Further Topics in Regression Analysis:
Topic 3: IV Estimation, Simultaneous Equations and Systems

Daniel Buncic

Institute of Mathematics & Statistics
University of St. Gallen
Switzerland

September 14, 2014
Version: [ftra3-a]

Homepage
www.danielbuncic.com
University of St. Gallen

Outline/Table of Contents

Outline

Method of Moment Estimation (MoM)

So far we have encountered OLS and MLE as "good estimators" (in a statistical sense) of the population parameters of interest based on a given random sample of data.

As we will analyse the problem of correlation between the regressors X and the error term u, it is useful here to introduce another 3^{rd} estimation principle, known as the Method of Moments (MoM)

- MoM is used when the most important assumption of $E(u_i|X_i) = 0$ in OLS is violated

The concept of "*moments*" and its use in finding parameter estimates was first introduced by Karl Pearson in the late 19^{th} and early 20^{th} century

- thus a bit earlier than MLE by Ronald Fisher

- Fisher's MLE in fact came as a response to Pearson's MoM, which Fisher regarded in many ways as sub-optimal
  - Fisher's famous quote "*wasting your time fitting curves by moments, eh!*" sums up his rather negative attitude towards MoM.

Although MLE is the only general class of estimators that is efficient because it reaches the Cramér-Rao lower bound, MLE has a cost.

- it requires a full probabilistic description of the data in terms of the Likelihood function
  - sometimes this is implicitly given by the model that one is trying to estimate
  - other times it needs to be assumed

The **advantage of MoM** estimation is that one only needs to specify (or know) a set of "**moment conditions**" to estimate the population parameters of interest.

- so the information (or assumption) requirement is much weaker than for MLE
- this comes at the cost of having a general loss of efficiency (MoM does **not** reach the Cramér-Rao lower bound in general!)
  - but there exist many situations when MoM and MLE not only yield the same estimator, but also have the same variance of the estimator
  - therefore, MoM **can** be as efficient as MLE.

Expected Values and Moments

Recall that in statistics, the expected value of a RV (or a functional transformation of a RV or sets of RVs) is called a **moment**.

- ie., $E(u_i)$, $E(X_i)$,
- $E(X_i^2)$, $E[\ln(u_i)]$
- $E(u_i X_i)$, $E(u_i^2 X_i)$ etc. are all examples of moments.

The k^{th} (raw) moment of RV y (taken at the origin) is defined as:

$$\begin{aligned} \mu'_k &= E(y^k) \qquad (1) \\ &= \int_y y^k \underbrace{f(y)}_{\text{PDF of } y} dy \end{aligned}$$

with a corresponding sample analogue, the (raw) sample moment or estimator generating equation (EGE) being

$$\hat{\mu}'_k = \frac{1}{n}\sum_{i=1}^{n} y^k. \tag{2}$$

Keep in mind here that μ'_k is the k^{th} "*moment at the origin*" and **not** the "*centered moment*".

The k^{th} centered moment is

$$E[(y - \mu'_1)^k]$$

where

$$\mu'_1 = E(y^1) = E(y)$$

ie., just the unconditional moment or first moment.

For instance, the second centered moment would be the variance of y, ie.

$$\begin{aligned}\mathrm{Var}(y) &= E[(y - \mu'_1)^2] \\ &= E[y^2] - [\mu'_1]^2 \\ &= \underbrace{\mu'_2}_{\text{second moment at origin } E(y^2)!} - \overbrace{\mu^2}^{\text{first moment at origin squared!}}.\end{aligned}$$

The MoM principle

The MoM principle effectively sets the population and the sample moments equal to one another.

This is done by defining a moment condition and setting its expectation equal to zero and then replacing the population moment condition with the EGE to get an estimate.

For the raw moment relation in (1), this means that we would specify the moment condition as:

$$m_i(\mu'_k) = y_i^k - \mu'_k$$

for $\{y_i\}_{i=1}^n$ an $i.i.d.$ sample and given k, and then form the population moment conditions as:

$$E[m_i(\mu'_k)] = 0 \tag{3}$$

$$E[y_i^k - \mu'_k] = 0$$
$$E[y_i^k] = \mu'_k$$

and then replacing $E[\cdot]$ by the EGE $\frac{1}{n}\sum_{i=1}^n[\cdot]$, yielding

$$\frac{1}{n}\sum_{i=1}^n [m_i(\hat{\mu}'_k)] = 0$$
$$\frac{1}{n}\sum_{i=1}^n [y_i^k] = \hat{\mu}'_k.$$

In general, one always specifies a moment condition which involves the data and the parameters of interest.

Notice that the MoM estimator requires "**equations to be solved**"

- ie., we set the population moment condition equal to zero in (3), where the population moments are functions of the parameters of interest!
- in general there will be M equations for p unknown parameters

There are three possible scenarios when we need to solve a set of M equations for p unknown parameters ($\#$ means Number)

1) $\#$ of unknowns $>$ $\#$ of equations $\Rightarrow$ unidentified system
2) $\#$ of unknowns $=$ $\#$ of equations $\Rightarrow$ just identified system
3) $\#$ of unknowns $<$ $\#$ of equations $\Rightarrow$ overidentified system

Scenario 1) we cannot estimate the model parameters.
$\Rightarrow$ no solution exists

Scenario 2) we have a just identified system which can be estimated by MoM.
$\Rightarrow$ exactly one solution exists

Scenario 3) we have more information than needed,
$\Rightarrow$ infinitely many solutions exist.

For the last problem in Scenario 3) we then have the following choices:

- we can either discard some equations until we have the same No. of equations as parameters
- or use an estimation techniques that compresses the extra information to the same dimension as the number of parameters

Fact: Asymptotic Normality and Consistency of MoM

Let $m_i(\theta)$ be an M dimensional vector of moment conditions such that $E[m_i(\theta)] = 0$ and let θ be a $p = M$ dimensional vector or parameters of interest. Also, let $\hat{\theta}_{\text{MoM}}$ be the method of moment (MoM) estimator of θ, given a sample of size $i = 1, \ldots, n$ observations. Then

$$\sqrt{n}(\hat{\theta}_{\text{MoM}} - \theta) \to N(0, \mathbb{V}(\theta)) \tag{4}$$

where

$$\mathbb{V}(\theta) = \left([d']^{-1}\text{Var}[m_i(\theta)][d]^{-1}\right)$$

and

$$d = E\left(\frac{\partial m_i(\theta)}{\partial \theta'}\right)$$

where we replace θ with its consistent estimate $\hat{\theta}_{\text{MoM}}$ in $\mathbb{V}(\theta)$ to obtain a feasible estimate of $\mathbb{V}(\theta)$.

Example 1: Bernoulli distribution

Let $\{x_i\}_{i=1}^n$ be a random ($i.i.d.$) sample of size n from a Bernoulli distribution, with $E(x_i) = p$ and $\text{Var}(x_i) = p(1-p)$, where $p \in [0, 1]$ is the probability of success parameter.

The moment condition is:

$$m_i(p) = x_i - p$$

which then yields the population moment condition

$$E[m_i(p)] = 0$$
$$\Leftrightarrow E(x_i - p) = 0$$

with the corresponding EGE being

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - p) = 0$$

MoM Estimation Examples

Example 1: Bernoulli distribution

$$\Leftrightarrow \qquad \frac{1}{n}\sum_{i=1}^{n} x_i = p.$$

So $\hat{p}_{\mathsf{MoM}} = \bar{x}$, ie. the sample mean. Now the variance is

$$\mathrm{Var}(\hat{p}_{\mathsf{MoM}}) = \frac{1}{n}[d']^{-1}\mathrm{Var}[m_i(p)][d]^{-1}$$

where

$$\begin{aligned}\mathrm{Var}[m_i(p)] &= \mathrm{Var}(x_i - p)\\ &= \mathrm{Var}(x_i)\\ &= p(1-p)\end{aligned}$$

MoM Estimation Examples

Example 1: Bernoulli distribution

and

$$d = E\left(\frac{\partial m_i(p)}{\partial p}\right) = -1.$$

This yields

$$\begin{aligned}\mathrm{Var}(\hat{p}_{\mathsf{MoM}}) &= \frac{1}{n}[-1]^{-1}[p(1-p)][-1]^{-1}\\ &= \frac{1}{n}p(1-p)\end{aligned}$$

and replacing p with a consistent estimate gives

$$\widehat{\mathrm{V}}\mathrm{ar}(\hat{p}_{\mathsf{MoM}}) = \frac{1}{n}\hat{p}_{\mathsf{MoM}}(1-\hat{p}_{\mathsf{MoM}})$$

as the estimate of the variance of $\hat{p}_{\mathsf{MoM}}$.

Example 2: Exponential distribution

Let $\{x_i\}_{i=1}^n$ be a random $(i.i.d.)$ sample of size n from an exponential distribution with $x_i \in [0, \infty)$, with $E(x_i) = \theta$ and $\mathrm{Var}(x_i) = \theta^2$, where $\theta > 0$ is a scale parameter.

The moment condition is:

$$m_i(\theta) = x_i - \theta$$

which then yields the population moment condition

$$\begin{aligned} E[m_i(\theta)] &= 0 \\ \Leftrightarrow E(x_i - \theta) &= 0 \end{aligned}$$

with the corresponding EGE being

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \theta) = 0$$

$$\Leftrightarrow \qquad \frac{1}{n}\sum_{i=1}^{n} x_i = \theta.$$

So $\hat{\theta}_{\text{MoM}} = \bar{x}$, ie. the sample mean again. Now for the variance we need

$$\begin{aligned} \mathrm{Var}[m_i(\theta)] &= \mathrm{Var}(x_i - \theta) \\ &= \mathrm{Var}(x_i) \\ &= \theta^2 \end{aligned}$$

and we have again

$$d = E\left(\frac{\partial m_i(\theta)}{\partial \theta}\right) = -1.$$

This yields

$$\begin{aligned}\text{Var}(\hat{\theta}_{\textsf{MoM}}) &= \frac{1}{n}[-1]^{-1}[\theta^2][-1]^{-1} \\ &= \frac{1}{n}\theta^2\end{aligned}$$

and replacing θ with a consistent estimate yields

$$\widehat{\text{Var}}(\hat{\theta}_{\textsf{MoM}}) = \frac{1}{n}\hat{\theta}^2_{\textsf{MoM}}$$

as the estimate of the variance of $\hat{\theta}_{\textsf{MoM}}$.

Example 3: Poisson distribution

Let $\{x_i\}_{i=1}^n$ be a random $(i.i.d.)$ sample of size n from a poisson distribution with $x_i = 0, 1, 2, \ldots$, where $\lambda > 0$ is a parameter that is related to the number of events in a fixed time interval.

Note that the first 4 **raw moments** of x_i are:

$$\begin{aligned}\mu'_1 &= \lambda \\ \mu'_2 &= \lambda(1+\lambda) \\ \mu'_3 &= \lambda(1+3\lambda+\lambda^2) \\ \mu'_4 &= \lambda(1+7\lambda+6\lambda^2+\lambda^3).\end{aligned}$$

To make things different this time, let us define $z = x^2$ so that

$$E(z) = E(x^2)$$

$$
\begin{aligned}
&= \mu_2' \\
&= \lambda(1+\lambda)
\end{aligned}
$$

and use the moment condition:

$$m_i(\lambda) = z_i - \lambda(1+\lambda)$$

which then yields the population moment condition

$$
\begin{aligned}
E[m_i(\lambda)] &= 0 \\
\Leftrightarrow E[z_i - \lambda(1+\lambda)] &= 0
\end{aligned}
$$

with the corresponding EGE being

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(z_i - (\lambda + \lambda^2)) &= 0 \\
\Leftrightarrow \qquad \lambda^2 + \lambda - \bar{z}_i &= 0.
\end{aligned}
$$

Now this is a quadratic in λ, so that there are two solutions:

$$
\begin{aligned}
\lambda_A &= \frac{1}{2}\sqrt{4\bar{z}_i + 1} - \frac{1}{2} \\
\lambda_B &= -\frac{1}{2}\sqrt{4\bar{z}_i + 1} - \frac{1}{2}.
\end{aligned}
$$

Recall that $\lambda > 0$ for a Poisson RV, thus we can rule out λ_B as $\bar{z}_i = \frac{1}{n}\sum_{i=1}^{n} x_i^2 > 0$, and hence the whole term is always negative, even if all $x_i = 0$ in our sample.

