High-Dimensional Data Analysis HomeworkII

統研一 106225020 董奕賢 *Dong, Yi-Shian*

1. Let Y_i =Wage and x_i = Age be our data(Wage data), $i = 1, 2, \cdots, 3000$.

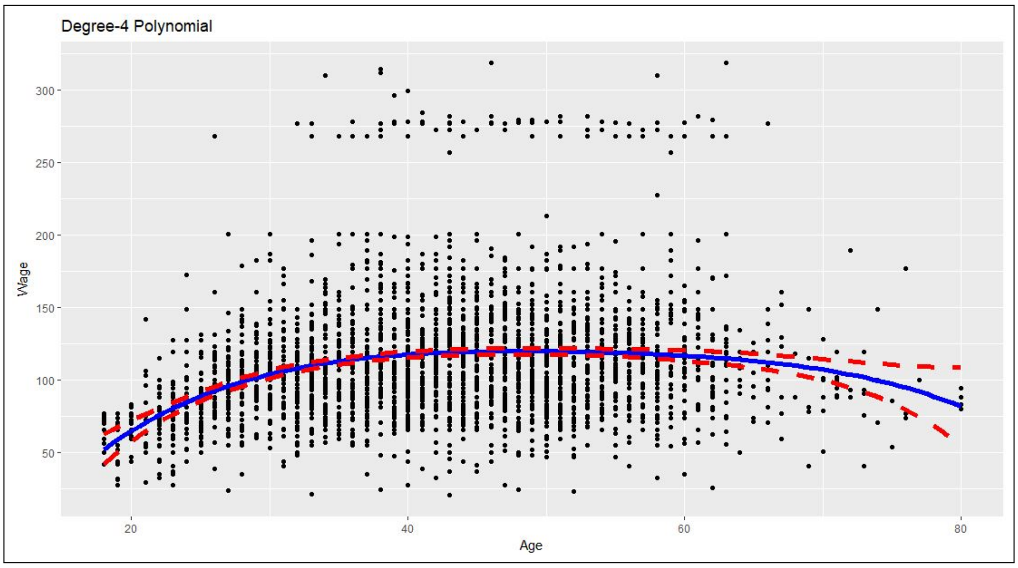
- Polynomial(following the textbook, here choose degree = 4.)
  If the true relation between wage and age is

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \epsilon_i,$$

where ϵ_i satisfy the Gaussian-Markov condition.
And the result of the following table are derived from R, with code given in the appendix.

	LSE	SE	95%C.I.		p-value
			lower	upper	
β_0	-184.1542	60.04038	-301.89	-66.43	0.0022
β_1	21.24552	5.886749	9.7030	32.788	0.0003
β_2	-0.563859	0.206108	-0.9680	-0.1597	0.0063
β_3	0.006811	0.003066	0.0008	0.01282	0.0264
β_4	-0.000032	0.000016	-6.4×10^{-5}	1.4×10^{-7}	0.0510

Figure 1: figure similar to **figure 7.1** in textbook



where the solid line are the predicted value of mean response, and the dotted line are the 95% C.I. of mean response.

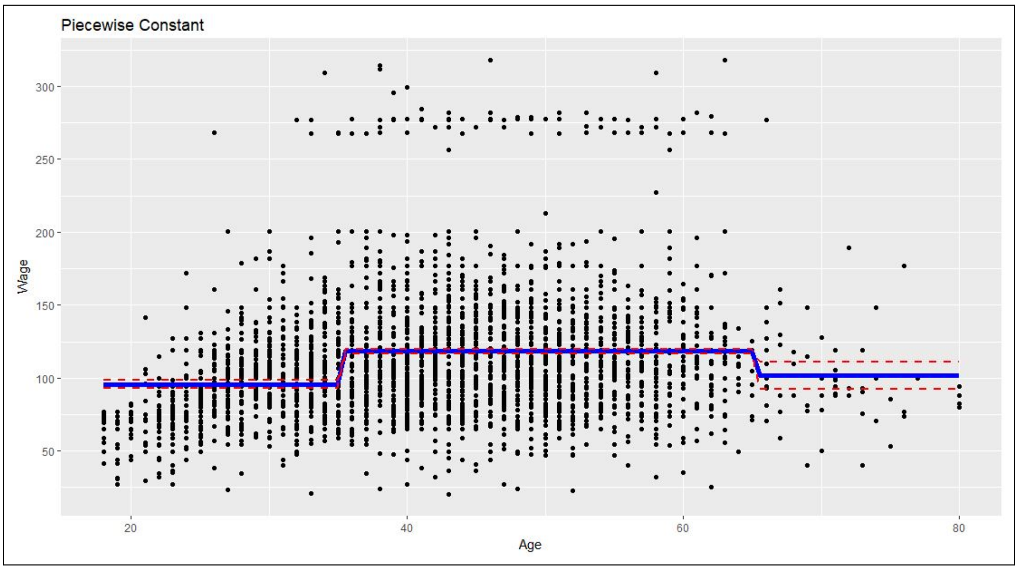
- Step function (Using the same setting as textbook)
  If the true relation between wage and age is

$$Y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \epsilon_i,$$

where $C_1(x) = I(35 \leq x < 65)$ and $C_2(x) = I(x \geq 65)$.
Here directly show the result derive from R.

