Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models

Review of Economics and Statistics 86(1) (2004)

James J. Heckman and Salvador Navarro
University of Chicago

Econon 312, Spring 2021

THE UNIVERSITY OF CHICAGO

Two potential outcomes (Y_0, Y_1).
$D = 1$ if Y_1 is selected. $D = 0$ if Y_0 is selected.
Let V be utility.

$$V = \mu_V(Z, U_V) \quad D = 1\,(V > 0) \tag{1}$$

Z : observed factors determining choices, U_V unobserved

$$Y_1 = \mu_1(X, U_1) \tag{2a}$$

$$Y_0 = \mu_0(X, U_0). \tag{2b}$$

U_0, U_1, U_V are (absolutely) continuous

$$\Delta = Y_1 - Y_0.$$

Additively Separable Case: For Familiarity. Not essential.

$$V = \mu_V(Z) + U_V \quad E(U_V) = 0 \tag{1'}$$

$$Y_1 = \mu_1(X) + U_1 \quad E(U_1) = 0 \tag{2a'}$$

$$Y_0 = \mu_0(X) + U_0 \quad E(U_0) = 0. \tag{2b'}$$

THE UNIVERSITY OF CHICAGO

$$ATE \quad : \quad E(Y_1 - Y_0|X) \text{ (Average Treatment Effect)}$$

$$TT \quad : \quad E(Y_1 - Y_0|X, D = 1) \text{ (Treatment on the Treated)}$$

$$MTE \quad : \quad E(Y_1 - Y_0|X, Z, V) \text{ (Marginal Treatment Effect)}$$

These are familiar but by no means the only parameters we could consider

From *MTE*, can identify many other parameters

(Recall IV Lectures)

Samples generated by choices:

$$E(Y|X,Z,D=1)=E(Y_1|X,Z,D=1)$$

$$E(Y|X,Z,D=0)=E(Y_0|X,Z,D=0)$$

Data:

$$\Pr(D=1|X,Z)$$

$$E(Y_1|X,D=1) \text{ and } E(Y_0|X,D=0).$$

From raw means, we get biases.
Can form $E(Y_1|X,D=1)-E(Y_0|X,D=0)$.

In General This Produces BIASES

TT :

$$\begin{aligned}\text{Bias } TT &= [E(Y|X,D=1)-E(Y|X,D=0)] \\ &\quad -E(Y_1-Y_0|X,D=1) \\ &= [E(Y_0|X,D=1)-E(Y_0|X,D=0)]\,.\end{aligned}$$

Under Additive Separability

$$\text{Bias } TT = E(U_0|X,D=1)-E(U_0|X,D=0)$$

ATE :

$$\begin{aligned}\text{Bias } ATE &= E(Y|X,D=1)-E(Y|X,D=0) \\ &\quad -E(Y_1-Y_0|X)\,.\end{aligned}$$

Under Additive Separability

$$\begin{aligned} \text{Bias } ATE &= [E(U_1|X, D=1) - E(U_1|X)] \\ &\quad - [E(U_0|X, D=0) - E(U_0|X)] \end{aligned}$$

MTE :

$$\begin{aligned} \text{Bias } MTE &= E(Y|X, Z, D=1) - E(Y|X, Z, D=0) \\ &\quad - E(Y_1 - Y_0|X, Z, V) \end{aligned}$$

Under Additive Separability

$$\begin{aligned} \text{Bias } MTE &= E(U_1|X, Z, D=1) - E(U_1|X, Z, V) \\ &\quad - [E(U_0|X, Z, D=0) - E(U_0|X, Z, V)] \end{aligned}$$

Matching

$$W = (X, Z)$$

$$(Y_1, Y_0) \perp\!\!\!\perp D | W \qquad \text{(M-1)}$$

"$\perp\!\!\!\perp$" denotes independence given W

$$0 < \Pr(D = 1|W) = P(W) < 1, \qquad \text{(M-2)}$$

Rosenbaum and Rubin (1983) show (M-1) and (M-2) imply

$$(Y_1, Y_0) \perp\!\!\!\perp D | P(W). \qquad \text{(M-3)}$$

$$E(Y_1|D=0,P(W)) = E(Y_1|D=1,P(W)) = E(Y_1|P(W))$$

$$E(Y_0|D=1,P(W)) = E(Y_0|D=0,P(W)) = E(Y_0|P(W)).$$

Dependence between U_V and (U_1, U_0) is eliminated by conditioning on W:

$$U_V \perp\!\!\!\perp (U_1, U_0)|W.$$

"Selection on Observables"

If $P(W) = 1$ or $P(W) = 0$, method breaks down for those values.

Extensions (Heckman, Ichimura, Smith and Todd)

Distinction between X and Z

Introducing Z allows one to solve the breakdown problem arising from

$$P(X, Z) = 1 \text{ or } P(X, Z) = 0$$

Thus if outcomes are defined in terms of X and

$$Support\,(X|Z) = Support\,(X)$$

If we can find another value Z' such that

$$\Pr(X, Z') \neq 1,$$

can match using this (IV assumption)

Require only weaker mean independence assumptions

$$
\begin{aligned}
E(Y_1|W, D=1) &= E(Y_1|W) \\
E(Y_0|W, D=0) &= E(Y_0|W).
\end{aligned}
$$

Can be used for Means.

Matching is "for free" (Gill and Robins (2001)):

$E(Y_0|D=1,W)$ is not observed.

Can just as well replace it by

$$E(Y_0|D=1,W) = E(Y_0|D=0,W)$$

However, the implied economic restrictions are not "for free".

Imposes that, conditional on X and Z, the marginal person is the same as the average person.

This is the same as a flat $MTE(X,U)$ in U.
(MTE does not depend on U)

Additively Separable Case

We observe left-hand sides of

$$E(Y_1|X,Z,D=1)=\mu_1(X)+E(U_1|X,Z,D=1)$$

$$E(Y_0|X,Z,D=0)=\mu_0(X)+E(U_0|X,Z,D=0).$$

If $(U_1,U_V) \perp\!\!\!\perp X,Z$

$$E(U_1|X,Z,D=1)=E(U_1|\mu_V(Z)>U_V)=K_1(P(X,Z)).$$

If $(U_1,U_V) \perp\!\!\!\perp X,Z$

$$E(U_0|X,Z,D=0)=E(U_0|\mu_V(Z)\leq U_V)=K_0(P(X,Z))$$

So, key assumption

$$(U_1, U_0, U_V) \perp\!\!\!\perp (X, Z).$$

Under this condition

$$E(Y_1|X, Z, D = 1) = \mu_1(X) + K_1(P(X, Z))$$

$$E(Y_0|X, Z, D = 0) = \mu_0(X) + K_0(P(X, Z))$$

Need Limit Set Results

$$\lim_{P \to 1} K_1(P) = 0 \text{ and } \lim_{P \to 0} K_0(P) = 0$$

- If there are limit sets $\mathbb{Z}_0$ and $\mathbb{Z}_1$ such that $\lim_{Z \to \mathbb{Z}_0} P(X, Z) = 0$ and $\lim_{Z \to \mathbb{Z}_1} P(X, Z) = 1$, then we can identify the constants.
- There are semiparametric versions of these estimators.
- Use polynomials in P; Local Linear Regression in P.

From this model can obviously identify

$$ATE = \mu_1(X) - \mu_0(X)$$

(As we have seen)

Plus,

$$\begin{aligned} TT &= \mu_1(X) - \mu_0(X) + E(U_1 - U_0 | X, Z, D = 1) \\ &= \mu_1(X) - \mu_0(X) + K_1(P(X, Z)) \\ &\quad + \left(\frac{1 - P}{P}\right) K_0(P(X, Z)) \end{aligned}$$

$$
\begin{aligned}
MTE &= \mu_1(X) - \mu_0(X) \\
&\quad + \frac{\partial \left[E\left(U_1 - U_0 | X, Z, D = 1\right) P\left(X, Z\right) \right]}{\partial P\left(X, Z\right)} \\
&= \mu_1(X) - \mu_0(X) \\
&\quad + \frac{\partial \left[P\left(X, Z\right) \left\{ K_1\left(P(X, Z\right) + \frac{1-P}{P} K_0\left(P(X, Z\right) \right\} \right]}{\partial P\left(X, Z\right)}.
\end{aligned}
$$

Marginal and Average are allowed to be different.

Problem: Show this for $P(X, Z) = U(p)$

THE UNIVERSITY OF CHICAGO

Both Matching and Control functions are defined only over

$$\text{Support}\,(X|D=1)\cap\text{Support}\,(X|D=0)$$

Method of control functions does not require

$$(U_0, U_1) \perp\!\!\!\perp U_V|\,(X, Z)$$

But Matching does.

Matching is a special case of control functions in the additively separable case.

