STOR 664: FALL 2021
Midterm Exam, October 7, 2021

Open book in-class exam: time limit 75 minutes.

This is a single multi-part question but each part will be graded independently of the other parts. You are allowed to consult course notes (printed or e-read), homework assignments and any personal notes you have made during the course. Other outside materials are not permitted. You may consult the instructor if the wording is unclear or if you think there might be an error, but the instructor will not give hints how to solve the exam. The university Honor Code is in effect at all times.

Consider the linear regression model

$$y_i \;=\; \beta_1 + x_i\beta_2 + x_i^3\beta_3 + \epsilon_i, i = 1, \ldots, n$$

where the ϵ_i are independent normally distributed random variables with mean 0 and a common unknown variance σ^2. The x_i are assumed to be symmetric about 0 in the following sense: $x_i = -x_{n+1-i}$ for $i = 1, \ldots, n$.

For notational convenience, define $S_k = \sum_{i=1}^n x_i^k,\ T_k = \sum_{i=1}^n y_i x_i^k$ for any $k \geq 0$.

(a) Find explicit expressions for the least squares estimators $\hat\beta_1$, $\hat\beta_2$, $\hat\beta_3$, and their variances. You should express the answer in terms of the values of S_k and T_k for $k \geq 0$, or any expressions derived from them. [**20 points**]

(b) It is suggested that $\beta_2 = 2\beta_3$. By rewriting the model in terms of β_1 and β_3 alone, find least squares estimators under this assumption, say $\tilde\beta_1$ and $\tilde\beta_3$, and expressions for their variances. [**15 points**]

(c) Now suppose we want to test the hypothesis $H_0: \ \beta_2 = 2\beta_3$ against the alternative $H_1: \ \beta_2 \neq 2\beta_3$. Find expressions for the regression sums of squares, RSS_0 and RSS_1, under H_0 and H_1 respectively. Hence derive an expression for the F statistic in this instance, and state the rejection region in which you would reject H_0 at significance level α. [**20 points**]

(d) Suppose that, in fact, $2\beta_3 - \beta_2 = M$ for some $M \neq 0$. Show that the noncentrality parameter in the distribution of the F statistic is

$$\lambda \;=\; \frac{S_2S_6 - S_4^2}{S_6 + 4S_4 + 4S_2} \cdot \frac{M^2}{\sigma^2}.$$

[**30 points**]

(e) Now consider the case $n = 11,\ x_i = i - 6$. In this case (assume without proof) $\frac{S_2S_6 - S_4^2}{S_6+4S_4+4S_2} = 21.76$ to two decimal places. Suppose $\sigma = 1,\ M = 1$ and we are considering a test of significance level 0.01 for the null hypothesis $\beta_2 = 2\beta_3$ against the alternative $\beta_2 \neq 2\beta_3$. Describe how you would calculate the power in this case. (A numerical answer is not required but you should describe explicitly how to obtain it, in terms of any relevant R functions.) [**15 points**]

Solutions

(a) Note that by the condition on the x_i, $S_k = 0$ for all odd k. We have: $X^TX = \begin{pmatrix} n & 0 & 0 \\ 0 & S_2 & S_4 \\ 0 & S_4 & S_6 \end{pmatrix}$ and $X^TY = \begin{pmatrix} T_0 \\ T_1 \\ T_3 \end{pmatrix}$. Defining $\Delta = S_2S_6 - S_4^2$, we have $(X^TX)^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{S_6}{\Delta} & -\frac{S_4}{\Delta} \\ 0 & -\frac{S_4}{\Delta} & \frac{S_2}{\Delta} \end{pmatrix}$. Hence $\hat{\beta}_1 = \frac{T_0}{n}$ ($\bar{y}$ in the conventional notation), $\hat{\beta}_2 = \frac{T_1S_6 - T_3S_4}{\Delta}$, $\hat{\beta}_3 = \frac{T_3S_2 - T_1S_4}{\Delta}$, and the variances are $\frac{\sigma^2}{n}$, $\frac{\sigma^2 S_6}{\Delta}$, $\frac{\sigma^2 S_2}{\Delta}$.

