Matching

Tadao Hoshino （星野匡郎）

ver. 2018 Fall Semester

Review of the Previous Lecture

Review of the Previous Lecture

- Let $T_i \in \{0, 1\}$ be a treatment variable, and Y_i be an observed outcome variable such that

$$Y_i = T_i \cdot Y_{1i} + (1 - T_i) \cdot Y_{0i},$$

  where Y_{ti} is a potential outcome when $T_i = t$.
- A naive estimator of the average treatmentl effect (ATE) $E[Y_{1i} - Y_{0i}]$ is simply to take the difference between the average outcome for the treatment and control groups:

$$\frac{\sum_{i=1}^{n} T_i Y_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} (1 - T_i) Y_i}{\sum_{i=1}^{n} (1 - T_i)}$$

- However, this estimator is biased for the causal effect when the selection bias is not zero:

$$E[Y_{0i} | T_i = 1] - E[Y_{0i} | T_i = 0] \neq 0.$$

Review of the Previous Lecture

- In randomized experiments, the treatment and control groups are randomly formed such that

$$(Y_{0i}, Y_{1i}) \perp\!\!\!\perp T_i. \tag{1}$$

- Since the selection bias is zero under (1), the simple difference estimator is consistent for the ATE (and also ATET) in randomized experiments.
- However, randomized experiments are often too expensive (financially and politically) and unethical.

Review of the Previous Lecture

- How can we estimate causal effects when randomized experiments are infeasible?
  $=>$ Assume "conditional independence" instead of the full independence (1).
- Under the conditional independence assumption, we can estimate ATET by using the so-called Matching method.

Conditional Independence

Conditional Independence

- Consider a job training program example where we are interested in estimating the causal effect of the training on earnings.
- Each individual chooses whether to participate in the program or not (i.e., non-random treatment assignment).
- Let T_i denote the program assignment, which takes the value 1 if individual i participated in the program and 0 otherwise.
- Each individual has two potential earnings: Y_{i1} for $T_i = 1$ and Y_{0i} for $T_i = 0$.
- In the following, we focus on the treatment effects for the training group, namely

$$ATET : E\left[Y_{1i} - Y_{0i} \mid T_i = 1\right].$$

Conditional Independence

- Individuals who anticipate larger gains from the training would be more willing to participate in the program.
  $=>$ The selection bias may exist.
- What types of people are willing to participate in the program?
  - Young workers with shorter work experience;
    Let X_{1i} be the work experience (years) of i.
  - Workers who are not satisfied with their current salary;
    Let X_{2i} be the pre-training salary (JPY) of i.
  - etc
- Since X_{1i} and X_{2i} are likely correlated not only with T_i but also with (Y_{0i}, Y_{1i}), the full independence (▸ full independence) may not hold.

Conditional Independence

- Alternatively, we assume the following condition: (Y_{1i}, Y_{0i}) and T_i are independent conditional on (X_{1i}, X_{2i}), namely

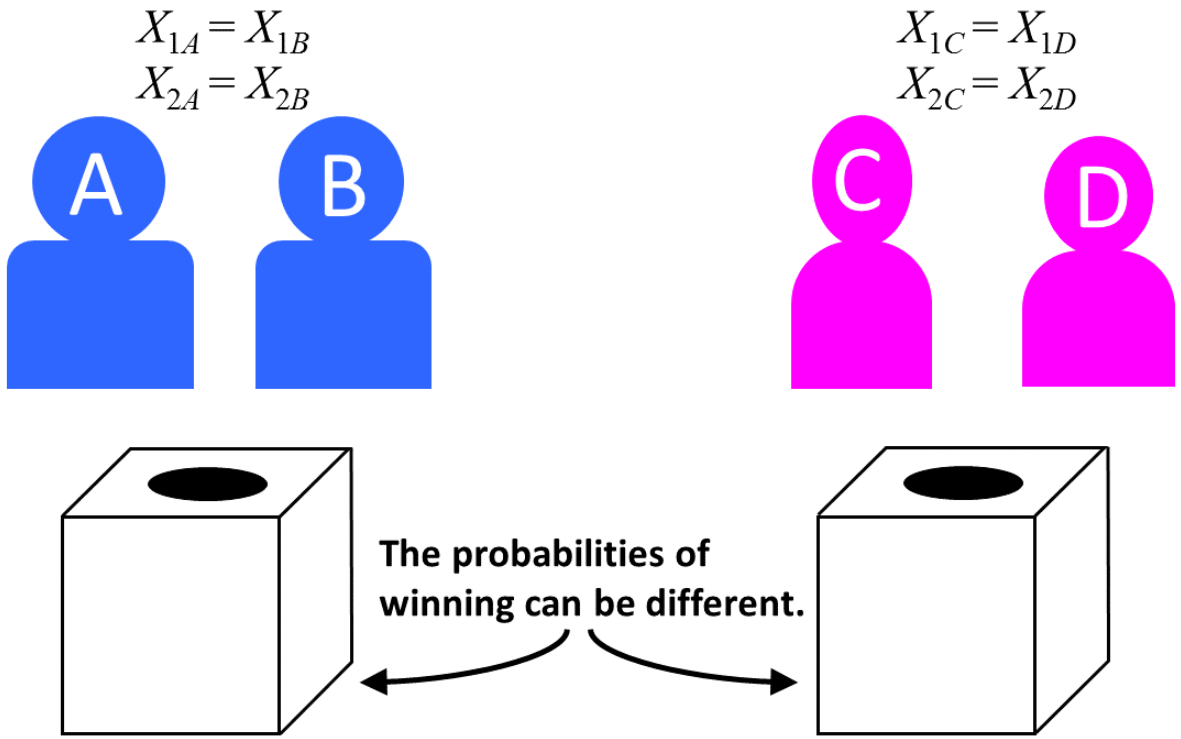
$$(Y_{0i}, Y_{1i}) \perp\!\!\!\perp T_i | X_{1i}, X_{2,i} \tag{2}$$

