Econometric Methods (Econometrics I)

Lecture 5:

Instrumental Variables Estimation of Single-Equation Linear Models

Prof. Dr. Kai Carstensen

Kiel University

Winter Term 2023/2024

Outline of this lecture

1. IV Estimation in the Single-Equation Single-Regressor Case: One Instrument
2. IV Estimation in the General Single-Equation Case: One Instrument
3. IV Estimation in the General Single-Equation Case: Multiple Instruments
4. Potential Pitfalls with 2SLS
5. IV Solutions to the Omitted Variables Problem

Reference: Wooldridge, Chapter 5; Greene, Chapter 8.

Example: Instrumental Variables for Education in a Wage Equation

- Consider a wage equation for the U.S. working population

$$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + u$$

- We have an omitted variable here: ability probably affects the wage and is correlated with education.
- As we left out ability, it is in the disturbance u.
- Therefore, u is correlated with *educ* and we expect an omitted variable bias.
- More factors are omitted (quality of education, family background) which may be correlated with *educ*.
- What can we do if we do not have a measure of ability or at least a good proxy?

1. Instrumental Variables Estimation in the Single-Equation Single-Regressor Case: One Instrument

Consistency of OLS when the regressors are exogenous

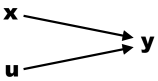
- Consider the scalar regression model with dependent variable y and single regressor x.
- A linear conditional mean model, without intercept for notational convenience, specifies

$$\mathrm{E}(y|x) = \beta x$$

- Using the error term u with $\mathrm{E}(u|x) = 0$ the OLS regression model is

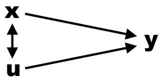
$$y = \beta x + u.$$

- $\mathrm{E}(u|x) = 0$ or the weaker assumption $\mathrm{Cov}(x, u) = 0$ imply consistency of the OLS estimator.
- In a causal interpretation the assumption means that x and u are unrelated causes of y:



The consequence of endogeneity

- Quite often, however, u is related to the observed regressor x (e.g., because of an omitted variable):



- As a consequence, on average u changes when x changes such that

$$\frac{d\,\mathrm{E}(u|x)}{dx} \neq 0.$$

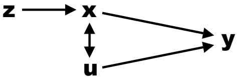
- This implies that the derivative of $\mathrm{E}(y|x)$ with respect to x is not β:

$$\frac{d\,\mathrm{E}(y|x)}{dx} = \frac{\beta dx + d\,\mathrm{E}(u|x)}{dx} = \beta + \frac{d\,\mathrm{E}(u|x)}{dx},$$

- Hence, OLS estimates the total effect $\beta + \frac{d\,\mathrm{E}(u|x)}{dx}$ instead of the partial effect β we are interested in.

Definition of an Instrument

- A popular way to estimate the partial effect from x to y even though x is endogenous is to find an instrument.
- An instrument z
  (1) does not directly affect y,
  (2) is uncorrelated with the error u (**exogeneity**), and
  (3) is correlated with the regressor x (**relevance**).
- Hence, a change in z causes an exogenous change in x which is used to assess the partial effect of a change in x on y:



Example 1

- To identify the partial effect of *educ* on the wage, use an instrument.
- Possibility 1: education of the mother (*motheduc*). This has no direct impact on the wage and is probably well correlated with *educ*. Is it uncorrelated with u? (Well, u contains *fatheduc* and *abil* which may be correlated with *motheduc*...)
- Possibility 2: last digit of one's social security number. This has no direct impact on the wage and is certainly exogenous (it is fully random). But is it relevant to explain *educ*?
- Possibility 3: quarter of birth. This has no direct impact on the wage. It is probably exogenous (unrelated to factors in u). And it explains *educ*! Why? Angrist and Krueger (1991) argue that compulsory school attendance laws induce a relationship with *educ*: at least some people are forced, by law, to attend school longer than they otherwise would, and this fact is correlated with quarter of birth.

Example 2

- Aim: identify partial effect of a price on demand.
- Clearly, in a market the price is endogenous.
- We need an instrument that affects the price but is uncorrelated with all kinds of demand drivers.
- Good idea: instrument that purely reflects exogenous supply conditions.
- Example: Weather conditions for agricultural products.

Example 3

- Aim: identify partial effect of institutions on GDP per capita.
- Most probably, institutions are endogenous and measured with error.
- We need an instrument that affects the institutional quality but is uncorrelated with all kinds of other current drivers of GDP (like demand shocks, tax changes, labor market reforms, monetary policy shifts) and to the current measurement error.
- Idea: 19th century settler mortality $\rightarrow$ type of institutions built up in the colony $\rightarrow$ has effect until today (but is uncorrelated with today's factors affecting GDP).
- Details: Acemoglu et al. (2001, AER).

