Causal Analysis

Impact Evaluation and Causal Machine Learning with Applications in R

Chapter 6: Instrumental Variables

Table of Contents

Motivation (1)

Why instruments?

- The selection-on-observables assumption fails if unobserved factors influence both treatment and outcome (even when controlling for covariates).
- Example: Training program with random assignment but imperfect compliance; some offered training choose not to participate.
- Compliance may depend on unobserved traits (e.g. ability, motivation) that also affect wages.
- Comparing treated vs. untreated does not identify the causal effect, even conditional on observed covariates X.

Motivation (2)

- If random assignment (denoted by Z) satisfies an exclusion restriction (affects Y only through treatment participation D), it can serve as an instrument, see Wright (1928).
- Intuition: randomization identifies the causal effect of Z on Y, which operates only through Z's effect on D.
- The ratio (effect of Z on Y divided by effect of Z on D) yields the causal effect of D on Y for compliers (those only taking the treatment when randomly assigned).
- This effect is the Complier Average Causal Effect (CACE), also called the Local Average Treatment Effect (LATE).

Compliance Types

- IV framework of Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996):
- Binary treatment variable D and instrument Z, such that $d, z \in \{0, 1\}$.
- Potential treatment decision $D(z)$ if the instrument Z takes the value $z \in \{0, 1\}$.
- Individuals satisfying $(D(1) = 1, D(0) = 0)$ are compliers.
- The remaining three groups are noncompliers.

Table 1: Compliance types.

$D(1)$	$D(0)$	Type
1	1	Always takers
1	0	Compliers
0	1	Defiers
0	0	Never takers

Identifying Assumptions (1)

- Potential outcomes $Y(d, z)$ given the treatment $d \in \{0, 1\}$ and instrument $z \in \{0, 1\}$.
- The following instrumental variable assumptions identify the LATE:

$$\{D(z), Y(z', d)\} \perp Z \text{ for } z, z', d \in \{0, 1\}, \quad Y(1, d) = Y(0, d) = Y(d),$$
$$\Pr(D(1) \geq D(0)) = 1, \quad E[D|Z = 1] - E[D|Z = 0] \neq 0. \tag{6.1}$$

- The listed identifying assumptions areindependence, exclusion restriction, monotonicity, and relevance.
- The first line of equation (6.1) is also referred to as IV validity, consisting of both the independence assumption and exclusion restriction.

Identifying Assumptions (2)

Independence

The instrument Z is independent of potential treatments and potential outcomes, such that there are no variables jointly affecting Z and D or Y:

$$\{D(z), Y(z', d)\} \perp Z \text{ for } z, z', d \in \{0, 1\}.$$

Exclusion restriction

The instrument Z does not affect the potential outcome conditional on the treatment, such that the instrument does not have have a direct effect on the outcome Y other than through the treatment D:

$$Y(1, d) = Y(0, d) = Y(d).$$

Identifying Assumptions (3)

Monotonicity

The potential treatment state $D(z)$ of any subject does not decrease in the instrument when switching Z from 0 to 1:

$$\Pr(D(1) \geq D(0)) = 1.$$

Relevance

A first-stage effect of the instrument Z on the treatment D exists:

$$E[D|Z = 1] - E[D|Z = 0] \neq 0.$$

Graphical Illustration

- Consider a causal graph with instrument Z, treatment D, and outcome Y.
- Unobserved variables U might jointly affect D and Y, also called treatment endogeneity.
- The following scenario satisfies IV validity:

