Lecture 1: Propensity Score Matching Method

Chung Choe
Department of Economics
Konkuk University
choechung@konkuk.ac.kr

Korea Small Business Institute
December 14, 2020 (2:00–3:30PM)

Motivation

- Both **person A** (Mr. Smart) and **person B** (Mr. Stupid) work for KOSBI in Seoul. Suppose person A took a job training program and person B didn't. (KOSBI didn't encourage A to take the training program and didn't discourage B not to take it.) After participating in the program, whereas B earns **€1,000**/mth, A earns **€1,200**/mth.
- In case policy makers (KOSBI CEO) would be interested in evaluating the job-training effectiveness, can we conclude that the job-training increased A's earning by **€200**?
- Why is this an incorrect measure of the training effects? Person A is **NOT** comparable to person B: age, education, language skills, etc.

Motivation

- Now assume that person B has the **same covariates** such as gender, education, age, etc.
  The **€200** would be a correct measure?
  Maybe not!!!
- What if person A is smart – real IQ that we cannot measure is 120 – and person B is not smart (IQ=80) and IQ is **positively** correlated with wages and participation in training?
  Over- or underestimate?
  The correct answer should be '**overestimate**'.
- Why?

Motivation

- What does make things difficult to estimate the **true treatment effect** (job-training effectiveness)?
- Basically we don't know about the **counterfactual** wage in case that person A does NOT take the job-training. In other words, we don't observe both wages **with** and **without** training for the person A.
- A similar story can be also applied when other treatments are of our interests such as returns to college education, impact of migration on wages, etc.

Unfamiliar Terminologies

- Potential/Counterfactual outcome
- Treatment/Control
- Unconfoundedness/Conditional independence assumption (CIA)
- Selection on observables
- Average Treatment Effect (ATE)
- Average Treatment Effect on the Treated (ATT)
- Propensity score
- Common support condition
- Balancing properties

Potential Outcomes Model

For units $i = 1, \ldots, n$,
$\mathbf{T_i} = \mathbf{0}, \mathbf{1}$: treatment (treatment might be 'received financial aid', 'went to college' or 'participated in job training')
$\mathbf{Y_i}(\mathbf{0})$: potential outcome under control
$\mathbf{Y_i}(\mathbf{1})$: potential outcome under treatment
$\mathbf{Y}_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$: observed outcome

The treatment effect for unit i is

$$\tau_i = Y_i(1) - Y_i(0)$$

In an experimental setting, the **average treatment effect (ATE)** for this population is

$$\tau = \mathbf{E}(\mathbf{Y_i}(\mathbf{1})) - \mathbf{E}(\mathbf{Y_i}(\mathbf{0}))$$

In words, **ATE** is the mean over the whole population of the expected outcome under treatment **less** the expected outcome under the alternative treatment (control).

Potential Outcomes Model

If treatment is **randomly** assigned, then it should be independent of potential outcomes ($T_i \perp (Y_i(1), Y_i(0))$). Then

$$\mathbf{E}(\mathbf{Y_i}(\mathbf{1}) | \mathbf{T_i} = \mathbf{1}) = \mathbf{E}(\mathbf{Y_i}(\mathbf{1})),$$

$$\mathbf{E}(\mathbf{Y_i}(\mathbf{0}) | \mathbf{T_i} = \mathbf{0}) = \mathbf{E}(\mathbf{Y_i}(\mathbf{0})),$$

and $\tau = E(Y_i(1) | T_i = 1) - E(Y_i(0) | T_i = 0) =$
$E(Y_i | T_i = 1) - E(Y_i | T_i = 0)$,
which can be estimated as follows:

$$\widehat{\tau_n} = E(\widehat{Y_i | T_i = 1}) - E(\widehat{Y_i | T_i = 0})$$

$$= \frac{\sum_{i=1}^{n} Y_i 1(T_i = 1)}{\sum_{i=1}^{n} 1(T_i = 1)} - \frac{\sum_{i=1}^{n} Y_i 1(T_i = 0)}{\sum_{i=1}^{n} 1(T_i = 0)}$$

A Simple Numerical Example (Job Training Effects)