Additive separability and control functions assumptions are central to this claim.

$$
\begin{aligned}
E(U_1|X,Z,D=1) &= E(U_1|X,Z) = E(U_1|P(W)) \\
E(U_0|X,Z,D=0) &= E(U_0|X,Z) = E(U_0|P(W)).
\end{aligned}
$$

If

$$\mu_1(W) = E(Y_1|W) \text{ and } \mu_0(W) = E(Y_0|W)$$

then

$$E(U_1|P(W)) = 0 \text{ and } E(U_0|P(W)) = 0$$

However, this is not strictly required

In the method of control functions,

$$\text{If } (X, Z) \perp\!\!\!\perp (U_0, U_1, U_V)$$

$$
\begin{aligned}
& E(Y|X, Z, D) \\
= \;& E(Y_1|X, Z, D = 1) D + E(Y_0|X, Z, D = 0)(1 - D) \\
= \;& \mu_0(X) + (\mu_1(X) - \mu_0(X)) D + E(U_1|X, Z, D = 1) D \\
& + E(U_0|P(X, Z), D = 0)(1 - D) \\
= \;& \mu_0(X) + (\mu_1(X) - \mu_0(X)) D + E(U_1|P(X, Z), D = 1) D \\
& + E(U_0|P(X, Z), D = 0)(1 - D) \\
= \;& \mu_0(X) \\
& + [\mu_1(X) - \mu_0(X) + K_1(P(X, Z)) - K_0(P(X, Z))] D \\
& + K_0(P(X, Z)).
\end{aligned}
$$

To identify

$$\mu_1(X) - \mu_0(X)$$

must isolate it from

$$K_1(P(X,Z))$$

and

$$K_0(P(X,Z)).$$

Tables 1 and 2 present sensitivity analysis for the case of

$$
\begin{aligned}
(U_1, U_0, U_V)' &\sim N(0, \Sigma) \\
corr(U_j, U_V) &= \rho_{jV} \\
var(U_j) &= \sigma_j^2; \quad j = \{0, 1\}.
\end{aligned}
$$

so

$$
\begin{aligned}
\text{Bias } TT(P(Z) = p) &= \sigma_0 \rho_{0V} M(p) \\
\text{Bias } ATE(P(Z) = p) &= M(p)\left[\sigma_1 \rho_{1V} \ (1 - p) + \sigma_0 \rho_{0V} \ p\right]
\end{aligned}
$$

where $M(p) = \frac{\phi\left(\Phi^{-1}(1-p)\right)}{p(1-p)}$

Problem: Using the Generalized Roy model derive these results for bias

THE UNIVERSITY OF CHICAGO

Table 1

Mean Bias for Treatment on the Treated

ρ_{0V}	Average Bias ($\sigma_0 = 1$)	Average Bias ($\sigma_0 = 2$)
-1.00	-1.7920	-3.5839
-0.75	-1.3440	-2.6879
-0.50	-0.8960	-1.7920
-0.25	-0.4480	-0.8960
0.00	0.0000	0.0000
0.25	0.4480	0.8960
0.50	0.8960	1.7920
0.75	1.3440	2.6879
1.00	1.7920	3.5839

$\text{BIASTT} = \rho_{0V} * \sigma_0 * M(p)$

$M(p) = \frac{\varphi(\Phi^{-1}(p))}{[p*(1-p)]}$

Table 2

Mean Bias for Average Treatment Effect

$(\sigma_0 = 1)$

ρ_{0V}	-1.00	-0.75	-0.50	-0.25	0	0.25	0.50	0.75	1.00
					$\rho_{1V}(\sigma_1 = 1)$				
-1.00	-1.7920	-1.5680	-1.3440	-1.1200	-0.8960	-0.6720	-0.4480	-0.2240	0
-0.75	-1.5680	-1.3440	-1.1200	-0.8960	-0.6720	-0.4480	-0.2240	0	0.2240
-0.50	-1.3440	-1.1200	-0.8960	-0.6720	-0.4480	-0.2240	0	0.2240	0.4480
-0.25	-1.1200	-0.8960	-0.6720	-0.4480	-0.2240	0	0.2240	0.4480	0.6720
0	-0.8960	-0.6720	-0.4480	-0.2240	0	0.2240	0.4480	0.6720	0.8960
0.25	-0.6720	-0.4480	-0.2240	0	0.2240	0.4480	0.6720	0.8960	1.1200
0.50	-0.4480	-0.2240	0	0.2240	0.4480	0.6720	0.8960	1.1200	1.3440
0.75	-0.2240	0	0.2240	0.4480	0.6720	0.8960	1.1200	1.3440	1.5680
1.00	0	0.2240	0.4480	0.6720	0.8960	1.1200	1.3440	1.5680	1.7920
					$\rho_{1V}(\sigma_1 = 2)$				
-1.00	-2.6879	-2.2399	-1.7920	-1.3440	-0.8960	-0.4480	0	0.4480	0.8960
-0.75	-2.4639	-2.0159	-1.5680	-1.1200	-0.6720	-0.2240	0.2240	0.6720	1.1200
-0.50	-2.2399	-1.7920	-1.3440	-0.8960	-0.4480	0	0.4480	0.8960	1.3440
-0.25	-2.0159	-1.5680	-1.1200	-0.6720	-0.2240	0.2240	0.6720	1.1200	1.5680
0	-1.7920	-1.3440	-0.8960	-0.4480	0	0.4480	0.8960	1.3440	1.7920
0.25	-1.5680	-1.1200	-0.6720	-0.2240	0.2240	0.6720	1.1200	1.5680	2.0159
0.50	-1.3440	-0.8960	-0.4480	0	0.4480	0.8960	1.3440	1.7920	2.2399
0.75	-1.1200	-0.6720	-0.2240	0.2240	0.6720	1.1200	1.5680	2.0159	2.4639
1.00	-0.8960	-0.4480	0	0.4480	0.8960	1.3440	1.7920	2.2399	2.6879

BIASATE $= \rho_{1V} * \sigma_1 * M_1(p) - \rho_{0V} * \sigma_0 * M_0(p)$

BIASMTE $=$ BIASATE $-\Phi^{-1}(1-p) * (\rho_{1V} * \sigma_1 - \rho_{0V} * \sigma_0)$

$M_1(p) = \frac{\varphi(\Phi^{-1}(p))}{p}$

$M_0(p) = \frac{-\varphi(\Phi^{-1}(p))}{[1-p]}$

THE UNIVERSITY OF CHICAGO

Matching and Method of control functions work with $E(Y|X,Z,D)$ and $\Pr(D=1|X,Z)$.

$$\begin{aligned} Y &= DY_1 + (1-D)Y_0 \\ &= \mu_0(X) + (\mu_1(X) - \mu_0(X) + U_1 - U_0)D + U_0 \\ &= \mu_0(X) + \Delta(X)D + U_0 \end{aligned}$$

If $U_1 = U_0$

$$E(U_0|P(X,Z),X) = E(U_0|X) \quad \text{(IV-1)}$$

$$\Pr(D=1|X,Z) \text{ is a nontrivial function of } Z \text{ for each } X. \quad \text{(IV-2)}$$

When

$$U_1 \neq U_0, \text{ but } D \perp\!\!\!\perp (U_1 - U_0) | X$$

or alternatively

$$U_V \perp\!\!\!\perp (U_1 - U_0 | X),$$

we have all three mean treatment effects are the same

$$\begin{aligned} ATE &= E(Y_1 - Y_0 | X) = E(\Delta(X) | X) \\ TT &= E(Y_1 - Y_0 | X, D = 1) = E(Y_1 - Y_0 | X) \\ &= MTE \end{aligned}$$

Analytically More Interesting Case

$U_1 \neq U_0$ and $D \not\perp\!\!\!\perp (U_1 - U_0)$

For ATE :

$$E\left(U_0 + D\left(U_1 - U_0\right)|P\left(X, Z\right), X\right) = E\left(U_0 + D\left(U_1 - U_0\right)|X\right) \qquad \text{(IV-3)}$$

For TT :

$$\begin{aligned} &E\left(U_0 + D\left(U_1 - U_0\right) - E\left(U_0 + D\left(U_1 - U_0\right)|X\right)|P\left(X, Z\right), X\right) \\ &\quad = E\left(U_0 + D\left(U_1 - U_0\right) - E\left(U_0 + D\left(U_1 - U_0\right)|X\right)|X\right) \end{aligned}$$

For ATE we can rewrite:

$$\begin{aligned} &E\left(U_0|P\left(X, Z\right), X\right) + E\left(U_1 - U_0|D = 1, P\left(X, Z\right), X\right)P\left(X, Z\right) \\ &\quad = E\left(U_0|X\right) + E\left(U_1 - U_0|D = 1, X\right)P\left(X, Z\right) \end{aligned}$$

All mean parameters are the same if $U_1 = U_0$, or $(U_1 - U_0) \perp\!\!\!\perp D|P\left(X, Z\right), X$

1. Method of Control Functions Models Dependence between (U_1, U_0) and V.
2. Matching assumes $(U_1, U_0) \perp\!\!\!\perp V \mid X, Z$.
3. Z independent of U_0, U_1 conditional on X.

Local Instrumental Variables (LIV) require that

$$\mu_D(Z) \quad \text{be a non-degenerate random variable given } X \text{ (existence of an exclusion restriction)} \tag{2}$$

$$(U_0, U_1, U_V) \perp\!\!\!\perp Z|X \tag{LIV-2}$$

$$0 < \Pr(D|X) < 1 \tag{LIV-3}$$

$$\text{Support } P(D|(X,Z)) = [0,1] \tag{LIV-4}$$

Under these conditions,

$$\frac{\partial E(Y|X, P(X,Z))}{\partial(P(Z))} = MTE(X, P(Z), V)$$

Table 3
Identifying Assumptions and Implicit Economic Assumptions Underlying the Four Methods Discussed in this Paper
Conditional on X and Z

Method	Exclusion Required?	Separability of Observables and Unobservables in Outcome Equations?	Functional Forms Required?	Marginal = Average? (Given X, Z)	Key Identification Condition for Means (assuming separability)
Matching*	No	No	No	Yes	$E(U_1\|X, D=1, Z) = E(U_1\|X, Z)$ $E(U_0\|X, D=0, Z) = E(U_0\|X, Z)$
Control Function**	Yes (for nonparametric identification)	Conventional, but not required	Conventional, but not required	No	$E(U_0\|X, D=0, Z)$ and $E(U_1\|X, D=1, Z)$ can be varied independently of $\mu_0(X)$ and $\mu_1(X)$, respectively and intercepts can be identified through limit arguments or symmetry assumptions
IV (conventional)	Yes	Yes	No	No (Yes in standard case)	$E(U_0 + D(U_1 - U_0)\|X, Z)$ $= E(U_0 + D(U_1 - U_0)\|X)$ (ATE) $E(U_0 + D(U_1 - U_0) - E(U_0 + D(U_1 - U_0)\|X)\|P(Z), X)$ $= E(U_0 + D(U_1 - U_0) - E(U_0 + D(U_1 - U_0)\|X)\|X)$ (TT)
LIV	Yes	No	No	No	$(U_0, U_1, U_v) \perp\!\!\!\perp Z\|X$ $\Pr(D=1\|Z, X)$ is a nontrivial function of Z for each X.

