
Kevin M. Esterling (UC – Riverside)

Presentation to GradQuant

April 29, 2014
Introduction

Rubin Causal Model (RCM) is a framework to identify causal effects from data

- “Identification” implies your estimand has enough constraints for a unique solution
- RCM states 5 assumptions that are necessary for identification
- Purely conceptual problem; estimation is a separate question entirely (“estimand” versus “estimator”)

RCM has revolutionized how biomedical and social scientists design research and analysis

- Directs attention to research design
- Simplifies analysis and improves credibility
- Assumptions clarify how hard it is to make causal statements; why social science is so hard to do
Plan for the talk

- Show traditional IV model
- Define “causality” and state the fundamental problem of causal inference
- State 5 RCM assumptions required to identify causal effect with data
  - Substantive examples of assumption violations
  - Notation widely adopted in the literature so useful to review
- Apply RCM to IV and demonstrate that the traditional interpretation of IV parameter is wrong
A substantive example

Variables
- $Y = \text{earnings}$
- $D = \text{training program}$
- $Z = \text{random assignment to training program}$
- $U = \text{unobserved willingness to take training}$

Relationships among variables
- $Y = \beta \times D + U + \epsilon_1 = \beta \times D + \omega$
- $D = D^* + U + \epsilon_2$
- $\text{cov}(D, \omega) \neq 0; \text{ positive or negative selection}$
Regression definition of IV

\[ \frac{\text{Cov}(Z,Y)}{\text{Var}(Z)} = \hat{\delta} \beta \]

\[ \frac{\text{Cov}(Z,D)}{\text{Var}(Z)} = \hat{\delta} \]

\[ \frac{\text{Cov}(D,Y)}{\text{Var}(D)} \neq \hat{\beta} \]

\[ \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)} = \hat{\beta} \]
Regression definition of IV

- $\hat{\beta}_{OLS} \neq \beta$ because $\text{cov}(D, U + \epsilon) \neq 0$ (\hat{\beta} “not identified”)
- IV identifies $\beta$
- Regression model is stated in terms of the observed data; $\delta$ and $\beta$ are the only missing parameters
- Implies that $\hat{\beta}$ is the ATE estimate but the RCM shows that this is unlikely to be true
- The “$\hat{\beta}$” that is identified in IV model does not have a natural interpretation in the regression framework, but a natural one in the RCM
Potential Outcomes Framework

- In order to make sense of $\beta$ one must use the RCM.
- Define *causality* at the individual level in terms of “potential outcomes”
  - A regression slope estimates causality at the population level.
  - Get this out of your head the best you can . . .
- Each unit/subject has two potential outcomes
  - $Y(Z = 0)$ is the outcome you would observe if the subject were assigned to control.
  - $Y(Z = 1)$ is the outcome you would observe if the subject were assigned to treatment.
- Causal Effect = $Y_i(Z = 1) - Y_i(Z = 0)$
- These potential outcomes are conceptually real before the experiment is run.
RCM: Defining causality at the individual level

Table: Before the experiment is run

<table>
<thead>
<tr>
<th>Unit</th>
<th>Z</th>
<th>D</th>
<th>Y(Z=0)</th>
<th>Y(Z=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Causal Effect = \( Y_i(Z = 1) - Y_i(Z = 0) \) (\( \Rightarrow \) “Fundamental problem of causal inference”)
Fundamental problem of causal inference; causality as missing data problem

<table>
<thead>
<tr>
<th>Unit</th>
<th>Z</th>
<th>D</th>
<th>Y(Z=0)</th>
<th>Y(Z=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
</tbody>
</table>

\[ ATE = E[Y_i(D = 1)] - E[Y_i(D = 0)] \]
Complications from using human subjects

- In many social science applications, we cannot compel people to take the treatment or not take the treatment.
- This yields four principal strata:
  - Complier
  - Never taker
  - Always taker
  - Defier
- In the presence of non-compliers, randomization is destroyed.
- Treatment effect estimates are biased if individual treatment effects vary.
- If principal strata were observable, then it would be easy to identify treatment effects; compare oranges to oranges.
Define “principal strata”

