Average Redistributional Effects

IFAI/IZA Conference on Labor Market Policy Evaluation

Geert Ridder, Department of Economics, University of Southern California.

October 10, 2006

Motivation

- Most papers in this conference considered on/off or 0-1 or treatment/control interventions.
- Some considered extensions to interventions in which more than 2 treatments are compared.
- In all these interventions there is no restriction on who gets treated, e.g. a typical counterfactual considered is the average outcome if all members of the population are treated and the average outcome if none of them is treated.
- Now consider a situation where there are as many treatment levels as there are individuals (or a distribution of treatment levels in the population) and constraints on treatment assignment.
- A particular case is when the treatment levels are fixed/indivisible. In that case, assigning a particular unit another treatment can only occur if we change the treatment of at least one other unit.
- This is a very simple example of a feedback effect in treatment assignment, i.e. the SUTVA assumption does not hold.
- The constraints on treatment assignment can take different forms, e.g. the treatment levels can be divisible and the total supply, i.e. the sum of the levels, is fixed.
- I will only consider the indivisible case where the treatment assignment matches units to treatments and different treatment assignments are reallocations of treatment levels among units.
- One reason this may be interesting is because these are type of treatments in the presence of peer effects or social interactions, e.g. roommates in college or gender composition of class.
- Literature on social interactions has concentrated on identifying the effect of social interactions (estimating social multiplier), but has not considered the effect of policy interventions.

- Problem considered here is different from that in Manski (2004) and Dehejia (2002) who consider optimal treatment rules (for 0-1 interventions) that are unconstrained (and do not violate SUTVA). Their question: who should be treated to maximize the average outcome in the population.

Overview

- Examples
- Average Redistributional Effect
- Some econometric issues
- Empirical illustration

Examples

Example 1: Production function (Athey and Stern, 1998))

- Notation
  - Y_i is output firm i
  - W_i input of firm i
  - X_i characteristic of firm i
- If i is assigned input w, then output is $Y_i(w)$ (causal).
- The relation between output and w is allowed to depend on X_i, e.g.

$$Y_i(w) = g(w, X_i) \tag{1}$$

- We assume that total supply of input is fixed, and that input is indivisible.
- Question: What is the effect on average output if we reallocate/redistribute inputs among the firms?
- In sequel we focus on average output we could also consider other summary measures or full distribution, e.g., inequality measures.

Example 2: Teacher assignment (Card and Krueger, 1992)

- Output Y_i is (average) test score of class i, W_i is (quality of) teacher assigned to class i, and X_i is 'quality' of class i, as e.g. measured by average test score in previous year.
- If pool of teachers fixed, then we can only consider reallocation.
- Question: What is the effect on the average (over the classes) test score of teacher reallocation.

Example 3: Roommates in college (Sacerdote, 2001)

- Y_i is GPA of student student i, W_i is high school GPA of roommate, X_i is high school GPA of student.
- Note that two inputs W and X act symmetrically.
- Effect of high school on college GPA depends on roommate high school GPA: peer effect.
- Question: What is the effect of roommate reallocation on average GPA?

Example 4: Parents' and child's education (Kremer, 1997)

- W_i is father's years of education and X_i is mother's years of education, and Y_i is child i's years of education.
- Like roommate example.

Example 5: Effect of group composition/diversity (Graham, Imbens, and Ridder, 2006)

- Y_i is test score of class i, and W_i fraction of girls in class i.
- Change the fraction girls in class (towards or away from equal representation c.q. segregation) holding total number of girls and class size distribution constant.
- Also reallocation problem, but under different restrictions, i.e. more like supply constrained divisible problem.

Average Redistributional Effect

- If we consider a large population then a feasible allocation is a joint distribution of inputs W, X over units consistent with given fixed marginal distributions of these inputs.

- Special feasible allocations

  - Current or status quo allocation. We do not assume that the status quo is necessarily optimal, i.e. no model for status quo allocation, except assumption that allows us to estimate the production function in (1).
  - There is small literature on analyzing the observed allocation as the outcome of choice by units (e.g. Fox (2006), Choo and Siow (2006)) and if that choice is strictly on the basis of the outcome variable that would potentially be helpful in identifying treatment effects as in the Roy model.
  - Random allocation, i.e. assignment of W independent of X_i.
  - Positive perfect sorting/assortative matching, i.e. the order of W in the reallocation is equal to that of X_i.
  - Negative perfect sorting/negative assortative matching, i.e. the order of W is the reverse of that of X_i.
  - Comparison of positive assortative and random or negative assortative matching gives a measure of complementarity of inputs.
  - Optimal allocation, i.e. the allocation that maximizes the average outcome.

