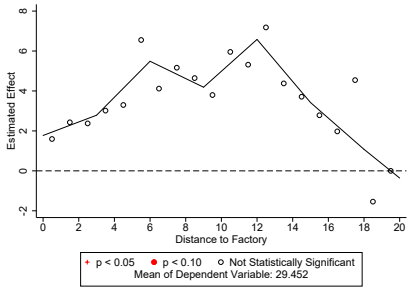
Contents

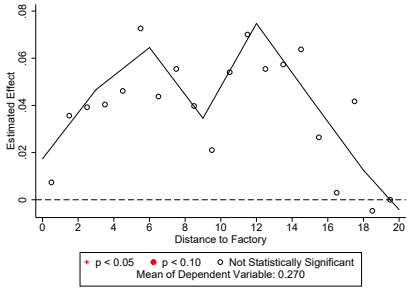
A Common Shifts Placebos

Figure A-1: Geography

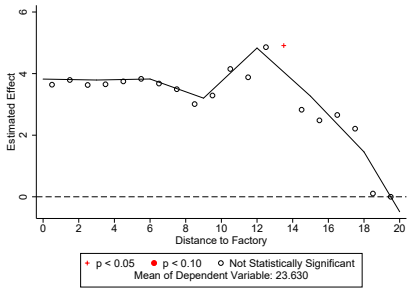
(a) Elevation



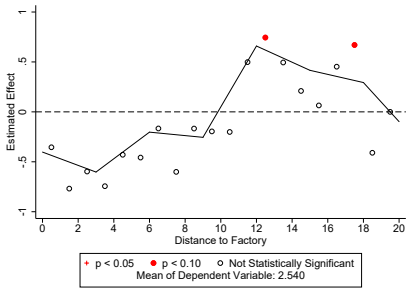
(b) Slope



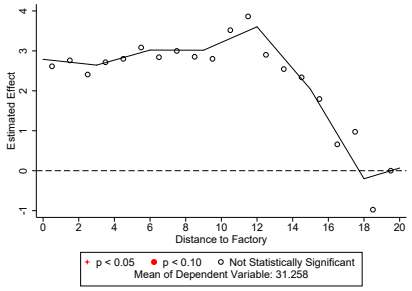
(c) Distance to Coast



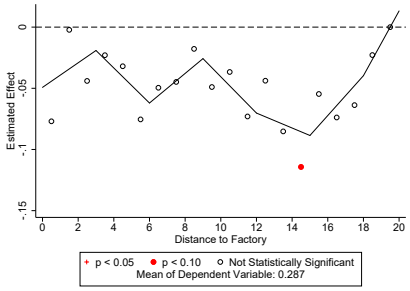
(d) Log Flow Accumulation



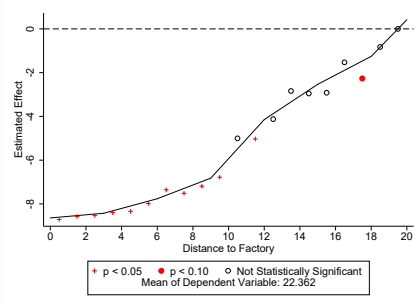
(e) Distance to Natural Harbor



(f) Distance to River



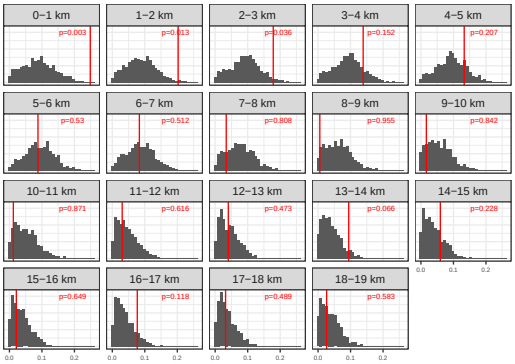
(g) Distance to Nearest 1830 Residency Capital



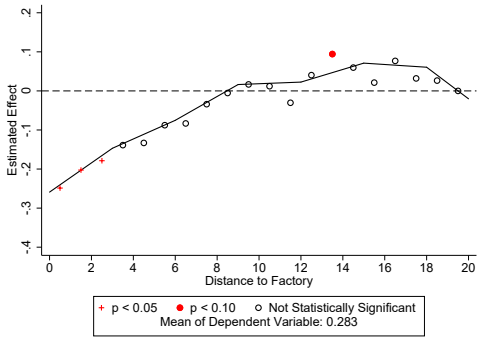
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

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

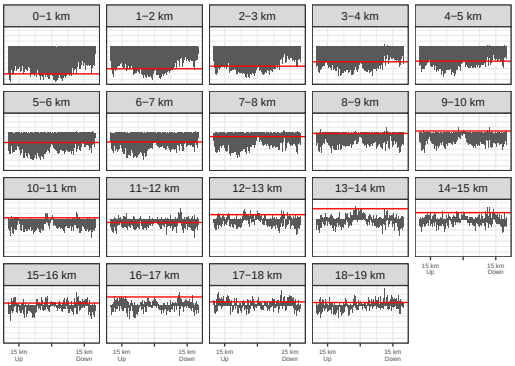
(a) Independent Shifts: Counterfactuals



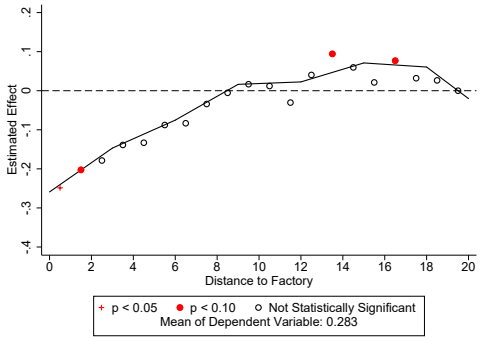
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



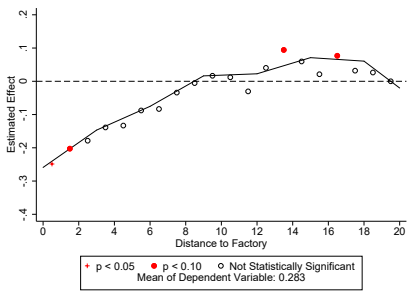
(d) Common Shifts: Plotted Coefficients



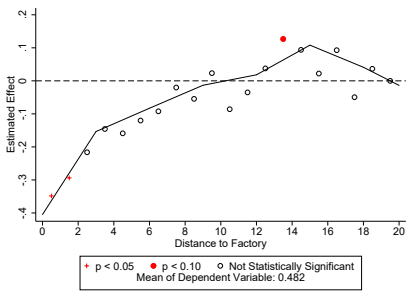
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure A-3: Industry and Agglomeration

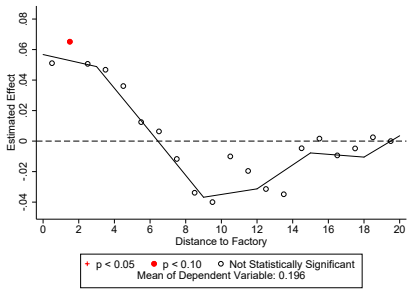
(a) Agriculture (Susenas 2001-11)



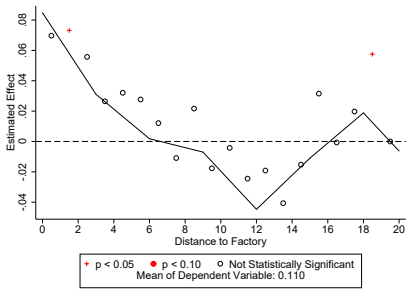
(b) Agriculture (Census 1980)



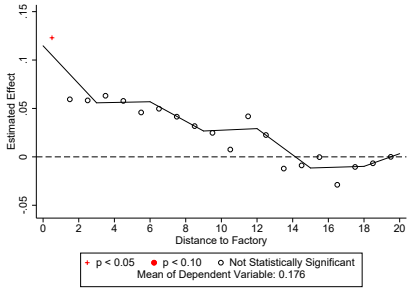
(c) Manufacturing (Susenas 2001-11)



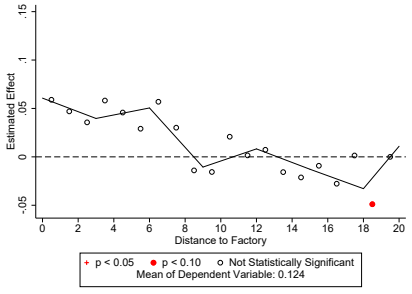
(d) Manufacturing (Census 1980)



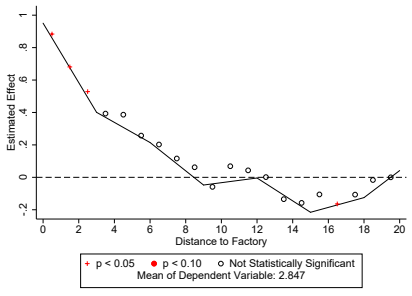
(e) Retail (Susenas 2001-11)



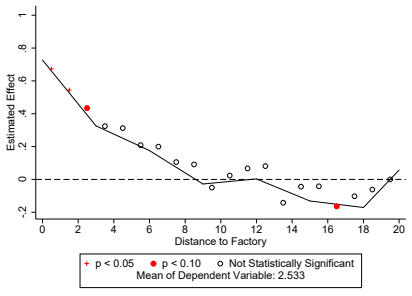
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



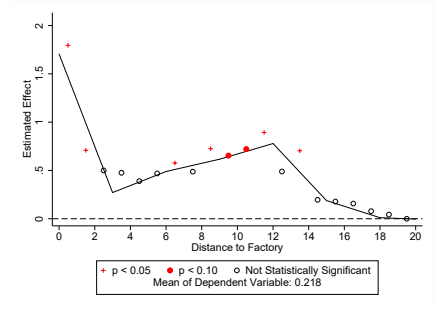
(h) Log Pop. Density (PODES 1980)



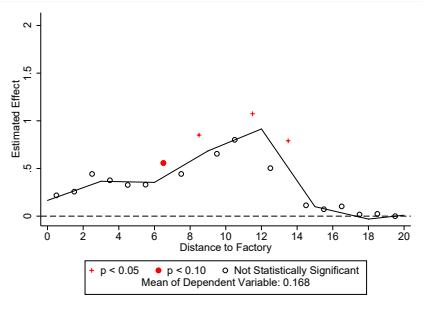
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

Figure A-4: Sugar and Linked Industries

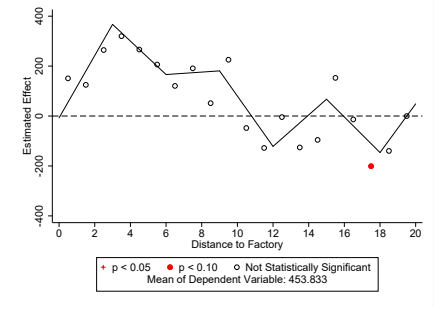
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



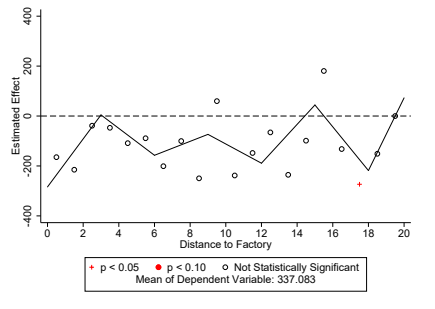
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



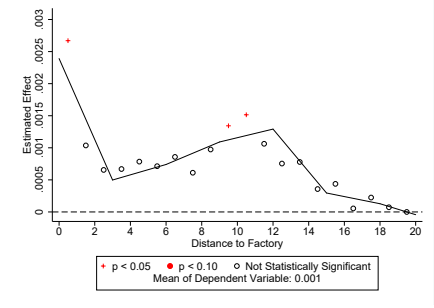
(c) Tons of Cane Grown (Full Sample, PODES 2003)



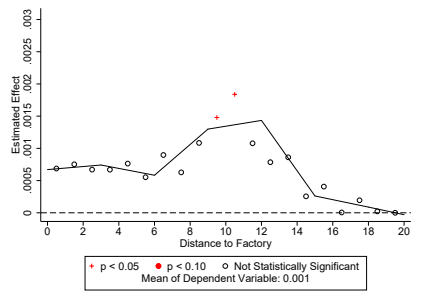
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



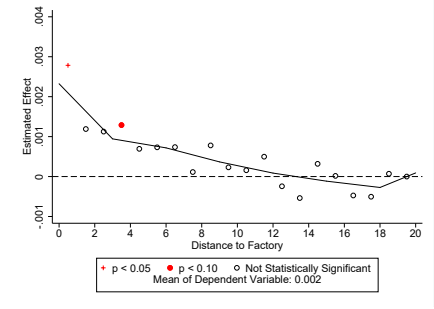
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



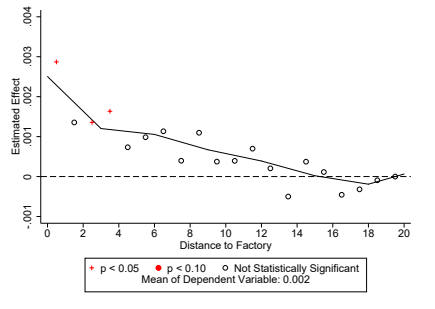
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



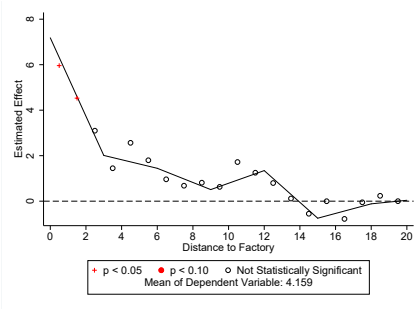
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



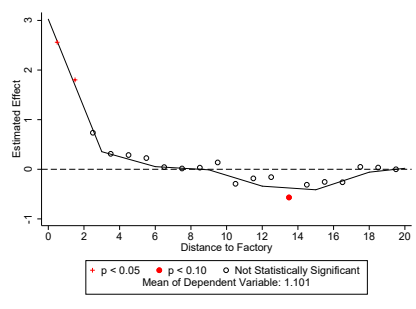
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

Figure A-5: Infrastructure

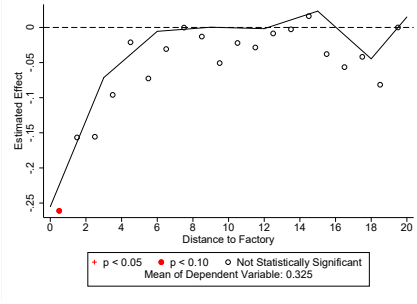
(a) Colonial Road Density (1900)



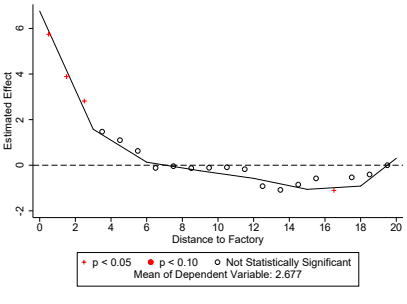
(b) Colonial Railroad Density (1900)



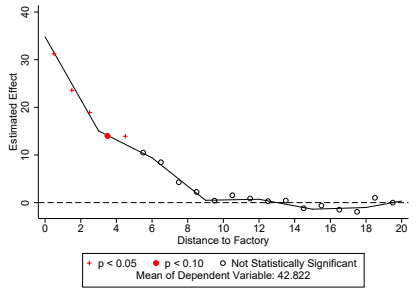
(c) Dirt Road (PODES 1980)



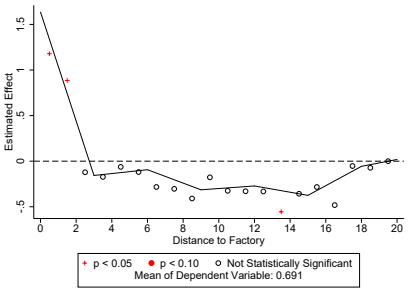
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



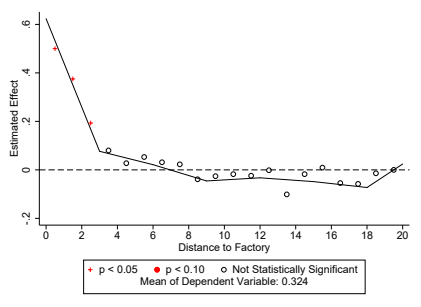
(f) Railroad Density (2017)



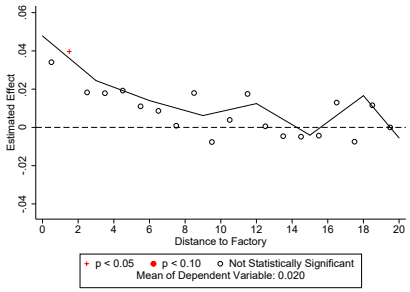
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

Figure A-6: Other Public Goods

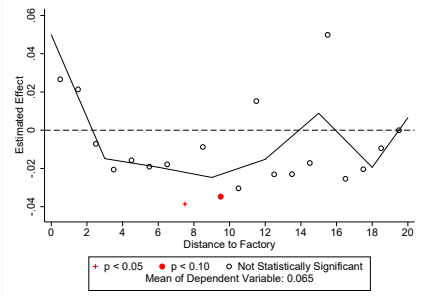
(a) Village Has Electricity (PODES 1980)



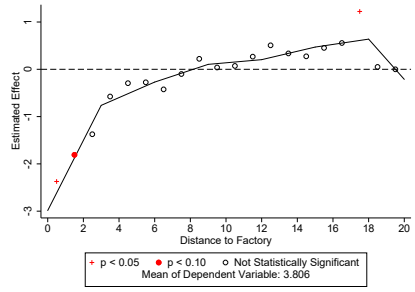
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



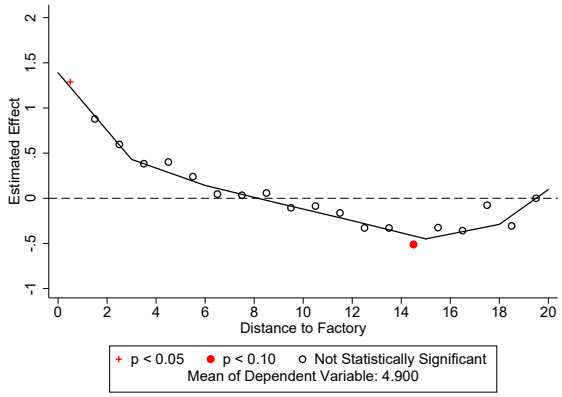
(d) Distance to Subdistrict Capital (2011 PODES)



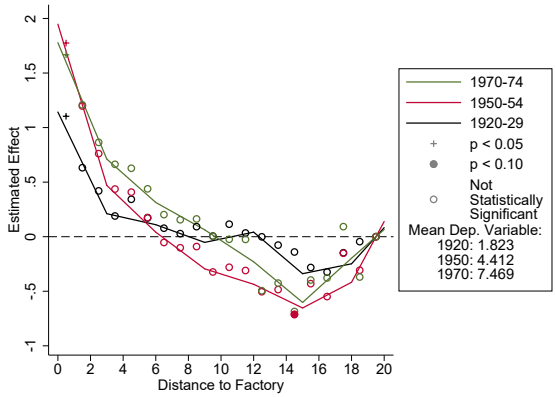
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

Figure A-7: Education

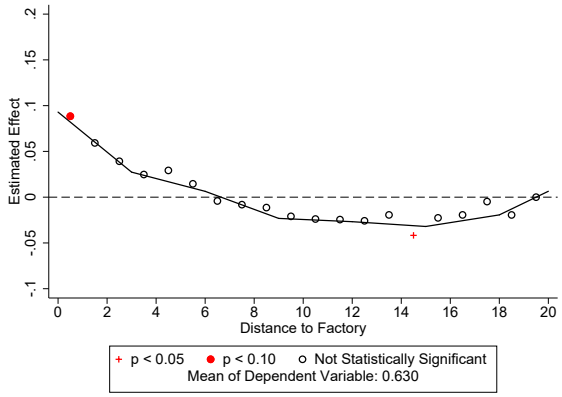
(a) Years Education (2000 Census)



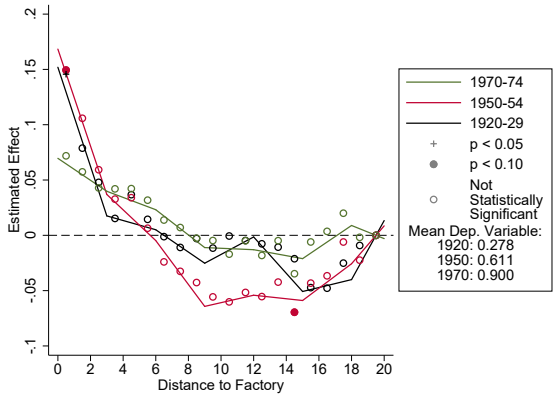
(b) Years Education by Cohort



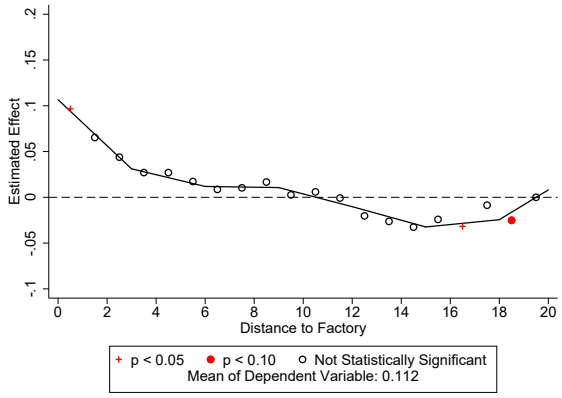
(c) Primary (2000 Census)



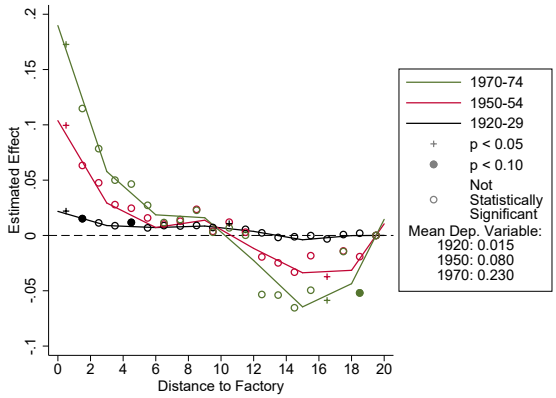
(d) Primary by Cohort



(e) High School (2000 Census)

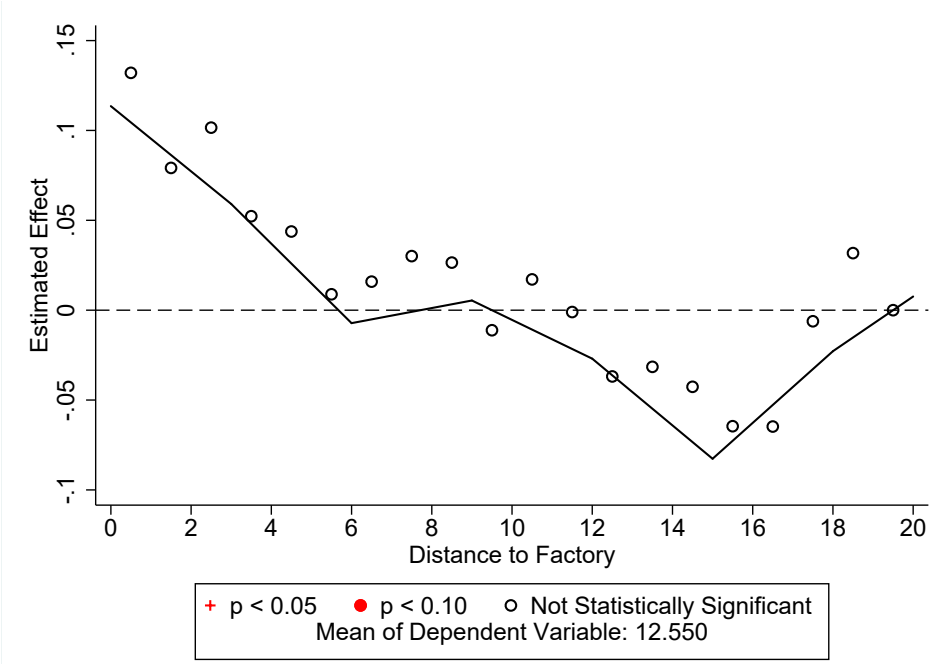


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

Figure A-8: Expenditure (2001-11)

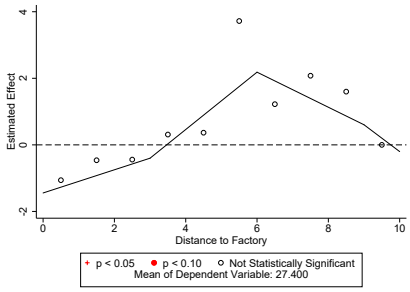


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 200 counterfactual factories located 10-20 km via river from the real factory.

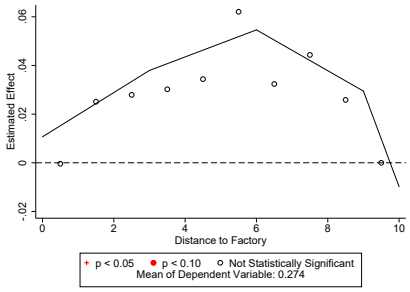
B 10 Kilometer Sample

Figure B-1: Geography

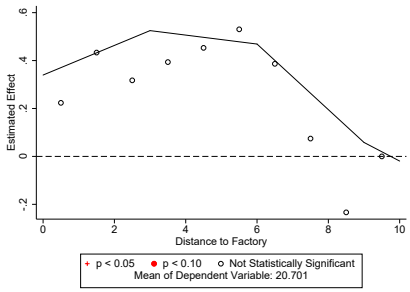
(a) Elevation



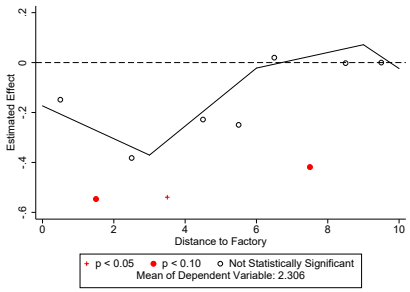
(b) Slope



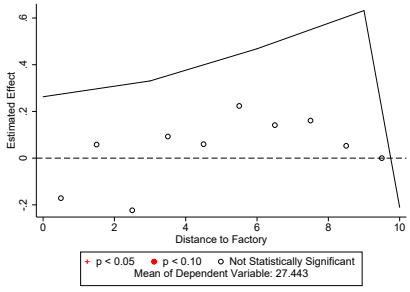
(c) Distance to Coast



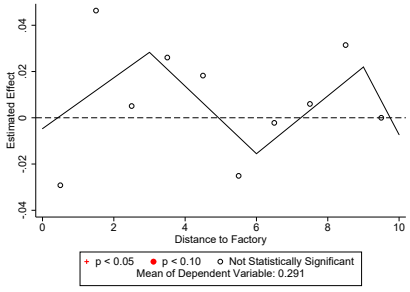
(d) Log Flow Accumulation



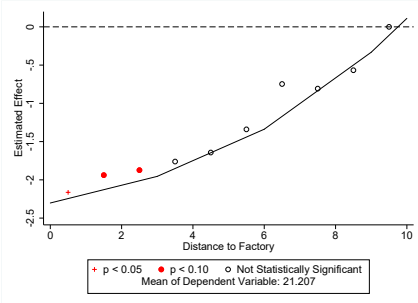
(e) Distance to Natural Harbor



(f) Distance to River



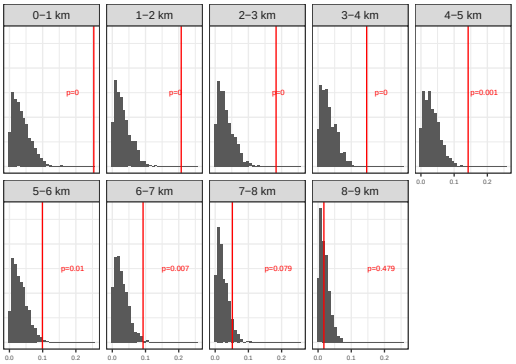
(g) Distance to Nearest 1830 Residency Capital



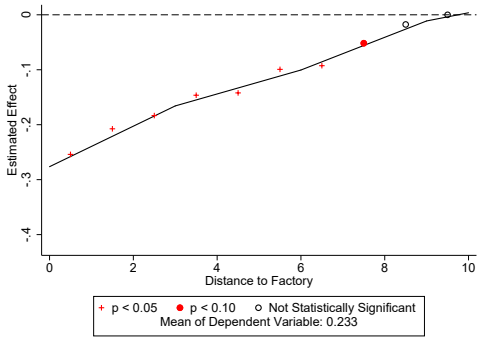
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of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

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

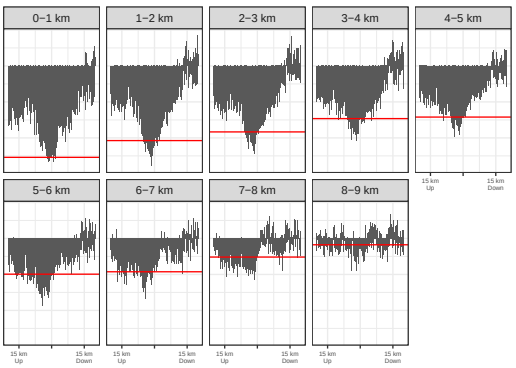
(a) Independent Shifts: Counterfactuals



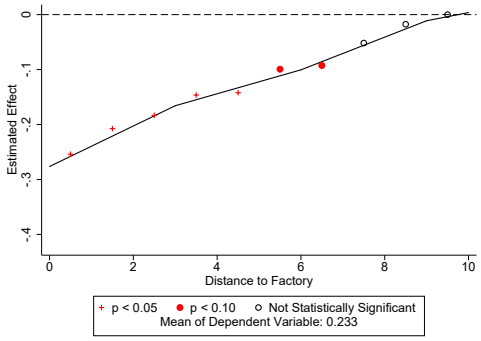
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



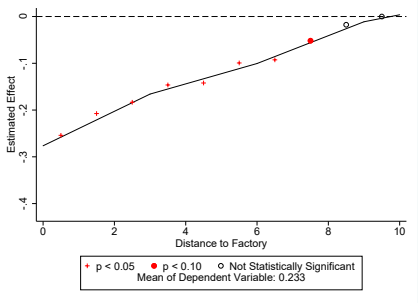
(d) Common Shifts: Plotted Coefficients



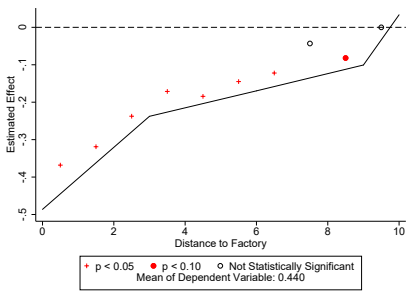
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure B-3: Industry and Agglomeration

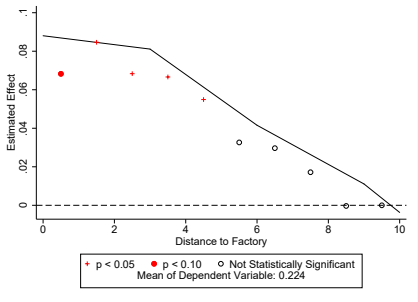
(a) Agriculture (Susenas 2001-11)



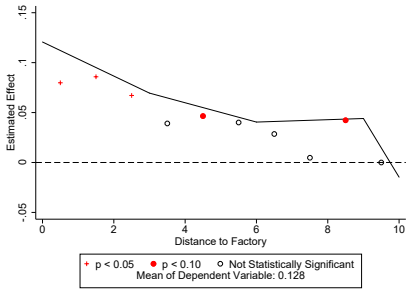
(b) Agriculture (Census 1980)



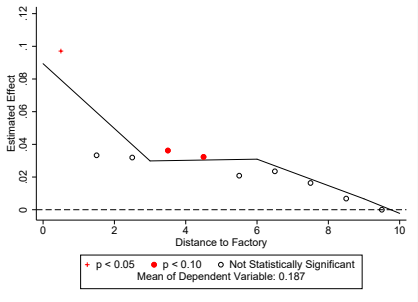
(c) Manufacturing (Susenas 2001-11)



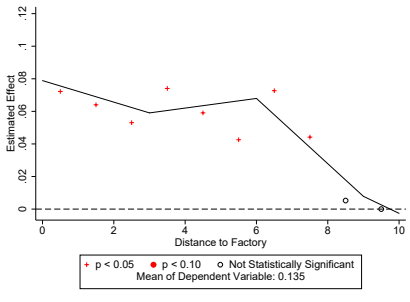
(d) Manufacturing (Census 1980)



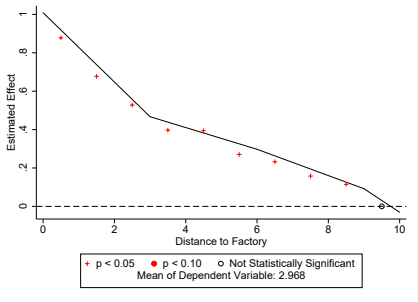
(e) Retail (Susenas 2001-11)



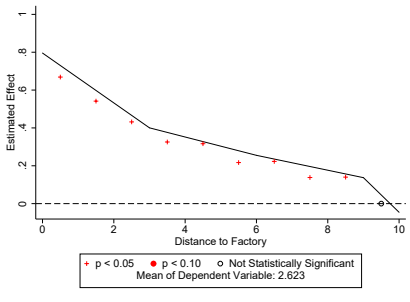
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



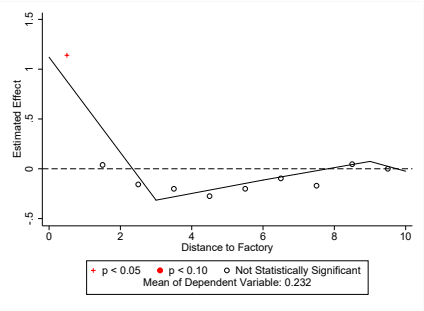
(h) Log Pop. Density (PODES 1980)



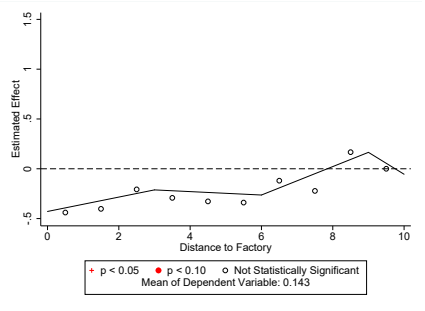
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure B-4: Sugar and Linked Industries

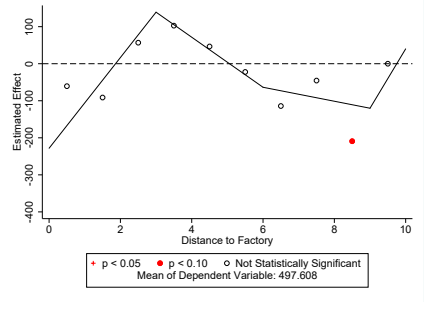
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



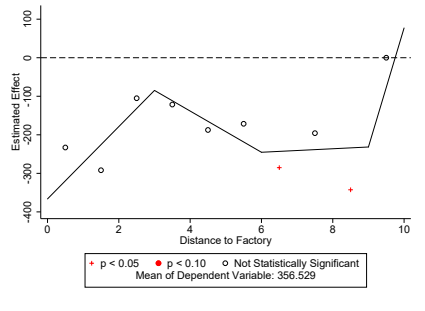
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



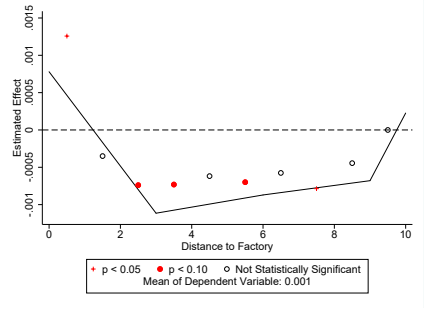
(c) Tons of Cane Grown (Full Sample, PODES 2003)



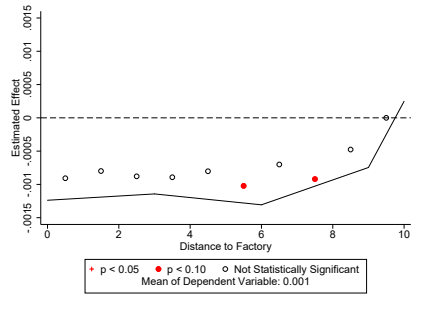
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



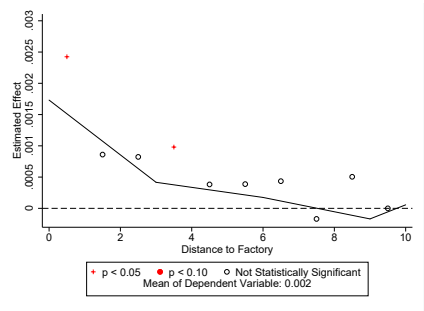
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



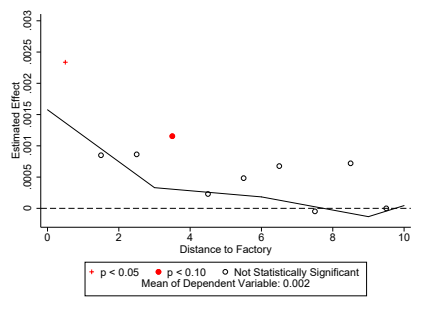
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



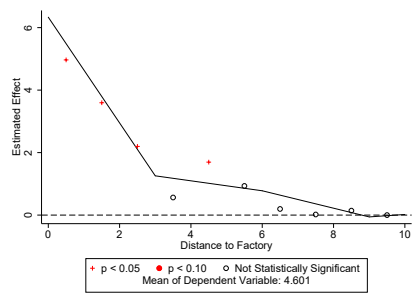
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



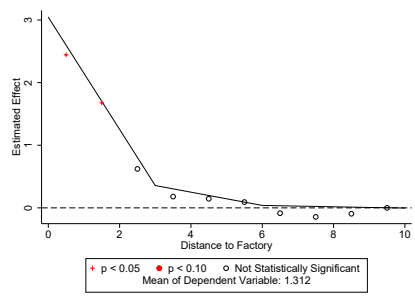
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure B-5: Infrastructure

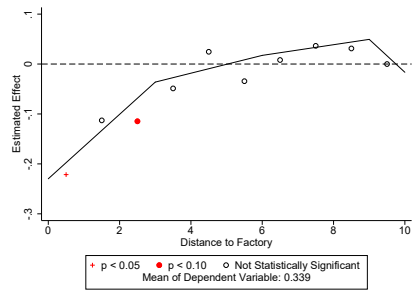
(a) Colonial Road Density (1900)



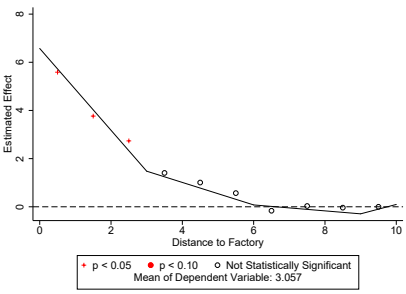
(b) Colonial Railroad Density (1900)



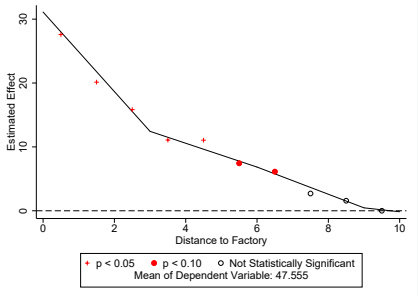
(c) Dirt Road (PODES 1980)



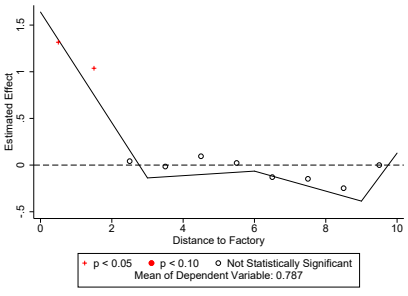
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



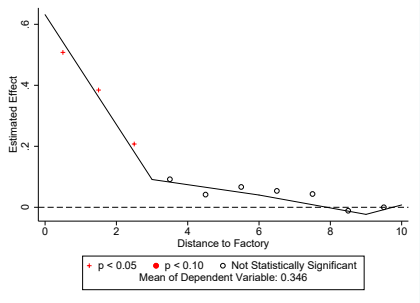
(f) Railroad Density (2017)



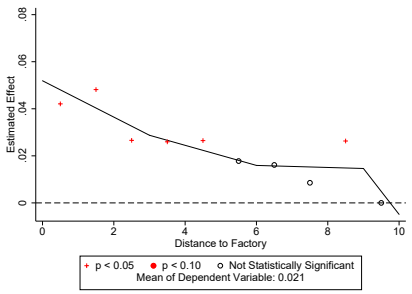
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure B-6: Other Public Goods

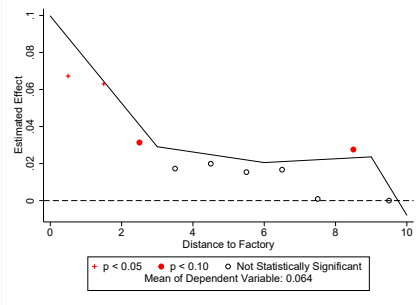
(a) Village Has Electricity (PODES 1980)



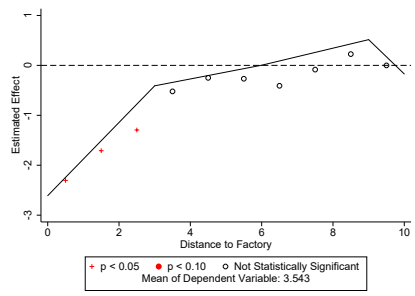
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



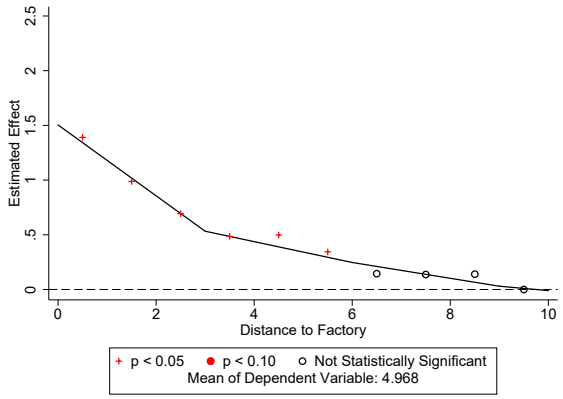
(d) Distance to Subdistrict Capital (2011 PODES)



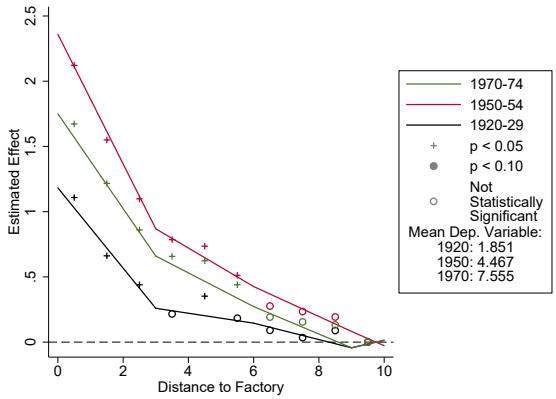
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure B-7: Education

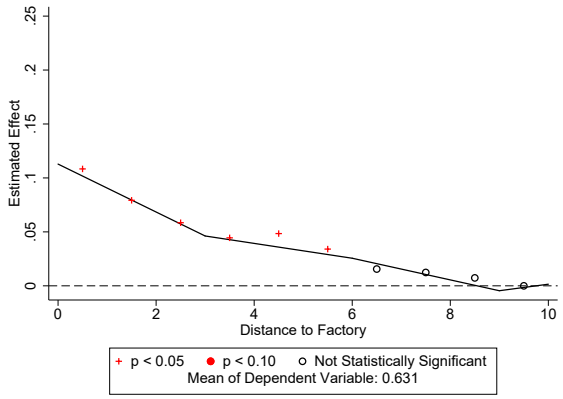
(a) Years Education (2000 Census)