The solution for λ_A gives then $\hat{\lambda}_{\text{MoM}} = \frac{1}{2}\sqrt{4\bar{z}_i + 1} - \frac{1}{2}$ as the MoM point estimate (Note that if $\bar{z}_i = 0$ then we get an inadmissable solution for λ).

Now for the variance we need to find

$$\begin{aligned}
\text{Var}[m_i(\lambda)] &= \text{Var}[z_i - \lambda(1+\lambda)] \\
&= \text{Var}(z_i) \\
&= E(z_i^2) - [E(z_i)]^2 \\
&= \underbrace{E(x_i^4)}_{=\mu_4'} - [\lambda(1+\lambda)]^2 \\
&= \lambda(1 + 7\lambda + 6\lambda^2 + \lambda^3) - [\lambda(1+\lambda)]^2 \\
&= \lambda\left(4\lambda^2 + 6\lambda + 1\right)
\end{aligned}$$

and we have

$$d = E\left(\frac{\partial m_i(\lambda)}{\partial \lambda}\right) = -(1 + 2\lambda).$$

This yields then

$$\begin{aligned}
\text{Var}(\hat{\lambda}_{\text{MoM}}) &= [d']^{-1}\text{Var}[m_i(\lambda)][d]^{-1} \\
&= \frac{1}{n}[-(1+2\lambda)]^{-1}[\lambda\left(4\lambda^2 + 6\lambda + 1\right)][-(1+2\lambda)]^{-1} \\
&= \frac{1}{n}(1+2\lambda)^{-2}\lambda\left(4\lambda^2 + 6\lambda + 1\right) \\
&= \frac{1}{n}\left[\lambda + \frac{2\lambda^2}{(1+2\lambda)^2}\right] \qquad (5)
\end{aligned}$$

where we would again replace λ with a consistent estimate $\hat{\lambda}_{\text{MoM}}$.

Now if we use the simple moment condition on the mean rather than the second moment μ_2', we would have

$$m_i(\lambda) = x_i - \lambda \tag{6}$$

and thus

$$\begin{aligned} E[m_i(\lambda)] &= 0 \\ \Leftrightarrow E[x_i - \lambda] &= 0 \end{aligned}$$

giving the EGE

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(x_i - \lambda) &= \\ \bar{x} &= \hat{\lambda}_{\mathsf{MoM}}. \end{aligned}$$

The variance of $\hat{\lambda}_{\mathsf{MoM}}$ using the first moment as in (6) is then given by

$$\begin{aligned} \mathrm{Var}[m_i(\lambda)] &= \mathrm{Var}(x_i - \lambda) \\ &= \mathrm{Var}(x_i) \\ &= \lambda \end{aligned}$$

yielding

$$\mathrm{Var}(\hat{\lambda}_{\mathsf{MoM}}) = \frac{1}{n}\,[\lambda] \tag{7}$$

with $d = [-1]$ as before.

We can see that the estimators are different not only in terms of how the point estimates are computed, but also in terms of their variances.

Nonetheless, both are consistent estimates of λ.

- the increase in the variance from using the moment condition $z_i - \lambda(1+\lambda)$ over $x_i - \lambda$ is $\frac{2\lambda^2}{(1+2\lambda)^2}$
- since $\lambda > 0$, this is bounded below by 0 and above by $\lim_{\lambda\to\infty} \frac{2\lambda^2}{(1+2\lambda)^2} = \frac{1}{2}$. (see also the plot of $\frac{2\lambda^2}{(1+2\lambda)^2}$ in the figure below)

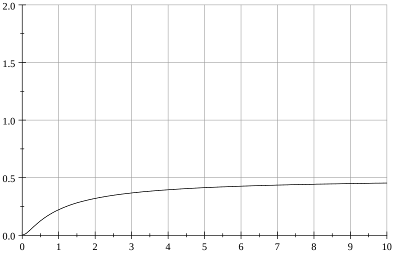


Figure 1: Amplification Factor for variance of moment condition $m_i = z_i - \lambda(1+\lambda)$.

Example 4: Regression model

Suppose we have an $i.i.d$ sample of data for $\{y_i, x_i\}_{i=1}^{n}$ from the regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{8}$$

where we make the (standard OLS) assumptions:

1) $E(u_i) = 0$ (zero error on average) with $\mathrm{Var}(u_i) = \sigma^2$
2) $E(u_i|x_i) = 0$ (conditional on x_i we have have zero errors on average).

Note that from 1) we have the population relations:

$$\begin{aligned} & E(u_i) = 0 \qquad (9) \\ \Leftrightarrow\ & E(y_i) = \beta_0 + \beta_1 E(x_i) \\ \Leftrightarrow\ & \quad \beta_0 = E(y_i) - \beta_1 E(x_i) \end{aligned}$$

which yields the EGE for β_0

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \tag{10}$$

The relation involving the conditional moment restriction in 2) yields

$$\begin{aligned} E(u_i|x_i) &= 0 \\ x_i E(u_i|x_i) &= \underbrace{x_i 0}_{=0} \\ E_x[x_i E(u_i|x_i)] &= E_x[0] \\ E[u_i x_i] &= 0, \qquad (11) \end{aligned}$$

where $E[u_i x_i] = 0$ from (11) can be used as the second (unconditional) moment restriction.

Notice here that in (8) we have two population parameters of interest (β_0 and β_1) so we need at least two moment conditions to solve for these two parameters.

- one was given in (9) and the second is (11)

Using (11) yields

$$\begin{aligned} E(u_i x_i) &= 0 \\ E[(y_i - \beta_0 + \beta_1 x_i)x_i] &= 0 \\ E(y_i x_i) &= E(\beta_0 x_i + \beta_1 x_i^2). \end{aligned} \tag{12}$$

Replacing the expectations again with the EGE and β_0 by $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, yields then for (12)

$$\overline{yx} = \overline{x^2}\hat{\beta}_1 + (\bar{y} - \hat{\beta}_1 \bar{x})\bar{x}$$

$$\underbrace{\overline{yx} - \bar{y}\bar{x}}_{\text{Sample Cov}(y_i, x_i)} = \hat{\beta}_1 \underbrace{\left(\overline{x^2} - \bar{x}^2\right)}_{\text{Sample Var}(x_i)} \tag{13}$$

where $\overline{yx} = \frac{1}{n}\sum_{i=1}^n y_i x_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$.

From (13) we than have

$$\begin{aligned} \hat{\beta}_1 &= \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2} \\ &= \frac{\text{Sample Cov}(y_i, x_i)}{\text{Sample Var}(x_i)} \end{aligned}$$

as for the classical OLS estimator and also for MLE.

MoM Estimation Examples

Example 4: Regression model

To find the variance, of the (two dimensional vector) $\hat{\boldsymbol{\beta}}_{\text{MoM}} = \left(\hat{\beta}_0 \ \hat{\beta}_1\right)'$ estimator, we need again the two terms $\text{Var}(m_i)$ as well as $E(\partial m_i/\partial\theta)$. With

$$m_i = \begin{bmatrix} u_i \\ u_i x_i \end{bmatrix}$$

having the property that

$$E(m_i) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

we can compute $\text{Var}(m_i)$ as

$$\begin{aligned} \text{Var}(m_i) &= E(m_i m_i') \\ &= E\left(\begin{bmatrix} u_i \\ u_i x_i \end{bmatrix} \begin{bmatrix} u_i & u_i x_i \end{bmatrix}\right) \end{aligned}$$

MoM Estimation Examples

Example 4: Regression model

$$\begin{aligned} &= E\left(\begin{bmatrix} u_i^2 & u_i^2 x_i \\ u_i^2 x_i & u_i^2 x_i^2 \end{bmatrix}\right) \\ &= \begin{bmatrix} \sigma^2 & \sigma^2 E(x_i) \\ \sigma^2 E(x_i) & \sigma^2 E(x_i^2) \end{bmatrix} \\ &= \sigma^2 \begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix} \end{aligned} \tag{14}$$

where the last two lines follow because of the conditional independence between u_i and x_i.

Now to find $d = E(\partial m_i/\partial\boldsymbol{\beta}')$, we need to look at

$$E\left[\frac{\partial m_i}{\partial\boldsymbol{\beta}'}\right] = E\begin{bmatrix} \partial u_i/\partial\boldsymbol{\beta}' \\ \partial u_i x_i/\partial\boldsymbol{\beta}' \end{bmatrix}$$

MoM Estimation Examples

Example 4: Regression model

$$= E\begin{bmatrix} \partial(y_i - \beta_0 - \beta_1 x_i)/\partial\boldsymbol{\beta}' \\ \partial(y_i x_i - \beta_0 x_i - \beta_1 x_i^2)/\partial\boldsymbol{\beta}' \end{bmatrix}$$

$$= E\begin{bmatrix} \partial(y_i - \beta_0 - \beta_1 x_i)/\partial\beta_0 & \partial(y_i - \beta_0 - \beta_1 x_i)/\partial\beta_1 \\ \partial(y_i x_i - \beta_0 x_i - \beta_1 x_i^2)/\partial\beta_0 & \partial(y_i x_i - \beta_0 x_i - \beta_1 x_i^2)/\partial\beta_1 \end{bmatrix} \quad (15)$$

$$= E\begin{bmatrix} -1 & -x_i \\ -x_i & -x_i^2 \end{bmatrix}$$

$$d = -\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix} \quad (16)$$

The variance of $\hat{\boldsymbol{\beta}}_{\text{MoM}}$ is then found from combining (16) and (14) as:

MoM Estimation Examples

Example 4: Regression model

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{MoM}}) = \frac{1}{n}[d']^{-1}\text{Var}[m_i(\lambda)][d]^{-1}$$

$$= \frac{1}{n}\left(-\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix}^{-1}\right)'\left(\sigma^2\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix}\right) \quad (17)$$

$$\times\left(-\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix}^{-1}\right) \quad (18)$$

$$= \frac{1}{n}\sigma^2\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix}^{-1}. \quad (19)$$

Now the inverse of 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ is } A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where $\det(A) = ad - cb$.

MoM Estimation Examples
Example 4: Regression model

For the inverse in (19) this means that

$$\begin{bmatrix} 1 & E(x_i) \\ E(x_i) & E(x_i^2) \end{bmatrix}^{-1} = \frac{1}{E(x_i^2) - [E(x_i)]^2} \begin{bmatrix} E(x_i^2) & -E(x_i) \\ -E(x_i) & 1 \end{bmatrix}.$$

Replacing $E(\cdot)$ with its EGE and σ^2 with a consistent estimate $\hat{\sigma}^2$, we get the familiar OLS variance and covariance results:

$$\begin{aligned} \mathrm{Var}(\hat{\beta}_1^{\mathsf{MoM}}) &= \frac{1}{n}\sigma^2 \widehat{\mathrm{V}}\mathrm{ar}(x_i)^{-1} \\ &= \hat{\sigma}^2 \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{-1} \end{aligned}$$

MoM Estimation Examples
Example 4: Regression model

$$\begin{aligned} \mathrm{Var}(\hat{\beta}_0^{\mathsf{MoM}}) &= \frac{1}{n}\sigma^2 \hat{E}(x_i^2) \widehat{\mathrm{V}}\mathrm{ar}(x_i)^{-1} \\ &= \frac{1}{n}\sigma^2 \sum_{i=1}^{n} x_i^2 \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{-1} \\ \mathrm{Cov}(\hat{\beta}_0^{\mathsf{MoM}}, \hat{\beta}_1^{\mathsf{MoM}}) &= -\frac{1}{n}\sigma^2 \hat{E}(x_i) \widehat{\mathrm{V}}\mathrm{ar}(x_i)^{-1} \\ &= -\frac{1}{n}\sigma^2 \sum_{i=1}^{n} x_i \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{-1} \\ &= -\sigma^2 \bar{x} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{-1} \end{aligned}$$

Failure of Moment Condition

We have previously seen that the most important OLS assumption is:

$$E(u_i|X_i) = 0 \tag{20}$$

which implies that

$$\begin{aligned} & \text{Cov}(u_i, X_i) = 0 \\ \Leftrightarrow \quad & E(u_i X_i) = 0 \end{aligned} \tag{21}$$

because it influences the consistency and unbiasedness of the OLS estimator.

