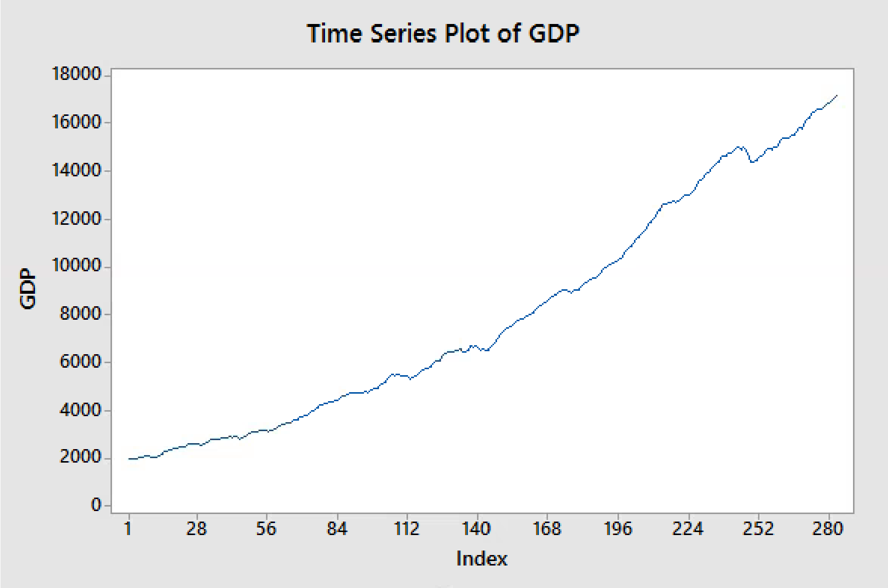
**HW1 Tianze Lin tl2718**

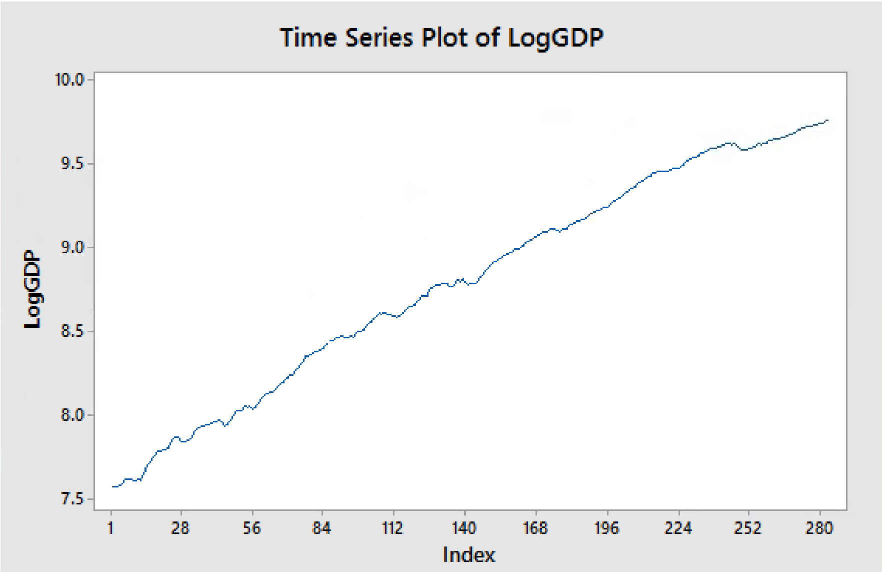
**Question 1**

**A.**




I think GDP seems to grow linearly over time because the GPD is increasing as the time proceeds.

**B.**




The log GDP appears to grow linearly over time. Changes in the natural logarithm are approximately equal to percentage changes of the original data series. Log GDP increases by 2 (7.6-9.6) over past 70 years, so the average growth of GPD is about 2/70 per year, equivalent to 2.86% per year.

**C.**

### Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 7.66121 | 0.00795 | 964.12 | 0.000 | |
| time | 0.007938 | 0.000049 | 163.65 | 0.000 | 1.00 |

