Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach[1]

Bryan S. Graham[†] Guido W. Imbens[‡] Geert Ridder[♮]

August 30, 2024

ABSTRACT: In this paper we nonparametrically analyze the effects of reallocating individuals across social groups in the presence of social spillovers. Individuals are either 'high' or 'low' types. Own outcomes may vary with the fraction of high types in one's social group. We characterize the average outcome and inequality effects of small increases in segregation by type. We also provide a measure of average spillover strength. We generalize the setup used by Benabou (1996) and others to study sorting in the presence of social spillovers by incorporating unobserved individual- and group-level heterogeneity. We relate our reallocation estimands to this theory. For each estimand we provide conditions for nonparametric identification, propose estimators, and characterize their large sample properties. We also consider the social planner's problem. We illustrate our approach by studying the effects of sex segregation in classrooms on mathematics achievement.

JEL CLASSIFICATION: C14, C31, D62, I21

KEY WORDS: Social Interactions, Peer Group Effects, Sorting, Segregation, Equity vs. Efficiency Trade-off, Allocation Problems, Semiparametric M-estimation.

[1]We thank David Ahn, Susan Athey, Alan Auerbach, Alberto Bisin, Raj Chetty, Rajeev Dehejia, Steven Durlauf, Raquel Fernández, Jack Porter, Jim Powell, Paul Ruud and, especially, Emmanuel Saez for helpful discussions. We would also like to thank participants in numerous seminars and conferences for helpful questions and suggestions. Financial support was generously provided by the NSF (SES #0820361), the Office of Naval Research for support under grant numbers N00014-17-1-2131 and N00014-19-1-2468 and Amazon for a gift. This revision has benefited from the feedback and suggestions of the anonymous associate editor and a referee. All the usual disclaimers apply. This is an updated version of our 2010 NBER Working Paper (No. 16499) with the same title.

[†]Department of Economics, University of California - Berkeley, 530 Evans Hall 3880, Berkeley, CA 94720-3880 and National Bureau of Economic Research. E-MAIL: bryangraham@berkeley.edu, WEB: `https://bryangraham.github.io/econometrics/`.

[‡]Graduate School of Business and Department of Economics, Stanford University, 655 Knight Way, Stanford, CA 94305 and the National Bureau of Economic Research. E-MAIL: imbens@stanford.edu, WEB: `https://www.gsb.stanford.edu/faculty-research/faculty/guido-w-imbens`.

[♮]Department of Economics, University of Southern California, Kaprielian Hall, University Park Campus, Los Angeles, CA 90089. E-MAIL: ridder@usc.edu, WEB: `https://dornsife.usc.edu/profile/geert-ridder/`.

1 Introduction

Debates about the social costs and benefits of segregation by socioeconomic status, ability, race or gender figure prominently in discussions of education, housing and other areas of social policy. In the late-1960s Coleman (1966) argued that racial isolation lowered the academic achievement of minority students. This claim immediately generated controversy, spawning a vast empirical literature in education, sociology and economics. Forty years later Rivkin and Welch (2006), surveying the resulting body of work, concluded that "the effect of integration on black students remains largely unsettled" (p. 1043). Schofield (1995), reviewing the education and sociology literature, comes to a similarly tentative conclusion, emphasizing the "methodological and other problems that typify work in this area" (p. 597). After almost six decades of research, school busing and other mandated desegregation policies remain controversial. Other unsettled debates touching on issues of 'segregation' include those on school vouchers, single-sex schooling, ability tracking and public housing policy.[2]

Each of these debates centers on a common question: would society be better off if social groups were configured differently? Are there welfare-increasing deviations from the status quo assignment of individuals to classrooms, schools or neighborhoods? How do average outcomes and inequality respond to 'reallocations' of individuals across groups? Durlauf (1996a) has termed such reallocating policies 'associational redistribution'.

Despite the long-standing controversy surrounding reallocation-inducing policies, econometric methods for framing and analyzing their effects are not widely available. Researchers interested in, for example, segregation in schools typically focus their efforts on identifying and estimating an average relationship between school racial composition and student achievement (e.g., Angrist and Lang, 2004; Guryan, 2004; Card and Rothstein, 2007). The optimality of segregation relative to integration is inferred by reference to this estimated relationship.[3] The target estimand of this literature, the average marginal effect of school

[2] Disagreements about the magnitude and relevance of 'cream-skimming' in response to widespread school choice figure prominently in the debate on educational vouchers (e.g. Hoxby, 2003; Ladd, 2003; Manski, 1992; Urquiola, 2005)

The evidence on the achievement effects of single-sex instruction is mixed (e.g., Morse, 1998; Mael, 2005), although this interpretation is debated by advocates of gender-separation (e.g., Sax, 2005). In 2006 the United States Department of Education, in a controversial decision, modified Title IX regulations to allow the formation of single-sex classrooms in public schools (Paulson and Teicher, 2006).

The literature on school tracking is enormous with supporting evidence available for both its advocates and opponents. For discussions see Oakes (1992) Epple et al. (2002), and Figlio and Page (2002).

Massey and Denton (1993, p. 231) advocate for increased use of housing vouchers and decreased use of public housing projects. The effects of housing vouchers are analyzed by Jacob (2004) and Kling et al. (2007).

[3] The original Coleman Report provides a particularly thoughtful example of this type of informal inference

racial composition on student achievement, does not correspond to an implementable policy. It would be impossible, for example, to engineer an increase in minority enrollment across all schools – the policy effect measured by this estimand – since an increase in such enrollment in one school necessarily requires a commensurate decrease in another. While knowledge of the (average) mapping between school racial composition and outcomes may be an ingredient to an evaluation of a particular race-based allocation of students to schools, it is not sufficient.[4]

In this paper we provide an initial exploration of the econometrics of reallocating individuals across groups in the presence of social spillovers. Our analysis emphasizes issues of measurement, that is, the definition of relevant target estimands. Additionally, we provide conditions for nonparametric identification, propose estimators and characterize their large sample properties. We implement our procedures using data from the randomized Tennessee class size reduction experiment, Project STAR. Following Whitmore (2005) we use these data to study the effects of classroom gender mix on student achievement.

Our setup generalizes that of a class of stylized locational sorting models developed by de Bartolome (1990), Benabou (1993, 1996), Becker and Murphy (2000) and others.[5] As in those papers, we consider a setting where individuals are either 'high' or 'low' types, with outcomes nonparametrically depending on the type composition of their social group. We add statistical content to this framework by introducing unobserved individual heterogeneity. We also allow for location-specific heterogeneity (both observed and unobserved). These extensions complicate our analysis but are, of course, essential for empirical relevance.

An example, which we develop empirically below, helps to clarify the various issues involved. Consider a setting where individuals are students, with high and lows types respectively denoting girls and boys. Students may differ in unobserved ways, for example in their ability. A social group is a classroom of students. Classrooms may also be heterogeneous, for example in observed and/or unobserved dimensions of teacher quality. This set-up is complicated because there are three distinct levels of heterogeneity: individual-level, peer-level and location-level. Any analysis of peer effects must keep track of, and impose conditions

process:

> "If a white pupil from a home that is strongly and effectively supportive of education is put in a school where most students do not come from such homes, his achievement will be little different than if he were in a school composed of others like himself. But if a minority pupil from a home without much educational strength is put with schoolmates with strong educational backgrounds, his achievement is likely to increase" Coleman (1966, p. 22).

[4]More generally the menu of program evaluation estimands surveyed by Imbens (2004), Heckman and Vytlacil (2007a,b) and others is, at best, only indirectly helpful for assessing the effects of reallocations. We justify this claim further below.

[5]Much of this theoretical literature is surveyed by Piketty (2000), Fernández (2003) and Durlauf (2004).

on, these three types of heterogeneities. Our approach involves imposing restrictions on the group formation process; both the mechanism whereby specific individuals sort together into groups, and that whereby such groups place themselves in specific locations. While we are restrictive regarding the process which generates the status quo allocation of individuals to groups, we are very flexible elsewhere. An alternative, complementary, approach would involve imposing more restrictions on, say, the 'production technology', in exchange for imposing fewer restrictions on the status quo assignment process (e.g., Calabrese et al., 2006; Ferreyra, 2007; Nesheim, 2002). We emphasize that our basic setup, in particular our estimands and characterization of the social planner's problem, is not linked to any specific approach to identification.

We develop three classes of estimands. The first class measures the average strength of any social spillovers. The central focus here is on what we call the *average spillover effect*, β^{ase}. Here our contribution is modest; we provide a nonparametric generalization of prior work on the measurement of spillovers (e.g., Manski, 1993; Brock and Durlauf, 2001; Moffitt, 2001; Glaeser and Scheinkman, 2003). In particular our measure of spillover strength can be viewed as a (simple) nonparametric generalization of Ciccone and Peri's (2006) 'constant composition' externality measure.

We view our second set of estimands as more innovative. This class includes the *local segregation outcome effect*, β^{lsoe}, which measures the effect of small increases in segregation (relative to the status quo) on average outcomes. We also develop a *local segregation inequality effect*, β^{lsie}, which measures the effect of a small increase in segregation on the the average outcome gap between high and low type individuals. These estimands provide a basis for characterizing any equity versus efficiency trade-offs associated with segregation-inducing policies.

Our final estimand allows us to assess the efficiency of the status quo allocation relative to an outcome-maximizing allocation. In our setup the social planner's problem is a functional optimization (i.e., infinite dimensional) one. Nevertheless we are able to characterize its solution quite generally. As we leave the (average) mapping from group composition to outcomes a priori unrestricted (and also allow for a large number of social groups) our result generalizes the social planner analyses of, for example, de Bartolome (1990), Benabou (1993, 1996) and Becker and Murphy (2000), in addition to providing them with statistical content.

Our framework offers several advantages over existing methods of characterizing social spillovers. First, our approach explicitly connects the data with many of the ideas emphasized in theoretical work on sorting in the presence of social spillovers. In particular, our estimands provide measures of segregation-induced inefficiencies, a key theme of the neighborhood sorting literature. For example, our *local segregation outcome effect* (LSOE) estimand has a representation as a weighted average of own and peer type *complementarity* and *curvature*. Benabou (1996), in the context of a stylized deterministic model, shows how the efficiency of segregation vis-a-vis integration depends on these two objects. Prior empirical work on social externalities generally only loosely connects to the relevant applied public finance theory. Fernández (2003), in her survey article, notes that "there has been very little work done to assess the significance of the inefficiencies [induced by segregation]," despite the growing body of empirical work that points to the importance of peer effects in a general way (p. 14). Piketty (2000) makes a similar point.

Second our focus on reallocations is novel. While we leave the microstructure of any social interactions processes unmodelled, our set-up allows us to think about reallocation-inducing policies in a straightforward way. Many controversial policies, such as busing, 'school choice' regimes or the provision of rental vouchers to public housing recipients, are fundamentally allocation mechanisms. Our estimands provide a partial basis for the evaluation of such policies.

Finally, unlike most work in this area, Brock and Durlauf (2007) being an important exception, our approach to identification and estimation is fully nonparametric.[6] We provide nonparametric estimators for our first two classes of estimands and also characterize their large sample properties.[7]

In recent years economists and other social scientists have made substantial progress on the identification and estimation of statistical models with social spillovers (e.g., Manski, 1993; Solon, 1999; Brock and Durlauf, 2001; Moffitt, 2001; Duncan and Raudenbush, 1993; Sampson et al., 2002; Glaeser and Scheinkman, 2003; Graham, 2008). Our work builds on this work inasmuch as the production technology is a component of each of our estimands. However our focus substantially differs from this prior work. Our goal is to develop estimands which *directly* characterize the effects of reallocations on the distribution of outcomes.

[6]Examples of formal identification analyses of parametric social interaction models include those of Manski (1993), Brock and Durlauf (2001), Moffitt (2001) and Graham (2008).

[7]A limitation of our framework is that it is not helpful for assessing the effects of non-reallocating interventions, such as providing subsidies to low types. Manski (1993), Brock and Durlauf (2001) and Durlauf (2004) discuss this class of policy interventions. The analysis of such interventions generally requires an explicit model of the social interaction process. Durlauf (2004) makes a compelling case for greater focus on the microeconomic foundations of social interaction processes. We are sympathetic to this perspective, but nevertheless have found it useful to leave such structure unspecified in the present setting. Lazear (2001) provides a nice, and now seminal, example of how a concrete microstructure of social interaction generate specific reduced form mappings from group structure into outcomes. Since we leave this mapping nonparametric, our approach is arguably consistent with a wide-variety of interaction microstructures. An important caveat to this claim, however, is that explicit microstructures of strategic interaction can generate a mapping from group composition into outcomes that exhibits discontinuities (e.g., Brock and Durlauf, 2001, 2007). Since we estimate this mapping using kernel smoothing methods, our approach may work poorly in such situations.

Related work in this vein includes that of Graham et al. (2007, 2014, 2018) and Bhattacharya (2009) (see Graham (2011) for a survey). More recently Hudgens and Halloran (2008) and Manski (2013) develop a notation for the study of treatment response in the presence of spillovers which shares features with our own setup.

Our work is also related to the mathematical programming and economic literature on resource allocation problems (e.g., Ginsberg (1974); Ibaraki and Katoh (1988); Luenberger (1969, 2005)). As noted above, in our setting the planner's problem is one of functional optimization. Our general characterization of the solution to this problem appears to be new.[8]

The statistical aspects of this paper are most closely connected to the literature of semi-parametric M-estimation as in Newey (1994a,b) and Newey and McFadden (1994). In particular our estimands share important features with weighted average derivatives as in Powell et al. (1989), Härdle and Stoker (1989), Newey and Stoker (1994) and others. While straightforward to compute, our estimators combine multiple first step nonparametrically estimated objects together in different ways. Most of our estimators, for example, require nonparametric estimation of two conditional expectation functions as well as their derivatives. Consequently characterizing their asymptotic properties, as we do below, is nontrivial.

Section 2, which follows next, describes our sampling structure and maintained identifying assumptions. The need to carefully keep track of all the sources of individual, peer and locational heterogeneity requires the development of a relatively elaborate set of notational conventions. For our purposes we have found a heavily modified potential outcomes notation to be the most convenient for representing our problem and stating our assumptions (Neyman, 1990; Rubin, 1974; Holland, 1986). To simplify the exposition we begin with the stylized case where all groups are (i) equally sized and (ii) there are no covariates beyond type (that status quo assignment of individuals to locations is, of course, known).

Section 3 presents our estimands. We begin by proposing a simple summary measure of the strength of social spillovers. We then present measures of the outcome and inequality effects of local reallocations of individuals across groups. Section 4 discusses estimation. Section 5 briefly considers how observed individual- and location-specific characteristics can be incorporated into our framework.

In Section 6 we discuss the planner's problem. By characterizing the solution to this

[8]The closest work of which we are is aware is that of Arnott and Rowse (1987) which uses parametric estimates of educational production functions and numerical programming methods to evaluate classroom assignment mechanisms based on student ability. Their methods are fundamentally parametric in nature and they do not discuss issues of identification, estimation or inference. Our analysis of the allocation problem is also related to the neighborhood sorting models of de Bartolome (1990), Benabou (1993, 1996), Durlauf (1996b,c), Epple and Romano (1998) and Becker and Murphy (2000).

problem we are able to show that the inefficiency of the status quo – the difference between the observed average outcome and that which would occur under an outcome-maximizing allocation – is identified under certain assumptions. In Section 7 we apply our methods, and compare them with parametric alternatives, in a study of the effect of classroom gender composition on student achievement using data collected in conjunction with the Tennessee Project STAR experiment (see Whitmore, 2005). Section 8 summarizes and suggests areas for future research. The proofs of our identification and representation results are contained in Appendix B. The pathwise derivative calculations underlying our large sample results are detailed in a Supplemental Web Appendix.

2 Setup and assumptions

In this section we present our statistical model and discuss the identifying assumptions we maintain in subsequent sections. Throughout we use upper case letters to denote random variables. Lower-case and calligraphic letters respectively denote specific realizations and the support of the corresponding distributions. That is Y, y and $\mathbb{Y}$ respectively denote a generic random draw of, a specific value of, and the support of, Y. A "0" subscript on a parameter denotes its population value and may be omitted when doing so causes no confusion.

2.1 Population framework

There exists a population of individuals (e.g., elementary school students) indexed by $i \in \mathcal{I} = \{1, \ldots, I_P\}$. Individuals are one of two observed types $T_i \in \{0, 1\}$, for example, boy or girl. Additional individual level heterogeneity is contained in the vector $A_i \in \mathcal{A}$. For reasons of exposition we refer to A_i as an individual's 'ability'. We also refer, without intending to be pejorative, to those individuals with $T_i = 1$ as 'high' types and those individuals with $T_i = 0$ as 'low' types. The population fraction of high types is given by p_H. We assume that T_i is non-manipulable, denoting a permanent characteristic such as race or sex assigned at birth. The outcome of interest, say, student achievement, is $Y_i \in \mathcal{Y}$ and may be discretely- or continuously-valued. For ease of exposition we initially assume there are no observed individual characteristics beyond type (we introduce observed individual-level attributes into our analysis in Section 5).

Individuals reside in different locations or, alternatively, 'attend' different 'schools'. Locations are indexed by $c \in \mathcal{C} = \{1, \ldots, C_P\}$. Associated with each location is a vector of unobserved characteristics $U_c \in \mathcal{U}$. If locations are, for example, schools, then U_c might capture heterogeneity in teacher quality and facilities (we introduce observed location characteristics into our analysis in Section 5).

At times it will be necessary to compute averages across the population of locations and, at others, ones across individuals. When we use a c subscript the relevant average is over locations, whereas an i subscript signals an average over individuals.[9]

Each individual's location of residence is given by the assignment indicator $G_i \in \mathcal{C}$. If individual i resides in location c, then $G_i = c$. To avoid double subscripting we use the notation $U_i = U_{G_i}$. An allocation is a feasible assignment of individuals to groups and is completely specified by a vector of group assignment indicators $\mathbf{G} = (G_1, \ldots, G_{I_P})'$. Feasibility of an assignment implies that (i) each individual is assigned to one, and only one, location and (ii) the capacity constraint associated with each location is respected (For example, if classroom c has twenty seats, then no more than twenty students are assigned to it). Feasibility is defined formally below.

Individuals assigned to a common location are neighbors. For ease of exposition we initially assume that all neighborhoods have room for exactly $N = I_P/C_P$ residents (we allow for unequally sized groups in Sections 5 and 6).

Individual $i's$ peer group includes those individuals also assigned to her location, i.e. the index set

$$p(i) = \{j : G_j = G_i,\ j \neq i\}.$$

These peers' types and abilities are given by the vectors

$$\underline{T}_{p(i)} = \left(T_{p(i),1}, \ldots, T_{p(i),N-1}\right)', \quad \underline{A}_{p(i)} = \left(A_{p(i),1}, \ldots, A_{p(i),N-1}\right)'.$$

where the subscripts $p(i), j$ with $j = 1, \ldots, N-1$ indicate the members of i's peer group in arbitrary order. Let $\underline{T}_i = (T_i, \underline{T}'_{p(i)})'$ and $\underline{A}_i = (A_i, \underline{A}'_{p(i)})'$ denote the vectors of types and abilities in $i's$ social group inclusive of herself.

The i^{th} individual's neighborhood quality, Q_i, depends on the type and ability of her peers as well as the vector of unobserved location characteristics U_i :

$$Q_i = (\underline{T}'_{p(i)}, \underline{A}'_{p(i)}, U'_i)'.$$

[9]In principle our "double use" of a single index notation could cause confusion; U_c and U_i denote the unobserved quality of location c and that of the location in which individual i resides respectively. This is clear. The meaning of U_{19}, in contrast, is ambiguous. We are careful to avoid such ambiguity in what follows.

2.2 Potential outcomes notation

Our focus is on characterizing different (summary) features of the mapping from allocations into outcomes. We assume that this mapping is individual-specific and given by

$$Y_i(\mathbf{g}), \quad \mathbf{g} \in \mathcal{G}, \tag{1}$$

where $\mathcal{G}$ denotes the set of all feasible allocations and the relation is individual specific due to its (implicit) dependence on T_i and A_i. The function $Y_i(\mathbf{g})$ gives the potential outcome for individual i associated with allocation $\mathbf{g} \in \mathcal{G}$.[10,11]

Tractability of our problem requires imposing restrictions on $Y_i(\mathbf{g})$. Our first restriction rules out cross location spillovers.

Assumption 2.1. (No Cross Neighborhood Spillovers) *Let* $\mathbf{g}$ *and* $\widetilde{\mathbf{g}}$ *denote two distinct feasible allocations with associated neighborhood qualities for individual* i *of* q_i *and* $\widetilde{q}_i$*. If* $q_i = \widetilde{q}_i$*, then*

$$Y_i(\mathbf{g}) = Y_i(\widetilde{\mathbf{g}}).$$

Assumption 2.1 means that individual outcomes depend only upon own characteristics and neighborhood quality; the type-structure, ability distribution, and location characteristics of, for example, adjacent neighborhoods do not affect outcomes. In the case where locations are spatially separated schools Assumption 2.1 may be reasonable. If locations represent residential neighborhoods the assumption of no cross location spillovers is considerably stronger. Nevertheless some restriction on the structure of dependence across locations is required for statistical analysis.

Under Assumption 2.1 we may write

$$Y_i(\mathbf{G}) = Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i) = Y_i(Q_i).$$

Our next assumption restricts the structure of peer influences within a neighborhood.

[10] Associated with each assignment is a mechanism by which it came about. For example assignment may be by lottery, tournament, or determined by a social planner. Implicit in (1) is the assumption that, conditional on the induced assignment, the mechanism by which it was achieved does not affect outcomes. If a court-ordered mandatory school busing plan induces the same allocation of students across schools as a lottery, then the associated outcome distributions will also be identical. This may be a strong assumption in certain settings. Schofield (1995), in her review of educational research on the impact of desegregation on black achievement, presents evidence suggesting that the desegregation mechanism matters. Similar (implicit) assumptions underlie the program evaluation literature (see Holland, 1986)

[11] The potential outcomes notation is convenient for our purposes, however, we could also use the 'production function' notation

$$Y_i = g\left(T_i, \mathbf{G}, A_i\right),$$

with A_i playing the role of a (non-separable) disturbance.

Let $N_i^H = \sum_{j=1}^{N-1} T_{p(i),j}$ and $N_i^L = \sum_{j=1}^{N-1}\left(1 - T_{p(i),j}\right)$ denote the total number of high and low type peers for individual i (here $T_{p(i),j}$ denotes the j^{th} element of $\underline{T}_{p(i)}$). Assume, without loss of generality, that $T_{p(i)}$ is ordered such that high types appear first, followed by low types (i.e., $T_{p(i)} = (1, \ldots, 1, 0, \ldots, 0)'$). The $N-1$ vector of peer 'abilities' is arranged conformably such that $\underline{A}_{p(i)} = (\underline{A}_{p(i)}^{H\prime}, \underline{A}_{p(i)}^{L\prime})'$, where $\underline{A}_{p(i)}^{H}$ equals the $N_i^H \times 1$ vector of abilities for each high type peer in individual $i's$ social group and $\underline{A}_{p(i)}^{L}$ equals the corresponding $N_i^L \times 1$ vector of low type peer abilities.

Assumption 2.2. (Within-Type Peer Exchangeability) *Let $\widetilde{\underline{A}}_{p(i)} = (\widetilde{\underline{A}}_{p(i)}^{H\prime}, \widetilde{\underline{A}}_{p(i)}^{L\prime})'$ where $\widetilde{\underline{A}}_{p(i)}^{H}$ and $\widetilde{\underline{A}}_{p(i)}^{L}$ are permutations of $\underline{A}_{p(i)}^{H}$ and $\underline{A}_{p(i)}^{L}$, and let $\widetilde{\underline{T}}_{p(i)}$ be a conformable re-ordering of $\underline{T}_{p(i)}$ (note that $\widetilde{\underline{T}}_{p(i)} = \underline{T}_{p(i)}$ by construction), for all such within-type permutations (i)*

$$Y_i(\widetilde{\underline{T}}_{p(i)}, \widetilde{\underline{A}}_{p(i)}, U_i) = Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i)$$

and (ii) the function $Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i)$ is a continuous function of $\left(\underline{A}_{p(i)}, U_i\right)$ for all $\underline{T}_{p(i)}$.

Assumption 2.2 implies that, among those of the same type, each of individual $i's$ peers are equally influential. This restriction follows from standard exchangeability arguments. As such it is a statement of researcher ignorance: *a priori* there is no reason to think that $i's$ 'first' high type neighbor affects her differently than her 'ninth' (Rubin, 1981). Manski (2000) and Durlauf (2001) have argued for improving data collection in order to avoid such restrictions. For example, if the researcher knew that $i's$ 'ninth' high type neighbor was across the street, while her 'first' was two blocks away, then Assumption 2.2 might be implausible. However, in most datasets, the structure of within-group social networks is unavailable and hence Assumption 2.2 is an appropriate, as well as unavoidable, representation of prior information.[12]

Let S_{-i} denote the fraction of i's peers that are high types (i.e., $S_i = N_i^H/N$ and $S_{-i} = \frac{N_i^H - T_i}{N-1}$. By Assumption 2.2 and the Weierstrass Theorem we can approximate the function $Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i)$ by

$$Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i) \approx Y_i(S_{-i}, \tau_{K_H}(\underline{A}_{p(i)}^{H}), \tau_{K_L}(\underline{A}_{p(i)}^{L}), U_i)$$

with $\tau_{K_H}(\underline{A}_{p(i)}^{H})$ denoting the vector of the first K_H symmetric polynomials in $\underline{A}_{p(i)}^{H}$ and

[12]Calvó-Armengol et al. (2009) provide a nice example of how richer network data can be used to study peer influences. Examples of this nature are more numerous today relative to when this paper was initially drafted. Note that, as may become apparent to careful readers below, exchangeability of the potential response function in the peer abilities' does *not* appear to be required for our main results. This follows because these latent variables are integrated out below under a particular density factorization.

$\tau_{K_L}(\underline{A}^L_{p(i)})$ defined similarly (see Altonji and Matzkin, 2005, pp. 1062 - 1063).[13]

We emphasize that Assumption 2.2 allows for individuals to be differentially affected by the ability structure of their high- and low-type peers. For example, outcomes may vary freely with the average ability of low type peers *and/or* the average ability of high type peers (rather than being restricted to vary with average ability taken across all peers). Some individuals, for example, may be particularly sensitive to variation in high-type peer ability, while others to variation in low-type peer ability.

Our final restriction on $Y_i(\mathbf{g})$ follows from being precise about the meaning of an agent's type.

