ONLINE APPENDIX

The Economics of Speed: The Electrification of the Streetcar System and the Decline of Mom-and-Pop Stores in Boston, 1885-1905

Wei You†

September 13, 2020

1 Supplemental Data

1.1 The Dun & Bradstreet Reference Book of American Business

The need for credit ratings stems from the first half of the nineteenth century, when commission merchants based in large urban cities were increasingly providing goods and supplies to rural merchants, jobbers, and general stores, but were unable to discriminate their credit-worthiness. Credit rating agencies established a network of local correspondents, who gathered business information on merchants and jobbers in their areas and reported it to the rating agency's headquarters.[1] The agency then sold this credit information to subscribers for a fee. R.G. Dun & Co was one of the most successful credit rating agencies of the era, and it merged with the company J.M. Bradstreet in 1933 to form the Dun & Bradstreet Corporation.

R.G. Dun & Co's reference books cover a wide range of businesses in the United States and Canada, containing their names, main product lines, pecuniary strengths (i.e., estimated net worth, grouped into 17 size categories), and credit ratings (8 classes).[2] These books were published bimonthly, and most of the issues are found in the Library of Congress, with the exceptions of those published between 1889 and 1898. I digitized the Boston sections of these books for September 1885 and July 1899.

†Peking University, The Institute of New Structural Economics, 5th Yiheyuan Road, Haidian District, Beijing, 100871, China. Email: weiyou@nsd.pku.edu.cn.

[1]Initially, there were no direct employees of these firms, but instead, the firms often used lawyers or postmasters who lived in the particular area. Later, the system relied on paid reporters, who worked exclusively for a particular agency.

[2]For a more detailed discussion of this data source, see Sarada and Ziebarth (2015).

I performed an assessment of the comprehensiveness of the business directory and the credit rating data by matching the two data sources in two directions. First, I randomly selected an 8% sample of firms from the credit rating reference books in 1885 and 1899, totaling 1,935 firms. I then manually matched these firms to those in the first main section of the *Boston Directories* in the corresponding years by both their names and products (occupations). I was able to match 1,736 of the total 1,935 firms, yielding a matching rate of 89.7%. Next, I randomly selected 826 firms from the business section of the *Boston Directories* in 1885 and 1899, and then matched them to the credit rating reference books, also by their names and products (occupations). 287 of these 826 firms could be matched, yielding a matching rate of 34.8%.[3] The much higher matching rate in the first direction suggests that the *Boston Directories* contain a more comprehensive list of firms, while the credit rating books probably selected businesses that catered to the needs of their subscribers. For this reason, as well as for the fact that the credit rating reference books are missing for the years 1889-1898, I used the *Boston Directories* as the main data source and drew on the credit rating records as a supplemental data source.

To document changes in aveage pecuniary strength and credit rating of businesses by geographic area, I also geocoded the firms matched between the Boston Directories and the credit rating books. I further restricted the sample to retail/wholesale food firms, resulting in a firm-level dataset of 813 observations. This data sample is used in the analysis in subsection 5.5 of the main text.

1.2 Linking Census Microdata to the Boston Directories

This subsection describes the construction of the individual-level panel data. The purpose of linking individuals over time is to understand the income sorting patterns following the streetcar electrification. As indicated in footnote 18 in subsection 3.1 of the main text, the relevant "size" of the market is the total dollar-weighted number of consumers. In Table 2 in the main text, I showed how the spatial distribution of the population changed in this period. In this subsection, I show how average incomes changed across different geographic areas in the city, exploiting the occupational income/earnings information in the Census microdata. [4]

[3] The matching rates in both directions are positively correlated with firm size. In the first direction, 100% of the firms with the highest net worth class in the credit rating reference books were matched to the *Boston Directories*, and this matching rate fell down to 76% for the lowest net worth class. In the second direction, 27.7% of the sole proprietorships in the *Boston Directories* were matched to the credit rating reference books, while this matching rate was 46.6% for the other legal forms.

