LATE and the Generalized Roy Model: Some Relationships

James J. Heckman
University of Chicago

Extract from: Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy
James J. Heckman (JEL 2010)

Econ 312, Spring 2019

Defining LATE

- **Question:*** Derive the MTE from the sample selection model. What parameters are identified by the selection model that are not identified by MTE? Explain the advantages and disadvantages of each approach.

*Answer after reading these slides

LATE

- LATE is defined by the variation of an instrument.
- The instrument in LATE plays the role of a randomized assignment.
- Randomized assignment is an instrument.
- Y_0 and Y_1 are potential *ex-post* outcomes.
- Instrument Z assumes values in $\mathcal{Z}$, $z \in \mathcal{Z}$.

- $D(z)$: indicator of hypothetical choice representing what choice the individual would have made had the individual's Z been exogenously set to z.
- $D(z) = 1$ if the person chooses (is assigned to) 1.
- $D(z) = 0$, otherwise.
- One can think of the values of z as fixed by an experiment or by some other mechanism independent of (Y_0, Y_1).
- All policies are assumed to operate through their effects on Z.
- It is assumed that Z can be varied conditional on X.

- Three assumptions define LATE.

Assumption 1

$(Y_0, Y_1, \{D(z)\}_{z \in \mathcal{Z}}) \perp\!\!\!\perp Z \mid X$

Assumption 2

$\Pr(D = 1 \mid Z = z)$ *is a nontrivial function of* z *conditional on* X.

Assumption 3

For any two values of Z, say $Z = z^1$ and $Z = z^2$, either $D(z^1) \geq D(z^2)$ for all persons, or $D(z^1) \leq D(z^2)$ for all persons.

- This condition is a statement *across* people.
- This condition does not require that for any other two values of Z, say z^3 and z^4, the direction of the inequalities on $D(z^3)$ and $D(z^4)$ have to be ordered in the same direction as they are for $D(z^1)$ and $D(z^2)$.
- It only requires that the direction of the inequalities are the *same across people*.
- Thus for any person, $D(z)$ need not be monotonic in z.

- Under LATE conditions, for two distinct values of Z, z^1 and z^2, IV applied to

$$\text{LATE}(z^2, z^1) = E(Y_1 - Y_0 \mid D(z^2) = 1, D(z^1) = 0),$$

  if the change from z^1 to z^2 induces people into the program $(D(z^2) \geq D(z^1))$.
- This is the mean return to participation in the program for people induced to switch treatment status by the change from z^1 to z^2.

- LATE does not identify which people are induced to change their treatment status by the change in the instrument.
- It leaves unanswered many policy questions.
- For example, if a proposed program changes the same components of vector Z as used to identify LATE but at different values of Z (say z^4, z^3), LATE(z^2,z^1) does not identify LATE(z^4, z^3).

- If the policy operates on different components of Z than are used to identify LATE, one cannot safely use LATE to identify marginal returns to the policy.
- It does not, in general, identify treatment on the treated, ATE or a variety of criteria.
- But using the implicit economics of the problem one can do better as I show below.

Identifying Policy Parameters

$$Y_1 = \mu_1(X) + U_1, \qquad Y_0 = \mu_0(X) + U_0, \qquad C = \mu_C(Z) + U_C, \quad (1)$$

- (X, Z) are observed by the analyst.
- U_0, U_1, U_C are unobserved.

- Define Z to include all of X.
- Variables in Z not in X are instruments.
- $I_D = E(Y_1 - Y_0 - C \mid \mathcal{I}) = \mu_D(Z) - V$
  $\mu_D(Z) = E(\mu_1(X) - \mu_0(X) - \mu_C(Z) \mid \mathcal{I})$
  $V = -E(U_1 - U_0 - U_C \mid \mathcal{I})$.
- Choice equation:

$$D = 1(\mu_D(Z) > V). \tag{2}$$

- Recall from Vytlacil's Theorem (2002) that (2) = Assumption 1–Assumption 3 **monotonicity**.
- In the early literature that implemented this approach $\mu_0(X)$, $\mu_1(X)$, and $\mu_C(Z)$ were assumed to be linear in the parameters, and the unobservables were assumed to be normal and distributed independently of X and Z.

