1 GMM

When we have the no. of instruments greater than the no. of regressor $L > K$, we have an overidentified model and MM cannot work. We have to use Generalised method of moments (GMM). In the GMM approach, we use this weighting matrix $\mathbf{W}_n$ and rewrite the moment conditions in quadratic form. Thus, we derive the GMM estimator by

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}}\{J_n(\hat{\beta}_0)\}, \tag{1}$$

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}}\{n\bar{\mathbf{g}}(\hat{\beta}_0)'\mathbf{W}_n\bar{\mathbf{g}}(\hat{\beta}_0)\}, \tag{2}$$

Note here $\mathbf{W}_n$ is a $L \times L$ symmetric and positive definite matrix. This matrix implicitly determines sample moments are important. For example, let's consider a very simple 2 moment condition case where

$$\bar{\mathbf{g}}(\hat{\beta}_0) = \begin{bmatrix} g_a \\ g_b \end{bmatrix}, \mathbf{W}_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

This implies $\bar{\mathbf{g}}(\hat{\beta}_0)'\mathbf{W}_n\bar{\mathbf{g}}(\hat{\beta}_0)$ to be

$$g_a^2 + g_b^2$$

which suggest both sample moment condition are equally important. However, if

$$\mathbf{W}_n = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Then $\bar{\mathbf{g}}(\hat{\beta}_0)'\mathbf{W}_n\bar{\mathbf{g}}(\hat{\beta}_0)$ becomes

$$2g_a^2 + g_b^2$$

which suggest the sample moment condition g_a to be more important than g_b.

Now, let's expand (2) and we know from the previously tutorials $\bar{\mathbf{g}}(\hat{\beta}_0) = \frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0$ and then

$$\hat{\beta}_{GMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}} n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0)'\mathbf{W}_n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0), \tag{3}$$

This is similar to minimising the SSR of the linear regression model to derive the OLS estimator. Taking the first-order condition respect to $\hat{\beta}_0$ and setting it equal to 0,

$$\frac{\partial}{\partial \hat{\beta}_0} n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0)'\mathbf{W}_n(\frac{1}{n}\mathbf{Z}'\mathbf{y} - \frac{1}{n}\mathbf{Z}'\mathbf{X}\hat{\beta}_0) = 0,$$

$$\frac{\partial}{\partial \hat{\beta}_0}(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\hat{\beta}_0)'\mathbf{W}_n(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\hat{\beta}_0) = 0$$

$$\frac{\partial}{\partial \hat{\beta}_0}(-\mathbf{y}'\mathbf{Z}'\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0 - \hat{\beta}_0\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y} + \hat{\beta}_0'\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0) = 0$$

We can use this use rule that $\mathbf{v} = \mathbf{A}\mathbf{u}$ is equivalent to $\mathbf{v} = \mathbf{u}'\mathbf{A}$, therefore

$$\frac{\partial}{\partial \hat{\beta}_0}(-2\mathbf{y}'\mathbf{Z}'\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0 + \hat{\beta}_0'\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0) = 0$$

Then we apply this next rule, $\psi(\mathbf{x}) = \mathbf{x}\mathbf{A}\mathbf{x}'$ and the $\frac{\partial\psi(\mathbf{x})}{\mathbf{x}} = \mathbf{x}'(\mathbf{A} + \mathbf{A}')$ and if $\mathbf{A}$ is a symmetric matrix then $\frac{\partial\psi(\mathbf{x})}{\mathbf{x}} = 2\mathbf{x}'\mathbf{A}$

$$(-2\mathbf{y}'\mathbf{Z}'\mathbf{W}_n\mathbf{Z}'\mathbf{X} + 2\hat{\beta}_0'\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}) = 0$$

$$\hat{\beta}_0'\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X} = \mathbf{y}'\mathbf{Z}'\mathbf{W}_n\mathbf{Z}'\mathbf{X}$$

Take transpose of both sides

$$\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0 = \mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y},$$

Then pre-multiply both sides by $(\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}$

$$(\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X}\hat{\beta}_0 = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y},$$

