Non-Linear Regression Analysis of a Typical Bivariate Dataset

Suryasis Jana
Roll no - 201447

March, 2022

(Under guidance of Dr. Debasis Kundu)

Contents

1 Introduction

Consider a statistical model as,

$$y_i = f(x_i, \theta) + \epsilon_i,\ \ i = 1, 2, \ldots, n$$

where
y_i : *the dependent variable*
x_i : *the independent variable(s)*
θ : *unknown parameter vector*
ϵ_i : *error random variables*, $\ \ i = 1, 2, \ldots, n$

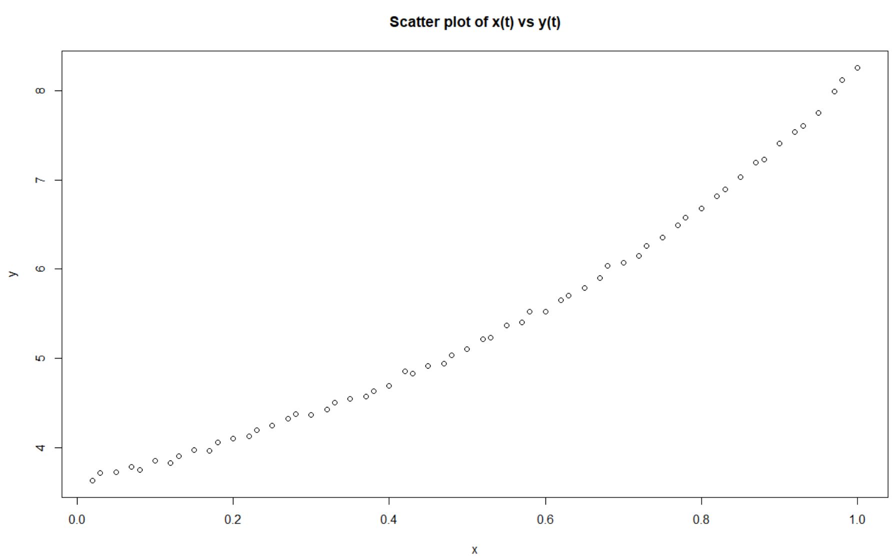
This model is considered as a Non Linear Model if f is a non linear function in θ.
Depending on the functional form of f there are several types of non linear models like Sum of Exponential Models, Ratio of Polynomial Models, Sum of Sinusoidal Model etc.
We are given a bivariate data set of the form

$$\{(x_t, y_t) : t = 1, 2, \ldots, n\},\ \ \textit{where}\ n = 60$$

and we are interested in fitting a non linear model in this data.

2 Plot of the given data

We have plot the given data in the following scatter plot



From the plot we can observe that the graph is of an increasing and a convex function.

3 Model

Now we wish to fit the data using the following non linear model:

$$y_t = \alpha_1^* + \alpha_2^* e^{\beta^* x_t} + \epsilon_t, \quad t = 1, 2, \ldots, n \qquad \ldots (1)$$

,where $\{\epsilon_t\}$ is a sequence of i.i.d. normal random variables with mean zero and finite variance say σ^2, and $\alpha_1^*, \alpha_2^*, \beta^*$ are the true value of the unknown parameters $\alpha_1, \alpha_2, \beta$. Here we assume $(\alpha_1, \alpha_2, \beta)^T \in \mathcal{R}^3$, which is our parameter space.
This model is in the form of a Sum of Exponential Model.

4 Estimation of Parameters

4.1 Method of Least Squares

The list square estimators of $(\alpha_1^*, \alpha_2^*, \beta^*)$ can be obtained by minimizing

$$Q(\alpha_1, \alpha_2, \beta) = \sum_{t=1}^{n} (y_t - \alpha_1 - \alpha_2 e^{\beta x_t})^2$$

with respect to $\alpha_1, \alpha_2, \beta$, i.e.,

$$\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \\ \hat{\beta} \end{bmatrix}_{LSE} = \arg \min_{\alpha_1, \alpha_2, \beta} Q(\alpha_1, \alpha_2, \beta)$$

4.2 Plot of residual sum of squares as a function of β

Now at first we want to plot the above residual sum of square function as a function of β. In order to do that we will minimize $Q(\alpha_1, \alpha_2, \beta)$ with respect to α_1, α_2 for a fixed β. Then we shall got estimates of α_1, α_2 as a function of β, viz., $\widehat{\alpha}_1(\beta), \widehat{\alpha}_2(\beta)$.

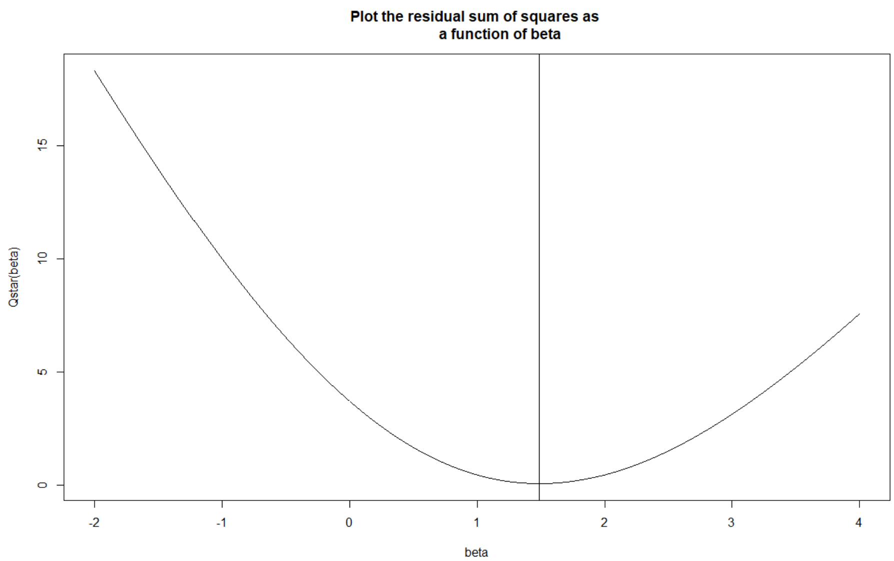
$$\begin{cases} \frac{\partial}{\partial \alpha_2} Q(\alpha_1, \alpha_2, \beta) = 0 \\ \frac{\partial}{\partial \alpha_1} Q(\alpha_1, \alpha_2, \beta) = 0 \end{cases}$$

