Appendix For
Black-Friendly Businesses in Cities During the Civil Rights Era

Last Updated: January 14, 2024

Maggie Jones
Emory University & NBER

Trevon Logan
The Ohio State University & NBER

David Rosé
Wilfrid Laurier University

Lisa Cook
Michigan State & NBER

Contents

A Data Appendix

A1 Additional Details on Geocoding Businesses

To geocode the historical addresses in the Green Books we began with an initial pass that ran all addresses through the U.S. Census Geocoder.[1] This produced either an exact match, a non-exact match, a tie, or no match. Exact matches occur when one unique address is found that matches the input address. About 50 percent of the addresses returned an exact match. Non-exact matches occur if the geocoder was able to locate a similar, but not exact match to the input address. Ties occur when there is more than one address that matches the input address. In some cases, the geocoder is not able to locate the input address. In the case of a tie or "no match," the input addresses were searched by hand.

A second pass was implemented by hand checking each address in the Green Books in Google Maps. If the distance between the Census Geocoder and Google Maps coordinates was less than 1 mile, the address was categorized as an exact match. Addresses that did not meet this accuracy threshold were rechecked for typos and historical context. This included addresses that did not return an initial exact match, which were then matched by hand through Google Maps. In these instances the remaining missing addresses were imputed using a majority rule. For example, if a listing had a street name but no street number and there were other establishments on the same street, then the unnumbered address was geocoded in close proximity to the other businesses (the midpoint of the address range). For addresses that are descriptions, for example, the "Corner 126th St. and 8th Ave" (Braddock Tavern in Harlem, NY) or "Rt. 301, 6 miles N. of Potomac River Bridge" (Blue Star Motel in Faulkner, MD), Google Maps and measurement tools were used to approximate the location. In instances where large construction projects, often the interstate highways, altered the landscape drastically, historical sources were consulted to inform an educated guess of the establishment location.

One important issue to contend with is that the level of detail about an establishment's address can vary significantly across years. In some years it could be missing, in others it could be a location description or an exact street address. Failure to account for this before combining Green Book locations with other data sources can introduce error to any ensuing empirical analyses. For example, we would not know whether an increase in Green Book establishments in a particular area reflects a real increase in Green Book listings or simply variation in address accuracy. To account for this, the most precise address that appears

[1] The Census Geocoder can be found here: `https://www.census.gov/geo/maps-data/data/geocoder.html`

across all publication years is assigned to each establishment.[2]

This requires that each establishment be uniquely identified across years. To generate unique panels with consistent addresses we proceeded in the following manner:

1. Match establishment names across years, matching exactly on state, city, and industry and probabilistically, using the Levenshtein distance, on name.[3]

2. Standardize addresses.[4]

3. Assign the modal establishment name conditional on standardized address.

4. Conduct a final check of the assigned modal name against the original establishment name using the Jarowinkler distance. If dissimilarity is high, this is flagged for human review to determine that it is indeed the same establishment.

5. The most accurate address is assigned to all observations within each panel. These cleaned panels are the inputs in the geocoding procedure described in the preceding section.

6. For addresses that had precise information (i.e. a street number or intersection) but an approximate match or a mid-point of the street were assigned, we had a research assistant do another search through sources on street name changes. This yielded some further improvement in our geolocating accuracy.

[2]We did not encounter instances of obvious conflict between a descriptive address and an exact street address.

[3]We supplement this with a human double-check to determine the similarity cut-off and adjudicate potential matches around it.

[4]We standardize abbreviations for road type, direction, capitalization, and punctuation.

Table A.1: Idiosyncratic Greenbook Entries

Example	Year	Establishment	Address	City	Type
Fresno Motel	1957 - 1962	Fresno Motel	Hwy. 99	Fresno, CA	Lodging
	1963	Fresno Hacienda	Hwy. 99 and Clinton	Fresno, CA	Lodging
Summer's Hotel & Resto.	1947-52	Jim Summers	719 S. Main St	Camden, AR	Restaurant
	"	Summers	715 1/2 S. Main Street	Camden, AR	Liquor store
	1953	Summer's Hotel & Court	721 Adams Street	Camden, AR	Lodging
	1954	Summer Hotel	754 1/2 Adams St. S. W.	Camden, AR	Lodging
	1955	Summer Hotel	740 Adams Ave. S.W.	Camden, AR	Lodging
	1956	Summers Hotel & Restaurant	740 Adams Avenue S.W.	Camden, AR	Lodging & Resto.
	1957	Summer's Hotel & Motel	750-754 1/2 Adams Ave.	Camden, AR	Lodging

Table A.2: Geocoding Accuracy by Region

	Midwest	Northeast	South	West	All Regions
Numbered Street	79.89	82.64	75.97	80.71	79.26
Intersection	6.99	7.26	5.46	5.29	6.32
Mid-point of Street	4.28	5.48	11.46	4.30	7.40
Descriptive Address	0.44	0.43	0.32	0.28	0.38
Approx. Location	1.60	0.18	1.33	1.14	1.04
County Centroid	6.79	4.01	5.46	8.31	5.6

Notes: This table reports the percent of establishments that were geocoded according to different methods.