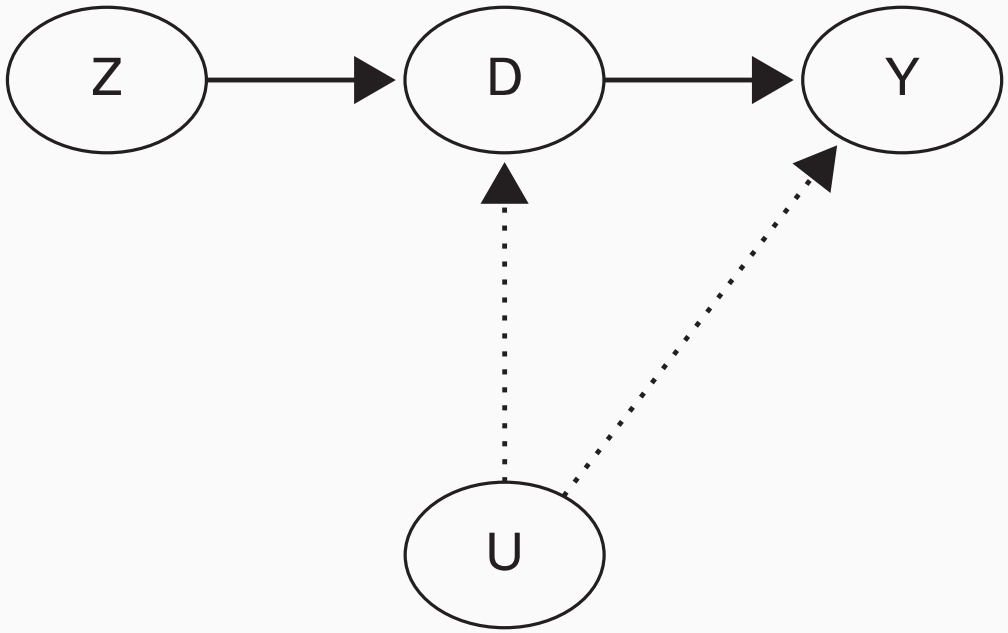

Figure 1: An instrumental variable framework.

Local Average Treatment Effect (1)

- Under the identifying assumptions in equation (6.1), we can identify the LATE among the compliers:

$$\Delta_{D(1)=1,D(0)=0} = E[Y(1) - Y(0)|D(1) = 1, D(0) = 0]. \tag{6.3}$$

- For estimation of an average effect like LATE, the identifying assumptions can be relaxed.
- For average effects, we do not have to impose full independence, but the weaker mean independence suffices: $E[Y(z,d)|D(1),D(0),Z=1] = E[Y(z,d)|D(1),D(0),Z=0] = E[Y(z,d)|D(1),D(0)]$ for $z,d \in \{0,1\}$.
- For average effects, also the exclusion restriction can be relaxed as follows: $E[Y(1,d)|D(1),D(0)] = E[Y(0,d)|D(1),D(0)] = E[Y(d)|D(1),D(0)]$.

Local Average Treatment Effect (2)

- Under the assumptions in equation (6.1), a binary instrument and treatment, the LATE is identified based on $E[Y|Z=1] - E[Y|Z=0]$ and $E[D|Z=1] - E[D|Z=0]$.
- The first expression corresponds to the intention to treat (ITT) effect of the instrument on the outcome.
- The second expression is the first-stage effect or share of compliers.
- The ITT corresponds to the first-stage effect multiplied by the LATE among the compliers:

$$E[Y|Z=1] - E[Y|Z=0] = \Delta_{D(1)=1,D(0)=0} \cdot (E[D|Z=1] - E[D|Z=0])$$
$$\Leftrightarrow \Delta_{D(1)=1,D(0)=0} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}. \quad (6.4)$$

Local Average Treatment Effect (3)

- As shown in equation (6.4), the LATE for compliers equals the ITT effect divided by the first-stage effect.
- This ratio is called the Wald estimand.
- Alternatively, one can (1) regress Y and D on a constant and Z and (2) divide the coefficient on Z in the outcome regression by the coefficient on Z in the treatment regression.
- Two-stage least squares (TSLS) regression is another alternative, regressing D on a constant and Z (first stage) and then Y on a constant and the predicted treatment from the first stage.
- All these approaches yield the same $\sqrt{n}$-consistent and asymptotically normal estimator of the LATE among the compliers under mild statistical conditions.

Practical Considerations

- TSLS has the advantage that it directly yields the standard error of the LATE estimate, which accounts for both estimation uncertainty in the first- and second-stage regression.
- Moreover, TSLS can be used to estimate the potential outcomes (rather than the effect) among compliers.
- Weak instrument problem: If the first-stage effect approaches zero, the Wald estimand goes to infinity, which implies that the variance of the LATE estimation explodes.
- Under a weak instrument, conventional standard errors and confidence intervals may be unreliable and misleading, particularly in small samples, see Staiger and Stock (1997).
- For alternative approaches to inference under weak instruments, see Anderson and Rubin (1949), Stock, Wright, and Yogo (2002), and Keane and Neal (2021), among others.

Table of Contents

Conditionally Valid Instruments

In many empirical applications, the identifying assumptions might not hold unconditionally, i.e., without controlling for covariates.

