Urban Growth Shadows
Online Appendix

David Cuberes
Clark University

Klaus Desmet
SMU, NBER and CEPR

Jordan Rappaport
Kansas City Fed

February 2021

A Additional Tables

A.1 Constant Distances across Time Periods

In the empirical analysis in the main text of the paper we specify our regressions to include indicators for 50km intervals out to a maximum distance that leaves at least 50 percent of the observations in the excluded category. As a robustness check, in what follows we report versions of the main tables, keeping the maximum distance constant across time periods.

Table A.1: Population Growth and the Presence of a Moderately Large Neighbor (using the same distance bins for all time periods)

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	1840-1860	1860-1880	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000	2000-2017
1 to 50 km	-0.63*	-0.77***	-1.10***	-0.19	0.15	0.99***	1.29***	0.59***	0.41**
	(0.35)	(0.28)	(0.21)	(0.15)	(0.14)	(0.17)	(0.25)	(0.18)	(0.20)
50 to 100 km	0.67***	-0.41**	-0.86***	-0.42***	-0.20*	0.04	0.31	0.35**	0.11
	(0.24)	(0.20)	(0.18)	(0.12)	(0.10)	(0.12)	(0.20)	(0.15)	(0.09)
100 to 150 km	-0.30***	-0.11	-0.63***	-0.27**	-0.24***	-0.22**	0.17	0.14	0.03
	(0.12)	(0.17)	(0.15)	(0.13)	(0.09)	(0.10)	(0.18)	(0.13)	(0.07)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.846	0.778	0.718	0.522	0.395	0.353	0.377	0.413	0.298
Incremental R^2	0.002	0.002	0.008	0.003	0.005	0.030	0.034	0.011	0.004

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. Specification is identical to that of Table 1 in the main text except that the same distance bins are used for all time periods. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As can be seen when comparing Table A.1 to Table 1 in the main text, the results are qualitatively similar. One minor difference is worth discussing: in Table A.1 the positive coefficient on the presence of a large neighbor less than 50km away for the period 1920-1940 is smaller and no longer statistically significant, whereas expanding the maximum distance to 150km now yields a negative and statistically significant coefficient on the 100 to 150km distance interval for both the 1920-1940 and 1940-1960 periods. These findings are related. With the economy transitioning from the negative regime to the positive regime during the period 1920-1940, we should not be surprised to see a negative coefficient on the presence of a large neighbor at more far-off distances. This negative coefficient on the 100 to 150km shifts down the coefficients at shorter distances, explaining the smaller coefficient on the presence of a large neighbor within 50km. We also generate versions of Tables 3 and 4 of the main text that keep the maximum distance constant across time periods. These are reported in Tables A.2 and A.3. The results are largely unchanged.

Table A.2: Population Growth and the Presence of a Very Large Neighbor (using the same distance bins for all time periods)

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance Nearest Neighbor with Pop $\geq$ 99th Pctile	1840-1860	1860-1880	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000	2000-2017
1 to 50 km	0.77	-1.51***	-0.78*	-0.22	0.66**	2.41***	3.20***	0.99*	-0.26
	(0.63)	(0.46)	(0.42)	(0.28)	(0.28)	(0.26)	(0.61)	(0.58)	(0.39)
50 to 100 km	-0.20	-1.18***	-0.86***	-0.58***	0.15	0.55***	0.97***	1.03***	0.65***
	(0.43)	(0.40)	(0.26)	(0.19)	(0.17)	(0.15)	(0.25)	(0.19)	(0.14)
100 to 150 km	0.18	-1.11***	-0.45**	-0.41**	0.01	0.14	0.30**	0.51***	0.35***
	(0.43)	(0.33)	(0.22)	(0.18)	(0.17)	(0.13)	(0.14)	(0.13)	(0.10)
150 to 200 km	0.39	-1.28***	-0.14	-0.47***	0.08	0.02	0.13	0.26***	0.24***
	(0.33)	(0.31)	(0.21)	(0.16)	(0.14)	(0.13)	(0.12)	(0.10)	(0.07)
200 to 250 km	-0.11	-0.96***	-0.02	-0.14	0.12	-0.12	0.03	0.10	0.18***
	(0.17)	(0.27)	(0.15)	(0.13)	(0.12)	(0.13)	(0.11)	(0.08)	(0.06)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.844	0.785	0.712	0.521	0.392	0.362	0.393	0.436	0.320
Incremental R^2	0.001	0.008	0.003	0.003	0.003	0.039	0.050	0.035	0.027

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 99th percentile is within the enumerated distance bin. Specification is identical to that of Table 3 in the main text except that the same distance bins are used for all time periods. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Regional Variation during the U.S. Westward Expansion

In this subsection we aim to understand to what extent the existence of urban shadows might have been related to the westward expansion of the U.S. during the 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries.