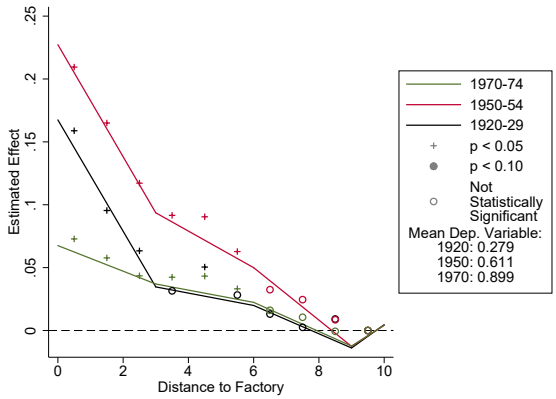
(b) Years Education by Cohort



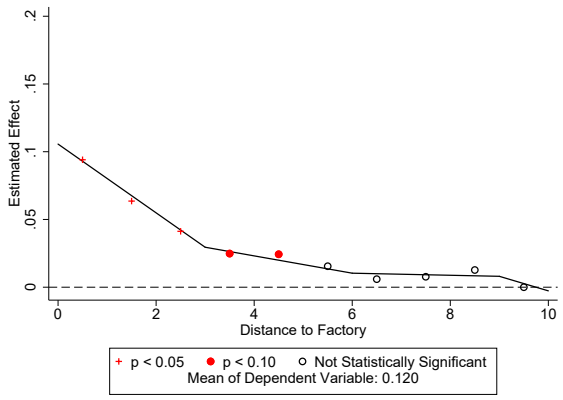
(c) Primary (2000 Census)



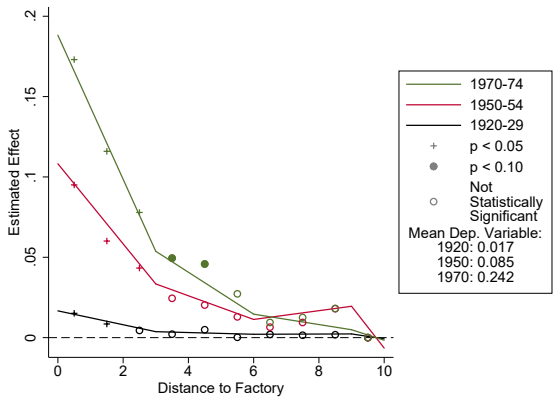
(d) Primary by Cohort



(e) High School (2000 Census)

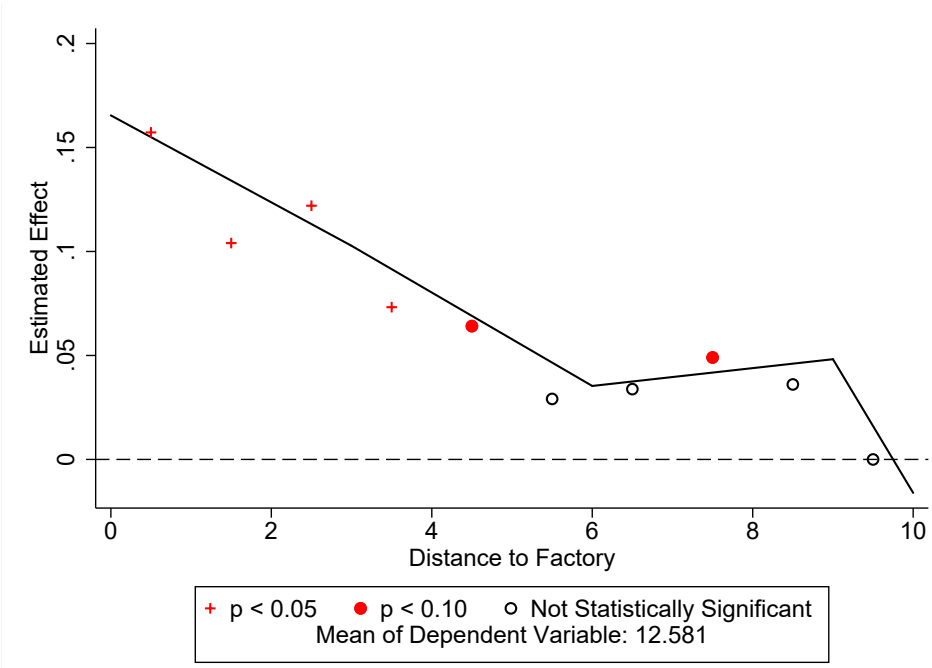


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure B-8: Expenditure (2001-11)

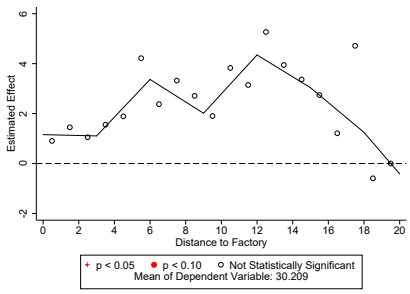


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

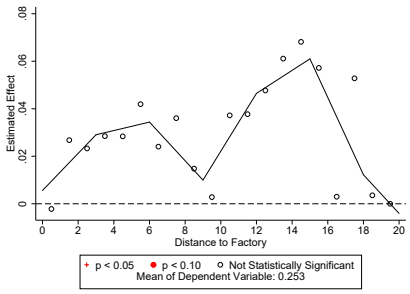
C Sample Includes Factories Near Modern Sugar Factories

Figure C-1: Geography

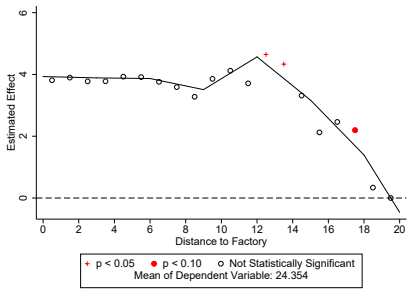
(a) Elevation



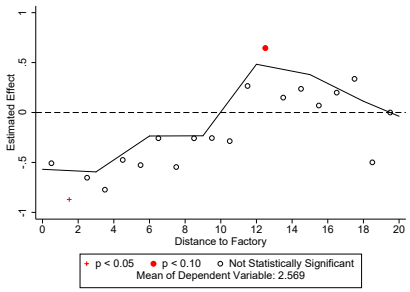
(b) Slope



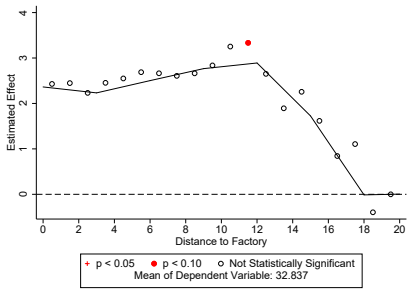
(c) Distance to Coast



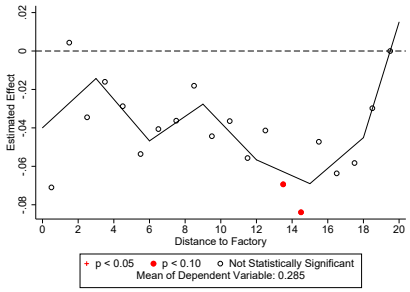
(d) Log Flow Accumulation



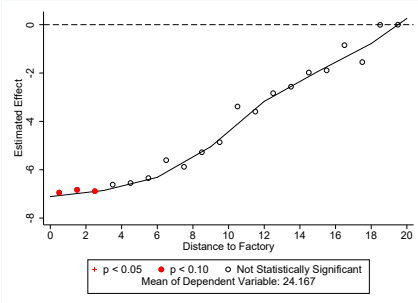
(e) Distance to Natural Harbor



(f) Distance to River



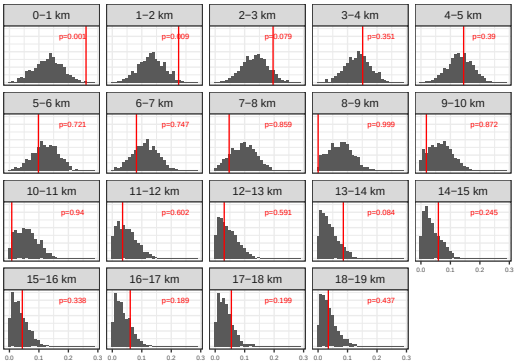
(g) Distance to Nearest 1830 Residency Capital



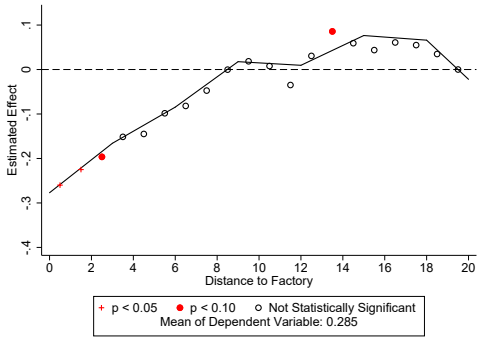
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of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

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

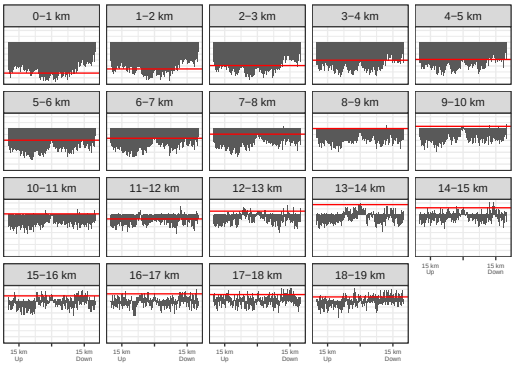
(a) Independent Shifts: Counterfactuals



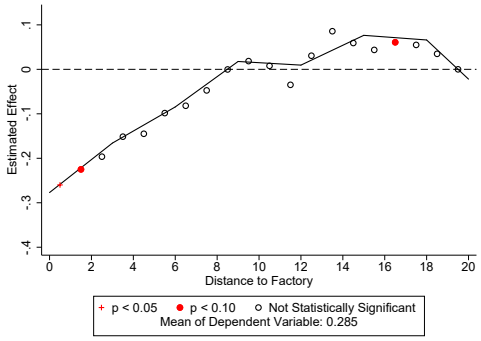
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



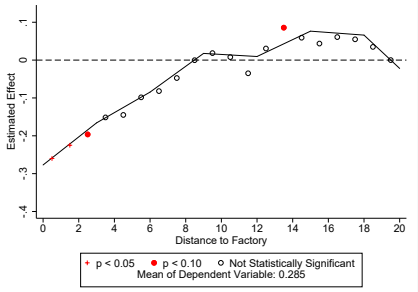
(d) Common Shifts: Plotted Coefficients



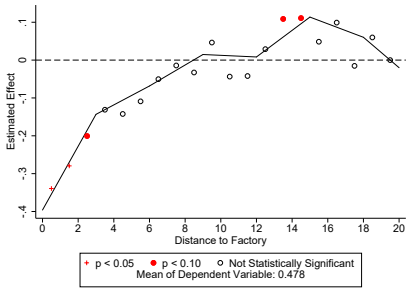
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure C-3: Industry and Agglomeration

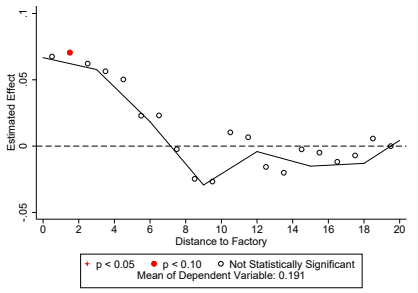
(a) Agriculture (Susenas 2001-11)



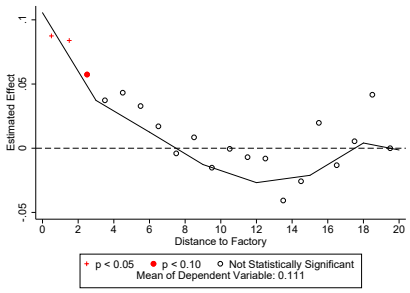
(b) Agriculture (Census 1980)



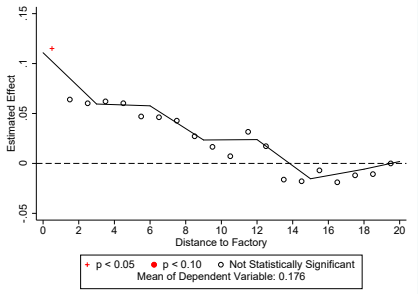
(c) Manufacturing (Susenas 2001-11)



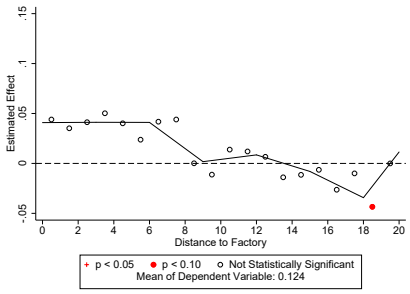
(d) Manufacturing (Census 1980)



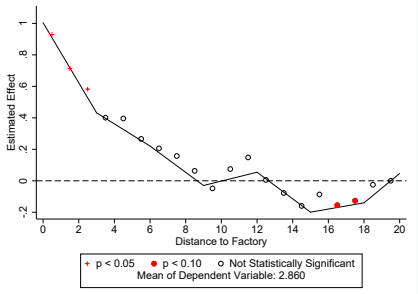
(e) Retail (Susenas 2001-11)



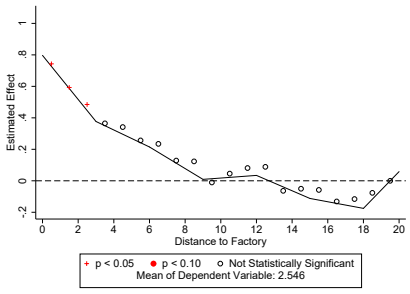
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



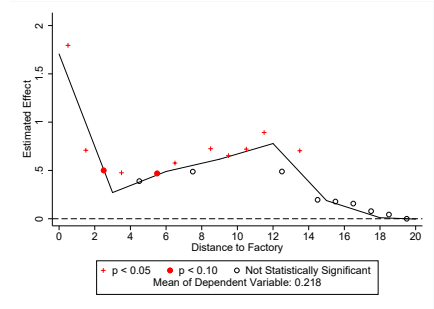
(h) Log Pop. Density (PODES 1980)



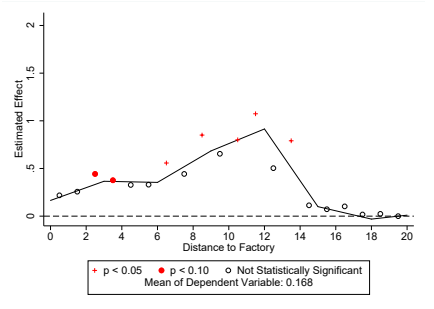
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure C-4: Sugar and Linked Industries

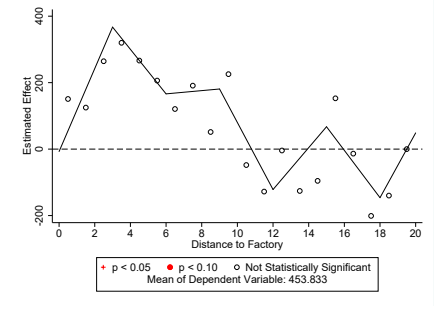
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



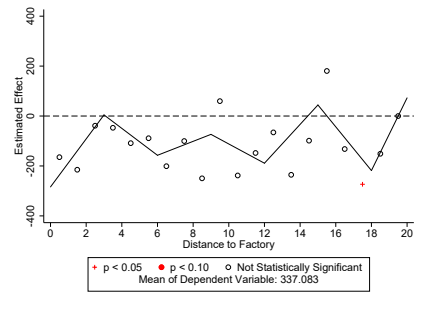
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



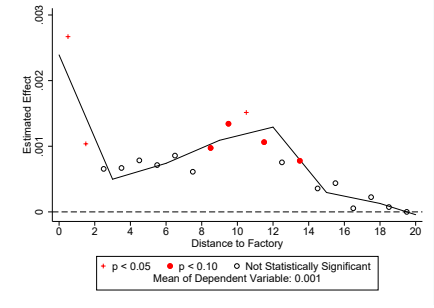
(c) Tons of Cane Grown (Full Sample, PODES 2003)



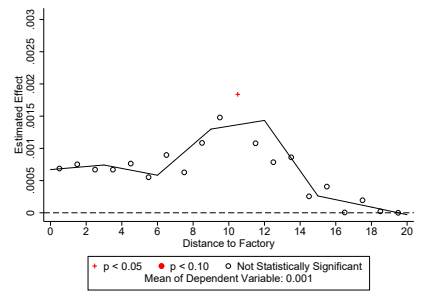
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



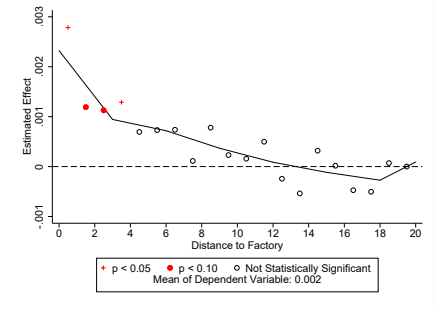
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



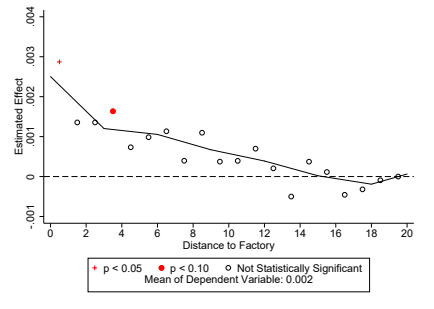
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



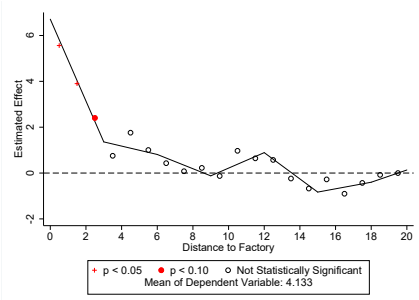
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



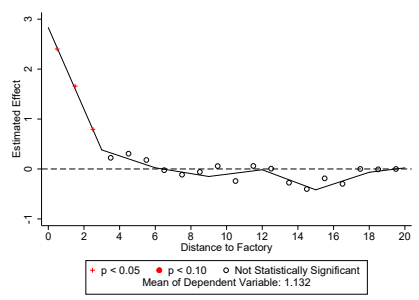
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure C-5: Infrastructure

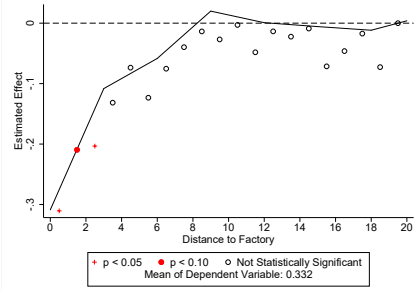
(a) Colonial Road Density (1900)



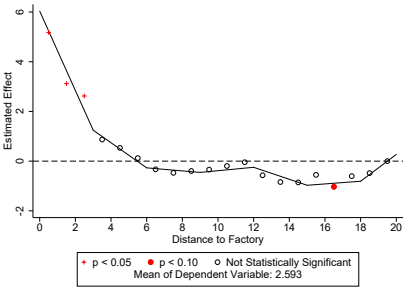
(b) Colonial Railroad Density (1900)



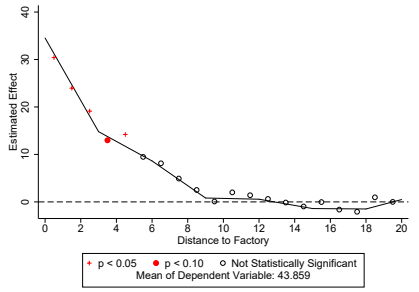
(c) Dirt Road (PODES 1980)



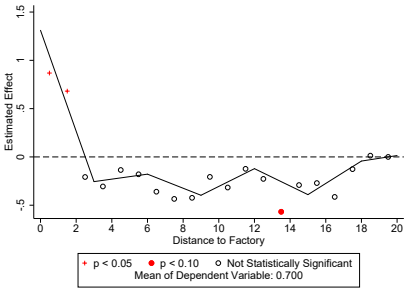
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



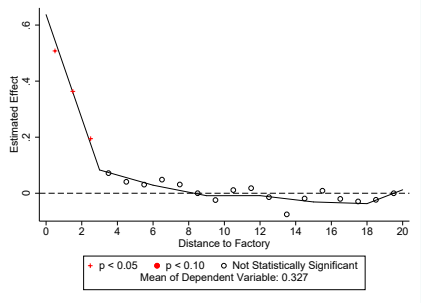
(f) Railroad Density (2017)



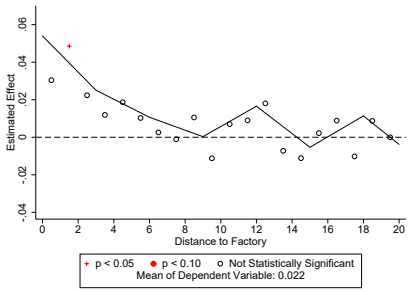
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure C-6: Other Public Goods

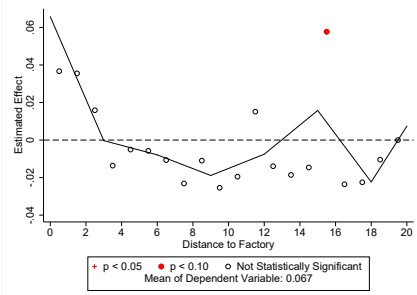
(a) Village Has Electricity (PODES 1980)



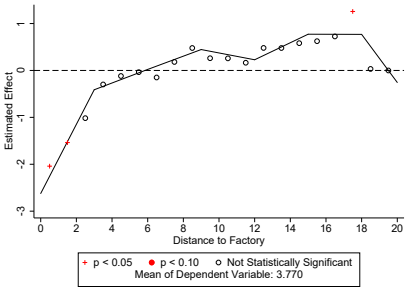
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



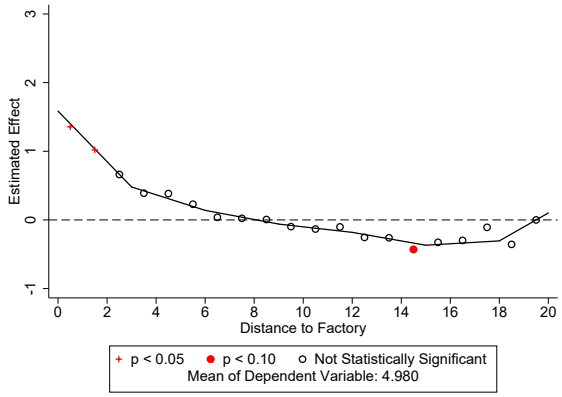
(d) Distance to Subdistrict Capital (2011 PODES)



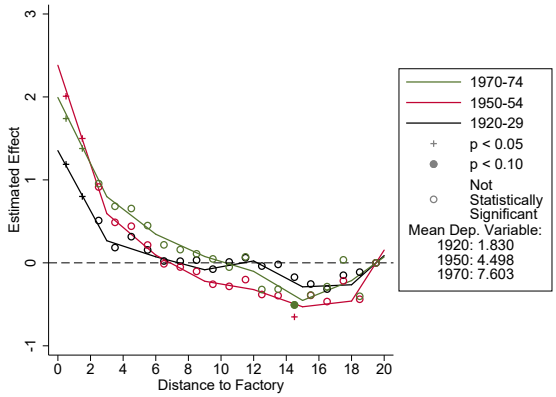
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure C-7: Education

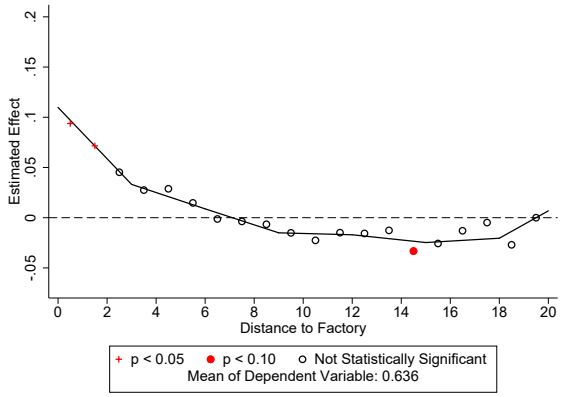
(a) Years Education (2000 Census)



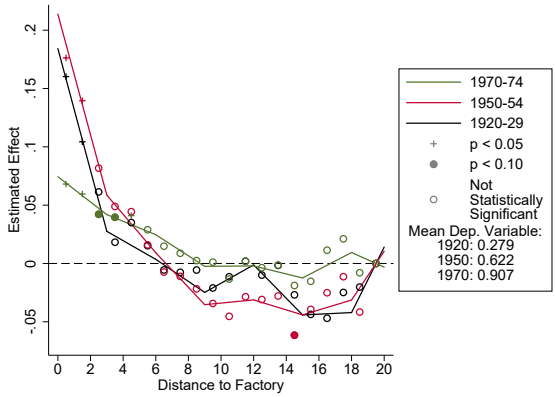
(b) Years Education by Cohort



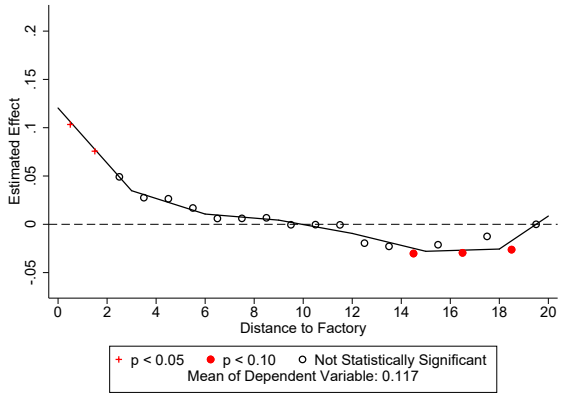
(c) Primary (2000 Census)



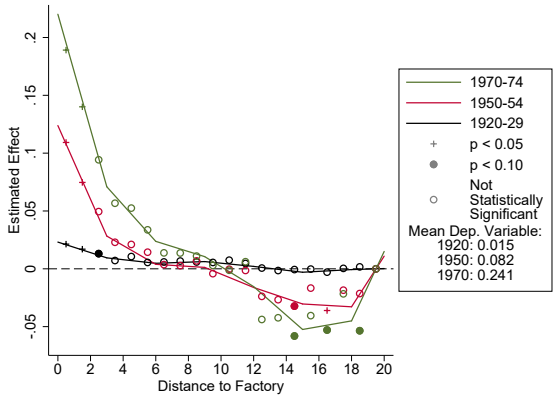
(d) Primary by Cohort



(e) High School (2000 Census)

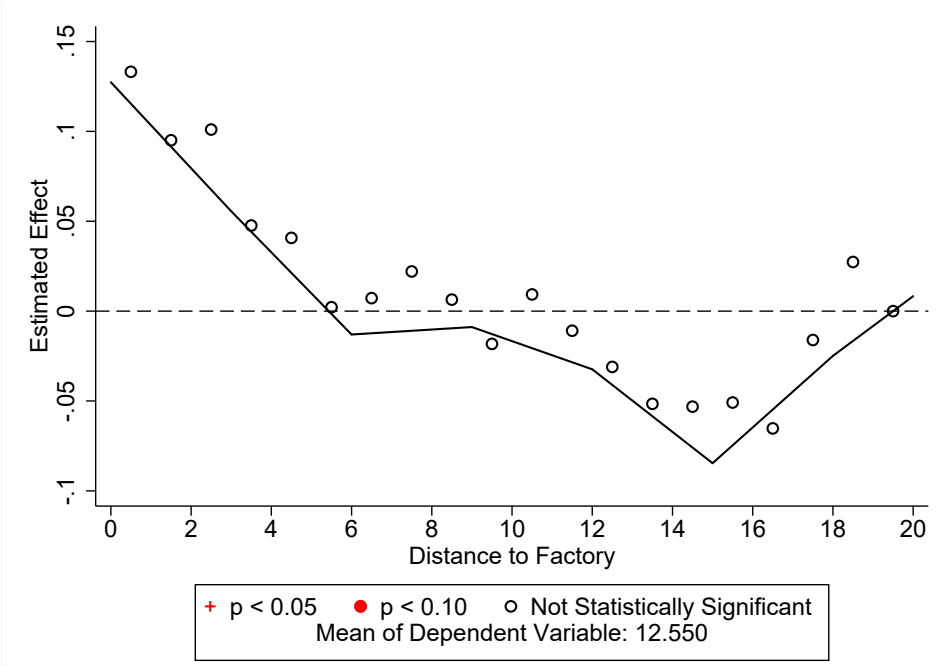


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure C-8: Expenditure (2001-11)

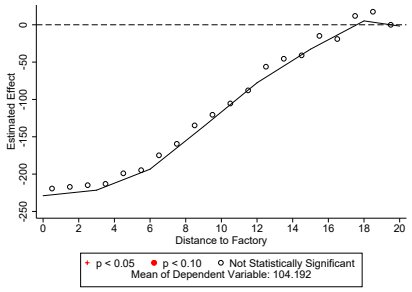


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

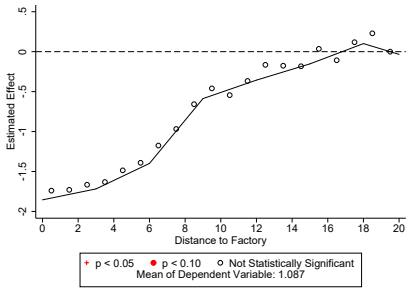
D No Village Restrictions

Figure D-1: Geography

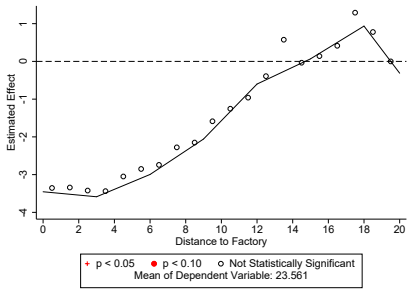
(a) Elevation



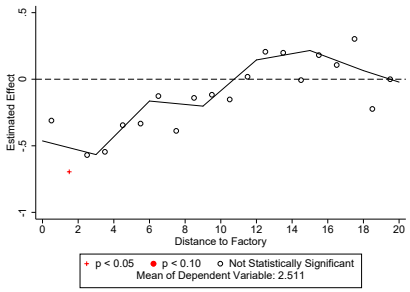
(b) Slope



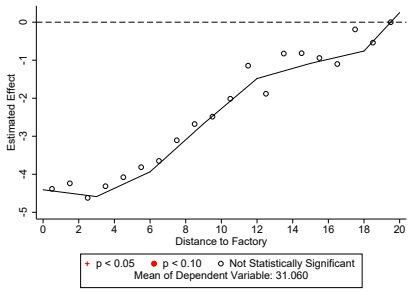
(c) Distance to Coast



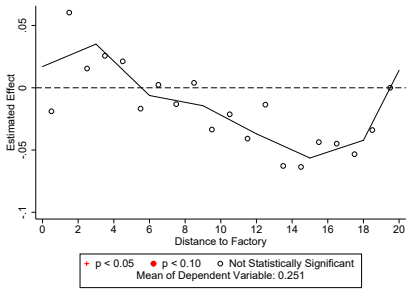
(d) Log Flow Accumulation



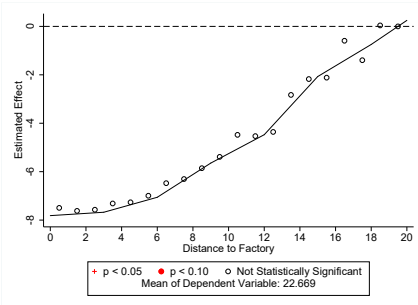
(e) Distance to Natural Harbor



(f) Distance to River



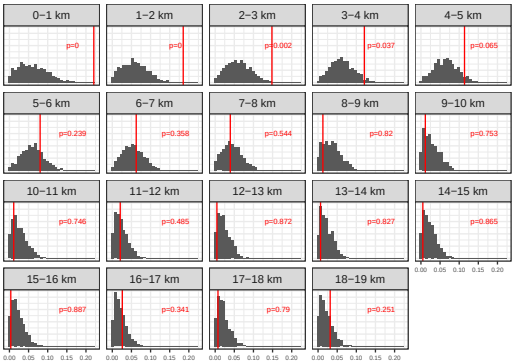
(g) Distance to Nearest 1830 Residency Capital



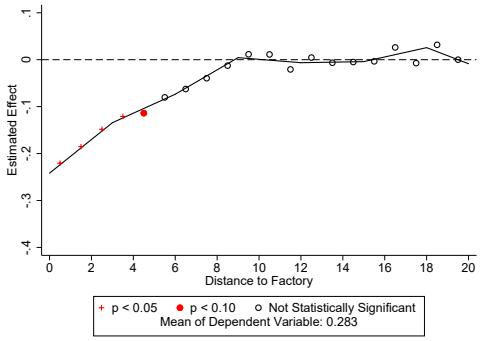
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of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

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

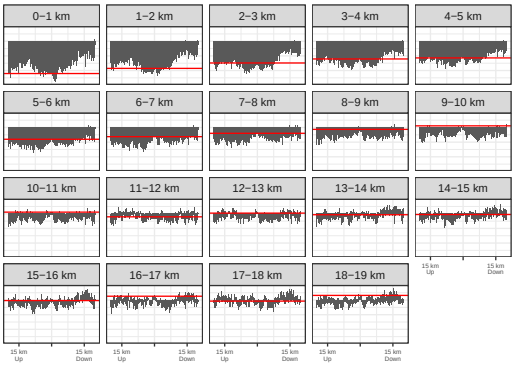
(a) Independent Shifts: Counterfactuals



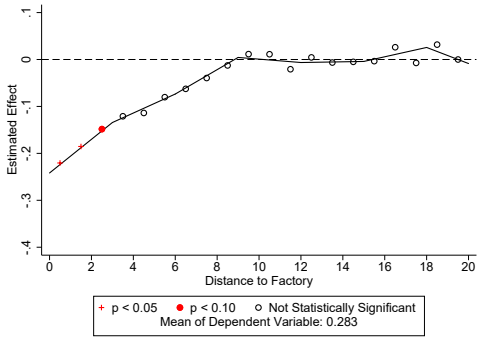
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



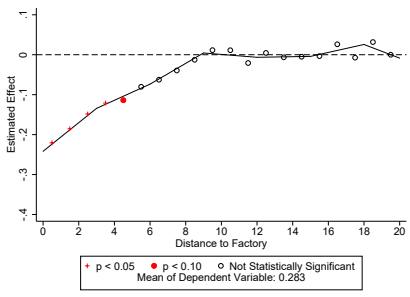
(d) Common Shifts: Plotted Coefficients



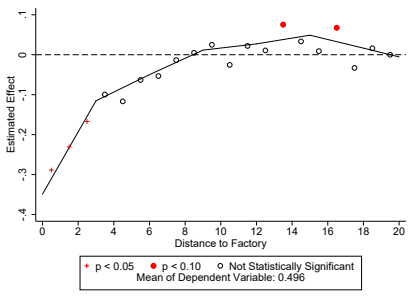
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure D-3: Industry and Agglomeration

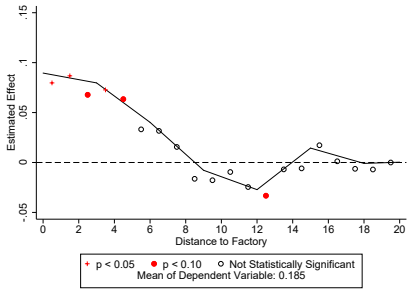
(a) Agriculture (Susenas 2001-11)



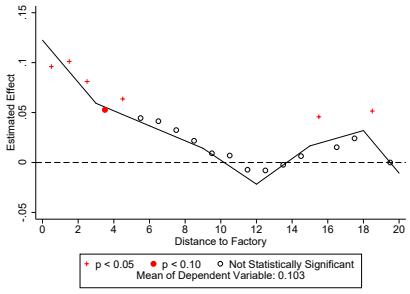
(b) Agriculture (Census 1980)



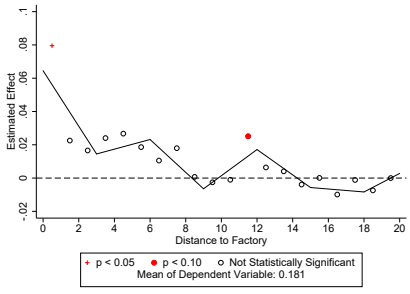
(c) Manufacturing (Susenas 2001-11)



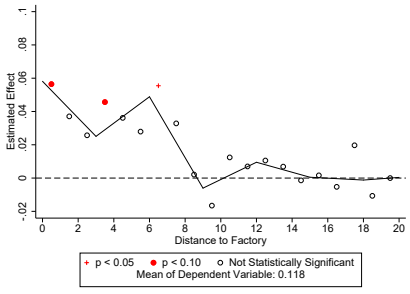
(d) Manufacturing (Census 1980)



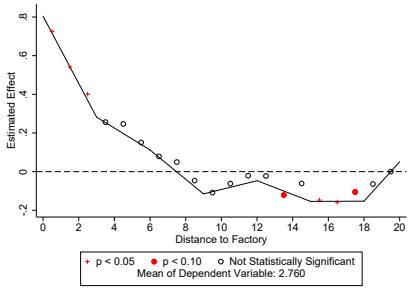
(e) Retail (Susenas 2001-11)



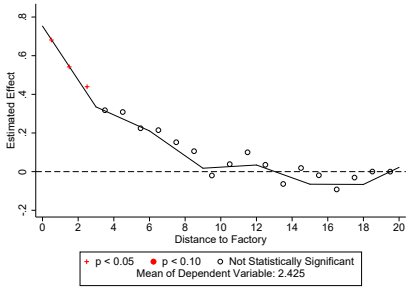
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



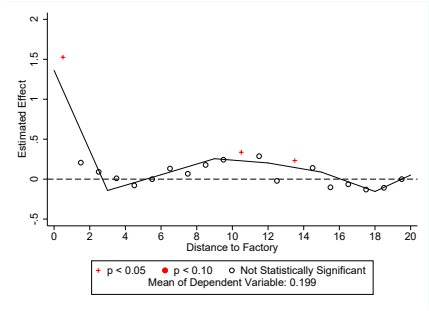
(h) Log Pop. Density (PODES 1980)



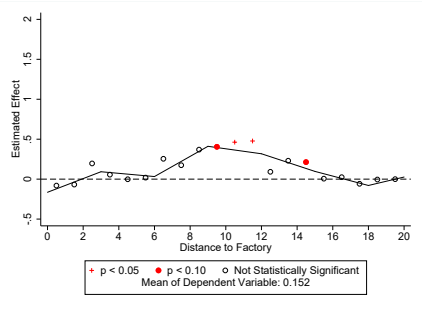
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure D-4: Sugar and Linked Industries

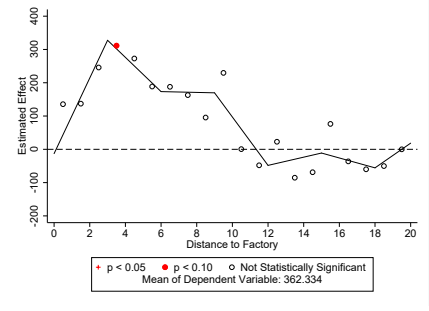
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



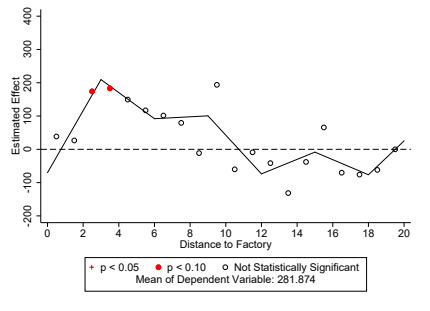
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



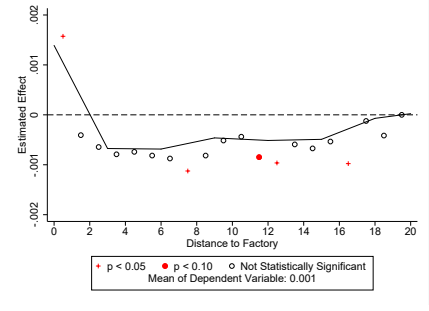
(c) Tons of Cane Grown (Full Sample, PODES 2003)



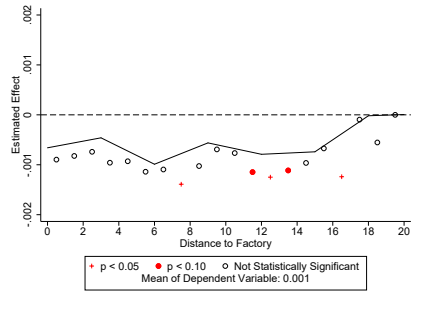
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



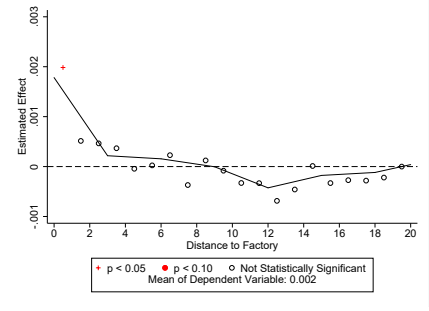
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



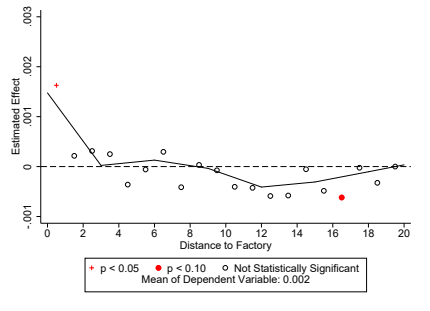
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



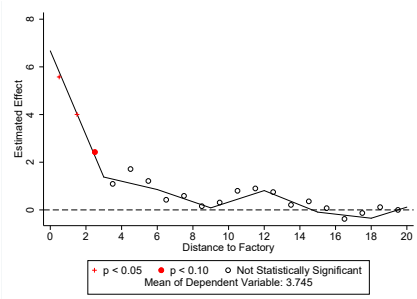
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



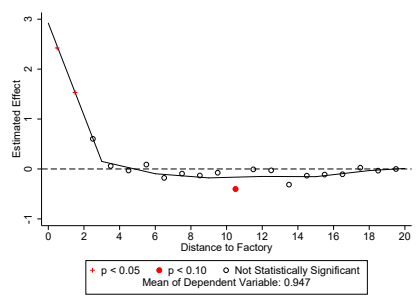
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure D-5: Infrastructure

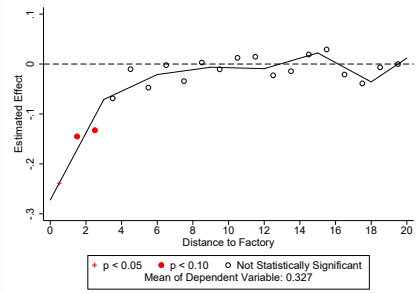
(a) Colonial Road Density (1900)



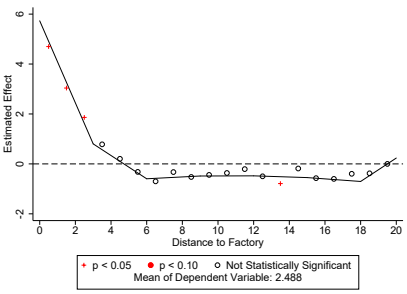
(b) Colonial Railroad Density (1900)



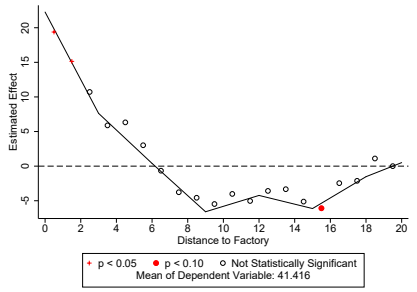
(c) Dirt Road (PODES 1980)



