CEF, OLS, and Causal Identification

Causal Inference

Spring 2025

Overview: From Statistics to Causality

- **Previously covered:**
  - The Potential Outcomes Framework ($Y(0), Y(1)$)
  - Identification and Estimation
- **Today's Goal:**
  - Linking (Causal) Identification and (Statistical) Estimation
  1. **Causal Estimand**: The Average Treatment Effect ($E[Y(1) - Y(0)]$).
  2. **Identification Assumptions**: Ignorability and Positivity.
  3. **Identified Statistical Object**: The CEF ($E[Y|D, X]$).
  4. **Estimation Strategy**: Using OLS to approximate the CEF.

Causal Estimand

- **The Causal Estimand** is the *theoretical* target in the *population*.
- It is what we want to know, independent of any specific data or estimator.
- **Average Treatment Effect (ATE):**

$$\tau_{ATE} = E[Y_i(1) - Y_i(0)]$$

- **Fundamental Problem of Causal Inference:** We never observe both $Y_i(1)$ and $Y_i(0)$ for the same unit.
- **Selection Bias**: Observed differences (SDM) do not equal the ATE unless $E[Y(0)|D = 1] = E[Y(0)|D = 0]$.

Identification Assumptions

- Identification is the process of expressing the **Causal Estimand** in terms of **Observed Data**.

1. **Conditional Ignorability (CIA)**:

$$\{Y(0), Y(1)\} \perp D \mid X$$

This implies: $E[Y(d)|X] = E[Y|D = d, X]$

2. **Positivity (Overlap)**:

$$0 < P(D = 1 \mid X = x) < 1 \quad \forall x$$

This ensures: We have both treated and control units across all values of X. If $P(D = 1|X) = 1$, we have no counterfactual.

Statistical Estimand

- Once identified, our target is no longer a "potential" outcome; it is a **Statistical Estimand**: the **Conditional Expectation Function (CEF)**.
- Under CIA, the ATE is identified as:

$$\tau_{ATE} = E_X[\underbrace{E[Y|D=1,X] - E[Y|D=0,X]}_{\delta(X) \text{ (The Stratum (conditional on } X \text{) Effect)}}]$$

- **Law of Iterated Expectations (LIE)** to recover population ATE

OLS as an Estimation Strategy

- $Y = \beta D + \gamma X + \epsilon$, where X is a set of strata dummies
- By the **Frisch-Waugh-Lovell Theorem**(QPM1, partialling out approach):
  - $D - E[D|X]$: "remaining" variation
  - The OLS estimator for β is:

$$
\begin{aligned}
\beta_{OLS} &= \frac{Cov(\text{residuals}_D, Y)}{Var(\text{residuals}_D)} \\
&= \frac{E[\text{residuals}_D Y] - E[\text{residuals}_D]E[y]}{Var(\text{residuals}_D)} \\
&= \frac{E[(D - E[D|X])Y]}{E[(D - E[D|X])^2]}
\end{aligned}
$$

- The denominator $E[(D - E[D|X])^2]$ is the expected conditional variance, $E_X[\text{Var}(D|X)]$.
- Since D is binary, $\text{Var}(D|X) = P(D = 1|X)(1 - P(D = 1|X))$.

OLS as an Estimation Strategy

- We can rewrite the numerator using the Law of Iterated Expectations:

$$E[(D - E[D|X])Y] = E_X[E[(D - E[D|X])Y \mid X]]$$

- Focus on the inner expectation $E[(D - E[D|X])Y \mid X]$:

$$E[DY \mid X] - E[D \mid X]E[Y \mid X]$$

- By the definition of $E[DY|X]$:

$$E[DY|X] = P(D = 1|X)E[1 \cdot Y|D = 1, X] + P(D = 0 \mid X)E[0 \cdot Y \mid D = 0, X]$$

- Since D is a binary, $E[D \mid X] = P(D = 1 \mid X)$
- Substitute the definition of $E[Y|X]$ for a binary D:

$$E[Y|X] = P(D = 1|X)E[Y|D = 1, X] + P(D = 0|X)E[Y|D = 0, X]$$

OLS as an Estimation Strategy

$$
\begin{aligned}
E[DY \mid X]-E[D \mid X]E[Y \mid X] \\
&= P(D=1|X)E[Y|D=1,X] - P(D=1 \mid X)E[Y|X] \\
&= P(D=1 \mid X)\,(E[Y \mid D=1,X] - E[Y \mid X]) \\
&= P(D=1 \mid X)\Big(E[Y \mid D=1,X] \\
&\qquad - P(D=1 \mid X)E[Y \mid D=1,X] - P(D=0 \mid X)E[Y \mid D=0,X]\Big) \\
&= P(D=1 \mid X)\Big(\\
&\quad P(D=1 \mid X)E[Y \mid D=1,X] + P(D=0 \mid X)E[Y \mid D=1,X] \\
&\qquad - P(D=1 \mid X)E[Y \mid D=1,X] - P(D=0 \mid X)E[Y \mid D=0,X] \\
&\quad \Big) \\
&= P(D=1 \mid X)P(D=0 \mid X)\Big(E[Y \mid D=1,X] - E[Y \mid D=0,X]\Big)
\end{aligned}
$$

OLS as an Estimation Strategy

- Back into the full formula:

$$\beta_{OLS} = \frac{E_X[\overbrace{P(D=1|X)(1-P(D=1|X))}^{\text{Var}(D|X)} \cdot \overbrace{(E[Y|D=1,X]-E[Y|D=0,X])}^{\delta(X)}]}{E_X[P(D=1|X)(1-P(D=1|X))]}$$

- If X is discrete, this is exactly a weighted average of $\delta(X)$:

$$\beta_{OLS} = \frac{\sum_x w(x)\delta(x)}{\sum_x w(x)} \quad \text{where } w(x) = P(X=x)\text{Var}(D|X=x)$$

- $\delta(x)$ is the treatment effect within a specific group $X = x$.
- $w(x)$ is the weight given to that group by OLS.

OLS (Estimation) vs. ATE (Estimand)

1. **ATE (Estimand)**:
   - Weights each $X = x$ strictly by its population size, $P(X = x)$.
   - $\tau_{ATE} = \sum_X P(X = x)\delta(X = x)$.
2. **OLS (Estimation Strategy)**:
   - Weights each $X = x$ by $P(X = x)Var(D \mid X = x)$
   - Linear approximation

R Application: Wage Gender Gap and Working Hours

- **Simple Difference**: `r mean(log_earn[female==1]) - mean(log_earn[female==0])`
  - Statistical Estimand: $E[Y|D=1] - E[Y|D=0]$.
  - Only causal if Selection Bias is zero.
- **Binned CEF**: `r group_by(hours) %>% summarise(mean(log_earn))`
  - Identified Statistical Object: $E[Y|X]$.
  - Does not assume linearity; shows non-linearities OLS might miss.
- **Multivariate Regression**: `r lm(log_earn ~ female + hours)`
  - Estimation Strategy: β_{OLS} is the variance-weighted average of the gender gap across different categories of hours worked.

Conclusion

- **Distinguish between**:
  - **Causal Estimand**: The ATE (The theoretical target).
  - **Identified Object**: The CEF (The target translated to data).
  - **Estimator**: OLS (The strategy to calculate the identified object).
- **The OLS Weighting Result**: OLS $\neq$ ATE if treatment effects $\delta(x)$ are **heterogeneous** across strata.
- OLS is an efficient estimator for a common effect, but a variance-weighted average for heterogeneous effects.