Table A.3: Population Growth and the Size of Large Neighbors (using the same distance bins for all time periods)

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance Nearest Neighbor with Pop	1840-1860	1860-1880	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000	2000-2017
$\geq$ 80th Pctile									
1 to 50 km	-0.40*	-0.66***	-0.21*	-0.26***	0.12	-0.05	0.06	0.14**	0.05
	(0.21)	(0.16)	(0.12)	(0.09)	(0.08)	(0.09)	(0.07)	(0.06)	(0.05)
$\geq$ 90th Pctile									
1 to 50 km	-0.35	0.21	-0.37**	-0.12	-0.11	0.39***	0.11	0.28***	0.00
	(0.34)	(0.24)	(0.19)	(0.19)	(0.16)	(0.15)	(0.10)	(0.11)	(0.08)
50 to 100 km	-0.29	-0.15	-0.29*	-0.35***	-0.09	0.13	-0.07	0.10	0.00
	(0.24)	(0.19)	(0.16)	(0.13)	(0.09)	(0.12)	(0.08)	(0.07)	(0.06)
$\geq$ 95th Pctile									
1 to 50 km	-0.19	-0.37	-0.74***	0.03	0.09	0.34*	0.80***	0.18	0.38*
	(0.44)	(0.31)	(0.23)	(0.23)	(0.17)	(0.20)	(0.20)	(0.17)	(0.21)
50 to 100 km	-0.47	-0.21	-0.54***	-0.10	-0.16	-0.17	0.14	0.13	0.03
	(0.30)	(0.25)	(0.18)	(0.15)	(0.12)	(0.15)	(0.19)	(0.15)	(0.09)
100 to 150 km	-0.35***	0.00	-0.60***	-0.21*	-0.24**	-0.27***	0.10	0.04	-0.01
	(0.13)	(0.20)	(0.16)	(0.12)	(0.10)	(0.10)	(0.18)	(0.12)	(0.06)
$\geq$ 99th Pctile									
1 to 50 km	0.91	-1.00**	-0.10	-0.23	0.49*	1.88***	2.42***	0.67	-0.61
	(0.78)	(0.45)	(0.47)	(0.29)	(0.26)	(0.28)	(0.58)	(0.50)	(0.39)
50 to 100 km	0.24	-0.86**	-0.54**	-0.44**	0.23	0.59***	0.98***	0.96***	0.63***
	(0.43)	(0.38)	(0.26)	(0.19)	(0.16)	(0.15)	(0.24)	(0.17)	(0.15)
100 to 150 km	0.37	-1.03***	-0.16	-0.29*	0.08	0.22*	0.26**	0.49***	0.35***
	(0.43)	(0.34)	(0.23)	(0.17)	(0.18)	(0.13)	(0.13)	(0.11)	(0.09)
150 to 200 km	0.49	-1.22***	-0.05	-0.42***	0.11	0.03	0.12	0.25***	0.23***
	(0.31)	(0.31)	(0.20)	(0.15)	(0.14)	(0.13)	(0.11)	(0.09)	(0.07)
200 to 250 km	-0.02	-0.87***	0.02	-0.15	0.13	-0.09	0.01	0.11	0.19***
	(0.16)	(0.26)	(0.15)	(0.13)	(0.12)	(0.13)	(0.10)	(0.07)	(0.06)
Additional Controls	48	52	52	52	52	52	52	52	52
N	691	1,328	1,844	2,110	2,357	2,387	2,283	2,104	1,895
Adjusted R^2	0.847	0.787	0.719	0.526	0.395	0.376	0.411	0.447	0.322
Incremental R^2	0.005	0.012	0.011	0.009	0.008	0.054	0.070	0.047	0.030

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above different thresholds is within the enumerated distance bin. These thresholds are nested: a neighbor that is above the 99th percentile is also above the 90th and 95th percentiles. Coefficients on these latter thresholds thus estimate the marginal boost to predicted growth compared to having a neighbor with population only above the next highest threshold. Specification is identical to that of Table 4 in the main text except that the same distance bins are used for all time periods. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

If during that time period locations in the West, much of them isolated, grew fast because they were becoming settled, this would contribute to a negative correlation between local growth and proximity to a large neighbor. To assess this possibility, we rerun our baseline regression for three separate regions: the East, corresponding to the states covering the area of the original 13 colonies; the Middle, corresponding to all other states east of the Mississippi River; and the West, corresponding to all states west of the Mississippi River.

