Lecture 6: Neighborhood Effects: Cross-City Research Designs

Bryan S. Graham, UC - Berkeley & NBER

September 4, 2015

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 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 25. 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. 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

$$\mathbb{E}\left[Y\right|T, s\left(Z\right), m_A\left(Z\right), U, A] = \alpha_0 + \beta_0 T + \gamma_0 s\left(Z\right) + m_A\left(Z\right)' \delta_0 + U'\kappa_0 + A'\lambda_0. \quad (1)$$

Equation (1) 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 if the outcome varies with changes in the neighborhood distribution of household types and/or backgrounds (i.e., $\gamma_0 \neq 0$ and/or $\delta_0 \neq 0$).

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 on the outcome.

I assume this equation is structural in the sense of Goldberger (1991). Since we are free to define A and U broadly, the assumption that (1) 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 (1) 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 (1996), associational redistributions. Equation (1) is helpful for understanding how "who lives with whom" influences the distribution of Y.

Let (1) 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:

$$\mathbb{E}\left[Y|\,T, s\left(Z\right)\right] = a_0 + b_0 T + c_0 s\left(Z\right). \tag{2}$$

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 (2) 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 (1). The linear form of (1) 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

$$A = \pi_0 + \phi_0 T + B,\ \mathbb{E}\left[B|\,T\right] = 0. \tag{3}$$

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 (3) is nothing more than a decomposition. Graham, Imbens and Ridder (2010) work with a variant of (3) 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 $\phi_0 = \mathbb{E}\left[A|\,T=1\right] - \mathbb{E}\left[A|\,T=0\right]$.

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:

$$\mathbb{E}\left[B|\,T, Z\right] = \mathbb{E}\left[B|\,T\right] = 0. \tag{4}$$

The second equality in (4) follows by construction. Condition (4) 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, than 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 Sealand, 1993; Duncan and Raudenbush, 2001; Kling, Liebman and Katz, 2007).

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

$$\mathbb{E}\left[U|\,T\right] = \mathbb{E}\left[U\right]. \tag{5}$$

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 (2), typically with additional individual- and location-level characteristics. How do these analyses relate to the structural model (1)? After tedious manipulation it is possible to show that the two slope coefficients in (2) equal[1]

$$\begin{aligned} b_0 &= \beta_0 + \phi_0'\lambda_0 \\ &\quad + \frac{1}{p\left(1-p\right)}\mathbb{C}\left(s\left(Z\right), \mathbb{E}\left[B|\,Z\right]\right)' \frac{\lambda_0}{1-\eta^2}. \end{aligned} \tag{6}$$

$$\begin{aligned} c_0 &= \gamma_0 + \phi_0'\delta_0 + \frac{1}{\eta^2}\left\{\frac{1}{p\left(1-p\right)}\mathbb{C}\left(s\left(Z\right), \mathbb{E}\left[B|\,Z\right]\right)'\left(\delta_0 + \frac{\lambda_0}{1-\eta^2}\right)\right. \\ &\quad \left. + \left(\mathbb{E}\left[U|\,T=1\right] - \mathbb{E}\left[U|\,T=0\right]\right)'\kappa_0\right\}. \end{aligned} \tag{7}$$

where $p = \mathbb{E}\left[T\right]$ is the population fraction Minority and η^2 the eta-squared index of segregation (e.g., Farley, 1977). The η^2 index coincides with the (linear) regression coefficient on T in the least squares fit of $s\left(Z\right)$ onto a constant and T.[2] 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) $\mathbb{C}\left(T - s\left(Z\right), s\left(Z\right)\right) = 0$, and (iii) $\mathbb{V}\left(T\right) = \mathbb{V}\left(s\left(Z\right)\right) + \mathbb{E}\left[\mathbb{V}\left(T|\,Z\right)\right]$ as well as

[1] To keep the calculations manageable I assume there is a continuum of neighborhoods. Calculation details are provided in Appendix A.

[2] Here and it 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).

some algebra gives

$$\eta^2 = \frac{C(T, s(Z))}{p(1-p)} = \frac{\mathrm{I} - p}{1-p} \tag{8}$$

where $\mathrm{I} = \mathbb{E}\left[\frac{s(Z)}{p} s(Z)\right]$ is a measure of minority isolation. 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 (8)) relative to the corresponding excess isolation that would be observed in a perfectly segregated city (the denominator in (8)).

