Inference in the CLNM

Based on S1-S4, we can compute the first two moments of $\widehat{\beta}$, prove the GM theorem, and compute the first moment of $\widehat{\sigma}^2$. We now investigate the consequence of adding

- S5: $u|X \sim N$

Rks:

- S3-S5 can be written succinctly as $u|X \sim N(0, \sigma^2 I_n)$. If we add S1, we get $y|X \sim N(X\beta, \sigma^2 I_n)$.
- The normality assumption is "dubious". The following material can be thought of as an example of finite sample theory, or as a guide to some of the results that will obtain based on asymptotic arguments (i.e. approximations that improve with sample size).

In what follows, I impose S1-S5, drop the qualifier "conditional on X" and refer to the results D0-D5 from the handout "Inference in the Classical Linear Normal Model".

Theorem R1: $\widehat{\beta}$ and $\hat{u}$ are joint normally distributed.

Proof:

$$\begin{bmatrix} \widehat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix} + \begin{bmatrix} L \\ M \end{bmatrix} u$$

where u is MVN. Use D1.

Theorem R2: $\widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$

Proof: We've already established the first two moments using S1-S4. From the theorem above, we get that the marginal of $\widehat{\beta}$ is MVN (apply $[I \ \ 0]$ to LHS).

Theorem R3: $\hat{u}'\hat{u}/\sigma^2 \sim \chi^2(n-K)$

Proof:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{u}{\sigma}'M\frac{u}{\sigma} = z'Mz$$

where $z \sim N(0, I)$. But by D3, $z'Mz \sim \chi^2(tr(M))$.

Theorem R4: $\hat{\beta}$ and $\hat{\sigma}^2$ are independent.

Proof:

$$E((\hat{\beta} - \beta)\hat{u}') = Luu'M = \sigma^2 LM = 0$$

So by R1 and D0, $\hat{\beta}$ and $\hat{u}'$ are independent. It follows that $\hat{\beta}$ and every function of $\hat{u}$ are also independent.

Theorem R5: $\hat{\beta}$ achieves the Cramer-Rao Lower Bound.

Rk: This means that under S1-S5, $\hat{\beta}$ has minimum variance in the class of all unbiased estimators (not just linear ones).

This provides a substantial improvement over the GM theorem.

Definition: A r.v. has a *Student's t-distribution* if it can be expressed as the ratio of a standard normal r.v. over the square root of an independent χ^2 r.v. divided by its d.f. Symbolically,

$$t(m) = \frac{z}{\sqrt{\chi^2(m)/m}}$$

where z and $\chi^2(m)$ are independent.

Definition: A r.v. has a *Snedecor's F-distribution* if it can be expressed as the ratio of two independent χ^2 r.v.'s each divided by its d.f. Symbolically,

$$F(m_1, m_2) = \frac{\chi_1^2(m_1)/m_1}{\chi_2^2(m_2)/m_2}$$

Theorem: $(\hat{\beta}_j - \beta_j)/s.e.(\hat{\beta}_j) \sim t(n-K)$

Proof:

$$\frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2((X'X)^{-1})_{jj}}}$$

$$= \frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2((X'X)^{-1})_{jj}}}}{\sqrt{\frac{\hat{u}'\hat{u}}{\sigma^2}/(n-K)}}$$

The numerator is a standard normal r.v., and the denominator is the square root of an indepent χ^2 r.v. divided by its d.f.

Before stating the next theorem, define

$V(R\hat{\beta}) = \sigma^2 R(X'X)^{-1}R'$ and $\hat{V}(R\hat{\beta}) = \hat{\sigma}^2 R(X'X)^{-1}R'$. Assume $R \in \mathbb{R}^{qxK}$, $r \in Sp(R)$, and $rank(R) = q$.

Theorem: Under $H_o : R\beta = r$,

$$(R\hat{\beta} - r)' \left[\hat{V}(R\hat{\beta}) \right]^{-1} (R\hat{\beta} - r)/q \sim F(q, n-K)$$

Proof: By D1

$$R\hat{\beta} - r \sim N(R\beta - r, V(R\hat{\beta}))$$

So, under H_o, $(R\hat{\beta} - r)' \left[V(R\hat{\beta}) \right]^{-1} (R\hat{\beta} - r) \sim \chi^2(q)$ by D2. $\therefore$

$$(R\hat{\beta} - r)' \left[\hat{V}(R\hat{\beta}) \right]^{-1} (R\hat{\beta} - r)/q = \frac{(R\hat{\beta} - r)' \left[V(R\hat{\beta}) \right]^{-1} (R\hat{\beta} - r)/q}{\hat{\sigma}^2/\sigma^2}$$

is the ratio of two indendent χ^2 r.v.s each divided by its d.f.

