Treatment Effects II

Whitney Newey

Fall 2007

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Continuous Intstrument

Heckman and Vytlacil (2000, Economics Letters)

Model assumptions: Drop the i subscript.

1. Z is continuously distributed

2. $D = 1(D^* > 0),\ D^* = \mu(Z) - V$, V is continuously distributed.

3. Z and (Y_1, Y_0, V) independent

4. $E|Y_1| < \infty,\ E|Y_0| < \infty.$

Condition (2) is index model assumption.

Vytlacil (2002, Econometrica) shows that index model (assumption 2.) is equivalent to LATE when 1), 3), 4) hold.

Can add covariates X by making assumptions hold conditional on X. Here drop X.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

$$D = 1(D^* > 0),\ D^* = \mu(Z) - V$$

Let

$$P(Z) = \Pr(D = 1|Z); U = F_V(V).$$

Then

$$\begin{aligned} P(Z) &= \Pr(V \leq \mu(Z)) = F_V(\mu(Z)) \\ &= \Pr(U \leq P(Z)). \end{aligned}$$

A key identified object is

$$E[Y_1 - Y_0 \mid U = u].$$

Marginal Treatment Effect (MTE)

Average treatment effect for those individuals who are whose particpation would be affected if $P(Z)$ were greater or lower than u. Similar to LATE.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

To show identification of MTE note that $D = 1$ if and only if $U \leq P(Z)$, so

$$\begin{aligned}
E[Y \quad | \quad & P(Z) = p] = E[DY_1 + (1-D)Y_0 \mid P(Z) = p] \\
= \quad & E[Y_0 \mid P(Z) = p] + E[D(Y_1 - Y_0) \mid P(Z) = p] \\
= \quad & E[Y_0] + E[Y_1 - Y_0 \mid D = 1, P(Z) = p] \Pr(D = 1 \mid P(Z) = p) \\
= \quad & E[Y_0] + E[Y_1 - Y_0 \mid U \leq p]p \\
= \quad & E[Y_0] + E[E[Y_1 - Y_0 \mid U] \mid U \leq p]p \\
= \quad & E[Y_0] + E[1(u \leq p)E[Y_1 - Y_0 \mid U]|U \leq p]p \\
= \quad & E[Y_0] + \int_0^p E[Y_1 - Y_0 \mid U = u]du.
\end{aligned}$$

Summary: For all p in the support of $P(Z)$,

$$\frac{\partial}{\partial p} E[Y \mid P(Z) = p] = E[Y_1 - Y_0 \mid U = p].$$

MTE is identified over the support (range) of $P(Z)$.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare
(http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

$$\frac{\partial}{\partial p} E[Y \mid P(Z) = p] = E[Y_1 - Y_0 \mid U = p].$$

Interpretation:

$$\begin{aligned} E[Y_1 - Y_0 \quad \mid \quad U = p] &= \lim_{\tilde{p} \to p} \frac{E[Y \mid P(Z) = \tilde{p}] - E[Y \mid P(Z) = p]}{\tilde{p} - p} \\ &= \lim_{\tilde{p} \to p} \frac{E[Y \mid P(Z) = \tilde{p}] - E[Y \mid P(Z) = p]}{E[D \mid P(Z) = \tilde{p}] - E[D \mid P(Z) = p]} \end{aligned}$$

Infinitesimal IV.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Other treatment effects parameters can be written in terms of MTE.

Average Treatment Effect.

$$\begin{aligned} ATE &= E[Y_1 - Y_0] = E[E[Y_1 - Y_0 \mid U]] = \int_0^1 E[Y_1 - Y_0 \mid U = u]du \\ &= \int_0^1 \frac{\partial E}{\partial p}[Y \mid P(Z) = p] = E[Y \mid P(Z) = 1] - E[Y \mid P(Z) = 0]. \end{aligned}$$

ATE is identified at the boundary; Sometimes referred to as identification at infinity.

Need support of $P(Z)$ to include 1 and 0 to have identification.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Average effect of Treatment on the Treated

$$\begin{aligned}
TT &= E[Y_1 - Y_0 \mid D = 1] = E[Y_1 - Y_0 \mid U \le P(Z)] \\
&= E[E[Y_1 - Y_0 \mid U, P(Z)] \mid U \le P(Z)] \\
&= E[E[Y_1 - Y_0 \mid U] \mid U \le P(Z)] \\
&= \frac{E[1(U \le P(Z))E[Y_1 - Y_0 \mid U]]}{\Pr(U \le P(Z))} \\
&= \frac{E[E[1(U \le P(Z))|U]E[Y_1 - Y_0 \mid U]]}{E[E[1(U \le P(Z))|U]]} \\
&= \frac{E[\{1 - F_P(U)\}E[Y_1 - Y_0 \mid U]]}{E[1 - F_P(U)]}
\end{aligned}$$

