# Instrumental variables estimation

Bernard Lejeune

Supplemental lecture notes VI
Advanced Econometrics
HEC-University of Liège
Academic year 2021-2022

- These lecture notes restate, in matrix form and with more details, the main results of Sections 15-2, 15-3 and 15-6 of Wooldridge (2016).

# 1. Regression and instrumental variables estimation

- Instrumental variables estimation provides a way to consistently estimate the parameters of a linear regression model when – for different possible reasons – one or more of its explanatory variables are endogenous, i.e., are correlated (have nonzero covariance) with the error term of the model. Instrumental variables methods may be used with cross-sectional data, time series data as well as panel data. We here focus on the cross-sectional case.

- We suppose that interest lies in estimating the parameters of an usual linear regression model as described in the following assumption:

  2SLS.1 Linearity in parameters

  The population model, describing the relationship between the dependent $y$ and the explanatory variables $(x_2, ..., x_k)$ for an individual $i$ draw at random from the population, can be written as:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i \\ \Leftrightarrow \quad y_i &= X_i \beta + u_i \end{aligned} \tag{1}$$

  where $X_i = (1, x_{i2}, ..., x_{ik})$ is a $1 \times k$ (row) vector (including a constant), $\beta = (\beta_1, \beta_2, ..., \beta_k)$ is a $k \times 1$ (column) vector of unknown parameters, and $u_i$ is an error term.

As usual with cross-sectional data, it is also supposed that the observations are obtained by random sampling:

2SLS.2 Random sampling
The available data are realizations of a random sample of size $n$, $\{(y_i, X_i, Z_i)$: $i = 1, ..., n\}$, following the population model in assumption 2SLS.1.

The supplemental vector $Z_i$ of data appearing in assumption 2SLS.2 stands for a $1 \times l$ (row) vector of variables, which usually overlaps with $X_i$, and whose the exact role will be explained below. Stacking all observations of a random sample of size $n$, let $Y$, $X$, $Z$ and $u$ stand for:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \quad \text{and} \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

where $Y$ and $u$ are $n \times 1$ vectors, $X$ is a $n \times k$ matrix (whose the $i$-th row is equal to $X_i$), and $Z$ is a $n \times l$ matrix (whose the $i$-th row is equal to $Z_i$), so that model (1) can as usual be written as:

$$Y = X\beta + u$$

- In model (1), it is usually assumed that $E(u_i|X_i) = 0$, i.e., that the systematic part $X_i\beta$ of model (1) is the conditional mean of $y_i$ given $X_i$: $E(y_i|X_i) = X_i\beta$. As we know, by the law of iterated expectations, the zero conditional mean assumption $E(u_i|X_i) = 0$ implies that each explanatory variable is uncorrelated with the error of the model:

$$E(X_i'u_i) = 0 \quad \Leftrightarrow \quad E(u_i) = 0 \text{ and } Cov(x_{ij}, u_i) = 0, \text{ for } j = 2, ..., k \tag{2}$$

From Property 6' in the supplemental lecture notes I (hereafter SLN-I), we also know that the zero correlation condition (2) is – along with random sampling and no perfect collinearity – actually enough for the OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$ to provide a consistent estimator of $\beta$ in model (1). As a matter of fact, following the method of moments approach to estimation[1], the OLS estimator $\hat{\beta}$ may actually be directly derived from the zero correlation condition (2). The method of moments approach simply suggests estimating $\beta$ based on the sample counterpart of the moment condition $E(X_i'u_i) = E[X_i'(y_i - X_i\beta)] = 0$, i.e., choosing as an estimator $\hat{\beta}$ of $\beta$ the solution of the sample moment condition:

$$\frac{1}{n}\sum_{i=1}^{n} X_i'(y_i - X_i\hat{\beta}) = 0$$

or equivalently:

$$\sum_{i=1}^{n} X_i'(y_i - X_i\hat{\beta}) = X'(Y - X\hat{\beta}) = 0$$

$$\Leftrightarrow \quad X'X\hat{\beta} = X'Y$$

[1] For a discussion of the method of moments approach to estimation, see Wooldridge (2016), Appendix C-4a.

whose the solution is indeed the OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$.

- Instrumental variables methods are concerned with the situation where the zero correlation condition (2) does not hold, i.e., the situation where one or more of the explanatory variables $x_{ij}$ – which are then referred to as endogenous – are correlated the error of the model. Two typical situations[2] where this happens are (i) when there is a measurement error in one (or more) explanatory variable(s), and (ii) when a relevant explanatory variable is omitted from the model and that this omitted variable is correlated with one (or more) of the included explanatory variable(s) :

  (i) In the classical errors-in-variables case, we are for example interested in the population model :

  $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x^*_{i3} + v_i \ , \ \text{where } E(v_i|x_{i2}, x^*_{i3}) = 0 \tag{3}$$

  but we do not observed $x^*_{i3}$. Instead, we observe $x_{i3}$, which is assumed to be an unbiased measurement of $x^*_{i3}$, unrelated to $x_{i2}$ :

  $$E(x_{i3}|x_{i2}, x^*_{i3}) = x^*_{i3} \quad \Leftrightarrow \quad x_{i3} = x^*_{i3} + e_i \ , \ \text{where } E(e_i|x_{i2}, x^*_{i3}) = 0 \tag{4}$$

  and the model actually estimated is :

  $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \ , \ \text{where } u_i = (v_i - \beta_3 e_i) \tag{5}$$

  Under (3) and (4), by the law of iterated expectations, we have :

  $$\begin{aligned} E(v_i) &= 0 \ , \ E(x_{i2}v_i) = 0 \ , \ E(x^*_{i3}v_i) = 0 \\ E(e_i) &= 0 \ , \ E(x_{i2}e_i) = 0 \ , \ E(x^*_{i3}e_i) = 0 \end{aligned} \tag{6}$$

  so that we have :

  $$E(u_i) = 0 \quad \text{and} \quad Cov(x_{i2}, u_i) = E(x_{i2}u_i) = 0$$

  but, because $x_{i3}$ is by assumption correlated[3] with $e_i$, we however have[4] :

  $$Cov(x_{i3}, u_i) = E(x_{i3}u_i) = E\left[x_{i3}(v_i - \beta_3 e_i)\right] \neq 0$$

  i.e., the estimated model (5) does not satisfy the zero correlation condition (2).

  (ii) In the omitted variable case, we are for example interested in the population model :

  $$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma q_i + v_i \ , \ \text{where } E(v_i|x_{i2}, x_{i3}, q_i) = 0 \tag{7}$$

---

[2] A third typical situation is when dealing with simultaneous equations models. We will not cover this case here. See Wooldridge (2016), Chapter 16.

[3] As a matter of fact, $E(x_{i3}e_i) = E[(x^*_{i3} + e_i)e_i] = E(x^*_{i3}e_i) + E(e_i^2) = \sigma_e^2$.

[4] In the population model (3), it is natural to assume that the mismeasured variable $x_{i3}$ is redundant, i.e., that $E(y_i|x_{i2}, x^*_{i3}, x_{i3}) = \beta_1 + \beta_2 x_{i2} + \beta_3 x^*_{i3} \Leftrightarrow E(v_i|x_{i2}, x^*_{i3}, x_{i3}) = 0$, so that, by the law of iterated expectations, we also have $E(x_{i3}v_i) = 0$.

where $q_i$ is a zero mean[5] unobserved variable, so that it is omitted from the model actually estimated :

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i \ , \ \text{where } u_i = (v_i + \gamma q_i) \tag{8}$$

As a concrete example, let $y_i$ stand for the wage of individual $i$, $x_{i2}$ for his working experience, $x_{i3}$ for his level of education and $q_i$ for his unobserved ability. Under (7), by the law of iterated expectations, we have:

$$E(v_i) = 0, \quad E(x_{i2}v_i) = 0 \quad \text{and} \quad E(x_{i3}v_i) = 0$$

If, according to the usual story, we assume that $x_{i2}$ is uncorrelated with $q_i$, but we allow that $x_{i3}$ is correlated with $q_i$, then we have :

$$E(u_i) = 0 \quad \text{and} \quad Cov(x_{i2}, u_i) = E(x_{i2}u_i) = 0$$

but, because $x_{i3}$ is correlated with $q_i$, we however have:

$$Cov(x_{i3}, u_i) = E(x_{i3}u_i) = E\left[x_{i3}(v_i + \gamma q_i)\right] \neq 0$$

i.e., as in the errors-in-variables case, the estimated model (8) does not satisfy the zero correlation condition (2).

