Microeconometrics II — Problem Set 02 — MTE

Vinícius de Almeida Nery Ferreira

October 24th, 2024

1 ATU as a Weighted Average of the MTE Function

1.1 Model

We first need to define our Generalized Roy Model. Let

- $Y(0)$ and $Y(1)$ be the potential outcomes;
- $Z \in \mathbf{Z}$ be the instrument;
- $V \in [0,1]$ be an unobservable variable (to the econometrician) representing latent heterogeneity, the "individual cost" of treatment or the selection process;
- $P : \mathbf{Z} \to \mathbb{R}$ be a function that denotes the propensity score of treatment;
- $D = \mathbb{1}\{P(Z) \geq V\}$ be the treatment indicator variable;
- $Y = D \cdot Y(1) + (1 - D) \cdot Y(0)$ be the observed outcome.

As in the lecture notes, we make the following assumptions:

Assumption 1 (IV Independence): $(Y(0), Y(1), V) \perp Z$

Assumption 2 (IV Strength): $P(Z)$ is a continuous random variable.

Assumption 3: V is a continuous random variable.

Note that we can normalize so that $V \sim U(0,1)$ by defining $V \coloneqq F_{\tilde{V}}\left(\tilde{V}\right)$. Similarly, $P(Z) \coloneqq F_{\tilde{V}}\left(\rho(Z)\right)$. Just as before, we then have that $D = \mathbb{1}\{P(Z) \geq V\}$, by applying $F_{\tilde{V}}(\cdot)$ (which is increasing) and renaming some variables.

Since $V \sim U(0,1)$, note that

$$
\begin{aligned}
\mathbb{P}\left(D = 1 \mid Z = z\right) &= \mathbb{P}\left(V \leq P(z) \mid Z = z\right) \\
&= \mathbb{P}\left(V \leq P(z)\right) && \text{(Assumption 1)} \\
&= P(z) && (V \sim U(0,1))
\end{aligned}
$$

Assumption 4 (Finite Means): $\mathbb{E}[|Y(1)|] < +\infty \quad \text{and} \quad \mathbb{E}[|Y(0)|] < +\infty$.

Assumption 5 (Common Support): $\mathbb{P}[D = 1] \in (0,1)$.

Finally, define the Marginal Treatment Effect (MTE) as

$$MTE : [0,1] \to \mathbb{R} \quad \text{such that} \quad MTE(v) = \mathbb{E}\left[Y(1) - Y(0) \mid V = v\right] \tag{MTE}$$

Note that, the smaller the v, the more likely the individuals self-select into treatment. In this sense, the MTE being bigger for higher values of v indicates that the treatment effect is larger for individuals less likely to take it.

1.2 Aggregation of the MTE into the ATU

Claim: The average treatment effect on the untreated (ATU) can be expressed as a weighted integral of the MTE:

$$ATU := \mathbb{E}\left[Y(1) - Y(0) \mid D = 0\right] = \int_0^1 \mathbb{E}\left[Y(1) - Y(0) \mid V = v\right] \cdot \omega(v)\, dv \quad \text{where} \quad \omega(v) = \frac{\int_0^v f_P(p)\, dp}{\mathbb{E}\left[1 - P(Z)\right]}$$

Proof.

