Econometrics II

Lecture 6
Instrumental Variables

Marie Paul

University of Duisburg-Essen
Ruhr Graduate School in Economics

Summer Semester 2020

Outline of lecture

4. Instrumental Variables
   - Motivation and Basics
   - IV in Treatment Effect Evaluation
   - Additional Topics in IV Estimation
   - IV Applications

Self-study plan for Lecture 6

- This lecture presents instrumental variables methods with a focus of the heterogenous treatment effects framework.
- For our next virtual meeting, please work through all slides of Lecture 6. Readings:
- You may want to refresh your knowledge on basic IV by reading the basic IV chapter in Cameron and Trivedi (2005, Chapter 4.8) or Wooldridge (2019, Chapter 15).
- Please read the chapter on instrumental variables by Angrist and Pischke (2009, Chapter 4) thorougly as long as this corresponds to the contents of the slides.
- Please prepare answers to the questions on the slides including the "your own research" questions and collect questions and topics to discuss them in our virtual meeting.

Outline

Endogeneity

One severe problem in many microeconometric applications is potentially inconsistent parameter estimation caused by **endogenous regressors**:

- Is the positive empirical relationship between earnings and years of schooling due to **returns to** or **selection into** education?
- Do cities grow **because of infrastructure** connection or does infrastructure improve due to growth?
- Are unemployed more unhealthy because they **lost their jobs** or did they lose their jobs because they were **unhealthy**?

Dealing with various types of **endogeneity** to justify **causal** statements is at the very core of the microeconometric agenda.

Instrumental variables are widely used in applications where parameters of interest are subject to endogeneity concerns.

Inconsistency of OLS

To exemplify how the OLS estimator may be subject to endogeneity, consider the scalar regression model with a single regressor

$$\mathsf{E}[y|x] = \beta x, \tag{4.1}$$

where interest lies in consistently estimating β to obtain the change in the CEF from an **exogenous** change in x.

The OLS regression model specifies

$$y = \beta x + u, \tag{4.2}$$

so that regressing y on x yields the OLS estimate $\widehat{\beta}$ of β.

The critical assumption for consistency is that x is uncorrelated with u

$$\begin{array}{ccc} x & \longrightarrow & y \\ & \nearrow & \\ u & & \end{array} \tag{4.3}$$

so that the only effect of x on y is a **direct** effect via βx since $dy/dx = \beta$.

Inconsistency of OLS

Suppose that x is **years of education** and y is **earnings**. Then β quantifies the average monetary return of another year of education.

Furthermore, suppose that u includes some general unobserved measure of productivity, likely to affect **both** schooling and earnings

$$\begin{array}{ccc} x & \longrightarrow & y \\ \uparrow & \nearrow & \\ u & & \end{array} \tag{4.4}$$

The path diagram implies that there is now both a **direct** and an **indirect** effect of x on y. Thus, β will be **inconsistent** since

$$y = \beta x + u(x). \tag{4.5}$$

Therefore, since $du/dx \neq 0$, OLS will now estimate the **total** effect

$$\frac{dy}{dx} = \beta + \frac{du}{dx} \neq \beta. \tag{4.6}$$

Inconsistency of OLS

The inconsistency of OLS is here due to **endogeneity** of x, meaning that changes in x are not only associated with changes in y but also in u.

- One solution is to attempt to **control** for the endogenous parts in u. However, since ability is **unobserved**, this is not possible.
- Another problem is **reverse causality** (or simultaneity) when y is also likely to have a direct impact on x ($x \longleftrightarrow y$).

It may still be possible to analyze the impact of x on y with the use of an **instrumental variable** z to create exogenous variation in x

$$\begin{array}{ccccc} z & \longrightarrow & x & \longleftrightarrow & y \\ & & \uparrow & \nearrow & \\ & & u & & \end{array} \tag{4.7}$$

i.e., where z have an impact on y, but only indirectly via x.

In other words, z is an **instrument** for x in the model $y = \beta x + u$ if it is (i) correlated with x [**relevance**] and (ii) uncorrelated with u [**validity**].

The IV estimator

Consider the linear regression model with K regressors $\mathbf{x}$

$$y = \mathbf{x}'\boldsymbol{\beta} + u, \tag{4.8}$$

where $\mathsf{E}[\mathbf{x}u] \neq 0$, so that OLS is inconsistent for $\boldsymbol{\beta}$.

