Microeconometrics Final Assignment

January 18, 2025

Microeconometrics

Batres, Gabriel - 673209

Dotchev, Emil - 693434

Contents

1 Introduction

Segregation is an important issue affecting society. Within American cities, it is a visible characteristic of certain neighborhoods. Cities with more segregation tend to have worse economic outcomes for the residents in the segregated part of town. And longstanding economic disparities are seen as the result of segregation. The paper "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality" by Elizabeth Oltmans Ananat (2011) explores a new theory of causal effect of segregation on city-wide poverty and inequality in the United States.

Although numerous other studies have previously suggested that neighborhoods with high levels of segregation tend to have worse economic outcomes, isolating the exact causal effects of segregation has remained a difficult task. One focus of research was to attempt to measure how segregation affects the labor market and human capital. Critics of this were skeptical that this approach suffered from omitted variable bias and endogenous migration effects. The way Ananat, 2011 addresses this problem is by exploring an instrumental variable approach, one of the most popular techniques in causal econometrics. In particular, she used the arrangement of railroad tracks in the nineteenth century as an instrument. The approach is novel, as it is based on geographical divisions preceding The Great Migration, which act as a "natural" catalyst for segregated living without the presence of inter-racial tensions. This allows Ananat to isolate the plausibly exogenous variation in the susceptibility of areas to segregation. Her results demonstrate that this divided way of living indeed increases metropolitan rates of black poverty and overall black-white income disparities while additionally decreasing rates of white poverty and inequality within the white population.

The purpose of this paper is to replicate and comment on the results of Ananat, 2011. We were provided data to reproduce methodology and evaluate the results that Ananat, 2011 was able to demonstrate. We utilized R to perform the statistical analysis. We will now showcase and discuss the main points of Ananat's analysis by reproducing figures 3 and 4, connected with the data description and Tables 1 through 4, which follow the research design.

2 Describing the data

In an ideal world, we could conduct a randomized experiment that would test how city characteristics differ when there is more or less segregation. Since this is not possible, the instrumental approach used by Ananat, 2011 tests the the exogenous variability of railroad construction impacts the individual treatment effect of segregation on resident outcomes. She also argues that this method also provides insight into selections effects of individuals who moved during the Great Migration.

Our first task in reproducing the analysis was to investigate the data provided. We were provided data for 121 cities. These cities were the ones that Ananat, 2011 focused on for this paper. All of the cities were located in either the Northeast, Midwest, or West regions. In addition to the city identifier, we had 63 columns with numeric values. We created a summary table of each variable to get a quick view of the statistical distribution of that variable. The appendix of Ananat, 2011 had a table of the "Mean Characteristics of Cities In and Out of Sample", this helped identify key variables and give confidence in the quality of the data provided. Additionally, without having direct data documentation, we used string matching to find columns of interest.

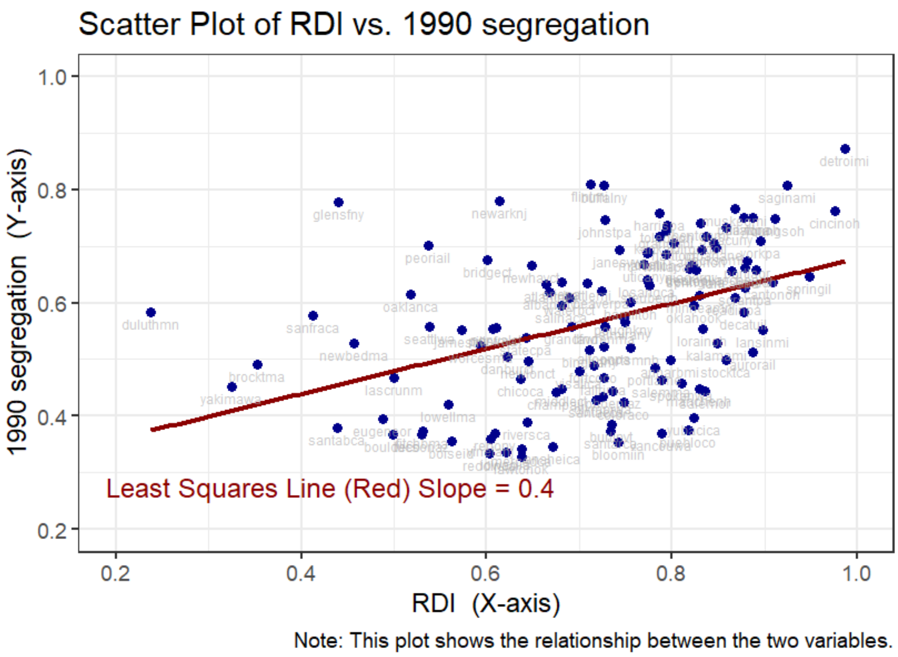

Figure 1: Full Sample Relationship between RDI and Segregation

In this first figure, we see the relationship between the 2 variables of interest, "railroad division index" (RDI) and segregation. Ananat, 2011 creates a measure named "railroad division index" (RDI) to measure the exogenous impact of railroad construction on segregation. The RDI measures how divided a city is by the areas created by the railroad construct. Segregation is measured by an index of dissimilarity, which calculates the concentration of black people in a census tract. In the figure we see a clear positive relationship between increasing RDI and increased segregation. Railroad tracks in cities with high segregation would have helped define neighborhoods where more black people live and neighborhoods where more white people live.