	LSE	SE	95%C.I.		p-value
			lower	upper	
β_0	95.707	1.409	92.944	98.470	≈ 0
β_1	22.621	1.663	19.361	25.882	≈ 0
β_2	6.091	4.976	-3.665	15.849	0.2210

Figure 2: figure similar to **figure 7.2** in textbook



where the solid line are the predicted value of mean response, and the dotted line are the 95% C.I. of mean response.

- Spline(In order to use knots with equal space, choose knots = 18, 49, 80.; since it cover the range (minimum is 18 and maximum is 80) of Wage data.) If the true relation between wage and age is

$$Y_i = \beta_{\lambda 1} M_1(x_i) + \beta_{\lambda 2} M_2(x_i) + \cdots + \beta_{\lambda 5} M_5(x_i) + \epsilon_i,$$

$$\text{where} \begin{cases} M_1(x) = -\frac{4I(\xi_1 \le x < \xi_2)}{\Delta} z_2(x)^3 \\ M_2(x) = \frac{I(\xi_1 \le x < \xi_2)}{2\Delta} \{7z_1(x)^3 - 18z_1(x)^2 + 12z_1(x)\} - \frac{I(\xi_2 \le x < \xi_3)}{2\Delta} z_3(x)^3 \\ M_3(x) = \frac{I(\xi_1 \le x < \xi_2)}{\Delta} \{-2z_1(x)^3 + 3z_1(x)^2\} + \frac{I(\xi_2 \le x < \xi_3)}{\Delta} \{2z_2(x)^3 - 3z_2(x)^2 + 1\} \\ M_4(x) = \frac{I(\xi_1 \le x < \xi_2)}{2\Delta} z_1(x)^3 + \frac{I(\xi_2 \le x < \xi_3)}{2\Delta} \{-7z_2(x)^3 + 3z_2(x)^2 + 3z_2(x) + 1\} \\ M_5(x) = \frac{4I(\xi_2 \le x < \xi_3)}{\Delta} z_2(x)^3 \end{cases},$$

and $z_i(x) = \frac{x - \xi_i}{\Delta}, i = 1, 2, 3;\ \ \xi_1 = 18, \xi_2 = 49, \xi_3 = 80 (\Delta = 31)$, also given that

$$\mathbf{\Omega} = \frac{1}{31^5} \begin{bmatrix} 192 & -132 & 24 & 12 & 0 \\ -132 & 96 & -24 & -12 & 12 \\ 24 & -24 & 24 & -24 & 24 \\ 12 & -12 & -24 & 96 & 132 \\ 0 & 12 & 24 & -132 & 192 \end{bmatrix}.$$