Under the no sorting on unobservables condition (4) 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

$$b_0 = \beta_0 + \phi_0' \lambda_0, \tag{9}$$

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.[3] When condition (4) 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 (6) 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 (4) fails), then 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 (2). It is this coefficient which purports to provide a measure of peer group or neighborhood effects. The third term in (7) 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 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 $\gamma_0 + \phi_0'\delta_0$,

[3] When 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)).

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

$$c_0 = \gamma_0 + \phi_0'\delta_0 + \frac{1}{\eta^2}\left(\mathbb{E}\left[U|\, T=1\right] - \mathbb{E}\left[U|\, T=0\right]\right)'\kappa_0. \tag{10}$$

The first two terms in (10) are invariant across reallocations, while the last is not. However if we additionally impose the no matching condition (5) the third term drops out leaving

$$c_0 = \gamma_0 + \phi_0'\delta_0, \tag{11}$$

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

I conclude that regression (2), is informative about neighborhood effects when conditions (4) and (5) 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.

Expressions (6) and (7) also provide a framework for understanding conventional neighborhood effects analyses based on observational data, where conditions (4) and (5) 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 $\mathbb{C}\left(s\left(Z\right), \mathbb{E}\left[B|\, Z\right]\right) < 0$. Sorting by background biases the coefficient on Minority downward relative to the no sorting benchmark case (cf., equation (9)). When the Minority population share is small (p small), and/or segregation by race large (η^2 big), the negative bias will be even larger.

Now consider the coefficient on fraction Minority in (2). Terms three and four in (7) are bias terms. The third term is proportional to the sorting bias term in (6). 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.

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 (2). 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. Of course, as in other areas of causal analysis, approaches based on covariate adjustment are not always compelling. This observation has led researchers to develop other research designs for neighborhood effects analysis.

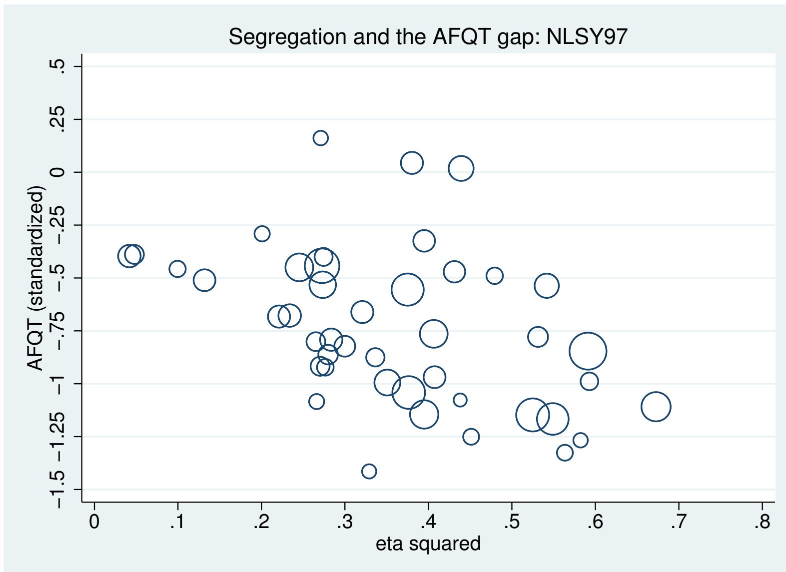
While conditions (4) and (5) are strong, there are examples of real world datasets where they are plausible. For example in the Project STAR class size reduction experiment, students were first randomly grouped into classes (no sorting), with classes then assigned randomly to teachers (no matching) (see Graham (2008) for additional details and caveats). The value of such experiments for enhancing our understanding of peer group and neighborhood effects is considerable.

Cross-city research designs

Several researchers have proposed estimation methods involving aggregation of (1) 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). 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 $(A', T)'$ various 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

Figure 1: 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 2,292 *male* 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 (baseline sampling weights used in this computation). The x-value corresponds to the eta-squared measure of Minority segregation in 2000. The size of the scatter points are proportional to the estimated precision of the corresponding Minority-White AFQT gap. A total of 44 MSAs are represented in the figure.

