Quantitative Methods in Economics
Causality and treatment effects

Maximilian Kasy

Harvard University, fall 2016
6) Instrumental Variables Estimation of Treatment Effects

(cf. “Mostly Harmless Econometrics,” chapter 4)
Motivation: Experiments with Imperfect Compliance

- Recall the JTPA experiment, where many experimental subjects did not comply with the randomized assignment:

<table>
<thead>
<tr>
<th></th>
<th>Enrolled in Training</th>
<th>Not Enrolled in Training</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned to Training</td>
<td>4804</td>
<td>2683</td>
<td>7487</td>
</tr>
<tr>
<td>Assigned to Control</td>
<td>54</td>
<td>3663</td>
<td>3717</td>
</tr>
<tr>
<td>Total</td>
<td>4858</td>
<td>6346</td>
<td>11204</td>
</tr>
</tbody>
</table>

- Units receiving training may differ from units that do not receive training.

- Still, randomized assignment has an effect on the probability of receiving training.

- Instrumental variables use the variation in receipt of training induced by the experiment to obtain an estimate of the effect of training.
Instrumental Variables

Instrumental variable models:

1. Traditional Econometric Framework:

\[ Y = \mu + \alpha D + X' \beta + u \]

\( \text{cov}(D, u) \neq 0 \)

\( \text{cov}(Z, D) \neq 0 \) and \( \text{cov}(Z, u) = 0 \)

⇒ Use 2SLS to estimate \( \alpha \). Without \( X \)'s:

\[ \alpha = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} \]

2. Potential Outcomes Model of Instrumental Variables

- Heterogeneous treatment effects
- Estimate Local Average Treatment Effect (LATE)
Randomized Experiments with Imperfect Compliance

Assignment

\[ Z = \begin{cases} 
1 & \text{if assigned to treatment group} \\
0 & \text{if assigned to control group} 
\end{cases} \]

Potential Treatments

- \( D_1 \): treatment status if assigned to treatment group
- \( D_0 \): treatment status if assigned to control group

Observed Treatment

\[ D = \begin{cases} 
D_1 & \text{if } Z = 1 \\
D_0 & \text{if } Z = 0 
\end{cases} \]

or, in a more compact notation: \( D = ZD_1 + (1 - Z)D_0 \).
Randomized Experiments with Imperfect Compliance

- Angrist, Imbens and Rubin (1996) define:
  - **Compliers**: $D_1 > D_0$ ($D_0 = 0$ and $D_1 = 1$)
  - **Always-takers**: $D_1 = D_0 = 1$
  - **Never-takers**: $D_1 = D_0 = 0$
  - **Defiers**: $D_1 < D_0$ ($D_0 = 1$ and $D_1 = 0$)

- Notice that for compliers, we still have a perfect experiment.

- However, only one of the potential treatment indicators, $(D_0, D_1)$, is observed, so we cannot identify which group any particular individual belongs to.
Identification with Instrumental Variables

Identification Assumptions

1. Independence: \((Y_0, Y_1, D_0, D_1) \perp \perp Z\)
2. First Stage: \(0 < P(Z = 1) < 1\) and \(P(D_1 = 1) \neq P(D_0 = 1)\)
3. Monotonicity: \(D_1 \geq D_0\)

Identification Result

Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996) prove:

\[
E[Y_1 - Y_0 | D_1 > D_0] = \frac{E[Y | Z = 1] - E[Y | Z = 0]}{E[D | Z = 1] - E[D | Z = 0]} \left( \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} \right)
\]
Identification with Instrumental Variables

