Topics in Time Series (6)

Yisen Du (Eason)

1 Question1

In the file "Prob1HW7DataStat153.csv", you will find simulated data (consisting of $n = 600$ observations) on two variables x and y. Your goal is to fit a smooth curve to this data by working out the following steps.

a) Fit a linear regression model of the form (below $\epsilon \sim N(0, \sigma^2)$)

$$y = \beta_0 + \beta_1 x + \sum_{j=1}^{99} \beta_{j+1} \left(x - \frac{j}{100} \right)_+ + \epsilon$$

using the standard regression methodology (from a Bayesian point of view, this corresponds to the prior $\beta_0, \beta_1, \ldots, \beta_{100}, \log \sigma \overset{\text{i.i.d}}{\sim} N(0, C)$ for a large C). On the scatter plot of the data, plot the fitted function:

$$\hat{f}^{ls}(x) := \beta_0^{ls} + \beta_1^{ls} x + \sum_{j=1}^{99} \beta_{j+1}^{ls} \left(x - \frac{j}{100} \right)_+ .$$

Comment on whether this least squares solution $\hat{f}^{ls}$ adequately summarizes the relationship between x and y. (4 points)

b) We shall now fit the model (1) with the prior (below $C = 10^6$)

$$\beta_0 \sim N(0, C) \text{ and } \beta_1, \ldots, \beta_{100} \overset{\text{i.i.d}}{\sim} N\left(0, \tau^2\right)$$

where τ is a tuning parameter. Calculate the loglikelihood of the observed data as a function of τ and σ. Evaluating the loglikelihood over a grid of possible values of τ and σ, calculate and report the Maximum Likelihood Estimates $\hat{\tau}$ and $\hat{\sigma}$ of τ and σ. Describe how you selected the range for placing this grid. (6 points)

c) Calculate the posterior means of $\beta_0, \ldots, \beta_{100}$ under the prior (2) with τ and σ set to the Maximum Likelihood estimates derived in the previous part. Denote these posterior means by $\tilde{\beta}_0^{\hat{\tau}}, \ldots, \tilde{\beta}_{100}^{\hat{\tau}}$ Plot the fitted function:

$$\tilde{f}^{\tau}(x) := \tilde{\beta}_0^{\tau} + \tilde{\beta}_1^{\tau} x + \sum_{j=1}^{99} \tilde{\beta}_{j+1}^{\tau} \left(x - \frac{j}{100} \right)_+$$

on the scatter plot of the data along with the least squares solution $\hat{f}^{ls}$. Comment on the differences between the two fitted functions $\tilde{f}^{\tau}$ and $\hat{f}^{ls}$. (6 points).

d) This data was actually simulated from the model:

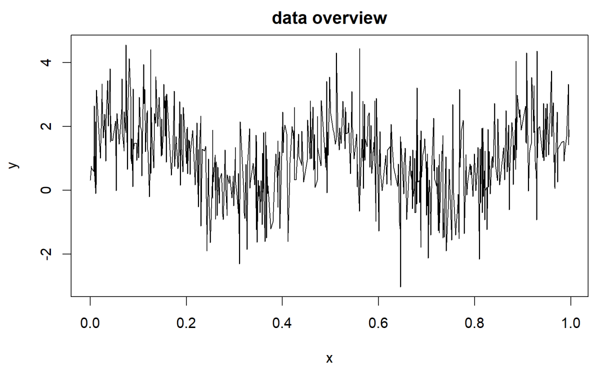
$$y_i = f^*(x_i) + \epsilon_i$$

with

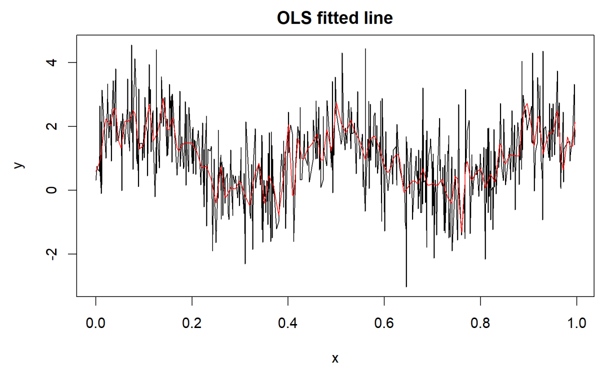
$$f^*(x) := \sin(15x) + \exp\left(-\frac{x^2}{2}\right) + \frac{1}{2}(x - 0.5)^2.$$

Comment on the closeness (or lack of closeness) of $\tilde{f}^{\tau}$ and $\hat{f}^{ls}$ to f^*. (3 points)

1.1 Answer of Q1(a)



(a) figure1



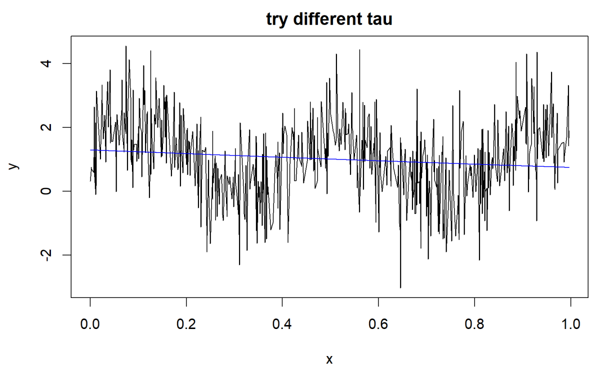
(b) figure2

The fitted values looks good but it's not smooth. There might be over-fitting problem. We need to fit some smoother model then.