<table>
<thead>
<tr>
<th>Unit</th>
<th>Z</th>
<th>D</th>
<th>Y(Z=0)</th>
<th>Y(Z=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Complier</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td><strong>Never Taker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td><strong>Always Taker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>=</td>
<td>=</td>
</tr>
<tr>
<td><strong>Defier</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>#</td>
<td>?</td>
</tr>
</tbody>
</table>

Table: Non-Compliance “Principal Strata”
RCM requires 5 assumptions to identify causal effects

**Notation**

- $Z$ is an n-dimensional assignment variable, $Z_i \in \{0, 1\}$; 1 means “assigned to treatment” and 0 means “assigned to control”
- $D_i(Z)$ is an indicator variable equal to 1 if subject $i$ would take the treatment given $Z$ and 0 if control; there are $2^n$ potential treatment exposures for each subject
- $Y_i(Z, D)$ are potential outcomes given $Z$ and $D$
  - e.g., with 100 subjects there are $2 \times 2^n$ potential outcomes per subject virtually all of which would be missing in the data
Assumption 1: Random Assignment
You did the randomization right... Usually not a strong assumption

Formal statement
\[ p(Z = c) = p(Z = c') \]
for all \( c, c' \) such that \( i^T c = i^T c' \)
e.g., \( c = [0, 1, 0, 1] \), \( c' = [0, 0, 1, 1] \)
Assumption 2: SUTVA

1. No interference between units – potential outcomes are not a function of who else is in the study
2. No versions of treatment – potential outcomes are constant for individuals

If both statements are true we can disregard the vector of assignments for each unit. Often a strong assumption: congestion and general equilibrium effects; spillover effects. Tested with more complex designs

Formal statement

If $Z_i = Z_i'$ then $D_i(Z) = D_i(Z')$; “no interference”
And if $Z_i = Z_i'$ and $D_i = D_i'$ then $Y_i(Z, D) = Y_i(Z', D')$; “no versions” and “no interference”
⇒ Then we can write: $Y_i(Z, D) = Y_i(Z_i, D_i)$
Assumption 3: Exclusion restriction

The assignment has no effect on outcomes other than through the treatment. Also a very strong assumption, e.g., disappointment, strategic behavior. Typically not testable with data.

Formal statement

If \( Y_i(1, d_i) = Y_i(0, d_i) \) for \( d_i = 0, 1 \)

Then we can write: \( Y_i(Z_i, D_i) = Y_i(D_i) \)
Assumption 4: Relevancy

Nonzero causal effect of $Z$ on $D$; instrument has to be correlated with treatment exposure. Testable with data

Formal statement

$$E[D_i(1) - D_i(0)] \neq 0$$
Assumption 5: Monotonicity
No “defiers” in the sample.

Formal statement
\[ p[D_i(1) - D_i(0) = -1] = 0 \]
Define the Intention to Treat (ITT) estimand

Table: ITT

<table>
<thead>
<tr>
<th>Unit</th>
<th>Z</th>
<th>Y(Z=0)</th>
<th>Y(Z=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
</tbody>
</table>

⇒ ITT = E(Y|Z = 1) − E(Y|Z = 0)
Define the “compliance” estimand

\[ p(\text{Complier}) = E(D|Z = 1) - E(D|Z = 0) \]

**Table: ITT**

<table>
<thead>
<tr>
<th>Unit</th>
<th>Z</th>
<th>D(Z=0)</th>
<th>D(Z=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>?</td>
<td>#</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>#</td>
<td>?</td>
</tr>
</tbody>
</table>
So, what the data can tell us . . .

Both the ITT and “compliance” estimands are defined in terms of (partially unobserved) potential outcomes, but the counterfactuals are well defined and hence can be estimated directly from the data.

