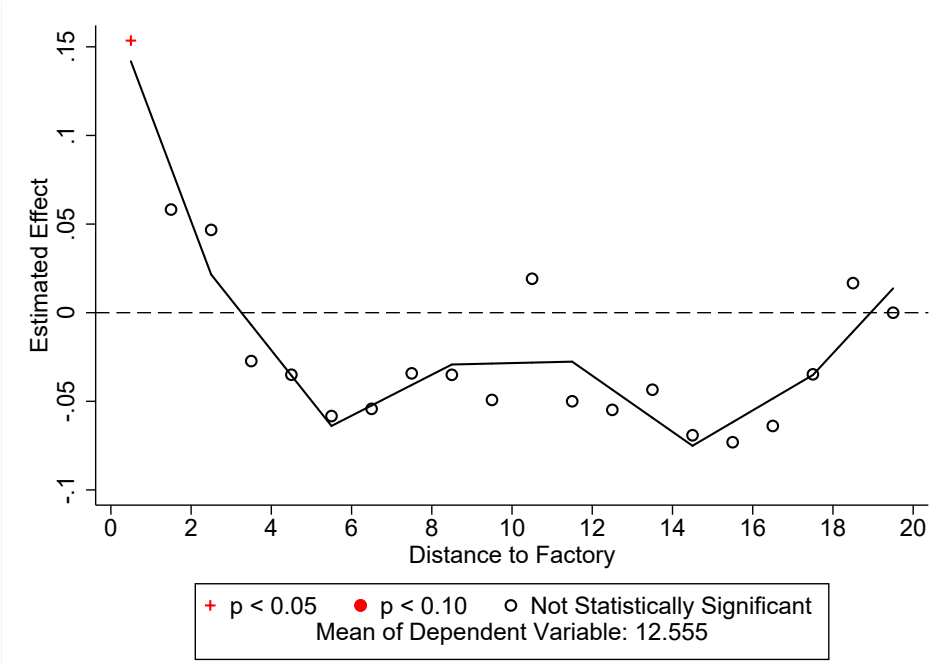


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure O-8: Expenditure (2001-11)

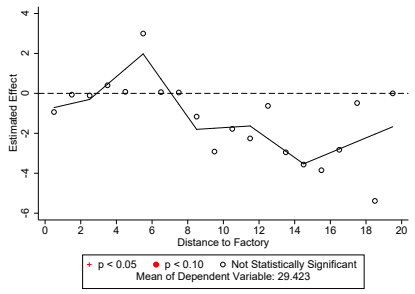


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

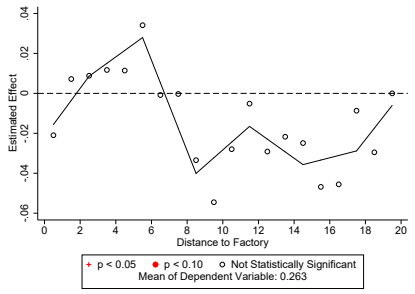
P Includes Catchment Indicator

Figure P-1: Geography

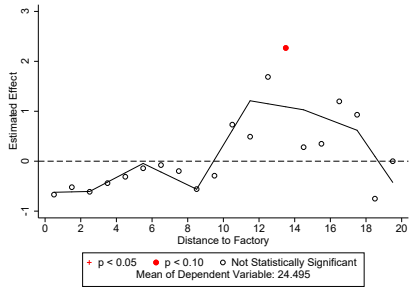
(a) Elevation



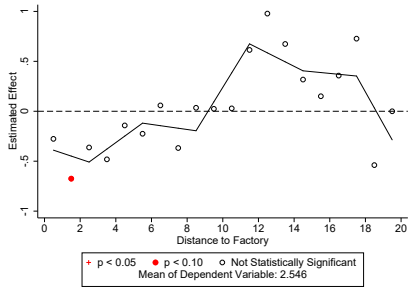
(b) Slope



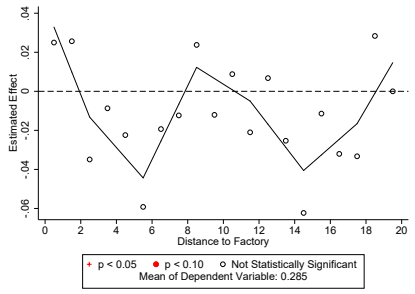
(c) Distance to Coast



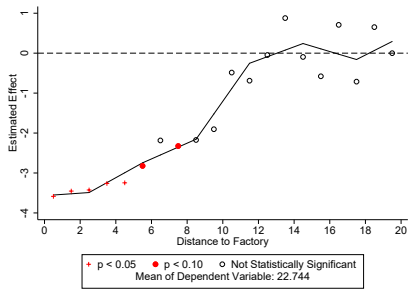
(d) Log Flow Accumulation



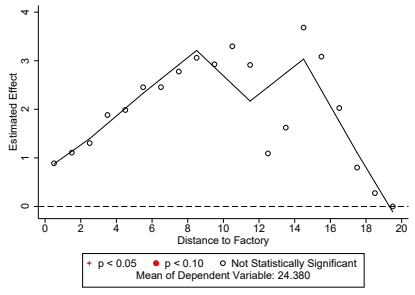
(e) Distance to River



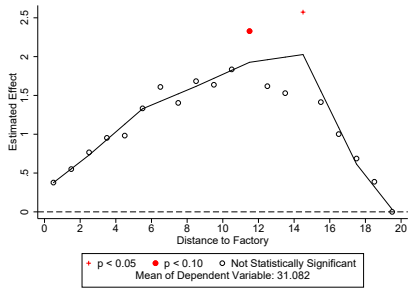
(f) Distance to Nearest 1830 Residency Capital



(g) Distance to Great Post Road



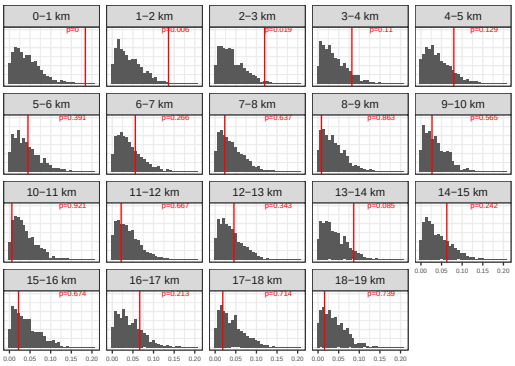
(h) Distance to VOC Port



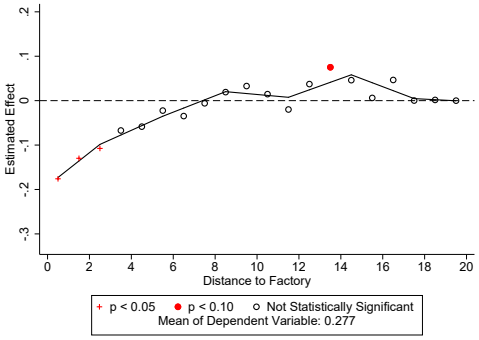
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy and nearest-factory fixed effects. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-2: Share in Agriculture (2001-11): Illustration of Methodology

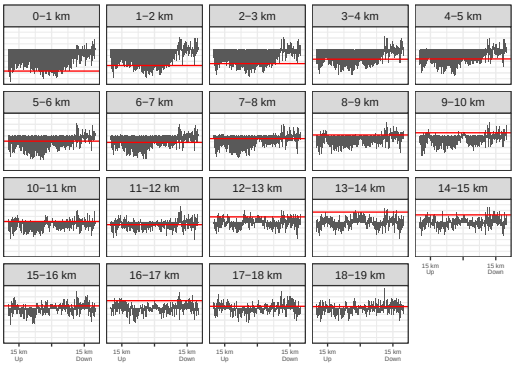
(a) Independent Shifts: Counterfactuals



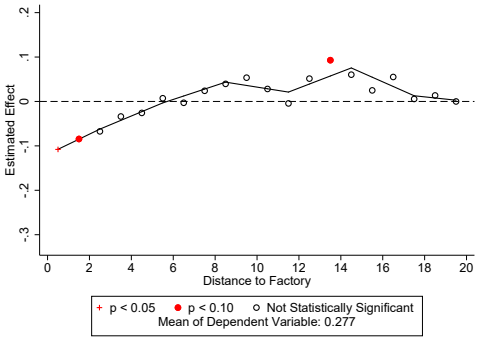
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



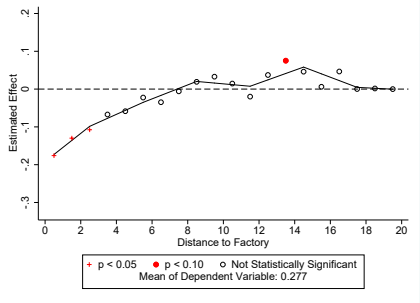
(d) Common Shifts: Plotted Coefficients



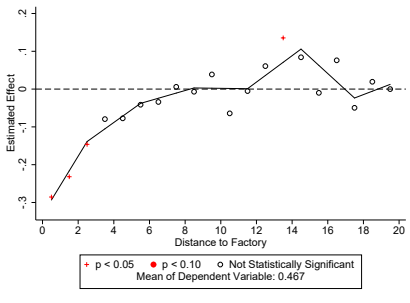
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for a catchment area dummy, nearest-factory fixed effects, geographic and pre-period characteristics, and survey year fixed effects. The sample is restricted to men aged 18 to 55. For each factory, a counterfactual was selected at random from the region of the river network that was sugar-suitable and within 5-20 km via river from the real factory. This procedure was repeated to construct 1,000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (a). Panel (c) plots coefficients on distance to counterfactual locations, where here placebos were chosen to be a specific distance upstream or downstream from the real factories. Real coefficients are shown as horizontal lines. Panel (d) plots differences between real coefficients for each bin and mean counterfactual coefficients, with the symbols indicating the real coefficients' positions in the distribution of counterfactual coefficients shown in panel (c).

Figure P-3: Industry and Agglomeration

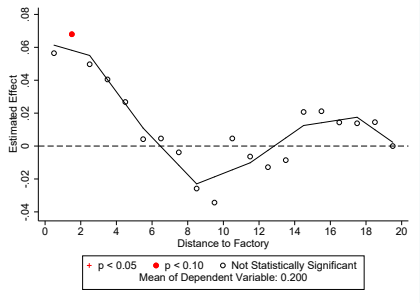
(a) Agriculture (Susenas 2001-11)



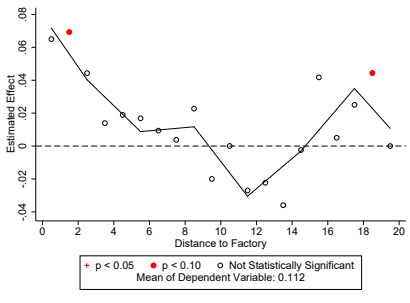
(b) Agriculture (Census 1980)



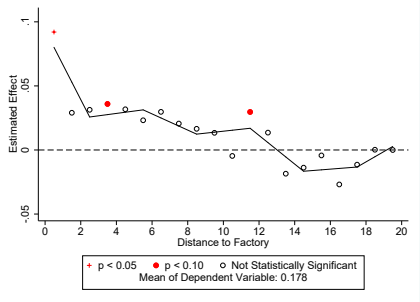
(c) Manufacturing (Susenas 2001-11)



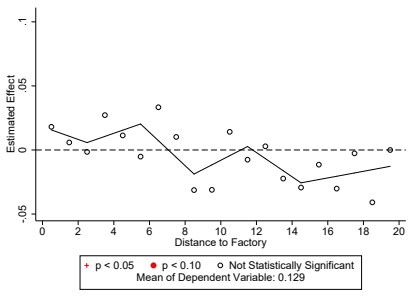
(d) Manufacturing (Census 1980)



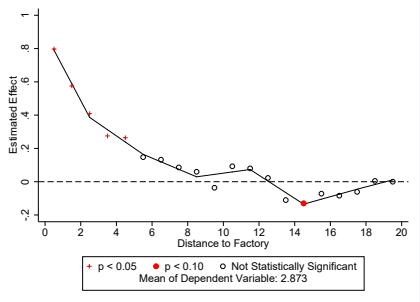
(e) Retail (Susenas 2001-11)



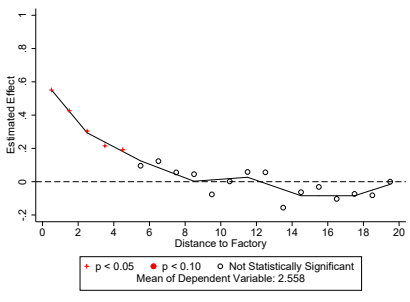
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



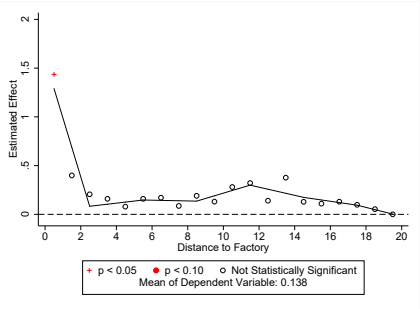
(h) Log Pop. Density (PODES 1980)



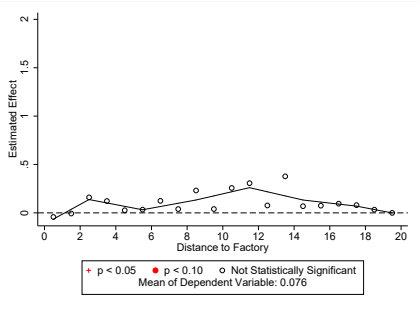
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearestcounterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-4: Sugar and Linked Industries

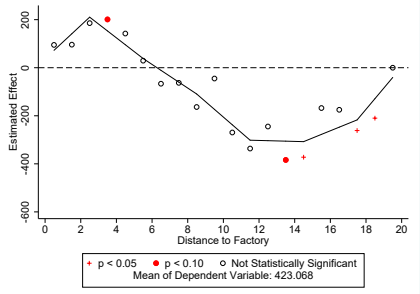
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



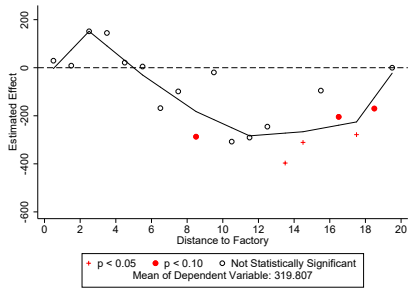
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



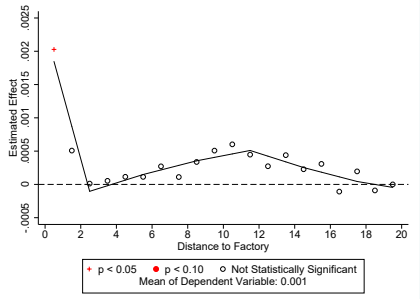
(c) Tons of Cane Grown (Full Sample, PODES 2003)



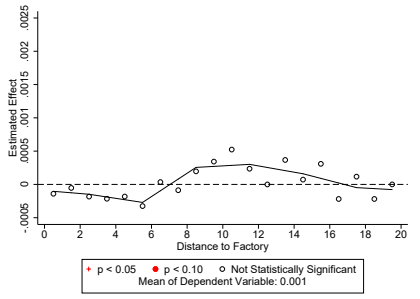
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



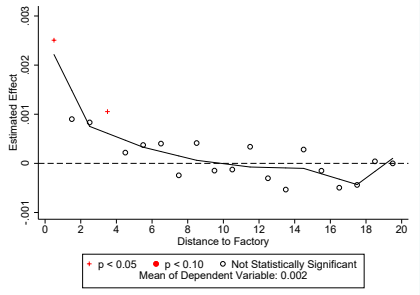
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



