MTHM504 - Summer 2023 Thesis

Methods of Stock Price Forecasting

Nga Man Luk

Abstract

Stock is one of the important elements in the financial market. It is very essential to have some stock market analysis and price forecasting since it helps the investor to retrieve information about when to buy or sell the stocks. This helps the investor to earn the maximum profit though the stock market. In additional, there are huge investment risks in the stock market because they can be affected by lots of uncertainties, interrelated economic and political factors in both locally and globally. Therefore, most of the investors own a few of stock to build their own investment portfolio. This paper is going to take TESLA, SHELL and HSBA as examples for demonstrating how to use two of the time series model, ARIMA and DLM to forecast their stock price.

Introduction

Researcher with finance domain field always interested in studying how to improve the existing forecasting model. The main reason is the investors or company would like to make best investment decision, have abilities to plan and establish effective strategies on their business. Since the nature of stock market, stock price prediction is always identified as the hardest task in financial forecasting. The investors look for a model that can guarantee the highest profit with minimizing the lowest risk from the market.

Stock market prediction involves attempting to forecast the future value of a company's stock or other financial instruments traded on an exchange. If the future price of a stock can be predicted accurately, there are potential for substantial profits. It can be an indicator to see the performance of a company. The effects of those newly revealed information can also be reflected by the stock price immediately. When the investor want to analysis the performance of a stock, both qualitative and quantitative aspects of the business will be taken into account.

Nowadays, there are a few models and techniques that has been investigated for the prediction of stock price, for example artificial neural networks (ANNS), text mining, deep learning, ARIMA model and DLM model. Different model got their own characteristics in forecasting. In this paper, we chose TESLA, SHELL and HSBA to do the prediction since all of them are from different industry, it means the model structure or intervention time may be different.

At first, We are going to discuss about a few methods for stock price prediction. Secondly, since Auto Regressive Integrated Moving Average (ARIMA) model and Dynamic Linear Model (DLM) will be used, we will talk about how we apply these two methods in forecasting the stock price of TESLA, SHELL and HSBA. Third, we will describe and explain the results of the model. Next, we will have some further discussion on this topic. At last, the paper will be concluded.

Literature Review

Different analysis for stock price prediction

In general, there are three board categories for prediction of stock price, which are fundamental analysis, technical analysis and machine learning.

The main objective of fundamental analysis is to find out the true value of the certain stock in the market based on the company past performance, e.g. the price-earning ratio. It performs the analysis with taking the macroeconomic environment, the corresponding industry and the company itself into account. Several financial ratios like the interest rates or exchange rates will be found and considered (Nti, Adekoya and Weyori, 2020). Researchers believe that the government financial policies and macroeconomics events have dynamic relationship with the stock price performances. For instance, the Arbitrage Pricing Theory (APT) that raised by Ross (1976), Chen et al. (1986), they advocated that the industrial production, changes in risk premiums, and changes in the term structure related to the stock return expectation in a positive way and the inflation rates related to the stock returns expectation negatively (Adam and Tweneboah, 2008). Therefore, fundamental analysis are more focus on different rates or indicators from the society.

In the technical analysis, it is more focus on modelling the probabilities of the stock price in the future movement. By using the past data (Time-series analysis), the patterns of the price can be predicted in the short-run. With increasing the availability of historical data, time series forecasting can be used in predicting the future values by ignoring the limitation of traditional forecasting, for example, complexity and time-consuming (Khan and Alghulaiakh, 2020). The price information, trading volume and different indicators like simple moving average and exponential moving average will set as the model input (Li and Bastos, 2020). ARIMA and DLM model are the most common method in the technical (statistical) analysis which can perform forecasting in a dynamic way.

Apart from the fundamental and technical analysis, machine learning is broadly used for the prediction in financial market. By using the artificial intelligence systems and the patterns of historical data, the prediction can be produced in the process of training and model validation can be made. Then, the model can be optimized with testing in order to improve

the performances. Artificial neural networks (ANNs) is one of the common machine learning tools. There are an input layer with a set of sensory nodes as the input nodes, a few of hidden layers for the computation nodes and the layer for showing the result after computation (Tsai, Wang, 2009). The layers are used to process the units of neutral network interconnection and adjusting the weights between each connection. After going through a few layers, the errors can be minimized and the output of nodes can be produced.

ARIMA Model in technical analysis

Auto Regressive Integrated Moving Average (ARIMA) model is one of the time-series forecasting method that have been used by lots of researcher. It is a linear model that can deal with stochastic series and had been broadly applied in different application. Stock price prediction is one of the big application. ARIMA is using the past value to predict the future value in time series format, especially in short-term prediction. It is combined from two parts, Autoregressive (AR) model and Moving average (MA) model which form Autoregressive Moving Average (ARMA) model with order (p,q). ARMA model is a special case of ARIMA model.

Under stationary series, the following are the general equations of AR(p), MA(q) and ARMA(p,q):

$$AR(p): x_t = \sum_{i=1}^{p} \alpha_i \, x_{t-i} + \varepsilon_i, \quad \varepsilon_t \sim N(0, \sigma_w^2)$$

$$MA(q): x_t = \sum_{i=1}^{q} \beta_i \, \varepsilon_{i-1}, \quad \varepsilon_t \sim N(0, \sigma_w^2)$$

$$X_t = \sum_{i=1}^{p} \alpha_i \, x_{t-i} + \sum_{i=0}^{q} \beta_i \, \varepsilon_{t-i}, \varepsilon_t \sim N(0, \sigma_w^2)$$

which X_t is the actual value, α is the regression coefficient, ε_t is the random noise term at time t, p and q are the parameters that indicate the autoregressive and moving average and σ_w^2 is the constant variance (Khan and Alghulaiakh, 2020).

However, only stationary series can be forecast by using the above model. Since the stock market can be affected by lots of different factors which cause large fluctuation over time, the stock price cannot be induced as a stationary series. Therefore, ARIMA model need to be used in order to include the non-stationary series of stock price. By applying differencing, we can transform the non-stationary series into a stationary one and this model was introduced by Box and Jenkins in 1970.

By using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), we can find the p and q of the model, which represent different features of the functions and help us to identify the type of model we should fit and the orders. The parameter p can be found when we saw the ACF tends to 0 as lag tends to infinity with PACF equal to 0 for lags > p. For the parameter q, the ACF equals to 0 for lags > q when PACF tends to 0 as lag tends to infinity. It can be concluded in the following table:

Model	ACF	PACF
AR(p)	->0 as lag -> ∞	Equal to 0 for lags > p
MA(q)	Equal to 0 for lags > q	-> 0 as lag -> ∞
ARMA(p,q)	->0 as lag -> ∞	->0 as lag -> ∞

In general, the general steps in building the ARIMA model are model identification, parameter estimation and diagnostic checking.

DLM model in technical analysis

Dynamic Linear Model (DLM) is another model that can capture the changes in time. It is a model that also fit for non-stationary series, time-varying parameters, multivariate time series, irregular temporal observations etc. DLM is not only modelling the environmental area, such as forecasting the precipitation or the strength of different current, but also broadly used in economics and financial area (Rivera, 2016).

The principles of DLM model is using the existing information as condition to forecast the future development in a time series approach. Here are the four variables in the general Normal DLm model:

$$\{F,G,V,W\}_t = \{F_t, G_t, V_t, W_t\}$$

where, F_t is a known (nxr) matrix, G_t a known (nxn) matrix, V_t a known (rxr) matrix and W_t a known (nxn) matrix.

The distribution can be written as follow:

$$(Y_t|\theta_t) \sim N[F'_t\theta_t, V_t]$$

$$(\theta_t|\theta_{t-1}) \sim N[G_t\theta_{t-1}, W_t]$$

From the above distribution, here are the general DLM equations:

$$ObservationEq: Y_t = F'_t\theta_t + v_t, \quad v_t \sim N[0, V_t]$$

$$SystemEq \quad \theta_t = G_t\theta_{t-1} + w_t, \quad w \sim N[0, W_t]$$

$$Initial \quad prior: (\theta_0|D_0) \sim N[m_t, C_0]$$

where the error v_t and w_t are independent and mutually independent and also independent with $(\mu_0|D_0)$ (Mike West, Jeff Harrison, 1989).

For the third component, it is the probabilities of the forecaster's beliefs at time t = 0 with given information that at the time, D_0. The mean m_0 is the estimation of the level and the variance C_0 about the uncertainty of the mean. With the initial information, the variances V_t and W_t for all t and the observation $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_1$, we can induce that the $D_t = \{Y_t, D_{t-1}\}$.

Methodology

Data

Since we need to perform some forecasting on the stock price of TESLA, SHELL and HSBA, time series model has been used in this paper. With working on the time series model, we need some past stock price data to build the model. Therefore, the past 5 years stock price data of TESLA, SHELL and HSBA had been taken from the website, Yahoo! Finance. For the dataset we downloaded, the start date is 2nd Jan 2018 and the end date is 30th Dec 2022. Given that we would like to forecast the stock price, we took the daily closing price of these three stocks. We also based on the daily closing price to calculate the average stock price each month in order to predict the future trend.

By modelling TESLA, SHELL and HSBA in this research, numerous time series forecasting method have been tried, including Auto Regressive Integrated Moving Average (ARIMA) and Dynamic Linear Model (DLM).

TESLA

With using the TESLA data to build the model, we tried to use ARIMA to fit the data that we had at the beginning. However, from the initial plot of the closing price of TESLA, the price between 2018 and 2020 are more stable since there are many factors that affect the stock price of TESLA after 2020, for example some speech from Elon Musk or some policies that announced by the US government. Therefore, we used the ARIMA to model the stock price from 2018 to 2020 initially.

By using the ARIMA model, we have two approach, one is using the auto ARIMA function, another one is using the customized parameters p,d and q to build the best fit forecasting model. For the manual ARIMA model, we have to find out the differecing of the model, which is parameter d, by stabilized the dataset. After the parameter d is found, we can use the dataset that has been differenced to plot the ACF and PACF. By using the theorem that has shown above, the parameter p and q can be found and the forecasting model can be built by

using the ARIMA function. After that, we can use this model to forecast the next 12 months of TESLA stock price.

