# Chapter 7

## 1. Assignment 1

Consider the two data files Chap\_7\_HW\_File\_2s.csv and Chap\_7\_HW\_File\_3s.csv found in the Files for Homework Assignments folder on BlackBoard. Each file consists of four time series, three predictor time series (P1, P2,P3) and a Data time series which may depend on the three predictors. For each file produce a linear regression model P1, P2,P3, P1 and P2, P1 and P3, P2 and P3, and all three. Use the first 90% of the Data to train the models and final 10% as a test case. Plot each model and the original Data. Analyze the models and let R decide which one is the best model. The first few lines of the file are shown below.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | Date | P1 | P2 | P3 | Data | | 1/1/2011 | -0.89547 | 0.02618 | 1 | 5.130708 | | 1/2/2011 | -0.69098 | 0.078531 | 1.8E-06 | 5.387548 | | 1/3/2011 | 1.5 | 0.130858 | 0.000005 | 4.630858 | | 1/4/2011 | 1.669131 | 0.183146 | 9.8E-06 | 4.852276 | |  |  |  |
|  |  |  |  |

## 2. Assignment 2

The annual population of Afghanistan is available in the global\_economy data set.

### a. Plot the data and comment on its features.

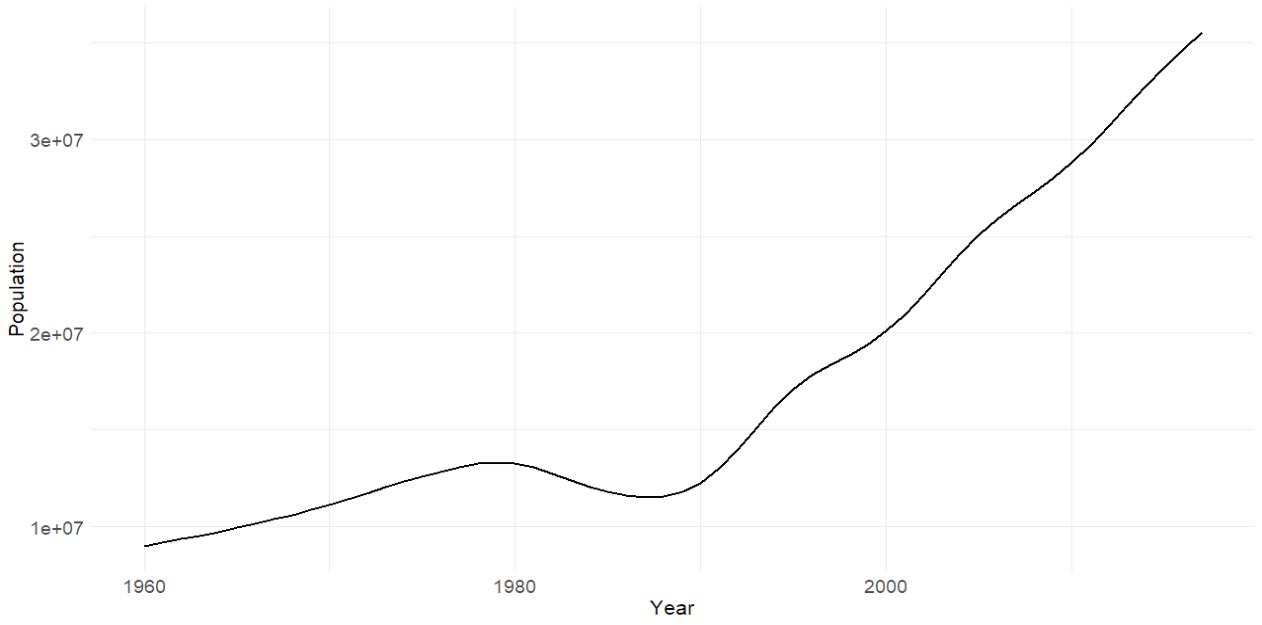


Figure : Time Series: Afghanistan Population vs Year

The population grew slowly from 1960 to 1980, then began to decline until 1990, after which it increased rapidly.