*For propensity score matching, (X, Z) are replaced with $P(X, Z)$ in defining parameters and conditioning sets.

**Conditions for writing the control function in terms of $P(X, Z)$ are given in the text.

THE UNIVERSITY OF CHICAGO

Fundamental Problem: Information of the Analyst often less than that of the Agent.

Definition 1

We say that $\sigma\left(I_{R^*}\right)$ is a **relevant information set** if its associated random variable, I_{R^*}, satisfies (M-1) so

$$\left(Y_1, Y_0\right) \perp\!\!\!\perp D | I_{R^*}$$

Definition 2

We say that $\sigma(I_R)$ is a **minimal relevant information set** if it is the intersection of all sets $\sigma(I_{R^*})$ and $(Y_1, Y_0) \perp\!\!\!\perp D | I_R$. The associated random variable I_R is the minimum amount of information that guarantees that (M-1) is satisfied. Intersection may be empty. May not be a unique minimal information set.

Definition 3

The agent's information set, $\sigma(I_A)$, is defined by the information I_A used by the agent when choosing among treatments. Accordingly, we call I_A the **agent's information**.

Definition 4

The econometrician's **full information set**, $\sigma\left(I_{E^*}\right)$, is defined by **all** the information **available** to the econometrician, I_{E^*}.

Definition 5

The econometrician's **information set**, $\sigma\left(I_E\right)$, is defined by the information **used** by the econometrician when analyzing the agent's choice of treatment, I_E.

Obvious Inclusions: $\sigma(I_R) \subseteq \sigma(I_{R^*}), \sigma(I_R) \subseteq \sigma(I_A)$ and $\sigma(I_E) \subseteq \sigma(I_{E^*})$

This assumes I_R exists.

Matching implies

$$\sigma(I_R) \subseteq \sigma(I_E)$$

Generalized Roy Examples (Assume Factor Structure for error terms)

- Consider bias from matching with different information sets (i.e., different p specifications).
- Remember the Rosenbaum and Rubin result.

$$\begin{aligned} V &= Z\gamma + U_V \\ &= Z\gamma + \alpha_{V1} f_1 + \alpha_{V2} f_2 + \varepsilon_V, \\ D &= 1 \text{ if } V \geq 0, \quad = 0 \text{ otherwise} \end{aligned}$$

$$\begin{aligned} Y_1 &= \mu_1 + U_1 = \mu_1 + \alpha_{11} f_1 + \alpha_{12} f_2 + \varepsilon_1 \\ Y_0 &= \mu_0 + U_0 = \mu_0 + \alpha_{01} f_1 + \alpha_{02} f_2 + \varepsilon_0, \end{aligned}$$

$(f_1, f_2, \varepsilon_V, \varepsilon_1, \varepsilon_0)$ mean zero random variables, mutually independent of each other and Z

The minimal relevant information set when factor loadings are not zero:

$$I_R = \{f_1, f_2\}.$$

Agent information sets may include different variables. If shocks to the outcomes not known, but other terms are:

$$I_A = \{f_1, f_2, Z, \varepsilon_V\}$$

Under perfect certainty, $I_A = \{f_1, f_2, Z, \varepsilon_V, \varepsilon_1, \varepsilon_0\}$.

Construct examples using:

$$(f_1, f_2, \varepsilon_V, \varepsilon_1, \varepsilon_0) \sim N(0, \Sigma),$$

$$diag(\Sigma) = \left(\sigma^2_{f_1}, \sigma^2_{f_2}, \sigma^2_{\varepsilon_V}, \sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_0}\right)$$

Suppose $I_E = \{Z, f_1, f_2\}$

$$\begin{aligned} & E(Y_1 | D = 1, I_E) - E(Y_0 | D = 0, I_E) \\ = \ & \mu_1 - \mu_0 + (\alpha_{11} - \alpha_{01}) f_1 + (\alpha_{12} - \alpha_{02}) f_2 \end{aligned}$$

Knowledge of (Z, f_1, f_2) and of $P(Z, f_1, f_2)$ equivalent

$$\begin{aligned} P(I_E) &= \Pr\left(\frac{\varepsilon_V}{\sigma_{\varepsilon_V}} > \frac{-Z\gamma - \alpha_{V1} f_1 - \alpha_{V2} f_2}{\sigma_{\varepsilon_V}}\right) \\ &= 1 - \Phi\left(\frac{-Z\gamma - \alpha_{V1} f_1 - \alpha_{V2} f_2}{\sigma_{\varepsilon_V}}\right) = p \end{aligned}$$

$$E\left(Y_1|D=1,P\left(I_E\right)=p\right)-E\left(Y_0|D=0,P\left(I_E\right)=p\right)$$

$$\begin{aligned}
&= \mu_1-\mu_0+E\left(U_1|D=1,P\left(I_E\right)=p\right) \\
&\quad -E\left(U_0|D=0,P\left(I_E\right)=p\right) \\
&= \mu_1-\mu_0+E\left(U_1|\frac{\varepsilon_V}{\sigma_{\varepsilon_V}}>\Phi^{-1}\left(1-p\right)\right) \\
&\quad -E\left(U_0|\frac{\varepsilon_V}{\sigma_{\varepsilon_V}}\leq\Phi^{-1}\left(1-p\right)\right) \\
&= \mu_1-\mu_0
\end{aligned}$$

All the treatment parameters equal

$(MTE = ATE = LATE = TT)$

$$E\left(U_1 | \frac{\varepsilon_V}{\sigma_{\varepsilon_V}} > \Phi^{-1}(1-p)\right) = \frac{COV(U_1, \varepsilon_V)}{\sigma_{\varepsilon_V}} M_1(P)$$
$$E\left(U_0 | \frac{\varepsilon_V}{\sigma_{\varepsilon_V}} \leq \Phi^{-1}(1-p)\right) = \frac{COV(U_0, \varepsilon_V)}{\sigma_{\varepsilon_V}} M_0(P)$$

where

$$M_1(P) = \frac{\phi(\Phi^{-1}(1-p))}{p} \quad \text{and} \quad M_0(P) = \frac{\phi(\Phi^{-1}(1-p))}{1-p}$$

because

$$COV(U_i, \varepsilon_V) = COV(\alpha_{i1} f_1 + \alpha_{i2} f_2 + \varepsilon_i, \varepsilon_V) = 0, \quad i = 0, 1.$$

$$I_E = \{Z\}$$

$$\frac{z\gamma + \alpha_{V1} f_1 + \alpha_{V2} f_2 + \varepsilon_V}{\sqrt{\alpha_{V1}^2 \sigma_{f_1}^2 + \alpha_{V2}^2 \sigma_{f_2}^2 + \sigma_{\varepsilon_V}^2}} = \Phi^{-1}(1-p).$$

$$\text{Bias } TT(P(Z) = p) = \beta_0 M(p),$$

$$M(P) = M_1(P) - M_0(P)$$

$$\text{Bias } ATE(P(Z) = p) = M(p)\left[\beta_1 (1-p) + \beta_0 p\right]$$

THE UNIVERSITY OF CHICAGO

$$\begin{aligned} \text{Bias } MTE\left(P\left(Z\right)=p\right) &= M(p)\left[\beta_1\left(1-p\right)+\beta_0 p\right] \\ &\quad -\Phi^{-1}\left(1-p\right)\left[\beta_1-\beta_0\right] \end{aligned}$$

where

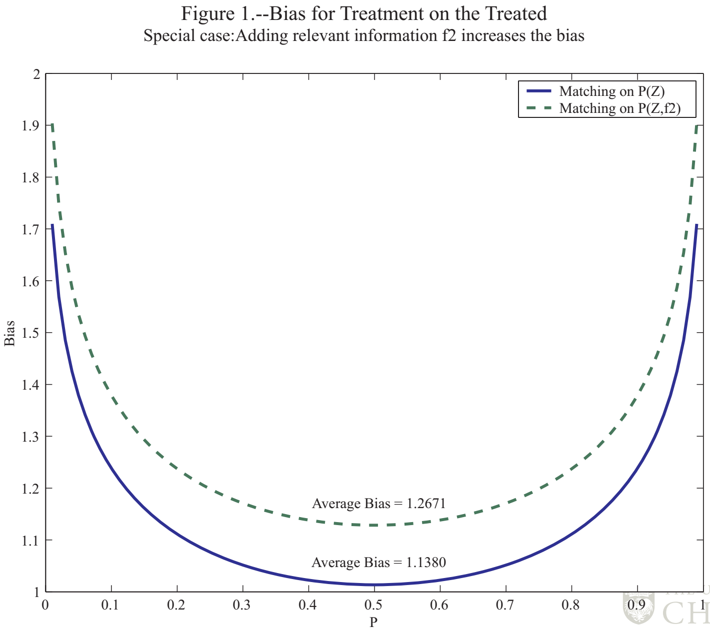
$$\begin{aligned} M(p) &= \frac{\phi\left(\Phi^{-1}\left(1-p\right)\right)}{p\left(1-p\right)} \\ \beta_1 &= \frac{\alpha_{V1}\alpha_{11}\sigma_{f_1}^2+\alpha_{V2}\alpha_{12}\sigma_{f_2}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2+\alpha_{V2}^2\sigma_{f_2}^2+\sigma_{\varepsilon_V}^2}} \\ \beta_0 &= \frac{\alpha_{V1}\alpha_{01}\sigma_{f_1}^2+\alpha_{V2}\alpha_{02}\sigma_{f_2}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2+\alpha_{V2}^2\sigma_{f_2}^2+\sigma_{\varepsilon_V}^2}} \end{aligned}$$

Problem: Verify the equations on all slides.

$$I'_E = \{Z, f_2\}$$

May raise or lower the bias.

