Generalized Roy Model and the Cost-Benefit Analysis of Social Programs

Philipp Eisenhauer, University of Mannheim,
James Heckman, University of Chicago,
and
Edward Vytlacil, Yale University

July, 2010

Treatment Effect Literature

Treatment effect literature:

- Seeks to identify effect of treatment on outcome, i.e., to identify the gross benefit of treatment
- Does not examine cost of treatment

Policy analysis requires knowledge of net benefit, and thus of cost of treatment, but we typically do not have direct measurements of costs.

Question of Interest

Imposing that agents select into treatment if the expected benefit exceeds the expected cost, and imposing that we observe outcomes but do not observe any direct information on costs, what can we learn about expected cost and expected net-benefit as perceived by the agents?

Summary of Analysis

- We impose a Generalized Roy Model
- We impose two exclusion restrictions:
  1. A variable that affects costs but not outcomes
  2. A variable that affects outcomes but not costs
- We identify MTE-like cost and net-benefit parameters, with the range of identified parameters depending on the strength of the exclusion restrictions.

Outline

1. Notation, Model: Generalized Roy Model
2. Parameters:
    1. Marginal benefit, cost, net-surplus
    2. Relationship between marginal parameters
3. Identification Analysis
    1. Local IV $\Rightarrow$ identify marginal benefit
    2. Marginal benefit + decision rule
       $\Rightarrow$ identify marginal cost
       $\Rightarrow$ identify marginal net surplus
4. Testable restrictions
5. Extension to imperfect foresight

Notation, Model

- Treatment Variables: $D = 1$ if treated, $= 0$ otherwise.
- Potential Outcomes: Y_0, Y_1, given by

$$Y_j = \mu_j(X) + U_j$$

- Observed Outcome: Y

$$\Rightarrow Y = DY_1 + (1 - D)\, Y_0.$$

- Cost of treatment: C, given by

$$C = \mu_C(Z) + U_C$$

Notation, Model (cont'd)

Surplus (net-benefit) of Treatment, S,

$$\begin{aligned} S &= (Y_1 - Y_0) - C \\ &= \{[\mu_1(X) - \mu_0(X)] - \mu_C(Z)\} - [U_C - (U_1 - U_0)] \\ &= \mu_S(X, Z) - V \end{aligned}$$

with

$$\mu_S(X, Z) = [\mu_1(X) - \mu_0(X)] - \mu_C(Z),$$

$$V = U_C - (U_1 - U_0).$$

Notation, Model (cont'd)

For ease of exposition, will assume Generalized Roy Model with perfect certainty:

$$D = 1 \quad \text{if} \quad S \geq 0 \,; \qquad D = 0 \quad \text{otherwise,} \tag{1}$$

Our analysis is more general, allowing for agents to enter treatment if expected benefit exceeds expected cost without requiring perfect certainty.

Assumptions

Assumption (Independence)

(U_0, U_1, U_C) *are independent of* (X, Z);

Assumption (Rank Condition)

The distribution of $\mu_C(Z)$ *conditional on* X *is absolutely continuous with respect to Lebesgue measure.*

Plus additional regularity conditions
(existence of moments, etc)

Normalization

Define:

$$P(X,Z) \equiv \Pr(D=1 \mid X,Z) = F_V(\mu_S(X,Z))$$

$$U_D = F_V(V)$$

Thus, we can rewrite the decision rule as:

$$D = \mathbf{1}[P(X,Z) \geq U_D]$$

with

$$U_D \sim \text{Unif}[0,1]$$

Parameters

In previous work, Heckman-Vytlacil focus on MTE parameter

$$\text{B-MTE}(X, U_D) = E(Y_1 - Y_0 | X, U_D)$$

B-MTE(X, U_D) is the expected benefit of treatment given covariates X and quantile of unobserved (non-)desire to select into treatment U_D.

Many treatment parameters of interest can be represented as a weighted average of B-MTE(X, U_D) .

Parameters (cont'd)

Likewise, can define cost or net surplus versions of MTE:

$$\begin{aligned}\text{C-MTE}(Z, U_D) &= E(C|Z, U_D) \\ \text{S-MTE}(X, Z, U_D) &= E(S|X, Z, U_D) \\ &= \text{B-MTE}(X, U_D) - \text{C-MTE}(Z, U_D)\end{aligned}$$

Many cost parameters and net surplus parameters of interest can be represented as as weighted averages of C-MTE(Z, U_D) and S-MTE(X, Z, U_D).