A2 Enumeration District-Census Tract Mapping

The Census Bureau published tables with census tract characteristics from the 1940 census during the mid-1940s. Census tracts were, at the time, the most granular level of geography for which data was published and they were meant to capture the sociodemographic characteristics of neighborhoods. In total, the Census Bureau published information about census tracts in 60 cities in 1940. These were digitized by a team led by Elizabeth Brogue and later processed and distributed by the NHGIS alongside GIS map files of the 1940 census tracts (Manson et al., 2020).[5]

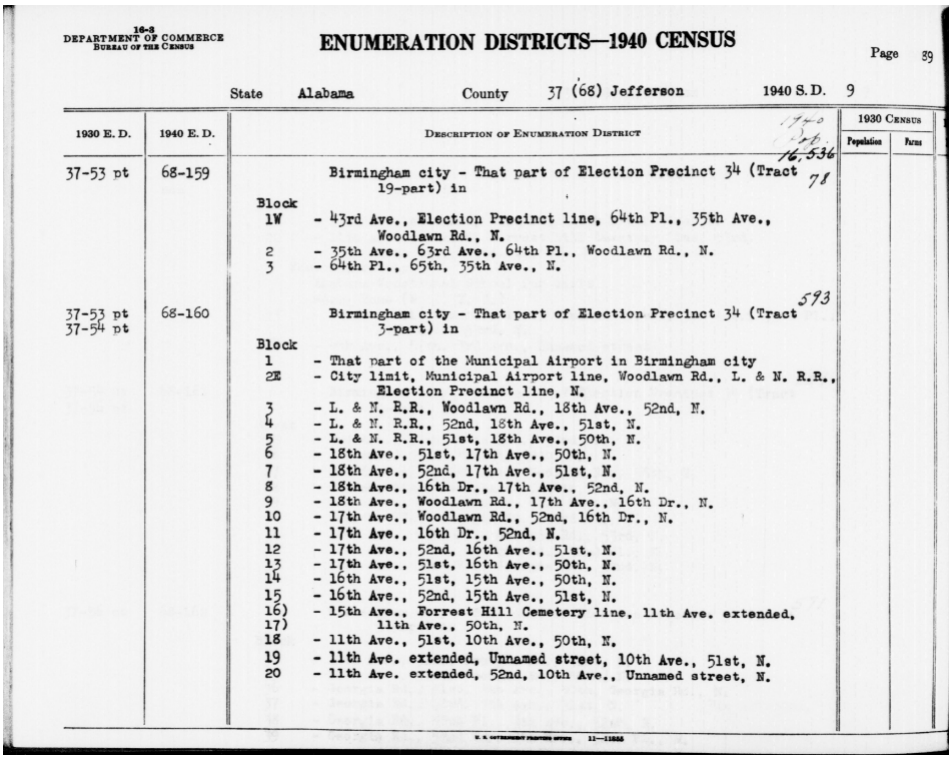
The socioeconomic information that can be gleaned about census tracts from the published tables is valuable but fairly limited for our purposes. In particular, there is very little information about characteristics along racial lines (with the notable exceptions of population counts and variables related to the number of dwellings by occupancy status - tenant or owner) and the economic information as it pertains to labor market participation is limited at the aggregate level as well. We create a crosswalk that allows us to generate census tract characteristics such as the share of the labor force with at least a grade 10 level of education, average income, the labor force participation rate, and more, *by race*.

To do this we created a crosswalk that allowed us to assign individuals in the 1940 full-count census (Ruggles et al., 2020) to census tracts using their enumeration district information. Fortunately, for our purposes, census tracts by their construction were comprised of groups of enumeration districts and these relationships are documented in the U.S. Census Bureau's Enumeration District Descriptions (see Figure A.1 for an example–it documents that Chicago's enumeration district 68-159 was part of Chicago's census tract 19). This raises the possibility of creating a mapping (or crosswalk) from enumeration districts to census tracts, which allows us to link records in the 1940 full-count census data with the NHGIS's spatial census tract data.[6] Our work was significantly aided by a website maintained by Steve Morse and Joel Weintraub (`https://stevemorse.org`) that contains transcriptions of most of the enumeration district descriptions as well as a tool to easily access the associated archival images in cases where consultation was required (in some instances the transcriptions are incomplete so we had to consult the Enumeration District

[5]For NYC, census tract shapefiles are not available through the NHGIS, in their place, maps of health areas (groupings of census tracts) that align with the main tables published by the Census Bureau are included.

[6]Apart from the aforementioned census tracts in New York City, we are unable to create a crosswalk for census tracts 51 through 60 in Oklahoma City because the Enumeration Descriptions are incomplete (they also do not appear in the main map of the enumeration districts in Oklahoma City–a note therein points to a supplementary map, but we were not successful in locating it).

Descriptions to collect the census tract information).[7] Using this cross-walk we then assign a census tract to individuals in the full count that lived in census tract cities.[8] This allows us to aggregate these observations to create a richer set of census tract-level characteristics.



16-3
DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

ENUMERATION DISTRICTS—1940 CENSUS

Page 89

State Alabama County 37 (68) Jefferson 1940 S. D. 9

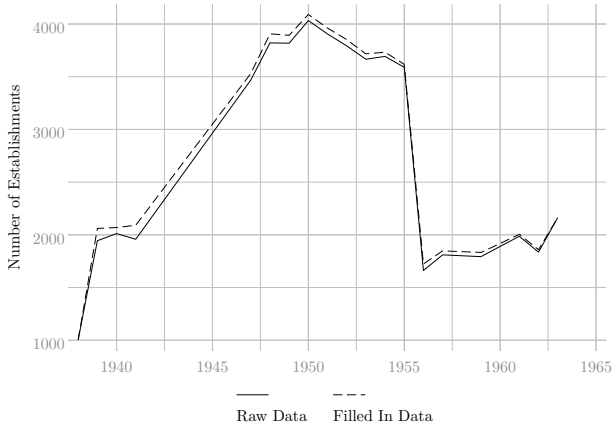
1930 E. D.	1940 E. D.	Description of Enumeration District	1930 Census: Population	1930 Census: Farms
37-53 pt	68-159	Birmingham city - That part of Election Precinct 34 (Tract 19-part) in Block 1W - 43rd Ave., Election Precinct line, 64th Pl., 35th Ave., Woodlawn Rd., N. 2 - 35th Ave., 63rd Ave., 64th Pl., Woodlawn Rd., N. 3 - 64th Pl., 65th, 35th Ave., N.	16,536	
37-53 pt 37-54 pt	68-160	Birmingham city - That part of Election Precinct 34 (Tract 3-part) in Block 1 - That part of the Municipal Airport in Birmingham city 2E - City limit, Municipal Airport line, Woodlawn Rd., L. & N. R.R., Election Precinct line, N. 3 - L. & N. R.R., Woodlawn Rd., 18th Ave., 52nd, N. 4 - L. & N. R.R., 52nd, 18th Ave., 51st, N. 5 - L. & N. R.R., 51st, 18th Ave., 50th, N. 6 - 18th Ave., 51st, 17th Ave., 50th, N. 7 - 18th Ave., 52nd, 17th Ave., 51st, N. 8 - 18th Ave., 16th Dr., 17th Ave., 52nd, N. 9 - 18th Ave., Woodlawn Rd., 17th Ave., 16th Dr., N. 10 - 17th Ave., Woodlawn Rd., 52nd, 16th Dr., N. 11 - 17th Ave., 16th Dr., 52nd, N. 12 - 17th Ave., 52nd, 16th Ave., 51st, N. 13 - 17th Ave., 51st, 16th Ave., 50th, N. 14 - 16th Ave., 51st, 15th Ave., 50th, N. 15 - 16th Ave., 52nd, 15th Ave., 51st, N. 16) - 15th Ave., Forrest Hill Cemetery line, 11th Ave. extended, 17) 11th Ave., 50th, N. 18 - 11th Ave., 51st, 10th Ave., 50th, N. 19 - 11th Ave. extended, Unnamed street, 10th Ave., 51st, N. 20 - 11th Ave. extended, 52nd, 10th Ave., Unnamed street, N.	593	

