# Instrumental Variable (IV) Methods

## Introduction to IVs

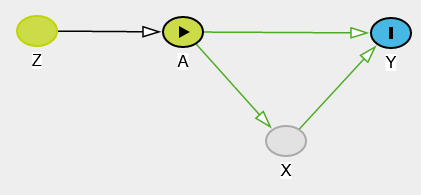
### Confounding

* A classic case of a confounding variable is some variable that directly affects both the treatment A and outcome Y.
* If X is observed, the matching methods or IPTW can be used to estimate the causal effect of A on Y.
* This is the case even if there are some risk factors, V, that only affect the outcome.

### Unmeasured Confounding

* In many cases, we may be concerned that there is unmeasured confounding i.e. variables we haven’t observed that affect both treatment and the outcome.
* This means the ignorability assumption is violated and treatment couldn’t be thought of as randomised conditional on the observed Xs.
* We then have biased estimates of causal effects as we can’t control for/condition on the confounders and average over the distribution.
  + If matching, you’re really pairing up subjects based on the covariates (X), estimating treatment effects within those pairs and averaging over the distribution of X. But you can’t directly match on unobserved confounders.
  + If performing IPTW, you’re creating a pseudo-population where there’s balance on X in the treated and control subjects. But balance can’t be generated on unmeasured confounders just be conditioning on X.

### IVs

* If you’re reasonably sure there’s unmeasured confounding, then the traditional methods aren’t going to work.
* Z is an IV that directly affects treatment but not the outcome i.e. no direct link between Z and Y.
  + 
* Z is ‘randomised’ encouragement where higher values promote a higher probability of treatment and vice versa for lower values.
* There’s some part being explained by a random factor, which provides some hope of being able to estimate a causal effect.
* Sometimes the IV is randomly assigned as part of the study.
* Other times, it’s naturally random e.g. quarter of birth, geographic distance to care provider.

### Example

* A: smoking during pregnancy
* Y: birthweight
* X: parity, mother’s age, weight etc.
* Concern: could be unmeasured confounders that help to more fully capture the decision to smoke and/or the baby’s birthweight.
* Challenge: Unethical to randomly assign smoking to pregnant women.

### Encouragement Design

* Carry out a randomised trial on pregnant smoking women.
  + A: smoking during pregnancy
  + Y: birthweight
  + X: parity, mother’s age, weight etc.
  + Z: randomise to either receive encouragement to stop smoking (intervention arm; Z = 1) or usual care (Z = 0).
* Not directly randomising women to stop smoking or not whilst pregnant, but randomising whether or not a pregnant woman is encouraged to stop smoking as part of her care.
* Z is an instrument because it’s not affecting birthweight directly other than via it’s affect on the exposure (i.e. on whether the subject is convinced to stop smoking or not).
* An intention-to-treat analysis focuses on the causal effect of encouragement:  
  + Compares the outcomes under the two intervention arms.
  + The average birthweight if all subjects had been encouraged to stop smoking vs the average birthweight if all subjects have just received their usual care.
  + Causal effect of encouragement on birthweight.
* IVs try to use this instrument/randomisation to get at the causal effect of smoking i.e the causal effect of the treatment itself.

## Randomised Trials with Noncompliance

### Setup

* Randomised trial:
  + Z: randomisation to treatment (1 if true otherwise 0).
  + A: treatment received (1 if true otherwise 0).
  + Y: outcome.
* Typically not every subject randomised to the treatment group will actually receive/take treatment i.e. non-compliance 100% of the time.
* More interested in the effects of actual treatment on the outcome than assigned treatment.

### DAG

* Non-compliance makes a randomised trial start to look like an observational study.
* Could be confounding based on actual treatment received.

### Potential Treatment

* Observed data: (Z, A, Y)
  + For every subject, we observed their assigned treatment, actual treatment and outcome.
* Just like we previously defined potential outcomes, we can also define potential treatment.
  + : value of treatment if randomised to Z = 1 (could be treatment or not treatment).
  + : value of treatment if randomised to Z = 0 (could be treatment or not treatment).
  + Even before the study begins, every subject can have a value for and even if they’re never observed simultaneously.

