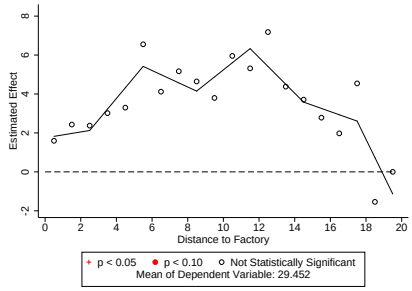
# Contents

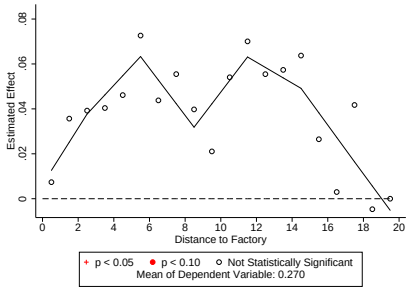
# A Common Shifts Placebos

Figure A-1: Geographic Controls

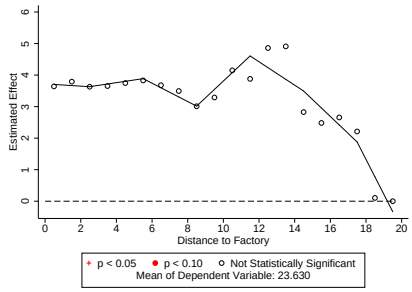
(a) *Elevation*



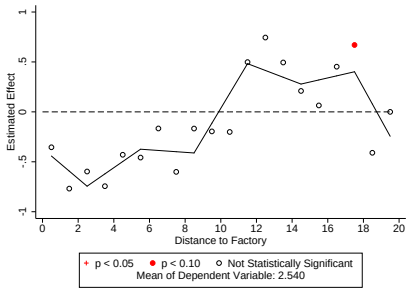

(b) *Slope*



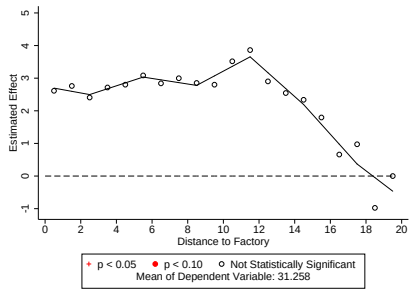

(c) *Distance to Coast*



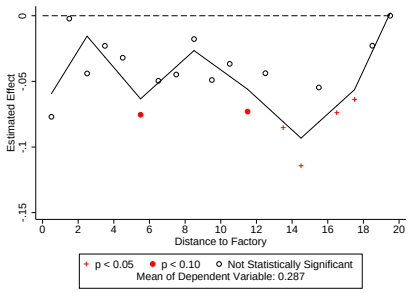

(d) *Log Flow Accumulation*



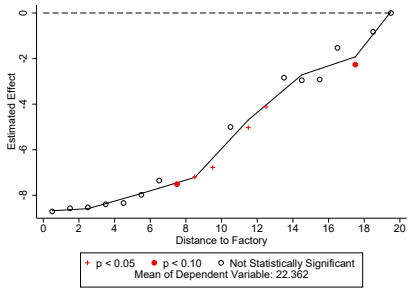

(e) *Distance to Natural Harbor*




(f) *Distance to River*



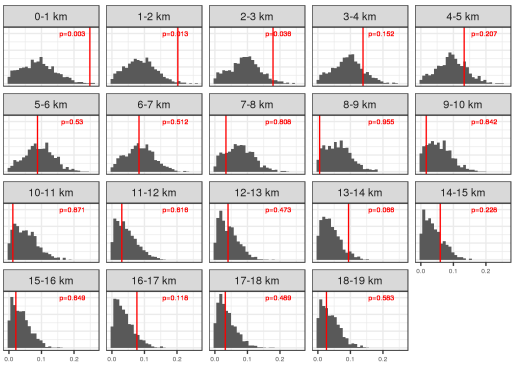

(g) *Distance to Nearest 1830 Residency Capital*



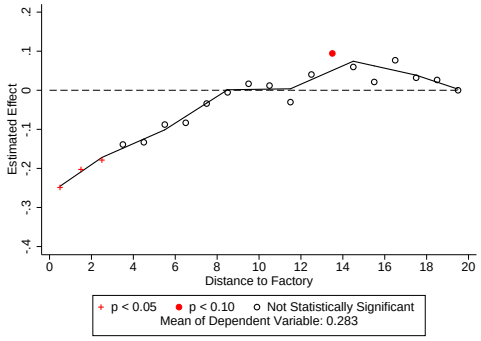

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact of proximity to 400 counterfactual factory locations.

Figure A-2: Share in Agriculture (Susenas 2001-11)

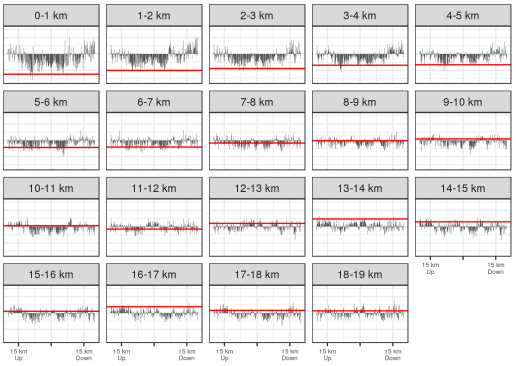
(a) *Independent Shifts*



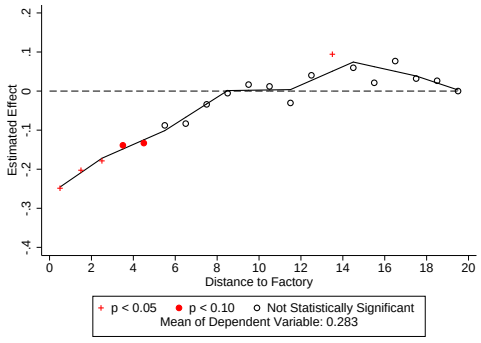

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



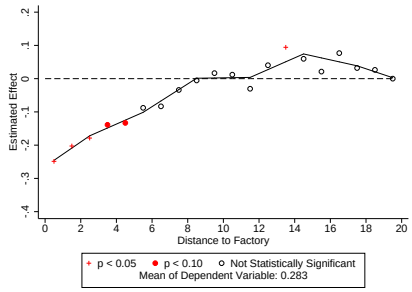

(d) *Coefficients, Common Shifts*



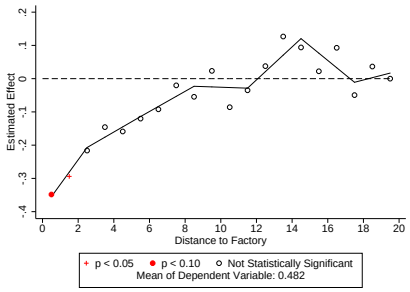

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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: Industrial Shares

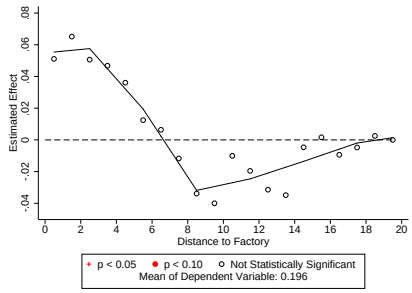
(a) *Agriculture (Susenas 2001-11)*



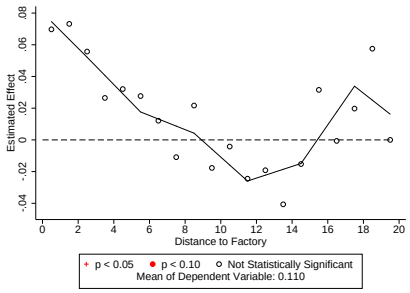

(b) *Agriculture (Census 1980)*



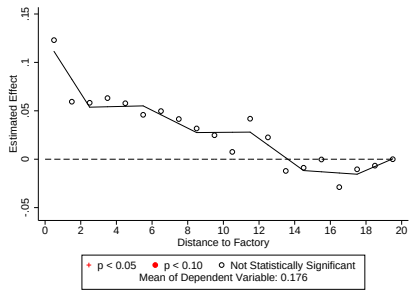

(c) *Manufacturing (Susenas 2001-11)*



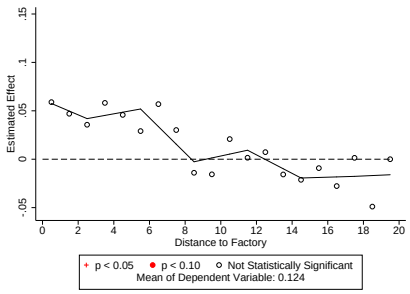

(d) *Manufacturing (Census 1980)*



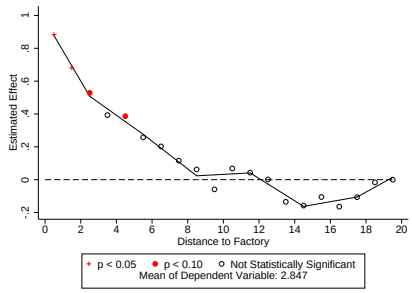

(e) *Retail (Susenas 2001-11)*



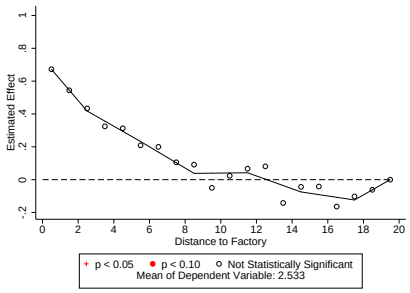

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



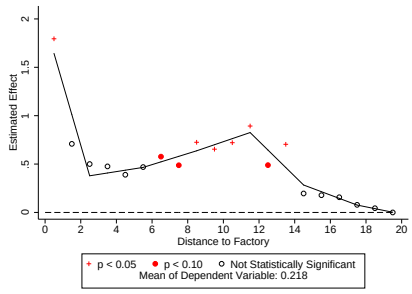

(h) *Log Population Density (PODES 2003)*



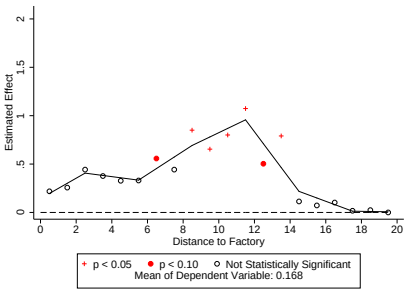

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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 400 counterfactual factory locations.

Figure A-4: Sugar Production Today

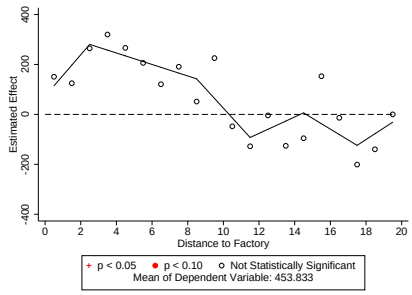
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



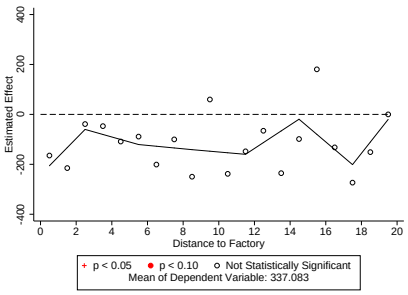

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



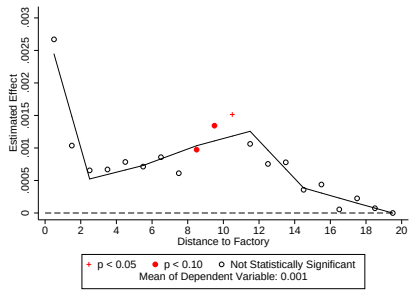

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



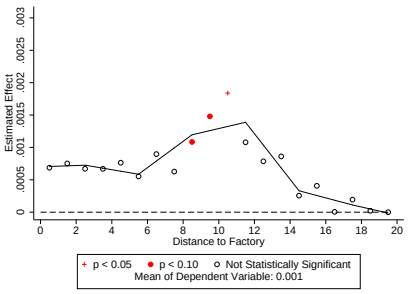

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



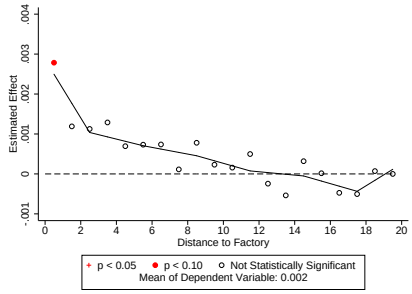

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



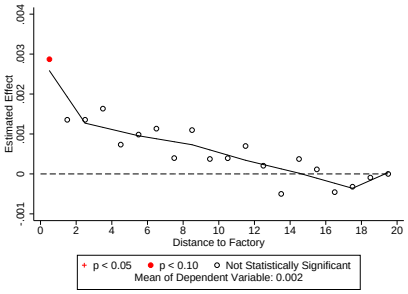

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



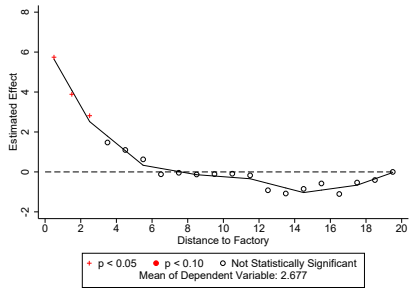

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



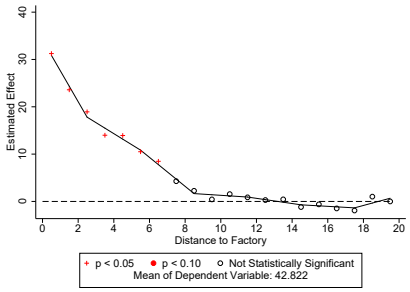

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 400 counterfactual factory locations.

Figure A-5: Public Goods

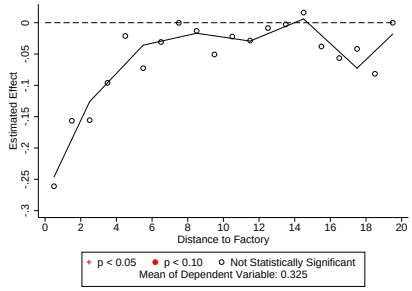
(a) *Intercity Road Density*



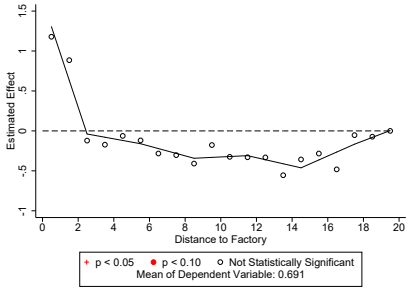

(b) *Local Road Density*



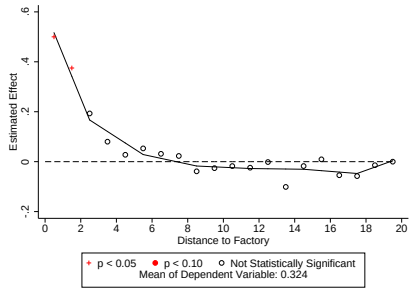

(c) *Dirt Road (PODES 1980)*



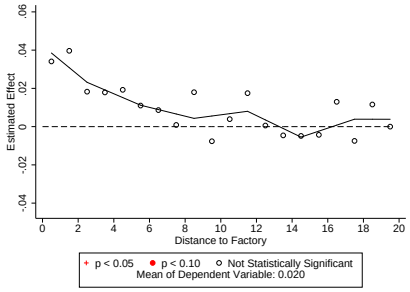

(d) *Railroad Density*



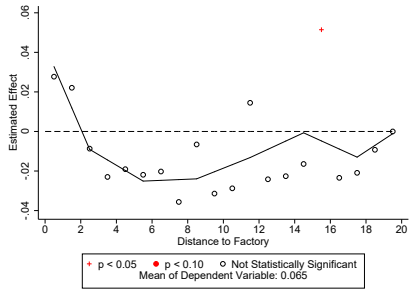

(e) *Village Has Electricity (PODES 1980)*



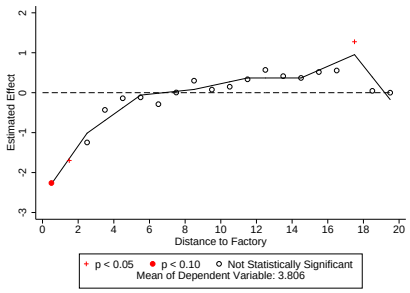

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



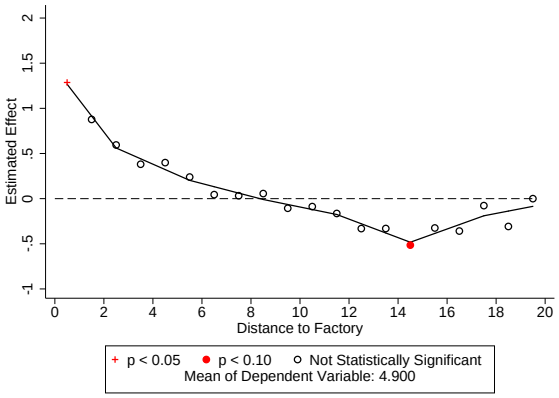

(h) *Subdistrict Capital (2011 PODES)*



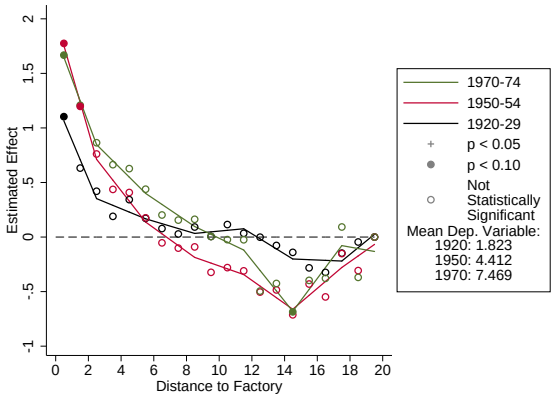

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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 400 counterfactual factory locations.

Figure A-6: Education

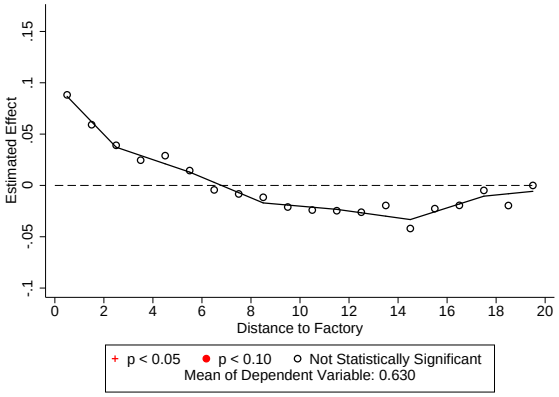
(a) *Years Education*



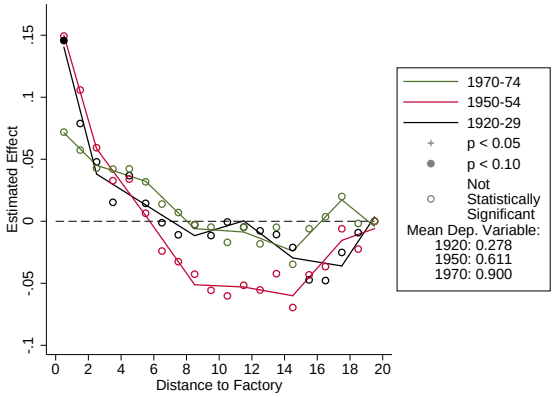

(b) *Years Education by Cohort*



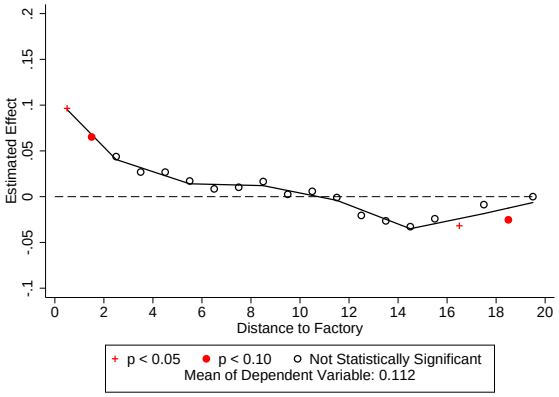

(c) *Primary*



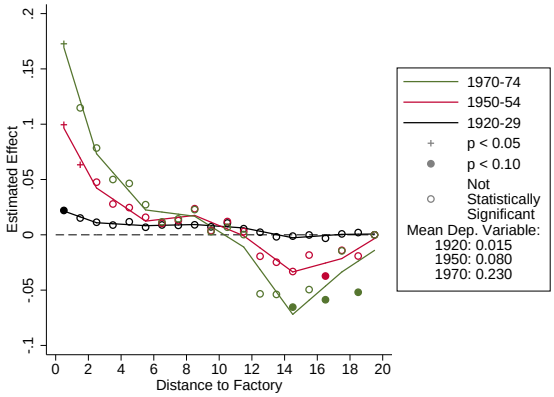

(d) *Primary by Cohort*




(e) *High School*

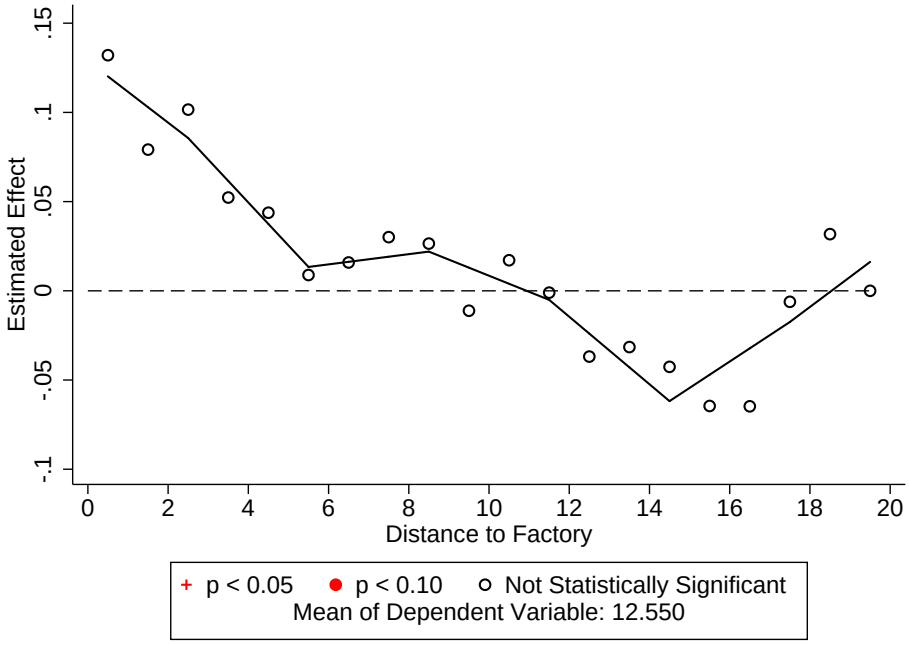



(f) *High School by Cohort*




**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 400 counterfactual factory locations.

Figure A-7: Expenditure (2001-11)

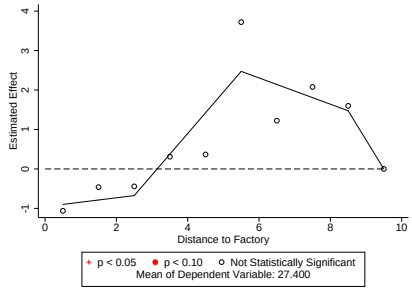



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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 400 counterfactual factory locations.

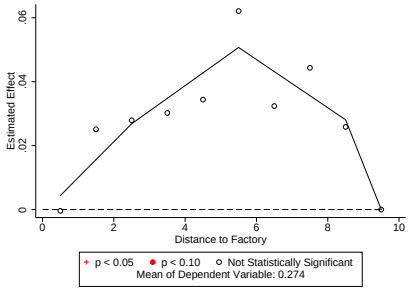
# B 10 Kilometer Sample

Figure B-1: Geographic Controls

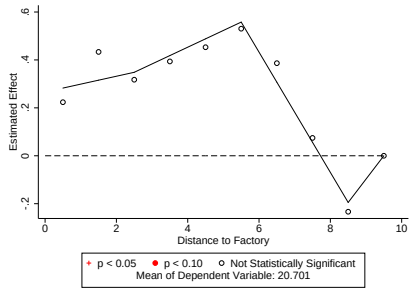
(a) *Elevation*



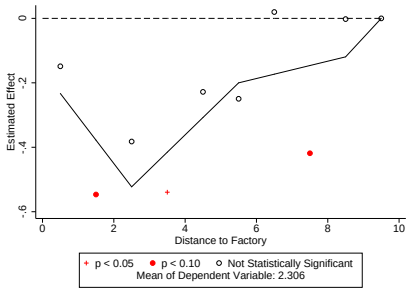

(b) *Slope*



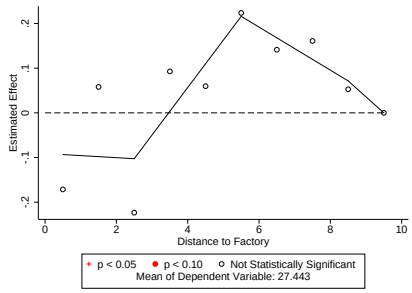

(c) *Distance to Coast*



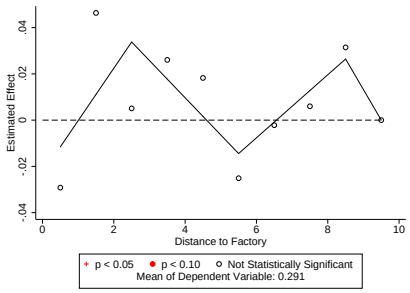

(d) *Log Flow Accumulation*



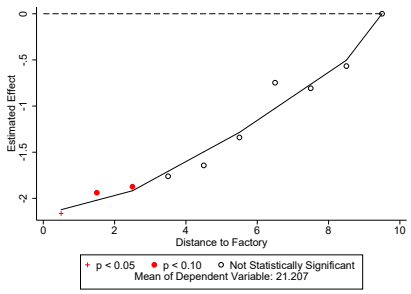

(e) *Distance to Natural Harbor*




(f) *Distance to River*



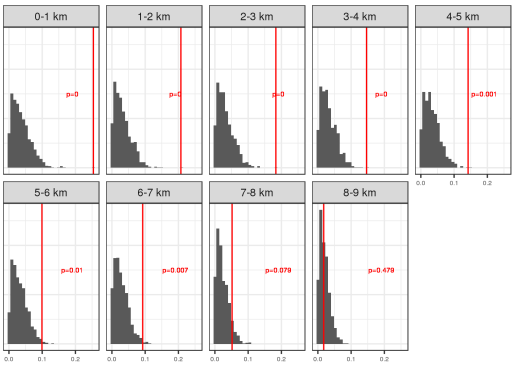

(g) *Distance to Nearest 1830 Residency Capital*



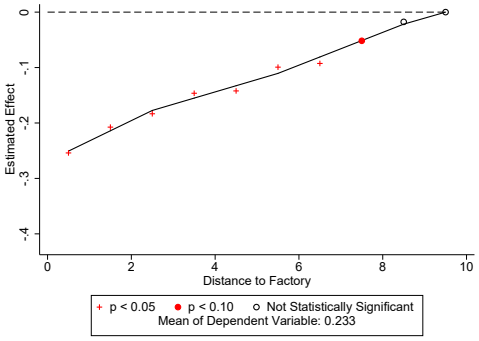

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure B-2: Share in Agriculture (Susenas 2001-11)

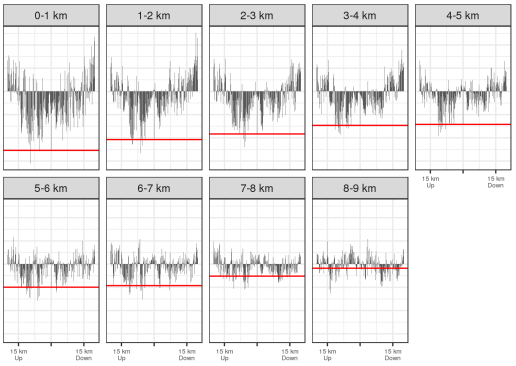
(a) *Independent Shifts*



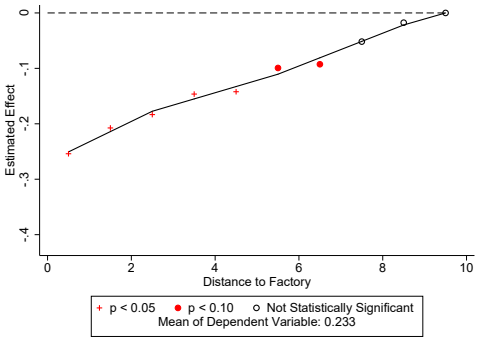

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



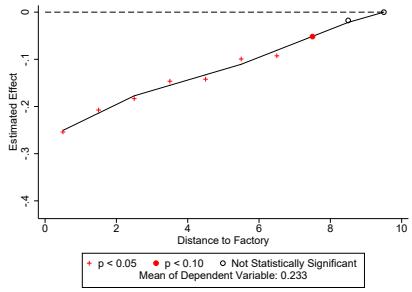

(d) *Coefficients, Common Shifts*



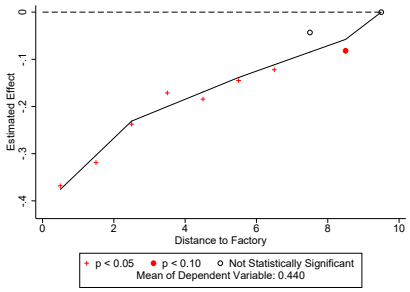

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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: Industrial Shares

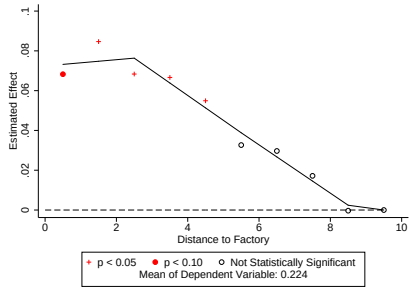
(a) *Agriculture (Susenas 2001-11)*



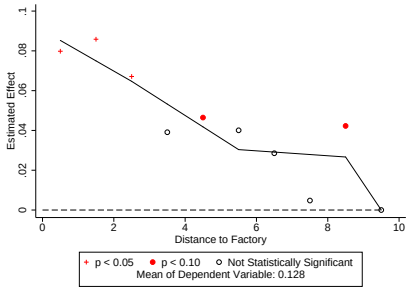

(b) *Agriculture (Census 1980)*



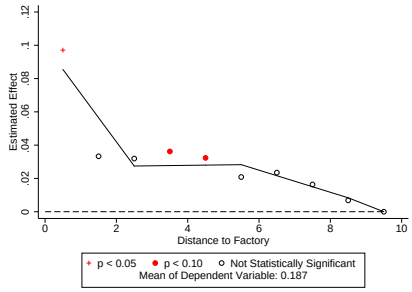

(c) *Manufacturing (Susenas 2001-11)*



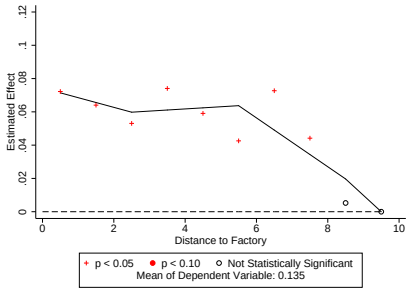

(d) *Manufacturing (Census 1980)*



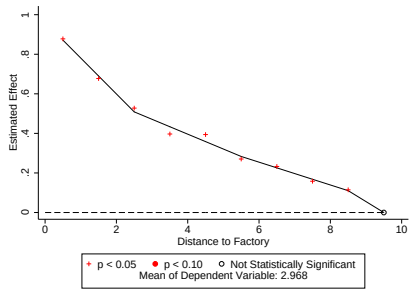

(e) *Retail (Susenas 2001-11)*



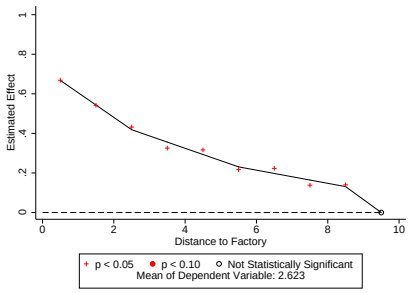

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



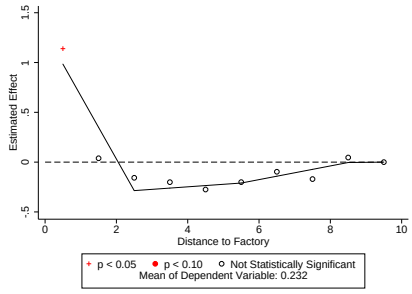

(h) *Log Population Density (PODES 2003)*



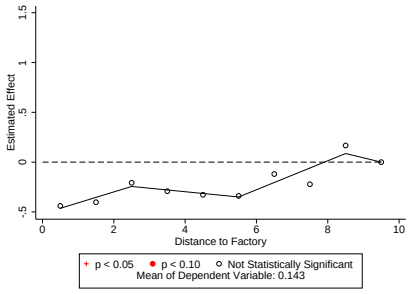

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

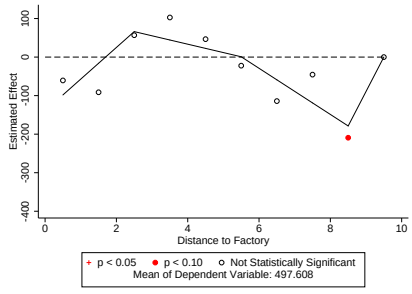
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



