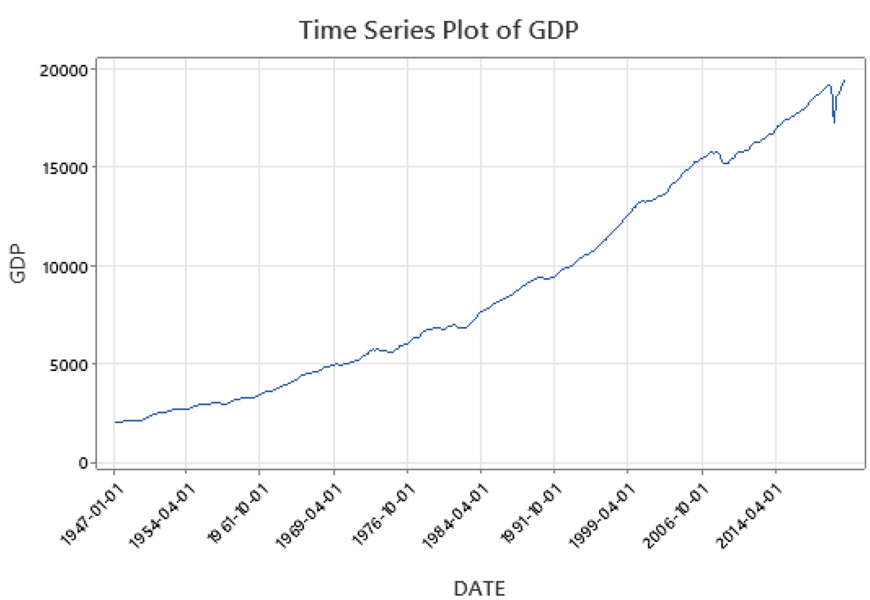
Forecasting Time Series Homework 1

Student Name	NYU ID
Yoon Tae Park	yp2201
Yeong Koh	yk2678
Yoobin Cheong	yc5206

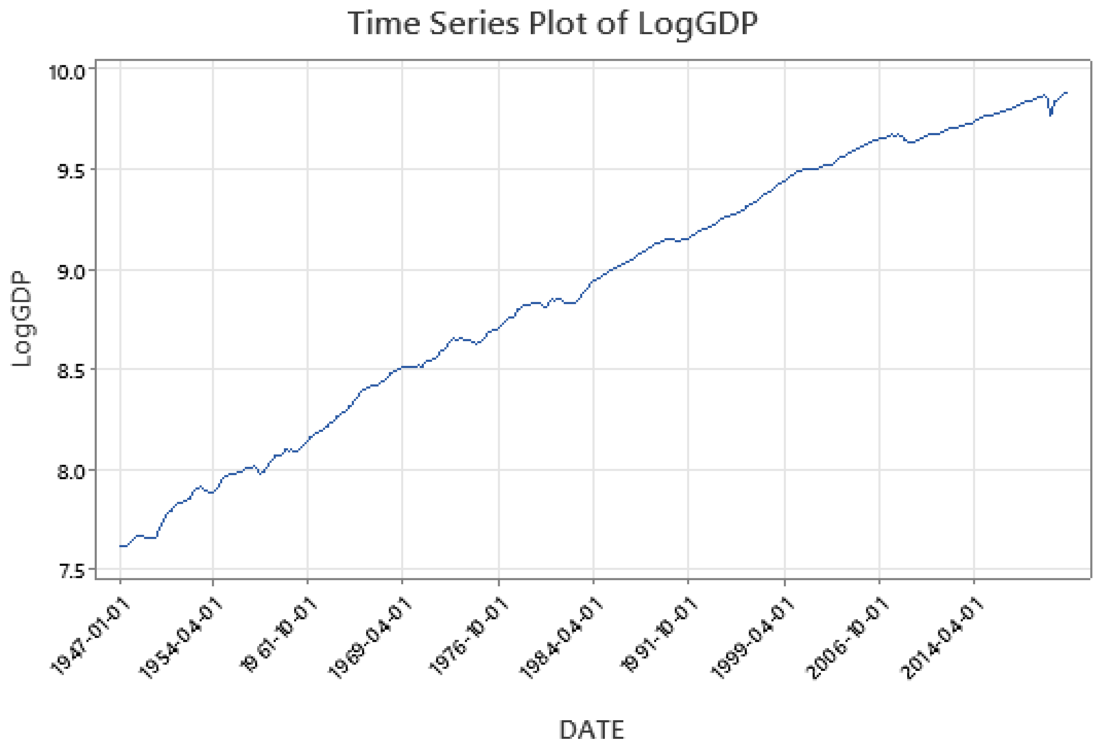
1)