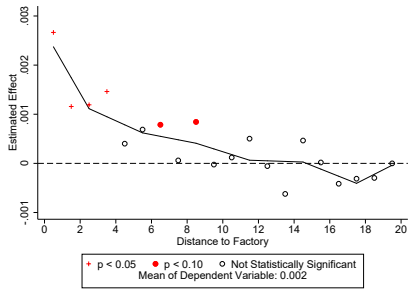
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



(g) Employment Share Downstream (Full Sample, Economic Census 2006)



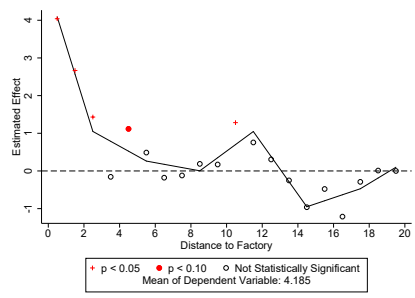
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



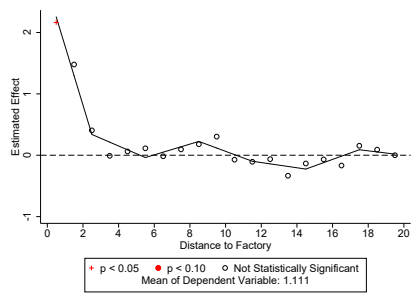
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-5: Infrastructure

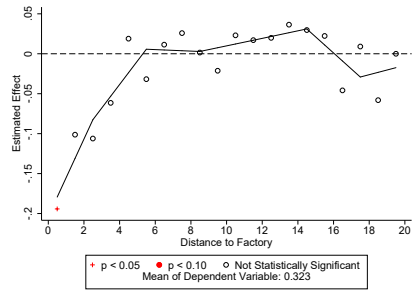
(a) Colonial Road Density (1900)



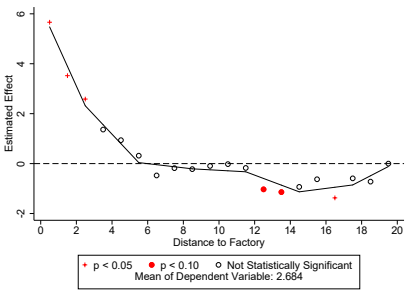
(b) Colonial Railroad Density (1900)



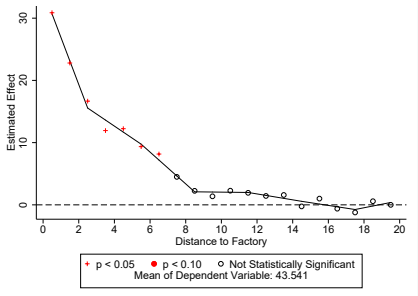
(c) Dirt Road (PODES 1980)



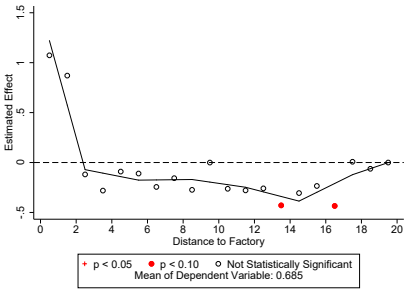
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



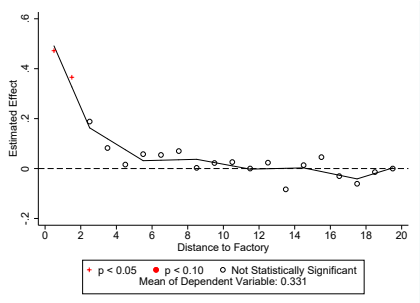
(f) Railroad Density (2017)



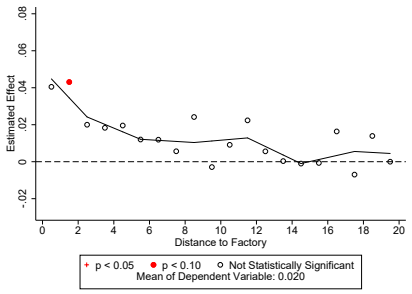
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-6: Other Public Goods

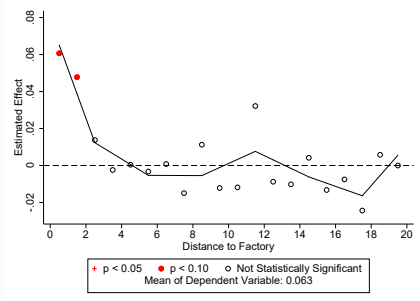
(a) Village Has Electricity (PODES 1980)



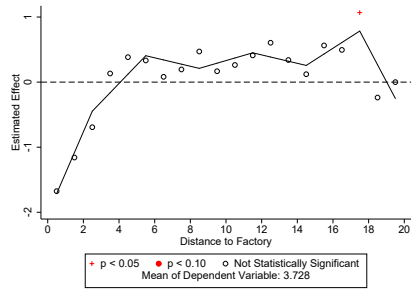
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



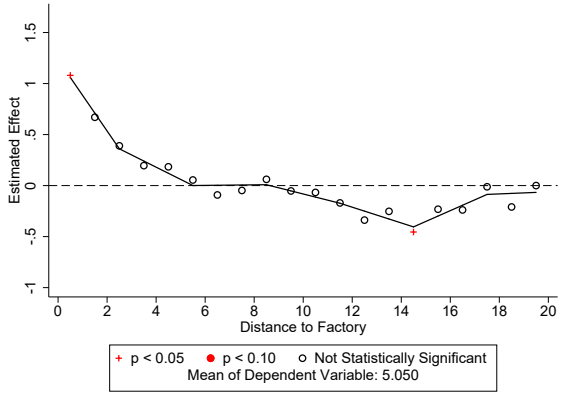
(d) Distance to Subdistrict Capital (2011 PODES)



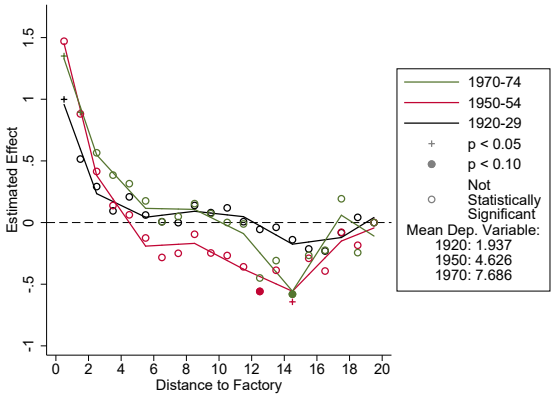
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, nearest-factory fixed effects and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Panel c) includes survey year fixed effects. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-7: Education

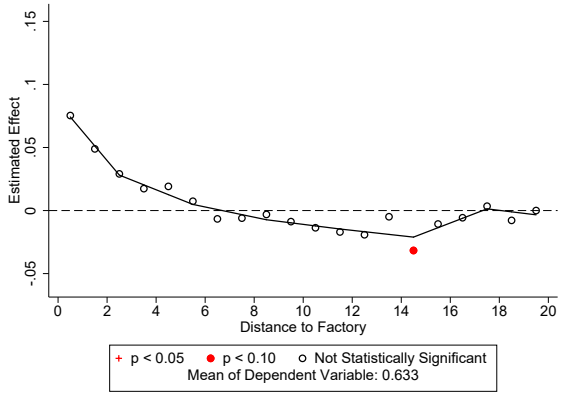
(a) Years Education (2000 Census)



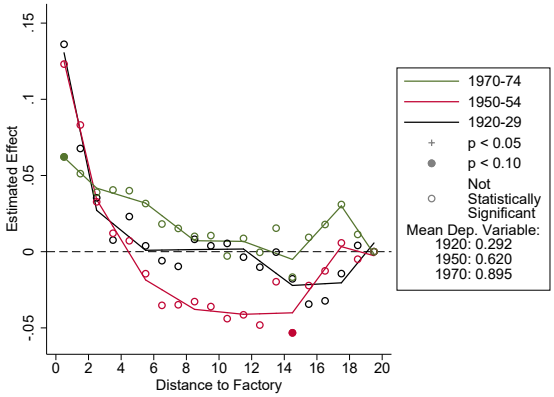
(b) Years Education by Cohort



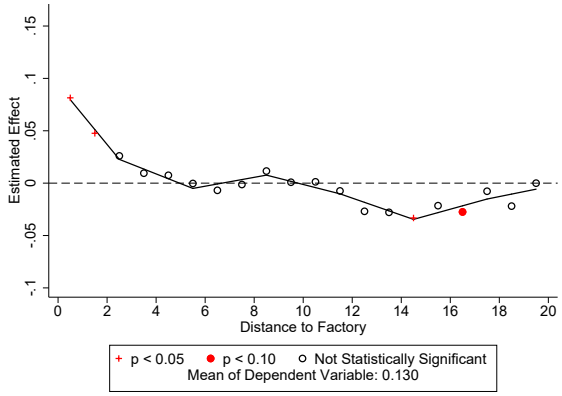
(c) Primary (2000 Census)



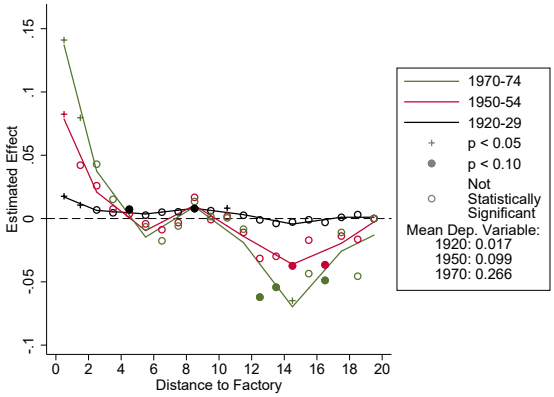
(d) Primary by Cohort



(e) High School (2000 Census)

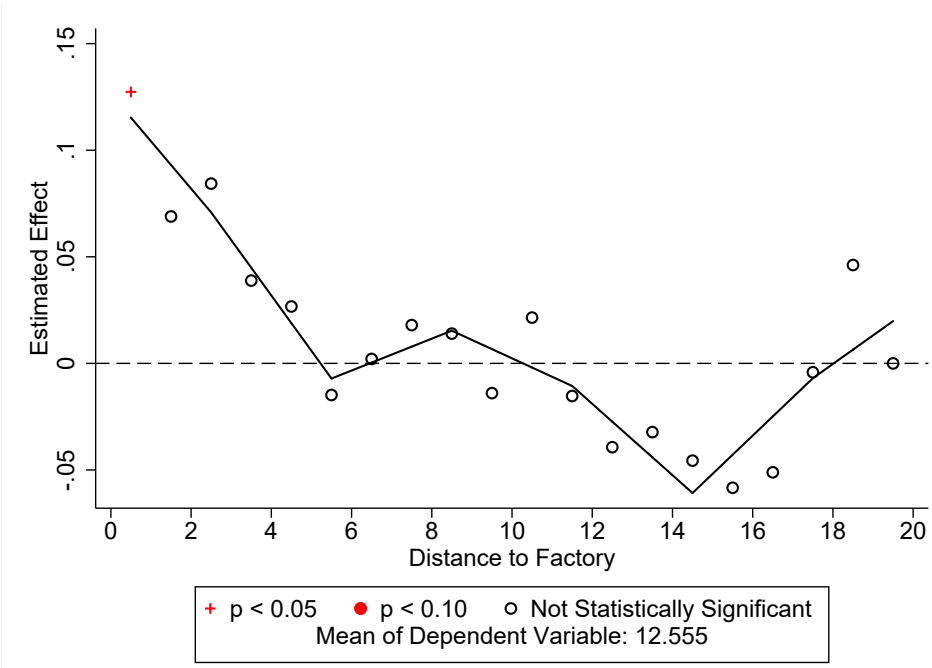


(f) High School by Cohort



Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Figure P-8: Expenditure (2001-11)



Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for a catchment area dummy, demographic variables, survey year fixed effects, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

Q Leontief-Weighted Outcomes

Our analysis uses outcome variables "Employment Share Upstream from Sugar" and "Employment Share Downstream from Sugar." The first component is employment share in industry i and village v:

$$s_{vi} = \frac{e_{vi}}{t_v}$$

Where e_{vi} is employment in industry i in village v, t_v is total employment in village v, and s_{vi} is the share of village v's employment that is in industry i. The source for e_{vi} is the 2006 Economic Census, which has data on employment at the firm level along with industry and village codes for each firm. We sum across firms within an industry and village to get e_{vi}. Since small firms (those with fewer than 20 employees) are surveyed, not censused, we use the provided sample weights to up-weight the employment numbers for small firms. When a village had no firms in a given industry, we count e_{vi} as zero. The source for t_v is the 2000 Population Census, from which we take the total number of individuals with non-missing industry as t_v.

The second component of the Leontief-weighted outcomes are the Leontief weights. We construct both upstream and downstream Leontief weights using the input-output table from the 2006 Economic Census. The input-output table shows the total value of sales between every pair of industries as well as final consumption sales for each industry. Using the input-output table, we construct the matrix A, where A_{ij} gives the sales from industry j to i divided by the total sales of industry i, or $\frac{sales_{j\to i}}{sales_i}$. Intuitively, this shows how important supplier j is to purchaser i, or the strength of the first-order upstream link between purchaser i and supplier j. We also use the input-output matrix to construct $\hat{A}$, where $\hat{A}_{ij}$ gives the sales from industry j to i divided by the total sales of industry j, or $\frac{sales_{j\to i}}{sales_j}$. Intuitively, this shows how important purchaser i is to supplier j, or the strength of the first-order downstream link between supplier j and purchaser i.

To capture higher-order upstream and downstream linkages, we construct $B = (I - A)^{-1}$ and $\hat{B} = (I - \hat{A})^{-1}$. These are the Leontief matrices. B_{ij} gives the strength of the upstream linkage between purchaser i and supplier j and $\hat{B}_{ij}$ gives the strength of the downstream linkage between supplier j and purchaser i. For both matrices, i is the purchaser and j is the supplier, but in B we see how important j is with respect to i while in $\hat{B}$ we see how important i is with respect to j.

Finally, we construct Leontief-weighted averages of employment share upstream and downstream from sugar processing. Let industry i be sugar processing:

$$s^u_{vi} = \frac{1}{\Sigma_B}\left(\sum_{j=1}^{N} s_{vj}B_{ij} - \mathbb{1}_{j=i}s_{vi}\right)$$
$$s^d_{vi} = \frac{1}{\Sigma_{\hat{B}}}\left(\sum_{j=1}^{N} s_{vj}\hat{B}_{ji} - \mathbb{1}_{j=i}s_{vi}\right)$$

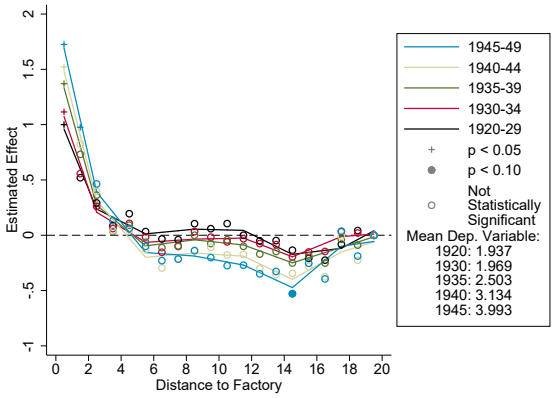
Here s^u_{vi} is the Leontief-weighted employment share upstream from sugar processing and s^d_{vi} is the Leontief-weighted employment share downstream from sugar processing. N is

the number of industries and Σ_B and $\Sigma_{\hat{B}}$ are the sums of the leontief weights: $\Sigma_B = \sum_{j=1}^{N} (B_{ij} - \mathbb{1}_{j=i})$ and $\Sigma_{\hat{B}} = \sum_{j=1}^{N} \left(\hat{B}_{ji} - \mathbb{1}_{j=i}\right)$. We subtract s_{vi} when $j = i$ to remove the own effect.

