COMPREHENSIVE WRITTEN EXAMINATION, PAPER III PART 1: FRIDAY AUGUST 12, 2022 9:00 A.M.–11:00 A.M. STOR 664 Theory Question (50 points)

Consider a regression with two predictors $x_{i1},\ x_{i2},\ i = 1, \ldots, n$, and assume the model

$$y_i \quad = \quad \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i,\ i = 1, \ldots, n \qquad (1)$$

where $\beta_0, \ldots, \beta_3$ are unknown parameters and $\epsilon_i \sim N[0, \sigma^2]$ are independent errors with common unknown variance σ^2. Note that the model has *just* the interaction term $\beta_3 x_{i1} x_{i2}$ but no terms in x_{i1}^2 or x_{i2}^2. It is natural to want to test whether $H_0 :\ \beta_3 = 0$.

Defining $S_{jk} = \sum_{i=1}^n x_{i1}^j x_{i2}^k$ let us further assume: $S_{10} = S_{01} = S_{11} = S_{12} = 0$ but that $S_{20},\ S_{02},\ S_{21}$ and S_{22} are not 0.

(a) Find explicit expressions for the least squares estimators $\hat{\beta}_0, \ldots, \hat{\beta}_3$ in terms of the S_{jk}'s and $\sum y_i,\ \sum y_i x_{i1},\ \sum y_i x_{i2}$ and $\sum y_i x_{i1} x_{i2}$. **[12 points]**

(b) Find expressions for the standard errors of $\hat{\beta}_0, \ldots, \hat{\beta}_3$, in terms of $n,\ S_{20},\ S_{02},\ S_{21},\ S_{22}$ and the residual standard deviation s (assuming that s^2 is the standard unbiased estimator of σ^2). **[5 points]**

(c) A t-test of significance level α will reject H_0 if $|\frac{\hat{\beta}_3}{s}| > C$ for some C which is a combination of $n,\ S_{20},\ S_{02},\ S_{21},\ S_{22}$ and α (alpha). Find C. (You may, if you wish, express your answer as an appropriate R function.) **[5 points]**

(d) What is the power of the test in (c) when $\beta_3 \neq 0$? You answer should be expressed in terms of the given parameters and relevant percentage points of the noncentral t or F distributions. (You may choose to express your answer as an R function though alternatives are also acceptable if the derivation behind your answer is clearly explained.) (*Hint:* First find the distribution of $\frac{\hat{\beta}_3^2}{s^2}$ when $\beta_3 \neq 0$.) **[16 points]**

(e) Show that the observation with highest leverage is the index i that maximizes

$$x_{i2}^2 (S_{20} S_{22} - S_{21}^2) + x_{i1}^2 S_{02} (S_{22} - 2 S_{21} x_{i2} + x_{i2}^2 S_{20}).$$

[12 points]