Analyzing the Relationship Between Dividends and Volumes in the Dow Jones Industrial Average (DJIA)

Gregory Albarian

December 19, 2020

## Background Information

The public stock market can help people finance their money by taking a small partial ownership in the company. As the company makes more money, they can sell the stock for more money or buy at a higher price. Sometimes inflation hits, and stocks go up because the overall prices from a certain product goes up. Sometimes there is a recession involving high unemployment or low wages which will mean people may sell stock to cover costs or buy while it is still low. People might not have money to buy the stocks either. As stocks go up and down companies must incentivize investors to take on these risks.

Thirty stocks are known to be a part of the Dow Jones Industrial Average (DJIA). The DJIA is considered a good representation of the American stock market. This program will use data from 2018 mainly because it was easily accessible. In 2018, the thirty stocks the DJIA composed of were Verizon, IBM, ExxonMobil, Chevron, Proctor & Gamble, Coca-Cola, Pfizer, JPMorgan Chase, Cisco Systems, McDonald's, Merck & Company, 3M, Johnson & Johnson, Intel, DowDuPont, Travelers, Caterpillar, Walgreens Boots Alliance, United Technologies, Walmart, The Home Depot, The Boeing Company, Microsoft, American Express, Goldman Sachs, Walt Disney Company, UnitedHealth Group, Apple, Nike, and Visa. Some companies pay dividends. This is money the company give throughout periods to its stockholders. The longer the investor stays with the company the more opportunity they must make with dividends. Companies can change their dividends at any time. One reason an investor likes dividends is it can give them money from the stocks without having to sell it. All companies in the DJIA pay dividends. On the other hand, if a company needs to pay a large dividend it may need to give it to keep investors because it is not expecting much in growth. The volume of the stock is how many times this is bought or sold on each trading day. The downside of this statistic is it may be difficult to know whether it is being bought or sold. Also, whether it is being bought or sold by a financial institution or individual investor. This data could be interesting, however it could not be found.

## Question:

Is there a relationship between a stock's average volume and dividend in the DJIA in 2018?

## Methods:

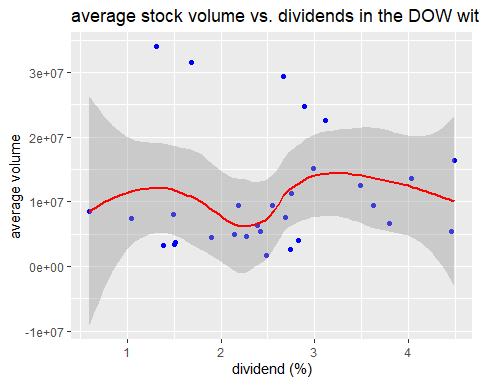
The stock volumes were provided by [Kaggle an open-source data site](https://www.kaggle.com/timoboz/stock-data-dow-jones?select=AAPL.csv). The dividends were provided by a web scraping a [website](https://www.reinisfischer.com/dow-jones-companies-yield-2018). While the web scraping could be very time consuming it was difficult to find anywhere else. The program was written an R to do the data cleaning and analysis. The average volumes of each stock were used. Each stock was contained within its own directory. This data was extracted and used only 2018 data. This was later converted into an easy-to-read data table with average volumes, dividends, and finding the ratio.

One important library used in this R script was [ggplot2](https://ggplot2.tidyverse.org/reference/geom_smooth.html). This library is used to do data visualizations. In the documentation for ggplot2, it says to use the loess method due to having less than 1,000 observations. The loess method is used because the DJIA only has thirty stocks in it. The Loess Method is regression method that was learned about to better understand what the program was doing. The Loess Method is very similar to a weighted least squares regression. It goes to each point and finds a subset of the points that are the closest to it. Each subset gets its own weighted least squares regression where the point that is currently being checked the next weights are assigned by how close they are to the original point. The Loess Method does this for each point. For closeness, is only measured in the x direction and not by a Euclidean Distance. When the Loess Method is at a certain point, there creates a new point of where that part of the weighted least squares regression is also at that same x value. For example, if the method is finding the value closest to the points where x = 2 and y = 3, and it takes from the nearby points and to create a weighted least squares regression where y = 4 at x = 2. Then (2,4) becomes a new used value. It goes through for all the points and all the new points until a smooth point is created (Cohen, 1999). The weighted least squares can be parabolic or linear. The R script created tried out both graphs. It also found the correlation.

