Technical Report

1. Exponential Smoothing

Data preparation: The given datasets for analyses are, **K54D** (Monthly average of private sector weekly pay), **EAFV** (Retail sales index, household goods, all businesses), **K226** (Extraction of crude petroleum and natural gas), **JQ2J** (The manufacturing and business sector of Great Britain, total turnover and orders). The data in Excel files contains original data in spreadsheet ("data") and calendar adjusted modified data in spreadsheet ("adjusted"). In Python files, the chosen data frame was transformed to logarithm (transformed back with exponential) and root square (transformed back raised with power of two) if analysis gives a sign that transformation might be useful.

1.1. K54D Preliminary Analysis, Model Selection and Evaluation

Preliminary Analysis: According to Figure 1.1.1 and 1.1.2, calendar adjustment does not improve the times series, peaks increased, so the analyses will be used by original data series. By Figure 1.1.1, 2009 is the output by financial crisis, but not significant for analysis. Figure 1.1.3 shows distribution of different data ranges, which is not perfect horizontal. Figure 1.1.4 gives logarithm data, which did not improve the distribution and did not reduce the variance of time series. Figure 1.1.5 is square root data, which did improve the distribution and did not reduce the variance as well, so non transformed data will be taken for analysis. Figure 1.1.6 gives a clear picture that there is multiplicative seasonality with peak on march. By figure 1.1.1 and 1.1.6 we can see that trend is additive, since there is no exponential trend. By preliminary analysis, the data would give the best forecasting by applying Holt Winter with trend additive and multiplicative seasonal with non-modified data.

Models chosen and evaluation: By given preliminary analysis, applying mean square error (later **MSE**) to check which model is the best. Since trend is additive for sure, the seasonality would be good to check if additive seasonal forecast would improve the model. MSE of Holt Winters' trend additive, seasonal multiplicate (Later **HWM**) - **51.344501**, meanwhile Holt Winters' trend additive, seasonal additive (Later **HWA**) - **52.389810**, so the final chosen model is HWM with lowest MSE. Figure 1.1.7 shows forecast by splitting data and forecasting 2019 year to compare with actual data. The figure gives that forecasting (red) is very accurate. Figure 1.1.8 gives a picture of HWM forecasting for 2020 and comparing with original data, which is very similar and forecasting looks very accurate. Figure 1.1.9 gives a picture that residuals have not perfectly, but significant white noise, since there is a lag as a first value. The output is in 2009, which we expected to have based on Figure 1.1.1. The whole steps gave an idea that chosen forecasting is realistic and accurate, there is positive trend which is followed, with increasing seasonal variance.

1.2. EAFV Preliminary Analysis, Model Selection and Evaluation

Preliminary Analysis: According to Figure 1.2.1 and 1.2.2, calendar adjustment does not improve the variance of times series, the analyses will be used by original data series. By Figure 1.2.1, during 2007-08 data decreased in same way as in 2018-19, which means we can take the whole time series for analysis, since the times series might have cyclical pattern. Figure 1.2.3 shows distribution of different data ranges, which in not perfect horizontal. Figure 1.2.4 gives logarithm data, which does not improve the distribution to be more horizontal, same as figure 1.2.5 by square root data. However, the transform being preferable as it appears the variability is less time dependent by figures 1.2.4 and 1.2.5. Figure 1.2.6 gives a clear picture that there is seasonality with peak at the last months, but clear is it additive or multiplicative. By figure 1.2.1 and 1.2.6 we can see that trend is additive, since there is no exponential trend. By preliminary analysis, the data would give the best forecasting by applying Holt Winter with trend additive. Both seasonal (additive and multiplicative) worth to check with transformation.

Models chosen and evaluation: By given preliminary analysis MSE is tested to choose the most suitable model. For original data MSE (HWM: **7.561303**, HWA: **7.764618**), for logarithm data MSE (HWM: **7.711013**, HWA: **7.632788**), for root square data (HWM: **7.581327**, HWA**: 7.601815**). The final chosen model is HWM with original data which has lowest MSE. The given figure 1.2.7 shows forecast by splitting data and forecasting 2019 year to check with actual data. The figure gives that forecasting (red) is very accurate by shape, but values are usually higher by similar ratio. Figure 1.2.8

gives a picture of HWM forecasting for 2020 and comparing with original data, the forecast is closely matching. Usually forecast gives lower values where was high spike, but higher values when spike is lower than usual. Figure 1.2.9 gives a picture that residuals have not perfectly but significant white noise, since acf have lags (13, 22, 32) but it does not have a repeating trend. The whole steps gave an idea that chosen forecasting is approximately good, since it follows reduced trend, and forecast is accurate, because residuals have significant white noise.

1.3. K226 Preliminary Analysis, Model Selection and Evaluation

Preliminary Analysis: According to Figure 1.3.1 and 1.3.2, calendar adjustment does not improve the times series, the data variance increase, so original data will be used for analysis. A trend exists in periods 2000-2013 and 2013-2019, so in order to get more accurate recent forecast, I will use data from 2013. Figure 1.3.3 gives data from 2013 where is a clear trend. Figures 1.3.3 and 1.3.7 shows that data does not have at all. Figure 1.3.4 shows distribution of different data ranges and it is not perfect horizontal. Figure 1.3.5 gives logarithm data, which improves the distribution to be little more horizontal, but variances of time series is similar. Figure 1.3.6 root square does not improve distribution and variance of spikes. By preliminary analysis, data would give the best forecasting by applying Holt's linear exponential (later **LES**). It is worth to check if logarithm can improvement the model.

Models chosen and evaluation: By given preliminary analysis MSE is tested to choose the most suitable model. For original data MSE LES: **25.339950**, for logarithm data MSE LSE **25.624028**. The final chosen model is LES with original data which have the lowest MSE. Figure 1.3.8 gives a picture of HWM forecasting for 2020 and comparing with original data, which is accurate for known data, but the forecast is more showing a trend, not variance. Since data series are less, acf can give better picture for white noise. According to figure 1.3.9 residuals have perfect white noise and error term does not have autocorrelation, since all values are below confidence level, no lags. Even the forecast for 2020 does not seem good, only trend line, HW can improve forecast, because they have more parameters, but it could be coincidence, since the K226 time series does not satisfy condition to have seasonality. Additionally, with this kind of data, it is hard to forecast variance of forecast, it is more reasonable to show trend, so by the trend, forecast can be accurate, since residuals have white noise.

1.4. JQ2J Preliminary Analysis, Model Selection and Evaluation

Preliminary Analysis: According to Figure 1.4.1 and 1.4.2, calendar adjustment improves graph to be more linear in small parts, but not significantly, the analyses will be used by original data series. By Figure 1.4.1, during 2009 was decrease, but it does not affect to analyse data from 2009, since nowadays, data can decrease again by pandemic and recession. Figure 1.4.3 shows distribution of different data ranges, which in not perfect horizontal. Figure 1.2.4 logarithm data together with figure 1.4.5 root square data reduces spike variances (variability), so both of transformations can be taken for analysis. By figure 1.4.1 and 1.4.6 we can see that trend is additive, since there is no exponential trend. The seasonality exists with peaks on March and September, and should be multivariate. By preliminary analysis, the logarithm data (reducing peak variance by a lot by figure 1.4.4) would give the best forecasting by applying Holt Winter with trend additive. Both seasonal (additive and multiplicative) worth to check, which performs well, as well with modified data.

Models chosen and evaluation: By given preliminary analysis MSE is tested to choose the most suitable model. For original data MSE (HWM: **363640.481972**, HWA: **419336.930934**), for logarithm data MSE (HWM: **359418.852407**, HWA: **372042.175780**), for root square data (HWM: **365119.372735**, HWA**: 386461.655300**). The final chosen model is HWM with logarithm data which has the lowest MSE. Given figure 1.4.7 shows forecast by splitting data and forecasting 2019 year to check with actual data. The figure gives that forecasting (red) is very accurate by shape and values. Figure 1.4.8 gives a picture of HWM forecasting for 2020 and whole period which is similar to original. During the high spikes, forecast usually reduces the values comparing with the original. Figure 1.4.9 gives a picture that residuals does not have a white noise, acf have a lot of lags in beginning by repeating series. The whole steps gave an idea that chosen forecasting is not accurate since residuals have white noise, but as forecasting performed good 2019 forecast in figure 1.4.7 and for the whole time series in figure 1.4.8, forecasting can be accurate.

2. ARIMA Forecasting

Data Preparation: Data is used K54D original, without any modifications based on results got in paragraph 1.1, which described which type of data have the best perform (original, non-modified).