$$\implies \begin{cases} \widehat{\alpha}_2(\beta) = & \frac{\sum_{t=1}^{n}(e^{\beta x_t}-\frac{1}{n}\sum_{i=1}^{n}e^{\beta x_i})(y_t-\bar{y})}{\sum_{t=1}^{n}(e^{\beta x_t}-\frac{1}{n}\sum_{i=1}^{n}e^{\beta x_i})^2} \\ \widehat{\alpha}_1(\beta) = & \bar{y}-\widehat{\alpha}_2(\beta)\frac{1}{n}\sum_{t=1}^{n}e^{\beta x_t} \end{cases}$$

Then inserting these estimates into Q we get the residual sum of squares only as a function of β as the following

$$Q^*(\beta) = Q(\widehat{\alpha}_1(\beta), \widehat{\alpha}_2(\beta), \beta) = \sum_{t=1}^{n}(y_t - \widehat{\alpha}_1(\beta) - \widehat{\alpha}_2(\beta)e^{\beta x_t})^2$$

Then the plot of $Q^*(\beta)$ versus β is the following



This Plot is actually I have generated using R, by taking 1000 equispaced values of beta from -2 to 4 and by calculating the values of the function $Q^*(\beta)$ for these values of β. This plot is of a function which has the global minima near 1.489489, which is got by looking at which of the 1000 equispaced values the minima of the function $Q^*(\beta)$ occurs, using R.
So, for the Gauss Newton method we can take our initial value of β as $\beta^{(0)} = 1.489489$. Similarly we shall take the initial values of α_1, α_2 as $\alpha_1^{(0)} = \widehat{\alpha}_1(\beta^{(0)}) = 2.284847, \alpha_2^{(0)} = \widehat{\alpha}_2(\beta^{(0)}) = 1.339692$

4.3 Vectorized form of our non linear model

Now we can express our model (1) in vectorized form as

$$y = f(\theta^*) + \epsilon \qquad \ldots(2)$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \theta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}, f(\theta) = \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \\ \vdots \\ f_n(\theta) \end{bmatrix}$ with $f_t(\theta) = f(x_t, \theta) = \alpha_1 + \alpha_2 e^{\beta x_t}, \quad t = 1, 2, \ldots, n$ and $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$ and θ^* is the true value of parameter θ. Also, $E(\epsilon) = 0, D(\epsilon) = \sigma^2 I$.

4.4 Gauss-Newton Method

Now proceed to obtain the least square estimate of the parameters by minimizing $Q(\alpha_1, \alpha_2, \beta)$ using the Gauss-Newton method.

We want to take initial values of the parameters $\alpha_1, \alpha_2, \beta$ as $\alpha_1^{(0)} = 2.284847, \alpha_2^{(0)} = 1.339692, \beta^{(0)} = 1.489489$.

Now we shall denote the $n \times 3$ matrix as

$$F.(\theta) = ((\frac{\partial}{\partial \theta_j} f_i(\theta))), i = 1, 2, \ldots, n, j = 1, 2, 3$$

So, in our model this matrix looks like

$$F.(\theta) = \begin{bmatrix} 1 & e^{\beta x_1} & \alpha_2 x_1 e^{\beta x_1} \\ 1 & e^{\beta x_2} & \alpha_2 x_2 e^{\beta x_2} \\ \vdots & \vdots & \vdots \\ 1 & e^{\beta x_n} & \alpha_2 x_n e^{\beta x_n} \end{bmatrix}_{n \times 3}$$

Using Gauss-Newton method, for initial value $\theta^{(0)}$ our iterating formula is the following

$$\theta^{(k+1)} = \theta^{(k)} + (F.^T(\theta^{(k)}) F.(\theta^{(k)}))^{-1} F.^T(\theta^{(k)}) r(\theta^{(k)})$$

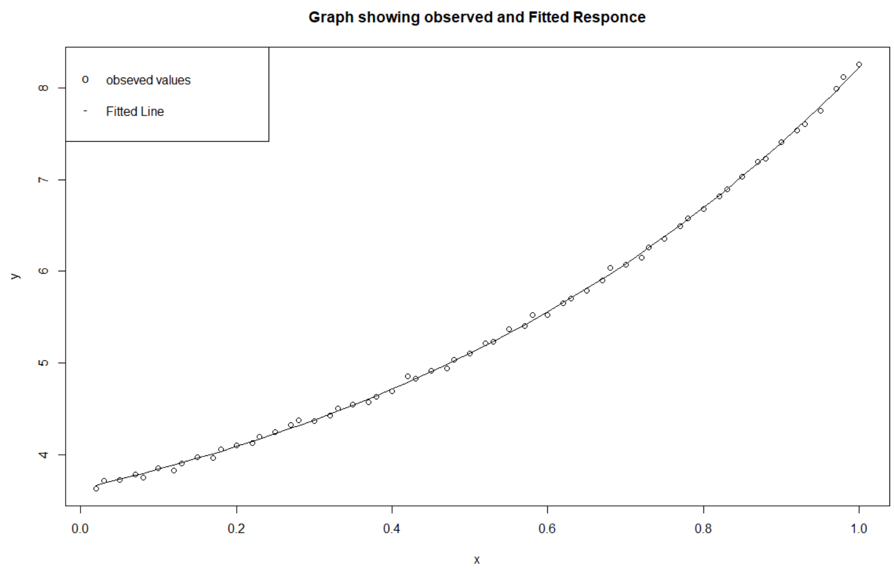
, $k = 0, 1, 2, \ldots$, where $r(\theta^{(k)}) = y - f(\theta^{(k)})$
and stop the iteration when the norm of $\theta^{(k+1)}$ and $\theta^{(k)}$ is very small.