Proof:

\[
\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]}
\]

\[
= \frac{E[Y_1D_1 + Y_0(1 - D_1)|Z = 1] - E[Y_1D_0 + Y_0(1 - D_0)|Z = 0]}{E[D_1|Z = 1] - E[D_0|Z = 0]}
\]

\[
= \frac{E[Y_1D_1 + Y_0(1 - D_1)] - E[Y_1D_0 + Y_0(1 - D_0)]}{E[D_1] - E[D_0]}
\]

\[
= \frac{E[(Y_1 - Y_0)(D_1 - D_0)]}{E[D_1 - D_0]}
\]

\[
= \frac{E[Y_1 - Y_0|D_1 - D_0 = 1]}{\Pr(D_1 - D_0 = 1)} \Pr(D_1 - D_0 = 1)
\]

\[
= E[Y_1 - Y_0|D_1 > D_0]
\]
LATE: Local Average Treatment Effect

\[ \alpha_{LATE} = E[Y_1 - Y_0| D_1 > D_0], \] the average treatment effect for compliers is often called Local Average Treatment Effect (LATE).

- Average effect of the treatment for the units affected in their treatment status by changes in the instrument.
- This parameter is different for different instruments, \( Z \).
- Whether LATE is interesting or not depends on the instrument.
Special Cases

- If $D$ is randomized, then $Z = D$ and everybody is a complier.
- One-sided noncompliance, $D_0 = 0$, then:

$$E[Y_1 - Y_0 | D_1 > D_0] = E[Y_1 - Y_0 | D_1 = 1]$$

$$= E[Y_1 - Y_0 | Z = 1, D_1 = 1]$$

$$= E[Y_1 - Y_0 | D = 1].$$

$$\Rightarrow \alpha_{LATE} = \alpha_{ATE}$$
Identification Assumptions

- **Independence:** $(Y_0, Y_1, D_0, D_1) \perp\!
\perp Z$
  - Implies that the instrument $Z$ is “as good as randomly assigned”
  - $Y_d$ implies an exclusion restriction: $Z$ has no direct effect on $Y_d$
  - $Z$ can only affect $Y$ through its effect on $D$

- **First Stage:** $0 < P(Z = 1) < 1$ and $P(D_1 = 1) \neq P(D_0 = 1)$
  - Implies that the instrument $Z$ induces variation in $D$
  - Testable by regressing $D$ on $Z$

- **Monotonicity:** $D_1 \geq D_0$
  - Rules out defiers
  - Often can be assessed from institutional knowledge
Instrumental Variable: Estimators

**LATE**

\[
E[Y_1 - Y_0 | D_1 > D_0] = \frac{E[Y | Z = 1] - E[Y | Z = 0]}{E[D | Z = 1] - E[D | Z = 0]} \left( = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} \right)
\]

**Wald Estimator**

The sample analog estimator is:

\[
\left( \frac{\sum_{i=1}^{N} Y_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} Y_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right) \bigg/ \left( \frac{\sum_{i=1}^{N} D_i Z_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} D_i (1 - Z_i)}{\sum_{i=1}^{N} (1 - Z_i)} \right)
\]
Instrumental Variable: Estimators

**LATE**

\[
E[Y_1 - Y_0|D_1 > D_0] = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} \left( = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} \right)
\]

**Wald Estimator as IV Regression**

Can also implement Wald Estimator using an IV regression:

\[
Y = \mu + \alpha D + u
\]

where \(\text{cov}(Z, u) = 0\), so \(\alpha = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} = \alpha_{LATE}\)

⇒ To estimate \(\alpha\) we run a simple IV regression (2SLS) of \(Y\) on a constant and \(D\), using \(Z\) as an instrument for \(D\).
Treatment effects

Instrumental Variable: Estimators

LATE

\[ E[Y_1 - Y_0|D_1 > D_0] = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} \left( = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} \right) \]

Two Stage Least Squares

If identification assumptions only hold after conditioning on \( X \), covariates are often introduced using 2SLS regression:

\[ Y = \mu + \alpha D + X' \beta + u, \]

where \( Z \) and \( u \) are uncorrelated. Now \( \alpha \) and \( \beta \) are computed regressing \( Y \) on \( D \) and \( X \), and using \( Z \) and \( X \) as instruments.

