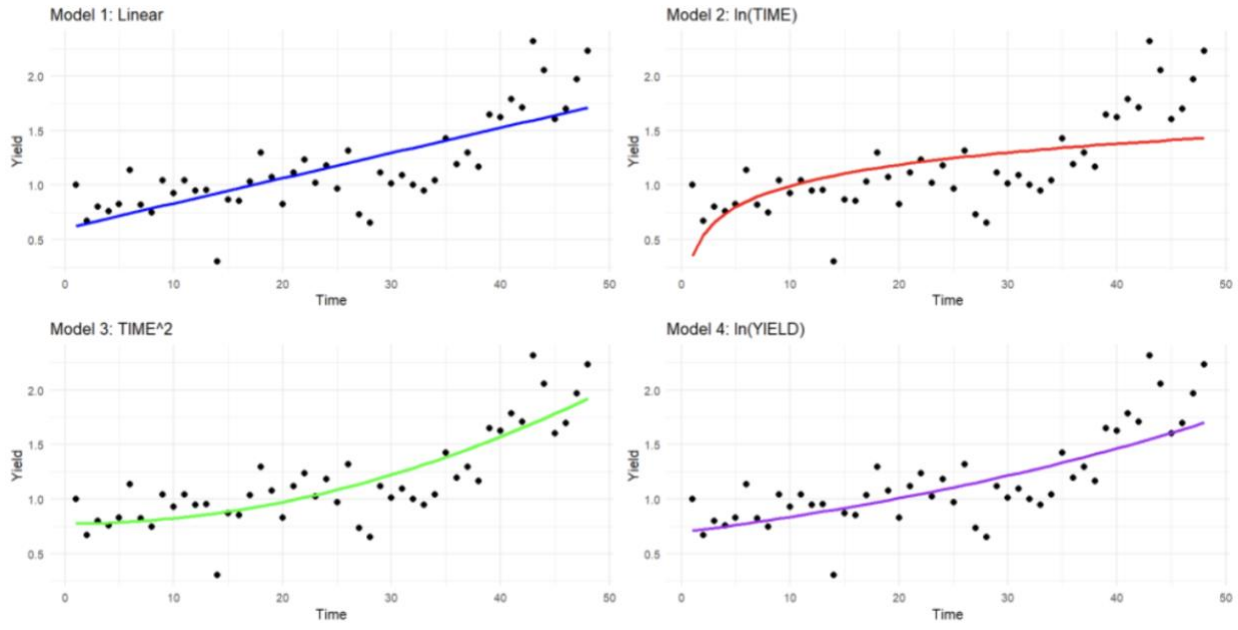
Name: Nguyen Quoc Nhan

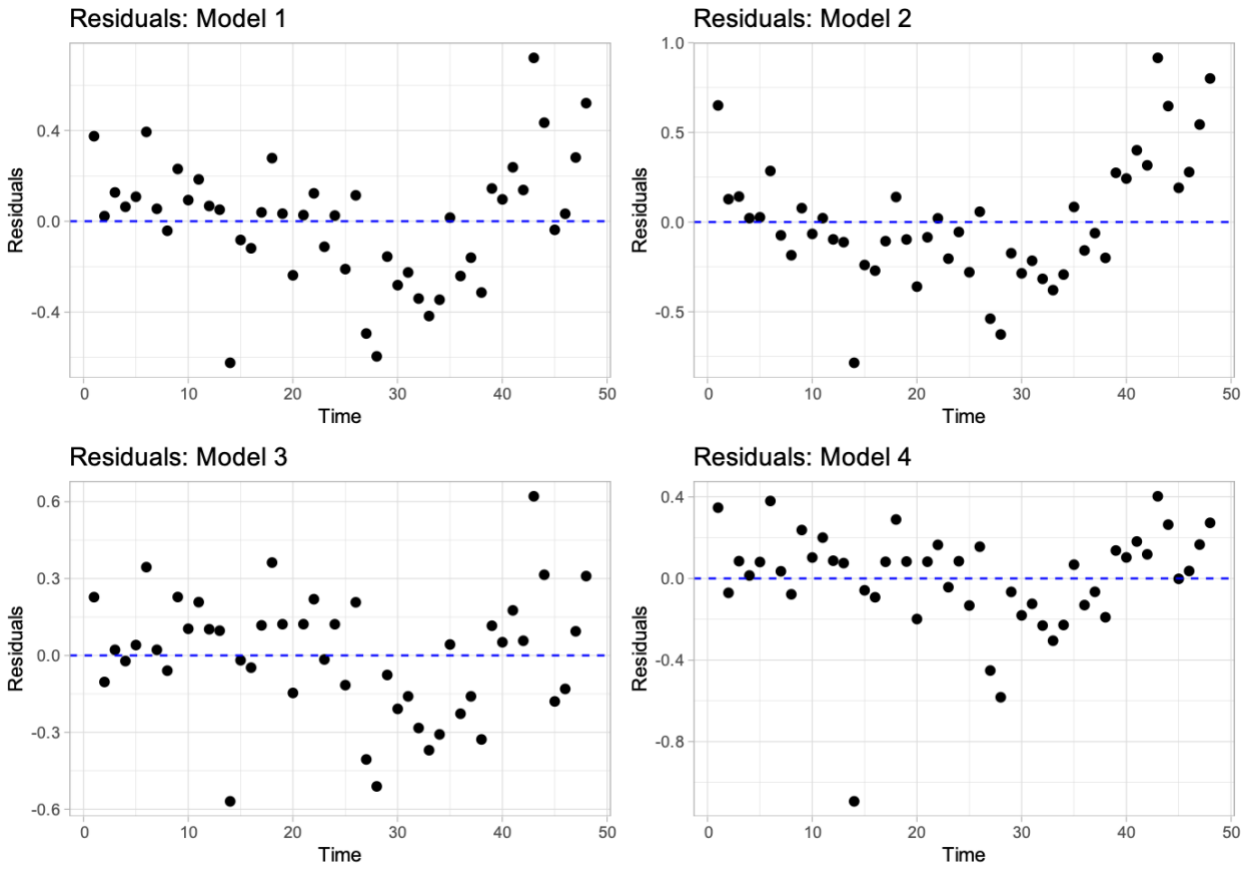
ID:413707009

HW0317

Question 28.

a&b






```
> cat("Shapiro-Wilk p-values for residuals:\n")
Shapiro-Wilk p-values for residuals:
> cat("Model 1:", shapiro1$p.value, "\n")
Model 1: 0.6792056
> cat("Model 2:", shapiro2$p.value, "\n")
Model 2: 0.1855502
> cat("Model 3:", shapiro3$p.value, "\n")
Model 3: 0.826645
> cat("Model 4:", shapiro4$p.value, "\n\n")
Model 4: 7.205319e-05
```

```
> cat("Adjusted R-squared values:\n")
Adjusted R-squared values:
> cat("Model 1:", adj_R2[1], "\n")
Model 1: 0.5686594
> cat("Model 2:", adj_R2[2], "\n")
Model 2: 0.3241945
> cat("Model 3:", adj_R2[3], "\n")
Model 3: 0.6822494
> cat("Model 4:", adj_R2[4], "\n\n")
Model 4: 0.4966469
```

Model 3 (using TIME²) appears to perform the best for several reasons:

- It has the highest adjusted R^2 value, indicating it accounts for the most variance.
- Its residuals exhibit the most favorable distribution, confirmed by the Shapiro–Wilk normality test ($p = 0.8266$).
- A visual inspection of the fit and residual plots further supports that Model 3 is the strongest choice.

c.

```
> cat("Indices of potential outliers:", outlier_idx, "\n")
Indices of potential outliers: 6 14 28 43 44 45 46 47 48
> cat("Corresponding years:", 1950 + outlier_idx - 1, "\n\n")
Corresponding years: 1955 1963 1977 1992 1993 1994 1995 1996 1997
```

d.

```
> cat("95% Prediction Interval for 1997 Yield:\n")
95% Prediction Interval for 1997 Yield:
> cat("Lower Bound:", pred_int[1, "lwr"], "\n")
Lower Bound: 1.372403
> cat("Upper Bound:", pred_int[1, "upr"], "\n")
Upper Bound: 2.389819
> cat("Predicted Yield:", pred_int[1, "fit"], "\n")
Predicted Yield: 1.881111
> cat("Actual Yield (1997):", actual_yield, "\n")
Actual Yield (1997): 2.2318
> cat("Does the interval cover the actual yield? ",
+     (actual_yield >= pred_int[1, "lwr"] & actual_yield <= pred_int[1, "upr"]), "\n")
Does the interval cover the actual yield?  TRUE
```