$c = 1, \ldots, N$ and sampled respondents within a city by $i = 1, \ldots, M_c$. 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)).

We begin by modifying (1) to incorporate a city-specific effect (i.e., intercept):

$$\begin{aligned} \mathbb{E}\left[Y_{ci} | T_{ci}, s(Z_{ci}), m_A(Z_{ci}), U_{ci}, A_{ci} \right] &= \alpha_c + \beta_0 T_{ci} + \gamma_0 s(Z_{ci}) \\ &\quad + m_A(Z_{ci})' \delta_0 + U_{ci}' \kappa_0 + A_{ci}' \lambda_0. \end{aligned} \tag{12}$$

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. Wooldridge (1997, Section 4) summarizes the basic properties of conditional linear predictors (CLPs). 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

$$\mathbb{E}^*\left[s(Z_{ci}) | T_{ci}; c \right] = \left(1 - \eta_c^2\right) p_c + \eta_c^2 T_{ci} \tag{13}$$

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 ($\eta_c^2 \to 1$) own race is very predictive of neighbors' race. In integrated cities ($\eta_c^2 \to 0$) the city-wide fraction Minority has more predictive value.[4]

Define

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

to be the Minority-White gap within city c in (i) "background" (ϕ_c), (ii) neighbors' background (υ_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 neighborhood. Specifically in the unobserved

[4] Note that

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

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

$$\begin{aligned}\mathbb{E}\left[Y_{ci}\right|T_{ci};c] - \mathbb{E}\left[Y_{ci}\right|c] &= \left\{\beta_0 + \left(\gamma_0 + \phi_c'\delta_0\right)\eta_c^2\right.\\ &\quad \left. + \upsilon_c'\delta_0 + \tau_c'\kappa_0 + \phi_c'\lambda_0\right\}\\ &\quad \times \left(T_{ci} - p_c\right). \end{aligned} \tag{14}$$

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 (14) indicates that the Minority-White *outcome* gap in city c – $\mathrm{GAP}_c = \mathbb{E}\left[Y_{ci}\right|T_{ci} = 1; c] - \mathbb{E}\left[Y_{ci}\right|T_{ci} = 0; c]$ – varies with its degree of segregation, as measured by the eta-squared, η_c^2, index:

$$\mathrm{GAP}_c = a_0 + \left(\gamma_0 + \phi_0'\delta_0\right)\eta_c^2 + V_c \tag{15}$$

with

$$\begin{aligned} a_0 &= \beta_0 + \upsilon_0'\delta_0 + \tau_0'\kappa_0 + \phi_0'\lambda_0 \\ V_c &= \left(\upsilon_c - \upsilon_0\right)'\delta_0 + \left(\tau_c - \tau_0\right)'\kappa_0 + \left(\phi_c - \phi_0\right)'\lambda_0 + \left(\phi_c - \phi_0\right)'\delta_0\eta_c^2. \end{aligned}$$

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

$$\mathbb{E}\left[V_c\eta_c^2\right] = 0 \tag{16}$$

then, by equations (14) and (15), an OLS 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 (2) under no sorting and matching (i.e., when conditions (4) and (5) hold).

Restriction (16) 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 (16) requires that, across cities, variation in the "background gap" is uncorrelated with variation in racial segregation (i.e, $\mathbb{C}\left(\phi_c, \eta_c^2\right) = 0$). 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 (2) is (see (7) above)

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

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

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

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 (16) requires maintaining additional (strong) assumptions. First, for $\delta_0 \neq 0$, we require zero covariance between measured racial segregation and $\left(\upsilon_c - \upsilon_0\right)'\delta_0$, 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 it to be similar across cities. When $\delta_0 \neq 0$ we also require strengthening the zero covariance condition $\mathbb{C}\left(\phi_c, \eta_c^2\right) = 0$. For example if ϕ_c is mean-independent of η_c^2, then $\left\{\left(\phi_c - \phi_0\right)'\delta_0\eta_c^2\right\}\eta_c^2$, 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 $\mathbb{C}\left(\phi_c, \eta_c^2\right) = 0$, then they ought to be willing to assume that $\mathbb{E}\left[\phi_c|\,\eta_c^2\right] = \phi_0$. 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 $\left(\tau_c - \tau_0\right)'\kappa_0$, 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 $E\left[\phi_c | \eta_c^2\right] = 0$ (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. 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