Average Redistributional Effect (ARE)

- In this paper we do not compute the optimal allocation and do inference for that allocation. Bhattacharya (2006) considers some issues regarding inference on optimal allocations . Computation of and inference for the optimal allocation are feasible if X, W take a fine number of values, e.g. years of education, but not if these variables are continuous, as explained later.

- Instead we consider a (two-parameter) class of allocations, the correlated matching allocations, that has all focal allocations as special case. This is a manageable subset of all allocations.

- ARE is average outcome for a particular correlated matching allocation. It is a treatment effect where the treatment is a re-allocation.

- All feasible allocations give the same average outcome if g is separable, i.e. $g(w, x) = g_1(w) + g_2(x)$.

Special case: X, W binary.

- Population is infinite, but X, W take only two values x_1, x_2 and w_1, w_2.
- An allocation is a joint distribution $r_{XW,ij}, i, j = 1, 2$ such that

$$\begin{aligned} r_{XW,11} + r_{XW,12} &= p_{X,1} && (2) \\ r_{XW,11} + r_{XW,21} &= p_{W,1} && (3) \\ r_{XW,11} + r_{XW,12} + r_{XW,21} + r_{XW,22} &= 1 && (4) \end{aligned}$$

  with $p_{XW,ij}, i, j = 1, 2$ the population, i.e. status quo, distribution.
- The allocations satisfy

$$\begin{aligned} r_{XW,12} &= p_{X,1} - r_{XW,11} \\ r_{XW,21} &= p_{W,1} - r_{XW,11} \\ r_{XW,22} &= 1 - p_{X,1} - p_{W,1} + r_{XW,11} \end{aligned}$$

  and hence are indexed by $r_{XW,11}$.

- Special allocations
  - Status quo allocation with $r_{XW,11} = p_{XW,11}$.
  - Random allocation with $r_{XW,11} = p_{X,1}p_{W,1}$.
  - Positive sorting allocation with $r_{XW,11} = \min\{p_{X,1}, p_{W,1}\}$.
  - Negative sorting allocation with $r_{XW,11} = p_{X,1} - \min\{p_{X,1}, p_{W,2}\}$.
- Optimal allocation maximizes

  $$g(x_1, w_1)r_{XW,11} + g(x_1, w_2)(p_{X,1} - r_{XW,11}) + g(x_2, w_1)(p_{W,1} - r_{XW,11})$$
  $$+g(x_2, w_2)(1 - p_{X,1} - p_{W,1} + r_{XW,11})$$

  and is equal to positive sorting allocation if

  $$g(x_1, w_1) - g(x_1, w_2) - g(x_2, w_1) + g(x_2, w_2) > 0$$

  and to the negative sorting allocation if reverse holds.

Extension to general X, W

- X and W take K: allocations indexed by $(K-1)^2$ parameters.
- Optimal allocation found by linear programming, i.e.

  $$\max \sum_{k=1}^{K} \sum_{l=1}^{K} g(x_k, w_l) r_{XW,kl}$$

  subject to

  $$\sum_{k=1}^{K} r_{XW,kl} = p_{W,l}$$

  $$\sum_{l=1}^{K} r_{XW,kl} = p_{X,k}$$

  $$r_{XW,kl} \geq 0$$

  for $l = 1, \ldots, L$ and $k = 1, \ldots, K$. This is a transportation problem.
- If we estimate the production function, then the optimal allocation is a consistent estimator of the population optimal allocation.
- W, X continuous variables: Allocations are joint distributions of X, W with given marginals. Optimal allocation is solution to infinitely dimensional LP problem (replace summations by integrals).
- Optimal allocations are considered in Imbens, Graham, and Ridder (2006).

Estimands and estimation

- Notation
  - $Y_i(w)$ = output for unit i is assigned $W_i = w$.
  - X_i = characteristic of i used in assignment of W_i (could be other input).
  - Z_i vector of other covariates.
- Data: $Y_i(W_i), W_i, X_i, Z_i$.
- Assumption: Unconfounded/exogenous assignment of W

$$Y(w) \perp W | X, Z$$

- Z contains all variables that affect both assignment of W and $Y(w)$.
- Credibility is same as in the 0-1 treatment case. Consider $Y(w)$ with w educational attainment child if mother has (counterfactual) level of education w and W is observed level of education mother.
- This is assumption on marriage market: matching on X, Z and other variables that are independent of $Y(w)$.
- Under this assumption

$$g(w, x, z,) = \mathrm{E}(Y(w)|W = w, X = x, Z = z) = \mathrm{E}(Y(W)|W = w, X = x, Z = z)$$

  i.e. we can obtain g as the average observed output given $W = w, X = x, Z = z$.

