Reproduction exercise for *The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality*

Paula Torres Higuera

May 27, 2022

Abstract

The Wrong Side(s) of the Tracks is a paper written by Professor Elizabeth Oltmans Ananat with the purpose of understanding the relationship between segregation and several population outcomes. The major contribution of Ananat's paper is her instrumental variables (IV) approach to establish causality between segregation and various city-level characteristics using the arrangements of railroad tracks in the nineteenth century to identify exogenous variation in a city's exposure to segregation. The instrument's validity is proved through several falsification checks and analyses that corroborate the fulfillment of IV assumptions. The purpose of the present project is to perform a replication exercise that allows to improve the reproducibility of Ananat's paper.

Key words: Segregation, Urban Economics, Replication

1 The Wrong Side(s) of the Tracks

Even though African Americans represent approximately 10% of the US population, the average African American lives in a neighborhood with a black population of approximately 50%. Furthermore, it has been documented that more segregated cities have worse population outcomes than segregated cities. This makes segregation one of the main suspects of persistent economic inequality between blacks and whites (Ananat, 2011). In *The Wrong Side(s) of the Tracks*, Professor Ananat attempts to answer the research question *How does segregation affect population characteristics?* from a causal perspective.

The econometric method is an Instrumental Variables specification to account for omitted variable bias (for example, unmeasured political attributes that directly affect both segregation and city outcomes) to correct endogeneity and therefore to identify the causal effect of segregation on net population outcomes. Ananat uses the railroad division index (RDI), defined as a measure of division of a city's land into subunits generated by railroads, as the instrument. The specification for the two-stage least squares (2SLS) estimation is defined as:

$$Seg = \alpha_1 RDI + \alpha_2 X + \mu \tag{1}$$

$$Y = \beta_1 Seg + \beta_2 + \varepsilon \tag{2}$$

Where Seg is the level of segregation, RDI is the instrument, Y is the outcome of interest and X is a vector of control variables. For RDI to be a valid instrument it must satisfy four major assumptions. First, there should exist a strong first stage which means that the index must generate a significant level of racial segregation. This is known as the relevance assumption. Second, a higher RDI index induces more segregation on cities affected by the instrument; this is known as the monotonicity assumption. In this case, RDI can affect segregation since a city with more subunits has more clear boundaries that serve to enforce segregation, and therefore implies RDI has a non-negative effect on segregation, which validates the previous

two assumptions. Third, the effect of RDI on population outcomes for each city must only be induced through segregation, which means that the exclusion restriction must be satisfied. The evidence presented in the paper shows that RDI did not alter city characteristics prior to the Great Migration, and only influences population outcomes in cities that experienced substantial black migration. Lastly, people's city choices during the Great Migration cannot be correlated with railroad division.

Ananat presents various tests, including first stage estimation, falsification checks and historical facts to demonstrate that her chosen instrument satisfies all four assumptions and is thus a valid instrument. Intuitively, railroads cover enough land to offer a clear boundary that allows for communal consensus on neighborhood borders by the community. Furthermore, railroads were constructed to optimize placement and variation in ground slope before 1900, when most African Americans lived in the South, and more than 20 years before the Great Migration, which means that they were not constructed to segregate or influence city outcomes.

Since it is demonstrated that RDI is a valid instrument, the results derived from estimating the model allow to identify casual relationships between segregation and city-level outcomes. The main results from the 2SLS model evidence the negative impact that race segregation has on population outcomes. It is shown that exogenous segregation leads cities to have higher poverty rates among African Americans and white communities with lower poverty rates. These results extend to other outcomes such as migration, housing demand and human capital. In all cases, segregation is seen as a disamenity. Therefore, segregation raises inequality between black and white communities in a city, mainly by lowering black peoples' outcomes, while inducing more equal outcomes within its white community.

2 Reproduction Score

The major findings, validation tests and robustness checks are presented in six tables. Results from these tables are replicable using the provided code. Ananat also presents six maps and two graphs that are not replicable with the provided code. Furthermore, although the results

from the tables are replicable, the file does not include codes to export the tables in any format. Therefore, the main code improvement done in this project is to include codes that allow to export tables in a *.tex* format that resembles that from the paper. Another important aspect of the original reproduction package is that the author only includes analytic data and not raw data. Taking all the previous factors into account, the global reproductibility score for this paper is Level 5 (L5) which means that there is analytic data and analysis code available that reproduce the same results from the tables, although this code does not export results. In terms of individual reproductibility scores, the tables are L5 for the reasons explained above, but the maps are L1 since there is neither code nor data to reproduce these figures. Finally, graphs would have a L3 reproductibility score since there is data available to reproduce them, but no code.

