

Identifying and estimating neighborhood effects

Bryan S. Graham*

August 24, 2016

Abstract

Residential segregation by race and income are enduring features of urban America. Understanding the effects of residential segregation on educational attainment, labor market outcomes, criminal activity and other outcomes has been a leading project of the social sciences for over half a century. This paper describes techniques for measuring the effects of neighborhood of residence on long run life outcomes.

*Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3888 and the National Bureau of Economic Research, e-mail: bgraham@econ.berkeley.edu, web: <http://bryangraham.github.io/econometrics/>. I am thankful to the editor, Steven Durlauf, for encouragement, support, intellectual input and patience. The final version of the paper reflects many improvements suggested by the six referees, each of whom prepared detailed and thoughtful reports. Financial support from NSF grant SES #1357499 is gratefully acknowledged. All the usual disclaimers apply. The stylized facts reported below are largely based on confidential geocoded versions of the National Longitudinal Survey of Youth 1979 and 1997 cohorts (i.e., NLSY79 & NLSY97). Details for replicating these statistics can be found in Graham and Sharkey (2013) and on the author's webpage.

Segregation is a pervasive and enduring feature of American society. Americans tend to live near, and attend school with, those with similar educational, socioeconomic, and racial backgrounds. In 2000 the average Black individual lived in a city where over 50 percent of her immediate neighbors were Black, although just 20 percent of residents city-wide were.¹ She was two and one half times as likely to have a Black neighbor relative to a counterfactual world without residential segregation by race.

Segregation, to borrow a phrase from Durlauf (1996b), generates associational inequality across identifiable sub-groups. The typical Black adolescent in America lives in a very different type of neighborhood, attends a very different type of school, and is embedded in a very different type of social network, than her White counterpart. Almost a half century after the civil rights era these differences remain poignant: an adolescent Black male at the turn of the century was three times as likely to have reported witnessing a shooting than his White counterpart (see Table 4 below).

Segregation, by making the lived experiences of identifiable sub-groups, on average, different is itself a type of inequality. Minority children tend to grow up in different social environments, than their White counterparts. This observation, while self-evident, is nevertheless important and, perhaps, too infrequently made. Society may legitimately have direct concerns about associational inequality. More often, however, concerns are indirect: worries about segregation are related to beliefs about its effects on other outcomes. By generating inequality in social networks and public goods important for the acquisition of human capital, segregation, so the argument goes, contributes to inequality in earnings, wealth and other outcomes. This argument has been frequently and elegantly made (e.g., Loury 1977, 2002). Its quantitative assessment has been a leading, and controversial, project of the social sciences at least since the Coleman report (Coleman et al., 1966).

Table 1 reports early adult outcomes for a subsample of White, Black and Hispanic respondents in the National Longitudinal Survey of Youth 1997 (NLSY97) cohort. This cohort was born between 1980 and 1984, went through adolescence in the late 1990s, and is now in their 30s. The subsample includes respondents living in 44 metro areas of the United States; these metro areas are broadly representative of urban America, especially as it exists in larger cities.² Black-White and Hispanic-White early adult outcome gaps are substantial. Blacks fail to complete high school at twice the rate of, complete college at only one-third

¹ Author's calculations using the *Neighborhood Change Database*. Relevant population is Blacks residing in 263 Metropolitan Statistical Areas (MSAs) in 2000.

² The NLSY97 sample consists of 8,984 respondents. Of these 7,263 resided in one of 126 different MSAs at baseline in 1997. The analytic sub-sample discussed in the text includes the 4,419 respondents that, at baseline, lived in an MSA with at least 5 White NLSY97 respondent households as well as at least 5 Black and/or 5 Hispanic respondent households.

the rate of, score almost one standard deviation lower on the Armed Forces Qualification Test (AFQT), earn about 74 percent less, and are incarcerated at almost three times the rate of, their White counterparts (Columns 1 and 2 of Table 1).³ Hispanic-White gaps in these outcomes are also substantial, albeit less pronounced.

The facts presented in Table 1 are at variance with many Americans own beliefs about mobility. In a 2011 survey conducted by the Pew Charitable Trust, 47 percent of respondents strongly agreed that economic success mostly depends on hard work and drive, while only 27 percent of respondents identified race and/or class background as severe limitations to upward mobility (Pew Charitable Trusts, 2011).⁴

From the vantage point of a social scientist, the gaps reported in Table 1 present a puzzle. The conjunction of large measured differences in outcomes across subgroups, and the relative isolation of these subgroups from one another due to residential segregation, motivates a focus on contextual or neighborhood effects as a possible driver of inter-group inequality. The hypothesis, in a very loose sense, is that “place” matters.

In thinking about this hypothesis it is useful to make three, policy-relevant, distinctions. Place may matter because expenditure per pupil, teacher quality, access to good hospitals and proximity to well-paying jobs varies across neighborhoods. In principle these types of neighborhood inequalities can be ameliorated by transferring resources across space. The wave of school finance reforms that began with the California Supreme Court’s ruling in *Serrano v. Priest* in 1971, generated transfers of this type. President Obama’s “Promise Zones” initiative also represents an effort to channel more resources into distressed neighborhoods. Kline and Moretti (2014) review the economics of “place-based” policies.

Second place may matter because the characteristics and behaviors of our neighbors directly influence key life outcomes. If employment depends, in part, on information and referrals from friends and neighbors, then living in a disadvantaged neighborhood, where few people are stably employed, makes the acquisition of a job that much harder. If learning depends, in part, on being surrounded by able peers, then a child in a classroom of advantaged children, will learn more quickly than an identical child in a classroom of disadvantaged children. These types of neighborhood effects are *peer group effects* or *social interaction* effects. This source of inequality can *not* be ameliorated by transferring financial resources across space.

³The reported earnings and incarceration statistics are for male respondents only.

⁴The exact wording of the question was: “Which of the following comes closest to your own point of view: A. Whether or not a person gets ahead economically in this country depends on drive, hard work and obtaining the right skills; OR B. Circumstances beyond any person’s control - a person’s race, gender, and parents as well as the quality of education available where they happen to live can severely limit a person’s ability to get ahead economically?” Respondents could strongly or not so strongly agree with either A or B; or say they didn’t know or were not sure.

Table 1: Early adult outcomes for MSA-resident White, Black and Hispanic youth born 1980 to 1984

	(1) White	(2) Black-White Difference	(3) Hispanic-White Difference
Youth Outcomes			
<i>Females & Males (N = 4,419)</i>			
Dropout/GED at age 24	0.1329 (0.0081)	0.1399 (0.0192)	0.1119 (0.0180)
College graduate at age 24	0.2746 (0.0109)	-0.1533 (0.0161)	-0.1761 (0.0151)
AFQT score at “age” 16	0.2552 (0.0256)	-0.8509 (0.0491)	-0.6885 (0.0491)
<i>Panel A: Males only (N = 2,292)</i>			
Log earnings (2007 to 2012)	10.3159 (0.0310)	-0.7435 (0.0792)	-0.1980 (0.0581)
Incarcerated by age 28	0.0766 (0.0083)	0.1198 (0.0223)	0.0396 (0.0178)

Source: National Longitudinal Survey of Youth 1997 (NLSY97) and author’s calculations..

Notes: Each row corresponds to a least squares fit of the listed outcome onto a constant, a Black dummy, and a Hispanic Dummy (with baseline NLSY97 sampling weights used). The NLSY97 sample consists of 8,984 youths, of which a total of 7,263 resided in an MSA at baseline (1,796 Hispanic respondents, 1,943 Black respondents, and 3,524 non-Hispanic, non-Black respondents). The estimation sample used here includes the 4,419 respondents who resided in a MSA with at least 5 White responding households as well as 5 Black and/or 5 Hispanic responding households (Panel A). Panel B is based on the subset of 2,292 male respondents. Standard errors, clustered at the household level, are reported in parentheses. “AFQT score at age 16” corresponds to the inverse normal CDF transform of the adjusted AFQT percentile score used in Altonji, Bharadwaj and Lange (2012). This score was normalized to the distribution of percentile scores across NLSY79 respondents aged 16 at the time of test-taking in 1980. Across respondents from the reference population of American Youth aged 15 to 23 in 1980 this transform of percentile scores is mean zero with unit variance. Since the Altonji, Bharadwaj and Lange (2012) percentile scores are normalized to a different reference group (both in terms of age-of-testing and year-of-birth), “AFQT score at age 16” need not be mean zero with unit variance (across all 7,002 respondents with valid AFQT scores its mean is -0.0001 and its standard deviation is 0.9893). AFQT and earnings were only available for a subset of the target sample. The AFQT row of Panel A is based on 3,291 respondents, while the earnings row of Panel B is based on the 1,879 male respondents with non-zero earnings.

Reducing peer group inequality requires people to move across space; in the words of Durlauf (1996b), *associational redistribution*.⁵

Third place may matter because it represents a partially isolated arena of social interaction. If individual decisions – to study hard, use drugs or engage in delinquent behavior – depend, in part on those of neighbors/peers, then there may exist multiple equilibrium distributions of actions within a neighborhood (Brock and Durlauf, 2001a,b, 2007). Across two ex ante identical neighborhoods, one may converge to a low drug use equilibria, another a high use equilibria. In the presence of multiple equilibria policies which attempt to engineer “equilibrium shifts” can change the distribution of outcomes within a neighborhood even in the absence of resource transfers or associational redistribution. Of course, the effects of these first two types of interventions may also vary with the structure of social interaction among neighborhood residents.

My typology mirrors that given by Manski (1993). Place may matter because of *(i) correlated*, *(ii) exogenous/contextual* and/or *(iii) endogenous* effects. In this paper I principally review econometric methods for the identification and estimation of neighborhood effects of the second type.⁶ The focus is methodological and no comprehensive review of the empirical literature is attempted, although empirical work will feature prominently in the exposition. Although peer group effects have been studied in other settings, most notably schools (e.g., Angrist and Lang, 2004; Graham, 2008; see Epplé and Romano (2011) for a recent review), my focus will be on their operation in the context of neighborhood of residence. This focus is partly expositional and partly substantive. Its expositional value is that it provides a common running example throughout what follows. The focus is substantive because sorting into neighborhoods is mediated by the housing market, for which we observe a price. Other types of sorting, for example into marriage and friendship, generally occurs outside a formal economic market. Studying the process of matching and network formation in these settings, as well as its effects on outcomes, is a topic of great interest but beyond the scope of this paper (see Goldsmith-Pinkham and Imbens (2013) and Graham (2015b) for two recent treatment of problems of this type and Graham (2011, 2015a) for surveys of other research). Nevertheless, many of the points made below will be relevant to researchers interested in “peer effects” broadly defined.

⁵See Piketty (2000, p. 466) for a related discussion. A referee made the valid point that while eliminating peer group inequality *itself* requires associational redistribution, other types of resource transfers could ameliorate the *effects* of peer group inequality. An example of such a transfer is Title I funding, which is provided to school districts with large numbers of economically disadvantaged students.

⁶The study of endogenous interactions raises distinct issues, some of which are closely related to those addressed by the burgeoning literature on the econometrics of games (see Aureo de Paula (2013) for a review). My omission of this material is not meant to suggest its irrelevance, merely that a careful treatment is best left for a separate survey. See Section 4 below for some additional comments.

Section 1 presents basic information on the scale of residential segregation by race in the contemporary United States. This section necessarily briefly reviews techniques for *measuring* segregation as well. After documenting the scale of residential segregation, I present, in Section 2, a series of correlations – or stylized facts – on its “effects” using geocoded data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY).

I establish (i) that the social contexts in which Black, Hispanic and White adolescents are embedded differ substantially (on average) and (ii) that these “contextual gaps” are larger for youth who reside in segregated cities (and, conversely, smaller for those who reside in more integrated cities).⁷ While the second fact follows almost axiomatically from the first, documenting them motivates the search for a causal connection between neighborhood of residence and long run life outcomes.

Having established the degree of residential segregation and its correlation with contextual measures, Section 3 turns to a review of the evidence on what drives segregation. I do *not* discuss the array of institutions, laws, and policies which have historically exerted a strong influence on neighborhood- and school-level stratification. Sharkey (2012) provides a recent synopsis of this material. Other useful references include Massey and Denton (1993), Schill and Wachter (1995) and Dawkins (2004). Instead, my discussion is on explanations for segregation which conceptualize it as an equilibrium outcome; one where overt discrimination does not play a role (e.g., Benabou, 1993, 1996; Durlauf 1996a; Epple and Platt, 1998; Nesheim, 2002; Fernandez, 2003). My intent is not to discount the historical or continuing relevance of “non-market” drivers of sorting (cf., Yinger, 1999; Ross and Yinger, 2002), but rather to narrow the focus to a domain where economics has a disciplinary advantage.

The focus on choice-based explanations of segregation also foreshadows my discussion of identification. In identifying the effect of neighborhoods on outcomes, the econometrician must confront the fact that the information set available to individuals (or their parents) when choosing their residence is generally larger than that available to her when she attempts to model their effects. Consequently, observed neighborhood characteristics may be correlated with unobserved individual characteristics, which themselves drive the outcome of interest (e.g., Mayer and Jencks, 1989; Manski, 1993).

Fortunately, the empirically-oriented economist can lean on a rich theoretical literature on locational sorting in order to build intuitions and guide research design. I attempt to pro-

⁷A note on nomenclature: first, in what follows I use the word “White” to denote any non-Hispanic, non-Black, individual. Thus “Whites” include Asian-Americans as well as Native Americans. Second, by “city” I mean a metropolitan statistical area (MSA) as defined by Office of Management and Budget (OMB). I use 1999 OMB definitions (with a few minor differences described in Graham and Sharkey (2013)). MSAs include central city cores as well as adjacent suburbs. They provide a convenient approximation to a local labor and housing market.

vide a synoptic view of this research which foreshadows the discussion of identification and estimation in Section 4.

Section 4 provides an overview of approaches to identifying and estimating neighborhood effects. Throughout I focus on measuring the effects of neighborhood racial composition on outcomes. Of course, residential stratification also occurs along other dimensions, most notably by income (e.g., Watson, 2009; Reardon and Bischoff, 2011). My focus on residential segregation by race is pedagogically motivated. Other types of segregation may be equally important or, more interestingly, interact with racial segregation. I ignore these and other complications.

I divide my discussion into two parts based on the type of data available to the researcher. First, I consider the case where the researcher has access to matched individual-neighborhood data, typically from a single city. A prototypical dataset of this type is the Los Angeles Family and Neighborhood Survey (L.A. FANS), which collects household- and neighborhood-level data from Los Angeles families. These data can be matched with tract level data from the U.S. Census (Sastry and Pebley, 2010). The data collected in conjunction with the Project on Human Development in Chicago Neighborhoods (PHDCN) is also of this type (Sampson, Sharkey and Raudenbush, 2008). I also discuss what can, and can not, be learned from housing mobility experiments, such as the Moving to Opportunity (MTO) program.

Second, I consider the case where the researcher has access to matched individual-city data, typically for many cities. Cutler and Glaeser (1997) provide an early and well-known example of this research design. Here the analysis involves correlating individual-level outcomes with city-level measures of sorting/segregation. This literature typically argues that aggregation reduces biases due to sorting (e.g., Card and Rothstein, 2007), I make this argument precise.

Section 4 briefly touches on other approaches to identifying neighborhood effects, such as methods based on structural locational sorting models. Section 5 summarizes my thoughts about what directions future work should take. No comprehensive review of empirical work is attempted, nor do I review all available research designs. My goal is narrower. Specifically by formalizing a stylized version of what empirical neighborhood effects researchers have often done in *practice*, I attempt to illuminate the strengths and weakness of common research designs, and suggest incremental ways to improve upon existing work. Readers interested in a more comprehensive survey of existing empirical work in economics and the other social sciences are directed to Durlauf (2004, 2006), Durlauf and Ioannides (2010), Harding et. al. (2011), Sharkey (2012), Sampson (2012b) and Sharkey and Faber (2014).

Table 2: Ten most racially segregated metropolitan areas in 2000

	(1)	(2)	(3)	(4)
	DI	η^2	Frac. URM	I
Detroit, MI	0.790	0.673	0.263	0.759
Newark, NJ	0.724	0.593	0.365	0.741
Milwaukee-Waukesha, WI	0.721	0.537	0.225	0.641
Cleveland-Lorain-Elyria, OH	0.719	0.582	0.224	0.676
Buffalo-Niagara Falls, NY	0.714	0.522	0.151	0.594
New York, NY	0.712	0.591	0.513	0.801
Cincinnati, OH-KY-IN	0.698	0.477	0.145	0.552
St. Louis, MO-IL	0.687	0.531	0.203	0.626
Gary, IN	0.686	0.563	0.306	0.697
<i>All MSAs ($N = 319$)</i>	0.533	0.337	0.372	0.601
<i>Large MSAs ($N = 99$)</i>	0.560	0.376	0.381	0.622

Source: Neighborhood Change Database (NCDB) and author’s calculations.

Notes: Measures of segregation for Underrepresented Minorities (URMs, Blacks & Hispanics) relative to Whites/Asians. Measures of neighborhood segregation were computed for a total of 319 MSAs in the year 2000. Of these 319 MSAs, a total of 99 had at least 500,000 residents in 2000. These two groups of MSAs constitute “All MSAs” and “Large MSAs” in the table. They are comprised of a total of 223,967,756 and 175,232,556 residents respectively. Listed MSAs correspond to those with the most segregated neighborhoods according to the Dissimilarity Index (DI) described in the main text (Column 1). Column 2 reports the “eta-squared” (η^2) index, Column 3 the fraction of residents belonging to a URM, and Column 4 the Isolation Index (I).

1 Residential segregation by race in contemporary urban America

Table 2 lists, among those with a year 2000 population of at least 500,000, the ten most racially segregated metropolitan areas in the United States. The ranking is based on the Dissimilarity Index (DI), a widely used, albeit often misinterpreted, measure of segregation in social science research (Iceland, Weinberg and Steinmetz, 2002). In order to provide a precise definition of the DI it is helpful to establish some basic notation. I will also use this notation in the exposition of research designs in Section 4 below.

Consider the population of individuals residing in a given metropolitan area (a “city”). Let $T \in \{0, 1\}$ denote whether a random draw from this city is Minority ($T = 1$) or not ($T = 0$). Let Z be an $N \times 1$ vector with a 1 in the i^{th} row and zeros elsewhere when the random draw resides in the i^{th} neighborhood ($i \in \{1, \dots, N\}$). Here N denotes the number of neighborhoods in the city. Let $s(z) = \Pr(T = 1 | Z = z)$ be the proportion of neighborhood $Z = z$ that is Minority. For example if z has a one in its third row and zeros elsewhere, then

$s(z)$ gives the fraction of residents in the third neighborhood that are Minority. Finally let $Q = \Pr(T = 1)$ be the city-wide frequency of Minority residents. The Dissimilarity Index (DI) equals

$$\text{DI} = \frac{1}{2} \mathbb{E} \left[\left| \frac{s(Z)}{Q} - \frac{1-s(Z)}{1-Q} \right| \right], \quad (1)$$

where, to be clear, the expectation in (1) is over the population of residents (not neighborhoods).⁸ In a perfectly integrated city each neighborhood's racial composition will mirror that of the city as whole (i.e., $s(Z) = Q$ for all neighborhoods) so that (1) equals zero. In a perfectly segregated city, where no Minority shares a neighborhood with a White, $s(Z)$ will equal 1 with probability Q and 0 with probability $(1 - Q)$ so that $\text{DI} = \frac{1}{2} \left[Q \cdot \frac{1}{Q} + (1 - Q) \cdot \frac{1}{1-Q} \right] = 1$.

Appendix A demonstrates that the Dissimilarity Index equals the city-wide proportion of minority residents who would need to move in order to achieve perfect integration, *relative* to the proportion that would need to move under a status quo of perfect segregation (cf., Jakubs, 1977). Researchers routinely omit the *relative* qualifier (e.g., Cutler and Glaeser, 1997; p. 837).

Massey and Denton (1993, p. 20) suggest that a DI of less than 30 corresponds to a relatively integrated city, one between 30 and 60 to a moderately segregated city, and one greater than 60 to a highly segregated city. Another point of reference is provided by considering the degree of segregation in urban South Africa at the end of the Apartheid era (where residential segregation was forcibly imposed). Christopher (1994), computing Black vs. non-Black dissimilarity indices using 1991 South African Census data, finds values of approximately 0.9. Since South African Census enumeration areas contain only one-fourth as many people as U.S. Census Tracts (on average), we might except South African segregation measures to be somewhat inflated relative to their U.S. counterparts for purely statistical reasons. With these benchmarks and caveats in mind, the 0.79 Dissimilarity Index for the Detroit metro area in 2000 is extraordinary and substantiates the “two societies” language famously used by the Kerner Commission to describe urban America almost a half century ago.

