Non-Random Sample Selection: Doing the Two-Step From Heckman to Instrumental Variables

Charlie Gibbons
Political Science 236

February 18, 2009

Outline

1 Heckman two-step
- Framework
- Estimation
- Issues

2 Instrumental variables
- Framework
- Identification
    - LATE
    - Compliers
- Assumptions

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association.* 91(434): 444–455.

Angrist, Joshua D. and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist’s Companion.*

Gibbons, Charles E., Juan Carlos Suárez Serrato, and Mike Urbancic. 2009. “LATE for School: Instrumental Variables and the Returns to Education.” Working paper.

Heckman, James J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica.* 47(1): 153–161.

The problem of selection

Economists have long recognized the problem of selection or endogeneity.

P.G. Wright (1928) was the first use of instrumental variables in order to overcome this problem in identifying parameters in supply and demand curves.

Loosely speaking, endogeneity means that there are unobserved qualities of individuals that generate different returns to treatment and thus individuals self-select into treatment levels based upon these unobserved characteristics.

Heckman two-step

Though the notion of instrumental variables has been around since the 1920s, wide application is a relatively recent development.

Heckman two-step

Though the notion of instrumental variables has been around since the 1920s, wide application is a relatively recent development.

Heckman pioneered thinking about selection and creating econometric models to deal with this issue, leading to the development of the *Heckman two-step* or *Heckit* selection model.

Heckman two-step

The canonical example is estimating the determinants of an individual's hours worked w.

Heckman two-step

The canonical example is estimating the determinants of an individual's hours worked w.

We only have w for employed individuals in our sample.

Heckman two-step

The canonical example is estimating the determinants of an individual's hours worked w.

We only have w for employed individuals in our sample.

How do we proceed?

1. Create a model of who is selected (*e.g.*, who is employed).
2. Then, given that model, model their outcomes (*e.g.*, hours worked).

Heckman two-step

The first stage is the *participation equation*:

$$\begin{aligned} D^* &= Z\gamma + \eta \\ D &= \begin{cases} 1 & \text{if } D^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ \Pr(D = 1|Z) &= \Phi(Z\gamma) \end{aligned}$$

Heckman two-step

The first stage is the *participation equation*:

$$\begin{aligned} D^* &= Z\gamma + \eta \\ D &= \begin{cases} 1 & \text{if } D^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ \Pr(D = 1|Z) &= \Phi(Z\gamma) \end{aligned}$$

The *outcome equation* is:

$$w^* = X\beta + \epsilon$$

Heckman two-step

The first stage is the *participation equation*:

$$\begin{aligned} D^* &= Z\gamma + \eta \\ D &= \begin{cases} 1 & \text{if } D^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ \Pr(D = 1|Z) &= \Phi(Z\gamma) \end{aligned}$$

The *outcome equation* is:

$$w^* = X\beta + \epsilon$$

We observe:

$$w = \begin{cases} w^* & \text{if } D = 1 \\ 0 & \text{if } D = 0 \end{cases}$$

Heckman two-step

Now, because we only have hours worked for employed people, we have the moment condition:

$$\mathbb{E}(w|D = 1) = X\beta + \mathbb{E}(\epsilon|D = 1)$$

Heckman two-step

Now, because we only have hours worked for employed people, we have the moment condition:

$$\mathbb{E}(w|D = 1) = X\beta + \mathbb{E}(\epsilon|D = 1)$$

The expectation of the error term is *not* 0 due to selection, so standard OLS would be biased.

Heckman two-step

Now, because we only have hours worked for employed people, we have the moment condition:

$$\mathbb{E}(w|D = 1) = X\beta + \mathbb{E}(\epsilon|D = 1)$$

The expectation of the error term is *not* 0 due to selection, so standard OLS would be biased.

Heckman's primary insight was to think of this as a missing variable problem.

Heckman two-step

Now, because we only have hours worked for employed people, we have the moment condition:

$$\mathbb{E}(w|D=1) = X\beta + \mathbb{E}(\epsilon|D=1)$$

The expectation of the error term is *not* 0 due to selection, so standard OLS would be biased.

