Does segregation matter for Latinos?

Jorge De la Roca[*‡]
University of Southern California

Ingrid Gould Ellen[*§]
New York University

Justin Steil[*¶]
Massachusetts Institute of Technology

This version, October 2017

Abstract: We estimate the effects of residential racial segregation on socio-economic outcomes for native-born Latino young adults over the past three decades. Using individual public use micro-data samples from the Census and a novel instrumental variable, we find that higher levels of metropolitan area segregation have negative effects on Latino young adults' likelihood of being either employed or in school, on the likelihood of working in a professional occupation, and on income. The negative effects of segregation are somewhat larger for Latinos than for African Americans. Controlling for Latino and white exposure to neighborhood poverty, neighbors with college degrees, and industries that saw large increases in high-skill employment explains between one half and two thirds of the association between Latino-white segregation and Latino-white gaps in outcomes.

Key words: racial segregation, Hispanics/Latinos, spatial inequality
JEL classification: J15, R23

[*] We thank Morgane Laouenan and Jonathan L. Rothbaum for valuable comments and discussions. We also thank Maxwell Austensen, Gerard Torrats Espinosa, and Justin Tyndall for their exceptional research assistance.

[‡] Corresponding author. Sol Price School of Public Policy, University of Southern California, 650 Childs Way RGL 326, Los Angeles, CA 90089, USA (e-mail: `jdelaroc@usc.edu`; website: `http://jorgedelaroca.name`).

[§] Robert F. Wagner Graduate School of Public Service, New York University, 295 Lafayette Street, New York, NY 10012, (e-mail: `ingrid.ellen@nyu.edu`; website: `https://wagner.nyu.edu/community/faculty/ingrid-gould-ellen`).

[¶] Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 9–515, Cambridge, MA 02139 (e-mail: `steil@mit.edu`; website: `https://steil.mit.edu/`).

1. Introduction

Between 1990 and 2010, the Latino population in the United States more than doubled, from 22.4 million to 50.5 million. As the Latino population has grown, levels of Latino-white residential segregation (as measured by the dissimilarity index) have remained relatively steady (at around 0.50), while levels of Latino isolation have risen (from 0.43 in 1990 to 0.46 in 2010) (De la Roca, Ellen, and O'Regan, 2014).[1] Despite this durable residential segregation, there has been little exploration of how that segregation affects the socio-economic outcomes of Latinos.

While existing research has found that black-white segregation negatively affects socio-economic outcomes for African Americans (e.g. Cutler and Glaeser, 1997, Ellen, 2000, Card and Rothstein, 2007), there are reasons to expect that segregation may not have the same negative consequences for Latinos. For instance, research on ethnic enclaves has suggested that ethnic concentration, in some circumstances, can improve employment outcomes by creating a market for ethnic goods and access to co-ethnic sources of capital (Portes and Sensenbrenner, 1993, Edin, Fredriksson, and Åslund, 2003, Cutler, Glaeser, and Vigdor, 2008). Residential segregation may still undermine the socio-economic outcomes of Latinos, however, through the same mechanisms that have been suggested to limit opportunities for blacks, by constraining Latinos to live in neighborhoods with less public investment, lower levels of human capital, or limited access to particular jobs and job networks (Kain, 1968, Loury, 1977, Borjas, 1995, Lou and Song, 2017).

Thus, we examine how levels of residential segregation affect the educational and labor market outcomes of Latino young adults and how those effects differ from the effects of segregation on the outcomes of black young adults. To address concerns regarding within-city sorting, we examine how metropolitan-level segregation affects the outcomes of individuals living anywhere in the metropolitan area. To mitigate bias from across-city sorting, we restrict our sample to native-born young adults and use the segregation level of the metropolitan area where they lived five years earlier, lag our measurement of segregation by ten years, estimate longitudinal models with metropolitan area fixed effects, and focus on variation in effects between Latino and white residents of the same metropolitan area, differencing out any residual unobserved attributes of the metropolitan area that may be related to segregation and affect outcomes. Finally, we also employ instrumental variables.

Specifically, we use a new instrument to predict Latino-white segregation, which captures the evenness of the distribution of single-family detached houses in relation to other types of housing in the metropolitan housing stock in 1970. The assumption is that the historical separation of single-family detached homes from other types of dwellings, such as attached homes or multi-family buildings, contributes to contemporary levels of Latino metropolitan area segregation because Latinos are less likely to live in detached, single-family homes than other types of housing (Weicher and Thibodeau, 1988, Brueckner and Rosenthal, 2009). This instrument is more predictive of Latino-white segregation than instruments that have been used for black-white segregation.

[1] Levels of black-white segregation over the same period declined somewhat (from a dissimilarity score of 0.68 to 0.59) but remained high. Levels of black residential isolation also declined, but remained high (declining from 0.55 to 0.46).

Using public-use decennial census data for 1990 and 2000 and data from the American Community Survey for 2007-2011, we examine how metropolitan area levels of segregation affect college graduation rates, employment rates, the likelihood of being in a professional occupation, and income for native-born Latino and African-American young adults between the ages of 25 and 30. The estimates from longitudinal models with metropolitan area fixed effects show that segregation is negatively associated with each of the measured socio-economic outcomes of both Latino and African-American young adults relative to whites.

These results, however, mask substantial heterogeneity in the link between segregation and outcomes for Latino groups of different ancestry and class status. Controlling for the heterogeneous experiences of different Latino ancestry groups, we find that segregation has a significant negative association with socio-economic outcomes for those who identify their ancestry in Mexico, South America, Central America, the Dominican Republic, or Puerto Rico, but not for those who identify as Cuban or of 'another Hispanic origin.'

The instrumental variable results add a more robust causal analysis and confirm that segregation has a negative effect on Latino young adults' likelihood of being employed or in school, on their likelihood of working in a professional occupation, and their income. Segregation widens the gaps in outcomes between Latinos and whites: in 2010, a one standard deviation increase in the metropolitan area level of segregation is associated with a decrease for Latinos relative to whites of 8 percentage points in college graduation rates and 15 percent in income, equivalent to a $4,219 annual income loss. The instrumental variable results also indicate that the wider gaps in socio-economic outcomes in more segregated metropolitan areas are driven in part by the fact that whites in those areas fare better than those in less segregated areas.

To understand why segregation has these effects, we examine potential mechanisms. We find that the exposure of white and Latino residents to neighborhood poverty, neighbors with college degrees, and high-employment growth industries between 1990–2010 together explain between one half and two thirds of the association between segregation and white-Latino gaps in outcomes.

2. Theoretical framework and hypotheses

The effects of residential segregation are theoretically ambiguous and have been found to vary significantly across groups and contexts. Residential segregation shapes access to neighborhoods, which in turn shape access to institutions, peers, and social networks, as well as exposure to crime and environmental benefits and hazards (Durlauf, 2004, Bayer, Ross, and Topa, 2008, Epple and Romano, 2011, Ludwig, Sanbonmatsu, Gennetian, Adam, Duncan, Katz, Kessler, Kling, Lindau, Whitaker, and McDade, 2011, Graham, 2016). But the resources and opportunities that racially or ethnically homogenous neighborhoods provide are likely to vary depending on the socio-economic attributes of the group. In general, groups with greater economic or other resources may benefit from segregation while those with fewer resources may be harmed.

Several studies have found that for immigrant groups with higher mean levels of human capital, ethnic concentration is associated with better outcomes in employment and earnings, while for groups with lower mean levels of human capital, segregation is linked to lesser benefits or negative effects (Borjas, 1995, Edin, Fredriksson, and Åslund, 2003, Cutler, Glaeser, and Vigdor, 2008). Human capital levels have been found to shape the effects of segregation for native-born blacks as well. For instance, increases in the proportion of college-educated African Americans in the metropolitan area reduce the negative effects of segregation on black youths' educational attainment (Bayer, Fang, and McMillan, 2014).

The average financial and political capital of a group also matters. Racial and ethnic groups with lower levels of financial and political capital may be less able to demand equal access to crucial municipal services, like school investment and community policing, to non-profit institutions that provide services and networks, and to private businesses that meet daily needs like child-care (Collins and Williams, 1999). Perhaps even more critically, violence tends to be disproportionately concentrated in low-income neighborhoods and even indirect exposure to neighborhood violence diminishes academic performance (Sharkey, Schwartz, Ellen, and Lacoe, 2014).

Latinos in the United States have lower than average levels of education and income, and arguably less political clout given lower citizenship rates than whites, which may translate into inferior neighborhood services and environmental amenities. Indeed, available measures of differences in neighborhood characteristics find that Latinos in more segregated metropolitan areas are exposed to fewer college educated neighbors, lower performing schools, and higher levels of violent crime than Latinos in less segregated cities (Steil, De la Roca, and Ellen, 2015).

There is of course, considerable variation in the socio-economic backgrounds of different Latino sub-groups in the United States. In 2010, nearly two thirds (63 percent) of the US Latino population identified as having Mexican ancestry, while 9 percent reported Puerto Rican, 8 percent Central American, 6 percent South American, 4 percent Cuban, 3 percent Dominican, and 8 percent 'another Hispanic origin' (United States Bureau of the Census, 2010). Mean educational attainment varies significantly by self-identified group of origin. For instance, 36 percent of Latinos in the United States who were 25 years and over and identified as having South American ancestry had a college degree or higher in 2013 while only 20 percent of those identifying Puerto Rican origins, 14 percent of those identifying Central American origins, and 11 percent of those identifying Mexican origins had college degrees. There is similar heterogeneity with regard to childhood poverty. In 2012, more than a third of those under 18 years of age with Puerto Rican (38 percent), Central American (36 percent), and Mexican (35 percent) origins lived below the poverty line compared to 22 percent of those of Cuban descent, and 20 percent of those of South American descent (United States Bureau of the Census, 2013). This heterogeneity of Latino experiences by ancestry is likely to contribute to variation in the effects of segregation.

While segregation's effects may vary across groups, they are also likely to vary over time. For example, the negative effects of black-white residential segregation on black educational attainment and employment rates did not emerge until the economic restructuring and dramatic neighborhood change of the 1970s (Collins and Margo, 2000). There are reasons to believe that the effects of segregation on Latinos may differ over time as well. For instance, as the Latino population in the United States has grown and Latinos have settled across a larger set of smaller metropolitan areas, the differences in neighborhood environments enjoyed by Latinos in high and low segregation areas may have diminished.

