ECMA 31100: Intro to Empirical Analysis II

Intro; Selection on Observables

Joe Hardwick

University of Chicago

Winter 2022

Causality

- So far we have considered counterfactuals using linear regression:
  - How much does earnings potential increase with another year of experience, holding all else fixed?
  - How much does an additional bedroom increase house price, holding floor area/lot size/air pollution (and all else) fixed?
- OLS estimates proportional to "partial correlations" between outcome and regressors.
  - 'Holding fixed' other observable factors in this way answers counterfactual question if $y = x'\beta + u$ and $\mathrm{E}(xu) = 0$.
  - $\mathrm{E}(xu) = 0 \implies \beta = \mathrm{E}(xx')^{-1}\mathrm{E}(xu)$: (Scaled) partial correlations are partial (causal) effects.
  - Partial correlations can be estimated with data.

Rubin Causal Model

- Can think about counterfactuals as 'potential outcomes' under different 'treatments':
  - Earnings with college degree, $y(1)$, and without $y(0)$.
  - Health outcome with drug A, $y(1)$, vs. drug B, $y(0)$.
  - Productivity with job training, $y(1)$ vs. without, $y(0)$.
- Define individual treatment effect for individual i as

$$y_i(1) - y_i(0).$$

- Denote treatment status of individual i by $D_i \in \{0, 1\}$.

Store Upgrades

- $y_i(1)$ denotes sales store i *would* observe with an upgrade.
- $y_i(0)$ denotes sales store i *would* observe without an upgrade.
- Outcome observed

$$Y_i = \begin{cases} y_i(1) & \text{if store } i \text{ is upgraded,} \\ y_i(0) & \text{if not.} \end{cases}$$

- In other words:

$$Y_i = y_i(1) D_i + y_i(0)(1 - D_i).$$

Store Upgrades

Store Upgrades

- We never observe the individual causal effect $y_i(1) - y_i(0)$ for any store because only one of the two potential outcomes is ever observed.
- Can try to measure average effects across treated and untreated stores by comparing averages:
  - e.g. Average Treatment Effect in store population (ATE): $\mathrm{E}(y(1) - y(0))$.
  - e.g. Average Treatment Effect on the treated (ATT): $\mathrm{E}(y(1) - y(0) | D = 1)$.
  - e.g. Average Treatment Effect on untreated (ATU): $\mathrm{E}(y(1) - y(0) | D = 0)$.

Which expectations can we estimate?

- First write out the conditional mean of $Y|D$:

$$\begin{aligned}
\mathrm{E}\left(Y|D=1\right) &= \mathrm{E}\left(y\left(1\right)D + y\left(0\right)\left(1-D\right)|D=1\right) \\
&= \mathrm{E}\left(y\left(1\right)|D=1\right); \\
\mathrm{E}\left(Y|D=0\right) &= \mathrm{E}\left(y\left(1\right)D + y\left(0\right)\left(1-D\right)|D=0\right) \\
&= \mathrm{E}\left(y\left(0\right)|D=0\right).
\end{aligned}$$

- We can observe $\mathrm{E}\left(y\left(1\right)|D=1\right)$ and $\mathrm{E}\left(y\left(0\right)|D=0\right)$. Suppose $\mathrm{E}\left(y\left(1\right)|D=1\right)=100$ and $\mathrm{E}\left(y\left(0\right)|D=0\right)=50$.

Average potential outcomes

Questions

- Should we upgrade stores that have not already been upgraded? -> ATU
- What was the effect of the upgrade on stores that have already been upgraded? -> ATT
- If upgraded all stores to begin with, what would be the average effect on sales be? -> ATE
- Data reveal an estimate of the 'naive comparison':

$$\mathrm{E}\left(y\left(1\right)|D=1\right)-\mathrm{E}\left(y\left(0\right)|D=0\right)=50.$$

Relationship between ATE,ATT,ATU

- We can decompose the ATE as follows:

$$\begin{aligned}\mathrm{E}\left(y\left(1\right)-y\left(0\right)\right) &= \mathrm{E}\left(y\left(1\right)-y\left(0\right)|D=1\right)\cdot\mathrm{P}\left(D=1\right)\\ &\quad+\mathrm{E}\left(y\left(1\right)-y\left(0\right)|D=0\right)\cdot\mathrm{P}\left(D=0\right)\\ &= ATT\cdot\mathrm{P}\left(D=1\right)+ATU\cdot\mathrm{P}\left(D=0\right)\end{aligned}$$

  by the law of iterated expectation.