Note that in empirical applications, endogeneity problems related to omitted variables is by far a much more prevalent concern than endogeneity problems related to errors-in-variables. Typically, concerns come from the fact that we are interested in the partial effect of a variable whose values result from individual choices – e.g., education – and that this variable is likely to be correlated with unobserved characteristics – e.g., ability or motivation – of the individuals.

- In both examples given above, OLS estimation would not provide a consistent estimator of $\beta$ because the zero correlation condition (2) does not hold for one of the explanatory variable – $x_{i3}$ – of the estimated model. The basic idea of instrumental variables methods is to replace the endogenous variables responsible of the failure of the zero correlation condition (2) by – a at least equal number of – another variables, called instrumental variables or more simply instruments, which satisfy the zero correlation condition (2), and then proceed by deriving an estimator based on the same method of moments approach than outlined above for the OLS estimator. In the two examples considered above, this means finding at least one instrumental variable $z_i$ which must be such that :

$$Cov(z_i, u_i) = E(z_i u_i) = 0 \tag{9}$$

i.e., a variable $z_i$ which is not correlated with the error $u_i$ of the estimated model. Note that finding an instrumental variable $z_i$ which is not correlated with $u_i$ is not enough: $z_i$ must also be related to the endogenous variable that it replaces. For now, we concentrate on the usually most difficult to fulfill requirement that $z_i$ must be uncorrelated with $u_i$, and examine what it means in the two examples considered above:

[5] As the model contains an intercept, it can be assumed that $E(q_i) = 0$ without loss of generality.

(i) In the classical errors-in-variables case, a natural variable to instrument – i.e., to be an instrumental variable for – the endogenous variable $x_{i3}$ is an additional measurement of the $x^*_{i3}$. To be a valid instrument, this additional measurement $z_i$ should be such that:

(i.a) $E(y_i|x_{i2}, x^*_{i3}, z_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x^*_{i3} \;\Leftrightarrow\; E(v_i|x_{i2}, x^*_{i3}, z_i) = 0$

(i.b) $E(z_i|x^*_{i3}, e_i) = x^*_{i3} \Leftrightarrow z_i = x^*_{i3} + r_i$ , where $E(r_i|x^*_{i3}, e_i) = 0$

Condition (i.a) requires that the additional measurement $z_i$ is redundant in the population model (3). Condition (i.b) supposes that the additional measurement $z_i$ is an unbiased measurement of $x^*_{i3}$, and is unrelated to the measurement error $e_i$ of the mismeasured variable $x_{i3}$ included in the estimated model (5). Under condition (i.a) and (i.b), by the law of iterated expectations, we have:

$$E(z_i v_i) = 0 \quad \text{and} \quad E(r_i e_i) = 0$$

so that:

$$\begin{aligned} Cov(z_i, u_i) &= E(z_i u_i) = E[z_i(v_i - \beta_3 e_i)] = -\beta_3 E(z_i e_i) \\ &= -\beta_3 E[(x^*_{i3} + r_i)e_i] = -\beta_3 E(x^*_{i3} e_i) = 0 \end{aligned}$$

where the last equality follows from (6). In words, under condition (i.a) and (i.b), the zero correlation condition (9) holds. Condition (i.a) and (i.b) are not very controversial: they basically require to have two independent measures ($x_{i3}$ and $z_i$) of the same variable $x^*_{i3}$. Also, by nature, the requirement that the instrumental variable $z_i$ and the endogenous variable $x_{i3}$ must be related follows from their shared dependence on $x^*_{i3}$. Yet, in practice, having two measures of the same variable is not common[6].

(ii) In the omitted variable case, to be a valid instrument for the endogenous variable $x_{i3}$, the instrumental variable $z_i$ should be such that:

(ii.a) $E(y_i|x_{i2}, x_{i3}, q_i, z_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma q_i$
$\Leftrightarrow\; E(v_i|x_{i2}, x_{i3}, q_i, z_i) = 0$

(ii.b) $E(z_i|q_i) = E(z_i)$

Condition (ii.a) likewise requires that the instrumental variable $z_i$ is redundant in the population model (7), i.e., that when controlling for $(x_{i2}, x_{i3}, q_i)$, $y_i$ is unrelated to $z_i$. Condition (ii.b) supposes that the instrumental variable $z_i$ is unrelated to the omitted variable $q_i$. Under condition (ii.a) and (ii.b), by the law of iterated expectations, we have[7]:

$$E(z_i v_i) = 0 \quad \text{and} \quad E(z_i q_i) = 0$$

so that:

$$Cov(z_i, u_i) = E(z_i u_i) = E[z_i(v_i + \gamma q_i)] = 0$$

[6] For more discussion, see Wooldridge (2016), Section 15-4.

[7] $E(z_i q_i) = E[E(z_i q_i|q_i)] = E[q_i E(z_i|q_i)] = E(q_i)E(z_i) = 0 \cdot E(z_i) = 0.$

In words, under conditions (ii.a) and (ii.b), the zero correlation condition (9) holds. Unlike the errors-in-variables case where it is – at least conceptually – straightforward, finding an instrumental variable $z_i$ which fulfills both[8] the redundancy condition (ii.a) and the no correlation condition (ii.b), while being at the same time related to the endogenous variable $x_{i3}$, is usually a very serious challenge, which may easily give rise to endless debates[9].

But there are some notable exceptions. Suppose for example that interest lies in evaluating the effect of the participation $x_{i3}$ to some training program ($x_{i3} = 1$ if individual $i$ participates, $x_{i3} = 0$ otherwise) on the wage $y_i$ of some population of individuals (let $x_{i2}$ stand for observable individual characteristics ; in practice several control variables should be considered). One might fear that individuals choose to participate to the training program partly based on their ability or motivation $q_i$, which is unobserved and thus omitted, so that the participation $x_{i3}$ is endogenous. If the access to the training program is subject to a randomized eligibility[10] $z_i$ ($z_i = 1$ if individual $i$ is eligible, $z_i = 0$ otherwise), then the eligibility variable $z_i$ may serve as a valid and indisputable instrument for the participation $x_{i3}$ : because $z_i$ is randomized, it is by construction both redundant in (7) and uncorrelated with $q_i$, and because only eligible individuals can participate to the training program, $z_i$ is also correlated with $x_{i3}$.

## 1.1. The IV estimator

• Suppose that in the model of interest defined in assumption 2SLS.1 :

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i \\ \Leftrightarrow \quad y_i &= X_i \beta + u_i \end{aligned} \tag{10}$$

all variables are exogenous – i.e., are uncorrelated with the error of the model – except the last variable $x_{ik}$, which is supposed endogenous, for example because it is correlated with an unobserved omitted variable. Suppose further that we managed to find a relevant instrumental variable $z_i$ for the endogenous variable $x_{ik}$ – i.e., a variable $z_i$ which satisfies the same redundancy and no correlation conditions[11] as (ii.a) and (ii.b) –, and let $Z_i$ be defined as the row vector :

$$Z_i = \begin{bmatrix} 1 & x_{i2} & \cdots & x_{ik-1} & z_i \end{bmatrix} \tag{11}$$

The vector $Z_i$ is the same as $X_i$, except that the endogenous variable $x_{ik}$ is replaced by its instrument $z_i$. The vector $Z_i$ is often referred to as the vector of instrumental variables or vector of instruments, because besides $z_i$ which acts

[8] In applied works, it is not uncommon that researchers concentrate on discussing the no correlation condition (ii.b). It is worth emphasizing that the redundancy condition (ii.a) is equally important.

[9] For numerous examples and some discussions, including our initial example where $y_i$ = wage, $x_{i2}$ = experience, $x_{i3}$ = education and $q_i$ = ability, see Wooldridge (2016), throughout Chapter 16.

[10] i.e., those who are proposed to participate to the training program are drawn at random.

[11] Just let $x_{i2}$ stand for the vector of variables $(x_{i2}, ..., x_{ik-1})$, $x_{i3}$ stand for $x_{ik}$, and accordingly redefine the vector of parameters $\beta$.

as an instrument for $x_{ik}$, the others variables $(1, x_{i2}, ..., x_{ik-1})$ may likewise be viewed as instruments for themselves.