However, with looking at the plot of TESLA stock price and market nature, we know that the stock price of TESLA is so fluctuated and the stock market can be affected by many different factors. These factors cause various intervention on the stock price, which changed the model structure. The ARIMA model may not be a good fit for building the forecasting model, therefore, we also tired to use the DLM model with adding intervention to fit the data.

By looking at the plot of TESLA stock price, second order polynomial model has been used at the initial stage. With the initial polynomial model, the parameter dV and dW can be found, which are the variances of observation and state vectors. After that, we can update the existing model with the estimation of dV and dW. Using the model with updated parameters, we can apply Kalman filter to compute the filtered values of state vectors, with their variances or covariance matrices and also one step ahead of the observations.

At the same time, since the second-order polynomial is being used with the level and growth parameters μ_t and β_t at time t, therefore $\theta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$. And we also got the standard observation and system matrix from the initial model.

$$F_t = [1 \ 0], \quad GG_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

After getting the standard observation, system matrix, a set of state vectors, variances and the observations, we would like to add the intervention effects in the model. Before adding the intervention event in the model, we have to decide which time t required the changes. Referring to the time-series plot of the stock price of TESLA, those time point t that having an abnormal rise or drop will be taken into account. After that, we can start to implement the intervention in the model.

In general, the intervention information can be written as follow:

$$I_t = \{h_t, H_t\}$$

Where h_t is the mean vector and H_t is the covariance matrix of the random intervention, ε_t following with $\varepsilon_t \sim N(h_t, H_t)$. Assuming the ε_t is not correlated with $(\theta_{t-1}|D_{t-1})$, w_t and $(\theta_t|D_{t-1})$, the noise term ε_t will be added to the θ_t after the basic prior distribution.

Here is the updated equation with the addition of ε_t

$$\theta_t = G_t m_{t-1} + w_t + \varepsilon_t$$

And

$$(\theta_t|I_t, D_{t-1}) \sim N[a_t^*, R_t^*]$$

With

$$a_t^* = a_t + h_t \quad and \quad R_t^* = R_t + H_t$$

For a_t^*, it is an arbitrary shifts in the mean vector of prior, which indicate there are a level and growth changes at a certain point t. Apart from the level and growth changes at a certain point, intervention can also be happened with increasing the variance through H_t. The intervention H_t will not occur any changes in the state θ_t but allows more flexibility to some of the components in the model from intervention.

From the initial model, we know the a_t at different point t and we can estimate the additional intervention level h_t by comparing the differences between the true value of the observation and the one-step-forecast value from the model. After that, we can modify our model by adding the h_t and get the new a_t^*. For the variance of intervention H_t, it is not easy to assign suitable value in general. Thus, we are using the trial and error method to identify the H_t for hedging with enough uncertainty in order to ensure the model are more adaptive to the future data.

After adding all the intervention events at different specific time point t, the filter model has been updated. We will have a more accurate one-step-forecast model for the stock price of TESLA. In the next step, we would like to calculate the standardized mean square error to look at the model performance after the adjustment. If the mean squared error is smaller, it means that the model has better fit on the data.

Here is the equation of the Mean Square Error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\frac{(True \quad Observation_t - Prediction \quad Observation_t)^2}{Prediction \quad of \quad \sigma_t}$$

SHELL

For modelling the stock price of SHELL, we used a regression model in DLM representation. Modelling with regression is concerning the relationship between the independent variables and the response in time series Y_t. It means that the current X_t will affect the current mean μ_t. In this study, we are going to use the price of Crude Oil to forecast the stock price of SHELL. We assume they have linear relationship between each other and can be explained in the following mathematical equation:

$$Y_{Stockprice of SHELL} = \alpha_t + \beta_t X_{Price of CrudeOil}$$

where α and β are the coefficients of the regression model and changing slowly by time with a simple random walk evolution as shown below.

$$\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + w_t$$

and w_t is the zero mean vector error term(Mike West, Jeff Harrison, 1989). On the other hand, , the system equation of the regression model in DLM has slight differences with the general DLM system equation. Here is the system equation for the Dynamic Regression Model

$$SystemEq \quad \theta_t = \theta_{t-1} + w_t, \quad w \sim N[0, W_t]$$

Using the dataset of the stock price of SHELL and the price of Crude Oil, we form the following matrix of the model. Let X_t be the price of crude oil at time t and n be the total number of the monthly Crude Oil price.

$$F_t = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \dots \\ 1 & X_n \end{bmatrix} \qquad\qquad \theta_t = \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix}$$

By using the F_t, θ_t, m_0 and C_0 with setting the dV and dW can only be positive value, we created the Dynamic regression model. Then, we can enhance the current model by incorporating the estimated values of dV and dW. With these updated model parameters, we can utilize the Kalman filter to calculate the filtered state vector values along with their associated variances or covariance matrices. From the derived model, we possess a collection of state vectors, variances, and observations.

After that, we would like to forecast the coming 12 months of the SHELL price based on the price of the Crude Oil. However, the forecast function is not working with the Dynamic regression model, therefore we have to use the k-step ahead distribution for finding the next 12 months of SHELL price. Here are the State and Forecast distribution for θ_{t+k} and Y_{t+k} with given D_t

$$StateDistribution: (\theta_{t+k}|D_t) \sim N[a_t(k), R_t(k)]$$

where

$$a_t(k) = a_t(k-1) \quad and \quad R_t(k) = R_t(k-1) + V_{t+k}$$

by given that $a_0 = m_t \quad and \quad R_t = C_t$

$$ForecastDistribution: (Y_{t+k}|D_t) \sim N[f_t(k), Q_t(k)]$$

where

$$f_t(k) = F'_t a_t(k) \quad and \quad Q_t(k) = F'_t R_t F_t + V_{t+k}$$

In addition, since the stock price of SHELL at time t is based on the price of the Crude Oil at time t, it is not possible for us to get the future price of Crude Oil and using model to forecast the stock price. Therefore, we are going to work on the scenario analysis in this case.

Scenario 1: The price of Crude Oil increased by 10% at each time t Scenario 2: The price of Crude Oil decreased by 10% at each time t

By using the condition of scenario 1 and 2, we can get the price of Crude Oil in the next 12 months, which helps us to develop the new matrix of F_t. Then, we can forecast the stock price of SHELL in the next 12 months under both scenarios.

For the validation of the SHELL model, we will look at the p-value for Ljung-Box statistic to determine whether the residuals of the time-series model has patterns or not. If the low p-value have been shown in the plot, it means that there is a pattern of the residuals in the time-series model.

HSBA

By referring the method that we used in modelling the stock price of TESLA, we used the ARIMA model to forecast the stock price of HSBA initially. Same as TESLA, there are two approach of using the ARIMA model, which are the auto ARIMA function and manually specifying the values for the parameters p, d, and q to construct a tailored forecasting model that best suits the data characteristics. As mentioned before, we have to find out the differencing by stabilizing the data first. After that, we can get the parameter p and q from the plot of ACF and PACF, which help us to fit in the manual model with the Arima function. Using the model, the stock price of HSBA in the next 12months can be forecasted.

However, we can see that the stock price of HSBA are quite fluctuated throughout the 5years. It is because there are many external factors that can affect the stock market, especially in Banking and Financial Industry. Therefore, taking the experiences with forecasting the stock price of TESLA, we are going to perform the DLM model with adding intervention events in order to build a better forecasting model for the stock price of HSBA.

At first, we started the model with second order polynomial function as the initial model of the stock price of HSBA. By using this model, we can find out the estimated parameter dV and dW for updating the model. Utilizing the model incorporating the revised parameters, we can employ the Kalman filter to calculate the filtered state vector values, along with their associated variances or covariance matrices, as well as the one-step-ahead predictions of the observations. The observation and system matrix also can be induced from the model

Using set of state vectors, variances and the observations in the model, we can try to add the intervention effects in the model. Again, before integrating the intervention event into the model, it is essential to pinpoint the precise time instances t that necessitate modifications.

To ascertain these moments, we analyze the time-series plot of HSBA stock prices, paying special attention to the points on the graph that exhibit significant increases or decreases.

After that, with the state vector at time t that we get from the model and the true value of the observation, we can find out the intervention level and growth h_t that we would like to adjust in the model. Apart from the level and growth adjustment, it is possible for us to adjust the flexibility of the updated mean vector by changing the H_t. It is because the addition of intervention may lead to more uncertainty to the model and the change forecast.

Upon incorporating intervention events at various specific time points t, the filter model has undergone updates. The refined model results in a more precise one-step-forecast model for HSBA's stock price. After that, we would like to calculate the standardized mean squared error, in order to evaluate the model's performance post-adjustment.