Regression

- Write out the conditional mean of $Y|D$:

$$\begin{aligned}\mathrm{E}\left(Y|D=1\right) &= \mathrm{E}\left(y\left(1\right)|D=1\right);\\ \mathrm{E}\left(Y|D=0\right) &= \mathrm{E}\left(y\left(0\right)|D=0\right).\end{aligned}$$

It follows that

$$\mathrm{E}\left(Y|D\right) = \beta_0 + \beta_1 D,$$

where

$$\begin{aligned}\beta_0 &= \mathrm{E}\left(y\left(0\right)|D=0\right)\\ \beta_1 &= \mathrm{E}\left(y\left(1\right)|D=1\right) - \mathrm{E}\left(y\left(0\right)|D=0\right).\end{aligned}$$

So β_1 equals the naive comparison, and we get an unbiased estimate by regressing Y on D.

Regression

- Recall: The slope coefficient in linear regression is proportional to correlation, and correlation may not reflect causality.
- What is β_1?

$$\begin{aligned}\beta_1 &= \mathrm{E}\left(y\left(1\right)|D=1\right) - \mathrm{E}\left(y\left(0\right)|D=0\right) \\ &= \underbrace{\mathrm{E}\left(y\left(1\right)|D=1\right) - \mathrm{E}\left(y\left(0\right)|D=1\right)}_{ATT} \\ &+ \underbrace{\mathrm{E}\left(y\left(0\right)|D=1\right) - \mathrm{E}\left(y\left(0\right)|D=0\right)}_{Selection\ Bias}\end{aligned}$$

- β_1 measures the average effect on stores that were upgraded (ATT) but is offset by "selection bias", which measures differences in average sales in the absence of upgrades.

Treatment Assignment

- Treatment assignment tells us how units are assigned to treatment.
- Random assignment:

$$D_i \perp y_i(0), y_i(1).$$

- This says treatment choice/assignment is independent of outcomes under treatment and no treatment.
  - Often violated when treatment is chosen: Management will try to choose stores for which effect on sales $y_i(1) - y_i(0)$ is greatest.
  - Plausible in experimental setting: Randomized controlled trials.

Random Assignment

- Under random assignment:

$$\mathrm{E}\left(y\left(0\right)|D\right) = \mathrm{E}\left(y\left(0\right)\right);$$
$$\mathrm{E}\left(y\left(1\right)|D\right) = \mathrm{E}\left(y\left(1\right)\right);$$

  i.e. the treatment and control groups are representative of the entire population. Follows that

$$\begin{aligned}\beta_1 &= \mathrm{E}\left(y\left(1\right)|D=1\right) - \mathrm{E}\left(y\left(0\right)|D=0\right)\\ &= \mathrm{E}\left(y\left(1\right)\right) - \mathrm{E}\left(y\left(0\right)\right)\\ &= ATE.\end{aligned}$$

- Conclusion: Under randomized treatment assignment estimating a linear regression produces an unbiased estimate of the ATE.

Multiple possible treatments

- Suppose in a randomized clinical trial there is a control group and k possible treatments for a particular health condition.
- Suppose participants are randomly assigned to either control or one of the treatments. Let

$$x_{i1} = \begin{cases} 1 & \text{if } i \text{ receives treatment } 1, \\ 0 & \text{otherwise;} \end{cases}$$

$$x_{i2} = \begin{cases} 1 & \text{if } i \text{ receives treatment } 2, \\ 0 & \text{otherwise;} \end{cases}$$

$$\vdots$$

$$x_{ik} = \begin{cases} 1 & \text{if } i \text{ receives treatment } k, \\ 0 & \text{otherwise;} \end{cases}$$

