**Normalization**

The scale of the variables vary significantly, thus normalization was used after extreme value and null value analysis to rescale the whole data.

**Seasonality**

Two methods were used to detect seasonality effects.

Decomposition

Additive decomposition procedure is used to display seasonality in this paper. The scale of the seasonal effect is quite small (range between -0.005 to 0.005), indicating the seasonal variates can be ignored. Figure

Scaled periodogram

Frequency can be estimated by plotting the scaled periodogram(Figure). The plot shows that the peak is obtained at 0, which verifies that there is no seasonality in the last-price series.

**Differencing**

Since one requirement for building an ARIMA model is that the series should be stationary, the next step is to determine the integration order. After first order differencing the series is stationary. Figure shows the ACF and PACF of the differenced series. Both the sample ACF and PACF seem to cut off after lag 2. The results of the augmented Dickey–Fuller test (ADF tests) for the original series and differenced one are shown in Table. The null hypothesis of the ADF test is that the series is non-stationary.

|  |  |  |
| --- | --- | --- |
| Series | p value | Conclusion |
| Original series | 0.6922 | Non-stationary |
| Differenced series | 0.01 | Stationary |

**ARIMA model**

According to the above ACF and PACF plots, parameters p and q can be 1 or 2. In order to obtain the best model as we can. We fit models when p or q are equal to 0, 1 ,2 ,3. Based on Akaike information criterion (AIC), ARIMA (0,1,2), which minimizes the AIC value, is the best model for the normalized last-price series. AIC values of models are shown in Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ARIMA (0,1,1) | ARIMA (0,1,2) | ARIMA (0,1,3) | ARIMA (1,1,0) | ARIMA (1,1,1) |
| -364121.3 | -364140.3\* | -364138.4 | -364122.4 | -364132.6 |
| ARIMA (1,1,2) | ARIMA (1,1,3) | ARIMA (2,1,0) | ARIMA (2,1,1) | ARIMA (2,1,2) |
| -364138.2 | -364136.4 | -364139.9 | -364138.0 | -364136.1 |
| ARIMA (2,1,3) | ARIMA (3,1,0) | ARIMA (3,1,1) | ARIMA (3,1,2) | ARIMA (3,1,3) |
| -364134.3 | -364138.3 | -364136.2 | -364134.3 | -364132.3 |

**Residuals**

After the ARIMA model fitting, residuals are analyzed to detect violation of assumptions. The residuals of a well-fitted model should be uncorrelated and normally distributed.

Correlation

Autocorrelation between residuals can be tested by the ACF and PACF plot (Figure) and Ljung-Box test. The residuals seem to be uncorrelated and the Ljung-Box test is passed.

Normality

Q-Q plot (Figure) and Jarque-Bera test is used to test the normality. Non-normal behavior is found in the Q-Q plot of residuals. Although most of the points are on the red line, tails seem to deviate from the normal distribution. Jarque-Bera test, whose null hypothesis is that the series follows normal distribution, is failed also.

Arch effect

The null hypothesis of arch test is that the series is homoscedastic. Since the results of arch test indicated strong evidence against the alternative hypothesis, being heteroscedastic is not contributed to their non-normality. The results of Ljung-Box test, Jarque-Bera test and arch test is shown in (Table).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Chi-squared | df | p value | Conclusion |
| Ljung-Box test | 38.99 | 40 | 0.5156 | Uncorrelated |
| Jarque-Bera test | 2281800000 | 2 | <2.2e-16 | Not normal |
| Arch test | 1.6067 | 12 | 0.9998 | Homoscedastic |

Thus, ARIMA (0,1,2) is not a good model as non-normality is detected in the residuals, which violates our assumption. We then conducted residual analysis on other models we fit in order to find one that passes the correlation and normality test. But the results all indicate the residuals are not under normal distribution. ARIMA is not a model structure as it fails to capture possible information in the series. Other new method is needed to fit the model and do the forecast.