Observation	Treatment	Outcome (wage)	Education	Gender
1	1	1100	CG	M
2	0	900	HG	F
3	1	900	HG	F
4	0	1000	CG	M
5	0	800	HG	F

$$\widehat{\tau}_n = \frac{\sum_{i=1}^{n} Y_i 1(T_i = 1)}{\sum_{i=1}^{n} 1(T_i = 1)} - \frac{\sum_{i=1}^{n} Y_i 1(T_i = 0)}{\sum_{i=1}^{n} 1(T_i = 0)}$$

$$= \frac{(1100 + 900)}{2} - \frac{(900 + 1000 + 800)}{3}$$

$$= 100$$

Nonrandom Assignment of Treatment

If the treatment randomly assigned, the ATE is identified which is '**€100**'. In many economic studies, however, we usually come across nonrandom situations.
In that sense, the estimator we used above does **not** consistently estimate **ATE**, $\tau = E(Y_i(1)) - E(Y_i(0))$, because

$$E(Y_i(1)|T_i = 1) \neq E(Y_i(1)),$$

$$E(Y_i(0)|T_i = 0) \neq E(Y_i(0)).$$

i.e. the potential outcomes of **self-selected** college graduates will **NOT** be equal to the ones of **randomly selected**.

Are there other assumptions that would be sufficient to identify the treatment effect?

Unconfoundedness/Conditional Independence Assumption

$$\mathbf{T_i} \perp (\mathbf{Y_i}(\mathbf{1}), \mathbf{Y_i}(\mathbf{0})) \mid \mathbf{X_i}$$

This says that T_i is independent of the potential outcomes conditional on X_i.

In an observational study, **CIA** means that T_i can be said to be "as good as randomly assigned", conditional on X_i.

Behavioral implication of this assumption is that, after controlling for the variation in outcomes induced by differences in X_i, participation in the treatment program does not depend on outcomes.

Unconfoundedness/Conditional Independence Assumption

Under unconfoundedness assumption:

$$(\mathbf{T_i} \perp (\mathbf{Y_i}(\mathbf{1}), \mathbf{Y_i}(\mathbf{0})) \mid \mathbf{X_i}),$$

we obtain

$$E(Y_i(1) \mid X_i, T_i = 0) = E(Y_i(1) \mid X_i, T_i = 1) = E(Y_i \mid X_i, \ T_i = 1)$$

i.e. If IQ (not ability) is controlled, the outcomes (= observed outcomes) of college graduates is the potential outcomes of high-school graduates.
What does that mean by 'IQ is controlled'?

Why is this concept called '**selection on observables**' not '**selection on unobservables**'?

Understanding unconfoundedness with a numerical example

Observation	Treatment	Outcome (wage)	Education	Gender
1	1	1100	CG	M
2	0	900	HG	F
3	1	900	HG	F
4	0	1000	CG	M
5	0	800	HG	F

1. See the first observation. T_1 is equal to 1 and the outcome corresponding to the treatment ($T_1 = 1$) is 1,100. We only observe the outcome when the first obs is treated.
2. We need to estimate the counterfactual outcome – $Y_1(0)$. In fact, one of the three observations (2, 4, or 5) in the data can be used to predict the potential outcome.
3. By the CIA assumption, self-selection bias can be removed. Controlling for X_1 (college graduate, male), $T_1$is independent of the potential outcomes ($Y_i(1), Y_i(0)$). Then the outcome of the 4th obs could be a good match to predict $Y_1(0)$.

Identification under Unconfoundedness

Subgroup Average Effect is the average effect for individuals with covariate value x:

$$ATE(x) = E(Y_i(1) \mid X_i = x) - E(Y_i(0) \mid X_i = x)$$

$$= E(Y_i | X_i = x,\ T_i = 1) - E(Y_i | X_i = x,\ T_i = 0).$$

Suppose $X_i = x$ is a discrete variable. Consider the estimator:

$$\widehat{\tau}(x) = \frac{\sum_{i=1}^{n} T_i 1(X_i = x) Y_i}{\sum_{i=1}^{n} T_i 1(X_i = x)} - \frac{\sum_{i=1}^{n} (1 - T_i) 1(X_i = x) Y_i}{\sum_{i=1}^{n} (1 - T_i) 1(X_i = x)}.$$