Intuitive derivation of the IV estimator

- Consider again the path diagram:

$$\begin{tikzcd}
\mathbf{z} \arrow[r] & \mathbf{x} \arrow[d, leftrightarrow] \arrow[dr] & \\
 & \mathbf{u} \arrow[r] & \mathbf{y}
\end{tikzcd}$$

- Since z is exogenous, OLS consistently estimates the indirect average effect from z to y, dy/dz, by regressing y on z.

- Similarly, OLS consistently estimates the direct average effect from z to x, dx/dz by regressing x on z.

- The partial effect from x to y, β, is the effect from z to y per unit change in x due to z:

$$\beta_{IV} = \frac{dy/dz}{dx/dz} = \frac{dy}{dx}$$

- This is consistently estimated by

$$\widehat{\beta}_{IV} = \frac{\widehat{\gamma}_{z\to y,OLS}}{\widehat{\gamma}_{z\to x,OLS}} = \frac{(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{y}}{(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\mathbf{x}} = (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{y}.$$

Formal derivation

- More formally, we use z to generate a consistent normal equation.
- Since by assumption and a WLLN

$$\text{plim}\, N^{-1}\mathbf{z}'\mathbf{u} = \mathrm{E}(z \cdot u) = 0,$$

  we use

$$N^{-1}\mathbf{z}'\widehat{\mathbf{u}} = N^{-1}\mathbf{z}'(\mathbf{y} - \widehat{\beta}_{IV}\mathbf{x}) = 0$$

  as a normal equation.
- Solving for $\widehat{\beta}_{IV}$ yields the IV estimator

$$\widehat{\beta}_{IV} = (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{y}.$$

2. Instrumental Variables Estimation in the General Single-Equation Case: One Instrument

The multiple regression model with one endogenous regressor

- Consider the linear population model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$$

with

$$\mathrm{E}(u) = 0, \qquad \mathrm{Cov}\,(x_j, u) = 0, \qquad j = 1, 2, \ldots, K-1$$

but where x_K might be correlated with u.

- In other words, the explanatory variables $x_1, x_2, \ldots, x_{K-1}$ are exogenous, but x_K is potentially endogenous.
- A possible reason is that there is an omitted variable q which is uncorrelated with all regressors except x_K.
- In general, this leads to inconsistent estimates of *all* the β's.
- Therefore, we again resort to instrumental variables estimation.

Properties of a good instrument

- Recall: an instrument z_1 must (1) not directly affect y, (2) be uncorrelated with the error u, and (3) be relevant to explain x_K.
- The first condition again implies that z_1 does not show up in the equation for y.
- The second condition is also as before:

$$\text{Cov}\,(z_1, u) = 0.$$

- The third condition (relevance) in this context implies that z_1 is *partially correlated* with x_K once the other exogenous variables $x_1, \ldots, x_{K-1}$ have been netted out.
- This means that a nonzero raw correlation $\text{Cov}(z_1, x_K) \neq 0$ is not sufficient.
- What does this mean?

Relevance of the instrument

- Relevance is precisely defined as follows. Consider the linear projection of x_K onto all the exogenous variables:

$$x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K.$$

- By definition of a linear projection error, $\mathrm{E}(r_K) = 0$ and r_K is uncorrelated with $x_1, x_2, \ldots, x_{K-1}$, and z_1.

- Relevance of z_1 means that the coefficient on z_1 is nonzero:

$$\theta_1 \neq 0$$

- Compare this to the single-regressor case discussed before: If x_K is the only explanatory variable in equation, then the linear projection is $x_K = \delta + \theta_1 z_1 + r_K$, where $\theta_1 = \mathrm{Cov}(z_1, x_K)/\mathrm{Var}(z_1)$, and so $\theta_1 \neq 0$ and $\mathrm{Cov}(z_1, x_K) \neq 0$ are the same.

Some terminology

- In general, we may call not only z_1 but also all exogenous regressors *instruments*.
  - This is so because $x_1, \ldots, x_{K-1}$ are already uncorrelated with u and thus serve as their own instruments.
  - In other words, the full list of instrumental variables is the same as the list of exogenous variables, but we often just refer to the instrument for the endogenous explanatory variable.
- The linear projection is called a *reduced form equation* or *first-stage regression* for the endogenous explanatory variable x_K.
  - The "reduced form" terminology comes from simultaneous equations analysis.
  - But it is used in all IV contexts because it is a concise way of stating that an endogenous variable has been linearly projected onto the exogenous variables.
  - The terminology also conveys that there is nothing necessarily structural about the reduced form equation.

