Regression

1. Find $\mathbb{E}[\mathbb{E}[\mathbb{E}[Y \mid X_1, X_2, X_3] \mid X_1, X_2] \mid X_1]$
2. Suppose that the random variables Y and X only take the values 0 and 1 , and have the following joint probability distribution

	$X = 0$	$X = 1$
$Y = 0$	.1	.2
$Y = 1$	.4	.3

   Find $\mathbb{E}[Y \mid X]$, $\mathbb{E}\left[Y^2 \mid X\right]$ and $\operatorname{var}[Y \mid X]$ for $X = 0$ and $X = 1$.

3. True or False?

   (a) If $Y = X\beta + e, X \in \mathbb{R}$, and $\mathbb{E}[e \mid X] = 0$, then $\mathbb{E}\left[X^2 e\right] = 0$.

   (b) If $Y = X\beta + e, X \in \mathbb{R}$, and $\mathbb{E}[Xe] = 0$, then $\mathbb{E}\left[X^2 e\right] = 0$.

   (c) If $Y = X'\beta + e$ and $\mathbb{E}[e \mid X] = 0$, then e is independent of X.

   (d) If $Y = X'\beta + e$ and $\mathbb{E}[Xe] = 0$, then $\mathbb{E}[e \mid X] = 0$.

   (e) If $Y = X'\beta + e$, $\mathbb{E}[e \mid X] = 0$, and $\mathbb{E}\left[e^2 \mid X\right] = \sigma^2$, then e is independent of X.
4. Take the homoskedastic model

$$\begin{aligned} Y &= X_1'\beta_1 + X_2'\beta_2 + e \\ \mathbb{E}[e \mid X_1, X_2] &= 0 \\ \mathbb{E}\left[e^2 \mid X_1, X_2\right] &= \sigma^2 \\ \mathbb{E}[X_2 \mid X_1] &= \Gamma X_1 \end{aligned}.$$

   Assume $\Gamma \neq 0$. Suppose the parameter β_1 is of interest. We know that the exclusion of X_2 creates omited variable bias in the projection coefficient on X_2. It also changes the equation error. Our question is: what is the effect on the homoskedasticity property of the induced equation error? Does the exclusion of X_2 induce heteroskedasticity or not? Be specific.
5. Let $\hat{\beta}_n = \left(\boldsymbol{X}_n'\boldsymbol{X}_n\right)^{-1}\boldsymbol{X}_n'\boldsymbol{Y}_n$ denote the OLS estimate when $\boldsymbol{Y}_n$ is $n \times 1$ and $\boldsymbol{X}_n$ is $n \times k$. A new observation (Y_{n+1}, X_{n+1}) becomes available. Prove that the OLS estimate computed using this additional observation is

$$\hat{\beta}_{n+1} = \hat{\beta}_n + \frac{1}{1+X_{n+1}'\left(\boldsymbol{X}_n'\boldsymbol{X}_n\right)^{-1}X_{n+1}}\left(\boldsymbol{X}_n'\boldsymbol{X}_n\right)^{-1}X_{n+1}\left(Y_{n+1} - X_{n+1}'\hat{\beta}_n\right).$$

6. Prove that R^2 is the square of the sample correlation between $\boldsymbol{Y}$ and $\widehat{\boldsymbol{Y}}$.
7. For the intercept-only model $Y_i = \beta + e_i$, show that the leave-one-out prediction error is

$$\tilde{e}_i = \left(\frac{n}{n-1}\right)\left(Y_i - \bar{Y}\right).$$

8. The observations are (Y_i, X_{1i}, X_{2i}), $i = 1, \ldots, n$. You estimate two least squares regressions.

$$\begin{aligned} Y_i &= X_{1i}'\tilde{\beta}_1 + \tilde{e}_i \\ Y_i &= X_{1i}'\hat{\beta}_1 + X_{2i}'\hat{\beta}_2 + \hat{e}_i' \end{aligned}$$

   and calculate the residual variance estimates

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{e}_i^2 \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 \end{aligned}.$$

   Show that for any $w \in (0,1)$, there is a constant $a \in (0,1)$ such that

$$\frac{1}{n}\sum_{i=1}^{n}(w\hat{e}_i + (1-w)\tilde{e}_i)^2 = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2.$$

Find the constant a.

9. Consider a regression of a $n \times 1$ vector of responses $\mathbf{Y}$ on a $n \times k$ matrix of explanatory variables $\mathbf{X}$. Let $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$, where $\mathbf{P}$ is the projection matrix $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, and $\mathbf{I}_n$ is the identity matrix of order n. Suppose one adds a new explanatory variable, $\mathbf{Z}$, to the regression, so there are now $(k+1)$ regressors. Show that the new residual sum of squares is given by $\mathbf{Y}'\mathbf{MY} - b^2\mathbf{Z}'\mathbf{MZ}$ where $b = \mathbf{Z}'\mathbf{MY}/\mathbf{Z}'\mathbf{MZ}$ is the OLS coefficient on $\mathbf{Z}$.