Results and Discussion

TESLA

**Closing Price of Tesla in Time Series**
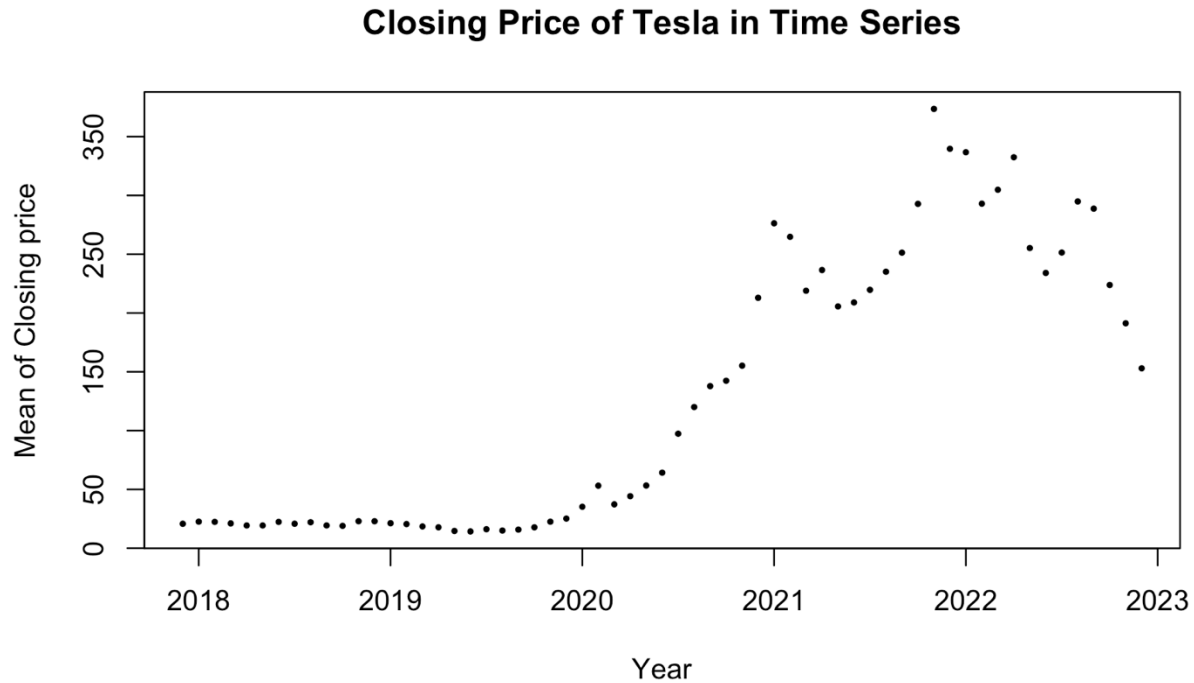
Fig. 1 Closing Price of TESLA in time-series

According to the time-series plot of TESLA throughout 5years (Fig. 1), the stock price of TESLA is more stable during 2018 and 2020, thus we used the ARIMA to build the model initially.

```
## Series: ts_month_tesla_2018to2020
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -0.6800
## s.e.   0.1278
##
## sigma^2 = 116.8:  log likelihood = -132.77
## AIC=269.54   AICc=269.92   BIC=272.65
```

By applying the auto ARIMA function into the time-series dataset of TESLA from 2018 to 2020, it showed that ARIMA(0,2,1) has the best fit and the AIC and BIC value are 262.91 and 265.97.
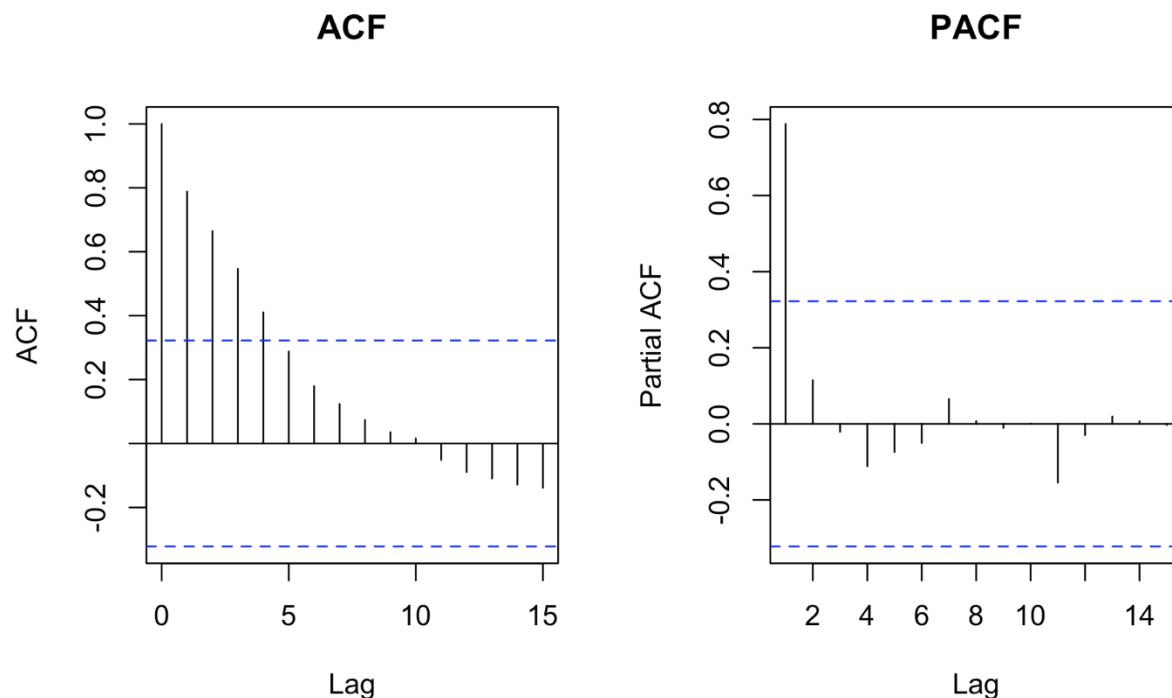
Fig. 2 ACF and PACF of TESLA in ARIMA model

```
## Series: (ts_month_tesla_2018to2020)
## ARIMA(2,2,10)
##
## Coefficients:
##            ar1      ar2      ma1      ma2     ma3      ma4      ma5      ma6
##        -0.0183  -0.2021  -0.6527  -0.0969   0.054  -0.1039   0.5916   0.0267
## s.e.    0.6816   0.3731   0.8228   0.6046   1.068   0.6655   0.3187   0.6616
##            ma7      ma8      ma9     ma10
##         0.0676  -0.0396  -0.6672   0.9111
## s.e.    0.8721   0.5483   0.7782   0.5558
##
## sigma^2 = 60.32:  log likelihood = -121.2
## AIC=268.4    AICc=285.73    BIC=288.61
```

For the customized one, we have to find out the differencing of the model first. After processing 2 differencing on the dataset, the curve seems to be stable, therefore, the parameter d is being set as 2. Secondly, the parameter p and q have to be found for the plot

of ACF and PACF plot. From the following plot (Fig. 2), ACF equals to 0 when the q is greater than 10 and PACF equals to 0 when p is greater than 2. Thirdly, we are going to apply ARIMA(2,2,10) for the manual ARIMA model and it shows that the AIC and BIC are 556.4 and 583.41, which is must higher than the auto arima one. Therefore, we can conclude that the auto arima with (0,2,1) has the best performance with using the ARIMA model approach.

```
##            Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
## Feb 2021         240.4344  226.5856  254.2833  219.2545  261.6144
## Mar 2021         267.9293  244.9953  290.8633  232.8548  303.0038
## Apr 2021         295.4242  263.1470  327.7013  246.0605  344.7878
## May 2021         322.9191  280.7455  365.0926  258.4201  387.4180
## Jun 2021         350.4139  297.7296  403.0982  269.8403  430.9876
## Jul 2021         377.9088  314.0953  441.7223  280.3144  475.5032
## Aug 2021         405.4037  329.8558  480.9516  289.8631  520.9443
## Sep 2021         432.8986  345.0301  520.7671  298.5153  567.2819
## Oct 2021         460.3934  359.6382  561.1487  306.3016  614.4853
## Nov 2021         487.8883  373.7000  602.0767  313.2523  662.5244
## Dec 2021         515.3832  387.2338  643.5326  319.3957  711.3707
## Jan 2022         542.8781  400.2572  685.4990  324.7582  760.9979
```

Table. 1 Table for forecasting of TESLA in ARIMA model

**Closing Price Prediction of Tesla in Time Series**
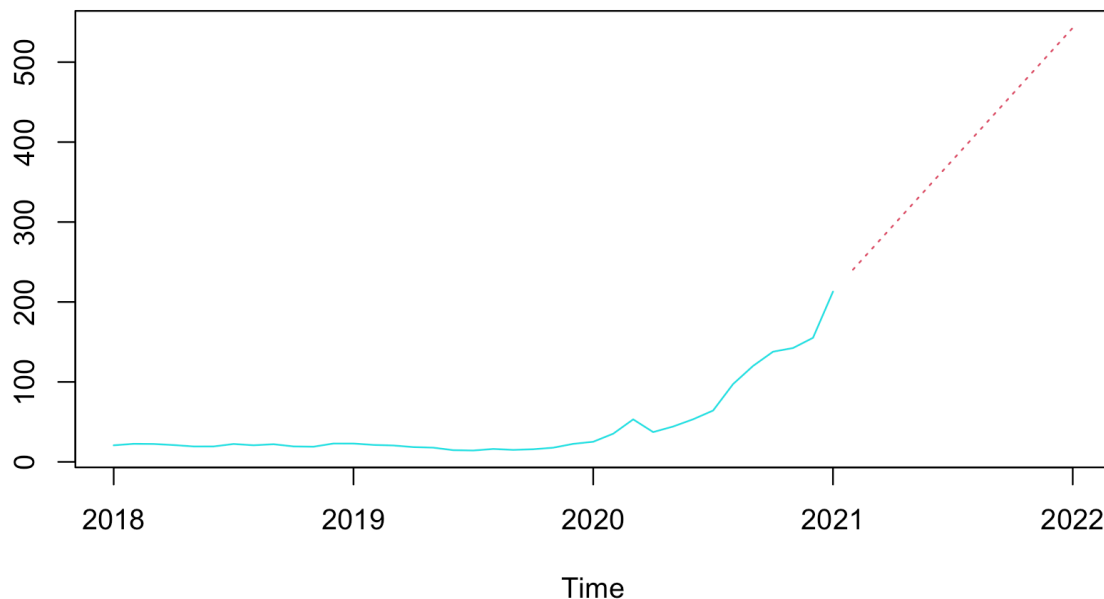
Fig. 3 12-months prediction for Closing price of TESLA in time-series

In Fig. 3, the plot used the auto ARIMA function and trying forecast the stock price of TESLA with 12months ahead. By comparing the real trend of the TESLA's performance, it is not a good model to predict the stock price since the model cannot include the intervention events that affecting the market.

