Measuring the average outcome and inequality effects of segregation in the presence of social spillovers[1]

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

INITIAL DRAFT: September 30, 2005
THIS DRAFT: September 22, 2008

Preliminary and Incomplete

ABSTRACT: In this paper we analyze the causal effects of reallocating individuals across social groups of varying sizes in the presence of social interactions or spillovers. We consider the case where individuals are either 'high' or 'low' types. Own outcomes may depend on the fraction of high types in one's social group. We characterize the average outcome effect and inter-type inequality effects of 'local' increases in segregation. We also characterize average outcome-maximizing and -minimizing allocations of individuals to groups. We relate our estimands to the theory of sorting in the presence of social interactions. For each estimand we provide conditions for identification. We also propose nonparametric estimators and characterize their large sample properties.

JEL CLASSIFICATION: C14, C31, D62, I21

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

[1]We thank David Ahn, Susan Athey, Alan Auerbach, 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 the Berkeley Labor Lunch, the January 2006 American Economic Association Meetings, the February 2006 meeting of MacArthur Network on Economic Inequality and Social Interactions, the June 2007 Causal Inference Conference hosted by the Department of Economics at Uppsala University, Sweden and the May 2008 conference in memory of Antoni Calvó-Armengol at UAT - Barcelona for feedback and suggestions. Comments from seminar participants at the University of California - Berkeley, University of California - Davis and the Institute for Fiscal Studies are also gratefully acknowledged. All the usual disclaimers apply.

[†]Department of Economics, University of California - Berkeley, 508-1 Evans Hall #3880, Berkeley, CA 94720 and the National Bureau of Economic Research. E-MAIL: bgraham@econ.berkeley.edu. WEB: http://www.econ.berkeley.edu/~bgraham/

[‡]Department of Economics, Harvard University, Littauer M-24, Cambridge, MA 02138 and the National Bureau of Economic Research. E-MAIL: imbens@fas.harvard.edu.WEB: http://www.economics.harvard.edu/faculty/imbens/imbens.html

[♮]Department of Economics, University of Southern California, Kaprielian Hall, University Park Campus, Los Angeles, CA 90089. E-MAIL: ridder@usc.edu.WEB: http://www-rcf.usc.edu/~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 et al. (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 effects 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 four 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 (1996c) 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 (AME) of school racial composition on student achievement, does not correspond to an implementable policy. It would be impossible, for example, to engineer

[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., Manski 1992, Hoxby 2003, Ladd 2003, 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 recent discussions see Oakes (1992), Epple, Newlon and Romano (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, Liebman and Katz (2007).

[3] The original Coleman Report provides a particularly thoughtful example of this type of informal inference 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 *et al* 1966, p. 22).

an increase in minority enrollment across all schools – the policy effect measured by the AME – 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 is clearly an essential ingredient to any evaluation of a particular race-based allocation of students to schools, it is not sufficient.[4]

In this paper we provide a comprehensive discussion 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. We also provide conditions for nonparametric identification, propose estimators and characterize their large sample properties. Paralleling treatments in the theoretical public finance literature we consider the simple setting where individuals are either 'high' or 'low' types (e.g., de Bartolome 1990, Benabou 1993, Becker and Murphy 2000).[5] Individual outcomes may depend on the type composition of their social group in a fully nonparametric way. We refer to such dependence as a social spillover or peer group effect. Within this setting we develop three classes of estimands. The first class measure the average strength any social spillovers. 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). We view our second set of estimands as more innovative. They measure the effects of small increases in segregation (relative to the status quo) on average outcomes as well as the average outcome gap between high and low type individuals. They provide a basis for characterizing any equity versus efficiency trade-offs associated with segregation. Our final class of estimands allow 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 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 such work. For example, our *local segregation outcome effect* (LSOE) estimand has a representation as an 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

[4] More generally the menu of program evaluation estimands surveyed by Imbens (2004, 2006), Heckman and Vytlacil (2007), 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).

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 vouches 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 recent exception, our approach to identification and estimation is fully nonparametric.[6] We provide nonparametric estimators for each of our proposed estimands and characterize their large sample properties.

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) and Weinberg (2006) provide nice examples of how concrete microstructures 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.[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, Moffit 2001, Duncan and Raudenbush 2001, Sampson, Morenoff and Gannon-Rowley 2002, Glaeser and Scheinkman 2003, Graham 2008a,b). 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. Related work in this vein includes that of Graham, Imbens and Ridder (2006, 2007) and Bhattacharya (2008).

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]

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

[7] An important caveat to this claim is that explicit microstructures of strategic interaction can generate a mapping from group composition into outcomes that exhibits discontinuities (cf., Brock and Durlauf 2001, 2007). Since we estimate this mapping using kernel smoothing methods, our approach may work poorly in such situations.

[8] The closet 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

The statistical aspects of this paper are most closely connected to the literature of semiparametric M-estimation as in Newey (1994a,b) and Newey and McFadden (1994). Our estimators while straightforward to compute 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 various 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 1923, Rubin 1974, Holland 1986a,b).

Section 3 presents our first class of estimands (and associated estimators); those which summarize the structure of social spillovers. We relate these estimands to other proposed measures of the strength of social spillovers such as the social multiplier (Manski 1993, Brock and Durlauf 2001, Glaeser and Scheinkman 2003).

Section 4 presents our local reallocation estimands and estimators. Section 5 discusses the planner's problem. By characterizing the solution to this problem we are able to show that the inefficiency of the status quo – the difference between observed average outcomes and that which would occur under an outcome-maximizing allocation – is identified under certain assumptions. The estimands, our identification results for them, and the proposed estimators of Sections 4 and 5 are the key contributions of this paper. Section 6 summarizes and suggests areas for future research. All proofs and derivations are collected in a series of appendices.

2 Setup and assumptions

In this section we present our statistical model and discuss the core 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 random variable.

2.1 Population framework

There is a population of individuals (e.g., elementary school students residing in a certain school district). Individuals are indexed by $i \in \mathcal{I} = \{1, \ldots, I_p\}$ and are one of two observed types $T_i \in \{0, 1\}$, for example, boy or girl. Additional, unobserved, individual-level heterogeneity is indexed by the scalar $A_i \in \mathcal{A}$. 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 $T = 1$ as 'high' types and

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 (1996a,b), Epple and Romano (1998) and Becker and Murphy (2000).

those individuals with $T = 0$ as 'low' types. The population fraction of high types is given by $p_H = \Pr(T = 1)$. We assume that T_i is non-manipulable, denoting a permanent characteristic such as race or sex. The outcome of interest, say, student achievement, is $Y_i \in \mathcal{Y}$ and may be discretely- or continuously-valued.

Individuals reside in different locations or, alternatively, 'attend' different 'schools'. Members of the population of available locations are indexed by $c \in \mathcal{C} = \{1, \ldots, C_p\}$. Associated with each location is a vector of unobserved characteristics $U_c \in \mathcal{U}$.[9] If locations are, for example, schools, then U_c might capture unobserved heterogeneity in teacher quality and facilities.

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 $\mathbf{G} = \left(G_1, \ldots, G_{I_p}\right)'$.

Individuals assigned to a common location are neighbors. Let $N_c = \sum_{i=1}^{I_p} \mathbf{1}(G_i = c) \in \mathcal{N} = \{n_1, \ldots, n_L\}$ equal the number of individuals assigned to group c (L equals the number of support points in the group size distribution).

Individual $i's$ peer group includes those individuals also assigned to her location

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

These peers' types and abilities are given by the vectors

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

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, ability and number of her peers as well as the vector of unobserved location characteristics U_i :

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

Depending on the context it will be useful to take expectations with respect to the population of locations or individuals. When a random variable is indexed by c, then the presumed population is that of locations, while when it is indexed by i, it is that of individuals. Hence $\mathbb{E}[N_c]$ equals average group size, while $\mathbb{E}[N_i]$ equals the expected group size for a randomly selected individual. If most individuals reside in a few large locations, then the two expectations may be very different.[10]

We can connect the two sets of expectations as follows. Observe that the probability that a randomly sampled individual resides in location c is given by $\Pr(G_i = c) = N_c/I_p$. By iterated

[9] Below we consider the case where some location attributes are observed.

[10] The average New Yorker lives in a very large city, by virtue of the fact that a large fraction of New York's population lives in New York City. However, the average size of a New York municipality, is far smaller.

expectations we then have

$$\mathbb{E}\left[N_i\right] = \sum\nolimits_{c=1}^{C_p} \Pr\left(G_i = c\right) N_c = \sum\nolimits_{c=1}^{C_p} \frac{N_c}{I_p} N_c = \frac{1}{C_p} \sum\nolimits_{c=1}^{C_p} \frac{N_c}{\mu_N} N_c = \mathbb{E}\left[\frac{N_c}{\mu_N} N_c\right],$$

where $\mathbb{E}\left[N_c\right] = I_p/C_p = \mu_N$ is mean location size. The N_c/μ_N weight appears frequently in the formulae which following. Its role is to convert an average over locations into one over individuals.

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. The function $Y_i(\mathbf{g})$ gives the potential outcome for individual i associated with allocation $\mathbf{g} \in \mathcal{G}$.[11,12]

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 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

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

[11] 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 court-ordered mandatory school busing plan induces the same allocation of students across schools as a voluntary one, then the associated outcome distributions will also be identical. This may be 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 (cf., Holland 1986a).

[12] 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.

Our next assumption restricts the structure of peer influences within a neighborhood. Let $N_i^H = \sum_{j=1}^{N_i-1} T_{p(i),j}$ and $N_i^L = \sum_{j=1}^{N_i-1} \left(1 - T_{p(i),j}\right)$ denote the total number of high and low type peers for individual 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 $N_i - 1$ 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* $\underline{A}_{p(i)}^{L}$, *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*

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

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.[13]

Assumption 2.2, and the Fundamental Theorem of Symmetric Functions (cf., Weyl 1946, Altonji and Matzkin 2005), implies a further restriction on the allocation response function of

$$Y_i(\underline{T}_{p(i)}, \underline{A}_{p(i)}, N_i, U_i) = Y_i(S_{-i}, \tau_H(\underline{A}_{p(i)}^{H}), \tau_L(\underline{A}_{p(i)}^{L}), N_i, U_i)$$

where $S_{-i} = (N_i - 1)^{-1} \sum_{j \in p(i)} T_j$ is the fraction of $i's$ peers who are high types and $\tau_H(\underline{A}_{p(i)}^{H})$ and $\tau_L(\underline{A}_{p(i)}^{L})$ are, respectively, the vectors of the first N_i^H and N_i^L elementary symmetric polynomials on $A_{p(i)}^{H}$ and $A_{p(i)}^{L}$ (see Weyl 1946, p. 30). We emphasize that Assumption 2.2 does allow individual's to respond differently to unobserved variation in the ability composition of their high versus low type 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

[13] Calvó-Armengol, Patacchini and Zenou (2006) provide a nice example of how richer network data can be used to study peer influences.

classmates. Not whether he learns more when surrounded by female classmates once we condition on their 'disruptiveness'. If 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]

Assumption 2.3 can always be imposed by a normalization. Assume that unnormalized ability is A_i^*, normalized ability is given by $A_i = F\left(A_i^* \middle| T_i\right)$. That is our definition of an individual's 'ability' is their rank amongst those of their own type.[15]

The allocation response function $Y_i(S_{-i}, \tau_H(\underline{A}^H_{p(i)}), \tau_L(\underline{A}^L_{p(i)}), N_i, U_i)$ defines an individual-specific mapping from peer types, ability, neighborhood size and other neighborhood characteristics into outcomes. In our framework the 'treatment' induced by a given allocation is a specific configuration of peers, as summarized by their type composition, S_{-i}, unobserved ability, $\tau_H(\underline{A}^H_{p(i)})$ and $\tau_L(\underline{A}^L_{p(i)})$, and number, N_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 1923, Rubin 1974, Holland 1986a,b). The *observed* 'treatment' is an assignment to a set of peers with a given type composition and size. However, because peers and locations are heterogeneous, observationally equivalent assignments may be associated with distinct potential outcomes. Assumptions 2.1 and 2.2 are not strong enough to satisfy the homogenous treatment restriction implicit in Rubin's Stable-Unit-Treatment-Value-Assumption (SUTVA) (cf., Holland 1986a,b, Rubin 1990).[16]

To deal with this issue we define an intermediate object: the expected allocation response function. Individual's $i's$ expected allocation response function is given by

$$Y_i^e\left(s_-, n\right) = \int \dots \int Y_i(s_-, \tau_H(\underline{a}^H_{p(i)}), \tau_H(\underline{a}^L_{p(i)}), n, u) \times \left\{ \prod_{j \in p(i)} f_A(\underline{a}_{p(i),j}) \mathrm{d}a_{p(i),j} \right\} f_U(u) \mathrm{d}u. \tag{2}$$

Equation (2) gives an individuals expected outcome when assigned to a group with peer composition

[14] If T_i indexes a manipulable 'treatment' then this assumption, of course, has more content. Our framework can be adapted to this case. This leads to a method of program evaluation that allows for treatment spillovers. This extension is a topic of ongoing research.

