Lecture 6: Instrumental variables (IV) estimation and Two Stage Least Squares (2SLS)

Prof. dr. Wolter Hassink
Utrecht University School of Economics
w.h.j.hassink@uu.nl

Contents:

- Motivation: omitted variables
- Proxy variables
- Methods of moments estimation
- Instrumental variables
- Instrumental variables: examples
- IV and multiple regression
- Example: OLS versus IV
- 2SLS and lagged dependent variables
- 2SLS and lagged dependent variables: example

Material:

Wooldridge:
Chapter 15: 15.1, 15.2, 15.3 until multiple explanatory variables, 15.4,
Appendix C.4 (moment estimation; or see slides)

Motivation

In Dutch there is an informal saying "*meten is weten*" which can be translated by "*measuring is knowing*"

However: wrongly measured phenomena implies building up the wrong (empirical) knowledge!

Simple examples:

- A thermometer indicates a temperature that is too high (= overestimate of the temperature)
- A scale gives a person's weight that is too low (= underestimate of the weight)

This is a fundamental issue in applied work

Motivation – Examples of omitted variables (see Lecture 1)

Specification 1. For a dataset of individual employees, the log(wage) depends on years of schooling (*educ*) and a set of additional explanatory variables (= *controls*)

$$\log(wage) = \beta_0 + \beta_1 educ + controls + v$$

where v contains the unobserved **ability of the individual**, which may be related to the years of schooling. Thus

$$E(v \mid educ, controls) \neq 0$$

Consequence of Ordinary Least Squares estimator (OLS):

- Positive bias/ overestimate (estimated effect is on too large):
  - Effect of unobserved ability on dependent variable wage is positive: +
  - Unobserved ability is positively correlated to education: +
  - "+" times "+": "+"
  - See Table 3.2 in Wooldridge

Specification 6 (of lecture 1). For a dataset of pupils at primary schools

$$grade = \beta_0 + \beta_1 class\_size + controls + u$$

where u contains the unobserved **parental motivation**.

$$E(u \mid class\_size, controls) \neq 0$$

Consequence of OLS:

- Negative bias/ underestimate (estimated parameter on class size is too small):
    - Effect of unobserved motivation on dependent variable grade is positive: +
    - Unobserved motivation is negatively correlated to class size:
    - "+" times "-": "-"
    - See Table 3.2 in Wooldridge

Two fundamental questions to start any empirical analysis

Question 1: is there any important confounding variable/ omitted variable that is not included as a control variable in the regression equation?

Question 2: in such an equation, does OLS lead to an overestimate or to an underestimate (a positive or negative bias of the estimated parameter)?

Introducing the concept of endogeneity

- Again, we consider the wage equation for working adults.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 abil + u \qquad (1)$$

- Because ability is not directly observed, it is tempting to estimate the equation:

$$\log(wage) = \beta_0 + \beta_1 educ + v \qquad (2)$$

- Because $v = \beta_2 abil + u$

  a regression (2) of log(*wage*) on *educ* will yield an inconsistent estimate of the parameter β_1 (the return to schooling), because *abil* is part of the error term *v*. The exogeneity assumption $E(v \mid x)$ is violated:

$$E(v \mid educ) = E(\beta_2 abil + u \mid educ) \neq 0$$

- Thus in equation (2), the variable *educ* is e**ndogenous**. We are confronted with the thorny issue of **endogeneity**.

- Definition of **endogeneity**: the error term *v* is correlated with the right-hand side variable

How can we reduce the bias of the estimated parameters by using the OLS estimator?

- A larger sample size *n* (more observations) does not reduce the bias. The estimated parameter remains too large (or too small) by expanding the data set.

- Additional explanatory variables (a broader set of controls) will partially solve the issue (a so-called "kitchen sink approach). It is often criticized since the problem of the bias remains unsolved.

- How can we solve this problem?
    - Panel data (weeks 4 and 5)
    - Use proxy variables for unobserved effects (in this case *ability*).
    - Instrumental variables estimation (weeks 6 and 7)

Proxy variables

Proxy variables (section 9.2 of Wooldridge)

Aim: to introduce proxy variables.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 abil + u \tag{3}$$

- In equation (3), we are primarily interested in β_1 and not β_0 or β_2.
- IQ seems to be a reasonable proxy variable for *abil*. Under some conditions, estimation of the following regression leads to a consistent estimate of β_1

$$\log(wage) = \beta_0^* + \beta_1 educ + \beta_2^* IQ + u^* \tag{4}$$

Theory of proxy variables

- Consider the following regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^* + u \tag{5}$$

where the explanatory variable x_2^* (e.g. ability) is unobserved.