:Testing hypotheses about a single parameter

Suppose we wish to test the simple null hypothesis $H_o : \beta_j = c$, where $c \in \mathbb{R}$.

Step 1. Determine the distribution of a test statistic under the null.

Given $(\widehat{\beta}_j - \beta_j)/s.e.(\widehat{\beta}_j) \sim t(n-K)$ where K denotes the total number of regressors (including the intercept), we see that under the null

$$\frac{\widehat{\beta}_j - c}{s.e.(\widehat{\beta}_j)} \sim t(n-K)$$

Rks:

- We like to choose a test statistic with known distribution.
- In theory, we should choose test statistics that do not involve a loss of information relative to the whole sample (i.e.

sufficient statistics) so as to achieve high efficiency (see below).

Step 2. Choose a *significance level* α.

Rk: Conventional choices are $\alpha \in \{.1, .05, .01\}$

Step 3. Choose an acceptance for the test statistic of Step 1 (⇔critical region) such that Pr(rejecting $H_o|H_o$ is true)$\leq \alpha$.

Rks:

- The *size* of a test is sup Pr(rejecting $H_o|H_o$ is true). For most tests, the size and significance level are the same.
- The *power* of a test is Pr(rejecting $H_o|H_o$ is false)
- The best choice of an acceptance/critical region given the significance level is one that maximizes the the power. The solution depends on the alternative hypothesis.

3a. Simple alternative: $H_1 : \beta_j = c'$

By the Neyman-Pearson Lemma, there's a test that maximizes power given the size. It's based on ordering values of the relative likelihood of the (sample) test statistic $\lambda = L_0/L_1$ and including all points with $\lambda \leq \lambda^*$ in the acceptance region, with λ^* chosen to satisfy the significance level constraint.

3b. Composite one sided-alternative $H_1 : \beta_j > c$

Under H_1, the test statistic $(\widehat{\beta}_j - c)/s.e.(\widehat{\beta}_j)$ will tend to be positive (more precisely, positive values of this test statistic are more likely under each point in the alternative than under the null; negative values are more likely under the null). So the critical region is chosen to reject if the realized value of the test statistic is a large, positive number.

The form of the test is to choose a *critical value*, t_α^*, and reject if the test statistic exceeds the critical value, where t_α^* is chosen to satisfy

$$\Pr\left(\frac{\widehat{\beta}_j - c}{s.e.(\widehat{\beta}_j)} \geq t_\alpha^* \right) = \alpha$$

Ex: Suppose we estimate the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$$

using $n = 32$ observation. We'll have $32 - 4 = 28$ d.f. Here's a table of conventional significance levels and the associated critical values for a one-sided alternative:

α	t_α^*
.10	1.321
.05	1.717
.01	2.508

- If the realized value of the test statistic is 1.9, we would say "H_0 is rejected at the 5% level, but not at the 1% level"
- If the realized value of the test statistic is 2.7, we would say "H_0 is rejected at the 1% level"
- If the realized value of the test statistic is 0.9, we would say "H_0 is not rejected at the 10% level"

Rks:

- Some authors use "accepted" as a synonym for "not rejected", others take great offence at this usage.

- Because the form of the best test is the same for all points under the alternative, this test is called *uniformly most powerful*, or UMP.
- If we had the composite null $H_o : \beta_j \leq c$ versus the composite one-sided alternative $H_1 : \beta_j > c$, we would follow exactly the procedure above. Intuitively, the point $\beta_j = c$ is the most difficult point under the null to reject, so we use it to "represent" the null.
- If the composite one-sided alternative is $H_1 : \beta_j < c$, we would proceed as above, but change the sign on the critical value (i.e. reject if the realized value of the test statistic is a large negative number).