Looking into the time-series plot of TESLA, we considered that the intervention events happened in December 2020 (t = 37), January 2021 (t = 38) and November 2021 (t = 48). The following figure showed the one-step-forecasts for the stock price of TESLA.

Forecast Closing Price of Tesla in Time Series
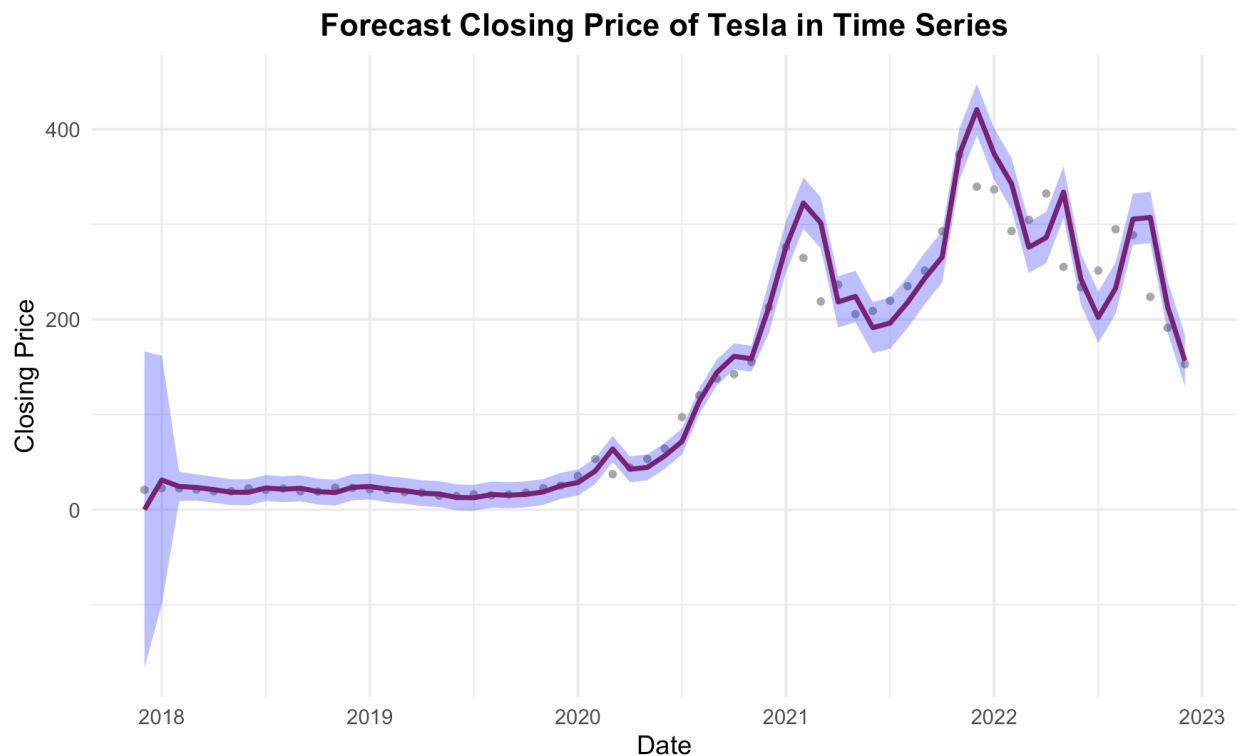
**Fig. 4 One-step ahead forecasts for TESLA with interventions in time-series**

(1) At t=37 (December 2020), there is a sharp rise in the stock price of TESLA. Since TESLA announced it delivered 499,500 vehicles in 2020, which over the market expectation. During that period, the sales of many automakers' went down because of the economic conditions were challenging (Noonan, 2021) and TESLA proved that they can maintain its remarkable performance, which also served as evidence of huge demand in the market. By taking the report announcement of TESLA into account of

the intervention event, we included the change in level and growth $h_{37} = \begin{bmatrix} 46 \\ 100 \end{bmatrix}$. On the other hand, since the intervention increased the uncertainty of the price, therefore, we added 100 to the observation variance in order to have higher variability on the stock price forecasting. The state vector and the observation variances have been updated to $a_{37}^* = \begin{bmatrix} 212.5145 \\ 110.5670 \end{bmatrix}$ and 134.2959.

(2) At t=38 (January 2021), the stock price of TESLA keeps increasing in a fast pace. The reason is probably based on the sales performances of TESLA in the last five straight quarters of profit which helps assisted the electric-car manufacturer in standing out within the global auto industry (Reuters, 2021). With the preferences of investor and the TESLA sales performances, another intervention is added in the model. There is a change in level but no change in growth since the growth looks like the same according to the time-series plot of TESLA's stock price. $h_{38} = \begin{bmatrix} 39 \\ 0 \end{bmatrix}$. was added in the model. Based on the previous time t (t=37), we increased the variability of the forecast based on the change in level so the observation variance at t=38 was added 100. In consequences, we got the state vector and the observation variances which are $a_{38}^* = \begin{bmatrix} 276.63040 \\ 34.45165 \end{bmatrix}$ and 134.2959.

(3) At t=48 (November 2021), TESLA's stock price was hit a record high among the period from 2018 to 2022. The cause of this peak happened is mainly because TESLA launched the non-TESLA Supercharger pilot in November 2021. The market shows there is strong buying in TESLA share along with this new product launch. Under this situation, we added the third intervention in the model. We set the change in level is 57 and the change in growth is 500, which is $h_{48} = \begin{bmatrix} 57 \\ 500 \end{bmatrix}$ in order to get a better forecasting model for TESLA. Same as the previous two intervention event, the observation variances got extra 100 to ensure the variability. At the end, the state vector is $a_{48}^* = \begin{bmatrix} 374.0458 \\ 529.9663 \end{bmatrix}$ and the observation variances is 134.2959.

```
##            Feb      Mar      Apr      May      Jun      Jul
## 2021 223.7174 249.2332 274.7490 300.2648 325.7806 351.2964
```

Table. 2 Table for Closing Price of TESLA after 1st interventions in time-series

Forecast Closing Price of Tesla after 1st intervention in Time Series
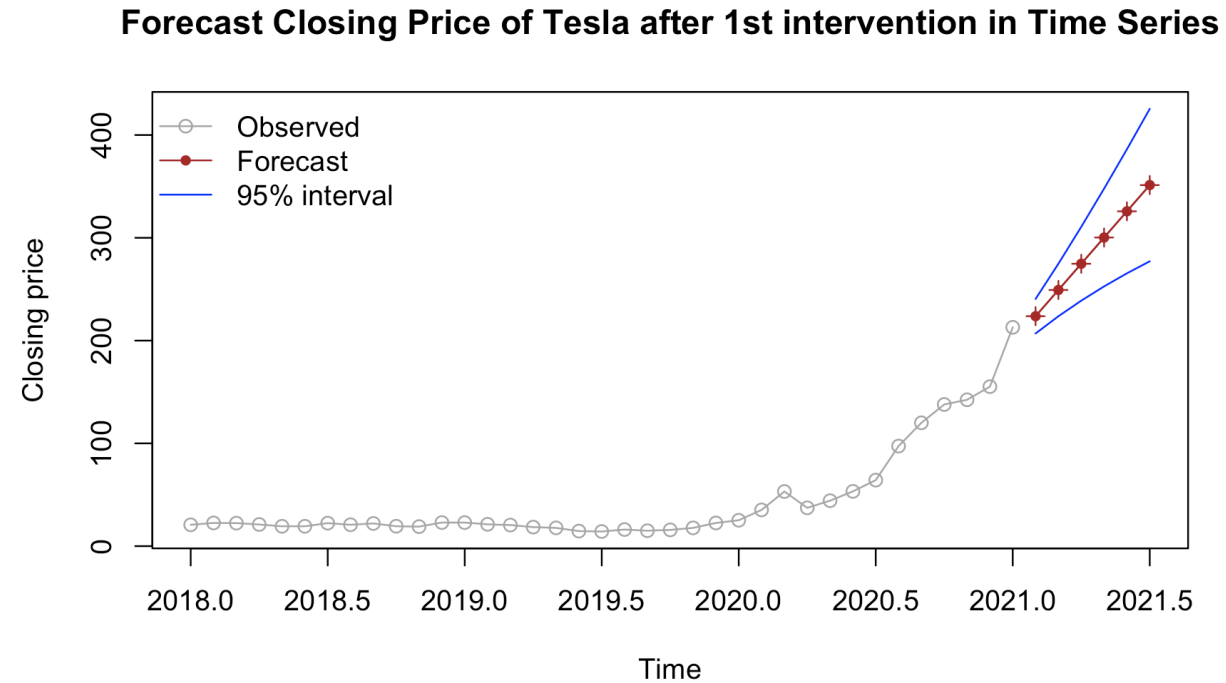
Fig. 5 Forecast Closing Price of TESLA after 1st interventions in time-series

From Table 2 and fig. 5, we can see that how the first intervention event affects the prediction of stock price. As mentioned before, the first intervention happened in December 2020 (t=37). There is a sharp rise in the stock price of TESLA with 46 increased in level and 100 increased in growth. In result, after December 2020 the prediction of stock price in the coming 6 months keeps increasing. This means that the intervention affected the prediction model with an uplifting trend.

