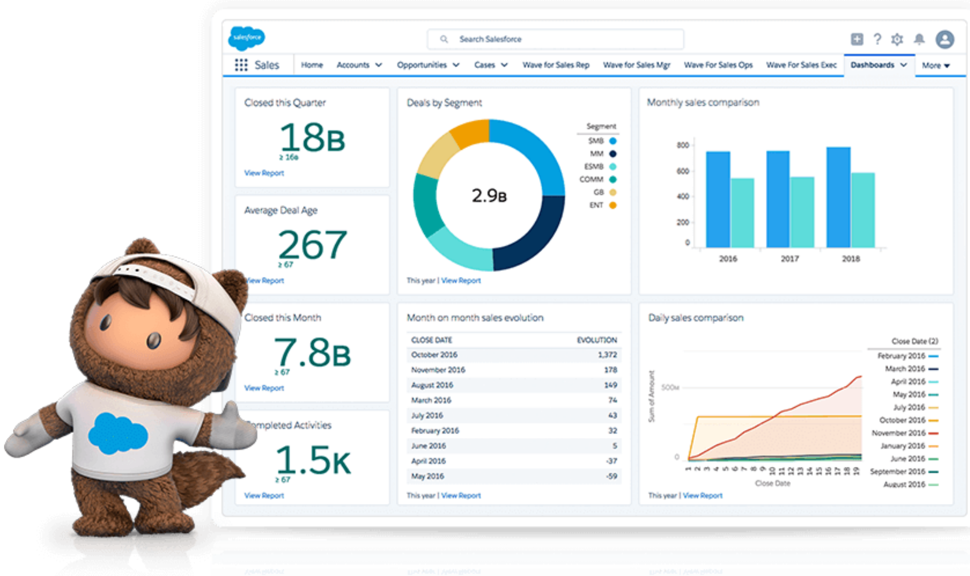
Forecasting Salesforce (CRM) Stock Daily Price using ARIMA and GARCH

Finnegan Nguyen

21 May 2020



Contents

1. Introduction

1.1. CRM and Question of Interest

Salesforce.com, Inc. (CRM) is a cloud-based software company headquartered in San Francisco, California. Salesforce provides customer relationship management (CRM) service and technology for managing the company's relationships and interactions with customers. More than 150,000 companies are using Salesforce software. For 2018, Salesforce reported earnings of $127 million.

The aim of this report is to predict the daily stock price of CRM in the last 20 days of 2019 base on the CRM daily stock price from January 2016 to December 2019 by choosing an adequate ARIMA model and GARCH model.

1.2. The Data

The data used in this report is the CRM stock daily closing prices which are collected every evening except weekends and holidays. The data is obtained from Yahoo Finance through the package quantmod in R.

Figure 1. Daily closing price of CRM in USD from January 2016 to December 2019.

Figure 1 shows that the prices of CRM increase every year. One share of CRM cost $65 in the beginning of 2016 then by the end of 2019 the price of one share is $160. In December 2015, Salesforce agreed to purchase Steelbrick Inc, for $360 million and that made the stock price to drop from $75 to $55 in January 2016. Salesforce stock dropped sharply in December 2018 because of the market sell-off, which means a large volume of stocks are sold in a short period of time. While there was no clear catalyst for the price drop at the end of 2018, the common causes of the market sell-off include the release of disappointing earnings reports, fears of increased competition, or the threat of technological disruption. Figure 1 also shows that the time series of daily price of CRM is not stationary.