1. Find $\mathbb{E}[\mathbb{E}[\mathbb{E}[Y \mid X_1, X_2, X_3] \mid X_1, X_2] \mid X_1]$

$$E[Y/X_1,X_2,X_3] = g(X_1,X_2,X_3)$$

$$E[E[Y/X_1,X_2,X_3]/X_1,X_2] = E[g(X_1,X_2,X_3)/X_1,X_2] = h(X_1,X_2)$$

$$E[E[E[Y/X_1,X_2,X_3]/X_1,X_2]/X_1] = E[h(X_1,X_2)/X_1] = f(X_1)$$

Ojo: Si $f(X_1,X_2) = aX_1 + bX_2$

$$\hookrightarrow E[f(X_1,X_2)/X_1] = aX_1 + bE(X_2)$$
$$= aX_1 + b\int X_2 \cdot f_{X_2/X_1}(X_2/X_1)\, dX_2$$

$$\Rightarrow E[E[E[Y/X_1,X_2,X_3]/X_1,X_2]/X_1] = E[h(X_1,X_2)/X_1]$$

¿Necesario? o solo Ley de formas iteradas

$$= \int E[E[Y/X_1,X_2,X_3]/X_1,X_2]\, f_{X_2/X_1}(X_2/X_1)\, dX_2$$

$$= \iint E[Y/X_1,X_2,X_3]\, f_{X_3/X_2,X_1}(X_3/X_2,X_1)\, dX_3\, f_{X_2/X_1}(X_2/X_1)\, dX_2$$

$$= \iiint Y \cdot f_{Y/X_1,X_2,X_3}(Y/X_1,X_2,X_3)\, dy\, f_{X_3/X_2,X_1}(X_3/X_2,X_1)\, dX_3\, f_{X_2/X_1}(X_2/X_1)\, dX_2$$

$$= \iiint Y\, f_{Y/X_1,X_2,X_3}(Y/X_1,X_2,X_3)\, f_{X_3/X_2,X_1}(X_3/X_2,X_1)\, f_{X_2/X_1}(X_2/X_1)\, dX_2\, dX_3\, dY$$

Ojo: Teorema de Bayes:

$$f_{Y/X_1,X_2,X_3}(Y/X_1,X_2,X_3) = \frac{f(Y,X_1,X_2,X_3)}{f(X_1,X_2,X_3)}$$

$$= \iiint Y\, \frac{f(Y,X_1,X_2,X_3)}{\cancel{f(X_1,X_2,X_3)}} \cdot \frac{\cancel{f(X_1,X_2,X_3)}}{\cancel{f(X_1,X_2)}} \cdot \frac{\cancel{f(X_1,X_2)}}{f(X_1)}\, dX_2\, dX_3\, dY$$

$$= \iiint Y\, \frac{f(Y,X_1,X_2,X_3)}{f(X_1)}\, dY\, dX_3\, dX_2$$

$$= \int Y\, dY \underbrace{\iint f(Y,X_1,X_2,X_3)\, dX_2\, dX_3}\cdot\left(\frac{1}{f(X_1)}\right)$$

$$= \int Y\, dY\, \frac{f(Y,X_1)}{f(X_1)} = \int Y\, f(Y/X_1)\, dy = E[Y/X_1]$$ //

2. Suppose that the random variables Y and X only take the values 0 and 1 , and have the following joint probability distribution

	$X = 0$	$X = 1$
$Y = 0$	.1	.2
$Y = 1$	.4	.3

Find $\mathbb{E}[Y \mid X]$, $\mathbb{E}[Y^2 \mid X]$ and $\mathrm{var}[Y \mid X]$ for $X = 0$ and $X = 1$.

$X=0 \rightarrow E[Y/X=X_0] = \sum Y \cdot P_{Y/X=X_0}(Y/X=X_0) = \sum Y \frac{P_Y(Y)}{P(X=X_0)}$

- $E[Y/X=0] = 1 \times \frac{0.4}{0.5} = 0.8$
- $E[Y^2/X=0] = 1^2 \times \frac{0.4}{0.5} = 0.8$
- $\mathrm{Var}[Y/X=0] = E[Y^2/X=0] - E[Y/X=0]^2$
  $= 0.8 - 0.8^2 = 0.16$

$X=1 \rightarrow E[Y/X=1] = \frac{0.3}{0.5} \times 1 = 0.6$

$\rightarrow E[Y^2/X=1] = \frac{0.3}{0.5} \times 1^2 = 0.6$

$\rightarrow \mathrm{Var}[Y/X=1] = E[Y^2/X=1] - E[Y/X=1]^2$
$= 0.6 - 0.6^2$
$= 0.24$

3. True or False?

(a) If $Y = X\beta + e$, $X \in \mathbb{R}$, and $\mathbb{E}[e \mid X] = 0$, then $\mathbb{E}[X^2 e] = 0$.

(b) If $Y = X\beta + e$, $X \in \mathbb{R}$, and $\mathbb{E}[Xe] = 0$, then $\mathbb{E}[X^2 e] = 0$.