If the variables $(1, x_{i2}, ..., x_{ik-1})$ are as assumed indeed exogenous[12] and the instrumental variable $z_i$ is indeed valid, then the following assumption holds :

2SLS.4 Exogenous instrumental variables
The $1 \times l$ (row) vector of instrumental variables $Z_i$, where $l \geq k$, is such that :

$$E(Z_i' u_i) = 0 \ , \quad i = 1, ..., n$$

Note the condition $l \geq k$ in assumption 2SLS.4 which requires that the vector of instrumental variables contains at least as many variables $(= l)$ as there are parameters to be estimated $(= k)$ in model (10). In the present considered case, we have $l = k$.

- As outlined above for the OLS estimator, the method of moments approach to estimation suggests estimating $\beta$ based on the sample counterpart of the moment condition :

$$E(Z_i' u_i) = E[Z_i'(y_i - X_i\beta)] = 0 \tag{12}$$

i.e., choosing as an estimator $\hat{\beta}_{IV}$ of $\beta$ the solution of the sample moment condition :

$$\frac{1}{n}\sum_{i=1}^{n} Z_i'(y_i - X_i\hat{\beta}_{IV}) = 0$$

or equivalently :

$$\sum_{i=1}^{n} Z_i'(y_i - X_i\hat{\beta}_{IV}) = Z'(Y - X\hat{\beta}_{IV}) = 0$$

$$\Leftrightarrow \quad Z'X\hat{\beta}_{IV} = Z'Y$$

so that the so-called instrumental variables (IV) estimator $\hat{\beta}_{IV}$ is given by :

$$\begin{aligned}\hat{\beta}_{IV} &= (Z'X)^{-1}Z'Y \\ &= \left(\sum_{i=1}^{n} Z_i'X_i\right)^{-1}\sum_{i=1}^{n} Z_i'y_i\end{aligned}$$

Note that the IV estimator $\hat{\beta}_{IV}$ contains as a special case the OLS estimator (when $Z = X$). Note also that for the simple regression model $y_i = \beta_1 + \beta_2 x_i + u_i$ with $Z_i = [1 \ \ z_i]$ as vector of instruments, the IV estimators of $\beta_1$ and $\beta_2$ are given by[13] :

$$\hat{\beta}_{IV_1} = \bar{y} - \hat{\beta}_{IV_2}\bar{x} \quad \text{and} \quad \hat{\beta}_{IV_2} = \frac{Cov_{spl}(z_i, y_i)}{Cov_{spl}(z_i, x_i)} \tag{13}$$

where $Cov_{spl}(z_i, y_i)$ and $Cov_{spl}(z_i, x_i)$ denote the sample covariance between, respectively, $z_i$ and $y_i$, and $z_i$ and $x_i$.

- For the vector of parameters $\beta$ in model (10) – which is sometimes called the

[12] Note that this is trivially the case for the intercept variable.
[13] See Wooldridge (2016), Section 15-1.

structural model[14] – to be estimable from the data, or equivalently for the IV estimator $\hat{\beta}_{IV}$ to be well defined, it is not enough that assumption 2SLS.4 holds, i.e., that we have at our disposal a proper instrumental variable $z_i$ to replace the endogenous variable $x_{ik}$. As already outlined, we also need the instrumental variable $z_i$ to be related to the endogenous variable $x_{ik}$ that it replaces, as well as quite naturally that there is no perfect collinearity among the variables in $Z_i$. This is formally stated in the following assumption:

2SLS.3 No perfect collinearity and rank condition

(a) There is no exact linear relationship among the variables (including the constant) in the vector $Z_i$ of instrumental variables.

(b) The population moment $M_{zx} = E(Z_i'X_i)$ is full column rank, i.e., $\text{rank}(M_{zx}) = k$.

The no perfect collinearity assumption 2SLS.3a is self-explanatory. The rank condition assumption 2SLS.3b ensures that the moment condition (12) identify[15] the vector of parameters $\beta$, and that the IV estimator $\hat{\beta}_{IV}$ is well defined[16]. Note that the so-called order condition $l \geq k$ included in assumption 2SLS.4 is a necessary condition[17] for the rank condition 2SLS.3b to hold.

Maintaining assumption 2SLS.3a, it may be shown that, for the simple regression model $y_i = \beta_1 + \beta_2 x_i + u_i$ with $Z_i = [\,1 \;\; z_i\,]$ as vector of instruments, the rank condition 2SLS.3b holds if and only if we have:

$$Cov(z_i, x_i) \neq 0$$

or equivalently, if and only if the parameter $\pi_2$ in the reduced form[18] population model:

$$x_i = \pi_1 + \pi_2 z_i + \nu_i \tag{14}$$

is different form zero, i.e., if and only if the instrumental variable $z_i$ is correlated with the endogenous variable $x_i$ (in the population). For the multiple regression model (10) with the vector of instrumental variables (11), it may be likewise be shown[19] that the rank condition 2SLS.3b holds if and only if the parameter $\pi_k$ in the reduced form population model:

$$x_{ik} = \pi_1 + \pi_2 x_{i2} + ... + \pi_{k-1} x_{ik-1} + \pi_k z_i + \nu_i \tag{15}$$

is different form zero, i.e., if and only if the instrumental variable $z_i$ is partially correlated[20] with the endogenous variable $x_{ik}$ (in the population).

[14] This terminology comes from the simultaneous equations models framework.

[15] $\beta$ is identified if there is an unique solution to the linear system of equation $E[Z_i'(y_i - X_i\beta)]) = 0 \Leftrightarrow E(Z_i'X_i)\beta = E(Z_i'y_i)$.

[16] For $\hat{\beta}_{IV}$ to be well defined, we need the square matrix $Z'X = \sum_{i=1}^n Z_i'X_i$ to be full rank, so that it is invertible. As from law of large numbers we have $\frac{1}{n}\sum_{i=1}^n Z_i'X_i \xrightarrow{p} E(Z_i'X_i) = M_{zx}$, it will be the case, at least for $n$ sufficiently large (with probability equal to one as $n \to \infty$), if $M_{zx}$ is, as assumed, full rank.

[17] We can not have $\text{rank}(M_{zx}) = k$ if $M_{zx}$ has less rows than columns, i.e., if $l < k$.

[18] This terminology again comes from the simultaneous equations models framework, where a reduced form equation refers to an equation expressing an endogenous variable in terms of all the exogenous variables.

[19] See Wooldridge (2010), Section 5.1.

[20] i.e., still correlated after partialling out the other assumed exogenous variables $(1, x_{i2}, ..., x_{ik-1})$.

- To conclude this section, note that the assumption that $z_i$ is a valid instrument for the endogenous variable $x_{ik}$ in model (10), so that assumption 2SLS.4 holds, can not be tested. But the rank condition assumption 2SLS.3b that the instrumental variable $z_i$ is sufficiently related to – i.e., is partially correlated with – the endogenous variable $x_{ik}$ can, and should always, be tested. This may simply be done by estimating by OLS the reduced form population model (15), and testing through a (preferably heteroskedasticity robust) $t$-test (or $F$-test) the nullity of the parameter $\pi_k$.

## 1.2. The 2SLS estimator

- The so-called two stage least squares (2SLS) estimator generalizes the IV estimator for the case where multiple instruments are available.

- Suppose again that we are interested in the model defined in assumption 2SLS.1 :

$$y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i$$
$$\Leftrightarrow \quad y_i = X_i\beta + u_i \tag{16}$$

where all variables are assumed to be exogenous except the last variable $x_{ik}$, which is supposed endogenous because it is correlated with an unobserved omitted variable. But we now suppose that we managed to find more than one, say $p$, relevant instrumental variables for the endogenous variable $x_{ik}$, i.e., $p$ different variables $(z_{i1}, ..., z_{ip})$ which each likewise satisfies the same redundancy and no correlation conditions as (ii.a) and (ii.b). In this generalized case with multiple instruments, the vector $Z_i$ of instrumental variables – which includes the assumed exogenous explanatory variables $(1, x_{i2}, ..., x_{ik-1})$ – is given by the $1 \times l$ vector (where $l = k + p - 1$) :

$$Z_i = \begin{bmatrix} 1 & x_{i2} & \cdots & x_{ik-1} & z_{i1} & \cdots & z_{ip} \end{bmatrix} \tag{17}$$

If the variables $(1, x_{i2}, ..., x_{ik-1})$ are as assumed indeed exogenous and the instrumental variables $(z_{i1}, ..., z_{ip})$ are indeed valid, then the vector $Z_i$ satisfies assumption 2SLS.4., and we again have :

$$E(Z_i' u_i) = 0 \ , \quad i = 1, ..., n \tag{18}$$

Simply, we now have $l > k$. This situation is usually referred to by saying that the structural model (16) is overidentified[21] and that there is $l - k = p - 1$ overidentifying restrictions, which means that the moment condition (18) contains $p - 1$ more moments than needed for just identifying $\beta$ and estimating it using the IV estimator.