Relationship Between Marginal Parameters

$$\begin{aligned} \text{S-MTE}(X, Z, U_D) &= E(\mu_S(X, Z) - V | X, Z, U_D) \\ &= \mu_S(X, Z) - E(V | U_D) \\ &= \mu_S(X, Z) - F_V^{-1}(U_D), \end{aligned}$$

using that $V = F_V^{-1}(U_D)$.

Relationship Between Marginal Parameters (cont'd)

Recall:

- S-MTE$(x, z, u) = \mu_S(x, z) - F_V^{-1}(u)$,
- $P(x, z) = F_V(\mu_S(x, z))$.

Thus

$$\begin{aligned} \text{S-MTE}(x, z, P(x, z)) &= \mu_S(x, z) - F_V^{-1}\left(P(x, z)\right) \\ &= \mu_S(x, z) - F_V^{-1}\left(F_V(\mu_S(x, z))\right) \\ &= 0. \end{aligned}$$

Relationship Between Marginal Parameters (cont'd)

$\text{S-MTE}(x, z, u) = \text{B-MTE}(x, u) - \text{C-MTE}(z, u)$, and thus

$$\begin{aligned} \text{B-MTE}(x,u) &> \text{C-MTE}(z,u) \quad \forall\ u < P(x,z) \\ \text{B-MTE}(x,u) &= \text{C-MTE}(z,u) \quad \quad u = P(x,z) \\ \text{B-MTE}(x,u) &< \text{C-MTE}(z,u) \quad \forall\ u > P(x,z) \end{aligned}$$

$\text{B-MTE}(x, P(x, z)) = \text{C-MTE}(z, P(x, z))$ will be key for how we identify the cost and net surplus parameters.

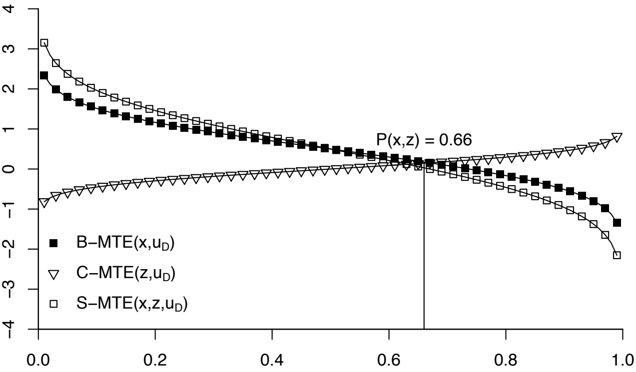

Figure: Relationship between MTE Parameters.

Identification Analysis: B-MTE

Following Heckman and Vytlacil (1999), we can use local IV (LIV) to identify MTE:

$$\frac{\partial}{\partial p} E(Y|X = x, P = p) = \text{B-MTE}(x, p).$$

For a fixed x, we identify B-MTE(x, u) for $u \in$ Supp$(P|X = x)$. The more variation in cost shifters, the greater the variation in propensity scores conditional on $X = x$, and the larger the set of evaluation points for which we identify B-MTE(x, u).

Identification Analysis: C-MTE

B-MTE$(x, P(x, z))$ = C-MTE$(z, P(x, z))$ implies

$$\frac{\partial}{\partial p} E(Y|X = x, P = p)\Big|_{p=P(x,z)} = \text{C-MTE}(z, P(x, z)).$$

For a fixed z, we identify C-MTE(z, u) for $u \in$ Supp$(P|Z = z)$. The more variation in regressors that shift outcomes, the more variation in propensity scores conditional on $Z = z$, and the larger the set of evaluation points for which we identify C-MTE(z, u).

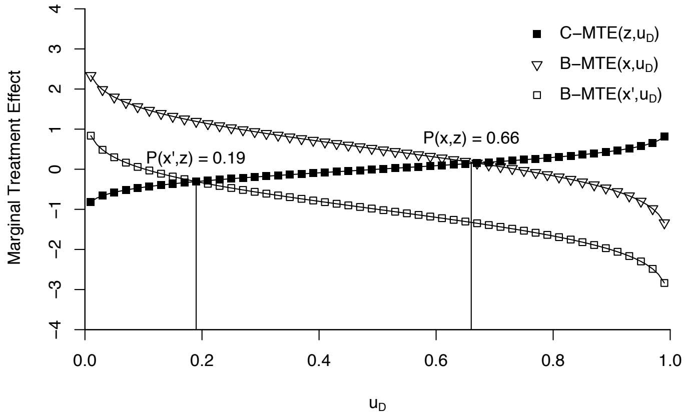

Figure: Identification of C-MTE

Identification Analysis: S-MTE

- We identify B-MTE(x, u) for $(x, u) \in \text{Supp}(X, P)$.
- We identify C-MTE(z, u) for $(z, u) \in \text{Supp}(Z, P)$.
- S-MTE$(x, z, u) =$ B-MTE$(x, u) -$ C-MTE(z, u),