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

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

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

$S_{-i} = s_{-}$ and size $N_i = n$ *when groups are formed in a certain way.* In particular high type peers are random draws from the subpopulation of high types, low type peers are random draws from the subpopulation of low types, and groups, so formed, are randomly assigned to specific locations.

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

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

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

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

where s is the overall fraction of high types in a group (inclusive of oneself). That is, we let $S_c = \sum_{i \in \{G_i = c\}} T_i / N_c$ 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 $m_H(s, n)$ and $m_L(s, n)$ feature in each of our estimands. Most of our identification results follow from identification of $m_H(s, n)$ and $m_L(s, n)$. These two functions respectively trace out the average outcome response associated with exogenous changes in group composition, s, and size, n, for high and low type individuals. The overall mean allocation response function is given by the composition weighted average

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

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 $(\underline{A}_c', \underline{Q}_c)'$. 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) (cf., Graham 2008a,b). This is a simple 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 1996a,b, 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 $m(s, n)$. A typical set of assumptions is that $m_H(s, n) - m_L(s, n) > 0$ for all $s \in \mathcal{S}$ and that $\partial^2 m(s, n) / \partial s^2$ 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 $m(s, n)$ completely unrestricted.

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

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

where

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

The derivative of $m(s,n)$ with respect to group composition consists of two parts. The first part, $p(s,n)$, 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. The second component, $e(s,n)$, measures the spillover or external effect associated with increasing s.

Replacing a low type individual with a high type individual raises the group average outcome for two reasons. First, irrespective of the presence of social spillovers, average outcomes will 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. Second, 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 core assumptions ensure that $m_H(s,n)$, $m_L(s,n)$ and their derivatives with respect to group composition, $\nabla_s m_H(s,n)$ and $\nabla_s m_L(s,n)$, are asymptotically identified. First we restrict the status quo assignment mechanism.

Assumption 2.4 (Double Randomization)

$$F_{\underline{A}|\underline{T},N,U}(\underline{A}_c|\underline{T}_c, N_c, U_c) = \prod_{i\in\{i:G_i=c\}} F_A(A_i)$$
$$F_{U|\underline{T},N}(U_c|\underline{T}_c, N_c) = F_U(U_c).$$

It is helpful to conceptualize Assumptions 2.4 as an implication of a specific assignment mechanism. The social planner first chooses a feasible distribution of group compositions and sizes

$$F^{\text{sq}}_{S,N}(s,n),$$

where the 'sq' superscript denotes 'status quo'.

Feasibility of the status quo allocation requires that it respects the type distribution of the population:

$$p_H = \frac{1}{I_p} \sum\nolimits_{c=1}^{C_p} N_c S_c^r = \frac{1}{C} \sum\nolimits_{c=1}^{C_p} \left\{ \frac{N_c}{\mu_N} \right\} S_c \tag{5}$$
$$= \mathbb{E}\left[\frac{N_c}{\mu_N} \mathbb{E}\left[S_c | N_c \right] \right] = \sum_{l=1}^{L} \left[\frac{n_l}{\mu_N} \int_0^1 s f_{S|N}^{\text{sq}}\left(s | n_l \right) \mathrm{d}s \right] \tau_l^{\text{sq}}$$

with $p_H = \Pr\left(T_i = 1\right), \mu_N = I_p/C_p = \mathbb{E}\left[N_c\right], \tau_l^{\text{sq}} = \Pr\left(N_c = n_l\right)$ and L, as noted above, being the number of support points in the location size distribution.

After choosing a feasible joint distribution for group composition and size the planner randomly assigns each chosen composition and size pair to a specific location. Random assignment at this stage ensures that the second part of Assumption 2.4 is satisfied. Finally the planner fills the high and low type spaces in each location by randomly sampling from the high and low type subpopulations. This ensures, along with Assumption 2.3, satisfaction of first part of Assumption 2.4.

Informally Assumption 2.4 rules out 'matching' and 'sorting' (cf., Graham 2008a). It does not, however, restrict the degree of status quo segregation or integration ($F_{S,N}^{\text{sq}}\left(s, n\right)$ is unrestricted). 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 requires 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 strong restrictions. Below we briefly discuss how to relax them in various ways. Of course, Assumption 2.4, will be most plausible when assignment of individuals to groups, and groups to locations, is actually randomized.

Assumption 2.4 is stronger than required for most of our results. Nevertheless we prefer it due to is simplicity and interpretability. Its role is to ensure identification of the average mapping from group composition and size to outcomes (i.e., $m_H\left(s, n\right)$ and $m_L\left(s, n\right)$). While there are other ways to achieve the same result, Assumption 2.4 is arguably the simplest and allows us to clearly highlight the main contributions of our paper. Double randomization evidently rules-out settings where the status quo allocation corresponds to a laissez faire equilibrium (where agents choose locations to maximize their outcomes). Other authors have discussed identification of (parametric) production functions exhibiting spillovers when agents choose neighborhoods endogenously (e.g., Neshiem 2002, Durlauf 2004, Graham 2008b).

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

Assumption 2.5 (Continuous Variation) *If $f_{S,N}^{\text{sq}}\left(s, n_l\right) > 0$ then $f_{S,N}^{\text{sq}}\left(s', n_l\right) > 0$ for all s' in a neighborhood of $s \in \mathcal{S}$.*

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 smallest possible group size (i.e., $\min\left(\mathcal{N}\right) = n_1$) 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) $\{\underline{Y}_c, \underline{T}_c, N_c\}_{c=1}^{C}$ *is a random sample of* $C < C_p$ *neighborhoods of* $I = \sum_{c=1}^{C} N_c < I_p$ *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 $m_H(s,n)$, $m_L(s,n)$ and their derivatives with respect to s are asymptotically identified.

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

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

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.1.

3 Characterizing social spillovers

Prior work on the empirics of social interactions has emphasized testing for their presence and/or measuring their average strength. Examples include Manski (1993), Brock and Durlauf (2001), Glaeser and Scheinkman (2003) and Graham (2008a). This work emphasizes the notion of a social multiplier or the ratio of the full effect of marginal changes in group composition to the private effect:

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

For simplicity, we instead focus on a direct measure of average spillover strength. Conditional on $S_i = s$ and $N_i = n$ the average external effect is given by $e(s,n)$. Averaging over individuals gives an overall *average spillover effect* (ASE) of

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

Equation (6) equals the mean external effect of an unit increase in the fraction of high type individuals in each group.

Below we show that the outcome effects of reallocations can be nontrivial even if $\beta^{\mathrm{ase}} = 0$. Nevertheless β^{ase} is a simple summary measure of spillover strength; being a nonparametric generalization of the target estimand of a large empirical literature (e.g., Coleman *et al* 1966, Mayer and Jencks 1989, Solon 1999, Angrist and Lang 2004, Ciccone and Peri 2006, Graham 2008a). While

β^{ase} is arguably of scientific interest it does not, since the peer structure of all individuals cannot be simultaneously improved, correspond to an implementable policy.

Identification of β^{ase} follows directly from Proposition 2.1 and random sampling. We propose to estimate β^{ase} using kernel smoothing methods. Let $\mathcal{K}(u)$ denote a kernel function which integrates to one and satisfies other conditions. Let $b > 0$ denote a bandwidth and define $K_b(u) = b^{-J}\mathcal{K}(u/b)$ with $J = \dim(u)$. Without loss of generality let the first I_1 individuals in the sample be high-types with the remaining $I_0 = I - I_1$ individuals being low-types.

The estimation procedure consists of two steps. In the first step the derivatives of $m_H(s,n)$ and $m_L(s,n)$ with-respect-to s are estimated by

$$\begin{aligned}\nabla_s \widehat{m}_H(s,n) &= \frac{1}{\widehat{g}_{2H}(s,n)}\left[\nabla_s \widehat{g}_{1H}(s,n) - \nabla_s \widehat{g}_{2H}(s,n)\,\widehat{m}_H(s,n)\right] \\ \nabla_s \widehat{m}_L(s,n) &= \frac{1}{\widehat{g}_{2L}(s,n)}\left[\nabla_s \widehat{g}_{1L}(s,n) - \nabla_s \widehat{g}_{2L}(s,n)\,\widehat{m}_L(s,n)\right].\end{aligned} \tag{7}$$

where

$$\widehat{m}_H(s,n) = \frac{\widehat{g}_{1H}(s,n)}{\widehat{g}_{2H}(s,n)}, \quad \widehat{m}_L(s,n) = \frac{\widehat{g}_{1L}(s,n)}{\widehat{g}_{2L}(s,n)} \tag{8}$$

are the corresponding conditional expectation function estimates and

$$\begin{aligned}\widehat{g}_{1H}(s,n) &= \frac{1}{I_1}\sum_{i=1}^{I_1} K_b(S_i - s, N_i - n)\,Y_i, \quad \widehat{g}_{1L}(s,n) = \frac{1}{I_0}\sum_{i=I_1+1}^{I} K_b(S_i - s, N_i - n)\,Y_i \\ \widehat{g}_{2H}(s,n) &= \frac{1}{I_1}\sum_{i=1}^{I_1} K_b(S_i - s, N_i - n), \quad \widehat{g}_{2L}(s,n) = \frac{1}{I_0}\sum_{i=I_1+1}^{n} K_b(S_i - s, N_i - n).\end{aligned}$$

These two derivatives are estimated for each (S_i, N_i) realization observed in the sample.

In the second step β^{ase} is estimated by

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

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

Proposition 3.1 *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*

$$\phi(Z_i) = \{e(S_i, N_i) - \beta^{\text{ase}}\} - \frac{\nabla_s f_{S,N}(S_i, N_i)}{f_{S,N}(S_i, N_i)}(Y_i - m(S_i, N_i)) - \left\{ \left[\frac{T_i Y_i}{S_i} - \frac{(1 - T_i) Y_i}{1 - S_i} \right] - [m_H(S_i, N_i) - m_L(S_i, N_i)] \right\}.$$

Proof *See Appendix C for a sketch.*

Observe that the asymptotic variance formula 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.[17] The form of the influence function is also instructive. The first term would be the influence function if $e(s, n)$ we known. The second two terms therefore capture the effects of first-step nonparametric estimation of $e(s, n)$. Of these two terms the first is identical to the correction term associated with semiparametric average derivative estimation (cf., Powell, Stock and Stoker 1989, Newey and McFadden 1994). This follows from re-expressing the estimand as the difference

$$\beta^{\text{ase}} = \mathbb{E}[\nabla_s m(S_i, N_i)] - \mathbb{E}[m_H(S_i, N_i) - m_L(S_i, N_i)].$$

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

Appendix C derives the form of $\phi(Z_i)$ using the methods described by Newey (1994a). It does not provide primitive conditions for $\sqrt{N}$ 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 observe that the support of (s, n) is bounded (the support of s is, at most, equal to the unit interval). Therefore, finiteness of the semiparametric efficiency bound for β^{ase} follows if $f_{S,N}(s, n)$ is bounded away from zero on this support. If this condition did not hold we would need to introduce 'fixed trimming' into the definition of our estimand.[18] A second concern is boundary bias in our first step estimates $\nabla_s \widehat{m}_H(s, n)$ and $\nabla_s \widehat{m}_L(s, n)$. Eliminating such bias is required for the remainder term from linearization (of our second step moment) to be small. This could be done by using one-sided kernels or local polynomial methods for points near the boundary of the support.

4 Identification of local average reallocation effects

In this section we introduce estimands which summarize different features of the effect of 'local' (to the status quo) reallocations on the distribution of outcomes. We begin by defining the general

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

[18]In practice, as in other semiparametric estimation problems, some sort of trimming may be required to deal with small denominator problems irrespective of whether $f_{S,N}(s, n)$ is bounded away from zero in the population.

class of reallocations under consideration. We assume that the social planner, or allocating agency, observes each individuals type, T_i, and the number of spaces in each location, N_c (i.e.,the planner observes $F^{\mathrm{sq}}_{S,N}(s,n)$, the joint distribution of (S_c, N_c) under the status quo). The planner also knows the high- and low-type mean allocation response functions $m_H(s,n)$ and $m_L(s,n)$. The planner does not observe A_i or U_c.