The spatial distribution of Minority and White residents in the ten cities listed in Table 2 is decidedly uneven. In each of these cities Minorities and Whites largely live in separate

⁸Let M_i denote the number of residents in neighborhood i and $n = \sum_{i=1}^N M_i$ the city-wide population, then (1) coincides with

$$\text{DI} = \frac{1}{2} \frac{1}{n} \sum_{i=1}^N \sum_{t=1}^{M_i} \left| \frac{s(Z_i)}{Q} - \frac{1-s(Z_i)}{1-Q} \right| = \frac{1}{2} \sum_{i=1}^N \frac{M_i}{n} \left| \frac{s(Z_i) - Q}{Q(1-Q)} \right|,$$

which, in turn, accords with the definition given by Massey and Denton (1988, p. 284).

neighborhoods; neighborhoods with differing access to schools, parks and other public goods, differing exposure to violence, and where individuals' potential networks of peers also differ.

Table 2 also lists the “eta-squared” (η^2) and isolation index (I) measures of segregation. The former has been widely used to characterize sorting by economists following Maskin and Kremer (1996). It has a longer history in other disciplines (e.g., Farley, 1977) and will also arise frequently in the econometric analysis undertaken in Section 4 below. To give a precise definition of η^2 , it is helpful to first define the *isolation* index (I). This index equals the fraction of one's neighbors who are Minority for the average Minority resident in a given city or

$$I = \mathbb{E}[s(Z) | X = 1].$$

While Minorities make up only 26 percent of the population in Detroit (Column 3), the neighborhood in which a typical minority resides in is over three-quarters minority (Column 4).⁹

The η^2 index coincides with the coefficient on T in the linear regression of $s(Z)$ onto a constant and T .¹⁰ As such it provides a measure of the degree to which own race predicts the race of one's neighbors. Using the fact that (i) T is binary, (ii) $\mathbb{C}(T - s(Z), s(Z)) = 0$, and (iii) $\mathbb{V}(T) = \mathbb{V}(s(Z)) + \mathbb{E}[\mathbb{V}(T | Z)]$, some algebra gives

$$\eta^2 = \frac{\mathbb{C}(T, s(Z))}{\mathbb{V}(T)} = \frac{I - Q}{1 - Q}. \quad (2)$$

The η^2 index therefore provides a scaled measure of minority isolation. It measures the excess isolation of Minorities in a city compared to perfect integration (the numerator in (2)) relative to the corresponding excess isolation that would be observed in a perfectly segregated city (the denominator in (2)).

The dramatic levels of Minority isolation in the ten cities listed in Table 2 do not characterize all large metro areas in the United States. Table 3 lists, again according to the Dissimilarity Index, the ten most integrated metro areas (among those with at least 500,000 residents in 2000). In the Vallejo-Fairfield-Napa metro area of Northern California almost one third of residents are Minorities. However these residents live in neighborhoods that are only slightly more than one third Minority (compare Columns 3 and 4). For some of the cities

⁹We can also define White *exposure* to minorities as

$$E = \mathbb{E}[s(Z) | X = 0].$$

Exposure coincides with the fraction of neighbors who are minority for the typical White resident.

¹⁰Here and in what follows I use “linear regression of Y on X ” to denote the mean squared error minimizing linear predictor of Y given a constant and X (e.g., Goldberger, 1991).

Table 3: Most racially integrated metro areas in 2000

	(1)	(2)	(3)	(4)
	DI _S	η_S^2	Frac. URM	I
Vallejo-Fairfield-Napa, CA	0.262	0.089	0.319	0.379
Honolulu, HI	0.286	0.064	0.100	0.158
Stockton-Lodi, CA	0.301	0.136	0.379	0.464
Colorado Springs, CO	0.332	0.106	0.188	0.274
Portland-Vancouver, OR-WA	0.347	0.106	0.107	0.201
Seattle-Bellevue-Everett, WA	0.356	0.091	0.105	0.186
Tacoma, WA	0.365	0.100	0.142	0.227
Sacramento, CA	0.382	0.149	0.231	0.346
Albuquerque, NM	0.385	0.198	0.447	0.556
<i>All MSAs (N = 319)</i>	0.533	0.337	0.372	0.601
<i>Large MSAs (N = 99)</i>	0.560	0.376	0.381	0.622

Source: Neighborhood Change Database (NCDB), and author’s calculations.

Notes: See notes to Table 2.

listed in Table 2, Hispanics constitute a sizable fraction of the total Minority population. Prior research has shown that Hispanics are less-segregated from Whites than Blacks, a distinction glossed over in the aggregate statistics report in Tables 2 and 3 (e.g., Logan and Turner, 2013). Nevertheless, as I will articulate more precisely below, the lived experiences of Minorities and Whites are measurably different in the cities listed in Table 2 than they are in those listed in Table 3. Whether these differences help explain other dimensions of racial inequality is a key question that will be evaluated below.

Although residential segregation by race is pronounced in the United States, it has undergone appreciable changes in recent decades (e.g., Cutler, Glaeser and Vigdor, 1999). Figure 1 plots exposure indices for Blacks and Hispanic metro area residents from 1970 to 2000. The exposure of Blacks to Blacks corresponds to the isolation index defined above. The exposure of Blacks to Hispanics corresponds to the fraction of Hispanic residents in the average Black person’s neighborhood, etc. Figure 1 indicates that Black isolation has declined from over 0.65 in 1970 to just over 0.50 today. This decline reflects a more or less equal rise in exposure to Hispanics (from 0.05 to just over 0.10) and Whites (from 0.30 to just over 0.35). The decline in Black exposure to all Minorities (i.e., to Blacks and Hispanics) has been far less pronounced: from 0.71 to 0.64 from 1970 to 2000. In 1970 the typical Black metro resident lived in a neighborhood that was a little more than two thirds Minority, in 2000 she lived in a neighborhood that was a little less than two thirds minority.

In contrast to that of Blacks, the experience of Hispanics from 1970 to 2000 has been one

of rising residential segregation.¹¹ While in 1970 Hispanic exposure to Whites was over 55 percent, it had fallen to almost 40 percent by 2000. Hispanic exposure to Blacks increased, albeit modestly, while own exposure, or isolation, increased sharply from just under 0.35 to just over 0.45. Overall Hispanic exposure to all Minorities increased from 0.44 to 0.58 between 1970 and 2000.

Segregation trends can be sensitive to the measure used. Figure 2 plots Minority (i.e., Black and Hispanic) vs. Majority (i.e., White and Asian) Dissimilarity and η^2 indices from 1970 to 2000. The population weighted DI declined by almost 0.1 from 1970 to 2000 (left panel, blue line). Likewise the η^2 index, which can be thought of as a composition adjusted measure of Minority isolation, fell from about 0.45 to 0.35 (right panel, blue line).

Although one’s conclusions about trends in residential segregation by race are somewhat sensitive to the measure used, high levels of segregation by race, even in the post Civil Rights era, appears to be a durable feature of urban America.

2 Contextual inequality by race in contemporary urban America

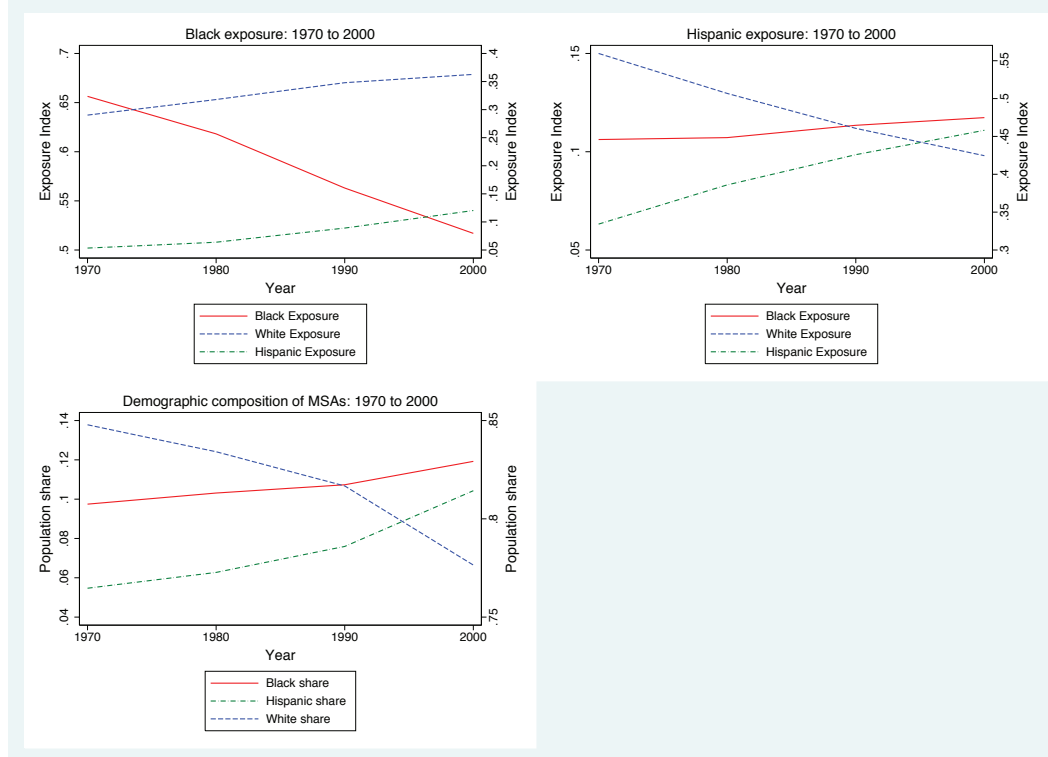
The previous section presented evidence indicating that in some of America’s largest cities Minority and White residents live in largely distinct neighborhoods. That extreme levels of residential sorting by race should be associated with commensurate inequalities in the social context in which Minorities and Whites live, work and socialize, should come as no surprise. Nevertheless, since differences in school quality, social services and exposure to violence (among other factors) across Minority and White neighborhoods are often asserted to be the mechanism by which “neighborhood effects” operate, it is worth documenting and measuring these differences more precisely.

Table 4 reports Black-White and Hispanic-White gaps in a few measures of contextual background for the subsample of NLSY97 respondents used to construct Table 1 above. The values of each of these variables are respondent reported, they measure youth perceptions, which may or may not be systematically biased.

Blacks and Hispanics are less likely to report feeling safe at school or that their peers have college plans (rows 1 and 2). Black and Hispanic males are almost twice as likely to report that significant numbers of their peers belong to gangs (row 3). Finally male Black youths

¹¹There is some nuance to this claim. The Hispanic population has increased sharply in many metro areas, this demographic change would increase Hispanic isolation even in a perfectly integrated society.

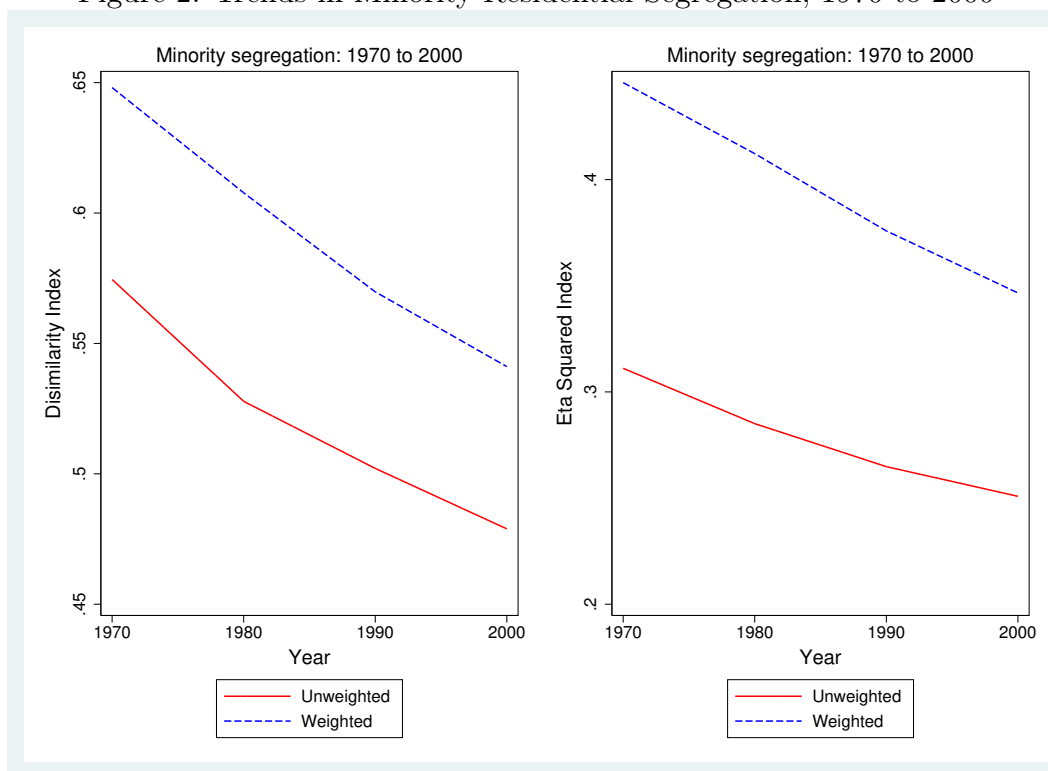
Figure 1: Trends in Hispanic and Black Exposure to other Races, 1970 to 2000



Source: Neighborhood Change Database (NCDB) and author's calculations.

Notes: Statistics computed using “normalized” 2000 census tract boundaries (Tatian, 2003). MSA definitions correspond to an author-defined cross-walk of Office of Management and Budget (OMB) 1981 SMSA and 1999 MSA/PMSA definitions (see Graham and Sharkey, 2013). Figures based on the 259 MSAs with complete data for all four census years. Panel A (Upper-left): Exposure indices (EIs) for Blacks. Black exposure to Blacks is plotted on the left y-axis, while their exposure to Whites and Hispanics is shown on the right y-axis. Panel B (Upper-right): Exposure indices (EIs) for Hispanics. Hispanic exposure to Blacks is plotted on the left y-axis, while their exposure to Whites and Hispanics is shown on the right y-axis. Panel C (Lower-left): Population shares by race. Black and Hispanic shares plotted on the left y-axis, White shares on the right y-axis. The indices plotted in Panels A and B correspond to, respectively, Black and Hispanic population weighted averages across MSAs. They therefore give the exposure of a randomly sampled Black and Hispanic MSA-resident at each time period. Shares plotted in Panel C correspond to unweighted averages across MSAs. They therefore give the demographic composition of the “average” MSA at each time period.

Figure 2: Trends in Minority Residential Segregation, 1970 to 2000



Source: Neighborhood Change Database (NCDB) and author's calculations.

Notes: See notes to Figure 1. Panel A (left): Dissimilarity index. Panel B (right): Eta-squared index. The weighted indices plotted in Panels A and B use MSA total population as weights.

Table 4: Contextual background for MSA-resident White, Black and Hispanic youth born 1980 to 1984

	(1) White	(2) Black-White Difference	(3) Hispanic-White Difference
Contextual background			
<i>Panel A: Females & Males (N = 4,419)</i>			
Feel safe at school	0.8991 (0.0068)	-0.1201 (0.0159)	-0.0438 (0.0152)
Peers with college plans	0.6499 (0.0111)	-0.1530 (0.0210)	-0.1671 (0.0216)
<i>Panel B: Males only (N = 2,292)</i>			
Peers in gangs	0.2338 (0.0131)	0.2192 (0.0282)	0.1674 (0.0293)
Witnessed a shooting	0.1573 (0.0123)	0.2911 (0.0276)	0.1590 (0.0269)

Source: National Longitudinal Survey of Youth 1997 (NLSY97), Neighborhood Change Database (NCDB) and author’s calculations.

Notes: Each row corresponds to the least squares fit of the stated outcome onto a constant (Column 1), a Black dummy (Column 2) and a Hispanic dummy (Column 3). Base-year NLSY97 sampling weights are used. Standard errors, clustered at the household level, are reported in parentheses. “Peers in gangs” takes a value of 1 for respondents who report that at least 25 percent of their peers belong to a gang and zero otherwise. “Peers with college plans” takes a value of 1 for respondents who report that at least 75 percent of their peers plan to go to college and zero otherwise. “Feel safe at school” takes a value of 1 for respondents who strongly agree or agree that school is safe and zero otherwise. “Shooting witness” equals one if a respondent reports witnessing a shooting at some point prior to age 18 and zero otherwise. Due to item non-response the samples sizes for the four rows in the table are, respectively, 4,411, 4,415, 2,288 and 2,040.

are three times more likely to report witnessing a shooting prior to turning 18 than their White counterparts; Hispanics report witnessing a shooting at twice the rate of Whites.

The choice of measures included in Table 4 reflects a combination of availability and potential relevance. The two explicit peer measures proxy for differential exposure by race to peers engaging in “delinquent” activities vs. “human capital building” activities. Models of peer influence hypothesize that peer behaviors can exert strong influences on adolescent behavior (e.g., Crane, 1991). Peer interactions have been hypothesized by a number of authors as a major neighborhood effects mechanism.

The second two contextual measures capture different levels of exposure to violence. In choosing these measures I was guided by two pieces (no doubt among others) of emerging evidence on the role of violence in “neighborhood effects”.

The first comes from the U.S. Department of Housing and Urban Development’s (HUD)

Moving to Opportunity (MTO) demonstration. A random subset of MTO participants were given access to a housing voucher to support relocation away from traditional public housing projects. Ludwig (2012) provides a basic overview of the MTO demonstration. While access to the housing voucher was randomly assigned, participation into the experiment was voluntary. Almost 80 percent of MTO participants listed a desire to “get away from gangs and drugs” as their primary or secondary reason for wanting to leave their current neighborhood. Violence and safety themes also emerge in various MTO follow-up studies. I will discuss what we can, and can not, learn about neighborhood effects from the MTO evidence below.

The second reason for focusing on violence exposure measures, is due to its hypothesized effect on various human capacities (see Massey (2004) for an interesting survey and commentary). For example, Sharkey (2010) compares scores on cognitive assessments given to youth just prior, or immediately after, a homicide in their neighborhood, finding that the performance of the latter group is significantly lower than the former. His sample consists of Project on Human Development in Chicago Neighborhoods (PHDCN) respondents. Youth exposure to violence is also correlated with lower levels of self reported health in early adulthood (e.g., Boynton-Jarrett et al. 2008).

Table 4 documents high levels of contextual inequality between Blacks, Hispanics and Whites. Since context is a local concept, one might conjecture that this inequality is greater (lower) in metro areas where these groups live apart from (with) one another. Tables 5 and 6 present evidence confirming such a pattern. Let $c = 1, \dots, C$ index cities and i NLSY97 respondents within cities. Define $T = \text{Black} + \text{Hispanic}$ to be a “Minority” dummy and let $s(z) = \Pr(T = 1 | Z = z)$ remain as defined above (i.e., the fraction of Minority residents in neighborhood $Z = z$). Let Y_{ci} be a given contextual measure and consider the following regression model

$$\begin{aligned} \mathbb{E}[Y_{ci} | W_{ci}] &= a_c + b_B \times \text{Black}_{ci} + b_H \times \text{Hispanic}_{ci} \\ &\quad + c_B \times \text{Black}_{ci} \times \{\mathbb{E}[s(Z) | \text{Black}_{ci} = 1] - \mathbb{E}[s(Z) | T_{ci} = 0]\} \\ &\quad + c_H \times \text{Hispanic}_{ci} \times \{\mathbb{E}[s(Z) | \text{Hispanic}_{ci} = 1] - \mathbb{E}[s(Z) | T_{ci} = 0]\}, \quad (3) \end{aligned}$$

with $\mathbb{E}[s(Z) | \text{Black}_{ci} = 1]$ measuring Black exposure to Minorities, $\mathbb{E}[s(Z) | \text{Hispanic}_{ci} = 1]$ Hispanic exposure to Minorities, $\mathbb{E}[s(Z) | T_{ci} = 0]$ White exposure to Minorities, and W_{ci} a vector containing all the regressors appearing on the right-hand-side of (3) (including a vector of city-specific dummy variables). Observe that $\mathbb{E}[s(Z) | \text{Black}_{ci} = 1] - \mathbb{E}[s(Z) | X_{ci} = 0]$ and $\mathbb{E}[s(Z) | \text{Hispanic}_{ci} = 1] - \mathbb{E}[s(Z) | T_{ci} = 0]$ are *exposure gaps*. In a perfectly integrated city these two terms will equal zero, in a perfectly segregated city, where Whites live separately from all Minorities, they will equal one. Hence b_B gives the expected Black-White “context

gap” in a perfectly integrated city and $b_B + c_B$ the expected gap in a perfectly segregated city. The interpretation of b_H and $b_H + c_H$ is analogous. These are predictive statements and, as such, can be viewed as “facts” (albeit with some parametric structure on the form of $\mathbb{E}[Y_{ci}|W_{ci}]$ maintained). Their implications will be discussed more critically below.¹²

The “a” columns in Tables 5 and 6 report the coefficients associated with the least squares fit of each of the four context variables onto MSA-specific constants and the Black and Hispanic dummy variables alone. The estimated context gaps are similar to those reported in Table 4, which made no allowance for cross-MSA differences in context. While the distribution of Blacks, Hispanics and Whites is not uniform across MSAs, these differences do not drive the contextual gaps documented in Table 4. Instead it is within-MSA differences which drive the differences in context documented in Table 4.

