Topic 18 Header Slide

- Topic 18. Analysis of instrumental variable estimators
  - Discussion point #1. Identification
  - Discussion point #2. Instrument validity and relevance
  - Discussion point #3. Statistical properties of instrumental variable estimators

---

Estimator	Notation	Under Complications	Extends to
1.	$\hat{\beta}_{OLS}$	$A2Nonlinear.additiveError$	$\rightarrow \hat{\beta}_{NLLS}$
2.	$\hat{\beta}_{LAD}$		
3.	$\hat{\beta}_{Lstar}$		
4.	$\hat{\beta}_{GMM}$	$A3EndogenousX/ExogenousW$	$\rightarrow \hat{\beta}_{IVE}$
5a.	$\hat{\beta}_{IGLS}$		
5b.	$\hat{\beta}_{FGLS}$		
6.	$\hat{\beta}_{MLE}$	$Multiple$	$\rightarrow \hat{\beta}_{FIMLE}$
7.	$\hat{\beta}_{IVE}$	$A2Nonlinear.AdditiveError$	$\rightarrow \hat{\beta}_{NLIVE}$

Basic Idea of Instrumental Variables

Consider $A2linear$ with two sets of regressors, X^G and X^B:

$$y = X^G\beta^G + X^B\beta^B + \epsilon^{true}$$

or, in observation-by-observation form:

$$y_s = x_s^{G'}\beta^G + x_s^{B'}\beta^B + x_s^{B'}\beta^B + \epsilon_s^{true} = x_s'\beta + \epsilon_s^{true}$$

Given $A1$ and $A2$, the Sampling Error Vector of OLS is:

$$SEV(\hat{\beta}_{ols}) = \left(\sum_s B_s\right)^{-1} \cdot \sum_s a_s^{ols} = \left(\sum_s x_s x_s'\right)^{-1} \cdot \sum_s x_s\epsilon_s$$

But:

$$Ea_s^{ols} = E\begin{bmatrix} x_s^G\epsilon_s \\ x_s^B\epsilon_s \end{bmatrix} = \begin{bmatrix} Ex_s^G\epsilon_s \\ Ex_s^B\epsilon_s \end{bmatrix} = \begin{bmatrix} = 0 \\ \neq 0 \end{bmatrix}$$

because we are told that the 'good' variables satisfy the weak exogeneity condition $Ex^G\epsilon_s = 0$, while the 'bad' variables do not (since they are *endogenous* w.r.t. to the error).
Therefore, OLS will be inconsistent for all the βs since in general X^G and X^B are correlated.

Reverse Engineering the Instrumental Variables Estimator

Suppose we can find a data matrix W of the same dimension as the original X and of full rank k. We then define:

$$W'\epsilon^{true} = \sum_s w_s \epsilon_s^{true} = \sum_s a_s^{ive}$$

such that:

$$Ea_s^{ive} = 0$$

at the true parameter values.
We now use the GMM idea and rely on the true *population* orthogonality conditions implied by the true model: $A2 : y = X\beta + \epsilon^{true}$ and

$$A3Rsru : W, X^G Ew_s \epsilon^{true} = 0, Ex_s^{G'} = 0$$

Therefore, we define the GMM=IVE by using the *sample* orthogonality conditions:

$$W'\hat{\epsilon}^{ive} = W'(y - X\hat{\beta}_{ive}) = 0$$

to mimic the population OCs. Finally, solving for $\hat{\beta}^{ive}$ we obtain:

$$\hat{\beta}_{ive} = (W'X)^{-1}W'y$$

because $W'X$ is square and invertible given that $rank(X) = rank(W) = k$.
In conclusion, the IVE will be consistent provided every column used to construct W is a weakly exogenous variable w.r.t. the true error (i.e., satisfies A3Rsru).

Overview/Discussion point #1.

- Discussion point #1. **Identification**
  - Identification refers to the *mathematical* ability to solve uniquely for all parameters. This is distinct from estimation, which aims to use *statistical* methods to learn about unknown parameters given a sample.
  - Parameter identification is a step in the theoretical analysis of the model given a specification; parameter estimation is a step in empirical analysis of the model given data.
  - We must assess whether our model is identified (so as to subsequently permit consistent estimation) in the presence of an exogenous-endogenous regressor dichotomy.
- Discussion point #2. **Instrument validity and relevance (and feasibility)**
  - Validity and relevance are the names given to our identification conditions. These conditions permit us to use IVE methods to consistently estimate parameters of interest.
- Discussion point #3. **Statistical properties of instrumental variable estimators**
  - The IVE method typically yields analytic estimators. This permits us to use the standard form of our (previously-seen) SEV in order to investigate exact and asymptotic properties.

Analysis of the SEV – Discussion point #2.