3c. Composite two sided-alternative $H_1 : \beta_j \neq c$

Both large positive and large negative values of the test statistic are much more likely under the alternative than under the null. So the form of the critical region will be to reject is the realized test statistic has absolute value that exceeds a critical value $t^*_{\alpha/2}$ that is chosen to satisfy

$$\Pr\left(abs\left(\frac{\hat{\beta}_j - c}{s.e.(\hat{\beta}_j)} \right) \geq t^*_{\alpha/2} \right) = \alpha$$

Using our example from section 3b above, we can expand the table to

α	t^*_α	$t^*_{\alpha/2}$
$.10$	1.321	1.717
$.05$	1.717	2.074
$.01$	2.508	2.819

So if the realized value of the test statistic is -1.9, we say "we reject H_0 at the 10% level but not at the 5% level".

Definition: A test is called unbiased if the probability of rejection under each point in the null is always less than the probability of rejection under each point in the alternative.

Rks:

- The test for a two sided alternative described above is UMP if we restrict attention to unbiased tests (i.e. it’s UMPU)

On tests, I expect you to give me the following information

1. State explicitly H_0 and H_1
2. Choose a test statistic and give its distribution under H_0
3. Compute the realized value of the test statistic
4. List the appropriate critical values at the conventional levels and state if the test statistic is significant or insignificant at the standard levels.

Rks:

- If not explicitly stated in the question, you should take the two-sided alternative as the default.

: Computing p-values

Suppose we compute the value of the significance level, call it α^*, at which we are just indifferent between accepting or rejecting H_0. We call α^* the p-value of the test.

For example, suppose the test statistic follows a $t(40)$ distribution, and its realized value is given by

$$abs\left(\frac{\hat{\beta}_j - c}{s.e.(\hat{\beta}_j)} \right) = 1.85$$

Using

$$\Pr\left(abs\left(\frac{\hat{\beta}_j - c}{s.e.(\hat{\beta}_j)} \right) \geq 1.85 \right) = 2(.0359) =.0718$$

we get a p-value of 7.18%

Rks:

- Choosing significance levels seems arbitrary. Reporting p-values is a way of letting the reader choose. For every significance level below the p-value, H_0 is not rejected. For every significance level above the p-value, H_0 is rejected.
- It's easy to compute p-values on a computer.
- With only tables available (as during tests), I prefer that you follow my procedure to approximate the p-value (i.e. say something like "we reject at the 10% but not at the 5% level", rather than just "we don't reject at the 5% level").

:Statistical versus Economic significance

- With a large enough sample, our test statistic will become arbitrarily large for any point under the alternative (power goes to 1 as sample size grows to infinity), i.e.

$$abs\left(\frac{\widehat{\beta}_j - c}{s.e.(\widehat{\beta}_j)}\right) \uparrow \infty \text{ w.p.1 for all } \beta_j \neq c$$

  so even alternatives arbitrarily close to *c* will lead to *statistical* rejection. This doesn't mean that the rejection is *economically significant*. For example, you may find that you can reject the hypothesis that returns are unpredictable ($\beta = 0$), but that doesn't mean necessarily that the predictable returns are large enough for you to make enough money to justify, say, trading costs.

- Don't forget that failure to reject has to be interpreted similarly. If we have very good power against alternatives $\beta_j = (1 + \epsilon)a$, where ϵ is small, then we would have much more confidence in H_0 then we would if our test has poor power even against distant alternatives (see the discussion on confidence intervals below).

:Confidence Intervals

Definition: A confidence interval gives an interval (rather than a point) estimate of a parameter.

To construct a confidence interval, we begin with a probability statement such as

$$\Pr\left(-t^*_{\alpha/2} \le \frac{\widehat{\beta}_j - \beta_j}{s.e.(\widehat{\beta}_j)} \le t^*_{\alpha/2}\right) = 1-\alpha$$

Then we manipulate this statement to obtain

$$\Pr\left(\widehat{\beta}_j - t^*_{\alpha/2} \bullet s.e.(\widehat{\beta}_j) \le \beta_j \le \widehat{\beta}_j + t^*_{\alpha/2} \bullet s.e.(\widehat{\beta}_j)\right) = 1-\alpha$$

This gives us a *random interval* (i.e. the endpoints are random) that will cover β_j $(1-\alpha)$% of the time (i.e. in repeated samples, β_j will be in the random interval $(1-\alpha)$% of the time).

