MTH686 Project

Kartik Diwan - 220504

Introduction

This report presents three models for predicting $y(t)$ based on various parameterizations and optimization techniques. Each model uses different approaches for estimating coefficients to fit the data.

Model 1

Define the parameter vectors:

$$\alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}_{3\times 1}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

The matrix $X(\beta)$ is structured as:

$$X(\beta) = \begin{bmatrix} 1 & e^{\beta_1 t_1} & e^{\beta_2 t_1} \\ 1 & e^{\beta_1 t_2} & e^{\beta_2 t_2} \\ \vdots & \vdots & \vdots \\ 1 & e^{\beta_1 t_n} & e^{\beta_2 t_n} \end{bmatrix}_{n\times 3}$$

The prediction equation becomes:

$$y_{\text{pred}} = X\alpha$$

We define the objective function:

$$Q(\alpha, \beta) = (Y - X\alpha)^T (Y - X\alpha)$$

For fixed β_1, β_2, this problem reduces to linear regression in terms of α:

$$\Rightarrow \hat{\alpha} = \left(X^T(\beta)X(\beta)\right)^{-1} X^T(\beta)Y$$

We have reduced the problem to a 2D optimization problem with respect to β_1 and β_2. The objective function in terms of β is:

$$Q(\beta) = \left(Y - X(\beta)\left(X^T(\beta)X(\beta)\right)^{-1} X^T(\beta)Y\right)^T \left(Y - X(\beta)\left(X^T(\beta)X(\beta)\right)^{-1} X^T(\beta)Y\right)$$

This reduces the problem to a 2D optimization over β_1 and β_2, for which we use a grid search:

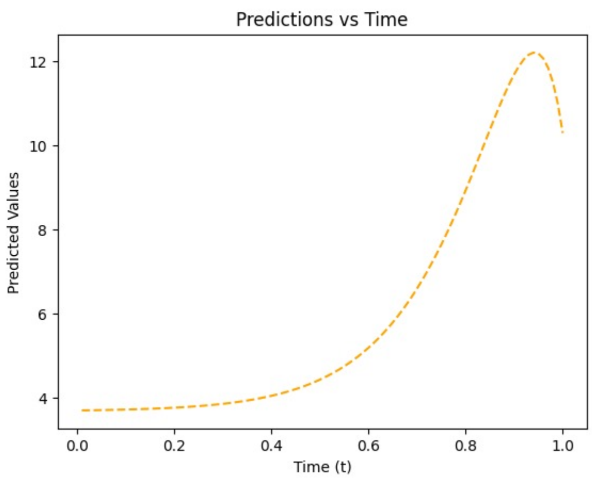
$$\beta_1 = \{-50, -49.999, \ldots, 50\}$$

$$\beta_2 = \{-50, -49.999, \ldots, 50\}$$

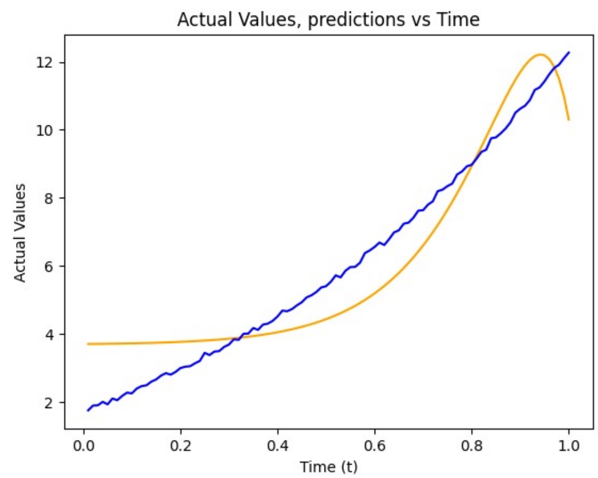
Estimated parameters:

$$\hat{\beta} = \begin{bmatrix} 11.999 \\ 7.529 \end{bmatrix}, \quad \hat{\alpha} = \begin{bmatrix} 3.678 \\ -1.77 \times 10^{-4} \\ 1.90 \times 10^{-2} \end{bmatrix}$$

The least squares error for Model 1 is 99.9849699867092.