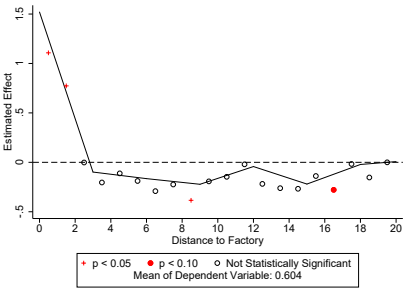
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



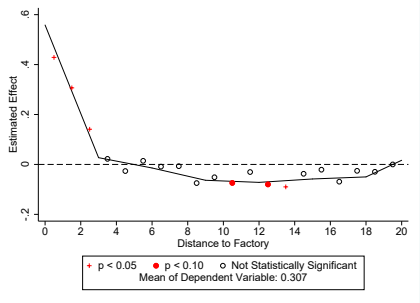
(f) Railroad Density (2017)



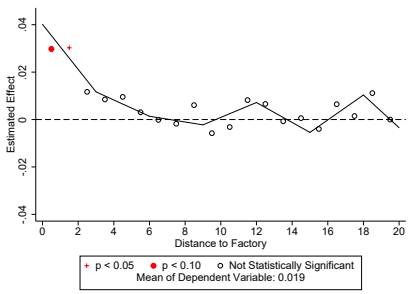
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure D-6: Other Public Goods

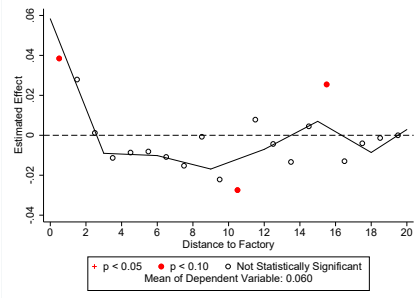
(a) Village Has Electricity (PODES 1980)



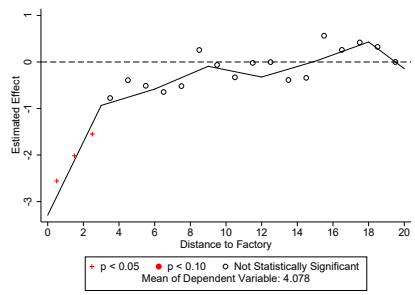
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



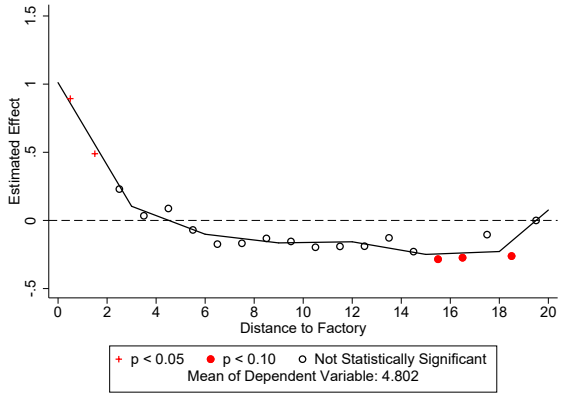
(d) Distance to Subdistrict Capital (2011 PODES)



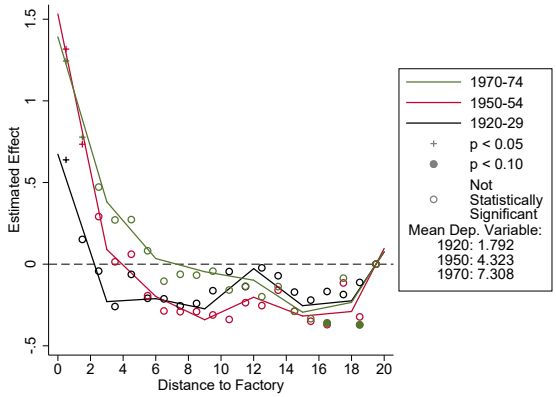
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure D-7: Education

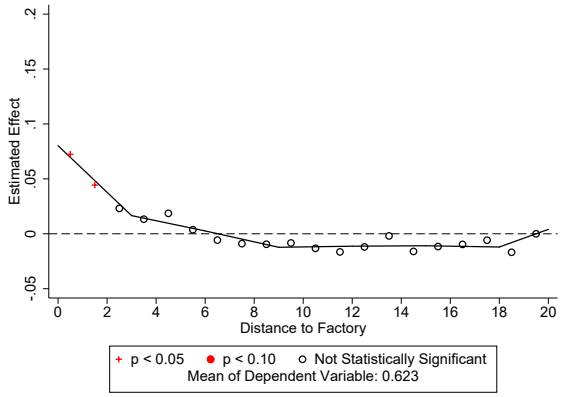
(a) Years Education (2000 Census)



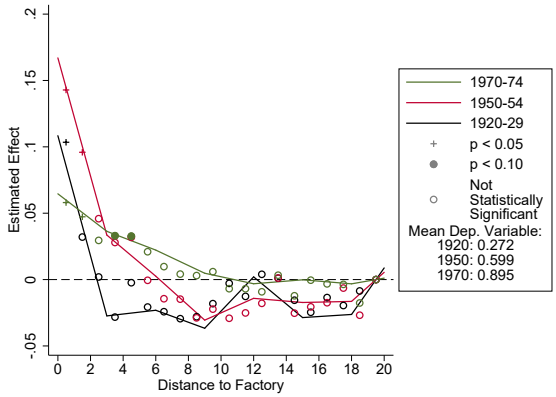
(b) Years Education by Cohort



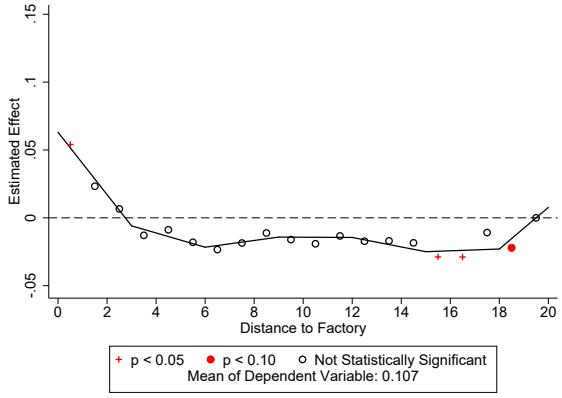
(c) Primary (2000 Census)



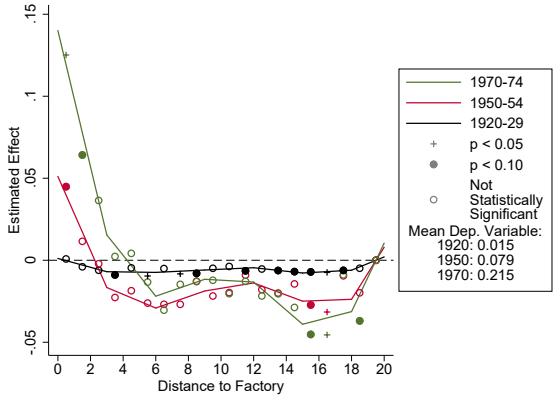
(d) Primary by Cohort



(e) High School (2000 Census)

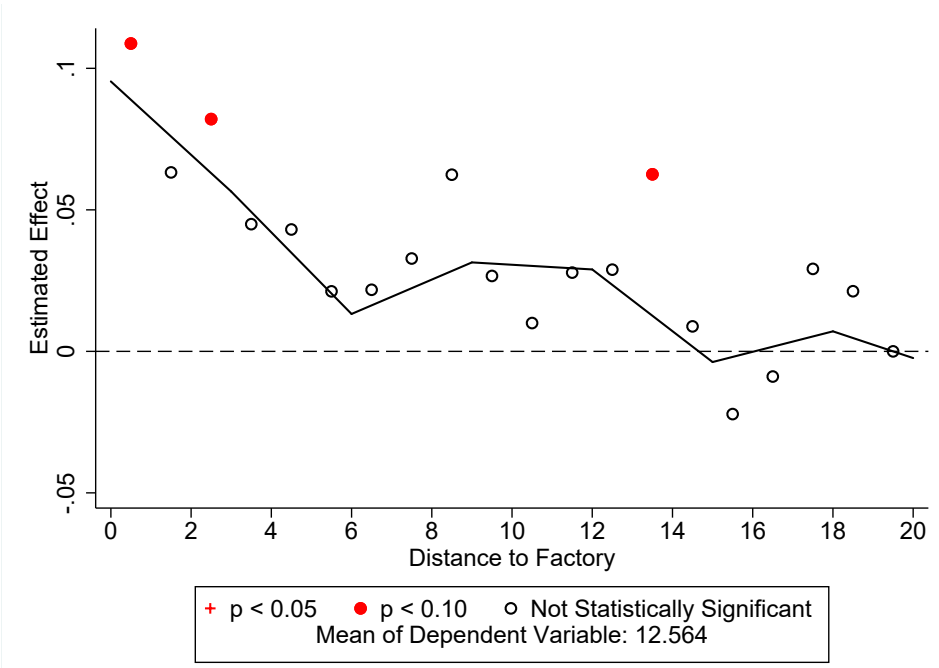


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure D-8: Expenditure (2001-11)

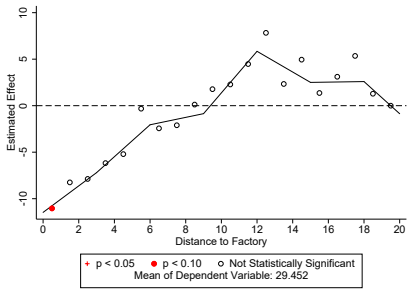


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

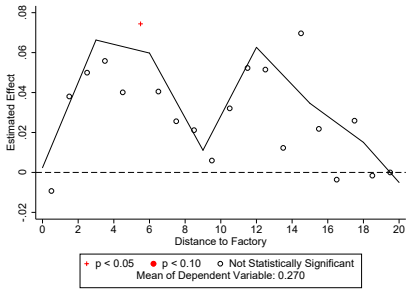
E No Factory Fixed Effects

Figure E-1: Geography

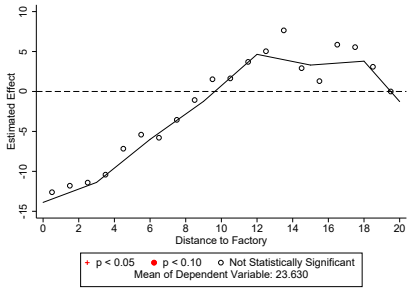
(a) Elevation



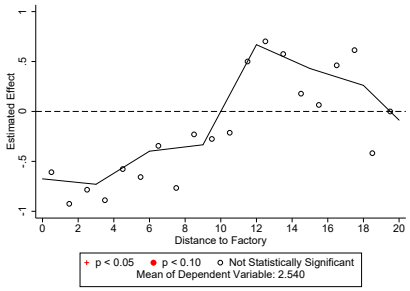
(b) Slope



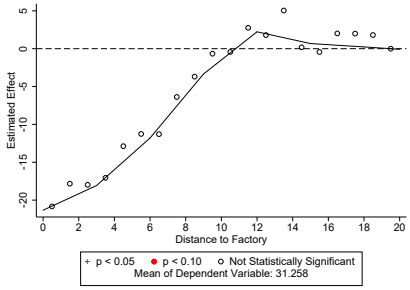
(c) Distance to Coast



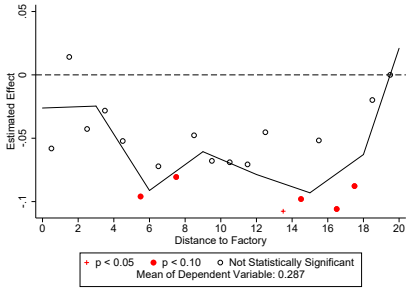
(d) Log Flow Accumulation



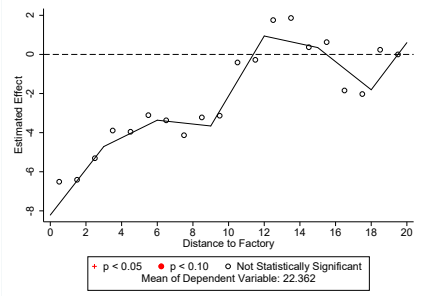
(e) Distance to Natural Harbor



(f) Distance to River



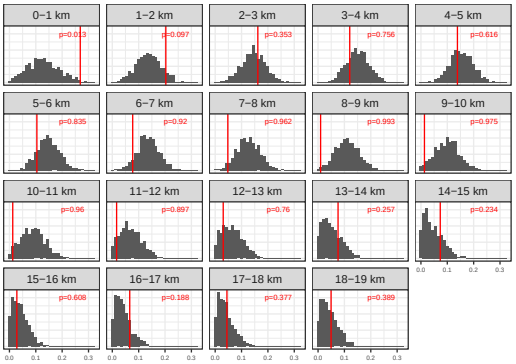
(g) Distance to Nearest 1830 Residency Capital



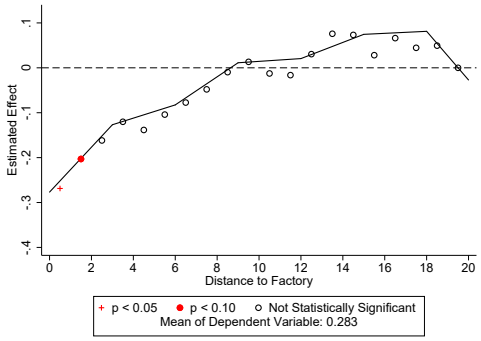
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

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

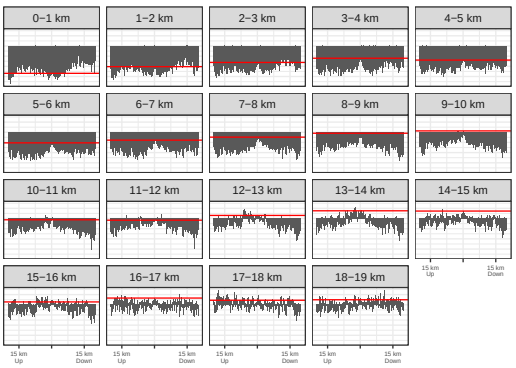
(a) Independent Shifts: Counterfactuals



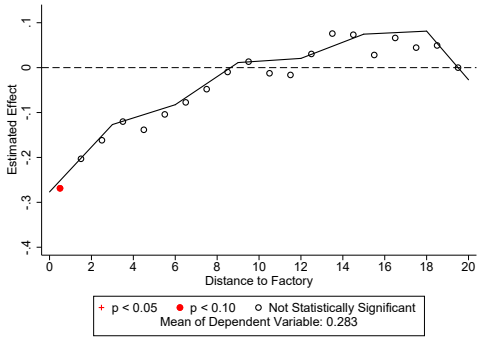
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



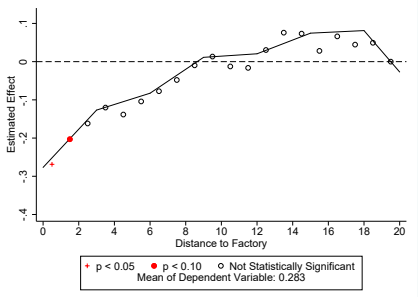
(d) Common Shifts: Plotted Coefficients



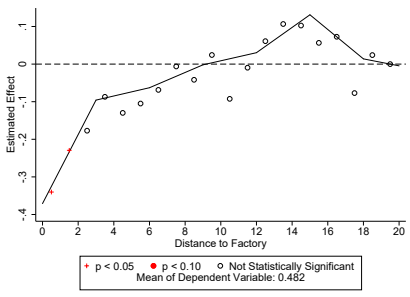
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year fixed effects. The sample is restricted 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure E-3: Industry and Agglomeration

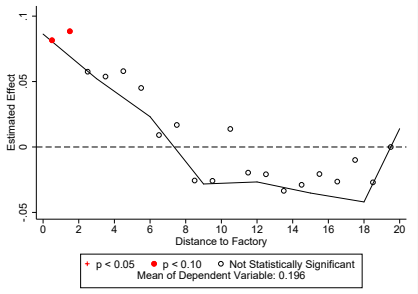
(a) Agriculture (Susenas 2001-11)



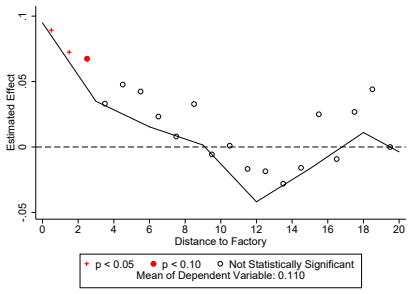
(b) Agriculture (Census 1980)



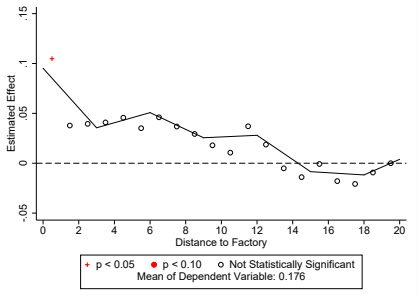
(c) Manufacturing (Susenas 2001-11)



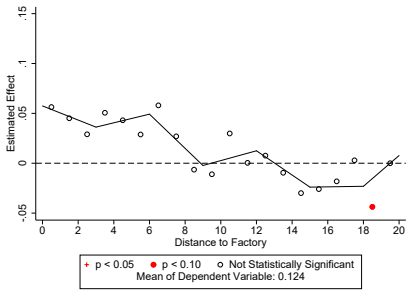
(d) Manufacturing (Census 1980)



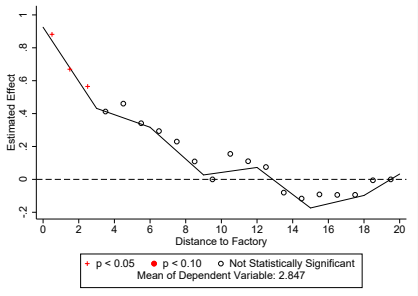
(e) Retail (Susenas 2001-11)



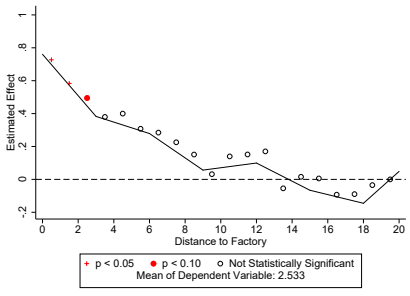
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



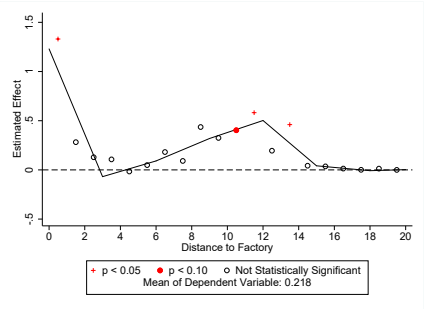
(h) Log Pop. Density (PODES 1980)



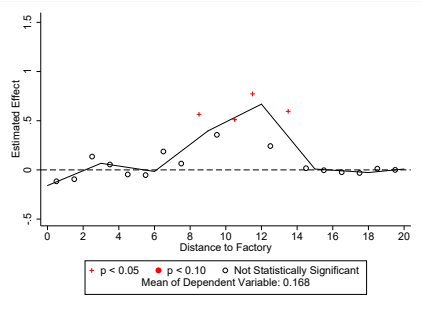
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure E-4: Sugar and Linked Industries

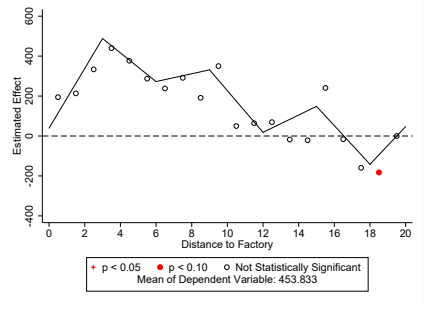
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



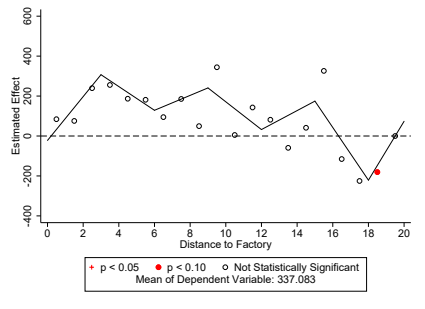
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



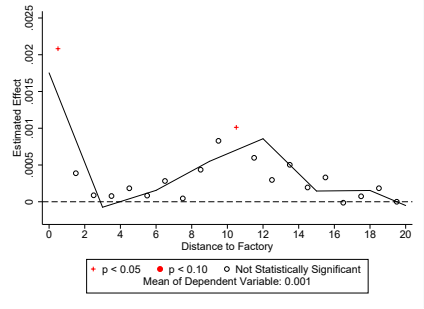
(c) Tons of Cane Grown (Full Sample, PODES 2003)



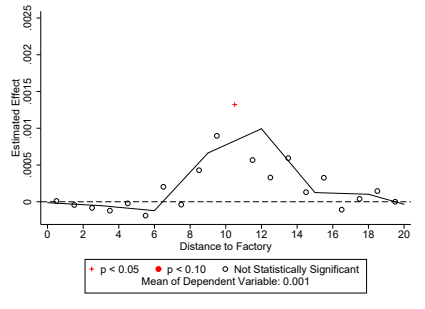
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



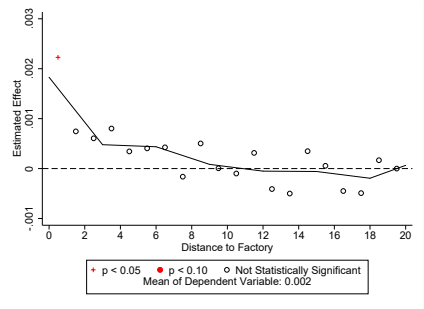
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



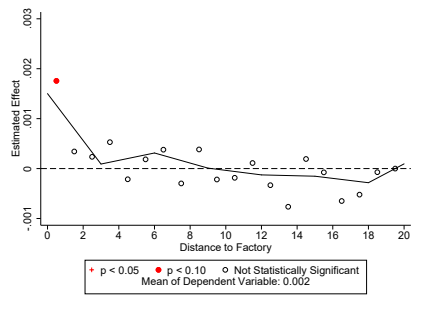
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



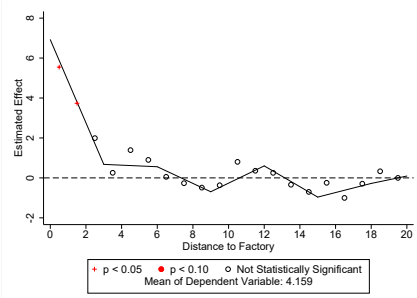
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



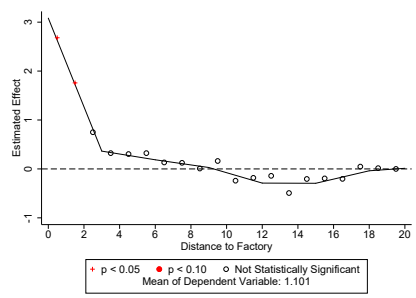
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure E-5: Infrastructure

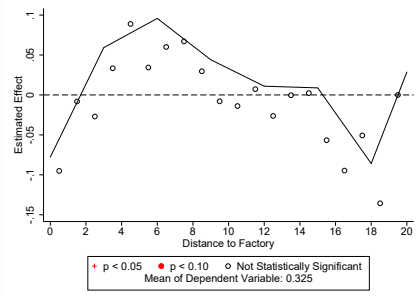
(a) Colonial Road Density (1900)



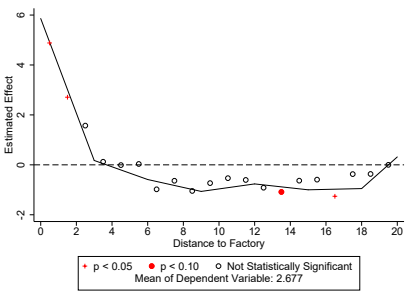
(b) Colonial Railroad Density (1900)



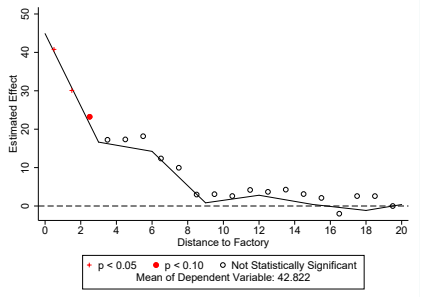
(c) Dirt Road (PODES 1980)



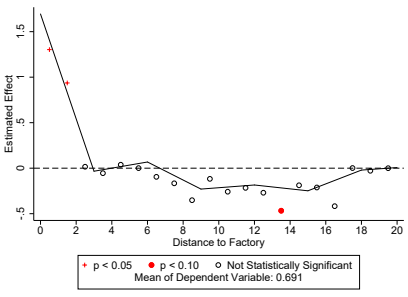
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



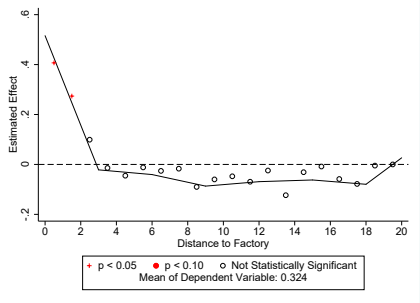
(f) Railroad Density (2017)



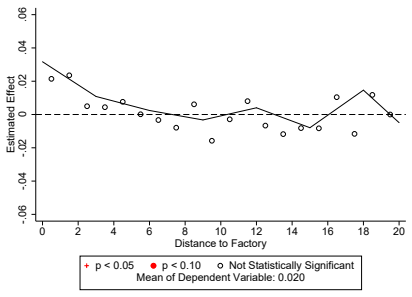
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure E-6: Other Public Goods

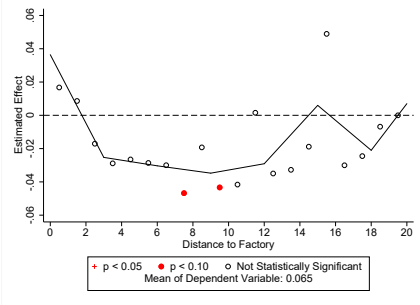
(a) Village Has Electricity (PODES 1980)



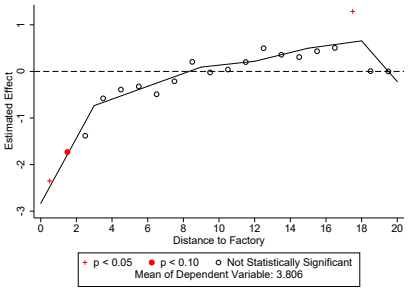
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



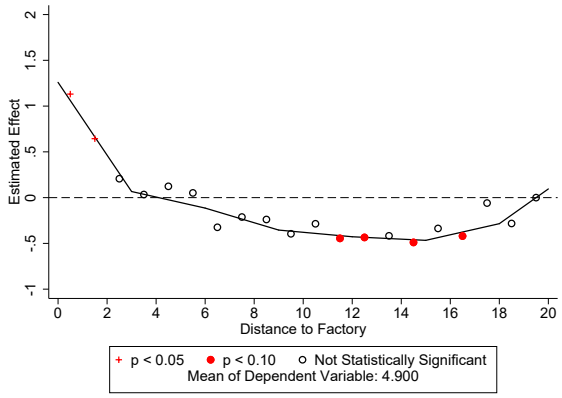
(d) Distance to Subdistrict Capital (2011 PODES)



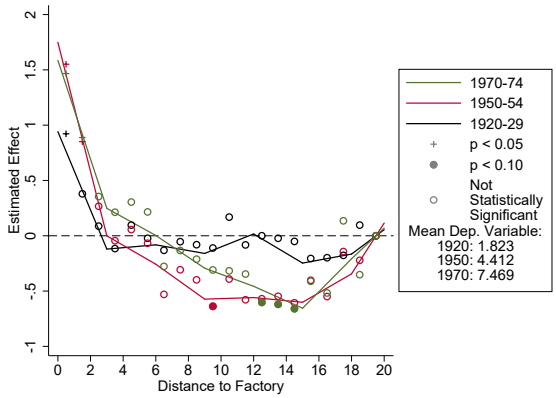
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure E-7: Education

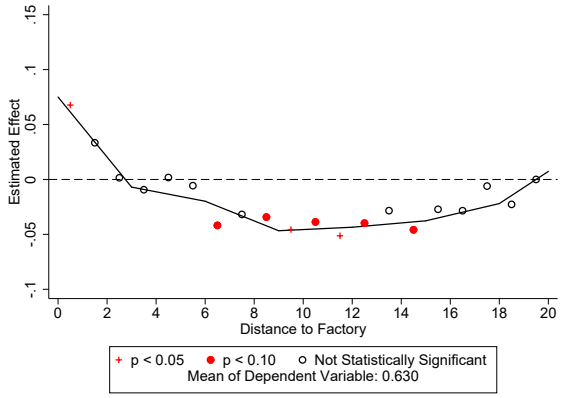
(a) Years Education (2000 Census)



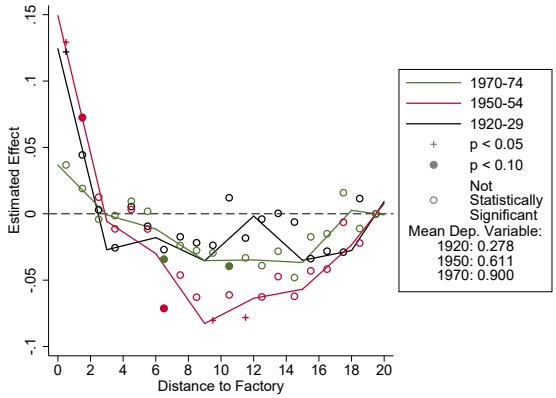
(b) Years Education by Cohort



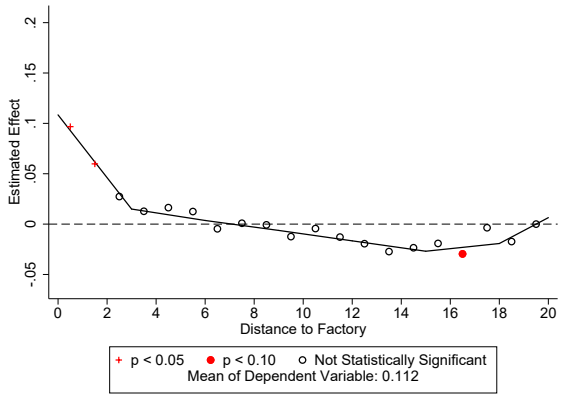
(c) Primary (2000 Census)



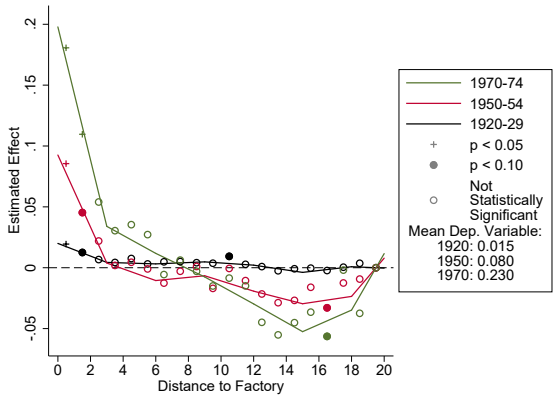
(d) Primary by Cohort



(e) High School (2000 Census)

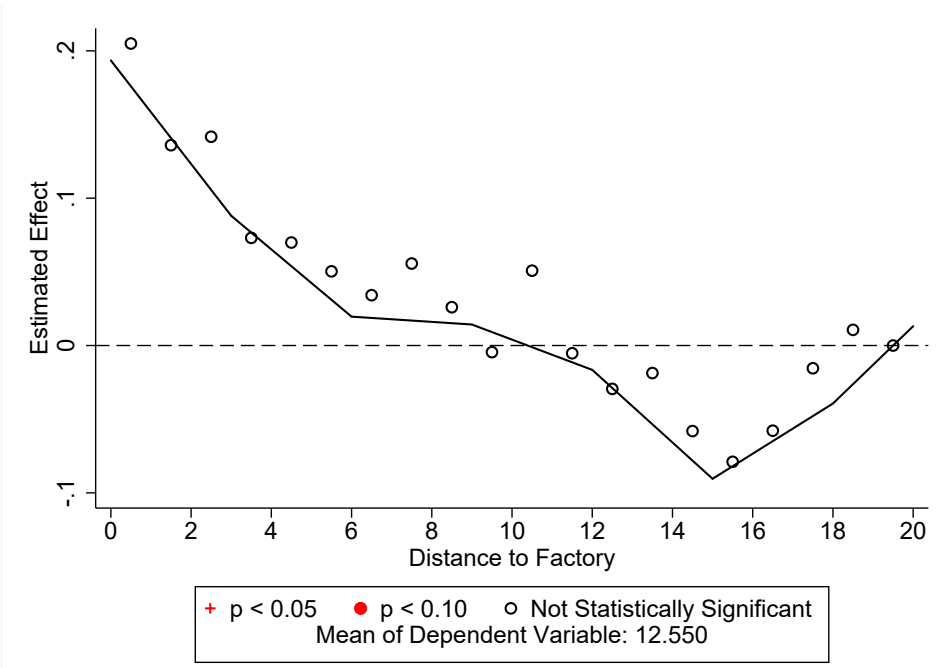


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure E-8: Expenditure (2001-11)

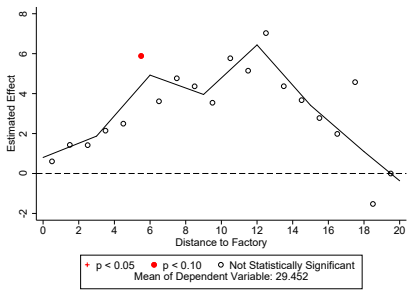


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

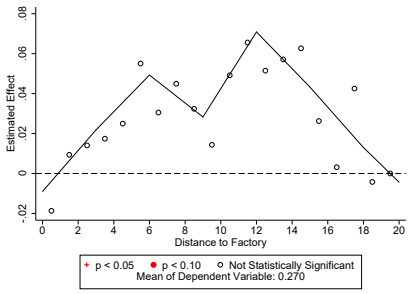
F Includes Catchment Dummy

Figure F-1: Geography

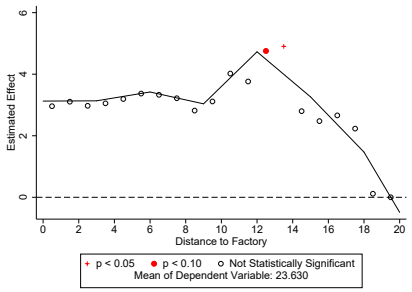
(a) Elevation



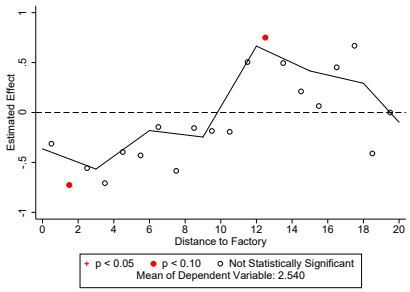
(b) Slope



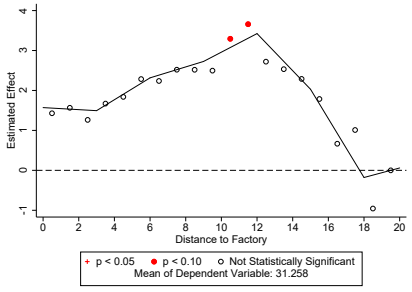
(c) Distance to Coast



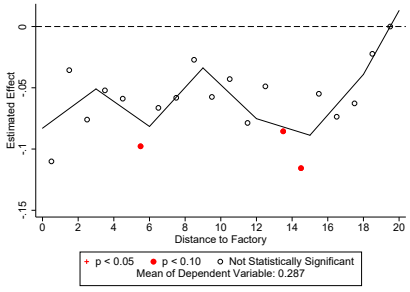
(d) Log Flow Accumulation



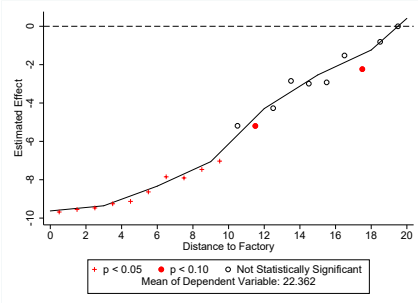
(e) Distance to Natural Harbor



(f) Distance to River



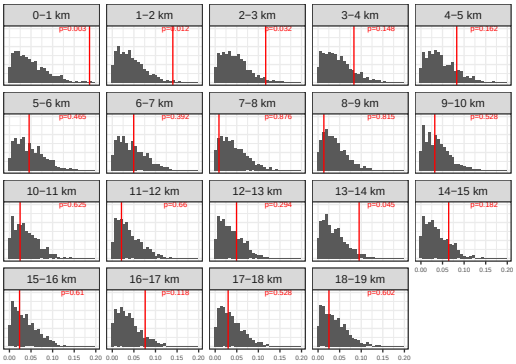
(g) Distance to Nearest 1830 Residency Capital



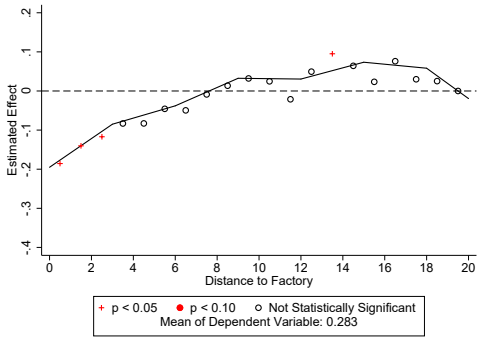
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects and a catchment area dummy. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

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

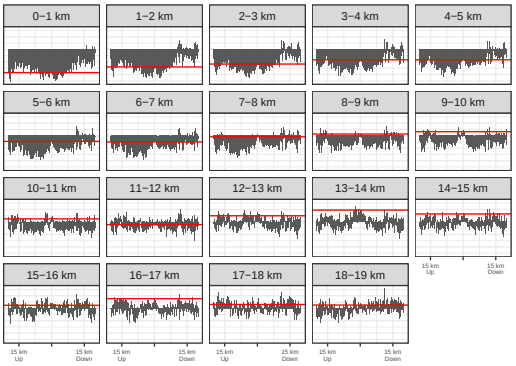
(a) Independent Shifts: Counterfactuals



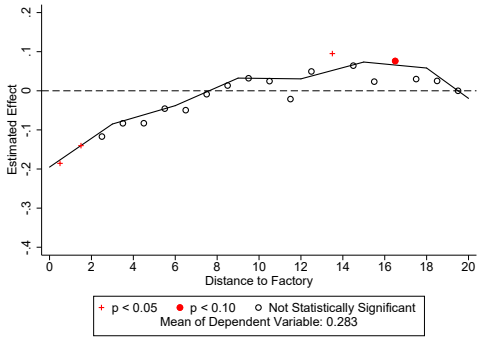
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



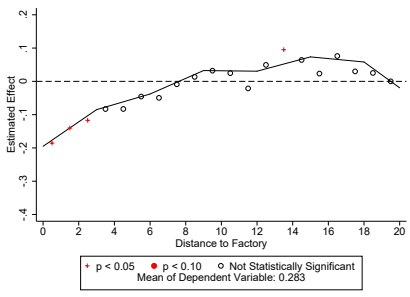
(d) Common Shifts: Plotted Coefficients



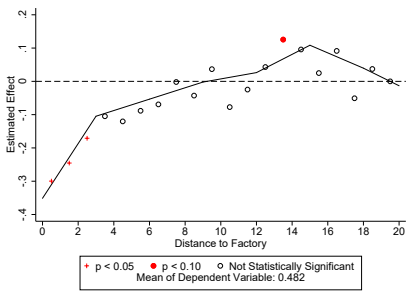
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, a catchment area dummy, geographic controls, a linear spline in distance to the nearest 1830 residency capital, 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure F-3: Industry and Agglomeration

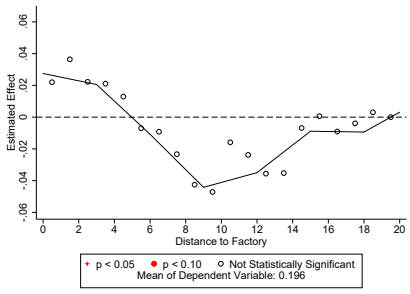
(a) Agriculture (Susenas 2001-11)



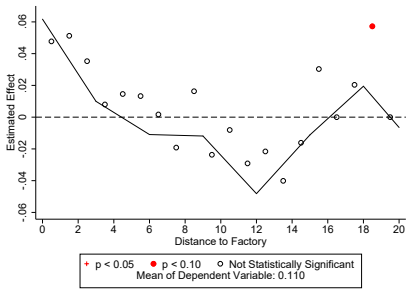
(b) Agriculture (Census 1980)



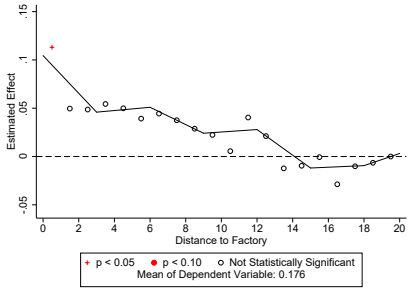
(c) Manufacturing (Susenas 2001-11)



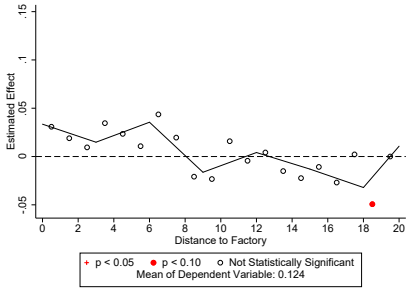
(d) Manufacturing (Census 1980)



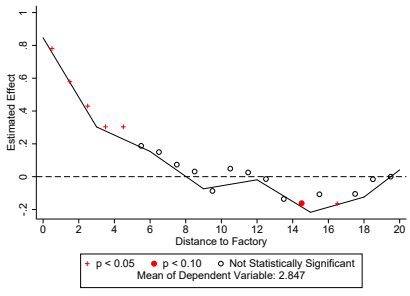
(e) Retail (Susenas 2001-11)



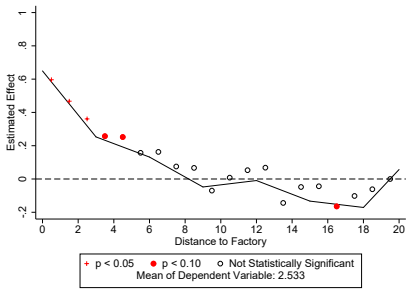
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



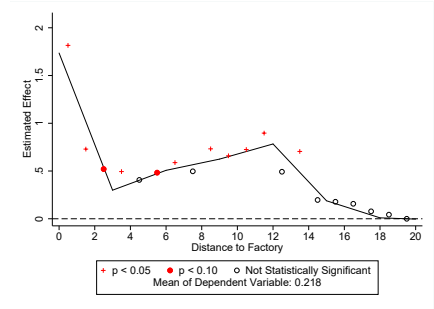
(h) Log Pop. Density (PODES 1980)



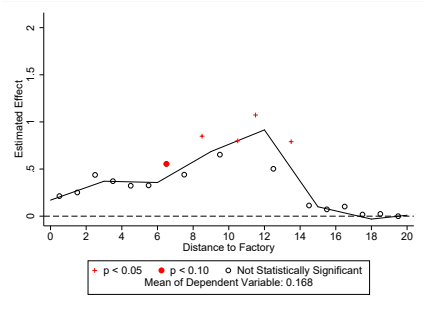
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. 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 data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure F-4: Sugar and Linked Industries

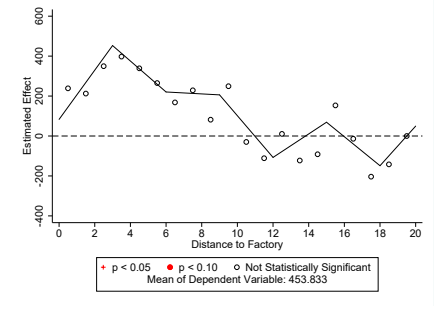
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