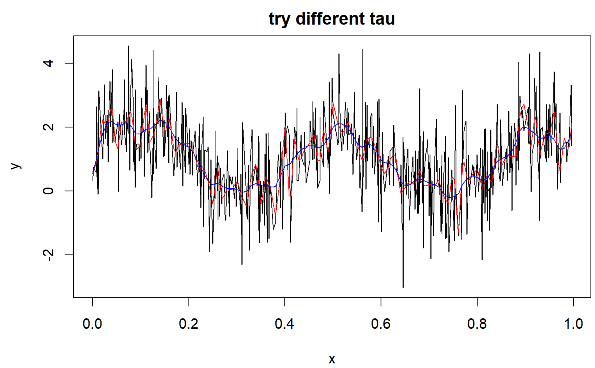
1.2 Answer of Q1(b)

To determine the range of σ, we refer to the residual standard error of OLS model. The value is 1.032. We could try different σ around it.
To determine the range of τ, we try different values first.

To select the range, we firstly try some different values to get wiggly model and underfitted model. Then, we have a rough idea of gird range. Then, we try the numbers in these range and check that whether the best values are lie on the boundary. If not, we use these range and do more granular grid. Otherwise, we adjust the ranges and repeat the above steps.



(a) figure1



(b) figure2

We try models in this range of τ. The best model is given by

$$\tau = 5.111$$
$$\sigma = 1.031$$

1.3 Answer of Q1(c)

The model of (b) is given by

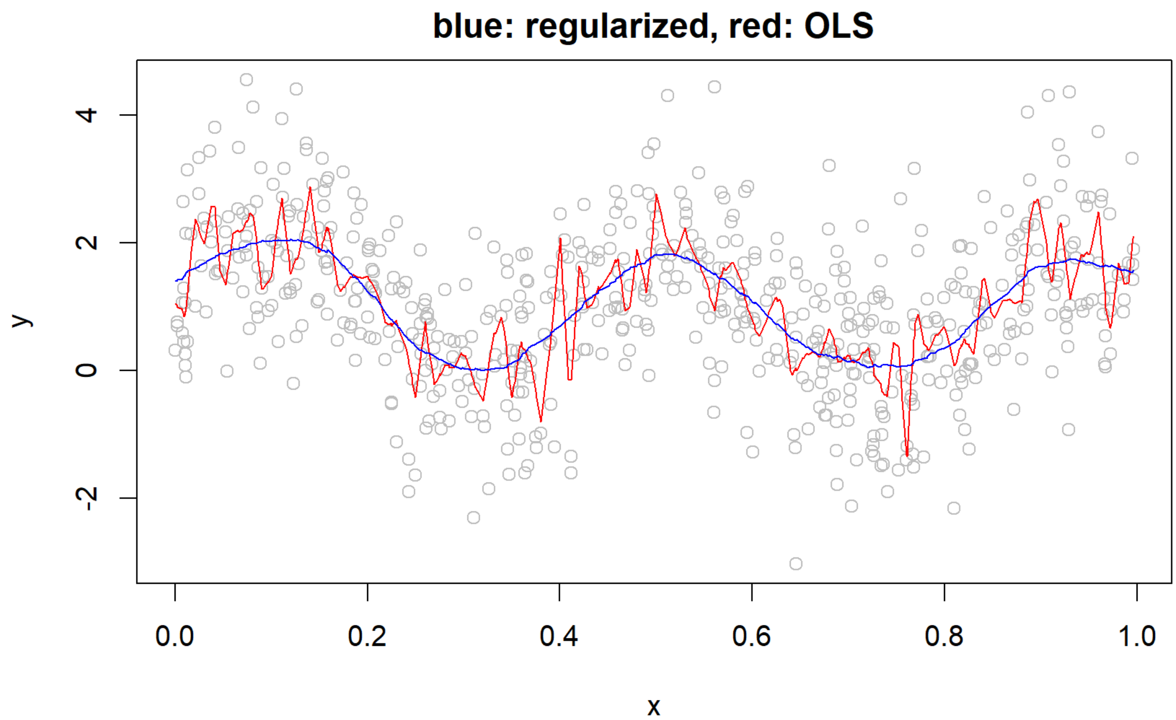


Figure 3: a figure

While both models shows similar trend, the second model is smooth and the first model is wiggly. It seems that the second model alleviate over-fitting issue and will have better prediction power.

