Instituto Tecnológico Autónomo de México

Prof. Cristián Sánchez (cristian.sanchez@itam.mx))

Advanced Microeconometrics: Exam 1

Luis Gerardo Martínez Valdés

Spring 2022

Contents

1. MTE, Treatment Effects Paramaters, and Weights

In the framework of Heckman and Vytlacil (1999, 2005, 2007), and Heckman, Urzua, and Vytlacil (2006):

1. Show that any discrete choice model, $I = \eta(Z) - V$, may be written in equivalent form as,

$$I \geq 0 \Leftrightarrow \eta(Z) > V$$

or,

$$P(Z) > U_D,$$

where $P(Z) = F_V(\eta(Z))$, $U_D = F_V(V)$, F_V is the cdf of V (assumed strictly increasing), and U_D is uniformly distributed.

Answer:
(WLG) we can rewrite the treatment assignment of $I = \eta(Z) - V$ where, in the framework of the class and of that proposed by Heckman and Vytlacil (1999, 2005, 2007), and Heckman, Urzua, and Vytlacil (2006):

$$\begin{aligned} D = \mathbb{I}(I \geq 0) = \mathbb{I}(\eta(Z) > V) &= \mathbb{I}(F_{V|X}(\eta(Z) - F_{V|X}(V) > 0) \\ &= \mathbb{I}(F_V(\eta(Z) - F_V(V) > 0) \\ &= \mathbb{I}(P(Z) > U_D), \end{aligned}$$

where $P(Z) = F_V(\eta(Z)) = \mathbb{P}[D = 1 \mid Z]$.
Therefore, the *discrete choice model* can be expressed as

$$I \geq 0 \Leftrightarrow \eta(Z) > V$$

or,

$$P(Z) > U_D.$$

2. Define the MTE (MTE$(X = x, U_D = u_D)$).

Answer:
In the context of potential outcomes model:

$$MTE(x, u_D) = \mathbb{E}[Y_1 - Y_0 \mid X = x, U_D = u_D] = \mathbb{E}[\Delta \mid X = x, U_D = u_D],$$

where $\Delta = Y_1 - Y_0$.

3. Define the LIV (Local Instrumental Variables) estimator. Assume Z is continuous conditional on X (variables in outcome equations). Show that LIV identifies the MTE if the Z are continuous.

Answer:
Following the proof given by Heckman, Urzua and Vytlacil (2006)[1], conditional on $X = x$ **LIV** is the derivative of the conditonal expectations of $Y = DY_1 + (1 - D)Y_0$ w.r. to propensity scores $P(Z) = p$, one defines LIV as

$$LIV(x,p) \equiv \frac{\partial \mathbb{E}[Y \mid X = x, P(Z) = p]}{\partial p}$$

Now, if the Z are continuous one can identify **MTE** using LIV. Take $\Delta = Y_1 - Y_0$ and leaving the conditioning on X implicit, then it is quite ease to see, following the definition of Y [1] that [**?**],

$$\begin{aligned}
\mathbb{E}[Y|Z = z] = \mathbb{E}[Y|P(Z) = p] &= \mathbb{E}(Y = DY_1 + (1 - D)Y_0|P(Z) = p) \\
&= \mathbb{E}[Y_0] + \mathbb{E}(D(Y_1 - Y_0)|P(Z) = p) \\
&= \mathbb{E}[Y_0] + \mathbb{E}((Y_1 - Y_0)|D = 1)p \\
&= \mathbb{E}[Y_0] + \int_0^1 \mathbb{E}(Y = (Y_1 - Y_0)|U_D = u_D)du_D,
\end{aligned}$$

Taking derivatives w.r. to p,

$$\frac{\partial \mathbb{E}[Y \mid X = x, P(Z) = p]}{\partial p} = \mathbb{E}[U_1 - U_0|U_D = p] = MTE(p).$$