We consider reallocations that leave the marginal distribution of groups-size fixed or reallocation densities that satisfy the factorization

$$f^{\mathrm{r}}_{S,N}(s,n_l) = f^{\mathrm{r}}_{S|N}(s|\,n_l)\tau^{\mathrm{sq}}_l, \quad \forall\ s \in \mathcal{S}, n_l \in \mathcal{N} \tag{9}$$

as well as the feasibility constraint

$$\sum_{l=1}^{L}\left[\frac{n_l}{\mu_N}\int_0^1 s f^{\mathrm{r}}_{S|N}(s|\,n_l)\,\mathrm{d}s\right]\tau^{\mathrm{sq}}_l = p_H. \tag{10}$$

The set of reallocations satisfying conditions (9) and (10) is very large. Our local reallocation estimands measure the effects of a particular parameterization of a small, segregation increasing (relative to the status quo), reallocation. In particular they determine the sign of a small such increase in segregation on average outcomes and inter-type inequality.

For reasons detailed below, we call the local reallocation under consideration an *exposure-decreasing reallocation*. The joint density of S_c and N_c after an exposure-decreasing reallocation is given by

$$f^{\mathrm{r}}_{S,N}(s,n_l;\lambda) = \frac{1}{1+\lambda}f^{\mathrm{sq}}_{S|N}\left(\left.\frac{s+\lambda p_H}{1+\lambda}\right|n_l\right)\tau^{\mathrm{sq}}_l. \tag{11}$$

This is equivalent to altering the composition of the c^{th} group according to the rule

$$S^{\mathrm{r}}_c = S_c + \lambda\left(S_c - p_H\right). \tag{12}$$

When $\lambda > 0$ reallocation (11) moves high type individuals from groups where the fraction of high types is below their population frequency ($S < p_H$), to groups where it is above that frequency ($S > p_H$). Such moves are accommodated by switching each high type with a corresponding low type individual. We assume that λ is small enough to ensure that $S^{\mathrm{r}}_c \in [0,1]$ for all groups.[19]

We can motivate (11) by reference to the literature on the measurement of segregation (e.g.,

[19] Feasibility of (11) follows from the fact that:

$$\begin{aligned}\sum_{l=1}^{L}\left[\frac{n_l}{\mu_N}\int_0^1 s f^{\mathrm{r}}_{S|N}(s|\,n_l)\,\mathrm{d}s\right]\tau^{\mathrm{sq}}_l &= \sum_{l=1}^{L}\left[\frac{n_l}{\mu_N}\int_0^1 \frac{s}{1+\lambda}f^{\mathrm{sq}}_{S|N}(\left.\frac{s+\lambda p_H}{1+\lambda}\right|n_l)\mathrm{d}s\right]\tau^{\mathrm{sq}}_l \\ &= \sum_{l=1}^{L}\left[\frac{n_l}{\mu_N}\int_0^1 \{(1+\lambda)v - \lambda p_H\}f^{\mathrm{sq}}_{S|N}(v|\,n_l)\mathrm{d}v\right]\tau^{\mathrm{sq}}_l \\ &= (1+\lambda)p_H - \lambda p_H = p_H.\end{aligned}$$

Massey and Denton 1988, 1993). The *exposure index* measures the proportion of low types in the social group occupied by the typical or average high type. In our notation the exposure index is

$$\mathrm{EI} = \mathbb{E}\left[\frac{N_c}{\mu_N}\frac{S_c\left(1-S_c\right)}{p_H}\right] = 1 - \mathbb{E}\left[\frac{N_c}{\mu_N}\frac{S_c^2}{p_H}\right]. \tag{13}$$

Under perfect integration the distribution of S_c is degenerate with mean p_H and (13) equals $1-p_H$. Under perfect segregation S_c is always zero or one with $\mathbb{E}\left[\frac{N_c}{\mu_N}\frac{S_c^2}{p_H}\right] = 1$ and hence the exposure index equal to zero. The index therefore ranges from 0 to $1-p_H$ with lower values implying greater levels of segregation (or isolation).

For $\lambda > 0$ (11) reinforces segregation or *reduces* high type 'exposure' to low types by the amount

$$\Delta\mathrm{EI}\left(\lambda\right) \overset{def}{\equiv} \left[\left(1+\lambda\right)^2 - 1\right]\left\{\mathrm{EI} - \left(1-p_H\right)\right\} < 0.$$

The decrease in exposure induced by the reallocation is proportional to the difference $\mathrm{EI} - (1-p_H)$ or the degree of status quo segregation relative to the case of perfect integration.

From (12) average outcomes after an exposure decreasing reallocation are given by

$$\mathbb{E}\left[\frac{N_c}{\mu_N}m\left(S_c^{\mathrm{r}}, N_c\right)\right] = \mathbb{E}\left[\frac{N_c}{\mu_N}m\left(S_c + \lambda\left(S_c - p_H\right), N_c\right)\right].$$

We are interested in the direction of the effect of implementing (11) 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):

$$\begin{aligned}\beta^{\mathrm{lsoe}} &= \mathbb{E}\left[\frac{N_c}{\mu_N}\nabla_s m\left(S_c, N_c\right)\left(S_c - p_H\right)\right] \\ &= \mathbb{C}\left(\nabla_s m\left(S_i, N_i\right), S_i\right).\end{aligned} \tag{14}$$

Equation (14) is an intuitive condition. If groups where the fraction of high type agents exceeds the population mean ($S_c > p_H$) tend also to be relatively responsive to changes in s (i.e., $\nabla_s m\left(S_c, N_c\right)$ 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.

The interpretation of β^{lsoe} requires some care. Decomposing β^{lsoe} gives

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

where

$$\alpha^{\mathrm{lppe}} = \mathbb{C}\left(p\left(S_i, N_i\right), S_i\right), \quad \alpha^{\mathrm{lepe}} = \mathbb{C}\left(e\left(S_i, N_i\right), S_i\right).$$

Local reallocations may alter population average outcomes for three distinct reasons. First, peer quality changes for those individuals who change groups as part of the reallocation. This is an internalizeable or private peer effect. Second, group size for those individuals who switch groups may also change. For reasons outlined below we call this the targeting effect (cf., Piketty and Valdenaire 2006). Finally, 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 targeting effect, if N_i and S_i co-vary under the status quo, then implementing (11) will alter the conditional distribution of N_i for high and low type individuals. For example, if under the status quo $\mathbb{C}(N_i, S_i) < 0$, then implementing an exposure-decreasing reallocation involves moving high types into groups with an above average fraction of high types *and* of below average size. To the extent that high and low types are differentially affected by changes in group size, this will alter expected outcomes even in the absence of social spillovers.

Consider once again the classroom setting. Krueger and Whitmore (2001) present evidence that black elementary school students benefit from class size reductions more than white students (in terms of increased test scores). If black students tend to be located in larger classrooms than whites then, even in the absence of peer group effects, integrating classrooms will raise average achievement by improving input targeting. Such reallocations reduce average class size for blacks while raising it for whites, but the outcome benefits of the former effect outweigh the outcome costs of the latter. If, on the other hand, class size tends to be smaller for blacks, then integrating classrooms will, in the absence of some compensating peer group effect, lower average achievement. In general, moves toward greater segregation or integration may improve or worsen the targeting of other inputs (such as teacher quality or class size).

Second, 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 (11) 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 *sum of* the private *and* targeting effect on average outcomes is captured by $\alpha^{\text{lppe}} = \mathbb{C}(p(S_i, N_i), S_i)$.

Finally 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., $\alpha^{\text{lepe}} = \mathbb{C}(e(S_i, N_i), S_i) > 0$), 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 (11).

The sign of β^{lsoe} locally determines what direction of reallocation will raise outcomes. If $\beta^{\text{lsoe}} > 0$ a small increase in segregation raises average outcomes, if $\beta^{\text{lsoe}} < 0$ a step toward integration does. This observation can be developed further in order to directly relate β^{lsoe} to the theoretical work on segregation and efficiency done by de Bartolome (1990), Benabou (1993, 1996), Becker and Murphy

(2000) and others. To facilitate our discussion it is convenient to make an additional assumption (cf., de Bartolome 1990).

Assumption 4.1 (Additive Separability) *The type-specific mean allocation response functions take the form*

$$m_H(s,n) = a_H(s) + b_N(n)$$
$$m_L(s,n) = a_L(s) + b_L(n).$$

Under the assumption of additive separability we have the following representation of β^{lsoe}.

Theorem 4.1 *Under Assumptions 2.1 to 2.6 and 4.1* $\beta^{\text{lsoe}} = \alpha^{\text{lppe}} + \alpha^{\text{lepe}}$ *with*

$$\begin{aligned}\alpha^{\text{lppe}} &= \mathbb{V}(S_i)\,\mathbb{E}\left[\omega(S_i)\left\{\nabla_s a_H(S_i) - \nabla_s a_L(S_i)\right\}\right] \\ &\quad + \mathbb{V}(S_i)\,\mathbb{E}\left[\omega(S_i)\left\{\nabla_s \mathbb{E}\left[b_H(N_i)|\,S_i\right] - \nabla_s \mathbb{E}\left[b_H(N_i)|\,S_i\right]\right\}\right] \\ \alpha^{\text{lepe}} &= \mathbb{V}(S_i)\,\mathbb{E}\left[\omega(S_i)\left\{\nabla_s a_H(S_i) - \nabla_s a_L(S_i) + S_i \nabla_{ss} a_H(S_i) + (1-S_i)\nabla_{ss} a_L(S_i)\right\}\right],\end{aligned}$$

where

$$\omega(s) = \frac{1}{f_S(s)} \frac{\mathbb{E}\left[S_i - p_H | S_i \geq s\right](1-F_S(s))}{\int_{v=0}^{v=1} \mathbb{E}\left[S_i - p_H | S_i \geq v\right](1-F_S(v))\,\mathrm{d}v}, \quad \mathbb{E}\left[\omega(S_i)\right] = 1.$$

The weight function $\omega(s)$ is maximal at $s = p_H$ and minimal at $s = 0$ and $s = 1$.

Proof See Appendix B.2.

Theorem 4.1 provides a mathematical representation of the private, targeting and external effects discussed above. The sum of the former two terms equals α^{lppe}, while the latter equals α^{lepe}. For simplicity it is helpful to abstract from targeting effects so that $\mathbb{E}\left[b_j(N_i)|\,S_i\right] = \mathbb{E}\left[b_j(N_i)\right]$ for $j = H, L$ (the introduction of targeting effects complicates the analysis without fundamentally changing it). In that case Theorem 4.1 implies that a small increase in segregation raises average outcomes if

$$2\mathbb{E}\left[\omega(S_i)\left\{\nabla_s a_H(S_i) - \nabla_s a_L(S_i)\right\}\right] + \mathbb{E}\left[\omega(S_i)\left\{S_i \nabla_{ss} a_H(S_i) + (1-S_i)\nabla_{ss} a_L(S_i)\right\}\right] > 0. \quad (15)$$

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 $\nabla_s a_H(s) - \nabla_s a_L(s) > 0$ 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.[20]

Theorem 4.1 indicates that a measure of *local average complementarity,*

$$\mathbb{E}\left[\omega\left(S_i\right)\left\{\nabla_s a_H\left(S_i\right)-\nabla_s a_L\left(S_i\right)\right\}\right],$$

– is important for determining whether small increases in segregation are outcome-raising. If, *in the neighborhood of* $s = p_H$, own and peers' type *tend to be* complementary, then the first term in (15) 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 further segregation relative to the status quo.[21]

The theory literature also discusses the importance of curvature for determining whether segregation is outcome-maximizing (Benabou 1996, Fernández 2003). Curvature, equal to $s\nabla_{ss}a_H(s) + (1-s)\nabla_{ss}a_L(s)$, determines whether there are diminishing returns to peer quality at the neighborhood level. Theoretical work emphasizes the case where curvature is such that $2\{\nabla_s a_H(s) - \nabla_s a_L(s)\} + s\nabla_{ss}a_H(s) + (1-s)\nabla_{ss}a_L(s)$ is negative for all $s \in (0,1)$ (i.e., global concavity of $m(s,n)$ 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 (cf., 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 4.1 indicates that a measure of *local average curvature,*

$$\mathbb{E}\left[\omega\left(S_i\right)\left\{S_i\nabla_{ss}a_H\left(S_i\right)+\left(1-S_i\right)\nabla_{ss}a_L\left(S_i\right)\right\}\right],$$

is important for determining whether segregation is outcome raising in the current context as well. If, again in the neighborhood of $s = p_H$, 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 (15) will be negative.

To summarize Theorem 4.1 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 $m(s,n)$ (beyond Assumption 4.1). The cost of such flexibility is that β^{lsoe} provides only local information about the relative average outcome effects of segregation versus integration.