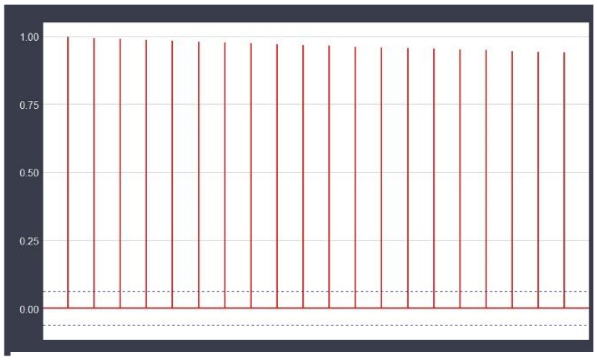


Figure 2. ACF of daily price of CRM

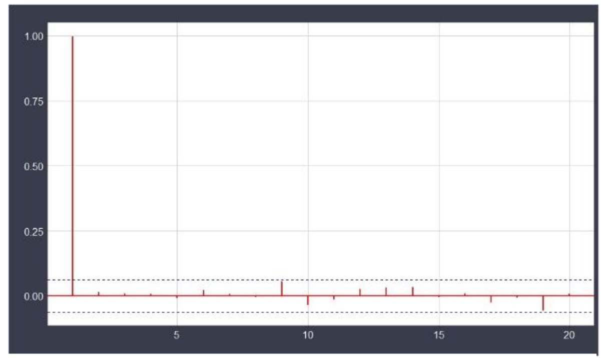


Figure 3. PACF of daily price of CRM

The ACF plot in Figure 2 shows there are significant autocorrelations with many lags in the time series. In Figure 3, the PACF plot only shows a spike at lag 1. By viewing the ACF and PACF, there is no evidence for seasonality.

1.3. The Daily Return of CRM

Transformation is necessary to achieve stationarity. The continuously compounded return on the t^{th} day of CRM which is defined as

$$r_t = log(p_t) - log(p_{t-1})$$

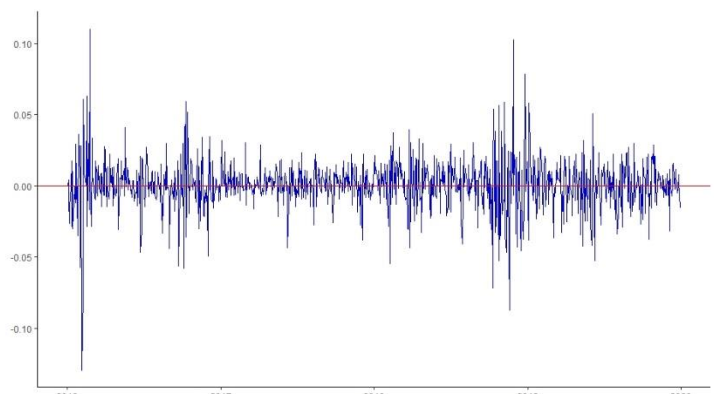


Figure 4. Daily return of CRM from January 2016 to December 2019 (Training Sample).

Training Sample: It consist of 985 times observations corresponding to 985 daily data points from January 2016 to December 2019.
Testing Sample: It consists of 20 times observations corresponding to 20 daily data points represent the last 20 days of December 2019.

By taking the first different of the logarithm of the data, the mean and variance are constant over time. Thus, the time series in Figure 4 is considered stationary and the Augmented Dickey-Fuller Test also confirm the stationarity with the p-value < 0.05.

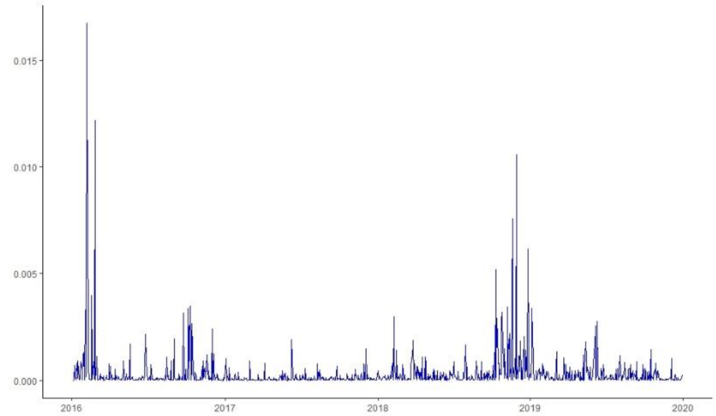


Figure 5. Squared daily return of CRM from January 2016 to December 2019.