A) Plot the GDP series versus time. Note that the data is quarterly, from 1946, Q4 to 2021, Q2, a total of $n = 299$ observations. (When talking about a plot of y versus x, it is traditional to give the response variable first, then the predictor variable. So I am asking for GDP on the y-axis, and time on the x-axis.) (In Minitab, to make a time series plot, you can use Stat $\rightarrow$ Time series $\rightarrow$ Time Series Plot. In the plot menu, I recommend that you click Data View and then un-check the box for Symbols. This removes the big dots that can obscure the fine structure of the plot.) Does GDP seem to grow linearly over time?



Answer: GDP seems to grow somewhat linearly over time in general, although it also exhibits some exponential growth and fluctuations especially towards the recent years.

B) Plot the log of GDP, in a time series plot. Does log GDP appear to grow linearly over time? If so, then what does this imply about the growth of GDP itself? To create the log of GDP in Minitab, you can use Calculator: Calc → Calculator → Store Result in Variable: LogGDP, Expression: log(GDP).



Answer: Log GDP appears to grow linearly over time. This implies that the growth of GDP itself is exponential. Level dependent volatility that we saw in part A is stabilized by taking the log.

C) Fit an ordinary linear regression model for LogGDP, using time as the predictor variable. (In Minitab, generate a column containing the predictor variable "time", taking values 1,2,3,...,299. To do this, use Calc → Make patterned data → simple set of numbers.) Based on the output, forecast the LogGDP for the fourth quarter of the year 2021. Based on the linear regression model, construct a 95% prediction interval to go with the above point forecast. (To get prediction intervals, after running the regression, use Stat → Regression → Regression → Predict, and enter the x-value at which you want to predict y. This x-value should be 301, since the 299th point was the final data value, corresponding to the second quarter of 2021.) Do you think this interval forecast (prediction interval, denoted by PI in the Minitab output) is valid? In your opinion, does it seem too wide, or too precise?

Regression Analysis: LogGDP versus time

Regression Equation

LogGDP = 7.72615 + 0.007761 time

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	7.72615	0.00858	900.14	0.000
time	0.007761	0.000050	156.48	0.000

Model Summary

S	R-sq	R-sq(adj)
0.0740232	98.80%	98.80%

GDP.MWX

Prediction for LogGDP

Regression Equation

LogGDP = 7.72615 + 0.007761 time

Settings

Variable	Setting
time	301

Prediction

Fit	SE Fit	95% CI	95% PI
10.0621	0.0086263	(10.0451, 10.0791)	(9.91545, 10.2088)

Answer: This interval forecast seems to be too wide. As we are taking logGDP to get a prediction, 95% prediction interval gets too wide when it is converted back into GDP. As our goal is to predict the actual GDP, the range presented by logGDP conversion seems to have a wider range than expected.

