Assignment 1

Rin Meng
Student ID: 51940633

October 1, 2024

1. (a) The linear regression model can be written as,

$$P = \beta_0 + \beta_1 C + \epsilon$$

where,

- P is the prime interest rate
- C is the core inflation rate
- β_0 is the intercept of the regression line
- β_1 is the slope of the regression line
- ϵ is the error term

Assumptions of the linear regression model are,

i. **Linearity:** The relationship between P and C is linear
ii. **Independence:** The error term ϵ is independent of C
iii. **Normality:** The error term ϵ is normally distributed
iv. **Homoscedasticity:** The error term ϵ has a constant variance across all levels of C

2. **Remark:** from ch2A.pdf slide 9 and 13,

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
- $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$

Let e_i be the ith residual term. Then for each observation, we can see that the residual for each observation i is defined as:

$$e_i = y_i - \hat{y_i}$$

then we can say that the predicted value $\hat{y_i}$ is equivalent to

$$\hat{y_i} = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

(a) $\sum_{i=1}^{n} e_i = 0$

Proof. LHS:

$$\begin{aligned}
\sum_{i=1}^{n} e_i &= \sum_{i=1}^{n}(y_i - \hat{y_i}) \\
&= \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= \sum_{i=1}^{n}(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) \\
&= \sum_{i=1}^{n} y_i - n\bar{y} + \hat{\beta}_1 n\bar{x} - \hat{\beta}_1 \sum_{i=1}^{n} x_i \\
&= n\bar{y} - n\bar{y} + \hat{\beta}_1 n\bar{x} - \hat{\beta}_1 n\bar{x} \\
&= 0
\end{aligned}$$

$$LHS = 0 = RHS$$

$\therefore$ It is true that the sum of residuals e_i is zero. $\square$

(b) $\sum_{i=1}^{n} x_i e_i = 0$

Proof. LHS:

$$\begin{aligned}
\sum_{i=1}^{n} x_i e_i &= \sum_{i=1}^{n} x_i(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= \sum_{i=1}^{n}(x_i y_i - x_i(\bar{y} - \hat{\beta}_1\bar{x}) - \hat{\beta}_1 x_i^2) \\
&= \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} + x_i\hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i^2) \\
&= \sum_{i=1}^{n}(x_i(y_i - \bar{y}) + \hat{\beta}_1(\bar{x}x_i - x_i^2)) \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right) \\
&= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) - \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \\
&= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}\right) - \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}\bar{x} - n\bar{x}\bar{x} + n\bar{x}\bar{x}\right) \\
&= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}\right) - \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}\bar{x} + n\bar{x}\bar{x}\right) \\
&= \left(\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y}\right) - \hat{\beta}_1\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2\right) \\
&= \sum_{i=1}^{n}\left(x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\bar{y}\right) - \hat{\beta}_1\sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\
&= \sum_{i=1}^{n}\left(x_i(y_i - \bar{y}) + \bar{x}(y_i - \bar{y})\right) - \hat{\beta}_1\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \sum_{i=1}^{n}(x_i + \bar{x})(y_i - \bar{y}) - \hat{\beta}_1\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= S_{xy} - \hat{\beta}_1 S_{xx} \\
&= S_{xy} - \frac{S_{xy}}{S_{xx}} S_{xx} \\
&= S_{xy} - S_{xy} \\
&= 0
\end{aligned}$$

$$LHS = 0 = RHS$$

□

$\therefore$ It is true that the independent variables x_i is completly uncorrelated to the residuals e_i.

(c) $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

Proof. LHS:

$$\begin{aligned}
\sum_{i=1}^{n} \hat{y}_i e_i &= \sum_{i=1}^{n} \hat{y}_i (y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n} (y_i \hat{y}_i - \hat{y}_i{}^2) \\
&= \sum_{i=1}^{n} (y_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i)^2) \\
&= \sum_{i=1}^{n} (y_i(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i) - (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i)^2) \\
&= \sum_{i=1}^{n} (y_i \bar{y} - y_i \hat{\beta}_1 \bar{x} + y_i \hat{\beta}_1 x_i \\
&- (\bar{y}^2 - 2\hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1{}^2 \bar{x}^2 + 2\hat{\beta}_1 \bar{y} x_i - 2\hat{\beta}_1{}^2 \bar{x} x_i + \hat{\beta}_1{}^2 x_i^2)) \\
&= (\sum_{i=1}^{n} y_i \bar{y} - \sum_{i=1}^{n} y_i \hat{\beta}_1 \bar{x} + \sum_{i=1}^{n} y_i \hat{\beta}_1 x_i \\
&- \sum_{i=1}^{n} (\bar{y}^2 - 2\hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1{}^2 \bar{x}^2 + 2\hat{\beta}_1 \bar{y} x_i - 2\hat{\beta}_1{}^2 \bar{x} x_i + \hat{\beta}_1{}^2 x_i^2)) \\
&= (n\bar{y}^2 - \hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1 \bar{y} \hat{\beta}_1 x_i \\
&- (\sum_{i=1}^{n} \bar{y}^2 - \sum_{i=1}^{n} 2\hat{\beta}_1 \bar{y}\bar{x} + \sum_{i=1}^{n} \hat{\beta}_1{}^2 \bar{x}^2 + \sum_{i=1}^{n} 2\hat{\beta}_1 \bar{y} x_i - \sum_{i=1}^{n} 2\hat{\beta}_1{}^2 \bar{x} x_i + \sum_{i=1}^{n} \hat{\beta}_1{}^2 x_i^2)) \\
&= (n\bar{y}^2 - (n\bar{y}^2 - 2\hat{\beta}_1 n\bar{y}\bar{x} + \hat{\beta}_1{}^2 n\bar{x}^2 + 2\hat{\beta}_1 n\bar{y}\bar{x} - 2\hat{\beta}_1{}^2 n\bar{x}\bar{x} + \hat{\beta}_1{}^2 \sum_{i=1}^{n} x_i^2))
\end{aligned}$$

proof continued.

$$\sum_{i=1}^{n} \hat{y}_i e_i = (n\bar{y}^2 - (n\bar{y}^2 - \hat{\beta_1}^2 n\bar{x}^2 + \hat{\beta_1}^2 n\bar{x}^2))$$
$$= (n\bar{y}^2 - n\bar{y}^2)$$
$$= 0$$

$$LHS = 0 = RHS$$

□

$\therefore$ It is true that the predicted values $\hat{y}_i$ is completely orthogonal to the residuals e_i.

3. (a) It is given that:
   - $\hat{\beta}_1 = \frac{S_{xx}}{S_{xy}}$
   - $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
   - $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$
   - $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i$
   - $\bar{y} = \beta_0 + \beta_1 \bar{x}$

   We want to show that

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{S_{xx}})$$

   Where β_1 is the mean of the distribution and $\frac{\sigma^2}{S_{xx}}$ is the variance.

   Proof. Let $k_i = \frac{x_i - \bar{x}}{S_{xx}}$ then we have

$$\hat{\beta}_1 = \sum k_i(y_i - \bar{y})$$

$$E[\hat{\beta}_1] = E\left[\sum k_i(y_i - \bar{y})\right]$$
$$= \sum k_i E[(y_i - \bar{y})]$$
$$= \sum k_i E[(\hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i - \beta_0 - \beta_1 \bar{x})]$$
$$= \sum k_i (E[\hat{\beta}_0] + E[\hat{\beta}_1 x_i] + E[\epsilon_i] - E[\beta_0] - E[\beta_1 \bar{x}])$$

$$= \sum k_i(0 + \beta_1 x_i + 0 - 0 - \beta_1 \bar{x})$$
$$= \sum k_i(\beta_1 x_i - \beta_1 \bar{x})$$
$$= \beta_1 \sum k_i(x_i - \bar{x})$$
$$= \beta_1 \sum \frac{x_i - \bar{x}}{S_{xx}}(x_i - \bar{x})$$
$$= \beta_1 \sum \frac{(x_i - \bar{x})^2}{(x_i - \bar{x})^2}$$
$$= \beta_1$$

$\therefore E[\hat{\beta}_1] = \beta_1$ □

Using k_i from earlier, we can proof that $Var[\hat{\beta}_1] = \frac{\sigma^2}{S_{xx}}$

$$Var[\hat{\beta}_1] = Var[\sum k_i(y_i - \bar{y})]$$
$$Var[\hat{\beta}_1] = \sum Var[k_i(y_i - \bar{y})]$$
$$Var[\hat{\beta}_1] = \sum k_i^2 Var[(y_i - \bar{y})]$$

4. Given the model:
$$y_i = \beta_1 + \beta_2 i + \epsilon_i$$

   (a) Let us derive the least-squares estimators for β_1 and β_2.