- Recall: identification here means that we can write β in terms of population moments in observable variables.
- To this end, define the $1 \times K$ vectors of regressors and exogenous variables

$$\mathbf{x} \equiv (1, x_2, \ldots, x_K) \qquad \text{and} \qquad \mathbf{z} \equiv (1, x_2, \ldots, x_{K-1}, z_1) \,.$$

- By the exogeneity assumption, we have the K population orthogonality conditions

$$\mathrm{E}\left(\mathbf{z}' u\right) = \mathbf{0}.$$

- By the relevance assumption ($\theta_1 \neq 0$), we have full rank and thus invertibility of $\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)$,

$$\operatorname{rank}\left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right] = K.$$

(To show this, plug the reduced form equation for x_K into $\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)$. Then it is straightforward to show that for $\theta_1 = 0$ the last column of $\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)$ is linearly dependent on the first $K - 1$ columns.)

- Now write the structural equation as

$$y = \mathbf{x}\boldsymbol{\beta} + u.$$

- Multiply it by the instruments $\mathbf{z}'$:

$$\mathbf{z}'y = \mathbf{z}'\mathbf{x}\boldsymbol{\beta} + \mathbf{z}'u$$

- Take expectations using the exogeneity assumption $\mathrm{E}\left(\mathbf{z}'u\right) = \mathbf{0}$:

$$\mathrm{E}\left(\mathbf{z}'y\right) = \mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\boldsymbol{\beta}.$$

- Solve for $\boldsymbol{\beta}$ using the relevance assumption (which implies $\operatorname{rank}\left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right] = K$):

$$\boldsymbol{\beta} = \left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right]^{-1}\mathrm{E}\left(\mathbf{z}'y\right).$$

- This shows that $\boldsymbol{\beta}$ can be represented in terms of population moments in observable variables.

The Instrumental Variables Estimator

- Consider a random sample $\{(\mathbf{x}_i, y_i, \mathbf{z}_i)\colon i = 1, \ldots, N\}$ from the population.
- Define the data matrices
  - $\mathbf{X}$: $N \times K$ matrix of the $\mathbf{x}_i$'s
  - $\mathbf{Z}$: $N \times K$ matrix of the $\mathbf{z}_i$'s
  - $\mathbf{Y}$: $N \times 1$ vector of the y_i's
- Then the **instrumental variables (IV) estimator** of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{IV} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i' y_i \right) = \left(\mathbf{Z}'\mathbf{X} \right)^{-1} \mathbf{Z}'\mathbf{Y}$$

Consistency of the Instrumental Variables Estimator

- The consistency of this estimator is shown in the same way as for the OLS estimator.
- Substituting the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$ into the IV estimator yields

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{IV} &= \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}) = \boldsymbol{\beta} + \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{U} \\ &= \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'u_i\right)\end{aligned}$$

- By the weak law of large numbers,

$$N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i \xrightarrow{\text{p}} \mathrm{E}\left(\mathbf{z}'\mathbf{x}\right) \quad \text{and} \quad N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'u_i \xrightarrow{\text{p}} \mathrm{E}\left(\mathbf{z}'u\right).$$

- Assumptions: $\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)$ is invertible (relevance) and $\mathrm{E}\left(\mathbf{z}'u\right) = 0$ (exogeneity).
- Hence,

$$\operatorname{plim}\hat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right]^{-1}\mathrm{E}\left(\mathbf{z}'u\right) = \boldsymbol{\beta} + \left[\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right]^{-1}\mathbf{0} = \boldsymbol{\beta}.$$

Asymptotic Normality of the Instrumental Variables Estimator

- As for the OLS estimator, asymptotic normality follows from the CLT, now applied to

$$N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{B})$$

  where

$$\mathbf{B} = \mathrm{Var}(\mathbf{z}_i' u_i) = \mathrm{E}(u_i^2 \mathbf{z}_i' \mathbf{z}_i).$$

- Furthermore, recall $N^{-1} \sum_{i=1}^{N} \mathbf{z}_i' \mathbf{x}_i \xrightarrow{\mathrm{p}} \mathrm{E}\left(\mathbf{z}' \mathbf{x}\right) \equiv \mathbf{A}$ with $\mathbf{A}$ being invertible.
- From this follows

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}) = \left(N^{-1} \sum_{i=1}^{N} \mathbf{z}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \right) \xrightarrow{\mathrm{d}} \mathrm{Normal}(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} (\mathbf{A}')^{-1})$$

3. Instrumental Variables Estimation in the General Single-Equation Case: Multiple Instruments

The problem posed by multiple instruments

- Consider again the linear population model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$$

  where x_K might be correlated with u.

- Now assume that we have more than one instrumental variable for x_K.

- Let $z_1, \ldots, z_M$ be the instrumental variables such that

$$\mathrm{Cov}\left(z_h, u\right) = 0, \qquad h = 1, 2, \ldots, M$$

  so that each z_h is exogenous.

- If each of these has some partial correlation with x_K, we could have M different IV estimators.