To recreate this image, we explored the data provided and identified individual variable with similar distributions to what was presented in the original paper. Once we could recreate the relationship, we validated the results of the output look similar. The plot created by R matches with what was presented in the paper. The results validate the relevance of RDI as an instrument.

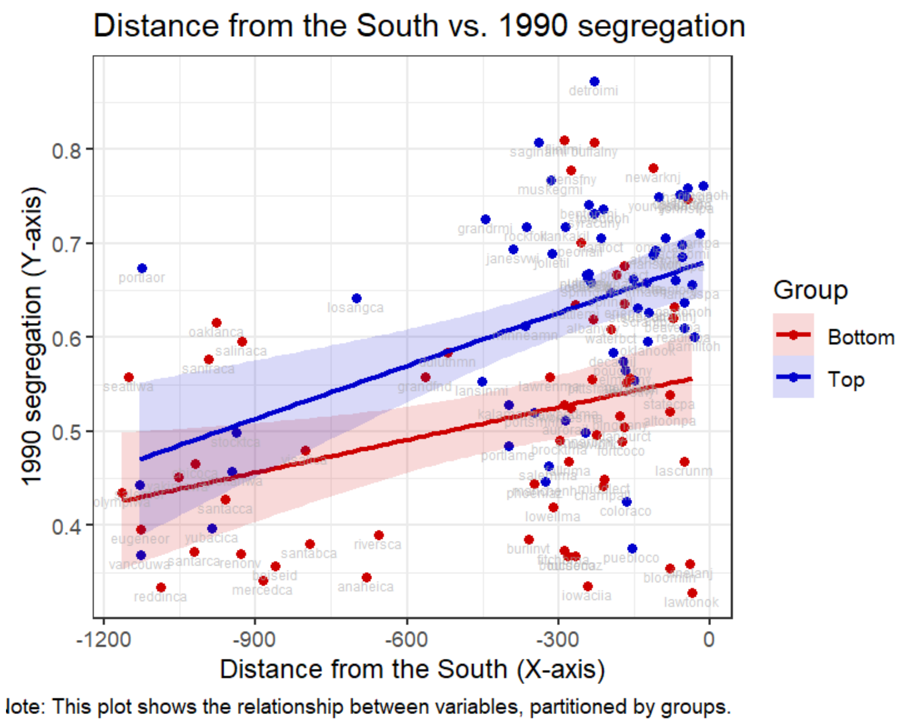

Figure 2: Relationship between Distance to the South and Segregation, by RDI

Cities closer to the south had higher black inflows, on average, during the Great Migration. As the author states, we could expect that these cities would be more segregated, and our instrument variable would show an increased causal relationship. The next figure we reproduced addresses this. In this figure we partition the data by cities closer to the south vs cities farther away. The author used a distance of 400 miles. This distance was selected because the it was in the 75th percentile of distances. In the figure we can see that cities farther from the south (Red) have less demand for segregation as the RDI increases, whereas cites closer to the south (Blue) desire more segregation. From these figures, we have a plausible hypothesis of railroad construction inducing segregation to test.

3 Analysis and Results

Table 1: Testing RDI as an Instrument

	Dependent variable:						
	1990 dissimilarity index	Physical area (1910)	Population (1910)	Ethnic dissimilarity index (1910)	Ethnic isolation index (1910)	Percent black (1910)	Street-cars per capita (1915)
	(1)	(2)	(3)	(4)	(5)	(6)	(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 length per km^2	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 Dep. Variable	0.569	14.626	1.527	0.311	0.055	0.014	0.179
Observations	121	58	121	49	49	121	13

	Dependent variable:						
	Percent black (1920)	Percent literate (1920)	Labor force participation (1920)	Percent in trade (1920)	Percent in manufacturing (1920)	Percent in railroads (1920)	1990 income segregation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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 length per km^2	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 Dep. Variable	0.016	0.959	0.419	0.058	0.462	0.003	0.217
Observations	121	121	121	121	121	121	69

Note: *p<0.1; **p<0.05; ***p<0.01

Robust standard errors in parentheses.
From Cutler-Glaeser-Vigdor data; sample limited to what that dataset provides.
Calculated from ipums.org; full sample represented.