1.4 Answer of Q1(d)

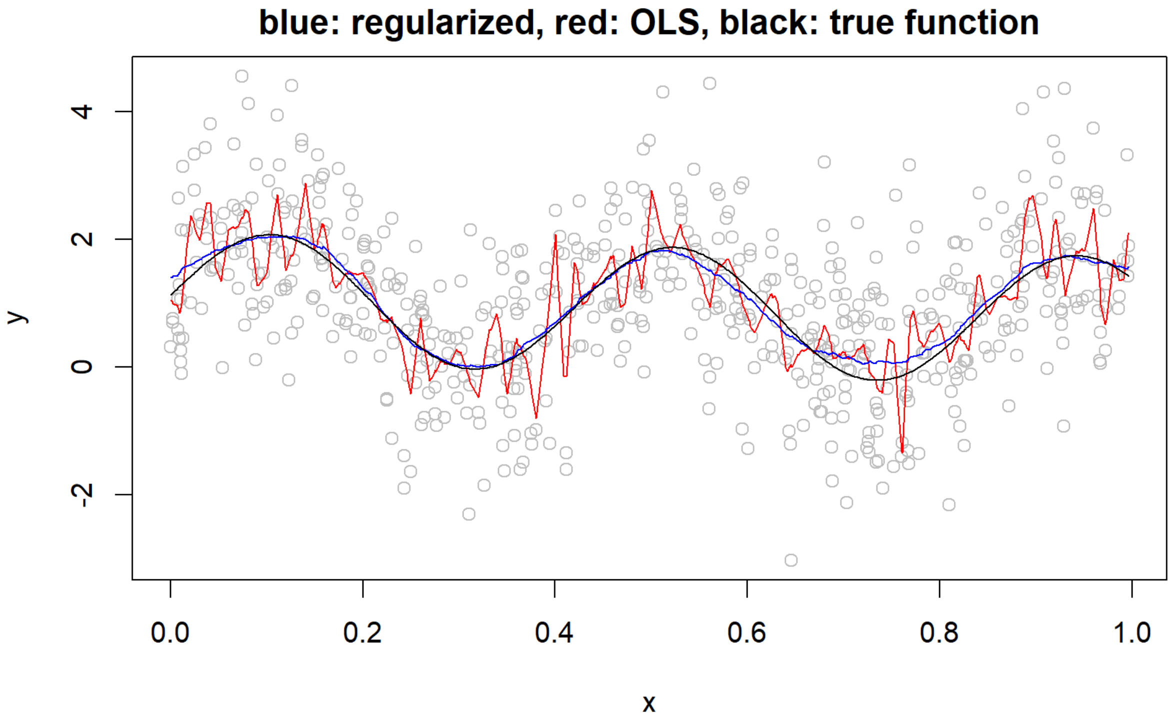


Figure 4: a figure

It's obvious that the second model is closer to true model. The regularized one better capture the whole trend while the OLS model has the over-fitting issue.

2 Question2

Download the google trends data (from trends.google.com) for the query facebook. This should be a monthly time series dataset that indicates the worldwide search popularity of facebook from January 2004 to November 2022. Plot and look at the data. Work out the following steps to fit a smooth curve to Y_t as a function of time t (we shall assume that t ranges from 1 to n where n is the length of the dataset).

a) Fit the model

$$Y_t = \beta_0 + \beta_1(t-1) + \sum_{j=2}^{n-1} \beta_j (t-j)_+ + Z_t$$

for $t = 1, \ldots, n$ with $Z_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. Note that this is a linear regression model with n data points and n coefficients so the usual linear regression will overfit the data and not give anything useful. We shall instead work with the prior:

$$\beta_0, \beta_1 \sim N(0, C) \text{ and } \beta_2, \ldots, \beta_{n-1} \overset{\text{i.i.d}}{\sim} N\left(0, \tau^2\right)$$

for a parameter $\tau > 0$ (as usual $C = 10^6$ is a constant). Calculate the loglikelihood of the observed data as a function of τ and σ. Evaluating the loglikelihood over a grid of possible values of τ and σ, calculate and report the Maximum Likelihood Estimates $\hat{\tau}$ and $\hat{\sigma}$ of τ and σ. Describe how you selected the range for placing this grid. (6 points)

b) Calculate the posterior means of $\beta_0, \ldots, \beta_{n-1}$ under the prior (3) with τ and σ set to the Maximum Likelihood estimates derived in the previous part. Denote these posterior means by $\tilde{\beta}_0^{\hat{\tau}}, \ldots, \tilde{\beta}_{n-1}^{\hat{\tau}}$ Plot the fitted function:

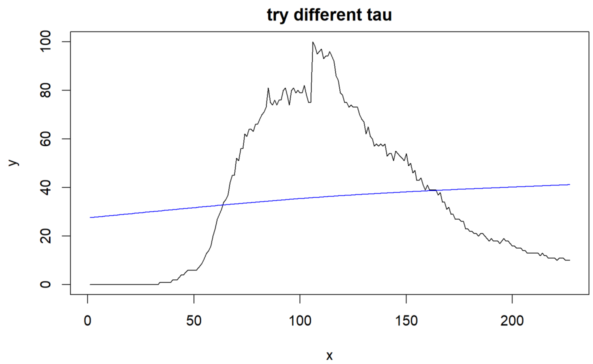
$$\tilde{f}^{\tau}(t) := \tilde{\beta}_0^{\tau} + \tilde{\beta}_1^{(\tau)}(t-1) + \sum_{j=2}^{n-1} \tilde{\beta}_j^{\tau}(t-j)_+$$

on the scatter plot of the data. (6 points).