From Figure 5, during the years 2016 and 2018-2019, there is spike in volatility indicating non-constant conditional volatility. There is evidence to use GARCH model as the time series are conditional heteroskedastic.

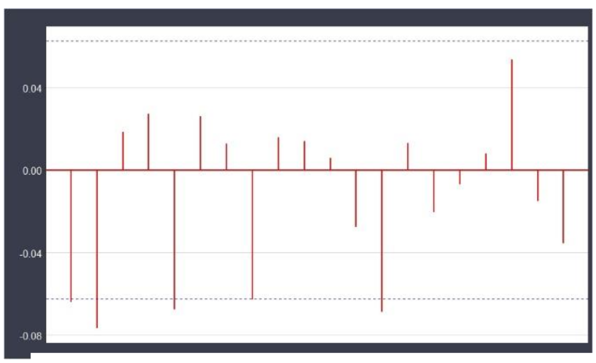


Figure 6. ACF of CRM daily return

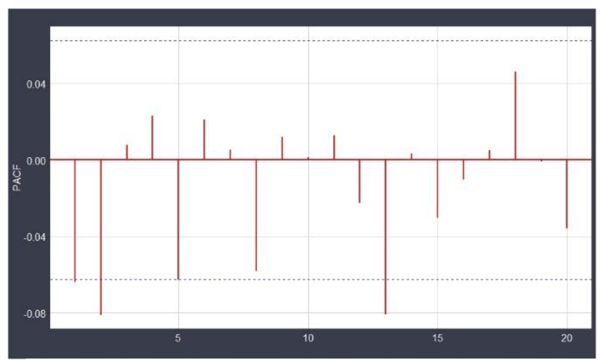


Figure 7. PACF of CRM daily return

2. Methods

2.1. Model Specification

2.1.1. ARIMA

The ACF in Figure 6 shows significant autocorrelation at lags 2. The PACF in Figure 7 shows significant autocorrelation at lags 2 and 13.

According to suggestions of the EACF in Figure 8, the algorithm is built to try 10 sets of different coefficient values. Based on comparing the Akaike Information Criteria (AIC), the best model to forecast daily CRM return is ARIMA(2,0,4) with the lowest AIC. The auto-arima function suggests ARIMA(0,0,2) but the AIC of the model is higher than other models.

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o x o o o  o  o  o
1 x o o o o o o x x o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x x o o o o o x o o o  o  o  o
4 x x o o o o o o o o o  o  o  o
5 x x x o x o o o x o o  o  o  o
6 x o x x o x o o x o o  o  o  o
7 o x x x o x o o o o o  o  o  o
```

Figure 8. EACF of CRM daily return.

ARIMA (p, d, q)	AIC
0,0,2	-5131.56
1,0,2	-5129.65
0,0,3	-5129.71
2,0,2	-5130.21
2,0,3	-5134.34
3,0,3	-5131.85
2,0,4	-5138.48

Table 1. Comparison of 7 models have the best AIC.

2.1.2. GARCH

According to suggestions of the EACF in Figure 9, the algorithm is built to try 6 sets of different coefficient values. Based on comparing the Akaike Information Criteria (AIC), the best model with the lowest AIC to forecast daily CRM return is the GARCH(2,1) with the estimated parameters of the fitted ARIMA(2,0,4).

```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x o o x  x  x  x
1 x o o o o x o o o o o  o  x  x
2 x x o o o x x o o o o  o  x  x
3 x x x o o x x o o o o  o  x  o
4 x o x o o x x o x o o  o  x  o
5 x o x x x x x o o o o  o  x  o
6 x o x x x x x o o o o  o  x  x
7 o o x x x o x x x o o  o  x  x
```

Figure 9. EACF of CRM squared daily return.

GARCH (p,q)	AIC
1,0	-5.270087
1,1	-5.487617
2,0	-5.364679
2,1	-5.489024
1,2	-5.485761
2,2	-5.487034

Table 2. Comparison of 6 models have the best AIC.

2.2. Model Fitting

2.2.1. ARIMA

The estimated parameter of the ARIMA(2,0,4) reflect the equation for AR(2) and MA(4) process

$$Y_t = -1.3770Y_{t-1} - 0.9621Y_{t-2} - 1.3265e_{t-1} - 0.8200e_{t-2} + 0.1730e_{t-3} + 0.0647e_{t-4} + e_t$$