A more complete proof follows[4],

$$\mathbb{E}[Y|P(Z) = p] = \mathbb{E}[Y_0|P(Z) = p] + \mathbb{E}[\Delta|P(Z) = p, D = 1]p,$$

Assuming the random vectors $(U_1, U_D), (U_0, U_D)$ are independent of $Z|X$, then,

$$\mathbb{E}[Y|P(Z) = p] = \mathbb{E}[Y_0|P(Z) = p] + \mathbb{E}[\Delta|p \geq U_D]p,$$

Applying Wald[2] estimator for $p \neq p'$ values from $P(Z)$[3],

[1] $Y = DY_1 + (1 - D)Y_0$
[2] IV for two different values (binary instrument)

$$\frac{\mathbb{E}[Y|P(Z)=p]-\mathbb{E}[Y|P(Z)=p']}{p-p'} = ATE + \frac{\mathbb{E}[U_1-U_0|p \geq U_D p]-\mathbb{E}[U_1-U_0|p' \geq U_D p']}{p-p'}$$

When $U_0 \equiv U_1$ or $(U_1 - U_0) \perp U_D$ the second term on the right hand of the last expression vanishes. Then,

$$\mathbb{E}[Y|P(Z)=p] = \mathbb{E}[Y_0] + ATE * p + \int_0^p + \mathbb{E}[U_1-U_0|U_D=u_D]du_D$$

differentiating w.r to p, the LIV and the Fundamental Theorem of Calculus to finally, identify the $MTE(p)$:

$$MTE(p) = \frac{\partial \mathbb{E}[Y \mid X=x, P(Z)=p]}{\partial p} = ATE(p) + \mathbb{E}[U_1-U_0|U_D=p]$$

4. Now, suppose that the Z are discrete instruments (i.e. are discrete valued variables). What does IV identify?

   Answer:
   Authors' analysis on the undertsanding of what IV identify extends to the case where the distribution of $P(Z)$ (conditional on X) is discrete i.e. discrete instruments. Presented in Heckman, Urzua, Vytlacil (2006) and Heckman, Vytlacil(2007) we argument that IV estimates a wieghted average of LATE's, which can be interpreted as weights of MTE's –recall that LATE can be represented as a function of MTE [3]– therefore, one can identify the margins associated by the discrete instruments.

5. Prove that,

$$\begin{aligned} ATE(X) &= \int_0^1 MTE(X, u_D)du_D, \\ TUT(X) &= \int_0^1 MTE(X, u_D)\omega^{TUT}(u_D)du_D, \end{aligned}$$

   where $\omega^{TUT}(u_D) \geq 0$ for all u_D in $[0,1]$ and $\int_0^1 \omega^{TUT}(u_D)du_D = 1$. Derive the expression for ω^{TUT}.

   Answer:
   It is pretty straightforward to see that:

$$ATE(x) = \mathbb{E}[\Delta|X=x] = \int_0^1 \mathbb{E}[Y_1-Y_0|X=x, U_D=u_D]du_D = \int_0^1 MTE(x, u_D)du_D$$

[3]See. Heckman, Vytlacil (2005) [2], p. 680

Now, to prove for Treatmen on th Untreated (TUT), taking $u_D = u$, we define $(1-P(Z))TUT(x, P(Z), D=0) = \int_0^1 \mathbb{E}[\Delta|X=x, U_D=u]du$, then,