Heckman's primary insight was to think of this as a missing variable problem.

His solution? Model the missingness.

Heckman two-step

Now, because we only have hours worked for employed people, we have the moment condition:

$$\mathbb{E}(w|D=1) = X\beta + \mathbb{E}(\epsilon|D=1)$$

The expectation of the error term is *not* 0 due to selection, so standard OLS would be biased.

Heckman's primary insight was to think of this as a missing variable problem.

His solution? Model the missingness.

Big assumption: The error from the models of D, η, and w, u, are distributed *jointly normal*.

Estimation

To understand the remainder of the method, we must know a statistical fact.

Suppose that X is a random variable distributed $\mathrm{N}(\mu, \sigma^2)$ and a is an arbitrary threshold value. Then, the following holds:

$$\begin{aligned}\mathbb{E}(X|X > a) &= \mu + \sigma \underbrace{\left[\frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right]}_{\text{Inverse Mills Ratio}} \\ &= \mu + \sigma\lambda\left(\frac{a-\mu}{\sigma}\right)\end{aligned}$$

Estimation

Now let's bring this back to the model. We have:

$$\begin{aligned}
\mathbb{E}(\epsilon|D=1) &= \mathbb{E}(\epsilon|Z\gamma+\eta \geq 0) \\
&= \mathbb{E}(\epsilon|\eta \geq -Z\beta) = \mathbb{E}(\sigma_{\epsilon,\eta}\eta+\xi|\eta \geq -Z\beta) \\
&= \frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}\left[\frac{\phi\left(\frac{-Z\gamma}{\sigma_\eta}\right)}{1-\Phi\left(\frac{-Z\gamma}{\sigma_\eta}\right)}\right] = \frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}\left[\frac{\phi\left(\frac{Z\gamma}{\sigma_\eta}\right)}{\Phi\left(\frac{Z\gamma}{\sigma_\eta}\right)}\right] \\
&= \frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}\lambda\left(\frac{-Z\gamma}{\sigma_\eta}\right)
\end{aligned}$$

Note that we use $\frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}$ rather than just σ_η to account for the correlation between ϵ and η (Note: ξ is independent of η).

Estimation

So, in the second stage, we run the regression:

$$w = X\beta + \lambda\left(\frac{Z\hat{\gamma}}{\sigma_\eta}\right)\delta,$$

where $\delta = \frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}$.

Estimation

So, in the second stage, we run the regression:

$$w = X\beta + \lambda\left(\frac{Z\hat{\gamma}}{\sigma_\eta}\right)\delta,$$

where $\delta = \frac{\sigma_{\epsilon,\eta}}{\sigma_\eta}$.

OLS gives unbiased estimates of β and δ, but they are not efficient. For efficient standard errors, you need to take into account the fact that γ is estimated and not the true value plus heteroskedasticity. For proofs, see, *e.g.*, Heckman (1979).

Estimation

To implement in `R`, use the `sampleSelection` package.

Issues

What are the downsides of this procedure?

- It is *very* dependent upon the joint normality assumption due to the incorporation of the inverse Mills ratio.
- Assumes that γ can be estimated consistantly; *i.e.*, there is no selection bias in the probit model. This (essentially) requires an instrument.

Issues

What are the downsides of this procedure?

- It is *very* dependent upon the joint normality assumption due to the incorporation of the inverse Mills ratio.
- Assumes that γ can be estimated consistantly; *i.e.*, there is no selection bias in the probit model. This (essentially) requires an instrument.

Issues

What are the downsides of this procedure?

- It is *very* dependent upon the joint normality assumption due to the incorporation of the inverse Mills ratio.
- Assumes that γ can be estimated consistantly; *i.e.*, there is no selection bias in the probit model. This (essentially) requires an instrument.

Issues

What are the downsides of this procedure?

- It is *very* dependent upon the joint normality assumption due to the incorporation of the inverse Mills ratio.
- Assumes that γ can be estimated consistantly; *i.e.*, there is no selection bias in the probit model. This (essentially) requires an instrument.