Assumption 2.3. (Inclusive Definition of Type) $T_i \perp A_i$

Independence of A_i from T_i follows by definition of the phenomena we seek to characterize. We are interested in whether, for example, an individual learns more when surrounded by female classmates. Not whether he learns more when surrounded by female classmates once we condition on their 'disruptiveness'. If, across the population under consideration, girls tend to be less disruptive than boys, then these two questions have different answers. For the first question the appropriate definition of A_i is precisely all individual heterogeneity that is independent of T_i. We want our notion of 'gender' to include, not exclude, systematic differences in behavior across boys and girls.[14]

If A_i is scalar Assumption 2.3 can always be imposed by a normalization. Assume that unnormalized ability is A_i^*, then normalized ability is given by $A_i = F\left(A_i^* \middle| T_i\right)$. That is our

[13] For the case where $\underline{A}^H_{p(i)}$ is scalar the elementary symmetric polynomials are of the form

$$\begin{aligned}
e_0\left(\underline{A}^H_{p(i)}\right) &= 1 \\
e_1\left(\underline{A}^H_{p(i)}\right) &= \sum_{1 \leq j \leq N_i^H} \underline{A}^H_{p(i),j} \\
e_2\left(\underline{A}^H_{p(i)}\right) &= \sum_{1 \leq j < k \leq N_i^H} \underline{A}^H_{p(i),j}\underline{A}^H_{p(i),k} \\
e_3\left(\underline{A}^H_{p(i)}\right) &= \sum_{1 \leq j < k < l \leq N_i^H} \underline{A}^H_{p(i),j}\underline{A}^H_{p(i),k}\underline{A}^H_{p(i),l} \\
&\vdots \\
e_{N_c}\left(\underline{A}^H_{p(i)}\right) &= \underline{A}^H_{p(i),1}\underline{A}^H_{p(i),2}\underline{A}^H_{p(i),3}\cdots\underline{A}^H_{p(i),N_i^H},
\end{aligned}$$

so that $\tau_{K_H}\left(\underline{A}^H_{p(i)}\right) = \left(e_0\left(\underline{A}^H_{p(i)}\right), e_1\left(\underline{A}^H_{p(i)}\right), \ldots, e_{K_H}\left(\underline{A}^H_{p(i)}\right)\right)'$. Weyl (1946) discusses such polynomials for the multivariate case.

[14] If T_i indexes a manipulable 'treatment' then this assumption, of course, has more content. Our framework can be adapted to this case (see Manski (2010) for an elegant development).

definition of an individual's 'ability' is their rank amongst those of their own type.[15] Let $Y_i = g(T_i, A_i, s_{-i}, \tau_{K_H}(\underline{a}^H_{p(i)}), \tau_{K_L}(\underline{a}^L_{p(i)}), u_i) = Y_i(s_{-i}, \tau_{K_H}(\underline{a}^H_{p(i)}), \tau_{K_L}(\underline{a}^L_{p(i)}), u_i)$ denote the i[th] individual's potential outcome given assignment to a group with fraction $S_{-i} = s_{-i}$ high type peers, peer abilities $\tau_{K_H}(\underline{A}^H_{p(i)}) = \tau_{K_H}(\underline{a}^H_{p(i)})$ and $\tau_{K_L}(\underline{A}^L_{p(i)}) = \tau_{K_L}(\underline{a}^L_{p(i)})$, and location attributes $U_i = u_i$ (here $g(\cdot)$ is the production function notation for the potential outcome function defined in footnote 11 above). Assuming the distribution of A_i does not depend on T_i does not restrict the conditional distribution of $Y_i(s_{-i}, \tau_{K_H}(\underline{a}^H_{p(i)}), \tau_{K_L}(\underline{a}^L_{p(i)}), u_i) \Big| T_i$ so that Assumption 2.3 can be made without loss of generality.

The allocation response function $Y_i(S_{-i}, \tau_{K_H}(\underline{A}^H_{p(i)}), \tau_{K_L}(\underline{A}^L_{p(i)}), U_i)$ defines an individual-specific mapping from peer types, ability, and neighborhood characteristics into outcomes. In our framework the 'treatment' induced by a given allocation is a specific configuration of peers, as summarized by their observed type composition, S_{-i}, and unobserved ability, $\tau_{K_H}(\underline{A}^H_{p(i)})$ and $\tau_{K_L}(\underline{A}^L_{p(i)})$. Residence in a specific location, where specificity is indexed by the vector of unobserved characteristics U_i, is also a feature of the 'treatment'.

The non-observability of $\underline{A}_{p(i)}$ and U_i generates complications, relative to the standard potential outcomes model of causal inference (Neyman, 1990; Rubin, 1974; Holland, 1986), because it implies that we do not observe the full 'treatment'. The *observed* treatment is an assignment to a set of peers with a given type composition. However, because peers and locations are heterogeneous, observationally equivalent assignments may be associated with distinct treatments (and hence potential outcomes). Assumptions 2.1 and 2.2 are not strong enough to ensure that the observed treatment satisfies the homogenous treatment assumption that is part of Rubin's Stable-Unit-Treatment-Value-Assumption (SUTVA) (see Rubin, 1990; Holland, 1986).[16]

To deal with this issue we define an intermediate object: the expected allocation response

[15] Many of our results extend straightforwardly to the case where unnormalized ability is a $J \times 1$ vector $A_i^* = (A_{1i}^*, \ldots, A_{Ji}^*)'$. In that case Assumption 2.3 is imposed by the one-to-one mapping

$$\begin{aligned} A_{1i} &= F\left(A_{1i}^* \middle| T_i\right) \\ A_{2i} &= F\left(A_{2i}^* \middle| A_{1i}^*, T_i\right) \\ &\vdots \\ A_{Ji} &= F\left(A_{Ji}^* \middle| A_{1i}^*, \ldots, A_{J-1i}^*, T_i\right). \end{aligned}$$

[16] In related work Sobel (2006a,b) conceptualizes neighborhood effects as violations of SUTVA.

function. Individual's $i's$ expected allocation response function is given by

$$Y_i^e(s_{-i}) = \int\int\cdots\int Y_i(s_{-i}, \tau_{K_H}(\underline{a}_{p(i)}^H), \tau_{K_L}(\underline{a}_{p(i)}^L), u)\left\{\prod_{j\in p(i)} f_A(a_{p(i),j})\mathrm{d}a_{p(i),j}\right\} f_U(u)\mathrm{d}u. \tag{2}$$

Equation (2) gives an individual's expected outcome when assigned to a group with peer composition $S_{-i} = s_{-i}$ *when groups are formed in a certain way.* The group formation process enters into the definition of $Y_i^e(s_{-i})$ because it is meant to measure the expected effect of exogenous changes in observed peer composition, s_{-i}. For this effect to have a causal interpretation it should be unconfounded by the effects of matching and/or sorting of peers.

Matching occurs if individuals choose (or are assigned to) a location on the basis of its *unobserved* attribute U_c *and* the utility derived from that choice depends on own attributes (T_i, A_i). Matching implies that the vector $(\underline{T}_i, \underline{A}_i)$ of individual peer and own attributes at the location of i is related to the unobserved location characteristic U_i. Hence there is no matching if

$$(\underline{T}_i, \underline{A}_i) \perp U_i,$$

which implies the density factorization

$$f_{\underline{A},U|\underline{T}}(\underline{a}_c, u_c|\underline{t}_c) = f_{\underline{A}|\underline{T}}(\underline{a}_c|\underline{t}_c)f_U(u_c).$$

Sorting is related to the distribution of $\underline{A}_c|\underline{T}_c$. Sorting occurs if, for example, an individual's *unobserved* ability, A_i, is related to those of her peers, $\underline{A}_{p(i)}$. Such a dependence would arise if an individual's preference for a location (or the assignment rule used) depends on the attributes and types of its residents *and* this preference varies systematically with (T_i, A_i). The absence of sorting therefore implies that

$$(\underline{T}_{p(i)}, \underline{A}_{p(i)}) \perp A_i \big| T_i,$$

so that, conditional on own type, own ability does not vary with the type or ability composition of one's peers. No sorting generates the density factorization (see Appendix B)

$$f_{\underline{A}|\underline{T}}(\underline{a}_c|\underline{t}_c) = \prod_{j=1}^{N} f_{A|T}(a_{cj}|t_{cj}) = \prod_{j=1}^{N} f_A(a_{cj}),$$

where the final equality is due to Assumption 2.3 and we use a double subscript notation with j indexing individuals within a group in arbitrary order. Note that sorting, as defined above, does not preclude high types seeking out peer groups composed of many other high types (i.e., sorting on observables is allowed). Consequently the distribution of peer composition across groups is not restricted by the absence of sorting. There is neither matching nor sorting if, for a group of a given type composition, high type members are random draws from the subpopulation of high types, low type members are random draws from the subpopulation of low types, and the group, so formed, is randomly assigned to a specific location.

In the absence of both matching and sorting the joint density of $\underline{A}_c, U_c$ given $\underline{T}_c$ factors into

$$f_{\underline{A},U|\underline{T}}(\underline{a}_c, u_c|\underline{t}_c) = \left\{ \prod_{j=1}^{N} f_A(a_{cj}) \right\} f_U(u_c),$$

which is the product of marginals being integrated over in (2), which defines $Y_i^e(s_{-i})$.

Averaging $Y_i^e(s_{-i})$ over the subpopulations of low and high types gives the type-specific mean allocation response functions

$$m_L^*(s_{-i}) = \mathbb{E}\left[Y_i^e(s_{-i}) \middle| T_i = 0 \right], \qquad m_H^*(s_{-i}) = \mathbb{E}\left[Y_i^e(s_{-i}) \middle| T_i = 1 \right].$$

In what follows it is convenient to instead work with the one-to-one mappings

$$m_L(s) = m_L^*\left(\frac{sN}{N-1} \right), \qquad m_H(s) = m_H^*\left(\frac{sN-1}{N-1} \right) \tag{3}$$

where s is the overall fraction of high types in a group (inclusive of oneself). That is, we let $S_c = \sum_{i=1}^{I_P} \frac{1(G_i=c)T_i}{N}$. denote the fraction of high types in location c. Henceforth we refer to S_c as a location c's group composition.

The type-specific *mean allocation response* functions $m_H(s)$ and $m_L(s)$ feature in each of our estimands. They equal the expected outcome, given exogenous assignment to a group of composition $S = s$, of a randomly selected member of, respectively, the subpopulation of high and low types *if* groups are formed without matching and sorting. Most of our identification results follow directly from identification of $m_H(s)$ and $m_L(s)$.

The overall mean allocation response function is given by the composition weighted average

$$m(s) = sm_H(s) + (1-s)\, m_L(s), \tag{4}$$

which is the expected outcome of a randomly selected member of the population when assigned to a group of composition $S = s$. This function is related to the average structural function of Blundell and Powell (2003). A direct application of their definition would replace the average in (2) with one over the joint distribution of $(\underline{A}_c', \underline{U}_c)'$. Such an average would not be causal in our setting as it would be contaminated by sorting (correlation in ability across group members) and matching (correlation between ability and location quality) (see

Graham, 2008, 2011, 2018; Graham et al., 2018). This is an example of how the presence of heterogeneity from *multiple* individuals (as well as locations) in the production function for *each* individual complicates analysis and requires extra care when defining estimands.

Equation (4) can be viewed as a statistical analog of the deterministic production technology that features prominently in the theoretical public finance literature on multi-community models (e.g., de Bartolome, 1990; Benabou, 1993, 1996; Durlauf, 1996b,c; Becker and Murphy, 2000). In order to provide a clean characterization of locational equilibrium as well as the solution to the social planner's problem, the multi-community literature has generally placed strong *a priori* restrictions on $m(s)$. A typical set of assumptions is that $m_H(s) - m_L(s) > 0$ for all $s \in \mathcal{S}$ and that $\partial^2 m(s)/\partial s^2$ is either positive or negative for all $s \in \mathcal{S}$. Fernández (2003) provides an extensive discussion of the role of these assumptions in this literature. In contrast, other than smoothness assumptions, we leave $m(s)$ (essentially) unrestricted.

Differentiating $m(s)$ with respect to s gives the marginal effect of changes in group composition on group average outcomes:

$$\nabla_s m(s) = p(s) + e(s),$$

where

$$p(s) = m_H(s) - m_L(s), \quad e(s) = s\nabla_s m_H(s) + (1-s)\nabla_s m_L(s).$$

The derivative of $m(s)$ with respect to group composition consists of two parts. The first part, $p(s)$, is the effect of changing group composition on expected outcomes holding spillover strength constant. It is the compositional effect of changing group composition on expected group average outcomes. Irrespective of the presence of social spillovers, average outcomes will often rise because the composition of the group has shifted toward high types. This effect is private, in the sense that it reflects benefits that are entirely confined to the entering high type.

The second component, $e(s)$, measures the spillover or external effect associated with increasing s. The introduction of an additional high type individual into the group creates a spillover which raises outcomes for all individuals in the group. Benabou (1996) and others have emphasized that, since agents do not internalize the second effect when choosing locations, decentralized equilibria may be inefficient.

Our final three main assumptions ensure that $m_H(s)$, $m_L(s)$ and their derivatives, $\nabla_s m_H(s)$ and $\nabla_s m_L(s)$, are *nonparametrically identified*. Nonparametric identification requires imposing strong assumptions on the group formation process. In particular, while we allow for matching and sorting on observables, we rule out the presence of these behaviors on unobservables (see Section 5). This assumption is easiest to justify when the assignment is administratively determined, but under certain information structures it may also hold when the assignment corresponds to a decentralized equilibrium.

We emphasize that semiparametric or parametric identification of $m_H(s)$ and $m_L(s)$ is generally possible under weaker assumptions on the group formation process (e.g., Nesheim, 2002). The trade-off between the identifying power of a priori restrictions on the production technology versus the assignment process is explored more fully in Graham (2011). Different researchers will find different combinations of assumptions appropriate depending on the application at hand. Our application, being based on a randomized experiment, allows us to leave $m_H(s)$ and $m_L(s)$ fully nonparametric. We consequently develop estimation and distribution theory appropriate to this case, although our estimands apply generally.

First we make an assumption on the status quo assignment mechanism. In particular, we assume the absence of matching and sorting on unobservables, as defined above.

Assumption 2.4. (No Matching and Sorting On Unobservables)

$$(\underline{T}_i, \underline{A}_i) \perp U_i, \qquad (\underline{T}_{p(i)}, \underline{A}_{p(i)}) \perp A_i \big| \, T_i.$$

Assumption 2.4 will be satisfied if groups are formed, and locations selected at random, (i.e. under a double randomization scheme). To describe this scheme assume that the social planner first chooses a feasible distribution of group compositions

$$F_S^{\mathrm{sq}}(s),$$

where the 'sq' superscript denotes 'status quo' and the density is across groups (i.e., it describes composition for the population of locations/groups). Feasibility of the status quo (as well as that of any other allocation), requires that it satisfies a restriction. Because the fraction high types p_H is fixed, and all groups are equally-sized, feasibility requires that

$$p_H = \int_0^1 s f_S^{\mathrm{sq}}(s) \mathrm{d}s, \tag{5}$$

where we treat S_c as a continuously-valued random variable (as would be appropriate if the common group size, N, is large).

After choosing a feasible joint distribution for group composition the planner fills high and low type spaces in each group by randomly sampling from the high and low type sub-populations. This ensures, along with Assumption 2.3, satisfaction of the second part of Assumption 2.4. The social groups, so formed, are then randomly assigned to a specific location. Random assignment at this stage ensures that the first part of Assumption 2.4 is satisfied.

As discussed above Assumption 2.4 rules out matching and sorting (on unobservables) (see Graham, 2008, 2011; Graham et al., 2018). It does not, however, restrict the degree of status quo segregation or integration ($F_S^{\mathrm{sq}}(s)$ is unrestricted beyond the requirement of feasibility). Consider the example where locations are schools and $T_i = 1$ for white students and $T_i = 0$ for black students. In that case Assumption 2.4 implies that the ability distribution of blacks is similar across schools regardless of the degree to which they are segregated. Furthermore it requires that unobserved teacher quality is independent of the degree to which a school is segregated. Clearly these are rather strong restrictions outside of explicitly experimental settings. Nevertheless, by initially maintaining Assumption 2.4 in what follows, we are able to develop some results on the effects reallocations in a reasonably straightforward way. In Section 5 we show how the presence of observable location-level attributes may be used to weaken Assumption 2.4.

Our next assumption ensures that the gradients, $\nabla_s m_H(s)$ and $\nabla_s m_L(s)$, are identified.

Assumption 2.5. (Continuous Variation) *If $f_S^{\mathrm{sq}}(s) > 0$ then $f_S^{\mathrm{sq}}(s') > 0$ for all s' in a neighborhood of $s \subset \mathcal{S}$.*

Assumption 2.5 only makes sense if it is legitimate to treat group composition, S_c, 'as if' it were a continuously distributed random variable. Such an approximation requires that the common group size, N, be relatively large. Thus our estimands and estimators are not appropriate for situations where groups are small (e.g., college roommates).

Finally we assume the availability of a random sample of locations.

Assumption 2.6. (Random Sampling) *$\{\underline{Y}_c, \underline{T}_c\}_{c=1}^{C}$ is a random sample of C neighborhoods of $I = CN$ individuals.*

These last three assumptions, as well as the restrictions on each individual's allocation response function implied by Assumptions 2.1 to 2.3, ensure that $m_H(s)$, $m_L(s)$ and their derivatives with respect to s are identified.

Proposition 2.1. *Under Assumptions 2.1 to 2.6 (i) $m_L(s)$ and $m_H(s)$ are identified for all s such that $f_S^{\mathrm{sq}}(s) > 0$ by the conditional expectation functions (CEFs):*

$$\mathbb{E}[Y_i|T_i = 0, S_i = s] = m_L(s), \qquad \mathbb{E}[Y_i|T_i = 1, S_i = s] = m_H(s),$$

and (ii) $\nabla_s m_L(s, n)$ and $\nabla_s m_H(s, n)$ are identified by the derivative of these CEFs with respect to s.

Proof See Appendix B.

We emphasize that Proposition 2.1 does not exhaust the set of all possible identification approaches. While Assumption 2.4 may be satisfied under a locational equilibrium of the type studied by Benabou (1996) and others, this will only be the case under strong informational assumptions and/or a priori restrictions on the production technology. For example if agents do not know A_i, nor observe $\underline{A}_{p(i)}$ and U_c, then all matching/sorting will be on observables as required. Alternatively if the derivative of $Y_i(s_{-i}, \tau_{K_H}(\underline{a}^H_{p(i)}), \tau_{K_L}(\underline{a}^L_{p(i)}), u_i)$ with respect to s_{-i} is the same across all individuals of the same type, then agents will have no incentive to match/sort on unobservables (if preferences depend on own outcomes alone).

An alternative approach to identifying $m_L(s)$ and $m_H(s)$ involves explicitly modelling the sorting/matching process (and hence taking an explicit stand on preferences). Unfortunately this approach will often require strong a priori restrictions on the form of $m_L(s)$ and $m_H(s)$. Nesheim (2002) provides some positive semiparametric identification results (see also Calabrese et al., 2006). Calabrese et al. (2006) and Ferreyra (2007) are examples of empirical peer effects studies that assume that the status quo assignment corresponds to a locational equilibrium.

The estimands we propose below will be of interest irrespective of the means by which $m_L(s)$ and $m_H(s)$ are identified and estimated (although details of our treatment of estimation are specific to the conditions of Proposition 2.1). The relative benefits of leaving $m_L(s)$ and $m_H(s)$ nonparametric, while imposing strong restrictions on the sorting/matching process, as we do here; versus imposing a priori restrictions on $m_L(s)$ and $m_H(s)$, but allowing for more complex sorting/matching processes will vary from application to application.

3 Characterizing the effects of social spillovers

In this section we introduce new estimands which characterize different features of the outcome effects of social spillovers. Prior work on the empirics of social interactions has emphasized testing for their presence and/or measuring their average strength. We therefore begin by proposing a simple measure of average spillover strength. The primary goal of this section, however, is to present summary measures of the effect of local reallocations on the distribution of outcomes. In particular we consider the outcome and inequality effects of a class of reallocations which increase segregation marginally.

3.1 Measuring spillover strength

Manski (1993), Brock and Durlauf (2001), Glaeser and Scheinkman (2003) and Graham (2008) emphasize the notion of a social multiplier or the ratio of the full effect of marginal changes in group composition to the private effect:

$$\frac{\nabla_s m(s)}{p(s)} = 1 + \frac{e(s)}{p(s)}, \quad \text{for } p(s) \neq 0.$$

The social multiplier is an intuitive measure of spillover strength and has the virtue of being unitless. Nevertheless, for simplicity, as well as technical reasons, we instead suggest a direct measure of average spillover strength. Conditional on $S_i = s$ the average external effect is given by $e(s)$. Averaging over individuals gives an overall *average spillover effect* (ASE)

$$\beta^{\text{ase}} = \mathbb{E}\left[d_\kappa(S_i)\, e(S_i)\right] = \mathbb{E}\left[d_\kappa(S_i)\left\{S_i \nabla_s m_H(S_i) + (1 - S_i)\nabla_s m_L(S_i)\right\}\right], \tag{6}$$

where $d_\kappa(s)$ is a fixed trimming function that gives zero weight to values of $e(s)$ near the boundary of the support of S, specifically,

$$d_\kappa(s) = \mathbf{1}(s > \underline{s} + \kappa)\,\mathbf{1}(s < \overline{s} - \kappa), \qquad \kappa \subset \mathcal{S} = \left[0, \frac{\underline{s} + \overline{s}}{2}\right).$$

The introduction of fixed trimming into the definition of β^{ase} is somewhat awkward, but is required to ensure that (i) the semiparametric efficiency bound for β^{ase} is non-zero and (ii) to avoid boundary bias problems associated with nonparametric estimation of $m_H(s)$ and $m_L(s)$ (see Newey and McFadden, 1994). We note that (6) is closely-related to the weighted average derivative estimand studied in Powell et al. (1989) and Newey and Stoker (1994) (among others); an observation that aided in undertaking the influence function calculations presented below.

Equation (6) equals the mean external effect, or spillover benefit, of an unit increase in the fraction of high type individuals in each group. Identification of β^{ase} follows directly from Proposition 2.1 and random sampling. While it is easy to construct examples where the outcome effects of reallocations are nontrivial even if $\beta^{\text{ase}} = 0$ (and vice versa), it is nevertheless a simple summary measure of spillover strength; being a nonparametric generalization of the target estimand of a large empirical literature (e.g., Coleman, 1966; Mayer and Jencks, 1989; Solon, 1999; Angrist and Lang, 2004; Ciccone and Peri, 2006; Graham, 2008). While β^{ase} is arguably of scientific interest it does not, since the peer structure of all individuals cannot be simultaneously improved, measure the effects of an implementable policy.

3.2 Measuring the effects of reallocations

The average spillover effect measures the outcome benefit of an infeasible increase in the population frequency of high types. In contrast reallocations of individuals across groups, since they leave the population type distribution unchanged, are, at least in principle, implementable policies. Before considering the effects of a reallocation of individuals across groups, we define the general class of reallocations under consideration. We assume that the social planner, or allocating agency, observes each individual's type, T_i and initial assignment (i.e., the planner observes $F_S^{\mathrm{sq}}(s)$, the distribution of S_c under the status quo, and knows $p_H \overset{def}{\equiv} \int_0^1 s f_S^{\mathrm{sq}}(s)\,\mathrm{d}s$). The planner also knows the high- and low-type mean allocation response functions $m_H(s)$ and $m_L(s)$. The planner does not observe A_i or U_c (or is institutionally constrained to not act on this knowledge).

We consider reallocations obeying the feasibility constraint

$$\int_0^1 s f_S^{\mathrm{r}}(s)\,\mathrm{d}s = p_H, \tag{7}$$

where $f_S^{\mathrm{r}}(s)$ is a valid probability density function. Equation (7) says that $F_S^{\mathrm{r}}(s)$ cannot imply an augmentation of resources, in this case the population frequency of high types. The set of reallocations satisfying condition (7) is very large. In Section 6 we characterize average outcome-maximizing reallocations. Here we consider estimands which characterize the effects of a specific class of local reallocations.

Our local reallocation estimands measure the effects of a particular parameterization of a small, segregation increasing (relative to the status quo), reallocation. Specifically they give the sign of a small such increase in segregation on average outcomes and inter-type inequality.

The reallocation density we consider takes the form

$$f_S^{\mathrm{r}}(s;\lambda,\kappa) = \frac{s}{1+\lambda d_\kappa(s)} f_S^{\mathrm{sq}}\left(\frac{s+\lambda d_\kappa(s)\,p_{H,\kappa}}{1+\lambda d_\kappa(s)}\right), \tag{8}$$

where $p_{H,\kappa} = \mathbb{E}\left[T_i \middle| d_\kappa(S_i) = 1\right]$ is the trimmed population frequency of high types (i.e., the frequency of high types with status quo assignments to groups with group compositions in the interior of $\mathcal{S}$). Appendix B demonstrates that (8) is a feasible reallocation.

Implementing the allocation defined by (8) is equivalent to altering the composition of the c^{th} group according to the rule

$$S_c^{\mathrm{r}} = S_c + \lambda d_\kappa(S_c)\left(S_c - p_{H,\kappa}\right), \tag{9}$$

so that (8) is effectively a mean-preserving spread of $F_S^{\text{sq}}(s)$ when $\lambda > 0$. For $\lambda > 0$ (8) increases segregation across those groups with status quo compositions, S_c, within the interval from $\underline{s} + \kappa$ to $\overline{s} - \kappa$. It leaves group composition unchanged across those groups that are initially highly segregated such that $S_c \leq \underline{s} + \kappa$ or $S_c > \overline{s} - \kappa$. Implementing (8) involves moving high type individuals from groups where the fraction of high types is below their trimmed population frequency ($S_c < p_{H,\kappa}$), to groups where it is above that frequency ($S_c > p_{H,\kappa}$). Such moves are accommodated by switching each high type with a corresponding low type individual. Highly segregated group compositions are left unchanged by (8) to (i) ensure feasibility (it is difficult to increase segregation in a group that is already very segregated) and (ii) for technical reasons. We assume that λ is small enough, or equivalently, κ large enough, to ensure that $S_c^{\text{r}} \in [0, 1]$ for all groups.

From (9) average outcomes after an segregation increasing reallocation are given by

$$\mathbb{E}\left[m\left(S_i^{\text{r}}\right)\right] = \mathbb{E}\left[m\left(S_i + \lambda d_\kappa\left(S_i\right)\left(S_i - p_{H,\kappa}\right)\right)\right].$$

We are interested in the direction of the effect of implementing (8) on average outcomes when $\lambda \to 0$. This corresponds to a small increase in segregation. Differentiating the above expression with respect to λ and evaluating at $\lambda = 0$ gives the desired *local segregation outcome effect* (LSOE):

$$\beta^{\text{lsoe}} = \mathbb{E}\left[d_\kappa\left(S_i\right)\nabla_s m\left(S_i\right)\left(S_i - p_{H,\kappa}\right)\right] = \pi_\kappa \mathbb{C}\left(\nabla_s m\left(S_i\right), S_i \middle| d_\kappa\left(S_i\right) = 1\right), \tag{10}$$

with $\pi_\kappa = \Pr\left(d_\kappa\left(S_i\right) = 1\right)$. Here $\mathbb{V}(Y)$ notes the variance of Y and $\mathbb{C}(X, Y)$ is covariance with X.