```
ARIMA(2,0,4) with non-zero mean

Coefficients:
          ar1      ar2     ma1     ma2      ma3      ma4   mean
      -1.3770  -0.9621  1.3265  0.8200  -0.1730  -0.0647  9e-04
s.e.   0.0105   0.0094  0.0290  0.0513   0.0539   0.0347  5e-04

sigma^2 estimated as 0.0003132:  log likelihood=2577.24
AIC=-5138.48   AICc=-5138.33   BIC=-5099.34
```

2.2.2. GARCH

The equation of GARCH(2,1) model is

$$\sigma_t^2 = 0.000018 + 0.096872e_{t-1}^2 + 0.086121e_{t-2}^2 + 0.76215\sigma_{t-1}^2$$

The mean model ARMA(2,4) is

$$Y_t = -1.3929Y_{t-1} - 0.9742Y_{t-2} - 1.3676e_{t-1} - 0.9250e_{t-2} + 0.0623e_{t-3} + 0.0148e_{t-4} + e_t$$

```
GARCH Model      : sGARCH(2,1)
Mean Model       : ARFIMA(2,0,4)
Distribution     : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
mu      0.001249    0.000417     2.9930 0.002762
ar1    -1.392926    0.002456  -567.1999 0.000000
ar2    -0.974148    0.005487  -177.5390 0.000000
ma1     1.367571    0.001079  1267.9311 0.000000
ma2     0.924996    0.000071 13089.1392 0.000000
ma3    -0.062311    0.001951   -31.9368 0.000000
ma4    -0.014791    0.002907    -5.0878 0.000000
omega   0.000018    0.000006     3.2747 0.001058
alpha1  0.096872    0.039657     2.4427 0.014577
alpha2  0.086121    0.054437     1.5820 0.113645
beta1   0.762150    0.049977    15.2501 0.000000
```

2.3. Model Diagnostics

