

CA 3

Redmond Xia

January 28, 2020

Executive Summary

In assignment, we will explore the effects on E/P of companies from lagged change in percent earnings for each stocks. The idea is that when earnings changes, so should the E/P especially because it is in the numerator of this expression. We expect when the lagged change effect should be positive with E/P since if change is positive, we should be earning more. We examine S & P 500 stocks from 2008 to 2013.

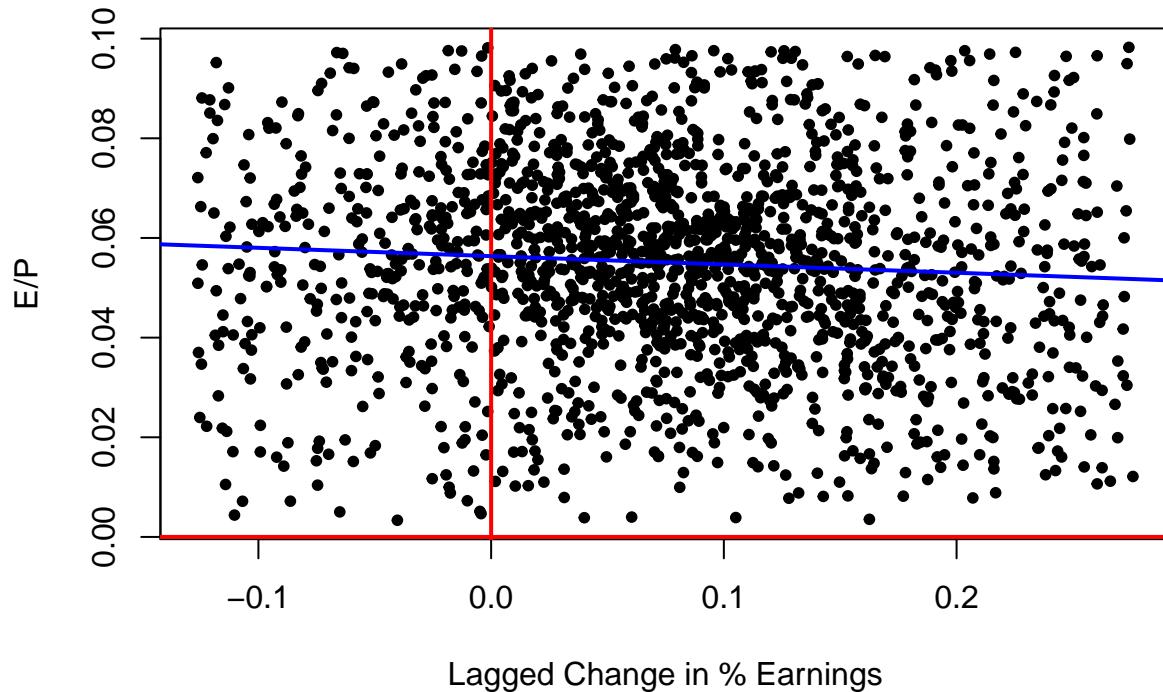
Assumptions and Methods

The data I used to was found in WRDS/Get Data/Compustat - Capital IQ/North America - Annual Updates/Fundamentals Annuals. Then I found the tickers of the S&P 500 through datahub <https://datahub.io/core/s-and-p-500-companies>. The variables that were relevant were EPS, EBIT, and Price for each tickers. We assume that if we divide that year's EPS with Price per share, we would get E/P. As noted, this may not be the most accurate way of measuring the yearly E/P. Next, I used the difference of log methods to estimate the changes in percent earnings. Since earnings could be negative sometimes, log isn't define for negative values. Thus, I transformed it to a positive value then changed it back to a negative change. This way $(1+r)^*x_{t-1}$ is the valid growth. After finding the correct r we take the rolling average of three years of the change, which starts our data in 2005. This will smooth out the curve. Then we omit 2005 to 2007 and start our data in 2008. Lastly, we take out the outlier from the bottom 10 percent and the top 10 percent of the data.

Results and Answers.

As you can see from the plot below, there is a lot of noise to the data which goes against the expectation as stated earlier. This is slightly surprising but this is mostly due to the fact that most earnings to price is positive because the S&P 500 is mostly matured firms, and changes in earnings could be negative. Thus we see this on the graph below. I ran the regression to if this result is consistent. From the regression we see that the best fit line is negatively sloping, which goes against the expectation mentioned in the summary. Next we analyze the regression summary.

