Title: Technical Report
Author: 31324878
Exponential Smoothing: 5 Pages (including references)
Appendix A: 1Page
Appendix B: 10 Pages
No. of Python Codes files: 19
Words Count: 2,815

1. Exponential Smoothing

1.1 EAFV data

Before the method is taken, the data is processed in the preparation stage. EAFV data, during this stage, is respectively transformed into training set (70% entire data), test set (30% hold out), and seasonal data forms for upcoming preliminary analysis. Lastly, in this phase, neither apparent error data point nor missing value is visually observed, while further confirmation will be made in the preliminary analysis (next) stage.

In the preliminary stage, no outlier is observed via Figure B0. “time plot”, “seasonal plot” (drawn from seasonal data), “autocorrelation (ACF)” and “partial autocorrelation (PACF)” plot are used to observe some trends and seasonality. The findings of time and seasonal plot indicate the trend and seasonality strongly exist with clear trend of data (by spikes which can never go down to zero soon) and characteristics repeating itself every 12 months. With such insight, Holt-Winter Method (H-W) accordingly stand out from other Exponential Smoothing Methods.

The types of H-W models can be various according to additive and multiplicative patterns. It is hence necessary to select one through comparing errors (in our context it refers to MSE, MAE and MAPE) of training and test set. For this, four potential models are validated, including [additive trend, additive seasonality] (Model 1), [multiplicative trend, additive seasonality] (Model 2), [multiplicative trend, multiplicative seasonality] (Model 3) and [additive trend, multiplicative seasonality] (Model 4). During the analysis, we adapt automatic optimised parameters calculated by function `.fit()`. As can be seen from Figure B1-4, Model 4 numerically has the least value over the training-set errors but has the poorest forecasting performance (largest errors on test set). Therefore, it is not an appropriate option. Graphically, the Model 2 and 4 mispredict the patterns will go up but in real data (test set) while it does not (Figure B1-3). This finally leads to higher test-set errors. Overall, the choice of forecasting model is Model 1 and 3, who have comparable test-set errors with one another. Based on Lewis (1982), both are “highly accurate” models with 6.15% and 6.22% in MAPE, respectively.

Here we present graphical forecasting result and residuals measurement for model 1 only. The ACF and PACF in residuals (obtained from fitting the entire dataset to the model) of Model 1 shows no trend in residuals but have interruption demonstrated in some unique points, which does not resemble fully white noise.

1.2 K226 Data

In preparation stage, the data is structured as seasonal data type and is split into training (70%) and test set (30%) for preliminary analysis and model selection. From dataset itself, there is no apparent outlier that can visually be observed and neither can missing/erroneous values. This information will be checked further in preliminary analysis.

In preliminary analysis, graphical method is the main skill used. we take boxplot and indeed find no outlier as shown in Figure B0. The time plot, ACF and PACF of original data (please see Figure B2-1 and B2-2) all reveal the strong trend (check the spikes and they do not go down to zero very soon) in general patterns while seasonality can hardly be diagnosed with. As a result, there are 6 potential models that can be tested: two Holt's Linear Exponential Smoothing Method (LES) and four H-W models. They will be validated over errors of training and test set for model selection.

In model selection, the parameters are automatically computed and are used. For errors of training set, six models are generally comparable with each other. Model 1 and 2 are LES with "additive trend" and "multiplicative trend" respectively. Model 3 to 6 include H-W with "additive trend; additive seasonality", "multiplicative trend; additive seasonality", "multiplicative trend; multiplicative seasonality" and "additive trend; multiplicative seasonality".
By Figure B2-3, we can check the errors. The training-set MSE of Model 1 and 2 are higher than others by approximately 2 units. As for training-set MAE and MAPE, there is of trivial difference among all the models. Contrastingly, in test-set error, the difference of prediction power is significantly different among models. Graphically (by Figure B2-4), Model 1, 3 and 6 make a totally opposite/wrong prediction on test set (real data trends up but it predicts down). The Model 2, 4 and 5 predict the bounce-back on the scope of test set where the 4 is one making least errors. Also, its MAPE is around 11.78 (%) which represent a "Good forecast" level according to Lewis (1982).

Out-of-sample forecasting and confidence interval, as a result, is made based on Model 4. The result can be checked in Figure B2-6. The residual is also estimated via ACF and PACF where we see no trend is shown. But unfortunately some seasonal components are not captured by the model (see the significant spikes in Figure B2-7).

1.3 K54D data

In preparation stage, the K54D data is divided into training (70%) and test set (30%) for testing model selection. Also, the seasonal type of data is created to build the seasonal plot in the preliminary analysis.

In preliminary analysis, the boxplot is adapted to check the outlier and finally no outlier is found (see Figure B0). Graphically, the trend and seasonality can both be observed via time plot, seasonal plot, ACF and PACF plot (Figure B3-1 and B3-2). Clearly, we acquire, from Figure B3-1, that the seasonal period is likely to be 12 months by the unique pattern repeating itself per 12 months from ACF plot. Overall, the appropriate model can be selected from four possible H-W models.

The selection of models is determined through four H-W models: Model 1: additive trend and multiplicative seasonality; Model 2: both multiplicative trend and seasonality; Model 3: additive trend and seasonality; Model 4: multiplicative trend and additive seasonality. From results shown in Figure B3-3, our final selected model is Model 1 who have least performance in MSE, MAE and MAPE in training set while predict best in test set (Model 3 has nearly no difference from 1, also belonging to a good forecaster). The 12-month forecasting results along with confidence interval can be seen in Figure B3-4. In the same graph, the ACF and PACF of residuals reveal some unique non-white noise components (some irregular significant spikes).

1.4 JQ2J data

Like what we have done beforehand, the seasonal data, training and test set are also generated or split from the original data.
In preliminary analysis, the boxplot points out some outliers in the original dataset. As indicated by Figure B0, these data points locates above the upper limit. For knowing what and where they are, the following measures are taken: (1) obtain

values of quartiles and upper/lower limits; (2) detect the outliers by logical label "Non-outlier" and "Outlier" and setting conditional algorithms: If the value locates within the upper and lower limit, Non-outlier, Otherwise, Outlier; (3) finally record the index/location where these values occur. These steps and results can be referenced in sheet *JQ2J_pre* in file *JQ2Jdata_31324878.xls*. By checking their location, some patterns appear------these values repeatedly occur at March and September ever since 2016 (one exception: Sept. 2018), which is line with features unveiled in seasonal plot (Figure B4-1). It is indicated some information may be hidden behind them and thus we choose to remain them, whilst it may result in difficulties of modelling. Through time plot, ACF and PACF plot, the trend and seasonality are revealed and hence 4 possible H-W models are going to be tested.

Model 1 refers to one with "both multiplicative trend and seasonality"; Model 2, "additive trend and multiplicative seasonality"; Model 3, "multiplicative trend and additive seasonality" and Model 4 "additive trend and multiplicative seasonality". Via test of MSE, MAE and MAPE, the Model 1 apparently stands out in terms of prediction on test set. The forecasting results and confidence interval are shown separately in Figure B4- and B4-3 (value of confidence interval is too large to be effectively graphed). Noteworthily, the key is the residuals under ACF and PACF plot which shows remaining seasonality and several non-white noise outcomes. This means (1) the confidence interval can hardly capture the interval where the real data point may probably occur and (2) the model cannot thoroughly seize the main features in JQ2J, which also affects the effectiveness on forecasting. In this case, the various transformation other than squared root and log are recommended in future analysis.

2. ARIMA

The data preparation on K54D we undertake covers the split of training/test set (70% vs 30%), the data transformation on square-root/logarithmic and the seasonal-type data, all of which is identical as have done at Exponential Smoothing part. In preliminary analysis, we identify the features of both trend and seasonality and further address them by taking 1^{st} differencing and seasonal differencing. Figure B5-1 shows time plot and ACF plot of 1^{st} differencing where the seasonal patterns remain. Therefore, seasonal differencing is subsequently operated. As can be seen in Figure B5-2 the exponential decay in PACF indicates a non-seasonal MA patterns. Via the spikes shown at lag 1 to 2, the "non-seasonal" MA(2) is suggested. Furthermore, the significant spike at lag 12 in ACF refers to a "seasonal" MA(2). With 1st and seasonal differencing already taken, the d and D refer both to 1. In sum, the ACF and PACF plot suggest a "tentative" model: ARIMA$(0, 1, 0)(0,1,2)_{12}$. As indicated by Makridakis, S. *et al*. (1998), an appropriate mixed ARIMA model can hardly be identified precisely, so a list of potential models based on this basic is thus prepared for further testing and selection.

