CHAPTER 19

Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers[1]

Bryan S. Graham[†]

Contents

1. Introduction 967
2. An overview of empirical matching models 969
   - 2.1 Some illustrative examples 970
   - 2.2 Econometric research on matching problems 972
3. Identification and estimation of one-to-one matching models when match output is observed 974
   - 3.1 The structure of feasible assignments in one-to-one matching problems 975
     - *3.1.1 Double randomization 978*
     - *3.1.2 Conditional double randomization 980*
     - *3.1.3 No matching on unobservables 981*
   - 3.2 Average reallocation effects (AREs) 984
     - *3.2.1 Definition of target estimands 984*
     - *3.2.2 Identification under double randomization 988*
     - *3.2.3 Identification under no matching on unobservables 990*
     - *3.2.4 Further thoughts on the identification of AREs 992*
   - 3.3 Continuously-valued match characteristics 993
   - 3.4 Estimation of AREs when match output is observed 997
     - *3.4.1 Estimation of the average match output function (AMF) 997*
     - *3.4.2 AREs with discretely-valued covariates 998*
     - *3.4.3 AREs with continuously-valued covariates 1005*

[1] I would like to thank Stephane Bonhomme, Charles Manski, Konrad Menzel, and Geert Ridder for helpful discussions. More generally my thinking on the material presented below has benefited from discussions with my co-authors, Guido Imbens and Geert Ridder, and with members of the former MacArthur Research Network on Economic Inequality and Social Interactions (especially Steven Durlauf). Some of the new material in this chapter was presented to the applied micro-economics group at the Federal Reserve Bank of New York to whom I am thankful for feedback. I have especially benefited from comments on an earlier draft by Pierre-André Chiappori, Jeremy Fox, Guido Imbens, Jim Powell, and Bernard Salanie. Financial support from the National Science Foundation (SES #0820361) is gratefully acknowledged. All the usual disclaimers apply.

[†] Department of Economics, New York University, 19 West 4th Street, 6FL, New York, NY 10012 and NBER. E-Mail: bryan.graham@nyu.edu, Web: https://files.nyu.edu/bsg1/public/.

Handbook of Social Economics, Volume 1B
ISSN 0169-7218, DOI: 10.1016/S0169-7218(11)01024-0

4. Identification and estimation of one-to-one matching models when match output is unobserved: equilibrium approaches 1007
4.1 A two-sided model of multinomial choice 1008
4.2 Parametric identification of AREs when match output is unobserved 1012
4.3 Nonparametric identification 1015
4.3.1 Pairwise logit identification 1016
4.3.2 Nonparametric identification in a 2 × 2 matching market 1018
4.3.3 Nonparametric identification in a K × L matching market 1021
4.4 Identification in one-to-one matching markets without transfers 1025
4.5 Estimation on the basis of match characteristics alone 1026
4.5.1 Testing for supermodularity 1026
4.5.2 Parametric modeling of average match surplus 1027
4.5.3 Pairwise logit estimation 1028
4.5.4 Pairwise maximum score estimation 1029
5. Segregation in the presence of social spillovers 1029
5.1 Population setup 1031
5.2 Global reallocations: the social planner's problem 1034
5.3 Local reallocations: movers versus stayers 1039
5.4 Parametric examples 1042
6. Treatment response with spillovers 1044
7. Areas for further research 1046
References 1047

Abstract

Concern over the distributional effects of policies which induce changes in peer group structure, or 'associational redistributions' (Durlauf, 1996c), motivates a substantial body of theoretical and empirical research in economics, sociology, psychology, and education. A growing collection of econometric methods for characterizing the effects of such policies are now available. This chapter surveys these methods. I discuss the identification and estimation of the distributional effects of partner reassignment in one-on-one matching models, the average outcome and inequality effects of segregation, and treatment response in the presence of spillovers.

JEL Classification: C14, C25, C60, C61, J4, J12, I21

Keywords

matching models
assignment problems
peer group effects
social spillovers
segregation
associational redistribution
complementarity
treatment effects
semiparametric M-estimation

1. INTRODUCTION

Individuals exercise substantial control over who they associate with. Sometimes directly, as is the case with spousal choice, and sometimes indirectly, as when a family, by choosing to reside in a certain neighborhood, gains access to specific schools (and hence peers) for their children. Associations or reference groups, such as families, co-workers, neighbors and classmates, define (partially) isolated environments in which social interaction takes places (Durlauf, 1996c). These interactions may, in turn, affect the acquisition of human capital, the availability of employment opportunities, or even influence one's aspirations and values. If this is the case, then inequality in social context may contribute to socioeconomic inequality.[2] Conversely, policies which alter the composition of social groups – 'associational redistributions' – either directly or indirectly (by changing the incentives governing their formation) can lessen inequality (Durlauf, 1996c).

Many of the most controversial social policies in the United States involve associational redistribution. Examples include affirmative action in admissions and hiring, school desegregation policies, school assignment policies, single-sex schooling within the public system, active labor market policies, and whether public housing should be concentrated in a small number of locations or dispersed throughout a metropolitan area. That these policies are controversial is understandable: many individuals view their choice of associates as beyond the (direct) purview of public policy (cf., Piketty, 2000). Their controversial nature, however, is not wholly political. It also stems from uncertainty surrounding their effects on average outcomes and inequality.

This chapter reviews econometric methods for evaluating the effects of reallocations on the distribution of outcomes. In the social economics context 'reallocations' coincide with associational redistributions. The methods outlined in this chapter, however, are also of relevance to researchers in the fields of empirical industrial organization, labor economics, public finance, educational studies, sociology, and public health.

Reallocations are distinguished from other policies by the fact that they involve no augmentation, only redistribution, of resources. The study of reallocations necessitates some foundational thinking. Consequently, this review devotes a substantial amount of space to issues of measurement. Particularly to defining and motivating estimands which measure the effects of reallocations.

Effective policy-making requires knowledge of the causal mapping from group composition into outcomes. Consider the design of a school voucher program. Calibrated theoretical models suggest that any meaningful school voucher program would generate large changes in the distribution of students, and hence peer groups, across schools (e.g., Manski, 1992; Nechyba, 2006; Ferreyra, 2007). The magnitude and structure of any peer group effect in learning would be an important determinant of

[2] Loury (1977, 2002), for example, argues that segregated social networks generate inequality across racial groups.

these changes. It would also determine their effects on the level and distribution of student achievement. For these reasons knowledge of the exact form of any peer spillover is required for optimal voucher design. Unfortunately, this information is not provided by extant empirical research (cf., Piketty, 2000; Fernández, 2003).

Knowledge of the average effect of a unit increase in measured peer quality, the target estimand of many papers, is only indirectly helpful (e.g., Angrist and Lang, 2004; Card and Rothstein, 2007). This estimand measures the effect of an infeasible policy. Not all individuals' peer groups may be improved simultaneously; raising peer quality in some schools or classrooms requires lowering it in others. If the target policy is a reassigning one, then this will necessarily influence the precise form econometric analysis should take.

As a second example, consider the relationship between teacher quality and student achievement. Few educators, parents, or students doubt the centrality of teachers in the learning process (Jacob and Lefgren, 2007). Furthermore measured teacher quality varies substantially across schools (e.g., Buddin and Zamarro, 2009). Yet, given the structure of a typical teacher labor market, is seems unlikely that the observed assignment of teachers to schools corresponds to one which, say, maximizes student achievement.

One response to these observations is to implement policies that attempt to change the distribution of teacher quality (e.g., policies which encourage highly able young people to enter the teaching profession). Another, not mutually exclusive approach, involves reassigning teachers across schools. Depending on the nature of the educational production function it may be possible to raises student achievement – *holding the population of available teachers fixed* – by such reassignments.

Assignment problems have been widely studied by economists as well as those in operations research (e.g., Koopmans and Beckmann, 1957; Gale, 1960; Roth and Sotomayor, 1990; Burkard, Dell'Amico and Martello, 2009). Adding statistical content to these problems in a manageable way is nontrivial. Doing so raises a number of interesting and challenging econometric issues which are explored below.

Sections 2 begins with a brief overview of empirical work on one-to-one matching problems. Econometric research on this class of models may be divided into two categories. In the first, which is the subject of Section 3, the econometrician observes the match outcome of interest in addition to match characteristics. The goal is to recover the match production function from these data and evaluate the effects of alternative assignments or 'matchings' on the distribution of outcomes. In the second, which is the subject of Section 4, only match characteristics are observed. Here the question is what does one's choice of match partner alone reveal about preferences? There are connections between this question and the revealed preference approach to single agent models of discrete choice (McFadden, 1974).[3]

[3] As in most situations they are unobserved by the researcher, this chapter does not explore the identifying content of transfers between agents.

Section 5 studies a setting where social groups consist of a large number of agents, for example neighborhoods or classrooms. Agents are binary-typed and heterogenous in unobserved ability. Outcomes may vary with the type composition of one's social group. de Bartolome (1990), Benabou (1993, 1996) and Durlauf (1996a,b) study non-stochastic versions of this set-up. While these papers have been influential in shaping economists' intuitions about the equity and efficiency implications of residential segregation, their effect on empirical work has been more indirect.[4] Section 5 outlines one way to bring these models to the data.

Section 6 studies treatment response in the presence of spillovers. Here influencing the structure of reference groups is not the policy-maker's goal. Instead, the policy-maker seeks to account for the effects of these groups when formulating an individualistic policy. A pro-typical example involves optimal vaccine policy (e.g., Manski, 2009a,b).

This chapter does not review the growing literature on identifying peer effects per se (e.g., Manski, 1993; Moffitt, 2001; Brock and Durlauf, 2001a; 2007; Glaeser and Scheinkman, 2001, 2003; Graham, 2008). This literature is, of course, very much related to the material surveyed here. Several good surveys of this material are now available; including those of Brock and Durlauf (2001b), Durlauf (2004), Epple and Romano (this Handbook), and Blume, Brock, Durlauf and Ioannides (this Handbook). I also ignore models where the study of strategic interaction *within* groups is central. Interactions of this type, which feature in the work of Manski (1993, 2010) and Brock and Durlauf (2001a), are likely to be relevant in practice and important for some policy questions, but a meaningful treatment of them would require a separate survey (cf., Blume, Brock, Durlauf and Ioannides in this Handbook).[5] Finally, while I often refer to empirical work in what follows, no comprehensive review is attempted.

2. AN OVERVIEW OF EMPIRICAL MATCHING MODELS

Consider a 'market' composed of two heterogeneous populations, say, 'firms' and 'workers' (i.e., men and women, teachers and students, etc.). Units in each population may either self-produce or costlessly seek out a partner from the other population to engage in joint production. Such settings generate one-to-one assignment or matching problems. Matching models play important roles in many areas of economics. They were famously used by Becker (1973, 1974) to characterize 'marriage markets' (cf., Mortesen, 1988; Chiappori and Oreffice, 2008). Other important applications include the study of job matching (e.g., Crawford and Knoer, 1981; Kelso and Crawford, 1982), housing markets (Shapley and Scarf, 1974), auctions (Hatfield and Milgrom,

[4] See Piketty (2000, pp. 462 – 467) and Fernández (2003, p.14) for related discussions.

[5] The econometric study of games is an important project of empirical industrial organization (e.g., Aradillas-López and Tamer, 2008).

2005; Edelman, Ostrovsky and Schwarz, 2007), supply chains (Ostrovsky, 2008), and the determinants of wage inequality (Sattinger, 1980, Kremer and Maskin, 1996).

Koopmans and Beckmann (1957) and Shapley and Shubik (1971) initiated the study of matching problems in economics. They considered the transferable utility case where, in addition to the assignment or matching, the division of match output between partners is determined in equilibrium.[6] Gale and Shapley (1962) studied the case where agents have preferences over different candidate partners, but are unable to make transfers to them.[7]

Each firm and worker has a utility function, allowing them to rank the desirability of different matches. When utility is transferable across match partners, an equivalent representation of agent utilities is in terms of a match-specific surplus and transfer. Theorists treat these objects as primitives. An econometrician, in contrast, might ask under what conditions they are identified by the joint distribution of match outcomes and/or partner characteristics. Given identifiablility questions of estimation and inference remain.

Identifying the form of agent preferences or, when utility is transferrable, the match surplus function allows the econometrician to undertake predictive exercises. Two types of predictions are of particular interest. First, one might want to characterize how counterfactual assignments (or policies which induce re-assignment as a by product), alter the distribution of outcomes. Second, one might want to understand how changes in the primitives of the market, for example the availability of certain types of workers or firms, affects the equilibrium assignment.

The first question involves reallocations. Reallocations, unlike many policies more widely studied in economics, *do not* involve changes in resource availability. Reallocations leave the distribution of agent characteristics unchanged, only agent pairings are changed. The second question *does* involve changes in the distribution of agent characteristics, but recognizes that the effects of such changes are filtered through an equilibrium assignment process.

2.1 Some illustrative examples

Some empirical examples, and associated policy questions, help to motivate the material that follows. As a first, canonical example, consider a firm that must assign distinct tasks to heterogenous workers. A set of characteristics for each task and worker are observed. Also available is a (historical) dataset with information on past worker characteristics, assigned tasks, and estimates of productivity. How should a social planner use this dataset to guide assignments? One organization with considerable interest in

[6] In what follows I will call a pairing of *two specific* agents a 'match' or a 'pairing'. I will call an assignment of *all* agents an 'allocation', 'assignment' or a 'matching'.

[7] The theoretical analysis of assignment problems remains an active research area in economics and operations research (e.g., Roth and Sotomayor, 1990; Burkard, Dell'Amico and Martello, 2009). Much of this literature focuses on variants of two questions. First, what form does a surplus-maximizing assignment take? Second, are there decentralized mechanisms, which lead to such an assignment? (cf., Roth and Vande Vate, 1990; Roth, 2008).

such questions is the United States Military. A modest literature, surveyed by Warner and Asch (1995), documents the relationship between various enlistee characteristics, such as Armed Forces Qualification Test (AFQT) score, and military performance (e.g., Fernandez, 1992). Such studies can inform the debate regarding the returns to increasing measured enlistee quality.

A different question is how can the Armed Forces best use those enlistees available right now? Optimally assigning enlistees to tasks could generate sizeable increases in military productivity (cf., Carrell, Fullerton and West, 2009). Implementing such a policy would require no augmentation of resources, the pool of available workers and set of tasks are left unchanged. The question is of more than intellectual interest: the military has substantial latitude over how it may employ its personnel (as do many other large organizations).

Personnel-assignment problems are widely studied in the field of operations research (e.g., Gale, 1960; Luenberger, 2005). The novelty here is statistical content: the mapping from match attributes, in this case worker and task characteristics, into outcomes is both stochastic and unknown.

A second example involves educational policy. Lankford, Loeb and Wyckoff (2003) document widespread differences in measured teacher quality across schools in New York City. These differences, in conjunction with residential segregation, generate substantial differences in average teacher quality across demographic groups. Understanding the mechanics of teacher-to-school matching could aid in the design of policies, which raise student achievement and/or reduce disparities in teacher quality across schools.

In a companion paper, Loeb, Boyd, Lankford and Wyckoff (2003) argue that the teacher labor market resembles a two-sided matching model without transfers.[8] They assume that assignment follows the deferred acceptance procedure of Gale and Shapley (1962): schools make offers to their most preferred candidate, candidates reject those offers which are either dominated by other available offers or unemployment. This process continues until all positions are filled or a school is unable to find an acceptable candidate among those still available.[9] They estimate the parameters of employer (school) and teacher utility functions by the method of simulated moments.

With employer and employee preference estimates in hand it becomes possible, at least in principle, to forecast the effects of alternative policies. For example, if teachers prefer small class sizes, changes in the distribution of class size across schools would change the equilibrium assignment of teachers to schools. The induced re-assignment of teachers to schools thus becomes a consideration in the formulation of class size policy.[10]

[8] They argue that collective bargaining agreements prevent school-specific wages from adjusting to equilibrate supply and demand.

[9] This algorithm leads to an employer-optimal stable matching (cf., Roth and Sotomayor, 1990).

[10] For example, a reduction in average class size in predominately minority schools might, indirectly, lead to an increase in average measured teacher quality in them.

A third example is provided by Baccara, Imrohoroglu, Wilson and Yariv (2009) who study the office choices of a group of academics who are connected through friendship and coauthorship networks. Since individuals may value physical proximity to those in their network, their choice of office affects the utility of others. An equilibrium assignment, even when transfers between agents are possible, need not be optimal. Under certain assumptions an individual's choice of office may provide information about her valuation of proximity to network partners. The efficiency of alternative assignments can then be compared to the status quo. The presence of externalities suggests that large welfare gains may be available via reallocation.

As a final example consider the empirical analysis of marriage markets (e.g., Kremer, 1997; Choo and Siow, 2006a,b; Chiappori and Oreffice, 2008). Men are rivals with one another when attempting to match with women and vice versa (Becker 1973, 1974). The distribution of men and women available for marriage, as well as the nature of any surplus generated by marriage, drives marriage patterns. These patterns influence, among other outcomes, the intra-household division of resources, the acquisition of human capital, fertility decisions and the evolution of inequality across generations. Empirical models of marriage markets consequently play important roles in many areas of family and household economics (cf., Weiss, 1997).

2.2 Econometric research on matching problems

Despite their prominent role in many areas of economics, comparatively little work explores the econometric implications of matching models. Formal research on the econometrics of matching is of relatively recent origin; with many of the key papers as yet unpublished.[11]

Graham, Imbens and Ridder (2007, 2009a) introduced reallocation problems to the econometrics literature. They study nonparametric estimation of, and inference on, *average reallocation effects* (AREs) – differences in expected outcomes across feasible assignments. Restrictions on the status quo assignment identify the match production function, which is then averaged over alternative allocations. Graham, Imbens and Ridder (2007) assume that match characteristics are discretely-value. This assumption makes inference on average outcome-maximizing allocations feasible. Bhattacharya (2009), also working in the discrete case, extends this work in a number of ways (e.g., by considering other notions of optimality). Graham, Imbens and Ridder (2009a) consider the case where match characteristics are continuously-valued. Since inference on optimal allocations is difficult in this case, they introduce a semiparametric family of reallocations, and present identification and estimation results for it.

[11] A sophisticated literature in empirical labor economics studies structural models of search and matching (e.g., Eckstein and Wolpin, 1990; Flinn, 2006). Here my focus is on frictionless assignment models. Non-stochastic versions of these models are widely-studied in the literature on linear and nonlinear programming (e.g., Luenberger, 2005). The game theoretic approach to such problems is summarized by Roth and Sotomayor (1990).

In some settings, match surplus may be difficult to observe and/or measure. In such situations, it is of interest to study what can be learned from data on the characteristics of paired agents alone. The analogy with the revealed preference approach to consumer behavior is quite sharp: under what conditions can an agent's choice of partner reveal the nature of her preferences? The two-sided nature of matching problems distinguish them from traditional discrete-choice models of consumer demand (e.g., McFadden, 1974; Domencich and McFadden, 1975). Both parties of a partnership are complicit in its formation, hence individuals are not unconstrained in their choice of partners.[12] This suggests that in the matching context choice data reveal less about preferences than in the textbook discrete choice model of consumer behavior.

Dagsvik (2000) appears to be the first in the econometrics literature to consider what the distribution of match characteristics alone reveals.[13] Choo and Siow (2006a,b) develop a closely related framework (henceforth the 'CS model' or 'CS framework'), which has been extended by Chiappori, Salanié and Weiss (2010), Galichon and Salanié (2009), and Siow (2009).[14] Associated with each unit is a discretely valued observed characteristic as well as an unobserved, continuously valued, characteristic.[15] The unobserved characteristic indexes heterogeneity in preferences for different types in the opposing population. By combining restrictions on how unobservables affect match production with the assumption that the observed assignment satisfies pairwise stability, they show that the net match surplus function is identified up to scale.[16] This result relies on parametric distributional assumptions.

Fox (2009a, b), in the first explicitly nonparametric treatment, also explores what can be learned from data on partner characteristics alone. His approach is based on a 'rank order property'. Consider two assignments, the rank order property states that the assignment which generates more surplus in a deterministic version of the model (i.e., one with no unobserved agent heterogeneity and/or match-specific output 'shocks'), will be more frequently observed in the data. Although the rank order property is intuitive, it can be difficult to justify primitively. Nevertheless his approach has already been used in several applied papers (e.g., Fox and Bajari, 2009; Yang, Shi and Goldfarb, 2009; Baccara, Imrohoroglu, Wilson and Yariv, 2009).

[12] Echenique, Lee and Shum (2010) make this point quite elegantly. In a matching market equilibrium an individual may choose A over B even if she prefers B. This is because B may be unavailable. Unlike in a single agent discrete choice model, revealed preference is ambiguous.

[13] In sociology, there is a small literature on two-sided logit models (Logan 1998, Logan, Hoff and Newton, 2008) which is also explicitly grounded in economic models of matching.

[14] Dagsvik (2000) considers the non-transferable utility case, while Choo and Siow (2006a,b) assume transferable utility.

[15] The discretely-value characteristic may be a a composite of multiple primitive characteristics (e.g., age, gender, years of schooling). The continuously valued characteristic is vector-valued (see below).

[16] A natural notion of equilibrium in matching problems with transfers is *pairwise stability*: an assignment (and associated set of transfers between units) corresponds to an equilibrium if no two pairs of agents can raise their total surplus by exchanging partners. In one-to-one matching games with transfers pairwise stable assignments are generically unique, although a continuum of transfers may sustain them (e.g., Shapley and Shubik, 1971; Roth and Sotomayor, 1990).

A goal of the sections that follow, is to bring out common features of each of the approaches mentioned above. To do this I break down a prototypical empirical matching problem into three parts. First, associated with each matching market is a set of feasible assignments. Two marginal distributions describe the distribution of observed and unobserved agent attributes on each side of the market. A feasible assignment is a joint distribution of partner attributes, *both observed and unobserved*, consistent with these two marginals (Graham, Imbens, and Ridder, 2007). Second, each match generates output or surplus. Properties of the match surplus function are important for counterfactual policy analysis. Third, there is a matching process that generates the observed assignment; this assignment may correspond to a decentralized equilibrium or be exogenously determined. Identifying the outcome effects of reallocations require restrictions on one or more of these parts of the problem. The necessary assumptions vary with the question begin asked, which is embodied in the target estimand. I assume that the econometrician knows the joint distribution of observed partner characteristics. In some situations, she may also know the conditional distribution of match outcomes given observed partner characteristics.

3. IDENTIFICATION AND ESTIMATION OF ONE-TO-ONE MATCHING MODELS WHEN MATCH OUTPUT IS OBSERVED

This section outlines econometric methods for the analysis of one-to-one matching problems appropriate for situations where match output, in addition to match characteristics is observed. For example, we may observe student achievement (the output) as well as measures of teacher and student quality (the match characteristics). Section 4 reviews methods appropriate for the case where only match characteristics are observed.

Section 3.1 begins with a discussion of assignments. An assignment is a feasible allocation of 'workers' to 'firms'. All empirical matching models, explicitly or implicitly, impose restrictions on the *status quo assignment*. Since the sampling process only reveals the distribution of observed match characteristics, restrictions on the conditional distribution of unobserved match characteristics are needed to identify the match surplus function. This point is explicit in Graham, Imbens and Ridder (2007, 2009a). There the fully nonparametric nature of the match surplus function necessitates rather strong restrictions on the status quo allocation. Less obviously, the structural model of Choo and Siow (2006a, b) also implies strong restrictions on the status quo allocation. There a priori restrictions on the match surplus function induce equilibrium (i.e., status quo) assignments where the conditional distribution of unobserved match characteristics takes a particular form.

Section 3.1 discusses two classes of assignments in detail: those which satisfy an 'as if' double randomization condition and those which satisfy a weaker no matching on unobservables condition. The idea of doubly randomized assignment, introduced in Graham (2008), is straightforward. The no matching on unobservables condition, introduced here, is more subtle. A theme of Section 3.1 is that data distributions in matching problems are conceptually more challenging than those induced by random sampling from a single population. In matching problems, the data distribution reflects constraints imposed by two separate, but interacting, populations. This material is tedious, but foundational for what follows.

Section 3.2 considers average reallocation effects (AREs). AREs measure the change in average outcomes induced by reassigning agents to different partners. The identification of AREs requires a combination of restrictions on the status quo assignment and the match surplus function. If the analyst wishes to leave the match surplus function nonparametric, then strong restrictions on the status quo assignment are required. Alternatively, imposing semiparametric restrictions on the match surplus function allows for identification under weaker assumptions on the status quo. However, when considering AREs a priori restrictions on the match output function should be imposed with considerable caution. Assuming that match output is separable in a specific firm and worker characteristic, for example, will imply that the distribution of match output is invariant across a potentially large set of distinct allocations. Since the evaluation of reallocations is a major motivation for undertaking empirical analysis this is undesirable.

For technical and/or pedagogical reasons Sections 3.1 to 3.2 emphasize settings where agent characteristics are discretely valued. Section 3.3 discusses the case of continuously valued agent characteristics as in Graham, Imbens and Ridder (2009a). Issues of estimation and inference are discussed in Section 3.4.

3.1 The structure of feasible assignments in one-to-one matching problems

Consider a market composed of two large populations. The first population consists of 'firms'. Associated with the i^{th} firm is the observed, discretely-valued, characteristic $W_i \in \{w_1, \ldots, w_K\} = \mathcal{W}$ and an unobserved characteristic, ε_i (which may be vector-valued). In many cases W_i will contain purely qualitative information, in which case we may set $w_k = k$ for $k = 1, \ldots, K$. The general notation, however, allows for the case where W_i has quantitative significance. Note that W_i may itself be a function of multiple underlying characteristics (e.g., it may enumerate age-by-location-by-industry cells). The population frequency of the k^{th} type of firm is $\Pr(W_i = w_k) = p_k$ with $\sum_{k=1}^{K} p_k = 1$. To simplify what follows assume that $p_k > 0$ for all k.

The second population is composed of 'workers'. Associated with the j^{th} worker is the observed, discretely-valued, characteristic $X^j \in \{x_1, \ldots, x_L\} = \mathcal{X}$ and unobserved heterogeneity, ν^j (which may be vector-valued).[17] The population frequency of the l^{th} type of worker is $\Pr(X^j = x_l) = q_l$ with $\sum_{l=1}^{L} q_l = 1$ and $q_l > 0$ for all l.

The education example introduced above helps to fix ideas. Here the population of 'firms' correspond to schools in a given metropolitan area. Schools vary according to their size, demographic composition, location, and type (e.g., magnet, charter, neighborhood, etc.). This vector of school attributes is observed by the econometrician and coded as

[17] To emphasize the presence of two distinct populations I use subscripts to index firms and superscripts to index workers.

$W_i \in \{w_1, \ldots, w_K\}$. Schools are also heterogenous in ways unobserved by the econometrician (e.g., in terms of principal quality), these characteristics are contained in ε_i. The population of 'workers' corresponds to teachers who vary in terms of their observed degree type, years of experience, gender, etc. These attributes are coded as $X^j \in \{x_1, \ldots, x_L\}$. Teachers also vary in unobservable ways, captured by v^j. For example some teachers may prefer to work in certain neighborhoods or in charter schools.

It is convenient to maintain an inclusive definition of firm and worker type. That is to assume that ε_i is independent of W_i and v^j is independent of X^j. When W_i and X^j are exogenous unit characteristics, which in the present context means they are unaffected by the assignment process, we can impose this restriction by normalization (cf., Graham, Imbens and Ridder, 2009b).

Assumption 3.1 (Inclusive Definition of Types)

$$\varepsilon_i \perp W_i, \quad v^j \perp X^j.$$

To see how Assumption 3.1 may be imposed by normalization let ε_i^* denote the unnormalized firm attribute. Defining $\varepsilon_i = F(\varepsilon_i^* | W_i)$ then yields ε_i independent of W_i as required. We interpret ε_i as a firm's ranking in the unobserved attribute *amongst* firms of its same type.[18] If there are two types of workers, those with college degrees and those without, Assumption 3.1 means that we absorb any differences in the distribution of unobserved ability across these two groups into our 'definitions' of types. From the standpoint of a firm, part of the benefit of hiring a random draw from the distribution of college-educated workers is her higher expected 'innate' ability. The econometrician adopts a similar perspective, the justification of which hinges on the class of policies under consideration. In contrast to other policies typically studied in empirical microeconomics, reallocations do not involve changes in the characteristics of agents and, consequently, we are not interested in their causal effects.

The assignment process matches workers with firms. Restrictions on this process drive the identification results reported below. Such restrictions may be directly imposed by the researcher, as in experiments. Alternatively, a particular decentralized assignment mechanism may be posited which induces status quo assignments with properties that facilitate identification. Regardless of whether the observed assignment was imposed by a centralized authority, or represents the equilibrium of a decentralized process, its properties will feature in any identification analysis. Therefore, before turning to the actual mechanics of assignment in Section 4, I discuss the mathematical structure of feasible matchings abstractly.

For simplicity, consider the case where the two populations are equally-sized and all units match. Define the assignment function $m(i) = j$ if the j^{th} worker matches with the i^{th} firm. Let $X^{m(i)} = X_i$ and $v^{m(i)} = v_i$ denote the observed and unobserved

[18] Graham, Imbens and Ridder (2009b) show how to extend this argument to the case where ε_i^* is vector-valued.

characteristics of the i^{th} firm's worker; hence X_i equals the type of worker assigned to firm i (i.e., i indexes both firms and matches). For clarity assume that the unobserved firm and worker characteristics are discretely-valued such that

$$\varepsilon_i \in \mathcal{E} = \{e_1, \ldots, e_F\}, \quad v^j \in \mathcal{V} = \{v_1, \ldots, v_G\},$$

with $f_\varepsilon(e_f)$ and $f_v(v_g)$ respectively denoting the marginal frequency of the events $\varepsilon_i = e_f$ and $v^j = v_g$. Let $f_\varepsilon = (f_\varepsilon(e_1), \ldots, f_\varepsilon(e_F))'$ and $f_v = (f_v(v_1), \ldots, f_v(v_G))'$ be the $F \times 1$ and $G \times 1$ vectors describing the marginal distributions of ε_i and v^j.

The discrete support assumption for the unobservables is not required for the results that follow; it makes their development more transparent. In particular, in highlights the important idea that, in large populations, allocations can be conceptualized as constrained joint distributions.

Under the maintained large matching market assumption, a feasible assignment is equivalent to any joint probability mass function

$$\Pr(W_i = w, X_i = x, \varepsilon_i = e, v_i = v) = h(w, x, e, v), \tag{1}$$

which satisfies the $KF + LG$ feasibility constraints

$$\begin{aligned} &\sum_{g=1}^{G}\sum_{l=1}^{L} h(w_k, x_l, e_f, v_g) = p_k\, f_\varepsilon(e_f), \quad k = 1, \ldots, K, \quad f = 1, \ldots, F \\ &\sum_{f=1}^{F}\sum_{k=1}^{K} h(w_k, x_l, e_f, v_g) = q_l\, f_v(v_g), \quad l = 1, \ldots, L, \quad j = 1, \ldots, G, \end{aligned} \tag{2}$$

and the adding-up condition

$$\sum_{f=1}^{F}\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{l=1}^{L} h(w_k, x_l, e_f, v_g) = 1. \tag{3}$$

Note that the product structure on the right-hand side of the equalities in (2) follows from Assumption 3.1. After eliminating the two redundant constraints using the marginal adding-up restrictions $-\sum_{f=1}^{F}\sum_{k=1}^{K} p_k f_\varepsilon(e_f) = \sum_{g=1}^{G}\sum_{l=1}^{L} q_l f_v(v_g) = 1$ − feasibility places a total of $KF + LG - 1$ constraints on the allocation probability mass function (1). Consequently, any feasible allocation may be defined in terms of a total of $(KF - 1) \times (LG - 1)$ probability masses. Let H be the $KF \times LG$ matrix of probabilities such that $h(w_k, x_l, e_f, v_g)$ is contained in the $(F(k - 1) + f, G(l - 1) + g)^{th}$ entry. An assignment is completely characterized by the form of H.[19]

Imposing additional structure on the assignment generates more parsimonious parameterizations for H. Such parsimony can facilitate identification. Here I want to

[19] In the large population context, where many agents may be identical in both observed and unobserved attributes, it is possible for distinct allocations to have identical values for H^a. For example, two identical pairs can switch partners without changing H^a. This lack of uniqueness has no substantive implications for empirical research.

