Yuyi Zhang

Jun 06, 2020

Spring 2020 CPS Quarter Term B

Instructor: Danielle Mauro

Project 3:   
Price Predictions Project

ALY 6050\_Introduction to Enterprise Analytics

# **Introduction**

In project 3, we are going to forecast the stock price of Honeywell International Incorporated for Apr 16, 2018 based on historical data from 2017 to 2018 in three different approaches: simple exponential smoothing forecasts, adjusted exponential smoothing forecast, and simple regression analysis. Then we are going to compare and discuss the three results with the true price on Apr 16, 2018.

# **Analysis**

## **Simple Exponential Smoothing Forecasts**

Here is the historical data that we have on hand from Oct 15, 2017 to Apr 15, 2018.

A close up of text on a white background

Description automatically generated

*Figure 1*. Plots of Honeywell Stock Prices

Perform simple exponential smoothing forecasts using α = 0.15, 0.35, 0.55, and 0.75, then the results listed below:

|  |  |  |
| --- | --- | --- |
| **α** | **Forecast Price** | **MSE** |
| 0.15 | 145.3851 | 7.8774 |
| 0.35 | 145.4597 | 4.5049 |
| 0.55 | 145.8634 | 3.3853 |
| 0.75 | 146.0815 | 2.9539 |

*Table 1*. Predicted Honeywell Stock Prices by Simple Exponential Smoothing Forecasts

Mean Squared Error (MSE) is one of the measures of forecast accuracy. Knowledge of a forecast’s accuracy helps on selection of forecasting methodology to be utilized and parameters of the forecast formula. In this case, we use 4 different smoothing parameters. The parameter that leads to the smallest MSE yields the most accurate predictions of future stock value. Therefore, we have

.

A close up of a map

Description automatically generated

*Figure 2*. Simple Exponential Smoothing for α = 0.75

## **Adjusted Exponential Smoothing Forecasts**

Perform adjusted exponential smoothing forecasts using α = 0.75, successive values of 0.15, 0.25, 0.45, and 0.85 for the trend parameter β, then the results listed below:

|  |  |  |
| --- | --- | --- |
| **β** | **Forecast Price** | **MSE** |
| 0.15 | 146.9439 | 3.2275 |
| 0.25 | 147.5793 | 3.3306 |
| 0.45 | 148.0460 | 3.6454 |
| 0.85 | 149.7483 | 4.0723 |

*Table 2*. Predicted Honeywell Stock Prices by Adjusted Exponential Smoothing Forecasts

In this case, we use 4 different trend parameters β. Then we’ve got a list of MSE. According to the same theory as problem 1, off the 4 trend parameters, β = 0.15, MSE = 3.2275 yields the most accurate predictions of future stock price. Thus, we have

A close up of a map

Description automatically generated

*Figure 3*. Adjusted Exponential Smoothing for α = 0.75, β = 0.15

## **Simple Regression Analysis**

Perform a simple regression on stock price to get a summary shown below. The residual section shows all the residual quantile information (the distance from the data to the fitted line). Ideally, they should be symmetrically distributed around the line. That is to say, the minimum value and the maximum value would be the same distance from zero, as well as 1Q and 3Q. But in our data, the absolute value of min(-8.9502) and max(14.0026) are quite different.

The std. error and t value are provided to show how the p-value were calculated. If the t value is equal to zero, it means that the variable doesn’t have much use in the model. In this case, we’ve got a significant t-value of period (time) 0.698, which means that out data is not linear distributed. Moreover, the p-value(0.4863) is larger than 0.05, which is not statistically significant.

A screenshot of a social media post

Description automatically generated

*Figure 4*. Simple Regression Summary

In the coefficients section, we got the least-squares estimates for the fitted line. In this case, we’ve got our equation as:

### Coefficients of correlation and determination, and the interpretations of their values

Coefficient of correlation and the coefficient of determination are two statistics that describe the adequacy of a model.

