Economics 240A, Section 4: Goldberger Ch.'s 19-22

Michael R. Roberts
Department of Economics and
Department of Statistics
University of California at Berkeley

April 21, 2000

1 Introduction

This handout reviews some of the key points regarding chapters 19-22 in Goldberger.

2 CNR Framework (σ^2 Known)

The idea now is that we add a distributional assumption to the CR framework. This allows us to conduct statistical inference (confidence intervals and hypothesis testing). The assumptions are now:

1. $y\tilde{\ } MVN\left(X\beta, \sigma^2 I\right)$
2. X nonstochastic and full rank.

Note that this is almost the same as the classical regression framework except for the normality assumption since

$$\begin{aligned} E(y) &= X\beta \\ V(y) &= \sigma^2 I \end{aligned}$$

2.1 Sampling Distributions

Let's consider the implied distributions for the OLS estimator b and corresponding sum of square residuals $e'e$.

1. Claim:

$$b\tilde{\ }MVN\left(\beta, \sigma^2 \left(X'X\right)^{-1}\right)$$

Proof:

$$b = \left(X'X\right)^{-1} X'y$$

b is a linear combination of the $y's$ which are $N\left(X\beta, \sigma^2 I\right)$. This implies the $b's$ are normal with expectation

$$\begin{aligned} E(b) &= E\left\{\left(X'X\right)^{-1} X'y\right\} \\ &= \left(X'X\right)^{-1} X'E\{y\} \\ &= \left(X'X\right)^{-1} X'X\beta \\ &= \beta \end{aligned}$$

and variance covariance matrix

$$\begin{aligned} V(b) &= V\left(\left(X'X\right)^{-1} X'y\right) \\ &= \left(X'X\right)^{-1} X'V(y) X \left(X'X\right)^{-1} \\ &= \left(X'X\right)^{-1} X'\sigma^2 I X \left(X'X\right)^{-1} \\ &= \sigma^2 \left(X'X\right)^{-1} X'X \left(X'X\right)^{-1} \\ &= \sigma^2 \left(X'X\right)^{-1} \end{aligned}$$

The key assumption here is that: σ^2 is known. If it isn't we get a Student's $t-$distribution.

Note that any nonstochastic linear combination of the parameter vector, Hb, will be normal with expectation $H\beta$ and variance $\sigma^2 H\left(X'X\right)^{-1} H'$ (assuming $H \in R^{p\times k}$ and $\rho(H) = p$).

2. Claim:

$$e'e/\sigma^2\tilde{\ }\chi^2_T$$

Proof: We'll use the general result that if $y \in R^n$ is distributed $MVN(\mu, \Sigma)$ then

$$(y-\mu)' \Sigma^{-1} (y-\mu)\tilde{\ }\chi^2_n$$

Since the residual vector has expectation 0,

$$\begin{aligned} e'e &= (y - X\beta)' (y - X\beta) \\ &= (y - X\beta)' \left[\sigma^2 I\right]^{-1} (y - X\beta) \times \sigma^2 \end{aligned}$$

So, $e'e/\sigma^2\tilde{\ }\chi^2_T$.

2.2 Confidence Intervals

In the CNR framework with σ^2 known, we form a confidence interval as

$$t \pm c\sigma_t$$

where $t = h'b$ is our estimated statistic, c is the appropriate critical value from the normal distribution (e.g. 1.96 for a 95% confidence interval, 1.00 for a 68% confidence interval, etc.) and $\sigma_t = \sqrt{h'V(b)h}$ is the standard error of t.

This set-up subsumes the more basic idea of a confidence interval for one parameter b_j. In that case, h is a vector of all $0's$ except for a 1 in the j^{th} position.

2.3 Joint Confidence Regions

We've got an unknown parameter vector $\theta = H\beta$ and we estiamte a sample value $t = Hb$ (we continue to assume knowledge of σ^2 which is an important assumption). From the results above

$$(t-\theta)'\left[\sigma^2 H\left(X'X\right)^{-1}H'\right]^{-1}(t-\theta)\,\tilde{}\,\chi^2_p$$

where p is the rank of the matrix H.(i.e. it's the number of linear restrictions). To form a confidence region for θ we would set

$$(t-\theta)'\left[\sigma^2 H\left(X'X\right)^{-1}H'\right]^{-1}(t-\theta) \leq c_p$$

where c_p is the critical value from the χ^2_p distribution. That is c_p is the number such that the area to the left of c_p under the χ^2_p pdf is equal to the relevant percentage. As a concrete example, consider a 95% confidence interval where the rank of H is 2. c_p would be $c_2 = 5.99$.