Weighted average of MTE:

$$TT = \int W(u)\frac{\partial E[Y \mid P(Z) = u]}{\partial}du, w(u) = \frac{1 - F_p(u)}{E[1 - F_p(u)]}$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Another formula good for estimation

$$
\begin{aligned}
TT &= E[E[Y_1 - Y_0 \mid U] \mid U \leq P(Z)] \\
&= E[\left.\frac{\partial E[Y|P(Z)=p]}{\partial p}\right|_{p=U} \mid U \leq P(Z)] \\
&= E[\int_0^{P(Z)} \frac{\partial E[Y|P(Z)=p]}{\partial p} dp]/E[E[1(U \leq P(Z))|Z]] \\
&= E[\{E[Y|P(Z)] - E[Y|P(Z)=0]\}/E[P(Z)] \\
&= \frac{E[Y] - E[Y \mid P(Z)=0]}{E[P(Z)]}
\end{aligned}
$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Policy Effects.

ATE or TT does not directly answer policy questions.

Could look at policy changes, ask how change affects average outcome.

$$E[Y|after] - E[Y|before]$$

Y could be utility but in practice usually just use some observed Y (like log earnings).

Extension of Stock (1989) to treatment effects model where there is endogeneity.

Stock (1989): $g_0(X) = E[Y|X]$

$$\int g_0(x)\tilde{F}(dx) - \int g_0(x)F_0(dx).$$

Heckman, Vytlacil consider class of policies that affect $P(Z)$, probability of participation, but does not affect (Y_1, Y_0, U).

Economic binary choice interpretation: $P(Z)$ is transformation of differences in observed utilitities, U is unobserved individual heterogeneity. So policy affects observed part of utility.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Suppose each individual has new probability, denoted $\tilde{P}$.

$\tilde{D} = 1(U \leq \tilde{P})$ is new treatment.

Policy treatment effect is (definition)

$$\mu(\tilde{P}) = \frac{\int E[Y \mid P(Z) = p]F_{\tilde{P}}(dp) - \int E[Y \mid P(Z) = p]F_P(dp)}{E[\tilde{P}] - E[P]}.$$

Denominator is normalization. Makes this a weighted average of MTE.

Note that $E[Y|P(Z) = p]$ is a function of $E[Y_1 - Y_0|U]$ and so does not depend on the distribution of Z, and so is stable across movements from P to $\tilde{P}$.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

$$\mu(\tilde{P}) = \frac{\int E[Y \mid P(Z) = p] F_{\tilde{P}}(dp) - \int E[Y \mid P(Z) = p] F_P(dp)}{E[\tilde{P}] - E[P]}.$$

Identification requires $\tilde{P}$ included in support of P.

Recall

$$Y = 1(U \leq P(Z))Y_1 + 1(U > P(Z))Y_0$$

Also (Y_1, Y_0, U) independent of $P(Z)$.

Thus changing $\tilde{P}$ is change in exogenous part of model.

Can only answer question of whether this is an interesting policy effect in the context of some economic model.

IV has been criticized because it is not this.

Could also ask whether this is an interesting policy parameter in economics models. Not answered yet.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Policy affect is weighted average of MTE.

Let $\underline{p}$ be smallest point in the support of $P(Z)$. Then

$$E[Y \mid P(Z)=p] = \int_{\underline{p}}^{p} \frac{\partial E[Y \mid P(Z)=t]dt}{\partial t} + E[Y \mid P(Z)=\underline{p}].$$

Also by $\frac{\partial E[Y|P(Z)=t]}{\partial t} = E[Y_1 - Y_0|U=t] = MTE(t)$ It follows that

$$\begin{aligned}
\int E[Y \quad \mid \quad & P(Z)=p]F_{\tilde{P}}(dp) - \int E[Y \mid P(Z)=p]F_P(dp) \\
&= \int \int_{\underline{p}}^{p} \frac{\partial E[Y \mid P(Z)=t]}{\partial t} dt F_{\tilde{P}}(dp) - \int \int_{\underline{p}}^{p} \frac{\partial E[Y \mid P(Z)=t]}{\partial t} dt F_P(dp) \\
&= \int_{\underline{p}}^{\bar{p}} [\int_t^{\bar{p}} F_{\tilde{P}}(dp)] MTE(t)dt - \int_{\underline{p}}^{\bar{p}} \left[\int_t^{\bar{p}} F_P(dp)\right] MTE(t)dt \\
&= \int_{\underline{p}}^{\bar{p}} \left[F_P(t) - F_{\tilde{P}}(t)\right] MTE(t)dt
\end{aligned}$$