This does not go away asymptotically, a the OLS estimator will converge to a pseudo true value, ie., a value that is the true value plus some bias.

Some common examples when the assumption $E(u_i|X_i) = 0 = E(u_i X_i)$ fails are under the following scenarios:

1) Measurement errors or errors in variables
2) Omitted variable bias
3) Simultaneous equation bias

Let us look at each scenario individually.

Measurement errors or errors in variables

The explanatory variable X_i is measured with an error (or imprecisely)

- this is pretty much a universal problem, especially when dealing with micro-level data on household behaviour
  - often micro level data is collected through surveys, either in written form or by telephone surveys
  - respondents frequently do not recall the exact figures correctly, or are not willing to disclose the actual figure
  - typically happens when respondents are asked about how much money is spent on certain types of goods, such as alcohol, cigarettes, or how much is earned, or how much time is dedicated to a task or leisure activity, or when a proxy variable is used for ability

- with macroeconomic data, a common example is the use of "*permanent income*" in relation to consumption.
  - permanent income is frequently proxied by current income.

As an illustration, consider the following simple example.

Suppose you are interested in a relationship that takes the form

$$y_i = \beta_0 + \beta_1 x_i^* + u_i \tag{22}$$

but the x_i^* variable is not observed directly, and all that is available is another random variable x_i which relates to x_i^* as:

$$x_i = x_i^* + e_i \tag{23}$$

where e_i is an $i.i.d.$ random variable with mean 0 and variance σ_e^2 (think Normal RV for example, but does not need to be).

So x_i is all that I have to proxy x_i^*, where the size of the error of the proxy is determined by the variance of e_i.

We need to make an assumption about the relation between x_i^* and e_i and u_i and e_i.

- it is common to assume $E(x_i^* e_i) = 0$ as in a "standard regression" problem.
- it is further frequently assumed that $E(u_i e_i) = 0$, ie, the two disturbances in (22) and (23) are uncorrelated.

Now using x_i instead of x_i^* in the regression relation in (22), we get

$$\begin{aligned} y_i &= \beta_0 + \beta_1 (x_i - e_i) + u_i \\ &= \beta_0 + \beta_1 x_i + \underbrace{u_i - \beta_1 e_i}_{=v_i} \\ &= \beta_0 + \beta_1 x_i + \underbrace{v_i.}_{=(u_i - \beta_1 e_i)} \end{aligned} \quad (24)$$

When using OLS to estimate the model in (24), we implicitly assume the OLS moment condition $E(x_i v_i) = 0$.

To see whether this is a valid assumption to make, we need to verify this moment condition from the assumptions of the model.

Evaluating $E(x_i v_i)$ yields

$$\begin{aligned} E(x_i v_i) &= E[(x_i^* + e_i)(u_i - \beta_1 e_i)] \\ &= E[x_i^*(u_i - \beta_1 e_i)] + E[e_i(u_i - \beta_1 e_i)] \quad (25) \\ &= E(x_i^* u_i) - \beta_1 E(x_i^* e_i) + E(e_i u_i) - \beta_1 E(e_i^2). \quad (26) \end{aligned}$$

Thus, we can see that even if we make the assumptions that $E(x_i^* u_i) = 0$, $E(x_i^* e_i) = 0$, and $E(e_i u_i) = 0$, it still follows that

$$\begin{aligned} E(x_i v_i) &= -\beta_1 E(e_i^2) \\ &= -\beta_1 \sigma_e^2 \neq 0. \end{aligned} \quad (27)$$

Notice that σ_e^2 is the variance of the measurement error in (23)!

- so the larger this variance, the larger will be the bias in the OLS estimator $\hat{\beta}_1$.

Recall from our standard expansion of the OLS estimator $\hat{\beta}_1$ for the relation in (24) (with $\beta_0 = 0$ for simplicity of exposition), that we had

$$E(\hat{\beta}_1) = \beta_1 + \underbrace{E\left[\frac{\sum_{i=1}^n x_i v_i}{\sum_{i=1}^n x_i^2}\right]}_{\text{bias generating term}} \quad (28)$$

$$= \beta_1(1 - \sigma_e^2/\sigma_x^2) \quad (29)$$

where $\frac{1}{n}\sum_{i=1}^n x_i^2 \xrightarrow{p} \sigma_x^2 = \mathrm{Var}(x_i)$, so that there will be a **wedge** between $E(\hat{\beta}_1)$ and β_1 which depends on σ_e^2/σ_x^2.

Omitted variable bias

We have already seen what happens when we forget to include an important regressor or explanatory variable in Topic 1.

Suppose we assume again the following simple "true" model (again with $\beta_0 = 0$ and $E(y_i) = E(x_{1i}) = E(x_{2i}) = 0$):

$$y_i = \beta_1 x_{1i} + \beta_2 x_{i2} + u_i \quad (30)$$

but we estimate

$$y_i = \beta_1 x_{1i} + e_i \quad (31)$$

$$\text{where } e_i = \beta_2 x_{2i} + u_i$$

Now to estimate β_1, OLS uses the moment condition $E(e_i x_i) = 0$, but we see that

$$E(e_i x_i) = \beta_2 \underbrace{E(x_{1i}x_{2i})}_{=\text{Cov}(x_{1i},x_{2i})} + \underbrace{E(x_{i1}u_i)}_{=0},$$

so this moment condition will only hold in the population if the two regressors x_{1i} and x_{2i} are orthogonal to one another, ie, are uncorrelated.

- this is only rarely the case, can be when the regressors that are used are constructed in that way
  - ie, the factors retrieved from PCA, residuals and regressors from regression output

Doing the same expansion as in (28) yields the relationship

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\text{Cov}(x_{1i}, x_{2i})}{\text{Var}(x_{1i})} \tag{32}$$

so how large the bias is depends on both $\frac{\text{Cov}(x_{1i},x_{2i})}{\text{Var}(x_{1i})}$ as well as the magnitude of β_2.

Simultaneous equation bias

Suppose you assume that

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{33}$$

but at the same time the x_i are determined by the relation

$$x_i = \alpha_0 + \alpha_1 y_i + e_i. \tag{34}$$

In such a situation, the explanatory variable x_i is said to be an "endogenous variable" or a variable that is "determined within the system".

If you do not know that the relation in (34) holds for x_i and y_i, and proceed to estimate β_1 by OLS you will implicitly again assume that the moment condition $E(x_i u_i) = 0$ is valid.

To see if this is the case, we again evaluate

$$\begin{aligned} E(x_i u_i) &= E[(\alpha_0 + \alpha_1 y_i + e_i)u_i] \\ &= \alpha_1 E(y_i u_i) \qquad \text{[with } E(u_i e_i) = 0\text{]} \\ &= \alpha_1 E[(\beta_0 + \beta_1 x_i + u_i)u_i] \\ &= \alpha_1 \beta_1 E(x_i u_i) + \alpha_1 \underbrace{E(u_i^2)}_{=\sigma_u^2} \end{aligned}$$

$$E(x_i u_i) = \frac{\alpha_1 \sigma_u^2}{(1 - \alpha_1 \beta_1)} \neq 0. \tag{35}$$

From (35) we can see that there will be a wedge between $E(\hat{\beta}_1)$ and β_1 which will be related to $\alpha_1$and σ_u^2 as well as β_1 and $\mathrm{Var}(x_i)$ (from denominator term).

Simultaneous equations models are a specialized field in econometrics, so that we will devote some more time to them towards the end of these lecture notes.

Most of terminology in IV literature comes from simultaneous equation models literature.

Two important things to take away from the above

1) when $E(u_i|x_i) \neq 0$, then we will have a biased and inconsistent estimator when using OLS.
   - ⇒ magnitude of bias will depend on the problem at hand and can sometimes be not so sever, but other times can be very severe
2) we **cannot test** whether the correlation between x_i and u_i is equal to zero!
   - ⇒ replacing the unobserved u_i with the fitted ones from OLS, ie., $\hat{u}_i$, yields by definition and construction of the OLS estimator zero, ie., we have:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

because that is what the FOC of OLS is!

- ⇒ any insights into whether the condition $E(u_i|x_i) = 0$ is violated needs to come from **economic theory**
- ⇒ need to think about any possible true (or alternative viable) model and evaluate algebraically by hand what the population moment is under the scenarios considered
- ⇒ this may again depend on the assumption of the alternative model considered if there is not strong theory to suggest a correct model structure
- ⇒ most of the time violation will fall in one of the above outlined scenarios.

Estimation with Instrumental variables

Let us now look at how to estimate the parameters of the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{36}$$

using Instrumental Variables (IVs) when the important OLS assumption $E(u_i x_i) = 0$ does not hold.

Note that we have two unknown parameters in (36), so that we need (at least) two population moment conditions to estimate β_0 and β_1.

One will be still

$$\begin{aligned} E(u_i) &= 0 \\ E(y_i - \beta_0 - \beta_1 x_i) &= 0 \\ &\Leftrightarrow \beta_0 = E(y_i) - \beta_1 E(x_i). \end{aligned} \tag{37}$$

Since $E(x_i u_i) = 0$ fails, we need to find another variable z_i for which the population moment condition $E(z_i u_i) = 0$ holds.

- this will be our second population moment condition which we need to be able to solve the system uniquely.
- variable z_i is called an "instrumental variable" (or instrument for short).

We can then form the second population moment condition as:

$$\begin{aligned} E(z_i u_i) &= 0 \\ E[z_i(y_i - \beta_0 - \beta_1 x_i)] &= 0 \\ E(z_i y_i) - \beta_0 E(z_i) - \beta_1 E(z_i x_i) &= 0 \end{aligned} \tag{38}$$

and substituting for $\beta_0 = E(y_i) - \beta_1 E(x_i)$ from (37) above, we get

$$E(z_i y_i) - [E(y_i) - \beta_1 E(x_i)]E(z_i) - \beta_1 E(z_i x_i) = 0$$
$$\underbrace{E(z_i y_i) - E(y_i)E(z_i)}_{=\mathrm{Cov}(z_i,y_i)} - \beta_1 \underbrace{[E(z_i x_i) - E(x_i)E(z_i)]}_{=\mathrm{Cov}(z_i,x_i)} = 0$$

so that

$$\beta_1 = \frac{\mathrm{Cov}(z_i, y_i)}{\mathrm{Cov}(z_i, x_i)}. \tag{39}$$

We can get estimates for β_0 and β_1 from (37) and (38) by replacing the population expectation with the sample analogues (ie. the EGE) which yields

$$\hat{\beta}_0^{\mathsf{IV}} = \bar{y} - \hat{\beta}_1^{\mathsf{IV}} \bar{x} \tag{40}$$

$$\hat{\beta}_1^{\mathsf{IV}} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})} \tag{41}$$

where the $\frac{1}{n}$ terms in the numerator and denominator of (41) cancel.