  $\Rightarrow$ identify S-MTE(x, z, u) for (x, z, u) s.t. $(x, u) \in \text{Supp}(X, P)$ and $(z, u) \in \text{Supp}(Z, P)$
- Following Heckman-Vytlacil,we can integrate these marginal parameters to obtain other average cost and net-surplus parameters.

Testable Restrictions

1. Suppose that $U_1 - U_0$ is degenerate. Then $E(Y|X = x, P = p)$ is linear in p.
2. Suppose $U_1 - U_0 \perp\!\!\!\perp U_C$. For a fixed x, consider a line $a + bp$, where $a = E(Y|X = x, P(X, Z) = 0)$ and $b = E(Y|X = x, P(X, Z) = 1) - E(Y|X = x, P(X, Z) = 0)$. Then $E(Y|X = x, P(X, Z) = p) \geq a + bp$ for all $p \in \text{Supp}(P|X = x)$.
3. Suppose $U_1 - U_0 \perp\!\!\!\perp U_C$, and suppose $U_1 - U_0$ and U_C have log concave densities. Then $E(Y|X = x, P(X, Z) = p)$ is a concave function of p.

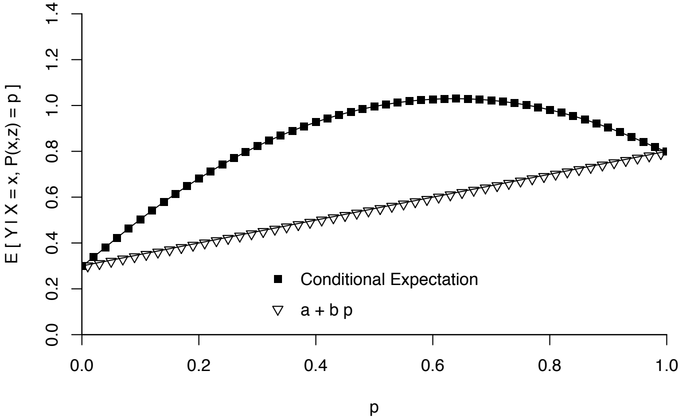

Figure: Testable Restriction

Extension to Imperfect Foresight

Have thus far assumed agents know (Y_1, Y_0, C), select into treatment if benefit exceeds the cost,

$$D = \mathbf{1}[Y_1 - Y_0 \geq C].$$

Same results extend to more general model with imperfect foresight.

Extension to Imperfect Foresight (cont'd)

Recall: $Y_j = \mu_j(X) + U_j,\ C = \mu_C(Z) + U_C$

- Suppose agents know (X_I, Z_I, U_I), where X_I is a subvector of X, Z_I of Z.
- Suppose $(X, Z) \perp\!\!\!\perp (U_1, U_0, U_C, U_I)$
- Suppose agents select into treatment if expected benefit exceeds expected cost,

$$D = \mathbf{1}[E(Y_1 - Y_0|X_I, U_I) \geq E(C|Z_I, U_I)]$$

- All results extend, though now need to project B-MTE as identified by LIV on information set of agent to obtain expected benefit as perceived by agent, which in turn allows identification of expected cost.

Application

- We perform cost-benefit analysis based on a sample of white males from the NLSY79.
- **Treatment:**
  D = 0 (high school dropouts and high school graduates)
  D = 1 (individuals with some college or more).
  Schooling is measured in 1991
  (individuals are between 28 and 34 years of age in 1991).
- **Outcome Variable.** Average of deflated (to 1983) non-missing hourly wages reported in 1989-1993.