[4] The 1900 occupational income/earnings scores are calculated by Integrated Public Use Microdata Series (IPUMS) according to the 1900 occupation data and the 1950 income data. Abramitsky et al. (2014) show that this measure is a reasonable proxy for the personal income in late-19th and early-20th century America.

To do so, I obtained a random sample 1,450 individuals (2%) in Boston in the 1900 census data. I restricted my attention to male household heads between the ages of 30 and 50 in 1900, an age range in which men are both young enough to be employed in 1900 and old enough to be in the workforce in 1887. I then manually matched between the census and the 1900 Boston Directory by individual name, occupation, and residence. The matching rate is 65%. Then, from these matched 942 individuals in 1900, I traced them back to the 1887 Boston Directory, obtaining their residence and occupation in 1887. I chose 1887 as the beginning year because at that time, the electrification of the streetcar system was almost unpredictable, and therefore we can avoid the possibility that people sorted in anticipation of this event. I chose to match individuals over time in the Boston Directories instead of in the censuses because: First, the 1890 census data was not available, while the Boston Directories are available annually. Second, the addresses in the Boston Directories can be used to precisely geolocate each individual, while the censuses do not have such detailed geographic information.

I matched individuals over time in the Boston Directories by name, occupation/product, and/or residence. Specifically, if two entries from two years have the same name, and the same occupation and/or the same residence, and this match is unique, I treated them as a match. If two entries from two years have the same name, but neither their occupations nor their residences match, I treated those two entries as unmatched, although there is a positive probability that the person changed both his/her occupation and residence. Such cases account for 5.8% of all the cases. Finally, if there are multiple matches such that more than two individuals have the same name and occupation or the same name and address. I classified them as unmatched due to the lack of information. Such cases account for 3.3% of all the cases. The 13-year matching rate using this method in the Boston Directories is 50%.[5] I chose 1887 as the beginning year, since the electrification of the streetcar system would have been highly unpredictable in that year, and thus, we can avoid the possibility that people might have sorted in anticipation of this future event. The final product is a panel data of 471 individuals who lived in Boston both in 1887 and 1900.

Using this panel dataset, I document the spatial sorting of income in Table 1 in this appendix. The proportional changes in income scores ranged between -7.6% and 2.0% across different distances to the rails and to the city center. Compared to Table 2 in the main text, which shows that the employment or the number of residents changed by between 35% and 113.8%, the changes in income scores were much smaller. Therefore, effective market size in our context can be approximated by population only.

[5] In comparison, in studies that use decennial census records to match individuals over time, the matching rate is typically below 30% over 20 years.

Table 1: Income Sorting

Geography	Occupational income score 1887	Occupational income score 1900	Occupational income score Growth	Occupational earnings score 1887	Occupational earnings score 1900	Occupational earnings score Growth
By distance to rails						
0-25m	32.6	32.0	-1.9%	65.7	63.9	-2.8%
25-75m	27.9	28.3	1.4%	60.6	60.4	-0.5%
>75m	28.5	28.8	1.2%	61.2	62.5	2.0%
By distance to city center						
Periphery	29.5	29.6	0.3%	62.8	62.8	0.0%
CBD	28.9	27.1	-6.2%	58.3	53.8	-7.6%

Notes: The sample is the matched 471 individuals between the Boston Directories and the Census microdata who showed up in both the 1887 and 1900 Boston Directories. Each observation is an individual. I calculate the distance to the streetcar rails and the distance to Boston City Hall using each individual's residence. The 1900 occupational income/earnings scores are calculated by Integrated Public Use Microdata Series (IPUMS) according to the 1900 occupation data and the 1950 income data.