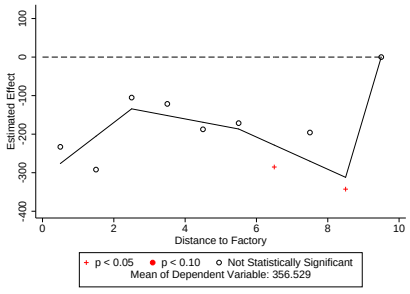

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



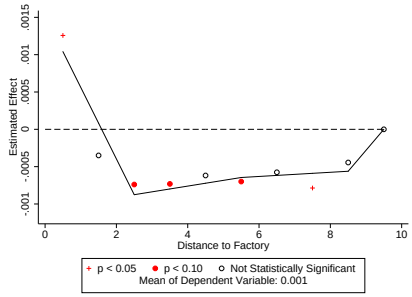

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



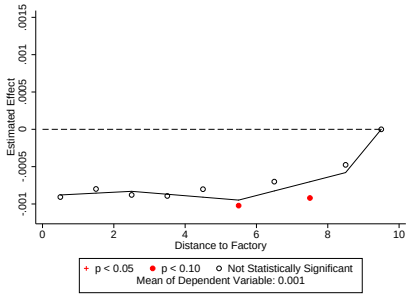

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



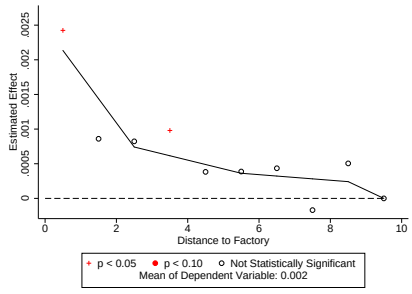

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



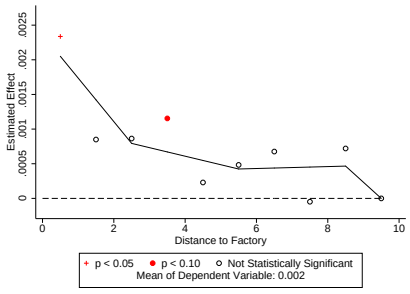

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



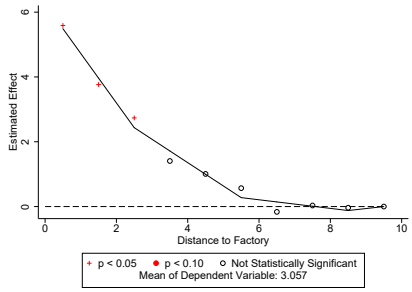

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



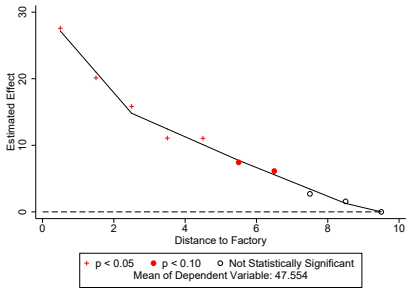

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 B-5: Public Goods

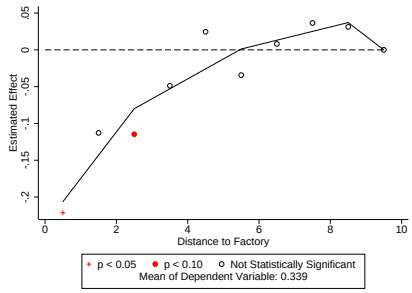
(a) *Intercity Road Density*



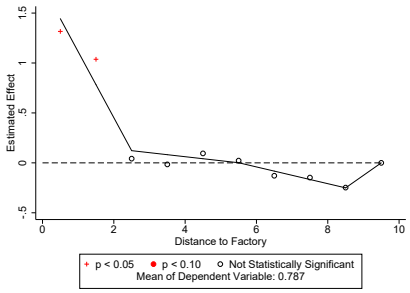

(b) *Local Road Density*



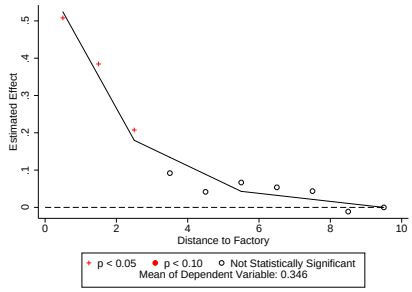

(c) *Dirt Road (PODES 1980)*



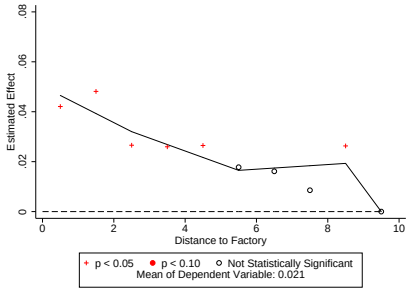

(d) *Railroad Density*



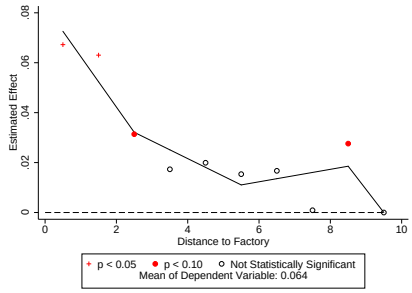

(e) *Village Has Electricity (PODES 1980)*



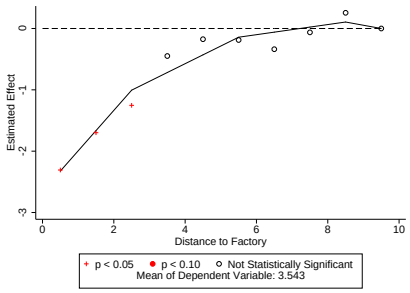

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



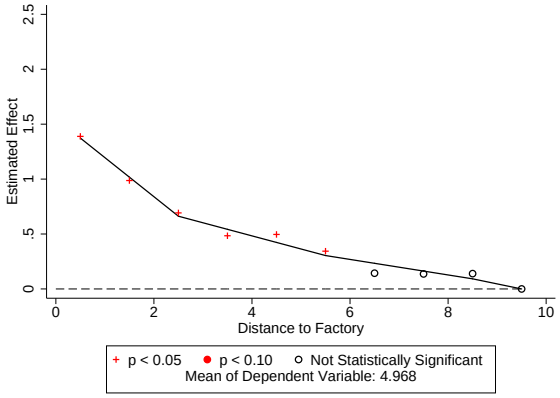

(h) *Subdistrict Capital (2011 PODES)*



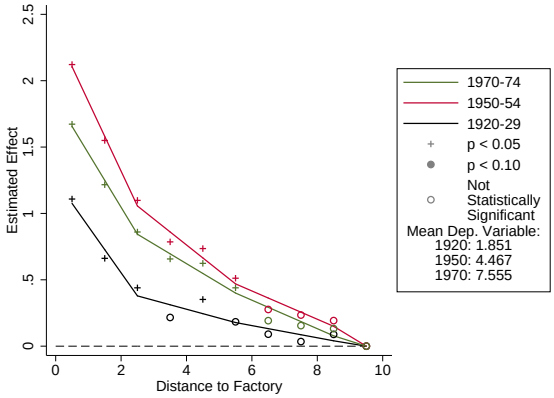

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

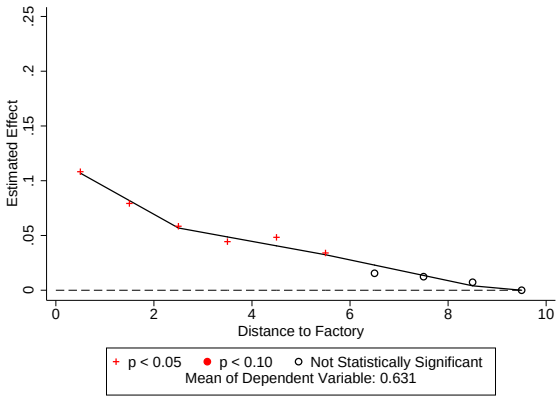
(a) *Years Education*



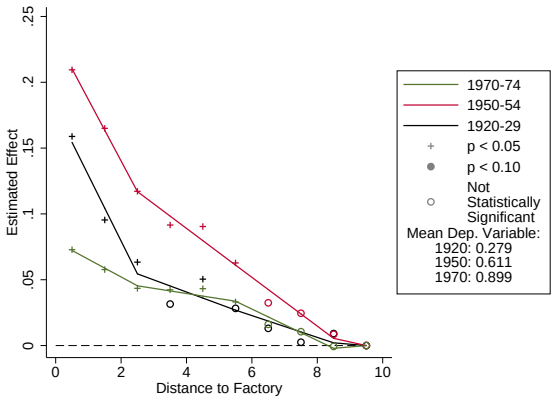

(b) *Years Education by Cohort*



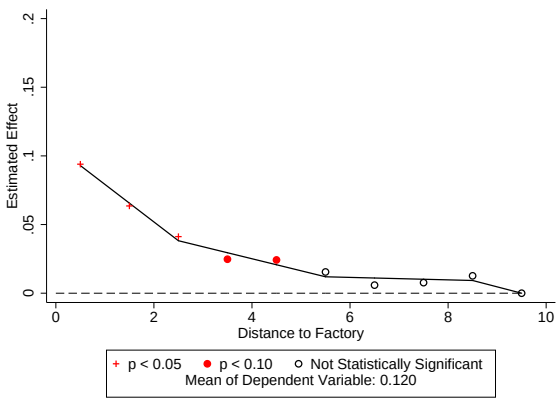

(c) *Primary*



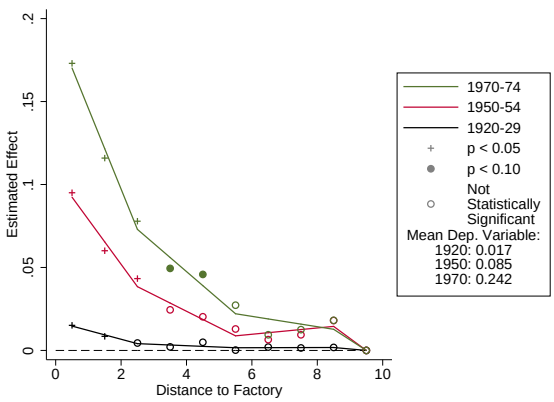

(d) *Primary by Cohort*




(e) *High School*

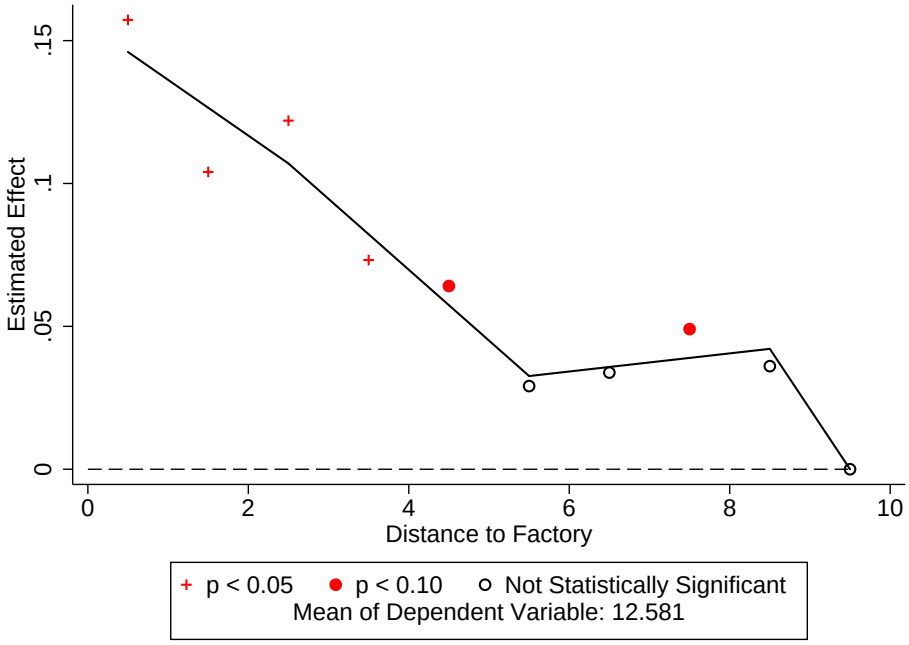



(f) *High School by Cohort*




**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-7: Expenditure (2001-11)

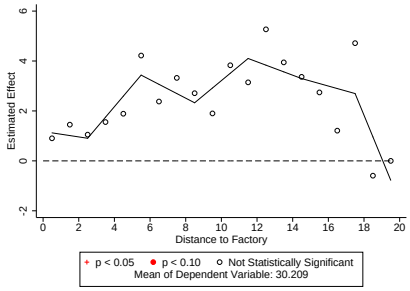



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

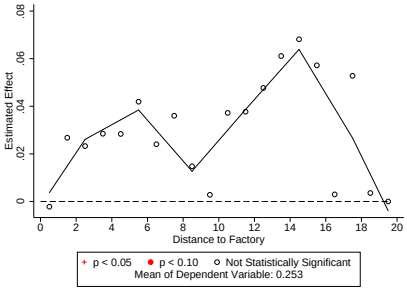
# C Sample Includes All Historical Factories

Figure C-1: Geographic Controls

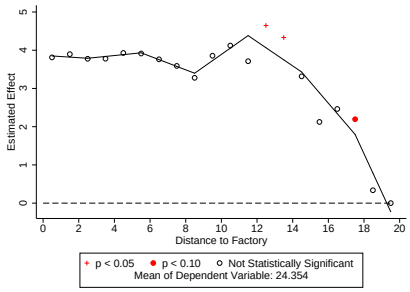
(a) *Elevation*



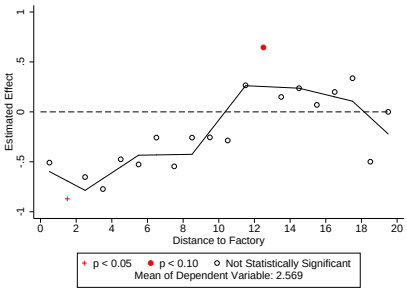

(b) *Slope*



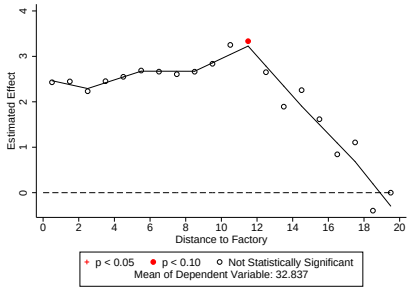

(c) *Distance to Coast*



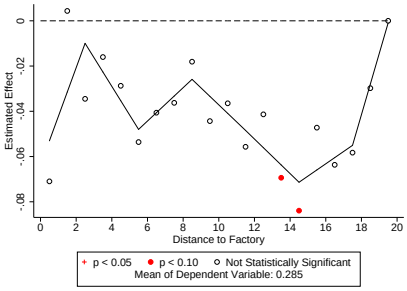

(d) *Log Flow Accumulation*



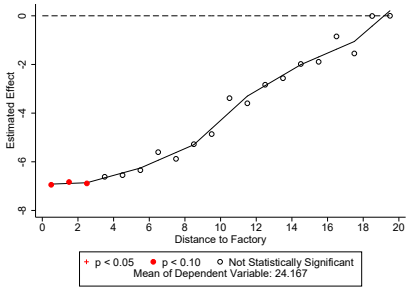

(e) *Distance to Natural Harbor*




(f) *Distance to River*



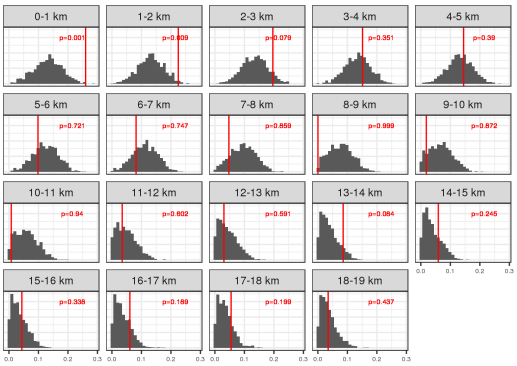

(g) *Distance to Nearest 1830 Residency Capital*



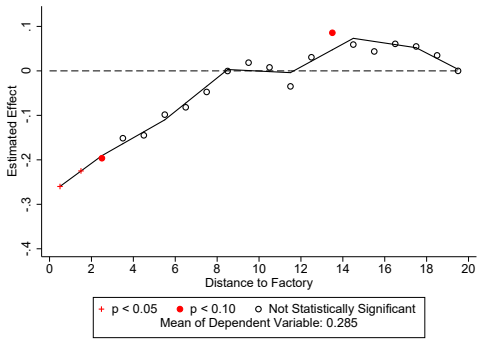

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure C-2: Share in Agriculture (Susenas 2001-11)

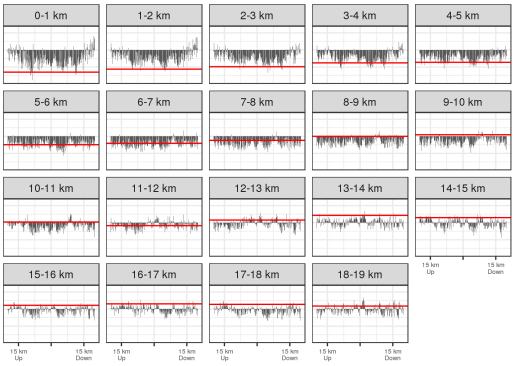
(a) *Independent Shifts*



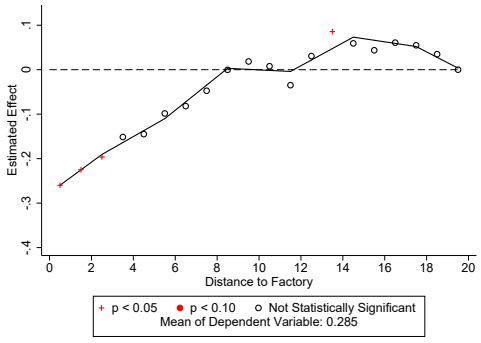

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



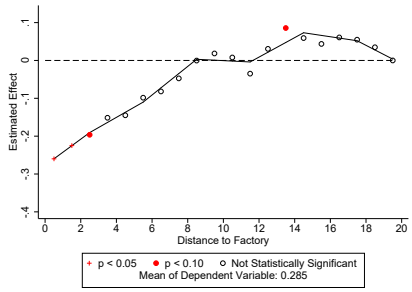

(d) *Coefficients, Common Shifts*



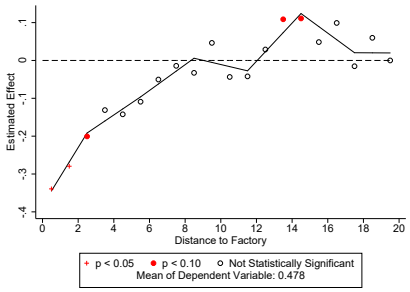

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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: Industrial Shares

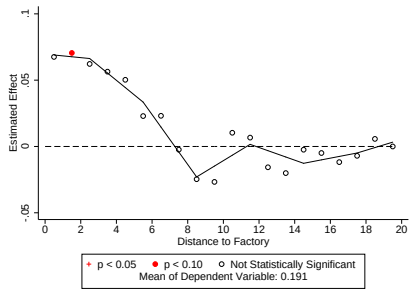
(a) *Agriculture (Susenas 2001-11)*



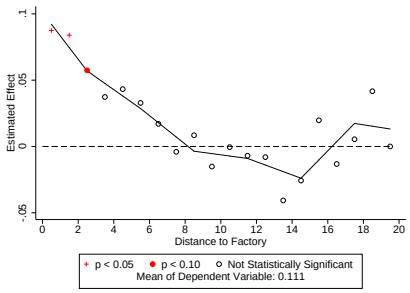

(b) *Agriculture (Census 1980)*



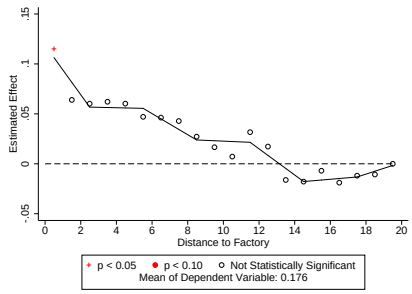

(c) *Manufacturing (Susenas 2001-11)*



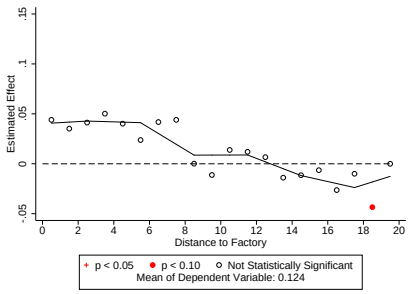

(d) *Manufacturing (Census 1980)*



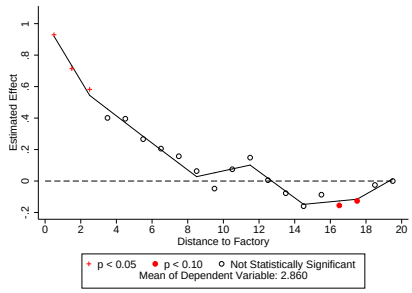

(e) *Retail (Susenas 2001-11)*



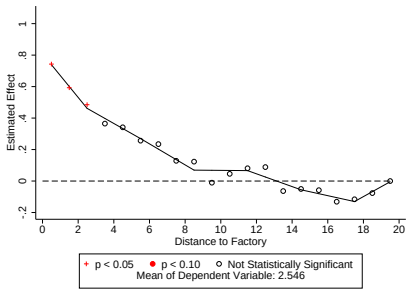

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



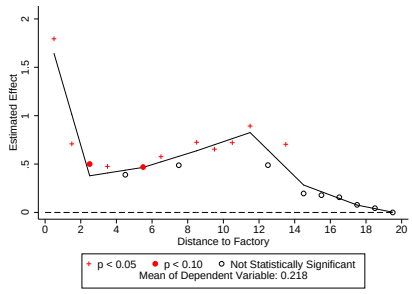

(h) *Log Population Density (PODES 2003)*



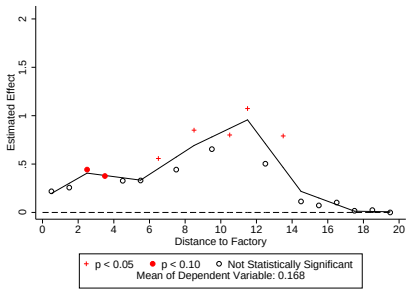

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

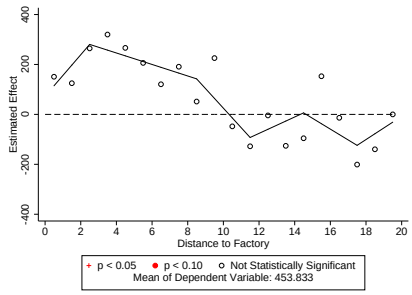
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



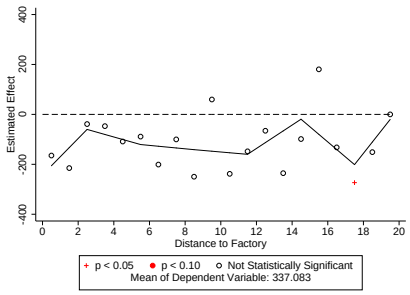

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



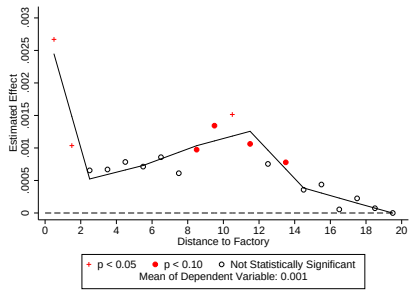

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



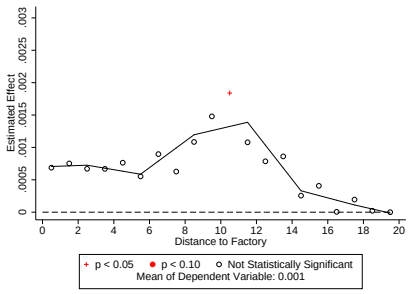

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



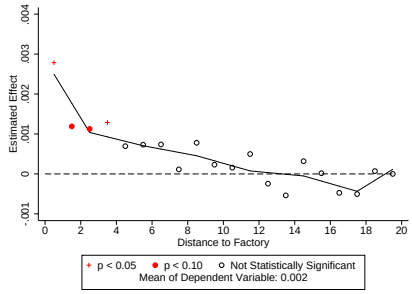

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



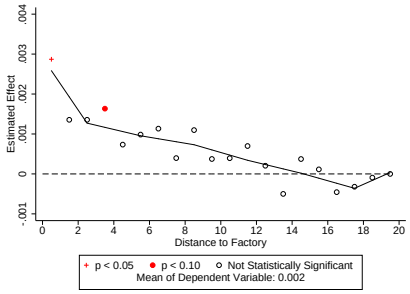

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



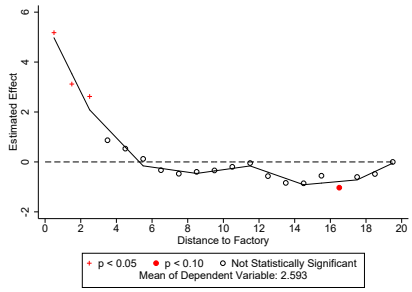

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



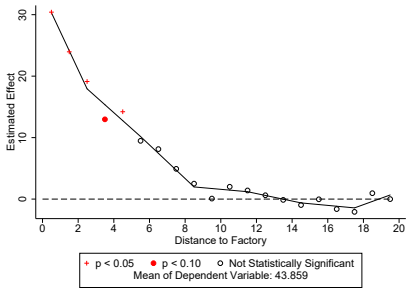

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 C-5: Public Goods

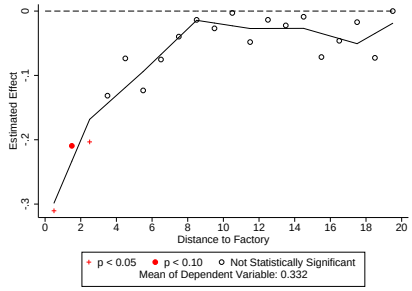
(a) *Intercity Road Density*



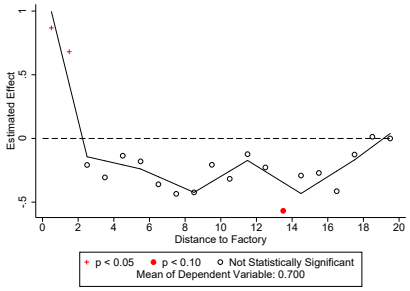

(b) *Local Road Density*



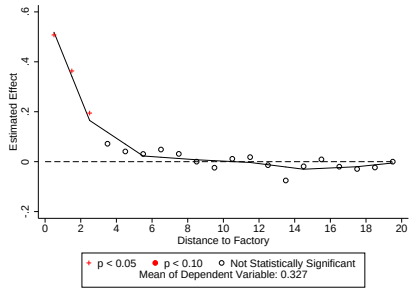

(c) *Dirt Road (PODES 1980)*



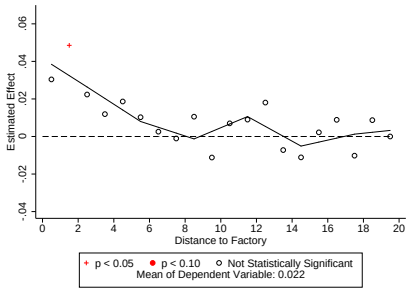

(d) *Railroad Density*



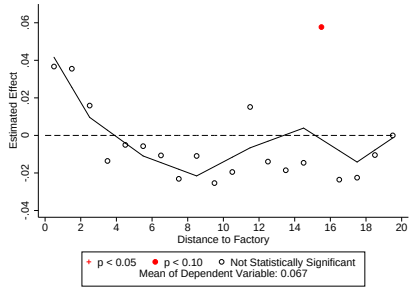

(e) *Village Has Electricity (PODES 1980)*



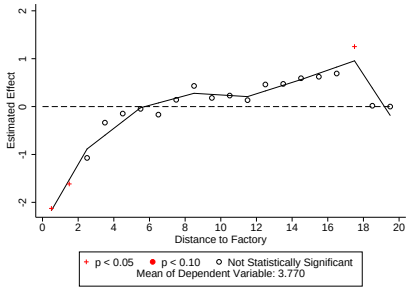

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



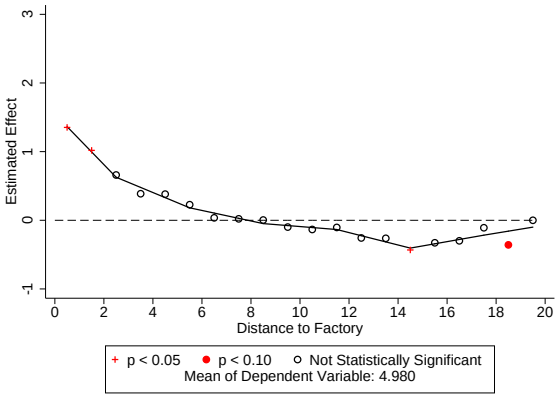

(h) *Subdistrict Capital (2011 PODES)*



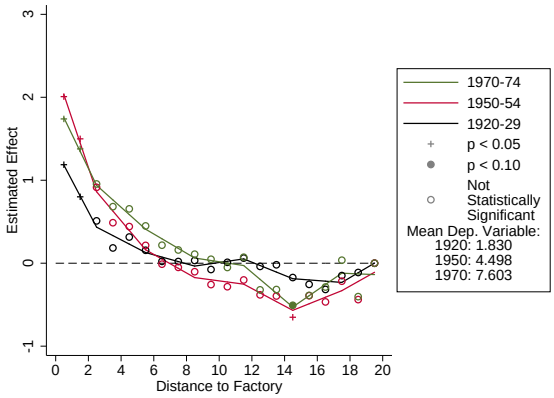

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

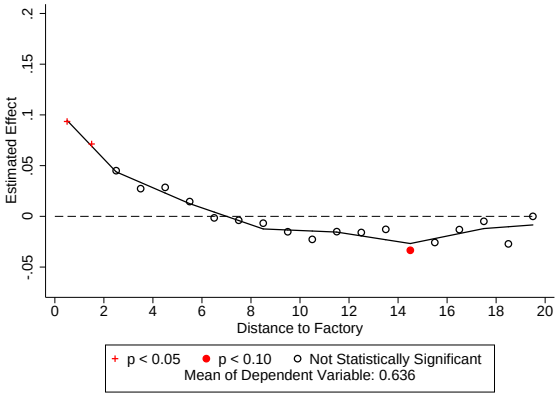
(a) *Years Education*



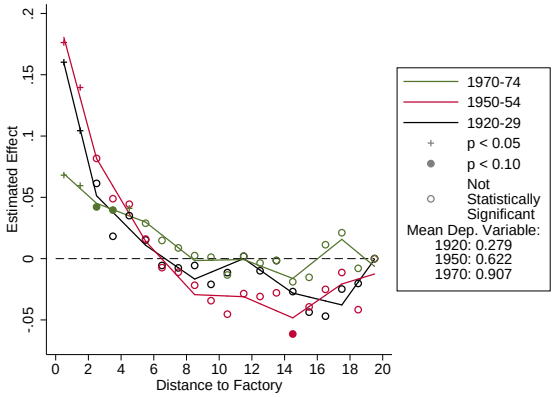

(b) *Years Education by Cohort*



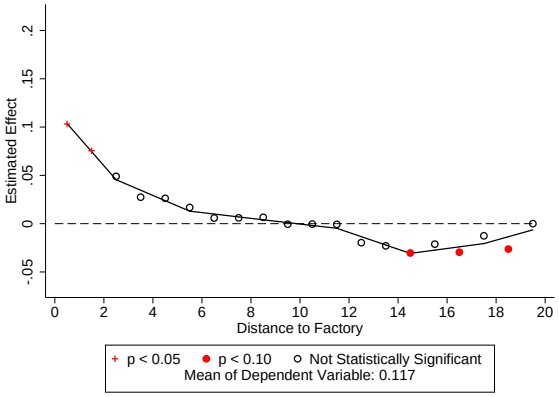

(c) *Primary*



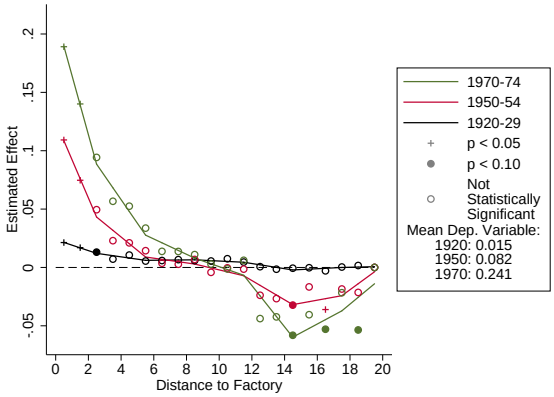

(d) *Primary by Cohort*




(e) *High School*

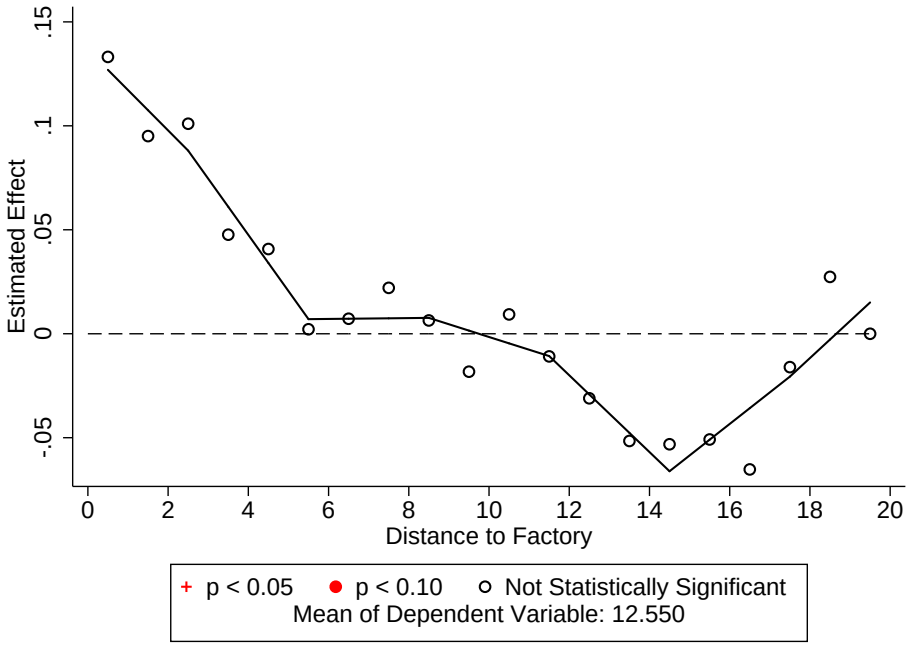



(f) *High School by Cohort*




**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-7: Expenditure (2001-11)

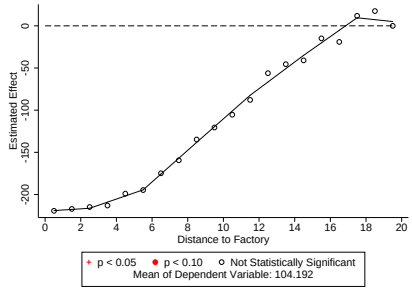



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

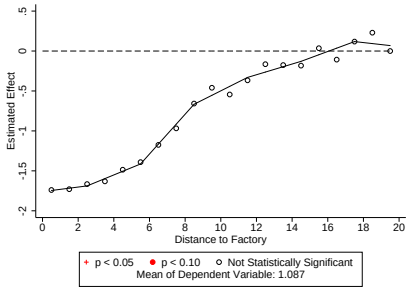
# D No Village Restrictions

Figure D-1: Geographic Controls

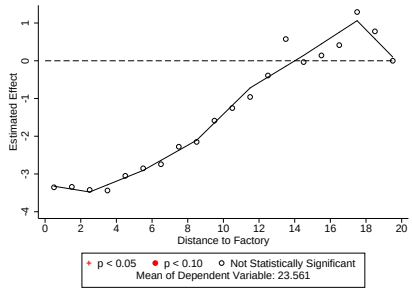
(a) *Elevation*



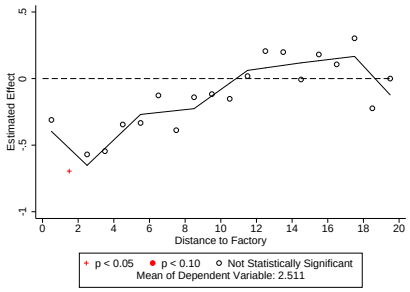

(b) *Slope*



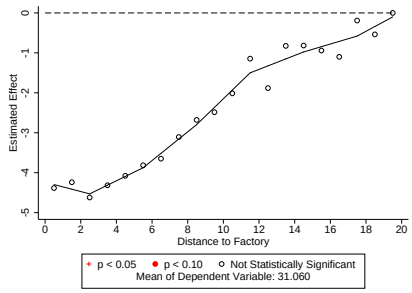

(c) *Distance to Coast*



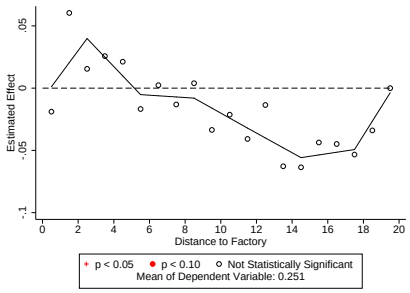

(d) *Log Flow Accumulation*



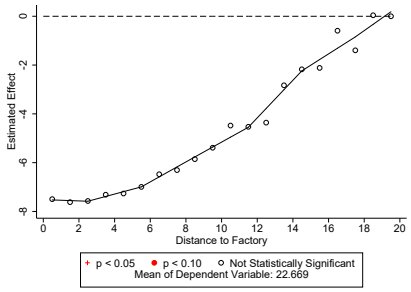

(e) *Distance to Natural Harbor*




(f) *Distance to River*



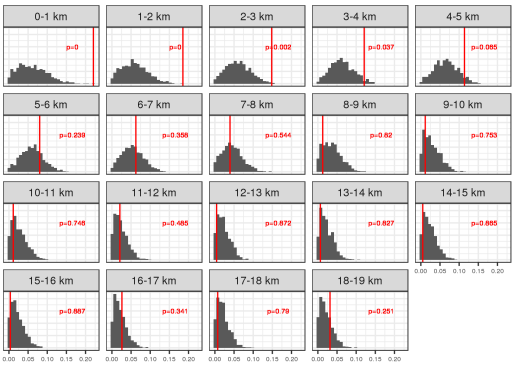

(g) *Distance to Nearest 1830 Residency Capital*



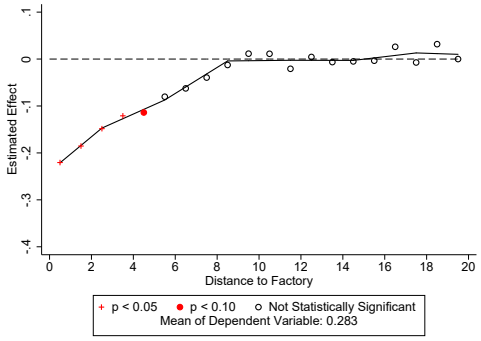

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure D-2: Share in Agriculture (Susenas 2001-11)

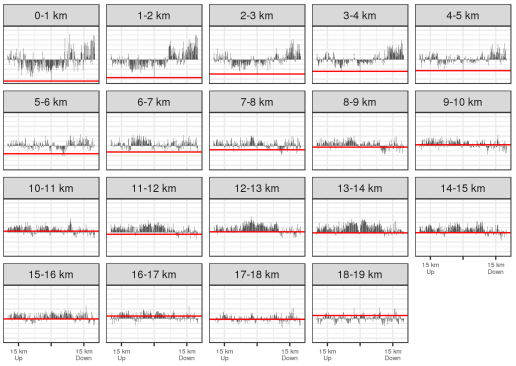
(a) *Independent Shifts*



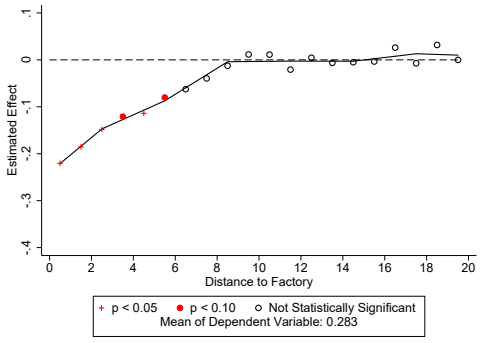

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



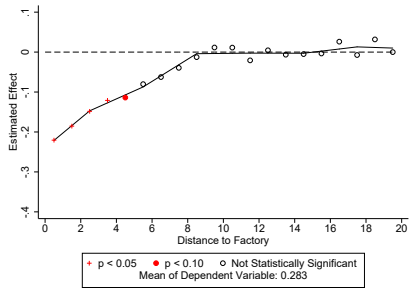

(d) *Coefficients, Common Shifts*



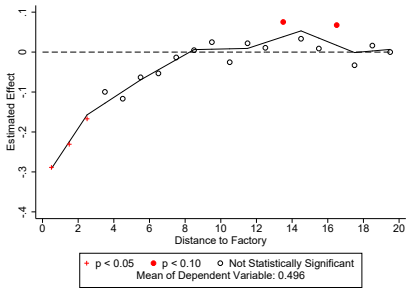

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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: Industrial Shares

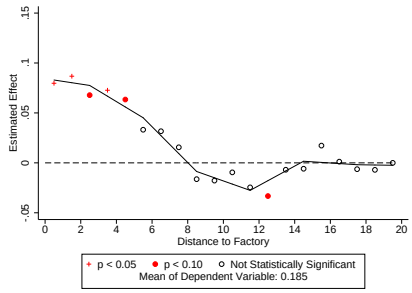
(a) *Agriculture (Susenas 2001-11)*



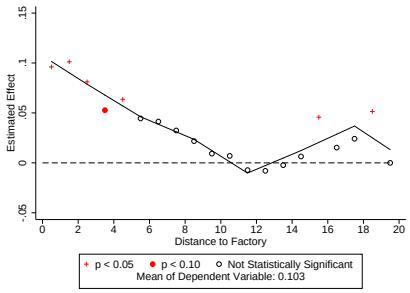

(b) *Agriculture (Census 1980)*



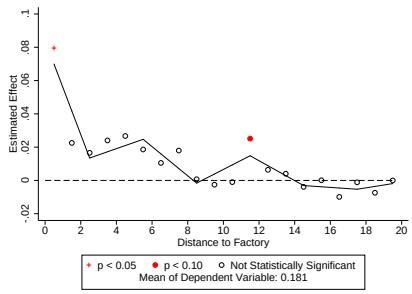

(c) *Manufacturing (Susenas 2001-11)*



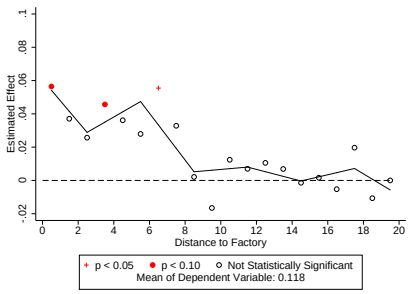

(d) *Manufacturing (Census 1980)*



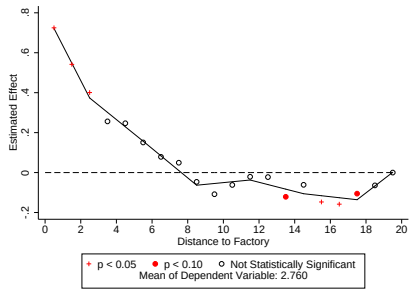

(e) *Retail (Susenas 2001-11)*



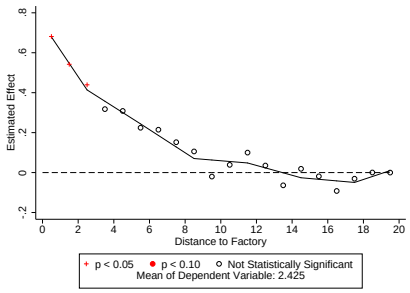

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



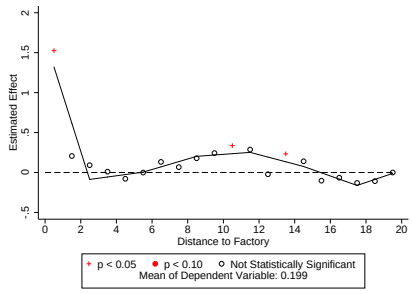

(h) *Log Population Density (PODES 2003)*



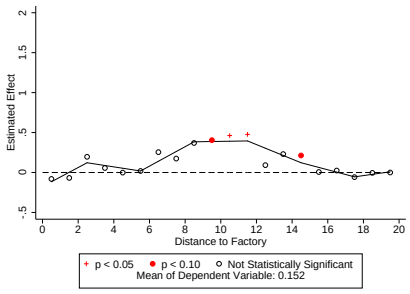

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

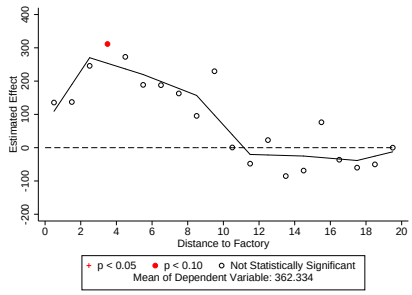
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