Column “b” presents estimates of the coefficients in equation (3). Consider the Black-White gap in peer gang exposure first (Column 1b of Table 6). Observe that the estimate of b_B is insignificantly different from zero while that of c_B is positive and significant. This indicates that Black-White differences in gang exposure are substantial in segregated cities and modest in more integrated cities. The lower panel of the Table reports estimates of the Black-White gang exposure gap in cities with low and high levels of Black-White separation. Clearly, and also unsurprisingly, contextual inequality is far greater in America’s most segregated cities.¹³

A similar pattern emerges for exposure to peers with college aspirations (Column 2b of Table 5). Blacks self report exposure to peers with college plans at about 70 percent of the rate at which Whites do. However in integrated cities there is only a small exposure gap, while in segregated cities Blacks report exposure at just over *one half the rate* Whites do (see Panel B of Table 5). Hispanic exposure to peers with college aspirations also varies systematically with the degree of residential segregation.

For the two violence exposure measures, the connection with residential segregation is both sharp and dramatic. Unconditionally Blacks report witnessing a shooting at three times the rate Whites do. Almost half of Black youths report witnessing a shooting by age 18. The Black-White gap in violence exposure varies substantially across cities, however. In relatively integrated cities Blacks witness shootings at just over twice the rate of their White counterparts. In segregated cities they do so at almost four times the rate (Table 6, Panel B).¹⁴

The evidence presented in Tables 4 to 6 indicates the Minority youth live in meaningfully

¹²As these “facts” are based on a random sample, it is worth reminding the reader that uncertainty, due to sampling variability, remains. However this component of uncertainty is quantifiable and captured by the standard errors reported throughout.

¹³For the gang exposure measure there is less evidence that residential segregation plays an important role

Table 5: Residential segregation and Black-White, Hispanic-White gaps in contextual background

	(1a)	(1b)	(2a)	(2b)
A: Females and Males (N = 4,419)	Feel safe at school		Peers with college plans	
Black	−0.1227 (0.0205)	0.0192 (0.0484)	−0.1868 (0.0237)	0.0271 (0.0482)
Hispanic	−0.0460 (0.0139)	0.0335 (0.0348)	−0.1701 (0.0231)	−0.0660 (0.0744)
Black $\times \{\mathbb{E}[s(Z) \text{Black}_{ci} = 1] - \mathbb{E}[s(Z) T_{ci} = 0]\}$		−0.3089 (0.1103)		−0.4649 (0.0946)
Hispanic $\times \{\mathbb{E}[s(Z) \text{Hispanic}_{ci} = 1] - \mathbb{E}[s(Z) T_{ci} = 0]\}$		−0.2565 (0.0976)		−0.3367 (0.1892)
Constant (White mean outcome)	0.8971 (0.0073)	0.9002 (0.0067)	0.6554 (0.0101)	0.6597 (0.0080)
MSA-specific intercepts?	Yes	Yes	Yes	Yes
N	4,411	4,411	4,415	4,415
B: Predicted Black-White & Hispanic-White Context Gaps at Low & High Segregation Levels				
Black-White (<u>Integrated</u> , 0.2 Minority Ex. Gap)		−0.0425 (0.0290)		−0.0659 (0.0315)
Black-White (<u>Segregated</u> , 0.40 Minority Ex. Gap)		−0.1816 (0.0318)		−0.2751 (0.0248)
Hispanic-White (<u>Integrated</u> , 0.15 Minority Ex. Gap)		−0.0050 (0.0218)		−0.1165 (0.0473)
Hispanic-White (<u>Segregated</u> , 0.40 Minority Ex. Gap)		−0.0691 (0.0142)		−0.2007 (0.0182)

Source: National Longitudinal Survey of Youth 1997 (NLSY97), Neighborhood Change Database (NCDB) and author's calculations..

Notes: See notes to Table 4. Each column corresponds to a least squares fit of the listed outcome onto (a) the Black and Hispanic indicator variables and a vector of MSA indicator variables, and (b) the addition of interactions of Black and Hispanic with, respectively, the Black-White and Hispanic-White minority exposure gap. Standard errors, reported in parentheses, are clustered at the MSA level. NLSY sampling weights are *not* used.

Table 6: Residential segregation and Black-White, Hispanic-White gaps in contextual background

	(1a)	(1b)	(2a)	(2b)
A: Males only (N = 2,292)	Peers in gangs		Witnessed a shooting	
Black	0.2262 (0.0305)	0.0385 (0.0768)	0.3214 (0.0369)	0.0829 (0.0995)
Hispanic	0.1763 (0.0361)	0.0285 (0.1031)	0.1680 (0.0352)	-0.0001 (0.0649)
Black $\times \{\mathbb{E}[s(Z) \text{Black}_{ci} = 1] - \mathbb{E}[s(Z) T_{ci} = 0]\}$		0.4080 (0.1477)		0.5200 (0.2039)
Hispanic $\times \{\mathbb{E}[s(Z) \text{Hispanic}_{ci} = 1] - \mathbb{E}[s(Z) T_{ci} = 0]\}$		0.4727 (0.2748)		0.5443 (0.2027)
Constant (White mean outcome)	0.2375 (0.0146)	0.2326 (0.0135)	0.1537 (0.0166)	0.1474 (0.0156)
MSA-specific intercepts?	Yes	Yes	Yes	Yes
N	2,288	2,288	2,040	2,040
B: Predicted Black-White & Hispanic-White Context Gaps at Low & High Segregation Levels				
Black-White (<u>Integrated</u> , 0.2 Minority Ex. Gap)		0.1201 (0.0504)		0.1869 (0.0623)
Black-White (<u>Segregated</u> , 0.40 Minority Ex. Gap)		0.3037 (0.0369)		0.4209 (0.0502)
Hispanic-White (Integrated, 0.15 Minority Ex. Gap)		0.0994 (0.0658)		0.0815 (0.0414)
Hispanic-White (<u>Segregated</u> , 0.40 Minority Ex. Gap)		0.2176 (0.0371)		0.2176 (0.0406)

Source: National Longitudinal Survey of Youth 1997 (NLSY97), Neighborhood Change Database (NCDB) and author's calculations.
Notes: See notes to Tables 4 and 5.

different types of neighborhoods than their White counterparts. They live with peers who, on average, are more likely to belong to gangs and less likely to have college plans. They attend schools that feel less safe and are exposed to violence at a greater rate. These contextual gaps, unsurprisingly, are more pronounced in segregated cities.

3 What drives segregation?

The previous section documented substantial levels of residential sorting by race and income. It also documented how measured sorting correlates with specific measures of the social contexts in which youths live. Blacks and Whites, for example, are exposed to different peers and levels of violence. These gaps are largest in segregated cities, where the two groups tend to live apart. These facts raise the question of why Blacks and Whites live apart? While there are very important historical, institutional and policy reasons for widespread segregation (cf., Sharkey, 2012), in this section I will focus on explanations of segregation as an *equilibrium* by-product of decentralized sorting by households.

At least since the work of Tiebout (1956), economists have modeled household neighborhood choice in terms of utility maximization. Tiebout (1956) concluded that neighborhood sorting was efficient: the combination of multiple communities, majority voting for public expenditures within communities, and heterogeneous valuations of public goods, allows households to choose a community with a level of public good provision (and associated tax rate) that closely aligns with their preferences. Households vote with their feet.

Bewley (1981) provided several counter examples which illustrated different ways in which a sorting equilibrium could diverge from a social planner’s assignment of households to communities. This theme was further developed, with a particular emphasis on the effects of social spillovers, in a series of papers published in the 1990s (e.g., de Bartolome, 1990; Benabou, 1993, 1996; Durlauf, 1996a; Becker and Murphy, 2000). Piketty (2000), Fernandez (2003) and Durlauf (2004) survey this literature. Here I provide a synoptic sketch of this work.

The model is a two period one.¹⁵ Household utility depends on first and second period consumption, $U(c_t, c_{t+1})$. There are two types of households, $t \in \{0, 1\}$. Without intending to be pejorative I will call $t = 1$ and $t = 0$ households, respectively, “high” and “low” type

in explaining Hispanic-White differences.

¹⁴The shooting exposure rates derived from the NLSY97 are consistent with those from other surveys. Finelhor et al (2009), using the National Survey of Childrens Exposure to Violence, report that 22.2 percent of 14 - 17 year olds were exposed to a shooting.

¹⁵See Benabou (1996) for an overlapping generations extension.

households (representing advantaged and disadvantage households respectively). In period one households choose a neighborhood of residence. All neighborhoods are ex ante identical, ex post they vary in the fraction of high type families which live in them. Let s_c denote the fraction of high type families in community c .

In period t households are exogenously endowed with income y_t . They choose their neighborhood of residence and optimal level of debt, d_t . Since housing is homogenous, and all neighborhoods are ex ante identical, the price of housing in community c depends only on the fraction of high type households there. The period t budget constraint is therefore

$$c_t + p(s) = y_t + d_t, \quad (4)$$

with $p(s)$ denoting the price of a home in a neighborhood with composition s .

In period $t + 1$ households realize income

$$y_{t+1} = tm_H(s) + (1 - t)m_L(s). \quad (5)$$

Period $t + 1$ income depends on the “quality” of neighborhood chosen in period t . Equation (5) captures the possibility that human capital acquisition may be facilitated by exposure to high type peers, which, in turn, increases future wages. Assume that both $m_H(s)$ and $m_L(s)$ are increasing in s , then period $t + 1$ income, holding own type constant, will be higher for those households who choose neighborhoods with many high type residents. Equation (5) provides a reduced form representation of how peer spillovers influence the human capital accumulation of children. This could be due to direct peer effects or because, say, the tax base in high type communities supports better schools.

Households face a simple trade-off: purchasing a home in a higher quality neighborhood increases future income (and hence second period consumption), but at the cost of reducing first period consumption. The presence of a credit market, which allows home purchases to be at least partially financed, complicates the analysis; but not essentially.

In period $t + 1$ households also pay off all debt (principal plus interest). Let r_H and r_L denote the interest rate available to high and low type borrowers. If credit markets are perfect $r_H = r_L = r$. The second period budget constraint is

$$c_{t+1} + d(1 + tr_H + (1 - t)r_L) = y_{t+1} = tm_H(s) + (1 - t)m_L(s). \quad (6)$$

Solving (4) and (6) for c_t and c_{t+1} and plugging into the utility function yields

$$U(c_t, c_{t+1}) = U(y_t + d_t - p, tm_H(s) + (1-t)m_L(s) - d_t(1 + tr_H + (1-t)r_L)). \quad (7)$$

To characterize how families trade-off housing prices (or equivalently period one consumption) and neighborhood quality we differentiate (7) with respect to housing price, p , and neighborhood composition, s , holding utility constant

$$-U_1 dp + U_2 \{t \nabla_s m_H(s) + (1-t) \nabla_s m_L(s)\} ds = 0.$$

Using the Euler equation for the optimal level of borrowing

$$\frac{U_2}{U_1} = t \frac{1}{1+r_H} + (1-t) \frac{1}{1+r_L},$$

and solving for dp/ds then yields

$$\begin{aligned} \frac{dp}{ds} &= t \frac{\nabla_s m_H(s)}{1+r_H} + (1-t) \frac{\nabla_s m_L(s)}{1+r_L} \\ &\stackrel{def}{=} R(t, s). \end{aligned} \quad (8)$$

Equation (8) characterizes a household's marginal rate of substitution between first period consumption and neighborhood quality. The higher $R(t, s)$ the more a household is willing-to-pay to live in a neighborhood with a higher fraction of high type households.

Equilibrium

Assume there is a continuum of equally-size neighborhoods. Let $f_S^{\text{eq}}(s)$ denote a candidate equilibrium neighborhood composition density. Feasibility of this equilibrium requires that

$$\int_0^1 s f_S^{\text{eq}}(s) ds = p_H, \quad (9)$$

with p_H the population frequency of high type households (cf., Graham, Imbens and Ridder, 2010).

Consider two neighborhoods, both with fraction s high types. Imagine a reallocation of households to neighborhoods such that s *increases* by ds in the first neighborhood and *decreases* by ds in second. Call the set of households who actually change neighborhoods as part of the reallocation *movers*. Each moving high type household trades homes with a mov-

ing low type households. Moving high type households enjoy a higher quality neighborhood post-move, while moving low type households enjoy a lower quality neighborhood post-move.

The net present value of the change in neighborhoods for high type movers is

$$\frac{m_H(s + ds) - m_H(s)}{1 + r_H}, \quad (10)$$

while that for low type movers is

$$\frac{m_L(s - ds) - m_L(s)}{1 + r_L}. \quad (11)$$

If the sum of (10) and (11) is positive, then the net gain to movers from the reallocation will be positive. Consequently there will be a set transfers between high and low type households that will sustain the reallocation (e.g., by selling each other their houses at differing prices). For ds infinitesimally small positivity of the sum of (10) and (11) is equivalent to the condition

$$\frac{\nabla_s m_H(s)}{1 + r_H} - \frac{\nabla_s m_L(s)}{1 + r_L} > 0. \quad (12)$$

The left-hand-side of (12) is a measure of *local complementarity* (in net present value terms) between own and neighbor's type. If, at neighborhood composition s , condition (12) holds, then *moving* high type households will gain more from an improvement in neighborhood quality, than *moving* low type households will lose from a deterioration in neighborhood quality. Condition (12) implies that there exists a set of transfers between a high and low type household such that both can be made better off by switching neighborhoods.

Note that Condition (12) coincides with

$$\tilde{\eta}(s) \stackrel{def}{=} R(1, s) - R(0, s) > 0. \quad (13)$$

which, if it holds for all $s \in [0, 1]$, is called a *single crossing* condition. Single crossing implies that high types' willingness to pay for a marginal improvement in neighborhood quality exceeds that of low types for all values of s . As Benabou (1996, Proposition 1) and others have shown, when (13) holds for all $s \in [0, 1]$ the equilibrium assignment of households to neighborhoods will be completely segregated. Conversely if $R(1, s) - R(0, s) < 0$ for all $s \in [0, 1]$ then perfect integration will be the unique stable equilibrium (under simple best-reply dynamics).

The characterization of equilibrium when $R(1, s) - R(0, s)$ changes sign is more complex, however, under best-reply dynamics, at any s on the interior of the unit interval for which

$f_S^{\text{eq}}(s)$ is positive, it must be the case that $R(1, s) - R(0, s) < 0$.

Condition (12) indicates that two forces drive sorting. First, even if the human capital benefit of a marginal improvement in neighborhood quality is the same for high and low type households (i.e., $\nabla_s m_H(s) = \nabla_s m_L(s)$), the equilibrium assignment of households to neighborhoods will be segregated if $r_H < r_L$. Credit market imperfections, which allow high type households to finance home purchases at more favorable terms than low type households, help to sustain a stratified equilibrium. Differential access to credit by Blacks and Whites has been an important driver of residential segregation by race historically (cf., Ross and Yinger, 2002). Under limited liability constraints, access to credit on favorable terms varies with socioeconomic status almost by construction.

Second, sorting may arise due to “technological” complementarity between own and neighbor’s type. That is, even in the absence of capital market imperfections, households will sort into homogenous neighborhoods if $\nabla_s m_H(s) > \nabla_s m_L(s)$ (i.e., if the marginal benefit, in terms of second period income, of an improvement in neighborhood quality is higher for high type households).

Efficiency

When two households “trade neighborhoods” in a decentralized housing market, the desirability of the trade is evaluated solely in terms of the costs and benefits of the trade to the directly involved parties. However, when a high type household leaves a neighborhood it imposes costs on the households left behind and provides benefits to households located in its new neighborhood. These spillover effects are not internalized by a decentralized housing market. The presence of these spillovers means that the *laissez faire* assignment of households to neighborhoods may be inefficient (even in the absence of credit market imperfections). The net benefits associated with a neighborhood trade depend on its implications both for the directly involved *movers*, which are fully internalized by the housing market, *as well as* the implications for *stayers*, which are not.

Since the social planner necessarily discounts the future income of low- and high-type households equally, we may assess the social efficiency of any candidate neighborhood assignment by studying its implications for the distribution of second period income. Let

$$m(s) = sm_H(s) + (1 - s)m_L(s)$$

equal the average second period income in a neighborhood of composition s . If, as before, we take two neighborhoods with ex ante composition s and move ds high types from one

neighborhood to the other (and ds low types in the opposite direction) the net change in average income will be

$$\frac{m(s + ds) + m(s - ds) - 2m(s)}{2}.$$

For ds infinitesimally small this change will be positive if

$$\nabla_{ss}m(s) = 2\eta(s) + \lambda(s) \tag{14}$$

is greater than zero for

$$\begin{aligned} \eta(s) &\stackrel{def}{=} \nabla_s m_H(s) - \nabla_s m_L(s) \\ \lambda(s) &\stackrel{def}{=} s \nabla_{ss} m_H(s) + (1-s) \nabla_{ss} m_L(s). \end{aligned}$$

The first term in (14) measures the “technological” *complementarity* between own and neighbor’s type. The second term is what Benabou (1996) calls *curvature*. If the marginal benefit of an additional high type is declining in the fraction of high types already present in the neighborhood, then reallocations which increase segregation may be inefficient (even if own and neighbor’s type are complementary).

Consider the following functional forms for $m_H(s)$ and $m_L(s)$

$$\begin{aligned} m_L(s) &= \alpha_L + \beta s - \gamma s^2 \\ m_H(s) &= \alpha_H + (\beta + \epsilon) s - \gamma s^2, \end{aligned}$$

with β , γ , and ϵ all positive. Under this technology the marginal benefit of an incremental increase in neighborhood quality is declining for both low- and high-types. Concavity of this type means that transfers of high types from low s to high s neighborhoods will tend to harm stayers on net. However the incentives for sorting will be driven by complementarity between own and peer type, in this case given by

$$\eta(s) = \epsilon,$$

so that for any $\epsilon > 0$ a high type household will always outbid a low type household for any marginal improvement in neighborhood quality: equilibrium will be stratified.¹⁶ However, for $0 < \epsilon < \gamma$, the overall mapping of neighborhood composition into average income is (globally) concave since $\nabla_{ss}m(s) = 2(\epsilon - \gamma) < 0$: stratification is inefficient.

¹⁶To keep things simple the discussion abstracts from credit market imperfections (i.e., I assume that $r_H = r_L = r$).

This example provides a rather dramatic demonstration of a general theme which emerges from the literature on locational sorting in the presence of social spillovers: when the private incentives to sort are misaligned with the social costs and benefits from doing so, the *laissez faire* assignment of families to neighborhoods need not be efficient and, in some cases, may be grossly inefficient. This possibility is an important motivation for associational redistribution (cf., Piketty, 2000).

Adding unobserved heterogeneity

In the simple locational sorting model outlined above, the only source of agent heterogeneity is their type, high vs. low. In practice there is within-type heterogeneity, generally unobserved by the econometrician. To get some sense of how heterogeneity affects sorting consider the following simple (expositional) example. Assume that the outcome, Y , given residence in a neighborhood with composition $S = s$ is

$$Y(s) = \begin{cases} \alpha_H + \beta_H \cdot s \cdot \exp(A_1) & , \text{ if } T = 1 \\ \alpha_L + \beta_L \cdot s \cdot \exp(A_1) & , \text{ if } T = 0 \end{cases}.$$

This model presumes that the outcome increase linearly in s for both high and low types, but at a rate which may be individual-specific (due to variation in A_1). Further allow agent utility to differ from the outcome of interest to the researcher. Specifically the utility associated with residence in a neighborhood with composition $S = s$ is

$$V(s) = \begin{cases} Y(s) - \gamma_H \cdot s \cdot \exp(A_2) & , \text{ if } T = 1 \\ Y(s) - \gamma_L \cdot s \cdot \exp(A_2) & , \text{ if } T = 0 \end{cases}.$$

There are two sources of unobserved heterogeneity in this model. Agents vary in the sensitivity of their outcome to neighborhood composition (variation in A_1) and may also vary in their taste for s for reasons other than its effect on Y (variation in A_2). Assume that

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \Big| T \sim N \left(\begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Under these assumptions we have average mappings from neighborhood composition into outcomes of

$$\begin{aligned} m_H(s) &= \mathbb{E}[Y(s) | T = 1] = \alpha_H + \beta_H s \\ m_L(s) &= \mathbb{E}[Y(s) | T = 0] = \alpha_L + \beta_L s \end{aligned}$$

and average mappings from neighborhood composition into utilities of

$$\begin{aligned} u_H(s) &= \mathbb{E}[V(s)|T=1] = \alpha_H + (\beta_H - \gamma_H)s \\ u_L(s) &= \mathbb{E}[V(s)|T=0] = \alpha_L + (\beta_L - \gamma_L)s. \end{aligned}$$

Agents observe A_2 – their tastes – prior to choosing their neighborhood of residence, but A_1 – which indexes heterogeneity in the outcome effect of neighborhood composition – is only revealed *ex post*. Under the normality assumption the *expected marginal utility* of an incremental increase in s is

$$\mathbb{E}[\nabla_s V(s)|T, A_2 = a_2] = \begin{cases} \beta_H \cdot \exp\left(-\frac{1}{2}(1-\rho) - \rho a_2 + \frac{1-\rho^2}{2}\right) - \gamma_H \cdot \exp(a_2) & , T = 1 \\ \beta_L \cdot \exp\left(-\frac{1}{2}(1-\rho) - \rho a_2 + \frac{1-\rho^2}{2}\right) - \gamma_L \cdot \exp(a_2) & , T = 0 \end{cases}.$$

Agents with a high marginal utility of neighborhood composition will outbid those with lower marginal utilities for spots in high s neighborhoods.¹⁷

There are two neighborhoods of size one. For simplicity, I also assume that there are equal numbers of high and low types. Let $F_H(\cdot)$ and $F_L(\cdot)$ denote the distribution function of expected marginal utility for high and low types respectively. Let c^* be the expected marginal utility level which satisfies

$$F_H(c^*) + F_L(c^*) = 1. \quad (15)$$

Note that exactly half the population (highs and lows together) has a marginal utility of neighborhood composition greater than c^* and half one less than or equal to c^* . The value of c^* which partitions the population into two halves determines the nature of sorting in this economy.