Notice from (40) and (41) how the estimate of the intercept term in the regression has the same structure as before from OLS estimation

- it is still a linear combination of the sample means of x and y
- but with the weight given by $\hat{\beta}_1^{\mathsf{IV}}$ and **not the OLS one** $\hat{\beta}_1^{\mathsf{OLS}}$.

Asymptotic Distribution of $\hat{\beta}_1^{\text{IV}}$

To see what the asymptotic distribution of $\hat{\beta}_1^{\text{IV}}$ looks like, let us subtract the unconditional mean of (36) to form

$$\begin{aligned} y_i - E(y_i) &= \beta_0 + \beta_1 x_i + u_i - E(\beta_0 + \beta_1 x_i + u_i) \\ y_i - E(y_i) &= \beta_1[x_i - E(x_i)] + u_i - \underbrace{E(u_i)}_{=0} \\ y_i - E(y_i) &= \beta_1[x_i - E(x_i)] + u_i. \qquad (42) \end{aligned}$$

Replacing $E(\cdot)$ with sample analogues again yields

$$(y_i - \bar{y}) = \beta_1(x_i - \bar{x}) + u_i. \qquad (43)$$

Now going back to the IV estimator for β_1 in (41), we can replace $(y_i - \bar{y})$ with the right hand side of (43) to yield

$$\begin{aligned} \hat{\beta}_1^{\text{IV}} &= \frac{\sum_{i=1}^n (z_i - \bar{z})(\beta_1(x_i - \bar{x}) + u_i)}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})} \\ &= \beta_1 + \frac{\sum_{i=1}^n (z_i - \bar{z})u_i}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})} \\ &= \beta_1 + \frac{\sum_{i=1}^n z_i u_i}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}. \qquad (44) \end{aligned}$$

To get the asymptotic distribution of $\hat{\beta}_1^{\text{IV}}$ we can again take probability limits of the numerator and denominator in (44) above.

Doing this yields:

$$\hat{\beta}_1^{\text{IV}} = \beta_1 + \frac{\overbrace{\text{plim}\, \frac{1}{n} \sum_{i=1}^{n} z_i u_i}^{=\text{plim}\, \bar{\nu} = E(\nu_i)}}{\underbrace{\text{plim}\, \frac{1}{n} \sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})}_{=\text{Cov}(z_i, x_i)}} \tag{45}$$

where $\nu_i = z_i u_i$, which is similar to the standard OLS set up we had earlier.

Due to the assumption of $E(z_i u_i) = E(\nu_i) = 0$ and $\text{plim}\, \bar{\nu} = E(\nu_i)$, we see that the IV estimator is consistent as long as $E(z_i u_i) = 0$ holds.

We can again invoke a CLT to show that asymptotically

$$\sqrt{n}(\hat{\beta}_1^{\text{IV}} - \beta_1) \xrightarrow{d} N\left(0, \text{Var}\left(\hat{\beta}_1^{\text{IV}}\right)\right)$$

where

$$\begin{aligned} \text{Var}(\hat{\beta}_1^{\text{IV}}) &= \text{Var}\left(\frac{\bar{\nu}}{\text{Cov}(x_i z_i)}\right) \\ &= \frac{\text{Var}\,(\bar{\nu})}{\text{Cov}(x_i z_i)^2} \\ &= \frac{\text{Var}\,(\nu_i)}{n\text{Cov}(x_i z_i)^2} \end{aligned} \tag{46}$$

Notice from (46) that this is the general heteroskedasticity robust formula.

If we further assume that the variance of u_i does not change with our instrument z_i, then we can write (46) as

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1^{\mathsf{IV}}) &= \frac{\mathrm{Var}(z_i u_i)}{n\mathrm{Cov}(x_i z_i)^2} \\ &= \frac{\mathrm{Var}(z_i)\mathrm{Var}(u_i)}{n\mathrm{Cov}(x_i z_i)^2} \\ &= \frac{\mathrm{Var}(u_i)}{n\mathrm{Var}(x_i)\underbrace{\frac{\mathrm{Cov}(x_i z_i)^2}{\mathrm{Var}(z_i)\mathrm{Var}(x_i)}}_{=\mathrm{Corr}(x_i, z_i)^2}} \end{aligned} \tag{47}$$

which can be estimated as:

$$\widehat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_1^{\mathsf{IV}}) = \frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})} \frac{1}{\hat{\rho}(x_i z_i)^2}. \tag{48}$$

where

- $\hat{\rho}(x_i z_i)^2$ is the squared sample correlation coefficient between x_i and z_i, which in this simple case coincides with the R^2 of a regression of x_i on a constant and z_i
- $\hat{\sigma}_u^2$ is the (sample) variance of the **IV residuals** $\hat{u}_i = y_i - \hat{\beta}_0^{\mathsf{IV}} - \hat{\beta}_1^{\mathsf{IV}} x_i$.

General Matrix notation of IV estimation

Recall that we had the general $k-$variable regression model specified as:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i \tag{49}$$

where

$$\underset{(1\times k)}{\mathbf{x}_i} = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1k} \end{bmatrix}, \quad \underset{(k\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

or in matrix form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{50}$$

where

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n\times k)}{\mathbf{X}} = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1k} \\ x_{20} & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{nk} \end{bmatrix}, \text{and } \underset{(n\times 1)}{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ y_n \end{bmatrix}$$

with the first entry of $\mathbf{X}$ being generally a column of ones and β defined as before.

The MoM estimator of the k unknown parameters would have the ($k-$dimensional) population moment condition

$$\underset{(k\times 1)}{E(\mathbf{x}_i' u_i)} = \mathbf{0} \tag{51}$$

and the corresponding EGE would be

$$\frac{1}{n}\mathbf{X}'\mathbf{u} = \mathbf{0} \tag{52}$$

$$= \underbrace{\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i0}u_i \\ \frac{1}{n}\sum_{i=1}^{n} x_{ii}u_i \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}u_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ik}u_i \end{bmatrix}}_{(k\times 1)} = \mathbf{0}.$$

Note that the $1/n$ term does not influence this moment condition and it is purely written to show its resemblance to the sample moment that we used earlier.

Therefore, (52) can equivalently be written as

$$\mathbf{X}'\mathbf{u} = \mathbf{0} \tag{53}$$

From (53) we can get the MoM estimator of $\boldsymbol{\beta}$ as

Instrumental Variable Estimation
General Matrix notation of IV estimation

$$\begin{aligned} \mathbf{X}'\mathbf{u} &= \mathbf{0} \\ \Leftrightarrow \mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{0} \\ \mathbf{X}'\mathbf{y} &= \mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ \Leftrightarrow \hat{\boldsymbol{\beta}}^{\text{MoM}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \end{aligned} \tag{54}$$

To get the asymptotic properties, expand (54) as before to yield

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\text{MoM}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1}\frac{1}{n}\mathbf{X}'\mathbf{u} \end{aligned} \tag{55}$$

and noting that

$$\text{plim } \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q_{xx}}$$
$$\text{plim } \frac{1}{n}\mathbf{X}'\mathbf{u} = E(\mathbf{x}_i' u_i)$$

with $\mathbf{Q_{xx}}$ being a full rank (ie invertible) matrix and $E(\mathbf{x}_i' u_i) = \mathbf{0}$.

With these assumptions in place, we can use a CLT to show that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{MoM}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(0, \text{Var}\left(\hat{\boldsymbol{\beta}}^{\text{MoM}}\right)\right)$$

where $\text{Var}\left(\hat{\boldsymbol{\beta}}^{\text{MoM}}\right) = \mathbf{Q_{xx}^{-1}}\text{Var}(\mathbf{x}_i' u_i)\mathbf{Q_{xx}^{-1}}$, which under the assumption of homoskedasticity reduces to

$$\text{Var}\left(\hat{\boldsymbol{\beta}}^{\text{MoM}}\right) = \mathbf{Q_{xx}^{-1}}\text{Var}(u_i)$$

and can be estimated as

$$\widehat{\text{Var}}\left(\hat{\boldsymbol{\beta}}^{\text{MoM}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\sigma}_u^2. \tag{56}$$

The IV Estimator has the EGE

$$\mathbf{Z}'\mathbf{u} = \mathbf{0} \tag{57}$$

in analogy to (53), where $\mathbf{Z}$ is the $n \times k$ matrix of instruments for $\mathbf{X}$. Expanding (57) we get

$$\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$
$$\mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X}\boldsymbol{\beta}$$
$$\Leftrightarrow \hat{\boldsymbol{\beta}}^{\text{IV}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}. \tag{58}$$

To get the asymptotic properties, we again take (58) and replace $\mathbf{y}$ with its right hand side to get

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\mathsf{IV}} &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{u}) \\
&= (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{X})\boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} \\
&= \boldsymbol{\beta} + \Bigg(\underbrace{\frac{1}{n}\mathbf{Z}'\mathbf{X}}_{\text{plim}=E(\mathbf{z}_i'\mathbf{x}_i)=\mathbf{Q}_{\mathbf{zx}}}\Bigg)^{-1}\overbrace{\frac{1}{n}\mathbf{Z}'\mathbf{u}}^{\text{plim}=E(\mathbf{z}_i'u_i)}. \qquad (59)
\end{aligned}$$

Instrumental Variable Estimation
General Matrix notation of IV estimation

Thus, as long as $E(\mathbf{z}_i'u_i) = \mathbf{0}$ holds, we have a consistent estimator. The asymptotic distribution we get from

$$\begin{aligned}
\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\mathsf{IV}} - \boldsymbol{\beta}\right) &= \underbrace{\left(\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)^{-1}}_{=\mathbf{Q}_{\mathbf{zx}}^{-1}} \underbrace{\sqrt{n}\left(\frac{1}{n}\mathbf{Z}'\mathbf{u}\right)}_{\xrightarrow{d} N\left(0,\text{Var}\left(\mathbf{z}_i'u_i\right)\right)} \\
\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\mathsf{IV}} - \boldsymbol{\beta}\right) &\xrightarrow{d} \mathbf{Q}_{\mathbf{zx}}^{-1} N\left(0, \text{Var}\left(\mathbf{z}_i'u_i\right)\right) \\
\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\mathsf{IV}} - \boldsymbol{\beta}\right) &\xrightarrow{d} N\left(0, \mathbf{Q}_{\mathbf{zx}}^{-1}\text{Var}\left(\mathbf{z}_i'u_i\right)\mathbf{Q}_{\mathbf{zx}}^{-1}\right) \qquad (60)
\end{aligned}$$

in general and if we want to make an independence assumption about the relationship between $\mathbf{z}_i$ and u_i as in the standard OLS case without heteroskedasticity, we get

$$\begin{aligned}\text{Var}\left(\mathbf{z}_i' u_i\right) &= E[\left(\mathbf{z}_i' u_i\right)\left(\mathbf{z}_i' u_i\right)'] \\ &= E[\mathbf{z}_i' u_i u_i \mathbf{z}_i] \\ &= \sigma_u^2 E(\mathbf{z}_i' \mathbf{z}_i)\end{aligned}$$

which can be estimated by $\hat{\sigma}_u^2 \frac{1}{n}\mathbf{Z}'\mathbf{Z}$, where $\hat{\sigma}_u^2$ is computed from the IV residuals

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{\text{IV}}.$$

Note that $\mathbf{Q}_{\mathbf{zx}}$ is the cross product moment matrix of $\mathbf{z}_i$ and $\mathbf{x}_i$ and therefore measure how much $\mathbf{z}_i$ and $\mathbf{x}_i$ co-vary (in a matrix sense).

When does IV estimation break down?

This is an important question that naturally arises with an estimator.