Multiple possible treatments

- There are now $k+1$ potential outcomes $y_i(0), y_i(1), \ldots, y_i(k)$ and

$$Y_i = y_i(0) + x_{i1}(y_i(1) - y_i(0)) + \cdots + x_{ik}(y_i(k) - y_i(0)).$$

See that:

$$\begin{aligned}
\mathrm{E}(Y_i | x_{i1} = 0, \ldots x_{ik} = 0) &= \mathrm{E}(y_i(0) | x_{i1} = 0, \ldots x_{ik} = 0) \\
(\text{Random Assignment }) &= \mathrm{E}(y_i(0)) \\
&\vdots \\
\mathrm{E}(Y_i | x_{i1} = 0, \ldots x_{ik} = 1) &= \mathrm{E}(y_i(k) | x_{i1} = 0, \ldots x_{ik} = 1) \\
(\text{Random Assignment }) &= \mathrm{E}(y_i(k)).
\end{aligned}$$

Multiple possible treatments

- Now let $x_i = (1, x_{i1}, \ldots, x_{ik})'$. We have

$$\mathrm{E}\left(Y_i | x_i\right) = \beta_0 + \beta_1 x_{i1} + \ldots \beta_k x_{ik},$$

where

$$\begin{aligned} \beta_0 &= \mathrm{E}\left(y\left(0\right)\right), \\ &\vdots \\ \beta_k &= \mathrm{E}\left(y\left(k\right) - y\left(0\right)\right). \end{aligned}$$

We can test whether these effects are statistically different from zero (or from each other) by forming appropriate hypotheses, e.g.

$$H_0 : \beta_k = 0; \quad \text{or} \quad H_0 : \beta_1 = \beta_2$$

Questions?

Job Training

- JTRAIN98 contains data on labor market outcomes for men, some of whom enrolled in a job training program in 1997.
- Data:
  - Whether or not individual took training in 1997: $train = 1$ if training taken
  - Earnings in 1998 (in \$1000s): *earn*98
  - Earnings in 1996 (in \$1000s): *earn*96
  - Whether or not individual is married: $married = 1$ if married
  - Education in years of schooling: *educ*
  - Age in years: *age*

Job Training

- Regression output meauring differences in labor market outcomes for those with and without training:

```
. reg earn98 train
```

Source	SS	df	MS	Number of obs	=	1,130
				F(1, 1128)	=	17.91
Model	1054.41369	1	1054.41369	Prob > F	=	0.0000
Residual	66408.4778	1,128	58.872764	R-squared	=	0.0156
				Adj R-squared	=	0.0148
Total	67462.8915	1,129	59.754554	Root MSE	=	7.6729

earn98	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
train	-2.050053	.4844142	-4.23	0.000	-3.000507	-1.099599
_cons	10.6099	.279429	37.97	0.000	10.06164	11.15816

Covariate Balance Test

- One way random assignment is justified statistically is by a "balance test".
- Idea: See if individuals with particular covariates are more likely to select into treatment.
- Practically: Regress treatment dummy D_i on observable characteristics x_i:

$$D_i = \gamma_0 + x_i'\gamma_1 + \epsilon_i.$$

- Conduct $F-$test of $H_0 : \gamma_1 = 0$ vs. $H_1 : \gamma_1 \neq 0$.

Covariate Balance Test

- Test of overall significance rejects with p-value ≈ 0:

```
. reg train age educ earn96 married
```

Source	SS	df	MS		
Model	62.3769648	4	15.5942412	Number of obs	= 1,130
Residual	188.511531	1,125	.167565805	F(4, 1125)	= 93.06
				Prob > F	= 0.0000
Total	250.888496	1,129	.222221874	R-squared	= 0.2486
				Adj R-squared	= 0.2460
				Root MSE	= .40935

train	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0050343	.0012841	-3.92	0.000	-.0075539	-.0025147
educ	-.0123111	.0043717	-2.82	0.005	-.0208886	-.0037336
earn96	-.019028	.0011428	-16.65	0.000	-.0212703	-.0167858
married	-.1240124	.0289627	-4.28	0.000	-.1808395	-.0671854
_cons	.9464928	.0731975	12.93	0.000	.8028737	1.090112