4.5 Least Square Estimation Results

Using R, considering the tolerance level of stopping criterion of the iteration as 0.000001, we have got the least square estimates of $\alpha_1^*, \alpha_2^*, \beta^*$ as

$$\hat{\alpha_1} = 2.288989, \hat{\alpha_2} = 1.336200, \hat{\beta} = 1.491535$$

For this estimated value, $\hat{\theta} = (2.288989, 1.336200, 1.491535)^T$ we have plot the observed values and fitted responses in a same graph, which is given below:



Note that in the above plot the line represents the graph of $y = \hat{\alpha_1} + \hat{\alpha_2} e^{\hat{\beta} x}$. So, we can observe that the observed values and fitted line is nicely matching. So this our model fits good to the given data.

To see whether our initial choices of parameters affects our final estimates or not, we have taken various initial choices $\theta^{(0)}$ as following,
For $\theta^{(0)} = (0, 0.5, 0.5)^T$ we have got $\hat{\theta} = (3.014, 0.678, -237.3)^T$
For $\theta^{(0)} = (1, 0.5, 0.6)^T$ we have got $\hat{\theta} = (7.464, -2.730, -31.401)^T$

So, we see that if we do not take the initial values close enough to the true values of the parameter then the Gauss-Newton algorithm may not converge to the true value of the parameter.

4.6 Estimation of σ^2

The MLE of σ^2 is given by

$$\hat{\sigma^2} = \frac{1}{n}(y - f(\hat{\theta}))^T(y - f(\hat{\theta})))$$

Using R, We have got the value as $\hat{\sigma^2} = 0.0009186769$
So the estimated standard error is $\hat{\sigma} = \sqrt{0.0009186769} = 0.03030968$, which is very small. So, our estimation is very good.

5 Interval Estimation

Again consider the vectorized form of the non inear model, i.e., model (2). Then the likelihood function is

$$L(\theta, \sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}(y-f(\theta))^T(y-f(\theta))}$$

The log-likeihood is

$$l(\theta, \sigma 2) = lnL(\theta, \sigma^2) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}(y-f(\theta))^T(y-f(\theta))$$

Let us denote $\gamma^{4\times1} = \begin{bmatrix} \theta^{3\times1} \\ \sigma^2 \end{bmatrix}$. Then the Fisher Information matrix is

$$I = -E(\frac{\partial^2 l}{\partial\gamma\partial\gamma^T})|_{\gamma=\gamma^*} = \begin{bmatrix} -E(\frac{\partial^2 l}{\partial\theta\partial\theta^T}) & -E(\frac{\partial^2 l}{\partial\sigma^2\partial\theta}) \\ -E(\frac{\partial^2 l}{\partial\sigma^2\partial\theta^T}) & -E(\frac{\partial^2 l}{\partial(\sigma^2)^2}) \end{bmatrix}_{\gamma=\gamma^*}$$

From the log likelihood function $l(\gamma)$ we get
$E(\frac{\partial^2 l}{\partial\theta\partial\theta^T})|_{\theta=\theta^*} = -\frac{1}{\sigma^2}F.^T(\theta^*)F.(\theta^*)$
$E(\frac{\partial^2 l}{\partial(\sigma^2)^2}) = -\frac{n}{2\sigma^4}$
$E(\frac{\partial^2 l}{\partial\sigma^2\partial\theta}) = 0$

So the Fisher Information matrix would be

$$I = \begin{bmatrix} \frac{1}{\sigma^2}F.^T(\theta^*)F.(\theta^*) & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

So, the inverse of the Fisher Information matrix is

$$I^{-1} = \begin{bmatrix} \sigma^2(F.^T(\theta^*)F.(\theta^*))^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

5.1 Confidence Intervals

So, the estimated inverse of Fisher Information Matrix is

$$\widehat{I^{-1}} = \begin{bmatrix} \hat{\sigma}^2(F.^T(\hat{\theta})F.(\hat{\theta}))^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

So, using R we have got the matrix as

$$\widehat{I^{-1}} = \begin{bmatrix} 0.002491743 & -0.0020520946 & 0.0011827468 & 0 \\ -0.002052095 & 0.0017147290 & -0.0009970732 & 0 \\ 0.001182747 & -0.0009970732 & 0.0005851230 & 0 \\ 0 & 0 & 0 & 2.813224e-08 \end{bmatrix}$$

