Solutions for Homework 2

1. Consider a bivariate linear regression model

$$Y_t = X'_t \beta^0 + \varepsilon_t, \quad t = 1, \cdots, n,$$

where $X_t = (X_{0t}, X_{1t}) = (1, X_{1t})'$, and ε_t is a regression error. Let $\hat{\rho}$ denote the sample correlation between Y_t and $X_1 t$; namely,

$$\hat{\rho} = \frac{\sum_{t=1}^n x_{1t} y_t}{\sqrt{\sum_{t=1}^n x_{1t}^2 \sum_{t=1}^n y_t^2}},$$

where $y_t = Y_t - \bar{Y}$, $x_{1t} = X_{1t} - \bar{X}_1$, and $\bar{Y}$ and $\bar{X}_1$ are the sample means of Y_t and $X_1 t$. Show $R^2 = \hat{\rho}^2$.

ANSWER:

$$y_t = Y_t - \bar{Y} = \beta_1^0 + X_{1t}\beta_2^0 + \varepsilon_t - \left(\beta_1^0 + \bar{X}_1 \beta_2^0\right) = x_{1t}\beta_2^0 + \varepsilon_t$$

$$\hat{\beta}_2^0 = \arg\min \sum \left(y_t - x_{1t}\beta_2^0\right)^2 = \sum_{t=1}^n x_{1t} y_t / \sum_{t=1}^n x_{1t}^2$$

$$R^2 = \frac{\sum\limits_{t=1}^n \left(\hat{Y}_t - \bar{Y}\right)^2}{\sum\limits_{t=1}^n \left(Y_t - \bar{Y}\right)^2} = \frac{\sum\limits_{t=1}^n \hat{y}_t^2}{\sum\limits_{t=1}^n y_t^2} = \frac{\sum\limits_{t=1}^n \left(x_{1t}\hat{\beta}_2^0\right)^2}{\sum\limits_{t=1}^n y_t^2} = \frac{\left(\hat{\beta}_2^0\right)^2 \sum\limits_{t=1}^n x_{1t}{}^2}{\sum\limits_{t=1}^n y_t^2} = \frac{\sum\limits_{t=1}^n \left(x_{1t} y_t\right)^2}{\sum\limits_{t=1}^n x_{1t}^2 \sum\limits_{t=1}^n y_t^2} = \hat{\rho}$$

This completes the proof.

2. Consider the following linear regression model

$$Y_t = X'_t \beta^0 + u_t, \quad t = 1, \cdots, n, \tag{1}$$

where

$$u_t = \sigma\left(X_t\right) \varepsilon_t,$$

where X_t is a nonstochastic process, and $\sigma\left(X_t\right)$ is a positive function of X_t such that

$$\Omega = \begin{bmatrix} \sigma^2\left(X_1\right) & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2\left(X_2\right) & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2\left(X_3\right) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sigma^2\left(X_n\right) \end{bmatrix} = \Omega^{\frac{1}{2}} \Omega^{\frac{1}{2}}.$$

with

$$\Omega^{\frac{1}{2}} = \begin{bmatrix} \sigma\left(X_1\right) & 0 & 0 & \cdots & 0 \\ 0 & \sigma\left(X_2\right) & 0 & \cdots & 0 \\ 0 & 0 & \sigma\left(X_3\right) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sigma\left(X_n\right) \end{bmatrix}$$

Assume that ε_t is i.i.d. $N(0,1)$. Then u_t is i.i.d. $N\left(0, \sigma^2\left(X_t\right)\right)$. This differs from Assumption 3.5 of the classical linear regression analysis, because now u_t exhibits conditional heteroskedasticity.

Let $\hat{\beta}$ denote the OLS estimator for β^0.

