

ECO395 Data Mining Final Project: Predicting analyst forecast accuracy under complex financial information.

Wen-Hsin(Molly) Chang

2021/5/1

Abstract

Analyst's earnings forecast has played an important role in the capital market since it facilitates investors' judgments. As such, this project aims to examine how financial analyst's forecast accuracy evolves under complex financial information (i.e., derivative instruments), and predict analyst's forecast accuracy under various supervised learning methods. I identify US "derivative user" firms by conducting web-scraping and textual analysis on each firm's annual report (10-K filings) through SEC's EDGAR database. The results show that derivatives instruments diminish analyst's ability to forecast EPS. Moreover, after comparing the out-of-sample performance between KNN, tree model, stepwise regression model, and linear model, I find that the stepwise regression model performs the best for predicting analyst's forecast accuracy in the face of derivative instruments. One possible explanation is probably due to the complex interactions between features. In the appendix, I establish a causal link between derivative instruments and analyst's forecast accuracy using a difference in difference method.

1.0 Introduction: Background or motivation

Analyst's earnings forecast has played an important role in the capital market since it facilitates investors' judgment, which in turn improves market efficiencies. Financial analysts have long been regarded as financial experts who are less likely to misinterpret complex information in the financial market. However, prior studies suggest that analysts' forecast uncertainty is associated with the amount and quality of a firm's information (Atiase 1985; Barron and Stuerke 1998). As such, this project is motivated by how analyst's earnings forecast accuracy evolves in the face of complex financial information in recent decades. Specifically, I focus on the financial reporting complexity caused by the emergence of derivative instruments (e.g., futures, options, swaps... etc.)

According to the Bank for International Settlement, the total notional amount of global derivatives usage has increased by 700 percent in the past twenty years. However, derivatives are difficult to comprehend for several reasons. Firstly, derivatives can be linked to virtually any type of underlying asset or liabilities. Due to the complex structure, the derivative contract often involves meticulous designs and difficult valuation methods. Secondly, the intent behind holding derivatives may vary (i.e., hedging versus speculative) and thus entail elusiveness. If a firm employs derivatives for hedging purposes, the risk exposure of the firm decreases, while if a firm holds derivatives for speculating purposes, the risk exposure of the firm increases. Therefore, firms may exercise considerable judgment when reporting relevant financial information to the public.

Chang et al. (2016) find that in the face of the complexity of derivatives, even sophisticated financial analysts routinely misjudge the earnings implications once a firm initializes a derivative program. Due to the complex nature of financial derivatives, the financial accounting standard board (FASB) has issued several accounting standards governing derivatives activities over the past few years, with the most recent one-SFAS No.161 in 2008. As derivatives have become more complex and pervasive, it is worthwhile to pay attention to the effect of derivatives on analyst forecast accuracy, and to predict analyst forecast accuracy under the influence of a firm's engagement in derivative instruments.

2.0 Data and Research Methods

2.1 Data

To obtain data regarding derivative usage, I conduct a textual analysis* on all the firm's annual reports (10-K filing) by scraping the SEC's EDGAR database. (*flagged all annual reports and search for specific terms: forward contract, option contract, futures contract, rate swap, swap agreement, currency exchange contract, foreign exchange contract, derivative instrument, and hedging instrument). I include U.S. sample for fiscal years 2004-2013, which is five years before and after the adoption of SFAS No.161 in November 2008. Samples are partitioned into either derivative USER or Non-USER.

Next, I merge the data above with firm's accounting variables from Compustat database, CEO option holdings from Execucomp database, and stock price information from CRSP database. After dropping observations without necessary data, the final sample consists of 5,820 firm-year observations, including 4,521 users and 1,299 non-user observations.

2.2 Research Methods

I conduct both visual analyses and supervised learning in this project. Firstly, in the visual analysis section, I provide several figures to illustrate (1) how derivatives usage affect analysts forecast accuracy, (2) how the US mandatory derivative disclosure requirements SFAS No.161 after 2009 improves analysts forecast accuracy, and (3) what's the determinants of analysts forecast accuracy. Next, in the supervised learning sections, I try to predict analyst's forecast accuracy using the following four methods: linear model, KNN model, Stepwise regression, and Tree model and determine which method performs the best in my research question. In the appendix, I conduct the standard difference-in-difference research design to validate the enhanced disclosure requirement of SFAS No.161 truly improves analyst's forecast accuracy for firms that engage in derivatives activities.