(a) Prediction vs Time



(b) Prediction and Actual vs Time

Figure 1: Model 1 Visualizations

Model 2

The model is defined as:

$$y(t) = \frac{\alpha_0 + \alpha_1 t}{\beta_0 + \beta_1 t} + \epsilon(t)$$

Using gradient descent, we estimate the parameters α_0, α_1, β_0, and β_1 to minimize:

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \left(y - \frac{\alpha_0 + \alpha_1 t}{\beta_0 + \beta_1 t} \right)^2 = \sum_{i=1}^{n} e_i$$

The partial derivatives for each parameter are:

$$\text{grad}_{\alpha_0} = \frac{\partial Q}{\partial \alpha_0}, \quad \text{grad}_{\beta_0} = \frac{\partial Q}{\partial \beta_0}$$

$$\text{grad}_{\alpha_1} = \frac{\partial Q}{\partial \alpha_1}, \quad \text{grad}_{\beta_1} = \frac{\partial Q}{\partial \beta_1}$$

Starting values for all parameters are initialized at 1.

Gradient Descent

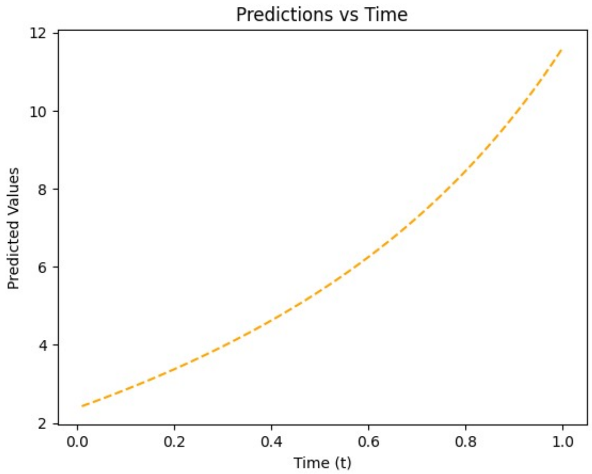
For i in $\{1, \ldots, 10000\}$:

$$\alpha_0^{(i)} = \alpha_0^{(i-1)} - \text{learning rate} \times \text{grad}_{\alpha_0}^{(i-1)}$$
$$\alpha_1^{(i)} = \alpha_1^{(i-1)} - \text{learning rate} \times \text{grad}_{\alpha_1}^{(i-1)}$$
$$\beta_0^{(i)} = \beta_0^{(i-1)} - \text{learning rate} \times \text{grad}_{\beta_0}^{(i-1)}$$
$$\beta_1^{(i)} = \beta_1^{(i-1)} - \text{learning rate} \times \text{grad}_{\beta_1}^{(i-1)}$$

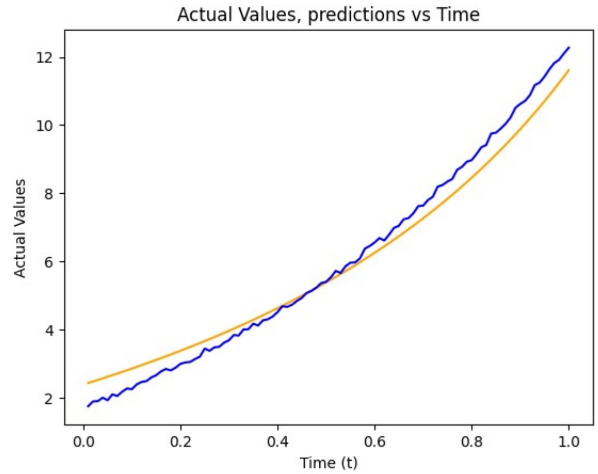
Estimated parameters:

$$\hat{\alpha}_0 = 1.687, \quad \hat{\alpha}_1 = 2.26, \quad \hat{\beta}_0 = 0.70, \quad \hat{\beta}_1 = -0.367$$

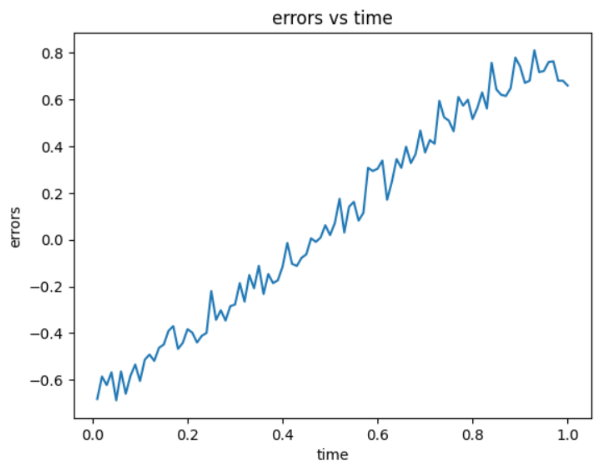
The least squares error for Model 2 is 21.198341135704933.