Identification of $\boldsymbol{\beta}_{IV}$ rests on two assumptions given model (4.8):

IV.1 $\mathsf{E}[\mathbf{z}u] = 0$. (validity)

IV.2 (a) rank $\mathsf{E}[\mathbf{z}\mathbf{z}'] = L$; (b) rank $\mathsf{E}[\mathbf{z}\mathbf{x}'] = K$ where $L \geqslant K$. (relevance)

$$\boldsymbol{\beta}_{IV} = \mathsf{E}[\mathbf{z}\mathbf{x}']^{-1}\mathsf{E}[\mathbf{z}y]. \tag{4.9}$$

The 2SLS estimator

The **two-stage least-squares (2SLS) estimator** chooses the **best linear prediction** of $\mathbf{x}$ from $\mathbf{z}$. It is defined by

$$\widehat{\beta}_{2SLS} = \left[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}\left[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}\right] = (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}, \tag{4.10}$$

where $\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ is the OLS prediction of $\mathbf{x}$ using instruments $\mathbf{z}$.

$\widehat{\beta}_{2SLS}$ is thus obtained by plugging the predictions obtained from the **first stage** in the regression.
In the just identified case 2SLS simplifies to the **IV-estimator**.

Examples of instrumental variables

There exist many potential instruments in **policy interventions** and **natural experiments**.

Finding an instrument that is likely to fulfill both relevance and validity is often **difficult** and subject to much **controversy**.

Returning to the motivating applications, we can define potential instruments for the endogenous regressors.

- **Proximity** to a university from the place of residence is often used as an instrument for years of schooling.
- **Historic infrastructure**, in particular if there is **an exogenous reason** for it's existence, is used to instrument recent infrastructure.
- **Plant closures** may generate exogenous variation in job loss to evaluate health effects of unemployment.

It is useful to think of an instrumental variable as a property of some hypothesized **assignment mechanism** that is subject to an **exclusion restriction**.

Examples of instrumental variables

Example 4.1 (Your own research)

Which instrumental variables are used in your research area?

IV Intuition: A Simple Example

Example 4.2 (Wald estimator)

The IV estimator **scales** the effect of a **change** in **z** on y (**reduced form**) by the **impact** on **x** induced by an exogenous shift in **z** (**first stage**).

$$\widehat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = \frac{(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}}{(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}} = \frac{\nabla y(\mathbf{z})}{\nabla \mathbf{x}(\mathbf{z})} \tag{4.11}$$

Consider a scalar **binary instrument** $z = \{0, 1\}$ and denote the subsample averages of y and x, $\bar{y}_z$ and $\bar{x}_z$, respectively. Then (4.11) collapses to

$$\widehat{\beta}_{\text{Wald}} = \frac{dy/dz}{dx/dz} = \frac{(\bar{y}_1 - \bar{y}_0)}{(\bar{x}_1 - \bar{x}_0)} \tag{4.12}$$

where $\Delta q/\Delta z = (\bar{q}_1 - \bar{q}_0) = \gamma_{qz}$, $q = \{y, x\}$, called the **Wald estimator**.

Let x be years of **schooling**, y annual **earnings**, z **distance** to a university and $(\gamma_{yz}, \gamma_{xz}) = (\$500, 0.2)$. Then the annual payoff for another year of schooling is

$$\widehat{\beta}_{\text{Wald}} = \frac{\Delta y/\Delta z}{\Delta x/\Delta z} = \frac{\$500}{0.2} = \$2,500 \tag{4.13}$$

IV Intuition: Another Simple Example

Example 4.3 (Angrist and Evans (1998))

Angrist and Evans (1998) use a sample of American mothers with at least two children. They find that having two children of the same sex (as opposed to having two children of different sex) increases the probability to have a further child by 6 percentage points. Furthermore, if the first two children have the same sex, mothers work 0.31 hours less than those who have a boy and a girl as their first two children. What is the IV-estimate for the effect of having a further child on hours worked?

The Wald estimator makes the "magic" of IV (y is not regressed on $\mathbf{x}$) visible!

Another intuition comes from 2SLS: take the exogenous part of the variation in $\mathbf{x}$ and use only that part in the regression.