2.3 Variable definitions

AFACC: Analyst's forecast accuracy, which is the accuracy of the last annual earnings forecast before earnings announcement date for year t . Calculated as the absolute value of forecast error multiplies -1000, scaled by the most recent stock price in the previous year. A larger AFACC implies more accurate forecasts.

SIZE: Defined as the natural logarithm of total assets

LEV: Defined as total liabilities divided by total assets at the beginning of the year

MB: Market to book ratio. Defined as market value divided by book ratio

ROE: Defined as income before extraordinary items scaled by lagged total equity

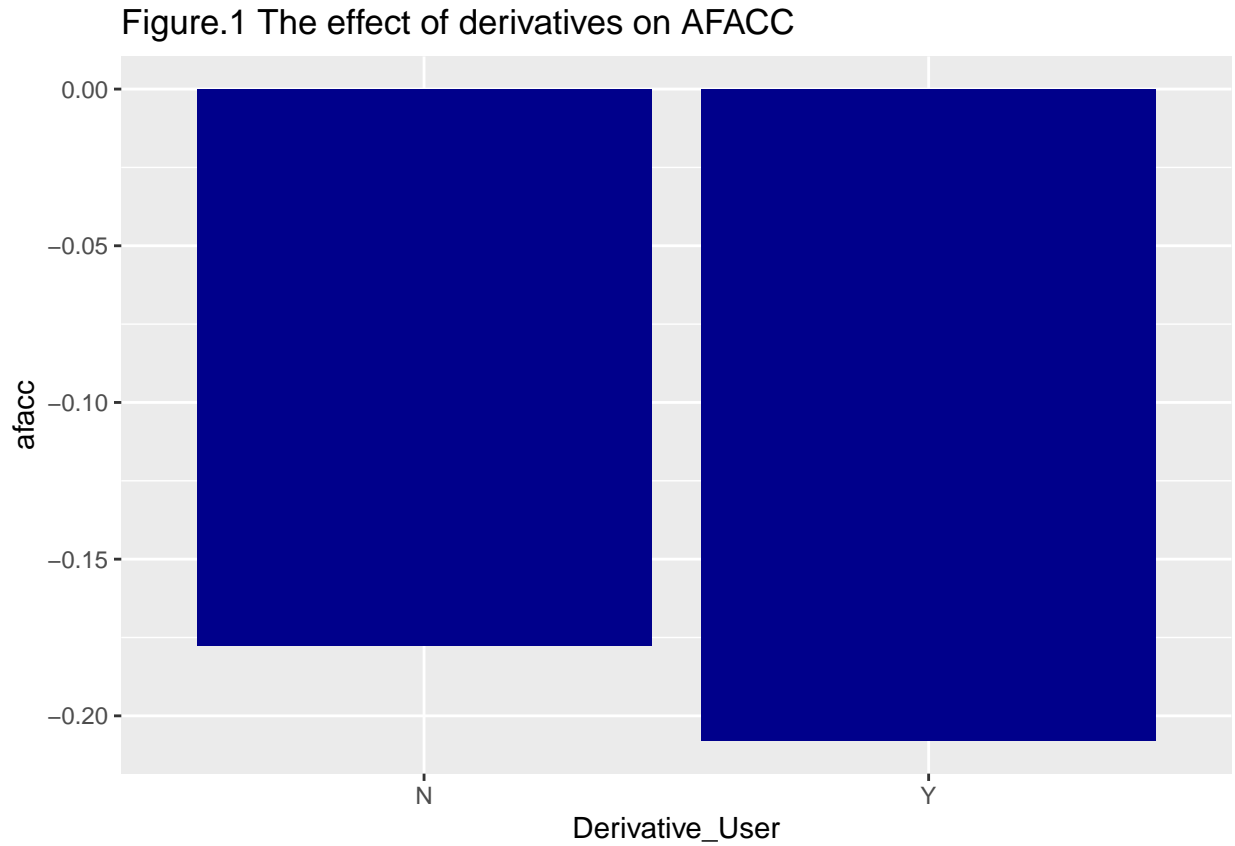
FOLLOW: Number of analysts following at the end of year t