- Or we could use linear combinations of $x_1, x_2, \ldots, x_{K-1}$, $z_1, z_2, \ldots, z_M$ as instruments, which are also uncorrelated with u.

- So which IV estimator should we use?

Two-stage least squares: idea

- Here is how the **two-stage least squares (2SLS) estimator** solves this challenge.
- Out of all possible linear combinations of the z's that can be used as an instrument for x_K, the method of 2SLS chooses that which is most highly correlated with x_K.
- How to find the linear combination that is most highly correlated with x_K?

Two-stage least squares: the reduced form in population

- Maximum correlation is achieved by the linear projection of x_K on all exogenous variables (also called the reduced form for x_K):

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_M z_M + r_K$$

- By definition, r_K has zero mean and is uncorrelated with each rhs variable.
- Endogeneity of x_K arises because r_K is correlated with u.
- As any linear combination of the exogenous variables is uncorrelated with u, also this one:

$$x_K^* \equiv \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_M z_M.$$

- x_K^* has maximum correlation with x_K among all linear combinations of z.
- x_K^* is that part of x_K that is uncorrelated with u.
- However, $\delta_0, \ldots, \delta_{K-1}, \theta_1, \ldots, \theta_M$ are unknown population parameters. Hence, x_K^* is unobserved.

Two-stage least squares: the first-stage regression

- The first-stage regression estimates by OLS the unknown parameters of the reduced form for x_K,

$$x_K = \delta_0 + \delta_1 x_1 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_M z_M + r_K.$$

- The OLS estimator is consistent, if there are no exact linear dependencies among the exogenous variables.
- The sample analogues of the x_{iK}^* for each observation i are simply the OLS fitted values:

$$\hat{x}_{iK} = \hat{\boldsymbol{\delta}}_0 + \hat{\boldsymbol{\delta}}_1 x_{i1} + \cdots + \hat{\boldsymbol{\delta}}_{K-1} x_{i,K-1} + \hat{\boldsymbol{\theta}}_1 z_{i1} + \cdots + \hat{\boldsymbol{\theta}}_M z_{iM}.$$

- Note that we effectively use all $L = K + M$ exogenous variables to construct a single instrument $\hat{x}_{iK}$ for the single endogenous regressor x_{iK}.

- To write this more compactly, define the $1 \times L$ vector of exogenous variables:

  $$\mathbf{z} \equiv (1, x_1, \ldots, x_{K-1}, z_1, \ldots, z_M)$$

  and the $1 \times (K+1)$ vector of regressors in the structural equation for y:

  $$\mathbf{x} \equiv (1, x_1, \ldots, x_K).$$

- Define the data matrices
  - $\mathbf{X}$: $N \times (K+1)$ matrix of the $\mathbf{x}_i$'s
  - $\mathbf{X}_K$: $N \times 1$ matrix of the $\mathbf{x}_{k,i}$'s, $k = 1, \ldots, K$.
  - $\mathbf{Z}$: $N \times (K+L)$ matrix of the $\mathbf{z}_i$'s
  - $\mathbf{Y}$: $N \times 1$ vector of the y_i's
- For later use, remember the projection matrix $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

- Write the first-stage regression in matrix form as

$$\mathbf{X}_K = \mathbf{Z}\gamma + \mathbf{R}.$$

- Then the OLS estimator is

$$\hat{\gamma} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_K.$$

- The predicted values are

$$\hat{\mathbf{X}}_K = \mathbf{Z}\hat{\gamma} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}_K = \mathbf{P}_Z\mathbf{X}_K.$$

Two-stage least squares: the second-stage regression

- In the second stage, estimate the structural equation for y by IV, using $\hat{x}_{iK}$ as instrument for the endogenous regressor x_{iK}.
- To this end, define the $1 \times (K+1)$ vector

$$\tilde{\mathbf{x}}_i \equiv (1, x_{i1}, \ldots, x_{i,K-1}, \hat{x}_{iK}), \qquad i = 1, \ldots, N.$$

  and the data matrix $\tilde{\mathbf{X}}$ which is the $N \times (K+1)$ matrix of the $\tilde{\mathbf{x}}_i$'s.
- The estimator thus is

$$\hat{\beta}_{2SLS} = \left(\sum_{i=1}^{n} \tilde{\mathbf{x}}_i' \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \tilde{\mathbf{x}}_i' y_i\right) = \left(\tilde{\mathbf{X}}'\mathbf{X}\right)^{-1} \tilde{\mathbf{X}}'\mathbf{Y}.$$

Two-stage least squares: the second-stage regression

- To better understand the 2SLS estimator, let us define the OLS predictions

$$\hat{\mathbf{X}} \equiv \mathbf{P_Z X} = [\mathbf{P_Z X}_0, \mathbf{P_Z X}_1, \ldots, \mathbf{P_Z X}_{K-1}, \mathbf{P_Z X}_K],$$

where $\mathbf{X}_0$ is a vector of ones.