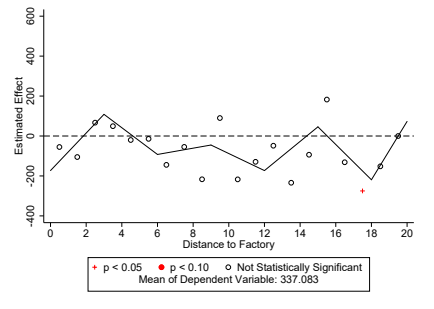
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



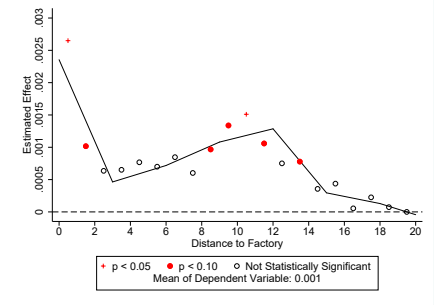
(c) Tons of Cane Grown (Full Sample, PODES 2003)



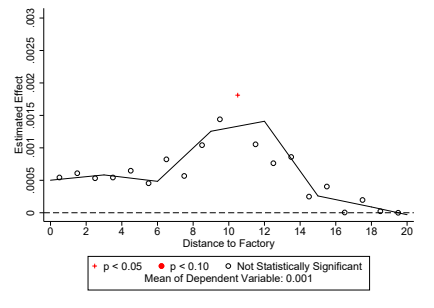
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



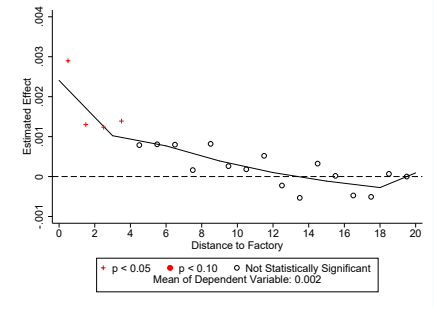
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



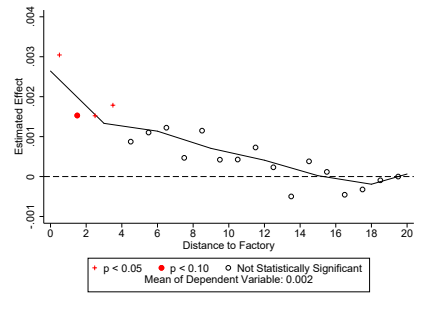
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



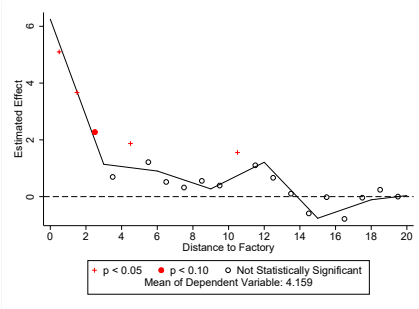
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



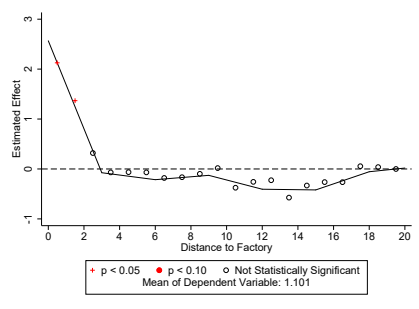
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure F-5: Infrastructure

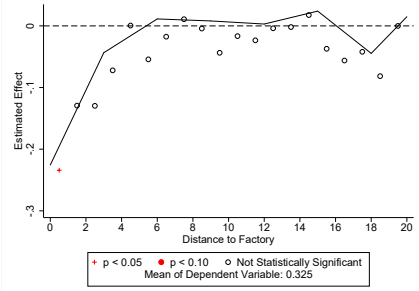
(a) Colonial Road Density (1900)



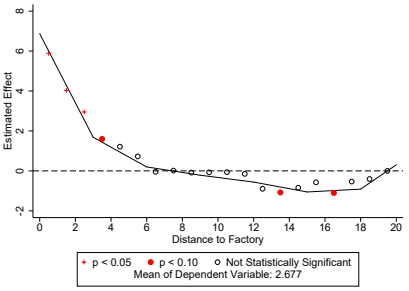
(b) Colonial Railroad Density (1900)



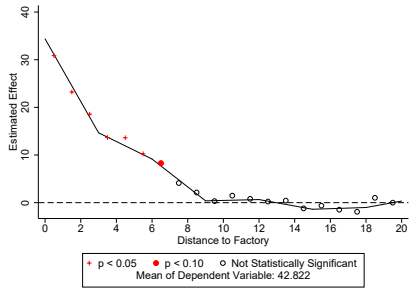
(c) Dirt Road (PODES 1980)



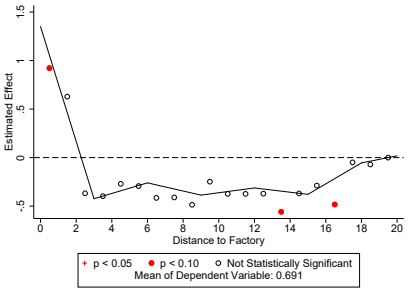
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



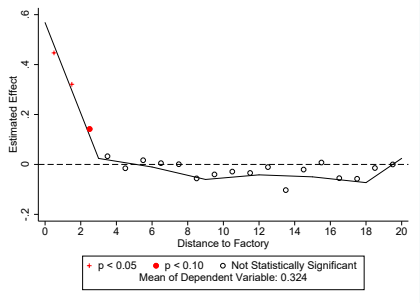
(f) Railroad Density (2017)



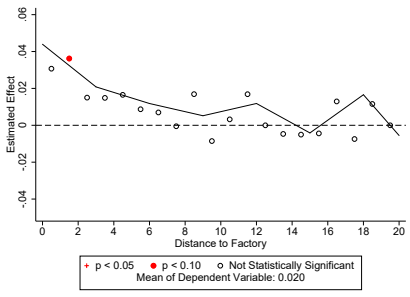
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure F-6: Other Public Goods

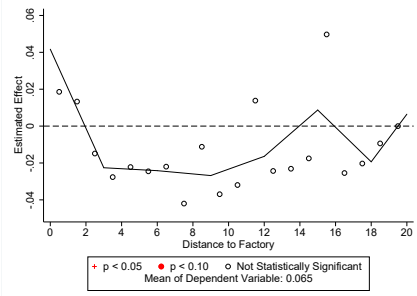
(a) Village Has Electricity (PODES 1980)



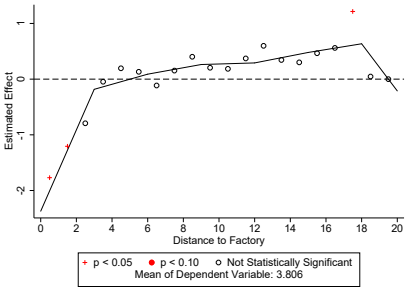
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



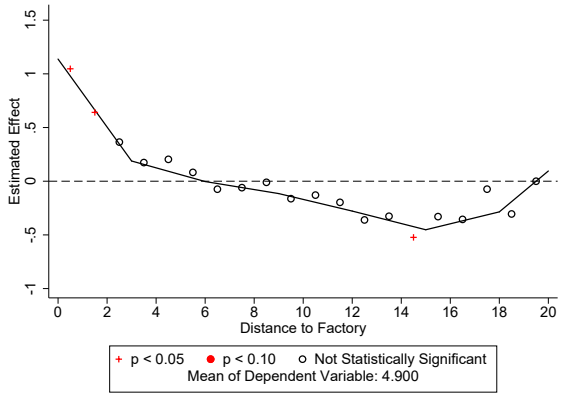
(d) Distance to Subdistrict Capital (2011 PODES)



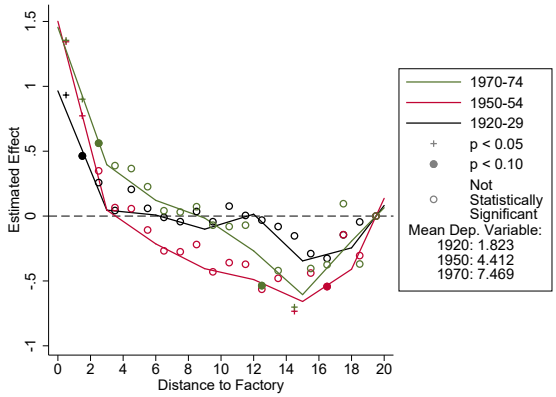
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure F-7: Education

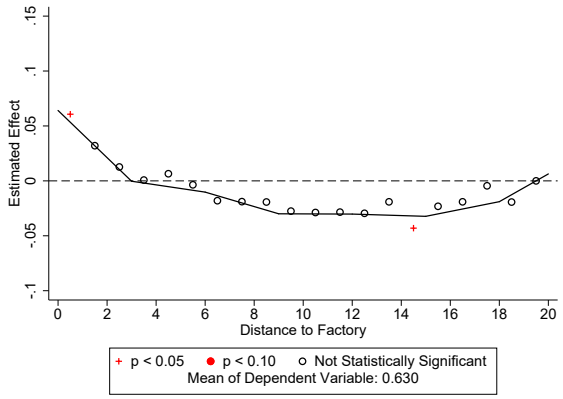
(a) Years Education (2000 Census)



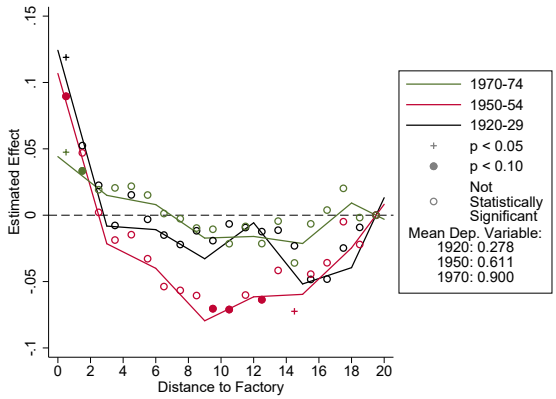
(b) Years Education by Cohort



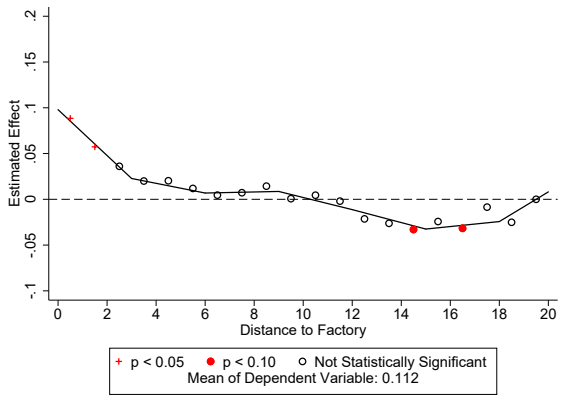
(c) Primary (2000 Census)



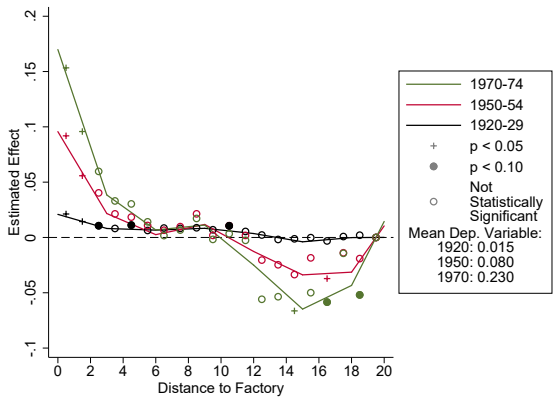
(d) Primary by Cohort



(e) High School (2000 Census)

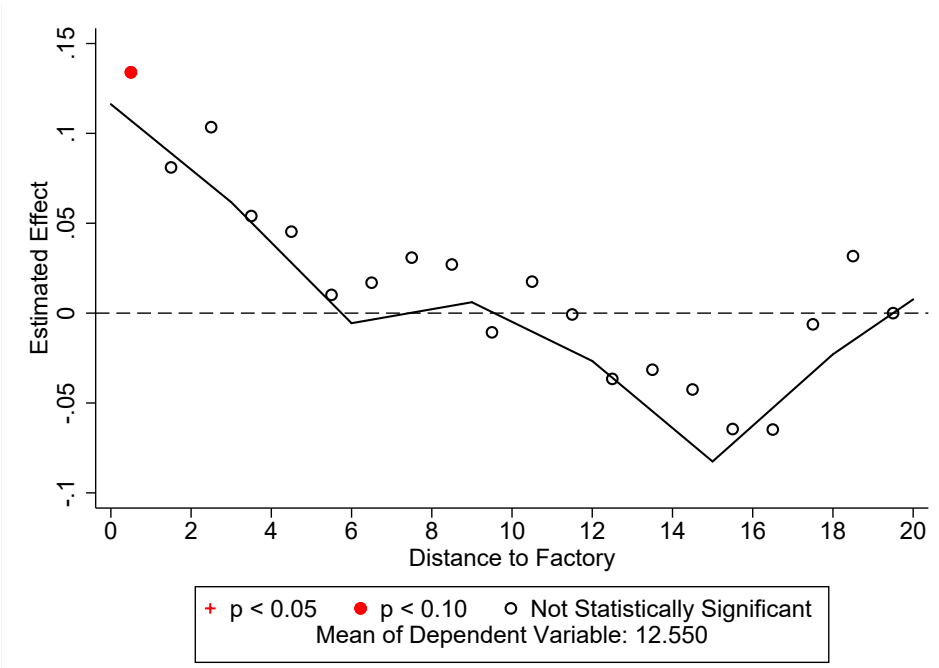


(f) High School by Cohort



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, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure F-8: Expenditure (2001-11)

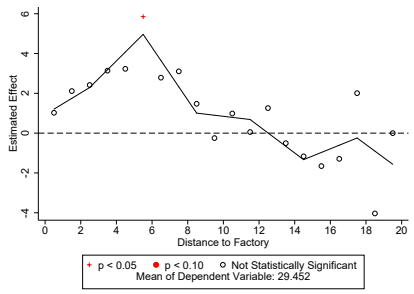


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, a catchment area dummy, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

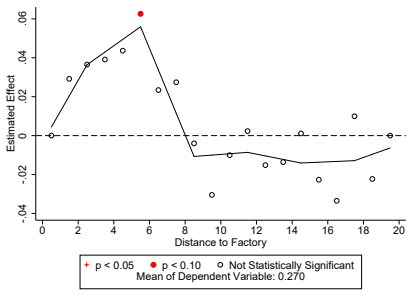
G Subtract Placebo Means

Figure G-1: Geography

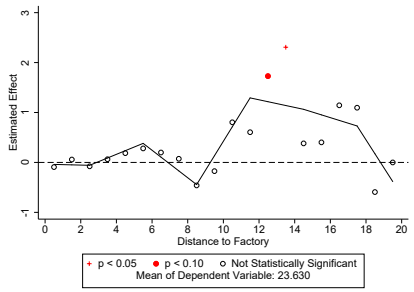
(a) Elevation



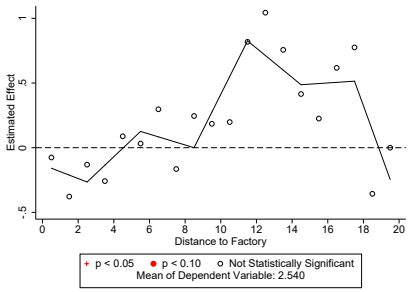
(b) Slope



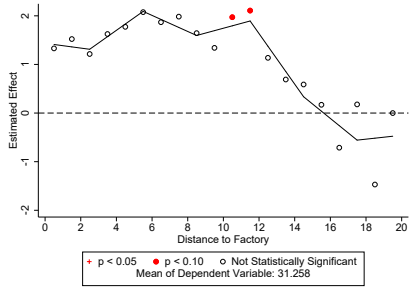
(c) Distance to Coast



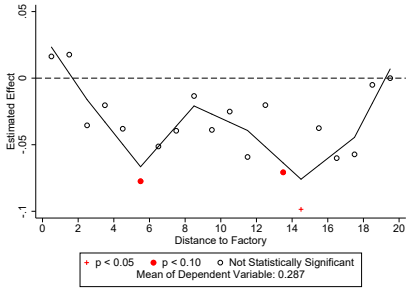
(d) Log Flow Accumulation



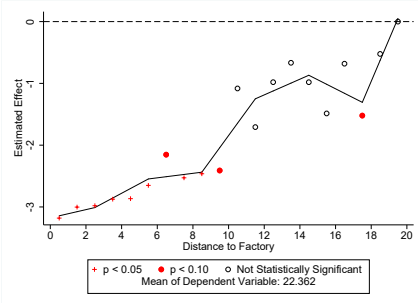
(e) Distance to Natural Harbor



(f) Distance to River



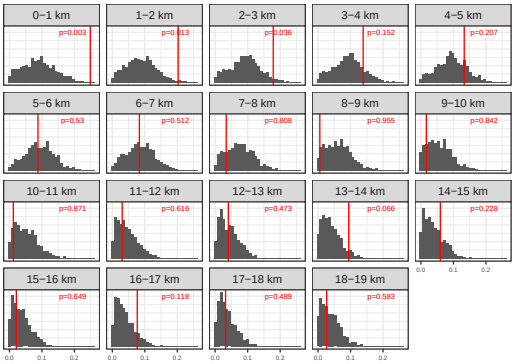
(g) Distance to Nearest 1830 Residency Capital



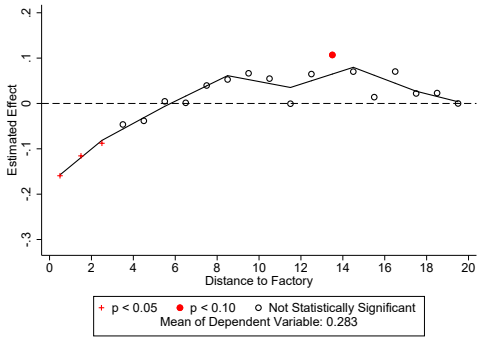
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of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

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

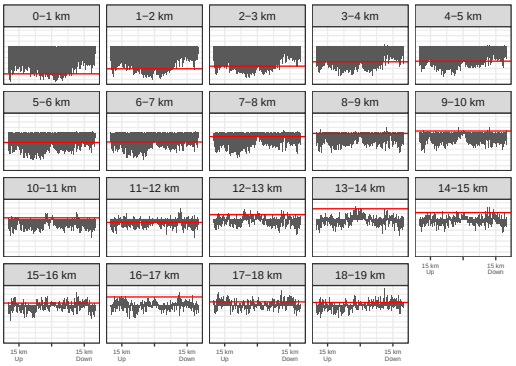
(a) Independent Shifts: Counterfactuals



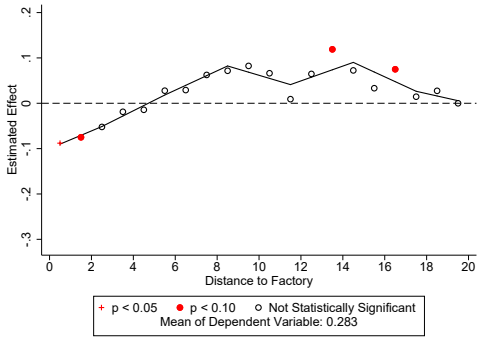
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



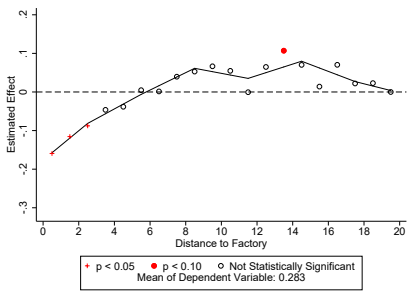
(d) Common Shifts: Plotted Coefficients



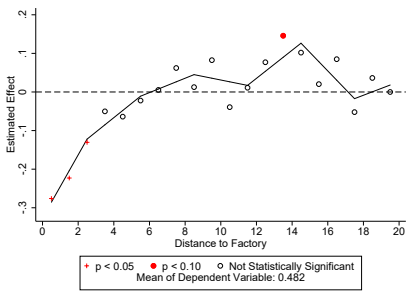
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure G-3: Industry and Agglomeration

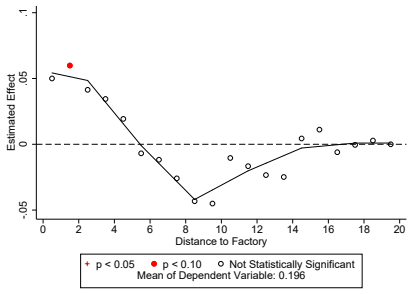
(a) Agriculture (Susenas 2001-11)



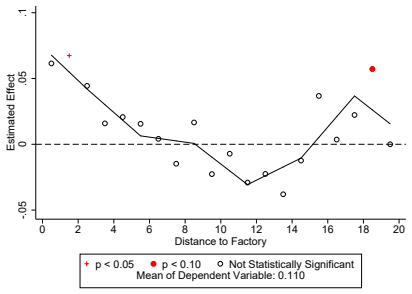
(b) Agriculture (Census 1980)



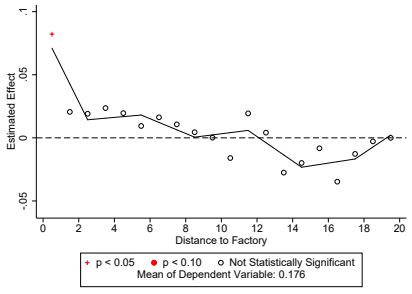
(c) Manufacturing (Susenas 2001-11)



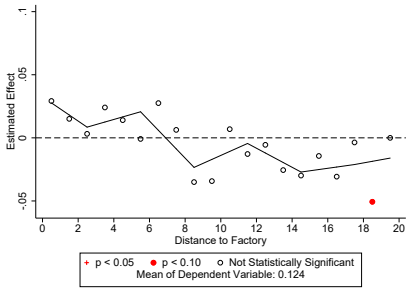
(d) Manufacturing (Census 1980)



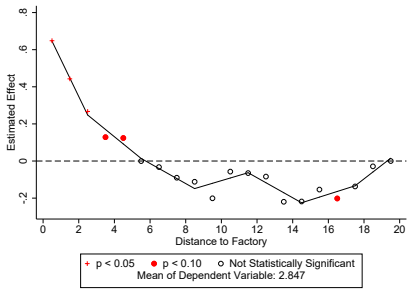
(e) Retail (Susenas 2001-11)



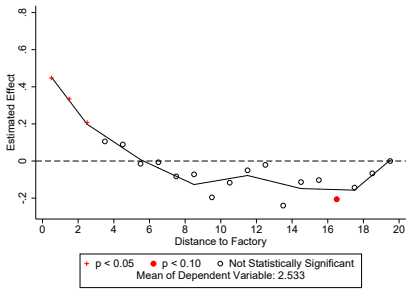
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



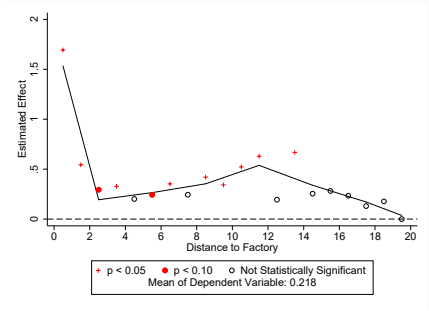
(h) Log Pop. Density (PODES 1980)



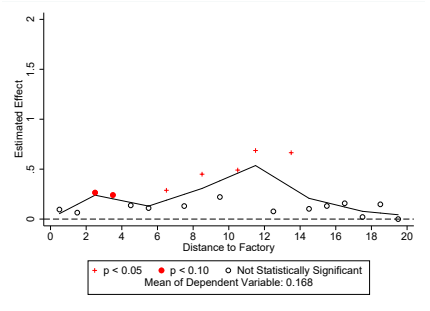
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure G-4: Sugar and Linked Industries

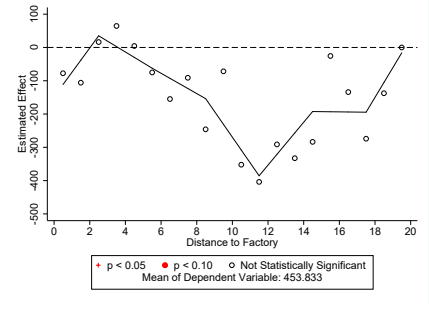
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



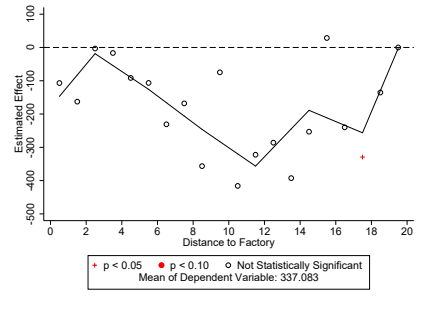
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



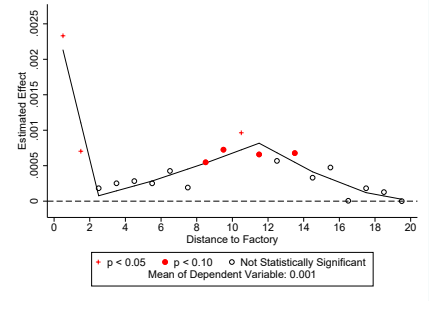
(c) Tons of Cane Grown (Full Sample, PODES 2003)



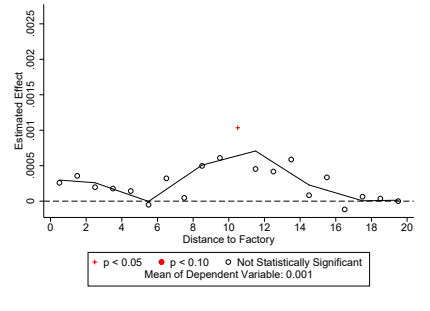
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



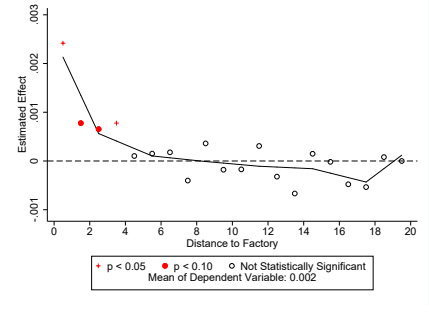
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



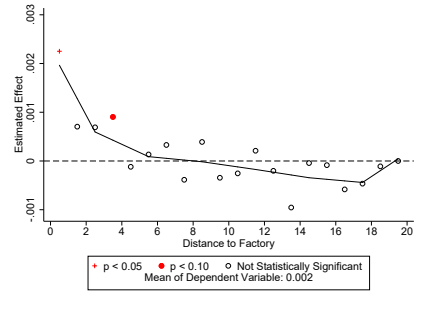
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



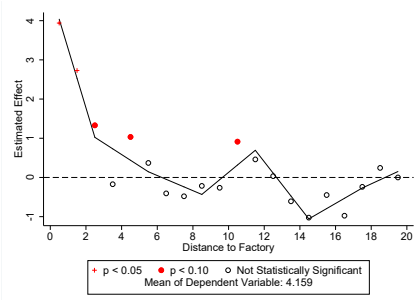
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



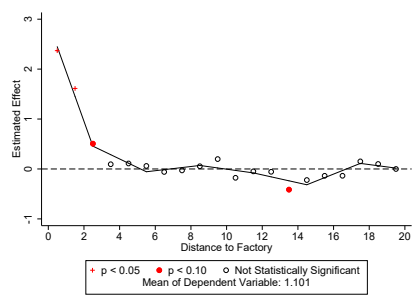
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure G-5: Infrastructure

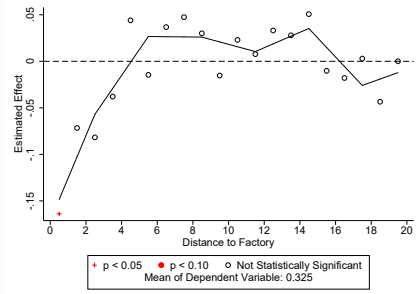
(a) Colonial Road Density (1900)



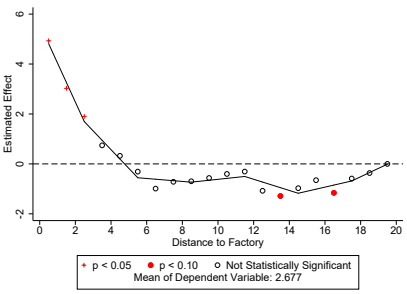
(b) Colonial Railroad Density (1900)



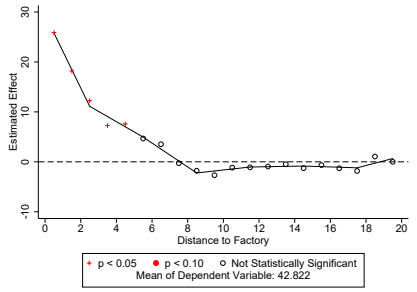
(c) Dirt Road (PODES 1980)



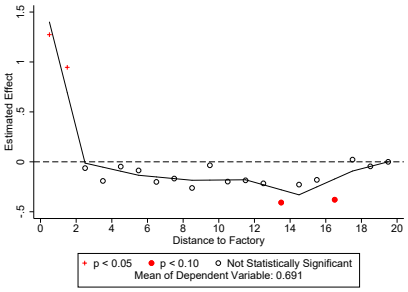
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



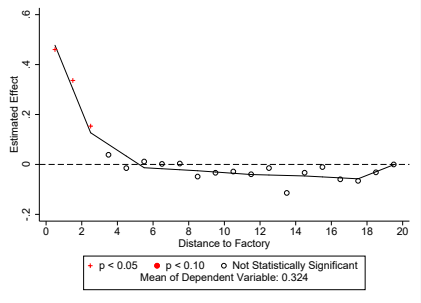
(f) Railroad Density (2017)



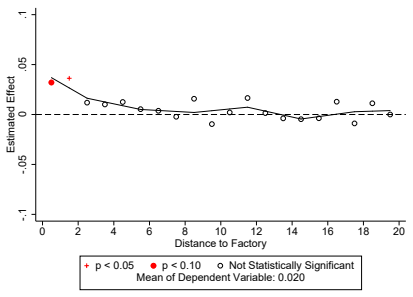
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure G-6: Other Public Goods

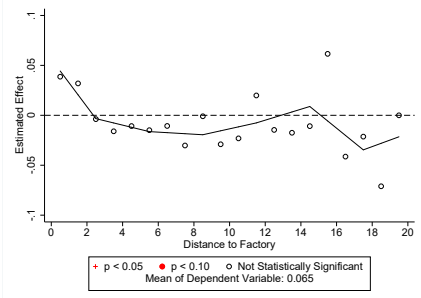
(a) Village Has Electricity (PODES 1980)



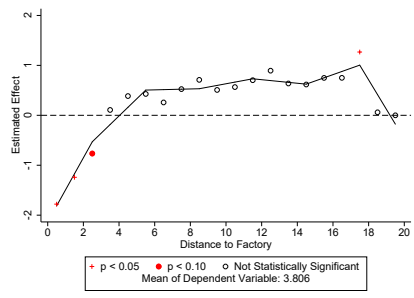
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



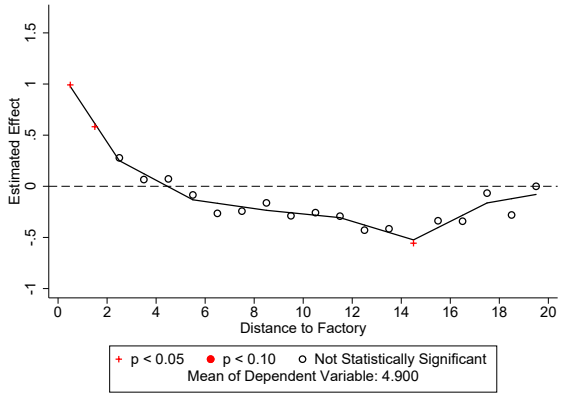
(d) Distance to Subdistrict Capital (2011 PODES)



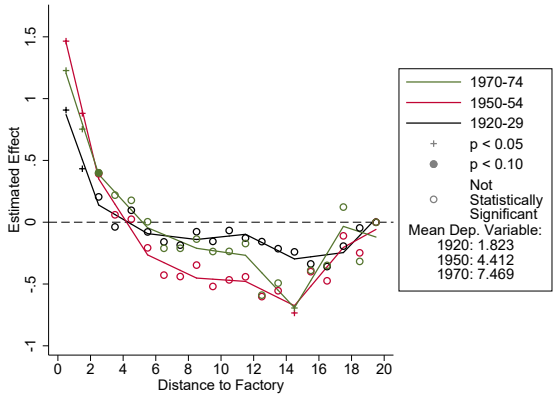
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure G-7: Education

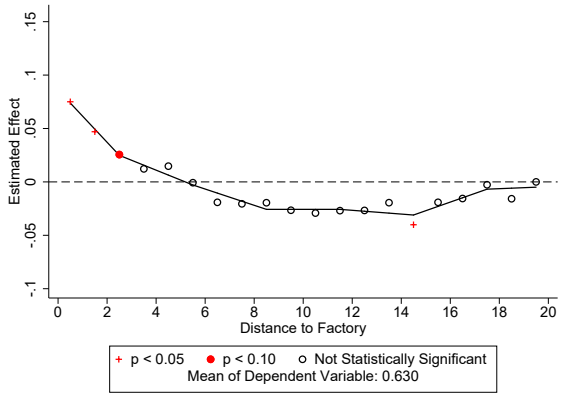
(a) Years Education (2000 Census)



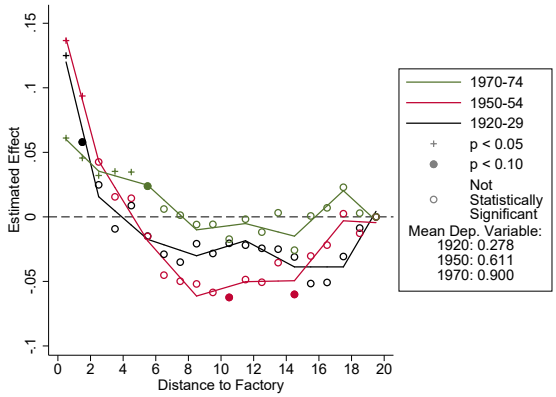
(b) Years Education by Cohort



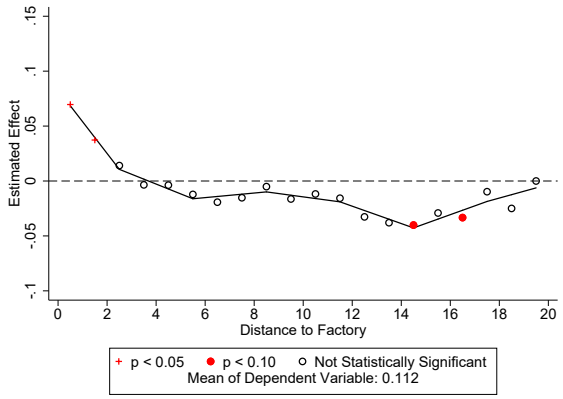
(c) Primary (2000 Census)



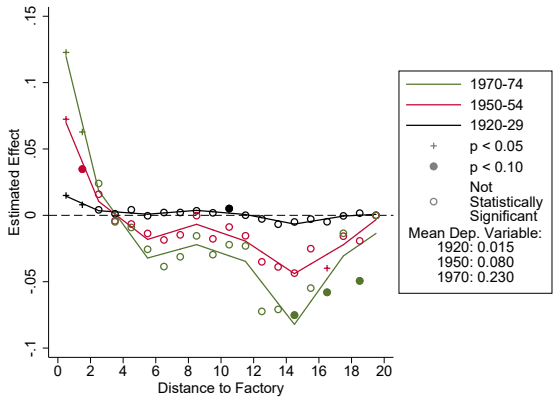
(d) Primary by Cohort



(e) High School (2000 Census)

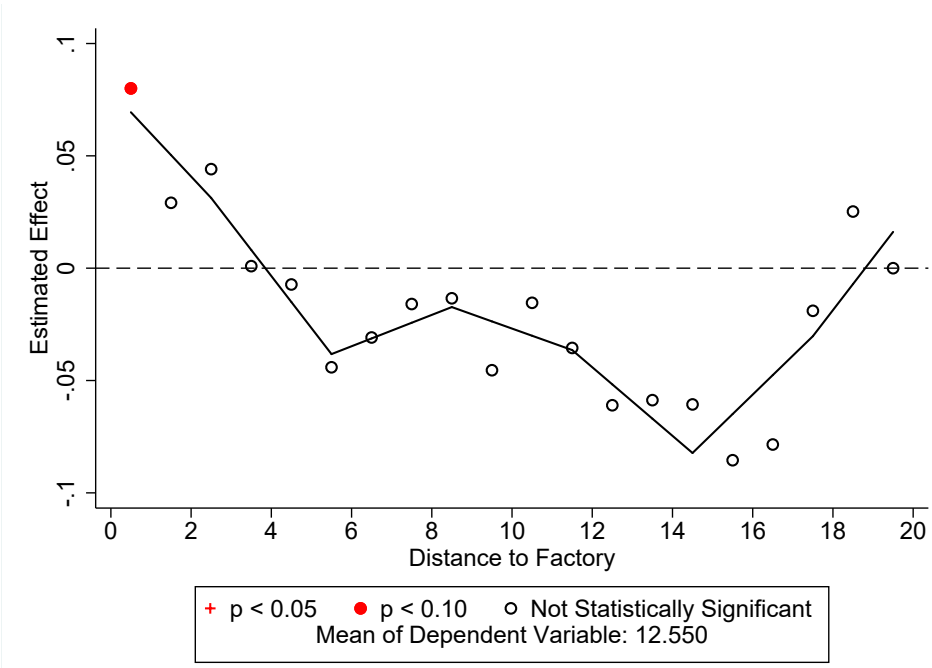


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure G-8: Expenditure (2001-11)

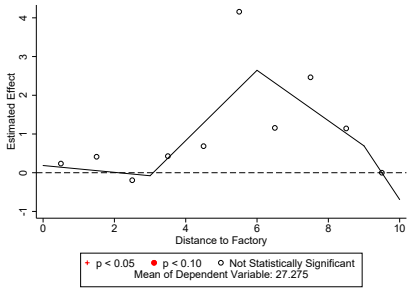


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

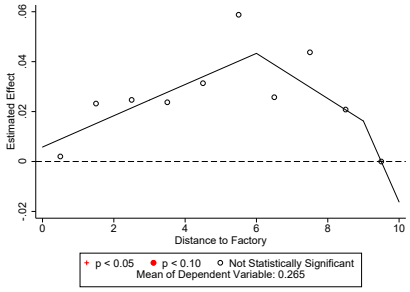
H Remove Factories Within 10 Km of Residency Capital

Figure H-1: Geography

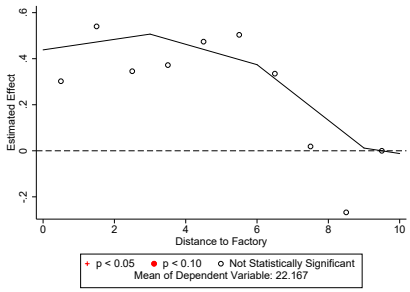
(a) Elevation



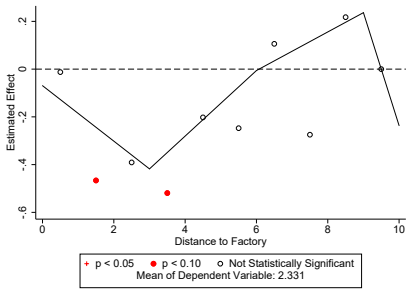
(b) Slope



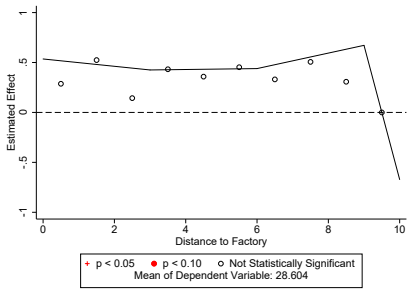
(c) Distance to Coast



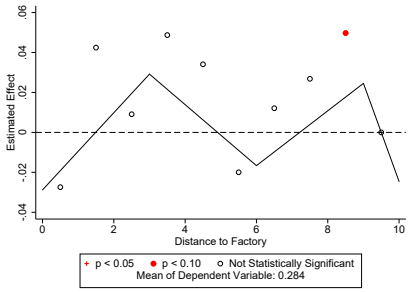
(d) Log Flow Accumulation



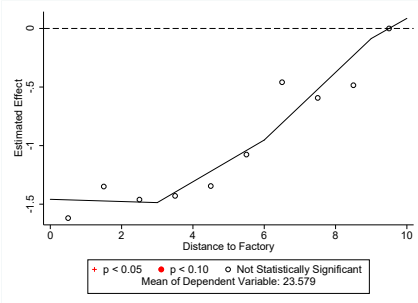
(e) Distance to Natural Harbor



(f) Distance to River



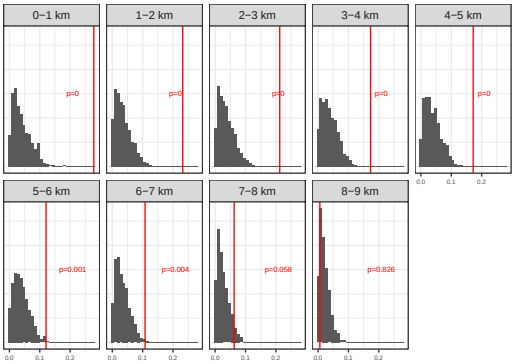
(g) Distance to Nearest 1830 Residency Capital



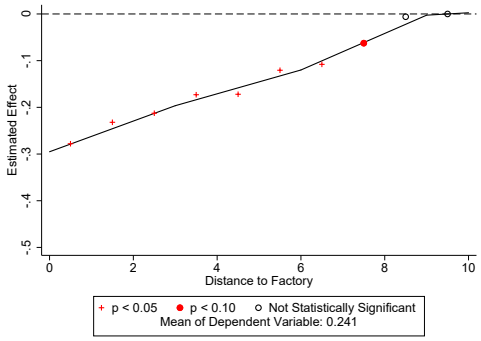
Notes: Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of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

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

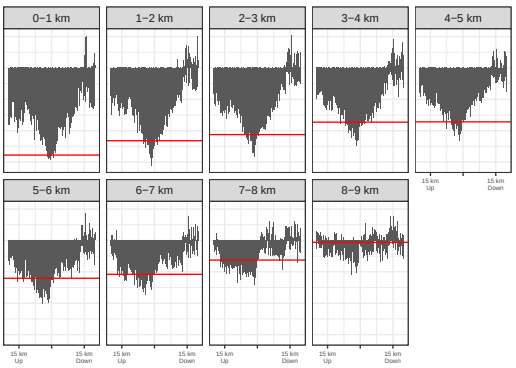
(a) Independent Shifts: Counterfactuals



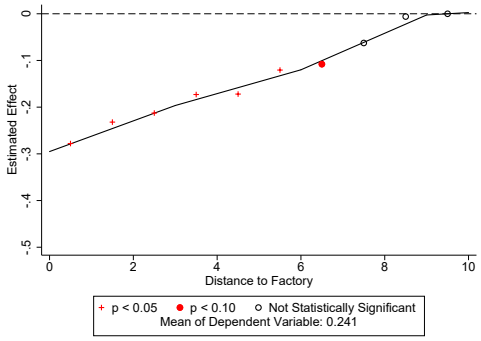
(b) Independent Shifts: Plotted Coefficients



(c) Common Shifts: Counterfactuals



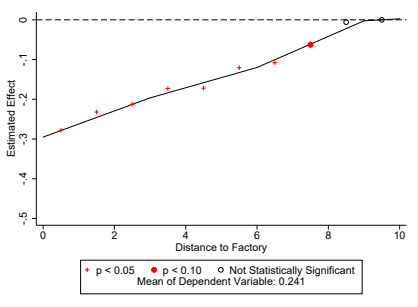
(d) Common Shifts: Plotted Coefficients



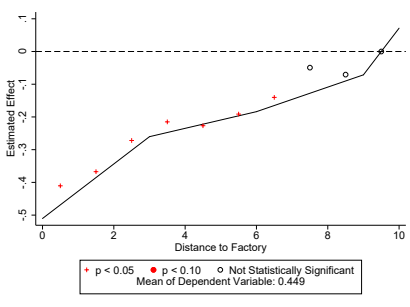
Notes: Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 1000 sets of counterfactual factories. The coefficients for distance to the real factories are shown as vertical lines. Panel (b) plots the real coefficients for each bin, with the symbols indicating their position 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 their position in the distribution of counterfactual coefficients shown in panel (c).