Since $(\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X} = \mathbf{I}$, then

$$\hat{\beta}_0 = \hat{\beta}_{GMM} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_n\mathbf{Z}'\mathbf{y}, \quad (4)$$

1.1 Two Step Efficient GMM

According to Hansen (1982) an the efficient GMM is achieved when the optimal weighting matrix $\mathbf{W}_n$ is set to be $\mathbf{S}^{-1}$ which is the inverse of the asymptotic variance of $\bar{g}_n$. This guarantees the GMM estimator have a minimum asymptotic variance. An example of a two step efficient GMM would be to choose an initial weighting matrix, e.g $\mathbf{W}_1 = \mathbf{I}$ or $\mathbf{W}_1 = (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}$, and find a consistent but inefficient first step GMM estimator

$$\hat{\beta}_1 = \underset{\hat{\beta}}{\operatorname{argmin}}\{n\bar{\mathbf{g}}(\hat{\beta})'\mathbf{W}_1\bar{\mathbf{g}}(\hat{\beta})\},$$

From this GMM estimator of $\hat{\beta}_1$ we can derive the optimal weighting matrix $\hat{\mathbf{W}}_2$. Then we can find an efficient estimator using this optimal weight matrix

$$\hat{\beta}_2 = \underset{\hat{\beta}}{\operatorname{argmin}}\{n\bar{\mathbf{g}}(\hat{\beta})'\hat{\mathbf{W}}_2\bar{\mathbf{g}}(\hat{\beta})\},$$

Hence, the $\hat{\beta}_2$ is the two step efficient GMM estimator. Also, efficiency is obtained if we set $\mathbf{S}$ to be

1. $\hat{\mathbf{S}}_{HC}$ the asymptotic estimator of the covariance in the presence of arbitrary heteroskedasticity.
2. $\hat{\mathbf{S}}_{CR}$ the asymptotic estimator of the covariance in the presence of arbitrary heteroskedasticity and within-cluster correlation.
3. $\hat{\mathbf{S}}_{HAC}$ the asymptotic estimator of the covariance in the presence of arbitrary heteroskedasticity and serial correlation.

and substitute it back into (4) of the GMM estimator.

1.2 Non-linear GMM

Similar approach to the standard GMM but now $\bar{\mathbf{g}}(\hat{\beta}_0)$ is nonlinear function of data and parameters. For example $\bar{\mathbf{g}}(\hat{\beta}_0) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i(y_i - (\mathbf{x}_i + \mathbf{x}_{i-1})^2\hat{\beta}_0)$. See attached notes about it.

2 Anderson-Rubin-test

Let's consider a linear model of

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \beta_2\mathbf{x}_2 + \epsilon, \tag{5}$$

$$\mathbf{x}_2 = \mathbf{X}_1\Pi_1 + \mathbf{Z}_2\Pi_1 + \mathbf{v},$$

where $\mathbf{Z}_1 = \mathbf{X}_1$ is your included instruments or exgenous regressor and $\mathbf{x}_2$ is the endogenous regressor. The Anderson-Rubin (AR) test can be specify as

$$H_0 : \beta_2 = \bar{\beta}_2 \text{ and } E(\mathbf{z}_i \epsilon_i) = 0,$$

$$H_1 : \beta_2 \neq \bar{\beta}_2 \text{ or } E(\mathbf{z}_i \epsilon_i) \neq 0$$

If $\bar{\beta}_2$ is close to the true β and weak exogeneity holds $E(\mathbf{z}_i \epsilon_i) = 0$, we won't be able to reject the null hyopthesis. Another way to conduct this test is subtract both sides by $\bar{\beta}_2 \mathbf{x}_2$ from (5)

$$\tilde{\mathbf{y}} = \mathbf{X}_1 \beta_1 + (\beta_2 - \bar{\beta}_2)\mathbf{x}_2 + \epsilon,$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \bar{\beta}_2 \mathbf{x}_2$ and we then substitute $\mathbf{x}_2 = \mathbf{X}_1 \Pi_1 + \mathbf{Z}_2 \Pi_1 + \mathbf{v}$ into the above equations

$$\tilde{\mathbf{y}} = \mathbf{X}_1 \beta_1 + (\beta_2 - \bar{\beta}_2)(\mathbf{X}_1 \Pi_1 + \mathbf{Z}_2 \Pi_1 + \mathbf{v}) + \epsilon,$$