Outline

4 Instrumental Variables
- Motivation and Basics
- IV in Treatment Effect Evaluation
- Additional Topics in IV Estimation
- IV Applications

Introduction

Treatment Effect Evaluation is concerned with the identification and estimation of **causal effects** and IV is an important **solution** when correlation between explanatory variables and the error term is suspected.

Very often the concern here is **selection into treatment based on unobservables**.

A leading case has a **continuous outcome** depending on a vector of regressors and an **endogeneous treatment dummy** variable that represents the decision to participate in the treatment. IV here provides alternatives to strongly parametric (non-linear) models.

Effect heterogeneity

The following is based on Angrist and Pischke (2009). The 2SLS estimator relies on the assumption of **treatment homogeneity**; a strong assumption unlikely to hold in many applications.

Consider the simple model

$$Y_i = \alpha + \rho_i D_i + \epsilon_i, \tag{4.14}$$

where treatment is determined by

$$D_i = \pi_0 + \pi_{1i} z_i + \zeta_i. \tag{4.15}$$

Thus, both treatment effects (ρ_i) and response to z_i (π_{1i}) may be individual...

...but normally we assume a **uniform** treatment effect ($\rho_i = \rho \; \forall \; i$).

Effect heterogeneity

Under **homogeneity**,

$$\begin{aligned} \mathbb{E}\left(Y_i|z_i=1\right) &= \alpha+\rho\mathbb{E}\left(D_i|z_i=1\right) \\ \mathbb{E}\left(Y_i|z_i=0\right) &= \alpha+\rho\mathbb{E}\left(D_i|z_i=0\right), \end{aligned} \tag{4.16}$$

and we may recover ρ from

$$\rho=\frac{\mathbb{E}\left(Y_i|z_i=1\right)-\mathbb{E}\left(Y_i|z_i=0\right)}{\mathbb{E}\left(D_i|z_i=1\right)-\mathbb{E}\left(D_i|z_i=0\right)}. \tag{4.17}$$

This leads to the **Wald estimator**, the simplest IV estimator...

...but, *again*, what if treatment effects are really **heterogeneous**? E.g. effect of training on wages? effect of additional child on hours worked? Heterogenity is particularly important when it comes to judging **external** validity.

This leads us to the **LATE IV estimator** of Imbens and Angrist (1994).

Independence

To explain the intuition we adjust the notation of the IV model to the notation of the Rubin model:

$Y_i^d(z)$**:** Potential outcome when treatment indicator equals $d \in \{0,1\}$ and instrument equals $z \in \{0,1\}$.

D_i^z**:** Treatment decision when instrument equals z.

Assumption LATE.1 (Independence)

The instrument z_i is independent of

- $Y_i^d(z)$ *for all* d, z.
- D_i^1 *and* D_i^0

Also called "as good as randomly assigned". Instrument is independent of vector of potential outcome and of potential treatment assignment. This allows to e.g. use those who have $z = 0$ to estimate the effect of z on D.

Independence II

Under **Independence**,

$$\begin{aligned}\mathbb{E}\left(D_i|z_i=1\right)-\mathbb{E}\left(D_i \mid z_i=0\right) &= \mathbb{E}\left(D_i^1 \mid z_i=1\right)-\mathbb{E}\left(D_i^0 \mid z_i=0\right) \\ &= \mathbb{E}\left(D_i^1-D_i^0\right).\end{aligned} \tag{4.18}$$

Likewise,

$$\begin{aligned}\mathbb{E}\left(Y_i \mid z_i=1\right)-\mathbb{E}\left(Y_i \mid z_i=0\right) =&\mathbb{E}\left(Y_i^{D_i^1}(z) \mid z_i=1\right) \\ &-\mathbb{E}\left(Y_i^{D_i^0}(z) \mid z_i=0\right) \\ =&\mathbb{E}\left(Y_i^{D_i^1}(1)\right)-\mathbb{E}\left(Y_i^{D_i^0}(0)\right).\end{aligned} \tag{4.19}$$

The independence assumption is thus sufficient for a causal interpretation of the **first stage** and the **reduced form**.

Exclusion

For the **IV interpretation**, it is **not** sufficient that the instrument is randomly assigned. We also need D to be the **only** channel through which z affects Y.