Preliminary Analysis: According to Figure 1.1.1 K54D data is not stationary, so the first difference for K54D is applied. The given Figure 2.1.1 1st difference data does not satisfy stationarity, since there is still seasonality. Figure 2.1.2 and 2.1.3 supports that data is still non stationary, seasonality exists, especially every 12 lag with spikes over 95% confidence level. To solve this problem, in figure 2.1.4 data modified by seasonality and 1st difference. As the result stationarity condition is satisfied and parameters are taken for preliminary analysis for Seasonal ARIMA (SARIMA). By satisfying stationarity, Figure 2.1.5 with support of Figure 2.1.6 shows MA exist, as values in beginning are negative and reducing closer to 0 with time. MA parameter in ARIMA will be given 1, while AR is 0, as usually it being opposite to MA and there are no signs in Figures 2.1.5 and 2.1.6. that AR exists. Additionally, we applied seasonal and first difference, so in ARIMA, these parameters will be given 1. Since we applied seasonality, ARIMA model need satisfy seasonal parameters. By given analysis, the preliminary model for SARIMAX is (0,1,**1**, 0,1,**1**, 12) with AIC **1557.6695993093**. Ljung-Box prob(Q) value is **0.57** which means we have to take zero hypothesis and data is not correlated with its own time, so we can take the model for forecasting.

Models chosen and evaluation: To make sure it is the most optional, it is good to try different combinations with parameters by varying by 1 value of each AR and MA in (0,1,1, 0,1,1) except the middle numbers, because it is critical to keep it, to satisfy stationarity. The criteria to select ARIMA is the lowest Akaike's Information Criterion (**AIC**) to make sure model is not modelling "randomness", the lowest is the best model. The new optimal ARIMA model is given (0,1,**2**, 0,1,**2**, 12) with the lowest AIC: **1547.7248873182734** with higher MA values (2) which will be chosen to forecast. Figures 2.2.1 and 2.2.2 shows that residuals have white noise and 2009 is outlier, but not significant. In figure 2.2.1 all values are in confidence level, which suggest data has white noise. Additionally, Ljung-Box prob(Q) value is **0.97** which is higher than previous model, which very confidentially shows white noise. Given figure 2.2.3 gives a forecast of ARIMA model with upper and lower bounds from 2001-02-01 together with 2020 forecasted period. Additionally, forecasted data from ARIMA model is taken from 2001-02-01 for showing forecast, because data was used for seasonal difference, so the first year could not be forecasted. In general forecasting is realistic comparing with actual data, since it is following positive trend with seasonality and similar values. Upper and lower bounds are not too big, so forecasted data might not vary a lot.

Comparing models in particular instance: As a comparison ARIMA and HWM forecasting, figure 2.3.1 shows that both models forecasting very similar. MSE is calculated by taking forecasted values from 2001-02-01, since reasonable ARIMA forecasting starts from mentioned year. As a result, MSE for ARIMA: **51.29553178471743**, while for MSE for HWD: **53.700476779490906**, which means ARIMA model have better forecasting. As for 2020 forecasting, figure 2.3.2 gives a view that both models have similar values, only HWM have little big higher variance between values. As comparison, with actual data, figure 2.3.3 shows split data forecasting for 2019. It gives a view that both models forecasting very accurate, during May-July, HWM have better forecasting, while for the rest of period, ARIMA. By taking small details, ARIMA gives better prediction.

Comparing models in general terms: ARIMA requires data to be stationary and in general model can be more flexible, since data does not need to have seasonality and trend (not like Holt Winters), as long as stationarity satisfied. ARIMA have more parameters to test, it might take more time, especially when comparing different SARIMAX parameters, but the result can be more accurate comparing with Holt Winters. Additionally, ARIMA can show upper and lower bounds, what variance regarding forecasted value might be. Holt Winters is specific case of ARIMA, have to be trend and seasonality, which can be easily and faster tested with different parameters (additive and multiplicative). To choose a model, it depends of data, but if there are enough resources and time, ARIMA model could give better forecast, but if forecast is needed to get a broad picture, a quick look, Holt Winters is good.

Regression prediction

Data Preparation: The given data set is **FTSE** (UK Footsie 100 share index) which is dependent variable where data start from 2000-01-01. All the other independent variable dataset was modified to match with FTSE, to have the same length data from 2000-01-01.

Preliminary Analysis: Figure 3.1.2 gives original FTSE data with trend while figure 3.1.3 shows calendar adjusted data which does not improve times series, data does not become more stable with less variance, so original data will be taken for analysis. Other important factor the last data with similar cyclical trend was from 2010 in figure 3.1.4, this length of data could give more accurate results. Before constructing regression model, figure 3.1.1 gives correlations between variables. There are correlations between ftse, k226, jq2, k54d each other, which means there will be multicollinearity. eafv does not have reasonable correlation with ftse and others, but since multicollinearity exists, all p-values might will not be true since independent variables affecting each other, so all of them are taken into regression model. If eafv for sure does not have impact at all to anybody, it will have high p-value. Additionally, dummy variables (months) are taken for analysis together with time, to test different regression models.

Models chosen: The priority to select the most suitable regression model to predict FTSE is the lowest AIC and reasonable R squared (proportion of dependent variable explained by independent variables, expecting to have more than 0.5 (50%). If variable exist with higher p value than significant 0.05, variable is taken out from regression model (backward selection) to improve AIC. According to table 1 in B appendix, 1st model with all variables gives r-square below 0.5, which is low and by applying backward selection, 2nd model gives only k54d and jq2j, with similar lower AIC and lower r-squared. To improve the model, time was added in 3rd model, where AIC decreased and increased r-square (50.2%), which improves the model. 4th model takes out not significant k54d, jq2j (multicollinearity), which improves model not very significantly but it is worth to test in analysis. 5th model add months as dummy variables, which should be treated as categorical variable and it is not appropriate for linear regression. For analysis it can be taken to test performance as different type of regression model, while 6th model improves AIC with all dummy variables by taking out k54d. 7th model gives data from 2010 to satisfy figure 3.1.4 positive trend condition. All variables are significant with much lower AIC and higher r-squared, it improves the model a lot. 9th model is applied backward selection for 8th model where data is from 2010 with added time and AIC is the lowest, r-squared the highest comparing with all given models. Models 4, 6, 7, 8 from table 1 were selected to test different significant linear models.

Final chosen model

Model	Formula	AIC	R-squares
4	ftse = b0 + b2***eafv** + b3***k226** + b5***time**	3844	0.501
6	ftse = b0 + b2***eafv** + b3***k226** + b4***jq2j** + b5***time** + b6***month1** + b7***month2** + … + b16***month11**	3851	0. 536
7 (from 2010)	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j**	1805	0.641
9 (from 2010)	ftse = b0 + b3***k226** + b5***time**	1730	0.801

Suitability for prediction: Figure 3.2.1 shows 6th regression model with months, lower and upper bounds (90% confidence level). Regression follows trend, but not short-time trends. Figure 3.2.2 is 4th model have similar shape as figure 3.2.1 but with lower variance. Dummy variables give more variance, but both models are not good for prediction, since it ignores short-time trends, not matching with actual data, only follows long-term trend. To have better forecast, figure 3.2.3 gives 7th model, while figure 3.2.4 gives 9th model. In general, figure 3.2.4 is a trend line, not cyclical trend it should be. Meanwhile Figure 3.2.3 seems to give the best prediction, but it has seasonality and variance big as cyclical trend. Figure 3.3.1 and figure 3.3.2 gives idea about residuals of 7th model which shows that prediction does not have white noise. Same conclusion for 9th model according to figures 3.3.3 and 3.3.4, the model does not have white noise. It gives a conclusion that given variables cannot predict FTSE by regression model, only a long-trem trend, because of problems as multicollinearity (especially with k54d, dummy variables and time), big upper and lower bounds, residuals do not have white noise and specific predicted values are too much different from actual. Figure 3.2.3 could reflect a true forecast, since it takes all 4 given variables, but they have multicollinearity between them and does not have white noise.

Appendix A

All Excel files have spreadsheets "data" with original data and "adjusted" with calendar adjusted data.

