Evaluating Causal Theories

Christopher Adolph

University of Washington, Seattle

April 5, 2010
Outline

Defining correlation & casusation

Writing causal theories as models

Testing causal theories
I used to think correlation implied causation.

Then I took a statistics class. Now I don't.

Sounds like the class helped.

\(\text{Well, maybe.}\)
"Correlation does not imply causation."

The above statement is:

1. Technically true, for the definition of “imply” used by logicians.
"Correlation does not imply causation."

The above statement is:

- Technically true, for the definition of "imply" used by logicians.
  - To a student of logic: “X implies Y” ($X \Rightarrow Y$) is translated as “if X is true, then Y must also be true.”

- Highly misleading, under the colloquial understanding of "imply":
  - To most English speakers, "imply" translates as "suggests".
  - Thus the statement becomes: "Correlation doesn't suggest the presence of causation."

- This is false. Correlation may not be sufficient evidence to conclude causation, but it is a clue!
"Correlation does not imply causation."

The above statement is:

- Technically true, for the definition of “imply” used by logicians.
  - To a student of logic: “X implies Y” ($X \Rightarrow Y$) is translated as “if $X$ is true, then $Y$ must also be true.”
  - Thus the statement becomes: “If two things are correlated, they are not necessarily causally related, but might be.”
“Correlation does not imply causation.”

The above statement is:

1. Technically true, for the definition of “imply” used by logicians.
   - To a student of logic: “X implies Y” \((X \Rightarrow Y)\) is translated as “if X is true, then Y must also be true.”
   - Thus the statement becomes: “If two things are correlated, they are not necessarily causally related, but might be.”
   - This is TRUE
“Correlation does not imply causation.”

The above statement is:

1. Technically true, for the definition of “imply” used by logicians.
   - To a student of logic: “X implies Y” ($X \implies Y$) is translated as “if X is true, then Y must also be true.”
   - Thus the statement becomes: “If two things are correlated, they are not necessarily causally related, but might be.”
   - This is TRUE

2. Highly misleading, under the colloquial understanding of “imply”:

Chris Adolph (UW)
"Correlation does not imply causation."

The above statement is:

1. Technically true, for the definition of “imply” used by logicians.
   - To a student of logic: “X implies Y” ($X \Rightarrow Y$) is translated as “if X is true, then Y must also be true.”
   - Thus the statement becomes: “If two things are correlated, they are not necessarily causally related, but might be.”
   - This is TRUE

2. Highly misleading, under the colloquial understanding of “imply”:
   - To most English speakers, “imply” translates as “suggests”.
   - Thus the statement becomes: “Correlation doesn’t suggest the presence of causation.”
“Correlation does not imply causation.”

The above statement is:

1. Technically true, for the definition of “imply” used by logicians.
   - To a student of logic: “X implies Y” \((X \Rightarrow Y)\) is translated as “if X is true, then Y must also be true.”
   - Thus the statement becomes: “If two things are correlated, they are not necessarily causally related, but might be.”
   - This is TRUE

2. Highly misleading, under the colloquial understanding of “imply”:
   - To most English speakers, “imply” translates as “suggests”.
   - Thus the statement becomes: “Correlation doesn’t suggest the presence of causation.”
   - This is FALSE. Correlation may not be sufficient evidence to conclude causation, but it is a clue!
Can Keynesian fiscal policy shorten the recession?

**Research question**  What is the effect of fiscal stimulus on recessions?
Can Keynesian fiscal policy shorten the recession?

Research question  What is the effect of fiscal stimulus on recessions?

Dependent Variable  Monthly national unemployment rate
Can Keynesian fiscal policy shorten the recession?

**Research question**  What is the effect of fiscal stimulus on recessions?

**Dependent Variable**  Monthly national unemployment rate

**Independent Variable**  Presence and size of stimulus packages
Can Keynesian fiscal policy shorten the recession?

**Research question**  What is the effect of fiscal stimulus on recessions?

**Dependent Variable**  Monthly national unemployment rate

**Independent Variable**  Presence and size of stimulus packages

**Hypothesis**  Larger fiscal stimulus lowers unemployment
Can Keynesian fiscal policy shorten the recession?

Research question  What is the effect of fiscal stimulus on recessions?