Assumption LATE.2 (Exclusion)

$$Y_i^d(0) = Y_i^d(1) \;\; \forall d \in \{0,1\}. \tag{4.20}$$

Thanks to exclusion, we get

$$\begin{aligned} Y_i &= Y_i^0(z_i) + \left[Y_i^1(z_i) - Y_i^0(z_i)\right] D_i \\ &= Y_i^0 + \left[Y_i^1 - Y_i^0\right] D_i \\ &= \alpha_i + \rho_i D_i + \epsilon_i. \end{aligned} \tag{4.21}$$

The term **exclusion restriction** refers to the case where the instrument is **excluded** from the structural outcome equation.
Instrument operates through a **single known causal channel**. Counterexample: training vouchers for evening course randomized. Vouchers increase probability of training but also make some receipients work less because they do not need to finance training.

Monotonicity

Finally, in models with heterogeneous treatment effects we need the assumption of **monotonicity** to arrive at a treatment effect for a well-defined subpopulation.

Assumption LATE.3 (Monotonicity)

All affected by the instrument are affected in the same direction (weakly positive):

$$D_i^1 - D_i^0 \geqslant 0, \forall i \tag{4.22}$$

Thus the theorem allows for

- **Always-takers** ($D_i^1 = D_i^0 = 1$),
- **Compliers** ($D_i^1 = 1, D_i^0 = 0$),
- **Never-takers** ($D_i^1 = D_i^0 = 0$),
- ...but **not** *Defiers* ($D_i^1 = 0, D_i^0 = 1$).

Individuals who would **not** have taken the treatment if assigned to it but **would** have done so if not assigned. The instrument **pushes them out** of treatment.

We cannot allow for both defiers and compliers **simultaneously**. Otherwise possible that reduced form is zero because effect of defiers and compliers **cancels** although all have positive treatment effect.

Heterogenous Effects

Example 4.4 (Angrist and Evans (1998) (contin.))

- Sample: mothers with at least two children.
- Causal effect of interest: effect of additional child on hours worked.
- Instrument: two first children of same sex.
- Heterogenous treatment effect? Assumptions? Define who is a complier etc.

The LATE Theorem

Theorem 4.5 (The LATE Theorem)

Under the assumptions of

1. **Independence**
2. **Exclusion**
3. **Monotonicity**...

...the Wald estimator identifies the treatment effect for a well-defined subpopulation.

The Local Average Treatment Effect

$$\rho = \frac{\mathbb{E}(Y_i \mid z_i = 1) - \mathbb{E}(Y_i \mid z_i = 0)}{\mathbb{E}(D_i \mid z_i = 1) - \mathbb{E}(D_i \mid z_i = 0)} \tag{4.23}$$

Numerator:

$$\begin{aligned}
&\mathbb{E}(Y_i \mid z_i = 1) - \mathbb{E}(Y_i \mid z_i = 0) \\
&\overset{(Exclusion)}{=} \mathbb{E}\left(Y_i^0 + \left(Y_i^1 - Y_i^0\right) D_i \mid z_i = 1\right) \\
&\qquad - \mathbb{E}\left(Y_i^0 + \left(Y_i^1 - Y_i^0\right) D_i \mid z_i = 0\right) \\
&\overset{(Independence)}{=} \mathbb{E}\left(Y_i^0 + \left(Y_i^1 - Y_i^0\right) D_i^1\right) - \mathbb{E}\left(Y_i^0 + \left(Y_i^1 - Y_i^0\right) D_i^0\right) \\
&= \mathbb{E}\left(\left(Y_i^1 - Y_i^0\right)\left(D_i^1 - D_i^0\right)\right) \\
&\overset{(Monotonicity)}{=} \mathbb{E}\left(Y_i^1 - Y_i^0 \mid D_i^1 > D_i^0\right) \times \Pr\left(D_i^1 > D_i^0\right)
\end{aligned} \tag{4.24}$$

The Local Average Treatment Effect II

$$\rho = \frac{\mathbb{E}\left(Y_i \mid z_i = 1\right) - \mathbb{E}\left(Y_i \mid z_i = 0\right)}{\mathbb{E}\left(D_i \mid z_i = 1\right) - \mathbb{E}\left(D_i \mid z_i = 0\right)} \tag{4.25}$$