The Model Diagnostics is concerned with testing the goodness of fit of a model which describes how well it fits a set of observations. In this report, the analysis of residuals needs to examine the Residual plots, Density histogram of residuals, Normal plot of residuals, ACF of the standardized residuals, and the Ljung-Box Test.

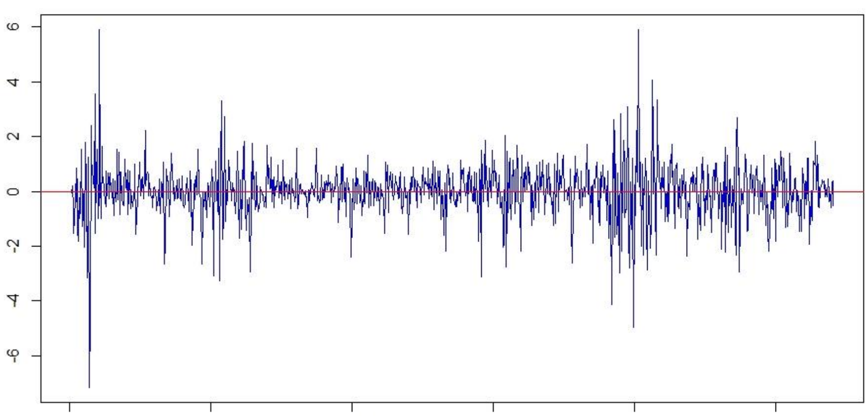


Figure 10. Plot of the standardized residuals (blue) from a fitted ARIMA(2,0,4) and a zero horizontal level (red).

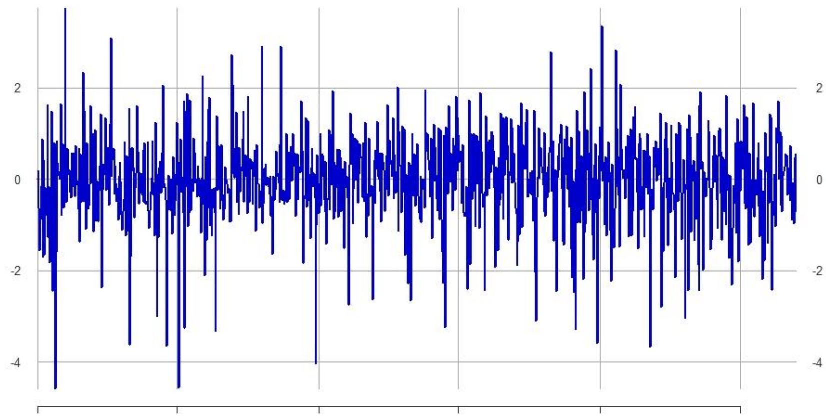


Figure 11. Plot of the standardized residuals (blue) from a fitted GARCH(2,1) and mean model ARIMA(2,0,4)

The standardized residuals in Figure 10 and Figure 11 look random with no trend whatsoever.

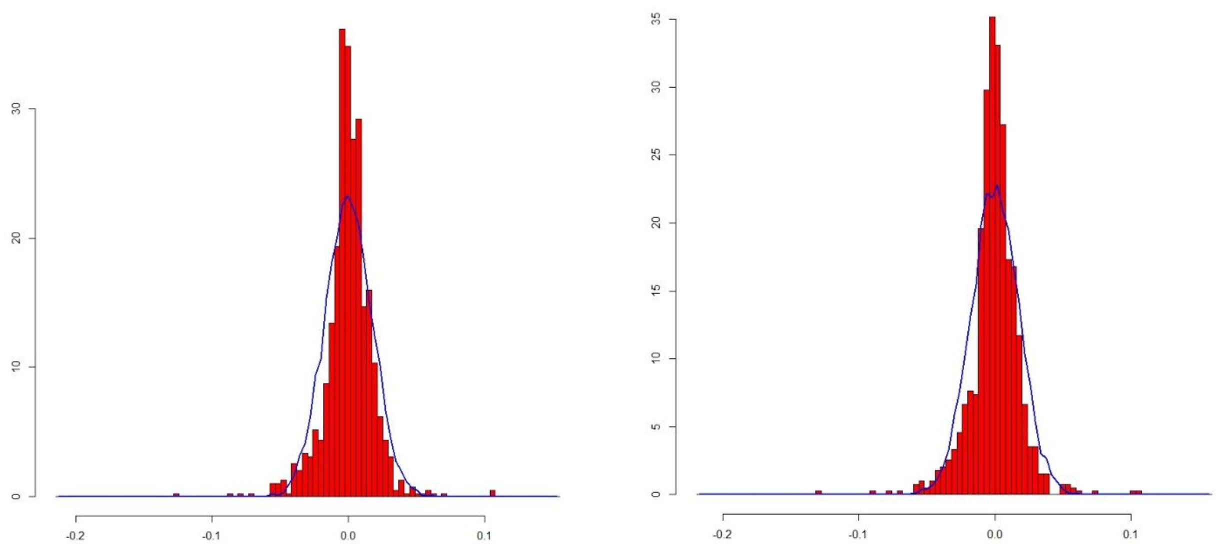


Figure 12. Density histogram of the residuals (red) and the density curve of the normal distribution with the same mean and variance (blue line). Left: ARIMA(2,0,4) model. Right: GARCH(2,1) model and the mean model ARIMA(2,0,4).

Figure 12 reflects the residuals are normally distributed. The ACF in Figure 13 and Figure 14 display the residuals look like white noise with other higher lags fall within the 95% confidence interval. The p-value for both ARIMA(2,0,4) and GARCH(2,1) all above 0.05 indicates that there is no evidence to against the null of no autocorrelation in the errors of the Ljung-Box Test. Therefore, ARIMA(2,0,4) and GARCH(2,1) with the mean model ARIMA(2,0,4) are adequate models.

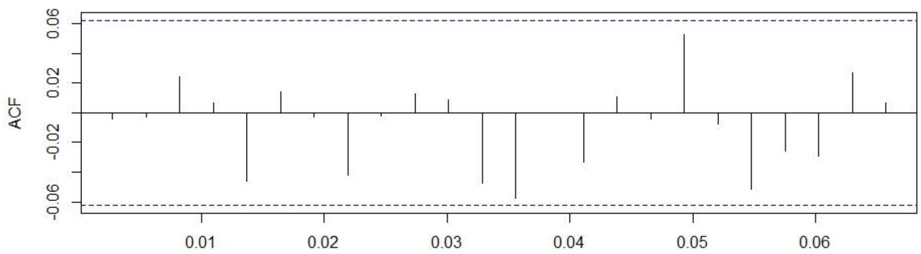


Figure 13. ACF of the standardized residuals of ARIMA(2,0,4).

