

Homework 2

David L. Olson

February 23, 2025

Contents

1 Problem 1 2

1.1 Solution 2

1.1.1 Part A 2

1.1.2 Part B 3

1.1.3 Part C 4

1.1.4 Part D 5

1.1.5 Part E 6

1.1.6 Part F 7

1.1.7 Part G 8

1.1.8 Part H 9

1.1.9 Part I 10

2 Problem 2 11

2.1 Solution 11

1 Problem 1

Exercise 1 in Section 2.7

1.1 Solution

1.1.1 Part A

The model

$$t_i = t_0 + s_2 x_i$$

can be approached as a discrete linear inverse problem. Six measurements of this system were provided which lead to casting the inverse problem in the form $G\mathbf{m} = \mathbf{d}$. The system matrix (i.e., operator) is defined as

$$G := \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_6 \end{bmatrix}$$

and data vector simply defined below.

$$\mathbf{d} := \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{bmatrix}$$

Because the standard deviation was provided for the measurement error, a weighted least-squares solution was utilized. Weighting was applied to the operator matrix G and data vector \mathbf{d} such that:

$$W = \frac{1}{\sigma} I_6$$

$$G_w = WG$$

$$\mathbf{d}_w = W\mathbf{d}$$

Then, the weighted least-squares solution given as:

$$\mathbf{m}_{L2} = (G_w^T G_w)^{-1} G_w^T \mathbf{d}_w$$

Using MATLAB[®], the solution \mathbf{m}_{L2} was computed to be:

$$\mathbf{m}_{L2} = \begin{bmatrix} t_0 \\ s_2 \end{bmatrix} = \begin{bmatrix} 2.032337 \\ 0.220281 \end{bmatrix}$$

Figure 1 contains the data, the fitted model, and the residuals.

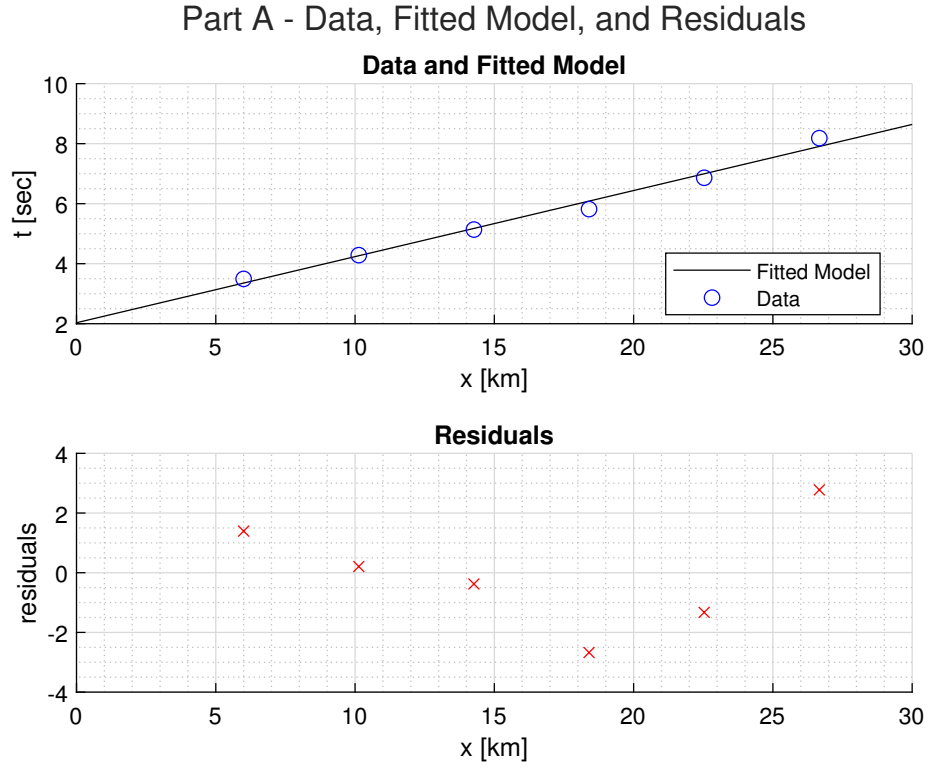


Figure 1: Part A - Data, Fitted Model, and Residuals

1.1.2 Part B

The model covariance matrix $C = \text{Cov}(\mathbf{m}_{L2})$ is computed as

$$A = (G_w^T G)^{-1} G_w^T$$

$$C = A A^T$$

Using MATLAB[®], the model covariance matrix was computed to be

$$C = \begin{bmatrix} 0.01059 & -0.0005463 \\ -0.0005463 & 3.3447 \times 10^{-5} \end{bmatrix}$$