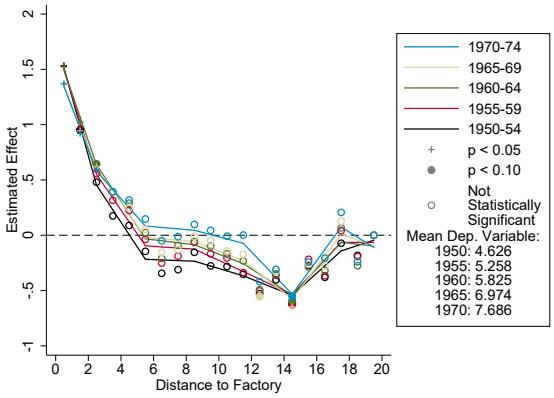
R Education for All Cohorts

Figure R-1: Education

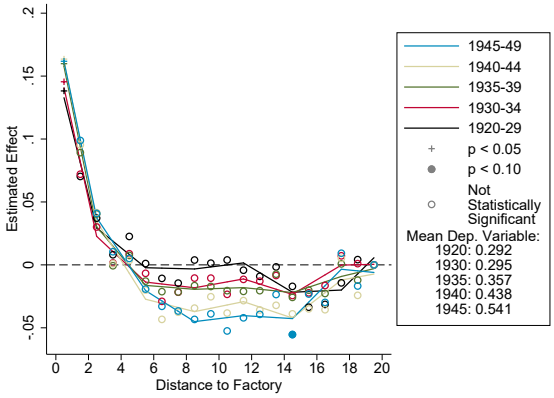
(a) Years Education: 1920-1949



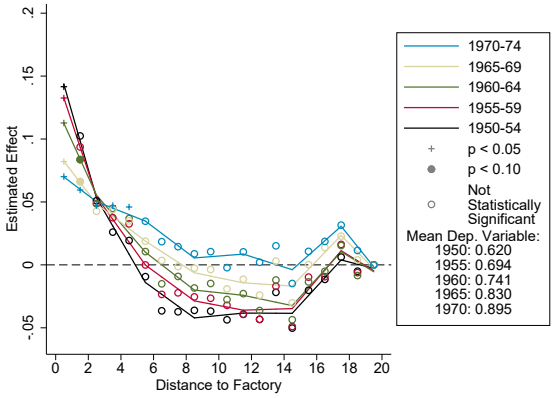
(b) Years Education: 1950-1974



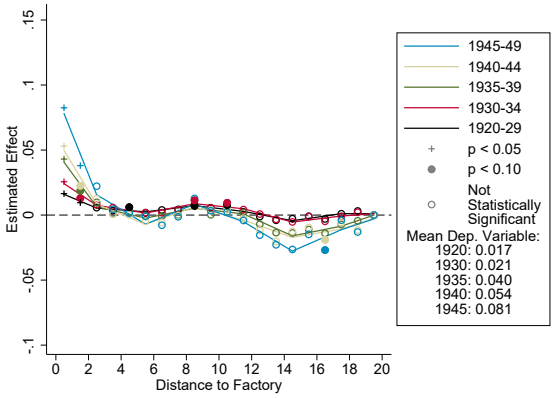
(c) Primary: 1920-1949



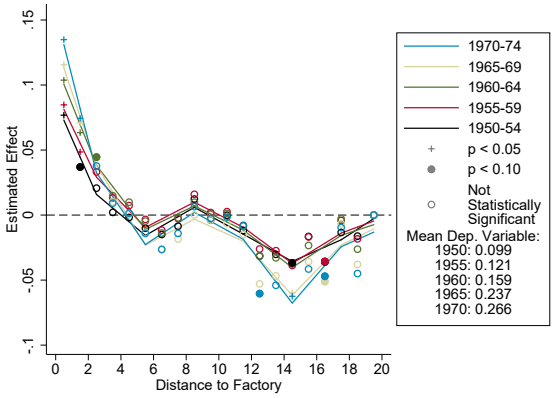
(d) Primary: 1950-1974



(e) High School: 1920-1949



(f) High School: 1950-1974

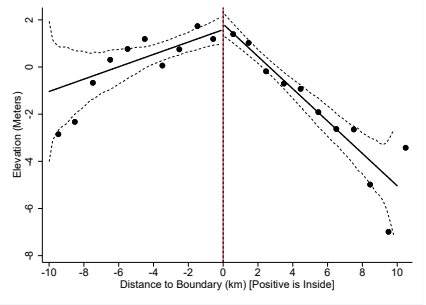


Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, and geographic and pre-period characteristics. The means of analogous estimates computed from 1,000 counterfactual factory configurations are subtracted from each coefficient. Each panel plots separate coefficients for five birth cohorts. The points are fit with a linear spline. p-values compare the effect of proximity to the nearest actual factory to the effects of proximity to the nearest counterfactual factory, computed from 1,000 counterfactual factory configurations.

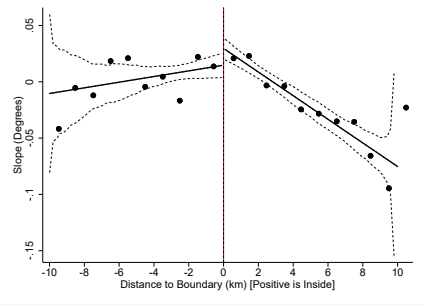
S RD Plots

Figure S-1: Geographic Characteristics

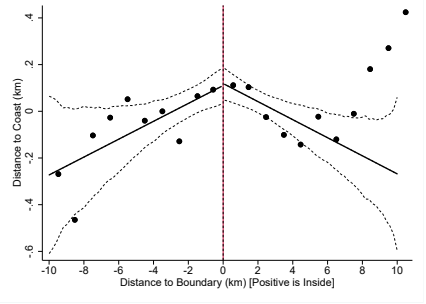
(a) Elevation



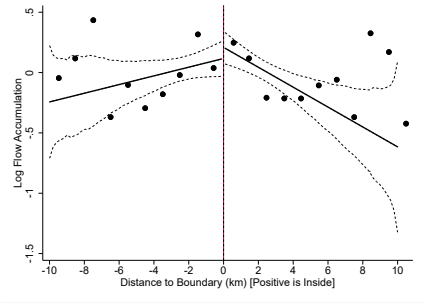
(b) Slope



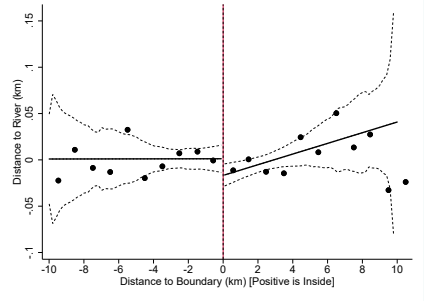
(c) Distance to Coast



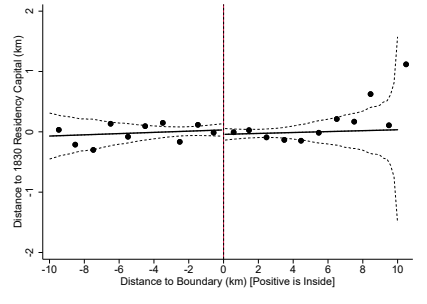
(d) Log Flow Accumulation



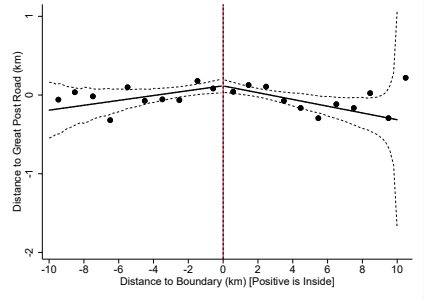
(e) Distance to River



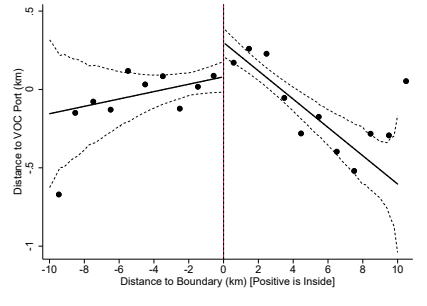
(f) Distance to 1830 Residency Capital



(g) Distance to Great Post Road



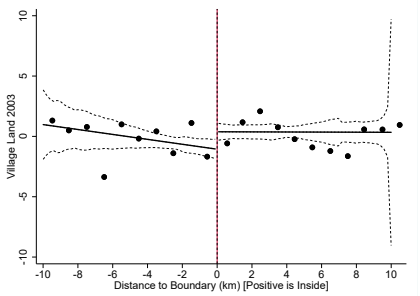
(h) Distance to VOC Port



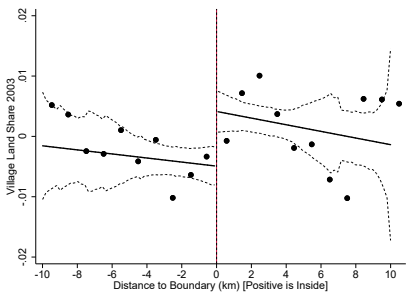
Notes: Points are binned residuals from a regression of the outcome variable on boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. Solid lines fit a local linear regression, estimated separately on either side of the threshold, and dashed lines fit 95% confidence intervals, computed using robust standard errors.

Figure S-2: Land Tenure

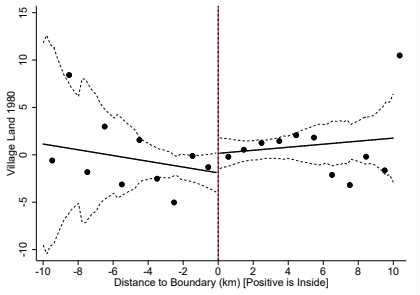
(a) Village Head 2003: Total Land



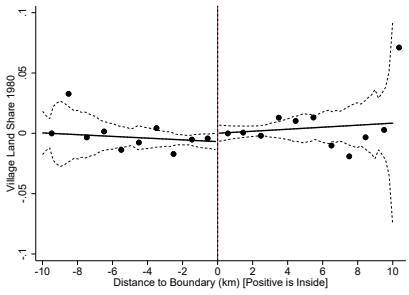
(b) Village Head 2003: Land Share



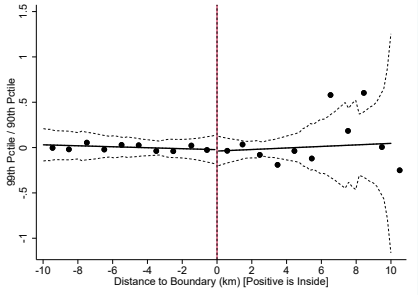
(c) Village Head 1980: Total Land



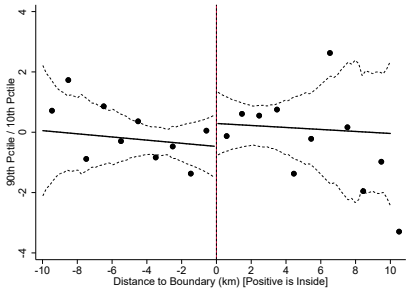
(d) Village Head 1980: Land Share



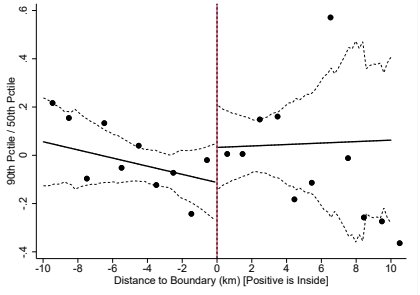
(e) 99th Pctile / 90th Pctile



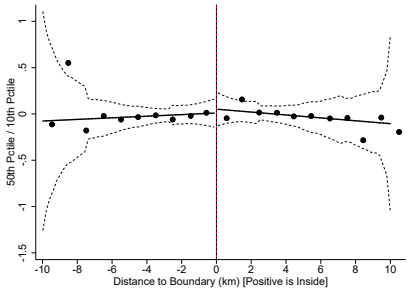
(f) 90th Pctile / 10th Pctile



(g) 90th Pctile / 50th Pctile



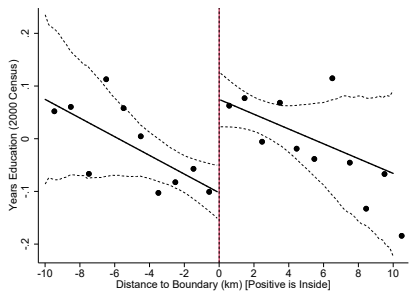
(h) 50th Pctile / 10th Pctile



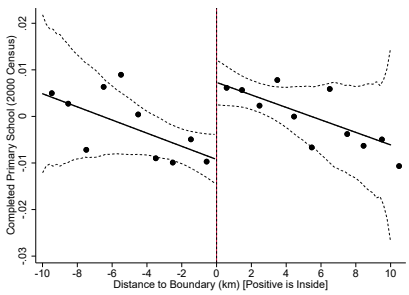
Notes: Points are binned residuals from a regression of the outcome variable on geographic and pre-period characteristics, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. Solid lines fit a local linear regression, estimated separately on either side of the threshold, and dashed lines fit 95% confidence intervals, computed using robust standard errors.

Figure S-3: Education

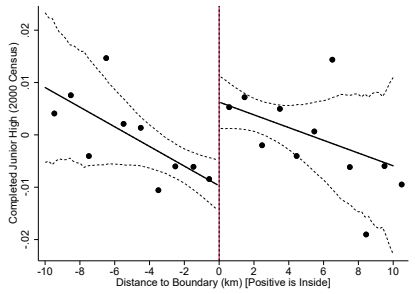
(a) Years Education 2000



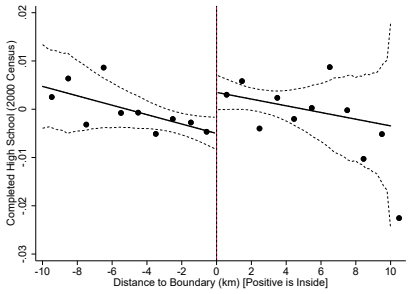
(b) Primary School 2000



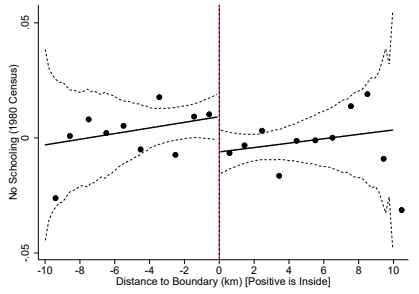
(c) Junior High 2000



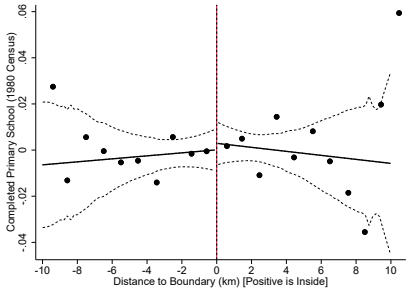
(d) High School 2000



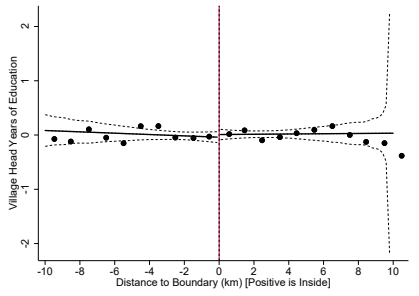
(e) No School 1980



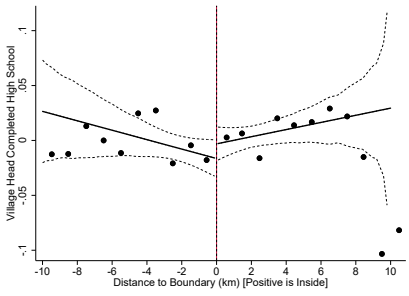
(f) Primary School 1980



(g) Village Head Years Education



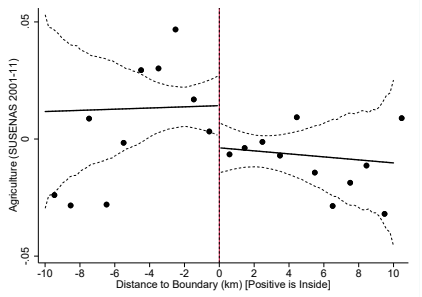
(h) Village Head High School



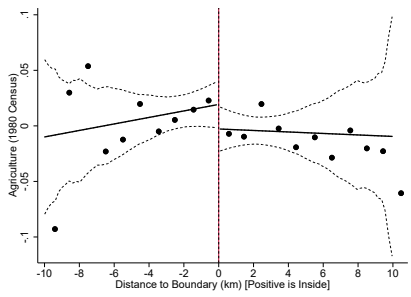
Notes: Points are binned residuals from a regression of the outcome variable on geographic and pre-period characteristics, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. Panels (a)-(f) include gender dummies and panels (g) and (h) include survey year fixed effects. Solid lines fit a local linear regression, estimated separately on either side of the threshold, and dashed lines fit 95% confidence intervals, computed using robust standard errors.