There are 256 extended models to be tested, including those with d and D = 1; p, P, q, Q $\in \{0, 1, 2, 3\}$; s = 12. Upper value "3" set in p, P, q, Q is set for capturing more possibilities that may happen. Finally, the appropriate model selected is ARIMA$(0, 1, 2)(2,1,1)_{12}$ due to its least AIC value, 1,549.013. By our diagnosis, there are another 8 models with similar but higher level of AICs (around 1,549). The results of the table show that any p, q, P, Q which equals to 3 seldom appear. This reflects Makridakis, S. *et al*. (1998, pp. 345)'s idea that the value in 0, 1, 2 can actually include various kinds of forecasting situations and those other than these are not practically necessary. The residual analysis is demonstrated in Figure B5-3. In that, some residual points are apparently outliers which are as shown in histogram part that demonstrates outliers located over range of 3σ. The upper and lower tails deviating from 45-degree line of QQ plot also reveal the existence of outliers. Lastly, 12-month forecasting results along with 95% confidence interval are then made and displayed in Figure B5-3.

For testing and comparing model's effectiveness, the final model is also implemented on training and test set. The MSE in training set is around 686.98 and that of test set, 26.49. For measuring accuracy, the MAPE of test set is 0.74, which reaches highly accurate level under Lewis (1982)' classification. According to errors, this (Seasonal) ARIMA model performs worse than H-W does in model fitting (H-W has lower training-set MSE, 64.56). Contrastingly, it has apparently better predictive ability based on test-set MSE and MAPE (H-W: MSE = 77.1, MAPE = 1.32). By comparing correlogram of residuals on training set (still Figure B5-3), ARIMA model is still better than H-W model on data K54D. For instance, the residuals of ARIMA model are far from significant and tend to be white noise. However, the H-W model has a spike at lag 2 that does not directly go down to zero and remains some significant spikes as well.

What makes the effectiveness of two models different is their nature. ARIMA model adapt parametric techniques involving two components of AR and MA integrated, which makes the data more stationary when the method is applicable. The Exponential Smoothing is more like the division of ARIMA model with non-parametric characteristics. This may sometimes lead to over-smoothing to the patterns as compared to ARIMA.

3. Regression Forecasting

In data preparation stage, the lengths of explanatory variables, EAFV, JQ2J, K54D and K226, are readjusted which means their length in this section have been different from the Exponential Smoothing part (section 1). The length of EAFV has been cut from 384 to 240 and as for K226, 300 to 240. The other explanatory variables have been required to start from January 2000 and so has dependant variable, FTSE. Based on abovementioned idea, the details of training (70% of all data) and test set (30% hold-out) is determined.

In addition, the dummy variables D1 to D11 are prepared. If the data point is referred to as the first month (of the year), its corresponding value of D1 will be set 1 and otherwise, 0; If the data is referred to as second month, the corresponding value of D2 will be set 1, otherwise, 0. This can be inferred to D11 (eleventh month of the year). We drop D12 from our data but make it a referenced variable for those dummy variables. Lastly, the variable, time, is added for indicating the index of month (from 1 to 240). Both "dummy variables" and "time" are created to reflect time effect in series data for regression modelling.

In preliminary analysis, the time plot of all variable is analysed. The aim of the time plot is to check if the model for independent variables, EAFV and K226, that we selected in section 1 can still be matched (because the length of these two variables have been changed). Generally, the patterns, based on Figure B6-1 and B6-2 can seemingly be similar whilst noteworthily, the trend and variation of EAFV get slightly flat than it was in section 1. In this case, it is "hypothesised" that the model used in section 1 can still be applied on the regression forecasting.

Also, the correlation plot is used. This reveal the relationships among the variables and how strong they are. As displayed in Figure B6-3, the relationship between FTSE (again, dependent variable) and K226 shows a "M" wave pattern, which is not a typical linear form. Similarly, the relationship between K54D and FTSE displays irregular pattern. This may adversely affect model fitting and prediction. Graphically, the linear relationship is sometimes strong within independent variables such as that between K54D and K226 and also, K54D and JQ2J. This may result in collinearity or multicollinearity which affect model's effectiveness.

In regression prediction/forecasting, two influential parts are suggested: Forecasting/prediction and regression part. In previous stage, the methodology for forecasting part has been selected. The model for EAFV is H-W with additive trend and seasonality; one for K226 is H-W with multiplicative trend and additive seasonality; one for JQ2J is H-W with multiplicative trend and seasonality and as for K54D, the H-W with multiplicative trend and additive seasonality is applied. For regression part, three potential models take forms:

$$FTSE\_Y = b0+b1*EAFV+b2*K226+b3*JQ2J+b4*K54D \quad (1)$$

$$FTSE\_Y = b0+b1*EAFV +b2*K226+b3*JQ2J+b4*K54D+D1+D2+D3+D4+D5+D6+D7+D8+D9+D10+D11 \quad (2)$$

$$FTSE\_Y = b0+b1*EAFV+b2*K226+b3*JQ2J+b4*K54D+D1+D2+D3+D4+D5+D6+D7+D8+D9+D10+D11+time \quad (3)$$

The results of regression are shown in Table B6-1. The model 1 (correspond to equation 1) owns the highest R square and adjusted R square. The model 2 has the second better R square value but worst in AIC and adjusted R sqaure. The AIC of model 1 is the best among all the models. It is hard to distinguish which can generally be the most appropriate model. Subsequently, three models are further compared using their performance in training and test set by MSE. Model 1 outperforms in training-set, with MSE around 1,207,067 and Model 3 does the worst. By Figure B6-4, the fitted results on training set of three models suggest different trends and variations but they poorly match the real/original data.

As for prediction power (by checking test set), all three models have very tremendous MSE values, which suggest the extremely wrong prediction, not to mention forecasting in future 12 months. In Figure B6-5, the prediction outcomes over test set go down drastically to the negative level throughout three models.

By diagnosing extreme down-going trend, it is found that the predictions go gradually biased with time being and even more so after reaching a specific point. Take model 2 for example, the 29^{th} estimated FTSE value is composed of underestimated EAFV, JQ2J and an overestimated K226, which refers to bias. We can basically conclude this as the composition of FTSE refers to the linear combination of forecasting/predicting results of dependent variables with multiplication of regression coefficients.

In sum, three models are all not appropriate to fit and predict FTSE data. As a result, the forecasting on 12 months is not done in this section as no model is selected during this stage. The biased estimation on FTSE indicates that the Exponential Smoothing Models we chose for regression forecasting are not appropriate. For recommendations, the regression forecasting should be used before the data transformation (e.g. logged, square-root or calendar transformation data) and detailed forecasting model selection (like Exponential Smoothing or even ARIMA models).

Reference

Brownlee, J. (2017) *How to Create an ARIMA Model for Time Series Forecasting in Python*. Available at: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/ (Accessed: 8 March 2020).
Brownlee, J. (2017). *A Gentle Introduction to SARIMA for Time Series Forecasting in Python*. Available at: https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/ (Accessed: 8 March 2020)
Lewis, C.D. (1982). *Industrial and business forecasting methods*. London: Butterworths.
Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998). *Forecasting: Methods and Applications*. Wiley.