HORIZON: Analyst forecast horizon

3.0 Results

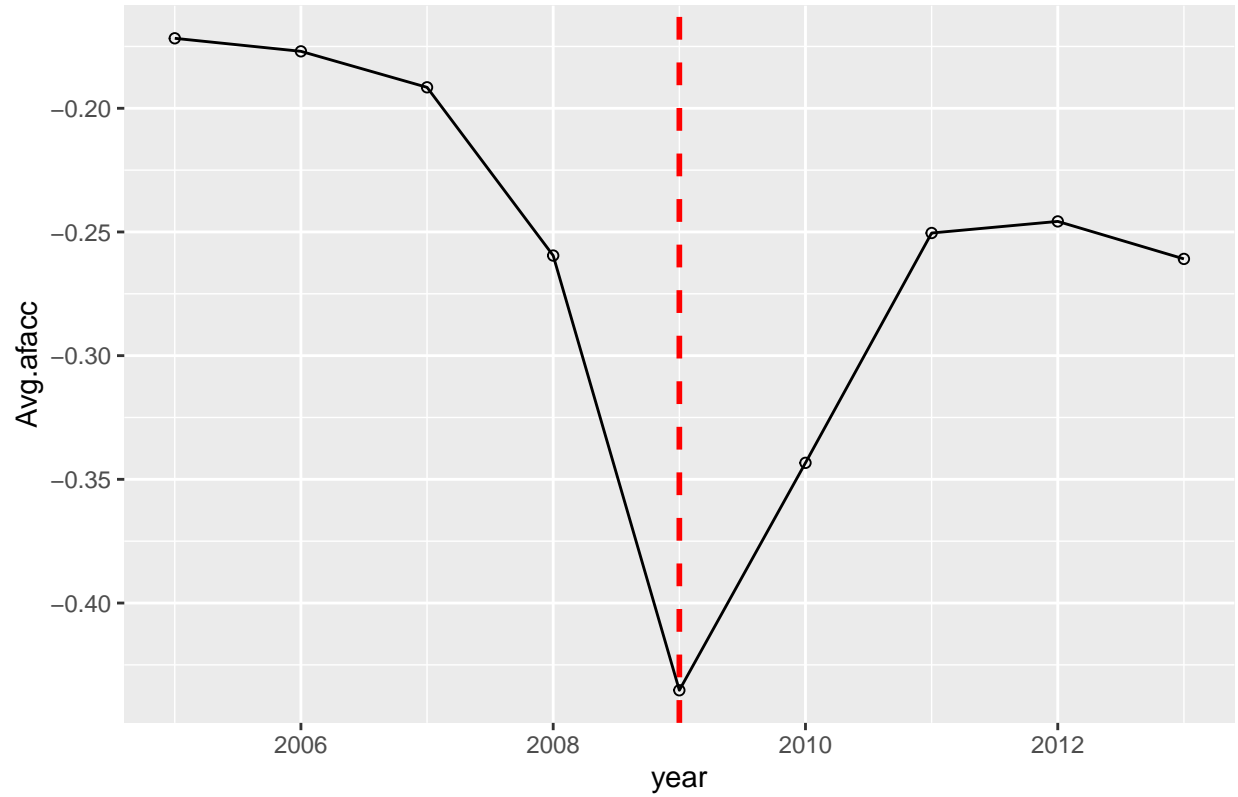
3.1 Visual Analysis

Figure 1. shows that for firms with derivatives usages, analyst's forecast accuracy is lower compared to firms without derivatives usage. The result provides preliminary evidence that the complex nature of derivatives instruments diminishes analyst's ability to forecast EPS.