1.3 Supplemental Tables

Table 2: The 25 Most Frequent Retail/Wholesale Products

Product	% of est.	Lbs./day	Product	% of est.	Lbs./day
Food related					
Grocers	16.1%	1.18	Fruits	2.8%	1.12
Liquors	10.0%	1.15	Produce	2.7%	4.45
Provisions	7.8%	1.10	Fish	2.6%	0.20
Restaurants	6.4%	N/A	Confectioners	1.4%	0.50
Bakers	4.0%	1.87	Milk	0.7%	2.55
Clothing related					
Boots & Shoes	8.6%		Clothing	2.7%	
Tailors	8.3%		Men's Furnishings	1.1%	
Dry Goods	3.4%		Hats, Caps, & Furs	0.8%	
Milliners	2.9%				
Others					
Apothecaries & Drugs	3.7%		Furniture	2.0%	
Cigar & Tabacco	3.0%		Jewelry	1.7%	
Books & Publishers	2.7%		Hardware	1.3%	
Leather	2.5%		Piano	0.6%	

Notes: These 25 products are taken from the 25 most common retail/wholesale products in the Boston Directories between 1880 and 1905. The total number of establishments is 47,879. I match the food related products to the product categories in the 1918-1919 Consumer Expenditure Survey from the US Bureau of Labor Statistics. Additionally, "% of est." is defined as the share in total number of establishments, and "Lbs./day" is defined as the weight (pounds) of food consumed by a family of five per day in 1918/1919. Restaurants are distinct from other food-related products, as consumption is primarily done on-site. When analyzing the results by product, I classify restaurants as nonfood products. I also test the robustness of the benchmark results by both including and excluding restaurants and report the results in Table 4.

Table 3: Robustness Checks

Dependent Variable:	(1)	(2)	(3)	(4)	(5)	(6)
Share of S.P. in total est.	300m block		500m block		1901 rails	
			Food Products			
<25m of the rails	-0.011	-0.017	-0.015	-0.022	0.037	0.027
	(0.032)	(0.038)	(0.033)	(0.034)	(0.029)	(0.033)
25-75m of the rails	0.044	0.031	0.030	0.004	0.019	0.011
	(0.034)	(0.042)	(0.025)	(0.026)	(0.024)	(0.031)
Post electrification	-0.030		-0.039		0.009	
	(0.033)		(0.034)		(0.033)	
<25m of the rails*Post	-0.134	-0.123	-0.156	-0.141	-0.214	-0.195
	(0.040)	(0.045)	(0.037)	(0.046)	(0.037)	(0.051)
25-75m of the rails*Post	-0.064	-0.039	-0.043	0.008	-0.065	-0.048
	(0.040)	(0.044)	(0.050)	(0.062)	(0.044)	(0.057)
CBD*Post	-0.126		-0.085		-0.124	
	(0.045)		(0.037)		(0.035)	
400m-Block FE	YES		YES		YES	
400m-Block*Year FE		YES		YES		YES
Observations	384	384	300	300	330	330
R-squared	0.759	0.830	0.864	0.900	0.846	0.887
			Nonfood Products			
<25m of the rails	0.021	0.024	0.036	0.039	-0.032	-0.018
	(0.073)	(0.084)	(0.055)	(0.057)	(0.055)	(0.064)
25-75m of the rails	0.074	0.071	0.089	0.085	0.025	0.035
	(0.056)	(0.066)	(0.041)	(0.039)	(0.035)	(0.041)
Post electrification	-0.060		-0.023		-0.019	
	(0.072)		(0.075)		(0.079)	
<25m of the rails*Post	0.018	0.012	-0.023	-0.029	0.000	-0.028
	(0.079)	(0.106)	(0.075)	(0.074)	(0.080)	(0.078)
25-75m of the rails*Post	0.000	0.007	-0.035	-0.028	0.012	-0.008
	(0.080)	(0.105)	(0.057)	(0.065)	(0.071)	(0.076)
CBD*Post	-0.003		-0.006		-0.045	
	(0.029)		(0.048)		(0.044)	
400m-Block FE	YES		YES		YES	
400m-Block*Year FE		YES		YES		YES
Observations	162	162	114	114	114	114
R-squared	0.760	0.794	0.845	0.903	0.871	0.914

Notes: Each observation is a location. For all specifications, the outcome variable is the share of sole proprietorship (S.P.) establishments of the location. Each location is weighted by the location's number of establishments in 1885. The regressions use only the 1885 and 1905 data. *Post* is a dummy for the post-electrification period, i.e., the year 1905. The treatment and control locations and the blocks are defined in Figure 3. In columns (1) and (2), the block sizes are 300m*300m. In columns (3) and (4), the block sizes are 500m*500m. In columns (5) and (6), I define the treatment bands based on the distance to the 1901 streetcar rails. The standard errors, clustered by block, are reported in parentheses.