$$S(\beta_1, \beta_2) = \sum (y_i - \beta_1 x_i - \beta_2 i)^2$$
$$\frac{\delta S}{\delta \beta_1} = -2 \sum (y_i - \beta_1 x_i - \beta_2 i) x_i = 0$$
$$= -2 \sum (y_i x_i - \beta_1 x_i^2 - \beta_2 i x_i) = 0$$
$$\sum y_i x_i = \beta_1 \sum x_i^2 + \beta_2 \sum x_i i$$
$$\frac{\delta S}{\delta \beta_2} = -2 \sum (y_i - \beta_1 x_i - \beta_2 i) i = 0$$
$$= -2 \sum (y_i i - \beta_1 x_i i - \beta_2 i^2) = 0$$
$$\sum y_i i = \beta_1 \sum x_i i + \beta_2 \sum i^2$$

Solving for β_2

$$\begin{aligned}
\sum y_i x_i &= \beta_1 \sum x_i^2 + \beta_2 \sum x_i i \\
\beta_1 \sum x_i^2 &= \sum y_i x_i - \beta_2 \sum x_i i \\
\beta_1 &= \frac{\sum y_i x_i - \beta_2 \sum x_i i}{\sum x_i^2} \\
\sum y_i i &= \beta_1 \sum x_i i + \beta_2 \sum i^2 \\
\beta_1 \sum x_i i &= \sum y_i i - \beta_2 \sum i^2 \\
\beta_1 &= \frac{\sum y_i i - \beta_2 \sum i^2}{\sum x_i i}
\end{aligned}$$

Now set them equal to each other and solve for β_2

$$\begin{aligned}
\frac{\sum y_i x_i - \beta_2 \sum x_i i}{\sum x_i^2} &= \frac{\sum y_i i - \beta_2 \sum i^2}{\sum x_i i} \\
(\sum y_i x_i - \beta_2 \sum x_i i)(\sum x_i i) &= (\sum x_i^2)(\sum y_i i - \beta_2 \sum i^2) \\
\sum y_i x_i (\sum x_i i) - \beta_2 (\sum x_i i)^2 &= \sum y_i i (\sum x_i^2) - \beta_2 \sum i^2 (\sum x_i^2) \\
\beta_2 \sum i^2 \sum x_i^2 - \beta_2 (\sum x_i i)^2 &= \sum y_i i \sum x_i^2 - \sum y_i x_i \sum x_i i \\
\beta_2 (\sum i^2 \sum x_i^2 - (\sum x_i i)^2) &= \sum y_i i \sum x_i^2 - \sum y_i x_i \sum x_i i
\end{aligned}$$

Finally, we have:

$$\beta_2 = \frac{\sum y_i i \sum x_i^2 - \sum y_i x_i \sum x_i i}{\sum i^2 \sum x_i^2 - (\sum x_i i)^2}$$

Solving for β_1 using same approach:

$$\begin{aligned}
\sum y_i x_i &= \beta_1 \sum x_i^2 + \beta_2 \sum x_i i \\
\beta_2 \sum x_i i &= \sum y_i x_i - \beta_1 \sum x_i^2 \\
\beta_2 &= \frac{\sum y_i x_i - \beta_1 \sum x_i^2}{\sum x_i i} \\
\sum y_i i &= \beta_1 \sum x_i i + \beta_2 \sum i^2 \\
\beta_2 \sum i^2 &= \sum y_i i - \beta_1 \sum x_i i \\
\beta_2 &= \frac{\sum y_i i - \beta_1 \sum x_i i}{\sum i^2}
\end{aligned}$$

Now set them equal to each other and solve for β_1

$$\begin{aligned}
\frac{\sum y_i x_i - \beta_1 \sum x_i^2}{\sum x_i i} &= \frac{\sum y_i i - \beta_1 \sum x_i i}{\sum i^2} \\
(\sum y_i x_i - \beta_1 \sum x_i^2)(\sum i^2) &= (\sum x_i i)(\sum y_i i - \beta_1 \sum x_i i) \\
\sum y_i x_i \sum i^2 - \beta_1 \sum x_i^2 \sum i^2 &= \sum y_i i \sum x_i i - \beta_1 (\sum x_i i)^2 \\
\beta_1 (\sum x_i i)^2 - \beta_1 \sum x_i^2 \sum i^2 &= \sum y_i i \sum x_i i - \sum y_i x_i \sum i^2 \\
\beta_1 ((\sum x_i i)^2 - \sum x_i^2 \sum i^2) &= \sum y_i i \sum x_i i - \sum y_i x_i \sum i^2
\end{aligned}$$