### Regression Equation

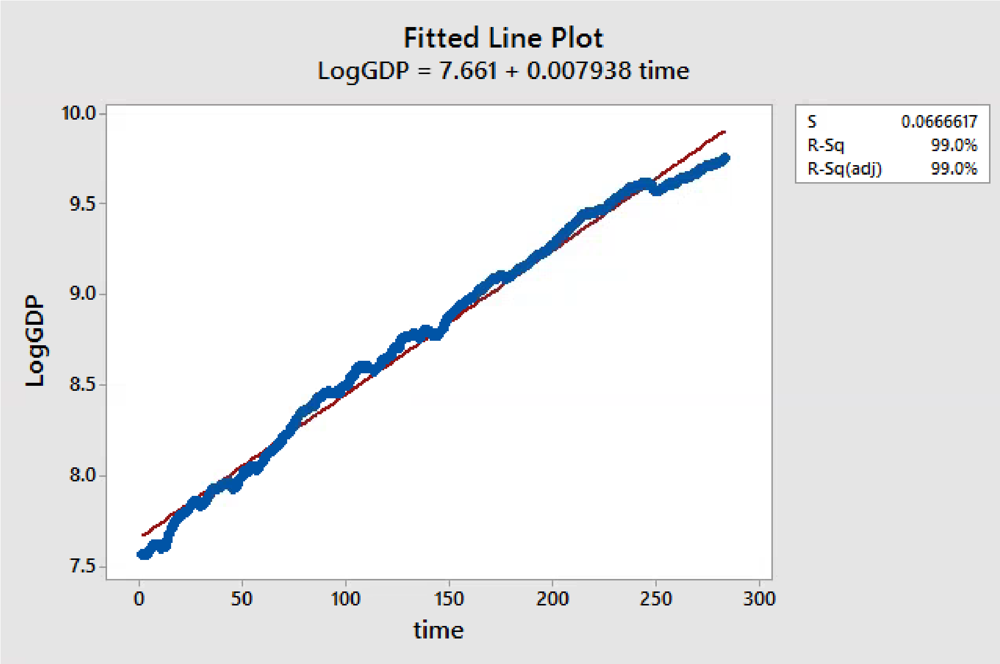
LogGDP = 7.66121 + 0.007938 time

## Prediction

| Fit | SE Fit | 95% CI | 95% PI |
|---|---|---|---|
| 9.92355 | 0.0079884 | (9.90783, 9.93928) | (9.79140, 10.0557) |

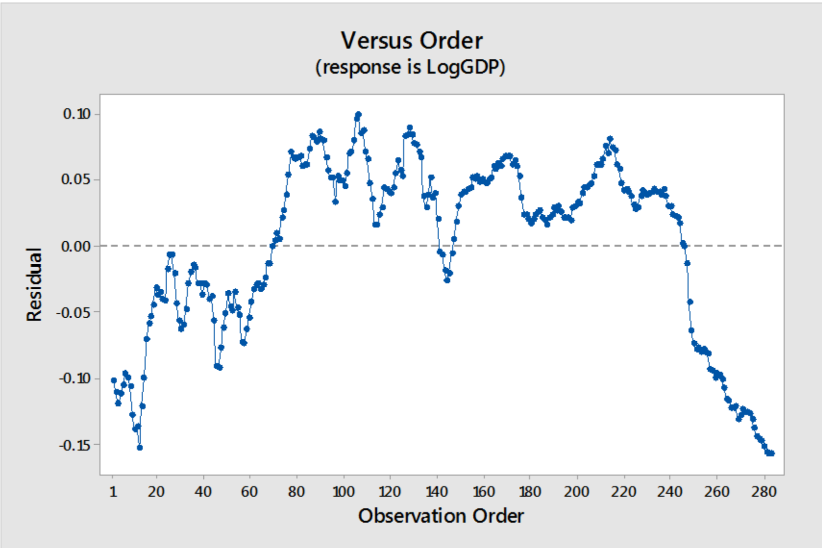
The 95% prediction interval is (9.79140, 10.0557). I don't think this prediction is effective because this prediction interval is too narrow.

**D.**




There are some data points below the fitted line in the beginning and the end of time data series, and this may be caused by trend or autocorrelation. However, the $R^2$ is 99%, which means the model explains 99% of the variability of the response data around its mean. Therefore, the line fits well.