- With more exogenous variables ($= l$) in $Z_i$ than parameters to be estimated ($= k$) in model (16), we can no longer use the IV estimator as it stands. The basic idea underlying 2SLS estimation is to first choose $k$ linear combinations of

[21] When $l = k$, the structural model is said just identified, and when $l < k$, it is said underidentified. This terminology again comes from the simultaneous equations models framework.

the original vector of instrumental variables $Z_i$, say $W_i = Z_i\Pi$, where $\Pi$ is a $l \times k$ matrix of constants, such that $W_i$ is now a $1 \times k$ vector, and then use this new vector of instrumental variables $W_i$ to estimate $\beta$ with the usual IV estimator. This means estimating $\beta$ using the estimator:

$$\hat{\beta}_{IV_{(\Pi)}} = \left(\sum_{i=1}^{n} W_i'X_i\right)^{-1} \sum_{i=1}^{n} W_i'y_i = \left(\sum_{i=1}^{n} \Pi'Z_i'X_i\right)^{-1} \sum_{i=1}^{n} \Pi'Z_i'y_i$$

i.e., in matrix form:

$$\hat{\beta}_{IV_{(\Pi)}} = (W'X)^{-1} W'Y = (\Pi'Z'X)^{-1} \Pi'Z'Y$$

where obviously $W$ is a $n \times k$ matrix whose the $i$-th row is equal to $W_i$. This approach is justified because if $Z_i$ satisfies assumption 2SLS.4, we also have:

$$E(W'u_i) = \Pi'E(Z_i'u_i) = 0 \ , \quad i = 1, ..., n$$

i.e., if $Z_i$ is uncorrelated with $u_i$, then any linear combination $W_i = Z_i\Pi$ of $Z_i$ is also uncorrelated with $u_i$.

- There are as many different estimators $\hat{\beta}_{IV_{(\Pi)}}$ as there are choices for the matrix $\Pi$. It may however be shown[22] that, under assumptions[23] 2SLS.1 – 2SLS.4 and the additional homoskedasticity[24] assumption 2SLS.5:

    2SLS.5 Homoskedasticity
    The error $u_i$ is such that $E(u_i^2|Z_i) = \sigma^2, \quad i = 1, ..., n$.

  there exists an optimal choice for $\Pi$, which is given by :

$$\Pi^* = M_{zz}^{-1} M_{zx} = E(Z_i'Z_i)^{-1} E(Z_i'X_i)$$

  This choice is optimal in the sense that it yields the most efficient estimator of $\beta$, i.e., the estimator of $\beta$ with the smallest[25] asymptotic variance among all estimators of the form $\hat{\beta}_{IV_{(\Pi)}} = (\Pi'Z'X)^{-1} \Pi'Z'Y$.

- The optimal choice $\Pi^*$ entails the population moments $M_{zz} = E(Z_i'Z_i)^{-1}$ and $M_{zx} = E(Z_i'X_i)$, which are unknown, but which can be consistently estimated. By the law of large numbers, consistent estimators of $M_{zz}$ and $M_{zx}$ are respectively given by $\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i = Z'Z/n$ and $\frac{1}{n}\sum_{i=1}^{n} Z_i'X_i = Z'X/n$, so that a consistent estimator $\hat{\Pi}$ of $\Pi^*$ is given by:

$$\hat{\Pi} = (Z'Z)^{-1} Z'X$$

  and a consistent estimator $\hat{W}$ of the optimal matrix of instruments $W^* = Z\Pi^*$ is given by:

$$\hat{W} = Z\hat{\Pi} = Z(Z'Z)^{-1}Z'X = P_Z X$$

[22] See Wooldridge (2010), Section 5.2.3, Theorem 5.3.

[23] as well as the assumption that the rank condition 2SLS.3b holds for $W_i$.

[24] Note that if it is assumed that $E(u_i|Z_i) = 0$ (rather than merely $E(Z_i'u_i) = 0$), then $E(u_i^2|Z_i) = Var(u_i|Z_i)$.

[25] in a matrix sense, just as in the Gauss-Markov theorem.

- The 2SLS estimator $\hat{\beta}_{2SLS}$, which is also called the generalized instrumental variables estimator (GIVE), is defined as the IV estimator $\hat{\beta}_{IV_{(\Pi)}}$ with $\Pi = \hat{\Pi}$, and thus $\hat{W} = P_Z X$ :

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'P_Z X)^{-1} X' P_Z Y \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1} X'Z(Z'Z)^{-1}Z'Y \end{aligned} \quad (19)$$

where[26] $P_Z = Z(Z'Z)^{-1}Z'$, i.e., in detailed form:

$$\begin{aligned}\hat{\beta}_{2SLS} &= \left[\left(\sum_{i=1}^{n} X_i'Z_i\right)\left(\sum_{i=1}^{n} Z_i'Z_i\right)^{-1}\left(\sum_{i=1}^{n} Z_i'X_i\right)\right]^{-1} \\ &\quad \times \left(\sum_{i=1}^{n} X_i'Z_i\right)\left(\sum_{i=1}^{n} Z_i'Z_i\right)^{-1}\left(\sum_{i=1}^{n} Z_i'y_i\right)\end{aligned}$$

- To understand what the 2SLS estimator $\hat{\beta}_{2SLS}$ does, we need to examine the matrix of instruments $\hat{W} = Z\hat{\Pi} = Z(Z'Z)^{-1}Z'X = P_Z X$ on which it relies. $\hat{\Pi} = (Z'Z)^{-1}Z'X$ looks like an OLS estimator. As a matter of fact, each of the $k$ columns of $\hat{\Pi}$ is equal to $\hat{\pi}^j = (Z'Z)^{-1}Z'X^j$, where $X^j$ denotes the $j$-th column of $X$, i.e., a $n \times 1$ vector containing the $n$ observations of the variable $x_{ij}$ in (16). In other words, each $\hat{\pi}^j$ is nothing but the OLS estimator of the reduced form regression:

$$X^j = Z\pi^j + \nu \quad (20)$$

i.e., the regression of the variable $x_{ij}$ in (16) on all the variables contained in the original vector of instrumental variables $Z_i$. Likewise, each of the $k$ columns of $\hat{W}$ is equal to $\hat{W}^j = Z\hat{\pi}^j$. In other words, each $\hat{W}^j$ is nothing but a $n \times 1$ vector containing the OLS fitted values $\hat{X}^j$ from the regression (20) of $X^j$ on $Z$. In sum, the matrix of instruments $\hat{W}$ on which relies $\hat{\beta}_{2SLS}$ is thus simply equal to the OLS fitted values $\hat{X} = P_Z X$ obtained from the $k$ regressions of the columns of the explanatory $X$ on the original matrix of instrumental variables $Z$. In a nutshell, the 2SLS estimator $\hat{\beta}_{2SLS}$ uses as instruments the best linear predictor (in the least squares sense) of $X_i$ based on the available exogenous variables $Z_i$. For all exogenous explanatory variables $(1, x_{i2}, ..., x_{ik-1})$ in (16), this simply means taking the variable itself as its own instrument[27], and this choice is optimal[28]. For the endogenous variable $x_{ik}$, with $Z_i$ as defined in (17), this means using as instrument the OLS fitted value $\hat{x}_{ik}$ from the reduced form population model:

$$x_{ik} = \pi_1 + \pi_2 x_{i2} + ... + \pi_{k-1} x_{ik-1} + \pi_k z_{i1} + ... + \pi_{k+p-1} z_{ip} + \nu_i \quad (21)$$

and this choice, which will usually entail using some linear combination of all exogenous explanatory variables and available instrumental variables $(1, x_{i2}, ...,$

[26] Note that $P_Z$ is symmetric (i.e., $P_Z = P_Z'$).

[27] If $X^j$ is part of the original matrix of instrumental variables $Z$, then the OLS fitted values $\hat{X}^j$ from the regression of $X^j$ on $Z$ is simply the variable $X^j$ itself: $\hat{X}^j = X^j$.

[28] provided that the homoskedacticity assumption 2SLS.5 holds.

$x_{ik-1}, z_{i1}, ...z_{ip})$, is likewise optimal[29]. Note that this implies that using only one instrument when multiple instruments are available is usually not optimal. This is the all point of using multiple instruments.