There are two important conditions that need to be met for IV estimation:

1) if we use z_i as an instrument for x_i, then z_i must be a valid instrument, such that $E(z_i u_i) = 0$ holds $\Rightarrow z_i$ is exogenous. This is known as **Instrument Validity** in the literature.

Note that if 1) does not hold, then there is no point in using IV estimation in the first place.

Also, as with the OLS scenario, if only one instrument is used for one endogenous regressor, then there is no way of testing for the validity of the instrument.

When we have more instrumental variables than regressors needing instruments (overidentified system), then we will be able to perform a test known as an **overidentification test**.

Failure of IV estimator

When does IV estimation break down?

This is a test that examines the validity (or exogeneity) of the **excess moment** conditions.

2) We need z_i to be a **relevant instrument**, such that $\text{Cov}(z_i, x_i) \neq 0$ or $\text{Corr}(z_i, x_i) \neq 0$. In fact, we can see from the variance of $\hat{\beta}_1^{\text{IV}}$ that we would actually want $\text{Corr}(z_i, x_i)$ to be as close to one (in absolute value) as possible, so that $\text{Corr}(z_i, x_i)^2$ is close to one.

Recall that

$$\text{Var}(\hat{\beta}_1^{\text{IV}}) = \frac{1}{\text{Corr}(x_i z_i)^2} \underbrace{\left(\frac{\hat{\sigma}^2_{u,\text{IV}}}{\sum_{i=}^{n}(x_i - \bar{x})^2} \right)}_{\text{looks like OLS Variance}}. \tag{61}$$

Looking at the equation in (61) it is clear that there are two terms that influence the variance of $\hat{\beta}_1^{\text{IV}}$.

Failure of IV estimator

When does IV estimation break down?

1) $\left(\frac{\hat{\sigma}^2_{u,\text{IV}}}{\sum_{i=}^{n}(x_i - \bar{x})^2} \right)$
2) $\frac{1}{\text{Corr}(x_i z_i)^2}$

Now, the first term looks like the estimator of the standard OLS variance of $\hat{\beta}_1$, but the important difference here is that $\hat{\sigma}^2_{u,\text{IV}}$ is computed from

$$\hat{u}_i = y_i - \hat{\beta}_0^{\text{IV}} - \hat{\beta}_1^{\text{IV}} x_i$$

and not from the usual OLS one which is

$$\hat{u}_i = y_i - \hat{\beta}_0^{\text{OLS}} - \hat{\beta}_1^{\text{OLS}} x_i$$

where evidently $\hat{\beta}_i^{\text{OLS}} \neq \hat{\beta}_i^{\text{IV}}$ for $i = 0, 1$, in general. The second term is just the squared correlation between x_i and z_i.

Let us assume that for simplicity of exposition that $\hat{\sigma}^2_{u,\mathrm{IV}}$ is equal to our OLS variance $\hat{\sigma}^2_{u,\mathrm{OLS}}$. If that is the case, then the only difference between $\mathrm{Var}(\hat{\beta}_1^{\mathrm{IV}})$ and $\mathrm{Var}(\hat{\beta}_1^{\mathrm{OLS}})$ comes from $\mathrm{Corr}(x_i z_i)$.

To see then how $\mathrm{Corr}(x_i z_i)$ influences the variance of the IV estimator, let us express it as:

$$\mathrm{Var}(\hat{\beta}_1^{\mathrm{IV}}) = \frac{\mathrm{Var}(\hat{\beta}_1^{\mathrm{OLS}})}{\mathrm{Corr}(x_i z_i)^2}. \tag{62}$$

Clearly, having $\mathrm{Corr}(x_i z_i)^2 = 0$ is not going to be a good thing.

When choosing an instrument z_i for x_i it is therefore crucial that the correlation between these two is as large as possible (ie, close to 1) to have a "precise" estimate of our $\hat{\beta}_1^{\mathrm{IV}}$.

For instance, consider the following correlation values for x_i and z_i.

$\mathrm{Corr}(x_i z_i)$	0.1	0.2	0.3	0.5	0.7
$1/\mathrm{Corr}(x_i z_i)^2$	100	25	≈ 11	4	≈ 2

So even with a rather strong instrument such that $\mathrm{Corr}(x_i, z_i) = 0.7$ (70%), the variance is about **double** the size of the OLS estimator.

In the context of the matrix version of the IV estimator shown in (58) and the large sample distribution shown in (60) we need $E(\mathbf{z}_i'\mathbf{x}_i) = \mathrm{plim}\ \frac{1}{n}\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{\mathbf{zx}}$ to be of full column rank, so that the inverse will exist.

- this is analogous to the condition of $\mathbf{Q}_{\mathbf{xx}}$ being of full column rank, so is a necessary condition for estimation.

Now it should be clear that, because x_i and z_i are random variables, $\mathrm{Corr}(x_i, z_i)$ will also be a random variable, so that when we try to gauge how weak (or strong) an instrument is (that is, its validity) we will need to use a statistical test on the "sample correlation" that we compute.

- in the simplest case this will involve running a regression of x_i on a constant and z_i and then "test" if the coefficient on z_i is statistically different from 0.
- the rule of thumb is that if the $t-$statistic on the z_i instrument variable is less than $\sqrt{10}$, then the instrument is considered to be weak and the coefficients and it standard errors are most likely useless for economic analysis.

Research in the weak instrument literature is still quite young and in progress. What is known is that when instruments are weak,

- not only are the standard errors large as evident from the simple example above
- but the distribution of the β_1^{IV} estimator can be highly non-normal, even if there is a very large number of observations available (ie. $30\,000$). See plot below taken from page 215 of Pagan and Robertson (1998).

Failure of IV estimator

When does IV estimation break down?

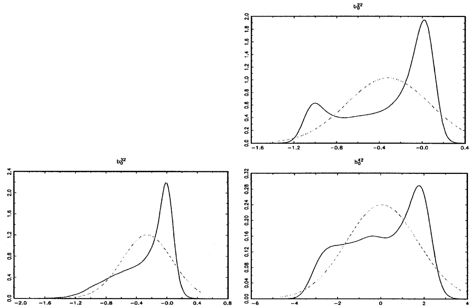


Figure 2: Large Sample distribution of IV estimator with weak instruments.

Generalised Instrumental Variable Estimator

More moment instruments than endogenous variables

More instruments than endogenous variables: Generalised IV Estimator (GIVE)

An obvious question that arises in the IV setting is what to do when there are more moment conditions than parameters to be estimated, or equivalently, when there are more instrumental variables available than endogenous regressors.

For example, suppose you have the model:

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{63}$$

where again $E(x_i u_i) \neq 0$ and there are two instrumental variables available, z_{1i} and z_{2i}, which can be used to estimate the unknown parameters.

With these two instruments, we then have a total of three moment conditions (one for the intercept term β_0) which are:

$$m_{i1} = E(u_i) = 0 \Rightarrow E(y_i - \beta_0 - \beta_1 x_i) = 0$$
$$m_{i2} = E(z_{1i} u_i) = 0 \Rightarrow E[z_{1i}(y_i - \beta_0 - \beta_1 x_i)] = 0$$
$$m_{i3} = E(z_{2i} u_i) = 0 \Rightarrow E[z_{2i}(y_i - \beta_0 - \beta_1 x_i)] = 0$$

so this translates into three equations to be solved for two unknown parameters β_0 and β_1.

This system is thus "**overidentified**", since we have more equations than unknowns.

The problem is now one of finding "**a good way**" of combining the excess of instruments/equations (or information)

- in a good way is meant in a statistical sense, translating in a consistent estimator with the smallest possible variance.
- one could of course always reduce the excess of instruments or moments by just using the exact number that we need, thus "throwing" the others away

- but intuition tells us that this is probably not a good way to go about it.

We are going to look at this problem in the context of a two step procedure known as **Two Stage Least Squares** (TSLS or 2SLS).

- general treatment of this falls under the heading of Generalised IV estimation (GIVE) and is related to Generalised Method of Moments (GMM) estimation
- the term "**Generalised**" addresses the problem of optimally "weighting" the excess of moment conditions (or instruments) without loosing any important information
- the weighting is done by looking at the information content of each moment condition (or instrument) and assigning a higher weight to those with more information

So how does GIVE work?

Recall that the moment condition for the intercept term always gives us the following simple EGE to solve

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \tag{64}$$

$$\Rightarrow \hat{\beta}_0^{\text{IV}} = \bar{y} - \hat{\beta}_1^{\text{IV}}\bar{x}.$$

So this is straight forward to solve, once $\hat{\beta}_1^{\text{IV}}$ has been found.

Now to get an estimate of β_0^{IV} with two instruments, we will first run a regression of x_i on a constant and the two instruments z_{1i} and z_{2i} to get the **predicted (or fitted) values** of x_i, that is, $\hat{x}_i$.

Then, form the EGE of the moment condition as:

$$\frac{1}{n}\sum_{i=1}^{n}[\hat{x}_i(y_i - \beta_0 - \beta_1 x_i)] = 0 \tag{65}$$

where, after plugging in the relation for $\hat{\beta}_0^{\text{IV}}$ from the EGE in (64), we get the expression for $\hat{\beta}_1^{\text{IV}}$

$$\hat{\beta}_1^{\text{IV}} = \frac{\sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}})(y_i - \bar{y})}{\sum_{i=1}^{n}(\hat{x}_i - \bar{\hat{x}})(x_i - \bar{x})}$$
$$= \frac{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})(x_i - \bar{x})} \tag{66}$$

where $\bar{\hat{x}} \equiv \bar{x}$ from OLS theory, because we always have that:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

and averaging on both sides yields

$$\begin{aligned}\bar{\hat{y}} &= \underbrace{\hat{\beta}_0}_{=\bar{y}-\hat{\beta}_1\bar{x}} + \hat{\beta}_1\bar{x} + \underbrace{\bar{\hat{u}}}_{=0 \text{ by OLS FOC}} \\ \bar{\hat{y}} &= \bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1\bar{x} \\ &= \bar{y}.\end{aligned}$$

We can see therefore, that running the first regression of x_i on z_{1i} and z_{2i} (and a constant) does the "**generalisation**" step of the GIVE estimator.

It further gives the instruments z_{1i} and z_{2i} their "optimal weights" from OLS in predicting x_i.

Generalised Instrumental Variable Estimator
More moment instruments than endogenous variables

Equivalent TSLS representation

The representation above was the GIVE representation, but we can find an equivalent representation in terms of the TSLS estimator. The sequence of operations is then:

1) regress x_i on z_{1i}, z_{2i} and a constant to get $\hat{x}_i$
2) use $\hat{x}_i$ directly in a regression of y_i and $\hat{x}_i$ and a constant, ie., estimate the equation

$$y_i = \beta_0 + \beta_1\hat{x}_i + u_i. \tag{67}$$

The $\hat{\beta}_i$, $i = 0, 1$ that we get from this TSLS procedure is the same as the GIVE estimator from above.

Asymptotic Distribution of TSLS/GIVE

To find asymptotic distribution of TSLS/GIVE estimators, again expand (66) as:

Replacing $(y_i - \bar{y}) = \beta_1(x_i - \bar{x}) + u_i$ in (66) yields

$$\begin{aligned}
\hat{\beta}_1^{\text{IV}} &= \frac{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})(\beta_1(x_i - \bar{x}) + u_i)}{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})(x_i - \bar{x})} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})u_i}{\sum_{i=1}^{n}(\hat{x}_i - \bar{x})(x_i - \bar{x})} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}(\hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i})u_i}{\sum_{i=1}^{n}\hat{x}_i x_i - \bar{x}^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}(\hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i})u_i}{\sum_{i=1}^{n}(\hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i})x_i - \bar{x}^2} \qquad (68)
\end{aligned}$$

where $\hat{x}_i = \hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i}$, ie., the fitted value from the first stage regression.