- We have a proxy variable x_2 (e.g. IQ) at our disposal.
- The unobserved x_2^* and the observed x_2 should be related. This relationship can be formalized as follows $(\delta_1 \neq 0)$:

$$x_2^* = \delta_0 + \delta_1 x_2 + v_2 \tag{6}$$

- A regression of *y* on x_1 and x_2 yields a consistent estimator of β_1 if the following assumptions hold:
    - **Assumption 1**. The error term *u* in equation (5) is uncorrelated with x_1 and x_2.
    - **Assumption 2.** The error term v_2 in equation (6) is uncorrelated with x_1 and x_2:

$$E(x_2^* \mid x_1, x_2) = \delta_0 + \delta_1 x_2 \tag{7}$$

- For our example, where IQ is a proxy for ability, equation (7) boils down to $E(ability \mid educ, IQ) = \delta_0 + \delta_1 IQ$

- Substituting (6) in (5) gives:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \underbrace{(\delta_0 + \delta_1 x_2 + v_2)}_{x_2^*} + u$$

$$= \beta_0 + \beta_2 \delta_0 + \beta_1 x_1 + \beta_2 \delta_1 x_2 + \beta_2 v_2 + u$$

$$= \beta_0^* + \beta_1 x_1 + \beta_2^* x_2 + u^* \qquad (8)$$

where $\beta_0^* = \beta_0 + \beta_2 \delta_0$; $\beta_2^* = \beta_2 \delta_1$; $u^* = u + \beta_2 v_2$

- Estimation of equation (8) yields a consistent estimate of β_1 if the error term u^* is uncorrelated with x_1 and x_2.
- Because u^* consists of u and v_2 ($u^* = u + \beta_2 v_2$), it is assumed that :
    - u is uncorrelated with x_1 and x_2 (assumption 1)
    - v_2 is uncorrelated with x_1 and x_2 (assumption 2)

Example 1: wage2.dta

. reg lwage educ exper tenure married south urban black

```
      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  7,   927) =   44.75
       Model |  41.8377619     7  5.97682312           Prob > F      =  0.0000
    Residual |  123.818521   927  .133569063           R-squared     =  0.2526
-------------+------------------------------           Adj R-squared =  0.2469
       Total |  165.656283   934  .177362188           Root MSE      =  .36547

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0654307   .0062504    10.47   0.000     .0531642    .0776973
       exper |    .014043   .0031852     4.41   0.000      .007792     .020294
      tenure |   .0117473    .002453     4.79   0.000     .0069333    .0165613
     married |   .1994171   .0390502     5.11   0.000     .1227801     .276054
       south |  -.0909036   .0262485    -3.46   0.001     -.142417   -.0393903
       urban |   .1839121   .0269583     6.82   0.000     .1310056    .2368185
       black |  -.1883499   .0376666    -5.00   0.000    -.2622717   -.1144281
       _cons |   5.395497    .113225    47.65   0.000      5.17329    5.617704
------------------------------------------------------------------------------
```

. reg lwage educ exper tenure married south urban black IQ

```
      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  8,   926) =   41.27
       Model |  43.5360162     8  5.44200202           Prob > F      =  0.0000
    Residual |  122.120267   926  .131879338           R-squared     =  0.2628
-------------+------------------------------           Adj R-squared =  0.2564
       Total |  165.656283   934  .177362188           Root MSE      =  .36315

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0544106   .0069285     7.85   0.000     .0408133     .068008
       exper |   .0141458   .0031651     4.47   0.000     .0079342    .0203575
      tenure |   .0113951   .0024394     4.67   0.000     .0066077    .0161825
     married |   .1997644   .0388025     5.15   0.000     .1236134    .2759154
       south |  -.0801695   .0262529    -3.05   0.002    -.1316916   -.0286473
       urban |   .1819463   .0267929     6.79   0.000     .1293645    .2345281
       black |  -.1431253   .0394925    -3.62   0.000    -.2206304   -.0656202
          IQ |   .0035591   .0009918     3.59   0.000     .0016127    .0055056
       _cons |   5.176439   .1280006    40.44   0.000     4.925234    5.427644
------------------------------------------------------------------------------
```

Note that without the proxy, *IQ*, the estimated parameter on *educ* is biased upwards (Corr(*abil*, *wage*)>0 and Corr(*educ*, *abil*)>0). See Table 3.2 of Wooldridge (pp. 91) and Lecture 1.

Methods of moments estimation

Methods of moments estimation

Aim: to introduce the method of moments estimation, which will improve understanding of IV.

- Consider the following bivariate regression model:

  $$y = \beta_0 + \beta_1 x + u \quad (9)$$

- The parameters of this model can be estimated by OLS
- OLS will yield consistent estimates if the CLM-assumptions are satisfied. These are:
  1) Linear model
  2) Random sampling
  3) Exogeneity
  4) No perfect multicollinearity (sampling variation in *x*).