Lagged Change in % Earnings on E/P from 2008 to 2013



Regression

Below, we can see that the correlation between the two is very low. After seeing the summary of the regression, we see that we have a low R^2 and standard error is fairly high. We conclude that the linear regression is not robust enough, even though lagged change in percent earnings is a good predictor. We also get less variability when we take only the middle 80% of confidence in the data. Next, let us explore the nature of our data.

```
## [1] "Regression Summary output"
##
## Call:
## lm(formula = sp500$epsfx ~ sp500$avgGrowthEbit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05380 -0.01320  0.00055  0.01411  0.04647
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0563526  0.0006476 87.020 < 2e-16 ***
## sp500$avgGrowthEbit -0.0166685  0.0054939 -3.034  0.00245 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02066 on 1733 degrees of freedom
## Multiple R-squared:  0.005284,    Adjusted R-squared:  0.00471
```

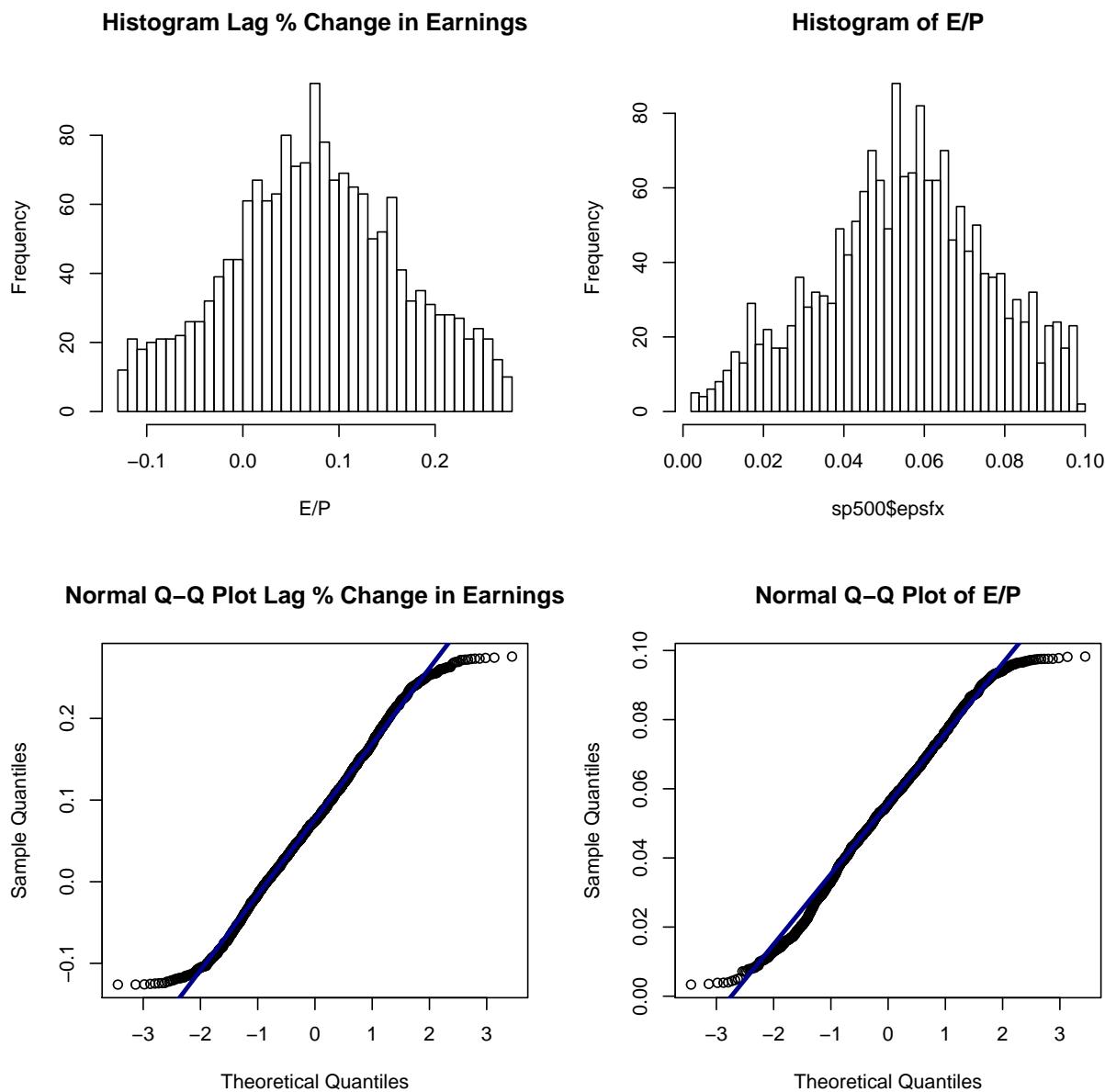
```

## F-statistic: 9.205 on 1 and 1733 DF, p-value: 0.002449
## Correlations between x and y: -0.07268897