Denominator:

$$\begin{aligned} \mathbb{E}\left(D_i \mid z_i = 1\right) - \mathbb{E}\left(D_i \mid z_i = 0\right) &= \mathbb{E}\left(D_i^1 \mid z_i = 1\right) - \mathbb{E}\left(D_i^0 \mid z_i = 0\right) \\ &\overset{(Independence)}{=} \mathbb{E}\left(D_i^1 - D_i^0\right) \overset{(Monotonicity)}{=} \Pr\left(D_i^1 > D_i^0\right) \end{aligned} \tag{4.26}$$

Therefore,

$$\begin{aligned} \rho_{IV} &= \frac{\mathbb{E}\left(Y_i^1 - Y_i^0 \mid D_i^1 > D_i^0\right) \times \Pr\left(D_i^1 > D_i^0\right)}{\Pr\left(D_i^1 > D_i^0\right)} \\ &= \mathbb{E}\left(Y_i^1 - Y_i^0 \mid D_i^1 > D_i^0\right). \end{aligned} \tag{4.27}$$

That is, the **average treatment effect** for the subpopulation of **treatment compliers**.

In Plain Words...

If there is Effect Heterogeneity (ρ_i)...

And **Independence**, **Monotonicity** and **Exclusion** assumptions are satisfied...

Then the **Wald estimator** identifies the average treatment effect for a subpopulation of **Compliers**.

N.B. This subpopulation is **instrument-specific**. For example, consider popular instruments for number of children on parental child investments.

- **Increase in parental leave benefits**: affects costs of having children.
- **Twin births**: having twins randomly increases number.
- **Same-sex siblings**: preference for mix.

The population margins affected may be quite different. In particular, the instruments affect different fertility levels (1,2,2+ children).

Different treatment effects

- ATE: Average Treatment Effect
- ATT: Average Treatment Effect on the Treated
- ATU: Average Treatment Effect on the Untreated
- LATE: Local Average Treatment Effect: effect on compliers.
- MTE: Marginal Treatment Effect (continuous instrument, see below)
- The **treated population** consists of always-takers and compliers (with the instrument switched on). The latter are **representative for all compliers**, because the instrument is as good as randomly assigned. By monotonicity there are **no defiers**.
- The **ATT** is a weighted average of **always-takers and compliers**.
- The **ATU** is a weighted average of **never-takers and compliers**.
- The **ATE** is a weighted average of **all three groups**.
- Special case: instruments that **allow no never-takers or no always-takers**.

Treatment Effects

Example 4.6 (Angrist and Evans,1998 (contin.))

- Define the LATE.
- Define the ATE, ATT, and ATU.

Special Cases with LATE equal to ATU

Example 4.7 (Twin Instrument of Angrist and Evans, 1998)

LATE is equal to ATU (not treated: two children only), because there are no never-takers (after first child plus twin birth virtually everyone has more than two children). So LATE ($=$ effect on compliers) is equal to ATU.

Special Cases with LATE equal to ATU

Example 4.8 (School Reform of Oreopoulus, 2006)

Compulsory school attendance law in Britain increased from age 14 to 15. Effect of interest: causal effect of another year of schooling i.e. finishing this additional school year is the treatment. LATE is equal to ATU (not treated: those dropping out at age 14), because there were no never-takers (for this specific reform everyone abided the law).
But: due to heterogenous effects we cannot obtain the effect on always-takers (the benefit of this school year for those who take this year more anyway) and thus (without additonal assumptions) not identify the ATT or ATE.

For a treatment with **no always-takers** (e.g. you cannot obtain the training without a voucher) LATE = ATT. But these are typically **very specific treatments** / causal questions and thus potentially of limited interest. Discuss if Chetty et. al (2016): Moving to Opportunity Program estimates an ATT.

IV in randomized trials

- **ITT**: Intention-to-treat effect. Causal effect of the **offer of treatment**. But this usually is not the causal effect of treatment: often some people are **not treated despite** they were assigned to the participant group or people **obtain the treatment despite** being assigned to the control group. In this case use IV:
- Instrument: randomly assigned **offer to treatment**.
- Reduced form: **ITT effect**.
- First stage: **complience rate**.
- Wald estimator: **LATE**.