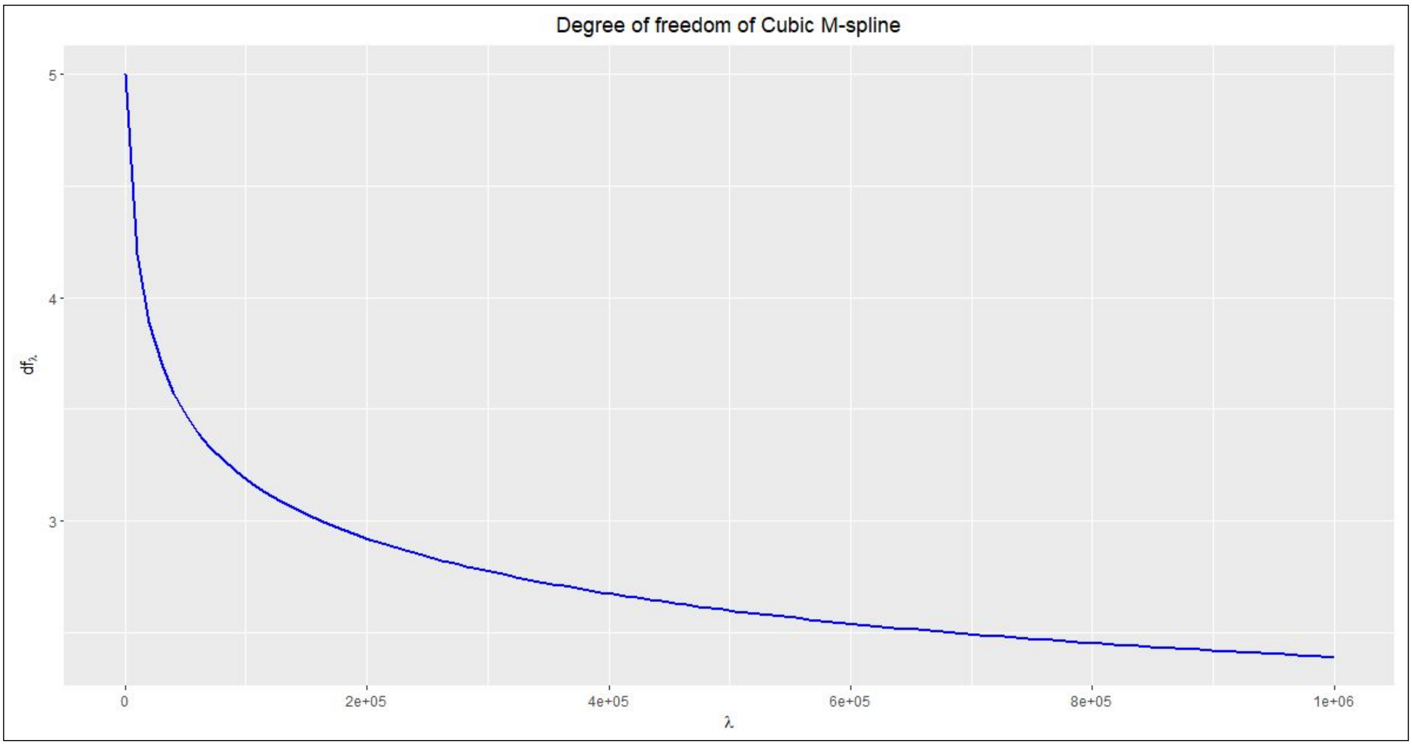
With the degree of freedom are

$$df_\lambda = tr\left[\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\mathbf{\Omega}\right)^{-1}\boldsymbol{X}^T\right],$$

$$\text{where } \boldsymbol{X} = \begin{bmatrix} M_1(x_1) & M_2(x_1) & M_3(x_1) & M_4(x_1) & M_5(x_1) \\ M_1(x_2) & M_2(x_2) & M_3(x_2) & M_4(x_2) & M_5(x_2) \\ & & \vdots & & \\ M_1(x_{3000}) & M_2(x_{3000}) & M_3(x_{3000}) & M_4(x_{3000}) & M_5(x_{3000}) \end{bmatrix}.$$

Here provide the plot of degree of freedom of this wage data, and we can observe that the degree of freedom is a monotone decreasing function, thus we can directly use the code **uniroot** in **R** to find the corresponding λ, and verify this solution is unique because of degree of freedom is a monotone function.

Figure 3: The plot of df_λ against λ of this wage data

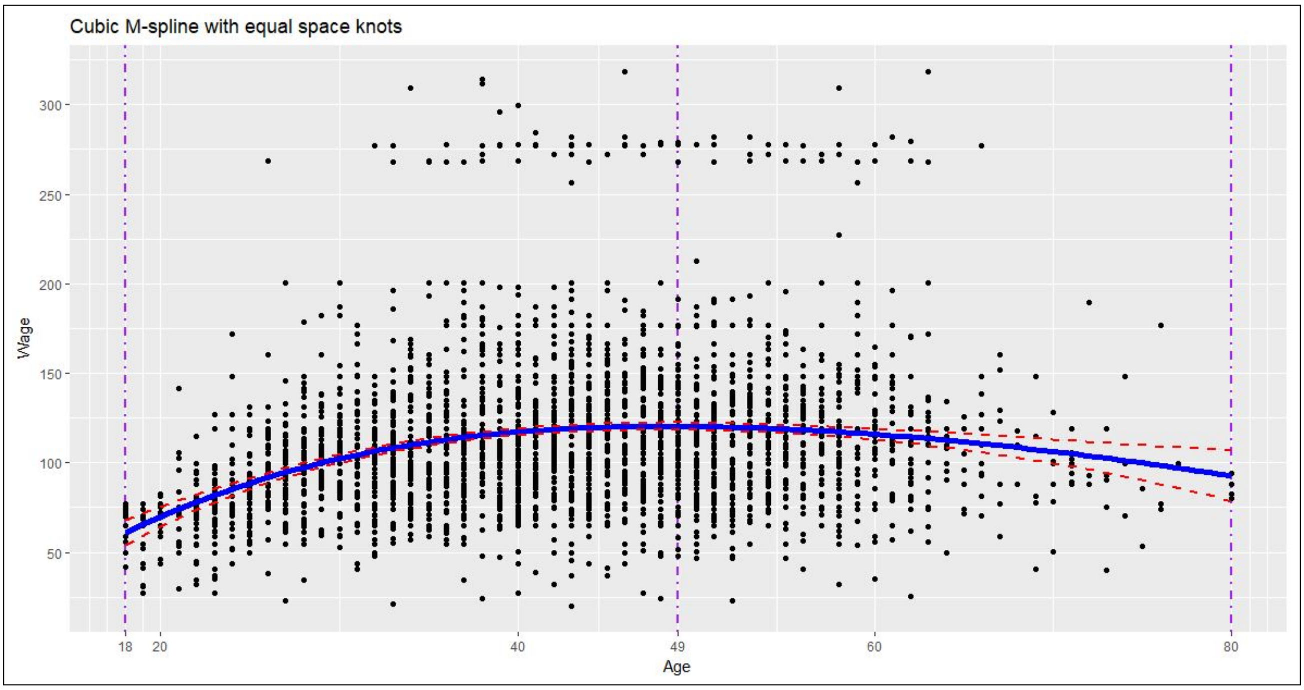



```
> ## about degree of freedom
> dg_lambda = function(lambda.par){
+   sum( diag( X_spline %*% solve( crossprod(X_spline) + lambda.par * Lambda) %*% t( X_spline )  ) )
+ }
> library(ggplot2)
> part_lambda = seq(0,1e6,length=100)
> ggplot() + geom_line( aes(x=part_lambda,y=sapply(part_lambda,dg_lambda) ) , size=1,colour = "blue" ) +
+   scale_x_continuous(name = expression(lambda) , breaks = seq(0,1e6,length=6),
+                      labels = c(0,200000,400000,600000,800000,1000000) )+
+   scale_y_continuous(name = expression(df[lambda]) ) +
+   ggtitle(label = "Degree of freedom of Cubic M-spline") +
+   theme(plot.title = element_text(hjust = 0.5)  )
> # determinant the value of lambda such that degree of freedom = 4
> uniroot( function(x) dg_lambda(x)-4,lower=0.1,upper=50000)$root
[1] 15727.51
```