$$\mathbb{E}^*\left[\text{GAP}_c | \eta_c^2\right] = \beta_0 + \phi_0'\lambda_0 + \Pi_\upsilon'\delta_0 + \Pi_\tau'\kappa_0 + \left(\gamma_0 + (\phi_0 + \Pi_{\upsilon\eta})'\delta_0 + \Pi_{\tau\eta}'\kappa_0\right)\eta_c^2 \quad (17)$$

where

$$\begin{aligned} \mathbb{E}^*\left[\upsilon_c | \eta_c^2\right] &= \Pi_\upsilon + \Pi_{\upsilon\eta}\eta_c^2 \\ \mathbb{E}^*\left[\tau_c | \eta_c^2\right] &= \Pi_\tau + \Pi_{\tau\eta}\eta_c^2. \end{aligned}$$

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 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. 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 1 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 (16) 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 $\gamma_0 + \phi_0'\delta_0$:

$$\widehat{\text{AFQTGAP}}_c = \underset{(0.0746)}{-0.3391} + \underset{(0.2044)}{-1.0649}\ \text{ETA}_c^2 .$$

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 (16) the data suggests that we can strongly reject the null that $\gamma_0 + \phi_0'\delta_0 = 0$. Again, maintaining condition (16), 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 (17), 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 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 1 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.

References

[1] 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.

[2] 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.

[3] Durlauf, Steven N. (1996). "Associational redistribution: a defense," *Politics & Society* 24 (4): 391 - 410.

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

[5] Culter, David M. and Edward L. Glaeser. (1997). "Are ghettos good or bad?" *Quarterly Journal of Economics* 112 (3): 827 - 872.

[6] 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.

[7] 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.

[8] 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.

[9] Goldberger, Arthur. (1991). *A Course in Econometrics.* Cambridge, MA: Harvard University Press.

[10] Graham, Bryan S. (2008). "Identifying social interactions through conditional variance restrictions," *Econometrica* 76 (3): 643 - 660.

[11] 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.

[12] 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.*

[13] Kling, Jeffrey R., Jeffrey B. Liebman, Lawrence F. Katz. (2007). "Experimental analysis of neighborhood effects," *Econometrica* 75 (1): 83 - 119.

[14] 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.

[15] Wooldridge, Jeffrey M. (1997). "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics* 90 (1): 77 - 97.

[16] 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.

A Derivations

To derive (6) and (7) 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

$$b_0 = \frac{1}{p(1-p)(1-\eta^2)} \{\mathbb{C}(T,Y) - \mathbb{C}(s(Z),Y)\} \tag{18}$$

$$c_0 = \frac{1}{p(1-p)\eta^2(1-\eta^2)} \left\{\mathbb{C}(s(Z),Y) - \eta^2\mathbb{C}(T,Y)\right\}. \tag{19}$$

To evaluate the terms in $\{\cdot\}$ in (18) and (19) first compute the covariances:

$$\begin{aligned}\mathbb{C}(T,A) &= \mathbb{E}[(T-p)A] \\ &= \mathbb{E}[TA] - p\mathbb{E}[A] \\ &= p\mathbb{E}[A|T=1] - p(p\mathbb{E}[A|T=1] + (1-p)\mathbb{E}[A|T=0]) \\ &= p(1-p)(\mathbb{E}[A|T=1] - \mathbb{E}[A|T=0]) \\ &= p(1-p)\phi_0,\end{aligned} \tag{20}$$

and

$$\begin{aligned}\mathbb{C}(T,U) &= \mathbb{E}\left[(T-p)U\right] \\ &= p(1-p)\left(\mathbb{E}\left[U|T=1\right]-\mathbb{E}\left[U|T=0\right]\right),\end{aligned} \tag{21}$$

and

$$\begin{aligned}\mathbb{C}(T,m_A(Z)) &= \mathbb{E}\left[(T-p)m_A(Z)\right] \\ &= \mathbb{E}\left[(T-p)\left\{\pi_0+\phi_0 s(Z)+\mathbb{E}\left[B|Z\right]\right\}\right] \\ &= \eta^2 p(1-p)\phi_0+\mathbb{E}\left[(s(Z)-p)\mathbb{E}\left[B|Z\right]\right],\end{aligned} \tag{22}$$