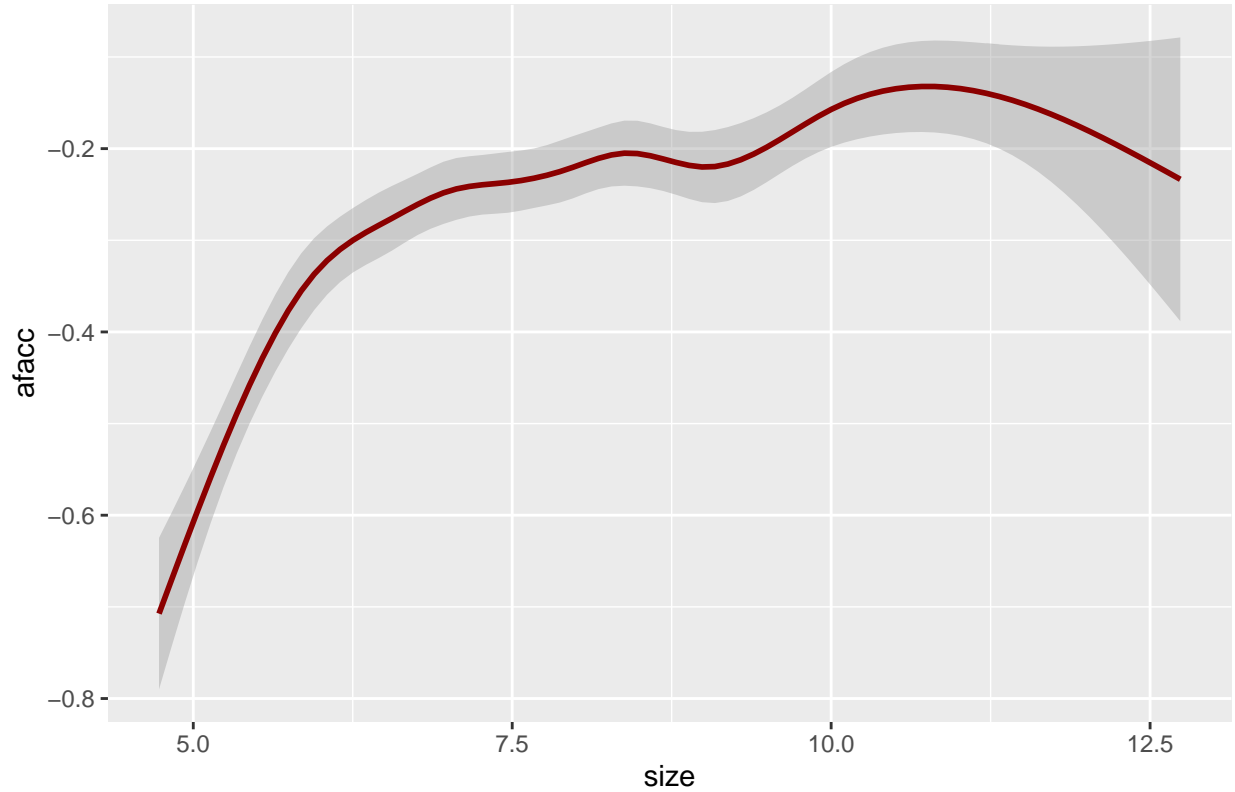
To further validate the effect of derivatives, I examine the time-series trend of analyst's forecast accuracy before and after the US mandatory derivatives disclosures (SFAS No.161). Figure 2. indicates that after the adoption of SFAS No.161, analyst's forecast accuracy reverse from the downward trend and improved quite significantly.

