Writing Draft

Maria Burzillo

3/28/2020

Abstract

Trounstine (2016) suggests that high levels of residential segregation are associated with increased political polarization and decreased public spending. In this analysis, I was able to successfully replicate Trounstine (2016)'s main results. Adding onto her original analysis, I impute a large amount of missing data from the original dataset and re-run the analysis. Finally, I add in income segregation as a predictor in her regressions to assess whether this could be a confounding variable.

Introduction

There is a large degree of variation in public goods spending across local governments. As a result, many scholars have worked to determine what factors may lead to the underprovision of public goods spending. Some scholars in the past have associated racial diversity or changes in levels of diversity with the under-provision of public goods (Baqir, Easterly, and Alesina 1999; Hopkins 2009). However, Trounstine (2016) argues that it is racial segregation, not diversity in and of itself, that results in the under-provision of public goods. Trounstine's analysis consists of two main parts. First, she uses election and demographic data from 25 of America's largest cities between 1990 and 2010 to run a multilevel mixed-effects linear regression with fixed effects for region and year and with random effects for cities in order to show that polarization increases with segregation. Trounstine measures segregation with Theil's H index, which measures the degree to which the diversity of a neighborhood differs from the diversity of the entire city. The second main part of Trounstine's analysis looks at the ability of the Theil's H segregation index to explain a variety of types of public expenditures by city using a sample of 2,637 cities with 13,742 city-year observations. Using linear regressions with fixed effects for cities and robust standard errors clustered by city, Trounstine finds a significant, negative correlation between segregation and public goods spending that is robust to the inclusion of a variety of relevant controls and an alternative specification in which the number of waterways is used as an instrument for segregation.

In this analysis, I first work to replicate the main results of Trounstine (2016) using R statistical software (R Core Team 2019). The original data and Stata code made publically available by the author were downloaded via the Harvard Dataverse (Trounstine 2015). I also make all of my code and analysis available on Github.[1] I was successfully able to replicate the main results of Trounstine (2016) in R with the exception of some of the marginal analyses, which I was nevertheless able to replicate in Stata.

One concern with the original analysis in Trounstine (2016) is the large amount of missing data, which substantially constrains the sample size used in the regression analysis. For example, the regression analysis in the main specification using the the original racial polarization exludes %0.5517241 of the observations in the original dataset, and the regression analysis in the main specification using the original financial segregation data excludes %0.9576097 of observations from the original dataset. As an extension of Trounstine (2016), I impute missing values in the original data using the mice package in R, which generates multivariate

[1] Link to my Github repository for this project.

imputations using chained equations (van Buuren and Groothuis-Oudshoorn 2011). Then, I use the imputed datasets to re-estimate the original models, pooling the results to final pooled regression coefficients and parameters.

Comparing the results of the original regressions and those done with the imputed data yields similar big picture results in terms of the direction of the signs on the coefficients on the segregation indices. Like in Trounstine (2016), I find that segregation is positively associated with political polarization and negatively associated with spending on public goods. However, the magnitude of the effects in most specifications has dminished and most results become statistically insignificant. While these findings do not necessarily challenge the results of Trounstine (2016), they do call into question the relative importance of segregation in determining public goods spending and political polarization and suggest that the results of Trounstine (2016) may not be quite as robust as once thought.

Literature Review

Despite some progress made towards racial equity in the U.S. on other fronts, residential racial segregation in U.S. neighborhoods continues to be prevasive and deeply entrenched in society (Fischer et al. 2004; Oliver 2010; Massey 1993). Research suggests that this kind of segregation has political consequences, as political cleavages in segregated cities tend to have racial as well as spatial dimensions (Massey 1993). Neighborhoods are often important actors within local politics because local governments provide many functions that are allocational in nature and concern geographical space (Trounstine 2016). Thus, when neighborhoods are divided on racial lines as well as spacial lines, it is natural to expect higher degrees of racial polarization as a result.

Studying residential segregation is difficult because its effects tend to differ by geographic level. On the neighborhood level, the kind of geographic racial isolation brought on by residential segregation has been associated with racial intolerance, resentment, and competition between racial groups (Oliver 2010). Living within segregated neighborhoods has also been associated with holding negative stereotypes and perceptions about out groups (Eric Oliver and Wong 2003). As a result, homogeneous neighborhoods have been associated with increased racial tension and political polarization in comparison to integrated, diverse neighborhoods. However, at the city or metropolitan level, the opposite seems to be true: when considering larger geographic areas, diversity and integration are correlated with racial tension, competition, prejudice, lower levels of cooperation, and lower spending on public goods (Oliver 2010; Baqir, Easterly, and Alesina 1999; Hopkins 2009). While these differences in the expected effect of segregation on the geographic level may seem confusing at first, they make sense as they suggest that the most severely segregated area is one that is diverse overall, but has many homogeneous neighborhoods. Thus, while people of different races co-exist within a highly segregated city, they live separately within their own neighborhoods, which creates an environment ripe for racial antagonism (Trounstine 2016). It is thus not simply the level of diversity or integration that matters for racial harmony and cooperation, but their patterns within a larger geographic framework (Trounstine 2016; Oliver 2010; Bharathi et al. 2018).