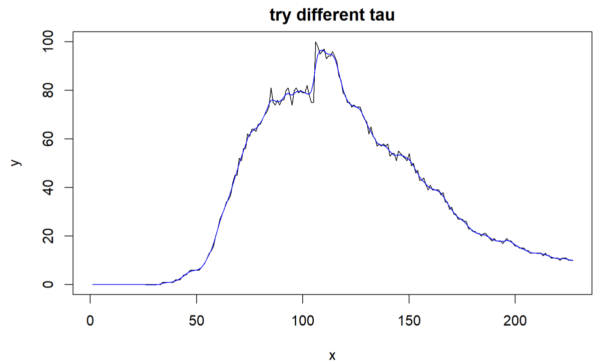
c) Comment on whether the estimated function $\tilde{f}^{\hat{\tau}}$ is capturing well the trends present in the data while being smooth. (3 points).

2.1 Answer of Q2(a)

To select the range, we firstly try some different values to get wiggly model and under-fitted model. Then, we have a rough idea of gird range. Then, we try the numbers in these range and check that whether the best values are lie in the boundary. If not, we use these range and do more granular grid. Otherwise, we adjust the ranges and repeat the above steps.



(a) $\tau = 0.0001,\ \sigma = 1.5$



(b) $\tau = 1,\ \sigma = 1.5$

2.2 Answer of Q2(b)

The best model is given by

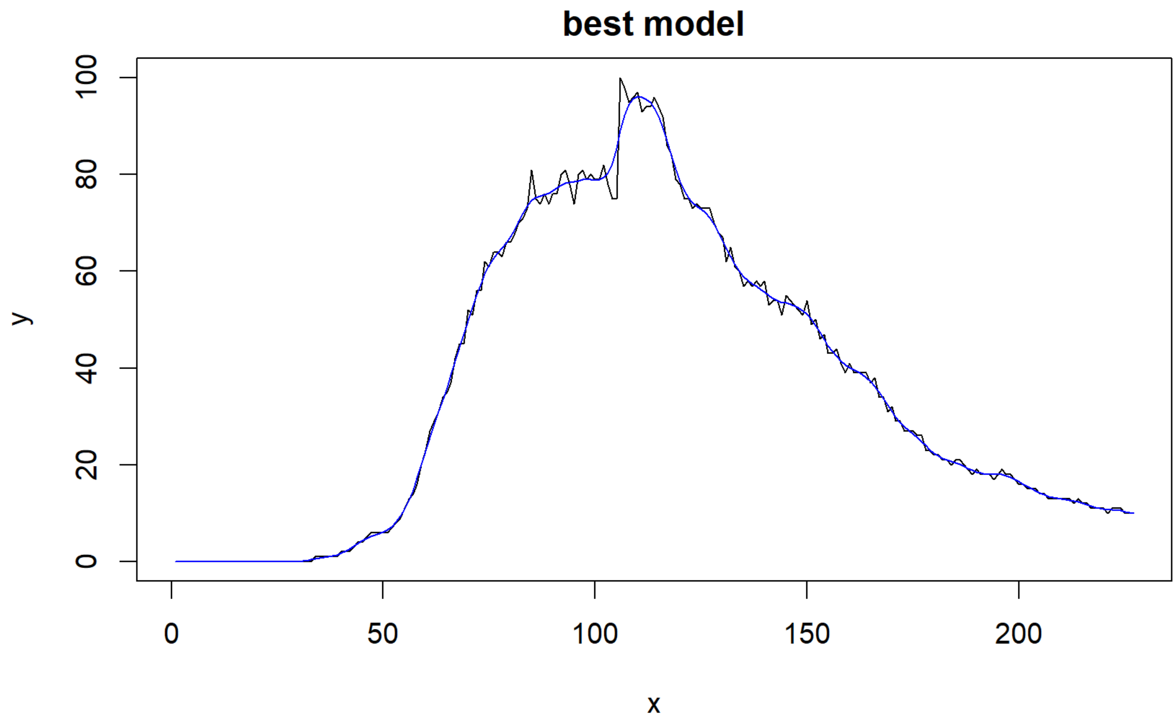


Figure 6: $\tau = 0.531, \ \sigma = 1.775$

2.3 Answer of Q2(c)

The model is smooth and it captures the trend. It omits some small changes and shows the clear trends in different periods.