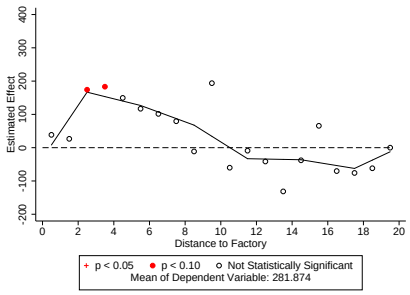

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



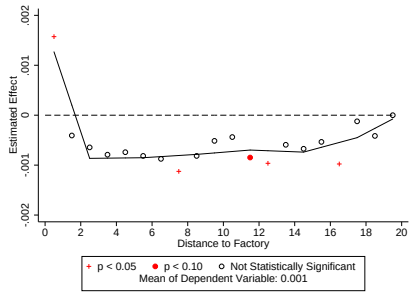

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



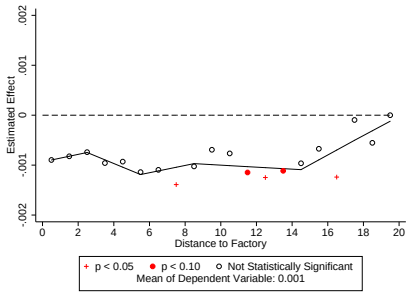

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



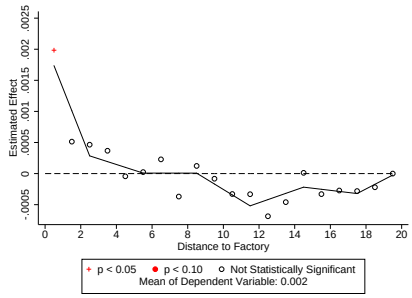

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



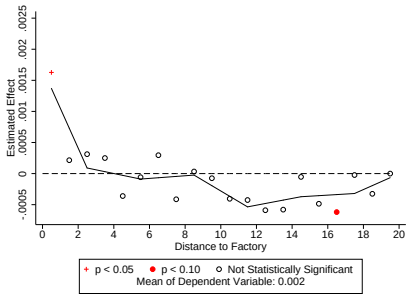

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



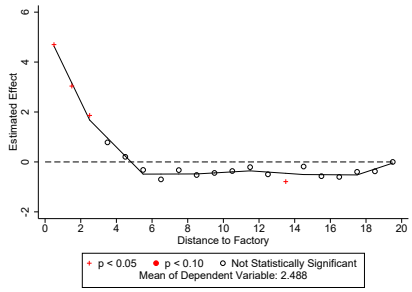

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



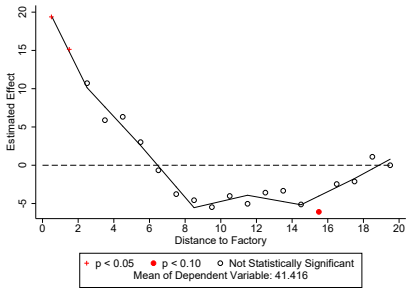

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 D-5: Public Goods

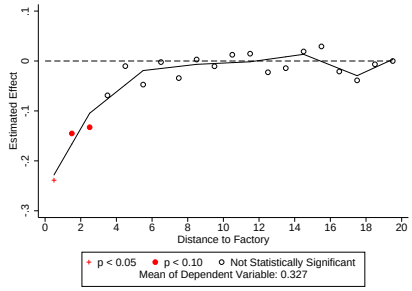
(a) *Intercity Road Density*



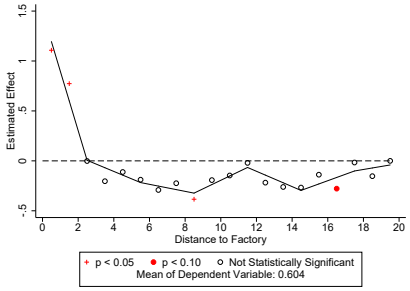

(b) *Local Road Density*



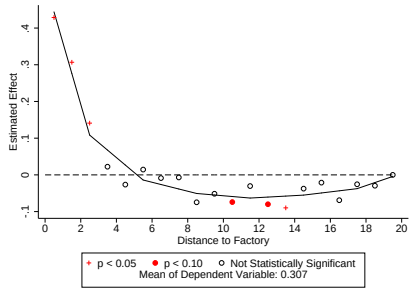

(c) *Dirt Road (PODES 1980)*



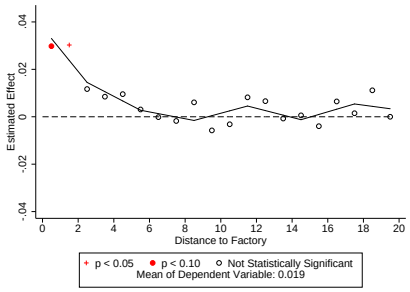

(d) *Railroad Density*



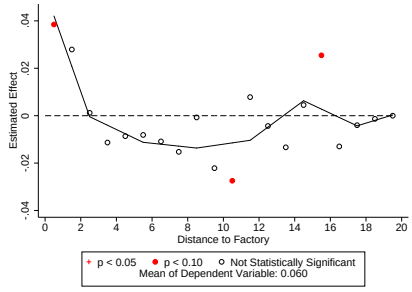

(e) *Village Has Electricity (PODES 1980)*



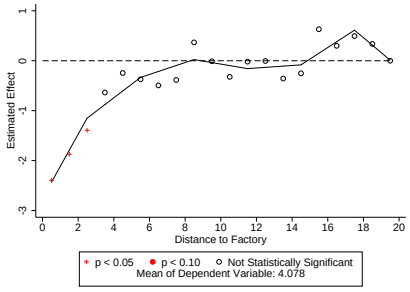

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



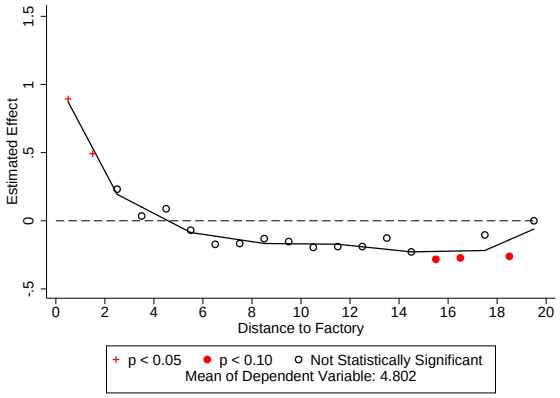

(h) *Subdistrict Capital (2011 PODES)*



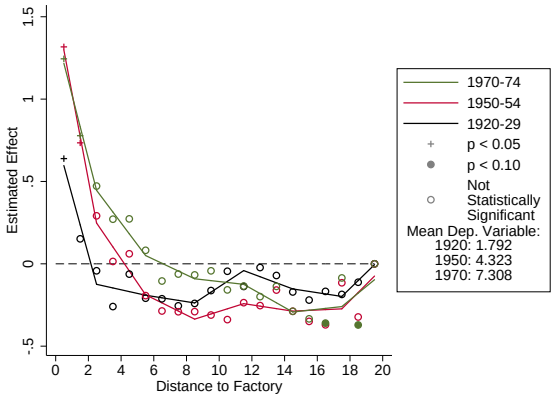

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

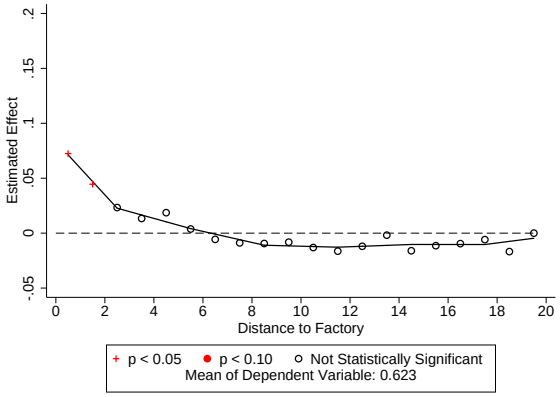
(a) *Years Education*



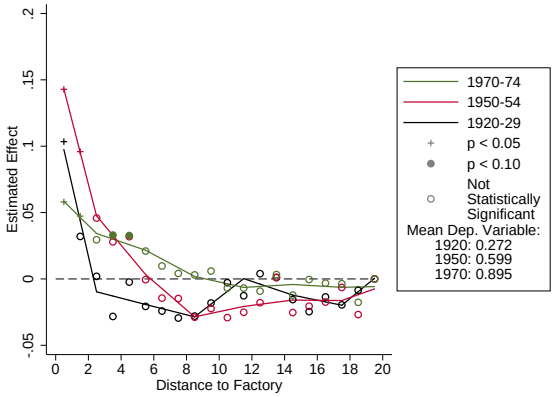

(b) *Years Education by Cohort*



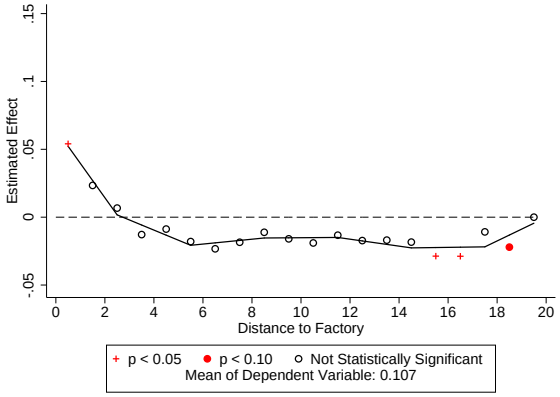

(c) *Primary*



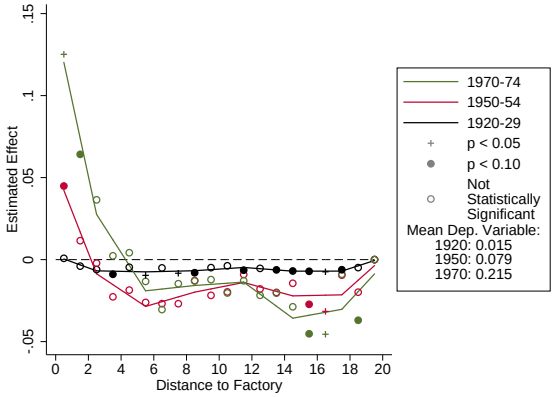

(d) *Primary by Cohort*




(e) *High School*

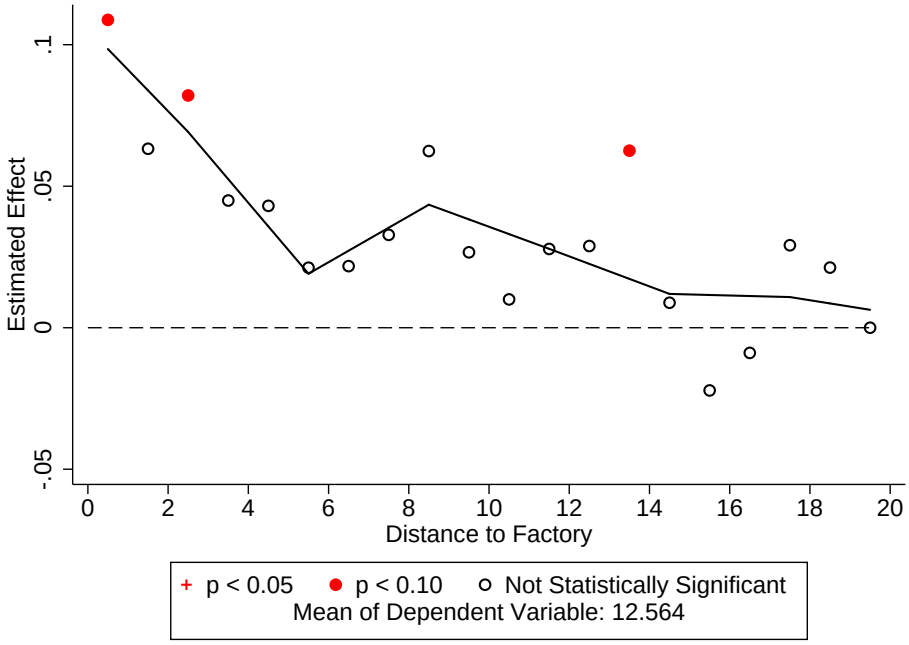



(f) *High School by Cohort*




**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-7: Expenditure (2001-11)

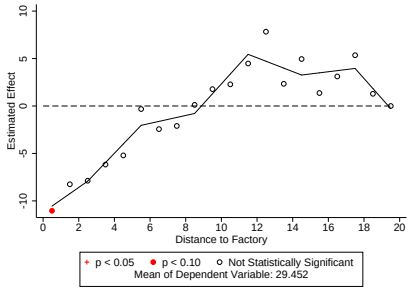



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

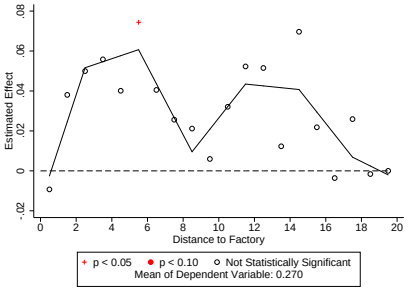
# E No Factory Fixed Effects

Figure E-1: Geographic Controls

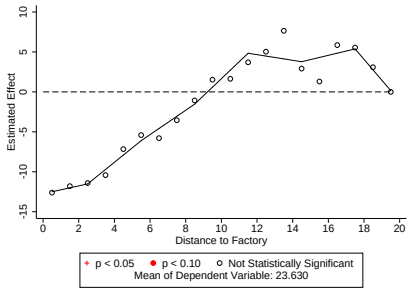
(a) *Elevation*



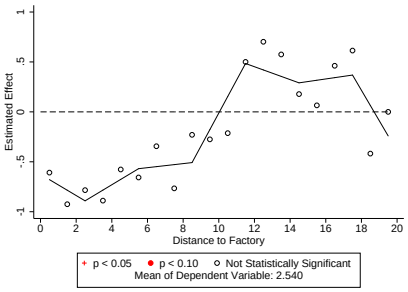

(b) *Slope*



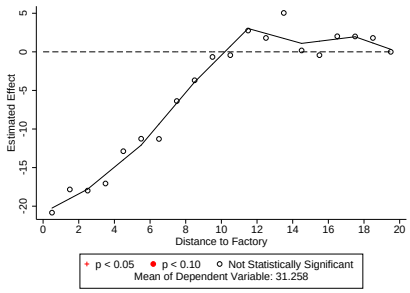

(c) *Distance to Coast*



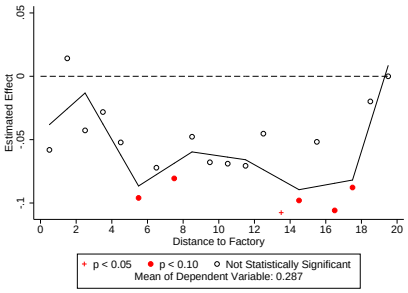

(d) *Log Flow Accumulation*



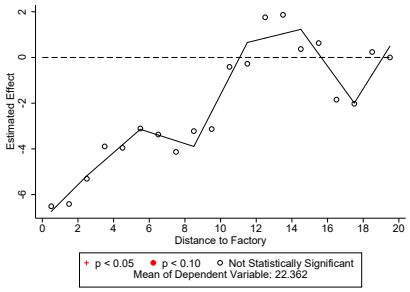

(e) *Distance to Natural Harbor*




(f) *Distance to River*



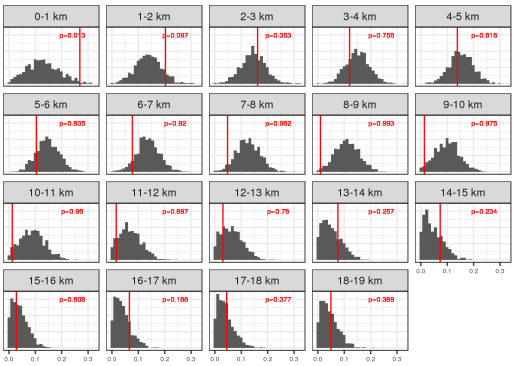

(g) *Distance to Nearest 1830 Residency Capital*



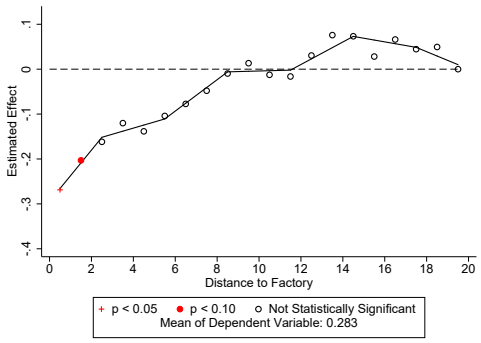

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure E-2: Share in Agriculture (Susenas 2001-11)

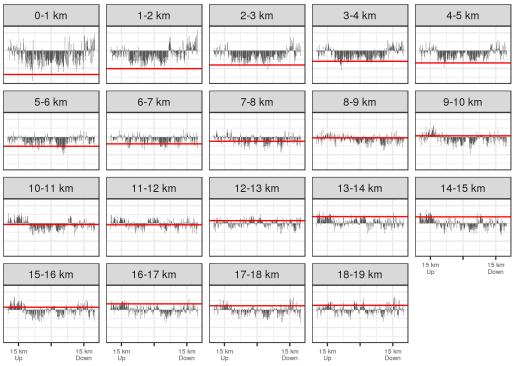
(a) *Independent Shifts*



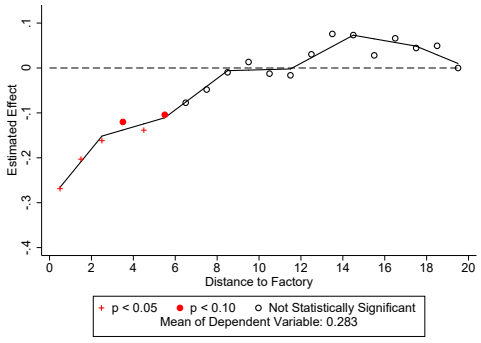

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



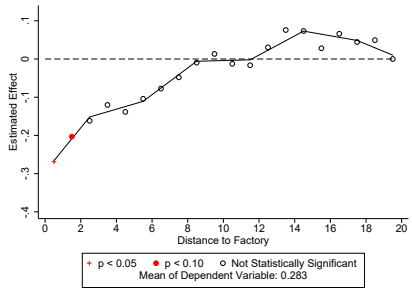

(d) *Coefficients, Common Shifts*



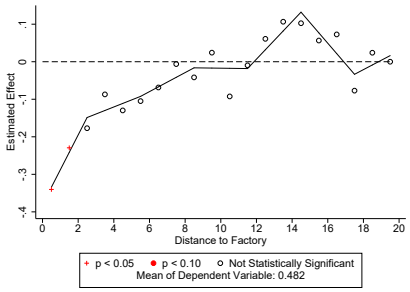

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 E-3: Industrial Shares

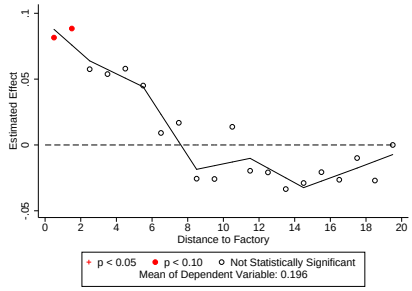
(a) *Agriculture (Susenas 2001-11)*



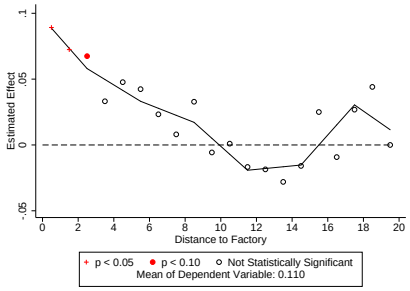

(b) *Agriculture (Census 1980)*



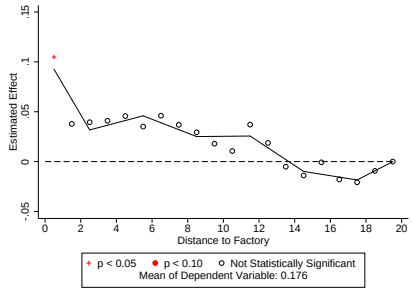

(c) *Manufacturing (Susenas 2001-11)*



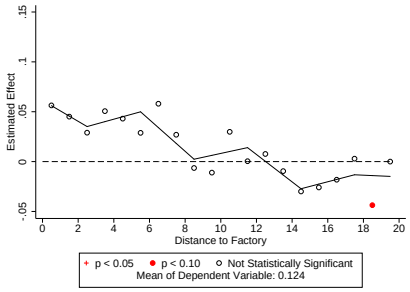

(d) *Manufacturing (Census 1980)*



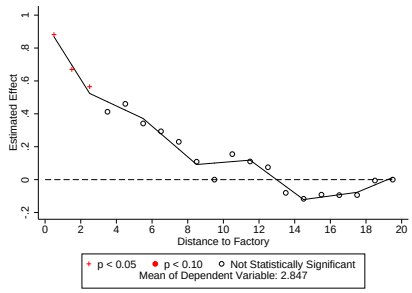

(e) *Retail (Susenas 2001-11)*



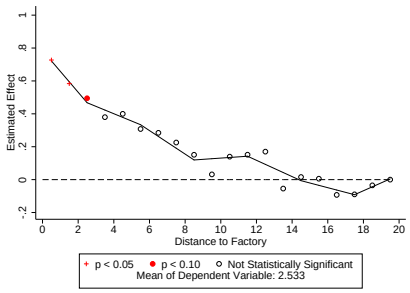

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



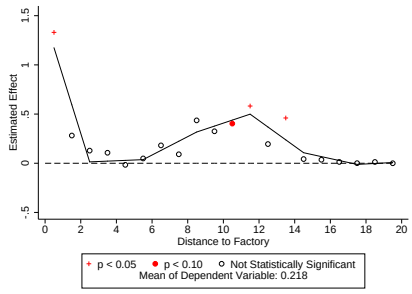

(h) *Log Population Density (PODES 2003)*



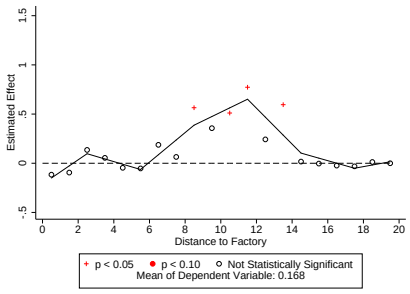

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

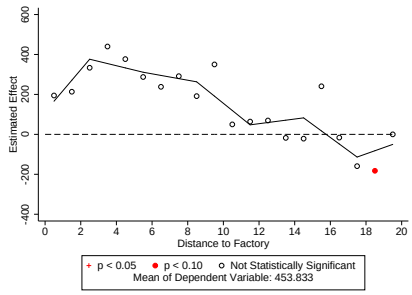
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



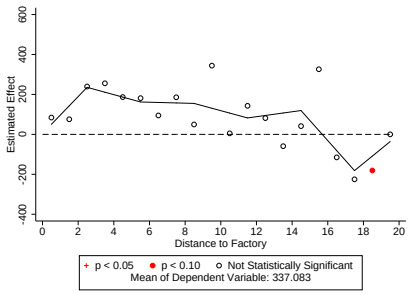

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



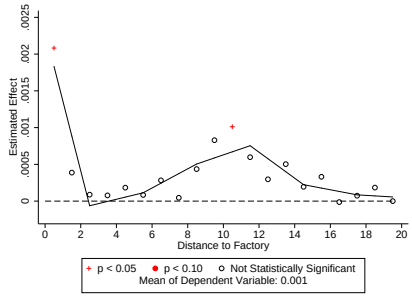

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



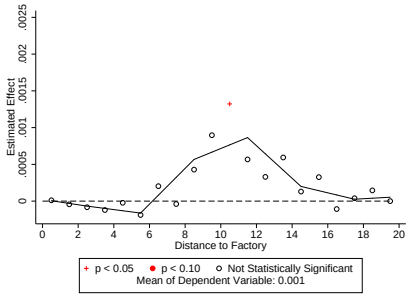

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



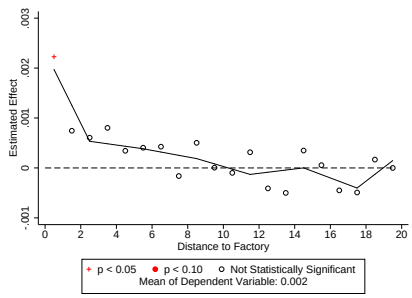

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



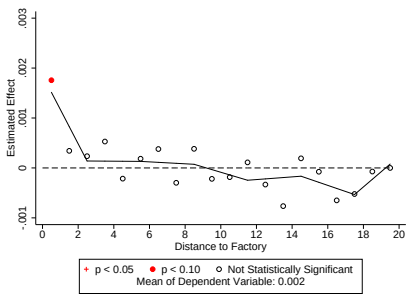

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



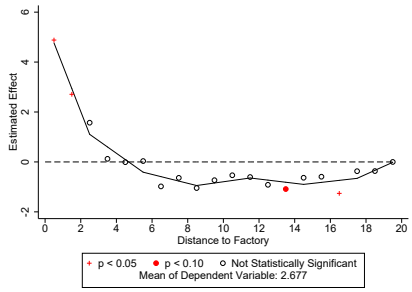

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



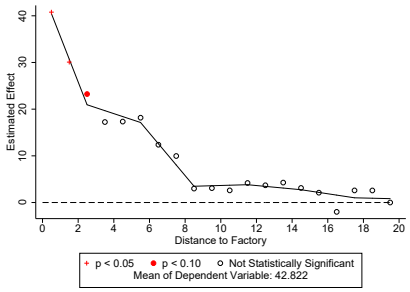

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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-5: Public Goods

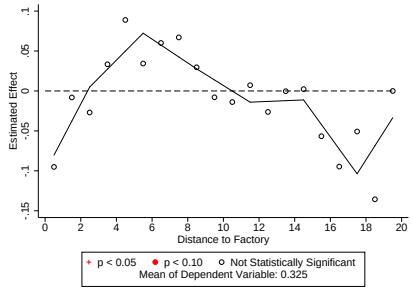
(a) *Intercity Road Density*



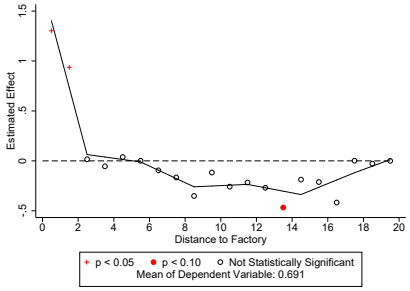

(b) *Local Road Density*



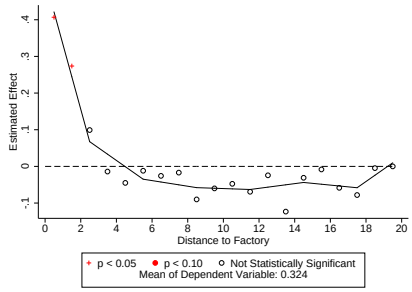

(c) *Dirt Road (PODES 1980)*



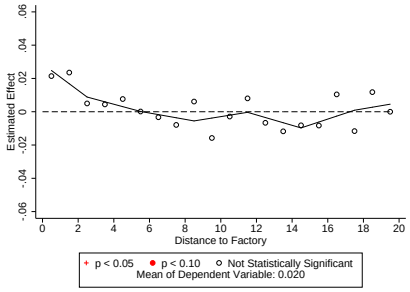

(d) *Railroad Density*



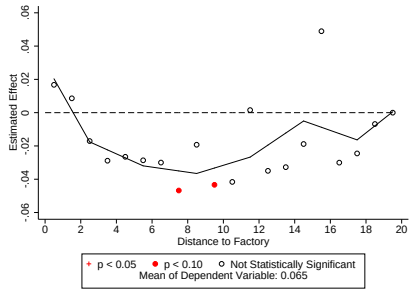

(e) *Village Has Electricity (PODES 1980)*



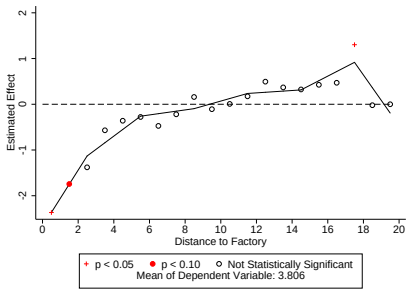

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



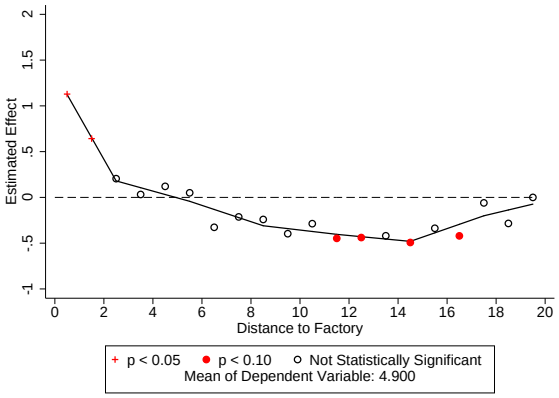

(h) *Subdistrict Capital (2011 PODES)*



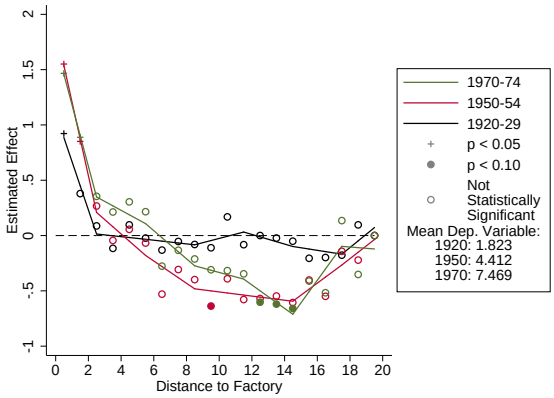

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

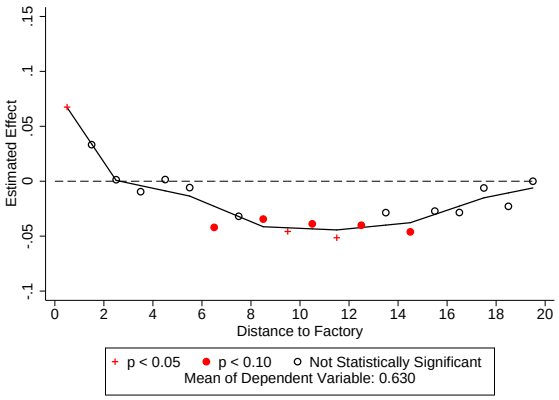
(a) *Years Education*



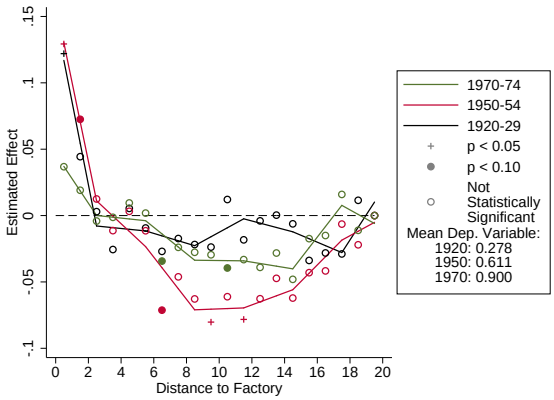

(b) *Years Education by Cohort*



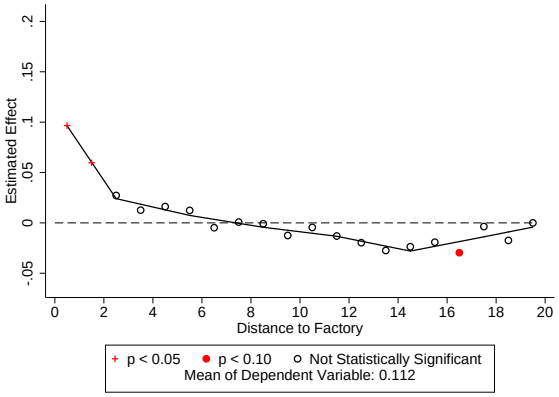

(c) *Primary*



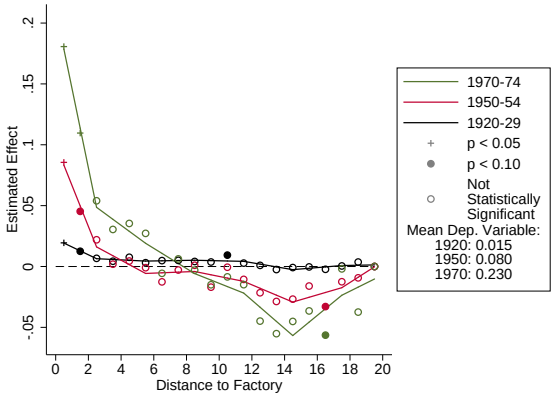

(d) *Primary by Cohort*




(e) *High School*

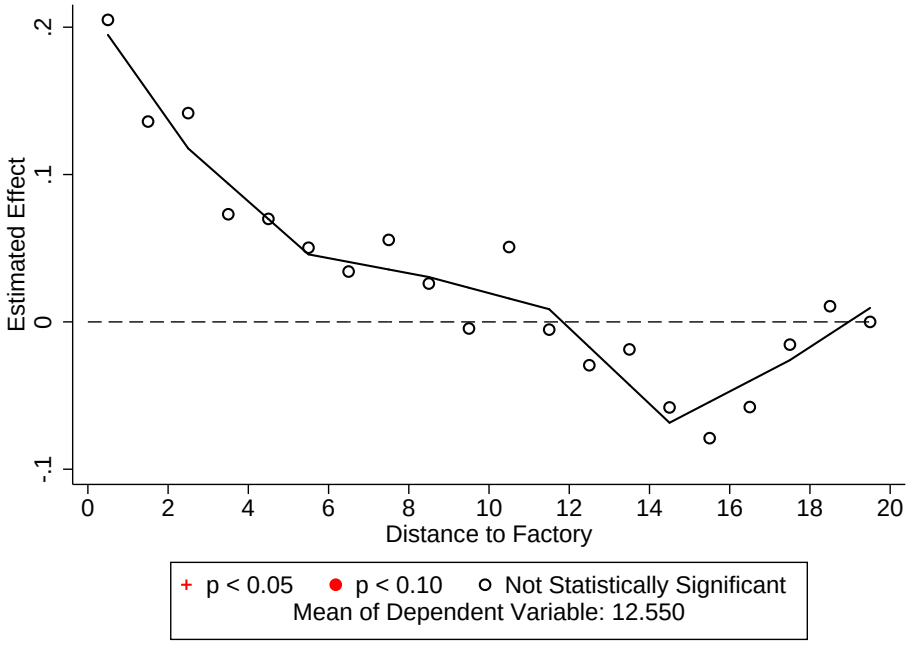



(f) *High School by Cohort*




**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 E-7: Expenditure (2001-11)

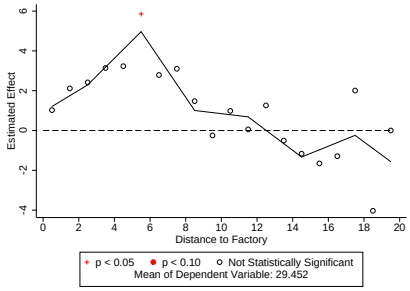



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

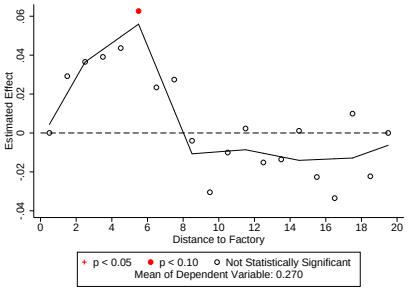
# F Coefficients that Subtract Placebo Means

Figure F-1: Geographic Controls

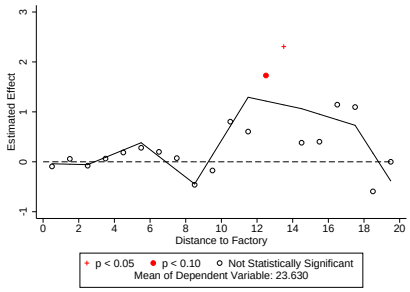
(a) *Elevation*



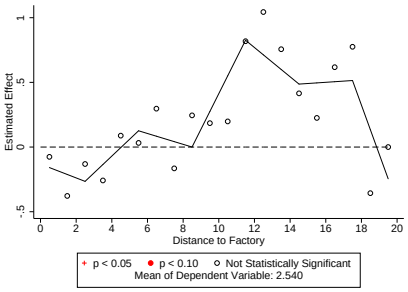

(b) *Slope*



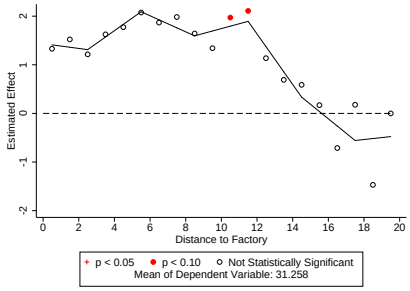

(c) *Distance to Coast*



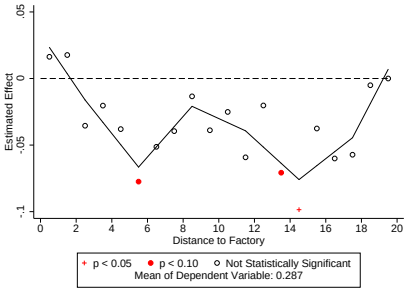

(d) *Log Flow Accumulation*



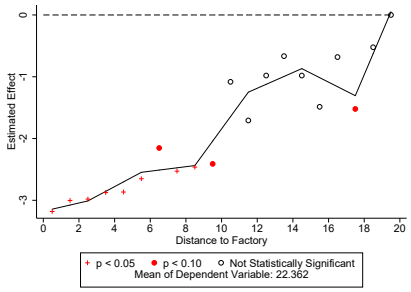

(e) *Distance to Natural Harbor*




(f) *Distance to River*



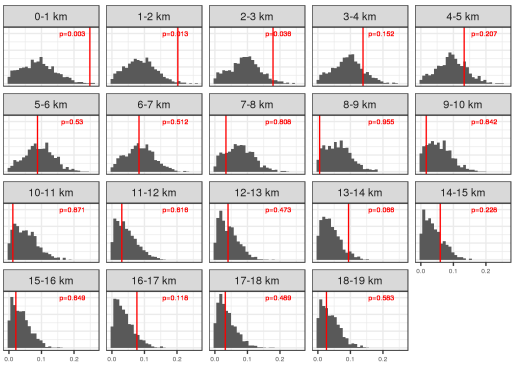

(g) *Distance to Nearest 1830 Residency Capital*



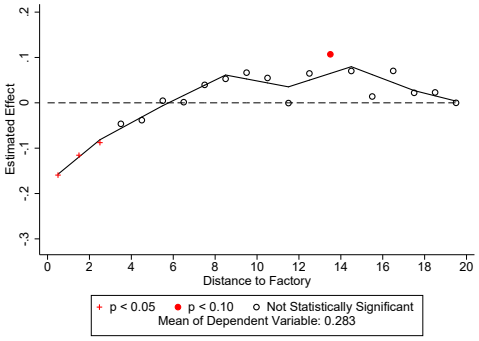

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure F-2: Share in Agriculture (Susenas 2001-11)

