Selection on Observables: Theory

ECON 31720 Applied Microeconometrics

Francesco Ruggieri

The University of Chicago

October 7, 2020

Framework for Selection on Observables

- $D \in \{0, 1\}$ is a **binary treatment**, $Y \in \mathbb{R}$ is an **outcome** of interest
- D and Y are linked by **potential outcomes** $Y(0), Y(1)$
- $X \in \mathbb{R}^{d_x}$ is a vector of predetermined, **observable** characteristics with support $\mathcal{X}$
- The treatment is **as-good-as randomly assigned conditional on observables**:

$$(Y(0), Y(1)) \perp\!\!\!\perp D | X = x \qquad \text{for all } x \in \mathcal{X}$$

- Three canonical **target parameters**:

$$\mathrm{ATE} \equiv \mathbb{E}\left[Y(1) - Y(0)\right]$$
$$\mathrm{ATT} \equiv \mathbb{E}\left[Y(1) - Y(0) | D = 1\right]$$
$$\mathrm{ATU} \equiv \mathbb{E}\left[Y(1) - Y(0) | D = 0\right]$$

Average Treatment Effect

Begin with the **Average Treatment Effect**:

$$
\begin{aligned}
\text{ATE} &\equiv \mathbb{E}[Y(1) - Y(0)] \\
&= \mathbb{E}[\mathbb{E}[Y(1) - Y(0)|X]] && \text{by the Law of Iterated Expectations} \\
&= \mathbb{E}[\mathbb{E}[Y(1)|X] - \mathbb{E}[Y(0)|X]] && \text{by the linearity of } \mathbb{E}[\cdot] \\
&= \mathbb{E}[\mathbb{E}[Y(1)|D=1,X] - \mathbb{E}[Y(0)|D=0,X]] && \text{because } (Y(0),Y(1)) \perp\!\!\!\perp D|X \\
&= \mathbb{E}[\mathbb{E}[Y|D=1,X] - \mathbb{E}[Y|D=0,X]] && \text{because, conditional on } D=d,\ Y=Y(d)
\end{aligned}
$$

Intuitively, the ATE can be backed out in **two steps**:

1. Compute $\mathbb{E}[Y|D=1,X=x] - \mathbb{E}[Y|D=0,X=x]$ for all $x \in \mathcal{X}$
2. Integrate $\mathbb{E}[Y|D=1,X] - \mathbb{E}[Y|D=0,X]$ using the unconditional distribution of X

Average Treatment Effect on the Treated/Untreated

Consider the **Average Treatment Effect on the Treated**:

$$\begin{aligned}
\mathrm{ATT} \equiv \mathbb{E}[Y(1)-Y(0)|D=1] &= \mathbb{E}[Y(1)|D=1]-\mathbb{E}[Y(0)|D=1] && \text{by the linearity of } \mathbb{E}[\cdot] \\
&= \mathbb{E}[Y|D=1]-\mathbb{E}[Y(0)|D=1] && \text{because, conditional on } D=1,\ Y=Y(1) \\
&= \mathbb{E}[Y|D=1]-\mathbb{E}[\mathbb{E}[Y(0)|D=1,X]\,|D=1] && \text{by the Law of Iterated Expectations} \\
&= \mathbb{E}[Y|D=1]-\mathbb{E}[\mathbb{E}[Y(0)|D=0,X]\,|D=1] && \text{because } Y(0) \perp\!\!\!\perp D|X \\
&= \mathbb{E}[Y|D=1]-\mathbb{E}[\mathbb{E}[Y|D=0,X]\,|D=1] && \text{because, conditional on } D=0,\ Y=Y(0)
\end{aligned}$$

Intuitively, the ATT can be backed out in **three steps**:

1. Compute $\mathbb{E}[Y|D=0,X=x]$ for all $x \in \mathcal{X}$
2. Integrate $\mathbb{E}[Y|D=0,X]$ using the distribution of X among units with $D=1$
3. Compute $\mathbb{E}[Y|D=1]$ and subtract the quantity in (2)

Average Treatment Effect on the Treated/Untreated

The **Average Treatment Effect on the Untreated** is symmetrically identified:

$$\mathrm{ATU} = \mathbb{E}\left[\mathbb{E}\left[Y|D=1,X\right]|D=0\right] - \mathbb{E}\left[Y|D=0\right]$$