Example

- Card (1995) uses geographic proximity to college as instrument Z for education D to assess earnings Y.
- Proximity likely reduces education costs, indicating a first-stage effect of proximity on education.
- But proximity reflects a neighborhood's socioeconomic status, likely influencing earnings.
- Consequently, identifying assumptions may not hold due to non-random instrument.
- Control for covariates (e.g., parent's education) affecting both Z and Y to make identifying assumptions more credible.

Identifying Assumptions

- Assumptions for identifying the LATE conditional on covariates X (Abadie, 2003):

$$\{D(z), Y(z', d)\} \perp Z|X \text{ for } z, z', d \in \{0,1\}, \Pr(Y(1,d) = Y(0,d) = Y(d)|X) = 1,$$
$$\Pr(D(1) \geq D(0)|X) = 1, \quad E[D|Z=1,X] - E[D|Z=0,X] \neq 0,$$
$$X(1) = X(0) = X, \quad 0 < P(Z=1|X) < 1. \tag{6.5}$$

- First line in equation (6.5) requires the IV validity assumptions of equation (6.1) to hold conditional on X.
- Second line in equation (6.5) rules out the existence defiers and requires the existence of compliers conditional on X.
- Third line, first expression in equation (6.5) requires that the treatment D does not affect X.
- Third line, second expression in equation (6.5) assumes common support in the instrument propensity score $P(Z=1|X)$.

Graphical Illustration

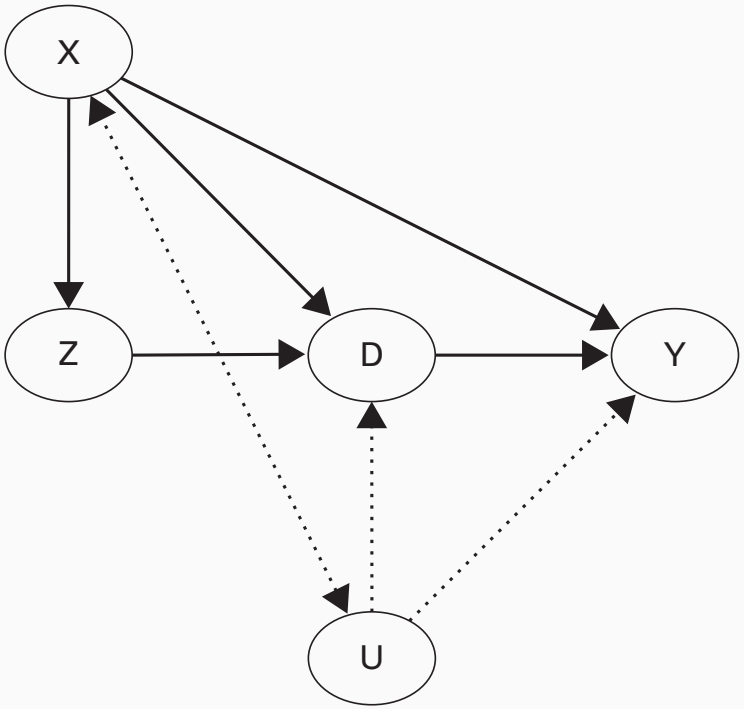

Figure 2: An instrumental variable framework with covariates.

- Unobservables U may affect X, or vice versa.
- Conditional on X, no unobservables jointly affect Z and Y.

LATE with Covariates (1)

LATE is identified based on the ratio of the ITT and first-stage effects:

$$\Delta_{D(1)=1,D(0)=0} = \frac{E[E[Y|Z=1,X] - E[Y|Z=0,X]]}{E[E[D|Z=1,X] - E[D|Z=0,X]]} \tag{6.7}$$
$$= \frac{E[Y \cdot Z/\Pr(Z=1|X) - Y \cdot (1-Z)/(1-\Pr(Z=1|X))]}{E[D \cdot Z/\Pr(Z=1|X) - D \cdot (1-Z)/(1-\Pr(Z=1|X))]} = \frac{\theta}{\gamma}$$

- Estimation may be based on matching (Frölich, 2007), conditional mean regression, IPW (Donald, Hsu, and Lieli, 2014), or DR (Tan, 2006) and is $\sqrt{n}$-consistent and asymptotically normal under certain regularity conditions.
- CML/DML for LATE estimation (Belloni, Chernozhukov, Fernández-Val, and Hansen, 2017) or investigating effect hetorgeneity (Athey, Tibshirani, and Wager, 2019).
- Frölich and Melly (2013) propose an estimator of local quantile treatment effects (LQTEs).