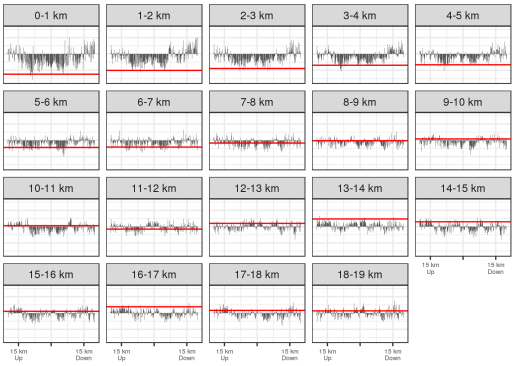
(a) *Independent Shifts*



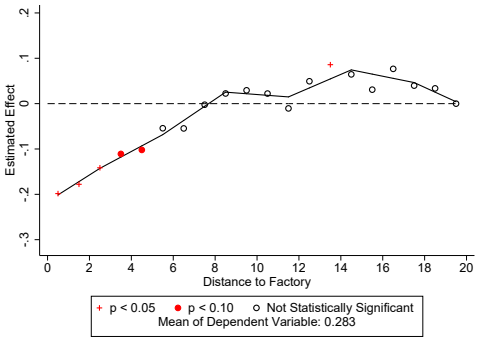

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



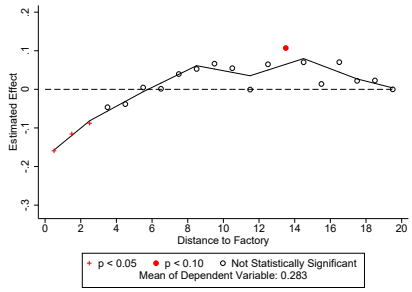

(d) *Coefficients, Common Shifts*



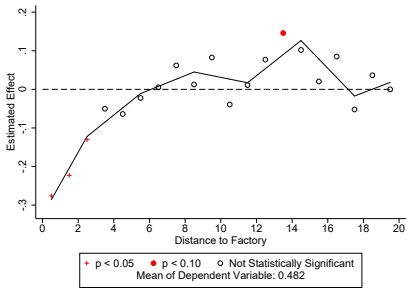

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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 F-3: Industrial Shares

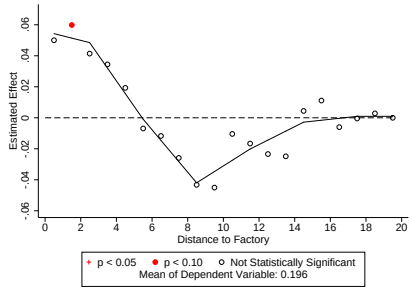
(a) *Agriculture (Susenas 2001-11)*



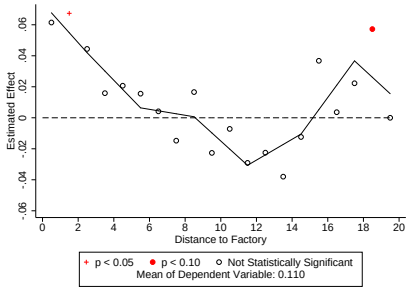

(b) *Agriculture (Census 1980)*



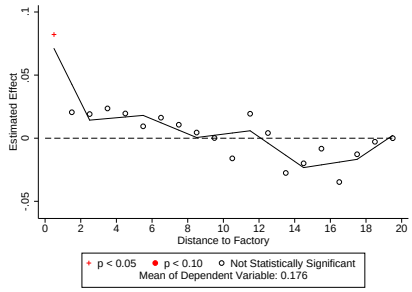

(c) *Manufacturing (Susenas 2001-11)*



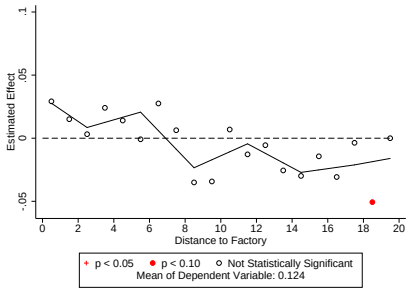

(d) *Manufacturing (Census 1980)*



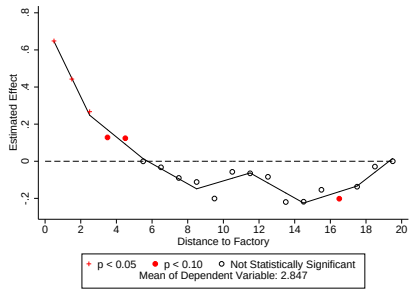

(e) *Retail (Susenas 2001-11)*



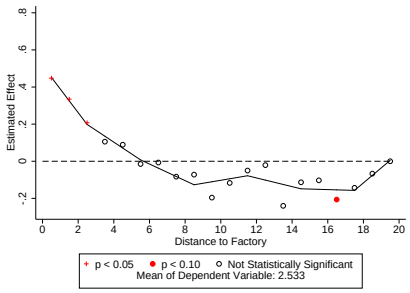

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



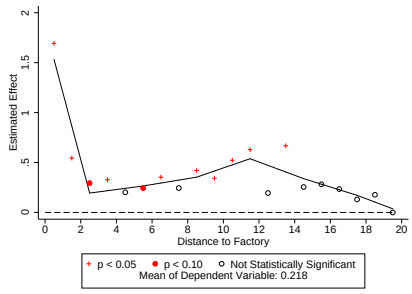

(h) *Log Population Density (PODES 2003)*



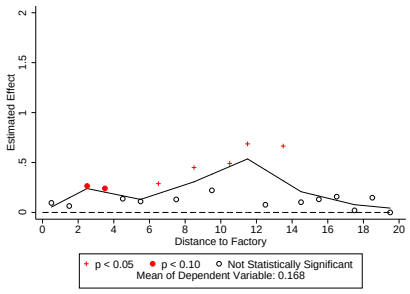

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

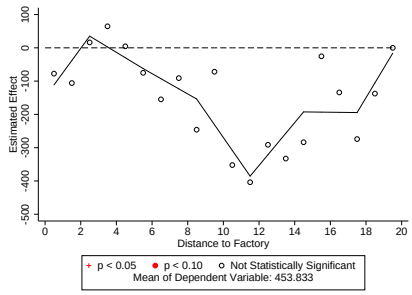
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



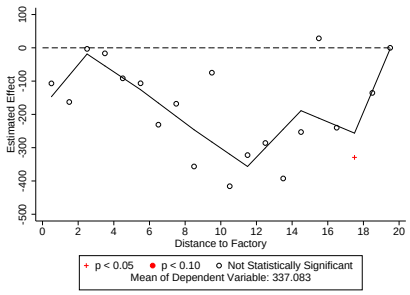

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



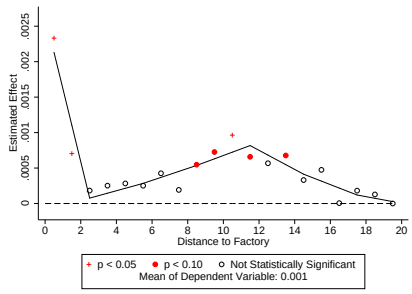

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



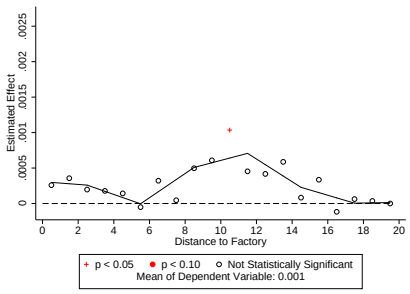

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



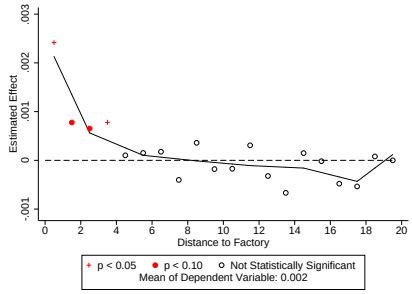

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



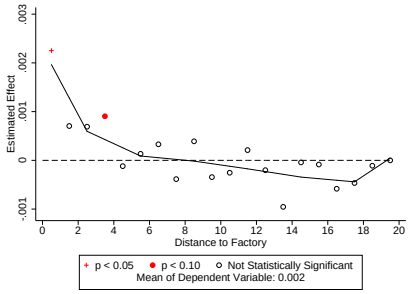

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



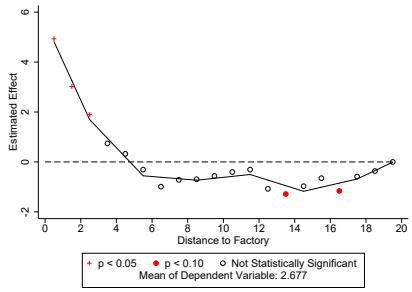

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



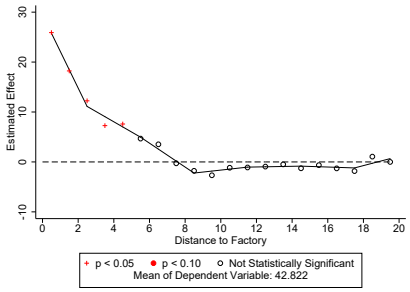

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 F-5: Public Goods

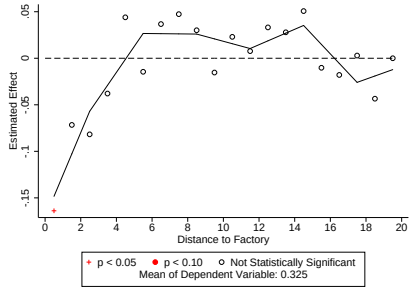
(a) *Intercity Road Density*



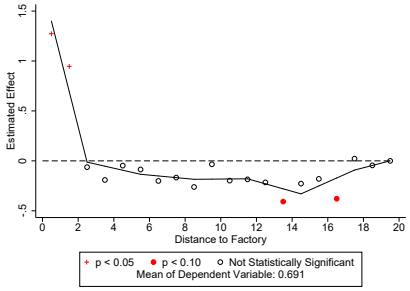

(b) *Local Road Density*



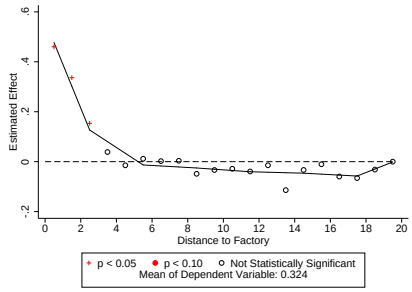

(c) *Dirt Road (PODES 1980)*



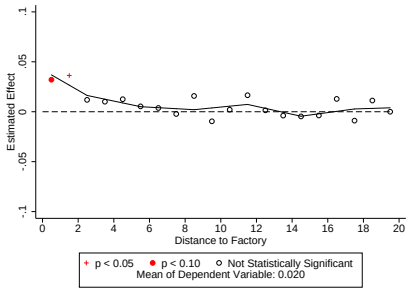

(d) *Railroad Density*



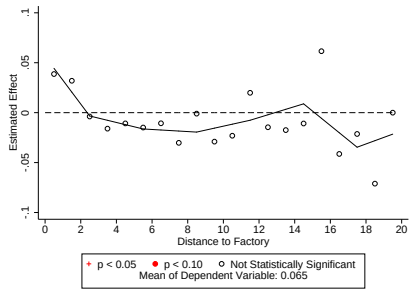

(e) *Village Has Electricity (PODES 1980)*



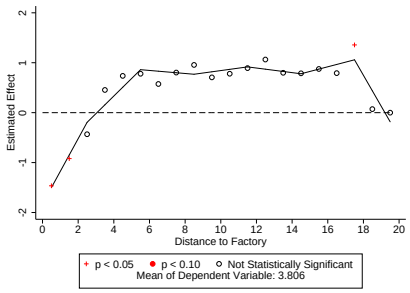

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



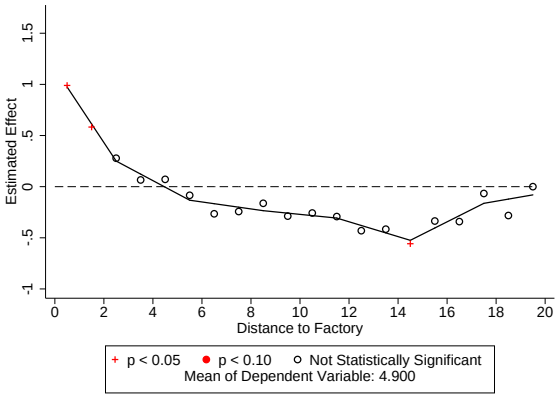

(h) *Subdistrict Capital (2011 PODES)*



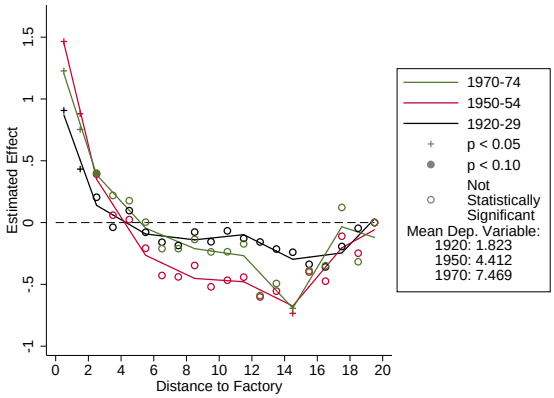

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

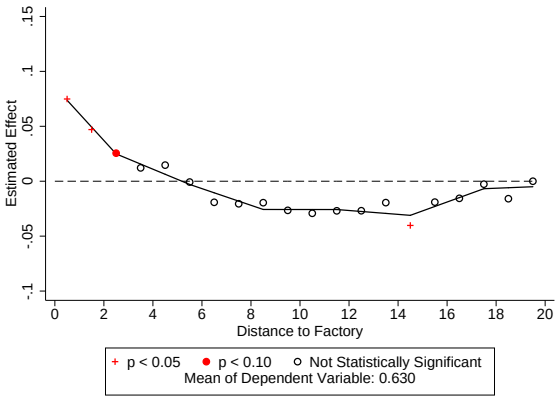
(a) *Years Education*



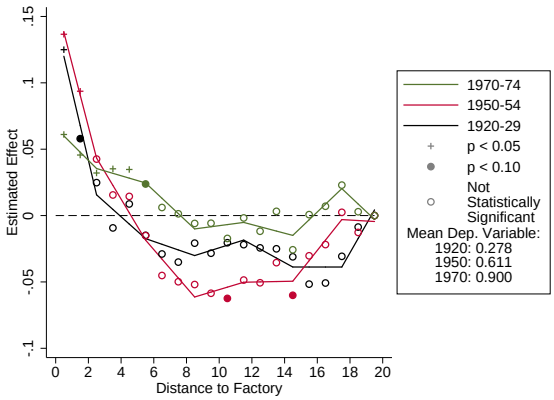

(b) *Years Education by Cohort*



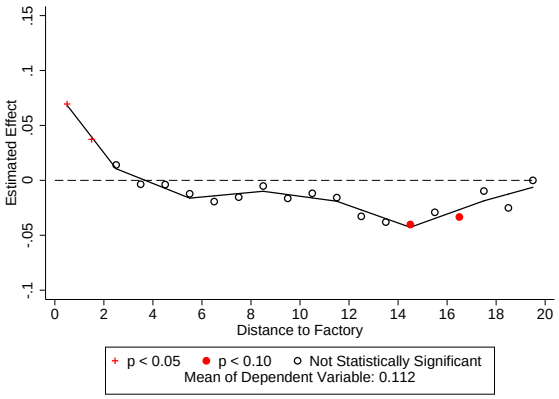

(c) *Primary*



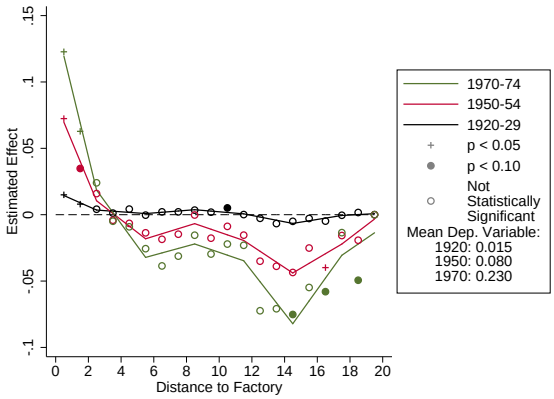

(d) *Primary by Cohort*




(e) *High School*

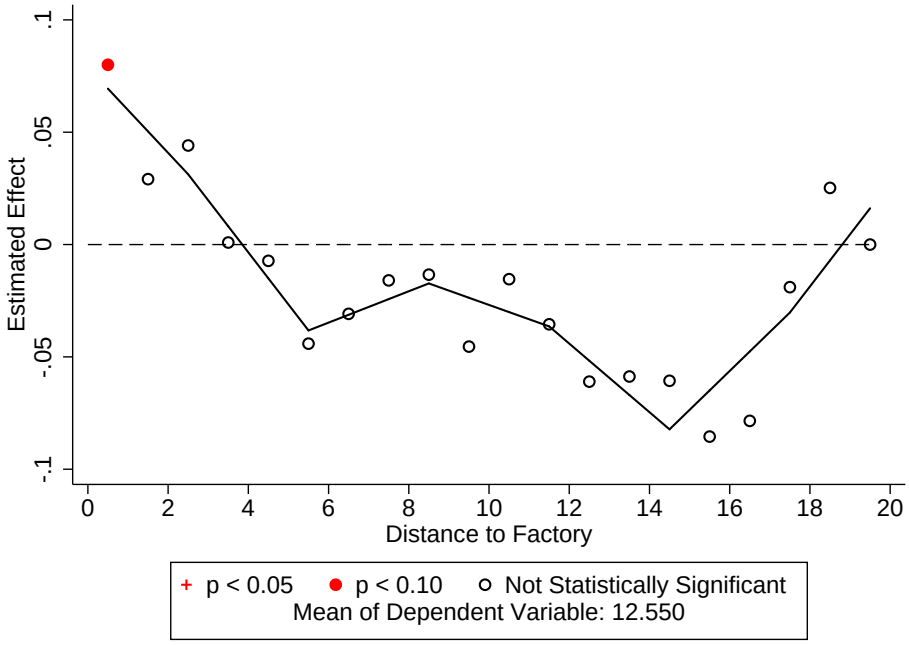



(f) *High School by Cohort*




**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 F-7: Expenditure (2001-11)

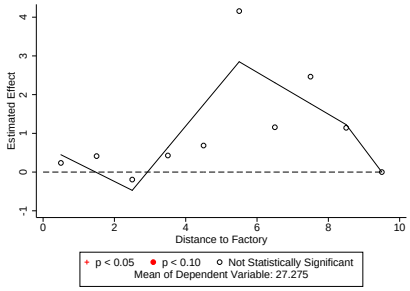



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

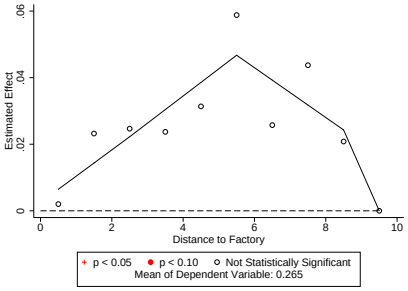
# G Removing Factories Within 10km of a Residency Capital

Figure G-1: Geographic Controls

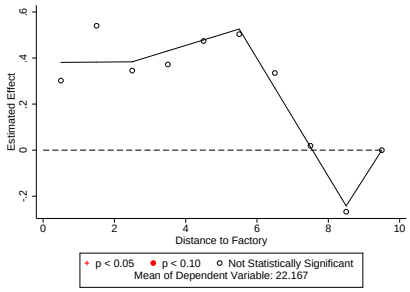
(a) *Elevation*



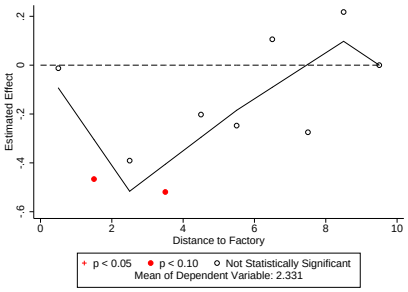

(b) *Slope*



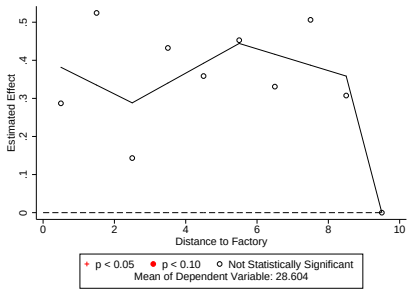

(c) *Distance to Coast*



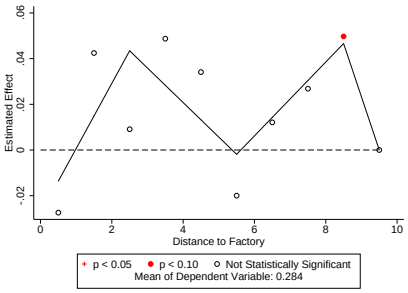

(d) *Log Flow Accumulation*



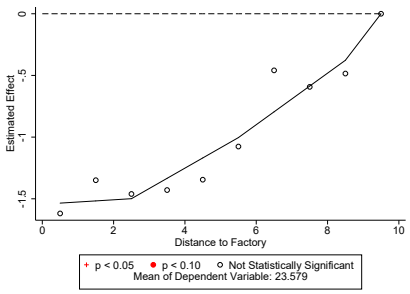

(e) *Distance to Natural Harbor*




(f) *Distance to River*



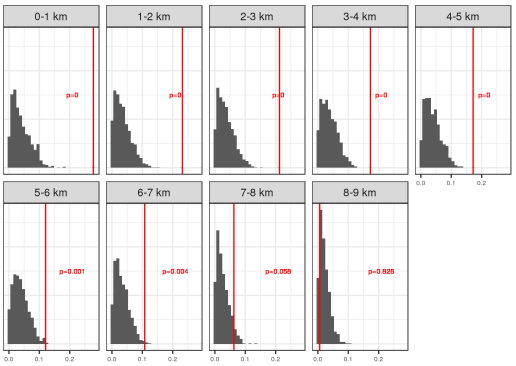

(g) *Distance to Nearest 1830 Residency Capital*



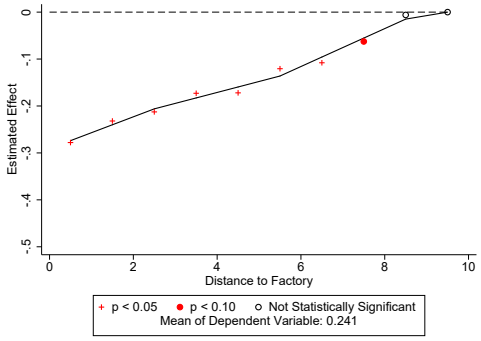

**Notes:** Points plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects. The data are fit with a linear spline. p-values compare the impact of proximity to actual factories to the impact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

Figure G-2: Share in Agriculture (Susenas 2001-11)

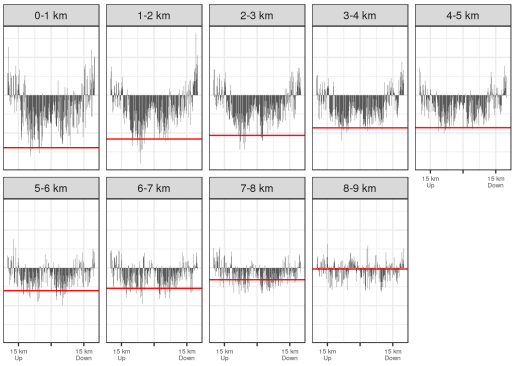
(a) *Independent Shifts*



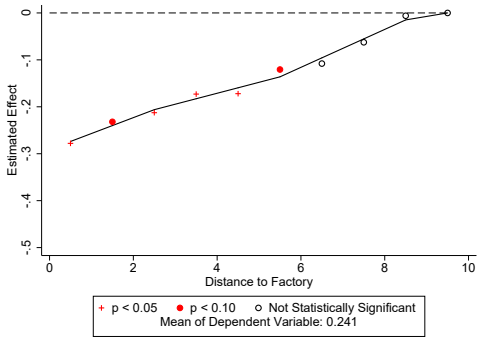

(b) *Coefficients, Independent Shifts*




(c) *Common Shifts*



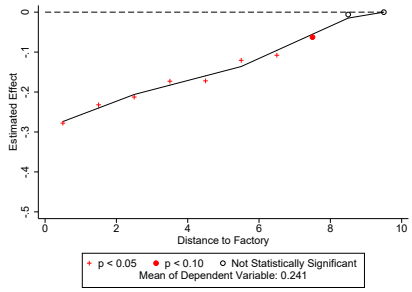

(d) *Coefficients, Common Shifts*



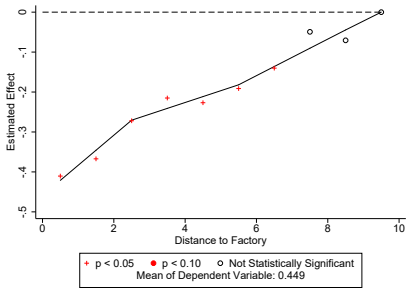

**Notes:** Panel (a) plots histograms of absolute coefficients from a regression of the outcome variable on bins in distance to counterfactual factories, controlling for nearest-factory fixed effects, geographic controls, a linear spline in distance to the nearest 1830 residency capital, and survey year 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: Industrial Shares

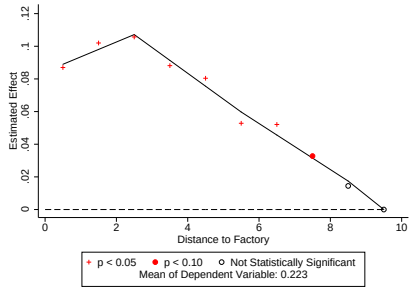
(a) *Agriculture (Susenas 2001-11)*



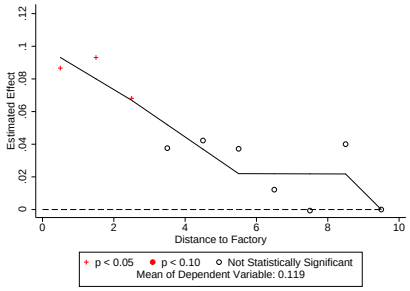

(b) *Agriculture (Census 1980)*



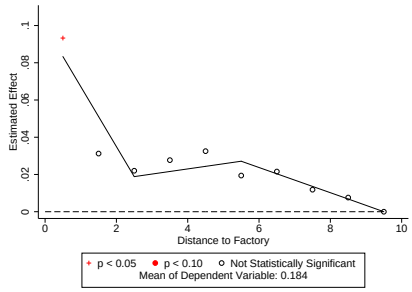

(c) *Manufacturing (Susenas 2001-11)*



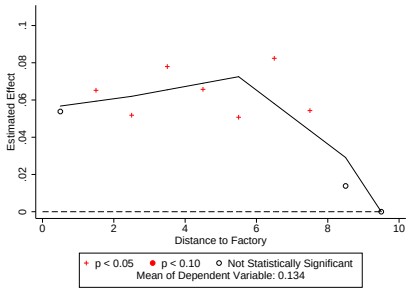

(d) *Manufacturing (Census 1980)*



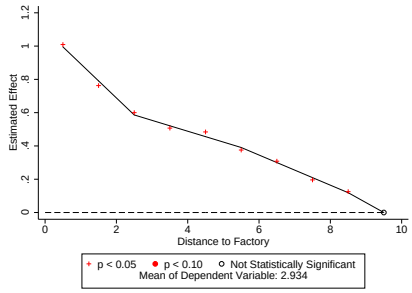

(e) *Retail (Susenas 2001-11)*



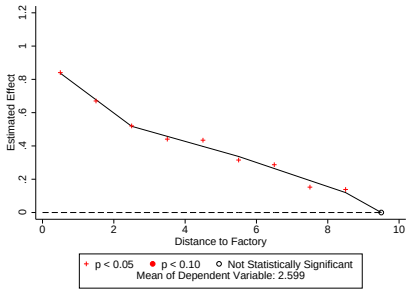

(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*



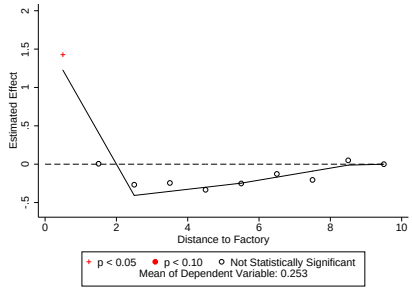

(h) *Log Population Density (PODES 2003)*



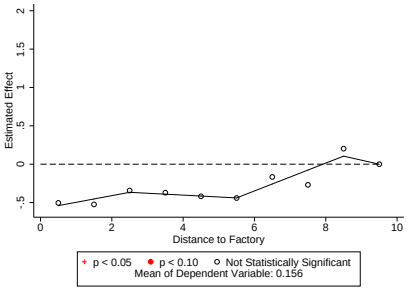

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 Residency capital. Panels (a), (c), and (e) also include 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-4: Sugar Production Today

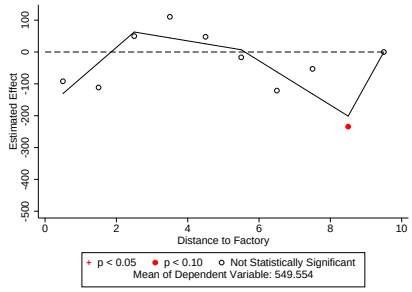
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



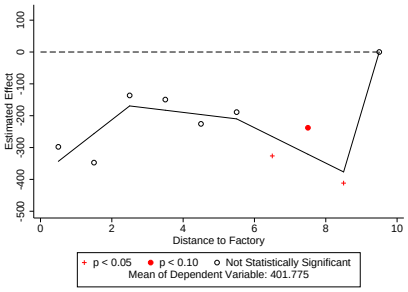

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



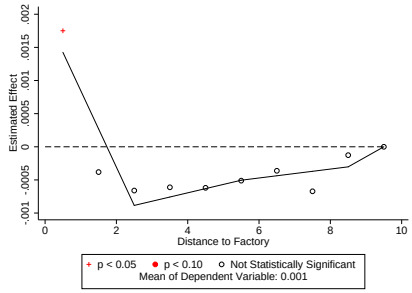

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



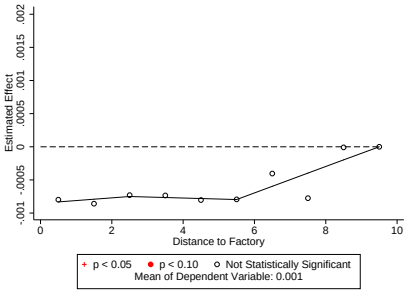

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



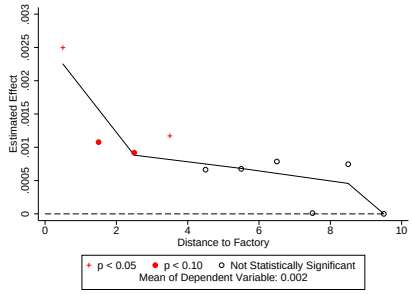

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*



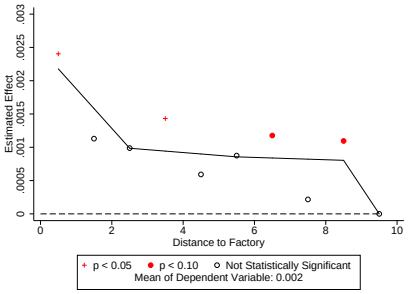

(f) *Emp Share Upstream (No Modern Factories, Economic Census 2006)*




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



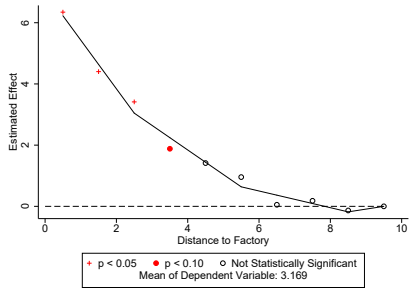

(h) *Emp Share Downstream (No Modern Factories, Economic Census 2006)*



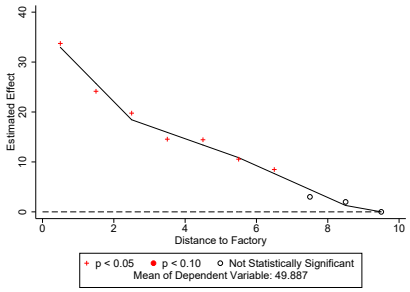

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. The left panels use all histrical sugar factories while the right panels exclude factories within 2km of a modern sugar 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 G-5: Public Goods

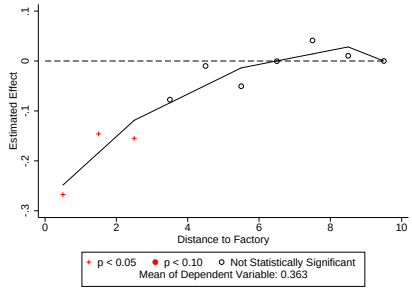
(a) *Intercity Road Density*



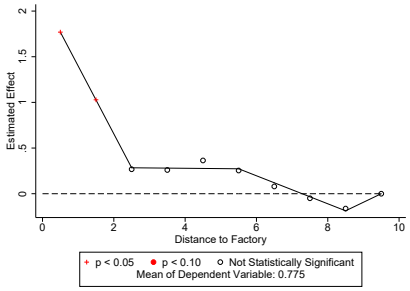

(b) *Local Road Density*



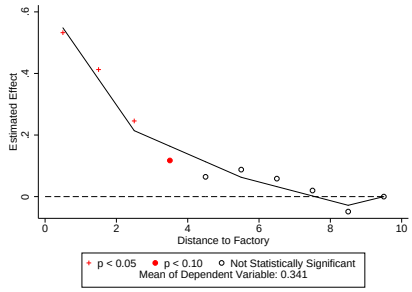

(c) *Dirt Road (PODES 1980)*



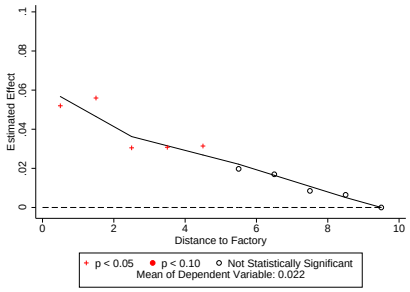

(d) *Railroad Density*



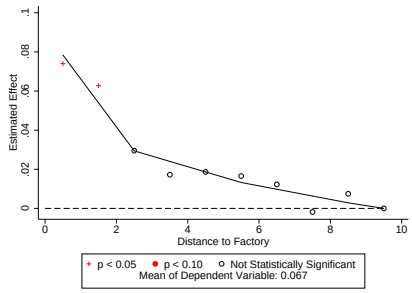

(e) *Village Has Electricity (PODES 1980)*



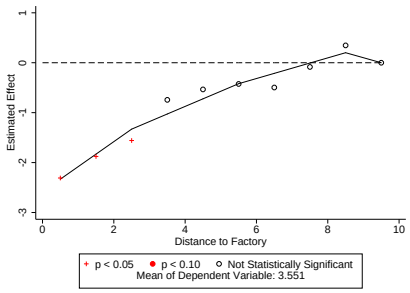

(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*



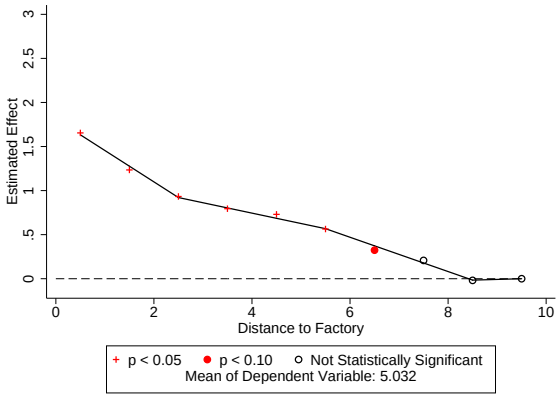

(h) *Subdistrict Capital (2011 PODES)*



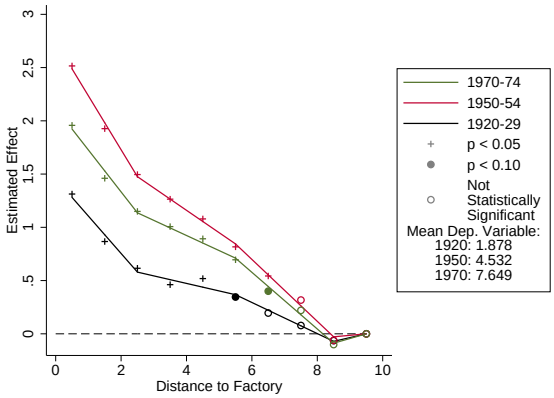

**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for nearest-factory fixed effects, geographic controls, and a linear spline in distance to 1830 Residency capital. Panel (g) also includes 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-6: Education

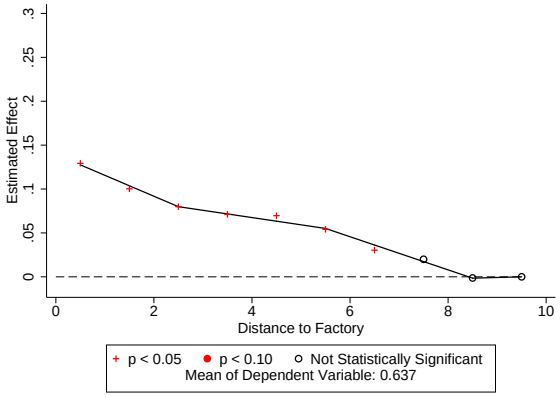
(a) *Years Education*



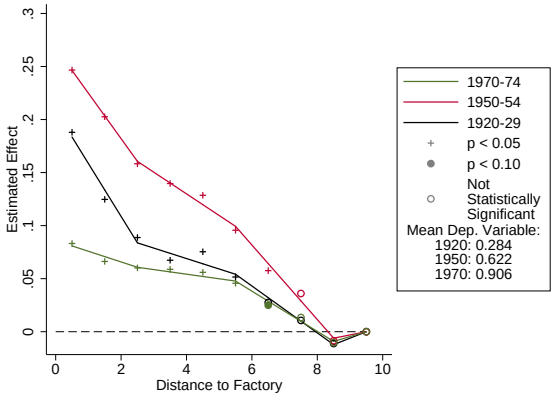

(b) *Years Education by Cohort*



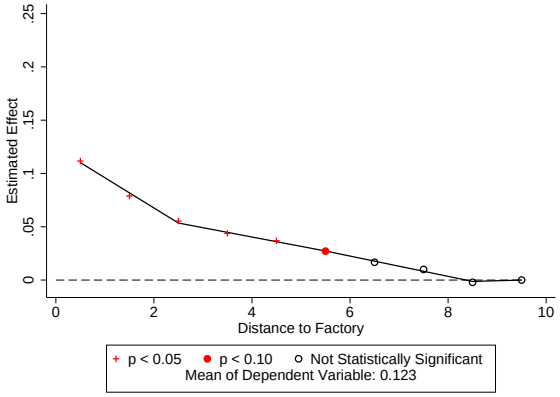

(c) *Primary*



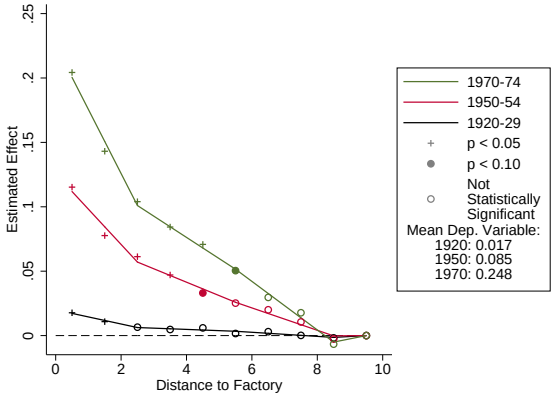

(d) *Primary by Cohort*




(e) *High School*

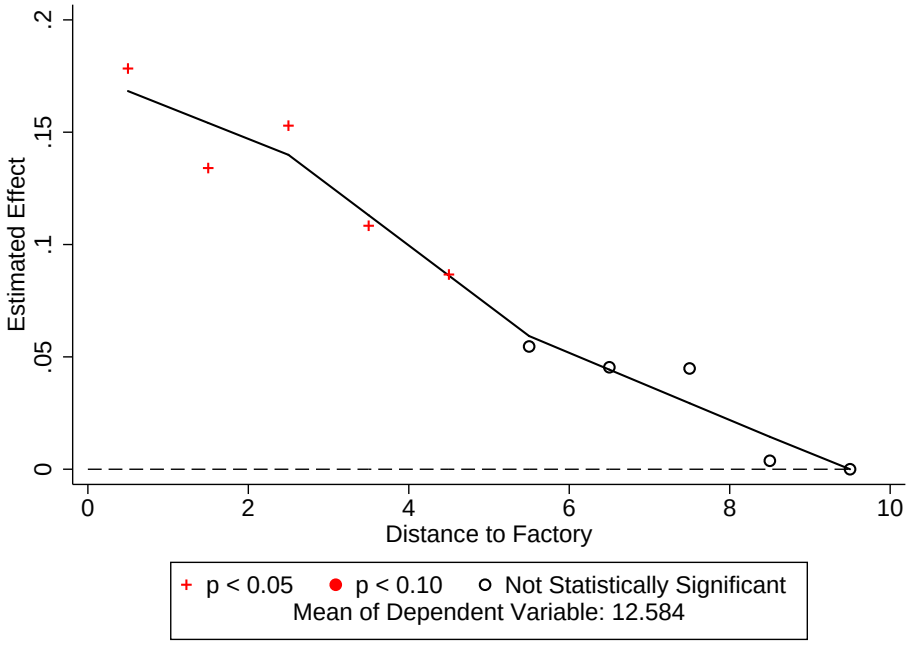



(f) *High School by Cohort*




**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-7: Expenditure (2001-11)

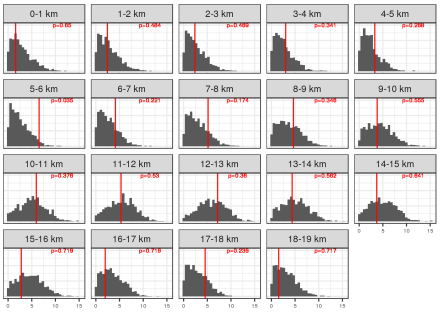



**Notes:** This figures plots coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for demographic variables, survey year fixed effects, nearest-factory fixed effects, geographic controls, and a linear spline in distance to 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.