Appendix A

1. ExponentialSmoothing_EAFVAutocorrelation_31324878.py:
The file includes the codes to generate time plot, ACF and PACF plot of EAFV.
2. ExponentialSmoothing_EAFVModelChoice_31324878.py:
The file includes the codes to generate four H-W potential models and calculation of errors we need for model selection.
3. ExponentialSmoothing_EAFVForecasting_31324878.py:
The file includes the codes to forecast future 12 months with 95% confident interval using the appropriate model for forecasting.
4. ExponentialSmoothing_K226Autocorrelation_31324878.py:
The file includes the codes to generate time plot, ACF and PACF plot of K226.
5. ExponentialSmoothing_K226ModelChoice_31324878.py:
The file includes the codes to generate six H-W potential models and calculation of errors we need for model selection.
6. ExponentialSmoothing_K226Forecasting_31324878.py:
The file includes the codes to forecast future 12 months with 95% confident interval using the appropriate model for forecasting.
7. ExponentialSmoothing_JQ2JAutocorrelation_31324878.py:
The file includes the codes to generate time plot, ACF and PACF plot of JQ2J.
8. ExponentialSmoothing_JQ2JModelChoice_31324878.py:
The file includes the codes to generate four H-W potential models and calculation of errors we need for model selection.
9. ExponentialSmoothing_JQ2JForecasting_31324878.py:
The file includes the codes to forecast future 12 months with 95% confident interval using the appropriate model for forecasting.
10. ExponentialSmoothing_K54DAutocorrelation_31324878.py:
The file includes the codes to generate time plot, ACF and PACF plot of K226.
11. ExponentialSmoothing_K54DModelChoice_31324878.py:
The file includes the codes to generate four H-W potential models and calculation of errors we need for model selection.
12. ExponentialSmoothing_K54DForecasting_31324878.py:
The file includes the codes to forecast future 12 months with 95% confident interval using the appropriate model for forecasting.
13. ARIMA_K54DDifferencing_31324878.py:
The files is to do 1^{st} and seasonal differencing on K54D and show their ACF and PACF graph.
14. ARIMA_ModelChoice_31324878.py:
The file is to calculate AIC of 256 potential models and helps us find the most appropriate one in this scope.
15. ARIMA_Modelling_31324878.py:
The file demonstrates codes to predict test set, evaluate the model and forecast future 12 months with 95% confidence interval.
16. Regression_Correlation_31324878.py
The codes is to generate the correlation plot and matrix of all regression variables.
17. Regression_ModelChoice_31324878.py
This is to compare three potential regression forecasting models and draw the time plot of training set along with real data.
18. Regression_TestSetPredict_31324878.py
This file is to use training set results to make the prediction on test set for test the model prediction (has visualisation).
19. Regression_Diagnosis_31324878
This code is to help us break down very biased prediction results and see when it starts to get extreme wrong prediction (Model 2 for example).

Appendix B

Section 0. Boxplot

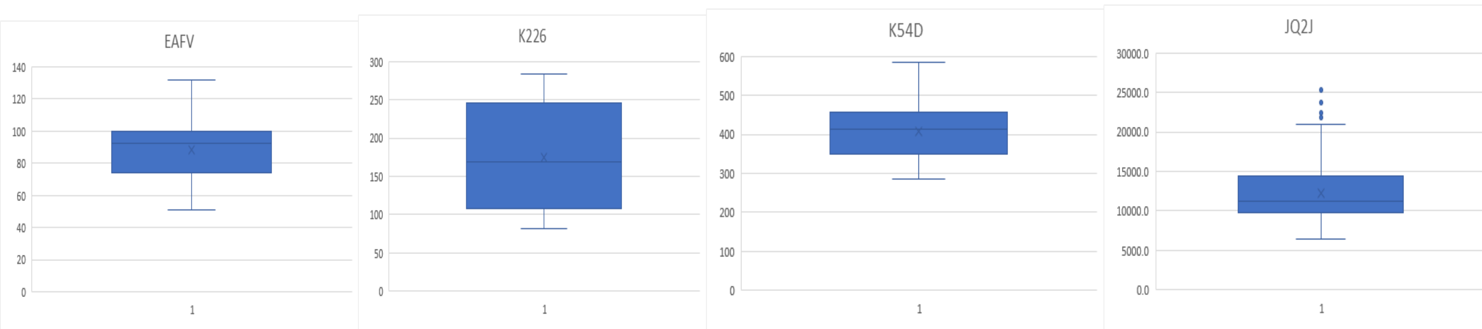

Figure B0. Boxplot for EAFV, K226, JQ2J and K54D (from left to right)