3 Question3

This problem asks you to prove two results that were used in Lectures 23 and 24. Suppose X and Y are two random vectors such that

$$X \sim N\left(m_0, Q_0\right) \quad \text{and} \quad Y \mid X = x \sim N(Bx, R).$$

Here Q_0 and R are fixed nonsingular matrices. Also m_0 is a fixed vector and B are fixed matrix. Prove then that

$$X \mid Y = y \sim N\left(m_1, Q_1\right)$$

where
$m_1 = \left(Q_0^{-1} + B^T R^{-1} B\right)^{-1} \left(Q_0^{-1} m_0 + B^T R^{-1} y\right)$ and $Q_1 = \left(Q_0^{-1} + B^T R^{-1} B\right)^{-1}$, by following the steps given below.

a) Show that (1 point)

$$f_{X|Y=y}(x) \propto \exp\left(-\frac{1}{2}\left[\left(x - m_0\right)^T Q_0^{-1}\left(x - m_0\right) + (y - Bx)^T R^{-1}(y - Bx)\right]\right)$$

b) Show that (3 points)

$$\left(x - m_0\right)^T Q_0^{-1}\left(x - m_0\right) + (y - Bx)^T R^{-1}(y - Bx) = \left(x - m_1\right)^T Q_1^{-1}\left(x - m_1\right) + E$$

where E is a term that does not depend on x.

c) Put the above two facts together to prove (5). (2 points).

Under the same assumptions (4), one can also show that

$$Y \sim N\left(Bm_0, BQ_0B^T + R\right).$$

Prove this by following the steps below.

a) Using $Y \mid X = x \sim N(Bx, R)$, argue that $Y - BX \mid X = x \sim N(0, R)$. (1 point)

b) Argue from the above that $Y - BX \sim N(0, R)$ and that $Y - BX$ is independent of X. (1 point)

c) Using the fact that if Z_1 and Z_2 are two multivariate normal random vectors that are independent, then their sum $Z_1 + Z_2$ is also multivariate normal, deduce (6). (3 points).

3.1 Answer of Q3(a)

By definition of multivariate Gaussian, we have

$$f_{Y|X=x}(y) \propto exp(-\frac{1}{2}(y-Bx)^T R^{-1}(y-Bx))$$
$$f_X(x) \propto exp(-\frac{1}{2}(x-m_0)^T Q_0^{-1}(x-m_0))$$

By the formula of conditional probability, we have

$$\begin{aligned} f_{X|Y=y}(x) &\propto f_X(x) f_{Y|X=x}(y) \\ &\propto exp(-\frac{1}{2}[(y-Bx)^T R^{-1}(y-Bx) + (x-m_0)^T Q_0^{-1}(x-m_0)]) \end{aligned}$$

3.2 Answer of Q3(b)

To prove

$$(x-m_0)^T Q_0^{-1}(x-m_0) + (y-Bx)^T R^{-1}(y-Bx) = (x-m_1)^T Q_1^{-1}(x-m_1) + E$$

We first consider LHS. Note that

$$\begin{aligned} &(y-Bx)^T R^{-1}(y-Bx) + (x-m_0)^T Q_0^{-1}(x-m_0) \\ =\, & x^T(Q_0^{-1} + B^T R^{-1} B)x - x^T(Q_0^{-1} m_0 + B^T R^{-1} y) - (m_0^T Q_0^{-1} + y^T R^{-1} B)x + m_0^T Q_0^{-1} m_0 + y^T R^{-1} y \end{aligned}$$

Let's define

$$A := Q_0^{-1} + B^T R^{-1} B$$
$$B := Q_0^{-1} m_0 + B^T R^{-1} y$$

Thus, we have

$$LHS = x^T A x - x^T B - B^T x + m_0^T Q_0^{-1} m_0 + y^T R^{-1} y$$

For RHS, we have

$$(x-m_1)^T Q_1^{-1}(x-m_1) + E = x^T A x - x^T B - B^T x + B^T A^{-1} B + E$$

Therefore, we have

$$LHS = RHS$$

3.3 Answer of Q3(c)

From (a) and (b), we have

$$\begin{aligned} f_{X|Y=y}(x) &\propto exp((x-m_1)^T \, Q_1^{-1} \, (x-m_1) + E) \\ &\propto exp((x-m_1)^T \, Q_1^{-1} \, (x-m_1)) \end{aligned}$$

By definition of Gaussian, we have

$$X|Y = y \sim N(m_1, Q_1)$$

3.4 Answer of Q3(d)

Note that

$$\begin{aligned} [Y - BX|X = x] &=_{s.t.} [Y|X = x] - [BX|X = x] \\ &=_{s.t.} [Y|X = x] - Bx \end{aligned}$$

By property of Normal,

$$[Y|X = x] - Bx \sim N(0, R)$$

Thus, we have

$$[Y - BX|X = x] \sim N(0, R)$$

3.5 Answer of Q3(e)

From

$$Y - BX|X = x \sim N(0, R) \forall x$$

Then we have

$$Y - BX|X \sim N(0, R)$$

Therefore,

$$\begin{aligned} &Y - BX|X \sim Y - BX \\ &\longrightarrow Pr\{Y - BX = m|X = m\} = Pr\{Y - BX = m\} \forall m, n \\ &\longrightarrow Pr\{Y - BX = m, X = m\} = Pr\{Y - BX = m\} Pr\{X = n\} \forall m, n \\ &\longrightarrow \text{By def. of indep, } Y - BX \perp X \end{aligned}$$

3.6 Answer of Q3(f)

Note that

$$X \sim N(m_0, Q_0)$$

By property of Gaussian (linear comb.),

$$BX \sim N(Bm_0, BQ_0B^T)$$

Since $Y - BX \perp BX$, by property of Gaussian (summation), we have

$$Y - BX + BX \sim N(Bm_0, R + BQ_0B^T)$$

Thus, proved.