(b) Rewrite the model as $y_i = \beta_1 + \beta_3(2x_i + x_i^3)$ where $\sum(2x_i + x_i^3)^2 = 4S_2 + 4S_4 + S_6$, so $X^TX = \begin{pmatrix} n & 0 \\ 0 & 4S_2 + 4S_4 + S_6 \end{pmatrix}$, $X^TY = \begin{pmatrix} T_0 \\ 2T_1 + T_3 \end{pmatrix}$, so $\tilde{\beta}_1 = \frac{T_0}{n} = \bar{y}$, $\tilde{\beta}_3 = \frac{2T_1 + T_3}{4S_2 + 4S_4 + S_6}$ with variances $\frac{\sigma^2}{n}$, $\frac{\sigma^2}{4S_2 + 4S_4 + S_6}$.

(c) Applying the formula $SSR = \sum(y_i - \bar{y})^2$ and noting that $\hat{\beta}_1$, we get

$$SSR_1 \quad = \quad \sum(\hat{\beta}_2 x_i + \hat{\beta}_3 x_i^3)^2 \quad = \quad \hat{\beta}_2^2 S_2 + 2\hat{\beta}_2\hat{\beta}_3 S_4 + \hat{\beta}_3^2 S_6,$$

where the subscript 1 denotes that this is for the original model which is now the alternative hypothesis H_1.

Similarly, under H_0 we have

$$SSR_0 \quad = \quad \sum\{\tilde{\beta}_3(2x_i + x_i^3)\}^2 \quad = \quad \tilde{\beta}_3^2(4S_2 + 4S_4 + S_6).$$

and therefore

$$SSR_1 - SSR_0 \quad = \quad \hat{\beta}_2^2 S_2 + 2\hat{\beta}_2\hat{\beta}_3 S_4 + \hat{\beta}_3^2 S_6 - \tilde{\beta}_3^2(4S_2 + 4S_4 + S_6). \tag{1}$$

Now, it turns out that with considerable extra work, we can show

$$SSR_1 - SSR_0 \quad = \quad \frac{S_6S_2 - S_4^2}{(4S_2 + 4S_4 + S_6)}\left(\hat{\beta}_2 - 2\hat{\beta}_3\right)^2. \tag{2}$$

When I originally set this question, I underestimated the difficulty of going from (1) to (2). Therefore, I'm prepared to accept (1), or something equivalent, as a complete solution to the question.

The rest of the construction of the F test follows the usual pattern. We also need to compute $SSE_1 = \sum(y_i - \hat{y}_i)^2 = \sum(y_i - \bar{y})^2 - SSR_1$ (either of those formulas is acceptable), then we have the null distribution of the F statistic:

$$\frac{SSE_0 - SSE_1}{SSE_1/(n-3)} \quad = \quad \frac{SSR_1 - SSR_0}{SSE_1/(n-3)} \quad \sim \quad F_{1,n-3} \text{ under } H_0$$

The test therefore rejects H_0 at significance level α if $(n-3)(SSR_1 - SSR_0)/SSE_1/(n-3) >$ `qf(n-3,1,1-alpha)` or some equivalent expression.

(d) If you succeeded in deriving formula (2) then the substitution rule says that $\lambda\sigma^2$ is the right hand side of (2) with $\left(\hat{\beta}_2 - 2\hat{\beta}_3\right)^2$ replaced by $(\beta_2 - 2\beta_3)^2 = M^2$, leading at once to the stated formula. However, we can also proceed directly using equation (3.42) of the course text (with δ^2 replaced by λ). The null hypothesis is of the form $C\boldsymbol{\beta} = 0$ where $C = \begin{pmatrix} 0 & 1 & -2 \end{pmatrix}$, so $C(X^TX)^{-1}C^T = \begin{pmatrix} 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{S_6}{\Delta} & -\frac{S_4}{\Delta} \\ 0 & -\frac{S_4}{\Delta} & \frac{S_2}{\Delta} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} = \frac{S_6+4S_4+4S_2}{\Delta}$. Recalling the formula for Δ and writing $h = 0,\ h' = M$, equation (3.42) becomes $\sigma^2\lambda = M^2/C(X^TX)^{-1}C^T = M^2(S_2S_6 - S_4^2)/(S_6 + 4S_4 + 4S_2)$, as required.