  This condition (2) is called conditional independence assumption (CIA), or unconfoundedness.
- The CIA (2) is interpreted as that among individuals with the same value of (X_{1i}, X_{2i}) program participation is independent of potential outcomes.
- In the sense that the treatment assignment probability can vary with (X_{1i}, X_{2i}), CIA is a relatively weak condition.

$$\text{Full independence} \Rightarrow \text{Conditional independence}$$
$$\text{Conditional independence} \not\Rightarrow \text{Full independence}$$

Conditional Independence

Under CIA (2), the treatments are assigned as if a randomized experiment was conducted within each group defined by the value of (X_{1i}, X_{2i}).



Conditional Independence

- If CIA (2) holds, since T_i and Y_{0i} are independent given (X_{1i}, X_{2i}),

$$E\left[Y_{0i}|T_i = 1, X_{1i}, X_{2i}\right] = E\left[Y_{0i}|T_i = 0, X_{1i}, X_{2i}\right]$$

  holds.
- This equality implies that the selection bias is zero conditional on the covariates (X_{1i}, X_{2i}). Hence, we have

$$\begin{aligned}
&E\left[Y_i|T_i = 1, X_{1i}, X_{2i}\right] - E\left[Y_i|T_i = 0, X_{1i}, X_{2i}\right] \\
&= E\left[Y_{1i}|T_i = 1, X_{1i}, X_{2i}\right] - E\left[Y_{0i}|T_i = 0, X_{1i}, X_{2i}\right] \\
&= E\left[Y_{1i} - Y_{0i}|T_i = 1, X_{1i}, X_{2i}\right] + \underbrace{E\left[Y_{0i}|T_i = 1, X_{1i}, X_{2i}\right] - E\left[Y_{0i}|T_i = 0, X_{1i}, X_{2i}\right]}_{\text{selection bias}=0} \\
&= E\left[Y_{1i} - Y_{0i}|T_i = 1, X_{1i}, X_{2i}\right].
\end{aligned} \tag{3}$$

Conditional Independence

- Furthermore, by the law of iterated expectations,

$$E\left\{E\left[Y_{1i}-Y_{0i}|T_i=1,X_{1i},X_{2i}\right]|T_i=1\right\}=E\left[Y_{1i}-Y_{0i}|T_i=1\right]=\textbf{ATET}$$
$$E\left\{E\left[Y_i|T_i=1,X_{1i},X_{2i}\right]|T_i=1\right\}=E\left\{Y_i|T_i=1\right\}$$

- Hence,

$$\begin{aligned}\textbf{ATET}&=E\left\{E\left[Y_{1i}-Y_{0i}|T_i=1,X_{1i},X_{2i}\right]|T_i=1\right\}\\&=E\left\{E\left[Y_i|T_i=1,X_{1i},X_{2i}\right]-E\left[Y_i|T_i=0,X_{1i},X_{2i}\right]|T_i=1\right\}\\&=E\left\{Y_i-E\left[Y_i|T_i=0,X_{1i},X_{2i}\right]|T_i=1\right\},\end{aligned}\tag{4}$$

where the second equality follows from (3).