- The essential aspect of the structural *approach* is joint modeling of outcome and choice equations.
- Structural econometricians have developed nonparametric identification analyses for the Roy and generalized Roy models.
- Central to the whole LATE enterprise is centrality of $Pr(D = 1|X, Z) = P$ (we keep X implicit).
- Remember $D = 1[F_V(M_D(Z)) \geq F_V(V)]$.

To Recapitulate

A useful fact: Assume $Z \perp\!\!\!\perp V$ (implied by Assumption 1)

$$\text{Then Choice Probability}: \quad \begin{aligned} P(z) &= \Pr(D = 1 \mid Z = z) \\ &= \Pr(\mu_D(z) \geq V) \\ &= \Pr\left(\frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V}\right) \end{aligned}$$

$$\begin{aligned} P(z) &= F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{\mu_D(z)}{\sigma_V}\right) \\ U_D &= F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{V}{\sigma_V}\right); \quad \text{Uniform}(0,1) \end{aligned}$$

$$
\begin{aligned}
P(z) &= \Pr\left(F_{\frac{V}{\sigma_V}}\left(\frac{\mu_D(z)}{\sigma_V}\right) \geq F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{V}{\sigma_V}\right)\right) \\
&= \Pr\left(P(z) \geq U_D\right)
\end{aligned}
$$

$P(z)$ is the $p(z)^{\text{th}}$ quantile of U_D.

Recall

$$
\begin{aligned}
Y &= DY_1 + (1-D)Y_0 \\
&= Y_0 + D(Y_1 - Y_0)
\end{aligned}
$$

Keep X implicit (condition on $X = x$)

$$
\begin{aligned}
E(Y \mid Z = z) &= E(Y_0) + \underbrace{E(Y_1 - Y_0 \mid D = 1, Z = z)P(z)}_{\text{from law of iterated expectations}} \\
&= E(Y_0) + E(Y_1 - Y_0 \mid P(z) \geq U_D)P(z)
\end{aligned}
$$

$\therefore$ It depends on Z only through $P(Z)$.

$$
E(Y \mid Z = z') = E(Y_0) + E(Y_1 - Y_0 \mid P(z') \geq U_D)P(z')
$$

- What is $E(Y_1 - Y_0 \mid P(z) \geq U_D)$? (Treatment on the treated)
- Assume (Y_1, Y_0, U_D) (absolutely) continuous.
- The joint density of $(Y_1 - Y_0, U_D)$: $f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D)$.
- Does not depend on Z.
- It may, in general, depend on X.
-

$$E(Y_1 - Y_0 \mid P(z) \geq U_D) = \frac{\int\limits_{-\infty}^{\infty} \int\limits_{0}^{P(z)} (y_1 - y_0) f_{y_1 - y_0, u_D}(y_1 - y_0, u_D)\, du_D d(y_1 - y_0)}{\Pr(P(z) \geq U_D)}$$

- Recall that

$$U_D = F_{\left(\frac{V}{\sigma_V}\right)}\left(\frac{V}{\sigma_V}\right).$$

- U_D is a quantile of the V/σ_V distribution.

- By construction, U_D is $\text{Uniform}(0,1)$ (this is the definition of a quantile).
- $\therefore f_{U_D}(u_D) = 1$.
- Also, $\Pr(P(z) \geq U_D) = P(z)$.
- By law of conditional probability,

$$f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D) = f_{Y_1 - Y_0, U_D}(y_1 - y_0 \mid U_D = u_D) \underbrace{f_{U_D}(u_D)}_{=1}.$$

$$E(Y_1 - Y_0 \mid P(z) \geq U_D) = \frac{\int_0^{P(z)} \int_{-\infty}^{\infty} (y_1 - y_0) f_{Y_1 - Y_0, U_D}(y_1 - y_0, u_D) \, d(y_1 - y_0) \, du_D}{P(z)}$$

$$E(Y_1 - Y_0 \mid P(z) \geq U_D) = \frac{\int_0^{P(z)} \int_{-\infty}^{\infty} (y_1 - y_0) f_{Y_1 - Y_0, U_D}(y_1 - y_0 \mid U_D = u_D) \, d(y_1 - y_0) \, du_D}{P(z)}$$

$$= \frac{\int_0^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) \, du_D}{P(z)}$$

$$\therefore E(Y \mid Z = z) = E(Y_0) + \int_0^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

$$\frac{\partial E(Y \mid Z = z)}{\partial P(z)} = \underbrace{E(Y_1 - Y_0 \mid U_D = P(z))}_{\substack{\text{marginal gains for} \\ \text{people with } U_D = P(z)}} = \text{MTE}(U_D) \text{ for } U_D = P(Z)$$

$$E(Y \mid Z = z') = E(Y_0) + \int_0^{P(z')} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