- Note that

$$\mathbf{P_Z X}_j = \mathbf{X}_j, \quad j = 0, \ldots, K-1,$$

because $\mathbf{X}_j$, $j = 0, \ldots, K-1$ is part of $\mathbf{Z}$ and thus $\mathbf{P_Z X}_j$ is the predicted value from a linear regression of $\mathbf{X}_j$ on itself (and some other variables). This is the variable $\mathbf{X}_j$ itself.

- This implies that

$$\mathbf{P_Z X} = \hat{\mathbf{X}} = [\mathbf{P_Z X}_0, \ldots, \mathbf{P_Z X}_{K-1}, \mathbf{P_Z X}_K] = [\mathbf{X}_0, \ldots, \mathbf{X}_{K-1}, \hat{\mathbf{X}}_K] = \tilde{\mathbf{X}}$$

and thus (using $\mathbf{P_Z} = \mathbf{P_Z P_Z}$)

$$\tilde{\mathbf{X}}'\mathbf{X} = \mathbf{X}'\mathbf{P_Z X} = \mathbf{X}'\mathbf{P_Z P_Z X} = \hat{\mathbf{X}}'\hat{\mathbf{X}}.$$

Two-stage least squares: the second-stage regression

- Therefore, we may write the 2SLS estimator equivalently as

$$\hat{\beta}_{2SLS} = \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{Y} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{Y} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\mathbf{Y}.$$

- The latter shows that the 2SLS estimator can be represented as an OLS estimator of a second-stage regression of y on the regressors $1, x_{i1}, \ldots, x_{i,K-1}, \hat{x}_{iK}$.
- Using the projection matrix $\mathbf{P_Z}$ explicitly, we do not need the first-stage regression but can compute the estimator in a single step:

$$\hat{\beta}_{2SLS} = (\mathbf{X}'\mathbf{P_Z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P_Z}\mathbf{Y} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

To prove consistency, we have to make two important assumptions.

ASSUMPTION 2SLS.1: For some $1 \times L$ vector $\mathbf{z}$, $\mathrm{E}(\mathbf{z}'u) = \mathbf{0}$.

ASSUMPTION 2SLS.2: (a) rank $\mathrm{E}(\mathbf{z}'\mathbf{z}) = L$; (b) rank $\mathrm{E}(\mathbf{z}'\mathbf{x}) = K + 1$.

- Assumption 2SLS.1 is the exogeneity assumption.
- Assumption 2SLS.2 (a) guarantees that the instruments are not perfectly correlated with each other.
- Assumption 2SLS.2 (b) is the relevance assumption. It is also called the *rank condition* for identification.
- Necessary (but not sufficient) for the rank condition is the *order condition*, $L \geq K + 1$: We must have at least as many instruments as we have explanatory variables.

- Identification is achieved when the assumptions 2SLS.1 and 2SLS.2 are satisfied.
- This can be show similarly as for the single-instrument case discussed above.
- First, write the linear projection of $\mathbf{x}$ onto $\mathbf{z}$ as $\mathbf{x}^* = \mathbf{z}\mathbf{\Pi}$, where $\mathbf{\Pi}$ is the $L \times (K+1)$ matrix $\mathbf{\Pi} = [\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\,\mathrm{E}(\mathbf{z}'\mathbf{x})$.
- In the next step, multiply the structural equation by the instruments $\mathbf{x}^{*\prime} = \mathbf{\Pi}'\mathbf{z}'$:

$$\mathbf{x}^{*\prime} y = \mathbf{x}^{*\prime}\mathbf{x}\boldsymbol{\beta} + \mathbf{x}^{*\prime} u \qquad \Rightarrow \qquad \mathbf{\Pi}'\mathbf{z}' y = \mathbf{\Pi}'\mathbf{z}'\mathbf{x}\boldsymbol{\beta} + \mathbf{\Pi}'\mathbf{z}' u.$$

- Now take expectations (using the exogeneity assumption $\mathrm{E}\left(\mathbf{z}'u\right) = \mathbf{0}$):

$$\mathbf{\Pi}'\,\mathrm{E}\left(\mathbf{z}'y\right) = \mathbf{\Pi}'\,\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\boldsymbol{\beta}.$$

- Finally, solve for $\boldsymbol{\beta}$ using the relevance assumption (which implies that $\mathbf{\Pi}'\,\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right) = \mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\,\mathrm{E}(\mathbf{z}'\mathbf{x})$ has rank $K+1$):

$$\boldsymbol{\beta} = \left[\mathbf{\Pi}'\,\mathrm{E}\left(\mathbf{z}'\mathbf{x}\right)\right]^{-1}\mathbf{\Pi}'\,\mathrm{E}\left(\mathbf{z}'y\right) = \left[\mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\,\mathrm{E}(\mathbf{z}'\mathbf{x})\right]^{-1}\mathrm{E}(\mathbf{x}'\mathbf{z})[\mathrm{E}(\mathbf{z}'\mathbf{z})]^{-1}\,\mathrm{E}(\mathbf{z}'y).$$

- This shows that $\boldsymbol{\beta}$ can be represented in terms of population moments in observable variables.