(c) If $Y = X'\beta + e$ and $\mathbb{E}[e \mid X] = 0$, then e is independent of X.

(d) If $Y = X'\beta + e$ and $\mathbb{E}[Xe] = 0$, then $\mathbb{E}[e \mid X] = 0$.

(e) If $Y = X'\beta + e$, $\mathbb{E}[e \mid X] = 0$, and $\mathbb{E}[e^2 \mid X] = \sigma^2$, then e is independent of X.

a) $Y = XB + e$ $\quad E[x^2e] = E[E[x^2e/x]] = E[x^2 \underbrace{E[e/x]}_{=0}] = 0$ $\quad$ Verdadero

$E[e/X] = 0$

b) $Y = XB + e$ $\quad E[x^2e] \rightarrow$ relación no lineal $\quad$ Falso

$E[xe] = 0$

↳ ortogonal, no hay corr (lineal)

c) $Y = x'B + e$

$E[e/x] = 0$ $\quad$ mean independence $\Big\}$ no es independencia !

d) $Y = x'B + e$ $\quad E[e/x] = 0$

$E[xe] = 0$ $\quad$ Falso

e) Falso, tiene q' ser en todos los momentos

4. Take the homoskedastic model

$$\begin{aligned} Y &= X_1'\beta_1 + X_2'\beta_2 + e \\ \mathbb{E}[e \mid X_1, X_2] &= 0 \quad \text{mean independence} \\ \mathbb{E}[e^2 \mid X_1, X_2] &= \sigma^2 \\ \mathbb{E}[X_2 \mid X_1] &= \Gamma X_1 \rightarrow \text{Corr}(X_2, X_1) \neq 0 \end{aligned}$$

Assume $\Gamma \neq 0$. Suppose the parameter β_1 is of interest. We know that the exclusion of X_2 creates omited variable bias in the projection coefficient on X_2. It also changes the equation error. Our question is: what is the effect on the homoskedasticity property of the induced equation error? Does the exclusion of X_2 induce heteroskedasticity or not? Be specific.

$Y = X_1'B_1 + X_2'B_2 + e \rightarrow e = Y - X_1'B_1 - X_2'B_2$

$E[e/X_1, X_2] = 0$

$E[e^2/X_1, X_2] = \sigma^2$

$E[X_2/X_1] = \Gamma X_1$

Si modelo $\quad Y = X_1'B_1 + e^* \rightarrow e^* = Y - X_1'B_1 = e + X_2'B_2$

$var(e^*/X_1) = E[e^* e^{*\prime}/X_1] - E[e^*/X_1]E[e^{*\prime}/X_1]$ $\quad$ $var(x) = E[x^2] - E[x]^2$

$= E[(e + X_2'B_2)(e' + B_2'X_2)/X_1]$

$= E[ee' + eB_2'X_2 + X_2'B_2e' + X_2'B_2B_2'X_2/X_1]$

$= E[ee'/X_1] + \underbrace{E[eB_2'X_2/X_1]}_{=0} + \underbrace{E[X_2'B_2e/X_1]}_{=0} + E[X_2'B_2B_2'X_2/X_1]$

$var(e^*/X_1) = \sigma^2 + E[X_2'B_2B_2'X_2/X_1]$

Como $E[X_2/X_1] = \Gamma X_1$ $\quad$ Ya no hay homocedasticidad !

$\hat{\beta}_{(-i)} = (X'X - X_iX_i')^{-1}(X'Y - X_iY_i)$. × $(X'X)^{-1}(X'X - x_i x_i')$

$$\hat{\beta}_{(-i)} = \Big(\sum_{j\neq i} X_jX_j'\Big)^{-1}\Big(\sum_{j\neq i} X_jY_j\Big) = (X'X - X_iX_i')^{-1}(X'Y - X_iY_i) = \big(X_{(-i)}'X_{(-i)}\big)^{-1}X_{(-i)}'Y_{(-i)}.$$

$$\hat{B}_{(-i)} - (X'X)^{-1}x_ix_i'\hat{B}_{(-i)} = (X'X)^{-1}\cancel{(X'X - x_ix_i')(X'X - x_ix_i')^{-1}}(X'Y - x_iY_i)$$

$$\hat{B}_{(-i)} - (X'X)^{-1}x_ix_i'\hat{B}_{(-i)} = \hat{B} - (X'X)^{-1}x_iY_i$$

$$\hat{B}_{(-i)} = \hat{B} - (X'X)^{-1}x_iY_i + (X'X)^{-1}x_ix_i'\hat{B}_{(-i)}$$

$$= \hat{B} - (X'X)^{-1}x_i\big(Y_i - x_i'\hat{B}_{(-i)}\big)$$

$$\hat{B}_{(-i)} = \hat{B} - (X'X)^{-1}x_i\tilde{e}_i$$

specific.