Equation (10) is an intuitive condition. If groups where the fraction of high type agents exceeds the trimmed population mean ($S_c > p_{H,\kappa}$) tend also to be relatively responsive to changes in s (i.e., $\nabla_s m(S_c)$ is larger than average), then reallocations that reinforce any existing segregation across groups will tend to raise average outcomes. In contrast, if groups with a low fraction of high type agents are very responsive to changes in s, then reallocations that reinforce existing segregation will tend to lower average outcomes.

To highlight the structure of β^{lsoe}, and connect it to theoretical work on neighborhood sorting, it is helpful to consider the decomposition

$$\beta^{\text{lsoe}} = \alpha^{\text{lppe}} + \alpha^{\text{lepe}},$$

where

$$\alpha^{\text{lppe}} = \pi_{\kappa}\mathbb{C}\left(p\left(S_i\right), S_i\right| d_{\kappa}\left(S_i\right) = 1\left.\right), \quad \alpha^{\text{lepe}} = \pi_{\kappa}\mathbb{C}\left(e\left(S_i\right), S_i\right| d_{\kappa}\left(S_i\right) = 1\left.\right).$$

Under the current setup, local reallocations may alter population average outcomes for two distinct reasons. First, peer quality changes for those individuals who change groups as part of the reallocation, called 'movers'. This is an internalizeable or private peer effect. Second, peer quality changes for those individuals who do not switch groups as part of the reallocation, called 'stayers', we call this the external peer effect.

First, consider the private peer effect. If the benefits of improved peer quality for high type movers entering groups with an initially above average fraction of high types exceed the costs for low type movers leaving such groups, then implementing (8) will tend to raise the average achievement of movers. Observe that the private peer effect will be zero when outcomes are separable in own and peer types (as is often assumed in empirical work), positive when they are complementary (as is typically assumed in theoretical work on sorting) and negative when they are substitutable. The sign of the private effect on average outcomes is captured by α^{lppe}. Positivity of α^{lppe} suggests the presence of private incentives for further, segregating-increasing, sorting.

Second consider the external peer effect. This term captures changes in average outcomes operating through the reallocation's effect on average spillover strength. If the marginal benefit of an additional high type on stayers is greater in groups with a large fraction of high types (i.e., $\alpha^{\text{lepe}} > 0$), then increased segregation will raise average outcomes by raising average spillover strength. This term is only non-zero in the presence of some form of social spillover. The sign of α^{lepe} determines the direction of the external effect associated with implementing (8). This effect is not internalized by individuals as they negotiate switches in group membership.

The next theorem makes the above statements more precise and explicitly connects β^{lsoe} to the theoretical work on segregation and efficiency done by de Bartolome (1990), Benabou (1993, 1996), Becker and Murphy (2000) and others.

Theorem 3.1. *Under Assumptions 2.1 to 2.6* $\beta^{\text{lsoe}} = \alpha^{\text{lppe}} + \alpha^{\text{lepe}}$ *with (i)*

$$\alpha^{\text{lppe}} = \pi_{\kappa}\mathbb{V}\left(S_i\right| d_{\kappa}\left(S_i\right) = 1\left.\right) \times \mathbb{E}\left[\omega(S_i)\left\{\nabla_s m_H\left(S_i\right) - \nabla_s m_L\left(S_i\right)\right\}\right| d_{\kappa}\left(S_i\right) = 1\left.\right]$$

$$\begin{aligned}\alpha^{\text{lepe}} &= \pi_{\kappa}\mathbb{V}\left(S_i\right| d_{\kappa}\left(S_i\right) = 1\left.\right) \\ &\times \mathbb{E}\left[\omega(S_i)\left\{\nabla_s m_H\left(S_i\right) - \nabla_s m_L\left(S_i\right) + S_i \nabla_{ss} m_H\left(S_i\right) + \left(1 - S_i\right)\nabla_{ss} m_L\left(S_i\right)\right\}\right| d_{\kappa}\left(S_i\right) = 1\left.\right]\right],\end{aligned}$$

where $\mathbb{E}\left[\omega(S_i)\right] = 1$ *with*

$$\omega\left(s\right) = \frac{1}{f_{S|d_\kappa(S)}\left(s|\, d_\kappa\left(S_i\right) = 1\right)} \times \frac{\mathbb{E}\left[S_i - p_{H,\kappa}|\, S_i > s, d_\kappa\left(S_i\right) = 1\right]\left(1 - F_{S|d_\kappa(S)}\left(s|\, d_\kappa\left(S_i\right) = 1\right)\right)}{\int_{v=0}^{v=1} \mathbb{E}\left[S_i - p_{H,\kappa}|\, S_i > v, d_\kappa\left(S_i\right) = 1\right]\left(1 - F_{S|d_\kappa(S)}\left(v|\, d_\kappa\left(S_i\right) = 1\right)\right) \mathrm{d}v},$$

and (ii) the averages α^{lppe} *and* α^{lepe} *give maximal weight to values at* $S = p_{H,\kappa}$ *and minimal weight to those at* $S = \underline{s} + \kappa$ *and* $S = \overline{s} - \kappa$*.*

Proof See Appendix B.

Theorem 3.1 provides a mathematical representation of the private and external effects discussed above. Theorem 3.1 implies that a small increase in segregation raises average outcomes if

$$\begin{aligned} &2\mathbb{E}\left[\omega(S_i)\left\{\nabla_s m_H\left(S_i\right) - \nabla_s m_L\left(S_i\right)\right\}|\, d_\kappa\left(S_i\right) = 1\right] \\ &+ \mathbb{E}\left[\omega(S_i)\left\{S_i \nabla_{ss} m_H\left(S_i\right) + \left(1 - S_i\right)\nabla_{ss} m_L\left(S_i\right)\right\}|\, d_\kappa\left(S_i\right) = 1\right] \end{aligned} \quad (11)$$

is greater than zero. The two terms in the above expression, to use the language of Benabou (1996), are respectively weighted averages of the degree of *complementarity* and *curvature.* They are *local statistical* analogs of identically named *global deterministic* objects discussed by Benabou (1996), Fernández (2003) and others.

Theoretical work has generally assumed that $\nabla_s m_H\left(s\right) - \nabla_s m_L\left(s\right) > 0$ for all $s \in (0, 1)$ or that own and peers' type are *global* complements. Global complementarity ensures that high type residents will *always* benefit more from improvements in peer quality than their low type neighbors. While the empirical evidence for such a strong form of complementarity is mixed, theoretical work nevertheless takes it as a primitive since it induces equilibrium stratification.[17]

Theorem 3.1 indicates that a measure of *local average complementarity,*

$$\mathbb{E}\left[\omega\left(S_i\right)\left\{\nabla_s m_H\left(S_i\right) - \nabla_s m_L\left(S_i\right)\right\}|\, d_\kappa\left(S_i\right) = 1\right],$$

is important for determining whether small increases in segregation raise the average outcome. If, *in the neighborhood of* $s = p_{H,\kappa}$, own and peers' type *tend to be* complementary, then the first term in (11) will be positive. This is a 'force' in favor of a local increases in segregation being outcome-raising. It is also suggestive of the existence of incentives for

[17] If the marginal benefit of an additional high type is greater for high types than it is for low types, then high types will be willing to pay more to live in high quality neighborhoods in equilibrium

further segregation relative to the status quo.

The theory literature also discusses the importance of curvature for determining whether segregation is outcome-maximizing. Curvature, equal to $s\nabla_{ss}m_H(s) + (1-s)\nabla_{ss}m_L(s)$, determines whether there are diminishing returns to peer quality at the neighborhood level. Theoretical work emphasizes the case where curvature is such that $2\{\nabla_s m_H(s) - \nabla_s m_L(s)\}$ $+ s\nabla_{ss}m_H(s) + (1-s)\nabla_{ss}m_L(s)$ is negative for all $s \in \mathcal{S}$ (i.e., global concavity of $m(s)$ in group composition). In that case complementarity of own and peer quality induces equilibrium segregation, but such segregation is inefficient in the sense that it does not maximize average outcomes (see Benabou, 1996, Proposition 7). In such a situation, within a neighborhood high types always benefit more from improvements in peer quality than do low types, while across neighborhoods areas with few high types benefit more from increases in peer quality than do areas with many high types. This situation, where the private and social incentives for sorting are misaligned has been emphasized by Benabou (1993, 1996) and others.

Theorem 3.1 indicates that a measure of *local average curvature,*

$$\mathbb{E}\left[\omega(S_i)\{S_i\nabla_{ss}m_H(S_i) + (1-S_i)\nabla_{ss}m_L(S_i)\}\middle| d_\kappa(S_i) = 1\right],$$

is important for determining whether segregation is outcome raising in the current context as well. If, again in the neighborhood of $s = p_{H,\kappa}$, the marginal benefit of an additional high type peer tends to decline more with s for high relative to low types, then the second term in (11) will be negative.

To summarize Theorem 3.1 indicates that the average outcome effects of small increases in segregation depend on the relative magnitudes of *local average complementarity* and *local average curvaturve.* These are statistical analogs of well-known deterministic objects from the multi-community models literature. The novelty here, besides the introduction of statistical content, is that the interpretation of β^{lsoe} does not depend on a priori restrictions on $m(s)$. The cost of such flexibility is that β^{lsoe} provides only local information about the relative average outcome effects of segregation versus integration.

The LSOE provides an indication of the likely effects of small increases in segregation on average outcomes. A longstanding concern of the literature on segregation, however, is the potential for an equity versus efficiency trade-off. Even if increases in segregation raise average outcomes, such efficiency gains may be unacceptable if they increase inequality across groups. On the other hand, reallocations which both reduce inter-type inequality and raise average outcomes are especially compelling.

Our next estimand measures the sign of the change in the high-low outcome gap associated with a segregation-increasing reallocation. This object, the *local segregation inequality effect* (LSIE), along with the LSOE defined above, allows one to test for the presence of a local equity-efficiency trade-off.

The average outcome of a high type individual under the status quo is given by, using iterated expectations,

$$\mathbb{E}\left[m_H\left(S_i\right)\middle| T_i=1\right]=\mathbb{E}\left[\frac{T_i m_H\left(S_i\right)}{p_H}\right]=\mathbb{E}\left[\frac{S_i}{p_H}m_H\left(S_i\right)\right],$$

with a similar expression holding for low types. Therefore, after reallocation the high-low outcome gap is given by

$$\begin{aligned}\mathbb{E}&\left[\frac{S_i^{\mathrm{r}}}{p_H}m_H\left(S_i^{\mathrm{r}}\right)\right]-\mathbb{E}\left[\frac{1-S_i^{\mathrm{r}}}{1-p_H}m_L\left(S_i^{\mathrm{r}}\right)\right]=\\&\mathbb{E}\left[\frac{\left(S_i+\lambda d_\kappa\left(S_i\right)\left(S_i-p_{H,\kappa}\right)\right)}{p_H}m_H\left(S_i+\lambda d_\kappa\left(S_i\right)\left(S_i-p_{H,\kappa}\right)\right)\right]\\&\quad-\mathbb{E}\left[\frac{\left(1-S_i-\lambda d_\kappa\left(S_i\right)\left(S_i-p_{H,\kappa}\right)\right)}{1-p_H}m_L\left(S_i+\lambda d_\kappa\left(S_i\right)\left(S_i-p_{H,\kappa}\right)\right)\right].\end{aligned}$$

Differentiating with respect to λ and evaluating at $\lambda=0$ gives a local segregation inequality effect of, or the sign of the reallocation's effect on the high versus low type average outcome gap equal to,

$$\begin{aligned}\beta^{\mathrm{lsie}}=&\ \mathbb{E}\left[\frac{d_\kappa\left(S_i\right)}{p_H}\left\{m_H\left(S_i\right)+S_i\nabla_s m_H\left(S_i\right)\right\}\left(S_i-p_{H,\kappa}\right)\right]\\&-\mathbb{E}\left[\frac{d_\kappa\left(S_i\right)}{1-p_H}\left\{-m_L\left(S_i\right)+\left(1-S_i\right)\nabla_s m_L\left(S_i\right)\right\}\left(S_i-p_{H,\kappa}\right)\right].\end{aligned}\tag{12}$$

4 Estimation

Our approach to estimation of β^{ase}, β^{lsoe} and β^{lsie} involves forming sample analogs of the right-hand-sides of, respectively, (6), (10) and (12) above. In order to do this we must replace $m_H\left(s\right)$, $m_L\left(s\right)$ and/or their derivatives with estimates (along with replacing $p_{H,\kappa}$ and, for the case of β^{lsie}, p_H with estimates). We propose to use kernel smoothing methods to estimate each of these objects.

Let $\mathcal{K}\left(u\right)$ denote a kernel function that integrates to one and satisfies other conditions. Define $K_b(s-S_i)=b^{-1}\mathcal{K}((s-S_i)/b)$. Our estimates of $m_H\left(s\right)$ and $m_L\left(s\right)$ are given by

$$\widehat{m}_H\left(s\right)=\frac{\widehat{g}_{1H}\left(s\right)}{\widehat{g}_{2H}\left(s\right)},\quad\widehat{m}_L\left(s\right)=\frac{\widehat{g}_{1L}\left(s\right)}{\widehat{g}_{2L}\left(s\right)}\tag{13}$$

where

$$\hat{g}_{1H}(s) = \frac{1}{I_1}\sum_{i=1}^{I_1} K_b(s-S_i)Y_i, \qquad \hat{g}_{1L}(s) = \frac{1}{I_0}\sum_{i=I_1+1}^{I} K_b(s-S_i)Y_i,$$
$$\hat{g}_{2H}(s) = \frac{1}{I_1}\sum_{i=1}^{I_1} K_b(s-S_i), \qquad \hat{g}_{2L}(s) = \frac{1}{I_0}\sum_{i=I_1+1}^{I} K_b(s-S_i).$$

We assume that the sample is ordered so that the I_1 high types appear first followed by the $I_0 = I - I_1$ low types.

We estimate the derivatives of $m_H(s)$ and $m_L(s)$ by the derivatives of their estimates:

$$\nabla_s \widehat{m}_H(s) = \frac{1}{\widehat{g}_{2H}(s)}\left[\nabla_s \widehat{g}_{1H}(s) - \nabla_s \widehat{g}_{2H}(s)\, \widehat{m}_H(s)\right] \tag{14}$$
$$\nabla_s \widehat{m}_L(s) = \frac{1}{\widehat{g}_{2L}(s)}\left[\nabla_s \widehat{g}_{1L}(s) - \nabla_s \widehat{g}_{2L}(s)\, \widehat{m}_L(s)\right].$$

Finally we estimate $p_{H,\kappa}$ and p_H by

$$\widehat{p}_{H,\kappa} = \frac{\frac{1}{I}\sum_{i=1}^{I} d_\kappa(S_i)\, T_i}{\frac{1}{I}\sum_{i=1}^{I} d_\kappa(S_i)}, \qquad \widehat{p}_H = \frac{1}{I}\sum_{i=1}^{I} T_i. \tag{15}$$

We begin by describing our average spillover effect estimator, which is

$$\widehat{\beta}^{\text{ase}} = \frac{1}{I}\sum_{i=1}^{I} d_\kappa(S_i)\left\{S_i \nabla_s \widehat{m}_H(S_i) + (1-S_i)\nabla_s \widehat{m}_L(S_i)\right\}.$$

The next proposition characterizes the large sample properties of $\widehat{\beta}^{\text{ase}}$.

Proposition 4.1. *Under regularity conditions $\widehat{\beta}^{\text{ase}}$ is $\sqrt{C}$ consistent with an asymptotic sampling distribution of*

$$\sqrt{C}\left(\widehat{\beta}^{\text{ase}} - \beta^{\text{ase}}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c \widetilde{\phi}_c\right]\right),$$

where, $\widetilde{\phi}_c = \sum_{i\in\{i:G_i=c\}} \phi(Z_i)$, and $\phi(Z_i)$, the efficient influence function, is given by

$$\phi(Z_i) = \frac{d_\kappa(S_i)}{N}\Bigg\{e(S_i) - \beta^{\text{ase}} - \frac{\nabla_s f_S(S_i)}{f_S(S_i)}(Y_i - m(S_i)) - \left(\left[\frac{T_i Y_i}{S_i} - \frac{(1-T_i)Y_i}{1-S_i}\right] - \left[m_H(S_i) - m_L(S_i)\right]\right)\Bigg\}.$$

Proof *See the Supplemental Web Appendix.*

Observe that the asymptotic variance formula $\widehat{\beta}^{\text{ase}}$ is of the 'clustered' variety. Independence of outcomes holds across groups but not within them due to the presence of unobserved locational heterogeneity, U_c.[18] The form of the influence function is also instructive. The first term would be the influence function if $e(s)$ we known. The second two terms therefore capture the effects of first-step nonparametric estimation of $e(s)$. Of these two terms the first is identical to the correction term associated with semiparametric average derivative estimation (e.g., Härdle and Stoker, 1989; Powell et al., 1989; Newey and McFadden, 1994). This follows from re-expressing the estimand as the difference

$$\beta^{\text{ase}} = \mathbb{E}\left[d_\kappa(S_i)\nabla_s m(S_i)\right] - \mathbb{E}\left[d_\kappa(S_i)\left\{m_H(S_i) - m_L(S_i)\right\}\right].$$

Thus the first of the two correction terms captures the sampling uncertainty from having to estimate $\nabla_s m(S_i)$, while the second is due to sampling error in the estimate of the difference $m_H(S_i) - m_L(S_i)$.

The Supplemental Web Appendix derives the form of $\phi(Z_i)$ using the methods described by Newey (1994a). It does not provide primitive conditions for $\sqrt{C}$ consistency and asymptotic normality. This can be done along the lines of Newey and McFadden (1994, Section 8). Here we make only a few comments that are particular to our problem. First, the weight function $d_\kappa(S_i)$ serves two distinct purposes. First, it ensures that the product $d_\kappa(s) f_S(s)$ is zero on the boundary of the support of S. The pathwise derivative calculations in the Supplemental Web Appendix make clear that such a condition is required for the semiparametric variance bound to be finite. Analogous weight functions play a similar role in average derivative estimation as elegantly explained in Newey and Stoker (1994, p. 1206). A second concern is boundary bias in our first step estimates $\nabla_s \widehat{m}_H(s)$ and $\nabla_s \widehat{m}_L(s)$. Eliminating such bias is required for the remainder term from linearization (of our second step moment) to be small. The $d_\kappa(s)$ weight effectively eliminates this problem by requiring us to only estimate $\nabla_s \widehat{m}_H(s)$ and $\nabla_s \widehat{m}_L(s)$ on the interior of the support of S. As is usual in semiparametric estimation, higher order kernels are required for bias reduction, although the use of such kernels in practice may be ill-advised.

Estimation of β^{lsoe} parallels that of β^{ase}. Using the first step estimates defined in (13), (14) and (15) above we form the sample analog of (10):

$$\widehat{\beta}^{\text{lsoe}} = \frac{1}{I}\sum_{i=1}^{I} d_\kappa(S_i)\left[\widehat{m}_H(S) - \widehat{m}_L(S_i) + S_i \nabla_s \widehat{m}_H(S_i) + (1 - S_i)\nabla_s \widehat{m}_L(S_i,)\right](S_i - \widehat{p}_{H,\kappa}).$$

[18] Newey (1994a, p. 1367) notes that dependence of this type does not affect the form of the efficient influence function.

Proposition 4.2. *Under regularity conditions* $\widehat{\beta}^{\text{lsoe}}$ *is* $\sqrt{C}$ *consistent with an asymptotic sampling distribution of*

$$\sqrt{C}\left(\widehat{\beta}^{\text{lsoe}} - \beta^{\text{lsoe}}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c\widetilde{\phi}_c\right]\right),$$

where, $\widetilde{\phi}_c = \sum_{i\in\{i:G_i=c\}} \phi\left(Z_i\right)$, *and* $\phi\left(Z_i\right)$, *the efficient influence function, is given by*

$$\begin{aligned}\phi\left(Z_i\right) = \frac{d_\kappa\left(S_i\right)}{N}&\left\{\nabla_s m\left(S_i\right)\left(S_i - p_{H,\kappa}\right) - \beta^{\text{lsoe}}\right.\\ &- \frac{\nabla_s f_S\left(S_i\right)}{f_S\left(S_i\right)}\left(Y_i - m\left(S_i\right)\right)\left(S_i - p_{H,\kappa}\right) - d_\kappa\left(S_i\right)\left[Y_i - m\left(S_i\right)\right]\\ &\left.-\mathbb{E}\left[\nabla_s m\left(S_i\right)\middle| d_\kappa\left(S_i\right) = 1\right]\left(T_i - p_{H,\kappa}\right)\right\}.\end{aligned}$$

Proof *See the Supplemental Web Appendix.*

As discussed in Section 3 above it is interesting to decompose β^{lsoe} into is private (mover), α^{lppe}, and spillover (stayer) components, α^{lepe}. These components may be estimated by

$$\begin{aligned}\widehat{\alpha}^{\text{lppe}} &= \frac{1}{I}\sum_{i=1}^{I} d_\kappa\left(S_i\right)\left[\widehat{m}_H\left(S_i\right) - \widehat{m}_L\left(S_i\right)\right]\left(S_i - \widehat{p}_{H,\kappa}\right)\\ \widehat{\alpha}^{\text{lepe}} &= \frac{1}{I}\sum_{i=1}^{I} d_\kappa\left(S_i\right)\left[S_i\nabla_s\widehat{m}_H\left(S_i\right) + \left(1 - S_i\right)\nabla_s\widehat{m}_L\left(S_i\right)\right]\left(S_i - \widehat{p}_{H,\kappa}\right).\end{aligned}$$

The next two propositions characterizes the large sample properties of these estimators.

Proposition 4.3. *Under regularity conditions* $\widehat{\alpha}^{\text{lppe}}$ *is* $\sqrt{C}$ *consistent with an asymptotic sampling distribution of*

$$\sqrt{C}\left(\widehat{\alpha}^{\text{lppe}} - \alpha^{\text{lppe}}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c\widetilde{\phi}_c\right]\right),$$

where, $\widetilde{\phi}_c = \sum_{i\in\{i:G_i=c\}} \phi\left(Z_i\right)$, *and* $\phi\left(Z_i\right)$, *the efficient influence function, is given by*

$$\begin{aligned}\phi\left(Z_i\right) = \frac{d_\kappa\left(S_i\right)}{N}&\left\{p\left(S_i\right)\left(S_i - p_{H,\kappa}\right) - \alpha^{\text{lppe}}\right.\\ &+\left\{\left(\frac{T_i}{S_i}\right)Y_i - m_H\left(S\right)\right\}\left(S_i - p_{H,\kappa}\right) - \left\{\left(\frac{1-T_i}{1-S_i}\right)Y_i - m_L\left(S_i\right)\right\}\left(S_i - p_{H,\kappa}\right)\\ &\left.-\mathbb{E}\left[p\left(S_i\right)\middle| d_\kappa\left(S_i\right) = 1\right]\left(T_i - p_{H,\kappa}\right)\right\}.\end{aligned}$$

Proof *See the Supplemental Web Appendix.*

Proposition 4.4. *Under regularity conditions* $\widehat{\alpha}^{\text{lepe}}$ *is* $\sqrt{C}$ *consistent with an asymptotic sampling distribution of*

$$\sqrt{C}\left(\widehat{\alpha}^{\text{lepe}} - \alpha^{\text{lepe}}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c \widetilde{\phi}_c\right]\right),$$

where, $\widetilde{\phi}_c = \sum_{i \in \{i: G_i = c\}} \phi(Z_i)$, *and* $\phi(Z_i)$, *the efficient influence function, is given by*

$$\begin{aligned}
\phi(Z_i) = \frac{d_\kappa(S_i)}{N} \Big\{ & e(S_i)(S_i - p_H) - \alpha^{\text{lepe}} \\
& - \frac{\nabla_s f_S(S_i)}{f_S(S_i)}(Y_i - m(S_i))(S_i - p_{H,\kappa}) - [Y_i - m(S_i)] \\
& - \left\{\left(\frac{T_i}{S_i}\right) Y_i - m_H(S_i)\right\}(S_i - p_{H,\kappa}) + \left\{\left(\frac{1-T_i}{1-S_i}\right) Y_i - m_L(S_i)\right\}(S_i - p_{H,\kappa}) \\
& - \mathbb{E}\left[e(S_i) | d_\kappa(S_i) = 1\right](T_i - p_{H,\kappa}) \Big\}.
\end{aligned}$$

Proof *See the Supplemental Web Appendix.*

Note that the sum of the influence functions for $\widehat{\alpha}^{\text{lppe}}$ and $\widehat{\alpha}^{\text{lepe}}$ equal that of $\widehat{\beta}^{\text{lsoe}}$.