An instrumental variables procedure overcomes the first issue and relies less on the modeling assumptions generally.

Issues

What are the downsides of this procedure?

- It is *very* dependent upon the joint normality assumption due to the incorporation of the inverse Mills ratio.
- Assumes that γ can be estimated consistantly; *i.e.*, there is no selection bias in the probit model. This (essentially) requires an instrument.

An instrumental variables procedure overcomes the first issue and relies less on the modeling assumptions generally.

Both permit using data with *selection on unobservables*.

Framework

Last semester, we looked at IV in the RCM framework. Here, we place it into a selection model framework.

Assume that, for an individual, there are two potential treatment types 0 and 1 and outcomes Y in each state is a (state-dependent) function of observed covariates X and unobserved factors U.

$$\begin{aligned} Y_1 &= \mu_1(X, U_1) \\ Y_0 &= \mu_0(X, U_0) \end{aligned}$$

Framework

Last semester, we looked at IV in the RCM framework. Here, we place it into a selection model framework.

Assume that, for an individual, there are two potential treatment types 0 and 1 and outcomes Y in each state is a (state-dependent) function of observed covariates X and unobserved factors U.

$$\begin{aligned} Y_1 &= \mu_1(X, U_1) \\ Y_0 &= \mu_0(X, U_0) \end{aligned}$$

An individual will opt into treatment 1 if $Y_1 \geq Y_0$.

Framework

Since U is unobservable, we must create a new model of Y, μ_D, which is a function only of observable covariates X and instruments Z.

Framework

Since U is unobservable, we must create a new model of Y, μ_D, which is a function only of observable covariates X and instruments Z.

$$\begin{array}{rcl} D^* & = & \mu_D(Z) - U_D \\ D & = & \begin{cases} 1 & \text{if } \; D^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \end{array}$$

Framework

U_D is normalized to be distributed $\text{Uniform}[0, 1]$ and μ_D is assumed to be an arbitrary and unknown function on the same interval.

Framework

U_D is normalized to be distributed Uniform$[0,1]$ and μ_D is assumed to be an arbitrary and unknown function on the same interval.

This implies that μ_D is an individual's *p-score*:

$$\begin{aligned}\Pr(D=1\,|X,Z) &= \Pr(D^*\geq 0\,|X,Z)=\Pr(\mu_D(Z)-U_D\geq 0\,|X,Z)\\ &= \Pr(U_D\leq \mu_D(Z)\,|X,Z)=\mu_D(Z)\end{aligned}$$

Framework

U_D is normalized to be distributed Uniform$[0,1]$ and μ_D is assumed to be an arbitrary and unknown function on the same interval.

This implies that μ_D is an individual's *p-score*:

$$\begin{aligned}\Pr(D=1\,|X,Z) &= \Pr(D^*\geq 0\,|X,Z)=\Pr(\mu_D(Z)-U_D\geq 0\,|X,Z)\\ &= \Pr(U_D\leq \mu_D(Z)\,|X,Z)=\mu_D(Z)\end{aligned}$$

Question: What happens when $U_D=0$? What does this mean?

LATE

Now let's consider treatment effects. Define $\Delta = Y_1 - Y_0$. Then,

$$\Delta^{ATE}(x) \equiv \mathbb{E}[\Delta | X = x].$$

LATE

Now let's consider treatment effects. Define $\Delta = Y_1 - Y_0$. Then,

$$\Delta^{ATE}(x) \equiv \mathbb{E}[\Delta | X = x].$$

This would yield a true treatment effect if treatment could be randomly assigned among individuals with $X = x$, under full compliance, and in the absence of general equilibrium effects. Under these assumptions, OLS estimates the ATE.

LATE

We need to remember how we are identifying the treatment effect—we are using variation in the *instrument*, not the variable of interest, to identify the effect. If the instrument doesn't change behavior, then we can't identify the treatment effect.

