Economics 421

Winter 2008

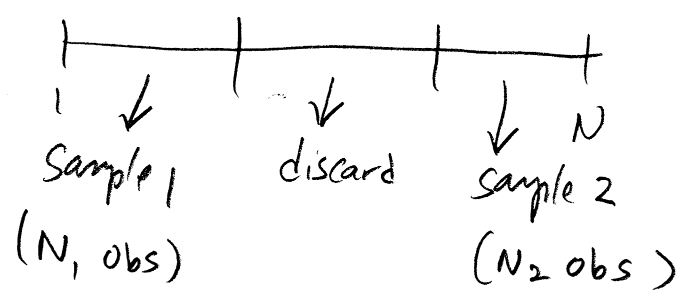
Final Exam

1

(1) (a) The steps for the Goldfeld-Quandt test are

1. Identify a variable, say X_3, thought to be related to the variance, σ_i^2, and arrange the data according to the magnitude of X_3 (e.g. smallest to largest).

2. Divide the sample into three parts



N = number of obs

3. Estimate two regressions, one for the first N_1 obs, and one for the last N_2 obs (often, $N_1 = N_2$, but not required).

4. From the estimates, construct

$$RSS_1 = \sum_{N_1} \hat{u}_i^2 \quad \text{(error sum of squares for model 1)}$$

$$RSS_2 = \sum_{N_2} \hat{u}_i^2 \quad \text{(error sum of squares for model 2)}.$$

[Cont.]

5. Compute the F-Statistic

$$F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}, \quad \text{where } \hat{\sigma}_2^2 = \frac{RSS_2}{N_2 - K}$$

$$\hat{\sigma}_1^2 = \frac{RSS_1}{N_1 - K}$$

$$K = 3$$

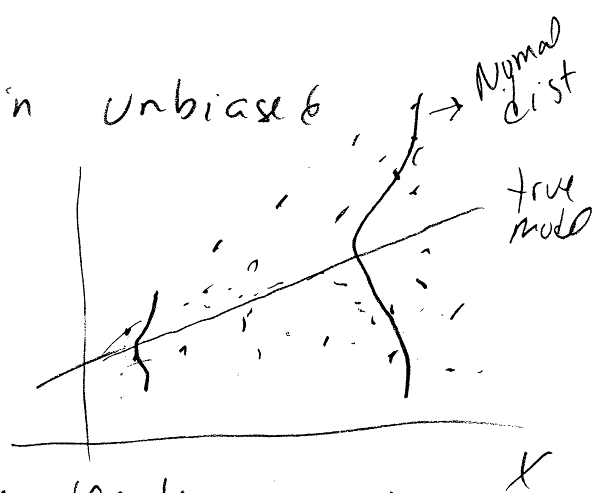
which is distributed as $F(N_2 - K, N_1 - K)$.

6. Compare the Calculated F statistic to an F-statistic from a table, call it F_c

If $F > F_c \rightarrow$ reject null that variances of two samples are equal ($H_0: \sigma_2^2 = \sigma_1^2$)

If $F < F_c \rightarrow$ fail to reject.

(b) The Coefficients remain unbiased. Heteroskedasticity changes the spread as $X \uparrow$, but not the central tendency (ie. Normal dist for errors still centered on zero). Since



Coefficients measure the central tendency, not the spread, and central points unchanged, $\hat{\beta}$'s are unbiased.

(2) (a) First, form the t-statistic $t = \frac{\hat{\beta}_4 - 1}{\hat{\sigma}_{\hat{\beta}_4}}$,

then look up the critical value of the t-statistic in a table (degrees of freedom are N−K). if $t > t_{critical}$, reject that $\beta_4 = 1$. If $t < t_{critical}$, fail to reject

(b) First, estimate the unrestricted model by OLS

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$, save RSS_{UR}

Then, estimate the restricted model,

$$y_i = \beta_1 + \beta_4 X_{4i} + u_i$$, again, save RSS_R

Form the F-statistic:

$$F = \frac{\frac{RSS_R - RSS_{UR}}{2}}{\frac{RSS_{UR}}{N-4}}$$. Compare to

critical $F(N-2, N-4)$, if $F > F_{critical}$, reject that both are zero, if $F < F_{critical}$, fail to reject.

[cont.]

2 (c) Again, first estimate the UR model, save RSS_{UR}. Next, impose the restriction that $\beta_2 = 2 + 4\beta_4$

$$y_i = \beta_1 + (2 + 4\beta_4)X_2 + \beta_3 X_3 + \beta_4 X_4$$

$$(Y_i - 2X_2) = \beta_1 + \beta_3 X_3 + \beta_4 (4X_2 + X_4)$$

→ form $y^* = Y_i - 2X_2$

$X^* = 4X_2 + X_4$

Run the restricted model:

$$Y^* = \beta_1 + \beta_3 X_3 + \beta_4 X^* + u_i$$

save RSS_R.