THE UNIVERSITY OF CHICAGO

- Suppose $P(z) > P(z')$

$$\therefore E(Y \mid Z = z) - E(Y \mid Z = z') =$$

$$= \int_{P(z')}^{P(z)} E(Y_1 - Y_0 \mid U_D = u_D) du_D$$

$$= E(Y_1 - Y_0 \mid P(z) \geq U_D \geq P(z')) \Pr(P(z) \geq U_D \geq P(z'))$$

Notice

$$\Pr(P(z) \geq U_D \geq P(z')) = \int_{P(z')}^{P(z)} du_D$$

$$= P(z) - P(z')$$

$$E(Y \mid Z = z) - E(Y \mid Z = z')$$

$$= \underbrace{E(Y_1 - Y_0 \mid P(z) \geq U_D \geq P(z'))}_{\text{LATE}}(P(z) - P(z'))$$

$$\frac{E(Y \mid Z = z) - E(Y \mid Z = z')}{P(z) - P(z')} = \mathrm{LATE}(z, z')$$

$$= \frac{\int_{P(z')}^{P(z)} \mathrm{MTE}(u_D) du_D}{P(z) - P(z')}$$

- **Question:** In what sense is $E(Y_1 - Y_0 \mid P(z) \geq U_D)$ a measure of surplus of agents for whom $P(z) \geq U_D$?

Appendix: The Generalized Roy Model for the Normal Case

$$
\begin{aligned}
Y_1 &= \mu_1(X) + U_1 \\
Y_0 &= \mu_0(X) + U_0 \\
C &= \mu_C(Z) + U_C \\
\text{Net Benefit: } I &= Y_1 - Y_0 - C \\
I &= \underbrace{\mu_1(X) - \mu_0(X) - \mu_C(Z)}_{\mu_D(Z)} + \underbrace{U_1 - U_0 - U_C}_{-V}
\end{aligned}
$$

$$
\begin{aligned}
(U_0, U_1, U_C) &\perp\!\!\!\perp (X, Z) \\
E(U_0, U_1, U_C) &= (0, 0, 0) \\
V &\perp\!\!\!\perp (X, Z)
\end{aligned}
$$

- Assume normally distributed errors.
- Assume Z contains X but may contain other variables (exclusions)

$$Y = DY_1 + (1 - D)Y_0 \qquad \text{observed } Y$$
$$D = 1(I \geq 0) = 1(\mu_D(Z) \geq V)$$

- Assume $V \sim N(0, \sigma_V^2)$

THE UNIVERSITY OF CHICAGO

- Propensity Score:

$$\Pr(D = 1 \mid Z = z) = \Phi\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$E(Y \mid D = 1, X = x, Z = z) = \mu_1(X) + \underbrace{E(U_1 \mid \mu_D(z) \geq V)}_{K_1(P(z))}$$

because $(X, Z) \perp\!\!\!\perp (U_1, V)$.

- Under normality we obtain

$$E\left(U_1 \middle| \frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V}\right) = \frac{\mathrm{Cov}(U_1, \frac{V}{\sigma_V})}{\mathrm{Var}(\frac{V}{\sigma_V})} \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

- Why?

$$U_1 = \text{Cov}\left(U_1, \frac{V}{\sigma_V}\right)\frac{V}{\sigma_V} + \varepsilon_1$$

$$\varepsilon_1 \perp\!\!\!\perp V$$

$$E\left(\frac{V}{\sigma_V} \mid \frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V}\right) = \frac{\int_{-\infty}^{\frac{\mu_D(z)}{\sigma_V}} t \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt}{\int_{-\infty}^{\frac{\mu_D(z)}{\sigma_V}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt} = \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$= \frac{\frac{-1}{\sqrt{2}\pi} e^{\left(-\frac{1}{2}\right)\left(\frac{\mu_D(z)}{\sigma_V}\right)^2}}{\Phi\left(\frac{\mu_D(z)}{\sigma_V}\right)} = \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right) = \frac{-\phi\left(\frac{\mu_D(z)}{\sigma_V}\right)}{\Phi\left(\frac{\mu_D(z)}{\sigma_V}\right)}$$