The model parameter correlation matrix ρ is essentially a scaled version of the model covariance matrix such that

$$\rho_{m_i, m_j} = \frac{\text{Cov}(m_i, m_j)}{\sqrt{\text{Var}(m_i) \text{Var}(m_j)}}$$

Using MATLAB[®], the model parameter correlation matrix was computed to be

$$\rho = \begin{bmatrix} 1 & -0.91794 \\ -0.91794 & 1 \end{bmatrix}$$

The off-diagonal terms indicate that the two model parameters have a high negative correlation!

1.1.3 Part C

The 95% confidence interval for the model parameters \mathbf{m}_{L2} given as

$$\mathbf{m}_{L2} \pm 1.96 \cdot \text{diag}(C)^{\frac{1}{2}} = \mathbf{m}_{L2} \pm \begin{bmatrix} 0.201696 \\ 0.011335 \end{bmatrix}$$

which results in the following 95% confidence interval which could be interpreted as a box.

$$1.8306 \leq t_0 \leq 2.234$$

$$0.20895 \leq s_2 \leq 0.23162$$

However, due to the high correlation between model parameters, interpreting the confidence as a box is insufficient. Instead, the eigenvectors and eigenvalues of C^{-1} are used to form an ellipse which bounds the 95% confidence interval. The eigenvectors indicate the directions of the semi-major and semi-minor axes, while each axis is scaled by a factor of $\frac{\Delta}{\sqrt{\lambda_i}}$. The region Δ^2 is the inverse χ^2 distribution for a given number of degrees of freedom ν . The degrees of freedom ν is the number of equations in G (i.e., m) minus the number of model parameters n . In our case, the number of degrees of freedom is $\nu = m - n = 4$.

Figure 2 shows the comparison between the box interpretation and the ellipse.

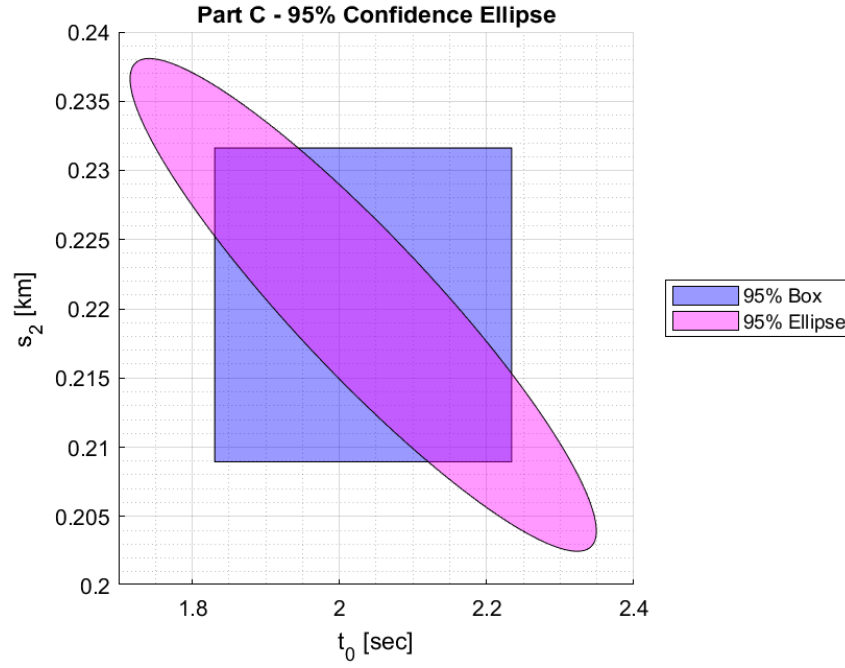


Figure 2: Part C - 95% Confidence Bounds

1.1.4 Part D

The observed χ_{obs}^2 value is defined as the sum of the residuals when scaled by σ such that:

$$\chi_{obs}^2 := \sum_{i=1}^m \frac{(d_i - (G\mathbf{m}_{L2})_i)^2}{\sigma_i^2}$$