Link



THE UNIVERSITY OF CHICAGO

Figure 2.--Bias for Average Treatment Effect
Special case: Adding relevant information f2 increases the bias

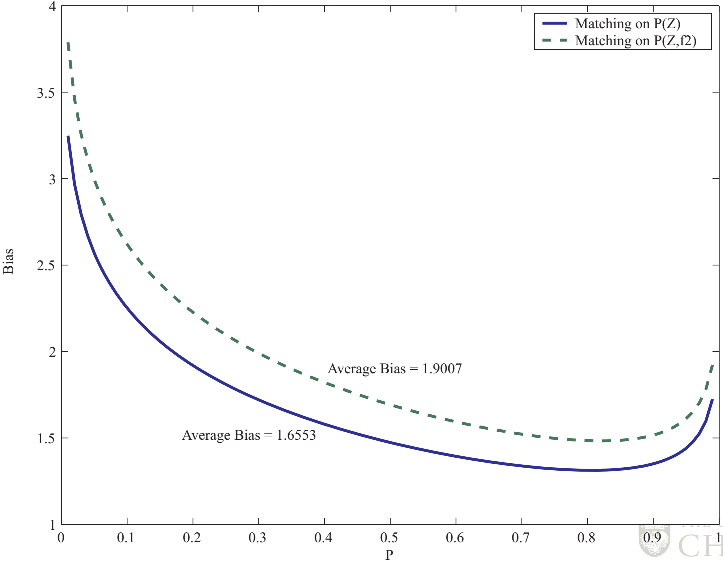
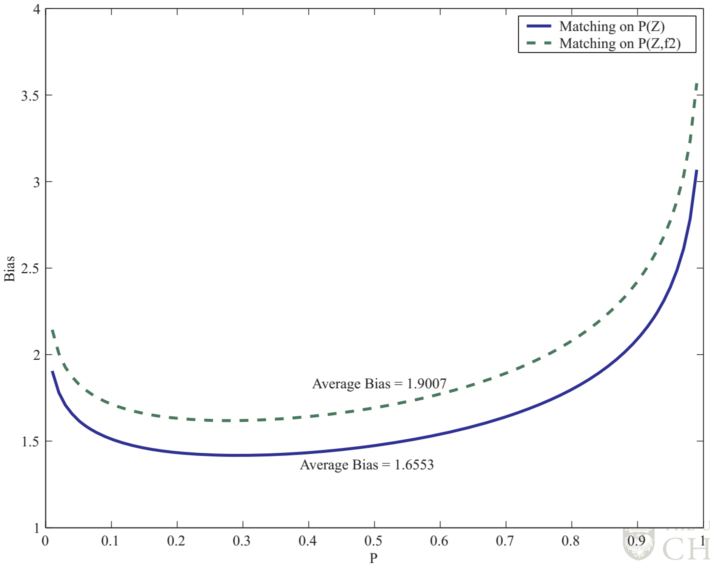


Figure 3.--Bias for Marginal Treatment Effect
Special case: Adding relevant information f2 increases the bias



THE UNIVERSITY OF CHICAGO

Control Function Method Models the Bias

In control function method, adding f_2 we simply change the control function. We go from

$$\begin{aligned} K_1\left(P\left(Z\right)=p\right) &= \beta_1 M_1\left(p\right) \\ K_0\left(P\left(Z\right)=p\right) &= -\beta_0 M_0\left(p\right) \end{aligned}$$

to

$$\begin{aligned} K_1'\left(P\left(Z,f_2\right)=p\right) &= \beta_1' M_1\left(p\right) \\ K_0'\left(P\left(Z,f_2\right)=p\right) &= -\beta_0' M_0\left(p\right) \end{aligned}$$

where $M_1\left(p\right)=\frac{\phi\left(\Phi^{-1}(1-p)\right)}{p}$ and $M_0\left(p\right)=\frac{\phi\left(\Phi^{-1}(1-p)\right)}{1-p}$

This protects us against misspecification.

THE UNIVERSITY OF CHICAGO

Suppose we do not know f_2, just proxy it by $\widetilde{Z}$

$$\widetilde{I}_{E^*} = \left\{ Z, \widetilde{Z} \right\}.$$

$$\text{Suppose } I_E = \widetilde{I}_{E^*}$$

Suppose further that

$$\begin{aligned} \widetilde{Z} &\sim N\left(0, \sigma^2_{\widetilde{Z}}\right) \\ corr\left(\widetilde{Z}, f_2\right) &= \rho, \text{ and } \widetilde{Z} \perp\!\!\!\perp \left(\varepsilon_0, \varepsilon_1, \varepsilon_V, f_1\right). \end{aligned}$$

THE UNIVERSITY OF CHICAGO

Expressions corresponding to β_0 and β_1 :

$$\widetilde{\beta}_1 = \frac{\alpha_{11}\alpha_{V1}\sigma_{f_1}^2 + \alpha_{12}\alpha_{V2}\left(1-\rho^2\right)\sigma_{f_2}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \alpha_{V2}^2\sigma_{f_2}^2\left(1-\rho^2\right) + \sigma_{\varepsilon_V}^2}}$$

$$\widetilde{\beta}_0 = \frac{\alpha_{01}\alpha_{V1}\sigma_{f_1}^2 + \alpha_{02}\alpha_{V2}\left(1-\rho^2\right)\sigma_{f_2}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \alpha_{V2}^2\sigma_{f_2}^2\left(1-\rho^2\right) + \sigma_{\varepsilon_V}^2}}$$

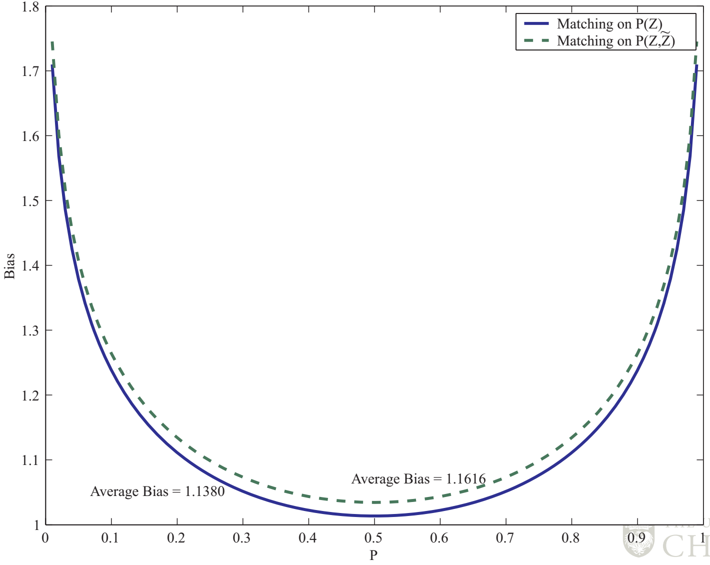
- By substituting I_E' for $\widetilde{I}_E$ and β_j' for $\widetilde{\beta}_j$ $(j = 0, 1)$ into Conditions (1), (2) and (3) we obtain equivalent results for this case. Whether $\widetilde{I}_E$ will be bias reducing depends on how well it spans I_R and the signs of the terms in the absolute values.
- In this case, there is another parameter ρ $(\rho = 0)$.

- The bias generated when the econometrician's information is $\widetilde{I}_E$ can also be smaller than when it is I_E'. It can be the case that knowing *the proxy* variable $\widetilde{Z}$ is *better* than knowing the actual variable f_2. Take treatment on the treated as the parameter. The bias is reduced when $\widetilde{Z}$ is used instead of f_2 if

$$\left| \frac{\alpha_{01}\alpha_{V1}\sigma_{f_1}^2 + \alpha_{02}\alpha_{V2}\left(1-\rho^2\right)\sigma_{f_2}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \alpha_{V2}^2\sigma_{f_2}^2\left(1-\rho^2\right) + \sigma_{\varepsilon_V}^2}} \right| < \left| \frac{\alpha_{01}\alpha_{V1}\sigma_{f_1}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \sigma_{\varepsilon_V}^2}} \right|.$$

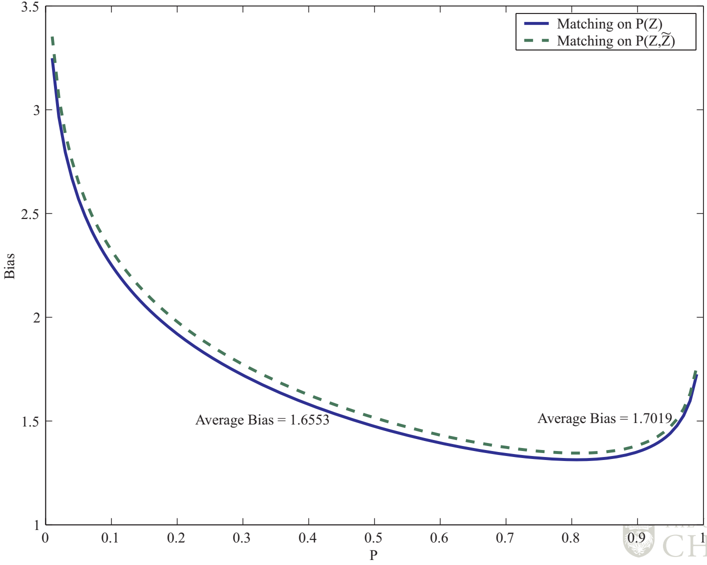
Problem: Prove this

Figure 4.--Bias for Treatment on the Treated
Special case: Adding irrelevant information $\widetilde{Z}$ increases the bias
correlation($\widetilde{Z}$,f2)=0.5