3. Data and methods

To examine how metropolitan area segregation affects individual socio-economic outcomes, we use public-use micro data gathered by the US Census and provided by IPUMS–USA of the University of Minnesota Population Center (Ruggles, Genadek, Goeken, Grover, and Sobek, 2015). We focus our analysis on data from the Decennial Censuses 5% samples in 1990 and 2000 and from the American Community Survey (ACS) 5-year estimates (2007-2011) to study the relationship between residential segregation and socio-economic outcomes of native-born Latinos between the ages of 25 and 30.[2]

We consider educational outcomes such as the probability of college graduation and labor market outcomes such as the probability of working in a professional occupation, income and the likelihood of being employed or in school. We focus on young adults because their metropolitan area of residence is more likely to be affected by parental location choices than that of older adults. In order to most accurately estimate the level of segregation to which an individual was exposed while growing up, we lag our segregation measures by 10 years and use the level of segregation in the metropolitan area in which the individual lived five years prior, for the 1990 and 2000 Census, and one year prior, for the 2007-2011 ACS.[3] We exclude the foreign born because the data do not provide precise information on their year of arrival and, hence, we cannot tell how long they have experienced segregation.

Our sample includes individuals living in 187 Core Based Statistical Areas (CBSAs) across the United States with a total population greater than 100,000 residents and a Latino population of at least 5,000 residents in 2010 (see appendix A for a detailed explanation on the assignment of individuals in IPUMS to CBSAs in each decade).[4] Throughout the study, we use the metropolitan area dissimilarity index from US2010, a joint project between the Russell Sage Foundation and Brown University, as our primary measure of Latino-white residential segregation.

Table 1 presents raw differences in socio-economic outcomes, pooled across 1990, 2000, and 2010, by quartile of metropolitan area segregation. The upper panel shows segregation quartiles based on the 2000 Latino-white dissimilarity index and the lower panel shows quartiles based on the 2000 black-white dissimilarity index. Higher levels of segregation are consistently associated with larger gaps in every outcome between whites and blacks and between whites and Latinos. Notably, the link between segregation and racial differences in outcomes appears to be driven both by better white outcomes and by worse black and Latino outcomes in more segregated areas.

Although these raw means by segregation quartile suggest a relationship between segregation and outcomes, determining how the level of segregation shapes individual socio-economic

[2]Selective ethnic attrition may produce some bias in estimates of Latino educational and labor market outcomes (Duncan and Trejo, 2011); however, the ACS does not allow us to control for immigrant generation or identify ethnicity other than through respondents' self-reporting.

[3]We drop individuals in the armed forces and those living in group quarters, and we also estimate robustness tests that exclude all those who recently moved across metropolitan areas.

[4]When looking at how metropolitan area segregation affects black young adults, our sample includes individuals living in 184 Core Based Statistical Areas with a total population greater than 100,000 residents and a black population of at least 5,000 residents in 2010.

Table 1: Relationship between segregation and outcomes, 1990–2010

	College graduation	Not idle	Professional occupation	Log earnings
	(1)	(2)	(3)	(4)
Whites				
All metropolitan areas	35.4%	89.4%	30.9%	9.94
Low segregation	30.3%	87.9%	26.7%	9.81
Moderate segregation	32.2%	88.9%	28.7%	9.87
High segregation	33.3%	89.3%	29.6%	9.92
Very high segregation	38.7%	89.9%	33.1%	10.01
Latinos				
All metropolitan areas	15.9%	84.1%	18.6%	9.77
Low segregation	21.1%	86.4%	21.0%	9.75
Moderate segregation	16.0%	84.8%	18.6%	9.70
High segregation	14.6%	84.5%	17.9%	9.73
Very high segregation	16.2%	83.7%	18.7%	9.79
White-Latino gap				
All metropolitan areas	19.5%	5.3%	12.3%	0.17
Low segregation	9.2%	1.5%	5.8%	0.06
Moderate segregation	16.2%	4.1%	10.1%	0.17
High segregation	18.7%	4.8%	11.7%	0.19
Very high segregation	22.5%	6.2%	14.5%	0.21
Whites				
All metropolitan areas	35.4%	89.4%	30.9%	9.94
Low segregation	28.4%	87.8%	26.9%	9.87
Moderate segregation	32.6%	88.8%	28.7%	9.86
High segregation	35.6%	90.2%	30.9%	9.94
Very high segregation	37.8%	89.5%	32.4%	9.98
Blacks				
All metropolitan areas	16.2%	83.5%	17.0%	9.56
Low segregation	15.8%	85.5%	17.1%	9.59
Moderate segregation	16.6%	85.2%	16.8%	9.54
High segregation	17.4%	85.1%	18.3%	9.61
Very high segregation	15.7%	82.4%	16.6%	9.55
White-black gap				
All metropolitan areas	19.2%	5.9%	13.8%	0.38
Low segregation	12.7%	2.3%	9.8%	0.27
Moderate segregation	16.0%	3.6%	12.0%	0.32
High segregation	18.1%	5.0%	12.7%	0.33
Very high segregation	22.1%	7.1%	15.8%	0.43

Notes: In the top (bottom) panel, Core Based Statistical Areas are classified into quartiles—low, moderate, high and very high—based on their 2000 Latino-white (black-white) dissimilarity index. Sample in the top (bottom) panel is restricted to native-born whites and Latinos (blacks) between 25 and 30 years living in 187 (184) metropolitan areas with population above 100,000 residents and more than 5,000 Latinos (blacks) in 2010. 'Not idle' takes value one if the individual is working or enrolled in school. Log annual income includes total income for the previous calendar year and is available only for individuals who report positive income.

outcomes is intrinsically difficult because people sort into cities and neighborhoods based on their tastes, preferences, and unobserved resources. To address sorting across neighborhoods, we measure segregation at the level of the metropolitan area rather than at the level of the neighborhood (Cutler and Glaeser, 1997). A metropolitan area level of analysis has the added strength of capturing metropolitan area wide restrictions on choice and of measuring how all members of a racial or ethnic group in a metropolitan area may be affected by levels of segregation that operate at a higher spatial level, even those who do not live in a racially or ethnically homogenous neighborhood themselves (Chetty, Hendren, Kline, and Saez, 2014). We focus on variation in effects across racial or ethnic groups to difference out any unobserved characteristics of a metropolitan area that shape economic outcomes and are correlated with segregation.

To learn how metropolitan area segregation affects Latinos, we regress an individual outcome such as the probability of college graduation or the likelihood of being employed or in school on a measure of Latino metropolitan area residential segregation (e.g. Latino-white dissimilarity index). Specifically we estimate the following specification:

$$Y_{ijt} = \alpha_1 + \beta_1 \,\mathrm{Seg}_{j,t-1} + \beta_2 \,\mathrm{Seg}_{j,t-1} \times \mathrm{Latino}_{ij} + \beta_3 \, X_{ijt} + \beta_4 \, Z_{jt} + T_t + \varepsilon_{ijt} \qquad (1)$$

where Y_{ijt} represents a socio-economic outcome for individual i in metropolitan area j in decade t, $\mathrm{Seg}_{j,t-1}$ is the dissimilarity index between Latinos and whites for metropolitan area j in the previous decade $t-1$, X_{ijt} is a vector of individual level characteristics, Z_{jt} is a vector of metropolitan level characteristics described below, and T_t is a decade time control. We let the coefficient on metropolitan area level of segregation—β_2 in equation (1)—differ for whites and Latinos ($\mathrm{Seg}_{j,t-1} \times \mathrm{Latino}_{ij}$). Therefore, we test whether segregation has a different association with socio-economic outcomes for Latinos relative to its association with outcomes for whites.[5] We lag segregation to help address concerns about reverse causality and to better capture the segregation levels present when young adults were growing up.

We include several individual variables as controls, including age indicator variables, gender, and a set of indicator variables for Latino groups of different origin (Mexicans, Puerto Ricans, Dominicans, Cubans, Central Americans, and South Americans). As discussed above, these ancestry groups exhibit substantial differences in levels of educational attainment, income, and presumably unobserved traits that could explain differences in outcomes among Latinos. By including these ancestry-group indicator variables we capture a share of the variance in outcomes that can be attributed to the fact that Latinos of specific subgroups, who may be concentrated in different metropolitan areas, bring different backgrounds and may experience different treatment.

We also include additional time-varying metropolitan area level controls, specifically metropolitan area population and median household income, the fraction of the metropolitan area population that is Latino, black, Asian, foreign born, over 65 years, under 15 years, and unemployed, as well as the share of the metropolitan area workers employed in the manufacturing sector and working in professional occupations, the share of residents with a college degree, the share of residents in poverty status, and census region-year indicator variables. We interact these metropolitan

[5]The sum of the coefficient on segregation and the interaction of segregation with the Latino indicator variable captures the total effect of segregation on Latinos.

area controls with a Latino indicator variable to let the effects of metro area characteristics differ for Latinos as compared to whites. Again, by including all of these metropolitan area level variables and interacting segregation with a Latino indicator variable, we test whether the level of segregation in a metropolitan area has a significantly different, independent effect on socio-economic outcomes for Latinos than it does for whites.

Earlier work exploring the impacts of metropolitan segregation on individual outcomes has only examined a single year of data (e.g. Cutler and Glaeser, 1997). Using multiple years of data allows us to introduce metropolitan area level fixed effects to examine how changes over time in the level of Latino segregation in a metropolitan area are associated with changes in outcomes, while controlling for other unobserved, time-invariant metropolitan area-level factors.

To minimize both potential endogeneity from omitted variables and the reverse causality that could come from the gap in socio-economic outcomes between Latinos and whites itself contributing to metropolitan area levels of segregation, we estimate two-stage least squares models. These models address reverse causation in which a segment of the population already living in a metropolitan area might cause future segregation, but, because the instrumental variable is itself correlated with segregation, their ability to fully address sorting from selection choices made by subsequent movers is more limited (Ananat, 2011, Rosenthal and Ross, 2015). Multiple instrumental variables have been developed to predict levels of black-white metropolitan area segregation, including rivers (Hoxby, 2000) and railroad tracks (Ananat, 2011), features of the natural or built environment that enabled the black-white segregation that became entrenched through the rise of Jim Crow, the Great Migration, and post-war suburbanization.