Issues with Balance Tests

- If the significance level is 5%, we will reject H_0 5% of the time even when $\gamma_1 = 0$.
  - This means calling into question validity of 5% of experimental results.
- If the balance test fails to reject, it doesn't mean randomization holds. D_i may just be independent of observed covariates, leading to $\gamma_1 = 0$.
- What is really needed is

$$D_i \perp y_i(0), y_i(1).$$

Selection on Observables

- Since individuals choose whether to take training, it is unreasonable to assume they are randomly assigned.
  - Previous labor market outcomes, age and work experience all likely to impact selection into training.
  - In other words: $y_i(1), y_i(0)$ likely correlated with D_i.
- Treatment assignment is unconfounded if, conditional on observed covariates x_i, treatment randomly assigned:

$$D_i \perp y_i(0), y_i(1) \,|\, x_i.$$

Selection on Observables: Example

Selection on Observables: Example

Selection on Observables

- Idea: If we fix prior labor market experience, age, etc. participation is as good as random.
- Implementation: Add in covariates to regression model to control for selection based on observable characteristics.
- Unconfounded treatment assignment implies:

$$\begin{aligned} \mathrm{E}\left(y\left(1\right)|D,x\right) &= \mathrm{E}\left(y\left(1\right)|x\right); \\ \mathrm{E}\left(y\left(0\right)|D,x\right) &= \mathrm{E}\left(y\left(0\right)|x\right). \end{aligned}$$

- We can estimate

$$\mathrm{E}\left(y\left(1\right)|D=1,x\right); \quad \mathrm{E}\left(y\left(0\right)|D=0,x\right).$$

Selection on Observables

- Now we can estimate the ATE by aggregating over x:

Selection on Observables

- We approximated conditional means with a linear model

$$\mathrm{E}\left(Y_i|D_i, x_i\right) \approx \beta_0 + \beta_1 D_i + x_i'\beta_2.$$

  This approximation gives

$$\mathrm{E}\left(Y|D=1, x\right) - \mathrm{E}\left(Y|D=0, x\right) \approx \beta_1.$$

  This implies that $\beta_1 = \mathrm{E}\left(y\left(1\right) - y\left(0\right)\right)$ (which is the ATE).
- Summary: Controlling for x allowed us to identify the ATE because treatment assignment is random conditional on x.

Selection on Observables

- The approximation works if:

$$\begin{aligned}\mathrm{E}\left(y\left(0\right)|x\right) &= \alpha_0 + x'\beta_2 \\ \mathrm{E}\left(y\left(1\right)|x\right) &= \alpha_1 + x'\beta_2\end{aligned}$$

so that by the LIE:

$$ATE = \mathrm{E}\left(y\left(1\right) - y\left(0\right)\right) = \alpha_1 - \alpha_0.$$

It follows by unconfoundedness that:

$$\begin{aligned}\mathrm{E}\left(Y|D,x\right) &= \mathrm{E}\left(y\left(0\right) + D\left(y\left(1\right) - y\left(0\right)\right)|D,x\right) \\ &= \underbrace{\alpha_0}_{\beta_0} + x'\beta_2 + D\underbrace{\left(\alpha_1 - \alpha_0\right)}_{\beta_1} \\ &= \beta_0 + \beta_1 D + x'\beta_2.\end{aligned}$$

Job Training

- Regression controlling for observables greatly alters treatment effect estimate:

```
. reg earn98 train earn96 educ age married
```

Source	SS	df	MS		
Model	27320.1797	5	5464.03593	Number of obs	= 1,130
Residual	40142.7118	1,124	35.7141564	F(5, 1124)	= 152.99
				Prob > F	= 0.0000
Total	67462.8915	1,129	59.754554	R-squared	= 0.4050
				Adj R-squared	= 0.4023
				Root MSE	= 5.9761

earn98	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
train	2.410547	.4352625	5.54	0.000	1.556528	3.264565
earn96	.3725384	.0186262	20.00	0.000	.3359923	.4090845
educ	.3628329	.064047	5.67	0.000	.2371678	.4884979
age	-.181046	.018875	-9.59	0.000	-.2180803	-.1440118
married	2.481719	.4262625	5.82	0.000	1.64536	3.318079
_cons	4.667042	1.145283	4.08	0.000	2.419908	6.914176