```

Data Exploration

Below are histograms and their Q-Q plots. We see that the nature of data is slightly skewed but looks fairly symmetrical around the mean. The E/P is left skewed which makes sense since the firm is trying to optimize earnings. Both of the data has positive mean, however it is interesting to note that they have fatter tails than a normal distribution as seen through the QQ plots. This means that it is very possible that firms are riskier since they sometime grows really fast or really slow. This can explain the extra noise in our data and the regression model. We conclude that it is hard to build a model of E/P based on Lag % Changes in Earnings. Another explanation to why this is hard is because difference in log is biased when the changes in x is large. Sometimes, a company has grown its earnings by triple amount, which is considered large so we introduce some bias.



Appendix (Codes)

```

# Maintaining, Clearning and Transforming data
suppressWarnings(suppressMessages(library(zoo)))
sp500 <- read.csv("sp500.csv")[-1]
sp500 <- na.omit(sp500)
tickers <- unique(sp500$tic)
sp500$epsfx <- sp500$epsfx / sp500$prcc_f
pct_chg <- c()

for(i in tickers) {
  stock <- sp500[sp500$tic == i,]

  # For negative Ebits, we turn it positive so that we can take the log
  negEbit <- stock$ebit < 0
  if(length(negEbit) != 0) {
    stock[negEbit,]$ebit <- -stock[negEbit,]$ebit
  }
  # The percent change in earnings as difference log method
  growth <- diff(log(stock$ebit))
  if(length(negEbit) != 0) {
    # Transforming the growth into their respective negative term
    growth[diff(negEbit) != 0] <- -(growth[diff(negEbit) != 0] + 1) - 1
  }
  growth <- na.omit(growth) # Running the above loop introduces NA in the last column

  # This takes the rolling mean of the three year back data.
  stock$avgGrowthEbit <- rollmean(c(0,growth), fill = 0, k = 3, align = "right")
  # Changing the original Ebit that were negative back to negative
  if(length(negEbit) != 0) {
    stock[negEbit,]$ebit <- -stock[negEbit,]$ebit
  }
  pct_chg <- rbind(pct_chg, stock)
}

# Taking out the years with no averages
sp500 <- pct_chg[-which(pct_chg$avgGrowthEbit == 0 | pct_chg$fyear == 2007) ,]

# Keeping the 80% confidence interval
conf_int90Eps <- quantile(sp500$epsfx,c(.1,.9))
sp500 <- sp500[which(sp500$epsfx > conf_int90Eps[1] & sp500$epsfx < conf_int90Eps[2]),]
conf_int90AvgGr <- quantile(sp500$avgGrowthEbit,c(.1,.9))
sp500 <- sp500[which(sp500$avgGrowthEbit > conf_int90AvgGr[1] &
                     sp500$avgGrowthEbit < conf_int90AvgGr[2]),]
rownames(sp500) <- 1:nrow(sp500)
plot(sp500$avgGrowthEbit,sp500$epsfx, pch = 20, xlab = "Lagged Change in % Earnings",
      ylab = "E/P", main = "Lagged Change in % Earnings on E/P from 2008 to 2013")
abline(h = 0, v = 0, col = "red", lwd = 2) # This gives the y and x axisd
abline(lm(sp500$epsfx~sp500$avgGrowthEbit), col = "blue", lwd = "2")
print("Regression Summary output")
summary(lm(sp500$epsfx~sp500$avgGrowthEbit))
cat("Correlations between x and y: ", cor(sp500$avgGrowthEbit,sp500$epsfx),"\n")
par(mfrow = c(2,2))
hist(sp500$avgGrowthEbit, xlab = "E/P", main = "Histogram Lag % Change in Earnings", breaks = 50)
hist(sp500$epsfx, main = "Histogram of E/P", breaks = 50)

```

```
qqnorm(sp500$avgGrowthEbit, main = "Normal Q-Q Plot Lag % Change in Earnings")
qqline(sp500$avgGrowthEbit,col = "darkblue", lwd = 3)
qqnorm(sp500$epsfx, main = "Normal Q-Q Plot of E/P")
qqline(sp500$epsfx,col = "darkblue", lwd = 3)
```