Heuristically, it can be backed out in **three steps**:

1. Compute $\mathbb{E}\left[Y|D=1,X=x\right]$ for all $x \in \mathcal{X}$
2. Integrate $\mathbb{E}\left[Y|D=1,X\right]$ using the distribution of X among units with $D=0$
3. Compute $\mathbb{E}\left[Y|D=0\right]$ and subtract it from the quantity in (2)

Notice that the ATT and the ATU require **weaker assumptions** than the ATE:

- **Ignorable treatment assignment**: $Y(0) \perp\!\!\!\perp D|X$ (ATT) and $Y(1) \perp\!\!\!\perp D|X$ (ATU)
- **Overlap**: $\forall x \in \mathcal{X}, \mathbb{P}\left(D=0|X=x\right) > 0$ (ATT) and $\mathbb{P}\left(D=1|X=x\right) > 0$ (ATU)

Estimation of Target Parameters

- To summarize, under the assumption of selection on observables:

$$\begin{aligned} \text{ATE} &= \mathbb{E}\left[\mathbb{E}\left[Y|D=1,X\right]-\mathbb{E}\left[Y|D=0,X\right]\right] \\ \text{ATT} &= \mathbb{E}\left[Y|D=1\right]-\mathbb{E}\left[\mathbb{E}\left[Y|D=0,X\right]|D=1\right] \\ \text{ATU} &= \mathbb{E}\left[\mathbb{E}\left[Y|D=1,X\right]|D=0\right]-\mathbb{E}\left[Y|D=0\right] \end{aligned}$$

- Let us focus on **alternative estimators** for these target parameters:
  1. Imputation if X is **discrete**
  2. Imputation with linear regression **with additive separability** between D and X
  3. Imputation with linear regression **without additive separability** between D and X

Imputation if X is Discrete

- Suppose $\mathcal{X} = \left\{x_1, \ldots, x_{\bar{j}}\right\}$ and $\mathbb{P}(X = x_j) > 0 \ \forall j \in \{1, \ldots, \bar{j}\}$
- In this case, the population target parameters can be expressed as

$$\mathrm{ATE} = \sum_{j=1}^{\bar{j}} \left(\mathbb{E}[Y|D=1, X=x_j] - \mathbb{E}[Y|D=0, X=x_j]\right) \times \mathbb{P}(X=x_j)$$

$$\mathrm{ATT} = \mathbb{E}[Y|D=1] - \sum_{j=1}^{\bar{j}} \mathbb{E}[Y|D=0, X=x_j] \times \mathbb{P}(X=x_j|D=1)$$

$$\mathrm{ATU} = \sum_{j=1}^{\bar{j}} \mathbb{E}[Y|D=1, X=x_j] \times \mathbb{P}(X=x_j|D=0) - \mathbb{E}[Y|D=0]$$

Imputation if X is Discrete

Consider a n-dimensional sample of i.i.d. random variables, $\{Y_i, D_i, X_i\}_{i=1}^n$. For $d \in \{0,1\}$:

$$\widehat{M}_d \equiv \frac{\sum_{i=1}^n Y_i \times \mathbb{I}[D_i = d]}{\sum_{i=1}^n \mathbb{I}[D_i = d]} \xrightarrow{p} \mathbb{E}[Y|D = d]$$

$$\widehat{M}_d(x_j) \equiv \frac{\sum_{i=1}^n Y_i \times \mathbb{I}[X_i = x_j, D_i = d]}{\sum_{i=1}^n \mathbb{I}[X_i = x_j, D_i = d]} \xrightarrow{p} \mathbb{E}[Y|X = x_j, D = d]$$

$$\overline{X}_j \equiv \frac{\sum_{i=1}^n \mathbb{I}[X_i = x_j]}{n} \xrightarrow{p} \mathbb{P}(X = x_j)$$

$$\overline{X}_{dj} \equiv \frac{\sum_{i=1}^n \mathbb{I}[X_i = x_j, D_i = d]}{\sum_{i=1}^n \mathbb{I}[D_i = d]} \xrightarrow{p} \mathbb{P}(X = x_j|D = d)$$