$$
\begin{aligned}
&TUT(x, D=0) && (1.1)\\
&= \int_0^1 \frac{1}{1-P(Z)} \left[\int_0^{P(Z)} \mathbb{E}[\Delta|X=x, U_D=u]du \right] dF_{1-P(Z)|X=x,D=0} && (1.2)\\
&= \frac{1}{1-P(Z)} \int_0^1 \left[\int_0^{P(Z)} \mathbb{E}[\Delta|X=x, U_D=u]du \right] dF_{1-P(Z)|X=x} && (1.3)\\
&= \frac{1}{1-P(Z)} \int_0^1 \left[\int_0^1 \mathbb{I}(u \leq 1-P(Z))\mathbb{E}[\Delta|X=x, U_D=u]du \right] dF_{1-P(Z)|X} && (1.4)\\
&= \frac{1}{\int_0^1 F_{1-P(Z)|X=x}(t)dt} \int_0^1 \left[\int_0^1 \mathbb{I}(P(Z) \leq 1-u)\mathbb{E}[\Delta|X=x, U_D=u]dF_{1-P(Z)|X} \right] du && (1.5)\\
&= \int_0^1 \mathbb{E}[\Delta|X=x, U_D=u] \left[\frac{F_{1-P(Z)|X}(u)}{\int F_{1-P(Z)|X}(u)dt} \right] du && (1.6)\\
&= \int_0^1 MTE(x,u)\omega^{TUT}(u)du && (1.7)
\end{aligned}
$$

Further analysis of this equation shows that $\mathbb{P}[D=0|X] = \mathbb{E}[1-P(Z)|X] = \int_0^1 1-(1-F_{1-P(Z))|X=x}(t))dt = \int_0^1 F_{1-P(Z))|X=x}(t)dt.$, where,

$$
\omega^{TUT}(u) = \frac{F_{1-P(Z)|X}(u)}{\int F_{1-P(Z)|X}(u)dt} = \frac{\mathbb{P}[P(Z) \leq 1-u|X]}{\mathbb{E}[1-P(Z)|X=x]} = \left[\int_0^u f(p|X=x)dp \right] \frac{1}{\mathbb{E}[1-P(Z)|X=x]}
$$

6. Define and derive the Policy Relevant Treatment Effect in terms of the MTE and compare it to ATE.

Answer:
In the context of Heckman, Vytlacil (2005) we denote a and a' two potential policies and D_a and $D_{a'}$ the choices that would be made under tis policies respectively. Further, we need to assume that the distribution of $(Y_{0,a}, Y_{1,a}, U_{D,a}) \mid X_a = x$ is the same as the distribution of $(Y_{0,a'}, Y_{1,a'}, U_{D,a'}) \mid X_{a'} = x$.

Using the Benthamite social welfare criterion $V(Y) = Y$, comparing policies using mean outcomes and considering the effect for individuales with $X = x$, we get the policy relevant treatment effect(PRTE),

$$
\begin{aligned}
PRTE(x) &= \mathbb{E}[Y_a|X=x] - \mathbb{E}[Y_{a'}|X=x] && (1.8)\\
&= \int_0^1 \mathbb{E}[\Delta|X=x, U_D=u_D]\Big[F_{P_{a'}|X}(U_D|x) - F_{P_a|X}(U_D|x)\Big]du_D && (1.9)\\
&= \int_0^1 MTE(x,u_D)\Big[F_{P_{a'}|X}(U_D|x) - F_{P_a|X}(U_D|x)\Big]du_D && (1.10)
\end{aligned}
$$

For this to make sense we will need to derive this result,

$$
\begin{aligned}
\mathbb{E}[Y_a|X=x] &= \int_0^1 \mathbb{E}[Y_a|X, P_a(Z_a)=p]dF_{P_a|X}(p)\\
&= \int_0^1 \Bigg[\int_0^1 \mathbb{I}_{(0,p)}(u_D)\mathbb{E}[Y_{1,a}|X, U_D=u_D]\\
&+ \mathbb{I}_{(p,1)}(u_D)\mathbb{E}[Y_{0,a}|X, U_D=u_D]du_D\Bigg]dF_{P_a|X}(p)\\
&= \int_0^1 \Bigg[\int_0^1 \mathbb{I}_{(u_D,1)}(p)\mathbb{E}[Y_{1,a}|X, U_D=u_D]\\
&+ \mathbb{I}_{(0,u_D)}(u_D)\mathbb{E}[Y_{0,a}|X, U_D=u_D]dF_{P_a|X}(p)\Bigg]du_D\\
&= \int_0^1 \Bigg[\int_0^1 (1-F_{P_a|X})(u_D)\mathbb{E}[Y_{1,a}|X, U_D=u_D]\\
&+ F_{P_a|X}(u_D)\mathbb{E}[Y_{0,a}|X, U_D=u_D]\Bigg]du_D
\end{aligned}
$$

