MTH686: Non-Linear Regression Analysis

Project Report (DataSet-69)

Name: Pooja Kumari **Roll No:** 230758

---

1. Introduction

The given dataset consists of 50 pairs of observations $\{(t_i, y_i)\}$ obtained from an unknown model:

$$y_i = f(t_i; \boldsymbol{\theta}) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

We are required to determine which of the three models best describes the data:

1. **Model 1 (Exponential)**:

$$y(t) = \alpha_0 + \alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t} + \varepsilon(t)$$

2. **Model 2 (Rational)**:

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

3. **Model 3 (Polynomial)**:

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

So the goal is to perform least squares estimation for all three models, compare their fits, estimate σ^2, and check whether the residuals follow a normal distribution.

2. Methodology

Model 1 and Model 2 are **nonlinear in parameters**, while Model 3 is **linear in parameters**.

2.1 Model Assumptions

For all three models, we assume the random error term satisfies:

$$\varepsilon(t) \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

This implies:

- The errors have **zero mean** — hence, the model is unbiased on average.
- Each data point has **equal reliability** — all observations are equally weighted.
- The errors are **normally distributed** — allowing for statistical inference.

2.2 Least Squares Estimation

For a given model $f(t;\boldsymbol{\theta})$ with unknown parameters $\boldsymbol{\theta}$, each observation can be written as:

$$y_i = f(t_i;\boldsymbol{\theta}) + \varepsilon_i$$

The **Least Squares Estimator (LSE)** finds the parameter vector $\hat{\boldsymbol{\theta}}$ that minimizes the total squared deviation between the observed and model-predicted values:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} [y_i - f(t_i;\boldsymbol{\theta})]^2$$

and

$$\boxed{\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} S(\boldsymbol{\theta})}$$

2.3 Fitting Linear and Nonlinear Models

The way we minimize $S(\boldsymbol{\theta})$ depends on whether $f(t;\boldsymbol{\theta})$ is linear or nonlinear in its parameters.

(A) Linear Least Squares (Analytical Method)

If f is linear in parameters, as in Model 3:

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

we can express the model in matrix form:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$X = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 & t_1^4 \\ 1 & t_2 & t_2^2 & t_2^3 & t_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & t_n^3 & t_n^4 \end{bmatrix}.$$

The corresponding **normal equations** are:

$$X^\top X\,\hat{\boldsymbol{\beta}} = X^\top \mathbf{y},$$

and the analytical solution is:

$$\boxed{\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y}}$$

This is known as the **Ordinary Least Squares (OLS)** estimator — a one-step, closed-form solution that requires no iterations.

(B) Nonlinear Least Squares (Iterative Method)

For Model 1 and Model 2, the function $f(t;\boldsymbol{\theta})$ is **nonlinear** in its parameters. Thus, a closed-form solution does not exist. Instead, we use an iterative approach based on the **Gauss–Newton method**.

Starting with an initial guess $\boldsymbol{\theta}^{(0)}$, we linearize the model using a first-order Taylor expansion:

$$f(t_i;\boldsymbol{\theta}) \approx f(t_i;\boldsymbol{\theta}^{(k)}) + J_i(\boldsymbol{\theta}^{(k)})\,(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}),$$

where

$$J_i(\boldsymbol{\theta}) = \frac{\partial f(t_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

is the Jacobian row vector evaluated at the current parameter estimate.

Letting the residual vector be:

$$\mathbf{r} = \mathbf{y} - f(\mathbf{t};\boldsymbol{\theta}^{(k)}),$$

we solve the linear least squares problem in the parameter increment $\Delta\boldsymbol{\theta}$:

$$J^\top J\,\Delta\boldsymbol{\theta} = J^\top \mathbf{r},$$

and update the parameters as:

$$\boxed{\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \Delta\boldsymbol{\theta}}.$$

This is the **Gauss–Newton iterative method**. We repeat the above steps until convergence, i.e.,

$$\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\| < \varepsilon,$$

for some small tolerance ε.