3 Code improvements

The main code improvements with which I contribute to this reproduction package are comments to the code so it is easier to understand the specifications and their correspondance with the paper. I also include labels to variables to better identify data, and codes to export main results to tables in *.tex* format. The revised reproduction package can be found in GitHub (Torres Higuera, 2022) (URL in the bibliography section). This package includes detailed comments on the improvements. The main Stata file is "aej_dofile" ("main" in the folder uploaded to Bloque Neón). This dofile includes comments for all the code.

4 Replication

As previously mentioned, the paper has six tables, two figures and six maps. Out of these results, only the tables are reproducible based on the replication code and data provided by the author. However, the code does not export the tables to any format. Therefore, the following tables are created using codes that I included to improve the replication package. The tables are as similar as possible to resemble those in the paper. However, they are not identical.

4.1 Replication of results from table 1

Table 1 replicates the results that the author presents in table 1 of her paper. In this table, Ananat presents the results for testing RDI as an instrument. It is important to note that the codes for the regressions provided by the author for the per 1,000 variables (columns 2, 3 and 7) did not produce the results presented in the paper. This was the case since the variables used were the originals and not the per 1,000. Therefore, three new variables were created to replicate the results from the paper for these three specifications.

The coefficients from columns 10, 11, 12 and 13 differ from the ones presented by Ananat, although not by much. It is not clear why in this case the codes provided by the author do not replicate her results in an exact way.

Table 1. Testing RDI as an Instrument

	First Stage	Falsification checks					
		1910 city characteristics					
	1990 dissimilarity index (1)	Physical area (square miles/ 1,000) (2)	Pob. (1,000s) (3)	Ethnic dissimilarity index (4)	Ethnic isolation index (5)	Percent black (6)	Street-cars per cap. (1,000) (1915) (7)
RDI	0.357***	-3.993	0.666	0.076	0.027	-0.001	-0.132
	(0.088)	(11.986)	(1.363)	(0.185)	(0.070)	(0.010)	(0.183)
Track lenght per square kilometer	18.514	-574.401	75.553	15.343	-12.439	9.236***	3.361
	(10.731)	(553.669)	(134.815)	(53.248)	(17.288)	(0.650)	(20.507)
Mean of dependent Variable	0.569	14.626	1.527	0.311	0.055	0.014	0.179
N	121	58	121	49	49	121	13

	Falsification checks						
	1920 city characteristics						
	Percent black (8)	Percent literate (9)	Labor force participation (10)	Percent of empl. in trade (11)	Percent of empl. in manufacturing (12)	Percent of empl. in railroads (13)	1990 income seg. (14)
RDI	0.013	0.053*	0.028	-0.080	0.191	-0.074	0.032
	(0.009)	(0.030)	(0.024)	(0.094)	(0.137)	(0.068)	(0.032)
Track lenght per square kilometer	9.119***	0.180	-3.427**	-0.152	18.400*	1.592	-2.504
	(0.615)	(0.880)	(1.500)	(2.910)	(10.911)	(2.428)	(1.626)
Mean of dependent Variable	0.016	0.959	0.419	0.058	0.462	0.003	0.217
N	121	121	121	121	121	121	69

Notes: Robust standard errors in parentheses

4.2 Replication of results from table 2

Table 2 replicates the results that correspond to those in the first panel of table 2 in the paper which are the effects of segregation on within race poverty and inequality among blacks and whites. The main findings of the paper correspond to columns 3 and 4. Coefficients perfectly match those presented by Ananat.

Table 2. The Effects of Segregation on Poverty and Inequality Amoung Blacks and Whites

	OLS: Effect of 1990 dissimilarity index		Main results: 2SLS RDI as instrument for 1990 dissimilarity		Falsification: Reduced form effect of RDI among cities far from the south	
	Whites (1)	Blacks (2)	Whites (3)	Blacks (4)	Whites (5)	Blacks (6)
Within-race poverty and inequality						
Gini index	-0.079*	0.459***	-0.334**	0.875*	-0.110	0.167
	(0.037)	(0.093)	(0.099)	(0.409)	(0.066)	(0.424)
Poverty rate	-0.073***	0.182***	-0.196**	0.258*	-0.036	-0.136
	(0.019)	(0.045)	(0.065)	(0.108)	(0.035)	(0.094)

Notes: 2SLS and reduced form estimates control for total track length. All outcomes except poverty rates are logged. Robust standard errors in parentheses.

Table 3 replicates the results present in the paper in the second panel of table 2 which are the effects of segregation on between race poverty and inequality among blacks and whites. The main findings of the paper correspond to those in column 2. Coefficients perfectly match those presented by Ananat.