Dependent Variable  Monthly national unemployment rate

Independent Variable  Presence and size of stimulus packages

Hypothesis  Larger fiscal stimulus lowers unemployment

Can we test this hypothesis?
Cause and Effect

**Correlation**  Suppose $X_i$ and $Y_i$ are correlated. Then:

- When $X_i$ is large, $Y_i$ is large, and vice versa.
- When $X_i$ is small, $Y_i$ is small, and vice versa.

Correlation is directionless.

\[
corr(X_i, Y_i) = corr(Y_i, X_i)
\]

If $X_i$ is correlated with $Y_i$, then $Y_i$ is correlated with $X_i$.

**Causation**  Suppose $X_i$ causes $Y_i$. Then:

- When $X_i$ is large, it causes $Y_i$ to be large.
- When $X_i$ is small, it causes $Y_i$ to be small.

Causation is directional.

$X_i \Rightarrow Y_i$ does not imply $Y_i \Rightarrow X_i$.
Cause and Effect

Correlation  Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and *vice versa*

Causation  Suppose $X_i$ causes $Y_i$. Then:

- when $X_i$ is large, it causes $Y_i$ to be large
- when $X_i$ is small, it causes $Y_i$ to be small

Causation is directional. $X_i \Rightarrow Y_i$ does not imply $Y_i \Rightarrow X_i$
Cause and Effect

Correlation Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and *vice versa*
- when $X_i$ is small, $Y_i$ is small, and *vice versa*
Cause and Effect

Correlation
Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and *vice versa*
- when $X_i$ is small, $Y_i$ is small, and *vice versa*
- Correlation is directionless. $\text{corr}(X_i, Y_i) = \text{corr}(Y_i, X_i)$
Cause and Effect

Correlation  Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and \textit{vice versa}
- when $X_i$ is small, $Y_i$ is small, and \textit{vice versa}
- Correlation is directionless. \( \text{corr}(X_i, Y_i) = \text{corr}(Y_i, X_i) \)
- If $X_i$ is correlated with $Y_i$, then $Y_i$ is correlated with $X_i$

Causation  Suppose $X_i$ causes $Y_i$. Then:

- when $X_i$ is large, it causes $Y_i$ to be large
Cause and Effect

Correlation  Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and *vice versa*
- when $X_i$ is small, $Y_i$ is small, and *vice versa*
- Correlation is directionless. $\text{corr}(X_i, Y_i) = \text{corr}(Y_i, X_i)$
- If $X_i$ is correlated with $Y_i$, then $Y_i$ is correlated with $X_i$

Causation  Suppose $X_i$ causes $Y_i$. Then:

- when $X_i$ is large, it causes $Y_i$ to be large
- when $X_i$ is small, it causes $Y_i$ to be small
**Cause and Effect**

**Correlation** Suppose $X_i$ and $Y_i$ are correlated. Then:

- when $X_i$ is large, $Y_i$ is large, and *vice versa*
- when $X_i$ is small, $Y_i$ is small, and *vice versa*
- Correlation is directionless. $\text{corr}(X_i, Y_i) = \text{corr}(Y_i, X_i)$
- If $X_i$ is correlated with $Y_i$, then $Y_i$ is correlated with $X_i$

**Causation** Suppose $X_i$ causes $Y_i$. Then:

- when $X_i$ is large, it causes $Y_i$ to be large
- when $X_i$ is small, it causes $Y_i$ to be small
- Causation is directional. $X_i \Rightarrow Y_i$ does not imply $Y_i \Rightarrow X_i$
Signal and noise

\[ Y_i = f(X_i) + \varepsilon_i \]

\(\varepsilon\) is a bit of random noise added to \(X_i\).
Signal and noise

\[ Y_i = f(X_i) + \varepsilon_i \]

\( \varepsilon \) is a bit of random noise added to \( X_i \).

The world isn’t usually deterministic, so relationships between variables are subject to random slippage.
**Signal and noise**

\[ Y_i = f(X_i) + \varepsilon_i \]

\( \varepsilon \) is a bit of random noise added to \( X_i \).