LATE

We need to remember how we are identifying the treatment effect—we are using variation in the *instrument*, not the variable of interest, to identify the effect. If the instrument doesn't change behavior, then we can't identify the treatment effect.

Instead of estimating the ATE, IV can only estimate the effect for individuals whose behaviors change as the instrument changes (*i.e.*, we can't say anything about always- or never-takers).

LATE

How do we see this? Let's go to the reduced-form framework. Suppose we set a dichotomous instrument equal to 1 for one individual and let $D_i(1)$ be his response and set the instrument to 0 for another. Then his treatment effect is:

$$
\begin{aligned}
Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right) &= \left[Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))\right] \\
&- \left[Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))\right] \\
&= \left(Y_i(1) - Y_i(0)\right) \cdot \left(D_i(1) - D_i(0)\right)
\end{aligned}
$$

LATE

How do we see this? Let's go to the reduced-form framework. Suppose we set a dichotomous instrument equal to 1 for one individual and let $D_i(1)$ be his response and set the instrument to 0 for another. Then his treatment effect is:

$$\begin{aligned} Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right) &= \left[Y_i(1) \cdot D_i(1) + Y_i(0) \cdot (1 - D_i(1))\right] \\ &- \left[Y_i(1) \cdot D_i(0) + Y_i(0) \cdot (1 - D_i(0))\right] \\ &= (Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0)) \end{aligned}$$

This is non-zero ($i.e.$, identified) when both the treatment effect is non-zero (first set of parentheses) and the instrument changes his behavior (second set of parentheses).

LATE

We can get the population expectation of the effect:

$$
\begin{aligned}
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0))\right] = \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = 1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = 1\right] + \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = -1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = -1\right]
\end{aligned}
$$

LATE

We can get the population expectation of the effect:

$$
\begin{aligned}
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0))\right] = \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = 1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = 1\right] + \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = -1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = -1\right]
\end{aligned}
$$

The first conditional expectation is for compliers and the second is for defiers.

LATE

We can get the population expectation of the effect:

$$\begin{aligned}
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \cdot (D_i(1) - D_i(0))\right] = \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = 1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = 1\right] + \\
&\mathbb{E}\left[(Y_i(1) - Y_i(0)) \,|\, (D_i(1) - D_i(0)) = -1\right] \\
&\cdot \Pr\left[(D_i(1) - D_i(0)) = -1\right]
\end{aligned}$$

The first conditional expectation is for compliers and the second is for defiers.

The monotonicity assumption (*i.e.*, the instrument moves everyone toward or away from treatment) lets us assume that defiers don't exist.

LATE

Now we can calculate the *intention to treat* estimator, which is the IV estimator:

$$\begin{aligned}
\Delta^{ITT} &= \frac{\mathbb{E}\left[Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right)\right]}{\mathbb{E}\left[D_i(1) - D_i(0)\right]} \\
&= \frac{\mathbb{E}\left[Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right)\right]}{\Pr\left[D_i(1) - D_i(0) = 1\right]} \\
&= \mathbb{E}\left[\left(Y_i(1) - Y_i(0)\right) \,|\, \left(D_i(1) - D_i(0)\right) = 1\right] \\
&\equiv \Delta^{LATE}
\end{aligned}$$

LATE

Now we can calculate the *intention to treat* estimator, which is the IV estimator:

$$\begin{aligned}
\Delta^{ITT} &= \frac{\mathbb{E}\left[Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right)\right]}{\mathbb{E}\left[D_i(1) - D_i(0)\right]} \\
&= \frac{\mathbb{E}\left[Y_i\left(D_i(1)\right) - Y_i\left(D_i(0)\right)\right]}{\Pr\left[D_i(1) - D_i(0) = 1\right]} \\
&= \mathbb{E}\left[\left(Y_i(1) - Y_i(0)\right) \mid \left(D_i(1) - D_i(0)\right) = 1\right] \\
&\equiv \Delta^{LATE}
\end{aligned}$$

IV gives us the *local average treatment effect* (LATE) and the value of our estimate depends upon the underlying unobserved variation.