The proof is analogous for $\mathbb{E}[Y_{a'}|X=x]$, now, changing the order of integration we will arrive at equation (1.10). If we assume weights of policy relevant treatment effect are all possitive (ω^{PRTE}) then they will be distributed in a vicinity of the ATE and then follow a downward parabolic shape from a cut point. Usually, more closely related to the TUT in the Roy model[4].

7. Derive the IV weights for IV(Z_1), where the vector of instruments is $Z = (Z_1, \ldots, Z_k)$, $k \geq 2$, and Z_1 is continuously distributed.

Answer:
A standard IV based on Z_1 can be written as,

$$
IV_{Z_1} = \int_0^1 MTE(u_D)\omega_{IV}^{Z_1}(u_D)du_D \qquad (1.11)
$$

[4]See Heckman, Vytlacil (2007) p. 4903

Take $\tilde{Z}_1 := Z_1 - \mathbb{E}[Z_1|X=x]$, then

$$
\begin{aligned}
&Cov(Z_1, Y|X=x) \\
&\quad = \mathbb{E}[(Z_1 - \mathbb{E}[Z_1|X=x])(D(Y_1 - Y_0)|X=x] \\
&\quad = \mathbb{E}[\tilde{Z}_1 \mathbb{I}[U_D \leq P(Z)]|X=x, U_D]\mathbb{E}[Y_1 - Y_0|X=x, U_D]|X=x] \\
&\quad = \int \left\{ \mathbb{E}[\tilde{Z}_1|X=x, u_D \leq P(Z)]Pr(u_D \leq P(Z)|X=x)\mathbb{E}[\Delta|X=x, U_D=u_D] \right\} du_D \\
&\quad = \int MTE(x, u_D)\mathbb{E}[\tilde{Z}_1|X=x, u_D \leq P(Z)]Pr(u_D \leq P(Z)|X=x)
\end{aligned}
$$

Considering the populationa analog of the IV estimator in our case $\frac{Cov(Z_1,Y|X)}{Cov(Z_1,D|X)}$, and equation (1.11) we can define the IV wieghts for a continuously distributed Z_1 as

$$\omega_{IV}(x, u_D) = \frac{\mathbb{E}[\tilde{Z}_1|X=x, u_D \leq P(Z)]Pr(u_D \leq P(Z)|X=x)}{Cov(Z_1, D|X)} \tag{1.12}$$

2. Semiparametric Estimation of the MTE

Consider the following model:

$$\begin{aligned} Y_1 &= \alpha_0 + \varphi + U_1 \\ Y_0 &= \alpha_0 + U_0, \end{aligned}$$

where U_0 and U_1 represent the unobservables in the potential outcome equations and φ represents the benefit associated with the treatment ($D = 1$). Individuals decide whether or not to receive the treatment ($D = 1$ or $D = 0$) based on a latent variable I:

$$I = Z\gamma - V,$$

where Z and V represent observables and unobservables, respectively. Thus, we can define a binary variable D indicating treatment status,

$$D = 1\left[I \geq 0\right].$$

Finally, we assume that the error terms in the model are not independent even conditioning on the observables, i.e. $U_1 U_0 V \mid Z$, but $(U_1, U_0, V)\, Z$.