Interaction Terms

- Previously we assumed

$$\begin{aligned}\mathrm{E}\left(y\left(0\right)|x\right) &= \alpha_0 + x'\beta_2 \\ \mathrm{E}\left(y\left(1\right)|x\right) &= \alpha_1 + x'\beta_2\end{aligned}$$

  which implied, under unconfoundedness

$$\begin{aligned}\mathrm{E}\left(Y|D,x\right) &= \mathrm{E}\left(y\left(0\right) + D\left(y\left(1\right) - y\left(0\right)\right)|D,x\right) \\ &= \alpha_0 + \left(\alpha_1 - \alpha_0\right)D + x'\beta_2\end{aligned}$$

Interaction Terms

- May think that potential outcomes should have different β:

$$\begin{aligned}\mathrm{E}\left(y\left(0\right)|x\right) &= \alpha_0 + x'\beta^0;\\ \mathrm{E}\left(y\left(1\right)|x\right) &= \alpha_1 + x'\beta^1.\end{aligned}$$

- The ATE (τ) is therefore written as:

$$\begin{aligned}\mathrm{E}\left(y\left(1\right) - y\left(0\right)\right) &= \mathrm{E}\left(\mathrm{E}\left(y\left(1\right)|x\right) - \mathrm{E}\left(y\left(0\right)|x\right)\right)\\ &= \mathrm{E}\left[\left(\alpha_1 - \alpha_0\right) + x'\left(\beta^1 - \beta^0\right)\right].\end{aligned}$$

Interaction Terms

- Under unconfoundedness:

$$\mathrm{E}\left(y\left(0\right)|D=0,x\right)=\alpha_0+x'\beta^0;$$
$$\mathrm{E}\left(y\left(1\right)|D=1,x\right)=\alpha_1+x'\beta^1.$$

- Then natural to estimate $\mathrm{E}\left(y\left(1\right)-y\left(0\right)\right)$ as

$$\hat{\tau}=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}_i\left(1\right)-\hat{y}_i\left(0\right)\right]=\hat{\alpha}_1-\hat{\alpha}_0+\bar{x}_n'\left(\hat{\beta}^1-\hat{\beta}^0\right).$$

Interaction Terms

- Having different slopes changes linear regression model:

$$\begin{aligned}\mathrm{E}\left(Y|D,x\right) &= \mathrm{E}\left(y\left(0\right)+D\left(y\left(1\right)-y\left(0\right)\right)|D,x\right)\\ &= \alpha_0+\left(\alpha_1-\alpha_0\right)D+x'\beta^0+Dx'\left(\beta^1-\beta^0\right)\end{aligned}$$

- Estimating this model with the entire sample is equivalent to esimating the two regressions separately.

The Overlap Condition

- Identified ATE as

$$ATE = \mathrm{E}(y_1 - y_0) = \mathrm{E}\left[\mathrm{E}(Y|D=1,x) - \mathrm{E}(Y|D=0,x)\right].$$

- Implicitly assumed overlap condition: $0 < \mathrm{P}(D=1|x=x') < 1$ for all x'.
- If for some x', $\mathrm{P}(D=1|x') = 0$, then there are no observations of Y for which $D=1$ and $x=x'$, so $\mathrm{E}(Y|D=1,x=x')$ is undefined.
- In practice, with several covariates, many pairs (d,x') will have very few observations (curse of dimensionality).

Non-parametric estimates of the ATE

- If x is discrete, we can use non-parametric estimate which averages Y for each (D, x) pair:

$$\begin{aligned}\mathrm{E}\left(Y|D=1, x=x'\right) &= \frac{\mathrm{E}\left(Y\mathbf{1}\left(D=1, x=x'\right)\right)}{\mathrm{P}\left(D=1, x=x'\right)}, \\ &\approx \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{1}\left(D=1, x=x'\right)}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(D=1, x=x'\right)} := \hat{\mu}_1\left(x'\right).\end{aligned}$$

  where

$$\mathbf{1}(A) = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

- Estimate $\mathrm{E}\left(Y|D=0, x=x'\right) \approx \hat{\mu}_0\left(x'\right)$ analogously. Then

$$\begin{aligned}\widehat{ATE} &:= \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_1\left(x_i\right) - \hat{\mu}_0\left(x_i\right) \\ &\approx \mathrm{E}\left[\mathrm{E}\left(Y|D=1, x\right) - \mathrm{E}\left(Y|D=0, x\right)\right].\end{aligned}$$

Non-parametric estimates of the ATE

- Suffers from curse of dimensionality if x is high-dimensional.
- If x is continuous, can use a non-parametric kernel estimator (compute average by including observations in a range of values for x, weighting $x \approx x'$ higher):

$$\hat{\mu}_1\left(x'\right) := \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{1}\left(D = 1, x \in [x' - h, x' + h]\right)}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(D = 1, x \in [x' - h, x' + h]\right)}$$

- h is called a "bandwidth" and $h \to 0$ as $n \to \infty$ to ensure consistency.

