Quantitative Economics 5 (2014), 29–66 1759-7331/20140029

Complementarity and aggregate implications of assortative matching: A nonparametric analysis

Bryan S. Graham
Department of Economics, University of California—Berkeley and NBER

Guido W. Imbens
Graduate School of Business, Stanford University

Geert Ridder
Department of Economics, University of Southern California

This paper presents econometric methods for measuring the average output effect of reallocating an indivisible input across production units. A distinctive feature of reallocations is that, by definition, they involve no augmentation of resources and, as such, leave the marginal distribution of the reallocated input unchanged. Nevertheless, if the production technology is nonseparable, they may alter average output. An example is the reallocation of teachers across classrooms composed of students of varying mean ability. We focus on the effects of reallocating one input, while holding the assignment of another, potentially complementary, input fixed. We introduce a class of such reallocations—correlated matching rules—that includes the status quo allocation, a random allocation, and both the perfect positive and negative assortative matching allocations as special cases. We also characterize the effects of small changes in the status quo allocation. Our analysis leaves the production technology nonparametric. Identification therefore requires conditional exogeneity of the input to be reallocated given the potentially complementary (and possibly other) input(s). We relate this exogeneity assumption to the pairwise stability concept used in the game theoretic literature on two-sided matching models with transfers. For estimation, we use a two-step approach. In the first step, we nonparametrically estimate the production function. In the second step, we average the estimated production function over the distribution of inputs induced by the new assignment rule. Our methods build upon the partial mean literature, but require extensions involving boundary issues and the fact that the weight function used in averaging is itself estimated. We derive the large-sample properties of our proposed estimators and assess their

Bryan S. Graham: bgraham@econ.berkeley.edu
Guido W. Imbens: imbens@stanford.edu
Geert Ridder: ridder@usc.edu
Financial support for this research was generously provided through NSF Grants SES 0136789, SES 0452590, and SES 0820361. We thank participants in seminars at Harvard–MIT, Princeton, UC—Berkeley, NBER Labor Studies, Cemmap, and Brown University for feedback. We also thank the co-editor and three anonymous referees for numerous comments and suggestions. Cristine Pinto provided excellent research assistance.

DOI: 10.3982/QE45

small-sample properties via a limited set of Monte Carlo experiments. Our characterization of the large-sample properties of estimated correlated matching rules uses a new result on kernel estimated "double averages," which may be of independent interest.

Keywords. Aggregate redistributional effects, complementarity, nonparametric estimation, partial mean, assortative matching, one-to-one matching with transfers, assignment problem, assignment game.

JEL classification. C14, C21, C52.

1. Introduction

Consider an educational production function. One factor of production—teacher quality—is intrinsically indivisible: a teacher may teach in one, and only one, classroom. Classrooms may be heterogeneous in mean student ability. While student achievement may be monotone in teacher quality across all classrooms, the magnitude of its responsiveness may not. Student achievement might rise more sharply with teacher quality in classrooms composed of students of above average ability. In such a situation, it may be possible to raise average student achievement by reallocating teachers across classrooms; in this case, by assigning high quality teachers to classrooms composed of students with above average ability.

Reallocations, unlike other policies, involve no augmentation of resources. Indivisibility of the input under consideration further complicates their analysis. If the input is indivisible and its aggregate stock is fixed, it is impossible to simultaneously raise the input level for all production units or firms. We cannot place a higher quality teacher into all classrooms if the population of teachers available for assignment remains fixed. In such cases, the achievement effects of reassigning teachers (the "input") across classrooms (the "firms") may be of interest.

Many organizations are interested in the output implications of reallocating, as opposed to augmenting, existing resources (cf. Graham (2011)). Here we investigate econometric methods for measuring the effects of such policies on average output. We will call the average causal effects of these policies aggregate redistributional effects (AREs). A key feature of the reallocations we consider is that, although they potentially alter input levels for each firm, they leave its marginal distribution across the population of firms unchanged.

The first contribution of our paper is to introduce a framework for considering such reallocations and to define estimands that capture their key features. These estimands include the effects of four focal reallocations, a semiparametric class of reallocations, and the effect of a small reallocation away from the status quo.

One focal reallocation redistributes the input across production units such that it has perfect rank correlation with a second input. This is the positive assortative matching (PAM) allocation. We also consider a negative assortative matching (NAM) allocation where the primary input is redistributed to have perfect negative rank correlation with the second input. A third allocation involves randomly assigning the input across firms. This allocation, by construction, ensures independence of the two inputs. A fourth allocation simply maintains the status quo assignment of the input. More generally, we consider a two parameter family of feasible reallocations that includes these four focal allocations as special cases.

Our family of reallocations, which we call *correlated matching rules*, traces a path from the positive to the negative assortative matching allocations. Each reallocation along this path keeps the marginal distribution of the two inputs fixed, but it induces a different level of correlation between the two inputs.

We also provide a local measure of complementarity that is identified under weaker conditions on the support of the input distribution. This estimand measures whether a small step away from the status quo toward the perfect assortative matching raises average output. It may be useful for assessing the probable effects of policies that induce reallocations that are close to the status quo.

Each of our estimands is a functional of the production technology and the marginal distributions of the two inputs.[1] In most applications, identification of the two input distribution functions will be straightforward, while that of the production technology, due to purposeful input choice on the part of firms, will be more difficult (Griliches and Mairesse (1998)). The second contribution of our paper is to study identification and estimation under an exogeneity condition on the input to be reallocated. This condition ensures that the production function is nonparametrically identified.

The effects of reallocations can be very sensitive to how the sign and magnitude of the cross-partial derivative of the production function varies with different input combinations (Graham (2011)). Since a priori parametric or semiparametric restrictions on the production function may impose substantial structure on the form of this cross-partial, such assumptions may inappropriately restrict the range of reallocation effects allowed. For this reason, our approach to identification and estimation invokes conditions that are sufficiently strong to allow for a nonparametric treatment of the production technology.[2]

We propose analog estimators based on our identification results and characterize their large-sample properties. Specifically, we propose an estimator for average output under all correlated matching allocations as well as for the local complementarity measure. Except for perfect negative and positive assortative matchings, these estimators converge at the usual parametric rate. For these two extreme matchings, the rate of convergence is slower, comparable to that of estimating a regression function with a scalar covariate at a point. In the first step of the estimation procedures, we use a nonparametric estimate of the production function. We modify existing kernel estimators to deal with boundary issues that arise in our setting.

Our focus on reallocation rules that keep the marginal distribution of the inputs fixed is appropriate in applications where the input is indivisible, such as in the allocation of

[1]Our local measure of complementarity also depends on certain features of the status quo joint input distribution.

[2]In settings where our exogeneity assumption is implausible, other approaches to semiparametrically identifying the production technology may be available (e.g., Matzkin (2008), Imbens and Newey (2009)). Our estimation theory would need to be modified for such cases. An interesting topic for future research would be to explore what can be learned about AREs from partially identified production technologies (cf. Kasy (2012)).

teachers to classes or managers to production units. In other settings, it may be more appropriate to consider allocation rules that leave the total amount of the input constant by fixing its average level. Such rules would require some modification of the methods considered in this paper (cf. Bhattacharya and Dupas (2012)).

Our methods may be useful in a variety of settings. One class of examples concerns complementarity between organizational form and technology (e.g., Athey and Stern (1998)). A second example concerns educational production functions. Loeb, Kalogrides, and Béteille (2012), using administrative data from Miami–Dade County Public Schools, presented evidence that suggests that “effective” schools attract and retain higher quality teachers. Teacher quality may improve outcomes for all students, but average outcomes may be higher or lower depending on whether, given a fixed supply of teachers, the best teachers are assigned to the least prepared students or vice versa. Parents concerned solely with outcomes for their own children may be most interested in the effect of raising teacher quality on expected outcomes. A school board, including the one in Miami–Dade County, however, may be more interested in maximizing expected outcomes, given a fixed set of classes and a fixed set of teachers, by optimally matching teachers to classes.

A third class of examples arises in settings with social interaction (cf. Manski (1993), Brock and Durlauf (2001)). Sacerdote (2001) studied peer effects in college by looking at the relationship between individual outcomes and roommate characteristics. From the perspective of the individual student, it may again be of interest whether having a roommate with different characteristics would, in expectation, lead to a different outcome. This is what Manski (1993) called an exogenous or contextual effect. The college, however, may be interested in a different effect, namely the effect on average outcomes of changing the procedures for assigning roommates. While a college may be unable to quickly change the distribution of characteristics in incoming classes, it is able to change the way roommates are assigned. In Graham, Imbens, and Ridder (2010), we studied the peer effect setting further, developing methods appropriate for social groups of arbitrary size when agents are binary typed. Our focus in that work is on the outcome and inequality effects of segregation.

If production functions are additive in inputs, the questions posed above have trivial answers: average outcomes are invariant to input reallocations. Although reallocations may raise outcomes for some units in that case, they will necessarily lower them by an offsetting amount for others. To generate nontrivial answers, one needs to allow for nonadditivity and nonlinearity in the production function. To reiterate, parametric or semiparametric assumptions on the production function, although useful for achieving identification, often imply very strong a priori restrictions on the structure of any reallocation effects. For this reason, our approach is fully nonparametric.

The current paper builds on the larger treatment effect and program evaluation literature.[3] More directly, it is complementary to the small literature on the effect of treatment assignment rules (Manski (2004), Dehejia (2005), Hirano and Porter (2009),

[3]For recent surveys, see Angrist and Krueger (1999), Heckman, Lalonde, and Smith (2000), and Imbens and Wooldridge (2009).

Stoye (2009), Tetenov (2012)). Our focus is different from these studies. First, in a conceptually straightforward extension, we allow for continuous rather than discrete or binary treatments. Second, and this is the main innovation relative to prior work, our assignment policies take into account resource constraints (by leaving unchanged the marginal distribution of the treatment), whereas in the previous papers, treatment assignment for one unit is not restricted by the assignments for other units. Stock (1989) studied policies that induce new input distributions. However, his work does not involve resource constraints. Our policies, in contrast, are redistributions.[4]

In the current paper, we focus on estimation and inference for specific assignment rules. It is also interesting to consider optimal rules as in the Manski, Dehejia, and Hirano–Porter studies. The class of feasible reallocations/redistributions includes all joint distributions of the two inputs with fixed marginal distributions. When the inputs are continuously valued, as we assume in the current paper, this class of potential rules is very large. Characterizing the optimal allocation within this class is, therefore, a nontrivial problem (cf. Chiappori, McCann, and Nesheim (2010)). When both inputs are discretely valued, the problem of finding the optimal allocation is tractable as the joint distribution of the inputs is characterized by a finite number of parameters (specifically, linear programming methods may be applied). In Graham, Imbens, and Ridder (2007), we considered optimal allocation rules when both inputs are discrete, allowing for general complementarity or substitutability of the inputs (cf. Bhattacharya (2009)).

Our paper is also related to recent work on identification and estimation of models of social interactions (e.g., Manski (1993), Brock and Durlauf (2001), Moffitt (2001), Graham (2008)). We do not focus on directly characterizing the within-group microstructure of social interactions, an important theme of this literature. Rather our goal is simply to estimate the average relationship between pair composition and outcomes. The average we estimate may reflect endogenous behavioral responses by the two agents to changes in each others' attributes; it may even equal a mixture over multiple equilibria (see Bajari, Hahn, Hong, and Ridder (2011)). Viewed in this light, our approach is reduced form in nature. However, it is sufficient for, say, a university administrator to characterize the outcome effects of alternative roommate assignment procedures as long as the average response to group composition remains unchanged across such procedures.

Finally, the approach taken here is complementary to recent work by Choo and Siow (2006a, 2006b), Fox (2010a, 2010b), Galichon and Salanié (2011), and Graham (2011) on the identification and estimation of one-to-one matching games with transfers. Fox (2010a) identified the *sign* of the cross-derivative of the production function at different pairs of input values using a stochastic analog of the deterministic theory result that any pairwise stable assignment will maximize the sum of production from all matches in the economy (e.g., Shapley and Shubik (1972), Becker (1973), Roth and Sotomayor (1990), Jackson (2008)). We discuss other connections to prior work below (see, especially, Section 3).

[4]Since the first drafts of this paper appeared in 2004, several researchers have studied problems related to those we explore here. Graham (2011) provided a recent review.

The econometric approach taken here builds on the partial mean literature (e.g., Newey (1994), Linton and Nielsen (1995)). In this literature, one first estimates a regression function nonparametrically. In the second stage, the regression function is averaged, possibly after some weighting with a known or estimable weight function, over some of the regressors. Here we also estimate the production function nonparametrically as the conditional mean of the outcome given the observed inputs. In the second stage, the averaging is over the distribution of the regressors induced by the new assignment rule. This typically involves the original marginal distribution for some of the regressors, but a different conditional distribution for others. Complications arise because this conditional covariate distribution may be degenerate, which will affect the rate of convergence for the estimator. In addition, the conditional covariate distribution itself may require nonparametric estimation through its dependence on the assignment rule. For the policies we consider, the assignment rule will involve distribution functions and their inverses similar to the way they enter in the changes-in-changes model of Athey and Imbens (2006).

The next section lays out our basic model and identifying assumptions. Section 3 relates our identifying exogeneity assumption to the notion of pairwise stability emphasized in theoretical work on matching. Section 4 then defines and motivates our estimands. Section 5 presents our estimators and derives their large-sample properties for the case where inputs are continuously valued. Section 6 presents the results from a small Monte Carlo exercise. Section 7 presents our conclusions. Appendixes A–C are in a supplementary file on the journal website, http://qeconomics.org/supp/45/supplement.pdf.

2. Model

In this section we present the basic setup and identifying assumptions. For clarity of exposition, we use the production function terminology. For firm i, for $i = 1, \dots, N$, the production function relates a triple of observed inputs (W_i, X_i, V_i) and an unobserved input ε_i to an output Y_i:

$$Y_i = k(W_i, X_i, V_i, \varepsilon_i). \tag{1}$$

The inputs W_i and X_i, and the output Y_i are scalars. The third observed input V_i and the unobserved input ε_i can both be vectors.