1. Show that,

$$E\left(Y \mid P(Z) = p\right) = \alpha_0 + K(p),$$

where $Y = DY_1 + (1 - D)Y_0$, $P(Z)$ is the propensity score or probability of selection, p is a particular evaluation value of the propensity score, and

$$K(p) = \varphi p + E\left(U_0 \mid P(Z) = p\right) + E\left(U_1 - U_0 | D = 1, P(Z) = p\right) p.$$

Answer:

$$\begin{aligned} \mathbb{E}[Y|P(Z) = p] &= \mathbb{E}[DY_1 + (1 - D)Y_0|P(Z) = p] \\ &= \mathbb{E}[Y_0|P(Z) = p] + \mathbb{E}[D(Y_1 - Y_0)|P(Z) = p] \\ &= \mathbb{E}[Y_0|P(Z) = p] + \mathbb{E}[(Y_1 - Y_0)|D = 1, P(Z) = p]\ p \\ &= \mathbb{E}[\alpha_0 + U_0|P(Z) = p] + \mathbb{E}[(\alpha_0 + \varphi + U_1 - \alpha_0 - U_0)|D = 1, P(Z) = p]\ p \\ &= \mathbb{E}[\alpha_0|P(Z) = p] + \mathbb{E}[]U_0|P(Z) = p] + \mathbb{E}[\varphi + (U_1 - U_0)|D = 1, P(Z) = p]\ p \\ &= \alpha_0 + \mathbb{E}[U_0|P(Z) = p] + \mathbb{E}[\varphi|D = 1]\ p + \mathbb{E}[(U_1 - U_0)|D = 1, P(Z) = p]\ p \\ &= \alpha_0 + \mathbb{E}[U_0|P(Z) = p] + \varphi p + \mathbb{E}[(U_1 - U_0)|D = 1, P(Z) = p]\ p \\ &= \alpha_0 + K(p) \end{aligned}$$

Considering a different control function, let it be $\hat{K}(p) = \varphi p + \mathbb{E}[(U_1 - U_0)|D = 1, P(Z) = p]\, p$ and applying the underlying assumptions of Heckman, Urzua, and Vytlacil (2006) where (U_0, U_1, V) are independent of $Z|X$, meaning $\mathbb{E}[Y_0] = \mathbb{E}[\alpha_0] + \mathbb{E}[U_0] = \alpha_0$, then

$$
\begin{aligned}
\mathbb{E}[Y|P(Z) = p] &= \mathbb{E}[DY_1 + (1 - D)Y_0|P(Z) = p] \\
&= \mathbb{E}[Y_0] + \mathbb{E}[D(Y_1 - Y_0)|P(Z) = p] \\
&= \mathbb{E}[Y_0] + \mathbb{E}[(Y_1 - Y_0)|D = 1, P(Z) = p]\; p \\
&= \mathbb{E}[\alpha_0 + U_0] + \mathbb{E}[(\alpha_0 + \varphi + U_1 - \alpha_0 - U_0)|D = 1, P(Z) = p]\; p \\
&= \mathbb{E}[\alpha_0] + \mathbb{E}[U_0] + \mathbb{E}[\varphi + (U_1 - U_0)|D = 1, P(Z) = p]\; p \\
&= \alpha_0 + \mathbb{E}[\varphi|D = 1]\; p + \mathbb{E}[(U_1 - U_0)|D = 1, P(Z) = p]\; p \\
&= \alpha_0 + \varphi p + \mathbb{E}[(U_1 - U_0)|D = 1, P(Z) = p]\; p \\
&= \alpha_0 + \hat{K}(p)
\end{aligned}
$$

2. Use these two expressions to explain the steps involved in the semiparametric estimation of the Marginal Treatment Effect as presented in Heckman, Urzua, and Vytlacil (2006).

Answer:
Using the previous results we can show that

$$
\left.\frac{\partial \mathbb{E}[Y|X = x, P(Z) = p]}{\partial p}\right|_{p=u_D} = \left.\frac{\partial K(p)}{\partial p}\right|_{p=u_D}
$$

We can see that in this particular model we **do not** have the coeficients β_0, β_1 we can look past the firts 5 steps of the semiparametric estimation procedure. Thus, the problem reduces itself to the estimation of $\frac{\partial \hat{K}(p)}{\partial p}$. This quantity can be estimated using standar non-parametric techinques. Take,

$$
\tilde{Y} = K(\hat{P}(Z)) + \tilde{\nu}
$$

Assume $\mathbb{E}[\tilde{\nu}|\hat{P}(Z), X] = 0, \hat{P}(Z) = \hat{p}$. Then, we would only need to estimate $\frac{\partial \hat{K}(p)}{\partial \hat{p}}$ where $K(\hat{P}(Z)) = \mathbb{E}[\tilde{Y}|\hat{P}(Z) = \hat{p}]$.