Finally our estimate of β^{lise}, the effect of a small increase in segregation on the high-low outcome gap, is given by

$$\begin{aligned}
\widehat{\beta}^{\text{lise}} = & \frac{1}{I}\sum_{i=1}^{I} \frac{d_\kappa(S_i)}{\widehat{p}_H}\left\{\widehat{m}_H(S_i) + S_i \nabla_s \widehat{m}_H(S_i)\right\}(S_i - \widehat{p}_{H,\kappa}) \\
& - \frac{1}{I}\sum_{i=1}^{I} \frac{d_\kappa(S_i)}{1-\widehat{p}_H}\left\{-\widehat{m}_L(S_i) + (1-S_i)\nabla_s \widehat{m}_L(S_i)\right\}(S_i - \widehat{p}_{H,\kappa}).
\end{aligned}$$

Proposition 4.5. *Under regularity conditions* $\widehat{\beta}^{\text{lise}}$ *is* $\sqrt{C}$ *consistent with an asymptotic sampling distribution of*

$$\sqrt{C}\left(\widehat{\beta}^{\text{lise}} - \beta^{\text{lise}}\right) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c \widetilde{\phi}_c\right]\right),$$

where, $\widetilde{\phi}_c = \sum_{i \in \{i : G_i = c\}} \phi(Z_i)$, *and* $\phi(Z_i)$, *the efficient influence function, is given by*

$$\begin{aligned}
\phi(Z_i) = \frac{d_\kappa(S_i)}{N} \Bigg\{ & \frac{1}{p_H} \{ m_H(S_i) + S_i \nabla_s m_H(S_i) \} (S_i - p_{H,\kappa}) \\
& - \frac{1}{1 - p_H} \{ -m_L(S_i) + (1 - S_i) \nabla_s m_L(S_i) \} (S_i - p_{H,\kappa}) - \beta^{\text{lise}} \\
& - \frac{1}{p_{H,\kappa}} \frac{\nabla_s f_S(S_i)}{f_S(S_i)} (T_i Y_i - S_i m_H(S_i)) (S_i - p_{H,\kappa}) \\
& + \frac{1}{1 - p_{H,\kappa}} \frac{\nabla_s f_S(S_i)}{f_S(S_i)} ((1 - T_i) Y_i - (1 - S_i) m_L(S_i)) (S_i - p_{H,\kappa}) \\
& - \frac{1}{p_{H,\kappa}} (T_i Y_i - S_i m_H(S_i)) + \frac{1}{1 - p_{H,\kappa}} ((1 - T_i) Y_i - (1 - S_i) m_L(S_i)) \\
& - \frac{1}{p_{H,\kappa}} \mathbb{E} \left[\frac{S}{p_{H,\kappa}} [m_H(S) + S \nabla_s m_H(S)] \middle| d_\kappa(S_i) = 1 \right] (T_i - p_{H,\kappa}) \\
& + \frac{1}{1 - p_{H,\kappa}} \mathbb{E} \left[\frac{1 - S}{1 - p_{H,\kappa}} [-m_L(S) + (1 - S) \nabla_s m_L(S)] \middle| d_\kappa(S_i) = 1 \right] (T_i - p_{H,\kappa}) \Bigg\}.
\end{aligned}$$

Proof *See the Supplemental Web Appendix.*

5 Incorporating additional covariates

The identification and estimation results presented so far maintain strong assumptions on the form of the status quo assignment. In this section we briefly discuss how the availability of individual- and location-level covariates may be used to accommodate richer patterns of matching and sorting in the status quo. Let W_i and X_c respectively denote vectors of observed individual- and location-level covariates (e.g., student and teacher characteristics or class size). We replace Assumptions 2.3 and 2.4 with the conditional analogs:

Assumption 5.1. (Inclusive Definition of Type) $T_i \perp A_i | W_i, X_i$.

Assumption 5.2. (No Matching and Sorting On Unobservables)

$$(\underline{T}_i, \underline{A}_i) \perp U_i | \underline{W}_i, X_i, \qquad (\underline{T}_{p(i)}, \underline{W}_{p(i)}, \underline{A}_{p(i)}) \perp A_i \big| T_i, W_i, X_i.$$

Assumption 5.1 establishes a different normalization for unobserved ability: we now conceptualize ‘ability’ as one’s rank within the subpopulation of individuals homogenous in type, T_i, other observed individual attributes, W_i, and observed location characteristics, X_i.

Assumption 5.2 substantively weakens the requirements placed on the status quo allocation. The first part of the assumption implies that location-specific unobservables, U_c, vary independently of the type and ability structure of a group $(\underline{A}_c, \underline{T}_c)$. This independence, however, now needs to hold only conditionally (on group members' observed characteristics, $\underline{W}_c$, and location-specific characteristics, X_c). This assumption rules out matching on unobservables, whereby groups with particular type and ability structures are able to systematically secure locations with particular unobserved characteristics. It does allow observed group member and location attributes (i.e., $\underline{W}_c$ and X_c) to covary with unobserved location quality. That is, it allows agents to match on observables.

The second part of the assumption implies that conditional on own- and location-specific observables each agent's ability is independent of the abilities, types and other characteristics of their peers. Conditional on own observed characteristics, individuals with higher ability, for example, are not able to sort into groups with peers of above average ability. Importantly, this assumption does allow for sorting on observables. For example, high type individuals may be more likely to co-locate with other high types and, similarly, W_i may covary with $\underline{W}_{p(i)}$.

One way to ensure the satisfaction of Assumption 5.2 is to adopt the following assignment scheme. The planner begins by choosing a feasible joint distribution for $(\underline{T}_c, \underline{W}_c, X_c)$. Second, the planner forms classrooms with specific gender $(\underline{T}_c)$ and socioeconomic configurations $(\underline{W}_c)$. These classes must obey the constraints imposed by the joint distribution of T_i and W_i in the population. Third, the planner assigns each class configuration to a certain type of teacher, defined in terms of their value for X_c (e.g., a measure of teaching experience). Fourth, to fill a $T_i = t$ and $W_i = w$ slot the planner draws a student at random from that subpopulation. To fill an $X_c = x$ teaching slot, the planner draws a teacher at random from that subpopulation.[19] Assumption 5.2 is also consistent with endogenous group formation under particular (and strong) informational structures (see Heckman and Vytlacil, 2007a,b).

Assumption 5.2 allows for richer assignment patterns. For example, blacks in predominately black classrooms may be poorer (i.e., more likely to be eligible for free or reduced price school lunch), than blacks in predominately white classrooms. Observed measures of teacher quality may also vary with class composition. Hence, it be that teachers in predominately minority classrooms are less experienced.

Adapting the argument used in the proof to Proposition 2.1 (see Appendix B) we can show that Assumptions 5.1 and 5.2 yield the density factorization:

$$f_{\underline{A},U|\underline{T},\underline{W},X}(\underline{a}_c, u_c|\underline{t}_c, \underline{w}_c, x_c) = \left\{ \prod_{j=1}^{N_c} f_A(a_{cj}|\, w_{cj}, x_c) \right\} f_U(u_c|\, \underline{w}_c, x_c),$$

[19] This scheme approximates that used by the Berkeley Unified School District for elementary school enrollment.

so that the regression function

$$\mathbb{E}\left[Y_i|\, T_i = 1, S_i = s, \underline{W}_i = \underline{w}, X_i = x\right] = g_H\left(s, \underline{w}, x\right), \tag{16}$$

gives the expected outcome for a high type individual with observed characteristic $W_i = w$, given exogenous assignment to a group of composition $S = s$ with observed peer and location characteristics $\underline{W}_{p(i)} = \underline{w}_{p(i)}$ and $X_i = x$. The proxy variable regression function for low types, $g_L\left(s, \underline{w}, x\right)$, is analogously defined.

The reallocation estimands defined in Section 3 remain valid after replacing $m_H\left(s\right)$ and $m_L\left(s\right)$ with $g_H\left(s, \underline{w}, x\right)$ and $g_L\left(s, \underline{w}, x\right)$. The influence functions given in Section 4 remain valid after replacing $m_H\left(s\right)$ and $m_L\left(s\right)$ with $g_H\left(s, \underline{w}, x\right)$ and $g_L\left(s, \underline{w}, x\right)$ and $f_S\left(s\right)$ with $f_{S,\underline{W},X}\left(s, \underline{w}, x\right)$.

Integrating over $(\underline{w}_c, x_c)$ and invoking Assumptions 5.1 and 2.3 yields

$$\begin{aligned}
&\int\int\ldots\int f_{\underline{A},U|\underline{T},\underline{W},X}(\underline{a}_c, u_c|\underline{t}_c, \underline{w}_c, x_c) f_{\underline{W},X}(\underline{w}_c, x_c)\mathrm{d}w_{c1}, \ldots, \mathrm{d}w_{cN}\mathrm{d}x_c \\
&= \int\int\ldots\int \left\{\prod_{j=1}^{N_c} f_A(a_{cj}|\, w_{cj}, x_c)\right\} f_U(u_c|\, \underline{w}_c, x_c) f_{\underline{W},X}(\underline{w}_c, x_c)\mathrm{d}w_{c1}, \ldots, \mathrm{d}w_{cN}\mathrm{d}x_c \\
&= \left\{\prod_{j=1}^{N_c} f_A(a_{cj})\right\} f_U(u_c),
\end{aligned}$$

so that we can recover $m_H\left(s\right)$ directly by the partial mean

$$m_H\left(s\right) = \mathbb{E}_{\underline{W},X}\left[\mathbb{E}\left[Y_i|\, T_i = 1, S_i = s, \underline{W}_i, X_i\right]\right]$$

under appropriate support conditions and similarly for $m_L\left(s\right)$.

6 The social planner's problem

In this section we characterize the structure of average outcome maximizing assignments of individuals to groups. We allow group size to vary, but only consider reallocations which leave the marginal distribution of group-size fixed. Let group size $N_c \in \{n_1, \ldots, n_J\}$ with $\tau_j = \Pr\left(N_c = n_j\right)$; the class of reallocations we study is completely characterized by the $j = 1, \ldots, J$ conditional group-composition cumulative distribution functions: $F_{S|N}\left(s|n_j\right)$. The social planner's problem is thus a functional (i.e., infinite-dimensional) optimization one. Such problems are typically quite difficult to solve, standard mathematical programming results being inapplicable.

In our case we show, by exploiting the special structure of the planner's problem and the feasibility constraint, that a direct solution is available, easily characterized and computationally feasible. This result allows us to identify the maximum average outcome level available via reallocation. A comparison of the maximum average outcome with that observed under the status quo provides a measure of efficiency of the status quo (see Bhattacharya, 2009).

Analysis of the planner's problem also provides insight into the interaction of the production technology and resource constraint (i.e., the fraction of high types in the population) in determining the optimal allocation. Below we provide examples where, holding technology fixed, the optimal allocation is either integrating or segregating depending on the type structure of the population. This highlights the danger of informally inferring the optimality of segregation versus integration by inspection of the production technology alone (as is common in practice).

We assume that the planner knows the mean allocation response function, $m(s,n)$, the status quo assignment, $F_{S,N}^{\mathrm{sq}}(s,n)$, and the population fraction of high types, p_H. Her problem is to choose an allocation which maximizes expected average outcomes:

$$\max_{F_{S|N}(\cdot|n_1),\ldots,F_{S|N}(\cdot|n_J)} \sum_{j=1}^{J} \left[\frac{n_j}{\mu_N} \int m(s,n_j) f_{S|N}(s|\,n_j)\,\mathrm{d}s \right] \tau_j \tag{17}$$

subject to the restriction

$$\sum_{j=1}^{J} \frac{n_j}{\mu_N} \left[\int s f_{S|N}(s|\,n_j)\,\mathrm{d}s \right] \tau_j = p_H, \tag{18}$$

with $\mu_N = \mathbb{E}[N_i]$. Weighting by n_j/μ_N ensures that the planner maximizes average individual outcomes (and not the average of mean group outcomes).

Our characterization of the solution to (17) involves two steps. First, we solve a simplified problem. In the simplified problem all groups are of the same size. In this case the only observable dimension distinguishing groups is their composition. We show that the optimizing planner chooses the allocation, $F_S^*(s)$, in a way that implicitly 'concavifies' the mean allocation response function, $m(s)$ (we suppress the n argument when discussing the simplified problem). One intuition for our result follows from the observation that an optimizing planner behaves similarly to that of a cost minimizing producer facing (possibly) nonconvex isoquants (McFadden, 1978).

Second, using our first step result we show that the original problem can be broken into two simple steps. Let σ_j denote the fraction of high types in the subpopulation of individuals assigned to groups of size n_j (as part of a candidate reallocation). Conditional on choosing such an allocation, the optimal conditional allocations $F_{S|N}(s|n_1), \ldots, F_{S|N}(s|n_J)$ are determined by our first result. Since $\sigma_j = \int s f_{S|N}(s|n_j)\,\mathrm{d}s$ we can re-write the feasibility constraint (18) as

$$\sum_{j=1}^{J} \frac{n_j}{\mu_N} \sigma_j \tau_l^{\mathrm{sq}} = p_H,$$

and hence show that the original problem is equivalent to a finite-dimensional optimization problem where the planner chooses the vector $\sigma = (\sigma_1, \ldots, \sigma_J)'$. Furthermore we show that the equivalent problem is a concave one and hence that the Kuhn-Tucker conditions are both necessary and sufficient. This allows us to provide a fairly complete characterization of the planner's problem. Numerical computation of an outcome maximizing allocation is straightforward. We can therefore estimate the maximum attainable average outcome. A similar argument can be used to characterize the problem of minimizing expected average outcomes.

The concave envelope of $m(s,n)$ plays an important role in our argument. The following definition, adapted from Horst et al. (2000), defines this object.

Definition 6.1. *Let $m : \mathcal{S} \to \mathbb{R}^1$ be a continuous function with $\mathcal{S} = [\underline{s}, \overline{s}]$ (a convex set in $\mathbb{R}^1$), then the concave envelope of $m(s)$ taken over $\mathcal{S}$ is a function $M(s)$ such that (i) $M(s)$ is concave on $\mathcal{S}$, (ii) $M(s) \geq m(s)$ for all $s \in \mathcal{S}$, (iii) if $h(s)$ is any concave function defined on $\mathcal{S}$ such that $h(s) \geq m(s)$ for all $s \in \mathcal{S}$, then $h(s) \geq M(s)$ for all $s \in \mathcal{S}$.*

Formally $M(s)$ is the function whose truncated lower epigraph coincides with the convex hull of the truncated lower epigraph of $m(s)$ (see Rockafellar, 1970). Intuitively it is the uniformly best concave overestimator of $m(s)$.

We begin by considering the planner's problem when all groups are equally-sized. Outcome maximizing allocations in that setting are characterized by the following theorem.

Theorem 6.1. *Consider the problem*

$$\max_{F_S(\cdot)\in\Gamma_S} \int m(s) f_S(s)\,\mathrm{d}s, \qquad s.t. \qquad \int s f_S(s)\,\mathrm{d}s = p_H, \tag{19}$$

where $s \in \mathcal{S} = [\underline{s}, \overline{s}]$ with $\underline{s} \geq 0$, $\overline{s} \leq 1$, Γ_S is the space of all probability measures on $\mathcal{S}$, and $p_H = \mathbb{E}[T_i]$, then, with $F_S^*(\cdot)$ denoting a solution to (19),

$$\int m(s) f_S^*(s)\,\mathrm{d}s = M(p_H) \tag{20}$$

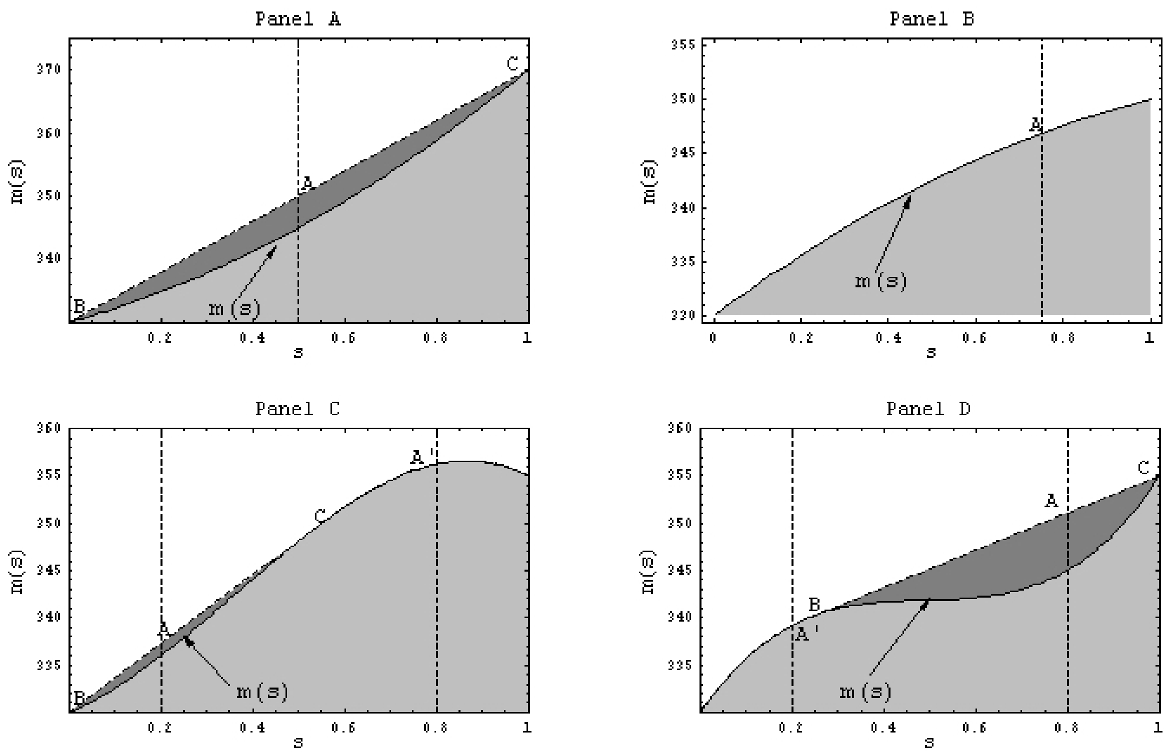

Figure 1: Optimal allocations for different $m(s)$ and p_H

NOTES: Each panel plots a different expected allocation response function, $m(s)$ (solid dark line). The concave envelopes of these expected allocation response functions, $M(s)$, are given by the dashed lines at or above $m(s)$. The vertical dashed lines indicate the population frequency of high types, p_H. For figures with two such lines the second line (i.e., the right-most line) gives the location of a second population frequency, p'_H. The point labeled A marks the location of $(p_H, M(p_H))$. The points labeled B and C mark the locations of, respectively, $(s_L, m(s_L))$ and $(s_U, m(s_U))$ (when $s_L \neq s_U$). The point labeled A', if present, marks the location of $(p'_H, M(p'_H))$.

and

$$F_S^*(s) = (1-\pi)\,\mathbf{1}(s \geq s_L) + \pi \mathbf{1}(s \geq s_U), \quad \pi = \begin{cases} \frac{p_H - s_L}{s_U - s_L} & s_L < s_U \\ 1/2 & s_L = s_U \end{cases} \tag{21}$$

where

$$s_L = \max\{s : s \geq \underline{s},\ s \leq p_H,\ M(s) = m(s)\}, \quad s_U = \min\{s : s \leq \overline{s},\ s \geq p_H,\ M(s) = m(s)\}.$$

Proof See Appendix B.

Theorem 6.1 shows that an outcome maximizing allocation may be constructed by a group composition density with just two mass points. The location of these mass points coincide with the s-axis values of the first extreme points to the 'right' and left' of $(p_H, M(p_H))$. To see why this is the case it is helpful to examine some examples in detail.[20] Figure 1 plots four different forms for $m(s)$. Consider Panel A of the figure. In that panel $m(s)$ is globally convex. The concave envelope of $m(s)$ is equal to the straight line passing through the points B, A and C. The vertical dashed line in this figure depicts the population frequency of high types, p_H. If 'production' on $M(s)$, the concave envelope of $m(s)$, were feasible, then, by Jensen's inequality, an optimal allocating would clearly be integrating: all groups would have a fraction of high types equal to p_H. While this is not possible, this same average outcome is achievable by a segregating allocation with groups of all low or high types. In Panel B of the figure, $m(s)$ is globally concave. In that case $m(s)$ and its concave envelope $M(s)$ coincide such that the integrated allocation maximizes average outcomes. These two cases correspond to those emphasized in the multi-community models literature.

Panels C and D depict more complicated examples. In Panel C $m(s)$ has both concave and convex regions. If $p_H = 0.2$, shown by the left-most vertical dashed line in the figure, then the social planner will form some groups with no high types (point B in the figure) and some partially integrated groups (point C in the figure). The proportion of each type of groups is determined by the feasibility constraint. This example illustrates the key idea of the theorem: because groups can be formed with different proportions of high types, the output level $M(p_H)$ is attainable. Since $M(s) \geq m(s)$ for all $s \in [0,1]$ and is concave it follows that $M(p_H)$ equals the maximal attainable average outcome level. Mathematically the result follows from that fact that any point on the convex hull of a set of points can be represented as a linear combination of extreme points on the hull.

Panel C highlights a second feature of our problem. As discussed above, when $p_H = 0.2$ (left-most vertical dashed line), $M(p_H) \geq m(p_H)$ so that the social planner will choose a

[20]We thank Emmanuel Saez for providing some of these examples. His intuitive insight was key for being able to formulate the proof of Theorem 6.1.

segregating allocation. In contrast when $p_H = 0.8$ (right-most vertical dashed line) $M(p_H) = m(p_H)$ so that the social planner will choose a perfectly integrated allocation. This provides a simple, albeit stylized, example of how knowledge of the production technology alone is not sufficient for solving the planner's problem. Panel D gives a further example of an average outcome response function with both convex and concave portions.

The solution to the original social planner's problem is characterized by the following corollary to Theorem 6.1.

Corollary 6.1. *A solution to the social planner's problem defined by (17) and (18) is given by*

$$F^*_{S|N}(s|\, n_j) = [1 - \pi(\sigma_j)]\, \mathbf{1}(s \geq s_L(\sigma_j)) + \pi(\sigma_j)\, \mathbf{1}(s \geq s_U(\sigma_j))$$

where

$$\pi(\sigma_j) = \begin{cases} \frac{\sigma_j - s_L(\sigma_j)}{s_U(\sigma_j) - s_L(\sigma_j)} & s_L(\sigma_j) < s_U(\sigma_j) \\ 1/2 & s_L(\sigma_j) = s_U(\sigma_j) \end{cases}$$

for $j = 1, \ldots, J$ *and*

$$s_L(\sigma_j) = \max\{s : s \geq \underline{s},\ s \leq \sigma_j,\ M(s, n_j) = m(s, n_j)\}$$
$$s_U(\sigma_j) = \min\{s : s \leq \overline{s},\ s \geq \sigma_j,\ M(s, n_j) = m(s, n_j)\},$$

with $M(s, n_j)$ *the concave envelope of* $m(s, n_j)$ *on* $s \in \mathcal{S}$ *and* $\sigma_1, \ldots, \sigma_J$ *the solution to the concave programming problem*

$$\max_{\sigma_1 \in \mathcal{S}, \ldots, \sigma_J \in \mathcal{S}} \sum_{j=1}^{J} \frac{n_j}{\mu_N} M(\sigma_j, n_j)\, \tau_j, \qquad s.t. \qquad \sum_{j=1}^{J} \frac{n_j}{\mu_N} \sigma_j \tau_j = p_H. \tag{22}$$

Proof See Appendix B.

Corollary 6.1 provides a simple algorithm for calculating the maximum attainable average outcome available via reallocation. First, compute $M(s, n_j)$ for each of the J group sizes. Second, solve the concave program (22). Third, compute the value of $\sum_{j=1}^{J} \frac{n_j}{\mu_N} M(\sigma_j, n_j)\, \tau_j$ at the solution.

Our final identification result follows directly:

Proposition 6.1. *If (i) Assumptions 2.1 to 2.6 hold and (ii)* $f^{\mathrm{sq}}_{S|N}(s|\, n_j) > 0$ *for all* $s \in \mathcal{S}$ *and* $j = 1, \ldots, J$*, then (a)* $F^*_{S|N}(s|\, n_j)$ *is identified and (b) so is the efficiency measure*

$$\beta^{\mathrm{esq}} = \sum_{j=1}^{J} \left[\frac{n_j}{\mu_N} \int m(s, n_j)\, f^*_{S|N}(s|\, n_j)\, \mathrm{d}s \right] \tau_j - \mathbb{E}[Y].$$

The *efficiency of the status quo measure* (ESQ), β^{esq}, equals the maximum average outcome gain, relative to the status quo, available via reallocation.

7 Empirical illustration

Here we apply our methods to an analysis of classroom gender composition on student achievement. The data were collected as part of a randomized study of the effects of class size on student performance (Project STAR). They have been previously analyzed by, among others, Whitmore (2005), Krueger and Whitmore (2001) and Graham (2008). The study involved randomized assignment of both teachers and students to classrooms, a design feature important to our analysis. We focus on the question of the effect of segregation by sex in classrooms.

We have information on 5,781 kindergarten students in 325 classrooms (on average 18 per class). We focus on math achievement as the outcome, normalized to have zero mean and unit variance. The average test score for girls is 0.08 and -0.08 for boys. Girls make up 49% of the sample. Figure 2 presents average achievement scores, averaged over all children in the class, as a function of the proportion of girls in the class. There is a clear upward slope in the regression function, implying that, on average, classes with more girls perform better than classes with few girls. The apparent nonlinearity of the curve is also suggestive of spillover effects. The second panel in the figure presents a histogram of the proportion of girls in the 325 classrooms, ranging from 0.28 to almost 0.80.

Figure 3 presents estimates of $m_H(s)$ and $m_L(s)$ (the regression functions for girls and boys respectively) separately. The estimation procedure is as described in Section 4 above with the modifications needed for the presence of additional observables implemented as described in Section 5. Other particulars of the estimation procedure are detailed in the notes to the tables and figures.

These nonparametric estimates underlie our estimates of the reallocation effects studied above. In Table 1 we present estimates of our various estimands. In Panel A we present the preferred nonparametric estimates. In the first row the results for the average spillover effect β^{ase} are reported. The estimates suggest that, on average (averaged over both girls and boys), kids benefit from having more girls in the class. This provides strong, nonparametric, evidence of peer spillovers from gender composition (see Whitmore, 2005). The next row of estimates gives the local segregation outcome effect, β^{lsoe}. It shows that, although on average kids benefit from having more girls as classmates, reallocating students to make classes slightly more segregrated by sex would not change average outcomes much. Girls would benefit from such segregation, but boys would suffer to approximately the same extent.