Figure S-4: Industrial Structure

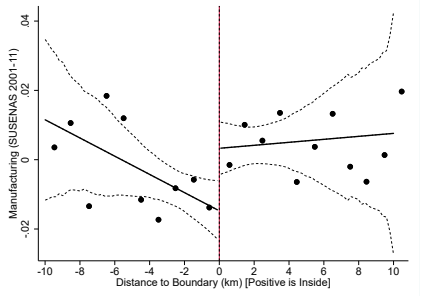
(a) Agriculture (SUSENAS)



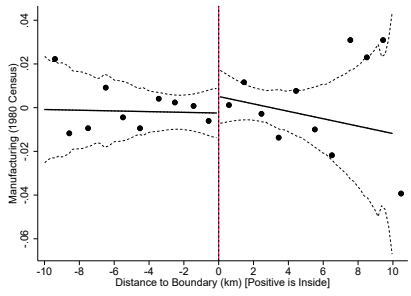
(b) Agriculture (1980 Census)



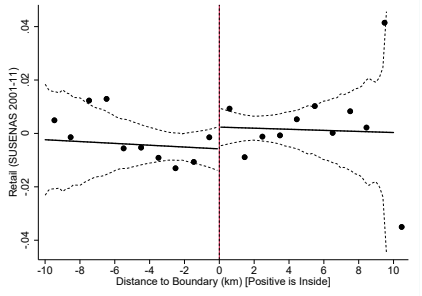
(c) Manufacturing (SUSENAS)



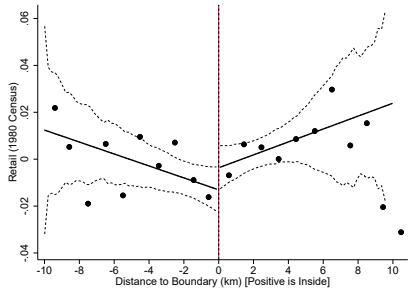
(d) Manufacturing (1980 Census)



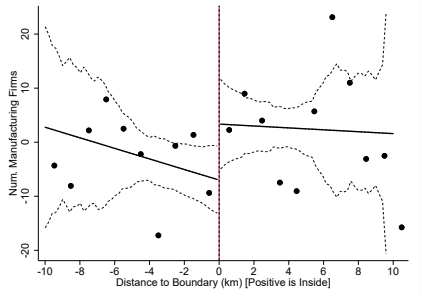
(e) Retail (SUSENAS)



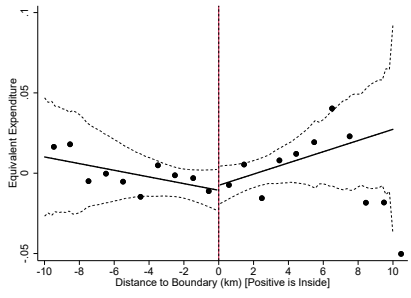
(f) Commerce (1980 Census)



(g) Number of Firms



(h) Log Equivalent Consumption



Notes: Points are binned residuals from a regression of the outcome variable on geographic and pre-period characteristics, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. The sample is restricted to men aged 18-55 in panels a) through f). Panels a), c), and e) include survey year fixed effects. Solid lines fit a local linear regression, estimated separately on either side of the threshold, and dashed lines fit 95% confidence intervals, computed using robust standard errors.

T Additional Public Goods Outcomes

We begin in Table T-1 by looking at the presence of schools in 1980, the earliest date for which systematic village level data on different types of schools are available through the Podes village census. 1980 was around the time of Indonesia's massive primary school building campaign - INPRES - which largely equalized access to basic primary schooling. The data break down primary schools into those that were built by the INPRES campaign and those that were not. Columns (1) and (2) focus on non-INPRES schools. There is not an impact on the number of buildings, whereas the effect on the number of teachers is positive and marginally significant.

Columns (3) and (4) show that cultivation villages received significantly fewer INPRES schools and teachers, almost surely an indicator of *greater* education beforehand since INPRES was targeted based on the availability of pre-existing primary education (Duflo, 2001). Subjected villages were around three percentage points less likely to receive an INPRES school, relative to a sample mean of 0.36. Moreover, subjected villages were 2 percentage points more likely to have a junior high school, relative to a sample mean of 6 percent of villages that had them (junior high schools were not covered by INPRES). There is not a statistically significant impact on the presence of a high school, which only two percent of villages had.

Table T-2 considers public goods more generally. Effects tend to be positive but not statistically significant.

Table T-1: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High	High
	Buildings	Teachers	Buildings	Teachers	Schools	Schools
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.011	0.172	-0.035	-0.212	0.019	0.007
	(0.019)	(0.104)	(0.020)	(0.081)	(0.009)	(0.006)
Obs	4,205	4,205	4,205	4,205	4,205	4,205
Clusters	380	380	380	380	380	380
Mean	0.43	2.81	0.36	1.37	0.06	0.02

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table T-2: Public Goods

	Intercity Road Density	Local Road Density	Railroad Density	Paved Road	Dirt Road	Has Electricty	Distance to Subdistrict Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cultivation	-0.045	0.622	0.171	0.053	-0.012	0.027	-0.472
	(0.266)	(1.293)	(0.113)	(0.025)	(0.014)	(0.026)	(0.201)
Obs	4,549	4,549	4,549	4,550	4,550	4,205	4,560
Clusters	383	383	383	383	383	380	383
Mean	2.72	44.18	0.76	0.69	0.06	0.32	3.77

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

U Crop Choice

This appendix examines the impacts of being subjected to forced labor on subsequent cropping choices, using data from the Indonesian Village Census. The table below shows that there are not economically or statistically significant differences in what villages near the catchment boundaries grow today. We consider cash crops (anything other than a staple cereal); crops that are specifically classified in the data as plantation crops (coconuts, coffee, rubber, cocoa, cloves, oil palm, sugar cane, cashews, tobacco, candlenuts, nutmeg, cinnamon, vanilla, and tea); sugar; and non-sugar plantation crops. We consider the extensive margin (whether anyone in the village grows the crop), as well as the share of village land devoted to the crop. Effects tend to be small relative to the mean and statistically insignificant.

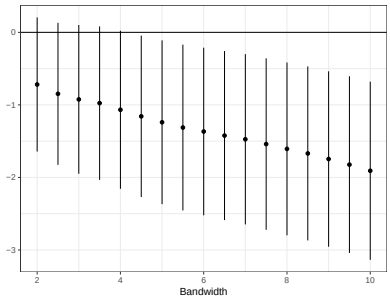
	All Cash Crops		Plantation Crops		Sugar		Non-Sugar Pl. Cr.	
	Any	Share	Any	Share	Any	Share	Any	Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.022	-0.016	0.002	-0.008	0.024	-0.001	-0.030	-0.007
	(0.025)	(0.011)	(0.025)	(0.007)	(0.024)	(0.005)	(0.022)	(0.004)
Obs	4,430	4,430	4,430	4,430	4,430	4,430	4,430	4,430
Clusters	379	379	379	379	379	379	379	379
Mean	0.71	0.13	0.49	0.08	0.35	0.05	0.23	0.03

V Robustness to Bandwidth Plots

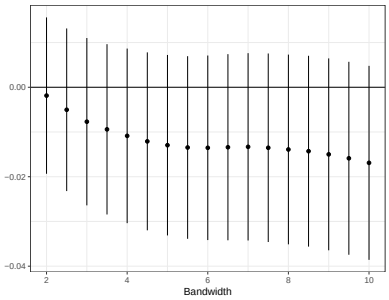
We are not aware of a well-accepted multi-dimensional RD optimal bandwidth, so we examine robustness to the full range of feasible alternative bandwidths. Each plot shows point estimates and confidence intervals of γ using the baseline RD specification and different bandwidths between 2 and 10 kilometers, with the bandwidth under consideration denoted on the x-axis and the point estimate on the y-axis. Since the catchment areas are small, treated villages are never more than 10 kilometers from the nearest catchment boundary. Results are broadly robust to the choice of bandwidth, with narrower bandwidths by construction tending to lead to somewhat noisier estimates.

Figure V-1: Geographic Characteristics

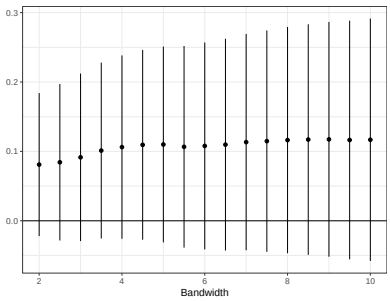
(a) Elevation



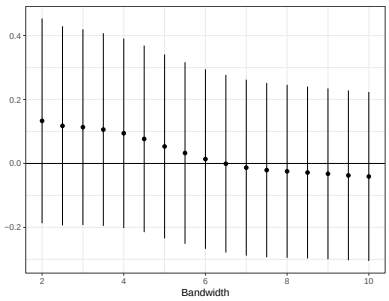
(b) Slope



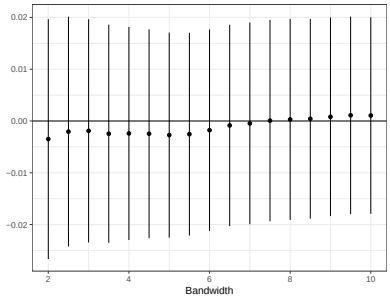
(c) Distance to Coast



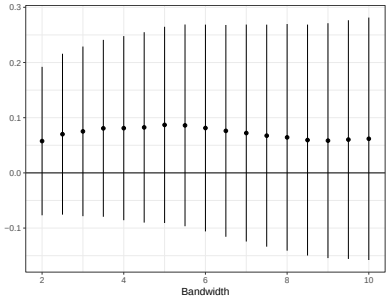
(d) Log Flow Accumulation



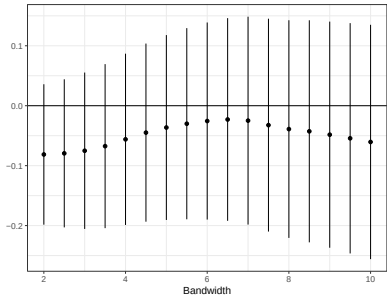
(e) Distance to River



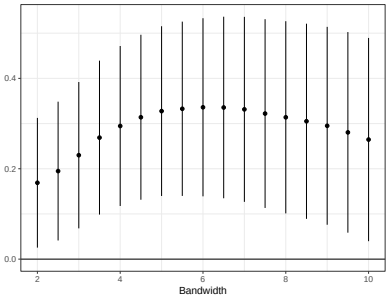
(f) Distance to 1830 Residency Capital



(g) Distance to Great Post Road



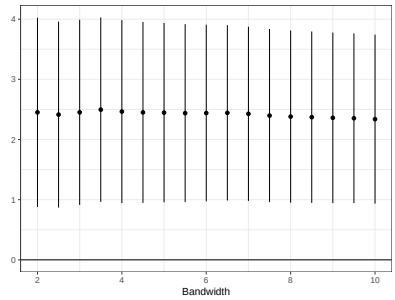
(h) Distance to VOC Port



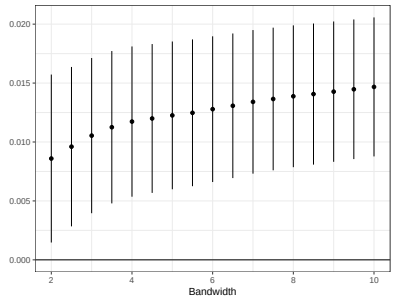
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Lines show 90% confidence intervals, constructed using standard errors clustered by subdistrict.

Figure V-2: Land Tenure

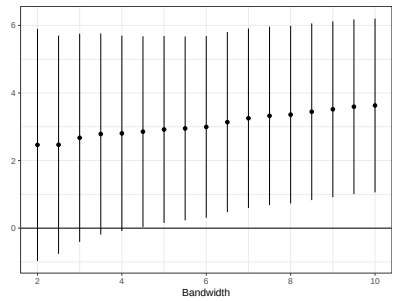
(a) Village Head 2003: Total Land



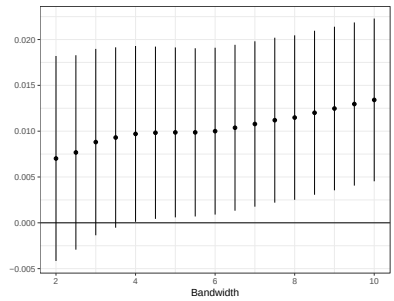
(b) Village Head 2003: Land Share



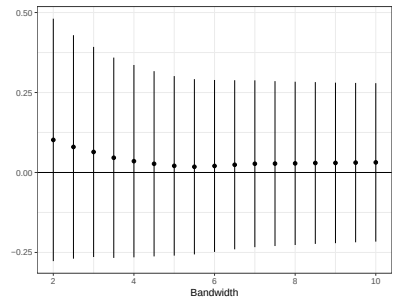
(c) Village Head 1980: Total Land



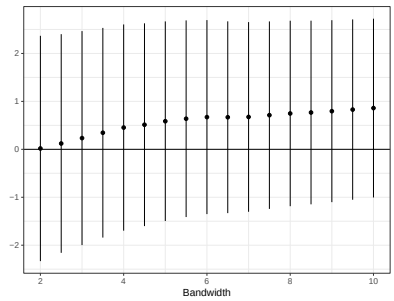
(d) Village Head 1980: Land Share



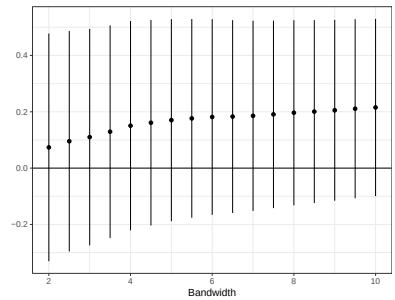
(e) 99th Pctile / 90th Pctile



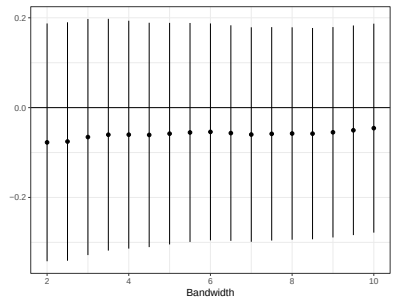
(f) 90th Pctile / 10th Pctile



(g) 90th Pctile / 50th Pctile



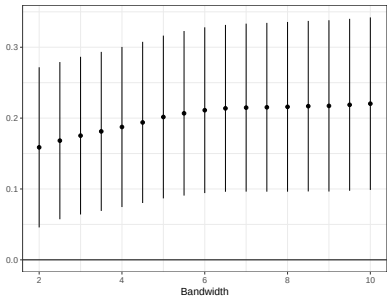
(h) 50th Pctile / 10th Pctile



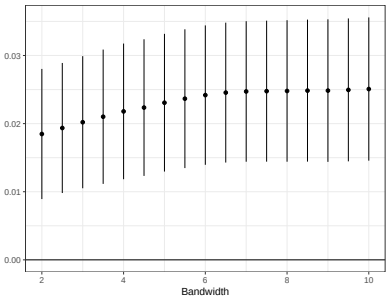
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic and pre-period characteristics, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Lines show 90% confidence intervals, constructed using standard errors clustered by subdistrict.