We are interested in reallocating the input W across production units. We focus on reallocations that hold the marginal distribution of W fixed. In the educational example, the "firm" would be a classroom. The variable input W would be teacher quality, and X would be a measure of class quality (e.g., the classroom average on an entrance exam). The second characteristic V could include other measures of the class (e.g., its age or gender composition) as elements. In the roommate example, the unit would be the individual, with W being the quality of the roommate (measured by, for example, a high school test score), and the characteristic X would be own quality. The second set of characteristics V could be other characteristics of the dorm or of either of the two roommates such as smoking habits (which may be used by university administrators in the assignment of roommates).

Our key identifying assumption is that conditional on firm characteristics (X, V), the assignment of W, the level of the input to be reallocated, is exogenous:

Assumption 2.1 (Exogeneity). *We have*

$$\varepsilon \perp W | X, V.$$

Let

$$g(w, x, v) = \mathbb{E}[Y | W = w, X = x, V = v] \tag{2}$$

denote expected output conditional on input level w and characteristics x and v. We often refer to the derivative of $g(w, x, v)$ with respect to w, which is denoted by

$$g_W(w, x, v) = \frac{\partial}{\partial w} g(w, x, v). \tag{3}$$

Under exogeneity, we have—in the population of firms with identical values of X and V—an equality between the counterfactual average output that we would observe if all firms in this subpopulation were assigned $W = w$, and the average output we observe for the subset of firms within this subpopulation that are in fact assigned $W = w$. Alternatively, the exogeneity assumption implies that the difference in $g(w, x, v)$ evaluated at two values of w (w_0 and w_1) has a causal interpretation as the average effect of assigning $W = w_1$ rather than $W = w_0$:

$$g(w_1, x, v) - g(w_0, x, v) = \mathbb{E}\big[k(w_1, X, V, \varepsilon) - k(w_0, X, V, \varepsilon) | X = x, V = v\big].$$

3. Exogeneity and pairwise stability

In the context of single agent production models, Assumption 2.1 is often controversial (cf. Heckman, Lalonde, and Smith (2000), Imbens and Wooldridge (2009)). It holds under conditional random assignment of W to units, as would occur in a randomized experiment. Randomized allocation mechanisms are also used by administrators in some institutional settings. For example, some universities match freshman roommates randomly conditional on responses to housing questionnaires (e.g., Sacerdote (2001)). This assignment mechanism is consistent with Assumption 2.1. In other settings, particularly where assignment is bureaucratic, as may be true in some educational settings, a plausible set of conditioning variables may be available.

While "as if" conditional random assignment may hold in some settings of empirical interest, in other settings it may be more appropriate to view input levels as agent choices. Purposeful input selection could generate a violation of Assumption 2.1 and complicate identification arguments (cf. Griliches and Mairesse (1998)). Nevertheless, maintaining Assumption 2.1 represents a natural starting point for studying reallocation effects, since a nonparametric treatment of the production technology is desired.

Extensions of our methods to settings where Assumption 2.1 does not hold would be a natural area for further research.

Importantly, purposeful input choice need not be inconsistent with our exogeneity assumption. An important class of models for which this is true is the aggregate matching setting studied by Choo and Siow (2006a, 2006b). Their work builds on game theoretic treatments of two-sided matching markets with transfers or so-called assignment games (e.g., Koopmans and Beckmann (1957), Shapley and Shubik (1972), Becker (1973)). In these games, an equilibrium matching corresponds to a pairwise stable assignment: no pair of agents can raise net utility by leaving their current partners and forming a new match.

In this section, by means of an extended example, we relate our exogeneity assumption to the pairwise stability equilibrium concept used in the matching literature.[5] While a general treatment of the various issues involved is beyond the scope of the current paper, our example demonstrates that an assignment can both (i) correspond to a pairwise stable equilibrium and (ii) exhibit a high degree of correlation, or assortativeness, between the two inputs, yet nevertheless satisfy Assumption 2.1.

Our example builds on the basic setup of Choo and Siow (2006a, 2006b), which they used to study "marriage markets" as pioneered by Becker (1973). We consider a single matching market composed of two large populations. Associated with each agent in the two populations is a vector of observed, discretely valued characteristics. The support of this characteristic vector defines a finite set of agent types. Across observationally identical agents (i.e., those of the same type), the utility attached to matching with each type in the opposing population is heterogeneous. Agents may transfer utility to one another. These transfers, which are unobserved, adjust to equilibrate the market. This setup generates a two-sided model of (stochastic) multinomial choice subject to a market clearing condition (cf. Graham (2011)).

While Choo and Siow (2006a, 2006b) made a number of strong distributional assumptions, they attractively assumed that the observed assignment corresponds to a pairwise stable matching equilibrium (cf. Fox (2010a)).[6]

Let i index firms with (for simplicity) the binary observable characteristic $W_i \in \mathbb{W} = \{w_L, w_H\}$. Let j index workers with binary observable characteristic $X^j \in \mathbb{X} = \{x_L, x_H\}$. The ith firm's *realized* utility or "profit" from matching with a random draw from the subpopulation of type l workers is given by, for $l \in \{L, H\}$,

$$\Pi_i(x_l) = U_i(x_l) - C_i^l, \tag{4}$$

where $U_i(x_l)$ is the firm's match output and C_i^l is a firm-specific unobserved cost-of-matching shifter. Note that while $\Pi_i(x_l)$ varies with the *type* of worker chosen, it does not vary with the *specific* worker chosen. From the vantage of a given firm, workers of the same type are perfect substitutes (cf. Chiappori, Salanié, and Weiss (2010), Galichon and Salanié (2011)).

[5] Roth and Sotomayor (1990) and Burkard, Dell'Amico, and Martello (2009), respectively, provided economics and operations research oriented surveys of this literature.

[6] See Chiappori, Salanié, and Weiss (2010), Galichon and Salanié (2011), and Graham (2011) for various extensions of the basic Choo and Siow (2006a, 2006b) model.

We assume that the firm's match output is given by

$$U_i(x_l) = \delta(W_i, x_l) - \tau(W_i, x_l) - \nu_i^l, \tag{5}$$

where $\delta(w_k, x_l)$ is the systematic output associated with a k-to-l match, $\tau(w_k, x_l)$ is the (equilibrium) transfer/wage a type k firm must pay to a type l worker (transfers may be negative), and ν_i^l is an output shock that captures firm-level heterogeneity in productivity.[7]

We assume that firms observe their own type, W_i, and the cost shifters $\mathbf{C}_i = (C_i^L, C_i^H)'$ when choosing workers. They do not observe $\boldsymbol{\nu}_i = (\nu_i^L, \nu_i^H)'$. Instead, they observe the signal S_i (which may be vector-valued). This signal is used to forecast $\boldsymbol{\nu}_i$, but conditional on S_i, neither W_i nor $\mathbf{C}_i$ has additional predictive power for $\boldsymbol{\nu}_i$ (i.e., $\boldsymbol{\nu}_i \perp (W_i, \mathbf{C}_i)|S_i$). This redundancy condition will be most plausible when S_i is a good proxy for $\boldsymbol{\nu}_i$ (cf. Wooldridge (2005)). Firm i's expected utility from matching with a type l worker is then

$$\mathbb{E}\big[\Pi_i(x_l)|W_i, S_i, \mathbf{C}_i\big] = \delta(W_i, x_l) - \tau(W_i, x_l) - \pi_l(S_i) - C_i^l,$$

with $\pi_l(S_i) = \mathbb{E}[\nu_i^l|S_i]$. Firms choose worker type to maximize *expected* utility, treating the equilibrium vector of transfers/wages as fixed. The probability that a type $W_i = w_L$ firm matches with a type $X_i = x_H$ worker is, therefore, given by

$$\begin{aligned}
&\Pr(X_i = x_H|W_i = w_L) \\
&\quad = \Pr\big(\mathbb{E}\big[\Pi_i(x_H)|W_i, S_i, \mathbf{C}_i\big] \geq \mathbb{E}\big[\Pi_i(x_L)|W_i, S_i, \mathbf{C}_i\big]\big) \\
&\quad = F_L\big(\delta(w_L, x_H) - \delta(w_L, x_L) - \big[\tau(w_L, x_H) - \tau(w_L, x_L)\big]\big),
\end{aligned}$$

where $F_L(\cdot)$ is the conditional distribution function of $\pi_H(S_i) - \pi_L(S_i) + C_i^H - C_i^L$ given $W_i = w_L$.

The probability that a type $W_i = w_H$ firm matches with a type $X_i = x_H$ worker is similarly given by

$$\begin{aligned}
&\Pr(X_i = x_H|W_i = w_H) \\
&\quad = F_H\big(\delta(w_H, x_H) - \delta(w_H, x_L) - \big[\tau(w_H, x_H) - \tau(w_H, x_L)\big]\big),
\end{aligned}$$

where $F_H(\cdot)$ is the conditional distribution function of $\pi_H(S_i) - \pi_L(S_i) + C_i^H - C_i^L$ given $W_i = w_H$.

Now consider the worker side of the matching process.[8] The jth worker's realized utility from matching with a random draw from the subpopulation of type k firms is given by, for $k \in \{L, H\}$,

$$\Xi^j(w_k) = Z^j(w_k) - D_k^j, \tag{6}$$

[7]Because, from the perspective of firms, workers of the same type are perfect substitutes (and likewise for firms from the workers' standpoint; see below), equilibrium transfers will vary with firm and worker type alone (cf. Graham (2011)).

[8]We use subscripts to index units drawn from the population of firms and use superscripts for units drawn from the population of workers.

where

$$Z^j(w_k) = \tau(w_k, X^j) - \xi_k^j \tag{7}$$

is the worker's match output, which consists of her wage, $\tau(w_k, x_l)$, and a worker-specific shock, ξ_k^j. The worker-specific cost-of-matching shifter is D_k^j. We assume that while the worker observes $\mathbf{D}^j = (D_L^j, D_H^j)$ prior to matching, she does not observe $\boldsymbol{\xi}^j = (\xi_L^j, \xi_H^j)$; instead, she observes the signal T^j. We assume that $\boldsymbol{\xi}^j$ is conditionally independent of $\mathbf{D}^j$ and X^j given T^j. Let $\omega_l(T^j) = \mathbb{E}[\xi_l^j | T^j]$.

The probability that, respectively, type $X^j = x_L$ and $X^j = x_H$ workers match with a type $W^j = w_H$ firm are

$$\begin{aligned}
&\Pr(W^j = w_H | X^j = x_L) = G_L(\tau(w_H, x_L) - \tau(w_L, x_L)), \\
&\Pr(W^j = w_H | X^j = x_H) = G_H(\tau(w_H, x_H) - \tau(w_L, x_H)),
\end{aligned}$$

where $G_L(\cdot)$ and $G_H(\cdot)$ are the conditional distribution functions of $\omega_H(T^j) - \omega_L(T^j) + D_H^j - D_L^j$ given, respectively, $X^j = x_L$ and $X^j = x_H$.

Transfers adjust so that in equilibrium the number of $W_i = w_H$ type firms wishing to match with a $X^j = x_H$ type worker (*demand*) coincides with the number of $X^j = x_H$ type workers wishing to match with a $W_i = w_H$ type firm (*supply*). Such equalities need to hold for all four types of matches. Within k-to-l cells, matching is at random, because firms (workers) are indifferent across workers (firms) of the same type.

The equilibrium must also be feasible. Let p_L denote the fraction of firms of type $W_i = w_L$ and let q_L denote the fraction of workers of type $X^j = x_L$. Let r_{kl} denote the share of k-to-l matches in equilibrium. Graham, Imbens, and Ridder (2007) showed that in the simple 2×2 case, all *feasible* assignments are indexed by r_{LL}.[9] Graham (2011) further showed that (under regularity conditions) the *equilibrium* assignment satisfies

$$\begin{aligned}
&F_H^{-1}\left(\frac{1 - p_L - q_L + r_{LL}}{1 - p_L}\right) - F_L^{-1}\left(\frac{p_L - r_{LL}}{p_L}\right) \\
&\quad + G_H^{-1}\left(\frac{1 - p_L - q_L + r_{LL}}{1 - p_L}\right) - G_L^{-1}\left(\frac{q_L - r_{LL}}{q_L}\right) \\
&= \delta(w_H, x_H) - \delta(w_H, x_L) - \left[\delta(w_L, x_H) - \delta(w_L, x_L)\right].
\end{aligned} \tag{8}$$

The right-hand side of (8) is a measure of systematic complementarity between firm and worker type; it is a discrete analog of a cross-partial derivative. When it is positive, the systematic component of the production technology exhibits so-called increasing differences (Topkis (1998)). The value of r_{LL} that solves (8) is increasing in the degree of complementarity in production. *Firms and workers will endogenously match in an assortative fashion when the systematic component of production exhibits increasing differences.*

[9]Given r_{LL}, the two marginal constraints ($p_L = r_{LL} + r_{LH}$, $q_L = r_{LL} + r_{HL}$) and the adding up constraint ($1 = r_{LL} + r_{LH} + r_{HL} + r_{HH}$) determine the frequency of the other three types of matches. Feasibility also requires that r_{LL} satisfy certain inequality constraints; we ignore those here (cf. Graham, Imbens, and Ridder (2007), Graham (2011)).

Let $n(i) = j$ if the ith firm matches with the jth worker in equilibrium. The sum of (5) and (7), or total match output, is

$$\begin{aligned}
Y_i &= U_i\big(X^{n(i)}\big) + Z^{n(i)}(W_i) \\
&= \delta\big(W_i, X^{n(i)}\big) - \big(\nu_i^L\big)^{\mathbf{1}(X^{n(i)}=x_L)}\big(\nu_i^H\big)^{\mathbf{1}(X^{n(i)}=x_H)} \\
&\quad - \big(\xi_L^{n(i)}\big)^{\mathbf{1}(W_i=w_L)}\big(\xi_H^{n(i)}\big)^{\mathbf{1}(W_i=w_H)}.
\end{aligned}$$

Note that Y_i excludes the firm- and worker-specific cost-of-matching shifters, and hence differs from total match surplus/utility (the sum of (4) and (6)). This is analogous to a farmer's interest in profit and the econometrician's interest in the production technology (e.g., Chamberlain (1984)).