Estimation of β^{lsoe} is straightforward. We once again propose a two step non-parametric estimator based on kernel smoothing. In the first step we estimate, p_H by $\widehat{p}_H = \sum_{i=1}^{I} T_i/I$, and $m_H(s,n)$,

[20] 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

[21] If utility is linear in outcomes, then this statement is formally correct.

$m_L(s,n), \nabla_s m_H(s,n)$ and $\nabla_s m_L(s,n)$ according to (8) and (7) above. We then estimate β^{lsoe} by

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

Proposition 4.1 *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, $\widetilde{\phi}_c = \sum_{i\in\{i:G_i=c\}}\phi(Z_i)$, *and $\phi(Z_i)$, the efficient influence function, is given by*

$$\begin{aligned}\phi(Z_i) = &\left\{\nabla_s m(S_i,N_i)(S_i - p_H) - \beta^{\text{lsoe}}\right\} \\ &- \frac{\nabla_s f_{S,N}(S_i,N_i)}{f_{S,N}(S_i,N_i)}(Y_i - m(S_i,N_i))(S_i - p_H) - [Y_i - m(S_i,N_i)] \\ &- \mathbb{E}[\nabla_s m(S_i,N_i)](T_i - p_H).\end{aligned}$$

Proof *See Appendix C for a sketch.*

In some situations it may be important to decompose β^{lsoe} into is private, α^{lppe}, and spillover components, α^{lepe}. These may be estimated by

$$\widehat{\alpha}^{\text{lppe}} = \frac{1}{I}\sum_{i=1}^{I}\left[\widehat{m}_H(S_i,N_i) - \widehat{m}_L(S_i,,N_i)\right](S_i - \widehat{p}_H)$$

$$\widehat{\alpha}^{\text{lepe}} = \frac{1}{I}\sum_{i=1}^{I}\left[S_i\nabla_s\widehat{m}_H(S_i,N_i) + (1-S_i)\nabla_s\widehat{m}_L(S_i,,N_i)\right](S_i - \widehat{p}_H).$$

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

Proposition 4.2 *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, $\widetilde{\phi}_c = \sum_{i\in\{i:G_i=c\}}\phi(Z_i)$, *and $\phi(Z_i)$, the efficient influence function, is given by*

$$\begin{aligned}\phi(Z_i) = &\left\{p(S_i,N_i)(S_i - p_H) - \alpha^{\text{lppe}}\right\} \\ &+ \left\{\left(\frac{T_i}{S_i}\right)Y_i - m_H(S_i,N_i)\right\}(S_i - p_H) - \left\{\left(\frac{1-T_i}{1-S_i}\right)Y_i - m_L(S_i,N_i)\right\}(S_i - p_H) \\ &- \mathbb{E}[p(S_i,N_i)](T_i - p_H).\end{aligned}$$

Proof *See Appendix C for a sketch.*

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

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

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

$$\begin{aligned}\phi\left(Z_i\right)=&\left\{e\left(S_i,N_i\right)\left(S_i-p_H\right)-\alpha^{\text{lepe}}\right\}\\&-\frac{\nabla_s f_{S,N}\left(S_i,N_i\right)}{f_{S,N}\left(S_i,N_i\right)}\left(Y_i-m\left(S_i,N_i\right)\right)\left(S_i-p_H\right)-\left[Y_i-m\left(S_i,N_i\right)\right]\\&-\left\{\left(\frac{T_i}{S_i}\right)Y_i-m_H\left(S_i,N_i\right)\right\}\left(S_i-p_H\right)+\left\{\left(\frac{1-T_i}{1-S_i}\right)Y_i-m_L\left(S_i,N_i\right)\right\}\left(S_i-p_H\right)\\&-\mathbb{E}\left[e\left(S_i,N_i\right)\right]\left(T_i-p_H\right).\end{aligned}$$

Proof *See Appendix C for a sketch.*

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

The LSOE provides an indication of the likely effects of small increases in segregation on average outcomes. An abiding 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 an exposure-decreasing 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.

After reallocation the high-low outcome gap is given by

$$\begin{aligned}&\mathbb{E}\left[\frac{\left(S_c+\lambda\left(S_c-p_H\right)\right)N_c}{p_H\mu_N}m_H\left(S_c+\lambda\left(S_c-p_H\right),N_c\right)\right]\\&\qquad-\mathbb{E}\left[\frac{\left(1-S_c-\lambda\left(S_c-p_H\right)\right)N_c}{\left(1-p_H\right)\mu_N}m_L\left(S_c+\lambda\left(S_c-p_H\right),N_c\right)\right].\end{aligned}$$

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,

$$\begin{aligned}\beta^{\text{lsie}}=&\,\mathbb{E}\left[\frac{N_c}{p_H\mu_N}\left\{m_H\left(S_c,N_c\right)+S_c\nabla_s m_H\left(S_c,N_c\right)\right\}\left(S_c-p_H\right)\right]\\&-\mathbb{E}\left[\frac{N_c}{\left(1-p_H\right)\mu_N}\left\{-m_L\left(S_c,N_c\right)+\left(1-S_c\right)\nabla_s m_L\left(S_c,N_c\right)\right\}\left(S_c-p_H\right)\right].\end{aligned}$$

We estimate β^{lsie} by

$$\widehat{\beta}^{\text{lsie}} = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{\widehat{p}_H}\left\{\widehat{m}_H\left(S_i,N_i\right)+S_i\nabla_s\widehat{m}_H\left(S_i,N_i\right)\right\}\left(S_i-\widehat{p}_H\right)$$
$$-\frac{1}{I}\sum_{i=1}^{I}\frac{1}{1-\widehat{p}_H}\left\{-\widehat{m}_L\left(S_i,N_i\right)+\left(1-S_i\right)\nabla_s\widehat{m}_L\left(S_i,N_i\right)\right\}\left(S_i-\widehat{p}_H\right),$$

where $\widehat{p}_H$, $\widehat{m}_H\left(s,n\right)$, $\widehat{m}_L\left(s,n\right)$, $\nabla_s\widehat{m}_H\left(s,n\right)$ and $\nabla_s\widehat{m}_L\left(s,n\right)$ are as described above.

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

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

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

$$\begin{aligned}
\phi\left(Z_i\right) = &\left\{p_H^{-1}\left[m_H\left(S_i,N_i\right)+S_i\nabla_s m_H\left(S_i,N_i\right)\right]\left(S_i-p_H\right)\right.\\
&\left.-\left(1-p_H\right)^{-1}\left[-m_L\left(S_i,N_i\right)+\left(1-S_i\right)\nabla_s m_L\left(S_i,N_i\right)\right]\left(S_i-p_H\right)-\beta^{\text{lsie}}\right\}\\
&-\frac{1}{p_H}\frac{\nabla_s f_{S,N}\left(S_i,N_i\right)}{f_{S,N}\left(S_i,N_i\right)}\left(T_iY_i-S_im_H\left(S_i,N_i\right)\right)\left(S_i-p_H\right)\\
&-\frac{1}{p_H}\left[T_iY_i-S_im_H\left(S_i,N_i\right)\right]\\
&-\mathbb{E}\left[\frac{S-p_H}{p_H^2}m_H\left(S,N\right)+\frac{S^2}{p_H^2}\nabla_s m_H\left(S,N\right)\right]\left(T_i-p_H\right)\\
&+\frac{1}{1-p_H}\frac{\nabla_s f_{S,N}\left(S_i,N_i\right)}{f_{S,N}\left(S_i,N_i\right)}\left(\left(1-T_i\right)Y_i-\left(1-S_i\right)m_L\left(S_i,N_i\right)\right)\left(S_i-p_H\right)\\
&+\frac{1}{1-p_H}\left(\left(1-T_i\right)Y_i-\left(1-S_i\right)m_L\left(S_i,N_i\right)\right)\\
&+\mathbb{E}\left[\frac{S-p_H}{\left(1-p_H\right)^2}m_L\left(S,N\right)+\left(\frac{1-S}{1-p_H}\right)^2\nabla_s m_L\left(S,N\right)\right]\left(T_i-p_H\right).
\end{aligned}$$

Proof *See Appendix C for a sketch.*

5 The social planner's problem

In this section we characterize the structure of average outcome maximizing assignments of individuals to groups. As before we consider reallocations which leave the joint distribution of group-size fixed. This class of reallocations is completely characterized by $l=1,\ldots,L$ conditional group-composition cumulative distribution functions: $F_{S|N}\left(s|n_l\right)$. 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 (cf., Bhattacharya 2008). Consider a school board pondering open enrollment. If current achievement levels are near the maximum attainable via reallocation, then costly reassignment policies are unattractive.

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.

We continue to maintain the assumptions that the planner knows the mean allocation response function, $m(s,n)$, as well as the status quo joint distribution, $F^{\text{sq}}_{S,N}(s,n)$ and population fraction of high types, p_H. Her problem is to choose an allocation which maximizes expected average outcomes:

$$\max_{F_{S|N}(\cdot|n_1),\ldots,F_{S|N}(\cdot|n_L)} \sum_{l=1}^{L} \left[\frac{n_l}{\mu_N} \int m(s,n_l) f_{S|N}(s|\, n_l)\, \mathrm{d}s \right] \tau_l^{\text{sq}} \tag{16}$$

subject to restriction (5). Weighting by n_l/μ_N ensures that the planner maximizes average individual outcomes (and not the average of mean group outcomes).

Our characterization of the solution to (16) 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_l denote the fraction of high types in the subpopulation of individuals assigned to groups of size n_l (as part of a candidate reallocation). Conditional on choosing such an allocation, the optimal conditional allocations $F_{S|N}(s|n_1), \ldots, F_{S|N}(s|n_L)$ are determined by our first result. Since $\sigma_l = \int s f_{S|N}(s|n_1)\, \mathrm{d}s$ we can re-write the feasibility constraint (5) as

$$\sum_{l=1}^{L} \frac{n_l}{\mu_N} \sigma_l \tau_l^{\text{sq}} = p_H,$$

and hence show that the original problem is equivalent to a finite-dimensional optimization problem where the planner chooses the vector $\sigma = (\sigma_1, \ldots, \sigma_L)'$. 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, Pardalos and Thoai (2000), defines this object.

Definition 5.1 *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 $m(s)$ (cf., 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 5.1 *Consider the problem*

$$\max_{F_S(\cdot)\in\Gamma_S} \int m(s) f_S(s)\,\mathrm{d}s, \qquad s.t. \qquad \int s f_S(s)\,\mathrm{d}s = p_H, \tag{17}$$

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 (17),

$$\int m(s) f_S^*(s)\,\mathrm{d}s = M(p_H) \tag{18}$$

and

$$F_S^*(s) = (1-\pi)\mathbf{1}(s \geq s_L) + \pi\mathbf{1}(s \geq s_U), \quad \pi = \begin{cases} \frac{p_H - s_L}{s_U - s_L} & s_L < s_U \\ 1/2 & s_L = s_U \end{cases} \tag{19}$$

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.3.

Theorem 5.1 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.[22] Figure 1 plots four different forms

[22] We thank Emmanuel Saez for providing some of these examples. His intuitive insight was key in being able to show Theorem 5.1.

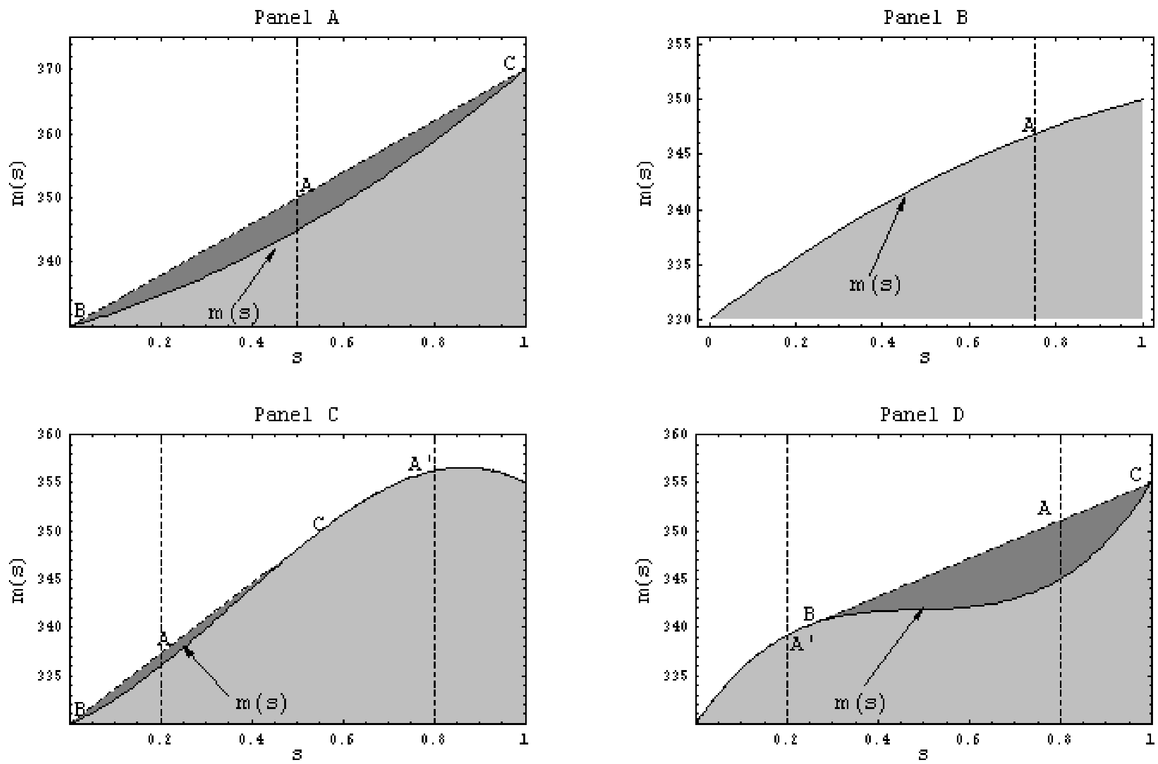

Figure 1: Optimal allocations for different $m(s)$ and p_H

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 the 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 $(p'_H, M(p'_H))$.

for $m(s)$. Consider Panel A of the figure. In that panel $m(s)$ is globally convex (on the support of s). 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 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 planners 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 5.1.