Since $\hat{\theta} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\beta} \end{bmatrix}$ so, approximately $\widehat{var}(\hat{\alpha}_1)$ is the $(1,1)^{th}$, $\widehat{var}(\hat{\alpha}_2)$ is the $(2,2)^{th}$, $\widehat{var}(\hat{\beta})$ is the $(3,3)^{th}$ element of $\widehat{I^{-1}}$.
So, approximately we have
$\widehat{SE}(\hat{\alpha}_1) = \sqrt{\widehat{var}(\hat{\alpha}_1)} = 0.04991736$
$\widehat{SE}(\hat{\alpha}_2) = \sqrt{\widehat{var}(\hat{\alpha}_2)} = 0.04140929$
$\widehat{SE}(\hat{\beta}) = \sqrt{\widehat{var}(\hat{\beta})} = 0.02418932.$
Since we know that under certain regularity conditions $\hat{\gamma}$ is approximately $\mathcal{N}_4(\gamma, I^{-1})$, so the approximate 95% confidence intervals are the following
CI of α_1^*: $(\hat{\alpha}_1 - \tau_{0.025}\widehat{SE}(\hat{\alpha}_1), \hat{\alpha}_1 + \tau_{0.025}\widehat{SE}(\hat{\alpha}_1)) = (2.191152, 2.386825)$
CI of α_2^*: $(\hat{\alpha}_2 - \tau_{0.025}\widehat{SE}(\hat{\alpha}_2), \hat{\alpha}_2 + \tau_{0.025}\widehat{SE}(\hat{\alpha}_2)) = (1.255040, 1.417361)$
CI of β^*: $(\hat{\beta} - \tau_{0.025}\widehat{SE}(\hat{\beta}), \hat{\beta} + \tau_{0.025}\widehat{SE}(\hat{\beta})) = (1.444125, 1.538946)$
,where τ_α is the $(1-\alpha)^{th}$ quantile of a Standard Normal Distribution.

5.2 Confidence sets

Now we know that under certain regularity conditions

$$\hat{\gamma} - \gamma^* \to \mathcal{N}_4(0, I^{-1})$$

So, $\hat{\theta} - \theta^* \to \mathcal{N}_3(0, \sigma^2(F.^T(\theta^*)F.(\theta^*))^{-1})$,
$\implies \frac{(\hat{\theta}-\theta^*)^T F.^T(\theta^*)F.(\theta^*)(\hat{\theta}-\theta^*)}{\sigma^2} \approx \chi_3^2$
which is independently distributed with $\hat{\sigma}^2$, where
$(n-3)\hat{\sigma}^2 = (y - f(\hat{\theta}))^T(y - f(\hat{\theta})) = y^T(I - P_{F.(\theta^*)})y \sim \sigma^2\chi_{n-3}^2$

where $P_{F.(\theta^*)} = F.(\theta^*)(F.^T(\theta^*)F.(\theta^*))^{-1}F.^T(\theta^*)$.

So, an approximate $100(1-\alpha)\%$ confidence interval for $\theta^* = \begin{bmatrix}\alpha_1^* \\ \alpha_2^* \\ \beta^*\end{bmatrix}$ is given by

$$\left\{\theta : \frac{n-3}{3}\frac{(\hat{\theta}-\theta)^T F.^T(\hat{\theta})F.(\hat{\theta})(\hat{\theta}-\theta)}{y^T(I-P_{F.(\hat{\theta})})y} \leq F_{3,n-3,\alpha}\right\}$$

Using R, We have got

$$\frac{n-3}{3} = 19,\ y^T(I-P_{F.(\hat{\theta})})y = 0.05512,\ F_{3,n-3,0.05} = 2.766,$$

$$F.^T(\hat{\theta})F.(\hat{\theta}) = \begin{bmatrix} 60.0000 & 140.2629 & 117.7317 \\ 140.2629 & 386.4976 & 375.0850 \\ 117.7317 & 375.0850 & 402.7513 \end{bmatrix}$$

So, our confidence set becomes

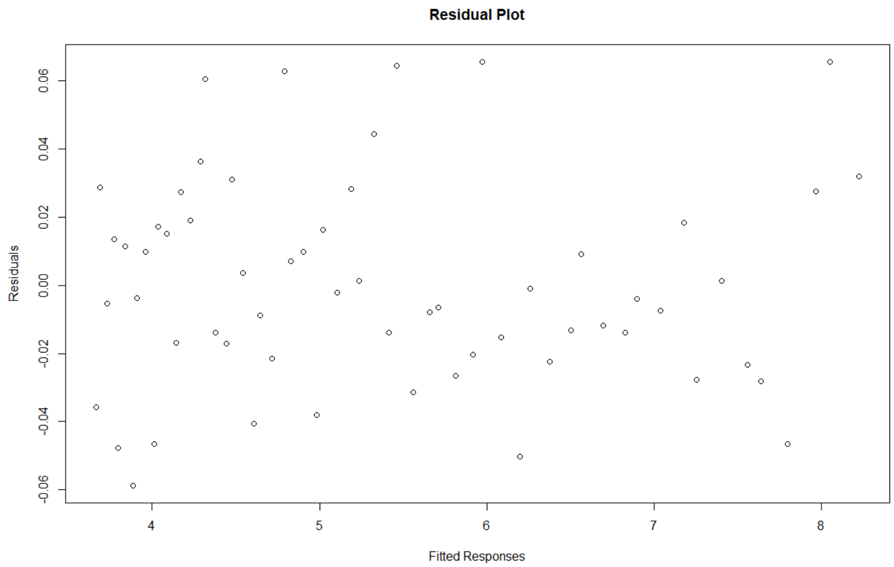
$$\left\{\theta : (\hat{\theta}-\theta)^T \begin{bmatrix} 60.0000 & 140.2629 & 117.7317 \\ 140.2629 & 386.4976 & 375.0850 \\ 117.7317 & 375.0850 & 402.7513 \end{bmatrix} (\hat{\theta}-\theta) \leq 0.008026\right\}$$

where $\hat{\theta} = (2.289, 1.336, 1492)^T$.

6 Model Diagnostics

6.1 Residual Plot

We have plot the residuals $\hat{\epsilon}_t = y_t - f_t(\hat{\theta})$ against the fitted values $\hat{y}_t = f_t(\hat{\theta}),\ t = 1, 2, \ldots, n$ in the following residual plot.



In this plot we see that the points are randomly scattered. Which indicates that our model has a good fit with the data.