Figure H-3: Industry and Agglomeration

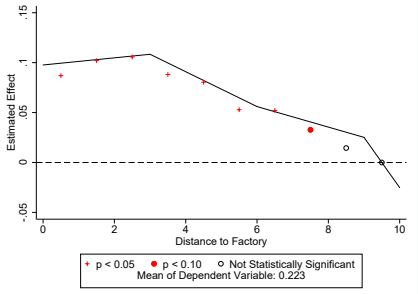
(a) Agriculture (Susenas 2001-11)



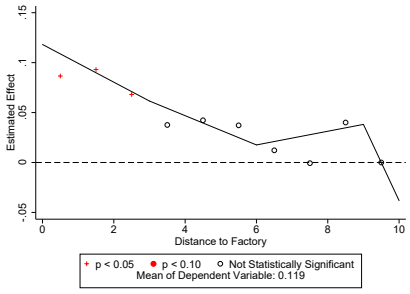
(b) Agriculture (Census 1980)



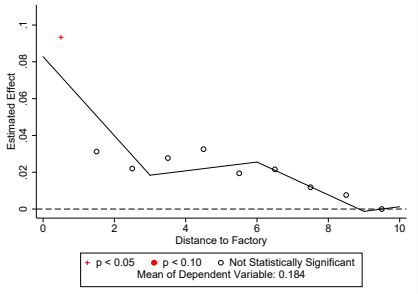
(c) Manufacturing (Susenas 2001-11)



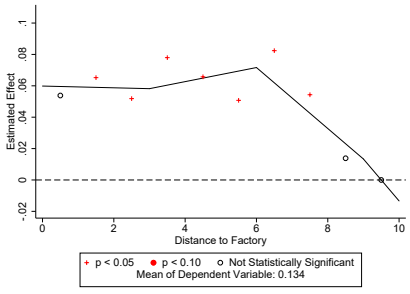
(d) Manufacturing (Census 1980)



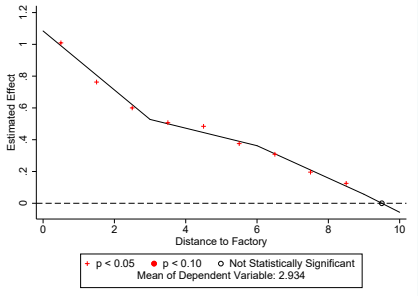
(e) Retail (Susenas 2001-11)



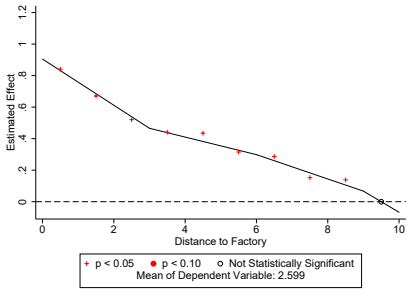
(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



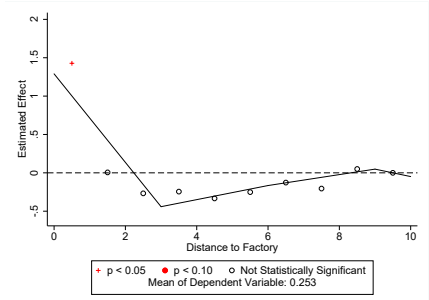
(h) Log Pop. Density (PODES 1980)



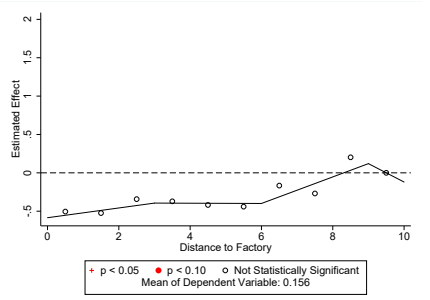
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panels a), c), and e) include survey year fixed effects. In panels a) through f), the sample is restricted to men aged 18 to 55. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure H-4: Sugar and Linked Industries

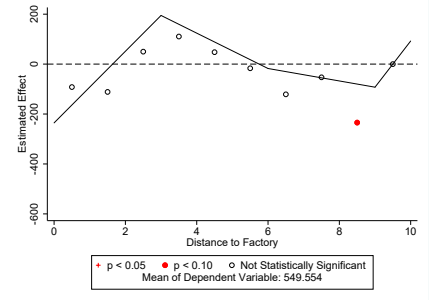
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



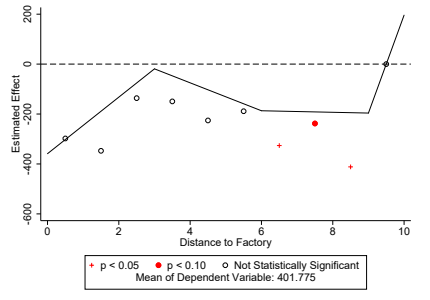
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



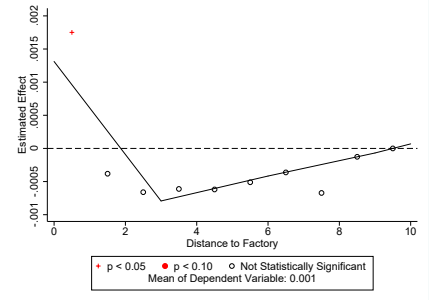
(c) Tons of Cane Grown (Full Sample, PODES 2003)



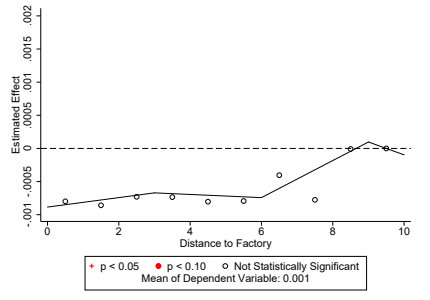
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



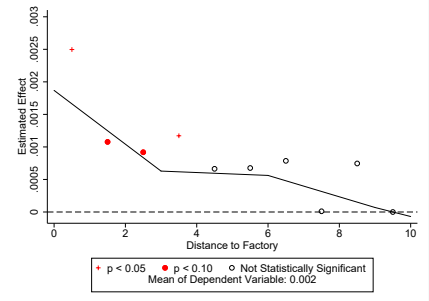
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



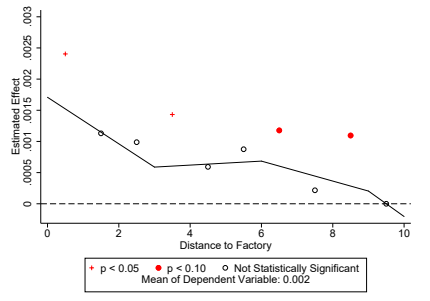
(f) Emp Share Upstream (No Modern Factories, Economic Census 2006)



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



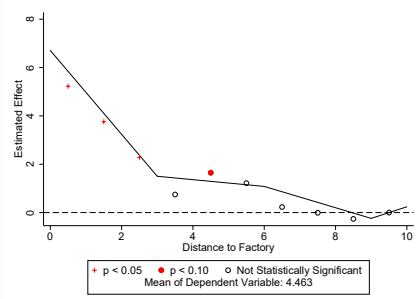
(h) Emp Share Downstream (No Modern Factories, Economic Census 2006)



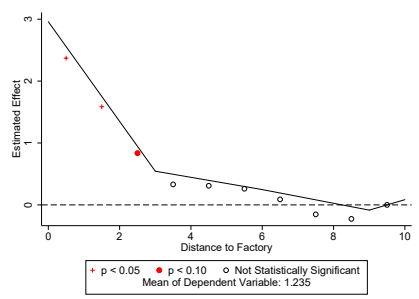
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure H-5: Infrastructure

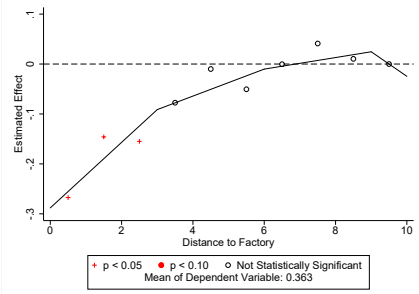
(a) Colonial Road Density (1900)



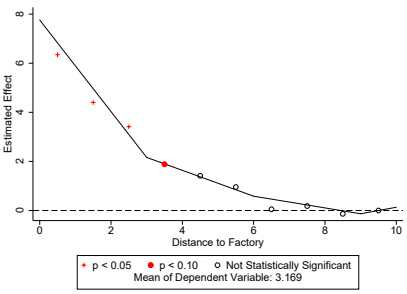
(b) Colonial Railroad Density (1900)



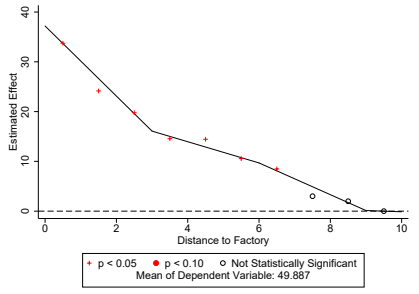
(c) Dirt Road (PODES 1980)



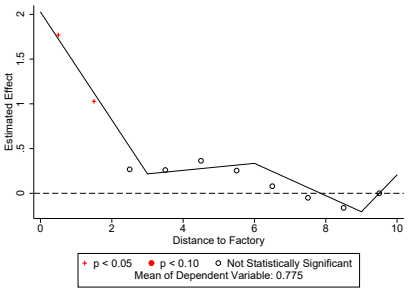
(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



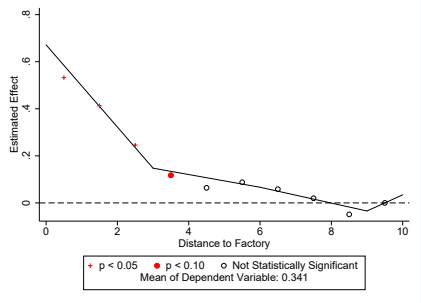
(f) Railroad Density (2017)



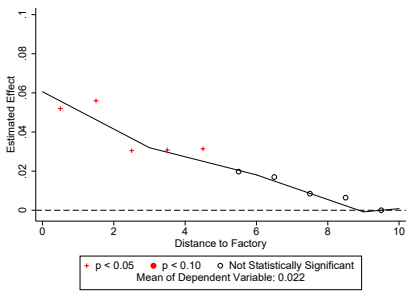
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure H-6: Other Public Goods

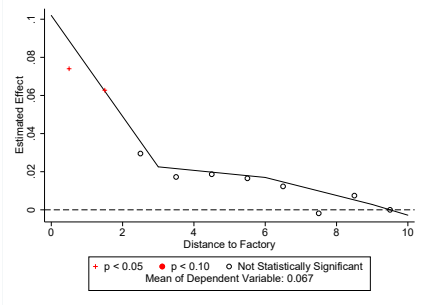
(a) Village Has Electricity (PODES 1980)



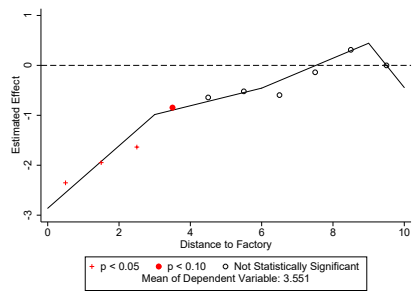
(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



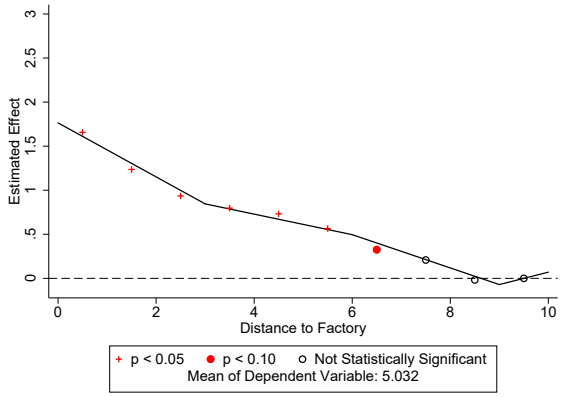
(d) Distance to Subdistrict Capital (2011 PODES)



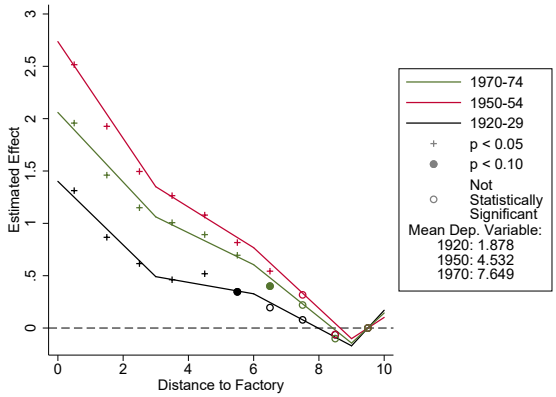
Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Panel c) includes survey year fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure H-7: Education

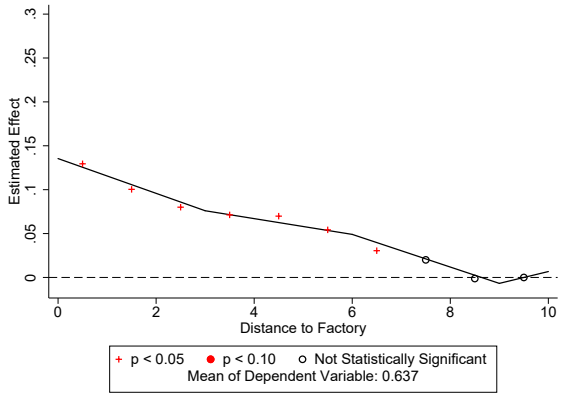
(a) Years Education (2000 Census)



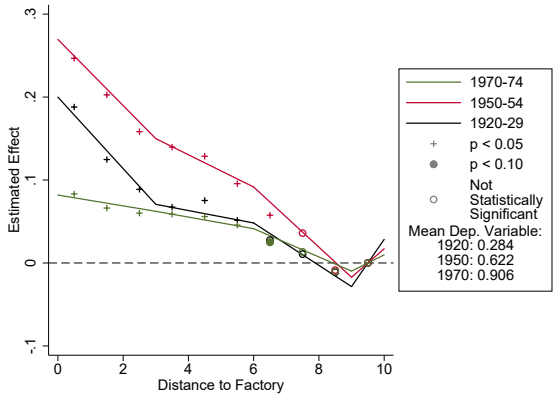
(b) Years Education by Cohort



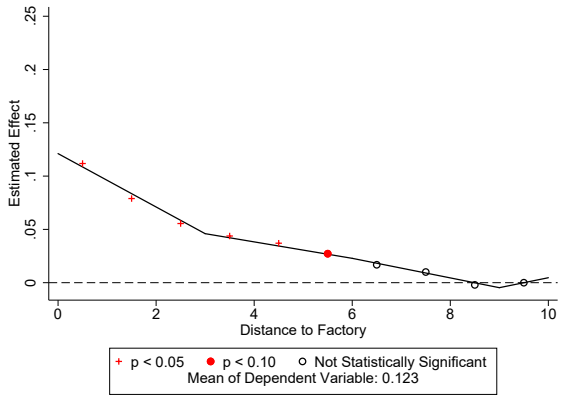
(c) Primary (2000 Census)



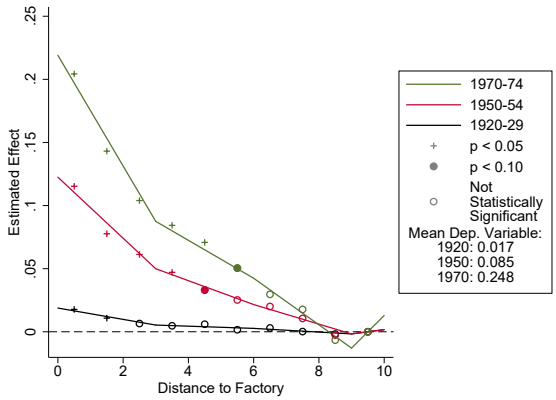
(d) Primary by Cohort



(e) High School (2000 Census)

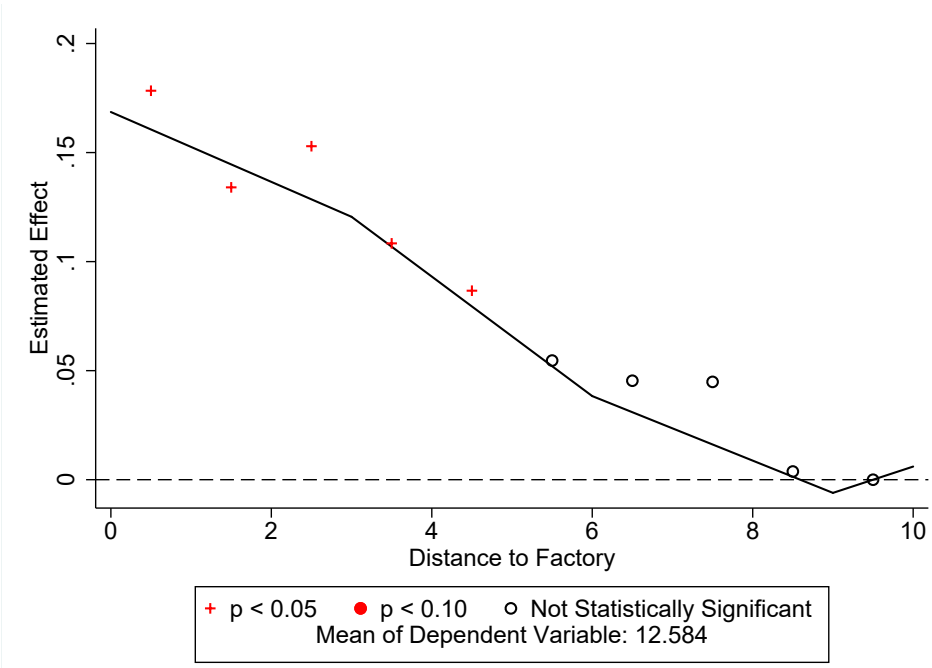


(f) High School by Cohort



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, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Left panels pool all birth cohorts and right panels plot separate coefficients for three birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

Figure H-8: Expenditure (2001-11)

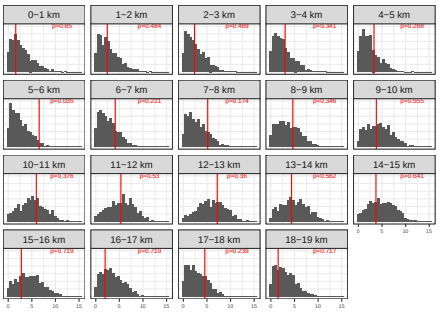


Notes: This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

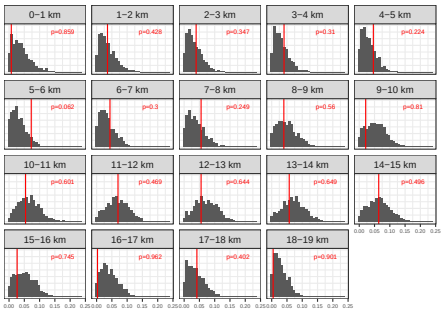
I Independent Shifts Distributions

Figure I-1: Geography

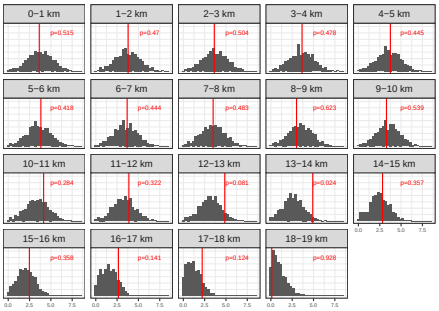
(a) Elevation



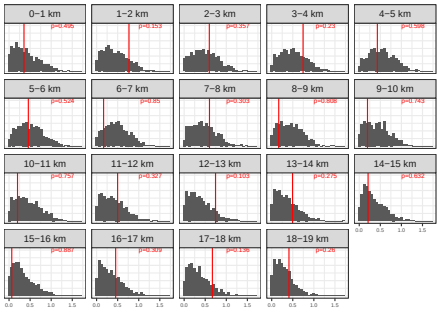
(b) Slope



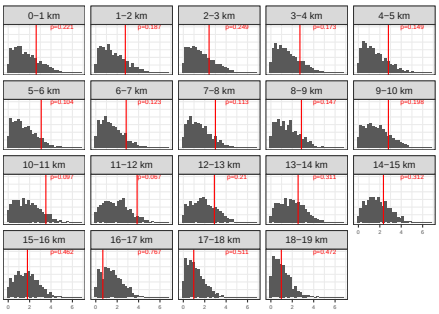
(c) Distance to Coast



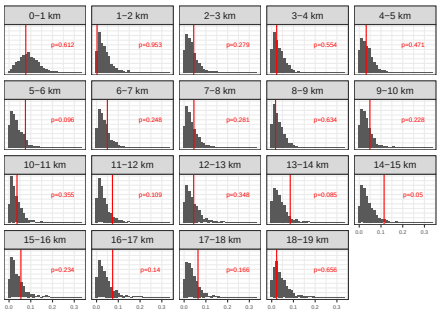
(d) Log Flow Accumulation



(e) Distance to Natural Harbor



(f) Distance to River



(g) Distance to Nearest 1830 Residency Capital

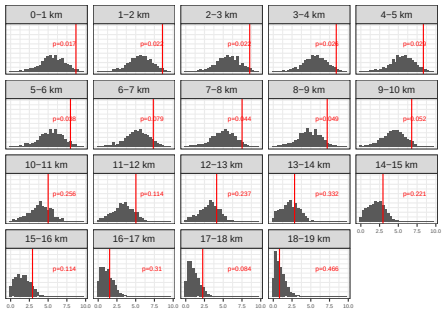
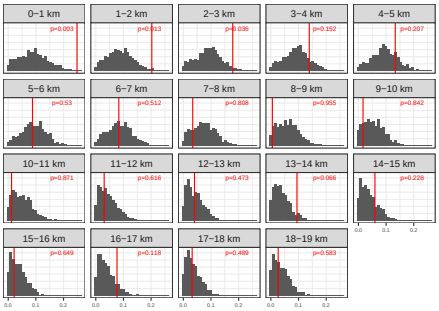
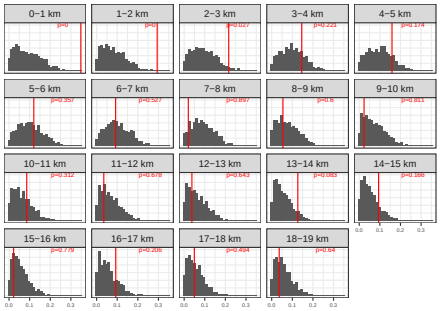


Figure I-2: Industry and Agglomeration

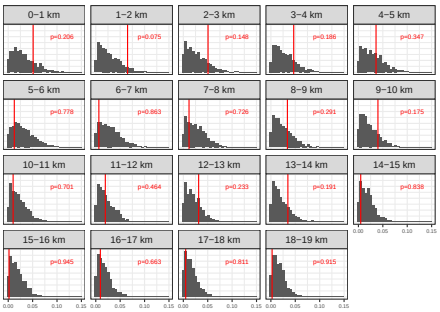
(a) Agriculture (Susenas 2001-11)



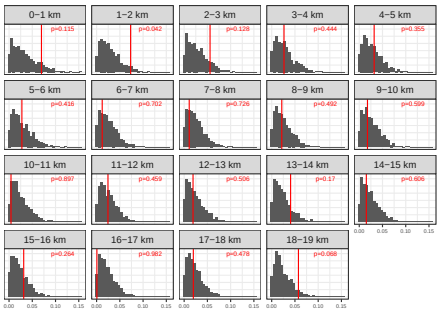
(b) Agriculture (Census 1980)



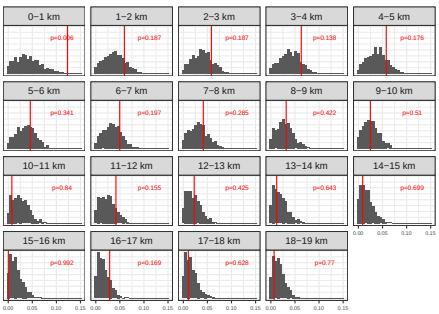
(c) Manufacturing (Susenas 2001-11)



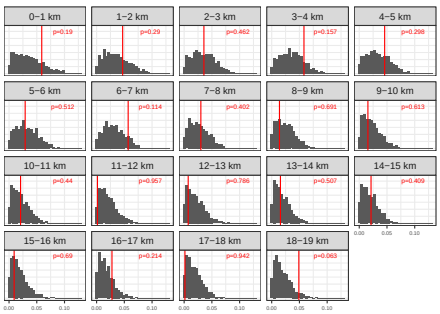
(d) Manufacturing (Census 1980)



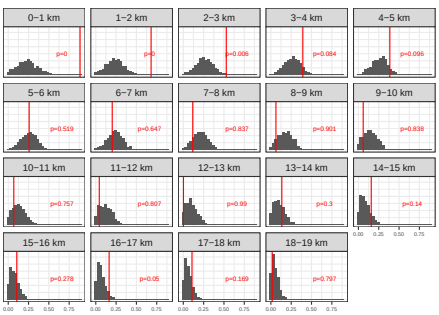
(e) Retail (Susenas 2001-11)



(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



(h) Log Pop. Density (PODES 1980)

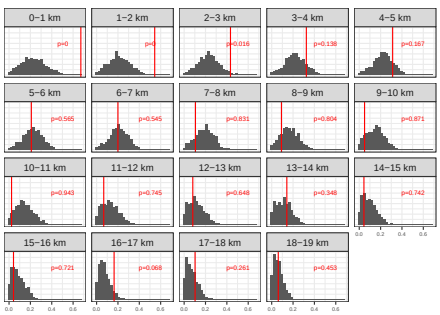
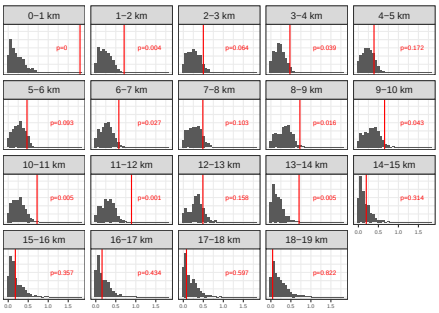
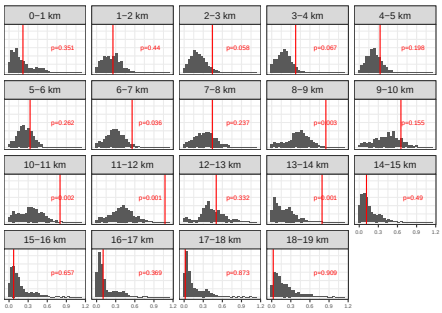


Figure I-3: Sugar and Linked Industries

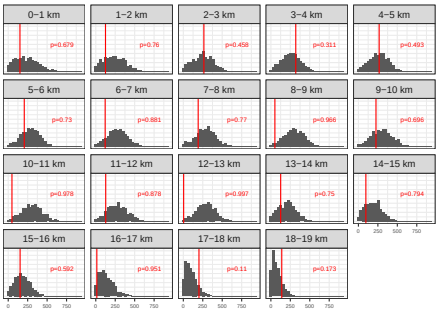
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



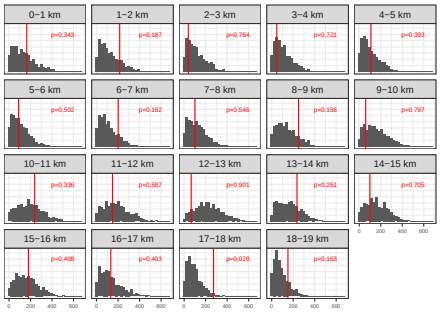
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



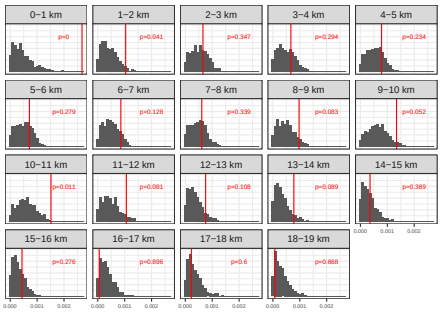
(c) Tons of Cane Grown (Full Sample, PODES 2003)



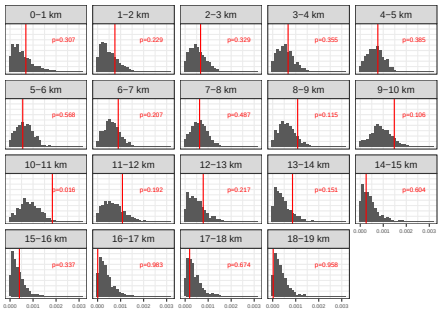
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



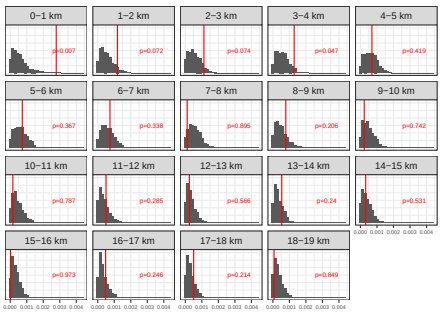
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



(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)

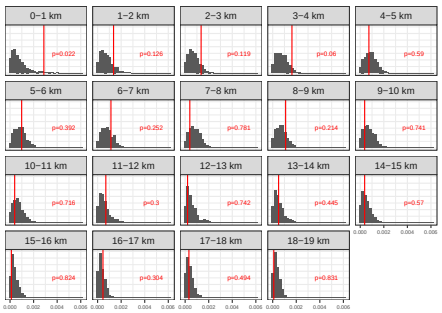
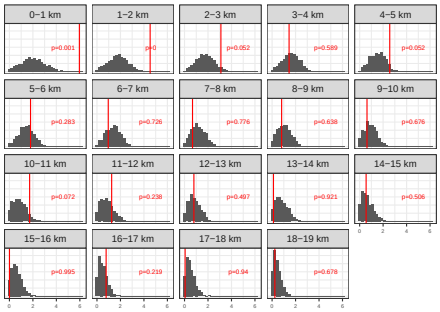
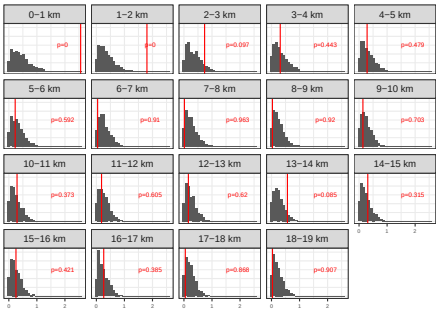


Figure I-4: Infrastructure

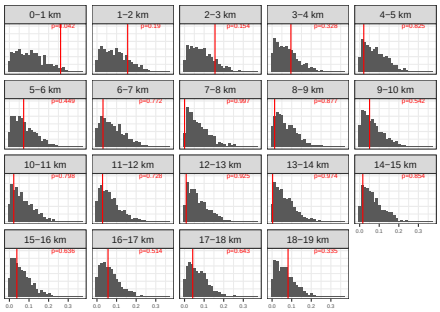
(a) Colonial Road Density (1900)



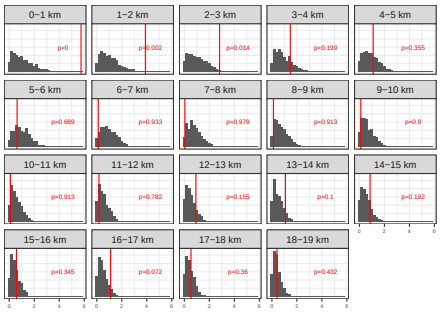
(b) Colonial Railroad Density (1900)



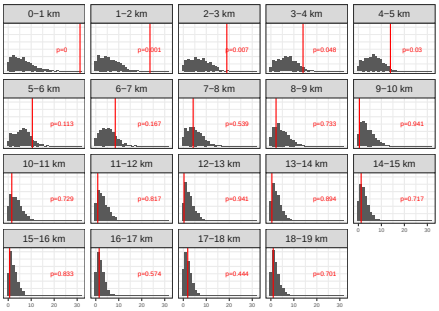
(c) Dirt Road (PODES 1980)



(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



(f) Railroad Density (2017)

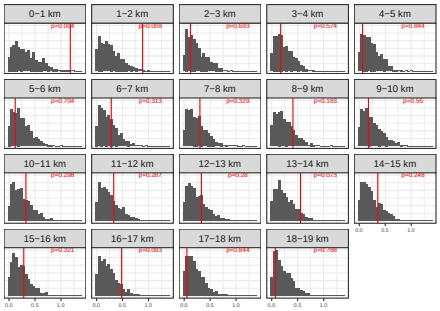
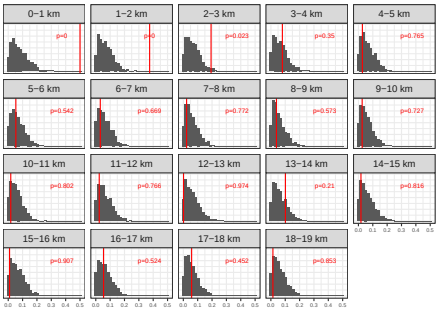
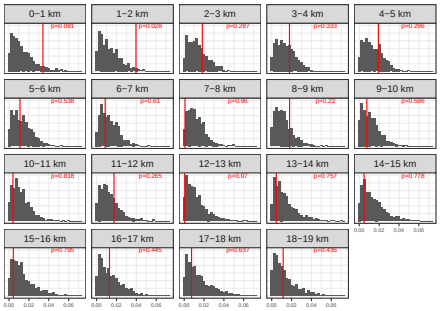


Figure I-5: Other Public Goods

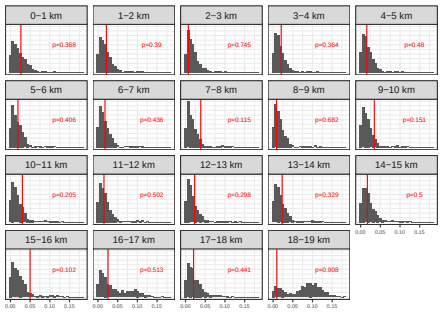
(a) Village Has Electricity (PODES 1980)



(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



(d) Distance to Subdistrict Capital (2011 PODES)

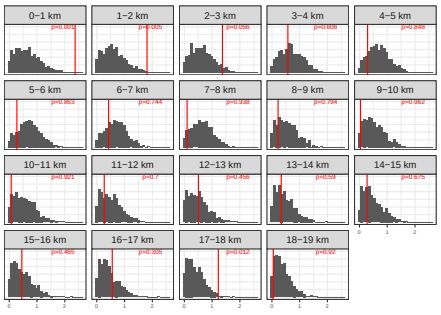
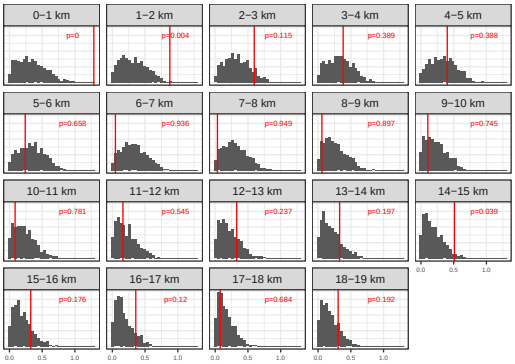
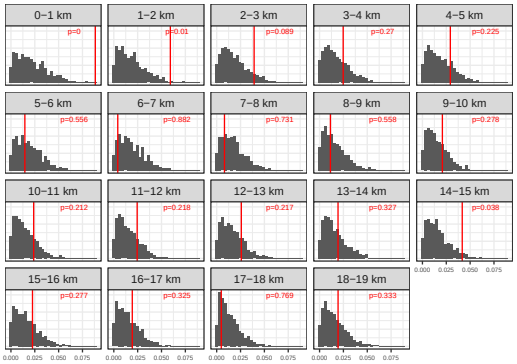


Figure I-6: Education

(a) Years Education



(b) Primary



(c) High School

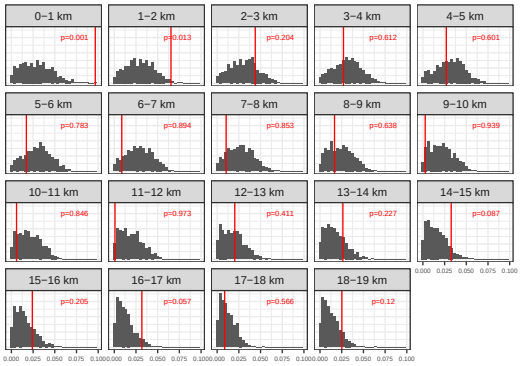
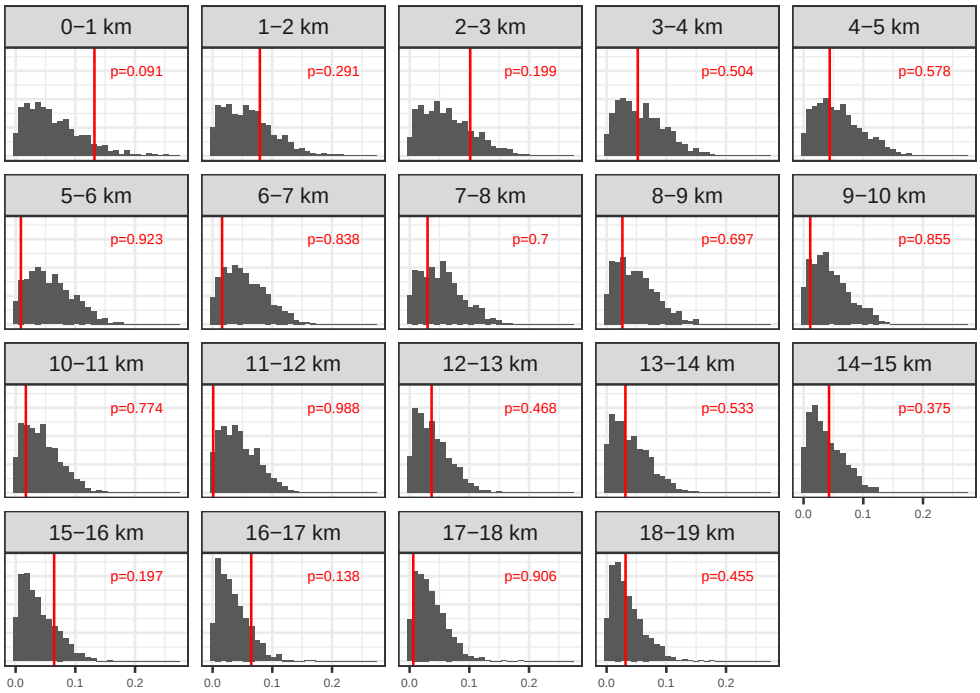


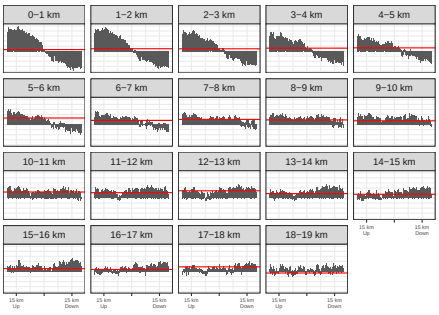
Figure I-7: Expenditure (2001-11)



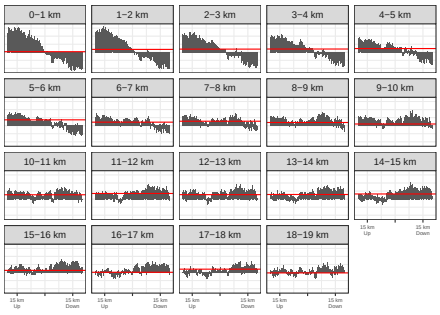
J Common Shifts Distributions

Figure J-1: Geography

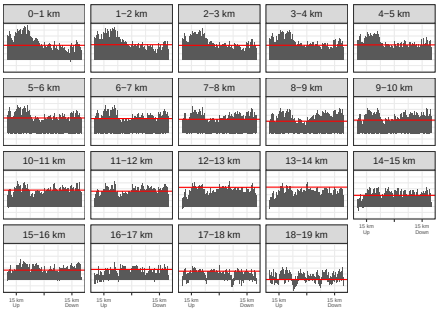
(a) Elevation



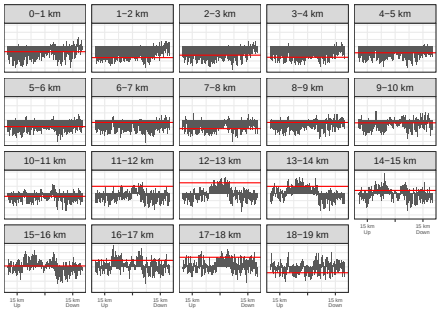
(b) Slope



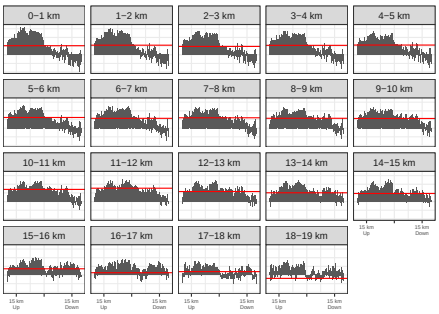
(c) Distance to Coast



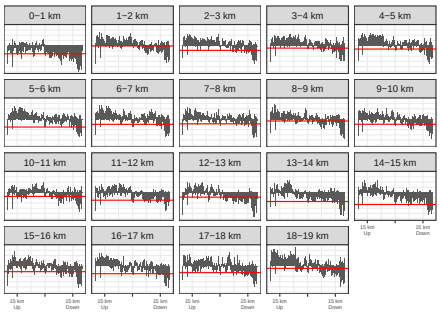
(d) Log Flow Accumulation



(e) Distance to Natural Harbor



(f) Distance to River



(g) Distance to Nearest 1830 Residency Capital

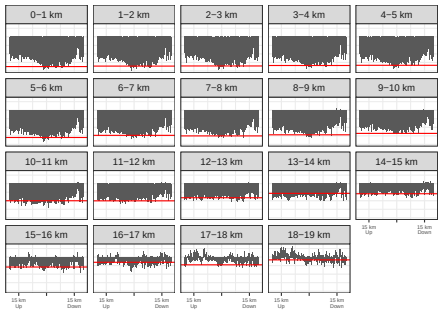
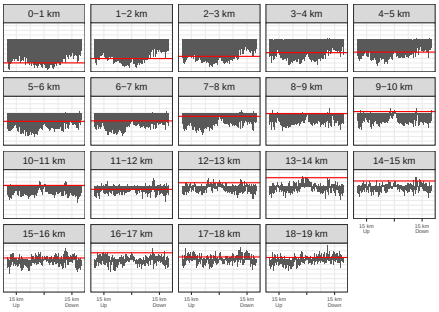
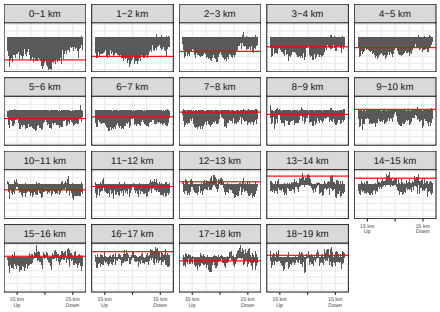


Figure J-2: Industry and Agglomeration

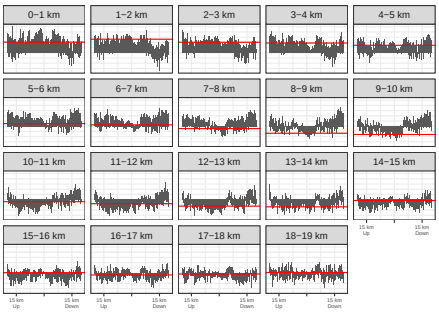
(a) Agriculture (Susenas 2001-11)