Conditional Independence

- Equation (4):

$$\underbrace{\textbf{ATET}}_{E[Y_{1i}-Y_{0i}|T_i=1]} = E\left\{Y_i - E\left[Y_i|T_i=0, X_{1i}, X_{2i}\right]|T_i=1\right\}$$

  implies that nuder CIA the causal effect ATET can be estimated by the right hand side of the above.

- In order to estimate ATET, we need to estimate

$$E\left[Y_i|T_i=0, X_{1i}, X_{2i}\right].$$

  When $T_i = 1$ is true, this term is interpreted as the predicted value of the unobserved potential outcome Y_{0i}.

- How can we estimate this term? $=>$ Matching method

Matching Method

Exact Matching

- Equation (4):

$$\textbf{ATET} = E\left\{Y_i - E\left[Y_i | T_i = 0, X_{1i}, X_{2i}\right] | T_i = 1\right\}$$

- If individual i is in the control group ($T_i = 0$), we can simply use his realized outcome Y_i for $E\left[Y_i | T_i = 0, X_{1i}, X_{2i}\right]$.
- If individual i is in the treatment group ($T_i = 1$), we find an individual, say j, who
  - belongs to the control group ($T_j = 0$), and
  - satisfies $X_{1i} = X_{1j}$ and $X_{2i} = X_{2j}$ (matching).
- Then, we substitute j's outcome Y_j for $E\left[Y_i | T_i = 0, X_{1i}, X_{2i}\right]$.

Exact Matching

- That is, define $\hat{Y}_{0i}$ as follows:

$$\hat{Y}_{0i} = \begin{cases} Y_i & \text{if } T_i = 0 \\ Y_j & \text{if } T_i = 1, T_j = 0, X_{1j} = X_{1i}, X_{2j} = X_{2i} \end{cases} \tag{5}$$

  Then, $\hat{Y}_{0i}$ can serve as an approximate value of $E\left[Y_i | T_i = 0, X_{1i}, X_{2i}\right]$.
- Using $\hat{Y}_{0i}$ in place of $E\left[Y_i | T_i = 0, X_{1i}, X_{2i}\right]$, we have

$$ATET \approx E\left[Y_i - \hat{Y}_{0i} | T_i = 1\right].$$

- Then, the exact matching estimator of ATET is

$$\widehat{ATET} = \frac{\sum_{i=1}^{n} T_i (Y_i - \hat{Y}_{0i})}{\sum_{i=1}^{n} T_i}. \tag{6}$$

Exact Matching

$T = 1$ (Trained)			
ID	Salary aftr training	Exp-erience	Salary bfr training
1	240,000	5	220,000
2	360,000	14	350,000
3	250,000	7	230,000
4	200,000	1	180,000
5	220,000	2	200,000
6	350,000	11	350,000

$T = 0$ (Not trained)			
ID	Salary aftr training	Exp-erience	Salary bfr training
7	190,000	1	180,000
8	240,000	6	240,000
9	220,000	5	220,000
10	320,000	12	310,000
11	230,000	5	230,000
12	310,000	9	300,000

$$\hat{Y}_{01} = 220{,}000$$
$$\hat{Y}_{04} = 190{,}000$$

Nearest Neighbor Matching

- One problem for implementing the exact matching is that when matching is done on the covariates (X_{1i}, X_{2i}), it is possible to end up with a very small available sample size. (Unmatched observations have to be excluded from the analysis.)
- As we add more covariates X_3, X_4, ..., exact matching becomes far more difficult.
- More critically, if some of the covariates are continuous, exact matching is theoretically infeasible (regardless of the dimension of the covariates) since $\Pr(X_{1j} = X_{1i}) = 0$ for $j \neq i$ when X_1 is continuous.