Figure 14. ACF of the standardized residuals of GARCH(2,1).

3.Results

According to the methods above, ARIMA(2,0,4) and GARCH(2,1) are both the good fitting model for forecasting Salesforce (CRM) daily price. Looking at the values on Table 3, the actual values fall within 80% forecast limits for both models except the value No. 3 (yellow highlight). When consider the square of the difference of the actual price and the forecast price for both models, the GARCH(2,1) is slightly better than the ARIMA(2,0,4). Therefore, GARCH(2,1) with the mean model ARIMA(2,0,4) is the best fitting model for forecasting Salesforce (CRM) stock daily price.

As the result, this report has successfully found the adequate model to predict the Salesforce stock daily price in the last 20 days of 2019 base on the daily prices from January 2016 to December 2019.

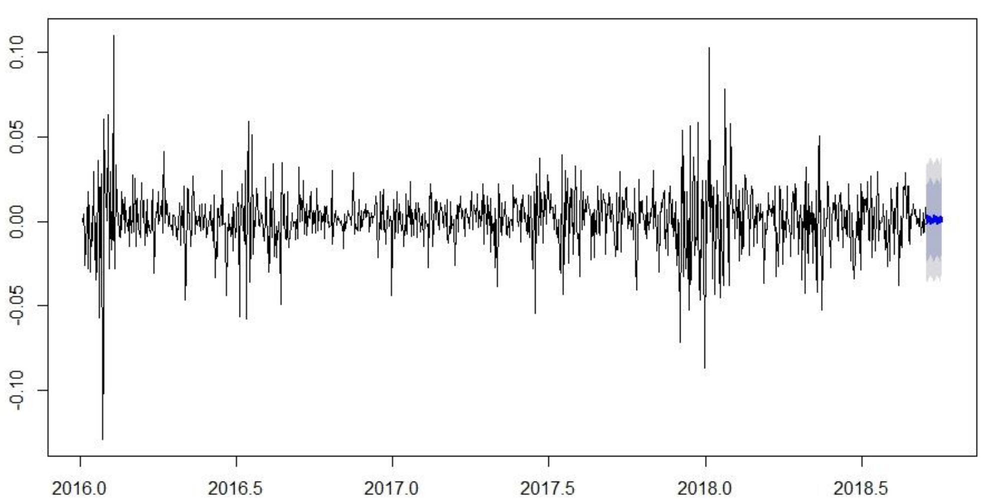


Figure 15. Forecasts from ARIMA(2,0,4) with non-zero mean (blue line), the 80% forecast limits (blue), and the 95% forecast limits (gray).

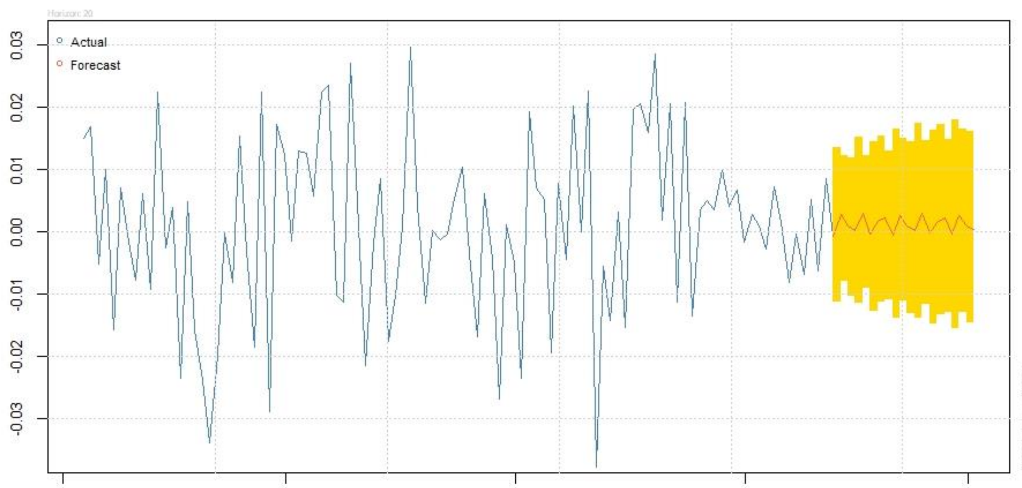


Figure 11. Forecasts from GARCH(2,1) and the mean model ARIMA(2,0,4) with the 95% forecast limits (yellow).

No.	Actual Price	ARIMA				GARCH			
		Forecast Price	80% lower	80% upper	(Actual – Forecast)^2	Forecast Price	80% lower	80% upper	(Actual – Forecast)^2
1	161.00	162.66	157.17	164.47	2.7556	160.87	158.72	163.05	0.0169
2	161.57	161.54	157.92	165.27	0.0009	161.45	159.24	163.68	0.0144