Figure V-3: Education

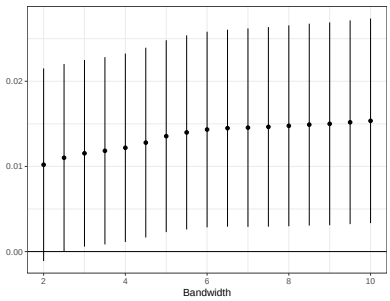
(a) Years Education 2000



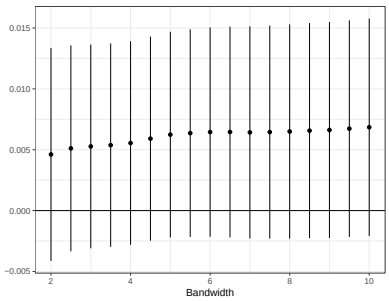
(b) Primary School 2000



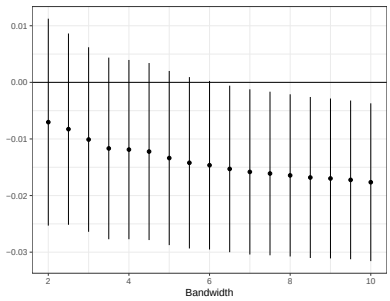
(c) Junior High 2000



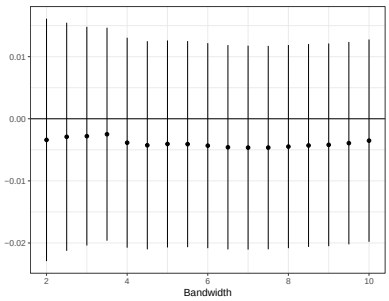
(d) High School 2000



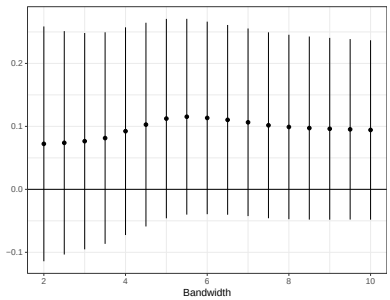
(e) No School 1980



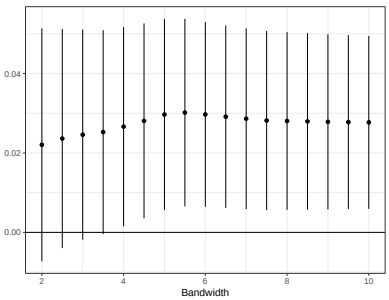
(f) Primary School 1980



(g) Village Head Years Education



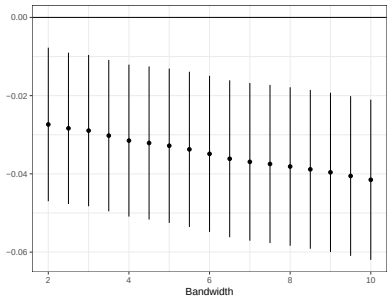
(h) Village Head High School



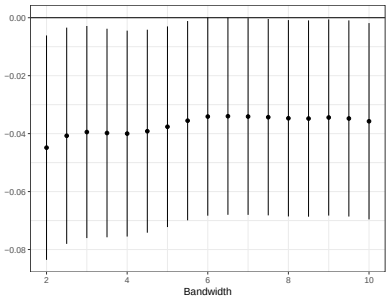
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic and pre-period characteristics, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Lines show 90% confidence intervals, constructed using standard errors clustered by subdistrict.

Figure V-4: Industrial Structure

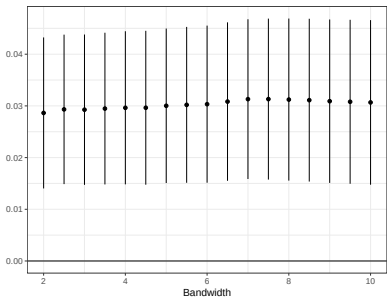
(a) Agriculture (SUSENAS)



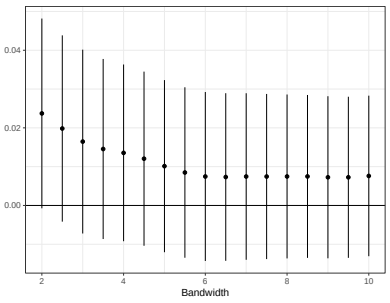
(b) Agriculture (1980 Census)



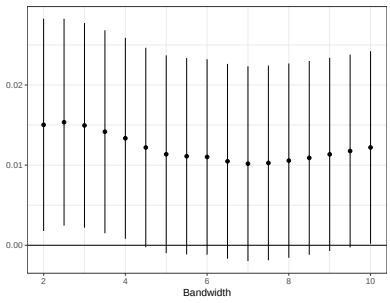
(c) Manufacturing (SUSENAS)



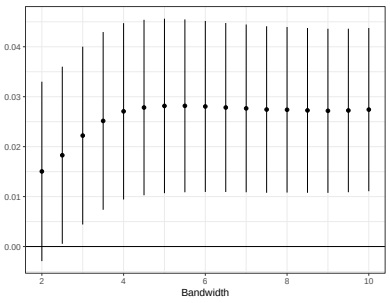
(d) Manufacturing (1980 Census)



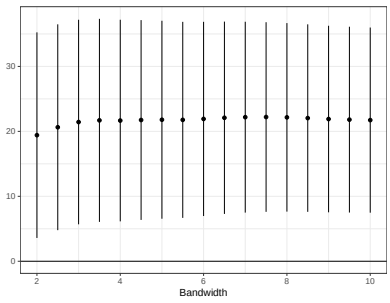
(e) Retail (SUSENAS)



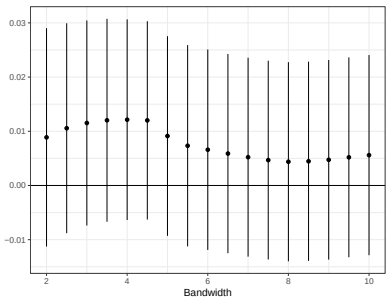
(f) Commerce (1980 Census)



(g) Number of Firms



(h) Log Equivalent Consumption



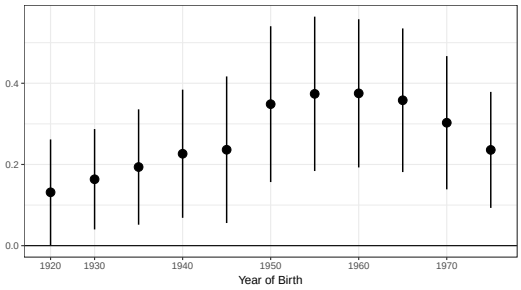
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic and pre-period characteristics, boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Lines show 90% confidence intervals, constructed using standard errors clustered by subdistrict.

W Single Linear Latitude-Longitude Polynomial

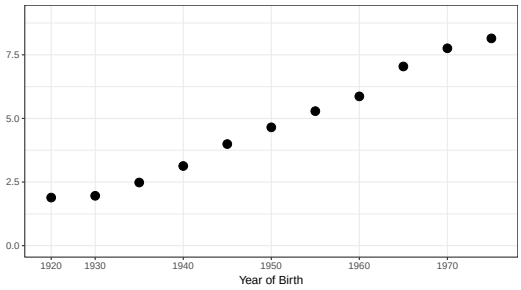
This section considers robustness to the specification of the RD polynomial. Results change little when a single multi-dimensional RD polynomial is used - rather than estimating the polynomial separately by catchment area

Figure W-1: Education by Cohort: Subjected Villages (2000 Census)

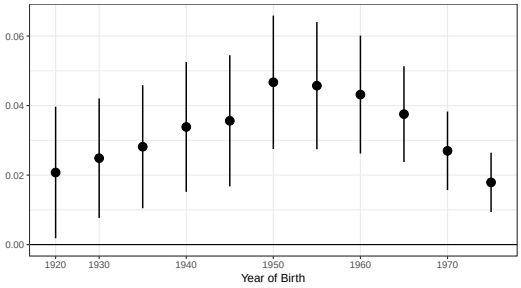
(a) Years of Schooling



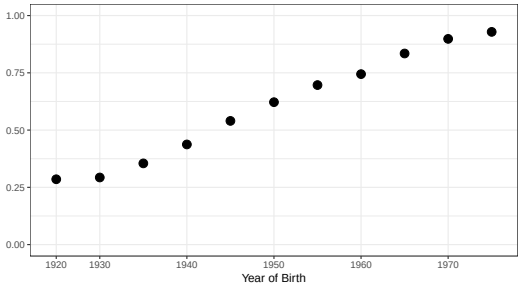
(b) Years of Schooling Levels



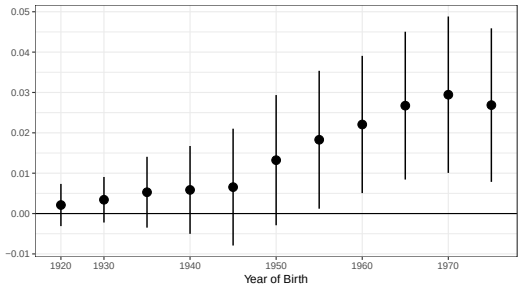
(c) Primary Completion



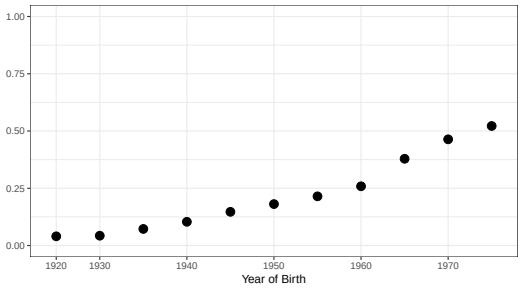
(d) Primary Completion Levels



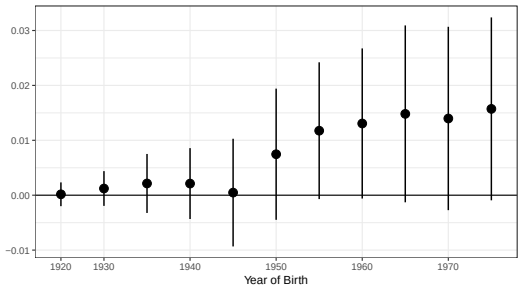
(e) Junior High Completion



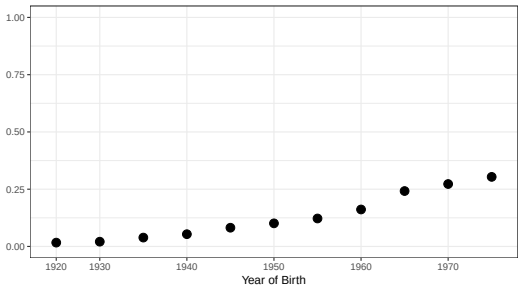
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude. Robust standard errors are clustered by subdistrict.

Table W-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.127	-0.015	0.054	-0.031	-0.004	0.063	-0.079	0.327
	(0.776)	(0.013)	(0.143)	(0.161)	(0.013)	(0.151)	(0.135)	(0.150)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table W-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷ 90th Pctile	90th Pctile ÷ 10th Pctile	90th Pctile ÷ 50th Pctile	50th Pctile ÷ 10th Pctile
	Total Land	Land Share	Total Land	Land Share				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.046	0.014	3.199	0.014	0.037	0.816	0.204	-0.037
	(0.826)	(0.003)	(1.503)	(0.005)	(0.153)	(1.100)	(0.186)	(0.133)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table W-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.219	0.025	0.015	0.007	-0.016	-0.003	0.094	0.026
	(0.077)	(0.007)	(0.008)	(0.006)	(0.008)	(0.010)	(0.088)	(0.013)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table W-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.041	0.030	0.013	-0.036	0.008	0.027	21.152	0.005
	(0.013)	(0.011)	(0.007)	(0.021)	(0.013)	(0.010)	(8.836)	(0.011)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

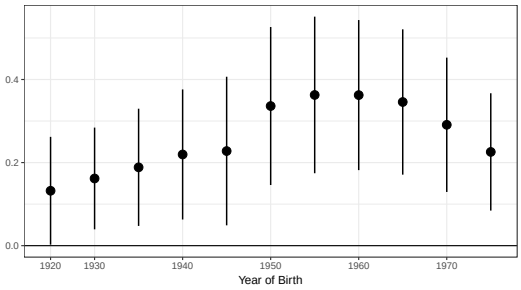
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

X Quadratic Latitude-Longitude Polynomial

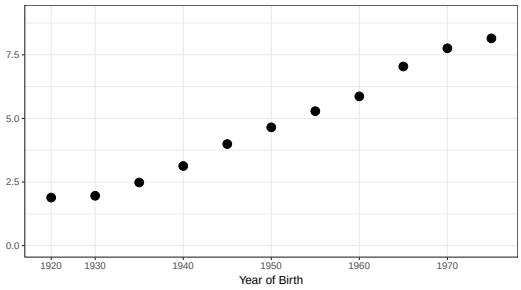
This appendix documents that estimates are very similar when the RD polynomial is specified as quadratic, instead of the linear specification used in the baseline.