- Notice

$$\lim_{\mu_D(z)\to\infty} \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right) = 0$$

$$\lim_{\mu_D(z)\to-\infty} \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right) = -\infty$$

- Propensity score:

$$P(z) = \Pr(D = 1 \mid Z = z) = \Phi\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$\therefore \left(\frac{\mu_D(z)}{\sigma_V}\right) = \Phi^{-1}\left(\Pr(D = 1 \mid Z = z)\right)$$

- Thus we can replace $\frac{\mu_D(z)}{\sigma_V}$ with a known function of $P(z)$

- Notice that because $(X, Z) \perp\!\!\!\perp (U, V)$, Z enters the model (conditional on X) only through $P(Z)$.
- This is called index sufficiency.

- Put all of these results together to obtain

$$E\left(Y \mid D=1, X=x, Z=z\right) = \mu_1(x) + \left(\frac{\mathrm{Cov}(U_1, \frac{V}{\sigma_V})}{\mathrm{Var}(\frac{V}{\sigma_V})}\right)\tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$= E\left(Y_1 \mid D=1, X=x, Z=z\right) = \mu_1(x) + \left(\frac{\mathrm{Cov}(U_1, \frac{V}{\sigma_V})}{\mathrm{Var}(\frac{V}{\sigma_V})}\right)\tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$\tilde{\lambda}(z) = E\left(\frac{V}{\sigma_V} \mid \frac{V}{\sigma_V} < \frac{\mu_D(z)}{\sigma_V}\right) < 0$$

$$\lambda(z) = E\left(\frac{V}{\sigma_V} \mid \frac{V}{\sigma_V} \geq \frac{\mu_D(z)}{\sigma_V}\right) > 0$$

$$E\left(Y \mid D=0, X=x, Z=z\right) = \mu_0(x) + \left(\frac{\mathrm{Cov}(U_0, \frac{V}{\sigma_V})}{\mathrm{Var}(\frac{V}{\sigma_V})}\right)\lambda\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

$$\mathrm{Var}\left(\frac{V}{\sigma_V}\right) = 1$$

$$\frac{V}{\sigma_V} = -\frac{(U_1 - U_0 - U_C)}{\sigma_V}$$

$$\text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) = -\text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) + \text{Cov}\left(U_0, \frac{V}{\sigma_V}\right) + \text{Cov}\left(U_C, \frac{V}{\sigma_V}\right)$$

In Roy model case ($U_C = 0$),

$$\begin{aligned} \text{Cov}\left(U_1, \frac{V}{\sigma_V}\right) &= -\text{Cov}\left(U_1, \frac{U_1 - U_0}{\sigma_V}\right) \\ &= -\frac{\text{Cov}\left(U_1 - U_0, U_1\right)}{\sqrt{\text{Var}(U_1 - U_0)}} \end{aligned}$$

- We can identify $\mu_1(x), \mu_0(x)$
- From Discrete Choice model we can identify

$$\frac{\mu_D(z)}{\sigma_V} = \frac{\mu_1(x) - \mu_0(x) - \mu_C(z)}{\sigma_V}$$

- If we have a regressor in X that does not affect $\mu_C(z)$ (say regressor x_j, so $\frac{\partial \mu_C(z)}{\partial x_j} = 0$), we can identify σ_V and $\mu_C(z)$.
- $\therefore$ We can identify the net benefit function and the cost function up to scale.
- $\therefore$ We can compute *ex-ante* subjective net gains.

