Two-Stage Least Squares (2SLS)

Consider a standard linear bivariate regression model \( y_i = \beta_0 + x_i \beta_1 + \epsilon_i \). In this case, begin with the assumption that \( E[\epsilon_i | x_i] \neq 0 \) which means that OLS estimates of \( \hat{\beta}_1 \) will produce biased and inconsistent estimates. Suppose however that there is a variable \( z_i \) that directly impacts \( x \) but it is uncorrelated with \( \epsilon_i \). This can be expressed as \( E[z_i | \epsilon_i] = 0 \). The variable \( z \) is known as an ‘instrument’ in the model. The implicit assumption is that \( z \) impacts \( x \) and \( x \) impacts \( y \). Therefore, if we were to somehow “shock” \( x \) with \( z \), we can infer the relationship between \( x \) and \( y \).

This is done by considering the relationship between \( x \) and \( z \) as a simple linear model. This “first-stage” relationship between \( x \) and \( z \) is given by the equation

\[
x_i = \theta_0 + z_i \theta_1 + u_i
\]

Where we assume \( z \) and \( u \) are uncorrelated. Note that \( x \) has two components. One is a deterministic one that is a function of \( z \), \( x_i^p = \theta_0 + z_i \theta_1 \). Because \( z \) is uncorrelated with \( \epsilon \) then this predictive component of \( x \) is also uncorrelated with \( \epsilon \). Since \( x \) is correlated with \( \epsilon \) and \( x_i^p \) is not, it must be that what is driving the correlation between \( x \) and \( \epsilon \) is then \( u \).

The suggestion then is instead of using \( x \) in the regression \( y_i = \beta_0 + x_i \beta_1 + \epsilon_i \)-- which we know would produce biased estimates -- why not use \( x_i^p \) instead. The \( x_i^p \) component of \( x \) contains information about \( x \), but it is NOT correlated with \( \epsilon \) so it should produce estimates with “nice” properties. That is a great idea except that we do not know \( \theta_0 \) or \( \theta_1 \) and therefore, we cannot use \( x_i^p \) in a regression. We can however use the next best thing: an unbiased estimate of \( x_i^p \) which is produced from a first-stage regression of \( x \) on \( z \).

This “two-stage least squares” procedure is very straightforward. Run a regression of the form

\[
x_i = \theta_0 + z_i \theta_1 + u_i
\]

It should be no surprise that the OLS estimates for the parameters \( \hat{\theta}_0 \) and \( \hat{\theta}_1 \) are

\[
\hat{\theta}_0 = \bar{x} - \bar{z} \hat{\theta}_1 \quad \text{and} \quad \hat{\theta}_1 = \frac{\sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}.
\]

With these estimates we can construct a predicted value for

\[
\hat{x}_i = \hat{\theta}_0 + z_i \hat{\theta}_1.
\]

Instead of using \( x \) in the estimates for \( \beta_1 \) -- which produces \( \hat{\beta}_{\text{ols}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \)
use the predicted value of $x$, $\hat{x}_i = \hat{\theta}_0 + z_i \hat{\theta}_1$, instead. The predicted value is a linear function of $z$ which is uncorrelated with $\varepsilon$ so the 2SLS estimate for $\beta_1$ should have “nice” properties. The 2SLS estimate for $\beta_1$ is therefore

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2}.$$

**Working with the 2SLS estimate**

Given that $\hat{x}_i = \hat{\theta}_0 + z_i \hat{\theta}_1$ it is clear that $\bar{x} = \hat{\theta}_0 + \bar{z} \hat{\theta}_1$ and hence $\hat{x}_i - \bar{x} = (z_i - \bar{z}) \hat{\theta}_1$. Substitute this into the equation for $\hat{\beta}_1^{2sls}$ and we obtain

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \hat{\theta}_1 (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n} \hat{\theta}_1^2 (z_i - \bar{z})^2} = \frac{\hat{\theta}_1 \sum_{i=1}^{n} (z_i - \bar{z})(y_i - \bar{y})}{\hat{\theta}_1 \sum_{i=1}^{n} (z_i - \bar{z})^2}.$$