These instruments, however, are not necessarily appropriate for the Latino-white segregation that has emerged with the growth of the Latino population in the United States since 1970, given the different historical context. To instrument for levels of Latino-white dissimilarity from 1990 to 2010, we rely on an instrument that captures features of the historical built environment that allowed for more segregation. Specifically, we create a variable measuring the dissimilarity index between single-family detached housing and other housing types in 1970. In 1970, there were 9.1 million individuals who identified as Latino in the United States, accounting for only 4.7 percent of the population. In the four decades after the passage of the 1965 Immigration and Nationality Act, more than 29 million immigrants from Latin America moved to the United States (Pew Research Center, 2015), and the relatively low incomes of those migrants constrained many to live in less-expensive multi-family housing. We hypothesize that when different types of housing are ex-ante placed in separate parts of the city, more segregation is likely to result, as Latinos are likely to disproportionately settle in multi-family or single-family attached housing because of their lower homeownership rates and lower average incomes (see Weicher and Thibodeau, 1988 and Brueckner and Rosenthal, 2009, for a related measure of the age of the housing stock).[6]

[6] To construct this instrument, we use the 1970 Neighborhood Change Database (NCDB) to calculate the dissimilarity index in 1970 between single-family detached housing and all other types of housing units (single-family attached dwellings, as well as all multi-family dwellings) for each Standard Metropolitan Statistical Area (SMSA), the 1970 definition of metropolitan areas. The source units of analysis are census tracts as defined in 1970. We have data on instruments for 142 out of the 187 CBSAs in the initial sample. Those CBSAs missing from the 2SLS sample are generally smaller and more recently recognized CBSAs.

Table 2: Types of housing units by racial/ethnic group, 1980–2010

	One family detached house	One family attached house	Building with 2 to 9 units	Building with 10+ units	Other (e.g. mobile home, boat)
Whites					
1980	60.2%	4.6%	16.2%	15.3%	3.6%
1990	59.4%	6.3%	14.3%	14.5%	5.5%
2000	62.7%	6.5%	12.7%	13.5%	4.7%
2010	64.9%	6.9%	11.2%	13.0%	4.0%
Latinos					
1980	42.9%	5.1%	24.8%	25.5%	1.7%
1990	41.7%	7.0%	22.9%	24.4%	3.9%
2000	42.9%	7.6%	21.8%	23.8%	3.9%
2010	48.0%	6.6%	20.3%	20.9%	4.2%
Blacks					
1980	39.2%	9.6%	26.9%	23.4%	0.9%
1990	39.3%	10.1%	24.9%	23.0%	2.6%
2000	42.5%	10.2%	23.4%	22.2%	1.8%
2010	44.8%	9.8%	21.8%	21.8%	1.6%

Notes: IPUMS-USA data for 1990 5% sample Decennial Census, 2000 5% sample Decennial Census and ACS 2007–2011. Race/ethnicity of household head is assigned to the type of housing unit. The sample for whites and Latinos is restricted to the 142 Core Based Statistical Areas (CBSAs) that are used in instrumental variable estimations, while the sample for blacks is restricted to the 147 CBSAs used in analogous estimations.

In table 2, we present the share of white, Latino, and black households living in different types of housing units by decade from 1980 to 2010. In each decade, Latinos are more likely to live in multi-family housing than whites. While this difference also exists for blacks, the dissimilarity of housing types is not as consistent a predictor of black-white segregation as it is of Latino-white segregation (results shown below in table 3), because of the existence of already historically established patterns of black-white segregation independent of housing type.

Figure 1 shows a scatterplot of the strong positive relationship between the 2000 Latino-white segregation and the 1970 dissimilarity index between single-family detached housing and all other housing types. For example, the New York NY-NJ-PA metropolitan area has simultaneously the highest level of single/multi-family housing dissimilarity index (0.793) and a very high score on the Latino-white dissimilarity index (0.656). At the other extreme, Modesto, CA has a very low housing type dissimilarity score (0.252) and also a low score on the Latino-white dissimilarity index (0.352).

We combine this measure of the dissimilarity of residential housing typology with two existing measures of the jurisdictional or fiscal environments that enable segregation—the number of local governments and the share of local revenue from federal or state transfers (Cutler and Glaeser, 1997), both from 1962, before the passage of the 1965 Immigration and Nationality Act and the rapid increase of the US Latino population. Following Tiebout (1956), a larger number of different

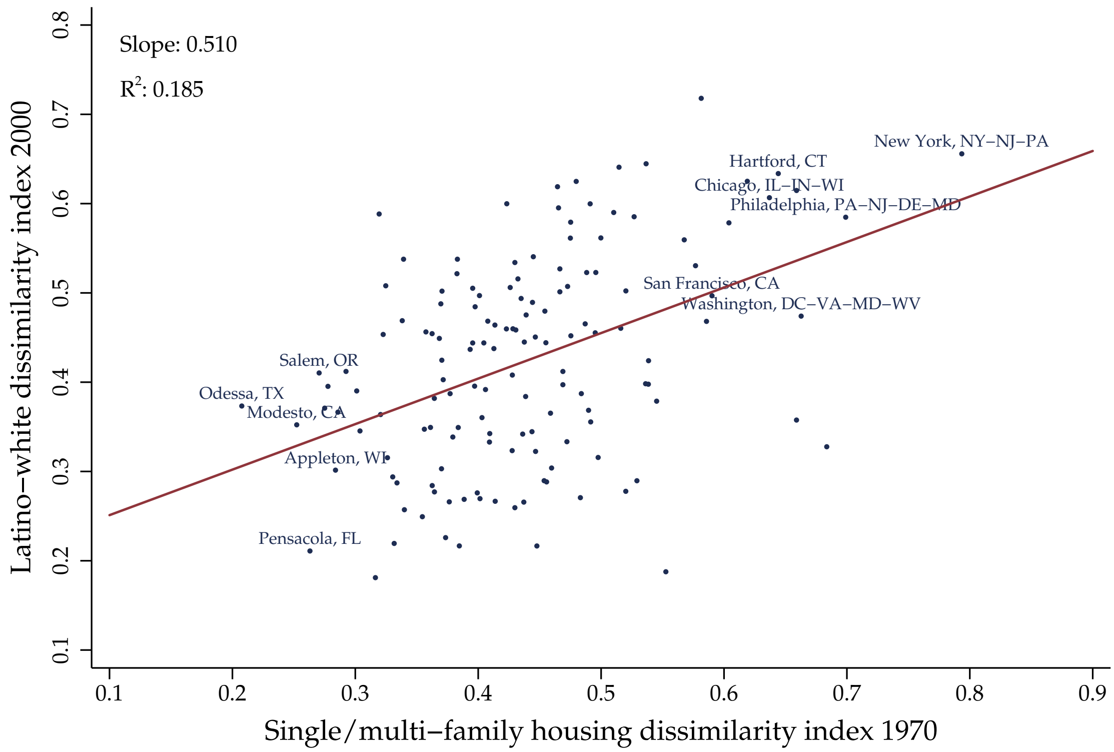

Figure 1: 2000 segregation and 1970 single-/multi-family dissimilarity index

municipalities within a given metropolitan area encourages greater sorting on the basis of municipal tax rates and service provision, thus facilitating greater segregation. Relatedly, the larger the share of local revenue from state or federal sources, the lower the variation in municipal tax rates and the greater the equality of public goods, therefore presumably the lower the incentive for sorting by jurisdiction.[7]

The first stage estimations of the Latino-white lagged dissimilarity index by decade from 1980 to 2000 are presented in table 3.[8] Note that in the 2SLS regressions, we include the black-white dissimilarity index as a control in order to capture historical levels of racial discrimination in a metropolitan area and related factors that change slowly over time, such as social, political, or educational institutions shaped by segregated norms.[9] As table 3 indicates, housing type dissimilarity has the expected relationship with segregation, even after including the lagged black-white segregation level and other metropolitan area controls. The number of local governments, a generally consistent predictor of black-white segregation, is not an effective predictor of Latino-white segregation in any of the three decades. The share of revenue from federal and state transfers does

[7]The data for both measures come from the 1962 Census of Governments Survey and are made available by the Inter-university Consortium for Political and Social Research at the University of Michigan (United States Department of Commerce, 2001, `http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/12`). Like Cutler and Glaeser (1997), we measure the share of intergovernmental transfers for the localities in a state as a whole in order to avoid including local endogenous factors and to better capture the relevant state political characteristics.

[8]Full results showing the effects of metropolitan area controls are available in online appendix tables B.10–B.12.

[9]Results are largely similar without this control.

Table 3: First-stage estimation of lagged dissimilarity indices

	1980 Latino-white dissimilarity index			1980 black-white dissimilarity index		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.547		0.455	-0.012		-0.135
	(0.115)***		(0.126)***	(0.108)		(0.100)
Log of local governments 1962		0.003	-0.005		0.013	0.014
		(0.014)	(0.015)		(0.011)	(0.011)
% of revenue from transfers 1962		-0.610	-0.440		-0.382	-0.423
		(0.171)***	(0.158)***		(0.104)***	(0.105)***
Black-white dissimilarity index 1980	0.298	0.078	0.168			
	(0.098)***	(0.139)	(0.112)			
Observations	511,547	511,547	511,547	546,378	546,378	546,378
R^2	0.739	0.716	0.758	0.648	0.704	0.711
	1990 Latino-white dissimilarity index			**1990 black-white dissimilarity index**		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.478		0.432	0.266		0.137
	(0.119)***		(0.128)***	(0.080)***		(0.079)*
Log of local governments 1962		0.002	-0.004		0.035	0.033
		(0.016)	(0.016)		(0.011)***	(0.011)***
% of revenue from transfers 1962		-0.401	-0.269		-0.279	-0.230
		(0.155)***	(0.143)*		(0.097)***	(0.099)**
Black-white dissimilarity index 1990	0.299	0.300	0.241			
	(0.096)***	(0.129)**	(0.108)**			
Observations	396,363	396,363	396,363	415,144	415,144	415,144
R^2	0.766	0.733	0.774	0.703	0.747	0.753
	2000 Latino-white dissimilarity index			**2000 black-white dissimilarity index**		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.281		0.277	0.465		0.241
	(0.093)***		(0.106)***	(0.104)***		(0.106)**
Log of local governments 1962		0.007	-0.0006		0.043	0.033
		(0.014)	(0.015)		(0.012)***	(0.012)***
% of revenue from transfers 1962		-0.094	-0.022		-0.258	-0.190
		(0.160)	(0.158)		(0.095)***	(0.104)*
Black-white dissimilarity index 2000	0.237	0.285	0.234			
	(0.078)***	(0.091)***	(0.091)***			
Observations	404,693	404,693	404,693	397,549	397,549	397,549
R^2	0.677	0.659	0.677	0.726	0.751	0.763
Number of CBSAS	142	142	142	147	147	147

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications include a constant term, a female indicator variable, age indicator variables and the following CBSA controls: log population, log median household income and shares of the population that are black, Latino, Asian, over 65 years, under 15 years, unemployed, working in manufacturing, in poverty status and with college degree. Columns (1)–(3) include Latino ancestry-group indicator variables and metropolitan area controls interacted with a Latino indicator variable. Columns (4)–(6) include a black indicator variable and metropolitan area controls interacted with a black indicator variable.

contribute to predicting Latino-white dissimilarity in 1980 and 1990. The results in column (3) indicate that the coefficient on the housing type dissimilarity measure remains highly significant and does not experience a large change in its magnitude in the presence of the other instruments. We take this as evidence that our proposed instrument is strong. Further, the *F*-statistic reported on the weak instruments identification test exceeds all thresholds proposed by Stock and Yogo (2005) for the maximal relative bias and maximal size in 1980 and 1990, though not in 2000.[10] As evidenced by the coefficients in columns (4) and (6), the housing type dissimilarity measure also predicts black-white dissimilarity in 1990 and 2000, though the strength of the prediction is not as strong for black-white segregation as it is for Latino-white segregation, once other instruments are included.