Table 1 A re-parameterized feasible joint density for observed match characteristics

W\X	x_1	...	x_{L-1}	x_L	$f_W(w)$
w_1	r_{11}	...	$r_{1\ L-1}$	$p_1 - \sum_{l=1}^{L-1} r_{1l}$	p_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
w_{K-1}	r_{K-11}		$r_{K-1\ L-1}$	$p_{K-1} - \sum_{l=1}^{L-1} r_{K-1l}$	p_{K-1}
w_K	$q_1 - \sum_{k=1}^{K-1} r_{k1}$	...	$q_{L-1} - \sum_{k=1}^{K-1} r_{kL-1}$	$1 - \sum_{k=1}^{K-1} p_k - \sum_{l=1}^{L-1} q_l + \sum_{k=1}^{K-1}\sum_{l=1}^{L-1} r_{kl}$	p_K
$f_X(x)$	q_1	...	q_{L-1}	q_L	

emphasize two classes of feasible assignments. The first is the class of assignments satisfying an 'as if' double randomization condition (cf., Graham, 2008). I also consider a natural extension of double randomization, which involves additional conditioning. A second class of assignments, which imposes fewer restrictions on H, satisfies a no matching on unobservables condition.

3.1.1 Double randomization

Under *double randomization* we have, letting $\Pr(W_i = w, X_i = x) = r_{WX}(w, x)$,

$$h(w, x, e, v) = r_{WX}(w, x) f_\varepsilon(e) f_v(v). \tag{4}$$

Allocations of the form (4) may be implemented as follows. First, choose $r_{WX}(w, x)$, a feasible joint allocation density for (W_i, X_i). Second, form type-k-firm-to-type-l-worker matches (henceforth k-to-l matches) by drawing a worker at random from the subpopulation of workers with $X^j = x_l$ and assigning her to a firm drawn at random from the subpopulation of firms with $W_i = w_k$. Under double randomization the joint distribution of unobserved agent attributes is the same across all types of matches, as defined in terms of observed attributes.

After eliminating redundant terms using the marginal constraints

$$\sum_{l=1}^{L} r_{WX}(w_k, x_l) = p_k,\ k = 1, \ldots, K$$

$$\sum_{k=1}^{K} r_{WX}(w_k, x_l) = q_l,\ l = 1, \ldots, L,$$

and the adding-up condition, the allocation density for the observed match characteristics can be represented in terms of $(K-1)\ (L-1)$ parameters. Let R_{WX} be the KL matrix with

$r_{WX}(w_k, x_l) = r_{kl}$ as its $(k, l)^{th}$ entry. Assume that the marginal constraints are used to reparameterize the K^{th} row and L^{th} column in terms of the other $(K-1)\ (L-1)$ joint probability masses and the $K+L$ marginal probability masses. This is illustrated in Table 1.

Using the above notation we can express an allocation which satisfies double randomization as a $KF \times LG$ matrix of the form

$$H = R_{WX} \otimes f_\varepsilon f_\nu{}' \tag{5}$$

subject to the restrictions that

$$R_{WX}\iota_L = p,\ \ R'_{WX}\iota_K = q,\ \ \iota'_K R_{WX}\iota_L = 1, \tag{6}$$

and a non-negativity constraint on each element of R_{WX} (ι_L denotes a L column vector of ones).

Note that double randomization does not restrict the degree of assortativeness in observed attributes embodied in R_{WX}: it may differ arbitrarily from the random allocation pq' (subject to the requirement that it is feasible). For example perfect positive assortative matching of W_i on X^j is not inconsistent with double randomization. Double randomization only restricts the matching process *within* k-by-l cells.

To get a feel for the structure of doubly randomized assignments consider the special case where $F = G = K = L = 2$. The marginal constraints on R_{WX} impose the $K + L - 1 = 3$ restrictions

$$r_{12} = p_1 - r_{11}, \quad r_{21} = q_1 - r_{11}, \quad r_{22} = 1 - p_1 - q_1 + r_{11},$$

Under a double randomized assignment we therefore have,

$$H(r_{11}) = \begin{pmatrix} r_{11} & p_1 - r_{11} \\ q_1 - r_{11} & 1 - p_1 - q_1 + r_{11} \end{pmatrix} \otimes \begin{pmatrix} f_\varepsilon(e_1)f_\nu(v_1) & f_\varepsilon(e_1)f_\nu(v_2) \\ f_\varepsilon(e_2)f_\nu(v_1) & f_\varepsilon(e_2)f_\nu(v_2) \end{pmatrix}. \tag{7}$$

Note that $r_{11} - p_1 q_1$ indexes the 'assortativeness' of the allocation or the frequency with which type $W = w_1$ firms are matched with type $X = x_1$ workers relative to the random matching benchmark.[20] Inspection of (7) illustrates that doubly randomized allocations allow for arbitrary amounts of sorting on observables, but no sorting or matching on unobservables.

Double randomization allows us to express the $KF \times LG$ mass points in H in terms of just the $(K-1)\ (L-1)$ parameters which uniquely define R_{WX}. This reduces the specification of H by $KL(FG-1) - K(F-1) - L(G-1)$ parameters relative to the imposition of feasibility alone. Doubly randomized allocations represent a small subset of the class of feasible allocations.

[20] Feasibility also requires that r_{11} satisfy the inequality $p_1 - \min\{p_1, q_1\} \le r_{11} \le \min\{p_1, q_1\}$ (cf., Graham, Imbens and Ridder, 2007).

Doubly randomized allocations are of interest from a policy perspective. Consider a social planner who is able to centrally assign workers to firms. If the planner is unable to observe ε_i and ν^j, or legally constrained from using such knowledge when making assignments, the class of reallocations available to her is of the form given by (5).

3.1.2 Conditional double randomization

A useful generalization of double randomization is conditional double randomization. The motivation for this extension is two-fold. First, such assignments may characterize some types of non-experimental matching market data. Second, it illustrates how the two-agent aspect of matching models complicates their analysis and generates new and interesting econometric issues.

Let Z_ε and Z_ν be observable proxies or signals for, respectively, ε and ν. I modify Assumption 3.1 so that firm type is independent of ε within subpopulations homogenous in the signal Z_ε. Likewise, worker type is conditionally independent of ν given Z_ν.

Assumption 3.2 (Conditional Inclusive Definition of Types)

$$\varepsilon_i \perp W_i \mid Z_{\varepsilon i}, \qquad \nu^j \perp X^j \mid Z_\nu^j.$$

A conditionally doubly randomized allocation is a joint density for W_i, X_i, ε_i, ν_i, $Z_{\varepsilon i}$ and $Z_{\nu i}$ of the form

$$h(w, x, z_\varepsilon, z_\nu, e, \nu) = r_{WXZ_\varepsilon Z_\nu}(w, x, z_\varepsilon, z_\nu) f_{\varepsilon \mid Z_\varepsilon}(e \mid z_\varepsilon) f_{\nu \mid Z_\nu}(\nu \mid z_\nu). \tag{8}$$

A member of this class of allocations may be formed as follows. First, the planner chooses a feasible joint distribution for W_i, X_i, $Z_{\varepsilon i}$ and $Z_{\nu i}$. Second, within each $W_i = w$ and $Z_{\varepsilon i} = z_\varepsilon$ by $X_i = x$ and $Z_{\nu i} = z_\nu$ cell, the required matches are formed by drawing workers at random from the subpopulation of workers with $X^j = x$ and $Z_{\nu i} = z_\nu$ and assigning them to firms drawn at random from the subpopulation of firms with $W_i = w$ and $Z_{\varepsilon i} = z_\varepsilon$.

This class of assignments allows for dependence between ε_i and ν^j. This is because dependence between $Z_{\varepsilon i}$ and $Z_{\nu i}$ induces dependence between ε_i and ν_i since, for $r_{Z_\varepsilon Z_\nu}(z_\varepsilon, z_\nu) = \sum_{k=1}^K \sum_{l=1}^L r_{WXZ_\varepsilon Z_\nu}(w_k, x_l, z_\varepsilon, z_\nu)$,

$$f_{\varepsilon,\nu}(e, \nu) = \sum_{d=1}^{D} \sum_{e=1}^{E} f_{\varepsilon \mid Z_\varepsilon}(e \mid z_{\varepsilon d}) f_{\nu \mid Z_\nu}(\nu \mid z_{\nu e})\, r_{Z_\varepsilon Z_\nu}(z_{\varepsilon d}, z_{\nu e}),$$

which does not equal $f_\varepsilon(e)\, f_\nu(\nu)$ unless $r_{Z_\varepsilon Z_\nu}(z_\varepsilon, z_\nu)$ coincides with the product of its two marginals. In contrast with pure double randomization, higher quality firms may match with higher quality workers.

The educational example introduced above suggests why conditional double randomization may be useful in practice. Recall that W_i and X^j denote observed characteristics of the school and teacher. Let ε_i and ν^j denote unobserved characteristics of the school and teacher, say average student ability and teacher quality, that are important determinants of student achievement (the outcome of interest to the econometrician). Let $Z_{\varepsilon i}$ and Z_v^j be proxies for these characteristics, such as the average student's intake test score and a teacher's licensure test score. Under a conditionally doubly randomized assignment of teachers to schools the joint distribution of W_i, X_i, $Z_{\varepsilon i}$ and Z_{vi} is restricted only by feasibility. However, within the subpopulation of *matches* homogenous in these observables, student ability is independent of teacher quality. Such an assumption can be plausible when $Z_{\varepsilon i}$ and Z_v^j closely approximate agents' information sets for ε_i and v^j; but is more difficult, though not impossible, to justify otherwise.[21]

3.1.3 No matching on unobservables

An alternative approach to relaxing the requirement of double randomization is to consider allocations that satisfy a 'no matching on unobservables' restriction. Unlike conditional double randomization, this extension does not require the introduction of proxy variables.

Let

$$s_{\varepsilon|X}(e|x) = \Pr(\varepsilon_i = e|X_i = x), \quad s_{v|w}(v|w) = \Pr(v_i = v|W_i = w),$$

denote the conditional densities of the unobserved firm and worker characteristics, say productivity and ability, given, respectively, observed worker and firm type. Allocations with the no matching on unobservables property have joint densities of the form

$$h(w,x,e,v) = r_{WX}(w,x)\, s_{\varepsilon|X}(e|x)\, s_{v|W}(v|w), \tag{9}$$

subject to the $F+G+K+L$ marginal constraints

$$\begin{aligned} &\textstyle\sum_{l=1}^{L} s_{\varepsilon|X}(e_f|x_l)q_l = f_\varepsilon(e_f), && \textstyle\sum_{k=1}^{K} s_{v|W}(v_g|w_k)p_k = f_v(v_g) \\ &\textstyle\sum_{l=1}^{L} r_{WX}(w_k,x_l) = p_k, && \textstyle\sum_{k=1}^{K} r_{WX}(w_k,x_l) = q_l \end{aligned} \tag{10}$$

and the $K+L+1$ adding up conditions

$$\begin{aligned} &\sum_{f=1}^{F} s_{\varepsilon|X}(e_f|x_l) = 1, \quad l = 1,\ldots,L \\ &\sum_{g=1}^{G} s_{v|W}(v_g|w_k) = 1, \quad k = 1,\ldots,K \\ &\sum_{f=1}^{F}\sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{l=1}^{L} r_{WX}(w_k,x_l)s_{\varepsilon|X}(e_f|x_l)s_{v|W}(v_g|w_k) = 1. \end{aligned} \tag{11}$$

[21] Rigorously justifying this claim is not attempted here. Heckman and Vytlacil (2007a, b) provide an extensive discussion of the role of informational assumptions in the econometric analysis of single agent models. Many of their insights should apply here as well.

Restrictions (10) and (11) can be compared with their double randomization counterparts (2) and (3).

Inspection of (9) indicates that assignments which satisfy the no matching on unobservables condition are special. Like doubly randomized assignments, they impose conditional independence of ε_i and ν_i given (W_i, X_i). However, unlike doubly randomized assignments, they do allow for a limited type of 'input endogeneity'. Consider the conditional distribution of ε_i given X_i (the analysis of the conditional distribution of ν_i is entirely parallel). The distribution of ε_i is allowed to arbitrarily vary with X_i. If we equate ε_i with firm productivity, then this allows productive firms to be more frequently matched with certain types of workers (defined in terms of their values of X^j). This is entirely analogous to a conventional production function problem where a firm's observed input choice may co-vary with its unobserved productivity (e.g., Griliches and Mairesse, 1998). This type of endogenous input choice is ruled out under doubly randomized assignments.

Note, however, that (9) requires that ε_i's conditional distribution be constant in W_i conditional on X_i. This implies that the relationship between a firm's (unobserved) productivity and its (observed) input level, is independent of its (observed) type. If the observed and unobserved firm characteristics, respectively W_i and ε_i, enter the production function non-separably, then it will generally be the case that firms with specific configurations of W_i *and* ε_i, as opposed to just ε_i alone, will differentially demand certain types of workers. Phrased in this way it is clear that the no matching on unobservables restriction is quite strong. Nevertheless it is weaker the requiring an allocation to satisfy the double randomization condition.

Consider once again the education example. Under a no matching on unobservables allocation we *do allow* schools with *unobserved* high quality principals to differentially hire teachers with certain types of *observed* qualifications. We *do not allow* them to differentially hire teachers with certain types of *unobserved* qualifications. Likewise we allow teachers with certain unobserved attributes to differentially work at schools with certain types of observed characteristics, but do not allow them to differentially work at schools with certain types of unobserved characteristics. Assume that both principal and teacher years of experience are observed but 'qualities' are not. No matching on unobservables implies that schools with high quality principals may hire more experienced teachers. Likewise high quality teachers may work at schools with more experienced principals. This, in turn, implies that the conditional distributions of principal and teacher quality will vary *across* subpopulations of matches defined in terms of observed principal and teacher experience. However, *within* such subpopulations, there is no matching on unobserved quality.

Let $s_{\varepsilon l} = (s_{\varepsilon|X}(e_1|x_l), \ldots, s_{\varepsilon|X}(e_F|x_l))'$ and $s_{\nu k} = (s_{\nu|W}(\nu_1|w_k), \ldots, s_{\nu|W}(\nu_G|w_k))'$. This gives $S_\varepsilon = (s_{\varepsilon 1}, \ldots, s_{\varepsilon L})$ and $S_\nu = (s_{\nu 1}, \ldots, s_{\nu K})$ equal to, respectively, the $F \times L$ and $G \times K$ matrices of probability masses which define the conditional distributions of ε_i given X_i and ν_i given W_i. Using this notation the $(k, l)^{th}$ block of H, for an allocation satisfying the no matching on unobservables assumption, takes the form $r_{kl}s_{\varepsilon l}s'_{\nu k}$ (which is of dimension $F \times G$). Therefore

$$H = (R_{WX} \otimes \iota_F \iota'_G) * \bar{S}_\varepsilon * \bar{S}'_\nu, \tag{12}$$

where the notation $\bar{S}_\varepsilon * \bar{S}'_\nu$ denotes the Hadamard, or entrywise product, of the matrices

$$\bar{S}_\varepsilon = \iota_K \otimes (S_\varepsilon \otimes \iota'_G), \qquad \bar{S}_\nu = \iota_L \otimes (S_\nu \otimes \iota'_F).$$

Restrictions (10) and (11) may be expressed in matrix form as

$$\begin{aligned} S_{\varepsilon q} = f_\varepsilon, \quad S_{\nu p} = f_\nu, \qquad R_{WX}\iota_L = p, \quad R'_{WX}\iota_K = q, \\ S'_\varepsilon \iota_F = \iota_L, \qquad S'_\nu \iota_G = \iota_K, \qquad \iota'_K R_{WX} \iota_L = 1. \end{aligned} \tag{13}$$

These conditions imply that S_ε admits a $(F-1) \times (L-1)$ parameterization and S_ν a $(G-1) \times (K-1)$ parameterization. This, along with a $(K-1) \times (L-1)$ parameterization of R_{WX} means that no matching on unobservables imposes

$$(KF-1) \times (LG-1) - (F-1) \times (L-1) - (G-1) \times (K-1) - (K-1) \times (L-1),$$

additional conditions on H beyond those required for feasibility. Relative to the double randomization condition, the no matching on unobservables condition adds a total of $(F-1) \times (L-1) + (G-1) \times (K-1)$ degrees of freedom to H.

Consider once again the special case where $F = G = K = L = 2$. Implementing a no matching on unobservables allocation requires choosing feasible values for S_ε, S_ν and R_{WX}. The latter choice was described above. The choice of S_ε involves selecting a 2×2 matrix with columns summing to one and rows summing to f_ε. This imposes 3 non-redundant constraints on S_ε (since $f_\varepsilon(e_1) + f_\varepsilon(e_2) = 1$). Let $s_{\varepsilon 11} = s_{\varepsilon|X}(e_1|x_1)$, the conditional frequency of type $\varepsilon_i = e_1$ firms among those matched to type $X_i = x_1$ workers, and $s_{\nu 11} = s_{\nu|W}(\nu_1|w_1)$, the conditional frequency of type $\nu^j = \nu_1$ workers among those matched to type $W^j = w_1$ firms. We have

$$S_\varepsilon(s_{\varepsilon 11}) = \begin{pmatrix} s_{\varepsilon 11} & \dfrac{f_\varepsilon(e_1) - s_{\varepsilon 11} q_1}{1 - q_1} \\ 1 - s_{\varepsilon 11} & 1 - \dfrac{f_\varepsilon(e_1) - s_{\varepsilon 11} q_1}{1 - q_1} \end{pmatrix}, \quad S_\nu(s_{\nu 11}) = \begin{pmatrix} s_{\nu 11} & \dfrac{f_\nu(\nu_1) - s_{\nu 11} p_1}{1 - p_1} \\ 1 - s_{\nu 11} & 1 - \dfrac{f_\nu(\nu_1) - s_{\nu 11} p_1}{1 - p_1} \end{pmatrix}.$$

To gauge the effects of $s_{\varepsilon 11}$ and $s_{\nu 11}$ on the properties of the assignment note that the average difference in productivity between firms who match with type 2 ('high') versus type 1 ('low') workers is

$$\mathbb{E}[\varepsilon_i \mid X_i = x_2] - \mathbb{E}[\varepsilon_i \mid X_i = x_1] = -\left(\frac{f_\varepsilon(e_1) - s_{\varepsilon 11}}{1 - q_1}\right)(e_2 - e_1) \gtreqless 0 \text{ as } s_{\varepsilon 11} \gtreqless f_\varepsilon(e_1).$$

If $s_{\varepsilon\ 11} > f_\varepsilon(e_1)$, then 'low' productivity firms (i.e., $\varepsilon_i = e_1$) more frequently choose 'low' type workers (i.e., $X_i = x_1$). In such an allocation *observed* worker type predicts *unobserved* firm productivity. Knowing a firm's type, however, does not help to predict *its* productivity.

Similarly the average difference in ability between workers who match with type 2 ('high') versus type 1 ('low') firms is

$$\mathbb{E}[v_i \mid W_i = w_2] - \mathbb{E}[v_i \mid W_i = w_1] = -\left(\frac{f_v(v_1) - s_{v11}}{1 - p_1}\right)(v_2 - v_1) \gtreqless 0 \text{ as } s_{v11} \gtreqless f_v(v_1),$$

so that if $s_{v11} > f_v(v_1)$, then 'low' ability workers (i.e., $v^j = v_1$) more frequently choose 'low' type firms (i.e., $W^j = w_1$). In such an allocation firm type predicts unobserved worker ability. Knowing a worker's type, however, does not help to predict *its* ability.

If $s_{\varepsilon 11} = f_\varepsilon(e_1)$ and $s_{v11} = f_v(v_1)$ we recover the doubly randomized assignment. Finally, as with double randomization, the no matching on unobservables requirement does not restrict the degree of assortativeness on observed firm and worker attributes.

3.2 Average reallocation effects (AREs)

A major motivation for the empirical analysis of matching markets is to predict the distribution of outcomes that would prevail under alternative feasible assignments of workers to firms. Here, following Graham, Imbens and Ridder (2007, 2009a), I consider identifying the change in average outcomes induced by a different allocations (i.e., *average reallocation effects* (AREs)). In some settings implementing a particular assignment, while in principle feasible, might be difficult in practice. For example, the private incentives for re-matching may be strong under certain assignments, suggesting that they would 'unravel' if actually implemented. I ignore these issues in what immediately follows. There are at least two motivations for doing so. First, an exploration of the outcome effects of alternative assignments is a prelude to more complete policy formulation. If the social benefits from a particular assignment are deemed large relative to the status quo, then further thought can be given to developing a decentralized mechanism which produces the desired assignment. In contrast, the mechanism design question is less interesting if the distribution of outcomes is largely invariant across different assignments. Second, in some institutional settings agents' exercise little control over whom they match with. In these settings the estimands introduced below are directly relevant.

3.2.1 Definition of target estimands

In order to identify AREs, knowledge of the match output function, or certain features of it, is required. A general form for the match output associated with the pairing of firm i with worker j is

$$Y_{(i,j)} = k(W_i, X^j, \varepsilon_i, v^j). \tag{14}$$

Equation (14) is a fully nonparametric specification of match output, being non-separable in the unobserved attributes of the matched firm and worker. Importantly, it allows for arbitrary complementarity or substitutability between observed and unobserved firm and worker attributes. No dimensionality or monotonicity assumptions on ε_i or ν^j are imposed.

Since ε_i and ν^j enter nonseparably, the identification of (14) may be too ambitious a goal. Instead, we might consider conditions under which we can identify the *average match output function* (AMF):

$$\begin{aligned} \delta(w,x) &= \iint k(w,x,e,v)\, f_\varepsilon(e)\, f_\nu(v)\, \mathrm{d}e\mathrm{d}v \\ &= \mathbb{E}_{\varepsilon_i}[\mathbb{E}_{\nu^j}[k(w,x,\varepsilon_i,\nu^j)]]. \end{aligned} \tag{15}$$

The second line of (15) establishes notation for an expectation taken with respect to the product of two marginals.

The AMF corresponds to the expected match output associated with a pairing of a type $W=w$ firm with a type $X=x$ worker *when both the firm and worker are independent and random draws from their respective subpopulations*.

The AMF is an average with respect to the product of two unobserved heterogeneity distributions, one for each side of the matching market. The analog to (15) in a single agent model would be the average structural function (ASF) of Blundell and Powell (2003). A direct application of the ASF definition to (14) would involve replacing $f_\varepsilon(e)$ $f_\nu(v)$ in (15) with $f_{\varepsilon,\nu}(e,v)$. Such a definition would not correspond to a structural object as $f_{\varepsilon,\nu}(e,v)$ is not invariant across feasible assignments; it is a by-product of such assignments.

Now consider the effect of a reassignment of workers-to-firms on the distribution of match output. As a benchmark consider the case where the planner is unable to use information on ε_i and ν^j when making her assignments; that is she is constrained to choose an allocation from among the set of doubly randomized allocations. Let

$$\gamma(R^{\mathrm{a}}_{WX}) = \iint \delta(w,x)\, r^{\mathrm{a}}_{WX}(w,x)\, \mathrm{d}w\mathrm{d}x$$

denote average match output under the alternative doubly randomized allocation R^a_{WX}; for W_i and X^j discretely-valued,

$$\begin{aligned} \gamma(R^{\mathrm{a}}_{WX}) = & \sum_{k=1}^{K-1}\sum_{l=1}^{L-1} r^{\mathrm{a}}_{kl} \times \{\delta(w_K,x_L) - \delta(w_K,x_l) - [\delta(w_k,x_L) - \delta(w_k,x_l)]\} \\ & + \sum_{l=1}^{L-1} ql\, \{\delta(w_K,x_l) - \delta(w_K,x_L)\} + \sum_{k=1}^{K-1} p_K\, \{\delta(w_k,x_L) - \delta(w_K,x_L)\} \\ & + \delta(w_K,x_L) \end{aligned}$$

The equality follows from using the $K+L-1$ non-redundant marginal constraints to re-parameterize R^{a}_{WX} (see Table 1). Let R^{sq}_{WX} denote the $K \times L$ matrix of mass probabilities which define the joint distribution of W_i and X_i under the status quo. The effect of choosing allocation R^{a}_{WX} on average match output relative to R^{sq}_{WX}, the *average reallocation effect* (ARE) of Graham, Imbens, and Ridder (2007, 2009a), is then

$$\beta(R^{\mathrm{a}}_{WX}) = \gamma(R^{\mathrm{a}}_{WX}) - \gamma(R^{\mathrm{sq}}_{WX}) = \sum_{k=1}^{K-1}\sum_{l=1}^{L-1}\left\{r^{\mathrm{a}}_{kl} - r^{\mathrm{sq}}_{kl}\right\} \times \phi_{KLkl}, \tag{16}$$

where

$$\phi_{KLkl} = \delta(w_K, x_L) - \delta(w_K, x_l) - [\delta(w_k, x_L) - \delta(w_k, x_l)], \tag{17}$$

is a measure of *average local complementarity* (ALC) between W and X. To see this observe that (17) measures the expected difference between the incremental return associated with hiring a type L versus l worker across type K versus k firms. Consider two randomly sampled firms and workers; if (17) is positive the sum of *expected* output across an assortative (K,L) and (k,l) assignment will exceed that across an anti-assortative (K,l) and (k,L) assignment.

It is important to understand that ϕ_{KLkl} is a local measure of complementarity. Consider $K > m > k$ and $L > n > l$ we may have $\phi_{KLkl} > 0$ and $\phi_{mnkl} < 0$; there is no presumption that $\delta(w,x)$ exhibits increasing differences or is supermodular (e.g., Topkis, 1998). This a priori flexibility vis-a-vis $\delta(w,x)$ is important for making the study of reallocations empirically interesting.

The representation of the average reallocation effect in terms of $(K-1)$ $(L-1)$ ALC parameters facilitates its identification since, as we shall see below, (17) may be identified even if $\delta(w,x)$ is not.

Some focal assignments Among the class of feasible allocations positive and negative assortative matchings have attracted substantial attention (e.g., Becker, 1973; Legros and Newman, 2002). These allocations maximize output when $\delta(w,x)$ is, respectively, super-modular and submodular. The definition of these allocations requires that $K=L$ with $w_{k+1} > w_k$ and $x_{k+1} > x_k$ for all $k = 1, \ldots, K-1$. In a *positive assortative matching* (pam) the rank order of W_i and X_i, to the extent that feasibility allows, coincide. The highest type firms are assigned to the highest type workers. In a *negative assortative matching* (nam) the opposite pattern occurs: the highest type firms are assigned to the lowest type workers. Mathematically these two allocations concentrate the maximal feasible amount of probability mass on, respectively, the primary and secondary diagonals of R^{a}_{WX}.

Recall that when $K=L=2$, the set of feasible allocations is indexed by r_{11}. In this case the positive assortative matching chooses $r^{\mathrm{pam}}_{11} = \min(p_1,\ q_1)$. This choice induces the maximal amount of feasible assortativeness yielding an ARE of

$$\beta^{\text{pam}} = \min\{p_1 - r_{11}^{\text{sq}}, q_1 - r_{11}^{\text{sq}}\}\ \phi_{2211}.$$

The negative assortative matching chooses $r_{11}^{\text{nam}} = \max\{0, p_1 + q_1 - 1\}$, which induces the maximal feasible amount of mixing. This yields an ARE of

$$\beta^{\text{nam}} = -\max\{r_{11}^{\text{sq}}, 1 - p_1 - q_1 + r_{11}^{\text{sq}}\}\phi_{2211}.$$

When $K = L = 3$ the set of feasible allocations is indexed by r_{11}, r_{12}, r_{21} and r_{22}. The positive assortative matching chooses the values

$$\begin{aligned}
r_{11}^{\text{pam}} &= \min(p_1, q_1) \\
r_{12}^{\text{pam}} &= \max(0, \min\{q_2, p_1 - r_{11}^{\text{pam}}\}) \\
r_{21}^{\text{pam}} &= \max(0, \min\{p_2, q_1 - r_{11}^{\text{pam}}\}) \\
r_{22}^{\text{pam}} &= \max(p_2 - r_{21}^{\text{pam}}, q_2 - r_{12}^{\text{pam}}),
\end{aligned}$$

with the negative assortative matching constructed analogously. For the general $K = L$ case a recursive algorithm can be used to construct the positive and negative assortative matchings.

An interesting feature of both the β^{pam} and β^{nam} estimands is their dependence on marginal distributions of firm and worker type as well as on the production technology. This dependence affects the sampling properties of their analog estimates (see Section 4.5). It also highlights, yet again, the interplay between the resource constraint, here the firm and worker type distributions, and the production technology in the analysis of matching problems.

A third focal allocation is the *random matching*, the ARE of which is given by

$$\beta^{\text{rm}} = \beta(pq') = \sum_{k=1}^{K-1}\sum_{l=1}^{L-1}\{p_k q_l - r_{kl}^{\text{sq}}\} \times \phi_{KLkl}.$$

This captures the outcome effects of switching to an assignment in which agents are randomly paired to one another.

In many situations it may be of interest to assess the maximum gain in average outcomes available via reallocation or the *maximum average reallocation effect* (MRE). One attractive feature of the assumption that W_i and X^j are discretely-valued is that this is a tractable estimand. The MRE, analyzed by Graham, Imbens and Ridder (2007) and Bhattacharya (2009), is

$$\beta^{\text{mre}} = \max_{R_{WX}^{\text{a}} \in \mathcal{R}} \sum_{k=1}^{K-1}\sum_{l=1}^{L-1}\{r_{kl}^{\text{a}} - r_{kl}^{\text{sq}}\} \times \phi_{KLkl}, \tag{18}$$

where $\mathcal{R}$ denotes the set of feasible allocations:

$$\mathcal{R} = \left\{R_{WX}^{\text{a}} : R_{WX}^{\text{a}}\iota_L = p, \quad R_{WX}^{\text{a}\prime}\iota_K = q, \quad \iota_K' R_{WX}^{\text{a}}\iota_L = 1, \quad r_{kl}^{\text{a}} \geq 0 \quad \text{for all } k, l\right\}.$$

This is a linear assignment program, a special case of the transportation problem (e.g., Luenberger, 2005, Chapter 5). While the maximizing allocation need not be unique, the value function will be. An allocation can be thought of as a $(K-1)\times(L-1)$ dimensional polyhedron, the extreme points of which correspond to extreme allocations. By the Fundamental Theorem of Linear Programming (Luenberger, 2005; Chapter 2.4) an optimal allocation will be one of the extreme allocations. The maximum reallocation effect, β^{mre}, can be used to assess the efficiency of the status quo (Bhattacharya, 2009).

When $K = L = 2$ the two extreme allocations respectively correspond to the positive and negative assortative matchings. Since the optimal allocation must be one of these two assignments we get the elegant solution (Graham, Imbens and Ridder, 2007)

$$r_{11}^{\mathrm{mre}} = \min(p_1, q_1)\cdot\mathbf{1}\,(\phi_{2211} \geq 0) + \max\{0, p_1 + p_2 - 1\}\cdot\mathbf{1}\,(\phi_{2211} < 0),$$

with

$$\beta^{\mathrm{mre}} = (r_{11}^{\mathrm{mre}} - r_{11}^{\mathrm{sq}}) \times \phi_{2211}.$$

When $K, L > 2$ the form of β^{mre} is more complicated, but is easily solved for numerically using linear programming methods. The solution in the $K = L = 3$ case is discussed in Graham, Imbens and Ridder (2007).