In general, \( \alpha \) estimated in this way does not have a clear causal interpretation unless we impose additional strong assumptions (Abadie, 2003).
Example: Effects of JTPA Training on Earnings

<table>
<thead>
<tr>
<th></th>
<th>Comparisons by Training Status</th>
<th>Comparisons by Assignment Status</th>
<th>Instrumental Variables Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Men</td>
<td>3970 (555)</td>
<td>1117 (569)</td>
<td>1825 (928)</td>
</tr>
<tr>
<td>B. Women</td>
<td>2133 (345)</td>
<td>1243 (359)</td>
<td>1942 (560)</td>
</tr>
</tbody>
</table>
Example: The Vietnam Draft Lottery (Angrist, 1990)

- Effect of military service on civilian earnings.
- Simple comparisons between Vietnam veterans and non-veterans are likely to be a biased measure.
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam.
- Draft eligibility is random and affected the probability of enrollment.
- Results suggest a negative effect of veteran status on earnings, but the estimates are quite imprecise.

<table>
<thead>
<tr>
<th>Year</th>
<th>Draft-Eligibility Effects in Current $</th>
<th>Service Effect in 1978 $</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FICA Earnings (1)</td>
<td>Adjusted FICA Earnings (2)</td>
</tr>
<tr>
<td>1950</td>
<td>1981</td>
<td>-435.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(210.5)</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>-320.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(235.8)</td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>-349.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(261.6)</td>
</tr>
<tr>
<td></td>
<td>1984</td>
<td>-484.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(286.8)</td>
</tr>
<tr>
<td>1951</td>
<td>1981</td>
<td>-358.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(203.6)</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>-117.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(229.1)</td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>-314.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(253.2)</td>
</tr>
<tr>
<td></td>
<td>1984</td>
<td>-398.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(279.2)</td>
</tr>
<tr>
<td>1952</td>
<td>1981</td>
<td>-342.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(206.8)</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>-235.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(232.3)</td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>-437.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(257.5)</td>
</tr>
<tr>
<td></td>
<td>1984</td>
<td>-436.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(281.9)</td>
</tr>
</tbody>
</table>
The goal of tax-deferred savings programs like 401(k) plans is to increase savings for retirement.

Does participation in 401(k) plans lead to additional savings or simply to the crowding out of other types of savings?

Participants presumably have stronger preferences for savings, so even in the absence of the program they would have saved more than those who do not participate.

Since 401(k) eligibility is decided by employers, unobserved preferences for savings may play a minor role in the determination of eligibility (Poterba, Venti and Wise, 1996).

This suggests using 401(k) eligibility as an instrument for 401(k) participation.

However, 401(k) eligibles and non-eligibles differ in observed characteristics that correlate with savings (earnings, marital status, etc.)

Then, 401(k) eligibility may be viewed as a plausible instrument only after controlling for earnings, marital status, etc.
### Table 1
Means and standard deviations

<table>
<thead>
<tr>
<th></th>
<th>Entire sample</th>
<th>By 401(k) participation</th>
<th>By 401(k) eligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants</td>
<td>Non-participants</td>
<td>Eligibles</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation in 401(k)</td>
<td>0.28</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.46)</td>
<td>(0.46)</td>
</tr>
<tr>
<td><strong>Instrument</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligibility for 401(k)</td>
<td>0.39</td>
<td>1.00</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td><strong>Outcome variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family net financial assets</td>
<td>19,071.68</td>
<td>38,472.96</td>
<td>11,667.22</td>
</tr>
<tr>
<td></td>
<td>(63,963.84)</td>
<td>(79,271.08)</td>
<td>(55,289.23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11,676.77</td>
</tr>
<tr>
<td>Participation in IRA</td>
<td>0.25</td>
<td>0.36</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.48)</td>
<td>(0.41)</td>
</tr>
<tr>
<td><strong>Covariates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family income</td>
<td>39,254.64</td>
<td>49,815.14</td>
<td>35,224.25</td>
</tr>
<tr>
<td></td>
<td>(24,090.00)</td>
<td>(26,814.24)</td>
<td>(21,649.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>34,066.10</td>
</tr>
<tr>
<td>Age</td>
<td>41.08</td>
<td>41.51</td>
<td>40.91</td>
</tr>
<tr>
<td></td>
<td>(10.30)</td>
<td>(9.65)</td>
<td>(10.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.82</td>
</tr>
<tr>
<td>Married</td>
<td>0.63</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.46)</td>
<td>(0.49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td>Family size</td>
<td>2.89</td>
<td>2.92</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>(1.53)</td>
<td>(1.47)</td>
<td>(1.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.87</td>
</tr>
</tbody>
</table>

