Problem Set 3

Due: 3/23

Part One: Hand-Written Exercise

1. We mentioned that the F statistic is given by:

$$F = \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)},$$

where SSR_r and SSR_{ur} are the residual sums of squares of restricted and unrestricted regressions respectively. Moreover, $(\mathrm{SSR}_r - \mathrm{SSR}_{ur})$ and SSR_{ur} are independent of each other.

(a) Given the fact that:

$$\frac{(n-k-1+q)\hat{\sigma}_r^2}{\sigma^2} - \frac{(n-k-1)\hat{\sigma}_{ur}^2}{\sigma^2} \sim \chi^2(q),$$

where $\hat{\sigma}_r^2$ and $\hat{\sigma}_{ur}^2$ are the OLS estimators of σ^2 of the restricted and unrestricted regressions respectively. Please show that

$$\frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)} \sim F(q, n-k-1).$$

(b) Show that the F statistic can also be written as the R-squared form

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)},$$

where R_r^2 and R_{ur}^2 are the R^2s of the restricted and unrestricted regressions, respectively.

2. Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\beta}_3 x_{3,i} + \hat{\beta}_4 x_{4,i} + \hat{\beta}_5 x_{5,i}$, with $i = 1, ..., 150$. Suppose we construct $\mathbf{R}\boldsymbol{\beta} = \mathbf{c}$ in order to test the following hypothesis jointly:

$$\begin{aligned}
&\text{hypothesis 1}: \ \beta_1 = \beta_2 = \beta_5 = 0 \\
&\text{hypothesis 2}: \ \beta_0 + 3\beta_3 - 6\beta_4 = 4 \\
&\text{hypothesis 3}: \ \beta_2 + 4\beta_4 + 1 = 2\beta_1 \\
&\text{hypothesis 4}: \ \beta_0 + 3\beta_1 + \beta_3 + 5\beta_5 = 7
\end{aligned}$$

(a) Please write out matrix $\mathbf{R}$ and $\mathbf{c}$.

(b) Construct the test statistics, and indicate which distribution does it follows (make sure the degree of freedom is clearly specified).

3. Consider two dummy variables:
$D_{i1} = 1$ if i is the top $\frac{1}{3}$ trading volume stock and $D_{i1} = 0$ otherwise;
$D_{i2} = 1$ if i is the bottom $\frac{1}{3}$ trading volume stock and $D_{i2} = 0$ otherwise.
If we assume that there are differences in intercept and slope between each group and reference group(middle size of trading volume stock).

   (a) Please write down the regression model with the setting above.

   (b) Please write down the regression models when $D_{i1} = 1, D_{i2} = 0$; $D_{i1} = 0, D_{i2} = 1$ and $D_{i1} = D_{i2} = 0$ respectively.

   (Note that there is another independent variable x_i)

4. The following model can be used to study whether campaign expenditures affect election outcomes:

$$\text{voteA} = \beta_0 + \beta_1 ln(\text{expendA}) + \beta_2 ln(\text{expendB}) + \beta_3 \text{prtystrA} + u,$$

   where "voteA" is the percentage of the vote received by candidate A, "expendA" and "expendB" are campaign expenditures by candidates A and B, and "prtystrA" is a measure of party strength for candidate A (the percentage of the most recent presidential vote that went to A's party). [modified from Wooldridge. Problem 1CE from chapter 4.]

   (a) What is the interpretation of β_1?

   (b) In terms of the parameters, state the null hypothesis that the effect of the increase in A's expenditure will be offset by the increase in B's expenditure.

   (c) Write the detailed procedure to do the hypothesis testing in (b).

   (d) If someone claims that both candidates' expenditures do not have any effect on the outcome, how can you specify a testing null hypothesis?

   (e) Write the detailed procedure to do the hypothesis testing in (d).

Part Two: Monte Carlo Simulation (Due: 3/30)

- Simulation design:
  - Sample sizes N:

(i) 10

(ii) 500

- Number of replications: 1000
- Data generating process (DGP):

(i) $y_i \sim N(0,1)$

(ii) $y_i \sim t(4)$

(iii) $y_i \sim t(1)$

- The statistics:

$$M_N = \frac{1}{\hat{\sigma}_N \sqrt{N}} \sum_{i=1}^{N} \phi(y_i), \quad \text{where} \quad \hat{\sigma}_N^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\phi(y_i) - \frac{1}{N} \sum_{i=1}^{N} \phi(y_i) \right)^2,$$

with the moment functions:

(i) $\phi(y_i) = y_i$

(ii) $\phi(y_i) = y_i^3$

(iii) $\phi(y_i) = sin(y_i)$

(iv) $\phi(y_i) = cos(y_i)$

- For the total 24 different ways of constructing M_N, please plot their corresponding histogram using 1000 replications. Open a new window and combine the 24 graphs on a single plot and place them as 6×4.
- Please compute the empirical frequencies of the events: $M_N > Z_{.99} = 2.3$ and $M_N > Z_{.95} = 1.6$ for each simulations(Z_q is the q^{th} quantile of standard normal distribution). Record them under their corresponding graphs, see if the frequencies are, respectively, sufficiently close to the 5% and 1% nominal levels.
- Please add the Gaussian kernel density estimate (KDE) of M_N as well as the probability density function (PDF) of $N(0,1)$ for each simulation graphs.