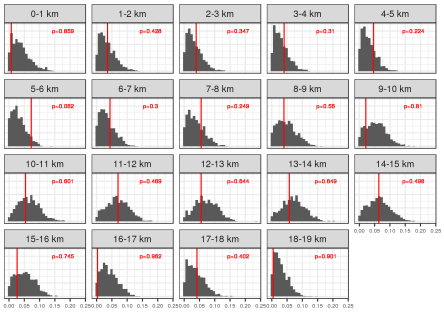
# H Independent Shifts Distributions

Figure H-1: Geographic Controls

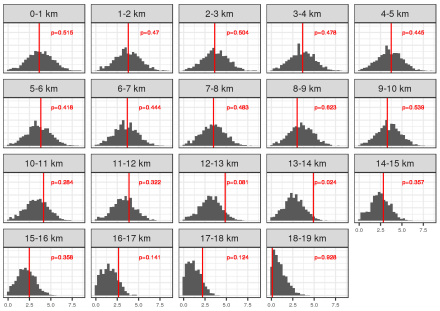
(a) *Elevation*



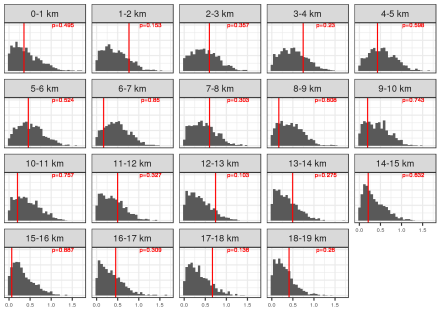

(b) *Slope*



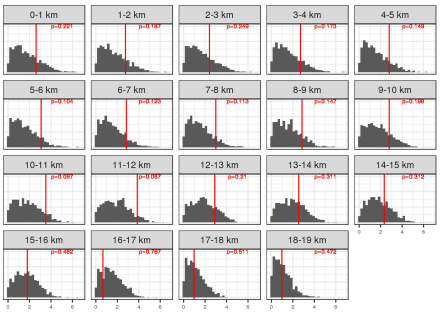

(c) *Distance to Coast*



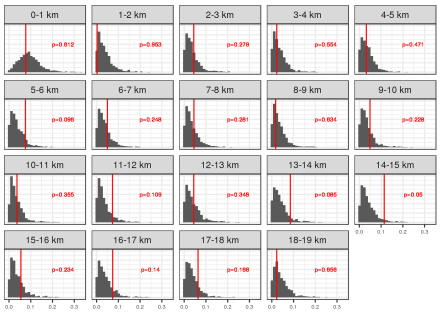

(d) *Log Flow Accumulation*




(e) *Distance to Natural Harbor*




(f) *Distance to River*




(g) *Distance to Nearest 1830 Residency Capital*

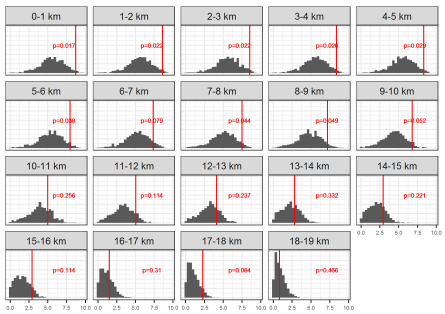
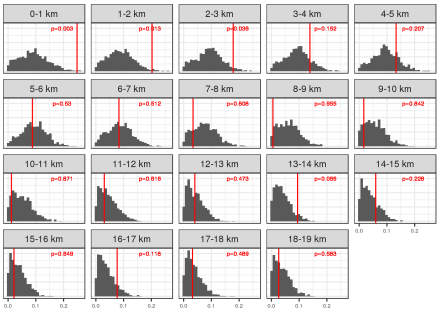
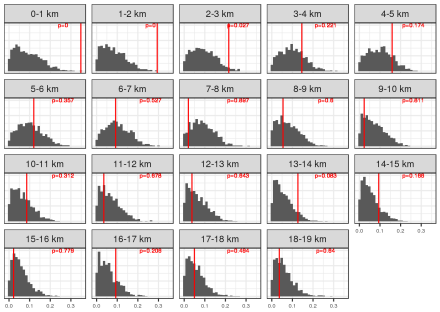


Figure H-2: Industrial Shares

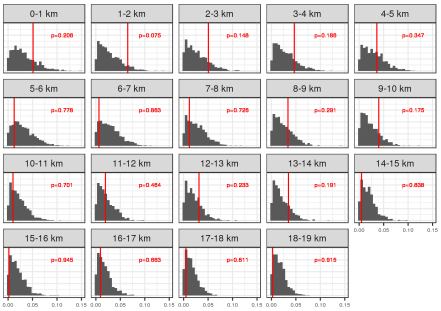
(a) *Agriculture (Susenas 2001-11)*



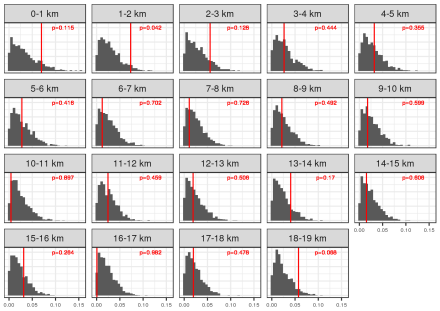

(b) *Agriculture (Census 1980)*



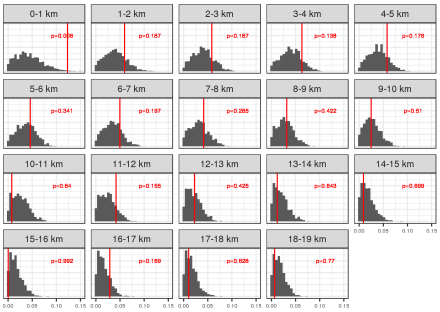

(c) *Manufacturing (Susenas 2001-11)*



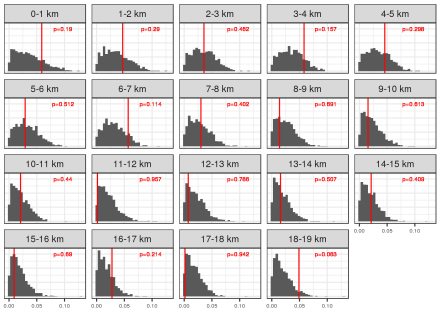

(d) *Manufacturing (Census 1980)*



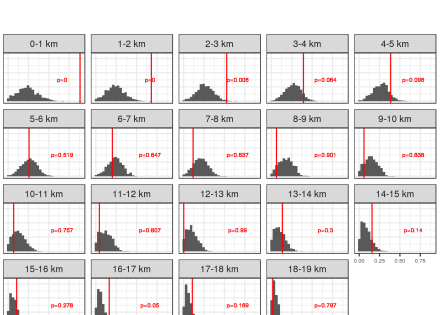

(e) *Retail (Susenas 2001-11)*




(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*




(h) *Log Population Density (PODES 2003)*

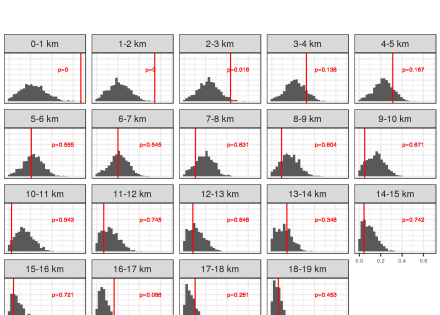
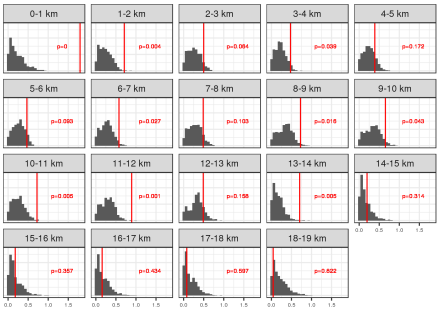
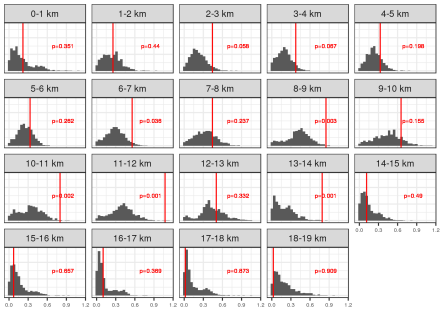


Figure H-3: Sugar Production Today

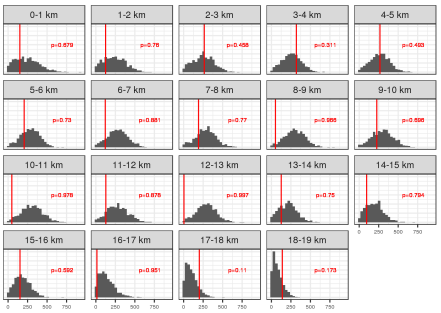
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



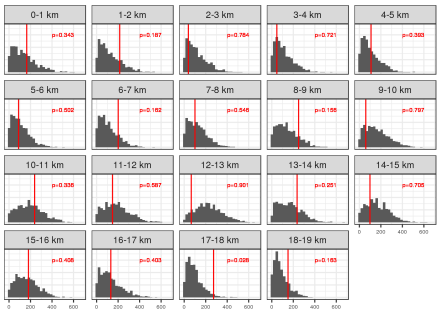

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



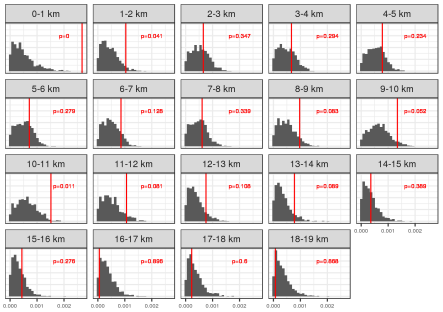

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



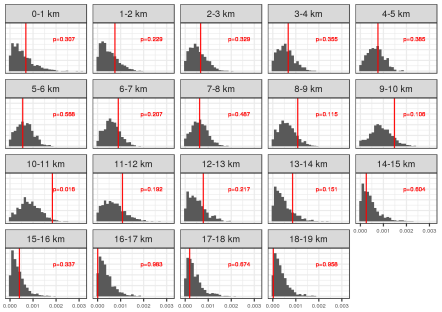

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



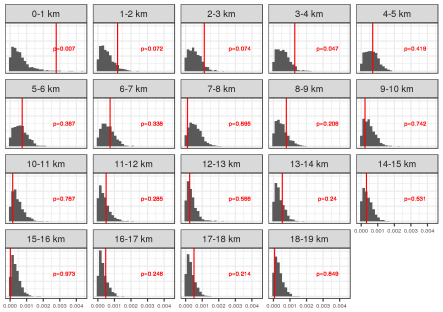

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*




(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)*

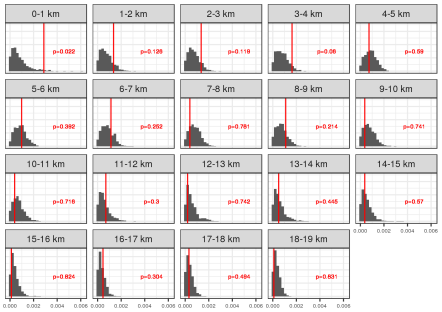
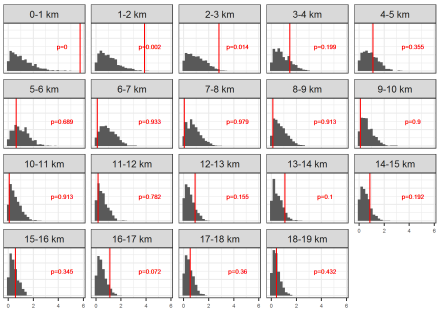
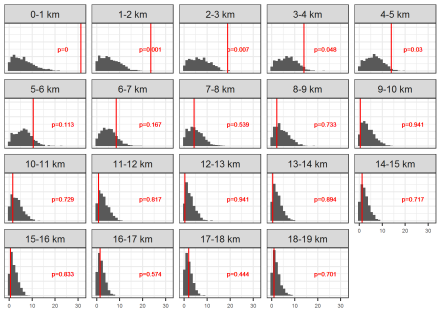


Figure H-4: Public Goods

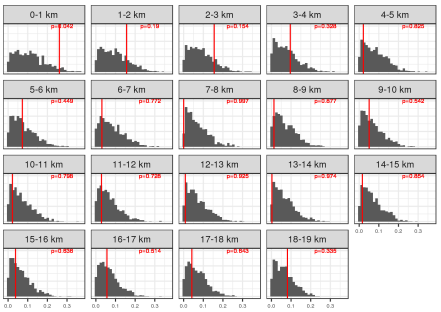
(a) *Intercity Road Density*



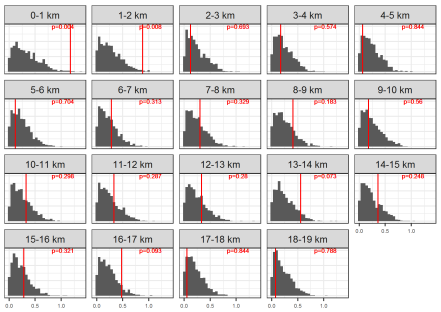

(b) *Local Road Density*



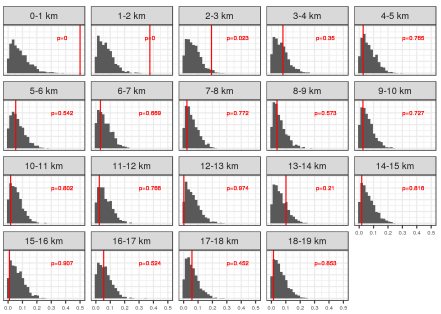

(c) *Dirt Road (PODES 1980)*



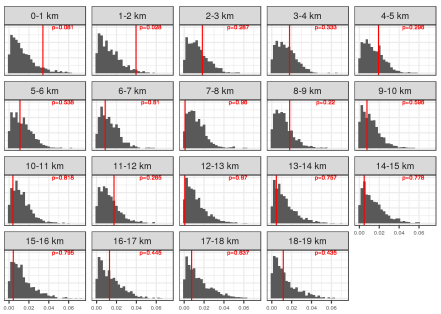

(d) *Railroad Density*



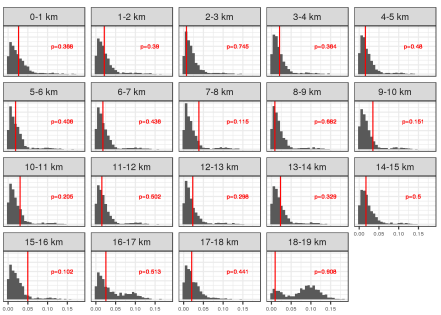

(e) *Village Has Electricity (PODES 1980)*




(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*




(h) *Subdistrict Capital (2011 PODES)*

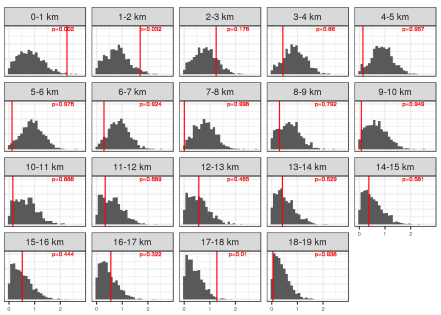
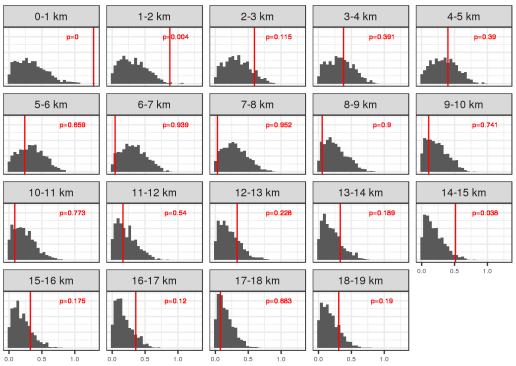
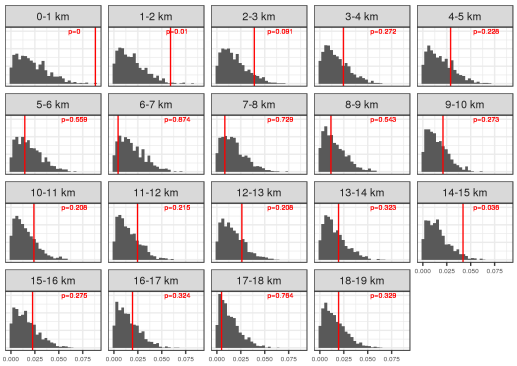


Figure H-5: Education

(a) *Years Education*




(b) *Primary*




(c) *High School*

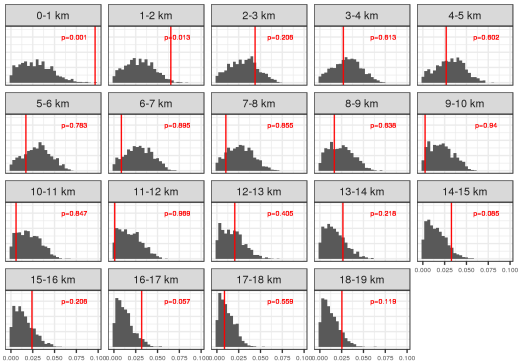
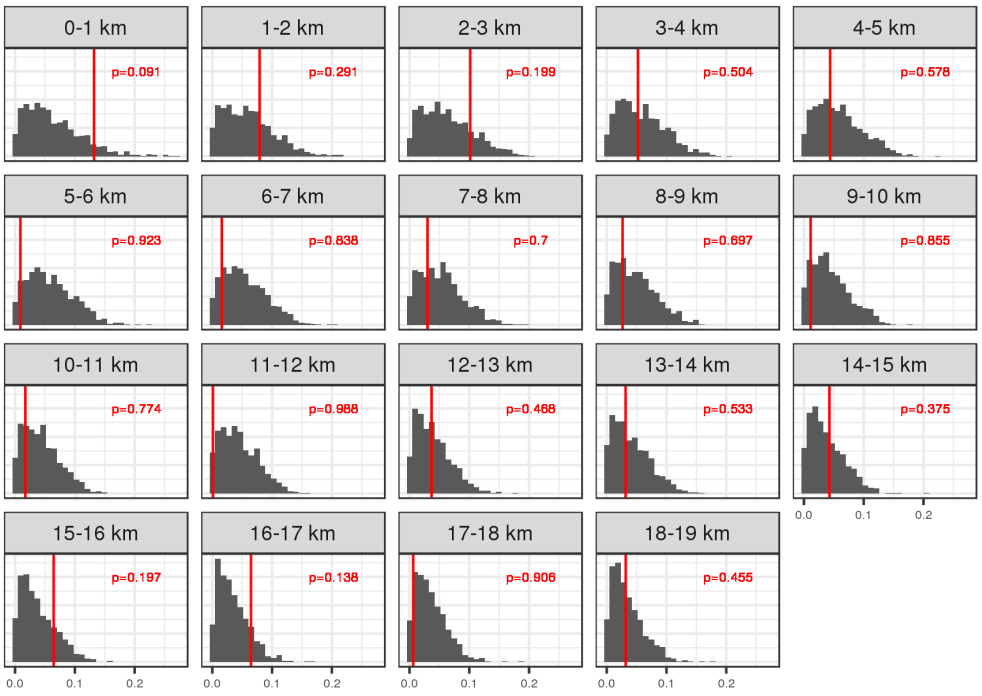


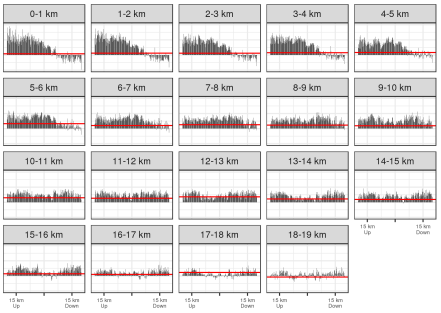
Figure H-6: Expenditure (2001-11)



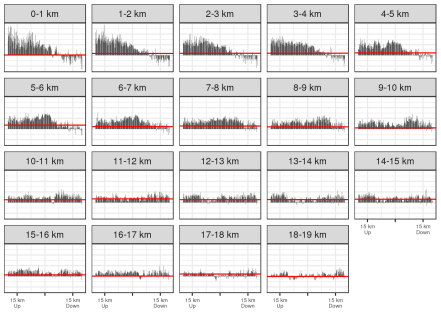
# I Common Shifts Distributions

Figure I-1: Geographic Controls

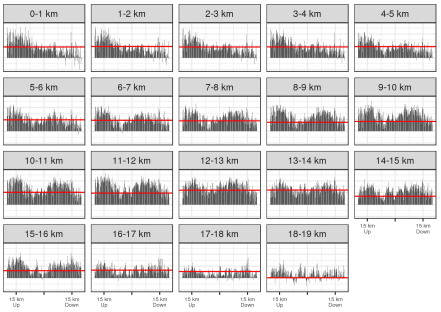
(a) *Elevation*



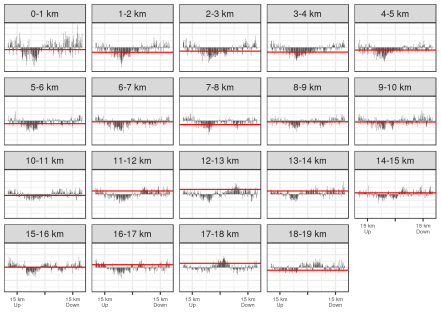

(b) *Slope*



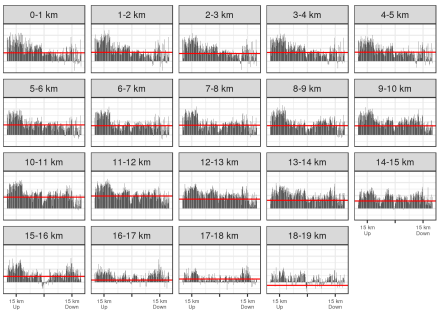

(c) *Distance to Coast*



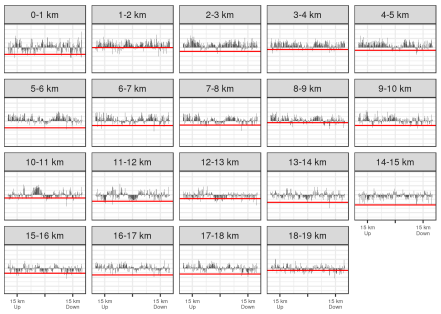

(d) *Log Flow Accumulation*




(e) *Distance to Natural Harbor*




(f) *Distance to River*




(g) *Distance to Nearest 1830 Residency Capital*

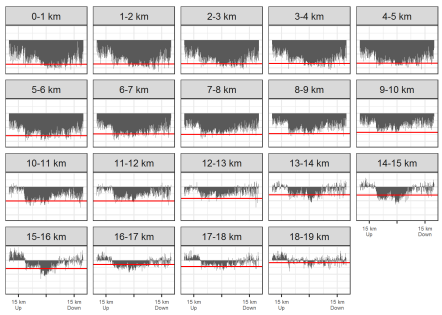
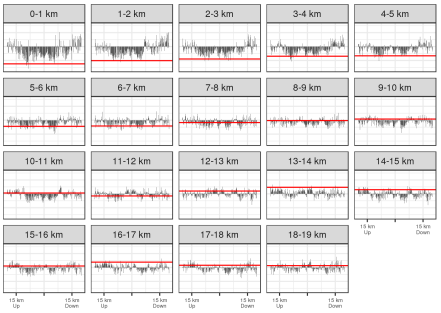
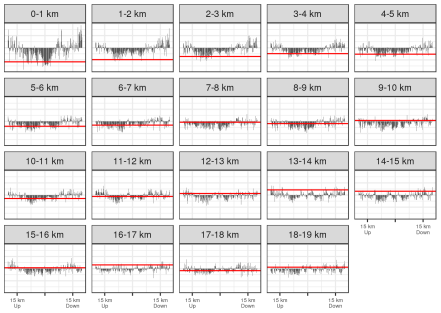


Figure I-2: Industrial Shares

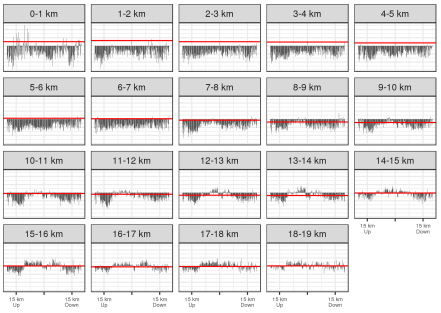
(a) *Agriculture (Susenas 2001-11)*



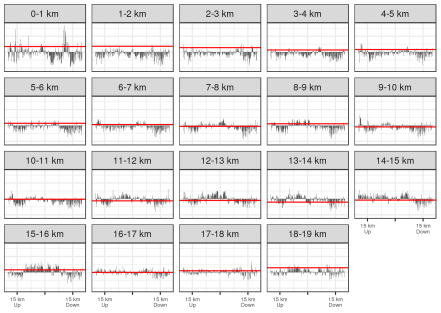

(b) *Agriculture (Census 1980)*



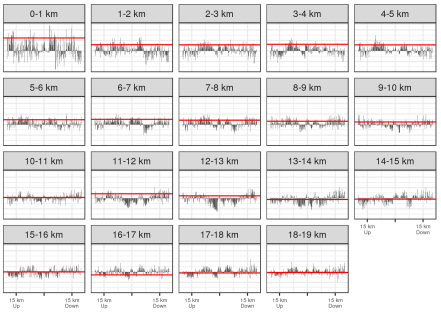

(c) *Manufacturing (Susenas 2001-11)*



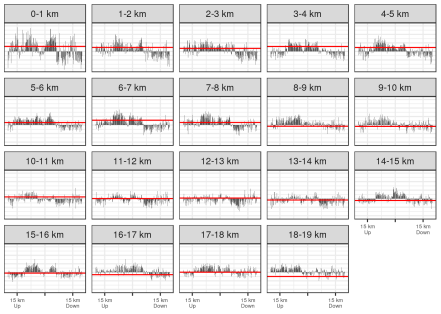

(d) *Manufacturing (Census 1980)*



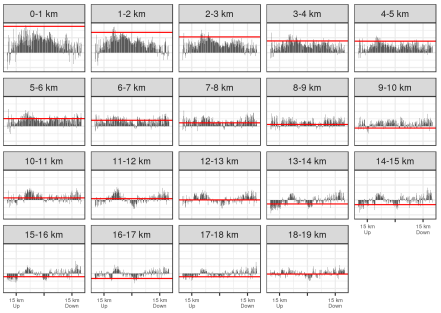

(e) *Retail (Susenas 2001-11)*




(f) *Retail (Census 1980)*




(g) *Log Population Density (PODES 2003)*




(h) *Log Population Density (PODES 2003)*

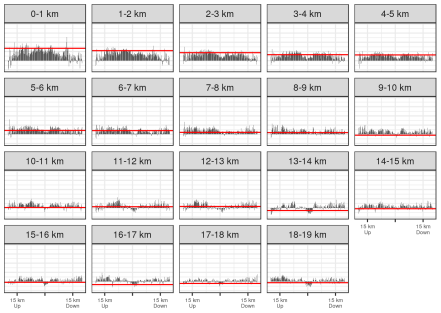
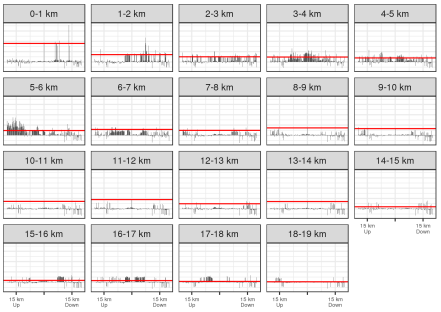
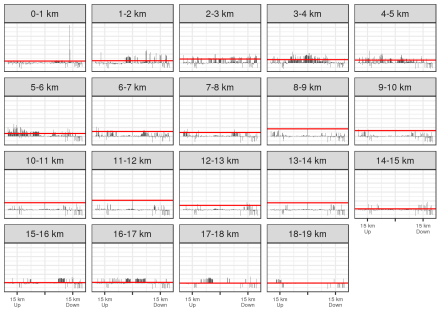


Figure I-3: Sugar Production Today

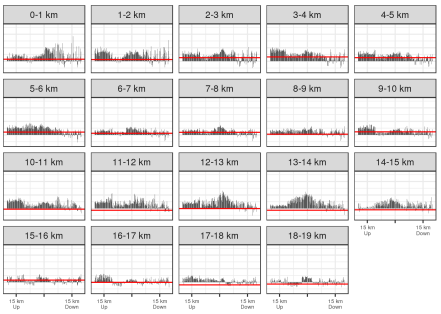
(a) *Log Value Sugar Processed (Full Sample, Ec Cen 2006)*



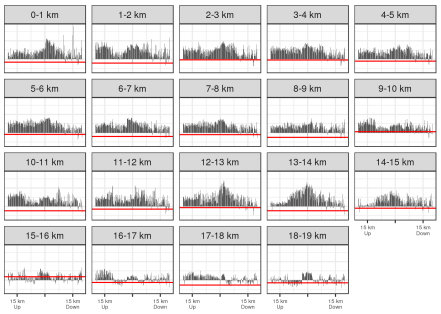

(b) *Log Value Sugar Processed (No Modern Factories, Ec Cen 2006)*



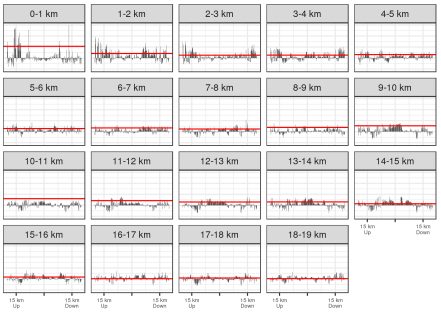

(c) *Tons of Cane Grown (Full Sample, PODES 2003)*



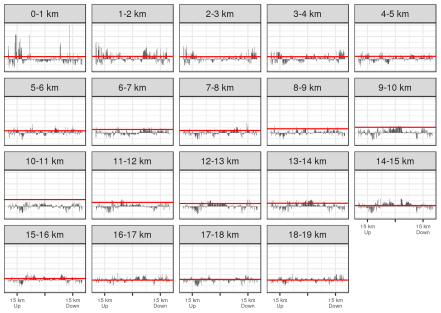

(d) *Tons of Cane Grown (No Modern Factories, PODES 2003)*



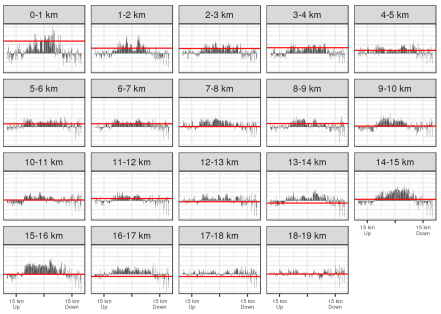

(e) *Employment Share Upstream (Full Sample, Economic Census 2006)*




(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)*

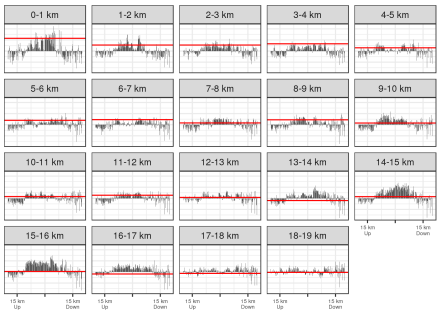
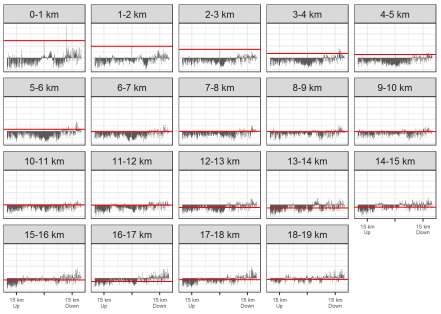
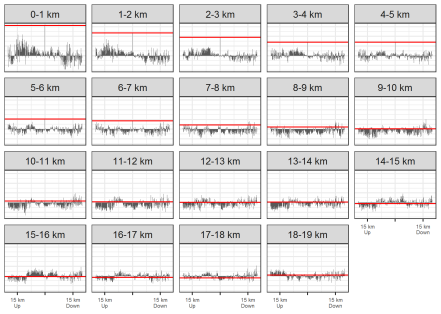


Figure I-4: Public Goods

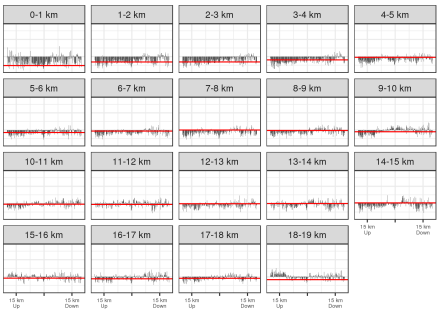
(a) *Intercity Road Density*



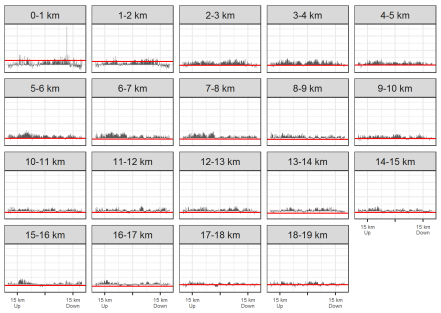

(b) *Local Road Density*



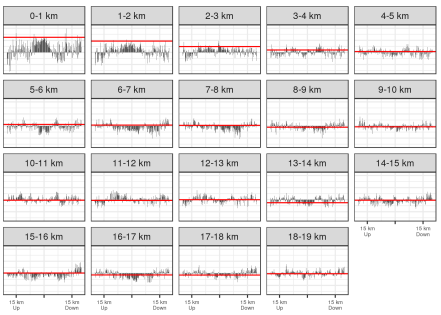

(c) *Dirt Road (PODES 1980)*



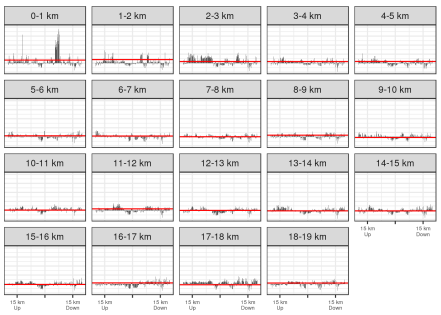

(d) *Railroad Density*



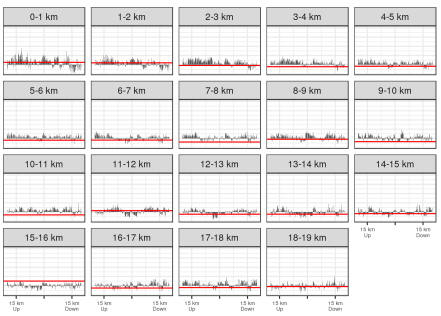

(e) *Village Has Electricity (PODES 1980)*




(f) *High Schools (PODES 1980)*




(g) *High Schools (PODES 1998-2011)*




(h) *Subdistrict Capital (2011 PODES)*

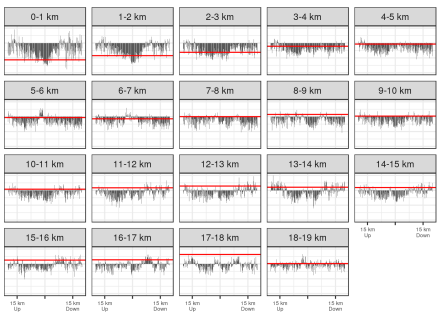
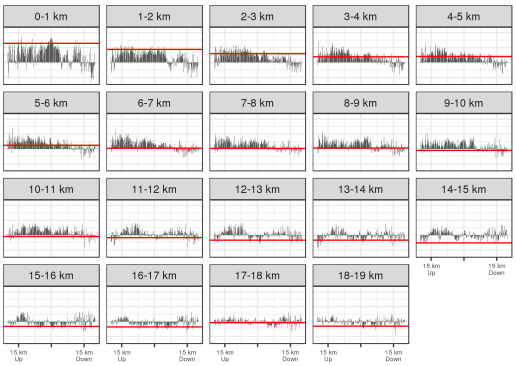
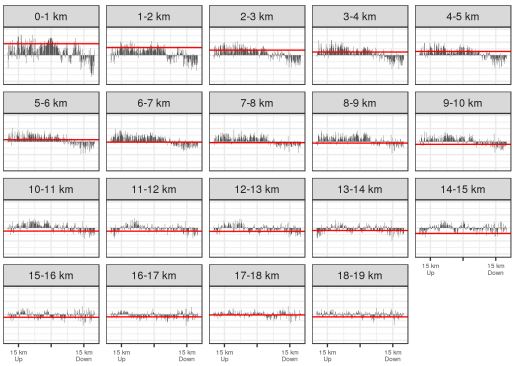