Table 4: Classifying Restaurants as Food

Dependent Variable:	(1)	(2)	(3)	(4)
Share of S.P. in total est.	Food Products		Nonfood Products	
<25m of the rails	0.009	0.013	-0.026	-0.001
	(0.025)	(0.031)	(0.080)	(0.085)
25-75m of the rails	0.018	0.018	0.039	0.049
	(0.023)	(0.029)	(0.046)	(0.052)
Post electrification	-0.025		-0.059	
	(0.030)		(0.084)	
<25m of the rails*Post	-0.143	-0.150	0.004	-0.045
	(0.028)	(0.042)	(0.091)	(0.089)
25-75m of the rails*Post	-0.040	-0.039	0.008	-0.013
	(0.037)	(0.048)	(0.077)	(0.085)
CBD*Post	-0.081		-0.004	
	(0.030)		(0.034)	
400m-Block FE	YES		YES	
400m-Block*Year FE		YES		YES
Observations	348	348	120	120
R-squared	0.855	0.891	0.850	0.887

Notes: Each observation is a location. For all specifications, the outcome variable is the share of sole proprietorship (S.P.) establishments of the location. Each location is weighted by the location's number of establishments in 1885. The regressions use only the 1885 and 1905 data. Additionally, *Post* is a dummy for the post-electrification period, i.e., the year 1905. The treatment and control locations and the blocks are defined in Figure 3. In columns (1) (2) and (3) (4), I calculate all the statistics using food and nonfood products (establishments), respectively. The standard errors, clustered by block, are reported in parentheses. Compared to Table 3 in the main text, I classify restaurants as food establishments here.

Table 5: Conley Standard Errors at Varying Assumed Distance Cutoffs

Dependent Variable: Share of S.P. in total est.	(1) Cluster	(2) 1000m	(3) 500m	(4) 250m	(5) Cluster	(6) 1000m	(7) 500m	(8) 250m
				Food Products				
<25m of the rails	-0.001	-0.001	-0.001	-0.001	-0.005	-0.005	-0.005	-0.005
	(0.028)	(0.025)	(0.026)	(0.024)	(0.032)	(0.026)	(0.026)	(0.025)
25-75m of the rails	0.020	0.020	0.020	0.020	0.012	0.012	0.012	0.012
	(0.023)	(0.019)	(0.021)	(0.024)	(0.028)	(0.019)	(0.022)	(0.025)
Post electrification	-0.021	-0.021	-0.021	-0.021				
	(0.034)	(0.020)	(0.023)	(0.023)				
<25m of the rails*Post	-0.180	-0.180	-0.180	-0.180	-0.172	-0.172	-0.172	-0.172
	(0.036)	(0.027)	(0.030)	(0.028)	(0.050)	(0.037)	(0.038)	(0.035)
25-75m of the rails*Post	-0.058	-0.058	-0.058	-0.058	-0.042	-0.042	-0.042	-0.042
	(0.038)	(0.022)	(0.027)	(0.029)	(0.054)	(0.030)	(0.035)	(0.038)
CBD*Post	-0.105	-0.105	-0.105	-0.105				
	(0.036)	(0.022)	(0.025)	(0.024)				
400m-Block FE	YES	YES	YES	YES				
400m-Block*Year FE					YES	YES	YES	YES
Observations	330	330	330	330	330	330	330	330
R-squared	0.850	0.979	0.979	0.979	0.895	0.986	0.986	0.986
				Nonfood Products				
<25m of the rails	-0.018	-0.018	-0.018	-0.018	0.005	0.005	0.005	0.005
	(0.064)	(0.037)	(0.055)	(0.053)	(0.071)	(0.037)	(0.056)	(0.053)
25-75m of the rails	0.037	0.037	0.037	0.037	0.048	0.048	0.048	0.048
	(0.037)	(0.027)	(0.037)	(0.039)	(0.043)	(0.027)	(0.036)	(0.038)
Post electrification	-0.027	-0.027	-0.027	-0.027				
	(0.065)	(0.042)	(0.055)	(0.056)				
<25m of the rails*Post	-0.024	-0.024	-0.024	-0.024	-0.070	-0.070	-0.070	-0.070
	(0.068)	(0.047)	(0.062)	(0.061)	(0.069)	(0.054)	(0.073)	(0.068)
25-75m of the rails*Post	-0.006	-0.006	-0.006	-0.006	-0.029	-0.029	-0.029	-0.029
	(0.061)	(0.032)	(0.045)	(0.052)	(0.069)	(0.035)	(0.050)	(0.054)
CBD*Post	-0.017	-0.017	-0.017	-0.017				
	(0.034)	(0.018)	(0.023)	(0.025)				
400m-Block FE	YES	YES	YES	YES				
400m-Block*Year FE					YES	YES	YES	YES
Observations	144	144	144	144	144	144	144	144
R-squared	0.862	0.982	0.982	0.982	0.903	0.987	0.987	0.987