*Note:* The sample includes 9275 observations from the SIPP of 1991. The observational units are household reference persons aged 25–64, and spouse if present, with *Family Income* in the $10,000–$200,000 interval. Other sample restrictions are the same as in Poterba et al. (1995).
Instrumental Variables Models with Covariates

- Like in the 401(k) examples, often IV identification assumptions may only hold once we condition on a set of pre-treatment characteristics $X$.

- This poses no additional problems for the traditional IV model. Because treatment effects are assumed to be constant we can just include $X$ in the 2SLS model.

- However, in the presence of heterogeneous treatment effects, 2SLS does not identify LATE or other well-defined average treatment effect.
IV Identification with Covariates

Identification assumption:

1. Conditional Independence: 
\((Y_0, Y_1, D_0, D_1) \perp \perp Z|X)\)

2. First Stage: 
\(0 < \Pr(Z = 1|X) < 1, \ Pr(D_1 = 1|X) > \Pr(D_0 = 1|X)\)

3. Monotonicity: 
\(\Pr(D_1 \geq D_0|X) = 1\)

Theorem (Abadie, 2003)

Let \(g(\cdot)\) be any function of \((Y, D, X)\) such that \(E|g(Y, D, X)| < \infty\). Define

\[
\kappa = 1 - \frac{D(1 - Z)}{\Pr(Z = 0|X)} - \frac{(1 - D)Z}{\Pr(Z = 1|X)}.\]

Given our identification assumptions we have that:

\[
E[g(Y, D, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa g(Y, D, X)].
\]

Notice that \(\Pr(D_1 > D_0) = E[\kappa]\).
Kappa: Intuition

We want to estimate $E[g(Y, D, X)|D_1 > D_0]$ but

$$E[g(Y, D, X)] = E[g(Y, D, X)|D_1 > D_0] \Pr(D_1 > D_0)$$

$$+ E[g(Y, D, X)|D_1 = D_0 = 1] \Pr(D_1 = D_0 = 1)$$

$$+ E[g(Y, D, X)|D_1 = D_0 = 0] \Pr(D_1 = D_0 = 0)$$

Therefore,

$$E[g(Y, D, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} \left\{ E[g(Y, D, X)] - E[g(Y, D, X)|D_1 = D_0 = 1] \Pr(D_1 = D_0 = 1) - E[g(Y, D, X)|D_1 = D_0 = 0] \Pr(D_1 = D_0 = 0) \right\}$$

Kappa:

$$\kappa = 1 - \frac{D(1-Z)}{\Pr(Z = 0|X)} - \frac{(1-D)Z}{\Pr(Z = 1|X)}$$
Local Average Response Function

Definition (LARF)

The local average response function is the regression of $Y$ on $X$ and $D$ for the subpopulation of compliers:

$$E[Y|X, D, D_1 > D_0].$$

Because $D$ and $Z$ are the same random variable for compliers, and because $Z$ is independent of $(Y_1, Y_0)$ given $X$, it follows that the LARF identifies the LATE conditional on $X$:

$$E[Y|X, D = 0, D_1 > D_0] = E[Y_0|X, Z = 0, D_1 > D_0]$$

$$= E[Y_0|X, D_1 > D_0],$$

and similarly, $E[Y|X, D = 1, D_1 > D_0] = E[Y_1|X, D_1 > D_0]$. Therefore:

$$E[Y|X, D = 1, D_1 > D_0] - E[Y|X, D = 0, D_1 > D_0] = E[Y_1 - Y_0|X, D_1 > D_0].$$
Estimation of LARF