The world isn’t usually deterministic, so relationships between variables are subject to random slippage.

But *on average*, \( X_i \) and \( Y_i \) will move together when correlated, even if a few cases with large \( \varepsilon_i \) are exceptions.
Signal and noise

\[ Y_i = f(X_i) + \varepsilon_i \]

\( \varepsilon \) is a bit of random noise added to \( X_i \).

The world isn’t usually deterministic, so relationships between variables are subject to random slippage.

But on average, \( X_i \) and \( Y_i \) will move together when correlated, even if a few cases with large \( \varepsilon_i \) are exceptions.

This means we can’t dismiss a relationship based on one or two counter-examples.
Signal and noise

\[ Y_i = f(X_i) + \varepsilon_i \]

\( \varepsilon \) is a bit of random noise added to \( X_i \).

The world isn’t usually deterministic, so relationships between variables are subject to random slippage.

But *on average*, \( X_i \) and \( Y_i \) will move together when correlated, even if a few cases with large \( \varepsilon_i \) are exceptions.

This means we can’t dismiss a relationship based on one or two counter-examples.

Suggests we need a lot of observations to study causality.
Correlation examples

1.0  0.8  0.4  0.0  -0.4  -0.8  -1.0

1.0  1.0  1.0  -1.0  -1.0  -1.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0

wavy  square  diamond  half-circle  cross  circle  dots
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be *spurious*?

- What if countries with severe endebtedness find it harder to borrow money for Keynesian stimulus, *and*
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?

- What if countries with severe indebtedness find it harder to borrow money for Keynesian stimulus, and

- Endebtedness independently lengthens recessions by discouraging consumer spending
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?

- What if countries with severe endebtedness find it harder to borrow money for Keynesian stimulus, and

- Endebtedness independently lengthens recessions by discouraging consumer spending

Would this threaten our inference of causality from correlation?
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?

- What if countries with severe endebtedness find it harder to borrow money for Keynesian stimulus, and
- Endebtedness independently lengthens recessions by discouraging consumer spending

Would this threaten our inference of causality from correlation?
Yes. Endebtedness may be a confounding variable
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?

- What if countries with severe indebtedness find it harder to borrow money for Keynesian stimulus, and

- Endebtedness independently lengthens recessions by discouraging consumer spending

Would this threaten our inference of causality from correlation? Yes. Endebtedness may be a confounding variable

How can we solve this problem?
Causes and confounders

We observe a correlation between fiscal stimulus & fast recovery from recession

Could this correlation be spurious?

- What if countries with severe indebtedness find it harder to borrow money for Keynesian stimulus, and
- Endebtedness independently lengthens recessions by discouraging consumer spending

Would this threaten our inference of causality from correlation?
Yes. Endebtedness may be a confounding variable

How can we solve this problem?
Compare countries with similar ex ante debt levels: Control for debt
A simple causal model

What is the causal relationship here?

Regular Exercise

Health
A simple causal model

What is the causal relationship here?

Which variables are correlated?

Regular Exercise

Health
A simple causal model

- Regular Exercise
- Health

What is the causal relationship here?

Which variables are correlated?

What is missing from this path diagram?
A simple causal model

We assume a random component affects each variable.
A simple causal model

We assume a random component affects each variable.

Here, both exercise and health are subject to random fluctuation.
A simple causal model

These random effects are uncorrelated with Health and Exercise

(why?)
A simple causal model

Regular Exercise \[ \epsilon \] Health

\( \epsilon \) means Exercise varies randomly from person to person.
A simple causal model

Health results from both Exercise and $\nu$
A simple causal model

\[ \varepsilon \]

Regular Exercise

Health

\[ \nu \]

Health results from both Exercise and \( \nu \)

\( \nu \) makes relationship between exercise & health probabilistic, *not* deterministic
A simple causal model

In most path diagrams, we will leave out random effects to reduce clutter.

What other variables might fit this pattern?
Mystery model 1

What is the causal relationship here?

- arm length
- genes
- leg length

Which variables are correlated?
Mystery model 1

arm length

genes

leg length

What is the causal relationship here?