For denominator, $E[P] = E[1(U \leq P)] = E[\{1 - F_P(U)\}]$ so

$$E[\tilde{P}] - E[P] = E[\left\{1 - F_{\tilde{P}}(U)\right\}] - E[\{1 - F_P(U)\}] = \int_{\underline{p}}^{\bar{p}} \left[F_P(t) - F_{\tilde{P}}(t)\right] dt$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Summarizing:

$$\int E[Y \quad | \quad P(Z)=p]F_{\tilde{P}}(dp) - \int E[Y \mid P(Z)=p]F_P(dp)$$
$$= \int_{\underline{p}}^{\bar{p}} \left[F_P(t) - F_{\tilde{P}}(t)\right] MTE(t)dt$$

$$E[\tilde{P}] - E[P] = E[\left\{1 - F_{\tilde{P}}(U)\right\}] - E[\{1 - F_P(U)\}] = \int_{\underline{p}}^{\bar{p}} \left[F_P(t) - F_{\tilde{P}}(t)\right] dt$$

So,

$$\mu(\tilde{P}) = \int w(t)MTE(t)dt, w(t) = \frac{F_{\tilde{P}}(t) - F_P(t)}{\int_{\underline{p}}^{\bar{p}} \left[F_{\tilde{P}}(u) - F_P(u)\right] du}$$

Identification requires that the support of $\tilde{P}$ be contained in the support of P.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

What does instrumental variables estimate.

Consider instrument that is some transformation $J(P)$ of P.

Numerator of plim of IV estimator is

$$\begin{aligned} Cov(J(P),Y) &= E[\left(J-\bar{J}\right)E[Y|P]] = \int \{J(p)-\bar{J}\}\{\int_{\underline{p}}^{p} MTE(t)dt\}f_P(p)dp \\ &= \int [\int_t^{\bar{p}} \{J(p)-\bar{J}\}f_P(p)dp]MTE(t)dt. \end{aligned}$$

Denominator of plim of IV is

$$\begin{aligned} Cov(J(P),D) &= E[\left(J-\bar{J}\right)E[D|P]] = E[\left(J-\bar{J}\right)P] \\ &= \int \{J(p)-\bar{J}\}\{\int_{\underline{p}}^{p} dt\}f_P(p)dp \\ &= \int [\int_t^{\bar{p}} \{J(p)-\bar{J}\}f_P(p)dp]dt \end{aligned}$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Summarizing

$$
\begin{aligned}
Cov(J(P),Y) &= \int [\int_t^{\bar{p}} \{J(p)-\bar{J}\} f_P(p)dp] MTE(t)dt, \\
Cov(J(P),D) &= \int [\int_t^{\bar{p}} \{J(p)-\bar{J}\} f_P(p)dp] dt.
\end{aligned}
$$

So,

$$
\begin{aligned}
\frac{Cov(J(P),Y)}{Cov(J(P),D)} &= \int w(t)MTE(t)dt, w(t) = \frac{v(t)}{\int v(s)ds}, \\
v(t) &= \int_t^{\bar{p}} \{J(p)-\bar{J}\} f_P(p)dp
\end{aligned}
$$

Weight is positive if $J(p)$ is monotonic increasing. Why?

Could have negative weight if instrument not monotonic in P.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Selection on Observables

Another model is where conditioning on (or identifiable) variables X_i makes treatment exogenous.

Like removing omitted variable bias.

Assume

$$E[Y_{i0}|X_i, D_i] = E[Y_{i0}|X_i].$$

That is, Y_{i0} is mean independent of D_i conditional on X_i.

Conditional version of $E[Y_{i0}|D_i] = E[Y_{i0}]$, being a conditional version of that hypothesis.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

$$E[Y_{i0}|X_i, D_i] = E[Y_{i0}|X_i].$$

Where does X_i come from?

Would be great to have a model.

In some applications the source of X_i not clear.

Identification is fragile, requiring specifying just the right X_i.

Conditional mean independence that holds for X_i need not hold for a subset of X_i nor when additional variables are added to X_i.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

$$E[Y_{i0}|X_i, D_i] = E[Y_{i0}|X_i].$$

Need additional condition for identification of TT.

Let $\mathcal{X}$ denote the support of X_i (the smallest closed set having probability one).

Let $\mathcal{X}_0$ and $\mathcal{X}_1$ the support of X_i conditional on $D_i = 0$ and $D_i = 1$ respectively.

Common support condition

$$\mathcal{X} = \mathcal{X}_0 = \mathcal{X}_1.$$

Necessary and sufficient for $E[Y_i|X_i, D_i = 1]$ and $E[Y_i|X_i, D_i = 0]$ to be well defined for all X_i.

It is like the rank condition for identification in this setting.

