Topic 14 Header Slide

- Topic 14. Logical relations between regression assumptions $A1$ through $A5$

  First part of discussion:
  Logical relations between $A2$ and $A3$ assumptions specifically. (This part of the discussion also includes an interesting historical example to motivate Topic 14.)

  Second part of discussion:
  Synthesis of all five assumptions. (This is the core of Topic 14.)

Estimator	Notation	Remark
1.	$\hat{\beta}_{OLS}$	
2.	$\hat{\beta}_{LAD}$	
3.	$\hat{\beta}_{Lstar}$	
4.	$\hat{\beta}_{GMM}$	
$5a$.	$\hat{\beta}_{IGLS}$	
$5b$.	$\hat{\beta}_{FGLS}$	
6.	$\hat{\beta}_{MLE}$	
7.	$\hat{\beta}_{IVE}$	(upcoming in Topic 18)

Joint consideration of $A2$ and $A3$

- Recall that $A2linear$ was such that it specified the following relationships between our vectors of random variables:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_S \end{pmatrix} = \begin{pmatrix} x'_1 \\ \vdots \\ x'_S \end{pmatrix} \begin{pmatrix} \beta_1^{true} \\ \vdots \\ \beta_k^{true} \end{pmatrix} + \begin{pmatrix} \varepsilon_1^{true} \\ \vdots \\ \varepsilon_S^{true} \end{pmatrix},$$

where x'_s represents the s-th row of $S \times k$ regressor matrix X, for $s = 1, ..., S$ and

$$\mathbb{E} \begin{pmatrix} \varepsilon_1^{true} \\ \vdots \\ \varepsilon_S^{true} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- The many different versions of $A3$ notwithstanding, the assumption goes hand-in-hand with $A2$. Indeed, the assumptions $A2$ and $A3$ are directly interrelated because making a mistake about one necessarily will imply that the other assumption is mis-specified.

- We explain by way of two examples below.

Theoretical Example

- Suppose the true data-generating process is

$$A2linear : y = X_A \beta_A^{true} + X_B \beta_B^{true} + \varepsilon^{true} \text{ and } \mathbb{E}(\varepsilon^{true}) = 0,$$

and

$$A3Rmi : \mathbb{E}(\varepsilon^{true} | X_A, X_B) = \mathbb{E}(\varepsilon^{true}).$$

- But we are guilty of mis-specification as per

$$A2linear.misspecified : y = X_A \beta_A^{true} + \eta$$

- In other words, in our mis-specified estimating model, we have a composite error given by:

$$\eta = X_B \beta_B^{true} + \varepsilon^{true}.$$

- It is clear that unless $\beta_B^{true} = 0$ or $X_A' X_B = 0$ or both, we will be unable to disentangle the error from the regressor, and no $A3$ can hold.

Real-life example

- French physician and anatomist, Paul Broca (1824–1880) collected around $S = 200$ corpses and performed autopsies $(s = 1, ..., 200)$.
- He recorded brain weight of corpse s in grams, W_s, for men, gender dummy $G_s = 1$, and women, gender dummy $G_s = 0$, and found a 16% higher average brain weight for men with a t statistic of around 20.
- The estimated specification can be expressed as

$$W_s = \hat{\beta}_{1,OLS} + \hat{\beta}_{2,OLS} G_s,$$

  where $\hat{\beta}_{2,OLS} = 0.16$ and $t \approx 20$.

- In terms of $A2$ and $A3$, we express the estimating model as

$$W_s = \beta_1^{true} + \beta_2^{true} G_s + \eta_s$$

  where η_s violates $A3$ because it includes important regressors left out by Paul Broca – some would say *intentionally so*!

Real-life example

- In particular, we have

$$\eta_s = \beta_3^{true} M_s + \beta_4^{true} A_s + ... + \varepsilon_s^{true},$$

where M_s is body mass/weight and A_s is age at death. The interpretation of β_3^{true} is as a "dinosaur effect" (positive) and β_4^{true} is as a "drying out effect" (negative)!

- Clearly, $A3$ (as applied with respect to η_s) is violated in multiple ways since:
  - Corr$(G_s, W_s) > 0$ (men are typically heavier than women); and
  - Corr$(G_s, A_s) < 0$ (men typically die younger).