### Causal Effect of Assignment on Receipt

* Average causal effect of treatment assignment on treatment received:
  + Proportion treated if all subjects had been assigned to receive treatment minus the proportion treated if no subject was assigned to receive treatment.
  + If there was perfect compliance this would be 1.
* Causal effect because we’re contrasting potential outcomes on a common population, and a population-level causal effect because we’re not conditioning on anything i.e. we’re considering the whole population.
* Generally estimable from the observed data as by randomisation (of Z) and consistency (; i.e. the potential outcome if the subject was assigned Z = 1 corresponds to the actual treatment if the subject was randomised to Z = 1):
  + - If we restrict to the sub-population actually assigned Z = 1 and take the mean of A for that group, it should be the same as the expected value of in the entire population because of to the randomisation.
    - If we stratify and only consider the sub-population that were assigned treatment, that sub-population should have the same characteristics as the whole population because we randomised.
    - Also estimable because the expectation only involves observed data.

### Causal Effect of Assignment on Outcome

* Average causal effect of treatment assignment on the outcome:
  + Average value of the outcome if all subjects were assigned to receive treatment minus the average outcome if no subject were assigned to receive treatment.
* Intention-to-treat effect because it’s effectively contrasting outcomes based on what the subject was meant to do.
* If there was perfect compliance, this would be the same as the causal effect of treatment.
* Can also be estimated from observed data from the randomisation and consistency assumptions:
  + - Average value of the outcome had all subjects in the population been assigned  
      Z = 1. But Z =1 has been randomised, so we can just condition on Z = 1 and take the mean of the observed outcome.
    - Identifiable because both Y and Z are observed data.
* Because we’ve directly randomised Z, we can estimate a causal effect of Z on anything i.e. on the treatment received or the outcome.
* Getting a causal effect of treatment assignment on actual treatment or on the outcome are relatively easy.

### Causal Effect of Treatment

* Typically of more interest is the causal effect of treatment received on the outcome.
* Z can be considered strong encouragement to receive treatment.
  + There could be some non-compliance, but typically there’s binding agreements etc so compliance tends to be reasonably high.
* Treatment assigned should affect treatment received, but we don’t necessarily think that the random assignment necessarily affects the outcome directly.
* The primary goal of IV methods is to exploit the IV randomisation to try and estimate the causal effect of treatment itself.

## Compliance Classes

### Potential Values of Treatment

* Recall that we can classify subjects based on potential treatment:
  + is the treatment received if the subject is randomised to the control condition, instrument was Z = 0, or the subject was randomised to not receive encouragement.
  + is the treatment received is randomised to the Z = 1 group.
* These exist for all subjects, even if they’re not observed.
* General approach known as principal stratification.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Label** | **Explanation** |
| 0 | 0 | Never-takers | No matter what they’re assigned, treatment never taken. |
| 0 | 1 | Compliers | Comply with whatever they’re assigned to do. |
| 1 | 0 | Defiers | Do the opposite of their assignment. |
| 1 | 1 | Always-takers | Always take treatment irrespective of assignment. |

### Sub-populations

* Consider each category as sub-populations of subject.
* Never-takers:
  + Never take treatment, irrespective of randomisation.
  + Encouragement never works.
  + If we knew who was in this population, we wouldn’t learn anything about the causal effect of treatment as there is no variation in treatment received (unless some strong assumptions are made).
    - Never observe an outcome under treatment for this group.
  + In general don’t have variability in treatment in this group.
* Compliers:
  + Take treatment when encouraged to do so and don’t otherwise i.e. treatment received is always the same as treatment assigned.
  + Treatment received is randomised.
  + A lot of promise for learning about the causal effect of treatment since we can randomise treatment received (via assignment).
* Defiers:
  + Will always do the opposite of what they’re encouraged to do.
  + Treatment received is still randomised, but in the opposite direction of the study design.
  + In principle, could also learn about the causal effect of treatment.
  + Typically, we’d like to think that this group is very small or non-existent.
  + In some cases, like randomised trials, the control group may not even have access to the treatment e.g. a new drug, which makes it very hard for this situation to occur.
* Always-takers:
  + Always take treatment, irrespective of randomisation/encouragement.
  + Like the never-takers, there is no variation in the treatment received and hence no information about the causal effect of treatment.