We carry out several checks to validate our first stage results. First, to confirm that housing type dissimilarity in 1970 was not the result of existing levels of Latino white segregation, we test the relationship between the 1970 Latino-white dissimilarity index and our housing type dissimilarity index, controlling for log population of the metropolitan area in 1970, and do not find a statistically significant association.[11] Second, higher levels of housing type dissimilarity in 1970 could have been more common in economically vulnerable and socially conservative metropolitan areas leading to more detrimental outcomes for minorities (Antecol and Cobb-Clark, 2008). The bivariate relationships between metropolitan area characteristics in 1970 and our housing type dissimilarity measure actually show that more affluent metropolitan areas exhibited higher levels of housing type dissimilarity.[12] While we are unclear theoretically why this relationship exists, we control for income, unemployment, college attainment, poverty rate and other metropolitan area characteristics in our first stage specifications. Third, we examine whether metropolitan areas with higher levels of housing type dissimilarity experienced larger subsequent inflows of Latinos. This could be a source of concern if Latinos disproportionately moved to these metropolitan areas based on the availability of multi-family housing or another unobserved attribute correlated with this type of housing (e.g. a booming construction sector). Specifically, we estimate regressions of our housing type dissimilarity index on the metropolitan area change in the share of Latinos between 1970 and 1980, controlling for initial metropolitan area population, and do not find a significant association.[13] Thus, historical housing configurations of some metropolitan areas contributed to the segregation of Latinos from whites; however, they did not necessarily attract larger inflows of

[10]The *F*-statistic (or Kleinberger-Papp rk Wald statistic) exceeds 16 in 1980, 12 in 1990 and 4 in 2000. Only in 2000, it falls slightly below the critical value for the 15% maximal IV size. This is in part due to the lack of significance of the two historic predictors of black-white segregation in column (3) of table 3 (bottom panel). Our estimates, shown in table 7, follow Wooldridge's IV estimation adjustment (see Wooldridge, 2002) that uses as 'instruments' the predicted value obtained for the Latino(black)-white dissimilarity index from the first-stage regression in table 3 and its interaction with a Latino(black) indicator variable.

[11]In complementary analogous estimations to our reported first-stage estimates, we control for other metropolitan area variables in 1970 such as the log average household income and the proportion of the population that are Latino, black, unemployed, with college degree, in poverty status, and working in manufacturing, and again do not find a statistically significant association. Results available upon request.

[12]Metropolitan areas with higher housing type dissimilarity indices have a higher share of residents with a bachelor's degree, higher average household income, lower poverty rates and a lower share of Latino residents. Yet, they do not have significant associations with the share of black residents or the proportion of workers in manufacturing.

[13]We obtain similar results when using the metropolitan area change in the share of Latinos between 1970 and 1990 as a dependent variable and when including a larger list of controls.

Table 4: Balancing tests of metropolitan area segregation on individual characteristics, 1990–2010

Dependent variable:	Latino-white lagged dissimilarity index		Black-white lagged dissimilarity index	
	(1)	(2)	(3)	(4)
Female	-0.012	-0.013	0.020	0.020
	(0.018)	(0.022)	(0.015)	(0.016)
Age 26	0.022	0.022	0.021	0.020
	(0.026)	(0.026)	(0.019)	(0.019)
Age 27	0.010	0.009	0.052	0.051
	(0.026)	(0.027)	(0.026)**	(0.027)*
Age 28	0.003	0.002	0.041	0.039
	(0.034)	(0.035)	(0.028)	(0.029)
Age 29	-0.023	-0.025	0.045	0.043
	(0.047)	(0.048)	(0.034)	(0.035)
Age 30	-0.041	-0.042	0.050	0.048
	(0.052)	(0.054)	(0.037)	(0.038)
High school completed		-0.056		-0.043
		(0.134)		(0.096)
Associate degree		-0.014		-0.195
		(0.116)		(0.106)*
College degree		0.071		0.309
		(0.157)		(0.124)**
F-test	0.920	0.12	1.12	2.09
P-value	0.483	0.946	0.351	0.103
Observations	1,395,238	1,395,238	1,430,120	1,430,120
R^2	0.677	0.677	0.737	0.737

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications have a constant term and census region-year indicator variables. Metropolitan area controls are listed in the notes of table 3. Sample in columns (1)–(2) is restricted to whites and Latinos. Sample in columns (3)–(4) is restricted to whites and blacks. See notes in table 1 for additional sample details. *F*-tests in columns (1) and (3) correspond to the joint effect of individual characteristics while in columns (2) and (4) correspond to the joint effect of the additional regressors on levels of education.

Latinos compared to other metropolitan areas.

Finally, we conduct balancing tests to examine the potential sorting of particular individuals into metropolitan areas with different levels of segregation. The idea is to test whether observable individual characteristics (e.g. gender, age, and educational outcomes) are correlated with measures of segregation across metropolitan areas. If we do not find statistically significant associations, then it is less likely that individuals sort into more or less segregated metropolitan areas based on unobserved characteristics (Bifulco, Fletcher, and Ross, 2011, Lou and Song, 2017). In table 4, we find no evidence that native-born white and Latino young adults have sorted on those observable characteristics. We also allow for interactions between a Latino indicator variable and individual characteristics (results not shown) and do not find any significant association. Therefore, we find no indication of certain types of Latino or white young adults sorting based on metropolitan area segregation.

4. Results

OLS results on the relation between segregation and individual outcomes

In table 5, we estimate ordinary least squares regressions of each individual outcome on metropolitan area levels of segregation, as well as individual and metropolitan area controls. We show results with contemporaneous and lagged segregation levels (in which 1990 outcomes are linked to 1980 segregation levels, etc.). We also show results for a regression with CBSA fixed effects with lagged segregation levels.[14] In each pair of rows in the first panel, the first row reports the coefficients on the metropolitan area Latino-white dissimilarity index and the second row the interaction between this index and a Latino indicator variable. For results in the top panel, the sample consists only of whites and Latinos, so the coefficient on the dissimilarity index can be interpreted as the association between Latino-white segregation and white outcomes, while the coefficient on the interaction between the dissimilarity index and the Latino indicator variable shows any difference in the association between segregation and outcomes for Latinos as compared to whites. Standard errors are clustered at the metropolitan area level.

Results reveal significant associations between metropolitan area segregation levels and every measured individual black and Latino outcome. Starting with the probability of having completed college for Latinos aged 25-30 in column (1), we find that the interaction coefficient is negative and statistically significant, indicating that, in more segregated metropolitan areas, Latinos are less likely to complete college relative to their white counterparts. The results are similar whether the dissimilarity index is lagged or not, and, when metropolitan area fixed effects are included, the interaction coefficient increases in magnitude. A one standard deviation increase in the Latino-white dissimilarity index is related to a decline in the probability of finishing college of 5.5 percentage points for Latinos relative to white graduation rates. The overall difference in the means in college graduation rates for whites and Latinos, pooled across the 1990-2010 study period, is 18.6 percentage points.

Looking at the incidence of being employed or in school again reveals that higher levels of segregation are consistently associated with a lower likelihood of being employed or in school for Latino young adults relative to whites. A one standard deviation increase in the level of segregation is associated with a decrease in the likelihood of being either employed or in school for Latino 25-30 year olds relative to whites of 2.1 percentage points (the overall difference between whites and Latinos in this age range is 5.1 percentage points).

As would be expected, the results with regard to professional occupations parallel the results with regard to college attainment. A one standard deviation increase in the Latino-white dissimilarity index is related to a decline in the probability of professional employment of 3.7 percentage points for Latinos relative to white graduation rates (the overall difference in professional employment rates between whites and Latinos is 11.7 percentage points).

As for income among 25-30 year olds, segregation is also associated with significantly larger Latino-white gaps. The results are consistent across all specifications and the magnitude is large.

[14] Results for CBSA fixed effects models are similar whether or not segregation is lagged.

Table 5: Estimation of the effect of segregation on individual outcomes, 1990–2010

Dependent variable:	College graduation	Not idle	Professional occupation	Log annual income
	(1)	(2)	(3)	(4)
Pooled OLS				
Latino-white lagged dissimilarity index	0.032	0.037	0.011	0.127
	(0.043)	(0.012)***	(0.029)	(0.081)
Latino-white lagged diss index × Latino	-0.289	-0.141	-0.211	-0.657
	(0.041)***	(0.019)***	(0.028)***	(0.086)***
Pooled OLS				
Latino-white dissimilarity index	0.031	0.036	0.009	0.146
	(0.045)	(0.013)***	(0.030)	(0.085)*
Latino-white diss index × Latino	-0.330	-0.157	-0.228	-0.670
	(0.046)***	(0.024)***	(0.032)***	(0.091)***
CBSA fixed-effects				
Latino-white lagged dissimilarity index	0.080	0.019	0.046	0.146
	(0.041)*	(0.015)	(0.029)	(0.081)*
Latino-white lagged diss index × Latino	-0.386	-0.149	-0.256	-0.653
	(0.051)***	(0.021)***	(0.035)***	(0.073)***
Observations	1,395,238	1,395,238	1,395,238	1,276,664
Number of CBSAS	187	187	187	187
Pooled OLS				
Black-white lagged dissimilarity index	0.105	0.026	0.059	0.047
	(0.046)**	(0.012)**	(0.035)*	(0.071)
Black-white lagged diss index × black	-0.259	-0.151	-0.136	-0.404
	(0.048)***	(0.022)***	(0.034)***	(0.091)***
Pooled OLS				
Black-white dissimilarity index	0.151	0.032	0.085	0.053
	(0.055)***	(0.013)**	(0.043)**	(0.086)
Black-white diss index × black	-0.280	-0.169	-0.129	-0.406
	(0.051)***	(0.025)***	(0.037)***	(0.098)***
CBSA fixed-effects				
Black-white lagged dissimilarity index	-0.003	0.033	-0.005	0.122
	(0.069)	(0.025)	(0.039)	(0.102)
Black-white lagged diss index × black	-0.240	-0.132	-0.137	-0.436
	(0.055)***	(0.021)***	(0.032)***	(0.078)***
Observations	1,430,120	1,430,120	1,430,120	1,307,648
Number of CBSAS	184	184	184	184

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. In the top (bottom) panel, the sample is restricted to native-born whites and Latinos (blacks) between 25 and 30 years. All specifications have a constant term, a female indicator variable, age and census region-year indicator variables. The top panel includes Latino ancestry-group indicator variables, while the bottom panel includes a black indicator variable. Additional CBSA controls include log population, log median household income and shares of population that are black, Latino, Asian, over 65 years, under 15 years, unemployed, working in manufacturing, in poverty status and with college degree. These CBSA controls are also interacted with a Latino or black indicator variable accordingly.