Political polarization along racial lines may lead to decreased public spending and goods provision because groups may have different preferences, which can make compromise hard, and because groups may preceive a disutility in out-groups receiving public goods expenditure Baqir, Easterly, and Alesina (1999). Einstein (2012) found evidence that racial segregation predicts large political divisions at the metropolitan level and that these divisions can create a lack of willingness to compromise and collaborate on local policy problems. Trounstine (2016) finds similar results at the city level: that residential racial segregation is associated with both increased political division and decreased public spending. Thus, these authors suggest that it is the combination of homogenous neighborhoods within a much larger, diverse geographic area that leads to increased political polarization and reduced public goods spending in local governments.

More recently, some scholars have called this hypothesis and its importance into question. For example, Lee (2018) finds evidence that larger inequalities within the political system favoring socially powerful groups, not local diversity patterns leading to decreased cooperation, may be a better explanation of failures in public goods provision in diverse areas. Other suggests that additional factors, such as income segregation, may be important confounding factors in public goods provision. An, Levy, and Hero (2018), for example, suggests that the more closely related income inequality is to racial inequality, the less investment is made in public goods, and that this interaciton was a better predictor of public goods spending patterns than measures of diversity. inequality and diversity jointly. A variety of evidence suggests that it is meaningful to consider the effects of income inequaltiy and diversity and segregation jointly (An, Levy, and Hero 2018; Massey 1993). Given the active debate in the literature over the relaitonships between diversity, segregation, public spending, and other factors, it is increasingly important to re-examine previously reported findings as a means of robustness checks.

Replication

For the most part, results from Trounstine (2016) were successfully replicated. All regressions and tables were fully replicated in R. However, I was unable to successfully replicate Trounstine (2016)'s marginal effects analyses and margins plots using R. There does not yet appear to be a built-in R function to calculate marginal or predicted effects or to generate margins plots from the complicated multi-level models employed in the original paper, and creating such a function was outside of the scope of this analysis. Nevertheless, these results were successfully replicated in Stata.

There was one interesting outcome from my attempt to replicate the original Stata code in R. Due to the differences in R and Stata in dealing with missing values, my first replication of analyses using the financial segregation dataset were slightly different than the results in Stata from the original paper. In order to exclude cities from her analysis with only one census tract, Trounstine conditions her regressions in Stata such that the value for the number of census tracts is greater than one. In Stata, this does not remove missing values, whereas in R, it does. Since this variable is used only as a conditional filter and not as a regression variable (and thus, the observations with missing values for number of census tracts are not dropped), Trounstine's analysis includes 14 cities and 58 observations with missing values for census tracts in addition to cities with two or more census tracts. This is a potential oversight on the part of the author, and I would suggest also dropping observations with missing census tract data, or else imputing them. Nevertheless, dropping these values did not have much of an effect on the subsequent results.

Extension

One concern with the original analysis in Trounstine (2016) is the large amount of missing data values in her original datasets. Because both R and Stata drop any observations with missing values for any of the variables used in a regression, this can exlcude a large portion of the data from the analysis and potentially bias the results if the data is not missing completely at random. A large amount of data is missing in both of the main datasets used in the analyses of Trounstine (2016). The main model specification using the racial polarization dataset exludes 112 of the original 203 observations, or %55.17 of the data. Additionally, the main model specification using the financial segregation data set excludes 3.10436×10^5 of the original 324178 observations, or %95.76 of observations from the original dataset.

In an attempt to better deal with the problem of missing data in Trounstine (2016), I impute missing values in the original data using the mice package in R, which generates multivariate imputations using chained equations (van Buuren and Groothuis-Oudshoorn 2011). While there are a variety of different imputation methods that could have been employed, multiple imputation (such as the multiple multivariate imputations generated by mice) is desireable because instead of inputting a single value such as the mean for missing values, it instead uses the distribution of the available data to estimate multiple potential values for the missing data. As a result, multiple imputation helps to account for the uncertainty inherent in the imputation process and allows for the calculation of standard errors around estimaotrs. As a result, multiple imputation allows the researcher to more accurately assess the of uncertainty in the analysis in general.