Nearest Neighbor Matching

- Even when exact matching is not feasible, we can still match each treated individual to the "nearest" untreated individual in terms of the value of (X_1, X_2)
  $=>$ Nearest-neighbor matching
- To do so, we need to define how to measure the distance between any two individuals.
- For example, consider Manhattan distance and Euclidean distance:

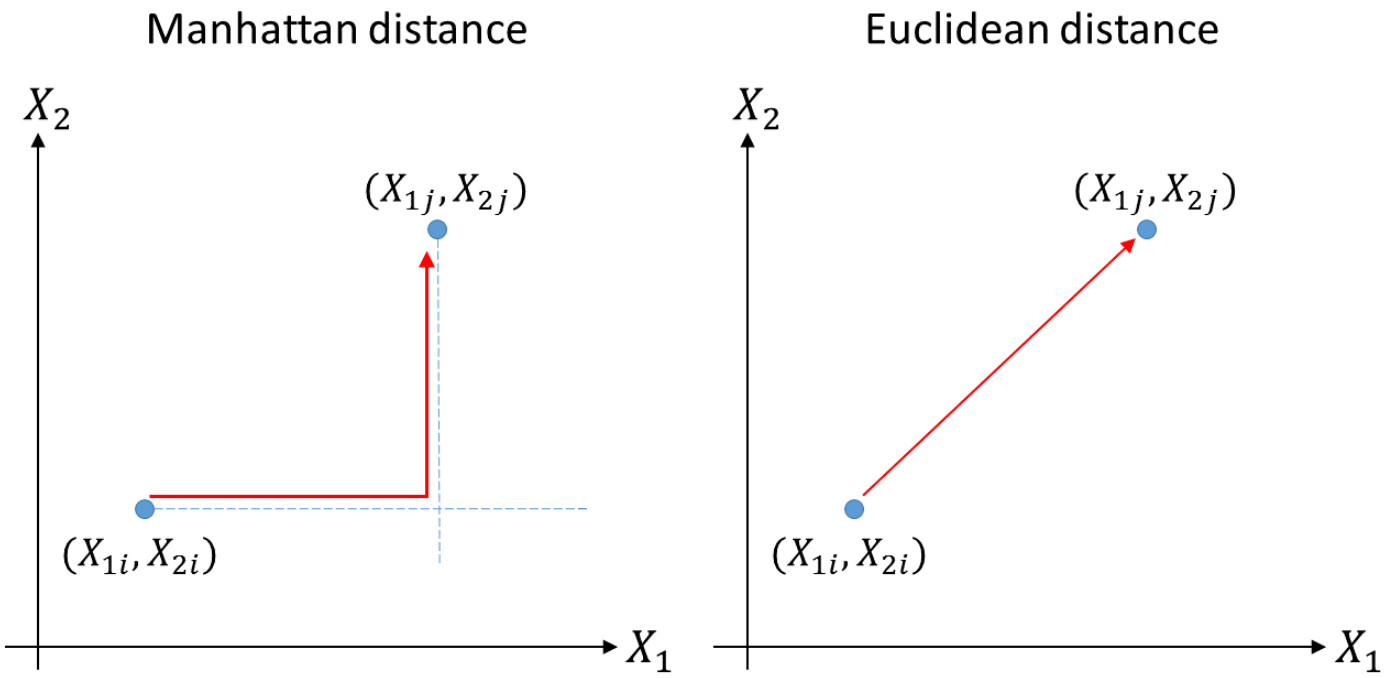
$$\text{Manhattan distance} \quad d(i,j) = \left|X_{1i} - X_{1j}\right| + \left|X_{2i} - X_{2j}\right|$$

$$\text{Euclidean distance} \quad d(i,j) = \sqrt{\left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2}$$

- If these distances are sufficiently small, individuals i and j are alike enough in their characteristics?

Nearest Neighbor Matching

- A drawback of Manhattan or Euclidean distance is that their value is heavily dependent on the unit of measurement.
  - Height (m): $X_{1i} - X_{1j} = 0.2$, Weight (g): $X_{2i} - X_{2j} = 1,000$.
- Some scale normalization is needed to balance out the contributions of each covariate.



Nearest Neighbor Matching

- Standardized Euclidean distance :

$$d(i,j) = \sqrt{\Delta_{1ij}^2 + \Delta_{2ij}^2}$$

- Mahalanobis distance :

$$d(i,j) = \sqrt{\frac{1}{1-\rho^2}\left[\Delta_{1ij}^2 + \Delta_{2ij}^2 - 2\rho\Delta_{1ij}\Delta_{2ij}\right]}$$

where

$$\Delta_{1ij} = \frac{X_{1i} - X_{1j}}{\sigma_1}, \quad \Delta_{2ij} = \frac{X_{2i} - X_{2j}}{\sigma_2},$$

σ_1 and σ_2 the standard deviations of X_1 and X_2, respectively, and ρ is the correlation coefficient between X_1 and X_2.