Available is a random sample $\{Y_i, W_i, X^{n(i)}, S_i, T^{n(i)}\}_{i=1}^N$ from a population of equilibrium matches (i.e., a population that satisfies (8)). Consider the conditional mean

$$\begin{aligned}
&\mathbb{E}\big[Y_i|W_i = w_H, X^{n(i)} = x_H, S_i = s, T^{n(i)} = t\big] \\
&\quad = \delta(w_H, x_H) - \mathbb{E}\big[\nu_i^H|W_i = w_H, X^{n(i)} = x_H, S_i = s, T^{n(i)} = t\big] \\
&\qquad - \mathbb{E}\big[\xi_H^{n(i)}|W_i = w_H, X^{n(i)} = x_H, S_i = s, T^{n(i)} = t\big].
\end{aligned}$$

Using the firm's choice rule, the assumption of random matching within k-by-l cells, and conditional independence of $(W_i, \mathbf{C}_i)$ and $\boldsymbol{\nu}_i$ given S_i, we get[10]

$$\mathbb{E}\big[\nu_i^H|W_i = w_H, X^{n(i)} = x_H, S_i = s, T^{n(i)} = t\big] = \mathbb{E}\big[\nu_i^H|S_i = s\big] = \pi_H(s). \tag{10}$$

[10] By random matching within k-by-l cells, we have the conditional independence relationship

$$f_{\boldsymbol{\nu},\mathbf{C},S,\boldsymbol{\xi},\mathbf{D},T|W,X}(\boldsymbol{\nu}, \mathbf{c}, s, \boldsymbol{\xi}, \mathbf{d}, t|w, x) = f_{\boldsymbol{\nu},\mathbf{C},S|W,X}(\boldsymbol{\nu}, \mathbf{c}, s|w, x) f_{\boldsymbol{\xi},\mathbf{D},T|W,X}(\boldsymbol{\xi}, \mathbf{d}, t|w, x).$$

This implies that

$$f_{\boldsymbol{\nu},\mathbf{C},S,T|W,X}(\boldsymbol{\nu}, \mathbf{c}, s, t|w, x) = f_{\boldsymbol{\nu},\mathbf{C},S|W,X}(\boldsymbol{\nu}, \mathbf{c}, s|w, x) f_{T|W,X}(t|w, x)$$

and hence that

$$f_{\boldsymbol{\nu}|\mathbf{C},S,T,W,X}(\boldsymbol{\nu}|\mathbf{c}, s, t, w, x) = \frac{f_{\boldsymbol{\nu},\mathbf{C},S,W,X}(\boldsymbol{\nu}, \mathbf{c}, s, w, x) f_{T|W,X}(t|w, x)}{f_{\mathbf{C},S,T,W,X}(\mathbf{c}, s, t, w, x)}. \tag{9}$$

Using the assumption of within-cell random matching, the denominator of (9) factors as

$$f_{\mathbf{C},S,T,W,X}(\mathbf{c}, s, t, w, x) = f_{\mathbf{C},S|W,X}(\mathbf{c}, s|w, x) f_{T|W,X}(t|w, x) f_{W,X}(w, x).$$

Rearranging (9), we get

$$f_{\boldsymbol{\nu}|\mathbf{C},S,T,W,X}(\boldsymbol{\nu}|\mathbf{c}, s, t, w, x) = f_{\boldsymbol{\nu}|\mathbf{C},S,W,X}(\boldsymbol{\nu}|\mathbf{c}, s, w, x).$$

Using the fact that the distribution of X_i is degenerate conditional on $\mathbf{C}_i$, S_i, and W_i, we further have

$$f_{\boldsymbol{\nu}|\mathbf{C},S,T,W,X}(\boldsymbol{\nu}|\mathbf{c}, s, t, w, x) = f_{\boldsymbol{\nu}|\mathbf{C},S,W}(\boldsymbol{\nu}|\mathbf{c}, s, w).$$

Finally, conditional independence of $\boldsymbol{\nu}_i$ and $(W_i, \mathbf{C}_i)$ given S_i yields

$$f_{\boldsymbol{\nu}|\mathbf{C},S,T,W,X}(\boldsymbol{\nu}|\mathbf{c}, s, t, w, x) = f_{\boldsymbol{\nu}|S}(\boldsymbol{\nu}|s).$$

By a similar argument, we also get

$$\mathbb{E}\big[\xi_H^{n(i)}|W_i=w_H, X^{n(i)}=x_H, S_i=s, T^{n(i)}=t\big]=\mathbb{E}\big[\xi_H^{n(i)}|T^{n(i)}=t\big]=\omega_H(t) \tag{11}$$

and hence

$$\begin{aligned}
&\mathbb{E}\big[Y_i|W_i=w_H, X^{n(i)}=x_H, S_i=s, T^{n(i)}=t\big] \\
&\quad=\delta(w_H,x_H)+\pi_H(s)+\omega_H(t) \\
&\quad=g(w_H,x_H,v)
\end{aligned} \tag{12}$$

for $v=(s',t')'$ as required.

The above example provides an example where input complementarity and utility maximization drive agents to "endogenously" match in an assortative fashion, yet Assumption 2.1 nevertheless holds. We conjecture that several features of our example likely apply more generally. First, it seems essential that the criterion function used by the agents when choosing match partners and the outcome of interest to the econometrician do not coincide. Second, informational assumptions are key. If $\mathbf{C}_i$ had predictive power for $\boldsymbol{\nu}_i$ given the signal S_i, then the equality in (10) would no longer hold and our exogeneity requirement would fail. Similar model features are important when evaluating the appropriateness of exogeneity assumptions in the context of single agent models (e.g., Heckman, Smith, and Clements (1997), Imbens (2004)).[11]

In actual applications, identification conditions will need to be carefully justified. The estimation and inference results we outline below are specific to an approach based on Assumption 2.1. This will be credible in some settings and not in others. In settings where it is not, other assumptions may be invoked to achieve (albeit perhaps partial) identification. For example, nonparametric instrumental variables methods might be used to identify the production function (e.g., Matzkin (2008), Imbens and Newey (2009)).

4. Aggregate redistributional effects

Much of the treatment effect literature (e.g., Angrist and Krueger (1999), Heckman, Lalonde, and Smith (2000), Manski (1990), Imbens and Wooldridge (2009)) has focused on the average effect of an increase in the value of the treatment. In particular, in the binary treatment case ($w\in\{0,1\}$), interest has centered on the average treatment effect

$$\mathbb{E}\big[g(1,X,V)-g(0,X,V)\big].$$

With continuous inputs, one may be interested in the full average output function $g(w,x,v)$ (Imbens (2000), Flores (2005)) or in its derivative with respect to the input

$$g_W(w,x,v)$$

[11] A classic and very elegant example is given by Chamberlain's (1984) application of strict exogeneity in the context of a panel data analysis of farm production. In Chamberlain's example, the farmer maximizes profit, while the econometrician studies output. Furthermore, the farmer makes input decisions prior to observing non-forecastable "weather."

at a point, or a weighted average

$$\mathbb{E}\big[\omega(W,X,V)\cdot g_W(W,X,V)\big].$$

See Powell, Stock, and Stoker (1989) or Hardle and Stoker (1989) for estimands of this type.

Here we are interested in a fundamentally different class of estimands. We focus on policies that redistribute the input W according to a rule based on the X characteristic of the unit. There are at least two limitations to our approach. First, we focus on comparing specific assignment rules, rather than searching for an optimal assignment rule. The latter problem is particularly demanding with continuously valued inputs as the optimal assignment for each unit depends on the characteristics of that unit as well as on the marginal distribution of characteristics in the population. When the inputs are discretely valued, both the problems of inference for a specific rule as well as the problem of finding the optimal rule become considerably more tractable. In that case, any rule that corresponds to a joint distribution of the inputs is characterized by a finite number of parameters. Maximizing estimated average output over all evaluated rules will then generally lead to the optimal rule. Graham, Imbens, and Ridder (2007) and, motivated by an early version of the current paper, Bhattacharya (2009), provided a discussion for the case with discrete covariates.

A second limitation is that the class of assignment rules we consider leaves all aspects of the marginal distribution of the inputs unchanged. This latter restriction is perfectly appropriate in cases where the inputs are indivisible, as, for example, in the social interactions and educational examples. In other cases, one need not be restricted to such assignment rules. A richer class of estimands would allow for assignment rules that maintain some aspects of the marginal distribution of inputs, but not others. An interesting class consists of assignment rules that maintain the average (and thus total) level of the input, but allow for its arbitrary distribution across units. This can be interpreted as assignment rules that “balance the budget.” In such cases, one might assign the maximum level of the input to some subpopulation and the minimum level of the input to the remainder of the population. Finally, one may wish to consider arbitrary decision rules where each unit can be assigned any level of the input within a set. In that case, interesting questions include the optimal assignment rule as a function of unit-level characteristics as well as average outcomes of specific assignment rules. In the binary treatment case, such problems have been studied by Dehejia (2005), Manski (2004), and Hirano and Porter (2009), among others.

Let $f_{W|X,V}(w|x,v)$ denote the conditional distribution of W given (X,V) in the data and let $\tilde{f}_{W|X,V}(w|x,v)$ denote a potentially different conditional distribution. We allow $\tilde{f}_{W|X,V}(w|x,v)$ to correspond to any distribution such that the implied marginal distribution for W remains unchanged, or

$$\int \tilde{f}_{W|X,V}(w|x,v)f_{X,V}(x,v)\,\mathrm{d}v\,\mathrm{d}x=\int f_{W|X,V}(w|x,v)f_{X,V}(x,v)\,\mathrm{d}v\,\mathrm{d}x$$

for all $w\in\mathbb{W}$. This includes degenerate conditional distributions. In general, we are interested in the average outcome that would result from the current distribution of

(X, V, ε) if the distribution of W given (X, V) were changed from its current distribution, $f_{W|X,V}(w|x, v)$ to $\tilde{f}_{W|X,V}(w|x, v)$. We denote the expected output given such a reallocation by

$$\beta_{\tilde{f}}^{\text{are}} = \int g(w, x, v)\tilde{f}_{W|X,V}(w|x, v) f_{X,V}(x, v)\,\mathrm{d}w\,\mathrm{d}x\,\mathrm{d}v. \tag{13}$$

In the next two subsections, we discuss some specific choices for $\tilde{f}(\cdot)$.

4.1 *Positive and negative assortive matching allocations*

The first estimand we consider is the expected average outcome given perfect assortative matching of W on X conditional on V,

$$\beta^{\text{pam}} = \mathbb{E}\big[g\big(F_{W|V}^{-1}\big(F_{X|V}(X|V)\big|V\big), X, V\big)\big], \tag{14}$$

where $F_{X|V}(X|V)$ denotes the conditional cumulative distribution function (CDF) of X given V and where $F_{W|V}^{-1}(q|V)$ is the qth quantile (for $q \in [0, 1]$) associated with the conditional distribution of W given V. Therefore, $F_{W|V}^{-1}(F_{X|V}(X|V)|V)$ computes a unit's location on the conditional CDF of X given V and reassigns it the corresponding quantile of the conditional distribution of W given V. Thus, among units with the same realization of V, those with the highest value of X are reassigned the highest value of W and so on.

For β^{pam} to be well defined, we need some conditions on the joint distribution of (Y, W, X, V). We do not state these conditions here explicitly. When we discuss estimation, in Section 5, we provide conditions for consistent estimation, including compact support and smooth distributions for (W, X), and moment conditions for the conditional distribution of Y given (W, X).[12] These conditions imply that β^{pam} is well defined.

The focus on reallocations within subpopulations defined by V, as opposed to population-wide reallocations, is motivated by the fact that the average outcome effects of such reallocations solely reflect complementarity or substitutability between W and X. To see why this is the case, consider the alternative estimand

$$\beta^{\text{pam-pop}} = \mathbb{E}\big[g\big(F_W^{-1}\big(F_X(X)\big), X, V\big)\big]. \tag{15}$$

This gives average output associated with population-wide perfect assortative matching of W on X. If, for example, X and V are correlated, then this reallocation, in addition to altering the joint distribution of W and X, will alter the joint distribution of W and V. Say V is also a scalar and is positively correlated with X. Population-wide positive assortative matching will induce perfect rank correlated between W and X, but it will also affect the degree of correlation between W and V. This complicates the interpretation of the estimand when $g(w, x, v)$ is nonseparable in w and v, as well as in w and x.

[12] Our formal distribution theory is developed in the setting without V.

An example helps to clarify the issues involved. Let W denote an observable measure of teacher quality, let X denote mean (beginning-of-year) achievement in a classroom, and let V denote the fraction of the classroom that is female. If beginning-of-year achievement varies with gender (say, with classes with a higher fraction of girls having higher average achievement), then X and V will be correlated. A reallocation that assigns high quality teachers to high achievement classrooms will also tend to assign such teachers to classrooms with an above average fraction of females. Average achievement increases observed after implementing such a reallocation may reflect complementarity between teacher quality and beginning-of-year student achievement or it may be that the effects of changes in teacher quality vary with gender and that, conditional on gender, there is no complementarity between teacher quality and achievement. By focusing on re-allocations of teachers across classrooms with similar gender mixes, but varying baseline achievement, (14) provides a more direct avenue to learning about complementarity between W and X.[13]

Both (14) and (15) may be policy relevant, depending on the circumstances, and both are identified under Assumption 2.1 and additional support conditions (which we make explicit below). Under the additional assumption that

$$g(w, x, v) = g_1(w, x) + g_2(v),$$

the estimands, although associated with different reallocations, also have the same basic interpretation. In the current paper, we focus on (14), although it is conceptually straightforward to extend our results to (15).

Our second estimand is the average outcome given negative assortative matching:

$$\beta^{\mathrm{nam}} = \mathbb{E}\big[g\big(F_{W|V}^{-1}\big(1 - F_{X|V}(X|V)\big|V\big), X, V\big)\big]. \tag{16}$$

If, within subpopulations homogeneous in V, the two inputs W and X are everywhere complements, then the difference $\beta^{\mathrm{pam}} - \beta^{\mathrm{nam}}$ provides a measure of the strength of input complementarity. When $g(\cdot)$ is not supermodular (i.e., its cross-derivative is not everywhere positive), the interpretation of this difference is not straightforward. In Theorem 4.1 below, we present a measure of "local" (relative to the status quo allocation) complementarity between X and W that is interpretable in such settings.