6.2 Test for Normality of Errors

6.2.1 Shapiro Wilk's Normality Test

Now we proceed to check whether the residuals $\hat{\epsilon}_t,\ t = 1, 2, \ldots, n$ come from a normal distribution or not. To check this we shall perform the Shapiro Wilk's Normality test for the residuals. Here we want to test

$$H_0 : \text{Residuals are normally distributed vs } H_1 : \text{not } H_0$$

The Shapiro Wilk's test statistic is given by,

$$W = \frac{(\sum_{i=1}^{n} a_i \hat{\epsilon}_{(i)})^2}{\sum_{i=1}^{n} (\hat{\epsilon}_i - \hat{\bar{\epsilon}})^2}$$

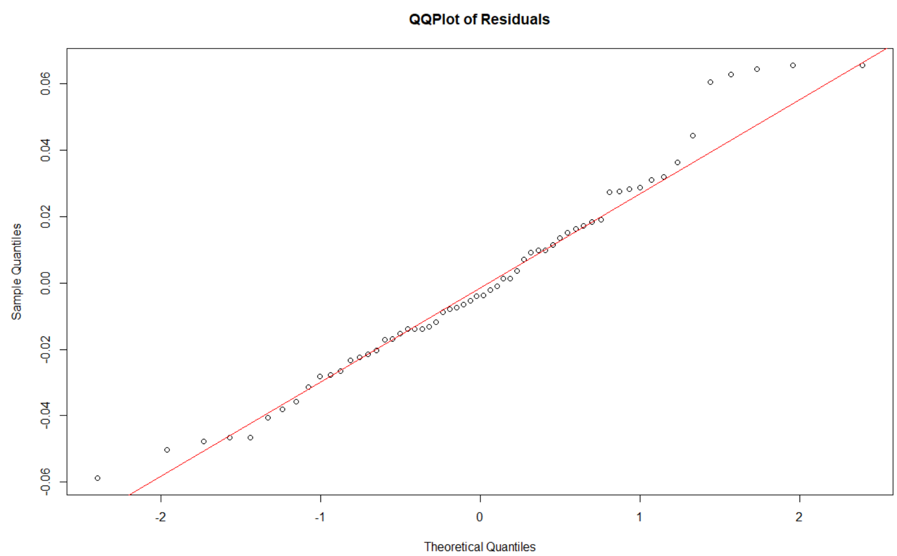
where $\hat{\epsilon}_{(i)}$ is the i^{th} order statistic of $\hat{\epsilon}_i s$ and $\hat{\bar{\epsilon}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i$
The coefficients $a_i s$ are given by $(a_1, a_2, \ldots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$, where $m^T = (m_1, m_2, \ldots, m_n)$ is made of the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and V is

the dispersion matrix of these order statistics.

Using R, we have got the p-value of the test $p - value = 0.1696$, which is greater than 0.05. So, we do not reject H_0 at level 0.05. So, we conclude that the residuals satisfies the normality assumptions.

6.2.2 QQ Plot of Residuals

We have also obtained the QQ Plot of the residuals and the plot looks like the following



From the above QQ Plot we see that the graph matches nicely with 45^o diagonal line. So, we conclude that the residuals are from a normal distribution.

7 Analysis Using NL2SOL Algorithm through R function "nls"

We also have used the 'NL2SOL' algorithm through the standard R function 'nls' to get the the least square estimates of the parameters $\alpha_1^*, \alpha_2^*, \beta^*$ with respect to the model (1).

7.1 Estimation of $\alpha_1, \alpha_2, \beta$

We have taken the same initial value as before, i.e., $\alpha_1^{(0)} = 2.284847, \alpha_2^{(0)} = 1.339692, \beta^{(0)} = 1.489489$.
Then using these initial values in the NL2SOL algorithm through the standard R function "nls" we have got the Least Square Estimate of the parameters as

$$\hat{\alpha_1} = 2.28899, \hat{\alpha_2} = 1.33620, \hat{\beta} = 1.49154$$

,which is exactly same as the results we got through the Gauss-Newton algorithm.

7.2 Estimation of σ^2

Using this 'nls' function we have got the estimate of the error variance $\hat{\sigma^2}$ as

$$\hat{\sigma^2} = \frac{1}{n}(y - f(\hat{\theta}))^T(y - f(\hat{\theta}))) \approx 0.0009189$$

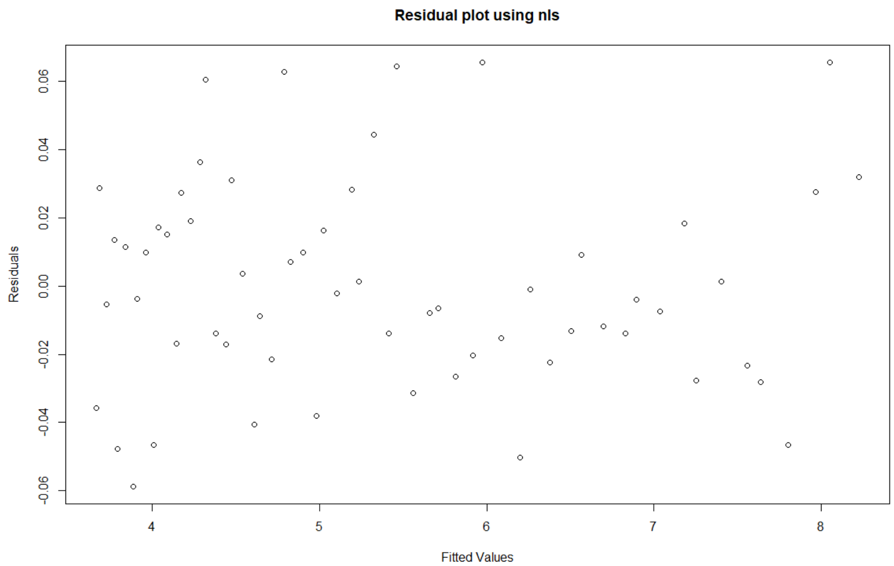
which is exactly same as the result we got before.