$$\begin{aligned}
ATU &= \mathbb{E}[Y(1) - Y(0) \mid D = 0] \\
&= \frac{\mathbb{E}[(1-D) \cdot (Y(1) - Y(0))]}{\mathbb{E}[1-D]} && \text{(Definition of } \mathbb{E}[\cdot \mid D=0] = \mathbb{E}[\cdot \mid (1-D)=1]) \\
&= \frac{\mathbb{E}[(1-D) \cdot (Y(1) - Y(0))]}{1 - \mathbb{E}[\mathbb{E}[D \mid Z]]} && \text{(Linearity and LIE on } Z) \\
&= \frac{\mathbb{E}[(1-D) \cdot (Y(1) - Y(0))]}{1 - \mathbb{E}[P(Z)]} && (\mathbb{E}[D \mid Z] = \mathbb{P}(D=1 \mid Z) = P(z) \text{ by A1, A3}) \\
&= \frac{\mathbb{E}[(1 - \mathbb{1}\{P(Z) \geq V\}) \cdot (Y(1) - Y(0))]}{1 - \mathbb{E}[P(Z)]} && \text{(Definition of } D \text{ and A3)} \\
&= \frac{\mathbb{E}_{Z,V}\left\{\mathbb{E}[(1 - \mathbb{1}\{P(Z) \geq V\}) \cdot (Y(1) - Y(0)) \mid Z, V]\right\}}{1 - \mathbb{E}[P(Z)]} && \text{(LIE on } Z \text{ and } V) \\
&= \frac{\mathbb{E}_{Z,V}\left\{(1 - \mathbb{1}\{P(Z) \geq V\}) \cdot \mathbb{E}[(Y(1) - Y(0)) \mid Z, V]\right\}}{1 - \mathbb{E}[P(Z)]} && \text{(Conditioning on } Z \text{ and } V) \\
&= \frac{\mathbb{E}_{Z,V}\left\{(1 - \mathbb{1}\{P(Z) \geq V\}) \cdot \mathbb{E}[(Y(1) - Y(0)) \mid V = v]\right\}}{1 - \mathbb{E}[P(Z)]} && \text{(A1: } (Y(0), Y(1), V) \perp Z) \\
&= \frac{\mathbb{E}_{Z,V}\left\{(1 - \mathbb{1}\{P(Z) \geq V\}) \cdot MTE(v)\right\}}{1 - \mathbb{E}[P(Z)]} && \text{(Definition of } MTE) \\
&= \frac{\int_0^1 \int_0^1 (1 - \mathbb{1}\{p \geq v\}) \cdot MTE(v) \cdot f_{P,V}(p,v) dp dv}{1 - \mathbb{E}[P(Z)]} && \text{(Definition of } \mathbb{E}_{Z,V}[\cdot] \text{ and A4)} \\
&= \frac{\int_0^1 \int_0^1 (1 - \mathbb{1}\{p \geq v\}) \cdot MTE(v) \cdot f_P(p) \cdot f_V(v) dp dv}{1 - \mathbb{E}[P(Z)]} && (f_{P,V}(p,v) = f_P(p) \cdot f_V(v) \text{ by A1}) \\
&= \frac{\int_0^1 \int_0^1 (1 - \mathbb{1}\{p \geq v\}) \cdot MTE(v) \cdot f_P(p) dp dv}{1 - \mathbb{E}[P(Z)]} && (f_V(v) = 1 \text{ by A3}) \\
&= \frac{\int_0^1 MTE(v) \left(\int_0^1 (1 - \mathbb{1}\{p \geq v\})\, f_P(p) dp\right) dv}{1 - \mathbb{E}[P(Z)]} && \text{(Rearranging)} \\
&= \int_0^1 MTE(v) \left(\frac{\int_0^v f_P(p) dp}{\mathbb{E}[1 - P(Z)]}\right) dv && \text{(Separating the integral at } v \text{ and using } \mathbb{1}\{\cdot\})
\end{aligned}$$

□

2 Estimating the MTE Function

Suppose the data generating process (DGP) is such that

- $V \sim U(a_V, b_V) \sim U(0,1)$;
- $Z \sim U(a_Z, b_Z) \sim U\left(0, \frac{1}{2}\right)$;
- $U \sim \mathcal{N}\left(0, \sigma_U^2\right) \sim \mathcal{N}(0,1)$
- $P(z) = \gamma \cdot Z = \frac{1}{b_Z} \cdot Z = 2 \cdot Z$;
- $D = \mathbb{1}\{P(Z) \geq V\}$;
- $Y(0) \sim \mathcal{N}\left(0, \sigma_0^2\right) \sim \mathcal{N}(0,1)$;
- $Y(1) = Y(0) + U + \beta_0 + \beta_1 \cdot V + \beta_2 \cdot V^2$, where $\beta_0 = 1$, $\beta_1 = 2$ and $\beta_2 = 3$.