Characterizing Compliers

- **Size** of complier group: first stage.
- **Proportion of the treated** who are compliers: (first stage * probability instrument is switched on) / proportion treated.
- We don't know if an **individual person** is a complier, because we never see the treatment variable D under **both** the instrument switched on and off.
- We can calculate the **relative probability** a complier is e.g. a college graduate by dividing the first stage for college graduates by the general first stage.

Covariates

- **Assumptions** may be more likely to hold if covariates are added.
- Covariates may increase **precision** of the estimator.
- The IV-estimate is then a **weighted average** of covariate specific LATEs.

Continuous instrument

Based on Cornelissen, Dustmann, Raute, Schönberg, 2016:

- The **marginal treatment effect (MTE)** is defined as the gain from treatment for individuals **shifted into or out** of treatment by a **marginal change** in the propensity score (predicted probability of treatment, function of the instrument). Concepts and estimators are developed in several papers by Heckman and Vytlacil.
- In case of a **continuous instrument** the IV-estimate (the overall LATE) is a **weighted average** of the effects of individuals that are compliers **at different values** of the instrument.
- An aggregate IV estimate **hides** which pairs of values of the instrument shift a particularly **large group** of individuals or a group of individuals with particularly **large treatment effects** into treatment.
- Roughly, the MTE is the **LATE** identified from a **small departure** of a certain value of the **propensity score** induced by the instrument. It is the **derivative** of the outcome with respect to the propensity score.
- The MTE can be estimated **parametrically** or **semiparametrically**.

IV in Treatment Effect Evaluation

Example 4.9 (Your own research)

Can your research question be formulated as a treatment effect? Is the treatment effect likely to be heterogenous? Which instrument might fullfill the LATE assumptions? Which effects can you identify and are they of interest?

Outline

4 Instrumental Variables

- Motivation and Basics
- IV in Treatment Effect Evaluation
- Additional Topics in IV Estimation
- IV Applications

Weak instruments

Weakness of an instrument means that instruments are **weakly correlated** with the variables being instrumented.

The upside is that this is a directly **testable** assumption. The downside is that there is no single definition of a **"weak" instrument**.

One common strategy to test the weakness of an instrument is to analyze the partial F-statistic in the **first stage** regression,

$$\mathbf{x} = \mathbf{z}_1 \boldsymbol{\pi}_1 + \mathbf{x}_2 \boldsymbol{\pi}_2 + v, \tag{4.28}$$

where a test is performed whether $\boldsymbol{\pi}_1 = \mathbf{0}$.

Staiger and Stock (1997) suggested based on simulations that a F-statistic below **10** may indicate problems with finite sample bias.

When an instrument is weak, it may also **exacerbate endogeneity issues** as even small violations of exogeneity can create substantial **parameter bias**.

Inconsistency and weak instruments

It is especially important that an instrument be exogenous if an instrument is weak since **even mild endogeneity** can lead to severe inconsistencies.

Consider the simple case of linear regression with one regressor and one instrument $y = \beta x + u$. Then

$$\frac{\text{plim } \widehat{\beta}_{IV} - \beta}{\text{plim } \widehat{\beta}_{OLS} - \beta} = \frac{\text{Cov}[z,u]/\text{Cov}[z,x]}{\text{Cov}[x,u]/\text{Var}[x]} = \frac{\text{Corr}[z,u]}{\text{Corr}[x,u]} \times \frac{1}{\text{Corr}[z,x]} \tag{4.29}$$

IV may be **more inconsistent than OLS** when the correlation between the instrument and the regressor is low.

Assume $\text{Corr}[z,x] = 0.1$, then

$$\frac{\text{Corr}[z,u]}{\text{Corr}[x,u]} = 0.1 \times \frac{\text{plim } \widehat{\beta}_{IV} - \beta}{\text{plim } \widehat{\beta}_{OLS} - \beta} \tag{4.30}$$

which means that the endogeneity of OLS can be **ten times** as large as the endogeneity of IV and still yield the same magnitude of inconsistency.

Precision

IV estimation always occur at the **cost of precision** compared to OLS. This is intuitive since an observed variable is being **predicted** using other variables.