Table A.4: Population Growth and Large Neighbors by Region, 1860 to 1920

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	East Region			Middle Region			West Region		
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	(1) 1860-1880	(2) 1880-1900	(3) 1900-1920	(4) 1860-1880	(5) 1880-1900	(6) 1900-1920	(7) 1860-1880	(8) 1880-1900	(9) 1900-1920
1 to 50 km	-0.28	-0.31***	0.50**	-0.58**	-0.62***	0.06	-2.11	-1.78***	-0.13
	(0.31)	(0.10)	(0.20)	(0.24)	(0.13)	(0.08)	(1.88)	(0.45)	(0.27)
50 to 100 km	-0.21			-0.46***	-0.52***		-2.08	-1.74***	-0.38*
	(0.26)			(0.17)	(0.14)		(1.48)	(0.31)	(0.20)
100 to 150 km							-2.59**	-1.41***	-0.07
							(1.29)	(0.28)	(0.18)
150 to 200 km							-2.30**	-1.17***	
							(1.03)	(0.26)	
200 to 250 km							-2.36***	-0.94***	
							(0.90)	(0.26)	
250 to 300 km							-1.77**	-0.53**	
							(0.70)	(0.23)	
300 to 350 km							-1.41***		
							(0.40)		
Additional Controls	41	41	41	42	42	42	42	43	43
N	346	379	390	479	567	600	503	898	1,120
Adjusted R^2	0.459	0.375	0.286	0.716	0.664	0.342	0.810	0.748	0.572
Incremental R^2	0.004	0.004	0.017	0.004	0.009	0.000	0.007	0.008	0.001

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. East region includes the states that were originally part of the thirteen colonies, excluding their western portions (Kentucky, Tennessee, Alabama, and Mississippi) and including Vermont. Middle region includes all other states east of the Mississippi River. West region includes all states west of the Mississippi River. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. The number of additional control variables varies between 41 and 43 depending on region and time period. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4 reports our findings for the time period 1860-1920. Although, as expected, the magnitudes of the correlations are stronger in the more newly settled portions of the U.S. (the West) compared to portions that were settled earlier (the East), we continue to observe evidence of urban shadows in all regions of the country. Consistent with this variation across regions, the switch from the negative to the positive regime occurs 20 years earlier in the East. Thereafter, the sign and magnitude of coefficients are similar to those for the nation.

A.3 Focusing on the East Coast of the U.S.

In this subsection we explore the local growth dynamics in the East. Focusing on the East has the advantage that the size of counties has not changed much over time. Tables A.5 and A.6 report how predictive a large neighbor (95th percentile and above) is for local growth. The first table sets the maximum distance as in the paper, requiring that at least 50 percent of the observations are in the excluded category, whereas the second table keeps the maximum distance constant across time periods. When comparing with Table 1 in the main text, the results for the East are similar to the ones for the U.S. as a whole. However, there are two differences worth pointing out. First, the growth shadows in the early time period are weaker, and second, the growth benefits from urban access in the later time period start two decades earlier.

Table A.5: East Region: Population Growth and the Presence of a Moderately Large Neighbor

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	1840-1860	1860-1880	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000	2000-2017
1 to 50 km	0.29	-0.28	-0.31***	0.50**	0.24**	0.78***	1.15***	0.34**	0.31
	(0.21)	(0.31)	(0.10)	(0.20)	(0.11)	(0.23)	(0.18)	(0.14)	(0.25)
50 to 100 km	-0.29*	-0.21							0.02
	(0.16)	(0.26)							(0.16)
Additional Controls	41	41	41	41	41	41	41	41	41
N	235	346	379	390	423	431	407	355	307
Adjusted R^2	0.465	0.459	0.375	0.286	0.339	0.360	0.310	0.439	0.269
Incremental R^2	0.014	0.004	0.004	0.017	0.005	0.035	0.093	0.005	0.002

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. East region includes the states that were originally part of the thirteen colonies, excluding their western portions (Kentucky, Tennessee, Alabama, and Mississippi) and including Vermont. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