If we replace the random endpoints with their realizations, we get a $(1 - \alpha)\%$ confidence interval.

Rks:

- Suppose the computed CI is $[-1.2, 2.7]$. We can't say "The probability that β_j falls in the interval $[-1.2, 2.7]$ is $1 - \alpha$. Either β_j falls in the interval or it does not, so the correct probability is either 0 or 1! The phrase "confidence interval" refers to our confidence in the *procedure* used to construct the interval.
- There is a close link between a CI and hypothesis testing. If $c \in \left[\widehat{\beta}_j \pm t^*_{\alpha/2} \cdot s.e.(\widehat{\beta}_j)\right]$ then we won't reject $H_0 : \beta_j = c$ at the $\alpha\%$ level; otherwise, we will.

- A short CI says that we have accurately estimated β_j; we feel very differently about not rejecting H_0 in such a situation then we would if the CI was very wide.
- I've shown you how to construct the shortest CI for β_j; other choices are possible and sometimes more interesting (eg. a one-sided CI).

:Testing hypotheses about a linear combination $\omega'\beta$

Suppose we wish to test $H_0 : \omega'\beta = a$ vs $H_1 : \omega'\beta < a$ where $\omega \in \mathbb{R}^K$

Under the null,

$$\frac{\omega'\widehat{\beta} - a}{s.e.(\omega'\widehat{\beta})} \sim t(n-K)$$

Therefore

$$\Pr\left(\frac{\omega'\widehat{\beta} - a}{s.e.(\omega'\widehat{\beta})} \leq -t_\alpha^*\right) = \alpha$$

so we can proceed just as we did with a single coefficient (the same is true for two sided tests, composite one-sided null versus composite one-sided alternative, etc.).

Recall $\omega' = (\omega_1, \omega_2, \cdots, \omega_K)$

so $\omega'\beta = \omega_1\beta_1 + \omega_2\beta_2 + \cdots + \omega_K\beta_K$

Two issues

:Examples of $\omega'\beta = a$

(i) $H_0 : \beta_1 = 0 \Leftrightarrow H_0 : \omega'\beta = a$

for $\omega' = (1, 0, 0, \ldots, 0)$ and $a = 0$

(ii) $H_0 : \beta_1 = \beta_2 \Leftrightarrow H_0 : \omega'\beta = a$

for $\omega' = (1, -1, 0, \ldots, 0)$ and $a = 0$

(iii) $H_0 : \beta_1 + \beta_2 = 1 \Leftrightarrow H_0 : \omega'\beta = a$

for $\omega' = (1, 1, 0, \ldots, 0)$ and $a = 1$

:Computing $s.e.(\omega'\widehat{\beta})$

$$\widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

$$\therefore \ \omega'\widehat{\beta} \sim N(\omega'\beta, \sigma^2 \omega'(X'X)^{-1})\omega$$

So $s.e.(\omega'\widehat{\beta}) = \left[\omega'\widehat{V}(\widehat{\beta})\omega\right]^{1/2}$

For example, if $\omega' = (1, -1, 0, \ldots, 0)$, then

$\omega'\widehat{V}(\widehat{\beta})\omega = \widehat{var}(\widehat{\beta}_1) - 2cov(\widehat{\beta}_1, \widehat{\beta}_2) + \widehat{var}(\widehat{\beta}_2)$

:An alternative procedure for testing l.c.

Suppose we wish to test $H_0 : \beta_1 = \beta_2$ vs. $H_0 : \beta_1 < \beta_2$.

Define $\theta_1 = \beta_1 - \beta_2$. Clearly, the original null is equivalent to $H_0 : \theta_1 = 1$ vs. $H_0 : \theta_1 < 0$.

Now, rewrite the model as

$$y_i = (\theta_1 + \beta_2)x_{1i} + \beta_2 x_{2i} + \cdots \beta_K x_{Ki} + u_i \qquad \Leftrightarrow$$

$$y_i = \theta_1 x_{1i} + \beta_2 \widetilde{x}_{2i} + \cdots \beta_K x_{Ki} + u_i$$

where $\widetilde{x}_{2i} \equiv x_{2i} + x_{1i}$. So testing the linear combination reduces to testing a single coefficient.