Then simplifying

$$\tilde{\mathbf{y}} = \mathbf{Z}_1 \theta_1 + \mathbf{Z}_2 \theta_2 + \mathbf{u},$$

where $\theta_1 = \beta_1 + (\beta_2 - \bar{\beta}_2)\Pi_1$, $\theta_2 = (\beta_2 - \bar{\beta}_2)\Pi_2$, and $\mathbf{u} = \epsilon + (\beta_2 - \bar{\beta}_2)\mathbf{v}$. Then the equivalent AR test can be conducted by

$$H_0 : \theta_2 = 0 \text{ and } E(\mathbf{z}_i \epsilon_i) = 0,$$

$$H_1 : \theta_2 \neq 0 \text{ or } E(\mathbf{z}_i \epsilon_i) \neq 0$$

Here if $\theta_2 = 0$ this implies $\beta_2 - \bar{\beta}_2 = 0$. Note here if β_2 is weakly identified, Π_2 is small. This means θ_2 will be small, which means we are less likely to reject the null hypothesis and more likely to believe that $\beta_2 = \bar{\beta}_2$. The weaker the identification, the wider the range of possible values of β_2 we will fail to reject.

3 Consider an OLS estimation in which the regressor of interest is treated as exogenous.

From the previous tutorial, we know that the OLS estimator is consistent if $E(\mathbf{g}_i) = E[\mathbf{x}_i\epsilon_i] = 0$, if the weak exogeneity assumption holds

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \Sigma_{xx}^{-1} E(\mathbf{g}_i),$$

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = 0,$$

However, when the weak exogeneity assumption $E(\mathbf{g}_i) = E[\mathbf{x}_i\epsilon_i] \neq 0$ (this implies endogeneity in the regressors and the errors) does not hold

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) \neq 0,$$

Then the OLS estimator is biased. According to tutorial 2, Q8, the sign of the biasedness is determined by

$$p \lim_{n \longrightarrow \infty} (\hat{\beta} - \beta) = \frac{cov[\mathbf{x}_i\epsilon_i]}{var[\mathbf{x}_i]}, \tag{6}$$

Since the denominator is always positive, the sign of the inconsistency in the OLS estimator is given by the numerator. If the regressor is positively correlated with the error, the OLS estimator is biased upwards; if the regressor is negatively correlated with the error, the OLS estimator is biased downwards.

3.1 Huassman test

We can use the Huassman test to test for endogeneity. The basic understanding of Huasman test is that consider we have two estimators, $\hat{\beta}_A$ and $\hat{\beta}_{AB}$, from a linear model. Then the hypothesis test will be

$$H_0 : \text{both} \hat{\beta}_A \text{ and } \hat{\beta}_{AB} \text{ are consistent estimators}$$

$$H_1 : \hat{\beta}_A \text{is a consistent estimator but not } \hat{\beta}_{AB}$$

Then the Huasman test statistics is calculated as

$$H = n(\hat{\beta}_A - \hat{\beta}_{AB})'(\mathbf{V}(\hat{\beta}_A) - \mathbf{V}(\hat{\beta}_{AB}))^{-1}(\hat{\beta}_A - \hat{\beta}_{AB})$$

where $\mathbf{V}(\hat{\beta}_A)$ and $\mathbf{V}(\hat{\beta}_{AB})$ are asymptotic variance of the estimator and $H \underset{d}{\longrightarrow} \chi^2(\nu)$ with a degree of freedom parameter ν.

3.2 GMM distance test

Another test we could consider is the GMM distance test that calculates test statistic based on the difference between the values of two minimized efficient GMM objective functions. For example

1. $\hat{\beta}_1 = \underset{\hat{\beta_A}}{\text{argmin}} J_n(\hat{\beta}_A)$ estimate a GMM estimator where we treat the regressor of interest x_i as exogenous that is the weak exogeneity assumption hold $E[x_i \epsilon_i] = 0$.