What might explain these differences? The statistically weaker urban shadows could partly be due to the smaller number of observations. They might also be related to East Coast cities being closer to their long-run steady state. In our framework, declining commuting costs make large cities more attractive, prompting them to absorb population from their hinterlands. The further those cities are from their steady-state population, the greater that absorptive capacity might be. While this paper does not focus on transitions, locations in the West were further away from their steady than locations in the East, a point emphasized by Desmet and Rappaport (2017). The earlier emergence of benefits from urban access may be due to East Coast counties being on average smaller than counties in the rest of the country. Our empirical results have already shown that the switch from the negative regime (lower growth in closeby places) to the positive regime (higher growth in closeby places) happens earlier for places "very close" to a large location than

Table A.6: East Region: Population Growth and the Presence of a Moderately Large Neighbor (using the same distance bins for all time periods)

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	1840-1860	1860-1880	1880-1900	1900-1920	1920-1940	1940-1960	1960-1980	1980-2000	2000-2017
1 to 50 km	0.29	-0.28	-0.18	0.67**	0.21*	0.84***	1.14***	0.39**	0.31
	(0.21)	(0.31)	(0.23)	(0.28)	(0.12)	(0.24)	(0.18)	(0.15)	(0.25)
50 to 100 km	-0.29*	-0.21	0.16	0.22	-0.03	0.08	-0.02	0.06	0.02
	(0.16)	(0.26)	(0.25)	(0.14)	(0.13)	(0.09)	(0.21)	(0.10)	(0.16)
Additional Controls	41	41	41	41	41	41	41	41	41
N	235	346	379	390	423	431	407	355	307
Adjusted R^2	0.465	0.459	0.375	0.288	0.337	0.358	0.308	0.438	0.269
Incremental R^2	0.014	0.004	0.006	0.020	0.005	0.035	0.093	0.005	0.002

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above the 95th percentile is within the enumerated distance bin. Specification is identical to that of Table A.5 in the main text except that the same distance bins are used for all time periods. East region includes the states that were originally part of the thirteen colonies, excluding their western portions (Kentucky, Tennessee, Alabama, and Mississippi) and including Vermont. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

for places slightly further away. Because counties in the East Coast tend to be smaller, there are more locations that are "very close", so that this shift becomes visible earlier.

A.4 Relative Size of Locations

In the main text of the paper we showed that urban shadows and urban access tend to be stronger if the bigger is the nearby large neighbor. Here we also explore the importance of the relative size, but rather than varying the size of the large neighbor, we vary the size of the location itself. We focus on two representative time periods, one for the positive regime, 1880 to 1900, and one for the negative regime, 1960 to 1980.

For each of the first four quintiles of locations, Table A.7 separately reports regressions of growth from 1880 to 1900 on the presence of a moderately large neighbor. All coefficients are estimated to be negative, with the magnitudes being lower for higher quintiles. When we consider growth of locations between the 80th and the 90th percentile, the magnitude of the correlation becomes even smaller, though it continues to be negative. Regressions for the other periods during the negative regime show similar results, both for the presence of moderately large and very large neighbors.[1] Table A.8 reports analogous regressions of growth from 1960 to 1980. All coefficients are estimated to be positive, with in this case too some tendency, albeit weaker, for the magnitudes

[1] The one exception concerns regressions of growth from 1880 to 1900 on the presence of very large neighbors. In contrast to the combined regression, estimated coefficients using the first quintile of observations are positive.

Table A.7: Population Growth by Location Size, 1880 to 1900.

	Average Annual Population Growth of Locations by Size, 1880 to 1900						
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	(1) quints 1 to 4	(2) quint 1	(3) quint 2	(4) quint 3	(5) quint 4	(6) decile 9	(7) decile 10
1 to 50 km	-1.10***	-4.78***	-1.17***	-0.35	-0.47***	-0.25*	0.07
	(0.21)	(1.46)	(0.26)	(0.25)	(0.11)	(0.13)	(0.16)
50 to 100 km	-0.86***	-1.13	-0.79**	-0.40***	-0.23**		
	(0.18)	(1.07)	(0.31)	(0.12)	(0.10)		
100 to 150 km	-0.63***	-1.84*	-0.39**				
	(0.15)	(1.00)	(0.19)				
150 to 200 km		-0.91					
		(0.97)					
200 to 250 km		-0.59					
		(0.96)					
250 to 300 km		-0.09					
		(0.96)					
300 to 350 km		-0.57					
		(1.01)					
Additional Controls	52	52	50	51	52	48	50
N	1,844	422	476	472	474	237	237
Adjusted R^2	0.718	0.721	0.441	0.403	0.451	0.436	0.469
Incremental R^2	0.008	0.008	0.016	0.013	0.011	0.004	0.000