4.2 *Correlated matching allocations*

The perfect positive and negative assortative allocations are focal allocations, being emphasized in the economic theory literature (e.g., Becker and Murphy (2000), Legros and Newman (2002, 2007)). There are many more possible allocations. Two others that are particularly important are the status quo allocation, and the random matching allocation. Average output under the status quo allocation is given by

$$\beta^{\mathrm{sq}} = \mathbb{E}[Y] = \mathbb{E}\big[g(W, X, V)\big].$$

[13]We make the connection to complementarity more explicit in Section 4.3.

Average output under the random matching allocation is given by

$$\beta^{\mathrm{rm}} = \int_v \left[\int_x \int_w g(w, x, v) f_{W|V}(w|v) f_{X|V}(x|v) \right] f_V(v) \, \mathrm{d}w \, \mathrm{d}x \, \mathrm{d}v.$$

This last estimand gives average output when W and X are independently assigned within subpopulations indexed by V.

These allocations are just four among the class of feasible allocations. This class comprises all joint distributions of inputs consistent with fixed marginal distributions (within subpopulations homogeneous in V). As noted in the Introduction, if the inputs are continuously distributed, this class of joint distributions is very large. For this reason, we only consider a subset of these joint distributions. To be specific, we concentrate on a family of the feasible allocations, indexed by two parameters τ and ρ, that includes as special cases the negative and positive assortative matching allocations, the independent allocation, and the status quo allocation. Let $\beta^{\mathrm{cm}}(\tau, \rho)$ denote average output under the allocation indexed by τ and ρ. By changing the two parameters, we trace out a "path" in two directions: farther from or closer to the status quo allocation and farther from or closer to the perfect sorting allocation. Borrowing a term from the literature on copulas, we call this class of feasible allocations comprehensive, because it contains all four focal allocations as a special case. For ease of exposition, we focus in the remainder of the paper on the case with no covariates beyond W and X, and so drop the argument V in the production function.[14]

To facilitate estimation, the correlated matching allocations are defined using a truncated bivariate normal copula. The truncation ensures that the denominator in the weights of the correlated matching ARE are bounded from 0, so that we do not require trimming to ensure that our estimand has a nonzero information bound (cf. Khan and Tamer (2010)). The bivariate standard normal probability density function (PDF) is

$$\phi(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(1/(2(1-\rho^2)))(x_1^2 - 2\rho x_1 x_2 + x_2^2)}, \quad -\infty < x_1, x_2 < \infty,$$

with a corresponding joint cumulative distribution function (CDF) denoted by $\Phi(x_1, x_2; \rho)$. Observe that

$$\begin{aligned} &\Pr(-c < x_1 \leq c, -c < x_2 \leq c) \\ &\quad = \Phi(c, c; \rho) - \Phi(c, -c; \rho) - \big[\Phi(-c, c; \rho) - \Phi(-c, -c; \rho)\big], \end{aligned}$$

so that the truncated standard bivariate normal PDF is given by

$$\begin{aligned} &\phi_c(x_1, x_2; \rho) \\ &\quad = \frac{\phi(x_1, x_2; \rho)}{\Phi(c, c; \rho) - \Phi(c, -c; \rho) - [\Phi(-c, c; \rho) - \Phi(-c, -c; \rho)]}, \quad -c < x_1, x_2 \leq c. \end{aligned}$$

Denote the truncated bivariate CDF by Φ_c.

[14] The distributional results stated below continue to hold if we maintain conditioning on V with (mostly) minor modification (e.g., replacing marginal with conditional densities in the appropriate places, etc.). A complete proof of Theorem 5.3 with explicit conditioning on V is available from the authors.

The truncated normal bivariate CDF gives a comprehensive copula, because the corresponding joint CDF

$$H_{W,X}(w,x)=\Phi_c\bigl(\Phi_c^{-1}\bigl(F_W(w)\bigr),\Phi_c^{-1}\bigl(F_X(x)\bigr);\rho\bigr)$$

has marginal CDFs equal to $H_{W,X}(w,\infty)=F_W(w)$ and $H_{W,X}(\infty,x)=F_X(x)$, it reaches the upper and lower Fréchet bounds on the joint CDF for $\rho=1$ and $\rho=-1$, respectively, and it has independent W, X as a special case for $\rho=0$.

To define $\beta^{\mathrm{cm}}(\rho,\tau)$, we note that the joint PDF associated with $H_{W,X}(w,x)$ equals

$$\begin{aligned}h_{W,X}(w,x)&=\phi_c\bigl(\Phi_c^{-1}\bigl(F_W(w)\bigr),\Phi_c^{-1}\bigl(F_X(x)\bigr);\rho\bigr)\\&\qquad\times\frac{f_W(w)f_X(x)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))}.\end{aligned}$$

Then we define $\beta^{\mathrm{cm}}(\rho,0)$ in terms of the truncated normal as

$$\begin{aligned}\beta^{\mathrm{cm}}(\rho,0)=\int_{w,x}&g(w,x)\frac{\phi_c(\Phi_c^{-1}(F_W(w)),\Phi_c^{-1}(F_X(x));\rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))}\\&\times f_W(w)f_X(x)\,\mathrm{d}w\,\mathrm{d}x.\end{aligned}\tag{17}$$

Average output under the correlated matching allocation is given by

$$\begin{aligned}\beta^{\mathrm{cm}}(\rho,\tau)&=\tau\cdot\mathbb{E}[Y]+(1-\tau)\cdot\beta^{\mathrm{cm}}(\rho,0)\\&=\tau\cdot\mathbb{E}[Y_i]\\&\quad+(1-\tau)\int_{x,w}g(w,x)\frac{\phi_c(\Phi_c^{-1}(F_W(w)),\Phi_c^{-1}(F_X(x));\rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))}\\&\quad\times f_W(w)f_X(x)\,\mathrm{d}w\,\mathrm{d}x\end{aligned}\tag{18}$$

for $\tau\in[0,1]$ and $\rho\in(-1,1)$.

The case with $\tau=1$ corresponds to the status quo:

$$\beta^{\mathrm{sq}}=\beta^{\mathrm{cm}}(\rho,1).$$

The case with $\tau=\rho=0$ corresponds to the random matching:

$$\beta^{\mathrm{rm}}=\beta^{\mathrm{cm}}(0,0)=\int_x\int_w g(w,x)\,\mathrm{d}F_W(w)\,\mathrm{d}F_X(x).$$

The cases with $(\tau=0,\rho\to 1)$ and $(\tau=0,\rho\to -1)$ correspond, respectively, to the perfect positive and negative assortative matching allocations:

$$\beta^{\mathrm{pam}}=\lim_{\rho\to 1}\beta^{\mathrm{cm}}(\rho,0)\quad\text{and}\quad\beta^{\mathrm{nam}}=\lim_{\rho\to -1}\beta^{\mathrm{cm}}(\rho,0).$$

More generally, with $\tau=0$, we allocate the inputs using a normal copula in a way that allows for arbitrary correlation between W and X indexed by the parameter ρ. It would be conceptually straightforward to use other copulas.

4.3 *Local measures of complementarity*

An empirical challenge posed by the correlated matching family of estimands $\beta^{\mathrm{cm}}(\rho, \tau)$, including the focal allocations β^{pam} and β^{nam}, is that the support requirements that allow for their precise estimation may be difficult to satisfy in practice. This is particularly relevant for allocations "distant" from the status quo. For example, if the status quo is characterized by a high degree of correlation between the inputs, evaluating the effect of allocations with a small or even negative correlation between inputs, such as random matching or negative assortative matching, can be difficult because such allocations rely on knowledge of the production function at pairs of input values (W, X) that are infrequently seen in the data. In such situations, a measure of complementarity between W and X in the vicinity of the status quo may be of interest, despite its more limited nature.

To this end, we next characterize the expected effect on output associated with a small increase toward either positive or negative assortative matching. This estimand should be most informative regarding the effects of "modest" policies (i.e., those that induce reallocations that stay close to the status quo). We derive this local measure by considering matching on a family of transformations of X and W, indexed by a scalar parameter λ, where for some values of λ, the matching is on W (corresponding to the status quo), and for other values of λ, the matching is on X or $-X$, corresponding to positive and negative assortative matching, respectively. We then focus on the derivative of the expected outcomes from matching on this family of transformations, evaluated at the value of λ that corresponds to the status quo.

For technical reasons, and to be consistent with the subsequent formal statistical analysis in Section 5 of the previously discussed estimands β^{pam} and β^{nam}, we assume that the support of X is the interval $[x_l, x_u]$ with midpoint $x_m = (x_u + x_l)/2$, and assume similarly that the support of W is the interval $[w_l, w_u]$ with midpoint $w_m = (w_u + w_l)/2$. Without loss of generality, we assume that $x_l = 0$, $x_m = 1/2$, $x_u = 1$, $w_l = 0$, $w_m = 1/2$, and $w_u = 1$. We continue to ignore the presence of additional covariates V. First define a smooth function $d(w)$ that goes to zero at the boundary of the support of W:

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

This function plays the role of a fixed weight/trimming function and ensures that the estimand introduced below has a finite semiparametric variance bound.[15]

We implement our local reallocation as follows: for $\lambda \in [-1, 1]$, define the random variable U_λ as a transformation of (X, W):

$$U_\lambda = \lambda \cdot X \cdot d(W)^{1-|\lambda|} + \left(\sqrt{1-\lambda^2}\right) \cdot W.$$

This gives us a parametric transformation of (W, X) that moves smoothly between $W = U_0$ and $X = U_1$. Now we consider reallocations based on positive assortative matching on U_λ, for a range of values of λ, as a smooth way to move from the status quo (matching on W) to positive assortative matching (matching on X). For general λ, the average output associated with positive assortative matching on U_λ is given by

$$\beta^{\mathrm{lr}}(\lambda) = \mathbb{E}\left[g\left(F_W^{-1}\left(F_{U_\lambda}(U_\lambda)\right), X\right)\right]. \tag{19}$$

[15] See Newey and Stoker (1993) for a related discussion in the context of weighted average derivatives.

For $\lambda = 0$ and $\lambda = 1$, we have $U_\lambda = W$ and $U_\lambda = X$, respectively, and hence $\beta^{\mathrm{lr}}(0) = \beta^{\mathrm{sq}}$ and $\beta^{\mathrm{lr}}(1) = \beta^{\mathrm{pam}}$. Perfect negative assortative matching is also nested in this framework since

$$\Pr(-X \le -x) = \Pr(X \ge x) = 1 - F_X(x),$$

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

We focus on the effect of a small reallocation as our local measure of complementarity:

$$\beta^{\mathrm{lc}} = \frac{\partial \beta^{\mathrm{lr}}}{\partial \lambda}(0). \tag{20}$$

This local complementarity measure has two interesting alternative representations, which are given in the following theorem. Before stating this result, we introduce one assumption. This assumption is stronger than needed for this theorem, but its full force will be used later. The required values of the parameters in this assumption, p and q, will be specified in the theorems.

Assumption 4.1 (Distribution of Data).

(i) *The vectors* $(Y_1, W_1, X_1), (Y_2, W_2, X_2), \dots, (Y_N, W_N, X_N)$ *are independent and identically distributed.*

(ii) *The support of* W *is* $\mathbb{W} = [w_l, w_u]$, *a compact subset of* $\mathbb{R}$.

(iii) *The support of* X *is* $\mathbb{X} = [x_l, x_u]$, *a compact subset of* $\mathbb{R}$.

(iv) *The joint probability density function of* W *and* X *is bounded and bounded away from zero, and* q *times continuously differentiable on* $\mathbb{W} \times \mathbb{X}$.

(v) *The function* $g(w, x)$ *is* q *times continuously differentiable with respect to* w *and* x *on* $\mathbb{W} \times \mathbb{X}$.

(vi) *The conditional expectation* $\mathbb{E}[|Y_i|^p | X_i = x]$ *is bounded.*[16]

The first representation is as the expected value of the conditional (on W) covariance of X and the returns to W, $g_W(w, x) = \frac{\partial g}{\partial w}(w, x)$, weighted by $d(W)$. The second representation is as a weighted average of the cross-derivative $\frac{\partial^2 g}{\partial w \partial x}(w, x)$. The following theorem formalizes these concepts.

Theorem 4.1. *Suppose Assumption* 4.1 *holds with* $q \ge 2$. *Then* β^{lc} *has two equivalent representations,*

$$\beta^{\mathrm{lc}} = \mathbb{E}\big[d(W) \cdot \mathrm{Cov}\big(g_W(W, X), X | W\big)\big] \tag{21}$$

[16] If we maintain conditioning on V, the above assumption would need to be modified to ensure that our four estimands are well defined. In particular, $f_{W|V=v}$ would need to be strictly positive on $\mathbb{W}$ for all $v \in \mathbb{V}$. A similar restriction would need to hold for $f_{X|V=v}$. We also require that the support of $(W, X)|V$ coincides with the product of its marginals (conditional on V).

and

$$\beta^{\mathrm{lc}} = \mathbb{E}\left[\delta(W, X) \cdot \frac{\partial^2 g}{\partial w \, \partial x}(W, X)\right], \tag{22}$$

where the weight function $\delta(w, x)$ *is nonnegative and has the form*

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

The proofs for the theorems given in the body of the text are presented in Appendix B in the supplementary file.

Representation (21), as we demonstrate below, suggests a straightforward nonparametric approach to estimating β^{lc}. Representation (22) is valuable for interpretation.

Equation (22) demonstrates that rejecting the null $H_0 : \beta^{\mathrm{lc}} = 0$ implies that there is at least one perturbation of the status quo assignment that raises average outcomes. Consequently, such a rejection implies that the status quo is inefficient (or rather not output maximizing). If $\beta^{\mathrm{lc}} > 0$, then in the vicinity of the status quo, W and X are complements; if $\beta^{\mathrm{lc}} < 0$, they are substitutes. The precise meaning of the "vicinity of the status quo" is implicit in the form of the weight function $\delta(w, x)$.