2. $\hat{\beta}_2 = \underset{\hat{\beta_{AB}}}{\text{argmin}} J_n(\hat{\beta}_{AB})$ estimate a GMM estimator where we treat the regressor of interest x_i as endogenous that is the weak exogeneity assumption does not hold $E[x_i \epsilon_i] \neq 0$.

3. The calculate the difference $D = J_n(\hat{\beta}_{AB}) - J_n(\hat{\beta}_A)$ between the two functions. If this is large, you can conclude that the regressors are actually endogenous (or something else is wrong).

For the GMM distance test, you have to assume a well specified IV/GMM estimation in which $\mathbf{x}_i$ is treated endogenous. Thus, this means at least one instrument z_i must satisfy the weak exogeneity assumption $E[z_i \epsilon_i] = 0$ and z_i is correlated with x_i (rank condition).

4 Consider an exactly-identied IV estimation in which the regressor of interest is treated as endogenous.

Under assumption 3.4, it assumes the $L \times K$ matrix of $\Sigma_{ZX} = E[\mathbf{z}_i \mathbf{x}_i]$ is full column rank. This full column rank means that each of the columns of Σ_{ZX} are linearly independent. If Σ_{ZX} is not full column rank then this implies that a column in Σ_{ZX} can determined by a linear combination of the other columns in Σ_{ZX} which implies some sort of colinearity exist between the columns. This rank assumption is important to ensure consistency for the $\hat{\beta}_{GMM}$ estimator. Recall from tutorial 2, our proof of consistency for the GMM estimator:

$$p \lim_{n \longrightarrow \infty} (\hat{\beta}_{GMM} - \beta) = p \lim_{n \longrightarrow \infty} ((\mathbf{S}_{ZX}^{'} \mathbf{W}_n \mathbf{S}_{ZX})^{-1}) \times p \lim_{n \longrightarrow \infty} (\mathbf{S}_{ZX}^{'}) \times p \lim_{n \longrightarrow \infty} (\mathbf{W}_n) \times p \lim_{n \longrightarrow \infty} (\bar{\mathbf{g}}), \tag{7}$$

Next, we apply the strong law of large numbers (SLLN) where the sample mean converges to the population mean and the continuous mapping theorem,

$$\mathbf{S}_{ZX}' \underset{a.s}{\longrightarrow} \Sigma_{ZX}', \Rightarrow \mathbf{S}_{ZX}' \underset{p}{\longrightarrow} \Sigma_{ZX}',$$

$$\bar{\mathbf{g}}_n \underset{a.s}{\longrightarrow} E(\bar{\mathbf{g}}_i), \Rightarrow \bar{\mathbf{g}}_n \underset{p}{\longrightarrow} E(\mathbf{g}_i),$$

$$\mathbf{W}_n \underset{a.s}{\longrightarrow} \mathbf{W}, \Rightarrow \mathbf{W}_n \underset{p}{\longrightarrow} \mathbf{W},$$

Note we also have a weak exogeneity assumption, that is $E(\mathbf{g}_i) = 0$. Note that $E(\mathbf{g}_i) = E(\mathbf{z}_i\epsilon_i) = 0$ has to hold to ensure consistency in the GMM estimator. This cannot be test since we assume the model $L = K$ is identified. Therefore,

$$p \lim_{n \longrightarrow \infty} (\hat{\beta}_{GMM} - \beta) = (\Sigma_{ZX}' \mathbf{W} \Sigma_{ZX})^{-1} \times \Sigma_{ZX}' \times \mathbf{W} \times 0, \tag{8}$$

$$p \lim_{n \longrightarrow \infty} (\hat{\beta}_{GMM} - \beta) = 0, \Leftrightarrow, \hat{\beta}_{GMM} \underset{p}{\longrightarrow} \beta,$$

