- Topic 10. Usefulness of the exact and approximate normal linear regression model (NLRM and ANLRM respectively)

Let us briefly recap our list of estimators so far:

Estimator	Notation
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.	$\hat{\beta}_{MLE}$

Recap from Topic 7: what is the NLRM?

We previously noted in Topic 7 that...

- The assumptions $A1, A2linear, \geq A3Rmi$, and some $A4$ are collectively referred to as a "linear regression model", or the LRM.
- If we have specifically $A4GM(iid)$, the above assumptions comprise the "classical" LRM; and if we have $A4\Omega$, they comprise the "generalised" LRM.
- If we leave $A4$ as general, but add $A5Gaussian$, the assumptions are said to collectively comprise the "normal" LRM, or the NLRM hereafter.

In the next slide, we recall the key result available (for inference) under the NLRM.

Recap from Topic 7: key result available under the NLRM

Suppose we have $A1, A2linear, \geq A3Rmi$, and either $A4\Omega$ or $A4GM(iid)$, and consider any of the linear in y methods previously encountered, generically denoted by

$$\hat{\beta}_{method} = C_X y,$$

where C_X is a $k \times S$ matrix characterising the specific linear in y estimation method, so that:

For $C_X \equiv A_X \equiv (X'X)^{-1}X'$, we obtain the first estimator, $\hat{\beta}_{method} = \hat{\beta}_{OLS}$.
For $C_X \equiv B_X$ where $B_X X = I_k$, we obtain the third estimator, $\hat{\beta}_{method} = \hat{\beta}_{Lstar}$.
For $C_X \equiv A_X$, we also have the fourth estimator, $\hat{\beta}_{method} = \hat{\beta}_{GMM}$.
For $C_X \equiv B_X^* \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$, we have the fifth estimator, $\hat{\beta}_{method} = \hat{\beta}_{IGLS}$.

Let us add $A5Gaussian$ whereby $\varepsilon^{true}|X \sim N(0, \Sigma)$ for $\Sigma \equiv \sigma^2 I_S$ under $A4GM(iid)$ and $\Sigma \equiv c^2\Omega$ under $A4\Omega$. Since multivariate Gaussianity is preserved under linear transformation,

$$\hat{\beta}_{method}|X \sim N(\mathbb{E}(\hat{\beta}_{method}), \mathbb{V}(\hat{\beta}_{method}|X)),$$ or more specifically,

$$\hat{\beta}_{method}|X \sim N(\beta^{true}, C_X \Sigma C_X'),$$

under the NLRM, so long as the method considered is linear in y.

Recap from Topic 7: some more specifics under the NLRM

Estimator	Scenario 1 – $A4GM(iid)$ and $A5Gaussian$
1. $R\hat{\beta}_{OLS}\|X \sim$	$N(R\beta^{true}, \sigma^2 R(X'X)^{-1}R')$
2. $R\hat{\beta}_{LAD}\|X \sim$	$???(?, ??)$
3. $R\hat{\beta}_{Lstar}\|X \sim$	$N(R\beta^{true}, \sigma^2 R B_X B_X' R')$
4. $R\hat{\beta}_{GMM}\|X \sim$	$see\ OLS$
5. $R\hat{\beta}_{IGLS}\|X \sim$	$N(R\beta^{true}, \sigma^2 R(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}I_S\Omega^{-1}X(X'\Omega^{-1}X)^{-1}R')$
6. $R\hat{\beta}_{MLE}\|X \sim$	$see\ OLS$ **BUE**

Estimator	Scenario 2 – $A4\Omega$ and $A5Gaussian$
1. $R\hat{\beta}_{OLS}\|X \sim$	$N(R\beta^{true}, c^2 R(X'X)^{-1}X'\Omega X(X'X)^{-1}R')$
2. $R\hat{\beta}_{LAD}\|X \sim$	$???(?, ??)$
3. $R\hat{\beta}_{Lstar}\|X \sim$	$N(R\beta^{true}, c^2 R B_X \Omega B_X' R')$
4. $R\hat{\beta}_{GMM}\|X \sim$	$see\ OLS$
5. $R\hat{\beta}_{IGLS}\|X \sim$	$N(R\beta^{true}, c^2 R(X'\Omega^{-1}X)^{-1}R')$
6. $R\hat{\beta}_{MLE}\|X \sim$	$see\ IGLS$ **BUE**

(Above, R is some $q \times k$ matrix with full row rank equal to $q \leq k$.)