And use the formula given in the class, we can drive the LSE and its SE of β_λ's, and **note that** here choose $\lambda = 15727.51$(numerical result given above), such that the degree of freedom equal to 4, which choosing the same as textbook.

	LSE	SE	95%C.I.		p-value
			lower	upper	
β_1	471.486	28.218	416.158	526.814	≈ 0
β_2	1704.009	44.051	1617.636	1790.383	≈ 0
β_3	2040.127	59.709	1923.053	2157.201	≈ 0
β_4	1675.927	59.520	1559.222	1792.632	≈ 0
β_5	719.573	56.388	609.009	830.136	≈ 0

Figure 4: figure similar to **figure 7.5** in textbook



where the solid line are the predicted value of mean response, the dotted line are the 95% C.I. of mean response, and the vertical dotted line are the (equal space) three knots.

2. Here using the same notations as in the course, and there are 3 knots $\xi_1 < \xi_2 < \xi_3$, where $\xi_3 - \xi_2 = \xi_2 - \xi_1 = \Delta$ (equal space). Define

$$\begin{cases} M_1(x) = -\frac{4I(\xi_1 \le x < \xi_2)}{\Delta} z_2(x)^3 \\ M_2(x) = \frac{I(\xi_1 \le x < \xi_2)}{2\Delta}\{7z_1(x)^3 - 18z_1(x)^2 + 12z_1(x)\} - \frac{I(\xi_2 \le x < \xi_3)}{2\Delta} z_3(x)^3 \\ M_3(x) = \frac{I(\xi_1 \le x < \xi_2)}{\Delta}\{-2z_1(x)^3 + 3z_1(x)^2\} + \frac{I(\xi_2 \le x < \xi_3)}{\Delta}\{2z_2(x)^3 - 3z_2(x)^2 + 1\} \\ M_4(x) = \frac{I(\xi_1 \le x < \xi_2)}{2\Delta} z_1(x)^3 + \frac{I(\xi_2 \le x < \xi_3)}{2\Delta}\{-7z_2(x)^3 + 3z_2(x)^2 + 3z_2(x) + 1\} \\ M_5(x) = \frac{4I(\xi_2 \le x < \xi_3)}{\Delta} z_2(x)^3 \end{cases},$$

where $z_i(x) = \frac{x - \xi_i}{\Delta}, i = 1, 2, 3$.

Define $\Omega_{ij} = \int_{\xi_1}^{\xi_3} M_i''(x) M_j''(x) dx$, the (i, j)-entry of $\mathbf{\Omega}$.

First, we can calculate the second derivative of $M_i(x)$'s as following:

$$\begin{cases} M_1''(x) = -24\frac{I(\xi_1 \le x < \xi_2)}{\Delta^3} z_2(x) = -24\frac{I(\xi_1 \le x < \xi_2)}{\Delta^3}\left(\frac{x - \xi_2}{\Delta}\right) \\ M_2''(x) = \frac{I(\xi_1 \le x < \xi_2)}{\Delta^3}\{21z_1(x) - 18\} - \frac{3I(\xi_2 \le x < \xi_3)}{\Delta^3} z_3(x) \\ M_3''(x) = \frac{I(\xi_1 \le x < \xi_2)}{\Delta^3}\{-12z_1(x) + 6\} + \frac{I(\xi_2 \le x < \xi_3)}{\Delta^3}\{12z_2(x) - 6\} \\ M_4''(x) = \frac{3I(\xi_1 \le x < \xi_2)}{\Delta^3} z_1(x) + \frac{I(\xi_2 \le x < \xi_3)}{\Delta^3}\{-21z_2(x) + 3\} \\ M_5''(x) = 24\frac{I(\xi_2 \le x < \xi_3)}{\Delta^3} z_2(x) \end{cases},$$

