Columbia University
Statistical Analysis and Time Series

IEOR-4709

A. Capponi **Problem Set #4** Issued: March 24, 2025
Spring 2025 Due: **BEFORE CLASS** April 2, 2025

Note: Please put the number of hours that you spent on this homework set on top of the first page of your homework. The TA in charge of this homework review session is Jose Sidaoui Gali.

The notation in the problem below is the same as the notation adopted in class.

Ex. 1.
Consider the simple linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

where ϵ_i is a zero-mean random variable. Moreover, ϵ_i's are independent Gaussian with mean zero and variance σ^2. Do the following:

- Find the MLE estimators for α and β given observed samples $x_{1:n} = (x_1, \ldots, x_n)$ and $y_{1:n} = (y_1, \ldots, y_n)$.
- Show that the MLE estimator of σ^2 is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2$.

Ex. 2.
Assume the error terms ϵ_i's in a linear regression model are independent Gaussian with mean zero and variance σ^2. Show that the variance and covariance estimators in the simple linear regression model are

$$Var(\hat{\beta}) = S_{xx}^{-1}\sigma^2$$
$$Cov(\hat{\alpha}, \hat{\beta}) = -\bar{x}S_{xx}^{-1}\sigma^2$$
$$Var(\hat{\alpha}) = \frac{\sigma^2}{n} + \bar{x}^2 S_{xx}^{-1}\sigma^2$$

Ex. 3.
Consider the following linear regression model:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

where x_1 and x_2 and the constant are three explanatory variables, y is the response variable, and ϵ is normally distributed noise. Suppose that 20 data points of $(x_{i,1}, x_{i,2}, y_i)$, $i = 1, \ldots, 20$ are observed. The following are some simple statistics:

$$\bar{y} := \frac{1}{20}\sum_{i=1}^{20} y_i = 5; \qquad \sum_{i=1}^{20}(y_i - \bar{y})^2 = 20; \qquad \sum_{i=1}^{20}\hat{\epsilon}_i^2 = 10$$

where ϵ_i's are the residuals of y_i's in the regression model. Do the following

(a) Find an unbiased estimate of the variance of ϵ.

(b) Test whether $\beta_1 = \beta_2 = 0$. Use 1% as the significance level.

(c) Calculate the R-squared of the model.

Ex. 4.
Consider a multiple linear regression model

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon.$$

It is desired to test the null hypothesis $H_0 : c_2\beta_2 + c_3\beta_3 = 0$ versus the alternative hypothesis $H_1 : c_2\beta_2 + c_3\beta_3 \neq 0$, where c_2, c_3 are real numbers. Provide the exact rule for rejecting the null hypothesis if the desired significance level of the test is ρ.

Ex. 5.
Consider the monthly return data of IBM and S&P500 index in period Janaury 2010 through December 2023. You can download the data set from Yahoo finance or any other sources (please, report the source). Do the following:

(a) Test December effect, i.e., test whether the returns of IBM in December has different alpha and beta from the returns in the other months.

(b) Provide a 95% confidence interval for the difference between the beta of IBM returns in December and the beta of IBM returns in the other months.