Notice that we are defining s this estimator as a function of p insted of $\hat{P}(Z)$ and taking $\nu_1(p)$ as the non-parametric estimator of $\frac{\partial K(p)}{\partial p}$. WHY? Because, unlike the case we had the coefficients β_0, β_1 and using local linear regression(LLR), we now use a subset of values of $\hat{P}(Z)$ to define the set of points p on which our estimador is being evaluated.

Taking $\mathcal{P}$ as this particular set of points p iwhere we are evaluating our estimators, we can see that $\mathcal{P}$ containse tha values of $\hat{p} = \hat{P}(Z)$ for which we obtain positive frequencies in both $D = 0$ and $D = 1$ samples [1]. Thus, we can calculate $\nu_1(p)$ like

$$\{\nu_0(p), \nu_{1(}p)\} = \underset{\{\nu_0,\nu_1\}}{argmin} \sum_{i=1}^{N} \left\{ \tilde{Y}(i) - \nu_0 - \nu_1(\hat{P}(Z_i) - p)^2 \ \Psi[(\hat{P}(Z(i)) - p)/h] \right\}$$

where $\Psi(\cdot), h$ are the kernel function and the bandwith respectivley.[1]

Finally, we can estimate the MTE using the LIV estimator by,

$$LIV(x, u_D) = \frac{\partial \mathbb{E}[Y|X = x, P(Z) = p]}{\partial p} = \widehat{\partial K(p)/\partial p}\Bigg|_{p=u_D} = \widehat{\text{MTE}}(x, u_D)$$

which is evaluated for each $p \in \mathcal{P}$.

3. Explain in detail why you can approximate $K(p)$ with a polynomial on p. Explain how you can estimate the Marginal Treatment Effect using this approximation.

Answer:
Basically, $K(p)$ is a function of p over a set of points. So there ar a few ways to approximate it. One could use functional forms and a curvature approach,in this case, we can actually get into some trouble as we are estimating the MTE by purely identifying it because of those distributional or functional form assumptions.

On the other hand, we can use Taylor expansion to approximate $K(p)$ since it is actually a function of p. Now, the assumptions we need for Taylor's Expansion are met, and this can be easily proven. The most easy to see is that $K(p)$ is differentiable on p and the most important of this (conditions) being the mapping of p over a set of points.Thus, We can create or more accurately, extend this set into a continuous domain of p, satisfying the necessary conditions for Taylor's Theorem. Once we can approximate $K(p)$ by a polynomial on p –and may be a linear polynomial (degree one on p) by intervals; that it, segmenting the domain and approximating by linear polynomials(this is not necessary, but sufficient)– one can quickly apply OLS or $2SLS$ or GMM or LLR in order to estimate the MTE. This kind of methods are persistent in dynamic systems and differential equations(they like to linearize everything)

IDEA:
Take $K(p) = \varphi p + E(U_0 \mid P(Z) = p) + E(U_1 - U_0 | D = 1, P(Z) = p)\, p$, approximate by linear polynomial

$$Pol(p) = K(0) + K'(0)p$$

[1]See. Heckman, Urzua, Vytlacil (2006) p. 428

Bibliography

[1] James J Heckman, Sergio Urzua, and Edward Vytlacil. Understanding instrumental variables in models with essential heterogeneity. *The review of economics and statistics*, 88(3):389–432, 2006.

[2] James J Heckman and Edward Vytlacil. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738, 2005.

[3] James J Heckman and Edward J Vytlacil. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, 96(8):4730–4734, 1999.

[4] James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6:4875–5143, 2007.