5. Let $\hat{\beta}_n = (X_n'X_n)^{-1}X_n'Y_n$ denote the OLS estimate when Y_n is $n \times 1$ and X_n is $n \times k$. A new observation (Y_{n+1}, X_{n+1}) becomes available. Prove that the OLS estimate computed using this additional observation is

$$\hat{\beta}_{n+1} = \hat{\beta}_n + \frac{1}{1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1}}(X_n'X_n)^{-1}X_{n+1}(Y_{n+1} - X_{n+1}'\hat{\beta}_n).$$

$$\Big(\hat{B}_{n+1} = (X_n'X_n + X_{n+1}X_{n+1}')^{-1}(X_n'Y_n + X_{n+1}Y_{n+1})\Big) \times (X_n'X_n)^{-1}(X_n'X_n + X_{n+1}X_{n+1}')$$

$$\hat{B}_{n+1} + (X_n'X_n)^{-1}(X_{n+1}X_{n+1}')\hat{B}_{n+1} = (X_n'X_n)^{-1}\underbrace{(X_n'X_n + X_{n+1}X_{n+1}')(X_n'X_n + X_{n+1}X_{n+1}')^{-1}}_{I_n}(X_n'Y_n + X_{n+1}Y_{n+1})$$

$$\hat{B}_{n+1} + (X_n'X_n)^{-1}(X_{n+1}X_{n+1}')\hat{B}_{n+1} = (X_n'X_n)^{-1}(X_n'Y_n + X_{n+1}Y_{n+1})$$

$$\hat{B}_{n+1} = (X_n'X_n)^{-1}(X_n'Y_n + X_{n+1}Y_{n+1}) - (X_n'X_n)^{-1}(X_{n+1}X_{n+1}')\hat{B}_{n+1}$$

$$\hat{B}_{n+1} = \hat{B}_n + (X_n'X_n)^{-1}X_{n+1}Y_{n+1} - (X_n'X_n)^{-1}(X_{n+1}X_{n+1}')\hat{B}_{n+1}$$

$$\hat{B}_{n+1} = \hat{B}_n + (X_n'X_n)^{-1}X_{n+1}\underbrace{(Y_{n+1} - X_{n+1}'\hat{B}_{n+1})}_{\hat{e}_{n+1}}$$

$$\times X_{n+1}'\ \Big(\hat{B}_{n+1} = \hat{B}_n + (X_n'X_n)^{-1}X_{n+1}\hat{e}_{n+1}\Big)$$

$$\hat{Y}_{n+1} = \tilde{Y}_{n+1} + X_{n+1}'(X_n'X_n)^{-1}X_{n+1}\hat{e}_{n+1}$$

$$(Y_{n+1} - \tilde{Y}_{n+1}) = \underbrace{(Y_{n+1} - \hat{Y}_{n+1})}_{\hat{e}_{n+1}} + X_{n+1}'(X_n'X_n)^{-1}X_{n+1}\hat{e}_{n+1}$$

$$\tilde{e}_{n+1} = (1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1})\hat{e}_{n+1}$$

$$\frac{\tilde{e}_{n+1}}{(1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1})} = \hat{e}_{n+1}$$

$$\hat{e}_{n+1} = \frac{Y_{n+1} - X_{n+1}'\hat{B}_n}{(1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1})}$$

$$\hat{B}_{n+1} = \hat{B}_n + (X_n'X_n)^{-1}X_{n+1}\left(\frac{Y_{n+1} - X_{n+1}'\hat{B}_n}{(1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1})}\right)$$

$$\hat{B}_{n+1} = \hat{B}_n + \left(\frac{1}{1 + X_{n+1}'(X_n'X_n)^{-1}X_{n+1}}\right)(X_n'X_n)^{-1}X_{n+1}(Y_{n+1} - X_{n+1}'\hat{B}_n)$$

6. Prove that R^2 is the square of the sample correlation between $\boldsymbol{Y}$ and $\widehat{\boldsymbol{Y}}$.

$$\left(\rho_{Y,\hat{Y}} = \frac{Cov(Y,\hat{Y})}{\sqrt{var(Y)}\sqrt{var(\hat{Y})}} \right)^2$$

$$\rho^2_{Y,\hat{Y}} = \frac{Cov(\hat{Y}+\hat{e},\hat{Y})^2}{v(Y)\ v(\hat{Y})} = \frac{[Cov(\hat{Y},\hat{Y}) + \underbrace{Cov(\hat{Y},\hat{e})}_{=0}]^2}{v(Y)\ v(\hat{Y})} = \frac{[Var(\hat{Y})]^{\cancel{2}}}{v(Y)\ \cancel{v(\hat{Y})}} = \frac{v(\hat{Y})}{v(Y)} = \frac{\overbrace{\frac{1}{n}\sum(\hat{Y}_i-\bar{Y})^2}^{SCE}}{\underbrace{\frac{1}{n}\sum(Y_i-\bar{Y})^2}_{SCT}} = R^2$$

