TP4 Correction : Heteroscedasticity

March 20, 2019

Question 1

Consider linear model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

The homoskedasticity assumption for multiple regression states that the variance of the unobservable error u, conditional on the explanatory variables, is constant. Homoskedasticity fails whenever the variance of the unobservables changes across different segments of the population, where the segments are determined by the different values of the explanatory variables.

Homoskedasticity is needed to justify the usual $t-$ tests, $F-$ tests, and confidence intervals for OLS estimation of the linear regression model, even with large sample sizes.
Testing the homoskedascticity. We start by performing regression

$$price = \beta_0 + \beta_1 bdrms + \beta_2 lotsize + \beta_3 sqrft + u.$$

We test

$$H_0 : Var(u|x_1, x_2, \cdots, x_k) = \sigma^2$$

Since we assume that u are zero mean, thus $Var(u|x) = \mathbb{E}(u^2|x)$ and

$$H_0 : \mathbb{E}(u^2|x_1, x_2, \cdots, x_k) = \mathbb{E}(u^2) = \sigma^2.$$

$$load\ \ hprice1.raw$$

$$y = hprice1(:, 1);$$

$$[n, k] = size(hprice1);$$

$$X = [ones(n, 1), hprice1(:, [3, 4, 5])];$$

$$[n, k] = size(X)$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

Thus, if H_0 is false, the expectation of u^2 may possibly be any function of x_j. We will test the model $\hat{u}^2 = \delta_0 + \delta_1 bdrms + \delta_2 lotsize + \delta_3 sqrft + \nu$, with the null hypothesis of homoskedasticity :

$$H0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0.$$

Test the homoskedasticity:

$$u2 = u.^2;$$

$$y = u2;$$

Unrestricted model:

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$SSR0 = u' * u$$

Restricted model (all the coefficients except the intercept are zeros)

$$X = [ones(n, 1)];$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$SSR1 = u' * u$$

$$F = ((SSR1 - SSR0)/SSR0) * (n - k)/3$$

$$p = fdis\_prb(F, 3, n - k)$$

We obtain $F = 5.3389$ and $p = 0.0020$, so we reject H_0 (with 2%).

Question 2

We will test the homoskedasticity for the logarithmic form:

$$log(price) = \beta_0 + \beta_1 bdrms + \beta_2 lotsize + \beta_3 sqrft + u.$$

.

$$y = hprice1(:, 1);$$

$$y = log(y)$$

$$[n, k] = size(hprice1);$$

$$X = [ones(n, 1), hprice1(:, [3, 4, 5])];$$

$$[n, k] = size(X)$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$u2 = u.^2;$$

$$y = u2;$$

Unrestricted model:

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$SSR0 = u' * u$$

Restricted model (all the coefficients except intercept are zeros)

$$X = [ones(n, 1)];$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$SSR1 = u' * u$$

$$F = ((SSR1 - SSR0)/SSR0) * (n - k)/3$$

$$p = fdis\_prb(F, 3, n - k)$$

We obtain $F = 1.17$ and $p = 0.3244$, so we can not reject H_0.

Correction by the method WLS

Question 3

We assume that we know the form of the heteroskedasticity $Var(u|x) = \sigma^2 h(x)$ with $h(x) = lotsize$ that is, the variance of the error is proportional to the size of the terrain. Thus

$$\frac{price}{\sqrt{lotsize}} = \frac{\beta_0}{\sqrt{lotsize}} + \frac{\beta_1 bdrms}{\sqrt{lotsize}} + \frac{\beta_2 lotsize}{\sqrt{lotsize}} + \frac{\beta_3 sqrft}{\sqrt{lotsize}} + \frac{u}{\sqrt{lotsize}}$$

satisfies the hypothesis the homoskedasticity.

$$y = hprice1(:,1);$$

$$X = [ones(n,1), hprice1(:,[3,4,5])];$$
$$lotsize = hprice1(:,4)$$
$$y = y./sqrt(lotsize)$$
$$for\ i = 1:k$$
$$X(:,i) = X(:,i)./sqrt(lotsize)$$
$$end$$
$$[n,k] = size(X)$$
$$beta = inv(X' * X) * X' * y$$
$$u = y - X * beta;$$

Test the homoskedasticity of the linear form:

$$u2 = u.^2;$$

$$y = u2;$$

Unrestricted model:

$$beta = inv(X' * X) * X' * y$$
$$u = y - X * beta;$$
$$SSR0 = u' * u$$

Restricted model (all the coefficients except intercept are zeros)

$$X = [ones(n,1)]./sqrt(lotsize);$$

$$beta = inv(X' * X) * X' * y$$
$$u = y - X * beta;$$
$$SSR1 = u' * u$$
$$F = ((SSR1 - SSR0)/SSR0) * (n-k)/3$$
$$p = fdis\_prb(F, 3, n-k)$$

We obtain $F = 1.7705$ and $p = 0.1590$, so we do not reject the H_0.

Question 4

Consider a linear model $sav = \beta_0 + \beta_1 inc + u$. We know *a priori*

$$\sigma_i^2 = Var(u_i|inc_i) = \sigma^2 inc_i$$

Conduct a correction by the method WLS. Test the homoskedasticity for modified model.

$$load\ \ saving.raw$$

$$y = saving(:, 1);$$

$$inc = saving(:, 2);$$

$$[n, k] = size(saving);$$

$$X = [ones(n, 1), inc];$$

$$[n, k] = size(X)$$

$$beta = inv(X' * X) * X' * y$$

$$u = y - X * beta;$$

$$sig2 = u' * u/(n - k)$$

$$std = sqrt(diag(sig2 * inv(X' * X)))$$

$$t = beta./std$$

Correction by the method WLS :

$$ys = y./sqrt(inc);$$

$$Xs = [ones(n, 1)./sqrt(inc)inc./sqrt(inc)];$$

$$beta = inv(Xs' * Xs) * Xs' * ys$$

$$u = ys - Xs * beta;$$

$$sig2 = u' * u/(n - k)$$

$$std = sqrt(diag(sig2 * inv(Xs' * Xs)))$$

$$t = beta./std$$

$$t = t(2)$$

$$tdis\_prb(t, n - 2)$$

$t = 3.0232$ and $p = 0.0032$. Thus, we do not reject (however weakly) H_0.