3	156.43	161.43	157.79	165.15	25	161.71	159.40	164.06	27.8784
4	158.22	156.52	152.99	160.13	2.89	156.46	154.13	158.82	3.0976
5	158.01	158.72	155.14	162.38	0.5041	158.71	156.26	161.19	0.49
6	157.48	157.72	154.17	161.37	0.0576	157.97	155.46	160.51	0.2401
7	156.40	157.88	154.32	161.52	2.1904	157.73	155.16	160.34	1.7689
8	156.39	156.61	153.07	160.23	0.0484	156.75	154.13	159.41	0.1296
9	158.59	156.20	152.68	159.81	5.7121	156.31	153.65	159.03	5.1984
10	161.13	159.14	155.54	162.81	3.9601	159.02	156.25	161.83	4.4521
11	161.96	161.05	157.41	164.77	0.8281	161.28	158.42	164.19	0.4624
12	161.63	162.04	158.38	165.78	0.1681	162.01	159.08	164.98	0.1444
13	161.48	162.10	158.44	165.85	0.3844	162.10	159.13	165.12	0.3844
14	163.33	161.26	157.61	164.99	4.2849	161.46	158.46	164.52	3.4969
15	164.55	163.69	159.99	167.47	0.7396	163.58	160.50	166.72	0.9409
16	163.74	164.78	161.05	168.59	1.0816	164.91	161.76	168.11	1.3689
17	163.25	163.58	159.88	167.37	0.1089	163.69	160.53	166.91	0.1936
18	164.51	163.75	160.04	167.54	0.5776	163.66	160.47	166.91	0.7225
19	164.98	164.48	160.76	168.29	0.25	164.67	161.43	167.98	0.0961
20	162.44	165.05	161.31	168.87	6.8121	165.04	161.76	168.38	6.76

Table 3. Summary results of forecasts from ARIMA(2,0,4) and GARCH(2,1) with the mean model ARIMA(2,0,4)

References

[1] Cryer, J. D., & Chan, K. (2011). Time series analysis with applications in R. New York: Springer.

[2] Salesforce. (2019). Retrieved from https://salesforce.com