We again see that the individual components of (68) matter for consistency of TSLS/GIVE estimators, ie., z_{1i} and z_{2i} must be uncorrelated with u_i in numerator.

In the denominator we have the covariance between $\hat{x}_i$ and x_i, which will be related to the covariance between the individual components in the fitted values $\hat{x}_i = \hat{\alpha}_0 - \hat{\alpha}_1 z_{1i} - \hat{\alpha}_2 z_{2i}$, that is, the instruments z_{1i} and z_{2i}.

An important point to note here is that the **variance of the residuals** is computed as

$$\sigma_{u,IV}^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$$

$$u_i = y_i - \hat{\beta}_1^{\text{IV}} - \hat{\beta}_1^{\text{IV}} x_i$$

and not $u_i = y_i - \hat{\beta}_1^{\text{IV}} - \hat{\beta}_1^{\text{IV}} \hat{x}_i$, so one has to be careful when constructing the residual variance by hand and most econometric software packages do this properly.

A few extra comments:

We outlined TSLS initially as an IV problem within the IV (or GIVE) framework was to highlight the importance of having strong instruments to have a "useful" estimator.

$\Rightarrow$ need to have high R^2 values of the first stage step in TSLS.

The hardest part in TSLS/GIVE is finding good instruments. Where do instruments come from?

- no quick answer to that. we need to use our understanding of economics and economic theory
- instruments have to be relevant as well as valid
- there are many examples in the literature and textbooks were later analysis show that instruments were **actually quite poor**, so **IV estimation results are useless**.

General Matrix version of TSLS/GIVE

The asymptotics are much easier to handle in a matrix set up. So let us look at TSLS/GIVE in matrix notation.

Let $\mathbf{X}$ be an $(n \times k)$ matrix of endogenous regressors needing instruments and $\mathbf{Z}$ an $(n \times m)$ matrix of instruments, where $m > k$ and n is the sample size.

The **first stage regression** in TSLS is:

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{E} \tag{69}$$

where $\mathbf{E}$ is an $(n \times k)$ matrix of residuals and $\boldsymbol{\alpha}$ is an $(m \times k)$ matrix of parameters that links the m instruments to the k endogenous regressors.

We get fitted values for (69) from

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\boldsymbol{\alpha}} \tag{70}$$

Generalised Instrumental Variable Estimator

More moment instruments than endogenous variables

$$\hat{\boldsymbol{\alpha}} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X}). \tag{71}$$

Now

$$\begin{aligned} \hat{\mathbf{X}} &= \boldsymbol{Z}\hat{\boldsymbol{\alpha}} \\ \hat{\mathbf{X}} &= \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X}) \\ &= \mathbf{P_z}\mathbf{X} \end{aligned}$$

where $\mathbf{P_z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is known as the **hat** or **projection matrix**, which has the properties

$$\mathbf{P_z}' = \mathbf{P_z}$$
$$\mathbf{P_z}\mathbf{P_z}' = \mathbf{P_z}$$
$$\mathbf{P_z}\mathbf{Z} = \mathbf{Z}.$$

Generalised Instrumental Variable Estimator

More moment instruments than endogenous variables

Given $\hat{\mathbf{X}}$, the **second regression** in TSLS would "regress" $\mathbf{y}$ on $\hat{\mathbf{X}}$ to yield

$$\hat{\boldsymbol{\beta}}^{\text{TSLS}} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}.$$

But

$$\begin{aligned} (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y} &= ([\mathbf{P_z}\mathbf{X}]'[\mathbf{P_z}\mathbf{X}])^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P_z}'\mathbf{P_z}\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{P_z}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ \Leftrightarrow \hat{\boldsymbol{\beta}}^{\text{TSLS}} &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \end{aligned} \tag{72}$$

The **GIVE estimator** sets up a "Generalised" moment condition of the form

$$\begin{aligned} \hat{\mathbf{X}}'\mathbf{u} &= \mathbf{0} \\ \Leftrightarrow \hat{\mathbf{X}}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{0} \\ \Leftrightarrow \hat{\boldsymbol{\beta}}^{\mathsf{GIVE}} &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}. \end{aligned} \tag{73}$$

Comparing (72) and (73) it is clear that the two solve the same optimality problem and hence yield the same estimator.

Generalised Instrumental Variable Estimator
More moment instruments than endogenous variables

Asymptotic Distribution of TSLS/GIVE in matrix form

Doing the standard expansion on (72) or (73), we get

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\mathsf{TSLS}} &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= \boldsymbol{\beta} + (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{u}. \end{aligned} \tag{74}$$

The relation in (74) implies again that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\mathsf{TSLS}} - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\frac{\sqrt{n}}{n}\hat{\mathbf{X}}'\mathbf{u} \tag{75}$$

where $\frac{1}{n}\hat{\mathbf{X}}'\mathbf{X}$ will go to $E(\hat{\mathbf{x}}_i'\mathbf{x}_i) = \underset{(k\times k)}{\mathbf{Q}_{\hat{\mathbf{x}}\mathbf{x}}}$ (some constant) in large samples and $\frac{\sqrt{n}}{n}\hat{\mathbf{X}}'\mathbf{u}$ will converge to a Normal random variable with mean $\underset{(k\times 1)}{E(\hat{\mathbf{x}}_i'u_i)} = \mathbf{0}$ (if assumptions hold) and variance

$$\begin{aligned}\text{Var}\left(\frac{\sqrt{n}}{n}\hat{\mathbf{X}}'\mathbf{u}\right) &= \text{Var}\left(\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\hat{\mathbf{x}}_i'u_i\right)\\ &= \frac{1}{n}\text{Var}\left(\sum_{i=1}^{n}\hat{\mathbf{x}}_i'u_i\right)\\ &= \frac{1}{n}n\text{Var}\left(\hat{\mathbf{x}}_i'u_i\right)\end{aligned}$$

$$\begin{aligned}&= E(\hat{\mathbf{x}}_i'u_iu_i\hat{\mathbf{x}}_i)\\ &= \sigma_u^2E(\hat{\mathbf{x}}_i'\hat{\mathbf{x}}_i)\end{aligned}$$

where the last result follows from the Homoskedasticity assumption of the residuals u_i.

Given these results, we can than take (75) and then see that the (asymptotic) distribution of

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{TSLS}} - \boldsymbol{\beta}\right) &\sim \mathbf{Q}_{\hat{\mathbf{x}}\mathbf{x}}^{-1} \times \underbrace{N\left(0, \sigma_u^2E(\hat{\mathbf{x}}_i'\hat{\mathbf{x}}_i)\right)}_{\text{Normal RV}}\\ &\sim N\left(0, \mathbf{Q}_{\hat{\mathbf{x}}\mathbf{x}}^{-1}\sigma_u^2E(\hat{\mathbf{x}}_i'\hat{\mathbf{x}}_i)\mathbf{Q}_{\hat{\mathbf{x}}\mathbf{x}}^{-1\prime}\right).\end{aligned}$$

Taking the fact from above, this simplifies to

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}^{\text{TSLS}} - \boldsymbol{\beta}\right) \sim N\left(0, \sigma_u^2\mathbf{Q}_{\hat{\mathbf{x}}\mathbf{x}}^{-1}\right)$$

where $\mathrm{Var}(\hat{\boldsymbol{\beta}}^{\mathsf{TSLS}})$ can be estimated from

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{\mathsf{TSLS}}) &= \frac{1}{n}\hat{\sigma}_u^2\left(\frac{1}{n}\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\\ &= \hat{\sigma}_u^2\left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\\ &= \hat{\sigma}_u^2\left(\mathbf{X}'\mathbf{P}_z\mathbf{X}\right)^{-1}\end{aligned}$$

with $\hat{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/n$ and $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ and not $\mathbf{y} - \hat{\mathbf{X}}\hat{\boldsymbol{\beta}}$.

Allowing for Exogenous variables

Suppose now that we have a more complicated regression model, taking the form

$$y_i = \beta_0 + \underbrace{\beta_1 x_{1i} + \ldots + \beta_B x_{Bi}}_{B \text{ endogenous variables}} + \underbrace{\gamma_1 w_{1i} + \ldots + \gamma_G w_{Gi}}_{G \text{ truly exogenous variables}} + u_i \tag{76}$$

- $\{w_{gi}\}_{g=1}^{G}$ are G truly exogenous variables that do **NOT** need an instrument
- $\{x_{bi}\}_{b=1}^{B}$ are B endogenous variables for which we need instruments, as the condition $E(x_{bi}u_i) = 0$ is violated.

Let $\{z_{li}\}_{l=1}^{L}$ be L available instrumental variables for $\{x_{bi}\}_{b=1}^{B}$.

To estimate the relation in (76) we need $L \geq B$ for the system to be identified. Recall from before that we had

a) a just identified case when $L = B$

b) an over-identified case when $L > B$.

To estimate **all** the γ and β parameters in (76) using TSLS (or GIVE) we need to again replace $\{x_{bi}\}_{b=1}^{B}$ with their fitted values $\{\hat{x}_{bi}\}_{b=1}^{B}$ and then run the regression:

$$y_i = \beta_0 + \beta_1 \hat{x}_{1i} + \ldots + \beta_B \hat{x}_{Bi} + \gamma_1 w_{1i} + \ldots + \gamma_G w_{Gi} + u_i. \tag{77}$$

What is different here to the previous treatment is how we obtain the $\{\hat{x}_{bi}\}_{b=1}^{B}$.

These will need to be estimated by regressing each of the endogenous variables $\{x_{bi}\}_{b=1}^{B}$ not only on all the instruments as before, but also on **all the truly exogenous regressors** $\{w_{gi}\}_{g=1}^{G}$.

The fitted values $\{\hat{x}_{bi}\}_{b=1}^{B}$ are then obtained as:

$$\hat{x}_{bi} = \hat{\theta}_{0b} + \underbrace{\hat{\theta}_{1i} z_{1i} + \ldots + \hat{\theta}_{Li} z_{1i}}_{L \text{ instruments for } x} + \underbrace{\hat{\alpha}_1 w_{1i} + \ldots + \hat{\alpha}_{Gi} w_{Gi}}_{\text{G truly exogenous variables}} \tag{78}$$

for all $b = 1, \ldots, B$, that is, for each endogenous variable.

Note here that we can again write this in the same matrix set up as before, but we now we need to add the $\mathbf{W}$ matrix of truly exogenous variables to the $\mathbf{Z}$ and $\mathbf{X}$ matrices, but the math is the same.

The point of this is to control for the effect of truly exogenous variables in the first stage regression when forming the predictions (or fitted values) of the endogenous $\{x_{bi}\}_{b=1}^{B}$.

This is important as the $x's$ and the $w's$ will in general be correlated and an allowance for this needs to be made.

A regression of the $\{x_{bi}\}_{b=1}^{B}$ on all exogenous as well as instrumental variables is often referred to as the **reduced form** of a model.

- terminology comes from the literature on simultaneous equation models
- will arrive at reduced from after "substituting out" the endogenous variables.

Testing the strength of the instruments

For **GIVE or TSLS to work well**, we saw earlier that we need two conditions to hold

i) Instrument validity: $E(z_i u_i) = 0$

ii) Instrument relevance: $\text{Corr}(x_i z_i) \neq 0$

If we assume for the moment that $i)$ holds, we can look at $ii)$ more carefully by examining how strongly our instruments z_i are correlated with the endogenous variables x_i. This is called **instrument strength**.

How do we test the strength of the instruments?

In the simplest example where we only have one endogenous variable x_i and one instrument z_i and no other "truly exogenous" regressor w_i

- run the "*first stage regression*" of x_i on a constant and z_i
- do a $t-$test on the coefficient of z_i

that is: estimate

$$x_i = \theta_0 + \theta_1 z_i + \varepsilon_i \tag{79}$$

and test the null hypothesis $\mathcal{H}_0 : \theta_1 = 0$ using a standard $t-$test

$$t - \text{statistic} = \frac{\hat{\theta}_1 - 0}{se(\hat{\theta}_1)} \tag{80}$$

The "*rule of thumb*" is (as mentioned before) that the size of the **$t-$statistic should be greater that $\sqrt{10}$** for the **instrument not to be considered weak**.

