COMPREHENSIVE WRITTEN EXAMINATION, PAPER III
PART 1: FRIDAY AUGUST 18, 2023 9:00 A.M.–11:00 A.M.
STOR 664 Theory Question (50 points)

This is a closed-book exam: no access to course materials or other (e.g. internet) resources is allowed. Answers in a blue book (provided). No communication is allowed with individuals either inside or outside the exam room; however, if you have queries about the exam, you may call or text the instructor at the phone number provided.

Many environmental variables display both seasonal variation and trends. A geophysicist collects monthly data for m years on an environmental variable Y and looks for evidence of both an annual sinusoidal effect and a linear trend. A plausible model for this purpose is

$$Y_t \;=\; \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t,\ 1 \le t \le 12m, \tag{1}$$

where Y_t is the observation in month t and $x_{1t} = t - \frac{12m+1}{2},\ x_{2t} = \cos\left(\frac{2\pi t}{12}\right),\ x_{3t} = \sin\left(\frac{2\pi t}{12}\right)$, and the errors ϵ_t are independent $\mathcal{N}[0, \sigma^2]$ as in the usual linear model assumptions. Note that with these definitions, each of $\sum_{t=1}^{12m} x_{jt} = 0,\ j = 1, 2, 3.$

In the following questions, you are asked to derive algebraic formulas for several estimators or tests from this model. Any algebraically correct answer will earn positive credit, but answers that show how to reduce the expressions to their simplest forms will earn the greatest credit.

(a) Write the above model in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, and show explicitly how $\mathbf{y},\ X,\ \boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ are derived from the quantities in (1). In particular, give expressions for the matrices $X,\ X^TX$ and $(X^TX)^{-1}$ in the simplest form you can derive, using the formula sheet at the end of this exam. (*Note:* For $(X^TX)^{-1}$ and in subsequent calculations that depend on $(X^TX)^{-1}$, it will suffice to give the answer in terms of the quantities a, b, c, d, e from the formula sheet, provided you explain clearly how these numbers are calculated.) **[12 points].**

(b) Using your results from (a), given explicit expressions for the least squares estimators $\hat{\beta}_j,\ j = 0, 1, 2, 3$, and give formulas for their standard error. **[8 points].**

(c) The geophysicist is convinced that there is periodic variation in the time series, but is not sure about the trend. Show how to construct a hypothesis test of $H_0:\ \beta_1 = 0$ against the alternative $H_1:\ \beta_1 \neq 0$, where $\beta_0,\ \beta_2$ and β_3 and the common sample variance σ^2 are unknown. Assume a two-sided test with significance level 0.05. **[6 points].**

(d) Alternatively, suppose the geophysicist accepts the existence of a trend but is unsure about the periodic component. This suggests testing $H_0:\ \beta_2 = \beta_3 = 0$ against the alternative H_1 that at least one of β_2 or β_3 is non-zero, again assuming $\beta_0,\ \beta_1$ and σ^2 are unrestricted. Describe the steps needed to construct a hypothesis test of H_0 against H_1. Assume a significance level of 0.05. **[12 points].**

(e) What is the power of the test in (d)? Describe explicitly the steps needed to calculate the power when $\beta_2,\ \beta_3$ and σ^2 are given. **[12 points].**

Formula Sheet

You may assume any of the following without proof.

$$
\begin{aligned}
\sum_{k=1}^{12m}\left(k-\frac{12m+1}{2}\right) &= 0,\\
\sum_{k=1}^{12m}\left(k-\frac{12m+1}{2}\right)^2 &= m(12m+1)(12m-1),\\
\sum_{k=1}^{12m}\cos\left(\frac{2\pi k}{12}\right) &= 0,\\
\sum_{k=1}^{12m}\sin\left(\frac{2\pi k}{12}\right) &= 0,\\
\sum_{k=1}^{12m}\cos^2\left(\frac{2\pi k}{12}\right) &= 6m,\\
\sum_{k=1}^{12m}\sin^2\left(\frac{2\pi k}{12}\right) &= 6m,\\
\sum_{k=1}^{12m}\cos\left(\frac{2\pi k}{12}\right)\sin\left(\frac{2\pi k}{12}\right) &= 0,\\
\sum_{k=1}^{12m}\left(k-\frac{12m+1}{2}\right)\cos\left(\frac{2\pi k}{12}\right) &= 6m,\\
\sum_{k=1}^{12m}\left(k-\frac{12m+1}{2}\right)\sin\left(\frac{2\pi k}{12}\right) &= -12m\left(1+\frac{\sqrt{3}}{2}\right).
\end{aligned}
$$

The inverse of the matrix $\begin{pmatrix} A & 0 & 0 & 0 \\ 0 & B & C & D \\ 0 & C & C & 0 \\ 0 & D & 0 & C \end{pmatrix}$ is of the form $\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & -b & c \\ 0 & -b & d & -c \\ 0 & c & -c & e \end{pmatrix}$ where

$$
\begin{aligned}
a &= \frac{1}{A},\\
b &= -\frac{C}{D^2+C^2-BC},\\
c &= \frac{D}{D^2+C^2-BC},\\
d &= \frac{D^2-CB}{C(D^2+C^2-BC)},\\
e &= \frac{C-B}{D^2+C^2-BC}.
\end{aligned}
$$