Before performing the multiple imputations on the datasets, I first examined the missing data for patterns. To better understand any potential patterns in missing data, I plotted the pattern of missingness for only and created a histogram showing the frequency of missing values for those variables with missing values in figure X1 for the racial polarization data and figure X2 for the financial segregation data. Looking at figure X1, the histogram shows that the variables for the largest vote split along racial lines and the variable indicating whether or not the election had candidates of more than one race had by far the largest percentages of missing data, missing %53.69 and %50.74, respectively. It is important to note that the variable for the largest vote split along racial lines is the dependent variable in the regresison analysis, and thus, we are missing a large percentage of this key variable. From the plot on the right, we can see that approximately 42% of observations are complete. There seem to be a correspondence between missing a value for biggest split and missing biracial. There also seem to be about 2% of values for wich most of the variables are missing. However, most observations are not missing more than 2-3 values.

In figure X2, we can similarly observe the trends for the financial segregation dataset.

Then, I use the imputed datasets to re-estimate the original models, pooling the results to final pooled regression coefficients and parameters.

One aspect of Trounstine's paper with room for improvement is that there is a large amount of missing data in her datasets upon which she bases her analyses. Using R's mice package (add citation), we can perform multiple imputation. By performing imputation multiple times, this helps account for the uncertainty inherent in the individual imputations. Before performing the multiple imputations, we will first look at the missing data to see if there are any patterns.

for all of the variables in the racial polarization data used in the analysis and for the main independent variable, the segregation index, used in the analysis with the financial segregation dataset. More of the data could not be imputed from the financial segregation dataset due to computing and time constraints for this project.

Bibliography

An, Brian, Morris Levy, and Rodney Hero. 2018. "It's Not Just Welfare: Racial Inequality and the Local Provision of Public Goods in the United States." *Urban Affairs Review* 54 (5): 833–65. https://doi.org/10.1177/1078087417752476.

Baqir, Reza, William Easterly, and Alberto Alesina. 1999. "Public Goods and Ethnic Divisions" 114 (4): 1243–84. https://doi.org/10.1162/003355399556269.

Bharathi, N. Manjula, Deepak V. Malghan, Sumit Mishra, and Andaleeb Rahman. 2018. "Public Goods, and Nested Subnational Units: Diversity, Segregation, or Hierarchy?" In. https://doi.org/10.2139/ssrn.3254095.

Einstein, Katherine Levine (Ph D. 2012. "Divided Regions: Race, Political Segregation, and the Fragmentation of American Metropolitan Policy." PhD thesis.

Eric Oliver, J., and Janelle Wong. 2003. "Intergroup Prejudice in Multiethnic Settings." *American Journal of Political Science* 47 (4): 567–82. https://doi.org/10.1111/1540-5907.00040.

Fischer, Claude, Gretchen Stockmayer, Jon Stiles, and Michael Hout. 2004. "Distinguishing the Geographic Levels and Social Dimensions of U.S. Metropolitan Segregation, 1960–2000." *Demography* 41 (1): 37–59. https://doi.org/10.1353/dem.2004.0002.

Hopkins, Daniel J. 2009. "The Diversity Discount: When Increasing Ethnic and Racial Diversity Prevents Tax Increases." *The Journal of Politics* 71 (1): 160–77. https://doi.org/10.1017/S0022381608090105.

Lee, Alexander. 2018. "Ethnic Diversity and Ethnic Discrimination: Explaining Local Public Goods Provision." *Comparative Political Studies* 51 (10): 1351–83. https://doi.org/10.1177/0010414017740604.

Massey, Douglas S. 1993. *American Apartheid: Segregation and the Making of the Underclass.* Democracy and Urban Landscapes. Cambridge, Mass.: Harvard University Press. http://nrs.harvard.edu/urn-3:hul.ebookbatch.ACLS_batch:20170426MIU01100000000000000000544.

Oliver, J. Eric. 2010. *The Paradoxes of Integration:Race, Neighborhood, and Civic Life in Multiethnic America.* Chicago ; London: University of Chicago Press.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Trounstine, Jessica. 2015. "Replication Data for: Segregation and Inequality in Public Goods." Harvard Dataverse. https://doi.org/10.7910/DVN/4LZXTY.

———. 2016. "Segregation and Inequality in Public Goods: SEGREGATION AND INEQUALITY IN PUBLIC GOODS." *American Journal of Political Science* 60 (3): 709–25. https://doi.org/10.1111/ajps.12227.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (3): 1–67. https://www.jstatsoft.org/v45/i03/.