7.3 Confidence Intervals

Since the estimates of the model parameters are exactly same as before, so we have got the 95% confidence intervals of $\alpha_1^*, \alpha_2^*, \beta^*$ exactly same as before, i.e.,
CI of α_1^*: $(\hat{\alpha}_1 - \tau_{0.025}\widehat{SE}(\hat{\alpha}_1), \hat{\alpha}_1 + \tau_{0.025}\widehat{SE}(\hat{\alpha}_1)) = (2.191152, 2.386825)$
CI of α_2^*: $(\hat{\alpha}_2 - \tau_{0.025}\widehat{SE}(\hat{\alpha}_2), \hat{\alpha}_2 + \tau_{0.025}\widehat{SE}(\hat{\alpha}_2)) = (1.255040, 1.417361)$
CI of β^*: $(\hat{\beta} - \tau_{0.025}\widehat{SE}(\hat{\beta}), \hat{\beta} + \tau_{0.025}\widehat{SE}(\hat{\beta})) = (1.444125, 1.538946)$

7.4 Residual Plot

We have got the following residual plot, which is also same as before:



In this plot the points are randomly scattered, indicating that our fit is good.

7.5 Test for Normality

It is also obvious that we have got the exactly same result as before for testing of normality of the residuals using Shapiro Wilk's Test and also the same QQ plot. So, the residuals are normally distributed.

8 Fitting Another Model

Now on the same data set let us consider another sum of exponential model

$$y_t = \alpha_1^* e^{\beta_1^* x_t} + \alpha_2^* e^{\beta_2^* x_t} + \epsilon_t, \quad t = 1, 2, \ldots, n \quad \ldots (3)$$

,where $\{\epsilon_t\}$ is a sequence of i.i.d. normal random variable with mean zero and finite variance say σ^2, and $\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*$ are the true value of the unknown parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$. Here we assume

$(\alpha_1, \alpha_2, \beta_1, \beta_2)^T \in \mathcal{R}^4$, which is our parameter space.
As previously done, for this model the vectorized version is given by

$$y = f(\theta^*) + \epsilon$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \theta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}, f(\theta) = \begin{bmatrix} f_1(\theta) \\ f_2(\theta) \\ \vdots \\ f_n(\theta) \end{bmatrix}$ with $f_t(\theta) = f(x_t, \theta) = \alpha_1 e^{\beta_1 x_t} + \alpha_2 e^{\beta_2 x_t}, \quad t = 1, 2, \ldots, n$ and $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$ and θ^* is the true value of parameter θ. Also, $E(\epsilon) = 0, D(\epsilon) = \sigma^2 I$.
Now we shall denote the $n \times 4$ matrix as

$$F.(\theta) = ((\frac{\partial}{\partial \theta_j} f_i(\theta))), i = 1, 2, \ldots, n, j = 1, 2, 3, 4$$

So, in our model this matrix looks like

$$F.(\theta) = \begin{bmatrix} e^{\beta_1 x_1} & e^{\beta_2 x_1} & \alpha_1 x_1 e^{\beta_1 x_1} & \alpha_2 x_1 e^{\beta_2 x_1} \\ e^{\beta_1 x_2} & e^{\beta_2 x_2} & \alpha_1 x_2 e^{\beta_1 x_2} & \alpha_2 x_2 e^{\beta_2 x_2} \\ \vdots & \vdots & \vdots & \vdots \\ e^{\beta_1 x_n} & e^{\beta_2 x_n} & \alpha_1 x_n e^{\beta_1 x_n} & \alpha_2 x_n e^{\beta_2 x_n} \end{bmatrix}_{n \times 4}$$

8.1 Estimation of Parameters

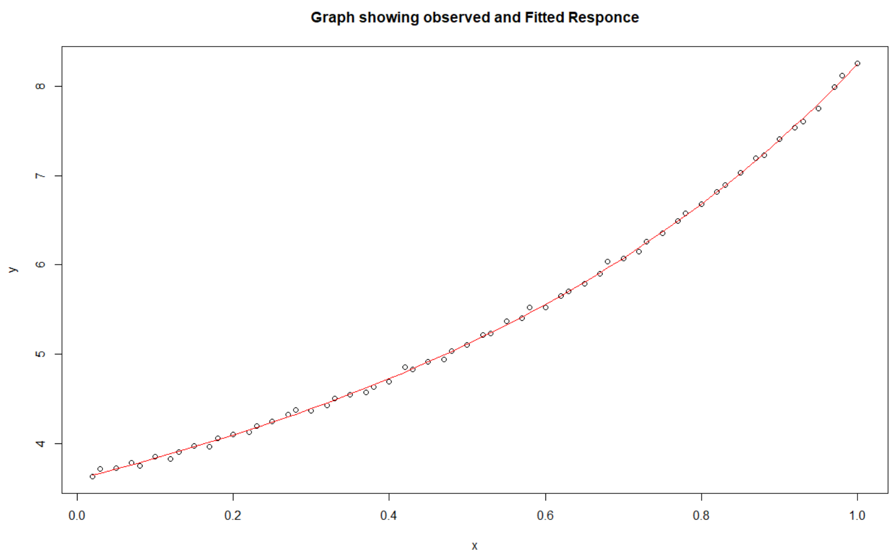
Using Gauss-Newton method, for initial value $\theta^{(0)}$ our iterating formula is the following

$$\theta^{(k+1)} = \theta^{(k)} + (F.^T(\theta^{(k)}) F.(\theta^{(k)}))^{-1} F.^T(\theta^{(k)}) r(\theta^{(k)})$$