- Method generalizes: Don't need normality
- Need "Large Support" assumption to identify ATE and TT
- $$E\left(Y \mid D=1, X=x, Z=z\right)=\mu_1(x)+\overbrace{K_1(P(z))}^{\text{control function}}$$
$$E\left(Y \mid D=0, X=x, Z=z\right)=\mu_0(x)+\underbrace{K_0(P(z))}_{\text{control function}}$$
$$\lim_{P(z)\to 1} E\left(Y \mid D=1, X=x, Z=z\right)=\mu_1(x)$$
$$\lim_{P(z)\to 0} E\left(Y \mid D=0, X=x, Z=z\right)=\mu_0(x)$$

- If we have this condition satisfied, we can identify ATE

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- ATE is defined in a limit set. This is true for any model with selection on unobservables (IV; selection models)

- What about treatment on the treated?

$$E(Y_1 - Y_0 \mid D = 1, X = x, Z = z)$$

a. From the data, we observe

$$E(Y_1 \mid D = 1, X = x, Z = z)$$

b. Can also create it from the model

c. $E(Y_0 \mid D = 1, X = x, Z = z)$ is a counterfactual

We know

$$E(Y_0 \mid D = 0, X = x, Z = z) = \mu_0(x) + \mathrm{Cov}\left(U_0, \frac{V}{\sigma_V}\right) \lambda\left(\frac{\mu_D(Z)}{\sigma_V}\right)$$

(this is data)

d. We seek

$$E(Y_0 \mid D = 1, X = x, Z = z) = \mu_0(x) + \mathrm{Cov}\left(U_0, \frac{V}{\sigma_V}\right) \tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$$

- But under normality, we know $\mathrm{Cov}\left(U_0, \frac{V}{\sigma_V}\right)$
- We know $\frac{\mu_D(Z)}{\sigma_V}$
- $\tilde{\lambda}(\cdot)$ is a known function.
- Can form $\tilde{\lambda}\left(\frac{\mu_D(z)}{\sigma_V}\right)$ and can construct counterfactual.

- More generally, without normality but with $(X, Z) \perp\!\!\!\perp (U, V)$,

$$E(Y_1 \mid D=1, X, Z) = E(Y \mid D=1, X=x, Z=z) = \mu_1(x) + K_1(P(z))$$

$$E(Y_0 \mid D=0, X, Z) = E(Y \mid D=0, X=x, Z=z) = \mu_0(x) + \tilde{K}_0(P(z))$$

where $K_1(P(z)) = E(U_1 \mid D=1, X=x, Z=z) = E\left(U_1 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$

$$\tilde{K}_1(P(z)) = E\left(U_1 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$$

$$\tilde{K}_0(P(z)) = E\left(U_0 \mid \frac{\mu_D(z)}{\sigma_V} > \frac{V}{\sigma_V}\right)$$

- Use the transformation

$$\frac{F_V}{\sigma_V}\left(\frac{\mu_D(z)}{\sigma_V}\right) = P(z)$$

$$\frac{F_V}{\sigma_V}\left(\frac{V}{\sigma_V}\right) = U_D \qquad \text{(a uniform random variable)}$$

$$D = 1\left(\frac{\mu_D(z)}{\sigma_V} \geq \frac{V}{\sigma_V}\right) = 1\left(P(z) \geq U_D\right)$$

$$K_1(P(z)) = E(U_1 \mid P(z) > U_D)$$

$$K_1(P(z))P(z) + \tilde{K}_1(P(z))(1 - P(z)) = 0$$

$$\therefore \text{ we can construct } \tilde{K}_1(P(z))$$

- Symmetrically

$$\tilde{K}_0(P(z)) = E(U_0 \mid P(z) \leq U_D)$$
$$K_0(P(z)) = E(U_0 \mid P(z) > U_D)$$
$$(1 - P(z))\tilde{K}_0(P(z)) + P(z)K_0(P(z)) = 0$$

- $\therefore$ If we have "identification at infinity," we can construct

$$E(Y_1 - Y_0 \mid X = x) = \mu_1(x) - \mu_0(x)$$

- We can construct TT

$$\begin{aligned} E(Y_1 - Y_0 \mid D = 1, X = x, Z = z) &= \\ &= \underbrace{[\mu_1(x) + K_1(P(z))]}_{\text{factual}} - \underbrace{[\mu_0(x) + K_0(P(z))]}_{\text{counterfactual}} \end{aligned}$$

- But we can form $\mu_1(x) + K_1(P(z))$ from data
- We get $\mu_0(x)$ from limit set $P(z) \to 0$ identifies $\mu_0(x)$
- We can form $K_0(P(z)) = -\tilde{K}_0(P(z))\frac{P(z)}{1-P(z)}$
- $\therefore$ Can construct the desired counterfactual mean.