We are simply taking the treatment and control averages for the subsample with X_i equal to a particular value. Then take sample analog to the equation $\mathbf{ATE} = \mathbf{E}[\mathbf{ATE}(\mathbf{x})]$:

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\tau}(X_i).$$

Identification under Unconfoundedness

The estimator of **Subgroup Average Effect**:

$$\widehat{\tau}(x) = \frac{\sum_{i=1}^{n} T_i 1(X_i = x) Y_i}{\sum_{i=1}^{n} T_i 1(X_i = x)} - \frac{\sum_{i=1}^{n} (1 - T_i) 1(X_i = x) Y_i}{\sum_{i=1}^{n} (1 - T_i) 1(X_i = x)}.$$

For each individual i we estimate $\widehat{\tau}(x)$, which can be estimated using one of the **nonparametric techniques** such as series estimator, series logit estimator, nearest neighbor estimator, and kernel estimator.

Note that regardless whether an individual i is treated or untreated, we must estimate both potential outcomes for treated and untreated, except the **nearest neighbor estimator**.

Average Treatment Effect on the Treated (ATT)

In many cases, it's more interesting to estimate the effects of policy or training for those **treated**. This leads us to an identification of the **average treatment effect for the treated population (ATT)**, which is a bit different concept from **ATE**:

$$\tau_{T=1} = E\left(Y_i(1) \mid T_i = 1\right) - E\left(Y_i(0) \mid T_i = 1\right).$$

This expression cannot be estimated directly, because $Y_i(0)$ is not observed for treated units. Then can we simply estimate ATT with the observed difference in outcomes,
$E\left(Y_i \mid T_i = 1\right) - E\left(Y_i \mid T_i = 0\right)$?
NO!!!

$$\underbrace{E\left(Y_i \mid T_i = 1\right) - E\left(Y_i \mid T_i = 0\right)}_{\textit{Observed difference in outcome}} = \underbrace{E\left(Y_i(1) - Y_i(0) \mid T_i = 1\right)}_{\textit{Average treatment effect on the treated}}$$

$$+ \underbrace{E\left(Y_i(0) \mid T_i = 1\right) - E\left(Y_i(0) \mid T_i = 0\right)}_{\textbf{Selection bias}}$$

Average Treatment Effect on the Treated (ATT)

Assuming unconfoundedness, however, we obtain

$$E\left(Y_i\left(0\right) \mid X_i, T_i = 1\right) = E\left(Y_i\left(0\right) \mid X_i, T_i = 0\right) = E\left(Y_i \mid X_i,\ T_i = 0\right).$$

This allows us to identify ATT,

$$\tau_{T=1} = E\left\{E\left(Y_i \mid X_i, T_i = 1\right) - E\left(Y_i \mid X_i, T_i = 0\right) \mid T_i = 1\right\}$$

where the outer expectation is over the distribution of $X_i | T_i = 1$.

In words, ATT is the mean over the whole population **who get treatment** of the expected outcome under treatment **less** the expected outcome under the alternative treatment (control).

Break

Let's have 5 min break!

Review of Notation and Terminology

$\mathbf{Y_i(0)}$: potential outcome under control

$\mathbf{Y_i(1)}$: potential outcome under treatment

$\mathbf{Y_i}= T_i Y_i(1) - (1 - T_i) Y_i(0) =$ observed outcome

Average treatment effect (ATE) for the population is:

$$\tau = E(Y_i(1)) - E(Y_i(0)).$$

Subgroup Average Effect is the average effect for individuals with covariate value x:

$$ATE(x) = E(Y_i(1) \mid X_i = x) - E(Y_i(0) \mid X_i = x)$$

$$= E(Y_i|X_i = x,\ T_i = 1) - E(Y_i|X_i = x,\ T_i = 0).$$

Then the ATE is represented as: $\mathbf{ATE} = \mathbf{E}[\mathbf{ATE}(\mathbf{x})]$.