3.2.2 Identification under double randomization

First, consider unrestricted match output functions of the form given in (14). If the status quo allocation satisfies the doubly randomized restriction the AMF is identified by

$$\mathbb{E}[Y_i \mid W_i = w_k, X_i = x_l] = \delta(w, x).$$

Identification of ϕ_{KLkl} and the average reallocation effect $\beta(R^{\mathrm{a}})$ follows directly under additional support conditions.

Proposition 3.1 (Identification Under Double Randomization) *If*
(i) $Y_{(i,j)} = k(W_i, X^j, \varepsilon_i, v^j)$,
(ii) *the status quo assignment is of the form given by (4), and*
(iii) $r_{kl}^{\mathrm{a}} > 0$ *only if* $r_{kl}^{\mathrm{sq}} > 0$, *then*

$$\mathbb{E}[Y_i \mid W_i = w_k, X_i = x_l] = \delta(w, x)$$

and $\beta(R_{WX}^{\mathrm{a}})$ *is identified. If* $r_{kl}^{\mathrm{sq}} > 0$ *for all* $k = 1, \ldots, K$ *and* $l = 1, \ldots, L$, *then* β^{mre} *is also identified.*

Condition (iii) requires the status quo and alternative allocations share a common support. Proposition 3.1 is implicit in Graham, Imbens and Ridder (2007).

Identification of $\delta(w, x)$ under conditional double randomization is more involved. Consider the proxy variable regression

$$\begin{aligned} g(w,x,z_\varepsilon,z_v) &= \mathbb{E}[Y_i \mid W_i = w, X_i = x, Z_\varepsilon = z_\varepsilon, Z_v = z_v] \\ &= \iint k(w,x,e,v) f_{\varepsilon \mid Z_\varepsilon}(e \mid z_\varepsilon) f_{v \mid Z_v}(v \mid z_v)\, \mathrm{d}m(e)\mathrm{d}m(v), \end{aligned}$$

with the second equality following from the conditional double randomization assumption. This suggests recovering $\delta(w,x)$ by

$$\delta(w,x) = \mathbb{E}_{Z_\varepsilon}[\mathbb{E}_{Z_v}[g(w,x,Z_\varepsilon,Z_v)]]. \tag{19}$$

Equation (19) is similar to the partial mean estimand introduced by Newey (1994) and widely-used in recent work on the nonparametric identification of single agent models (e.g., Blundell and Powell, 2003; Imbens, 2007a). It is distinctive in that it involves two marginal averages, as opposed to one. To understand the importance of sequentially averaging over the two marginal distributions, note that the conventional partial mean

$$\mathbb{E}_{Z_\varepsilon,Z_v}[g(w,x,Z_\varepsilon,Z_v)],$$

does not equal the AMF. This is because any unobserved dependence between ε_i and v^j will be mirrored in the observed dependence between Z_ε and Z_v. Therefore averaging over the joint distribution of the latter will not recover the AMF. The idea that unobserved heterogeneity may be 'averaged out' in a two-agent model by averaging over the product of two proxy variable marginal distributions, one for each agent, appears to be new (cf., Graham, Imbens, Ridder, 2009b).

This idea is summarized by Proposition 3.2.

Proposition 3.2 (Identification Under Conditional Double Randomization) *If*
(i) $Y_{(i,j)} = k(W_i, X^j, \varepsilon_i, v^j)$,
(ii) *the status quo assignment is of the form given by (8),*
(iii) *the support of W_i, X_i given $Z_\varepsilon = z_\varepsilon$, $Z_v = z_v$, does not depend on (z_ε, z_v),*
(iv) *the joint support of Z_ε and Z_v coincides with the product of its two marginals' supports,*
(iv) $r_{kl}^{\mathrm{a}} > 0$ *only if* $r_{kl}^{\mathrm{sq}} > 0$, *then,*

$$\delta(w,x) = \mathbb{E}_{Z_\varepsilon}[\mathbb{E}_{Z_v}[g(w,x,Z_\varepsilon,Z_v)]],$$

for $g(w,x,z_\varepsilon,z_v) = \mathbb{E}[Y_i \mid W_i = w, X_i = x, Z_\varepsilon = z_\varepsilon, Z_v = z_v]$ and $\beta(R_{WX}^{\mathrm{a}})$ *is identified. If $r_{kl}^{\mathrm{sq}} > 0$ for all $k = 1, \ldots, K$ and $l = 1, \ldots, L$, then β^{mre} is also identified.*

Condition (iii) ensures that for any configuration of proxies all types of matches, defined in terms of their (X_i, W_i) value, are observed. This is similar to the 'overlap' assumption in the program evaluation literature. Condition (iv) is specific to the matching context. It requires that their not be too much dependence between $Z_{\varepsilon i}$ and Z_{vi}. Averaging over the the two marginal distributions in (19) eliminates any bias due to observed matching on $Z_{\varepsilon i}$ and Z_{vi} (which proxies for unobserved matching on ε and v). If there is two much dependence between $Z_{\varepsilon i}$ and Z_{vi} in the status quo this double averaging cannot be performed. Conditions (iii) and (iv) are strong conditions.

In situations where point identification fails, $\beta(R^{\mathrm{a}}_{WX})$ and β^{mre} may still be set identified. This possibility is not explored here.

3.2.3 Identification under no matching on unobservables

Propositions 3.1 and 3.2 leave the match output function unrestricted, but make strong a priori assumptions about the form of the status quo assignment. If we restrict the match output function positive identification results are possible without assuming double randomization. Consider the following restricted match outcome function

$$k(W_i, X^j, \varepsilon_i, v^j) = \delta(W_i, X^j) + \lambda(\varepsilon_i, X^j) + \rho(W_i, v^j), \tag{20}$$

where

$$\mathbb{E}[\lambda(\varepsilon_i, x)] = \mathbb{E}[\rho(w, v^j)] = 0. \tag{21}$$

These normalizations imply that that $\delta(w,x)$ is the AMF as defined in (15) above. Restriction (20) will also feature in the analysis of equilibrium matching data when agent characteristics alone are observed (Section 4 below).

Equation (20) is restrictive. Holding unobserved firm and worker complementarity fixed we have, for $w' > w$ and $x' > x$,

$$\begin{aligned} &k(w', x', \varepsilon, v) - k(w, x, \varepsilon, v) - [k(w, x', \varepsilon, v) - k(w, x, \varepsilon, v)] \\ &\quad = \delta(w', x') - \delta(w, x) - [\delta(w, x') - \delta(w, x)]. \end{aligned}$$

The degree of complementarity between *observed* firm and worker attributes does not vary with *unobserved* firm and worker attributes.

Similarly, holding observed match characteristics fixed, for $\varepsilon' > \varepsilon$ and $v' > v$,

$$k(w, x, \varepsilon', v') - k(w, x, \varepsilon', v) - [k(w, x, \varepsilon, v') - k(w, x, \varepsilon, v)] = 0,$$

which rules out complementarity in unobserved agent attributes.

The match surplus function does allow for complementarity between ε_i and X^j as well as W_i and v^j. Specifically

$$\begin{aligned} &k(w, x', \varepsilon', v) - k(w, x, \varepsilon', v) - [k(w, x', \varepsilon, v) - k(w, x, \varepsilon, v)] \\ &\quad = \lambda(\varepsilon', x') - \lambda(\varepsilon', x) - [\lambda(\varepsilon, x') - \lambda(\varepsilon, x)], \end{aligned}$$

and

$$\begin{aligned} &k(w', x, \varepsilon, v') - k(w', x, \varepsilon, v) - [k(w, x, \varepsilon, v') - k(w, x, \varepsilon, v)] \\ &\quad = \rho(w', v') - \rho(w', v) - [\rho(w, v') - \rho(w, v)], \end{aligned}$$

may be non-zero.

The form of the surplus function drives matching incentives. Restriction (20) has strong implications for these incentives. While it does allow for complementarity between observed attributes, generating incentives for matching on observables, it does *not* allow for complementarity between unobserved attributes. This eliminates any incentive to sort on unobservables within k-by-l assignment cells. It also restricts the types of complementarity allowed between observed and unobserved inputs. While it allows for complementarity between ε_i and X^j, so that firms will seek out different types of workers depending on their value for ε_i, this complementarity is constant in firm type (W_i). This suggests that the intensity of any matching of ε_i on X^j will not vary with W_i. Complementarity between v^j and W_i is similarly restricted.

Under (20) and the no matching on unobservables restriction we have

$$\mathbb{E}[Y_i \mid W_i = w, X_i = x] = \delta(w, x) + \bar{\lambda}(x) + \bar{\rho}(w), \tag{22}$$

for $\bar{\lambda}(x) = \mathbb{E}[\lambda(\varepsilon_i, X_i) \mid X_i = x]$ and $\bar{\rho}(w) = \mathbb{E}[\rho(W_i, v^{m(i)}) \mid W_i = w]$[22].

Unlike the case of double randomization, average output across matches with $W_i = w$ and $X_i = x$ does not coincide with $\delta(w,x)$. The two additional terms, $\bar{\lambda}(x)$ and $\bar{\rho}(w)$, are due to selection bias. Under the no matching on unobservables assumption it is still possible that firms with different values of ε_i differentially match with workers of type $X^j = x$. This matching of ε_i on X^j is captured by $\bar{\lambda}(x)$. Likewise workers with different values of v^j may differentially match with firms of type $W_i = w$. This matching of v^j on W_i is captured by $\bar{\rho}(w)$.

Although allocations which satisfy the no matching on unobservables restriction do not allow for identification of the AMF, they do allow for identification of average local complementarity. To see this note that (22) implies that

$$\begin{aligned}&\mathbb{E}[Y_i \mid W_i = w', X_i = x'] - \mathbb{E}[Y_i \mid W_i = w', X_i = x] \\ &\qquad - (\mathbb{E}[Y_i \mid W_i = w, X_i = x'] - \mathbb{E}[Y_i \mid W_i = w, X_i = x]) \\ &\quad = \delta(w', x') - \delta(w, x) - [\delta(w, x') - \delta(w, x)].\end{aligned}$$

The 'difference-in-differences' structure of the ALC estimand means that any selection bias allowed for by the no matching on unobservables restriction is differenced away (assuming the production function is given by (20) above).

Since $\beta(R^a)$ only depends on $\delta(w,x)$ through the ALC terms it is also identified under additional support conditions.

Proposition 3.3 (Identification Under No Matching on Unobservables) *If*
(i) $Y_{(i,j)} = \delta(W_i,X^j) + \lambda(\varepsilon_i,X^j) + \rho(W_i,v^j)$,

[22] To see this note that

$$\begin{aligned}\mathbb{E}[Y_i|W_i = w, X_i = x] &= \sum\nolimits_{f=1}^{F}\sum\nolimits_{g=1}^{G} k(w, x, e_f, v_g) s_{\varepsilon|X}(e_f|x) s_{v|W}(v_g|w) \\ &= \delta(w, x) + \sum\nolimits_{f=1}^{F} \lambda(e_f, x) s_{\varepsilon|X}(e_f|x) + \sum\nolimits_{g=1}^{G} \rho(w, v_g) s_{v|W}(v_g|w) \\ &= \delta(w, x) + \mathbb{E}[\lambda(\varepsilon_i, X_i)|X_i = x] + \mathbb{E}\left[\rho\left(W_i, v^{m(i)}\right)\middle| W_i = w\right] \\ &\quad \delta(w, x) + \bar{\lambda}(x) + \bar{\rho}(w).\end{aligned}$$

(ii) *the status quo assignment is of the form given by (9), and*
(iii) $r_{kl}^{\mathrm{a}} > 0$ *only if* $r_{kl}^{\mathrm{sq}} > 0$, *then*

$$\phi_{KLkl} = \mathbb{E}[Y_i \mid W_i = w_K, X_i = x_L] - \mathbb{E}[Y_i \mid W_i = w_K, X_i = x_l] \\ - (\mathbb{E}[Y_i \mid W_i = w_k, X_i = x_L] - \mathbb{E}[Y_i \mid W_i = w_k, X_i = x_l]),$$

and $\beta(R_{WX}^{\mathrm{a}})$ *is identified. If* $r_{kl}^{\mathrm{sq}} > 0$ *for all* $k = 1, \ldots, K$ *and* $l = 1, \ldots, L$, *then* β^{mre} *is also identified.*

Proposition 3.3 is new. It is a consequence of (i) the 'difference-in-differences' or 'increasing differences' structure of the ALC estimand and (ii) the type of selection bias allowed under the no matching on unobservables assumption. Consider the difference in match output across matches with workers of type $X_i = x'$ versus $X_i = x$:

$$\mathbb{E}[Y_i \mid W_i = w', X_i = x'] - \mathbb{E}[Y_i \mid W_i = w', X_i = x] = \delta(w', x') - \delta(w', x) + \bar{\lambda}(x') - \bar{\lambda}(x). \tag{23}$$

The first term $-\ \delta(w',x')\ -\ \delta(w',x)\ -$ is the systematic return that a type $W_i = w'$ firm gets from hiring a type $X_i = x'$ versus $X_i = x$ worker. The second term $-\bar{\lambda}(x') - \bar{\lambda}(x)-$ captures the difference in average firm productivity across the two types of matches. This may arise from selective input choice on the part of the firm. The key point is that the combination of the restricted match output function (20) and the no matching on unobservables assumption implies that this latter term is constant in firm type. Consequently we also have

$$\mathbb{E}[Y_i \mid W_i = w, X_i = x'] - \mathbb{E}[Y_i \mid W_i = w, X_i = x] = \delta(w, x') - \delta(w, x) + \bar{\lambda}(x') - \bar{\lambda}(x). \tag{24}$$

Subtracting (24) from (23) yields the ALC. Underlying Proposition 3.3 are strong assumptions the appropriateness of which will vary from application to application. The proposition does highlight the gains from directly searching for restrictions which identify ALC (as opposed to first identifying the AMF and then computing ALC).

3.2.4 Further thoughts on the identification of AREs

The definition of the AMF as an average over the product of two unobserved heterogeneity distributions makes identification of reallocation effects particularly challenging. This section has outlined two approaches. The first leaves the match surplus function nonparametric but imposes strong restrictions on the status quo assignment. These restrictions can be weakened somewhat by additional conditioning (Proposition 3.2). The second approach involves imposing separability assumptions on the match output function. As noted above these restrictions are strong, particularly in the context of allocation problems where complementarity properties are paramount. However, such assumptions allow for the formulation of positive identification results under weaker restrictions on the status quo assignment.

An area that merits further thought is the value of partially identifying restrictions of the type discussed in Manski (2003). For example sign, monotonicity or other restrictions on the ALC may be both well motivated and informative about reallocation effects in some situations.

3.3 Continuously-valued match characteristics

Graham, Imbens, and Ridder (2009a) study identification and estimation of AREs when match characteristics are continuously-valued. Continuity of agent characteristics makes some features of assignment problems simpler. For example the definitions of the positive and negative assortative matchings are less clumsy in this case. Other aspects of the problem become more challenging. The class of feasible assignments becomes very large, making identification of optimal allocations difficult. Formally the planner's problem is a nonconvex functional (i.e., infinite dimensional) optimization problem. Such problems, unlike linear programs, are quite hard to solve in general (e.g., Luenberger, 1969).

As before match output is given by $Y_{(i,j)} = k(W_i, X^j, \varepsilon_i, \nu^j)$. However, now both W_i and X^j are continuously-valued. The average match output function is

$$\delta(w,x) = \int\int k(w,x,\varepsilon,\nu) f_\varepsilon(e) f_\nu(\nu) \mathrm{d}e\mathrm{d}\nu$$

with $f_\varepsilon(e)$ and $f_\nu(\nu)$ the marginal density functions for, respectively, firm productivity and worker ability.

To keep the exposition simple assume double randomization such that the status quo assignment *density* is given by

$$h^{\mathrm{sq}}(w,x,e,\nu) = r_{WX}^{\mathrm{sq}}(w,x) f_\varepsilon(e) f_\nu(\nu),$$

so that the AMF is identified by $\delta(w,x) = \mathbb{E}[Y_i \mid W_i = w, X_i = x]$.

Feasibility of an assignment density, $r_{WX}(w, x)$, requires that

$$\int_w r_{WX}(w,x)\ \mathrm{d}w = f_X(x), \qquad \int_x r_{WX}(w,x)\ \mathrm{d}x = f_W(w), \tag{25}$$

with $f_W(w)$ and $f_X(x)$ the marginal density functions for, respectively, the firm and worker attributes (or types). The class of reallocations studied by Graham, Imbens and Ridder (2009a) consists of all joint densities satisfying (25).

The first estimand they consider is expected outcome gain from perfect assortative matching of W_i on X^j:

$$\beta^{\mathrm{pam}} = \mathbb{E}[\delta(F_W^{-1}(F_X(X_i)), X_i) - Y_i], \tag{26}$$

where $F_X(X^j)$ denotes the CDF of X^j, and $F_W^{-1}(q)$ is the q-th quantile of the distribution of W_i. Therefore $F_W^{-1}(F_X(X_i))$ computes the location of match $i's$ worker on the

CDF of X^j and reassigns her to a firm on the corresponding quantile of the distribution of W_i. Those workers with the highest value of X^j are reassigned to firms with the highest value of W_i, and so on.

The average outcome gain from negative assortative matching follows similarly with

$$\beta^{\text{nam}} = \mathbb{E}[\delta(F_W^{-1}(1 - F_X(X_i)), X_i) - Y_i]. \tag{27}$$

Note that (26) and (27) are related to the partial mean estimand introduced by Newey (1994). While the average match function $\delta(w,x)$ is a bivariate function, both (26) and (27) are averages over a single random variable. This reflects the fact that under perfectly assortative matchings the conditional distributions of W_i given X_i is degenerate: knowledge of X_i implies knowledge of W_i. This feature of the β^{pam} and β^{nam} estimands affects the limiting distribution of their sample analogs (see Section 3.4).

Because the class of feasible allocations is so large, Graham, Imbens and Ridder (2009a) do not attempt to identify optimal allocations. Instead they introduce a two parameter family of feasible allocations called *correlated matchings*. Let $H(\cdot, \cdot; \rho)$ denote the CDF of a standard bivariate normal random variable with correlation coefficient ρ.[23] Let $R^{\text{sq}}(w, x)$, in a change of notation, denote the CDF of the joint distribution of (W_i, X_i) under the status quo. Correlating matchings are given by

$$R^{\text{cm}}(w, x; \tau, \rho) = \tau R^{\text{sq}}(w, x) + (1 - \tau) H(\Phi^{-1}(F_W(w)), \Phi^{-1}(F_X(x)); \rho),$$

for τ between zero and one and ρ between -1 and 1.

The effect of implementing a correlated matching on average outcomes is thus

$$\beta^{\text{cm}}(\rho, \tau) = (1 - \tau) \times \left\{ \int_w \int_x \delta(w, x) \frac{\phi(\Phi^{-1}(F_W(w)), \Phi^{-1}(F_X(x)); \rho)}{\phi(\Phi^{-1}(F_W(w)))\phi(\Phi^{-1}(F_X(x)))} f_W(w) f_X(x)\, \mathrm{d}w\mathrm{d}x - \mathbb{E}[Y_i] \right\}. \tag{28}$$

By varying ρ from 1 to -1 for $\tau = 0$ correlated matchings trace a path from the positive, through the random ($\rho = 0$), to the negative assortative matching. By setting $\tau = 1$ they can reproduce the status quo assignment. Note that unless $\delta(w,x)$ is supermodular, $\beta^{\text{cm}}(\rho,\tau)$ need not vary monotonically with ρ. Furthermore there is no guarantee that an average outcome maximizing allocation corresponds to a correlated matching. Indeed, it would be surprising if it did.

One way to conceptualize correlated matchings is to view them as particular perturbations of the random allocation. To see this note that

$$\beta^{\text{cm}}(\rho, \tau) = (1 - \tau)\left\{\mathbb{E}_{W_i}[\mathbb{E}_{X^j}[\delta(W_i, X^j)\eta(W_i, X^j; \rho)]] - \mathbb{E}[Y_i]\right\},$$

[23] To avoid small denominator problems, they actually work with a truncated bivariate normal distribution function.

for

$$\eta\left(W_i, X^j; \rho\right) = \frac{\phi(\Phi^{-1}(F_W(W_i)), \Phi^{-1}(F_X(X^j)); \rho)}{\phi(\Phi^{-1}(F_W(W_i)))\phi(\Phi^{-1}(F_X(X^j)))}.$$

If $\rho = 0$, the weight function $\eta(w, x; \rho)$ is identically equal to one for all w and x and $\beta^{\mathrm{cm}}(0, \tau) = (1 - \tau)\{\mathbb{E}_{W_i}[\mathbb{E}_{X^j}[\delta(W_i, X^j)]] - \mathbb{E}[Y_i]\}$. If $\rho > 0$, then $\eta(w, x; \rho)$ will be larger for pairs of W_i and X^j that correspond to similar quantiles of their respective marginal distributions. Likewise if $\rho < 0$, then $\eta(w, x; \rho)$ will be larger for pairs of W_i and X^j that correspond to very different quantiles. In the limit as $\rho \to 1$ the weight function $\eta(W_i, X^j; \rho)$ is only non-zero when the quantiles of W_i and X^j coincide, as in the perfect positive assortative matching. When $\rho \to -1$ we recover the negative assortative matching.

Identification of β^{pam}, β^{nam} and $\beta^{\mathrm{cm}}(\rho, \tau)$ requires strong support conditions. A sufficient condition for identification is that the joint support of the status quo assignment $R^{\mathrm{sq}}(w,x)$ coincides with the product of its marginals' support. This condition allows the econometrician to learn about $\delta(w,x)$ at all conceptually possible combinations of w and x. Given this knowledge the average outcome across *any* feasible assignment can be computed by integration. Unfortunately in many datasets this support condition will fail (or effectively fail) to hold. For example, if under the status quo there is strong positive dependence between W_i and X_i, then it will be difficult to identify β^{nam}. This is because β^{nam} is an average of $\delta(w,x)$ over pairs of w and x where w is large (small) and x is small (large). These are precisely the match types that are infrequently observed in a status quo with 'lots of' positive dependence.

In addition to being difficult to identify, assignments that are distant from the status quo may be less policy relevant. Policies which represent incremental modifications of the status quo may be politically and/or logistically easier to implement than large reallocations.[24] Motivated by these issues Graham, Imbens and Ridder (2009a) also study *local reallocation effects* (LREs) (i.e., the effects of small reallocations 'away' from the status quo and 'toward' the perfect positive assortative matching).

They derive the LRE by matching on a family of transformations of X_i and W_i, indexed by a scalar parameter λ, where for some values of λ the matching is on W_i (corresponding to the status quo), and for other values of λ the matching is on X_i or $-X_i$, corresponding to positive and negative assortative matching respectively. Assume that the supports of W_i and X_i are given by the intervals $[w_l, w_u]$ and $[x_l, x_u]$. Let $d(w)$ be the following smooth function that goes to zero at the boundary of the support of W_i:

$$d(w) = 1_{w > w_m} \cdot (w_u - w) + 1_{w \leq w_m} \cdot (w - w_l),$$

where w_m is the midpoint of the support of W.

[24] Of course some organizations, like the military, have a greater ability to implement radical organizational changes.

For $\lambda \in [-1, 1]$, define the random variable $U_{\lambda i}$ as the following transformation of (X_i, W_i):

$$U_{\lambda i} = \lambda \cdot X_i \cdot d(W_i)^{1-|\lambda|} + (\sqrt{1-\lambda^2}) \cdot W_i.$$

Now we consider reallocations based on positive assortative matching on $U_{\lambda i}$:

$$\beta^{\mathrm{lr}}(\lambda) = \mathbb{E}[\delta(F_W^{-1}(F_{U_\lambda}(U_{\lambda i})), X_i)]. \tag{29}$$

For $\lambda = 0$ and $\lambda = 1$ we have $U_{\lambda i} = W_i$ and $U_{\lambda i} = X_i$ respectively, and hence $\beta^{\mathrm{lr}}(0)$ recovers the status quo average outcome $\mathbb{E}[Y_i]$ and $\beta^{\mathrm{lr}}(1) = \beta^{\mathrm{pam}}$. The negative assortative matching is also nested in this framework since

$$\Pr(-X^j \leq -x) = \Pr(X^j \geq x) = 1 - F_X(x),$$

and hence for $\lambda = -1$ we have $\beta^{\mathrm{lr}}(-1) = \beta^{\mathrm{nam}}$. Values of λ close to zero induce reallocations of W_i that are 'local' to the status quo, with $\lambda > 0$ and $\lambda < 0$ generating shifts toward positive and negative assortative matching respectively.

Graham, Imbens and Ridder (2009a) consider the direction of the effect on average outcomes associated with a small step toward the positive assortative matching:

$$\beta^{\mathrm{lc}} = \frac{\partial \beta^{\mathrm{lr}}}{\partial \lambda}(0). \tag{30}$$

Theorem 3.1 of their paper presents two representations of this derivative:

$$\beta^{\mathrm{lc}} = \mathbb{E}\left[d(W) \cdot \mathbb{C}\left(\frac{\partial}{\partial_w}\delta(W, X), X \mid W\right)\right], \tag{31}$$

and

$$\beta^{\mathrm{lc}} = \mathbb{E}\left[\omega(W, X) \cdot \frac{\partial^2}{\partial w \partial x}\delta(W, X)\right], \tag{32}$$

where the weight function $\omega(w,x)$ is non-negative and has the form

$$\omega(w,x) = d(w) \cdot \frac{F_{X|W}(x|w) \cdot (1 - F_{X|W}(x|w))}{f_{X|W}(x|w)} \cdot (\mathbb{E}[X|X > x, W = w] - \mathbb{E}[X|X \leq x, W = w]).$$

The first representation of β^{lc} motivates their approach to estimation. It implies that β^{lc} will be positive if the conditional covariance between $\frac{\partial}{\partial w}\delta(W, X)$ and X is positive. This will occur if the return to increases in firm type, $\frac{\partial}{\partial w}\delta(W, X)$, tends to be larger when X exceeds its conditional mean $\mathbb{E}[X \mid W]$. Intuitively this suggests that increasing assortativeness should raise average outcomes. Representation (31) makes this intuition precise.

The second representation shows that β^{lc} is a weighted average measure of complementarity between firm and worker type. This provides a connection between the presence of complementarity and reallocation effects.

3.4 Estimation of AREs when match output is observed

The estimation of matching models raises complex and interesting statistical issues. Some of these issues are well-known from the literature on semiparametric estimation (Powell, 1994), others are less familiar. Consider the structure of the ARE estimand. It is a function of three primitives (i) the production technology, (ii) the marginal distribution of the two inputs (firm and worker type) and (iii) the chosen allocation. This intertwined aspect of the target parameter can generate surprises. For example, Graham, Imbens and Ridder (2007) show that the limiting distribution of β^{pam}, when firm and worker type are discretely-valued, changes discontinuously in the marginal distributions on W and X. Discontinuities in limit distributions arise elsewhere in econometrics, for example in the literature on unit roots, weak instruments, and moment inequalities, but their presence in assignment problems is, at least a priori, surprising. When W and X are continuously-valued Graham, Imbens and Ridder (2009a) show that the rate of convergence is slower for their estimates of the two extremal allocations β^{pam} and β^{nam}, than for non-extremal correlated matchings β^{cm}.

3.4.1 Estimation of the average match output function (AMF)

Consider a setting where the status quo assignment is 'as if' conditionally doubly randomized as described in Section 3.1. If all covariates are discretely-valued the estimated proxy variable regression introduced in Proposition 3.2 is given by the cell mean

$$\hat{g}(w,x,z_\varepsilon,z_\nu)=\frac{\sum_{i=1}^{N} Y_i \mathbf{1}(W_i=w,X_i=x,Z_{\varepsilon i}=z_\varepsilon,Z_{\nu i}=z_\nu)}{\sum_{i=1}^{N} \mathbf{1}(W_i=w,X_i=x,X_{\varepsilon i}=z_\varepsilon,Z_{\nu i}=z_\nu)}.$$

When covariates are continuously-valued $g(w, x, z_\varepsilon, z_\nu)$ can be estimated by kernel regression. Graham, Imbens and Ridder (2009a) use the Nearest Interior Point (NIP) kernel estimator of Imbens and Ridder (2009). This estimator eliminates boundary bias present in the standard Nadaraya-Watson estimate.

In either case the average match output function (AMF) is recovered by separately averaging over the sample distributions of $Z_{\varepsilon i}$ and $Z_{\nu j}$:

$$\hat{\delta}(w,x)=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(w,x,Z_{\varepsilon i},Z_{\nu j}). \tag{33}$$

Equation (33) is similar to the partial mean estimator of Newey (1994b). It differs in that it averages over a product of two marginals instead of the joint distribution of

$(Z_{\varepsilon i}, Z_{\nu j})$. Its asymptotic properties are unknown, although they should be straightforward to derive.

For simplicity in what follows I will assume the status quo assignment is doubly randomized. Many of the results will also hold if instead it satisfies the no matching on unobservables condition *and* the match production function is given by (20). When this is the case should be obvious by the context. Under double randomization, with discretely-valued covariates, we may estimate the AMF by

$$\hat{\delta}(w,x) = \frac{\sum_{i=1}^{N} Y_i \mathbf{1}(W_i = w, X_i = x)}{\sum_{i=1}^{N} \mathbf{1}(W_i = w, X_i = x)}, \tag{34}$$

With continuously-valued covariates $\delta(w,x)$ may be estimated by NIP kernel regression.

3.4.2 AREs with discretely-valued covariates

Fixed 'interior' allocations Inference on AREs for fixed interior allocations when match characteristics are discretely-valued is straightforward. To illustrate I consider only the simple $K = L = 2$ case. Generalizing what follows to allow for K, $L > 2$ is straightforward, albeit tedious.

Recall from Section 3.1 that when $K = L = 2$ we have the one parameter representation

$$R_{WX}^{\mathrm{a}}(r_{11}^{\mathrm{a}}) = \begin{pmatrix} r_{11}^{\mathrm{a}} & p_1 - r_{11}^{\mathrm{a}} \\ q_1 - r_{11}^{\mathrm{a}} & 1 - p_1 - q_1 + r_{11}^{\mathrm{a}} \end{pmatrix},$$

for all r_{11}^{a} such that R_{WX}^{a} is a valid joint distribution or

$$p_1 - \min\{p_1, q_1\} \leq r_{11}^{\mathrm{a}} \leq \min\{p_1, q_1\}. \tag{35}$$

Interior allocations consist of all allocations where r_{11}^{a} is non-stochastic and the inequalities in (35) are strict.

Letting $\beta(R^{\mathrm{a}}) = \beta^{\mathrm{a}}$ the analog estimator is

$$\hat{\beta}^{\mathrm{a}} = \left\{ r_{11}^{\mathrm{a}} - \hat{r}_{11}^{\mathrm{sq}} \right\} \times \hat{\phi},$$

with

$$\hat{\phi} = \hat{\delta}_{22} - \hat{\delta}_{21} - \left(\hat{\delta}_{12} - \hat{\delta}_{11} \right),$$

and $\hat{\delta}_{kl} = \hat{\delta}(w_k, x_l)$ given by (34) above. Note that sampling error in $\hat{\beta}^{\mathrm{a}}$ will reflect both uncertainty in (i) the form of the match production technology (in this case ϕ, the complementarity measure) and (ii) the status quo assignment distribution (in this case r_{11}^{sq}).

The delta method gives

$$\sqrt{N}(\hat{\phi}-\phi_0)\xrightarrow{D}\mathcal{N}\left(0,\frac{\sigma_{11}^2}{r_{11}^{\mathrm{sq}}}+\frac{\sigma_{12}^2}{p_1-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{21}^2}{q_1-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{22}^2}{1-p_1-q_1+r_{11}^{\mathrm{sq}}}\right),$$

where $\sigma_{kl}^2=\mathbb{V}(Y_i\mid W_i=w_k,X_i=x_l)$.