Each column presents the results from a regression of average annual population growth of locations in enumerated population rangebetween 1880 and 1900 on categorical indicators if the nearest neighbor with population at or above the 95th percentile of the national distribution is within the enumerated distance bin. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

to decline with a location's size. These findings confirm the results of Table 4 in the main text: the bigger the size difference between the large neighbor and the location, the greater the growth impact of the large neighbor.

Table A.8: Population Growth by Location Size, 1960 to 1980.

	Average Annual Population Growth of Locations by Size, 1960 to 1980						
Distance Nearest Neighbor with Pop $\geq$ 95th Pctile	(1) quints 1 to 4	(2) quint 1	(3) quint 2	(4) quint 3	(5) quint 4	(6) decile 9	(7) decile 10
1 to 50 km	1.19***	1.69**	1.34***	1.03***	1.16***	0.27*	0.12
	(0.19)	(0.83)	(0.34)	(0.26)	(0.19)	(0.15)	(0.11)
50 to 100 km	0.21*	1.12**	0.15	0.18			
	(0.11)	(0.47)	(0.13)	(0.12)			
100 to 150 km		0.56*					
		(0.31)					
150 to 200 km		0.60*					
		(0.31)					
Additional Controls	52	51	50	51	52	51	52
N	2,283	571	570	571	571	285	285
Adjusted R^2	0.376	0.470	0.397	0.404	0.349	0.298	0.566
Incremental R^2	0.033	0.022	0.032	0.034	0.086	0.005	0.001

Each column presents the results from a regression of average annual population growth of locations in enumerated population range between 1960 and 1980 on categorical indicators if the nearest neighbor with population at or above the 95th percentile of the national distribution is within the enumerated distance bin. All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.5 Streetcars

Specification. In the main text we analyzed how the introduction of streetcars in the neighboring large location might affect the strength of urban shadows. Here, instead of considering the introduction of streetcars, we focus on the presence of streetcars. That is, we analyze the effect of the *level* of streetcars, rather than the effect of the *change* in streetcars. To do so, we take the baseline specification and introduce an interaction term between the presence of a large neighbor and that large neighbor having a streetcar. We also control for whether the own county has a streetcar or not. This yields the following specification:

$$g_\ell = \mathbf{I}_\ell^{\tilde{L}}\,\beta + \mathbf{I}_\ell^{\tilde{L}} \cdot \mathbf{S}_\ell^{\tilde{L}}\,\alpha + s_\ell\theta + \mathbf{L}_\ell\,\gamma + \mathbf{x}_\ell\,\delta + \varepsilon_\ell, \tag{A.1}$$

where $\mathbf{S}_\ell^{\tilde{L}}$ are indicator variables that take a value of one if the corresponding large locations in $\mathbf{I}_\ell^{\tilde{L}}$ have a streetcar and s_ℓ is an indicator variable that measures whether ℓ has a streetcar or not. In addition to the presence of streetcars, we are also interested in the length of streetcars. We therefore consider an alternative, where $\mathbf{S}_\ell^{\tilde{L}}$ measures the miles of streetcars of the corresponding large locations in $\mathbf{I}_\ell^{\tilde{L}}$ and s_ℓ measures the miles of streetcars in ℓ. As in the main text, when measuring miles, we consider the log of one plus the mileage. Since we have two years of data on streetcars, we will run (A.1) for 1900-1920 using the streetcar data of 1902 and for 1910-1920 using the streetcar data of 1907.

Results on streetcars. Table A.9 reports our findings from running specification (A.1). Column (1) analyzes how the presence of streetcars in 1902 affects growth between 1900 and 1920. The predicted population growth of small and medium locations was slower if they had a moderately large neighbor with a streetcar, and it was larger if they themselves had a streetcar, although the former correlation is not statistically significant. Column (2) considers the length of streetcars, rather than its mere presence, and reports similar results. The lack of statistical significance of the coefficient on the presence of streetcars in the large neighbor might partly be due to the fact that almost all large locations already had streetcars in 1902, so that the differential effect of having a streetcar is hard to identify.