Nearest Neighbor Matching

Mahalanobis distance

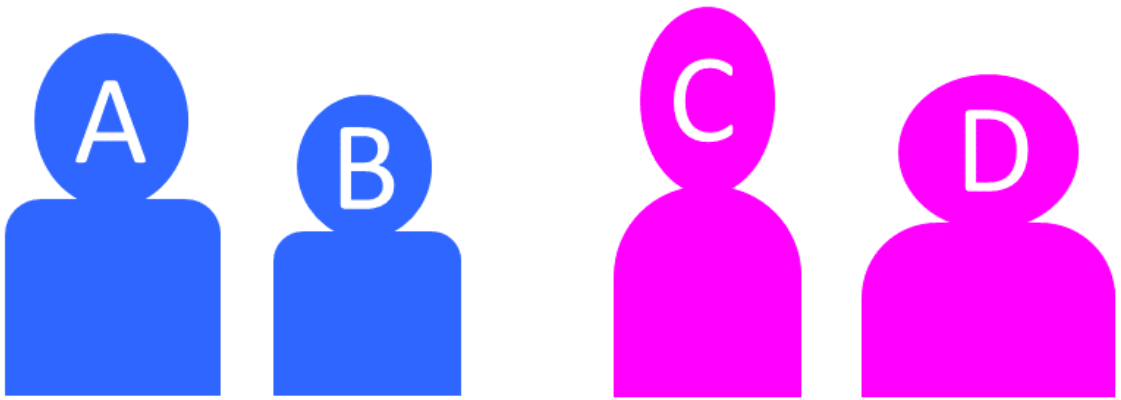
- The Mahalanobis distance is a generalization of the standardized Euclidean distance in that it accounts for the correlation between the covariates.
  - When $\rho = 0$, the Mahalanobis distance reduces to the standardized Euclidean distance.
- Write the squared Mahalanobis distance between i and j as

$$d^2(i,j) = \frac{\Delta_{1ij}^2 + \Delta_{2ij}^2}{1-\rho^2} + \frac{-2\rho\Delta_{1ij}\Delta_{2ij}}{1-\rho^2}$$

  - The first term on the rhs is proportional to the squared standardized Euclidean distance.
  - Suppose that $\rho > 0$ (e.g., X_1: height, X_2: weight). The second term increases as i and j become more dissimilar such that $\Delta_{1ij}\Delta_{2ij} < 0$.

Nearest Neighbor Matching

$$X_{1A} - X_{1B} = 1 \qquad X_{1C} - X_{1D} = 1$$
$$X_{2A} - X_{2B} = 1 \qquad X_{2C} - X_{2D} = -1$$



Euclidean	$d_E(A,B) = 1.414$	$d_E(C,D) = 1.414$
Mahalanobis	$d_M(A,B) = 1.118$	$d_M(C,D) = 2.236$

$$\sigma_1 = \sigma_2 = 1$$
$$\rho = 0.6$$

Nearest Neighbor Matching

- Suppose that individual i is treated, and let j^* be the most "similar" untreated individual to i such that $d(i,j)$ is minimized at $j = j^*$ over all untreated j's .
- As an estimator for $E[Y_i|T_i = 0, X_{1i}, X_{2i}]$ we can use Y_{j^*}:

$$\hat{Y}_{0i} = \begin{cases} Y_i & \text{if } T_i = 0 \\ Y_{j^*} & \text{if } T_i = 1 \end{cases}$$

- Then, we can estimate ATET following (6).
- The exact matching is a special case of the nearest neighbor matching where j is matched to i only when $d(i,j) = 0$ is satisfied.

Nearest Neighbor Matching

- K-nearest neighbor matching: Alternatively to the 1-to-1 nearest neighbor matching, match each treated individual to the K most similar untreated individuals.
- The nearest neighbor matching is a special case of K-nearest neighbor matching where $K = 1$.
- Specifically, letting $j(m)$ be i's m-th nearest untreated individual:

$$d(i, j(1)) \leq d(i, j(2)) \leq \cdots \leq d(i, j(m)) \leq d(i, j(m+1)) \leq \cdots,$$

we define $\hat{Y}_{0i}$ as

$$\hat{Y}_{0i} = \begin{cases} Y_i & \text{if } T_i = 0 \\ \frac{1}{K} \sum_{m=1}^{K} Y_{j(m)} & \text{if } T_i = 1. \end{cases}$$

- Then, estimate ATET by (6).

A remark on the conditional independence (2)

- The conditioning variables in the CIA are all observables.
- When there are unobservable factors affecting both the treatment status and the potential outcomes, we cannot identify the treatment effects by matching methods.
- This problem does not occur in randomized experiments because regardless of their observability all the factors that can influence treatment and outcome are randomized.
- Even when a randomized experiment is infeasible and some unobservable confounders may exist, the treatment effects can still be estimated by using the method of instrumental variables.