- Notice how we can get Effect of Treatment for People at the Margin:

$$E(Y_1 - Y_0 \mid I = 0, X = x, Z = z)$$

- Under normality we have (as a result of independence and normality)

$$\begin{aligned}
&E(Y_1 - Y_0 \mid I = 0, X = x, Z = z) \\
&\quad = \mu_1(x) - \mu_0(x) + E\left(U_1 - U_0 \mid \frac{\mu_D(z)}{\sigma_V} = \frac{V}{\sigma_V}, X = x, Z = z\right) \\
&\quad = \mu_1(x) - \mu_0(x) + \mathrm{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right)\frac{\mu_D(z)}{\sigma_V}
\end{aligned}$$

In the Roy model case where $U_C = 0$ but $\mu_C(z) \neq 0$

$$\begin{aligned}
&= \mu_1(x) - \mu_0(x) - \sigma_V\left(\frac{\mu_D(z)}{\sigma_V}\right) \\
&= \mu_1(x) - \mu_0(x) - \mu_D(z) \\
&= \mu_C(z)
\end{aligned}$$

(marginal gain $=$ marginal cost)

- MTE is

$$E(Y_1 - Y_0 \mid V = v, X = x, Z = z) =$$
$$= \mu_1(x) - \mu_0(x) + \operatorname{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right) v$$

- Effect of Treatment for People at the Margin picks $v = \frac{\mu_D(z)}{\sigma_V}$
- Notice we can use the result that

$$\frac{\mu_D(z)}{\sigma_V} = F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(P(z))$$
$$V = F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(U_D)$$

- Effect of Treatment for People at Margin of Indifference Between Taking Treatment and Not:

$$E(Y_1 - Y_0 \mid I = 0, X = x, Z = z) =$$

$$= \mu_1(x) - \mu_0(x) + \text{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right) F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(P(z))$$

- MTE:

$$E(Y_1 - Y_0 \mid U_D = u_D, X = x, Z = z) =$$

$$= \mu_1(x) - \mu_0(x) + \text{Cov}\left(U_1 - U_0, \frac{V}{\sigma_V}\right) F^{-1}_{\left(\frac{V}{\sigma_V}\right)}(U_D)$$

THE UNIVERSITY OF CHICAGO

- Recent (1987 and Later!) Advances in Econometrics:
  1. Relax normality
  2. Do not assume linearity of $\mu_1(X)$ and $\mu_0(X)$ in terms of X
  3. Do not require identification at infinity but only because they abandon pursuit of ATE, TT, TUT or else assume that $(Y_1, Y_0) \perp\!\!\!\perp D \mid X$ (matching assumption)
  4. Identification at infinity in some version or the other is required for ATE, TT, TUT as long as there is selection on unobservables (i.e., $(Y_1, Y_0) \not\!\perp\!\!\!\perp D \mid X$)

THE UNIVERSITY OF CHICAGO

End of Example of Normal Model

Appendix: Nonparametric Identification of the Roy Model

- (Y_0, Y_1) potential outcomes
- $I^* = Y_1 - Y_0$ choice index
- Observe Y_1 if $Y_1 \geq Y_0$.
- Observe Y_0 if $Y_1 < Y_0$.
- Cannot simultaneously observe Y_0 and Y_1.
- Generalized Roy model: $I = Y_1 - Y_0 - C$.
- C depends on Z.

- Heuristically, we can conduct an identification analysis assuming we know

$$I = \frac{I^*}{\sigma_{Y_1 - Y_0}} = \frac{Y_1 - Y_0}{\sigma_{Y_1 - Y_0}}$$

  for each person where $D = \mathbf{1}(I > 0)$.
- See Cosslett (1983), Manski (1988), Matzkin (1992).
- Assumes there is an an instrument Z that shifts C.
- Even though we do not ever observe I, we observe (Y_0, D) and (Y_1, D). We never observe the full triple (Y_0, Y_1, D) for anyone.
- We only observe some components of C.