7. For the intercept-only model $Y_i = \beta + e_i$, show that the leave-one-out prediction error is

$$\tilde{e}_i = \left(\frac{n}{n-1}\right)(Y_i - \bar{Y}).$$

$$Y_i = B + e_i \longrightarrow \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} B + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$Y = X B + e$$

$$h_{ii} = X_i'(X'X)^{-1}X_i$$

$$= 1' \cdot \Big[\underbrace{\underset{(1\times n)}{(1 \cdots 1)} \underset{(n\times 1)}{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}}_{1+1+\cdots+1 \text{ n veces}} \Big]^{-1} \cdot 1$$

$$(1\ 1\ 1 \cdots 1)\ n^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{1}{n}$$

$h_{ii} = \frac{1}{n} = \bar{h}$ todos los elementos de la diagonal de P son $\frac{1}{n}$

recordar: $\tilde{e}_i = (1-h_{ii})^{-1}\hat{e}_i$ Fórmula

$$= (1-\bar{h})^{-1}(Y_i - \hat{B})$$

Entonces

$$SSE = \sum \overbrace{(Y_i - \hat{B})}^{\hat{e}_i}{}^2$$

CPO: $\frac{\partial SSE}{\partial B} : 2\sum(Y_i - \hat{B})(-1) = 0$

$$\sum(Y_i - \hat{B}) = 0$$

$$\sum Y_i - \sum \hat{B} = 0$$

$= n\left(\frac{1}{n}\right)\sum Y_i$

$= n\bar{Y}$

$$\left. n\bar{Y} - n\hat{B} = 0 \right\} \bar{Y} = \hat{B}$$

Luego

$\hat{e}_i = (Y_i - \underset{\bar{Y}}{\hat{B}})$

$$\tilde{e}_i = (1-\bar{h})^{-1}\hat{e}_i$$

$$= \left(1-\frac{1}{n}\right)^{-1}(Y_i - \bar{Y})$$

$$\tilde{e}_i = \left(\frac{n}{n-1}\right)(Y_i - \bar{Y}) \quad //$$

8. The observations are (Y_i, X_{1i}, X_{2i}), $i = 1, \dots, n$. You estimate two least squares regressions.

$$Y_i = X_{1i}'\tilde{\beta}_1 + \tilde{e}_i$$
$$Y_i = X_{1i}'\hat{\beta}_1 + X_{2i}'\hat{\beta}_2 + \hat{e}_i' \rightarrow \tilde{e}_i = X_{2i}'\hat{\beta}_2 + \hat{e}_i$$

$$(\quad)_{1\times k}\ (\quad)_{k\times 1}$$

and calculate the residual variance estimates

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{e}_i^2$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2$$

Show that for any $w \in (0,1)$, there is a constant $a \in (0,1)$ such that

$$\frac{1}{n}\sum_{i=1}^{n}(w\hat{e}_i + (1-w)\tilde{e}_i)^2 = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2.$$

Find the constant a.

$$\frac{1}{n}\sum_{i=1}^{n}\left(w^2\hat{e}_i^2 + 2w(1-w)\hat{e}_i\tilde{e}_i + (1-w)^2\tilde{e}_i^2\right) = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$\frac{1}{n}\sum w^2\hat{e}_i^2 + \left(\frac{1}{n}\right)2\sum w(1-w)\hat{e}_i\tilde{e}_i + \left(\frac{1}{n}\right)\sum(1-w)^2\tilde{e}_i^2 = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$w^2\underbrace{\left(\frac{1}{n}\right)\sum\hat{e}_i^2}_{\hat{\sigma}^2} + 2w(1-w)\left(\frac{1}{n}\right)\sum\hat{e}_i\tilde{e}_i + (1-w)^2\underbrace{\left(\frac{1}{n}\right)\sum\tilde{e}_i^2}_{\tilde{\sigma}^2} = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$w^2\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 + 2w(1-w)\underline{\left(\frac{1}{n}\right)\sum\hat{e}_i\tilde{e}_i} = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

Como $\tilde{e}_i = X_{2i}'\hat{\beta}_2 + \hat{e}_i$

$$\sum\hat{e}_i\tilde{e}_i = \sum\hat{e}_i(X_{2i}'\hat{\beta}_2 + \hat{e}_i)$$
$$= \sum\hat{e}_i X_{2i}'\hat{\beta}_2 + \sum\hat{e}_i\hat{e}_i$$
$$= \hat{\beta}_2\sum\hat{e}_i X_{2i}' + \sum\hat{e}_i^2$$

$$w^2\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 + 2w(1-w)\left(\frac{1}{n}\right)\left(\hat{\beta}_2\sum\hat{e}_i X_{2i}' + \sum\hat{e}_i^2\right) = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$w^2\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 + 2w(1-w)\left(\frac{1}{n}\right)\sum\hat{e}_i^2 + 2w(1-w)\left(\frac{1}{n}\right)\hat{\beta}_2\sum\hat{e}_i X_{2i}' = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$w^2\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 + 2w(1-w)\hat{\sigma}^2 + 2w(1-w)\left(\frac{1}{n}\right)\hat{\beta}_2\sum\hat{e}_i X_{2i}' = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$[w^2 + 2w(1-w)]\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 + 2w(1-w)\left(\frac{1}{n}\right)\hat{\beta}_2\underbrace{\sum\hat{e}_i X_{2i}'}_{=0} = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$[w^2 + 2w(1-w)]\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$! Preguntar.