- The exogeneity assumption is crucial. It states that:

  $$E(u \mid x) = 0 \quad (10)$$

- The implications of exogeneity is that the covariance between *u* and *x* is zero. Equation (10) implies that:
  - $E(u) = 0$
  - $E(xu) = 0$, which implies that *u* and *x* are uncorrelated
    - $Cov(x,u) = E(x - Ex)(u - Eu) = Exu - ExEu = Exu$ (since $E(u) = 0$)
- Equation (9) may be used to demonstrate both implications:
  - $E(u) = E(y - \beta_0 - \beta_1 x) = 0$ (11a)
  - $E(xu) = E(x(y - \beta_0 - \beta_1 x)) = 0$ (11b)
- Equations (11a,b) is a set of two so-called 'moment conditions' that apply to the population. The conditions are in terms of expectations on the error term and the explanatory variables (moments) that are implied by the population regression model.

- When using the methods of moments estimation, we apply both moment conditions on the estimator through the corresponding sampling moments. That is, the moment estimators $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ for $\beta = (\beta_0, \beta_1)$ are solved from
    - $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ (12a)
    - $\frac{1}{n}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ (12b)
- To solve a system of **normal equations** (12a and 12b), which have two linear equations with two unknowns, $\hat{\beta}_0$ and $\hat{\beta}_1$, the following estimators are required:
    - $\overline{y}_i - \hat{\beta}_1 \overline{x} = \hat{\beta}_0$ (13a)
    - $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$ (13b)
- Note that the estimators (13a) and (13b) are identical to the OLS estimators.
- This is because the set of normal equations (12a) and (12b) also follow from the same first-order conditions underlying ordinary least squares optimization (see equations (2.19) and (2.17)).
- Hence, the OLS estimators can also be considered methods of moments estimators.

- In the case of the multivariate regression model (k + 1 parameters to be estimated):
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- the exogeneity assumption
$$E(u \mid x_1, \ldots, x_k) = 0$$
implies the following k +1 moment conditions
$$E(u) = E\{y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)\} = 0$$
$$E(x_j u) = E\{x_j [y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)]\} = 0 \quad j = 1, \ldots, k \quad (14)$$

- the moment estimators $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ follow from solving the corresponding sample moments of the k + 1 equations (14)
$$\frac{1}{n} \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki})) = 0$$
$$\frac{1}{n} \sum_{i=1}^{n} x_{ji} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki})) = 0 \qquad j = 1, \ldots, k \quad (15)$$

Instrumental variables

Instrumental variable methods

Aim: to introduce instrumental variables.

- Consider again the following:
  $$\log(wage) = \beta_0 + \beta_1 educ + u \tag{15}$$
  $$u = \beta_2 abil + v$$
- OLS only yields consistent estimates if the following conditions hold:
  Condition 1: $E(u) = E\{\log(wage) - (\beta_0 + \beta_1 educ)\} = 0$
  Condition 2: $E(educ * u) = E\{educ[\log(wage) - (\beta_0 + \beta_1 educ)]\} = 0$
- However, since ability is part of the error term *u* in equation (15), a regression of log(*wage*) on *educ* will yield an inconsistent estimate of β_1, because $E(u \mid educ) = E(\beta_2 abil + v \mid educ) \neq 0$, violating Condition 2. In other words: *educ* is **endogenous** in equation (15), so that it is a specific RHS-variable that is correlated to the error term *u*.
- Formally: $E(educ * u) = E\{educ[\log(wage) - (\beta_0 + \beta_1 educ)]\} \neq 0$
- From this point, we replace *y* with log(*wage*) and *x* with *educ*, so that equation (15) can be written as:
  $$y = \beta_0 + \beta_1 x + u \tag{16}$$
- Suppose that we have an instrumental variable (IV), *z*, which replaces the endogenous right-hand side variable, *x*, in equation (16).
- The IV estimator can be considered a method of moments estimator that is based on the following moments:
  - $E(u) = E(y - (\beta_0 + \beta_1 x)) = 0$ (17a)
  - $E(zu) = E\{z[y - (\beta_0 + \beta_1 x)]\} = 0$ (17b)

- Equations (17a) and (17b) can be rewritten as:
    - $\beta_1 = \dfrac{Cov(z,y)}{Cov(z,x)}$ (18a)
    - $\beta_0 = E(y) - \dfrac{Cov(z,y)}{Cov(z,x)} E(x)$ (18b)
- Note that the OLS estimator uses the following formulas:

$$\beta_1 = \frac{Cov(x,y)}{Var(x)} \tag{18c}$$

$$\beta_0 = E(y) - \frac{Cov(x,y)}{Var(x)} E(x) \tag{18d}$$

Note that the IV-estimator (18a) is equal to OLS-estimator (18c) if $z = x$

{since $Cov(z,x) = ...(\text{if } z = x)... = Cov(x,x) = Var(x)$ }

Relevance and exogeneity of instrumental variables

Aim: to discuss relevance and exogeneity.