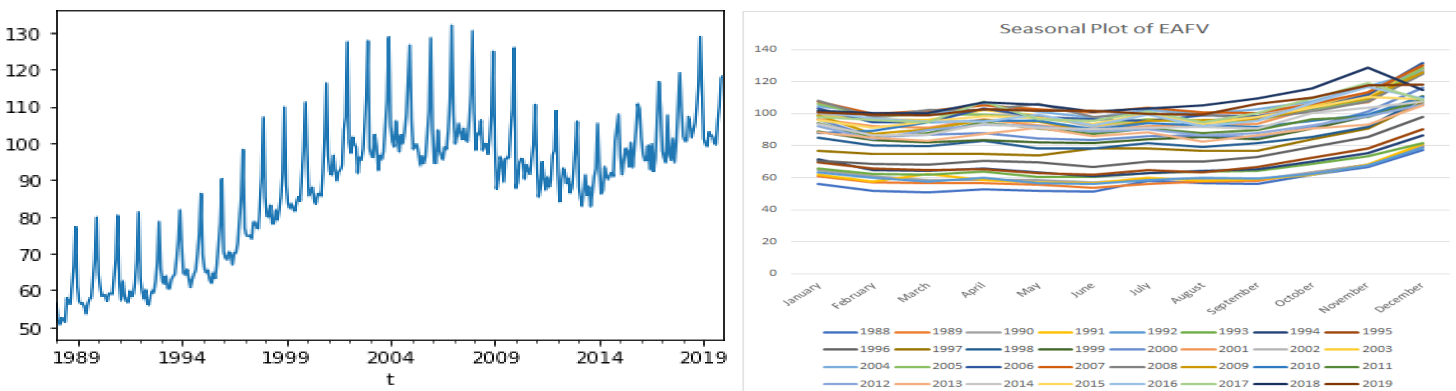

Figure B1-1. Time Plot and Seasonal Plot of EAFV

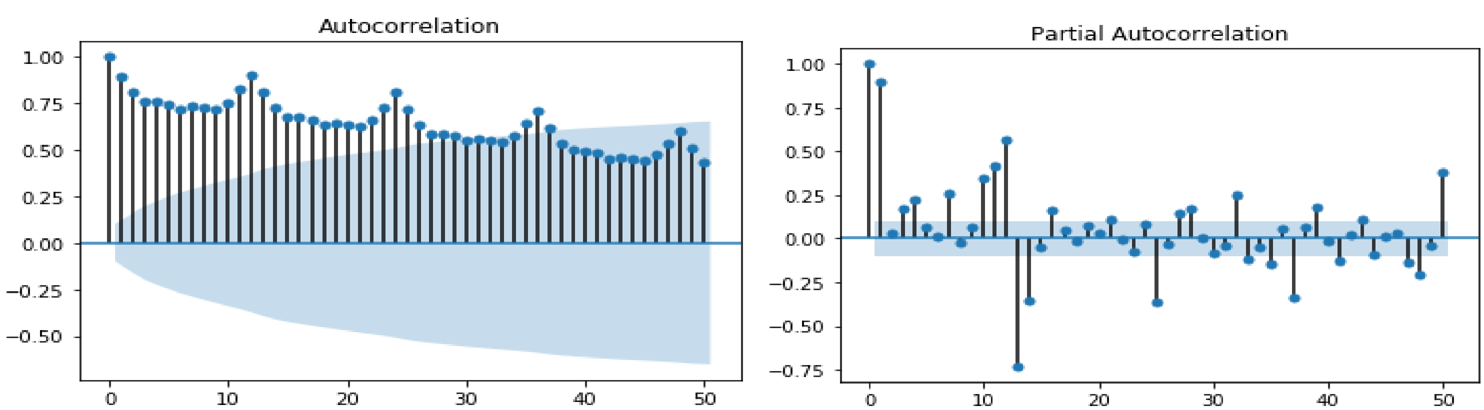

Figure B1-2. ACF and PACF of EAFV

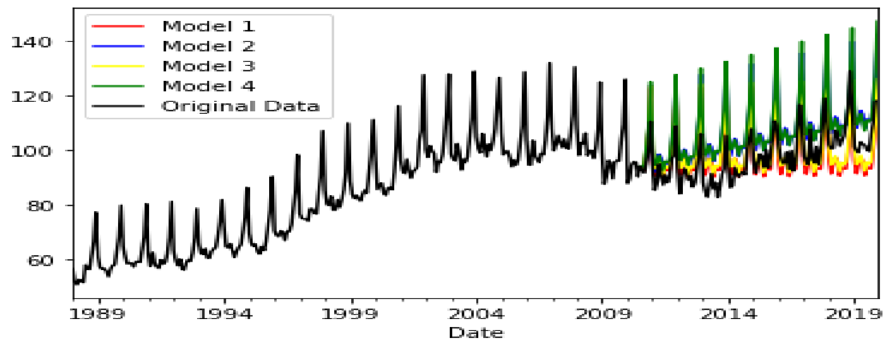

Figure B1-3. Training and Testing Set

```
Training Set of Model 1: MSE, MAE and MAPE are (5.574569989988328, 1.7148817434030768, 1.9996585698526919)
Training Set of Model 2: MSE, MAE and MAPE are (5.585452072405356, 1.7082518337491142, 1.9885758639169187)
Training Set of Model 3 MSE, MAE and MAPE are (4.72563088843821, 1.5791868655223655, 1.892743842297191)
Training Set of Model 4: MSE, MAE and MAPE are (4.671172917138801, 1.5683231906624295, 1.8755819393775144)
Test Set of Model 1: MSE, MAE and MAPE are (59.31707725327434, 6.19889440578868, 6.152337484187479)
Test Set of Model 2: MSE, MAE and MAPE are (117.54286041916072, 9.244763355939952, 9.4907263193728)
Test Set of Model 3: MSE, MAE and MAPE are (59.69457203470568, 6.191847479105143, 6.216402868814972)
Test Set of Model 4: MSE, MAE and MAPE are (145.09366143086814, 10.151873817800979, 10.398563441030065)
```

Figure B1-4. Errors

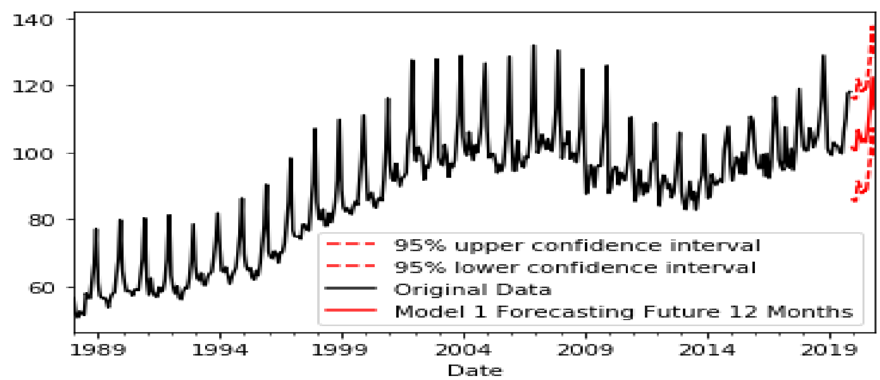

Figure B1-5. 12 Months Forecasting by Model 1 with 95% Confidence Interval

Section 2: K226

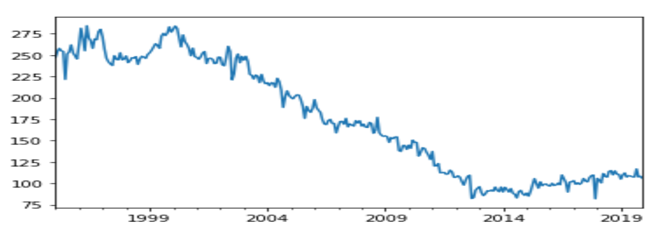

Figure B2-1. K226 Time Plot

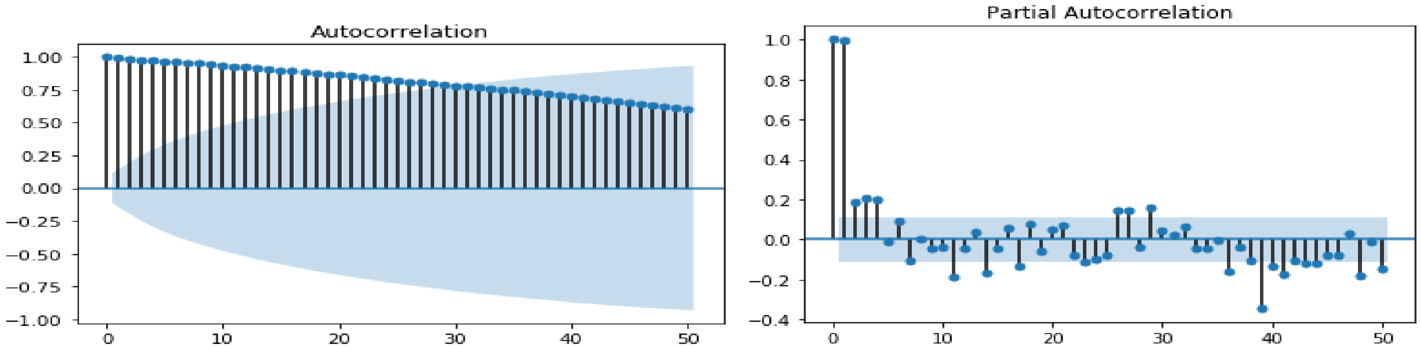

Figure B2-2. ACF and PACF of K226

```
Training Set of Model 1: MSE, MAE and MAPE are (62.835358107297566, 5.770007925230744, 2.846763027746834)
Training Set of Model 2: MSE, MAE and MAPE are (62.70546565061978, 5.798331046156437, 2.8621455463601166)
Training Set of Model 3 MSE, MAE and MAPE are (60.36792737920041, 5.72582037305l103, 2.826812225532l164)
Training Set of Model 4: MSE, MAE and MAPE are (60.397332933165046, 5.733313093889439, 2.8354568154347826)
Training Set of Model 5: MSE, MAE and MAPE are (60.174534461517204, 5.760549133597518, 2.844568928180792)
Training Set of Model 6: MSE, MAE and MAPE are (60.233399426564695, 5.748225765229088, 2.835066685831028)
Test Set of Model 1: MSE, MAE and MAPE are (2847.244270053646, 43.452975296797035, 41.828367201982566)
Test Set of Model 2: MSE, MAE and MAPE are (306.61800806209914, 14.537132230301332, 14.11261182340276)
Test Set of Model 3: MSE, MAE and MAPE are (2858.0182311076064, 43.54319265371228, 41.91304291956389)
Test Set of Model 4: MSE, MAE and MAPE are (192.51611970627013, 11.9145655709021, 11.777380169852075)
Test Set of Model 5: MSE, MAE and MAPE are (293.41545571762373, 14.340388436786522, 13.954753386173952)
Test Set of Model 6: MSE, MAE and MAPE are (2898.056951575575, 43.84685879971297, 42.208986414738)
```