- If there is no Z such that assignment is unconfounded, we could identify g using instruments. 'Strong' instrument needed for non-parametric identification, otherwise bounds.

Estimands (continuous X, W)

- The correlated matching allocations are defined using a truncated bivariate normal cupola

$$\phi_c(x_1, x_2; \rho) = \frac{\phi(x_1, x_2; \rho)}{\Phi(c, c; \rho) - \Phi(c, -c; \rho) - [\Phi(-c, c; \rho) - \Phi(-c, -c; \rho)]}$$

  with $-c < x_1, x_2 \leq c$. Denote the truncated bivariate cdf by Φ_c.

- The truncated normal bivariate distribution gives comprehensive allocations, because the corresponding joint cdf

$$H_{W,X|Z}(w, x|z) = \Phi_c\left(\Phi_c^{-1}(F_{W|Z}(w|z)), \Phi_c^{-1}(F_{X|Z}(x|z)); \rho\right)$$

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

- Note that with additional covariates Z we hold the marginal distributions of $X|Z$ and $W|Z$ fixed. This is necessary, if $g(w, x, z)$ is not separable in z and we want to concentrate on the effect of reallocation of W among units characterized by X, Z We need to hold the relation between W and Z and X and Z constant.

- To obtain an estimate of $\beta^{cm}(\rho, \tau)$ we note

$$\beta^{cm}(\rho, \tau) = \tau \mathrm{E}(Y) +$$

$$(1-\tau) \int_x \int_w \int_z g(w, x, z) \frac{\phi_c\left(\Phi_c^{-1}(F_{W|Z}(w|z)), \Phi_c^{-1}(F_{X|Z}(x|z)); \rho\right)}{\phi_c\left(\Phi_c^{-1}(F_{W|Z}(w|z))\right) \phi_c\left(\Phi_c^{-1}(F_{X|Z}(x|z))\right)}$$

$$\cdot f_{W|Z}(w|z) f_{X|Z}(x|z) f(z) \mathrm{d}w \mathrm{d}x \mathrm{d}z.$$

Special cases

- Status quo assignment

$$\beta^{cm}(\rho, 1) = \mathrm{E}(Y)$$

- Positive perfect sorting

$$\beta^{cm}(1, 0) = \mathbb{E}\left[g\left(F^{-1}_{W|Z}(F_{X|Z}(X|Z)|Z), X, Z\right)\right]$$

- Negative perfect sorting

$$\beta^{cm}(-1, 0) = \mathbb{E}\left[g\left(F^{-1}_{W|Z}(1 - F_{X|Z}(X|Z)|Z), X, Z\right)\right].$$

- Random sorting

$$\beta^{cm}(0, 0) = \int_x \int_w \int_z g(w, x, z) dF_{W|Z}(w|z) dF_{X|Z}(x|z) dF_Z(z).$$

Estimators

- The 'production function' g is estimated nonparametrically. Kernel regression estimator of g

$$\hat{g}(w,x,z) = \sum_{i=1}^{N} y_i \cdot K((v_i - v)/h) \Big/ \sum_{i=1}^{N} K((v_i - v)/h).$$

- Correlated sorting

$$\hat{\beta}^{cm}(\rho,0) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{g}(w_i,x_j,z_j)\frac{\phi\left(\Phi^{-1}(\hat{F}_W(w_i)),\Phi^{-1}(\hat{F}_X(x_j));\rho\right)}{\phi\left(\Phi^{-1}(\hat{F}_W(w_i))\right)\phi\left(\Phi^{-1}(\hat{F}_X(x_j))\right)}$$

- Positive perfect sorting

$$\hat{\beta}^{cm}(1,0) = \frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(\hat{F}_W^{-1}(\hat{F}_X(x_i)),x_i,z_i\right).$$

- Negative perfect sorting

$$\hat{\beta}^{cm}(-1,0) = \frac{1}{N}\sum_{i=1}^{N}\hat{g}\left(\hat{F}_W^{-1}(1-\hat{F}_X(x_i)),x_i,z_i\right).$$

Asymptotic properties of estimators

- Estimators are averages of nonparametric regression estimators over their arguments.
- Relevant asymptotic theory is partial means as in Newey (1994) and Linton and Nielsen (1995).
- $\hat{\beta}^{cm}(1, 0)$ has parametric rate of convergence if we use a higher order kernel and the appropriate bandwidth sequence.
- For $\hat{\beta}^{cm}(1, 0)$ and $\hat{\beta}^{cm}(-1, 0)$ we average over a singular distribution, i.e. W is a function of X, so that the rate of convergence is slower than parametric. This singular case is also the reason that we use a kernel and not a series estimator for g.
- Because we take averages over a compact support, we have to deal with boundary bias in the kernel estimators. Can be dealt with by trimming or by the boundary kernel modification in Imbens and Ridder (2006) that combine kernel and series estimators and applies to densities and their derivatives and accommodates higher-order kernels.