- An instrumental variable z should satisfy two important criteria:
  - Instrument exogeneity: z and the error term u should be uncorrelated (see equation (17b)). $Cov(z,u)=0$
  - Instrument relevance: the variable x (which is **endogenous**) and the instrument z should be correlated (otherwise the denominator of (18a) is zero and β_1 cannot be computed. $Cov(z,x)\neq 0$
- Instrument exogeneity cannot be tested, only by counterexamples can it be claimed that an instrument is not exogenous.
- It is possible to check for instrument relevance by running the regression:

$$x=\pi_0+\pi_1 z+v \tag{19}$$

  - If $Cov(z,x)\neq 0$, $\pi_1\neq 0$
- Estimation using instrumental variables is basically a method of moments estimation: the IV-estimator $\hat{\beta}_0^{IV}$ and $\hat{\beta}_1^{IV}$ follow from the sample counterparts of the moment conditions (17a) and (17b). In other words,
  - $$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\beta}_0^{IV}-\hat{\beta}_1^{IV}x_i)=0 \tag{20a}$$
  - $$\frac{1}{n}\sum_{i=1}^{n}z_i(y_i-\hat{\beta}_0^{IV}-\hat{\beta}_1^{IV}x_i)=0 \tag{20b}$$
- This yields the following IV-estimators

$$\hat{\beta}_1^{IV}=\frac{\sum_{i=1}^{n}(z_i-\overline{z})(y_i-\overline{y})}{\sum_{i=1}^{n}(z_i-\overline{z})(x_i-\overline{x})} \tag{21a}$$

$$\hat{\beta}_0^{IV}=\overline{y}-\hat{\beta}_1^{IV}\overline{x} \tag{21b}$$

Consistency of IV-estimator

Aim: to discuss the consistency of IV.

$$y = \beta_0 + \beta_1 x + u$$

- It can be shown that

$$\text{plim } \hat{\beta}_1^{IV} = \beta_1 + \frac{Cov(z,u)}{Cov(z,x)} \frac{\sigma_u}{\sigma_x} \tag{22}$$

- In other words, the estimator will be consistent if:
    - $Cov(z,u) = 0$ (instrument exogeneity)
    - $Cov(z,x) \neq 0$ (instrument relevance)

- When weak instrument: The denominator $Cov(z,x)$ is relatively small, so that the IV-estimator is inconsistent.

Standard error of IV-estimator

Aim: to discuss the implications of IV for standard error.

- Recall from Chapter 2 (Wooldridge) that the standard error of the OLS-estimator $\hat{\beta}_1$ is

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\hat{\sigma}^2}{SST_x} \tag{23}$$

- It can be shown that the standard error of the IV-estimator $\hat{\beta}_1^{IV}$:

$$Var(\hat{\beta}_1^{IV}) = \frac{\hat{\sigma}^2}{SST_x R_{x,z}^2} \tag{24}$$

Where $R_{x,z}^2$ is the R^2 from the regression of *x* on *z* (equation (19)).

- Note that (23) equals (24) only if $R_{x,z}^2$ =1
    - IV will yield a smaller *t*-value for $\hat{\beta}_1^{IV}$ than the *t*-value for $\hat{\beta}_1$ obtained by OLS!
    - If the instrument *z* is weak (when $R_{x,z}^2$ is small, so that the denominator of (24) is small), $Var(\hat{\beta}_1^{IV})$ will be larger.
    - Conclusion: if the correlation between the explanatory variable *x* and the instrumental variable *z* is relatively weak, $R_{x,z}^2$ will be small.
    - As a result, $Var(\hat{\beta}_1^{IV})$ will be large. In other words, the *t*-value for $\hat{\beta}_1^{IV}$ will be small.

Instrumental variables: examples

Example 2: Application of IV to the wage equation and ability bias (I) using *wage2.dta*

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

- Question: what is a reasonable instrument for *educ*?
- The proxy variable *IQ* is NOT a valid instrument for education:
    - *IQ* satisfies the criterion of instrument relevance: $Cov(z, x) \neq 0$. *IQ* is strongly correlated with education.
        - Check this using an *F*-test on *IQ* in equation (19): $educ = \pi_0 + \pi_1 IQ + v$

. reg educ IQ

```
      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  1,   933) =  338.02
       Model |  1198.55887     1  1198.55887           Prob > F      =  0.0000
    Residual |  3308.26038   933  3.54583106           R-squared     =  0.2659
-------------+------------------------------           Adj R-squared =  0.2652
       Total |  4506.81925   934  4.82528828           Root MSE      =   1.883

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   .0752564   .0040933    18.39   0.000     .0672233    .0832896
       _cons |     5.8463   .4191273    13.95   0.000     5.023758    6.668841
------------------------------------------------------------------------------
```

. test IQ

```
 ( 1)  IQ = 0

       F(  1,   933) =  338.02
            Prob > F =    0.0000
```

- We check for relevance by examining the value of the F-statistic of the instrumental variable in the regression of the endogenous variable *Education* on *IQ*. The *F*-statistic (338.02) is larger than 10.
- *IQ* does not satisfy the criterion of instrument exogeneity:

$Cov(z,u) = Cov(IQ,u) \neq 0$, since ability is part of the error term *u*.