Application (cont'd)

- **Parametric/Semiparametric Structure**

$$\begin{aligned}\mu_1(X) &= X\beta_1 \\ \mu_0(X) &= X\beta_0, \\ \mu_C(Z) &= Z\gamma\end{aligned}$$

$$V \sim N(0, \sigma^2)$$

Given this structure, we

1. Estimate $P(X, Z)$ by a probit
2. Estimate MTE based on a semi-parametric partially linear model, following Carneiro et al (2009).
3. Use estimated probit coefficients and MTE estimates to back out cost-MTE

Application (cont'd)

- **Variables Common to Choice, Cost and Outcome Equations:** linear and quadratic terms in mother's education, number of siblings as well as dummy variables indicating urban residence at age 14, and cohort dummies.
- **Variables in Cost/Choice Equations, Excluded form the Outcome Equation:** (a) the presence of a four year college in the county of residence at age 14 , (b) local wage in the county of residence at age 17, (c) local unemployment in the state of residence at age 17, and (d) average tuition in public 4 year colleges in the county of residence at age 17.

Application (cont'd)

- **Variables in Outcome Equation, Excluded form the Cost Equation.** Measures of permanent local labor market conditions: average earnings and unemployment between 1973 and 2000 for each location of residence at 17. Thus, short-run fluctuations in labor market conditions only affect the cost of treatment without effects on benefits, while long-run labor market environment affect the choice of treatment only as determinants of the benefits from participation.

Application (cont'd)

- **Variables in Outcome Equation, Unknown to Agent at Time of Selection:** We include the average log earnings in the county of residence in 1991, and the average unemployment rate in the state of residence in 1991 in the outcome equation but not in the choice equation.

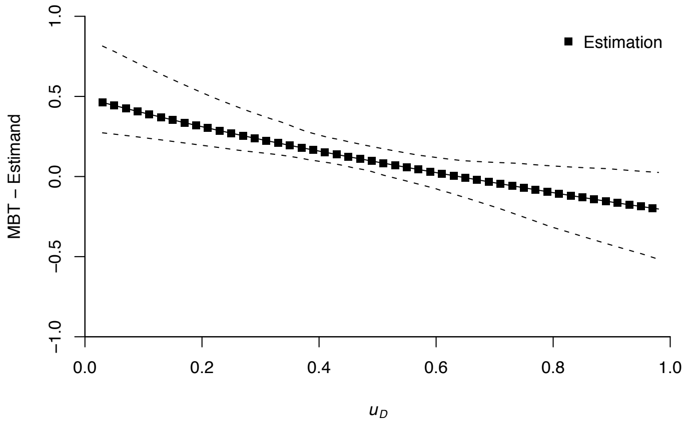

Figure: Estimated B-MTE.

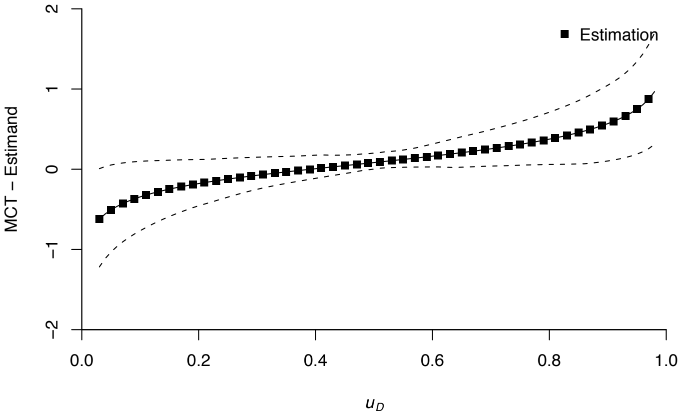

Figure: Estimated C-MTE.

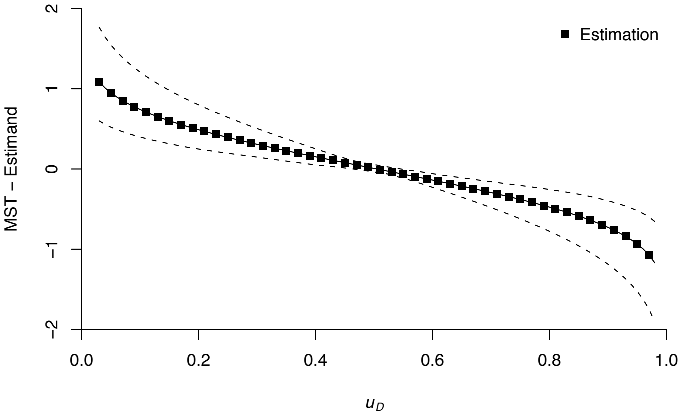

Figure: Estimated S-MTE.