Coefficient of correlation is a numerical descriptive measure of the linear relationship between two variables, in this case, stock prices and time. In the summary we could easily calculate that,

A value of *r* nearby 0 implies little relationship between two variables.

Coefficient of determination is a numerical descriptive measure of the contribution of one variable in predicting the other. If the independence value does contribute information for the prediction of the dependence variable, the absolute value of *r*2 would approach to 1. If not, then *r*2 will approach to 0. In this case, *r*2 = 0.003981, this means that only 0.3% of stock price could be explained by using time to predict in the straight-line model.

### A histogram of the regression residuals, and the interpretation of its shape

A screenshot of a social media post

Description automatically generated

*Figure 5*. Histogram of Residuals

This chart is skewed to the right with very few outliers. The distribution is unimodal with only one main cluster. There are 5 samples on horizontal axis, while middle 2 bars are higher than the rest, which are about 40. In another words, most of residuals are allocated between -5 to 5.

### A Chi-squared normality test of the residuals, and the interpretation of its outputs

Setup a Chi-squared normality test with the level of significance: α = 0.05. State null and alternative hypotheses as below:

Null hypothesis H0: the residuals follow the Normal distribution

Alternative hypothesis H1: the residuals do not follow the Normal distribution

A screenshot of a cell phone

Description automatically generated

*Figure 6*. Chi-squared Test Result

Compare p-value and significance value. In this case, the p-value (1) is larger than the significance level (α = 0.05). So, there isn’t sufficient evidence to reject H0. In another word, the residuals are normally distributed.

1. A Normal probability plot of the residuals

Prepare a normal probability plot by qqnorm() and qqline(), we have

A close up of a map

Description automatically generated

*Figure 7*. Normal Probability Plots

As we can see from the qq plots, the theoretical quantiles vs. sample quantiles are about to fit in a straight line. Therefore, the residual is normally distributed.

1. A scatter plot of residuals versus time to study their independency, and the interpretation of the shape of the scatter plot

A close up of a map

Description automatically generated

*Figure 8*. Residual Scatter Plot versus Time

The independency of residuals could be tested by residual scatter plots versus time. To be satisfied as independent, the residual should be randomly spread of the plot around zero. But in this case, we could see an increasing trend from 0 to 70 and decreasing trend in the rest. Therefore, the residuals are not independent.

1. A scatter plot of residuals versus predicted stock values to study their homoscedasticity, and the interpretation of the shape of the scatter plot

A close up of a map

Description automatically generated

*Figure 9*. Residual Scatter Plot versus Predicted Stock Values

Similar with problem 3.e, we generate a scatter plot of residuals versus predicted stock values to test their homoscedasticity. To be satisfied as homoscedasticity, the residual should also be randomly spread of the plot around zero. But in this case, we could see a similar trend up and down. Therefore, the residuals are not homoscedastic.

# **Conclusions**

All three forecasts result as listed below.

|  |  |
| --- | --- |
| **Forecasting Methods** | **Forecast Price** |
| Simple Exponential Smoothing Forecast | 146.0815 |
| Adjusted Exponential Smoothing Forecast | 146.9439 |
| Simple Regression Analysis | 151.5336 |

*Table 3*. Predicted Honeywell Stock Prices

According to Yahoo Finance, the actual Honeywell stock value on Apr 16, 2018 is $144.31. Therefore, the simple exponential smoothing forecast provides the most accurate result in this case. The simple regression has the least accurate forecast because the stock price is not liner distributed. The difference between simple and adjusted exponential smoothing forecast is that the adjusted exponential smoothing forecast explicitly recognize the trend in time series. But in this case, forecast without trend component perform a better result.

# 

# Reference

1. Simple Regression and Normal Probability Plot of Residuals. Retrieved from <https://northeastern.blackboard.com/webapps/blackboard/content/listContent.jsp?course_id=_2608793_1&content_id=_22434671_1>
2. Yahoo finance (Apr 16, 2018). *Honeywell International Inc.(HON)*. Retrieved from <https://finance.yahoo.com/quote/HON/>