Then, the p-value is defined as:

$$p := \int_{\chi_{obs}^2}^{\inf} f_{\chi^2} x dx$$

Using MATLAB[®], each value was evaluated as:

$$\chi_{obs}^2 = 18.750184$$

$$p = 0.000874$$

1.1.5 Part E

A Monte-Carlo simulation of 1000 runs was conducted. Figures 3 and 4 show histograms for model parameters t_0 and s_2 , as well as the histogram for χ^2 .

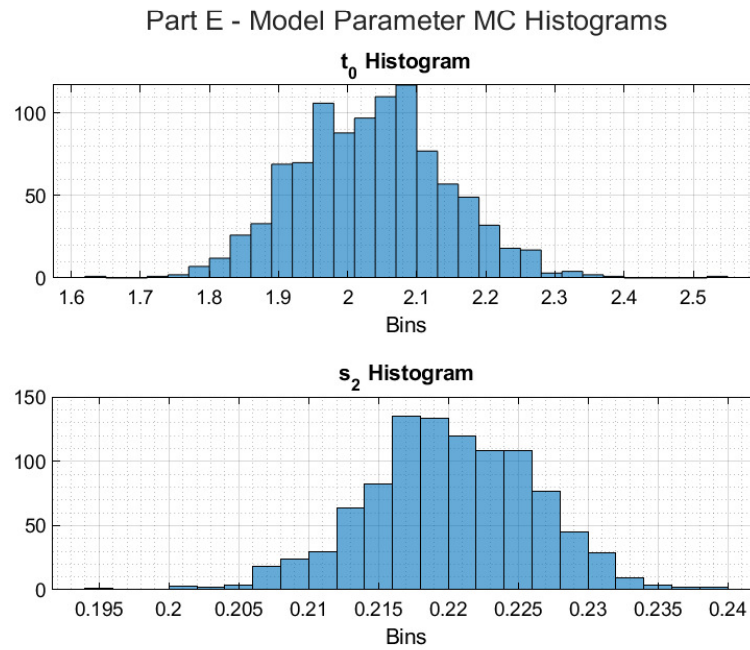
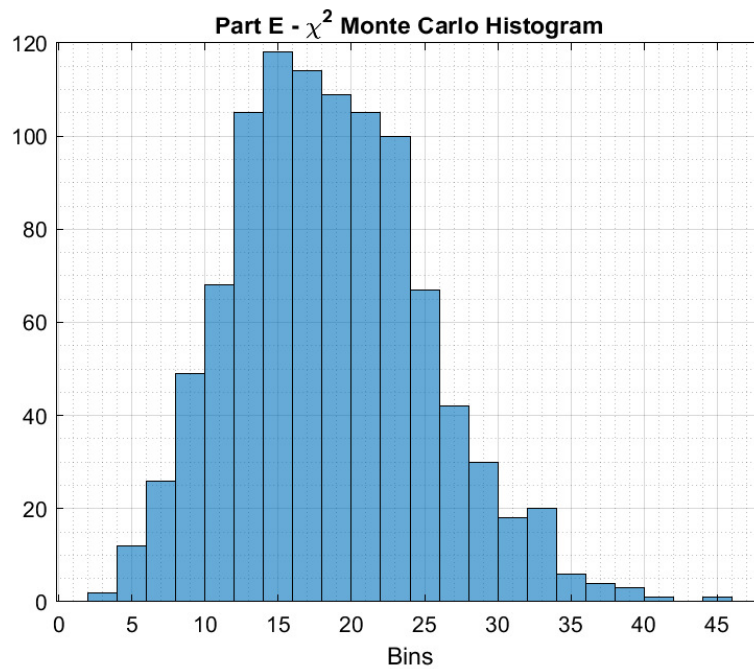


Figure 3: Part E - Model Parameter Histograms

Figure 4: Part E - χ^2 Histogram

As expected, the model parameter histograms in figure 3 appear to be normally distributed, while the χ^2 histogram in figure 4 appears to keep its expected shape with a long tail to the right.