And note that

$$\begin{cases} dz_1 = \frac{d}{dx} z_1(x) dx = \frac{d}{dx}\frac{x - \xi_1}{\Delta} dx = \frac{dx}{\Delta} \\ z_1(x) = \frac{x - \xi_1}{\Delta} = \frac{x - \xi_2}{\Delta} + \frac{\xi_2 - \xi_1}{\Delta} = z_2(x) + 1 \Rightarrow dz_1 = dz_2 \\ z_2(x) = z_3(x) + 1 \Rightarrow dz_2 = dz_3 \end{cases}$$

then,

- $$\Omega_{11} = \int_{\xi_1}^{\xi_3} M_1''(x) M_1''(x) dx = \int_{\xi_1}^{\xi_3} \left(-24 \frac{I(\xi_1 \leq x < \xi_2)}{\Delta^3} z_2(x) \right)^2 dx$$
$$= \frac{24^2}{\Delta^6} \int_{\xi_1}^{\xi_2} z_2(x)^2 dx = \frac{24^2}{\Delta^6} \int_{-1}^{0} z_2^2 \Delta dz_2 = \frac{24^2}{\Delta^5} \left. \frac{z_2^3}{3} \right|_{z_2=-1}^{z_2=0} = \frac{192}{\Delta^5}$$
- $$\Omega_{12} = \int_{\xi_1}^{\xi_3} M_1''(x) M_2''(x) dx = \frac{1}{\Delta^6} \int_{\xi_1}^{\xi_3} I^2(\xi_1 \leq x < \xi_2)(-24 z_2(x))(21 z_1(x) - 18) dx$$
(Here automatically set the integral term equal to zero if the region of indicator function are not overlapping.)
$$= \frac{1}{\Delta^6} \int_{\xi_1}^{\xi_2} (-24 z_2(x))(21 z_2(x) + 21 - 18) dx = \frac{1}{\Delta^5} \int_{-1}^{0} (-24 z_2)(21 z_2 + 3) dz_2$$
$$= \frac{-24}{\Delta^5} \int_{-1}^{0} 21 z_2^2 + 3 z_2 dz_2 = \frac{-24}{\Delta^5} \left. \left(7 z_2^3 + \frac{3}{2} z_2^2 \right) \right|_{z_2=-1}^{z_2=0} = \frac{-24}{\Delta^5} \left(\frac{11}{2} \right) = \frac{-132}{\Delta^5}$$
- $$\Omega_{22} = \int_{\xi_1}^{\xi_3} M_2''(x) M_2''(x) dx = \frac{1}{\Delta^6} \int_{\xi_1}^{\xi_2} (21 z_1(x) - 18)^2 dx + \frac{9}{\Delta^6} \int_{\xi_2}^{\xi_3} z_3(x)^2 dx$$
$$= \frac{1}{\Delta^5} \int_{0}^{1} (21 z_1 - 18)^2 dz_1 + \frac{9}{\Delta^5} \int_{-1}^{0} z_3^2 dz_3 = \frac{1}{\Delta^5} \left. \frac{(21 z_1 - 18)^3}{63} \right|_{z_1=0}^{z_1=1} + \frac{9}{\Delta^5} \left. \frac{z_3^3}{3} \right|_{z_3=-1}^{z_3=0}$$
$$= \frac{1}{\Delta^5} \times \frac{5859}{63} + \frac{9}{\Delta^5} \times \frac{1}{3} = \frac{93}{\Delta^5} + \frac{3}{\Delta^5} = \frac{96}{\Delta^5}$$
- $$\Omega_{13} = \int_{\xi_1}^{\xi_3} M_1''(x) M_3''(x) dx = \frac{24}{\Delta^6} \int_{\xi_1}^{\xi_2} 12 z_1(x) z_2(x) - 6 z_2(x) dx$$
$$= \frac{24}{\Delta^5} \int_{-1}^{0} 12 z_2^2 + 6 z_2 dz_2 = \frac{24}{\Delta^5} \left. \left(4 z_2^3 + 3 z_2^2 \right) \right|_{z_2=-1}^{z_2=0} = \frac{24}{\Delta^5}$$
- $$\Omega_{23} = \int_{\xi_1}^{\xi_3} M_2''(x) M_3''(x) dx = \int_{\xi_1}^{\xi_2} \frac{-(21 z_1(x) - 18)(12 z_1(x) - 6)}{\Delta^6} dx - \int_{\xi_2}^{\xi_3} \frac{3(12 z_2(x) z_3(x) - 6 z_3(x))}{\Delta^6} dx$$
$$= \frac{-1}{\Delta^5} \int_{0}^{1} (21 z_1 - 18)(12 z_1 - 6) dz_1 - \frac{3}{\Delta^5} \int_{-1}^{0} 12 z_3^2 + 6 z_3 dz_3$$
$$= \frac{-1}{\Delta^5} \left. \left(84 z_1^3 - 171 z_1^2 + 108 z_1 \right) \right|_{z_1=0}^{z_1=1} - \frac{3}{\Delta^5} \left. \left(4 z_3^3 + 3 z_3^2 \right) \right|_{z_3=-1}^{z_3=0} = \frac{-21}{\Delta^5} - \frac{3}{\Delta^5} = \frac{-24}{\Delta^5}$$
- $$\Omega_{33} = \int_{\xi_1}^{\xi_3} M_3''(x) M_3''(x) dx = \frac{1}{\Delta^5} \int_{0}^{1} (12 z_1 - 6)^2 dz_1 + \frac{1}{\Delta^5} \int_{0}^{1} (12 z_2 - 6)^2 dz_2$$
$$= 2 \times \frac{1}{\Delta^5} \left. \frac{(12 z_1 - 6)^3}{24} \right|_{z_1=0}^{z_1=1} = 2 \times \frac{1}{\Delta^5} \times \frac{432}{24} = \frac{72}{\Delta^5}$$