Note that $(t-\theta)'\left[\sigma^2 H\left(X'X\right)^{-1}H'\right]^{-1}(t-\theta)$ can be written more generally as $(t-\theta)'\left[V(t)\right]^{-1}(t-\theta)$.

Exercise 19.1: The CNR model applies with $k = 4, X'X = I, \sigma^2 = 2$, and $\beta = 0$. Let $t = b'b$. Find the number $c : \Pr(t > c) = 0.10$.

$$b'b = \sigma^2\left\{b'\left[\sigma^2 I\right]^{-1}b\right\}$$

The term in brackets is distributed χ^2_4 so we need to find the c :

$$\Pr\{t > 2c\} = 0.10$$

Using the χ^2 table and the fact that $\Pr\{t \leq 2c\} = 0.90$, we get $2c = 7.78$ or $c = 3.89$.

2.4 Hypothesis Testing

2.4.1 Univariate

Consider testing whether a particular parameter,β_j, is equal to β_j^0. The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \beta_j = \beta_j^0 \\ H_1 &: \beta_j \neq \beta_j^0 \end{aligned}$$

Our test is a simple two-tail z-test,

$$z = \frac{b_j - \beta_j^0}{\sigma_j} \tilde{} N(0,1)$$

Assuming our significance level is 5%, if $|z| > 1.96$, then we reject the null hypothesis $H_0 : \beta_j = \beta_j^0$. If $|z| \leq 1.96$, then we fail to reject the null.

We can just as easily test a linear combination of parameters with

$$\frac{(t - \theta^0)}{\sigma_t} \tilde{} N(0,1)$$

where $t = hb$ and $\sigma_t = \sqrt{V(t)} = \sqrt{h'V(b)h}$.

Example: Consider the following model

$$y = x_1\beta_1 + x_2\beta_2 + \varepsilon$$

under the assumptions of the CNR model. We want to test:

$$\begin{aligned} H_0 &: \beta_1 + \beta_2 = 1 \\ H_1 &: \beta_1 + \beta_2 \neq 1 \end{aligned}$$

Then

$$\begin{aligned} h &= (1,1)' \\ b &= (b_1, b_2) \\ \theta^0 &= 1 \end{aligned}$$

2.4.2 Multivariate

What about testing a set of parameters? We need a joint null hypothesis about β. Let $\theta = H\beta$ where H is a non-random $p \times k$ matrix with rank p (i.e. p linear restrictions on the parameters). The hypotheses are

$$\begin{aligned} H_0 &: \theta = \theta^0 \\ H_1 &: \theta \neq \theta^0 \end{aligned}$$

where θ^0 is a vector of hypothesized values (numbers).

Consider testing at the 5% significance level. We will accept the null (or more accurately fail to reject the null) if θ^0 lies within the 95% confidence region for θ :

$$w = (\theta - t)' [V(t)]^{-1} (\theta - t) \leq c_p$$

and reject otherwise. Here, $t = Hb$ while c_p is the 5% critical value from the χ^2_p table. We can equivalently think about rejecting the null if $w > c_p$ and accepting the null if $w \leq c_p$.

Example: Consider the following model

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \varepsilon$$

under the assumptions of the CNR model. We want to test:

$$\begin{aligned} H_0 &: \beta_1 = 2; \beta_2 - 2\beta_3 = 0 \\ H_1 &: \beta_1 \neq 2; \beta_2 - 2\beta_3 \neq 0 \end{aligned}$$

Then

$$\begin{aligned} H &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \end{pmatrix} \\ b &= (b_1, b_2, b_3) \\ \theta^0 &= \begin{pmatrix} 2 \\ 0 \end{pmatrix} \end{aligned}$$

3 CNR Framework (σ^2 Unknown)

The set-up is as before except now σ^2 is not assumed known. It therefore must be estimated and the usual estimator is

$$\hat{\sigma}^2 = \frac{e'e}{T - k}$$

1. Claim:

   $$\hat{\sigma}^2 = \chi^2_{T-k}$$

   Proof: See Goldberger pp. 223-224

2. Claim:

   $$b \text{ is independent of } e$$

   Proof: See Goldberger p. 224

   Therefore, any function of b is independent of any function of e (This is a basic fact of math-stat you should be familiar with).