### Causal Effects

* A motivation for IV methods in general is concern about possible unmeasured confounding.
  + If unmeasured confounding exists, we can’t average/marginalise over all confounders (via matching, IPTW etc).
* This is important because previous methods focused on causal effects in the whole population. This is hard to do if we don’t know all the confounders.
* IV methods aim to produce a valid estimate of a causal effect in the presence of unmeasured confounding.
* Achieved by focusing on a local average treatment effect instead of a population-level effect.

### Local Average Treatment Effect

* Target of inference:
* Contrasting means of potential outcomes (treatment assignment): Z = 1 or Z = 0.
* Valid causal effect because we’re comparing the same sub-populations i.e. what we’re conditioning on is exactly the same.
  + Note that this is the sub-population of compliers.
* We’re really looking at the average causal effect of treatment on the sub-population of compliers.
  + Local hence means an inference about a sub-population.
* Switching from indexing by to means we’re comparing potential outcomes based on treatment received instead of treatment assigned. We can do this because:
  + Compliers always do what they’re told so is always the same as and the same for / .
* As long as we restrict to the sub-population of compliers, we can use Z and a interchangeably.
* Complier average causal effect (CACE):
  + Causal effect in a sub-population. Local average treatment effect is the more general term for causal effects in sub-populations, but in the case of a randomised trial with non-compliance present, the more correct term is CACE.
  + ‘Local’ causal effect.
  + No inference about any other sub-population (defiers, never-takers, always-takers).
* An IV analysis is only valid for compliers.

### Observed Data

* We have a target of inference, but that involves potential outcomes and we need to estimate this.
* For every subject, we observe an and , not .

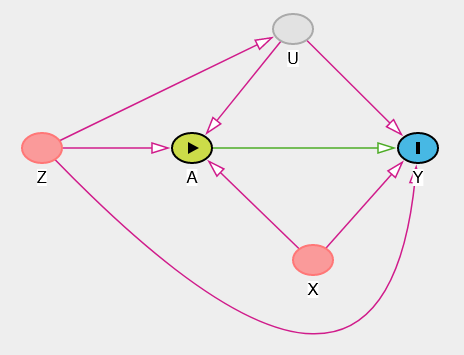
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Z (treatment assigned)** | **A (treatment received)** | **(potential treatment if assigned to control)** | **(potential treatment if assigned to treatment)** | **Class** | **Explanation** |
| 0 | 0 | 0 | ? | Never-takers or compliers | Actually assigned to the control group, and did what they were told.  But don’t know what they would have done had they been assigned to the treatment group. |
| 0 | 1 | 1 | ? | Always-takers or defiers | Actually assigned to the control group but actually took treatment. |
| 1 | 0 | ? | 0 | Never-takers or defiers | Actually assigned to the treatment group, but refused/didn’t take treatment. |
| 1 | 1 | ? | 1 | Always-takers or compliers | Actually assigned to the treatment group and took treatment. |

* Without additional assumption, can’t narrow down the options any further with observed data.

### Identifiability

* Compliance classes also known as principal strata.
  + Classes/labels based on the characteristics of the subjects relative to treatment assignment and treatment taken.
* In order to get the causal effect we want, we stratify on compliers.
  + Latent and not directly observable.
  + So we don’t know for sure which subjects are compliers (or any of the other classes for that matter).
  + Which makes estimation of CACE challenging.

## Assumptions

* A variable is an IV iff:
  + Associated with the treatment.
  + Affects the outcome only through its effect on treatment: exclusion restriction.
    - Cannot directly affect the outcome.
    - In a DAG, there cannot be a direct arrow between Z and Y.
* One reason we could be interested in an IV analysis is because we think there are some unmeasured confounders U.
  + The exclusion restriction also means that Z cannot directly or indirectly affect the outcome via U i.e. Z is affecting Y via a pathway other than that through treatment.
  + 
    - Under the exclusion restriction, the arrows from Z to U and Z to Y cannot exist if we want Z to be a valid instrument.