- Recall any of the seven leading cases of endogeneity. In other words, suppose we have a violation of even the weakest form of exogeneity (i.e. even $A3Rsru$ does not hold).
- Recall analysis of $SEV(\hat{\beta}_{OLS})$. Under $A2linear$ and with the definition of OLS, we have

$$\hat{\beta}_{OLS} - \beta = \left(\frac{1}{S} \sum_{s=1}^{S} x_s x_s' \right)^{-1} \frac{1}{S} \sum_{s=1}^{S} x_s \varepsilon_s,$$

so that by Slutsky's theorem, we obtain

$$\operatorname*{plim}_{S \to \infty} \hat{\beta}_{OLS} - \beta = \left(\operatorname*{plim}_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} x_s x_s' \right)^{-1} \operatorname*{plim}_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} x_s \varepsilon_s,$$

which by "suitable" LLNs evaluates to,

$$\operatorname*{plim}_{S \to \infty} \hat{\beta}_{OLS} - \beta = \left(\mathbb{E}(x_s x_s') \right)^{-1} \mathbb{E}(x_s \varepsilon_s),$$

a quantity that is only well-defined and equal to zero under the critical assumptions that:

1. There exists a finite non-singular matrix, B_∞^0, such that $\mathbb{E}(x_s x_s') = B_\infty^0$
2. $\left(\sum_{s=1}^{S} x_s x_s' \right)^{-1}$ exists, or $A1$ holds (otherwise the first step of the proof fails)
3. $\mathbb{E}(x_s \varepsilon_s) = \mathbb{E}(a_s^{ols}) = 0$, or $A3Rsru$ holds (otherwise the final step of the proof fails)

Considering the loss of $A3Rsru$ – Discussion point #2.

- The loss of $A3Rsru$ is clearly fatal as far as our OLS estimator is concerned.
- But suppose we could (in principle, or in practice, and ideally both!) avail of data on k-dimensional vector w_s (i.e. for $s = 1, ..., S$) that happened to exhibit weak exogeneity in respect of ε_s. Let us denote this new assumption by $A3Rsru.W : \mathbb{E}(w_s \varepsilon_s) = \mathbb{E}(a_s^{ive}) = 0$.
- How could we exploit this windfall source of exogenous variation?
- We cannot blindly replace x_s with w_s in the previous proof – this would constitute a major mis-specification of the original model. (Remember that no one is interested in estimating the partial effect of w_s on y_s. What we care about is the partial effect of x_s on y_s.)
- One could, however, develop a whole new approach (OLS simply cannot be rescued) as per the following slide. Let us first consider the algebra and then consider the intuition.

SEV-based introduction to IVE (1 of 4) – Discussion point #2.

- Consider again our consistency proof, but this time we will try to avail of the windfall data on w_s for $s = 1, ..., S$. Consider a thought-experiment in which we analyse the properties of

$$\hat{\beta}_{novel} \equiv \left(\frac{1}{S} \sum_{s=1}^{S} w_s x_s' \right)^{-1} \frac{1}{S} \sum_{s=1}^{S} w_s y_s$$

- (Ultimately, this "novel" estimator will be exactly our IV estimator, and we will drop the odd-sounding name. But let us just work through the mechanics first so that we can define what we mean by an instrument.)

SEV-based introduction to IVE (2 of 4) – Discussion point #2.

- Consider the analysis of $SEV(\hat{\beta}_{novel})$. Under $A2linear$ and given the definition of our novel estimator, we have

$$\hat{\beta}_{novel} - \beta = \left(\frac{1}{S} \sum_{s=1}^{S} w_s x_s' \right)^{-1} \frac{1}{S} \sum_{s=1}^{S} w_s \varepsilon_s,$$

so that by Slutsky's theorem, we obtain

$$\underset{S\to\infty}{plim}\, \hat{\beta}_{novel} - \beta = \left(\underset{S\to\infty}{plim}\, \frac{1}{S} \sum_{s=1}^{S} w_s x_s' \right)^{-1} \underset{S\to\infty}{plim}\, \frac{1}{S} \sum_{s=1}^{S} w_s \varepsilon_s,$$

which by a suitable LLN evaluates to,

$$\underset{S\to\infty}{plim}\, \hat{\beta}_{novel} - \beta = \left(\mathbb{E}(w_s x_s') \right)^{-1} \mathbb{E}(w_s \varepsilon_s),$$

which is only well-defined and equal to zero under the critical assumptions that:

1. There exists a finite non-singular matrix, B_∞^0, such that $\mathbb{E}(w_s x_s') = B_\infty^0$
2. $\left(\sum_{s=1}^{S} w_s x_s' \right)^{-1}$ exists (otherwise the first step of the proof fails)
3. $\mathbb{E}(w_s \varepsilon_s) = \mathbb{E}(a_s^{ive}) = 0$, or $A3Rsru.W$ holds (otherwise the final step of the proof fails)

SEV-based introduction to IVE (3 of 4) – Discussion point #2.