4 Question4

4. Consider the ex1029 dataset from the R package Sleuth3 that we used in Lectures 23 and 24. Take a random sample of $n = 500$ observations from the full dataset and call this dataset D_1. Take another random sample of size 100 and set this dataset D_2 as a test dataset. We shall build models on D_1 and test their performance via prediction on D_2.

a) Let $y = \log($ Earnings $)$ and $x_1 =$ years of experience. Fit the following model to D_1 :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - 1)_+ + \beta_3 (x_1 - 2)_+ + \cdots + \beta_{64} (x_1 - 63)_+ + Z$$

using the standard 1 m function (this corresponds to the prior $\beta_0, \beta_1, \ldots, \beta_{64}, \log \sigma \overset{\text{i.i.d}}{\sim} N(0, C)$ or Unif $(-C, C)$ for a large C, say $C = 10^6$). Use your fitted model to obtain point predictions for the response variable y in the test dataset D_2 and report the average squared prediction error. (3 points)

b) As in our analysis in class, fit the model (7) with the prior:

$$\beta_0, \beta_1 \overset{\text{i.i.d}}{\sim} N(0, C) \text{ and } \beta_2, \ldots, \beta_{64} \overset{\text{i.i.d}}{\sim} N\left(0, \tau^2\right)$$

where τ and σ (σ^2 is the variance of Z) are treated as unknown parameters. Fit this model to D_1 and obtain point estimates of τ, σ and use them to obtain point estimates of $\beta_0, \beta_1, \ldots, \beta_{64}$. Use your fitted model to obtain point predictions for the response variable y in the test dataset D_2 and report the average squared prediction error. (4 points)

c) Let x_2 denote the variable years of education. Fit the following model to D_1 :

$$y = \gamma_0 + \gamma_1 x_2 + \gamma_2 (x_2 - 1)_+ + \cdots + \gamma_{17} (x_2 - 17)_+ + Z$$

using the prior

$$\gamma_0, \gamma_1 \overset{\text{i.i.d}}{\sim} N(0, C) \text{ and } \gamma_2, \ldots, \gamma_{17} \overset{\text{i.i.d}}{\sim} N\left(0, \tau^2\right)$$

Repeat the analysis from part (b) for this model and obtain point predictions for the response variable y in the test dataset D_2 and report the average squared prediction error. (4 points)

d) Now fit the model:

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 (x_1 - 1)_+ + \beta_3 (x_1 - 2)_+ + \cdots + \beta_{64} (x_1 - 63)_+ \\ & + \gamma_1 x_2 + \gamma_2 (x_2 - 1)_+ + \cdots + \gamma_{17} (x_2 - 17)_+ + Z \end{aligned}$$

using the prior

$\beta_0, \beta_1, \gamma_1 \overset{\text{i.i.d}}{\sim} N(0, C)$ and $\beta_2, \ldots, \beta_{64} \overset{\text{i.i.d}}{\sim} N(0, \tau_1^2)$ and $\gamma_2, \ldots, \gamma_{17} \overset{\text{i.i.d}}{\sim} N(0, \tau_2^2)$

Treat τ_1, τ_2, σ as three parameters and obtain point estimates of them. Use them to obtain point estimates of $\beta_0, \ldots, \beta_{64}, \gamma_1, \ldots, \gamma_{17}$. Use your fitted model to obtain point predictions for the response variable y in the test dataset D_2 and report the average squared prediction error. (6 points)

e) Which of these models has the best prediction performance, and is it intuitively clear why the best model is performing better than the others? (2 points)