Assume that $\beta_H > \beta_L$, $\gamma_H < \gamma_L$ and $\rho \leq 0$ such that, on average, high types value an incremental increase in s more than low types. Under these conditions the high type neighborhood will have fraction $s' = 1 - F_H(c^*)$ high types while the low type neighborhood will have fraction $s = F_H(c^*)$.

Now say the econometrician attempts to infer the marginal returns to an increase in s for

¹⁷To keep things simple I assume that utility is directly transferable across agents who trade locations.

high types by comparing outcomes for this group across the two neighborhoods.

$$\frac{\mathbb{E}[Y|S = s', T = 1] - \mathbb{E}[Y|S = s, T = 1]}{s' - s} = \underbrace{\beta_H}_{\text{structural effect}} + \underbrace{\beta_H \{s' \cdot \mathbb{E}[\exp(A_1)|\nabla_s V(s) \geq c^*, T = 1] - s \cdot \mathbb{E}[\exp(A_1)|\nabla_s V(s) < c^*, T = 1]\}}_{\text{sorting bias}}$$

High types in the two neighborhoods have different observed outcomes, on average, for two reasons. First, those individuals located in the high $S = s'$ neighborhood benefit from greater exposure to high type neighbors. Second, the distribution of A_1 differs across the two neighborhoods. High types in the high $S = s'$ neighborhood tend to benefit more from their high type neighbors than their counterparts in the low $S = s$ neighborhood would. This is *sorting bias*: the distribution of unobserved agent ability varies with observed neighborhood composition. Note that if $\rho = 0$, so that knowledge of A_2 provides no information about the value of A_1 , then the sorting bias term will be zero. Assessing households' information sets is often an important step in justifying particular identification conditions (cf., Graham, Imbens and Ridder, 2010).

Developing research designs which ameliorate the effects of sorting bias has been a preoccupation, perhaps the central preoccupation, of neighborhood effects researchers for at least three decades (e.g., Mayer and Jencks, 1989). The logic of simple sorting models, particularly when augmented with unobserved agent-level heterogeneity, can be useful for developing intuitions about the plausibility of different research designs in particular settings.

4 Research designs for identifying neighborhood effects

Consider a cohort of individuals who reside in a common metropolitan area during their formative years, for example, individuals born in the early 1960s who lived in the Bay Area of California as teenagers. Diversity in this population is characterized by the pair (A', T) . Here A is a vector of *unobserved* individual attributes, measured prior to, or at the onset of, adolescence. The elements of A capture, among (many) other attributes, innate cognitive ability, health endowments, and family and ethnic background. An agent's observed *type* is given, as earlier, by the binary indicator T . In practice T might be a race dummy or an indicator for whether a child's parents graduated from college or not.

Let Y be an outcome of interest. This outcome is measured post-adolescence; examples

include adult earnings, eventual educational attainment and incarceration status by age 28. All exogenous agent characteristics measurable at the onset of adolescence and relevant for the determination of Y are captured by (A', T) . Since A is unmeasured, and may be of arbitrarily high dimension, this is without loss of generality. For convenience I will call A “background”. It should be recognized that this reification is simply a shorthand for what is typically a large bundle of both acquired and innate attributes which help to determine Y . Continuing to keep the exposition concrete I will also refer to $T = 1$ households as Minorities and $T = 0$ households as Whites. Of course, other running examples are possible.

Individuals reside in one of $i \in \{1, \dots, N\}$ neighborhoods. Let Z be an $N \times 1$ vector of neighborhood dummies. Let $s(z) = \Pr(T = 1 | Z = z)$ be the fraction Minority in neighborhood $Z = z$, $m_A(z) = \mathbb{E}[A | Z = z]$ the mean background of residents living in neighborhood $Z = z$, and U additional unobserved neighborhood-level characteristics. Throughout what follows I will take the joint distribution of $(A', T)'$ as given and invariant across policies. In practice this means that the elements of $(A', T)'$ are non-manipulable, at least over the time frame in which the outcome is being measured. We conceptualize $(A', T)'$ as a bundle of fixed characteristics that a household brings with them as they move from neighborhood to neighborhood. Behaviors, for example parenting style, which may change with neighborhood of residence, are not elements of A .

We will also view U as a vector of *exogenous* location specific characteristics (e.g., micro-climate, proximity to the city-center and so on). Its marginal distribution is invariant across policies. Other location-specific characteristics, for example the mill rate on assessed property values, may vary with the mix of residents living in a location (e.g., via the mechanism of majority voting). The influence of these composition-induced policy changes on the outcome of interest are “peer effects”, broadly defined.

A prototypical analysis of the effects of residential segregation by race on Y begins with the linear regression model

$$\mathbb{E}[Y | T, s(Z), m_A(Z), U, A] = \alpha_0 + \beta_0 T + \gamma_0 s(Z) + m_A(Z)' \delta_0 + U' \kappa_0 + A' \lambda_0. \quad (16)$$

Equation (16) provides a mapping from the neighborhood distribution of household types and backgrounds, given own type, background and exogenous neighborhood-level characteristics, into outcomes. There are neighborhood effects, or more precisely, *spillovers*, if the outcome varies with changes in the neighborhood distribution of household types and/or backgrounds (i.e., $\gamma_0 \neq 0$ and/or $\delta_0 \neq 0$). There are *place effects* if any of γ_0 , δ_0 and κ_0 differ from zero.

The predictive value of own type and background, T and A , on the outcome is indexed by

β_0 and λ_0 . The parameter γ_0 measures how the expected outcome changes with the fraction Minority in one’s neighborhood, while δ_0 does so for changes in mean neighbor “background”. Finally κ_0 measures the influence of U , neighborhood “amenities”, on the outcome.

I assume this equation is structural in the sense of Goldberger (1991); specifically that its invariant across associational redistributing policies. Since we are free to define A and U broadly, the assumption that (16) describes a causal relationship, the linear conditional mean assumption aside, is not especially restrictive (cf., Wooldridge, 2005). Recall that the joint distribution of $(A', T)'$, and the marginal distribution of U is assumed invariant. Therefore equation (16) is structural in that it provides accurate predictions for an individual’s outcome given an exogenous change in her neighborhood environment. Put differently, the policies of interest are reallocations or, in the words of Durlauf (1996b), associational redistributions. Equation (16) is helpful for understanding how “who lives with whom” influences the distribution of Y .

Let (16) be the long regression of interest. The econometrician can compute the sample analog of the short regression of the outcome onto a constant, own Minority status and the fraction Minority in one’s neighborhood of residence:

$$\mathbb{E}[Y|T, s(Z)] = a_0 + b_0T + c_0s(Z). \quad (17)$$

Prior reviewers of the empirical literature on neighborhood effects have outlined a variety of reasons for why c_0 need not coincide with γ_0 and, more generally, why knowledge of relationship (17) may not be useful for understanding the effects of reallocations on the distribution of outcomes.

Sorting and matching

Two concerns commonly raised by reviewers of the neighborhood effects literature are biases due to *sorting* and *matching*. Graham, Imbens and Ridder (2010) formally define sorting and matching on unobservables in the context of a non-linear, non-separable, version of regression (16). The linear form of (16) allows for a more compact development of the key ideas, but this is not without a loss of generality. I begin with the mean regression representation of A

$$A = \pi_0 + \phi_0T + B, \quad \mathbb{E}[B|T] = 0. \quad (18)$$

Here B is the component of “background” that does not vary, on average, with race. It is convenient to partition unobserved agent-level heterogeneity into a component that varies

with type and one which is (mean) independent of it. Equation (18) is nothing more than a decomposition. Graham, Imbens and Ridder (2010) work with a variant of (18) appropriate for models where A enters non-separably; they provide additional discussion and motivation. Note that the “background gap” between Minorities and Whites is given by $\phi_0 = \mathbb{E}[A|T = 1] - \mathbb{E}[A|T = 0]$.

There is *no sorting unobservables* if an individual’s neighborhood of residence, Z , is not predictive of the B component of her background conditional on her observed type, T :

$$\mathbb{E}[B|T, Z] = \mathbb{E}[B|T] = 0. \quad (19)$$

The second equality in (19) follows by construction. Condition (19) implies that the distribution of “background” (specifically its mean) among, say, Minorities is similar across neighborhoods. If this is not the case, then we say there is sorting on unobservables. For example, it may be that Minority families living in predominately White neighborhoods differ systematically in terms of A , from their counterparts in predominately Minority neighborhoods (e.g., their adult members may have graduated from more elite colleges). This captures the intuition that observed neighborhood characteristics may be correlated with the unobserved characteristics of its residents; a commonly articulated concern in empirical analyses of neighborhood effects (e.g., Brooks-Gunn, Duncan, Klebanov and Sealander, 1993; Duncan and Raudenbush, 2001; Sampson, Morenoff and Gannon-Rowley, 2002; Kling, Liebman and Katz, 2007). The previous section presented an example of how this could occur in a sorting equilibrium.

Matching is a related but distinct process. There is matching when the unobserved *exogenous* attributes of one’s neighborhood (e.g., proximity to the city-center) can be predicted by own type. There is *no matching on U* if

$$\mathbb{E}[U|T] = \mathbb{E}[U]. \quad (20)$$

If, for example, Minorities are less likely to live in neighborhoods adjacent to employment districts, then we say there is matching on U .

Anatomy of the short regression

A large empirical literature fits models of the form given in (17), typically with additional individual- and location-level characteristics (see below for more on the role of controls). How do these analyses relate to the structural model (16)? After tedious manipulation it is

possible to show that the two slope coefficients in (17) equal¹⁸

$$b_0 = \beta_0 + \phi'_0 \lambda_0 - \frac{1}{p(1-p)} \mathbb{C}(s(Z), \mathbb{E}[B|Z])' \frac{\lambda_0}{1-\eta^2}. \quad (21)$$

$$c_0 = \gamma_0 + \phi'_0 \delta_0 + \frac{1}{\eta^2} \left\{ \frac{1}{p(1-p)} \mathbb{C}(s(Z), \mathbb{E}[B|Z])' \left(\delta_0 + \frac{\lambda_0}{1-\eta^2} \right) + (\mathbb{E}[U|T=1] - \mathbb{E}[U|T=0])' \kappa_0 \right\}. \quad (22)$$

where $p = \mathbb{E}[T]$ is the population fraction Minority, η^2 the eta-squared index of segregation defined earlier and $\mathbb{C}(X, Y)$ denotes the covariance of X with Y . Under the no sorting on unobservables condition (19) we have $\mathbb{C}(s(Z), \mathbb{E}[B|Z]) = 0$ since

$$(s(Z) - p) \mathbb{E}[B|Z] = (s(Z) - p) \mathbb{E}[\mathbb{E}[B|T, Z]|Z] = (s(Z) - p) \mathbb{E}[\mathbb{E}[B|T]|Z] = 0$$

and hence a coefficient on Minority equal to

$$b_0 = \beta_0 + \phi'_0 \lambda_0, \quad (23)$$

so that knowledge of a person's race alters one's prediction of their realization of Y via a direct adjustment, β_0 , as well as an indirect adjustment capturing the population average difference in A across the two groups, $\phi'_0 \lambda_0 = (\mathbb{E}[A|T=1] - \mathbb{E}[A|T=0])' \lambda_0$. Because the distribution of $(A', T)'$ is assumed invariant across reallocations, b_0 , is a structural, albeit composite, parameter when there is no sorting on unobservables.¹⁹ When condition (19) fails to hold, b_0 is not structural, in the sense that its value is not policy invariant. As is clear from inspection, the third component of (21) varies with the joint distribution of $(A', T, Z)'$, whose manipulation is precisely the goal of a reallocation policy. If neighborhood of residence predicts unobserved background conditional on observables (i.e., condition (19) fails), then b_0 will not be useful for predicting the effects of reallocative policies.

A similar set of observations apply to c_0 , the coefficient on fraction Minority, in (17). It is this coefficient which purports to provide a measure of peer group or neighborhood effects. The third term in (22) depends on the covariance between fraction minority in a neighborhood and mean neighbors' background. Although B is mean independent of T , its neighborhood average need not be. Consider a city where low B households, irrespectively of race, sort into

¹⁸To keep the calculations manageable I assume there is a continuum of neighborhoods. Calculation details are provided in Appendix B.

¹⁹When the distribution of $(A', T)'$ is invariant across the policies of interest maintaining an "inclusive definition of type" is without loss of generality (cf., Graham, Imbens and Ridder, 2010).

predominately Minority neighborhoods and high B households into predominately White neighborhoods. Under this type of sorting pattern $\mathbb{C}(s(Z), \mathbb{E}[B|Z]) = \mathbb{E}[\mathbb{E}[B|Z]|T=1] - \mathbb{E}[\mathbb{E}[B|Z]|T=0]$ will be negative. This will, in turn, bias c_0 downward relative to $\gamma_0 + \phi'_0 \delta_0$, making exposure to Minority neighbors appear more detrimental for Y than it would be in the absence of sorting.

When there is no sorting on unobservables the coefficient on fraction Minority simplifies to

$$c_0 = \gamma_0 + \phi'_0 \delta_0 + \frac{1}{\eta^2} (\mathbb{E}[U|T=1] - \mathbb{E}[U|T=0])' \kappa_0. \quad (24)$$

The first two terms in (24) are invariant across reallocations, while the last is not.²⁰ However if we additionally impose the no matching condition (20) the third term drops out leaving

$$c_0 = \gamma_0 + \phi'_0 \delta_0, \quad (25)$$

which does measure the causal effect of exogenous changes in fraction Minority on outcomes.

I conclude that regression (17), is informative about neighborhood effects when conditions (19) and (20) hold. Specifically b_0 and c_0 are useful for predicting the effects of “doubly randomized” reallocations of households across neighborhoods (Graham 2008, 2011; Graham, Imbens and Ridder, 2010). First, the social planner selects a feasible distribution of Minority fraction across neighborhoods. If the status quo assignment is heavily segregated, the planner may choose a more integrated distribution. Second, she fills Minority and White “spots” in each neighborhood by taking independent random draws from the populations of Minorities or Whites as appropriate. Third, neighborhoods, so formed, are assigned at random to locations. Steps two and three of this procedure ensure that the new neighborhood assignment obeys the no sorting and matching restrictions. No sorting and no matching requires that individuals, *conditional on their type*, are “as if” randomly assigned to groups and, further, that groups so formed are “as if” randomly assigned to locations.

Expressions (21) and (22) also provide a framework for understanding conventional neighborhood effects analyses based on observational data, where conditions (19) and (20) are unlikely to hold. Consider the effect of sorting on unobservables on the coefficient on Minority status. If high “background” households, irrespective of type, tend to be concentrated in predominately White neighborhoods, then $\mathbb{C}(s(Z), \mathbb{E}[B|Z]) < 0$. Sorting by background biases the coefficient on Minority upward relative to the no sorting benchmark case (cf.,

²⁰In practice we could replace the no sorting on unobservables condition with the weaker requirement that $\mathbb{C}(s(Z), \mathbb{E}[B|Z]) = 0$. However the stronger condition more naturally generalizes to semiparametric models and is maintained here (cf., Graham, Imbens and Ridder, 2010).

equation (23)). When the Minority population share is small (p small), and/or segregation by race large (η^2 big), the bias will be even larger.

Now consider the coefficient on fraction Minority in (17). Terms three and four in (22) are bias terms. The third term is proportional to the sorting bias term in (21). If sorting follows the form sketched in the previous paragraph this term will be negative. The fourth term is due to Minority-White gaps in location-specific amenities. If predominately Minority neighborhoods have characteristics which otherwise tend to lower outcomes, then this term will be negative as well. Collectively sorting and matching will exaggerate any negative, or attenuate any positive, impact of Minority exposure on the outcome.

These conclusions are specific to the assumptions made about sorting on unobservables and matching; other assumptions could lead to opposite results. The point of walking through a specific example is to show how the structure of the prototypical neighborhood effects regression analysis may be utilized to be more precise about (i) assumptions needed for causal inferences to be valid and (ii) how to think about likely biases when causal inference is not warranted.

While conditions (19) and (20) are strong, there are examples of real world datasets where they are plausible. For example in the Project STAR class size reduction experiment, students were first randomly grouped into classes (no sorting), with classes then assigned randomly to teachers (no matching) (see Graham (2008) for additional details and caveats). The value of such experiments for enhancing our understanding of peer group and neighborhood effects is considerable. It is illuminating to consider a hypothetical housing mobility experiment satisfying double randomization. Consider a public housing authority that has constructed new housing units in multiple locations of a city. Among all households allocated to a new unit, the authority could (i) randomly group sets of households together (no sorting) and then (ii) randomly assign these groups to different newly constructed locations (no matching). This is not a completely implausible scenario, but it is also far from the structure of any type of experiment implemented to date.

Conditioning on observed household- and neighborhood-level attributes

In the context of observational neighborhood effect analyses, researchers' reactions to the biases caused by sorting and matching has typically been to improve measurement. Specifically to add proxies for A and U to equation (17). This approach has led to innovative data collection strategies (e.g., the “ecometrics” espoused by Sampson and Raudenbusch (1999)). Datasets like the L.A. FANS include a rich array of theoretically motivated and carefully

measured family- and neighborhood-level proxies for A and U .

To understand how this approach work in practice, let W be a vector of predetermined household-level attributes (e.g., parents' schooling) and X a vector of exogenous neighborhood-level attributes (e.g., distance to closest major employment center). Redefine a household's unobserved "background", B , as

$$A = \pi_0(W, X) + \phi_0(W, X)T + B, \quad \mathbb{E}[B|T, W] = 0.$$

There is no sorting on unobservables, *conditional* on W and X if

$$\mathbb{E}[B|T, W, X, Z] = \mathbb{E}[B|T, W, X] = 0 \quad (26)$$

Condition (26) implies that among, say, Minority households (i) living in neighborhoods with the same observable amenities (X) and (ii) having identical measured backgrounds (W), one's neighborhood of residence, Z , is not predictive of the unobserved household-level determinants of Y . More concretely, it implies, for example, that among Minority households with college-educated heads living in neighborhoods near a major employment center, the average unobserved component of background, B , is similar across neighborhoods.

There is no matching on unobservables, *conditional* on W and X if

$$\mathbb{E}[U|T, W, X] = \mathbb{E}[U|W, X] = 0, \quad (27)$$

or if, among households homogenous in W and X , Minority status is not predictive of unobserved neighborhood amenities. Again, concretely, it implies that college-educated households living near major employment centers, enjoy similar unobserved neighborhood amenities irrespective of their Minority status.

Let $\mathbb{E}^*[Y|X; Z]$ denote the best linear predictor of Y given X conditional on Z . Wooldridge (1997, Section 4) summarizes the basic properties of conditional linear predictors (CLPs). Under (26) and (27) the conditional linear predictor of Y given T and $s(Z)$, conditional on W and X , is given by

$$\begin{aligned} \mathbb{E}[Y|T, s(Z); W, X] &= \alpha_0 + \pi_0(W, X)'(\beta_0 + \gamma_0) + (\beta_0 + \phi_0(W, X)'\lambda_0)T \\ &\quad + (\gamma_0 + \phi_0(W, X)'\delta_0)s(Z) \end{aligned} \quad (28)$$

In practice researchers generally impose the additional restrictions that $\phi_0(W, X) = \phi_0$ (i.e., is constant in W and X) and $\pi_0(W, X) = \Pi_W W + \Pi_X X$ (i.e., is linear in W and X). With these two extra functional-formal assumptions, conditions (26) and (27), provide a

formal justification for least squares “kitchen-sink” neighborhood-effects regression analyses (e.g., Datcher, 1982; Brooks-Gunn, Duncan, Klebanov and Sealant, 1993; Leventhal and Brooks-Gunn, 2000).