Table 1 presents the test of RDI as a relevant instrumental variable. The significant correlation with the variable of interest, namely the 1990 dissimilarity measure, is confirmed to be, even in the presence of a control variable for the length of the rail track, greater than 0.35. This is in line with the regression slope of Figure 1 and the result is statistically significant at the 1% level which lends support to the relevance of the IV. To further establish that RDI can be used as an instrument, Ananat, 2011 performs additional tests on 13 target variables describing city characteristics. The goal is to showcase that RDI is not directly correlated to any of the target variables. Thus, it can be used to isolate the effects of the variation of segregation alone. In our recreation of the table, we mostly obtain the same or similar results. Small numeric differences can be found in the bottom part of the table likely due to some inconsistencies in the data and/or due to the inclusion of further unspecified control variables. Importantly, significance testing with robust standard errors renders equivalent results as the ones in Ananat, 2011. Our code for this part can be retrieved in Listing 1.

Table 2: The Effects of Segregation on Poverty and Inequality among Blacks and Whites

	OLS		2SLS		Falsification	
Outcome	Whites	Blacks	Whites	Blacks	Whites	Blacks
Gini Index	-0.079**	0.459***	-0.334***	0.875**	-0.11	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)*

	OLS	2SLS	Falsification
Outcome	White:black ratios	White:black ratios	White:black ratios
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)
10 white: 90 black	-0.234*	-0.807**	0.13
	(0.131)	(0.384)	(0.248)

Table 2 explores the causal effect of segregation (measured as 1990 dissimilarity) on the poverty rate and inequality (measured by the Gini index) by race. The main results of interest are in the 2SLS column as it represents the two-stage least squares model using RDI as an instrument (total track length is taken as control). Consistently with the findings of Ananat, 2011 we estimate that a one-standard-deviation (14 point) increase in segregation causes roughly $-0.334 * 14 = 4.676 \approx 4.7$ percent decrease in the white gini index and by contrast a $0.875 * 14 = 12.25 \approx 12.3$ percent increase in the black Gini index. The same increase in dissimilarity leads to a ≈ 2.7 percent decrease and a ≈ 3.6 percent increase in white and black poverty rates respectively. The OLS part of the upper table serves to show that OLS, although still showing statistically significant results in the right directions, underestimates the magnitude of this causal effect.

The bottom part of Table 2 examines the effect of segregation on between-racial inequalities. Outcomes of interest are ratios of different percentiles of the income distributions of blacks and whites. Identical to Ananat, 2011 we estimate no significant effect of segregation on income disparities between 90th percentiles of blacks and whites (the most well-off). By contrast, whites at the tenth percentile are discovered to have an income 107 percent higher on average than blacks at the tenth percentile. A one-standard-deviation increase in segregation in this setting, widens the gap to $107*(2.727*14) \approx 107*1.38 \approx 148$ percent. Similiarly, the gap between the worst-off blacks and well-off whites increases while the one between the well-off blacks and worst-off whites narrows in response. OLS is again found to be underestimating these effects.

The Falsification columns for both panels check for the presence of a reduced form (direct) effect of the instrument RDI on poverty and Gini in cities that are at least 400 miles away from the South. The lack of significant results here is consistent with our observations from Figure 2 and indicates that no meaningful relationship exists between RDI and the income distribution in cities where RDI wouldn't be influencing segregation. This lends credibility to the instrument.

The results we obtained are fully equivalent with the ones of Ananat, 2011. Our code for this part can be retrieved in Listing 2.

Table 3: The Effects of Segregation on Poverty and Inequality among Blacks and Whites

	Outcome: Gini Index		Outcome: Poverty rate	
	Whites	Blacks	Whites	Blacks
With contols 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**	0.36**
	(0.171)	*(0.547)*	*(0.097)*	*(0.141)*
Education	-0.361**	0.887	-0.162**	0.222
	(0.148)	(0.664)	(0.08)	(0.174)
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)	(0.04)	(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 contols for 1920 city characteristics				
Population 1920	-0.374***	0.899**	-0.214***	0.281**
	(0.106)	(0.442)	(0.071)	(0.115)
Percent Black 1920	-0.364***	0.896**	-0.199***	0.296***
	(0.114)	(0.434)	(0.069)	(0.109)
Literacy	-0.312***	1.029**	-0.163***	0.27**
	(0.107)	(0.47)	(0.061)	(0.124)
Manufacturing 1920	-0.401***	0.904*	-0.213***	0.307**
	(0.132)	(0.483)	(0.081)	(0.122)
Labor Force Participation 1920	-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)

Table 3 presents robustness checks that replicate the models from the third and fourth columns of the top panel of Table 2, while also controlling for a variety of city characteristics. The top panel presents 1990 characteristics, while the bottom focuses on similar ones available in 1920. The majority are expected to exhibit a correlation with both segregation and poverty and inequality. The study's goal is to establish that RDI affects poverty and inequality only through segregation and not any other channels. In our estimation we obtain mostly identical results with some small differences in the 1920 estimates for literacy, employment manufacturing share and labor force participation. This could be due to some missing and unspecified additional controls or small alterations in the data used. Overall, however, significance tests yield equivalent results close to the estimates of Table 2. This provides confidence and credibility to the robustness of the IV estimate.