Matching

- Matching provides an alternate method of estimating ATE based on unconfoundedness.
- Idea: Match unit with $D = 0, x = x'$ to one with $D = 1, x = x'$.
- For x continuous, can match units with $x_i \approx x_j$ and construct a proxy for $y_i(d)$ using, e.g. k nearest neighbors:

$$\hat{y}_i(d) = \begin{cases} y_i & \text{if } D = d, \\ \frac{1}{k} \sum_{j \in \mathcal{M}_i} y_j & \text{if } D \neq d, \end{cases}$$

where $\mathcal{M}_i$ is the set of points "closest" to i.

Propensity Score Matching

- Asserting $\mathrm{E}\left(y\left(d\right)|x\right) = \alpha_d + x'\beta_d$ alleviates curse of dimensionality (approximate conditional means with linear regression to avoid the "small-bin" problem).
- Another way to reduce the dimension of the set of conditioning variables is propensity score matching.
- Rosenbaum, Rubin (1983):

$$D_i \perp y_i\left(0\right), y_i\left(1\right)|x_i \implies D_i \perp y_i\left(0\right), y_i\left(1\right)|p\left(x_i\right),$$

where

$$p\left(x'\right) = \mathrm{P}\left(D = 1|x = x'\right).$$

is the propensity score.

Propensity Score Matching

- Adding/removing covariates can violate unconfoundedness, but this works because:

$$\begin{aligned}
\mathrm{P}\left(D=1|y_0,y_1,p\left(x\right)\right) &= \mathrm{E}\left(\mathrm{P}\left(D=1|y_0,y_1,p\left(x\right),x\right)|y_0,y_1,p\left(x\right)\right) \\
&= \mathrm{E}\left(\mathrm{P}\left(D=1|y_0,y_1,x\right)|y_0,y_1,p\left(x\right)\right) \\
&= \mathrm{E}\left(\mathrm{P}\left(D=1|x\right)|y_0,y_1,p\left(x\right)\right) \\
&= \mathrm{E}\left(p\left(x\right)|y_0,y_1,p\left(x\right)\right) \\
&= p\left(x\right).
\end{aligned}$$

- After conditioning on $p\left(x\right)$, D is independent of $\left(y_0,y_1\right)$.
- Identification of ATE via propensity score:

$$ATE = \mathrm{E}\left[\mathrm{E}\left(Y|D=1,p\left(x\right)\right)-\mathrm{E}\left(Y|D=0,p\left(x\right)\right)\right]$$

Propensity Score Matching

- Avoid curse of dimensionality because $p(x)$ is scalar even if x high dimensional.
- However, p still needs to be estimated, so curse of dimensionality now features in estimation of p.
- Exception: In a randomized experiment, $p(x)$ is known (by design).
- Suggests estimate:

$$\mathrm{E}\left(Y|D=1,p(x)=p'\right) \approx \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{1}\left(D=1,p(x)=p'\right)}{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(D=1,p(x)=p'\right)}.$$

Propensity Score Matching

- Hirano, Imbens and Ridder (2003) use representation of ATE as unconditional weighted average:

$$ATE = \mathrm{E}\left(\frac{Y\left(D - p\left(x\right)\right)}{p\left(x\right)\left(1 - p\left(x\right)\right)}\right),$$

  estimated by sample analog principle:

$$\widehat{ATE} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i\left(D_i - \hat{p}\left(x_i\right)\right)}{\hat{p}\left(x_i\right)\left(1 - \hat{p}\left(x_i\right)\right)}.$$

- Show that (under regularity conditions) this estimate is asymptotically normal and efficient.

Propensity Score Matching

Propensity Score Matching

Questions?