An interesting example of identification via conditioning is provided by Bayer, Ross, and Topa (2008). In that paper a neighborhood corresponds to a census block, while W corresponds to a vector of census tract indicators. Recall that census tracts contain multiple census blocks. In their setup condition (26) rules out within-track sorting on unobservables; an assumption they motivate by appealing to frictions in the housing market (i.e., households may be able to choose their tract of residence, but probably not their exact block). Condition (27) requires, effectively, that the Minority and White census blocks, within a given tract, have similar neighborhood amenities. Hoxby (2000) and Ammermueller and Pischke (2009) employ an analogous research design to study peer effects at the classroom level. With the increasing availability, albeit generally by special agreement, of rich geocoded datasets, this approach to identification may be useful to empirical researchers.

My own view is that approaches based on covariate adjustment, as in other areas of causal analysis, have a meaningful role to play in neighborhood effects research. Such approaches can motivate the collection of better data, which often has other auxiliary benefits. Of course, as in other areas of causal analysis, approaches based on covariate adjustment are not always compelling (cf., Durlauf, 2006). This observation has led researchers to develop other research designs for neighborhood effects analysis.

Within- and between-neighborhood decomposition

Graham and Hahn (2005) emphasize the panel data structure of conventional neighborhood effects analysis. Under (16) the between-neighborhood outcome equation equals

$$\begin{aligned}\mathbb{E}[Y|Z] &= \alpha_0 + \pi'_0(\delta_0 + \lambda_0) + (\beta_0 + \gamma_0 + \phi'_0(\delta_0 + \lambda_0))s(Z) \\ &\quad + \mathbb{E}[B|Z]'(\delta_0 + \lambda_0) + U'\kappa_0,\end{aligned}\tag{29}$$

while the within-neighborhood outcome equation equals

$$Y - \mathbb{E}[Y|Z] = (\beta_0 + \phi'_0\lambda_0)(T - s(Z)) + (B - \mathbb{E}[B|Z])'\lambda_0.\tag{30}$$

Let b_W equal the coefficient on T in the within-neighborhood population regression and b_B the coefficient on $s(Z)$ in the between-neighborhood regression. A bit of algebra (or a little

introspection; cf., Mundlak (1978)), gives

$$\begin{aligned}b_0 &= b_W \\c_0 &= b_B - b_W.\end{aligned}$$

The within- and between-neighborhood regression functions provide no additional information beyond that already contained in the short regression function (17). Nevertheless the decomposition has pedagogic value and can facilitate the implementation of certain instrumental variables methods. Graham and Hahn (2005) provide details.

Moving to Opportunity (MTO) and other instrumental variables methods

A large body of quantitative observational studies on neighborhood effects is now available. Early studies typically matched household-level data from, for example, the Panel Study of Income Dynamics (PSID), with neighborhood-level variables constructed from the Census (e.g., tract-level poverty rate). An early example, within economics, is provided by Datcher (1982). More recent studies often utilize special-purpose datasets, such as the L.A. FANS.

Prior reviewers have generally been dismissive of this work. For example, Durlauf (2006), argues that “with very few exceptions...empirical studies of neighborhood effects based on observational data have failed to deal seriously with the possible statistical biases induced by self-selection into neighborhoods” (p. 159). This assessment continues to be true today, but it does not, in my view, imply that such work is without value. The establishment of robust conditional correlations provides the library of “stylized facts” that theoretical, as well as more sophisticated empirical work, should seek to explain. For example, it is a *fact* that the Black-White gap in many outcomes is greater in more segregated cities. Explaining this pattern is a legitimate project for the social sciences. At the same time it is correct that basic observational studies are arguably of only limited value in justifying, for example, policies of school and/ residential desegregation.

It is in this context that projects such as the Moving to Opportunity (MTO) demonstration have arisen. The MTO program provided randomly assigned opportunities for some households to leave public housing and move to a low poverty neighborhood with the help of rent subsidies. Ludwig (2012) provides an useful overview of the MTO research program, and Kling, Liebman and Katz (2007) and Ludwig et. al. (2008, 2012) summaries of its main findings. In what follows, since my purposes is primarily pedagogical, I abstract from many

nuances of the experiment's design, instead taking a more synoptic approach.

Let $W \in \{0, 1\}$ now be an indicator for whether a household has access to a rent subsidy. To keep the exposition more policy-relevant, and in line with the MTO example, let T now indicate whether a household falls below (some multiple of) the poverty line or not. Assume that that W is independent of B . This allows for random assignment of vouchers to households, but with differing assignment probabilities across poor and non-poor households (including an assignment probability of zero for non-poor households).

Let $\Xi_S = (\mathbb{E}[s(Z)|W = 1] - \mathbb{E}[s(Z)|W = 0])$ be the average difference in neighborhood poverty across voucher and non-voucher households. Similarly define

$$\begin{aligned}\Xi_{\bar{B}} &= (\mathbb{E}[\mathbb{E}[B|Z]|W = 1] - \mathbb{E}[\mathbb{E}[B|Z]|W = 0]) \\ \Xi_U &= \mathbb{E}[U|W = 1] - \mathbb{E}[U|W = 0]\end{aligned}$$

as the average difference in neighbors' background and neighborhood amenities respectively across voucher and non-voucher households. All these contrasts are post-assignment, and hence post-move (if applicable). Let $p_W = \mathbb{E}[W]$ equal the marginal frequency of voucher receipt.

Consider the linear instrumental variables fit of Y onto a constant, T and $s(Z)$, using W as an excluded instrument for $s(Z)$. Note that, since W is randomly assigned conditional on T , this would seem to be a promising and valid application of the method of instrumental variables. The coefficient on $s(Z)$ in this fit has a probability limit of (see Appendix B)

$$\begin{aligned}c_{IV} &= (\gamma_0 + \phi'_0 \delta_0) \\ &+ \frac{p_W(1 - p_W)\Xi_{\bar{B}} - (\mathbb{E}[\mathbb{E}[B|Z]|T = 1] - \mathbb{E}[\mathbb{E}[B|Z]|T = 0])p(\mathbb{E}[W|T = 1] - p_W)'}{p_W(1 - p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T = 1] - p_W)} \delta_0 \\ &+ \frac{p_W(1 - p_W)\Xi_U - (\mathbb{E}[U|T = 1] - \mathbb{E}[U|T = 0])p(\mathbb{E}[W|T = 1] - p_W)'}{p_W(1 - p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T = 1] - p_W)} \kappa_0,\end{aligned}\tag{31}$$

which is *not* free of sorting and matching bias.

Although W is conditionally randomly assigned given T and the underlying outcome model is linear, the instrumental variables fit in this setting does not recover an interpretable causal parameter. The reason is that T is not exogenous. Under sorting and matching a household's type will covary with unobserved neighbors' attributes $m_A(Z)$ and neighborhood amenities U . Equation (31) suggests that the application of instrumental variables methods to neighborhood effects regression analysis may involve some subtlety.

In the Moving To Opportunity (MTO) demonstration experiment, vouchers were only avail-

able with positive probability to $T = 1$ households.²¹

In this case $\mathbb{E}[W|T = 1] = p_W$ so that (31) simplifies to

$$c_{MTO} = (\gamma_0 + \phi'_0 \delta_0) + \frac{\Xi'_B \delta_0}{\Xi_S} + \frac{\Xi'_U \kappa_0}{\Xi_S}. \quad (32)$$

Equation (32) *is* a measure of the “effect of place” *free* of sorting and matching bias. Expression (32) is also consistent with the interpretation of instrumental variable fits provided by neighborhood effects researchers (e.g., Ludwig et. al. 2012). Importantly, it *is not* a measure of neighborhood-level spillovers. Furthermore, as a composite parameter, it may be close to zero even if many of its components are non-zero due to offsetting effects.

Whether (32) is an interesting parameter or not, depends upon the question of interest. It *is not* a parameter that can be used to predict the effects of (doubly randomized) associational redistributions. The intent to treat (ITT) effects of voucher receipt reported by MTO researchers are likely good estimates of the benefits of a marginal increase in voucher availability among the population of MTO participants.

The instrumental variables coefficient on $s(Z)$, (32), varies with the effect of voucher receipt on the change in (i) neighborhood fraction poor (Ξ_S) and (ii) unobserved neighbors’ background (Ξ_B) and unobserved neighborhood amenities (Ξ_U). While voucher receipt appears to have had rather durable effects on households’ exposure to concentrated poverty, other aspects of neighborhood changed much less for MTO voucher recipients (cf., Sampson, 2012a). For example, various (measurable) dimensions of school quality appeared to have been little affected by voucher receipt.

While a rejection of the null that $c_{MTO} = 0$ requires that at least one of γ_0 , δ_0 or κ_0 differ from zero and hence that “place matters”. The interpretation of such a rejection cannot be equated with the presence of neighborhood spillovers (i.e., γ_0 and/or δ_0 different from zero), since such a rejection may arise from a pure amenity effect ($\kappa_0 \neq 0$). Furthermore a failure to reject the null could easily reflect the operation of various offsetting effects. An MTO-type-induced move may improve some dimensions of neighborhood quality, while worsening others (cf., Durlauf and Ioannides, 2010; as an example, lower poverty neighborhoods with available rental units could be located in neighborhoods far from employment centers, with less transit options, etc.). Furthermore the relative importance of these different dimensions may vary with the outcome under study. Expression (32) may be helpful for understanding

²¹In fact MTO participants were a self-selected subgroup of eligible households. While this fact has important implications for the interpretation of the MTO experimental results, it is less relevant to the methodological points I aim to make here.

patterns of cross-site heterogeneity found by MTO researchers as well as the findings of researchers exploiting closely related designs (e.g., Oreopoulos, 2003).

Cross-city research designs

Several researchers have proposed estimation methods involving aggregation of (16) to the city-level as a remedy for sorting bias (e.g., Evans, Oates and Schwab, 1992; Cutler and Glaeser, 1997; Card and Rothstein, 2007; Ananat, 2011). This approach requires observations from a cross-section of cities, typically operationalized by a metropolitan statistical area (MSA), and restrictions on how the joint distribution of $(A', T)'$ varies across cities. A formalization and illustration of this approach is provided here.

The researcher has access to a nationally representative sample of (T, Y) pairs. Assume further that this sample is geocoded, such that each observation can be assigned to a specific MSA. As a concrete example, the NLSY79 and NLSY97 restricted-use geocode files, available from the Bureau of Labor Statistics (BLS) by special agreement, may be used to assign each respondent to an MSA of residence (at the time of adolescence). Index MSAs by $c = 1, \dots, N$ and sampled respondents within a city by $i = 1, \dots, M_c$. Assume that various measures of residential segregation by race, corresponding to the period coinciding with the respondent's adolescence, are available for each city (e.g., computed using information in the Neighborhood Change Database (NCDB)).²²

We begin by modifying (16) to incorporate a city-specific effect (i.e., intercept):

$$\begin{aligned} \mathbb{E}[Y_{ci} | T_{ci}, s(Z_{ci}), m_A(Z_{ci}), U_{ci}, A_{ci}] &= \alpha_c + \beta_0 T_{ci} + \gamma_0 s(Z_{ci}) \\ &+ m_A(Z_{ci})' \delta_0 + U_{ci}' \kappa_0 + A_{ci}' \lambda_0. \end{aligned} \quad (33)$$

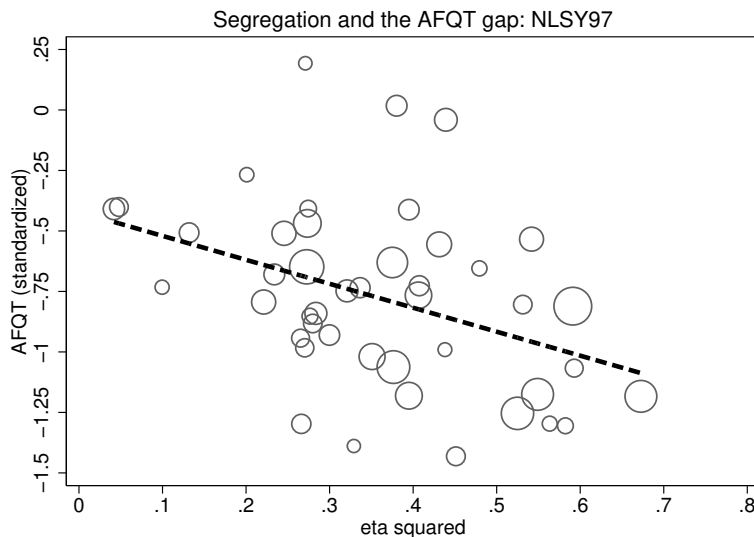
The presence of α_c allows the mean outcome to vary across cities for reasons unrelated to segregation.

Let $\mathbb{E}^*[Y | X; c]$ denote the best linear predictor of Y given X conditional on residence in city c . Let $\mathbb{V}(Y | c)$ and $C(X, Y | c)$ denote city-specific variances and covariances. Some basic algebra gives a CLP of fraction minority in one's neighborhood given own Minority status of

$$\mathbb{E}^*[s(Z_{ci}) | T_{ci}; c] = (1 - \eta_c^2) p_c + \eta_c^2 T_{ci} \quad (34)$$

²²In what follows I assume that these segregation measures are known precisely (i.e., not up to some sampling error). If these segregation measures are themselves estimated, then this fact will need to be incorporated into inference procedures. In the empirical illustration below η_c^2 is constructed from tract-level information in the NCDB; sampling variation is not an issue.

Figure 3: Minority vs. White AFQT gap and residential segregation across 44 MSAs, NLSY97



Source: National Longitudinal Survey of Youth 1997 (NLSY97), Neighborhood Change Database (NCDB) and author’s calculations.

Notes: The NLSY97 sample consists of 8,985 youths, of which a total of 7,263 resided in an MSA at baseline (1,796 Hispanic respondents, 1,943 Black respondents, and 3,524 non-Hispanic, non-Black respondents). The estimation sample used here includes the 4,419 respondents who resided in a city with at least 5 White responding households as well as 5 Black and/or Hispanic responding households. “AFQT score at age 16” corresponds to the inverse normal CDF transform of the adjusted AFQT percentile score used in Altonji, Bharadwaj and Lange (2012). This score was normalized to the distribution of percentile scores across NLSY79 respondents aged 16 at the time of test-taking in 1980. Across respondents from the reference population of American Youth aged 15 to 23 in 1980 this transform of percentile scores is mean zero with unit variance. Since the Altonji, Bharadwaj and Lange (2012) percentile scores are normalized to a different reference group (both in terms of age-of-testing and year-of-birth), “AFQT score at age 16” need not be mean zero with unit variance (across all 7,002 respondents with valid AFQT scores its mean is -0.0001 and its standard deviation is 0.9893). AFQT scores were only available for a subset of the target sample of 7,263 MSA-residents. The y-value of each point in the figure corresponds to the coefficient on a dummy variable for Minority (i.e., Black or Hispanic) in the least squares fit of standardized AFQT score onto a constant and Minority using observations from a single MSA. The x-value corresponds to the eta-squared measure of Minority segregation in 2000. The size of the scatter points are increasing in the estimated precision of the corresponding Minority-White AFQT gap. A total of 44 MSAs are represented in the figure.

where η_c^2 is the eta squared segregation measure for city c and $p_c = \Pr(T = 1|c)$ is the city-wide fraction Minority. In highly segregated cities ($\eta_c^2 \rightarrow 1$) own race is very predictive of neighbors' race. In integrated cities ($\eta_c^2 \rightarrow 0$) the city-wide fraction Minority has more predictive value.²³

Define

$$\begin{aligned}\phi_c &= \mathbb{E}[A_{ci}|T_{ci} = 1, c] - \mathbb{E}[A_{ci}|T_{ci} = 0, c] \\ v_c &= \mathbb{E}[\mathbb{E}[B_{ci}|Z_{ci}]|T_{ci} = 1, c] - \mathbb{E}[\mathbb{E}[B_{ci}|Z_{ci}]|T_{ci} = 0, c] \\ \tau_c &= \mathbb{E}[U_{ci}|T_{ci} = 1, c] - \mathbb{E}[U_{ci}|T_{ci} = 0, c]\end{aligned}$$

to be the Minority-White gap within city c in (i) “background” (ϕ_c), (ii) neighbors' background (v_c), and (iii) neighborhood amenities (τ_c) respectively. The first of these terms is a vector of average differences between Minorities and Whites. The latter two terms are vectors of average differences in features of their neighborhoods. Specifically in the unobserved attributes of their neighbors and the unobserved non-composition-based characteristics of their neighborhoods. Both of these are measures of average differences in neighborhood quality between Minorities and Whites. To the extent that these components of neighborhood quality directly influence the outcome of interest (i.e., $\delta_0 \neq 0$ and/or $\kappa_0 \neq 0$), they both are drivers of neighborhood effects, broadly defined.

Using (33), (34) and the notation defined above we get a city-specific short regression, deviated from city-specific means, of

$$\begin{aligned}\mathbb{E}[Y_{ci}|T_{ci}; c] - \mathbb{E}[Y_{ci}|c] &= \{\beta_0 + (\gamma_0 + \phi'_c \delta_0) \eta_c^2 \\ &\quad + v'_c \delta_0 + \tau'_c \kappa_0 + \phi'_c \lambda_0\} \\ &\quad \times (T_{ci} - p_c).\end{aligned}\tag{35}$$

Let ϕ_0 now equal the mean of ϕ_c across cities (i.e., $\phi_0 = \mathbb{E}[\phi_c]$) with ν_0 and τ_0 analogously defined. Equation (35) indicates that the Minority-White *outcome* gap in city c – $\text{GAP}_c = \mathbb{E}[Y_{ci}|T_{ci} = 1; c] - \mathbb{E}[Y_{ci}|T_{ci} = 0; c]$ – varies with its degree of segregation, as measured by the eta-squared, η_c^2 , index:

$$\text{GAP}_c = a_0 + (\gamma_0 + \phi'_0 \delta_0) \eta_c^2 + V_c\tag{36}$$

²³Note that

$$\frac{\mathbb{V}(s(Z))}{p(1-p)} = \frac{\mathbb{E}[s(Z)^2] - p^2}{p(1-p)} = \frac{\mathbb{E}\left[\frac{s(Z)}{p}s(Z)\right] - p}{1-p} = \frac{1-p}{1-p} = \eta^2.$$

with

$$\begin{aligned} a_0 &= \beta_0 + v'_0 \delta_0 + \tau'_0 \kappa_0 + \phi'_0 \lambda_0 \\ V_c &= (v_c - v_0)' \delta_0 + (\tau_c - \tau_0)' \kappa_0 + (\phi_c - \phi_0)' \lambda_0 + (\phi_c - \phi_0)' \delta_0 \eta_c^2. \end{aligned}$$

Here V_c varies with a city's Minority-White gap in (i) neighbors' unobserved "background", (ii) exogenous neighborhood attributes, (iii) own "background", and (iv) the interaction of own-background with measured segregation. If

$$\mathbb{E} [V_c \eta_c^2] = 0 \tag{37}$$

then, by equations (35) and (36), a least squares fit of the city-specific measure of the Minority-White outcome gap, GAP_c , onto a constant and η_c^2 will provide a consistent estimate of $\gamma_0 + \phi'_0 \delta_0$. Observe that this coincides with the coefficient on fraction minority in the prototypical neighborhood effects regression (17) under no sorting and matching (i.e., when conditions (19) and (20) hold).

Restriction (37) is sometimes referred to as a no "differential sorting" across cities assumption (e.g., Card and Rothstein, 2007). To understand this language consider the common case where δ_0 and κ_0 are *a priori* presumed equal to zero. In that case condition (37) requires that, across cities, variation in the "background gap" is uncorrelated with variation in racial segregation (i.e., $\mathbb{C}(\phi_c, \eta_c^2) = 0$). Sorting across cities *is* allowed in the sense that differences in average "background" across cities do not threaten identification. However the background-gap across cities, if it varies, must do so independently of measured segregation.

Note that, again *in the special case* where $\delta_0 = \kappa_0 = 0$, aggregation *does* eliminate biases due to sorting on unobservables. In that case the coefficient on $s(Z)$ in the *within-city* neighborhood effects regression (17) is (see (22) above)

$$c_0 = \gamma_0 + \frac{1}{\eta^2(1 - \eta^2)} (\mathbb{E} [\mathbb{E} [B|Z] | T = 1] - \mathbb{E} [\mathbb{E} [B|Z] | T = 0])' \lambda_0$$

with the second term due to sorting and where I have also made use of the equality

$$\frac{\mathbb{C}(s(Z), \mathbb{E}[B|Z])}{p(1 - p)} = \mathbb{E} [\mathbb{E} [B|Z] | T = 1] - \mathbb{E} [\mathbb{E} [B|Z] | T = 0].$$

In contrast, the coefficient on η_c^2 in the *cross-city* regression of the Minority-White outcome gap onto a constant and η_c^2 is equal to γ_0 alone. Under these conditions aggregation does eliminate biases due to sorting, as is often asserted in empirical work.