Finally, we have:

$$\beta_1 = \frac{\sum y_i i \sum x_i i - \sum y_i x_i \sum i^2}{(\sum x_i i)^2 - \sum x_i^2 \sum i^2}$$

To find the conditions where x_i makes the estimators not well-defined, we let $x_i = i$. so then we have our β_1,

$$\begin{aligned}
\beta_1 &= \frac{\sum y_i i \sum x_i i - \sum y_i x_i \sum i^2}{(\sum x_i)^2 - \sum x_i^2 \sum i^2} \\
&= \frac{\sum y_i i \sum i^2 - \sum y_i i \sum i^2}{(\sum i^2)^2 - \sum i^2 \sum i^2} \\
&= \frac{\sum y_i i \sum i^2 - \sum y_i i \sum i^2}{\sum i^2 \sum i^2 - \sum i^2 \sum i^2} \\
&= \frac{0}{0}
\end{aligned}$$

and then our β_2,

$$\begin{aligned}
\beta_2 &= \frac{\sum y_i i \sum x_i^2 - \sum y_i x_i \sum x_i i}{\sum i^2 \sum x_i^2 - (\sum x_i i)^2} \\
&= \frac{\sum y_i i \sum i^2 - \sum y_i i \sum i^2}{(\sum i^2)^2 - \sum i^2 \sum i^2} \\
&= \frac{\sum y_i i \sum i^2 - \sum y_i i \sum i^2}{\sum i^2 \sum i^2 - \sum i^2 \sum i^2} \\
&= \frac{0}{0}
\end{aligned}$$

$\therefore$ The estimator β_1 and β_2 is not well-defined at $x_i = i$.

(b) For the case where the coefficient estimators are well-defined, the unbiased estimator for σ^2 is:

$$E[\sum \epsilon^2] = (n-2)\sigma^2$$
$$\sigma^2 = \frac{\sum \epsilon^2}{n-2}$$

Derived from, Question 3b.

(c) *Proof.* We want to show that MSE is an unbiased estimator σ^2 then we want to show that:

$$E[MSE] = E\left[\frac{1}{n-1}\sum(y_i - \bar{y})^2\right]$$

so then we can start by recalling the model,

$$y_i = \alpha + \epsilon_i$$

then it is true that the sample mean is,

$$\bar{y} = \alpha + \bar{\epsilon}$$

then, $y_i - \bar{y}$ can be rewritten as,

$$y_i - \bar{y} = (\alpha + \epsilon_i) - (\alpha + \bar{\epsilon}) = \epsilon_i - \bar{\epsilon}$$

Now, we have

$$\begin{aligned} E[MSE] &= E\left[\frac{1}{n-1}\sum(\epsilon_i - \bar{\epsilon})^2\right] \\ &= \frac{1}{n-1}E\left[\left(\sum \epsilon_i^2 - \sum \bar{\epsilon}^2\right)\right] \\ &= \frac{1}{n-1}\left(\sum E[\epsilon_i^2] - E[n\bar{\epsilon}^2]\right) \end{aligned}$$

We know that

$$E[\epsilon_i] = 0 \Rightarrow E[\epsilon_i^2] = Var[\epsilon_i] + E[\epsilon_i]^2 = \sigma^2$$

$$\sum E[\epsilon_i] = n\sigma^2$$

now we find that,

$$
\begin{aligned}
E[n\bar{\epsilon}^2] = nE[\bar{\epsilon}^2] \\
&\Rightarrow E[\bar{\epsilon}^2] = Var[\bar{\epsilon}] + E[\bar{\epsilon}]^2 \\
&= \frac{\sigma^2}{n} \\
&\Leftrightarrow Var[\bar{\epsilon}] = Var[\frac{1}{n}\sum \epsilon_i] \\
&= \frac{1}{n^2}\sum Var[\epsilon_i] \\
&= \frac{1}{n^2}\sum \sigma^2 \\
&= \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n} \\
&\Leftrightarrow E[\bar{\epsilon}]^2 = E[\frac{1}{n}\sum e_i]^2 \\
&= (\frac{1}{n}\sum E[e_i])^2 \\
&= (\frac{1}{n}\cdot 0)^2 = 0 \\
E[n\bar{\epsilon}^2] = n\cdot\frac{\sigma^2}{n} = \sigma^2
\end{aligned}
$$

so then finally we have:

$$E[MSE] = \frac{1}{n-1}(n\sigma^2 - \sigma^2)$$

$$= \frac{1}{n-1}(n-1)\sigma^2$$
$$= \sigma^2$$

$\therefore MSE$ is an unbiased estimator for σ^2. $\square$