Let's focus on the term $(\Sigma_{ZX}' \mathbf{W} \Sigma_{ZX})^{-1}$on (8). This term will give us a $K \times K$ matrix. However if Σ_{ZX} is not full rank, that is $rank(\Sigma_{ZX}) < K$ (underidentified), then this also implies $rank(\Sigma_{ZX}' \mathbf{W} \Sigma_{ZX}) < K$. If the matrix $(\Sigma_{ZX}' \mathbf{W} \Sigma_{ZX})$ is not full rank, then this matrix is a singular matrix which means is not invertible or the inverse does not exist (Recall matrix is invertible if it is non-singular). Thus, the proof in (8) fails since Σ_{ZX} is not full rank.

Let's consider a very simple two variables linear model with no constant

$$y_i = x_{1,i}\beta_1 + x_{2,i}\beta_2 + \epsilon_i,$$

and let's assume $x_{2,i}$ is the endogenous regressor. We can $z_{1,i} = x_{1,i}$ is the "included instruments" and $z_{2,i}$ is the "excluded instruments". What we want to do next is run an OLS on

$$x_{2,i} = z_{1,i}\delta_1 + z_{2,i}\delta_2 + v_i,$$

Then we test for significance δ_2 whether it is zero or not. If we find that statistically $\delta_2 \neq 0$ then this implies the rank condition is satisfied. For a multivariate case

$$\mathbf{X}_2 = \mathbf{Z}\Delta + \mathbf{v} = \mathbf{Z}_1\Delta_1 + \mathbf{Z}_2\Delta_2 + \mathbf{v},$$

where$\mathbf{X}_2$ is a vector of all endogenous regressors in the linear model, $\mathbf{Z}_1$ is a matrix that contains all the "included instruments" and $\mathbf{Z}_2$ is a matrix that contains all the "excluded instruments". Then if $rank(\Delta) = K$ and $rank(\Delta_2) = K_2$ (K_2 is no. of excluded instruments) then model is identified and the rank condition is satisfied. To test whether a regressor is exogenous, we want to have statistical evidence that $\delta_2 = 0$, this implies that $x_{2,i}$ are linearly independent compared to the other regressors.

4.1 Testing for the weak exogeneity assumption

We know from our proof of consistency for the IV estimator to be

$$p \lim_{n \longrightarrow \infty} (\hat{\beta_{iv}} - \beta) = \frac{cov(z_i, \epsilon_i)}{cov(z_i, x_i)}, \tag{9}$$

Thus, assuming a large sample case and the model is exactly identified ($L = K$), then we have

$$\hat{\beta_{iv}} = \beta + \frac{cov(z_i, \epsilon_i)}{cov(z_i, x_i)} = \frac{cov(z_i, y_i)}{cov(z_i, x_i)},$$

Note if $y_i = x_i\beta + \epsilon_i$, then $cov(z_i, y_i) = cov(z_i, y_i - x_i\beta) = cov(z_i, y_i) + \beta cov(z_i, x_i)$. Next, let's $v_i = y_i - x_i\hat{\beta_{iv}}$ to be the IV residuals, and $v_i = \epsilon_i$ if z_i is a valid instrument. Next, we want find what is $cov(z_i, v_i)$?

$$cov(z_i, v_i) = cov(z_i, y_i - x_i\hat{\beta_{iv}}) = cov(z_i, y_i) - \hat{\beta_{iv}} cov(z_i, x_i),$$

We can then substitute $\hat{\beta_{iv}} = \frac{cov(z_i, y_i)}{cov(z_i, x_i)}$ into the above equation

$$cov(z_i, v_i) = cov(z_i, y_i) - \frac{cov(z_i, y_i)}{cov(z_i, x_i)} cov(z_i, x_i),$$

$$cov(z_i, v_i) = cov(z_i, y_i) - cov(z_i, y_i) = 0,$$

Therefore, in exactly identified model, z_i is always perfectly uncorrelated with the IV residuals by construction regardless of whether the weak exogeneity assumption holds or not. Therefore, the weak exogeneity assumption cannot be tested when no. of instruments is equal to the no. of regressor. However, if we have two instruments $L = 2$ and $K = 1$ regressor, then the IV/GMM estimators will be