Notes: Columns (1) and (5) correspond exactly to Table 3 in the main text, which serve as basis of comparison. The column titles "1000m", "500m", "250m" indicate that standard errors are Conley standard errors (Conley 1999) based on assumed distance cutoffs of 1000m, 500m, 250m, respectively.

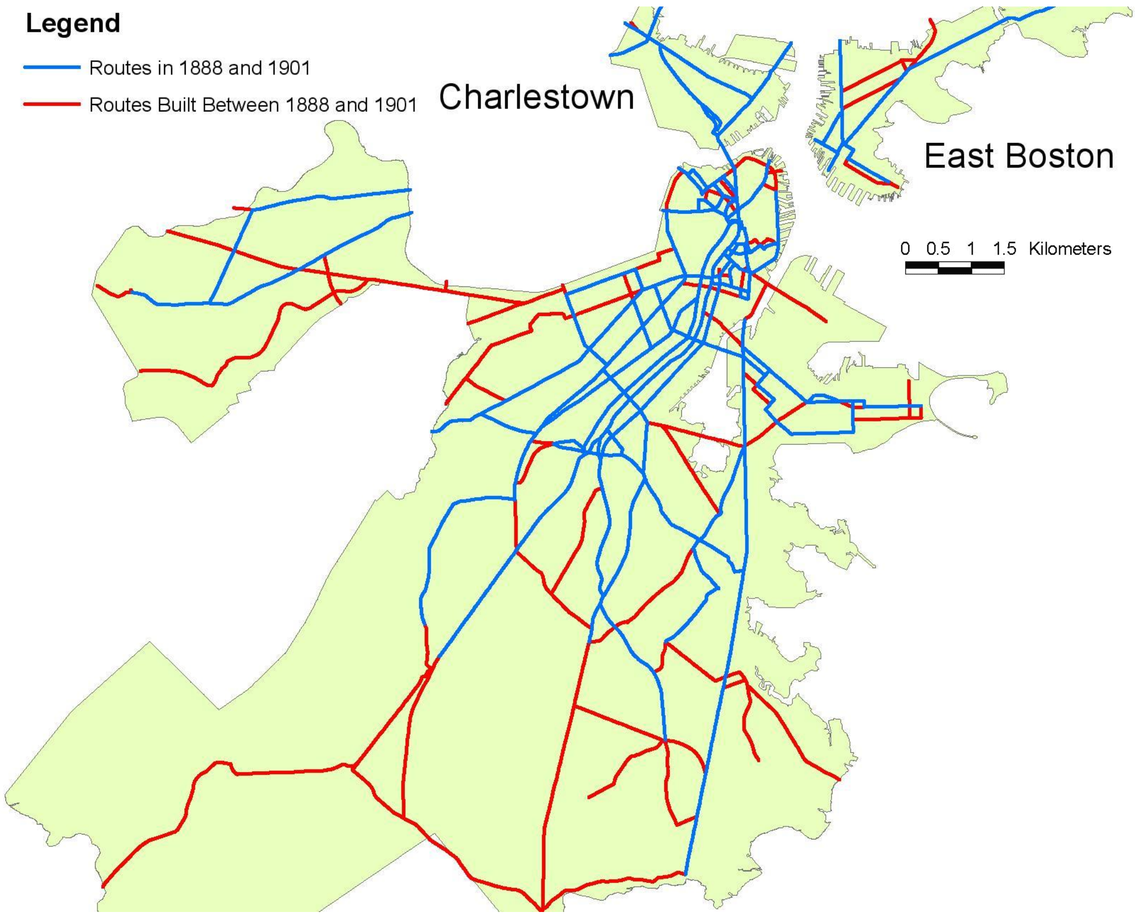
Table 6: Heterogeneity by Purchase Frequency