However, when δ_0 and κ_0 possibly differ from zero, condition (37) requires maintaining additional (strong) assumptions. First, for $\delta_0 \neq 0$, we require zero covariance between measured racial segregation and $(v_c - v_0)' \delta_0$, the first element of V_c . This implies that the Minority-White gap in neighbors' background is uncorrelated with the degree of metropolitan-area segregation. Very loosely speaking this assumption does not rule out within-city sorting on unobservables, but it does constrain such sorting to be similar across cities. When $\delta_0 \neq 0$ we also require strengthening the zero covariance condition $\mathbb{C}(\phi_c, \eta_c^2) = 0$. For example if ϕ_c is mean-independent of η_c^2 , then $\{(\phi_c - \phi_0)' \delta_0 \eta_c^2\} \eta_c^2$, the last term in $V_c \eta_c^2$, will be mean zero (along with the third term). In practice if the researcher is willing to assume that $\mathbb{C}(\phi_c, \eta_c^2) = 0$, then they ought to be willing to assume that $\mathbb{E}[\phi_c | \eta_c^2] = \phi_0$. It is difficult to imagine plausible sorting processes that imply the weaker (former) condition, but not the stronger (latter) condition.

Second, for $\kappa_0 \neq 0$, we require zero covariance between η_c^2 and $(\tau_c - \tau_0)' \kappa_0$, the second element of V_c . This implies that the degree to which race predicts exogenous neighborhood characteristics is uncorrelated with the level of racial segregation.

In practice, it seems likely that own race will be a better predictor of unobserved neighborhood attributes in highly segregated cities, for no other reason than in such cities Minorities and Whites live apart. These considerations suggest that recovering a consistent estimate of γ_0 from the cross-city correlation of GAP_c and η_c^2 is difficult. At the same time, one may be able to learn about the “effects of place” broadly defined from such an analysis.

To understand this claim, considering maintaining the assumption that $E[\phi_c | \eta_c^2] = 0$ (i.e., that the city-specific Minority-White background-gap is mean independent of measured segregation). This is a strong assumption, but is one that can be indirectly assessed (see Figure 4 for some mixed support in the present case). For example, one could correlate gaps in measured components of family background with segregation. Under this assumption the least squares fit of G_c onto a constant and η_c^2 is consistent for

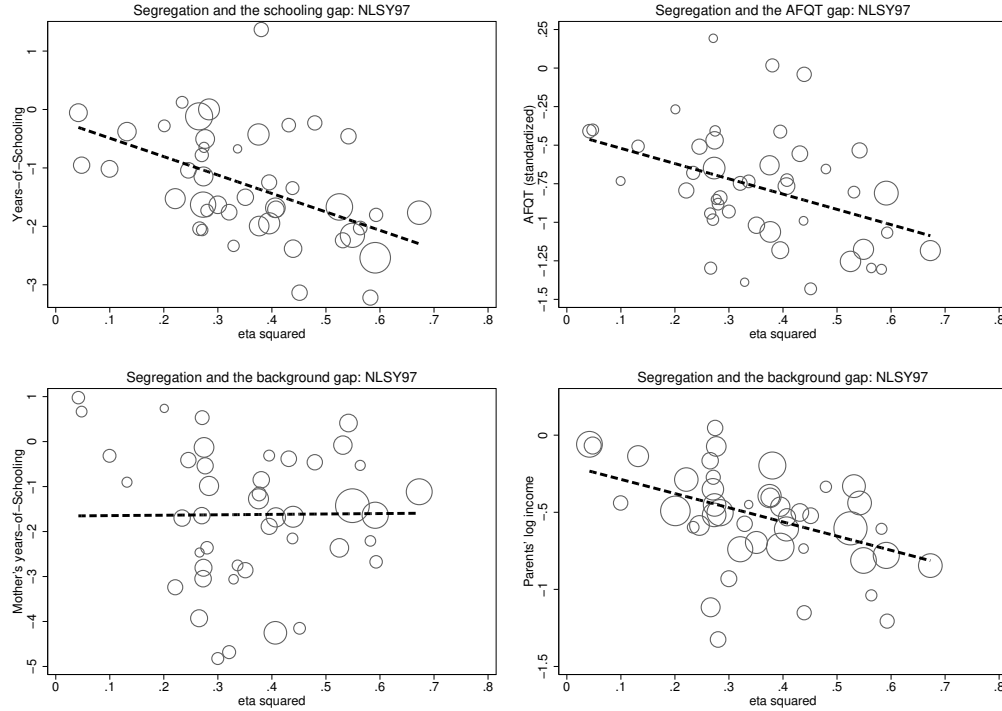
$$\mathbb{E}^* [\text{GAP}_c | \eta_c^2] = \beta_0 + \phi_0' \lambda_0 + \Pi_v' \delta_0 + \Pi_\tau' \kappa_0 + (\gamma_0 + (\phi_0 + \Pi_{v\eta})' \delta_0 + \Pi_{\tau\eta}' \kappa_0) \eta_c^2 \quad (38)$$

where

$$\begin{aligned} \mathbb{E}^* [v_c | \eta_c^2] &= \Pi_v + \Pi_{v\eta} \eta_c^2 \\ \mathbb{E}^* [\tau_c | \eta_c^2] &= \Pi_\tau + \Pi_{\tau\eta} \eta_c^2. \end{aligned}$$

Here the coefficient on η_c^2 is not structural. In particular it is not consistent for γ_0 (or $\gamma_0 + \phi_0' \delta_0$) and hence informative about the effects of reallocations on outcomes. The coefficient

Figure 4: Evidence for no differential sorting assumption, NLSY79



Source: National Longitudinal Survey of Youth 1997 (NLSY97), Neighborhood Change Database (NCDB) and author's calculations.

Notes: The top two figures plot the relationship between segregation and city-specific estimates of the Minority-White AFQT and years-of-completed schooling gaps in the NLSY79 cohort. The former of these is also shown in Figure 3 above. The bottom two figures show the relationship between city-specific estimates of the Minority-White gap in two background measures, mother's years-of-schooling and family income during adolescence, and segregation. While the bottom left-hand figure is consistent with the no differential sorting assumption, the bottom right-hand figure is not.

is, however, a measure of the effects of “place” on outcomes. It captures the collective effects of differences in the racial composition of ones’ neighborhoods, their average background, and neighborhood amenities on the Minority-White outcome gap. Importantly, *it is not affected by sorting*. I conclude that while a structural interpretation of cross-city correlations between outcome gaps and segregation requires strong assumptions, a looser interpretation – as a measure of the “effects of place” – may be compelling in some circumstances (similar to the interpretation of the IV-MTO estimand (32) above). Of course, the policy-relevance of this measure, unlike that of the composite parameter $\gamma_0 + \phi'_0\delta_0$, which can be used to predict the effects of doubly randomized reallocations, is not immediately clear.

Figure 3 depicts the relationship between Minority (Black or Hispanic) - White gaps in AFQT scores and residential segregation across 44 large MSAs (see the notes to the Figure for details on the estimation sample). The Minority-White gap in test scores is substantial and significantly larger in highly segregated cities. Under condition (37) the variance-weighted least squares (OLS) fit of the estimated AFQT gap onto a constant and η_c^2 will provide a consistent estimate of $\gamma_0 + \phi'_0\delta_0$:²⁴

$$\widehat{\text{AFQTGAP}}_c = \frac{-0.4214}{(0.0857)} + \frac{-0.9901}{(0.2364)} \text{ETA}_c^2.$$

This fit indicates that the AFQT gap averages under 0.5 standard deviations in America’s least segregated cities but over 1 standard deviations in its most segregated cities. These differences are precisely determined. Under condition (37) the data suggests that we can strongly reject the null that $\gamma_0 + \phi'_0\delta_0 = 0$. Again, maintaining condition (37), the fit suggests that a student, if contrary to fact, was instead raised in a neighborhood with a fraction Minority 10 percent lower, her AFQT score would, in expectation, be about 0.1 standard deviations higher.

As a pure peer effect, a value of $\gamma_0 + \phi'_0\delta_0$ equal to -1 seems implausibly large. If we instead interpret the coefficient on the eta-square measure in terms of regression (38), our view changes. In that case the measured relationship is not presumed to be a pure peer effect. Rather the effect measures the cumulative impact of a bundle of human-capital producing amenities that differ, on average, across Minority and White neighborhoods. These could include, among of things, systematic differences in school and teacher quality, differences in environmental quality (e.g., indoor air quality, noise, etc.), differences in exposure to violence and other risks, as well as a pure peer effect. Each of these items has been shown

²⁴Since I am weighting I technically need the mean independence analog of (37). Unweighted results are not appreciably different.

to independently influence academic performance. When many drivers of human capital accumulation vary across neighborhoods, the estimated effect of place can be plausibly large. While the interpretation of the pattern shown in Figure 3 is not straightforward, what is without dispute is that the AFQT gap varies significantly with measured segregation. Minority adolescents fare worse relative to their White counterparts, in terms of AFQT scores, in cities where they live in separate neighborhoods. Again, whether this is due to neighborhood effects (i.e., $\gamma_0 + \phi'_0 \delta_0 < 0$), or has some other cause, is open to debate.

Additional identification challenges and opportunities

Policies other than associational redistribution

The discussion above has presumed that the researchers' interest lies in understanding the likely effects of an associational redistribution. Associational redistributions are resource neutral and, as such, leave the joint distribution of T and A unchanged. In certain settings, however, interest may center on the effects of policies which *do* change the conditional distribution of T given A . For example T may be an indicator for whether an individual has completed high school or college. In such a setting the composite parameter, $\gamma_0 + \phi'_0 \delta_0$ would not be invariant to policies which, say, increase the rate of college completion. In order to understand the effects of such a policy on the distribution of outcomes, knowledge of γ_0 is required. Acemoglu and Angrist (2000), and Ciccone and Peri (2006) discuss identification and estimation of γ_0 in such settings. They also provide empirical illustrations.

Endogenous effects and social multipliers

In model (16) all social spillovers, using the typology of Manski (1993), are of the exogenous type. The presence of endogenous interactions, whereby the propensity of an individual to engage in some activity covaries with the frequency of participation among her peers, can raise new issues.

One feature of (16) is that all individuals within a group, in our case, a neighborhood, are equally influenced by one another. Furthermore, there are no cross group linkages. This simple model can be viewed as the reduced form of a model also admitting endogenous effects. This point is developed carefully in Manski (1993) and Brock and Durlauf (2001a). If the policies of interest are associational redistributions, then knowledge of the reduced form will be sufficient.

Once one departs from the benchmark “linear-in-means” model, however, the endogenous vs. exogenous spillovers distinction has content. One direction of departure involves maintaining linearity, but assuming that agents are embedded in a non-trivial network (cf., Calvo-Armengol, Patacchini and Zenou, 2009; Bramoullé, Djebbari and Fortin, 2009; Blume, Brock, Durlauf and Jayaraman, 2015). Each individual is connected to only a finite subset of all other individuals. In such a setting, the presence of an endogenous effect, implies that an intervention acting upon a single-individual can have social multiplier effects which percolate across the entire network (cf., Graham, 2015c). By picking certain, more “central”, individuals within a network to intervene upon, indirect spillover effects can be maximized (see Kim et al. (2015) for an interesting empirical study related to technology adoption). In contrast, if only exogenous effects are present, then the effects of interventions stay local; with spillover effects being confined to direct friends alone. The non-trivial network structure also generates new avenues for identification (cf., Bramoullé, Djebbari and Fortin, 2009; Blume, Brock, Durlauf and Jayaraman, 2015).

Another direction of departure involves introducing nonlinearity, while maintaining the simple group interaction structure. Brock and Durlauf (2001a,b) provide an early, and seminal, analysis of such models. In nonlinear models, endogenous effects may result in multiple equilibrium action configurations. Multiple equilibria introduce the possibility of engineering improvements in the distribution of outcomes without having to (i) engage in associational redistribution or (ii) change the conditional distribution of T given A .

In addition to their potential policy implications, nonlinear models of strategic interaction raise a number of challenging, albeit also interesting, identification and estimation issues. de Paula (2013) provides a useful survey of the relevant literature, mostly inspired by empirical settings common in industrial organization. Brock and Durlauf (2007) present some interesting identification results specific to the neighborhood effects setting.

My own view is that while it is now possible to point to a modest collection of studies which provide relatively strong evidence for the existence of social spillovers (e.g., Graham, 2008; Bayer, Hjalmarsson and Pozen, 2009), there are few studies which characterize the relative importance of endogenous and exogenous effects. A priori we might expect endogenous interactions to be relatively important for explaining some types of behaviors, such as the initiation of sexual activity, or drug use, among adolescents (cf., Card and Giuliano, 2013). For other outcomes, such as measured academic achievement (i.e., grades or test scores), the a priori case for endogenous effects is, in my view, weaker.

Given the rather profound policy implications of endogenous interactions, as well as the interesting identification challenges and opportunities their presence creates, empirical work

in this area would be an interesting undertaking. Boucher, Bramoulle, Djebbari and Fortin (2014) and Liu, Patacchini, and Zenou (2014) are two recent, and careful, attempts to study endogenous social effects. Unfortunately, many existing discussions of endogenous versus exogenous effects in the empirical literature are often either non-credible, confused or both.

Nonlinear exogenous effects

Even without endogenous effects, nonlinearity raises complex identification, estimation and policy issues. Henderson, Mieszkowski and Sauvageau (1978) provide an early example of an empirical study which takes such nonlinearities seriously (both in terms of empirical, as well as policy, analysis). Graham, Imbens, Ridder (2010) provide a discussion of identification, estimation and social planning issues in nonparametric (i.e., nonlinear) models with exogenous social spillovers. The presentation of the more basic linear model in this survey, as well as the development of the no sorting and matching condition, was designed to make the more general results in Graham, Imbens, Ridder (2010) accessible to readers of this survey. Graham (2011, Section 5) provides additional material.

Structural models of locational sorting

The analysis thus far has viewed sorting and matching biases as nuisances best dealt with by an appropriate choice of research design. An alternative approach utilizes structural features of a locational sorting equilibrium to facilitate identification. Of course, taking a stand on the sorting process does make inference fragile to violations of the additional maintained assumptions. However explicitly modeling the sorting process can help to (i) increase efficiency, (ii) generate testable restrictions and (iii) facilitate the implementation of counterfactual policy analysis. Nesheim (2002) provides a nice illustration this approach (cf., Bayer, Ferreira, McMillan, 2007; Calabrese, Epplé, Romer and Sieg, 2006). Sieg and Holmes (2015) survey the literature on structural estimation of models of locational equilibrium, albeit not with a specific focus on models with spillover-induced sorting (which do raise special issues). Ioannides (2011) provides a survey with an explicit neighborhood effects focus.

My view is that structural models have an important role to play in the literature on neighborhood effects and, further, that there remains considerable scope for additional empirical work in this area. In justifying this view it is perhaps helpful to step back and take a bird's eye view of the identification problem. The key empirical challenge faced by neighborhood effects researchers in economics is to identify the average mapping from own and neighborhood

attributes into outcomes. Call this mapping the production function. The mean regression of the outcome onto own and neighborhood attributes, regression (17) above, typically does not identify the production function due to sorting and matching. A model of locational equilibrium allows the researcher to explicitly characterize sorting and matching bias *under maintained assumptions*. Furthermore such models can, again under certain assumptions, be used to correct for sorting and matching bias. Conceptually the idea is analogous to modeling input demand in order to identify a firm’s production function (Griliches and Mairrese, 1998; Olley and Pakes, 1996). When one further considers that the process of locational sorting is itself of intrinsic scientific and policy interest, the dividends to modeling sorting become even greater. Altonji and Mansfield (2014) provide a recent and interesting take on this line of reasoning.

At the same time I also wish to emphasize that fully nonparametric identification of the production function is always preferable. In a fully nonparametric/non-linear model, non-parametric identification requires exogenous variation in neighborhood quality of the type generated by, for example, (conditionally) double random assignment as described in Graham (2008), Graham, Imbens and Ridder (2010) and Graham (2011). An ideal empirical approach would therefore combine experimental estimation of the production technology with a model of locational sorting fitted to observational data. Here the experimentally-identified production function could be used to validate the sorting model, while the sorting model could be used to simulate the effects of various counter-factual policies. This last point is especially important when we consider that most realistic policies, by changing the incentives to sort, only indirectly implement associational redistributions (cf., Durlauf, 2006).

Neighborhood effects over the life course

Most empirical models of neighborhood effects are static. An individual’s neighborhood environment during, for example, childhood is modeled as fixed. In practice individuals’ neighborhood contexts generally evolve throughout childhood, whether due to the explicit movement of households from neighborhood to neighborhood, or due to secular changes in one’s environment due to, for example, gentrification.

In practice, therefore, it is likely better to conceptualize individuals’ as experiencing a *sequence* of neighborhood environments during childhood. Furthermore it seems sensible to allow for feedback from current neighborhood quality and (intermediate) outcomes into future neighborhood quality. For example parents who feel as though their children are poorly served by their neighborhood school, may seek out new schooling options for their children. In such a setting current test scores may be influenced by past neighborhood environment,

while at the same time also drive changes in future neighborhood environment. A dynamic perspective also allows researchers to ask when neighborhood context is most salient (e.g., early childhood vs. adolescence). One can also discriminate between episodic and persistent exposure to various types of neighborhood environments.

Relatively few researchers have taken a dynamic approach to neighborhood effects analysis (e.g., Wodtke Harding and Elwert, 2011; Sharkey and Elwert, 2011; Wodtke, 2013). This work utilizes methods of covariate adjustment under settings of sequential exogeneity developed by Jamie Robins in a series of papers (e.g., Robins 1986, 1997, 1999, 2000). These methods are neither conceptually, nor in terms of implementation, straightforward. Nevertheless they arguable have considerable merit in the context of observational studies of neighborhood effects. Graham (2015d) provides a “textbook” introduction to these methods. The scope of their application, given the increasing availability of large administrative databases with a longitudinal structure is considerable. For example, Chetty, Hendren and Katz (forthcoming), observe features of an individual’s neighborhood environment through the school age years, the intermediate outcomes of test scores, as well as the final outcome of adult earnings.

5 What next?

Prior reviewers of the empirical literature on social spillovers have been remarkably (and consistently) pessimistic. For example, Durlauf (2006), an otherwise enthusiastic proponent of the (potential) role of social interactions in various social and economic domains, argues that extant empirical work “suffers from sufficiently serious problems of interpretation that it should not strongly influence one’s prior beliefs...” (p. 163). Thankfully, the situation has improved somewhat in the intervening decade. It is now possible to point to a handful of empirical studies which, at least in my view, are of sufficiently high quality that one should use them to update their prior beliefs (e.g., Graham, 2008; Carrell, Fullerton and West, 2009; Gould, Lavy and Paserman, 2004, 2009, 2011).

There is also a growing collection of high quality studies which report null effects for many outcomes (e.g., Oreopoulos, 2003; Angrist and Lang, 2004; Ludwig et al. 2012). Many of these are housing mobility studies with strong, highly persuasive, research designs. Some researchers have interpreted this body of evidence as supportive of the null that neighborhoods and/or peers are relatively unimportant for a wide range of life outcomes (e.g., Angrist, 2014, p. 106). I have not attempted to survey this literature exhaustively, but my selective reading of it leads me to a rather different conclusion.

Consider Oreopoulos (2003), who measures the long run effects of growing up in “high density” public housing versus “low density” public housing in Toronto. Oreopoulos (2003) reports that the percent of households below the poverty line in neighborhoods with “high density” public housing was about 18 to 36 percentage points higher than in neighborhoods with “low density” housing (depending on the definition of “neighborhood” used; Table II, Row 3). Oreopoulos (2003) then estimates that male youths growing up in the smaller, low density, housing projects had average earnings 5 percent higher than their counterparts in high density projects, with a 95 percent confidence interval of (roughly) -5 percent to 15 percent (Table VI, Row 2, Columns 5 & 6). A 15 percent increase in earnings, for a 18 to 36 percentage point reduction in childhood exposure to neighborhood poverty, strikes me as a rather large, perhaps implausibly large “place effect”. Put differently Oreopoulos’ (2003) 95 percent confidence interval more or less spans the magnitude of effects that I would place positive probability density on *a priori*. While the Oreopoulos (2003) study is superbly designed and executed, its effect on my posterior views of the long-run earnings effects of exposure to impoverished neighborhoods, has been limited.

My read of Angrist and Lang (2004), who study the effects of classroom exposure to inner city youth on the academic achievement of suburban youth, is similar. When an inner city youth enters a suburban classroom she affects average measured achievement in the two ways. The first effect is purely compositional, the second effect arises from any spillovers onto other students. The ratio of the total-to-compositional effect is the *social multiplier*, a unitless summary measure of spillover strength (e.g., Glaeser, Sacerdote and Sacerdote, 2003; Graham, Imbens and Ridder, 2010). Using the coefficient estimates reported in Column 5 of Panel A of Tables 5 and 6 of Angrist and Lang (2004) I compute social multiplier point estimates for reading, mathematics and language achievement of 1.16, 1.26 and 1.59. While, as Angrist and Lang (2004) correctly note, many observational studies report even larger effects, these point estimates are nevertheless meaningfully large. The tables lack sufficient information to calculate confidence intervals for these social multipliers, but the available information suggests they would be quite large.