Testing for local reallocation effects

- Correlated sorting gives the effect of reallocations between perfect positive and negative assortative matching.
- Other approach is to consider small reallocations from the status quo in the direction of positive assortative matching.
- Define $W_\rho = \rho X + \sqrt{1-\rho^2} W$, so that if $\rho = 0$, $W_\rho = X$ and if $\rho = 1$, $W_\rho = W$.
- Define

$$\beta(\rho) = \mathbb{E}\left[g(F^{-1}_{W|Z}(F_{W_\rho}(W_\rho|Z)|Z), X, Z) \right].$$

  and consider

$$\gamma = \frac{\partial \beta}{\partial \rho}(0).$$

- We find

$$\gamma = \mathbb{E}\left[\frac{\partial g}{\partial w}(W, X, Z) \cdot (X - \mathbb{E}[X|W, Z]) \right].$$

  (average measure of local complementarity)

- Advantage is that this test is not affected by potential support problems.

Application

- Kremer (1997) considered effect of change in correlation of parents' education on inequality of education among children,
- He specifies a linear relation between parents' and child's education, i.e. no effect of redistribution on average education of children.
- We use data on 10272 children from the NLSY. For now only education of father and mother (X, W) and education of child Y.

- Summary statistics

Table 1: Years of education NLSY; $N = 12272$

	Mean	Std. dev.
Ed. child	13.06	2.38
Ed. mother	11.20	2.87
Ed. father	11.20	3.64

Correlation of parents' education is 0.6.

- Regression of education child on education father and mother, squares and interaction

Table 2: Regression of education of child on education parents; NLSY, $N = 10272$

	Coefficient	Standard err.
Constant	11.27	.19
Ed. mother	-.041	.036
Ed. father	-.077	.029
Ed. mother^2	.011	.0023
Ed. father^2	.011	.0015
Ed. mother $\times$ Ed. father	.0014	.0029
R^2	.22	