Example 3: Application of IV to the wage equation and ability bias (II)

- A valid instrument for education may be the education of each individuals' father. We can show this by:
  - Instrument relevance: $Cov(educ, educf) \neq 0$
    - Check equation (19): $educ = \pi_0 + \pi_1 feduc + v$

```
. reg  educ feduc

      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  1,   739) =  164.72
       Model |   678.68578     1   678.68578           Prob > F      =  0.0000
    Residual |  3044.92826   739  4.12033593           R-squared     =  0.1823
-------------+------------------------------           Adj R-squared =  0.1812
       Total |  3723.61404   740  5.03191086           Root MSE      =  2.0299

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       feduc |   .2901432   .0226071    12.83   0.000     .2457615     .334525
       _cons |   10.64956   .2427211    43.88   0.000     10.17306    11.12607
------------------------------------------------------------------------------

. test feduc

 ( 1)  feduc = 0

       F(  1,   739) =  164.72
            Prob > F =    0.0000
```

- *educf* must also satisfy the criterion of instrument exogeneity, making it is also necessary to assume that the father's education is uncorrelated with ability. However,
  - $Cov(educf, u) \neq 0$ if ability of child is partly due to the education of father.
  - Therefore, *educf* is not a valid instrument.

Example 4: Application to the wage equation and ability bias (III)

- In this case the Instrumental variable is the month of birth
  - Instrument relevance: born late in year: first day at school at later age; years of schooling ↓
  - However, this is also a weak instrument as the relation between month of birth and education is low. $Cov(z, x)$ is small (equation (22)), so that the IV-estimator is inconsistent.
  - Instrument exogeneity: correlation between month of birth and ability is zero.

- Conclusion: valid instruments are hard to find.

Example 5: loneliness (ongoing research Wolter Hassink and Reneé van Eyden (university of Pretoria))

We consider the regression equation

$$Employment_{it} = \alpha_i + \beta Loneliness_{it} + Controls + \delta_t + u_{it}$$

For which the dependent variable is Employed (yes/no (see lecture 8)). *Loneliness* is a self-reported variable, which is endogenous. The confounding variable is the quality of the network. Applied to survey data from South Africa.

Instrumental variable *z*: trust of stranger to return a lost wallet containing R200 and a name card of owner.

IV and multiple regression

IV of the multiple regression model

- The IV estimator for the bivariate model is easily extended to the case with multiple RHS-variables. Consider the case of two RHS-variables:

  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$

  - Where z_1 is an exogenous RHS-variable $E(u_1 \mid z_1) = 0$, so that $E(u_1) = 0$ and $Cov(z_1, u_1) = 0$
  - y_2 is an endogenous RHS-variable: $E(u_1 \mid y_2) \neq 0$ or $Cov(u_1, y_2) \neq 0$
- Suppose there is a suitable instrument, z_2, for y_2. This means that the instrumental variable z_2 satisfies the following criteria:
  - Instrument exogeneity: z_2 and u_1 are uncorrelated. $E(u_1 z_2) = 0$ or $Cov(z_2, u_1) = 0$
  - Note again that we cannot test for instrument exogeneity.
  - Instrument relevance: the instrumental variable z_2 should predict y_2. In other words, consider the following regression model:

    $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$

    and we assume that $\pi_2 \neq 0$
- Given the three moment conditions
  - $E(u_1) = 0$
  - $Cov(z_1, u_1) = 0$ or $E(z_1 u_1) = 0$ (exogenous RHS)
  - $Cov(z_2, u_1) = 0$ or $E(z_2 u_1) = 0$ (instrument exogeneity)

- The moment estimators (instrumental estimators) can be constructed by solving the sample counterpart of the moment conditions.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - (\hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} y_{2i} + \hat{\beta}_2^{IV} z_{1i})) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} z_{1i}(y_i - (\hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} y_{2i} + \hat{\beta}_2^{IV} z_{1i})) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} z_{2i}(y_i - (\hat{\beta}_0^{IV} + \hat{\beta}_1^{IV} y_{2i} + \hat{\beta}_2^{IV} z_{1i})) = 0$$

Two Stage Least Squares estimator (2SLS)

Aim: to introduce 2SLS-estimator

- Again, we consider an equation with two RHS-variables.

  $$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1 \tag{25}$$

  which is also referred to as a structural equation.
- Assume that the endogenous y_2 is related to the instrument variable z_2. $Cov(y_2, z_2) \neq 0$.
- We can construct what is known as a reduced-form equation for y_2, which depends only on exogenous variables z_1 and z_2:

  $$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2 \tag{26}$$

  where $\pi_2 \neq 0$ and it is assumed that the explanatory variables are uncorrelated with the error term. $E(v_2) = 0$, $Cov(z_1, v_2) = 0$, and $Cov(z_2, v_2) = 0$.