The status quo assignment, which enters the definition of β^{a} through r_{11}^{sq}, is assumed unknown. However it may be consistently estimated by

$$\hat{p}_1=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(W_i=w_1),\quad \hat{q}_1=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(X_i=x_1),\quad \hat{r}_{11}^{\mathrm{sq}}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(W_i=w_1,X_i=x_1),$$

with a large sample distribution equal to

$$\sqrt{N}\begin{pmatrix}\hat{p}_1-p_1\\ \hat{q}_1-q_1\\ \hat{r}_{11}^{\mathrm{sq}}-r_{11}^{\mathrm{sq}}\end{pmatrix}\xrightarrow{D}\mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}p_1(1-p_1) & r_{11}^{\mathrm{sq}}-p_1q_1 & r_{11}^{\mathrm{sq}}(1-p_1)\\ r_{11}^{\mathrm{sq}}-p_1q_1 & q_1(1-q_1) & r_{11}^{\mathrm{sq}}(1-q_1)\\ r_{11}^{\mathrm{sq}}(1-p_1) & r_{11}^{\mathrm{sq}}(1-q_1) & r_{11}^{\mathrm{sq}}(1-r_{11}^{\mathrm{sq}})\end{pmatrix}\right).$$

Note that sampling error in the estimate of R^{sq} is asymptotically independent of that in $\hat{\phi}$.

Since $\hat{\beta}^{\mathrm{a}}$ is a continuous function of sample averages $\hat{\beta}^{\mathrm{a}}$ is consistent for β^{a}. A second application of the delta method then gives an asymptotic sampling distribution of

$$\sqrt{N}(\hat{\beta}^{\mathrm{a}}-\beta^{\mathrm{a}})\xrightarrow{D}Z_a \tag{36}$$

where Z_a is the normally distributed random variable

$$Z_a\sim\mathcal{N}\left(0,(r_{11}^{\mathrm{a}}-r_{11}^{\mathrm{sq}})^2\left\{\frac{\sigma_{11}^2}{r_{11}^{\mathrm{sq}}}+\frac{\sigma_{12}^2}{p_1-r_{11}^{\mathrm{sq}}}+.\ \frac{\sigma_{21}^2}{q_1-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{22}^2}{1-p_1-q_1+r_{11}^{\mathrm{sq}}}\right\}+\phi_0^2 r_{11}^{\mathrm{sq}}(1-r_{11}^{\mathrm{sq}})\right).$$

The asymptotic sampling variance of $\hat{\beta}^{\mathrm{a}}$ has two components. The first reflects sampling error in $\hat{\phi}$, the degree of average local complementarity between W and X, *as well as* the distance between the new allocation and the status quo, $r_{11}^{\mathrm{a}}-r_{11}^{\mathrm{sq}}$. The greater the distance between the counterfactual assignment of interest and the status quo, the greater our uncertainty about the magnitude of the ARE. The second source of sampling error in $\hat{\beta}^{\mathrm{a}}$ arises from our imperfect knowledge of the status quo assignment distribution; sampling error in $\hat{r}_{11}^{\mathrm{sq}}$. Since all the components entering its asymptotic variance are consistently estimable, conducting large sample inference on $\hat{\beta}^{\mathrm{a}}$ is straightforward.

Extreme allocations Estimation of, and inference on, extremal allocations raises new and interesting issues. First consider the case of the positive assortative matching. When $K = L = 2$ the positive assortative matching chooses $r_{11}^{\text{pam}} = \min\{p_1, q_1\}$, so that

$$\beta^{\text{pam}} = \min\{(p_1 - r_{11}^{\text{sq}})\phi, (q_1 - r_{11}^{\text{sq}})\phi\}. \tag{37}$$

The analog estimator is

$$\hat{\beta}^{\text{pam}} = \min\{(\hat{p}_1 - \hat{r}_{11}^{\text{sq}})\hat{\phi}, (\hat{q}_1 - \hat{r}_{11}^{\text{sq}})\hat{\phi}\}.$$

Since the function $\min\{(p_1 - r_{11}^{\text{sq}})\phi, (q_1 - r_{11}^{\text{sq}})\phi\}$ is continuous we have, by the continuous mapping theorem, $\hat{\beta}^{\text{pam}} \xrightarrow{p} \beta^{\text{pam}}$. While the demonstration of consistency is straightforward, characterizing the asymptotic sampling distribution of $\hat{\beta}^{\text{pam}}$ is more difficult. This is because the definition of β^{pam} depends on unknown features of the population distribution of firm and worker types. In particular $\hat{\beta}^{\text{pam}}$ has three possible limit distributions depending on whether (i) $p_1 > q_1$, (ii) $p_1 < q_1$, or (iii) $p_1 = q_1$. Since we do not know which case obtains a priori Graham, Imbens and Ridder (2007) suggest a conservative approach to inference. To describe this approach to inference we first need to characterize the three limiting distributions.

Case 1 $p_1 > q_1$: When $p_1 > q_1$, so that type 1 firms are more numerous than type 1 workers, we have

$$\begin{aligned}\sqrt{N}(\min\{\hat{p}_1, \hat{q}_1\} - \min\{p_1, q_1\}) &= \sqrt{N}(\min\{\hat{p}_1, \hat{q}_1\} - q_1) \\ &= \min\{\sqrt{N}(\hat{p}_1 - p_1) + \sqrt{N}(p_1 - q_1), \sqrt{N}(\hat{q}_1 - q_1)\}.\end{aligned}$$

Since $\sqrt{N}(p_1 - q_1) > 0$ this gives

$$|\sqrt{N}(\min\{\hat{p}_1, \hat{q}_1\} - \min\{p_1, q_1\}) - \sqrt{N}(\hat{q}_1 - q_1)| \xrightarrow{p} 0,$$

which allows us to replace $\min\left\{(\hat{p}_1 - \hat{r}_{11}^{\text{sq}})\hat{\phi}, (\hat{q}_1 - \hat{r}_{11}^{\text{sq}})\hat{\phi}\right\}$ in (79) with $(\hat{q}_1 - \hat{r}_{11}^{\text{sq}})\hat{\phi}$. Following a sequence of steps analogous to those described above for fixed interior allocations we then get

$$\sqrt{N}\left(\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}\right) \xrightarrow{D} Z_{q-r}, \tag{38}$$

with Z_{q-r} the normal random variable

$$\begin{aligned}Z_{q-r} \sim \mathcal{N}\Bigg(0, (q_1 - r_{11}^{\text{sq}})^2 \Bigg\{ &\frac{\sigma_{11}^2}{r_{11}^{\text{sq}}} + \frac{\sigma_{12}^2}{p_1 - r_{11}^{\text{sq}}} + \frac{\sigma_{21}^2}{q_1 - r_{11}^{\text{sq}}} \\ &+ \frac{\sigma_{22}^2}{1 - p_1 - q_1 + r_{11}^{\text{sq}}} \Bigg\} + \phi^2(1 - q_1 - r_{11}^{\text{sq}})(q_1 - r_{11}^{\text{sq}})\Bigg).\end{aligned}$$

Note that uncertainty in two features of the status quo, q_1 and r_{11}^{sq}, are reflected in the sampling variance of Z_{q-r}. This is because q_1 enters in the definition of β^{pam}.

Case 2 $p_1 < q_1$: When $p_1 < q_1$, so that type 1 firms are less numerous than type 1 workers, we have, following an argument parallel to case 1 above,

$$\sqrt{N}(\hat{\beta}^{\mathrm{pam}} - \beta^{\mathrm{pam}}) \xrightarrow{D} Z_{p-r} \tag{39}$$

with Z_{p-r} the normal random variable

$$Z_{p-r} \sim \mathcal{N}\left(0, (p_1 - r_{11}^{\mathrm{sq}})^2 \left\{ \frac{\sigma_{11}^2}{r_{11}^{\mathrm{sq}}} + \frac{\sigma_{12}^2}{p_1 - r_{11}^{\mathrm{sq}}} + \frac{\sigma_{21}^2}{q_1 - r_{11}^{\mathrm{sq}}} + \frac{\sigma_{22}^2}{1 - p_1 - q_1 + r_{11}^{\mathrm{sq}}} \right\} + \phi^2 (1 - p_1 - r_{11}^{\mathrm{sq}})(p_1 - r_{11}^{\mathrm{sq}}) \right).$$

Case 3 $p_1 = q_1$: The limit distribution for the third case, which corresponds to the marginal distributions of W and X coinciding, is nonstandard. To see this note that when $p_1 = q_1$ we have

$$\sqrt{N}(\min\{\hat{p}_1, \hat{q}_1\} - \min\{p_1, q_1\}) = \min\left\{\sqrt{N}(\hat{p}_1 - p_1), \sqrt{N}(\hat{q}_1 - q_1)\right\} \xrightarrow{D} \min(U_p, U_q),$$

with (U_p, U_q) the bivariate normal random variable:

$$\begin{pmatrix} U_p \\ U_q \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} p_1(1-p_1) & r_{11}^{\mathrm{sq}} - p_1^2 \\ r_{11}^{\mathrm{sq}} - p_1^2 & p_1(1-p_1) \end{pmatrix} \right).$$

Recalling the definition of β^{pam} we then get

$$\sqrt{N}(\hat{\beta}^{\mathrm{pam}} - \beta^{\mathrm{pam}}) \xrightarrow{D} \min\{Z_{p-r}, Z_{q-r}\},$$

with Z_{q-r} and Z_{p-r} the normal random variables defined by (38) and (39) above. Their covariance is given by

$$(p_1 - r_{11}^{\mathrm{sq}})(q_1 - r_{11}^{\mathrm{sq}}) \left\{ \frac{\sigma_{11}^2}{r_{11}^{\mathrm{sq}}} + \frac{\sigma_{12}^2}{p_1 - r_{11}^{\mathrm{sq}}} + \frac{\sigma_{21}^2}{q_1 - r_{11}^{\mathrm{sq}}} + \frac{\sigma_{22}^2}{1 - p_1 - q_1 + r_{11}^{\mathrm{sq}}} \right\} - \phi^2 (p_1 - r_{11}^{\mathrm{sq}})(q_1 - r_{11}^{\mathrm{sq}}).$$

While the distribution of $\min\{Z_{p-r}, Z_{q-r}\}$ is difficult to characterize analytically, it is straightforward to simulate from: (i) draw Z_{p-r} and Z_{q-r} jointly, (ii) calculate their minimum, and (iii) repeat. If we knew that $p_1 = q_1$ in the population, then we could simulate critical values for testing hypotheses. Consider the null hypothesis

$\beta^{\text{pam}} = \beta_0^{\text{pam}}$ versus the alternative $\beta^{\text{pam}} \neq \beta_0^{\text{pam}}$. The proposal is to construct the statistic $T_N = \sqrt{N}\left(\hat{\beta}^{\text{pam}} - \beta_0^{\text{pam}}\right)$ and reject if $|T_N| > C^{1-\alpha}$ where $C^{1-\alpha}$ is the $1-\alpha$ quantile of the simulated distribution of $|\min\{Z_{p-r}, Z_{q-r}\}|$. To obtain a $1-\alpha$ confidence level we invert the test.

In practice we do not know which state of the world obtains: $p_1 > q_1$, $p_1 < q_1$ or $p_1 = q_1$. Graham, Imbens and Ridder (2007) suggest calculating the critical value for each case and choosing the largest of the three. That is reject if $|T_N| > C^{1-\alpha}$ with $C^{1-\alpha}$ such that

$$\sup_{p_1>q_1, p_1<q_1, p_1=q_1} \left\{ \lim_{N\to\infty} \Pr\left(\left| \sqrt{N}(\hat{\beta}^{\text{pam}} - \beta_0^{\text{pam}}) \right| > C^{1-\alpha} \right) \leq \alpha \right\}.$$

An interesting feature of the above analysis is that attributes of the distribution of the 'regressors' feature centrally in the inferential procedure. This is quite different from textbook hypothesis testing. The difference arises from the nature of the estimand: the distribution of agent characteristics features directly in the definition of β^{pam}.

Inference on the negative assortative matching is essentially the same as in the positive case. The negative assortative matching sets

$$r_{11}^{\text{nam}} = \max\{0, p_1 + q_1 - 1\},$$

yielding the estimand

$$\beta^{\text{nam}} = \max\{-r_{11}^{\text{sq}}\phi, -(1 - p_1 - q_1 + r_{11}^{\text{sq}})\phi\}.$$

As before we must consider three cases: (i) $p_1 > 1 - q_1$ (which also corresponds to the $p_1 < q_1$ case above), (ii) $p_1 < 1 - q_1$ (which corresponds to $p_1 > q_1$ case above) and (iii) $p_1 = 1 - q_1$ (which need not coincide with the $p_1 = q_1$ case discussed above).

In the first case $p_1 > 1 - q_1$ so the limiting distribution of $\hat{\beta}^{\text{nam}}$ coincides with that of $-(1 - \hat{p}_1 - \hat{q}_1 + \hat{r}_{11}^{\text{sq}})\hat{\phi}$ giving

$$\sqrt{N}\left(\hat{\beta}^{\text{nam}} - \beta^{\text{nam}}\right) \xrightarrow{D} Z_{1-p-q+r} \tag{40}$$

for $Z_{1-p-q+r}$ the normal random variable

$$Z_{1-p-q+r} \sim \mathcal{N}\Bigg(0, (1 - p_1 - q_1 + r_{11}^{\text{sq}})^2 \Bigg\{ \frac{\sigma_{11}^2}{r_{11}^{\text{sq}}} + \frac{\sigma_{12}^2}{p_1 - r_{11}^{\text{sq}}} + \frac{\sigma_{21}^2}{q_1 - r_{11}^{\text{sq}}} + \frac{\sigma_{22}^2}{1 - p_1 - q_1 + r_{11}^{\text{sq}}} \Bigg\} + \phi^2 (p_1 + q_1 - r_{11}^{\text{sq}})(1 - p_1 - q_1 + r_{11}^{\text{sq}}) \Bigg).$$

In the second case $p_1 < 1 - q_1$ so the limiting distribution of $\hat{\beta}^{\text{nam}}$ coincides with that of $-\hat{r}_{11}^{\text{sq}}\hat{\phi}$ giving

$$\sqrt{N}\left(\hat{\beta}^{\mathrm{nam}}-\beta^{\mathrm{nam}}\right)\xrightarrow{D}Z_{-r} \tag{41}$$

for Z_{-r} the normal random variable

$$Z_{-r}\sim\mathcal{N}\left(0,\left(r_{11}^{\mathrm{sq}}\right)^{2}\left\{\frac{\sigma_{11}^{2}}{r_{11}^{\mathrm{sq}}}+\frac{\sigma_{12}^{2}}{p_{1}-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{21}^{2}}{q_{1}-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{22}^{2}}{1-p_{1}-q_{1}+r_{11}^{\mathrm{sq}}}\right\}+\phi^{2}r_{11}^{\mathrm{sq}}\left(1-r_{11}^{\mathrm{sq}}\right)\right).$$

In the third case $p_1 = 1 - q_1$ so that the limit distribution coincides with

$$\sqrt{N}\left(\hat{\beta}^{\mathrm{nam}}-\beta^{\mathrm{nam}}\right)\xrightarrow{D}\max\left\{Z_{-r},Z_{1-p-q+r}\right\},$$

with $Z_{1-p-q+r}$ and Z_{-r} as defined in (41) and (40) with a covariance of

$$r_{11}^{\mathrm{sq}}\left(1-p_{1}-q_{1}+r_{11}^{\mathrm{sq}}\right)\left\{\frac{\sigma_{11}^{2}}{r_{11}^{\mathrm{sq}}}+\frac{\sigma_{12}^{2}}{p_{1}-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{21}^{2}}{q_{1}-r_{11}^{\mathrm{sq}}}+\frac{\sigma_{22}^{2}}{1-p_{1}-q_{1}+r_{11}^{\mathrm{sq}}}\right\}-\phi^{2}r_{11}^{\mathrm{sq}}\left(1-p_{1}-q_{1}+r_{11}^{\mathrm{sq}}\right).$$

We can conduct inference on β^{nam} in a manner analogous to the method described for the positive assortative matching above.

Optimal allocations The average reallocation effect associated with an optimal assignment is given by, continuing with the $K = L = 2$ case, $\beta^{\mathrm{mre}} = \max\{\beta^{\mathrm{pam}}, \beta^{\mathrm{nam}}\}$. It is consistently estimated by

$$\hat{\beta}^{\mathrm{mre}}=\max\left\{\hat{\beta}^{\mathrm{pam}},\hat{\beta}^{\mathrm{nam}}\right\}.$$

Let Z^{pam} denote a random variable whose distribution coincides with the limiting distribution of $\sqrt{N}(\hat{\beta}^{\mathrm{pam}} - \beta^{\mathrm{pam}})$. Recall from the discussion above that this limit distribution may take any of three forms. If $\beta^{\mathrm{pam}} > \beta^{\mathrm{nam}}$ in the population then we have $\sqrt{N}(\hat{\beta}^{\mathrm{mre}} - \beta^{\mathrm{mre}}) \xrightarrow{D} Z^{\mathrm{pam}}$. Let Z^{nam} denote the limiting distribution of $\sqrt{N}(\hat{\beta}^{\mathrm{nam}} - \beta^{\mathrm{nam}})$, which may also take three forms. If $\beta^{\mathrm{pam}} < \beta^{\mathrm{nam}}$, then we have $\sqrt{N}(\hat{\beta}^{\mathrm{mre}} - \beta^{\mathrm{mre}}) \xrightarrow{D} Z^{\mathrm{nam}}$. Finally consider the degenerate case where $\beta^{\mathrm{pam}} = \beta^{\mathrm{nam}}$ (this occurs if the ALC is identically equal to zero). In the degenerate case the limit distribution is given by $\sqrt{N}(\hat{\beta}^{\mathrm{mre}} - \beta^{\mathrm{mre}}) \xrightarrow{D} \max\{Z^{\mathrm{pam}}, Z^{\mathrm{nam}}\}$.

As in the case of the two extremal allocations conservative tests and confidence intervals may be constructed by computing critical values under all possible cases and

Table 2 Possible limit distributions for the maximum average reallocation effect when K = L = 2

Marginal Type Distribution	Limiting distributions for Z^{pam}	Z^{nam}
Panel A: Positive assortative matching optimal $\sqrt{N}\left(\hat{\beta}^{mre} - \beta^{mre}\right) \xrightarrow{D} Z^{pam}$ with:		
$p_1 > q_1$	Z_{q-r}	n.a.
$p_1 < q_1$	Z_{p-r}	n.a.
$p_1 = q_1$	$\min\{Z_{p-r}, Z_{q-r}\}$	n.a.
Panel B: Negative assortative matching optimal $\sqrt{N}\left(\hat{\beta}^{mre} - \beta^{mre}\right) \xrightarrow{D} Z^{nam}$ with:		
$p_1 < 1 - q_1$	n.a.	Z_{-r}
$p_1 > 1 - q_1$	n.a.	$Z_{1-p-q+r}$
$p_1 = 1 - q_1$	n.a.	$\max\{Z_{-r}, Z_{1-p-q+r}\}$
Panel C: Degenerate case $\sqrt{N}\left(\hat{\beta}^{mre} - \beta^{mre}\right) \xrightarrow{D} \max\{Z^{pam}, Z^{nam}\}$ with:		
$p_1 > q_1$ & $p_1 < 1 - q_1$	Z_{q-r}	Z_{-r}
$p_1 < q_1$ & $p_1 > 1 - q_1$	Z_{p-r}	$Z_{1-p-q+r}$
$p_1 = q_1$ & $p_1 < 1 - q_1$	$\min\{Z_{p-r}, Z_{q-r}\}$	Z_{-r}
$p_1 = q_1$ & $p_1 > 1 - q_1$	$\min\{Z_{p-r}, Z_{q-r}\}$	$Z_{1-p-q+r}$
$p_1 > q_1$ & $p_1 = 1 - q_1$	Z_{q-r}	$\max\{Z_{-r}, Z_{1-p-q+r}\}$
$p_1 < q_1$ & $p_1 = 1 - q_1$	Z_{p-r}	$\max\{Z_{-r}, Z_{1-p-q+r}\}$

picking the largest one. Now however there are now a total of 12 possible limit distributions to consider (see Table 2).[25]

The number of cases that must be considered will increase with K and/or L. Given its conservative nature the approach to inference outlined above is likely to have low power for modestly large K and/or L (this is consistent with the pattern found by Bhattacharya (2009) in his empirical application).

An alternative method of inference would adopt a Bayesian perspective. The planner would formulate a prior distribution for the parameters characterizing the status quo assignment as well as those of the production function. Inference would then be based on the resulting posterior distribution. Chamberlain (2009) considers some

[25] Bhattacharya (2009) also considers inference on β^{mre}, however he assumes that the marginal type distributions are known so he needs to only consider a total of three limit distributions when $K = L = 2$.

Bayesian aspects of unconstrained treatment choice. A number of his insights might be applicable here. Graham, Imbens, and Ridder (2007) provide a preliminary treatment of some decision theoretic issues.

3.4.3 AREs with continuously-valued covariates

Fixed 'interior' allocations The starting point for estimating the average outcome gain associated with implementing a correlated matching is equation (28) above. Note that $\beta^{\text{cm}}(\rho, \tau)$ is an integral over the product of the marginal pdfs of W and X, not the joint. Graham, Imbens, and Ridder (2009a) estimate $\beta^{\text{cm}}(\rho, \tau)$ by replacing these integrals with sums over the two empirical distribution functions to get the analog estimator

$$\hat{\beta}^{\text{cm}}(\rho,\tau) = (1-\tau)\left\{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\delta}(W_i, X_j)\frac{\phi\left(\Phi^{-1}(\hat{F}_W(W_i)), \Phi^{-1}(\hat{F}_X(X_j)); \rho\right)}{\phi\left(\Phi^{-1}(\hat{F}_W(W_i))\right)\phi\left(\Phi^{-1}(\hat{F}_X(X_j))\right)} - \frac{1}{N}\sum_{i=1}^{N}Y_i\right\}.$$

This estimator would be a standard second order V-statistic if $\delta(w,x)$, $F_W(w)$ and $F_X(x)$ were known. Instead $\hat{\beta}^{\text{cm}}(\rho,\tau)$ depends on nonparametric estimates of each of these objects. Sampling error in these nuisance parameters affects $\hat{\beta}^{\text{cm}}(\rho,\tau)$'s sampling properties.

To characterize the large sample properties of $\hat{\beta}^{\text{cm}}(\rho,\tau)$ Graham, Imbens and Ridder (2009a) first formulate a general theorem for double averages of kernel estimates (Theorem A.3). Their results demonstrate that the average outcome effects of correlated matchings are estimable at the regular $\sqrt{N}$ parametric rate. The influence function for their estimator is complicated with functions of the production technology, the marginal distributions of firm and worker types, the status quo assignment, and the precise correlated matching under consideration entering.

Extreme allocations Graham, Imbens and Ridder (2009a) also present estimation results for the extremal positive and negative assortative matchings. Their estimates are the sample analogs of (26) and (27) above, namely,

$$\hat{\beta}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\delta}\left(\hat{F}_W^{-1}(\hat{F}_X(X_i)), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}Y_i, \tag{42}$$

and

$$\hat{\beta}^{\text{nam}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\delta}\left(\hat{F}_W^{-1}(1-\hat{F}_X(X_i)), X_i\right) - \frac{1}{N}\sum_{i=1}^{N}Y_i. \tag{43}$$

It is straightforward to demonstrate consistency of these estimates. The nonparametric estimates $\hat{\delta}(w,x)$, $\hat{F}_W(w)$, and $\hat{F}_X(x)$ are uniformly consistent under their assumptions. Consistency then directly follows. The derivation of their sampling distributions is more involved. In contrast with correlated matchings, both $\hat{\beta}^{\text{pam}}$ and $\hat{\beta}^{\text{nam}}$ involve only a single average. This is because the conditional distribution of W_i given X_i is degenerate under an extremal assignment. Since in the first stage $\hat{\delta}(w,x)$ is estimated with two arguments, but in the second stage averaged over only one, its sampling error dominates the asymptotic variances of $\hat{\beta}^{\text{pam}}$ and $\hat{\beta}^{\text{nam}}$. This property is shared by the partial mean estimator of Newey (1994b). Graham, Imbens and Ridder (2009a) nevertheless propose accounting for asymptotically dominated terms when conducting inference. Their Monte Carlo results suggest that this idea has some merit.

Local reallocations The local reallocation effect is estimated by

$$\hat{\beta}^{\text{lc}} = \frac{1}{N}\sum\nolimits_{i=1}^{N} \frac{\partial}{\partial w}\hat{\delta}(W_i, X_i) \cdot d(W_i) \cdot (X_i - \hat{m}(W_i)), \tag{44}$$

where $\hat{m}(W_i)$ is the NIP kernel regression estimate of $\mathbb{E}[X_i \mid W_i]$. This estimator is similar to the class of weighted average derivative estimators surveyed by Powell (1994) and Newey and McFadden (1994). To see this note that if $X_i - \hat{m}(W_i)$ is removed from (44), then the estimator coincides with an weighted average derivative estimator (with $d(W_i)$ equalling the known weight function). The derivation of $\hat{\beta}^{\text{lc}}$'s asymptotic sampling distribution closely parallels that of weighted average derivatives. However the covariance structure of the estimator, as well as the additional nonparametrically estimated object, $\hat{m}(W_i)$, renders its influence function more complicated (see Theorem 4.4 of Graham, Imbens and Ridder, 2009a).

Implementation issues A precise implementation of the methods described in Graham, Imbens and Ridder (2009a) would prove challenging to the typical empirical researcher. The absence of easily usable software implementing the NIP kernel estimator of Imbens and Ridder (2009) is one barrier. An additional issue is that an analog approach to variance estimation would require non-parametric estimation of many objects. Finally, as in much of the semiparametric literature, the issue of bandwidth selection is left unaddressed.

Nevertheless the simple analog structure of the estimators suggests several natural shortcuts that *may* be appropriate for empirical work (at least experimentally). First $\delta(W_i, X_i)$, $\frac{\partial}{\partial w}\delta(W_i, X_i)$ and $m(W_i)$ may be estimated by local linear regression methods. One can estimate $F_X(X_i)$ by the empirical CDF and $F_W^{-1}(\tau)$ by the τ^{th} sample quantile. With these objects in hand the computation of $\hat{\beta}^{\text{cm}}(\rho,\tau)$, $\hat{\beta}^{\text{pam}}$, $\hat{\beta}^{\text{nam}}$ and $\hat{\beta}^{\text{lc}}$ involves only summation. For this last step it may be advisable to trim observations that are near the boundary of the joint support of W_i and X_i. No nonlinear optimization is involved. Each of these steps can be performed with commercial software. Finally, while Graham, Imbens and Ridder (2009a) provide no formal justification for it, the bootstrap can be used to conduct inference.

4. IDENTIFICATION AND ESTIMATION OF ONE-TO-ONE MATCHING MODELS WHEN MATCH OUTPUT IS UNOBSERVED: EQUILIBRIUM APPROACHES

This section considers what features of the match output function are identified when only match attributes are observed. What does the choice of one's partner alone reveal about preferences and/or match output? As noted above, this question shares important similarities with those which motivated the development of single agent discrete choice models (McFadden, 1974; 1981). The two-sided nature of the matching problem, however, complicates the identification challenge. I emphasize choice in a decentralized market where agents are 'rivals to match' and transfers between agents adjust to clear the market (e.g., Becker, 1973). Conveniently this allows the econometrician to focus directly on match output, as opposed to the separate utility functions of the two agents. This is the case considered by Choo and Siow (2006a,b), which are the key references.

When match output is observed, as was assumed in the previous section, identifying AREs requires identifying (features of) the average match output function (AMF). This is difficult because units may purposely select their match partners; hence the observed 'inputs' in the production function will covary with the unobserved ones. When match output is unobserved, in contrast, the challenge is not to 'correct for' the effects of purposeful matching, but rather to draw inferences directly from it. This requires an explicit behavioral model of partner choice or a 'structural' matching model.[26]

Section 4.1 outlines one such model. Under specific distributional assumptions the model corresponds to the one introduced by Choo and Siow (2006a, b); this point is developed in Section 4.2. Matching is one-to-one. Each firm matches with one worker and vice versa; hence firms are rivals (as are workers). Associated with each match is an unobserved transfer from the firm to the worker. The level of this transfer, which may be negative, is an equilibrium outcome. For concreteness, I will sometimes refer to this transfer as a wage. The equilibrium concept is that of *pairwise stability*: in equilibrium no firm is willing to pay the wage required to match with a different worker and no worker is willing to accept the wage offered by a different firm (e.g., Roth and Sotomayor, 1990).

The matching market consists of a finite number of 'firm' and 'worker' types. However there are a large number of firms and workers of each type. This large market assumption is important. It effectively transforms the matching problem into an applied general equilibrium one.

Section 4.3 considers the identifying content of micro-data pairwise comparisons without distributional assumptions as first proposed by Fox (2009a,b). A key aspect

[26] Structural model of partner choice may also be helpful in settings where match output is observed and, of course, (features of) the matching mechanism are central to any approach to identification.

of his approach is the interpretation of assortativeness in the data, the tendency for, say, high quality teachers to match with high quality schools, as evidence of complementarity. This is also implicit in Choo and Siow (2006a,b).[27,28] Writing down explicit data generating process under which this inference is valid is non-trivial.

As a prelude to nonparametric analysis, Section 4.3 begins by showing that the Choo and Siow (2006a,b) model can be identified and estimated via micro-data pairwise comparisons. This point, while straightforward to show, appears to be new. I term the resulting estimation procedure 'pairwise logit'. Pairwise logit is intriguingly similar to the fixed effects conditional logit estimator for binary choice panel data models (e.g., Chamberlain, 1980). However there are important differences. For example, the pairwise logit estimate is the minimum of a U-process, to which standard M-estimation theory does not directly apply.

The pairwise implications of the CS model are interesting for at least three reasons. First, they highlight that the method interprets local assortativeness as evidence of local complementarity. Second, they provide a way to estimate the model without aggregate data. Third, they show that method relies on the same insight which underlies Fox's (2009a,b) semiparametric approach: pairwise stability implies that if we draw any two pairs of matches at random, then switching match partners should not raise welfare.

In addition to its potential empirical applicability, the pairwise logit estimator suggests natural semiparametric extensions. Section 4.3 explores these extensions under one set of primitive conditions. The main result is given in Theorem 4.1. The primary contribution of Theorem 4.1 is to provide a primitive justification for a specific version of the pairwise method first suggested by Fox (2009a,b). An explicit data generating process, with unobserved heterogeneity on both sides of the market, is specified from which the identifying population restriction is formally derived.

The assumption of transferable utility is not tenable in some settings. Some thoughts on the non-transferable case, as well as a brief survey of existing results, are provided in Section 4.4. Issues of estimation and inference are discussed in Section 4.5.

4.1 A two-sided model of multinomial choice

Consider an assignment where firm i matches with worker j. The firm's profit from such a match is assumed to be

$$\Pi(W_i, X^j, \varepsilon_i) = \delta(W_i, X^j) + \lambda(\varepsilon_i, X^j) - \tau(W_i, X^j), \tag{45}$$

[27] The complementarity measure discussed by Galichon and Salanié (2009) and Siow (2009), for example, is the logarithm of the likelihood ratio measure of dependence (e.g., Lehmann, 1966).

[28] This interpretation of dependence in the data has obvious pitfalls. A simple example illustrates the problem. Say agents on both sides of the market are characterized by two binary attributes only the first of which is observed. These two attributes are positively dependent so that if an agent has the first attribute, she is more likely to have the second and vice versa. Now say there is complementarity in match output between the first pair of attributes and substitutability between the second. If the substitutability in the unobserved agent attribute is strong enough we may observe negative assortative matching on the observed attribute despite the underlying structural complementarity.

where $\delta(W_i,X^j) + \lambda(\varepsilon_i,X^j)$ is the firm's match output and $\mathbb{E}[\lambda(\varepsilon_i, x)] = 0$. This output consists of two parts: (i) a deterministic or average component, $\delta(W_i,X^j)$, and (ii) a firm-specific component, $\lambda(\varepsilon_i,X^j)$. The transfer/wage paid by the firm to the worker is given by $\tau(W_i,X^j)$. Note that the transfer function, $\tau(W_i, X^j)$, depends only on observed firm and worker characteristics. That this is an equilibrium feature of the model will become apparent below (cf., Galichon and Salanié, 2009).