- Recall the the 2SLS estimator can be written as

$$\hat{\beta}_{2SLS} = \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$

- Substituting the model yields

$$\hat{\beta}_{2SLS} = \beta + \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{U}.$$

- Using sums, this is equivalent to

$$\begin{aligned}\hat{\beta}_{2SLS} \quad &= \quad \beta + \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{x}_i\right)\right]^{-1} \\ &\quad \cdot \left(N^{-1}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{z}_i\right)\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i'u_i\right)\end{aligned}$$

- Define the moment matrices $\mathbf{A}_{xz} \equiv \mathrm{E}(\mathbf{x}'\mathbf{z})$, $\mathbf{A}_{zx} \equiv \mathrm{E}(\mathbf{z}'\mathbf{x})$ and $\mathbf{A}_{zz} \equiv \mathrm{E}(\mathbf{z}'\mathbf{z})$.
- By the weak law of large numbers and Slutsky's theorem,

$$\hat{\boldsymbol{\beta}}_{2SLS} \xrightarrow{\text{p}} \boldsymbol{\beta} + \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\,\mathrm{E}(\mathbf{z}'u)$$

- Using $\mathrm{E}(\mathbf{z}'u) = \mathbf{0}$ yields

$$\hat{\boldsymbol{\beta}}_{2SLS} \xrightarrow{\text{p}} \boldsymbol{\beta}.$$

- Asymptotic normality follows from the CLT applied to

$$N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{B})$$

where

$$\mathbf{B} = \text{Var}(\mathbf{z}_i' u_i) = \text{E}(u_i^2 \mathbf{z}_i' \mathbf{z}_i).$$

- From this follows

$$\sqrt{N}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{\text{d}} \text{Normal}\left(\mathbf{0}, \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1} \mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{B}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx} \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}\right)$$

- Under homoscedasticity, we have

$$\mathbf{B} = \operatorname{Var}(\mathbf{z}_i' u_i) = \operatorname{E}(u_i^2 \mathbf{z}_i' \mathbf{z}_i) = \sigma^2 \operatorname{E}(\mathbf{z}_i' \mathbf{z}_i) = \sigma^2 \mathbf{A}_{zz}.$$

- This simplifies the asymptotic covariance matrix to

$$\left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1} \mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\sigma^2\mathbf{A}_{zz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx} \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1} = \sigma^2 \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}$$

- Hence, the asymptotic distribution under homoscedasticity is

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{\mathrm{d}} \text{Normal}\left(\mathbf{0}, \sigma^2 \left[\mathbf{A}_{xz}\mathbf{A}_{zz}^{-1}\mathbf{A}_{zx}\right]^{-1}\right)$$

- To estimate the asymptotic variance of the 2SLS estimator, be sure to start from the correct residuals

$$\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\beta}_{2SLS}.$$

- Do not use $\mathbf{Y} - \hat{\mathbf{X}}\hat{\beta}_{2SLS}$ as residuals!
- Then proceed as always: use the sample counterparts.
- For example, under heteroscedasticity

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N} \hat{u}_i^2 \mathbf{z}_i' \mathbf{z}_i.$$

- The other moment matrices are estimated as

$$\hat{\mathbf{A}}_{xz} = N^{-1}\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{z}_i, \quad \hat{\mathbf{A}}_{zx} = \hat{\mathbf{A}}_{xz}', \quad \hat{\mathbf{A}}_{zz} = N^{-1}\sum_{i=1}^{N} \mathbf{z}_i' \mathbf{z}_i.$$

- Like always,

$$\text{Avar}(\hat{\boldsymbol{\beta}}_{2SLS}) = \text{Avar}\left(\sqrt{N}(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta})\right)/N.$$

- Using some algebra, it turns out that the estimated asymptotic variance can also be represented as

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{2SLS}) = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\left(\sum_{i=1}^{N}\hat{\mathbf{u}}_i^2\hat{\mathbf{x}}_i'\hat{\mathbf{x}}_i\right)\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}.$$

- Under homoscedasticity, this simplifies to

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{2SLS}) = \hat{\sigma}^2\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}.$$

- Sometimes this matrix is multiplied by $N/(N-K)$ as a degrees-of-freedom adjustment.

- The standard errors for the individual $\hat{\beta}_j$ are taken as the square roots of the respective element of the main diagonal of $\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{2SLS})$.