A one standard deviation increase in Latino-white segregation is associated with a 9.9 percent increase in the gap between Latino incomes relative to whites. In the sample, annual income for whites exceeds those for Latinos by 18.9 percent.

As shown in the second panel, the relationship between metropolitan area segregation and outcomes among African American young adults is similar to that for Latinos. In more segregated metropolitan areas, black young adults are less likely to graduate from college, to be either in school or employed, and to work in a professional occupation, and have lower incomes, relative to whites. The results are again robust to lagged dissimilarity and metropolitan area fixed effects.

In sum, higher levels of segregation are associated with worse educational and employment outcomes for both black and Latino young adults. The magnitudes of these negative associations are larger for Latinos in every case except for the likelihood of being simultaneously out of work and out of school.[15]

We carried out alternative estimations that use the isolation index as the measure of metropolitan area segregation, estimated the same specifications for a younger sample of adults between the ages of 20 and 24, and excluded recent (domestic) migrants from the sample. Results from all of these robustness tests, available upon request, are similar both in terms of significance and magnitude of the effects. We also estimated regressions of black-white segregation on Latino outcomes and of Latino-white segregation on black outcomes and found no significant associations, suggesting that these results are not artifacts of unobserved metropolitan area characteristics associated with higher levels of residential segregation in general.[16] In sum, our findings indicate that Latino-white segregation has consistent negative associations with socio-economic outcomes for Latino young adults relative to whites and black-white segregation has consistent negative associations with socio-economic outcomes for black young adults relative to whites.

The link between segregation and individual outcomes by ancestry

Examining the association between segregation and individual outcomes by ancestry in 2010 reveals considerable heterogeneity across groups.[17] In table 6, we include interactions between the dissimilarity index and seven ancestry groups (Cuban, Mexican, South American, Central American, Puerto Rican, Dominican and those who identified as 'Other Hispanic').[18] Thus, the total effect of segregation in each of these groups is the sum of the general Latino interaction coefficient and the ancestry group of interest. The association between segregation and outcomes is generally largest for Latinos who self-report having Puerto Rican or Dominican ancestry. For

[15]Note that the standard deviations of the Latino-white (0.144) and black-white (0.130) dissimilarity indices are similar. Thus, the magnitudes of the coefficients can be compared for both samples.

[16]Results available upon request.

[17]The question on Latino ancestry identification varied in 2000 and this prevents us from establishing consistent ancestry groups over time (Logan and Turner, 2013). We focus our analysis in 2010, when the Latino population in the United States is at its largest, most heterogeneous, and most geographically extensive.

[18]Sample sizes do not allow us to construct more narrow ancestry groups within Central and South America. The total sample size for the Latino-white analyses in 2010 is 432,756 native born young adults, of which 359,903 (83.2%) are whites. Of the Latinos in the sample, 47,256 (64.9%) identify their ancestry as Mexican, 11,590 (15.9%) as Puerto Rican, 2,551 (3.5%) as Cuban, 2,902 (4.0%) as Central American, 1,482 (2.0%) as Dominican, 2,543 (3.5%) as South American, and 4,529 (6.2%) as 'Other.'

Table 6: Estimation of the effect of segregation by Latino ancestry, 2010

Dependent variable:	College graduation	Not idle	Professional occupation	Log annual income
	(1)	(2)	(3)	(4)
Latino-white dissimilarity index	0.066	0.069	0.040	0.096
	(0.068)	(0.017)***	(0.045)	(0.061)
Latino-white lagged DI × other Latino	-0.138	-0.063	-0.023	-0.303
	(0.086)	(0.055)	(0.072)	(0.160)*
Latino-white lagged DI × Cuban	-0.242	-0.059	-0.039	-0.050
	(0.159)	(0.062)	(0.095)	(0.174)
Latino-white lagged DI × Mexican	-0.344	-0.110	-0.227	-0.516
	(0.068)***	(0.034)***	(0.049)***	(0.093)***
Latino-white lagged DI × South American	-0.511	-0.070	-0.413	-0.834
	(0.143)***	(0.040)*	(0.101)***	(0.135)***
Latino-white lagged DI × Central American	-0.545	-0.105	-0.333	-0.935
	(0.114)***	(0.061)*	(0.083)***	(0.171)***
Latino-white lagged DI × Puerto Rican	-0.614	-0.282	-0.451	-0.947
	(0.084)***	(0.034)***	(0.055)***	(0.117)***
Latino-white lagged DI × Dominican Republic	-0.685	-0.140	-0.584	-0.627
	(0.157)***	(0.086)	(0.121)***	(0.212)***
Latino ancestry-group indicators	Yes	Yes	Yes	Yes
Observations	432,756	432,756	432,756	395,742
R^2	0.093	0.033	0.048	0.064

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is restricted to native-born whites and Latinos between 25 and 30 years in 2010. Additional controls listed in notes of table 5 are included. The 'other Latino' category includes those Latinos who self-report 'Spaniard' or 'Other, not specified' ancestry. DI stands for dissimilarity index.

instance, a one standard deviation increase in the dissimilarity index is associated for Puerto Ricans with an 8.4 percentage-point decrease in the likelihood of attending college relative to whites, a 4.3 percentage-point decrease in the likelihood of being employed or in school relative to whites, and a 13.7 percentage reduction in income relative to whites. This stronger association may reflect the larger share of Puerto Ricans and Dominicans who are poor and identify or are perceived as black.

We see roughly similar associations for those who self-identified as having Central American ancestry in terms of college graduation and income, but smaller associations between segregation and professional occupation as well as the likelihood of being simultaneously out of work and out of school. Segregation also has large negative associations with income and the likelihood of being in a professional occupation for those who identify as having South American ancestry. The association between segregation and socio-economic outcomes is somewhat more modest for those who identify as having Mexican ancestry, though the association with an increased likelihood of being simultaneously out of school and out of work is large. Segregation has no negative association with educational or labor market outcomes for those who identified Cuban ancestry and almost no negative association for those who identified their ethnicity as Hispanic, but their ancestry as 'Other.'

Instrumental variable results

Table 7 presents the IV estimation of the effect of the Latino-white dissimilarity index on socio-economic outcomes in 1990, 2000, and 2010. The first two columns show results for 1990, with the first column repeating the OLS estimation for the 142 CBSAs included in the sample and the second column showing IV estimates using all three instrumental variables discussed above for Latino-white segregation. The two subsequent rows within each panel show analogous results for blacks for a subset of 147 CBSAs and IV estimates that use the same set of instrumental variables for black-white segregation. IV estimations use Wooldridge's IV adjustment given that the same instruments are used to predict the coefficient on segregation and the interaction with the minority indicator variable (see table note). Columns (3) and (4) show results for 2000, and columns (5) and (6) for 2010, all following the same pattern of OLS estimations followed by IV estimations.

Instrumental variable estimates of the causal effect of segregation on racial or ethnic gaps in college graduation show that segregation widens the gap in outcomes between whites and both blacks and Latinos in all three decades. But in some years (2000 for blacks and 2010 for Latinos), the total effect of segregation (adding the coefficient on the dissimilarity index and the coefficient on the interaction term) is zero or positive. Using the instrumental variable estimation, a one standard deviation increase in metropolitan area segregation in 2010 had the effect of widening the gap in college graduation rates between whites and Latinos by 8 percentage points, slightly more than the 5 percentage-point gap in the OLS regression.

Regarding the likelihood of being either employed or in school, the IV results are significant and negative for Latinos in both 2000 and 2010 (as well as slightly larger than the OLS results) but significant for blacks only in 2010. The 2 percentage-point gap between whites and Latinos caused by a one standard deviation increase in segregation is similar to the 1.6 percentage-point gap found in the OLS estimation.

For both blacks and Latinos, segregation widens the gap with whites in the likelihood of professional occupation in all three decades. Further, when looking at point estimates the effects of segregation on gaps in access to professional occupations were wider in 2000 and 2010 than they were in 1990. The magnitudes of the IV results are again larger than the OLS estimates—a 6.9 percentage-point gap caused by a one standard deviation increase in segregation compared to a 3.5 percentage-point gap in the OLS estimation.

Finally, with regard to income, the IV results are large, negative, and significant for both blacks and Latinos in all three decades. In fact, IV estimates indicate that a one standard deviation increase in the Latino-white dissimilarity index in 2010 almost doubles the income gap between Latinos and whites compared to OLS estimates, from 8.1 to 15.3 percent. This causal effect of segregation accounts for 56 percent of the total gap in earnings between Latinos and whites in 2010 (27.5 percent). These earnings gaps have been remarkably persistent along the three decades.