**ITT assumptions:** randomization, exclusion restriction, SUTVA

**Relevancy assumptions:** randomization, monotonicity

Now use the data and assumptions to identify the causal effect from IV....
**IV estiamand derivation**

Decompose the ITT into each subgroup's (principal strata) contribution

**IV Decomposition**

\[
ITT = ITT_{Complier} p(Complier) + ITT_{NeverTaker} p(NeverTaker) + ITT_{AlwaysTaker} p(AlwaysTaker) + ITT_{Defier} p(Defier)
\]

♦ 1 equation and 8 parameters (!), but

- \(ITT_{NeverTaker} = 0\); by exclusion restriction
- \(ITT_{AlwaysTaker} = 0\)
- \(p(Defier) = 0\) by monotonicity
**IV estimand derivation**

Decompose the ITT into each subgroup’s (principal strata) contribution

**IV Decomposition**

\[
\text{ITT} = \text{ITT}_{\text{Complier}} p(\text{Complier}) + \text{ITT}_{\text{NeverTaker}} p(\text{NeverTaker}) + \text{ITT}_{\text{AlwaysTaker}} p(\text{AlwaysTaker}) + \text{ITT}_{\text{Defier}} p(\text{Defier})
\]

◇ 1 equation and 8 parameters (!), but

- \(\text{ITT}_{\text{NeverTaker}} = 0\); by exclusion restriction
- \(\text{ITT}_{\text{AlwaysTaker}} = 0\)
- \(p(\text{Defier}) = 0\) by monotonicity
IV estimand derivation

Decompose the ITT into each subgroup’s (principal strata) contribution

IV Decomposition

\[ ITT = ITT_{Complier} p(Complier) + ITT_{NeverTaker} p(NeverTaker) + ITT_{AlwaysTaker} p(AlwaysTaker) + ITT_{Defier} p(Defier) \]

◊ 1 equation and 8 parameters (!), but

- \( ITT_{NeverTaker} = 0 \); by exclusion restriction
- \( ITT_{AlwaysTaker} = 0 \)
- \( p(Defier) = 0 \) by monotonicity
IV estimand derivation

Decompose the ITT into each subgroup’s (principal strata) contribution

\[ \text{ITT} = \text{ITT}_{\text{Complier}} p(\text{Complier}) + \text{ITT}_{\text{NeverTaker}} p(\text{NeverTaker}) + \]
\[ \text{ITT}_{\text{AlwaysTaker}} p(\text{AlwaysTaker}) + \text{ITT}_{\text{Defier}} p(\text{Defier}) \]

\( \diamond \) 1 equation and 8 parameters (!), but

- \( \text{ITT}_{\text{NeverTaker}} = 0 \); by exclusion restriction
- \( \text{ITT}_{\text{AlwaysTaker}} = 0 \)
- \( p(\text{Defier}) = 0 \) by monotonicity
### IV estimand derivation, cont.

\[ \text{ITT}_{\text{Complier}} = \frac{\text{ITT}}{p(\text{Complier})} = \text{LATE} \]

⇒ R.H.S. terms are estimated from the data by some method

### IV estimand interpretation

IV method identifies the “complier average causal effect” (CACE) or the “local average treatment effect” (LATE) but not the “average treatment effect” (ATE)
Five assumptions often not met in practice

- Randomization of assignment (*)
- SUTVA (*)
  - No interference
  - No versions of treatments
- Exclusion restriction (*)
- Relevancy
- Monotonicity
Principal Stratification


- Extends IV to the case of noncompliance and missing outcome data
- Assumes missing data are MAR after conditioning on the stratifying variable and covariates
- Requires “compound exclusion restriction” = outcomes and response not affected by the assignment
Principal stratification with measured compliance


- For IV/PS, compliance type is only partially observable
- ENL proposes an experimental design to measure compliance from subjects’ participation behavior
- Observing compliance:
  - Improves efficiency of treatment effect estimates; less missing data
  - Identifies ATE as well as LATE
  - Enables a formal test of the exclusion restriction (a deeply problematic assumption)
  - Allows estimate of treatment effect heterogeneity as a function of compliance type
  - Allows estimation of the latent correlation between treatment compliance and outcomes; formal test of balance after conditioning on covariates