Recap from Topic 7: what is the ANLRM?

Suppose we are unwilling to assume $A5Gaussian$, or in fact, any $A5specific$ at all.

Obviously, this would scupper any inferential objectives we might have. However, rather than give up, we can appeal to asymptotic theory. That is, we can develop inferential techniques that are at least asymptotically valid as $S \to \infty$.

In practice, what this means is that we assume our sample size, S, is sufficiently large to justify our belief that limiting results hold at least approximately in our empirical context. (The approximation is typically thought of as only improving as the sample size increases.)

So what is "A" in the acronym "ANLRM"? Depending on how one wishes to structure the exposition, one can think of A as standing either for:

- "Asymptotic" (i.e., as the sample size passes to infinity); or
- "Approximate" (i.e., where the approximation is justified for "sufficiently large" samples).

To summarise, the ANLRM refers to the asymptotic (as $S \to \infty$) or approximate (for large S) normal linear regression model. (Asymptotic/approximate normality is ushered in by a CLT.)

Extensions from Topic 7: key result available under the ANLRM

Suppose we have $A1, A2linear, \geq A3Rsru, A4GM(iid)$ or $A4\Omega$, and $S \to \infty$ or S large.

Consider the linear-in-y estimators:

Estimator 1: $\hat{\beta}_{OLS}$ where $\hat{\beta}_{OLS} = A_X y$
Estimator 3: $\hat{\beta}_{Lstar}$ where $\hat{\beta}_{Lstar} = B_X y$ such that $B_X X = I_K$
Estimator 4: $\hat{\beta}_{GMM} = \hat{\beta}_{OLS}$
Estimator 5: $\hat{\beta}_{IGLS} = B_X^* y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ with $B_X^* X = I_K$
Estimator 6linear: $\hat{\beta}_{MLE}$ in special cases that give linear-in-y estimators (e.g., Gaussianity)
and the nonlinear-in-y estimator:
Estimator 6nonlinear: $\hat{\beta}_{nlMLE}$ in other"regular" cases that give nonlinear-in-y MLE (e.g., A5Logistic)

The ANLRM will prevail in the sense that all listed estimators (some of which nonlinear-in-y) will be distributed as:

$$\hat{\beta}_{method}|X \overset{approx.}{\sim} N(\beta^{true}, \mathbb{V}(\hat{\beta}_{method}|X)) \text{ for large } S,$$

where the expression for $\mathbb{V}(\hat{\beta}_{method}|X)$ depends on $method$ and specific form of $A4$ and $A5$, the approximation becoming better the larger is the sample size S.

- It follows under the ANLRM that we have

$$R\hat{\beta}_{method}|X \overset{approx.}{\sim} N(R\beta^{true}, R\mathbb{V}(\hat{\beta}_{method}|X)R') \text{ for large } S,$$

for some R, a $q \times k$ matrix with full row rank equal to $q \leq k$.

- As was the case with the exact NLRM, it is important to note that the inferential results alluded to in the previous bullet only hold due to preservation of multivariate Gaussianity under linear transformation. (Bear in mind that this is a special property of the multivariate Gaussian.)