Figure A.1: The 1940 Enumeration District Description for two enumeration districts in Chicago.

[7]This website contains invaluable tools for genealogical (and, it turns out, economic history) research. We thank David Van Riper at IPUMS for making us aware of this tremendous resource.

[8]In 1940, 60 cities were assigned census tracts, and the group of cities expanded over subsequent decades.

B The Drop in Establishments After 1955



(a) Filled in series

(b) Drop cities that exit post-1955

Figure B.1: The top panel shows the raw number of Green Book establishments over time (solid line) alongside a series that "fills in" establishments that disappear in one year and reappear in another year (dashed line). The bottom panel shows the raw number of Green Book establishments over time (solid line) alongside a series that excludes establishments in cities that exit the Green Books after 1955 (dashed line).

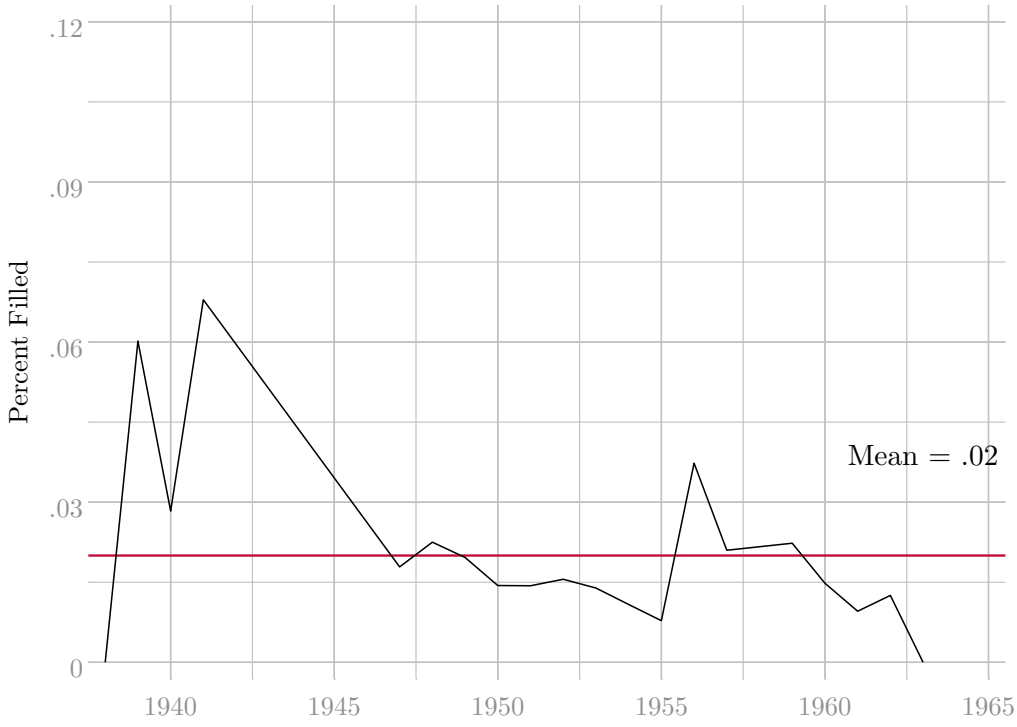
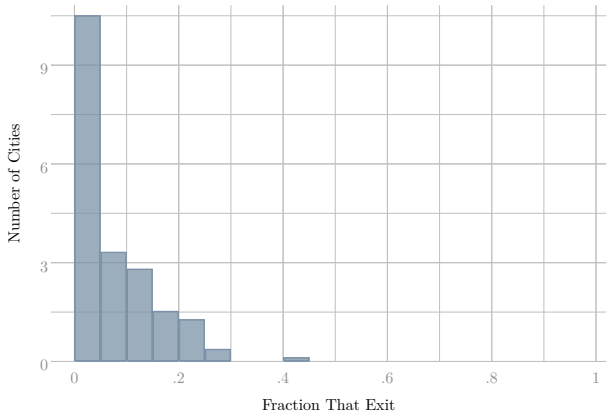
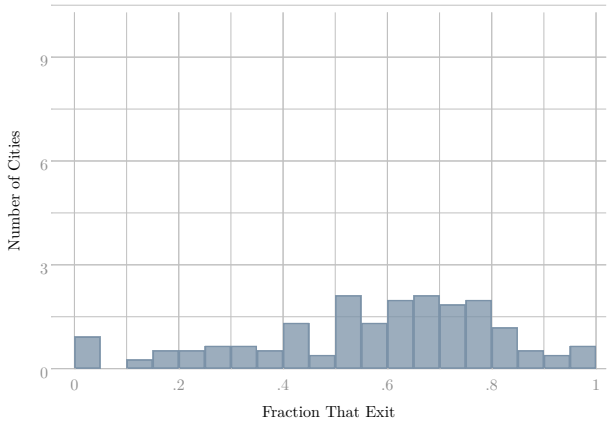