We have a sample size of $N = 10,000$ and we will do $M = 1,000$ Monte Carlo simulations. Our evaluation grid for V is $\mathcal{V} = \{0.1, 0.2, \cdots, 0.9\}$. We are interested in the relative bias of the MTE estimator:

$$RB(v) = \frac{\widehat{MTE}(v) - MTE(v)}{MTE(v)}$$

2.1 Parametric Estimator

We first assume a parametric form for the MTE function:

$$MTE(v) = h(v; \beta)$$

where h is known up to the unknown parameters β. Given the nature of our DGP, it is very natural that the MTE function be

$$\boldsymbol{h(v;\beta) = \beta_0 + \beta_1 \cdot V + \beta_2 \cdot V^2}$$

Since $Y = (1-D) \cdot Y(0) + D \cdot Y(1) = Y(0) + D\left[Y(1) - Y(0)\right]$ and $D = \mathbb{1}\{P(Z) \geq V\}$, we can write

$$\begin{aligned}
\mathbb{E}[Y \mid P(Z) = p] &= \mathbb{E}[Y(0) + D\left(Y(1) - Y(0)\right) \mid P(Z) = p] \\
&= \mathbb{E}[Y(0) \mid P(Z) = p] + \mathbb{E}[\mathbb{1}\{P(Z) \geq V\}\left(Y(1) - Y(0)\right) \mid P(Z) = p]
\end{aligned}$$

By Assumption 1, $\mathbb{E}[Y(0) \mid P(Z) = p] = \mathbb{E}[Y(0)]$. For the second term, since we are conditioning on $P(Z) = p$,

$$\mathbb{E}[\mathbb{1}\{P(Z) \geq V\}\left(Y(1) - Y(0)\right) \mid P(Z) = p] = \mathbb{E}[\mathbb{1}\{p \geq V\}\left(Y(1) - Y(0)\right) \mid P(Z) = p]$$

and now we can apply Assumption 1 to drop the conditional on $P(Z) = p$. Then, by using the LIE with respect to V,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}\{p \geq V\}\left(Y(1) - Y(0)\right) \mid P(Z) = p] &= \mathbb{E}[\mathbb{1}\{p \geq V\}\left(Y(1) - Y(0)\right)] \\
&= \mathbb{E}_V\left\{\mathbb{E}[\mathbb{1}\{p \geq V\}\left(Y(1) - Y(0)\right) \mid V = v]\right\} \\
&= \mathbb{E}_V\left\{\mathbb{1}\{p \geq V\}\mathbb{E}[Y(1) - Y(0) \mid V = v]\right\} \\
&= \mathbb{E}_V\left\{\mathbb{1}\{p \geq V\}MTE(v)\right\}
\end{aligned}$$

By applying the definition of $\mathbb{E}$ and substituting $MTE(v) = h(v;\beta)$, we get

$$\begin{aligned}
\mathbb{E}_V\left\{\mathbb{1}\{p \geq V\} MTE(v)\right\} &= \int_0^1 \mathbb{1}\{p \geq V\} MTE(v) dv \\
&= \int_0^p \mathbb{1}\{p \geq V\} MTE(v) dv + \int_p^1 \mathbb{1}\{p \geq V\} MTE(v) dv \\
&= \int_0^p 1 \cdot MTE(v) dv + \int_p^1 0 \cdot MTE(v) dv \\
&= \int_0^p h(v;\beta) dv
\end{aligned}$$