Figure B2-3. Errors

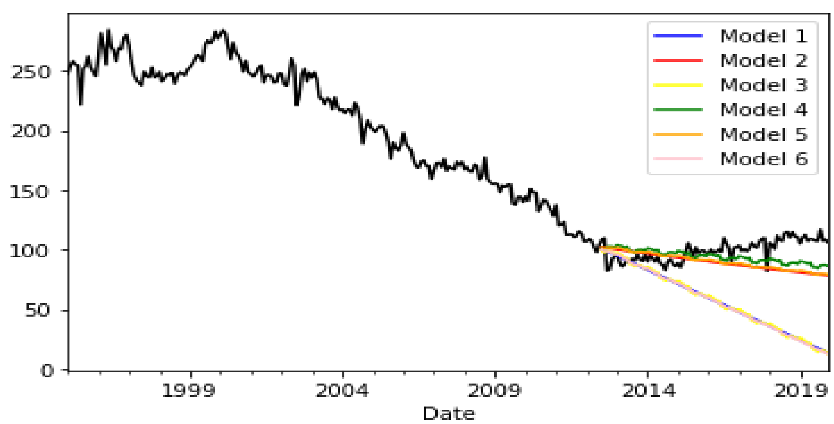

Figure B2-4. Training and Testing Set

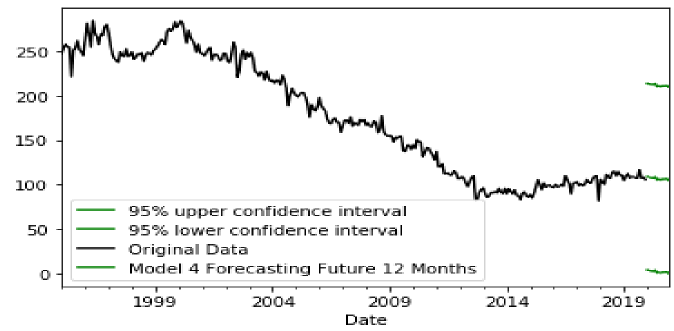

Figure B2-6. 12 Months Forecasting by Model 4 with 95% Confidence Interval

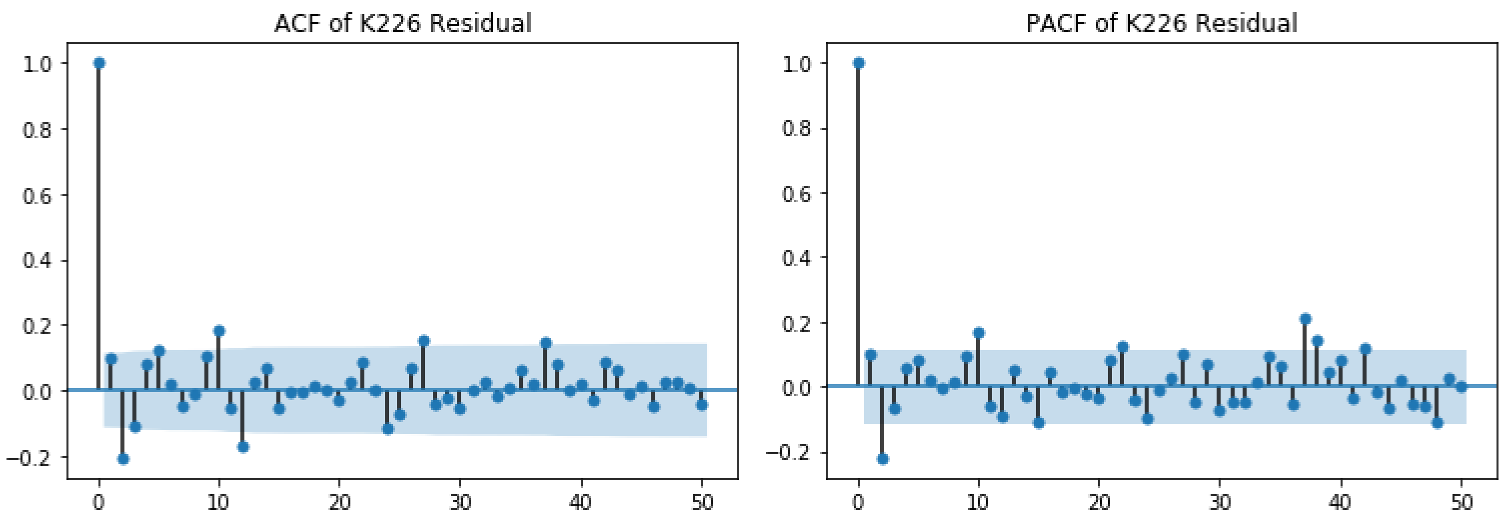

Figure B2-7. ACF and PACF of Residuals of Model 4

Section 3: K54D

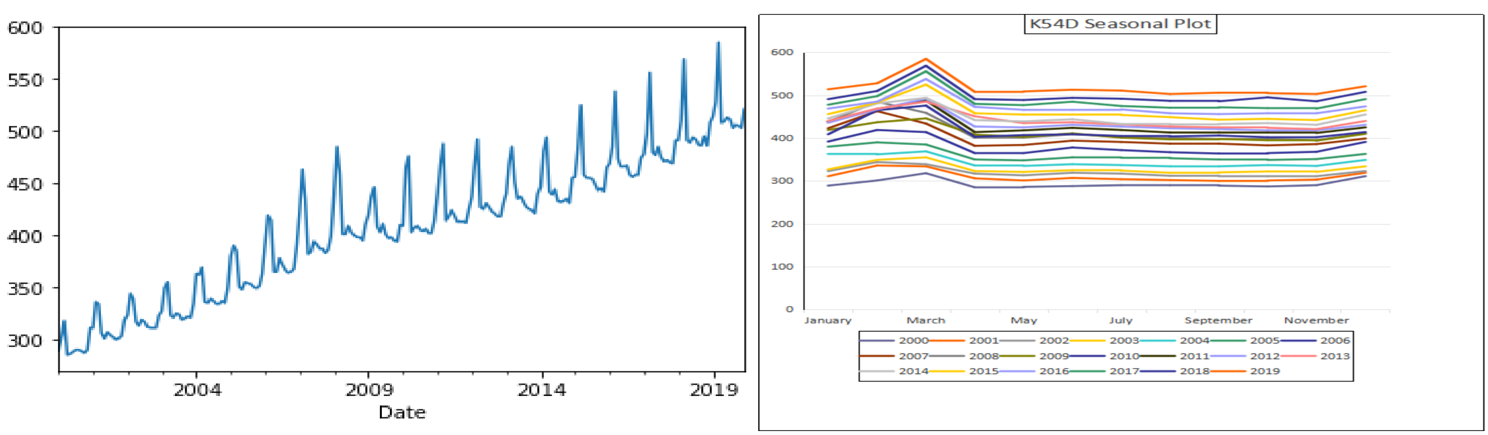

Figure B3-1. Time Plot and Seasonal Plot of K54D

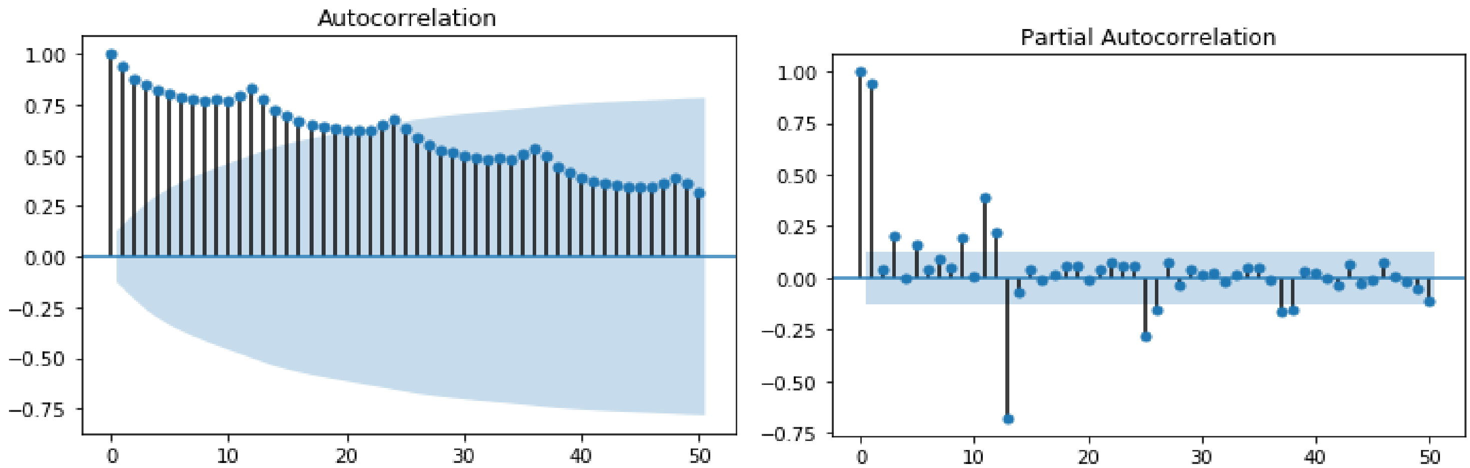

Figure B3-2. ACF and PACF of K54D

Training Set of Model 1: MSE, MAE and MAPE are (62.39027621016282, 4.5697989964630095, 1.155821052462358)
Training Set of Model 2: MSE, MAE and MAPE are (63.841240965775384, 4.722443376963556, 1.1964528782076664)
Training Set of Model 3: MSE, MAE and MAPE are (64.25384536886365, 4.677683750938316, 1.1824314721014642)
Training Set of Model 4: MSE, MAE and MAPE are (65.32510447059477, 4.758332067345933, 1.205685891868177)
Test Set of Model 1: MSE, MAE and MAPE are (86.31377451347548, 7.2037900347900665, 1.4520772562100877)
Test Set of Model 2: MSE, MAE and MAPE are (258.7869754628733, 12.786742377586936, 2.578241031728562)
Test Set of Model 3: MSE, MAE and MAPE are (86.31377451347548, 7.2037900347900665, 1.4520772562100877)
Test Set of Model 4: MSE, MAE and MAPE are (258.7869754628733, 12.786742377586936, 2.578241031728562)