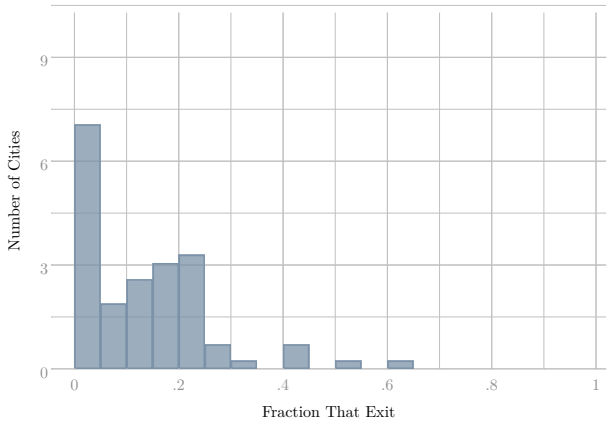
Figure B.2: The percent of establishments that are filled in, by year.



(a) 1954



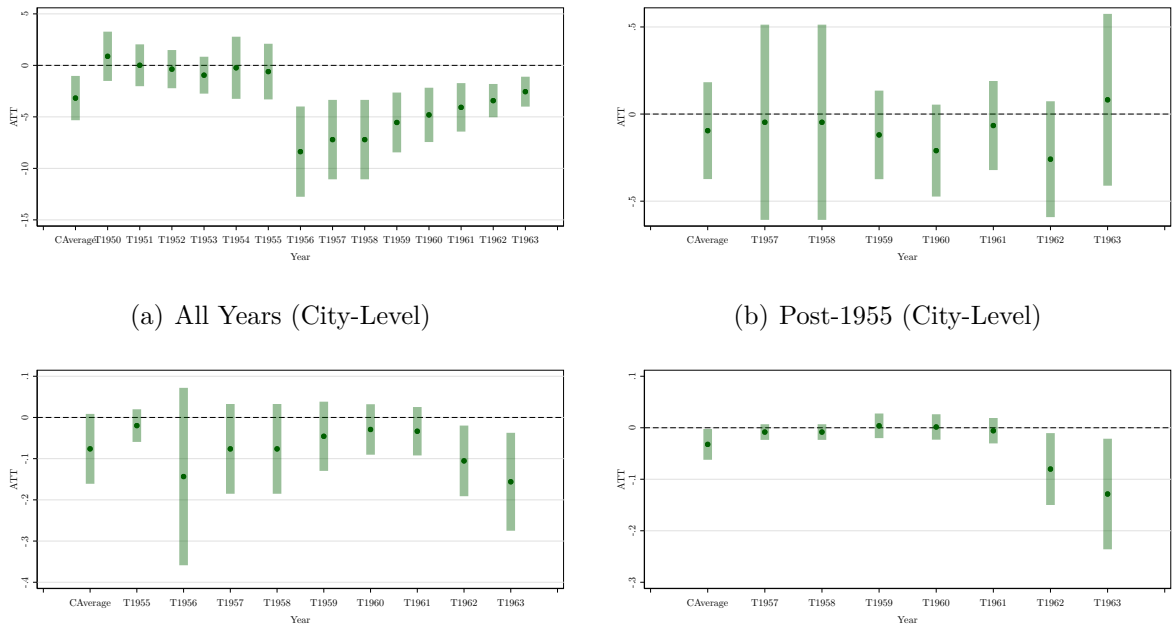
(b) 1955



(c) 1956

Figure B.3: These histograms display the frequency of exit from the Green Books, by year. The horizontal axis displays bins that represent the fraction of Green Book establishments that existed in a city in 1955 that do not reappear in 1956. The vertical axis displays the number of cities that fall within each bin. These plots only include cities that had at least 5 Green Book establishments in 1955.

C Additional Urban Renewal Results



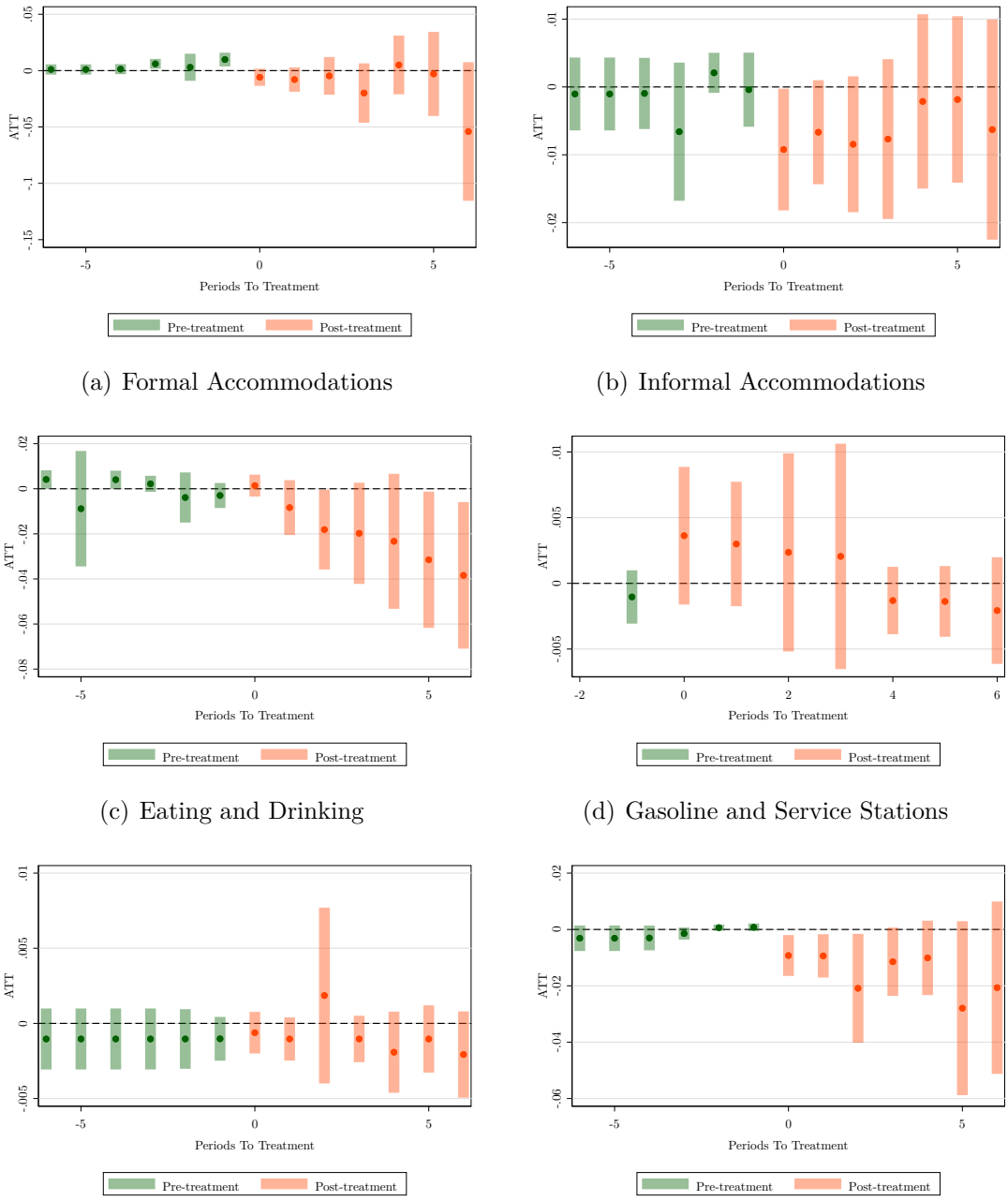
(a) All Years (City-Level)