Deviations of β^{lc} from zero imply that the status quo allocation does not maximize average outcomes. For $\beta^{\mathrm{lc}} > 0$, a shift toward positive assortative matching will raise average outcomes, while for $\beta^{\mathrm{lc}} < 0$, a shift toward negative assortative matching will do so. Theorem 4.1 therefore provides the basis of a test of the null hypothesis that the status quo allocation is locally efficient. Of course, acceptance of the above null does not mean there is no output raising perturbation of the status quo (i.e., the test may have low power in some directions).

5. Estimation and inference with continuously valued inputs

In this section, we discuss estimation and inference. The estimators are all (variants of) weighted averages of (derivatives of) nonparametric estimates of the regression function. These are what Newey (1994) called full and partial means and derivatives. First, in Section 5.1, we describe our nonparametric regression function estimator. We use a new nonparametric kernel estimator introduced in Imbens and Ridder (2009). The usual Nadaraya–Watson kernel regression estimator has a uniform rate of convergence on the internal region of support of the conditioning variable (e.g., Newey (1994)). In the boundary region of support, the estimator exhibits additional bias, resulting in a slower rate of convergence. Newey (1994) dealt with this issue by introducing a fixed trimming function into his partial mean estimand. This function ensures that averaging only occurs across observations that lie on the interior of the support. In our setting, fixed trimming methods are unattractive because they change the nature of the estimands. The nonparametric regression estimator of Imbens and Ridder (2009) includes a correction for bias in the boundary region of support, leading to a uniform convergence rate across the entire support of the conditioning variable.

Next, in Section 5.2, we present estimators for the first pair of estimands, β^{pam} and β^{nam}. In Section 5.3, we discuss estimation and inference for β^{cm} (including β^{rm}), and in Section 5.4, we discuss β^{lc}. Estimation of and inference for the status quo allocation β^{sq} is straightforward. As this estimand is a simple expectation, it is consistently estimable by a sample average.

5.1 *Estimating the production and distribution functions*

For the two distributions functions, we use the empirical distribution functions

$$\hat{F}_W(w) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{W_i \leq w} \quad \text{and} \quad \hat{F}_X(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i \leq x}.$$

For the inverse distribution functions, we use the definition

$$\hat{F}_W^{-1}(q) = \inf_{w\in\mathbb{W}} \mathbf{1}_{\hat{F}_W(w)\geq q} \quad \text{and} \quad \hat{F}_X^{-1}(q) = \inf_{x\in\mathbb{X}} \mathbf{1}_{\hat{F}_X(x)\geq q}.$$

The estimands we consider in this paper depend on the regression function $g(w,x)$ (in the case of β^{pam}, β^{nam}, and β^{cm}) or its derivative in the case of β^{lc}. The latter also depends on the regression function $m(w)$, defined as

$$m(w) = \mathbb{E}[X|W=w]. \tag{23}$$

To estimate these objects, we need estimators for the regression functions $m(w)$ and $g(w,x)$, and the derivative $g_W(w,x)$. Write the regression function as

$$g(w,x) = \mathbb{E}[Y|W=w, X=x] = \frac{h_2(w,x)}{h_1(w,x)},$$

where

$$h_1(w,x) = f_{WX}(w,x) \quad \text{and} \quad h_2(w,x) = g(w,x)\cdot f_{WX}(w,x).$$

To simplify the following discussion, we rewrite $h_1(w,x)$ and $h_2(w,x)$ as

$$h_m(w,x) = \mathbb{E}[\tilde{Y}_m|W=w, X=x]\cdot f_{WX}(w,x) \tag{24}$$

for $m=1,2$, where $\tilde{Y} = (\tilde{Y}_1\ \tilde{Y}_2)'$, with $\tilde{Y}_1 = 1$ and $\tilde{Y}_2 = Y$.

We focus on estimators for $h_m(w,x)$, and use those to estimate $g(w,x)$ and its derivatives. The standard Nadaraya–Watson (NW) estimator for $h_m(w,x)$ is, for some bivariate kernel $K(\cdot,\cdot)$,

$$\hat{h}_{\text{nw},m}(w,x) = \frac{1}{N\cdot b^2}\sum_{i=1}^{N}\tilde{Y}_{im}\cdot K\left(\frac{W_i - w}{b}, \frac{X_i - x}{b}\right). \tag{25}$$

We denote the resulting nonparametric estimator by $\hat{g}(w,x)$.

Because the support of (W,X) is assumed to be bounded, we have to deal with boundary bias of the kernel estimators. Because we also need bias reduction, by using higher order kernels, we adopt the nearest interior point (NIP) estimator of Imbens and Ridder (2009). This estimator divides, for given bandwidth b, the support of (W, X) into an internal region and a boundary region. On the internal region, the uniform convergence of the standard NW kernel estimators holds, but the estimators must be modified on the boundary region of the support. The NIP estimator coincides with the usual NW kernel estimator on the internal set, but it is equal to a polynomial on the boundary set. The coefficients of this polynomial are those of a Taylor series expansion around a point of the internal set.

To obtain a compact expression for the NIP estimator, we adopt the following notation. The vector $z = (w\ x)'$ has $L = 2$ components. Some of the results below are stated for general L, although we only use the case with $L = 2$. Let $\mathbb{Z} = \mathbb{W} \times \mathbb{X}$ denote the (compact) support of Z. Let λ denote an L vector of nonnegative integers, with $|\lambda| = \sum_{l=1}^{L} \lambda_l$ and $\lambda! = \prod_{l=1}^{L} \lambda_l!$. For L vectors of nonnegative integers λ and μ, let $\mu \leq \lambda$ be equivalent to $\mu_l \leq \lambda_l$ for all $l = 1, \dots, L$, and define

$$\binom{\lambda}{\mu} = \frac{\lambda!}{\mu!(\lambda - \mu)!} = \prod_{l=1}^{L} \frac{\lambda_l!}{\mu_l!(\lambda_l - \mu_l)!} = \prod_{l=1}^{L} \binom{\lambda_l}{\mu_l}.$$

For L vectors λ and z, let $z^\lambda = \prod_{l=1}^{L} z_l^{\lambda_l}$. As shorthand for partial derivatives of some function g, we use $g^{(\lambda)}(z)$:

$$g^{(\lambda)}(z) = \frac{\partial g^{|\lambda|}}{\partial z^\lambda}(z). \tag{26}$$

The definition of the internal region depends on the support of the kernel. Let $K : \mathbb{R}^L \mapsto \mathbb{R}$ denote the kernel function. We assume that $K(u) = 0$ for $u \notin \mathbb{U}$ with $\mathbb{U}$ compact and that $K(u)$ is bounded. For the bandwidth b, define the internal set of the support $\mathbb{Z}$ as the subset of $\mathbb{Z}$ such that all $\tilde{z}$ with a distance of up to b times the support of the kernel from z are also in $\mathbb{Z}$:

$$\mathbb{Z}_b^I = \left\{ z \in \mathbb{Z} \middle| \left\{ \tilde{z} \in \mathbb{R}^L \middle| \frac{z - \tilde{z}}{b} \in \mathbb{U} \right\} \subset \mathbb{Z} \right\}. \tag{27}$$

This is a compact subset of the interior of $\mathbb{Z}$ that contains all points that are sufficiently far away from the boundary that the standard kernel density estimator at those points is not affected by any potential discontinuity of the density at the boundary. If $\mathbb{U} = [-1, 1]^L$ and $\mathbb{Z} = \bigotimes_{l=1}^{L} [z_{ll}, z_{ul}]$, we have $\mathbb{Z}_b^I = \bigotimes_{l=1}^{L} [z_{ll} + b, z_{ul} - b]$.[17] The complement of the internal region is the boundary region

$$\mathbb{Z}_b^B = \mathbb{Z} / \mathbb{Z}_b^I = \left\{ z \in \mathbb{Z} \middle| \exists \tilde{z} \notin \mathbb{Z} \text{ s.t. } \frac{z - \tilde{z}}{b} \in \mathbb{U} \right\}. \tag{28}$$

Next, we need to develop some notation for Taylor series approximations. Define for a given q times differentiable function $g : \mathbb{Z} \mapsto \mathbb{R}$, a point $r \in \mathbb{R}^L$, and an integer $s \leq q$,

[17]The set $[-1, 1]^L$ is the set of L vectors with components that are between -1 and 1. The set $\bigotimes_{l=1}^{L} [z_{ll}, z_{ul}]$ is the set of L vectors with the lth component between z_{ll} and z_{ul}.

the $(s-1)$th order polynomial function $t:\mathbb{Z} \mapsto \mathbb{R}$, which is based on the Taylor series expansion, of order $s-1$, of $g(z)$ around the point $r \in \mathbb{Z}$:

$$t(z; g, r, s) = \sum_{j=0}^{s-1} \sum_{|\lambda|=j} \frac{1}{\lambda!} \cdot g^{(\lambda)}(r) \cdot (z-r)^{\lambda}. \tag{29}$$

Because the function $g(z)$ is $q \geq s$ times continuously differentiable on $\mathbb{Z}$, the remainder term in the Taylor series expansion is

$$g(z) - t(z, g, r, s) = \sum_{|\lambda|=s} \frac{1}{\lambda!} g^{(\lambda)}\big(\bar{r}(z)\big) \cdot (z-r)^{\lambda},$$

with $\bar{r}(z)$ intermediate between z and r. Because $\mathbb{Z}$ is compact and the sth order is continuous, the sth order derivative must be bounded and, therefore, this remainder term is bounded by $C|z-r|^s$. For the NIP estimator, we use this Taylor series expansion around a point that depends on z and the bandwidth. Specifically, we take the expansion around $r_b(z)$, the projection on the internal region:

$$r_b(z) = \operatorname*{arg\,min}_{r \in \mathbb{Z}_b^I} \|z - r\|. \tag{30}$$

With this preliminary discussion, the NIP estimator of order s of $h_m(z)$ can be defined as

$$\hat{h}_{m,\text{nip},s}(z) = \sum_{j=0}^{s-1} \sum_{|\lambda|=j} \frac{1}{\lambda!} \cdot \hat{h}_{m,\text{nw}}^{(\lambda)}\big(r_b(z)\big)\big(z - r_b(z)\big)^{\lambda}, \tag{31}$$

with $\hat{h}_{m,\text{nw}}^{(\lambda)}$ the λth derivative of the kernel estimator $\hat{h}_{m,\text{nw}}$. For values of z in the internal region $\mathbb{Z}_b^I$, the NIP estimator is identical to the NW kernel estimator, $\hat{h}_{m,\text{nip},s}(z) = \hat{h}_{m,nw}(z)$. It is only in the boundary region that a $(s-1)$th order Taylor series expansion is used to address the poor properties of the NW estimator in that region.

Now the NIP estimator for $g(w, x)$ is

$$\hat{g}_{\text{nip},s}(w, x) = \frac{\hat{h}_{2,\text{nip},s}(w, x)}{\hat{h}_{1,\text{nip},s}(w, x)}, \tag{32}$$

and the NIP estimator for the first derivative of $g(w, x)$ with respect to w is

$$\widehat{\frac{\partial g_{\text{nip},s}}{\partial w}}(w, x) = \frac{\frac{\partial}{\partial w}\hat{h}_{2,\text{nip},s}(w, x)}{\hat{h}_{1,\text{nip},s}(w, x)} - \frac{\hat{h}_{2,\text{nip},s}(w, x) \cdot \frac{\partial}{\partial w}\hat{h}_{1,\text{nip},s}(w, x)}{(\hat{h}_{1,\text{nip},s}(w, x))^2}. \tag{33}$$

Unlike the NW kernel estimator, the NIP estimator is uniformly consistent. Its properties are discussed in more detail in Imbens and Ridder (2009). Formal statements of the relevant properties for our discussion are given in Lemmas A.9, A.10, and A.11, and Theorems A.1, A.2, and A.3 in Appendix A.

In the remainder of the paper, we drop the subscripts from the estimator of the regression function. Unless specifically mentioned, $\hat{g}(w,x)$ is used to denote $\hat{g}_{\mathrm{nip},s}(w,x)$ for s equal to the order of the kernel, with its value stated in the lemmas and theorems.

Next we introduce two more assumptions. Assumption 5.1 describes the properties of the kernel function, and Assumption 5.2 gives the rate on the bandwidth. Before stating the next assumption, we need to introduce a class of restrictions on kernel functions. The restrictions govern the rate at which the kernel, which is assumed to have compact support, goes to zero on the boundary of its support. This property allows us to deal with some of the boundary issues. Such properties have previously been used in, for example, Powell, Stock, and Stoker (1989).

Definition 5.1 (Derivative Order of a Kernel). A kernel function $K:\mathbb{U}\mapsto\mathbb{R}$ is of derivative order d if, for all u in the boundary of the set $\mathbb{U}$ and all $|\lambda|\le d-1$,

$$\lim_{v\to u}\frac{\partial^\lambda}{\partial u^\lambda}K(v)=0.$$

Assumption 5.1 (Kernel). *The kernel satisfies*:

(i) $K:\mathbb{R}^L\mapsto\mathbb{R}$, *with* $K(u)=\prod_{l=1}^L\mathcal{K}(u_l)$.

(ii) $K(u)=0$ *for* $u\notin\mathbb{U}$, *with* $\mathbb{U}=[-1,1]^L$.

(iii) $K(\cdot)$ *is r times continuously differentiable, with the rth derivative bounded on the interior of* $\mathbb{U}$.

(iv) $K(\cdot)$ *is a kernel of order s, so that* $\int_{\mathbb{U}}K(u)\,\mathrm{d}u=1$ *and* $\int_{\mathbb{U}}u^\lambda K(u)\,\mathrm{d}u=0$ *for all* λ *such that* $0<|\lambda|<s$ *for some* $s\ge 1$.

(v) K *is a kernel of derivative order* d.

We refer to a kernel that satisfies Assumption 5.2 as a derivative kernel of order (s,d).

Assumption 5.2 (Bandwidth). *The bandwidth* $b_N=N^{-\delta}$ *for some* $\delta>0$.