Figure.2 The effect of mandatory derivatives disclosures on AFACC



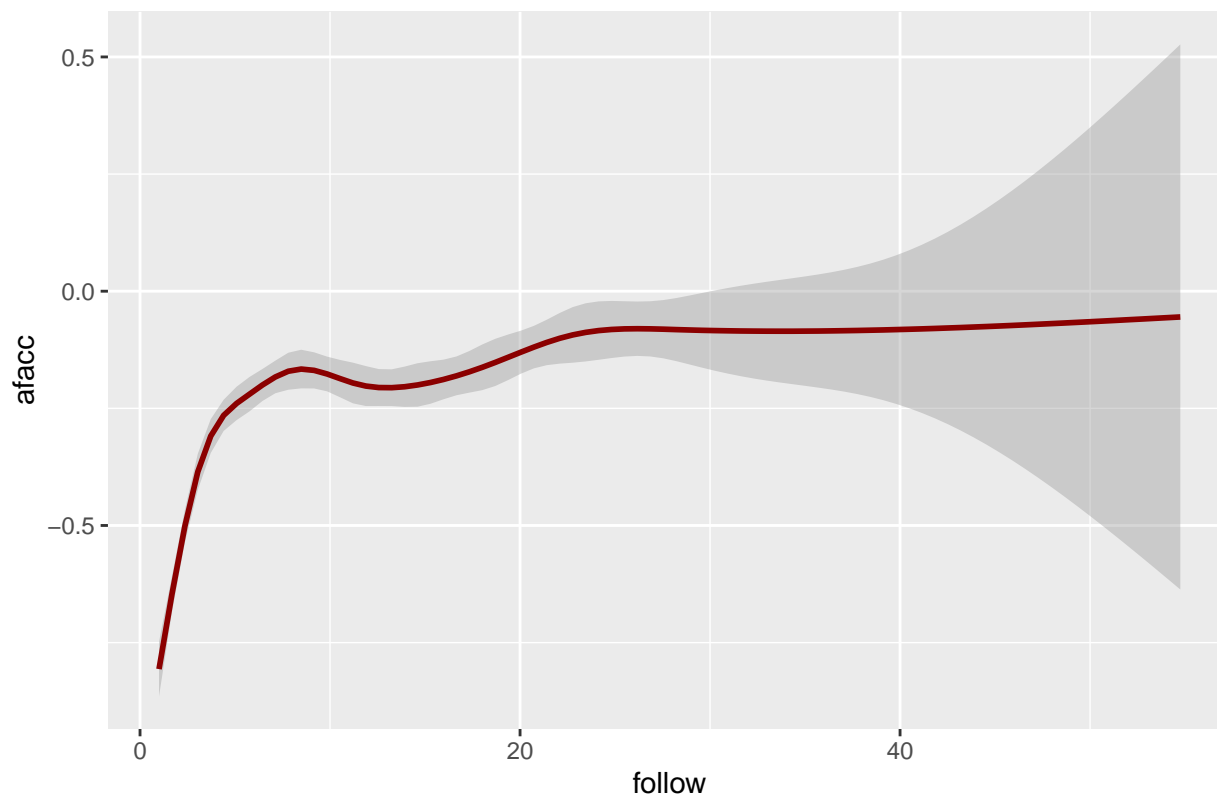
Next, I examine other factors that may affect analysts forecast accuracy. Figure 3 indicates that the larger the firm is, the more accurate analysts will perform. This is quite intuitive since large firms will have more media coverage, and therefore more information available for analysts.

Figure.3 The effect of firm size on AFACC



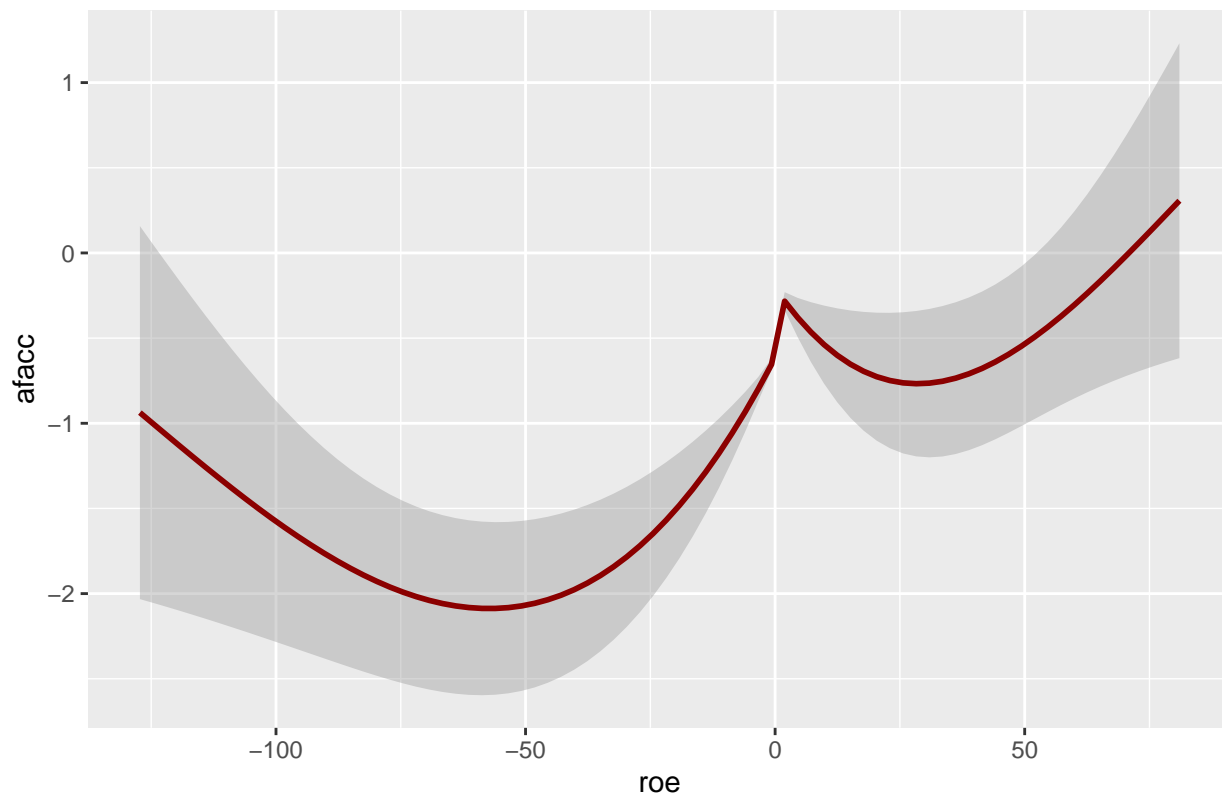
In Figure 4, I look at whether the number of analysts following (covering) is positively related to analyst's forecast accuracy. Prior studies suggest that firms with more analyst coverage are under pressure to release more transparent information during the conference call. Figure 4 supports such a notion.