4.1 Answer of Q4(a)

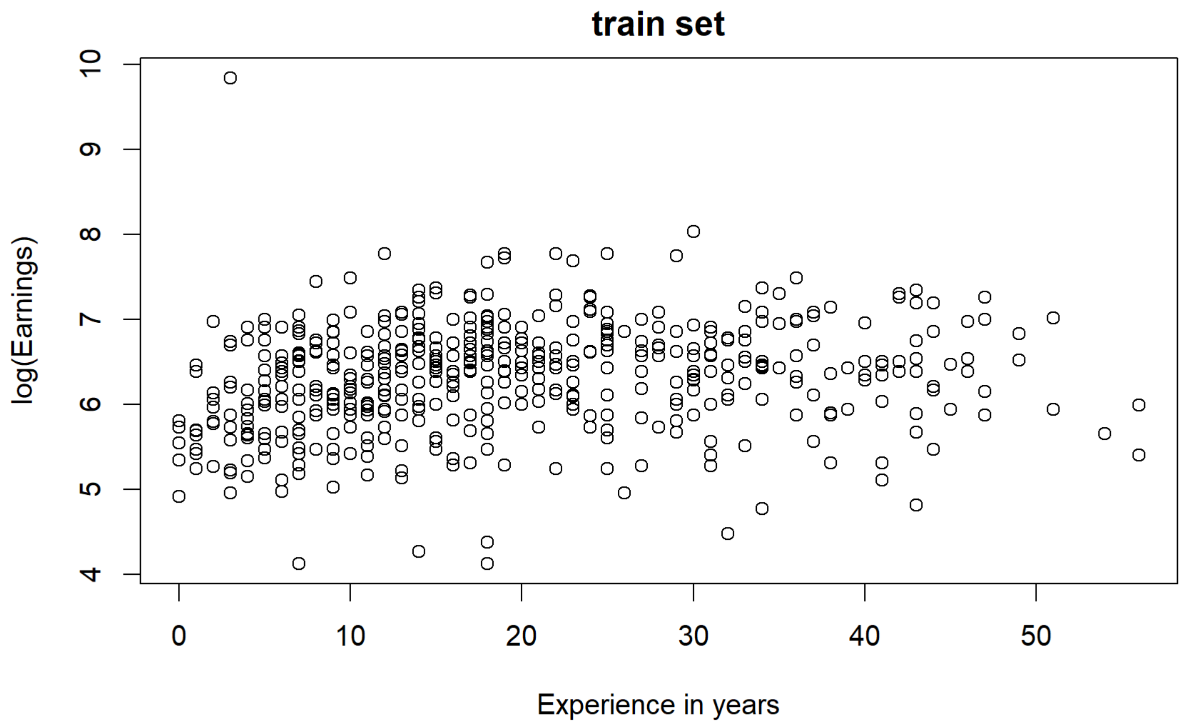


Figure 7: Overview

By OLS, the coefficients of model are give by

```
         x1          x2          x3          x4          x5          x6 
 5.45826191  0.23812276  0.03424308 -0.41637170  0.32562856 -0.07234652 
         x7          x8          x9         x10         x11         x12 
-0.38278596  0.41379235  0.14167878 -0.55981044  0.57248688 -0.39629521 
        x13         x14         x15         x16         x17         x18 
 0.19887463 -0.23995008  0.30387491  0.13113447 -0.31828583 -0.17682764 
        x19         x20         x21         x22         x23         x24 
 0.42438513 -0.37775153  0.19556240  0.04903348  0.17183701 -0.69385728 
        x25         x26         x27         x28         x29         x30 
 0.47764015  0.02699331 -0.19005511  0.35588400 -0.38316700  0.09630776 
        x31         x32         x33         x34         x35         x36 
 1.03278042 -1.51286654  0.16214988  0.83213617 -0.96944958  0.46168793 
        x37         x38         x39         x40         x41         x42 
-0.05576406  0.38166832 -0.30787363 -0.17317297  0.47462437 -0.27934664 
        x43         x44         x45         x46         x47         x48 
-0.51224533  0.61156775  0.18183799 -0.19961338 -0.14283461 -0.70500486 
        x49         x50         x51         x52         x53         x54 
 1.12232412 -0.28140098 -1.13560221  3.19264264 -2.61652624          NA 
        x55         x56         x57         x58         x59         x60 
         NA  0.38350588          NA          NA          NA          NA 
        x61         x62         x63         x64         x65 
         NA          NA          NA          NA          NA 
```

Figure 8: a figure

The average squared error is 3555.308.

4.2 Answer of Q4(b)

For this model, the coefficients of model are give by

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	5.6775423648	0.0613474771	-0.0008690940	-0.0019066525	-0.0022057507
[2,]	-0.0017191632	-0.0020002439	-0.0014942257	-0.0013624864	-0.0012682509
[3,]	-0.0025904403	-0.0021295161	-0.0019225547	-0.0016046090	-0.0004029702
[4,]	-0.0029774197	-0.0046814538	-0.0048263654	-0.0041593316	-0.0038762284
[5,]	-0.0028381918	-0.0031118212	-0.0027830311	-0.0021095995	-0.0019606748
[6,]	-0.0007150928	-0.0003314433	-0.0003227125	-0.0003139818	-0.0003052510
[7,]	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000