An interesting addition of a control is the one of a propensity score variable, which is calculated based on all the 1920 (pre-Great Migration) city characteristics. The score measures the probability of having an above-median RDI. The results are again similar to the ones of Table 2, although estimates for the black community become only marginally significant. Our code for the execution of this part can be retrieved in Listing 3.

Table 4: 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	
White (1)	Black (2)	White (3)	Black (4)	White (5)	Black (6)	White (7)	Black (8)
OLS							
-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							
-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							
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: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

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.

Since individuals moving between cities could indicate a selection effect, the author investigates how RDI impacts city choice. The data we have has direct evidence of in-migration for both white and black residence. If the black population increased, we would expect that neighborhoods would become overpopulated and black people would move into other neighborhoods. For segregation to be maintained, we expect that neighborhoods defined by railroads would have increased concentration of black people during the Great Migration. This would affect the housing market and living conditions of people living in those cities.

Table 4 explores the migration, housing, and living conditions by race for the 1990's census. We are using the 1990's as the conclusion of the experiment to measure the ultimate outcomes. In our data, we found all the variables relating to this analysis. The first and last row are standard OLS regressions, first measuring segregation directly and last a falsification test that explores just the impact of railroad construction. The first row only looks at how segregation affects the outcomes. The second row performs 2-stage least regression to get the main results evaluating the instrument variable of RDI. We use track length as a control variable for the instrument. For falsification we look at only the effect of RDI, controlled for by track length, for cities far from the south.

The results of the 2-stage least regression are in the second row of the table. We observe that the more RDI-induced segregation a city has, the less new residents it gains. These cities may be less desirable for people moving, they may still be growing at a similar rate as other cities with higher in-migration due to current residents not moving away. This is where the housing prices suggest that there is less market demand for housing. A final consideration is that the quantity consumed may be different in each city. The last two columns addresses this by looking at how crowded houses are. We see that people living in cities with more RDI-induced segregation live in less crowded houses. This again shows less demand for cities with higher RDI-induced segregation. The author argues that these observation present plausible explanations for selection effects of migration. The interesting result is that demand for these cities is reduced for both white and black people, indicating a distaste for segregation by both groups.

The original paper goes on to test for individual treatment effects of segregation on human capital by focusing on young adults born shortly after the Great Migration. The estimation relies on data from 1980 which we unfortunately couldn't access. Verifying the results of Table 5 from Ananat, 2011 was thus not possible.

4 Conclusion

In conclusion, Ananat, 2011 provides compelling evidence of the causal effects of racial segregation on urban poverty and inequality in the United States. We have managed to successfully reproduce the main results of her study and follow the implementation of her IV design. How she created her models was clear to follow, making it easy to code in R. Overall we find that the use of RDI, controlling for track length, provides a valid instrument for outcomes of segregation. The analysis clearly showcases the RDI's relevancy and conducts numerous tests to its robustness and exclusivity. It seems plausible that the exogeneity condition is also fulfilled.

The results of the paper are interesting as they showcase a clear dynamic in which segregation causes white communities to be better off, while black communities experience higher levels of poverty and inequality. Moreover, they present the idea that segregated cities become less attractive to live in and suggest an almost universal distaste for segregation.