$$w[w + 2(1-w)]\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 =$$
$$w(2-w)\hat{\sigma}^2 + (1-w)^2\tilde{\sigma}^2 =$$
$$(2w - w^2)\hat{\sigma}^2 + \underbrace{(1-w)^2}_{a}\tilde{\sigma}^2 = (1-a)\hat{\sigma}^2 + a\tilde{\sigma}^2$$

$$1-a = 1-(1-w)^2 = 1-(1^2-2w+w^2) = 1-1+2w-w^2 = 2w-w^2 \checkmark$$

entonces $a = 1-w$

Find the constant a.

9. Consider a regression of a $n \times 1$ vector of responses $\mathbf{Y}$ on a $n \times k$ matrix of explanatory variables $\mathbf{X}$. Let $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$, where $\mathbf{P}$ is the projection matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{I}_n$ is the identity matrix of order n. Suppose one adds a new explanatory variable, $\mathbf{Z}$, to the regression, so there are now $(k+1)$ regressors. Show that the new residual sum of squares is given by $\mathbf{Y}'\mathbf{M}\mathbf{Y} - b^2\mathbf{Z}'\mathbf{M}\mathbf{Z}$ where $b = \mathbf{Z}'\mathbf{M}\mathbf{Y}/\mathbf{Z}'\mathbf{M}\mathbf{Z}$ is the OLS coefficient on $\mathbf{Z}$.

proyección	residuos
$PX = X$	$MX = 0$
$PY = \hat{Y}$	$MY = \hat{e}$

$$SSE_1 = \hat{e}'\hat{e} = (MY)'(MY) = Y'MY$$

nuevo regresor Z:
- $PZ \rightarrow$ proyección de Z sobre el espacio generado por X
- $MZ \rightarrow$ parte de "Z" que no está relacionada con X } aporta información nueva para explicar Y

$$(bMZ)'(bMZ)' = b^2Z'MZ$$

$\tilde{e}_* = My - bMz$

$$SSE_z = (My - bMz)'(My - bMz)$$

$$SSE_z = y'My - 2bz'My + b^2 z'Mz \rightarrow y'My - 2\frac{z'My}{z'Mz} z'My + \left(\frac{z'My}{z'Mz}\right)^2 z'Mz$$

$$= y'My - 2\frac{(z'My)(z'My)}{z'Mz} + \frac{(z'My)(z'My)}{(z'Mz)\cancel{(z'Mz)}} \cancel{(z'Mz)}$$

$$= y'My - \frac{(z'My)(z'My)}{z'Mz} \quad ; \quad b^2 = \frac{(z'My)(z'My)}{(z'Mz)(z'Mz)}$$

$$b^2 (z'Mz) = \frac{(z'My)(z'My)}{z'Mz}$$

$$SSE_z = y'My - b^2(z'Mz)$$

$$\frac{\partial SSE_z}{\partial b} = -\cancel{2} z'My + \cancel{2}bz'Mz = 0$$

$$b = \frac{z'My}{z'Mz}$$

Regression (Cont.)

1. For some integer k, set $\mu_k = \mathbb{E}\left[Y^k\right]$.

   (a) Construct an estimator $\hat{\mu}_k$ for μ_k.

   (b) Show that $\hat{\mu}_k$ is unbiased for μ_k.

   (c) Calculate the variance of $\hat{\mu}_k$, say var$[\hat{\mu}_k]$. What assumption is needed for var$[\hat{\mu}_k]$ to be finite?

   (d) Propose an estimator of var$[\hat{\mu}_k]$.

2. Let $\mu = \mathbb{E}[Y], \sigma^2 = \mathbb{E}\left[(Y-\mu)^2\right]$ and $\mu_3 = \mathbb{E}\left[(Y-\mu)^3\right]$ and consider the sample mean $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Find $\mathbb{E}\left[(\bar{Y}-\mu)^3\right]$ as a function of μ, σ^2, μ_3 and n.
3. Now assume that

   $$\sqrt{n}\begin{pmatrix}\hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2\end{pmatrix} \underset{d}{\longrightarrow} \mathrm{N}(0, \Sigma)$$

   where $\| \Sigma \| < \infty$, then find the asymptotic distribution of the following statistics

   (a) $\hat{\theta}_1\hat{\theta}_2$
   (b) $\exp(\hat{\theta}_1 + \hat{\theta}_2)$
   (c) If $\theta_2 \neq 0$, $\hat{\theta}_1/\hat{\theta}_2^2$
   (d) $\hat{\theta}_1^3 + \hat{\theta}_1\hat{\theta}_2^2$
4. Take a regression model $Y = X\beta + e$ with $\mathbb{E}[e \mid X] = 0$ and i.i.d. observations (Y_i, X_i) and scalar X. The parameter of interest is $\theta = \beta^2$. Consider the OLS estimators $\hat{\beta}$ and $\hat{\theta} = \hat{\beta}^2$.