For corresponding methodology: **ExponentialSmoothing** (Let call **EXS**), all specific tasks for K54D, EAFV, K226, JQ2J (Let call **X**) are the same just for different dataset, (Let call student ID, **ID**).
EXS_XTimePlot_ID.py – Plotting time series of given data together with trend from decomposition.
EXS_K2262013Plot_ID.py – Plotting times series of K226 from 2013 with trend from decomposition.
EXS_XSeasonsPlot_ID.py – Transforming data into time series by months and plotting by pink colour. The darker pink, the more year is closer to 2019, while every line gives a year next to it.
EXS_XCalendar_ID.py – Taking calendar adjusted data and plotting time series.
EXS_Xhistogram_ID.py – Taking original data and plotting histogram of different range of value.
EXP_XLogarithm_ID.py – Taking original data, transforming to logarithm and plotting time series with histogram of different range of values.
EXP_Xsqrt_ID.py - Taking original data, transforming to square root and plotting time series with histogram of different range of values.
EXP_XMSE_ID.py – Forecasting by selected exponential forecasting models, calculating errors and printing all MSE values.
EXS_XForecast19_ID.py (K226 does not have) – Splitting data by not taking last 2019-year, training Exponential smoothing model to predict 2019 and plotting together with true 2019 data.
EXS_XForecast20_ID.py – forecasting by exponential smoothing method and plotting for the whole data series and for 2020. In the same graph, plotting original data series.
EXS_XResiduals_ID.py – Taking errors from forecasting and plotting time series with acf of residuals.
Summary of ARIMA Python description
ARIMA_1stDiffAcfPacf_ID.py – Taking first difference of K54D, plotting acf, pacf and times series of result.
ARIMA_ SeasDiffAcfPacf_ID.py – Taking seasonal difference with first difference of K54D, plotting acf, pacf and times series of result.
ARIMA_OptimalSarimax_ID.py – Calculating automatically optimal SARIMAX parameters by prioritising lowest AIC.
ARIMA_Sarimax-Summary_ID.py – two SARIMAX models by chosen and optimal parameters with summary for getting Ljung-box probability Q.
ARIMA_ResidualsPlot_ID.py – Plotting residuals of ARIMA model which has optimal parameters
ARIMA_SarimaxPlot_ID.py – Plotting ARIMA forecasting for all series and 2020 with upper and lower bounds.
ARIMA_Comparing2020_ID.py – 1st graph with models ARIMA, HWM fby orecasting only 2020. 2nd graph is with models ARIMA, HWM by forecasting 2020 and plotting all years including 2020.
ARIMA_Comparing2019_ID.py – Plotting one graph with ARIMA, HWM forecasting 2019 and comparing with original data.
ARIMA_ComparingMSE_ID.py – Calculating and printing MSE of ARIMA and HWM for K54D.
Regression Python. Description of Regression models can be found in Appendix B, 15 page, Table 1.
Regression_FTSEPlot_ID.py – FTSE time series plot with trend from decomposition
Regression_FTSECalendar_ID.py – calendar adjusted FTSE time series plot
Regression_FTSE2010Plot_ID.py – FTSE time series plot from 2010 with trend from decomposition
Regression_Correlation_ID.py – Plotting scatter of all times series together with correlation between
Regression_RegressionModel4_ID.py – Regression models and backward selection to get Model 4
Regression_RegressionModel6_ID.py – Regression models and backward selection to get Model 6
Regression_RegressionModel7_ID.py – Regression models and backward selection to get Model 7
Regression_RegressionModel9_ID.py – Regression models and backward selection to get Model 9
Regression_Model4Forecast_ID.py – Regression models 4 forecast calculation and plot
Regression_Model6Forecast_ID.py – Regression models 6 forecast calculation and plot
Regression_Model7Forecast_ID.py – Regression models 7 forecast calculation and plot
Regression_Model9Forecast_ID.py – Regression models 9 forecast calculation and plot
Regression_Model7Residuals_ID.py – Regression models 7 forecast acf and time series of residuals
Regression_Model9Residuals_ID.py – Regression models 9 forecast acf and time series of residuals