- Clearly, under conditions 1–3 on the previous slide, our novel estimator is consistent. (One can arguably think of these conditions as reverse-engineered to ensure consistency!)
- All that remains is for us to give w_s a scientific name; that is, an instrument.
- Indeed, an instrument is defined precisely such that conditions 1–3 hold. These defining conditions, in turn, are so crucial that they bear their own names. Let us review them:
  1. There exists a finite non-singular matrix, B_∞^0, such that $\mathbb{E}(w_s x_s') = B_\infty^0$
  2. $\left(\sum_{s=1}^{S} w_s x_s'\right)^{-1}$ exists (otherwise the first step of the proof fails)
  3. $\mathbb{E}(w_s \varepsilon_s) = \mathbb{E}(a_s^{ive}) = 0$, or $A3Rsru.W$ holds (otherwise the final step of the proof fails)
- Conventional nomenclature for the bullets above is as follows:
  1. Condition 1 is none other than "relevance". It ensures the inverse term in the final step of the consistency proof is well-defined.
  2. Condition 2 is "in-sample relevance" or "feasibility". It ensures the IV estimator exists to begin with. (For instance, it precludes perfect collinearity among the w_s variables.)
  3. Condition 3 is none other than "validity". It ushers in the final consistency result.

SEV-based introduction to IVE (4 of 4) – Discussion point #2.

We are now in a position to (i) define what an instrument means; and (ii) concurrently move the discussion forwards by setting up the notation to generalise our framework from the so-called "just-identified" setting to the "over-identified" setting (which we will analyse in detail shortly).

- A q-dimensional vector w_s is said to be instrumental for k-dimensional vector of (potentially endogenous) regressors x_s if w_s satisfies the twin conditions of validity and relevance.
  - When $q < k$, the model is said to be "under-identified". (No solution is possible.)
  - When $q > k$, the model is said to be "over-identified". (We will study 2SLS shortly.)
  - When $q = k$, the model is said to be "just-identified". (Our current focus – i.e., IVE.)
- So long as $q = k$, we can define the IV estimator as:

$$\hat{\beta}_{IVE} \equiv (W'X)^{-1}W'y = \left(\sum_{s=1}^{S} w_s x_s'\right)^{-1} \sum_{s=1}^{S} w_s y_s, \text{ where } W \equiv \begin{pmatrix} w_1' \\ w_2' \\ \vdots \\ w_S' \end{pmatrix} \text{ is an } S \times k \text{ matrix}$$

  satisfying the needed feasibility condition that $(W'X)^{-1}$ exists.
- Note that when $q > k$, the over-identified setting, the inverse of $(W'X)$ cannot exist (at least not in a conventional sense) because it is not a square matrix.

Asymptotic analysis – Discussion point #3.

- Recall the common structure of the SEV for OLS and IVE estimators.
- As a reminder, we have

$$\left(\sum_{s=1}^{S} B_s\right)^{-1} \sum_{s=1}^{S} a_s = (B_S)^{-1} a_S,$$

  where each B_s is a $k \times k$ invertible matrix and each a_s is a $k \times 1$ vector, with matrix B_S and vector a_S defined as the sums over S terms accordingly.
- For *both* consistency and asymptotic normality, the first requirement is that an LLN must apply to matrix B_S/S so that it has a well-defined probability limit:

$$\underset{S\to\infty}{p\lim} \, (B_S/S) = B_\infty^0,$$

  where B_∞^0 is a finite non-singular limiting matrix.

A closer look at consistency – Discussion point #3.

- For consistency, specifically, there are two additional requirements involving vector a_S.
- First, an LLN must apply to vector a_S/S so that it has a well-defined probability limit:

$$\underset{S\to\infty}{p\lim}\,(a_S/S) = a_\infty^0$$

- Second, the true model must satisfy conditions that guarantee that $a_\infty^0 = 0$.
- Given these two conditions, the first continuous mapping theorem ("CMT1") – alternatively known as Slutsky's theorem – applied twice will guarantee that the probability limit of the SEV under consideration will be:

$$\begin{aligned}\underset{S\to\infty}{p\lim}\left(\left(\frac{B_S}{S}\right)^{-1}\frac{a_S}{S}\right) &= \underset{S\to\infty}{p\lim}\left(\left(\frac{B_S}{S}\right)^{-1}\right)\cdot\underset{S\to\infty}{p\lim}\left(\frac{a_S}{S}\right)\\ &= \left(\underset{S\to\infty}{p\lim}\left(\frac{B_S}{S}\right)\right)^{-1}\cdot\underset{S\to\infty}{p\lim}\left(\frac{a_S}{S}\right) = (B_\infty^0)^{-1}\cdot a_\infty^0\end{aligned}$$

  Hence, the estimator in question will be consistent provided $a_\infty^0 = 0$.

A closer look at asymptotic normality – Discussion point #3.

- For asymptotic normality, specifically, there is one additional requirement involving vector a_S. That is, a CLT must apply to vector $\sqrt{S} \cdot (a_S/S) = a_S/\sqrt{S}$ so that it exhibits a convergence in distribution given by:

$$a_S/\sqrt{S} \xrightarrow{d} Z_0 \text{ as } S \to \infty$$

  for $k \times 1$ vector Z_0 such that $Z_0 \sim \mathcal{N}(0, \Sigma_Z)$.