(b) Post-1955 (City-Level)

(c) All Years (Census Tract)

(d) Post-1955 (Census Tract)

Figure C.1: Coefficient estimates and 95% confidence intervals for the average treatment effect across each year of the analysis. In panel a and b the ATTs estimate the differential change in the number of Green Book establishments in cities with urban renewal projects compared to those without urban renewal projects. In panel c and d the ATTs estimate the differential change in the number of Green Book establishments in the portion of census tracts that intersect with urban renewal areas compared to the portion of the same census tracts that do not intersect with urban renewal areas following the start of an urban renewal project. Estimates are constructed using the method of Callaway and Sant'Anna (2021). The Green Book data span 1947-1963 in panel (a) and (c) and 1956-1963 in panel (b) and (d). Panels (a) and (b) present results from the city-level analysis and panels (c) and (d) present results from the census tract-level analysis.



(a) Formal Accommodations

(b) Informal Accommodations

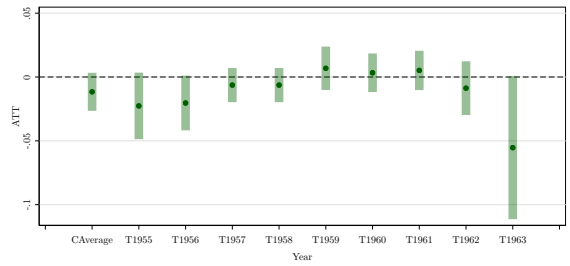
(c) Eating and Drinking

(d) Gasoline and Service Stations

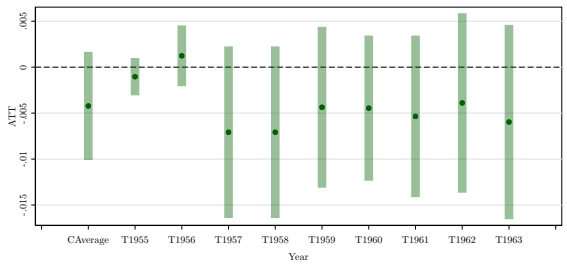
(e) Barber Shops and Beauty Parlors

(f) Other Establishments

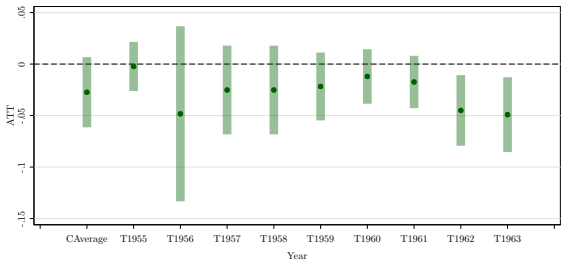
Figure C.2: Coefficient estimates and 95% confidence intervals from event studies that estimate the effect of urban renewal projects. Panel a and b show the differential change in the number of Green Book establishments in cities with urban renewal projects compared to those without urban renewal projects. Panel c and d show the differential change in the number of Green Book establishments in the portion of census tracts that intersect with urban renewal areas compared to the portion of the same census tracts that do not intersect with urban renewal areas following the start of an urban renewal project. Estimates are constructed using the method of Callaway and Sant'Anna (2021). The Green Book data span 1955-1963.