Residual Plot of TESLA model
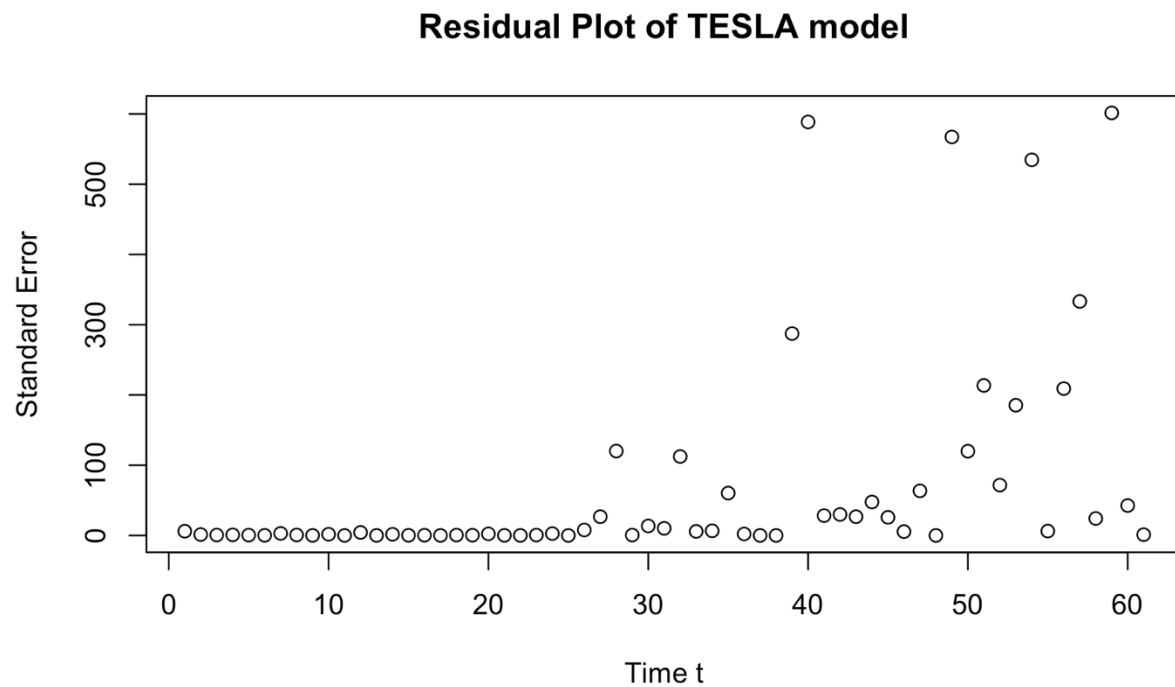
Fig. 6 Residual plot of TESLA model

```
## Mean Square Error of TESLA model:  72.25187
```

After the model has been updated, the mean squared error of the model is 72.25187. From fig. 6, we can see that most of the points from t = 0 to t = 25 are lying around zero. After t = 25, some of the points are around zero or dispersed with different value. We can also conclude that there are no patterns for the residual points among all the time point t.
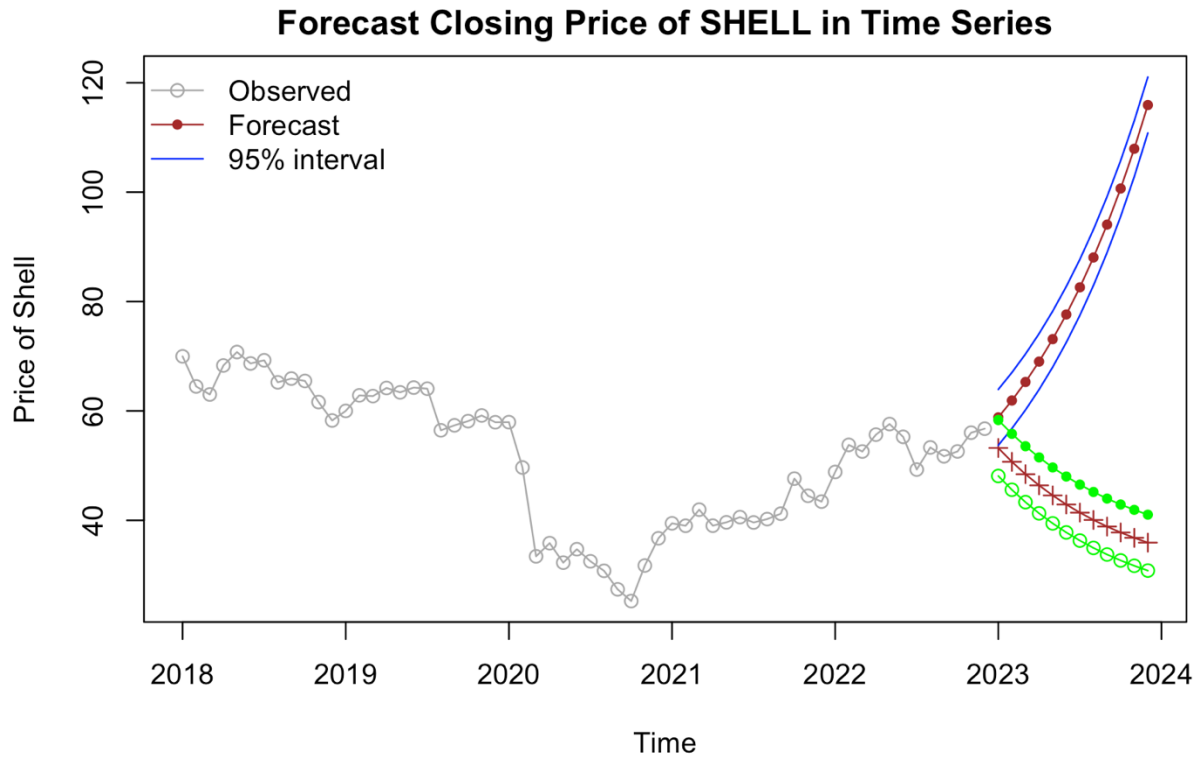
SHELL
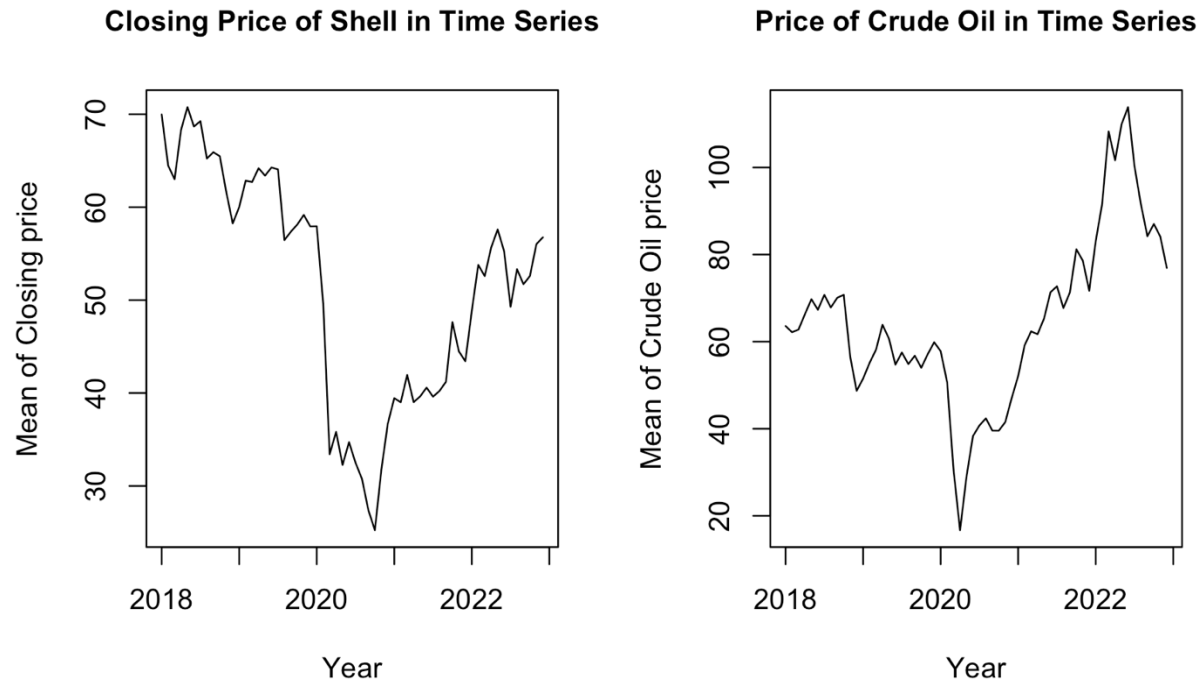
Fig. 7 Closing Price of SHELL and Price of Crude Oil in time-series

From the time-series plot of the stock price of SHELL and the price of Crude Oil (Fig. 7), we can see that they seems have linear relationship between each other. Therefore, we used the Dynamic Regression model and try to forecast the stock price of SHELL by using the price of Crude Oil. We have set two scenarios, (1) The price of Crude Oil increased by 10% at each time t and (2) The price of Crude Oil decreased by 10% at each time t. The following is the forecasting plot for both scenarios.

```
##             Jan       Feb       Mar       Apr       May       Jun       Jul
## 2023   58.82500  61.90631  65.29575  69.02414  73.12537  77.63671  82.59919
##             Aug       Sep       Oct       Nov       Dec
## 2023   88.05793  94.06253 100.66759 107.93316 115.92529

##             Jan       Feb       Mar       Apr       May       Jun       Jul       A
ug
## 2023 53.22262 50.70154 48.43258 46.39051 44.55265 42.89857 41.40990 40.070
10
##             Sep       Oct       Nov       Dec
## 2023 38.86428 37.77904 36.80233 35.92328
```

Table. 3 Table for forecasting of SHELL in DLM model

Fig. 8 Forecast Closing Price of SHELL in time-series

(1) In Fig.8, the curve is describing scenario 1. It shows that the stock price of the SHELL is increasing for each time point t. Since the price of Crude Oil has been set to be increased by 10% for each time point t, the raise of the SHELL's stock price is increased by each t.

(2) When examining the curve associated with scenario 2 (Fig.8), it becomes evident that the SHELL stock price experiences a downward trend gradually across each time point t. This trend aligns with the predetermined condition of a 10% decrease in the Crude Oil price at each time point, which subsequently leads to a corresponding fall in the SHELL's stock price at each instance of t.