- Given this condition, the second continuous mapping theorem ("CMT2") – alternatively known as Cramér's theorem – will guarantee that the normalised SEV given by

$$(B_S/S)^{-1} \cdot a_S/\sqrt{S}$$

  will obey $\mathcal{N}\left(0, \left(B^0_\infty\right)^{-1} \Sigma_Z \left(B^0_\infty\right)^{-1'}\right)$ as its limiting distribution.

Consistency of the OLS estimator (1 of 2) – Discussion point #3.

Let us apply what we have learned to the OLS case:

- In the OLS case, we have (as per our 'common structure' notation):

$$a_S/S = \frac{1}{S}\sum_{s=1}^{S} x_s\varepsilon_s, \text{ and } B_S/S = \frac{1}{S}\sum_{s=1}^{S} x_s x_s'$$

- For consistency, we said that the first requirement is that an LLN must apply to matrix B_S/S so that it has a well-defined probability limit:

$$\underset{S\to\infty}{p\lim}\,(B_S/S) = B_\infty^0,$$

  where B_∞^0 is a finite non-singular limiting matrix.

- In particular, for OLS, what we will need is for an LLN to apply to matrix $\frac{1}{S}\sum_{s=1}^{S} x_s x_s'$ so that it has a well-defined probability limit:

$$\underset{S\to\infty}{p\lim}\,\frac{1}{S}\sum_{s=1}^{S} x_s x_s' = \mathbb{E}(x_s x_s'),$$

  where $\mathbb{E}(x_s x_s')$ is a finite non-singular matrix. This precludes perfect linear relationships among the regressors (in the population).

Consistency of the OLS estimator (2 of 2) – Discussion point #3.

- Further, we said that an LLN must apply to vector a_S/S so that it too has a well-defined probability limit:

$$\underset{S\to\infty}{p\lim}\,(a_S/S) = a_\infty^0$$

- Moreover, the true model must not betray the guarantee that $a_\infty^0 = 0$.

- Applying the above to the OLS case, we will require that vector a_S/S has probability limit

$$\underset{S\to\infty}{p\lim}\,\frac{1}{S}\sum_{s=1}^{S} x_s\varepsilon_s = \mathbb{E}(x_s\varepsilon_s),$$

  which equals zero only if $A3Rsru$ holds.

- A violation of weak exogeneity of one or more regressors leads to inconsistency of the OLS estimator. It will not be CUAN in general.

Asy. normality of the OLS estimator (1 of 2) – Discussion point #3.

- For asymptotic normality of the OLS estimator, let us examine the 'additional' requirement (over and above the limiting behaviour of B_S/S) involving vector $a_S = \sum_{s=1}^{S} x_s\varepsilon_s$.
- In particular, a CLT must apply to vector $\sqrt{S} \cdot (\sum_{s=1}^{S} x_s\varepsilon_s/S) = \sum_{s=1}^{S} x_s\varepsilon_s/\sqrt{S}$ so that it exhibits a convergence in distribution given by:

$$\left(\frac{1}{\sqrt{S}}\sum_{s=1}^{S} x_s\varepsilon_s\right) \xrightarrow{d} Z_0 \text{ as } S \to \infty$$

  for $k \times 1$ vector Z_0 such that $Z_0 \sim \mathcal{N}(0, \Sigma_Z)$.
- Under $A4GM(iid)$, we will obtain $\Sigma_Z = \underset{S\to\infty}{p\lim}\ (X'X/S)$
- Under $A4\Omega$, we will obtain $\Sigma_Z = \underset{S\to\infty}{p\lim}\ (X'\Omega X/S)$

Asy. normality of the OLS estimator (2 of 2) – Discussion point #3.

- Subsequent to the discussion on the previous slide, we apply the second continuous mapping theorem ("CMT2") – alternatively known as Cramér's theorem – which will guarantee that the normalised SEV will converge in distribution as per:

$$\sqrt{S}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \left(\underset{S\to\infty}{plim}\ (X'X/S)\right)^{-1}\right)$$

  as S passes to infinity under $A4GM(iid)$.

- The corresponding expression for the asymptotic variance becomes

$$c^2 \left(\underset{S\to\infty}{plim}\ (X'X/S)\right)^{-1} \underset{S\to\infty}{plim}\ (X'\Omega X/S) \left(\underset{S\to\infty}{plim}\ (X'X/S)\right)^{-1}$$

  under $A4\Omega$.

Asymptotics of the IV estimator – Discussion point #3.

The key results in the IVE case, given $S \times k$ instrument matrix W are as follows:

(i) Under relevance, an LLN ensures that $\underset{S\to\infty}{plim}\ (W'X/S)$ is a finite non-singular $k \times k$ matrix.

(ii) Under validity, an LLN ensures that $\underset{S\to\infty}{plim}\ (W'\varepsilon/S)$ is the k-dimensional zero vector.

(iii) Under the feasibility condition, the SEV of the estimator, $(W'X)^{-1}W'\varepsilon$, is well-defined.