Inference

- Confidence intervals, t tests and Wald tests can be applied in the same way as known from the OLS case.
- Like always, when you are unsure whether there is heteroscedasticity better use robust estimates of the covariance matrix.
- In Stata, use the options `robust` or `vce(robust)`.

Test of overidentifying orthogonality restrictions

- If $L > K+1$, we can test whether the sample deviation of $\sum_{i=1}^{N} \mathbf{z}_i' \hat{u}_i$ from zero is significant or not.
- In terms of the population, the question is whether the orthogonality conditions are valid or not.
- Hypotheses: $H_0 : \mathrm{E}(\mathbf{z}'u) = \mathbf{0}$ against $H_1 : \mathrm{E}(\mathbf{z}'u) \neq \mathbf{0}$
- Test statistic à la Sargan (1958):

$$S \equiv \left[N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \right]' \hat{\mathbf{B}}^{-1} \left[N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \right] \xrightarrow{\mathrm{d}} \chi^2_{L-K-1}$$

- Test decision: If $S > \chi^2_{L-K-1}(1-\alpha)$, reject $H_0 \Rightarrow$ invalid orthogonality conditions
- Important: The test does not tell us which of the orthogonality conditions are invalid.
- In Stata, use the postestimation command `estat overid`.

Under H_0, the moment conditions are valid and the CLT applies:

$$\mathbf{h} \equiv N^{-1/2} \sum_{i=1}^{N} \mathbf{z}_i' u_i \xrightarrow{\text{d}} \text{Normal}(\mathbf{0}, \mathbf{B}).$$

Hence, using the rules for quadratic forms,

$$\mathbf{h}' \mathbf{B}^{-1} \mathbf{h} \xrightarrow{\text{d}} \chi^2_L.$$

Replacing u_i by $\hat{u}_i$ “eats up” $K + 1$ degrees of freedom, hence

$$S \equiv \hat{\mathbf{h}}' \hat{\mathbf{B}}^{-1} \hat{\mathbf{h}} \xrightarrow{\text{d}} \chi^2_{L-K-1}.$$

4. Potential Pitfalls with 2SLS

Overview

When applying and interpreting the 2SLS estimator, always keep in mind that the 2SLS estimator

- is always biased,
- often has a large variance,
- in certain cases follows a finite-sample distribution that is not well-approximated by the asymptotic distribution we derived.

- While the 2SLS estimator is consistent, it is not unbiased even if we assume $\mathrm{E}[\mathbf{U}|\mathbf{Z}] = 0$.

- Recall that OLS is unbiased when we assume $\mathrm{E}[\mathbf{U}|\mathbf{X}] = \mathbf{0}$:

$$
\begin{aligned}
\mathrm{E}[\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}] &= \mathrm{E}\left\{\mathrm{E}[\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}|\mathbf{X}]\right\} = \mathrm{E}\left\{\mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}|\mathbf{X}]\right\} \\
&= \mathrm{E}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{E}[\mathbf{U}|\mathbf{X}]\right\} = \mathrm{E}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0}\right\} = \mathbf{0}.
\end{aligned}
$$

- However, for 2SLS conditioning on $\mathbf{X}$ does not work because $\mathbf{X}$ is endogenous. Conditioning on $\mathbf{Z}$ is feasible, but

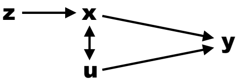
$$
\mathrm{E}[\widehat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}] = \mathrm{E}\left\{\mathrm{E}[\widehat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}|\mathbf{Z}]\right\} = \mathrm{E}\left\{\mathrm{E}[(\mathbf{X}'\mathbf{P_Z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P_Z}\mathbf{U}|\mathbf{Z}]\right\}
$$

  cannot be further simplified.

- Hence, the 2SLS estimator is biased in finite samples.

Large variance of the 2SLS estimator

- Another problem with the 2SLS estimator is its potentially low estimation precision.
- To see this, consider the scalar regression model with one endogenous regressor x and one instrument z.
- Intuitively, z is a "treatment" that leads to exogenous movement in x:



- But as there are many (endogenous) reasons why x moves, the *exogenous* regressor variance is small which leads to low estimation precision.
- This implies that the imprecision increases as the correlation of the instrument with the endogenous regressor decreases.

Large variance of the 2SLS estimator

- This is easily seen in the case of iid (homoscedastic) disturbances and a single regressor and instrument (here: both in mead-adjusted form). Then the asymptotic variance is estimated as

$$\widehat{\text{Var}}[\widehat{\beta}_{2SLS}] = \hat{\sigma}^2 \left[\hat{\mathbf{X}}'\hat{\mathbf{X}}\right]^{-1} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} / \frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{\mathbf{X}'\mathbf{X}} = \widehat{\text{Var}}[\widehat{\beta}_{OLS}]/R^2_{xz}$$

  where

$$R^2_{xz} = \frac{\hat{\mathbf{X}}'\hat{\mathbf{X}}}{\mathbf{X}'\mathbf{X}}$$

  is the R-squared of the first-stage regression.