Average treatment effect for the treated population (ATT) is:

$$\tau_{T=1} = E(Y_i(1) \mid T_i = 1) - E(Y_i(0) \mid T_i = 1).$$

Assuming **unconfoundedness**, we identify ATT as

$$\tau_{T=1} = E\{E(Y_i|\ X_i, T_i = 1) - E(Y_i|\ X_i, T_i = 0) \mid T_i = 1\}.$$

Relationship between POM and Linear Regression Model

What does the parameter of the dummy variable in a typical linear regression function represent – is it either **ATE** or **ATT**? Suppose **Conditional Expectation Function (CEF)** is linear as:

$$E\left(Y_i|\ X_i, T_i\right) = \beta_1 + T_i\beta_2 + X_i'\beta_3 + \left(T_iX_i\right)'\beta_4.$$

This implies:

$$E\left(Y_i|\ X_i, T_i = 0\right) = \beta_1 + X_i'\beta_3,$$

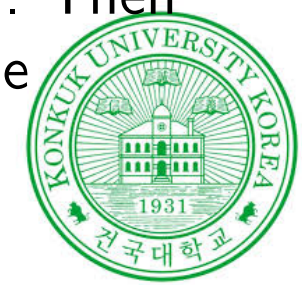
$$E\left(Y_i|\ X_i, T_i = 1\right) = \left(\beta_1 + \beta_2\right) + X_i'\left(\beta_3 + \beta_4\right).$$

We could then estimate this regression function by OLS, and then estimate $\tau(x)$ by:

$\widehat{\tau}(x) = E\left(Y|\widehat{X = x}, T = 1\right) - E\left(Y|\widehat{X = x}, T = 0\right) = \widehat{\beta}_2 + x'\widehat{\beta}_4.$

Note that we would have different $ATE(x)$, depending on x. Then the estimate of the overall average treatment effect would be

$$\widehat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\beta}_2 + X_i'\widehat{\beta}_4\right) = \widehat{\beta}_2 + \overline{X}'\widehat{\beta}_4.$$



Relationship between POM and Linear Regression Model

If we run the regression with $(X_i - \overline{X})$,

$$E(Y_i|\ X_i, T_i) = \beta_1 + T_i\beta_2 + (X_i - \overline{X})'\beta_3 + T_i(X_i - \overline{X})'\beta_4.$$

Then we will have

$$\widehat{\tau} = \widehat{\beta}_2.$$

How do we estimate ATT?

$$\widehat{\tau}_{T=1} = \frac{1}{\sum_{i=1}^{n} 1(T_i = 1)} \sum_{i=1}^{n} 1(T_i = 1) \cdot \left(\widehat{\beta}_2 + X_i'\widehat{\beta}_4\right).$$

Note that if $\overline{X} = \overline{X}_{T=1}$, then ATE would be equal to ATT.

Relationship between POM and Linear Regression Model

Any intuition why...

if $\overline{X} = \overline{X}_{T=1}$, then ATE would be equal to ATT.

In the observational studies, treatment is random conditional on covariates (X).

In other words, the treatment effects of individuals are determined by the covariates.

Therefore, if $\overline{X} = \overline{X}_{T=1}$, average treatment effects would be equivalent to average treatment effects on the treated.

Note also that $\overline{X} = \overline{X}_{T=1}$ means $\overline{X}_{T=1} = \overline{X}_{T=0}$ ...

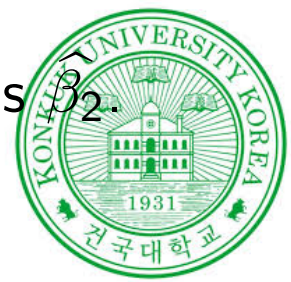
Relationship between POM and Linear Regression Model

In typical regression analysis including dummy variables, we do NOT include the interactions between T and X as below.

$$E(Y_i|\ X_i, T_i) = \beta_1 + T_i\beta_2 + X_i'\beta_3.$$

What does this say?
It implicitly implies that there is no difference in the coefficients of X between treated and control (1). We can interpret $\widehat{\beta_2}$ as an estimate of ATE under a restricted model.