- Approximate LARF within some class of functions, e.g., linear:

\[(\alpha, \beta) = \arg\min_{a, b} E \left[ \left( E[Y|X, D, D_1 > D_0] - (aD + X'b) \right)^2 \bigg| D_1 > D_0 \right] \]

- From the regression handout, we know

\[(\alpha, \beta) = \arg\min_{a, b} E \left[ \left( Y - (aD + X'b) \right)^2 \bigg| D_1 > D_0 \right] \]

- Applying the kappa theorem yields:

\[(\alpha, \beta) = \arg\min_{a, b} E \left[ \kappa \left( Y - (aD + X'b) \right)^2 \right] \]

- With analog estimator:

\[(\hat{\alpha}, \hat{\beta}) = \arg\min_{a, b} \frac{1}{N} \sum_{i=1}^{N} \kappa_i \left( Y_i - aD_i - X_i'b \right)^2 \]
Estimation of LARF

- $\kappa_i$ is not known:

\[
\kappa_i = 1 - \frac{D_i(1 - Z_i)}{\Pr(Z_i = 0 | X_i)} - \frac{(1 - D_i)Z_i}{\Pr(Z_i = 1 | X_i)}.
\]

- Proceed in two steps:
  1. Estimate $\Pr(Z_i = 1 | X_i)$ (e.g., using Probit, Logit, nonparametric regression). Obtain the fitted values $\hat{\Pr}(Z_i = 1 | X_i)$
  2. Estimate the LARF using:

\[
(\hat{\alpha}, \hat{\beta}) = \arg\min_{a,b} \frac{1}{N} \sum_{i=1}^{N} \hat{\kappa}_i \left( Y_i - aD_i - X_i' b \right)^2
\]

where

\[
\hat{\kappa}_i = 1 - \frac{D_i(1 - Z_i)}{\hat{\Pr}(Z_i = 0 | X_i)} - \frac{(1 - D_i)Z_i}{\hat{\Pr}(Z_i = 1 | X_i)}.
\]

- Abadie, Angrist, and Imbens (2002) apply the same ideas to quantile regression.
Example: Effect of Participation in 401(k) Saving Plans

Variables:

*Treatment* (*D*): Participation in 401(k)

*Instrument* (*Z*): Eligibility for 401(k)

*Outcome variables* (*Y*): Family Net Financial Assets

*Covariates* (*X*): Family Income
Age
Marital Status
Family Size
### Example: Effect on Family Net Financial Assets (in $)

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Ordinary least squares (1)</th>
<th>Endogenous treatment</th>
<th>Least squares treated (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First stage (2)</td>
<td>Second stage (3)</td>
<td></td>
</tr>
<tr>
<td>Participation in 401(k)</td>
<td>13,527.05 (1,810.27)</td>
<td>9,418.83 (2,152.89)</td>
<td>10,800.25 (2,261.55)</td>
</tr>
<tr>
<td>Constant</td>
<td>−23,549.00 (2,178.08)</td>
<td>−0.0306 (0.0087)</td>
<td>−27,133.56 (3,212.35)</td>
</tr>
<tr>
<td>Family income (in thousand $)</td>
<td>976.93 (83.37)</td>
<td>997.19 (83.86)</td>
<td>982.37 (106.65)</td>
</tr>
<tr>
<td>Age (minus 25)</td>
<td>−376.17 (236.98)</td>
<td>−0.0022 (0.0010)</td>
<td>312.30 (371.76)</td>
</tr>
<tr>
<td>Age (minus 25) squared</td>
<td>38.70 (7.67)</td>
<td>0.0001 (0.0000)</td>
<td>24.44 (11.40)</td>
</tr>
<tr>
<td>Married</td>
<td>−8,369.47 (1,829.93)</td>
<td>−0.0005 (0.0079)</td>
<td>−6,646.69 (2,742.77)</td>
</tr>
<tr>
<td>Family size</td>
<td>−785.65 (410.78)</td>
<td>0.0001 (0.0024)</td>
<td>−1,234.25 (647.42)</td>
</tr>
<tr>
<td>Eligibility for 401(k)</td>
<td>0.6883 (0.0080)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* The dependent variable in column (2) is *Participation in 401(k)*. The sample includes 9275 observations from the SIPP of 1991. The observational units are household reference persons aged 25–64, and spouse if present, with *Family Income* in the $10,000–$200,000 interval. Other sample restrictions are the same as in Poterba et al. (1995). Robust standard errors are reported in parentheses.
Let