Corollary 5.1 *A solution to the social planner's problem defined by (16) and (5) is given by*

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

where

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

for $l = 1, \ldots, L$ and

$$\begin{aligned} s_L(\sigma_l) &= \max\{s : s \geq \underline{s},\ s \leq \sigma_l,\ M(s, n_l) = m(s, n_l)\} \\ s_U(\sigma_l) &= \min\{s : s \leq \overline{s},\ s \geq \sigma_l,\ M(s, n_l) = m(s, n_l)\}, \end{aligned}$$

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

$$\max_{\sigma_1 \in \mathcal{S}, \ldots, \sigma_L \in \mathcal{S}} \sum_{l=1}^{L} \frac{n_l}{\mu_N} M(\sigma_l, n_l)\, \tau_l^{\text{sq}}, \qquad s.t. \qquad \sum_{l=1}^{L} \frac{n_l}{\mu_N} \sigma_l \tau_l^{\text{sq}} = p_H. \tag{20}$$

Proof See Appendix B.4.

Corollary 5.1 provides a simple algorithm for calculating the maximum attainable average outcome available via reallocation. First, compute $M(s, n_l)$ for each of the L group sizes. Second, solve the concave program (20). Third, compute the value of $\sum_{l=1}^{L} \frac{n_l}{\mu_N} M(\sigma_l, n_l)\, \tau_l^{\text{sq}}$ at the solution.

Our final identification result is:

Proposition 5.1 *If (i) Assumptions 2.1 to 2.6 hold and (ii)* $f^{\mathrm{sq}}_{S|N}\left(s|\, n_l\right) > 0$ *for all* $s \in \mathcal{S}$ *and* $l = 1, \ldots, L$*, then (a)* $F^{*}_{S|N}\left(s|\, n_l\right)$ *is identified and (b) so is the efficiency measure*

$$\beta^{\mathrm{esq}} = \sum_{l=1}^{L} \left[\frac{n_l}{\mu_N} \int m\left(s, n_l\right) f^{*}_{S|N}\left(s|\, n_l\right) \mathrm{d}s \right] \tau^{\mathrm{sq}}_{l} - \mathbb{E}\left[Y\right].$$

Proof See Appendix B.5.

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

6 Summary

Appendices

A Some preliminary results

Lemma A.1 *For X, a continuous random variable, with (i) compact support $X \in [a, b]$, (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. *β 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, Graddy and Imbens (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

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

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

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

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

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

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

$$\int_x^b [1 - F_X(t)]\,\mathrm{d}t = [1 - F_X(t)]\,t|_x^b + \int_x^b t f_X(t)\,\mathrm{d}t \tag{21}$$
$$= -[1 - F_X(x)]\,x + \int_u^b t f_X(t)\,\mathrm{d}t.$$

We then write

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

Using (21) to substitute for $\frac{\partial}{\partial x}\int_x^b x\ f_X(t)\,\mathrm{d}t$ gives

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

This gives $\frac{\partial}{\partial x}\{\mathbb{E}[X - \mu_X | X \geq x](1 - F_X(x))\} = 0$ 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

B.1 Proof of Proposition 2.1

Under Assumptions 2.1 and 2.2 we have $Y_i = Y_i(S_{-i}, \tau_H(A^H_{p(i)}), \tau_L(A^L_{p(i)}), N_i, U_i)$. Writing

$$Y_i(S_{-i}, \tau_H(A^H_{p(i)}), \tau_L(A^L_{p(i)}), N_i, U_i) = g(T_i, A_i, S_{-i}, \tau_H(A^H_{p(i)}), \tau_L(A^L_{p(i)}), N_i, U_i)$$

we therefore have

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

where the second equality follows from Assumptions 2.3 and 2.4. Let the integral in outermost set of $\{\cdot\}$ equal $g^e(1, a, s_-, n)$. Observe that $g^e(T_i, A_i, s_-, n) = Y_i^e(s_-, n)$, therefore by Assumption 2.3 we have

$$\int g^e(1, a, s_-, n)\,f_A(a)\,\mathrm{d}a = \mathbb{E}[Y_i^e(s_-, n) | T = 1] = m_H(s, n),$$

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

B.2 Proof of Theorem 4.1

Under Assumption 4.1 we have

$$\begin{aligned}\beta^{\text{lsoe}} &= \mathbb{C}\left(a_H\left(S_i\right)-a_L\left(S_i\right), S_i\right)+\mathbb{C}\left(b_H\left(N_i\right)-b_H\left(N_i\right), S_i\right) \\ &+\mathbb{C}\left(S_i \nabla_s a_H\left(S_i\right)+\left(1-S_i\right) \nabla_s a_L\left(S_i\right)\right).\end{aligned}$$

Applying iterated expectations to the middle term gives

$$\mathbb{C}\left(b_H\left(N_i\right)-b_H\left(N_i\right), S_i\right)=\mathbb{E}\left[\mathbb{E}\left[b_H\left(N_i\right)-b_H\left(N_i\right) \mid S_i\right]\left(S_i-p_H\right)\right].$$

The result then follows direct from Lemma A.1 above.

B.3 Proof of Theorem 5.1

Consider the problem

$$\max_{F_S(\cdot) \in \Gamma_S} \int M(s) f_S(s) \mathrm{d} s, \qquad \text{s.t.} \qquad \int s f_S(s) \mathrm{d} s=p_H, \tag{22}$$

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

$$\max_{F_S(\cdot) \in \Gamma_S} \int M(s) f_S(s) \mathrm{d} s \leq M(p_H). \tag{23}$$

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

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 $M\left(p_H\right)=\int m(s) f_S^*(s) \mathrm{d} s$ 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 $s \in\left[s_L, s_U\right]$, i.e.,

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

with

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

This gives

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

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.

B.4 Proof of Corollary 5.1

Conditional on setting the fraction of high types assigned to groups of size n_l equal to σ_l we know, by Theorem 5.1, that $F^*_{S|N}(s|\, n_l)$ is an outcome-maximizing allocation. Since, again conditional on σ_l, $\int m(s, n_l) f^*_{S|N}(s|\, n_l)\,\mathrm{d}s = M(\sigma_l)$, we may therefore choose $\sigma_1, \ldots, \sigma_L$ by solving (20) which is concave by inspection.

B.5 Proof of Proposition 5.1

C Influence function derivations

This appendix details the derivation of the influence functions associated with the estimators described in Sections 3 and 4. 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 $Z = (Y, T, S, N)'$.[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 $\sqrt{N}$ consistency and asymptotic normality of our proposed estimators, our calculations do provide a formula for their large sample variance. In this sense our approach is similar in spirit and implementation to that of Newey and Stoker (1993) 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

$$\nabla_\eta \beta(\eta)|_{\eta=\eta_0} = \mathbb{E}\left[\phi(Z)\,\mathbb{S}_\eta(Z)'\right] \tag{24}$$

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

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

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 (1994a) 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 (cf., 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

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

[24] The form of (24) and a simple argument due to Newey (1990, pp. 106 - 107) 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 (24) 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 $\mathcal{T} = \{\mathbb{S}_\eta(Z) : \mathbb{E}[\mathbb{S}_\eta(Z)] = 0\}$. 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.

simplify notation; each our estimators can be defined as the solution to

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

where $\psi\left(Z,\beta,h\right)$ is some known function, $\widetilde{h}$ is a preliminary 'first step' nonparametric estimate of h and $\widehat{\mu}_N=I/C$.

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

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

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

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

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

$$\nabla_\eta\mathbb{E}\left[\psi\left(Z,h\left(\eta\right)\right)\right]=\mathbb{E}_\eta\left[\delta\left(Z\right)\mathbb{S}_\eta\left(Z\right)'\right],\tag{25}$$

then the influence function for any regular estimator of β, by the results of Newey (1990, 1994a) and equation (24) above, will be

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

As explained by Newey (1994a) and also Newey and McFadden (1994), the function $\delta\left(Z\right)$ may be viewed a correction term which accounts for first step estimation of h. Below we use the structure of (25) to calculate the appropriate correction term for each of our estimators. In particular we begin by linearizing $\psi\left(z,h\left(\eta\right)\right)$ around the truth h_0. With $\psi\left(z,h\right)-\psi\left(z,h_0\right)\simeq\Psi\left(z,h-h_0\right)$, $\Psi\left(z,h\right)$ linear in h, and (25) we then have

$$\nabla_\eta\mathbb{E}\left[\psi\left(Z,h\left(\eta\right)-h_0\right)\right]=\nabla_\eta\mathbb{E}\left[\Psi\left(Z,h\left(\eta\right)\right)\right]=\mathbb{E}_\eta\left[\delta\left(Z\right)\mathbb{S}_\eta\left(Z\right)'\right].\tag{26}$$

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

Once the form of $\delta\left(Z\right)$ has been calculated, the asymptotic variance formulae given in Sections 3 and 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 $\phi\left(Z\right)$, therefore accounting for it is no more complicated than accounting for group-dependence in the context of a standard ordinary least squares regression problem. In particular defining $\phi_c=\sum_{i\in\{i:G_i=c\}}\phi\left(Z_i\right)$, the appropriate asymptotic sampling distribution is

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

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

We begin with the local segregation outcome effect (LSOE) defined in Section 4:

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

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

$$\begin{aligned} h_{10}(r) &= f_{S,N}(s,n)\,\mathbb{E}\left[TY|\,S=s,N=n\right]=f_{S,N}(s,n)\,sm_H(s,n) \\ h_{20}(r) &= f_{S,N}(s,n)\,\mathbb{E}\left[(1-T)\,Y|\,S=s,N=n\right]=f_{S,N}(s,n)\,(1-s)\,m_L(s,n) \\ h_{30}(r) &= f_{S,N}(s,n) \\ h_{40}(r) &= \mathbb{E}\left[T\right]=p_H. \end{aligned} \tag{27}$$

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

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

$$\psi\left(r,\beta^{\text{lsoe}},h\right)=\nabla_s\left\{\frac{h_1(r)+h_2(r)}{h_3(r)}\right\}\times(s-h_4(r))-\beta^{\text{lsoe}}.$$

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 ratio 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

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

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

$$\Psi(r,h(r))=a_0(r)'h(r)+\nabla_s h(r)'b_0(r)+c_0(r)'h(r), \tag{28}$$

where

$$\begin{aligned} a_0(r) &= \frac{s-p_H}{f_{S,N}(s,n)}\left(-k(r),-k(r),-\nabla_s m(s,n)+m(s,n)\,k(r),0\right)' \\ b_0(r) &= \frac{s-p_H}{f_{S,N}(s,n)}\left(1,1,-m(s,n),0\right)' \\ c_0(r) &= -\nabla_s m(s,n)\,(0,0,0,1)' \end{aligned}$$

with

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

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 (28) and (26) 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 (28),

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

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

$$v_1(r) = (s-p_H)\left\{-\nabla_s m(s,n) + m(s,n)k(r), 0, -k(r), -k(r)\right\}'.$$

To evaluate the second term of (28) we use integration by parts as in Powell, Stock and Stoker (1989) (with $u(r) = f_0(r)\, b_0(r)'$ and $v(r) = h(r;\eta)$) to obtain a representation directly in terms of $h(r;\eta)$:

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

with

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

This follows from the fact that

$$\begin{aligned}
\nabla_s\left[f_0(r)\, b_0(r)'\right] &= \nabla_s\left[(s-p_H)\, s^t(1-s)^{1-t}\{1, 1, -m(s,n), 0\}'\right] \\
&= (s-p_H)\, s^t(1-s)^{1-t}\{0, 0, -\nabla_s m(s,n), 0\}' \\
&\quad + \left[\left((s-p_H)\frac{t-s}{s(1-s)} s^t(1-s)^{1-t} + s^t(1-s)^{1-t}\right)\{1, 1, -m(s,n), 0\}'\right],
\end{aligned}$$

and also that

$$\sum\nolimits_{t=0,1} (s-p_H)\frac{t-s}{s(1-s)} s^t(1-s)^{1-t} + s^t(1-s)^{1-t} = 1.$$

Finally we take the expectation of the final term in (28):

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

where

$$v_3(R) = \{0, -\mathbb{E}\left[\nabla_s m(s,n)\right], 0, 0\}'.$$

Combining terms gives

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

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

$$
\begin{aligned}
v(r) = \{ & m(s,n) + (s - p_H) m(s,n) k(r), \\
& -\mathbb{E}[\nabla_s m(s,n)], -(s - p_H) k(r) - 1, -(s - p_H) k(r) - 1\}.
\end{aligned}
$$

Differentiating with respect to η gives

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

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

$$
\begin{aligned}
\delta^{\text{lsoe}}(z) = & -\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)} (y - m(s,n)) (s - p_H) \\
& - (y - m(s,n)) - \mathbb{E}[\nabla_s m(s,n)] (t - p_H),
\end{aligned} \tag{29}
$$

as claimed.