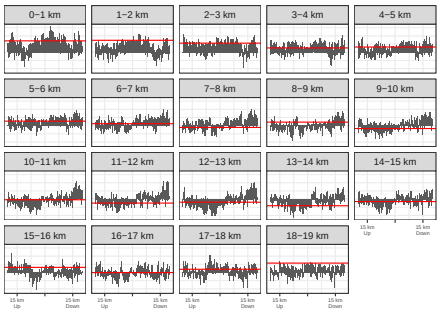
(b) Agriculture (Census 1980)



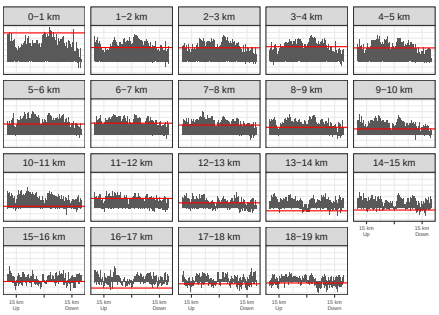
(c) Manufacturing (Susenas 2001-11)



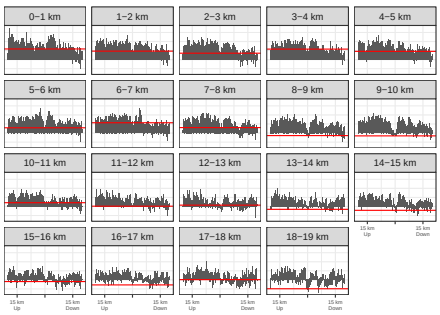
(d) Manufacturing (Census 1980)



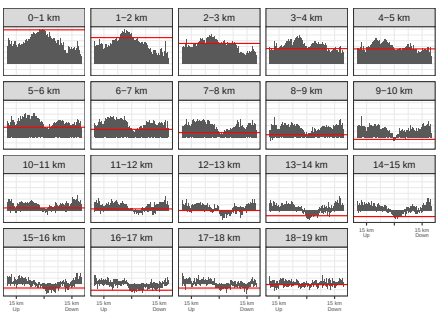
(e) Retail (Susenas 2001-11)



(f) Retail (Census 1980)



(g) Log Pop. Density (PODES 2003)



(h) Log Pop. Density (PODES 1980)

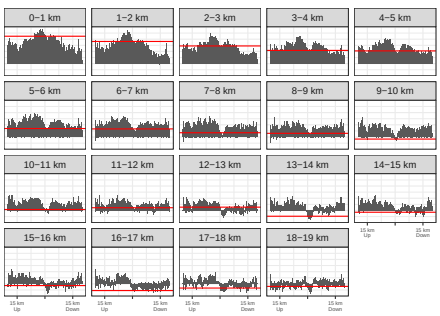
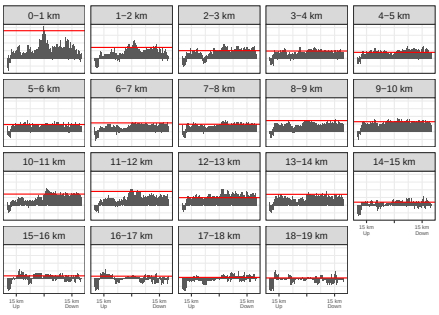
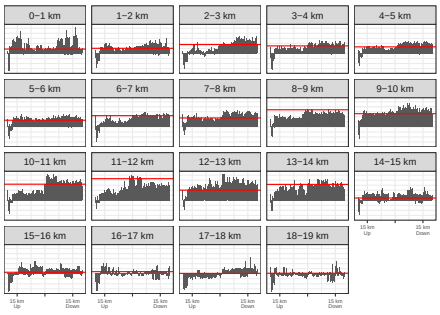


Figure J-3: Sugar and Linked Industries

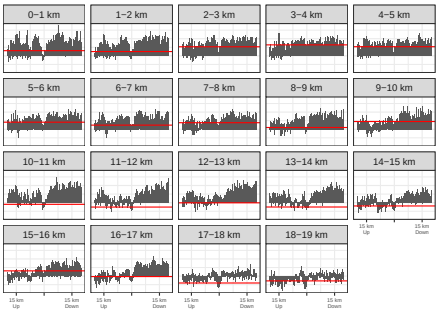
(a) Log Value Sugar Processed (Full Sample, Economic Census 2006)



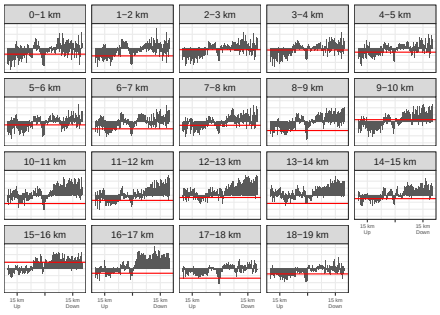
(b) Log Value Sugar Processed (No Modern Factories, Econ Census 2006)



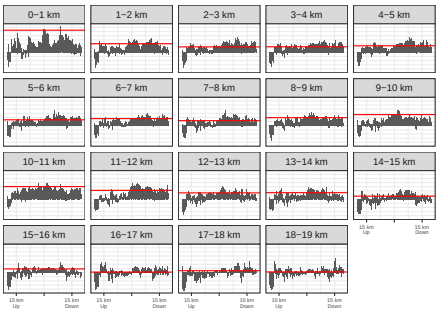
(c) Tons of Cane Grown (Full Sample, PODES 2003)



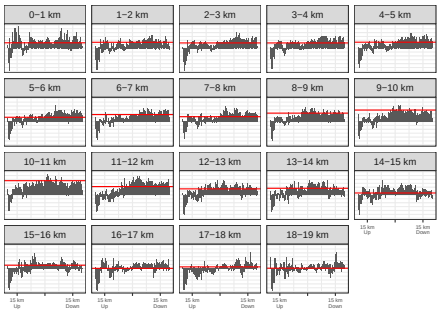
(d) Tons of Cane Grown (No Modern Factories, PODES 2003)



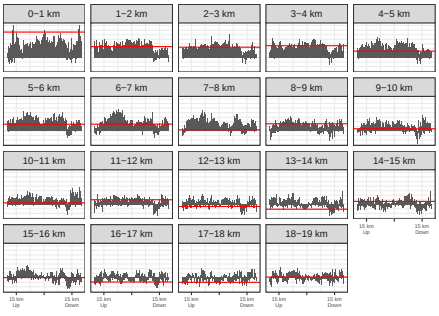
(e) Employment Share Upstream (Full Sample, Economic Census 2006)



(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)

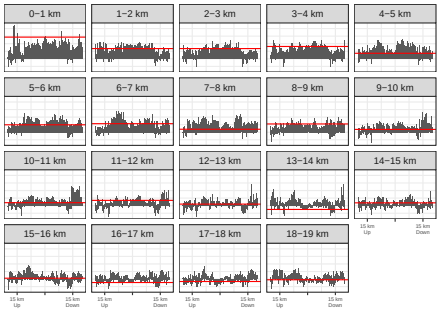
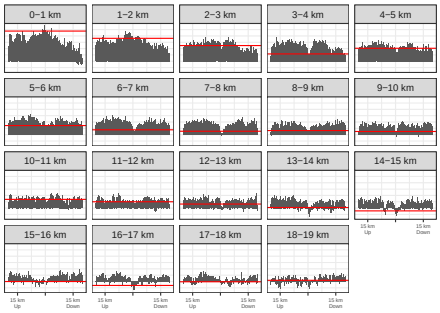
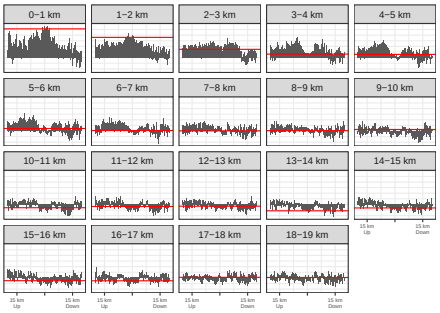


Figure J-4: Infrastructure

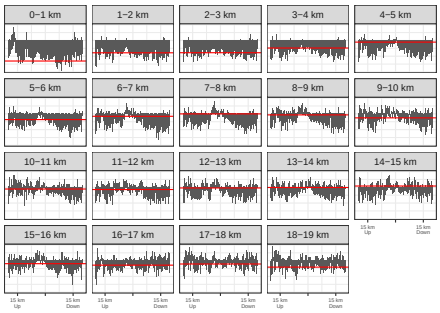
(a) Colonial Road Density (1900)



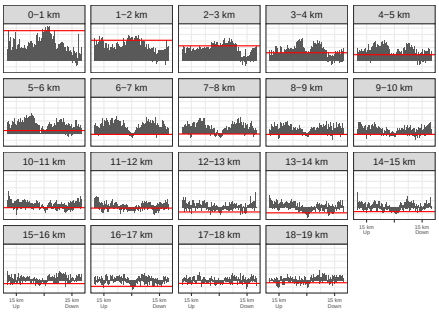
(b) Colonial Railroad Density (1900)



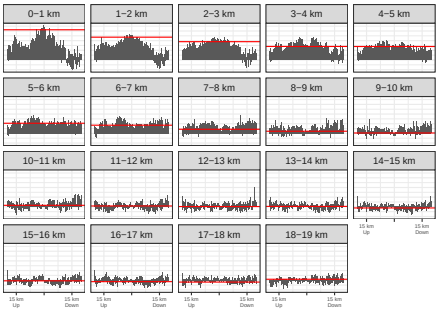
(c) Dirt Road (PODES 1980)



(d) Intercity Road Density (2017)



(e) Local Road Density (2017)



(f) Railroad Density (2017)

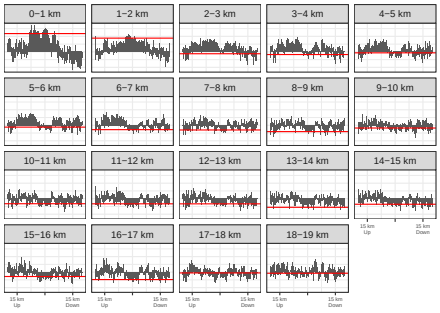
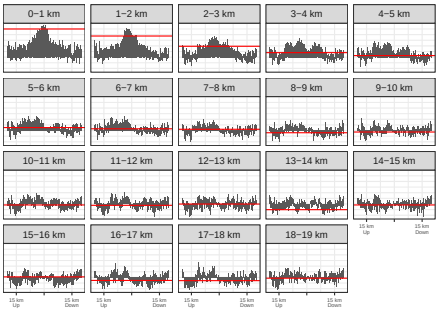
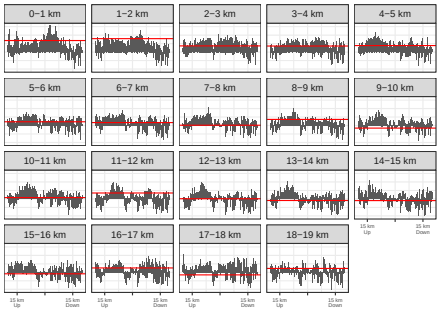


Figure J-5: Other Public Goods

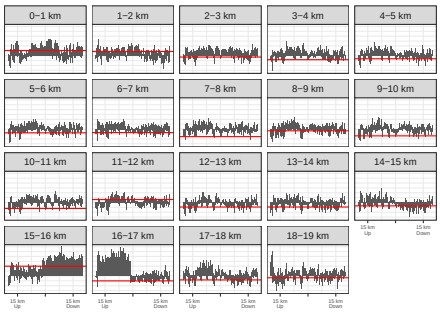
(a) Village Has Electricity (PODES 1980)



(b) High Schools (PODES 1980)



(c) High Schools (PODES 1996-2011)



(d) Distance to Subdistrict Capital (2011 PODES)

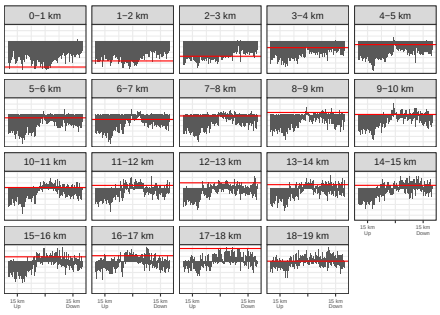
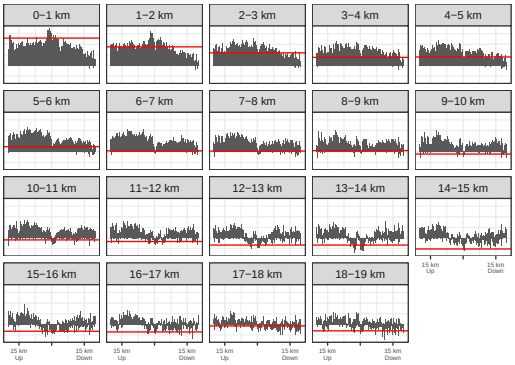
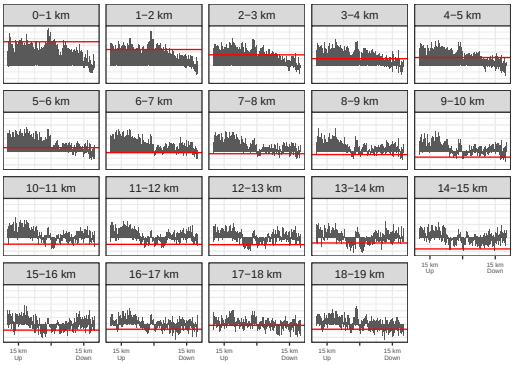


Figure J-6: Education

(a) Years Education



(b) Primary



(c) High School

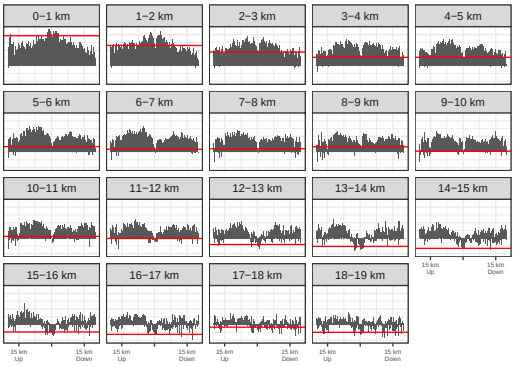
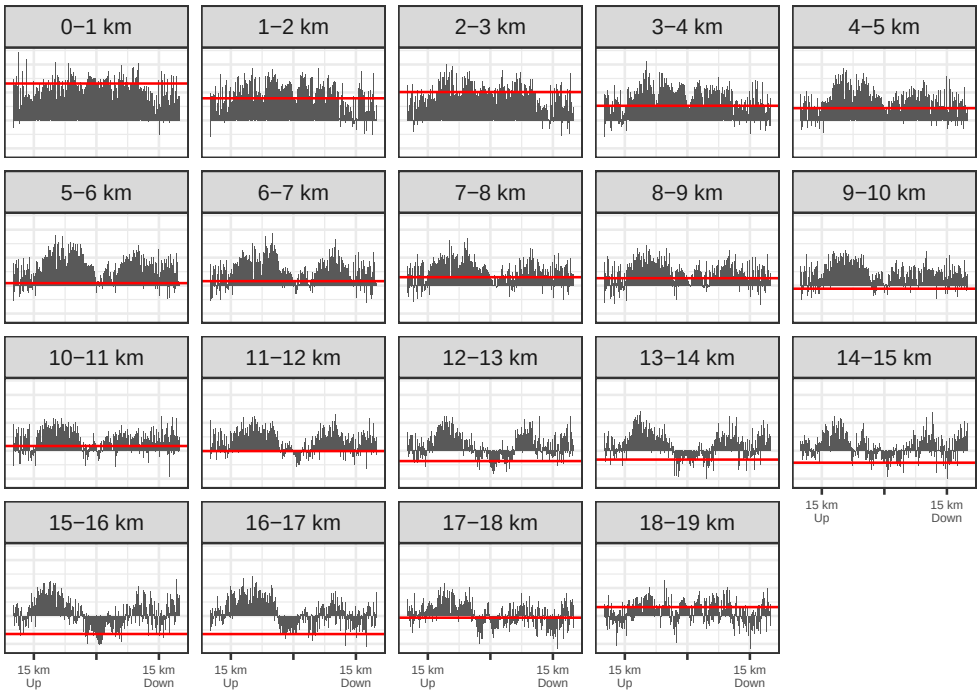


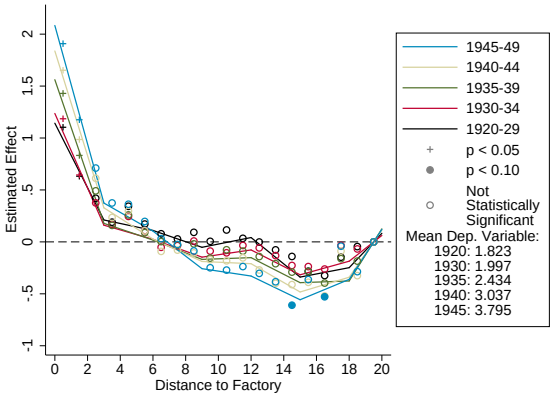
Figure J-7: Expenditure (2001-11)



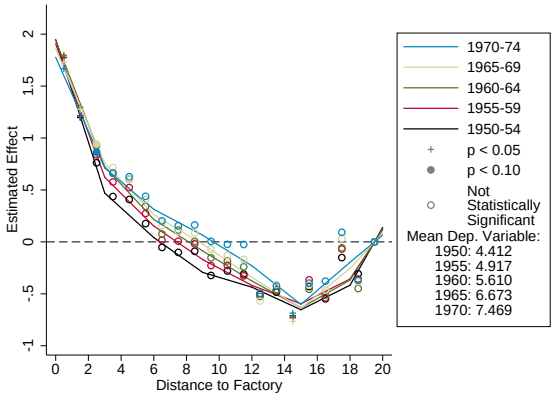
K Education for All Cohorts

Figure K-1: Education

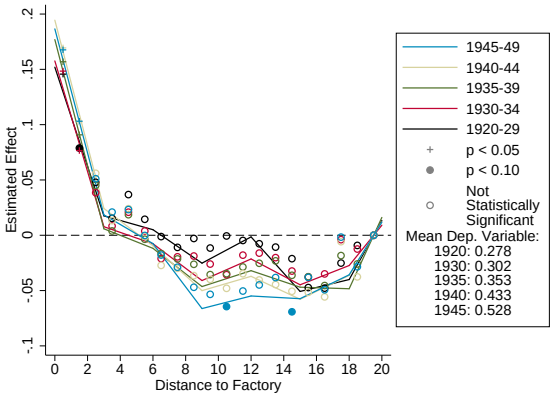
(a) Years Education: 1920-1949



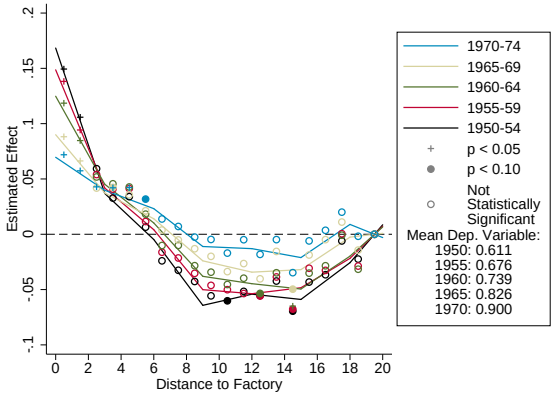
(b) Years Education: 1950-1974



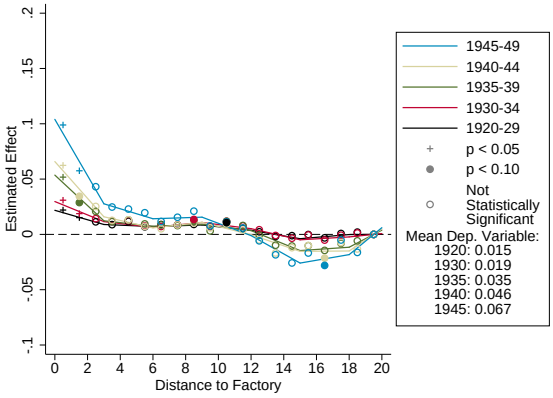
(c) Primary: 1920-1949



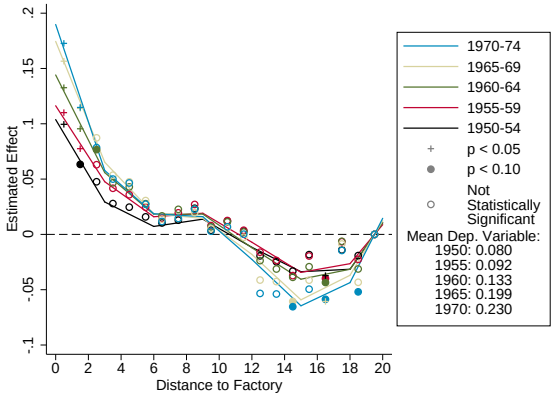
(d) Primary: 1950-1974



(e) High School: 1920-1949



(f) High School: 1950-1974

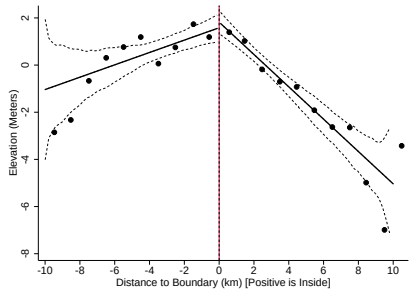


Notes: These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Each panel plots separate coefficients for five birth cohorts. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 1,000 counterfactual factory locations.

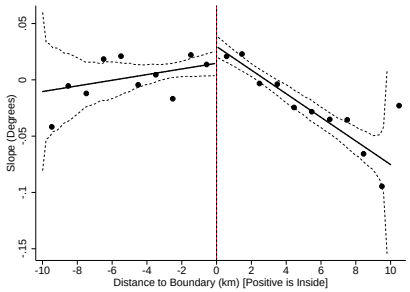
L RD Plots

Figure L-1: Geographic Characteristics

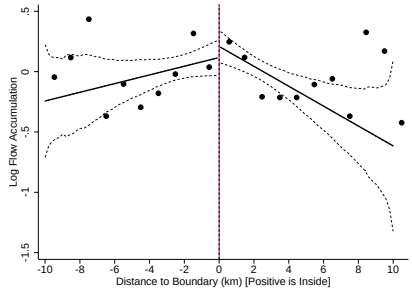
(a) Elevation



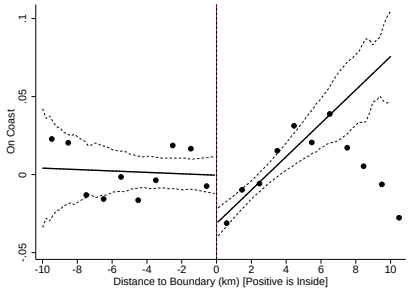
(b) Slope



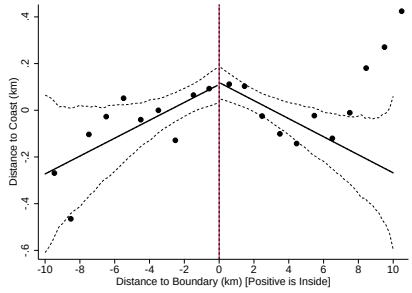
(c) Log Flow Accumulation



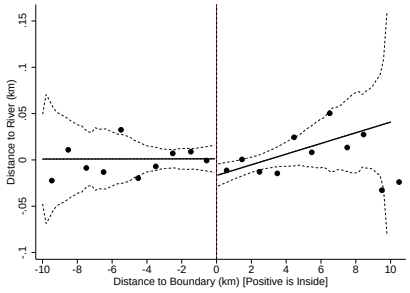
(d) On Coast



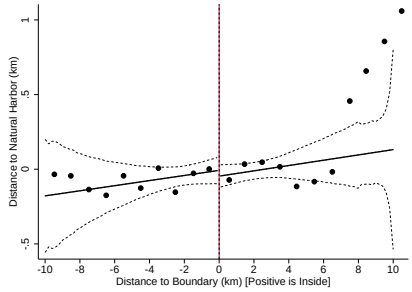
(e) Distance to Coast



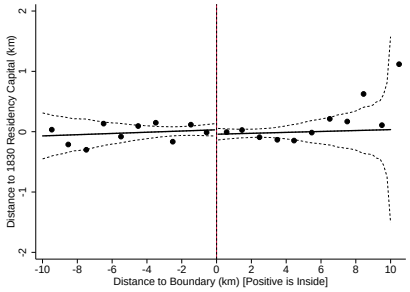
(f) Distance to River



(g) Distance to Natural Harbor



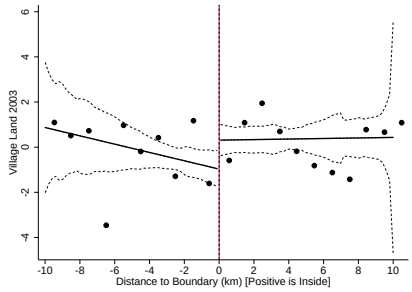
(h) Distance to 1830 Residency Capital



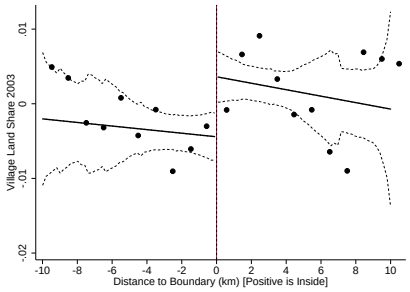
Notes: Points are binned residuals from a regression of the outcome variable on boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. Solid lines fit a local linear regression,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

Figure L-2: Land Tenure

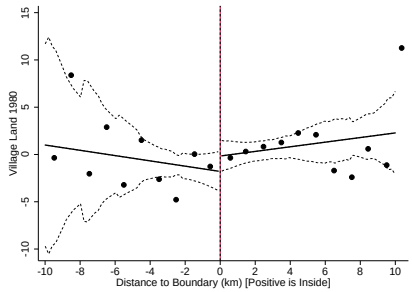
(a) Village Head 2003: Total Land



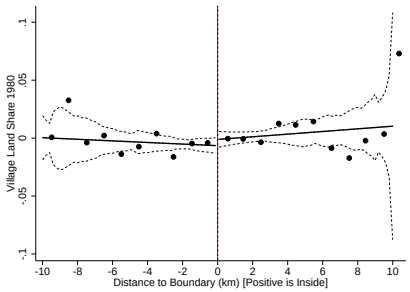
(b) Village Head 2003: Land Share



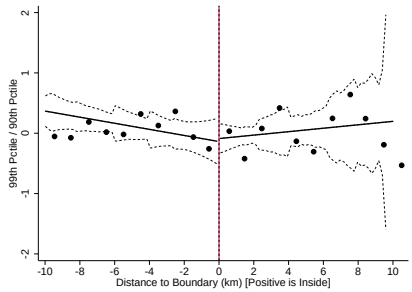
(c) Village Head 1980: Total Land



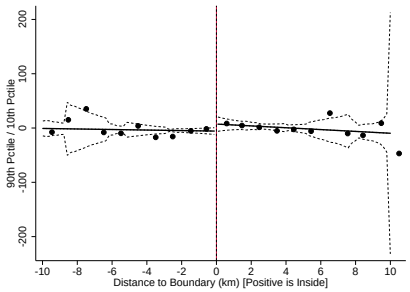
(d) Village Head 1980: Land Share



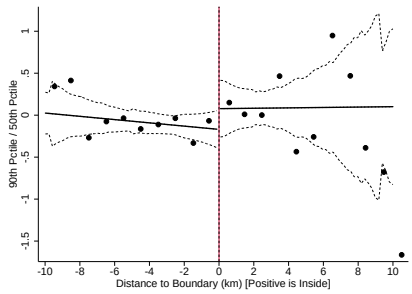
(e) 99th Pctile / 90th Pctile



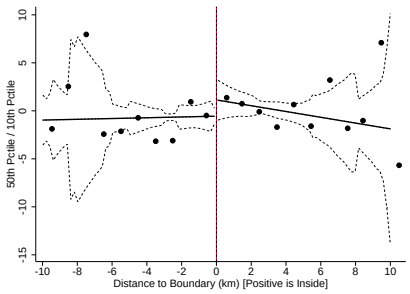
(f) 90th Pctile / 10th Pctile



(g) 90th Pctile / 50th Pctile



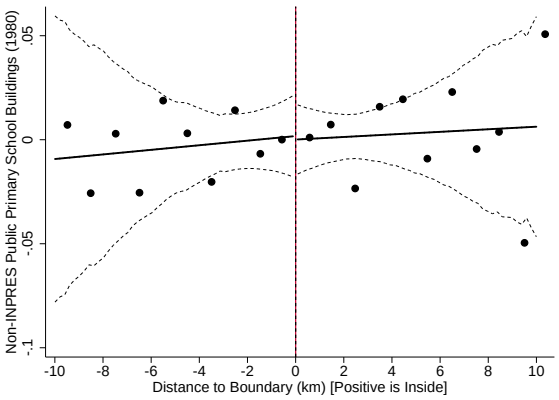
(h) 50th Pctile / 10th Pctile



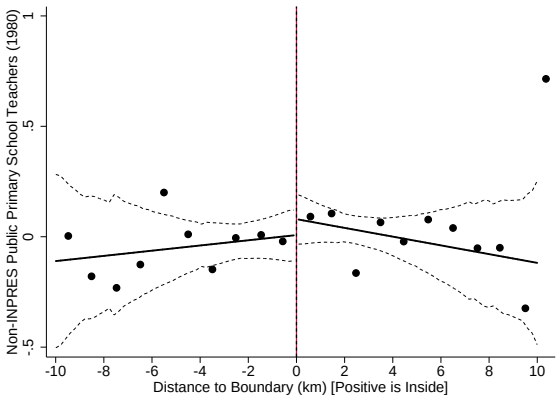
Notes: Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest residency capital. 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 L-3: Schools (1980)

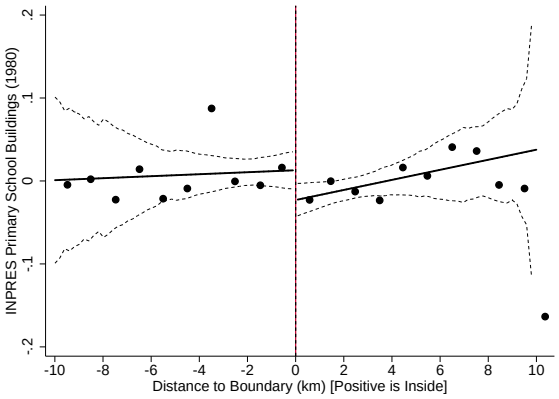
(a) Public Non-INPRES Primary Buildings



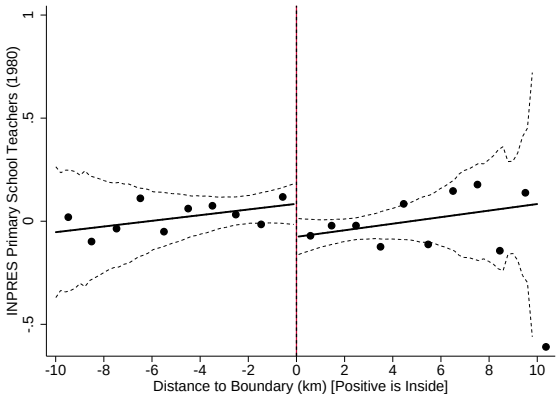
(b) Public Non-INPRES Primary Teachers



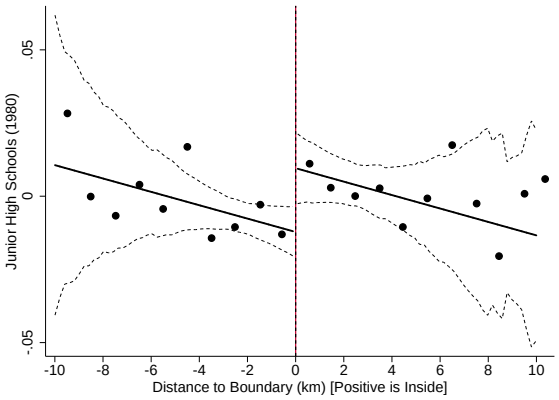
(c) INPRES Primary Buildings



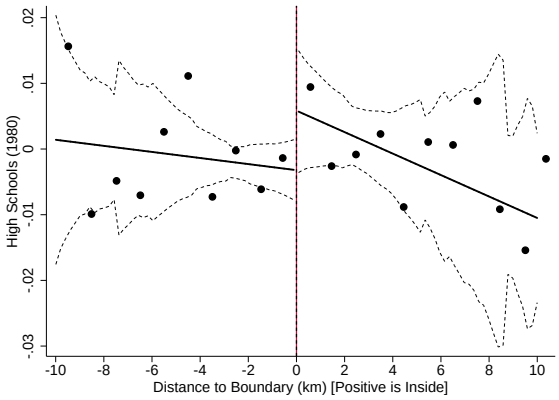
(d) INPRES Primary Teachers



(e) Junior High Schools



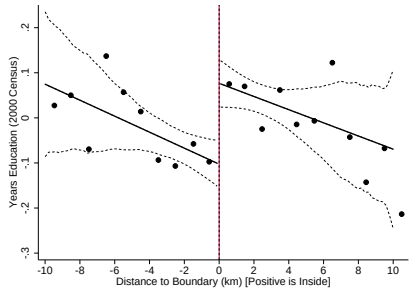
(f) High Schools



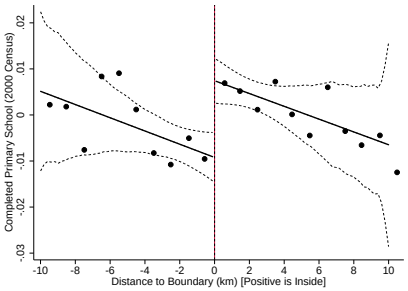
Notes: Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest residency capital. 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 L-4: Education

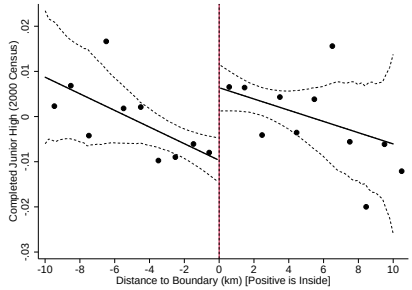
(a) Years Education 2000



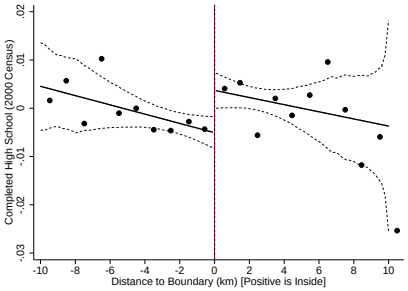
(b) Primary School 2000



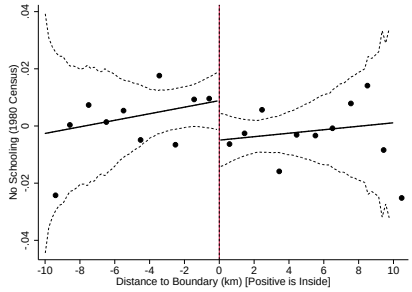
(c) Junior High 2000



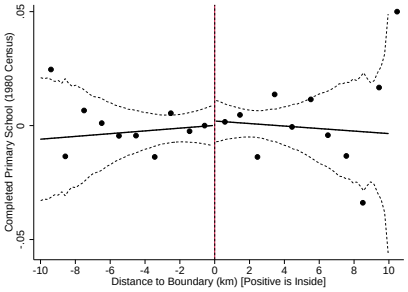
(d) High School 2000



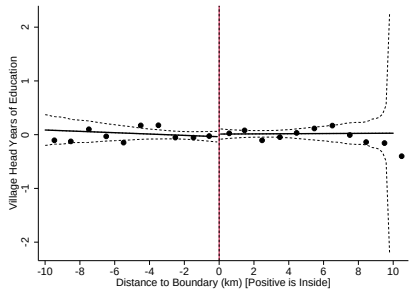
(e) No School 1980



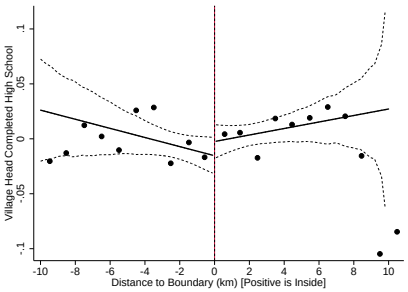
(f) Primary School 1980



(g) Village Head Years Education



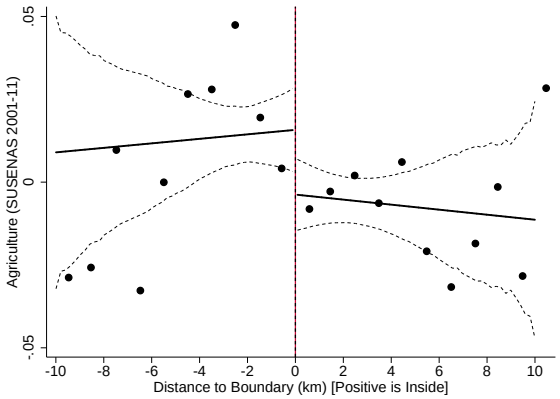
(h) Village Head High School



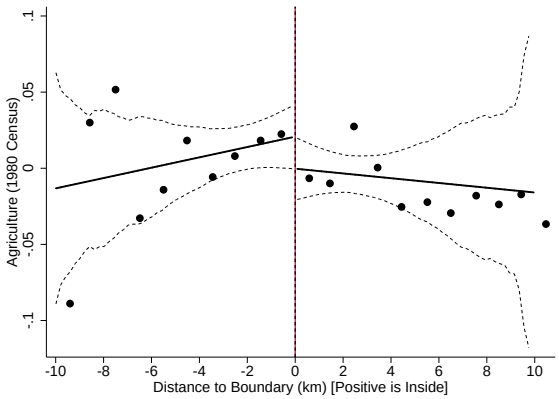
Notes: Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest residency capital. 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 L-5: Industrial Structure

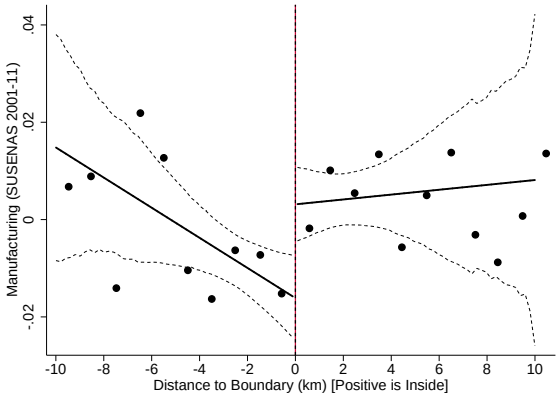
(a) Agriculture (SUSENAS)



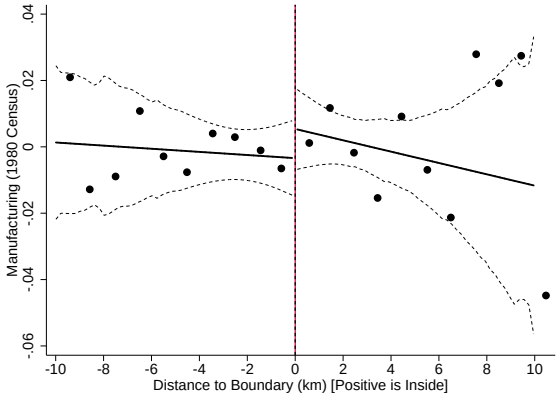
(b) Agriculture (1980 Census)



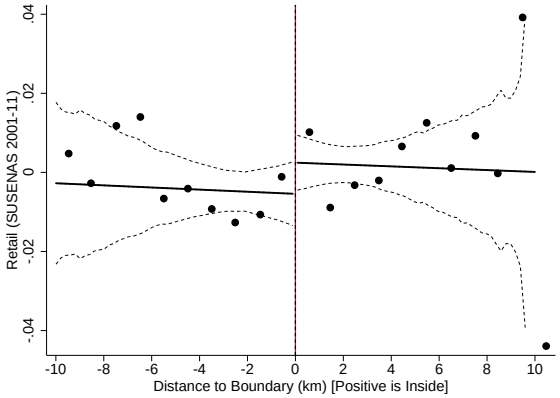
(c) Manufacturing (SUSENAS)



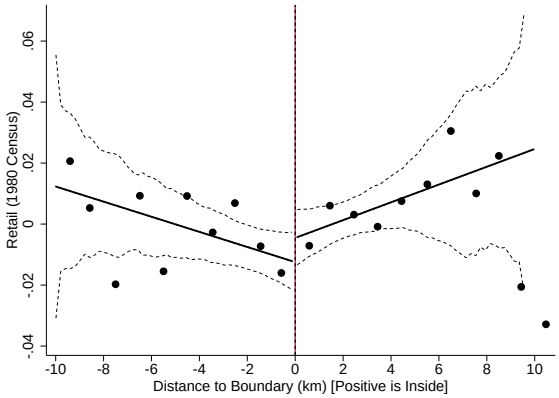
(d) Manufacturing (1980 Census)



(e) Retail (SUSENAS)



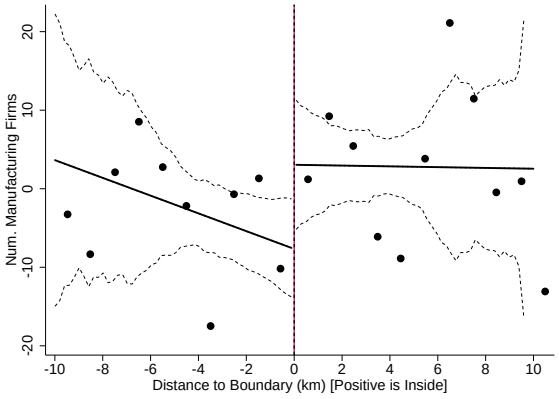
(f) Commerce (1980 Census)



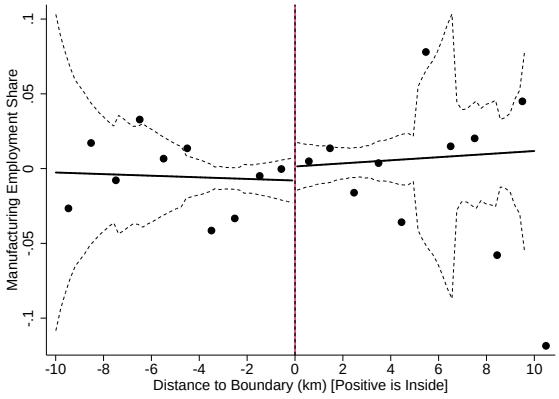
Notes: Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest residency capital. The sample is restricted to men aged 18-55. Left panels 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 L-6: Firms, Population, and Consumption

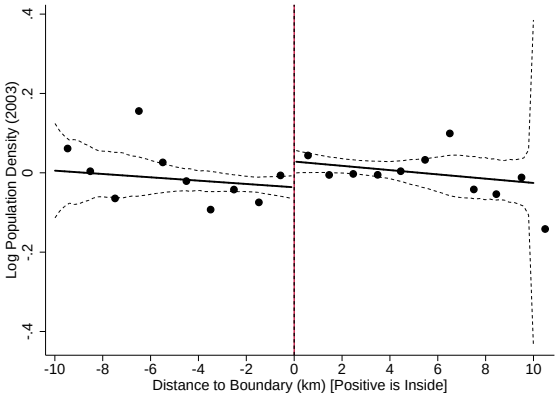
(a) Number of Firms



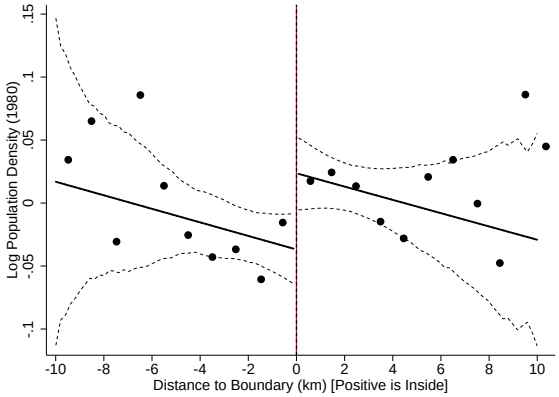
(b) Employment Share



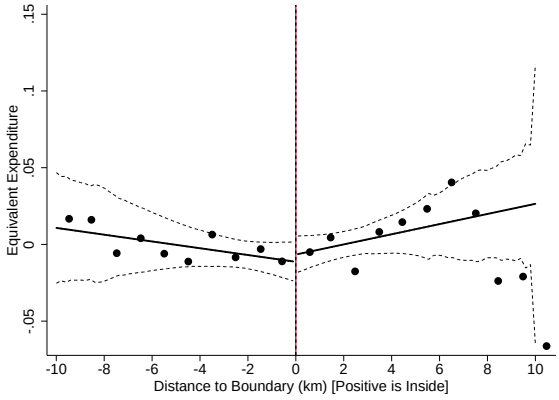
(c) Log Population Density (2003)



(d) Log Population Density (1980)



(e) Log Equivalent Consumption

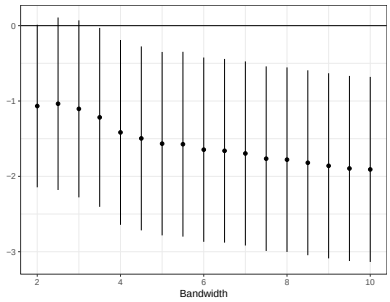


Notes: Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest residency capital. Panel (e) includes demographic controls and 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.