5.2 *Estimation and inference for* $\widehat{\beta}^{\mathrm{pam}}$ *and* $\widehat{\beta}^{\mathrm{nam}}$

In this section, we introduce the estimators for β^{pam} and β^{nam}, and present results on the large-sample properties of the estimators. We estimate β^{pam} and β^{nam} by substituting nonparametric estimators for the unknown functions $g(w,x)$, $F_W(w)$, and $F_X(x)$:

$$\widehat{\beta}^{\mathrm{pam}}=\frac{1}{N}\sum_{i=1}^N\hat{g}\bigl(\hat{F}_W^{-1}\bigl(\hat{F}_X(X_i)\bigr),X_i\bigr)\tag{34}$$

and

$$\widehat{\beta}^{\mathrm{nam}}=\frac{1}{N}\sum_{i=1}^N\hat{g}\bigl(\hat{F}_W^{-1}\bigl(1-\hat{F}_X(X_i)\bigr),X_i\bigr).\tag{35}$$

It is straightforward to demonstrate consistency for these estimators. The nonparametric estimators $\hat{g}$, $\hat{F}_W$, and $\hat{F}_X$ are uniformly consistent under our assumptions, and consistency of $\hat{\beta}^{\text{pam}}$ follows directly from that. It is more difficult to derive the large-sample distributions for these estimators. There are four components to their asymptotic approximations. Here we discuss the decomposition for $\hat{\beta}^{\text{pam}}$. A similar argument holds for $\hat{\beta}^{\text{nam}}$. In both cases, the first component corresponds to the estimation error in $g(w, x)$. This component converges at a rate slower than the regular parametric (root-N) rate. This is because, in the first stage, we estimate a nonparametric regression function with more arguments than we average over in the second stage. As a result, $\hat{\beta}^{\text{pam}}$ (and $\hat{\beta}^{\text{nam}}$) is a partial (as opposed to a full) mean in the terminology of Newey (1994). The other three terms converge faster, that is, at the regular root-N rate. There is one term each that corresponds to the estimation error in $F_W(w)$ and $F_X(x)$, respectively, and one each that corresponds to the difference between the average of $g(F_W^{-1}(F_X(X_i)), X_i)$ and its expectation. In describing the large-sample properties, we include all four of these terms, which leaves a remainder that is $o_p(N^{-1/2})$. In principle, one could ignore the three terms of $O_p(N^{-1/2})$, since they will become dominated by the term that describes the uncertainty stemming from estimation of $g(w, x)$, but including the additional terms is likely to lead to more accurate confidence intervals.[18] We provide evidence for this in the simulations in Section 6.

To describe the formal properties of the estimator $\hat{\beta}^{\text{pam}}$, it is useful to introduce notation for an intermediate quantity and some additional functions. Define the average with the true regression function $g(w, x)$ (but still the estimated distribution functions $\hat{F}_W$ and $\hat{F}_X$),

$$\tilde{\beta}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N} g\big(\hat{F}_W^{-1}\big(\hat{F}_X(X_i)\big), X_i\big), \tag{36}$$

so that we can write $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} = (\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}}) + (\tilde{\beta}^{\text{pam}} - \beta^{\text{pam}})$. Then the first term $\tilde{\beta}^{\text{pam}} - \beta^{\text{pam}} = O_p(N^{-1/2})$ and the second term $\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}} = O_p(N^{-1/2} b_N^{-1/2})$. Recall the notation for the derivative of $g(w, x)$ with respect to w,

$$g_W(w, x) = \frac{\partial g}{\partial w}(w, x),$$

and define

$$\begin{aligned}
q^{\text{pam}}(w, x) &= \frac{g_W(F_W^{-1}(F_X(x)), x)}{f_W(F_W^{-1}(F_X(x)))} \cdot \big(\mathbf{1}_{F_W(w) \le F_X(x)} - F_X(x)\big), \\
\psi_W^{\text{pam}}(w) &= \mathbb{E}\big[q^{\text{pam}}(w, X)\big], \\
r^{\text{pam}}(x, z) &= \frac{g_W(F_W^{-1}(F_X(z)), z)}{f_W(F_W^{-1}(F_X(z)))} \cdot \big(\mathbf{1}_{x \le z} - F_X(z)\big),
\end{aligned}$$

[18]Note that in the presence of additional inputs V, we would be required to construct estimates of the conditional CDFs $F_{W|V}(w|v)$ and $F_{X|V}(x|v)$. Such estimates would also converge to their population values at a slower than root-N rate and hence contribute to the asymptotic sampling variance of $\hat{\beta}^{\text{pam}}$.

and

$$\psi_X^{\text{pam}}(x) = \mathbb{E}\bigl[r^{\text{pam}}(x, X)\bigr].$$

Theorem 5.1 (Large Sample Properties of $\hat{\beta}^{\text{pam}}$). *Suppose Assumptions* 2.1, 4.1, 5.1, *and* 5.2 *hold, with* $q \geq 2s+1$, $r \geq s+3$, $p \geq 4$, $d \geq s-1$, *and* $1/(2s) < \delta < 1/8$. *Then*

$$\sqrt{N} \cdot \begin{pmatrix} b_N^{1/2}(\hat{\beta}^{\text{pam}} - \tilde{\beta}^{\text{pam}}) \\ \tilde{\beta}^{\text{pam}} - \beta^{\text{pam}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11}^{\text{pam}} & 0 \\ 0 & \Omega_{22}^{\text{pam}} \end{pmatrix}\right),$$

where

$$\begin{aligned} \Omega_{11}^{\text{pam}} = \mathbb{E}\biggl[& \sigma^2\bigl(F_W^{-1}\bigl(F_X(X)\bigr), X_i\bigr) \\ & \cdot \int_{u_1}\left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W(F_W^{-1}(F_X(X)))} \cdot u_2, u_2\right) \mathrm{d}u_2\right)^2 \mathrm{d}u_1 \\ & \cdot f_{W|X}\bigl(F_W^{-1}\bigl(F_X(X)\bigr)|X\bigr)\biggr] \end{aligned} \tag{37}$$

and

$$\Omega_{22}^{\text{pam}} = \mathbb{E}\bigl[\bigl(\psi_W^{\text{pam}}(W) + \psi_X^{\text{pam}}(X) + g\bigl(F_W^{-1}\bigl(F_X(X)\bigr), X\bigr) - \beta^{\text{pam}}\bigr)^2\bigr].$$

In the expression for the large-sample variance, ψ_X^{pam} captures the uncertainty that results from estimation of $F_X(x)$, and ψ_W^{pam} captures the uncertainty that results from estimation of $F_W(w)$.

Note that the component of the variance that captures the uncertainty from estimation of $g(w, x)$, Ω_{11}^{pam}, depends on the kernel in a way that involves the distribution of the data. Often when one estimates nonparametric functionals at parametric rates, the dependence on the kernel vanishes asymptotically if one undersmoothes. Here the kernel shows up in the leading term. This is also the case in the discussion of partial means in Newey (1994).

Suppose we wish to construct a 95% confidence interval for β^{pam}. In that case, we approximate the variance of $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$ by $\hat{\mathbb{V}} = \hat{\Omega}_{11}^{\text{pam}} \cdot N^{-1} \cdot b_N^{-1} + \hat{\Omega}_{22}^{\text{pam}} \cdot N^{-1}$, using suitable plug-in estimators $\hat{\Omega}_{11}^{\text{pam}}$ and $\hat{\Omega}_{22}^{\text{pam}}$, and construct the confidence interval as $(\hat{\beta}^{\text{pam}} - 1.96 \cdot \sqrt{\hat{\mathbb{V}}}, \hat{\beta}^{\text{pam}} + 1.96 \cdot \sqrt{\hat{\mathbb{V}}})$. Although the first term in $\hat{\mathbb{V}}$ dominates the second term in large samples, in finite samples, the second term may still be important. We shall see this in the simulations in Section 6.

Similar results hold for β^{nam}, with some appropriately redefined terms:

$$\tilde{\beta}^{\text{nam}} = \frac{1}{N}\sum_{i=1}^{N} g\bigl(\hat{F}_W^{-1}\bigl(1 - \hat{F}_X(X_i)\bigr), X_i\bigr),$$

$$q^{\text{nam}}(w, x) = \frac{g_W(F_W^{-1}(1 - F_X(x)), x)}{f_W(F_W^{-1}(1 - F_X(x)))} \cdot \bigl(\mathbf{1}_{F_W(w) \leq F_X(x)} - F_X(x)\bigr), \tag{38}$$

$$\psi_W^{\text{nam}}(w) = \mathbb{E}\big[q_{WX}^{\text{nam}}(w, X)\big],$$

$$r^{\text{nam}}(x, z) = \frac{g_W(F_W^{-1}(1 - F_X(z)), z)}{f_W(F_W^{-1}(1 - F_X(z)))} \cdot \big(\mathbf{1}_{x \le z} - F_X(z)\big),$$

and

$$\psi_X^{\text{nam}}(x) = \mathbb{E}\big[r_{XZ}^{\text{nam}}(x, X)\big].$$

Theorem 5.2 (Large Sample Properties of $\hat{\beta}^{\text{nam}}$). *Suppose Assumptions* 2.1, 4.1, 5.1, *and* 5.2 *hold, with* $q \ge 2s + 1$, $r \ge s + 3$, $p \ge 4$, $d \ge s - 1$, *and* $1/(2s) < \delta < 1/8$. *Then*

$$\sqrt{N} \cdot \begin{pmatrix} b_N^{1/2}(\hat{\beta}^{\text{nam}} - \tilde{\beta}^{\text{nam}}) \\ \tilde{\beta}^{\text{nam}} - \beta^{\text{nam}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{11}^{\text{nam}} & 0 \\ 0 & \Omega_{22}^{\text{nam}} \end{pmatrix}\right),$$

where

$$\begin{aligned}\Omega_{11}^{\text{nam}} = \mathbb{E}\bigg[&\sigma^2\big(F_W^{-1}\big(1 - F_X(X)\big), X\big) \\ &\cdot \int_{u_1}\left(\int_{u_2} K\left(u_1 + \frac{f_X(X)}{f_W(F_W^{-1}(1 - F_X(X)))} \cdot u_2, u_2\right) \mathrm{d}u_2\right)^2 \mathrm{d}u_1 \\ &\cdot f_{W|X}\big(F_W^{-1}\big(1 - F_X(X)\big)|X\big)\bigg]\end{aligned}$$

and

$$\Omega_{22}^{\text{nam}} = \mathbb{E}\big[\big(\psi_W^{\text{nam}}(W) + \psi_X^{\text{nam}}(X) + g(W, X) - \beta^{\text{nam}}\big)^2\big].$$

5.3 *Estimation and inference for* $\beta^{\text{cm}}(\rho, \tau)$

The starting point for estimation of β^{cm} is the representation of $\beta^{\text{cm}}(\rho, 0)$ in equation (17):

$$\beta^{\text{cm}}(\rho, 0) = \int_{w,x} g(w, x) \frac{\phi_c(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))} f_W(w) f_X(x) \,\mathrm{d}w \,\mathrm{d}x.$$

Note that this expression is an integral over the product of the marginal PDFs of W and X, not the joint. We estimate this by replacing the integrals with sums over the two empirical distribution functions to get the analog estimator

$$\widehat{\beta}^{\text{cm}}(\rho, 0) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{g}(W_i, X_j) \frac{\phi_c(\Phi_c^{-1}(\hat{F}_W(W_i)), \Phi_c^{-1}(\hat{F}_X(X_j)); \rho)}{\phi_c(\Phi_c^{-1}(\hat{F}_W(W_i)))\phi_c(\Phi_c^{-1}(\hat{F}_X(X_j)))}.$$

This estimator would be a standard second order V statistic if we had the true regression function and the true distribution functions. The dependence on the estimated regression function complicates its analysis.

Observe that if $\rho = 0$ (random matching), the ratio of densities on the right-hand side is equal to 1, so that

$$\hat{\beta}^{\mathrm{rm}} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(W_i, X_j).$$

For $\tau > 0$, the $\beta^{\mathrm{cm}}(\rho, \tau)$ estimand is a convex combination of average output under the status quo and a correlated matching allocation. The corresponding sample analog is

$$\widehat{\beta}^{\mathrm{cm}}(\rho, \tau) = \tau \cdot \widehat{\beta}^{\mathrm{sq}} + (1-\tau) \cdot \widehat{\beta}^{\mathrm{cm}}(\rho, 0),$$

where $\widehat{\beta}^{\mathrm{sq}} = \overline{Y} = \sum_{i=1}^{N} Y_i/N$, the average outcome. This estimator is linear in the nonparametric regression estimator $\hat{g}$ and nonlinear in the empirical CDFs of X and W.

An insightful representation of $\beta^{\mathrm{cm}}(\rho, 0)$ is as an average of partial means. This representation provides intuition about both the structure of the estimand and its large-sample properties. Fixing W at $W = w$, but averaging over the distribution of X, we get the partial mean

$$\eta(w; \rho) = \mathbb{E}_X\bigl[g(w, X) \cdot \omega(w, X; \rho)\bigr], \tag{39}$$

where

$$\omega(w, x; \rho) = \frac{\phi_c(\Phi_c^{-1}(F_W(w)), \Phi_c^{-1}(F_X(x)); \rho)}{\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))}. \tag{40}$$

Observe that (39) is a weighted averaged of the production function over the distribution of X, holding the value of the input to be reallocated (W) fixed at $W = w$. The weight function $\omega(w, x)$ depends on the truncated normal copula. In particular, the weights give greater emphasis to realizations of $g(w, X)$ that are associated with values of X that will be assigned a value of W close to w as part of the correlated matching reallocation. Thus (39) equals the average post-reallocation output for those firms being assigned $W = w$. To give a concrete example, (39) is the post-reallocation expected achievement of those classrooms that will be assigned a teacher of quality $W = w$.

Equation (39) also highlights the value of using the truncated normal copula. Doing so ensures that the denominators of the copula "weights" in (39) are bounded from zero.