- As suggested by the outlined above instrumental variable $\hat{x}_{ik}$ actually used by the 2SLS estimator $\hat{\beta}_{2SLS}$ to instrument the endogenous variable $x_{ik}$ in (16), the 2SLS estimator will only work if at least one of the available instrumental variables $(z_{i1}, ...z_{ip})$ is partially correlated with $x_{ik}$. This is exactly what is required for the rank condition assumption 2SLS.3b to hold in the present case. Maintaining assumption 2SLS.3a, it may indeed be shown[30] that, for the multiple regression model (16) with the vector of instrumental variables (17), the rank condition 2SLS.3b holds if and only if at least one of the $p$ parameters $(\pi_k, ..., \pi_{k+p-1})$ in the reduced form population model (21) is different from zero. As in the case of the usual IV estimator, this can, and should always, be formally tested. In practice, this may simply be done by estimating by OLS the reduced form population model (21), and testing through a (preferably heteroskedasticity robust) $F$-test the joint nullity of the parameters $(\pi_k, ..., \pi_{k+p-1})$.

- When there is more than one endogenous variable, the 2SLS estimation mechanics is essentially the same. If there are for example two endogenous variables in the structural model (16), say $x_{ik-1}$ and $x_{ik}$, then at least two relevant exogenous – i.e., uncorrelated with $u_i$ – instrumental variables $(z_{i1}, z_{i2})$ are needed. If $p \geq 2$ relevant exogenous instrumental variables $(z_{i1}, ..., z_{ip})$ are available, then the vector $Z_i$ of instrumental variables – which includes the assumed exogenous explanatory variables $(1, x_{i2}, ..., x_{ik-2})$ – is given by the $1 \times l$ vector (where $l = k + p - 2$):

$$Z_i = \begin{bmatrix} 1 & x_{i2} & \cdots & x_{ik-2} & z_{i1} & \cdots & z_{ip} \end{bmatrix} \tag{22}$$

and the 2SLS estimator will actually use as matrix of instruments $\hat{W} = \hat{X} = P_Z X$, for all exogenous explanatory variables $(1, x_{i2}, ..., x_{ik-2})$ in (16), the variables themselves (each of them is its own optimal instrument), and for the endogenous variables $x_{ik-1}$ and $x_{ik}$, the OLS fitted values $\hat{x}_{ik-1}$ and $\hat{x}_{ik}$ from the reduced form population models:

$$x_{ik-1} = \pi_1^a + \pi_2^a x_{i2} + ... + \pi_{k-2}^a x_{ik-2} + \pi_{k-1}^a z_{i1} + ... + \pi_{k+p-2}^a z_{ip} + \nu_i \tag{23}$$

and

$$x_{ik} = \pi_1^b + \pi_2^b x_{i2} + ... + \pi_{k-2}^b x_{ik-2} + \pi_{k-1}^b z_{i1} + ... + \pi_{k+p-2}^b z_{ip} + \nu_i \tag{24}$$

For the 2SLS estimator to work, it is necessary that the endogenous instrumental variables $(z_{i1}, ..., z_{ip})$ be partially correlated with the endogenous variables $x_{ik-1}$ and $x_{ik}$. This requires that at least some of the parameters $(\pi_{k-1}^a, ..., \pi_{k+p-2}^a)$ and $(\pi_{k-1}^b, ..., \pi_{k+p-2}^b)$ in respectively (23) and (24) be different from zero. This may, and always should, be formally tested through (preferably heteroskedasticity robust) $F$-tests of their joint nullity. Note however that if this is necessary, this

[29] provided again that the homoskedacticity assumption 2SLS.5 holds.
[30] See Wooldridge (2010), Section 5.2.

is actually not sufficient for the rank condition assumption 2SLS.3b to hold[31]. More than two endogenous variables is similarly handled.

- If the 2SLS estimator $\hat{\beta}_{2SLS}$ is used to estimate a just identified structural model, i.e., whenever the vector of instrumental variables $Z_i$ contains the same number of exogenous variables ($= l$) as there are parameters to be estimated ($= k$) – or equivalently the number of available instrumental variables is exactly equal to the number of endogenous variables in the structural model –, then it is simply equal to the IV estimator $\hat{\beta}_{IV}$. As a matter of fact, if $l = k$, then $Z'X$ is a square and invertible matrix, so that we have[32] :

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'P_ZX)^{-1}X'P_ZY = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= (Z'X)^{-1}Z'Z(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= (Z'X)^{-1}Z'Y
\end{aligned}$$

In other words, the 2SLS estimator $\hat{\beta}_{2SLS}$ contains as a special case (when $l = k$) the IV estimator $\hat{\beta}_{IV}$, and thus also the OLS estimator (when $Z = X$).

- To conclude this section, a last remark. Because $P_Z = Z(Z'Z)^{-1}Z'$ is a symmetric ($P_Z' = P_Z$) and idempotent ($P_ZP_Z = P_Z$) matrix, the 2SLS estimator $\hat{\beta}_{2SLS}$ may be rewritten as :

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (X'P_ZX)^{-1}X'P_ZY = ((P_ZX)'P_ZX)^{-1}(P_ZX)'Y \\
&= (\hat{X}'\hat{X})^{-1}\hat{X}Y
\end{aligned}$$

where $\hat{X} = P_ZX$. In other words, the 2SLS estimator $\hat{\beta}_{2SLS}$ may numerically be obtained as the OLS estimator of the regression[33] $Y = \hat{X}\beta +$ residuals, where $\hat{X}$ may itself be obtained, as outlined above, as the OLS fitted values from the $k$ regressions (20) of $X^j$ on $Z$. The name 'two stage least squares' comes from this procedure. In practice, there is however no need to perform manually any OLS regression to perform 2SLS estimation : all modern econometric software provide built-in 2SLS estimation routines which automatically do the job.

# 2. Asymptotic properties of the 2SLS estimator

- Hereafter, we derive the asymptotic properties of the 2SLS estimator $\hat{\beta}_{2SLS}$, i.e., its sampling properties as the sample size $n$ goes to infinity. Because the IV estimator $\hat{\beta}_{IV}$ is just a special case of the 2SLS estimator $\hat{\beta}_{2SLS}$, its properties just follow from those of $\hat{\beta}_{2SLS}$. Note that the 2SLS estimator $\hat{\beta}_{2SLS}$ has no exact in finite sample properties. For example, it is not unbiased.

[31] As a matter of fact, if for example there is only one of the instrumental variables $(z_{i1}, ..., z_{ip})$ which has a parameter different from zero in each reduced form population model, and that this is the parameter of the same variable, then the rank condition 2SLS.3 does not hold. For more details, see Wooldridge (2010), Section 5.2.

[32] As a reminder, if $A$ and $B$ are nonsingular matrices, then : $(AB)^{-1} = B^{-1}A^{-1}$.

[33] Be careful : the usual OLS standard errors and tests obtained from this regression are not valid.

## 2.1. Consistency

- We have the following property:

Property 20 Consistency of $\hat{\beta}_{2SLS}$
Under assumptions 2SLS.1 – 2SLS.4, the 2SLS estimator $\hat{\beta}_{2SLS}$ is a consistent estimator of $\beta$:

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta$$

A sketch of the proof is as follow. Under assumptions 2SLS.1 – 2SLS.3, we have:

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'P_Z X)^{-1}X'P_Z Y = (X'P_Z X)^{-1}X'P_Z(X\beta + u) \\ &= \beta + (X'P_Z X)^{-1}X'P_Z u \\ &= \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'u\end{aligned}$$

i.e., in detailed form:

$$\begin{aligned}\hat{\beta}_{2SLS} &= \beta + \left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'X_i\right)\right]^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^{n} X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i'u_i\right) \qquad (25)\end{aligned}$$

Under random sampling, $X_i'Z_i$, $Z_i'Z_i$, $Z_i'X_i$ and $Z_i'u_i$ are i.i.d. across $i$, so that $\frac{1}{n}\sum_{i=1}^{n} X_i'Z_i$, $\frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i$, $\frac{1}{n}\sum_{i=1}^{n} Z_i'X_i$ and $\frac{1}{n}\sum_{i=1}^{n} Z_i'u_i$ are sample average to which the law of large numbers[34] (LLN) can be applied. From the LLN, we have:

$$\frac{1}{n}\sum_{i=1}^{n} X_i'Z_i \xrightarrow{p} E(X_i'Z_i) = M_{xz}, \qquad \frac{1}{n}\sum_{i=1}^{n} Z_i'Z_i \xrightarrow{p} E(Z_i'Z_i) = M_{zz}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Z_i'X_i \xrightarrow{p} E(Z_i'X_i) = M_{zx} = M_{xz}'$$

Under assumption 2SLS.4, we have $E(Z_i'u_i) = 0$. From the LLN, we thus also have:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i'u_i \xrightarrow{p} 0$$

so that, from (25), we finally have:

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta + \left[M_{xz}M_{zz}^{-1}M_{zx}\right]^{-1} M_{xz}M_{zz}^{-1} \cdot 0 = \beta$$

[34] As a reminder, if $\{W_i\colon\, i = 1, ..., n\}$ are i.i.d. random variables with $E(W_i) = m$, then by the LLN we have: $\bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow{p} m$. This holds for $W_i$ being a scalar, a vector or a matrix.