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

The local private peer effect (LPPE) of Section 4 is given by

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

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r))'$ as defined in (27) above. Linearizing the implied moment function gives

$$
\Psi(r, h(r)) = a_0(r)' h(r), \tag{30}
$$

where

$$
a_0(r) = \frac{1}{f_{S,N}(s,n)} \left\{ \frac{s - p_H}{s}, -\frac{s - p_H}{1 - s}, -[m_H(s,n) - m_L(s,n)] (s - p_H), -[m_H(s,n) - m_L(s,n)] f_{S,N}(s,n) \right\}.
$$

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

$$
\begin{aligned}
\mathbb{E}\left[a_0(R)' h(R;\eta)\right] = & \int \frac{1}{f_{S,N}(s,n)} \left\{ \frac{s - p_H}{s}, -\frac{s - p_H}{1 - s}, -[m_H(s,n) - m_L(s,n)] (s - p_H)\right., \\
& \left. -[m_H(s,n) - m_L(s,n)] f_{S,N}(s,n) \right\} \\
& \times \begin{pmatrix} f_{S,N}(s,n;\eta) \mathbb{E}_\eta [TY | S = s, N = n] \\ f_{S,N}(s,n;\eta) \mathbb{E}_\eta [(1-T) Y | S = s, N = n] \\ f_{S,N}(s,n;\eta) \\ p_H(\eta) \end{pmatrix} f_{S,N}(s,n) \, \mathrm{d}s \mathrm{d}n \\
& \int \left\{ \frac{s - p_H}{s}, -\frac{s - p_H}{1 - s}, -[m_H(s,n) - m_L(s,n)] (s - p_H), -[m_H(s,n) - m_L(s,n)] f_{S,N}(s,n) \right\} \\
& \times \begin{pmatrix} \mathbb{E}_\eta [TY | S = s, N = n] \\ \mathbb{E}_\eta [(1-T) Y | S = s, N = n] \\ 1 \\ \mathbb{E}_\eta [T] / f_{S,N}(s,n;\eta) \end{pmatrix} f_{S,N}(s,n;\eta) \, \mathrm{d}s \mathrm{d}n \\
= & \, \mathbb{E}_\eta \left[v(R)' \{1, T, TY, (1-T) Y\} \right],
\end{aligned}
$$

where

$$
v(r) = \left\{ -[m_H(s,n) - m_L(s,n)] (s - p_H), -\mathbb{E}[m_H(S,N) - m_L(S,N)], \frac{s - p_H}{s}, -\frac{s - p_H}{1 - s} \right\}.
$$

This follows because of the equality

$$
-\int [m_H(s,n) - m_L(s,n)] f_{S,N}(s,n) \mathbb{E}_\eta [T] \, \mathrm{d}s \mathrm{d}n = -\mathbb{E}[m_H(S,N) - m_L(S,N)] \mathbb{E}_\eta [T].
$$

Using (30) and (26), these calculations suggest a correction term of the form

$$\delta^{\text{lppe}}(z) = \left\{ \left(\frac{t}{s} \right) y - m_H(s,n) \right\} (s - p_H) - \left\{ \left(\frac{1-t}{1-s} \right) y - m_L(s,n) \right\} (s - p_H) \tag{31}$$
$$- \mathbb{E}\left[m_L(s,n) - m_L(s,n) \right] (t - p_H),$$

as claimed.

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

The local external peer effect (LEPE) of Section 4 is given by

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

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r))'$ as defined in (27) above. Linearizing the implied moment function gives

$$\Psi(r, h(r)) = s\nabla_s \left\{ \frac{1}{f_{S,N}(s,n)} \left\{ \frac{1}{s}, -m_H(s,n) \right\} \begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix} \right\} (s - p_H)$$
$$+ (1-s)\nabla_s \left\{ \frac{1}{f_{S,N}(s,n)} \left\{ \frac{1}{1-s}, -m_L(s,n) \right\} \begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix} \right\} (s - p_H)$$
$$- (s\nabla_s m_H(s,n) + (1-s)\nabla_s m_H(s,n))(h_4(r) - h_{40}(r)).$$

By the chain rule we have

$$s\nabla_s \left\{ \frac{1}{f_{S,N}(s,n)} \left\{ \frac{1}{s}, -m_H(s,n) \right\} \begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix} \right\} (s - p_H)$$
$$= \nabla_s \begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix}' \left\{ \frac{s - p_H}{f_{S,N}(s,n)} \left[1, -sm_H(s,n) \right] \right\}'$$
$$+ \frac{s - p_H}{f_{S,N}(s,n)} \left\{ \left[-\frac{1}{s} - k(r), k(r)\, sm_H(s,n) - s\nabla_s m_H(s,n) \right] \right\} \begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix}$$

where

$$k(r) = \frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}.$$

Similarly we have

$$(1-s)\nabla_s \left\{ \frac{1}{f_{S,N}(s,n)} \left\{ \frac{1}{1-s}, -m_L(s,n) \right\} \begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix} \right\} (s - p_H)$$
$$= \nabla_s \begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix}' \left\{ \frac{s - p_H}{f_{S,N}(s,n)} \left[1, -(1-s)\, m_L(s,n) \right] \right\}'$$
$$+ \frac{s - p_H}{f_{S,N}(s,n)} \nabla_s \left\{ \frac{1}{1-s} - k(r), k(r)(1-s)\, m_L(s,n) - (1-s)\nabla_s m_L(s,n) \right\} \begin{pmatrix} h_2(r) - h_{20}(r) \\ h_3(r) - h_{30}(r) \end{pmatrix}.$$

Collecting terms and reorganizing yields the linearization

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

where

$$a_0(r) = \left\{ \frac{s - p_H}{f_{S,N}(s,n)} \left[-\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r)\, m(s,n) - e(s,n) \right], \right.$$
$$\left. -e(s,n) \right\}$$
$$b_0(r) = \frac{s - p_H}{f_{S,N}(s,n)} \left\{ 1, 1, -m(s,n), 0 \right\},$$

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

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

$$
\begin{aligned}
\mathbb{E}\left[a_0(R)' h(R;\eta)\right] = & \int \frac{s-p_H}{f_{S,N}(s,n)} \left[-\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r)\, m(s,n) - e(s,n)\right], \\
& -e(s,n)\} \\
& \times \begin{pmatrix} f_{S,N}(s,n;\eta)\, \mathbb{E}_\eta\left[TY|\, S=s, N=n\right] \\ f_{S,N}(s,n;\eta)\, \mathbb{E}_\eta\left[(1-T)\, Y|\, S=s, N=n\right] \\ f_{S,N}(s,n;\eta) \\ p_H(\eta) \end{pmatrix} f_{S,N}(s,n)\, \mathrm{d}s\mathrm{d}n \\
= & \int (s-p_H) \left[-\frac{1}{s} - k(r), \frac{1}{1-s} - k(r), k(r)\, m(s,n) - e(s,n)\right] \\
& -e(s,n)\, f_{S,N}(s,n)\} \\
& \times \begin{pmatrix} \mathbb{E}_\eta\left[TY|\, S=s, N=n\right] \\ \mathbb{E}_\eta\left[(1-T)\, Y|\, S=s, N=n\right] \\ 1 \\ p_H(\eta)/f_{S,N}(s,n;\eta) \end{pmatrix} f_{S,N}(s,n;\eta)\, \mathrm{d}s\mathrm{d}n \\
= & \, \mathbb{E}_\eta\left[v_1(R)' \{1, T, TY, (1-T)\, Y\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
v_1(r) = \Bigg\{ & (s-p_H)\left[\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)} m(s,n) - e(s,n)\right], -\mathbb{E}\left[e(S,N)\right], \\
& (s-p_H)\left(-\frac{1}{s} - k(r)\right), (s-p_H)\left(\frac{1}{1-s} - k(r)\right)\Bigg\}.
\end{aligned}
$$

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

$$
\mathbb{E}\left[\nabla_s h(R;\eta)' b_0(R)\right] = \mathbb{E}_\eta\left[v_2(R)\{1, T, TY, (1-T)\, Y\}'\right],
$$

with

$$
v_2(r) = (s-p_H)\{\nabla_s m(s,n), 0, 0, 0\}' + \{m(s,n), 0, -1, -1\}'.
$$

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

$$
\begin{aligned}
\delta^{\text{lepe}}(z) = & -\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)} (y - m(s,n))(s-p_H) \\
& - (y - m(s,n)) \\
& - \left\{\left(\frac{t}{s}\right) y - m_H(s,n)\right\}(s-p_H) + \left\{\left(\frac{1-t}{1-s}\right) y - m_L(s,n)\right\}(s-p_H) \\
& - \mathbb{E}\left[e(S,N)\right](t-p_H),
\end{aligned}
$$

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

C.4 Influence function derivation for $\widehat{\beta}^{\text{ase}}$

The average spillover effect is given by

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

where $h_{10}(r)$, $h_{10}(r)$ and $h_{10}(r)$ are as defined in (27) above. Linearizing the implied moment function gives

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

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

$$\Psi(r, h(r)) = a_0(r)' h(r) + \nabla_s h(r)' b_0(r) + c_0(r)' h(r) \tag{33}$$

with $h(r) = (h_1(r), h_2(r), h_3(r))'$ and

$$a_0(r) = \frac{1}{f_{S,N}(s,n)} \left(-k(r), -k(r), -\nabla_s m(s,n) + m(s,n) k(r)\right)'$$
$$b_0(r) = \frac{1}{f_{S,N}(s,n)} (1, 1, -m(s,n))'$$
$$c_0(r) = -\frac{1}{f_{S,N}(s,n)} \left(\frac{1}{s}, -\frac{1}{1-s}, -[m_H(s,n) - m_L(s,n)]\right)$$

where

$$k(r) = \frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}, \quad m(s,n) = \frac{h_{10}(r) + h_{20}(r)}{h_{30}(r)}, \quad m_H(s,n) = \frac{h_{10}(r)}{sh_{30}(r)}, \quad m_L(s,n) = \frac{h_{20}(r)}{(1-s) h_{30}(r)}.$$

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

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

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

$$v_1(r) = \{-\nabla_s m(s,n) + m(s,n) k(r), 0, -k(r), -k(r)\}.$$

To evaluate the second term of (33) we use integration by parts (with $u(r) = f_0(r) b_0(r)'$ and $v(r) = h(r;\eta)$) to obtain a representation directly in terms of $h(r;\eta)$:

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

with

$$v_2(r) = \{\nabla_s m(s,n), 0, 0, 0\}'.$$

This follows from the fact that

$$\begin{aligned}\nabla_s \left[f_0(r) b_0(r)'\right] &= \nabla_s \left[s^t (1-s)^{1-t} \{1, 1, -m(s,n)\}'\right] \\ &= s^t (1-s)^{1-t} \{0, 0, -\nabla_s m(s,n)\}' \\ &\quad + \left[\frac{t-s}{s(1-s)} s^t (1-s)^{1-t} \{1, 1, -m(s,n)\}'\right]\end{aligned}$$

and also that

$$\sum\nolimits_{t=0,1} \frac{t-s}{s(1-s)} s^t (1-s)^{1-t} = -1 + 1 = 0.$$

Evaluating the expectation of the third term in (33) gives

$$\begin{aligned}\mathbb{E}\left[c_0(R)' h(R;\eta)\right] &= -\int \frac{1}{f_{S,N}(s,n)} \left(\frac{1}{s}, -\frac{1}{1-s}, -\left[m_H(s,n) - m_L(s,n)\right]\right) \\ &\quad \times \begin{pmatrix} f_{S,N}(s,n;\eta)\, s m_H(s,n;\eta) \\ f_{S,N}(s,n;\eta)(1-s) m_L(s,n;\eta) \\ f_{S,N}(s,n;\eta) \end{pmatrix} f_{S,N}(s,n)\,\mathrm{d}s\mathrm{d}n \\ &= -\int \left(\frac{1}{s}, -\frac{1}{1-s}, -\left[m_H(s,n) - m_L(s,n)\right]\right) \\ &\quad \times \begin{pmatrix} \mathbb{E}_\eta\left[TY \mid S=s, N=n\right] \\ \mathbb{E}_\eta\left[(1-T)Y \mid S=s, N=n\right] \\ 1 \end{pmatrix} f_{S,N}(s,n;\eta)\,\mathrm{d}s\mathrm{d}n \\ &= \mathbb{E}_\eta\left[v_3(R)' \{1, T, TY, (1-T)Y\}\right],\end{aligned}$$

with

$$v_3(r) = \left\{m_H(s,n) - m_L(s,n), 0, -\frac{1}{s}, \frac{1}{1-s}\right\}.$$

Together these calculations suggest a correction term of the form

$$\begin{aligned}\delta^{\text{ase}}(z) &= -\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)} (y - m(s,n)) \\ &\quad - \left[\left\{\left(\frac{t}{s}\right) y - m_H(s,n)\right\} - \left\{\left(\frac{1-t}{1-s}\right) y - m_L(s,n)\right\}\right]\end{aligned} \tag{34}$$

as claimed.

C.5 Influence function derivation for $\widehat{\beta}^{\text{lsie}}$

The local segregation inequality effect (LSIE) is given by

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

where

$$\begin{aligned}\beta_H^{\text{lsie}} &= \mathbb{E}\left[\frac{1}{p_H}\left\{m_H(S,N) + S\nabla_s m_H(S,N)\right\}(S - p_H)\right] \\ &= \mathbb{E}\left[\frac{1}{h_{40}(R)}\left(\frac{h_{10}(R)}{S h_{30}(R)} + S\nabla_s \left\{\frac{h_{10}(R)}{S h_{30}(R)}\right\}\right)(S - h_{40}(R))\right]\end{aligned}$$

and

$$\beta_L^{\text{lsie}} = \mathbb{E}\left[\frac{1}{1-p_H}\left\{-m_L(S,N) + (1-S)\nabla_s m_L(S,N)\right\}(S-p_H)\right]$$
$$= \mathbb{E}\left[\frac{1}{1-h_{40}(R)}\left(-\frac{h_{20}(R)}{(1-S)h_{30}(R)} + (1-S)\nabla_s\left\{\frac{h_{20}(R)}{(1-S)h_{30}(R)}\right\}\right)(S-h_{40}(R))\right],$$

with $h(r) = (h_1(r), h_2(r), h_3(r), h_4(r))'$ as defined in (27) above.

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

$$\Psi(r, h(r) - h_0(r)) = \left\{\frac{1}{f_{S,N}(s,n)s}\frac{s-p_H}{p_H}, -\frac{1}{f_{S,N}(s,n)}\frac{s-p_H}{p_H}m_H(s,n), -\frac{s-p_H}{p_H^2}m_H(s,n) - \frac{s^2}{p_H^2}\nabla_s m_H(s,n)\right\}$$
$$\times\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r) \\ h_4(r) - h_{40}(r)\end{pmatrix}$$
$$+ \frac{s}{p_H}\nabla_s\left\{\frac{1}{f_{S,N}(s,n)}\left\{\frac{1}{s}, -m_H(s,n)\right\}\begin{pmatrix} h_1(r) - h_{10}(r) \\ h_3(r) - h_{30}(r)\end{pmatrix}\right\}(s-p_H).$$

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

$$\frac{s}{p_H}\nabla_s\left\{\frac{1}{f_{S,N}(s,n)}\left\{\frac{1}{s}, -m_H(s,n)\right\}\begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}\right\}(s-p_H)$$
$$= \nabla_s\begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}'\left[\frac{s}{p_H}\frac{s-p_H}{f_{S,N}(s,n)}\left\{\frac{1}{s}, -m_H(s,n)\right\}\right]'$$
$$+ \begin{pmatrix} h_1(r) \\ h_3(r)\end{pmatrix}'\frac{s}{p_H}\frac{s-p_H}{f_{S,N}(s,n)s}\left\{-1/s - k(r), k(r)sm_H(s,n) - s\nabla_s m_H(s,n)\right\}',$$

where

$$k(r) = \frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}.$$

Collecting terms allows us to write

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

with

$$a_0(r) = \left\{-\frac{1}{f_{S,N}(s,n)}\frac{s-p_H}{p_H}k(r), 0, \frac{1}{f_{S,N}(s,n)}\frac{s-p_H}{p_H}\left[-m_H(s,n) + sk(r)m_H(s,n) - s\nabla_s m_H(s,n)\right],\right.$$
$$\left.-\frac{s-p_H}{p_H^2}m_H(s,n) - \frac{s^2}{p_H^2}\nabla_s m_H(s,n)\right\}$$
$$b_0(r) = \frac{1}{f_{S,N}(s,n)}\frac{s-p_H}{p_H}\left\{1, 0, -sm_H(s,n), 0\right\}.$$

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

$$
\begin{aligned}
&\mathbb{E}\left[a_0(R)' h(R;\eta)\right] \\
&= \int \sum_{t=0,1} \left\{ -\frac{1}{f_{S,N}(s,n)} \frac{s-p_H}{p_H} k(r), \frac{1}{f_{S,N}(s,n)} \frac{s-p_H}{p_H} \left[-m_H(s,n) + sk(r) m_H(s,n) - s\nabla_s m_H(s,n) \right], \right. \\
&\left. -\frac{s-p_H}{p_H^2} m_H(s,n) - \frac{s^2}{p_H^2} \nabla_s m_H(s,n) \right\} \\
&\times \begin{pmatrix} f_{S,N}(s,n;\eta)\, \mathbb{E}_\eta\left[TY | S=s, N=n\right] \\ f_{S,N}(s,n;\eta) \\ p_H(\eta) \end{pmatrix} s^t (1-s)^{1-t} f_{S,N}(s,n) \mathrm{d}s\mathrm{d}n \\
&= \int \left\{ -\frac{s-p_H}{p_H} k(r), \frac{s-p_H}{p_H} \left[-m_H(s,n) + sk(r) m_H(s,n) - s\nabla_s m_H(s,n) \right], \right. \\
&\left. \left[-\frac{s-p_H}{p_H^2} m_H(s,n) - \frac{s^2}{p_H^2} \nabla_s m_H(s,n) \right] f_{S,N}(s,n) \right\} \\
&\times \begin{pmatrix} \mathbb{E}_\eta\left[TY | S=s, N=n\right] \\ 1 \\ \frac{p_H(\eta)}{f_{S,N}(s,n;\eta)} \end{pmatrix} f_{S,N}(s,n;\eta) \mathrm{d}r \\
&= \mathbb{E}_\eta\left[v_1(R)\{1, T, TY, (1-T)Y\}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
v_1(r) = &\left\{ \frac{s-p_H}{p_H} \left[-m_H(s,n) + sk(r) m_H(s,n) - s\nabla_s m_H(s,n) \right], \right. \\
&\left. -\mathbb{E}\left[\frac{S-p_H}{p_H^2} m_H(S,N) + \frac{S^2}{p_H^2} \nabla_s m_H(S,N) \right], -\frac{s-p_H}{p_H} k(r), 0 \right\}.
\end{aligned}
$$

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

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

with

$$
v_2(r) = \frac{s-p_H}{p_H}\{m_H(s,n) + s\nabla_s m_H(s,n), 0, 0, 0\}' + \frac{1}{p_H}\{s m_H(s,n), 0, -1, 0\}'.
$$

This follows from the fact that

$$
\begin{aligned}
\nabla_s \left[f_0(r) b_0(r)'\right] &= \nabla_s \left[s^t (1-s)^{1-t} f_{S,N}(s,n) \frac{1}{f_{S,N}(s,n)} \frac{s-p_H}{p_H} \{1, 0, -s m_H(s,n), 0\} \right] \\
&= \nabla_s \left[s^t (1-s)^{1-t} \frac{s-p_H}{p_H} \{1, 0, -s m_H(s,n), 0\} \right] \\
&= s^t (1-s)^{1-t} \frac{s-p_H}{p_H} \{0, 0, -m_H(s,n) - s\nabla_s m_H(s,n), 0\} \\
&+ \left[\frac{1}{p_H} \left((s-p_H) \frac{t-s}{s(1-s)} s^t (1-s)^{1-t} + s^t (1-s)^{1-t} \right) \{1, 0, -s m_H(s,n), 0\} \right],
\end{aligned}
$$

and also that

$$
\frac{1}{p_H} \sum_{t=0,1} (s-p_H) \frac{t-s}{s(1-s)} s^t (1-s)^{1-t} + s^t (1-s)^{1-t} = \frac{1}{p_H}.
$$

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

$$\begin{aligned}
\delta_H^{\text{lsie}}(z) = &-\frac{1}{p_H}\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}(ty - sm_H(s,n))(s-p_H) \\
&-\frac{1}{p_H}(ty - sm_H(s,n)) \\
&-\mathbb{E}\left[\frac{S-p_H}{p_H^2}m_H(S,N) + \frac{S^2}{p_H^2}\nabla_s m_H(S,N)\right](t-p_H).
\end{aligned}$$

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

$$\begin{aligned}
\delta_L^{\text{lsie}}(z) = &-\frac{1}{1-p_H}\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}((1-t)y - (1-s)m_L(s,n))(s-p_H) \\
&-\frac{1}{1-p_H}((1-t)y - (1-s)m_L(s,n)) \\
&-\mathbb{E}\left[\frac{S-p_H}{(1-p_H)^2}m_L(S,N) + \left(\frac{1-S}{1-p_H}\right)^2\nabla_s m_L(S,N)\right](t-p_H).
\end{aligned}$$

and hence $\delta(z) = \delta_H(z) - \delta_L(z)$ equal to

$$\begin{aligned}
\delta^{\text{lsie}}(z) &= \delta_H^{\text{lsie}}(z) - \delta_L^{\text{lsie}}(z) \\
&= -\frac{1}{p_H}\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}(ty - sm_H(s,n))(s-p_H) \\
&+\frac{1}{1-p_H}\frac{\nabla_s f_{S,N}(s,n)}{f_{S,N}(s,n)}((1-t)y - (1-s)m_L(s,n))(s-p_H) \\
&-\frac{1}{p_H}(ty - sm_H(s,n)) + \frac{1}{1-p_H}((1-t)y - (1-s)m_L(s,n)) \\
&-\mathbb{E}\left[\frac{S-p_H}{p_H^2}m_H(S,N) + \left(\frac{S}{p_H}\right)^2\nabla_s m_H(S,N)\right](t-p_H) \\
&+\mathbb{E}\left[\frac{S-p_H}{(1-p_H)^2}m_L(S,N) + \left(\frac{1-S}{1-p_H}\right)^2\nabla_s m_L(S,N)\right](t-p_H),
\end{aligned}$$

as claimed.

References

[1] Altonji, Joseph G. and Rosa L. Matzkin (2005). "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica* 73 (4): 1053 - 1102.

[2] Angrist, Joshua D., Kathryn Graddy and Guido W. Imbens. (2000). "The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish," *Review of Economics Studies* 67 (3): 499 - 527.

[3] Angrist, Joshua D. and Kevin Lang. (2004). "Does school integration generate peer effects? Evidence from Boston's Metco Program," *American Economic Review* 94 (5): 1613 - 1634.

[4] Arnott, Richard and John Rowse. (1987). "Peer group effects and educational attainment," *Journal of Public Economics* 32 (3): 287 - 305.

[5] Becker, Gary S. and Kevin M. Murphy. (2000). *Social Economics: Market Behavior in a Social Environment.* Cambridge, MA: Harvard University Press.

[6] Benabou, Roland. (1993). "Workings of a city: location, education, and production," *Quarterly Journal of Economics* 108 (3): 619 - 652.