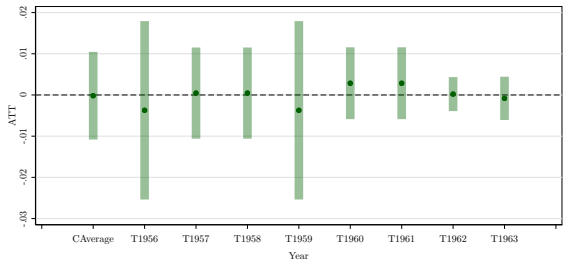
(a) Formal Accommodations



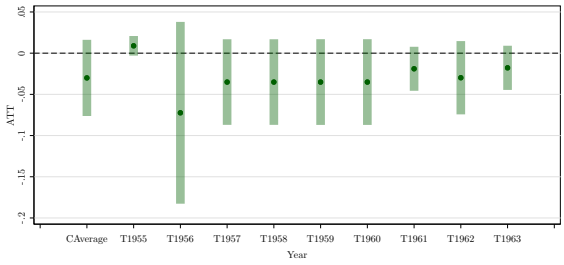
(b) Informal Accommodations



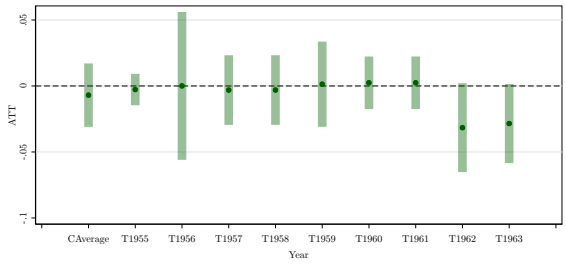
(c) Eating and Drinking



(d) Gasoline and Service Stations



(e) Barber Shops and Beauty Parlors



(f) Other Establishments

Figure C.3: Coefficient estimates and 95% confidence intervals for the average treatment effect across each year of the analysis. The ATTs estimate the differential change in the number of Green Book establishments in the portion of census tracts that intersect with urban renewal areas compared to the portion of the same census tracts that do not intersect with urban renewal areas following the start of an urban renewal project. Estimates are constructed using the method of Callaway and Sant'Anna (2021). The Green Book data span 1947-1963.

D Other Additional Figures

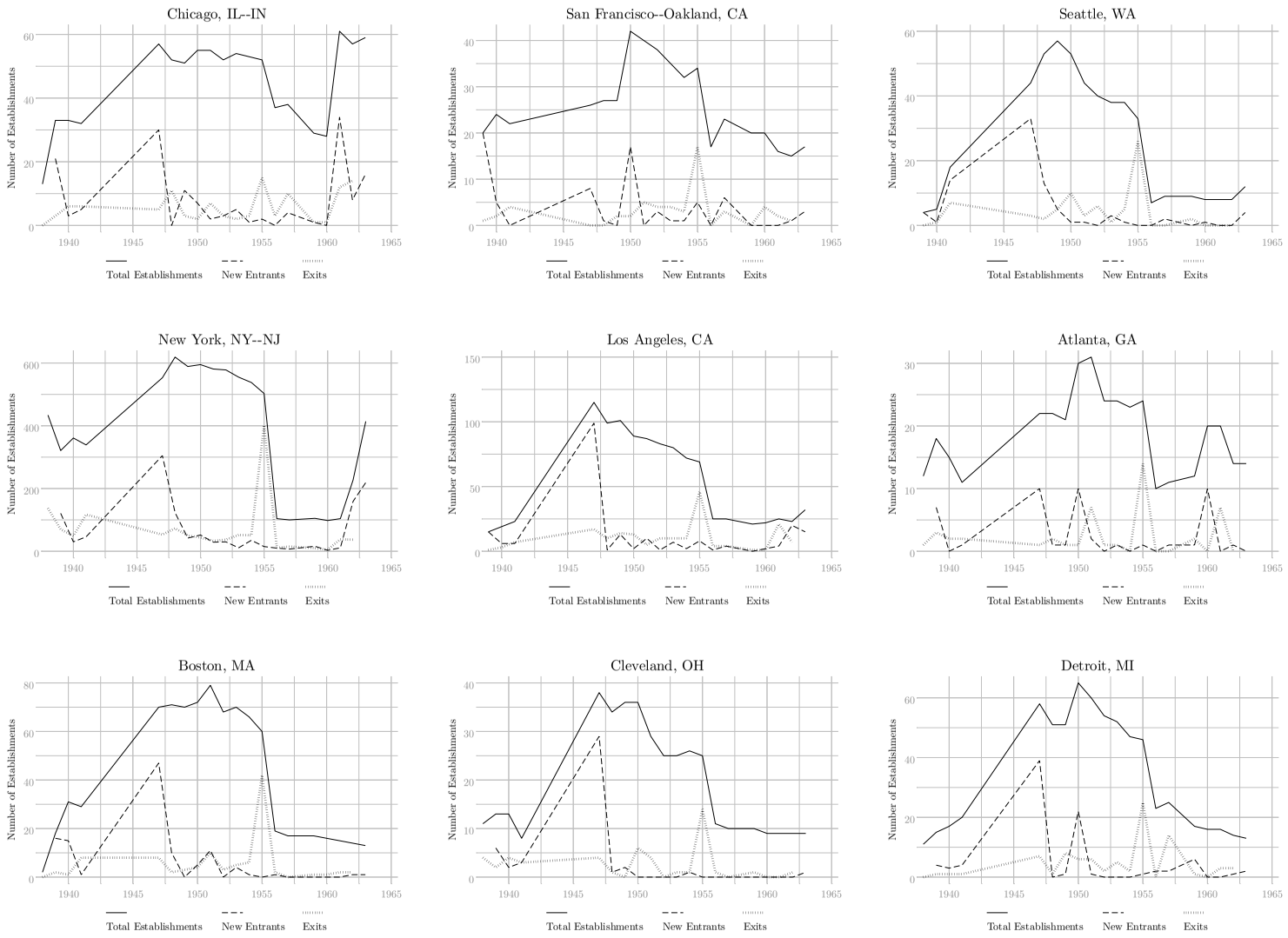


Figure D.1: SMA-level trends in the number of Green Book establishments, the number of establishments that exit, and the number of establishments that enter, for major urban areas in the United States. All values are calculated using urban area boundaries defined as of 1990.