- $$\Omega_{14} = \int_{\xi_1}^{\xi_3} M_1''(x)M_4''(x)dx = \frac{1}{\Delta^6}\int_{\xi_1}^{\xi_2} -24z_2(x) \times 3z_1(x)dx$$

  $$= \frac{-72}{\Delta^5}\int_0^1 z_1(z_1-1)dz_1 = \frac{-72}{\Delta^5}\left(\frac{z_1^3}{3} - \frac{z_1^2}{2}\right)\bigg|_{z_1=0}^{z_1=1} = \frac{-72}{\Delta^5} \times \frac{-1}{6} = \frac{12}{\Delta^5}$$

- $$\Omega_{24} = \int_{\xi_1}^{\xi_3} M_2''(x)M_4''(x)dx = \frac{1}{\Delta^6}\int_{\xi_1}^{\xi_2} 63z_1(x)^2 - 54z_1(x)dx + \frac{1}{\Delta^6}\int_{\xi_2}^{\xi_3} 63z_2(x)z_3(x) - 9z_3(x)dx$$

  $$= \frac{1}{\Delta^5}\int_0^1 63z_1^2 - 54z_1dz_1 + \frac{1}{\Delta^5}\int_{-1}^0 63z_3^2 + 54z_3dz_3$$

  $$= \frac{1}{\Delta^5}\left(21z_1^3 - 27z_1^2\right)\bigg|_{z_1=0}^{z_1=1} + \frac{1}{\Delta^5}\left(21z_3^3 + 27z_3^2\right)\bigg|_{z_3=-1}^{z_3=0} = \frac{-6}{\Delta^5} + \frac{-6}{\Delta^5} = \frac{-12}{\Delta^5}$$

- $$\Omega_{34} = \int_{\xi_1}^{\xi_3} M_3''(x)M_4''(x)dx = \frac{-1}{\Delta^6}\int_{\xi_1}^{\xi_2} 36z_1(x)^2 - 18z_1(x)dx - \frac{1}{\Delta^6}\int_{\xi_2}^{\xi_3}(12z_2(x)-6)(21z_2(x)-3)dx$$

  $$= \frac{-1}{\Delta^5}\int_0^1 36z_1^2 - 18z_1dz_1 - \frac{1}{\Delta^5}\int_0^1(12z_2-6)(21z_2-3)dz_2$$

  $$= \frac{-1}{\Delta^5}\left(12z_1^3 - 9z_1^2\right)\bigg|_{z_1=0}^{z_1=1} - \frac{1}{\Delta^5}\left(84z_2^3 - 81z_2^2 + 18z_2\right)\bigg|_{z_2=0}^{z_2=1} = \frac{-3}{\Delta^5} - \frac{21}{\Delta^5} = \frac{-24}{\Delta^5}$$

- $$\Omega_{44} = \int_{\xi_1}^{\xi_3} M_4''(x)M_4''(x)dx = \frac{1}{\Delta^6}\int_{\xi_1}^{\xi_2} 9z_1(x)^2dx + \frac{1}{\Delta^6}\int_{\xi_2}^{\xi_3}(21z_2(x)-3)^2dx$$

  $$= \frac{1}{\Delta^5}\int_0^1 9z_1^2dz_1 + \frac{1}{\Delta^5}\int_0^1(21z_2-3)^2dz_2 = \frac{3z_1^3}{\Delta^5}\bigg|_{z_1=0}^{z_1=1} + \frac{(21z_2-3)^3}{63\Delta^5}\bigg|_{z_2=0}^{z_2=1} = \frac{3}{\Delta^5} + \frac{93}{\Delta^5} = \frac{96}{\Delta^5}$$

- $$\Omega_{15} = \int_{\xi_1}^{\xi_3} M_1''(x)M_5''(x)dx = \frac{-24^2}{\Delta^6}\int_{\xi_1}^{\xi_3} \underline{I(\xi_1 \le x < \xi_2)I(\xi_2 \le x < \xi_3)}z_2(x)^2dx = 0$$