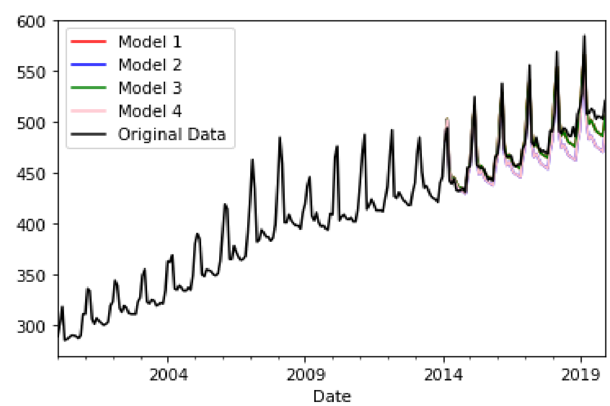

Figure B3-3 Training and Testing Errors, plus, Graphical Results of K54D

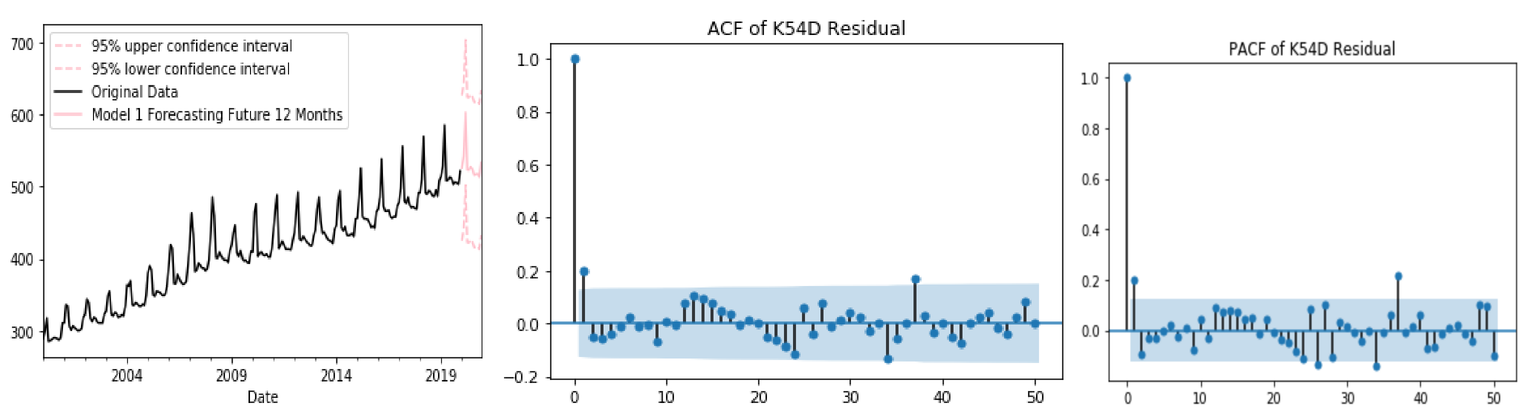

Figure B3-4. 12 Months Forecasting Results (95% Confidence Interval), ACF and PACF of Residuals from Model 1

Section 4: JQ2J

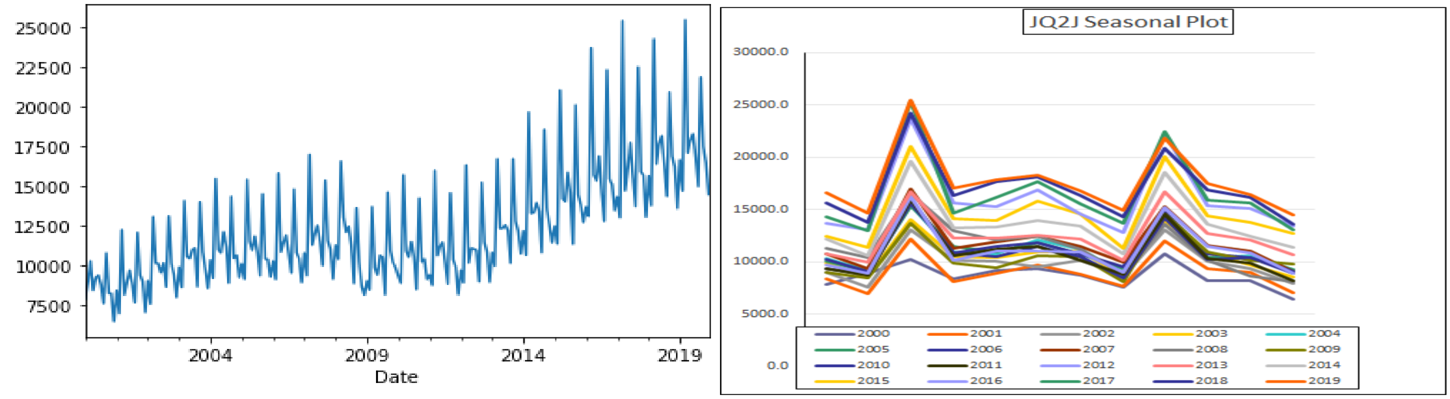

Figure B4-1. Time Plot and Seasonal Plot of JQ2J

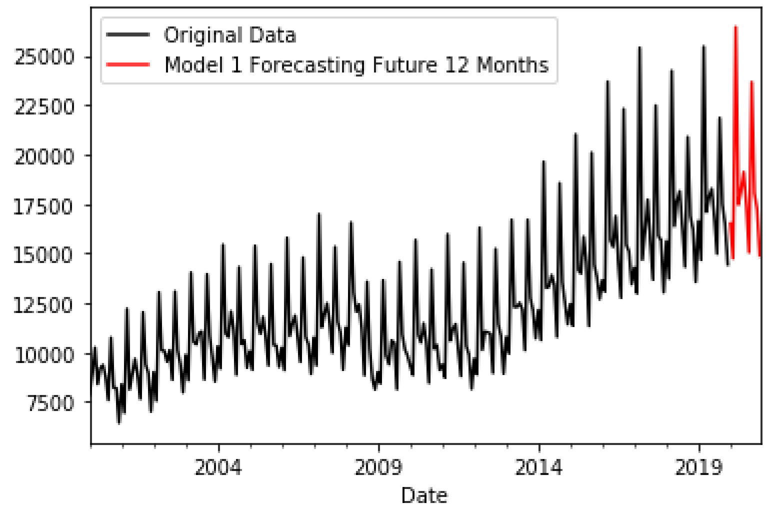

Figure B4-2. JQ2J Forecasting Results of Future 12 Months

2020-01-01	731763.278777	2020-01-01	-698754.008154
2020-02-01	730021.538033	2020-02-01	-700495.748898
2020-03-01	741711.684518	2020-03-01	-688805.602413
2020-04-01	732715.999815	2020-04-01	-697801.287116
2020-05-01	733454.410711	2020-05-01	-697062.876221
2020-06-01	734368.589595	2020-06-01	-696148.697336
2020-07-01	732657.277322	2020-07-01	-697860.009610
2020-08-01	730302.330132	2020-08-01	-700214.956799
2020-09-01	738922.396002	2020-09-01	-691594.890929
2020-10-01	733308.567514	2020-10-01	-697208.719417
2020-11-01	732489.819146	2020-11-01	-698027.467785
2020-12-01	730150.959571	2020-12-01	-700366.327360

Figure B4-3. Upper (left panel) and Lower (right panel) Bound of 95% Confidence Interval of Forecasting