Appendix B
1. Exponential Smoothing

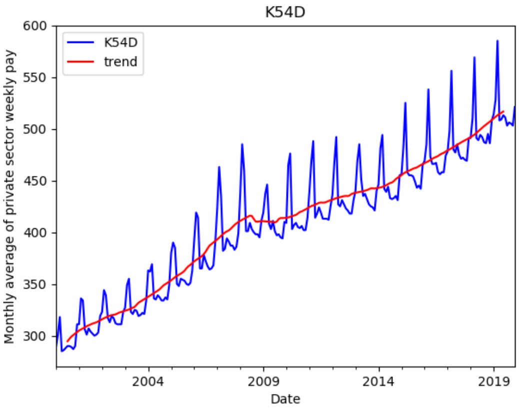


Figure 1.1.1 K54D Original data

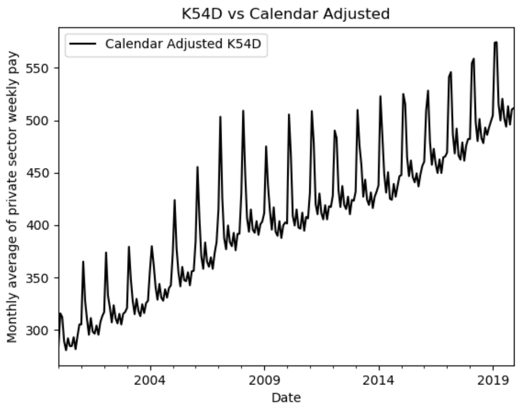


Figure 1.1.2 K54D Calendar Adjusted data

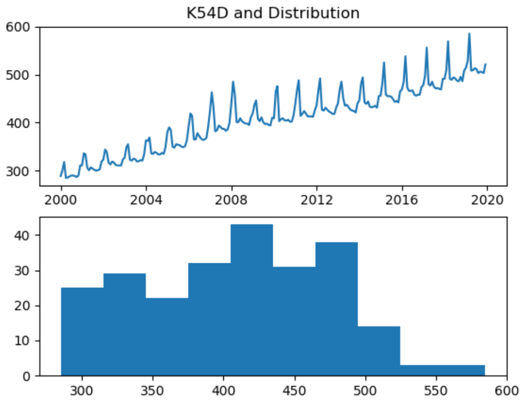


Figure 1.1.3 K54D data Distribution

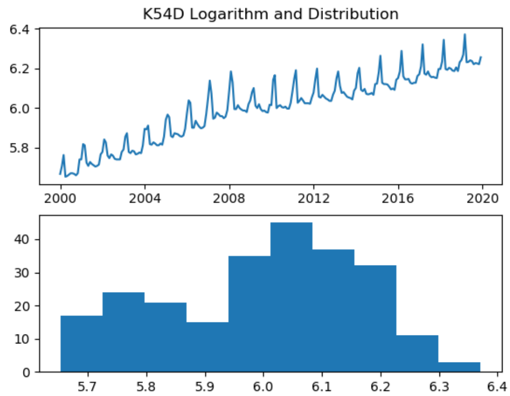


Figure 1.1.4 K54D Logarithm Distribution

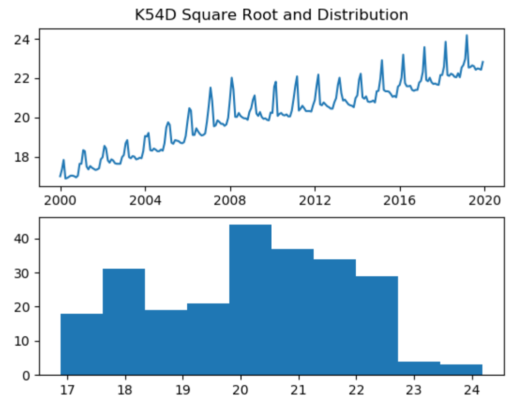


Figure 1.1.5 K54D Square Root Distribution

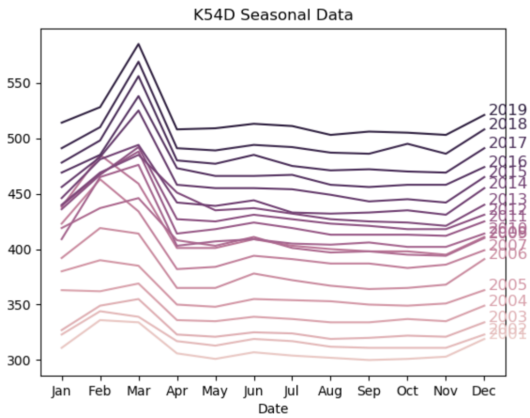


Figure 1.1.6 K54D Decomposition

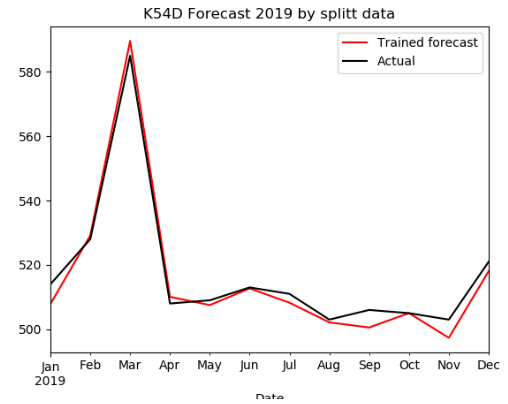


Figure 1.1.7 K54D Train Test Data Forecast

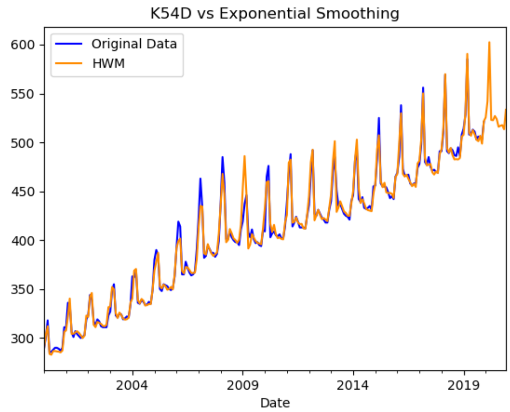


Figure 1.1.8 K54D Forecast 2020

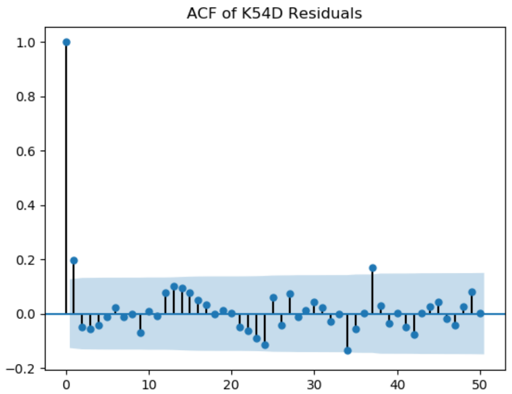


Figure 1.1.9 ACF of K54D Residuals

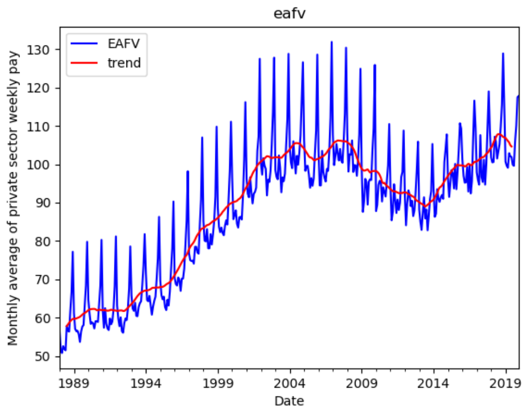


Figure 1.2.1 EAFV Original data

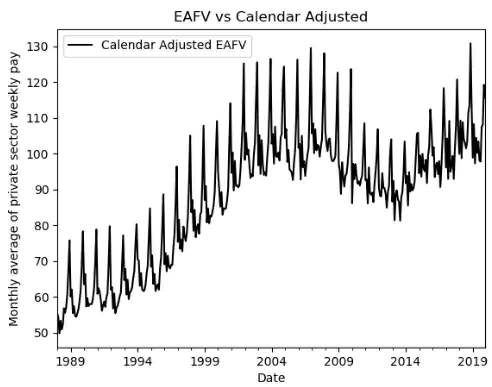


Figure 1.2.2 EAFV Calendar Adjusted data

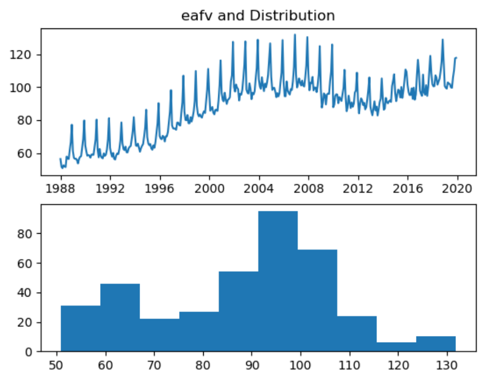


Figure 1.2.3 EAFV Square Root Distribution

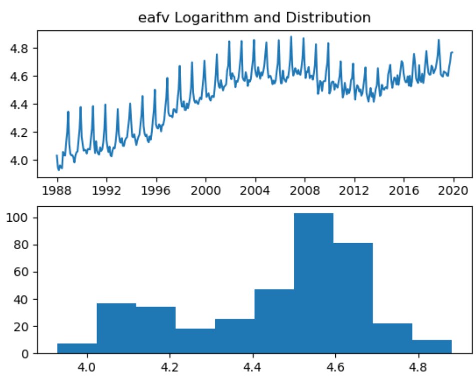


Figure 1.2.4 EAFV Logarithm Distribution

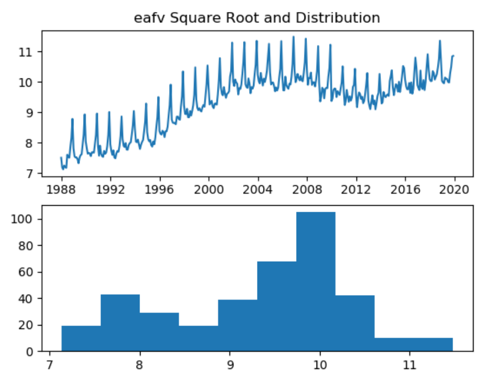


Figure 1.2.5 EAFV Square Root Distribution