Overall, the IV results present a relatively consistent story of negative effects of metropolitan area segregation on socio-economic outcomes for both Latino and black young adults. Somewhat surprisingly, the magnitudes of the effects of segregation are generally larger for Latinos than for African-Americans. Also surprisingly, the negative effects of segregation for black young adults

Table 7: Instrumental variable estimation of the effect of segregation on individual outcomes

	1990		2000		2010	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
College graduation						
Latino-white lagged diss index	0.072 (0.054)	0.433 (0.117)***	0.046 (0.070)	0.535 (0.197)***	0.014 (0.076)	0.827 (0.374)**
Latino-white lagged DI × Latino	-0.280 (0.053)***	-0.607 (0.139)***	-0.422 (0.063)***	-0.737 (0.165)***	-0.421 (0.075)***	-0.681 (0.221)***
Black-white lagged diss index	0.048 (0.056)	0.413 (0.162)**	0.169 (0.054)***	0.704 (0.156)***	0.211 (0.071)***	0.531 (0.128)***
Black-white lagged DI × black	-0.149 (0.055)***	-0.457 (0.144)***	-0.343 (0.066)***	-0.701 (0.155)***	-0.372 (0.085)***	-0.919 (0.188)***
Not idle						
Latino-white lagged diss index	0.028 (0.014)**	0.067 (0.027)**	0.021 (0.016)	0.060 (0.042)	0.023 (0.020)	0.019 (0.081)
Latino-white lagged DI × Latino	-0.144 (0.043)***	-0.011 (0.100)	-0.193 (0.030)***	-0.271 (0.058)***	-0.138 (0.033)***	-0.171 (0.059)***
Black-white lagged diss index	0.069 (0.018)***	0.164 (0.043)***	0.040 (0.015)**	0.109 (0.030)***	0.050 (0.018)***	0.088 (0.039)**
Black-white lagged DI × black	-0.114 (0.037)***	-0.066 (0.147)	-0.144 (0.028)***	-0.076 (0.072)	-0.114 (0.047)**	-0.210 (0.121)*
Professional occupation						
Latino-white lagged diss index	-0.005 (0.031)	0.216 (0.070)***	0.001 (0.039)	0.301 (0.119)**	0.002 (0.049)	0.269 (0.176)
Latino-white lagged DI × Latino	-0.181 (0.046)***	-0.361 (0.101)***	-0.313 (0.040)***	-0.671 (0.118)***	-0.293 (0.048)***	-0.583 (0.130)***
Black-white lagged diss index	0.036 (0.032)	0.222 (0.088)**	0.095 (0.036)***	0.400 (0.089)***	0.101 (0.043)**	0.237 (0.080)***
Black-white lagged DI × black	-0.131 (0.040)***	-0.277 (0.105)***	-0.181 (0.044)***	-0.458 (0.127)***	-0.207 (0.058)***	-0.475 (0.115)***
Log annual income						
Latino-white lagged diss index	0.215 (0.092)**	0.696 (0.122)***	0.207 (0.079)***	0.837 (0.184)***	0.065 (0.067)	0.323 (0.322)
Latino-white lagged DI × Latino	-0.966 (0.145)***	-1.707 (0.309)***	-0.816 (0.107)***	-1.567 (0.234)***	-0.657 (0.088)***	-1.205 (0.209)***
Black-white lagged diss index	0.030 (0.098)	0.363 (0.198)*	0.201 (0.084)**	0.867 (0.211)***	0.197 (0.077)**	0.497 (0.164)***
Black-white lagged DI × black	-0.509 (0.144)***	-0.884 (0.467)*	-0.592 (0.132)***	-0.680 (0.344)**	-0.445 (0.116)***	-0.975 (0.256)***

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications include a constant term, a female indicator variable and age indicator variables. Controls included for CBSAs and their interactions with a Latino or black indicator variable are the ones listed in table B.12. The IV specifications for Latino-white and black-white dissimilarity indices are columns (3) and (6) of table B.12, respectively. Regressions follow Wooldridge's IV estimation adjustment (see Wooldridge, 2002) that uses as 'instruments' the predicted value obtained for the Latino (black)-white dissimilarity index from the first-stage regression in table B.12 and its interaction with a Latino (black) indicator variable. DI stands for dissimilarity index.

are largest in 2010, while the negative effects for Latino young adults are generally largest in 2000.

The results also suggest some benefit to whites of segregated metropolitan areas for all four measures of socio-economic outcomes in most specifications. These potential benefits for whites are more significant for black-white segregation than Latino-white segregation, especially in 2010, and are particularly strong for college graduation and earnings. These findings are consistent with Cutler and Glaeser (1997), who found that young white adults benefited from segregation in 1990, at least with respect to college attainment. Segregation, by generating inequality in public goods and in social networks, is likely to both reinforce advantage and cumulate disadvantage by widening inequality and facilitating resource hoarding (Durlauf, 2004, Graham, 2016).

Mechanisms

Table 8 examines mechanisms that can help explain how residential segregation translates into unequal individual outcomes. Using the Neighborhood Change Database developed by GeoLytics and the Urban Institute, we construct weighted averages of neighborhood socio-economic characteristics. These weighted averages or exposure rates show the extent to which the average person of a specific race or ethnicity is exposed to a neighborhood characteristic. We construct measures of exposure to poverty and exposure to neighbors with college degrees. We also use IPUMS data to calculate the average growth in college graduate employment by three-digit industry in the nation as a whole between 1990 and 2010.[19] We then calculate how exposed workers of different races or ethnicities in each CBSA were to subsequent growing or declining industries. We subtract from all our exposure measures the overall mean in the metropolitan area (calculated for all workers regardless of race) to avoid capturing differences in levels across metropolitan areas.

The exposure of white young adults to neighbors in poverty is associated with a decline in the odds of being employed or in school. White exposure to neighbors with college degrees is associated with an increase in college graduation and, relatedly, in the likelihood of working in a professional occupation.

The exposure of Latino young adults to individuals in poverty is associated with worse outcomes across the board, while exposure to neighbors with a bachelor degree is associated with an increased likelihood of college graduation. Latinos also benefit from being exposed to sectors that experienced notable skilled employment growth, in terms of college graduation, likelihood of professional employment, and earnings.

Once we control for these exposures, the coefficient on Latino-white segregation interacted with a Latino indicator variable falls by between 48 and 67 percent, depending on the outcome. Large unexplained effects remain for all outcomes.

The most consistent and largest IV results are the negative effects of segregation on black and Latino income in comparison to whites. To see how much of this effect could be explained by segregation's effects on educational attainment, we re-estimate our regressions of income, after adding a set of binary variables indicating the educational level of the individual and their self-

[19] We construct time-consistent three-digit industry codes using the crosswalk provided in Autor and Dorn (2013).

Table 8: Potential mechanisms for the effects of Latino-white segregation, 2010

Dependent variable:	College graduation	Not idle	Professional occupation	Log annual income
	(1)	(2)	(3)	(4)
OLS				
Latino-white lagged dissimilarity index	0.009	0.051	0.012	0.048
	(0.061)	(0.016)***	(0.039)	(0.056)
Latino-white lagged diss index × Latino	-0.364	-0.131	-0.253	-0.580
	(0.064)***	(0.030)***	(0.043)***	(0.085)***
OLS including mechanisms				
Latino-white lagged dissimilarity index	-0.010	0.031	0.000	0.025
	(0.062)	(0.018)*	(0.040)	(0.060)
Latino-white lagged diss index × Latino	-0.119	-0.068	-0.094	-0.217
	(0.083)	(0.041)*	(0.063)	(0.120)*
White exposure to poverty × white	1.171	-0.459	0.783	-0.478
	(0.802)	(0.233)**	(0.563)	(0.890)
White exposure to college × white	1.696	-0.041	1.064	-0.073
	(0.382)***	(0.137)	(0.270)***	(0.419)
White exposure to industry growth × white	-0.154	0.126	-0.012	-0.421
	(0.820)	(0.180)	(0.501)	(0.866)
Latino exposure to poverty × Latino	-0.559	-0.386	-0.428	-1.881
	(0.274)**	(0.164)**	(0.216)**	(0.477)***
Latino exposure to college × Latino	0.422	-0.156	0.188	-0.300
	(0.244)*	(0.100)	(0.171)	(0.261)
Latino exposure to industry growth × Latino	0.199	0.032	0.177	0.386
	(0.088)**	(0.064)	(0.080)**	(0.189)**
Reduction in DI coefficient for Latinos	67%	48%	63%	63%

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. Sample is restricted to native-born whites and Latinos between 25 and 30 years in 2010. Additional controls listed in notes of table 5 are included. See main text for an explanation on mechanisms.

reported English proficiency (whether they speak only English at home or speak it very well as compared to not well or not at all). As shown in column (2) of table 9, the inclusion of education and English proficiency explains just under 40 percent of the differences between blacks and whites in annual income and half of the difference between Latinos and whites. When we shift from examining log annual income to examining log hourly income, this still leaves significant differences between whites and Latinos in hourly income (though not between whites and blacks), even after controlling for education. In sum, a portion of the wider differences in income between whites and minorities in more segregated metropolitan areas can be explained by differences in educational attainment in those areas, and, for blacks, some of the differences can also be explained by differences in participation or hours worked. For Latinos, we find that hourly income is significantly shaped by metropolitan area segregation, even after taking into account English proficiency, education and number of hours worked.

Table 9: Estimation of the effect of segregation on income, 1990–2010

Dependent variable:	Log annual income		Log hourly income	
	(1)	(2)	(3)	(4)
Pooled OLS				
Latino-white lagged dissimilarity index	0.127	0.100	0.023	0.012
	(0.081)	(0.083)	(0.072)	(0.073)
Latino-white lagged diss index × Latino	-0.657	-0.368	-0.275	-0.138
	(0.086)***	(0.081)***	(0.067)***	(0.067)**
Pooled OLS				
Latino-white dissimilarity index	0.146	0.129	0.049	0.042
	(0.085)*	(0.089)	(0.075)	(0.077)
Latino-white diss index × Latino	-0.670	-0.334	-0.293	-0.131
	(0.091)***	(0.084)***	(0.072)***	(0.073)*
CBSA fixed-effects				
Latino-white lagged dissimilarity index	0.146	0.083	0.072	0.036
	(0.081)*	(0.073)	(0.053)	(0.049)
Latino-white lagged diss index × Latino	-0.653	-0.309	-0.264	-0.093
	(0.073)***	(0.054)***	(0.044)***	(0.034)***
Education categories	No	Yes	No	Yes
English proficiency	No	Yes	No	Yes
Observations	1,276,664	1,276,664	1,216,458	1,216,458
Number of CBSAs	187	187	187	187
Pooled OLS				
Black-white lagged dissimilarity index	0.047	-0.0002	-0.008	-0.038
	(0.071)	(0.064)	(0.069)	(0.063)
Black-white lagged diss index × black	-0.404	-0.223	-0.012	0.081
	(0.091)***	(0.074)***	(0.063)	(0.055)
Pooled OLS				
Black-white dissimilarity index	0.053	-0.024	-0.005	-0.053
	(0.086)	(0.074)	(0.083)	(0.075)
Black-white diss index × black	-0.406	-0.204	0.051	0.153
	(0.098)***	(0.080)**	(0.070)	(0.061)**
CBSA fixed-effects				
Black-white lagged dissimilarity index	0.122	0.118	0.037	0.043
	(0.102)	(0.102)	(0.065)	(0.067)
Black-white lagged diss index × black	-0.436	-0.271	-0.082	0.001
	(0.078)***	(0.060)***	(0.048)*	(0.038)
Education categories	No	Yes	No	Yes
English proficiency	No	Yes	No	Yes
Observations	1,307,648	1,307,648	1,237,895	1,237,895
Number of CBSAs	184	184	184	184

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. The same sample composition criteria and additional controls specified in notes of table 5 apply.