## 2.2. Asymptotic normality of $\hat{\beta}_{2SLS}$ and inference

- We have the following property:

  Property 21 Asymptotic normality of $\hat{\beta}_{2SLS}$
  Under assumptions 2SLS.1 − 2SLS.5, the 2SLS estimator $\hat{\beta}_{2SLS}$ is asymptotically normally distributed:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, \sigma^2 A^{-1}) \tag{26}$$

  where:

$$A = E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)$$

  A sketch of the proof is as follows. Under assumptions 2SLS.1 − 2SLS.3, from (25), we have:

$$\begin{aligned} \hat{\beta}_{2SLS} &= \beta + \left[\left(\frac{1}{n}\sum_{i=1}^n X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_i'X_i\right)\right]^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_i'u_i\right) \end{aligned}$$

$$\begin{aligned} \Leftrightarrow \quad \sqrt{n}(\hat{\beta}_{2SLS} - \beta) &= \left[\left(\frac{1}{n}\sum_{i=1}^n X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_i'Z_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_i'X_i\right)\right]^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_i'Z_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_i'Z_i\right)^{-1}\left(n^{-\frac{1}{2}}\sum_{i=1}^n Z_i'u_i\right) \end{aligned}$$

  Under random sampling, $X_i'Z_i$, $Z_i'Z_i$, $Z_i'X_i$ and $Z_i'u_i$ are i.i.d. across $i$. As already outlined, from the LLN, we have $\frac{1}{n}\sum_{i=1}^n X_i'Z_i \xrightarrow{p} E(X_i'Z_i) = M_{xz}$, $\frac{1}{n}\sum_{i=1}^n Z_i'Z_i \xrightarrow{p} E(Z_i'Z_i) = M_{zz}$, $\frac{1}{n}\sum_{i=1}^n Z_i'X_i \xrightarrow{p} E(Z_i'X_i) = M_{zx} = M_{xz}'$, and we can write:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \overset{as}{=} A^{-1}\left(M_{xz}M_{zz}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n Z_i'u_i\right) \tag{27}$$

  where $A = E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)$ and $\overset{as}{=}$ means 'asymptotically equivalent', so that $\sqrt{n}(\hat{\beta}_{2SLS} - \beta)$ has asymptotically the same distribution as $A^{-1}\left(M_{xz}M_{zz}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n Z_i'u_i\right)$. As also already outlined, under assumption 2SLS.4, we have $E(Z_i'u_i) = 0$. From the central limit theorem[35] (CLT), we thus have:

$$n^{-\frac{1}{2}}\sum_{i=1}^n Z_i'u_i \xrightarrow{d} N(0, C), \quad \text{where } C = V(Z_i'u_i) = E(u_i^2 Z_i'Z_i)$$

---

[35] if $\{W_i\colon\ i = 1, ..., n\}$ are i.i.d. $(l \times 1)$ random vectors with $E(W_i) = m$ and $V(W_i) = \Sigma$, then by the CLT we have: $\sqrt{n}(\bar{W}_n - m) = n^{-\frac{1}{2}}\sum_{i=1}^n (W_i - m) \xrightarrow{d} N(0, \Sigma)$.

and thus[36] :

$$M_{xz}M_{zz}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}Z_i'u_i \xrightarrow{d} N(0,B)$$

where $B = E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(u_i^2Z_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)$, so that[37] :

$$A^{-1}\left(M_{xz}M_{zz}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}Z_i'u_i\right) \xrightarrow{d} N(0,A^{-1}BA^{-1})$$

and thus, from (27) :

$$\sqrt{n}(\hat{\beta}_{2SLS}-\beta) \xrightarrow{d} N(0,A^{-1}BA^{-1}) \tag{28}$$

To complete the proof, it remains to show that, under the homoskedasticity assumption 2SLS.5 $E(u_i^2|Z_i)=\sigma^2$, we have $B=\sigma^2A$. Under assumption 2SLS.5, by the law of iterated expectations, we have :

$$\begin{aligned} C &= V(Z_i'u_i) = E(u_i^2Z_i'Z_i) = E[E(u_i^2Z_i'Z_i|Z_i)] \\ &= E[E(u_i^2|Z_i)Z_i'Z_i] = E[\sigma^2Z_i'Z_i] = \sigma^2E(Z_i'Z_i) \end{aligned}$$

and thus :

$$\begin{aligned} B &= E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(u_i^2Z_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i) \\ &= \sigma^2E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i) \\ &= \sigma^2E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i) = \sigma^2A \end{aligned}$$

so that, from (28), we finally have :

$$\sqrt{n}(\hat{\beta}_{2SLS}-\beta) \xrightarrow{d} N(0,\sigma^2A^{-1})$$

- The limiting distributional result (26) is similar to the limiting distributional result given by Property 7 in SLN-I for the OLS estimator[38]. As for the OLS estimator, it provides an approximate finite sample distribution for the 2SLS estimator $\hat{\beta}_{2SLS}$ :

$$\begin{aligned} &\sqrt{n}(\hat{\beta}_{2SLS}-\beta) \approx N(0,\sigma^2A^{-1}) \\ \Leftrightarrow\ &\hat{\beta}_{2SLS} \approx N(\beta,\sigma^2A^{-1}/n) \end{aligned} \tag{29}$$

  which can be used – when $n$ is sufficiently large – for performing inference (confidence interval, hypothesis testing) without having to rely on any other assumption than assumptions 2SLS.1 – 2SLS.5.

- For inference based on the limiting distributional result (26), or equivalently on the approximate distributional result (29), we need an estimator of the asymp-

[36] because a linear function of jointly normally distributed random variables is itself normally distributed, and $M_{zz}^{-1}=E(Z_i'Z_i)^{-1}$ is a symmetric matrix.
[37] likewise because a linear function of jointly normally distributed random variables is itself normally distributed, and $A^{-1}=[E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)]^{-1}$ is a symmetric matrix.
[38] As a matter of fact, Property 7 in SLN-I is just a special case of Property 21, where $Z_i=X_i$.

totic variance $Avar(\hat{\beta}_{2SLS}) = \sigma^2 A^{-1}/n$. This requires consistent estimators of $\sigma^2$ and $A = E(X_i' Z_i)E(Z_i' Z_i)^{-1}E(Z_i' X_i)$. From the LLN, consistent estimators of $E(X_i' Z_i)$, $E(Z_i' Z_i)$ and $E(Z_i' X_i)$ are respectively given by $\frac{1}{n}\sum_{i=1}^n X_i' Z_i = X'Z/n$, $\frac{1}{n}\sum_{i=1}^n Z_i' Z_i = Z'Z/n$ and $\frac{1}{n}\sum_{i=1}^n Z_i' X_i = Z'X/n$, so that a consistent estimator of $A$ is given by $X'Z(Z'Z)^{-1}Z'X/n$. A consistent estimator of $\sigma^2$ is likewise given[39] by $\hat{s}^2 = \frac{1}{n-k}\sum_{i=1}^n \hat{u}_i^2$, where $\hat{u}_i = y_i - X_i\hat{\beta}_{2SLS}$, so that an estimator of $Avar(\hat{\beta}_{2SLS}) = \sigma^2 A^{-1}/n$ is given by:

$$\begin{aligned}\hat{V}_{2SLS}(\hat{\beta}_{2SLS}) &= \hat{s}^2\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} \\ &= \hat{s}^2(X'P_Z X)^{-1} \qquad (30)\end{aligned}$$