(iv) We will also require a 'well-behaved' population second moment matrix for instruments W as captured by existence and non-singularity of $\underset{S\to\infty}{plim}\ (W'W/S)$.

(v) Under these conditions, a CLT ensures that $W'\varepsilon/\sqrt{S} \xrightarrow{d} \mathcal{N}(0,\Sigma)$ as S passes to infinity, where $\Sigma = \underset{S\to\infty}{plim}\ (W'W/S)$ under $A4GM(iid)$ and $\Sigma = \underset{S\to\infty}{plim}\ (W'\Omega W/S)$ under $A4\Omega$.

(vi) It follows from (i)–(iii) and Slutsky's theorem that $SEV(\hat{\beta}_{IVE}) \xrightarrow{p} 0$ as $S \to \infty$; and it follows from (i)–(v) and Cramér's theorem that $\sqrt{S} \cdot SEV(\hat{\beta}_{IVE}) \xrightarrow{d} Z_0$ as $S \to \infty$, for $k \times 1$ vector Z_0, where $Z_0 \sim \mathcal{N}(0,\Sigma_Z)$ with the definition of Σ_Z as per:

$$c^2 \left(\underset{S\to\infty}{plim}\ (W'X/S)\right)^{-1} \underset{S\to\infty}{plim}\ (W'\Omega W/S) \left(\underset{S\to\infty}{plim}\ (X'W/S)\right)^{-1}$$

under $A4\Omega$.

A mix of endogenous-exogenous regressors (the model)

- Consider a situation in which we have a single endogenous regressor and several hundred exogenous regressors. The temptation might be to think that the exogenous variation, in some sense, swamps or drowns-out the endogenous variation. However, this reasoning is incorrect. We see why below.

- Suppose we have a linear regression model with data (y, X) of sample size S. The k regressors are grouped in two parts, X_A and X_B, of dimensions $S \times k_A$ and $S \times k_B$ respectively, with $k_A + k_B = k$. Suppose that the model satisfies the following assumptions:

$$
\begin{aligned}
&A1: && rank(X) = k < S \\
&A2: && y = X\beta + \varepsilon = \underbrace{X_A\beta^A}_{X^{Good}\beta^{Good}} + \underbrace{X_B\beta^B}_{X^{Bad}\beta^{Bad}} + \varepsilon = \underbrace{Z^I\beta^G}_{X^{Good}\beta^{Good}} + \underbrace{X_B\beta^B}_{X^{Bad}\beta^{Bad}} + \varepsilon \\
&A3Rmi.X_A: && \mathbb{E}(\varepsilon|X_A) = \mathbb{E}(\varepsilon) \\
&A3.X_B: && \varepsilon \text{ and } X_B \text{ correlated for all } s = 1, ..., S \\
&A4\Omega: && \mathbb{E}(\varepsilon\varepsilon'|X) = c^2\Omega \\
&A5Gaussian: && \varepsilon_s|X \sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

  In other words, regressors X_B are endogenous with respect to the true error.

- We are particularly interested in the true coefficients of the X_A variables, β_A.

A mix of endogenous and exogenous regressors (the problem)

- The sampling error vectors for OLS and IGLS estimators is respectively:

$$\begin{pmatrix} \hat{\beta}_A^{OLS} \\ \hat{\beta}_B^{OLS} \end{pmatrix} - \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} = \begin{pmatrix} X_A'X_A & X_A'X_B \\ X_B'X_A & X_B'X_B \end{pmatrix}^{-1} \begin{pmatrix} X_A'\varepsilon \\ X_B'\varepsilon \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_A^{IGLS} \\ \hat{\beta}_B^{IGLS} \end{pmatrix} - \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} = \begin{pmatrix} X_A'\Omega^{-1}X_A & X_A'\Omega^{-1}X_B \\ X_B'\Omega^{-1}X_A & X_B'\Omega^{-1}X_B \end{pmatrix}^{-1} \begin{pmatrix} X_A'\Omega^{-1}\varepsilon \\ X_B'\Omega^{-1}\varepsilon \end{pmatrix}$$

- Given these expressions, it is clear that both estimators will be biased and inconsistent for coefficients on both sets of regressors (i.e., sets A and B). This is because:

  1. the endogeneity of regressors X_B implies that the terms $X_B'\varepsilon$ and $X_B'\Omega^{-1}\varepsilon$ will not be zero in expectation, nor will $X_B'\varepsilon/S$ and $X_B'\Omega^{-1}\varepsilon/S$ converge to zero asymptotically;

  2. the bias/inconsistency carries over even to the estimated coefficients on the X_A variables because in general neither $X_A'X_B$ nor $X_A'\Omega^{-1}X_B$ will vanish.

- Detailed formal algebraic explanations (i.e., based on partitioned regression formulas) of the above statements are provided in the extended notes. (These are DrRS's old teaching notes based on the associated problem set question – i.e., the same one that was ear-marked for you to submit to your class teachers. Compare his solution with your own...? Hope it helps.)

$$y = X_A \beta_A + X_B \beta_B + \varepsilon$$

where X_A is $T \times (K-1)$ and

X_B is $T \times 1$ and $E[x_B' \varepsilon] \neq 0$.