## b. Fit a linear trend model and compare this to a piecewise linear trend model with knots at 1980 and 1989.

### Linear trend model

Series: Population

Model: TSLM

Residuals:

Min 1Q Median 3Q Max

-5794518 -2582559 744761 2259222 6036280

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4798904 848259 5.657 5.45e-07 \*\*\*

trend() 425774 25008 17.025 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3188000 on 56 degrees of freedom

Multiple R-squared: 0.8381, Adjusted R-squared: 0.8352

F-statistic: 289.9 on 1 and 56 DF, p-value: < 2.22e-16

### Piecewise linear trend model

Series: Population

Model: TSLM

Residuals:

Min 1Q Median 3Q Max

-577590 -174198 -16784 187226 679947

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8697573 131122 66.33 <2e-16 \*\*\*

trend(knots = c(1980, 1989))trend 224372 9623 23.32 <2e-16 \*\*\*

trend(knots = c(1980, 1989))trend\_21 -456804 24498 -18.65 <2e-16 \*\*\*

trend(knots = c(1980, 1989))trend\_30 1082782 21418 50.55 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 300900 on 54 degrees of freedom

Multiple R-squared: 0.9986, Adjusted R-squared: 0.9985

F-statistic: 1.293e+04 on 3 and 54 DF, p-value: < 2.22e-16

### Compare the models visually



Figure : Linear trend model vs Piecewise linear trend model

Based on the graph and from the reports, the piecewise linear model explains 99.86% of the variance in the data, much higher than the 83.81% explained by the simple linear model. Its residual error is also about 90% lower, indicating much better accuracy.

### c. Generate forecasts from these two models for the five years after the end of the data, and comment on the results.

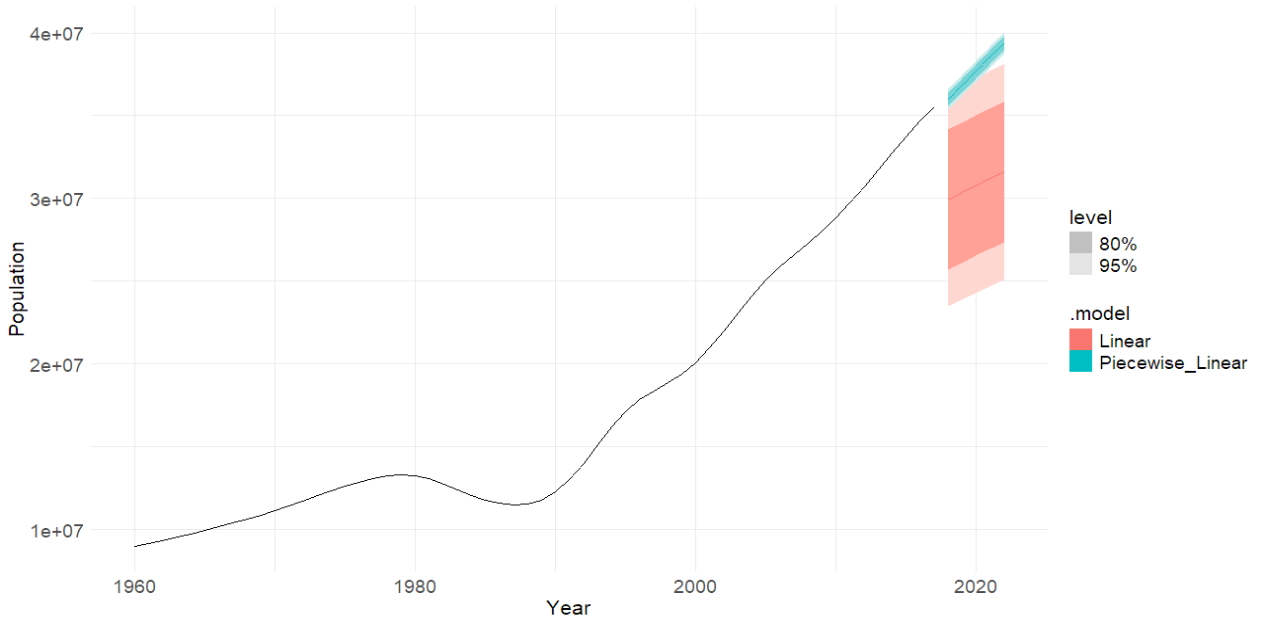


Figure : Forecasts from two models

The linear model appears inadequate, as it produces excessively wide prediction intervals and systematically underestimates future values.

In contrast, the piecewise linear model provides a better fit to the data. However, its prediction intervals are likely too narrow, reflecting overconfidence in the model's assumptions.