Note also that if we want to *impose* the restriction $\beta_1 = \beta_2$, we could do so by dropping x_1 from the regression.

:Testing Multiple Linear Restrictions

Suppose we wish to test

$$H_0 : R\beta = r$$

$$\text{vs } H_1 : R\beta \neq r$$

where $R \in \mathbb{R}^{qxK}$, $rank(R) = q$, and $r \in Sp(R)$.

- Each row of R and r corresponds to a single linear restriction of the form $H_0 : \omega' \beta = a$,

$$R\beta = r \Leftrightarrow \begin{cases} R_{11}\beta_1 + R_{12}\beta_2 + \cdots + R_{1K}\beta_K & = r_1 \\ R_{21}\beta_1 + R_{22}\beta_2 + \cdots + R_{2K}\beta_K & = r_2 \\ \vdots & \vdots \\ R_{q1}\beta_1 + R_{q2}\beta_2 + \cdots + R_{qK}\beta_K & = r_K \end{cases}$$

- $rank(R) = q \Leftrightarrow$"no redundant restrictions", i.e. we *don't* have

$$H_0 : \beta_1 = 0, \beta_2 = 0, \beta_1 + \beta_2 = 0$$

- $r \in Sp(R) \Leftrightarrow$ restrictions are consistent so there is a solution, i.e. we rule out

$$H_0 : \beta_1 = 5, \beta_2 = 3, \beta_1 + \beta_2 = 4$$

- There are many algebraically equivalent ways to express the restrictions embodied in $H_0 : R\beta = r$. The geometric approach shows that it doesn't matter which "basis" we pick to represent the restrictions, we'll get the same results.
- The case where the null maintains some l.c. hold with equality while others hold with inequality does not have an obviously best test and will not be considered here.

:Test statistics

Under the null, $R\beta = r$

$$R\widehat{\beta} \sim N(r, RV(\widehat{\beta})R') \qquad \text{where } V(\widehat{\beta}) = \sigma^2(X'X)^{-1}$$

$$\therefore \left(R\widehat{\beta} - r\right)' \left[RV(\widehat{\beta})R'\right]^{-1} \left(R\widehat{\beta} - r\right) \sim \chi^2(q)$$

Rk:

- If σ^2 is unknown, an asymptotic approach (Wald Test) replaces $V(\widehat{\beta})$ with a consistent estimator (see next chapter)
- In the CLNM, if σ^2 is unknown, an exact finite sample test statistic can be constructed because

$$\frac{\left(R\widehat{\beta} - r\right)' \left[RV(\widehat{\beta})R'\right]^{-1} \left(R\widehat{\beta} - r\right)/q}{(n-K)\widehat{\sigma}^2/\sigma^2/(n-K)}$$

is obviously the ratio of two independent χ^2 r.v.'s each divided by its d.f. Therefore

$$(**) \quad \left(R\widehat{\beta} - r\right)'\left[R\widehat{V}(\widehat{\beta})R'\right]^{-1}\left(R\widehat{\beta} - r\right)/q \sim F(q, n-K)$$

where $\widehat{V}(\widehat{\beta}) = \widehat{\sigma}^2(X'X)^{-1}$

:Restricted Least squares estimator (RLS)

Define the (RLS) estimator as

$$\widehat{\beta}_* = \arg \min_{\widetilde{\beta} \in \mathbb{R}^K : R\widetilde{\beta} = r} (y - X\widetilde{\beta})'(y - X\widetilde{\beta})$$

where $rank(R) = q$ and $r \in Sp(R)$. The solution is

$$\widehat{\beta}_* = \widehat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\widehat{\beta})$$

Rks:

- This form of the estimator is more useful for deriving properties than for computation.
- Let $\widehat{y}_* = X\widehat{\beta}_*$ and $\widehat{u}_* = y - \widehat{y}_*$
- Easy to show that if the restrictions are valid then

$$E(\widehat{\beta}_*|X) = \beta$$

- Exercise: Compute $V(\widehat{\beta}_*|X)$ and compare to $V(\widehat{\beta}|X)$

Proof of the formula for $\widehat{\beta}_*$ (for your enjoyment)