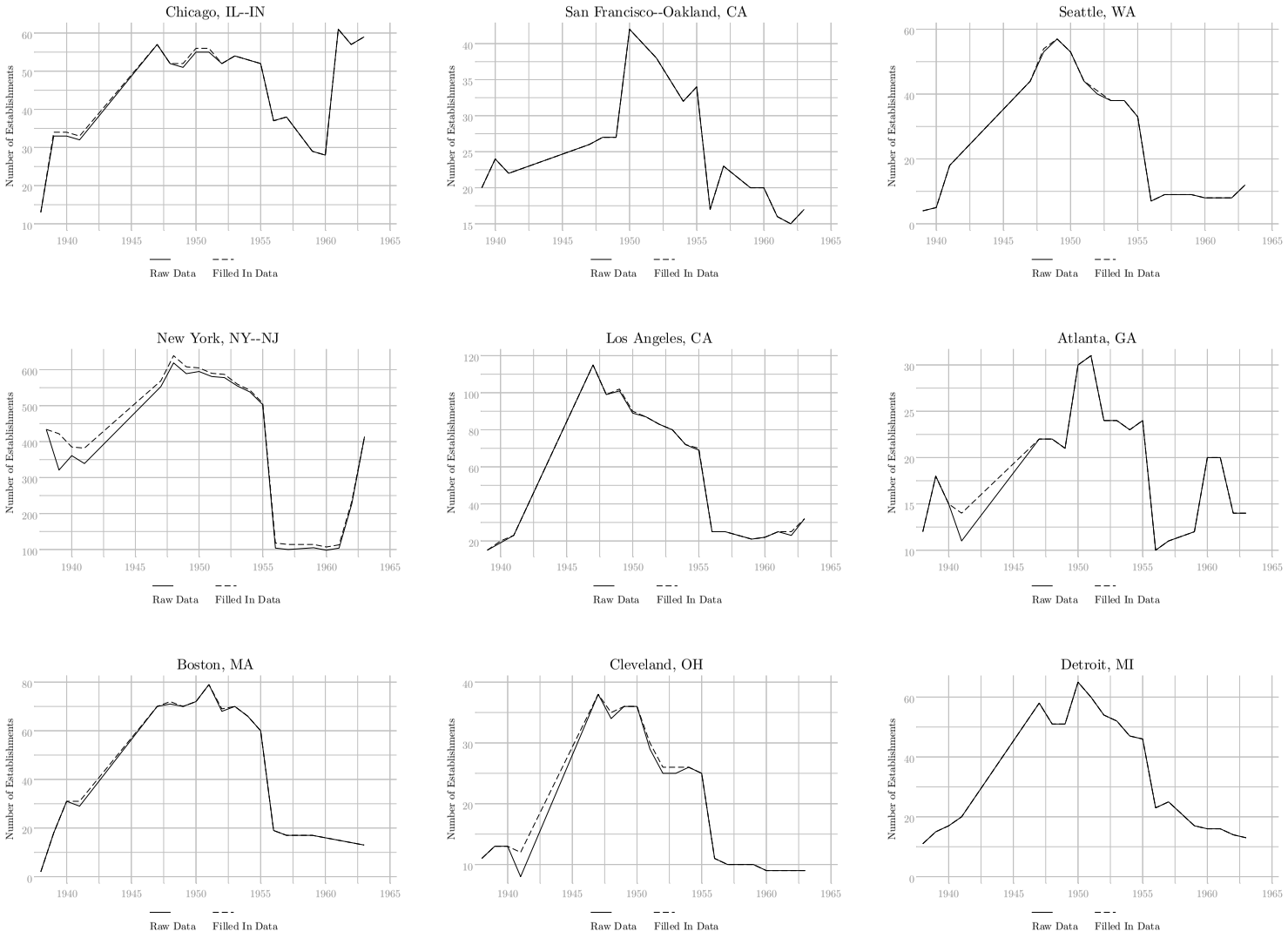
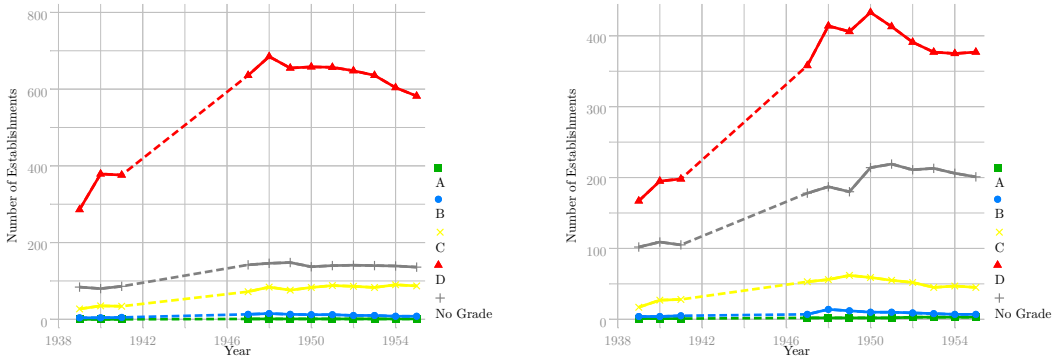


Figure D.2: SMA-level trends in the number of Green Book establishments for major urban areas in the United States. All values are calculated using urban area boundaries defined as of 1990.



(a) Northeast

(b) South

(c) Midwest

(d) West

Figure D.3: Number of establishments over time by HOLC grade and by region.

E Other Additional Tables

Table E.1: The Correlation Between Green Book Establishments and Population

	(1) All	(2) Eating	(3) Formal	(4) Informal	(5) Barber	(6) Gasoline	(7) Other
Log(White Pop)	-0.0483***	-0.0124***	-0.0120***	-0.00746***	-0.00648***	-0.00612***	-0.00396***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Log(Black Pop)	0.0634***	0.0139***	0.0179***	0.0118***	0.0112***	0.00421***	0.00433***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Constant	0.285***	0.0765***	0.0665***	0.0418***	0.0338**	0.0413***	0.0249***
	(0.050)	(0.018)	(0.019)	(0.015)	(0.014)	(0.007)	(0.008)
Observations	6749	6749	6749	6749	6749	6749	6749
Adjusted R^2	0.101	0.062	0.049	0.059	0.042	0.033	0.030

Standard errors in parentheses

All columns include city fixed effects. Standard errors are included in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.2: The Correlation Between Green Book Establishments and Labor Market Characteristics