$$\hat{\beta_{iv}^{(1)}} = \beta + \frac{cov(z_1, \epsilon_i)}{cov(z_1, x_i)},$$

$$\hat{\beta_{iv}^{(2)}} = \beta + \frac{cov(z_2, \epsilon_i)}{cov(z_2, x_i)},$$

Then

$$\hat{\beta_{iv}^{(1)}} - \hat{\beta_{iv}^{(2)}} = \frac{cov(z_1, \epsilon_i)}{cov(z_1, x_i)} - \frac{cov(z_2, \epsilon_i)}{cov(z_2, x_i)},$$

where $\hat{\beta_{iv}^{(1)}} - \hat{\beta_{iv}^{(2)}} = 0$ implies both $cov(z_1, \epsilon_i) = cov(z_2, \epsilon_i) = 0$ and z_1& z_2 are valid instruments. However, $\hat{\beta_{iv}^{(1)}} - \hat{\beta_{iv}^{(2)}} \neq 0$, then this implies at least one instrument, either z_1 or z_2, must be invalid. Therefore, we can test the weak exogeneity assumption case when we have more instruments than regressors.

4.2 Using IV when in fact the regressor could be treated as exogenous and you could instead use OLS.

If the regressor is actually exogenous but you use IV, the consequence is that you are using an inefficent estimator. If both OLS and IV are consistent, the OLS estimator is more e¢ cient (more precise, smaller variance) and hence preferable.

$$var(\hat{\beta}_{OLS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

$$var(\hat{\beta}_{IV}) = \sigma^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{X})^{-1}.$$

If $\mathbf{X}$ contains all exogenous regressor, then you can see why $var(\hat{\beta}_{OLS})$ will have a lower variance than the $var(\hat{\beta}_{IV})$ since it has less terms.

4.3 Comparison

In section 2, the researcher starts with OLS which means that the researcher believes that the regressors are exogenous and then no valid instruments. But in the section 3, the researcher starts IV which means that the researcher has a priori believe that some of the regressors are endogenous and they have some valid instrument to correct for this issue.

5 Sargan–Hansen test

The Sargan–Hansen (HS) test is a test for over-identifying restrictions that is $L > K$. Recall a feasible GMM estimator is when we set optimal weighting matrix $\mathbf{W}_n$ to be $\mathbf{S}^{-1}$ which is the inverse of the asymptotic variance of $\bar{g}_n$. Thus in notation

$$\hat{\beta}_{EGMM} = \underset{\hat{\beta}_0}{\operatorname{argmin}}\{J_n(\hat{\beta}_0)\},$$

where $J_n(\hat{\beta}_0) = n\bar{\mathbf{g}}(\hat{\beta}_0)'\mathbf{S}^{-1}\bar{\mathbf{g}}(\hat{\beta}_0)$ and the HS test statistic is compute by the value of the minimised GMM objective function $J_n(\hat{\beta}_{EGMM})$. This

$$J_n(\hat{\beta}_{EGMM}) = n\bar{\mathbf{g}}(\hat{\beta}_{EGMM})'\mathbf{S}^{-1}\bar{\mathbf{g}}(\hat{\beta}_{EGMM}) \underset{d}{\longrightarrow} \chi^2(L-K)$$

where $\bar{\mathbf{g}}(\hat{\beta}_{EGMM}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\hat{\epsilon}_i$. A large HS test statistics implies some part of the model is wrong. This indicates a failure of the orthogonality conditions, i.e., some of the variables we are treating as exogenous are actually endogenous (correlated with the disturbance). The opposite is the case with a small HS test statistics. The reasons for large HS test

1. The model is correctly specified but the excluded instruments $\mathbf{z}_i$ are not exogenous and correlated with error.
2. The model is not correctly specified but some of the excluded instruments $\mathbf{z}_i$ should be included as the regressors.
3. The entire model is wrong.
4. The estimator $\hat{\mathbf{S}}$ for $\mathbf{S}$ could be inconsistent.