Imputation if X is Discrete

Imputation estimators for the three target parameters of interest are the following:

$$\widehat{\text{ATE}} \equiv \sum_{j=1}^{\bar{j}} \left[\widehat{M}_1(x_j) - \widehat{M}_0(x_j) \right] \times \overline{X}_j \xrightarrow{p} \text{ATE}$$

$$\widehat{\text{ATT}} \equiv \widehat{M}_1 - \sum_{j=1}^{\bar{j}} \widehat{M}_0(x_j) \times \overline{X}_{1j} \xrightarrow{p} \text{ATT}$$

$$\widehat{\text{ATU}} \equiv \sum_{j=1}^{\bar{j}} \widehat{M}_1(x_j) \times \overline{X}_{0j} - \widehat{M}_0 \xrightarrow{p} \text{ATU}$$

However, X is often **high-dimensional** and/or has components with **continuous support**...

Imputation with Linear Regression

- If $X \in \mathbb{R}$, the **curse of dimensionality** will typically kick in quickly
  - It may be **infeasible** to compute conditional outcome means within bins implied by X
- Within-bin conditional outcome means may be **approximated with linear regression**
- A key choice in this regard is whether to **allow for separability** between D and X:
  1. If D and X are assumed to be **additively separable**, then

$$Y = \alpha^* + \beta^* D + \gamma^* X + U \quad \text{with} \quad \mathbb{E}[U] = \mathbb{E}[DU] = \mathbb{E}[XU] = 0$$

  2. If D and X are **not** assumed to be **additively separable**, then

$$Y = \alpha^* + \beta^* D + \gamma^* X + \delta^* DX + U \quad \text{with} \quad \mathbb{E}[U] = \mathbb{E}[DU] = \mathbb{E}[XU] = \mathbb{E}[DXU] = 0$$

1. Framework for Selection on Observables
2. Identification of Target Parameters
3. Estimation of Target Parameters
   - Imputation if X is Discrete
   - Imputation with Linear Regression with Additive Separability between D and X
   - Imputation with Linear Regression without Additive Separability between D and X
4. Summary

Imputation with Linear Regression with Additive Separability

If one assumes **additive separability between $\boldsymbol{D}$ and $\boldsymbol{X}$**, then

$$Y = \alpha^* + \beta^* D + \gamma^* X + U \qquad \text{with } \mathbb{E}[U] = \mathbb{E}[DU] = \mathbb{E}[XU] = 0$$

Thus, **conditional ATEs** can be **approximated** as follows:

$$\begin{aligned}
\mathrm{ATE}(x) &\equiv \mathbb{E}[Y(1) - Y(0)|X = x] \\
&= \mathbb{E}[Y(1)|X = x] - \mathbb{E}[Y(0)|X = x] \qquad \text{by the linearity of } \mathbb{E}[\cdot] \\
&= \mathbb{E}[Y(1)|D = 1, X = x] - \mathbb{E}[Y(0)|D = 0, X = x] \qquad \text{because } (Y(0), Y(1)) \perp\!\!\!\perp D|X \\
&= \mathbb{E}[Y|D = 1, X = x] - \mathbb{E}[Y|D = 0, X = x] \qquad \text{because, condit. on } D = d,\ Y = Y(d) \\
&\approx \mathbb{L}(Y|D = 1, X = x) - \mathbb{L}(Y|D = 0, X = x) \\
&= (\alpha^* + \beta^* + \gamma^* x) - (\alpha^* + \gamma^* x) \\
&= \beta^*
\end{aligned}$$

Imputation with Linear Regression with Additive Separability

- With additive separability, ATEs are **constant across bins** implied by the observables:

$$\mathrm{ATE}(x) \approx \beta^* \quad \forall x \in \mathcal{X}$$

- Because $(Y(0), Y(1)) \perp\!\!\!\perp D|X$

$$\begin{aligned}\mathrm{ATE}(X) \equiv \mathbb{E}[Y(1) - Y(0)|X] &= \mathbb{E}[Y(1) - Y(0)|D = 1, X] \equiv \mathrm{ATT}(X) \\ &= \mathbb{E}[Y(1) - Y(0)|D = 0, X] \equiv \mathrm{ATU}(X)\end{aligned}$$