- The reason we focus on the distribution of $R\hat{\beta}_{method}|X$ is because the latter is what is needed for all three fundamental aims of statistical inference (see Topic 8) including interval estimation, out-of-sample prediction, and hypothesis testing.
- The distributional results about $R\hat{\beta}_{method}$, that were exact for the NLRM for any finite sample size S, now become *approximate* for very large S under weaker conditions ($A3Rsru$ instead of $A3Rmi$, $A5$ non-Gaussian, unknown, but "regular") and even for "regular" nonlinear estimators (e.g., nonlinear MLE).

Four key facts about the ANLRM

Fact 1: Results hold even with weakest $A3$ and $A5$ *unknown but "regular"*

Recall the two fundamental distributional results we have already discussed:

Exact NLRM (for any finite S):
If $A1$, $A2linear$, $\geq A3Rmi$, $A4GM(iid)$ or $A4\Omega$, and $A5Gaussian$, then:

$$R\hat{\beta}_{method}|X) \sim N(R\beta^{true}, R\mathbb{V}(\hat{\beta}_{method}|X)R') \text{ for large } S,$$

and
ANLRM (for large/infinite S):
If $A1$, $A2linear$, $\geq A3Rsru$, $A4GM(iid)$ or $A4\Omega$, and $S \to \infty$ or S large, then:

$$R\hat{\beta}_{method}|X \overset{approx.}{\sim} N(R\beta^{true}, R\mathbb{V}(\hat{\beta}_{method}|X)R') \text{ for large S,}$$

When $S \to \infty$ or S large, (a) Strong exogeneity ($A3Rmi$ or stronger) is relaxed to Weak exogeneity ($A3Rsru$) and (b) $A5Gaussian$ is relaxed to any "regular" A5, known or unknown.

Fact 2: Results hold for MLE even with $A5nonGaussian$, provided "regular"

Suppose a linear regression model satisfies:
$A1$, $A2linear$, $\geq$ $A3Rsru$, $A4.independent$, and $A5nonGaussian$ "regular" with $pdf(y_s|X;\theta) = f(y_s, X, \theta)$
When $S \to \infty$ or S large, the MLE satisfies:

$$R\hat{\beta}_{mle}|X \overset{approx.}{\sim} N(R\beta^{true}, R\mathbb{V}(\hat{\beta}_{mle}|X)R')$$

where:

$$\mathbb{V}(\hat{\beta}_{mle}|X) = -\left(\sum_{s=1}^{S} \frac{\partial^2 \ln f(y_s, X, \theta)}{\partial\theta \cdot \partial\theta'}\right)^{-1}$$

It should be noted that in these circumstances:

1. The $\hat{\beta}_{mle}$ will be nonlinear in y. And:
2. An example of $A5nonGaussian$ that is "regular" is $A5Logistic$. In contrast, $A5DoubleExponential/Laplace$ or $A5Uniform$ do *not* satisfy "regularity" conditions.

Fact 3: The Delta method changes results to Approximate/Asymptotic, whether starting from NLRM or ANLRM

Suppose we consider a particular estimator $\hat{\beta}_{method}$ that either satisfies the Exact NLRM:

$$\hat{\beta}_{method}|X \sim N(\beta^{true}, \mathbb{V}(\hat{\beta}_{method}|X))$$

for any finite S,
or it satisfies the ANLRM:

$$\hat{\beta}_{method}|X \overset{approx.}{\sim} N(\beta^{true}, \mathbb{V}(\hat{\beta}_{method}|X))$$

for large or infinite S.
Consider a continuous nonlinear function $g(\cdot)$. In *both* cases, the Delta method will imply that:

$$g(\hat{\beta}_{method}|X) \overset{approx.}{\sim} N\left(g(\beta^{true}), \left(\frac{\partial g(\cdot)}{\partial \beta}\right) \mathbb{V}(\hat{\beta}_{method}|X) \left(\frac{\partial g(\cdot)}{\partial \beta}\right)'\right)$$

Fact 4: How to pass back and forth between the "A=Approximate for large S" and "A=Asymptotic for $S \to \infty$" results