Synthesis of all five assumptions

Below, we will consider five examples of settings, or "situations", in which what we assume (or do not assume) in one category (e.g., in the Broca example, $A2$) has consequences for what we are able to assume (or not able to assume) in another category (e.g., in the Broca example, $A3$).

Situation 1: Perfect collinearity

Causes:

- Suppose we have a 'dummy variable trap' situation. This leads to rank-deficiency in $X'X$.
- An example would be the inclusion of 4 quarterly dummies, say $\{Q1, Q2, Q3, Q4\}$, as regressors along with a constant in the regression specification. This leads to a perfect linear relationship among the regressors, which is the source of the rank-deficiency.

Consequences:

- $(X'X)^{-1}$ does not exist, and accordingly neither does $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$.

Solutions:

- Think carefully before being over-zealous with dummies. Either drop the constant or choose a dummy to drop. (If you are judicious about choice of which dummy to drop, you might be able to facilitate easy interpretation of the remaining coefficients.)
  Of course, if you drop regressor variables (intercept? dummies?) then you redefine the X matrix and hence you will change the $A2$ specification.

Situation 2: Near-perfect collinearity

Causes:

- Suppose $X'X$ is non-singular, but such that its minimum eigenvalue is very small (i.e., only very slightly larger than 0).
- An example would be where there is an extremely high correlation between two regressors.

Consequences:

- Inversion of $(X'X)$ is numerically unstable because computation of $(X'X)^{-1}$ involves division by a determinant (product of eigenvalues) that is very close to zero. $\hat{\beta}_{OLS}$ will be estimated with high variance, as evidenced by $\mathbb{V}(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$ (under $A4GM$).

Solutions:

- One cannot do much, since the specific form of $X'X$ is usually a property of the data and typically beyond our control in observational (rather than experimental) settings.

Situation 3.benign: $A2$ where $\mathbb{E}(\varepsilon^{true}) \neq 0$ but $A3Rsru$ holds

Causes:

- Suppose we have

$$A2 : y_s = \beta_1^{true} + \beta_2^{true} x_{s2} + ... + \beta_k^{true} x_{sk} + \varepsilon_s^{true}, \quad \mathbb{E}(\varepsilon_s^{true}) = c \neq 0$$

Consequences:

- Defining $u_s = \varepsilon_s^{true} - c$ and substituting, we have

$$A2^* : y_s = \beta_1^* + \beta_2^{true} x_{s2} + ... + \beta_k^{true} x_{sk} + u_s, \quad \mathbb{E}(u_s) = 0$$

  where $\beta_1^* = \beta_1^{true} + c$. Hence, the starred model satisfies all assumptions and OLS will thus be BLUE for all parameters of the *starred* model.

- To summarise, the estimated intercept $\hat{\beta}_{1,OLS}^*$ will be biased for β_1^{true} but estimators of all other coefficients, β_2^{true}, ..., β_k^{true}, should be unbiased.

Solutions:

- Always include an intercept.

Situation 4.Benign: Non-normality of errors

Causes:

- This arises when for some reasons, empirical or theoretical, we cannot assume $A5Gaussian$; perhaps we are more willing to assume some other distribution denoted by $A5specific$. (For example, LDE, Logistic, etc.)

Consequences:

- As long as the error distribution has finite moments, OLS will still be unbiased. But OLS will only be BLUE under $A1, A2linear, \geq A3Rmi, A4GM(iid)$, but not BUE. Furthermore, as $S \to \infty$ the OLS will be consistent and asympotically normally distributed provided the error distribution is "regular" (hence(it will be CUAN). However, it will *not* be the best CUAN — that optimality being reserved for the correctly specified MLE.
- MLE can be obtained from deriving and maximising the likelihood function under $A1, A2linear, \geq A3Rmi, A4GMiid$ and $A5specific$. MLE will be BUE if it can be shown to be unbiased in finite samples.
- If we are not able to show that MLE is unbiased for any S, then we can try to check if it is asymptotically unbiased as $S \to \infty$. Indeed, this will typically be true for all "regular" estimation problems.
- With In this case the MLE will be the best CUAN, linear or nonlinear, in that it will attain the CRLB. In fact, this optimality of MLE will hold even with relaxation of $A3Rmi$ to $A3Rsru$.