- Interpretation: if the R-squared of the first-stage regression (regression of of x on z) is small, then the 2SLS standard error is large.

- Example: if the R-squared of the first-stage regression equals 0.1, then 2SLS standard errors are $\sqrt{10}$ times those of OLS.

- Consider the scalar regression model with one endogenous regressor x and one instrument z.

- It is obvious that an R-squared of the first-stage regression of exactly zero leads to a breakdown of the 2SLS estimator because this implies $\hat{\mathbf{X}} = 0$ and thus

$$\hat{\beta}_{2SLS} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{Y}$$

  cannot be computed.

- Similarly, a population correlation between x and z of zero is highly problematic. Then the relevance condition is violated on which our proofs of the consistency and asymptotic normality of the 2SLS estimator hinge.

- But even the case of a very small but nonzero correlation between x and z leads to problems. This is the so-called *weak instrument* case.

The problem of weak instruments

- Under weak instruments (small but nonzero correlation between x and z), the asymptotic distribution of the 2SLS estimator is a very bad proxy for its unknown finite-sample distribution.
- Hence, inference based on the asymptotic distribution is incorrect.
- In addition, the 2SLS estimator is biased towards the (inconsistent) OLS estimator and exhibits a large variance.
- Unfortunately, even 300,000 observations may not be enough to make the asymptotic distribution approximate the finite-sample distribution well, see Bound, J., D. A. Jaeger, and R. M. Baker (1995), Problems with Instrumental Variables Estimation When the Correlation between the Instruments and Endogenous Explanatory Variables Is Weak, Journal of the American Statistical Association 90, 443-450.

The problem of weak instruments

- How to shield against the weak instrument problem in practice?
- In the first-stage regression

$$x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \cdots + \theta_M z_M + r_K$$

  the F statistic for excluding the instruments $z_1, \ldots, z_M$ should be large.
- What is large? As a rule of thumb, the F statistic should exceed 10, preferably by a large amount, see Stock, Wright and Yogo (2002), A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, Journal of Business and Economic Statistics 20, 518-529.
- Fortunately, not all is lost when the F statistic does not exceed 10.
- What to do in this case and more on weak instruments: lecture on Microeconometrics (next semester).

5. IV Solutions to the Omitted Variables Problem

The simple IV solution

- Consider again the omitted variable model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma q + v$$

  where q represents the omitted variable and $\mathrm{E}(v|\mathbf{x}, q) = 0$.

- We can estimate the parameters $\beta_0, \ldots \beta_K$ consistently by 2SLS.

- What we have to do is to find instruments for any element of $\mathbf{x}$ that is correlated with q.

- These instruments must satisfy the following requirements:

  (1) they are redundant in the structural model,

  (2) they are uncorrelated with the omitted variable, q; and

  (3) they are sufficiently correlated with the endogenous elements of $\mathbf{x}$.

Solutions using indicator variables

- When you have at least two indicators, you can combine the indicator and IV approaches.
- This is especially helpful for the special case of a measurement error: you need two measurements with uncorrelated measurement error. Then you can use one measurement as the instrument for the other measurement you plug into the structural equation.
- For details, see Wooldrigde (p. 105-107).

6. Example

Question

Acemoglu, Johnson, Robinson (2001), The Colonial Origins of Comparative Development: An Empirical Investigation, AER 91(5), 1369-1401

- How does the quality of institutions affect output per capita?
- Problem: measure of institutions?
- AJR: regressor $x =$ protection against expropriation (score between 0 and 10, Political Risk Services).
- But x may suffer from endogeneity (e.g., reverse causality) and measurement error ($\rightarrow$ attrition).
- Idea: use $z =$ settler mortality as instrument. Surprising?

Regression results

- OLS of log GDP per capita on protection against expropriation score: $\hat{\beta} = 0.52$
- IV of log GDP per capita on protection against expropriation score: $\hat{\beta} = 0.94$
- So roughly, IV yields an estimate that is twice as strong.
- Economically relevant? Compare Nigeria and Chile (which are very near the regression line, GDP of 1995):
    - Score difference: $x_{Chi} - x_{Nig} = 2.24$.
    - Predicted difference: $lGDP_{Chi} - lGDP_{Nig} = 0.94 \times 2.26 \approx 2.1$.
    - Exponentiating: $\exp(lGDP_{Chi} - lGDP_{Nig}) = GDP_{Chi}/GDP_{Nig} = \exp(2.1) \approx 8$.
    - Hence, the model predicts that Chile's GDP per capita should be 8 times as large as Nigeria's.
- In contrast, using the OLS estimate yields only the factor 3.2!