A Appendix: Code snippets

Listing 1: Producing Table 1

```
# Table 1 - Testing RDI as an Instrument ----------------------------
# Divide column by 1,000
data$area1910 <- data$area1910 / 1000
data$count1910 <- data$count1910 / 1000
data$passpc <- data$passpc / 1000

# Dependent Variables
focus_variables1 <- c("dism1990")
false_variables1 <- c("area1910", "count1910", "ethseg10", "ethiso10", "black1910", "passpc", "black1920", "ctyliterate1920", "lfp1920", "ctytrade_wkrs1920", "ctymanuf_wkrs1920", "ctyrail_wkrs1920", "incseg")
vars <- c(focus_variables1, false_variables1)

# Run OLS regressions for each variable
models <- lapply(vars, function(var) lm(as.formula(paste(var, "~ herf + lenper")), data = data))

# Compute robust standard errors
robust_ses <- lapply(models, function(model) coeftest(model, vcov = vcovHC(model, type = "HC1")))

# Compute means
means <- sapply(vars, function(var) round(mean(data[[var]], na.rm = TRUE), 3))

means1 <- c("Mean of Dependent Variable", means[1:7])
names(means1) <- NULL

means2 <- c("Mean of Dependent Variable", means[8:14])
names(means2) <- NULL

# Generate the upper table
stargazer(models[1:7],
          omit = "(Constant)",
          title = "Testing RDI as an Instrument",
          align = TRUE,
          type = "latex",
          se = lapply(robust_ses[1:7], function(se) se[, 2]),
          dep.var.labels.include = FALSE,
          column.labels = c("1990 dissimilarity index", "Physical area (1910)", "Population (1910)", "Ethnic dissimilarity index (1910)", "Ethnic isolation index (1910)", "Percent black (1910)", "Street-cars per capita (1915)"),
          add.lines = list(means1),
          omit.stat = c("rsq", "adj.rsq", "ser", "f")
)
```

```

# Generate the lower table
stargazer(models[8:14],
          omit = "(Constant)",
          title = "Testing RDI as an Instrument",
          align = TRUE,
          type = "latex",
          se = lapply(robust_ses[8:14], function(se) se[,2]),
          dep.var.labels.include = FALSE,
          column.labels = c("Percent black (1920)", "Percent literate (1920)", "Labor force participation (1920)", "Percent in trade (1920)", "Percent in manufacturing (1920)", "Percent in railroads (1920)", "1990 income segregation"),
          add.lines = list(means2),
          omit.stat = c("rsq", "adj.rsq", "ser", "f"),
          notes = c("Robust standard errors in parentheses.", "From Cutler-Glaeser-Vigdor data; sample limited to what that dataset provides.", "Calculated from ipums.org; full sample represented.")
)
```