Oreopoulos (2003) and Angrist and Lang (2004) are not the only high quality studies to report null findings of neighborhood and/or peer effects, although they are two of the more widely cited of such studies. Indeed, they are both excellent studies. They also present results completely consistent with sizable neighborhood/peer effects (as well as results consistent with the absence of such effects). Unless one places substantial *a priori* probability on the event that neighborhood effects are extraordinarily large, the Oreopoulos (2003) and Angrist and Lang (2004) should cause only a modest update in your beliefs.

It is important to contextualize these observations. Several observational peer and neighborhood effects studies in the 1990s reported what were, at least in retrospect, implausibly large effects (e.g., Crane, 1991). In addition to well-understood problems of unmeasured “confounders” emphasized in the discussion above, many of these studies suffered from much deeper methodological problems (e.g., they attempted to fit what we would now call incomplete models; cf., Manski, 1993; Angrist, 2014). The discipline’s reaction to these weaknesses was to undertake studies with stronger research designs. Unfortunately many of these studies have been underpowered; failing to reject zero effects, but also sometimes quite large effects. Relatedly these studies have generally focused on very disadvantaged populations where few interventions of *any* kind have been shown to improve outcomes. Nevertheless the corrective has been useful; recent evidence suggest, for example, that the return to less exposure to concentrated poverty is probably not as high as, for example, earning a college degree.

I have learned a different set of lessons from Moving To Opportunity (MTO) demonstration project. It is undeniable that many scholars have found the apparent failure of MTO to find any effects of improved neighborhood quality on the economic outcomes of adult movers deeply disappointing (cf., Rothwell, 2015). This finding seems to cast doubt on, for example, the belief that spatial mismatch is an important driver of poor minority employment outcomes. However, Ludwig et al. (2012, Table 1) report that only about 6 percent of MTO participants listed employment goals as a primary or secondary reason for wanting to move at baseline. Perhaps a lack of large economic effects among the population targeted by MTO should not have been so surprising? In contrast, participants did list violence avoidance, housing quality and better schools for children, as main motivations for wanting to move. MTO was largely successful at meeting the first two of these self-reported goals.

Chetty, Hendren and Katz (2016), undertaking the longest term follow-up to date, with arguably the best data as well, find that MTO youth who moved early in life experience better labor market outcomes in early adulthood than their non-moving counterparts. This finding is consistent with their being a human capital penalty associated with sustained exposure to concentrated poverty. It will be interesting to track this cohort in the future.

Where does this leave us? Perhaps disappointingly there remains substantial scope to conduct studies whose primary aim is simply to test for the presence, and measure the magnitude of, neighborhood effects. There is not yet even a loose disciplinary consensus on the rough importance of neighborhoods on life outcomes. We are even further away from answering the deep question motivating many neighborhood effects researchers: what are the social costs and benefits of residential segregation? It is possible to motivate associational redistribution policies solely by appeals to equity (i.e., by placing intrinsic social value on diverse, inte-

grated, neighborhoods); however, given the controversy often associated with such policies, it is also important to understand their efficiency implications as well (cf., Piketty, 2000). The optimal assignment of individuals to groups in the presence of social spillovers, in addition to depending on the form of preferences, typically hinges on quite subtle features of the mapping from group composition into outcomes. This was emphasized in the context of a stylized model by Benabou (1996) and is an important theme of Graham, Imbens and Ridder (2010). Hopefully the next wave of work on neighborhood effects will combine the considerable progress that has been made on understanding the strengths and weakness of various research designs with a sharper focus on questions raised by the theory on sorting and segregation.

References

- [1] Acemoglu, Daron and Joshua Angrist. (2000). "How large are human capital externalities? Evidence from compulsory schooling laws," *NBER Macroeconomics Annual* 15: 9 - 59.
- [2] Altonji, Joseph G., Prashant Bharadwaj and Fabian Lange. (2012). "Changes in the characteristics of American youth: implications for adult outcomes," *Journal of Labor Economics* 30 (4): 783 - 828.
- [3] Altonji, Joseph G. and Richard Mansfield. (2014). "Group average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects," *Mimeo*.
- [4] Ananat, Elizabeth Oltmans. (2011). "The wrong side(s) of the tracks: the causal effects of racial segregation on urban poverty and inequality," *American Economic Journal: Applied Economics* 3 (2): 34-66.
- [5] Angrist, Joshua D. (2014). "The perils of peer effects," *Labour Economics* 30 (1): 94 - 108.
- [6] Angrist, Joshua D. and Kevin Lang. (2004). "Does school integration generate peer effects? Evidence from Boston's Metco program," *American Economic Review* 94 (5): 1613 - 1634.
- [7] Ammermueller, Andreas and Jörn-Steffen Pischke. (2009). "Peer effects in European primary schools: evidence from PIRLS," *Journal of Labor Economics* 27 (3): 315 - 348.

- [8] Bayer, Patrick, Fernando Ferreira and Robert McMillan. (2007). "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy* 115 (4): 588 - 638.
- [9] Bayer, Patrick, Randi Hjalmarsson and David Pozen (2009). "Building criminal capital behind bars: peer effects in juvenile corrections" *Quarterly Journal of Economics* 124 (1): 105 - 147.
- [10] Bayer, Patrick, Stephen Ross and Giorgio Topa. (2008). "Place of work, place of residence: informal hiring networks and labor market outcomes," *Journal of Political Economy* 116 (6): 1150 -1196.
- [11] Becker, Gary S. and Kevin M. Murphy. (2000). "Segregation and Integration in Neighborhoods," *Social Economics*. Cambridge, MA: Harvard University Press.
- [12] Benabou, Roland. (1993). "Workings of a city: location, education, and production," *Quarterly Journal of Economics* 108 (3): 619 - 652.
- [13] Benabou, Roland. (1996). "Equity and efficiency in human capital investment: the local connection," *Review of Economic Studies* 63 (2): 237 - 264.
- [14] Bewley, Truman F. (1981). "A critique of Tiebout's theory of local public expenditures," *Econometrica* 49 (3): 713 - 740.
- [15] Blume, Lawrence E., William A. Brock, Steven N. Durlauf and Rajshri Jayaraman. (2015). "Linear social interactions models," *Journal of Political Economy* 123 (2): 444 - 496.
- [16] Boucher, Vincent, Yann Bramoullé, Habiba Djebbari and Bernard Fortin. (2014). "Do peers affect student achievement? Evidence from Canada using group size variation," *Journal of Applied Econometrics* 29 (1): 91 - 109.
- [17] Boyton-Jarrett, Renee et. al. (2008). "Cumulative violence exposure and self-rated health: longitudinal study of adolescents in the United States," *Pediatrics* 122 (5): 961 - 70.
- [18] Bramoullé, Yann, Habiba Djebbari and Bernard Fortin. (2009). "Identification of peer effects through social networks," *Journal of Econometrics* 150 (1): 41 - 55.
- [19] Brock, William A. and Steven N. Durlauf. (2001a). "Interactions-based models," *Handbook of Economics* 5: 3297 - 3380 (J.J. Heckman & E. Leamer, Eds.). Amsterdam: North-Holland.

- [20] Brock, William A. and Steven N. Durlauf. (2001b) “Discrete choice with social interactions,” *Review of Economic Studies* 68 (2): 235 - 260.
- [21] Brock, William A. and Steven N. Durlauf. (2007). “Identification of binary choice models with social interactions,” *Journal of Econometrics* 140 (1): 52 - 75.
- [22] Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Kato Klebanov and Naomi Sealand. (1993). “Do neighborhoods influence child and adolescent development,” *American Journal of Sociology* 99 (2): 353 - 395.
- [23] Calabrese, Stephen, Dennis Epple, Thomas Romer, Holger Sieg. (2006). “Local public good provision: voting, peer effects, and mobility,” *Journal of Public Economics* 90 (6–7): 959 - 981.
- [24] Calvo-Armengol, Antoni, Eleonora Patacchini and Yves Zenou. (2009). “Peer effects and social networks in education,” *Review of Economic Studies* 76 (4): 1239 - 1267.
- [25] Card, David and Laura Giuliano. (2013). “Peer effects and multiple equilibria in the risky behavior of friends,” *Review of Economics and Statistics* 95 (4): 1130 - 1149.
- [26] Card, David and Jesse Rothstein. (2007). “Racial segregation and the black-white test score gap,” *Journal of Public Economics* 91 (11-12): 2158 - 2184.
- [27] Carrell, Scott E., Richard L. Fullerton, James E. West. (2009). “Does your cohort matter? Measuring peer effects in college achievement,” *Journal of Labor Economics* 27 (3): 439 - 464.
- [28] Chetty, Raj, Nathaniel Hendren and Lawrence Katz. (2016). “The effects of exposure to better neighborhoods on children: new evidence from the Moving to Opportunity experiment,” *American Economics Review* 104 (4): 855 - 902.
- [29] Christopher, A. J. (1994). “Segregation levels in the late-apartheid city 1985–1991,” *Tijdschrift voor economische en sociale geografie* 85 (1): 15 - 24.
- [30] Ciccone, Antonio and Giovanni Peri. (2006). “Identifying human-capital externalities: theory with applications,” *Review of Economic Studies* 73 (2): 381 - 412.
- [31] Coleman, James S. et al. (1966). *Equality of Educational Opportunity*. Washington D.C.: U.S. Department of Health, Education and Welfare.
- [32] Cutler, David M. and Edward L. Glaeser. (1997). “Are ghettos good or bad?” *Quarterly Journal of Economics* 112 (3): 827 - 872.

- [33] Cutler, David M., Edward L. Glaeser and Jacob L. Vigdor. (1999). "The rise and decline of the American ghetto," *Journal of Political Economy* 107 (3): 455 - 506.
- [34] Crane, Jonathan. (1991). "The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing," *American Journal of Sociology* 96 (5): 1226 - 1259.
- [35] Datcher, Linda. (1982). "Effects of community and family background on achievement," *Review of Economics and Statistics* 64 (1):32 - 41.
- [36] Dawkins, Casey J. (2004). "Recent evidence on the continuing cause of Black-White residential segregation," *Journal of Urban Affairs* 26 (3): 379 - 400.
- [37] de Bartolome, Charles A. M. (1990). "Equilibrium and inefficiency in a community model with peer group effects," *Journal of Political Economy* 98 (1): 110 - 133.
- [38] de Paula, Aureo. (2013). "Econometric analysis of games with multiple equilibria," *Annual Review of Economics* 5: 107 - 131.
- [39] Duncan, Greg J. and Stephen Raudenbusch. (2001). "Neighborhoods and adolescent development: how can we determine the links," *Does it Take a Village? Community Effects on Children, Adolescents, and Families*: 1005 - 1036 (A. Booth & A.C. Crouter, Eds.). State College, PA: Pennsylvania State University Press.
- [40] Durlauf, Steven N. (1996a). "A theory of persistent income inequality," *Journal of Economic Growth* 1 (1): 75 - 93.
- [41] Durlauf, Steven N. (1996b). "Associational redistribution: a defense," *Politics & Society* 24 (4): 391 - 410.
- [42] Durlauf, Steven N. (2004). "Neighborhood Effects," *Handbook of Regional and Urban Economics* 4: 2173 - 2242 (J.V. Henderson & J-F. Thisse, Eds.). Amsterdam: North-Holland.
- [43] Durlauf, Steven N. (2006). "Groups, social influences, and inequality," *Poverty Traps*: 141 - 175 (S. Bowels, S. N. Durlauf, & K. Hoff, Eds.). Princeton, NJ: Princeton University Press.
- [44] Durlauf, Steven N. and Yannis Ioannides. (2010). "Social Interactions," *Annual Review of Economics* 2: 451 - 478.

- [45] Epple, Dennis and Glenn J. Platt. (1998). "Equilibrium and local redistribution in an urban economy when households differ by preferences and incomes," *Journal of Urban Economics* 43 (1): 23-51.
- [46] Epple, Dennis and Richard E. Romano (2011). "Peer effects in education: a survey of theory and evidence," *Handbook of Social Economics* 1B: 1053 - 1164 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.
- [47] Evans, William N., Wallace E. Oates and Robert M. Schwab. (1992). "Measuring peer group effects: a study of teenage behavior," *Journal of Political Economy* 100 (5): 966 - 991.
- [48] Farley, Reynolds. (1977). "Residential segregation in urbanized areas of the United States in 1970: an analysis of social class and racial differences," *Demography* 14 (4): 497 - 518.
- [49] Fernandez, Raquel. (2003). "Sorting, education and inequality," *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* 2: 1 - 40 (M. Dewatripont & L.P. Hansen, Eds.). Cambridge: Cambridge University Press.
- [50] Finkelhor, Daniel et al. (2009). "Violence, abuse, and crime exposure in a national sample of children and youth," *Pediatrics* 124 (5): 1411 -1423.
- [51] Glaeser, Edward L., Bruce I. Sacerdote, and Jose A. Scheinkman. (2003). "The social multiplier," *Journal of the European Economic Association* 1 (2-3): 345 - 353.
- [52] Goldberger, Arthur. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
- [53] Goldsmith-Pinkham, Paul and Guido W. Imbens. (2013). "Social networks and the identification of peer effects," *Journal of Business and Economic Statistics* 31 (3): 253 - 264.
- [54] Gould, Eric D., Victor Lavy, M. Daniele Paserman. (2004). "Immigrating to opportunity: estimating the effect of school quality using a natural experiment on Ethiopians in Israel," *Quarterly Journal of Economics* 119 (2): 489 - 526.
- [55] Gould, Eric D., Victor Lavy, M. Daniele Paserman. (2009). "Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence," *Economic Journal* 119 (540): 1243-1269

- [56] Gould, Eric D., Victor Lavy, M. Daniele Paserman. (2011). Sixty years after the magic carpet ride: the long-run effect of the early childhood environment on social and economic outcomes," *Review of Economic Studies* 78 (3): 938 - 973.
- [57] Graham, Bryan S. (2008). "Identifying social interactions through conditional variance restrictions," *Econometrica* 76 (3): 643 - 660.
- [58] Graham, Bryan S. (2011). "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers," *Handbook of Social Economics* 1B: 965 - 1052 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.
- [59] Graham, Bryan S. (2015a) "Methods of identification in social networks," *Annual Review of Economics* 7: 465 - 485.
- [60] Graham, Bryan S. (2015b). "An econometric model of link formation with degree heterogeneity," *Cemmap Working Paper CWP43/15*.
- [61] Graham, Bryan S. (2015c). "Lecture 8: Using network structure to identify peer spillovers," *Empirical Methods for Program Evaluation and Networks*, CEMFI, August 24 - 28, 2015.
- [62] Graham, Bryan S. (2015d). "Lecture 2: Covariate adjustment with time-varying policy variables," *Empirical Methods for Program Evaluation and Networks*, CEMFI, August 24 - 28, 2015.
- [63] Graham, Bryan S. and Jingyong Hahn. (2005). "Identification and estimation of the linear-in-means model of social interactions," *Economics Letters* 88 (1): 1 - 6.
- [64] Graham, Bryan S., Guido W. Imbens and Geert Ridder. (2010). "Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach," *NBER Working Paper No. 16499*.
- [65] Graham, Bryan S. and Patrick Sharkey. (2013). *Mobility and the Metropolis: How Communities Factor into Economic Mobility, A Report from the Pew Charitable Trusts*. Washington D.C.: The Pew Charitable Trusts.
- [66] Griliches, Zvi and Jacques Mairesse. (1998). "Production functions: the search for identification," *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Memorial Symposium*: 169 - 203 (S. Strom, Ed.). Cambridge: Cambridge University Press.

- [67] Harding, David et al. (2011). “Unpacking neighborhood influences on education outcomes: setting the stage for future research,” *Whither Opportunity? Rising Inequality, Schools, and Children’s Life Chances*: 277 - 298 (G. Duncan & R. Murnane, Eds.). New York: Russell Sage Foundation.
- [68] Henderson, Vernon, Peter Mieszkowski and Yvon Sauvageau. (1978). “Peer group effects and educational production functions,” *Journal of Public Economics* 10 (1): 97 – 106.
- [69] Hoxby, Caroline M. (2000). “Peer effects in the classroom: learning from gender and race variation,” *NBER Working Paper No. 7867*.
- [70] Ioannides, Yannis M. (2011). “Neighborhood Effects And Housing,” *Handbook of Social Economics* 1B: 1281 - 1340 (J. Benhabib, A. Bisin, & M. Jackson, Eds.). Amsterdam: North-Holland.
- [71] Iceland, John, Daniel H. Weinberg, and Erika Steinmetz. (2002). *Racial and Ethnic Residential Segregation in the United States: 1980 - 2000*. Washington D.C.: U.S. Census Bureau.
- [72] Jakubs, John F. (1977). “Residential segregation the Taeuber index reconsidered,” *Journal of Regional Science* 17 (2): 281 - 283.
- [73] Kim, David A. et al. (2015). “Social network targeting to maximise population behaviour change: a cluster randomised controlled trial,” *Lancet* 386 (9989): 145 – 153.
- [74] Kline, Patrick and Enrico Moretti. (2014). “People, places, and public policy: some simple welfare economics of local economic development policies,” *Annual Review of Economics* 6: 629 - 662.
- [75] Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz. (2007). “Experimental analysis of neighborhood effects,” *Econometrica* 75 (1): 83 - 119.
- [76] Kremer, Michael and Erin Maskin. (1996). “Wage inequality and segregation by skill,” *NBER Working Paper No. 5718*.
- [77] Leventhal, Tama and Jeanne Brooks-Gunn. (2000). “The neighborhoods they live in: the effects of neighborhood of residence on child and adolescent outcomes,” *Psychological Bulletin* 126 (2): 309 - 337.
- [78] Liu, Xiaodong, Eleonora Patacchini and Yves Zenou. (2014). “Endogenous peer effects: local aggregate or local average?” *Journal of Economic Behavior & Organization* 103: 39 – 59.

- [79] Logan, John R. and Richard N. Turner. (2013). *Hispanics in the United States: Not Only Mexicans*. Providence: US2010 Project, Brown University.
- [80] Loury, Glenn C. (1977). "A dynamic theory of racial income differences," *Women, Minorities and Employment Discrimination*: 153 - 186 (P.A. Wallace & A LeMond, Eds.). Lexington, MA: Lexington Books.
- [81] Loury, Glenn C. (2002). *The Anatomy of Racial Inequality*. Cambridge, MA: Harvard University Press.
- [82] Ludwig, Jens (2012). "Guest Editor's Introduction," *Cityscape: A Journal of Policy Development and Research* 14 (2): 1 - 28.
- [83] Ludwig, Jens, et al. (2008). "What can we learn about neighborhood effects from the Moving to Opportunity Experiment?" *American Journal of Sociology* 14 (1): 144 - 188.
- [84] Ludwig, Jens et al. (2012). "Neighborhood effects on the long-term well-being of low-income adults," *Science* 337 (6101): 1505 - 1510.
- [85] Manski, Charles F. (1993). "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies* 60 (3): 531 - 542.
- [86] Massey, Douglas S. (2004). "Segregation and stratification," *Du Bois Review* 1 (1): 7 - 25.
- [87] Massey, Douglas S. and Nancy A. Denton. (1988). "The dimensions of segregation," *Social Forces* 67 (2): 281 - 315.
- [88] Massey, Douglas S. and Nancy A. Denton. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.
- [89] Mayer, Susan E. and Christopher Jencks. (1989). "Growing up in poor neighborhoods: how much does it matter?" *Science* 243 (4897): 1441 - 1445.
- [90] Mundlak, Yair. (1978). "On the pooling of time series and cross section data," *Econometrica* 46 (1): 69 - 85.
- [91] Nesheim, Lars. (2002). "Equilibrium sorting of heterogenous consumers across locations: theory and empirical implications," *Cemmap Working Paper CWP08/02*.
- [92] Olley, G. Steven and Ariel Pakes. (1996). "The dynamics of productivity in the telecommunications equipment industry," *Econometrica* 64 (6): 1263 - 1297.

- [93] Oreopoulos, Phil. The Long-Run Consequences of Living in a Poor Neighborhood *The Quarterly Journal of Economics* 118.4 (2003): 1533-575
- [94] Pew Charitable Trusts. (2011). "Poll results: Economic mobility and the American Dream," accessed on 4 February 2014 at <http://www.pewstates.org/research/data-visualizations>.
- [95] Piketty, Thomas. (2000). "Theories of persistent inequality and intergenerational mobility," *Handbook of Income Distribution* 1: 429 - 476 (A. B. Atkinson & F. Bourguignon, Eds.). Amsterdam: North-Holland.
- [96] Reardon, Sean F. and Kendra Bischoff. (2011). "Income inequality and income segregation," *American Journal of Sociology* 116 (4): 1092 - 1153.
- [97] Robins, James. (1986). "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical Modelling* 7 (9–12): 1393 - 1512.
- [98] Robins, James M. (1997). "Causal inference from complex longitudinal data," *Latent Variable Modeling and its Applications to Causality*: 69 -117 (M. Berkane, Ed.). New York: Springer-Verlag.
- [99] Robins, James M. (1999). "Association, causation, and marginal structural models," *Synthese* 121 (1-2): 151 - 179.
- [100] Robins, James M. (2000). "Marginal structural models versus structural nested models as tools for causal inference," *Statistical Models in Epidemiology, the Environment, and Clinical Trials: The IMA Volumes in Mathematics and its Applications* 116: 95 - 133 (M. E. Halloran & D. Berry, Eds.). New York: Springer.
- [101] Ross, Stephen L. and John Yinger. (2002). *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. Cambridge, MA: The MIT Press.
- [102] Rothwell, Jonathan. (2015). "Sociology's revenge: Moving to Opportunity (MTO) revisited," *Brookings Social Mobility Memo*, Accessed on 28 January 2016 at <http://www.brookings.edu/blogs/social-mobility-memos/posts/2015/05/06-moving-to-opportunity-revisited-rothwell>
- [103] Sampson, Robert J. (2012a). "Moving and the neighborhood glass ceiling," *Science* 337 (6101): 1464 - 1465.