- $$\Omega_{25} = \int_{\xi_1}^{\xi_3} M_2''(x)M_5''(x)dx = \frac{1}{\Delta^6}\int_{\xi_2}^{\xi_3} -72z_2(x)z_3(x)dx = \frac{-72}{\Delta^5}\int_0^1 z_2(z_2-1)dz_2$$

  $$= \frac{-72}{\Delta^5}\left(\frac{z_2^3}{3} - \frac{z_2^2}{2}\right)\bigg|_{z_2=0}^{z_2=1} = \frac{-72}{\Delta^5} \times \frac{-1}{6} = \frac{12}{\Delta^5}$$

- $$\Omega_{35} = \int_{\xi_1}^{\xi_3} M_3''(x)M_5''(x)dx = \frac{24}{\Delta^6}\int_{\xi_2}^{\xi_3} 12z_2(x)^2 - 6z_2(x)dx$$
$$= \frac{24}{\Delta^5}\int_0^1 12z_2^2 - 6z_2dz_2 = \frac{24}{\Delta^5}\left(4z_2^3 - 3z_2^2\right)\Big|_{z_2=0}^{z_2=1} = \frac{24}{\Delta^5}$$

- $$\Omega_{45} = \int_{\xi_1}^{\xi_3} M_4''(x)M_5''(x)dx = \frac{24}{\Delta^6}\int_{\xi_2}^{\xi_3} -21z_2(x)^2 + 3z_2(x)dx$$
$$= \frac{24}{\Delta^5}\int_0^1 -21z_2^2 + 3z_2dz_2 = \frac{24}{\Delta^5}\left(-7z_2^3 + \frac{3}{2}z_2^2\right)\Bigg|_{z_2=0}^{z_2=1} = \frac{24}{\Delta^5} \times \frac{-11}{2} = \frac{-132}{\Delta^5}$$

- $$\Omega_{55} = \int_{\xi_1}^{\xi_3} M_5''(x)M_5''(x)dx = \frac{24^2}{\Delta^6}\int_{\xi_2}^{\xi_3} z_2(x)^2dx = \frac{24^2}{\Delta^5}\int_0^1 z_2^2dz_2 = \frac{576}{\Delta^5}\times\frac{1}{3} = \frac{192}{\Delta^5}$$

- Since $\Omega_{ij} = \int_{\xi_1}^{\xi_3} M_i''(x)M_j''(x)dx = \int_{\xi_1}^{\xi_3} M_j''(x)M_i''(x)dx = \Omega_{ji}\ \forall i,j = 1,2,\cdots,5$, thus we have calculated $\mathbf{\Omega}$ as following,

$$\mathbf{\Omega} = \frac{1}{\Delta^5}\begin{bmatrix} 192 & -132 & 24 & 12 & 0 \\ -132 & 96 & -24 & -12 & 12 \\ 24 & -24 & 24 & -24 & 24 \\ 12 & -12 & -24 & 96 & 132 \\ 0 & 12 & 24 & -132 & 192 \end{bmatrix}$$