Substitute the definition of $\hat{\theta}_1$ into this equation and we obtain

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^{n} (z_i - \bar{z})(y_i - \bar{y})}{\hat{\theta}_1 \sum_{i=1}^{n} (z_i - \bar{z})^2} = \frac{\sum_{i=1}^{n} \frac{(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n} (z_i - \bar{z})^2} \sum_{i=1}^{n} (z_i - \bar{z})^2}{\sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})}.$$

Divide the numerator and denominator by (n-1) and we note the 2SLS estimate is therefore the ratio of the covariate between $y$ and $z$ divided by the covariance between $x$ and $z$.

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^{n} (z_i - \bar{z})(y_i - \bar{y})/(n-1)}{\sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})/(n-1)} = \frac{\hat{\sigma}_{yz}}{\hat{\sigma}_{zx}}.$$

Is $\hat{\beta}_1^{2sls}$ an unbiased estimate?
Anytime we have a new random variable, we always ask what is the expected value and the variance? This is a difficult question. In the end, we are unable to show that \( \hat{\beta}_{1}^{2SLS} \) is an unbiased estimate in this context? Why? In order to demonstrate the properties of the estimate, we always substitute the truth back into the model. Start with the estimate for \( \hat{\beta}_{1}^{2SLS} \) and note that we can drop one of the means in both the numerator and denominator.

\[
\hat{\beta}_{1}^{2SLS} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})(y_{i} - \overline{y})}{\sum_{i=1}^{n} (z_{i} - \overline{z})(x_{i} - \overline{x})} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})y_{i}}{\sum_{i=1}^{n} (z_{i} - \overline{z})x_{i}}
\]

At this point, we usually substitute in the definition of \( y \) into the numerator. However, \( x \) is also a random variable that is a function of the error \( \varepsilon \) so we would need to substitute that definition in as well.

\[
\hat{\beta}_{1}^{2SLS} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})y_{i}}{\sum_{i=1}^{n} (z_{i} - \overline{z})x_{i}} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})(\beta_{0} + x_{i}\beta_{1} + \varepsilon_{i})}{\sum_{i=1}^{n} (z_{i} - \overline{z})(\theta_{0} + z_{i}\theta_{1} + \mu_{i})}
\]

The problem is that we now have random variables in the numerator and the denominator. If \( a \) and \( b \) are random variables \( E[a/b] \neq E[a] / E[b] \). Without knowing the joint distribution of the numerator and denominator, we cannot figure out the expected value of the estimate. Our only option is then to consider whether the estimate is consistent.

**The consistency of 2SLS estimates**

To show whether the 2SLS estimate is consistent, start with the definition \( \hat{\beta}_{1}^{2SLS} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})(y_{i} - \overline{y})}{\sum_{i=1}^{n} (x_{i} - \overline{x})(z_{i} - \overline{z})} \)

and substitute in the definition of \( y \). This produces \( \hat{\beta}_{1}^{2SLS} = \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})(\beta_{0} + x_{i}\beta_{1} + \varepsilon_{i})}{\sum_{i=1}^{n} (x_{i} - \overline{x})(z_{i} - \overline{z})} \). Expand the numerator, and we get

\[
\hat{\beta}_{1}^{2SLS} = \frac{\beta_{0} \sum_{i=1}^{n} (z_{i} - \overline{z})}{\sum_{i=1}^{n} (x_{i} - \overline{x})(z_{i} - \overline{z})} + \frac{\beta_{1} \sum_{i=1}^{n} (z_{i} - \overline{z})x_{i}}{\sum_{i=1}^{n} (x_{i} - \overline{x})(z_{i} - \overline{z})} + \frac{\sum_{i=1}^{n} (z_{i} - \overline{z})\varepsilon_{i}}{\sum_{i=1}^{n} (x_{i} - \overline{x})(z_{i} - \overline{z})}.
\]
term equals zero because sums of deviation from means equal zero. The second term equals \( \beta_1 \) and the whole term reduces to \( \hat{\beta}_1^{2SLS} = \beta_1 + \frac{\sum_{i=1}^{n} (z_i - \bar{z}) \varepsilon_i}{\sum_{i=1}^{n} (x_i - \bar{x})(z_i - \bar{z})} \). Divide the numerator and denominator of the right hand term by \( (n-1) \) and you get:

\[
\hat{\beta}_1^{2SLS} = \beta_1 + \frac{\sum_{i=1}^{n} (z_i - \bar{z}) \varepsilon_i / (n-1)}{\sum_{i=1}^{n} (x_i - \bar{x})(z_i - \bar{z}) / (n-1)} = \beta_1 + \frac{\hat{\sigma}_{z\varepsilon}}{\hat{\sigma}_{zx}}.
\]

Now we ask what happens to this estimate as the sample sizes increases to infinity. Note that if \( a \) and \( b \) are estimates, then

\[
\lim(a/b) = \lim(a)/\lim(b).
\]

Note also that the \( \lim(\hat{\sigma}_{z\varepsilon}) = \sigma_{z\varepsilon} \) and \( \lim(\hat{\sigma}_{zx}) = \sigma_{zx} \), and therefore, it is no surprise that \( \lim(\hat{\beta}_1^{2SLS}) = \beta_1 + \frac{\sigma_{z\varepsilon}}{\sigma_{zx}} \). In this case, the estimate \( \hat{\beta}_1^{2SLS} \) will only be consistent if

\[
\lim(\hat{\sigma}_{z\varepsilon}) = \sigma_{z\varepsilon} = 0.
\]

To produce consistent estimates, it must be the case that \( z \) only produces a change in \( y \) through an initial change in \( z \). If \( z \) has a direct effect on \( y \) then \( \lim(\hat{\sigma}_{z\varepsilon}) = \sigma_{z\varepsilon} \neq 0 \) and the model will produce inconsistent estimates.

### An alternative interpretation of 2SLS estimates

The structural equation of interest is \( y_i = \beta_0 + x_i \beta_1 + \varepsilon_i \). We anticipate that \( x \) is correlated with \( \varepsilon \) so we cannot estimate this model by OLS. We do however have a variable \( z \) that we believe is predictive of \( x \) but it is not directly correlated with \( \varepsilon \). The first-stage equation that relates \( x \) to \( z \) is given by the equation

\[
x_i = \theta_0 + z_i \theta_1 + u_i.
\]

Substitute this equation into the equation for \( y \) and one obtains the equation

\[
y_i = \beta_0 + (\theta_0 + z_i \theta_1 + u_i) \beta_1 + \varepsilon_i = (\beta_0 + \theta_0 \beta_1) + z_i \theta_1 \beta_1 + u_i \beta_1 + \varepsilon_i.
\]

In this model, define \( (\beta_0 + \theta_0 \beta_1) = \pi_0 \) let \( \theta_1 \beta_1 = \pi_1 \) and let \( u_i \beta_1 + \varepsilon_i = \nu_i \). This allows us to write the equation for \( y \) as \( y_i = \pi_0 + z_i \pi_1 + \nu_i \). This equation is referred to as the “reduced-form” and it represents the correlation between \( z \) and \( y \). The coefficient is \( \frac{\partial y}{\partial z} = \pi_1 \). Note that the coefficient has an interesting interpretation. The model we are suggesting is that \( y \) is a function of \( x \) and \( x \) is a function of \( z \). Therefore \( y = f(x(z)) \). If we take the derivative of \( y \) with respect to \( z \), because we assume \( z \) only changes \( y \) through a
change in x, we obtain \( \frac{\partial y}{\partial z} = \frac{\partial y}{\partial x} \frac{\partial x}{\partial z} \). Note that \( \pi_1 = \beta_1 \theta_1 \), which is exactly \( \frac{\partial y}{\partial x} \frac{\partial x}{\partial z} \). Therefore, if we take the reduced form estimate for \( \pi_1 \), which equals \( \frac{\partial y}{\partial x} \frac{\partial x}{\partial z} = \beta_1 \theta_1 \) and divide it by the 1st stage coefficient \( \theta_1 \), we get \( \beta_1 \). It is also the case that

\[
\hat{\beta}_1^{2sls} = \frac{\hat{\pi}_1}{\hat{\theta}_1}
\]

which is exactly the same estimate we got above.