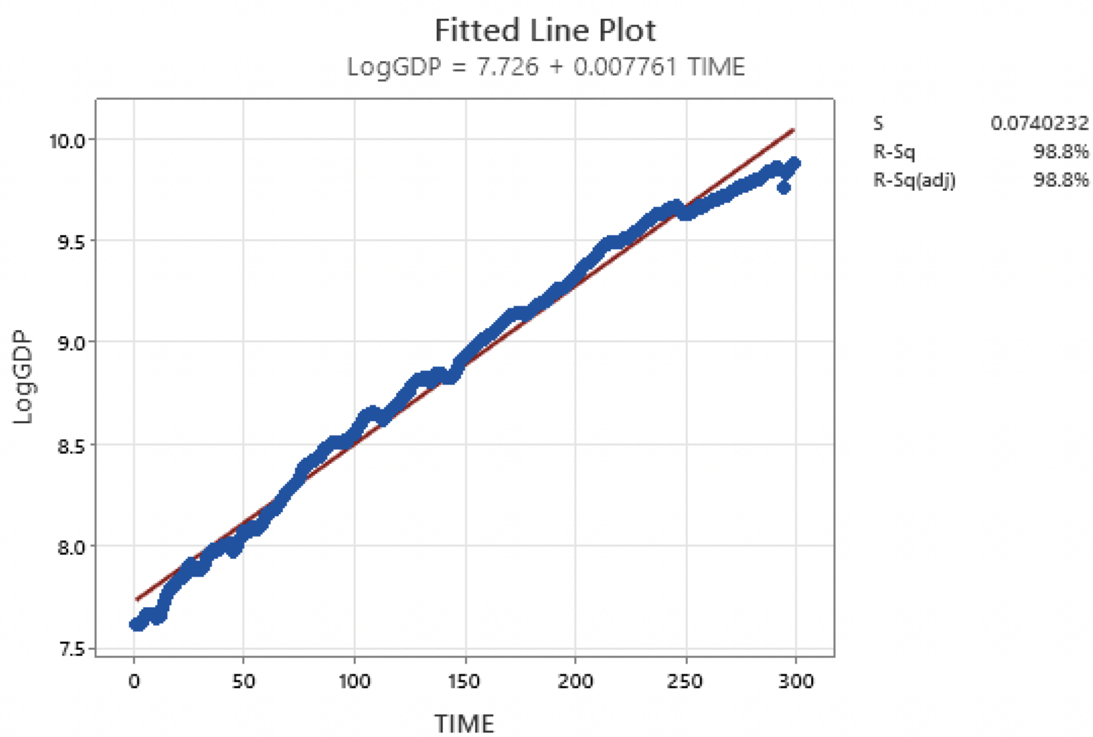
D) On a single plot, superimpose the LogGDP series and the fitted line. (For this, you can use Stat → Regression → Fitted Line Plot.) Does the line fit well?

Model Summary

S	R-sq	R-sq(adj)
0.0740232	98.80%	98.80%

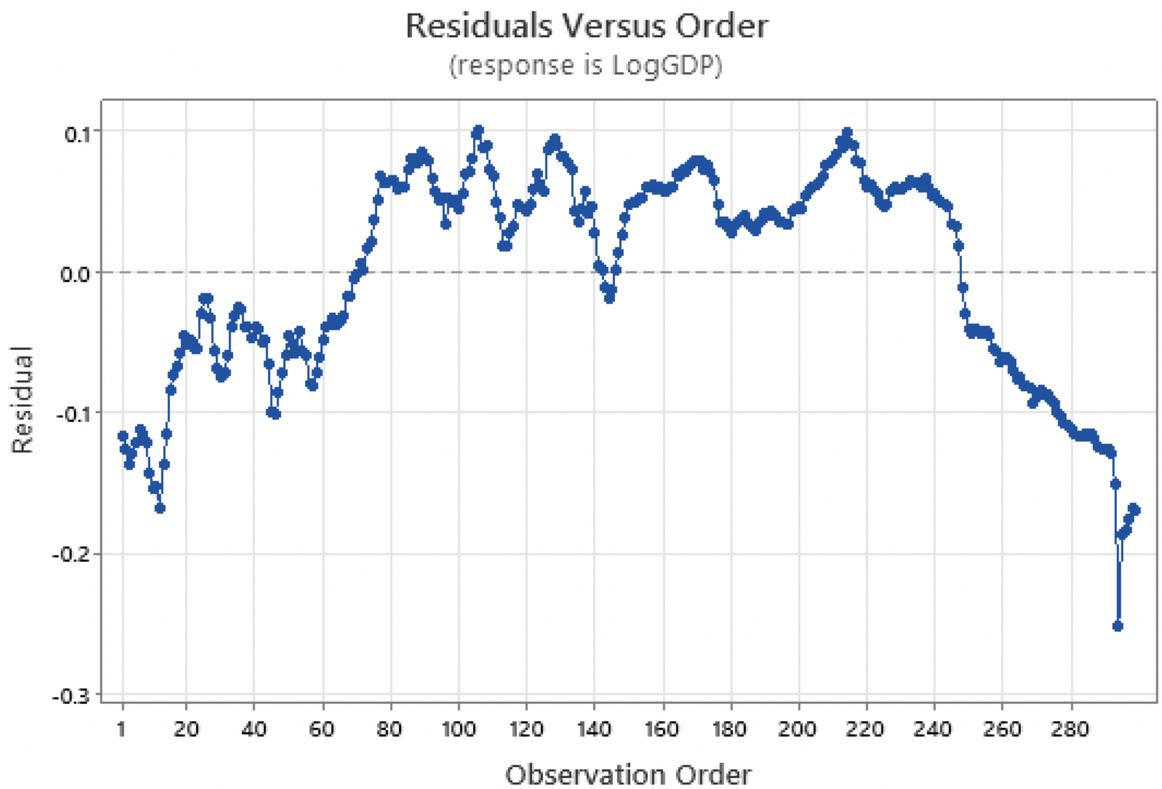
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	134.161	134.161	24484.49	0.000
Error	297	1.627	0.005		
Total	298	135.788			