Table 3. The Effects of Segregation on Poverty and Inequality Amoung Blacks and Whites

	OLS: Effect of 1990 dissimilarity index	Main results: 2SLS RDI as instrument for 1990 dissimilarity	Falsification: Reduced form effect of RDI among cities far from the south
	(1)	(2)	(3)
Between-race inequality			
90 white: 90 black	0.111	-0.131	-0.443
	(0.086)	(0.312)	(0.217)
10 white: 10 black	1.295***	2.727**	-0.135
	(0.249)	(0.867)	(0.532)
90 white: 10 black	1.172***	1.789*	-0.449
	(0.282)	(0.758)	(0.558)
90 black: 10 white	-0.234	-0.807*	0.130
	(0.131)	(0.384)	(0.248)

Notes: 2SLS and reduced form estimates control for total track length. All outcomes except poverty rates are logged. Robust standard errors in parentheses. N = 121 for columns (1) and (2). N = 29.

Table 4 replicates the robustness checks that the author presents in table 3 of her paper. Most of the results in table 4 replicate those presented by Ananat. However, the coefficients for the regressions controlling for 1920 share employed in manufacturing do not match those

presented in the paper (except the coefficient associated to white poverty rate). It is not clear why these coefficients do not match since the specifiations used correspond to those provided by the author in the replication code.

Table 4. Robustness Checks: 2SLS Effects of 1990 Segregation, Controlling for City-Level Characteristics

	Outcome: Gini Index		Outcome: Poverty rate	
	Whites (1)	Blacks (2)	Whites (3)	Blacks (4)
With controls for 1990 city characteristics				
Population	-0.371***	0.898*	-0.212**	0.291**
	(0.107)	(0.434)	(0.068)	(0.109)
Percent black	-0.473***	0.886	-0.241**	0.360**
	(0.171)	(0.547)	(0.097)	(0.141)
Education	-0.361**	0.887	-0.162**	0.222
	(0.148)	(0.664)	(0.080)	(0.174)
Share employed in manufacturing	-0.359**	1.106	-0.272**	0.219
	(0.175)	(0.777)	(0.124)	(0.195)
Labor force participation	-0.295***	0.907**	-0.142***	0.321***
	(0.092)	(0.393)	(0.040)	(0.105)
Number of local governments (N = 69)	-0.386*	0.792***	-0.118	0.519***
	(0.203)	(0.277)	(0.077)	(0.169)
With controls for 1920 city characteristics				
Population	-0.374***	0.899**	-0.214***	0.281**
	(0.106)	(0.442)	(0.071)	(0.115)
Percent black	-0.364***	0.896**	-0.199***	0.296***
	(0.114)	(0.434)	(0.069)	(0.109)
Literacy	-0.312***	1.029**	-0.163***	0.270**
	(0.107)	(0.470)	(0.061)	(0.124)
Share employed in manufacturing	-0.401***	0.904*	-0.213***	0.307**
	(0.132)	(0.483)	(0.081)	(0.122)
Labor force participation	-0.305***	0.849**	-0.187***	0.243**
	(0.085)	(0.372)	(0.061)	(0.104)
Control for propensity score	-0.412**	1.038	-0.189**	0.304*
	(0.181)	(0.639)	(0.094)	(0.177)

Notes: *N=121* except where noted. All estimates control for total track length. Gini indexes are logged. Robust standard errors in parentheses. 1990 education controls include percent of blacks and of whites who have less than a high school diploma, exactly a high school diploma, some college, and college diploma. Propensity score to have an above-median RDI generated based on: 1920 population, 1920 percent black, 1920 share of employment in manufacturing, 1920 literacy, 1920 labor force participation, and distance from the South.

4.3 Replication of results from table 4

Table 5 replicates the results on the effects of 1990 segregation on 1990 city demand which are presented in table 4 of the paper. The results presented in this table are the same as those presented by the author in her paper.