External Validity (1)

External validity

External validity refers to how representative a causal effect (such as the LATE) is for effects in other populations. In other words, it concerns the generalizability of the effect to other populations.

- Limitations of LATE: The LATE among compliers may be of little relevance because it only refers to the subpopulation of compliers.
- ATE in the total population is typically more interesting.
- Based on covariate information, one might attempt to generalize from the effect among compliers to draw conclusions about the effect in the full population.

External Validity (2)

Verifying covariate distributions

- One approach to evaluating the plausibility of external validity consists in comparing the distribution of covariates among compliers with that of the total population.
- To this end, the κ-weighting approach of Abadie (2003) for the identification of complier-related statistics can be applied:

$$\kappa = 1 - \frac{D \cdot (1 - Z)}{1 - \Pr(Z = 1|X)} - \frac{(1 - D) \cdot Z}{\Pr(Z = 1|X)} \tag{6.8}$$

- For instance, $\frac{E[\kappa \cdot X]}{E[\kappa]} = E[X \mid D(1) = 1, D(0) = 0]$ gives the covariate means among compliers.
- Similar covariates may increase confidence that effects among compliers might be comparable to effects in total population.
- However, treatment effects might also vary with unobserved factors not included in X.

External Validity (3)

Homogeneity in effects

- Another approach consists in assuming homogeneous average effects across compliance types conditional on X, see see Angrist and Fernández-Val (2010) and Aronow and Carnegie (2013):

$$E[Y(1) - Y(0)|D(1), D(0), X] = E[Y(1) - Y(0)|X]. \tag{6.9}$$

- Assumptions in equation (6.5) and (6.9) permit identifying ATE:

$$\Delta = E[\Delta_{D(1)=1,D(0)=0,X}] = E\left[\frac{E[Y|Z=1,X] - E[Y|Z=0,X]}{E[D|Z=1,X] - E[D|Z=0,X]}\right]. \tag{6.10}$$

- In the presence of multiple instruments, conditional effect homogeneity is testable.

External Validity (4)

Homogeneity in potential outcomes

- Further assumption establishing external validity of the LATE: different compliance types have the same mean potential outcomes, at least conditional on X, see Angrist (2004) and de Luna and Johansson (2014).
- This resembles a selection-on-observables assumption for the treatment.
- A statistically significant association of Z and Y, conditional on D and X, suggests a violation of homogeneous potential outcomes.
- Donald, Hsu & Lieli (2014) suggest a related test of whether the IV-based LATE differs from the selection-on-observables-based ATET (see chapter 4) under one-sided noncompliance, ruling out always takers and defiers.

External Validity (5)

Rank invariance or similarity, see Chernozhukov and Hansen (2005)

- Assumption: $\text{rank}(Y(1)) = \text{rank}(Y(0))$ (rank invariance), or at least no systematic differences (rank similarity).
- Permits identifying quantile and average treatment effects on continuous outcomes in the total population, even without imposing monotonicity.
- However, such assumptions substantially restrict treatment effect heterogeneity.
- Example: Consider education choice (college vs. vocational training) as treatment. Rank invariance implies that a college graduate would have the same rank in the wage distribution under vocational training.
- This may be unrealistic if individuals are systematically more competitive in one track than in the other.

Table of Contents

Multivalued or Continuous Instrumental Variables

- Consider multivalued instead of binary instrument, while treatment remains binary.
- Multivalued instruments may possibly be continuous.
- We may asses the LATE for any pair of instrument values z' and z (which might be different from 0 and 1) that satisfy the IV assumptions.
- For example, consider a medical treatment instrumented by randomized cash incentives with values z' (e.g. 20 USD) and z (10 USD).

Propensity Score-Based Approach

- Instead of directly using instrument Z, we may alternatively consider the treatment propensity score as instrument, defined as $p(Z, X) = \Pr(D = 1|Z, X)$.
- Using the propensity score approach, the LATE is identified by:

$$\Delta_{D(1)=1,D(0)=0} = \frac{E[E[Y|p(Z,X) = p(z',X), X] - E[Y|p(Z,X) = p(z,X), X]]}{E[E[D|p(Z,X) = p(z',X), X] - E[D|p(Z,X) = p(z,X), X]]} \quad (6.11)$$

Propensity Score with Multiple Instruments

- Propensity score-based approach appears attractive if Z consists of multiple instruments (e.g., cash transfers and geographic proximity), which are collapsed into a single score $p(Z, X)$.
- However, monotonicity and common support must hold for the newly created instrument $p(Z, X)$, not just one instrument.
- Problem in terms of monotonicity: rules out cases where subjects comply with only one instrument.
- Mogstad, Torgovitsky, and Walters (2020) propose and discuss identification under weaker partial monotonicity, i.e., monotonicity in one instrument conditional on the other.