1.1.6 Part F

First, the Monte-Carlo realizations for the model parameters are consistent with the 95% confidence ellipse from Part C.

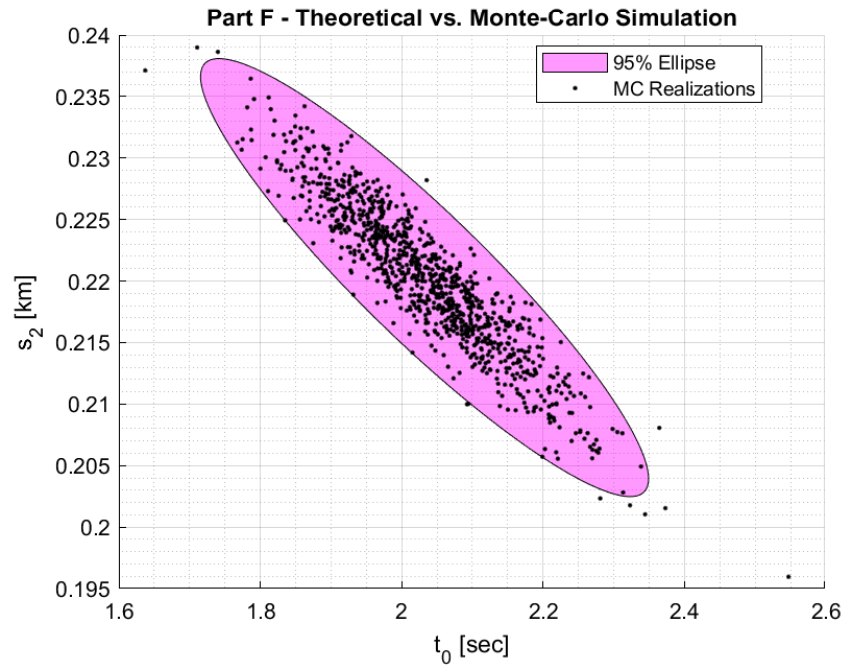


Figure 5: Part F - Model Parameter MC Realization vs. 95% Confidence Ellipse

Additionally, there is agreement between the theoretical means of computing χ_{obs}^2 vs. the histogram in figure 4. Notice that the mode of the histogram (i.e., bin with the greatest number of occurrences) is near theoretical value of $\chi_{obs}^2 = 18.750184$.

I am surprised that with 1000 runs, the two modes are not a closer match. However, I repeated the Monte-Carlo simulation with 100000 runs and got much better agreement. Since I re-seed the random number generator each time the script is called, perhaps I was just "unlucky" and got an unfavorable stochastic run?

Finally, the computed p-value indicates a strong "goodness of fit" measure which is consistent with figure 5 which shows that the majority of points are contained within the ellipse.

1.1.7 Part G

L1 regression using iterative reweighted least squares (IRLS) was performed using the same data set. The `irls()` function provided in the `Lib` folder was utilized. This resulted in the following comparison to the L2 regression.