(e) There was an error here because the correct value for $(S_2S_6 - S_4^2)/(S_6 + 4S_4 + 4S_2)$ is 13.78 not 21.56! However, based on the stated value, $\lambda = \frac{21.76M^2}{\sigma^2} = 21.76$. In R, we have

```
c=qf(0.99,1,8)
pf(c,1,8,ncp=21.76,lower.tail=F)
```

which, if evaluated numerically, has $c = 11.25862$ and power 0.86 to two decimal places. (With the correct value $\lambda = 13.78$, the power is 0.64.)

Footnote: proof that (1) implies (2).

Using the expressions for $\hat{\beta}_2,\ \hat{\beta}_3,\ \tilde{\beta}_2$ in parts (a) and (b), let's try to write

$$\tilde{\beta}_3 = A\hat{\beta}_2 + B\hat{\beta}_3 \tag{3}$$

where A and B are constants (independent of $y_1, \ldots, y_n$, and hence of T_1 and T_3). Equating the coefficients of $T_1,\ T_3$ we deduce

$$\begin{aligned} \frac{2}{4S_2 + 4S_4 + S_6} &= \frac{AS_6}{\Delta} - \frac{BS_4}{\Delta} \\ \frac{1}{4S_2 + 4S_4 + S_6} &= -\frac{AS_4}{\Delta} + \frac{BS_2}{\Delta} \end{aligned}$$

The solution of this pair of simultaneous equations is given by

$$\begin{aligned} \frac{2S_4 + S_6}{4S_2 + 4S_4 + S_6} &= B \cdot \frac{S_2S_6 - S_4^2}{\Delta} = B, \\ \frac{2S_2 + S_4}{4S_2 + 4S_4 + S_6} &= A \cdot \frac{S_2S_6 - S_4^2}{\Delta} = A. \end{aligned}$$

From (1) we have,

$$\begin{aligned} \mathrm{SSR}_1 - \mathrm{SSR}_0 &= \hat{\beta}_2^2S_2 + 2\hat{\beta}_2\hat{\beta}_3S_4 + \hat{\beta}_3^2S_6 - \tilde{\beta}_3^2(4S_2 + 4S_4 + S_6) \\ &= \hat{\beta}_2^2S_2 + 2\hat{\beta}_2\hat{\beta}_3S_4 + \hat{\beta}_3^2S_6 - (A\hat{\beta}_2 + B\hat{\beta}_3)^2(4S_2 + 4S_4 + S_6) \\ &= \hat{\beta}_2^2S_2 + 2\hat{\beta}_2\hat{\beta}_3S_4 + \hat{\beta}_3^2S_6 - \frac{(2S_2 + S_4)^2\hat{\beta}_2^2 + 2(2S_2 + S_4)(2S_4 + S_6)\hat{\beta}_2\hat{\beta}_3 + (2S_4 + S_6)^2\hat{\beta}_3^2}{(4S_2 + 4S_4 + S_6)} \end{aligned}$$

which reduces after some algebraic manipulation to

$$\mathrm{SSR}_1 - \mathrm{SSR}_0 = \frac{S_2S_6 - S_4^2}{(4S_2 + 4S_4 + S_6)}\hat{\beta}_2^2 - 4\frac{S_2S_6 - S_4^2)}{(4S_2 + 4S_4 + S_6)}\hat{\beta}_2\hat{\beta}_3 + 4\frac{S_2S_6 - S_4^2}{(4S_2 + 4S_4 + S_6)}\hat{\beta}_3^2$$

which is the same as (2).