LATE

Returning to the structural framework that we used earlier, we can write this expression more generally:

$$\begin{aligned}\Delta^{LATE}\left(x,\mu_D(z),\mu_D\left(z'\right)\right) &\equiv \mathbb{E}[\Delta|X=x,D_z=1,D_{z'}=0] \\ &= \mathbb{E}\left[\Delta|X=x,\mu_D\left(z'\right)<U_D\leq\mu_D(z)\right]\end{aligned}$$

LATE

Returning to the structural framework that we used earlier, we can write this expression more generally:

$$\begin{aligned}\Delta^{LATE}\left(x, \mu_D(z), \mu_D\left(z'\right)\right) &\equiv \mathbb{E}[\Delta | X = x, D_z = 1, D_{z'} = 0] \\ &= \mathbb{E}\left[\Delta | X = x, \mu_D\left(z'\right) < U_D \leq \mu_D(z)\right]\end{aligned}$$

It is estimated for the compliers, defined here as

$\{U_D : \mu_D\left(z'\right) < U_D \leq \mu_D(z)\}$.

LATE

We can write the ATE and the LATE using the *marginal treatment effect* (MTE) for individuals with given levels of observed and unobserved covariates:

$$\Delta^{ATE}(x) = \int_0^1 \Delta^{MTE}(x, u_D)\, du_D, \text{ and}$$

$$\Delta^{LATE}\left(x, \mu_D(z), \mu_D\left(z'\right)\right) =$$

$$\frac{1}{\mu_D(z) - \mu_D\left(z'\right)} \int_{\mu_D(z')}^{\mu_D(z)} \Delta^{MTE}(x, u_D)\, du_D$$

LATE

We can write the ATE and the LATE using the *marginal treatment effect* (MTE) for individuals with given levels of observed and unobserved covariates:

$$\Delta^{ATE}(x) = \int_0^1 \Delta^{MTE}(x, u_D)\, du_D, \quad \text{and}$$

$$\Delta^{LATE}\left(x, \mu_D(z), \mu_D\left(z'\right)\right) =$$

$$\underbrace{\frac{1}{\mu_D(z) - \mu_D\left(z'\right)}}_{\frac{1}{\mathbb{E}[D_i(1) - D_i(0)]}} \underbrace{\int_{\mu_D(z')}^{\mu_D(z)} \Delta^{MTE}(x, u_D)\, du_D}_{\mathbb{E}[Y_i(D_i(1)) - Y_i(D_i(0))]}$$

It is entirely analogous to our reduced-form result.

LATE

The ATE equals the LATE when there is no endogeneity (*i.e.*, the MTE does not depend on the unobserved u_D) or they could equal by happenstance.

LATE

The ATE equals the LATE when there is no endogeneity (*i.e.*, the MTE does not depend on the unobserved u_D) or they could equal by happenstance.

We get the following theorem: *if there is no endogeneity, then the LATE equals the ATE.* And as a contrapositive, if the LATE does not equal the ATE, then there is endogeneity.

Thus a demonstration of *inequality* of the OLS (ATE) and IV (LATE) estimates can prove endogeneity, but *equality* is not clear evidence of exogeneity.

LATE

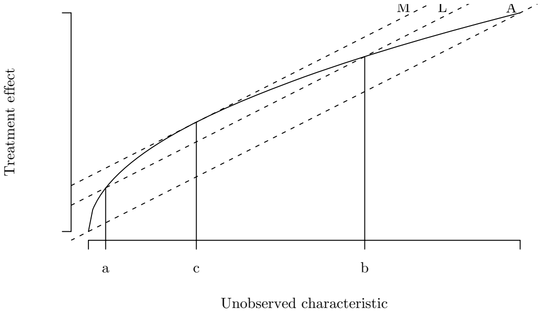

Figure: The response curve to treatment with examples of a MTE and LATE equivalent to the ATE (see Gibbons et al. (2009))

Compliers

So the LATE is integrated over an unobservable set of compliers. We don't know who they are, but can we say something about them as a group?