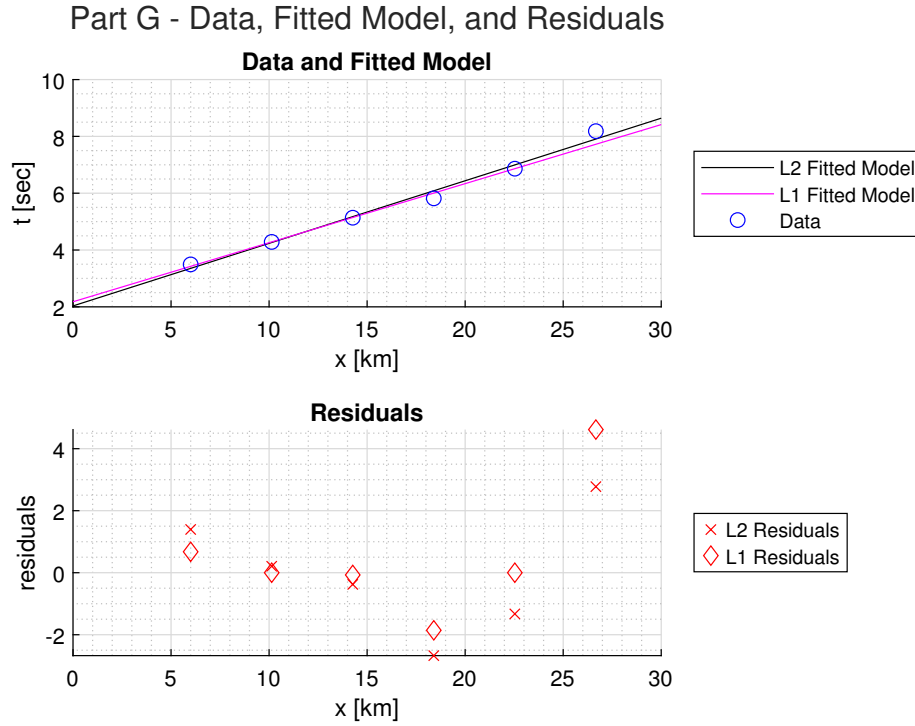


Figure 6: Part G - L1 Regression Data, Fitted Model, and Residuals

1.1.8 Part H

The model covariance resulting from L1 regression was estimated empirically from the 1000 Monte-Carlo trials. With the number of trials q , the empirical covariance matrix C_{L1} is computed as

$$A_{L1} = \mathbf{m}_{L1,mc} - \bar{\mathbf{m}}_{L1,mc}$$

$$C_{L1} = \frac{A_{L1}^T A_{L1}}{q}$$

Using MATLAB®, C_{L1} evaluates to

$$C_{L1} = \begin{bmatrix} 0.047204 & -0.0019812 \\ -0.0019812 & 9.2902 \times 10^{-5} \end{bmatrix}$$

Interestingly enough, the model parameter correlation matrix showed a slightly higher correlation than the L2 regression solution.

$$\rho_{L1} = \begin{bmatrix} 1 & -0.9461 \\ -0.9461 & 1 \end{bmatrix}$$

When comparing the Monte-Carlo realizations with the resulting 95% confidence ellipse, there are a few values which fall far outside the ellipse.

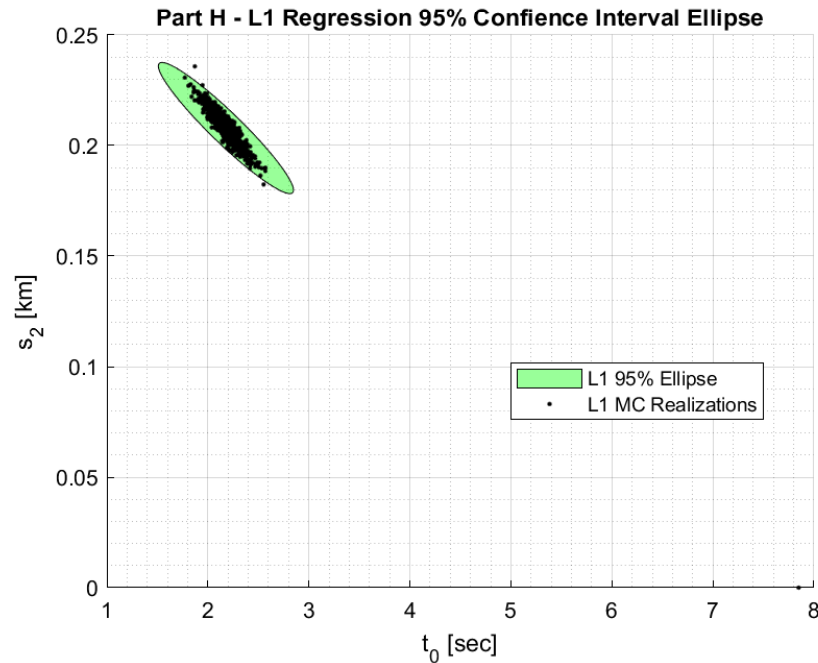


Figure 7: Part H - L1 Monte-Carlo Realizations vs. 95% Confidence Ellipse

Notice that one Monte-Carlo realization sits in the far right-bottom corner.

1.1.9 Part I

Recall in figure 6 that the L1 regression residuals were all less than the L2 regression residuals for the first 5 data points, but not the 6th. This implies that the L1 regression assigned "greater importance" to the first 5 data points, which signifies that the 6th data point was in fact an outlier.