Table 5. The Effects of 1990 Segregation on 1990 City Demand

	Outcome: Percent of residents who are in-migrants		Outcome: Median rent		Outcome: Median rent as a percent of income		Outcome: Share of households with more than one person per room	
	Whites (1)	Blacks (2)	Whites (3)	Blacks (4)	Whites (5)	Blacks (6)	Whites (7)	Blacks (8)
OLS								
Dissimilarity Index	-0.153***	-0.294***	-313.851***	-391.813***	-8.535***	-3.490	-0.062***	-0.103***
	(0.032)	(0.052)	(83.934)	(75.643)	(1.337)	(2.676)	(0.014)	(0.022)
IV								
Dissimilarity Index	-0.155**	-0.271**	-636.453**	-623.642***	-16.666***	-3.416	-0.116***	-0.165***
	(0.073)	(0.115)	(276.151)	(156.969)	(3.643)	(5.387)	(0.037)	(0.047)
Falsification: Reduced form effect of RDI among cities far from the South								
RDI	0.019	0.058	295.092	326.160**	0.427	3.660	0.034	0.062
	(0.063)	(0.158)	(275.735)	(158.217)	(2.061)	(3.572)	(0.038)	(0.048)

Notes: Robust standard errors in parentheses. All 2SLS and reduced form regressions control for total track length per square kilometer. N = 121 for top two panels; N = 29 for falsification check on subset of cities at least 400 miles from the South.

4.4 Replication of results from table 5

Table 6 replicates the results on the effects of 1980 dissimilarity on human capital of 22-to 30-year-olds in 1980 which are presented in table 5 of the paper. The results presented in this table are the same as those presented by the author in her paper.

Table 6. The Effects of 1980 Dissimilarity on Human Capital of 22- To 30-Year-Olds in 1980 (OLS)

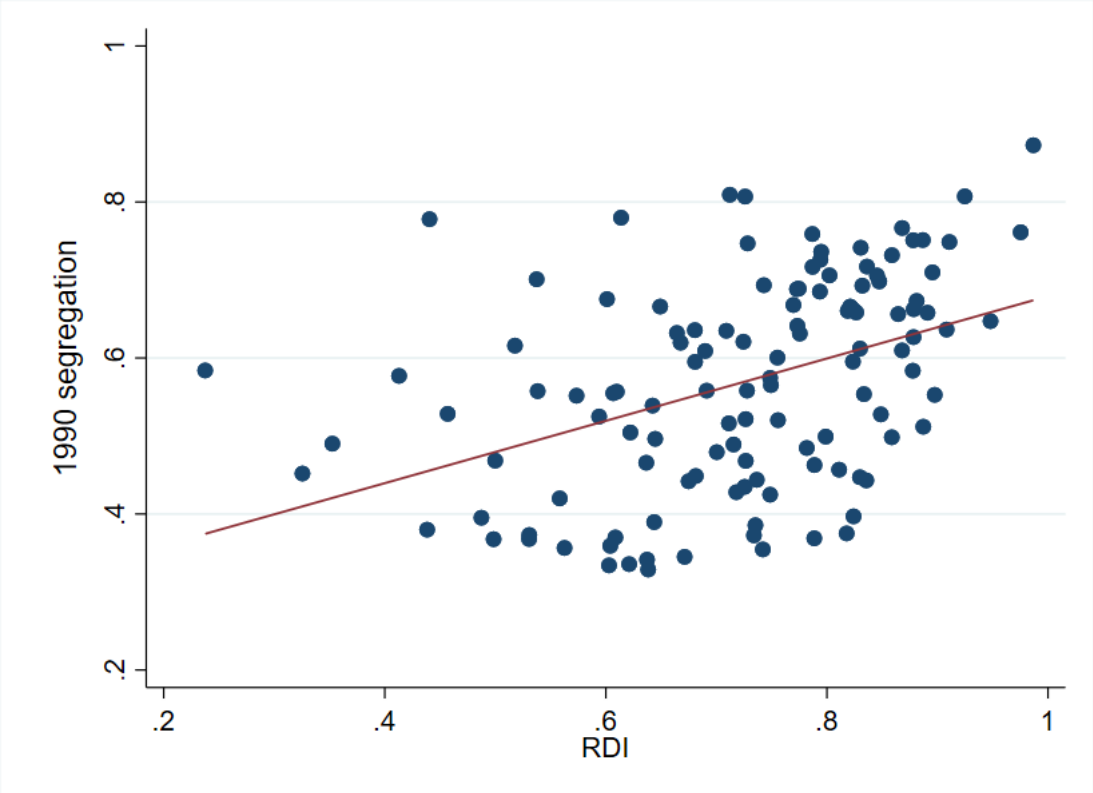
	Outcome: Share who are high school dropouts		Outcome: Share who are high school graduates		Outcome: Share who have some college		Outcome: Share who are college graduates	
	Whites (1)	Blacks (2)	Whites (3)	Blacks (4)	Whites (5)	Blacks (6)	Whites (7)	Blacks (8)
OLS								
Dissimilarity Index	0.025	0.354***	0.249***	0.262	-0.153***	-0.355*	-0.121	-0.262*
	(0.039)	(0.087)	(0.060)	(0.137)	(0.043)	(0.139)	(0.089)	(0.126)
IV								
Dissimilarity Index	-0.144	0.431**	0.458**	0.652*	-0.174	-0.786**	-0.140	-0.297
	(0.149)	(0.201)	(0.174)	(0.333)	(0.110)	(0.321)	(0.152)	(0.219)
Falsification: Reduced form effect of RDI among cities far from the South								
RDI	0.089	0.100	-0.003	0.320	-0.021	-0.274	-0.065	-0.146
	(0.085)	(0.235)	(0.086)	(0.292)	(0.064)	(0.300)	(0.079)	(0.232)