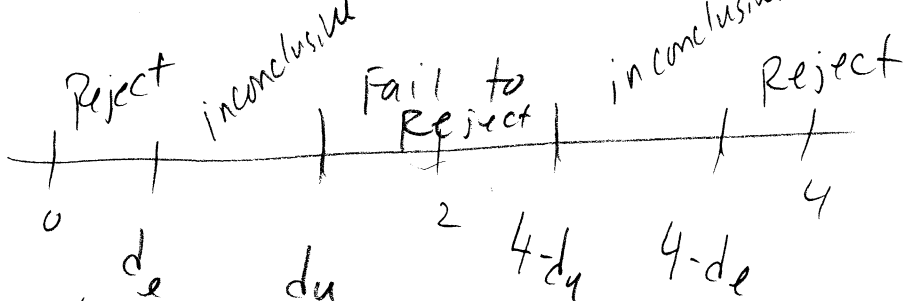
Next, form $F = \dfrac{\dfrac{RSS_R - RSS_{UR}}{N-1}}{\dfrac{RSS_{UR}}{N-4}}$

and compare this to $F(N-1, N-4)$ as explained in part (b).

3 (a) For illustration, suppose the model is

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

1. Estimate by OLS, save $\hat{u}_t$.
2. Form $$d = \frac{\sum_{2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{1}^{T} \hat{u}_t^2}$$
3. Look up critical values for test of $\rho = 0$ in $u_t = \rho u_{t-1} + \varepsilon_t$, will get d_L and d_U. (df = T and K)



pos correlation $0 < d < 2$

neg corr $2 < d < 4$

4. Find where the calculated d stat appears, and reject, fail to reject, or find it's inconclusive as appropriate.

3 (b) Suppose that

$$y_t = \alpha Y_{t-1} + \beta X_t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t.$$

Then the RHS variable Y_{t-1} is correlated with the error (through u_{t-1}) $\rightarrow$ biased coefficients $\rightarrow$ biased $\hat{u}_t$ $\rightarrow$ biased test. Instead, use Durbin's h-test.

4 (a) Suppose the model is $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ and the test is for whether X_3 should be in the model.

1. Estimate the restricted model
$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i, \text{ save } \hat{u}_i.$$
2. Regress $\hat{u}_i$ on constant, X_2 and X_3
3. Test stat is $NR^2 \sim \chi^2$(# of restrictions) ↓ one here.
4. So, calculate NR^2, compare to critical value. Reject if $NR^2 >$ critical, fail to reject if $NR^2 <$ critical.

[cont.]

4 (b) Omitting a variable that should be included:

If uncorrelated with X's in the model, constant biased, but slope coefficients are unbiased/consistent

If correlated with X's, then coefficient estimates are biased and inconsistent.

5 $Y_t = \beta_0 + \beta_1 X^* + u$, sub in for X^* using $X^* = X - w$

$$Y_t = \beta_0 + \beta_1 (X - w) + u$$

$$Y_t = \beta_0 + \beta_1 X + \underbrace{(u - \beta_1 w)}_{\text{call this } v.}$$

Then $$\hat{\beta}_1 = \beta_1 + \frac{\sum (X - \bar{X})(V - \bar{V})}{\sum (X - \bar{X})^2} = \beta_1 + \frac{\frac{1}{N} \sum (X - \bar{X})(V - \bar{V})}{\frac{1}{N} \sum (X - \bar{X})^2}$$

Take plim:
$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(X, V)}{\text{Var}(X)}$$

[cont.]

$Cov(X, v) = Cov(X^* + w, u - \beta_1 w) = -\beta_1 \sigma_w^2$

$Var(X) = \sigma_{X^*}^2 + \sigma_w^2$

$$\Rightarrow \text{plim } \hat{\beta}_1 = \beta_1 + \frac{-\beta_1 \sigma_w^2}{\sigma_{X^*}^2 + \sigma_w^2} \Rightarrow \text{inconsistent.}$$

To overcome, estimate with IV.

6 (a) Equation 1 : X_2, X_4 excluded.

$G-1 = 1 \Rightarrow$ over ID

Equation 2 : X_2 excluded, $\# = G-1$

$\Rightarrow$ exactly ID

(b) Stage 1

Regress Y_2 on a constant, X1, X2, X3, and X4 (i.e. all the X^s), save $\hat{Y}_2$, the predicted value.

Stage 2

use OLS to estimate

$$Y_1 = \alpha_0 + \alpha_1 \hat{Y}_2 + \alpha_2 X_1 + \alpha_3 X_3 + u$$

note the "hat"

7 The main problem is that the coefficients are estimated imprecisely. OLS is still BLUE, but because it cannot distinguish the x's very easily (due to their correlation), the std errors are generally large

To detect, look for a high degree of correlation among x's (e.g. $|\rho| > .8$ is one rule of thumb).

Two points of view on what to do about it. One point of view says that the model is correctly specified, so only solution is to get more data. As $N \uparrow$, coefficients are estimated more precisely, so with enough data we can overcome the lack of precision

(cont.)

Another solution is to drop one of the variables. If they are very highly correlated, they are essentially the same variable anyway, so dropping one of the two won't be very costly. If collecting more data is not possible, this may be the only way to get any precision at all.

(8) (a) This is a model such as

$$Y_i = \alpha + \beta X_i + u_i$$

Probability of an event, e.g. purchase of a car

explanatory variables for probability.

When data are for individuals,

[cont.]

It is usually coded $y=1$ (yes) and $Y=0$ (no). Thus, dependent variable is binary. Since Y_t is binary, (1) it cannot be Normal (same with errors) $\Rightarrow$ non-normal $\rightarrow$ biased tests. Also, (2) can show model is heteroskedastic, and that (3) the predicted probabilities don't necessarily lie in the $[0,1]$ range.

(b) As an example, suppose that whenever $Y_i < 0$, we only observe $Y=0$ (ie. data coded $Y=0$ when collected).

Can see from the diagram that the censoring will bias the coefficients because the data are no longer centered on the true line.