- Thus, $\mathrm{ATE}(x) = \mathrm{ATT}(x) = \mathrm{ATU}(x) = \beta^* \ \forall x \in \mathcal{X}$
  - This may be **restrictive**, especially if agents make choices based on observable characteristics

Imputation with Linear Regression with Additive Separability

- **Unconditional target parameters** can be easily backed out:

$$\begin{aligned}
\text{ATE} &= \mathbb{E}\left[\text{ATE}(X)\right] \approx \mathbb{E}\left[\beta^*\right] = \beta^* \\
\text{ATT} &= \mathbb{E}\left[\text{ATT}(X)|D=1\right] = \mathbb{E}\left[\text{ATE}(X)|D=1\right] \approx \mathbb{E}\left[\beta^*|D=1\right] = \beta^* \\
\text{ATU} &= \mathbb{E}\left[\text{ATU}(X)|D=0\right] = \mathbb{E}\left[\text{ATE}(X)|D=0\right] \approx \mathbb{E}\left[\beta^*|D=0\right] = \beta^*
\end{aligned}$$

- Thus, with additive separability, $\mathbf{ATE} = \mathbf{ATT} = \mathbf{ATU} \approx \boldsymbol{\beta^*}$

- β^* can be consistently estimated with **Ordinary Least Squares**

Imputation with Linear Regression with Additive Separability

- Suppose **additive separability** between D and X were a **plausible assumption**
- If $\mathrm{ATE}(x) = c\ \forall x \in \mathcal{X}$, then probably not unrealistic to assume that $Y(1) - Y(0) = \tau$
- If **treatment effects** are **homogeneous**, is linear regression an ideal tool?

$$Y = \alpha^* + \beta^* D + \gamma^* X + U \quad \text{with} \quad \mathbb{E}[U] = \mathbb{E}[DU] = \mathbb{E}[XU] = 0$$

- In other words, how does $\boldsymbol{\beta^*}$ relate to $\boldsymbol{\tau}$?

Imputation with Linear Regression with Additive Separability

Let us focus on β^* and apply the Frisch-Waugh Theorem:

$$\beta^* = \frac{\mathbb{E}\left[\widetilde{D}Y\right]}{\mathbb{E}\left[\widetilde{D}^2\right]} \quad \text{where} \quad \widetilde{D} \equiv D - \mathbb{L}\left(D|1, X\right) = D - \pi_0^* - \pi_1^* X$$

Because D is regressed onto a constant, $\mathbb{E}\left[\widetilde{D}\right] = 0$. Thus:

$$\beta^* = \frac{\mathbb{E}\left[\widetilde{D}Y\right]}{\mathbb{E}\left[\widetilde{D}^2\right]} = \frac{\mathbb{E}\left[\left(\widetilde{D} - \mathbb{E}\left[\widetilde{D}\right]\right)Y\right]}{\mathbb{E}\left[\widetilde{D}^2\right] - \mathbb{E}\left[\widetilde{D}\right]^2} = \frac{\mathrm{Cov}\left[\widetilde{D}, Y\right]}{\mathrm{Var}\left[\widetilde{D}\right]}$$

Imputation with Linear Regression with Additive Separability

Apply the switching equation and exploit the orthogonality between $\widetilde{D}$ and X:

$$\beta^* = \frac{\mathrm{Cov}\left[\widetilde{D}, Y(0) + (Y(1) - Y(0))\, D\right]}{\mathrm{Var}\left[\widetilde{D}\right]} = \frac{\mathrm{Cov}\left[\widetilde{D}, Y(0) + \tau D\right]}{\mathrm{Var}\left[\widetilde{D}\right]}$$

$$= \frac{\mathrm{Cov}\left[\widetilde{D}, Y(0)\right]}{\mathrm{Var}\left[\widetilde{D}\right]} + \tau \frac{\mathrm{Cov}\left[\widetilde{D}, D\right]}{\mathrm{Var}\left[\widetilde{D}\right]} = \frac{\mathrm{Cov}\left[\widetilde{D}, Y(0)\right]}{\mathrm{Var}\left[\widetilde{D}\right]} + \tau \frac{\mathrm{Cov}\left[\widetilde{D}, D - \mathbb{L}\left(D | 1, X\right)\right]}{\mathrm{Var}\left[\widetilde{D}\right]}$$

$$= \tau + \frac{\mathrm{Cov}\left[\widetilde{D}, Y(0)\right]}{\mathrm{Var}\left[\widetilde{D}\right]}$$