5. Discussion and conclusion

In summary, in our models with metropolitan-area fixed effects, segregation has only a weak or non-existent association with the outcomes of whites, but it has a strong, negative association with the educational and labor market outcomes of Latinos and blacks. As the level of segregation in a metropolitan area increases, the socio-economic outcomes of black and Latino young adults living in that metropolitan area deteriorate both absolutely and relative to whites. Among Latinos, segregation has a particularly negative association with the outcomes of young adults of Puerto Rican and Dominican ancestry.

The instrumental variables results generally confirm the negative effects of segregation on black and Latino young adults' employment outcomes and indicate that, if anything, the OLS results understate the negative effects of segregation. These findings suggest that Latinos in more segregated metropolitan areas have developed some means to mitigate the negative consequences of segregation, yet, despite this attenuation, segregation's effects remain large.

The instrumental variables results also suggest that whites in metropolitan areas with higher levels of segregation are more likely to graduate from college. These positive effects of segregation on whites are most apparent in 2010, while the negative effects of segregation on Latino outcomes exhibit little variation between 2000 and 2010. Higher levels of black-white segregation also appear to lead to improved labor market outcomes for whites. On the one hand, whites may benefit from segregation through the opportunity it affords to hoard resources, such as access to high-performing schools or neighborhoods with more highly educated peers. On the other hand, economic opportunities may happen to be greater and labor markets more robust in areas with greater levels of segregation, but Latinos and blacks may be unable to access those benefits because of the physical and social barriers that residential segregation creates.

In short, our work makes clear that segregation heightens inequality between whites and Latinos. While the precise mechanisms are unclear, we provide suggestive evidence that part of the story is that residential segregation appears to lead to both differential exposure to growing industries as well as differential exposure to peers and social networks, as proxied by neighbors' poverty and educational attainment.

References

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

Antecol, Heather and Deborah A. Cobb-Clark. 2008. Racial and ethnic discrimination in local consumer markets: Exploiting the army's procedures for matching personnel to duty locations. *Journal of Urban Economics* 64(2):496–509.

Autor, David and David Dorn. 2013. The growth of low skill service jobs and the polarization of the US labor market. *American Economic Review* 103(5):1553–1597.

Bayer, Patrick, Hanming Fang, and Robert McMillan. 2014. Separate when equal? Racial inequality and residential segregation. *Journal of Urban Economics* 82:32–48.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy* 116(6):1150–1196.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross. 2011. The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy* 3(1):25–53.

Borjas, George. 1995. Ethnicity, neighborhoods, and human capital externalities. *American Economic Review* 85(3):365–390.

Brueckner, Jan K. and Stuart S. Rosenthal. 2009. Gentrification and neighborhood housing cycles: will America's future downtowns be rich? *Review of Economics and Statistics* 91(4):725–743.

Card, David and Jesse Rothstein. 2007. Racial segregation and the black-white test score gap. *Journal of Public Economics* 91(11–12):2158–2184.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics* 129(4):1553–1623.

Collins, Chiquita A. and David R. Williams. 1999. Segregation and mortality: The deadly effects of racism? *Sociological Forum* 14(3):495–523.

Collins, William J. and Robert A. Margo. 2000. Residential segregation and socioeconomic outcomes: When did ghettos go bad? *Economics Letters* 69(2):239–243.

Cutler, David M. and Edward L. Glaeser. 1997. Are ghettos good or bad? *Quarterly Journal of Economics* 112(3):827–872.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. When are ghettos bad? Lessons from immigrant segregation in the United States. *Journal of Urban Economics* 63(3):759–774.

De la Roca, Jorge, Ingrid Gould Ellen, and Katherine M. O'Regan. 2014. Race and neighborhoods in the 21st century: What does segregation mean today? *Regional Science and Urban Economics* 47:138–151.

Duncan, Brian and Stephen J. Trejo. 2011. Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans. *Journal of Labor Economics* 29(2):195–227.

Durlauf, Steven N. 2004. Neighborhood effects. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2173–2242.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment. *Quarterly Journal of Economics* 118(1):329–357.

Ellen, Ingrid Gould. 2000. Is segregation bad for your health? The case of low birth weight. *Brookings-Wharton Papers on Urban Affairs* :203–238.

Epple, Dennis and Richard E. Romano. 2011. Peer effects in education: a survey of the theory and evidence. In Jess Benhabib, Alberto Bisin, and Matthew O. Jackson (eds.) *Handbook of Social Economics*, volume 1. Amsterdam: North-Holland, 1053–1163.

Graham, Bryan S. 2016. Identifying and estimating neighborhood effects. Working Paper 22575, National Bureau of Economic Research.

Hoxby, Caroline M. 2000. Does competition among public schools benefit students and taxpayers? *American Economic Review* 90(5):1209–1238.

Kain, John F. 1968. Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics* 82(2):175–197.

Logan, John R. and Richard N. Turner. 2013. Hispanics in the United States: Not only Mexicans. Report, US2010 Project.

Lou, Tian and Tao Song. 2017. Ethnic segregation, education, and immigrants' labor market outcomes. Processed, University of Connecticut.

Loury, Glenn C. 1977. A dynamic theory of racial income differences. In Phyllis A. Wallace and Anette M. LaMond (eds.) *Women, Minorities and Employment Discrimination*. Lexington, MA: Lexington Books, 153–186.

Ludwig, Jens, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade. 2011. Neighborhoods, obesity and diabetes—A randomized social experiment. *New England Journal of Medicine* 365(16):1509–1519.

Pew Research Center. 2015. Modern immigration wave brings 59 million to U.S., driving population growth and change through 2065: Views of immigration's impact on U.S. society mixed. Technical report, Pew Research Center, Washington, DC.

Portes, Alejandro and Julia Sensenbrenner. 1993. Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology* 98(6):1320–1350.

Rosenthal, Stuart S. and Stephen L. Ross. 2015. Change and persistence in the economic status of neighborhoods and cities. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5. Amsterdam: North-Holland, 1047–1120.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. *Integrated Public Use Microdata Series: Version 6.0*. Minneapolis, MN: University of Minnesota.

Sharkey, Patrick, Amy E. Schwartz, Ingrid Gould Ellen, and Johanna Lacoe. 2014. High stakes in the classroom, high stakes on the street: The effects of community violence on student's standardized test performance. *Sociological Science* 1:199–220.

Steil, Justin P., Jorge De la Roca, and Ingrid Gould Ellen. 2015. Desvinculado y desigual: Is segregation harmful to Latinos? *The Annals of the American Academy of Political and Social Science* 660(1):57–76.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In James H. Stock and Donald W. K. Andrews (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*. Cambridge: Cambridge University Press, 109–120.

Tiebout, Charles M. 1956. A pure theory of local expenditures. *Journal of Political Economy* 64(5):416–424.

United States Bureau of the Census. 2010. Hispanic or Latino by specific origin. 2010 Decennial Census Summary File 1. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_SF1_PCT11&prodType=table.

United States Bureau of the Census. 2013. Current Population Survey. Annual Social and Economic Supplement (ASEC) Supplement Survey, 2013. Retrieved from https://www.census.gov/topics/population/hispanic-origin/data/tables.2013.html.

United States Department of Commerce, Bureau of the Census. 2001. *Census of Governments, 1962 and 1967*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, https://doi.org/10.3886/ICPSR00017.V1.

Weicher, John C. and Thomas G. Thibodeau. 1988. Filtering and housing markets: An empirical analysis. *Journal of Urban Economics* 23(1):21–40.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Appendix A. Assigning individuals in IPUMS to a Core Based Statistical Area (CBSA)

Barriers to using PUMAs

The smallest level of geography that can be identified in the Integrated Public Use Microdata Sample (IPUMS) is the Public Use Microdata Area (PUMA). PUMAs contain no fewer than 100,000 residents, are nested within state boundaries, and cover the entire United States. The Census Bureau oversees the delineation of PUMAs, which are redrawn for every decennial census. Because PUMAs boundaries are not consistent across decennial censuses, they cannot be used to identify location patterns over time.

Barriers to using MSAs or CBSAs

Although IPUMS assigns observations to a 'metropolitan area,' the boundaries of any particular metropolitan area change between censuses (generally becoming larger as the region grows). Therefore, we need to construct geographic units that generally correspond to metropolitan areas and are identical across the 1990, and 2000 Decennial Censuses, as well as the 2007–2011 American Community Survey's 5-year estimates.

Consistent PUMAs

We use publicly available IPUMS 5 percent unweighted samples. IPUMS-USA assigns every individual observation to a PUMA and a 'Consistent PUMA.' Consistent PUMAs are the smallest geographic units that do not change over time. Therefore, we can use them to link locations in the decennial censuses and ACS samples used in this study. The Consistent PUMAs contain no fewer than 400,000 residents, making them a significantly larger unit of geography than PUMAs.

The Missouri Census Data Center provides a crosswalk from 2000 PUMA delineations to 2008 Cored Based Statistical Areas (CBSAs). The crosswalk generates allocation factors for each PUMA, indicating what percentage of the PUMA population falls within a particular CBSA. We assign an entire PUMA to a CBSA if more than 50% of the PUMA's population lives within that CBSA. As a result, we build up a crosswalk that assigns each 2000 PUMA to a single CBSA or non-metropolitan area. A total of 1,694 PUMAS are linked to 353 CBSAS.

As discussed, PUMAS cannot be used across years due to their boundaries being redrawn in each census. However, given that all 2000 PUMAS also have a Consistent PUMA identifier, we can create a crosswalk that links Consistent PUMAS to CBSAS. Using this crosswalk we can then assign individual observations in the 1990 Census and the 2007–2011 ACS data to a CBSA through their Consistent PUMA identifier.

Assigning Consistent PUMAS to CBSAS requires an additional level of assumptions. While some Consistent PUMAS are fully contained in large metropolitan areas, other Consistent PUMAS overlap several metropolitan areas or non-metropolitan areas. Using our crosswalk of 2000 PUMAS to 2008 CBSAS and the population of each PUMA in 2000, we can calculate the share of the population for each Consistent PUMA that lives in a particular CBSA. We then assign the whole population of the Consistent PUMA to the CBSA with the highest population share (as long as this share exceeds 50 percent of the Consistent PUMA's population). We thus exclude those Consistent PUMAS where no single CBSA contains more than 50 percent of the Consistent PUMA population. Of the 453 Consistent PUMAS that encompass some portion of a CBSA, we end up with 392 Consistent PUMAS in 242 CBSAS. Most of the CBSAS that we lose have small populations.