Figure 4. The effect of the number of analysts following on AFACC



Finally, I provide a visual analysis on whether firm's return on equity (ROE) affects analyst's forecast accuracy. Figure 5 shows that the effect of ROE on forecast accuracy is not monotone. It decreases at first and bounces back after a certain threshold.

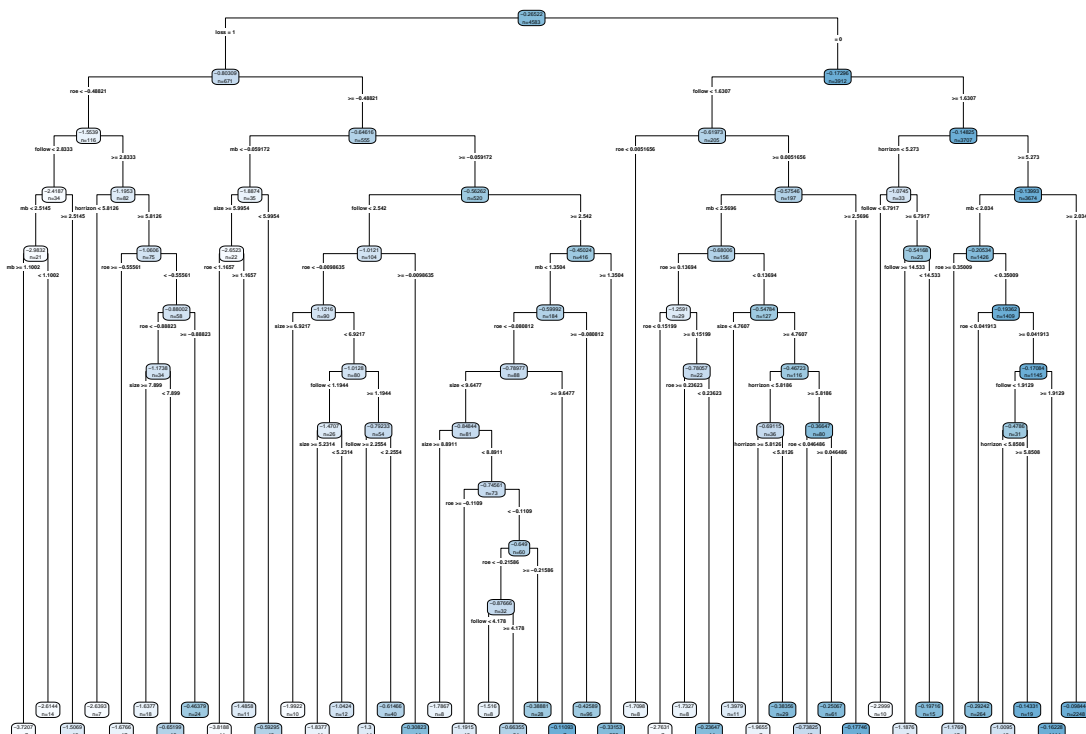
Figure.5 The effect of ROE on AFACC



3.2 Prediction/Supervised learning

After conducting visual analysis, I turn to supervised learning methods, aiming to predict analyst's forecast accuracy in the face of complex derivative instruments. First, I use a **forward stepwise regression** technique to get a better sense of which variable should be included in the model since I have numerous determinants. The stepwise regression result shows that the optimal combination is “**user + loss + horizon + follow + size + loss:follow + user:horizon + user:loss + loss:size + horizon:size + follow:size + horizon:follow + user:size**”, which is pretty much related to my findings in the previous visual analysis section.