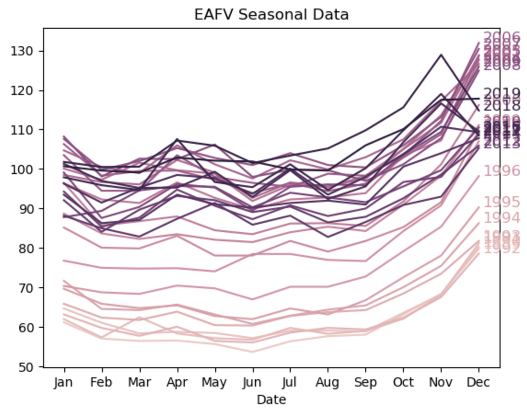


Figure 1.2.6 EAFV Decomposition

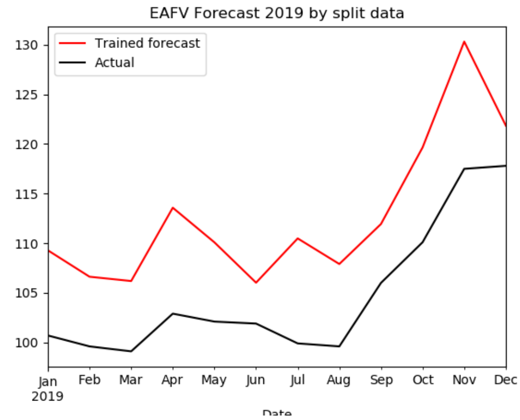


Figure 1.2.7 EAFV Train Test Data Forecast

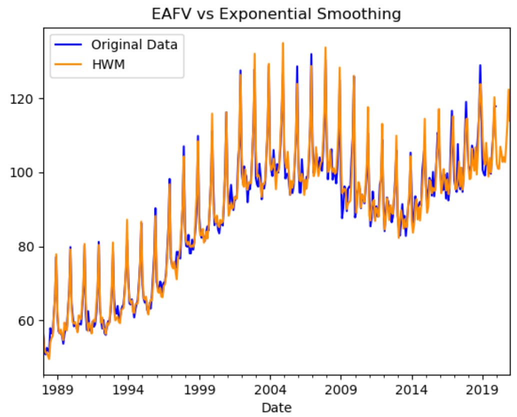


Figure 1.2.8 EAFV Forecast 2020

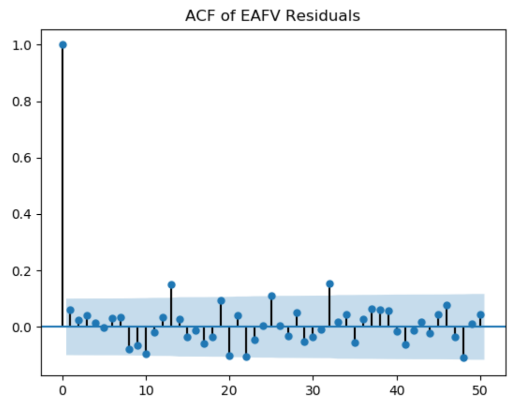


Figure 1.2.9 ACF of EAFV residuals

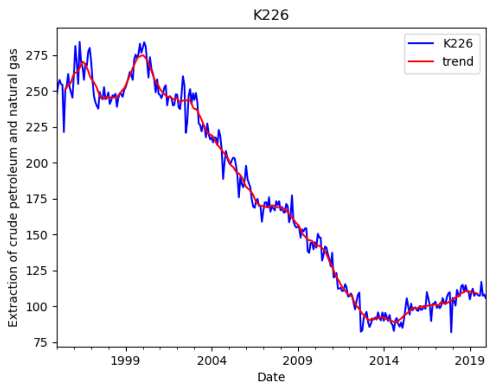


Figure 1.3.1 K226 Original data

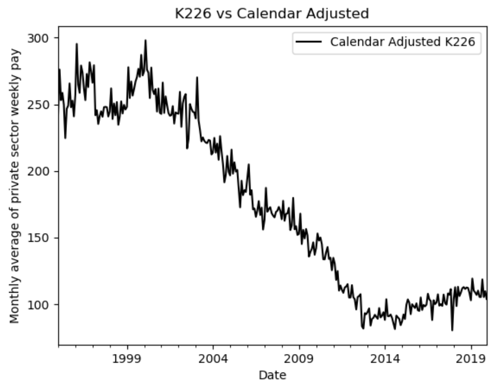


Figure 1.3.2 K226 Calendar Adjusted data

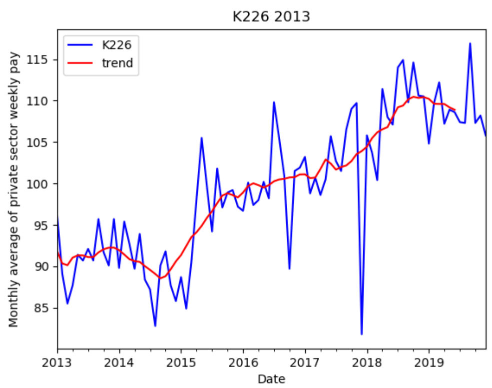


Figure 1.3.3 K226 data from 2013

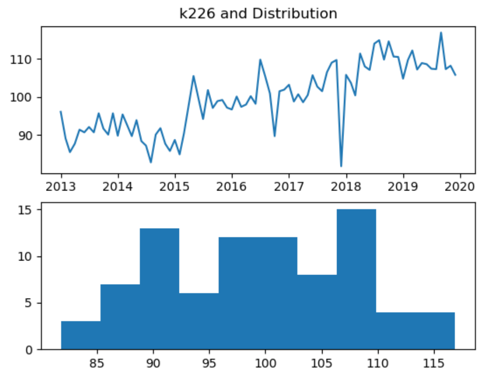


Figure 1.3.4 K226 data Distribution

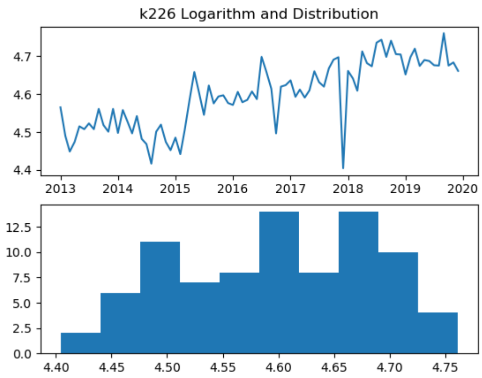


Figure 1.3.5 K226 Logarithm Distribution

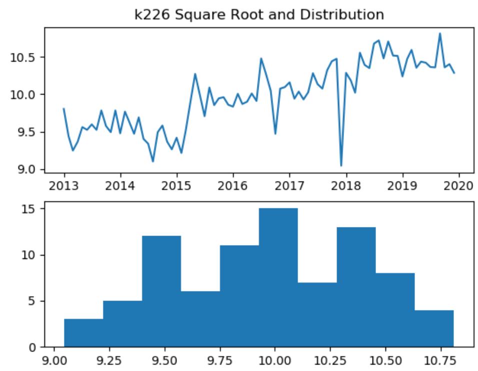


Figure 1.3.6 K226 Square Root Distribution

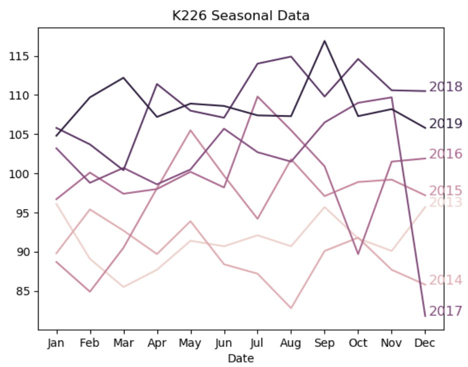


Figure 1.3.7 K226 Decomposition

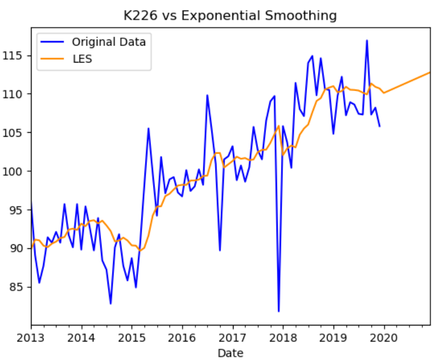


Figure 1.3.8 K226 Forecast 2020

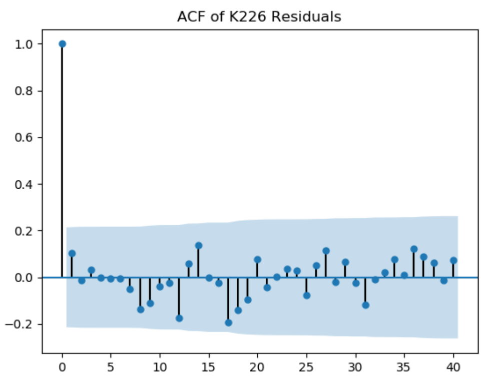


Figure 1.3.9 K226 ACF of Residuals

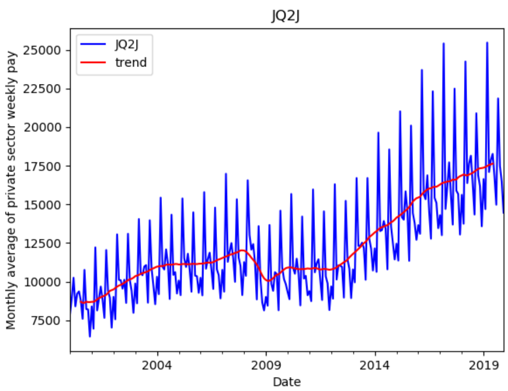


Figure 1.4.1 JQ2J Original data

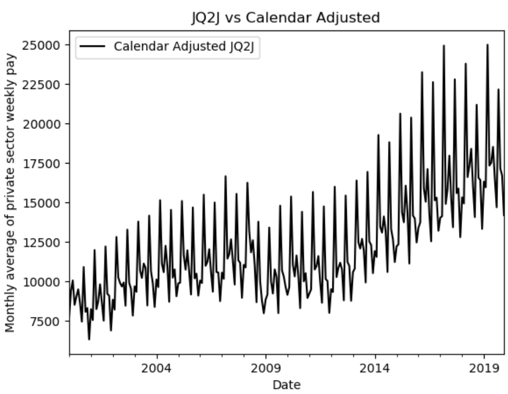


Figure 1.4.2 JQ2J Calendar Adjusted data