3. The test statistic

$$v = (t-\theta)' \left[\hat{V}(t)\right]^{-1} (t-\theta)/p$$

is distributed $F(p, T-k)$ where

$$\begin{aligned} t &= Hb \\ \hat{V}(t) &= \hat{\sigma}^2 H \left(X'X\right)^{-1} H' \end{aligned}$$

If we recall from Section 3 Handout, an $F(p, T-k)$ random variable takes the form

$$f = \frac{x/n}{y/d}$$

where $x \tilde{} \chi^2_n$ independently of $y \tilde{} \chi^2_d$. Rewriting v, this distributional result becomes immediately clear.

$$v = \frac{(t-\theta)' \left[H(X'X)^{-1} H'\right]^{-1} (t-\theta) \diagup \sigma^2 p}{\left[e'e/T-k\right]/\sigma^2}$$

The numerator is a χ^2_p random variable divided by its degrees of freedom p. It is also random only through its dependence on b. The denominator is a χ^2_{T-k} random variable and is random only through e. As noted above, e and b are independent as are any functions of these two random variables. The result follows.

4. The test statistic

$$u = \frac{(b_j - \beta_j)}{\hat{\sigma}_{b_j}}$$

is distributed t_{T-k}. Again, from section 3 handout, we know a t random variable is the ratio of a standard normal to a χ^2 divided by its degrees of freedom where the random variables are independent of one another. Rewriting u below, we see this is clearly the case.

$$u = \frac{(b_j - \beta_j)/\sigma_{b_j}}{\sqrt{\left[e'e/T-k\right]/\sigma^2_{b_j}}}$$

3.1 Confidence Intervals and Regions

To find confidence intervals, the methodology is exactly the same except now we use the t_{T-k} distribution to find the critical values.

$$t \pm c\sigma_t$$

For $(T-k) > 50$ the difference between the t and normal distribution is negligible. It's even pretty close for $(T-k) > 25$.

Confidence regions are found similarly using the $F_{p,T-k}$ distribution for the critical values.

$$(t-\theta)' \left[\hat{\sigma}^2 H \left(X'X\right)^{-1} H'\right]^{-1} (t-\theta) \leq c_p$$

3.2 Hypothesis Testing

3.2.1 Univariate

This is the standard $t-$test situation. Consider testing one parameter,

$$\begin{aligned} H_0 &: \beta_j = \beta_j^0 \\ H_1 &: \beta_j \neq \beta_j^0 \end{aligned}$$

Our test statistic is as before except σ_{b_j} is replaced by its estimate $\hat{\sigma}_{b_j}$.

$$t = \frac{b_j - \beta_j^0}{\hat{\sigma}_{b_j}}$$

which now has the t_{T-k} distribution.

3.2.2 Multivariate

As with confidence intervals, the procedure and test statistic are the same except we use our estimator for σ^2 and the $F_{p,T-k}$ distribution for defining the rejection region.

3.2.3 Zero Null Subvector Hypothesis

This subsection discusses the situation where we want to test wether a subvector of the $\beta' s$ are equal to 0. The idea is to relate this testing situation to the short regressions discussed earlier. For illustrative purposes, assume it is the last k_2 elements of the following regression

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where $X_1 \in R^{T\times k_1}, X_2 \in R^{T\times k_2}, \beta_1 \in R^{k_1}$ and $\beta_2 \in R^{k_2}$. The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_1 &: \beta_2 \neq 0 \end{aligned}$$

Using our standard hypothesis testing framework from above, we can write

$$\begin{aligned} t &= Hb = b_2 \\ \theta &= H\beta = \beta_2 \end{aligned}$$

where $H = [0_{k_2\times k_1}; I_{k_2\times k_2}]$. The estimated variance of t is simply, $\hat{V}(t) = \hat{\sigma}^2 H (X'X)^{-1} H'$. If we partition the $(X'X)^{-1}$ matrix according to the subvectors we see

$$H\left(X'X\right)^{-1}H' = (0, I)\begin{pmatrix} Q^{11} & Q^{12} \\ Q^{21} & Q^{22} \end{pmatrix}\begin{pmatrix} 0 \\ I \end{pmatrix} = Q^{22}$$

Recall our test statistic,

$$w = (t-\theta)' \left[\hat{V}(t)\right]^{-1} (t-\theta)/p$$

which can now be written

$$v = b_2' \left[\hat{\sigma}^2 Q^{22}\right]^{-1} b_2 \diagup k_2$$

Using the results from the FWL theorem (or simply the inverse of a partitioned matrix), we can write

$$\left[Q^{22}\right]^{-1} = X_2' M_1 X_2$$

so our statistic becomes

$$v = b_2' X_2' M_1 X_2 b_2 \diagup \hat{\sigma}^2 k_2$$

Residual Sum of Squares: An alternative way of writing this test statistic is to recognize that

$$e^{*\prime} e^* = e'e + b_2' X_2' M_1 X_2 b_2$$

(see Section 3 handout). Therefore

$$\begin{aligned} v &= \left(e^{*\prime}e^* - e'e\right)/\hat{\sigma}^2 k_2 \\ &= \frac{(T-k)}{k_2}\frac{\left(e^{*\prime}e^* - e'e\right)}{e'e} \end{aligned}$$

Result 1 *To calculate the test statistic:*

1. *Run a short (restricted) regression of y on X_1 and compute the sum of square residuals, $e^{*\prime}e^*$.*
2. *Run the long (unrestricted) regression of y on X_1 and X_2 and compute the sum of square residuals, $e'e$.*
3. *Using 1) and 2) compute v.*

The intuition is as follows. A large value of v leads to a rejection of the null (i.e. $\beta_2 \neq 0$) which occurs when the relative difference between the restricted and unrestricted sum of squares is large. This is saying the fit is significantly better when the X_2 matrix is included in the regression.