What about ATT?
Here we assume that Treatment Effect is invariant across individuals (2) so that ATE should be equal to ATT, which is β_2.



The Propensity Score

Unfortunately, we run into problems if the covariate vector X is high-dimensional (curse of dimensionality). Unless the sample size is huge, it's hard to find two observations that are really "close" to each other along every dimension of X.

Theorem

Theorem 1 (Rosenbaum and Rubin 1983): Let $p(X_i)$ *be the probability of unit* i *having been assigned to treatment, defined as* $p(X_i) \equiv \Pr(T_i = 1 \mid X_i) = \Pr(T_i \mid X_i)$. *Suppose that* $T_i \perp (Y_i(1), Y_i(0)) \mid X_i$ *and* $0 < Pr(T_i = 1 \mid X_i) < 1$ *for all* X_i. *Then*

$$T_i \perp (Y_i(1), Y_i(0)) \mid p(X_i).$$

The Propensity Score

Corollary

If $T_i \perp (Y_i(1), Y_i(0)) \mid X_i$ *and assumptions of Theorem 1 hold, then*

$$\tau_{T=1} = E\{E(Y_i|p(X_i), T_i = 1) - E(Y_i|p(X_i), T_i = 0) \mid T_i = 1\},$$

assuming that the expectations are defined. The outer expectation is over the distribution of $p(X_i)|T_i = 1$.

The earlier version for ATT without P-score is:

$$\tau_{T=1} = E\{E(Y_i|\, X_i, T_i = 1) - E(Y_i|\, X_i, T_i = 0) \mid T_i = 1\}.$$

Estimation Strategy of ATT using P-Score Matching

When identifying $\tau_{\mathbf{T=1}}$, $E\left\{E\left(Y_i|p\left(X_i\right), T_i = 1\right) \mid T_i = 1\right\}$ is easily estimated, but **counterfactual outcomes** need to be estimated for $E\left\{E\left(Y_i|p\left(X_i\right), T_i = 0\right) \mid T_i = 1\right\}$.

Let $\mathbf{s_0}\left(\mathbf{p}\right) = \mathbf{E}\left(\mathbf{Y_i}\left(\mathbf{0}\right)|\mathbf{p}\left(\mathbf{X_i}\right) = \mathbf{p}\right)$. By Theorem 1,

$$s_0\left(p\right) = E\left(Y_i|T_i = 0, p\left(X_i\right) = p\right).$$

The idea is to compare individuals who based on observables have a very similar probability of receiving treatment (similar propensity score). We estimate potential outcomes of the treatment group by matching with the outcomes of control group, based on the **propensity score**.

Then we can form the following estimator of the ATT:

$$\widehat{\tau}_{T=1} = \frac{\sum_{i=1}^{n} Y_i 1\left(T_i = 1\right)}{\sum_{i=1}^{n} 1\left(T_i = 1\right)} - \frac{\sum_{i=1}^{n} \widehat{s_0}\left(p\left(X_i\right)\right) 1\left(T_i = 1\right)}{\sum_{i=1}^{n} 1\left(T_i = 1\right)}$$

Estimation Strategy of ATT using P-Score Matching

Think about the **single-nearest neighbor estimator** of $s_0(p)$.
Then this amounts to finding the untreated individual with propensity score **closest** to treated individual i.
Now we are "matching" individuals based on the scalar propensity score, instead of the multidimensional variable X_i.

Then how can we estimate ATE?
Let $\mathbf{s_1(p) = E(Y_i(1) | p(X_i) = p)}$. Again, by Theorem 1,

$$s_1(p) = E(Y_i | T_i = 1, p(X_i) = p).$$

Hence, we can form the following estimator of the **Average Treatment Effect**:

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} [\widehat{s_1}(p(X_i)) - \widehat{s_0}(p(X_i))]$$

Estimating Treatment Effects using P-Score Matching

Average Treatment Effect on the Treated (ATT):