If we average these partial means over the marginal distribution of W, we get $\beta^{\mathrm{cm}}(\rho, 0)$, since

$$\beta^{\mathrm{cm}}(\rho, 0) = \mathbb{E}_W\bigl[\eta(W; \rho)\bigr],$$

yielding average output under the correlated matching reallocation.

From the above discussion, it is clear that our correlated matching estimator can be viewed as a semiparametric two-step method-of-moments estimator with a moment function of

$$m\bigl(Y, W, \beta^{\mathrm{cm}}(\rho, \tau), \eta(W; \rho)\bigr) = \tau Y + (1-\tau)\eta(W; \rho) - \beta^{\mathrm{cm}}(\rho, \tau).$$

Our estimator, $\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)$, is the feasible generalized method of moments (GMM) estimator based on the above moment function after replacing the partial mean ($\eta(w;\rho)$ defined in (39)) with a consistent estimate. While the above representation is less useful for deriving the asymptotic properties of $\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)$, it does provide some insight as to why we are able to achieve parametric rates of convergence.

To state the large-sample properties of the correlated matching estimator, we need some additional notation. Define

$$\begin{aligned} e_W(w,x) &= \frac{\rho\phi_c(\Phi_c^{-1}(F_W(w)),\Phi_c^{-1}(F_X(x));\rho)}{(1-\rho^2)\phi_c(\Phi_c^{-1}(F_W(w)))^2\phi_c(\Phi_c^{-1}(F_X(x)))} \\ &\quad \times \big[\Phi_c^{-1}\big(F_X(x)\big)-\rho\Phi_c^{-1}\big(F_W(w)\big)\big], \\ e_X(w,x) &= \frac{\rho\phi_c(\Phi_c^{-1}(F_W(w)),\Phi_c^{-1}(F_X(x));\rho)}{(1-\rho^2)\phi_c(\Phi_c^{-1}(F_W(w)))\phi_c(\Phi_c^{-1}(F_X(x)))^2} \\ &\quad \times \big[\Phi_c^{-1}\big(F_W(w)\big)-\rho\Phi_c^{-1}\big(F_X(x)\big)\big], \end{aligned}$$

$$\begin{aligned} \psi_0^{\mathrm{cm}}(y,w,x) &= \big(\mathbb{E}\big[g(W,x)\cdot\omega(W,x)\big]-\beta^{\mathrm{cm}}(\rho,0)\big) \\ &\quad + \big(\mathbb{E}\big[g(w,X)\cdot\omega(w,X)\big]-\beta^{\mathrm{cm}}(\rho,0)\big), \end{aligned} \tag{41}$$

$$\psi_g^{\mathrm{cm}}(y,w,x) = \frac{f_W(w)\cdot f_X(x)}{f_{WX}(w,x)}\big(y-g(w,x)\big)\omega(w,x), \tag{42}$$

$$\psi_W^{\mathrm{cm}}(y,w,x) = \int\int g(s,t)e_W(s,t)\big(\mathbf{1}_{w\le s}-F_W(s)\big)f_W(s)f_X(t)\,\mathrm{d}s\,\mathrm{d}t, \tag{43}$$

and

$$\psi_X^{\mathrm{cm}}(y,w,x) = \int\int g(s,t)e_X(s,t)\big(\mathbf{1}_{x\le t}-F_X(t)\big)f_W(s)f_X(t)\,\mathrm{d}s\,\mathrm{d}t. \tag{44}$$

Theorem 5.3. *Suppose Assumptions* 2.1, 4.1, 5.1, *and* 5.2 *hold with* $q\ge 2s-1$, $r\ge s+1$, $p\ge 3$, $d\ge s-1$, *and* $(1/2s)<\delta<1/4$. *Then*

$$\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)\xrightarrow{p}\beta^{\mathrm{cm}}(\rho,\tau)$$

and

$$\sqrt{N}\big(\widehat{\beta}^{\mathrm{cm}}(\rho,\tau)-\beta^{\mathrm{cm}}(\rho,\tau)\big)\xrightarrow{d}\mathcal{N}\big(0,\Omega^{\mathrm{cm}}\big),$$

where

$$\Omega^{\mathrm{cm}} = \mathbb{E}\big[\big(\tau\big(Y-\beta^{\mathrm{sq}}\big)+(1-\tau)\psi^{\mathrm{cm}}(Y,W,X)\big)^2\big]$$

and

$$\psi^{\mathrm{cm}}(y,w,x) = \psi_0^{\mathrm{cm}}(y,w,x)+\psi_g^{\mathrm{cm}}(y,w,x)+\psi_W^{\mathrm{cm}}(y,w,x)+\psi_X^{\mathrm{cm}}(y,w,x). \tag{45}$$

Note that this estimator is root-N consistent, unlike $\hat{\beta}^{\mathrm{pam}}$ and $\hat{\beta}^{\mathrm{nam}}$.

If there was no estimation error in $\hat{g}(w,x)$, $\hat{F}_W(w)$, and $\hat{F}_X(x)$, the estimator would be root-N consistent with normalized asymptotic variance equal to $\mathbb{E}[(\psi_0^{\text{cm}}(Y_i, W_i, X_i))^2]$. The remaining terms in the influence function, $\psi_W^{\text{cm}}(y,w,x)$, $\psi_X^{\text{cm}}(y,w,x)$, and $\psi_g^{\text{cm}}(y,w,x)$, capture the uncertainty coming from estimation of $F_W(w)$, $F_X(x)$, and $g(w,x)$, respectively.

The proof of Theorem 5.3 is based on a general result for doubly averaged estimands given in Appendix A (Theorem A.3). This result may be of independent interest.

5.4 *Estimation and inference for* β^{lc}

Estimation of β^{lc} proceeds in two steps. First, we estimate $g(w,x) = \mathbb{E}[Y|W=w, X=x]$ (and its derivative with respect to w) and $m(w) = \mathbb{E}[X|W=w]$ using kernel methods as in Section 5.1. In the second step, we estimate β^{lc} by method-of-moments using the sample analog of the moment condition

$$\mathbb{E}\left[\frac{\partial}{\partial w} g(W,X)\cdot d(W)\cdot\big(X - m(W)\big) - \beta^{\text{lc}}\right] = 0.$$

Thus,

$$\hat{\beta}^{\text{lc}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial w}\widehat{g}(W_i, X_i)\cdot d(W_i)\cdot\big(X_i - \widehat{m}(W_i)\big). \tag{46}$$

Define

$$\begin{aligned}\psi_g^{\text{lc}}(y,w,x) = &-\frac{1}{f_{W,X}(w,x)}\frac{\partial f_{W,X}(w,x)}{\partial W} d(w)\big(y - g(w,x)\big)\big(x - m(w)\big)\\ &-\frac{\partial m}{\partial w}(w) d(w)\big(y - g(w,x)\big)\\ &-\frac{\partial d}{\partial w}(w)\big(y - g(w,x)\big)\big(x - m(w)\big)\end{aligned}$$

and

$$\psi_m^{\text{lc}}(y,w,x) = \mathbb{E}\left[\frac{\partial}{\partial w} g(w,X)\middle| W = w\right]\cdot d(w)\cdot\big(x - m(w)\big).$$

As in the previous results, the ψ^{lc} are the influence functions, with $\psi_g^{\text{lc}}(y,w,x)$ capturing the uncertainty from estimation of $g(w,x)$ and $\psi_m^{\text{lc}}(y,w,x)$ capturing the uncertainty from estimation of $m(w)$.

The asymptotic properties of $\hat{\beta}^{\text{lc}}$ are summarized by Theorem 5.4.

Theorem 5.4. *Suppose Assumptions* 2.1, 4.1, 5.1, *and* 5.2 *hold with* $q \geq 2s+1$, $r \geq s+1$, $p \geq 4$, $d \geq s-1$, *and* $1/(2s) < \delta < 1/12$. *Then*

$$\hat{\beta}^{\text{lc}} \xrightarrow{p} \beta^{\text{lc}}$$

and

$$\sqrt{N}(\hat{\beta}^{\mathrm{lc}} - \beta^{\mathrm{lc}}) \xrightarrow{d} \mathcal{N}(0, \Omega^{\mathrm{lc}}),$$

where

$$\Omega^{\mathrm{lc}} = \mathbb{E}\left[\left(\left(\frac{\partial}{\partial w} g(W,X) \cdot d(W) \cdot \big(X - m(W)\big) - \beta^{\mathrm{lc}}\right) + \psi_g^{\mathrm{lc}}(Y,W,X) + \psi_m^{\mathrm{lc}}(Y,W,X)\right)^2\right].$$

Note that β^{lc} is an average over the joint distribution of (W, X), not the product of the two marginals, as is the case for correlated matching rules. Consequently, it is estimable at parametric rates under weaker support conditions.

6. A Monte Carlo study

To assess whether the asymptotic properties derived in Section 5 provide useful approximations to finite sample distributions, we carry out a small simulation study. In the interest of brevity, we focus on β^{pam} and β^{lc}. We consider the following data generating process. The pair (W_i^*, X_i^*) is drawn from a bivariate normal distribution with both means equal to 0, both variances equal to 1, and correlation coefficient equal to ζ. The two covariates W_i and X_i are then constructed as $W_i = 2 \cdot \Phi(W_i^*) - 1$ and $X_i = 2 \cdot \Phi(X_i^*) - 1$, so that both W_i and X_i have a uniform distribution on $[-1, 1]$, with potentially some correlation between them. The outcome is generated as

$$Y_i = W_i + X_i + W_i \cdot X_i + \varepsilon_i, \quad \varepsilon_i | W_i, X_i \sim \mathcal{N}(0, 0.25).$$

Under this data generating process, $\beta^{\mathrm{pam}} = 0.3333$, irrespective of the value of the correlation between the covariates, ζ. The expected outcome under the current allocation is $\mathbb{E}[Y] = 0$ if $\zeta = 0$ and $\mathbb{E}[Y] = 0.1212$ if $\zeta = 0.5$. We fix the weight function $d(w)$ in the definition of the local complementarity measure at $d(w) = 1 - |w|$. The value of the local reallocation parameter is $\beta^{\mathrm{lc}} = 0.1667$ if $\zeta = 0$ and $\beta^{\mathrm{lc}} = 0.1355$ if $\zeta = 0.5$.

We estimate β^{pam} using equation (35) and β^{lc} using equation (46). We use a rectangular kernel on $[-1, 1]$ and local linear regression for estimating $g(w, x)$. The bandwidth for the regression estimation is chosen using (leave-one-out least squares) cross-validation, after which we divide the bandwidth by 2 to ensure some undersmoothing. For density estimation, we use the Silverman rule of thumb, modified for a uniform kernel. For univariate density estimation, this leads to

$$b_N = 1.84 \cdot \sigma \cdot N^{-1/5}.$$

To estimate the bivariate density, we use a bivariate uniform kernel, with the bandwidths in each direction equal to

$$b_N' = 1.84 \cdot \sigma \cdot N^{-1/6},$$

Table 1. Simulation results for $\hat{\beta}^{\text{pam}}$ and $\hat{\beta}^{\text{lc}}$: 10,000 simulations.

	$N = 200$				$N = 1000$			
	$\zeta = 0.0$		$\zeta = 0.5$		$\zeta = 0.0$		$\zeta = 0.5$	
	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$	$\hat{\beta}^{\text{pam}}$	$\hat{\beta}^{\text{lc}}$
Mean bias	−0.009	−0.018	−0.002	−0.016	−0.003	−0.011	−0.000	−0.010
Median bias	−0.010	−0.020	−0.003	−0.017	−0.003	−0.011	−0.001	−0.010
s.d.	0.093	0.039	0.088	0.043	0.040	0.013	0.039	0.013
Ave s.e.	0.085	0.256	0.088	0.578	0.041	0.418	0.044	0.306
Median s.e.	0.085	0.051	0.087	0.064	0.040	0.020	0.044	0.039
r.m.s.e.	0.093	0.043	0.088	0.046	0.040	0.017	0.039	0.016
m.a.e.	0.061	0.027	0.060	0.028	0.028	0.013	0.027	0.012
Coverage rate 90% c.i.	0.871	0.938	0.897	0.959	0.905	0.931	0.935	0.991
Coverage rate 95% c.i.	0.929	0.968	0.947	0.980	0.953	0.965	0.971	0.997

where the σ is estimated on the data and so may differ in the two directions for the bivariate kernel.

We consider four designs, based on two sample sizes, $N = 200$ and $N = 1000$, and two dependence structures, $\zeta = 0$ and $\zeta = 0.5$. For both designs, we calculate the two estimators $\hat{\beta}_{\text{pam}}$ and $\hat{\beta}_{\text{lc}}$, and their variances. In Table 1, we report some summary statistics from the simulations. We report the average and median bias, the standard deviation, the average of the standard errors, the root mean squared error, the median absolute error, and the coverage rates for the nominal 90% and 95% confidence intervals (c.i.s). The estimators appear to work fairly well. Note that the average standard error for $\hat{\beta}^{\text{lc}}$ is large relative to its standard deviation (the ratio is greater than 6). The reason is that occasionally the estimated standard error is very large. This happens with low probability, so the median standard error is not affected, and the coverage rate is also fine.

The estimators have a complicated structure, with the asymptotic distribution relying on a number of approximations. We further investigate these approximations in Table 2. Define

$$\hat{\beta}^{\text{pam}}_{g} = \frac{1}{N}\sum_{i=1}^{N} \hat{g}\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big), \tag{47}$$

$$\hat{\beta}^{\text{pam}}_{W} = \frac{1}{N}\sum_{i=1}^{N} g\big(\hat{F}_W^{-1}\big(F_X(X_i)\big), X_i\big), \tag{48}$$

$$\hat{\beta}^{\text{pam}}_{X} = \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(\hat{F}_X(X_i)\big), X_i\big), \tag{49}$$

and

$$\overline{g}^{\text{pam}} = \frac{1}{N}\sum_{i=1}^{N} g\big(F_W^{-1}\big(F_X(X_i)\big), X_i\big). \tag{50}$$

Table 2. Simulation results: assessing the adequacy of the asymptotic approximations for $\hat{\beta}^{\text{pam}}$ ($N = 1000$, $\zeta = 0.0$).