### Realistic?

* Need to carefully consider if these assumptions are reasonable especially the exclusion restriction assumption (this one tends to be the subject of most debate and/or requires the greatest justification).
* Consider a randomised trial:
  + If Z is random treatment assignment, we should be very confident it affects treatment assignment.
    - Can be checked with data.
  + If we think of the randomisation as a coin flip, there should be no reason it affects the outcome or unmeasured confounders other than via the effect on treatment.
    - But if subjects aren’t blinded i.e. know which group they were assigned to, this could impact their uptake of the treatment/outcome.
    - Similarly, if the researchers aren’t blinded then they could change other aspects of the study e.g. other ways in which the subject is treated.
* Whether the exclusion restriction assumption holds/is reasonable depends on the context of the study.
* Also not technically checkable from data; need to use subject-matter expertise to make a decision similar to determining whether or not all the confounders have been captured.

### Causal Effects

* Assuming the IV is valid, we can use it to estimate the LATE/CACE:
  + The causal effect of treatment among subjects who only take the treatment if they’re randomised to i.e. the causal effect of treatment among compliers.

### Identification Challenge

* Challenge with the identification of causal effects.
* Can label subjects based on their pairs of potential outcome values.
* But we don’t actually know who the compliers are as we don’t actually know their behaviour under the alternative treatment assignment. Can narrow it down to two options for each option.
  + Fundamental problem of causal inference.

### Monotonicity Assumption

* Motivates the need for an additional (common) assumption: there are no defiers.
  + No subject consistently does the opposite of what they’re assigned.
* Really assuming that the probability of treatment increases with encouragement i.e. a positive relationship.
* Thus far have only considered binary IVs, but they can also be continuous.
* Can be a reasonable assumption in many situations.
* Changes the table to the following (and the identifiability problem gets a little easier):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Z** | **A** |  |  | **Class** |
| 0 | 0 | 0 | ? | Never-takers or compliers |
| 0 | 1 | 1 | **1** | Always-takers ~~or defiers~~ |
| 1 | 0 | **0** | 0 | Never-takers ~~or defiers~~ |
| 1 | 1 | ? | 1 | Always-takers or compliers |

* + Two situations now fully identified.
  + Problem is easier to the point that we have enough information to identify the causal effect we’re interested in – causal effect among compliers.

## Causal Effect Identification and Estimation

### Identification of Causal Effects

* Goal is to estimate:
* All the usual assumptions discussed previously:
  + Z is a valid instrument.
  + Monotonicity/no defiers.
  + Exclusion restriction.
* Let’s start with the intention-to-treat (ITT) effect:
  + Contrasting potential outcomes based on treatment assignment i.e. the expected value if all subjects were assigned treatment minus the expected value if no one was assigned treatment.
  + Equality holds due to randomisation.
* Can also express this conditional expectation in terms of the compliance classes:
* Expected value of Y among subjects assigned treatment is a weighted average of the expected value of Y give Z = 1 in the three sub-populations.
* Uses standard rules about expectations and probabilities i.e. can take an expectation conditional on something and then multiply by the probability and sum overall all possibilities to essentially condition and average over it.
  + Weight the expected values (mean) by the contribution of the respective sub-population i.e. the size of that sub-population.
* Note that among aways takers and never takers, Z does nothing (because these sub-populations always do the same thing irrespective of encouragement).
* Can hence drop the conditioning on 1.
* Also, by randomisation, (and the same for never takers).
  + Z and the size of these sub-populations have no relationship to each other as the complier classes pre-date the study/already exist in the target population.
* Simplified:
* Therefore:
  + Average value of the outcome among subjects assigned treatment among the compliers multiplied by the probability that the subject is a complier. And the same goes for the subjects assigned to the control.
* Which implies (dividing both sides by

= CACE

* The third step follows because Z was randomised, we’re looking at the subpopulation of compliers, so by randomising we’re randomising .
* And this is what we wanted – the CACE.
  + is just the proportion of subjects who are either always takers or compliers.
    - A is binary (treatment/control) and the expected value of a binary variable is just a probability.
    - , which is the proportion of people who would take treatment if they were assigned it (which in turn are always takers and compliers).
  + is just the proportion of subjects who are always takers (there are no defiers).
  + Difference is the proportion of compliers as the always takers cancel out.
  + Can estimate the numerator from data as we observe both Y and Z.
  + Can also estimate the denominator from data because we randomised Z so A and Z are observed.
  + Even though we don’t know for sure who the compliers are.
* If there is perfect compliance, CACE = ITT.
* Denominator is always between 0 and 1, so CACE will always be at least as large as ITT.
  + The ITT is an underestimate of the CACE as some people assigned to treatment didn’t take it.
  + The dominator is just an estimate of the proportion of compliers.