To recap, we call the initial equation the “structural equation of interest”:

\[
y_i = \beta_0 + x_i \beta_1 + \varepsilon_i
\]

To estimate this via 2SLS or indirect least squares, we need a “first-stage” relationship

\[
x_i = \theta_0 + z_i \theta_1 + u_i
\]

The direct relationship between the instrument and the outcome of interest is called the “reduced-form”.

\[
y_i = \pi_0 + z_i \pi_1 + v_i
\]

The variance of \( \hat{\beta}_1^{2sls} \)

Note that if we were to estimate the regression \( y_i = \beta_0 + x_i \beta_1 + \varepsilon_i \) by OLS, the variance for \( \hat{\beta}_1^{ols} \) would be

\[
Var(\hat{\beta}_1^{ols}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.
\]

The variance is proportional to how much variance in x is used to produce the estimate. In the case of 2SLS, we are not use x but instead, \( \hat{x}_i \). Therefore \( Var(\hat{\beta}_1^{2sls}) = \frac{\sigma^2}{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2} \).

Recall that in a regression, SST=SSM+SSE and in this case

\[
\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2 + \sum_{i=1}^n \hat{v}_i^2
\]

As a result, by construction, \( \sum_{i=1}^n (x_i - \bar{x})^2 > \sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2 \) which means also that by construction

\[
Var(\hat{\beta}_1^{ols}) < Var(\hat{\beta}_1^{2sls})
\]

The benefit of 2SLS is that we are only using a part of the variation in x – that part
that is uncorrelated with $\varepsilon$ -- so our estimates are consistent. However, because we are using smaller variation in $x$ than in the OLS case, the cost of 2SLS is a reduction in precision -- the value $\text{Var}(\hat{\beta}_1^{2sls})$ increases considerably.

What will produce smaller values for $\text{Var}(\hat{\beta}_1^{2sls})$? Start with the definition $\text{Var}(\hat{\beta}_1^{2sls}) = \frac{\sigma^2_{\varepsilon}}{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})^2}$ and remember that $\hat{x}_i - \bar{x} = (z_i - \bar{z})\hat{\theta}_i$. Substitute this in and we get $\text{Var}(\hat{\beta}_1^{2sls}) = \frac{\sigma^2_{\varepsilon}}{\hat{\theta}_i^2 \sum_{i=1}^{n}(z_i - \bar{z})^2}$. Note that

$$\hat{\theta}_i = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

and substitute this into the variance equation and you get

$$\text{Var}(\hat{\beta}_1^{2sls}) = \frac{\sigma^2_{\varepsilon} \sum_{i=1}^{n}(z_i - \bar{z})^2}{(\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x}))^2} \cdot \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

The denominator in this term $(\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x}))$ is nothing more that $(n-1)\hat{\sigma}_{xz}$ while the numerator can be thought of as $\sum_{i=1}^{n}(z_i - \bar{z})^2 = (n-1)\hat{\sigma}_{z}^2$. Therefore,

$$\text{Var}(\hat{\beta}_1^{2sls}) = \frac{\sigma^2_{\varepsilon} \sum_{i=1}^{n}(z_i - \bar{z})^2}{(\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x}))^2} = \frac{\sigma^2_{\varepsilon} (n-1)\hat{\sigma}_{z}^2}{(n-1)\hat{\sigma}_{xz}^2} = \frac{\sigma^2_{\varepsilon} \hat{\sigma}_{z}^2}{(n-1)\hat{\sigma}_{xz}^2}$$

and we will reduce the $\text{Var}(\hat{\beta}_1^{2sls})$ is if we can find a variable $z$ that is strongly correlated with $x$. At the other end of the spectrum, suppose $\hat{\sigma}_{xz}$ approaches zero -- this means $z$ does not explain much of $x$. What is the cost? In this case, if $z$ does not explain much of $x$, then we cannot learn $x$’s impact on $y$ and the $\text{Var}(\hat{\beta}_1^{2sls})$ will explode.