Which variables are correlated?
Common cause; spurious correlation

Arm length & leg length are correlated but not causally related
Common cause; spurious correlation

Arm length & leg length are correlated but not causally related.

What if we didn’t know about genes?
Common cause; spurious correlation

Might be misled by correlation between observed arm and leg lengths into perceiving a causal relationship.

Watch out for omitted common causes!
Common cause; spurious correlation

As before, we are assuming random effects on arms and legs.

As before, we are assuming random effects on arms and legs.
As before, we are assuming random effects on arms and legs. Recall that \( \varepsilon \) and \( \nu \) are uncorrelated. What does that mean here?
Common cause; spurious correlation

What other variables might fit in this model?
Mystery model 2

per capita gross domestic product

Female education

Fertility rate

per capita gross domestic product

Female education

What is the causal relationship here?
Mystery model 2

What is the causal relationship here?

Which variables are correlated?

<table>
<thead>
<tr>
<th>per capita gross domestic product</th>
<th>Female education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fertility rate</td>
<td></td>
</tr>
</tbody>
</table>

Chris Adolph (UW)
Course Introduction
April 5, 2010
Indirect relationship

Economic development leads to female education

- Economic development
- Female education
- Per capita gross domestic product

Fertility rate

- Educated females
- Use contraception
- Have fewer kids
Indirect relationship

Economic development leads to female education

Educated females use contraception, have fewer kids
Rising GDP lowers fertility, but not directly
Indirect relationship

Rising GDP lowers fertility, but not directly

GDP and fertility are (inversely) correlated
Indirect relationship

A → X → Y

What other variables fit this model?
Mystery model 3

Democratic Governor
Democratic Legislature

What is the causal relationship here?

Welfare Spending
Writing causal theories as models

Mystery model 3

Democratic Governor

Democratic Legislature

Welfare Spending

What is the causal relationship here?

Which variables are correlated?
Conditional relationship

Effect of Democratic governors depends on party in control of leg, & vice versa
Conditional relationship

Effect of Democratic governors depends on party in control of leg, & vice versa

Welfare might stay low unless Democrats have united control
Conditional relationship

Effects of party control not additive but multiplicative
Conditional relationship

Effects of party control not additive but multiplicative. Known as an “interactive effect.”
Conditional relationship

Can we think of variables that would fit in an interactive model?
## Mystery model 4

<table>
<thead>
<tr>
<th>Campaign Contributions</th>
<th>Expected Vote Percentage</th>
</tr>
</thead>
</table>

What is the causal relationship here?
Mystery model 4

What is the causal relationship here?

Which variables are correlated?
Reciprocal Causation

Campaign Contributions

Better funded candidates do better, all else equal.
Reciprocal Causation

Better funded candidates do better, all else equal

But in close elections even losers raise a lot
Reciprocal Causation

Hard to separate these effects empirically

Campaign Contributions

Expected Vote Percentage
Reciprocal Causation

Hard to separate these effects empirically

Often people “find” campaign contributions & victory are inversely correlated!
Reciprocal Causation

Campaign Contributions

Expected Vote Percentage

Implies candidates should shun campaign contributions!
Reciprocal Causation

Campaign Contributions

Expected Vote Percentage

Implies candidates should shun campaign contributions!

Is this a theoretically plausible causal finding?
Reciprocal Causation

Can we think of other variables that have a reciprocal causal relationship?

X

Y
A complex causal model

We’ll start with a simple model of Vote Choice and build up.
A complex causal model

We’ll start with a simple model of Vote Choice and build up

Strongest predictor of Vote Choice: Party identification
Writing causal theories as models

A complex causal model

- Ideology
- Party ID
- Vote

Ideology might have an independent impact on Vote

Are we missing any arrows?
A complex causal model

Ideology

Vote

Party ID

Ideology and Party ID surely influence each other

Reciprocal causation again!

What else is missing?
What about parents?
Party ID and Ideology tend to be set in childhood & early adulthood
Any missing arrows?

A complex causal model
A complex causal model

If Party ID and ideology mutually cause each other for subjects, they should mutually cause for parents.

Any other missing arrows?
A complex causal model

Perhaps parents’ ideology causes subject party ID?
(Thoughts?)