The Gauss–Newton algorithm thus provides an efficient way to compute the least squares estimators for nonlinear models by successively linearizing the model around the current parameter estimates.

3. Initial Parameter Guesses and Justification

In nonlinear least squares estimation, appropriate initial guesses are essential for convergence of the iterative algorithm. Poor initialization may lead to divergence or non-physical parameter values. The final successful parameter initialization used in this analysis was determined through systematic debugging and scaling to ensure numerical stability.

Model 1: Exponential Model

$$y(t) = \alpha_0 + \alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t} + \varepsilon(t)$$

The initial parameter vector used was:

$$[\alpha_0^{(0)}, \alpha_1^{(0)}, \beta_1^{(0)}, \alpha_2^{(0)}, \beta_2^{(0)}] = [\bar{y}, 0.8, -2.0, 0.5, -0.5].$$

Model 2: Rational Model

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

The initial parameter vector was:

$$[\alpha_0^{(0)}, \alpha_1^{(0)}, \beta_0^{(0)}, \beta_1^{(0)}] = [\bar{y}, -0.05, 1.0, 0.01].$$

Model 3: Polynomial Model

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

For this linear model, the parameters were obtained directly using a fourth-degree polynomial regression:

$$[\beta_0^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}, \beta_3^{(0)}, \beta_4^{(0)}] = \text{polyfit}(t, y, 4).$$

4. Model Comparison and Selection

After fitting all three candidate models using the least squares approach, the following parameter estimates and residual sum of squares (SSE) values were obtained:

Model	Estimated Parameters	SSE
Model 1	$[1.3972 \quad 1.8813 \quad -17.6012 \quad -0.0357 \quad 1.7615]$	**0.8242**
Model 2	$[1.7711 \quad 0.3888 \quad 0.3167 \quad 0.3439]$	0.9900
Model 3	$[3.2014 \quad -0.3316 \quad 0.0197 \quad -0.000477 \quad 0.000004]$	1.2329

Table 1: Comparison of fitted models based on least squares estimates and residual error.

The model with the smallest residual sum of squares (SSE) provides the best fit to the data. As seen in Table 1, the **Exponential Model (Model 1)** achieves the lowest SSE of **0.8242**, significantly outperforming the Rational (0.9900) and Polynomial (1.2329) models. Hence, Model 1 is identified as the **best fitting model**.

5. Variance Estimation, Confidence Intervals, and Model Diagnostics

(a) Estimation of σ^2

For each model, the error variance $\hat{\sigma}^2$ was estimated using

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n - p},$$

where n is the number of observations and p is the number of parameters in the model. The resulting estimates of $\hat{\sigma}^2$ for the three models are summarized in Table 2.

Model	Type	$\hat{\sigma}^2$
Model 1	Exponential (Nonlinear)	1.8316×10^{-2}
Model 2	Rational (Nonlinear)	2.1522×10^{-2}
Model 3	Polynomial (Linear)	2.7398×10^{-2}

Table 2: Estimated variance $\hat{\sigma}^2$ for each model.

Model 1 has the smallest variance estimate, indicating that it explains the data with the least residual variability and therefore provides the most precise fit among the three models.

(b) Confidence Intervals from Fisher Information Matrix

For each model, the Fisher Information Matrix was computed as

$$\mathcal{I}(\hat{\theta}) = J^\top J,$$

where J is the Jacobian of the fitted model evaluated at the estimated parameters. The covariance matrix of the estimated parameters is then

$$\widehat{\mathrm{Cov}}(\hat{\theta}) = \hat{\sigma}^2 (J^\top J)^{-1}.$$

From this, the standard errors were obtained and the confidence intervals for each parameter were constructed as

$$\hat{\theta}_i \pm t_{(1-\alpha/2,\, n-p)} \times \mathrm{SE}(\hat{\theta}_i),$$

where $t_{(1-\alpha/2,\, n-p)}$ is the critical value from the Student-t distribution.