, $k = 0, 1, 2, \ldots$, where $r(\theta^{(k)}) = y - f(\theta^{(k)})$
and stop the iteration when the norm of $\theta^{(k+1)}$ and $\theta^{(k)}$ is very small.
For this iteration from our previous knowledge, we shall take the initial values as $\alpha_1^{(0)} = 1, \beta_1^{(0)} = 1, \alpha_2^{(0)} = 1, \beta_2^{(0)} = 1$.
As the same method used previously, we have got the least square estimates of $\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*$ as

$$\hat{\alpha_1} = 3.2003, \hat{\alpha_2} = 0.4005, \hat{\beta_1} = 0.4199, \hat{\beta_2} = 2.1330$$

For this estimated value, $\hat{\theta} = (3.2003, 0.4005, 0.4199, 2.1330)^T$ we have plot the observed values and fitted responses in a same graph, which is given below:



From the above plot we see that the observed values and the fitted values are matches very nicely. So, the filling of model (3) is also very satisfactory.

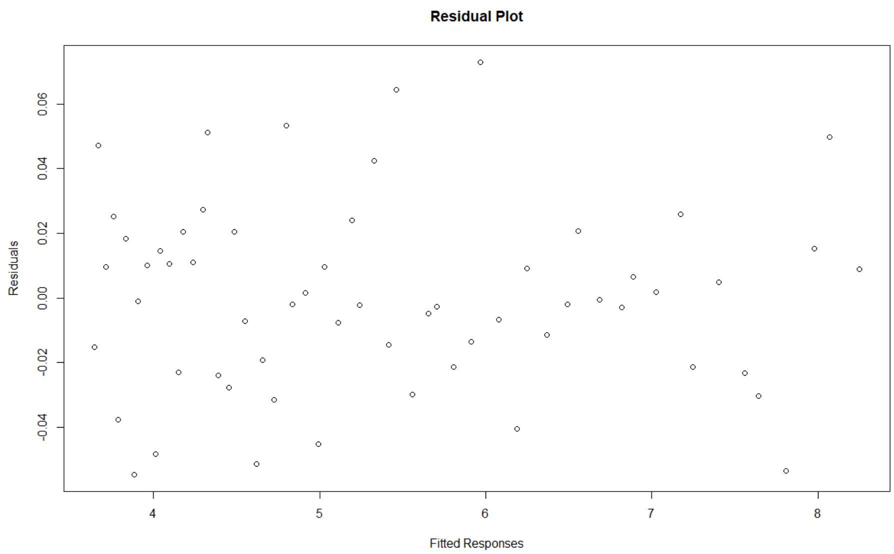
8.2 Model Diagnostics

8.2.1 Estimation of σ^2

Similarly as the previous analysis we have got the value of estimates error variance as $\hat{\sigma^2} = 0.0008380$.
So, estimated standard error is $\hat{\sigma} = 0.02894897$, which is very small, so our fitting is very good.

8.2.2 Residual Plot

We have also obtained the following residual plot



In this plot we see that the points are randomly scattered, which indicates that our model has a good fit with the data.

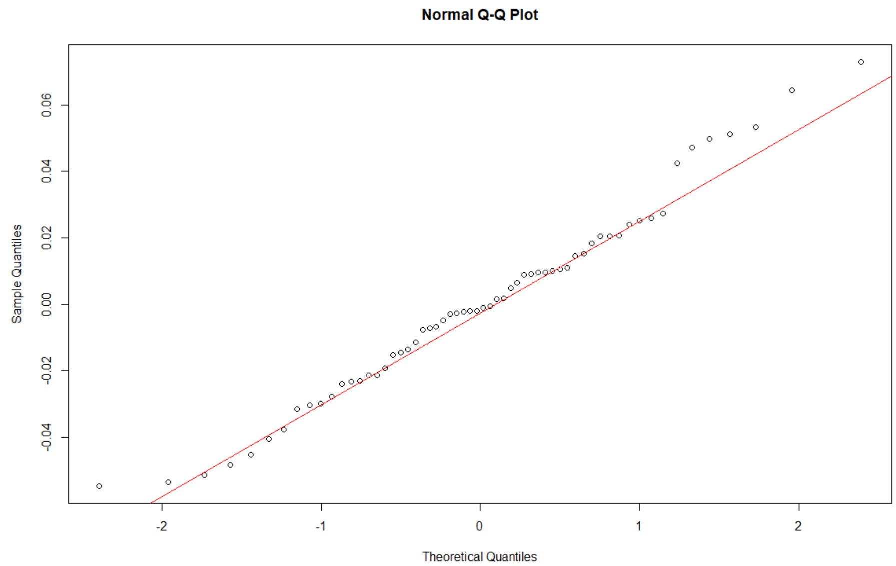
8.2.3 Test for Normality

Shapiro Wilk's Test

Next to check the normality of errors, we have performed the Shapiro Wilk's test on the residuals and the p-value came out to be $p - value = 0.4621 > 0.05$, So we conclude that the errors are also normally distributed.

QQ Plot

We have also obtained the QQ Plot of the residuals and the plot looks like the following



From the above QQ Plot we see that the graph matches nicely with 45^o diagonal line. So, we conclude that the residuals are from a normal distribution.

9 Conclusion

So, we have seen that both model (1) and model (3) fits the data very well. For both the models the model diagnostics measures are very satisfactory. Also, we have noted that the value of estimated standard error for model (3) is slightly less that that of model (1). This is due to model (3) has more number of parameters than model (1).