Table A.9: Population Growth and Large Neighbors, 1900 to 1920: Streetcars.

	Average Annual Population Growth of Small and Medium Locations (Quintiles 1-4)							
	Neighbor with Pop $\geq$ 95th Pctile 1900-1920		Neighbor with Pop $\geq$ 90th Pctile 1900-1920		Neighbor with Pop $\geq$ 95th Pctile 1910-1920		Neighbor with Pop $\geq$ 90th Pctile 1910-1920	
Distance Nearest Neighbor	(1) Streetcar Presence	(2) Streetcar Log Miles	(3) Streetcar Presence	(4) Streetcar Log Miles	(5) Streetcar Presence	(6) Streetcar Log Miles	(7) Streetcar Presence	(8) Streetcar Log Miles
1 to 50 km	0.02	0.05	0.11	-0.14	0.15	-0.02	0.86^{*}	0.32^{*}
	(0.44)	(0.37)	(0.29)	(0.21)	(0.32)	(0.28)	(0.45)	(0.19)
Streetcar 1-50 km	-0.17	-0.05	-0.45^{*}	-0.05	-0.20	-0.01	-0.84^{*}	-0.07
	(0.43)	(0.08)	(0.26)	(0.05)	(0.31)	(0.06)	(0.43)	(0.05)
50 to 100 km	-0.26	-0.14	-0.32	-0.24	0.20	0.26	0.29	0.30
	(0.21)	(0.16)	(0.21)	(0.19)	(0.23)	(0.19)	(0.41)	(0.21)
Streetcar 50-100 km	-0.02	-0.04	-0.12	-0.06	-0.41^{*}	-0.11^{***}	-0.44	-0.12^{**}
	(0.22)	(0.04)	(0.18)	(0.05)	(0.23)	(0.04)	(0.41)	(0.05)
Own Streetcar	0.69^{***}	0.34^{***}	0.71^{***}	0.35^{***}	0.23^{*}	0.12^{**}	0.20^{*}	0.12^{**}
	(0.14)	(0.06)	(0.15)	(0.06)	(0.12)	(0.06)	(0.12)	(0.06)
Additional Controls	52	52	52	52	52	52	52	52
N	2,110	2,110	2,110	2,110	2,238	2,238	2,238	2,238
Adjusted R^2	0.525	0.525	0.528	0.528	0.154	0.156	0.155	0.156

Each column presents the results from a regression of average annual population growth of those with population at or below the 80th percentile over the enumerated period on categorical indicators if the nearest neighbor with population at or above either the 95th or the 90th percentile is within the enumerated distance bin. The regressions include an interaction of these categorical indicators with either streetcar presence or streetcar miles (expressed as log(1+mileage of streetcars). All regressions include a constant and control for initial population and additional geographic, weather, and topographical control variables, as described in the text. Columns 1-4 are based on streetcar data of 1902 and have annualized growth between 1900 and 1920 as dependent variable, Columns 5-8 are based on streetcar data of 1907 and have annualized growth between 1910 and 1920 as dependent variable. Standard errors, in parentheses, are robust to spatial correlation based on Conley (1999). The incremental R^2 refers to the difference between the R^2 of the regression and the R^2 of a regression on only the additional control variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To partly address this issue, columns (3) and (4) lower the threshold of what we consider a large neighbor to locations with a population at the 90th percentile or above. The results are similar, though the coefficient on the presence of streetcars in a large neighbor at a distance up to 50km is now statistically significant at the 10% level. Columns (5) through (8) are analogous to columns (1) through (4), with two differences: the dependent variable is growth between 1910 and 1920 and the streetcar data are for 1907. The results are similar to the ones for the period 1900-1920 in the sense that the coefficients on the presence or length of streetcars in the large neighbors are always negative. One difference is that more coefficients are now statistically significant at either the 5% or 10% level. Also as before, locations that themselves have a streetcar are predicted to grow faster.

Table A.9 suggests that streetcars strengthen the urban shadows of large locations. We reach this conclusion by observing that in all but one case the coefficients on the presence (or length) of streetcars in neighboring large locations are negative. This conclusion is suggestive in the sense that the negative coefficients are often not statistically significant. Needless to say, our results cannot be interpreted as causal. While it could of course be that better commuting infrastructure in the neighboring large city attracts population from neighboring locations, it is also possible that the absorption of closeby population by the large city prompts it to build better commuting infrastructure. Overall, these findings are consistent with Stylized Fact 5 in the main text.