Since we assumed that the stock price of SHELL and the price of Crude oil have linear relationship between each other, the changes in the price of Crude Oil will affect the stock price of SHELL in a direct way.
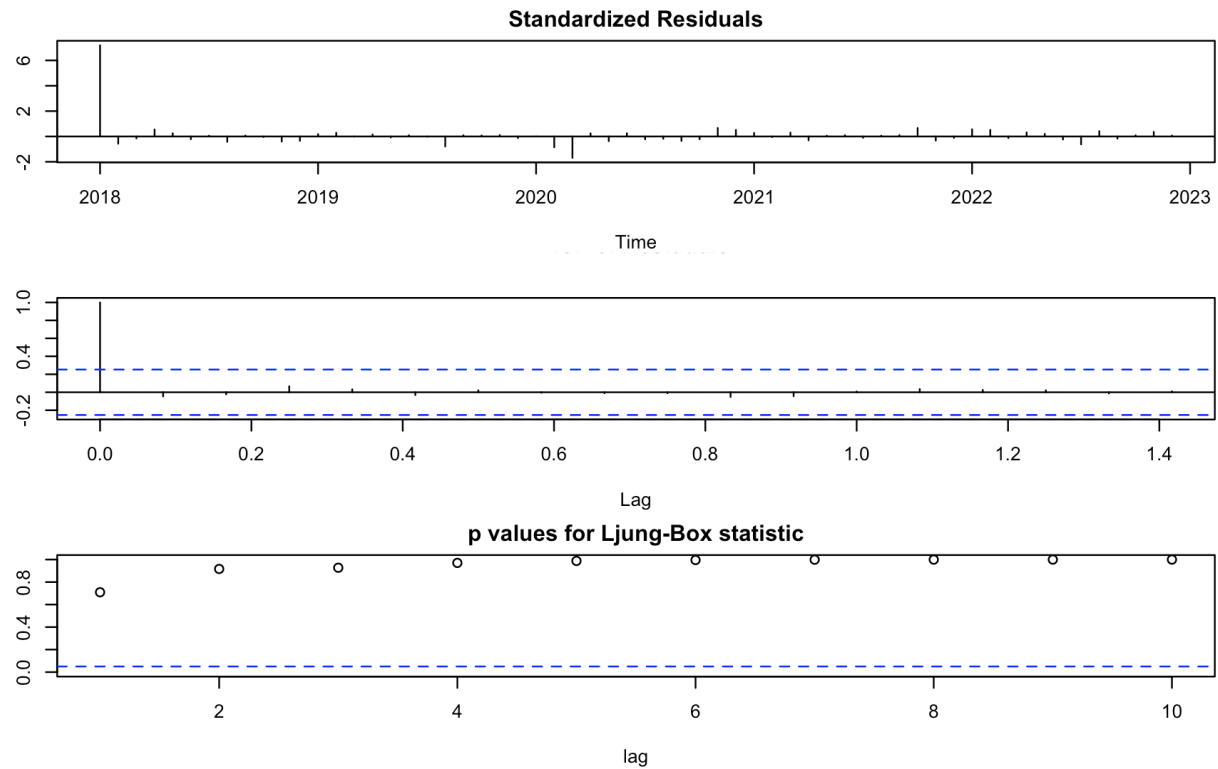
Fig. 9 Validation of DLM model of SHELL

After building the forecasting model of the stock price of SHELL by using the price of Crude Oil, we got the validation plot of the model (Fig. 9). From the p-value for Ljung-Box statistic, we can see that all the points in different lags are much larger than 0.05. It means that the residuals in the time-series model did not have any patterns.

HSBA

**Closing Price of HSBA in Time Series**
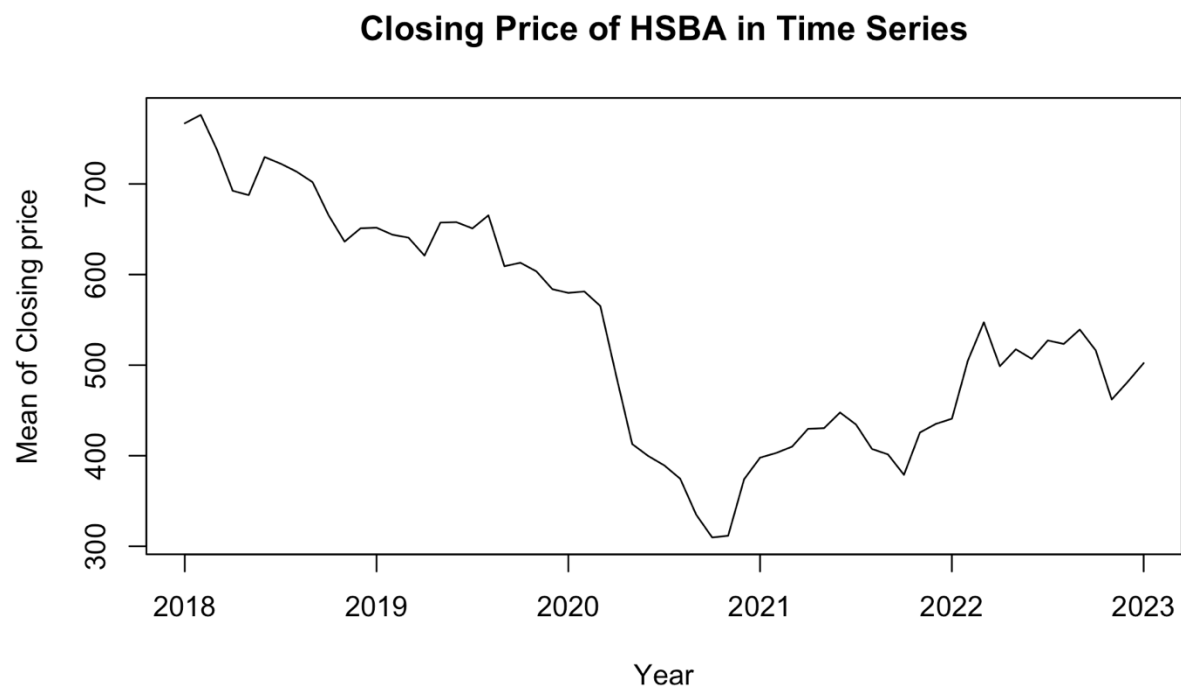
Fig. 10 Forecast Closing Price of HSBA in time-series

```
## Series: ts_month_hsba
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##       0.2766
## s.e.  0.1317
##
## sigma^2 = 807.3:  log likelihood = -285.48
## AIC=574.97    AICc=575.18    BIC=579.15
```

By applying the auto ARIMA function into the time-series dataset of HSBA from 2018 to 2022, it showed that ARIMA(0,1,1) has the best fit and the AIC and BIC value are 574.97 and 579.15.
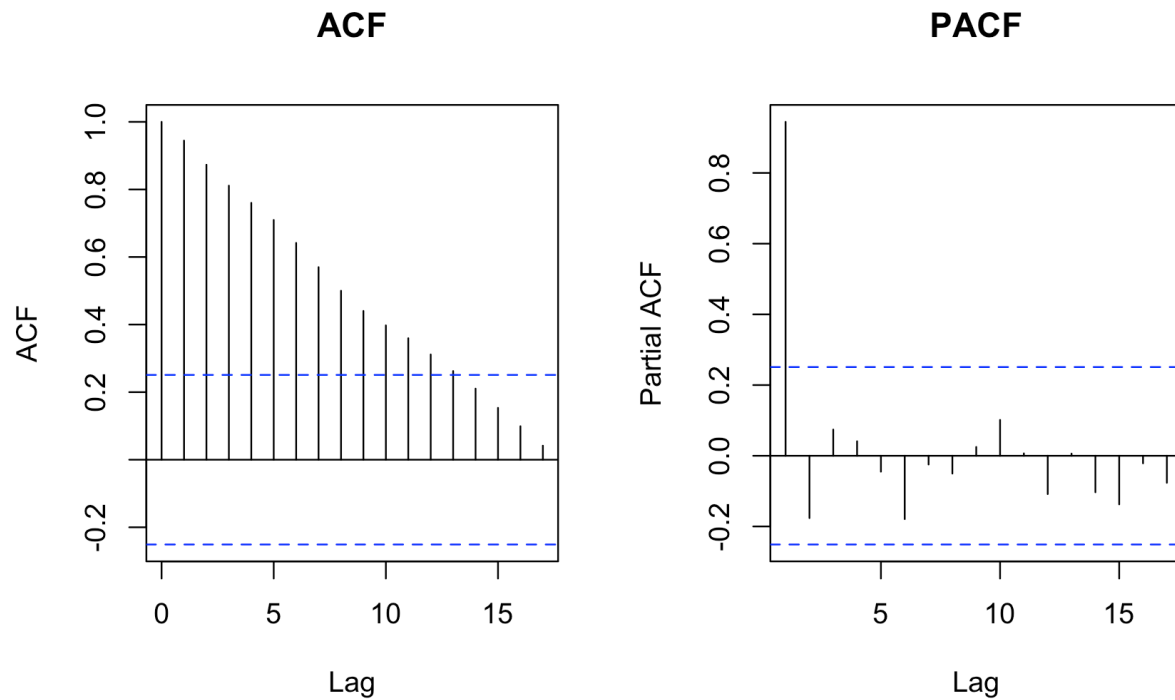
Fig. 10 ACF and PACF of HSBA in ARIMA model

```
## Series: (ts_month_hsba)
## ARIMA(1,2,18)
##
## Coefficients:
##           ar1      ma1      ma2      ma3      ma4      ma5      ma6       ma
7
##       -0.2014  -0.5272  -0.3147  -0.0036  -0.1015   0.2074  -0.1389  -0.023
0
## s.e.   0.9897   1.1305   1.1571   0.7601   0.7541   0.7239   0.6539   0.560
1
##           ma8      ma9     ma10     ma11     ma12     ma13     ma14      ma15
##        0.0246  -0.1619  -0.1478   0.1081  -0.0974   0.3390   0.3100  -0.0924
## s.e.   0.5798   0.5232   0.6630   0.5694   0.5078   0.4844   0.5049   0.4527
##          ma16     ma17     ma18
##       -0.6821   0.1651   0.1654
## s.e.   0.5203   0.5733   0.3538
##
## sigma^2 = 802.7:  log likelihood = -277.69
## AIC=595.38    AICc=617.49    BIC=636.93
```