Answer: The line seems to fit well. 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 98.8%, which means the model explains 98.8% of the data (LogGDP).

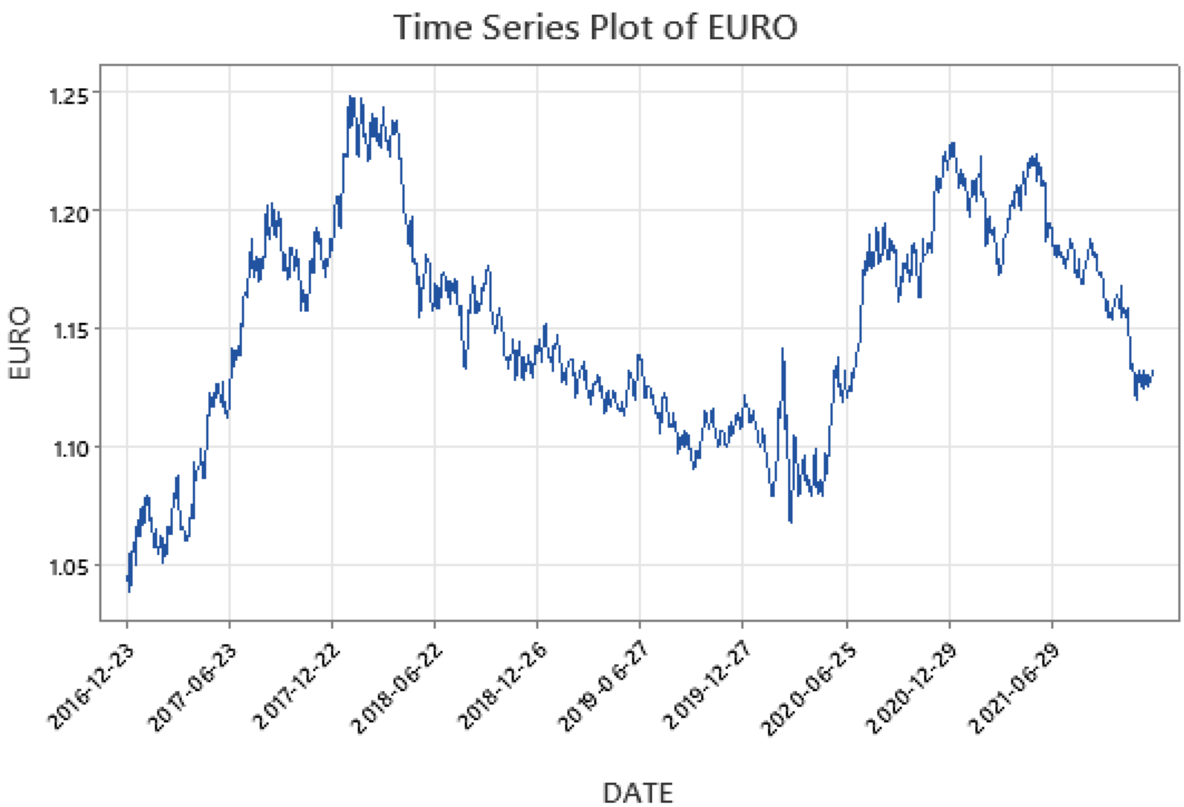
E) Plot the residuals from the fitted line versus time. (In Minitab, this can be done by selecting "Graphs" in the regression dialog box, and then checking "Residuals versus order"). What potential problems with the linear model are indicated by this plot? Do you think these problems could spoil the validity of the forecast interval?