We have A1, A2, A3Rmi.X_A, A4Ω, A5N.

(a) $\hat{\beta}_{OLS} := (X'X)^{-1} X'y$ where $X := [X_A \; X_B]$.

$\hat{\beta}_{IGLS} := (X'\hat{\Omega}^{-1}X)^{-1} X'\hat{\Omega}^{-1}y$.

let $\beta := (\beta_A', \beta_B)'$, $\hat{\beta}_{OLS} := (\hat{\beta}_{A,OLS}', \hat{\beta}_{B,OLS})'$ and

$\hat{\beta}_{IGLS} := (\hat{\beta}_{A,IGLS}', \hat{\beta}_{B,IGLS})'$.

"Explain carefully... unbiasedness"

Consider $(\hat{\beta}_{A,OLS} - \beta_A) = (X_A' M_B X_A)^{-1} X_A' M_B \varepsilon$ where $M_B := [I_T - X_B (X_B' X_B)^{-1} X_B']$.

Now, $E[\hat{\beta}_{A,OLS} - \beta_A] \overset{LIE}{=} E[(X_A' M_B X_A)^{-1} X_A' M_B E[\varepsilon | X]] \overset{A3.X_B}{\neq} 0$, even under A3Rmi.$X_A$

Similarly,

$(\hat{\beta}_{A,IGLS} - \hat{\beta}_A) = (\tilde{X}_A' \tilde{M}_B \tilde{X}_A)^{-1} \tilde{X}_A' \tilde{M}_B \tilde{\varepsilon}$ where

$\tilde{X}_A := \hat{\Omega}^{-1/2} X_A$

$\tilde{\varepsilon} := \hat{\Omega}^{-1/2} \varepsilon$

$\tilde{M}_B := [I_T - \hat{\Omega}^{-1/2} X_B ((\hat{\Omega}^{-1/2} X_B)' (\hat{\Omega}^{-1/2} X_B))^{-1} (\hat{\Omega}^{-1/2} X_B)']$.

So $E[\hat{\beta}_{A,IGLS} - \beta_A] \overset{LIE}{=} E[(\tilde{X}_A' \tilde{M}_B \tilde{X}_A)^{-1} \tilde{X}_A' \tilde{M}_B \hat{\Omega}^{-1/2} E[\varepsilon | X]] \overset{A3.X_B}{\neq} 0$

even under A3Rmi.X_A.

"Explain carefully... Consistency."

For some K-1 by K-1 finite P.D. matrix J,

$$\operatorname*{plim}_{T\to\infty}\left(\hat{\beta}_{A,OLS}-\beta_A\right) \overset{ST}{=} \left[\operatorname*{plim}_{T\to\infty}\left(T^{-1}X_A'M_BX_A\right)\right]^{-1}\operatorname*{plim}_{T\to\infty}T^{-1}X_A'M_B\varepsilon$$

$$= J\cdot\operatorname*{plim}_{T\to\infty}T^{-1}\left[X_A'\varepsilon - X_A'X_B\left(X_B'X_B\right)^{-1}X_B'\varepsilon\right]$$

$\neq 0$ even if $\operatorname*{plim}_{T\to\infty}\left(T^{-1}X_A'\varepsilon\right) \overset{A3Rmi.X_A}{=} 0$ since

$\operatorname*{plim}_{T\to\infty}\left(T^{-1}X_B'\varepsilon\right) \overset{A3.X_B}{\neq} 0$ unless $\operatorname*{plim}_{T\to\infty}\frac{1}{T}\left(X_A'X_B\right)=0.$

Similarly, since $\operatorname*{plim}_{T\to\infty}T^{-1}\left(\Omega^{-1/2}X_B\right)'\Omega^{-1/2}\varepsilon \neq 0$, we also have that $\operatorname*{plim}_{T\to\infty}\left(\hat{\beta}_{A,IGLS}-\beta_A\right)\neq 0.$

ASK YOURSELF: What's the key message of these slides?

A mix of endogenous and exogenous regressors (IVE – 1 of 2)

As a possible response to the problem of endogeneity of X_B, econometricians proposed the IVE-based solution, defined by:

$$\hat{\beta}_{IVE} = (W'X)^{-1}W'y, \text{ where}$$
$$SEV(\hat{\beta}_{IVE}) = \hat{\beta}_{IVE} - \beta^{true} = (W'X)^{-1}W'\varepsilon$$

Here is a recap of the key points; now from a practitioner's (rather than a theorist's) perspective:

- The matrix W should be of the same dimension as X and have full rank k. Further, $W'X$ should have full rank $k = k_A + k_B$.
- Further, W should consist only of exogenous variables, implying that X_A can be used.
- Since this disallows the use of the endogenous regressor X_B, in order to make the method feasible, we must find k_E additional "instrument" variables ($k_E \geq k_B$) to construct W.
- As a reminder, W should consist solely of valid instruments or linear combinations of such valid instruments. An instrument variable z is termed "valid" if it is weakly exogenous w.r.t. to the error term – i.e., $\mathbb{E}(z_s\varepsilon_s) = 0$.
- An instrument variable z is termed "relevant" if it has a high correlation with the endogenous variables of the model, in this case the endogenous regressors X_B.