For the customized one, we have to find out the differencing of the model first. After processing 2 differencing on the dataset, the curve seems to be stable, therefore, the

parameter d is being set as 2. Secondly, the parameter p and q have to be found for the plot of ACF and PACF plot. From Fig.10, ACF equals to 0 when the q is greater than 18 and PACF equals to 0 when p is greater than 1. Thirdly, we are going to apply ARIMA(1,2,18) for the manual ARIMA model and it shows that the AIC and BIC are 595.38 and 636.93, which is must higher than the auto arima one. Therefore, we can conclude that the auto arima with (0,1,1) has the best performance with using the ARIMA model approach.

```
##              Point Forecast     Lo 80      Hi 80     Lo 95      Hi 95
## Feb 2023            505.4733  469.0601   541.8865  449.7842   561.1625
## Mar 2023            505.4733  446.4253   564.5213  415.1672   595.7794
## Apr 2023            505.4733  430.3240   580.6226  390.5424   620.4042
## May 2023            505.4733  417.1095   593.8371  370.3326   640.6141
## Jun 2023            505.4733  405.6289   605.3177  352.7745   658.1722
## Jul 2023            505.4733  395.3386   615.6080  337.0368   673.9098
## Aug 2023            505.4733  385.9308   625.0158  322.6489   688.2977
## Sep 2023            505.4733  377.2113   633.7353  309.3135   701.6332
## Oct 2023            505.4733  369.0479   641.8987  296.8286   714.1180
## Nov 2023            505.4733  361.3461   649.6005  285.0498   725.8968
## Dec 2023            505.4733  354.0356   656.9111  273.8693   737.0773
## Jan 2024            505.4733  347.0620   663.8846  263.2042   747.7425
```

Table. 4 Table for forecasting of HSBA in ARIMA model

Closing Price Prediction of Hsba in Time Series
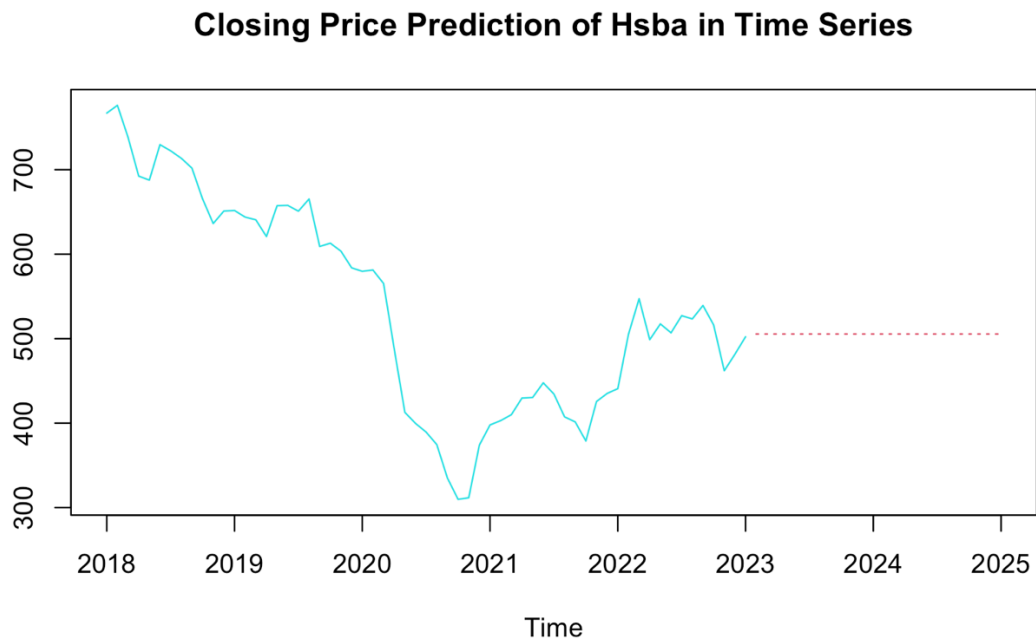
Fig. 11 Closing Price Prediction of HSBA in Time Series

From Table 4 and Fig.11, it is using the auto ARIMA function and trying to forecast the stock price of HSBA with 12months ahead. The model does not provide any useful informaion for forecasting, since the forecasted value are constant, which cannot predict how the price changes in the next 12 months. At the same time, by comparing the real trend of the HSBA's performance, it also shows that it is not a good model to predict the stock price since the model cannot include the intervention events that affecting the market.

Looking into the time-series plot of HSBA (Fig. 10), we considered that the intervention events happened in March 2020 (t = 28), April 2020 (t = 29), May 2021 (t = 30) , November 2021 (t = 36) and January 2022 (t = 50).
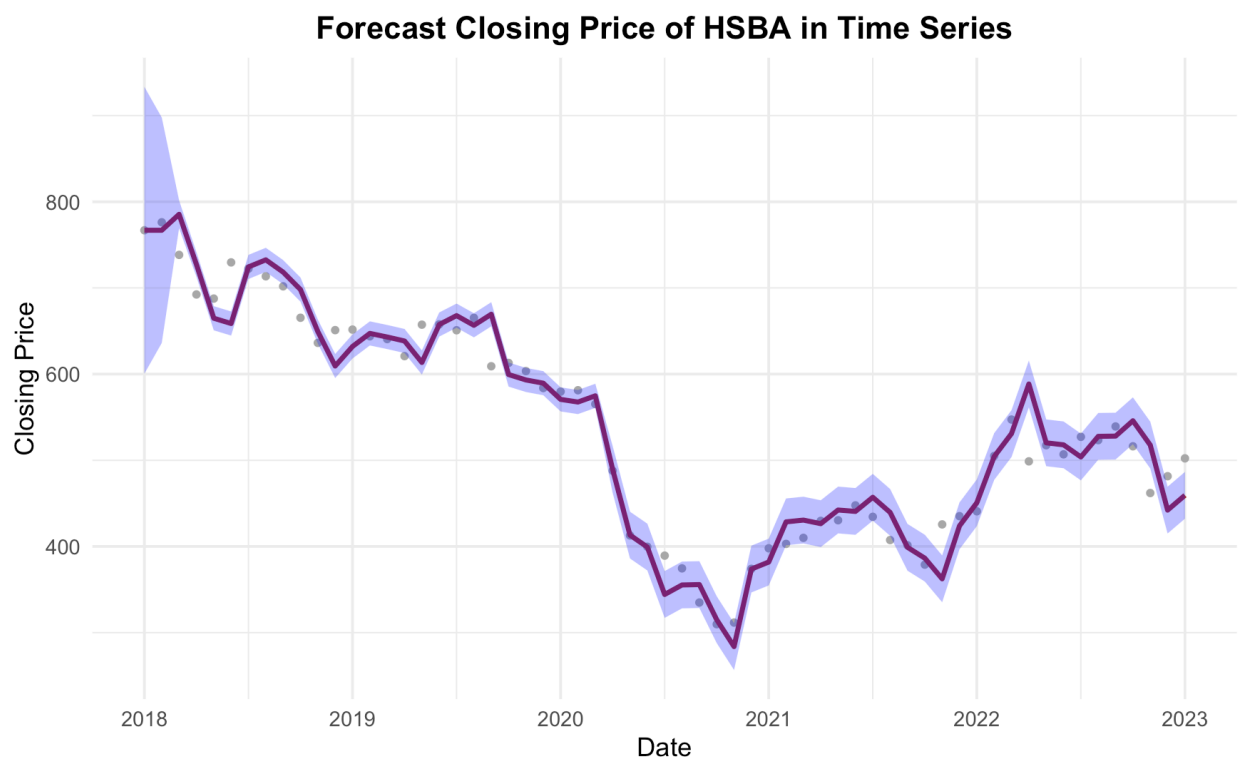
Fig. 12 One-step ahead forecasts for HSBA with interventions in time-series

(1) At t=28 (April 2020), the stock price of HSBA was hammered. Global financial institutions were suffering with financial setbacks as the impact of the COVID-19 pandemic persists. HSBC revealed a significant 50% decline in profits directly attributed to the outbreak (Frost, 2020). This news was announced in the first quarter of 2020, which is March, therefore it apparently affects the share price of HSBA. By

including this intervention event, we set 70 as the change in level and no change in growth, $h_{28} = \begin{bmatrix} -70 \\ 0 \end{bmatrix}$. Since the intervention event cause the uncertainty of the HSBA's stock price, the observation variance has been added extra 100 to enhance the variability of the forecasting. $a_{28}^* = \begin{bmatrix} 489.373914 \\ -8.026525 \end{bmatrix}$ and 136.1282 are the updated state vector and the observation variances.

(2) At t=29 (May 2020), the continuous effect of COVID-19 pandemic along with the first quarter report announcement, the stock price of HSBA continued their downward trend. The change in level is set to be 48 without the change in growth and we got $h_{29} = \begin{bmatrix} -48 \\ 0 \end{bmatrix}$. Considering the flexibility, the observation variance were added extra 100. Therefore, the updated state vector and the observation variances is $a_{29}^* = \begin{bmatrix} 413.18064 \\ -42.87211 \end{bmatrix}$ and 136.1282.

(3) At t=30 (June 2020), the stock price of HSBA stopped falling sharply as the previous two months, since the interim report of HSBC was announced. From the report, there are good performances in the second quarter of 2020. Around 55% adjusted revenue has grown in the global market and the reported pre-tax profit is 3.6 billion (Motley Fool Transcribers , 2020). This announcement leads to an intervention to the market. Although the stock price of HSBA was still falling, the intervention reduced the fall with 42 changes in level, which is $h_{30} = \begin{bmatrix} 42 \\ 0 \end{bmatrix}$. Thus, the updated state vector and observation variance were $a_{30}^* = \begin{bmatrix} 399.28492 \\ -66.45512 \end{bmatrix}$ and 136.1282.