	(1)	(2)	(3)
Dependent Variable: Share of S.P. in total establishments			
<25m of the rails	-0.001	0.002	-0.002
	(0.028)	(0.028)	(0.029)
25-75m of the rails	0.020	0.022	0.019
	(0.023)	(0.023)	(0.024)
Post electrification	-0.021	-0.023	-0.022
	(0.034)	(0.032)	(0.037)
<25m of the rails*Post	-0.180	0.182	-0.178
	(0.036)	(0.101)	(0.041)
<25m*Post*Purchase frequency		-0.521	
		(0.141)	
25-75m of the rails*Post	-0.058	-0.118	-0.057
	(0.038)	(0.137)	(0.044)
25-75m*Post*Purchase frequency		0.079	
		(0.159)	
CBD*Post	-0.105	-0.103	-0.095
	(0.036)	(0.033)	(0.080)
CBD*Post*Purchase frequency			-0.016
			(0.121)
400m-Block FE	YES	YES	YES
Observations	330	330	330
R-squared	0.850	0.857	0.850

Notes: Column (1) here corresponds exactly to the column (1) in Table 3 in the main text, which serve as basis of comparison. "Purchase frequency" measures the average purchase frequency of the products sold by the establishments in that location. This measure has a mean of 0.78 and standard deviation of 0.19 across locations. The key coefficient is on *<25m*Post*Purchase frequency*. A negative value indicates that the treatment effect is larger when the purchase frequency of the products is higher.