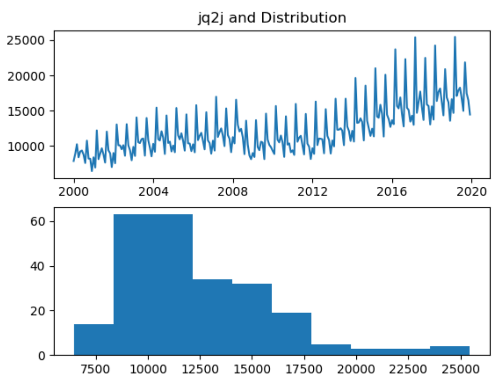


Figure 1.4.3 JQ2J data Distribution

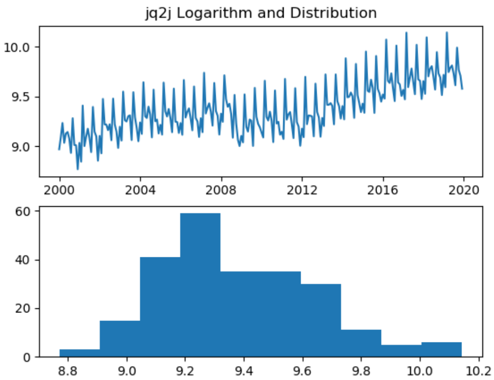


Figure 1.4.4 JQ2J Logarithm Distribution

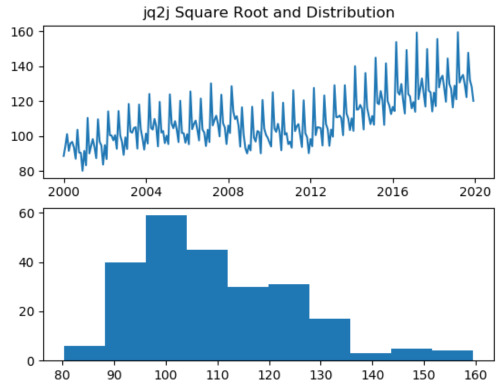


Figure 1.4.5 JQ2J Square Root Distribution

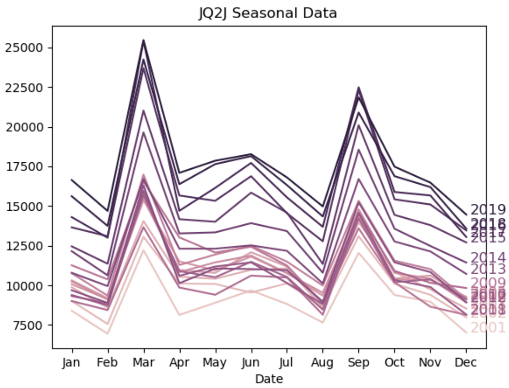


Figure 1.4.6 JQ2J Decomposition

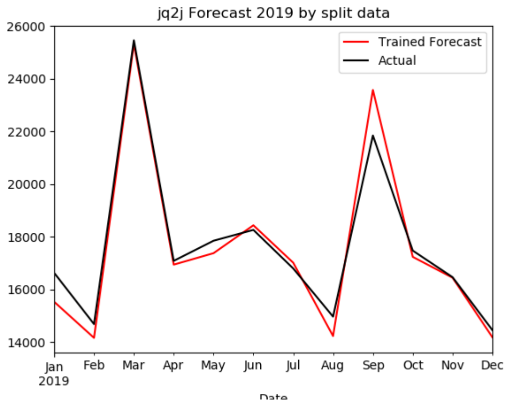


Figure 1.4.7 JQ2J Train Test Data Forecast

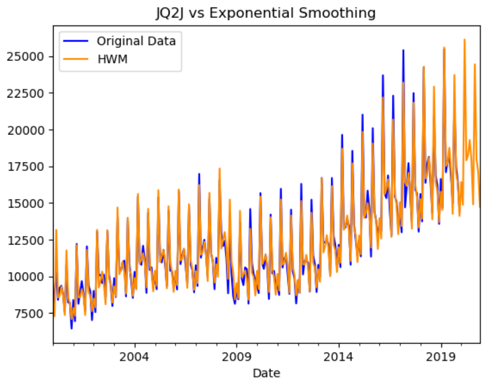


Figure 1.4.8 JQ2J Forecast 2020

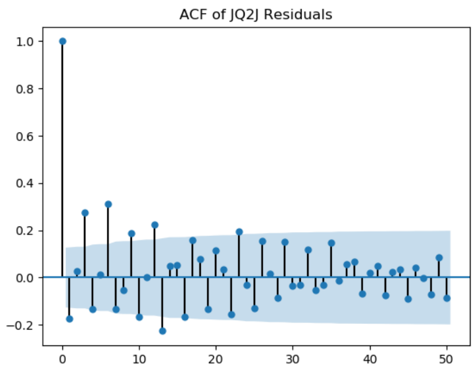


Figure 1.4.9 JQ2J Time plot of Residuals

2. ARIMA

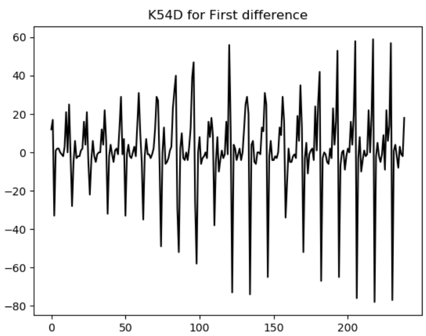


Figure 2.1.1 K54D 1st difference

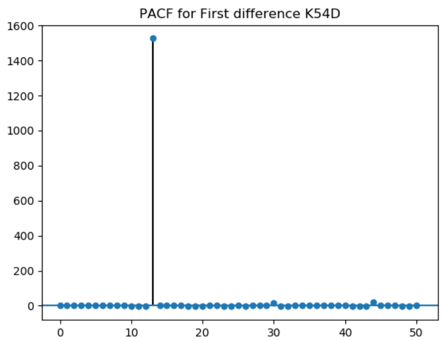


Figure 2.1.2 PACF Residuals, 1st difference

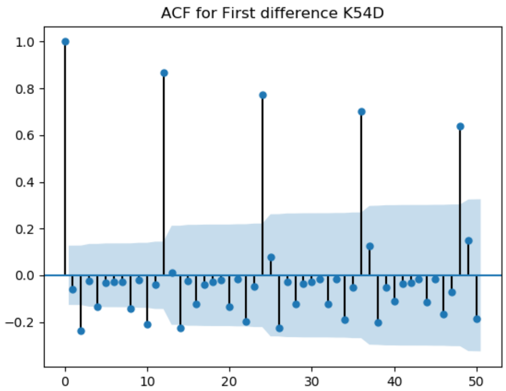


Figure 2.1.3 ACF Residuals, 1st difference

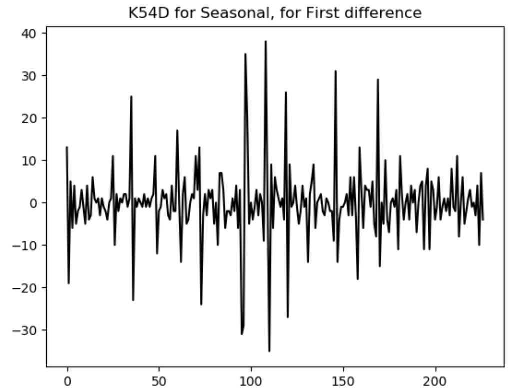


Figure 2.1.4. K54D seasonal 1st difference

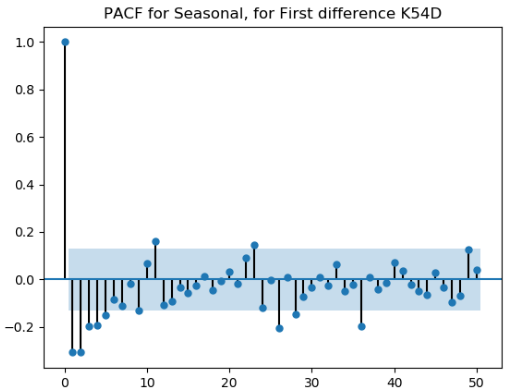


Figure 2.1.5 PACF Residuals, seasonal 1st difference

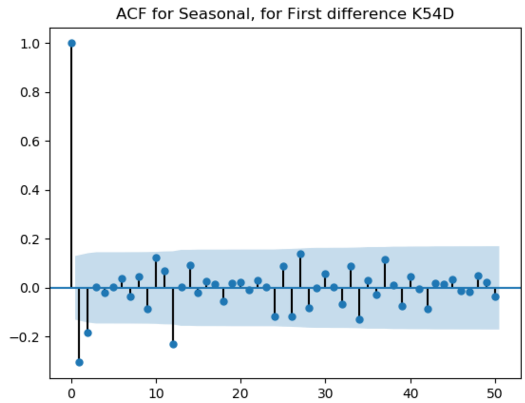


Figure 2.1.6 ACF Residuals, seasonal 1st difference

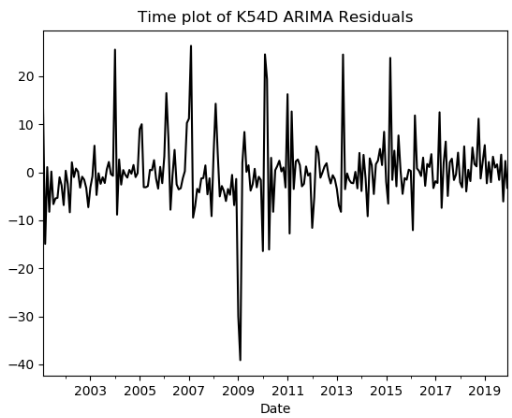


Figure 2.2.1. ARIMA Residuals time series