[7] Benabou, Roland. (1996). "Equity and efficiency in human capital investment: the local connection," *Review of Economic Studies* 63 (2): 237 - 264.

[8] Bhattacharya, Debopam. (2008). "Inferring optimal resource allocations from experimental data," *Journal of the American Statistical Association*, forthcoming.

[9] Blundell, Richard W. and James L. Powell. (2003). "Endogeneity in nonparametric and semiparametric regression models," *Advances in Economics and Econometrics: Theory and Applications II*: 312 - 357. (M. Dewatripont, L.P. Hansen, S. J. Turnovsky, Eds.). Cambridge: Cambridge University Press.

[10] Brock, William A. and Steven N. Durlauf. (2001). "Interactions-based Models," *Handbook of Econometrics* 5: 3297 - 3380 (J. Heckman & E. Leamer, Eds.). Amsterdam: North-Holland.

[11] Brock, William A. and Steven N. Durlauf. (2007). "Identification of binary choice models with social interactions," *Journal of Econometrics* 140 (1): 52 - 75.

[12] Calvó-Armengol, Antoni, Eleonora Patacchini and Yves Zenou. (2006). "Peer effects and social networks in education," *Mimeo.*

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

[14] Ciccone, Antonio and Giovanni Peri. (2006). "Identifying human-capital externalities: theory with applications," *Review of Economic Studies* 73 (2): 381 - 412.

[15] Coleman, James S. et al. (1966). *Equality of Educational Opportunity.* Washington: U.S. Department of Health, Education and Welfare.

[16] de Bartolome, Charles A. M. (1990). "Equilibrium and inefficiency in a community model with peer group effects," *Journal of Political Economy* 98 (1): 110 - 133.

[17] Duncan, Greg J. and Stephen Raudenbush. (2001). "Neighborhoods and adolescent development: how can we determine the links?" *Does it Take a Village? Community Effects on Children, Adolescents, and Families*: 1005 - 136 (A. Booth and A. C. Crouter, Eds.). State College, PA: Pennsylvania State University Press.

[18] Durlauf, Steven N. (1996a). "Neighborhood feedbacks, endogenous stratification, and income inequality," *Dynamic Disequilibrium Modeling*: 505 - 534 (W. A. Barnett et al., Eds.). Cambridge: Cambridge University Press.

[19] Durlauf, Steven N. (1996b). "A theory of persistent income inequality," *Journal of Economic Growth* 1 (1): 75 - 93.

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

[21] Durlauf, Steven N. (2001). "A framework for the study of individual behavior and social interactions," *Social Methodology* 31 (1): 47 - 87.

[22] Durlauf, Steven. N. (2004). "Neighborhood effects," *Handbook of Regional and Urban Economics* 4: 2173 - 2242 (J.V. Henderson & J.F. Thisse). Amsterdam: North Holland.

[23] Epple, Dennis and Richard Romano. (1998). "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review* 88 (1): 33 - 62.

[24] Epple, Dennis, Elizabeth Newlon and Richard Romano. (2002). "Ability tracking, school competition, and the distribution of educational benefits," *Journal of Public Economics* 83 (1): 1 - 48.

[25] Fernández, Raquel. (2003). "Sorting, education and inequality," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* 2: 1 - 40 (M. Dewatripont, L. P. Hansen, S.J. Turnovsky, Eds.). Cambridge: Cambridge University Press.

[26] Figlio, David N. and Marianne E. Page. (2002). "School choice and the distributional effects of ability tracking: does separation increase inequality?" *Journal of Urban Economics* 51 (3): 497 - 514.

[27] Ginsberg, William. (1974). "The multiplant firm with increasing returns to scale," *Journal of Economic Theory* 9 (3): 283 - 292.

[28] Glaeser, Edward and José A. Scheinkman. (2003). "Nonmarket interactions," *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress* 1: 339 - 369 (M. Dewatripont, L. P. Hansen, S.J. Turnovsky, Eds.). Cambridge: Cambridge University Press.

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

[30] Graham, Bryan S. (2008b). "On the identification of neighborhood effects in the presence of endogenous neighborhood selection," *Mimeo.*

[31] Graham, Bryan S., Geert Ridder and Guido W. Imbens (2006). "The effect of reallocating an indivisible input with discrete values," *Mimeo.*

[32] Graham, Bryan S., Geert Ridder and Guido W. Imbens (2007). "Complementarity and aggregate implications of assortative matching: a nonparametric analysis," *Mimeo.*

[33] Guryan, Jonathan. (2004). "Desegregation and black dropout rates," *American Economic Review* 94 (4): 919 - 943.

[34] Heckman, James J. and Edward J. Vytlacil. (2007). "Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation," *Handbook of Econometrics* 6 (2): 4779 - 4874 (J. J. Heckman & E. Leamer, Eds.). Amsterdam: North - Holland.

[35] Holland, Paul W. (1986a). "Statistics and causal inference," *Journal of the American Statistical Association* 81 (396): 945 - 960.

[36] Holland, Paul W. (1986b). "Statistics and causal inference: rejoinder," *Journal of the American Statistical Association* 81 (396): 968 - 970.

[37] Horst, Reiner, Panos M. Pardalos and Nguyen V. Thoai. (2000). *Introduction to Global Optimization*, 2nd Ed. Dordrecht: Kluwer Academic Publishers.

[38] Hoxby, Caroline M. (2003). "School choice and school competition: evidence from the United States," *Swedish Economic Policy Review* 10 (1): 9 - 65.

[39] Ibaraki, Toshihide and Naoki Katoh. (1988). *Resource Allocation Problems: Algorithmic Approaches.* Cambridge, MA: The MIT Press.

[40] Imbens, Guido W. (2004). "Nonparametric estimation of average treatment effects under exogeneity: a review," *Review of Economics and Statistics* 86 (1): 4 - 29.

[41] Imbens, Guido W. (2006). "Nonadditive models with endogenous regressors," *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress* 3: 17 - 46 (R. Blundell, W. Newey & T. Persson, Eds.). Cambridge: Cambridge University Press.

[42] Jacob, Brian A. (2004). "Public housing, housing vouchers and student achievement: evidence from public housing demolitions in Chicago," *American Economic Review* 94 (1): 233 - 258.

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

[44] Krueger, Alan B. and Diane M. Whitmore. (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.

[45] Ladd, Helen F. (2003). "Comment on Caroline M. Hoxby: School choice and school competition: evidence from the United States," *Swedish Economic Policy Review* 10 (1): 67 - 76.

[46] Lazear, Edward P. (2001). "Educational production," *Quarterly Journal of Economics* 116 (3): 777 - 803.

[47] Luenberger, David G. (1969). *Optimization by Vector Space Methods.* New York: John Wiley & Sons.

[48] Luenberger, David G. (2005). *Linear and Nonlinear Programming*, 2nd Ed. New York: Springer.

[49] Mael, Fred et al. (2005). *Single-sex versus coeducational schooling: a systematic review.* Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

[50] Manski, Charles F. (1992). "Educational choice (vouchers) and social mobility," *Economics of Education Review* 11 (4): 351 - 369.

[51] Manski, Charles F. (1993). "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies* 60 (3): 531 - 542.

[52] Manski, Charles F. (2000). "Economic analysis of social interactions," *Journal of Economic Perspectives* 14 (3): 115 - 136.

[53] Massey, Douglas S. and Nancy A. Denton. (1988). "The dimensions of residential segregation," *Social Forces* 67 (2): 281 - 315.

[54] Massey, Douglas S. and Nancy A. Denton. (1993). *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA: Harvard University Press.

[55] Mayer, Susan E. and Christopher Jencks. (1989). "Growing up in poor neighborhoods: how much does it matter," *Science* 243 (4897): 1441 - 1445.

[56] McFadden, Daniel L. (1978). "Cost, revenue, and profit functions," *Production Economics: A Dual Approach to Theory and Applications, Volume I: The Theory of Production*: 3 - 109 (M. Fuss and D. L. McFadden, Eds.). Amsterdam: North-Holland.

[57] Moffitt, Robert A. (2001). "Policy interventions, low-level equilibria and social interactions," *Social Dynamics*: 45 - 82 (S. Durlauf & P. Young, Eds.). Cambridge, MA: MIT Press.

[58] Morse, Susan, Ed. (1998). *Separated by Sex: A Critical Look at Single Sex Education for Girls.* Washington, D.C.: American Association of University Women Educational Foundation.

[59] Nesheim, Lars. (2002). "Equilibrium sorting of heterogenous consumers across locations: theory and empirical implications," *CEMMAP Working Paper* CWP08/02.

[60] Newey, Whitney K. (1994a). "The asymptotic variance of semiparametric estimators," *Econometrica* 62 (6): 1349 - 1382.

[61] Newey, Whitney K. (1994b). "Kernel estimation of partial means and a general variance estimator," *Econometric Theory* 10 (2): 233 - 253.

[62] Newey, Whitney K. and Daniel McFadden. (1994). "Large sample estimation and hypothesis testing," *Handbook of Econometrics 4*: 2111 - 2245 (R.F. Engle & D.L. McFadden). Amsterdam: North Holland.

[63] Newey, Whitney K. and Thomas M. Stoker. (1993). "Efficiency of weighted average derivative estimators and index models," *Econometrica* 61 (5): 1199 - 1223.

[64] Neyman, Jerzy. (1923) 1990. "On the application of probability theory to agricultural experiments. Essays on principles. Section 9," (D.M. Dabrowska & T.P. Speed, Trans), *Statistical Science* 5 (4): 465 - 472.

[65] Oakes, Jeannie. (1992). "Can tracking research inform practice? Technical, normative, and political considerations," *Educational Researcher* 21 (4): 12 - 21.

[66] Paulson, Amanda and Stacy A. Teicher. (2006). "Move to single-sex classes fans debate," *Christian Science Monitor*: October 26.

[67] Piketty, Thomas and Mathieu Valdenaire. (2006). "L'Impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises : une estimation à partir du panel primaire 1997 et du panel secondaire 1995," *Les Dossiers Enseignement Scolaire 173.* Paris: Ministère de l'Èducation nationale.

[68] Piketty, Thomas. (2000). "Theories of persistent inequality and intergenerational mobility," *Handbook of Income Distribution* 1: 430 - 476 (A. Atkinson & F. Bourguignon). Amsterdam: North Holland.

[69] Powell, James L, James H. Stock and Thomas M. Stoker. (1989). "Semiparametric estimation of index coefficients," *Econometrica* 57 (6): 1403 -1430.

[70] Rivkin, Steven and Finis Welch (2006). "Has school desegregation improved academic and economic outcomes for blacks," *Handbook of the Economics of Education* 2: 1019 - 1049 (E. Hanushek & F. Welch, Eds.). Amsterdam: North-Holland.

[71] Rockafellar, R. Tyrrell. (1970). *Convex Analysis.* Princeton, NJ: Princeton University Press.

[72] Rubin, Donald B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology* 66 (5): 688 - 701.

[73] Rubin, Donald B. (1981). "Estimation in parallel randomized experiments," *Journal of Educational Statistics* 6 (4): 377 - 400.

[74] Rubin, Donald B. (1990). "Formal mode of statistical inference for causal effects," *Journal of Statistical Planning and Inference* 25 (3): 279 - 292.

[75] Sampson, Robert J., Jeffrey Morenoff and Thomas Gannon-Rowley. (2002). "Assessing neighborhood effects: social processes and new directions in research," *Annual Review of Sociology* 28 (1): 443 - 478.

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

[77] Schofield, Janet Ward. (1995). "Review of research on school desegregation's impact on elementary and secondary school students," *Handbook of Research on Multicultural Education*: 597 - 616 (J.A. Banks & C.A.M. Banks). New York: Macmillan Publishing USA.

[78] Sobel, Michael E. (2006a). "Spatial concentration and social stratification: does the clustering of disadvantage 'beget' bad outcomes?" *Poverty Traps*: 204 - 229 (S. Bowles, S. N. & K. Hoff, Eds.). Princeton, NJ: Princeton University Press.

[79] Sobel, Michael 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.

[80] Solon, Gary. (1999). "Intergenerational mobility in the labor market," *Handbook of Labor Economics* 3: 1761 - 1800 (O. Ashenfelter & D. Card, Eds.). Amsterdam: North-Holland.

[81] Urquiola, Miguel. (2005). "Does school choice lead to sorting? Evidence from Tiebout variation," *American Economic Review* 95 (4): 1310 - 1326.

[82] Weinberg, Bruce. (2006). "Social interactions and endogenous association," *Mimeo.*

[83] Weyl, Hermann. (1946). *The Classical Groups: Their Invariants and Representations.* Princeton, NJ: Princeton University Press.