- Under conditions specified in the literature, $F(Y_0, I|X,Z)$ and $F(Y_1, I|X,Z)$ are identified where

$$\begin{aligned} Y_0 &= \mu_0(X) + U_0 & E(Y_0 \mid X) &= \mu_0(X) \\ Y_1 &= \mu_1(X) + U_1 & E(Y_1 \mid X) &= \mu_1(X) \\ I^* &= \mu_I(X,Z) + U_I & & \\ I &= \frac{\mu_I(X,Z)}{\sigma_{U_I}} + \frac{U_I}{\sigma_{U_I}} & & \end{aligned}$$

- Source: Heckman (1990), Heckman and Honoré (1990).
- The key idea in these papers is "sufficient" variation in Z holding X fixed.

THE UNIVERSITY OF CHICAGO

Sketch of the Proof

- From the left-hand side of

$$\Pr(D = 1|X, Z) = \Pr(\mu_I(X, Z) + U_I \geq 0|X, Z),$$

  we can identify the distribution of $\frac{U_I}{\sigma_{U_I}}$, as well as $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$.
- This is true under normality or any assumed form for the distribution of $\frac{U_I}{\sigma_{U_I}}$.
- It is also true more generally.
- One does not have to assume the distribution of U_I is known or that the functional form of $\mu_I(X, Z)$ is linear, e.g., $\mu_I(X, Z) = X\beta_I + Z\gamma$.
- See the conditions in the Matzkin (1992) paper and the survey in Matzkin, 2007, *Handbook of Econometrics*.

- This more general claim requires full support of Z and restrictions on $\mu_I(X, Z)$. See the "Matzkin conditions" in Cunha, Heckman, and Navarro (2007, IER).
- A key condition is

$$\text{Support}\left(\frac{\mu_I(X, Z)}{\sigma_{U_I}}\right) \supseteq \text{Support}\left(\frac{U_I}{\sigma_{U_I}}\right)$$

and other regularity conditions.
- Commonly it is assumed that for a fixed X

$$\text{Support}\left(\frac{\mu_I(X, Z)}{\sigma_{U_I}}\right) = (-\infty, \infty).$$

- This is called "identification at infinity." When we vary Z over its conditional support (for each X) we trace out the full support of $\frac{U_I}{\sigma_{U_I}}$.

Identifying the Joint Distribution of (Y_0, I)

- From data, we know the conditional distribution of Y_0:

$$F(Y_0 \mid D = 0, X, Z) = \Pr(Y_0 \leq y_0 \mid \mu_I(X, Z) + U_I \leq 0, X, Z)$$

- Multiply this by $\Pr(D = 0 \mid X, Z)$:

$$F(Y_0 \mid D = 0, X, Z) \Pr(D = 0 \mid X, Z) = \Pr(Y_0 \leq y_0, I^* \leq 0 \mid X, Z) \tag{*}$$

- Follow the analysis of Heckman (1990), Heckman and Smith (1998), and Carneiro, Hansen, and Heckman (2003).

- Left hand side of (*) is known from the data.
- Right hand side:

$$\Pr\left(Y_0 \leq y_0, \frac{U_I}{\sigma_{U_I}} < -\frac{\mu_I(X,Z)}{\sigma_{U_I}} \mid X, Z\right)$$

- Since we know $\dfrac{\mu_I(X,Z)}{\sigma_{U_I}}$ from the previous analysis, we can vary it for each fixed X.

- If $\mu_I(X, Z)$ gets small ($\mu_I(X, Z) \to -\infty$), recover the marginal distribution Y and in this limit set we can identify the marginal distribution of

$$Y_0 = \mu_0(X) + U_0 \quad \therefore \quad \text{can identify } \mu_0(X) \text{ in limit.}$$

- (See Heckman, 1990, and Heckman and Vytlacil, 2007.)
- More generally, we can form:

$$\Pr\left(U_0 \le y_0 - \mu_0(X), \frac{U_I}{\sigma_{U_I}} \le \frac{-\mu_I(X, Z)}{\sigma_{U_I}} \mid X, Z\right)$$

- X and Z can be varied and y_0 is a number. We can trace out joint distribution of $\left(U_0, \frac{U_I}{\sigma_{U_I}}\right)$ by varying (Y_0, Z) for each fixed X.