Solutions:

- Use MLE or OLS, as appropriate, with our choice between the two estimators depending on how credible is our $A5specific$ assumption and how large is our sample size, S.

Situation 4.Serious: Non-normality of errors

Causes:

- Much greater problems arise in regression estimation if there are reasons to believe that the disturbances follow distributions with "thick tails" like the t-distribution with very few degrees of freedom (say less than 4); or the Cauchy distribution — (which is the t-distribution with a single degree of freedom); or the Pareto distribution with tail parameter λ taking certain problematic values. For example, in financial econometrics certain asset returns have been found to be t-distributed with fewer than four degrees of freedom.

Consequences:

- Consider the extreme case of the Cauchy distribution as an illustration. Then the population odd moments of the disturbances, $E(\epsilon_s)^{2p-1}$, p positive integer, are undefined and the even moments $E(\epsilon_s)^{2p}$, are infinite. Consequently, $A2$, $A3$, and $A4GM/\Omega$ all fail because of the undefinedness/infiniteness of moments. In such case, OLS will be biased, inconsistent, and not CUAN. Consequently, OLS should be avoided in this situation.
  To see this, recall the SEV analysis of OLS:
  $$SEV(OLS) = \left(\sum_s B_s\right)^{-1} \cdot \sum_s a_s = \left(\frac{1}{S}\sum_s x_s x_s'\right)^{-1} \cdot \frac{1}{S}\sum_s x_s \epsilon_s$$
  Laws of Large numbers will *fail* on the $Ea_s = Ex_s\epsilon_s$ term since $E\epsilon_s$ is undefined and $E\epsilon_s^2$ is infinite for Cauchy distributed errors.

Solutions:

- Use appropriate MLE or other estimators that are "robust" to non-existence of finite moments. Examples are LAD or other "quantile regression" methods.
- OLS should be avoided irrespective of the sample size.

Situation $5.benign$: $A4\Omega$ instead of $A4GM$ (other assumptions holding)

Causes:

- Conditional heteroskedasticity in ε_s^{true} (e.g., Marylebone cond. heteroskedasticity model).
- Alternatively, autocorrelation in ε_s^{true} (e.g., stable $AR(1)$, or $MA(2)$ model).

Consequences:

- Non-sphericity of errors – i.e., $\mathbb{E}(\varepsilon\varepsilon'|X) = c^2\Omega \neq \sigma_\varepsilon^2 I_S$. OLS is a LUE but not the BLUE. IGLS is BLUE according to the Aitken/GM2 theorem.

Solutions:

- Option 1. Use OLS and estimate consistently $\mathbb{V}(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$. This would be called the use of OLS with "robust SEs". For example, in the conditional heteroskedasticity case, we could use the Huber-White-Eicker estimator of standard errors; in the autocorrelation case, we could use the Newey-West estimator of standard errors.
- Option 2. Use IGLS (if Ω is known), whereby $\mathbb{V}(\hat{\beta}_{IGLS}|X) = c^2(X'\Omega^{-1}X)^{-1}$. In practice, Ω is unknown, so we are forced to use FGLS using consistent estimators of the parameters on which $\Omega = \Omega(\lambda)$ depends. Since the FGLS estimator will typically be a non-linear function of ε^{true}, it will not be unbiased for finite S. Indeed, FGLS is only asymptotically BLUE.
- To summarise, use Option 2 in case of large S, and Option 1 in case of small S. In the latter case, our estimator will not be BLUE but at least it will be unbiased.

Situation 5.serious — Endogeneity: $A4\Omega$ instead of $A4GM$

Causes:

- The presence of true state dependence in our model (e.g., inclusion of a lagged dependent variable on the RHS) alongside error persistence (e.g., AR(1) autocorrelation in ε_s^{true}).
- For example, consider $y_s = \beta^{true} y_{s-1} + \varepsilon_s^{true}$, where $\text{Cov}(\varepsilon_{s-1}^{true}, \varepsilon_s^{true}) \neq 0$. We have $\text{Cov}(y_{s-1}, \varepsilon_s^{true}) = \text{Cov}(\beta^{true} y_{s-2} + \varepsilon_{s-1}^{true}, \varepsilon_s^{true}) \neq 0$ even if ε_s^{true} is news w.r.t. y_{s-1}.
- Endogeneity = forced failure of $A3Rsru$.