Combining our previous results, we then have that

$$\begin{aligned}
\mathbb{E}[Y \mid P(Z) = p] &= \mathbb{E}[Y(0) + D\left(Y(1) - Y(0)\right) \mid P(Z) = p] \\
&= \mathbb{E}[Y(0) \mid P(Z) = p] + \mathbb{E}[\mathbb{1}\{P(Z) \geq V\}\left(Y(1) - Y(0)\right) \mid P(Z) = p] \\
&= \mathbb{E}[Y(0)] + \int_0^p h(v;\beta) dv
\end{aligned}$$

We can substitute our DGP and our proposed h to get

$$\begin{aligned}
\mathbb{E}[Y \mid P(Z) = p] &= \mathbb{E}[Y(0)] + \int_0^p h(v;\beta) dv \\
&= 0 + \int_0^p \beta_0 + \beta_1 \cdot v + \beta_2 \cdot v^2 dv \\
&= \left| \beta_0 \cdot v + \beta_1 \cdot \frac{v^2}{2} + \beta_2 \cdot \frac{v^3}{3} \right|_{v=0}^{p} \\
&= \boldsymbol{\beta_0 \cdot p + \beta_1 \cdot \frac{p^2}{2} + \beta_2 \cdot \frac{p^3}{3}},
\end{aligned}$$

which can be estimated using standard OLS. In computational terms, we just use the classic `lm` command, running a regression of Y on $P(Z)$, $\frac{P(Z)^2}{2}$ and $\frac{P(Z)^3}{3}$.

2.2 Nonparametric Estimator

We showed above that, in a general, nonparametric, case,

$$\mathbb{E}[Y \mid P(Z) = p] = \mathbb{E}[Y(0)] + \int_0^p MTE(v) dv$$

which implies that we can identify the MTE by applying Leibniz' rule to this expression:

$$MTE(p) = \frac{d\mathbb{E}[Y \mid P(Z) = p]}{dp}$$

In computational terms, we can use the `lprobust` of the `nprobust` package with `y =` Y, `x =` $P(z)$, `deriv = 1` and `eval` $= \mathcal{V}$. Since we want a local quadratic regression, we set `p = 2`. All other parameters are set to their default.

2.3 Simulation Results

We show our simulation results in Figures 1 and 2. Confidence intervals (CIs) are constructed using the distribution of our $1,000$ simulations (bands are the 5th and 95th quantiles).

In Figure 1, we show the true and estimated MTE functions: both models do a good job in estimating the true MTE function at all values of $\mathcal{V}$, specially the parametric one.

Figure 2 shows that the parametric model indeed has a lower relative bias relative (multiplied by 100, in p.p.) to the nonparametric specification. This is because we actually have the parametric form of the DGP and we specified it correctly, so these estimates are consistent and relatively more precise.