Appendix(R code).

```
> library(ISLR)
> age_wage =  Wage[,c("age","wage")]
>
> ###polynomial
> X = cbind( 1, age_wage[,1] ,age_wage[,1]^2,age_wage[,1]^3,age_wage[,1]^4)
> library(strucchange)  # using "solveCrossprod" instead of "solve"
>
> ( beta_hat = solveCrossprod( X ,method="chol")  %*% t(X) %*% age_wage[,2] )
              [,1]
[1,] -1.841542e+02
[2,]  2.124552e+01
[3,] -5.638593e-01
[4,]  6.810688e-03
[5,] -3.203830e-05
> sigma_hat = t(age_wage[,2] - X %*% beta_hat ) %*% (age_wage[,2] - X %*% beta_hat ) / (3000 - 5)
> ( se = sqrt( sigma_hat *  diag( solveCrossprod( X )  ) ) )# standard error of beta's
[1] 6.004038e+01 5.886748e+00 2.061083e-01 3.065931e-03 1.641359e-05
```

```
> # p - value
> apply( pt(beta_hat / se , df=2995 )  , 1, function(x) 2*min(x,1-x))
[1] 0.0021802539 0.0003123618 0.0062606445 0.0263977517 0.0510386497
> # 95% C.I
> for( i in 1:5){
+   cat("95%CI of beta",i-1, "is", beta_hat[i] - qt(0.975,2995) * se[i] , beta_hat[i] + qt(0.975,2995) * se[i],"¥n")
+ }
95%CI of beta 0 is -301.8787 -66.42963
95%CI of beta 1 is 9.703041 32.788
95%CI of beta 2 is -0.9679874 -0.1597312
95%CI of beta 3 is 0.0007991437 0.01282223
95%CI of beta 4 is -6.422135e-05 1.447431e-07
```

```
> #####################   Step function
> X_step = cbind(1, ifelse(age_wage[,1]>=35 & age_wage[,1]<65,1,0),ifelse(age_wage[,1]>=65 ,1,0)  )
> ( beta_hat_step = solveCrossprod( X_step ,method="chol")  %*% t(X_step) %*% age_wage[,2] )
          [,1]
[1,] 95.707068
[2,] 22.621790
[3,]  6.091915
> sigma_hat_step = crossprod(age_wage[,2] - X_step %*% beta_hat_step )  / (3000 - 5)
> (se_step = sqrt( sigma_hat_step *  diag( solveCrossprod( X_step )  ) ) )# standard error of beta's
> # p - value
> apply( pt(beta_hat_step / se_step , df=2995 )  , 1, function(x) 2*min(x,1-x))
[1] 0.0000000 0.0000000 0.2209564
> # 95% C.I
> for( i in 1:3){
+   cat("95%CI of beta",i-1, "is", beta_hat_step[i] - qt(0.975,2995) * se_step[i] ,
+       beta_hat_step[i] + qt(0.975,2995) * se_step[i],"¥n")
+ }
95%CI of beta 0 is 92.94434 98.46979
95%CI of beta 1 is 19.36112 25.88246
95%CI of beta 2 is -3.664931 15.84876
```

```
> ###############spline ,(xi1,xi2,xi3=18,49,80)
> z1 = function(x){ (x-18)/31 }    ;  z2 = function(x){ (x-49)/31 } ;  z3 = function(x){ (x-80)/31 }
> M1 = function(x){ ifelse( x >=18 & x<49,-4/31* z2(x)^3 ,0 )  }
> M2 = function(x){ ifelse( x >=18 & x<49,
+                          0.5/31*( 7*z1(x)^3 - 18 * z1(x)^2 + 12 *z1(x) ) ,    -0.5/31 * z3(x)^3 )  }
> M3 = function(x){ ifelse( x >=18 & x<49,
+                          ( -2*z1(x)^3 + 3 * z1(x)^2 )/31 , ( 2 * z2(x)^3 - 3 * z2(x)^2 + 1 )/31  )  }
> M4 = function(x){ ifelse( x >=18 & x<49,
+                          0.5/31 * z1(x)^3 , 0.5/31 * ( -7*z2(x)^3 + 3 * z2(x)^2 + 3 *z2(x) + 1 )  ) }
> M5 = function(x){ ifelse( x >=49 & x<= 80 ,4/31* z2(x)^3 ,0 )  }
> X_spline = cbind( sapply(age_wage[,1],M1),sapply(age_wage[,1],M2),sapply(age_wage[,1],M3),
+                   sapply(age_wage[,1],M4),sapply(age_wage[,1],M5) )
> Lambda = 1/31^5 * matrix( c(192,-132,24,12,0,-132,96,-24,-12,12,24,-24,24,-24,24,12,-12,-24,96,
+                          -132,0,12,24,-132,192),byrow=T,ncol=5 )
> dg_lambda = function(lambda.par){
+   sum( diag(   X_spline %*% solve( crossprod(X_spline) + lambda.par * Lambda) %*% t( X_spline )  ) )
+ }
> # determinant the value of lambda such that degree of freedom = 4
> uniroot( function(x) dg_lambda(x)-4,lower=0.1,upper=50000)$root
[1] 15727.51
>
> ( beta_hat_spline = solve( crossprod(X_spline) + 15727.51 * Lambda) %*% t( X_spline ) %*% age_wage[,2] )
          [,1]
[1,]  471.4864
[2,] 1704.0094
[3,] 2040.1272
[4,] 1675.9270
[5,]  719.5725
> sigma_hat_spline = ( crossprod(age_wage[,2] - X_spline %*% beta_hat_spline ) +
+                      15727.51 * t(beta_hat_spline) %*% Lambda %*% beta_hat_spline )/(3000-4)
> ( se_spline = sqrt( sigma_hat_spline * diag(    solve( crossprod(X_spline) + 15727.51 * Lambda) %*%
+                          crossprod(X_spline)   %*% solve( crossprod(X_spline) + 15727.51 * Lambda) ) ) )
[1] 28.21766 44.05113 59.70845 59.52036 56.38805
```

```
> # p - value
> apply( pt(beta_hat_spline / se_spline , df=2996 )  , 1, function(x) 2*min(x,1-x))
[1] 0 0 0 0 0
> # 95% C.I
> for( i in 1:5){
+   cat("95%CI of beta",i, "is", beta_hat_spline[i] - qt(0.975,2996) * se_spline[i] ,
+       beta_hat_spline[i] + qt(0.975,2996) * se_spline[i],"¥n")
+ }
95%CI of beta 1 is 416.1584 526.8143
95%CI of beta 2 is 1617.636 1790.383
95%CI of beta 3 is 1923.053 2157.201
95%CI of beta 4 is 1559.222 1792.632
95%CI of beta 5 is 609.0093 830.1357
```