Theorem: Suppose Ω is a p.d. matrix, R has full column rank, and $\theta \in Sp(R)$. Then

$$\min_{s.t.\ Rc=\theta} c'\Omega c = \theta'[R\Omega^{-1}R']^{-1}\theta$$

and is achieved at $c_* = \Omega^{-1}R'[R\Omega^{-1}R']^{-1}\theta$

Proof:

$$Rc_* = R\Omega^{-1}R'[R\Omega^{-1}R']^{-1}\theta = \theta$$

so c_* satisfies the constraint. It yields the value

$$\begin{aligned} c_*'\Omega c_* &= \theta'[R\Omega^{-1}R']^{-1}R\Omega^{-1}\Omega\Omega^{-1}R'[R\Omega^{-1}R']^{-1}\theta \\ &= \theta'[R\Omega^{-1}R']^{-1}\theta \end{aligned}$$

Let $\Omega^{1/2}$ be a square root matrix for Ω, i.e. $\Omega = (\Omega^{1/2})'\Omega^{1/2}$.
Let $\tilde{c}$ be any other solution of $Rc = \theta$.

Define

$$Y_1 = \Omega^{1/2}\tilde{c}$$
$$Y_2 = \Omega^{1/2}c_*$$

By Cauchy-Schwartz,

$$Y_1' Y_2 \le \|Y_1\| \bullet \|Y_2\|$$
$$\Leftrightarrow \tilde{c}'\Omega c_* \le (\tilde{c}'\Omega\tilde{c})^{1/2}(c_*'\Omega c_*)^{1/2}$$

But

$$\tilde{c}'\Omega c_* = \tilde{c}' R' [R\Omega^{-1}R']^{-1}\theta$$
$$= \theta' [R\Omega^{-1}R']^{-1}\theta = (c_*'\Omega c_*)^{1/2}$$

Therefore

$$c_*'\Omega c_* \le \tilde{c}'\Omega\tilde{c} \qquad \forall \tilde{c} : R\tilde{c} = \theta$$

Application:

$$\begin{aligned}\widehat{\beta}_* &= \arg \min_{\widetilde{\beta}\in\mathbb{R}^K : R\widetilde{\beta}=r} (y - X\widetilde{\beta})'(y - X\widetilde{\beta}) \\ &= \arg \min_{\widetilde{\beta}\in\mathbb{R}^K : R\widetilde{\beta}=r} \widehat{u}'\widehat{u} + (X\widehat{\beta} - X\widetilde{\beta})'(X\widehat{\beta} - X\widetilde{\beta}) \\ &= \arg \min_{\widetilde{\beta}\in\mathbb{R}^K : R(\widetilde{\beta}-\widehat{\beta})=r-R\widehat{\beta}} (\widehat{\beta} - \widetilde{\beta})'X'X(\widehat{\beta} - \widetilde{\beta})\end{aligned}$$

So apply the theorem above with $\Omega = X'X$, $c = \widetilde{\beta} - \widehat{\beta}$, $\theta = r - R\widehat{\beta}$, to get formula for RLS given above.

:Alternative forms of the test statistic $(**)$

$$
\begin{aligned}
\widehat{u}_*'\widehat{u}_* &= \widehat{u}'\widehat{u} + (\widehat{y} - \widehat{y}_*)'(\widehat{y} - \widehat{y}_*) \\
&= \widehat{u}'\widehat{u} + (X\widehat{\beta} - X\widehat{\beta}_*)'(X\widehat{\beta} - X\widehat{\beta}_*) \\
&= \widehat{u}'\widehat{u} + (\widehat{\beta} - \widehat{\beta}_*)'X'X(\widehat{\beta} - \widehat{\beta}_*) \\
&= \widehat{u}'\widehat{u} + \left(R\widehat{\beta} - r\right)'\left[R(X'X)^{-1}R'\right]^{-1}\left(R\widehat{\beta} - r\right)
\end{aligned}
$$

$$
\therefore\ (**) \ \equiv\ \left(R\widehat{\beta} - r\right)'\left[R\widehat{V}(\widehat{\beta})R'\right]^{-1}\left(R\widehat{\beta} - r\right)/q
$$

$$
= \frac{\left(\widehat{u}_*'\widehat{u}_* - \widehat{u}'\widehat{u}\right)/q}{\widehat{u}'\widehat{u}/(n-K)} \equiv (***)
$$