  There are two stages:
- **First stage.** Regress y_2 on z_1 and z_2 (equation (26)) by OLS and determine the fitted value of y_2, using the estimated parameters:

  $$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2$$
- **Second stage:** Regress the structural equation (25), using the fitted value $\hat{y}_2$, instead of its actual value:

  $$y_1 \text{ on } \hat{y}_2 \text{ and } z_1 \tag{27}$$
- Note that the *t*-values for (27) are wrong using the above procedure, because the standard error of $\hat{y}_2$ is incorrect. 2SLS estimation commands in software packages corrects this fault.

Example 6: 2SLS

- The structural model:
  $\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u_1$
- To illustrate, we assume that *feduc* (father's education) may be used to instrument *educ* in a wage equation.
- In other words, we assume that Corr(*feduc*,*u*)=0.
- Estimate the parameters of the reduced-form equation by OLS:
- $educ = \pi_0 + \pi_1 exper + \pi_2\, feduc + v_2$

```
. reg educ exper feduc

      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  2,   738) =  162.65
       Model |  1139.16105     2  569.580524           Prob > F      =  0.0000
    Residual |  2584.45299   738  3.50196882           R-squared     =  0.3059
-------------+------------------------------           Adj R-squared =  0.3040
       Total |  3723.61404   740  5.03191086           Root MSE      =  1.8714

------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |  -.1898551   .0165568   -11.47   0.000     -.222359   -.1573511
       feduc |   .2266428   .0215649    10.51   0.000     .1843069    .2689786
       _cons |   13.45928   .3318292    40.56   0.000     12.80784    14.11072
------------------------------------------------------------------------------
```

- From the above *F*-test we may conclude that *feduc* is a relevant instrument.

```
. test feduc

 ( 1)  feduc = 0

       F(  1,   738) =  110.46
            Prob > F =    0.0000
```

- Conclusion: *feduc* is a relevant instrument, since the F-statistic is larger than 10.

- After the first-stage equation, we can determine the predicted value of the endogenous variable *educ*:

```
. predict p_educ
(option xb assumed; fitted values)
```

- If *educ* (p_educ) is included in the structural equation,

  $$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + u_1$$

  it will give the incorrect *t*-values.

- In order to obtain the right *t*-values 2SLS commands must be used directly in Stata.

```
. ivreg lwage (educ=feduc) exper, first
```

```
First-stage regressions
-----------------------
      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  2,   738) =  162.65
       Model |  1139.16105     2  569.580524           Prob > F      =  0.0000
    Residual |  2584.45299   738  3.50196882           R-squared     =  0.3059
-------------+------------------------------           Adj R-squared =  0.3040
       Total |  3723.61404   740  5.03191086           Root MSE      =  1.8714
------------------------------------------------------------------------------
        educ |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |  -.1898551   .0165568   -11.47   0.000     -.222359   -.1573511
       feduc |   .2266428   .0215649    10.51   0.000     .1843069    .2689786
       _cons |   13.45928   .3318292    40.56   0.000     12.80784    14.11072
------------------------------------------------------------------------------

Instrumental variables (2SLS) regression
      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  2,   738) =   23.92
       Model |  5.40902478     2  2.70451239           Prob > F      =  0.0000
    Residual |  123.641133   738  .167535411           R-squared     =  0.0419
-------------+------------------------------           Adj R-squared =  0.0393
       Total |  129.050158   740  .174392106           Root MSE      =  .40931
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1410919   .0208115     6.78   0.000     .1002352    .1819487
       exper |   .0378939   .0060062     6.31   0.000     .0261026    .0496852
       _cons |   4.447279   .3415488    13.02   0.000     3.776756    5.117802
------------------------------------------------------------------------------
Instrumented:  educ
Instruments:   exper feduc
------------------------------------------------------------------------------
```

- The parameter estimate on education indicates that for each addition year of schooling, education increases by 14.1 percent, keeping constant experience.

Example: OLS versus IV

Example 7: OLS versus IV

```
Instrumental variables (2SLS) regression
      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  2,   738) =   23.92
       Model |  5.40902478     2  2.70451239           Prob > F      =  0.0000
    Residual |  123.641133   738  .167535411           R-squared     =  0.0419
-------------+------------------------------           Adj R-squared =  0.0393
       Total |  129.050158   740  .174392106           Root MSE      =  .40931
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1410919   .0208115     6.78   0.000     .1002352    .1819487
       exper |   .0378939   .0060062     6.31   0.000     .0261026    .0496852
       _cons |   4.447279   .3415488    13.02   0.000     3.776756    5.117802
```