- $\therefore$ Recover joint distribution of

$$(Y_0, I) = \left(\mu_0(X) + U_0, \frac{\mu_I(X,Z) + U_I}{\sigma_{U_I}}\right).$$

- **Three key ingredients:**
  1. The independence of (U_0, U_I) and (X, Z).
  2. The assumption that we can set $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$ to be very small (so we get the marginal distribution of Y_0 and hence $\mu_0(X)$).
  3. The assumption that $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$ can be varied independently of $\mu_0(X)$.
- Trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_{U_I}}\right)$. Result generalizes easily to the vector case. (Carneiro, Hansen, and Heckman, 2003, IER; Heckman and Vytlacil, Part I).

THE UNIVERSITY OF CHICAGO

- Another way to see this is to write:

$$F(Y_0 \mid D = 0, X, Z)\,\Pr(D = 0 \mid X, Z)$$

- This is a function of $\mu_0(X)$ and $\dfrac{\mu_I(X, Z)}{\sigma_{U_I}}$ (Index sufficiency)

THE UNIVERSITY OF CHICAGO

- Varying the $\mu_0(X)$ and $\dfrac{\mu_I(X,Z)}{\sigma_{U_I}}$ traces out the distribution of $\left(U_0, \dfrac{U_I}{\sigma_{U_I}}\right)$.
- Effectively we observe the pairs $\left(\frac{I}{\sigma_{U_I}}, Y_1\right)$ and $\left(\frac{I}{\sigma_{U_I}}, Y_0\right)$.
- We never observe the triple $\left(\frac{I}{\sigma_{U_I}}, Y_0, Y_1\right)$.

- Use the intuition that we "know" I.
- Actually we observe

$$F(Y_0 \mid I < 0, X, Z)$$

and

$$F(Y_1 \mid I \geq 0, X, Z)$$

and

$$\Pr(I \geq 0 \mid X, Z)$$

- Can construct the joint distributions $F(Y_0, I \mid X, Z)$ and $F(Y_1, I \mid X, Z)$.

Roy Case

- Armed with normality (or the nonparametric assumptions in Heckman and Honoré, 1990), we can estimate

$$\mathrm{Cov}(I, Y_1) = \frac{\mathrm{Var}(Y_1) - \mathrm{Cov}(Y_0, Y_1)}{\sigma^2_{Y_1} + \sigma^2_{Y_0} - 2\sigma_{Y_1,Y_0}}$$
$$\mathrm{Cov}(I, Y_0) = -\frac{\mathrm{Var}(Y_0) - \mathrm{Cov}(Y_0, Y_1)}{\sigma^2_{Y_1} + \sigma^2_{Y_0} - 2\sigma_{Y_1,Y_0}}.$$

- We know Var Y_1, Var Y_0 (e.g. normal selection model or use limit sets).
- $\therefore$ $\mathrm{Cov}(Y_0, Y_1)$ is identified (actually over-identified).
- This line of argument does not generalize if we add a cost component (C) that is unobserved (or partly so).
- It carries through exactly if $C(Z)$ is solely a function of Z.

Intuition

- In the Roy model the decision rule is generated solely by (Y_1, Y_0).
- Knowing agent choices we observe the relative order (and magnitude) of Y_1 and Y_0.
- Thus we get a second valuable piece of information from agent choices. This information is ignored in statistical approaches to program evaluation.
- But does this analysis generalize?

Generalized Roy Model

- Add cost

$$I = Y_1 - Y_0 - C$$

- Assume that we do not directly observe C.

$$\text{Observe } Y_1 \mid I > 0,$$
$$\text{Observe } Y_0 \mid I < 0,$$

$$I = \frac{Y_1 - Y_0 - C}{\sqrt{\text{Var}(Y_1 - Y_0 - C)}}.$$

THE UNIVERSITY OF CHICAGO

- We can identify Var Y_1 and can identify Var Y_0.
- But we cannot directly identify $\text{Cov}(Y_0, Y_1)$ which measures comparative advantage in Willis-Rosen model.
- Notice, however, we can determine if

$$E(Y_1 \mid I > 0) > E(Y_1)$$
$$E(Y_0 \mid I < 0) > E(Y_0)$$

- (Are people who work in a sector above average for the sector?)