   (a) Find $\mathbb{E}[\hat{\theta} \mid \boldsymbol{X}]$ using our knowledge of $\mathbb{E}[\hat{\beta} \mid \boldsymbol{X}]$ and $V_{\hat{\beta}} = \text{var}[\hat{\beta} \mid \boldsymbol{X}]$. Is $\hat{\theta}$ biased for θ ?

   (b) Suggest an (approximate) biased-corrected estimator $\hat{\theta}^*$ using an estimator $\hat{V}_{\hat{\beta}}$ for $V_{\hat{\beta}}$.

   (c) For $\hat{\theta}^*$ to be potentially unbiased, which estimator of $V_{\hat{\beta}}$ is most appropriate?

   Under which conditions is $\hat{\theta}^*$ unbiased?
5. Consider an i.i.d. sample $\{Y_i, X_i\} i = 1, \ldots, n$ where X_i is $k \times 1$. Assume the linear conditional expectation model $Y = X'\beta + e$ with $\mathbb{E}[e \mid X] = 0$. Assume that $n^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}_k$ (orthonormal regressors). Consider the OLS estimator $\hat{\beta}$.

   (a) Find $\boldsymbol{V}_{\hat{\beta}} = \text{var}[\hat{\beta}]$

   (b) In general, are $\hat{\beta}_j$ and $\hat{\beta}_\ell$ for $j \neq \ell$ correlated or uncorrelated?

   (c) Find a sufficient condition so that $\hat{\beta}_j$ and $\hat{\beta}_\ell$ for $j \neq \ell$ are uncorrelated.
6. Take the model in vector notation

   $$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\beta + \boldsymbol{e} \\ \mathbb{E}[\boldsymbol{e} \mid \boldsymbol{X}] &= 0 \\ \mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right] &= \Omega \end{aligned}.$$

   Assume for simplicity that Ω is known. Consider the OLS and GLS estimators $\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right)$ and $\tilde{\beta} = \left(\boldsymbol{X}'\Omega^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\Omega^{-1}\boldsymbol{Y}\right)$. Compute the (conditional) covariance between $\hat{\beta}$ and $\tilde{\beta}$:

   $$\mathbb{E}\left[(\hat{\beta}-\beta)(\tilde{\beta}-\beta)' \mid \boldsymbol{X}\right].$$

   Find the (conditional) covariance matrix for $\hat{\beta} - \tilde{\beta}$:

   $$\mathbb{E}\left[(\hat{\beta}-\tilde{\beta})(\hat{\beta}-\beta)' \mid \boldsymbol{X}\right].$$
7. The model is $Y = X\beta + e$ with $\mathbb{E}[e \mid X] = 0$ and $X \in \mathbb{R}$. Consider the two estimators

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

$$\tilde{\beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{X_i}$$

(a) Under the stated assumptions are both estimators consistent for β ?

(b) Are there conditions under which either estimator is efficient?

8. Take the linear model $Y = X\beta + e$ with $\mathbb{E}[e \mid X] = 0$ and $X_i \in \mathbb{R}$. Consider the estimator

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i^3 Y_i}{\sum_{i=1}^{n} X_i^4}.$$

Find the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$.

9. The model is $y = X'\beta + e$ with $\mathbb{E}[e \mid X] = 0$. An econometrician is worried about the impact of some unusually large values of the regressors. The model is thus estimated on the subsample for which $|X_i| \leq c$ for some fixed c. Let $\tilde{\beta}$ denote the OLS estimator on this subsample. It equals

$$\tilde{\beta} = \left(\sum_{i=1}^{n} X_i X_i' \mathbb{1}\{|X_i| \leq c\}\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i \mathbb{1}\{|X_i| \leq c\}\right).$$

(a) Show that $\tilde{\beta} \xrightarrow[p]{} \beta$.

(b) Find the asymptotic distribution of $\sqrt{n}(\tilde{\beta} - \beta)$.

Binary Response

1. For a binary response y, let $\bar{y}$ be the proportion of ones in the sample (which is equal to the sample average of the y_i). Let $\hat{q}_0$ be the percent correctly predicted for the outcome $y = 0$ and let $\hat{q}_1$ be the percent correctly predicted for the outcome $y = 1$. If $\hat{p}$ is the overall percent correctly predicted, show that $\hat{p}$ is a weighted average of $\hat{q}_0$ and $\hat{q}_1$:

$$\hat{p} = (1 - \bar{y})\hat{q}_0 + \bar{y}\hat{q}_1.$$

   In a sample of 300 , suppose that $\bar{y} = .70$, so that there are 210 outcomes with $y_i = 1$ and 90 with $y_i = 0$. Suppose that the percent correctly predicted when $y = 0$ is 80 , and the percent correctly predicted when $y = 1$ is 40 . Find the overall percent correctly predicted.