Clearly, the bigger the $t-$statistic, the stronger the instrument.

When we have more instruments than regressors $(\dim(z_i) > \dim(x_i))$ as well as other truly exogenous variables in the model, it is important to account (or control) for the influence of the other exogenous regressors on the model.

This means that when we test for the strength of the instruments, we will need to include the truly exogenous regressors w_i as controls in the "*first stage regression*".

For the b^{th} endogenous variable x_{bi} that we need to instrument, we would run the "*first stage regression*":

$$x_{bi} = \theta_{0b} + \underbrace{\theta_{1i}z_{1i} + \ldots + \theta_{Li}z_{1i}}_{L \text{ instruments for } x} + \underbrace{\alpha_1 w_{1i} + \ldots + \alpha_{Gi}w_{Gi}}_{\text{G truly exogenous variables}} + \varepsilon_i \tag{81}$$

for all $b = 1, \ldots, B$, that is, for each endogenous variable.

In the (big) "*first stage regression*" in (81) we explicitly account for the influence of the truly exogenous w_i variables on x_i when examining the strength of the instruments z_i.

This is important, because if most of the variation in the endogenous variables x_i is explained by the truly exogenous variables w_i, then the instruments are useless and we will not be able to figure out what the impact is of x_i on y_i in (76)

- most of the variation will be related to w_i but we also include w_i in the regression in (76)
- not enough "information" to determine cause and effect from x_i and w_i.

To test the **strength of the instruments** in the more general set up with many instruments as well as truly exogenous variables we need to run the "*first stage regression*" (81) and then test the join null hypothesis (with an $F-$test)

$$\mathcal{H}_0: \ \theta_1 = \theta_2 = \ldots = \theta_L = 0 \tag{82}$$

against the alternative that at least one instrument is "*significant*" in the regression.

For instruments not to be considered weak, we need the $F-$statistic from (82) to be **greater than 10**.

The value of 10 was determined by Stock and Yogo (2005) (see the paper for details).

Note that we are only testing θ and not α in the $F-$test in (82).

Testing for Endogeneity

Recall that $E(x_i u_i) = 0$ cannot be tested directly using the OLS residuals $\hat{u}_i$ in place of u_i. Note that this value is always set equal 0 by the OLS first order conditions

Because OLS is more efficient than TSLS/GIVE when endogeneity is not a problem, we would still like to test somehow for the endogeneity of the regressor x_i to determine whether we need to use IV estimation in the first place.

Intuition tells us that we could determine if endogeneity is a problem by computing $\hat{\beta}_1^{\text{OLS}}$ and $\hat{\beta}_1^{\text{TSLS}}$ and then looking at how similar these values are to one another.

Since we are dealing with **random variables** here, we would like to see how different $\hat{\beta}_1^{\text{OLS}}$ and $\hat{\beta}_1^{\text{TSLS}}$ are, but taking into account the random variation of these two RVs.

Hauseman test for Endogeneity

The Hausman (1978) test for endogeneity does exactly that. It is defined as:

$$h - \text{statistic} = \frac{\left(\hat{\beta}_1^{\text{TSLS}} - \hat{\beta}_1^{\text{OLS}}\right)^2}{\left[\text{Var}(\hat{\beta}_1^{\text{TSLS}}) - \text{Var}(\hat{\beta}_1^{\text{OLS}})\right]}. \tag{83}$$

If we define $q = \left(\hat{\beta}_1^{\text{TSLS}} - \hat{\beta}_1^{\text{OLS}}\right)$, then (83) looks like

$$h - \text{statistic} = \frac{q^2}{\text{Var}(q)} \tag{84}$$

which looks like a standard $t-$test, squared and under the null hypothesis of $q = 0$.

Recall that a $t-$statistic goes towards a $z-$ statistic (where z is standard normal) as the sample size and thus the degrees of freedom go to infinity.

Also, if you take the square of the $z-$statistic, you will get a new RV that will be a χ^2 RV and if you take the ratio of two χ^2 RVs divided by their degrees of freedom, you will end up with an F RV.

The Hausman test looks at how close the $\hat{\beta}_1^{\text{OLS}}$ is to $\hat{\beta}_1^{\text{TSLS}}$ and adjusts for the variation in the series by dividing by the variance.

A word of caution on how $\text{Var}(q)$ looks.

Recall that we always have the rule:

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y). \tag{85}$$

The result of the variance of the difference being equal to the difference of the variances is due to:

$$\text{Cov}(\hat{\beta}_1^{\text{TSLS}}, \hat{\beta}_1^{\text{OLS}}) = \text{Var}(\hat{\beta}_1^{\text{OLS}}) \tag{86}$$

so that we end up with

$$\begin{aligned}
\text{Var}(q) &= \text{Var}\left(\hat{\beta}_1^{\text{TSLS}} - \hat{\beta}_1^{\text{OLS}}\right) \\
&= \text{Var}(\hat{\beta}_1^{\text{TSLS}}) + \text{Var}(\hat{\beta}_1^{\text{OLS}}) - 2\text{Cov}(\hat{\beta}_1^{\text{TSLS}}, \hat{\beta}_1^{\text{OLS}}) \qquad (87) \\
&= \text{Var}(\hat{\beta}_1^{\text{TSLS}}) + \text{Var}(\hat{\beta}_1^{\text{OLS}}) - 2\text{Var}(\hat{\beta}_1^{\text{OLS}}) \\
&= \text{Var}(\hat{\beta}_1^{\text{TSLS}}) - \text{Var}(\hat{\beta}_1^{\text{OLS}}).
\end{aligned}$$

Hausman's Result on Covariance

This comes from Lemma 2.1 in Hausman (1978) (see page 1253) which says that if you have two consistent estimators (both are consistent under the null hypothesis of no endogeneity, even the OLS one), of which one reaches the Cramer-Rao lower bound (ie., is the most efficient estimator), then the covariance between the efficient estimator and the difference between the efficient and inefficient estimator is 0.

With simple algebra you can show that this translates into the relation given in (86).

A convenient way to implement the Hausman test is to do the following:

Suppose you have the model

$$y_i = \beta_0 + \beta_1 x_i + \gamma_1 w_i + u_i \qquad (88)$$

where w_i is a truly exogenous variable, and you want to test if x_i is correlated with u_i and thus endogenous.

Suppose you have two instruments z_{1i} and z_{2i}. Then you would run the (reduced form) regression:

$$x_i = \theta_0 + \theta_1 z_{1i} + \theta_2 z_{2i} + \alpha_1 w_i + \varepsilon_i. \qquad (89)$$

and compute the predicted values $\hat{x}_i$.

Note that once you have run the regression in (89) you can decompose x_i as follows:

$$x_i = \hat{x}_i + \hat{\varepsilon}_i \tag{90}$$

where $\hat{\varepsilon}_i$ is the fitted regression error or residual that you get from OLS, and $\hat{x}_i$ and $\hat{\varepsilon}_i$ are, by **construction** of the OLS residuals and predicted values, **uncorrelated**.

Now, since

$$\begin{aligned}\hat{x}_i &\equiv E(x_i|z_{1i}, z_{2i}, w_i) \\ &= \hat{\theta}_0 + \hat{\theta}_1 z_{1i} + \hat{\theta}_2 z_{2i} + \hat{\alpha}_1 w_i\end{aligned}$$

by definition, it captures the correlation between the endogenous regressor x_i and the instruments z_{1i}, z_{2i} as well as w_i (the truly exogenous variable).

Instrumental Variable Estimation
Testing for endogeneity

The fitted residual $\hat{\varepsilon}_i$ can thus be thought of as the left-over bit of the endogenous variable x_i after taking out the effect of z_{1i}, z_{2i} and w_i which **are not correlated** with u_i in (88) above.

Thus, the left-over bit $\hat{\varepsilon}_i$ has to contain that part of x_i that is correlated with u_i. That is, if $E(x_i u_i) \neq 0$, then $E(u_i \hat{\varepsilon}_i) \neq 0$, because $\hat{\varepsilon}_i = (x_i - \hat{x}_i)$, so that

$$\begin{aligned}E(u_i \hat{\varepsilon}_i) &= E[u_i(x_i - \hat{x}_i)] \\ &= E(u_i x_i) - \underbrace{E(u_i \hat{x}_i)}_{=0 \text{ by construction/assumption}}\end{aligned} \tag{91}$$

$E(u_i \hat{x}_i)$ follows because z_{1i}, z_{2i} and w_i are by construction of the problem or our definition of instruments and truly exogenous variables uncorrelated with u_i.

An intuitive implementation of the Hausman test is then to take the $\hat{\varepsilon}_i$ from the fitted regression in (89) and plug them into the regression in (88) for y_i as an extra regressor.

That is, we form

$$y_i = \beta_0 + \beta_1 x_i + \gamma_1 w_i + \delta\hat{\varepsilon}_i + u_i \tag{92}$$

and test the null hypothesis $\mathcal{H}_0 : \delta = 0$ (which is the null of no correlation between x_i and u_i and thus no endogeneity problem).

If we **reject the null hypothesis** that $\delta = 0$, then we have an **endogeneity problem**.

Note that the relation in (92) is a convenient way to implement the test because ultimately we want to test if $\hat{\beta}_1^{\mathsf{TSLS}} = \hat{\beta}_1^{\mathsf{OLS}}$ and when we expand $\hat{\varepsilon}_i = (x_i - \hat{x}_i)$ we would get $\beta_1 x_i$ as well as $-\delta\hat{x}_i$ in the regression, where the latter is the TSLS and the former the OLS estimate.

If we have more than one endogenous variable to test for, say x_{1i} as well as x_{2i}, then we would simply get predicted values from two separate reduced form regressions, as in (89) for both x_{1i}, x_{2i} and obtain corresponding (fitted) regression errors $\hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_{2i}$.

Including $\hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_{2i}$ in the regression of y_i on a constant, x_i, as well as w_i and using an $F-$test on the coefficients of $\hat{\varepsilon}_{1i}$ and $\hat{\varepsilon}_{2i}$ would then tell us whether both suspected endogenous variables are actually correlated with u_i.

So again, if the null hypothesis $H_0 : \delta_1 = \delta_2 = 0$ is rejected, then we have an endogeneity problem with respect to either x_{1i}, x_{2i} or both.

Testing instrument Validity (over-identification testing)

Recall that we use B to denote the No. of endogenous regressors $(\dim(x_i))$ and L to denote the No. of instruments $(\dim(z_i))$.

One problem with testing the validity of instruments $(E(z_i u_i) = 0)$ is that one cannot perform a test when $L = B$, that is, when we have a just identified case.

Nevertheless, it is possible to test whether an **excess of instruments is valid**, but this is only possible if and only if $L > B$. Testing the excess of instruments can then help us try to weed out the instruments that are not valid, ie, find those where $E(z_i u_i) \neq 0$.

Recall that if the instruments are not valid, then TSLS is biased just as OLS is, but at the same time it can be highly inefficient. So there are two potential downsides and it is important to work out how likely it is that an excess of instrument(s) is not valid.

To test the excess instruments for validity, we can again look at a regression approach to implement this. Suppose we are interested in the model

$$y_i = \beta_0 + \beta_1 x_i + \gamma_1 w_i + u_i \tag{93}$$

where $E(u_i x_i) \neq 0$ and we have instruments z_{1i} and z_{2i}, but we are not sure if they are valid instruments.

Then, to implement an over-identification test we proceed as follows.