	$\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$	$\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}$	$\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}$	$\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}$	$\overline{g}^{\text{pam}} - \beta^{\text{pam}}$	Remainder Panel A
Panel A						
Mean	−0.003	−0.002	−0.000	0.001	0.000	−0.001
s.d.	0.040	0.031	0.019	0.019	0.038	0.004
	$\hat{t}_{\text{pam}}$	$\hat{t}_g$	$\hat{t}_W$	$\hat{t}_X$	$\hat{t}_{\overline{g}}$	Nominal
Panel B						
Mean	−0.074	−0.078	−0.015	0.049	0.002	0.000
s.d.	0.989	1.018	1.005	1.018	1.008	1.000
$\Pr(\lvert T\rvert \geq 1.645)$	0.095	0.105	0.102	0.111	0.107	0.100
$\Pr(\lvert T\rvert \geq 1.96)$	0.047	0.055	0.053	0.058	0.052	0.050
$\Pr(T \geq 1.645)$	0.039	0.043	0.052	0.062	0.053	0.050
$\Pr(T \leq -1.645)$	0.056	0.062	0.050	0.049	0.054	0.050
$\Pr(T \geq 1.96)$	0.018	0.021	0.026	0.033	0.027	0.025
$\Pr(T \leq -1.96)$	0.029	0.034	0.027	0.025	0.025	0.025

Then, as stated formally in Appendix A, Lemma A.15,

$$\begin{aligned}\hat{\beta}^{\text{pam}} - \beta^{\text{pam}} &= \left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) \\ &\quad + \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right) + o_p\left(N^{-1/2}\right).\end{aligned} \tag{51}$$

In Panel A of Table 2, we show the mean and standard deviation of $\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}$, $\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}$, $\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}$, $\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}$, and the remainder term,

$$\begin{aligned}\text{rem} &= \left(\hat{\beta}^{\text{pam}} - \beta^{\text{pam}}\right) \\ &\quad - \left(\left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) + \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right)\right).\end{aligned}$$

The results in Panel A of Table 2 suggest that the remainder term is indeed small compared to the terms that are taken into account in the asymptotic distribution. Moreover, the relative magnitude of the $O_p(N^{-1/2})$ terms are supportive of the fact that we take into account these terms, not just the leading term, which is $N^{-1/2}b_N^{-1/2}$.

In the Appendix, we also show that

$$N^{1/2}b_N^{1/2} \cdot \left(\hat{\beta}_g^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_{11}^{\text{pam}}\right), \tag{52}$$

where Ω_{11}^{pam} is defined in (37),

$$N^{1/2} \cdot \left(\hat{\beta}_W^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_W^{\text{pam}}(W)^2\right]\right), \tag{53}$$

$$N^{1/2} \cdot \left(\hat{\beta}_X^{\text{pam}} - \overline{g}^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_X^{\text{pam}}(X)^2\right]\right), \tag{54}$$

and

$$N^{1/2} \cdot \left(\overline{g}^{\text{pam}} - \beta^{\text{pam}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\left(g\left(F_W^{-1}\left(F_X(X_i), X_i\right)\right) - \beta^{\text{pam}}\right)^2\right]\right). \tag{55}$$

To assess the normal approximations, we calculate the t-statistics based on these distributions (the point estimates divided by estimates of the standard deviations) and report in Panel B of Table 2 summary statistics for these random variables, which should have approximate normal distributions. The summary statistics we report are averages, standard deviations, and tail frequencies. We find that the actual means, standard deviations, and tail frequencies are close to the nominal ones from the normal distribution.

7. Conclusions

In this paper, we introduce a new class of estimands that involve reallocation of inputs, and develop statistical methods for analyzing them. We consider a class of problems where a fixed set of inputs is reallocated to a fixed set of units. Whereas a large part of the literature in econometrics has focused on estimating the causal effects of changing inputs for all units or for a subset of units, here we focus on reallocation rules that take into account resource constraints by keeping the distribution of the inputs fixed. The effects we focus on depend critically on the degree of complementarity between inputs. We therefore follow a flexible approach where the nature of the complementarity is not restricted to a parametric or even semiparametric form. We propose estimators for the effects of various reallocation rules and derive the asymptotic properties of these estimators.

Our work could be extended in a number of ways. One direction for future research would involve relaxing our exogeneity assumption. If the production function was left nonparametric, this would lead naturally to a partial identification analysis. Alternatively one could impose a priori restrictions on the production technology, but be less restrictive on input assignment.

A second extension would involve studying optimal assignments. For continuously valued inputs, this is computationally challenging (cf. Chiappori, McCann, and Nesheim (2010)) and likely also challenging in terms of inference (cf. Graham, Imbens, and Ridder (2007)).

Our theorems also invoke strong conditions on the status quo distribution of the inputs. To nonparametrically identify the output effect of a reallocation, we need to be able to identify the production function at all relevant input combinations. It may be that some relevant input combinations are either unobserved or only infrequently observed under the status quo. This could lead to, respectively, a failure of identification or a reduction in the rate of convergence of our estimators (i.e., irregular identification; cf. Khan and Tamer (2010)).

Finally, it would be interesting to apply our proposed methods to a real world empirical problem.

References

Angrist, J. D. and A. B. Krueger (1999), "Empirical strategies in labor economics." In *Handbook of Labor Economics*, Vol. 3A (O. Ashenfelter and D. Card, eds.), 1277–1366, Elsevier Science, New York. [32, 40]

Athey, S. and G. Imbens (2006), "Identification and inference in nonlinear difference-in-differences models." *Econometrica*, 74 (2), 431–497. [34]

Athey, S. and S. Stern (1998), "An empirical framework for testing theories about complementarity in organizational design." Working Paper 6600, NBER. [32]

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011), "A note on semiparametric estimation of finite mixtures of discrete choice modes with an application to game theoretic models." *International Economic Review*, 52 (3), 807–824. [33]

Becker, G. S. (1973), "A theory of marriage: Part I." *Journal of Political Economy*, 81 (4), 813–846. [33, 36]

Becker, G. S. and K. M. Murphy (2000), *Social Economics: Market Behavior in a Social Environment*. Harvard University Press, Cambridge, Massachusetts. [43]

Bhattacharya, D. (2009), "Inferring optimal peer assignment from experimental data." *Journal of the American Statistical Association*, 104 (486), 486–500. [33, 41]

Bhattacharya, D. and P. Dupas (2012), "Inferring welfare maximizing treatment assignment under budget constraints." *Journal of Econometrics*, 167 (1), 168–196. [32]

Brock, W. and S. Durlauf (2001), "Interactions-based models." In *Handbook of Econometrics*, Vol. 5 (J. Heckman and E. Leamer, eds.), 3297–3380, North-Holland, Amsterdam. [32, 33]

Burkard, R., M. Dell'Amico, and S. Martello (2009), *Assignment Problems*. Society for Industrial and Applied Mathematics, Philadelphia. [36]

Chamberlain, G. (1984), "Panel data." In *Handbook of Econometrics*, Vol. 2 (Z. Griliches and M. D. Intriligator, eds.), 1247–1318, North-Holland, Amsterdam. [39, 40]

Chiappori, P. A., R. J. McCann, and L. P. Nesheim (2010), "Hedonic price equilibria, stable matching, and optimal transport: Equivalence, topology and uniqueness." *Economic Theory*, 42 (2), 317–354. [33, 62]

Chiappori, P. A., B. Salanie, and Y. Weiss (2010), "Partner choice and the marital college premium." Report. [36]

Choo, E. and A. Siow (2006a), "Who marries whom and why?" *Journal of Political Economy*, 114 (1), 175–201. [33, 36]

Choo, E. and A. Siow (2006b), "Estimating a matching model with spillover effects." *Demography*, 42 (3), 463–490. [33, 36]

Dehejia, R. (2005), "Program evaluation as a decision problem." *Journal of Econometrics*, 125 (1–2), 141–173. [32, 41]

Flores, C. (2005), "Estimation of dose-response functions and optimal doses with a continuous treatment." Report. [40]

Fox, J. (2010a), "Identification in matching games." *Quantitative Economics*, 1 (2), 203–254. [33, 36]

Fox, J. (2010b), "Estimating matching games with transfers." Report. [33]

Galichon, A. and B. Salanié (2011), "Cupid's invisible hand: Social surplus and identification in matching models." Report. [33, 36]

Graham, B. (2008), "Identifying social interactions through conditional variance restrictions." *Econometrica*, 76 (3), 643–660. [33]

Graham, B., G. Imbens, and G. Ridder (2007), "Redistributive effects for discretely-valued inputs." IEPR Working Paper 07.7, University of Southern California. [33, 38, 41, 62]

Graham, B., G. Imbens, and G. Ridder (2010), "Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach." Working Paper 16499, NBER. [32]

Graham, B. S. (2011), "Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers." In *Handbook of Social Economics*, Vol. 1B (J. Benhabib, A. Bisin, and M. Jackson, eds.), 965–1052, North-Holland, Amsterdam. [30, 31, 33, 36, 37, 38]

Griliches, Z. and J. Mairesse (1998), "Production functions the search for identification." In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centenniel Symposium* (S. Strøm, ed.), 169–203, Cambridge University Press, Cambridge. [31, 35]

Hardle, W. and T. M. Stoker (1989), "Investigating smooth multiple regression by the method of average derivatives." *Journal of the American Statistical Association*, 84 (408), 986–995. [41]

Heckman, J., R. Lalonde, and J. Smith (2000), "The economics and econometrics of active labor market programs." In *Handbook of Labor Economics*, Vol. 3A (O. Ashenfelter and D. Card, eds.), 1865–2097, Elsevier Science, New York. [32, 35, 40]

Heckman, J., J. Smith, and N. Clements (1997), "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts." *Review of Economic Studies*, 64 (4), 487–535. [40]

Hirano, K. and J. Porter (2009), "Asymptotics for statistical treatment rules." *Econometrica*, 77 (5), 1683–1701. [32, 41]

Imbens, G. (2000), "The role of the propensity score in estimating dose-response functions." *Biometrika*, 87 (3), 706–710. [40]

Imbens, G. (2004), "Nonparametric estimation of average treatment effects under exogeneity: A survey." *Review of Economics and Statistics*, 86 (1), 4–30. [40]

Imbens, G. and W. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77 (5), 1481–1512. [31, 40]

Imbens, G. and G. Ridder (2009), "Estimation and inference for generalized full and partial means and derivatives." Unpublished manuscript, Department of Economics, Harvard University. [48, 50, 51]

Imbens, G. and J. Wooldridge (2009), "New developments in the econometrics of program evaluation." *Journal of Economic Literature*, 47 (1), 5–86. [32, 35, 40]

Jackson, M. (2008), *Social and Economic Networks*. Princeton University Press, Princeton, New Jersey. [33]

Kasy, M. (2012), "Partial identification, distributional preferences, and the welfare ranking of policies." Report. [31]

Khan, S. and E. Tamer (2010), "Irregular identification, support conditions, and inverse weight estimation." *Econometrica*, 78 (6), 2021–2042. [44, 62]

Koopmans, T. C. and M. Beckmann (1957), "Assignment problems and the location of economic activities." *Econometrica*, 25 (1), 53–76. [36]

Legros, P. and A. Newman (2002), "Monotone matching in perfect and imperfect worlds." *Review of Economic Studies*, 69 (4), 925–942. [43]

Legros, P. and A. Newman (2007), "Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities." *Econometrica*, 75 (4), 1073–1102. [43]

Linton, O. and J. Nielsen (1995), "A kernel method of estimating structured nonparametric regression based on marginal integration." *Biometrika*, 82 (1), 93–100. [34]

Loeb, S., D. Kalogrides, and T. Béteille (2012), "Effective schools: Teacher hiring, assignment, development, and retention." *Education Finance and Policy*, 7 (3), 269–304. [32]

Manski, C. (1990), "Nonparametric bounds on treatment effects." *American Economic Review*, 80 (2), 319–323. [40]

Manski, C. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60 (3), 531–542. [32, 33]

Manski, C. (2004), "Statistical treatment rules for heterogeneous populations." *Econometrica*, 72 (4), 1221–1246. [32, 41]

Matzkin, R. (2008), "Identification in nonparametric simultaneous equation models." *Econometrica*, 76 (5), 945–978. [31, 40]

Moffitt, R. (2001), "Policy interventions, low-level equilibria and social interactions." In *Social Dynamics* (S. Durlauf and P. Young, eds.), 45–82, MIT Press, Cambridge, Massachusetts. [33]

Newey, W. (1994), "Kernel estimation of partial means and a general variance estimator." *Econometric Theory*, 10 (2), 233–253. [34, 48, 53, 54]

Newey, W. and T. Stoker (1993), "Efficiency of weighted average derivatives and index models." *Econometrica*, 61 (5), 1199–1223. [46]

Powell, J., J. Stock, and T. Stoker (1989), "Semiparametric estimation of index coefficients." *Econometrica*, 57 (6), 1403–1430. [41, 52]

Roth, A. E. and M. Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, Cambridge. [33, 36]

Sacerdote, B. (2001), "Peer effects with random assignment: Results for Dartmouth roommates." *Quarterly Journal of Economics*, 116 (2), 681–704. [32, 35]

Shapley, L. S. and M. Shubik (1972), "The assignment game I: The core." *International Journal of Game Theory*, 1 (1), 111–130. [33, 36]

Stock, J. H. (1989), "Nonparametric policy analysis." *Journal of the American Statistical Association*, 84 (406), 567–575. [33]

Stoye, J. (2009), "Minimax regret treatment choice with finite samples." *Journal of Econometrics*, 151 (1), 70–81. [33]

Tetenov, A. (2012), "Statistical treatment choice based on asymmetric minimax regret criteria." *Journal of Econometrics*, 166 (1), 157–165. [33]

Topkis, D. M. (1998), *Supermodularity and Complementarity*. Princeton University Press, Princeton. [38]

Wooldridge, J. M. (2005), "Unobserved heterogeneity and estimation of average partial effects." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (D. W. K. Andrews and J. H. Stock, eds.), 27–55, Cambridge University Press, Cambridge. [37]

Submitted December, 2009. Final version accepted May, 2013.