The effect of ε_i is to generate heterogeneity, across observationally identical firms, in the incremental return to matching with a type x' instead of a type x worker:

$$\Pi(w, x', \varepsilon) - \Pi(w, x, \varepsilon) = \delta(w, x') - \delta(w, x) + [\lambda(\varepsilon, x') - \lambda(\varepsilon, x)].$$

Such heterogeneity ensures that the conditional distribution of observed worker type, given observed firm type will be non-degenerate in equilibrium (cf., Galichon and Salanié, 2009; Siow, 2009).

Equation (45) imposes a strong restriction: firm match profits are constant in unobserved worker characteristics, v^j. Put differently neither firm output, or the wage paid, depends on the *particular* worker employed, *only her type matters*. From the perspective of firms, workers, conditional on their type, are homogenous inputs or perfectly substitutable. This assumption will be reasonable in some settings and strain credulity in others. Choo and Siow (2006a,b) and, especially, Chiappori, Salanié, and Weiss (2010) and Galichon and Salanié (2009) discuss this assumption further.

Let $\boldsymbol{\tau}_k = (\tau(w_k, x_1), \ldots, \tau(w_k, x_L))'$ denote the L vector of wages/transfers at which a type $W_i = w_k$ firm can 'hire' each of the L types of workers. Firm's treat these transfers as fixed when matching. Therefore, under the maintained assumption of profit maximization, the type of worker hired by firm i, X_i, is equal to

$$X_i = \underset{x \in \mathcal{X}}{\arg\max} \{\Pi(W_i, x, \varepsilon_i)\},$$

with $\Pi(W_i, x, \varepsilon_i)$ as defined by (45).

Consider an assignment where worker j matches with firm i. The worker's utility from such a match is

$$V(W_i, X^j, v^j) = \tau(W_i, X^j) + \rho(W_i, v^j), \tag{46}$$

where $\rho(W_i, v^j)$ is a utility shifter such that $\mathbb{E}[\rho(w, v^j)] = 0$. Observationally identical workers may rank the desirability of matching with different types of firms differently. That is, a worker's utility is individual-specific, but analogous to firm profits, does not depend on the specific firm at which she works, only its type.

Let $\boldsymbol{\tau}_l = (\tau(w_1, x_l), \ldots, \tau(w_K, x_l))'$ denote the K vector of wages/transfers available to a worker of type $X^j = x_l$ in exchange for matching with each of the K types of firms. Utility maximization implies that worker j will match with a firm of type

$$W^j = \underset{x \in \mathcal{W}}{\arg\max}\, V(w, X^j, v^j)$$

with $V(w, X^j, v^j)$ as defined by (46).

Total match surplus $Y_{(i,j)}$ is equal to the sum of (45) and (46),

$$Y_{(i,j)} = \delta(W_i, X^j) + \lambda(\varepsilon_i, X^j) + \rho(W_i, v^j),$$

which is identical to the restricted match surplus function (20) discussed in Section 3.2 above.[29]

In the absence of an outside option with exogenously-specified utility (i.e., the ability not to match), it is clear that an across the board increase or decrease in equilibrium transfers will leave both firm and worker preferences unchanged. In particular changing transfers to ${\boldsymbol{\tau}^*}_k = \boldsymbol{\tau}_k + t$ leaves type $W_i = w_k$ firms' rankings over worker types unchanged. Likewise $\boldsymbol{\tau}_l^* = \boldsymbol{\tau}_l + t$ leaves type $X^j = x_l$ workers' rankings unchanged. Consequently the equilibrium transfer vector will be non-unique. Therefore I normalize the first element of $\boldsymbol{\tau}_k$ and $\boldsymbol{\tau}_l$ to zero for all firm and worker types. This generates $K + L - 1$ non-redundant normalizations.[30]

Note that the restriction that firms (workers) are indifferent across workers (firms) of the same type implies that in equilibrium the transfer function will vary with W_i and X^j alone. Firms are unwilling to pay a premium for workers with different realizations of v^j, neither are they required to compensate workers for variability in their own realization of ε_i. This means that some firms and workers will earn inframarginal rents in equilibrium.

Associated with each firm is a vector of L productivities: one specific to each of the L types of workers with which it may match. This vector is independently and identically distributed across firms:

$$(\lambda(\varepsilon_i, x_1), \ldots, \lambda(\varepsilon_i, x_L))' \sim F_\lambda. \tag{47}$$

Consistent with Condition 3.1 this distribution is constant in firm type.

Associated with each worker is a vector of K utility shifters: one specific to each of the K types of firms with which it may match:

$$(\rho(w_1, v^j) \ldots, \rho(w_K, v^j))' \sim F_\rho. \tag{48}$$

This vector is independently and identically distributed across all workers.

[29] The interpretation of the two sources of unobserved heterogeneity, $\lambda(\varepsilon_i, X^j)$ and $\rho(W_i, v^j)$, is somewhat more flexible then suggested by the language adopted here (cf., Chiappori, Salanié, and Weiss, 2010).

[30] The introduction of an outside option for each firm and worker type (with an exogenously given utility level), will eliminate this indeterminancy. The parallel with the role of an outside good in discrete choice models of demand is quite close (e.g., Nevo, 2000).

For what follows it will sometimes be convenient, as well as conceptually helpful, to use the abbreviated notation

$$(\lambda_{1i}, \ldots, \lambda_{Li})' \sim F_\lambda, \quad (\rho_1^j, \ldots, \rho_K^j)' \sim F_\rho,$$

where $\lambda_{li} = \lambda(\varepsilon_i, x_l)$ and $\rho_k^j = \rho(w_k, v^j)$.

The equilibrium assignment of workers to firms is determined by the interaction of three primitives of the model: (i) the marginal distributions of firm and worker types, respectively $p = (p_1, \ldots p_K)'$ and $q = (q_1, \ldots q_L)'$, (ii) the distribution functions of the firm productivities and worker utilities, respectively F_λ and F_ρ, and (iii) the production function at each possible (w,x) pair $\delta = (\delta_1', \ldots \delta_K')'$ (where $\delta_k = (\delta(w_k, x_1), \ldots, \delta(w_k, x_L))'$ denotes the 'deterministic' component of the L vector of outputs available to a type k firm).

Consider a type k firm facing the wage/transfer vector $\boldsymbol{\tau}_k$. This vector contains the wage a type k firm must pay in order to match with each of the L types of workers. The total *demand* for matches with type l workers by type k firms is

$$r_{kl}^{\mathrm{D}} = \Pr(X_i = x_l \mid W_i = w_k; \boldsymbol{\tau}_k, \delta_k, F_\lambda) \times p_k. \tag{49}$$

The first term in (49) is the conditional probability that a type k firm's most preferred match is with a type l worker given the vector of prevailing wages $\boldsymbol{\tau}_k$. The precise form of this conditional probability will depend on the joint distribution of unobserved productivities, F_λ. The second term in (49) is the marginal frequency of type k firms in the population. The product of the two terms gives the total demand for k-to-l matches.

Now consider a type l worker facing the wage/transfer vector $\boldsymbol{\tau}_l$. This vector contains the wages available to a type l worker in exchange for matching with each of the K types of firms. The total supply of matches with type k firms by type l workers is

$$r_{kl}^{\mathrm{S}} = \Pr(W^j = w_k \mid X^j = x_l; \boldsymbol{\tau}_l, G_\rho) \times q_l. \tag{50}$$

The first term in (50) is the conditional probability that a type l worker's most preferred match is with a type k firm given the vector of available wages $\boldsymbol{\tau}_l$. The second is the marginal frequency of type l workers in the population. Their product gives the total supply of k-to-l matches.

The $(K-1) \times (L-1)$ non-normalized transfers adjust to equate supply and demand for each of the $K \times L$ types of matches. That is $\boldsymbol{\tau}_k^{\mathrm{eq}}$ and $\boldsymbol{\tau}_l^{\mathrm{eq}}$ adjust to satisfy

$$\Pr(X_i = x_l \mid W_i = w_k; \boldsymbol{\tau}_k^{\mathrm{eq}}, \delta_k, F_\lambda) \times p_k = \Pr(W^j = w_k \mid X^j = x_l; \boldsymbol{\tau}_l^{\mathrm{eq}}, F_\rho) \times q_l, \tag{51}$$

for $k = 1, \ldots, K$ and $l = 1, \ldots, L$. After eliminating the $K + L - 1$ redundant conditions, we are left with $(K-1) \times (L-1)$ equilibrium conditions which pin down the $(K-1) \times (L-1)$ transfers.

Given the equilibrium transfer vectors, the *equilibrium* frequency of k-to-l matches, r_{kl}^{eq}, is given by (49) or (50) after substituting in $\boldsymbol{\tau}_k^{\mathrm{eq}}$ or $\boldsymbol{\tau}_l^{\mathrm{eq}}$.

4.2 Parametric identification of AREs when match output is unobserved

If F_λ and F_ρ belong to parametric families indexed by parameter η, then the analysis of identification is conceptually straightforward (although, as in single agent multinomial choice models, the details may be involved). The problem is one of multinomial choice subject to the $(K - 1) \times (L - 1)$ market clearing conditions (51). The parametric assumptions on F_λ and F_ρ induce specific functional forms for the conditional choice probabilities (49) and (50). Fixing $\theta = (\boldsymbol{\delta}', \eta)'$ one can therefore use (51) to solve for the set of transfers which will clear the market $\boldsymbol{\tau}(\theta)$. This vector is then plugged back into the firm's conditional demand equation. Finally, θ is chosen to align the predicted with the actual match-type 'market shares'.

Specifically, let $G_{kl}^D(\theta, \boldsymbol{\tau})$ and $G_{kl}^S(\theta, \boldsymbol{\tau})$ be the parametric forms for, respectively, type k firms' demand for type l workers and type l workers' supply to type k firms. These forms are induced by the parametric assumptions on F_λ and F_ρ. In 'step 1' we find, fixing θ, the vector of transfers $\boldsymbol{\tau}(\theta)$ which solve the $k = 1, \ldots, K - 1$ and $l = 1, \ldots, L - 1$ market clearing conditions

$$G_{kl}^D(\theta, \boldsymbol{\tau}(\theta)) \times p_k = G_{kl}^S(\theta, \boldsymbol{\tau}(\theta)) \times q_l. \tag{52}$$

In 'step 2' we choose θ such that

$$G_{kl}^D(\theta, \boldsymbol{\tau}(\theta)) \times p_k = r_{kl} \tag{53}$$

for all $k = 1, \ldots, K - 1$ and $l = 1, \ldots, L - 1$. Here r_{kl} denotes the equilibrium or status quo frequency of k-to-l matches (I drop the 'sq' superscript to simplify the notation).

Under conventional distributional assumptions 'demand' $G_{kl}^D(\theta, \boldsymbol{\tau})$ will be strictly decreasing in 'price' τ_{kl} and 'supply' $G_{kl}^S(\theta, \boldsymbol{\tau})$ will be strictly increasing in 'price' τ_{kl}; consequently (52) should be straightforward to solve.[31] However without additional assumptions θ is not point identified. Equation (53) provides only $(K - 1) \times (L - 1)$ equations for $\dim(\theta) = \dim(\boldsymbol{\delta}) + \dim(\eta) = KL + \dim(\eta)$ unknowns. Point identification requires additional assumptions. Two basic options are available. First, we might impose extra structure on the KL match-specific surpluses so that $\boldsymbol{\delta} = \boldsymbol{\delta}(\beta)$ for some low dimensional β. If the support points of W_i and X^j have a natural ordering then such structure can be quite natural (e.g., following from smoothness assumptions on $\delta(w, x)$).[32] A priori restrictions on $\delta(w, x)$ allow for more flexibility in the specification

[31] Depending on the precise distributional assumptions uniqueness issues could arise. Goeree, Holt and Palfrey (2005) give examples of distributional assumptions that lead to choice probabilities which are non-monotone in their indices.

[32] If W_i and X^j are themselves functions of multiple underlying characteristics then separability assumptions could also be imposed on $\delta(w, x)$.

of the joint distributions of unobserved firm productivities and worker preferences (i.e., a higher dimensional η). Even if F_λ and F_ρ are known, additional assumptions will be required in order to identify the KL match-specific surpluses $\boldsymbol{\delta}$. One approach, which may be both useful and empirically relevant, is to allow agents not to match. Introducing an outside option for each firm and worker type, and normalizing its profit/utility to zero, will sometimes allow for the identification of $\boldsymbol{\delta}$.

As an example of a parametric treatment consider the work of Choo and Siow (2006a,b). They study the model outlined above under the additional assumption that $\lambda(\varepsilon_i,x_l)$ and $\rho(w_k,v^j)$ are independently and identically distributed centered Type I extreme value random variables. Galichon and Salanié (2009) relax this assumption by allowing the common scale parameter for the firm heterogeneity to differ from that associated with the worker heterogeneity. Adopting this latter formulation assume that

$$\begin{aligned} F_\lambda(\lambda_1,\ldots,\lambda_L) &= \textstyle\prod_{l=1}^{L} \exp\left(-\exp\left(-\frac{\lambda_l}{\sigma_\lambda}\right)\right) \\ F_\rho(\rho_1,\ldots,\rho_K) &= \textstyle\prod_{k=1}^{K} \exp\left(-\exp\left(-\frac{\rho_k}{\sigma_\rho}\right)\right). \end{aligned} \tag{54}$$

Under this assumption McFadden (1974) shows that there exist closed form expressions for the equilibrium firm demand equations (49) of, letting $\tau_{kl}^{\text{eq}} = \tau_{kl}^{\text{eq}}(w_k,x_l)$ and $\delta_{kl} = \delta(w_k,x_l)$,

$$r_{k1} = p_k \frac{1}{1+\sum_{m=2}^{L}\exp\left(\sigma_\lambda^{-1}[\delta_{km}-\delta_{k1}-\tau_{km}^{\text{eq}}]\right)}, \quad l=1, \tag{55}$$

$$r_{kl} = p_k \frac{\exp\left(\sigma_\lambda^{-1}[\delta_{kl}-\delta_{k1}-\tau_{kl}^{\text{eq}}]\right)}{1+\sum_{m=2}^{L}\exp\left(\sigma_\lambda^{-1}[\delta_{km}-\delta_{k1}-\tau_{km}^{\text{eq}}]\right)}, \quad l=2,\ldots,L \tag{56}$$

for $k = 1, \ldots, K$ (recall that $\tau_{k1} = 0$ by normalization). Taking logarithms of (55) and (56) and subtracting then gives

$$\sigma_\lambda \ln(r_{kl}/r_{k1}) = \delta_{kl} - \delta_{k1} - \tau_{kl}^{\text{eq}}. \tag{57}$$

Similarly the equilibrium worker supply equations (50) take the form

$$r_{1l} = q_l \frac{1}{1 + \sum_{m=2}^{K} \exp\left(\sigma_\rho^{-1}[\tau_{ml}^{\text{eq}}]\right)}, \quad k = 1 \tag{58}$$

$$r_{kl} = q_l \frac{\exp\left(\sigma_\rho^{-1}[\tau_{kl}^{\text{eq}}]\right)}{1 + \sum_{m=2}^{K} \exp\left(\sigma_\rho^{-1}[\tau_{ml}^{\text{eq}}]\right)}, \quad k = 2, \ldots, K, \tag{59}$$

for $l = 1, \ldots, L$ (recall that $\tau_{1l} = 0$ by normalization). Taking logarithms of (58) and (59) and subtracting then gives

$$\sigma_\rho \ln(r_{kl}/r_{1l}) = \tau_{kl}^{\text{eq}}. \tag{60}$$

Finally, adding (57) and (60) yields (cf., Equation (10) of Choo and Siow (2006a)):

$$\sigma_\lambda \ln(r_{kl}/r_{k1}) + \sigma_\rho \ln(r_{kl}/r_{1l}) = \delta_{kl} - \delta_{k1}. \tag{61}$$

Further manipulation then gives (Galichon and Salanié, 2009; Siow, 2009):

$$\ln\left(\frac{r_{KL}}{r_{Kl}}\frac{r_{kl}}{r_{kL}}\right) = \frac{\delta_{KL} - \delta_{Kl} - [\delta_{kL} - \delta_{kl}]}{\sigma_\lambda + \sigma_\rho} = \frac{\phi_{KLkl}}{\sigma_\lambda + \sigma_\rho}, \quad k = 1, \ldots, K-1, \quad l = 1, \ldots, L-1 \tag{62}$$

which identifies the average local complementarity (ALC) between W_i and and X^j up to scale. Knowledge of ϕ_{KLkl} up to scale is sufficient to identify $\beta(R^{\text{a}})$ up to scale (see Section 3.2).

In the CS model the $(K-1) \times (L-1)$ average local complementarity (ALC) parameters are *parametrically just identified* by the $(K-1) \times (L-1)$ non-redundant entries in R^{sq}. Embodied in the setup are a number of strong, a priori, restrictions. To see this consider the case where W_i and X^j equal male and female years of completed schooling. Consider a man with k years of schooling, it seems likely that if his idiosyncratic valuation of women with l years of schooling, λ_{li}, is above average, then so is his valuation of women with $l+1$ years of schooling, λ_{l+1i}. Men who are particularly attracted to college educated women may be similarly attracted to those with graduate degrees. The CS model rules out such correlations in unobserved tastes. Furthermore, in the absence of placing additional structure elsewhere, such correlations are unidentified. Unfortunately, if they are present in the population, the model will generate poor forecasts of, say, the effect of increasing the fraction of women who are college graduates on the equilibrium pattern of marriages. The problem is analogous to McFadden's (1981) well-known Red bus/Blue bus problem.[33]

As noted above if $\boldsymbol{\delta} = \boldsymbol{\delta}(\beta)$ for some low dimensional β, then it will typically be possible to make the parametric assumption on F_λ and F_ρ less restrictive.[34] Methods refined in the empirical industrial organization (IO) literature on differentiated product demand may be helpful in this regard (e.g., Berry, Levinsohn and Pakes, 1995; Nevo, 2000; Ackerberg, Benkard, Berry and Pakes, 2007). In recent work, Chiappori, Salanie and Weiss (2010), have emphasized the identifying power of observing multiple markets where (i) the distribution of agent preferences in the same across markets, but (ii) the distribution of agent types varies.

4.3 Nonparametric identification

Fox (2009a) initiated the study of nonparametric identification in transferable utility matching games when only agent characteristics are observed. In addition to approaching the problem nonparametrically, his results, unlike Choo and Siow (2006a,b), who work with aggregate 'market share' data, rely only on match-level pairwise comparisons. Fox's (2009a) theorems require that a 'rank order property' hold. This property ensures that, across a population of observationally identical markets, assignments which yield more surplus when the stochastic component of match surplus is ignored will be more frequently observed. While this assumption is intuitive, and analogous to those underlying single agent discrete choice models, Fox (2009a) notes that it is difficult to write down data generating processes under which it holds.

A virtue of Choo and Siow's (2006a,b) likelihood-based approach is its complete specification of the data generating process. Unfortunately its heavy reliance on the conditional logit model is unattractive. The discussion in Section 4 clarifies that the CS model is perhaps best viewed as particular specification of a two-sided multinomial discrete choice problem subject to market clearing conditions. Manski (1975) demonstrated semiparametric identification of a single agent multinomial choice model and proposed an associated 'maximum score' estimator (see also Lee, 1995; Matzkin, 2007; Powell and Ruud, 2008). The discrete-choice structure of the CS model suggests that it too may have semiparametric analog.

To explore this possibility this section begins by developing some 'pairwise implications' of the CS model. This leads naturally to a semiparametric approach based on pairwise comparisons similar to those first suggested by Fox (2009a). The valued-added

[33] In recent, pedagogically-oriented work, Imbens (2007b) has re-cast this as the Chez Panise/Lalime's problem. Debreu (1960) is the first published account of this problem.

[34] Galichon and Salanié (2009) consider restrictions of the form $\boldsymbol{\delta} = \boldsymbol{\delta}(\beta)$; however they do not use the resulting extra degrees of freedom to relax the Type I extreme value forms for F_λ and F_ρ. Instead they use the extra restrictions for specification testing.

here, relative to Fox (2009a), is in providing a primitive justification for these comparisons. A limitation is that this justification hinges on agent characteristics being discretely-valued (which allows me to adopt a two-sided multinomial modelling approach). In contrast Fox (2009a), under the maintained rank order property, can accommodate continuously-valued agent characteristics. Fox (2009a) is also able to accommodate situations where matching is many-to-many.

4.3.1 Pairwise logit identification

Consider the subpopulation of type k and m firms that choose to match with either type l or n workers. Likewise consider the subpopulation of type l and n workers that choose to match with either type k or m firms. This defines a conditional analog of the simple 2×2 assignment problem discussed above. Consider two matches, say match i and match j, which are independent random draws from a population of equilibrium matches. Let

$$A_{ij}^{klmn} = \mathbf{1}(W_i \in \{w_k, w_m\})\mathbf{1}(W_j \in \{w_k, w_m\})\mathbf{1}(X_i \in \{x_l, x_n\})\mathbf{1}(X_j \in \{x_l, x_n\}), \quad (63)$$

be a binary indicator for the event that both matches belong to the *klmn sub-allocation* (i.e., belong to the set of k-to-l, k-to-n, m-to-l and m-to-n matches). There are a total of $J = \binom{K}{2} \times \binom{L}{2}$ such sub-allocations.[35] Each randomly sampled pair of matches will belong to at least one sub-allocation. If $W_i \neq W_j$ and $X_i \neq X_j$, an event I will condition on below, then they will belong to a unique sub-allocation.

We can imagine locally reallocating workers across firms within the *klmn* sub-allocation of matches. If we normalize

$$\begin{aligned} r^{klmn} &= \frac{r_{kl}}{r_{kl} + r_{kn} + r_{ml} + r_{mn}} \\ p^{klmn} &= \frac{r_{kl} + r_{kn}}{r_{kl} + r_{kn} + r_{ml} + r_{mn}} \\ q^{klmn} &= \frac{r_{kl} + r_{ml}}{r_{kl} + r_{kn} + r_{ml} + r_{mn}}, \end{aligned} \quad (64)$$

then the set of feasible sub-reallocations is summarized in Table 3. The *klmn* sub-allocation may be made more assortative by increasing r^{klmn}, and less so by decreasing it.

Now, for randomly sampled matches i and j, define

$$S_{ij} = sgn\{(W_i - W_j)(X_i - X_j)\}. \quad (65)$$

[35] Note that this definition allows two sub-allocations to be over-lapping. For example the *KLkl* and *KLmn* sub-allocations overlap.

Table 3 The set of feasible klmn sub-allocations

W\X	$X = x_l$	$X = x_n$	
$W = w_k$	r^{klmn}	$p^{klmn} - r^{klmn}$	p^{klmn}
$W = W_m$	$q^{klmn} - r^{klmn}$	$1 - p^{klmn} - q^{klmn} + r^{klmn}$	$1 - p^{klmn}$
	q^{klmn}	$1 - q^{klmn}$	

If match i and j are assortatively paired, consisting of, for example, one (k, l) and one (K, L) match, then $S_{ij} = 1$. If, in contrast, the two matches are 'integrated' or anti-assortatively paired, consisting of, for example, one (k, L) and one (K, l) match, then $S_{ij} = -1$. If either firm or worker type (or both) are the same across the two drawn matches then $S_{ij} = 0$. In such cases a switch of workers by the two firms would leave the joint distribution of (W_i, X_i) unchanged. This case corresponds to sampling matches which belong to multiple sub-allocations. The above discussion assumes, as is conventional, that the support points of W_i and X^j are ordered in increasing magnitude.

Consider the probability of drawing an assortatively matched pair conditional on (i) the draw being either assortative or anti-assortative and (ii) the pair belonging to the *klmn* sub-allocation. By the definition of conditional probability we have

$$\Pr(S_{ij} = 1 \,|\, S_{ij} \in \{-1, 1\}, A_{ij}^{klmn} = 1) \\ = \frac{r^{klmn}\left(1 - p^{klmn} - q^{klmn} + r^{klmn}\right)}{r^{klmn}\left(1 - p^{klmn} - q^{klmn} + r^{klmn}\right) + \left(p^{klmn} - r^{klmn}\right)\left(q^{klmn} - r^{klmn}\right)}.$$

By equation (62) above the right-hand-side of this expression, under the maintained assumptions of the CS model, is

$$\Pr(S_{ij} = 1 \,|\, S_{ij} \in \{-1, 1\}, A_{ij}^{klmn} = 1) = \frac{\exp\left(\frac{\delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl})}{\sigma_\lambda + \sigma_\rho}\right)}{1 + \exp\left(\frac{\delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl})}{\sigma_\lambda + \sigma_\rho}\right)}.$$

Recall that $\phi_{mnkl} = \delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl})$ equals average local complementarity (ALC) in the *klmn* sub-allocation. Let $\mathbf{A}_{ij}$ be the $J \times 1$ vector of all sub-allocation indicators and ϕ the corresponding set of average local complementarities. We have shown

$$\Pr(S_{ij} = 1 \,|\, S_{ij} \in \{-1, 1\}, \mathbf{A}_{ij}) = \frac{\exp\left((\sigma_\lambda + \sigma_\rho)^{-1}\mathbf{A}_{ij}'\phi\right)}{1 + \exp\left((\sigma_\lambda + \sigma_\rho)^{-1}\mathbf{A}_{ij}'\phi\right)}. \tag{66}$$

Equation (66) has the form a two-period panel data conditional logit probability (e.g., Chamberlain, 1980). This is a consequence of its pairwise formulation and event conditioning (i.e., discarding $S_{ij} = 0$ draws). Insights from the panel data literature on discrete choice will prove useful below (Manski, 1987; Chamberlain, 2010).

Galichon and Salanié (2009) propose the following parameterization of the average match surplus function

$$\delta(w,x) = a(w) + b(x) + e(w,x)'\beta, \tag{67}$$

for $e(w,x)$ a known low dimension vector of basis functions and $a(w)$ and $b(x)$ arbitrary.[36] Define

$$m_{mnkl} = e(w_m, x_n) - e(w_m, x_l) - [e(w_k, x_n) - e(w_k, x_l)],$$

and let $\mathbf{M}$ be the matrix composed of J rows of the form m'_{mnkl}. This gives $\phi = \mathbf{M}\beta$ and hence

$$\Pr(S_{ij} = 1 \mid S_{ij} \in \{-1,1\}, \mathbf{A}_{ij}) = \frac{\exp\left((\sigma_\lambda + \sigma_\rho)^{-1}\mathbf{A}'_{ij}\mathbf{M}\beta\right)}{1 + \exp\left((\sigma_\lambda + \sigma_\rho)^{-1}\mathbf{A}'_{ij}\mathbf{M}\beta\right)}.$$

Now, imposing the scale normalization $\sigma_\lambda + \sigma_\rho = 1$, consider the the criterion function

$$L_N(\beta) = \sum_{i=1}^{N}\sum_{j<i} |S_{ij}| \left\{ S_{ij}\mathbf{A}'_{ij}\mathbf{M}\beta - \ln\left[1 + \exp\left(S_{ij}\mathbf{A}'_{ij}\mathbf{M}\beta\right)\right]\right\}. \tag{68}$$

Assuming that in the population $r_{kl} > 0$ for all k and l the minimizer of (68), the pairwise logit estimate $\hat{\beta}_{PL}$, will be consistent for β. This estimate is the minimizer of a second order U-Statistic. This class of estimators was introduced by Huber (1964). Honoré and Powell (1994) provide distribution theory for minimizers of U-processes.[37] The sampling properties of the minimizer of (68) are outlined in Section 4.5 below.

4.3.2 Nonparametric identification in a 2 × 2 matching market

This section demonstrates that the sign of ϕ is identified in semiparametric analog of the CS model. Fox (2009a,b) also shows that the sign of local complementarity is identified, but under non-primitive assumptions about the data generating process. The derivation given below provides a primitive justification of Fox's (2009a,b) approach.

To keep the analysis simple initially consider a market with just two types of firms and two types of workers (i.e., $W_i \in \{w_k, w_m\}$ and $X^j \in \{x_l, x_n\}$). This allows a

[36] There is a close connection between (67) and models used to parameterize $K \times L$ ordinal contingency tables (e.g., Goodman, 1979).

[37] Honoré and Powell (2005) characterize the large sample properties of minimizers of kernel weighted U-statistics. In the statistics literature Bose (1998, 2002) studies the asymptotic properties of U-statistic minimizers.

demonstration of identification that formally resembles Manski's (1987) extension of maximum score to two period binary choice panel data models.

As in Section 4 firm profits are given by

$$\Pi\left(W_i, X^j, \varepsilon_i\right) = \delta(W_i, X^j) + \lambda(\varepsilon_i, X^j) - \tau(W_i, X^j)$$

and worker utility by

$$V(W_i, X^j, \nu^j) = \tau(W_i, X^j) + \rho(W_i, \nu^j).$$

There are four types of matches in a 2×2 market. To show identification it suffices to consider just two of them. Consider first matches between $W_i = w_m$ firms and type $X^j = x_n$ workers. Firms in such matches, *at the equilibrium vector of transfers*, prefer type n to type l workers. This generates the inequality, first recalling the notation $\delta_{kl} = \delta(w_k, x_l)$, $\tau_{kl} = \tau(w_k, x_l)$, $\lambda_{li} = \lambda(\varepsilon_i, x_l)$ and $\rho_k^j = \rho(w_k, \nu^j)$,

$$\delta_{mn} - \tau_{mn} + \lambda_{ni} \geq \delta_{ml} - \tau_{ml} + \lambda_{li}. \tag{69}$$

The left-hand-side equals firm profits when matched with a type n worker, the right-hand-side profits when matched with a type l worker. Rearranging yields the equivalent expression

$$\lambda_{li} - \lambda_{ni} \leq \delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml}),$$

which says that a firm chooses a type n worker if the systematic 'gains', $\delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml})$, exceed the idiosyncratic 'losses', $\lambda_{li} - \lambda_{ni}$, from doing so. Since transfers adjust such that all type m firms who prefer type n workers may match with one in equilibrium we have

$$F_{\lambda_l - \lambda_n}(\delta_{mn} - \delta_{ml} - (\tau_{mn\cdot} - \tau_{ml})) = r_{mn}/p_m, \tag{70}$$

with $F_{\lambda_l - \lambda_n}(\cdot)$ the (unknown) distribution of $\lambda_{li} - \lambda_{ni}$ and r_{mn}/p_m the fraction of type m firms matching with type n workers in equilibrium.

Now consider the opposite side of the market. For type n workers who choose to match with type m firms we must have

$$\tau_{mn} + \rho_m^{m(i)} \geq \tau_{kn} + \rho_k^{m(i)},$$

so that

$$F_{\rho_k - \rho_m}(\tau_{mn} - \tau_{kn}) = r_{mn}/q_n, \tag{71}$$

with $F_{\rho_k - \rho_m}(\cdot)$ the (unknown) distribution of $\rho_k^j - \rho_m^j$ and r_{mn}/q_n the fraction of type n workers matching with type m firms in equilibrium.