↑ Continue to Table 2 ↑

Listing 2: Producing Table 2

```r

# Table 2 - The Effects of Segregation on Poverty and Inequality among Blacks and Whites ---------------------
# Define a function to run a model and calculate robust statistics
run_model <- function(formula, data, is_ivreg = FALSE) {
  model <- if (is_ivreg) ivreg(formula, data = data) else lm(formula, data = data)
  coeftest(model, vcov = vcovHC(model, type = "HC1"))
}

# Define formulas for the models
ols_formulas <- list(
  lngini_w ~ dism1990,
  povrate_w ~ dism1990,
  lngini_b ~ dism1990,
  povrate_b ~ dism1990
)
iv_formulas <- list(
  lngini_w ~ dism1990 + lenper | herf + lenper,
  povrate_w ~ dism1990 + lenper | herf + lenper,
  lngini_b ~ dism1990 + lenper | herf + lenper,
  povrate_b ~ dism1990 + lenper | herf + lenper
)
falsification_formulas <- list(
  lngini_w ~ herf + lenper,
  povrate_w ~ herf + lenper,
  lngini_b ~ herf + lenper,
  povrate_b ~ herf + lenper
)

# Run the models and store results in lists
ols_results <- lapply(ols_formulas, function(f) run_model(f, data))
iv_results <- lapply(iv_formulas, function(f) run_model(f, data, is_ivreg = TRUE))
falsification_results <- lapply(falsification_formulas, function(f) run_model(f, data %>% filter(closeness <= -400)))

# Function to format coefficients with significance stars
format_with_stars <- function(coef, pval) {
  if (pval < 0.01) return(paste0(round(coef, 3), "***"))
  if (pval < 0.05) return(paste0(round(coef, 3), "**"))
  if (pval < 0.1) return(paste0(round(coef, 3), "*"))
  return(round(coef, 3))
}

# Extract results from models
results <- data.frame(
  Outcome = c("Gini Index", "", "Poverty Rate", ""),
  Whites_OLS = c(
    format_with_stars(coef(ols_results[[1]])["dism1990"],
```

```
    ols_results[[1]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[1]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(ols_results[[2]])["dism1990"], ols_results[[2]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[2]]["dism1990", "Std. Error"], 3), ")")
  ),
  Blacks_OLS = c(
    format_with_stars(coef(ols_results[[3]])["dism1990"], ols_results[[3]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[3]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(ols_results[[4]])["dism1990"], ols_results[[4]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[4]]["dism1990", "Std. Error"], 3), ")")
  ),
  Whites_2SLS = c(
    format_with_stars(coef(iv_results[[1]])["dism1990"], iv_results[[1]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[1]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(iv_results[[2]])["dism1990"], iv_results[[2]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[2]]["dism1990", "Std. Error"], 3), ")")
  ),
  Blacks_2SLS = c(
    format_with_stars(coef(iv_results[[3]])["dism1990"], iv_results[[3]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[3]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(iv_results[[4]])["dism1990"], iv_results[[4]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[4]]["dism1990", "Std. Error"], 3), ")")
  ),
  Whites_Falsification = c(
    format_with_stars(coef(falsification_results[[1]])["herf"], falsification_results[[1]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[1]]["herf", "Std. Error"], 3), ")"),
    format_with_stars(coef(falsification_results[[2]])["herf"], falsification_results[[2]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[2]]["herf", "Std. Error"], 3), ")")
  ),
  Blacks_Falsification = c(
    format_with_stars(coef(falsification_results[[3]])["herf"], falsification_results[[3]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[3]]["herf", "Std. Error"], 3), ")"),
    format_with_stars(coef(falsification_results[[4]])["herf
```

```
"], falsification_results[[4]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[4]]["herf", "Std. Error"], 3), ")")
  )
)

# Create the kableExtra table
latex_output <- results %>%
  kbl(
    caption = "The Effects of Segregation on Poverty and Inequality among Blacks and Whites",
    col.names = c("Outcome", "Whites", "Blacks", "Whites", "Blacks", "Whites", "Blacks"),
    booktabs = TRUE, format = "latex", align = "lcccccc"
  ) %>%
  add_header_above(c(" " = 1, "OLS" = 2, "2SLS" = 2, "Falsification" = 2)) %>%
  kable_styling(latex_options = c("hold_position", "scale_down"), font_size = 10) %>%
  column_spec(1, bold = TRUE) %>%
  row_spec(c(2, 4), italic = TRUE)

# Export to tex
writeLines(latex_output, "table2_top.tex")

# Bottom part
ols_formulas <- list(
  ln90w90b ~ dism1990,
  ln10w10b ~ dism1990,
  ln90w10b ~ dism1990,
  ln90b10w ~ dism1990
)
iv_formulas <- list(
  ln90w90b ~ dism1990 + lenper | herf + lenper,
  ln10w10b ~ dism1990 + lenper | herf + lenper,
  ln90w10b ~ dism1990 + lenper | herf + lenper,
  ln90b10w ~ dism1990 + lenper | herf + lenper
)
falsification_formulas <- list(
  ln90w90b ~ herf + lenper,
  ln10w10b ~ herf + lenper,
  ln90w10b ~ herf + lenper,
  ln90b10w ~ herf + lenper
)

ols_results <- lapply(ols_formulas, function(f) run_model(f, data))
iv_results <- lapply(iv_formulas, function(f) run_model(f, data, is_ivreg = TRUE))
falsification_results <- lapply(falsification_formulas, function(f) run_model(f, data %>% filter(closeness <= -400)))

# Extract results from models
```

```r
results <- data.frame(
  Outcome = c("90 white: 90 black", "", "10 white: 10 black", "", "90 white: 10 black", "", "10 white: 90 black", ""),
  OLS = c(
    format_with_stars(coef(ols_results[[1]])["dism1990"], ols_results[[1]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[1]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(ols_results[[2]])["dism1990"], ols_results[[2]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[2]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(ols_results[[3]])["dism1990"], ols_results[[3]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[3]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(ols_results[[4]])["dism1990"], ols_results[[4]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(ols_results[[4]]["dism1990", "Std. Error"], 3), ")")
  ),
  `2SLS` = c(
    format_with_stars(coef(iv_results[[1]])["dism1990"], iv_results[[1]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[1]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(iv_results[[2]])["dism1990"], iv_results[[2]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[2]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(iv_results[[3]])["dism1990"], iv_results[[3]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[3]]["dism1990", "Std. Error"], 3), ")"),
    format_with_stars(coef(iv_results[[4]])["dism1990"], iv_results[[4]]["dism1990", "Pr(>|t|)"]),
    paste0("(", round(iv_results[[4]]["dism1990", "Std. Error"], 3), ")")
  ),
  Falsification = c(
    format_with_stars(coef(falsification_results[[1]])["herf"], falsification_results[[1]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[1]]["herf", "Std. Error"], 3), ")"),
    format_with_stars(coef(falsification_results[[2]])["herf"], falsification_results[[2]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[2]]["herf", "Std. Error"], 3), ")"),
    format_with_stars(coef(falsification_results[[3]])["herf"], falsification_results[[3]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[3]]["herf", "Std. Error"], 3), ")"),
    format_with_stars(coef(falsification_results[[4]])["herf
```

```
"], falsification_results[[4]]["herf", "Pr(>|t|)"]),
    paste0("(", round(falsification_results[[4]]["herf", "Std. Error"], 3), ")")
  )
)

# Create the kableExtra table
latex_output <- results %>%
  kbl(
    caption = "The Effects of Segregation on Poverty and Inequality for Various Groups",
    col.names = c("Outcome", "Whites", "Blacks", "Whites", "Blacks", "Whites", "Blacks"),
    booktabs = TRUE, format = "latex", align = "lcccccc"
  ) %>%
  add_header_above(c(" " = 1, "OLS" = 2, "2SLS" = 2, "Falsification" = 2)) %>%
  kable_styling(latex_options = c("hold_position", "scale_down"), font_size = 10) %>%
  column_spec(1, bold = TRUE) %>%
  row_spec(c(2, 4), italic = TRUE)

# Export to tex
writeLines(latex_output, "table2_bottom.tex")
```