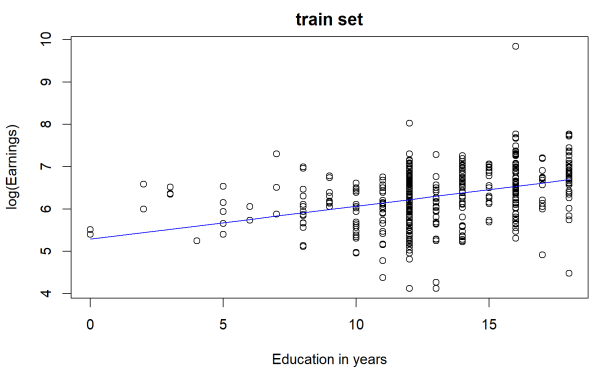
	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	-2.702198e-03	-2.797282e-03	-2.014434e-03	-0.0020019604	-0.002395046
[2,]	-2.350933e-03	-3.748224e-03	-4.208780e-03	-0.0035561392	-0.003466011
[3,]	7.514564e-05	1.329072e-04	1.879834e-05	-0.0008965574	-0.001726964
[4,]	-2.869899e-03	-2.320274e-03	-2.142723e-03	-0.0025319452	-0.002744715
[5,]	-2.391879e-03	-2.461734e-03	-2.174252e-03	-0.0014454921	-0.001070797
[6,]	-1.572136e-04	-7.860681e-05	0.000000e+00	0.0000000000	0.000000000
[7,]	5.677542e+00	6.134748e-02	-8.690940e-04	-0.0019066525	-0.002205751

Figure 9: a figure

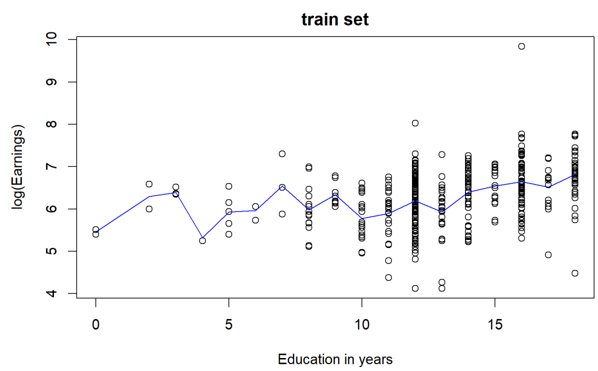
The average squared error is 1655.212

4.3 Answer of Q4(c)

For this model, let's try some σ and τ first to observe the data.



(a) figure1



(b) figure2

For this model, the coefficients of model are give by

```
             [,1]         [,2]         [,3]         [,4]
[1,] 5.317101e+00 7.584740e-02 7.001027e-07 1.400205e-06
[2,] 2.216077e-06 8.231010e-06 1.309734e-05 2.458728e-05
[3,] 2.970040e-05 3.678889e-05 5.071813e-05 6.721476e-05
[4,] 8.282705e-05 9.680246e-05 8.382841e-05 6.262925e-05
[5,] 4.633498e-05 2.183118e-05 9.611057e-06 5.317101e+00
```

Figure 11: a figure

The average squared error is 0.331

4.4 Answer of Q4(d)

For this model, the coefficients of model are give by

```
              [,1]          [,2]          [,3]          [,4]          [,5]
[1,]  4.859798e+00  1.862439e-02 -1.120022e-06 -2.237909e-06 -8.456712e-06
[2,] -2.049446e-04 -2.314018e-04 -2.572188e-04 -2.768015e-04 -2.831390e-04
[3,] -3.588966e-04 -3.484891e-04 -3.371730e-04 -3.279670e-04 -3.157351e-04
[4,] -1.292587e-04 -1.038019e-04 -8.105709e-05 -5.904911e-05 -3.766863e-05
[5,] -2.150075e-07  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
[6,] -1.243811e-06  3.125494e-06  8.770058e-06  1.420393e-05  2.546476e-05
              [,6]          [,7]          [,8]          [,9]         [,10]
[1,] -1.029509e-05 -2.021882e-05 -3.460813e-05 -5.139127e-05 -6.928726e-05
[2,] -2.966989e-04 -3.072649e-04 -3.158667e-04 -3.268669e-04 -3.390536e-04
[3,] -2.943562e-04 -2.821121e-04 -2.624198e-04 -2.435847e-04 -2.286428e-04
[4,] -1.826601e-05 -1.055144e-05 -3.589074e-06 -1.412410e-06 -1.261830e-06
[5,]  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  8.307811e-02
[6,]  3.545919e-05  5.164430e-05  5.295665e-05  4.799951e-05  4.581911e-05
             [,11]         [,12]         [,13]         [,14]
[1,] -9.545674e-05 -1.187444e-04 -1.495670e-04 -1.749125e-04
[2,] -3.482972e-04 -3.580004e-04 -3.595359e-04 -3.585522e-04
[3,] -2.087520e-04 -1.897332e-04 -1.676816e-04 -1.506405e-04
[4,] -1.060930e-06 -8.600299e-07 -6.450224e-07 -4.300150e-07
[5,]  2.615137e-06  4.548850e-06  4.323817e-06  1.496748e-06
[6,]  3.657711e-05  1.074883e-05  4.439944e-06  4.859798e+00
```

Figure 12: a figure

The average squared error is 0.24

4.5 Answer of Q4(e)

Model in (d) gives the best prediction.

- We include two variables education and experience in this model. The multiple linear regression has stronger power to depict the data.
- We use τ_1 and τ_2 to avoid over-fitting so the model has better predictions.