Consequences:

- OLS/GLS will be inconsistent. The reason is that our error, ε_s^{true} is correlated with our regressor, y_{s-1}, since both are affected by an omitted (unobserved) variable ε_{s-1}^{true}.
  We have forced a violation of $A3Rsru$.

- The intuition is that, in this situation, there exist two sources of linkage between the present and the past. The first is true state dependence whereby y_s depends on y_{s-1}. The second is error persistence whereby ε_s^{true} depends on ε_{s-1}^{true}. Since errors are unobserved, we will never be able to disentangle these two sources of dynamics from each other.

Solutions:

- Instrumental variable methods (which we will study soon) may provide a consistent estimator (feasible for $MA(q)$ processes but not for $AR(p)$ processes. Alternative estimation strategies may include Quasi MLE methods.

Situation 6.serious.OV — Endogeneity: omitted variables

Causes:

- Suppose we have
$$A2 : y = X\beta^{true} + Z\gamma^{true} + \varepsilon^{true},$$
but we mistakenly only run y on X (i.e., we omit the Z regressors). This implies that the estimating model is given by the starred version:
$$A2^* : y = X\beta^{true} + u, \quad \text{where } u = Z\gamma^{true} + \varepsilon^{true}$$

- Endogeneity = forced failure of $A3Rsru$.

Consequences:

- $A3Rsru$ will be violated and OLS and (GLS) will be inconsistent.

Solutions:

- We need to think. One option is to find the missing variables and include them. Another is to trade omitted variable bias off for measurement error bias by including a proxy for Z. Yet another option is to find an instrument for X and use IV methods instead of OLS.

- A third option is to try to sign (i.e., $\pm$) the bias of the OLS estimator (though this may not always be possible). To understand this point, let us return to the Broca example (see the next two slides).

Situation 6.serious.OV — Endogeneity: (cont'd.): the Broca example

Going by the Broca example, we have

$$x_s' = (1, G_s); \quad z_s' = (M_s, A_s),$$

where (as a reminder), we had G_s ,an identifier for male gender dummy, G_s, the body weight/-mass, and A_s, the age at death.

- The bias vector (i.e., the systematic component of the sampling error of our estimator that remains non-zero in expectation over hypothetical repeated experiments) is given by

$$\mathbb{E}(\hat{\beta}_{OLS}|X, Z) - \beta^{true} = (X'X)^{-1}X'Z\gamma^{true}.$$

- Let $\hat{\delta}_{W,OLS}$ denote the OLS estimator of the slope coefficient obtained from a regression of W_s on $x_s = (1, G_s)'$, and let $\hat{\delta}_{A,OLS}$ arise analogously from a regression of A_s on $x_s = (1, G_s)'$. Further let $\gamma^{true} = (\gamma_W^{true}, \gamma_A^{true})'$. We can then visualise the bias as:

$$\begin{aligned}\mathbb{E}(\hat{\beta}_{OLS}|X, Z) - \beta^{true} &= (X'X)^{-1}X'Z\gamma^{true} \\ &= \begin{pmatrix} \cdot & \cdot \\ \hat{\delta}_{W,OLS} & \hat{\delta}_{A,OLS} \end{pmatrix} \begin{pmatrix} \gamma_W^{true} \\ \gamma_A^{true} \end{pmatrix}\end{aligned}$$

Situation 6.serious.OV (cont'd.): the Broca example

- The bias, which is a 2×1 vector, is such that the first element relates to bias in the intercept; but the second element, bias in the coefficient of interest, is

$$\hat{\delta}_{W,OLS} \times \gamma_W^{true} + \hat{\delta}_{A,OLS} \times \gamma_A^{true}.$$

  Recall what we know about the signs:
  - $\hat{\delta}_{W,OLS} > 0$ (men are typically heavier than women)
  - $\gamma_W^{true} > 0$ (the "dinosaur effect")
  - $\hat{\delta}_{A,OLS} < 0$ (men typically die younger than women)
  - $\gamma_A^{true} < 0$ (the "drying out effect")

- The upshot is that overall bias in the estimated coefficient on G_s will be upwards/positive.