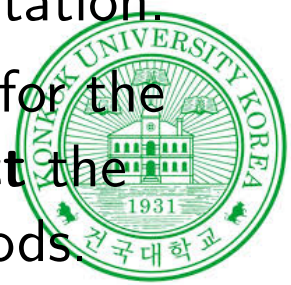
$$\widehat{\tau}_{T=1} = \frac{\sum_{i=1}^{n} Y_i 1(T_i = 1)}{\sum_{i=1}^{n} 1(T_i = 1)} - \frac{\sum_{i=1}^{n} \widehat{s}_0(p(X_i)) 1(T_i = 1)}{\sum_{i=1}^{n} 1(T_i = 1)}$$

Average Treatment Effect (ATE):

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} [\widehat{s}_1(p(X_i)) - \widehat{s}_0(p(X_i))]$$

Note the **difference** in estimation between ATE and ATT.
Note also that

1. We do not mention about **standard error** in this presentation.
2. An observation in the untreated (as well as the treated for the case of ATE) can be used more than **once to construct** the counterfactual outcomes, depending on matching methods.



Covariates for the estimation of the propensity score

In practice, what variables should be included to estimate propensity score?

1. Basically, the covariates should be **predetermined**. If your covariates are affected by the treatment, they will be bad controls. Bad controls are variables that are themselves outcome variables in the notional experiment at hand. That is, **bad controls** might just as well be dependent variables too. **Good controls** are variables that we can think of as having been fixed at the time the regressor of interest was determined.
2. The variables in the estimation of propensity score are supposed to partly determine both **outcome** and **treatment**. [INTUITION: 1) There is no need to include variables unrelated to the treatment variable - those will have no explanatory power. 2) If some variables are uncorrelated with the outcome variable, they are not useful to resolve selection issue even if included.]

Covariates for the estimation of the propensity score

We need to include variables to satisfy **CIA** conditions. Say we'd like to estimate the **return to college education**. **CIA** implies that potential wages are **uncorrelated** with the decision to go to college, if variables to proxy motivations (or innate ability) are controlled.

Covariates for the estimation of the propensity score

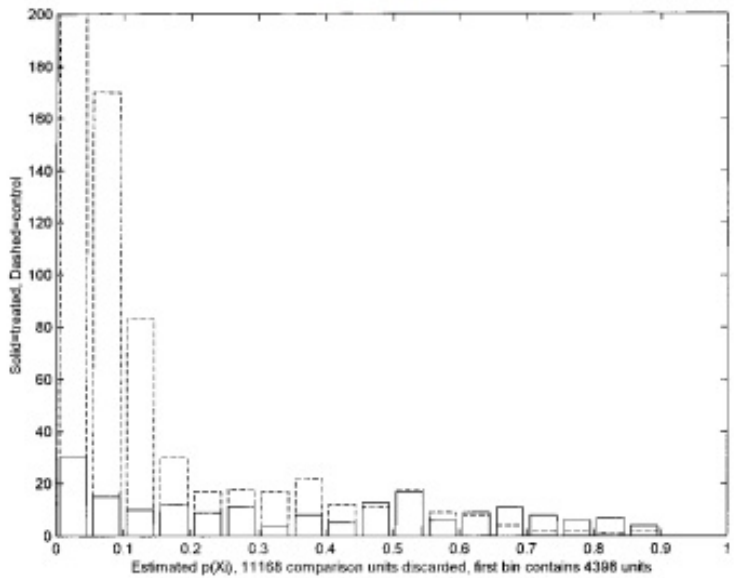
- What do you mean by the controls of motivations? Under the same level of motivations, it's unlikely that there exists correlation between **entrance of college** (T_i) and **wages** (potential outcome: $Y_i(T_i)$).
- Note that X's which influence both **outcomes** and **treatment** can be included. On the other hand, the information - **distance to college** – cannot be used to satisfy CIA, because the distance does not directly influence the outcomes.
- Then would it be okay to include distance along with proxy for motivations? No, propensity score matching method removes bias from selection on **observables**, not unobservables. If you still think that unobservables determine both treatment and outcome (even after conditioning on a rich set of observables), we may have to apply different methods, such as **IV methods** (2SLS).