Now consider type k firms who choose to match with type l workers. For such firms we have

$$\delta_{kl} - \tau_{kl} + \lambda_{li} \geq \delta_{kn} - \tau_{kn} + \lambda_{ni},$$

so that

$$1 - F_{\lambda_l - \lambda_n}(\delta_{kn} - \delta_{kl} - (\tau_{kn} - \tau_{kl})) = r_{kl}/p_k. \tag{72}$$

Finally for type l workers who choose to match with type k firms we have

$$\tau_{kl} + \rho_k^{m(i)} \geq \tau_{ml} + \rho_m^{m(i)},$$

so that

$$1 - F_{\rho_k - \rho_m}(\tau_{ml} - \tau_{kl}) = r_{kl}/q_l. \tag{73}$$

Assume that both $F_{\lambda_l - \lambda_n}(\cdot)$ and $F_{\rho_k - \rho_m}(\cdot)$ are strictly increasing on the entire real line with continuous, bounded derivatives. Subtracting the sum of the inverses of (72) and (73) from the sum of the inverses of (70) and (71) yields.

$$\begin{aligned} &F^{-1}_{\lambda_l - \lambda_n}\left(\frac{1 - p_k - q_l + r_{kl}}{1 - p_k}\right) + F_{\rho_k - \rho_m}\left(\frac{1 - p_k - q_l + r_{kl}}{1 - q_l}\right) \\ -&F^{-1}_{\lambda_l - \lambda_n}\left(\frac{p_k - r_{kl}}{p_k}\right) - F^{-1}_{\rho_k - \rho_m}\left(\frac{q_l - r_{kl}}{q_l}\right) = \delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl}), \end{aligned} \tag{74}$$

where the feasibility conditions $p_m = 1 - p_k$, $q_n = 1 - q_l$ and $r_{mn} = 1 - p_k - q_l + r_{kl}$ are also substituted in.

Equation (74) shows that the average local complementarity parameter $\phi_{mnkl} = \delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl})$ can be written in terms of the observed allocation and the unobserved heterogeneity distribution functions. Under the extreme value assumption of Section 4 equation (74) is equivalent to the complementarity measure derived by Galichon and Salanié (2009) and Siow (2009) (see (62) above).

Observe that if $r_{kl} = p_k\, q_l$ the left-hand-side of (74) evaluates to

$$F^{-1}_{\lambda_l - \lambda_n}(1 - q_l) + F^{-1}_{\rho_k - \rho_m} - 1(1 - p_k) - F^{-1}_{\lambda_l - \lambda_n}(1 - q_l) - F^{-1}_{\lambda_l - \lambda m}(1 - p_k) = 0.$$

If the status quo allocation is the random allocation we may conclude that $\phi_{mnkl} = 0$. Differentiating with respect to r_{kl} yields

$$\begin{aligned} &\frac{1}{f_{\lambda_l - \lambda_n}\left(\frac{1 - p_k - q_l + r_{kl}}{1 - p_k}\right)} \frac{1}{1 - p_k} + \frac{1}{f_{\rho_k - \rho_m}\left(\frac{1 - p_k - q_l + r_{kl}}{1 - q_l}\right)} \frac{1}{1 - q_l} \\ &\qquad + \frac{1}{f_{\lambda_l - \lambda_n}\left(\frac{p_k - r_{kl}}{p_k}\right)} \frac{1}{p_k} + \frac{1}{f_{\rho_k - \rho_m}\left(\frac{q_l - r_{kl}}{q_l}\right)} \frac{1}{q_l} > 0. \end{aligned}$$

So that if $r_{kl} > p_k q_l$ we may conclude that $\phi_{mnkl} > 0$ and if $r_{kl} < p_k q_l$ we can conclude the opposite. Summarizing we have

$$sgn\{r_{kl} - p_k q_l\} = sgn\{\phi_{mnkl}\}, \tag{75}$$

so that assortativeness implies complementarity and mixing implies substitutability.

Now consider the median of S_{ij} conditional on it equalling 1 or -1. Since S_{ij} always equals 1 or -1 its median is necessarily one or the other with

$$med(S_{ij} \mid S_{ij} \in \{-1,1\}) = 1 \Leftrightarrow (1 - p_k - q_l + r_{kl})r_{kl} > (1 - p_k)(1 - q_l)$$
$$med(S_{ij} \mid S_{ij} \in \{-1,1\}) = -1 \Leftrightarrow (1 - p_k - q_l + r_{kl})r_{kl} < (1 - p_k)(1 - q_l).$$

Since $(1 - p_k - q_l + r_{kl})r_{kl} > (1 - p_k)(1 - q_l) \Leftrightarrow r_{kl} > p_k q_l$ we have therefore have

$$med(S_{ij} \mid S_{ij} \in \{-1,1\}) = sgn(\phi_{mnkl}).$$

4.3.3 Nonparametric identification in a $K \times L$ matching market

Generalizing the argument outlined above to general $K \times L$ matching markets requires imposing additional structure on the distributions of firm and worker heterogeneity. Theorem 4.1 provides a formal result based on one sufficient set of conditions. The proof shows how monotonicity of the firm demand and worker supply probabilities in their indexes, combined with the assumption that the matching market clears, delivers a conditional quantile restriction that can be used to identify the sign of ϕ_{mnkl}.

Theorem 4.1 (Semiparametric Identification)

Consider the two-sided multinomial discrete choice model described in Section 4.1 with firm profits given by (45) and worker utilities by (46). If

(i) *r_{kl} is known for all $k = 1, \ldots, K$ and $l = 1, \ldots, L$,*

(ii) *$X_i = x_l$ implies that $\Pi(W_i, x_l, \varepsilon_i) \geq \Pi(W_i, x_n, \varepsilon_i)$ for all $n = 1, \ldots, L$ and $W^j = w_k$ implies that $V(w_k, X^j, v^j) \geq V(w_m, X^j, v^j)$ for all $m = 1, \ldots, K$,*

(iii) *at the equilibrium wage schedule all firms hire their preferred type of worker, and all workers are employed by their preferred firm type,*

(iv) $F_\lambda(\lambda_{1i}, \ldots, \lambda_{Li} \mid W_i) = \prod_{l=1}^{L} F_\lambda(\lambda_{li}),\ F_\rho(\rho_1^j, \ldots, \rho_K^j \mid X^j) = \prod_{l=1}^{K} F_\rho(\rho_k^j)$, *and*

(v) *$F_\lambda(\lambda_{li})$ and $F_\rho(\rho_k^j)$ are strictly increasing on the entire real line, with bounded, continuous derivatives, then the sign of average local complementarity (ALC)*

$$\phi_{mnkl} = \delta_{mn} - \delta_{ml} - (\delta_{kn} - \delta_{kl})$$

is identified, for r^{klmn}, p^{klmn} and q^{klmn} as defined in (64) above, by

$$sgn\{r^{klmn} - p^{klmn} q^{klmn}\} = sgn\{\phi_{mnkl}\},$$

for all k, m (with $k \neq m$) and all l, n (with $l \neq n$).

Proof. The proof consists of four parts. First, following Manski (1975), I use (iv) and (v) to show monotonicity of the choice probabilities in the deterministic firm and worker payoffs (so that if, for example, $\delta_{mn} - \tau_{mn} > \delta_{mo} - \tau_{mo}$, then $\Pr(X_i = x_n | W_i = w_m) > \Pr(X_i = x_o | W_i = w_m)$). Second, following Fox (2009c), I show that monotonicity holds within subsets. Parts one and two of the proof are entirely standard and included only for completeness. Let $F_{\lambda_l - \lambda_n}(\cdot \,|\, W_i = w_m, X_i \in \{x_l, x_n\})$ be the distribution function for $\lambda_{li} - \lambda_{ni}$ in the subpopulation of type m firms who choose to match with either type l or n workers in equilibrium and let $F_{\lambda_l - \lambda_n}(\cdot \,|\, W_i = w_k, X_i \in \{x_l, x_n\})$ be the same function for type k firms. The third part of the proof uses monotonicity and market clearing (iii) to show that $F_{\lambda_l - \lambda_n}(\cdot \,|\, W_i = w_m, X_i \in \{x_l, x_n\})$ and $F_{\lambda_l - \lambda_n}(\cdot \,|\, W_i = w_k, X_i \in \{x_l, x_n\})$ cross just once and at that this point of crossing is their $1 - q^{klmn}$ quantile (q^{klmn} is defined by (64) above). Finally, this conditional quantile restriction is then used to show the main result in an adaptation of the simple argument developed for the 2×2 case in the previous section.

Part 1: The conditional probability that a type m firm chooses a type n worker is, using (ii) and (iv), given by

$$\begin{aligned}\Pr(X_i = x_n \,|\, W_i = w_m) &= \Pr(\lambda_o - \lambda_n < \delta_{mn} - \delta_{mo} - (\tau_{mn} - \tau_{mo}), o = 1, \ldots, L, o \neq n) \\ &= \int_{-\infty}^{\infty} \prod_{o=1, o\neq n}^{L} F_\lambda(\lambda_n + \delta_{mn} - \delta_{mo} - (\tau_{mn} - \tau_{mo})) f_\lambda(\lambda_n)\, \mathrm{d}\lambda_n,\end{aligned}$$

so that for all $l \neq n$

$$\begin{aligned}&\Pr(X_i = x_n \,|\, W_i = w_m) - \Pr(X_i = x_l \,|\, W_i = w_m) \\ &= \int_{-\infty}^{\infty} \Bigg[\prod_{o=1, o\neq n}^{L} F_\lambda(\lambda + \delta_{mn} - \delta_{mo} - (\tau_{mn} - \tau_{mo})) \\ &\quad - \prod_{o=1, o\neq l}^{L} F_\lambda(\lambda + \delta_{ml} - \delta_{mo} - (\tau_{ml} - \tau_{mo})) \Bigg] f_\lambda(\lambda) \mathrm{d}\lambda.\end{aligned}$$

This gives, using (v),

$$\Pr(X_i = x_n \,|\, W_i = w_m) \gtreqless \Pr(X_i = x_l \,|\, W_i = w_m) \tag{76}$$

according to whether

$$\delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml}) \gtreqless 0.$$

Part 2: Dividing both sides of (76) by $\Pr(X_i = x_n | \, W_i = w_m) + \Pr(X_i = x_l | W_i = w_m)$ does not change the inequality that so that by the definition of a conditional probability we have

$$\Pr(X_i = x_n \,|\, W_i = w_m, X_i \in \{x_l, x_n\}) \gtreqless \Pr(X_i = x_l \,|\, W_i = w_m, X_i \in \{x_l, x_n\})$$

according to whether $\delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml}) \gtreqless 0$. Replicating the above arguments also gives

$$\Pr(W^j = w_m \mid X^j = x_n, W^j \in \{w_k, w_m\}) \gtreqless \Pr(W^j = w_k \mid X^j = x_n, W^j \in \{w_k, w_m\})$$

according to whether $\tau_{mn} - \tau_{kn} \gtreqless 0$.

Part 3: Note that

$$\Pr(X_i = x_n \mid W_i = w_m, X_i \in \{x_l, x_n\}) = F_{\lambda_l - \lambda_n}(\delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml}) \mid W_i = w_m, X_i \in \{x_l, x_n\})$$
$$\Pr(X_i = x_n \mid W_i = w_k, X_i \in \{x_l, x_n\}) = F_{\lambda_l - \lambda_n}(\delta_{kn} - \delta_{kl} - (\tau_{kn} - \tau_{kl}) \mid W_i = w_k, X_i \in \{x_l, x_n\})$$

Conditional monotonicity implies that these two CDFs cross just once. Furthermore, market clearing, or hypothesis (iii), implies the sub-allocation feasibility condition:

$$\begin{aligned}(1 - p^{klmn})F_{\lambda_l - \lambda_n}(\delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml}) \mid W_i = w_m, X_i \in \{x_l, x_n\}) \\ + p^{klmn} F_{\lambda_l - \lambda_n}(\delta_{kn} - \delta_{kl} - (\tau_{kn} - \tau_{kl}) \mid W_i = w_k, X_i \in \{x_l, x_n\}) = 1 - q^{klmn}.\end{aligned}$$

That is, within the *klmn* suballocation, the 'demand' for matches with type *n* workers equals the available 'supply'. This gives the conditional quantile restriction

$$F^{-1}_{\lambda_l - \lambda_n}(1 - q^{klmn} \mid W_i = w_m, X_i \in \{x_l, x_n\}) = F^{-1}_{\lambda_l - \lambda_n}(1 - q^{klmn} \mid W_i = w_k, X_i \in \{x_l, x_n\}). \tag{77}$$

A parallel argument gives

$$F^{-1}_{\rho_k - \rho_m}(1 - p^{klmn} \mid X^j = x_n, W^j \in \{w_k, w_m\}) = F^{-1}_{\rho_k - \rho_m}(1 - p^{klmn} \mid X^j = x_l, \cdot W^j \in \{w_k, w_m\}). \tag{78}$$

Part 4: Inverting the conditional 'demands' and 'supplies' yields

$$F^{-1}_{\lambda_l - \lambda_n}\left(\frac{1 - p^{klmn} - q^{klmn} + r^{klmn}}{1 - p^{klmn}} \middle| W_i = w_m, X_i \in \{x_l, x_n\}\right) = \delta_{mn} - \delta_{ml} - (\tau_{mn} - \tau_{ml})$$
$$F^{-1}_{\rho_k - \rho_m}\left(\frac{1 - p^{klmn} - q^{klmn} + r^{klmn}}{1 - q^{klmn}} \middle| X^j = x_n, W^j \in \{w_k, w_m\}\right) = \tau_{mn} - \tau_{kn}$$
$$F^{-1}_{\lambda_l - \lambda_L}\left(\frac{p^{klmn} - r^{klmn}}{p^{klmn}} \middle| W_i = w_k, X_i \in \{x_l, x_n\}\right) = \delta_{kn} - \delta_{kl} - (\tau_{kn} - \tau_{kl})$$
$$F^{-1}_{\rho_k - \rho_m}\left(\frac{q^{klmn} - r^{klmn}}{q^{klmn}} \middle| X^j = x_l, W^j \in \{w_k, w_m\}\right) = \tau_{ml} - \tau_{kl}.$$

Following the derivation beginning with Equation (74) in the discussion of the 2 × 2 case gives the result. ■

The assumption of independence across the elements of $F_\lambda(\lambda_1,\ldots,\lambda_L)$ and $F_\rho(\rho_1,\ldots,\rho_K)$ could be relaxed to exchangeability (cf., Goeree, Holt and Palfrey, 2005; Fox, 2009c).

The conclusion of Theorem 4.1 is a discrete analog of Fox's (2009a, Theorem 5.1) derivative-based approach to identifying local complementarity for continuously-valued inputs. The underlying intuition behind the two results coincide: assortativeness suggests complementarity. The value-added of Theorem 4.1 is that it is an implication of an explicitly specified data generating process, whereas Fox's (2009a) result is not primitively justified.

The availability of a large number of each type of firm and worker is essential for the conclusion of Theorem 4.1. Market thickness ensures that equilibrium transfers depend only on firm and worker types. This allows the econometrician to construct functions of the data that are invariant to these transfers. Identification requires the observation of only a single market. In contrast, Fox (2009a) formally considers identification in a population of many small markets. If markets are truly small, with only a few agents on each side, one can, using the linear programming representation of the equilibrium matching and a parametric specification of F_λ and F_ρ, write down a likelihood for the market-level assignment (cf., Fox, 2009b). When markets are medium-sized this approach is less tractable numerically. In such situations the large market result of Theorem 4.1 may be an useful approximation.

Theorem 4.1 generates, recalling the definitions of S_{ij}, $\mathbf{A}_{ij}$, and ϕ given in the discussion of pairwise logit above, the following conditional median restriction

$$med(S_{ij} \mid S_{ij} \in \{-1,1\}, \mathbf{A}_{ij}) = sgn(\mathbf{A}_{ij}'\phi). \tag{79}$$

Note that (79) has the form of the conditional median restriction derived by Manski (1987) in the context of a two period binary choice panel data model. In the absence of additional structure only the signs of the ALC parameters are identified. This loss of point identification relative to the pairwise logit case is intriguingly analogous to Chamberlain's (2010) identification analysis for binary choice panel data.

If we assume that $\phi = \mathbf{M}\beta$, then (79) suggests choosing $\hat{\beta}_{MRC}$ to maximize the rank correlation criterion suggested by Han (1987):

$$L_N(\beta) = \sum_{i=1}^{N}\sum_{j<i} sgn(\mathbf{A}_{ij}'\mathbf{M}\beta)S_{ij}. \tag{80}$$

This criterion was first advocated by Fox (2009a,b) and Fox and Bajari (2009) in the matching context.

Since $\mathbf{A}_{ij}'\mathbf{M}$ is discretely-valued $\hat{\beta}_{MRC}$ will be set-valued and this will remain true as N grows large. Consequently β is only set identified, however if its dimension is small relative to ϕ, then the identified set may be quite small (see Cavanagh and Sherman, 1998, Section 5).

4.4 Identification in one-to-one matching markets without transfers

In some contexts it may be difficult for match partners to make transfers to one another. For example the institutional structure of the teacher labor market in New York limits the amount of variation in wages across schools (Loeb, Boyd, Lanford and Wyckoff, 2003). Theorists, starting with the seminal paper by Gale and Shapley (1962), have extensively studied two-sided matching problems without transfers (e.g., Roth and Sotomayor, 1990). Little econometric work on these models has been undertaken.

In a change of notation assume that firm utility is given by

$$U(W_i, X^j, \varepsilon_i) = \vartheta(W_i, X^j) + \lambda(\varepsilon_i, X^j),$$

and worker utility by

$$V(W_i, X^j, v^j) = \phi(W_i, X^j) + \rho(W_i, v^j).$$

Consider two matches, i and j, that are assortatively configured. That is matches i and j consist of, respectively type w and w' firms and x and x' workers with $w < w'$ and $x < x'$. For this configuration to be stable we require that either the firm in match i or the worker in match j (or both) prefer the status quo (i.e., $U(w,x,\varepsilon_i) > U(w,x',\varepsilon_i)$ and/or $V(w',x',v^j) > V(w,x',v^j)$). If this were not the case then this pair could block the assignment by leaving their partners and forming a new match. Similarly stability requires that either the worker in match i or the firm in match j (or both) prefer the status quo (i.e., $U(w',x',\varepsilon_j) > U(w',x,\varepsilon_j)$ and/or $V(w,x,v^i) > V(w',x,v^i)$).[38]

These stability conditions, which rule out so called *blocking pairs*, are considerably more complicated than those needed when utility is transferable. One implication of the absence of transfers is that the relationship between complementarity and assortativeness is weakened (cf., Becker and Murphy, 2000). Consider the case where $\vartheta(w, x) = \vartheta_w(w) + \vartheta_x(x)$ and $\phi(w,x) = \phi_w(w) + \phi_x(x)$ such that there is no complementarity. If $\vartheta_x(x') > \vartheta_x(x)$ and $\phi_w(w') > \phi_w(w)$ agents will nevertheless assortatively match (assuming, as maintained above, that $\lambda(\varepsilon_i,x')$ and $\lambda(\varepsilon_i,x)$ are identically distributed and similarly for $\rho(w',v^j)$ and $\rho(w,v^j)$).

When agents make transfers to one another the equilibrium assignment is (i) generically unique and (ii) surplus maximizing.[39] In the absence of transfers neither of these two conditions typically holds. Multiplicity of equilibria complicate empirical modelling. These considerations suggests that the recovery of agent preferences from match characteristics alone is likely to be even more difficult than in the case with transfers.

[38] Stability also requires that each matched agent prefer their assignment to the always available alternative of not matching at all. For simplicity assume that this condition holds in what follows.

[39] If externalities are present, as in Baccara, Imrohoroglu, Wilson and Yariv (2009), then multiple equilibria are possible even when transfers between agents are allowed.

While identification in two-sided matching problems without transfers has not been formally studied, several papers have implemented different estimation procedures. These papers make various simplifying assumptions. Loeb, Boyd, Lankford and Wyckoff (2003) rule out multiplicity by assuming the status quo assignment is the product of the Gale and Shapley (1962) deferred acceptance algorithm (with firms proposing). Gordon and Knight (2009), in contrast, restrict preferences to ensure uniqueness. Sørensen (2007) and Logan, Hoff and Newton (2008) use Bayesian methods. The latter paper attempts to sidestep the issue of multiplicity by choosing the distribution of firm and worker utilities to maximize the probability that the observed assignment is stable. Unfortunately it seems likely that 'the' maximizing distribution of preferences is not unique, particularly if the matching market is small.

4.5 Estimation on the basis of match characteristics alone

Variants of the simple two-sided conditional logit model outlined in Section 4.2 underlie a growing body of empirical work. In an early application Dagsvik, Brunborg and Flaaten (2001) fit the model with population register data from Norway. Their point estimates suggest substantial decline in the net returns to marriage over the 1986 to 1994 period. Chiappori, Salanié and Weiss (2010) use another variant of the model to study changes in the 'returns' to education on the U.S. marriage market since the 1970s. Siow (2008) studies the effects of sex ratio imbalances generated by a famine associated with the 'Great Leap Forward' on the marriage market in Sichuan, China.

These applications notwithstanding, a systematic approach to inference in the CS model has yet to be developed. Fortunately the close connection between the model and a $K \times L$ contingency table, suggests that the development of the required estimation and inference theory is likely to be straightforward. Indeed some of the required results are provided in Siow (2009) and Galichon and Salanié (2009). Some additional results, based on its pairwise logit representation, are given below.

When firm and worker preferences are not parametrically specified an estimator based on Theorem 4.1 can be used. This leads to a simple rank regression estimator. The asymptotic properties of rank regression are well-known (Han, 1987; Cavanagh and Sherman, 1998). The lack of point identification in the matching context results in some inferential challenges. These may be solved using methods recently developed for models defined by moment inequalities (e.g., Rosen, 2008; Andrews and Soares, 2010; Romano and Shaikh, 2010).

4.5.1 Testing for supermodularity

Siow (2009) exploits methods developed for contingency table analysis to test for super-modularity of the match surplus function. Assessing this hypothesis has been a focus of empirical research since Becker (1973). Equation (62) shows that the two-sided conditional logit model equates average local complementarity (ALC) with local dependence as measured by the local log-odds ratio:

$$\ln\left(\frac{r_{k+1l+1}}{r_{k+1l}}\frac{r_{kl}}{r_{kl+1}}\right)=\frac{\delta_{k+1l+1}-\delta_{k+1l}-[\delta_{kl+1}-\delta_{kl}]}{\sigma_\lambda+\sigma_\rho},\quad k=1,\ldots,K-1,\ l=1,\ldots,L-1.$$

Positivity of all $(K-1)$ $(L-1)$ of these local log odds ratios implies that $\delta(w,x)$ exhibits 'increasing differences' or is supermodular (Topkis, 1998). Assessing the supermodularity hypothesis therefore corresponds to a multivariate one-sided testing problem of the type first studied by Kudo (1963).[40] The supermodularity null corresponds to

$$H_0:\delta_{k+1l+1}-\delta_{k+1l}-[\delta_{kl+1}-\delta_{kl}]>0,\quad k=1,\ldots,K-1,\ l=1,\ldots,L-1,$$

with the alternative that the inequality is weak or reversed for at least one k and l pair. Siow (2009) notes that this null is formally equivalent to testing whether a $K\times L$ contingency table is totally positive of order 2 (TP2) (cf., Douglas et al. 1990):

$$H_0:\ln(r_{k+1l+1})-\ln(r_{k+1l})-[\ln(r_{kl+1})-\ln(r_{kl})]>0,\ k=1,\ldots,K-1,\ l=1,\ldots,L-1. \tag{81}$$

Siow (2009) conceptualizes his data as a N draw random draws from a multinomial population with the probabilities for each match type given by the $K\times L$ assignment matrix R_{WX} introduced in Section 3.1 above. This allows him to form a likelihood ratio statistic for the supermodularity null. Unfortunately the sampling distribution of this statistic is not χ^2 (it follows a 'chi-bar-square' distribution). Dardanoni and Forcina (1998) describe how to compute critical values.

4.5.2 Parametric modeling of average match surplus

Recall the $\phi=\mathbf{M}\beta$ parameterization for the ALC terms proposed by Galichon and Salanié (2009) and discussed in Section 4.3 above. From (62) we have, after imposing the scale normalization $\sigma_\lambda+\sigma_\rho=1$, the equality

$$\ln\left(\frac{r_{mn}}{r_{ml}}\frac{r_{kl}}{r_{kn}}\right)=\phi_{mnkl}.$$

The left-hand-side of this expression is consistently estimable from a random sample of matches (with a asymptotic sampling variance that is also consistently estimable). Galichon and Salanié (2009) then suggest estimating the structural parameters β in a second step by minimum distance (e.g., Chamberlain 1982, 1984).[41] To be precise

[40] The Silvapulle and Sen (2005) monograph summarizes the extensive statistics literature on this, and related, testing problems.

[41] Section 5.2 of their paper also introduces an alternative moment matching estimator.

let $\hat{\phi}$ be vector of all estimated local odds ratios and $\hat{V}$ an estimate of their asymptotic sampling variance. An efficient minimum distance estimator is

$$\hat{\beta} = \arg\min_{\beta}\left(\hat{\phi} - M\beta\right)'\hat{V}^{-1}\left(\hat{\phi} - M\beta\right) = \left(M'\hat{V}^{-1}M\right)^{-1}\left(M'\hat{V}^{-1}\hat{\phi}\right).$$

4.5.3 Pairwise logit estimation

An alternative to minimum distance estimation is the pairwise logit procedure introduced in Section 4.3. The 'one step' nature of the pairwise logit procedure is attractive, as is its direct connection to the pairwise stability concept emphasized by Fox (2009a,b). In very large datasets, however, the minimum distance approach may be preferable for computational reasons (since it requires no nonlinear optimization). A comparison of the asymptotic properties of the two procedures is beyond the scope of this chapter.

Recall the suggested criterion function

$$L_N(\beta) = \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j<i} | S_{ij} | \left\{S_{ij}\mathbf{A}_{ij}'M\beta - \ln\left[1 + \exp\left(S_{ij}\mathbf{A}_{ij}'M\beta\right)\right]\right\}. \tag{82}$$

Let

$$s(Z_i,\beta_0) = \mathbb{E}\left[| S_{ij} | M'\mathbf{A}_{ij}\left\{\mathbf{1}(S_{ij}=1) - \frac{\exp\left(\mathbf{A}_{ij}'M\beta_0\right)}{1+\exp\left(\mathbf{A}_{ij}'M\beta_0\right)}\right\}\middle| W_i, X_i\right],$$

with $\mathbf{\Omega}_0 = \mathbb{E}[s(Z_i,\beta_0)s(Z_i,\beta_0)']$ and

$$\mathbf{\Gamma}_0 = \lim_{N\to\infty} M'\left\{\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j<i} | S_{ij} | \frac{\exp\left(\mathbf{A}_{ij}'M\beta_0\right)}{\left[1+\exp\left(\mathbf{A}_{ij}'M\beta_0\right)\right]^2}\mathbf{A}_{ij}\mathbf{A}_{ij}'\right\}M.$$

The results of Honoré and Powell (1994, 2005) yield a large sample distribution for the minimizer of (82) equal to

$$\sqrt{N}\left(\hat{\beta}_{PL} - \beta_0\right) \to N(0,\Lambda_0), \quad \Lambda_0 = 4\mathbf{\Gamma}_0^{-1}\mathbf{\Omega}_0\Gamma_0^{-1}. \tag{83}$$

This limiting variance may be estimated by $\hat{\Lambda} = 4\hat{\mathbf{\Gamma}}^{-1}\hat{\mathbf{\Omega}}\hat{\mathbf{\Gamma}}^{-1}$ with

$$\hat{\mathbf{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} s_N\left(Z_i,\hat{\beta}_{PL}\right)s_N\left(Z_i,\hat{\beta}_{PL}\right)',$$

for

$$s_N(Z_i,\beta) = M'\left\{\frac{1}{N-1}\sum_{j=1,j\neq i}^{N} | S_{ij} | \mathbf{A}_{ij}\left\{\mathbf{1}(S_{ij}=1) - \frac{\exp\left(\mathbf{A}_{ij}'M\beta\right)}{1+\exp\left(\mathbf{A}_{ij}'M\beta\right)}\right\}\right\},$$

and

$$\hat{\Gamma} = -M'\left\{\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j<i}|S_{ij}|\frac{\exp\left(\mathbf{A}_{ij}'M\hat{\beta}_{PL}\right)}{\left[1+\exp\left(\mathbf{A}_{ij}'M\hat{\beta}_{PL}\right)\right]^2}\mathbf{A}_{ij}\mathbf{A}_{ij}'\right\}M.$$

Formulating primitive regularity conditions under which (83) holds is beyond the scope of this chapter. However to get a feel for the small sample properties of the method Tables 4 and 5 summarize the results of a simple Monte Carlo experiment. I consider two designs. In both cases $K = L = 3$. In the first design the ALCs take the form

$$\phi_{mnkl} = \beta_1(w_m - w_k)(x_n - x_l),$$

with $\beta_1 = 1$. This implies that $\delta(w,x)$ is supermodular. As a result the equilibrium assignment, reported in Panel B of Table 4, is highly assortative.

In the second design I set

$$\phi_{mnkl} = \beta_1(w_m - w_k)(x_n - x_l) + \beta_2\{1(w_m = 3, x_n = 3) - 1(w_m = 3, x_l = 3) \\ -[1(w_k = 3, x_n = 3) - 1(w_k = 3, x_l = 3)]\},$$

with $\beta_1 = 1$ and $\beta_2 = -2$. This induces a more complicated equilibrium assignment. The upper-left-hand 2×2 portion of the assignment matrix is assortatively matched, while the lower-right-hand 2×2 portion is anti-assortatively matched. Other features of the two designs are summarized in Panel A of Table 4.

For both designs I draw 1000 samples of 1000 matches each from the equilibrium assignment distributions. For each draw I estimate β by pairwise logit, form an estimated standard error, and a 95% asymptotic confidence interval. Table 5 summarizes the results. Across both designs the pairwise logit estimates are, up to simulation error, mean and median unbiased. Their Monte Carlo sampling distributions are also well-approximated by the asymptotic distribution theory sketched above.

4.5.4 Pairwise maximum score estimation

Fox (2009b) studies semiparametric estimation of matching models. His criterion function is similar to the one suggested by Theorem 4.1 above. While set identification is not generic in his set-up, it can occur. For this reason he recommends the use of subsampling methods for conducting inference (e.g., Delgado, Rodríguez-Poo and Wolf, 2001; Romano and Shaikh, 2010). Fox and Bajari (2009) provide an empirical illustration.

5. SEGREGATION IN THE PRESENCE OF SOCIAL SPILLOVERS

Debates about the social costs and benefits of ‘segregation’ are present in many areas of social policy. As a famous example consider the effect of racial segregation in schools on academic achievement. Coleman et al. (1966), in research which helped initiate

Table 4 Pairwise logit Monte Carlo designs

Panel A: Parameterizations of Monte Carlo Designs

	$a(w)$	$b(x)$	$e(w,x)$	$e_2(w,x)$	β_1	β_2	(p_1, p_2, p_3)	(q_1, q_2, q_3)
Design 1	w	x	$w \cdot x$	–	1	–	(1/3, 1/3, 1/3)	(1/3, 1/3, 1/3)
Design 2	w	x	$w \cdot x$	$1(w = 3, x = 3)$	1	−2	(1/3, 1/3, 1/3)	(1/3, 1/3, 1/3)

Panel B: Average Match Surplus & Equilibrium Assignment

Design 1

Average Match Surplus				Equilibrium Assignment			
$W \backslash X$	1	2	3	$W \backslash X$	1	2	3
1	3	5	7	1	0.205	0.100	0.028
2	5	8	11	2	0.100	0.133	0.100
3	7	11	15	3	0.028	0.100	0.205

Design 2

Average Match Surplus				Equilibrium Assignment			
$W \backslash X$	1	2	3	$W \backslash X$	1	2	3
1	1	2	3	1	0.198	0.080	0.055
2	2	4	5	2	0.080	0.088	0.165
3	3	6	7	3	0.055	0.165	0.113

NOTES: Panel A summarizes the two data generating processes. Panel B reports the average match surplus for each w and x combination as well as the equilibrium assignment, R_{WX}^{sq}. The equilibrium assignments were computed using the algorithm given in Section 6 of Galichon and Salanié (2009). These allocations are then checked against equation (62) above.