Continuous Instrument and Marginal Treatment Effect

- Marginal change in a continuous instrument yields the marginal treatment effect (MTE) (Heckman and Vytlacil, 1999, 2001, 2005).
- MTE is the average treatment effect conditional on the covariates X and unobserved term V affecting treatment decision:

$$\Delta_{x,v} = E[Y(1) - Y(0)|X = x, V = v]. \tag{6.12}$$

- MTE in equation (6.12) can be estimated by the local IV (LIV):

$$\Delta_{X=x,\bar{V}=p(z,x)} = \frac{\partial E[Y|X = x, p(Z, X) = p(z, x)]}{\partial p(z, x)}. \tag{6.13}$$

- MTEs are identified under the assumptions in equation (6.5).
- Very strong, continuous instruments allow assessing MTEs for all feasible values of X and V (yielding the ATE), but are hard to find.
- Consequently, MTE is generally only identified over the common support of $p(Z, X)$ across all values of X.

Graphical Illustration

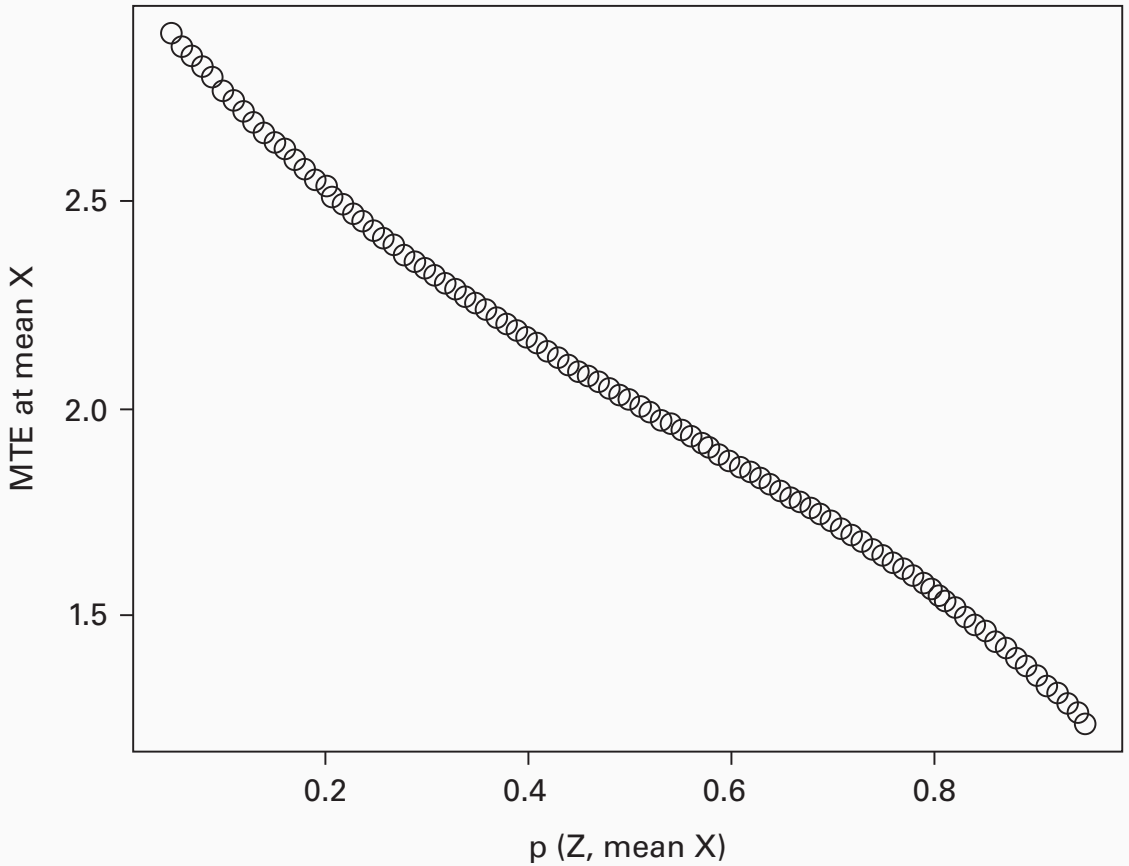

Figure 3: MTEs

Identifying the ATE Despite Limited Common Support

Despite common support issues one might identify the ATE:

- By replacing the independence assumption $\{D(z), Y(z', d)\} \perp Z|X$ by a much stronger version (Carneiro, Heckman, and Vytlacil, 2011):

$$\{D(z), Y(z', d)\} \perp (Z, X) \text{ for } z, z' \text{ in the support of } Z, \tag{6.14}$$

  and thus imposing independence between X and U.
- By imposing parametric assumptions like a linear change of the MTE across values of $p(Z, X)$ (Brinch, Mogstad, and Wiswall, 2017).
- By imposing additive separability in treatment effect heterogeneity caused by covariates X on the one hand and unobserved characteristics on the other hand.

Multivalued Treatments

- Consider an ordered treatment with multiple values $D \in \{0, 1, 2, ..., J\}$, while maintaining a binary instrument.
- Evaluating the effects among complier groups at specific treatment values is not straightforward.
- However, if $\Pr(D(1) \geq j > D(0)) > 0$ for some treatment value j such that compliers exist at some treatment margin, then the Wald estimand equals (Angrist and Imbens, 1995):

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} = \sum_{j=1}^{J} w_j \cdot E[Y(j) - Y(j-1)|D(1) \geq j > D(0)] \quad (6.15)$$

  with weights $w_j = \frac{\Pr(D(1) \geq j > D(0))}{\sum_{j=1}^{J} \Pr(D(1) \geq j > D(0))}$, implying that $0 \leq w_j \leq 1$ and $\sum_{j=1}^{J} w_j = 1$.
- The complier groups contributing to the weighted effect might be overlapping and some compliers might be accounted for multiple times, compromising interpretability.

Binarized Treatments

- Temptation with a multivalued treatment: reduce it to a binary treatment (e.g., high vs. low training; tertiary vs. no tertiary education).
- However, binarization violates the IV exclusion restriction if the instrument affects the treatment at margins not captured by the binarized treatment (Andresen and Huber, 2021).
- Example: Instrument affects the decision of upper vs. lower secondary education, while the binarized treatment captures tertiary vs. no tertiary education.

Unordered Treatments

- Unordered treatments are equivalent to multiple mutually exclusive options (e.g., $D = 1$: IT course, $D = 2$: sales training).
- They pose challenges to (the credibility of) monotonicity assumptions.
- Behaghel, Crépon, and Gurgand (2013) consider three-valued $(D, Z) \in \{0, 1, 2\}$ and consider the following monotonicity assumption:
  - $Z : 0 \to 1$ affects choice 1 vs. 0, not 2.
  - $Z : 0 \to 2$ affects choice 2 vs. 0, not 1.
- Heckman and Pinto (2018) assume that if some move into (out of) a treatment when Z changes, no one moves out of (into) it simultaneously.
- These examples demonstrate that monotonicity conditions with multiple unordered treatments are more complex than in the binary case.

Table of Contents

Nonrandom Outcome Attrition and Sample Selection

Related to chapter 4.11, nonrandom outcome attrition and sample selection can complicate treatment evaluation:

- Nonrandom outcome attrition, e.g., nonresponses in follow-up survey in which the outcome is measured.
- Sample selection, e.g, when wage outcomes are only observed conditional on selection into employment.

Impose further assumptions to tackle this issue:

- Missing-at-random (MAR): Conditional independence of the attrition or sample selection and Y, given Z, D, and X.
- Latent ignorability (LI; Frangakis and Rubin, 1999): Conditional independence given compliance type (complier, always, or never taker).
- Combination of MAR and LI: Conditional independence given both observed characteristics and the compliance type.

IV Approach with Attrition and Sample Selection

- Nonignorable nonresponse or Heckman-type sample selection models allow for more general association of attrition or sample selection (denoted by O) and unobservables affecting the outcome than LI (and its combination with MAR).
- But they generally require an additional instrument (denoted by Q) for O which does not affect the outcome Y.
- LATE evaluation requires further assumptions like parametric restrictions on the outcome model, specific (e.g., monotonicity) conditions concerning the effect of instrument Q, or both.
- For instance, Fricke, Frölich, Huber, and Lechner (2020) consider a continuous instrument Q for sample selection/attrition O, in addition to the binary instrument Z for treatment D.