Section 5: ARIMA

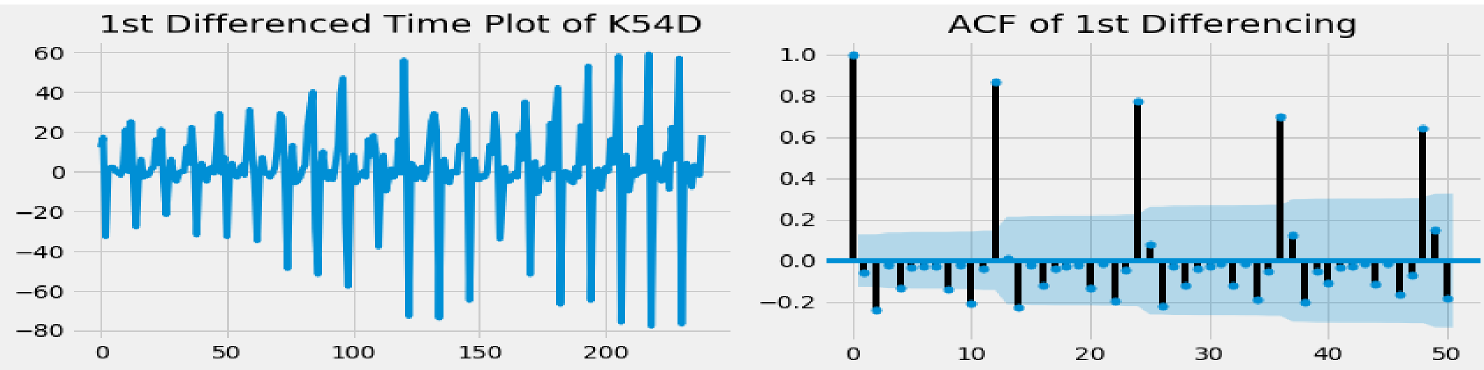

Figure B5-1. Time plot and ACF of 1st Differencing on K54D

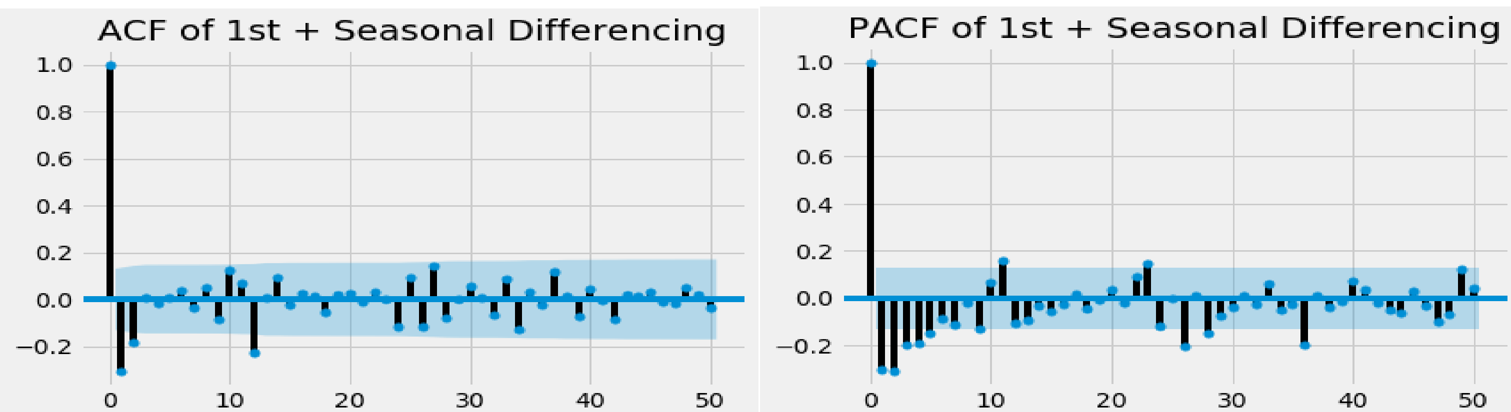

Figure B5-2. ACF and PACF Plot of 1st_seasonal Differencing K54D

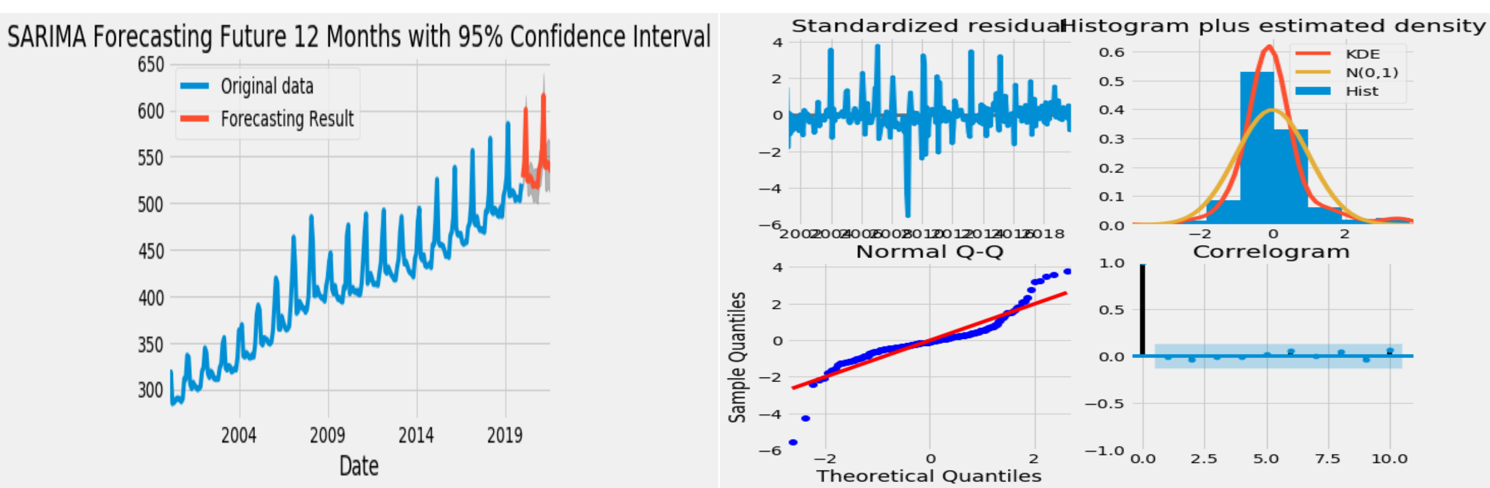

Figure B5-3. (S)ARIMA Forecasting with 95% Confidence Interval (left panel) and Residual Analysis (right panel). The residual analysis demonstrates the time plot, histogram, QQ plot and ACF plot of residuals to show fitted level of final model on K54D.

Section5: Regression

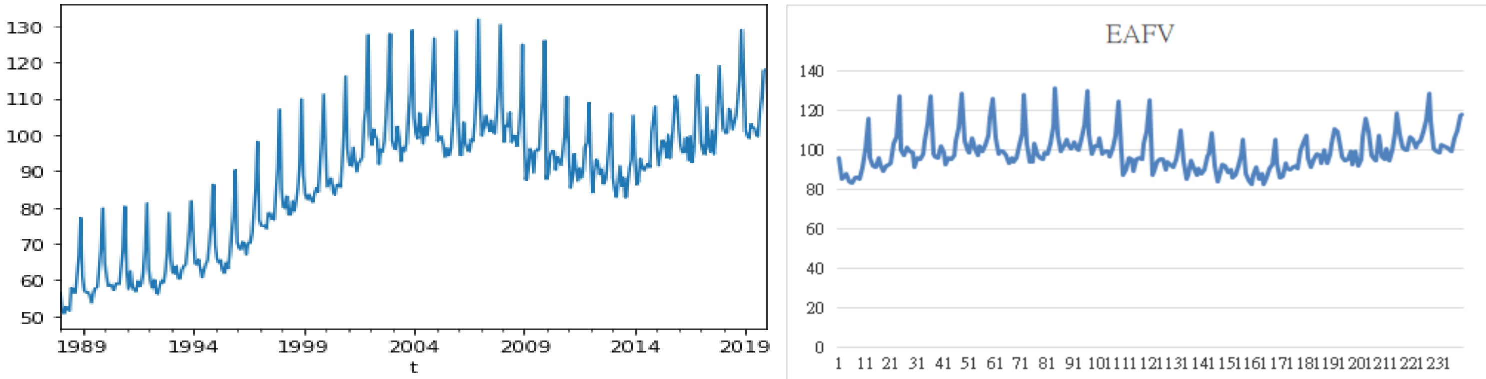

Figure B6-1. The Original EAFV (starts from Jan 1988) vs. EAFV Data for Regression (starts from Jan. 2000)