- **Compare with OLS:**

```
. reg lwage educ exper
      Source |       SS       df       MS              Number of obs =     741
-------------+------------------------------           F(  2,   738) =   58.88
       Model |   17.759082     2  8.87954102           Prob > F      =  0.0000
    Residual |  111.291076   738  .150800916           R-squared     =  0.1376
-------------+------------------------------           Adj R-squared =  0.1353
       Total |  129.050158   740  .174392106           Root MSE      =  .38833
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .076621   .0071241    10.76   0.000     .0626351     .090607
       exper |    .022773   .0037172     6.13   0.000     .0154754    .0300706
       _cons |   5.497093   .1228384    44.75   0.000     5.255939    5.738248
------------------------------------------------------------------------------
```

- Note that the interpretation of the estimated parameters does not differ between OLS and 2SLS.
- Note that the $\hat{\beta}_1$ of the 2SLS estimate is larger than the $\hat{\beta}_1$ estimate. This is unexpected, as it has been shown before that the $\hat{\beta}_1$ of OLS is an overestimate of the true unknown parameter β_1.
    - See Slide 4 above - the estimated parameter on *educ* is biased upwards (Corr(*abil*, *wage*)>0 and Corr(*educ*, *abil*)>0). See Table 3.2 (page 91) for the direction of the bias (lecture 1).
- Note that the standard errors of the estimated parameters with IV are substantially larger than those of OLS (compare equations (24) and (23)).

2SLS and lagged dependent variables

Application of IV to regression model with lagged dependent variable

Aim: to reconsider IV for a panel data model with a lagged dependent variable

- IV can be applied to panel data. We will consider a model with a lagged dependent variable.
- Remember from lecture 4 that the assumption of strict exogeneity is violated in dynamic regression models, such as:

$$\log(wage_{it}) = \gamma \log(wage_{it-1}) + a_i + u_{it} \qquad i = 1,...,N; t = 2,...,T \qquad (28)$$

- The above model addresses the following question: how persistent are wages after controlling for unobserved heterogeneity (a_i)?
- We assume that $y_{it-1} = \log(wage_{it-1})$ is a contemporaneously exogenous variable.
- Obviously, lagged log wage is correlated with a_i.
- In addition, u_{it} is correlated with $x_{it+1} = y_{it}$, meaning that y_{it-1} is NOT a strictly exogenous variable.
- It can be demonstrated that for a panel data model with a lagged dependent variable, the pooled OLS estimator, the fixed effects estimators and the first-difference estimator are inconsistent estimators that deliver biased parameter estimates.

First-difference estimator with IV

Aim: to show the correct procedure for a panel data model with a lagged dependent variable.

The wage equation is

$$\log(wage_{it}) = \gamma \log(wage_{it-1}) + a_i + u_{it} \qquad i = 1,...,N; t = 2,...,T \qquad (28)$$

That we rewrite as

$$y_{it} = \gamma y_{it-1} + a_i + u_{it} \qquad i = 1,...,N; t = 2,...,T \qquad (28')$$

- The IV approach can be applied to get a consistent estimator of γ.
- The starting position is the model in first differences (FD):
  $(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + (u_{it} - u_{it-1}) \quad i = 1,...,N; t = 3,...,T \qquad (29)$
- An FD-estimator on (29) is biased, since
  $E(u_{it} - u_{it-1})(y_{it-1} - y_{it-2}) \neq 0$

- Instead, we can instrument the endogenous RHS-variable $y_{it-1} - y_{it-2}$ with y_{it-2}, since we have also assumed that u_{it} is i.i.d.:
  - Instrument exogeneity:
    $Cov(y_{it-2}, u_{it} - u_{it-1}) = Cov(y_{it-2}, u_{it}) - Cov(y_{it-2}, u_{it-1}) = 0$
  - Instrument relevance:
    $Cov(y_{it-2}, y_{it-1} - y_{it-2}) = Cov(y_{it-2}, y_{it-1}) - Cov(y_{it-2}, y_{it-2}) =$
    $= Cov(y_{it-2}, y_{it-1}) - Var(y_{it-2}) \neq 0$
- Thus we can apply 2SLS on
  $(y_{it} - y_{it-1}) = \gamma(y_{it-1} - y_{it-2}) + (u_{it} - u_{it-1})$
- where y_{it-2} is used as an instrument for $(y_{it-1} - y_{it-2})$.
- y_{it-2} can also be referred to as an **internal instrumental variable**. I.e., a lagged value of the variable *y* will be used as an instrumental variable. In the applications before (e.g. father's education), we were applying **external instrumental variables**.