SOLUTIONS

(a) $X = \begin{pmatrix} 1 & 1-\frac{12m+1}{2} & \cos\left(\frac{2\pi}{12}\right) & \sin\left(\frac{2\pi}{12}\right) \\ 1 & 2-\frac{12m+1}{2} & \cos\left(\frac{4\pi}{12}\right) & \sin\left(\frac{4\pi}{12}\right) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 12m-\frac{12m+1}{2} & \cos\left(\frac{24m\pi}{12}\right) & \sin\left(\frac{24m\pi}{12}\right) \end{pmatrix}$;

$X^TX = \begin{pmatrix} 12m & 0 & 0 & 0 \\ 0 & m(12m+1)(12m-1) & 6m & -12m\left(1+\frac{\sqrt{3}}{2}\right) \\ 0 & 6m & 6m & 0 \\ 0 & -12m\left(1+\frac{\sqrt{3}}{2}\right) & 0 & 6m \end{pmatrix}$;

Defining $A = 12m,\ B = m(12m+1)(12m-1),\ C = 6m,\ D = -12m\left(1+\frac{\sqrt{3}}{2}\right)$ and a, b, c, d, e by the expressions on the last five lines of the formula sheet,

$(X^TX)^{-1} = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & -b & c \\ 0 & -b & d & -c \\ 0 & c & -c & e \end{pmatrix}$.

(b) The estimators are given by

$$\begin{aligned} \hat{\beta}_0 &= a\sum Y = \frac{1}{12m}\sum Y_t = \bar{Y}, \text{ (any of these is acceptable)} \\ \hat{\beta}_1 &= b\sum Y_t x_{1t} - b\sum Y_t x_{2t} + c\sum Y_t x_{2t}, \\ \hat{\beta}_2 &= -b\sum Y_t x_{1t} + d\sum Y_t x_{2t} - c\sum Y_t x_{2t}, \\ \hat{\beta}_3 &= c\sum Y_t x_{1t} - c\sum Y_t x_{2t} + e\sum Y_t x_{2t}. \end{aligned}$$

The standard errors are respectively $\sqrt{a}s$, $\sqrt{b}s$, $\sqrt{d}s$, $\sqrt{e}s$, where s is the sample standard deviation.

(c) A t-test for the null hypothesis $H_0:\ \beta_1 = 0$ will reject H_0 when

$$\left|\frac{\hat{\beta}_1}{\sqrt{b}s}\right| > \texttt{qt(0.975, 12*m-4)}$$

where the right hand side is R notation for the 0.975-probability point of the t_{12m-4} distribution (other standard notations that are equivalent to this will be accepted).

(d) When H_0 is true, the model contains just β_0 and β_1, $X^TX = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$, $(X^TX)^{-1} = \begin{pmatrix} 1/A & 0 \\ 0 & 1/B \end{pmatrix}$, and the estimators are $\tilde{\beta}_0 = \bar{Y}$ as before, and $\tilde{\beta}_1 = \sum Y_t x_{1t}/B$. We define SSE_0 and SSE_1 to be the error sums of squares under H_0 and H_1 respectively, given by $\text{SSE}_0 = \sum\left(Y_t - \bar{Y} - \tilde{\beta}_1 x_{1t}\right)^2$ and $\text{SSE}_1 = \sum\left(Y_t - \bar{Y} - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t}\right)^2$. The F test is based on

$$F = \frac{(\text{SSE}_0 - \text{SSE}_1)/q}{(\text{SSE}_1)/(n-p)}$$

which has the $F_{q,n-p}$ distribution when H_0 is true, where for this problem, $q = 2,\ n = 12m,\ p = 4$. Therefore, we reject H_0 at significance level 0.05 when

$$\frac{(6m-2)(\mathrm{SSE}_0 - \mathrm{SSE}_1)}{\mathrm{SSE}_1} > \mathtt{qf}(0.95, 2, 12*\mathtt{m}-4) \tag{2}$$

where we have again used R notation for the probability point of the F distribution; alternative expressions equivalent to (2) will earn full credit.

(e) The null hypothesis may be written in the form $C\boldsymbol{\beta} = \mathbf{h}$ where $C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{pmatrix}^T$, $\mathbf{h} = \begin{pmatrix} 0 & 0 \end{pmatrix}^T$. The alternative hypothesis $C\boldsymbol{\beta} = \mathbf{h}'$ corresponds to $\mathbf{h}' = \begin{pmatrix} \beta_2 & \beta_3 \end{pmatrix}^T$. We calculate

$$C(X^TX)^{-1}C^T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & -b & c \\ 0 & -b & d & -c \\ 0 & c & -c & e \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} d & -c \\ -c & e \end{pmatrix}$$

and hence

$$\left\{C(X^TX)^{-1}C^T\right\}^{-1} = \frac{1}{de-c^2} \begin{pmatrix} e & c \\ c & d \end{pmatrix}.$$

So the noncentrality parameter is defined by

$$\sigma^2\lambda = \frac{\beta_2^2 e + 2c\beta_2\beta_3 + d\beta_3^2}{de-c^2} \tag{3}$$

(or write δ^2 in place of λ; equivalent notation). The power of the test is then given in R notation by `1-pf(fstar,2,12*m-4,lambda)` where `fstar=qf(0.95,2,12*m-4)` in (2), and `lambda` is the same as λ in (3).