A mix of endogenous and exogenous regressors (IVE – 2 of 2)

- Notice the potential for over-identification alluded to in the previous slide. If k_E were *strictly* larger than k_B, we would have to eliminate the extra instruments so as to maintain feasibility.
- However, wasting information is never a good idea. Formally, there are efficiency gains to be exploited by (linearly) combining the information across all $k_z = k_I + k_E$ available instruments.
- So what is the optimal choice, say W^*, in the over-identified setting? We explore this next.

Dealing with over-identification

- Say we have available an $S \times k_z$ matrix of (valid and relevant) instruments, Z, where $k_z > k$.
- Then, the optimal choice would entail construction of an $S \times k$ matrix, W^*, that constitutes effectively a $k = (k_A + k_B) < k_z = (k_A + k_E)$ dimensional fitted (via least squares) version of all available weakly exogenous variables, given by $Z = (Z^I, Z^E) = (X_A, Z^E)$.
- Indeed, the optimal combination – one that maximises the correlations between W^* and the original regressors (X_A, X_B) – is obtained via the fitted/predicted values from a 'multivariate' regression, using OLS, of $S \times (k_A + k_B)$ matrix (X_A, X_B) on $S \times (k_A + k_E)$ matrix (X_A, Z^E).
- Let us denote these fitted/predicted values as $S \times k$ matrix $\hat{X} = (\hat{X}_A, \hat{X}_B) = (X_A, \hat{X}_B)$.
- The multivariate regression described on this slide is known as the "1st stage regression".
- To summarise we have our $S \times k$ optimal instrument matrix, W^*, given by:

$$W^* = \hat{X} = Z\hat{\pi} = Z(Z'Z)^{-1}Z'X,$$

  where $\hat{\pi} = (Z'Z)^{-1}Z'X$ is a $k_z \times k$ matrix of so-called 1st stage regression coefficients and $\hat{Z} = Z\hat{\pi}$ is the $S \times k$ (i.e. lower-dimensional) projection.

- The so-called optimal IVE is then obtained as

$$\hat{\beta}_{Opt.IVE} = (W^{*'}X)^{-1}W^{*'}y$$

A matter of interpretation: Robert Basmann

- There exists an interesting mathematical equivalence between two distinct approaches to solving the over-identification problem:

  Approach 1. (Opt.IVE) This is what we considered on the previous slide.
  To recap, we have

  $$\hat{\beta}_{Opt.IVE} = (W^{*'}X)^{-1}W^{*'}y \text{ where } W^* = Z(Z'Z)^{-1}Z'X$$

  – We can interpret the second equation above as a 1st stage regression.

  – We can interpret the first equation above as a subsequent 2nd stage regression in which the fitted values from the 1st stage serve as a (suitably-dimensioned) instrument matrix for endogenous regressor matrix, X.

  – Such an interpretation for the solution to the endogeneity problem (i.e., undertaking IVE in the second stage using W^*) is due to Basmann (1957).

A matter of interpretation: Henri Theil

- There exists an interesting mathematical equivalence between two distinct approaches to solving the over-identification problem:

  Approach 2. (2SLS) An alternative approach is as follows:

  $$\hat{\beta}_{2SLS} = (W^{*'}W^{*})^{-1}W^{*'}y \text{ where } W^{*} = Z(Z'Z)^{-1}Z'X$$

  - We still interpret the second equation above as a 1st stage regression.
  - But we now interpret the first equation above as a subsequent 2nd stage regression in which the fitted values from the 1st stage serve as an asymptotically weakly exogenous (by construction) regressor.
  - Such an interpretation for the solution to the endogeneity problem (i.e., undertaking OLS in the second stage using W^{*}) is due to Theil (1953).

- The two approaches are numerically equivalent in the sense that

  $$\hat{\beta}_{Opt.IVE} = \hat{\beta}_{2SLS}$$

  because, defining $P_Z \equiv Z(Z'Z)^{-1}Z'$, we see that

  $$(W^{*'}X) = (\hat{X}'X) = (P_ZX)'X = X'P_Z'P_ZX = (\hat{X}'\hat{X}) = (W^{*'}W^{*})$$

  by symmetry and idempotence of P_Z.