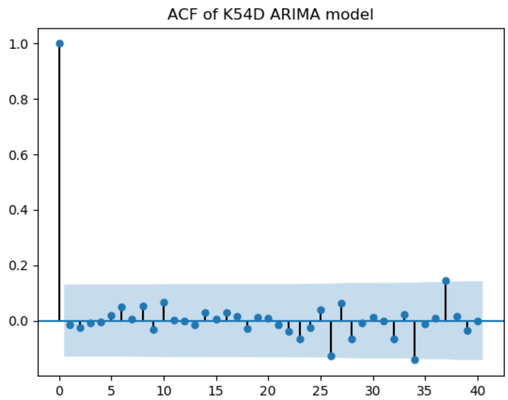


Figure 2.2.2. ARIMA ACF of Residuals

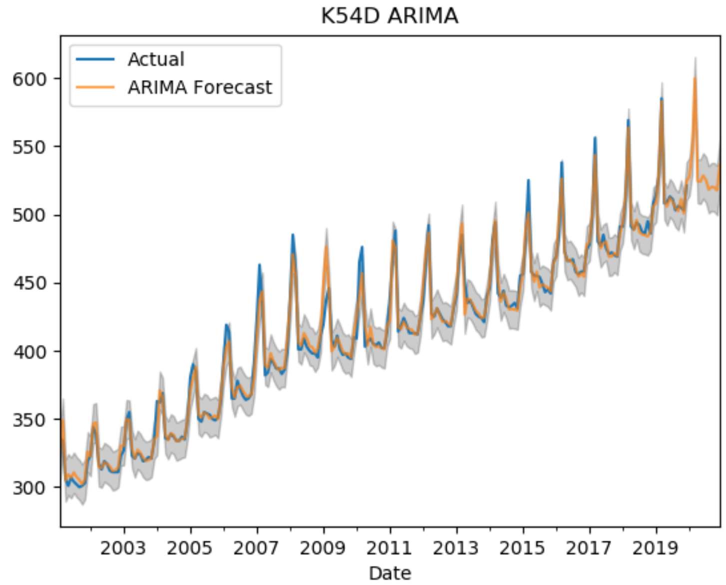


Figure 2.2.3. Forecasting K54D by ARIMA with upper and lower bounds

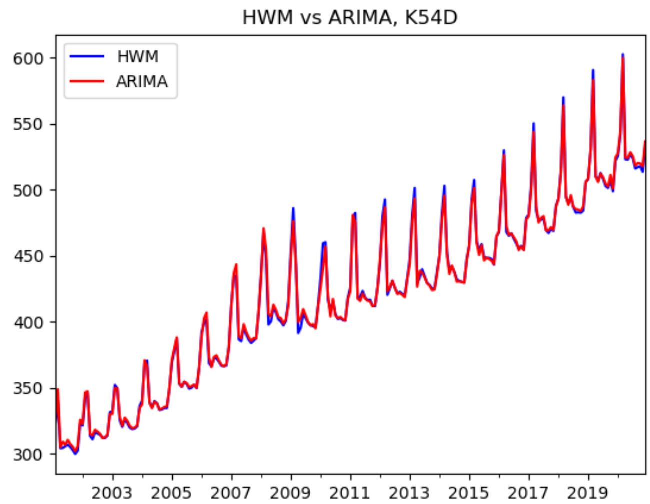


Figure 2.3.1. K54D: ARIMA vs HWM

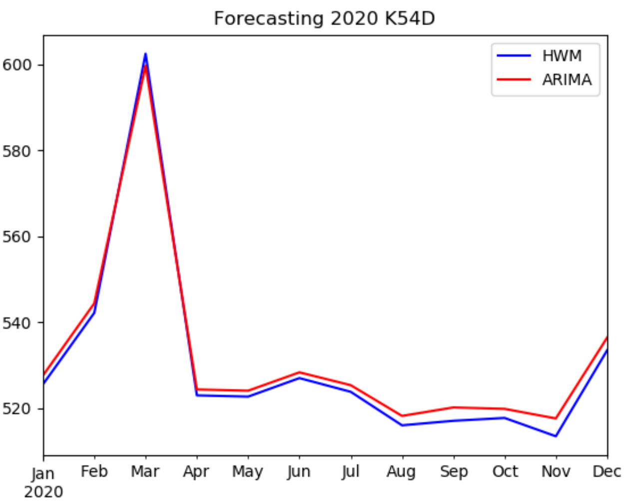


Figure 2.3.2. K54D: ARIMA vs HWM 2020

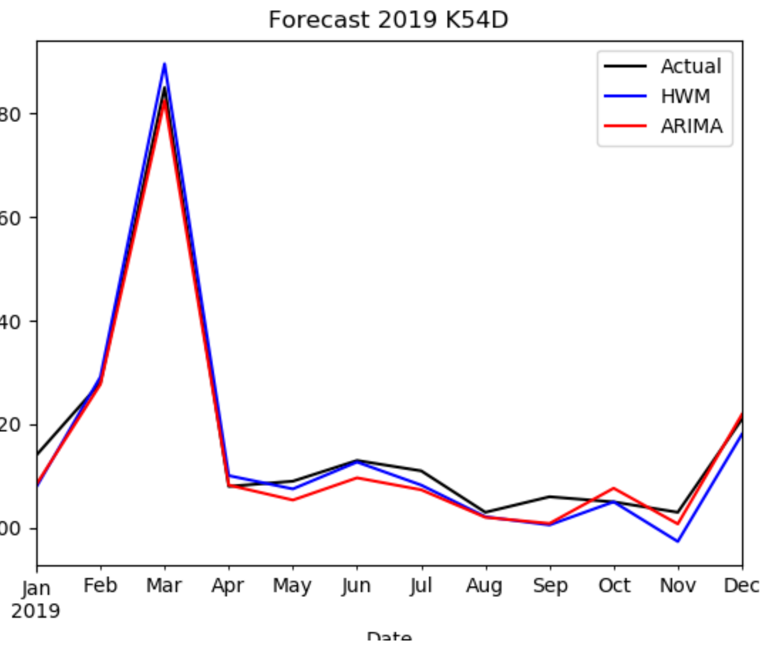


Figure 2.3.3. K54D: ARIMA vs HWM 2019, Split data

3. Regression Methods

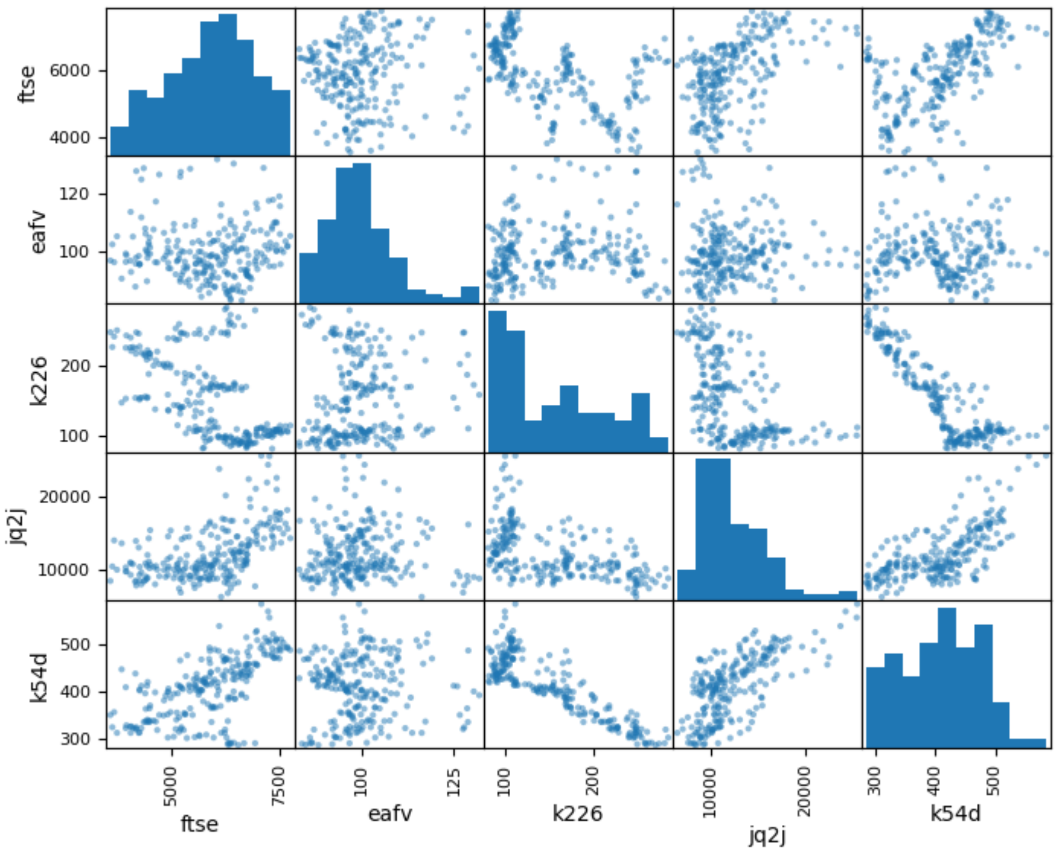


Figure 3.1.1. Correlations and distributions between given variables

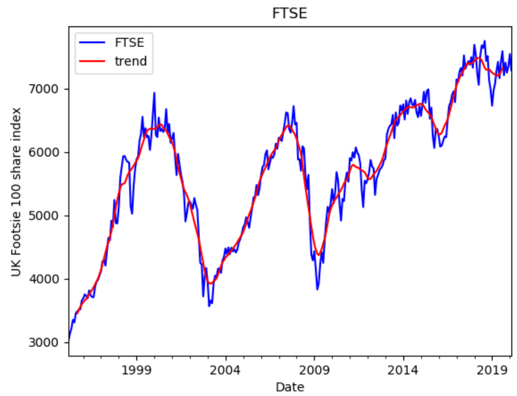


Figure 3.1.2. FTSE original with trend

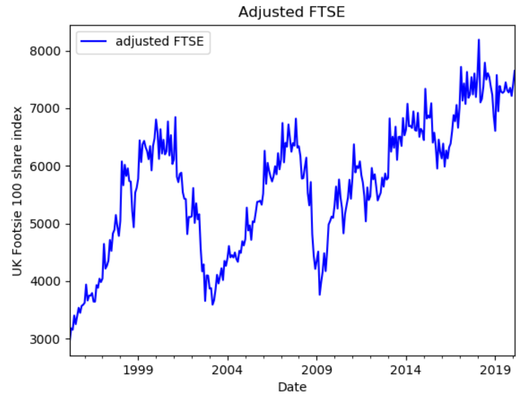


Figure 3.1.3 FTSE Calendar Adjusted data

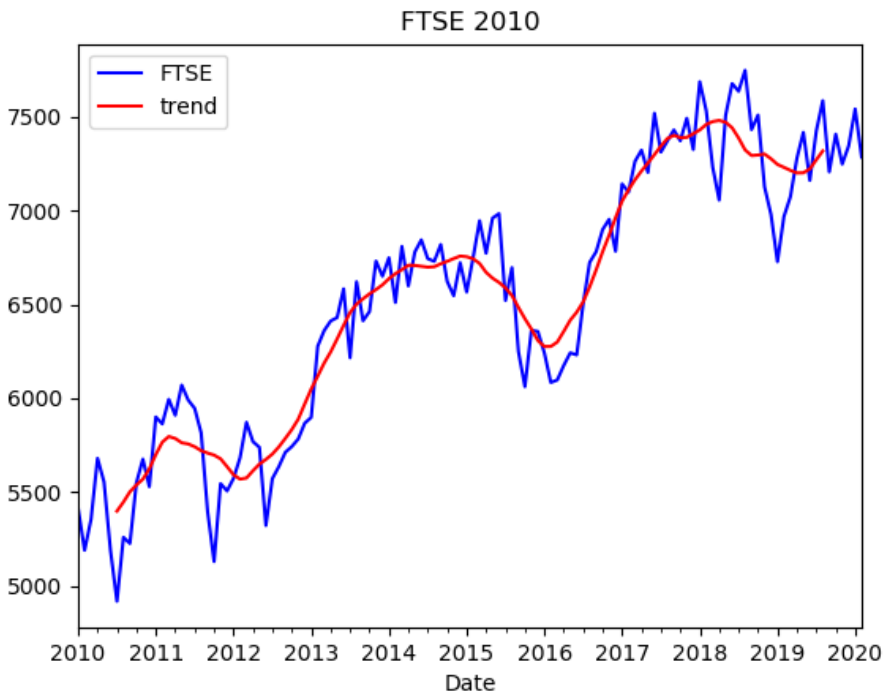


Figure 3.1.4 FTSE Data from 2010

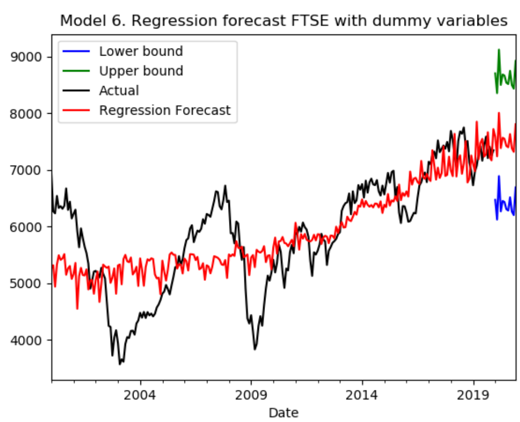


Figure 3.2.1. Model 6 (eafv, k226, jq2j, months)

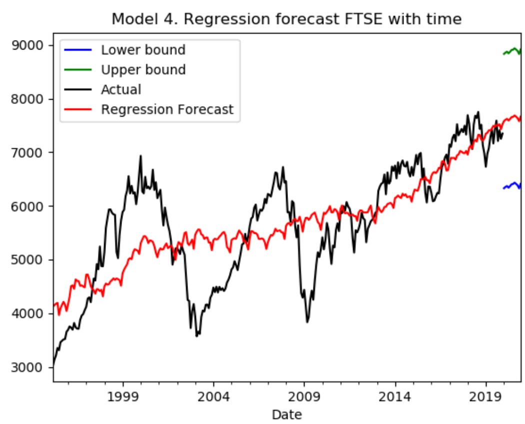
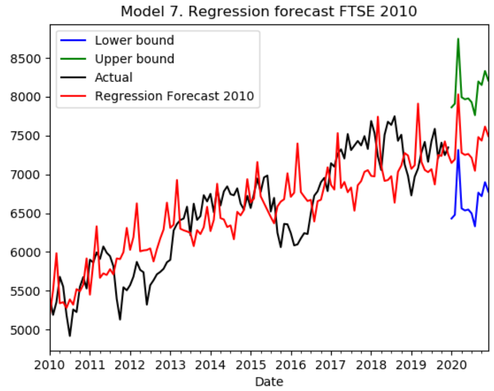


Figure 3.2.2. Model 4 Regression (eafv, k226, time)



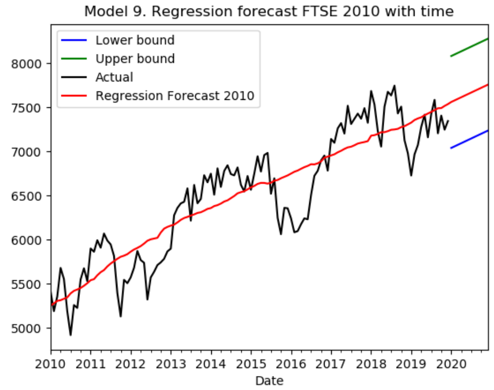
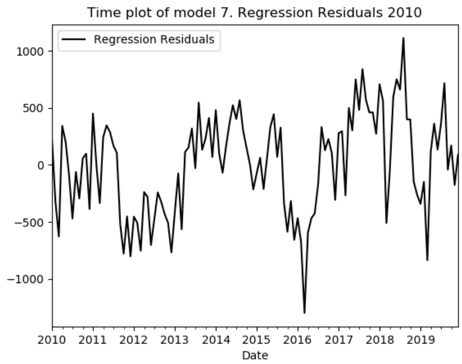


Figure 3.2.3. Model 7 (eafv, k226, jq2j, k54d)   Figure 3.2.4. Model 7 (k226, time)