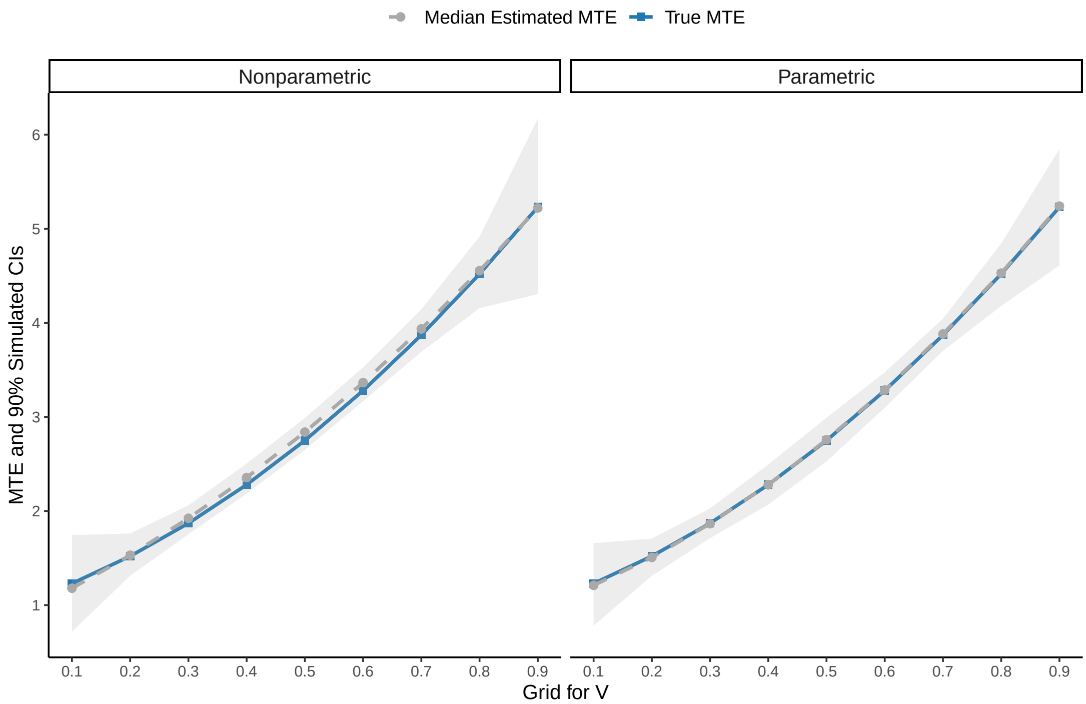

Figure 1: True and Estimate MTEs

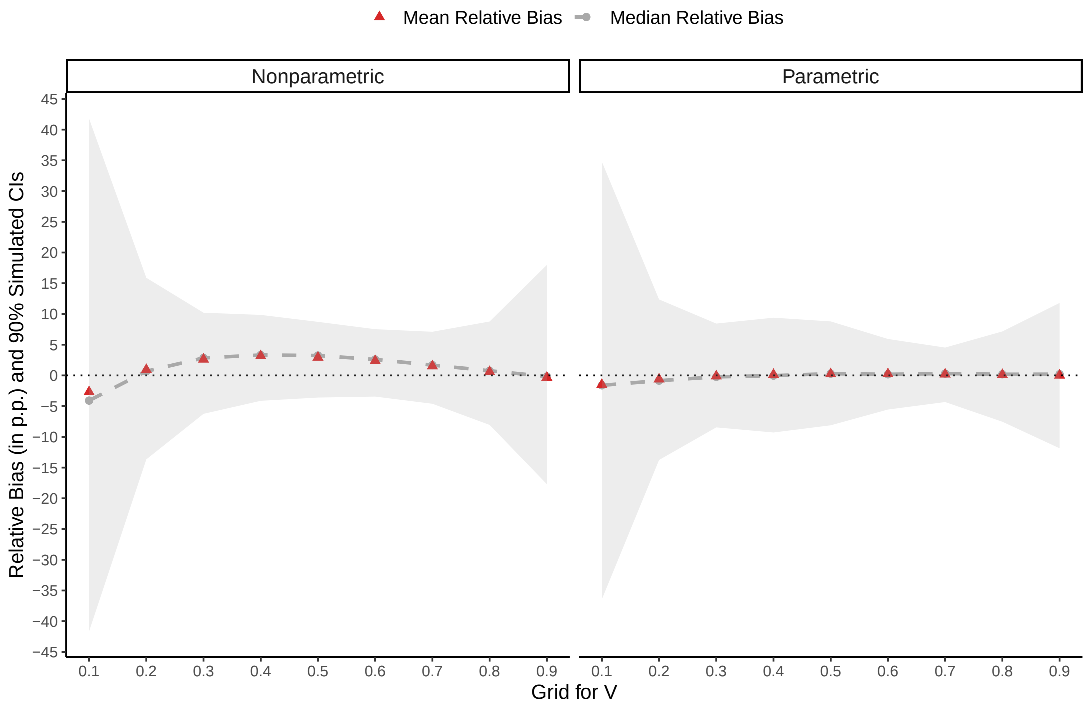

Figure 2: Median and Mean Relative Bias

3 Interpreting the MTE Function

We will interpret Figure IV(B) of Agan, Doleac and Harvey (2023). They focus on the effect of non-prosecution (treatment) into criminal recidivism (outcome). We reproduce the figure below:

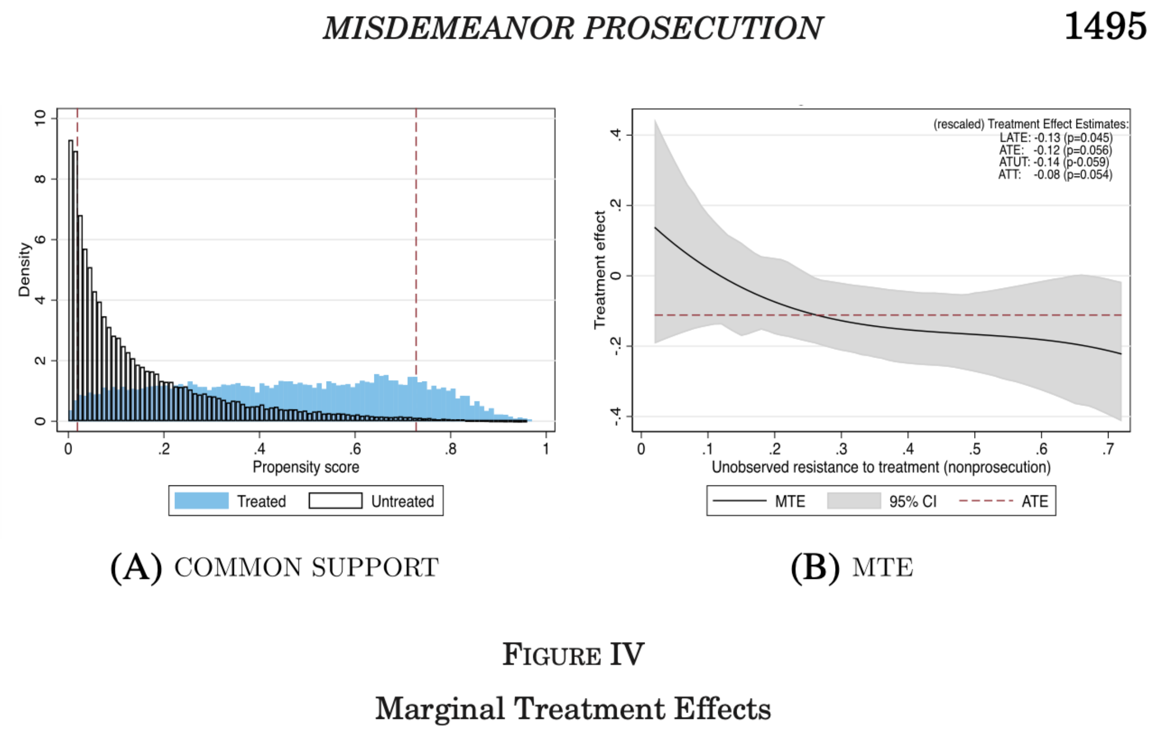

Figure 3: Figure IV of Agan, Doleac and Harvey (2023)

In this case, the MTE function is decreasing: treatment effects are higher for individuals with lower v, that is, those that are more likely to get treatment (in the case of the paper, of not being prosecuted).

In other words, those more likely to be prosecuted (less likely to be non-prosecuted, right of the graph) experience decreases in recidivism. In turn, those less likely to be prosecuted (more likely to be non-prosecuted, left of the graph) have a higher chance of subsequent crime, which might reflect loss of faith in the justice system.

As the authors write, this implies that increasing the leniency of (non)prosecution decisions would not cause increases in recidivism.