## IVs in Observational Studies

### Overview

* Previously discussed IVs only in the context of randomised trials, but IVs can be used in observational (non-randomised) studies.
* Generally view Z as encouragement to receive treatment i.e. it’s not randomised but something encouraging subjects to receive treatment or not.
  + Binary: Yes or no.
  + Continuous: ‘Dose’ of encouragement.

### Identifying an IV

* Z hasn’t been randomised by the researcher, but in nature as a sort of natural experiment.
* Key challenge is to find a variable like that i.e. that affects the treatment but doesn’t affect the outcome.
  + Are only able to check the assumption that the proposed instrument affects treatment.
  + The validity of the exclusion restriction assumption largely relies on subject matter knowledge.

### Example 1: Calendar Time

* Used or proposed in the literature.
* Could be a change in treatment preferences over time.
* Suppose there are two drugs that treat the same condition, but early on Drug A is more likely to be used and later on, Drug B is more likely to be used.
  + May occur if a new drug is introduced into the market and take-up is high over a relatively short period of time.
* At any point in time, who receives Drug A and who receives Drug B is likely not random and dependent on patient characteristics.
* But, calendar time could be essentially randomising as most other covariates could be fairly similar between two time periods if they aren’t very far apart except for which drug is used.
* Is it a valid IV?
  + Certainly associated with treatment received.
  + Exclusion restriction: Possibly met. Could be associated with the outcome if other treatment practices or subject behaviours changed between selected time periods.
    - Stronger argument if the time periods are very close together.
* In general, a lot of focus and critique will be on the exclusion restriction when considering a candidate IV.

### Example 2: Distance

* (Geographic) distance to a speciality centre has been used as an IV for the effect of care on health outcomes.
  + Shorter distance is stronger encouragement (to attend).
  + Is distance likely associated with outcomes in other ways. For example, could people deliberately have moved closer to a care centre? Could people living close to a centre differ from people not living close to a centre?
* Debate would be on whether differential travel times could be associated with the outcome via a pathway other than treatment.

### Other Examples

* Mendelian randomisation: some genetic variant is associated with some behaviour e.g. alcohol use, but isn’t assume to be associated with the outcome of interest.
* Provider preference: use treatment prescribed to previous patients as an IV for current patient. The idea here is that the previous decision should be associated with the current decision, but the previous decision shouldn’t directly affect the outcome.
  + Tend to be seen in medication studies.
  + Provider might be biased/have a preference for a particular drug.
* Quarter of birth: years in school – income.
  + Quarter of birth has been shown to be associated with the number of years in school.
  + Also could be considered randomised.
  + Don’t randomise the number of years in school, but it is probably very strongly associated with lots of other things like socioeconomic status, parental education levels etc.
* Typically involves clever ideas about why it’s a valid instrument, but hence a lot of discussion about whether the exclusion restriction is reasonable.

### Compliers

* Can still think about compliance in these settings.
* What we mean by encouraged really has to do with the instrument used in practice rather than actual compliance with a request/order by a researcher.
* If the instrument is such that the subject was ‘encouraged’ and the subject did receive the treatment, then that is compliance in an observational setting.

## Two-Stage Least Squares

### OLS

* Usual assumption in a standard model is that the error term and explanatory variables are independent. Not particularly important assumption in a correlational setting/model, but critical for causal inference.
* If there is confounding, they are correlated. (In an economics setting, we would say that the independent variables are endogenous/there is endogeneity). So OLS fails.
* OLS still fails if known confounders are controlled for, but there is unmeasured confounding.

### Two Stage Least Squares

* Method for estimating a causal effect in the IV setting.
* Assume Z is a valid instrument (affects treatment and the exclusion restriction is reasonable).

### Stage 1

* Regress treatment received , on the IV :
  + Assume error term is independent with mean 0 and constant variance.
  + By randomisation (of ), and are independent.
* Obtain for each subject using a standard OLS model.
  + This is the predicted value of given (and only ) i.e. predicted treatment given a subject’s characteristics.
* Use a standard OLS to get a predicted value for treatment received.