Note that $\hat{V}_{2SLS}(\hat{\beta}_{2SLS})$ contains as a special case (when $Z = X$) the estimator $\hat{V}(\hat{\beta}) = \hat{s}^2\left(X'X\right)^{-1}$ of the asymptotic variance of the OLS estimator. As usual, the diagonal elements $V\hat{a}r_{2SLS}(\hat{\beta}_{2SLS_j})$ of the $k \times k$ matrix estimator $\hat{V}_{2SLS}(\hat{\beta}_{2SLS})$ being the estimators of the variance $Avar(\hat{\beta}_{2SLS_j})$ of the estimator $\hat{\beta}_{2SLS_j}$ of different parameters $\beta_j$ $(j = 1, ..., k)$, natural estimators of the asymptotic standard error $As.e.(\hat{\beta}_{2SLS_j}) = \sqrt{Avar(\hat{\beta}_{2SLS_j})}$ of the estimator $\hat{\beta}_{2SLS_j}$ of the different parameters $\beta_j$, as well as a natural estimator of the asymptotic standard error $As.e.(R_0\hat{\beta}_{2SLS}) = \sqrt{Avar(R_0\hat{\beta}_{2SLS})} = \sqrt{R_0 Avar(\hat{\beta}_{2SLS})R_0'}$ of the estimator $R_0\hat{\beta}_{2SLS}$ of a single linear combination $R_0\beta$ of $\beta$, are likewise given by:

$$s.\hat{e}._{2SLS}(\hat{\beta}_{2SLS_j}) = \sqrt{V\hat{a}r_{2SLS}(\hat{\beta}_{2SLS_j})}, \quad j = 1, ..., k \qquad (31)$$

and

$$s.\hat{e}._{2SLS}(R_0\hat{\beta}_{2SLS}) = \sqrt{R_0\hat{V}_{2SLS}(\hat{\beta}_{2SLS})R_0'} \qquad (32)$$

where $R_0$ is a $1 \times k$ (row) vector of constants.

- As for the OLS estimator, the limiting distributional result (26), or equivalently the approximate distributional result (29), and the estimators $\hat{V}_{2SLS}(\hat{\beta}_{2SLS})$, $s.\hat{e}._{2SLS}(\hat{\beta}_{2SLS_j})$ and $s.\hat{e}._{2SLS}(R_0\hat{\beta}_{2SLS})$ given above in respectively (30), (31) and (32), provide all which is needed for performing inference after 2SLS estimation. Following exactly the same reasoning as in Section 4.3 and Section 4.4.2 of SLN-I, it may readily be checked that if in all the usual OLS inference procedures – confidence interval for $\beta_j$ or a single linear combination $R_0\beta$, two-sided and one-sided $t$-tests of $\beta_j$ or a single linear combination $R_0\beta$, $F$-test (or Wald test) of multiple linear restrictions – we replace the usual estimators $\hat{V}(\hat{\beta})$, $s.\hat{e}.(\hat{\beta}_j)$ and $s.\hat{e}.(R_0\hat{\beta})$ by their 2SLS versions $\hat{V}_{2SLS}(\hat{\beta}_{2SLS})$, $s.\hat{e}._{2SLS}(\hat{\beta}_{2SLS_j})$ and $s.\hat{e}._{2SLS}(R_0\hat{\beta}_{2SLS})$, then we obtain inference procedures that are asymptotically valid – i.e., approximately valid for $n$ sufficiently large – under assumptions 2SLS.1 – 2SLS.5.

[39] It is standard to use a degrees of freedom correction, i.e., to use $1/(n-k)$ rather than $1/n$. Asymptotically, it actually does not matter.

## 2.3. Heteroskedasticity robust inference

- The outlined above asymptotic properties of the 2SLS estimator require homoskedasticity. If the homoskedasticity assumption 2SLS.5 does not hold, then the outlined above usual-like inference procedures are no longer valid. However, as for the OLS estimator, it is again possible to derive inference procedures which are valid in the presence of heteroskedasticity of unknown form, i.e., which are robust to arbitrary form of heteroskedasticity.

- Heteroskedasticity robust inference procedures basically rely on the following property, which outlines the asymptotic properties of the 2SLS estimator $\hat{\beta}_{2SLS}$ under only assumptions 2SLS.1 – 2SLS.4 (Linearity in parameters, random sampling, no perfect collinearity and rank condition, and exogenous instrumental variables) :

  Property 22  Asymptotic properties of $\hat{\beta}_{2SLS}$ without homoskedasticity
  Under assumptions 2SLS.1 – 2SLS.4, the 2SLS estimator $\hat{\beta}_{2SLS}$ is a consistent estimator of $\beta$ :
  $$\hat{\beta}_{2SLS} \xrightarrow{p} \beta$$
  and is asymptotically normally distributed as :
  $$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, A^{-1}BA^{-1}) \tag{33}$$
  where :
  $$A = E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)$$
  and
  $$B = E(X_i'Z_i)E(Z_i'Z_i)^{-1}E(u_i^2 Z_i'Z_i)E(Z_i'Z_i)^{-1}E(Z_i'X_i)$$

  The fact that $\hat{\beta}_{2SLS}$ is consistent for $\beta$ under assumptions 2SLS.1 – 2SLS.4 was already outlined in Property 20. On the other hand, the limiting distribution result (33) has already be shown to hold likewise under assumptions 2SLS.1 – 2SLS.4 in the sketch of the proof of Property 21 : see the intermediary result (28).

- The limiting distribution result (33) is similar to the limiting distributional result given by Property 9 in the supplemental lecture notes III (hereafter SLN-III) for the OLS estimator[40]. As for the OLS estimator, it provides an approximate finite sample distribution for the 2SLS estimator $\hat{\beta}_{2SLS}$ :
  $$\begin{aligned} &\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, A^{-1}BA^{-1}) \\ \Leftrightarrow \quad &\hat{\beta}_{2SLS} \approx N(\beta, A^{-1}BA^{-1}/n) \end{aligned} \tag{34}$$
  which can be used – when $n$ is sufficiently large – for performing robust inference (confidence interval, hypothesis testing) without having to rely on any other assumption than assumptions 2SLS.1 – 2SLS.4, i.e., in particular without having

[40] As a matter of fact, Property 9 in SLN-III is just a special case of Property 22, where $Z_i = X_i$.

to rely on the homoskedasticity assumption 2SLS.5.

- For inference based on the limiting distributional result (33), or equivalently on the approximate distributional result (34), we need an estimator of the asymptotic variance $Avar(\hat{\beta}_{2SLS}) = A^{-1}BA^{-1}/n$. This requires consistent estimators of $A$ and $B$. We already outlined that $X'Z/n$, $Z'Z/n$ and $Z'X/n$ are consistent estimators of respectively $E(X_i'Z_i)$, $E(Z_i'Z_i)$ and $E(Z_i'X_i)$, so that a consistent estimator of $A$ is given by $X'Z(Z'Z)^{-1}Z'X/n = X'P_ZX/n$. For $B$, we additionally need a consistent estimator of $E(u_i^2Z_i'Z_i)$, which is simply given[41] by $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 Z_i'Z_i$, where obviously $\hat{u}_i = y_i - X_i\hat{\beta}_{2SLS}$. A consistent estimator of $B$ is then given by $X'Z(Z'Z)^{-1}(\sum_{i=1}^n \hat{u}_i^2 Z_i'Z_i)(Z'Z)^{-1}Z'X/n$, so that an estimator of $Avar(\hat{\beta}_{2SLS}) = A^{-1}BA^{-1}/n$ is given by:
$$\begin{aligned}\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS}) &= (X'P_ZX)^{-1}X'Z(Z'Z)^{-1}\left(\sum_{i=1}^n \hat{u}_i^2 Z_i'Z_i\right) \\ &\quad\times(Z'Z)^{-1}Z'X(X'P_ZX)^{-1} \qquad (35)\end{aligned}$$
where the subscript '$HC$' simply stands for 'Heteroskedasticity Consistent'. Note that $\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS})$ again contains as a special case (when $Z = X$) the heteroskedasticity robust estimator $\hat{V}_{HC}(\hat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^n \hat{u}_i^2 X_i'X_i\right)(X'X)^{-1}$ of the asymptotic variance of the OLS estimator. As usual, the diagonal elements $V\hat{a}r_{2SLS_{HC}}(\hat{\beta}_{2SLS_j})$ of the $k \times k$ matrix estimator $\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS})$ being the estimators of the variance $Avar(\hat{\beta}_{2SLS_j})$ of the estimator $\hat{\beta}_{2SLS_j}$ of the different parameters $\beta_j$ $(j = 1, ..., k)$, natural estimators of the asymptotic standard error $As.e.(\hat{\beta}_{2SLS_j}) = \sqrt{Avar(\hat{\beta}_{2SLS_j})}$ of the estimator $\hat{\beta}_{2SLS_j}$ of different parameters $\beta_j$, as well as a natural estimator of the asymptotic standard error $As.e.(R_0\hat{\beta}_{2SLS}) = \sqrt{Avar(R_0\hat{\beta}_{2SLS})} = \sqrt{R_0 Avar(\hat{\beta}_{2SLS})R_0'}$ of the estimator $R_0\hat{\beta}_{2SLS}$ of a single linear combination $R_0\beta$ of $\beta$, are likewise given by:
$$s.\hat{e}._{2SLS_{HC}}(\hat{\beta}_{2SLS_j}) = \sqrt{V\hat{a}r_{2SLS_{HC}}(\hat{\beta}_{2SLS_j})}, \quad j = 1, ..., k \qquad (36)$$
and
$$s.\hat{e}._{2SLS_{HC}}(R_0\hat{\beta}_{2SLS}) = \sqrt{R_0\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS})R_0'} \qquad (37)$$
where $R_0$ is a $1 \times k$ (row) vector of constants.