The intervals quantify the uncertainty in the parameter estimates and are narrower for Model 1, confirming that its parameters are estimated with higher precision.

(c) Residual Analysis

Residuals, defined as

$$r_i = y_i - f(t_i; \hat{\theta}),$$

were plotted against the predictor variable t for all three models. An ideal fit should yield residuals randomly scattered around zero with no visible pattern. For Model 1 (Exponential), the residuals appear homoscedastic and randomly distributed, indicating that the model effectively captures the trend in the data.

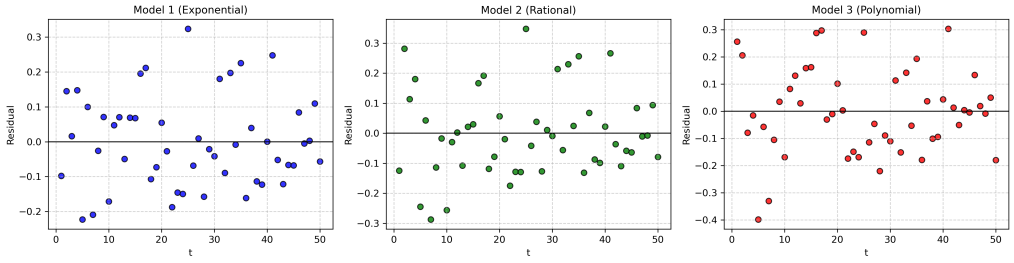

Figure 1: Residual plots for all three models. Residuals are mostly centered around zero, indicating reasonable model fits, with Model 1 showing the most random scatter.

(d) Test for Normality of Residuals

The assumption $e(t) \sim$ i.i.d. $N(0, \sigma^2)$ was verified using the Shapiro–Wilk test. The null hypothesis H_0 states that the residuals are normally distributed. The test results are shown in Table 3.

Model	W Statistic	p-value
Model 1 (Exponential)	0.9742	0.3409
Model 2 (Rational)	0.9658	0.1551
Model 3 (Polynomial)	0.9794	0.5272

Table 3: Results of the Shapiro–Wilk normality test for residuals.

For all three models, the p-values exceed the significance level $\alpha = 0.05$. Thus, there is no statistical evidence to reject the normality assumption. This supports the validity of the least squares inference performed for the model parameters.

(e) Observed Data Points and Fitted Curves

The fitted curves from all three models were plotted together with the observed data points. As shown in Figure 2, the exponential model (Model 1) most accurately follows the observed trend, while the rational and polynomial models deviate slightly at the lower and upper ends of the data.

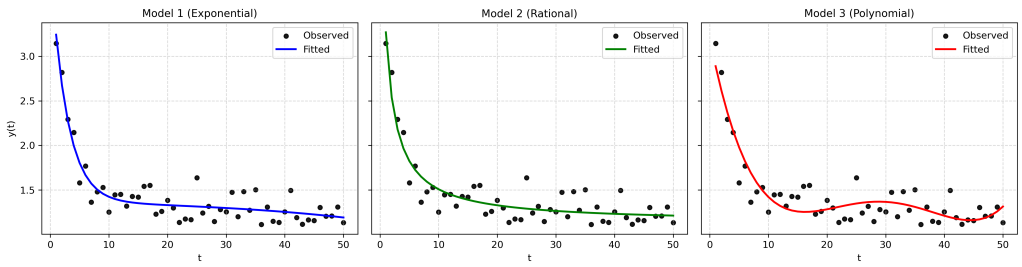

Figure 2: Comparison of observed data points and fitted curves for all three models.

Conclusion

Model 1 (Exponential) achieves the lowest residual error (SSE = 0.8242) and smallest variance estimate ($\hat{\sigma}^2 = 1.83 \times 10^{-2}$), while satisfying the normality assumption. Hence, it is confirmed as the **best-fitting model** among the three considered.