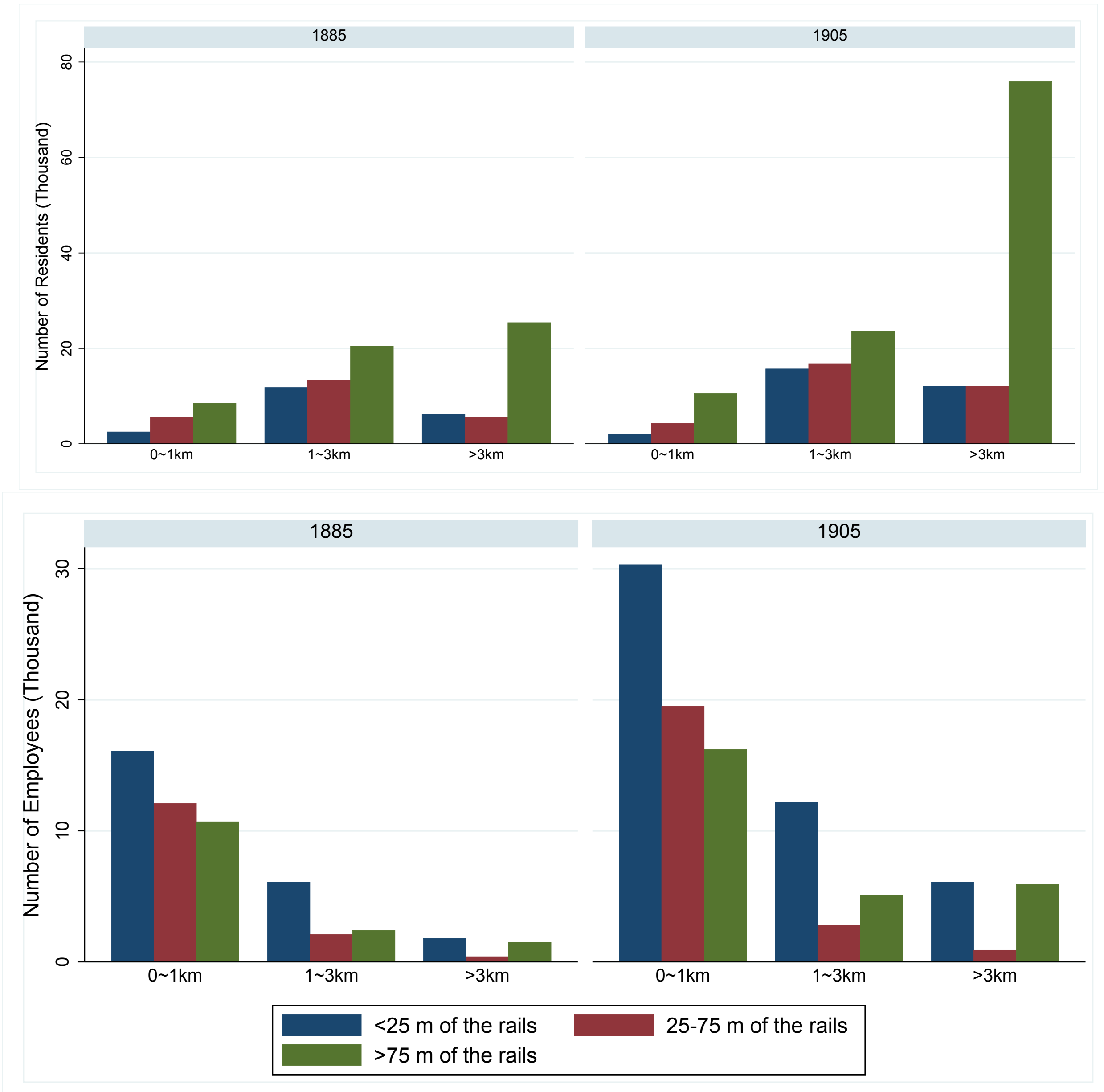
2 Supplementary Figures

Figure 1: The Streetcar Routes in 1888 and 1901



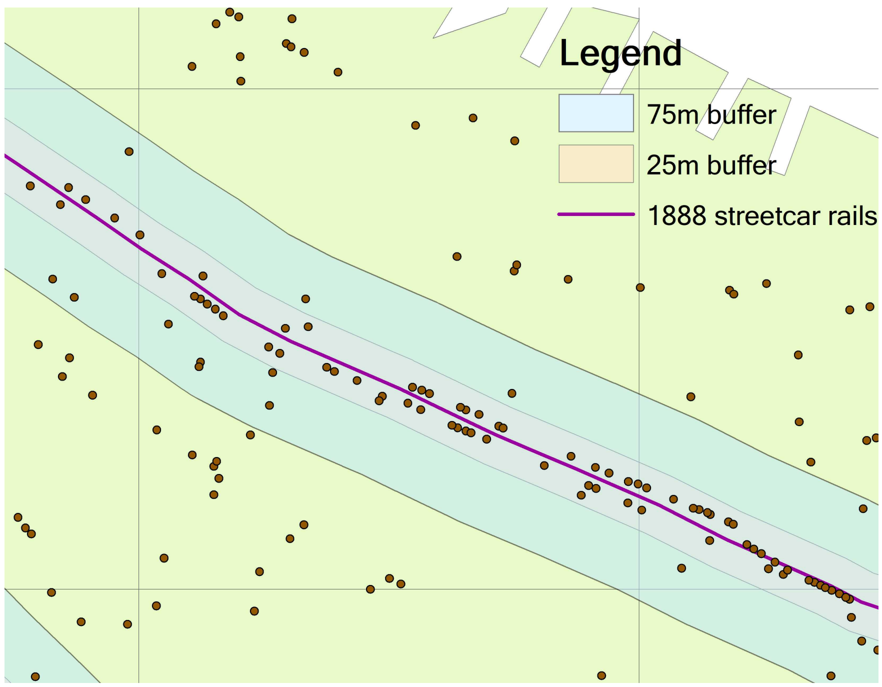
Source: Digitized Boston city maps from the online David Rumsey Historical Map Collection.