Figure X-1: Education by Cohort: Subjected Villages (2000 Census)

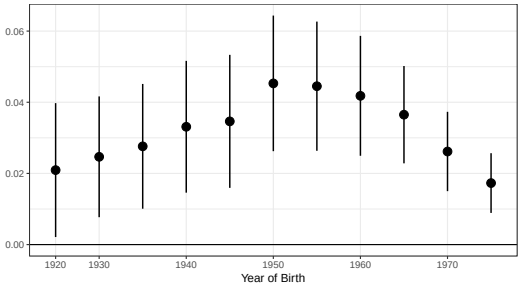
(a) Years of Schooling



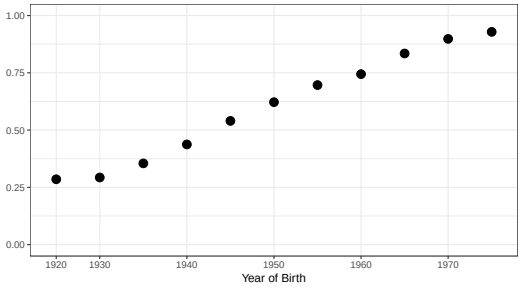
(b) Years of Schooling Levels



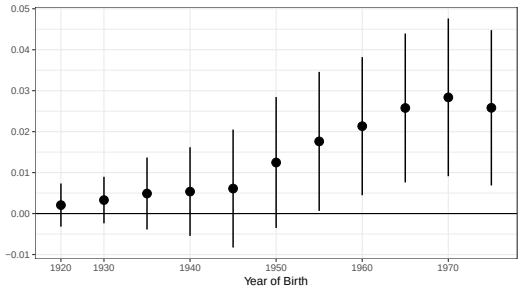
(c) Primary Completion



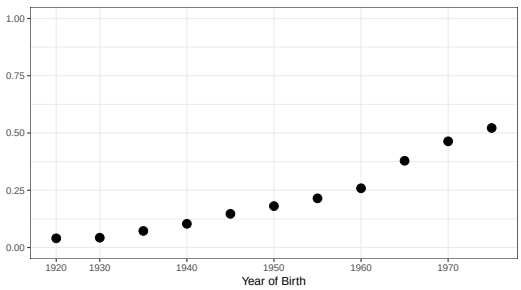
(d) Primary Completion Levels



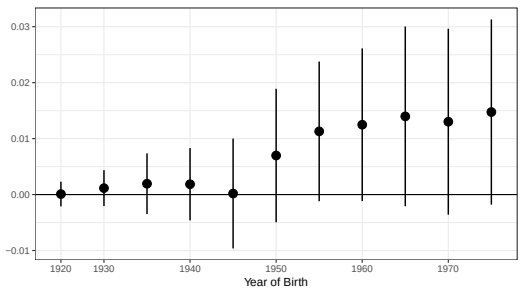
(e) Junior High Completion



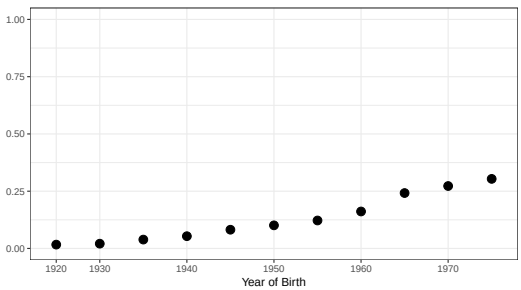
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a quadratic polynomial in latitude and longitude. Robust standard errors are clustered by subdistrict.

Table X-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.151	-0.016	0.089	-0.026	-0.004	0.083	-0.110	0.332
	(0.769)	(0.013)	(0.126)	(0.161)	(0.013)	(0.149)	(0.127)	(0.150)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a quadratic polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table X-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.121	0.014	3.251	0.014	0.048	0.824	0.205	-0.038
	(0.837)	(0.003)	(1.512)	(0.005)	(0.152)	(1.098)	(0.186)	(0.131)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a quadratic polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table X-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.209	0.024	0.014	0.006	-0.016	-0.003	0.090	0.026
	(0.076)	(0.006)	(0.008)	(0.006)	(0.008)	(0.010)	(0.088)	(0.013)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a quadratic polynomial in latitude and longitude. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table X-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.039	0.029	0.014	-0.034	0.007	0.027	20.752	0.006
	(0.013)	(0.011)	(0.007)	(0.021)	(0.013)	(0.010)	(8.780)	(0.011)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

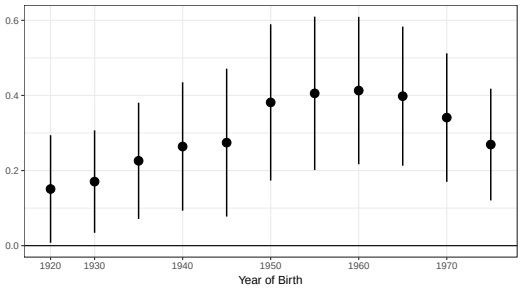
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a quadratic polynomial in latitude and longitude. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Y One-Dimensional Linear RD Polynomial

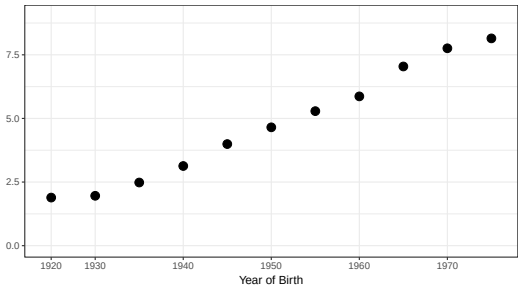
This appendix illustrates that results are broadly similar when a linear polynomial in distance to the boundary, estimated separately on either side of the threshold, is included instead of the multi-dimensional RD polynomial used in the baseline.

Figure Y-1: Education by Cohort: Subjected Villages (2000 Census)

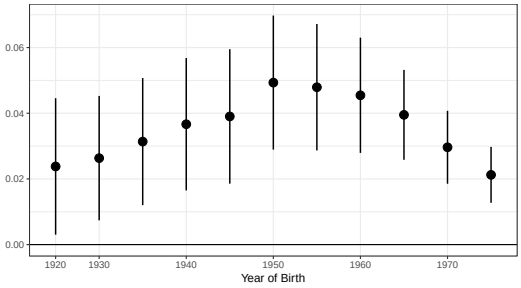
(a) Years of Schooling



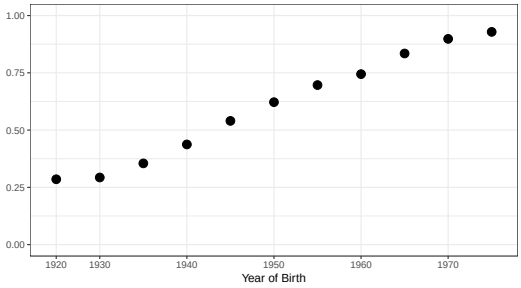
(b) Years of Schooling Levels



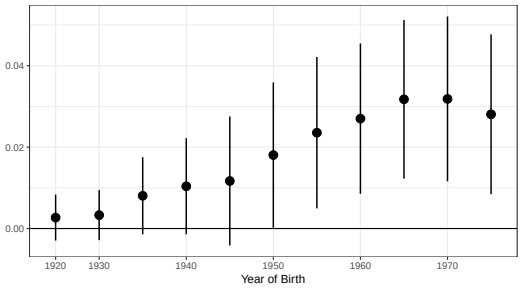
(c) Primary Completion



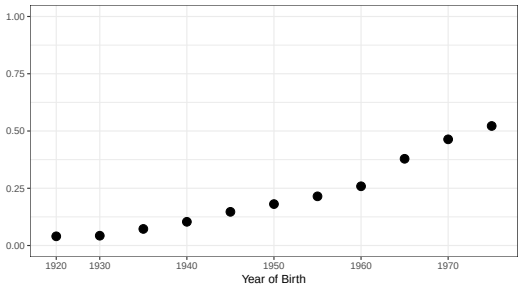
(d) Primary Completion Levels



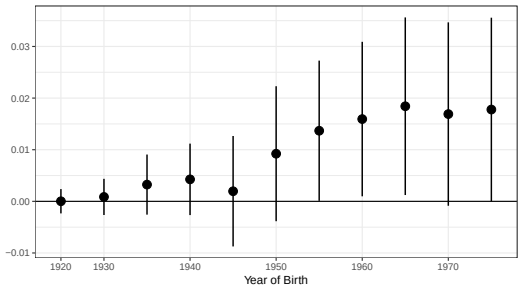
(e) Junior High Completion



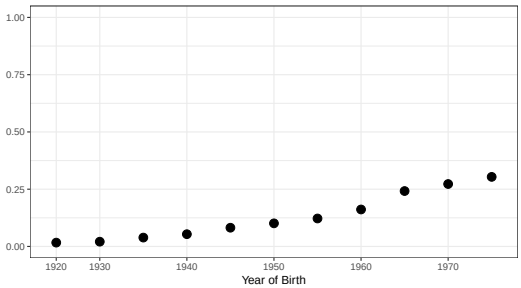
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in distance to the boundary. Robust standard errors are clustered by subdistrict.

Table Y-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.076	0.014	0.103	0.075	-0.020	0.007	0.180	0.391
	(0.777)	(0.013)	(0.171)	(0.174)	(0.014)	(0.175)	(0.165)	(0.176)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in distance to the boundary. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Y-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	1.981	0.013	3.078	0.011	-0.005	0.741	0.175	-0.003
	(0.895)	(0.004)	(1.765)	(0.006)	(0.159)	(1.174)	(0.195)	(0.158)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in distance to the boundary. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Y-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.240	0.026	0.018	0.008	-0.016	-0.001	0.086	0.022
	(0.081)	(0.007)	(0.008)	(0.006)	(0.009)	(0.011)	(0.090)	(0.014)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in distance to the boundary. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Y-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.040	0.026	0.016	-0.035	0.014	0.021	14.960	0.007
	(0.014)	(0.011)	(0.008)	(0.021)	(0.013)	(0.011)	(9.183)	(0.012)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

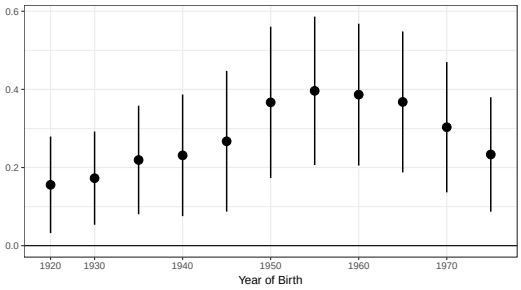
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in distance to the boundary. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Z 25 Km Boundary Segment Fixed Effects

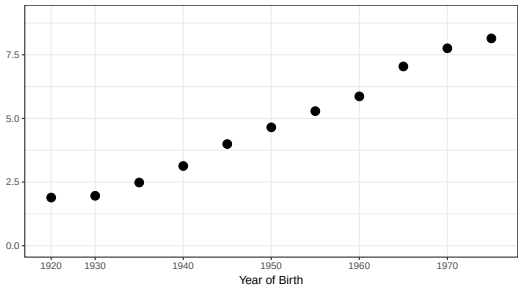
This appendix shows that results are robust to changing the length of the boundary segment fixed effects to 25 kilometers, as compared to the 10 kilometer segments used in the baseline.

Figure Z-1: Education by Cohort: Subjected Villages (2000 Census)

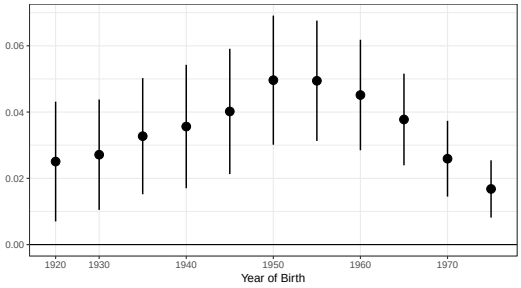
(a) Years of Schooling



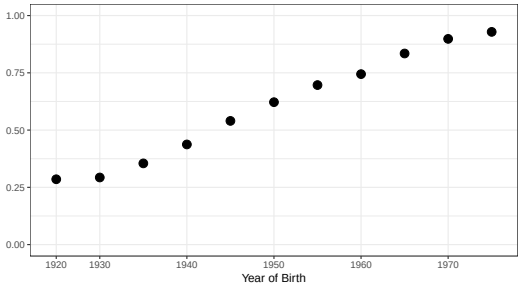
(b) Years of Schooling Levels



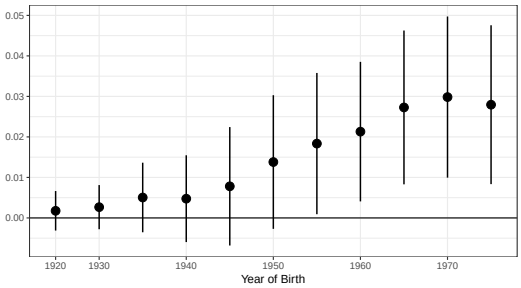
(c) Primary Completion



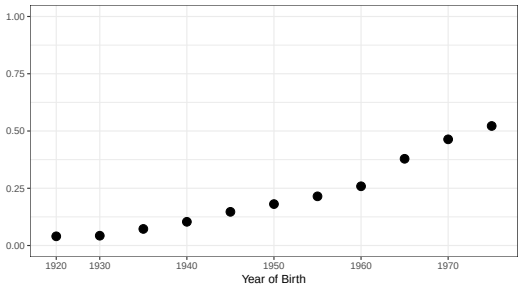
(d) Primary Completion Levels



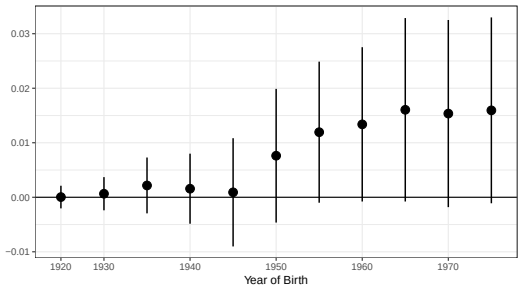
(e) Junior High Completion



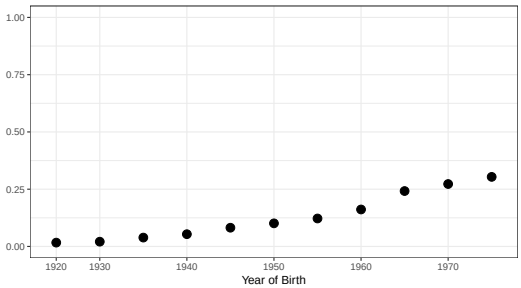
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table Z-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.036	-0.016	0.326	-0.049	0.000	0.038	0.038	0.570
	(0.826)	(0.014)	(0.123)	(0.150)	(0.012)	(0.167)	(0.153)	(0.170)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Z-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.536	0.016	2.855	0.009	0.108	1.036	0.261	-0.057
	(0.904)	(0.004)	(1.584)	(0.006)	(0.145)	(1.052)	(0.177)	(0.130)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Z-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.226	0.026	0.015	0.007	-0.015	-0.005	0.079	0.024
	(0.075)	(0.006)	(0.008)	(0.006)	(0.008)	(0.009)	(0.083)	(0.014)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Z-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.045	0.026	0.018	-0.040	0.013	0.027	20.639	0.004
	(0.013)	(0.012)	(0.007)	(0.018)	(0.011)	(0.010)	(8.590)	(0.011)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

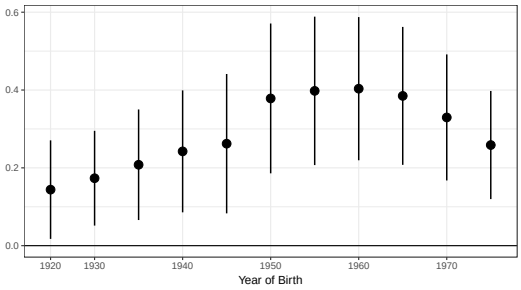
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AA No Geographic or Pre-Period Controls

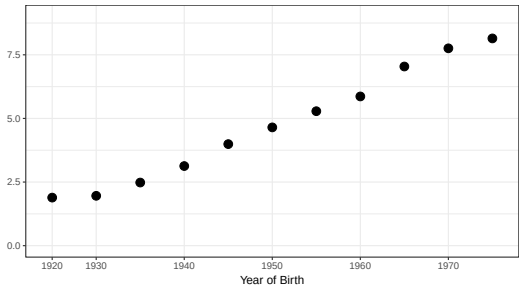
This appendix documents the robustness of the results to excluding the geographic and pre-period controls.