Figure I-5: Education

(a) *Years Education*




(b) *Primary*




(c) *High School*

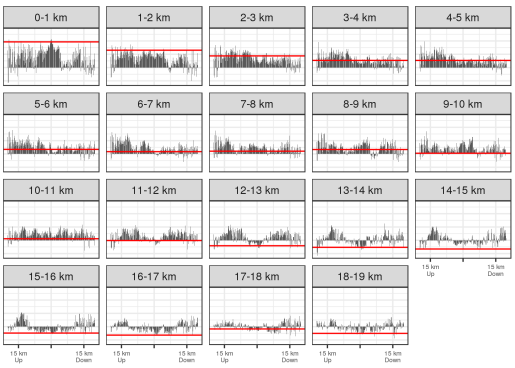
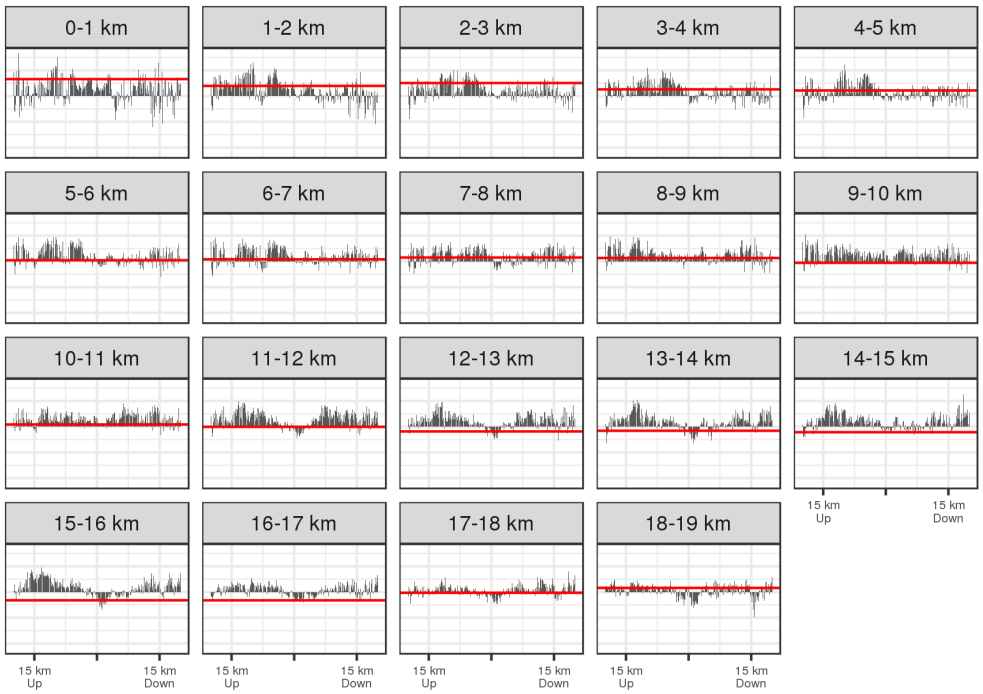


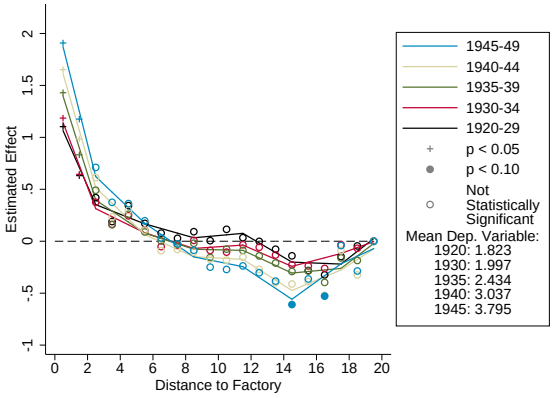
Figure I-6: Expenditure (2001-11)



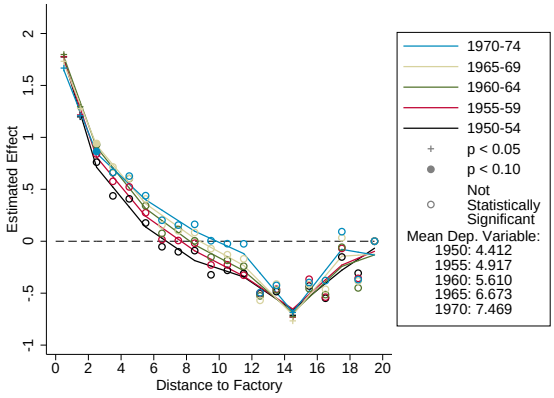
# J Education for All Cohorts

Figure J-1: Education

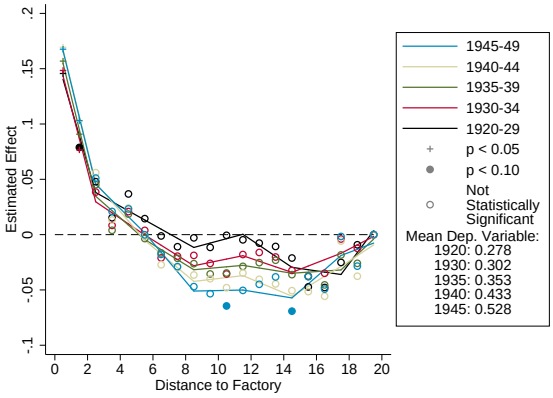
(a) *Years Education: 1920-1949*



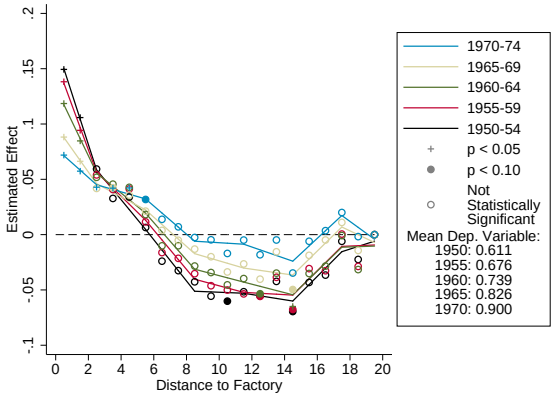

(b) *Years Education: 1950-1974*



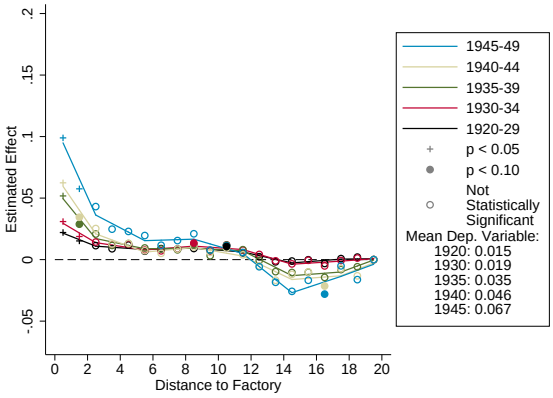

(c) *Primary: 1920-1949*



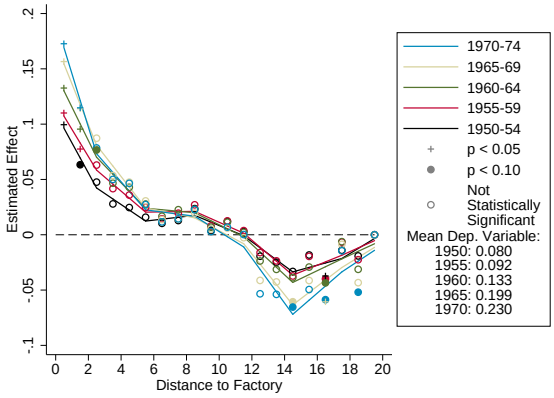

(d) *Primary: 1950-1974*




(e) *High School: 1920-1949*




(f) *High School: 1950-1974*

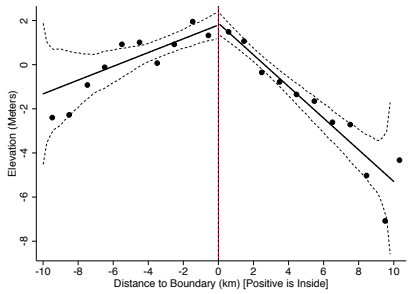



**Notes:** These figures plot coefficients estimated from regressing the outcome variable on 1-km bins of distance to the nearest historical factory, controlling for gender, nearest-factory fixed effects, geographic controls, and a linear spline in distance to the nearest 1830 residency capital. Each 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.

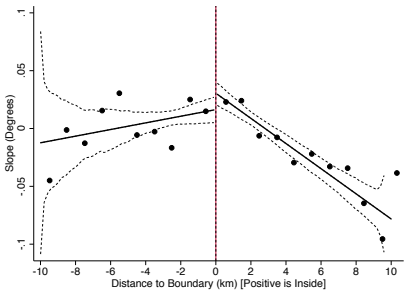
# K RD Plots

Figure K-1: Geographic Characteristics

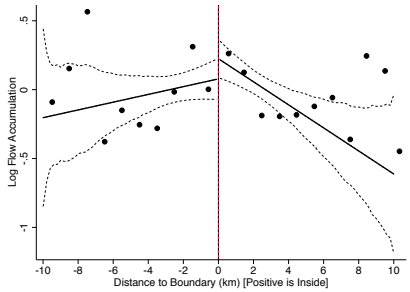
(a) *Elevation*



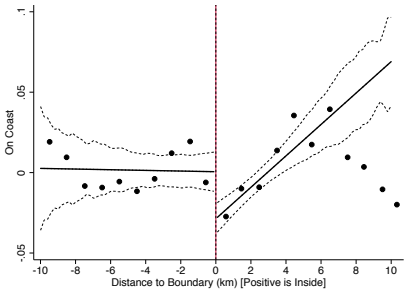

(b) *Slope*



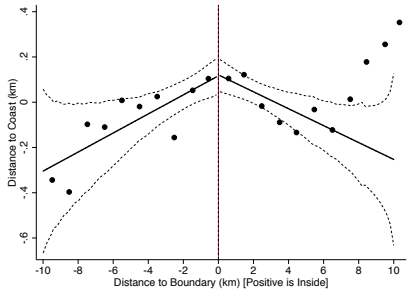

(c) *Log Flow Accumulation*



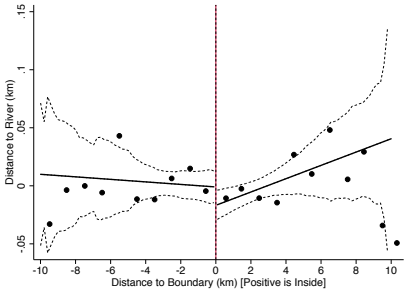

(d) *On Coast*



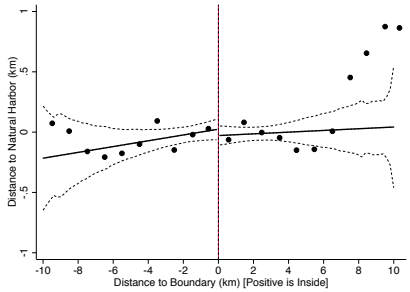

(e) *Distance to Coast*



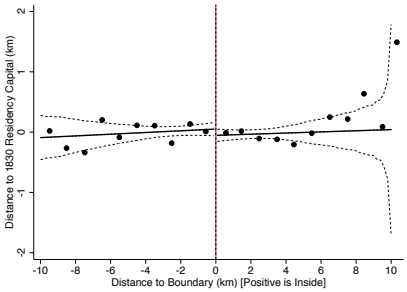

(f) *Distance to River*




(g) *Distance to Natural Harbor*



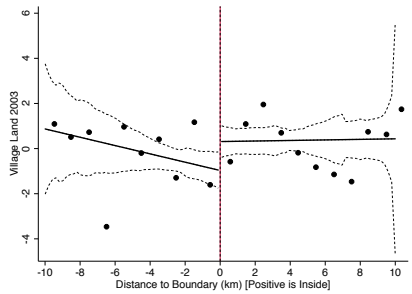

(h) *Distance to 1830 Residency Capital*



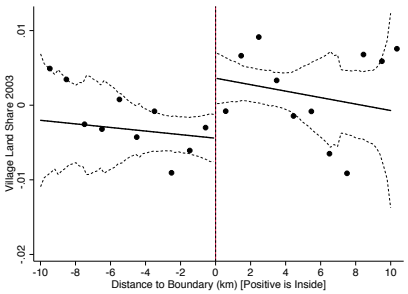

**Notes:** Points are binned residuals from a regression of the outcome variable on boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and a linear spline in distance to the nearest historical factory. Solid lines fit a local linear regression,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

Figure K-2: Land Tenure

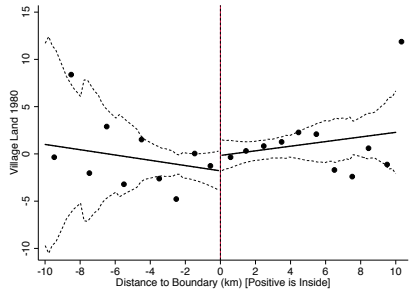
(a) *Village Head 2003: Total Land*



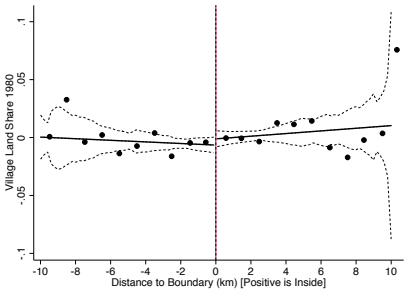

(b) *Village Head 2003: Land Share*



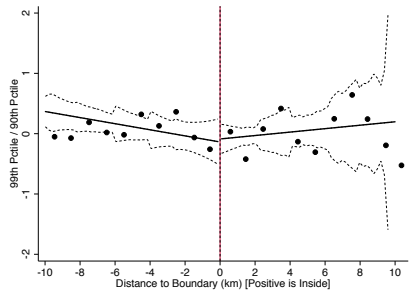

(c) *Village Head 1980: Total Land*



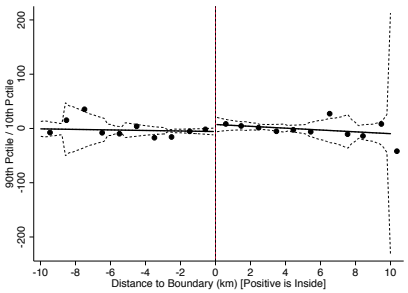

(d) *Village Head 1980: Land Share*



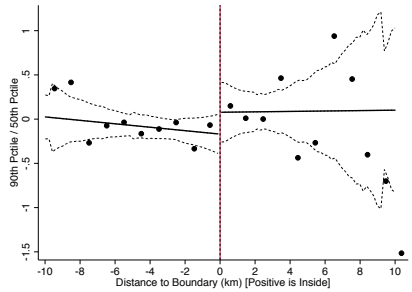

(e) *99th Pctile / 90th Pctile*



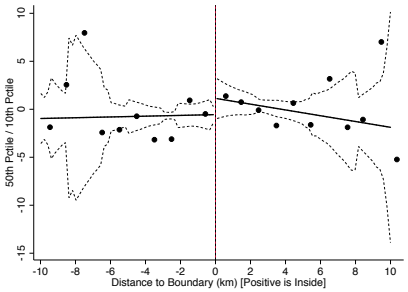

(f) *90th Pctile / 10th Pctile*




(g) *90th Pctile / 50th Pctile*



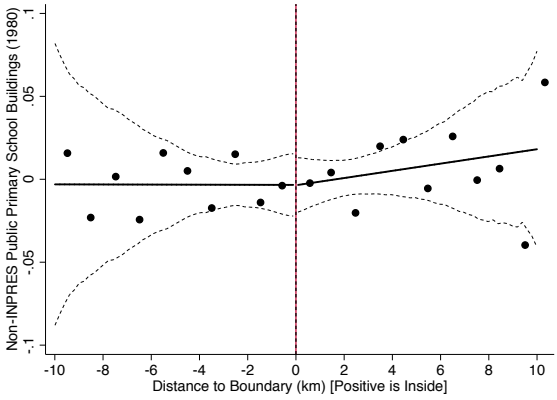

(h) *50th Pctile / 10th Pctile*



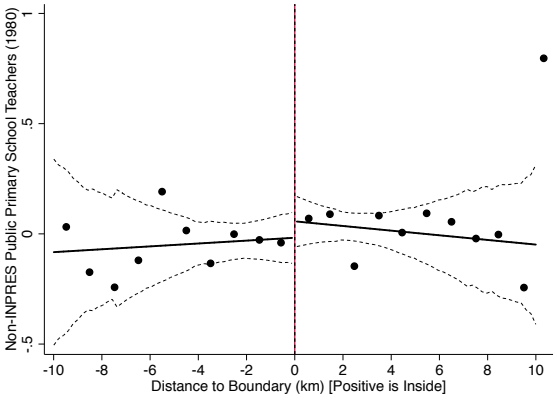

**Notes:** Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest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

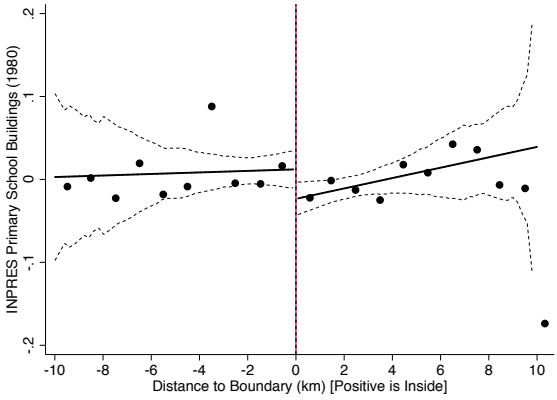
(a) *Public Non-INPRES Primary Buildings*



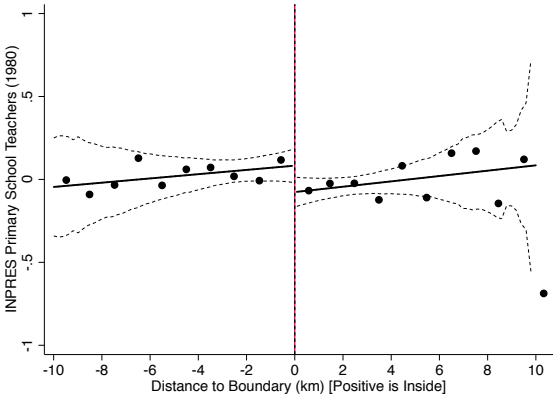

(b) *Public Non-INPRES Primary Teachers*



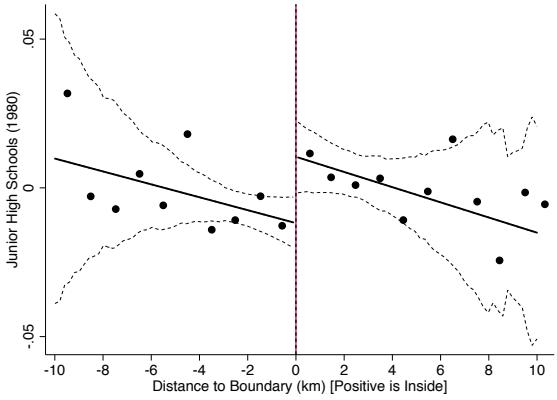

(c) *INPRES Primary Buildings*



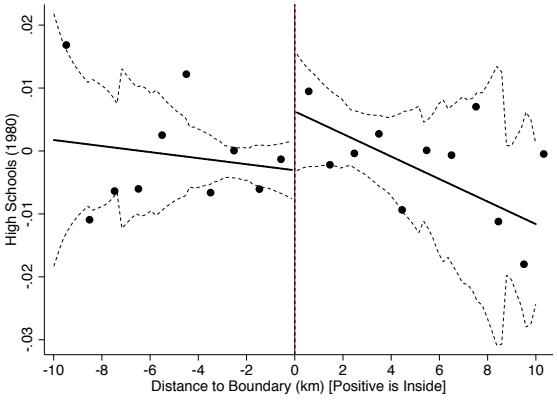

(d) *INPRES Primary Teachers*




(e) *Junior High Schools*



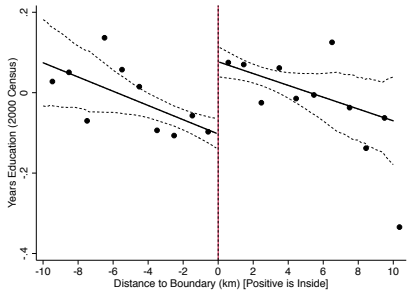

(f) *High Schools*



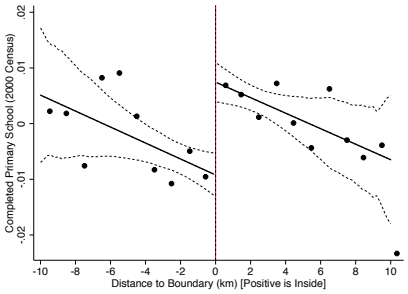

**Notes:** Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest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

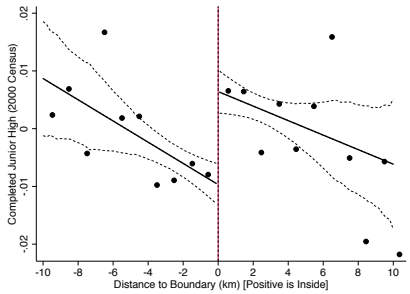
(a) *Years Education 2000*



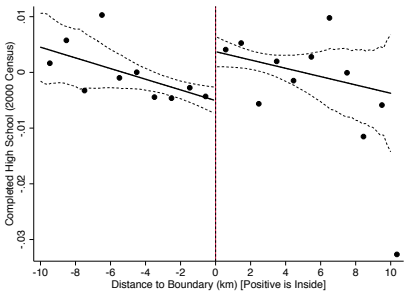

(b) *Primary School 2000*



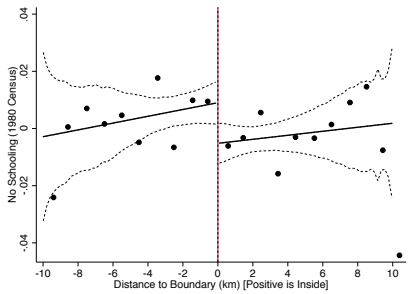

(c) *Junior High 2000*



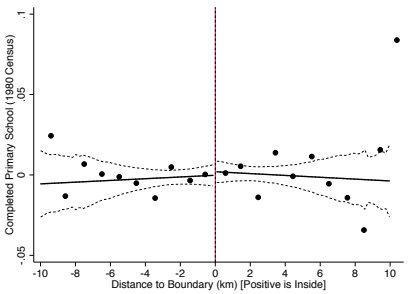

(d) *High School 2000*



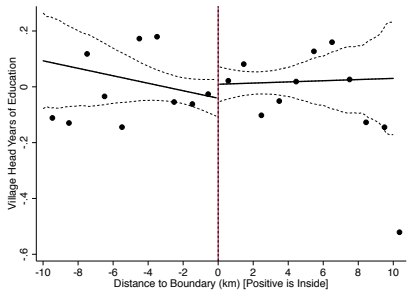

(e) *No School 1980*



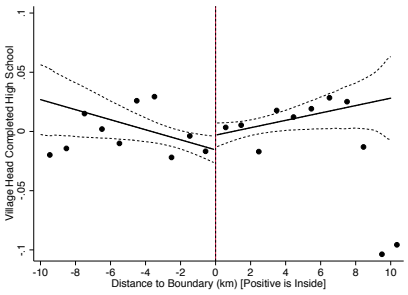

(f) *Primary School 1980*




(g) *Village Head Years Education*



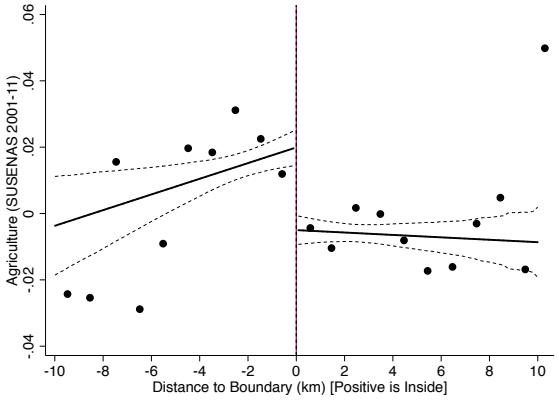

(h) *Village Head High School*



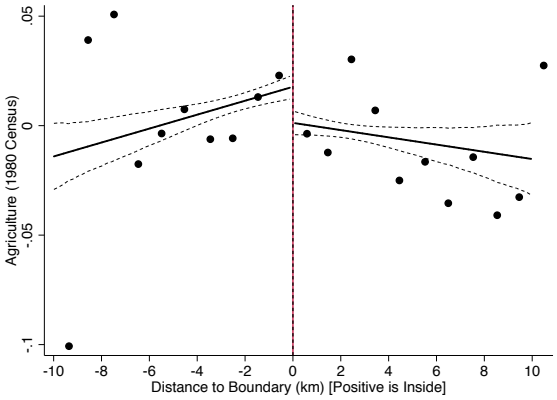

**Notes:** Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest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

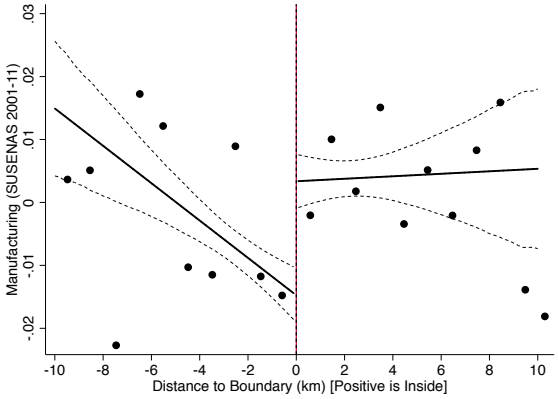
(a) *Agriculture (SUSENAS)*



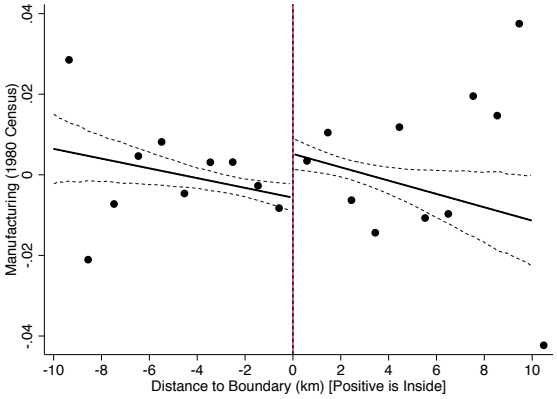

(b) *Agriculture (1980 Census)*



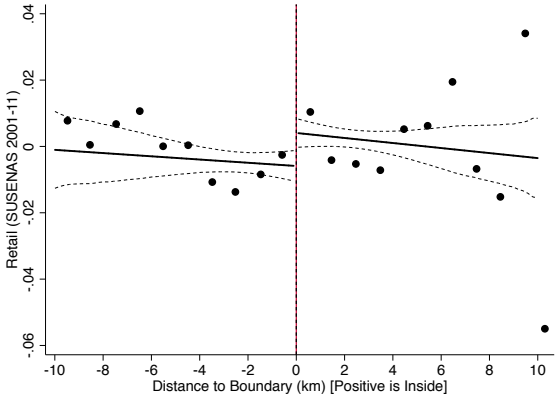

(c) *Manufacturing (SUSENAS)*



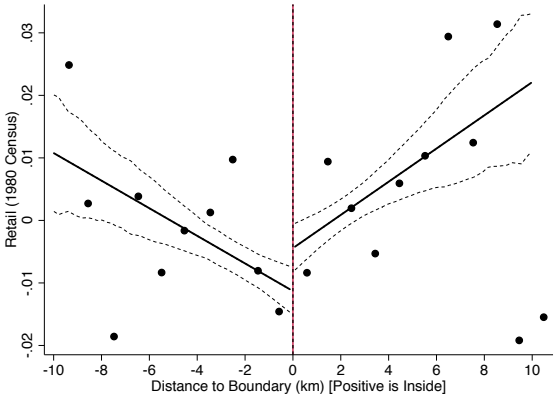

(d) *Manufacturing (1980 Census)*




(e) *Retail (SUSENAS)*



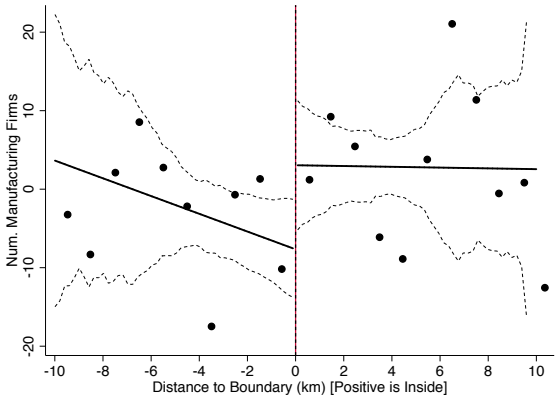

(f) *Commerce (1980 Census)*



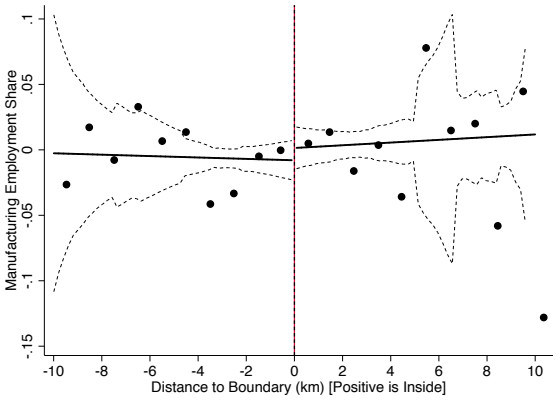

**Notes:** Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest 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 K-6: Firms, Population, and Consumption

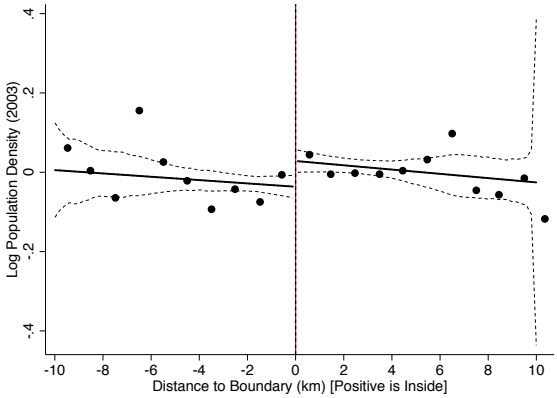
(a) *Number of Firms*



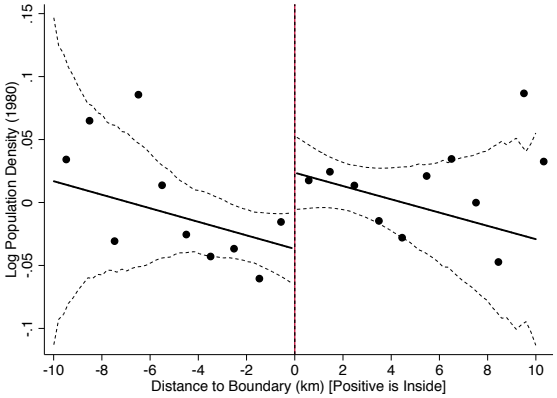

(b) *Employment Share*



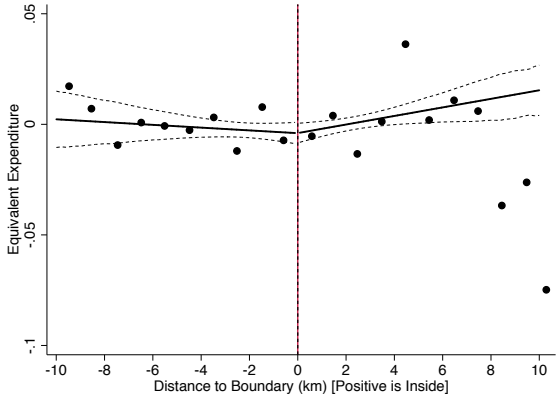

(c) *Log Population Density (2003)*




(d) *Log Population Density (1980)*




(e) *Log Equivalent Consumption*

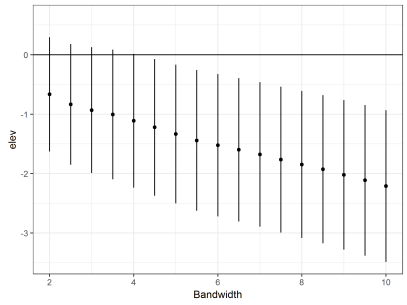



**Notes:** Points are binned residuals from a regression of the outcome variable on geographic controls, boundary segment fixed effects, a linear polynomial in latitude and longitude estimated separately by catchment area, and linear splines in distance to the nearest historical factory and to the nearest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

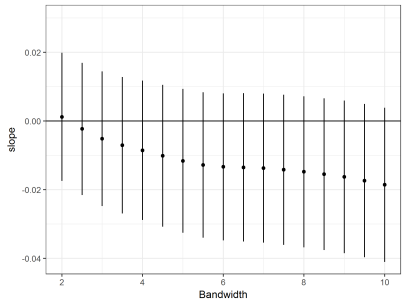
# L Robustness to Bandwidth Plots

Figure L-1: Geographic Characteristics

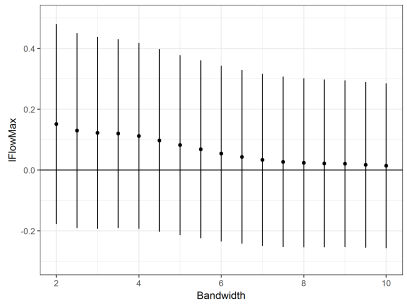
(a) *Elevation*



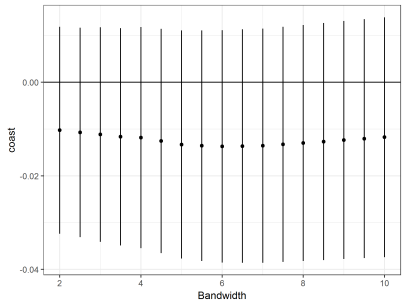

(b) *Slope*



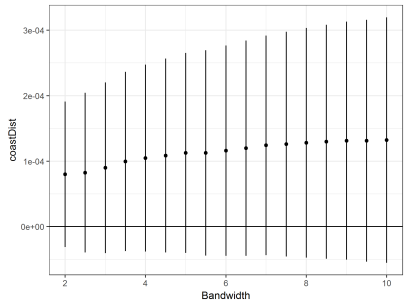

(c) *Log Flow Accumulation*



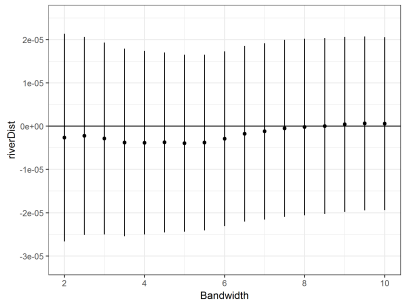

(d) *On Coast*



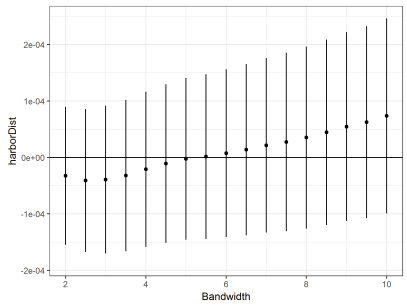

(e) *Distance to Coast*



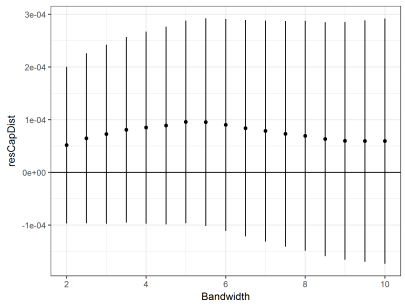

(f) *Distance to River*




(g) *Distance to Natural Harbor*



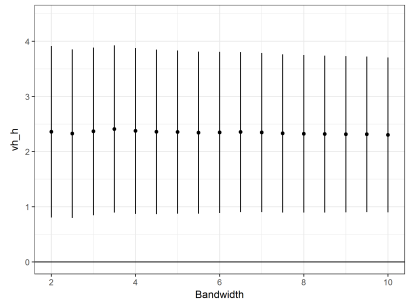

(h) *Distance to 1830 Residency Capital*



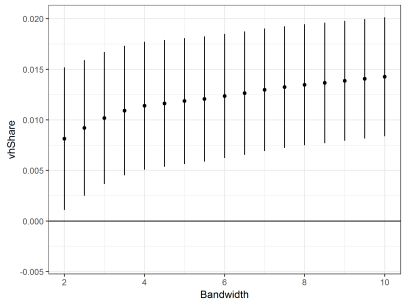

**Notes:** The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes boundary segment fixed effects, a spline in distance to the nearest historical factory with kinks each 3km, and a linear polynomial in latitude and longitude estimated separately for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