Common Support Condition

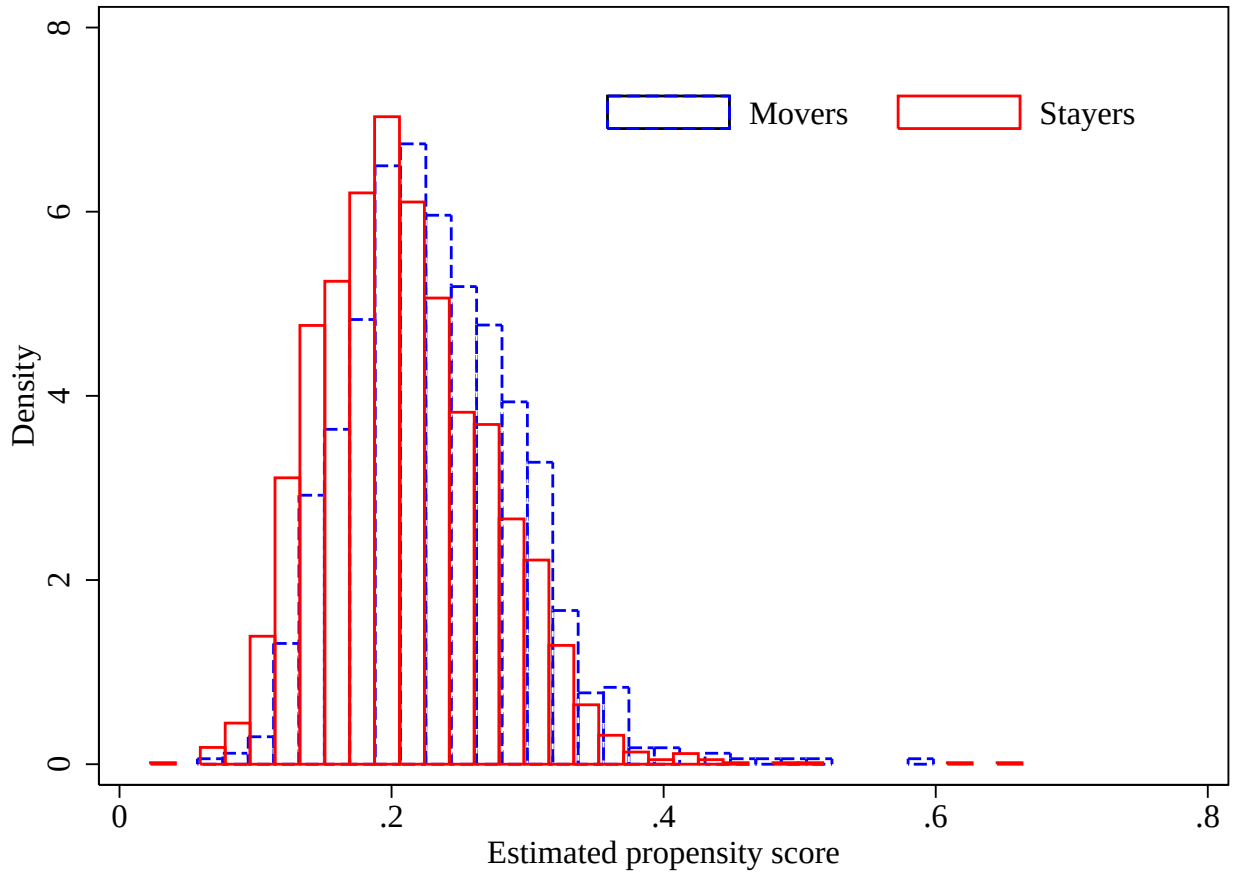
While CIA **cannot** be tested (its plausibility all depends on our argument), we have a condition and a property to be tested.

Common support condition checks the existence of the p-score overlap between treated and untreated, by that we obtain potential outcomes for the untreated. For the binary case, we can implement this by graphical illustration.

FIGURE 1.—HISTOGRAM OF ESTIMATED PROPENSITY SCORE, NSW AND CPS



Common Support Condition: migrants vs. stayers using GOMS



Balancing properties

In order to reduce the dimensionality of covariate vector X, **Rosenbaum and Rubin (2002)** suggest the propensity score approach (Theorem 1).