An Empirical Example:
Causal effects of the ESA, Ferraro et al. (2007)

Causal effects of the ESA, Ferraro et al. (2007)

- Conservation of biodiversity is a global concern.
  - Ethical reason: every species has some intrinsic (existence) value.
  - Economic reason: we can obtain many medicinal and industrial products from plants and animals.
  - Prevention of epidemics and natural disasters, etc.
- The Endangered Species Act (ESA) is the most important biodiversity legislation in the United States.
- However, its effectiveness has been debated by many researchers. Previous studies that have attempted to evaluate the effectiveness of the ESA showed conflicting results.

Causal effects of the ESA, Ferraro et al. (2007)

- Under the ESA, the government must list species as "endangered" or "threatened". The act gives strong legal tools to the government for the protection of listed species, including regulations on land use.
- Many endangered species occur on private lands.
- The ESA can impose perverse incentives on private landowners to undertake preemptive actions to harm species and their habitat in order to avoid the regulatory burdens.
  - For example, Lueck and Michael (2003)[1] found that landowners whose lands were near the habitat of the endangered red-cockaded woodpecker prematurely harvested trees more often than other landowners.

[1] Lueck, D., and Michael, J. A. (2003) Preemptive habitat destruction under the Endangered Species Act. The Journal of Law and Economics, 46(1), 27-60.

Causal effects of the ESA, Ferraro et al. (2007)

- More rigorous examination of the effectiveness of the ESA is needed.
- Ferraro et al. (2007)[2] used matching methods to estimate the causal impacts of the ESA on species recovery.
- For each species i, the outcome variable Y_i is "change in endangerment status from 1993 to 2004", which is defined based on the *NatureServe Conservation Status*[3].
- A positive (resp. negative) value of Y_i means that the species i is in a recovery (resp. deteriorating) state.

---

[2]Ferraro, P. J., McIntosh, C., and Ospina, M. (2007) The effectiveness of the US endangered species act: An econometric analysis using matching methods. Journal of Environmental Economics and Management, 54(3), 245-261.

[3]NatureServe is a non-profit conservation organization whose mission is to provide the scientific basis for effective conservation action. `http://www.natureserve.org/`

Causal effects of the ESA, Ferraro et al. (2007)

- As for the treatment variable T_i, Ferraro et al. (2007) considered the following three treatments:

  (1) being listed under the ESA
  (2) being listed and receiving "substantial" federal and state funds for recovery
  (3) being listed but not receiving "substantial" federal and state funds for recovery

- The control group is the set of species not listed under the ESA.
- If the ESA listing process is random with respect to the endangerment risk of species, we can estimate the effects of the ESA by simply taking the difference between the treatment and control groups.
- However, such assumption is unrealistic.

Causal effects of the ESA, Ferraro et al. (2007)

- Ferraro et al. (2007) considered the conditional independence

$$(Y_{0i}, Y_{1i}) \perp\!\!\!\perp T_i | X_{1i}, ..., X_{ki},$$

  where the set of conditioning covariates include, for example, the number of journal citations to each species, taxonomic dummies, dummy for being a monotype.
- Then, using matching methods, Ferraro et al. (2007) estimated the ATET of each treatment.

Causal effects of the ESA, Ferraro et al. (2007)

ATET estimates : extracted from Ferraro et al. (2007) Table 3

	ATET		
Treatment T	(1)	(2)	(3)
Nearest-neighbor : Standardized Euclidean	-0.0191 (0.839)	0.4537 (0.001)	-0.2128 (0.027)
Nearest-neighbor : Mahalanobis	-0.0189 (0.823)	0.4091 (0.001)	-0.1806 (0.047)
sample size	430	329	396

(Standard errors in the parentheses.)

Recall:

(1) being listed under the ESA
(2) being listed and receiving "substantial" federal and state funds for recovery
(3) being listed but not receiving "substantial" federal and state funds for recovery

Causal effects of the ESA, Ferraro et al. (2007)

- The estimated ATET of listing alone (1) is negative but not statistically significant, and that of listing with small funding (2) is also negative significantly.
- In contrast, the effect of listing with substantial funding (3) is statistically significantly positive.

> Listing a species under the ESA can be harmful to species recovery if not combined with substantial government funds.

- The level of expenditure is a more essential factor in recovering species, rather than merely listing endangered species.

Glossary I