(a) Is $\hat{\beta}$ unbiased for β^0?
(b) Show that var $\left(\hat{\beta}\right) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$.
Consider an alternative estimator

$$\begin{aligned}\tilde{\beta} &= (X'\Omega X)^{-1}X'\Omega Y \\ &= \left[\sum_{t=1}^{n}\sigma^{-2}(X_t)X_tX'_t\right]^{-1}\sum_{t=1}^{n}\sigma^{-2}(X_t)X_tY\end{aligned}$$

(c) Is $\tilde{\beta}$ unbiased for β^0? (d) Show that var $\left(\tilde{\beta}\right) = (X'\Omega X)^{-1}$. (e) Is var $\left(\hat{\beta}\right)$ − var $\left(\tilde{\beta}\right)$ negative semi-definite (n.s.d)? Which estimator, $\hat{\beta}$ or $\tilde{\beta}$, is more efficient? (f) Is $\tilde{\beta}$ the Linear Best Unbiased Estimator (BLUE) for β^0? [Hint: There are several approaches to this question. A simple one is to consider the transformed model

$$Y_t^* = X_t^{*\prime}\beta^0 + \varepsilon_t, \quad t = 1, \cdots, n, \tag{2}$$

where $Y_t^* = Y_t/\sigma(X_t)$, $X_t^* = X_t/\sigma(X_t)$. This model is obtained from model (1) after dividing by $\sigma(X_t)$. In matrix notation, model (2) can be written as

$$Y^* = X^{*\prime}\beta^0 + \varepsilon,$$

where the $n \times 1$ vector $Y^* = \Omega^{-\frac{1}{2}}Y$ and the $n \times k$ matrix $X^* = \Omega^{-\frac{1}{2}}X$.]
(g) Construct two test statistics for the null hypothesis of interest $H_0 : \beta_2^0 = 0$. One test is based on $\hat{\beta}$, and the other test is based on $\tilde{\beta}$. What are the finite sample distributions of your test statistics under H_0? Can you tell which test is better?
(h) Construct two test statistics for the null hypothesis of interest $H_0 : R\beta^0 = r$, where R is a $J \times k$ matrix with $J > 0$. One test is based on $\hat{\beta}$, the other test is based on $\tilde{\beta}$. What are the finite sample distributions of your test statistics under H_0?

ANSWER: (a)

$$\begin{aligned}E[\hat{\beta}] &= (X'X)^{-1}X'E[X'\beta^0 + u\,|X] \\ &= (X'X)^{-1}X'X\beta^0 + 0 \\ &= \beta^0\end{aligned}$$

thus, $\hat{\beta}$ is unbiased.

(b)

$$\begin{aligned}
Var\left(\hat{\beta}\right) &= E[(\hat{\beta}-\beta^0)(\hat{\beta}-\beta^0)' \,|X\,] \\
&= E[(X'X)^{-1}X'uu'X(X'X)^{-1} \,|X\,] \\
&= (X'X)^{-1}X'E[uu' \,|X\,]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega X(X'X)^{-1}
\end{aligned}$$

(c)

$$\begin{aligned}
E[\tilde{\beta}] &= E[\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}X'\beta^0 + u\,|X\,] \\
&= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}X\beta^0 + 0 \\
&= \beta^0
\end{aligned}$$

thus, $\tilde{\beta}$ is unbiased.
(d)

$$\begin{aligned}
Var\left(\tilde{\beta}\right) &= E[(\tilde{\beta}-\beta^0)(\tilde{\beta}-\beta^0)' \,|X\,] \\
&= E[\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}uu'\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} \,|X\,] \\
&= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}E[uu' \,|X\,]\Omega^{-1}X\left(X'\Omega^{-1}X\right)^{-1} \\
&= \left(X'\Omega^{-1}X\right)^{-1}
\end{aligned}$$

(e) Let $c = \Omega^{-\frac{1}{2}}X(X'X)^{-1}$

$$\begin{aligned}
Var\left(\hat{\beta}\right) - Var\left(\tilde{\beta}\right) &= (X'X)^{-1}X'\Omega X(X'X)^{-1} - \left(X'\Omega^{-1}X\right)^{-1} \\
&= c'c - c'\Omega^{-\frac{1}{2}}X\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-\frac{1}{2}}c \\
&= c'\left(I - \Omega^{-\frac{1}{2}}X\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-\frac{1}{2}}\right)c
\end{aligned}$$

It is easy to proof that $I - \Omega^{-\frac{1}{2}}X\left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-\frac{1}{2}}$ is an idempotent matrix. So $Var\left(\hat{\beta}\right) - Var\left(\tilde{\beta}\right)$ is p.s.d. and $\tilde{\beta}$ is more efficient.
(f) For model

$$Y^* = X^*\beta^0 + \varepsilon$$

$\hat{\beta}^*$ is BLUE. And we have

$$\begin{aligned}\hat{\beta}^* &= \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}Y^* \\ &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}Y \\ &= \tilde{\beta}\end{aligned}$$