$$\kappa(1) = D \frac{Z - \Pr(Z = 1|X)}{\Pr(Z = 0|X) \Pr(Z = 1|X)}$$

$$\kappa(0) = (1 - D) \frac{(1 - Z) - \Pr(Z = 0|X)}{\Pr(Z = 0|X) \Pr(Z = 1|X)}.$$

For any function $g(\cdot)$, under the previous assumptions:

$$E[g(Y_1, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa(1)g(Y, X)]$$

$$E[g(Y_0, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa(0)g(Y, X)]$$

(Abadie, 2003).
Treatment effects

\[ E[Y_1 \mid D_1 > D_0], \ E[Y_0 \mid D_1 > D_0] \text{ and } E[Y_1 - Y_0 \mid D_1 > D_0] \]

In particular, making \( g(Y, X) = Y \) we obtain:

\[
E[Y_1 \mid D_1 > D_0] = \frac{E[\kappa_1 Y]}{E[\kappa_1]}, \\
E[Y_0 \mid D_1 > D_0] = \frac{E[\kappa_0 Y]}{E[\kappa_0]}.
\]

Therefore,

\[
E[Y_1 - Y_0 \mid D_1 > D_0] = \frac{E[\kappa_1 Y]}{E[\kappa_1]} - \frac{E[\kappa_0 Y]}{E[\kappa_0]} = E \left[ Y \frac{Z - \Pr(Z = 1 \mid X)}{\Pr(Z = 0 \mid X) \Pr(Z = 1 \mid X)} \right] = E \left[ D \frac{Z - \Pr(Z = 1 \mid X)}{\Pr(Z = 0 \mid X) \Pr(Z = 1 \mid X)} \right].
\]
Treatment effects

$E[Y_1 \mid D_1 > D_0]$, $E[Y_0 \mid D_1 > D_0]$ and $E[Y_1 - Y_0 \mid D_1 > D_0]$

Apply the analogy principle to obtain estimators:

\[
\hat{E}[Y_1 \mid D_1 > D_0] = \frac{1}{N} \sum_{i=1}^{N} \hat{\kappa}_{(1)i} Y_i,
\]

\[
\hat{E}[Y_0 \mid D_1 > D_0] = \frac{1}{N} \sum_{i=1}^{N} \hat{\kappa}_{(0)i} Y_i,
\]

\[
\hat{E}[Y_1 - Y_0 \mid D_1 > D_0] = \frac{1}{N} \sum_{i=1}^{N} \frac{Z_i - \hat{\Pr}(Z_i = 1 \mid X_i)}{\hat{\Pr}(Z_i = 0 \mid X_i)\hat{\Pr}(Z_i = 1 \mid X_i)}.
\]
Describing Compliers

- We can estimate average treatment effects for compliers, even when we cannot identify compliers individually.
- So, who are these compliers? The kappa result allows us to describe them. For example, using $g(Y, D, X) = X$:

$$E[X|D_1 > D_0] = \frac{1}{Pr(D_1 > D_0)} E[\kappa X]$$

$$= \frac{E[\kappa X]}{E[\kappa]}.$$ 

- We can estimate:

$$\hat{E}[X|D_1 > D_0] = \frac{1}{N} \sum_{i=1}^{N} \hat{\kappa}_i X_i.$$