Coefficient of Determination: When an intercept is included in both the restricted and unrestricted regressions, the R^2 is well-defined. Recall

$$R^2 = 1 - \frac{e'e}{y'M_i y}$$

where M_i projects into the orthocomplement of the summer vector space (it de-means things). This suggests another way of writing our test statistic,

$$v = \frac{(T-k)}{k_2}\frac{\left(R^2 - R^{2*}\right)}{\left(1-R^2\right)}$$

where R^{2*} is the R^2 from the restricted regression.

Result 2 *To calculate this test statistic:*

1. *Run a short (restricted) regression of* y *on* X_1 *and compute the* $R^2\left(\equiv R^{2*}\right)$
2. *Run the long (unrestricted) regression of* y *on* X_1 *and* X_2 *and compute the* R^2.
3. *Using 1) and 2) compute* v.

As a special case, consider testing whether all the slope coefficients were 0. That is, all coefficients except for the intercept. Our test statistic can be written as

$$\frac{(T-k)}{k-1}\frac{R^2}{1-R^2}$$

since the restricted regression sum of square residuals is $e^{*\prime}e^{*}=\sum\left(y_t-\overline{y}\right)^2=y'M_iy$ implying R^{2*} is in effect 0 since

$$R^{2*}=1-\frac{e^{*\prime}e^{*}}{y'M_iy}=1-\frac{y'M_iy}{y'M_iy}=0$$

3.3 General Linear Hypotheses

Consider the following problem

$$y=\beta_0+x_1\beta_1+x_2\beta_2+x_3\beta_3+\varepsilon$$

where $x_i, i=1,2,3$ are $T\times 1$ column vectors.. Now consider testing the following hypotheses

$$\begin{aligned} H_0 &: \beta_3=-\beta_1;\beta_1=\beta_2 \\ H_1 &: \beta_3\neq-\beta_1;\beta_1\neq\beta_2 \end{aligned}$$

We can run this test in the usual manner by constructing the test statistic

$$(\theta-t)'\left[\hat{V}\left(t\right)\right]^{-1}(\theta-t)\,\tilde{}\,F_{p,T-k}$$

where

$$\begin{aligned} t &= Hb=\begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \\ \theta &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \hat{V}\left(t\right) &= \hat{\sigma}^2H\left(X'X\right)^{-1}H' \\ p &= 2 \\ k &= 3 \end{aligned}$$

The idea this section attempts to illustrate is that any general linear hypothesis can be converted into a zero-null subvector hypothesis. That is, we can solve out the restrictions, run a short regression and use methods zero subvector null hypotheses. For the above example we see the first restriction $\beta_3 = -\beta_1$ implies

$$y = \beta_0 + \beta_1 (x_1 - x_3) + x_2\beta_2 + \varepsilon$$

The second restriction, $\beta_1 = \beta_2$, implies

$$y = \beta_1 (x_1 - x_3 + x_2) + \varepsilon$$

So our short regression is simply

$$y = \gamma_1 z + \varepsilon$$

where $z = x_1 - x_3 + x_2$.

Another example is to consider

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \varepsilon$$

and the hypothesis $\beta_1 + \beta_2 + \beta_3 = 1$. But his implies $\beta_1 = 1 - \beta_2 - \beta_3$ so

$$\begin{aligned} y &= \beta_0 + x_1 (1 - \beta_2 - \beta_3) + x_2\beta_2 + x_3\beta_3 + \varepsilon \\ y &= \beta_0 + x_1 + \beta_2 (x_2 - x_1) + \beta_3 (x_3 - x_1) + \varepsilon \\ y - x_1 &= \beta_0 + \beta_2 (x_2 - x_1) + \beta_3 (x_3 - x_1) + \varepsilon \end{aligned}$$

Our short regression is thus

$$y^* = \gamma_0 + \gamma_2 z_1 + \gamma_3 z_2 + \varepsilon$$

where $y^* = y - x_1$, $z_1 = x_2 - x_1$ and $z_2 = x_3 - x_1$.