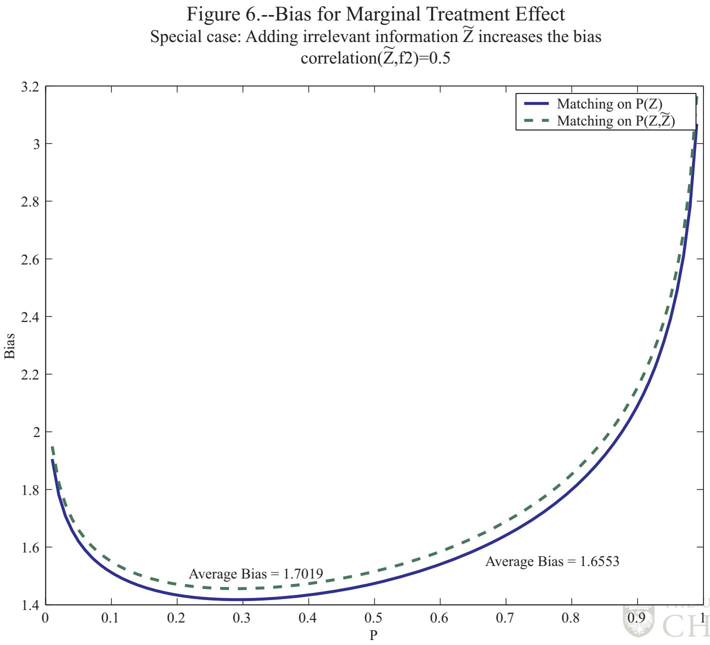


THE UNIVERSITY OF CHICAGO

Figure 5.--Bias for Average Treatment Effect
Special case: Adding irrelevant information $\widetilde{Z}$ increases the bias
correlation($\widetilde{Z}$,f2)=0.5

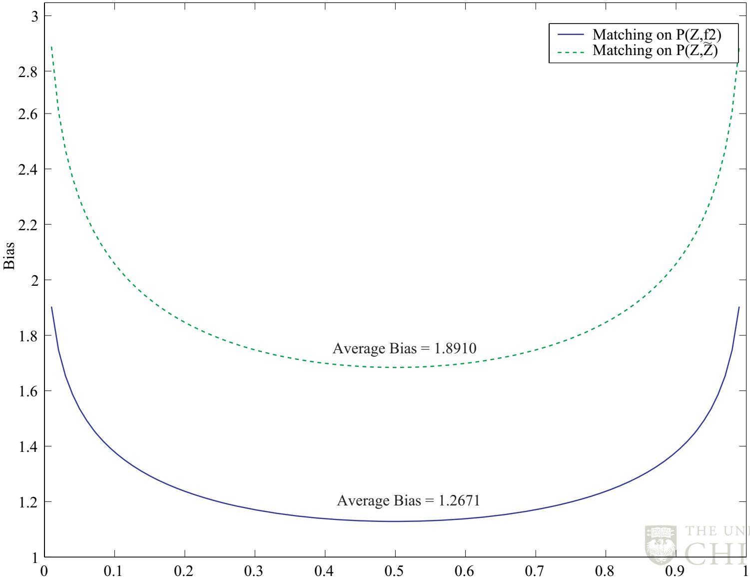


THE UNIVERSITY OF CHICAGO

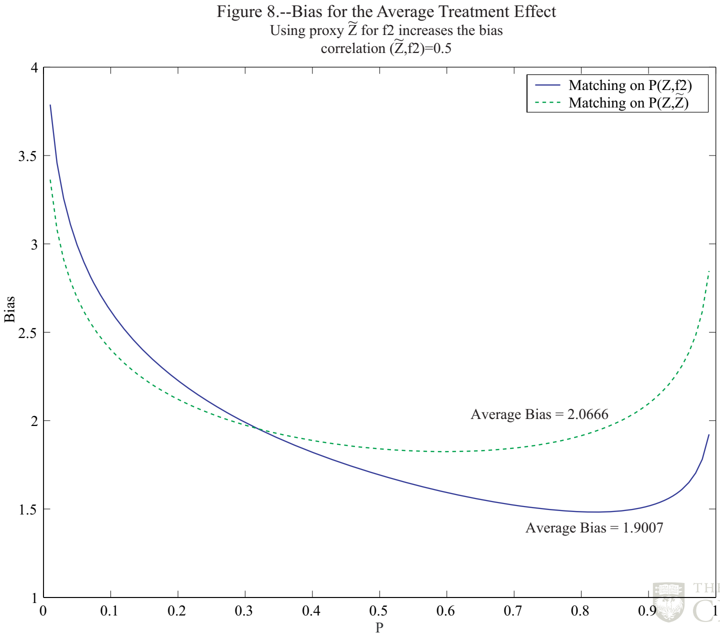


THE UNIVERSITY OF CHICAGO

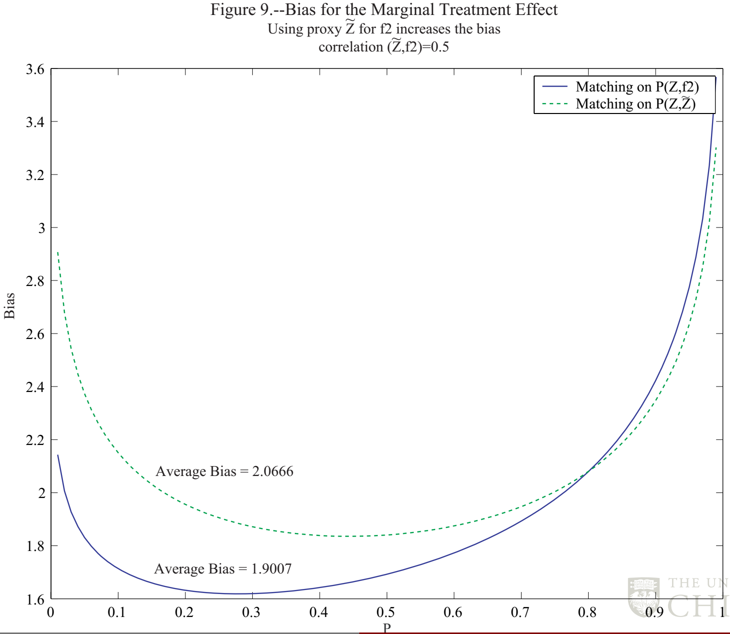
Figure 7.--Bias for Treatment on the Treated
Using proxy $\widetilde{Z}$ for f2 increases the bias
correlation $(\widetilde{Z}, f2)=0.5$



THE UNIVERSITY OF CHICAGO



THE UNIVERSITY OF CHICAGO



THE UNIVERSITY OF CHICAGO

- Adding more variables to the Information Set may increase Bias.
- How to choose the relevant W variables?
- Standard methods on model selection criteria fail.
- An implicit assumption underlying such procedures is that the added conditioning variables C are exogenous in the following sense

$$(Y_0, Y_1) \perp\!\!\!\perp D | I_E, C \tag{M-4}$$

(I_E is the list of initial variables used as conditioning variables.)

- Sometimes procedures suggested "Add variables when t ratios big in propensity score"
- Improve Fit
- Such procedures can raise the bias.

Consider the following example:

$$\widetilde{\widetilde{I}}_E = \{Z, S\}$$

where

$$\begin{aligned} & S = V - Z\gamma + \eta \\ & \eta \sim N\left(0, \sigma_\eta^2\right) \\ & \eta \perp\!\!\!\perp \left(f_1, f_2, \varepsilon_0, \varepsilon_1, \varepsilon_V\right). \end{aligned}$$

S might be an elicitation from a questionaire.

Same expressions for the biases using $\tilde{\tilde{\beta}}_j$ $(j = 0, 1)$ instead of β_j where:

$$\tilde{\tilde{\beta}}_1 = \frac{\left(\alpha_{11}\alpha_{V1}\sigma^2_{f_1} + \alpha_{12}\alpha_{V2}\sigma^2_{f_2}\right)}{\sqrt{\alpha^2_{V1}\sigma^2_{f_1} + \alpha^2_{V2}\sigma^2_{f_2} + \sigma^2_{\varepsilon_V}} + \sigma^2_{\eta}}$$

$$\tilde{\tilde{\beta}}_0 = \frac{\left(\alpha_{01}\alpha_{V1}\sigma^2_{f_1} + \alpha_{02}\alpha_{V2}\sigma^2_{f_2}\right)}{\sqrt{\alpha^2_{V1}\sigma^2_{f_1} + \alpha^2_{V2}\sigma^2_{f_2} + \sigma^2_{\varepsilon_V}} + \sigma^2_{\eta}}$$

In general, these expressions are not zero. Bias can be increased *or* decreased.

Problem: Derive these conditions

When $\sigma_\eta^2 = 0$ we can perfectly predict D. (Will pass a goodness-of-fit criterion)
Thus for

$$2\varepsilon > \sigma_\eta^2 > \varepsilon > 0$$

It follows that

$$\begin{aligned}
\lim_{\varepsilon \to 0} \Pr\left(D = 1 | V - Z\gamma + \eta\right) &= 1 \text{ for } V > Z\gamma \\
\lim_{\varepsilon \to 0} \Pr\left(D = 1 | V - Z\gamma + \eta\right) &= 0 \text{ for } V < Z\gamma.
\end{aligned}$$

Assumption (M-2) is violated and matching breaks down
Making σ_η^2 arbitrarily small, we can predict D arbitrarily well.
Can improve over the fit with $(f_{1,} f_2)$ in the set which produces no bias.