Answer: The potential problems with the above plot is that the residuals are not randomly distributed around zero, suggesting a lack of independence. This might be caused by the autocorrelation, thus it could spoil the validity of the forecast interval. The non-constant variance can also spoil the validity of the forecast interval.

2) Consider the daily U.S. Dollar/Euro exchange rate, daily, 23 Dec 2016 to 23 Dec 2021 (n=1247). The numbers are given in the file EURO.mwx on the course website.

A) Create a time-series plot of Euro. Does a straight-line model seem appropriate?



Answer) No, straight-line model doesn't seem appropriate, as time-series plot shows relatively high variance that any straight-line models are not good enough to fit the data.

B) Get point and interval predictions for 23 Dec 2021, using two different methods: First, based on fitting a straight line to observations 1 to 300; Second, based on fitting a straight line to observations 301 to 1246. Did both of the forecast intervals succeed in containing the actual value for 23 Dec 2021? If not, then use what you learned from Problem 1 to give a statistical explanation of what went wrong.

Answer:

- **Fitting a straight line to observations 1 to 300**

Regression Equation

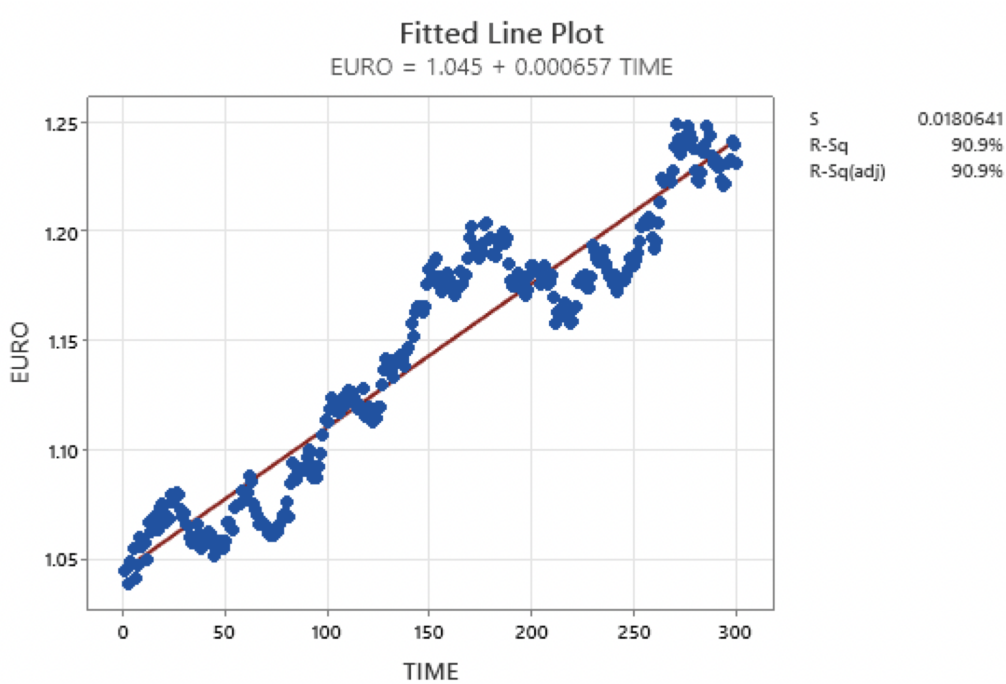
EURO = 1.04465 + 0.000657 TIME

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.04465	0.00209	499.57	0.000	
TIME	0.000657	0.000012	54.58	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0180641	90.91%	90.88%	90.81%