and

$$\begin{aligned}\mathbb{E}\left[(s(Z)-p)A\right] &= \mathbb{E}\left[(s(Z)-p)(\pi_0+\phi_0 T+B)\right] \\ &= \eta^2 p(1-p)\phi_0+\mathbb{E}\left[(s(Z)-p)B\right] \\ &= \eta^2 p(1-p)\phi_0+\mathbb{E}\left[(s(Z)-p)\mathbb{E}\left[B|Z\right]\right] \qquad (23)\\ \mathbb{E}\left[(s(Z)-p)m_A(Z)\right] &= \mathbb{E}\left[(s(Z)-p)(\pi_0+\phi_0 s(Z)+\mathbb{E}\left[B|Z\right])\right] \\ &= \eta^2 p(1-p)\phi_0+\mathbb{E}\left[(s(Z)-p)(\mathbb{E}\left[B|Z\right])\right] \qquad (24)\\ \mathbb{E}\left[(s(Z)-p)U\right] &= p(1-p)\left(\mathbb{E}\left[U|T=1\right]-\mathbb{E}\left[U|T=0\right]\right). \qquad (25)\end{aligned}$$

Using (20), (21), (22), (23), (24) and (25) I evaluate

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

and

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

which yields, after subtracting,

$$\begin{aligned}
\mathbb{C}(T,Y)-\mathbb{C}(s(Z),Y) &= \left(1-\eta^2\right)p(1-p)\beta_0+\phi_0'\lambda_0 p(1-p)-\mathbb{C}(s(Z),A)'\lambda_0 \\
&= \left(1-\eta^2\right)p(1-p)\beta_0+\phi_0'\lambda_0 p(1-p) \\
&\quad -\eta^2 p(1-p)\phi_0'\lambda_0+\mathbb{E}\left[(s(Z)-p)\mathbb{E}[B|Z]\right]'\lambda_0 \\
&= \left(1-\eta^2\right)p(1-p)\{\beta_0+\phi_0'\lambda_0\}-\mathbb{E}\left[(s(Z)-p)\mathbb{E}[B|Z]\right]'\lambda_0.
\end{aligned}$$

Substituting this expression in (18) above then gives (6) of the main text:

$$b_0=\beta_0+\phi_0'\lambda_0+\frac{1}{1-\eta^2}\frac{1}{p(1-p)}\mathbb{E}\left[(s(Z)-p)\mathbb{E}[B|Z]\right]'\lambda_0.$$

To derive (7) I evaluate the difference:

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

which, after substituting into (19), gives (7) of the main text:

$$
\begin{aligned}
c_0 &= \gamma_0 + \phi_0'\delta_0 \\
&\quad + \frac{1}{p(1-p)\eta^2(1-\eta^2)}\left\{\left(1-\eta^2\right)\mathbb{E}\left[\left(s(Z)-p\right)\mathbb{E}\left[B|\,Z\right]\right]'\delta_0\right. \\
&\quad \left. + \mathbb{E}\left[\left(s(Z)-p\right)\mathbb{E}\left[B|\,Z\right]\right]'\lambda_0 + \left(1-\eta^2\right)\mathbb{C}\left(s(Z),U\right)'\kappa_0\right\} \\
&= \gamma_0 + \phi_0'\delta_0 + \frac{1}{\eta^2}\left\{\frac{1}{p(1-p)}\mathbb{E}\left[\left(s(Z)-p\right)\mathbb{E}\left[B|\,Z\right]\right]'\delta_0\right. \\
&\quad + \frac{1}{1-\eta^2}\frac{1}{p(1-p)}\mathbb{E}\left[\left(s(Z)-p\right)\mathbb{E}\left[B|\,Z\right]\right]'\lambda_0 \\
&\quad \left. + \left(\mathbb{E}\left[U|\,T=1\right]-\mathbb{E}\left[U|\,T=0\right]\right)'\kappa_0\right\}.
\end{aligned}
$$

Equation (14) in the main text is easily calculated using the following terms:

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