Figure L-2: Land Tenure

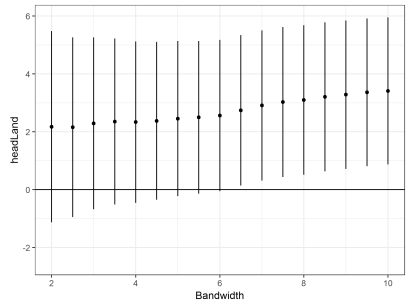
(a) *Village Head 2003: Total Land*



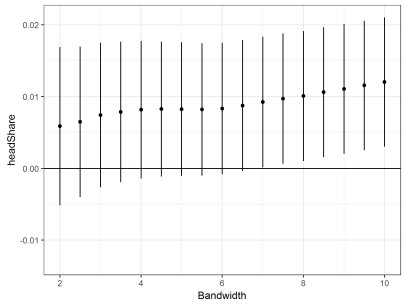

(b) *Village Head 2003: Land Share*



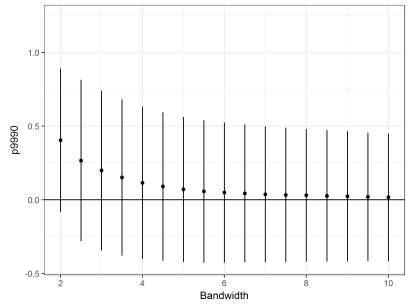

(c) *Village Head 1980: Total Land*



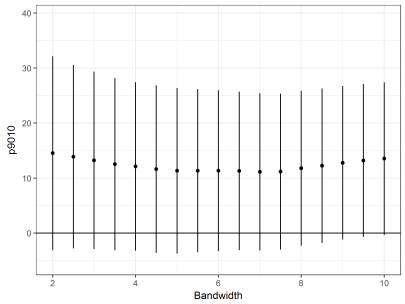

(d) *Village Head 1980: Land Share*



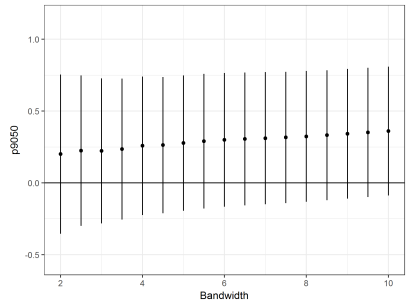

(e) *99th Pctile / 90th Pctile*



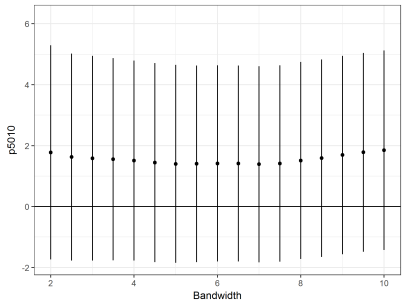

(f) *90th Pctile / 10th Pctile*




(g) *90th Pctile / 50th Pctile*



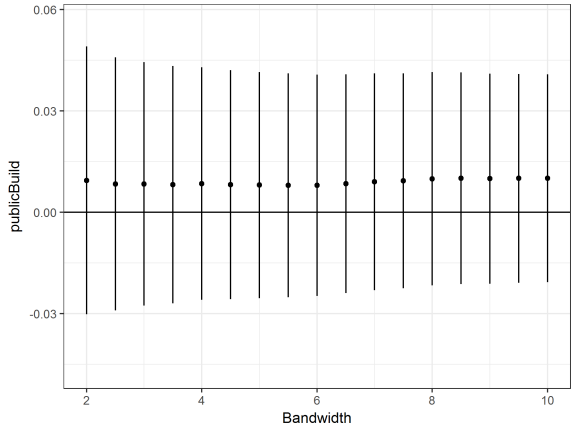

(h) *50th Pctile / 10th Pctile*



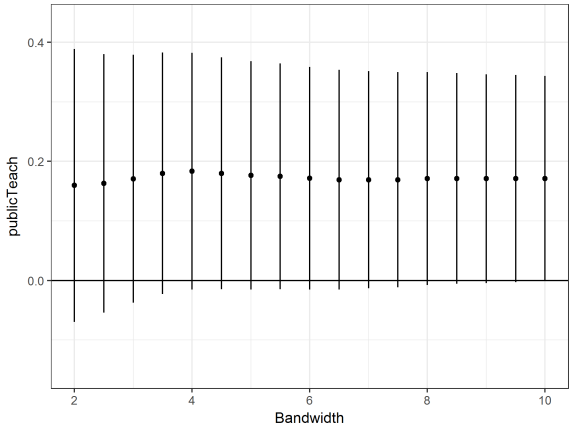

**Notes:** The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks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

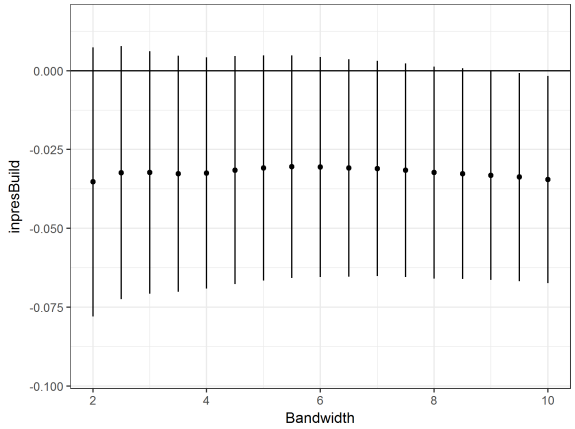
(a) *Public Non-INPRES Primary Buildings*



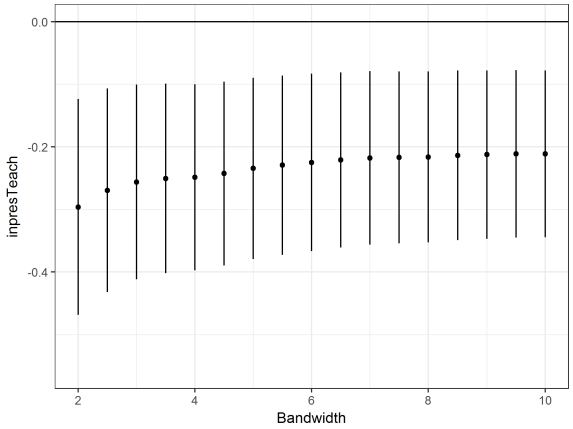

(b) *Public Non-INPRES Primary Teachers*



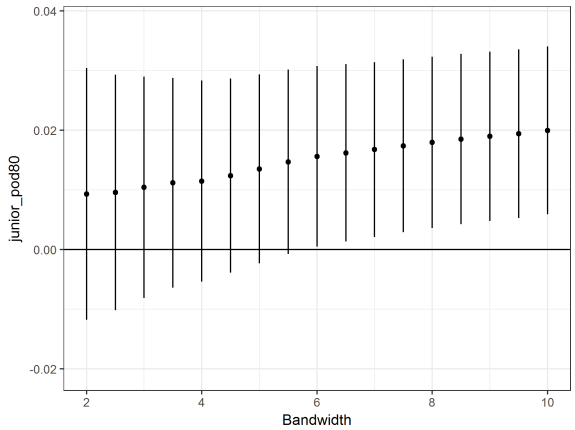

(c) *INPRES Primary Buildings*



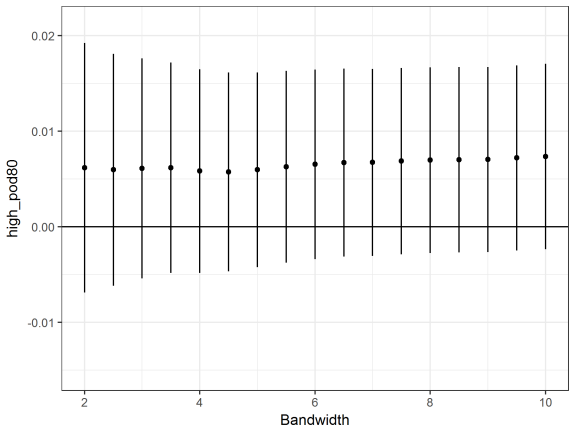

(d) *INPRES Primary Teachers*




(e) *Junior High Schools*



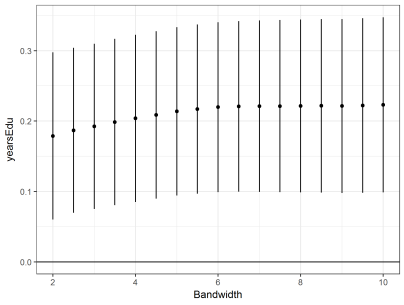

(f) *High Schools*



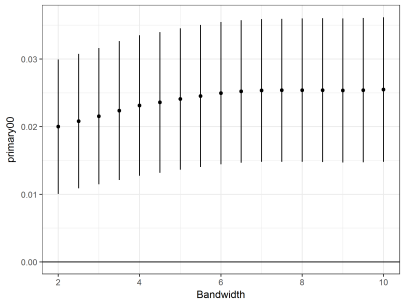

**Notes:** The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks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

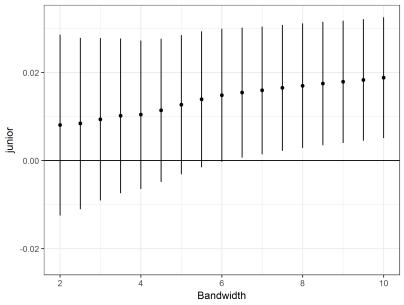
(a) *Years Education 2000*



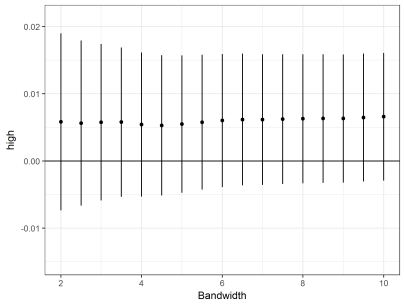

(b) *Primary School 2000*



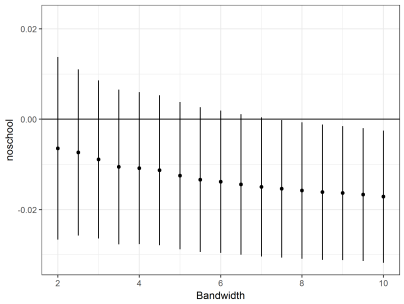

(c) *Junior High 2000*



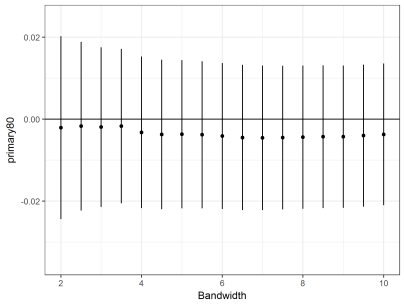

(d) *High School 2000*



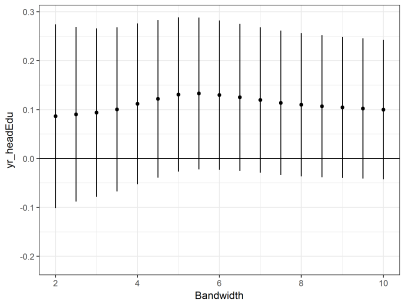

(e) *No School 1980*



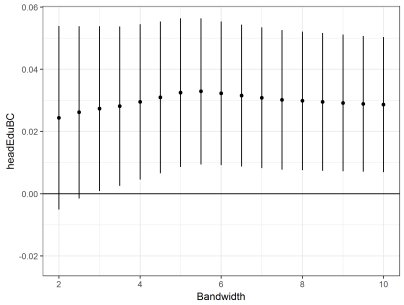

(f) *Primary School 1980*




(g) *Village Head Years Education*



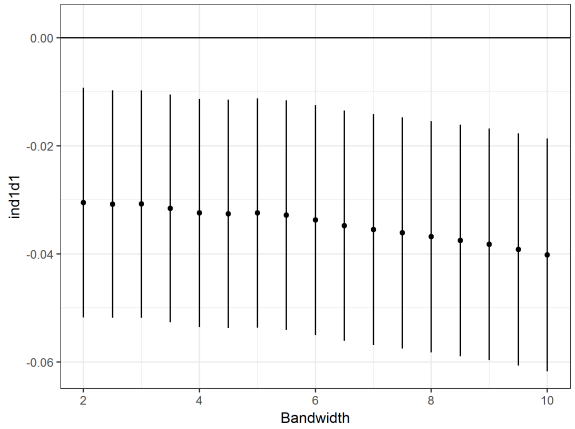

(h) *Village Head High School*



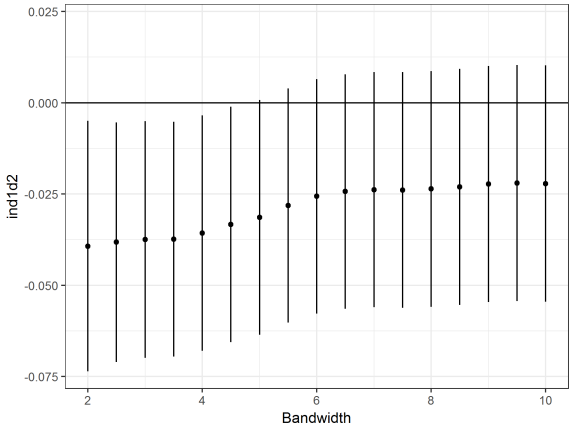

**Notes:** The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks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

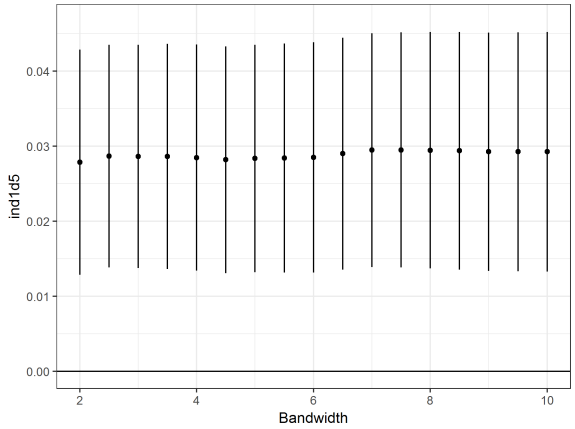
(a) *Agriculture (SUSENAS)*



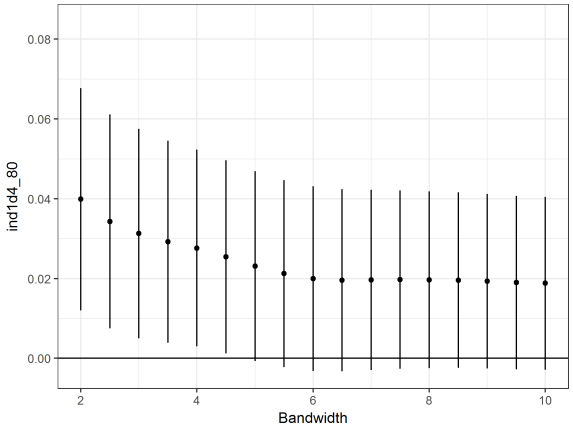

(b) *Agriculture (1980 Census)*



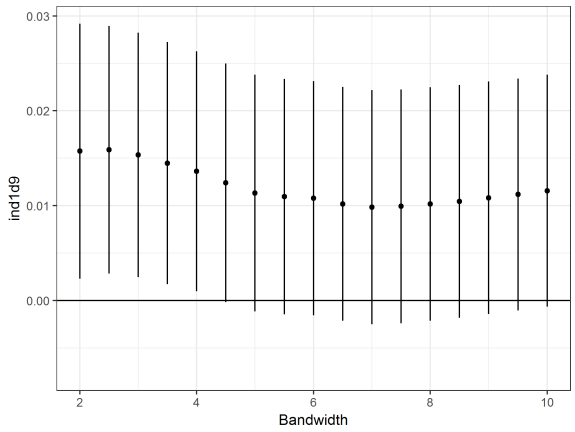

(c) *Manufacturing (SUSENAS)*



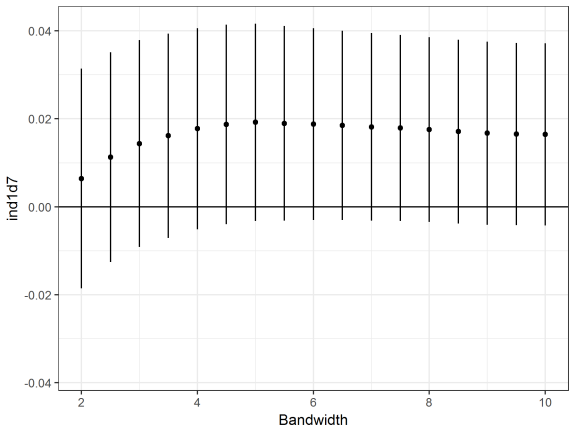

(d) *Manufacturing (1980 Census)*




(e) *Retail (SUSENAS)*



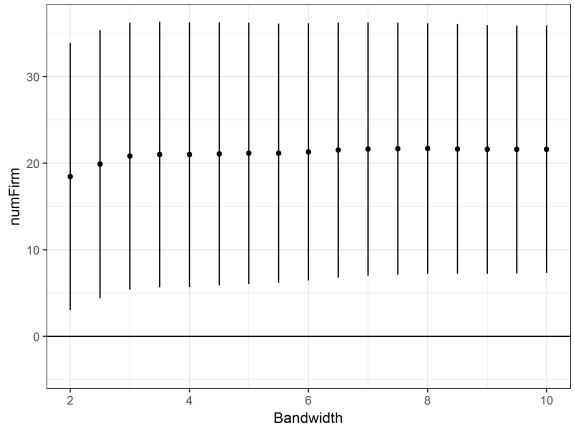

(f) *Commerce (1980 Census)*



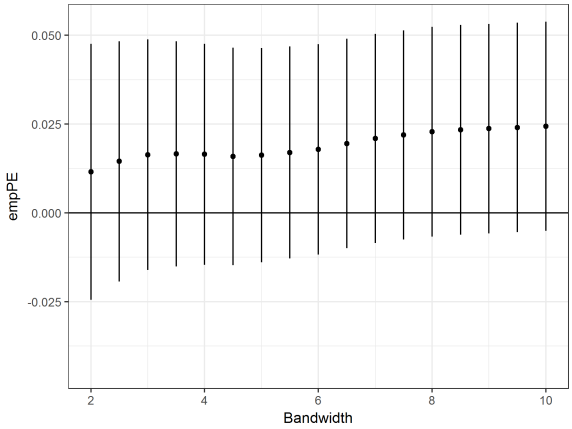

**Notes:** The x axis plots bandwidth, and the y axis plots regression coefficients from equation (2), which includes geographic controls, boundary segment fixed effects, a spline in distance to the nearest historical factory and a spline in distance to the nearest 1830 residency capital, with kinks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

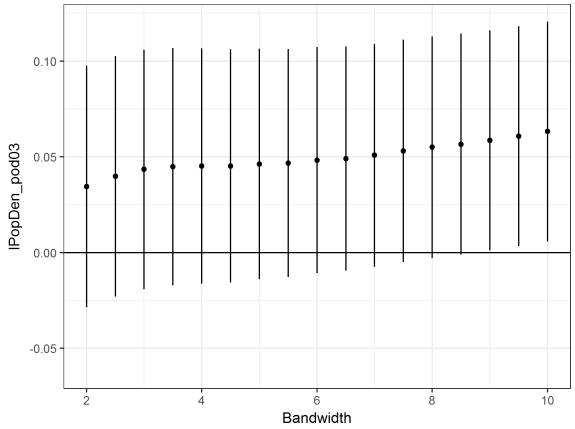
(a) *Number of Firms*



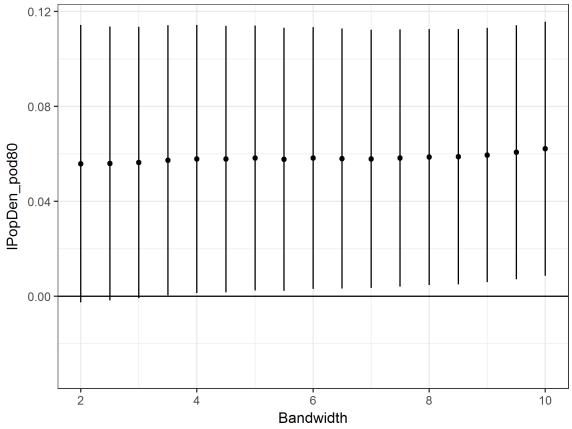

(b) *Employment Share*



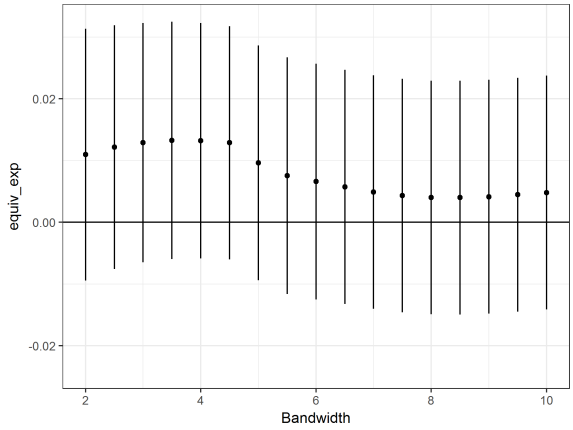

(c) *Log Population Density (2003)*




(d) *Log Population Density (1980)*




(e) *Log Equivalent Consumption*




**Notes:** 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.

# M Single Linear Latitude-Longitude Polynomial

Table M-1: Geographic Characteristics

| | 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 | -2.403 | -0.017 | 0.020 | -0.005 | 0.063 | -0.002 | 0.144 | 0.070 |
| | (0.788)*** | (0.014) | (0.166) | (0.015) | (0.149) | (0.013) | (0.137) | (0.159) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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 M-2: Land Tenure

| | 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 | 1.893 | 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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

Table M-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

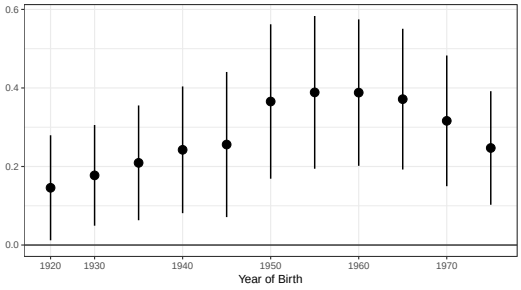
# Table M-4: Education

| | 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 |

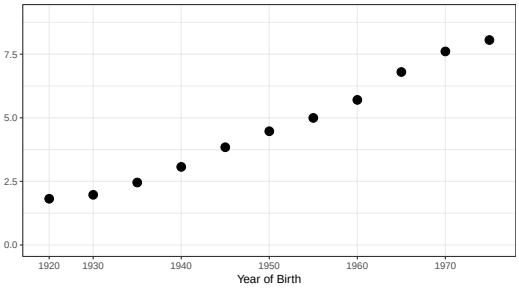
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude. Columns (1) through (6) include gender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

Figure M-1: Education by Cohort (2000 Census)

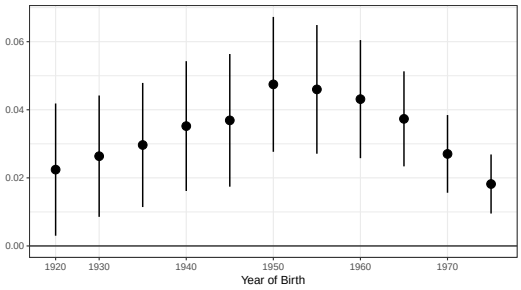
(a) *Years of Schooling*



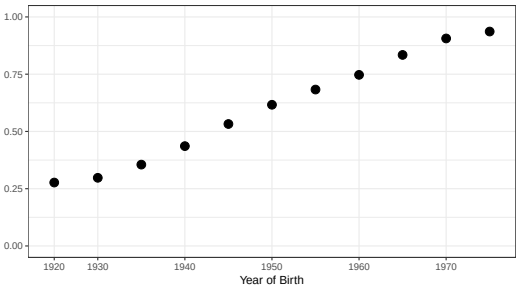

(b) *Years of Schooling Levels*



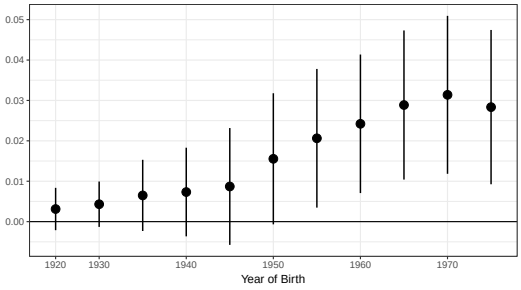

(c) *Primary Completion*



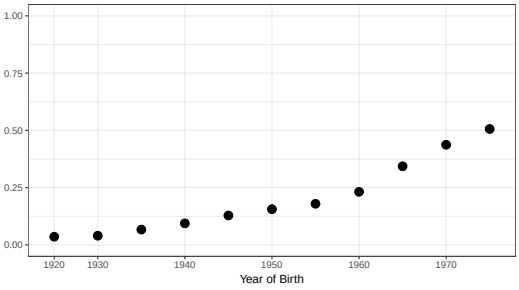

(d) *Primary Completion Levels*



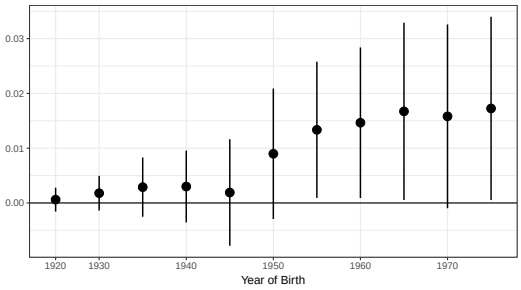

(e) *Junior High Completion*



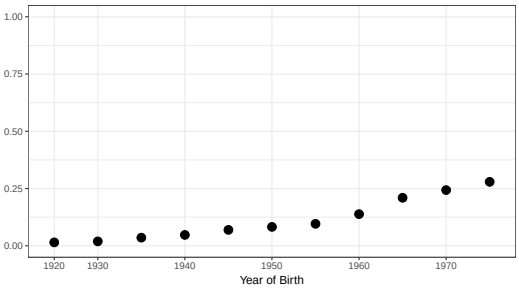

(f) *Junior High Completion Levels*




(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 M-5: Industry

| | 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 M-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# N Quadratic Latitude-Longitude Polynomial

## Table N-1: Geographic Characteristics

| | 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 | -2.429 | -0.018 | 0.027 | -0.006 | 0.092 | -0.003 | 0.117 | 0.092 |
| | (0.786)*** | (0.014) | (0.165) | (0.015) | (0.131) | (0.013) | (0.125) | (0.155) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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 N-2: Land Tenure

| | 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 | 1.998 | 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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

## Table N-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High Schools | High Schools |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | | |
| | (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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

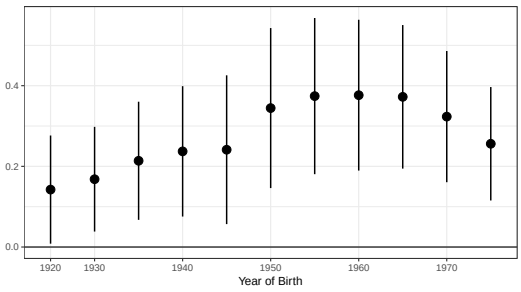
# Table N-4: Education

| | 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 |

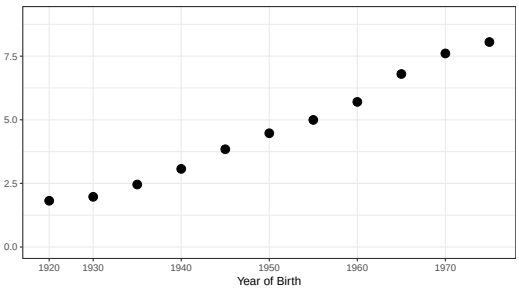
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a quadratic polynomial in latitude and longitude. Columns (1) through (6) include gender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

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

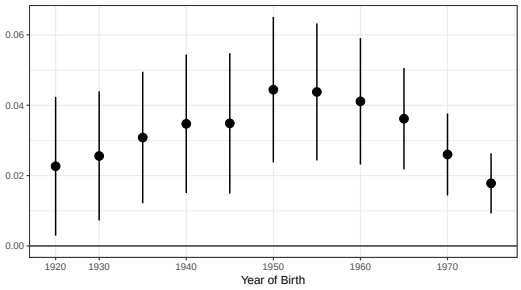
(a) *Years of Schooling*



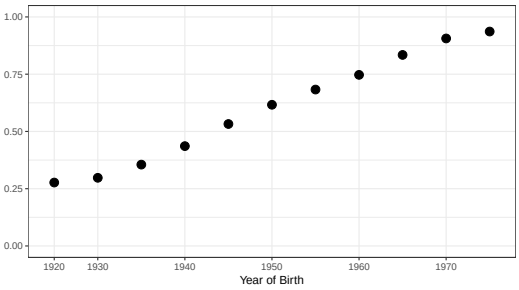

(b) *Years of Schooling Levels*



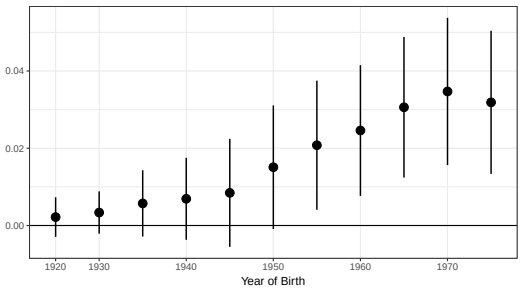

(c) *Primary Completion*



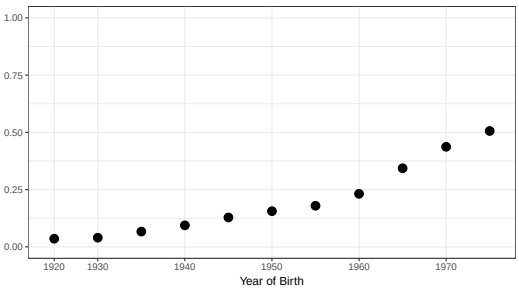

(d) *Primary Completion Levels*



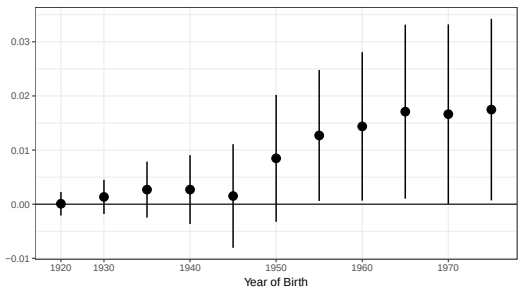

(e) *Junior High Completion*



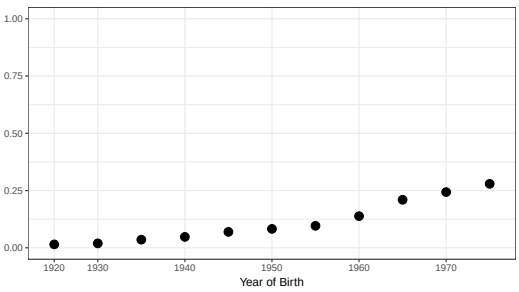

(f) *Junior High Completion Levels*




(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 N-5: Industrial Structure

| | 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 N-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# O One-Dimensional Linear RD Polynomial

Table O-1: Geographic Characteristics

| | 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 | -0.045 | 0.015 | 0.135 | -0.039 | 0.085 | -0.018 | 0.219 | -0.034 |
| | (0.779) | (0.013) | (0.176) | (0.016)** | (0.178) | (0.015) | (0.183) | (0.182) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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 O-2: Land Tenure

| | 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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

## Table O-3: Schools (1980)

| | 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 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. Robust standard errors, clustered by subdistrict, are in parentheses.

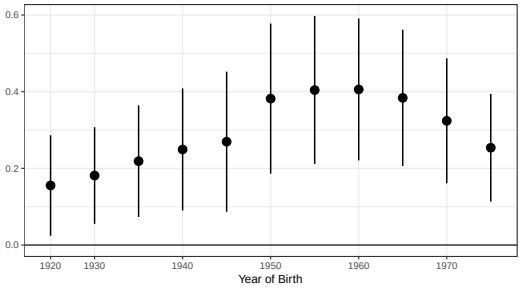
## Table O-4: Education

| | 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 |

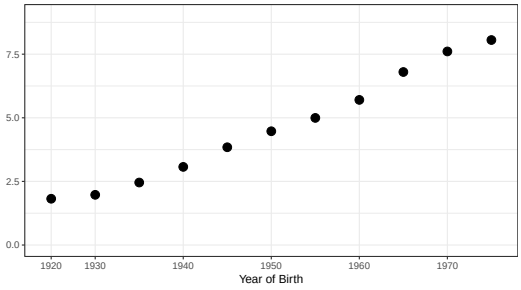
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in distance to the boundary. Columns (1) through (6) include gender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

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

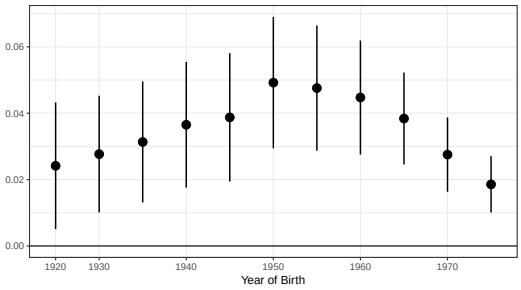
(a) *Years of Schooling*



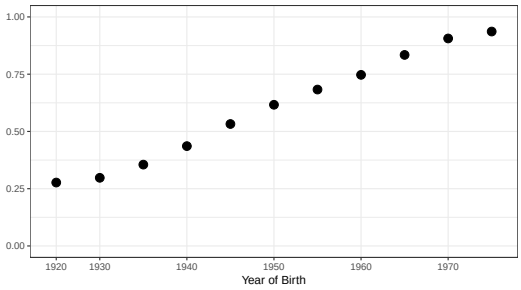

(b) *Years of Schooling Levels*



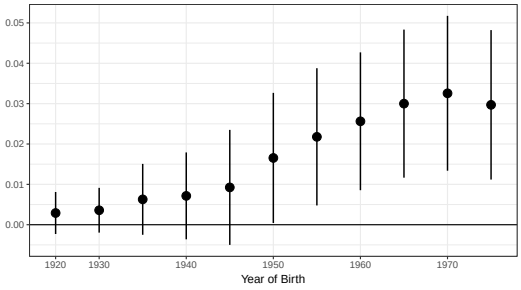

(c) *Primary Completion*



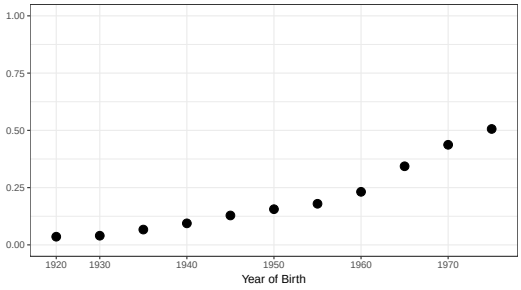

(d) *Primary Completion Levels*



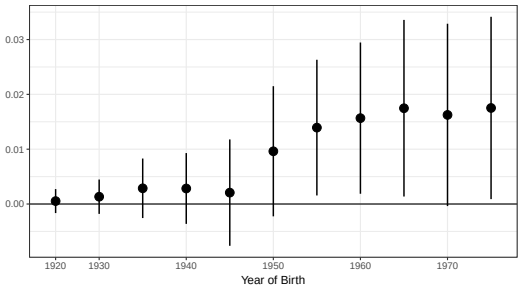

(e) *Junior High Completion*



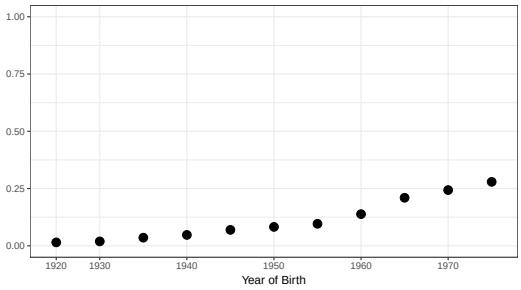

(f) *Junior High Completion Levels*




(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 O-5: Industrial Structure

| | 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 O-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# P 25 Km Boundary Segment Fixed Effects

Table P-1: Geographic Characteristics

| | 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 | -2.097 | -0.014 | 0.026 | -0.023 | 0.328 | -0.001 | 0.380 | 0.021 |
| | (0.845)** | (0.015) | (0.155) | (0.015) | (0.130)** | (0.012) | (0.160)** | (0.177) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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 P-2: Land Tenure

| | 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 | 2.384 | 0.015 | 2.609 | 0.008 | 0.240 | 15.758 | 0.434 | 2.376 |
| | (0.904)*** | (0.003)*** | (1.551)* | (0.006) | (0.297) | (7.429)** | (0.260)* | (1.693) |
| 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 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.

## Table P-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Cultivation | 0.014 | 0.167 | -0.036 | -0.224 | 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 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.

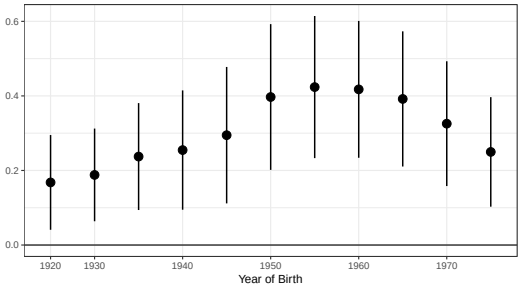
## Table P-4: Education

| | 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 | 0.027 | 0.017 | 0.009 | -0.016 | -0.004 | 0.086 | 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 |

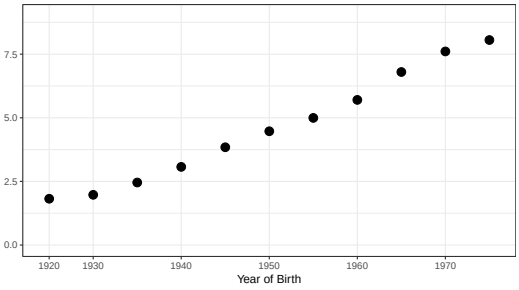
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment 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.

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

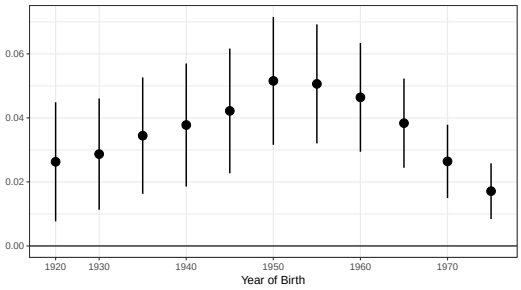
(a) *Years of Schooling*



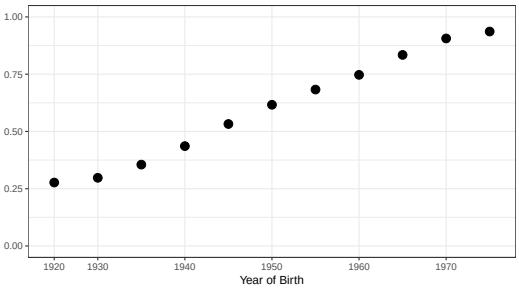

(b) *Years of Schooling Levels*



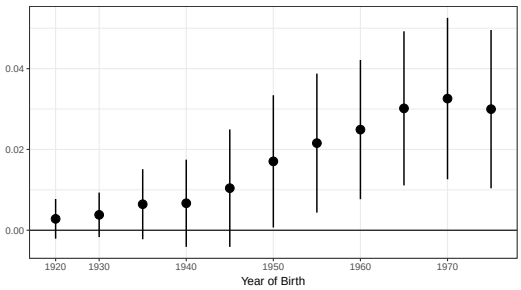

(c) *Primary Completion*



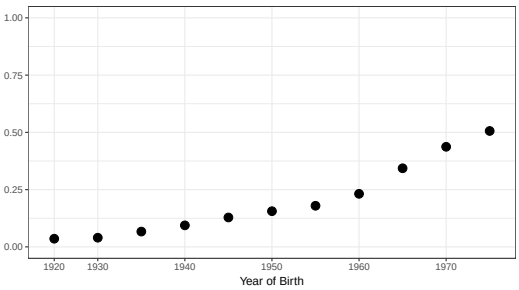

(d) *Primary Completion Levels*



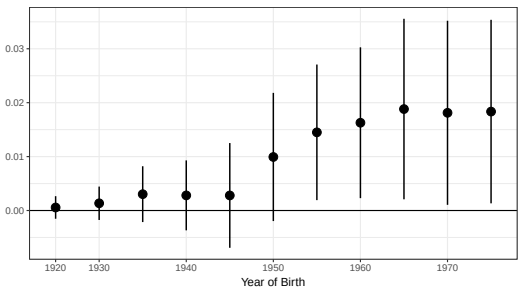

(e) *Junior High Completion*



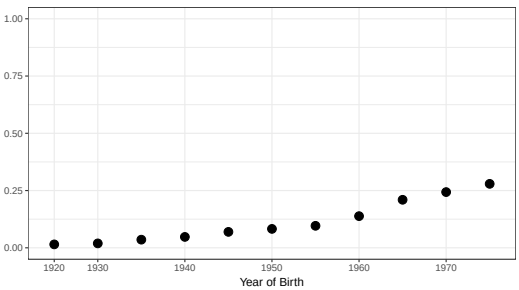

(f) *Junior High Completion Levels*




(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 P-5: Industrial Structure

| | 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 P-6: Firms, Population, and Consumption

| | Num. Manuf. Firms | Manuf. Emp. Share | Log Population Density (2003) | Log Population Density (1980) | Log Equiv. Consumption |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Cultivation | 20.996 | 0.020 | 0.099 | 0.067 | 0.007 |
| | (8.502)** | (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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# Q No Geographic Controls

Table Q-1: Geographic Characteristics

| | 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 | -2.209 | -0.019 | 0.014 | -0.012 | 0.132 | 0.001 | 0.074 | 0.059 |
| | (0.774)*** | (0.014) | (0.164) | (0.016) | (0.114) | (0.012) | (0.105) | (0.141) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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

| | Village Land 2003 | | Village Land 1980 | | 99th Pctile ÷ | 90th Pctile ÷ | 90th Pctile ÷ | 50th Pctile ÷ |
|---|---|---|---|---|---|---|---|---|
| | Total Land | Land Share | Total Land | Land Share | 90th Pctile | 10th Pctile | 50th Pctile | 10th Pctile |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Cultivation | 2.341 | 0.015 | 3.344 | 0.014 | 0.015 | 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 Q-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (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, 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.