4.6 *A Discussion of "Instrument Validity" and "Instrument Relevance"*

A colleague proposes the following instrument variables:

Variable z_1 is the sum of the first three regressor variables from the X_A group i.e., $z_{t1} = x_{A,t1}+x_{A,t2}+x_{A,t3}$

Since z_{t1} is a perfect linear combination of the first three regressor variables of the A group, the W matrix will not have a full rank of k. Hence $W'X$ will not be invertible and the IV method will break down. Note that z_{t1} will be weakly exogenous since each constituent variable is weakly exogenous, so it will be "valid" in the strict sense of the term. But it will be useless as an instrument since it cannot be used for the IVE construction. Note also that it would appear as highly "relevant" if we simply checked its correlation with the endogenous regressor x_B., since the X_A variables are typically going to be correlated with x_B.

Variable z_2 is the square of the fourth regressor variable from the X_A group i.e., $z_{t2} = x_{A,t5}^2$

This may be a decent instrument depending on the precise nature of the $x_{A,t5}$ regressor: it will be valid since it can be expected to be weakly exogenous because $x_{A,t5}$ is believed to be weakly exogenous and its square will be also. And it will be relevant to some extent because it can be expected to be correlated with x_B if $x_{A,t5}$ is correlated. And finally, it will not violate the rank condition since the correlation with the original $x_{A,t5}$ will not be perfect.

Variable z_3 is a measure of sunspot activity in period t z_3 will clearly be a "valid" instrument in the sense of being weakly exogenous w.r.t. the error term, since sunspot activity can safely be assumed exogenous w.r.t. basically everything! But it will be completely "irrelevant" in that there is no reason to expect it to be correlated wth the endogenous regressor x_B.

An Analysis of Consistency — Contrasting Three Estimators: OLS, IVE, and MLE

An Explanation as to Why Misspecified MLE will be Inconsistent in General [SKIM!]

Recall the Consistency proof for any "regular", analytic ML estimator, the SEV of which has the following Taylor linearization expansion:

$$SEV(\hat{\theta}_{MLE}) = \hat{\theta}_{MLE} - \theta^{true} \approx (B_S)^{-1} a_s = (\frac{1}{S}\sum_{s=1}^{S} \ell_{\theta\theta s}(\theta^{true}))^{-1} \frac{1}{S}\sum_{s=1}^{S} \ell_{\theta s}(\theta^{true})$$

When an appropriate LLN applies to the last term, its probability limit will be:

$$p\lim_{S\to\infty} \frac{1}{S}\sum_{s=1}^{S} \ell_{\theta s}(\theta^{true}) = E(\ell_{\theta s}(y_s \mid X; \theta^{true})$$

When the loglikelihood contribution and its derivatives (first and second) are all correctly specified:

$$\ell_s(y_s \mid X; \theta^{true}) \equiv \ln f(y_s \mid X; \theta^{true})$$

the expectation of the first-derivative vector, taken with respect to the true distribution of $y_s|X$, will equal 0_{px1} as we have proved in the previous section:

$$\int_{-\infty}^{\infty} \ell_{\theta s}(y_s \mid X; \theta^{true}) f(y_s \mid X; \theta^{true}) dy_s \equiv E(\ell_{\theta s}(y_s \mid X; \theta^{true}) = 0_{px1}$$

But suppose that our A5 assumption is *incorrect*, i.e., there is a *discrepancy* between the assumed distribution used to define the LLF contributions — denoted by $f^{employed}(\cdot|\cdot)$:

$$\ell_s^{employed}(y_s \mid X; \cdot) \equiv \ln f^{employed}(y_s \mid X; \cdot)$$

and the *true* conditional distribution of $y_x|X$, denoted by $f^{true}(y_s \mid X; \theta^{true})$. Now consider the integral of the Expectation of the last term probability limit:

$$E_{wrt\ f^{true}(y|\cdot)}(\ell_{\theta s}^{employed}(y_s \mid X; \theta^{true}) = \int_{-\infty}^{\infty} \ell_{\theta s}^{employed}(y_s \mid X; \theta^{true}) f^{true}(y_s \mid X; \theta^{true}) dy_s \neq 0_{px1}$$

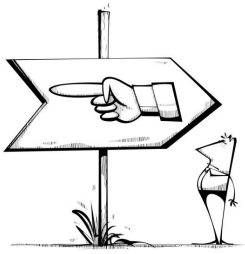
In general, this will *NOT* equal 0_{px1} because the first derivatives $\ell_{\theta s}^{employed}(\cdot)$ are defined using $f^{emplyed}(\cdot)$, the pdf used to define the LLF, while the expectation is taken over the true pdf $f^{true}(\cdot)$. This discrepancy can be expected to result in inconsistency of the misspecified MLE.

Review Quiz for Topic 18

Question 1. Please try to present a step-by-step flawlessly-executed formal and rigorous consistency proof for the OLS/IVE estimators under the classical LRM.

Question 2. Please try to present a step-by-step flawlessly-executed formal and rigorous asymptotic normality proof for the OLS/IVE estimators under the classical LRM.

Question 3. Please review your answer to the IVE question that you submitted to your class teachers for feedback and see if you can strengthen it using material provided in this slide pack.



Signpost 18

Verbeek has a very good exposition of IVE and 2SLS. Verbeek refers to optimal IVE as "GIVE". Cameron and Trivedi (PDF available online) is also an excellent resource.