- As for the OLS estimator, the limiting distributional result (33), or equivalently the approximate distributional result (34), and the estimators $\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS})$, $s.\hat{e}._{2SLS_{HC}}(\hat{\beta}_{2SLS_j})$ and $s.\hat{e}._{2SLS_{HC}}(R_0\hat{\beta}_{2SLS})$ given above in respectively (35), (36) and (37), provide all which is needed for performing robust inference after 2SLS estimation. Following exactly the same reasoning as in Section 4.3 and Section 4.4.2 of SLN-I, it may again readily be checked that if in all the usual OLS inference procedures – confidence interval for $\beta_j$ or a single linear combi-

[41] From the LLN, $\frac{1}{n}\sum_{i=1}^n u_i^2 Z_i'Z_i \xrightarrow{p} E(u_i^2 Z_i'Z_i)$. As $\hat{u}_i$ converges to $u_i$ (because $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$), we also have $\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 Z_i'Z_i \xrightarrow{p} E(u_i^2 Z_i'Z_i)$.

nation $R_0\beta$, two-sided and one-sided $t$-tests of $\beta_j$ or a single linear combination $R_0\beta$, $F$-test (or Wald test) of multiple linear restrictions – we replace the usual estimators $\hat{V}(\hat{\beta})$, $s.\hat{e}.(\hat{\beta}_j)$ and $s.\hat{e}.(R_0\hat{\beta})$ by their 2SLS heteroskedasticity robust versions $\hat{V}_{2SLS_{HC}}(\hat{\beta}_{2SLS})$, $s.\hat{e}._{2SLS_{HC}}(\hat{\beta}_{2SLS_j})$ and $s.\hat{e}._{2SLS_{HC}}(R_0\hat{\beta}_{2SLS})$, then we obtain inference procedures that are asymptotically valid – i.e., approximately valid for $n$ sufficiently large – under only assumptions 2SLS.1 – 2SLS.4, i.e., without having to rely on the homoskedasticity assumption 2SLS.5.

- Modern econometric software usually provide options to compute heteroskedasticity robust standard errors and to perform heteroskedasticity robust tests after 2SLS estimation. In applied works, heteroskedasticity robust standard errors and tests should systematically be considered, at least for comparison.

## 2.4. Remarks

- In practice, 2SLS estimation tends to yield estimates with large standard errors, typically much larger than standard errors from OLS estimation. This is all the more so true that the instrumental variables used are weak, i.e. that the instrumental variables used are only weakly partially correlated with the endogenous explanatory variable(s) of the estimated model. See Wooldridge (2016), Sections 15-1a and 15-3b for details[42].

- Weak instruments do not only give rise to rather imprecise estimates. When the instruments are weak:

  - even small violations of the validity of the instruments, i.e., small violations of the exogenous instrumental variables assumption 2SLS.4, may cause the 2SLS estimator to have a large asymptotic bias[43].

  - maintaining that the instruments are valid (so that there is no asymptotic bias), the 2SLS estimator may be severely biased in finite sample, and its asymptotic normal distribution may provide a very poor approximation of its true finite sample distribution, leading to unreliable inference (confidence interval, hypothesis testing), even with (very) large sample size[44].

  For details, as well as some guidelines about how to detect weak instruments, see Wooldridge (2016), Sections 15-1b and 15-3c[45].

- Assuming that the available instrumental variables are valid, it is possible to formally test through a simple variable addition test whether an explanatory variable is endogenous, i.e., to test if 2SLS estimation is actually necessary. In a nutshell, this may simply be done by augmenting the considered structural model (16) with the OLS estimated residuals $\hat{\nu}_i$ from the reduced form population model (21) of explanatory variable $x_{it}$ whose the endogeneity is tested, and

[42] For more details, see Wooldridge (2010), Section 5.2.6.
[43] i.e., a bias which does not vanish as $n \to \infty$.
[44] Note that using too many instruments may have the same effect.
[45] For more details, see again Wooldridge (2010), Section 5.2.6.

after OLS estimation, testing through a (preferably heteroskedasticity robust) $t$-test the nullity of the coefficient of the added OLS estimated residuals variable $\hat{\nu}_i$. The mechanics is the same for jointly testing the endogeneity of multiple explanatory variables[46]. Note that this variable addition test is equivalent to a so-called Hausman test based on the direct comparison of the 2SLS and OLS estimates of the considered structural model. See Wooldridge (2016), Section 15-5a, for details[47].

- When the considered structural model is just identified, i.e., when the vector of instrumental variables $Z_i$ contains the same number of exogenous variables $(= l)$ as there are parameters to be estimated $(= k)$[48] , there is no room for testing the validity the available instrumental variables, i.e. for testing the exogenous instrumental variables assumption 2SLS.4. However, when the model is overidentified, i.e., when $l > k$, then there is $l - k$ overidentifying restrictions, which means that the moment condition $E(Z_i'u_i) = 0$ contains $l - k$ more moments than actually needed for just identifying $\beta$, and it is now possible to globally test the exogenous instrumental variables assumption 2SLS.4. This test is known as an overidentifying restrictions test or Sargan test. In a nutshell, the test statistic $S$ is equal to $n$ times the $R$-squared $(= R^2_{\hat{u}})$ of the auxiliary regression of the 2SLS residuals $\hat{u}_i$ on all the variables contained in the vector of instrumental variables $Z_i$, and the decision rule of the test is: reject the null hypothesis $H_0$ that $E(Z_i'u_i) = 0$ if $S = nR^2_{\hat{u}} > \chi^2_{l-k;1-\alpha}$ and do not reject otherwise, where $\chi^2_{l-k;1-\alpha}$ is the quantile of order $1 - \alpha$ of the $\chi^2(l - k)$ distribution[49]. Note that this regression-based test is equivalent to a so-called Hausman test based on the direct comparison of the 2SLS estimate of the model obtained using all available instrumental variables and the 2SLS estimate of the model obtained using only a number of the instrumental variables equal to the number of the number of endogenous variables in the structural model (the test is invariant to the chosen subset of instrumental variables). Note also that this regression-based test is only valid if the homoskedasticity assumption 2SLS.5 holds. See Wooldridge (2016), Section 15-5b, for details[50].

# Reference

Wooldridge J.M. (2010), *Econometric Analysis of Cross-Section and Panel Data*, Second Edition, MIT Press.

---

[46] It suffices to augment the considered structural model with the OLS estimated residuals from the reduced form population model of each of the explanatory variables whose the endogeneity is tested, and after OLS estimation, testing through a (preferably heteroskedasticity robust) $F$-test the joint nullity of the coefficients of the added OLS estimated residuals variables.

[47] For more details, in particular regarding how to test the endogeneity of some variables, while allowing other variables to be endogenous, see Wooldridge (2010), Section 6.3.1.

[48] or equivalently the number of available instrumental variables is exactly equal to the number of endogenous variables in the structural model.

[49] The $p$-value of the test, for a value $S^*$ of the test statistic obtained in a particular sample, is given by: $p_S = \mathbb{P}(v > S^*)$, where $v \sim \chi^2(l - k)$.

[50] For more details, in particular regarding how to compute an heteroskedasticity robust version of the test, see Wooldridge (2010), Section 6.3.2.

Wooldridge J.M. (2016), *Introductory Econometrics : A Modern Approach*, 6th Edition, Cengage Learning.