Table 5 Monte Carlo results for Pairwise Logit estimator

	(1) Mean $\hat{\beta}^{(m)}$	(2) Median $\hat{\beta}^{(m)}$	(3) Std. Dev. $\hat{\beta}^{(m)}$	(4) Mean $ase(\hat{\beta}^{(m)})$	(5) Median $ase(\hat{\beta}^{(m)})$	(6) Coverage (nom. 0.95)
Design 1						
$\beta_1(=1)$	0.998	0.998	0.0667	0.0669	0.0667	0.9480
Design 2						
$\beta_1(=1)$	1.002	1.001	0.0723	0.0753	0.0751	0.958
$\beta_2(=-2)$	−2.002	−2.010	0.1939	0.2002	0.2001	0.956

NOTES: Monte Carlo results based on 1,000 samples of size $N = 1,000$ drawn from the two equilibrium assignments listed in Table 4. Columns 1 through 3 report the mean, median and standard deviation of $\hat{\beta}^{(m)}$ across the Monte Carlo replications. Columns 4 and 5 report the mean and median estimated asymptotic standard error of $\hat{\beta}^{(m)}$. Column 6 reports the actual coverage of an asymptotic 95% confidence interval.

widespread court-ordered desegregation in the 1970s, argued that racial isolation lowered the academic achievement of black students. Despite the substantial body of subsequent social science research, there remains widespread disagreement about the effects of segregation in schools. The absence of a consensus opinion among social scientists is, at least partially, due to methodological difficulties (Schofield, 1995).

This section outlines a framework for measuring the 'equity and efficiency' implications of segregation in the presence of social spillovers based on Graham, Imbens and Ridder (2009b). The applied theory literature on segregation in the presence of social spillovers is rich (see the Epple and Romano chapter in this Handbook for a review). As in other successful applications of economic theory to real world problems, much of this work is highly stylized. For example Benabou (1996) models agents as binary-typed but otherwise identical (cf., Becker and Murphy, 2000). While the resulting analysis is insightful and elegant, in particular allowing for a sharp characterization of the laissez faire assignment against which the planner's solution may be compared, it is not obvious how to empirically evaluate it.

Section 5.1 extends the basic set-up employed by Benabou (1996) to include unobserved individual- and neighborhood-level heterogeneity. Sections 5.2 and 5.3 then outline two sets of estimands. The first measure the effects of global reallocations. The second measure the effects of reallocations which only slightly perturb the status quo. Some parametric examples are explored in Section 5.4.

5.1 Population setup

Consider a population of individuals ('students') indexed by $i \in I$. Individuals are binary-typed, $T_i \in \{0, 1\}$, and heterogenous in unobserved ability, A_i. As in Section 2 I maintain an inclusive definition of type such that T_i and A_i are independent.

A second population of locations ('classrooms') indexed by $c \in C$ also exists. These locations are where reference groups 'reside'. Locations are also heterogenous with their unobserved quality given by U_c. For example, teacher quality might vary across classrooms.

Let $G_i = c$ if individual i is assigned to location c. To avoid double-subscripting let, in an abuse of notation, $U_i = U_{G_i}$ denote the quality of individual i's location of residence. The assignment vector $\mathbf{G} = (G_1, \ldots, G_I)'$ completely summarizes the population assignment of individuals to groups. An individual's peer or reference group consists of all other individuals who reside in her location or the index set:

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

An individual's overall neighborhood environment is completely characterized by (i) the types and abilities of her peers and (ii) her location's quality. Let $\underline{T}_{p(i)}, \underline{A}_{p(i)}$ be the vectors of individual i's peers' types and abilities. Without loss of generality assume that peers are sorted such that low types come first, followed by high types, so that $\underline{A}_{p(i)} = (\underline{A}^L_{p(i)}, \underline{A}^H_{p(i)})$, with $\underline{A}^L_{p(i)}$ the ability vector for i's low type peers and $\underline{A}^H_{p(i)}$ for her high type peers. Collecting terms let $Q_i = (\underline{T}_{p(i)}, \underline{A}_{p(i)}, U_i)$ be an individual's overall 'neighborhood quality'. Observe that there are two sources of variation in unobserved neighborhood quality: (i) that due to variation in the ability structure of peers and (ii) that due to variation in location-specific characteristics.

Graham, Imbens and Ridder (2009b) posit the existence of an individual-specific allocation response function

$$Y_i(\mathbf{g}), \quad \mathbf{g} \in \mathcal{G}, \tag{84}$$

where $\mathcal{G}$ denotes the set of all feasible assignments; $Y_i(\mathbf{g})$ gives the potential outcome for individual i associated with allocation $\mathbf{g} \in \mathcal{G}$. Let $\mathbf{G}$ denote the observed assignment, then the observed outcome Y_i coincides with $Y_i(\mathbf{G})$. To make further progress Graham, Imbens and Ridder (2009b) impose two additional restrictions on $Y_i(\mathbf{g})$. First they rule out spillovers across groups. Second they assume peers of the same type are exchangeable within groups (i.e., equally influential).

The first assumption implies that if the two assignments $\mathbf{g}$ and $\tilde{\mathbf{g}}$ are such that $q_i = \tilde{q}_i$, then $Y_i(\mathbf{g}) = Y_i(\tilde{\mathbf{g}})$. This implies that the allocation response function varies with q_i alone:

$$Y_i(\mathbf{g}) = Y_i(q_i) = Y_i(\underline{t}_{p(i)}, \underline{a}_{p(i)}, u_i). \tag{85}$$

The second assumption implies that

$$Y_i(q_i) = Y_i(s_{-i}, \tau_{N_L}(\underline{a}^L_{p(i)}), \tau_{N_H}(\underline{a}^H_{p(i)}), u_i), \tag{86}$$

where $N = N_L + N_H$ is total group size, $s_{-i} = \frac{1}{N-1}\sum_{j\in p(i)} T_j$ is the fraction of high type peers, $\tau_{N_L}\left(\underline{a}^L_{p(i)}\right)$ is the vector of the first N_L elementary symmetric polynomials on $\underline{a}^L_{p(i)}$, and $\tau_{N_H}\left(\underline{a}^H_{p(i)}\right)$ is defined analogously.[42]

Equation (86) indicates the allocation response is an individual-specific function of peers' types, peers' abilities and location quality. There are three distinct sources of unobserved heterogeneity in this setup: (i) own ability, (ii) peers' ability and (iii) locational quality. This heterogeneity arises from two distinct populations: that of individuals and that of locations. This represents a substantial complication over a conventional single agent econometric model. Identification arguments in such a setting will necessarily involve restrictions on the conditional distribution of the three sources of unobserved heterogeneity given observables.

To keep what follows simple I will assume a version of double randomization holds. To understand the required condition consider a social planner who must assign individuals to locations. Assignment is done on the basis of the individual-level binary characteristic. No additional individual- or location-level characteristics are used by the planner. For simplicity assume that each location accommodates the same number of individuals. Since the planner only acts on knowledge of T_i we may assume that each high type individual in a group is an independent random draw from the subpopulation of high types (and likewise for low types). Groups, so formed, are randomly assigned to a location.

Under a doubly randomized assignment mechanism individual i's expected outcome given assignment to a group where s_{-i} of her peers' are high types is

$$Y_i^e(S_{-i}) = \int\left[\int\cdots\int Y_i\left(s_{-i}, \tau_{N_L}\left(\underline{a}^L_{p(i)}\right), \tau_{N_H}\left(\underline{a}^H_{p(i)}\right), u_i\right)\prod_{j\in p(i)} f_A(a_j)\,\mathrm{d}a_j\right] f_U(u_i)\,\mathrm{d}u_i. \quad (87)$$

Note that $Y_i^e(s_{-i})$ is an expectation over a product of marginals. This is because double randomization ensures independence between own and peer ability and own ability and location quality. Loosely speaking it rules out 'sorting' and 'matching' on unobservables. If the planner is constrained, either informationally or institutionally, to implement only double randomized allocations, then she only requires knowledge of the distribution of $Y_i^e(s_{-i})$.

Restricting counterfactual assignments to be doubly randomized is reasonable; an econometrician cannot do social planning calculations if knowledge of individual ability and/or location quality is required. Assuming the status quo assignment is doubly randomized is harder to motivate (outside of the important special case of experimental

[42] Allowing for multiple group-sizes is straightforward but ignored here in order to keep the notation simple.

settings). Much of what follows can be extended to settings where double randomization does not hold, at least if auxiliary restrictions on the production technology are also imposed. These extensions are important for empirical work, but substantially complicate both the notation and analysis (see Graham, Imbens and Ridder, 2009b). Here I wish to focus on a more basic issue: that of measurement. In particular how one can define estimands that give statistical content to the various 'equity versus efficiency' questions that typically arise when considering desegregation policies?

5.2 Global reallocations: the social planner's problem

Let $m_H(s)$ denote the expected outcome for a high type individual in a group with composition s, *when groups are formed according to the doubly randomized mechanism*. Let $m_L(s)$ be the corresponding expected outcome for low types. If $m_L^*(s_{-i}) = \mathbb{E}[Y_i^e(s_{-i}) \mid T_i = 0]$ and $m_H^*(s_{-i}) = \mathbb{E}[Y_i^e(s_{-i}) \mid T_i = 1]$ with $Y_i^e(s_{-i})$ given by (87) above, then

$$m_L(s) = m_L^*\left(\frac{sN}{N-1}\right), \quad m_L(s) = m_L^*\left(\frac{sN-1}{N-1}\right).$$

The composition weighted average

$$m(s) = sm_H(s) + (1-s)m_L(s), \tag{88}$$

gives the expected average outcome, irrespective of type, in a group of composition s. If the status quo is doubly randomized then

$$\mathbb{E}[Y_i \mid S_i = s, T_i = 1] = m_H(s), \quad \mathbb{E}[Y_i \mid S_i = s, T_i = 0] = m_L(s). \tag{89}$$

Equation (89) is intuitive, perhaps even obvious, under doubly randomized assignment, but showing the equality holds formally requires some work. This is because $m_H(s)$ and $m_L(s)$ are averages over the products of several marginal distributions: one ability distribution for *each* group member and one locational quality distribution (cf., Graham, Imbens and Ridder, 2009b). In what follows I assume that (89) holds. In practice other ways of identifying $m_H(s)$ and $m_L(s)$, which might involve introducing separability assumptions of the type explored in the context of two-sided matching models above, may be important.

Now consider a planner who, given knowledge of $m(s)$ and the population frequency of high types, p_H, chooses a distribution of group composition, $F_S(s)$, to maximize the average outcome:

$$\max_{F_S(\cdot)} \int m(s) f_S(s) \mathrm{d}s, \tag{90}$$

subject to the feasibility constraint

$$\int s f_S(s) \mathrm{d}s = p_H. \tag{91}$$

The problem defined by (90) and (91) is one of functional optimization. Although in general such problems are quite difficult to solve, Graham, Imbens and Ridder (2009b) show that the above problem has a simple solution. Before considering their solution it is helpful to underscore why the planner's problem is, at least on the surface, a challenging one. Consider the case where both $m_H(s)$ and $m_L(s)$ are constant in s. This corresponds to a complete absence of social spillovers. In this case we have the maximand in (90) equal to

$$\int [sm_H + (1-s)m_L\,] f_S\,(s)\mathrm{d}s = p_H m_H + (1-p_H)m_L,$$

where the equality follows by substituting in the feasibility constraint (91). In the absence of social spillovers the average outcome is invariant across all feasible assignments. For the assignment problem to be interesting we require the presence of social spillovers. The presence of such spillovers also makes the planner's problem non-trivial. This is because the form of the social spillover is left nonparametric.

It is helpful to begin by analyzing the planner's problem under two special cases: (i) $m(s)$ is globally concave and (ii) $m(s)$ is globally convex (i.e., $\nabla_{ss}m(s)$ is respectively less than and greater than zero for all $s \in [0, 1]$). These two cases are the focus of an extensive applied theory literature on multi-community models (e.g., Benabou, 1993, 1996; Becker and Murphy, 2000; de Bartolome, 1990; Durlauf 1996a,b, 2004; Epple and Romano, 1998; Fernández, 2003).

When $m(s)$ is globally concave Jensen's inequality implies that

$$m(p_H) \geq \mathbb{E}_{F_S}[m(S)],$$

for any feasible $F_S(\cdot)$. Since the restriction holds with equality for the degenerate distribution concentrated at p_H, corresponding to the perfectly integrated assignment, global concavity of $m(s)$ implies that integration maximizes the average outcome.

If $m(s)$ is globally convex, then a mean value expansion and the feasibility constraint (91) gives

$$\mathbb{E}_{F_S}[m(S)] = m(p_H) + \mathbb{E}_{F_S}[\nabla_{ss}m(\bar{S})(S-p_H)^2],$$

for $\bar{S}$ an intermediate value between S and p_H. Convexity of $m(s)$ yields the upper and lower bounds

$$m(p_H) + \left[\min_{\bar{s}\in[0,1]} \nabla_{ss}m(\bar{s})\right] \cdot \mathbb{V}_{F_S}(S) \leq \mathbb{E}_{F_S}[m(S)] \leq m(p_H) + \left[\max_{\bar{s}\in[0,1]} \nabla_{ss}m(\bar{s})\right] \cdot \mathbb{V}_{F_S}(S).$$

These bounds are maximized at the most dispersed distribution on [0, 1] with mean p_H. This distribution, which corresponds to the perfectly segregated assignment, places a mass of p_H on one and a mass of $1 - p_H$ on zero. Global convexity of $m(s)$ therefore implies that segregation maximizes the average outcome. Figure 1 illustrates the planner's solution when $m(s)$ is convex or concave.

The difficult case corresponds to situations where $\nabla_{ss}m(s)$ may change signs on [0, 1]; that is where $m(s)$ is neither concave or convex. Let $M(s)$ be the concave envelope of $m(s)$. Formally $M(s)$ is a function whose truncated lower epigraph coincides with the convex hull of the truncated lower epigraph of $m(s)$ (e.g., Rockafellar,

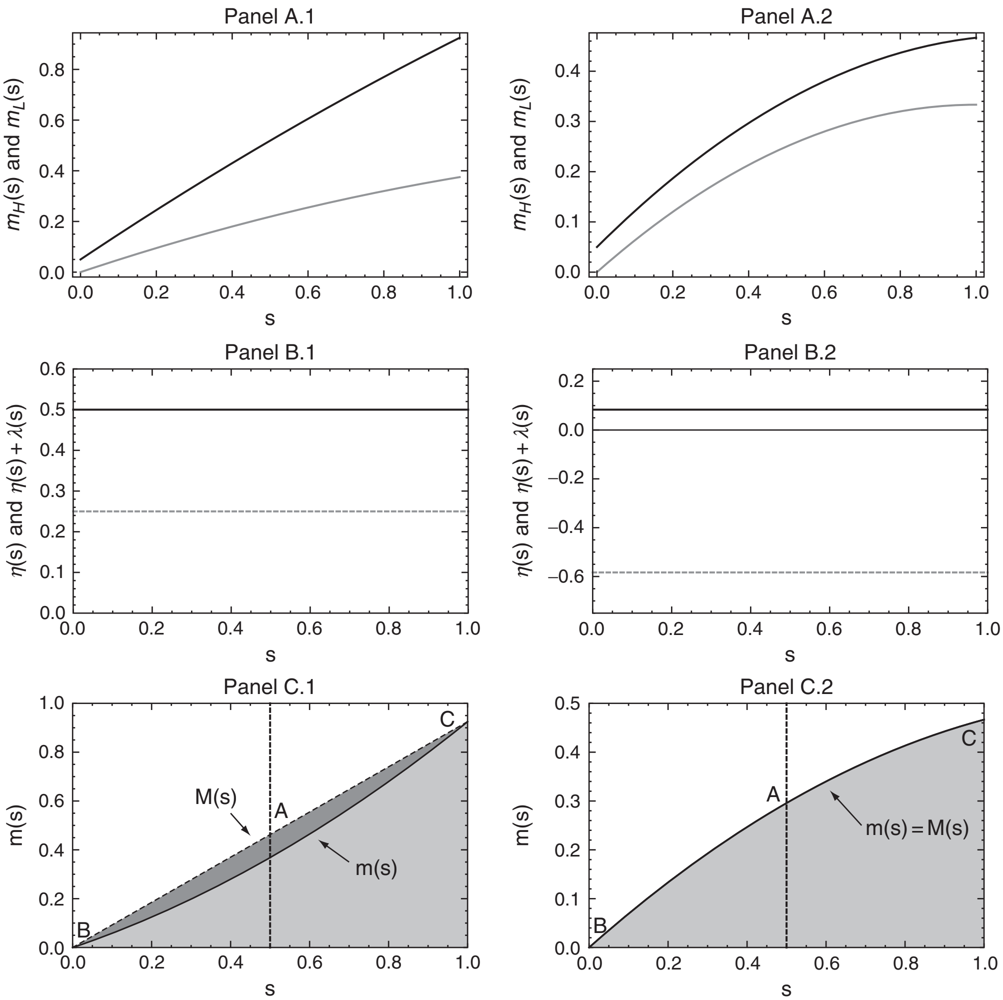

Figure 1 Social planner's assignment when $m(s)$ is globally convex and concave.
Notes: Panel A of the figure plots the type specific group composition response functions $m_H(s)$ (black line) and $m_L(s)$ (gray line). Panel B plots the associated $\eta(s)$ (solid line) and $\eta(s) + \lambda(s)$ (dashed line) functions. The sign of these functions indicate whether a local increase in segregation at point s would raise, respectively, movers' and stayers' average outcomes. Panel C plots $m(s)$, its concave envelope, $M(s)$, and the maximal attainable average outcome (point A).

1970; Horst, Pardalos and Thoai, 2000). Intuitively it is the uniformly best concave overestimator of $m(s)$. If the planner could 'produce on' $M(s)$, then an optimal assignment, by virtue of concavity, would be the perfectly integrated one. This assignment would yield an average outcome of $M(p_H)$. This observation yields the following inequality on the planner's maximand

$$\mathbb{E}_{F_S}[m(S)] \leq M(p_H),$$

for any feasible $F_S(\cdot)$. Now let s_L and s_U be the first points respectively below and above p_H where $m(s)$ and its envelope coincide:

$$s_L = \max\{s : 0 \leq s \leq p_H, M(s) = m(s)\}$$
$$s_U = \min\{s : p_H \leq s \leq 1, M(s) = m(s)\}.$$

Since $M(s)$ is linear on the interval $[s_L, s_U]$ the upper bound on the maximand is attained by the assignment

$$F_S^{\text{opt}}(s) = (1-\pi)\mathbf{1}(s \geq s_L) + \pi\mathbf{1}(s \geq s_U), \quad \pi = \begin{cases} \dfrac{p_H - s_L}{s_U - s_L} & s_L < s_U \\ \dfrac{1}{2} & s_L = s_U \end{cases}.$$

Panels C.1 & C.2 of Figure 2 illustrate the planner's solution. In the figures point A corresponds to $(p_H, M(p_H))$. To achieve this bound the planner forms two types of groups. The first has a fraction of high types equal to the s-axis value of point B in the figure, corresponding to s_L; the second the s-axis value of point C, corresponding to s_U. The feasibility constraint determines the relative frequency of the two types of groups. It is then not hard to see that the average outcome associated with this assignment is equal to $M(p_H)$. Since $M(p_H)$ is an upper bound to the planner's objective function this assignment is indeed an optimal one.

With the planner's solution characterized we can define the maximum change in the average outcome available via reassignment as

$$\beta^{\text{mre}} = \frac{1}{s_U - s_L}[(s_U - p_H)\, m(s_L) + (p_H - s_L)\, m(s_U)] - \mathbb{E}[Y]. \tag{92}$$

Given knowledge of the extreme points of the concave envelope of $m(s)$ the maximum reallocation effect is straightforward to compute. Unfortunately finding the concave envelope of $m(s)$ requires knowledge of $m(s)$ at all points on the unit interval. This requires that the status quo assignment, $F_S^{\text{sq}}(\cdot)$, have support along the entire unit interval.

Graham, Imbens and Ridder (2009b) do not develop estimation and inference results for β^{mre}. Showing consistency of a simple plug in estimator should be straightforward. If $\hat{m}(s)$ is some uniformly consistent nonparametric estimate of $m(s)$, say the NIP estimate used by Graham, Imbens and Ridder (2009a), then s_L and s_U should be consistently estimable and hence so should β^{mre}. Characterizing the asymptotic

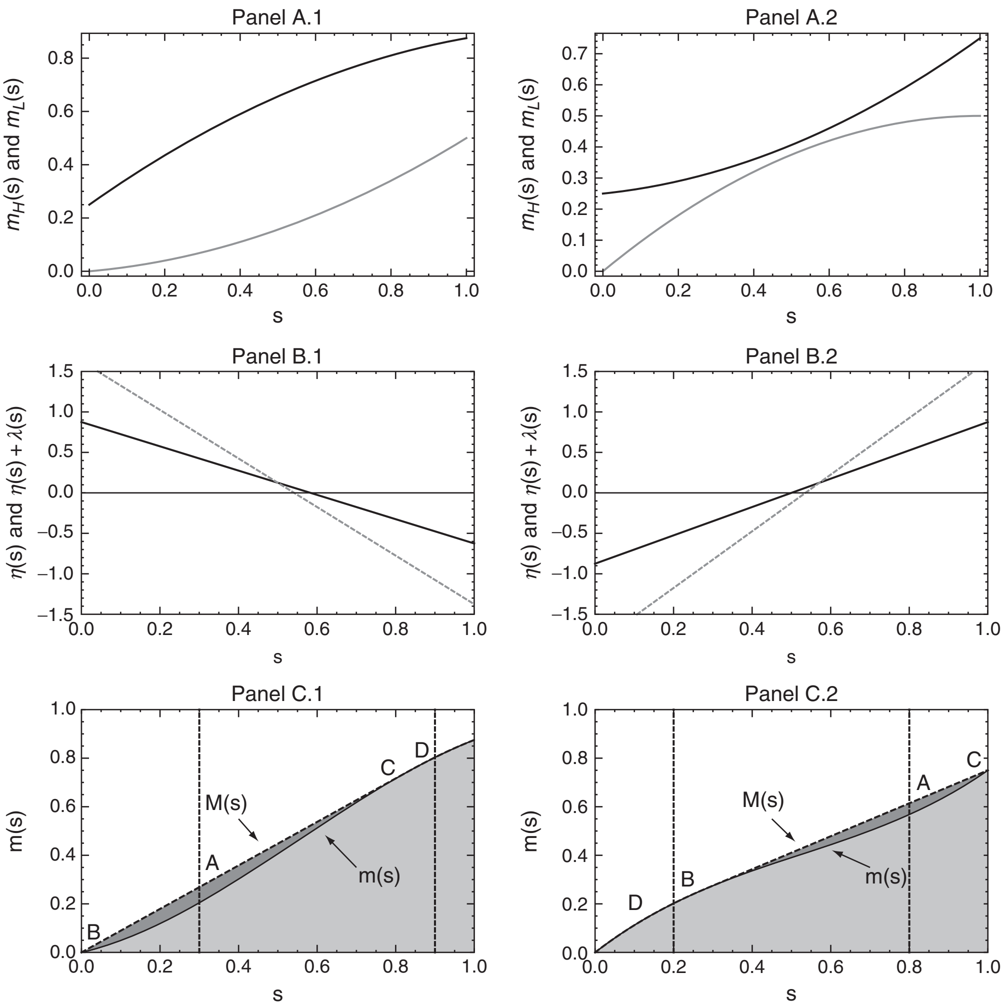

Figure 2 Social planner's assignment when *m*(*s*) is neither convex nor concave.
Notes: Panel A of the figure plots the type specific group composition response functions $m_H(s)$ (black line) and $m_L(s)$ (gray line). Panel B plots the associated $\eta(s)$ (solid line) and $\eta(s) + \lambda(s)$ (dashed line) functions. The sign of these functions indicate whether a local increase in segregation at point *s* would raise, respectively, movers' and stayers' average outcomes. Panel C plots *m*(*s*), its concave envelope, *M*(*s*), and the maximal attainable average outcome (point *A*).

sampling properties of such an estimator would be more challenging. Consider the case where s_L, s_U and p_H are known, then $\hat{\beta}^{\mathrm{mre}}$, which would be a function of $\hat{m}(s_L)$ and $\hat{m}(s_U)$, would behave similarly to a nonparametrically estimated conditional mean. When s_L, s_U and p_H are replaced with estimates, the effects of the additional sampling error would need to be ascertained.

5.3 Local reallocations: movers versus stayers

Studying the effects of small, local, reallocations on outcomes generates additional insights. Consider the set of groups with group composition s. As before, high- and low-type individuals in these groups are random draws from their respective subpopulations and assignment to locations is random. Now consider the following local reallocation. In half of the groups the fraction high is increased from s to $s + \mathrm{d}s$. This change is accommodated by reducing the fraction high from s to to $s - \mathrm{d}s$ in the remaining groups. Implementing this reallocation requires low type individuals in the first half of groups to switch places with high type individuals in the second half of groups. Call the set of individuals who actually switch groups as part of the reallocation *movers*.

The change in average outcomes for high type movers is given by

$$m_H(s + \mathrm{d}s) - m_H(s), \tag{93}$$

while that for low type movers is given by

$$m_L(s - \mathrm{d}s) - m_L(s). \tag{94}$$

The average outcome of movers will increase if the sum of (93) and (94) is positive. For $\mathrm{d}s$ infinitesimally small this is equivalent to the condition

$$\nabla_s m_H(s) - \nabla_s m_L(s) > 0. \tag{95}$$

Equation (95) is a measure of *local* complementarity between own and peers' type. If, at group composition s, high type individuals gain more from small increases in peer quality than do low type individuals, then (95) will be positive. In such situations, a high type and a low type can raise their total expected outcome by switching groups. Note that such a switch will often involve a 'winner' and a 'loser'. For example if $\nabla_s m_H(s)$ and $\nabla_s m_L(s)$ are both positive, then (95) implies that the outcome gain for the high type mover exceeds the outcome loss for the low type mover.[43]

Those individuals who do not move as part of the reallocation, or *stayers*, also experience changes in expected outcomes. This is because the reallocation changes these individuals' peer groups. To see this note that the total change in expected outcomes, movers and stayers inclusive, is given by

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

[43] The locational sorting literature emphasizes the case where (95) is positive for all $s \in [0, 1]$ (e.g., Benabou, 1996). This corresponds to assuming that own and peer types are global complements. Assuming that (i) utility is linear in the expected outcome and (ii) that there is a functioning market for 'seats' or 'residences' in groups, this condition generates strong incentives for sorting. Since a high type individual initially in a group with composition s gains more from moving to a group with composition $s + \mathrm{d}s$ than a low type loses from making the reverse move, high types will outbid low types for spots in high s groups.

For ds infinitesimally small this change will be positive if

$$\nabla_{ss} m(s) > 0.$$

Local convexity of $m(s)$ implies that locally increasing segregation will raise average outcomes. Now observe that

$$\nabla_{ss} m(s) = 2\eta(s) + \lambda(s). \tag{96}$$

for

$$\eta(s) = \nabla_s m_H(s) - \nabla_s m_L(s), \quad \lambda(s) = s\nabla_{ss} m_H(s) + (1-s)\nabla_{ss} m_L(s).$$

The first term in (96), $\eta(s)$, equals the local complementarity term of (95) above. If it is positive movers will benefit, on net, from local increases in segregation. The second term, $\lambda(s)$, is what Benabou (1996) has termed 'curvature'. The direction of the reallocation's impact on the stayers' average outcome depends on the relative magnitudes of complementarity and curvature. In particular if $\eta(s) + \lambda(s)$ is negative then stayers will be hurt, on net, by local increases in segregation. Panels B of Figures 1 and 2 plot $\eta(s)$ and $\eta(s) + \lambda(s)$ for four different pairs of $m_H(s)$ and $m_L(s)$. In three of these four examples the effects of local increases in segregation on movers and stayers are opposite signed for at least some values of s. Situations where $\eta(s) + \lambda(s)$ and $\eta(s)$ do not have the same sign are of particular interest as they suggest that the private and social returns to segregation are grossly misaligned. Even when the two terms share a common sign, there will exist a wedge between the private and social returns to sorting. This wedge arises because movers have no incentive to internalize the effects of their actions on stayers.

Translating the above analysis into meaningful estimands is challenging. One approach would involve estimating $m_H(s)$ and $m_L(s)$, as well as their first and second derivatives, pointwise and constructing sample analogs of Figures 1 and 2. Figures of this type would give some indication of the likely effects of small increases in segregation at various values of s. This approach is conceptually simple, but would likely result in noisy inferences (e.g., the maximal feasible rate of convergence for $\lambda(s)$ would be that of a one dimensional second derivative). Graham, Imbens and Ridder (2009b) propose an alternative method. They introduce a family of feasible reallocations that involves perturbing the group composition distribution across its entire support. They then study the effects of such reallocations on average outcomes and inequality. They focus on the case where the resulting reallocation becomes vanishingly close to the status quo. While this approach has obvious limitations it does lead to an intuitive estimand and a tractable estimator (which converges at the normal parametric rate). Their approach is most likely to be informative about the effects of modest changes in the degree of segregation or integration.

Decompose the derivative of $m(s)$ as

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

with

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

The first term, $p(s)$, captures the private, or compositional effect, of a small change in group composition on the average outcome. The second term, $e(s)$, in an external effect. It captures the effect of changes in group composition on the outcomes of stayers.

Now consider implementing a reallocation of the following form

$$\Delta S_c = \lambda(S_c - p_{H,\kappa}) \times d_\kappa(S_c), \quad \lambda > 0 \tag{97}$$

with $d_\kappa(s) = \mathbf{1}\ (s > \kappa)\ \mathbf{1}\ (s < 1-\kappa)$ for some $\kappa \in (0, 1)$ and $p_{H,\kappa} = \mathbb{E}[T_i \mid d_\kappa(S_i) = 1]$. This allocation takes high type individuals in predominately low type groups and switches them with low type individuals in predominately high type groups. The function $d_\kappa(S_c)$ excludes groups with compositions close to zero or one from the reallocation. This ensures that, for small enough λ, (97) is feasible. The $d_\kappa(s)$ also serves as a fixed trimming device for estimation purposes.

Graham, Imbens and Ridder (2009b) show that the *sign* of the effect on average outcomes from implementing (97) in the limit where λ approaches zero coincides with the sign of

$$\beta^{\text{lsoe}} = \mathbb{E}[d_\kappa(S_i)\nabla_s m(S_i)(S_i - p_{H,\kappa})]. \tag{98}$$

They term β^{lsoe} the *local segregation outcome effect* (LSOE). The form of β^{lsoe} is intuitive. If $\nabla_s m(S_i)$ tends to be larger for $S_i < p_{H,\kappa}$, then increasing segregation should lower average outcomes. This is because the returns to group composition are highest in those groups that have few high types. Increasing segregation therefore involves taking high types away from groups where the return to their presence is highest.

The LSOE can be broken down into the two components

$$\alpha^{\text{lmse}} = \mathbb{E}[d_\kappa(S_i)p(S_i)(S_i - p_{H,\kappa})]$$
$$\alpha^{\text{lsse}} = \mathbb{E}[d_\kappa(S_i)e(S_i)(S_i - p_{H,\kappa})],$$

with $\beta^{\text{lsoe}} = \alpha^{\text{lmse}} + \alpha^{\text{lsse}}$. The signs of α^{lmse} and α^{lsse} respectively coincide with the signs of the reallocation's effect on the average outcomes of movers and stayers. To show this Graham, Imbens and Ridder (2009b) provide the following representations

$$\alpha^{\text{lmse}} = \pi_\kappa \mathbb{V}(S_i \mid d_\kappa(S_i) = 1) \times \mathbb{E}[\omega(S_i)\eta(S_i) \mid d_\kappa(S_i) = 1]$$
$$\alpha^{\text{lsse}} = \pi_\kappa \mathbb{V}(S_i \mid d_\kappa(S_i) = 1) \times \mathbb{E}[\omega(S_i)\{\eta(S_i) + \lambda(S_i)\} \mid d_\kappa(S_i) = 1],$$

for $\pi_\kappa = \mathbb{E}[d_\kappa(S_i)]$ and $\omega(S_i)$ a mean one weight function. That is α^{lmse} is equal to a weighted average of the local complementarity measure $\eta(s)$. Recall that it is the sign of $\eta(s)$ that determines whether movers gain from local increases in segregation in the neighborhood of s. A weighted average of these local measures gives the overall direction of implementing (97) on movers. Likewise α^{lsse} is equal to a weighted average of $\eta(s) + \lambda(s)$, the sign of which determines the effect of local increases in segregation on stayers' average outcomes.