Checking **balancing properties** is to double-check if this replacement works. That is, **conditional on the propensity score**, the covariates are independent of assignment to treatment, as in a randomized experiment.

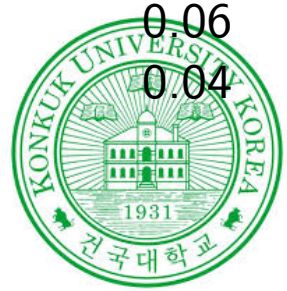
For the binary case, we do this as follows:

1. Split the sample into **k** equally spaced intervals of the propensity score (strata), in most cases we divide into 5.
2. Within each interval, test for statistically significant differences between the distribution of covariates for treated and comparison units.

Balancing Property Test

Table: Balancing property tests

	Difference	Paired t statistics	95% Conf	Interval
Age when starting job 2	0.08	0.66	-0.15	0.30
Age squared/100	0.04	0.68	-0.08	0.16
Junior College	-0.01	-0.47	-0.05	0.03
Female	-0.04	-1.72	-0.09	0.01
Married	0.02	1.14	-0.01	0.05
Log of wage on job 1	-0.00	-0.02	-0.04	0.04
Tenure of job 1 (years)	0.01	0.19	-0.10	0.12
Professional occ on job 1	0.01	0.72	-0.02	0.05
Father w/ college degree	0.02	0.75	-0.02	0.06
Mother w/ college degree	0.01	0.93	-0.02	0.04



Sensitivity Analysis

Ichino et al. (2008) allows assessment of the sensitivity of ATT matching estimates.

1. They derive point estimates of the ATT under different possible scenarios of deviation from unconfoundedness.
   - To do so they impose values of the parameters that characterize the distribution of U.
   - Given these parameters, the value of the confounding factor for each treated and control subject is predicted and the ATT is reestimated now including the influence of the simulated U.
   - By changing the assumptions about the distribution of U, they can assess the robustness of the ATT with respect to different hypotheses on the nature of the confounding factor.
2. Their approach also allows one to verify whether there exists a set of plausible assumptions on U under which the estimated ATT would be driven to zero by the inclusion of U.
   - By modelling the nature of U based on already existing variables, it is possible to assess the robustness of the estimates with respect to deviations from unconfoundedness that would occur if observed factors were omitted from the matching set.

Some Drawbacks of Propensity Score Matching

1. It requires two steps-that is, **matching** and **averaging**.
2. Thus, estimating **standard errors** of the resulting estimates may not be straightforward, either.
3. A third consideration is that the **two-way contrast** does not always do full justice to the problem at hand. If T_i can take on more than two values such as "**years of education**", there are separate average casual effects for each possible increment in treatment, which also must summarized in some way.

These considerations lead us back to **regression**.

Why Propensity Score Matching?

Do you have an example when we are able to maximize the benefits of application of p-score approaches? Is it always better than ordinary regressions?
There are a couple of things of which we can take advantage from the application of the p-score matching approach.

1. Robust to the **specification errors**.
2. **Nonlinearity**: provide a more flexible specification of the relationship between the covariates and the outcome.
3. **Heterogeneity**: doesn't have to assume a constant treatment effect across individuals.
4. By imposing the **common support condition**, we can drop observations for which there are no comparable individuals across different treatment levels.

Whether it's because of the reasons above or not, it's been shown that the propensity score matching relatively better to replicate experimental results **Dehejia and Wahba (2002)**.

Final Notes

- Propensity score matching methods are NOT supposed to resolve the endogeneity problem.
- They are simply econometric methods which assume selection on observables.
- In case you are suspicious of endogeneity, you may want to look for instrumental variables or attempt to apply other econometric models such as fixed effects estimator.
- To see how to implement P-score matching in STATA, please refer Becker and Ichino (2002).
- For more practical guidance, please refer Caliendo and Kopeinig (2008).

Bibiography

Becker, S. O. and Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal*, 2(4):358–377.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22:31–72.

Dehejia, R. H. and Wahba, S. (2002). Propensity Score-Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics*, 84(1):151–161.

Ichino, A., Mealli, F., and Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, 23(3):305–327.

Rosenbaum, P. R. and Rubin, D. B. (2002). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.