As an illustration of the steps involved, suppose that the (weaker) conditions are satisfied for $\hat{\beta}_{ols}$ to satisfy the ANRLM, i.e.,

$$\hat{\beta}_{ols}|X \overset{approx.}{\sim} N\left(\beta^{true}, \sigma^2(X'X)^{-1}\right)$$

for large S. Since the matrix $X'X$ is defined by the summation relationship:

$$X'X \equiv \sum_{s=1}^{S} x_s x_s'$$

when the sample grows in size, the more and more data will mean bigger and bigger sum of squares, and thus $X'X$ will eventually blow up to infinity, which means its inverse will implode to 0. Therefore the VCov matrix of the OLS estimator will collapse to zero, meaning that $\hat{\beta}_{ols}$ will collapse to β^{true} as $S \to \infty$ — this is known as a *Mean-Squared-Error-consistent* estimator of β^{true}.

Given that $X'X$ explodes to infinity as the sample size grows bigger and bigger, and the inverse $(X'X)^{-1}$ implodes to 0, this makes it very intractable to work with, both computationally and theoretically. Therefore, we rewrite the matrix as:

$$(X'X) \equiv S \cdot (X'X/S)$$

and its inverse:

$$(X'X)^{-1} \equiv \frac{1}{S} \cdot (X'X/S)^{-1}$$

Therefore, we may rewrite the OLS ANLRM result as:

$$\hat{\beta}_{ols}|X \overset{approx.}{\sim} N\left(\beta^{true}, \frac{1}{S} \cdot \sigma^2 (\frac{1}{S} X'X)^{-1}\right)$$

Why did we do this? Because the averaged term $X'X/S$ will *converge* by a Law of Large Numbers (LLN) to a well-defined, finite, non-singular probability limit:

$$\frac{1}{S}X'X = \frac{1}{S}\sum_{s=1}^{S} x_s x_s' \rightarrow plim \frac{1}{S}\sum_{s=1}^{S} x_s x_s' = E x_s x_s'$$

when the DGP is identically distributed (either $i.i.d.$ or $ni.i.d.$).

Four key facts about the ANLRM

If S is sufficiently large, we can replace $\frac{1}{S}X'X$ by its probability limit to get:

$$\hat{\beta}_{ols}|X \overset{approx.}{\sim} N\left(\beta^{true}, \frac{1}{S}\cdot\sigma^2(plim\frac{1}{S}X'X)^{-1}\right) = N\left(\beta^{true}, \frac{1}{S}\cdot\sigma^2(Ex_sx_s')^{-1}\right)$$

If we subtract β^{true} from both sides, the mean changes to zero:

$$\hat{\beta}_{ols} - \beta^{true} \overset{approx.}{\sim} N\left(0, \frac{1}{S}\cdot\sigma^2(Ex_sx_s')^{-1}\right)$$

This illustrates the problem perfectly: as $S \to \infty$, $\frac{1}{S} \to 0$ and so the VCov will implode to 0. To overcome this difficulty, we multiply both sides by $\sqrt{S}$ to obtain:

$$\sqrt{S}\cdot(\hat{\beta}_{ols} - \beta^{true}) \overset{approx.}{\sim} N\left(0, \sigma^2(Ex_sx_s')^{-1}\right)$$

since the VCov is multiplied by $(\sqrt{S})^2 = S$, which cancels the $\frac{1}{S}$.
In the limit when $S \to \infty$, the approximation will become perfect and thus the *asymptotic* result will obtain:

$$\sqrt{S}\cdot(\hat{\beta}_{ols} - \beta^{true}) \overset{asymptotically}{\sim} N\left(0, \sigma^2(Ex_sx_s')^{-1}\right)$$

When we mention "statistical tables" in the slides above, we are not always referring to the table for the standard normal distribution. Indeed, the usual distributions that feature in statistical inference include also the t, χ^2 and F distributions. See a review note below (covering sampling distributions) from DrVH.