- Average education child by education father and mother

Figure 1: Average years of education child by education father and mother

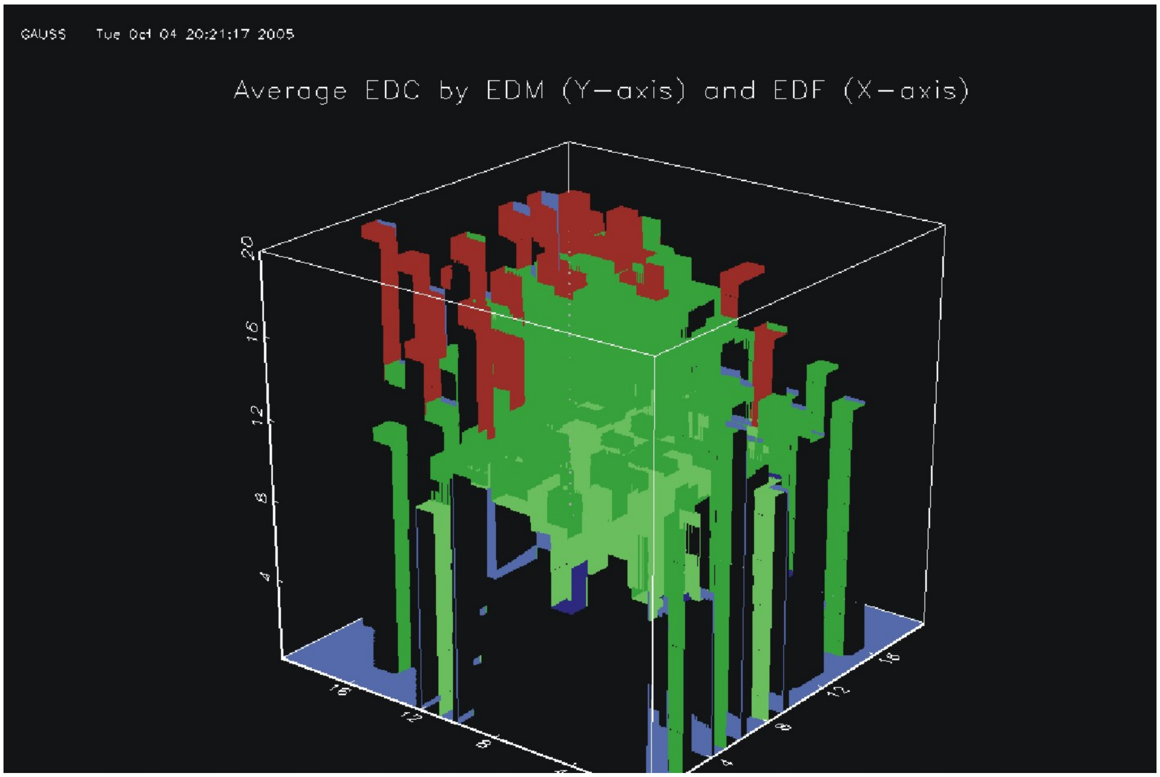
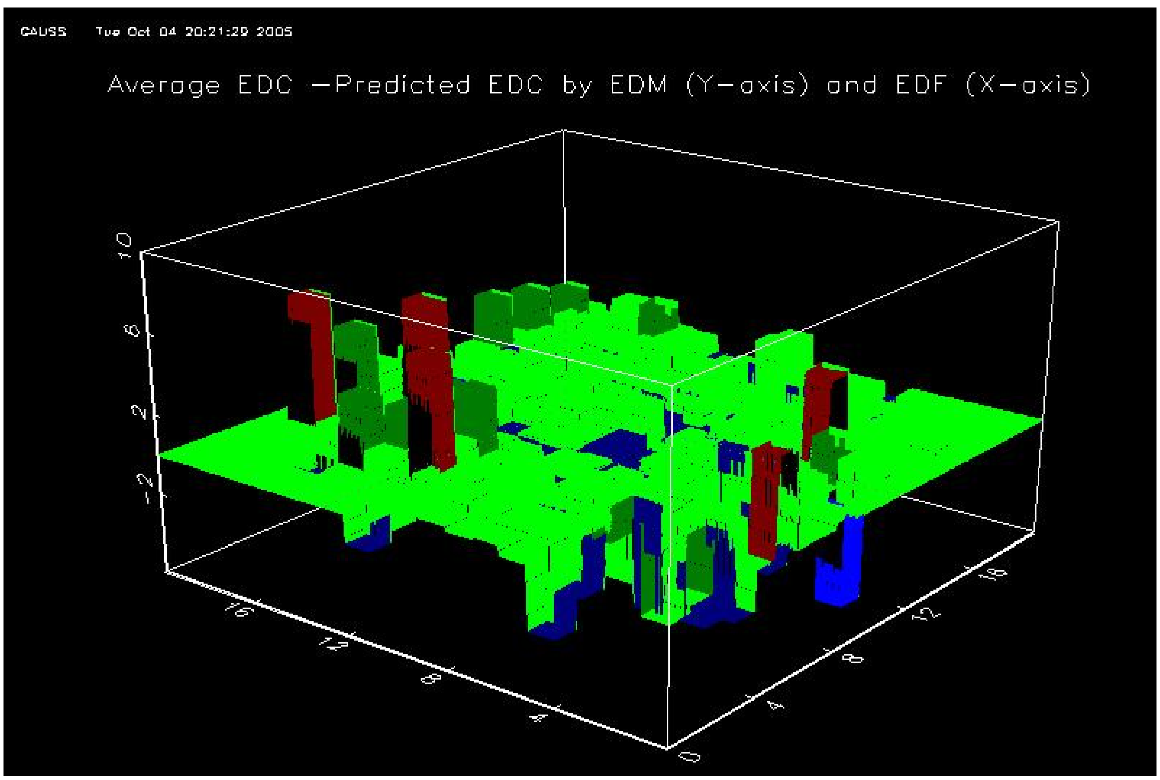


Figure 2: Average years of education child minus predicted by education father and mother



- Correlated sorting effect

Table 3: Average education given correlated (ρ) sorting

ρ	$\hat{\beta}_{cs}$	Std($\hat{\beta}_{cs}$)
-.99	11.5	.069
-.8	11.7	.048
-.6	11.9	.040
-.4	12.1	.037
-.2	12.4	.034
0.	12.6	.033
.2	12.8	.031
.4	12.9	.030
.6	13.0	.029
.8	13.0	.029
.99	13.1	.039

Figure 3: Average years of education child given correlated sorting; 95% error bands