Problem Set 3

Due Oct 17, before class

You may use the following results:

- The inverse of block matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

- For $X = (X_1, X_2)$

$$(X^T X)^{-1} = \begin{pmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{pmatrix}^{-1} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

where

$$\begin{aligned} S_{11} &= (X_1^T X_1)^{-1} + (X_1^T X_1)^{-1} X_1^T X_2 (\tilde{X}_2^T \tilde{X}_2)^{-1} X_2^T X_1 (X_1^T X_1)^{-1}, \\ S_{12} &= -(X_1^T X_1)^{-1} X_1^T X_2 (\tilde{X}_2^T \tilde{X}_2)^{-1}, \\ S_{21} &= S_{12}^T, \\ S_{22} &= (\tilde{X}_2^T \tilde{X}_2)^{-1} \end{aligned}$$

with $\tilde{X}_2 = \left\{ I - X_1 (X_1^T X_1)^{-1} X_1^T \right\} X_2$.

1. Projection matrix decomposition

Show that $H = H_1 + \tilde{H}_2$, where $H = X(X^T X)^{-1}X^T$, $H_1 = X_1(X_1^T X_1)^{-1}X_1^T$, and $\tilde{H}_2 = \tilde{X}_2(\tilde{X}_2^T \tilde{X}_2)^{-1}\tilde{X}_2^T$.

2. Alternative formula for the EHW standard error

Consider the partition regression $Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\epsilon}$, where X_1 is an $n \times (p-1)$ matrix and X_2 is an n dimensional vector. So $\hat{\beta}_2$ is a scalar and the (p,p)-th element of $\hat{V}_{\text{EHW}}$ equals $\hat{\text{se}}^2_{\text{EHW},2}$, the squared EHW standard error for $\hat{\beta}_2$. Define

$$\tilde{X}_2 \;=\; \left\{I - X_1(X_1^T X_1)^{-1}X_1^T\right\} X_2 \;=\; \begin{pmatrix} \tilde{x}_{12} \\ \vdots \\ \tilde{x}_{n2} \end{pmatrix}.$$

Show that under the heteroskedastic linear model,

$$\text{var}(\hat{\beta}_2) \;=\; \sum_{i=1}^n w_i \sigma_i^2, \quad \hat{\text{se}}^2_{\text{EHW},2} \;=\; \sum_{i=1}^n w_i \hat{\epsilon}_i^2,$$

where

$$w_i \;=\; \frac{\tilde{x}_{i2}^2}{(\sum_{i=1}^n \tilde{x}_{i2}^2)^2}.$$

3. QR decomposition and the leverage score

Based on the QR decomposition of X, show that

$$H \;=\; QQ^T,$$

and h_{ii} equals the squared length of the i-th row of Q.

4. Formula of the partial correlation coefficient

For $Y, X, W \in \mathbb{R}^n$, show that

$$\hat{\rho}_{yx|w} = \frac{\hat{\rho}_{yx} - \hat{\rho}_{yw}\hat{\rho}_{xw}}{\sqrt{1-\hat{\rho}_{yw}^2}\sqrt{1-\hat{\rho}_{xw}^2}}.$$

5. R^2 and the sample Pearson correlation coefficient

Show that $R^2 = \hat{\rho}_{y\hat{y}}^2$, where

$$\hat{\rho}_{y\hat{y}} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}}.$$

6. Relationship between the standardized and studentized residuals

Under the Gauss-Markov model, let $\hat{\sigma}^2_{[-i]}$ denote the unbiased variance estimator for σ^2 based on the leave-i-out OLS. Show that

(a) $(n-p-1)\hat{\sigma}^2_{[-i]} = (n-p)\hat{\sigma}^2 - \hat{\epsilon}_i^2/(1-h_{ii})$

(b) There is a monotone relationship between the standardized and studen- tized residual

$$\text{studr}_i = \text{standr}_i \sqrt{\frac{n-p-1}{n-p-\text{standr}_i^2}}$$

Remark: The right hands of the equality in (a) must be non-negative, which implies

$$h_{ii} + \frac{\hat{\epsilon}_i^2}{\sum_{k=1}^n \hat{\epsilon}_k^2} \le 1.$$

Therefore, if $h_{ii} = 1$, then $\hat{\epsilon}_i = 0$, which further implies that $h_{ij} = 0$ for all $j \ne i$.

7. More on the leverage score

Show that

$$\det(X_{[-i]}^T X_{[-i]}) \;=\; (1 - h_{ii})\det(X^T X).$$

Hint: you can use the following result

$$\det(A + uv^T) \;=\; (1 + v^T A^{-1} u) \cdot \det(A).$$

Remark: If $h_{ii} = 1$, then $X_{[-i]}^T X_{[-i]}$ is degenerate with determinant 0. Therefore, if we delete an observation i with leverage score 1, the columns of the covariate matrix become linearly dependent.

8. Measurement error and Frisch's bounds

We consider population OLS in this problem. Given scalar random variables x and y, we can obtain the population OLS coefficients (α, β) of y on $(1, x)$. However, x and y may be measured with errors. We only observe $x^* = x + u$ and $y^* = y + v$, where u and v are mean zero error terms satisfying $u \perp\!\!\!\perp v$ and $(u, v) \perp\!\!\!\perp (x, y)$. We can obtain the population OLS coefficient (α^*, β^*) of y^* on $(1, x^*)$ and the population OLS coefficient (a^*, b^*) of x^* on $(1, y^*)$. Show that if $\beta = 0$, then $\beta^* = b^* = 0$; if $\beta \neq 0$, then

$$|\beta^*| \;\leq\; |\beta| \;\leq\; 1/|b^*|.$$

9. Real data analysis: leave-one-out prediction

Under the Normal linear model, obtain the leave-one-out prediction intervals for the Boston Housing data (You need to install R package *mlbench* to get the data). That is, for each observation i in the data, obtain the prediction interval for y_i (variable name: “medv”) based on the OLS with the other $n-1$ observations. Plot the prediction intervals for all observations.

Hint: You can use the leave-one-out formulas and problem 6(a) to avoid fitting n OLS.