Note that the following calculations rely upon D and Z being dichotomous. Generalizations are possible.

Compliers

What fraction of observations are compliers:

$$\begin{aligned}\Pr\left[D_i(1) > D_i(0)\right] &= \mathbb{E}\left[D_i(1) - D_i(0)\right] \\ &= \mathbb{E}\left[D_i(1)\right] - \mathbb{E}\left[D_i(0)\right] \\ &= \mathbb{E}\left[D_i|Z = 1\right] - \mathbb{E}\left[D_i|Z = 0\right]\end{aligned}$$

This is just the coefficient on the instrument dummy in the first-stage regression.

Compliers

What fraction of the treated are compliers:

$$\begin{aligned}
&\Pr\left[D_i(1) > D_i(0) | D_i(1) = 1\right] \\
&= \frac{\Pr\left[D_i(1) | D_i(1) > D_i(0)\right] \Pr\left[D_i(1) > D_i(0)\right]}{\Pr[D_i(1)]} \\
&= \frac{\Pr\left[Z_i(1) | D_i(1) > D_i(0)\right] \Pr\left[D_i(1) > D_i(0)\right]}{\Pr[D_i(1)]} \\
&= \frac{\Pr\left[Z_i(1)\right] \left[\mathbb{E}\left[D_i | Z = 1\right] - \mathbb{E}\left[D_i | Z = 0\right]\right]}{\Pr[D_i(1)]}
\end{aligned}$$

By the definition of conditional probability, assumption of the problem, and independence. Notice that all these quantities are calculable.

Compliers

What fraction of the compliers have a dichotomous covariate $x = 1$?

$$
\begin{aligned}
&\Pr[x = 1 | D_i(1) > D_i(0)] \\
&= \Pr[D_i(1) > D_i(0) | x = 1] \frac{\Pr[x = 1]}{\Pr[D_i(1) > D_i(0)]} \\
&= \Pr[x = 1] \frac{\mathbb{E}\left[D_i | Z = 1, x = 1\right] - \mathbb{E}\left[D_i \ Z = 0, x = 1\right]}{\mathbb{E}\left[D_i | Z = 1\right] - \mathbb{E}\left[D_i \ Z = 0\right]}
\end{aligned}
$$

Compliers

Note that this doesn't really give us the information that we want and the LATE may not be the parameter of interest (in fact, it is probably not very relevant at all). So, if there is heterogeneity in the treatment effect, then IV is not that informative.

The advantage of structural models (*e.g.*, Heckman) is that we are forced to consider precisely what we want to estimate and the model lets us estimate that parameter.

Assumptions

Assumptions of IV:

- SUTVA holds
  $Y_i(Z_j) = Y_i \quad \forall\, Z_j :\ j \neq i$
- Ignorable instrument assignment: The instrument is uncorrelated with the unobserved variation
  $\mathrm{Cov}(Z, U_D) = 0$
- Inclusion restriction: The instrument is correlated with treatment
  $\mathrm{Cov}(Z, X) \neq 0$
- Exclusion restriction: The instrument only acts on the outcome by altering treatment status
  $Y|X \perp Z$
- Monotonicity
  $\mu(z) \geq \mu(z')$ or $\mu(z) \leq \mu(z') \quad \forall\, z \geq z'$

Assumptions

We can test the inclusion restriction by examining first-state F-statistics. If these are “large enough” (typically over 10), then we can assume that the instrument is sufficiently correlated with treatment.

Assumptions

We can test the inclusion restriction by examining first-state F-statistics. If these are “large enough” (typically over 10), then we can assume that the instrument is sufficiently correlated with treatment.

Weak instruments generate large standard errors, which can be corrected by using Imbens-Rosenbaum standard errors.

Assumptions

We can test the inclusion restriction by examining first-state F-statistics. If these are "large enough" (typically over 10), then we can assume that the instrument is sufficiently correlated with treatment.

Weak instruments generate large standard errors, which can be corrected by using Imbens-Rosenbaum standard errors.

The other assumptions are *untestable*.