## Results:

#### graph for y = x for each weighted squares analysis within the Loess Method

# graph  
graph<-ggplot(conjoined\_data, aes(x=dividends, y=average\_volume))  
graph\_title <- "average stock volume vs. dividends in the DOW with y = x loess method"  
graph <- graph + labs(title = graph\_title, x = "dividend (%)", y = "average volume")  
graph <- graph + geom\_point(colour="blue")  
graph <- graph + geom\_smooth(method = "loess", color = "red", formula = y ~ x)  
print(graph)



#### graph for y = x^2 for each weighted squares analysis within the Loess Method

graph<-ggplot(conjoined\_data, aes(x=dividends, y=average\_volume))  
graph\_title <- "average stock volume vs. dividends in the DOW with y = x^2 loess method"  
graph <- graph + labs(title = graph\_title, x = "dividend (%)", y = "average volume")  
graph <- graph + geom\_point(colour="blue")  
graph <- graph + geom\_smooth(method = "loess", color = "red", formula = y ~ poly(x,2))  
print(graph)

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at -2.8549 -1.2985

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 3.562

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 1.3817e-016

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 21.167

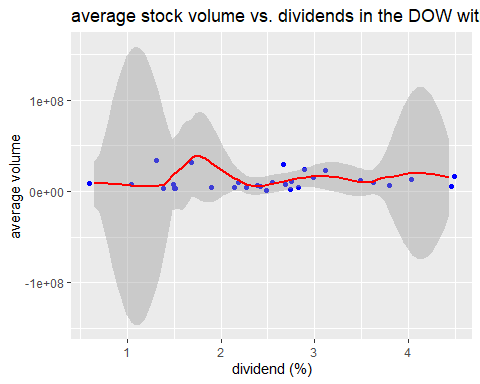
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## -2.8549 -1.2985

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 3.562

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 1.3817e-016

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other near  
## singularities as well. 21.167

## Warning: Removed 4 rows containing missing values (geom\_smooth).

 ### points on graph

conjoined\_data

## symbol average\_volume dividends ratio  
## 1: VZ 16473433 4.49 3668916.0  
## 2: IBM 5455029 4.46 1223100.7  
## 3: XOM 13640679 4.03 3384784.0  
## 4: CVX 6661690 3.80 1753076.4  
## 5: PG 9494895 3.63 2615673.5  
## 6: KO 12531564 3.49 3590706.1  
## 7: PFE 22633934 3.11 7277792.2  
## 8: JPM 15178733 2.99 5076499.4  
## 9: CSCO 24713430 2.89 8551359.9  
## 10: MCD 4068239 2.83 1437540.3  
## 11: MRK 11327173 2.75 4118972.0  
## 12: MMM 2607011 2.74 951464.0  
## 13: JNJ 7670742 2.69 2851576.8  
## 14: INTC 29403515 2.67 11012552.3  
## 15: DWDP 9510672 2.55 3729675.2  
## 16: TRV 1650737 2.49 662946.6  
## 17: CAT 5504695 2.42 2274667.4  
## 18: WBA 6399083 2.39 2677440.5  
## 19: UTX 4620617 2.27 2035513.9  
## 20: WMT 9469565 2.19 4324002.5  
## 21: HD 4983339 2.14 2328663.0  
## 22: BA 4459564 1.90 2347139.1  
## 23: MSFT 31591199 1.68 18804284.9  
## 24: AXP 3684340 1.51 2439960.0  
## 25: GS 3384592 1.50 2256394.4  
## 26: DIS 8042372 1.49 5397564.9  
## 27: UNH 3297169 1.39 2372063.9  
## 28: AAPL 34021450 1.31 25970572.2  
## 29: NKE 7418171 1.05 7064925.1  
## 30: V 8557225 0.60 14262040.9  
## symbol average\_volume dividends ratio

dividends are display as a percent #### correlation between dividends and average volumes

correlation <- cor(conjoined\_data$average\_volume, conjoined\_data$dividends)  
correlation

## [1] 0.04718087

## Conclusion:

The correlation between stock dividends and volumes was extremely low. While the parabolic graph may look like it fits the data better this could still be overfitting the data. This high volume could be due in stock dividends what are known as ex-dividend dates. The ex-date is the latest an investor can buy the stock and still get the dividend. I could not find this data either for when the ex-dates were for 2018 in these stock dividends. Some investors can buy exactly at the ex-date and sell once they get the dividend for what could possibly be a quick payoff. However, this technique may not work if the stock fall more than it can pay in dividends.

## Bibliography

Cohen, R. A. (1999, April). An introduction to PROC LOESS for local regression. In Proceedings of the twenty-fourth annual SAS users group international conference, Paper (Vol. 273).