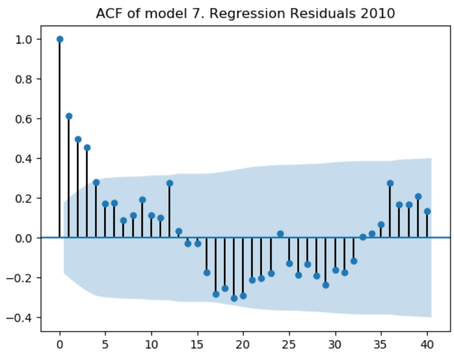
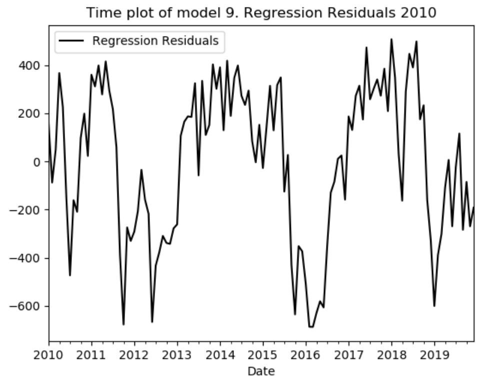


Figure 3.3.3. Model 7 Regression (Residuals)   Figure 3.3.4. Model 7 Regression (ACF of residuals)



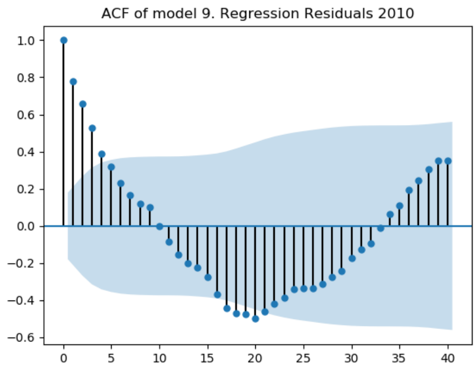


Figure 3.3.3. Model 9 Regression (Residuals)   Figure 3.3.4. Model 9 Regression (ACF of residuals)

Linear Model	Formula	AIC	R-square
1	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j**	3883	0.417
2	ftse = b0 + b1***k54d** + b4***jq2j**	3880	0.415
3	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j** + b5***time**	3847	0.502
4	ftse = b0 + b2***eafv** + b3***k226** + b5***time**	3844	0.501
5	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j** + b5***time** + b6***month1** + … + b16***month11**	3852	0.538
6	ftse = b0 + b2***eafv** + b3***k226** + b4***jq2j** + b5***time** + b6***month1** + b7***month2** … + b16***month11**	3851	0. 536
7 (from 2010)	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j**	1805	0.641
8 (from 2010)	ftse = b0 + b1***k54d** + b2***eafv** + b3***k226** + b4***jq2j** + b5***time**	1734	0.805
9 (from 2010)	ftse = b0 + b3***k226** + b5***time**	1730	0.801

Table 1. Linear regression models and summary