Figure AA-1: Education by Cohort: Subjected Villages (2000 Census)

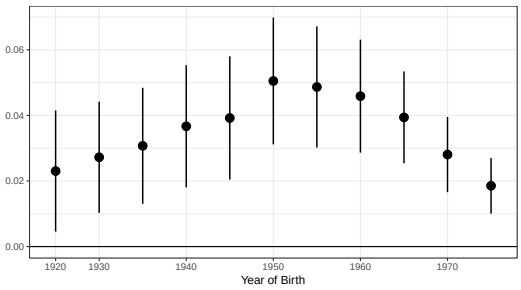
(a) Years of Schooling



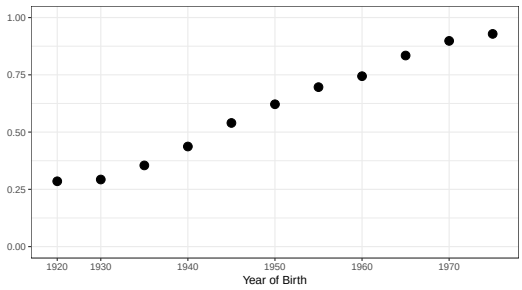
(b) Years of Schooling Levels



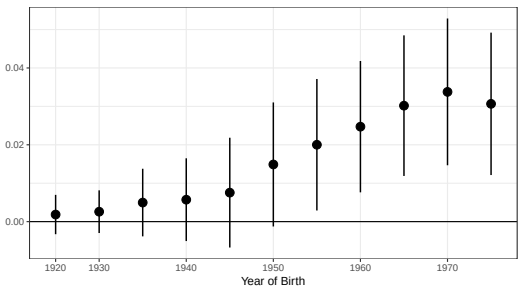
(c) Primary Completion



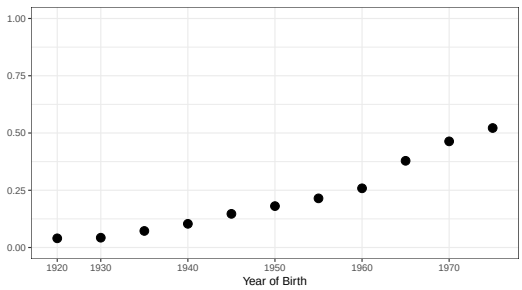
(d) Primary Completion Levels



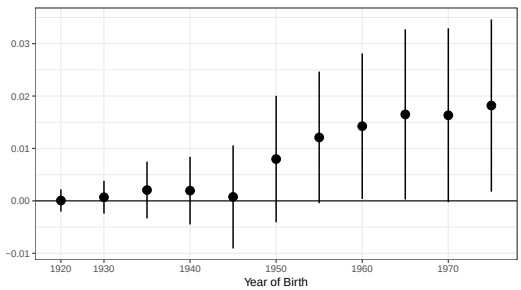
(e) Junior High Completion



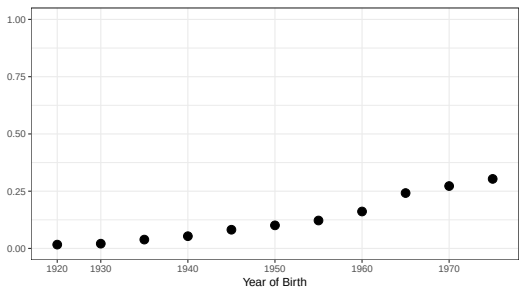
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table AA-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-1.908	-0.017	0.117	-0.041	0.001	0.062	-0.060	0.265
	(0.744)	(0.013)	(0.106)	(0.160)	(0.012)	(0.133)	(0.118)	(0.136)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AA-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.341	0.015	3.344	0.014	0.039	0.706	0.195	-0.043
	(0.857)	(0.004)	(1.495)	(0.005)	(0.151)	(1.131)	(0.195)	(0.136)
Obs	4,554	4,554	4,209	4,110	4,206	4,206	4,206	4,206
Clusters	383	383	380	380	381	381	381	381
Mean	18.60	0.09	23.93	0.11	3.60	17.05	3.79	4.28

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AA-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.240	0.027	0.017	0.008	-0.018	-0.002	0.103	0.030
	(0.076)	(0.007)	(0.007)	(0.006)	(0.009)	(0.010)	(0.086)	(0.013)
Obs	16,139,613	16,139,613	16,139,613	16,139,613	653,443	653,318	26,653	26,653
Clusters	383	383	383	383	358	358	383	383
Mean	5.09	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AA-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.046	0.032	0.013	-0.037	0.008	0.026	18.919	0.007
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)	(8.544)	(0.012)
Obs	130,479	130,479	130,479	127,901	127,901	127,901	4,553	144,178
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.74	12.55

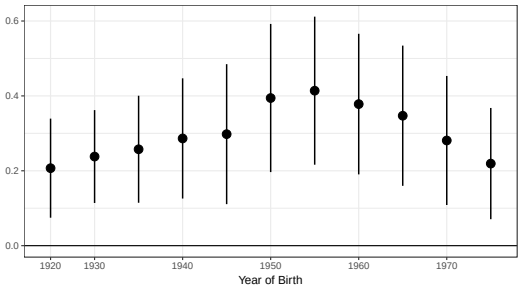
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AB Alternative Sample Restriction

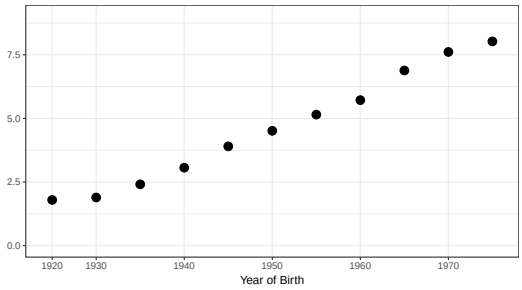
In the baseline we limit the sample to villages suitable for sugar cultivation - since non-suitable places are not a plausible counterfactual - but estimates are robust to varying the sample inclusion criteria. An alternative is to limit to boundary segments that are similarly suitable for sugar cultivation on either side of the boundary segment. This appendix shows that estimates are broadly robust to this alternative sample restriction.

Figure AB-1: Education by Cohort: Subjected Villages (2000 Census)

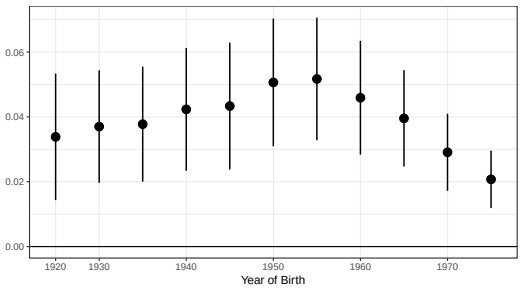
(a) Years of Schooling



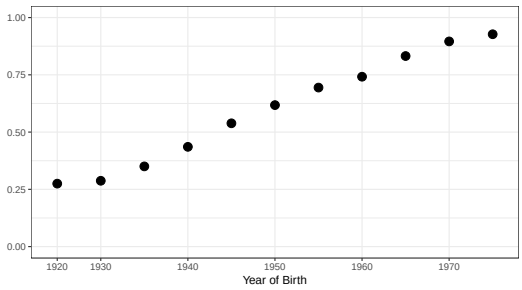
(b) Years of Schooling Levels



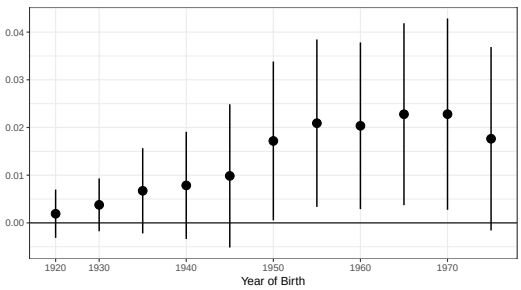
(c) Primary Completion



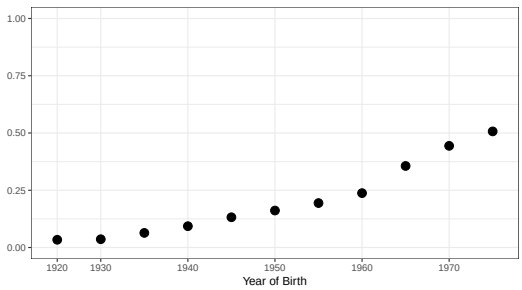
(d) Primary Completion Levels



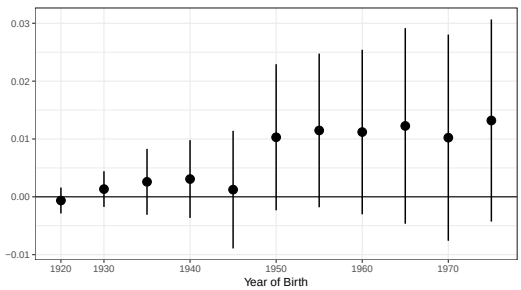
(e) Junior High Completion



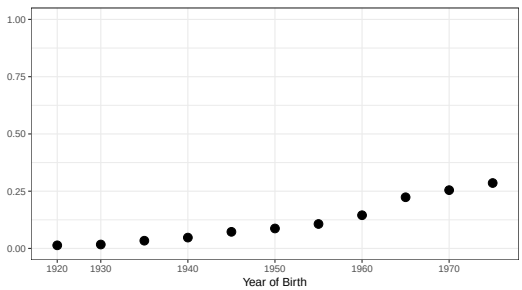
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table AB-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	1.165	0.064	0.341	0.024	-0.011	0.118	-0.029	0.440
	(0.789)	(0.028)	(0.096)	(0.179)	(0.014)	(0.145)	(0.108)	(0.148)
Obs	3,290	3,290	3,293	3,290	3,293	3,293	3,293	3,293
Clusters	330	330	330	330	330	330	330	330
Mean	41.11	0.34	28.37	2.64	0.28	24.75	35.59	33.04

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AB-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	1.117	0.010	4.204	0.020	0.127	0.529	0.210	-0.093
	(1.054)	(0.005)	(1.848)	(0.006)	(0.194)	(1.287)	(0.213)	(0.175)
Obs	3,288	3,288	3,048	2,993	3,107	3,107	3,107	3,107
Clusters	330	330	326	326	328	328	328	328
Mean	20.04	0.09	25.09	0.11	3.54	16.87	3.74	4.30

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AB-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.221	0.027	0.013	0.006	-0.009	-0.006	0.173	0.042
	(0.080)	(0.007)	(0.008)	(0.006)	(0.009)	(0.010)	(0.102)	(0.016)
Obs	11,429,344	11,429,344	11,429,344	11,429,344	478,199	478,122	19,305	19,305
Clusters	330	330	330	330	308	308	330	330
Mean	4.97	0.63	0.26	0.12	0.42	0.19	11.90	0.75

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AB-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.040	0.026	0.018	-0.052	0.018	0.026	10.730	0.011
	(0.014)	(0.010)	(0.008)	(0.020)	(0.013)	(0.010)	(8.936)	(0.012)
Obs	87,866	87,866	87,866	94,064	94,064	94,064	3,289	99,095
Clusters	323	323	323	308	308	308	330	323
Mean	0.33	0.18	0.17	0.52	0.10	0.12	69.73	12.51

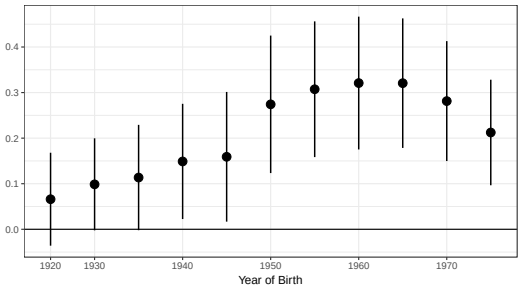
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AC Full Sample

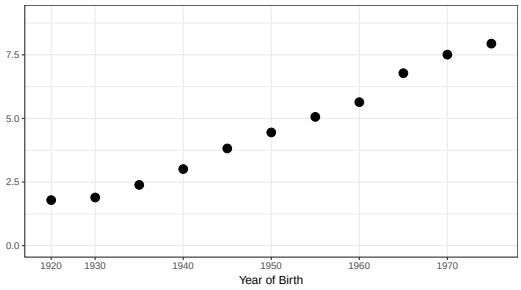
In the baseline we limit the sample to villages suitable for sugar cultivation - since non-suitable places are not a plausible counterfactual - but estimates are robust to varying the sample inclusion criteria. This appendix documents that estimates are similar when all villages are included in the sample, regardless of their sugar suitability.

Figure AC-1: Education by Cohort: Subjected Villages (2000 Census)

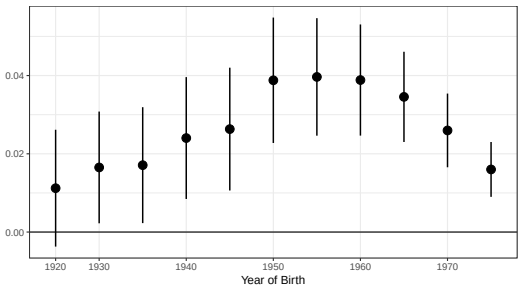
(a) Years of Schooling



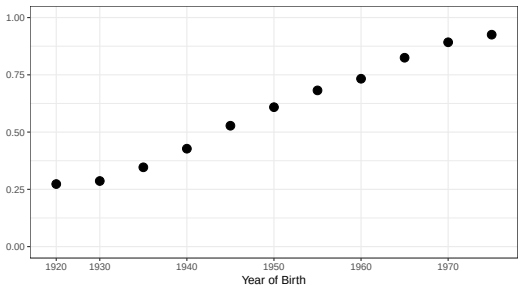
(b) Years of Schooling Levels



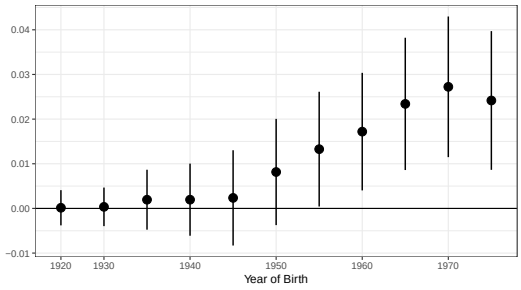
(c) Primary Completion



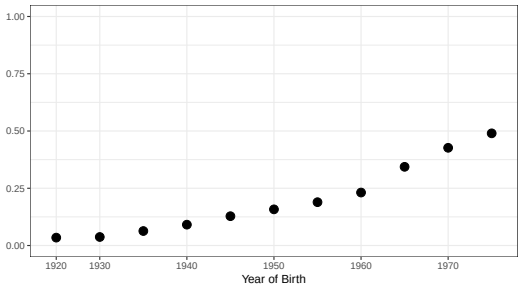
(d) Primary Completion Levels



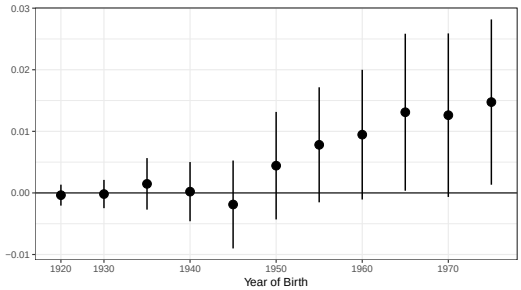
(e) Junior High Completion



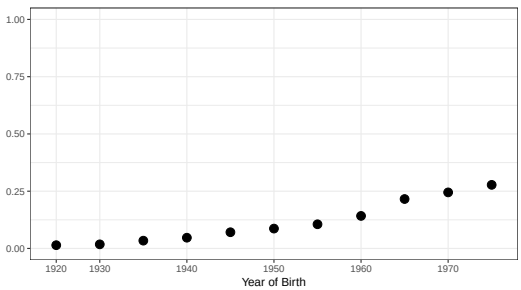
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table AC-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-10.489	-0.139	0.133	-0.018	-0.004	0.004	-0.110	0.298
	(3.068)	(0.051)	(0.080)	(0.121)	(0.009)	(0.108)	(0.095)	(0.107)
Obs	6,898	6,898	6,906	6,893	6,906	6,906	6,906	6,906
Clusters	524	524	524	524	524	524	524	524
Mean	114.02	1.11	26.38	2.56	0.26	25.23	30.71	32.91