10 Appendix

The R codes which I have written to execute this project is given below.

```
    rm(list = ls())
library(Matrix)
A=read.csv("mydata.csv") # given data file is saved as mydata.csv
colnames(A)=c("x","y")
A
n=60
p=3
plot(A, main="Scatter plot of x(t) vs y(t)")
attach(A)
x
y


Q=function(beta){
  alpha2 = cov(exp(beta*x),y)/var(exp(beta*x))
  alpha1 = mean(y)-alpha2*mean(exp(beta*x))
  sum((y-alpha1-alpha2*exp(beta*x))^2)
}


beta=seq(-2,4, length=1000)
k=1
q=array(0)
for (i in beta) {
  q[k]=Q(i)
  k=k+1
}

plot(beta, q, main = "Plot the residual sum of squares as
            a function of beta", ylab = "Qstar(beta)",type="l")


beta = beta[which.min(q)]
beta
alpha2 = cov(exp(beta*x),y)/var(exp(beta*x))
```

```
alpha2
alpha1 = mean(y)-alpha2*mean(exp(beta*x))
alpha1
abline(v=beta)


l2norm=function(v){
  sqrt(sum(v^2))
}
theta0=c(alpha1,alpha2,beta) # original initial choice

#  theta0=c(1,0.5,0.6) #testing initial choice
theta=c(0,0,0)
F.= matrix(0,nrow=n,ncol=p)
F.[,1]=1

iter=0
while (l2norm(theta-theta0)>1e-6 ) {

F.[,2]=exp(theta0[3]*x)
F.[,3]=theta0[2]*x*exp(theta0[3]*x)
r=y-theta0[2]*exp(theta0[3]*x)-theta0[1]
delta= solve(t(F.)%*%F.)%*%t(F.)%*%matrix(r)

theta=theta0
theta0=theta0+delta

iter=iter+1
print(l2norm(theta-theta0))
}

theta
iter

v=theta[1]+theta[2]*exp(theta[3]*x)

plot(A, main="Graph showing observed and Fitted Responce")
lines(x,v,type="l")
```

```
F.[,2]=exp(theta[3]*x)
F.[,3]=theta[2]*x*exp(theta[3]*x)
F.
P=F.%*%solve(t(F.)%*%F.)%*%t(F.)
Y=matrix(y)
I=diag(c(rep(1,60)))
t(Y)%*%(I-P)%*%Y
t(F.)%*%F.
qf(1-0.05,3,n-3)
2.766438*0.05512061/19

v
sighs = sum((y-v)^2)/60 # sigma hat square
sighs

S11 = sighs*solve(t(F.)%*%F.)
S11
se = sqrt(diag(S11))

theta
se
CI=matrix(c(theta-qnorm(1-0.025)*se,theta+qnorm(1-0.025)*se)
                                        ,nrow = 3, byrow = F)
CI

2*sighs^2/n
r=y-v
plot(v,r,main="Residual Plot",xlab = "Fitted Responses", ylab = "Residuals")

shapiro.test(r)

qqnorm(r,main="QQPlot of Residuals")
qqline(r,col="red")

####Question 8
data=data.frame(x,y)
model1.nls=nls(y ~ a1 + a2*exp(b*x),data = data,
               start=list(a1 = 2.284847,a2 = 1.339692,b=1.489489),
               algorithm = "port")
```

```
# algorithm = "port" will use NL2SOL

summary(model1.nls)
s2=(0.0311^2)*57/60
s2
resi = residuals(model1.nls)
pred = predict(model1.nls)
plot(pred, resi, ylab = "Residuals", xlab = "Fitted Values",
                            main = "Residual plot using nls")


### Question 9
theta0=c(2,1,0.1,1)
theta=c(0,0,0,0)
F.= matrix(0,nrow=n,ncol=4)

iter=0
while (l2norm(theta-theta0)>1e-6 ) {

  F.[,1]=exp(theta0[3]*x)
  F.[,2]=exp(theta0[4]*x)
  F.[,3]=theta0[1]*x*exp(theta0[3]*x)
  F.[,4]=theta0[2]*x*exp(theta0[4]*x)

  r=y-theta0[1]*exp(theta0[3]*x)-theta0[2]*exp(theta0[4]*x)
  delta= solve(t(F.)%*%F.)%*%t(F.)%*%matrix(r)

  theta=theta0
  theta0=theta0+delta

  iter=iter+1
  print(l2norm(theta-theta0))
}

theta
iter


data=data.frame(x,y)
```

```
model2.nls=nls(y ~ a1*exp(b1*x) + a2*exp(b2*x),data = data,
            start=list(a1 = 1,a2 = 1,b1 = 0.1,b2 = 1))
summary(model2.nls)

model3.nls=nls(y ~ a1*exp(b1*x) + a2*exp(b2*x),data = data,
               start=list(a1 = 1,a2 = 1,b1 = 0.1,b2 = 1),algorithm = "port" )
summary(model3.nls)

plot(A, main="Graph showing observed and Fitted Responce")
v=theta[1]*exp(theta[3]*x)+theta[2]*exp(theta[4]*x)
lines(x,v,type="l",col="red")

sum((y-v)^2)/60 # sigma hat square

r=y-v
plot(v,r,main="Residual Plot",xlab = "Fitted Responses", ylab = "Residuals")
shapiro.test(r)
qqnorm(r)
qqline(r,col="red")
```

11 References

1. Class notes of Dr. Debasis Kundu of the course MTH686A: Non-Linear Regression.
2. G.A.F. Seber and C.J. Wild (1989), Nonlinear Regression, Wiley Interscience.

12 Acknowledgement

Project is like a bridge between the theoretical learning and practical working. With this will, I started this project. I am feeling obliged in taking the opportunity to express my sincere thanks and gratitude to our professor **Dr. Debasis Kundu** who gave me the golden opportunity to execute this Non Linear Regression Project and also helped to complete it.