↑ continue to Table 3 ↑

Listing 3: Producing Table 3

```
# Table 3 -------------------------------

# Population
population_formulas <- list(
  "lngini_w ~ dism1990 + lenper + I(pop1990/1000) | herf + lenper + I(pop1990/1000)",
  "lngini_b ~ dism1990 + lenper + I(pop1990/1000) | herf + lenper + I(pop1990/1000)",
  "povrate_w ~ dism1990 + lenper + I(pop1990/1000) | herf + lenper + I(pop1990/1000)",
  "povrate_b ~ dism1990 + lenper + I(pop1990/1000) | herf + lenper + I(pop1990/1000)"
)

# Percent black
percent_black_formulas <- list(
  "lngini_w ~ dism1990 + lenper + pctbk1990 | herf + lenper + pctbk1990",
  "lngini_b ~ dism1990 + lenper + pctbk1990 | herf + lenper + pctbk1990",
  "povrate_w ~ dism1990 + lenper + pctbk1990 | herf + lenper + pctbk1990",
  "povrate_b ~ dism1990 + lenper + pctbk1990 | herf + lenper + pctbk1990"
)

# Education
education_formulas <- list(
  "lngini_w ~ dism1990 + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b | herf + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b",
  "lngini_b ~ dism1990 + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b | herf + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b",
  "povrate_w ~ dism1990 + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b | herf + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b",
  "povrate_b ~ dism1990 + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b | herf + lenper + hsdrop_w + hsdrop_b + hsgrad_w + hsgrad_b + somecoll_w + somecoll_b + collgrad_w + collgrad_b"
)

# Share employed in manufacturing
manufacturing_formulas <- list(
  "lngini_w ~ dism1990 + lenper + manshr | herf + lenper +
```

```
    manshr",
  "lngini_b ~ dism1990 + lenper + manshr | herf + lenper + manshr",
  "povrate_w ~ dism1990 + lenper + manshr | herf + lenper + manshr",
  "povrate_b ~ dism1990 + lenper + manshr | herf + lenper + manshr"
)

# Labor force participation
lfp_formulas <- list(
  "lngini_w ~ dism1990 + lenper + lfp_w + lfp_b | herf + lenper + lfp_w + lfp_b",
  "lngini_b ~ dism1990 + lenper + lfp_w + lfp_b | herf + lenper + lfp_w + lfp_b",
  "povrate_w ~ dism1990 + lenper + lfp_w + lfp_b | herf + lenper + lfp_w + lfp_b",
  "povrate_b ~ dism1990 + lenper + lfp_w + lfp_b | herf + lenper + lfp_w + lfp_b"
)

# Number of local governments
local_governments_formulas <- list(
  "lngini_w ~ dism1990 + lenper + ngov62 | herf + lenper + ngov62",
  "lngini_b ~ dism1990 + lenper + ngov62 | herf + lenper + ngov62",
  "povrate_w ~ dism1990 + lenper + ngov62 | herf + lenper + ngov62",
  "povrate_b ~ dism1990 + lenper + ngov62 | herf + lenper + ngov62"
)

# 1920 Controls
# Population
population_1920_formulas <- list(
  "lngini_w ~ dism1990 + lenper + count1920 | herf + lenper + count1920",
  "lngini_b ~ dism1990 + lenper + count1920 | herf + lenper + count1920",
  "povrate_w ~ dism1990 + lenper + count1920 | herf + lenper + count1920",
  "povrate_b ~ dism1990 + lenper + count1920 | herf + lenper + count1920"
)

# Percent black 1920
percent_black_1920_formulas <- list(
  "lngini_w ~ dism1990 + lenper + black1920 | herf + lenper + black1920",
  "lngini_b ~ dism1990 + lenper + black1920 | herf + lenper + black1920",
  "povrate_w ~ dism1990 + lenper + black1920 | herf + lenper + black1920",
```

```
  "povrate_b ~ dism1990 + lenper + black1920 | herf + lenper
      + black1920"
)

# Literacy
literacy_formulas <- list(
  "lngini_w ~ dism1990 + lenper + ctyliterate1920 | herf +
     lenper + ctyliterate1920",
  "lngini_b ~ dism1990 + lenper + ctyliterate1920 | herf +
     lenper + ctyliterate1920",
  "povrate_w ~ dism1990 + lenper + ctyliterate1920 | herf +
     lenper + ctyliterate1920",
  "povrate_b ~ dism1990 + lenper + ctyliterate1920 | herf +
     lenper + ctyliterate1920"
)

# Share employed in manufacturing 1920
manufacturing_1920_formulas <- list(
  "lngini_w ~ dism1990 + lenper + ctymanuf_wkrs1920 | herf +
      lenper + ctymanuf_wkrs1920",
  "lngini_b ~ dism1990 + lenper + ctymanuf_wkrs1920 | herf +
      lenper + ctymanuf_wkrs1920",
  "povrate_w ~ dism1990 + lenper + ctymanuf_wkrs1920 | herf