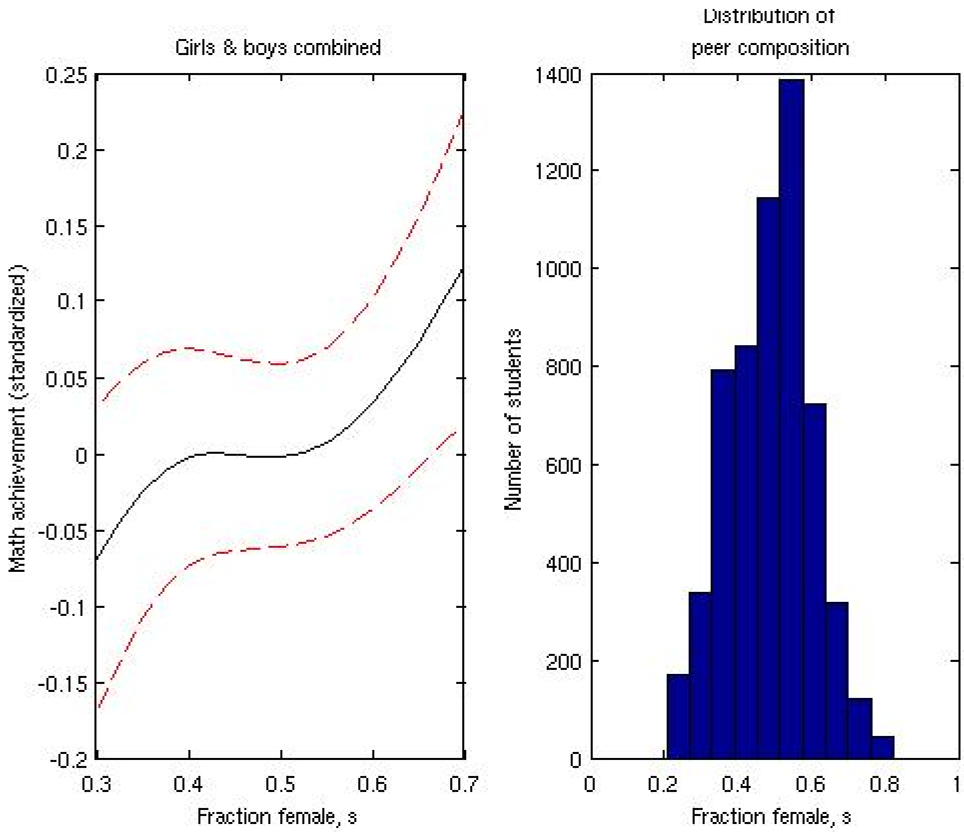

Figure 2: Average math achievement and classroom gender composition, Project STAR Kindergarten Students

NOTES: The left-hand-side of the figure plots kernel partial mean estimates of $m(s) = \mathbb{E}\left[sg_H\left(s, \underline{W}_i, X_i\right) + (1-s)\, g_L\left(s, \underline{W}_i, X_i\right)\right]$ where $\underline{W}_i$ is empty and X_i includes total school enrollment, fraction female in the school, and class size. A multivariate standard normal kernel was used with a bandwidth matrix proportional to the covariance matrix of the regressors. The degree of proportionality was chosen by leave-own-school-out cross-validation. The dashed lines are pointwise 90 percent confidence intervals calculated using the approach of Newey (1994b) (modified to allow for within-school dependence across observations). Units attending schools with enrollments below 50 or above 150 and/or those in schools with fraction female below 0.35 or above 0.65 were trimmed when forming the partial mean (about 9 percent of the students). Valid test scores, standardized to be mean zero with unit variance, were available for $I = 5,871$ students in $C = 325$ classrooms located across 79 different schools. The right-hand-side of the figure plots a histogram of peer composition at the individual level.

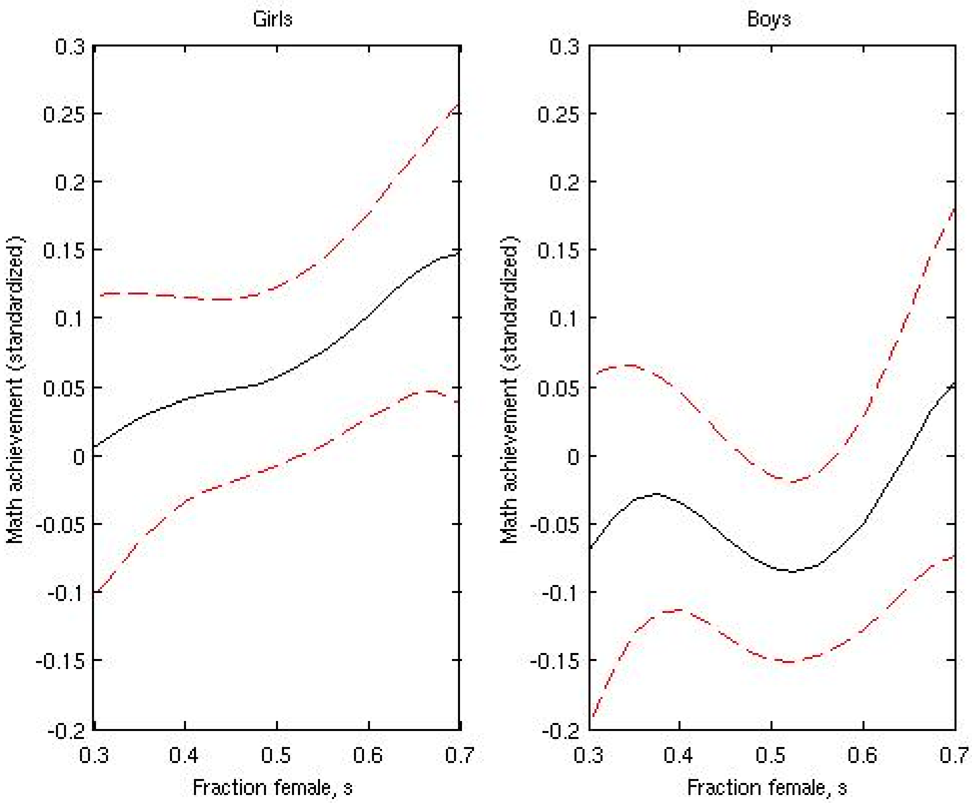

Figure 3: Average math achievement by gender and classroom gender composition, Project STAR Kindergarten Students

NOTES: The figure plots kernel partial mean estimates of $m_H(s) = \mathbb{E}[g_H(s, \underline{W}_i, X_i)]$ and $m_L(s) = \mathbb{E}[g_L(s, \underline{W}_i, X_i)]$. Bandwidths, regressors, trimming and confidence intervals are as described in the notes to Figure 2. A total of 2,857 students are used to compute the girls' figure and 3014 students for the boys' figure.

The latter is shown directly by the last estimand, the local segregation inequality effect, β^{lsie}. Increasing segregation by a small amount increases the average test score difference between girls and (lower performing) boys. Because the average effect of additional segregation on outcomes is close to zero, it is not surprising that breaking this down into a private and public component β^{lppe} and β^{lepe} does not show much of an effect either.

The last two columns of the table present estimates based on parametric models. The first of these is the widely used linear-in-means model. In this model the effect of the class composition is identical for all children. As a result the reallocation effects, β^{lsoe}, β^{lppe} and β^{lepe} are constrained to equal zero. Although consistent with our nonparametric estimates, the linear-in-means estimate of the average spillover effect is considerably larger. The same holds for the second parametric model that allows the for heterogeneity in peer effects by type (e.g., Angrist and Lang, 2004).

Taking the partial mean estimates of the conditional means $m_H(s)$ and $m_L(s)$ as given, we can informally solve the social planner's problem discussed in Section 5. Because we have no classrooms in the sample with very small or large fractions of girls, we restrict the allowable allocations to those with fractions girls in the range $[0.3, 0.7]$. Inspection of Figure 2 suggests that average math achievement will be maximized when approximately two thirds of classrooms are 40% girls and 60% boys, and the remaining one third of classrooms 70% girls and 30% boys (as segregrated as allowed).[21] This would raise average test scores by about 0.04 standard deviations relative to the status quo.

8 Summary

In this paper we have developed a unified framework for the analysis of the effects of segregation in the presence of social spillovers. We provide nonparametric identification and estimation results for our proposed estimands when matching and sorting is on observables alone. We also explore features of the social planner's problem and illustrate our methods by studying the effects of sex segregation in kindergarten classrooms.

Several areas potentially merit further study. The approach taken in this paper has been to leave $m_H(s)$ and $m_L(s)$ nonparametric. The price for this flexibility is that our identification and estimation procedures require strong conditions on the status quo assignment mechanism. It would be interesting to explore whether the imposition of various a priori restrictions on $m_H(s)$ and $m_L(s)$ might facilitate the development of positive identification results under less stringent restrictions on the status quo (c.f., Nesheim, 2002; Epple et al., 2010). An exploration of partially identifying assumptions, as done by Manski (2013) in a

[21] For simplicity we assume that all classes have the average number of students (about 18).

Table 1: Nonparametric and parametric estimates of spillover strength and reallocation effects (math achievement, Project STAR Kindergarten Students)

	Panel A: Nonparametric				**Panel B: Parametric**	
	(1)	(2)	(3)	(4)	Linear-in-Means	Type-Specific Linear-in-Means
β^{ase}	0.2273	0.2600	0.3065	0.3469	0.4433	0.4728
	(0.1432)	(0.1352)	(0.1345)	(0.1341)	(0.1995)	(0.2007)
β^{lsoe}	−0.0104	−0.0108	−0.0075	−0.0004	–	0.0006
	(0.0238)	(0.0236)	(0.0229)	(0.0205)		(0.0067)
α^{lppe}	−0.0003	0.0002	0.0004	0.0009	–	−0.0003
	(0.0027)	(0.0027)	(0.0027)	(0.0027)		(0.0034)
α^{lepe}	−0.0101	−0.0111	−0.0079	−0.0012	–	0.0009
	(0.0240)	(0.0238)	(0.0230)	(0.0206)		(0.0036)
β^{lsie}	0.0552	0.0521	0.0474	0.0548	0.0628	0.0656
	(0.0250)	(0.0274)	(0.0296)	(0.0266)	(0.0264)	(0.0266)
h	CV/RT	5/6 of CV/RT	2/3 of CV/RT	1/2 of CV/RT	–	–

NOTES: The estimates reported in Panel A of the Table were calculated use the kernel procedure outlined in the main text. Estimated standard errors are in parentheses. A multivariate standard normal kernel was used with a bandwidth matrix proportional to the covariance matrix of the regressors (fraction female in the classroom, total school enrollment, fraction female in the entire school and class size). In the first column of Panel A the degrees of proportionality used for estimating $g_H(s, \underline{w}, x)$and $g_L(s, \underline{w}, x)$ were chosen by leave-own-school-out cross validation. The bandwidths for $\nabla_s g_H(s, \underline{w}, x)$and $\nabla_s g_L(s, \underline{w}, x)$ were then taken to be rescaled versions of the corresponding cross-validated ones. The chosen rescaling reflects the differential MSE-optimal bandwidth for pointwise conditional mean and derivative estimation. The estimated standard errors are calculated using nonparametric estimates of the relevant influence functions. The bandwidth used for the joint density of $f_{S,\underline{W},X}(s, \underline{w}, x)$, which appears in the influence functions, is a multivariate version of Silverman's 'rule-of-thumb' bandwidth (see Wand and Jones, 1995, p. 111). The bandwidth used for $\nabla_s f_{S,\underline{W},X}(s, \underline{w}, x)$ is a rescaling of this rule-of-thumb bandwidth. Columns 2 through 4 report undersmoothed estimates based on bandwidth values equal to, respectively, 5/6, 2/3 and 1/2 of the column one bandwidth values. Panel B of the table reports estimates based on parametric models for $g_H(s, \underline{w}, x)$ and $g_L(s, \underline{w}, x)$. Standard errors were calculated taking into account the sequential nature of the estimation procedure. In both the nonparametric and parametric cases standard errors appropriately account for arbitrary within-school dependence in outcomes across individuals. See the notes to Figure 2 for additional details on the estimation sample.

related context, might also be fruitful.

Finally, we have not formally developed an estimator for β^{esq}, our measure of the efficiency of the status quo assignment. While showing consistency of the simple plug in estimator (informally) used in the application should be straightforward, the characterization of its asymptotic sampling properties appears more difficult (see Graham, 2011).

9 Postscript

The onset of the research reported in this paper began in the summer of 2005, with drafts circulated as early as 2007. In this section we touch on several developments in econometrics, statistics and machine learning in the intervening period that connect with the work reported on above.

'As if' double randomization, whereby individuals – conditional on their type – are randomly assigned to groups and groups, so formed, are randomly assigned to locations, drives our main identification results. Also necessary for identification are some support conditions on the status quo assignment. For our local reallocation estimands, the necessary support conditions are modest (see Assumption 2.5 above). The required support conditions for the *efficiency of the status quo* measure are substantially stronger (see Proposition 6.1).

Some real world indication of the practical importance of these conditions is provided by the work of Carrell et al. (2013). Carrell et al. (2013) measure the effect of peer group composition on academic achievement among Air Force Academy cadets. Informed by a fitted flexible parametric model, they changed the peer group assignment mechanism at the Air Force Academy in an attempt to maximize average academic achievement. In practice academic achievement declined under their policy. Their work provides a cautionary tale on the dangers of extrapolation. It also suggests that identification of our *efficiency of the status quo* measure may be difficult in practice. In contrast, we expect that our local reallocation estimands will be informative about the effects of modest perturbations of the status quo under conditions where 'as if' double randomization holds. The challenges associated with extrapolation in other areas of causal inference are well understood (see, for example, Imbens and Rubin (2015)). We expect such challenges to be especially acute when considering reallocation effects.

Empirical research on sorting and peer effects, which connects to and/or builds upon ideas in this paper, includes work by Angrist and Lang (2004), Friesen and Krauth (2007, 2010), Booij et al. (2017), Li et al. (2019) and, especially, Graham et al. (2023). Angrist (2014) and Graham (2018) provide contrasting assessments of, and suggested ways forward for, empirical research on peer effects.

Our work also intersects with ideas in the causal inference literature on interference and spillovers. Our exchangeability restriction on the potential response function shares features with ideas introduced by Manski (2013) in the treatment effects context. See also Ferracci et al. (2014), Kasy (2016), Baird et al. (2018) and Viviano and Rudder (2024) for related applications and/or methodological developments. Of particular interest are analyses which allow for partial identification. Our no sorting and no matching conditions may be difficult to defend outside of overtly experimental settings; formulating relaxations of these conditions and studying their implications for identification would be of considerable value.

A related avenue of exploration would involve imposing additional restrictions on the production technology and exploring whether such restrictions allow for the relaxation of conditions on the status quo assignment. The referee, for example, has suggested that the imposition of a partially linear structure on, for example, $g_L(s, \underline{w}, x)$ and $g_H(s, \underline{w}, x)$ might allow for a weakening of our no matching/sorting on observables conditions. Such explorations could be very important for bringing the ideas of this paper to observational data.

Our focus on kernel-based estimation methods reflects the widespread reliance on such methods by econometricians in the early 2000s. It also reflects the influence of Jim Powell's seminal work on weighted average derivative estimation on the analysis of our own local reallocation effects (Powell et al., 1989).

Since the initial drafting of our paper, developments in computer science, statistics and econometrics have substantially expanded the range of estimation procedures which can be applied to semiparametric two-step estimators like our own. Compare Newey (1994a) with, for example, Chernozhukov et al. (2022). We expect these innovations in semiparametric estimation to be of relevance to empirical researchers wishing to study rellocation effects. Formal work on methods of estimation for the estimands introduced above might also be of interest.

In Graham et al. (2018) we explore covariate adjustment and semiparametric efficiency bounds in the related context introduced in Graham et al. (2014). Covariate adjustment in the setting introduced in this paper appears to be comparatively more complicated; the basic results reported in Section 5 above provide a basic starting point for further research. Covariates raise important conceptual issues even at the level of estimand definition. Consider a status quo assignment where both peers and other outcome-enhancing resources (e.g., teaching quality) are unequally distributed. In this setting a reallocation may change the distribution of achievement both because of changes in peer group composition and because, for example, average teaching quality changes for low types. These complications are scientifically interesting as well as policy-relevant. Some additional discussion of this, and related issues, can be found in the survey paper by Graham (2018) as well as in Graham et al. (2023).[22]

Understanding the effects of segregation remains as important today as it was when James Samuel Coleman undertook his pioneering research over 50 years ago. Segregation, by race and income, remains widespread in American neighborhoods and schools. Our understanding of the consequences of such segregation remains incomplete.

[22]We thank the anonymous referee for some shrewd discussion on this issue.

References

Altonji, J. G. and Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053 – 1102.

Angrist, J. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Angrist, J. D., Graddy, K., and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67(3):499 – 527.

Angrist, J. D. and Lang, K. (2004). Does school integration generate peer effects? evidence from boston's metco program. *American Economic Review*, 94(5):1613 – 1634.

Arnott, R. and Rowse, J. (1987). Peer group effects and educational attainment. *Journal of Public Economics*, 32(3):287 – 305.

Baird, S., Bohren, J. A., McIntosh, C., and Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844 – 860.

Becker, G. S. and Murphy, K. M. (2000). *Social Economics: Market Behavior in a Social Environment*. Harvard University Press, Cambridge, MA.

Benabou, R. (1993). Workings of a city: location, education, and production. *Quarterly Journal of Economics*, 1-8(3):619 – 652.

Benabou, R. (1996). Equity and efficiency in human capital investment: the local connection. *Review of Economic Studies*, 63(2):237 – 264.

Bhattacharya, D. (2009). Inferring optimal peer assignment from experimental data. *Journal of the American Statistical Association*, 104(486):486 – 500.

Blundell, R. and Powell, J. L. (2003). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, chapter Endogeneity in nonparametric and semiparametric regression models, pages 312 – 357. Cambridge University Press.

Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability peer effects in university: evidence from a randomized experiment. *Review of Economic Studies*, 84(2):547 – 578.

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1):57–75.

Brock, W. A. and Durlauf, S. N. (2001). *Handbook of Econometrics*, volume 5, chapter Interactions-based models, pages 3297 – 3380. North-Holland, Amsterdam.

Calabrese, S., Epple, D., Romer, T., and Sieg, H. (2006). Local public good provision: voting, peer effects, and mobility. *Journal of Public Economics*, 90(6-7):959 – 981.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, 76(4):1239 – 1267.

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

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855 – 882.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4):1501 – 1535.

Ciccone, A. and Peri, G. (2006). Identifying human-capital externalities: theory with applications. *Review of Economic Studies*, 73(2):381 – 412.

Coleman, J. S. (1966). Equality of educational opportunity. Technical report, U.S. Department of Health, Education and Welfare, Washington D.C.

de Bartolome, C. A. M. (1990). Equilibrium and inefficiency in a community model with peer group effects. *Journal of Political Economy*, 98(1):110 – 133.

Duncan, G. J. and Raudenbush, S. (1993). *Does it Take a Village? Community Effects on Children, Adolescents, and Families*, chapter Neighborhoods and adolescent development: how can we determine the links?, pages 105 – 136. Pennsylvania State University Press, State College, PA.

Durlauf, S. N. (1996a). Associational redistribution: A defense. *Politics & Society*, 24(4):391 – 410.

Durlauf, S. N. (1996b). *Dynamic Disequilibrium Modeling*, chapter Neighborhood feedbacks, endogenous stratification, and income inequality, pages 505 – 534. Cambridge University Press, Cambridge.

Durlauf, S. N. (1996c). A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75 – 93.

Durlauf, S. N. (2001). A framework for the study of individual behavior and social interactions. *Sociological Methodology*, 31(1):47 – 87.

Durlauf, S. N. (2004). *Handbook of Regional and Urban Economics*, volume 4, chapter Neighborhood effects, pages 2173 – 2242. North-Holland, Amsterdam.

Epple, D., Newlon, E., and Romano, R. (2002). Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics*, 83(1):1 – 48.

Epple, D., Peress, M., and Sieg, H. (2010). Identification and semiparametric estimation of equilibrium models of local jurisdictions. *American Economic Journal: Microeconomics*, 2(4):195 – 220.

Epple, D. and Romano, R. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1):33 – 62.

Fernández, . R. (2003). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, chapter Sorting, education and inequality, pages 1 – 40. Cambridge University Press, Cambridge.

Ferracci, M., Jolivet, G., and van den Berg, G. J. (2014). Evidence of treatment spillovers within markets. *Review of Economics and Statistics*, 96(5):812 – 823.

Ferreyra, M. M. (2007). Estimating the effects of private school vouches in mutidistrict economies. *American Economic Review*, 97(3):790 – 817.

Figlio, D. N. and Page, M. E. (2002). School choice and the distributional effects of ability tracking: does separation increase inequality? *Journal of Urban Economics*, 51(3):497 – 514.

Friesen, J. and Krauth, B. (2007). Sorting and inequality in canadian schools. *Journal of Public Economics*, 91(11-12):2185–2212.

Friesen, J. and Krauth, B. (2010). Sorting, peers, and achievement of aboriginal students in british columbia. *Canadian Journal of Economics*, 43(4):1273–1301.

Ginsberg, W. (1974). The multiplant firm with increasing returns to scale. *Journal of Economic Theory*, 9(3):283 – 292.

Glaeser, E. L. and Scheinkman, J. A. (2003). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 1, chapter Non-market interactions, pages 339 – 369. Cambridge University Press, Cambridge.

Graham, B. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):643–660.

Graham, B. S. (2011). *Handbook of Social Economics*, volume 1, chapter Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers, pages 965 – 1052. North-Holland, Amsterdam.

Graham, B. S. (2018). Identifying and estimating neighborhood effects. *Journal of Economic Literature*, 56(2):450 – 500.

Graham, B. S., Imbens, G. W., and Ridder, G. (2007). Redistributive effects for discretely-valued inputs. IEPR Working Paper 07.7, University of Southern California.

Graham, B. S., Imbens, G. W., and Ridder, G. (2014). Complementarity and aggregate implications of assortative matching: a nonparametric analysis. *Quantitative Economics*, 5(1):29 – 66.

Graham, B. S., Imbens, G. W., and Ridder, G. (2018). Identification and efficiency bounds for the average match function under conditionally exogenous matching. *Journal of Business and Economic Statistics*.

Graham, B. S., Ridder, G., Thiemann, P., and Zamarro, G. (2023). Teacher-to-classroom assignment and student achievement. *Journal of Business and Economic Statistics*, 41(4):1328 – 1340. Lund University.

Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review*, 94(4):919 – 943.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of American Statistical Association*, 84(408):986 – 995.

Heckman, J. J. and Vytlacil, E. J. (2007a). *Handbook of Econometrics*, volume 6B, chapter Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation, pages 4779 – 4874. North-Holland, Amsterdam.

Heckman, J. J. and Vytlacil, E. J. (2007b). *Handbook of Econometrics*, volume 6B, chapter Econometric evaluation of social programs, part II: using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments, pages 4875 – 5143. North-Holland, Amsterdam.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945 – 960.

Horst, R., Pardalos, P. M., and Thoai, Nguyen V., E. . (2000). *Introduction to Global Optimization.* Kluwer Academic Publishers., Dordrecht, 2nd edition.

Hoxby, C. M. (2003). School choice and school competition: evidence from the united states. *Swedish Economic Policy*, 10(1):9 – 65.

Hudgens, M. G. and Halloran, E. (2008). Toward a causal inference with interference. *Journal of American Statistical Association*, 103(482):832 – 842.

Ibaraki, T. and Katoh, N. (1988). *Resource Allocation Problems: Algorithmic Approaches*. The MIT Press, Cambridge, MA.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics*, 86(1):4 – 29.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, Cambridge.

Jacob, B. A. (2004). Public housing, housing vouchers and student achievement: evidence from public housing demolitions in chicago. *American Economic Review*, 94(1):233 – 258.

Kasy, M. (2016). Partial identification, distributional preferences, and the welfare ranking of policies. *Review of Economics and Statistics*, 98(1):111 – 131.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83 – 119.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: evidence from project star. *Economic Journal*, 111(468):1 – 28.

Ladd, H. F. (2003). Comment on caroline m. hoxby: School choice and school competition: evidence from the united states. *Swedish Economic Policy*, 10(1):67 – 76.

Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics*, 116(3):777 – 803.

Li, X., Ding, P., Lin, Q., Yang, D., and Liu, J. S. (2019). Randomization inference for peer effects. *Journal of American Statistical Association*, 114(528):1651 – 1664.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York.

Luenberger, D. G. (2005). *Linear and Nonlinear Programming*. Springer, New York, 2nd edition.

Mael, F. e. a. (2005). Single-sex versus coeducational schooling: a systematic review. Technical report, U.S. Department of Education, Office of Planning, Evaluation and Policy Development., Washington D.C.

Manski, C. F. (1992). Educational choice (vouchers) and social mobility. *Economics of Education Review*, 11(4):351 – 369.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem,. *Review of Economic Studies*, 60(3):531 – 542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115 – 136.

Manski, C. F. (2013). Identification of treatment response with social interactions. *Econometrics Journal*, 16(1):S1–S23.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press, Cambridge, MA.

Mayer, S. E. and Jencks, C. (1989). Growing up in poor neighborhoods: how much does it matter? *Science*, 243(4897):1441 – 1445.

McFadden, D. L. (1978). *Production Economics: A Dual Approach to Theory and Applications*, volume 1: The Theory of Production, chapter Cost, revenue, and profit functions, pages 3 – 109. North-Holland, Amsterdam.

Moffitt, R. (2001). Policy interventions, low-level equilibria and social interactions. In Durlauf, S. and Young, P., editors, *Social Dynamics*. MIT Press.

Morse, S. (1998). Separated by sex: A critical look at single sex education for girls. Technical report, American Association of University Women Educational Foundation, Washington, D.C.

Nesheim, L. (2002). Equilibrium sorting of heterogenous consumers across locations: theory and empirical implications. Technical Report CWP08/02, CeMMAP.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99 – 135.

Newey, W. K. (1994a). The asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349 – 1382.

Newey, W. K. (1994b). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2):233 – 253.

Newey, W. K. and McFadden, D. (1994). *Handbook of Econometrics*, volume 4, chapter Large sample estimation and hypothesis testing, pages 2111 – 2245. North-Holland, Amsterdam.

Newey, W. K. and Stoker, T. M. (1994). Efficiency of weighted average derivative estimators and index models. *Econometrica*, 61(5):1199 – 1223.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. essays on principles. section 9. *Statistical Science*, 5(4):465 – 472.

Oakes, J. (1992). Can tracking research inform practice? technical, normative, and political considerations. *Educational Researcher*, 21(4):12 – 21.

Paulson, A. and Teicher, S. A. (2006). Move to single-sex classes fans debate. *Christian Science Monitor*.

Piketty, T. (2000). *Handbook of Income Distribution*, volume 1, chapter Theories of persistent inequality and intergenerational mobility, pages 430 – 476. North-Holland, Amsterdam.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of weighted average derivatives. *Econometrica*, 57(6):1403 – 1430.

Rivkin, S. and Welch, F. (2006). *Handbook of the Economics of Education*, volume 2, chapter Has school desegregation improved academic and economic outcomes for blacks, pages 1019 – 1049. North-Holland, Amsterdam.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688 – 701.

Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, 6(4):377 – 400.

Rubin, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25(3):279 – 292.

Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. (2002). Assessing “neighborhood effects”: social processes and new directions in research. *Annual Review of Sociology*, 28:443 – 478.

Sax, L. (2005). *Why Gender Matters: What Parents and Teachers Need to Know about the Emerging Science of Sex Differences*. Doubleday, New York.

Schofield, J. W. (1995). *Handbook of Research on Multicultural Education*, chapter Review of research on school desegregation’s impact on elementary and secondary school students, pages 597 – 616. Macmillan Publishing USA, New York.

Sobel, M. E. (2006a). *Poverty Traps*, chapter Spatial concentration and social stratification: does the clustering of disadvantage ‘beget’ bad outcomes?, pages 204 – 229. Princeton University Press, Princeton, NJ.

Sobel, M. E. (2006b). What do randomized studies of housing mobility demonstrate?: causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398 – 1407.

Solon, G. (1999). *Handbook of Labor Economics*, volume 3, chapter Intergenerational mobility in the labor market, pages 1761 – 1800. North-Holland, Amsterdam.