2. Emily estimates a probit regression setting her dependent variable to equal *Y = 1* for a purchase and *Y = 0* for no purchase. Using the same data and regressors, Jacob estimates a probit regression setting the dependent variable to equal *Y = 1* if there is no purchase and *Y = 0* for a purchase. What is the difference in their estimated slope coefficients?
3. Jackson estimates a logit regression where the primary regressor is measured in dollars. Julie estimates a logit regression with the same sample and dependent variable, but measures the primary regressor in thousands of dollars. What is the difference in the estimated slope coefficients?
4. For the logistic distribution $\Lambda(x) = (1 + \exp(-x))^{-1}$ verify that
   (a) $\frac{d}{dx}\Lambda(x) = \Lambda(x)(1 - \Lambda(x))$.
   (b) $h_{\text{logit}}(x) = \frac{d}{dx}\log\Lambda(x) = 1 - \Lambda(x)$.
   (c) $H_{\text{logit}}(x) = -\frac{d^2}{dx^2}\log\Lambda(x) = \Lambda(x)(1 - \Lambda(x))$.
   (d) $\left|H_{\text{logit}}(x)\right| \leq 1$.
5. For the normal distribution $\Phi(x)$ verify that
   (a) $h_{\text{probit}}(x) = \frac{d}{dx}\log\Phi(x) = \lambda(x)$ where $\lambda(x) = \phi(x)/\Phi(x)$.
   (b) $H_{\text{probit}}(x) = -\frac{d^2}{dx^2}\log\Phi(x) = \lambda(x)(x + \lambda(x))$.

Multinomial Response

1.

a. For estimating the mean of a nonnegative random variable y using a random sample $\{y_i\}_{i=1}^n$, the Poisson quasi-log likelihood for a random draw is $\gamma_i(\mu) = y_i \log(\mu) - \mu, \mu > 0$ (where terms not depending on μ have been dropped). Letting $\mu_0 \equiv \mathbb{E}(y_i)$, we have $\mathbb{E}[\ell_i(\mu)] = \mu_0 \log(\mu) - \mu$. Show that this function is uniquely maximized at $\mu = \mu_o$.

b. The gamma (exponential) quasi-log likelihood is $\ell_i(\mu) = -y_i/\mu - \log(\mu), \mu > 0$. Show that $\mathbb{E}[\ell_i(\mu)]$ is uniquely maximized at $\mu = \mu_o$.

2. In a balanced panel data setting, i.e. $\{\{y_{it}, \mathbf{x}_{it}\}_{t=1}^T\}_{i=1}^n$, consider an unobserved effects model for count data with exponential regression function

$$\mathbb{E}(y_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = c_i \exp(\mathbf{x}_{ii}\boldsymbol{\beta}).$$

If $\mathbb{E}(c_i \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = \exp(\alpha + \bar{\mathbf{x}}_i\gamma)$, find $\mathbb{E}(y_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ where $\bar{\mathbf{x}}_i = T^{-1}(\mathbf{x}_{i1} + \cdots + \mathbf{x}_{iT})$.

3. Let patents be the number of patents applied for by a firm during a given year. Assume that the conditional expectation of patents given sales and RD is

$$\mathbb{E}(\ patents \mid sales\ \ , RD) = \exp[\beta_0 + \beta_1 \log(\ sales\ ) + \beta_2 RD + \beta_3 RD^2],$$

where $sales$ is annual firm sales and RD is total spending on research and development over the past 10 years.

a. How would you estimate the β_j ? Justify your answer by discussing the nature of patents.

b. How do you interpret β_1 ?

c. Find the partial effect of RD on $\mathbb{E}(\ patents \ \mid \ sales, RD)$.

High-Dimensional Models

1. Take the model $Y = X'\beta + e$ with $\mathbb{E}[Xe] = 0$. Define the ridge regression estimator

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i' + \lambda \boldsymbol{I}_k\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right),$$

   here $\lambda > 0$ is a fixed constant.
   (a) Find $\mathbb{E}[\hat{\beta} \mid \boldsymbol{X}]$.
   (b) Is $\hat{\beta}$ biased for β ?
   (c) Find the probability limit of $\hat{\beta}$ as $n \to \infty$.
   (d) Is $\hat{\beta}$ consistent for β ?
   (e) Derive $\operatorname{var}\left[\hat{\beta} \mid \boldsymbol{X}\right]$.
2. Show that the ridge regression estimator can be computed as least squares applied to an augmented data set. Take the original data $(\boldsymbol{Y}, \boldsymbol{X})$. Add p 0's to $\boldsymbol{Y}$ and p rows of $\sqrt{\lambda}\boldsymbol{I}_p$ to $\boldsymbol{X}$, apply least squares, and show that this equals $\hat{\beta}_{\text{ridge}}$.
3. Which estimator produces a higher regression R^2, least squares or ridge regression?