Rk: The approach that leads to $(**)$ generalizes to the case where $V(\widehat{\beta}) = \sigma^2\Omega$ with $\Omega \neq I$ but not $(***)$

If we have "intercepts" (i.e. $\iota \in Sp(X)$) under both the restricted and unrestricted models, then we can write

$$(***) = \left[\frac{\frac{\hat{u}_*'\hat{u}_*}{y'Ay} - \frac{\hat{u}'\hat{u}}{y'Ay}}{\frac{\hat{u}'\hat{u}}{y'Ay}} \right] \left[\frac{n-K}{q} \right]$$

$$= \left[\frac{(1-R_*^2) - (1-R^2)}{(1-R^2)} \right] \left[\frac{n-K}{q} \right]$$

$$= \left[\frac{R^2 - R_*^2}{(1-R^2)} \right] \left[\frac{n-K}{q} \right] \equiv (****)$$

Rk: The interpretation of the test statistic for $(****)$ is easy, but it is the least generalizable.

:Computing RLS estimator and its covariance matrix

The general solution to $R\beta = r$ is given by

$$\beta = \beta_0 + C\theta$$

where β_0 is any particular solution (i.e. satisfies $R\beta_0 = r$) and the columns of C form a basis for the null space of R (i.e. $Rz = 0$ iff $z = C\theta$ for some θ). Write

$$\begin{aligned} y &= X\beta + u \\ &= X(\beta_0 + C\theta) + u \end{aligned}$$

Because β_0 is known, we can subtract $X\beta_0$ from both sides of the equation to obtain

$$y - X\beta_0 = XC\theta + u \Leftrightarrow \widetilde{y} = \widetilde{X}\theta + u$$

The transformed model satisfies MLR.1-MLR.6 and θ is unrestricted, so we can estimate it by OLS. Then compute $\widehat{\beta}_* = \beta_0 + C\widehat{\theta}$, and $V(\widehat{\beta}_*) = CV(\widehat{\theta})C'$

:Examples

1. Suppose we wish to test $H_0 : \beta_1 = 0.$ This corresponds to the choices If $R = (1, 0, 0, \ldots, 0)$ and $r = 0.$ Then $(**)$ is given by

$$\frac{\left(R\widehat{\beta} - r\right)' [R(X'X)^{-1}R']^{-1} \left(R\widehat{\beta} - r\right)/1}{\widehat{\sigma}^2}$$

$$= \frac{\widehat{\beta}_i^2}{\widehat{\sigma}^2 [(X'X)^{-1}]_{ii}} = \left(\frac{\widehat{\beta}_i}{s.e.(\widehat{\beta}_i)} \right)^2$$

So the F-test statistic for the null is exactly the square of the t-test statistic for the same null (against a two-sided alternative).

2. Suppose the model is (W 4.31)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_5 x_{i5} + u_i$$

OLS estimation with $n = 353$ yields $SSR = 183.186$ and $R^2 = .6278$. (NOTE: $K = 6$). We wish to impose the restrictions $\beta_3 = \beta_4 = \beta_5 = 0$. The restricted model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Estimating the restricted model by OLS yields $SSR_* = 198.311$ and $R_*^2 = .5971$. Using the form $(***)$ of the test statistic, we see that the realized value is

$$\frac{(198.311 - 183.186)/3}{183.186/(353 - 6)} = 9.55$$

Large values of the test statistic lead to rejection. Critical values for the $F(3, 347)$ are not provided in the table. But using $Pr(F(3, 347) \geq 3.95) < Pr(F(3, 120) \geq 3.95) = .01$, we conclude that we reject at the 1% level.

3. Testing for overall significance

Suppose we have a model with an intercept and we wish to test if all the coefficients except the intercept are zero. The restricted model is

$$y = \beta_0 + u$$

Clearly $SSR_* \equiv \hat{u}_*'\hat{u}_* = y'Ay$ where $A = I - (1/n)\iota\iota'$ and $R_*^2 = 0$. Therefore, using the form $(****)$, the test statistic reduces to

$$\left[\frac{R^2}{(1-R^2)}\right]\left[\frac{n-K}{K-1}\right]$$

Rk: STATA (and most other packages) routinely report this F-statistic as part of their OLS results.