- [104] Sampson, Robert J. (2012b). *Great American City*. Chicago: University of Chicago Press.
- [105] Sampson, Robert J., Jeffrey D. Morenoff and Thomas Gannon-Rowley. (2002). "Assessing 'neighborhood effects': social processes and new directions for research," *Annual Review of Sociology* 28 (10): 443 - 478.
- [106] Sampson, Robert J., Patrick Sharkey and Stephen W. Raudenbush. (2008). "Durable effects of concentrated disadvantage on verbal ability among African-American children," *Proceedings of the National Academy of Sciences* 105 (3): 845 - 852
- [107] Sampson, Robert J, and Stephen W Raudenbush. (1999). "Systematic social observation of public spaces: a new look at disorder in urban neighborhoods," *American Journal of Sociology* 105 (3): 603-651
- [108] Sastry, Narayan and Anne R. Pebley. (2010). "Family and neighborhood sources of socioeconomic inequality in children's achievement," *Demography* 47 (3): 777 - 800.
- [109] Schill, Michael H. and Susan M. Wachter. (1995). "Housing market constraints and spatial stratification by income and race," *Housing Policy Debate* 6 (1): 141 - 167.
- [110] Sharkey, Patrick. (2010). "The acute effect of local homicides on children's cognitive performance," *Proceedings of the National Academy of Sciences* 107 (26): 11733 - 11738.
- [111] Sharkey, Patrick. (2012). *Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality*. Chicago: University of Chicago Press.
- [112] Sharkey, Patrick and Felix Elwert. (2011). "The legacy of disadvantage: multigenerational neighborhood effects on cognitive ability," *American Journal of Sociology* 116 (6): 1934 - 81.
- [113] Sharkey, Patrick and Jacob W. Faber. (2014). "Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects," *Annual Review of Sociology* 40: 559 - 579.
- [114] Sieg, Holger and Thomas J. Holmes. (2015). "Structural estimation in urban economics," *Handbook of Regional and Urban Economics* 69 - 114 (G. Duranton, J.V. Henderson and W.C. Strange, Eds.). Amsterdam: North-Holland.
- [115] Tatian, Peter A. (2003). *Neighborhood Change Database (NCDB) 1970 - 2000 Tract Data: Data User's Guide*. Washington, D.C.: The Urban Institute.

- [116] Tiebout, Charles M. (1956). “A pure theory of local expenditures,” *Journal of Political Economy* 64 (5): 416 – 424.
- [117] Watson, Tara. (2009). “Inequality and the measurement of residential segregation by income,” *Review of Income and Wealth* 55 (3): 820 - 844.
- [118] Wodtke, Geoffrey T. (2013). “Duration and timing of exposure to neighborhood poverty and the risk of adolescent parenthood,” *Demography* 50 (5): 1765 - 1788.
- [119] Wodtke, Geoffrey T., David J. Harding and Felix Elwert. (2011). “ Neighborhood effects in temporal perspective: the impact of long-term exposure to concentrated disadvantage on high school graduation,” *American Sociological Review* (76 (5): 713 - 736.
- [120] Wooldridge, Jeffrey M. (1997). “Distribution-free estimation of some nonlinear panel data models,” *Journal of Econometrics* 90 (1): 77 - 97.
- [121] Wooldridge, Jeffrey M. (2005). “Unobserved heterogeneity and estimation of average partial effects,” *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*: 27 - 55 (D.W.K. Andrews & J.H. Stock, Eds.). Cambridge: Cambridge University Press.
- [122] Yinger, John. (1999). “Sustaining the Fair Housing Act,” *Cityscape* 4 (3): 93 - 106.

A Properties of the dissimilarity index (DI)

To get additional insight into the interpretation of the dissimilarity index (DI) rewrite (1) as

$$DI = \frac{1}{2} \mathbb{E} \left[\left| \frac{s(Z) - Q}{Q(1 - Q)} \right| \right]. \quad (39)$$

The numerator inside the expectation in (39) gives the proportion of a neighborhood’s residents who would have to move in order to align its demographic structure with the city-wide one. If $s(Z) > Q$, then movers are Minority residents who will be replaced with arriving White residents (with the opposite occurring when $s(Z) < Q$). Following this logic,

$$\frac{\mathbb{E} [|s(Z) - Q| \cdot \mathbf{1}(s(Z) - Q \geq 0)]}{Q} = \frac{\mathbb{E} [|s(Z) - Q| \mid s(Z) \geq Q]}{Q} \Pr(p(Z) \geq Q) \quad (40)$$

equals the city-wide proportion of minority residents who must move in order to achieve perfect integration.²⁵ Likewise $\mathbb{E} \left[\left| \frac{s(Z)-Q}{1-Q} \right| \middle| s(Z) < Q \right] \Pr(s(Z) < Q)$ gives the city-wide proportion of White residents who must move. Since the aggregate number of Minority and White moves must coincide we have the equality

$$Q \mathbb{E} \left[\left| \frac{s(Z)-Q}{Q} \right| \middle| s(Z) \geq Q \right] \Pr(s(Z) \geq Q) = (1-Q) \mathbb{E} \left[\left| \frac{s(Z)-Q}{1-Q} \right| \middle| s(Z) < Q \right] \times \Pr(s(Z) < Q)$$

Iterated expectations applied to (39) then gives, after substituting in the above equality and re-arranging,

$$\text{DI} = \frac{Q^{-1} \mathbb{E} [|s(Z) - Q| \cdot \mathbf{1}(s(Z) - Q \geq 0)]}{1 - Q}. \quad (41)$$

Equation (41) demonstrates that the Dissimilarity Index equals the city-wide proportion of minority residents who would need to move in order to achieve perfect integration, *relative* to the proportion that would need to move under a status quo of perfect segregation (cf., Jakubs, 1977).²⁶

²⁵Let M_i denote neighborhood size and n city size. Let n_1 and n_0 respectively equal the Minority and White population of the city such that $Q = n_1/n$. We can re-write (40) as

$$Q^{-1} \mathbb{E} [|s(Z) - Q| \cdot \mathbf{1}(s(Z_i) - Q \geq 0)] = \sum_{i=1}^N \frac{M_i}{n_1} |s(Z_i) - Q| \cdot \mathbf{1}(s(Z_i) - Q \geq 0),$$

from which the claimed interpretation of (40) is easy to verify.

²⁶The denominator in (41), $1 - Q$, equals the proportion of Minority residents that would need to move under a status quo of complete residential separation. In that case $1 - Q$ fraction of residents would need to leave each Minority neighborhood in order to integrate all neighborhoods.

B Derivations

To derive (21) and (22) of the main text begin by observing that by the definition of the linear predictor

$$\begin{aligned}
\begin{pmatrix} b_0 \\ c_0 \end{pmatrix} &= \mathbb{V} \begin{pmatrix} T \\ s(Z) \end{pmatrix}^{-1} \mathbb{C} \left(\begin{pmatrix} T \\ s(Z) \end{pmatrix}, Y \right) \\
&= \begin{pmatrix} p(1-p) & \eta^2 p(1-p) \\ \eta^2 p(1-p) & \eta^2 p(1-p) \end{pmatrix}^{-1} \mathbb{C} \left(\begin{pmatrix} T \\ s(Z) \end{pmatrix}, Y \right) \\
&= \frac{1}{p(1-p)\eta^2(1-\eta^2)} \begin{pmatrix} \eta^2 & -\eta^2 \\ -\eta^2 & 1 \end{pmatrix} \mathbb{C} \left(\begin{pmatrix} T \\ s(Z) \end{pmatrix}, Y \right),
\end{aligned}$$

which after evaluating both rows yields

$$b_0 = \frac{1}{p(1-p)(1-\eta^2)} \{ \mathbb{C}(T, Y) - \mathbb{C}(s(Z), Y) \} \quad (42)$$

$$c_0 = \frac{1}{p(1-p)\eta^2(1-\eta^2)} \{ \mathbb{C}(s(Z), Y) - \eta^2 \mathbb{C}(T, Y) \}. \quad (43)$$

To evaluate the terms in $\{\cdot\}$ in (42) and (43) first compute the covariances:

$$\begin{aligned}
\mathbb{C}(T, A) &= \mathbb{E}[(T-p)A] \\
&= \mathbb{E}[TA] - p\mathbb{E}[A] \\
&= p\mathbb{E}[A|T=1] - p(p\mathbb{E}[A|T=1] + (1-p)\mathbb{E}[A|T=0]) \\
&= p(1-p)(\mathbb{E}[A|T=1] - \mathbb{E}[A|T=0]) \\
&= p(1-p)\phi_0,
\end{aligned} \quad (44)$$

and

$$\begin{aligned}
\mathbb{C}(T, U) &= \mathbb{E}[(T-p)U] \\
&= p(1-p)(\mathbb{E}[U|T=1] - \mathbb{E}[U|T=0]),
\end{aligned} \quad (45)$$

and

$$\begin{aligned}
\mathbb{C}(T, m_A(Z)) &= \mathbb{E}[(T-p)m_A(Z)] \\
&= \mathbb{E}[(T-p)\{\pi_0 + \phi_0 s(Z) + \mathbb{E}[B|Z]\}] \\
&= \eta^2 p(1-p)\phi_0 + \mathbb{E}[(s(Z)-p)\mathbb{E}[B|Z]],
\end{aligned} \quad (46)$$

and

$$\begin{aligned}
\mathbb{E}[(s(Z) - p)A] &= \mathbb{E}[(s(Z) - p)(\pi_0 + \phi_0 T + B)] \\
&= \eta^2 p(1 - p)\phi_0 + \mathbb{E}[(s(Z) - p)B] \\
&= \eta^2 p(1 - p)\phi_0 + \mathbb{E}[(s(Z) - p)\mathbb{E}[B|Z]] \tag{47}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[(s(Z) - p)m_A(Z)] &= \mathbb{E}[(s(Z) - p)(\pi_0 + \phi_0 s(Z) + \mathbb{E}[B|Z])] \\
&= \eta^2 p(1 - p)\phi_0 + \mathbb{E}[(s(Z) - p)(\mathbb{E}[B|Z])] \tag{48}
\end{aligned}$$

$$\mathbb{E}[(s(Z) - p)U] = p(1 - p)(\mathbb{E}[U|T = 1] - \mathbb{E}[U|T = 0]). \tag{49}$$

Using (44), (45), (46), (47), (48) and (49) I evaluate

$$\begin{aligned}
\mathbb{C}(T, Y) &= \mathbb{C}(T, \beta_0 T + \gamma_0 s(Z) + m_A(Z))' \delta_0 + U' \kappa_0 + A' \lambda_0 \\
&= \beta_0 p(1 - p) + \gamma_0 \eta^2 p(1 - p) \\
&\quad + \mathbb{C}(T, m_A(Z))' \delta_0 + \mathbb{C}(T, U)' \kappa_0 + \mathbb{C}(T, A)' \lambda_0 \\
&= (\beta_0 + \phi'_0 \lambda_0) p(1 - p) + \gamma_0 \eta^2 p(1 - p) \\
&\quad + \mathbb{C}(T, m_A(Z))' \delta_0 + \mathbb{C}(T, U)' \kappa_0
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}(s(Z), Y) &= \mathbb{C}(s(Z), \beta_0 T + \gamma_0 s(Z) + m_A(Z))' \delta_0 + U' \kappa_0 + A' \lambda_0 \\
&= (\beta_0 + \gamma_0) \eta^2 p(1 - p) \\
&\quad + \mathbb{C}(s(Z), m_A(Z))' \delta_0 + \mathbb{C}(s(Z), U)' \kappa_0 + \mathbb{C}(s(Z), A)' \lambda_0,
\end{aligned}$$

which yields, after subtracting,

$$\begin{aligned}
\mathbb{C}(T, Y) - \mathbb{C}(s(Z), Y) &= (1 - \eta^2) p(1 - p) \beta_0 + \phi'_0 \lambda_0 p(1 - p) - \mathbb{C}(s(Z), A)' \lambda_0 \\
&= (1 - \eta^2) p(1 - p) \beta_0 + \phi'_0 \lambda_0 p(1 - p) \\
&\quad - \eta^2 p(1 - p) \phi'_0 \lambda_0 + \mathbb{E}[(s(Z) - p)\mathbb{E}[B|Z]]' \lambda_0 \\
&= (1 - \eta^2) p(1 - p) \{\beta_0 + \phi'_0 \lambda_0\} - \mathbb{E}[(s(Z) - p)\mathbb{E}[B|Z]]' \lambda_0.
\end{aligned}$$

Substituting this expression in (42) above then gives (21) of the main text:

$$b_0 = \beta_0 + \phi'_0 \lambda_0 - \frac{1}{1 - \eta^2} \frac{1}{p(1 - p)} \mathbb{E}[(s(Z) - p)\mathbb{E}[B|Z]]' \lambda_0.$$

To derive (22) I evaluate the difference:

$$\begin{aligned}
& \mathbb{C}(s(Z), Y) - \eta^2 \mathbb{C}(T, Y) \\
&= (\beta_0 + \gamma_0) \eta^2 p(1-p) + \mathbb{C}(s(Z), m_A(Z))' \delta_0 + \mathbb{C}(s(Z), U)' \kappa_0 + \mathbb{C}(s(Z), A)' \lambda_0 \\
&\quad - \eta^2 \{(\beta_0 + \phi'_0 \lambda_0) p(1-p) + \gamma_0 \eta^2 p(1-p) + \mathbb{C}(T, m_A(Z))' \delta_0 + \mathbb{C}(T, U)' \kappa_0\} \\
&= \gamma_0 \eta^2 (1 - \eta^2) p(1-p) + (1 - \eta^2) \mathbb{C}(s(Z), m_A(Z))' \delta_0 + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0 \\
&\quad + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&= \gamma_0 \eta^2 (1 - \eta^2) p(1-p) + (1 - \eta^2) \mathbb{C}(s(Z), m_A(Z))' \delta_0 + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0 \\
&\quad - \eta^2 \phi'_0 \lambda_0 p(1-p) + \eta^2 p(1-p) \phi'_0 \lambda_0 + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&= \gamma_0 \eta^2 (1 - \eta^2) p(1-p) + (1 - \eta^2) \mathbb{C}(s(Z), m_A(Z))' \delta_0 + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0 \\
&\quad + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&= \gamma_0 \eta^2 (1 - \eta^2) p(1-p) + (1 - \eta^2) \{ \eta^2 p(1-p) \phi_0 + \mathbb{E}[(s(Z) - p) (\mathbb{E}[B|Z])] \}' \delta_0 \\
&\quad + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0 + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&= (\gamma_0 + \phi'_0 \delta_0) \eta^2 (1 - \eta^2) p(1-p) \\
&\quad + (1 - \eta^2) \mathbb{E}[(s(Z) - p) (\mathbb{E}[B|Z])] \delta_0 \\
&\quad + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&\quad + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0,
\end{aligned}$$

which, after substituting into (43), gives (22) of the main text:

$$\begin{aligned}
c_0 &= \gamma_0 + \phi'_0 \delta_0 \\
&\quad + \frac{1}{p(1-p) \eta^2 (1 - \eta^2)} \{ (1 - \eta^2) \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \delta_0 \\
&\quad + \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 + (1 - \eta^2) \mathbb{C}(s(Z), U)' \kappa_0 \} \\
&= \gamma_0 + \phi'_0 \delta_0 + \frac{1}{\eta^2} \left\{ \frac{1}{p(1-p)} \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \delta_0 \right. \\
&\quad + \frac{1}{1 - \eta^2} \frac{1}{p(1-p)} \mathbb{E}[(s(Z) - p) \mathbb{E}[B|Z]]' \lambda_0 \\
&\quad \left. + (\mathbb{E}[U|T=1] - \mathbb{E}[U|T=0])' \kappa_0 \right\}.
\end{aligned}$$

Equation (35) in the main text is easily calculated using the following terms:

$$\begin{aligned}
\mathbb{E}[s(Z_{ci})|T_{ci};c] &\propto \eta_c^2 T_{ci} \\
\mathbb{E}[m_A(Z_{ci})|T_{ci};c] &\propto \left\{ \eta_c^2 \phi_c + \frac{\mathbb{C}(s(Z_{ci}), \mathbb{E}[B_{ci}|Z_{ci}];c)}{p_c(1-p_c)} \right\} T_{ci} \\
&= \{ \eta_c^2 \phi_c + v_c \} T_{ci} \\
\mathbb{E}[U_{ci}|T_{ci};c] &\propto (\mathbb{E}[U_{ci}|T_{ci}=1,c] - \mathbb{E}[U_{ci}|T_{ci}=0,c]) T_{ci} \\
&= \tau_c T_{ci} \\
\mathbb{E}[A_{ci}|T_{ci};c] &\propto (\mathbb{E}[A_{ci}|T_{ci}=1,c] - \mathbb{E}[A_{ci}|T_{ci}=0,c]) T_{ci} \\
&= \phi_c T_{ci}.
\end{aligned}$$

To derive the probability limit of the instrumental variables estimator (31) I first compute

$$\begin{aligned}
\begin{pmatrix} b_{IV} \\ c_{IV} \end{pmatrix} &= \mathbb{C} \left(\begin{pmatrix} T \\ W \end{pmatrix} \begin{pmatrix} T \\ s(Z) \end{pmatrix}' \right)^{-1} \mathbb{C} \left(\begin{pmatrix} T \\ W \end{pmatrix}, Y \right) \\
&= \begin{pmatrix} p(1-p) & \eta^2 p(1-p) \\ \mathbb{C}(T, W) & \mathbb{C}(s(Z), W) \end{pmatrix}^{-1} \mathbb{C} \left(\begin{pmatrix} T \\ W \end{pmatrix}, Y \right) \\
&= \frac{1}{p(1-p)p_W(1-p_W)\Xi_S - \eta^2 p(1-p)p(\mathbb{E}[W|T=1] - p_W)} \\
&\quad \times \begin{pmatrix} p_W(1-p_W)\Xi_S & -\eta^2 p(1-p) \\ -p(\mathbb{E}[W|T=1] - p_W) & p(1-p) \end{pmatrix} \mathbb{C} \left(\begin{pmatrix} T \\ W \end{pmatrix}, Y \right) \\
&= \frac{1}{p_W(1-p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T=1] - p_W)} \begin{pmatrix} \frac{p_W(1-p_W)\Xi_S}{p(1-p)} & -\eta^2 \\ -\frac{\mathbb{E}[W|T=1] - p_W}{1-p} & 1 \end{pmatrix} \mathbb{C} \left(\begin{pmatrix} T \\ W \end{pmatrix}, Y \right)
\end{aligned}$$

with $p_W = \Pr(W=1)$ and Ξ_T and Ξ_S as defined in the main text.

Evaluating the second row of the expression above gives

$$c_{IV} = \frac{1}{p_W(1-p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T=1] - p_W)} \left\{ \mathbb{C}(W, Y) - \mathbb{C}(T, Y) \frac{\mathbb{E}[W|T=1] - p_W}{1-p} \right\}.$$

Using the fact that $\mathbb{C}(W, Y) = p_W(1-p_W)(\mathbb{E}[Y|W=1] - \mathbb{E}[Y|W=0])$ and the expres-

sion for $\mathbb{C}(T, Y)$ derived above I get

$$\begin{aligned}
c_{IV} &= (\gamma_0 + \phi'_0 \delta_0) \\
&+ \frac{1}{p_W(1-p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T=1] - p_W)} \\
&\times \left\{ p_W(1-p_W)\Xi_{\bar{B}} - \mathbb{C}(s(Z), \mathbb{E}[B|Z]) \frac{\mathbb{E}[W|T=1] - p_W}{1-p} \right\}' \delta_0 \\
&+ \frac{1}{p_W(1-p_W)\Xi_S - \eta^2 p(\mathbb{E}[W|T=1] - p_W)} \\
&\times \left\{ p_W(1-p_W)\Xi_U - \mathbb{C}(s(Z), U) \frac{\mathbb{E}[W|T=1] - p_W}{1-p} \right\}' \kappa_0
\end{aligned}$$

from which (31) and (32) follow after some manipulation.