Related Distributions to the Gaussian (SKIM)

(Relevant for NLRM.$A4GM(iid)$ and NLRM.$A4\Omega$)

Preliminary Fact 1:

If $n \times 1$ random vector $z \sim N(\mu, \Sigma)$ (with VCov $n \times n$ non-singular matrix Σ), then the $n \times 1$ random vector

$$w \equiv \Sigma^{-1/2}(z - \mu) \sim N(0, I_n)$$

because w will have VCov $\Sigma^{-1/2}\Sigma^{1/2}\Sigma^{1/2\prime}\Sigma^{-1/2\prime} = I_n I_n = I_n$

Preliminary Fact 2:

If $n \times 1$ random vector $z \sim N(0, I_n)$, then the scalar r.v.

$$q \equiv z'z \sim \chi^2(n)$$

Preliminary Fact 3:

If $n \times 1$ random vector $z \sim N(0, I_n)$ and $n \times n$ matrix M is idempotent of with $\rho(M) = r < n$, then

$$q \equiv z'Mz \sim \chi^2(r)$$

Preliminary Fact 4:

If (i) r.v. $Z \sim N(0, I_n)$, (ii) r.v. $X \sim \chi^2(r)$, and (iii) Z and X are statistically independent, then the r.v.

$$\tau \equiv Z/\sqrt{X/r}$$

will be distributed as $student - t(r)$

Preliminary Fact 5:

If (i) r.v. $R_1 \sim \chi^2(r_1)$, (ii) r.v. $R_2 \sim \chi^2(r_2)$, and (iii) R_1 and R_2 are statistically independent, then the r.v.

$$f \equiv \frac{R_1/r_1}{R_2/r_2}$$

will be distributed as $F(r_1, r_2)$

Preliminary Fact 6:

If r.v. τ distributed as $student - t(r)$, then the r.v. $w \equiv \tau^2 \sim F(1, r)$

Review Quiz for Topic 10

Question 1. What are sampling distributions? Why are sampling distributions needed – i.e., from a theoretical perspective? How do empiricists use sampling distributions – i.e., from a practical perspective? Provide (one or more) example(s) of how you have personally used the standard normal and/or the t, χ^2 or F distribution in any data analysis exercise(s) that you have undertaken previously. How did you decide, in your example, that your choice of sampling distribution was the right choice?

Question 2. Suppose we have the NLRM with $A4GM(iid)$ and our OofS prediction is given by

$$\hat{y}_{S+1} = x'_{S+1}\hat{\beta}_{OLS}.$$

Show that the (conditional) prediction error variance of $\hat{y}_{S+1} - y_{S+1}$, given $S \times k$ matrix X and a $k \times 1$ vector of new observations x_{S+1}, is:

$$\mathbb{V}(\hat{y}_{S+1} - y_{S+1}|X, x_{S+1}) = (1 + x'_{S+1}(X'X)^{-1}x_{S+1})\sigma^2_\varepsilon$$

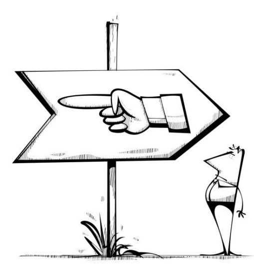
Question 3. In Question 2. above, we focussed on predicting y_{S+1}. Now, suppose we wish to predict

$$m_{S+1} = \mathbb{E}(y_{S+1}|X, x_{S+1}).$$

Consider $\hat{m}_{S+1} = x'_{S+1}\hat{\beta}_{OLS}$ and show that its (conditional) prediction error variance is:

$$\mathbb{V}(\hat{m}_{S+1} - m_{S+1}|X, x_{S+1}) = (x'_{S+1}(X'X)^{-1}x_{S+1})\sigma^2_\varepsilon.$$

Provide intuition for the differences in the (conditional) prediction error variance expressions obtained under Questions 2 and 3.



Signpost 10

[To be confirmed.]