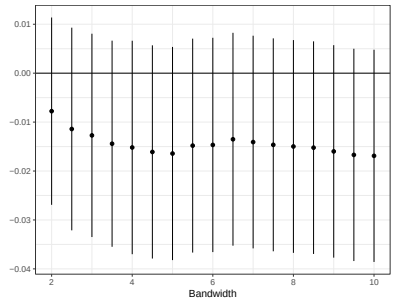
M Robustness to Bandwidth Plots

Figure M-1: Geographic Characteristics

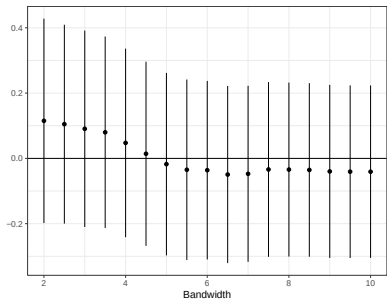
(a) Elevation



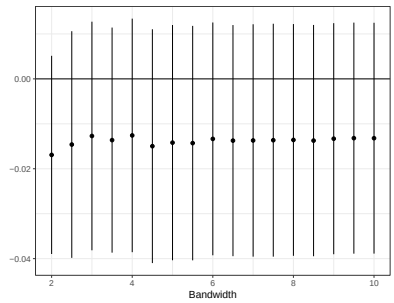
(b) Slope



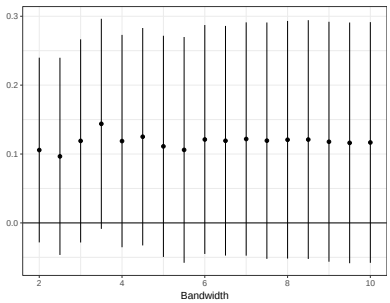
(c) Log Flow Accumulation



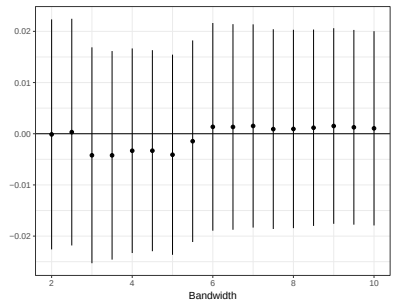
(d) On Coast



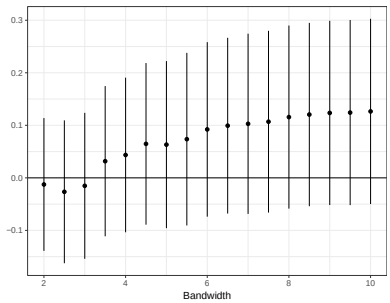
(e) Distance to Coast



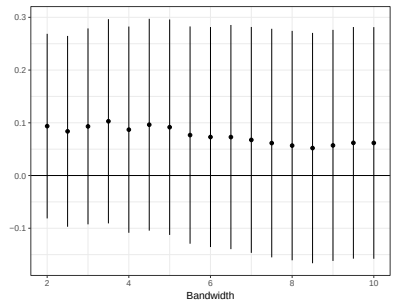
(f) Distance to River



(g) Distance to Natural Harbor



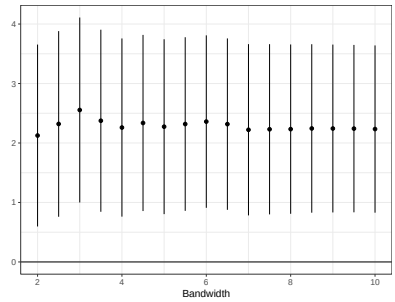
(h) Distance to 1830 Residency Capital



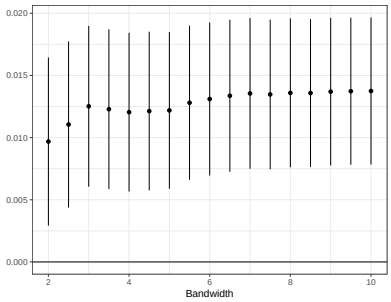
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for each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

Figure M-2: Land Tenure

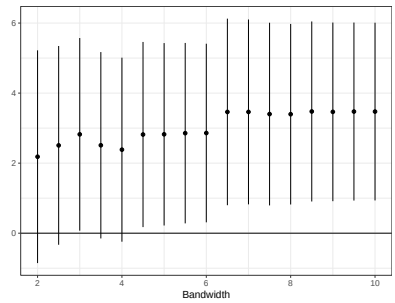
(a) Village Head 2003: Total Land



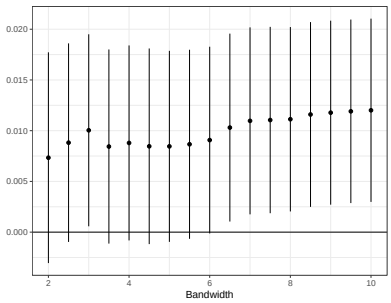
(b) Village Head 2003: Land Share



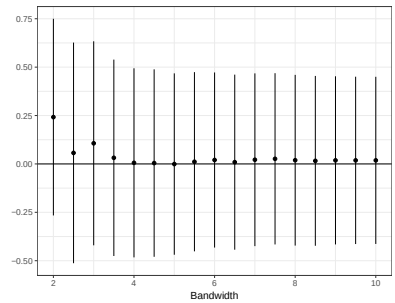
(c) Village Head 1980: Total Land



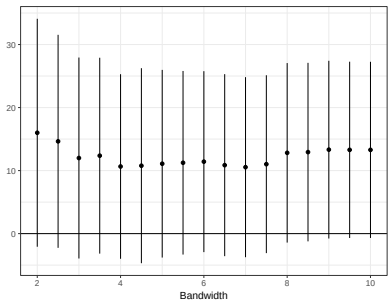
(d) Village Head 1980: Land Share



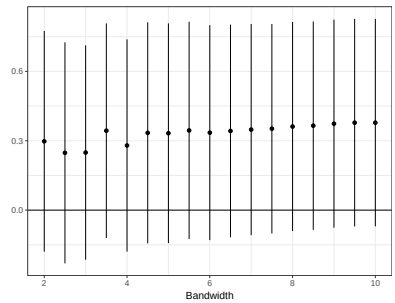
(e) 99th Pctile / 90th Pctile



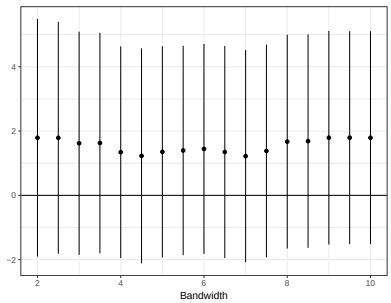
(f) 90th Pctile / 10th Pctile



(g) 90th Pctile / 50th Pctile



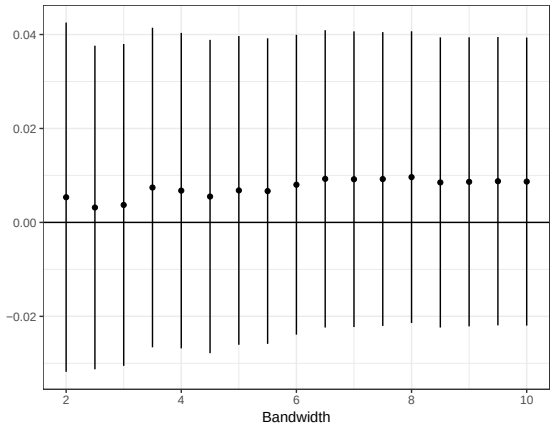
(h) 50th Pctile / 10th Pctile



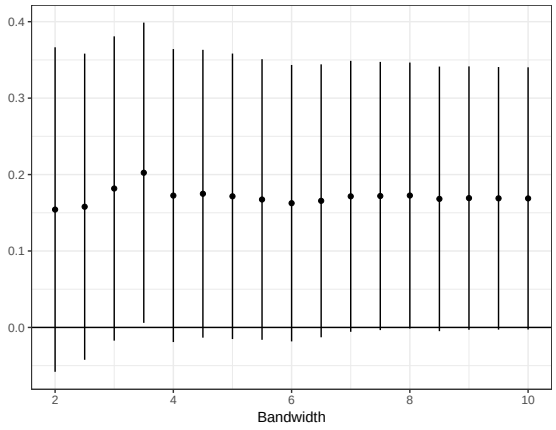
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks each 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 M-3: Schools (1980)

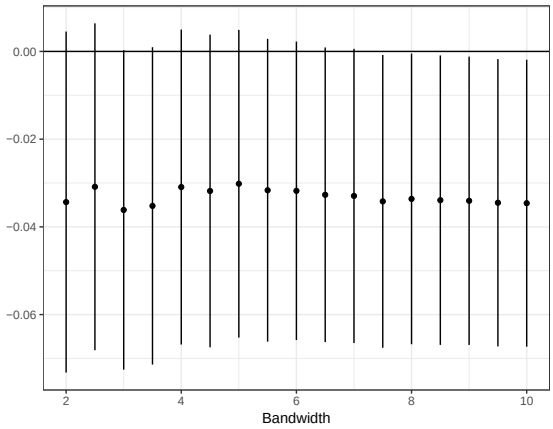
(a) Public Non-INPRES Primary Buildings



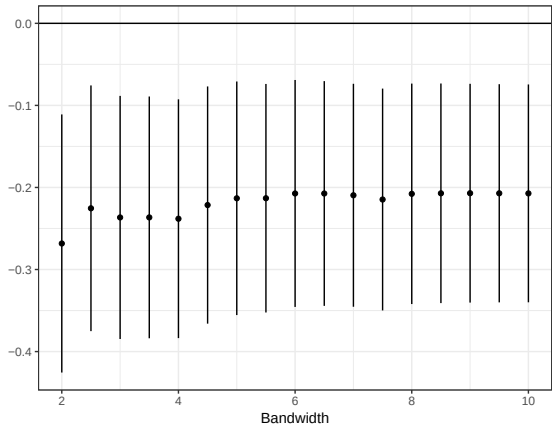
(b) Public Non-INPRES Primary Teachers



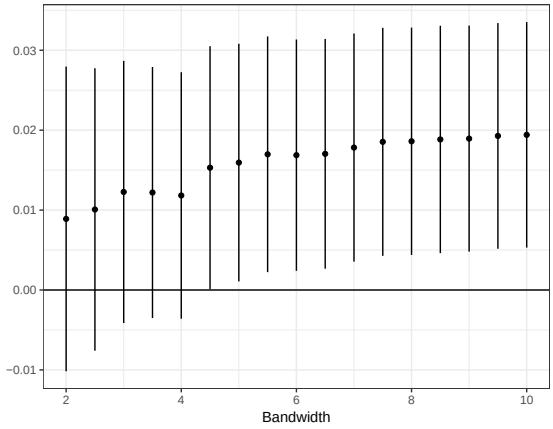
(c) INPRES Primary Buildings



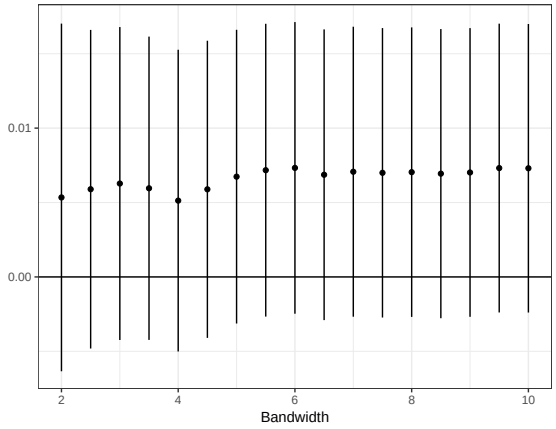
(d) INPRES Primary Teachers



(e) Junior High Schools



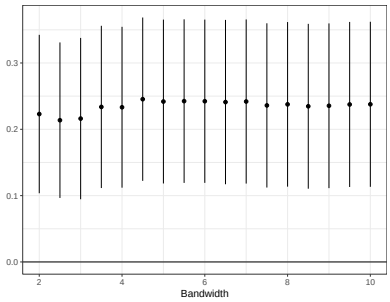
(f) High Schools



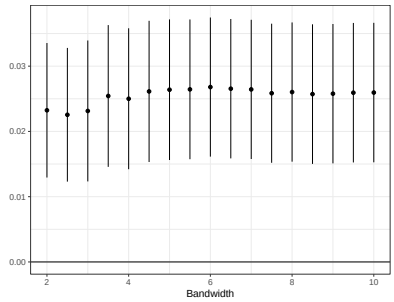
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks each 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 M-4: Education

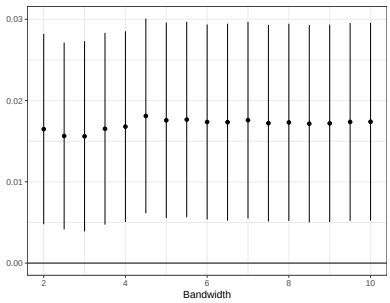
(a) Years Education 2000



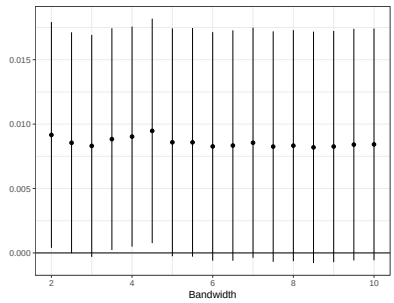
(b) Primary School 2000



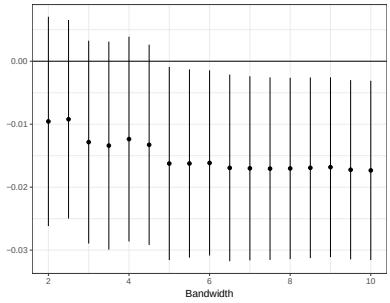
(c) Junior High 2000



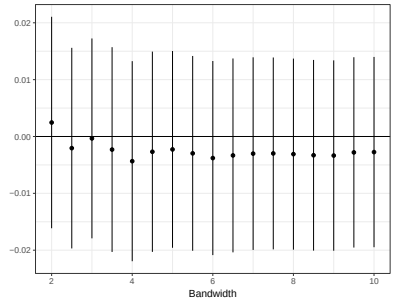
(d) High School 2000



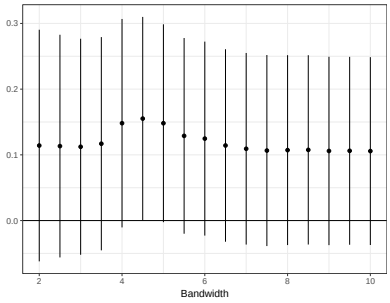
(e) No School 1980



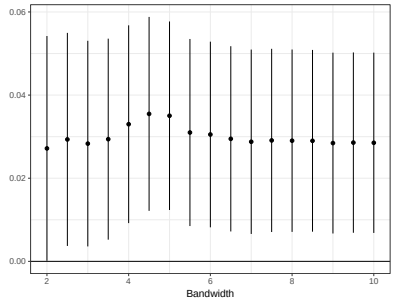
(f) Primary School 1980



(g) Village Head Years Education



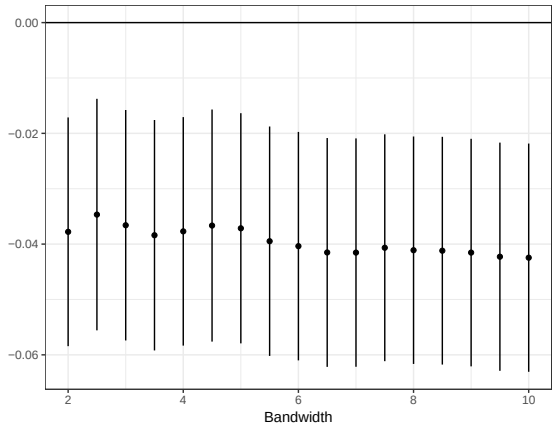
(h) Village Head High School



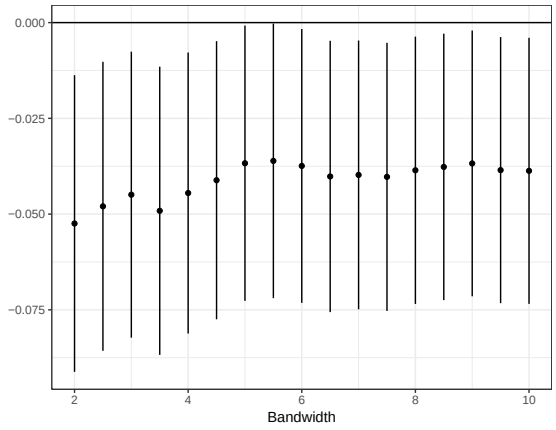
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks each 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 M-5: Industrial Structure

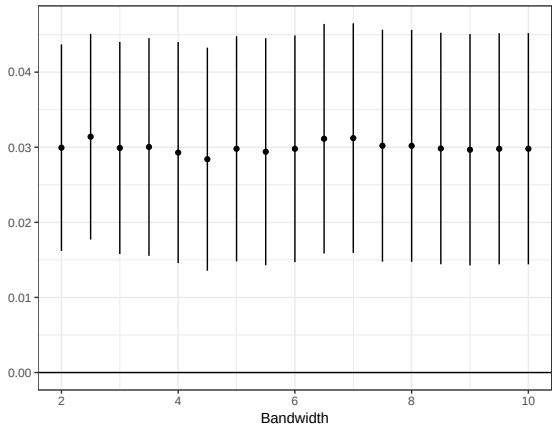
(a) Agriculture (SUSENAS)



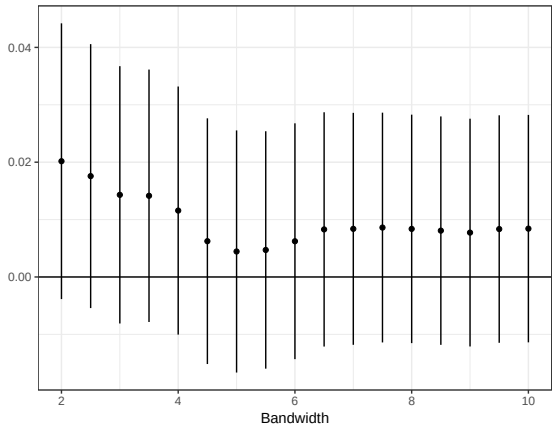
(b) Agriculture (1980 Census)



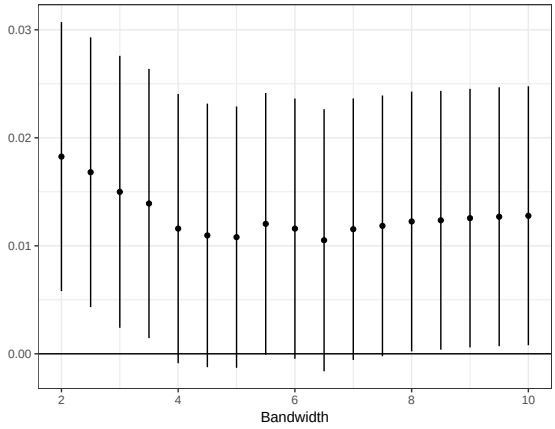
(c) Manufacturing (SUSENAS)



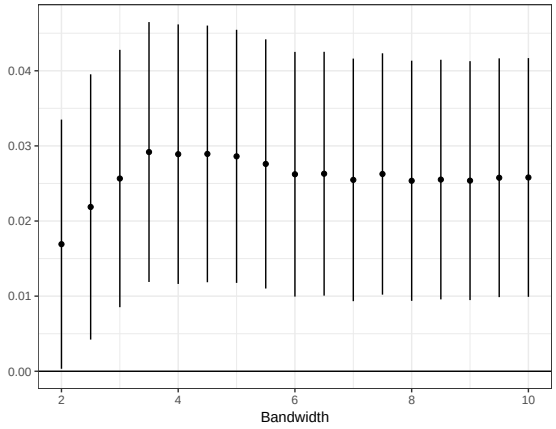
(d) Manufacturing (1980 Census)



(e) Retail (SUSENAS)



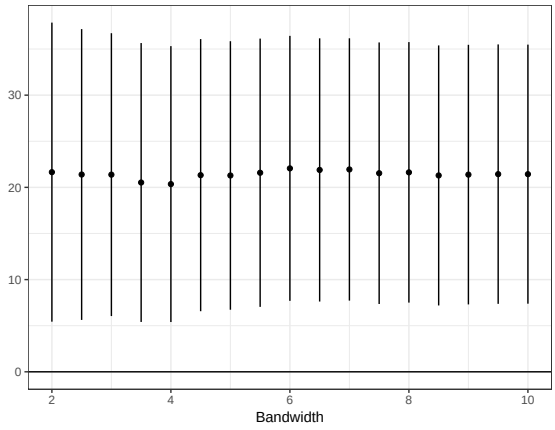
(f) Commerce (1980 Census)



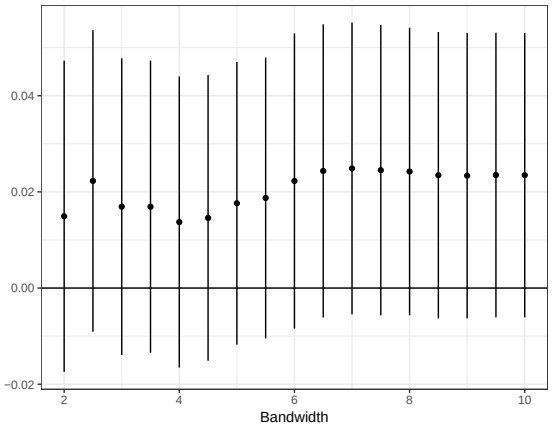
Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks each 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 M-6: Firms, Population, and Consumption

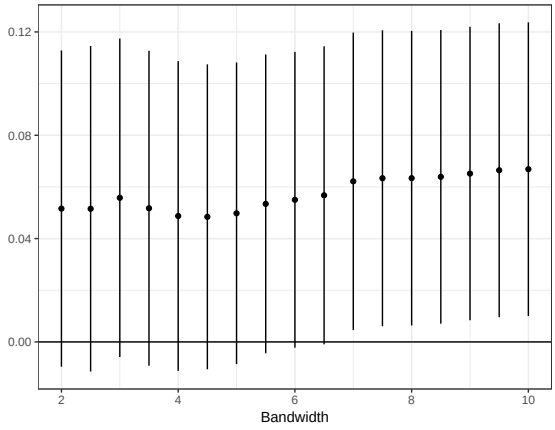
(a) Number of Firms



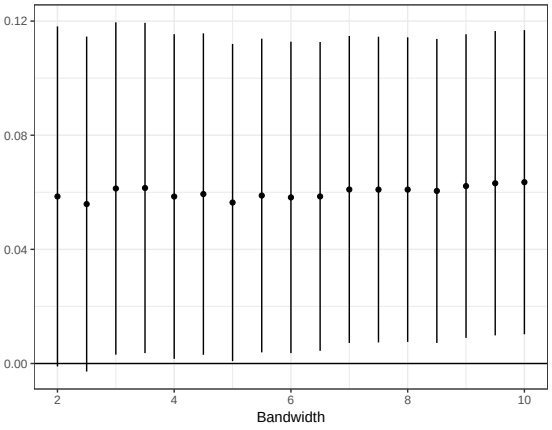
(b) Employment Share



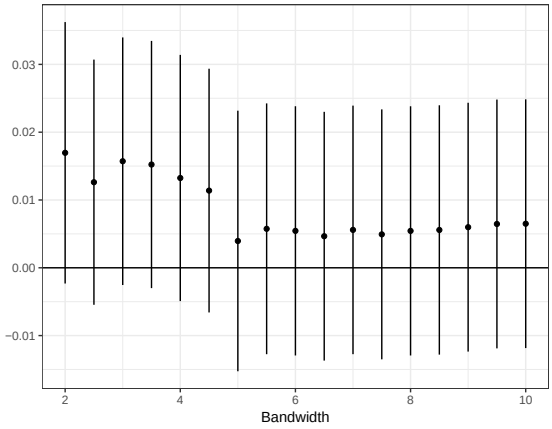
(c) Log Population Density (2003)



(d) Log Population Density (1980)



(e) Log Equivalent Consumption

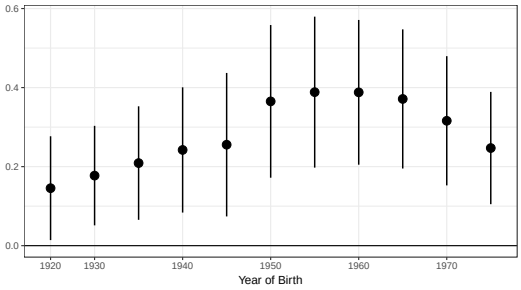


Notes: The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks each 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.

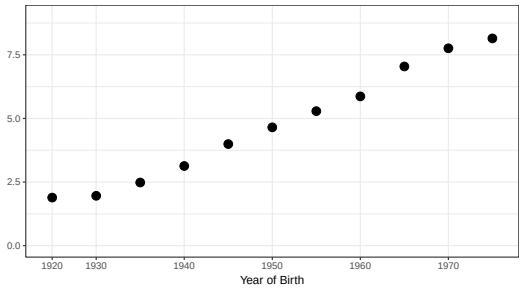
N Single Linear Latitude-Longitude Polynomial

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

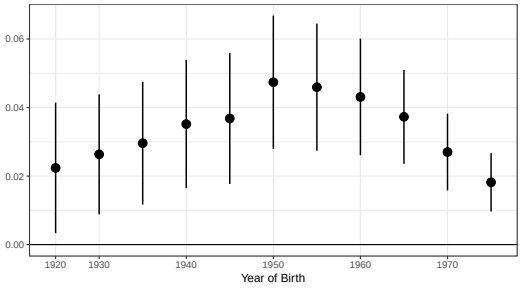
(a) Years of Schooling



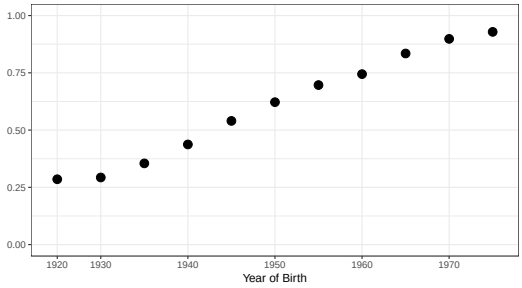
(b) Years of Schooling Levels



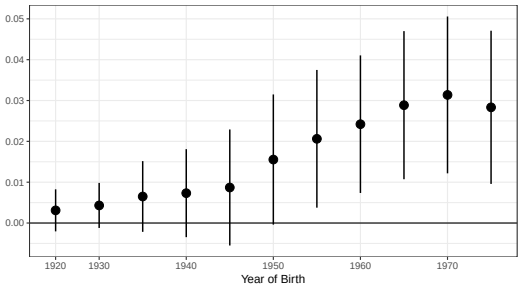
(c) Primary Completion



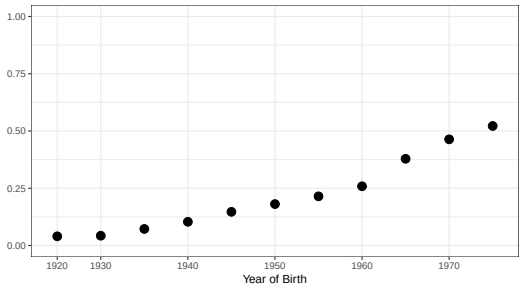
(d) Primary Completion Levels



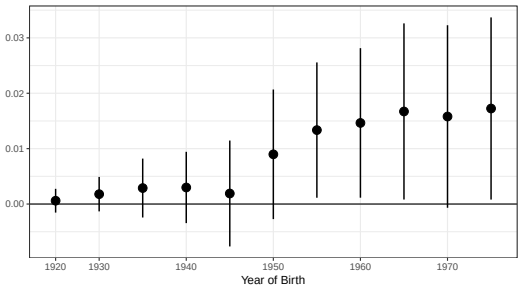
(e) Junior High Completion



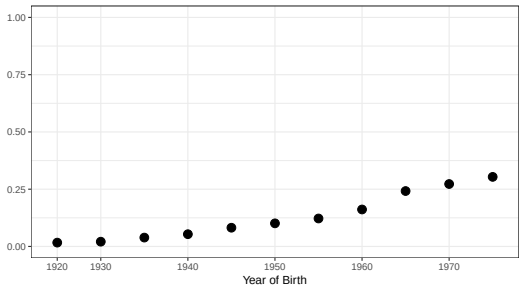
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude. Robust standard errors are clustered by subdistrict.

Table N-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.127	-0.015	-0.031	-0.007	0.054	-0.004	0.169	0.063
	(0.776)	(0.013)	(0.161)	(0.016)	(0.143)	(0.013)	(0.138)	(0.151)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 N-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	1.893	0.013	2.972	0.012	0.003	11.994	0.368	1.608
	(0.824)	(0.003)	(1.469)	(0.005)	(0.289)	(8.363)	(0.266)	(1.890)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table N-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.008	0.169	-0.042	-0.215	0.023	0.009
	(0.019)	(0.103)	(0.020)	(0.080)	(0.008)	(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table N-4: 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	0.229	0.025	0.016	0.008	-0.016	-0.002	0.095	0.026
	(0.077)	(0.007)	(0.008)	(0.006)	(0.008)	(0.010)	(0.087)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 N-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.041	0.029	0.014	-0.038	0.009	0.026
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table N-6: Firms, Population, and Consumption: Subjected Villages

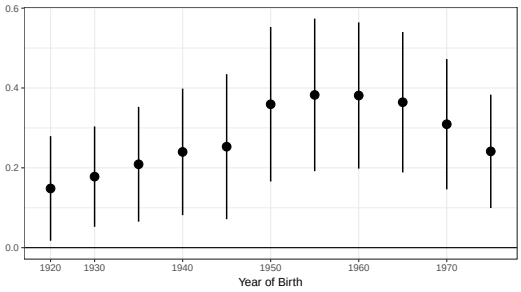
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	21.201	0.024	0.078	0.071	0.006
	(8.760)	(0.018)	(0.034)	(0.034)	(0.011)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

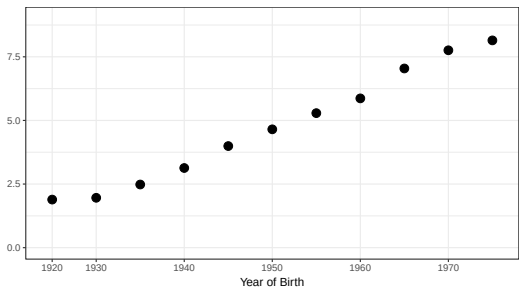
O Quadratic Latitude-Longitude Polynomial

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

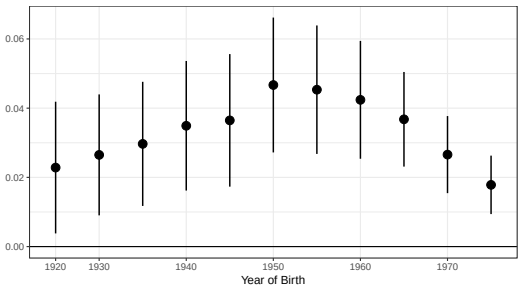
(a) Years of Schooling



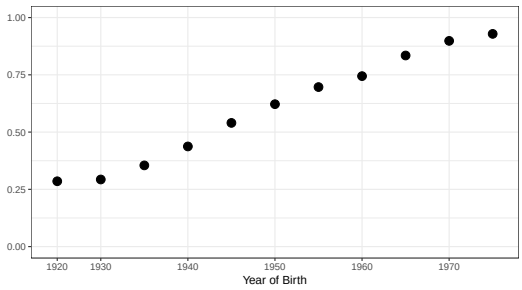
(b) Years of Schooling Levels



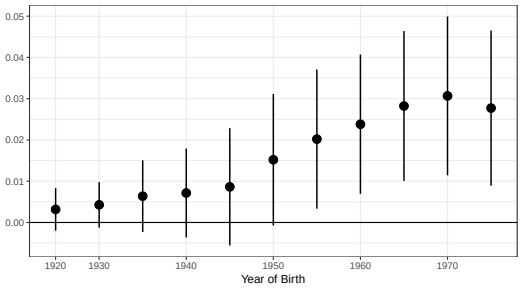
(c) Primary Completion



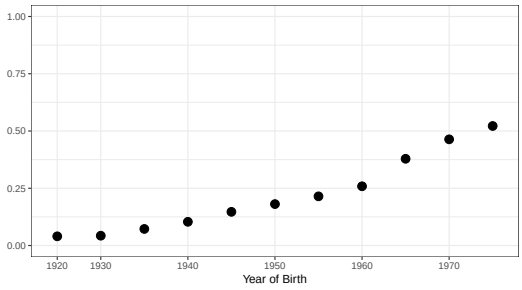
(d) Primary Completion Levels



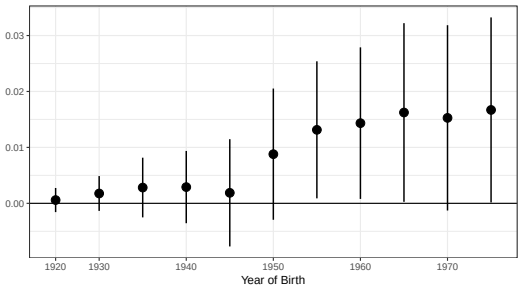
(e) Junior High Completion



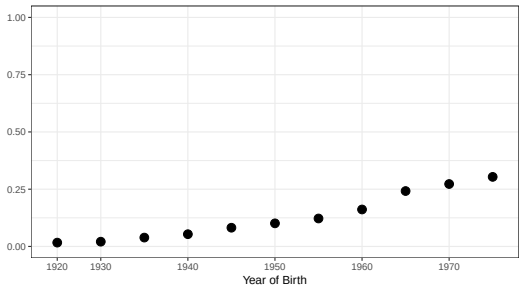
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a quadratic polynomial in latitude and longitude. Robust standard errors are clustered by subdistrict.

Table O-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.151	-0.016	-0.026	-0.008	0.089	-0.004	0.177	0.083
	(0.769)	(0.013)	(0.161)	(0.015)	(0.126)	(0.013)	(0.127)	(0.149)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 O-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	1.998	0.013	3.103	0.012	0.017	12.445	0.362	1.700
	(0.837)	(0.004)	(1.483)	(0.005)	(0.285)	(8.508)	(0.268)	(1.907)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table O-3: 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	0.009	0.171	-0.041	-0.217	0.022	0.008
	(0.018)	(0.103)	(0.020)	(0.080)	(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a quadratic polynomial in latitude and longitude. Robust standard errors, clustered by subdistrict, are in parentheses.

Table O-4: 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	0.224	0.025	0.016	0.008	-0.016	-0.002	0.093	0.026
	(0.077)	(0.007)	(0.008)	(0.006)	(0.009)	(0.010)	(0.087)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 O-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.040	0.028	0.014	-0.037	0.007	0.026
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table O-6: Firms, Population, and Consumption: Subjected Villages

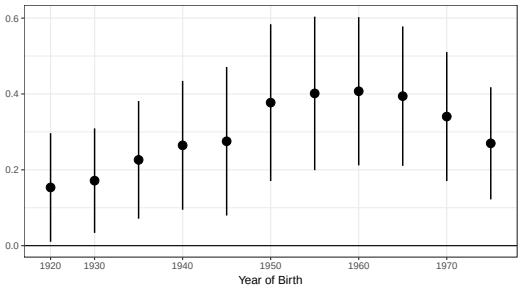
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	20.703	0.024	0.076	0.070	0.006
	(8.679)	(0.018)	(0.034)	(0.034)	(0.011)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a quadratic polynomial in latitude and longitude. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

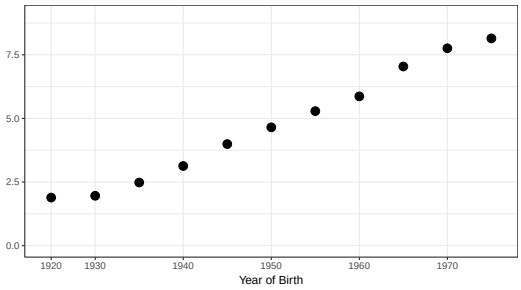
P One-Dimensional Linear RD Polynomial

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

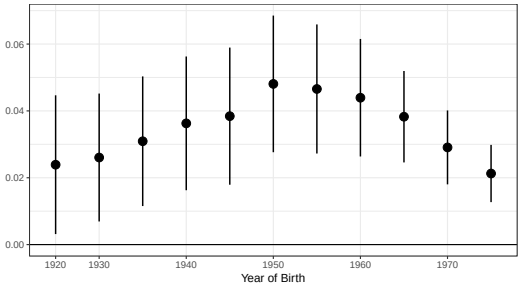
(a) Years of Schooling



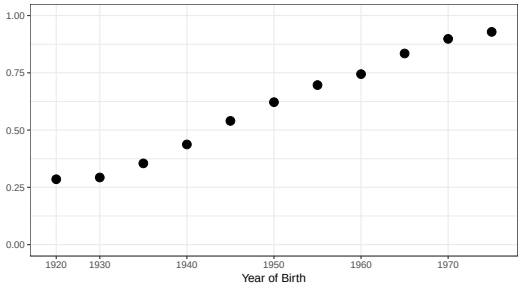
(b) Years of Schooling Levels



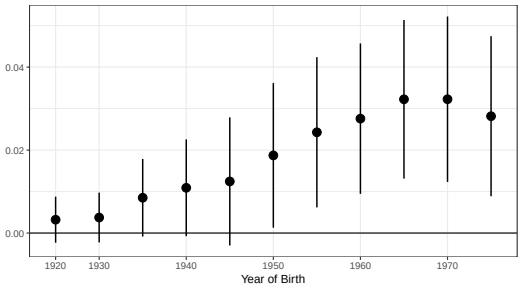
(c) Primary Completion



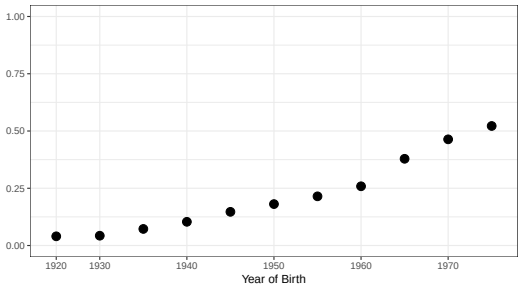
(d) Primary Completion Levels



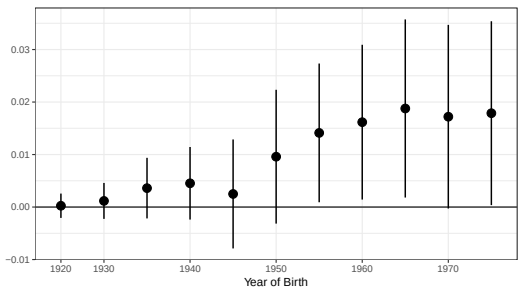
(e) Junior High Completion



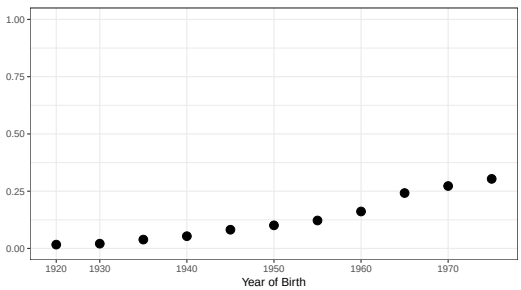
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in distance to the boundary. Robust standard errors are clustered by subdistrict.

Table P-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	0.076	0.014	0.075	-0.041	0.103	-0.020	0.245	0.007
	(0.777)	(0.013)	(0.174)	(0.016)	(0.171)	(0.014)	(0.177)	(0.175)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 P-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	1.728	0.011	2.729	0.009	0.024	14.740	0.318	2.012
	(0.892)	(0.004)	(1.732)	(0.006)	(0.281)	(9.096)	(0.270)	(2.148)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table P-3: 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	0.001	0.106	-0.044	-0.214	0.028	0.012
	(0.020)	(0.109)	(0.022)	(0.089)	(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in distance to the boundary. Robust standard errors, clustered by subdistrict, are in parentheses.