Urquiola, M. (2005). Does school choice lead to sorting? evidence from tiebout variation. *American Economic Review*, 95(4):1310 – 1326.

Viviano, D. and Rudder, J. (2024). Policy design in experiments with unknown interference. Technical report, Harvard University.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall/CRC, Boca Raton.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: evidence from a randomized experiment. *American Economic Review*, 95(2):199 – 203.

Appendices

A Some preliminary results

Lemma A.1. *For X, a continuous random variable, with (i) compact support $\mathcal{X} = [a, b]$, (ii) cumulative distribution function $F_X(X)$, and (iii) $g(\cdot)$ a continuously differentiable function on the support of X:*

1. *The slope coefficient of the (mean squared error minimizing) linear predictor (LP) of $g(X)$ given X has a weighted average derivative representation of*

$$\mathrm{B} = \frac{\mathbb{C}(g(X), X)}{\mathbb{V}(X)} = \mathbb{E}\left[\omega(X)\frac{\partial g(X)}{\partial x}\right],$$

where

$$\omega(x) = \frac{1}{f_X(x)} \frac{\mathbb{E}[X - \mu_X | X \geq x](1 - F_X(x))}{\int_{v=a}^{v=b} \mathbb{E}[X - \mu_X | X \geq v](1 - F_X(v))\,\mathrm{d}v}, \quad \mathbb{E}[\omega(X)] = 1,$$

and

2. B *gives maximum weight to values of $\frac{\partial g(X)}{\partial x}$ for X close to its mean, $\mu_X = \mathbb{E}[X]$, and minimum weight when X is near the boundaries of its support.*

The proof for the first result of the Lemma is similar to that of Lemma 5 of Angrist et al. (2000). The second result of the Lemma, i.e., the precise characterization of the weighting process follows from a simple integration by parts argument. Observe that $g(X) - g(a) = \int_{u=a}^{u=X} \frac{\partial g(u)}{\partial x}\mathrm{d}u$ and that $\mathbb{E}[g(a)(X - \mu_X)] = 0$. Under weak conditions we therefore have

$$\begin{aligned}
\mathbb{C}(g(X), X) &= \mathbb{E}[g(X)(X - \mu_X)] \\
&= \mathbb{E}\left[\int_{u=a}^{u=X} \frac{\partial g(u)}{\partial x}(X - \mu_X)\,\mathrm{d}u\right] \\
&= \mathbb{E}\left[\int_{u=a}^{u=b} \frac{\partial g(u)}{\partial x}(X \geq u)(X - \mu_X)\,\mathrm{d}u\right] \\
&= \int_{u=a}^{u=b} \frac{\partial g(u)}{\partial x}\mathbb{E}[(X \geq u)(X - \mu_X)]\,\mathrm{d}u \\
&= \int_{u=a}^{u=b} \frac{\partial g(u)}{\partial x}\mathbb{E}[X - \mu_X | X \geq u](1 - F_X(u))\,\mathrm{d}u.
\end{aligned}$$

The variance of X can be written as

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}\left[X(X-\mu_X)'\right] \\
&= \mathbb{E}\left[\int_{v=a}^{v=X} 1(X-\mu_X)\,\mathrm{d}v\right] \\
&= \int_{v=a}^{v=b} \mathbb{E}\left[X-\mu_X|X\geq v\right](1-F(v))\,\mathrm{d}v.
\end{aligned}$$

The first result follows for $\omega(x)$ as given in the Lemma. To show the second result, that the weighted average derivative representation of B gives the most emphasis to values of $\frac{\partial g(X)}{\partial x}$ for X close to its mean, begin by noting that

$$\mathbb{E}\left[\omega(X)\frac{\partial g(X)}{\partial x}\right] = \frac{\int_{u=a}^{u=b}\frac{\partial g(u)}{\partial x}\mathbb{E}\left[X-\mu_X|X\geq u\right](1-F_X(u))\,\mathrm{d}u}{\int_{v=a}^{v=b}\mathbb{E}\left[X-\mu_X|X\geq v\right](1-F_X(v))\,\mathrm{d}v}.$$

Therefore the size of the weight on $\frac{\partial g(x)}{\partial x}$ is proportional to

$$\mathbb{E}\left[X-\mu_X|X\geq x\right](1-F_X(x)).$$

Integration by parts (with $u=1-F_X(t)$ and $v=t$) gives

$$\begin{aligned}
\int_x^b \left[1-F_X(t)\right]\mathrm{d}t &= \left[1-F_X(t)\right]t\Big|_x^b + \int_x^b tf_X(t)\,\mathrm{d}t \\
&= -\left[1-F_X(x)\right]x + \int_u^b tf_X(t)\,\mathrm{d}t.
\end{aligned} \tag{23}$$

We then write

$$\begin{aligned}
\frac{\partial}{\partial x}\left\{\mathbb{E}\left[X-\mu_X|X\geq x\right](1-F_X(x))\right\} &= \frac{\partial}{\partial x}\int_x^b x\ f_X(t)\,\mathrm{d}t - \frac{\partial}{\partial x}\left[1-F_X(x)\right]\mu_X \\
&= \frac{\partial}{\partial x}\int_x^b x\ f_X(t)\,\mathrm{d}t + \mu_X f_X(x)
\end{aligned}$$

Using (23) to substitute for $\frac{\partial}{\partial x}\int_x^b x\ f_X(t)\,\mathrm{d}t$ gives

$$\begin{aligned}
\frac{\partial}{\partial x}\left\{\mathbb{E}\left[X-\mu_X|X\geq x\right](1-F_X(x))\right\} &= \frac{\partial}{\partial x}\left\{\left[1-F_X(x)\right]x + \int_x^b \left[1-F_X(t)\right]\mathrm{d}t\right\} \\
&+ \mu_X f_X(x) \\
&= \left[1-F_X(x)\right] + \frac{\partial}{\partial x}\int_x^b \left[1-F_X(t)\right]\mathrm{d}t \\
&- (x-\mu_X)f_X(x) \\
&= \left[1-F_X(x)\right] - \left[1-F_X(x)\right] - (x-\mu_X)f_X(x) \\
&= -(x-\mu_X)f_X(x).
\end{aligned}$$

This gives $\frac{\partial}{\partial x}\left\{\mathbb{E}\left[X-\mu_X|X\geq x\right]\left(1-F_X\left(x\right)\right)\right\}=0$ at $x=\mu_X$. This derivative is negative for $x>\mu_X$ and positive for $x<\mu_X$, hence it attains a maximum at $x=\mu_X$ and its minimum at the boundaries of the support of X.

B Identification proofs

Proof of Proposition 2.1: First we show that Assumptions 2.3 and 2.4 imply the density factorization

$$f_{\underline{A},U|\underline{T}}(\underline{a}_c,u_c|\underline{t}_c)=\left\{\prod_{j=1}^{N}f_A(a_{cj})\right\}f_U(u_c), \tag{24}$$

as claimed in the main text. We begin by considering the implications of the second part of Assumption 2.4. Start with the case where $N=2$ (i.e., groups consist of two individuals). In this case the second part of Assumption 2.4 becomes

$$A_{c2},T_{c2}\perp A_{c1}|\,T_{c1},\qquad A_{c1},T_{c1}\perp A_{c2}|\,T_{c2}.$$

The first of these conditions generates the density factorization

$$f_{\underline{A},\underline{T}}\left(a_{c1},a_{c2},t_{c1},t_{c2}\right)=f\left(a_{c1}|\,t_{c1}\right)f\left(a_{c2},t_{c1},t_{c2}\right),$$

while the second the factorization

$$f_{\underline{A},\underline{T}}\left(a_{c1},a_{c2},t_{c1},t_{c2}\right)=f\left(a_{c2}|\,t_{c2}\right)f\left(a_{c1},t_{c1},t_{c2}\right).$$

These two factorizations give the conditional density factorizations

$$f_{\underline{A}|\underline{T}}\left(a_{c1},a_{c2}|\,t_{c1},t_{c2}\right)=f\left(a_{c1}|\,t_{c1}\right)f\left(a_{c2}|\,t_{c1},t_{c2}\right)=f\left(a_{c2}|\,t_{c2}\right)f\left(a_{c1}|\,t_{c1},t_{c2}\right).$$

Integrating over a_{c1}we get

$$f\left(a_{c2}|\,t_{c1},t_{c2}\right)=f\left(a_{c2}|\,t_{c2}\right),$$

which after substitution yields

$$f_{\underline{A}|\underline{T}}\left(a_{c1},a_{c2}|\,t_{c1},t_{c2}\right)=f\left(a_{c1}|\,t_{c1}\right)f\left(a_{c2}|\,t_{c2}\right).$$

Now consider the $N=3$ case. The second part of Assumption 2.4 now implies

$$\begin{aligned}f\left(a_{c1},a_{c2},a_{c3},t_{c1},t_{c2},t_{c3}\right)&=f\left(a_{c1}|\,t_{c1}\right)f\left(a_{c2},a_{c3},t_{c1},t_{c2},t_{c3}\right)\\ f\left(a_{c1},a_{c2},a_{c3},t_{c1},t_{c2},t_{c3}\right)&=f\left(a_{c2}|\,t_{c2}\right)f\left(a_{c1},a_{c3},t_{c1},t_{c2},t_{c3}\right),\end{aligned}$$

so that after dividing through by $f\left(t_{c1},t_{c2},t_{c3}\right)$ and integrating over a_{c3} we get the equalities

$$\begin{aligned}f\left(a_{c1},a_{c2}|\,t_{c1},t_{c2},t_{c3}\right)&=f\left(a_{c1}|\,t_{c1}\right)f\left(a_{c2}|\,t_{c1},t_{c2},t_{c3}\right)\\ f\left(a_{c1},a_{c2}|\,t_{c1},t_{c2},t_{c3}\right)&=f\left(a_{c2}|\,t_{c2}\right)f\left(a_{c1}|\,t_{c1},t_{c2},t_{c3}\right).\end{aligned}$$

Integrating over a_{c1}then yields the equality

$$f\left(a_{c2}|\, t_{c1}, t_{c2}, t_{c3}\right) = f\left(a_{c2}|\, t_{c2}\right).$$

The same argument also gives the equalities

$$\begin{aligned} f\left(a_{c1}|\, t_{c1}, t_{c2}, t_{c3}\right) &= f\left(a_{c1}|\, t_{c1}\right) \\ f\left(a_{c3}|\, t_{c1}, t_{c2}, t_{c3}\right) &= f\left(a_{c3}|\, t_{c3}\right). \end{aligned}$$

Now observe that, using the results above,

$$\begin{aligned} f_{\underline{A}|\underline{T}}\left(a_{c1}, a_{c2}, a_{c3}|\, t_{c1}, t_{c2}, t_{c3}\right) &= f\left(a_{c1}|\, t_{c1}\right) f\left(a_{c2}, a_{c3}|\, t_{c1}, t_{c2}, t_{c3}\right) \\ &= f\left(a_{c1}|\, t_{c1}\right) f\left(a_{c2}|\, t_{c2}\right) f\left(a_{c3}|\, t_{c1}, t_{c2}, t_{c3}\right) \\ &= f\left(a_{c1}|\, t_{c1}\right) f\left(a_{c2}|\, t_{c2}\right) f\left(a_{c3}|\, t_{c3}\right). \end{aligned}$$

Finally Assumption 2.3 gives

$$f_{\underline{A}|\underline{T}}\left(a_{c1}, a_{c2}, a_{c3}|\, t_{c1}, t_{c2}, t_{c3}\right) = f\left(a_{c1}\right) f\left(a_{c2}\right) f\left(a_{c3}\right).$$

The above argument can be generalized to groups of arbitrary size such that under Assumption 2.4 we have

$$f_{\underline{A}|\underline{T}}(\underline{a}_c|\underline{t}_c) = \prod_{j=1}^{N} f_A(a_{cj}).$$

Now using the first part of Assumption 2.4 we get

$$f_{\underline{A},U|\underline{T}}\left(\underline{a}_c, u_c|\, \underline{t}_c\right) = f_{\underline{A}|\underline{T}}(\underline{a}_c|\underline{t}_c) f_U\left(u_c\right),$$

which along with the factorization of $f_{\underline{A}|\underline{T}}(\underline{a}_c|\underline{t}_c)$ derived above gives the needed result.

Now using (24) we can write

$$\begin{aligned} f_{\underline{A},U|\underline{T}}\left(\underline{a}_c, u_c|\, \underline{t}_c\right) &= f\left(a, a_{p(\cdot)}, u \middle|\, t, t_{p(\cdot)}\right) \\ &= f\left(a, a_{p(\cdot)}, u \middle|\, t, t_{p(\cdot)}, s\right) \\ &= f_A\left(a\right) \prod_{j \in p(\cdot)} f_A\left(a_{p(\cdot),j}\right) f_U\left(u\right) \\ &= f\left(a, a_{p(\cdot)}, u \middle|\, t, s\right), \end{aligned} \tag{25}$$

where the second equality follows from that fact that s_{-} is a deterministic function of $t_{p(\cdot)}$, the third equality from (24), and the fourth by inspection of the second and third density representations.

Recall that Assumptions 2.1 and 2.2 give $Y_i = Y_i(S_{-i}, \tau_{K_H}(A^H_{p(i)}), \tau_{K_L}(A^L_{p(i)}), U_i)$. Writing

$$Y_i(S_{-i}, \tau_{K_H}(A^H_{p(i)}), \tau_{K_L}(A^L_{p(i)}), U_i) = g(T_i, A_i, S_{-i}, \tau_{K_H}(A^H_{p(i)}), \tau_{K_L}(A^L_{p(i)}), U_i)$$

we therefore have, using (25),

$$
\begin{aligned}
\mathbb{E}\left[Y_i \middle| T_i = 1, S_i = s\right] &= \mathbb{E}\left[g(T_i, A_i, S_{-i}, \tau_{K_H}(A^H_{p(i)}), \tau_{K_L}(A^L_{p(i)}), U_i) \middle| T_i = 1, S_i = s\right] \\
&= \int \ldots \int g(1, a, s_-, \tau_{K_H}(a^H_{p(\cdot)}), \tau_{K_L}(a^L_{p(\cdot)}), u) f\left(a, a_{p(\cdot)}, u \middle| t, s\right) \mathrm{d}u \left(\prod_{j \in p(\cdot)} \mathrm{d}a_{p(\cdot),j} \right) \mathrm{d}a \\
&= \int \left\{ \int \ldots \int g(1, a, s_-, \tau_{K_H}(a^H_{p(\cdot)}), \tau_{K_L}(a^L_{p(\cdot)}), u) \right. \\
&\quad \times \left. \left\{ \prod_{j \in p(\cdot)} f_A\left(a_{p(\cdot),j}\right) \mathrm{d}a_{p(\cdot),j} \right\} f_U(u) \,\mathrm{d}u \right\} f_A(a) \,\mathrm{d}a.
\end{aligned}
$$

Let

$$
g^e(t, a, s_{-i}) = \int \ldots \int g(t, a, s_-, \tau_{K_H}(a^H_{p(\cdot)}), \tau_{K_L}(a^L_{p(\cdot)}), u) \left\{ \prod_{j \in p(\cdot)} f_A\left(a_{p(\cdot),j}\right) \mathrm{d}a_{p(\cdot),j} \right\} f_U(u) \,\mathrm{d}u.
$$

Observe that $g^e(T_i, A_i, s_{-i}) = Y_i^e(s_{-i})$, therefore by Assumption 2.3 we have

$$
\begin{aligned}
\int g^e(1, a, s_{-i}) f_A(a) \,\mathrm{d}a &= \int g^e(1, a, s_{-i}) f_{A|T}(a | T = 1) \,\mathrm{d}a \\
&= \mathbb{E}\left[Y_i^e(s_{-i}) \middle| T = 1\right] = m_H(s),
\end{aligned}
$$

as claimed. The result for $m_L(s)$ follows analogously. Identification of the two gradient function then follows directly from Assumption 2.5.

Feasibility of local reallocation density: Feasibility of (8) follows from the fact that, making the change of variables $v = (s + \lambda p_{H,\kappa}) / (1 + \lambda)$, and decomposing the integral,

$$
\begin{aligned}
\int_0^1 s f^{\mathrm{r}}_{S_c}(s; \lambda, \kappa) \mathrm{d}s &= \int_0^1 \frac{s}{1 + \lambda d_\kappa(s)} f^{\mathrm{sq}}_{S_c}\left(\frac{s + \lambda d_\kappa(s) p_{H,\kappa}}{1 + \lambda d_\kappa(s)} \right) \mathrm{d}s \\
&= \int_0^{\underline{s}+\kappa} s f^{\mathrm{sq}}_{S_c}(s) \,\mathrm{d}s + \int_{\underline{s}+\kappa}^{\overline{s}-\kappa} \frac{s}{1+\lambda} f^{\mathrm{sq}}_{S_c}\left(\frac{s + \lambda p_{H,\kappa}}{1+\lambda} \right) \mathrm{d}s + \int_{\overline{s}-\kappa}^1 s f^{\mathrm{sq}}_{S_c}(s) \,\mathrm{d}s \\
&= \Pr(S_i \le \underline{s} + \kappa) \,\mathbb{E}\left[T_i \middle| S_i \le \underline{s} + \kappa\right] \\
&+ \int_{\underline{s}+\kappa}^{\overline{s}-\kappa} \left\{ (1+\lambda) v - \lambda p_{H,\kappa} \right\} f^{\mathrm{sq}}_{S_c}(v) \mathrm{d}v \\
&+ \Pr(S_i \ge \overline{s} - \kappa) \,\mathbb{E}\left[T_i \middle| S_i \ge \overline{s} - \kappa\right] \\
&= \mathbb{E}\left[d_\kappa(S_i) \left\{ (1+\lambda) S_i - \lambda p_{H,\kappa} \right\} \right] \\
&+ \Pr(S_i \ge \overline{s} - \kappa) \,\mathbb{E}\left[T_i \middle| S_i \ge \overline{s} - \kappa\right] \\
&= p_H,
\end{aligned}
$$

as required.

Proof of Theorem 3.1: The result follows directly from Lemma A.1 above.

Proof of Theorem 6.1: Consider the problem

$$\max_{F_S(\cdot)\in\Gamma_S} \int M(s)\, f_S(s)\,\mathrm{d}s, \qquad \text{s.t.} \qquad \int s f_S(s)\,\mathrm{d}s = p_H, \tag{26}$$

where $M(s)$ is the concave envelope of $m(s)$ on $\mathcal{S}$. By concavity of $M(s)$ and Jensen's inequality we have

$$\int M(s)\, f_S(s)\,\mathrm{d}s \leq M\left(\mathbb{E}_F[S]\right).$$

Feasibility requires that $\mathbb{E}_F[S] = p_H$, therefore

$$\max_{F_S(\cdot)\in\Gamma_S} \int M(s)\, f_S(s)\,\mathrm{d}s \leq M(p_H). \tag{27}$$

Observe that this upper bound is attained by the degenerate distribution concentrated at p_H (i.e., $M^* = M(p_H)$).

Since $M(s) \geq m(s)$ for all $s \in \mathcal{S}$ we have the inequalities

$$M(p_H) \geq \int M(s)\, f_S(s)\,\mathrm{d}s \geq \int m(s)\, f_S(s)\,\mathrm{d}s,$$

for all feasible $F_S(\cdot)$. Therefore any feasible $F_S^*(s)$ such that $M(p_H) = \int m(s)\, f_S^*(s)\,\mathrm{d}s$ must be a solution to the planner's problem.

By the definition of $M(s)$, s_L and s_U we have that $M(s)$ is linear on the interval $s \in [s_L, s_U]$, i.e.,

$$M(s) = a + bs, \qquad s \in [s_L, s_U]$$

with

$$a = m(s_L) - \left(\frac{m(s_U) - m(s_L)}{s_U - s_L}\right) s_L, \qquad b = \frac{m(s_U) - m(s_L)}{s_U - s_L}.$$

This gives

$$\begin{aligned}
M(p_H) &= m(s_L) - \left(\frac{m(s_U) - m(s_L)}{s_U - s_L}\right) s_L + \left(\frac{m(s_U) - m(s_L)}{s_U - s_L}\right) p_H \\
&= m(s_L)\left(1 - \frac{p_H - s_L}{s_U - s_L}\right) + m(s_U)\,\frac{p_H - s_L}{s_U - s_L} \\
&= (1-\pi)\, m(s_L) + \pi m(s_L) \\
&= \int m(s)\, f_S^*(s)\,\mathrm{d}s.
\end{aligned}$$

Since $\int s f_S^*(s)\,\mathrm{d}s = p_H$, and therefore $F_S^*(s)$ feasible, we have that $F_S^*(s)$ is a solution to the planner's problem as claimed.

Proof of Corollary 6.1: Conditional on setting the fraction of high types assigned to groups of size n_j equal to σ_j we know, by Theorem 6.1, that $F_{S|N}^*(s|\, n_j)$ is an outcome-maximizing allocation. Since, again conditional on σ_j, $\int m(s, n_j)\, f_{S|N}^*(s|\, n_j)\,\mathrm{d}s = M(\sigma_j)$, we may therefore choose $\sigma_1, \ldots, \sigma_J$ by solving (22) which is concave by inspection.

Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach, supplemental material: proofs of Propositions 4.1 to 4.5

This appendix details the derivation of the influence functions associated with the estimators described in Section 4 of the main paper. Equation number continues in sequence with that of the main paper. All notation is as established in the main paper unless stated otherwise. In this appendix all expectations are with respect to the population of individuals unless noted otherwise. The i subscripts on random variables are omitted to simplify the notation.

We begin by noting that β^{ase}, β^{lsoe} and β^{lsie} are unrestricted parameters in the sense that their definitions do not place substantive restrictions on the joint distribution of $Z = (Y, T, S)'$.[23] Newey (1990, pp. 106 - 107) notes that the pathwise derivative of such unrestricted parameters will be unique. This implies that any regular estimator will have an influence function equal to the unique pathwise derivative. Furthermore, as described in Newey (1994a), the semiparametric efficiency bound for such parameters can be calculated as the variance of the pathwise derivative of the parameter with respect to the distribution of the data. The large sample characterization of the two-step M-estimators described in the main text follows from these observations. While we do not provide regularity conditions ensuring $\sqrt{C}$ consistency and asymptotic normality of our proposed estimators, our calculations do provide a formula for their large sample variance. Note that we do provide conditions that guarantee finiteness of the semiparametric efficiency bound, hence $\sqrt{C}$ consistency is achievable under suitable regularity conditions. Our approach is similar in spirit and implementation to that of Newey and Stoker (1994) in their analysis of weighted average derivatives.

To describe our calculations further we let $f(z)$ denote the true density of $Z = z$. A parametric submodel or path is a parametric family of densities $f(z; \eta)$ containing the 'truth' (i.e., $f(z; \eta_0) = f(z)$ for some η_0). Let $\beta(\eta)$ denote the population value of the parameter in question when Z is distributed according to $f(z; \eta)$. The pathwise derivative is the function $\phi(Z)$ such that

$$\nabla_\eta \beta(\eta)|_{\eta=\eta_0} = \mathbb{E}\left[\phi(Z)\, \mathbb{S}_\eta(Z)'\right] \tag{28}$$

where $\mathbb{S}_\eta(z) = \nabla_\eta f(z; \eta_0) / f(z; \eta_0)$ denotes the score of $f(z; \eta)$ at $\eta = \eta_0$.[24] By the delta method the Cramer-Rao variance bound for $\beta(\eta)$ in the parametric submodel is

$$\nabla_\eta \beta(\eta)\, \mathbb{E}\left[\mathbb{S}_\eta(Z)\, \mathbb{S}_\eta(Z)'\right]^{-1} \nabla_\eta \beta(\eta)' = \mathbb{E}\left[\phi(Z)\, \mathbb{S}_\eta(Z)'\right] \mathbb{E}\left[\mathbb{S}_\eta(Z)\, \mathbb{S}_\eta(Z)'\right]^{-1} \mathbb{E}\left[\mathbb{S}_\eta(Z)\, \phi(Z)'\right].$$

[23] In such models the allowable set of scores can approximate any mean zero function of Z (with finite variance).

[24] The form of (28) and a simple argument due to Newey (1990; ?) shows why $\phi(Z)$ is unique when β is an unrestricted parameter. Let $\phi(Z)$ and $\widetilde{\phi}(Z)$ denote two pathwise derivatives (centered to be mean zero), by (28) we have

$$\mathbb{E}\left[\left\{\phi(Z) - \widetilde{\phi}(Z)\right\} \mathbb{S}_\eta(Z)'\right] = 0.$$

When β is an unrestricted parameter the set of valid scores, or the tangent set, for the model is given by

Since $\mathbb{S}_\eta(z)$ is unrestricted the supremum of all such Cramer-Rao bounds, or the semiparametric variance bound, is obviously

$$\mathbb{E}\left[\phi(Z)\,\phi(Z)'\right].$$

By the arguments of Newey (1994) the asymptotic variance of any regular estimator of β is given by this bound.

The specific structure of each of our estimators can be used to simplify the calculation of $\phi(Z)$. In particular each of our estimators can be formulated as a two-step M-estimator with a nonparametric first step (see Newey and McFadden, 1994)). As shown by Newey (1994a) such problems have certain features which can be exploited in order to calculate the pathwise derivative. Let h be a function of Z, the arguments of which are suppressed in order to simplify notation; each our estimators can be defined as the solution to

$$\frac{1}{I}\sum\nolimits_{i=1}^{I}\psi(Z_i,\widehat{\beta},\widetilde{h})=0,$$

where $\psi(Z,\beta,h)$ is some known function and $\widetilde{h}$ is a preliminary 'first step' nonparametric estimate of h.