Situation 7.serious.ONL — Endogeneity: omitted non-linearities

Causes:

- The true model is $A2 : y = g(X, \beta^{true}) + \varepsilon^{true}$, but the estimating model is $A2^* : y = X\beta^{true} + u$, where $u = g(X, \beta^{true}) - X\beta^{true} + \varepsilon^{true}$
- Endogeneity = forced failure of $A3Rsru$.

Consequences:

- We have in general that $\mathbb{E}(u) \neq 0$ and, more crucially, $\mathbb{E}(u|X) \neq \mathbb{E}(u)$. In other words, $A3Rsru$ is violated.

Solutions:

- We need to identify the correct $g(X, \beta^{true})$ function and apply suitable non-linear least squares (NLLS) methods.

Situation 8.serious.ME — Endogeneity: regressor measurement errors

Causes:

- Suppose we have error-ridden versions of (a subset of) our regressors:

$$X_1 = X_1^* + V_1,$$

  where X_1 is observed whereas X_1^* and V_1 are not.

- The true model is $A2^* : y = X_1^* \beta_1^{true} + X_2 \beta_2^{true} + \varepsilon^{true}$, but the estimating model is $A2 : y = X_1 \beta_1^{true} + X_2 \beta_2^{true} + u$, where $u = \varepsilon^{true} - V_1 \beta_1^{true}$.
- Endogeneity = forced failure of $A3Rsru$.

Consequences:

- The problem in $A2$ is that both the error-ridden regressor, X_1, and the model error, u, contain V_1, the measurement error, so that $A3Rsru$ will be violated.

Solutions:

- Apply IVE if suitable instruments can be found for X_1, which is seldom possible. Alternative *ad hoc* methods exist in case additional imperfect measures of X_1^* are available.

Situation 9.serious.SE — Endogeneity: simultaneous equations

Causes:

- Consider the (dual-equation) structural model given by

$$y = Y_1\beta_1 + Z_1\gamma_1 + \varepsilon$$
$$Y_1 = Z_1\pi_{11} + Z_2\pi_{12} + \nu_1,$$

  where y or Y notation denotes an endogenous variable whereas Z notation denotes an exogenous variable. Note that structural equations jointly describe the behaviour of economic agents. There may be (as in the case above) both exogenous and endogenous variables on the RHS of such equations. In contrast, equations in which all regressors are Z variables are referred to as reduced form equations.
- Endogeneity $=$ forced failure of $A3Rsru$.

Consequences:

- Since Y_{1s} and y_s are simultaneously determined through the two equations, it follows that Y_{1s} and ε_s are correlated as long as ε_s and ν_{1s} are correlated. $A3Rsru$ is violated.

- OLS is an appropriate unbiased and consistent estimation strategy for reduced form equations, but gives biased and inconsistent results when applied to a structural equation. So the π parameters can be consistently estimated by OLS from the Y_1 equation, but the β_1 and γ_1 parameters in the y equation will not be consistently estimated.

Situation 9.serious.SE — Endogeneity: simultaneous equations (cont'd.)

Solutions:

- We can use IVE (see Topic 18 ahead), using Z_1 and Z_2 as instruments for Z_1 and Y_1.
- For IVE of β_1 and γ_1 to be feasible, a necessary (but not sufficient) condition is that the y equation be identified, which requires that there be at least as many columns in Z_2, our excluded exogenous variables, as columns in Y_1, our included endogenous regressors.

Situation 10.serious.LDVexogenous: Probit etc.

See Case 6, Topic 15: Leading Causes of Endogeneity

- Omitted nonlinearity caused by Limited Dependent Variables
- Endogeneity $=$ forced failure of $A3Rsru$.

Situation 11.serious.LDVendog: Sample Selectivity

See Case 7, Topic 15: Leading Causes of Endogeneity

- Omitted nonlinearity *plus* Simultaneity caused by Endogenous Sample Selection
- Endogeneity $=$ forced failure of $A3Rsru$.