Table 4

	Goodness of fit statistics		Average Bias		
Variables in Probit	Correct in-sample prediction rate	Pseudo R^2	TT	ATE	MTE
Z	66.88%	0.1284	1.1380	1.6553	1.6553
Z, f_2	75.02%	0.2791	1.2671	1.9007	1.9007
Z, f_1, f_2	83.45%	0.4844	0.0000	0.0000	0.0000
Z, S_1	77.38%	0.3282	0.9612	1.3981	1.4070
Z, S_2	92.25%	0.7498	0.9997	1.4541	1.4590

Model:

$$\begin{array}{llll} V = Z + f_1 + f_2 + \varepsilon_v & \varepsilon_v \sim N(0,1) & S_1 = V + u_1 & u_1 \sim N(0,4) \\ Y_1 = 2f_1 + 0.1f_2 + \varepsilon_1 & \varepsilon_1 \sim N(0,1) & S_2 = V + u_2 & u_2 \sim N(0,0.25) \\ Y_0 = f_1 + 0.1f_2 + \varepsilon_0 & \varepsilon_0 \sim N(0,1) & & \\ & f_1 \sim N(0,1) & & \\ & f_2 \sim N(0,1) & & \end{array}$$

A More General Example

Considers use of a proxy regressor

$$Q = \alpha_{Q_Z} Z + \alpha_{Q_1} f_1 + \alpha_{Q_2} f_2 + \tau + \eta$$

- $Z \perp\!\!\!\perp (f_1, f_2, \tau, \eta)$;
- (f_1, f_2, τ, η) has mean zero
- $f_1 \perp\!\!\!\perp f_2$, $\tau \perp\!\!\!\perp \eta$ and $(f_1, f_2) \perp\!\!\!\perp (\tau, \eta)$;
- τ possibly dependent on ε_V in the latent variable generating the treatment choice
- η is measurement error.
- For different levels of dependence between τ and ε_V, and different weights on Z, f_1, f_2 and on the scale of measurement error, Q can be a better predictor of D than f_1, f_2 or even f_1, f_2, Z.

- However, in general, $(Y_1, Y_0) \not\perp\!\!\!\perp D \mid Q$ because Q is an imperfect proxy for the combinations of f_1 and f_2 entering Y_1 and Y_0.
- Thus conditioning on Q can produce a better fit for D but greater bias for the treatment parameters.

Consider the following example where Y is an outcome and I is an index

$$\begin{aligned} Y &= \theta + \varepsilon_Y \\ I &= \theta + \varepsilon_I \end{aligned}$$

where

$$\theta \perp\!\!\!\perp (\varepsilon_Y, \varepsilon_I),$$
$$\varepsilon_Y \perp\!\!\!\perp \varepsilon_I.$$

Obviously $Y \perp\!\!\!\perp I \mid \theta$.

- Suppose instead that we have a candidate conditioning variable $Q = \alpha_\theta \theta + \eta + \tau$.
- Suppose that all variables are normal with zero mean and are mutually independent.
- Then we may write

$$I = \pi_I Q + \varepsilon_Q$$

where

$$\pi_I = \frac{\alpha_\theta \sigma_\theta^2 + \sigma_{\tau,\varepsilon_I}}{\alpha_\theta^2 \sigma_\theta^2 + \sigma_\eta^2 + \sigma_\tau^2}.$$

- It is assumed that η is independent of all other error components on the right hand sides of the equations for Q, I and y.
- From normal regression theory we know that conditioning is equivalent to residualizing.
- Constructing the residuals we obtain

$$I - \pi_I Q = \theta\left(1 - \alpha_\theta \pi_I\right) + \varepsilon_I - \pi_I\left(\eta + \tau\right).$$

- By a parallel argument

$$Y - \pi_Y Q = \theta\left(1 - \alpha_\theta \pi_Y\right) + \varepsilon_Y - \pi_Y\left(\eta + \tau\right)$$

$Y \perp\!\!\!\perp I \mid Q$ requires that $I - \pi_I Q$ and $Y - \pi_Y Q$ be uncorrelated, which in general does not happen.

Conclusion

- Letting the dependence between τ and ε_I get large, and setting α_θ to suitable values, we can predict I better (in the sense of R^2) with Q than with θ.
- Letting $D = 1(I > 0)$ produces a simple version of the example because better prediction of I produces better prediction of D.

Comparable to $\beta_1$and β_0 above, we can define

$$\beta_1' = \frac{\alpha_{V1}\alpha_{11}\sigma_{f_1}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \sigma_{\varepsilon_V}^2}}$$

$$\beta_0' = \frac{\alpha_{V1}\alpha_{01}\sigma_{f_1}^2}{\sqrt{\alpha_{V1}^2\sigma_{f_1}^2 + \sigma_{\varepsilon_V}^2}}.$$

Condition 1 *The bias produced by using matching to estimate TT is smaller in absolute value for any given p when the new information set* $\sigma\left(I_E'\right)$ *is used if*

$$|\beta_0| > |\beta_0'| .$$

Condition 2 *The bias produced by using matching to estimate ATE is smaller in absolute value for any given p when the new information set* $\sigma\left(I_E'\right)$ *is used if*

$$|\beta_1(1-p) + \beta_0 p| > |\beta_1'(1-p) + \beta_0' p| .$$

Condition 3 *The bias produced by using matching to estimate MTE is smaller in absolute value for any given p when the new information set $\sigma\left(I'_E\right)$ is used if*

$$\begin{aligned} & \left| M(p)\left[\beta_1\left(1-p\right)+\beta_0 p\right]-\Phi^{-1}\left(1-p\right)\left[\beta_1-\beta_0\right]\right| \\ > & \left| M(p)\left[\beta'_1\left(1-p\right)+\beta'_0 p\right]-\Phi^{-1}\left(1-p\right)\left[\beta'_1-\beta'_0\right]\right| \end{aligned}$$

Proof of Condition 1

Suppose

$$\beta_0 = \underbrace{\frac{\alpha_{V1}\alpha_{01}\sigma^2_{f_1} + (\alpha^2_{V2})\left(\frac{\alpha_{02}}{\alpha_{V2}}\right)\sigma^2_{f_2}}{\sqrt{\alpha^2_{V1}\sigma^2_{f_1} + \alpha^2_{V2}\sigma^2_{f_2} + \sigma^2_{\varepsilon_V}}}}_{\text{when } f_2 \text{ is in information set}} > \underbrace{\frac{\alpha_{V1}\alpha_{01}\sigma^2_{f_1}}{\sqrt{\alpha^2_{V1}\sigma^2_{f_1} + \sigma^2_{\varepsilon_V}}}}_{\substack{\text{when } f_2 \text{ is not} \\ \text{in information set}}} = \beta_0'.$$

When $\left(\frac{\alpha_{02}}{\alpha_{V2}}\right) = 0$, $\beta_0 < \beta_0'$.

$$\frac{\partial \beta_0}{\partial\left(\frac{\alpha_{02}}{\alpha_{V2}}\right)} = \frac{\alpha^2_{V2}\sigma^2_{f_2}}{\sqrt{\alpha^2_{V1}\sigma^2_{f_1} + \alpha^2_{V2}\sigma^2_{f_2} + \sigma^2_{\varepsilon_V}}} > 0.$$

There is some critical value α_{02}^{*} beyond which $\beta_0 > \beta_0'$

Assume

$$\begin{aligned} \alpha_{01} &= \alpha_{V_1} = \alpha_{V_2} = 1 \\ \alpha_{02} &= \alpha_{12} = 1 \\ \alpha_{11} &= 2 \end{aligned}$$

Return to Text

$$
\begin{aligned}
Y_j^* &= \mu_j + U_j \\
U_j &= \alpha_{j1} f_1 + \alpha_{j2} f_2 + \varepsilon_j \ , \ j = 0, 1 \\
Y_j &= 1 \text{ if } Y_j^* \geq 0, \quad = 0 \text{ otherwise,}
\end{aligned}
$$

People receive treatment according to the rule

$$
\begin{aligned}
V &= \mu_V + U_V \\
U_V &= \alpha_{V_1} f_1 + \alpha_{V_2} f_2 + \varepsilon_V \\
D &= 1 \text{ if } V \geq 0, \quad = 0 \text{ otherwise;}
\end{aligned}
$$

$$
f_1 \perp\!\!\!\perp f_2, \ \varepsilon_0 \perp\!\!\!\perp \varepsilon_1 \perp\!\!\!\perp \varepsilon_V, \qquad (f_1, f_2) \perp\!\!\!\perp (\varepsilon_0, \varepsilon_1, \varepsilon_V)
$$

The effect of treatment is given by:

$$\Delta_1 (I_E) = \frac{\Pr(Y_1 = 1, D = 1 | I_E)}{\Pr(Y_0 = 1, D = 1 | I_E)}.$$

A second definition works with odds ratios:

$$\Delta_2 (I_E) = \frac{\frac{\Pr(Y_1 = 1, D = 1 | I_E)}{\Pr(Y_1 = 0, D = 1 | I_E)}}{\frac{\Pr(Y_0 = 1, D = 1 | I_E)}{\Pr(Y_0 = 0, D = 1 | I_E)}}$$

One could also work with $\log \Delta$. Under the null hypothesis of no effect of treatment $\Delta_1 = \Delta_2 = 1$.

$$\widehat{\Delta}_1 (I_E) = \frac{\Pr(Y_1 = 1, D = 1 | I_E)}{\Pr(Y_0 = 1, D = 0 | I_E)}.$$

The denominator replaces the desired probability

$$\Pr(Y_0 = 1, D = 1 | I_E)$$

by

$$\Pr(Y_0 = 1, D = 0 | I_E).$$

Under the Null Hypothesis of no "real" effect of treatment

$$\mu_1 = \mu_0 = \mu$$
$$F_{U_1} = F_{U_0} = F_U$$

can be generated by

$$\alpha_{11} = \alpha_{01} = \alpha_1$$
$$\alpha_{12} = \alpha_{02} = \alpha_2$$
$$F_{\varepsilon_1} = F_{\varepsilon_0} = F_{\varepsilon}$$

Assume initially that

$$I_E = \{f_1, f_2\}.$$

$$\widehat{\Delta}_1(I_E) = \frac{\Pr(Y_1 = 1, D = 1 | f_1, f_2)}{\Pr(Y_0 = 1, D = 0 | f_1, f_2)} = \frac{\Pr(Y_1 = 1 | f_1, f_2)}{\Pr(Y_0 = 1 | f_1, f_2)} = \Delta_1(I_E).$$