### Stage 2

* Regress the outcome on the fitted value from stage 1 :
  + Looks like a standard regression except that is predicted instead of actual treatment.
  + Where the error term has mean 0 and constant variance.
* By the exclusion restriction, should be independent of given . only affects through its effect on treatment, so if you condition on treatment, shouldn’t be affecting .
  + Technically is a projection of onto the space spanned by .
  + In other words, should be independent of .
  + There is confounding between and , but is just determined by , which has the exclusion restriction.
* The estimate of is an estimate of the causal effect.
* Both stages are relatively simple to implement in practice if

### 2SLS Estimator

* Delve a little more into why this approach produces a valid causal effect.
* Consider the situation where and are both binary.
  + Stage 1: is an estimate of i.e. the probability of treatment given a particular value of .
  + Stage 2:
* is this contrast of expected values:
* By the time you reach the 2nd stage model, can only take two values: and .
  + This assumes complete compliance, otherwise they will not be 0 and 1 (but somewhere between those lower and upper bounds).
* What we observe is going from to in .
  + Occurs when we go from Z = 0 to Z = 1.
  + Observe a mean difference of (the ITT effect) with a unit change in .
  + is a change in of a full unit, which means is essentially inflating the ITT effect by .
* If we observe a change in the mean of for a unit change in , then we should see a unit change in the mean of for a 1 unit change in .
* The 2SLS estimator is a consistent estimator of the CACE:
  + The denominator is – the slope of the stage 1 model.

### 2SLS More Generally

* Can also be used with covariates and for non-binary data (for example a continuous instruments).
* Stage 1: regress A on Z and covariates X.
  + Obtain fitted value of A:
* Stage 2: regress Y on and X.
  + Coefficient of is the causal effect.
* The covariates to be controlled for are included in both regression equations.

### Sensitivity Analysis

* Good practice to conduct a sensitivity analysis whenever using IV analysis due to the importance of the assumptions for valid results.
  + Sensitivity analysis is generally determining by how much the assumptions are violated for the conclusions to change.
* Methods have been developed for each of the IV assumptions.
  + Exclusion restriction: If Z does directly affect Y by an amount , would the conclusions change? Vary .
  + Monotonicity: If the proportion of defiers was , would my conclusions change?
* Beyond the scope of this course.

## Weak Instruments

### Strength of IVs

* How well an instrument predicts treatment received.
* Strong instruments are highly predictive of treatment.
  + Encouragement greatly increases the probability of treatment.
* Weak instruments are weakly predictive of treatment.
  + Encouragement increases the probability of treatment very little.
* Can measure the strength of an instrument.
  + Estimate the proportion of compliers: .
    - Probability of receiving treatment given you were encouraged minus the probability of receiving treatment given you weren’t encouraged.
  + Can just use the observed proportions of treated subjects for and .
    - Contrast of two proportions.
    - Can get an actual number, which tells us the strength of our instrument.
    - Bounded between 0 and 1, where 1 is a strong instrument and 0 is a weak instrument.

### Weak Instruments

* Problematic.
* Suppose only 1% of the population are compliers.
  + Getting encouraged will only increase the probability of actually getting treatment by a very small amount.
* Recall that in IV analysis, we’re interested in estimating local average treatment effects i.e. local causal effects or CACEs.
  + This means we’re focussed on the subset of compliers.
* If the sample size of the dataset is n, then only 1% of n have useful information about treatment.
* Hence, we’ll have noisy estimates of causal effects, which means very large variances.
* So the estimates are unstable and potentially biased estimates, but the latter is of lesser importance.
* IV analysis may not be the best option.
  + Confidence intervals could be too wide to be useful.
* One of the first things to be done when considering an IV analysis is to check the strength of the relationship between the proposed instrument and the treatment.
* An area of active research are methods to strengthen weak instruments.
  + For example, near/far matching, where matching is performed to make the covariates (confounders) similar, but the instrument is very different (in value).
  + i.e. a matched pair of subjects is similar on their covariates, but have as different as possible values for the instrument.
  + Similar to a randomised trial setting where subjects should have similar covariate distributions but greatly different levels of encouragement.
  + Also newer statistical methods.