Notes: Robust standard errors, corrected for arbitrary nonindependence of observations within a city, are in parentheses. All 2SLS and reduced form regressions control for total track length per square kilometer. Instrumented variable is 1980 dissimilarity; instrument is RDI. Observations are shares in each educational category within city × single year of age × race cells.

4.5 Replication of results from figure 3

Figure 1 replicates figure 3 from the paper. The code to create this figure was not included by the author. Therefore, the code to create this graph represents an important addition to the replication package.

Figure 1. Full Sample Relationship between RDI and Segregation



4.6 Replication of results from table A

Table 7 replicates results from table A presented in the appendix by Ananat. This table only includes 27 out of the 44 variables presented in the original table. This is the case since the dataset provided by the author in the original replication package includes 113 variables without labels. Therefore, I was only able to identify 27 variables based on their names. All the results in the following table match those presented by the author.

Table 7. Mean Characteristics of Cities In and Out of Sample

Cutler-Glaeser-Vigor variable	Not in sample	(Standard error)	In sample	(Standard error)	Difference in means	*p*-value
Isolation index-1890	0.049	0.007	0.053	0.008	-0.004	0.698
Isolation index-1940 tract	0.355	0.053	0.318	0.043	0.037	0.586
Isolation index-1940 ward	0.234	0.034	0.198	0.023	0.036	0.361
Isolation index-1970	0.343	0.034	0.365	0.023	-0.022	0.578
Isolation index-1990	0.229	0.022	0.214	0.017	0.015	0.586
Dissimilarity index-1890	0.385	0.032	0.383	0.024	0.002	0.956
Dissimilarity index-1940 tract	0.736	0.029	0.742	0.019	-0.006	0.862
Dissimilarity index-1940 ward	0.570	0.032	0.570	0.022	0.000	0.990
Dissimilarity index-1970	0.744	0.015	0.740	0.012	0.004	0.843
Dissimilarity index-1990	0.574	0.016	0.569	0.012	0.005	0.798
Percent black-1890	0.030	0.005	0.027	0.003	0.004	0.532
Percent black-1940	0.058	0.007	0.041	0.005	0.018	0.034
Percent black-1970	0.056	0.006	0.062	0.005	-0.006	0.477
Percent black-1990	0.067	0.006	0.061	0.005	0.005	0.480
Population-1890	129,828.891	56,323.791	66,043.966	19,199.357	63,784.925	0.242
Population-1940	390,895.128	170,643.415	203,676.338	40,731.773	187,218.790	0.206
Population-1970	919,238.686	261,007.290	681,598.716	129,697.523	237,639.970	0.375
Population-1990	689,768.265	135,048.715	590,188.496	96,574.448	99,579.769	0.538
Number of wards-1890	17.778	3.724	13.421	1.731	4.357	0.288
Number of wards-1940	15.929	2.641	14.122	1.440	1.807	0.519
Number of tracts-1940	146.059	53.033	103.348	21.920	42.711	0.417
Number of tracts-1970	211.118	61.515	161.811	29.971	49.307	0.432
Number of tracts-1990	203.687	44.957	137.496	20.894	66.191	0.131
Total area-1900	19,283.332	5,711.826	11,764.322	1,755.343	7,519.011	0.147
Total area-1940	32,855.273	9,499.416	27,137.362	6,214.421	5,717.911	0.610
Total area-1970	2,344.362	604.402	1,615.082	201.438	729.280	0.184
Total area-1990	2,386.990	469.458	1,825.886	262.051	561.104	0.262

4.7 Content not included

It is important to note that the code provided by the author is not sufficient to replicate all tables and figures. For example, figure 4 could not be replicated since there is no code provided and since it is hard to infer from the paper the process behind the creation of said graph. Also, there is neither data nor codes provided to replicate the maps that are presented in the paper.

References

Ananat, E. O. (2011). The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2):34–66.

Torres Higuera, P. (2022). Revised reproduction package for oltmans 2011. https://github.com/ptorresh/Revised-reproduction-package-for-Oltmans-2011-.