MA 550 – Homework 1

Due Wed, 11:30am, September 16, 2020

Homework Assignment Policy and Guidelines

(a) Homework assignment should be submitted electronically through `Canvas`, and your submission should be combined into ONE PDF file. Credit will not be given for homework turned in late.

(b) Homework assignments should be well organized. It is required that you show your work in order to receive credit.

(c) It's highly recommended to use R Markdown to write up your homework solutions, especially for problems that involve data analysis. But this is not required.

(d) When writing up solutions via R Markdown, please refrain from showing lengthy results/output for data analysis problems. Please display only relevant results.

(e) If you are not to use R Markdown, please keep the R code and/or R output for all the relevant problems in a *well organized appendix*, and please refrain from copying and pasting R code and/or R output directly into your answers.

(f) For problems that involve technical writing, strive to be clear, concise, and cogent. Organize figures and tables in an efficient manner while maintaining clarity.

(g) For problems that involve data analysis, always interpret the results in the context of the study. This may include comments on whether the results are meaningful or not. Think of your answers as a mini technical report of your findings and follow the guidelines on technical writing.

(h) You may discuss most homework problems with others including your peers and instructor, but you must write up your homework solutions by yourself in order to receive credit. Similarly, you must write your own computer code and obtain your computer output independently.

1. Consider Slides #1, Page 28.

(a) Find the covariance matrices $Cov(\boldsymbol{e}, \boldsymbol{Y})$ and $Cov(\boldsymbol{e}, \hat{\boldsymbol{Y}})$.

(b) The estimator of σ^2 is $\widehat{\sigma}^2 = (n-p)^{-1}\boldsymbol{e}'\boldsymbol{e}$. Show that $\mathrm{E}(\widehat{\sigma}^2) = \sigma^2$ and $\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}$ follows a Chi-square distribution with degrees of freedom $n-p$. *Hint: check Rencher & Schaalje (2008)*[1] *Page 107 Theorem 5.2a and Page118 Corollary 2.*

(c) Show that $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\widehat{\sigma}^2$ are independent. *Hint: check Rencher & Schaalje (2008) Page 119 Theorem 5.6a.*

2. Prove the following results in Slides #1:

[1] Rencher and Schaalje (2008): Linear Models in Statistics (2nd ed), Wiley.

(a) Page #26: $\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$;

(b) Page #27: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ and $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$;

(c) Page #29: $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$;

(d) Page #30: $\frac{\hat{\beta}_k - \beta_k}{S\{\hat{\beta}_k\}} \sim T_{n-p}$. *Hint: consider Problem 1 (b) and (c).*

3. Consider the model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ where $\varepsilon_i \sim \text{ind}N(0, \sigma_i^2)$ for observations $i = 1, \ldots, n$. Suppose $X_i > 0$ and $\sigma_i^2 = \sigma^2 X_i$. Let $\tilde{Q}(\boldsymbol{\beta}) = \sum_{i=1}^n \sigma_i^{-2}\{Y_i - (\beta_0 + \beta_1 X_i)\}^2$ denote the weighted error sum of squares. Define

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$: $n \times 1$ vector of response variables

$\boldsymbol{X}$: $n \times 2$ design matrix with 1's in the first column and $(X_1, \ldots, X_n)'$ in the second column.

$\boldsymbol{\beta} = (\beta_0, \beta_1)'$: 2×1 vector of regression coefficients

$\boldsymbol{V} = \text{diag}\{X_1, \ldots, X_n\}$: $n \times n$ diagonal matrix with $X_1, \ldots, X_n$ along the diagonal

(a) Derive the distribution of $\boldsymbol{Y}$ including the analytical forms of the mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

(b) Write the weighted error sum of squares $\tilde{Q}(\boldsymbol{\beta})$ in matrix terms.

(c) Let $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_0, \tilde{\beta}_1)'$ denote the weighted least squares estimates of $\boldsymbol{\beta}$. Derive $\tilde{\boldsymbol{\beta}}$ by minimizing $\tilde{Q}(\boldsymbol{\beta})$.

(d) Derive the mean vector and the variance-covariance matrix of $\tilde{\boldsymbol{\beta}}$.

(e) Despite the weighted variances, derive the ordinary least squares estimates of $\boldsymbol{\beta}$ by minimizing the (unweighted) error sum of squares $Q(\boldsymbol{\beta}) = \sum_{i=1}^n \{Y_i - (\beta_0 + \beta_1 X_i)\}^2$ all in matrix terms. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$ denote the (ordinary) least squares estimates of $\boldsymbol{\beta}$.

(f) Derive the mean vector and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$.

(g) Draw a connection between $Var(\hat{\beta}_1)$ and $Var(\tilde{\beta}_1)$.

4. The file 'MortBond.txt' contains monthly effective interest rate z_t for conventional single-family mortgages from Jan. 1973 to Dec. 2003.

(a) Draw threes plots: time plot of the time series, scatter plots of z_{t+1} versus z_t and of z_{t+2} versus z_t.

(b) Based on the plots in (a), do you think that the series is autocorrelated? Comment on whether you think the time series is stationary.

(c) Plot the sample autocorrelation for this series up to lag 40. Relate this plot to the plots in (a).

5. Consider the daily simple return of CRSP equal-weighted index from January 1980 to December 1999 in the file 'DEWRet.csv'. Use a regression model to study the effects of trading days on the index return. What is the fitted model? Are the weekday effects significant in the returns at the 5% level? Use the Newey-West estimator of the covariance matrix to obtain the t-values of regression estimates. Does it change the conclusion of weekday effect?

6. Consider the daily returns of S&P composite index from January 3, 2000 to December 31, 2003. The data are in the file 'SP.csv'. Perform all tests using the 5% significance level, and answer the following questions:

(a) Is there a Friday effect on the daily simple returns of S&P composite index? You may employ a simple linear regression model to answer this question. Estimate the model via Newey-West estimator and test the hypothesis that there is no Friday effect. Draw your conclusion.

(b) Check the residual serial correlations. Are there any significant serial correlations in the residuals?