	(1) All	(2) Eating	(3) Formal	(4) Informal	(5) Barber	(6) Gasoline	(7) Other
Log(White Income)	-0.0423	-0.0129	-0.0181	-0.00940	-0.00673	-0.00405	0.00888
	(0.040)	(0.014)	(0.016)	(0.012)	(0.011)	(0.006)	(0.006)
% White Pop in LF	0.449***	0.0548	0.286***	0.0133	0.0413	0.0171	0.0364
	(0.164)	(0.059)	(0.065)	(0.049)	(0.045)	(0.023)	(0.026)
Log(White Pop)	-0.0662***	-0.0186***	-0.0142***	-0.00978***	-0.00873***	-0.00821***	-0.00658***
	(0.010)	(0.004)	(0.004)	(0.003)	(0.003)	(0.001)	(0.002)
Log(Black Income)	-0.0114	-0.00203	-0.0101	0.00360	-0.00164	0.000148	-0.00135
	(0.023)	(0.008)	(0.009)	(0.007)	(0.006)	(0.003)	(0.004)
% Black Pop in LF	0.357***	0.0813***	0.0911***	0.0769***	0.0670***	0.0232***	0.0171*
	(0.063)	(0.023)	(0.025)	(0.019)	(0.017)	(0.009)	(0.010)
Log(Black Pop)	0.102***	0.0223***	0.0283***	0.0192***	0.0185***	0.00655***	0.00716***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
_cons	0.0830	0.0980	-0.0114	0.00463	0.000945	0.0468	-0.0561
	(0.341)	(0.124)	(0.135)	(0.102)	(0.094)	(0.048)	(0.053)
Observations	4532	4532	4532	4532	4532	4532	4532
Adjusted R^2	0.131	0.078	0.066	0.073	0.054	0.036	0.038

Standard errors in parentheses

All columns include city fixed effects. Standard errors are included in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.3: The Correlation Between Green Book Establishments and Competition (Restricted Sample)

	(1) All	(2) Eating	(3) Formal	(4) Informal	(5) Barber	(6) Gasoline	(7) Other
Total # Est 1939	0.00339						
	(0.007)						
Eating # Est 1939		-0.0144*					
		(0.009)					
Formal # Est 1939			0.00373				
			(0.006)				
Informal # Est 1939				-0.0503***			
				(0.005)			
Barber # Est 1939					0.00364		
					(0.015)		
Gas # Est 1939						-0.0145	
						(0.012)	
Other # Est 1939							-0.393***
							(0.010)
Log(White Income)	-0.0406**	-0.00408	-0.0225***	-0.00196	-0.0112	-0.00365	0.00602
	(0.018)	(0.008)	(0.006)	(0.004)	(0.009)	(0.004)	(0.005)
% White Pop in LF	0.0644	0.0512	-0.00990	0.0348**	0.0204	0.00357	-0.00466
	(0.073)	(0.033)	(0.026)	(0.017)	(0.036)	(0.016)	(0.021)
Log(White Pop)	-0.0114**	-0.00782***	0.00252	0.000964	-0.00499**	-0.00317***	-0.00289**
	(0.005)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Log(Black Income)	0.00635	0.00532	-0.00140	-0.00208	0.00400	0.00198	-0.000367
	(0.011)	(0.005)	(0.004)	(0.003)	(0.006)	(0.002)	(0.003)
% Black Pop in LF	0.0752***	0.0199	0.0199**	0.00290	0.0379***	0.00707	0.00136
	(0.029)	(0.013)	(0.010)	(0.007)	(0.014)	(0.006)	(0.008)
Log(Black Pop)	0.0172***	0.00413***	0.00311***	0.00189***	0.00865***	0.00195***	0.00238***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	3891	3891	3891	3891	3891	3891	3891
Adjusted R^2	0.034	0.016	0.007	0.048	0.040	0.018	0.309

Notes: The dependent variable in columns (1) and (2) is the difference in the number of Green Book establishments in 1940 and 1939. In column (3) the dependent variable is the percent change in the number of Green Book establishments between 1939 and 1940. All columns include city fixed effects. Standard errors are included in parentheses. Column (1) includes the full sample, and columns (2) and (3) only includes census tracts that had a Green Book establishment in 1939. The sample is restricted to census tracts where we have the full set of controls. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.4: The Correlation Between Green Book Establishments and Competition

	(1) Levels	(2) Levels	(3) Percent
Total # Est 1939	0.00380	-0.0641*	-0.0452**
	(0.006)	(0.034)	(0.021)
Log(White Income)	-0.0350**	-0.249	-0.0842
	(0.015)	(0.330)	(0.210)
% White Pop in LF	0.0429	-0.554	-0.279
	(0.062)	(0.823)	(0.524)
Log(White Pop)	-0.0103***	-0.0389	-0.0798*
	(0.004)	(0.069)	(0.044)
Log(Black Income)	0.000241	0.138	-0.0860
	(0.009)	(0.412)	(0.262)
% Black Pop in LF	0.0812***	1.919*	1.200*
	(0.024)	(1.032)	(0.657)
Log(Black Pop)	0.0159***	0.161**	0.107***
	(0.002)	(0.062)	(0.040)
_cons	0.197	-0.717	0.518
	(0.129)	(3.347)	(2.130)
Observations	4532	240	240
Adjusted R^2	0.033	0.125	0.148

The dependent variable in columns (1) and (2) is the difference in the number of Green Book establishments in 1940 and 1939. In column (3) the dependent variable is the percent change in the number of Green Book establishments between 1939 and 1940. All columns include city fixed effects. Standard errors are included in parentheses. Column (1) includes the full sample, and columns (2) and (3) only includes census tracts that had a Green Book establishment in 1939. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230. Themed Issue: Treatment Effect 1.

Manson, S., J. Schroeder, D. V. Riper, T. Kugler, and S. Ruggles (2020). Ipums national historical geographic information system: Version 15.0 [dataset]. Data retrieved from IPUMS USA, `https://doi.org/10.18128/D010.V10.0`.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2020). IPUMS USA: version 10.0 [dataset].