     + lenper + ctymanuf_wkrs1920",
  "povrate_b ~ dism1990 + lenper + ctymanuf_wkrs1920 | herf
     + lenper + ctymanuf_wkrs1920"
)

# Labor force participation 1920
lfp_1920_formulas <- list(
  "lngini_w ~ dism1990 + lenper + lfp1920 | herf + lenper +
     lfp1920",
  "lngini_b ~ dism1990 + lenper + lfp1920 | herf + lenper +
     lfp1920",
  "povrate_w ~ dism1990 + lenper + lfp1920 | herf + lenper +
      lfp1920",
  "povrate_b ~ dism1990 + lenper + lfp1920 | herf + lenper +
      lfp1920"
)

# Control for propensity score
propensity_score_formulas <- list(
  "lngini_w ~ dism1990 + lenper + herfscore | herf + lenper
     + herfscore",
  "lngini_b ~ dism1990 + lenper + herfscore | herf + lenper
     + herfscore",
  "povrate_w ~ dism1990 + lenper + herfscore | herf + lenper
      + herfscore",
  "povrate_b ~ dism1990 + lenper + herfscore | herf + lenper
      + herfscore"
)


# Create a list of formula categories
```

```
formula_lists <- list(
  population_formulas,
  percent_black_formulas,
  education_formulas,
  manufacturing_formulas,
  lfp_formulas,
  local_governments_formulas,
  population_1920_formulas,
  percent_black_1920_formulas,
  literacy_formulas,
  manufacturing_1920_formulas,
  lfp_1920_formulas,
  propensity_score_formulas
)

# List names for easy referencing
list_names <- c(
  "population_results",
  "percent_black_results",
  "education_results",
  "manufacturing_results",
  "lfp_results",
  "local_governments_results",
  "population_1920_results",
  "percent_black_1920_results",
  "literacy_results",
  "manufacturing_1920_results",
  "lfp_1920_results",
  "propensity_score_results"
)

# Apply run_model to each formula list and store results in
    named variables
results <- setNames(
  lapply(formula_lists, function(f) lapply(f, function(
      formula) run_model(formula, data, is_ivreg = TRUE))),
  list_names
)

attach(results)

# Use a loop and lapply to generate the vector for each
    result list
result_vectors <- list()

for (i in 1:4) {
  result_vectors[[paste0("N", i)]] <- unlist(lapply(list_
      names, function(result_name) {
    coef_data <- coef(get(result_name)[[i]])["dism1990"]
    p_value <- get(result_name)[[i]]["dism1990", "Pr(>|t|)"]
    std_error <- get(result_name)[[i]]["dism1990", "Std.
        Error"]

    c(
```

```
      format_with_stars(coef_data, p_value),
      paste0("(", round(std_error, 3), ")")
    )
  }))
}

# Access each result vector
result_vectors$N1
result_vectors$N2
result_vectors$N3
result_vectors$N4

# Extract results from models
results <- data.frame(
  Outcome = c(
    "Population", "",
    "Percent Black", "",
    "Education", "",
    "Manufacturing", "",
    "Labor Force Participation", "",
    "Number of Local Governments", "",
    "Population 1920", "",
    "Percent Black 1920", "",
    "Literacy", "",
    "Manufacturing 1920", "",
    "Labor Force Participation 1920", "",
    "Propensity Score", ""
  )
  ,
  Whites_Gini = result_vectors$N1,
  Blacks_Gini = result_vectors$N2,
  Whites_Poverty = result_vectors$N3,
  Blacks_Poverty = result_vectors$N4
)


# Create the kableExtra table
latex_output <- results %>%
  kbl(
    caption = "The Effects of Segregation on Poverty and Inequality among Blacks and Whites",
    col.names = c("",
                  "Whites", "Blacks",
                  "Whites", "Blacks"),
    booktabs = TRUE,  # Use LaTeX booktabs styling
    format = "latex",
    align = "lcccccc" # Align columns (left for Outcome, center for others)
  ) %>%
  add_header_above(
    c(" " = 1, "Outcome: Gini Index" = 2, "Outcome: Poverty rate" = 2) # Group column headers
  ) %>%
  kable_styling(
```

```
    latex_options = c("hold_position", "scale_down"), # Adjust table to fit
    font_size = 10                                    # Optional: Adjust font size
  ) %>%
  column_spec(1, bold = TRUE) %>%                     # Make the first column bold
  row_spec(c(2, 4), italic = TRUE)                    # Italicize rows for standard errors

# Export to tex
writeLines(latex_output, "table3.tex")
```

↑ Continue to Table 4 ↑