Imputation with Linear Regression with Additive Separability

Finally, apply the definition of regression residual:

$$\beta^* = \tau + \frac{\mathrm{Cov}\left[D - \pi_0^* - \pi_1^* X, Y(0)\right]}{\mathrm{Var}\left[\widetilde{D}\right]} = \tau + \frac{\mathrm{Cov}\left[D - \pi_1^* X, Y(0)\right]}{\mathrm{Var}\left[\widetilde{D}\right]}$$

- The **denominator** is **strictly positive** because $\widetilde{D}$ is a nondegenerate random variable
- The **numerator** is of **ambiguous sign** as it depends on $\mathrm{Cov}\left[D, Y(0)\right]$, $\mathrm{Cov}\left[X, Y(0)\right]$, π_1^*
- Even if one assumes homogeneous treatment effects, then in general $\boldsymbol{\beta^* \neq \tau}$
  - They will be equal as long as $\mathrm{Cov}\left[D, Y(0)\right] = \pi_1^* \mathrm{Cov}\left[X, Y(0)\right]$, which is untestable

1. Framework for Selection on Observables
2. Identification of Target Parameters
3. Estimation of Target Parameters
   - Imputation if X is Discrete
   - Imputation with Linear Regression with Additive Separability between D and X
   - Imputation with Linear Regression without Additive Separability between D and X
4. Summary

Imputation with Linear Regression without Additive Separability

If one does **not assume additive separability** between D and X, then

$$Y = \alpha^* + \beta^* D + \gamma^* X + \delta^* DX + U \quad \text{with} \quad \mathbb{E}[U] = \mathbb{E}[DU] = \mathbb{E}[XU] = \mathbb{E}[DXU] = 0$$

Notice that this is numerically **equivalent** to regressing Y on X among units with $D = d$:

$$\begin{aligned} D = 0: \quad & Y = \alpha^* + \gamma^* X + V & \text{with} \quad & \mathbb{E}[V] = \mathbb{E}[XV] = 0 \\ D = 1: \quad & Y = \underbrace{(\alpha^* + \beta^*)}_{\lambda^*} + \underbrace{(\gamma^* + \delta^*)}_{\eta^*} X + W & \text{with} \quad & \mathbb{E}[W] = \mathbb{E}[XW] = 0 \end{aligned}$$

Thus, the linear regression at the top is analogous to the approach adopted by Imbens (2004)

Imputation with Linear Regression without Additive Separability

Conditional ATEs can be **approximated** as follows:

$$
\begin{aligned}
\mathrm{ATE}(x) &\equiv \mathbb{E}[Y(1) - Y(0)|X = x] \\
&= \mathbb{E}[Y(1)|X = x] - \mathbb{E}[Y(0)|X = x] \qquad \text{by the linearity of } \mathbb{E}[\cdot] \\
&= \mathbb{E}[Y(1)|D = 1, X = x] - \mathbb{E}[Y(0)|D = 0, X = x] \qquad \text{because } (Y(0), Y(1)) \perp\!\!\!\perp D|X \\
&= \mathbb{E}[Y|D = 1, X = x] - \mathbb{E}[Y|D = 0, X = x] \qquad \text{because, condit. on } D = d,\ Y = Y(d) \\
&\approx \mathbb{L}(Y|D = 1, X = x) - \mathbb{L}(Y|D = 0, X = x) \\
&= (\alpha^* + \beta^* + \gamma^* x + \delta^* x) - (\alpha^* + \gamma^* x) \\
&= \beta^* + \delta^* x
\end{aligned}
$$

Without additive separability between D and X, ATEs **vary across bins** implied by X:

$$\mathrm{ATE}(x) \approx \beta^* + \delta^* x \qquad \forall x \in \mathcal{X}$$