Perhaps parents’ ideology causes subject vote?
(Thoughts?)
A complex causal model

What’s correlated in this model?
Uncorrelated?
If our main interest is modeling Vote, what are we still missing?

Parents’ Ideology → Ideology
Parents’ Party ID → Party ID
Ideology → Vote
A complex causal model

We could add the subject’s income.

Do we need more arrows?

That is, does Income cause anything else besides vote choice?

Parents’ Ideology

Parents’ Party ID

Party ID

Income

Vote

Ideology

Chris Adolph  (UW)  
Course Introduction  
April 5, 2010  
41 / 56
A complex causal model

Parents’ Ideology → Ideology
Parents’ Party ID → Ideology

Income probably shapes Ideology & Party ID directly

Does Income suggest any other missing variables?
A complex causal model

Parents’ Income a likely cause of subject’s income

Parents’ Income should shape Parents’ Ideol. & Party ID

Other missing variables?
A complex causal model

How about subject’s Education? (Is this a simple relationship?)

Are we missing any arrows?
A complex causal model

Lots: Education should influence income, ideology, and Party ID

You can probably guess the last variable in our diagram
A complex causal model

Once again, we include the parents’ characteristics.

What happens if we omit?

Misattribute effects of parental education!
A complex causal model

What’s causally related in this diagram?

What is (and isn’t) correlated?
How do we test causal theories?

To tease out the difference between correlation & causation, we’ll need some notation:

**Definitions**

- $Y_i$: dependent variable observed for case $i$
- $X_i$: treatment variable, as assigned for case $i$
- $Z_i$: a confounding variable for case $i$; it may or may not be observed
- $E(A|B = b)$: expected value of $A$ when $B = b$
Correlation redux

Finding whether two variables are correlated is easy

Simple formula for correlation (we’ll discuss at much greater length when we get to quantitative methods):

$$\text{corr}(X, Y) = \frac{E((X - E(X))(Y - E(Y)))}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$$

where

$$\text{var}(X) = E\left((X - E(X))^2\right)$$

Finding whether two variables are causally related is hard
### Correlation examples

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>0.8</th>
<th>0.4</th>
<th>0.0</th>
<th>-0.4</th>
<th>-0.8</th>
<th>-1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

![Correlation Examples](image)
Fundamental problem of causal inference

Average causal effect of $X_i$ on $Y_i = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$

In words, the effect of $X_i$ is the difference between $Y_i$ when $X_i$ is present and $Y_i$ when $X_i$ is absent.
Fundamental problem of causal inference

Average causal effect of $X_i$ on $Y_i = E(Y_i|X_i = 1) - E(Y_i|X_i = 0)$

In words, the effect of $X_i$ is the difference between $Y_i$ when $X_i$ is present and $Y_i$ when $X_i$ is absent

But we only ever observe one of these!

The other is counterfactual, and can only be estimated

Why are we taking expectations?
Because there is random noise around each case
We’re interested in the average causal effect, which removes the random noise
Fundamental problem of causal inference

Would the economy be better or worse had Obama and the Democratic Congress not implemented last year’s stimulus package?

If we could rerun time like a tape (er, DVD? Blu-Ray? Video file?), we could change $X_i$ for the same unit and see the change in $Y_i$

That is, we could remove the stimulus package, and track unemployment over the last year

Lacking a time machine, and the power to change history, FPC is unsolvable
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

- Unit homogeneity
  - Fundamental assumption in science: regular behavior
  - Can discount random variation: this is about the expected effect.
  - Points to importance of defining scope of theory correctly.