So, $\tilde{\beta}$ is also BLUE.
(g) we have proved that

$$\begin{aligned}\hat{\beta} &\sim N\left(\beta^0, (X'X)^{-1}X'\Omega X(X'X)^{-1}\right) \\ \tilde{\beta} &\sim N\left(\beta^0, (X'\Omega X)^{-1}\right)\end{aligned}$$

Define $R = (0, 0, 1, 0, \cdots, 0)$, then

$$H_0: \quad R\beta^0 = 0$$

If Ω is known, then

$$\frac{R\hat{\beta} - r}{\sqrt{R(X'X)^{-1}X'\Omega X(X'X)^{-1}R'}} \sim N(0, 1)$$

$$\frac{R\tilde{\beta} - r}{\sqrt{R\left(X'\hat{\Omega}^{-1}X\right)^{-1}R'}} \sim N(0, 1)$$

If Ω is unknown, then

$$t_{\hat{\beta}} = \frac{R\hat{\beta} - r}{\sqrt{R(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}R'}}$$

$$t_{\tilde{\beta}} = \frac{R\tilde{\beta} - r}{\sqrt{R\left(X'\hat{\Omega}^{-1}X\right)^{-1}R'}}$$

Both $t_{\hat{\beta}}$ and $t_{\tilde{\beta}}$ valued for large sample, they have no exact distribution for finite sample. So, we can not tell which one is better.
(h)

$$H_0: \quad R\beta^0 = r$$

For Ω is known,

$$F_{\hat{\beta}} = \left(R\hat{\beta} - r\right)'\left[R(X'X)^{-1}X'\Omega X(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta} - r\right) \sim \chi^2_J$$

$$F_{\tilde{\beta}} = \left(R\tilde{\beta} - r\right)'\left[R\left(X'\Omega^{-1}X\right)^{-1}R'\right]^{-1}\left(R\tilde{\beta} - r\right) \sim \chi^2_J$$

For Ω is unknown,

$$W_{\hat{\beta}} = \left(R\hat{\beta} - r\right)'\left[R(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta} - r\right) \xrightarrow{d} \chi^2_J$$

$$W_{\tilde{\beta}} = \left(R\tilde{\beta} - r\right)' \left[R\left(X'\hat{\Omega}^{-1}X\right)^{-1}R'\right]^{-1} \left(R\tilde{\beta} - r\right) \xrightarrow{d} \chi_J^2$$

Both $W_{\hat{\beta}}$ and $W_{\tilde{\beta}}$ converge in distribution to χ_J^2. They have no exact distribution for finite sample, So we can not tell which one is better.

3. Suppose $X'X$ is a $K \times K$ matrix, and V is a $N \times n$ matrix, and both $X'X$ and V are symmetric and nonsingular, with the minimum eigenvalue $\lambda_{\min}(X'X) \to \infty$ as $n \to \infty$ and $0 < c \leq \lambda_{\max}(V) \leq C < \infty$. Show that for any $\tau \in R^K$ such that $\tau'\tau = 1$,

$$\tau'\text{var}\left(\hat{\beta}\,|X\right)\tau = \sigma^2\tau'(X'X)^{-1}X'VX(X'X)^{-1}\tau \to 0$$

as $n \to \infty$. Thus, var $\left(\hat{\beta}\,|X\right)$ vanishes to zero as $n \to \infty$ under conditional heteroskedasticity.

ANSWER:

$$\begin{aligned}
\tau'(X'X)^{-1}X'VX(X'X)^{-1}\tau &\leq \tau'(X'X)^{-1}X'VX(X'X)^{-1}\tau\lambda_{\max}(V) \\
&= \lambda_{\max}(V)\,\tau'(X'X)^{-1}\tau \\
&\leq \lambda_{\max}(V)\,\lambda_{\max}(X'X)^{-1} \\
&= \frac{\lambda_{\max}(V)}{\lambda_{\min}(X'X)} \\
&\leq \ \tfrac{C}{\lambda_{\min}(X'X)} \to 0 \quad\quad as \quad n \to \infty
\end{aligned}$$

Thus $\tau'(X'X)^{-1}X'VX(X'X)^{-1}\tau \to 0$ as $n \to \infty$. This completes the proof.