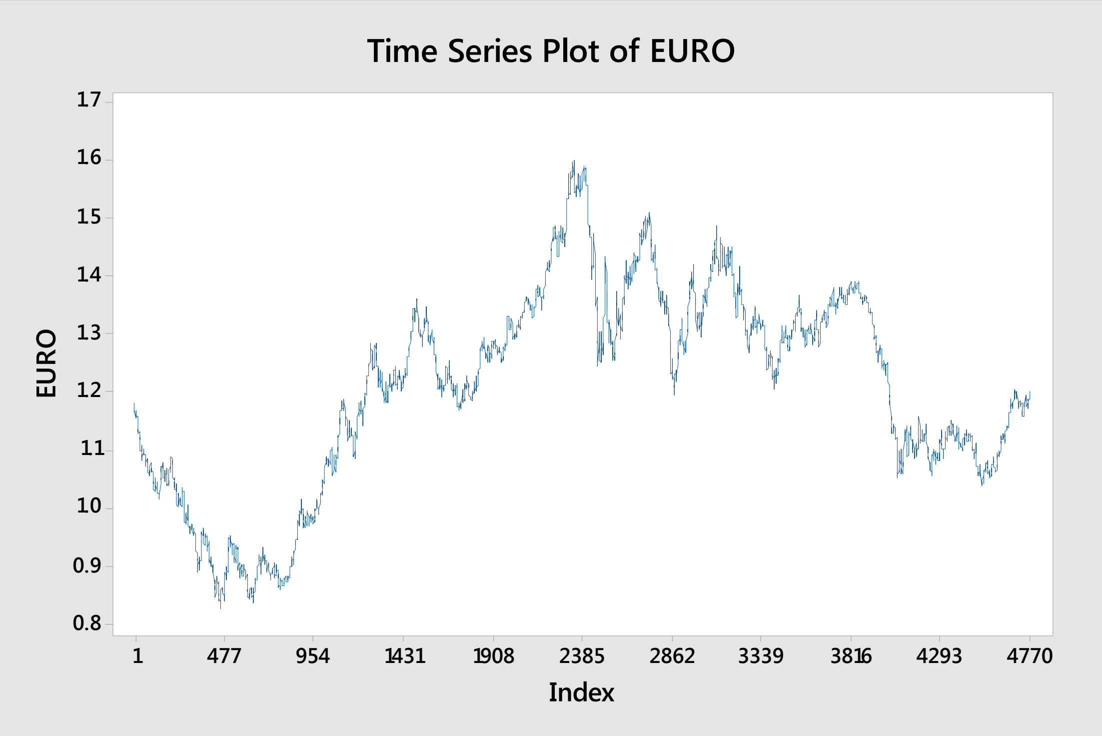
**E.**




From the residual plot, we see the residuals are not randomly distributed around 0, and they are not independent and follow some pattern This is caused by autocorrelation. These problems could spoil the validity of the forecast interval.

**Question 2**

**A.**




Clearly, the straight-line model doesn't seem appropriate.

**B.**

Model of the first 700 data points:

## Regression Analysis: EURO_2 versus Time2

### Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 4.1180 | 4.11800 | 2937.14 | 0.000 |
| Time2 | 1 | 4.1180 | 4.11800 | 2937.14 | 0.000 |
| Error | 698 | 0.9786 | 0.00140 | | |
| Total | 699 | 5.0966 | | | |

### Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 0.0374439 | 80.80% | 80.77% | 80.68% |

### Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 1.09976 | 0.00283 | 388.12 | 0.000 | |
| Time2 | -0.000380 | 0.000007 | -54.20 | 0.000 | 1.00 |

## Regression Equation

EURO_2 = 1.09976 - 0.000380 Time2

## Settings

| Variable | Setting |
|---|---|
| Time2 | 4773 |

## Prediction

| Fit | SE Fit | 95% CI | 95% PI | |
|---|---|---|---|---|
| -0.711912 | 0.0310060 | (-0.772788, -0.651036) | (-0.807361, -0.616463) | XX |

*XX denotes an extremely unusual point relative to predictor levels used to fit the model.*

Model of 701-4772:

# Regression Analysis: EURO_1 versus Time1

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 1.429 | 1.42876 | 65.79 | 0.000 |
| Time1 | 1 | 1.429 | 1.42876 | 65.79 | 0.000 |
| Error | 4070 | 88.386 | 0.02172 | | |
| Total | 4071 | 89.815 | | | |

## Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 0.147365 | 1.59% | 1.57% | 1.48% |

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 1.20613 | 0.00585 | 206.14 | 0.000 | |
| Time1 | 0.000016 | 0.000002 | 8.11 | 0.000 | 1.00 |

## Regression Equation

EURO_1 = 1.20613 + 0.000016 Time1

## Settings

| Variable | Setting |
|---|---|
| Time1 | 4773 |

## Prediction

| Fit | SE Fit | 95% CI | 95% PI |
|---|---|---|---|
| 1.28219 | 0.0046196 | (1.27314, 1.29125) | (0.993135, 1.57125) |

The forecast interval from first model doesn’t succeed, but the second one is successful. The first predication interval fails since 1) When regression is performed on time series data, errors are autocorrelated, so errors will follow some pattern, which violates the assumptions of linear regression. 2) In the regression equation, the slope term is negative, which means the USD/Euro exchange rate will always decrease as time increases. Therefore, when the time gets really large, the exchange rate will become negative, which is impossible in real-life situations.