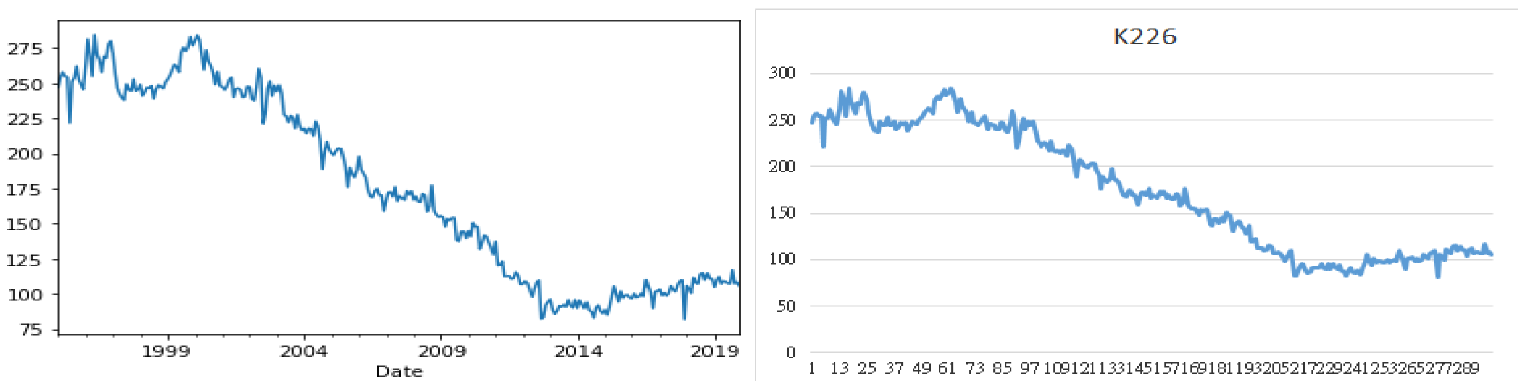

Figure B6-2. The Original K226 (starts from Jan 1995) vs. K226 Data for Regression (starts from Jan. 2000)

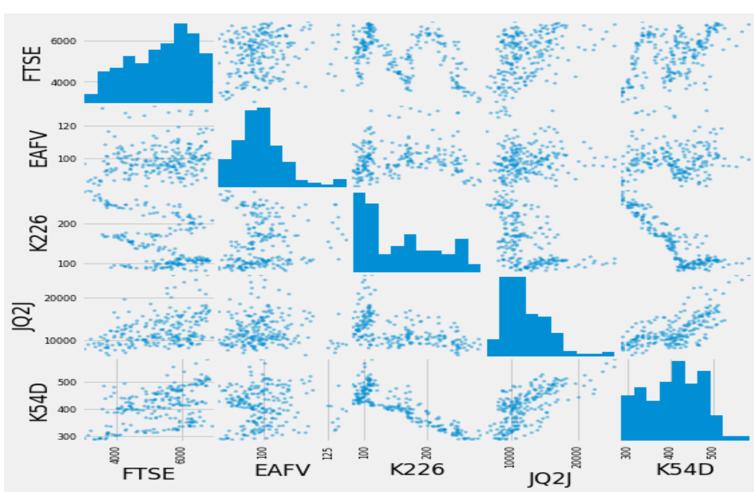

Figure B6-3. Correlation Plot

Variables included /Model	Model 1	Model 2	Model 3
EAFV	✓	✓	✓
K226	✓	✓	✓
JQ2J	✓	✓	✓
K54D	✓	✓	✓
D1		✓	✓
D2		✓	✓
D3		✓	✓
D4		✓	✓
D5		✓	✓
D6		✓	✓
D7		✓	✓
D8		✓	✓
D9		✓	✓
D10		✓	✓
D11		✓	✓
time			✓
R Square	0.273	0.304	0.319
Adj. R Square	0.260	0.257	0.270
AIC	3,916	3,928	3925
MSE in Training Set	1,207,066.97	3716553.11	6396019.74
MSE in Test Set	1,684,981,981.79	6,726,380,072.85	3,000,624,841.68

Table B6-1. Basic Information Table for Model Selection

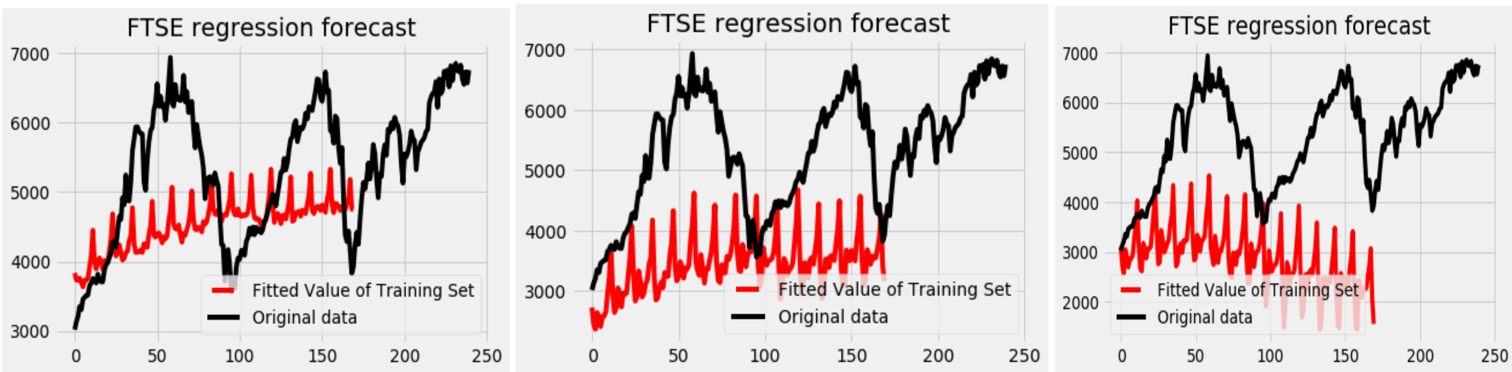

Figure B6-4. Regression Forecasting Fitting Results on Training Set of FTSE

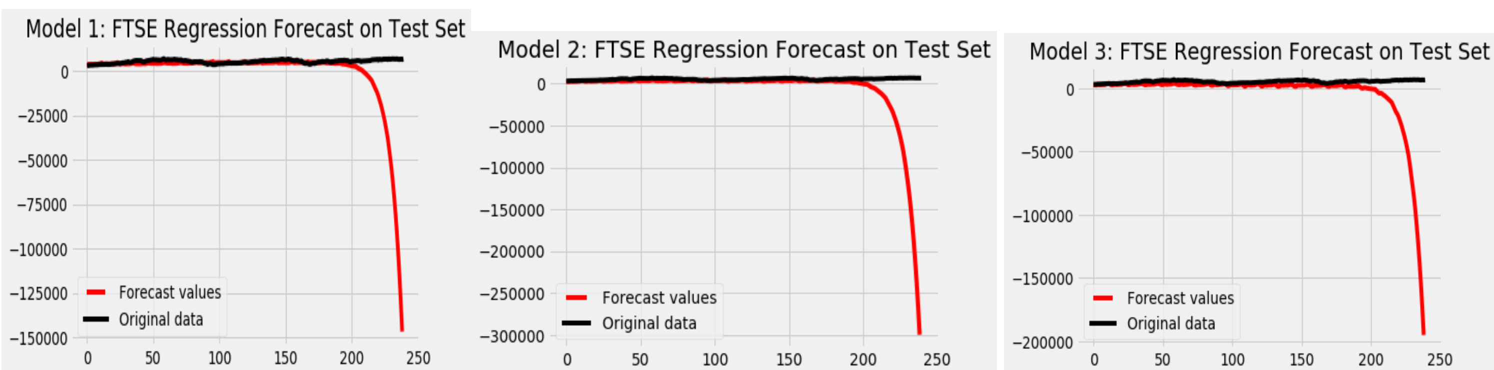

Figure B6-5. Regression Forecasting Prediction on Test Set of FTSE