1-300

Prediction for EURO

Regression Equation

EURO = 1.04465 + 0.000657 TIME

Settings

Variable	Setting
TIME	1247

Prediction

Fit	SE Fit	95% CI	95% PI	
1.86429	0.0132461	(1.83822, 1.89035)	(1.82020, 1.90837)	XX

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

- **Fitting a straight line to observations 301 to 1246**

301-1246

Regression Analysis: EURO versus TIME

Regression Equation

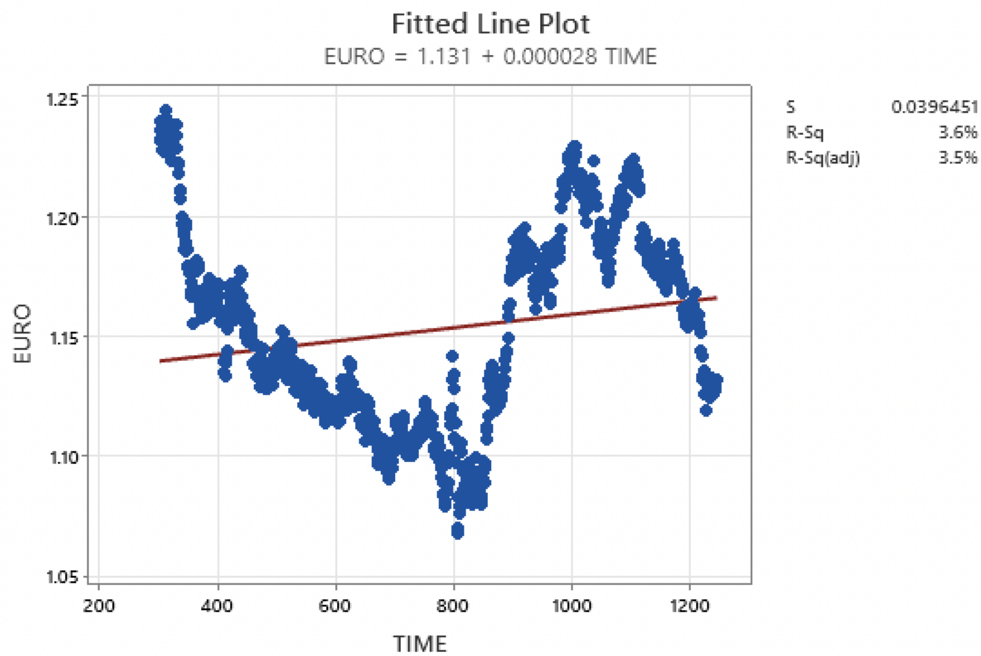
EURO = 1.13140 + 0.000028 TIME

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.13140	0.00387	292.22	0.000	
TIME	0.000028	0.000005	5.90	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0396451	3.56%	3.46%	3.14%



301-1246

Prediction for EURO

Regression Equation

EURO = 1.13140 + 0.000028 TIME

Settings

Variable	Setting
TIME	1247

Prediction

Fit	SE Fit	95% CI	95% PI
1.16615	0.0025800	(1.16109, 1.17121)	(1.08818, 1.24412)

Answer> The forecast using observations 1 to 300 (model 1) failed to contain the actual value in the prediction interval, while the forecast using observations 301 to 1246 (model 2) succeeded in containing the actual value in the prediction interval. The possible cause for model 1 could be having only an increasing trend in the beginning of the data (1-300), so it fails to predict the fluctuations followed by the time 300. Model 1 does not capture the degree of randomness that exists in the future data. On the other hand, Model 2 is better at reflecting signal and noise by including both positive and negative trends, and has more observations than Model 1. Therefore Model 2 can be more robust compared to Model 1 and have better predictions.