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AC-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	3.634	0.015	3.162	0.015	0.179	0.622	0.152	-0.005
	(1.359)	(0.003)	(1.260)	(0.004)	(0.122)	(0.803)	(0.135)	(0.097)
Obs	6,891	6,891	6,360	6,252	6,495	6,495	6,495	6,495
Clusters	524	524	523	523	523	523	523	523
Mean	20.07	0.08	22.78	0.10	3.39	16.28	3.62	4.29

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AC-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.186	0.021	0.013	0.006	-0.012	-0.006	0.134	0.037
	(0.061)	(0.005)	(0.006)	(0.004)	(0.007)	(0.008)	(0.072)	(0.011)
Obs	23,388,214	23,388,214	23,388,214	23,388,214	974,723	974,535	40,345	40,345
Clusters	524	524	524	524	508	508	524	524
Mean	4.92	0.63	0.25	0.12	0.41	0.19	11.70	0.71

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AC-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.037	0.030	0.010	-0.026	0.007	0.018	22.944	0.008
	(0.010)	(0.008)	(0.006)	(0.017)	(0.010)	(0.008)	(6.754)	(0.009)
Obs	189,152	189,152	189,152	193,946	193,946	193,946	6,891	211,544
Clusters	519	519	519	508	508	508	524	519
Mean	0.32	0.18	0.16	0.53	0.10	0.11	75.91	12.50

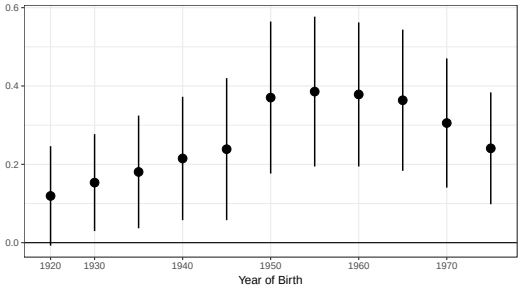
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AD No Triangular Kernel Weighting

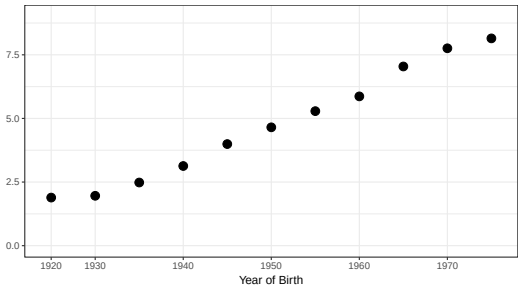
The baseline specification uses a triangular kernel such that the weight given to each observation decays with distance from the threshold. This appendix shows that estimates change little when no weighting is used.

Figure AD-1: Education by Cohort: Subjected Villages (2000 Census)

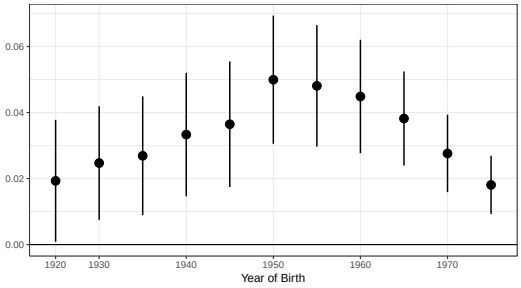
(a) Years of Schooling



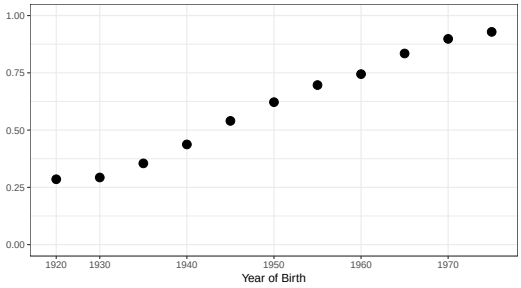
(b) Years of Schooling Levels



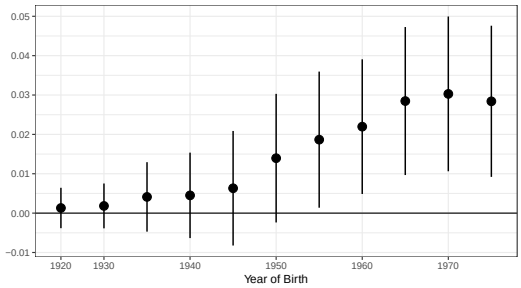
(c) Primary Completion



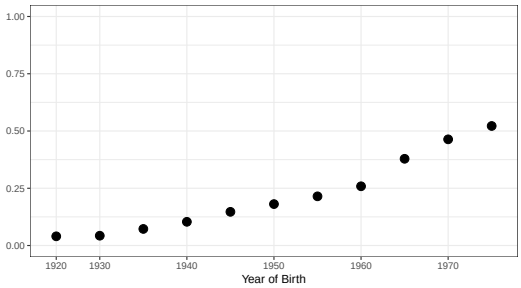
(d) Primary Completion Levels



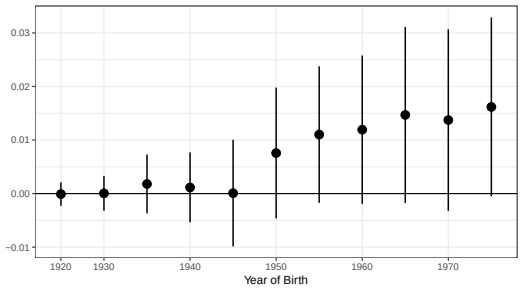
(e) Junior High Completion



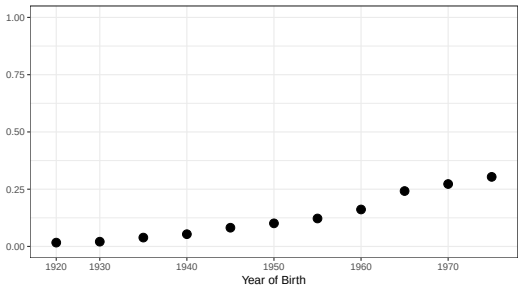
(f) Junior High Completion Levels



(g) High School Completion



(h) High School Completion Levels



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table AD-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.709	-0.024	0.142	-0.079	-0.000	0.045	-0.100	0.149
	(0.851)	(0.015)	(0.126)	(0.157)	(0.012)	(0.161)	(0.145)	(0.158)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AD-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	2.141	0.017	3.813	0.018	0.033	1.163	0.248	0.015
	(0.851)	(0.004)	(1.583)	(0.006)	(0.145)	(1.053)	(0.174)	(0.141)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AD-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.219	0.025	0.015	0.006	-0.022	-0.001	0.076	0.025
	(0.077)	(0.007)	(0.008)	(0.006)	(0.008)	(0.010)	(0.085)	(0.013)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AD-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.051	0.031	0.017	-0.042	0.011	0.029	21.833	0.005
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)	(8.429)	(0.012)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

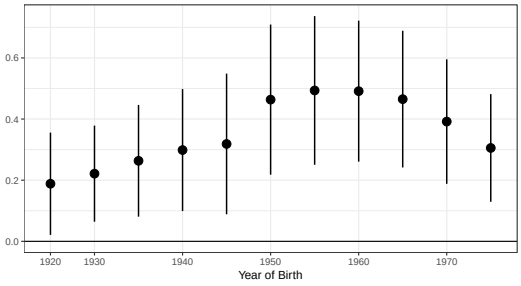
Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AE Instrumental Variables

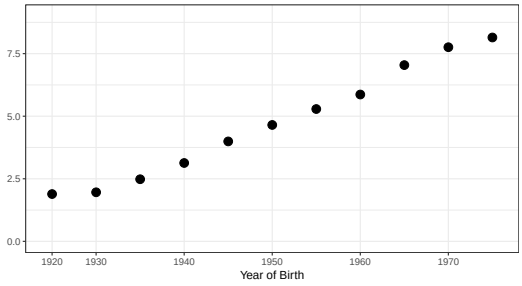
Recall that not all villages within the constructed catchment zones can be matched to historical villages, since some villages have disappeared or changed their names. This appendix designates matched villages as treated and instruments these with being inside the catchment areas. The first stage F-stat is over 1300, and estimates remain quite similar to the baseline.

Figure AE-1: Education by Cohort: Subjected Villages (2000 Census)

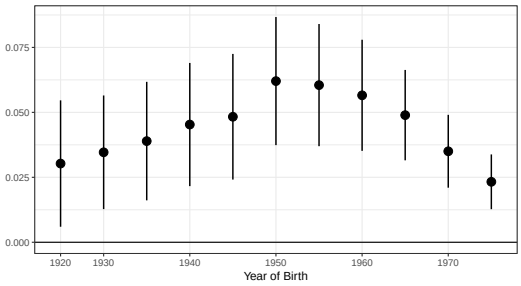
(a) Years of Schooling



(b) Years of Schooling Levels

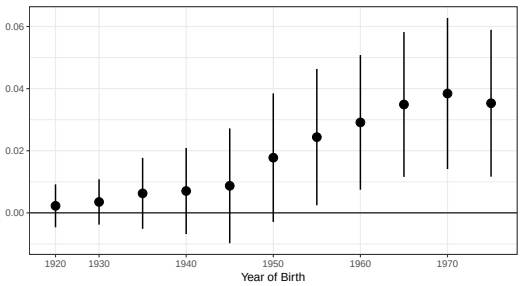


(c) Primary Completion

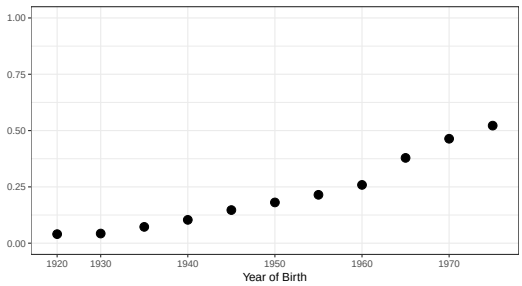


(d) Primary Completion Levels

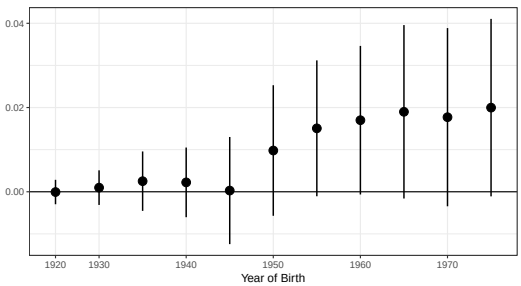
(e) Junior High Completion



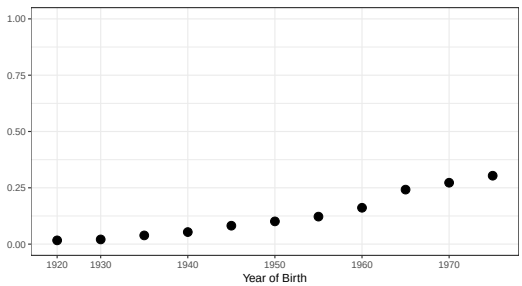
(f) Junior High Completion Levels



(g) High School Completion



(h) *High School Completion Levels*



Notes: In the left panels, each point plots a separate regression coefficient for different birth cohorts (1920-1929, 1930-1934, 1935-1939, ..., 1975-1979). Lines show 90% confidence intervals. In the right panels, points plot means. The unit of analysis is the individual, and the specification includes gender dummies, geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table AE-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance 1830 Residency Capital	Distance To Great Post Road	Distance To To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.517	-0.022	0.154	-0.054	0.001	0.082	-0.080	0.349
	(0.939)	(0.017)	(0.135)	(0.203)	(0.015)	(0.169)	(0.150)	(0.173)
Obs	4,553	4,553	4,553	4,549	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
F stat	1463	1463	1463	1476	1463	1463	1463	1463
Mean	31.17	0.26	24.90	2.56	0.29	24.80	25.16	30.23

Notes: The unit of observation is the village. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AE-2: Land Tenure: Subjected Villages

	Village Land 2003		Village Land 1980		99th Pctile ÷	90th Pctile ÷	90th Pctile ÷	50th Pctile ÷
	Total Land	Land Share	Total Land	Land Share	90th Pctile	10th Pctile	50th Pctile	10th Pctile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	3.095	0.019	4.744	0.017	0.042	1.133	0.284	-0.060
	(1.069)	(0.005)	(1.941)	(0.007)	(0.189)	(1.431)	(0.243)	(0.178)
Obs	4,550	4,550	4,205	4,107	4,202	4,202	4,202	4,202
Clusters	383	383	380	380	381	381	381	381
F stat	1432	1432	1269	1404	1439	1439	1439	1439
Mean	18.61	0.09	23.95	0.11	3.60	17.00	3.78	4.27

Notes: The unit of observation is the village. Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AE-3: Education: Subjected Villages

	2000 Population Census				1980 Census		Village Head	
	Years Education	Primary School	Junior High	High School	No School	Primary School	Years Education	High School
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.287	0.033	0.020	0.009	-0.023	-0.005	0.125	0.037
	(0.095)	(0.008)	(0.009)	(0.007)	(0.011)	(0.013)	(0.113)	(0.017)
Obs	16,125,747	16,125,747	16,125,747	16,125,747	653,313	653,188	26,630	26,630
Clusters	383	383	383	383	358	358	383	383
F stat	1340	1340	1340	1340	739	739	1580	1580
Mean	5.10	0.64	0.27	0.13	0.41	0.19	11.87	0.74

Notes: The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, geographic and pre-period characteristics, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender dummies, and columns (7) and (8) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

Table AE-4: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census			Num. Manuf.	Log Equiv.
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce	Firms	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-0.055	0.041	0.016	-0.047	0.010	0.036	28.769	0.007
	(0.016)	(0.013)	(0.010)	(0.027)	(0.017)	(0.013)	(10.869)	(0.015)
Obs	130,335	130,335	130,335	127,873	127,873	127,873	4,549	144,046
Clusters	381	381	381	358	358	358	383	381
F stat	916	916	916	716	716	716	1429	1033
Mean	0.27	0.21	0.18	0.48	0.11	0.12	71.72	12.55

Notes: The unit of observation is the individual in columns (1) through (6), the village in column (7), and the household in column (8). The sample is restricted to men age 18-55 in columns (1) through (6). Regressions include geographic and pre-period characteristics, boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (3) include survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

AF Grid Cells

Table AF-1: Geographic Characteristics

	Elevation	Slope	Distance To Coast	Log Flow Accumulation	Distance To River	Distance To Residency Capital	Distance To Great Post Road	Distance To VOC Port
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-4.443	-0.004	0.310	-0.182	-0.079	-0.056	-0.359	-0.004
	(2.932)	(0.052)	(0.566)	(0.362)	(0.090)	(0.699)	(0.633)	(0.780)
Obs	526	526	526	526	526	526	526	526
Clusters	70	70	70	70	70	70	70	70
Mean	31.49	0.28	24.36	5.55	0.27	27.24	24.43	30.01

Notes: Observations are centroids of a 5km x 5km grid. Regressions include boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by cells of a 25km x 25km grid, are in parentheses.