Thus, in terms of the 2SLS-estimator we do the following:

- We apply a structural model (equation (25), excluded z_1):

$$y_1 = \beta_0 + \beta_1 y_2 + u_1 \tag{30}$$

- The relation between equations (28') and (30) is as follows:
    - The dependent-variable $(y_{it} - y_{it-1})$ of equation (28') corresponds to y_1 of equation (30)
    - The RHS endogenous variable $(y_{it-1} - y_{it-2})$ corresponds to y_2
    - The instrument y_{it-2} corresponds to z_2
    - The error term $(u_{it} - u_{it-1})$ corresponds to u_1

2SLS and lagged dependent variables: example

Example 8: application of IV to a dynamic wage equation

```
. xtset nr year
       panel variable:  nr (strongly balanced)
        time variable:  year, 1980 to 1987
                delta:  1 unit
```

- Pooled OLS (with Newey West clustered s.e.; yields biased estimates)

```
. reg lwage l.lwage, cluster(nr)

Linear regression                                      Number of obs =    3815
                                                       F(  1,   544) =  591.68
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4162
                                                       Root MSE      =  .39648

                                   (Std. Err. adjusted for 545 clusters in nr)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .6265667   .0257587    24.32   0.000     .5759681    .6771653
       _cons |   .6718241    .041685    16.12   0.000     .5899409    .7537073
------------------------------------------------------------------------------
```

- Fixed-effects estimator (yields biased estimates)

```
. xtreg lwage l.lwage, fe

Fixed-effects (within) regression               Number of obs      =      3815
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.0366                         Obs per group: min =         7
       between = 0.9541                                        avg =       7.0
       overall = 0.4162                                        max =         7

                                                F(1,3269)          =    124.21
corr(u_i, Xb)  = 0.7176                         Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .1740662   .0156184    11.14   0.000     .1434433    .2046891
       _cons |   1.404015   .0258827    54.25   0.000     1.353267    1.454763
-------------+----------------------------------------------------------------
     sigma_u |  .33713398
     sigma_e |  .34516481
         rho |  .48823138   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3269) =     3.24           Prob > F = 0.0000
```

- First-difference estimator:

```
. reg d.lwage d.l.lwage, cluster(nr)

Linear regression                                      Number of obs =    3270
                                                       F(  1,   544) =  321.49
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1784
                                                       Root MSE      =  .37962

                                   (Std. Err. adjusted for 545 clusters in nr)
------------------------------------------------------------------------------
             |               Robust
     D.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         LD. |  -.3908602   .0217989   -17.93   0.000    -.4336805   -.3480399
       _cons |   .0853993   .0045781    18.65   0.000     .0764064    .0943922
------------------------------------------------------------------------------
```

- 2SLS on within with $\log(wage_{it-2})$ as the instrumental variable:

```
. ivreg d.lwage (d.l.lwage=l2.lwage), cluster(nr) first

First-stage regressions
-----------------------
      Source |       SS       df       MS              Number of obs =    3270
-------------+------------------------------           F(  1,  3268) =  823.42
       Model |  134.747747     1  134.747747           Prob > F      =  0.0000
    Residual |  534.790562  3268  .163644603           R-squared     =  0.2013
-------------+------------------------------           Adj R-squared =  0.2010
       Total |  669.538309  3269  .204814411           Root MSE      =  .40453
------------------------------------------------------------------------------
    LD.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L2. |  -.3820328   .0133135   -28.70   0.000    -.4081363   -.3559292
       _cons |   .6743101   .0222918    30.25   0.000     .6306028    .7180173
------------------------------------------------------------------------------

Instrumental variables (2SLS) regression               Number of obs =    3270
                                                       F(  1,   544) =   11.32
                                                       Prob > F      =  0.0008
                                                       R-squared     =       .
                                                       Root MSE      =  .44385
                                   (Std. Err. adjusted for 545 clusters in nr)
------------------------------------------------------------------------------
             |               Robust
     D.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         LD. |   .1172337   .0348386     3.37   0.001      .048799    .1856683
       _cons |   .0509978   .0039317    12.97   0.000     .0432747     .058721
------------------------------------------------------------------------------
Instrumented:  LD.lwage
Instruments:   L2.lwage
------------------------------------------------------------------------------
```

- Using the first-stage regression we can conclude that l2.lwage is a relevant variable, since *F*-test on l2.lwage is larger than 10. Apply the command test l2.lwage after first stage. See example above.

Winding up: following steps in 2SLS

Step 1: Are there any important confounding variables/ omitted variables in the regression equation?

Step 2: Is there any endogenous variable on the right-hand side of the regression equation.

Step 3: OLS: positive bias or negative bias (Table 3.2 from Wooldridge)?

Step 4: Introduce an (internal or external) instrumental variable that is potentially correlated with the endogenous variable on the right-hand side of the regression equation. The instrumental variable should be uncorrelated with the error term of the regression equation.

Step 5: check the F-statistic of the first-stage (a regression of the endogenous variable on the instrumental variable, corrected for all further variables on the right-hand side of the equation). F-statistic larger than 10?

Step 6: Apply the 2SLS estimator

a) Compare the estimated parameter of 2SLS with the estimated OLS-parameter (should move in the right direction).

b) Compare the standard error of 2SLS with the standard error of the OLS-parameter (should become larger).