- Conditional independence
  - No reverse causality; $Y$ does not cause $X$.
  - Selection process is related to size of outcome variable.
  - Points to importance of controls for selection process in observational studies.
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity** Expected causal effect of $X$ on $Y$ is same for all $i$:

Fundamental assumption in science: regular behavior
Can discount random variation: this is about the expected effect
Points to importance of defining scope of theory correctly

Conditional independence
No reverse causality; $Y$ does not cause $X$:
Selection process is related to size of outcome variable
Points to importance of controls for selection process in observational studies
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity**  Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity** Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
- Can discount random variation: this is about the *expected* effect

Points to importance of defining scope of theory correctly

Conditional independence
- No reverse causality; $Y$ does not cause $X$
- Selection process is related to size of outcome variable
- Points to importance of controls for selection process in observational studies
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity**  Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
- Can discount random variation: this is about the *expected* effect
- Points to importance of defining scope of theory correctly
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity** Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
- Can discount random variation: this is about the *expected* effect
- Points to importance of defining scope of theory correctly

**Conditional independence** No reverse causality; $Y$ does not cause $X$: 
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity** Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
- Can discount random variation: this is about the *expected* effect
- Points to importance of defining scope of theory correctly

**Conditional independence** No reverse causality; $Y$ does not cause $X$:
- Selection process is related to size of outcome variable
Coping with Causality

Can we make some assumptions that would help us tackle FPC?

We need two:

**Unit homogeneity**  Expected causal effect of $X$ on $Y$ is same for all $i$:
- Fundamental assumption in science: regular behavior
- Can discount random variation: this is about the *expected* effect
- Points to importance of defining scope of theory correctly

**Conditional independence**  No reverse causality; $Y$ does not cause $X$:
- Selection process is related to size of outcome variable
- Points to importance of controls for selection process in observational studies
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies

But what if strength of theoretical effects vary over cases?
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies.

But what if strength of theoretical effects vary over cases?

E.g., what if Keynesian stimulus is stronger in some cases? (Ideas?)
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies.

But what if strength of theoretical effects vary over cases?

E.g., what if Keynesian stimulus is stronger in some cases? (Ideas?)

Solution:

1. Identify variables mediating strength of effect
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies.

But what if strength of theoretical effects vary over cases?

E.g., what if Keynesian stimulus is stronger in some cases? (Ideas?)

Solution:

1. Identify variables mediating strength of effect

2. Test whether the correlation of stimulus and recession varies systematically with these conditioning variables
How do we satisfy unit homogeneity?

Need to restrict the universe of observations to cases where we strongly believe theory applies.

But what if strength of theoretical effects vary over cases?

E.g., what if Keynesian stimulus is stronger in some cases? (Ideas?)

Solution:

1. Identify variables mediating strength of effect

2. Test whether the correlation of stimulus and recession varies systematically with these conditioning variables

Unit homogeneity is at least partly *testable* with enough data.
How do we satisfy conditional independence?

**Controlled experiment**  Randomly assign observations to two groups, then (artificially) expose the first group to the treatment. Completely solves.

**Matching**  Relying on natural variation in treatment, match cases with and without treatment that are identical, or very similar, in observables.
How do we satisfy conditional independence?

**Controlled experiment**  Randomly assign observations to two groups, then (artificially) expose the first group to the treatment. Completely solves.

**Matching**  Relying on natural variation in treatment, match cases with and without treatment that are identical, or very similar, in observables.

May solve, but never certain. Why?
How do we satisfy conditional independence?

**Controlled experiment** Randomly assign observations to two groups, then (artificially) expose the first group to the treatment. Completely solves.

**Matching** Relying on natural variation in treatment, match cases with and without treatment that are identical, or very similar, in observables.

May solve, but never certain. Why? What if you missed some key independent variables?

**Regression** Relying on natural variation in treatment, mathematically control for effects of observable confounders, leaving only the effect of $X$ on $Y$
How do we satisfy conditional independence?

**Controlled experiment** Randomly assign observations to two groups, then (artificially) expose the first group to the treatment. Completely solves.

**Matching** Relying on natural variation in treatment, match cases with and without treatment that are identical, or very similar, in observables.

May solve, but never certain. Why?
What if you missed some key independent variables?

**Regression** Relying on natural variation in treatment, mathematically control for effects of observable confounders, leaving only the effect of X on Y

May solve, but never certain. Similar vulnerabilities as Matching.

Can we imagine applying these methods to the effectiveness of Keynesian stimulus?
I used to think correlation implied causation.

Then I took a statistics class. Now I don’t.

Sounds like the class helped.

Well, maybe.

http://xkcd.com/552/