Graphical Illustration

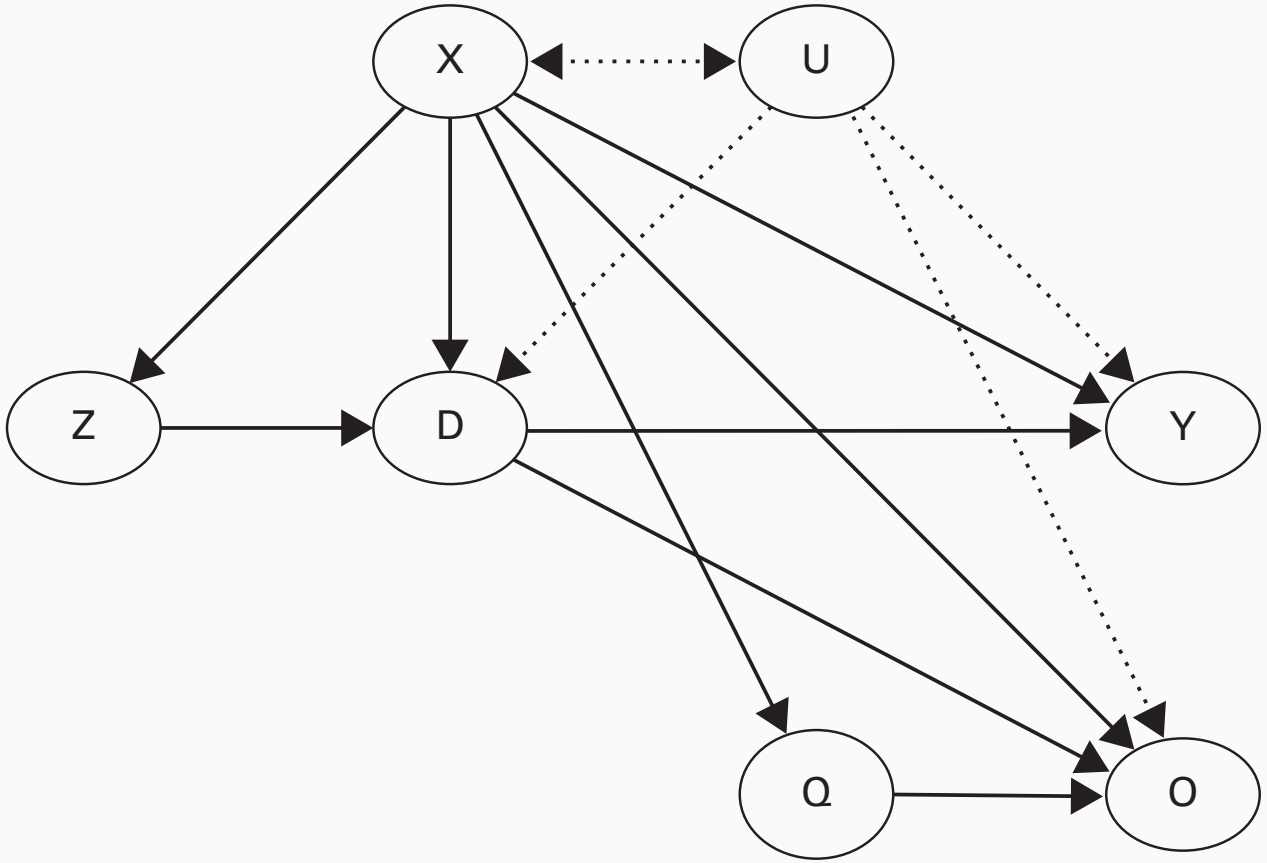

Figure 4: Causal paths with two separate instruments for the treatment and attrition.

IV Approach with Multiple Treatments

- As in chapter 4.9, consider evaluating the impact of several sequentially assigned treatments.
- This generally requires multiple instruments for each treatment and further assumptions.
- In a multiple treatment framework, the different treatments are not assigned sequentially, but in the same period.
- Identification generally requires different instruments for each treatment, too (Blackwell, 2015) .
- As in chapter 4.10, consider disentangling the direct effect and indirect effect (through a mediator) of a treatment on an outcome.
- Identification in general requires distinct instruments for the treatment and the mediator (Frölich and Huber, 2017), along with further assumptions.