Consider the variance in the case of a single regressor and instrument under homoscedasticity

$$\begin{aligned} \mathsf{V}[\hat{\beta}_{\mathsf{IV}}] &= \sigma^2(\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{z}(\mathbf{z}'\mathbf{x})^{-1} \\ &= \sigma^2(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{x}(\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{z}(\mathbf{z}'\mathbf{x})^{-1} \\ &= [\sigma^2(\mathbf{x}'\mathbf{x})^{-1}][(\mathbf{x}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{x}]^{-1}[(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{x}]^{-1} \\ &= \mathsf{V}[\hat{\beta}_{\mathsf{OLS}}]/r^2_{xz} \end{aligned} \tag{4.31}$$

where r^2_{xz} is the **squared sample correlation coefficient** of the regression of x on z.

Thus, the IV estimator **always** has **larger variance** than the OLS estimator unless $\mathsf{Corr}[x, z] = 1$. A similar result holds for the 2SLS case

$$\mathsf{se}[\hat{\beta}_{1,\mathsf{2SLS}}] = \mathsf{se}[\hat{\beta}_{1,\mathsf{OLS}}]/R^2_p. \tag{4.32}$$

Outline

Application: Effects of Rural Electrification on Employment

Example 4.10 (Dinkelman, 2011, Am. Econ. Rev.)

- Investigate impact of electrification on labor market participation and wages by analyzing mass roll-out of electricity to rural households in South Africa between 1996 and 2001.
- One part of the analysis uses community panel data and regresses e.g. the female employment rate in a community on whether electricity has already arrived in the community.
- Instrument: average community land gradient. Assumptions?
- Discuss: role of research design (natural experiment) and role of estimator (IV).

IV Application from my own Work

Example 4.11 (Dehos and Paul (2019): The Effect of After-School Programs on Maternal Employment)

- Primary schools end around at lunch-time in Germany. The transition from kindergarten to school involves a worsening of the child-care situation for many families. Major investment in after-schools programs (ASP): in 2002 1% of children in West Germany attended an ASP, already 22% in 2012.
- Instrument: share of primary schools that have received a grant to invest in building up (or extending) an ASP in a county and year.
- ASP-question in SOEP: *Does the child usually attend the school all-day?*
- Sample: mothers with primary school child (years 2003 to 2012).
- **Two-Sample-2SLS**: reduced form with microzensus.

How is ASP participation related to maternal employment?

$$hours_{it} = \delta ASP_{it} + X_{it}'\beta + \epsilon_{it}$$

Table 1: OLS Results

	(1)	(2)	(3)	(4)
	OLS *(SOEP)*			
Hours worked				
ASP attendance	6.905 (0.928)***	5.408 (0.833)***	5.612 (0.835)***	2.154 (0.745)***
Employment				
ASP attendance	0.191 (0.030)***	0.154 (0.027)***	0.160 (0.027)***	0.058 (0.028)**
Observations	5 103	5 103	5 103	4 005
Time dummies	yes	yes	yes	yes
State dummies	yes	yes		
Individual covariates		yes	yes	yes
County covariates		yes	yes	yes
County fixed effects			yes	yes
Past labor market attach.				yes

First Stage

1^{st} stage with SOEP			
z on ASP	0.27	0.24	0.23
	(0.04)***	(0.03)***	(0.04)***
F-statistic	41	47	38
Years	yes	yes	yes
States	yes	yes	
Covariates		yes	yes
Counties			yes

Notes: Covariates include individual and regional time-varying characteristics. Standard errors are clustered at the **county level**. 5103 observations. 230 counties and 10 years.

Reduced Form and IV

Reduced form with Microcensus			
z on hours	0.23 (0.35)	0.18 (0.25)	0.12 (0.27)
z on emp	-0.01 (0.01)	0.00 (0.01)	-0.00 (0.01)
TS2SLS with Microcensus and SOEP			
ASP on hours	0.86 (1.32)	0.76 (1.09)	0.51 (1.18)
ASP on emp	-0.04 (0.05)	0.01 (0.04)	- 0.00 (0.04)
Years	yes	yes	yes
States	yes	yes	
Covariates		yes	yes
Counties			yes

Notes: Covariates include individual and regional time-varying characteristics. Standard errors are clustered at the county level. 140,071 observations. 230 counties and 10 years.