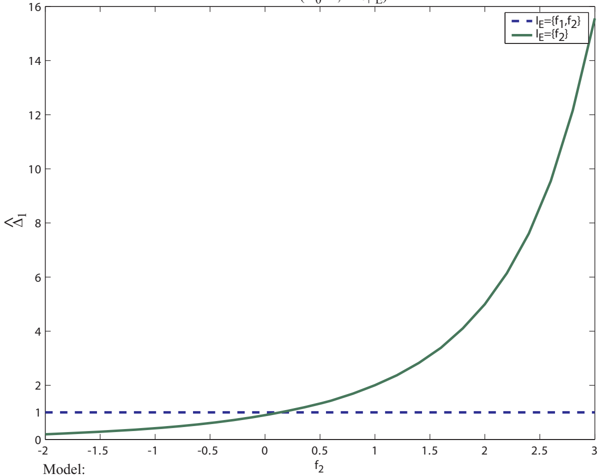
In general:

$$\Delta_1(I_E) = \frac{\Pr(Y_1 = 1, D = 1 | I_E)}{\Pr(Y_0 = 1, D = 1 | I_E)} \neq \frac{\Pr(Y_1 = 1, D = 1 | I_E)}{\Pr(Y_0 = 1, D = 0 | I_E)} = \widehat{\Delta}_1(I_E)$$

Figure 10.--Estimated Effect of Treatment under Different Information Sets

No Effect of Treatment and $\alpha_{v2}=1$

$$\hat{\Delta}_1 = \frac{\Pr(Y_1=1, D=1 | I_E)}{\Pr(Y_0=1, D=0 | I_E)}$$



Model:

$V = -1+f_1+f_2+\varepsilon_v$ $\quad \varepsilon_v \sim N(0,1)$

$Y^*_1 = -1+f_1+f_2+\varepsilon_1$ $\quad \varepsilon_1 \sim N(0,1)$

$Y^*_0 = -1+f_1+f_2+\varepsilon_0$ $\quad \varepsilon_0 \sim N(0,1)$

$Y_1 = 1(Y^*_1>0)$ $\quad f_1 \sim N(0,1)$

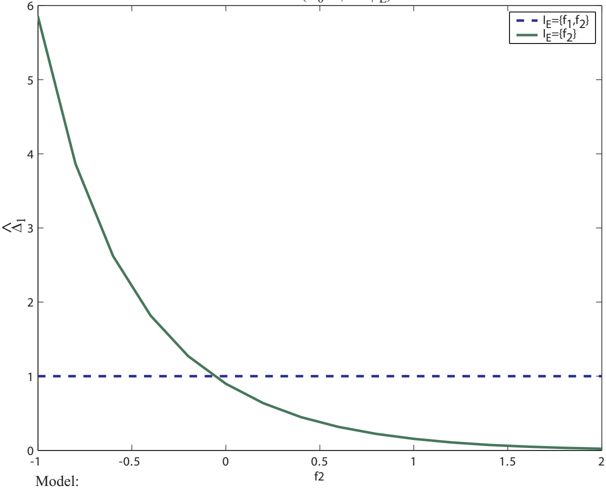
$Y_0 = 1(Y^*_0>0)$ $\quad f_2 \sim N(0,1)$

THE UNIVERSITY OF CHICAGO

Figure 11.--Estimated Effect of Treatment under Different Information Sets

No Effect of Treatment and $\alpha_{V2}=-1$

$$\hat{\Delta}_1 = \frac{Pr(Y_1=1,D=1|I_E)}{Pr(Y_0=1,D=0|I_E)}$$



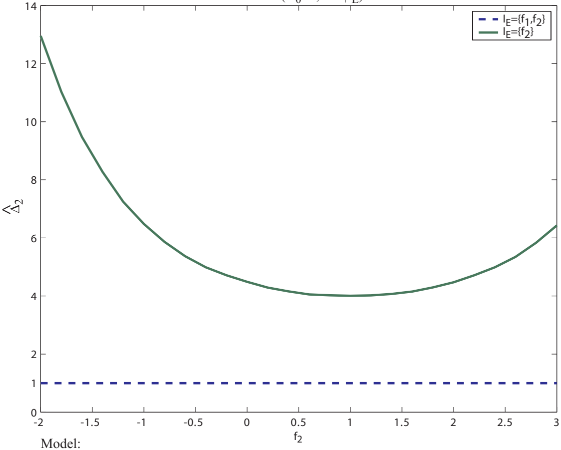
Model:

$V=-1+f_1-f_2+\varepsilon_V$	$\varepsilon_V \sim N(0,1)$
$Y^*_1=-1+f_1+f_2+\varepsilon_1$	$\varepsilon_1 \sim N(0,1)$
$Y^*_0=-1+f_1+f_2+\varepsilon_0$	$\varepsilon_0 \sim N(0,1)$
$Y_1=1(Y^*_1>0)$	$f_1 \sim N(0,1)$

Figure 12.--Estimated Effect of Treatment under Different Information Sets

No Effect of Treatment and $\alpha_{V2}=1$

$$\hat{\Delta}_2=\frac{\dfrac{\Pr(Y_1=1,D=1|I_E)}{\Pr(Y_1=0,D=1|I_E)}}{\dfrac{\Pr(Y_0=1,D=0|I_E)}{\Pr(Y_0=0,D=0|I_E)}}$$



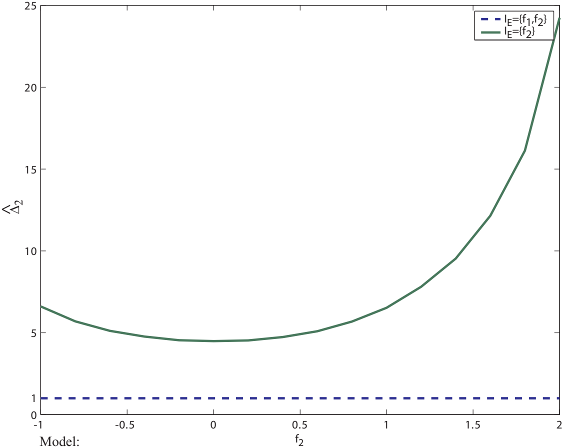
Model:

$V= -1+f_1+f_2+\varepsilon_V$	$\varepsilon_V\sim N(0,1)$
$Y^*_1=-1+f_1+f_2+\varepsilon_1$	$\varepsilon_1\sim N(0,1)$
$Y^*_0=-1+f_1+f_2+\varepsilon_0$	$\varepsilon_0\sim N(0,1)$
$Y_1=1(Y^*_1>0)$	$f_1\sim N(0,1)$
$Y_0=1(Y^*_0>0)$	$f_2\sim N(0,1)$

Figure 13.--Estimated Effect of Treatment under Different Information Sets

No Effect of Treatment and α_{V2}=-1

$$\hat{\Delta}_2 = \frac{\dfrac{\Pr(Y_1=1,D=1|I_E)}{\Pr(Y_1=0,D=1|I_E)}}{\dfrac{\Pr(Y_0=1,D=0|I_E)}{\Pr(Y_0=0,D=0|I_E)}}$$



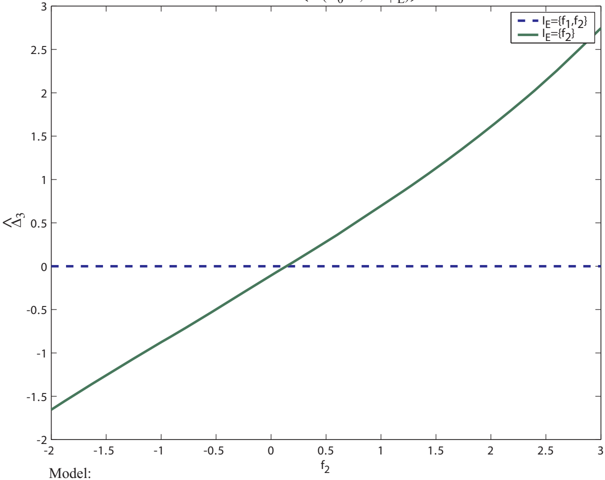
Model:

$V= -1+f_1-f_2+\varepsilon_V$	$\varepsilon_V \sim N(0,1)$
$Y^*_1=-1+f_1+f_2+\varepsilon_1$	$\varepsilon_1 \sim N(0,1)$
$Y^*_0=-1+f_1+f_2+\varepsilon_0$	$\varepsilon_0 \sim N(0,1)$
$Y_1=1(Y^*_1>0)$	$f_1 \sim N(0,1)$
$Y_0=1(Y^*_0>0)$	$f_2 \sim N(0,1)$

Figure 14.--Estimated Effects of Treatment under Different Information Sets

No Effect of Treatment and $\alpha_{v2}=1$

$$\hat{\Delta}_3 = \text{Log}\left(\frac{\Pr(Y_1=1, D=1 | I_E)}{\Pr(Y_0=1, D=0 | I_E)}\right)$$



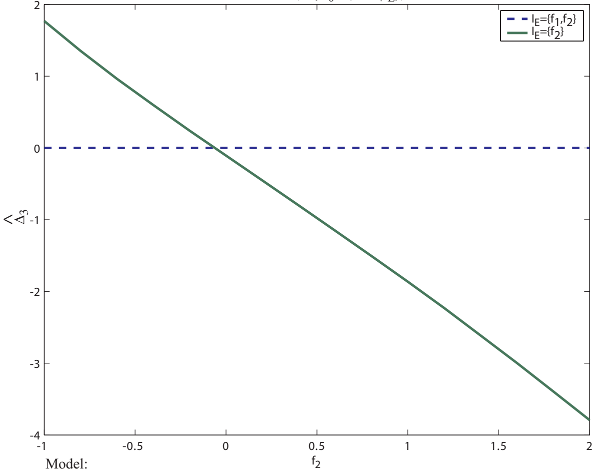
Model:

$V= -1+f_1+f_2+\varepsilon_v$	$\varepsilon_v \sim N(0,1)$
$Y^*_1=-1+f_1+f_2+\varepsilon_1$	$\varepsilon_1 \sim N(0,1)$
$Y^*_0=-1+f_1+f_2+\varepsilon_0$	$\varepsilon_0 \sim N(0,1)$
$Y_1=1(Y^*_1>0)$	$f_1 \sim N(0,1)$
$Y_0=1(Y^*_0>0)$	$f_2 \sim N(0,1)$