The assignment of individuals to CBSAS through Consistent PUMA identifiers results in geographic areas that are comparable over time. However, the CBSA boundaries we obtain do not correspond exactly to census CBSA boundaries. For example, a Consistent PUMA may have 51% of its population within a CBSA but 49% within a non-metropolitan area; our method would essentially redraw the CBSA boundary to include this non-metropolitan area. Likewise, if a Consistent PUMA overlaps two CBSAS, then our method would entirely allocate this Consistent PUMA to a single CBSA as long as the share of its population in this CBSA exceeds 50%. Furthermore, some parts of CBSAS are ignored when Consistent PUMAS partially overlap multiple geographic areas and do not have half or more of their population within any one CBSA.

One potential concern is that the use of Consistent PUMAS may result in improper assignment of observations to CBSAS due to the large size of Consistent PUMAS. We can examine the extent of this problem for the year 2000. As already mentioned, the crosswalk provided by the Missouri Census Data Center links 2000 IPUMS data to CBSAS. We estimate our specifications for 2000 using both PUMA to CBSA assignment and the coarser Consistent PUMA to CBSA assignment. The two methods give generally consistent results.

Appendix B. Online appendix

Table B.10: First-stage estimation of lagged dissimilarity indices, 1980

Dependent variable:	1980 Latino-white dissimilarity index			1980 black-white dissimilarity index		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.547		0.455	-0.012		-0.135
	(0.115)***		(0.126)***	(0.108)		(0.100)
Log of local governments 1962		0.003	-0.005		0.013	0.014
		(0.014)	(0.015)		(0.011)	(0.011)
% of revenue from transfers 1962		-0.610	-0.440		-0.382	-0.423
		(0.171)***	(0.158)***		(0.104)***	(0.105)***
Black-white DI 1980	0.298	0.078	0.168			
	(0.098)***	(0.139)	(0.112)			
Log population	-0.018	0.015	-0.005	0.069	0.050	0.054
	(0.015)	(0.018)	(0.019)	(0.009)***	(0.013)***	(0.013)***
% black population	-0.004	0.398	0.135	0.151	0.277	0.344
	(0.198)	(0.212)*	(0.198)	(0.101)	(0.094)***	(0.094)***
% Latino population	0.928	0.827	0.854	-0.122	-0.069	-0.079
	(0.135)***	(0.161)***	(0.148)***	(0.094)	(0.113)	(0.107)
% Asian population	-0.445	-0.255	-0.399	-0.321	-0.141	-0.096
	(0.238)*	(0.276)	(0.243)	(0.129)**	(0.138)	(0.144)
% population over 65	1.887	2.372	1.741	0.486	0.120	0.281
	(0.572)***	(0.527)***	(0.538)***	(0.450)	(0.406)	(0.396)
% population under 15	0.851	1.257	0.817	-0.141	-0.328	-0.229
	(0.579)	(0.529)**	(0.523)	(0.534)	(0.462)	(0.469)
% unemployed	1.085	1.329	0.826	0.853	0.622	0.777
	(0.904)	(0.890)	(0.904)	(0.734)	(0.747)	(0.723)
% working in manufacturing	0.620	0.665	0.777	0.425	0.461	0.416
	(0.255)**	(0.305)**	(0.261)***	(0.242)*	(0.245)*	(0.248)*
% in poverty status	-0.422	-0.547	-0.190	0.089	0.181	0.070
	(0.488)	(0.554)	(0.511)	(0.443)	(0.449)	(0.448)
% with bachelor's degree	0.478	0.789	0.491	-0.388	-0.379	-0.291
	(0.361)	(0.393)**	(0.339)	(0.237)	(0.226)*	(0.225)
Log median household income	0.034	0.069	0.036	0.011	-0.013	-0.003
	(0.046)	(0.049)	(0.043)	(0.027)	(0.023)	(0.027)
Observations	511,547	511,547	511,547	546,378	546,378	546,378
Number of CBSAs	142	142	142	147	147	147
R^2	0.739	0.716	0.758	0.648	0.704	0.711

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications include a constant term, a female indicator variable and age indicator variables. Columns (1)–(3) include Latino ancestry-group indicator variables and CBSA controls interacted with a Latino indicator variable. Columns (4)–(6) include a black indicator variable and CBSA controls interacted with a black indicator variable. DI stands for dissimilarity index.

Table B.11: First-stage estimation of lagged dissimilarity indices, 1990

Dependent variable:	1990 Latino-white dissimilarity index			1990 black-white dissimilarity index		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.478		0.432	0.266		0.137
	(0.119)***		(0.128)***	(0.080)***		(0.079)*
Log of local governments 1962		0.002	-0.004		0.035	0.033
		(0.016)	(0.016)		(0.011)***	(0.011)***
% of revenue from transfers 1962		-0.401	-0.269		-0.279	-0.230
		(0.155)***	(0.143)*		(0.097)***	(0.099)**
Black-white DI 1990	0.299	0.300	0.241			
	(0.096)***	(0.129)**	(0.108)**			
Log population	-0.017	0.001	-0.010	0.064	0.038	0.033
	(0.012)	(0.016)	(0.014)	(0.009)***	(0.013)***	(0.013)**
% black population	-0.016	0.168	0.059	0.158	0.382	0.344
	(0.130)	(0.128)	(0.120)	(0.096)*	(0.094)***	(0.094)***
% Latino population	0.799	0.666	0.745	-0.098	0.002	0.032
	(0.110)***	(0.116)***	(0.113)***	(0.088)	(0.108)	(0.108)
% Asian population	-0.382	-0.220	-0.354	-0.299	0.043	0.004
	(0.155)**	(0.182)	(0.168)**	(0.115)***	(0.128)	(0.135)
% population over 65	1.977	2.288	1.824	1.239	1.279	1.127
	(0.628)***	(0.610)***	(0.620)***	(0.548)**	(0.498)**	(0.494)**
% population under 15	0.951	1.382	0.915	-0.312	-0.067	-0.175
	(0.713)	(0.762)*	(0.678)	(0.849)	(0.724)	(0.719)
% unemployed	0.504	1.527	0.481	-0.266	-0.421	-0.780
	(1.151)	(1.024)	(1.096)	(1.071)	(0.858)	(0.864)
% working in manufacturing	0.678	0.540	0.758	1.024	0.680	0.731
	(0.213)***	(0.268)**	(0.253)***	(0.231)***	(0.256)***	(0.248)***
% in poverty status	0.126	0.055	0.256	0.187	0.083	0.144
	(0.400)	(0.421)	(0.442)	(0.471)	(0.447)	(0.427)
% with bachelor's degree	0.669	1.005	0.686	-0.164	-0.090	-0.184
	(0.262)**	(0.305)***	(0.260)***	(0.186)	(0.179)	(0.174)
Log median household income	0.084	0.115	0.087	-0.047	-0.047	-0.055
	(0.057)	(0.059)**	(0.052)*	(0.030)	(0.028)*	(0.029)*
Observations	396,363	396,363	396,363	415,144	415,144	415,144
Number of CBSAs	142	142	142	147	147	147
R^2	0.766	0.733	0.774	0.703	0.747	0.753

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications include a constant term, a female indicator variable and age indicator variables. Columns (1)–(3) include Latino ancestry-group indicator variables and CBSA controls interacted with a Latino indicator variable. Columns (4)–(6) include a black indicator variable and CBSA controls interacted with a black indicator variable. DI stands for dissimilarity index.

Table B.12: First-stage estimation of lagged dissimilarity indices, 2000

Dependent variable:	2000 Latino-white dissimilarity index			2000 black-white dissimilarity index		
	(1)	(2)	(3)	(4)	(5)	(6)
Single/multi-family housing DI 1970	0.281 (0.093)***		0.277 (0.106)***	0.465 (0.104)***		0.241 (0.106)**
Log of local governments 1962		0.007 (0.014)	-0.0006 (0.015)		0.043 (0.012)***	0.033 (0.012)***
% of revenue from transfers 1962		-0.094 (0.160)	-0.022 (0.158)		-0.258 (0.095)***	-0.190 (0.104)*
Black-white DI 2000	0.237 (0.078)***	0.285 (0.091)***	0.234 (0.091)***			
Log population	0.003 (0.012)	0.002 (0.013)	0.003 (0.013)	0.052 (0.009)***	0.020 (0.014)	0.020 (0.013)
% black population	0.060 (0.133)	0.115 (0.129)	0.068 (0.134)	0.353 (0.108)***	0.557 (0.125)***	0.493 (0.122)***
% Latino population	0.546 (0.101)***	0.463 (0.122)***	0.541 (0.120)***	0.081 (0.100)	0.073 (0.111)	0.134 (0.109)
% Asian population	-0.357 (0.184)*	-0.339 (0.208)	-0.356 (0.194)*	-0.193 (0.129)	-0.014 (0.170)	-0.026 (0.155)
% population over 65	1.452 (0.588)**	1.896 (0.592)***	1.456 (0.604)**	0.908 (0.808)	1.500 (0.630)**	1.087 (0.645)*
% population under 15	1.245 (0.689)*	1.425 (0.705)**	1.252 (0.694)*	-0.198 (0.942)	-0.037 (0.841)	-0.106 (0.778)
% unemployed	-0.322 (0.669)	-0.840 (0.687)	-0.339 (0.701)	0.677 (0.700)	0.104 (0.646)	0.522 (0.665)
% working in manufacturing	0.868 (0.197)***	0.713 (0.241)***	0.877 (0.243)***	1.174 (0.258)***	0.697 (0.263)***	0.805 (0.260)***
% in poverty status	0.377 (0.540)	1.071 (0.632)*	0.399 (0.585)	-0.123 (0.774)	0.861 (0.633)	0.259 (0.656)
% with bachelor's degree	0.521 (0.281)*	0.344 (0.292)	0.521 (0.270)*	0.412 (0.283)	0.046 (0.259)	0.213 (0.269)
Log median household income	0.148 (0.169)	0.393 (0.174)**	0.152 (0.162)	-0.300 (0.173)*	0.119 (0.138)	-0.096 (0.162)
Observations	404,693	404,693	404,693	397,549	397,549	397,549
Number of CBSAs	142	142	142	147	147	147
R^2	0.677	0.659	0.677	0.726	0.751	0.763

Notes: Coefficients are reported with robust standard errors in parenthesis, which are clustered by Core Based Statistical Area (CBSA). ***, **, and * indicate significance at the 1, 5, and 10 percent levels. All specifications include a constant term, a female indicator variable and age indicator variables. Columns (1)–(3) include Latino ancestry-group indicator variables and CBSA controls interacted with a Latino indicator variable. Columns (4)–(6) include a black indicator variable and CBSA controls interacted with a black indicator variable. DI stands for dissimilarity index.