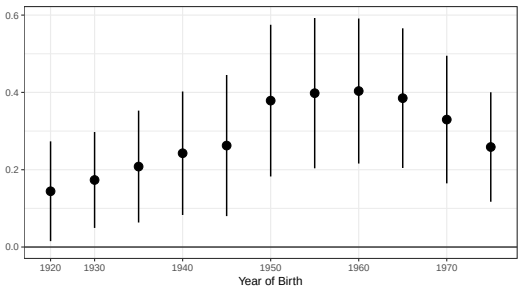
# Table Q-4: Education

| | 2000 Population Census | | | | 1980 Census | | Village Head | |
|---|---|---|---|---|---|---|---|---|
| | Years Education (1) | Primary School (2) | Junior High (3) | High School (4) | No School (5) | Primary School (6) | Years Education (7) | High School (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 |

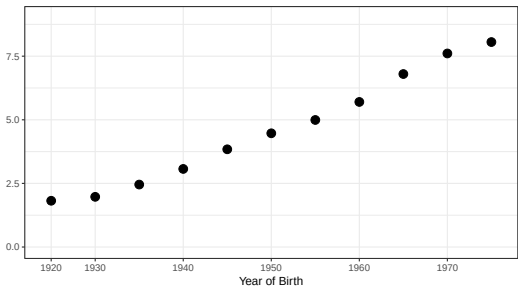
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, a linear spline in distance to the nearest historical factory, and a linear polynomial in latitude and longitude estimated separately for each catchment area. Columns (1) through (6) include gender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

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

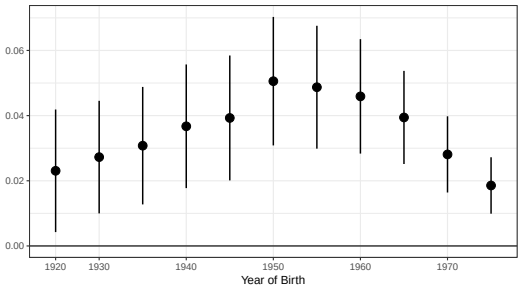
(a) *Years of Schooling*



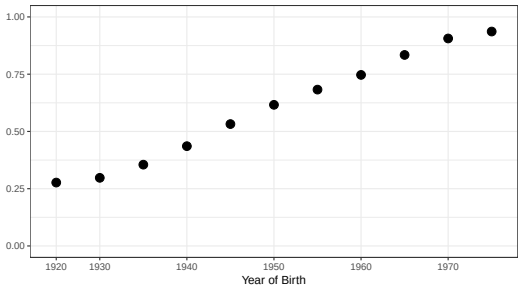

(b) *Years of Schooling Levels*



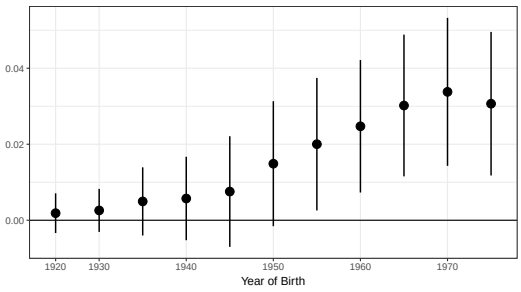

(c) *Primary Completion*



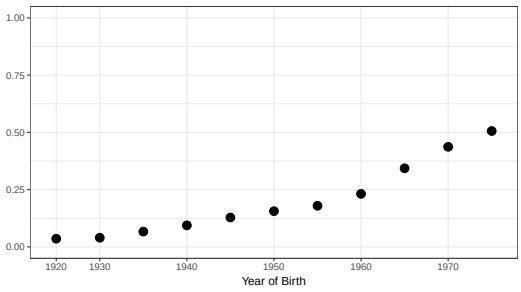

(d) *Primary Completion Levels*



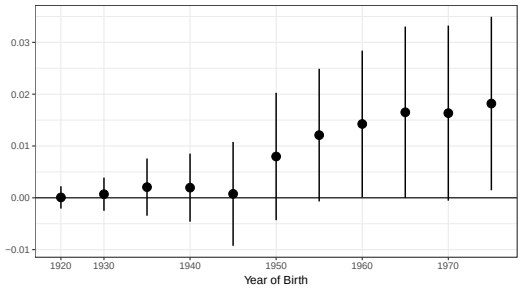

(e) *Junior High Completion*



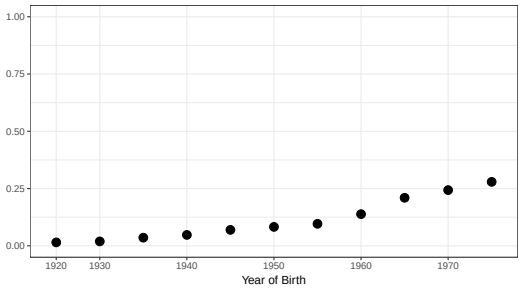

(f) *Junior High Completion Levels*




(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 Q-5: Industrial Structure

| | 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 geographic controls, 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 Q-6: Firms, Population, and Consumption

| | 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)-(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 demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# R Alternative Sample Restriction

Table R-1: Geographic Characteristics

| | 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 | 1.107 | 0.066 | 0.097 | -0.024 | 0.367 | -0.012 | 0.137 | 0.113 |
| | (0.862) | (0.030)** | (0.182) | (0.019) | (0.101)*** | (0.014) | (0.094) | (0.153) |
| Obs | 3,049 | 3,049 | 3,049 | 3,052 | 3,052 | 3,052 | 3,052 | 3,052 |
| Clusters | 326 | 326 | 326 | 326 | 326 | 326 | 326 | 326 |
| Mean | 42.08 | 0.34 | 2.63 | 0.05 | 29.17 | 0.28 | 44.07 | 24.85 |

**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

| | 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 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.

## Table R-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (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 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.

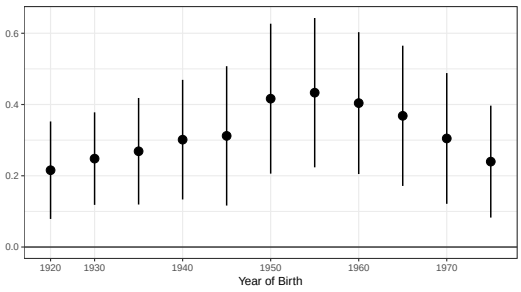
## Table R-4: Education

| | 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 |

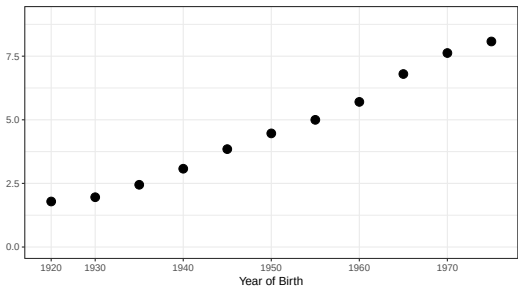
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment 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.

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

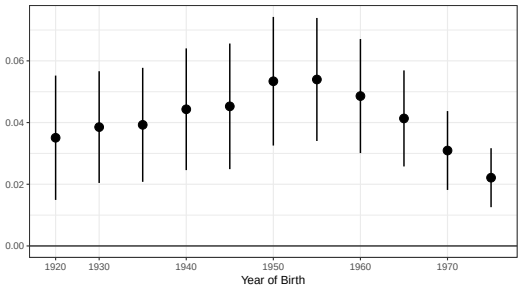
(a) *Years of Schooling*



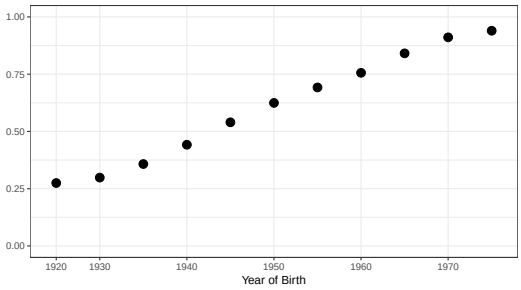

(b) *Years of Schooling Levels*



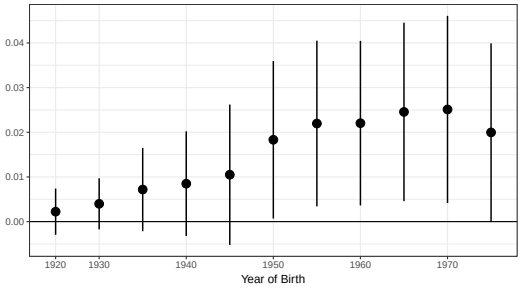

(c) *Primary Completion*



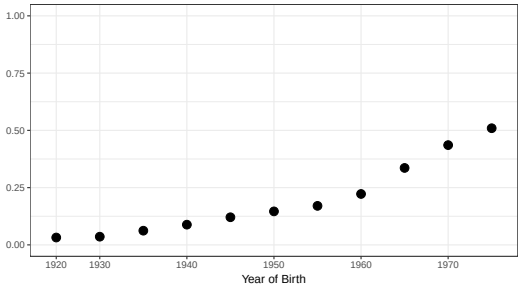

(d) *Primary Completion Levels*



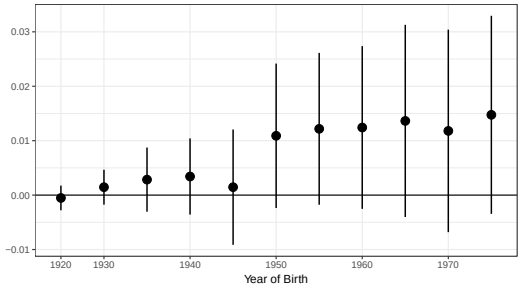

(e) *Junior High Completion*



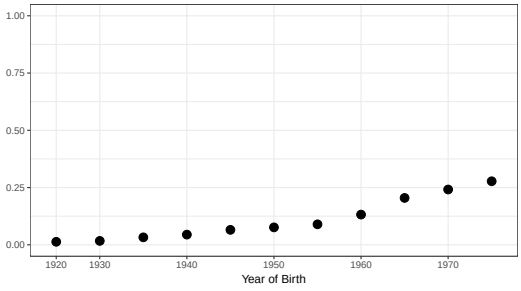

(f) *Junior High Completion Levels*




(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 R-5: Industrial Structure

| | 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 R-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# S Full Sample

Table S-1: Geographic Characteristics

| | 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 | -10.218 | -0.116 | 0.032 | -0.005 | 0.145 | -0.004 | 0.173 | 0.021 |
| | (3.046)*** | (0.049)** | (0.125) | (0.012) | (0.085)* | (0.010) | (0.086)** | (0.115) |
| Obs | 6,367 | 6,367 | 6,362 | 6,375 | 6,375 | 6,375 | 6,375 | 6,375 |
| Clusters | 523 | 523 | 523 | 523 | 523 | 523 | 523 | 523 |
| Mean | 112.06 | 1.08 | 2.55 | 0.04 | 26.88 | 0.26 | 39.29 | 25.30 |

**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

| | 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 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.

Table S-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (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 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.

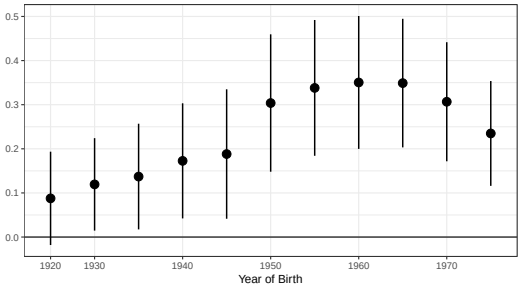
# Table S-4: Education

| | 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 |

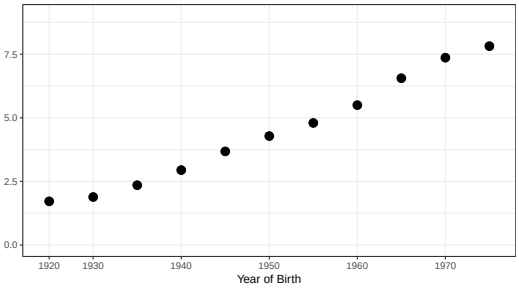
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment 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.

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

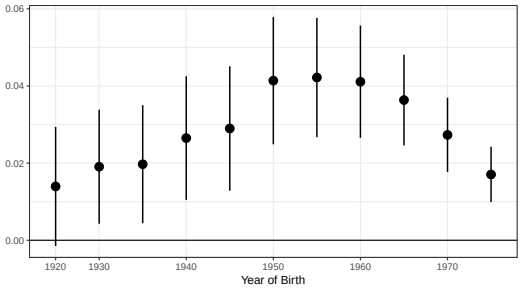
(a) *Years of Schooling*



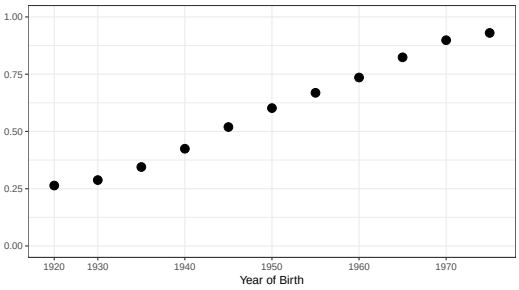

(b) *Years of Schooling Levels*



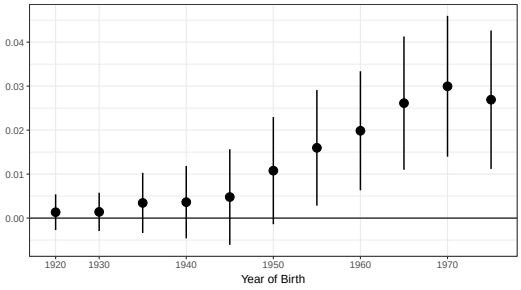

(c) *Primary Completion*



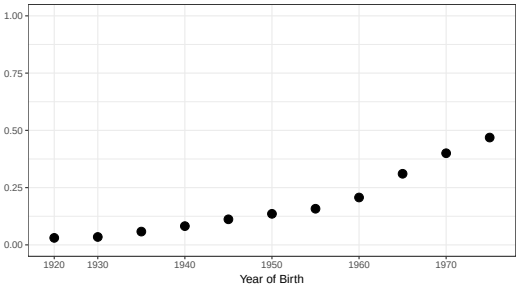

(d) *Primary Completion Levels*



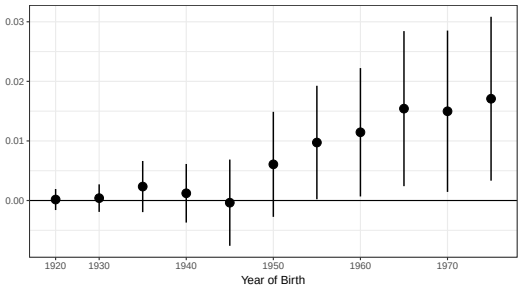

(e) *Junior High Completion*



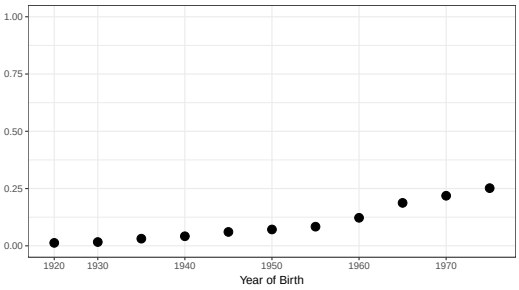

(f) *Junior High Completion Levels*




(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-5: Industrial Structure

| | 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 S-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# T No Triangular Kernel Weighting

Table T-1: Geographic Characteristics

| | 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 | -3.183 | -0.028 | -0.030 | -0.008 | 0.188 | -0.003 | 0.230 | 0.035 |
| | (0.888)*** | (0.015)* | (0.161) | (0.016) | (0.134) | (0.013) | (0.134)* | (0.170) |
| Obs | 4,229 | 4,229 | 4,225 | 4,229 | 4,229 | 4,229 | 4,229 | 4,229 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 31.54 | 0.26 | 2.54 | 0.06 | 25.54 | 0.29 | 32.93 | 24.95 |

**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

| | 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 | 2.019 | 0.016 | 3.744 | 0.017 | -0.029 | 15.769 | 0.444 | 2.479 |
| | (0.861)** | (0.004)*** | (1.582)** | (0.006)*** | (0.248) | (8.435)* | (0.274) | (2.080) |
| Obs | 4,573 | 4,573 | 4,225 | 4,127 | 4,111 | 4,101 | 4,110 | 4,101 |
| Clusters | 383 | 383 | 380 | 380 | 381 | 381 | 381 | 381 |
| Mean | 18.59 | 0.09 | 23.97 | 0.11 | 3.53 | 38.53 | 4.34 | 7.83 |

**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.

Table T-3: Schools (1980)

| | Public Non-INPRES Primary | | INPRES Primary | | Junior High | High |
|---|---|---|---|---|---|---|
| | Buildings | Teachers | Buildings | Teachers | Schools | Schools |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Cultivation | 0.012 | 0.192 | -0.046 | -0.226 | 0.022 | 0.007 |
| | (0.018) | (0.100)* | (0.019)** | (0.078)*** | (0.008)** | (0.006) |
| Obs | 4,225 | 4,225 | 4,225 | 4,225 | 4,225 | 4,225 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 |
| Mean | 0.43 | 2.82 | 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, 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.

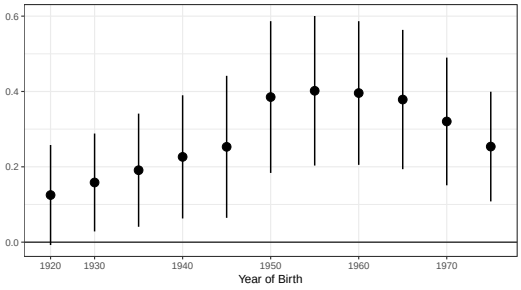
# Table T-4: Education

| | 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 | 0.227 | 0.026 | 0.016 | 0.007 | -0.022 | -0.001 | 0.069 | 0.025 |
| | (0.078)*** | (0.007)*** | (0.008)** | (0.006) | (0.009)** | (0.010) | (0.085) | (0.013)* |
| Obs | 16,187,628 | 16,187,628 | 16,187,628 | 16,187,628 | 655,852 | 655,727 | 26,767 | 26,767 |
| 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 |

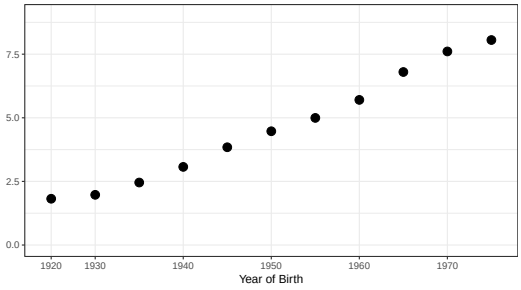
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment 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.

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

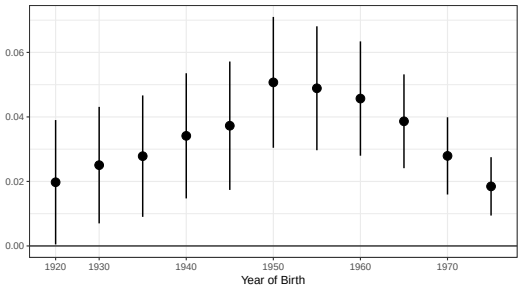
(a) *Years of Schooling*



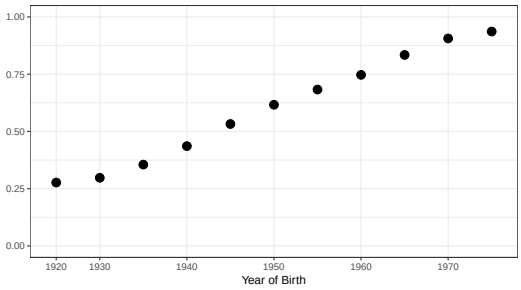

(b) *Years of Schooling Levels*



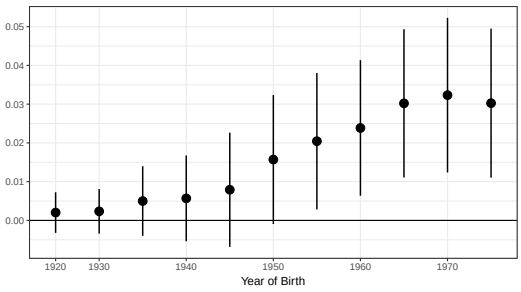

(c) *Primary Completion*



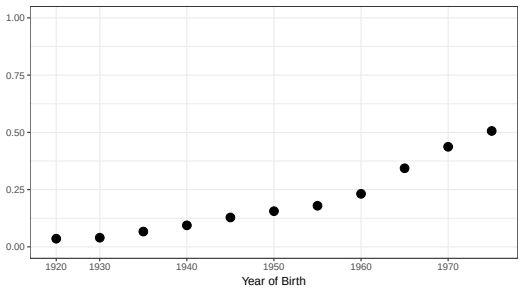

(d) *Primary Completion Levels*



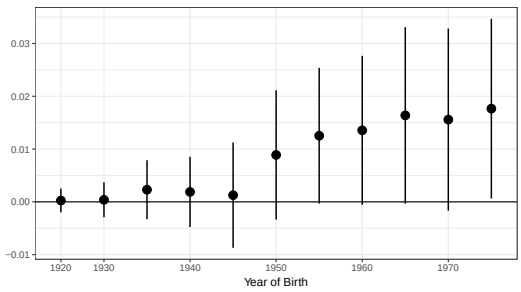

(e) *Junior High Completion*



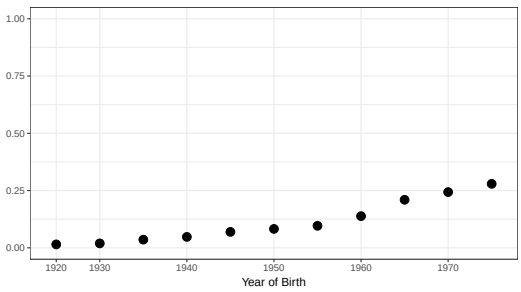

(f) *Junior High Completion Levels*




(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-5: Industrial Structure

| | 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 | 130,810 | 130,810 | 130,810 | 128,393 | 128,393 | 128,393 |
| 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 T-6: Firms, Population, and Consumption

| | Num. Manuf. Firms | Manuf. Emp. Share | Log Population Density (2003) | Log Population Density (1980) | Log Equiv. Consumption |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Cultivation | 22.631 | 0.031 | 0.090 | 0.078 | 0.006 |
| | (8.382)*** | (0.019) | (0.036)** | (0.034)** | (0.011) |
| Obs | 4,572 | 4,572 | 4,573 | 4,127 | 144,573 |
| Clusters | 383 | 383 | 383 | 380 | 381 |
| Mean | 71.50 | 0.16 | 2.87 | 2.54 | 12.55 |

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

# U Instrumental Variables

Table U-1: Geographic Characteristics

| | 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 | -2.874 | -0.024 | 0.019 | -0.015 | 0.172 | 0.001 | 0.096 | 0.077 |
| | (0.960)*** | (0.017) | (0.205) | (0.019) | (0.142) | (0.015) | (0.131) | (0.176) |
| Obs | 4,209 | 4,209 | 4,205 | 4,209 | 4,209 | 4,209 | 4,209 | 4,209 |
| Clusters | 380 | 380 | 380 | 380 | 380 | 380 | 380 | 380 |
| F stat | 1308 | 1308 | 1321 | 1308 | 1308 | 1308 | 1308 | 1308 |
| Mean | 31.60 | 0.26 | 2.54 | 0.06 | 25.50 | 0.29 | 32.94 | 24.89 |

**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

| | 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 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.

## Table U-3: Schools (1980)

| | 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 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.

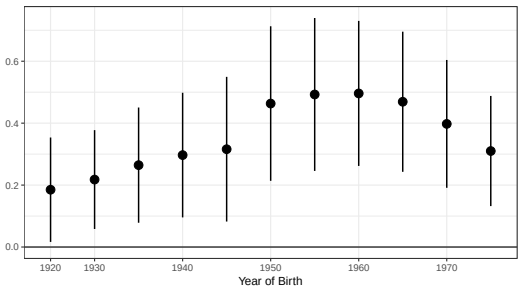
## Table U-4: Education

| | 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 |

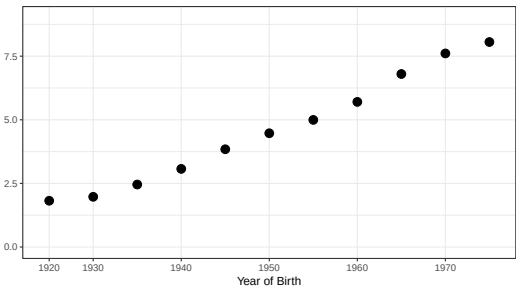
**Notes:** The unit of observation is the individual. Regressions include boundary segment fixed effects, linear splines in distance to the nearest historical factory and residency capital, geographic controls, and a linear polynomial in latitude and longitude estimated separately for each catchment 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.

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

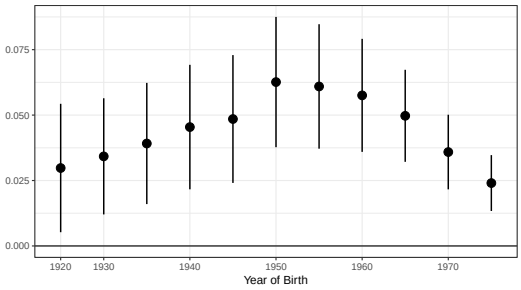
(a) *Years of Schooling*



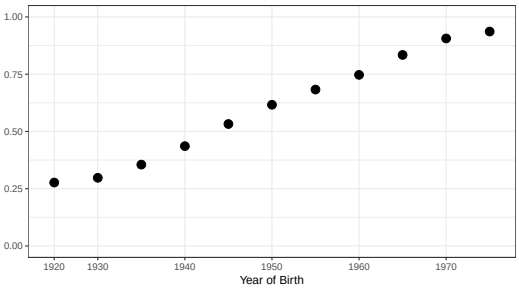

(b) *Years of Schooling Levels*



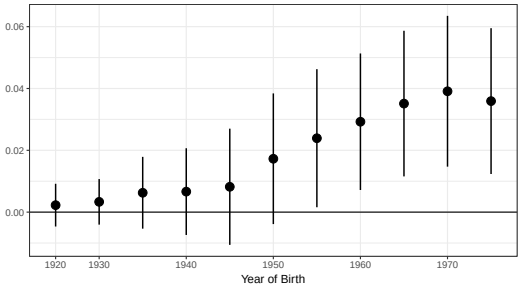

(c) *Primary Completion*



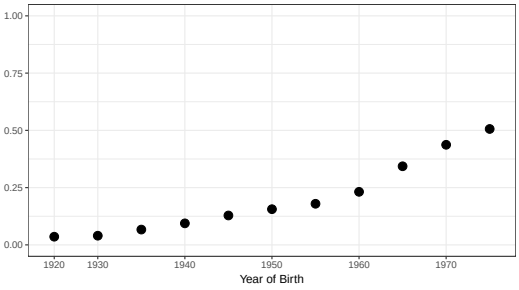

(d) *Primary Completion Levels*



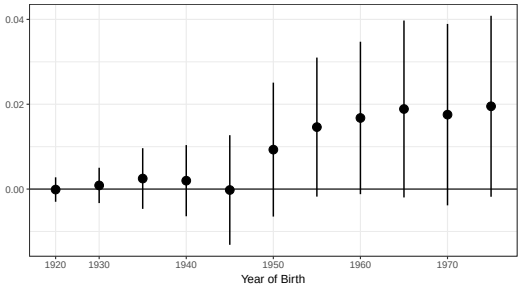

(e) *Junior High Completion*



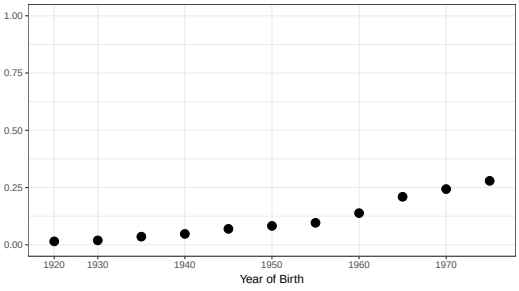

(f) *Junior High Completion Levels*




(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-5: Industrial Structure

| | 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 U-6: Firms, Population, and Consumption

| | 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)-(4) and the household in column (5). 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. Column (5) includes demographic controls and survey year fixed effects. Robust standard errors, clustered by subdistrict, are in parentheses.

# V Grid Cells

Table V-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 | -6.132 | -0.050 | -0.457 | -0.048 | 0.515 | -0.088 | 0.191 |
| | (2.995)** | (0.064) | (0.396) | (0.075) | (0.572) | (0.084) | (0.617) |
| Obs | 553 | 553 | 553 | 553 | 553 | 553 | 553 |
| Clusters | 71 | 71 | 71 | 71 | 71 | 71 | 71 |
| Mean | 33.65 | 0.31 | 5.23 | 0.10 | 24.04 | 0.27 | 32.64 |

**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.

# W Additional Public Goods Outcomes

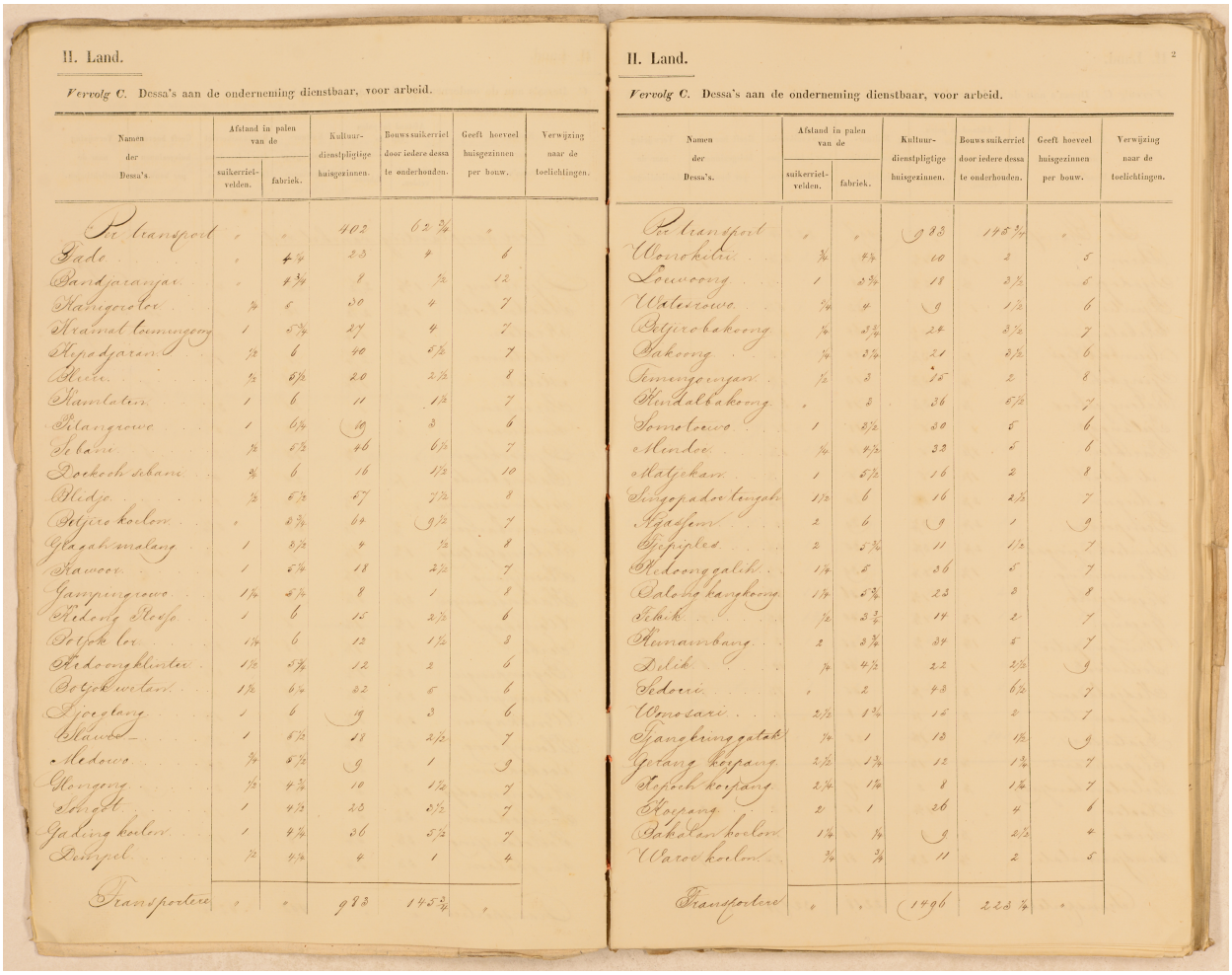
Table W-1: Public Goods

| | 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 | -0.468 |
| | (0.272) | (1.255) | (0.115) | (0.025)** | (0.014) | (0.026) | (0.200)** |
| 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.

# X Data Appendix

Figure X-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. |
|---|---|---|---|---|---|---|
| Per transport | " | " | 402 | 62 3/4 | " | |
| Dado | " | 4 1/4 | 23 | 4 | 6 | |
| Pandjarangan | " | 4 3/4 | 9 | 1/2 | 12 | |
| Hahngerlor | 1/4 | 5 | 30 | 4 | 7 | |
| Kramat toemoengoeng | 1 | 5 3/4 | 27 | 4 | 7 | |
| Kepadjaran | 1/2 | 6 | 40 | 5 1/2 | 7 | |
| Hoete | 1/2 | 6 1/2 | 20 | 2 1/2 | 8 | |
| Kamlaten | 1 | 6 | 11 | 1 1/2 | 7 | |
| Silangroewe | 1 | 6 1/4 | 19 | 3 | 6 | |
| Sebane | 1/2 | 6 1/2 | 46 | 6 1/2 | 7 | |
| Dockoeh sebane | 3/4 | 6 | 16 | 1 1/2 | 10 | |
| Blidjo | 1/2 | 6 1/2 | 57 | 7 1/2 | 8 | |
| Peljerekoelon | " | 2 3/4 | 44 | 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 | |
| Kedang Rejo | 1 | 6 | 15 | 2 1/2 | 6 | |
| Pottok lor | 1 1/4 | 6 | 12 | 1 1/2 | 8 | |
| Kedoongklinter | 1 1/2 | 5 3/4 | 12 | 2 | 6 | |
| Pogokwetan | 1 1/2 | 6 1/4 | 22 | 5 | 6 | |
| Tjoeplang | 1 | 6 | 19 | 3 | 6 | |
| Klaboe | 1 | 6 1/2 | 18 | 2 1/2 | 7 | |
| Medowo | 3/4 | 6 1/2 | 9 | 1 | 9 | |
| Rongeng | 1/2 | 4 3/4 | 10 | 1 1/2 | 7 | |
| Schrott | 1 | 4 1/2 | 23 | 3 1/2 | 7 | |
| Gading koelon | 1 | 4 3/4 | 36 | 5 1/2 | 7 | |
| Dawoeh | 1/2 | 4 1/4 | 4 | 1 | 4 | |
| Transporteren | " | " | 983 | 145 3/4 | " | |

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. |
|---|---|---|---|---|---|---|
| Per transport | " | " | 983 | 145 3/4 | " | |
| Wonokoloo | 3/4 | 4 1/4 | 40 | 3 | 5 | |
| Lowoong | 1 | 2 3/4 | 18 | 2 1/2 | 6 | |
| Waterowe | 3/4 | 4 | 9 | 1 1/2 | 6 | |
| Peljerobakoong | 1/4 | 2 3/4 | 24 | 3 1/2 | 7 | |
| Bakoong | 1/4 | 2 1/4 | 21 | 3 1/2 | 6 | |
| Temengoengan | 1/2 | 3 | 15 | 2 | 8 | |
| Kendalbakoong | " | 3 | 36 | 5 1/2 | 7 | |
| Somolowo | 1 | 2 1/2 | 30 | 5 | 6 | |
| Mindoo | 1/4 | 4 1/2 | 33 | 5 | 6 | |
| Malyekam | 1 | 5 1/2 | 16 | 2 | 8 | |
| Singopadoelingah | 1 1/2 | 6 | 16 | 2 1/2 | 7 | |
| Ngadjeman | 2 | 6 | 9 | 1 | 9 | |
| Tjipiples | 2 | 5 3/4 | 11 | 1 1/2 | 7 | |
| Kedoonggalih | 1 1/4 | 5 | 36 | 5 | 7 | |
| Balolokawoong | 1 1/4 | 5 3/4 | 23 | 3 | 8 | |
| Schuk | 1/2 | 3 3/4 | 14 | 2 | 7 | |
| Kenambang | 2 | 3 3/4 | 34 | 5 | 7 | |
| Dolok | 1/4 | 4 1/2 | 22 | 2 1/2 | 9 | |
| Sedoeni | " | 2 | 43 | 6 1/2 | 7 | |
| Wonosaxi | 2 1/2 | 1 3/4 | 15 | 2 | 7 | |
| Tjangkringgatak | 1/4 | 1 | 13 | 1 1/2 | 9 | |
| Gedang koepang | 2 1/2 | 1 3/4 | 12 | 1 3/4 | 7 | |
| Soepoch 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 | |
| Warokoelon | 3/4 | 3/4 | 11 | 2 | 5 | |
| Transporteren | " | " | 1496 | 223 1/4 | " | |

# Figure X-2: 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 | | 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 | BPS | Hectares of Agricultural Land Used |
| Atlas van Nederlandsch-Indie | 1856 | Harvard Library | 1830 Residency and 1856 Regency Capitals |
| Commissie Umbgrove | 1858 | | Subjected Villages, Distance to Historical Factory |
| Author's Calculations (Based on Commissie Umbgrove) | N/A | | Catchment Areas |

# Y 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 $\Sigma_B$ 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.