1) Construct
$$\hat{u}_i^{\text{TSLS}} = y_i - \hat{\beta}_0^{\text{TSLS}} - \hat{\beta}_1^{\text{TSLS}} x_i - \hat{\gamma}_1^{\text{TSLS}} w_i$$
where $\hat{\beta}_j^{\text{TSLS}}$ for all $j = 0, 1$ and $\hat{\gamma}_1^{\text{TSLS}}$ are the two stage least squares estimates of the parameters in (93) (obtained from a normal TSLS regression)
2) Regress $\hat{u}_i^{\text{TSLS}}$ on z_{1i}, z_{2i} and w_i (all instruments and truly exogenous regressors), that is:
$$\hat{u}_i^{\text{TSLS}} = \phi_0 + \phi_1 z_{1i} + \phi_2 z_{2i} + \phi_3 \omega_i + \nu_i \tag{94}$$
with ν_i being some regression error of this equation.

3) Compute the $F-$statistic of the regression restriction that $\hat{\phi}_1$ and $\hat{\phi}_2$ are equal to zero, ie., test the null hypothesis $\mathcal{H}_0 : \phi_1 = \phi_2 = 0$. The over-identifying restriction test statistic, commonly referred to as the $J-$test, is then constructed as

$$J - \text{statistic} = L \times F - \text{statistic} \tag{95}$$

where $L =$ No. of instruments, which in this case is equal to 2.

The asymptotic distribution of the $J-$statistic is a $\chi^2_{(L-B)}$ where $(L - B)$ is the degrees of freedom.

Note here that the degrees of freedom in the $J-$test is equal to the **excess of the number of instruments** $(L - B)$.

To illustrate this, note that in (94) we have two instruments z_{1i} and z_{2i} for one endogenous variable x_i so the null hypothesis for the $F-$test is $\phi_1 = \phi_2 = 0$ against the alternative of at least one not being zero.

The $J-$test is thus computed as $2 \times F-$statistic but the critical values come from a χ^2 distribution with $(L - B) = (2 - 1)$ degrees of freedom. This is important to keep in mind.

Note here once again that the over-identification test (the $J-$test) can only be implemented when $L > B$, that is, when we have more instruments than endogenous variables (an over-identified system).

We **cannot implement** the $J-$test when $L = B$.

So what happens when we have $J-$statistic $> \chi^2_{(L-B)}$ critical value?

Then we would reject the null hypothesis $\mathcal{H}_0 : \phi_1 = \phi_2 = 0$ which translates into rejecting the validity of the $(L - B) = (2 - 1) = 1$ surplus instruments, in favour of the alternative.

The alternative means here that either z_{1i} or z_{2i} is not a valid instrument, as it is still correlated with u_i in (93), ie., the error term of the economic model of interest that we wanted to estimate with OLS.

One obvious problem with the $J-$test is that it does not tell which instrument is invalid. So we will again need to use our intuition about the economic model and an understanding of the variable to figure our which instrument the trouble maker is.

Simultaneous Equation Models: Background

Simultaneous Equation Models (SEMs) are in fact a much broader class of models and we are only going to look at a few specific cases, mainly to illustrate the basics of this literature.

Greene (2011) has in fact two chapters on this topic and in many ways we can think of the Vector-Autoregression (VAR) literature in Macro-economics also falling into this broad class.

The origins of SEMs, and standard examples always come back to the classic Demand-Supply curve relations in economics. To understand the SEM literature and to outline the original problem in estimating Demand-Supply curve relations, let us consider such a simple scenario.

The Demand-Supply curve relation

Suppose you have the following Demand-Supply relation

$$\text{Demand: } Q_i = \alpha_1 P_i + \alpha_2 x_i + u_i^D \tag{96a}$$

$$\text{Supply: } Q_i = \beta_1 P_i + u_i^S. \tag{96b}$$

Both, Q_i and P_i are endogenous variables here, where u_i^D and u_i^S are "*shocks*" to demand and supply, and x_i is some exogenous variable.

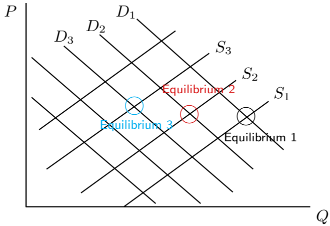
Note that (96b) could also have been written down as

$$\text{Inverse Supply: } P_i = b_1 P_i + \epsilon_i^S \tag{97}$$

that is, with P_i on the right hand side to make the "endogeneity" of P_i more obvious.

This relation is known as the Inverse Supply equation in micro-economics.

Because the system in (96) is assumed to be an equilibrium relationship, we only ever observe P_i and Q_i when this equilibrium relationship holds. See figure below.



For a given supply curve S_1, prices P (are assumed to) adjust so that "*markets clear*" and equilibrium is reached.

This occurs by **moving along** the given supply and demand relations until an intersection point is reached, the equilibrium point.

Recall from your study of microeconomics, that one always moves along the demand and supply relations, when there are changes in the price (or quantity) of the product, and that the demand and supply relations **shift whenever a factor other than price (or quantity) change**.

Whenever both supply and demand shift at the same time, new equilibrium points will be recorded and these will create a scatter of equilibrium values that will look like a "*random scatter*".

Thus trying to use these "scatter" data points to estimate the α and β parameters in (96) will be unsuccessful.

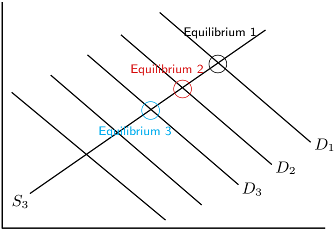
What we clearly need to do is to fix of one of the curves and induce shifts in the other.

From, the relation in (96), we can see that if x_i changes (the exogenous variable), then this will result in shifts in the demand curve. The supply curve will remain fixed as it is not a function of x_i.

These shifts in the demand curve due to changes in x_i which will again result in new equilibrium points that we can observe and record would make it possible to "identify" the supply curve and hence "estimate" the parameter β.

We can illustrate this first with a diagram (see the figure below). Note here again that we know that only the demand curve will shift when x_i changes.

The supply curve remains fixed as it is not a function of x_i.



Forming the "Reduced Form" Equations

Let us now look at what happens algebraically when we have this classic demand-supply curve model with one exogenous variable x_i in the demand equation as in (96).

To clear up a bit of the terminology used in SEM, the original formulation in (96) is called the **structural model**, containing **structural parameters** α_2, α_2 and β, which are the actual **parameters of interest**.

The **reduced form model** represents all **endogenous variables** as a function of all the **exogenous variables** only.

In the TSLS examples discussed so far, the exogenous variables were the z_i (the instruments) and the w_i (truly exogenous) variables.

The **reduced form equations** for the SEM in (96) can be arrived at by setting the demand and supply equations equal to one another, yielding

$$\begin{aligned} \text{Demand } Q &= \text{Supply } Q \\ \beta_1 P_i + u_i^S &= \alpha_1 P_i + \alpha_2 x_i + u_i^D \\ (\beta_1 - \alpha_1) P_i &= \alpha_2 x_i + u_i^D - u_i^S \\ P_i &= \frac{\alpha_2}{(\beta_1 - \alpha_1)} x_i + \frac{\left(u_i^D - u_i^S\right)}{(\beta_1 - \alpha_1)} \end{aligned} \tag{98}$$

so the reduced form for the endogenous variable price P_i is:

$$P_i = \pi_1 x_i + e_{1i} \tag{99}$$

where

$$\pi_1 = \frac{\alpha_2}{(\beta_1 - \alpha_1)} \text{ and } e_{1i} = \frac{\left(u_i^D - u_i^S\right)}{(\beta_1 - \alpha_1)}.$$

Similarly, taking $P_i = \pi_1 x_i + e_{1i}$ in (99) and plugging it into the supply equation in (96b) yields:

$$\begin{aligned} Q_i &= \beta_1 P_i + u_i^S \\ &= \beta_1 \left[\pi_1 x_i + e_{1i}\right] + u_i^S \\ &= \underbrace{\frac{\beta_1 \alpha_2}{(\beta_1 - \alpha_1)}}_{=\pi_2} x_i + \beta_1 \frac{\left(u_i^D - u_i^S\right)}{(\beta_1 - \alpha_1)} + \frac{(\beta_1 - \alpha_1)}{(\beta_1 - \alpha_1)} u_i^S \\ &= \pi_2 x_i + \underbrace{\frac{\beta_1 u_i^D - \alpha_1 u_i^S}{(\beta_1 - \alpha_1)}}_{=e_{2i}} \\ Q_i &= \pi_2 x_i + e_{2i}. \end{aligned} \tag{100}$$

So this is the second reduced form equation for Q_i.

Note here that π_1 and π_2 are the **reduced form parameters**, and that e_{1i} and e_{2i} are the **reduced form shocks** or errors.

More importantly, note here also that π_1 and π_2 can be estimated consistently by OLS because the x_i variable is (assumed to be) exogenous. So we can always find π_1 and π_2.

The identification problem

You can see that

$$\pi_1 = \frac{\alpha_2}{(\beta_1 - \alpha_1)} \text{ and } \pi_2 = \frac{\beta_1 \alpha_2}{(\beta_1 - \alpha_1)} \quad (101)$$

so that once we run OLS of P_i on x_i and Q_i on x_i we will get consistent estimates of π_1 and π_2.

But we don't care about the **reduced form parameters** π_1 and π_2 and are instead interested in the **structural parameters** α_1, α_2 and β_1.

From (101) you can see that once we replace π_1 and π_2 with their OLS estimates, we are faced with the problem of **solving for 3 unknowns** α_1, α_2 and β_1 with only **2 equations**.

This is a classical example of an **un-identified system** of equations.

In this particular case we can see that

$$\begin{aligned}\frac{\pi_2}{\pi_1} &= \frac{\beta_1\alpha_2}{(\beta_1-\alpha_1)} \div \frac{\alpha_2}{(\beta_1-\alpha_1)} \\ &= \frac{\beta_1\alpha_2}{(\beta_1-\alpha_1)} \times \frac{(\beta_1-\alpha_1)}{\alpha_2} \\ &= \beta_1.\end{aligned}$$

So replacing $\frac{\pi_2}{\pi_1}$ by their OLS estimates allows us to recover an estimate for β_1 in the supply equation (96b) $Q_i = \beta_1 P_i + u_i^S$.

In this setting, it is the demand equation $Q_i = \alpha_1 P_i + \alpha_2 x_i + u_i^D$ that is not identified in the system, and we will not be able to estimate α_1 and α_2 without some other exogenous variables.

What is needed for identification?

The necessary condition for identification in a system of G simultaneous equations is as follows: Suppose you have G endogenous variables. Then, at least $(G-1)$ variables must be absent (excluded) from any particular equation for it to be identified.

If the particular equation is identified, then we can use **TSLS** to estimate the **structural parameters** of that equation.

In our simple example above, we had $G = 2$, and thus $(G-1) = 1$. The supply equation had exactly one excluded variable, that is, x_i. So we can perform "*standard*" TSLS as we have done before in the context of IV regression.

This is implemented as follows:

a) Run the reduced form regression of P_i on x_i to get $\hat{\pi}_1$ and form the predicted or fitted value $\hat{P}_i = \hat{\pi}_1 x_i$.

b) Then, get an estimate of the structural parameter β_1 by regressing Q_i on $\hat{P}_i$.

References

Greene, William H. (2011): *Econometric Analysis*, *7th Edition*, Prentice Hall.

Hausman, Jerry A. (1978): "Specification Tests in Econometrics," *Econometrica*, **46**(6), 1251–1271.

Pagan, Adrian R. and John C. Robertson (1998): "Structural Models of the Liquidity Effect," *The Review of Economics and Statistics*, **80**(2), 202–217.

Stock, James H. and Motohiro Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by Donald W. K. Andrews and James H. Stock, Cambridge University Press, 80–108.