Next, I turn to the tree model because there are several features in my dataset, and I suspect there are more interactions between the features that can't be solved using linear models. I plot my tree model below with $cp=0.00015$ and $type=4$.



For comparison, I also run a KNN method to see if it perform better in my data set. After several runs of analysis, I choose $k=100$ in my KNN method.

Finally, I compare the out-of-sample performance of predicting analyst's forecast accuracy between linear model, KNN model, Stepwise model, and Tree model based on 30 folds cross validations. After multiple examinations, I find that the stepwise regression model seems to attain the lowest RMSE compared to all the other methods. In my opinion, the stepwise regression model is suitable in my setting because, as suggested by prior literature, various features and interactions will determine analyst's forecast accuracy.

```
## [1] "The model with lowest RMSE: Stepwise Regression Model "
```

4.0 Conclusion

This project aims to examine how financial analyst’s forecast accuracy evolves under complex financial information (i.e., derivative instruments), and predict analyst’s forecast accuracy under various supervised learning methods. I identify US derivative user firms by conducting web-scraping and textual analysis on each firm’s annual report (10-K filings) through SEC’s EDGAR database. The results show that derivatives instruments diminish analyst’s ability to forecast EPS. Moreover, after comparing the out-of-sample performance between KNN, tree model, stepwise model, and linear model, I find that the stepwise regression model performs the best for predicting analyst’s forecast accuracy in the face of derivative instruments. One possible explanation is probably due to the complex interactions between features.

5.0 Appendix. Establish a causal link using Difference-in-Difference method

In the appendix, I establish a causal link between derivative instruments and analyst's forecast accuracy using a difference in difference method. The FASB's mandatory derivatives disclosures (SFAS No.161) in the US serves as an exogenous shock to the derivative users, which can be considered as the treatment group. The result below shows that the coefficient of the interaction term "post x treat" is extremely significant, indicating that the improvement of derivative disclosure truly affects analyst's forecast accuracy, and thus we should indeed take derivative usage into account when predicting analyst's forecast accuracy.

```
##
## Call:
## lm(formula = afacc ~ user + post + post_treat + size + loss +
##      roa + roe + mb + altz, data = Audrey_new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43846 -0.00326  0.00731  0.01712  0.18089
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.498e-02  4.266e-03 -12.886  < 2e-16 ***
## user         2.557e-03  1.697e-03   1.507  0.131741
## post        -1.934e-02  1.787e-03 -10.820  < 2e-16 ***
## post_treat   1.282e-02  2.014e-03   6.365  2.10e-10 ***
## size         4.704e-03  4.880e-04   9.639  < 2e-16 ***
## loss        -4.709e-02  2.659e-03 -17.709  < 2e-16 ***
## roa          1.493e-02  8.781e-03   1.700  0.089161 .
## roe          8.908e-04  2.599e-04   3.427  0.000614 ***
## mb          -1.894e-05  2.182e-05  -0.868  0.385365
## altz         1.155e-03  2.040e-04   5.661  1.58e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0552 on 5594 degrees of freedom
## (216 observations deleted due to missingness)
## Multiple R-squared:  0.1552, Adjusted R-squared:  0.1538
## F-statistic: 114.2 on 9 and 5594 DF,  p-value: < 2.2e-16
```

Reference

Atiase, R. 1985. Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research* 23(1): 21–36.

Barron, O. E., and P. S. Stuerke. 1998. Dispersion in analysts' earnings forecasts as a measure of uncertainty. *Journal of Accounting, Auditing and Finance* 13 (Summer): 245–270.

Chang, H., M. Donohoe, and T. Sougianis. 2016. Do analysts understand the economic and reporting complexities of derivatives? *Journal of Accounting and Economics* 61(2-3): 584-604.