It is verifiable; not often satisfied in practice.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Common support condition and conditional mean independence give

$$
\begin{aligned}
E[Y_i|X_i, D_i &= 1] - E[Y_i|X_i, D_i = 0] \\
&= E[\alpha_i + \beta_i|X_i, D_i = 1] - E[\alpha_i|X_i, D_i = 0] \\
&= E[\alpha_i|X_i, D_i = 1] - E[\alpha_i|X_i, D_i = 0] + E[\beta_i|X_i, D_i = 1] \\
&= E[\beta_i|X_i, D_i = 1].
\end{aligned}
$$

$E[\beta_i|X_i, D_i = 1]$ is TT conditional on X.

By iterated expectations

$$TT = E[\beta_i|D_i = 1] = E[\{E[Y_i|X_i, D_i = 1] - E[Y_i|X_i, D_i = 0]\}|D_i = 1].$$

Different notation:

$$E[Y|X, D] = \alpha(X) + \beta(X)D.$$

Then $E[Y_i|X_i, D_i = 1] - E[Y_i|X_i, D_i = 0] = \beta(X)$, so

$$TT = E[\beta(X_i)|D_i = 1].$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

For the ATE, assume Y_{i1} is mean independent of D_i conditional on X_i. In that case

$$\begin{aligned} E[\beta_i|X_i, D_i &= 1] = E[Y_{i1}|X_i, D_i = 1] - E[Y_{i0}|X_i, D_i = 1] \\ &= E[Y_{i1}|X_i,] - E[Y_{i0}|X_i] = E[\beta_i|X_i]. \end{aligned}$$

Therefore

$$\begin{aligned} ATE &= E[\beta_i] = E[\{E[Y_i|X_i, D_i = 1] - E[Y_i|X_i, D_i = 0]\}] \\ &= E[\beta(X_i)] \end{aligned}$$

ATE is a different function of the data than ATT.

$$\begin{aligned} TT &= E[\beta(X_i)|D_i = 1], \\ ATE &= E[\beta(X_i)]. \end{aligned}$$

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Estimating ATT and ATE:

Use nonparametric regression to get

$$\hat{\beta}(X) = \hat{E}[Y|X, D = 1] - \hat{E}[Y|X, D = 0].$$

Then

$$\begin{aligned} \widehat{TT} &= \sum_{i=1}^{n} D_i \hat{\beta}(X_i) / \sum_{i=1}^{n} D_i, \\ \widehat{ATE} &= \sum_{i=1}^{n} \hat{\beta}(X_i)/n. \end{aligned}$$

Root-n consistent under regularity conditions.

See Hahn (2004), Imbens and Ridder (2006) for different estimators.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Subject to curse of dimensionality when X_i is large dimensional.

Try to curb the curse using propensity score

$$P(X_i) = \Pr(D_i = 1|X_i) = E[D_i|X_i].$$

Have Rosenbaum and Rubin result that for $0 < P(X_i) < 1$,

$$E[\alpha_i|X_i, D_i] = E[\alpha_i|X_i] \Longrightarrow E[\alpha_i|P(X_i), D_i] = E[\alpha_i|P(X_i)].$$

For same reason as before, under $E[\alpha_i|X_i, D_i] = E[\alpha_i|X_i]$.

$$TT = E[\{E[Y_i|P(X_i), D_i = 1] - E[Y_i|P(X_i), D_i = 0]\}|D_i = 1].$$

Similarly, if in addition, $E[Y_{i0}|X_i, D_i] = E[Y_{i0}|X_i]$ then

$$ATE = E[\{E[Y_i|P(X_i), D_i = 1] - E[Y_i|P(X_i), D_i = 0]\}].$$

Now have one dimensional nonparametric regression (with dummy) if $P(X)$ known.

If $P(X_i)$ is completely unknown and unrestricted there is no known advantage for this approach.

It appears that advantage of using propensity score depends on knowing something about $P(X)$.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

To show propensity score result, let $P_i = P(X_i)$.

Theorem: For any W_i, $E[W_i|X_i, D_i] = E[W_i|X_i] \Longrightarrow E[W_i|P_i, D_i] = E[W_i|P_i]$

Proof: By iterated expectations,

$$E[D_i|P_i] = E[E[D_i|X_i]|P_i] = P_i$$

By iterated expectations again,

$$\begin{aligned}
E[W_i|P_i, D_i &= 1] = E[E[W_i|X_i, D_i = 1]|P_i, D_i = 1] = E[E[W_i|X_i]|P_i, D_i = 1] \\
&= \frac{E[D_i E[W_i|X_i]|P_i]}{E[D_i|P_i]} = \frac{E[P_i E[W_i|X_i]|P_i]}{P_i} \\
&= E[E[W_i|X_i]|P_i] = E[W_i|P_i].
\end{aligned}$$

Similarly, $E[W_i|P_i, D_i = 0] = E[W_i|P_i]$.

Cite as: Whitney Newey, course materials for 14.385 Nonlinear Econometric Analysis, Fall 2007. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].