Figure 15.--Estimated Effects of Treatment under Different Information Sets

No Effect of Treatment and α_{V2}=-1

$$\hat{\Delta}_3 = \text{Log}\left(\frac{\Pr(Y_1=1,D=1|I_E)}{\Pr(Y_0=1,D=0|I_E)}\right)$$



Model:

$V = -1 + f_1 - f_2 + \varepsilon_V$ $\quad \varepsilon_V \sim N(0,1)$

$Y^*_1 = -1 + f_1 + f_2 + \varepsilon_1$ $\quad \varepsilon_1 \sim N(0,1)$

$Y^*_0 = -1 + f_1 + f_2 + \varepsilon_0$ $\quad \varepsilon_0 \sim N(0,1)$

$Y_1 = 1(Y^*_1 > 0)$ $\quad f_1 \sim N(0,1)$

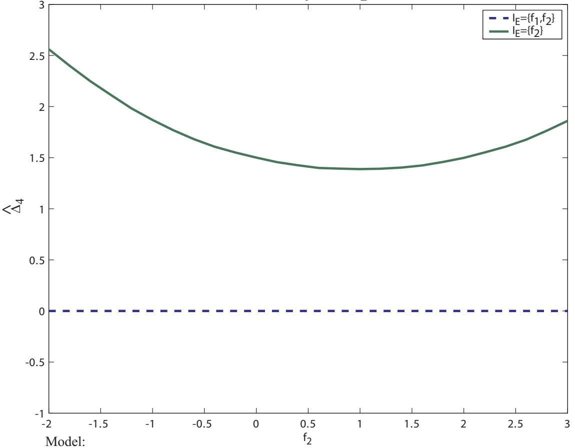
$Y_0 = 1(Y^*_0 > 0)$ $\quad f_2 \sim N(0,1)$

$D = 1(V > 0)$

Figure 16.--Estimated Effect of Treatment under Different Infomation Sets

No Effect of Treatment and $\alpha_{v2}=1$

$$\hat{\Delta}_4 = \text{Log}\left(\frac{\frac{\Pr(Y_1=1,D=1|I_E)}{\Pr(Y_1=0,D=1|I_E)}}{\frac{\Pr(Y_0=1,D=0|I_E)}{\Pr(Y_0=0,D=0|I_E)}} \right)$$



Model:

$V = -1+f_1+f_2+\varepsilon_V$ $\quad \varepsilon_V \sim N(0,1)$

$Y^*_1 = -1+f_1+f_2+\varepsilon_1$ $\quad \varepsilon_1 \sim N(0,1)$

$Y^*_0 = -1+f_1+f_2+\varepsilon_0$ $\quad \varepsilon_0 \sim N(0,1)$

$Y_1 = 1(Y^*_1>0)$ $\quad f_1 \sim N(0,1)$

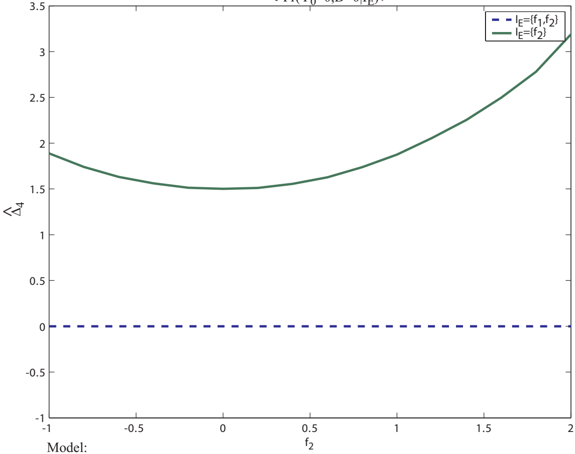
$Y_0 = 1(Y^*_0>0)$ $\quad f_2 \sim N(0,1)$

$D = 1(V>0)$

Figure 17.--Estimated Effects of Treatment under Different Information Sets

No Effect of Treatment and $\alpha_{v2}=-1$

$$\hat{\Delta}_4=\text{Log}\left(\frac{\frac{\Pr(Y_1=1,D=1|I_E)}{\Pr(Y_1=0,D=1|I_E)}}{\frac{\Pr(Y_0=1,D=0|I_E)}{\Pr(Y_0=0,D=0|I_E)}}\right)$$



Model:

$V= -1+f_1-f_2+\varepsilon_v$ $\quad\varepsilon_v\sim N(0,1)$

$Y^*_1=-1+f_1+f_2+\varepsilon_1$ $\quad\varepsilon_1\sim N(0,1)$

$Y^*_0=-1+f_1+f_2+\varepsilon_0$ $\quad\varepsilon_0\sim N(0,1)$

$Y_1=1(Y^*_1>0)$ $\quad f_1\sim N(0,1)$

$Y_0=1(Y^*_0>0)$ $\quad f_2\sim N(0,1)$

THE UNIVERSITY OF CHICAGO

Consider a general model of the form:

$$\begin{aligned} Y_1 &= \mu_1 + U_1 \\ Y_0 &= \mu_0 + U_0 \\ V &= \mu_V(Z) + U_V \\ D &= 1 \text{ if } V \geq 0, \quad = 0 \text{ otherwise} \\ Y &= DY_1 + (1-D)Y_0. \end{aligned}$$

where

$$\begin{aligned} (U_1, U_0, U_V)' &\sim N(0, \Sigma) \\ var(U_i) &= \sigma_i^2 \\ cov(U_i, U_j) &= \sigma_{ij} \\ i = 0; j = 1 & \\ cov(U_1, V) &= \sigma_{1V} \\ cov(U_0, V) &= \sigma_{0V} \end{aligned}$$

Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the pdf and the cdf of a standard normal random variable. Then, the propensity score for this model is given by:

$$\begin{aligned}\Pr(V > 0 | \mu_V(Z)) &= P(\mu_V(Z)) = \Pr(U_V > -\mu_V(Z)) = p \\ &= 1 - \Phi\left(\frac{-\mu_V(Z)}{\sigma_V}\right) = p\end{aligned}$$

so

$$\frac{-\mu_V(Z)}{\sigma_V} = \Phi^{-1}(1-p).$$

Since the event $\left(V \gtreqqless 0, P\left(\mu_V\left(Z\right)\right) = p \right)$ can be written as

$$\frac{U_V}{\sigma_V} \lesseqqgtr -\frac{\mu_V\left(Z\right)}{\sigma_V}$$
$$\frac{U_V}{\sigma_V} \lesseqqgtr \Phi^{-1}\left(1 - p\right)$$

we can write the conditional expectations required to get the biases as a function of p.

For U_1 :

$$E\left(U_1|V>0,P\left(\mu_V\left(Z\right)\right)=p\right)$$
$$=\frac{\sigma_{1V}}{\sigma_V}E\left(\frac{U_V}{\sigma_V}|\frac{U_V}{\sigma_V}>\frac{-\mu_V\left(Z\right)}{\sigma_V},P\left(\mu_V\left(Z\right)\right)=p\right)$$
$$=\frac{\sigma_{1V}}{\sigma_V}E\left(\frac{U_V}{\sigma_V}|\frac{U_V}{\sigma_V}>\Phi^{-1}\left(1-p\right)\right)$$
$$=\beta_1 M_1(p)$$

$$E\left(U_1|V=0,P\left(\mu_V\left(Z\right)\right)=p\right)$$
$$=\frac{\sigma_{1V}}{\sigma_V}E\left(\frac{U_V}{\sigma_V}|\frac{U_V}{\sigma_V}=\frac{-\mu_V\left(Z\right)}{\sigma_V},P\left(\mu_V\left(Z\right)\right)=p\right)$$
$$=\frac{\sigma_{1V}}{\sigma_V}E\left(\frac{U_V}{\sigma_V}|\frac{U_V}{\sigma_V}=\Phi^{-1}\left(1-p\right),P\left(\mu_V\left(Z\right)\right)=p\right)$$
$$=\beta_1\Phi^{-1}(1-p)$$

THE UNIVERSITY OF CHICAGO

Where

$$\beta_1 = \frac{\sigma_{1V}}{\sigma_V}$$

Similarly for U_0 :

$$\begin{aligned} E\left(U_0 | V > 0, P\left(\mu_V\right) = p\right) &= \beta_0 M_1(p) \\ E\left(U_0 | V < 0, P\left(\mu_V\right) = p\right) &= \beta_0 M_0(p) \\ E\left(U_0 | V = 0, P\left(\mu_V\right) = p\right) &= \beta_0 \Phi^{-1}\left(1 - p\right) \end{aligned}$$

Where

$$\beta_0 = \frac{\sigma_{0V}}{\sigma_V}.$$

and

$$\begin{aligned} M_1(p) &= \frac{\phi\left(\Phi^{-1}(1-p)\right)}{p} \\ M_0(p) &= -\frac{\phi\left(\Phi^{-1}(1-p)\right)}{(1-p)} \end{aligned}$$

are inverse Mills ratio terms.

Substituting these into the expressions for the biases

$$\begin{aligned} \text{Bias } TT(p) &= \beta_0 M_1(p) - \beta_0 M_0(p) \\ &= \beta_0 M(p) \end{aligned}$$

$$\begin{aligned} \text{Bias } ATE(p) &= \beta_1 M_1(p) - \beta_0 M_0(p) \\ &= M(p)\left(\beta_1(1-p) + \beta_0 p\right) \end{aligned}$$

$$\begin{aligned} \text{Bias } MTE &= \beta_1 M_1(p) - \beta_0 M_0(p) \\ &\quad -\beta_1 \Phi^{-1}(1-p) + \beta_0 \Phi^{-1}(1-p) \\ &= M(p)\left(\beta_1(1-p) + \beta_0 p\right) \\ &\quad -\Phi^{-1}(1-p)\left[\beta_1 - \beta_0\right]. \end{aligned}$$

where

$$M(p) = M_1(p) - M_0(p) = \frac{\phi(\Phi^{-1}(1-p))}{p(1-p)}$$