(4) At t=36 (November 2020), the stock price of HSBA climbed a bit after it fell to its lowest level since 1995 and the third quarter results on cost savings are better than expectation. At the same time, the investor are optimistic with the financial stock because of the Covid-19 vaccine (BBC News, 2020). It leads to the change in level without any growth, which is $h_{36} = \begin{bmatrix} 82 \\ 0 \end{bmatrix}$. On the other hand, the observation variance

has been updated extra 100 to the model. $a_{36}^* = \begin{bmatrix} 373.76307 \\ -13.58666 \end{bmatrix}$ and 136.1282 are the updated state vector and the observation variances.

(5) At t=50 (January 2022), there was a sharp rise from the stock price data of HSBA. The main reason of this result probably is because the profit of HSBC in 2021 was skyrocketed with 14.7 in billion U.S. dollars, which was a double more than the profit in 2020. This leads to more investors were positive to the stock price of HSBA in 2022. With this intervention, we added extra 28 for the change in level, which is $h_{50} = \begin{bmatrix} 48 \\ 0 \end{bmatrix}$. Same as the prevous intervention events, observation variance was increased extra 100. Therefore, the updated state vector and the observation variances are $a_{50}^* = \begin{bmatrix} 504.30319 \\ 13.27305 \end{bmatrix}$ and 136.1282.

```
##            May       Jun       Jul       Aug       Sep       Oct
## 2020  500.5362  479.6008  458.6654  437.7300  416.7946  395.8591
```

Table. 5 Table for Closing Price of SHELL after 1st interventions in time-series
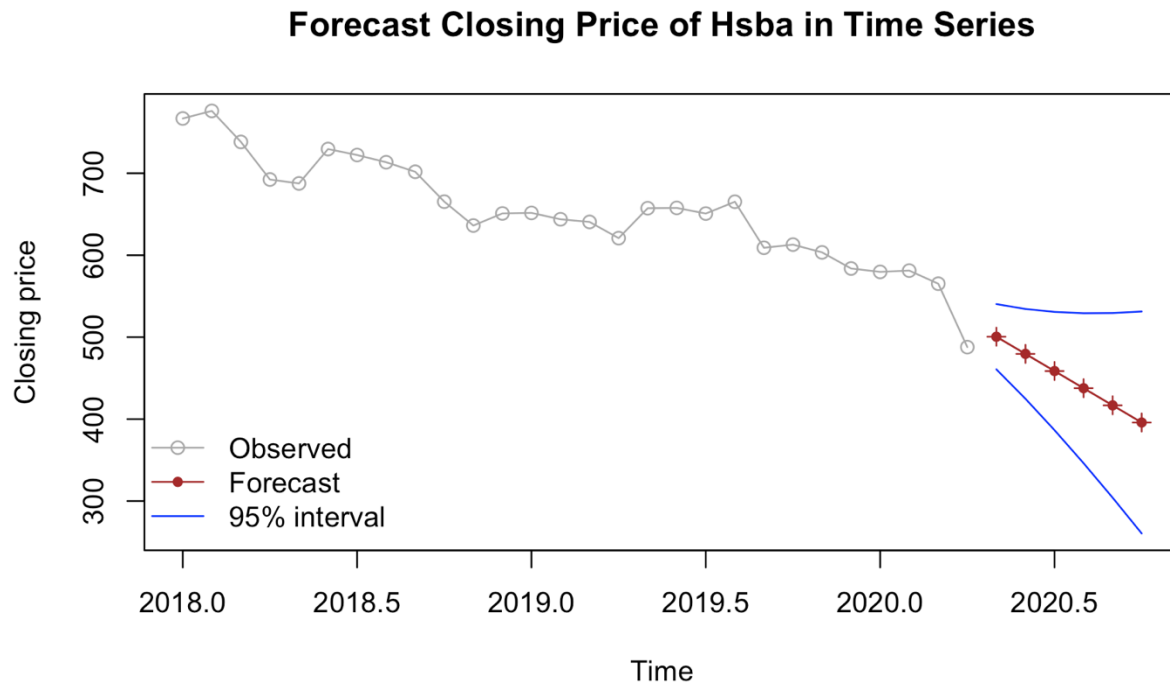
Fig. 13 Forecast Closing Price of TESLA after 1st interventions in time-series

From Table 5 and Fig. 13, we can see that how the first intervention event affects the prediction of stock price in HSBA. As mentioned before, the first intervention happened in April 2020 (t=28). The stock price HSBA fell like a stone with 70 decreased in level. In consequences, the coming 6 months after April 2020 got a falling trend in the stock price of HSBA. It indicated that the corresponding intervention affected the forecasting model with a declining trend.

Residual Plot of HSBA model
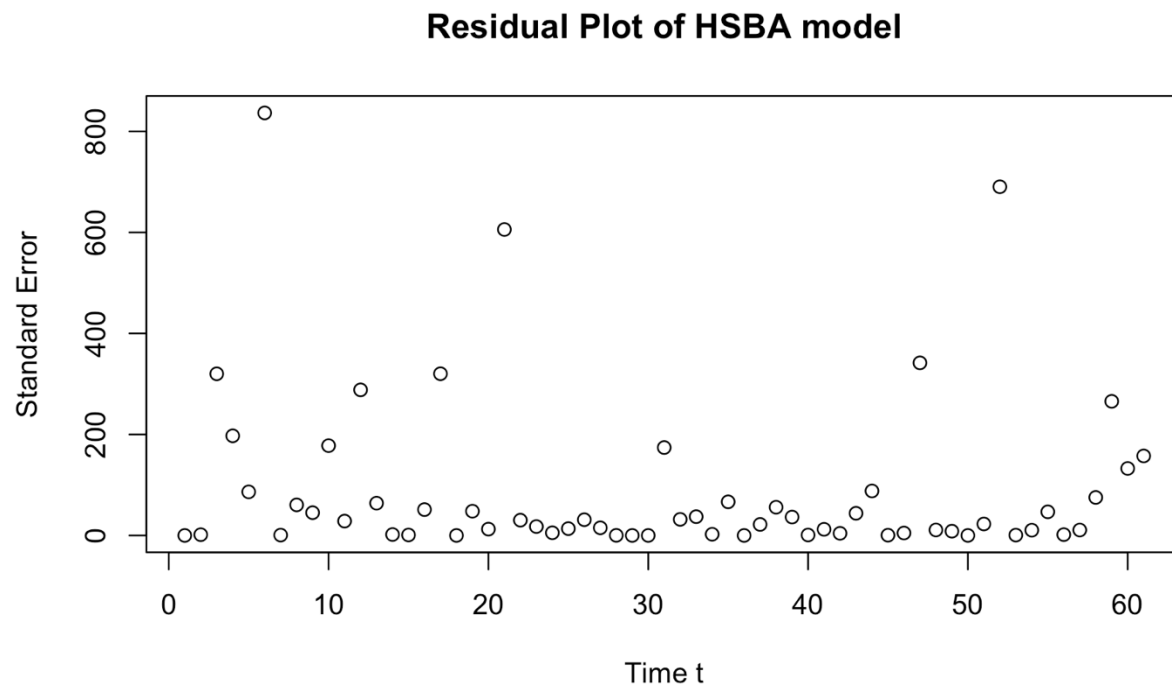
Fig. 14 Residual plot of HSBA model

```
## Mean Square Error of HSBA model:  92.06708
```

After the model has been updated, the mean squared error of the model is 92.06708 and the above plot showed that how the residuals in the dataset dispersed. Most of the points are located around 0 and other residual points do not have any pattern among all the time point.

Discussion

In this study, there are a few of limitation that has been found and some further suggestions in the future study.

Firstly, the financial market prediction is a very complicated topic. An accurate prediction of stock price involves both quantitative and qualitative analysis. Starting with the background of the company and the trend of the corresponding industry to the potentials, they all required lots of data science and statistics, with supporting sufficient financial area knowledge. However, I do not have related studies before. Therefore, the prediction of stock price can only be focus on using the time series modelling to do the prediction, which cannot be fully applied in the real world. In consequences, we suggest that the project can involve some people that own basic financial background, which help to analysis the market or company in a more comprehensive way.

Secondly, most of the investors will not put all the eggs in one basket, which mean they prefer to invest different stock at the same time. Therefore, they usually build up their own portfolio in order to get the highest profit in the market. For building up their own portfolio, it is essential to perform portfolio risk optimization after getting the results from the forecasting model. However, there is limitation of time, we do not have sufficient time to perform the risk optimization with the result. Thus, we suggest this project can perform a further study in portfolio risk optimization.

Thirdly, there are lots of techniques that can forecast the stock price. However, only Auto Regressive Integrated Moving Average model (ARIMA) and Dynamic Linear Model (DLM) have being choose to build the prediction model. Other models like text mining, different types of neural network or deep learning have not been trying to use in this study. Thus, we suggest this study can be continuous trying different methods to build the forecasting model in order to get the best fit one for prediction.

Conclusion

In conclusion, we used the ARIMA and DLM model to perform the stock price prediction in this study. At first, we applied ARIMA model to both TESLA and HSBA dataset. However, the ARIMA model cannot produce a good forecasting because there are some interventions in the stock market which affect the performance of stock price. Therefore, we changed to use DLM model with adding a few of intervention values inside in order to get the better model. After modifying the two models, we got the one step ahead forecasts. On the other hand, we performed the impact of the first intervention event in both models affect the movement of the price. For the stock price of SHELL, we used the DLM with regression model to perform the forecasting since we assumed there are a linear relationship between the price of Crude Oil and SHELL's stock price. There are two scenarios that predict how the stock price performs based on the movement of Crude Oil prices – one where the Crude Oil price is increasing and another where it is decreasing.

References