Let $\psi(Z,h)=\psi(Z,\beta_0,h)$. Application of the chain rule yields

$$\begin{aligned}\nabla_\eta\mathbb{E}_{\eta_0}\left[\psi(Z,h(\eta))\right]&=\int\nabla_\eta\psi(z,h(\eta))\,f(z)\,\mathrm{d}z+\int\psi(z,h_0)\,\mathbb{S}_\eta(z)'\,f(z)\,\mathrm{d}z\\&=\nabla_\eta\mathbb{E}_{\eta_0}\left[\psi(Z,h(\eta))\right]+\mathbb{E}_{\eta_0}\left[\psi(Z,h_0)\,\mathbb{S}_\eta(Z)'\right],\end{aligned}$$

where $\mathbb{E}_\eta[\cdot]$ denotes expectations taken with respect to the density $f(z;\eta)$ (throughout $\mathbb{E}[\cdot]=\mathbb{E}_{\eta_0}[\cdot]$). Noting that $\mathbb{E}_{\eta_0}\left[\psi(Z,\beta(\eta),h(\eta))\right]\big|_{\eta=\eta_0}=0$ a direct application of the implicit function theorem and the previous result then gives

$$\begin{aligned}\nabla_\eta\beta(\eta)\big|_{\eta=\eta_0}&=-\left[\nabla_\beta\mathbb{E}_{\eta_0}\left[\psi(Z,\beta_0,h(\eta_0))\right]\right]^{-1}\times\nabla_\eta\mathbb{E}_{\eta_0}\left[\psi(Z,\beta_0,h(\eta_0))\right]\\&=-\Gamma^{-1}\left\{\mathbb{E}\left[\psi(Z,h_0)\,\mathbb{S}_\eta(z)'\right]+\nabla_\eta\mathbb{E}\left[\psi(Z,h(\eta_0))\right]\right\}\end{aligned}$$

with $\Gamma=\nabla_\beta\mathbb{E}\left[\psi(Z,\beta,h_0)\right]\big|_{\beta=\beta_0}$ (assumed nonsingular). If we can find at function $\delta(z)$ such that

$$\nabla_\eta\mathbb{E}\left[\psi(Z,h(\eta))\right]=\mathbb{E}_\eta\left[\delta(Z)\,\mathbb{S}_\eta(Z)'\right],\tag{29}$$

then the influence function for any regular estimator of β, by the results of Newey (1990, 1994a) and equation (28) above, will be

$$\phi(Z)=-\Gamma^{-1}\left\{\psi(Z,h_0)+\delta(Z)\right\}.$$

$\mathcal{T}=\{\mathbb{S}_\eta(Z):\mathbb{E}[\mathbb{S}_\eta(Z)]=0\}$. Since $\phi(Z)-\widetilde{\phi}(Z)$ belongs to this set orthogonality requires that

$$\mathbb{E}\left[\left\{\phi(Z)-\widetilde{\phi}(Z)\right\}'\left\{\phi(Z)-\widetilde{\phi}(Z)\right\}\right]=0$$

or, equivalently, the equality $\phi(Z)=\widetilde{\phi}(Z)$. A simple intuition for this result, also due to Newey (1990), is that when the model places no restrictions on the distribution of the data β is just identified.

As explained by Newey (1994a) and also Newey and McFadden (1994), the function $\delta(Z)$ may be viewed a correction term which accounts for first step estimation of h. Below we use the structure of (29) to calculate the appropriate correction term for each of our estimators. In particular we begin by linearizing $\psi(z, h(\eta))$ around the truth h_0. With $\psi(z,h) - \psi(z,h_0) \simeq \Psi(z, h-h_0)$, $\Psi(z,h)$ linear in h, and (29) we then have

$$\nabla_\eta \mathbb{E}\left[\psi\left(Z, h(\eta) - h_0\right)\right] = \nabla_\eta \mathbb{E}\left[\Psi\left(Z, h(\eta)\right)\right] = \mathbb{E}_\eta\left[\delta(Z)\, \mathbb{S}_\eta(Z)'\right]. \tag{30}$$

Finding the form of $\delta(z)$ thus involves finding an 'integral representation' for $\mathbb{E}\left[\Psi\left(Z, h(\eta)\right)\right]$. The bulk of our derivations detailed below are devoted to this step.

Once the form of $\delta(Z)$ has been calculated, the asymptotic variance formulae given in Section 4 follow directly. A minor complication involves appropriately accounting for within-group dependence in the data induced by the presence of unobserved location-specific attributes. As noted by Newey (1994a, p. 1367), such dependence does not affect the form of $\delta(Z)$ and so can be accounted for relatively easily. Note that $\widehat{\beta}$ can be equivalently expressed as the solution to

$$\frac{1}{C}\sum_{c=1}^{C}\sum_{i\in\{i:G_i=c\}}\psi(Z_i,\widehat{\beta},\widetilde{h}) = 0,$$

with independence across groups so that the second step moment function is the within-group summation $g(Z_i,\beta,h) = \sum_{i\in\{i:G_i=c\}}\psi(Z_i,\beta,h)$. Let

$$G = \nabla_\beta \mathbb{E}\left[\sum\nolimits_{i\in\{i:G_i=c\}}\psi(Z_i,\beta,h_0)\right]\Bigg|_{\beta=\beta_0},$$

and

$$\widetilde{\phi}_c = -G^{-1}\left\{\sum\nolimits_{i\in\{i:G_i=c\}}\psi\left(Z_i,h_0\right) + \delta\left(Z_i\right)\right\},$$

so that the appropriate asymptotic sampling distribution is

$$\sqrt{C}(\widehat{\beta}-\beta_0) \xrightarrow{D} \mathcal{N}\left(0, \mathbb{E}\left[\widetilde{\phi}_c\widetilde{\phi}_c'\right]\right).$$

In all of the estimators considered here $G = -\mathbb{E}\left[N_i\right] = -\mu_N$, so that $\widetilde{\phi}_c = \mu_N^{-1}\left\{\sum\nolimits_{i\in\{i:G_i=c\}}\psi\left(Z_i,h_0\right)+\delta\left(Z_i\right)\right\}$.

B.1 Influence function derivation for $\widehat{\beta}^{\text{lsoe}}$

We begin with the local segregation outcome effect (LSOE) defined in Section 3:

$$\beta_0^{\text{lsoe}} = \mathbb{E}\left[d_\kappa(S)\nabla_s m(S)\left(S - p_{H,\kappa}\right)\right] = \mathbb{E}\left[d_\kappa(S)\nabla_s\left\{\frac{h_{10}(R)+h_{20}(R)}{h_{30}(R)}\right\}\left(S - \frac{h_{40}(R)}{h_{50}(R)}\right)\right],$$

where $R=(T,S)'$ (such that $Z=(Y,R')'$) and

$$
\begin{aligned}
h_{10}(r) &= f_S(s)\,\mathbb{E}\left[TY|\,S=s\right]=f_S(s)\,sm_H(s) \\
h_{20}(r) &= f_S(s)\,\mathbb{E}\left[(1-T)\,Y|\,S=s\right]=f_S(s)\,(1-s)\,m_L(s) \\
h_{30}(r) &= f_S(s) \\
h_{40}(r) &= \mathbb{E}\left[d_\kappa(S)\,T\right] \\
h_{50}(r) &= \mathbb{E}\left[d_\kappa(S)\right].
\end{aligned}
\tag{31}
$$

Let $h(r)=(h_1(r),h_2(r),h_3(r),h_4(r),h_5(r))'$. For what follows it is helpful to note that $f_{T|S}(t|\,s)=s^t(1-s)^{1-t}$.

The second step moment restriction defining β_0^{lsoe} is

$$\mathbb{E}\left[\psi\left(R,\beta_0^{\text{lsoe}},h_0\right)\right]=0,$$

with

$$\psi\left(r,\beta^{\text{lsoe}},h\right)=d_\kappa(s)\,\nabla_s\left\{\frac{h_1(r)+h_2(r)}{h_3(r)}\right\}\times\left(s-\frac{h_{40}(R)}{h_{50}(R)}\right)-\beta^{\text{lsoe}}.$$

Let $\psi\left(r,\beta_0^{\text{lsoe}},h\right)=\psi(r,h)$, linearizing $\psi(r,h)$ about h_0 gives

$$\psi(r,h)-\psi(r,h_0)\simeq\Psi(r,h-h_0),$$

where $\Psi(r,h-h_0)$ is linear in $h-h_0$. The precise form of $\Psi(r,h-h_0)$ is obtained by expanding the two ratios entering $\psi(R,h)$ pointwise. Since $a/b-a_0/b_0=b_0^{-1}\left[1-b^{-1}(b-b_0)\right]\left[(a-a_0)-(a_0/b_0)(b-b_0)\right]$, the linearization of a/b around a_0/b_0 is given by $b_0^{-1}\left[(a-a_0)-(a_0/b_0)(b-b_0)\right]$. This fact and the product rule allow us to write

$$
\begin{aligned}
\Psi(r,h-h_0) = {} & d_\kappa(s)\,\nabla_s\left\{\frac{1}{h_{30}(r)}\left[1,1,-\frac{h_{10}(r)+h_{20}(r)}{h_{30}(r)}\right]\begin{pmatrix} h_1(r)-h_{10}(r)\\ h_2(r)-h_{20}(r)\\ h_3(r)-h_{30}(r)\end{pmatrix}\right\} \\
& \times\left(s-\frac{h_{40}(r)}{h_{50}(r)}\right) \\
& -d_\kappa(s)\,\nabla_s\left\{\frac{h_{10}(r)+h_{20}(r)}{h_{30}(r)}\right\}\times\frac{1}{h_{50}(r)}\left\{1,-\frac{h_{40}(r)}{h_{50}(r)}\right\}\begin{pmatrix} h_4(r)-h_{40}(r)\\ h_5(r)-h_{50}(r)\end{pmatrix}.
\end{aligned}
$$

Differentiating the first term in $\{\cdot\}$ with respect to s, collecting terms, and rearranging yields

$$\Psi(r,h(r))=a_0(r)'\,h(r)+\nabla_s h(r)'\,b_0(r)+c_0(r)'\,h(r),\tag{32}$$

where

$$a_0(r) = d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} (-k(r), -k(r), -\nabla_s m(s) + m(s) k(r), 0, 0)'$$
$$b_0(r) = d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} (1, 1, -m(s), 0, 0)'$$
$$c_0(r) = -\frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]} \nabla_s m(s) (0, 0, 0, 1, -\mathbb{E}[T | d_\kappa(S) = 1])'$$

with

$$k(r) = \frac{\nabla_s f_S(s)}{f_S(s)}, \quad m(s) = \frac{h_{10}(r) + h_{20}(r)}{h_{30}(r)}.$$

As noted above the influence function for $\widehat{\beta}^{\text{lsoe}}$ will take the form $\psi(R, \gamma_0, h_0) + \delta(Z)$, where $\delta(Z)$ is the term which 'corrects' for first stage nonparametric estimation. From (32) and (30) this term solves

$$\nabla_\eta \mathbb{E}\left[a_0(R)' h(R;\eta)\right] + \nabla_\eta \mathbb{E}\left[\nabla_s h(R;\eta)' b_0(R)\right] + \nabla_\eta \mathbb{E}\left[c_0(R)' h(R;\eta)\right] = \mathbb{E}_\eta\left[\delta(Z) \mathbb{S}_\eta(Z)'\right].$$

To apply this result we begin by evaluating the expectations of on the left-hand-side of the above equation term-by-term. By iterated expectations we have, for the first term in (32),

$$\begin{aligned}
\mathbb{E}\left[a_0(R)' h(R;\eta)\right] &= \int d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} (-k(r), -k(r), -\nabla_s m(s) + m(s) k(r)) \\
&\times \begin{pmatrix} f_S(s;\eta) s \mathbb{E}_\eta[Y | T = 1, S = s] \\ f_S(s;\eta)(1-s)\mathbb{E}_\eta[Y | T = 0, S = s] \\ f_S(s;\eta) \end{pmatrix} f_0(r) \, \mathrm{d}r \\
&= \int d_\kappa(s)(s - p_{H,\kappa})(-k(r), -k(r), -\nabla_s m(s) + m(s) k(r)) \\
&\times \begin{pmatrix} \mathbb{E}_\eta[TY | S = s] \\ \mathbb{E}_\eta[(1-T) Y | S = s] \\ 1 \end{pmatrix} f_S(s;\eta) \, \mathrm{d}s \\
&= \mathbb{E}_\eta\left[v_1(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}'\right],
\end{aligned}$$

where the second equality follows from the fact that $f_{T|S}(t | s;\eta) = s^t (1-s)^{1-t}$ does not depend on η and

$$v_1(r) = (s - p_{H,\kappa}) \{-\nabla_s m(s) + m(s) k(r), 0, -k(r), -k(r)\}'.$$

To evaluate the second term of (32) we use integration by parts as in Powell et al. (1989) (with $u(r) = f_0(r) b_0(r)'$ and $v(r) = h(r;\eta)$) to obtain a representation directly in terms of $h(r;\eta)$. Using the fact that $b_0(r)$ and $h(r;\eta)$ vary in s alone, as well as the density factorization $f_0(r) = s^t (1-s)^{1-t} f_0(s)$, we have

$$
\begin{aligned}
\mathbb{E}\left[\nabla_s h(R;\eta)' b_0(R)\right] &= \int f_0(r)\, b_0(r)' \left[\nabla_s h(r;\eta)\right] \mathrm{d}r \\
&= \int_{s=0}^{s=1} \sum_{t=0,1} s^t (1-s)^{1-t} f_0(s)\, b_0(r)' \left[\nabla_s h(r;\eta)\right] \mathrm{d}s \\
&= \int_{s=0}^{s=1} f_0(s)\, b_0(r)' \left[\nabla_s h(r;\eta)\right] \mathrm{d}s \\
&= \left[f_0(s)\, b_0(r)' h(r;\eta)\right]_0^1 - \int_{s=0}^{s=1} \nabla_s \left[f_0(s)\, b_0(r)'\right] h(r;\eta)\, \mathrm{d}s \\
&= 0 - \int_{s=0}^{s=1} \nabla_s \left[f_0(s)\, b_0(r)'\right] h(r;\eta)\, \mathrm{d}r \\
&= \mathbb{E}_\eta\left[v_2(R)\, d_\kappa(S) \left\{1, T, TY, (1-T)Y\right\}'\right],
\end{aligned}
$$

with

$$
v_2(r) = (s - p_{H,\kappa}) \left\{\nabla_s m(s), 0, 0, 0\right\}' + \left\{m(s), 0, -1, -1\right\}'.
$$

This follows from the fact that $f_0(s)\, b_0(r) = 0$ at $s = 0, 1$ since $d_\kappa(0) = d_\kappa(1) = 0$ and also that

$$
\begin{aligned}
\nabla_s \left[f_0(s)\, b_0(r)'\right] &= \nabla_s \left[d_\kappa(s)(s - p_{H,\kappa}) \left\{1, 1, -m(s), 0, 0\right\}'\right] \\
&= d_\kappa(s)(s - p_{H,\kappa}) \left\{0, 0, -\nabla_s m(s), 0, 0\right\}' \\
&+ d_\kappa(s) \left\{1, 1, -m(s), 0, 0\right\}'.
\end{aligned}
$$

Finally we take the expectation of the final term in (32):

$$
\begin{aligned}
\mathbb{E}\left[c_0(R)' h(R;\eta)\right] &= -\int \frac{d_\kappa(s)}{\mathbb{E}\left[d_\kappa(S)\right]} \nabla_s m(s) \left(0, 0, 0, 1, -\mathbb{E}\left[T \mid d_\kappa(S) = 1\right]\right)' h(r;\eta)\, f_0(r)\, \mathrm{d}r \\
&= -\int \frac{d_\kappa(s)}{\mathbb{E}\left[d_\kappa(S)\right]} \nabla_s m(s) \left(\mathbb{E}_\eta\left[d_\kappa(S)\, T\right] - \mathbb{E}_\eta\left[d_\kappa(S)\right] \mathbb{E}\left[T \mid d_\kappa(S) = 1\right]\right) f_0(r)\, \mathrm{d}r \\
&= -\left\{\int \frac{d_\kappa(s)}{\mathbb{E}\left[d_\kappa(S)\right]} \nabla_s m(s)\, f_0(r)\, \mathrm{d}r\right\} \mathbb{E}_\eta\left[d_\kappa(S)(T - p_{H,\kappa})\right] \\
&= -\mathbb{E}\left[\nabla_s m(s) \mid d_\kappa(S) = 1\right] \mathbb{E}_\eta\left[d_\kappa(S)(T - p_{H,\kappa})\right] \\
&= \mathbb{E}_\eta\left[v_3(R)\, d_\kappa(S) \left\{1, T, TY, (1-T)Y\right\}'\right],
\end{aligned}
$$

where

$$
v_3(R) = \mathbb{E}\left[\nabla_s m(s) \mid d_\kappa(S) = 1\right] \left\{p_{H,\kappa}, -1, 0, 0\right\}'.
$$

Combining terms gives

$$
\mathbb{E}\left[\Psi(R, h_0)\right] = \mathbb{E}_\eta\left[v(R)\, d_\kappa(S) \left\{1, T, TY, (1-T)Y\right\}'\right],
$$

with $v(r) = v_1(r) + v_2(r) + v_3(r)$ or, equivalently,

$$v(r) = \{m(s) + (s - p_{H,\kappa}) m(s) k(r) + \mathbb{E}[\nabla_s m(s) | d_\kappa(S) = 1] p_{H,\kappa},$$
$$-\mathbb{E}[\nabla_s m(s) | d_\kappa(S) = 1], -(s - p_{H,\kappa}) k(r) - 1, -(s - p_{H,\kappa}) k(r) - 1\}.$$

Differentiating with respect to η gives

$$\nabla_\eta \mathbb{E}_\eta\left[v(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}'\right] = \mathbb{E}_\eta\left[v(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}' \mathbb{S}'_\eta\right]$$

and hence a correction term of $\delta(Z) = v(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}'$ or

$$\delta^{\text{lsoe}}(z) = -d_\kappa(s) \frac{\nabla_s f_S(s)}{f_S(s)} (y - m(s)) (s - p_{H,\kappa}) - d_\kappa(s) (y - m(s)) - \mathbb{E}[\nabla_s m(s) | d_\kappa(S) = 1] d_\kappa(s) (t - p_{H,\kappa}), \tag{33}$$

as claimed.

B.2 Influence function derivation for $\widehat{\alpha}^{\text{lppe}}$

The local private peer effect (LPPE) of Section 3 is given by

$$\alpha_0^{\text{lppe}} = \mathbb{E}\left[d_\kappa(S) \left(\frac{h_{10}(R)}{s h_{30}(R)} - \frac{h_{20}(R)}{(1-s) h_{30}(R)}\right) \left(S - \frac{h_{40}(R)}{h_{50}(R)}\right)\right],$$

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r), h_5(r))'$ as defined in (31) above. Linearizing the implied moment function gives

$$\Psi(r, h(r)) = a_0(r)' h(r) + b_0(r)' h(r), \tag{34}$$

where

$$a_0(r) = \frac{d_\kappa(s)}{f_S(s)} \left\{\frac{s - p_{H,\kappa}}{s}, -\frac{s - p_{H,\kappa}}{1-s}, -[m_H(s) - m_L(s)] (s - p_{H,\kappa}), 0, 0\right\},$$
$$b_0(r) = -d_\kappa(s) \frac{m_H(s) - m_L(s)}{\mathbb{E}[d_\kappa(S)]} \{0, 0, 0, 1, -p_{H,\kappa}\}.$$

Taking expectations of $a_0(R)' h(R;\eta)$ yields

$$\begin{aligned}
\mathbb{E}\left[a_0(R)' h(R;\eta)\right] &= \int \frac{d_\kappa(s)}{f_S(s)} \left\{ \frac{s-p_{H,\kappa}}{s}, -\frac{s-p_{H,\kappa}}{1-s}, -\left[m_H(s)-m_L(s)\right](s-p_{H,\kappa}) \right\} \\
&\quad \times \begin{pmatrix} f_S(s;\eta)\,\mathbb{E}_\eta\left[TY|\,S=s\right] \\ f_S(s;\eta)\,\mathbb{E}_\eta\left[(1-T)\,Y|\,S=s\right] \\ f_S(s;\eta) \end{pmatrix} f_S(s)\,\mathrm{d}s \\
&= \int d_\kappa(s) \left\{ \frac{s-p_{H,\kappa}}{s}, -\frac{s-p_{H,\kappa}}{1-s}, -\left[m_H(s)-m_L(s)\right](s-p_{H,\kappa}) \right\} \\
&\quad \times \begin{pmatrix} \mathbb{E}_\eta\left[TY|\,S=s\right] \\ \mathbb{E}_\eta\left[(1-T)\,Y|\,S=s\right] \\ 1 \end{pmatrix} f_S(s;\eta)\,\mathrm{d}s \\
&= \mathbb{E}_\eta\left[\upsilon_1(R)\,d_\kappa(S)\left\{1,T,TY,(1-T)\,Y\right\}'\right],
\end{aligned}$$

where

$$\upsilon_1(r) = \left\{ -\left[m_H(s)-m_L(s)\right](s-p_{H,\kappa}), 0, \frac{s-p_{H,\kappa}}{s}, -\frac{s-p_{H,\kappa}}{1-s} \right\}.$$

Now taking expectations of $b_0(R)' h(R;\eta)$ we get

$$\begin{aligned}
\mathbb{E}\left[b_0(R)' h(R;\eta)\right] &= -\int d_\kappa(s)\,\frac{m_H(s)-m_L(s)}{\mathbb{E}\left[d_\kappa(S)\right]}\left\{1,-p_{H,\kappa}\right\} \begin{pmatrix} \mathbb{E}_\eta\left[d_\kappa(S)\,T\right] \\ \mathbb{E}_\eta\left[d_\kappa(S)\right] \end{pmatrix} f_S(s)\,\mathrm{d}s \\
&= -\left[\int d_\kappa(s)\,\frac{m_H(s)-m_L(s)}{\mathbb{E}\left[d_\kappa(S)\right]} f_S(s)\,\mathrm{d}s\right] \times \mathbb{E}_\eta\left[d_\kappa(S)\,(T-p_{H,\kappa})\right] \\
&= -\mathbb{E}\left[\frac{d_\kappa(S)}{\mathbb{E}\left[d_\kappa(S)\right]}\left(m_H(S)-m_L(S)\right)\right] \times \mathbb{E}_\eta\left[d_\kappa(S)\,(T-p_{H,\kappa})\right] \\
&= -\mathbb{E}\left[m_H(S)-m_L(S)|\,d_\kappa(S)=1\right] \times \mathbb{E}_\eta\left[d_\kappa(S)\,(T-p_{H,\kappa})\right] \\
&= -\mathbb{E}_\eta\left[\upsilon_2(R)\,d_\kappa(S)\left\{1,T,TY,(1-T)\,Y\right\}'\right],
\end{aligned}$$

with

$$\upsilon_2(r) = \mathbb{E}\left[m_H(S)-m_L(S)|\,d_\kappa(S)=1\right]\left\{-p_{H,\kappa},1\right\}.$$

Using (34) and (30), these calculations suggest a correction term of the form

$$\begin{aligned}
\delta^{\mathrm{lppe}}(z) &= d_\kappa(s)\left\{\left(\frac{t}{s}\right)y - m_H(s)\right\}(s-p_{H,\kappa}) - d_\kappa(s)\left\{\left(\frac{1-t}{1-s}\right)y - m_L(s)\right\}(s-p_{H,\kappa}) \\
&\quad - \mathbb{E}\left[m_H(S)-m_L(S)|\,d_\kappa(S)=1\right] d_\kappa(s)\,(t-p_{H,\kappa}),
\end{aligned} \tag{35}$$

as claimed.

B.3 Influence function derivation for $\widehat{\alpha}^{\text{lepe}}$

The local external peer effect (LEPE) of Section 3 is given by

$$\alpha_0^{\text{lepe}} = \mathbb{E}\left[d_\kappa (S)\left(S\nabla_s\left\{\frac{h_{10}(R)}{Sh_{30}(R)}\right\} + (1-S)\nabla_s\left\{\frac{h_{20}(R)}{(1-s)\,h_{30}(R)}\right\}\right)\left(S - \frac{h_{40}(R)}{h_{50}(R)}\right)\right]$$

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r), h_5(r))'$ as defined in (31) above. Linearizing the implied moment function gives

$$\begin{aligned}\Psi(r, h(r)) = {} & d_\kappa(s)\, s\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{s}, -m_H(s)\right\}\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s - p_{H,\kappa}) \\ & + d_\kappa(s)(1-s)\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{1-s}, -m_L(s)\right\}\begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s - p_{H,\kappa}) \\ & - d_\kappa(s)\,(s\nabla_s m_H(s) + (1-s)\nabla_s m_H(s))\,\frac{1}{h_{50}(r)}\left\{1, -\frac{h_{40}(r)}{h_{50}(r)}\right\}\begin{pmatrix} h_4(r) - h_{40}(r) \\ h_5(r) - h_{50}(r)\end{pmatrix}.\end{aligned}$$

By the chain rule we have

$$\begin{aligned} & d_\kappa(s)\, s\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{s}, -m_H(s)\right\}\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s - p_{H,\kappa}) \\ & \qquad = d_\kappa(s)\,\nabla_s\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}'\left\{\frac{s - p_{H,\kappa}}{f_S(s)}\,[1, -s m_H(s)]\right\}' \\ & \quad + d_\kappa(s)\,\frac{s - p_{H,\kappa}}{f_S(s)}\left\{\left[-\frac{1}{s} - k(r), k(r)\, s m_H(s) - s\nabla_s m_H(s)\right]\right\}\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix},\end{aligned}$$

where $k(r) = \nabla_s f_S(s)/f_S(s)$ as above. Similarly we have

$$\begin{aligned} & d_\kappa(s)(1-s)\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{1-s}, -m_L(s)\right\}\begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s - p_{H,\kappa}) \\ & \qquad = d_\kappa(s)\,\nabla_s\begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}'\left\{\frac{s - p_{H,\kappa}}{f_S(s)}\,[1, -(1-s)\, m_L(s)]\right\}' \\ & + d_\kappa(s)\,\frac{s - p_{H,\kappa}}{f_S(s)}\left\{\frac{1}{1-s} - k(r), k(r)(1-s)\, m_L(s) - (1-s)\nabla_s m_L(s)\right\}\begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}.\end{aligned}$$

Collecting terms and reorganizing yields the linearization

$$\Psi(r, h(r)) = a_0(r)'\, h(r) + \nabla_s h(r)'\, b_0(r) + c_0(r)'\, h(r), \tag{36}$$

where

$$a_0(r) = d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} \left\{ -\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r) m(s) - e(s), 0, 0 \right\},$$

$$b_0(r) = d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} \{1, 1, -m(s), 0, 0\},$$

$$c_0(r) = -\frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]} e(s) \{0, 0, 0, 1, -p_{H,\kappa}\},$$

recalling that $e(s) = s\nabla_s m_H(s) + (1-s)\nabla_s m_L(s)$.