Figure 2: Population and Employment Distribution



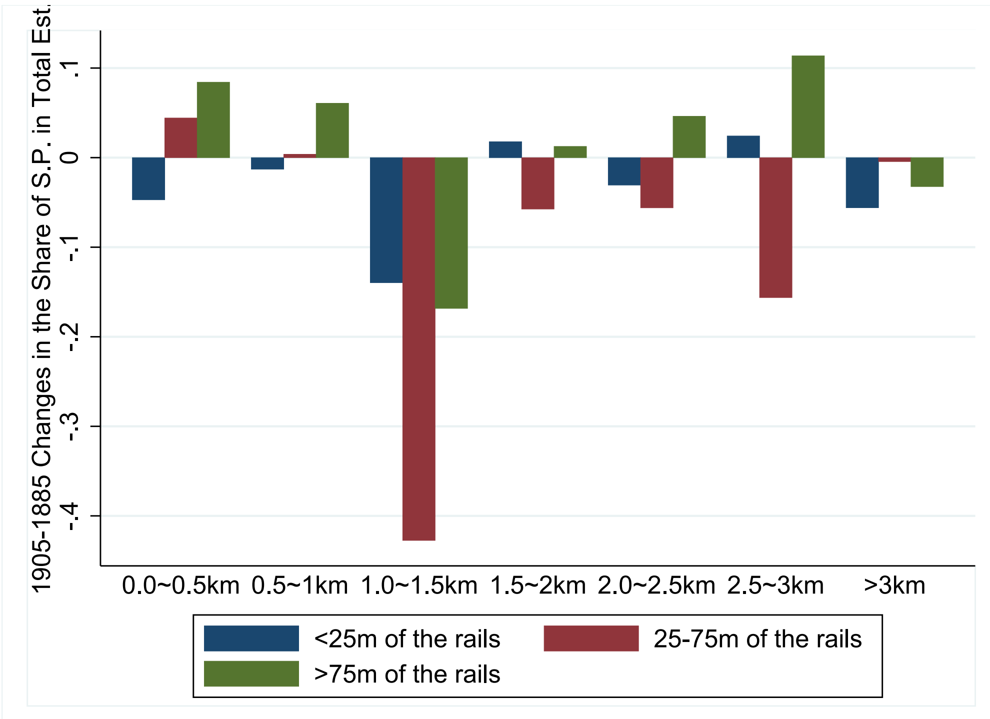
Notes: The horizontal axis represents the distance from the city center. The bars in different colors indicate different distances from the 1888 streetcar rails. The spatial data of population and employment came from the geocoded 1% random sample of the inhabitants in the 1885 and 1905 *Boston Directories*, which contain information on both the place of work and the place of residence.

Figure 3: Illustration of Treatment and Control "Locations", and "Blocks"



Notes: The above figure illustrates the definition of the treatment locations and control locations, as well as the blocks. The grids in this figure are $400m \times 400m$, called blocks. The purple areas are within 25 m of the streetcar lines, which are the most treated locations; the light blue areas are between 25 and 75 m from the streetcar lines, which are the less treated locations; and the remainder of the areas within the block, indicated by the light green areas, are the control locations.

Figure 4: Anomaly in S.P. Trends



Notes: Author's calculation from the *Boston Directories*. The sample includes only nonfood establishments.

Figure 5: Sanborn Fire Insurance Map with Geo-located Points

Plot-maps, such as the one above, are georeferenced to the 1930 Boston street centerline map. The red dots correspond to each building/address. The green line represents a portion of the streetcar line.