2 Problem 2

Exercise 5 in Section 2.7

2.1 Solution

Consider a polynomial with degree 19 containing 20 terms such that

$$y_i = a_0 + a_1x_i + a_2x_i^2 + \dots + a_{19}x_i^{19}$$

where elements $x_i \in \mathbf{x}$ and $y_i \in \mathbf{y}$ are sampled from -1 to 1 in steps of 0.1 . This problem can be formulated as a discrete linear inverse problem of the form $G\mathbf{m} = \mathbf{d}$. The operator G of size 21×20 is formed as

$$G = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{19} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{19} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{21} & x_{21}^2 & \cdots & x_{21}^{19} \end{bmatrix}$$

and model parameters $\mathbf{m} \in \mathbb{R}^{20}$ are the polynomial coefficients. The normal equations are applied as

$$\mathbf{m}_{L2} = (G^T G)^{-1} G^T \mathbf{y}$$

which should ideally result in $a_1 = 1$ and all other coefficients equal zero. Evaluating the normal equations in MATLAB[®] reveal model parameters that are close but not exact to ideal as shown in figure 8.

```
L2 Regression Model Parameters
-6.5730e-11
 1.0000e+00
 1.2503e-08
 4.5346e-06
-4.6706e-07
-1.1948e-04
 6.6599e-06
 1.3954e-03
-4.6491e-05
-8.6041e-03
 1.7814e-04
 3.0378e-02
-3.9212e-04
-6.3207e-02
 4.9233e-04
 7.6254e-02
-3.2648e-04
-4.9121e-02
 8.8419e-05
 1.3021e-02
```

Figure 8: Model Parameters

While a_1 equals one as expected, all other terms are *close to zero, but not exactly zero*. 20 coefficients are far too many for fitting a line, but the result is interesting. Figure 9 shows the residuals in the L2 regression which take an interesting shape.

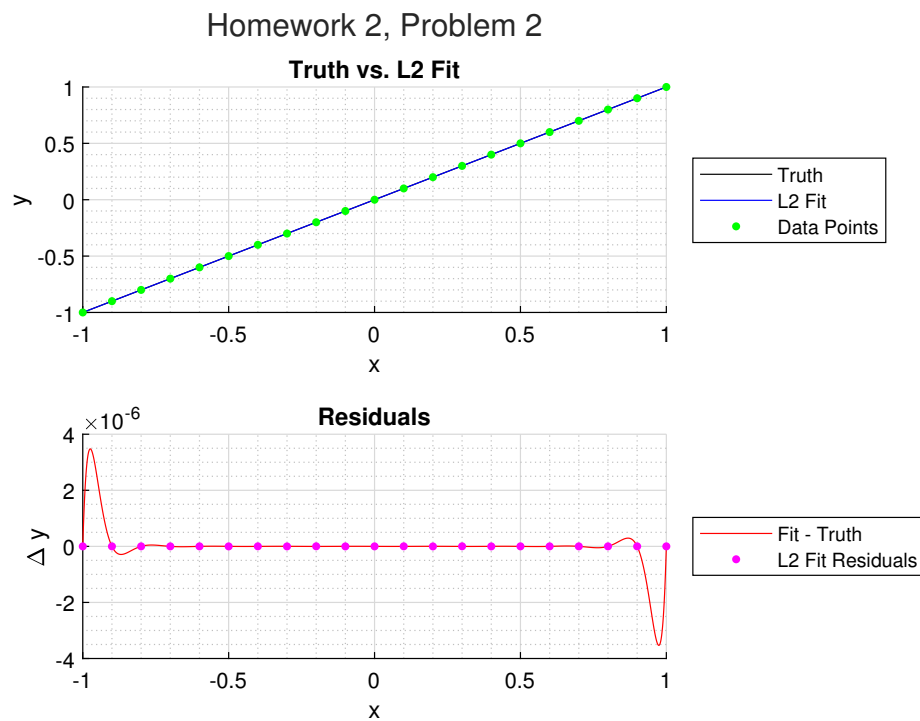


Figure 9: Data, Fitted Model, and Residuals

This is an extreme example of over-fitting the data. By simple observation of the data, there are 18 extra unnecessary parameters to fit this linear trend.