Evaluating the expectation of $\mathbb{E}\left[a_0(R)' h(R;\eta)\right]$ yields

$$\begin{aligned}
\mathbb{E}\left[a_0(R)' h(R;\eta)\right] &= \int d_\kappa(s) \frac{s - p_{H,\kappa}}{f_S(s)} \left[-\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r) m(s) - e(s) \right], \\
&\times \begin{pmatrix} f_S(s;\eta) \mathbb{E}_\eta[TY | S = s] \\ f_S(s;\eta) \mathbb{E}_\eta[(1-T) Y | S = s] \\ f_S(s;\eta) \end{pmatrix} f_S(s) \,\mathrm{d}s \\
&= \int d_\kappa(s)(s - p_{H,\kappa}) \left[-\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r) m(s) - e(s) \right] \\
&\times \begin{pmatrix} \mathbb{E}_\eta[TY | S = s] \\ \mathbb{E}_\eta[(1-T) Y | S = s] \\ 1 \end{pmatrix} f_S(s;\eta) \,\mathrm{d}s \\
&= \mathbb{E}_\eta\left[\upsilon_1(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}' \right],
\end{aligned}$$

where

$$\begin{aligned}
\upsilon_1(r) = \Bigg\{ &(s - p_H) \left[\frac{\nabla_s f_S(s)}{f_S(s)} m(s) - e(s) \right], 0, \\
&(s - p_H) \left(-\frac{1}{s} - \frac{\nabla_s f_S(s)}{f_S(s)} \right), (s - p_H) \left(\frac{1}{1-s} - \frac{\nabla_s f_S(s)}{f_S(s)} \right) \Bigg\}.
\end{aligned}$$

From the analysis of $\widehat{\beta}^{\text{lsoe}}$ above we have

$$\mathbb{E}\left[\nabla_s h(R;\eta)' b_0(R)\right] = \mathbb{E}_\eta\left[\upsilon_2(R) d_\kappa(S) \{1, T, TY, (1-T) Y\}' \right],$$

with

$$\upsilon_2(r) = (s - p_{H,\kappa}) \{\nabla_s m(s), 0, 0, 0\}' + \{m(s), 0, -1, -1\}'.$$

Finally we evaluate the expectation of $c_0(R)' h(R;\eta)$:

$$
\begin{aligned}
\mathbb{E}\left[c_0(R)' h(R;\eta)\right] &= -\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(s)]} e(s)\left(0,0,0,1,-\mathbb{E}[T|\, d_\kappa(S)=1]\right)' h(r;\eta)\, f_0(r)\,\mathrm{d}r \\
&= -\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(s)]} e(s)\left(\mathbb{E}_\eta[d_\kappa(S)\,T] - \mathbb{E}_\eta[d_\kappa(S)]\,\mathbb{E}[T|\, d_\kappa(S)=1]\right) f_0(r)\,\mathrm{d}r \\
&= -\left\{\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(s)]} e(s)\, f_0(r)\,\mathrm{d}r\right\} \mathbb{E}_\eta\left[d_\kappa(S)\,(T-p_{H,\kappa})\right] \\
&= -\mathbb{E}[e(s)|\, d_\kappa(s)=1]\,\mathbb{E}_\eta\left[d_\kappa(S)\,(T-p_{H,\kappa})\right] \\
&= \mathbb{E}_\eta\left[v_3(R)\, d_\kappa(S)\left\{1,T,TY,(1-T)\,Y\right\}'\right],
\end{aligned}
$$

where

$$
v_3(R) = \mathbb{E}[e(S)|\, d_\kappa(S)=1]\left\{p_{H,\kappa},-1,0,0\right\}'.
$$

The form of $v_1(r)$, $v_2(r)$ and $v_3(r)$ together imply a correction term of

$$
\begin{aligned}
\delta^{\mathrm{lepe}}(z) = &-d_\kappa(s)\,\frac{\nabla_s f_S(s)}{f_S(s)}\,(y-m(s))\,(s-p_{H,\kappa}) \\
&- d_\kappa(s)\,(y-m(s)) \\
&- d_\kappa(s)\left\{\left(\frac{t}{s}\right) y - m_H(s)\right\}(s-p_{H,\kappa}) + d_\kappa(s)\left\{\left(\frac{1-t}{1-s}\right) y - m_L(s)\right\}(s-p_{H,\kappa}) \\
&- \mathbb{E}[e(S)|\, d_\kappa(S)=1]\, d_\kappa(s)\,(t-p_{H,\kappa}),
\end{aligned}
$$

as claimed. Note that $\delta^{\mathrm{lppe}}(z) + \delta^{\mathrm{lepe}}(z) = \delta^{\mathrm{lsoe}}(z)$ as would be expected.

B.4 Influence function derivation for $\widehat{\beta}^{\mathrm{ase}}$

The average spillover effect is given by

$$
\begin{aligned}
\beta_0^{\mathrm{ase}} &= \mathbb{E}[d_\kappa(S)\, e(S)] \\
&= \mathbb{E}\left[d_\kappa(S)\,\nabla_s\left\{\frac{h_{10}(R)+h_{20}(R)}{h_{30}(R)}\right\} - \frac{d_\kappa(S)}{h_{30}(R)}\left\{\frac{h_{10}(R)}{S} - \frac{h_{20}(R)}{1-S}\right\}\right],
\end{aligned}
$$

where $h_{10}(r)$, $h_{20}(r)$ and $h_{30}(r)$ are as defined in (31) above. Linearizing the implied moment function gives

$$
\begin{aligned}
\Psi(r, h(r)-h_0(r)) = {} & d_\kappa(s)\,\nabla_s\left\{\frac{1}{h_{30}(r)}\left[1,1,-\frac{h_{10}(r)+h_{20}(r)}{h_{30}(r)}\right]\begin{pmatrix} h_1(r)-h_{10}(r) \\ h_2(r)-h_{20}(r) \\ h_3(r)-h_{30}(r)\end{pmatrix}\right\} \\
&- \frac{d_\kappa(s)}{h_{30}(r)}\left\{\frac{1}{s},-\frac{1}{1-s},-\left[\frac{h_{10}(r)}{s h_{30}(r)} - \frac{h_{20}(r)}{(1-s)\,h_{30}(r)}\right]\right\}\left\{\begin{matrix} h_1(r)-h_{10}(r) \\ h_2(r)-h_{20}(r) \\ h_3(r)-h_{30}(r)\end{matrix}\right\}.
\end{aligned}
$$

Differentiating the first term in $\{\cdot\}$ with respect to s and collecting terms yields

$$\Psi\left(r,h\left(r\right)\right)=a_{0}\left(r\right)'h\left(r\right)+\nabla_{s}h\left(r\right)'b_{0}\left(r\right)+c_{0}\left(r\right)'h\left(r\right) \tag{37}$$

with $h\left(r\right)=\left(h_{1}\left(r\right),h_{2}\left(r\right),h_{3}\left(r\right)\right)'$ and

$$\begin{aligned}
a_{0}\left(r\right)&=\frac{d_{\kappa}\left(s\right)}{f_{S}\left(s\right)}\left(-k\left(r\right),-k\left(r\right),-\nabla_{s}m\left(s\right)+m\left(s\right)k\left(r\right)\right)',\\
b_{0}\left(r\right)&=\frac{d_{\kappa}\left(s\right)}{f_{S}\left(s\right)}\left(1,1,-m\left(s\right)\right)',\\
c_{0}\left(r\right)&=-\frac{d_{\kappa}\left(s\right)}{f_{S}\left(s\right)}\left(\frac{1}{s},-\frac{1}{1-s},-\left[m_{H}\left(s\right)-m_{L}\left(s\right)\right]\right),
\end{aligned}$$

where

$$k\left(r\right)=\frac{\nabla_{s}f_{S}\left(s\right)}{f_{S}\left(s\right)},\quad m\left(s\right)=\frac{h_{10}\left(r\right)+h_{20}\left(r\right)}{h_{30}\left(r\right)},\quad m_{H}\left(s\right)=\frac{h_{10}\left(r\right)}{sh_{30}\left(r\right)},\quad m_{L}\left(s\right)=\frac{h_{20}\left(r\right)}{\left(1-s\right)h_{30}\left(r\right)}.$$

Taking expectations of the first term in $\Psi\left(r,h\left(r\right)\right)$ we have

$$\begin{aligned}
\mathbb{E}\left[a_{0}\left(R\right)'h\left(R;\eta\right)\right]&=\int\frac{d_{\kappa}\left(s\right)}{f_{S}\left(s\right)}\left(-k\left(r\right),-k\left(r\right),-\nabla_{s}m\left(s\right)+m\left(s\right)k\left(r\right)\right)\\
&\quad\times\begin{pmatrix}f_{S}\left(s;\eta\right)s\mathbb{E}_{\eta}\left[Y\right|T=1,S=s]\\ f_{S}\left(s;\eta\right)\left(1-s\right)\mathbb{E}_{\eta}\left[Y\right|T=0,S=s]\\ f_{S}\left(s;\eta\right)\end{pmatrix}f_{0}\left(r\right)\mathrm{d}r\\
&=\int d_{\kappa}\left(s\right)\left(-k\left(r\right),-k\left(r\right),-\nabla_{s}m\left(s\right)+m\left(s\right)k\left(r\right)\right)\\
&\quad\times\begin{pmatrix}\mathbb{E}_{\eta}\left[TY\right|S=s]\\ \mathbb{E}_{\eta}\left[\left(1-T\right)Y\right|S=s]\\ 1\end{pmatrix}f_{S}\left(s;\eta\right)\mathrm{d}s\\
&=\mathbb{E}_{\eta}\left[v_{1}\left(R\right)d_{\kappa}\left(S\right)\left\{1,T,TY,\left(1-T\right)Y\right\}'\right],
\end{aligned}$$

where the second equality follows from the fact that $f_{T|S}\left(t|\,s;\eta\right)=s^{t}\left(1-s\right)^{1-t}$ does not depend on η and

$$v_{1}\left(r\right)=\left\{-\nabla_{s}m\left(s\right)+m\left(s\right)k\left(r\right),0,-k\left(r\right),-k\left(r\right)\right\}.$$

To evaluate the second term of (37) we use integration by parts (with $u(r)=f_{0}\left(r\right)b_{0}\left(r\right)'$ and $v\left(r\right)=h\left(r;\eta\right)$) to obtain a representation directly in terms of $h\left(r;\eta\right)$. As in our analysis of β^{lsoe} above we use the fact that $b_{0}\left(r\right)$ and $h\left(r;\eta\right)$ vary in s alone, as well as the density factorization $f_{0}\left(r\right)=s^{t}\left(1-s\right)^{1-t}f_{0}\left(s\right)$, to get

$$\begin{aligned}
\mathbb{E}\left[\nabla_{s}h\left(R;\eta\right)'b_{0}\left(R\right)\right]&=0-\int\nabla_{s}\left[f_{0}\left(s\right)b_{0}\left(r\right)'\right]h\left(r;\eta\right)\mathrm{d}s\\
&=\mathbb{E}_{\eta}\left[v_{2}\left(R\right)d_{\kappa}\left(S\right)\left\{1,T,TY,\left(1-T\right)Y\right\}'\right],
\end{aligned}$$

with

$$\upsilon_2(r) = \{\nabla_s m(s), 0, 0, 0\}'.$$

This follows from the fact that

$$\begin{aligned}\nabla_s \left[f_0(s) b_0(r)'\right] &= \nabla_s \left[d_\kappa(s)(1, 1, -m(s))'\right] \\ &= d_\kappa(s)(0, 0, -\nabla_s m(s))'.\end{aligned}$$

Evaluating the expectation of the third term in (37) gives

$$\begin{aligned}\mathbb{E}\left[c_0(R)' h(R;\eta)\right] &= -\int \frac{d_\kappa(s)}{f_S(s)} \left(\frac{1}{s}, -\frac{1}{1-s}, -\left[m_H(s) - m_L(s)\right]\right) \\ &\quad \times \begin{pmatrix} f_S(s;\eta) s m_H(s;\eta) \\ f_S(s;\eta)(1-s) m_L(s;\eta) \\ f_S(s;\eta) \end{pmatrix} f_S(s)\,\mathrm{d}s \\ &= -\int d_\kappa(s) \left(\frac{1}{s}, -\frac{1}{1-s}, -\left[m_H(s) - m_L(s)\right]\right) \\ &\quad \times \begin{pmatrix} \mathbb{E}_\eta\left[TY \mid S = s\right] \\ \mathbb{E}_\eta\left[(1-T)Y \mid S = s\right] \\ 1 \end{pmatrix} f_S(s;\eta)\,\mathrm{d}s \\ &= \mathbb{E}_\eta\left[\upsilon_3(R) d_\kappa(S) \{1, T, TY, (1-T)Y\}'\right],\end{aligned}$$

with

$$\upsilon_3(r) = \left\{m_H(s) - m_L(s), 0, -\frac{1}{s}, \frac{1}{1-s}\right\}.$$

Together these calculations suggest a correction term of the form

$$\begin{aligned}\delta^{\mathrm{ase}}(z) &= -d_\kappa(s) \frac{\nabla_s f_S(s)}{f_S(s)} (y - m(s)) \\ &\quad - d_\kappa(s) \left[\left\{\left(\frac{t}{s}\right) y - m_H(s)\right\} - \left\{\left(\frac{1-t}{1-s}\right) y - m_L(s)\right\}\right]\end{aligned} \tag{38}$$

as claimed.

B.5 Influence function derivation for $\widehat{\beta}^{\mathrm{lsie}}$

The local segregation inequality effect (LSIE) is given by

$$\beta_0^{\mathrm{lsie}} = \beta_H^{\mathrm{lsie}} - \beta_L^{\mathrm{lsie}}$$

where

$$\beta_H^{\text{lsie}} = \mathbb{E}\left[\frac{d_\kappa(S)}{p_{H,\kappa}}\{m_H(S) + S\nabla_s m_H(S)\}(S - p_{H,\kappa})\right]$$
$$= \mathbb{E}\left[d_\kappa(S)\left\{\frac{1}{h_{40}(R)/h_{50}(R)}\left(\frac{h_{10}(R)}{Sh_{30}(R)} + S\nabla_s\left\{\frac{h_{10}(R)}{Sh_{30}(R)}\right\}\right)\left(S - \frac{h_{40}(R)}{h_{50}(R)}\right)\right\}\right]$$

and

$$\beta_L^{\text{lsie}} = \mathbb{E}\left[\frac{d_\kappa(S)}{1 - p_{H,\kappa}}\{-m_L(S) + (1-S)\nabla_s m_L(S)\}(S - p_{H,\kappa})\right]$$
$$= \mathbb{E}\left[\frac{d_\kappa(S)}{1 - h_{40}(R)/h_{50}(R)}\left(-\frac{h_{20}(R)}{(1-S)h_{30}(R)} + (1-S)\nabla_s\left\{\frac{h_{20}(R)}{(1-S)h_{30}(R)}\right\}\right)\left(S - \frac{h_{40}(R)}{h_{50}(R)}\right)\right],$$

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r), h_5(r))'$ as defined in (31) above.

We begin by analyzing the first component of the estimand, β_H^{lsie}. Linearizing the moment defining β_H^{lsie} we get

$$\begin{aligned}
\Psi(r, h(r) - h_0(r)) = \\
& d_\kappa(s)\left\{\frac{1}{f_S(s)\,s}\frac{s - p_{H,\kappa}}{p_{H,\kappa}}, -\frac{1}{f_S(s)}\frac{s - p_{H,\kappa}}{p_{H,\kappa}}m_H(s),\right. \\
& \left.-\frac{1}{\mathbb{E}[d_\kappa(S)]}\frac{s}{p_{H,\kappa}^2}[m_H(s) + s\nabla_s m_H(s)], \frac{1}{\mathbb{E}[d_\kappa(S)]}\frac{s}{p_{H,\kappa}}[m_H(s) + s\nabla_s m_H(s)]\right\} \\
& \times\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \\ h_4(r) - h_{40}(r) \\ h_5(r) - h_{50}(r)\end{pmatrix} \\
& + d_\kappa(s)\frac{s}{p_{H,\kappa}}\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{s}, -m_H(s)\right\}\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s - p_{H,\kappa}).
\end{aligned}$$

Differentiating the second term in $\{\cdot\}$ with respect to s yields

$$\begin{aligned}
d_\kappa(s)\frac{s}{p_{H,\kappa}}\nabla_s\left\{\frac{1}{f_S(s)}\left\{\frac{1}{s}, -m_H(s)\right\}\begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}\right\}(s - p_{H,\kappa}) \\
= \nabla_s\begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}'\left[d_\kappa(s)\frac{s}{p_{H,\kappa}}\frac{s - p_H}{f_S(s)}\left\{\frac{1}{s}, -m_H(s)\right\}\right]' \\
+ \begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}' d_\kappa(s)\frac{s}{p_{H,\kappa}}\frac{s - p_{H,\kappa}}{f_S(s)\,s}\{-1/s - k(r), k(r)\,s m_H(s) - s\nabla_s m_H(s)\}',
\end{aligned}$$

where $k(r) = \nabla_s f_S(s)/f_S(s)$ as above.

Collecting terms allows us to write

$$\Psi(r, h(r)) = a_0(r)' h(r) + \nabla_s h(r)' b_0(r) + c_0(r)' h(r),$$

with

$$a_0(r) = d_\kappa(s)\left\{-\frac{1}{p_{H,\kappa}}\frac{s-p_{H,\kappa}}{f_S(s)}k(r), 0, \frac{1}{p_{H,\kappa}}\frac{s-p_{H,\kappa}}{f_S(s)}(k(r)sm_H(s) - s\nabla_s m_H(s) - m_H(s)), 0, 0\right\}$$
$$b_0(r) = d_\kappa(s)\frac{1}{p_{H,\kappa}}\frac{s-p_H}{f_S(s)}\{1, 0, -sm_H(s), 0, 0\}$$
$$c_0(r) = -\frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]}\frac{s}{p_{H,\kappa}}[m_H(s) + s\nabla_s m_H(s)]\left\{0, 0, 0, \frac{1}{p_{H,\kappa}}, -1\right\}.$$

Taking the expectation of the first component of $\Psi(R, h(R;\eta))$ yields

$$\begin{aligned}
&\mathbb{E}\left[a_0(R)' h(R;\eta)\right] \\
&= \int\sum_{t=0,1} d_\kappa(s)\left\{-\frac{1}{p_{H,\kappa}}\frac{s-p_{H,\kappa}}{f_S(s)}k(r), \frac{1}{p_{H,\kappa}}\frac{s-p_{H,\kappa}}{f_S(s)}(k(r)sm_H(s) - s\nabla_s m_H(s) - m_H(s))\right\} \\
&\times \begin{pmatrix} f_S(s;\eta)\, s\mathbb{E}_\eta[Y|\, T=1, S=s] \\ f_S(s;\eta)\end{pmatrix} s^t(1-s)^{1-t} f_S(s)\,\mathrm{d}s \\
&= \int d_\kappa(s)\left\{-\frac{s-p_{H,\kappa}}{p_{H,\kappa}}k(r), \frac{s-p_{H,\kappa}}{p_{H,\kappa}}(k(r)sm_H(s) - s\nabla_s m_H(s) - m_H(s))\right\} \\
&\times \begin{pmatrix} s\mathbb{E}_\eta[Y|\, T=1, S=s] \\ 1\end{pmatrix} f_S(s;\eta)\,\mathrm{d}s \\
&= \mathbb{E}_\eta[v_1(R)\, d_\kappa(S)\{1, T, TY, (1-T)Y\}],
\end{aligned}$$

where

$$v_1(r) = \left\{\frac{s-p_{H,\kappa}}{p_{H,\kappa}}[-m_H(s) + sk(r)m_H(s) - s\nabla_s m_H(s)], 0, -\frac{s-p_{H,\kappa}}{p_{H,\kappa}}k(r), 0\right\}.$$

To take the expectation of the second component of $\Psi(r, h(r))$ we use integration by parts:

$$\begin{aligned}
\mathbb{E}\left[\nabla_s h(R;\eta)' b_0(R)\right] &= \int f_0(r)\, b_0(r)' [\nabla_s h(r;\eta)]\,\mathrm{d}r \\
&= \int_{s=0}^{s=1} f_0(s)\, b_0(r)' [\nabla_s h(r;\eta)]\,\mathrm{d}s \\
&= \left[f_0(s)\, b_0(r)' h(r;\eta)\right]_0^1 - \int \nabla_s\left[f_0(s)\, b_0(r)'\right] h(r;\eta)\,\mathrm{d}r \\
&= 0 - \int \nabla_s\left[f_0(s)\, b_0(r)'\right] h(r;\eta)\,\mathrm{d}r \\
&= \mathbb{E}_\eta\left[v_2(R)\, d_\kappa(S)\{1, T, TY, (1-T)Y\}'\right],
\end{aligned}$$

with

$$v_2(r) = \frac{1}{p_{H,\kappa}}(s-p_{H,\kappa})\{m_H(s) + s\nabla_s m_H(s), 0, 0, 0\}' + \frac{d_\kappa(s)}{p_{H,\kappa}}\{sm_H(s), 0, -1, 0\}'.$$

This follows from the fact that

$$\begin{aligned}
\nabla_s \left[f_0(s)\, b_0(r)'\right] &= \nabla_s \left[f_S(s)\, d_\kappa(s) \frac{1}{p_{H,\kappa}} \frac{s-p_{H,\kappa}}{f_S(s)} \{1,0,-sm_H(s),0,0\}\right] \\
&= \nabla_s \left[d_\kappa(s) \frac{s-p_{H,\kappa}}{p_{H,\kappa}} \{1,0,-sm_H(s),0,0\}\right] \\
&= \frac{d_\kappa(s)}{p_{H,\kappa}} (s-p_H) \{0,0,-m_H(s) - s\nabla_s m_H(s),0,0\} \\
&+ \frac{d_\kappa(s)}{p_{H,\kappa}} \{1,0,-sm_H(s),0,0\}.
\end{aligned}$$

Finally the expectation of the third term is given by

$$\begin{aligned}
\mathbb{E}\left[c_0(R)'\, h(R;\eta)\right] &= -\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]} \frac{s}{p_{H,\kappa}} \left[m_H(s) + s\nabla_s m_H(s)\right] \left\{\frac{1}{p_{H,\kappa}}, -1\right\} \begin{pmatrix} \mathbb{E}_\eta[d_\kappa(S)\,T] \\ \mathbb{E}_\eta[d_\kappa(S)] \end{pmatrix} f_0(r)\,\mathrm{d}r \\
&= -\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]} \frac{s}{p_{H,\kappa}} \left[m_H(s) + s\nabla_s m_H(s)\right] \left(\frac{\mathbb{E}_\eta[d_\kappa(S)\,T]}{p_{H,\kappa}} - \mathbb{E}_\eta[d_\kappa(S)]\right) f_0(r)\,\mathrm{d}r \\
&= -\left[\int \frac{d_\kappa(s)}{\mathbb{E}[d_\kappa(S)]} \frac{s}{p_{H,\kappa}^2} \left[m_H(s) + s\nabla_s m_H(s)\right] f_0(r)\,\mathrm{d}r\right] \mathbb{E}_\eta\left[d_\kappa(S)(T-p_{H,\kappa})\right] \\
&= -\frac{1}{p_{H,\kappa}} \mathbb{E}\left[\frac{S}{p_{H,\kappa}}\left[m_H(S) + S\nabla_s m_H(S)\right] \middle| d_\kappa(S)=1\right] \mathbb{E}_\eta\left[d_\kappa(S)(T-p_{H,\kappa})\right] \\
&= \mathbb{E}_\eta\left[v_3(R)\, d_\kappa(S) \{1,T,TY,(1-T)Y\}'\right],
\end{aligned}$$

with

$$v_3(r) = -\frac{1}{p_{H,\kappa}} \mathbb{E}\left[\frac{S}{p_{H,\kappa}}\left[m_H(S) + S\nabla_s m_H(S)\right] \middle| d_\kappa(S)=1\right] \{-p_{H,\kappa},1,0,0,0\}.$$

The correction term portion of the efficient influence function will take the form $\delta^{\text{lsie}}(z) = \delta_H^{\text{lsie}}(z) - \delta_L^{\text{lsie}}(z)$. The forms for $v_1(r)$, $v_2(r)$ and $v_3(r)$ given above suggest that

$$\begin{aligned}
\delta_H^{\text{lsie}}(z) &= -\frac{d_\kappa(s)}{p_{H,\kappa}} \frac{\nabla_s f_S(s)}{f_S(s)} (ty - sm_H(s))(s-p_{H,\kappa}) \\
&- \frac{d_\kappa(s)}{p_{H,\kappa}} (ty - sm_H(s)) \\
&- \frac{d_\kappa(s)}{p_{H,\kappa}} \mathbb{E}\left[\frac{S}{p_{H,\kappa}}\left[m_H(S) + S\nabla_s m_H(S)\right] \middle| d_\kappa(S)=1\right] (t-p_{H,\kappa}).
\end{aligned}$$

The second part of the correction term, $\delta_L^{\text{lsie}}(z)$, can be derived similarly to the first. This derivation, which is omitted, yields

$$
\begin{aligned}
\delta_L^{\text{lsie}}(z) = & -\frac{d_\kappa(s)}{1-p_{H,\kappa}} \frac{\nabla_s f_S(s)}{f_S(s)} \left((1-t)\,y-(1-s)\,m_L(s)\right)(s-p_{H,\kappa}) \\
& - \frac{d_\kappa(s)}{1-p_{H,\kappa}} \left((1-t)\,y-(1-s)\,m_L(s)\right) \\
& - .\frac{d_\kappa(s)}{1-p_{H,\kappa}} \mathbb{E}\left[\frac{1-S}{1-p_{H,\kappa}} \left[-m_L(S)+(1-S)\nabla_s m_L(S)\right] \middle| d_\kappa(S)=1 \right](t-p_{H,\kappa}),
\end{aligned}
$$

and hence $\delta^{\text{lsie}}(z) = \delta_H^{\text{lsie}}(z) - \delta_L^{\text{lsie}}(z)$ equal to

$$
\begin{aligned}
\delta^{\text{lsie}}(z) &= \delta_H^{\text{lsie}}(z) - \delta_L^{\text{lsie}}(z) \\
&= -\frac{d_\kappa(s)}{p_{H,\kappa}} \frac{\nabla_s f_S(s)}{f_S(s)} \left(ty - s m_H(s)\right)(s-p_{H,\kappa}) \\
&+ \frac{d_\kappa(s)}{1-p_{H,\kappa}} \frac{\nabla_s f_S(s)}{f_S(s)} \left((1-t)\,y-(1-s)\,m_L(s)\right)(s-p_{H,\kappa}) \\
&- \frac{d_\kappa(s)}{p_{H,\kappa}} \left(ty - s m_H(s)\right) + \frac{d_\kappa(s)}{1-p_{H,\kappa}} \left((1-t)\,y-(1-s)\,m_L(s)\right) \\
&- \frac{d_\kappa(s)}{p_{H,\kappa}} \mathbb{E}\left[\frac{S}{p_{H,\kappa}} \left[m_H(S)+S\nabla_s m_H(S)\right] \middle| d_\kappa(S)=1 \right](t-p_{H,\kappa}) \\
&+ \frac{d_\kappa(s)}{1-p_{H,\kappa}} \mathbb{E}\left[\frac{1-S}{1-p_{H,\kappa}} \left[-m_L(S)+(1-S)\nabla_s m_L(S)\right] \middle| d_\kappa(S)=1 \right](t-p_{H,\kappa})
\end{aligned}
$$

as claimed.