Under random sampling analog estimation of β^{lsoe} is straightforward; $\nabla_s m(S_i)$ may be estimated by differentiating the Nadarya-Watson kernel regression estimate of $m(S_i)$ and $p_{H,\kappa}$ by the trimmed sample mean of T_i. With these estimates in hand $\hat{\beta}^{\text{lsoe}}$ is given by the sample analog of (98) after replacing $\nabla_s m(S_i)$ and $p_{H,\kappa}$ with their estimates. Characterizing the limiting distribution of the resulting estimator is more involved. Graham, Imbens and Ridder (2009b) derive the appropriate influence function and show how to properly take into account the group-structure of the data when conducting inference.

Policy debates which touch on issues of segregation often center on its implications for inter-type inequality. Graham, Imbens and Ridder (2009b) show that the sign of the effect of a local increase in segregation on inter-type inequality is given by

$$\beta^{\text{lsie}} = \mathbb{E}\left[\frac{d_\kappa(S_i)}{p_H}\{m_H(S_i) + S_i \nabla_s m_H(S_i)\}(S_i - p_{H,\kappa})\right] - \mathbb{E}\left[\frac{d_\kappa(S_i)}{1 - p_H}\{-m_L(S_i) + (1 - S_i)\nabla_s m_L(S_i)\}(S_i - p_{H,\kappa})\right].$$

They call β^{lsie} the local segregation inequality effect (LSIE). As with the LSOE they propose an analog estimator and characterize its large sample properties.

5.4 Parametric examples

Empirical work on the socioeconomic effects of segregation generally assumes parametric forms for $m_H(s)$ and $m_L(s)$. This section evaluates several widely-used parametric models in light of the material reported in Sections 5.2 and 5.3 above.

Perhaps the most common empirical peer effects model is the linear-in-means one analyzed by Manski (1993). A reduced form version of this model amounts to assuming that

$$m_H(s) = \alpha_H + \beta s, \quad m_L(s) = \alpha_L + \beta s.$$

This model restricts the marginal effect of group composition to be the same for high and low type individuals. Since $\nabla_s m_H(s) = \nabla_s m_L(s) = \beta$ and $\nabla_{ss} m_H(s) = \nabla_{ss} m_L(s) = 0$ this-model implies that the LSOE is zero by construction. All assignments lead to the same average outcome. The model does allow for segregation to generate inequality. The LSIE, for example, equals

$$\beta^{\text{lsie}} = \mathbb{V}(S_i) \times \frac{2\beta}{p_H(1-p_H)},$$

where, to simplify the expression, I set $\kappa = 0$ such that $d_\kappa(s) = 1$. Note that segregation always increases inequality (if $\beta > 0$). Although the linear-in-means model is widely-used (e.g., Graham, 2008) it is clearly inappropriate for studying the effects of alternative assignments to groups.

A slight generalization of the linear-in-means model allows the return to group composition to vary by type. This model is frequently employed in analyses of the effects of racial segregation on student achievement (e.g., Schofield, 1995; Angrist and Lang, 2004; Guryan, 2004; Card and Rothstein, 2007). We have

$$m_H(s) = \alpha_H + \beta_H s$$
$$m_L(s) = \alpha_L + \beta_L s.$$

This model, while more flexible than the linear-in-means one, also rules out curvature by construction: $\lambda(s) = 0$. This means that the interests of movers and stayers are perfectly aligned. This is an important limitation given the focus on inefficient segregation in theoretical work.

In this model the LSOE equals

$$\beta^{\text{lsoe}} = \mathbb{V}(S_i) \times 2(\beta_H - \beta_L).$$

Note that the sign of β^{lsoe} coincides with the sign of $\beta_H - \beta_L$. Features of the status quo group composition distribution cannot alter the LSOE's sign. The LSIE effect is given by

$$\beta^{\text{lsie}} = \mathbb{V}(S_i) \times 2\left\{\frac{(1-p_H)\beta_H + p_H\beta_L}{p_H(1-p_H)}\right\}.$$

As with the outcome effect, the direction of the effect of local increases in segregation on inter-type inequality is independent of the status quo.

A simple parametric model that allows for curvature is the 'quadratic-in-means' model:

$$m_H(s) = \alpha_H + \beta_H s + \gamma_H s^2$$
$$m_L(s) = \alpha_L + \beta_L s + \gamma_L s^2.$$

In this model complementarity equals $\eta(s) = (\beta_H - \beta_L) - (\gamma_H - \gamma_L)\, s$ and curvature is given by $\lambda(s) = -\,[s\,\gamma_H + (1-s)\,\gamma_L]$. Global complementarity is not assumed a priori and the model is flexible enough to allow for a misalignment between the interests of movers and stayers (see Figures 1 and 2 above for examples).

The LSOE in this model is given by

$$\beta^{\text{lsoe}} = \mathbb{V}(S_i) \times \left[2(\beta_H - \beta_L + \gamma_L) + \left[2p_H + \mathbb{V}(S_i)^{1/2}\mathbb{S}(S_i)\right] \times 3(\gamma_H - \gamma_L)\right],$$

where $\mathbb{S}(S_i)$ is the skewness of S_i. Note that features of the status quo assignment meaningfully enter the above expression. The sign of $2p_H + \mathbb{V}(S_i)^{1/2}\mathbb{S}(S_i)$ will vary depending on the form of $F_S^{\text{sq}}(\cdot)$. Thus, even if the production function is the same across two 'societies', the effect of small increases in segregation need not be. The interplay between the production technology and the distribution of types in the population is a fundamental feature of assignment problems. This feature is often obscured when simple parametric forms for the production technology are employed. It motivates the focus on nonparametric approaches throughout this chapter.

6. TREATMENT RESPONSE WITH SPILLOVERS

The conventional approach to causal inference assumes the absence of interference between units (Holland, 1986). In practice many interventions are likely to generate spillover effects. For example, providing college scholarships to a few students may increase the college attendance of non-recipients within the same school. The probability of viral infection for unvaccinated individuals may vary with the fraction of their close peers who are vaccinated (e.g., Ali et al. 2005). A common reaction to 'interference' among statisticians has been to treat it as a nuisance; something to be avoided by experimental design or method of data collection (see Rosenbaum (2007) for a related discussion). This approach is problematic since social planning requires knowledge of the entire treatment response; inclusive of any spillover effects. This point is elegantly made by Manski (2009a,b).

In recent work Hudgens and Halloran (2008) and Manski (2010) have developed frameworks for analyzing treatment response in the presence of social spillovers. The exposition here follows Manski (2010). Let $i \in I$ index individuals in a population who receive treatment $T_i \in \mathcal{T}$. In the most general setup an individual's potential outcome may vary with the entire population vector of treatment assignments. This yields a treatment response function of

$$Y_i(t^I), \tag{99}$$

where t^I is one element of the Cartesian product of $\mathcal{T}$ across the entire population: $\mathcal{T}^I = \times_{i \in I}\mathcal{T}$. Note the similarity, in terms of the degree of interconnectedness across agents, with the allocation response function introduced in the previous section (equation (84) above).

Manski (2010) emphasizes two sets of restrictions on (99). The first, which he calls constant treatment response (CTR), assumes that outcomes remain constant when t^I

varies within specified subsets of $\mathcal{T}^I$. For example an individual's outcome might only depend on own and peers' treatments. Reference groups may be person-specific, but are assumed to be non-manipulable.[44] The second restriction is that interactions are distributional. This corresponds to exchangeability of peers within reference groups. These two assumptions play roles similar to those of the no spillovers across groups and peer exchangeability assumptions made by Graham, Imbens and Ridder (2009b).

Let $c_i(\cdot)$ be a function mapping the population treatment vector onto a set. The CTR assumptions is that, for treatment vectors t^I and s^I,

$$c_i(t^I) = c_i(s^I) \Rightarrow Y_i(t^I) = Y_i(s^I). \tag{100}$$

Let $G_i \subset I$ be individual i's reference group. If (99) varies only with own and peers' treatments, then $c_i(t^I) = t^{G_i}$.

Let G_{-i} denote the leave-own-out reference group and $Q(t^{G_{-i}})$ the distribution of treatments across i's peers. If interactions occur only within-groups and are additionally distributional then

$$c_i(t^I) = \begin{pmatrix} t_i \\ Q(t^{G_{-i}}) \end{pmatrix}. \tag{101}$$

If the treatment is binary-valued then the within-group distribution of treatment effects is entirely summarized by the fraction of one's peers who are treated (implicit in Manski's formulation of distributional interactions is that reference group size does not matter). This means that if T_i is binary (100) and (101) give

$$Y_i(t^I) = Y_i(t_i, \bar{t}_i). \tag{102}$$

Note that, in contrast to the allocation response function introduced in the previous section, peer unobservables do not directly enter (102). This simplification arises because here peer groups are non-manipulable.

The law of total probability yields an identification region for $P[Y_i(t_i, \bar{t}_i)]$, the distribution of $Y_i(t_i, \bar{t}_i)$, equal to

$$\begin{aligned}\mathbf{H}\{P[Y_i(t_i,\bar{t}_i)]\} = \{&P[Y_i(t_i,\bar{t}_i) \mid T_i = t_i, \bar{T}_i = \bar{t}_i]\cdot P[T_i = t_i, \bar{T}_i = \bar{t}_i] \\ &+ \delta\cdot P[T_i \neq t_i \text{or } \bar{T}_i \neq \bar{t}_i],\ \delta \in \Delta_Y\},\end{aligned}$$

where Δ_Y denotes the space of all probability distributions on Y. Manski (2010) studies a variety of additional restrictions that may tighten $\mathbf{H}\{P[Y_i(t_i, \bar{t}_i)]\}$.

[44] That is reference groups may be directional in the sense that i may belong to j's reference group but not vice-versa (i.e., reference groups may be overlapping). This differs from Graham, Imbens and Ridder (2009b) who require that reference groups form a non-overlapping partition of the population.

Manski (2009a,b) studies social planning problems in the presence of treatment spillovers. Let $u_i(t_i, \bar{t}_i) = u_i[Y_i(t_i, \bar{t}_i), t_i, \bar{t}_i]$ be the utility of individual i when she is assigned treatment $T_i = t_i$ and $\bar{t}_i$ of her reference group receives treatment. Define

$$\alpha(\bar{t}_i) = \mathbb{E}[u_i(0, \bar{t}_i)], \quad \beta(\bar{t}_i) = \mathbb{E}[u_i(1, \bar{t}_i)],$$

to be the average utility of a non-treated and treated individual, respectively, when $\bar{t}_i$ of reference group members are treated. For example $\beta(\bar{t}_i)$ and $\alpha(\bar{t}_i)$ might be the average probability of infection given vaccination and non-vaccination when $\bar{t}_i$ of the reference group is vaccinated. Assume that $\beta(\bar{t}_i)$ adjusts for the cost of vaccination as well as any vaccine side-effects.

What fraction of the reference group should be vaccinated? The planner's criterion function is

$$W(\bar{t}_i) = (1 - \bar{t}_i)\alpha(\bar{t}_i) + \bar{t}_i\beta(\bar{t}_i). \tag{103}$$

If the vaccine is fully effective, then $\beta(\bar{t}_i) = \beta$. Assume further that spillovers onto the untreated are linear

$$\alpha(\bar{t}_i) = (1 - \bar{t}_i)\alpha_0 + \bar{t}_i\alpha_1.$$

Under these assumptions Manski (2009a) shows that the optimal vaccine rate is

$$\bar{t}_i^* = \max\left\{0, \min\left\{\frac{1}{2} + \frac{1}{2}\frac{\beta + \alpha_0}{\alpha_1 - \alpha_0}, 1\right\}\right\}.$$

Knowledge of the structure of any treatment spillovers, in this case, α_0 and α_1, is required to implement this rule. Manski (2009a,b) also studies social planning under ambiguity. The point I wish to make here is that social planning requires knowledge of the external effect.

7. AREAS FOR FURTHER RESEARCH

Several areas touched on in this chapter merit further study. Developing identification results and estimation methods for nontransferable utility matching models is one. Since the set of stable matchings can be large it seems likely that the parameters of these models will be set identified in absence of additional restrictions on the matching mechanism. Results for more complex matching structures, such as those with one-to-many and many-to-many matching, are also needed. Fox (2009a,b) provides some results along these lines. Relatedly, tractable, yet microtheory-founded, econometric models of network formation would help empirical researchers in their analyses of (increasingly available) social network data (e.g., Christakis, Fowler, Imbens and Kalyanaraman, 2010).

The material surveyed above has unduly focused on restrictions which result in point identification of the parameter of interest. The underlying economics of the class of models surveyed above naturally generates partially identifying restrictions (e.g., stability inequalities). Developing results based on these restrictions would help to extend the material surveyed above to a larger number of empirical problems.

REFERENCES

Ackerberg, D.l., Benkard, L.C., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Economics. 6A, North-Holland, Amsterdam, pp. 4171–4275.

Ali, M., et al., 2005. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis. Lancet 366 (9479), 44–49.

Andrews, D.W.K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78 (1), 119–157.

Angrist, J.D., Lang, K., 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. Am. Econ. Rev. 94 (5), 1613–1634.

Aradillas-López, A., Tamer, E., 2008. The identification power of equilibrium in simple games. J. Bus. Econ. Stat. 26 (3), 261–283.

Baccara, M., Imrohoroglu, A., Wilson, A., Yariv, L., 2009. A field study on matching with network externalities. Mimeo.

Becker, G.S., 1973. A theory of marriage: part I. J. Polit. Econ. 81 (4), 813–846.

Becker, G.S., 1974. A theory of marriage: part II. J. Polit. Econ. 82 (2.2), S11–S26.

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

Benabou, R., 1993. Workings of a city: location, education, and production. Q. J. Econ. 108 (3), 619–652.

Benabou, R., 1996. Equity and efficiency in human capital investment: the local connection. Rev. Econ. Stud. 63 (2), 237–264.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.

Bhattacharya, D., 2009. Inferring optimal peer assignment from experimental data. J. Am. Stat. Assoc. 104 (486), 486–500.

Blume, L.E., Brock, W.A., Durlauf, S.N., Ioannides, Y.M., 2010. Identification of social interactions. this Handbook.

Blundell, R.W., Powell, J.L., 2003. Endogeneity in nonparametric and semiparametric regression models. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), Advances in Economics and Econometrics: Theory and Applications. II, Cambridge University Press, Cambridge, pp. 312–357.

Bose, A., 1998. Bahadur representation of M_m estimates. Ann. Stat. 26 (2), 771–777.

Bose, A., 2002. U statistics and M_m estimates. In: Misra, J.C. (Ed.), Uncertainty and Optimality: Probability, Statistics and Operations Research. World Scientific Publishing Co. Pte. Ltd, Singapore, pp. 257–292.

Brock, W.A., Durlauf, S.N., 2001a. Discrete choice with social interactions. Rev. Econ. Stud. 68 (2), 235–260.

Brock, W.A., Durlauf, S.N., 2001b. Interactions-based Models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. 5, North-Holland, Amsterdam, pp. 3297–3380.

Brock, W.A., Durlauf, S.N., 2007. Identification of binary choice models with social interactions. J. Econom. 140 (1), 52–75.

Buddin, R., Zamarroa, G., 2009. Teacher qualifications and student achievement in urban elementary schools. J. Urban Econ. 66 (2), 103–115.

Burkard, R., Mauro, D., Martello, S., 2009. Assignment Problems. Society for Industrial and Applied Mathematics, Philadelphia.

Card, D., Rothstein, J., 2007. Racial segregation and the black–white test score gap. J. Public Econ. 91 (11–12), 2158–2184.
Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does your cohort matter? Measuring peer effects in college achievement. J. Labor Econ. 27 (3), 439–464.
Cavanagh, C., Sherman, R.P., 1998. Rank estimators for monotonic index models. J. Econom. 84 (2), 351–381.
Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econ. Stud. 47 (1), 225–238.
Chamberlain, G., 1982. Multivariate regression models for panel data. J. Econom. 18 (1), 5–46.
Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics. 2, North-Holland, Amsterdam, pp. 1247–1318.
Chamberlain, G., 2009. Bayesian aspects of treatment choice. Mimeo.
Chamberlain, G., 2010. Binary response models for panel data: identification and information. Econometrica 78 (1), 159–168.
Chiappori, P.A., Oreffice, S., 2008. Birth control and female empowerment: an equilibrium analysis. J. Polit. Econ. 116 (1), 113–139.
Chiappori, P.-A., Salanie, B., Weiss, Y., 2010. Partner choice and the marital college premium. Mimeo.
Choo, E., Siow, A., 2006a. Who marries whom and why. J. Polit. Econ. 114 (1), 175–201.
Choo, E., Siow, A., 2006b. Estimating a marriage matching model with spillover effects. Demography 43 (3), 463–490.
Coleman, J.S., et al., 1966. Equality of Educational Opportunity. Department of Health, Education and Welfare, Washington, U.S.
Crawford, V.P., Knoer, E.M., 1981. Job matching with heterogeneous firms and workers. Econometrica 49 (2), 437–450.
Christakis, N., Fowler, J., Imbens, G.W., Kalyanaraman, K., 2010. An empirical model for strategic network formulation. Mimeo.
Dagsvik, J.K., 2000. Aggregation in matching markets. Int. Econ. Rev. (Philadelphia) 41 (1), 27–57.
Dagsvik, J.K., Brunborg, H., Flaatten, A.S., 2001. A behavioral two-sex marriage model. Math. Popul. Stud. 9 (2), 97–121.
Dardanoni, V., Forcina, A., 1998. A unified approach to likelihood inference on stochastic orderings in a nonparametric context. J. Am. Stat. Assoc. 93 (443), 1112–1123.
de Bartolome, Charles, A.M., 1990. Equilibrium and inefficiency in a community model with peer group effects. J. Polit. Econ. 98 (1), 110–133.
Debreu, G., 1960. Review of 'Individual Choice Behavior: A Theoretical Analysis' by R. Duncan Luce. Am. Econ. Rev. 50 (1), 186–188.
Delgado, M.A., Rodríguez-Poo, J.M., Wolf, M., 2001. Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. Econ. Lett. 73 (2), 241–250.
Domencich, T.A., McFadden, D., 1975. Urban Travel Demand: A Behavioral Analysis. North Holland, Amsterdam.
Douglas, R.E.F., Lee, M.T., Sampson, A.R., Whitaker, L.R., 1990. Positive dependence concepts for ordinal contingency tables. In: Block, H.W., Sampson, A.R., Savits, T.H. (Eds.), Topics in statistical dependence. Institute of Mathematical Statistics, Hayward, CA, pp. 189–202.
Durlauf, S.N., 1996a. Neighborhood feedbacks, endogenous stratification, and income inequality. In: Barnett, W.A. et al., (Ed.), Dynamic Disequilibrium Modeling. Cambridge University Press, Cambridge, pp. 505–534.
Durlauf, S.N., 1996b. A theory of persistent income inequality. J. Econ. Growth 1 (1), 75–93.
Durlauf, S.N., 1996c. Associational redistribution: a defense. Polit. Soc. 24 (4), 391–410.
Durlauf, S.N., 2004. Neighborhood effects. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics. 4, North-Holland, Amsterdam, pp. 2173–2242.
Echenique, F., Lee, S.M., Shum, M., 2010. Aggregate matchings. California Institute of Technology Social Sciences Working Paper 1316.
Eckstein, Z., Wolpin, K.I., 1990. Estimating a market equilibrium search model from panel data on individuals. Econometrica 58 (4), 783–808.
Edelman, B., Ostrovsky, M., Schwarz, M., 2007. Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. Am. Econ. Rev. 97 (1), 242–259.

Epple, D., Romano, R., 1998. Competition between private and public schools, vouchers, and peer-group effects. Am. Econ. Rev. 88 (1), 33–62.
Epple, D., Romano, R., 2010. Peer effects in education: a survey of the theory and evidence. this Handbook.
Fernandez, J.C., 1992. Soldier quality and job performance in team tasks. Soc. Sci. Q. 73 (2), 253–265.
Fernández, R., 2003. Sorting, education and inequality. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress. 2, Cambridge University Press, Cambridge, pp. 1–40.
Ferreyra, M.M., 2007. Estimating the effects of private school vouchers in multidistrict economies. Am. Econ. Rev. 97 (3), 789–817.
Flinn, C.J., 2006. Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. Econometrica 74 (4), 1013–1062.
Fox, J.T., 2009a. Identification in matching games. Mimeo.
Fox, J.T., 2009b. Estimating matching games with transfers. Mimeo.
Fox, J.T., 2009c. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. Rand J. Econ. 38 (4), 1002–1019.
Fox, J.T., Bajari, P.k., 2009. Measuring the efficiency of an FCC spectrum auction. Mimeo.
Gale, D., 1960. The Theory of Linear Economic Models. McGraw-Hill Book Company, New York.
Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. Am. Math. Mon. 69 (1), 9–15.
Galichon, A., Salanié, B., 2009. Matching with trade-offs. Mimeo.
Glaeser, E., Scheinkman, J., 2001. Measuring social interactions. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. MIT Press, Cambridge, MA, pp. 83–130.
Glaeser, E., Scheinkman, J.A., 2003. Nonmarket interactions. In: Dewatripont, M. (Ed.), Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress. 1, Cambridge University Press, Cambridge, pp. 339–369.
Goeree, J.K., Holt, C.A., Palfrey, T.R., 2005. Regular quantal response equilibrium. Experimental Economics 8 (4), 347–367.
Goodman, L.A., 1979. Simple models for the analysis of association in cross-classifications having ordered categories. J. Am. Stat. Assoc. 74 (367), 537–552.
Gordon, N., Knight, B., 2009. A spatial merger estimator with an application to school district consolidation. J. Public Econ. 93 (5–6), 752–765.
Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. Econometrica 76 (3), 643–660.
Graham, B.S., Ridder, G., Imbens, G.W., 2007. The effect of reallocating an indivisible input with discrete values. Mimeo.
Graham, B.S., Ridder, G., Imbens, G.W., 2009a. Complementarity and aggregate implications of assortative matching: a nonparametric analysis. Mimeo.
Graham, B.S., Ridder, G., Imbens, G.W., 2009b. Measuring the average outcome and inequality effects of segregation in the presence of social spillovers. Mimeo.
Griliches, Z., Mairesse, J., 1998. Production functions: the search for identification. In: Strøm, S. (Ed.), Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium. Cambridge University Press, Cambridge, pp. 169–203.
Guryan, J., 2004. Desegregation and black dropout rates. Am. Econ. Rev. 94 (4), 919–943.
Han, A.K., 1987. Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. J. Econom. 35 (2–3), 303–316.
Hatfield, J.W., Milgrom, P.R., 2005. Matching with contracts. Am. Econ. Rev. 95 (4), 913–935.
Heckman, J.J., Vytlacil, E.J., 2007a. Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, vol. 6. (2), North-Holland, Amsterdam, pp. 4779–4874.
Heckman, J.J., Vytlacil, E.J., 2007b. Econometric evaluation of social programs, part II: using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, vol. 6. (2), North-Holland, Amsterdam, pp. 4875–5143.

Holland, P.W., 1986. Statistics and causal inference. J. Am. Stat. Assoc. 81 (396), 945–960.
Honoré, B.E., Powell, J.L., 1994. Pairwise difference estimators of censored and truncated regression models. J. Econom. 64 (1–2), 241–278.
Honoré, B.E., Powell, J.L., 2005. Pairwise difference estimators for nonlinear models. In: Andrews, D.W. K., Stock, J.H. (Eds.), Identification and Inference for Econometric Models. Cambridge University Press, Cambridge, pp. 520–553.
Horst, R., Pardalos, P.M., Thoai, N.V., 2000. Introduction to Global Optimization, second ed. Kluwer Academic Publishers, Dordrecht.
Huber, P.J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics 35 (1), 73–101.
Hudgens, M.G., Halloran, E., 2008. Toward a causal inference with interference. J. Am. Stat. Assoc. 103 (482), 832–842.
Imbens, G.W., 2007a. Nonadditive models with endogenous regressors. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress. III, Cambridge University Press, Cambridge, pp. 17–46.
Imbens, G.W., 2007b. What's News in Econometrics?. Unpublished lecture notes.
Imbens, G.W., Ridder, G., 2009. Estimation and inference for generalized full and partial means and average derivatives. Mimeo.
Jacob, B.A., Lefgren, L., 2007. What do parents value in education? An empirical investigation of parents' revealed preferences for teachers. Q. J. Econ. 122 (4), 1603–1638.
Kelso, A.S., Crawford, V.P., 1982. Job matching, coalition formation, and gross substitutes. Econometrica 50 (6), 1483–1504.
Koopmans, T.C., Beckmann, M., 1957. Assignment problems and the location of economic activities. Econometrica 25 (1), 53–76.
Kremer, M., 1997. How much does sorting increase inequality? Q. J. Econ. 112 (1), 115–139.
Kremer, M., Maskin, E., 1996. Wage inequality and segregation by skill. NBER Working Paper No. W5718.
Kudo, A., 1963. A multivariate analogue of the one-sided test. Biometrika 50 (3–4), 403–418.
Lankford, H., Loeb, S., Wyckoff, J., 2003. Teacher sorting and the plight of urban schools: a descriptive analysis. Educ. Eval. Policy Anal. 24 (1), 37–62.
Lee, L., 1995. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. J. Econom. 65 (2), 381–428.
Lehmann, E.L., 1966. Some concepts of dependence. Annals of Mathematical Statistics 7 (5), 1137–1153.
Legros, P.k., Newman, A.F., 2002. Monotone matching in perfect and imperfect worlds. Rev. Econ. Stud. 69 (4), 925–942.
Loeb, S., Boyd, D., Lankford, H., Wyckoff, J., 2003. Analyzing determinants of the matching of public school teachers to jobs: estimating compensating differentials in imperfect labor markets. NBER Working Paper 9878.
Logan, J.A., 1998. Estimating two-sided logit models. Sociol. Methodol. 28 (1), 139–173.
Logan, J.A., Hoff, P.D., Newton, M.A., 2008. Two-sided estimation of mate preferences for similarities in age, education, and religion. J. Am. Stat. Assoc. 103 (482), 559–569.
Loury, G.C., 1977. A dynamic theory of racial income differences. In: Wallace, P.A., LaMond, A. (Eds.), Women, Minorities and Employment Discrimination. Lexington Books, Lexington, MA, pp. 153–186.
Loury, G.C., 2002. The Anatomy of Racial Inequality. Harvard University Press, Cambridge, MA.
Luenberger, D.G., 1969. Optimization by Vector Space Methods. John Wiley & Sons, New York.
Luenberger, D.G., 2005. Linear and Nonlinear Programming, second ed. Springer, New York.
Matzkin, R.L., 2007. Heterogenous choice. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Theory and Applications. III, Cambridge University Press, Cambridge, pp. 75–110.
Manski, C.F., 1975. Maximum score estimation of the stochastic utility model of choice. J. Econom. 3 (3), 205–228.
Manski, C.F., 1987. Semiparametric analysis of random effects linear models from binary panel data. Econometrica 55 (2), 357–362.

Manski, C.F., 1992. Educational choice (vouchers) and social mobility. Econ. Educ. Rev. 11 (4), 351–369.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60 (3), 531–542.
Manski, C.F., 2003. Partial Identification of Probability Distributions. Springer-Verlag, New York.
Manski, C.F., 2009a. Social planning with partial knowledge of social interactions. J. Inst. Theor. Econ. 165 (1), 146–158.
Manski, C.F., 2009b. Vaccination with partial knowledge of external effectiveness. Proc. Natl. Acad. Sci. U. S. A. 107 (9), 3953–3960.
Manski, C.F., 2010. Identification of treatment response with social interactions. Mimeo.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.
McFadden, D., 1981. Econometric models of probabilistic choice. In: Manski, C.F., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA, pp. 198–272.
Moffitt, R.A., 2001. Policy interventions, low-level equilibria and social interactions. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. MIT Press, Cambridge, MA, pp. 45–82.
Mortensen, D.T., 1988. Matching: finding a partner for life or otherwise. A.J.S. 94 (Suppl.), S215–S240.
Nechyba, T.J., 2006. Income and peer quality sorting in public and private schools. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education. 2, North-Holland, Amsterdam, pp. 1327–1368.
Nevo, A., 2000. A practioner's guide to estimation of random-coefficient logit models of demand. J. Econ. Manag. Strategy 9 (4), 513–548.
Newey, W.K., 1994. Kernel estimation of partial means and a general variance estimator. Econometric Theory 10 (2), 233–253.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. 4, North Holland, Amsterdam, pp. 2111–2245.
Ostrovsky, M., 2008. Stability in supply chain networks. Am. Econ. Rev. 98 (3), 897–923.
Piketty, T., 2000. Theories of persistent inequality and intergenerational mobility. In: Atkinson, A., Bourguignon, F. (Eds.), Handbook of Income Distribution. 1, North Holland, Amsterdam, pp. 430–476.
Powell, J.L., 1994. Estimation of semiparametric models. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. 4, North Holland, Amsterdam, pp. 2443–2521.
Powell, J.L., Ruud, P.A., 2008. Simple estimators for semiparametric multinomial choice models. Mimeo.
Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.
Romano, J.P., Shaikh, A., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78 (1), 169–211.
Rosen, A.M., 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. J. Econom. 146 (1), 107–117.
Rosenbaum, P.R., 2007. Interference between units in randomized experiments. J. Am. Stat. Assoc. 102 (477), 191–200.
Roth, A.E., 2008. What have we learned from market design? Economic Journal 118 (527), 285–310.
Roth, A.E., Sotomayor, M., 1990. Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Cambridge University Press, Cambridge.
Roth, A.E., Vate, J.H.V., 1990. Random paths to stability in two-sided matching. Econometrica 58 (6), 1475–1480.
Sattinger, M., 1980. Capital and the Distribution of Labor Earnings. Amsterdam, North-Holland.
Schofield, J.W., 1995. Review of research on school desegregation's impact on elementary and secondary school students. In: Banks, J.A., Banks, C.A.M. (Eds.), Handbook of Research on Multicultural Education. Macmillan Publishing USA, New York, pp. 597–616.
Shapley, L.S., Shubik, M., 1971. The assignment game I: the core. International Journal of Game Theory 1 (1), 111–130.
Shapley, L., Scarf, H., 1974. On cores and indivisibility. Journal of Mathematical Economics 1 (1), 23–37.

Siow, A., 2008. How does the marriage market clear? An empirical framework. Can. J. Econ. 41 (4), 1121–1155.
Siow, A., 2009. Testing Becker's theory of positive assortative matching. University of Toronto, Department of Economics Working Paper 356.
Silvapulle, M.J., Sen, P.K., 2005. Constrained Statistical Inference: Inequality, Order, and Shape Restrictions. John Wiley & Sons, Inc, Hoboken, N.J.
Sørensen, M., 2007. How smart is smart money? A two-sided matching model of venture capital. J. Finance 62 (6), 2725–2762.
Topkis, Donald M., 1998. Supermodularity and Complementarity. Princeton University Press, Princeton, NJ.
Warner, J.T., Asch, B.J., 1995. The economics of military manpower. In: Hartley, K., Sandler, T. (Eds.), Handbook of Defense Economics. 1, North-Holland, Amsterdam, pp. 347–398.
Weiss, Y., 1997. The formation and dissolution of families: Why marry? Who marries whom? And what happens upon divorce. In: Rosenzweig, M.R., Stark, O. (Eds.), Handbook of Population and Family Economics, vol. 1. (1), North-Holland, Amsterdam, pp. 81–123.
Yang, Y., Shi, M., Goldfarb, A., 2009. Estimating the value of brand alliances in professional team sports. Marketing Science 28 (6), 1095–1111.