(a) Prediction vs Time



(b) Prediction and Actual vs Time



(c) Error vs Time

Figure 2: Model 2 Visualizations

Model 3

The polynomial model is given by:

$$y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \varepsilon(t)$$

The objective function for least squares is:

$$Q(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = \sum_{i=1}^{m} \left(y(t_i) - \left(\beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \beta_4 t_i^4 \right) \right)^2$$

Define the vector of coefficients:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

The vector Y is:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

The matrix X is:

$$X = \begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 & t_1^4 \\ 1 & t_2 & t_2^2 & t_2^3 & t_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 & t_m^3 & t_m^4 \end{pmatrix}$$

The model becomes:

$$Y = X\beta + \varepsilon$$

The least squares estimate for $\hat{\beta}$ is:

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

Estimated parameters:

$$\hat{\beta}_0 = 1.706, \quad \hat{\beta}_1 = 5.899, \quad \hat{\beta}_2 = 1.277, \quad \hat{\beta}_3 = 3.91, \quad \hat{\beta}_4 = -0.49$$

The least squares error for Model 3 is 0.2684896837690878.

Based on the least squares error for each model, we conclude that Model 3 provides the best fit, with a least squares error of 0.2684896837690878 with a variance of 0.002684896837690879.

Normality Test

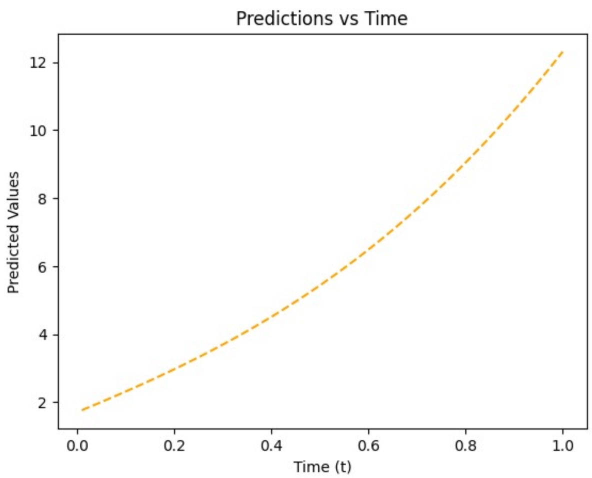
In the Shapiro-Wilk test, the p-value is 0.0258. Since $p < 0.05$, we reject the null hypothesis H_0 that the errors are normally distributed, and conclude that the errors are not normally distributed.

Fisher Information Matrix

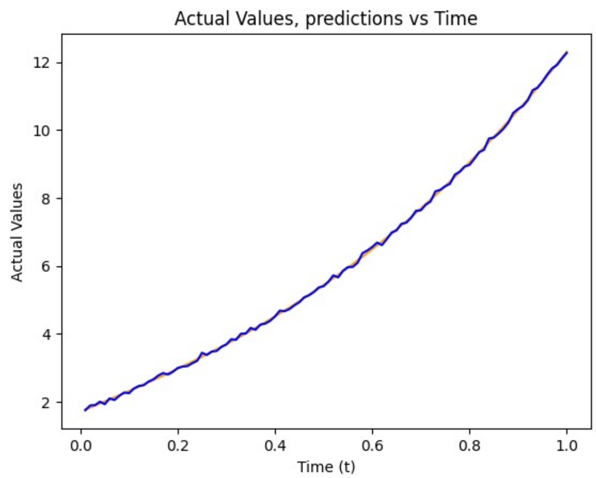
$$\text{Fisher Information Matrix} = \frac{1}{\text{variance}} X^T X$$

Using this, the associated confidence interval is as follows:

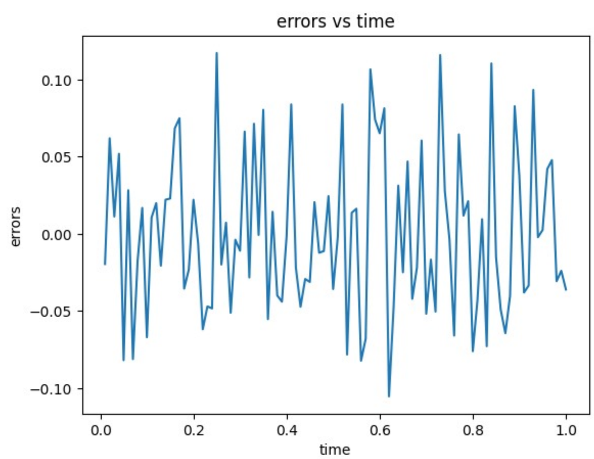
Parameters	**Lower Bound**	**Upper Bound**
β_0	1.652	1.76
β_1	5.167	6.63
β_2	-1.652	4.207
β_3	-0.934	8.261
β_4	-2.625	1.645



(a) Prediction vs Time



(b) Prediction and Actual vs Time



(c) Error vs Time

Figure 3: Model 3 Visualizations