Note on the solution to (e)

For this example, I think the method based on $\left\{C(X^TX)^{-1}C^T\right\}^{-1}$ is the simplest way of calculating the noncentrality parameter, but the alternative method (substitution rule) also works, even if it is a little tricky to prove that the two methods in fact lead to the same answer. To apply the substitution rule in this case, begin by writing

$$\begin{aligned}
\sum \left(Y_t - \bar{Y} - \tilde{\beta}_1 x_{1t}\right)^2 &= \sum \left(Y_t - \bar{Y} - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t} + (\hat{\beta}_1 - \tilde{\beta}_1)x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t}\right)^2 \\
&= \sum \left(Y_t - \bar{Y} - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t}\right)^2 + \sum \left((\hat{\beta}_1 - \tilde{\beta}_1)x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t}\right)^2 \\
&\quad + 2\sum \left(Y_t - \bar{Y} - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t}\right)\left((\hat{\beta}_1 - \tilde{\beta}_1)x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t}\right)
\end{aligned}$$

which we may also write as

$$\begin{aligned}
\mathrm{SSE}_0 - \mathrm{SSE}_1 &= \sum \left((\hat{\beta}_1 - \tilde{\beta}_1)x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t}\right)^2 \\
&\quad + 2\sum \left(Y_t - \bar{Y} - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t} - \hat{\beta}_3 x_{3t}\right)\left((\hat{\beta}_1 - \tilde{\beta}_1)x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t}\right).
\end{aligned} \tag{4}$$

According to the substitution rule, the right hand side of (4) is evaluated where each value of Y_t is replaced by its expected value under the alternative hypothesis, which is $\beta_0+\beta_1 x_{1t}+\beta_2 x_{2t}+\beta_3 x_{3t}$. Note that this immediately implies that each of $\hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\beta}_2,\ \hat{\beta}_3$ is equal to the corresponding true parameter value (because, for example, each of these estimators is unbiased under H_1, so if we remove all the random errors, they must be equal to their true values). From this argument, it follows immediately that the second term in the right hand side of (4) is 0.

The first term, however, is a little more complicated, because under the substitution rule, $\tilde{\beta}_1$ is not the same as β_1 when either β_2 or β_3 is non-zero. In fact, $\tilde{\beta}_1=\frac{1}{B}\sum_t Y_t x_{1t}$ is replaced by

$$\frac{1}{B}\sum_t\left(\beta_0+\beta_1 x_{1t}+\beta_2 x_{2t}+\beta_3 x_{3t}\right)x_{1t} \quad = \quad \frac{1}{B}\left(\beta_1 B+\beta_2 C+\beta_3 D\right)$$

and hence $\sum\left((\hat{\beta}_1-\tilde{\beta}_1)x_{1t}+\hat{\beta}_2 x_{2t}+\hat{\beta}_3 x_{3t}\right)^2$ becomes

$$\begin{aligned} & \sum\left(-\frac{\beta_2 C+\beta_3 D}{B}x_{1t}+\beta_2 x_{2t}+\beta_3 x_{3t}\right)^2 \\ = \quad & \frac{(\beta_2 C+\beta_3 D)^2}{B}+\beta_2^2 C+\beta_3^2 C-2C\frac{\beta_2 C+\beta_3 D}{B}\beta_2-2D\frac{\beta_2 C+\beta_3 D}{B}\beta_3. \end{aligned} \tag{5}$$

By the substitution rule, the right hand side of (5) is $\sigma^2\lambda$.

This would be a fully acceptable answer, expressing $\sigma^2\lambda$ as a function of the unknowns $\beta_2,\ \beta_3$ and the known quantities B, C, D. Naturally, however, we would like to show that this is the same answer as the one derived from (3), so we now do that.

First, let us rearrange (5) into the form

$$\sigma^2\lambda \quad = \quad \frac{BC-C^2}{B}\beta_2^2-\frac{2CD}{B}\beta_2\beta_3+\left(C-\frac{D^2}{B}\right)\beta_3^2. \tag{6}$$

Starting from (3), we manipulate the algebraic expressions for c, d, e to derive

$$de-c^2 \quad = \quad -\frac{B}{C(D^2+C^2-BC)}.$$

Hence

$$\frac{e}{de-c^2} \quad = \quad -\frac{C-B}{D^2+C^2-BC}\cdot\frac{C(D^2+C^2-BC)}{B} \quad = \quad \frac{C(B-C)}{B}.$$

This is the same as the coefficient of β_2^2 in (6). Similar manipulations apply to the coefficients of $2\beta_2\beta_3$ and β_3^2, proving the equivalence of (3) and (6).