Table P-4: 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	0.239	0.026	0.018	0.009	-0.015	-0.003	0.083	0.022
	(0.082)	(0.007)	(0.008)	(0.006)	(0.009)	(0.011)	(0.089)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 P-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.040	0.026	0.016	-0.035	0.016	0.020
	(0.014)	(0.011)	(0.008)	(0.022)	(0.013)	(0.010)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table P-6: Firms, Population, and Consumption: Subjected Villages

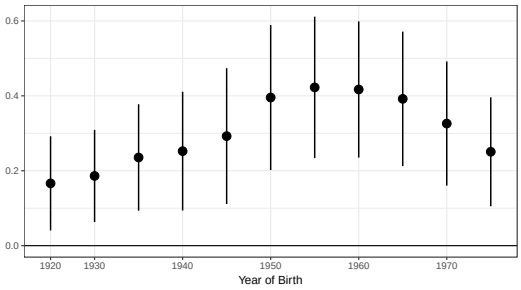
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	15.407	0.010	0.087	0.080	0.007
	(9.186)	(0.018)	(0.036)	(0.036)	(0.012)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in distance to the boundary. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

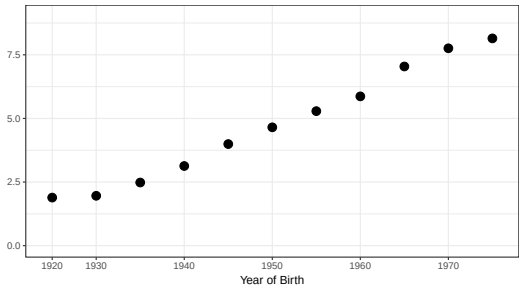
Q 25 Km Boundary Segment Fixed Effects

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

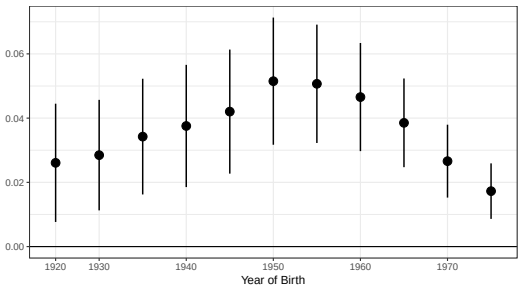
(a) Years of Schooling



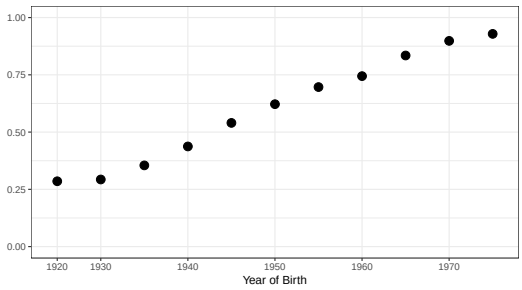
(b) Years of Schooling Levels



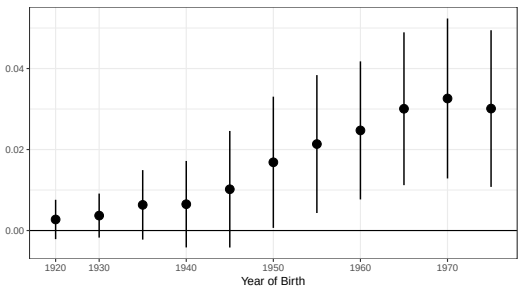
(c) Primary Completion



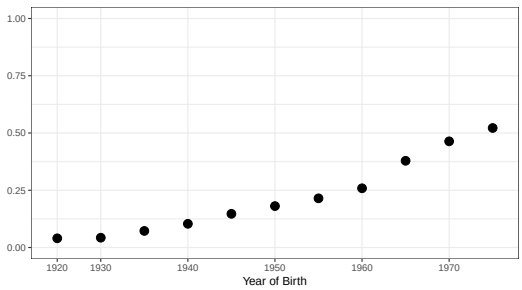
(d) Primary Completion Levels



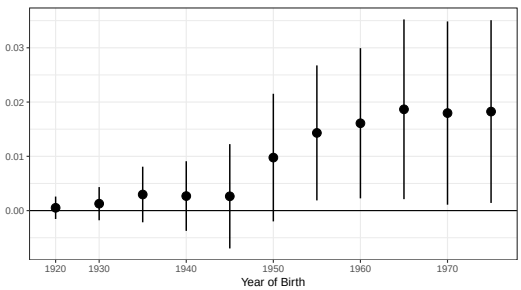
(e) Junior High Completion



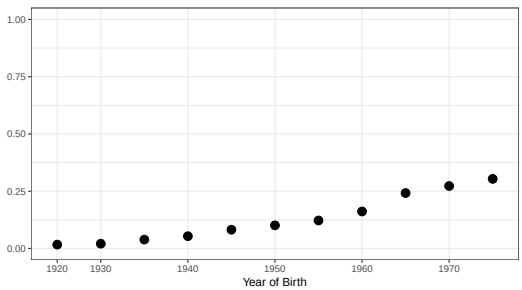
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table Q-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.036	-0.016	-0.049	-0.022	0.326	0.000	0.413	0.038
	(0.826)	(0.014)	(0.150)	(0.015)	(0.123)	(0.012)	(0.166)	(0.167)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 Q-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	2.385	0.015	2.608	0.008	0.234	15.714	0.434	2.369
	(0.904)	(0.003)	(1.553)	(0.006)	(0.295)	(7.397)	(0.260)	(1.691)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table Q-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.014	0.168	-0.036	-0.225	0.026	0.008
	(0.018)	(0.098)	(0.019)	(0.077)	(0.008)	(0.005)
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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table Q-4: 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	0.243	0.027	0.017	0.009	-0.016	-0.004	0.085	0.024
	(0.075)	(0.006)	(0.008)	(0.006)	(0.008)	(0.009)	(0.082)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 Q-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.046	0.025	0.018	-0.043	0.013	0.026
	(0.013)	(0.011)	(0.007)	(0.019)	(0.011)	(0.010)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table Q-6: Firms, Population, and Consumption: Subjected Villages

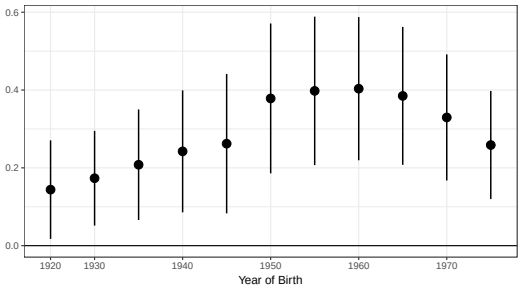
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	20.981	0.020	0.098	0.066	0.007
	(8.510)	(0.018)	(0.034)	(0.033)	(0.011)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

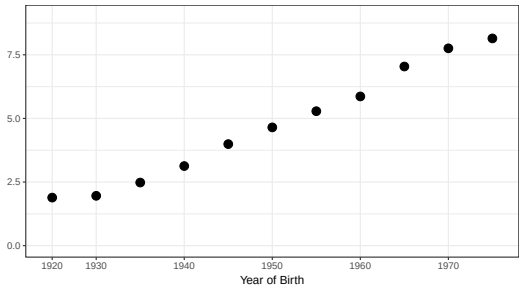
R No Geographic Controls

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

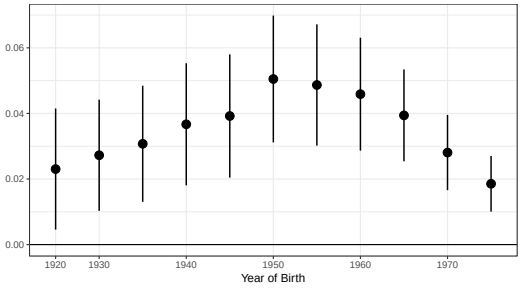
(a) Years of Schooling



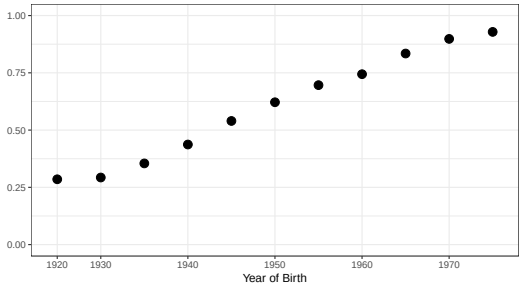
(b) Years of Schooling Levels



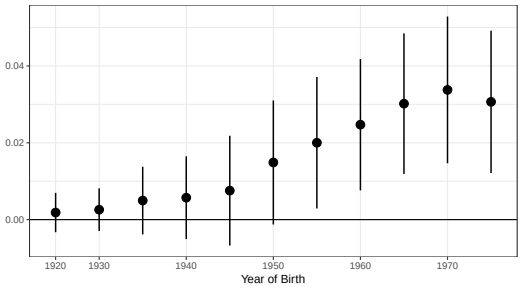
(c) Primary Completion



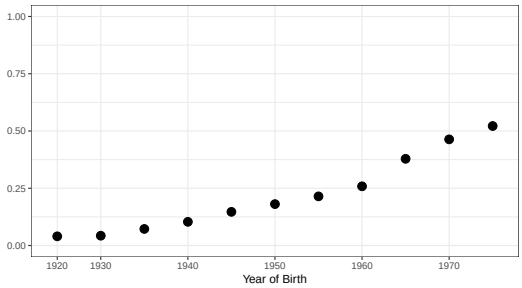
(d) Primary Completion Levels



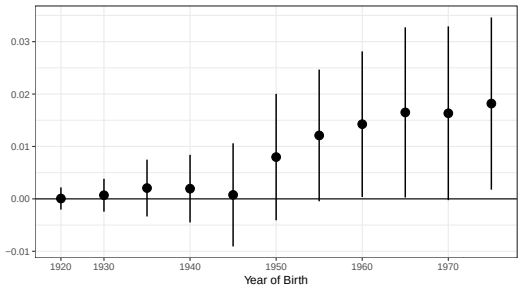
(e) Junior High Completion



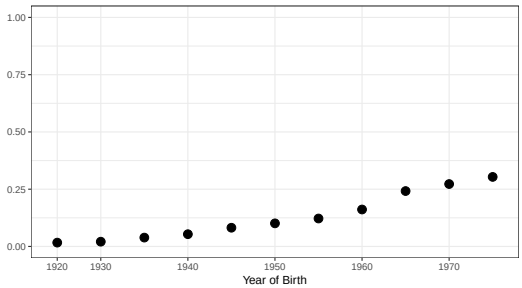
(f) Junior High Completion Levels



(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 R-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-1.908	-0.017	-0.041	-0.013	0.117	0.001	0.126	0.062
	(0.744)	(0.013)	(0.160)	(0.016)	(0.106)	(0.012)	(0.107)	(0.133)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 R-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	2.341	0.015	3.344	0.014	0.015	13.971	0.373	1.912
	(0.857)	(0.004)	(1.495)	(0.005)	(0.265)	(8.368)	(0.268)	(1.959)
Obs	4,554	4,554	4,209	4,110	4,093	4,084	4,092	4,084
Clusters	383	383	380	380	381	381	381	381
Mean	18.60	0.09	23.93	0.11	3.53	38.70	4.35	7.86

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 R-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.014	0.199	-0.033	-0.206	0.019	0.007
	(0.019)	(0.106)	(0.020)	(0.081)	(0.008)	(0.006)
Obs	4,209	4,209	4,209	4,209	4,209	4,209
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 boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table R-4: 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 R-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.046	0.032	0.013	-0.037	0.008	0.026
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)
Obs	130,479	130,479	130,479	127,901	127,901	127,901
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55.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.

Table R-6: Firms, Population, and Consumption: Subjected Villages

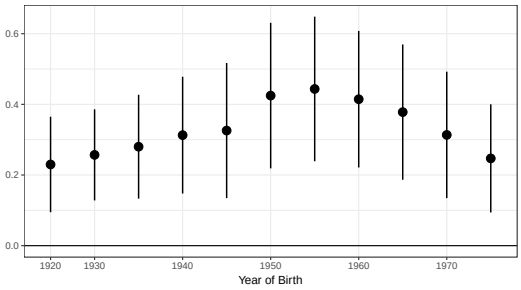
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	18.919	0.025	0.066	0.071	0.007
	(8.544)	(0.019)	(0.036)	(0.033)	(0.012)
Obs	4,553	4,553	4,554	4,110	144,178
Clusters	383	383	383	380	381
Mean	71.74	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). 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. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

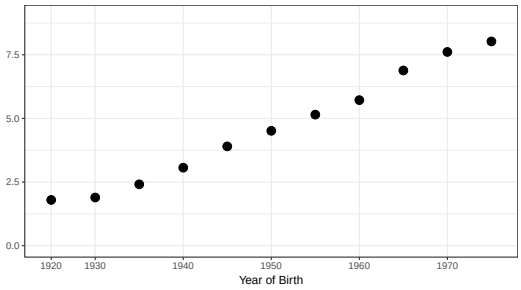
S Alternative Sample Restriction

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

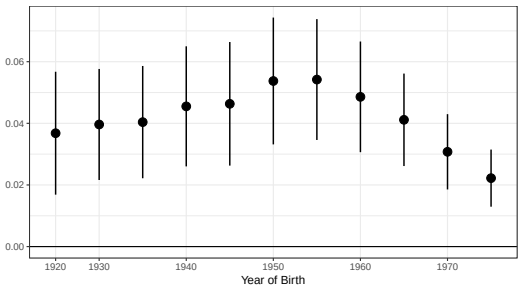
(a) Years of Schooling



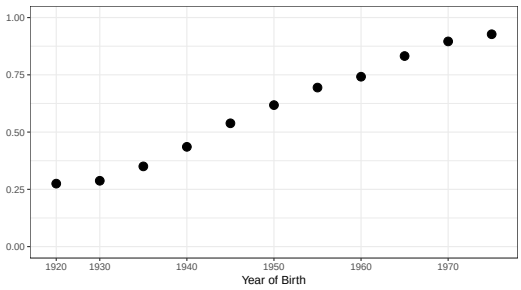
(b) Years of Schooling Levels



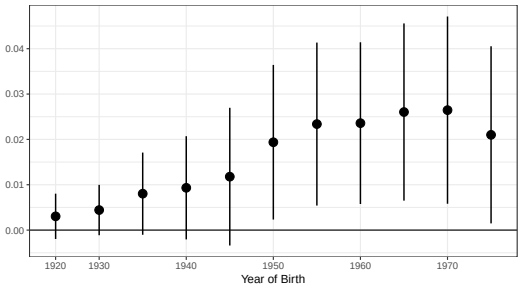
(c) Primary Completion



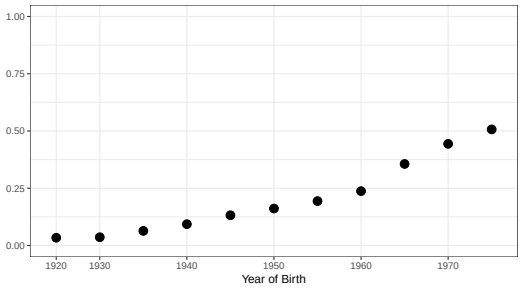
(d) Primary Completion Levels



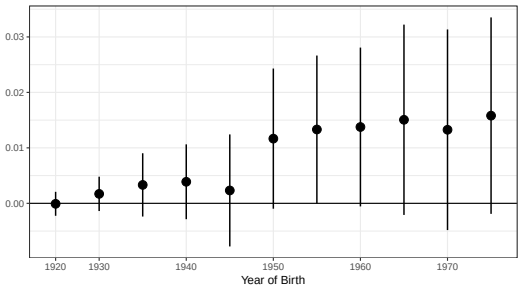
(e) Junior High Completion



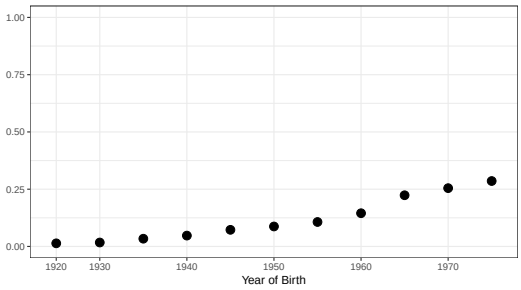
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table S-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	1.165	0.064	0.024	-0.025	0.341	-0.011	0.215	0.118
	(0.789)	(0.028)	(0.179)	(0.018)	(0.096)	(0.014)	(0.099)	(0.145)
Obs	3,290	3,290	3,290	3,293	3,293	3,293	3,293	3,293
Clusters	330	330	330	330	330	330	330	330
Mean	41.11	0.34	2.64	0.05	28.37	0.28	43.68	24.75

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 S-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	1.110	0.010	4.051	0.018	0.117	17.330	0.525	1.899
	(1.051)	(0.005)	(1.806)	(0.006)	(0.189)	(9.525)	(0.328)	(2.372)
Obs	3,288	3,288	3,048	2,993	3,041	3,027	3,039	3,027
Clusters	330	330	326	326	327	327	327	327
Mean	20.04	0.09	25.09	0.11	3.45	40.28	4.35	8.10

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

Table S-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.007	0.131	-0.032	-0.134	0.013	0.010
	(0.021)	(0.126)	(0.022)	(0.098)	(0.009)	(0.006)
Obs	3,048	3,048	3,048	3,048	3,048	3,048
Clusters	326	326	326	326	326	326
Mean	0.45	2.97	0.35	1.33	0.06	0.02

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

Table S-4: 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	0.244	0.029	0.015	0.007	-0.008	-0.006	0.200	0.045
	(0.084)	(0.007)	(0.008)	(0.006)	(0.009)	(0.010)	(0.104)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 S-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.043	0.027	0.018	-0.054	0.018	0.025
	(0.014)	(0.010)	(0.008)	(0.021)	(0.013)	(0.010)
Obs	87,866	87,866	87,866	94,064	94,064	94,064
Clusters	323	323	323	308	308	308
Mean	0.33	0.18	0.17	0.52	0.10	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table S-6: Firms, Population, and Consumption: Subjected Villages

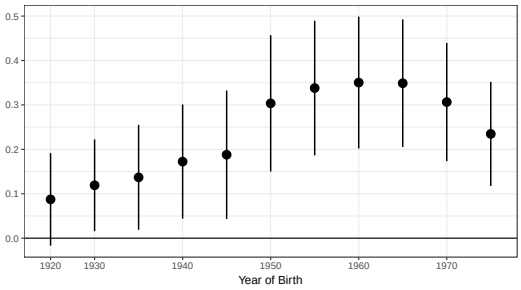
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	9.675	0.006	0.097	0.084	0.010
	(8.731)	(0.018)	(0.033)	(0.033)	(0.012)
Obs	3,289	3,289	3,288	2,993	99,095
Clusters	330	330	330	326	323
Mean	69.73	0.14	2.80	2.49	12.51

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

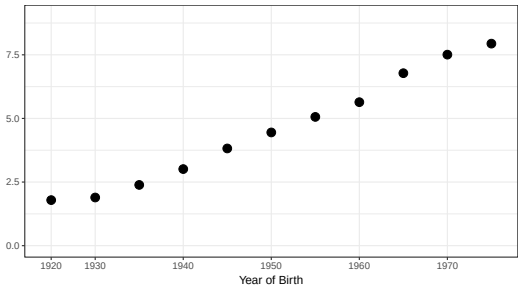
T Full Sample

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

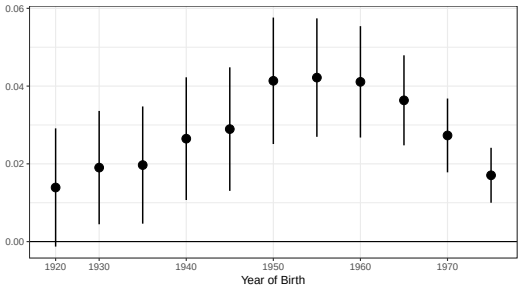
(a) Years of Schooling



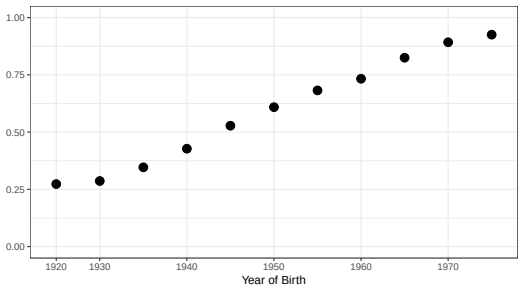
(b) Years of Schooling Levels



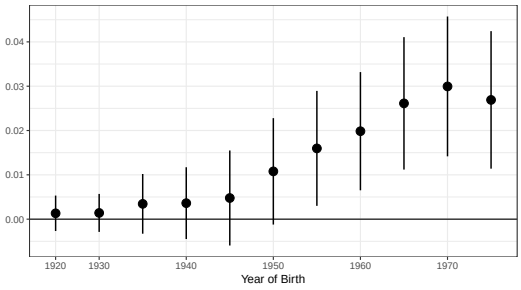
(c) Primary Completion



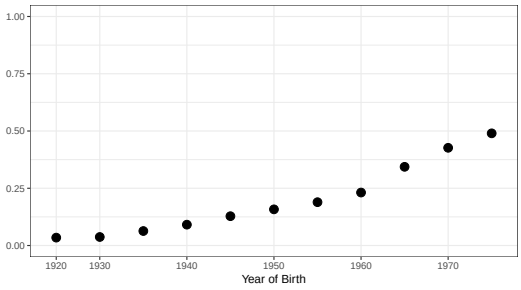
(d) Primary Completion Levels



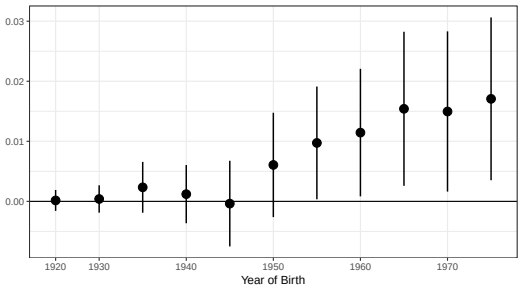
(e) Junior High Completion



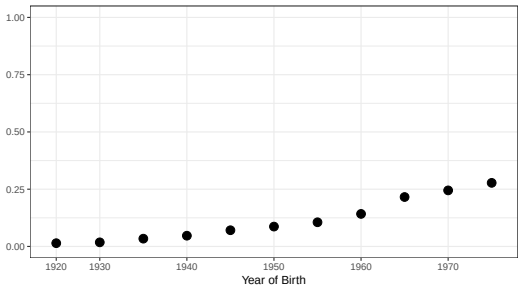
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table T-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-10.489	-0.139	-0.018	-0.004	0.133	-0.004	0.205	0.004
	(3.068)	(0.051)	(0.121)	(0.012)	(0.080)	(0.009)	(0.089)	(0.108)
Obs	6,898	6,898	6,893	6,906	6,906	6,906	6,906	6,906
Clusters	524	524	524	524	524	524	524	524
Mean	114.02	1.11	2.56	0.04	26.38	0.26	39.26	25.23

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 T-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	3.625	0.015	3.054	0.014	0.064	11.390	0.366	1.389
	(1.348)	(0.003)	(1.241)	(0.005)	(0.211)	(6.054)	(0.208)	(1.388)
Obs	6,891	6,891	6,360	6,252	6,390	6,369	6,388	6,369
Clusters	524	524	523	523	522	522	522	522
Mean	20.07	0.08	22.78	0.10	3.31	32.91	4.07	7.08

Notes: The unit of observation is the village. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, 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-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.012	0.178	-0.036	-0.187	0.013	0.004
	(0.015)	(0.091)	(0.017)	(0.070)	(0.007)	(0.004)
Obs	6,360	6,360	6,360	6,360	6,360	6,360
Clusters	523	523	523	523	523	523
Mean	0.45	2.87	0.36	1.38	0.05	0.02

Notes: The unit of observation is the village. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, 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-4: 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	0.207	0.023	0.015	0.007	-0.014	-0.004	0.151	0.039
	(0.062)	(0.006)	(0.006)	(0.004)	(0.007)	(0.008)	(0.073)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 T-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.038	0.029	0.010	-0.029	0.008	0.018
	(0.010)	(0.007)	(0.006)	(0.017)	(0.009)	(0.008)
Obs	189,152	189,152	189,152	193,946	193,946	193,946
Clusters	519	519	519	508	508	508
Mean	0.32	0.18	0.16	0.53	0.10	0.11

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table T-6: Firms, Population, and Consumption: Subjected Villages

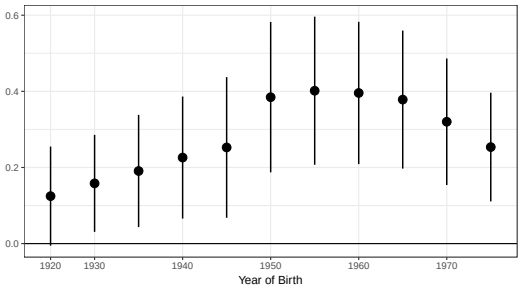
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	22.400	0.028	0.093	0.080	0.010
	(6.634)	(0.014)	(0.027)	(0.027)	(0.009)
Obs	6,891	6,891	6,891	6,252	211,544
Clusters	524	524	524	523	519
Mean	75.91	0.15	2.72	2.40	12.50

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

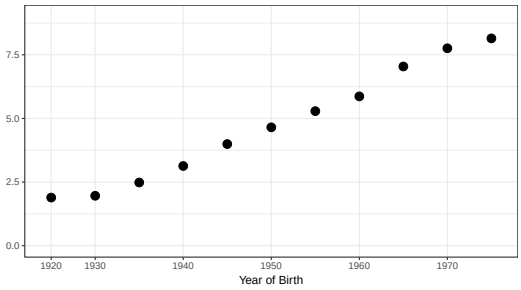
U No Triangular Kernel Weighting

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

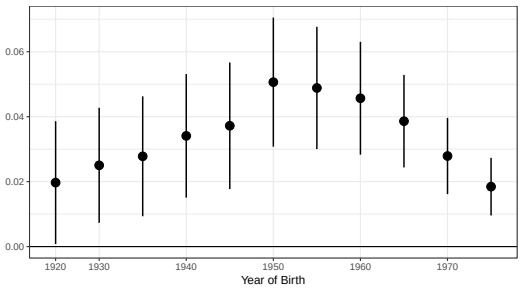
(a) Years of Schooling



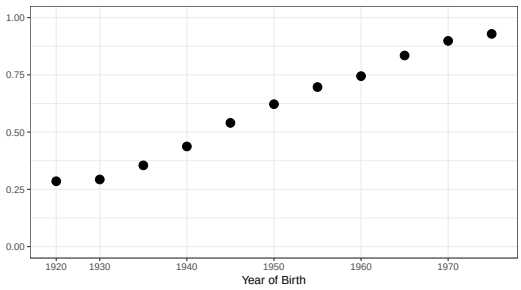
(b) Years of Schooling Levels



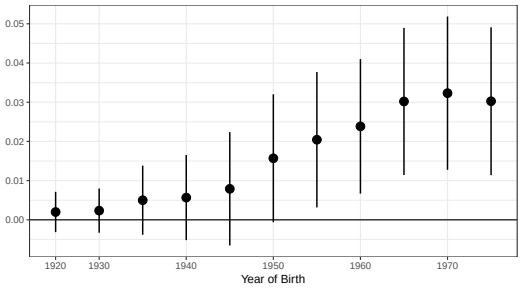
(c) Primary Completion



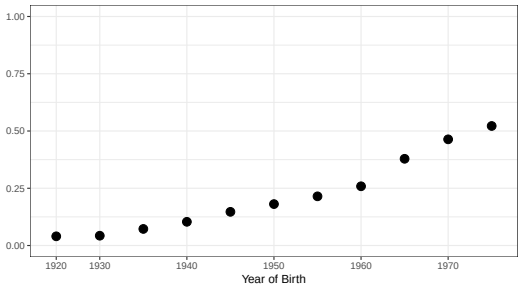
(d) Primary Completion Levels



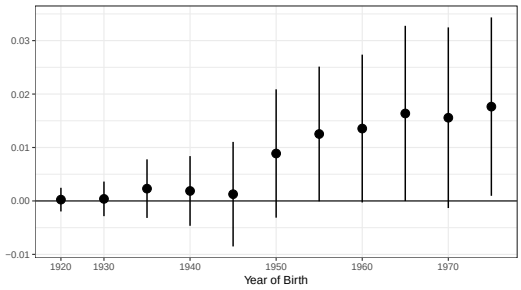
(e) Junior High Completion



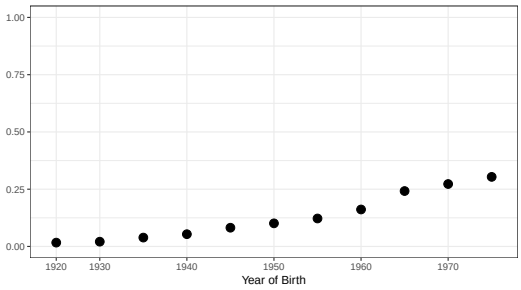
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table U-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.709	-0.024	-0.079	-0.008	0.142	-0.000	0.280	0.045
	(0.851)	(0.015)	(0.157)	(0.016)	(0.126)	(0.012)	(0.138)	(0.161)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 U-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	2.058	0.016	3.623	0.016	-0.021	16.318	0.469	2.533
	(0.859)	(0.004)	(1.572)	(0.006)	(0.249)	(8.464)	(0.275)	(2.087)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table U-3: Schools (1980): Subjected Villages

	Public Non-INPRES Primary		INPRES Primary		Junior High Schools	High Schools
	Buildings	Teachers	Buildings	Teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	0.011	0.186	-0.046	-0.221	0.022	0.007
	(0.018)	(0.100)	(0.019)	(0.078)	(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Table U-4: 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	0.230	0.026	0.017	0.007	-0.022	-0.001	0.074	0.024
	(0.078)	(0.007)	(0.008)	(0.006)	(0.009)	(0.010)	(0.086)	(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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 U-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.051	0.031	0.017	-0.041	0.011	0.027
	(0.013)	(0.010)	(0.007)	(0.021)	(0.012)	(0.010)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table U-6: Firms, Population, and Consumption: Subjected Villages

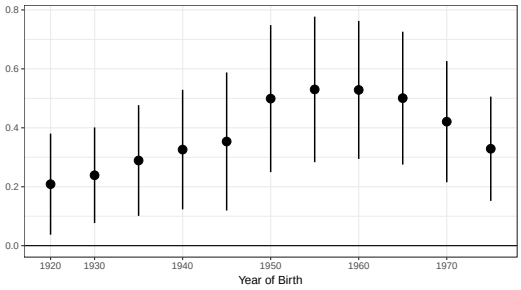
	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	22.683	0.032	0.092	0.078	0.006
	(8.390)	(0.019)	(0.036)	(0.034)	(0.011)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

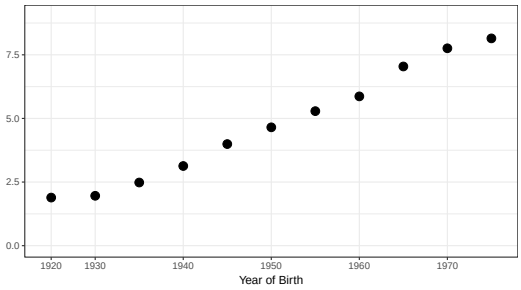
V Instrumental Variables

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

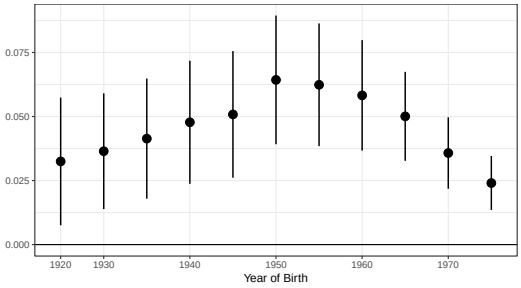
(a) Years of Schooling



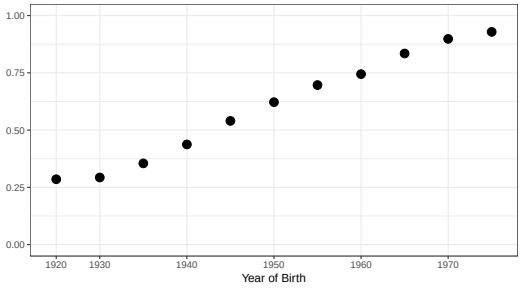
(b) Years of Schooling Levels



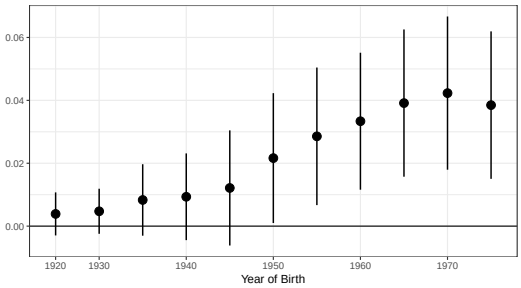
(c) Primary Completion



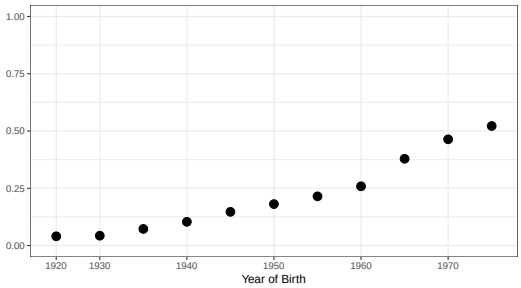
(d) Primary Completion Levels



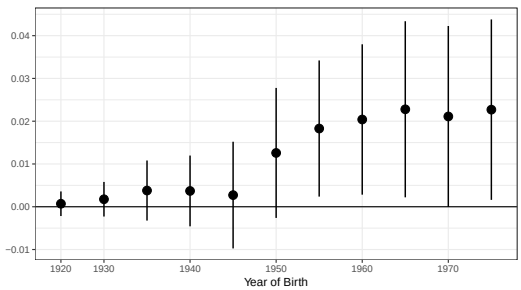
(e) Junior High Completion



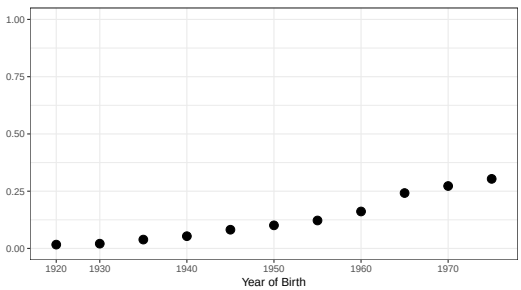
(f) Junior High Completion Levels



(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 controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors are clustered by subdistrict.

Table V-1: Geographic Characteristics: Subjected Villages

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor	Distance To 1830 Residency Capital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cultivation	-2.517	-0.022	-0.054	-0.017	0.154	0.001	0.167	0.082
	(0.939)	(0.017)	(0.203)	(0.020)	(0.135)	(0.015)	(0.136)	(0.169)
Obs	4,553	4,553	4,549	4,553	4,553	4,553	4,553	4,553
Clusters	383	383	383	383	383	383	383	383
F stat	1463	1463	1476	1463	1463	1463	1463	1463
Mean	31.17	0.26	2.56	0.06	24.90	0.29	33.26	24.80

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 V-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	2.951	0.018	4.536	0.015	0.024	17.497	0.499	2.360
	(1.069)	(0.005)	(1.914)	(0.007)	(0.330)	(10.648)	(0.345)	(2.521)
Obs	4,550	4,550	4,205	4,107	4,089	4,080	4,088	4,080
Clusters	383	383	380	380	381	381	381	381
F stat	1472	1472	1301	1429	1428	1476	1452	1476
Mean	18.61	0.09	23.95	0.11	3.53	38.58	4.34	7.84

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

Table V-3: 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	0.220	-0.045	-0.271	0.025	0.010
	(0.023)	(0.130)	(0.025)	(0.101)	(0.011)	(0.007)
Obs	4,205	4,205	4,205	4,205	4,205	4,205
Clusters	380	380	380	380	380	380
F stat	1301	1301	1301	1301	1301	1301
Mean	0.43	2.81	0.36	1.37	0.06	0.02

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

Table V-4: 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	0.309	0.034	0.023	0.011	-0.023	-0.004	0.139	0.038
	(0.097)	(0.008)	(0.010)	(0.007)	(0.012)	(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	1381	1381	1381	1381	772	772	1617	1617
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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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 V-5: Industrial Structure: Subjected Villages

	SUSENAS (2001-11)			1980 Population Census		
	Ag.	Manuf.	Retail	Ag.	Manuf.	Commerce
	(1)	(2)	(3)	(4)	(5)	(6)
Cultivation	-0.056	0.039	0.017	-0.051	0.011	0.034
	(0.016)	(0.012)	(0.010)	(0.028)	(0.016)	(0.013)
Obs	130,335	130,335	130,335	127,873	127,873	127,873
Clusters	381	381	381	358	358	358
F stat	966	966	966	745	745	745
Mean	0.27	0.21	0.18	0.48	0.11	0.12

Notes: The unit of observation is the individual. The sample is restricted to men age 18-55. Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, 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.

Table V-6: Firms, Population, and Consumption: Subjected Villages

	Num. Manuf. Firms	Manuf. Emp. Share	Log Population Density (2003)	Log Population Density (1980)	Log Equiv. Consumption
	(1)	(2)	(3)	(4)	(5)
Cultivation	28.313	0.031	0.088	0.082	0.009
	(10.673)	(0.023)	(0.044)	(0.040)	(0.015)
Obs	4,549	4,549	4,550	4,107	144,046
Clusters	383	383	383	380	381
F stat	1469	1469	1472	1429	1077
Mean	71.72	0.16	2.87	2.54	12.55

Notes: The unit of observation is the village in columns (1) through (4) and the household in column (5). Regressions include geographic controls, boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Column (5) includes year fixed effects and household demographic controls. Robust standard errors, clustered by subdistrict, are in parentheses.

W Grid Cells

Table W-1: Geographic Characteristics

	Elevation	Slope	Log Flow Accumulation	On Coast	Distance To Coast	Distance To River	Distance To Natural Harbor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cultivation	-3.704	-0.031	-0.018	-0.046	0.131	-0.123	-0.302
	(2.501)	(0.058)	(0.353)	(0.086)	(0.464)	(0.086)	(0.525)
Obs	559	559	559	559	559	559	559
Clusters	72	72	72	72	72	72	72
Mean	30.44	0.28	5.49	0.12	23.43	0.29	32.16

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.

X Additional Public Goods Outcomes

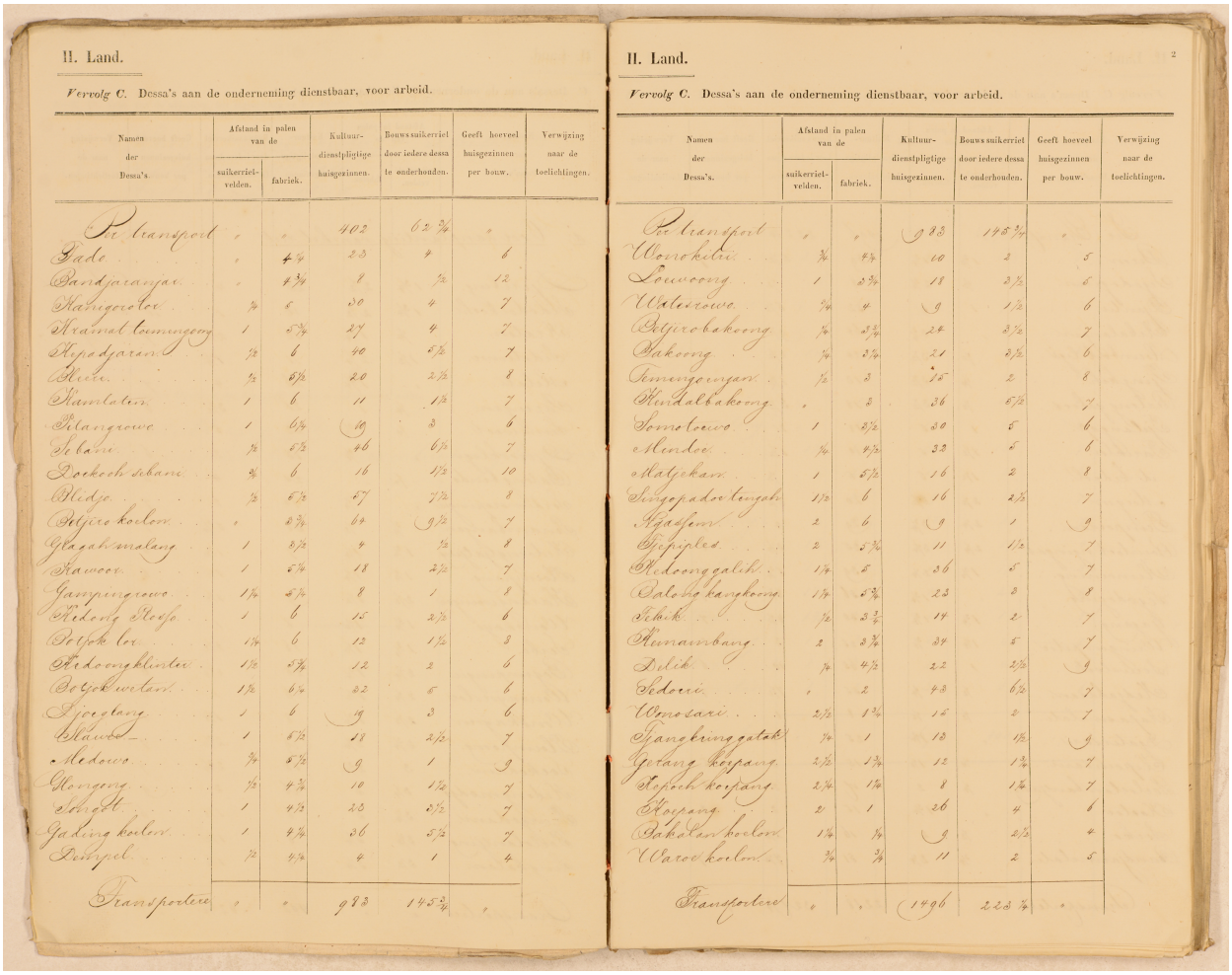
Table X-1: Public Goods

	Intercity Road Density (1)	Local Road Density (2)	Railroad Density (3)	Paved Road (4)	Dirt Road (5)	Has Electricty (6)	Distance to Subdistrict Capital (7)
Cultivation	-0.022	0.696	0.183	0.052	-0.012	0.033	-0.472
	(0.272)	(1.255)	(0.115)	(0.025)	(0.014)	(0.026)	(0.197)
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, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Robust standard errors, clustered by subdistrict, are in parentheses.

Y Data Appendix

Figure Y-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 suikerriet-velden.	Afstand in palen van de fabriek.	Kultuur-dienstpligtige huisgezinnen.	Bouws suikerriet door iedere dessa te onderhouden.	Geeft hoeveel huisgezinnen per bouw.	Verwijzing naar de toelichtingen.
Pr. transport	"	"	402	62 3/4	"	
Dado	"	4 1/4	23	4	6	
Bandjaranjar	"	4 3/4	8	1/2	12	
Kabigorobor	1/4	5	30	4	7	
Kraual Lemungoeng	1	5 3/4	27	4	7	
Kepadjaran	1/2	6	40	5 1/2	7	
Kreto	1/2	6 1/2	20	2 1/2	8	
Kamalaten	1	6	11	1 1/2	7	
Pilangrowo	1	6 1/4	19	3	6	
Sobani	1/2	5 1/2	46	6 1/2	7	
Doekoh sobani	3/4	6	16	1 1/2	10	
Bledjo	1/2	6 1/2	57	7 1/2	8	
Petjirokoelon	"	3 3/4	64	9 1/2	7	
Glagahmalang	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 Kroso	1	6	15	2 1/2	6	
Petjok lor	1 1/4	6	12	1 1/2	8	
Pedoongblimbing	1 1/2	5 3/4	12	2	6	
Petjokwetan	1 1/2	6 1/2	22	5	6	
Sjoglang	1	6	9	3	6	
Klawe	1	5 1/2	18	2 1/2	7	
Medowo	3/4	5 1/2	9	1	9	
Klongeng	1/2	4 3/4	10	1 1/2	7	
Sidogdol	1	4 1/2	23	3 1/2	7	
Gading koelon	1	4 1/4	36	5 1/2	7	
Pedjapel	1/2	4 1/2	4	1	4	
Transporteren	"	"	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 suikerriet-velden.	Afstand in palen van de fabriek.	Kultuur-dienstpligtige huisgezinnen.	Bouws suikerriet door iedere dessa te onderhouden.	Geeft hoeveel huisgezinnen per bouw.	Verwijzing naar de toelichtingen.
Pr. transport	"	"	983	145 3/4	"	
Wonokelir	3/4	4 1/4	10	2	5	
Louwoong	1	5 3/4	18	3 1/2	6	
Waterrowo	3/4	4	9	1 1/2	6	
Petjirobakoong	1/4	5 3/4	24	3 1/2	7	
Bakoong	1/4	5 1/4	21	3 1/2	6	
Semengangan	1/2	5	15	2	8	
Kendalbakoong	"	5	36	5 1/2	7	
Somolowo	1	5 1/2	30	5	6	
Mendoe	1/4	4 1/2	32	5	6	
Malyekarr	1	5 1/2	16	2	8	
Sengapadoelangan	1 1/2	6	16	2 1/2	7	
Ngaglem	2	6	9	1	9	
Gepipples	2	5 3/4	11	1 1/2	7	
Kedoonggalih	1/4	5	36	5	7	
Salokbangkoong	1/4	5 3/4	23	3	8	
Tekik	1/2	5 3/4	14	2	7	
Kunambang	2	5 3/4	34	5	7	
Dalih	1/4	4 1/2	22	2 1/2	9	
Sedoeri	"	2	43	6 1/2	7	
Wonosari	2 1/2	1 3/4	12	2	7	
Gangbinggatak	1/4	1	13	1 1/2	9	
Gotang koepang	2 1/2	1 3/4	12	1 3/4	7	
Sopoh 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	
Transporteren	"	"	1496	223 1/4	"	

Table Y-1: Data Sources

Data Source	**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	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	N/A	Hydrosheds	Elevation, Slope, Flow Accumulation
Author's Calculations	N/A	N/A	Distance to Coast
Digital Atlas of Southeast Asia	N/A	USGS	Distance to Natural Harbor (Perennial lake or river within 10km of coast)
Indonesian Government Topographical Map	N/A	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	1830 Residency 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

Z 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.