Imputation with Linear Regression without Additive Separability

- As in the previous case, conditional target parameters are equal to each other:

$$\text{ATE}(X) = \text{ATT}(X) = \text{ATU}(X) \approx \beta^* + \delta^* X$$

- **Unconditional target parameters** can be backed out as follows:

$$\begin{aligned}
\text{ATE} &= \mathbb{E}[\text{ATE}(X)] \approx \mathbb{E}[\beta^* + \delta^* X] = \beta^* + \delta^* \mathbb{E}[X] \\
\text{ATT} &= \mathbb{E}[\text{ATT}(X) | D = 1] \approx \mathbb{E}[\beta^* + \delta^* X | D = 1] = \beta^* + \delta^* \mathbb{E}[X | D = 1] \\
\text{ATU} &= \mathbb{E}[\text{ATU}(X) | D = 0] \approx \mathbb{E}[\beta^* + \delta^* X | D = 0] = \beta^* + \delta^* \mathbb{E}[X | D = 0]
\end{aligned}$$

- Without additive separability, target parameters will generally **differ from each other**
  - $\text{ATE} = \text{ATT} = \text{ATU} \approx \beta^* + \delta^* \mathbb{E}[X]$ in the case in which $\mathbb{E}[X|D] = \mathbb{E}[X]$

Imputation with Linear Regression without Additive Separability

Consider a n-dimensional sample of i.i.d. random variables, $\{Y_i, D_i, X_i\}_{i=1}^n$:

- β^* and δ^* can be consistently estimated with **Ordinary Least Squares**

- The remaining components of the target parameters can be estimated as

$$\overline{X} \equiv \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p} \mathbb{E}[X]$$
$$\overline{X}_1 \equiv \frac{\sum_{i=1}^n X_i D_i}{\sum_{i=1}^n D_i} \xrightarrow{p} \mathbb{E}[X|D=1]$$
$$\overline{X}_0 \equiv \frac{\sum_{i=1}^n X_i (1-D_i)}{\sum_{i=1}^n (1-D_i)} \xrightarrow{p} \mathbb{E}[X|D=0]$$

- ATE, ATT, ATU are consistently estimated via **Continuous Mapping Theorem**

1. Framework for Selection on Observables

2. Identification of Target Parameters

3. Estimation of Target Parameters
   - Imputation if X is Discrete
   - Imputation with Linear Regression with Additive Separability between D and X
   - Imputation with Linear Regression without Additive Separability between D and X

4. Summary

Summary

Under the assumption of selection on observables, target parameters can be estimated as follows:

Parameter	Identification	Discrete X
ATE	$\mathbb{E}\left[\mathbb{E}\left[Y\vert D=1,X\right]-\mathbb{E}\left[Y\vert D=0,X\right]\right]$	$\sum_{j=1}^{J}\left[\widehat{M}_1\left(x_j\right)-\widehat{M}_0\left(x_j\right)\right]\overline{X}_j$
ATT	$\mathbb{E}\left[Y\vert D=1\right]-\mathbb{E}\left[\mathbb{E}\left[Y\vert D=0,X\right]\vert D=1\right]$	$\widehat{M}_1-\sum_{j=1}^{J}\widehat{M}_0\left(x_j\right)\overline{X}_{1j}$
ATU	$\mathbb{E}\left[\mathbb{E}\left[Y\vert D=1,X\right]\vert D=0\right]-\mathbb{E}\left[Y\vert D=0\right]$	$\sum_{j=1}^{J}\widehat{M}_1\left(x_j\right)\overline{X}_{0j}-\widehat{M}_0$

Parameter	Imputation with LR, with Separability	Imputation with LR, without Separability
ATE	$\approx\widehat{B}_{\text{OLS}}$	$\approx\widehat{B}_{\text{OLS}}+\widehat{D}_{\text{OLS}}\overline{X}$
ATT	$\approx\widehat{B}_{\text{OLS}}$	$\approx\widehat{B}_{\text{OLS}}+\widehat{D}_{\text{OLS}}\overline{X}_1$
ATU	$\approx\widehat{B}_{\text{OLS}}$	$\approx\widehat{B}_{\text{OLS}}+\widehat{D}_{\text{OLS}}\overline{X}_0$