FACTORS AFFECTING INVESTMENT RESEARCH

A STUDY ON THE IMPACT OF MIFID II ON ANALYST COVERAGE

SHIH JUNG YAPE

Bachelor Thesis

Stockholm School of Economics

2020

Factors Affecting Investment Research – A Study on the Impact of MiFID II on Analyst Coverage

Abstract:

This paper studies the effects of MiFID II and firm's performances on analyst coverage. The financial regulation that was imposed in Europe, January 3, 2018, drove analysts to become more transparent with the presentation of their cost structure, leading to unbundling of research and trading costs. Consequently, investment firms had to re-evaluate their research strategy since the margins had decreased. My study applies a Difference-in-Differences Method and Multiple Linear Regression to examine how MiFID II has affected the coverage on Earnings per Share for European Economic Area and North American firms. Consistent with previous papers published in 2019 I present empirical evidence that the amount of firm coverage had decreased after the implementation of MiFID II. However, my results deviate from the preceding studies by showing an increased forecast error compared to pre-MiFID II.

Keywords:

Analyst Coverage, European Economic Area, Forecast Error, Investment Research, MiFID II

Authors:

Shih Jung Yape (23699)

Tutors:

Olga Obizhaeva, Assistant Professor, Department of Finance

Examiner:

Adrien d'Avernas, Assistant Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

Acknowledgements

I would like to express my deep gratitude to my supervisor Assistant Professor Olga Obizhaeva for sharing her knowledge within the field of finance and econometrics. Without Obizahaeva's guidance and her flexible schedule, the thesis would have posed a tougher challenge than it should have. I could not have imagined a more suitable candidate in counselling me than Obizhaeva and hope that future students will value her effort and expertise as much as I did. Beside my supervisor, I would also like to acknowledge my family for their ever continuous support throughout my education. Their inspiration and love have given me the extra drive to complete my Bachelor Thesis in Stockholm School of Economics.

Jung Yape
Stockholm School of Economics
Stockholm, May 14, 2020

This page is intentionally left blank

Table of Contents

1 Introduction

The Markets in Financial Instrument Directive (MiFID) was the initial legislative framework imposed by the European Union (EU) member states in 2007. The goal was to increase the transparency in the financial market so that investors would receive better protection. In the aftermath of the 2008 financial crisis EU saw further needs to revise and improve the regulations in the financial markets which led to The Markets in Financial Instruments Directive II (MiFID II). In similarity to its predecessor, MiFID II intended to ensure and solidify the transparency between service providers in the market and investors. One of the more significant regulatory changes imposed in MiFID II is that investment firms in both buy- and sell-side must associate all their costs to the services they provide, thus minimizing the exposure that investors have from potential frauds (The European Union 2014).

Prior to MiFID II lawmakers were concerned with the bundled payments that investment firms received from their clients. Brokers and banks were bundling research with trading costs and received commissions as payment ("soft dollars"), thus inducing investment fund to direct trading towards the brokerage. Previous studies provides empirical evidence that analysts may strategically put more effort in generating "soft dollars" when research is bundled with transaction costs (Harford et al. 2018). As a consequence, analysts may have been producing inaccurate and biased estimates to attract more clients. (Hong and Kubik 2003; Hong and Kacperczyk 2010; Karmaziene 2019).

As of January 3, 2018, brokerages and investment firms had to unbundle their research costs from trading and execution costs, forcing them to present a more transparent cost structure (The European Union 2014). This gave investors wider choices of investment research without being locked to tempting offers from brokerages (Lang et al. 2019).

Investment firms offer various services in their operation. Analyst coverage are firm forecasts and estimates conducted by either buy-side or sell-side analysts on behalf of their own internal unit or private investors (Cheng et al. 2003). On the one hand buy-side analysts determines how promising investments are and typically work for institutional investors such as hedge funds, mutual funds and private equity firms. Sell-side analysts on the other hand has a strong focus on issuing recommendations for investors to short, hold or long positions in securities and are employed by brokerages (Hobbs and Singh 2015).

Researches published independently in 2019, concludes that MiFID II has had both a negative and positive impact on investment research. Furthermore, research by Fang et al. (2019) demonstrates that the forecast error shrunk in the aftermath of MiFID II. While the quantity of sell-side analyst coverage had decreased, the overall quality had increased (Guo and Mota 2019). This phenomena could be explained by the increase of the transparency in investment firms' cost structure and the unbundling of research costs. Investors have a wider selection of competitors in the market, thus leading to higher pressure on analysts to give close to perfect predictions (Fang et al. 2019).

Considering the recent regulations that have been passed in the financial market, and in particular for investment firms, this thesis intend to further study what impact MiFID II and other factors have had on analyst coverage. I will, with inspiration from previous papers published by Lang et al. (2019), Fang et al. (2019), and Guo and Mota (2019), conduct my research on European Economic Area (EEA) and North American (NA) firms. The main difference with this thesis is that the data set is extended to 2019 and that the models will include more explanatory variables.

1.1 Report Structure

The remaining paper will proceed as follows. Section 1.2 will set forth the purpose and problem formulation of the thesis, while Section 1.3 presents the scope. In Section 2, I will introduce the literature associated with this field of study. The empirical framework used in this thesis will be provided for the reader in Section 3. I intend to go through the data set and methodology in Section 4. In Section 5 and 6, I lay forth my results and analysis of the results. Lastly, Section 7 and 8 will be devoted to concluding this thesis and suggesting the readers future research topics.

1.2 Purpose and Problem Formulation

The purpose of this thesis is to gain a better understanding of how MiFID II has affected investment research of firms in the EEA and NA. Only a handful of researches has been conducted in this field that dates back to the beginning of the regulation in January, 2018, emphasizing the need for further studies with an extended data set. Furthermore, this thesis intend gain understanding of how firm's financial measures affects the number of forecasts. The areas of interest that I choose to study will mainly be analyzed with the help of mathematical tools such as multiple linear regression analysis and the difference-in-differences method. Following research questions will be studied,

Research Question 1 (RQ1): *What impact has MiFID II had on analyst coverage?*

Research Question 2 (RQ2): *How does firms' financial measures affect analyst coverage?*

1.3 Scope

With inspiration from previous studies conducted by Lang et al. (2019), Fang et al. (2019), and Guo and Mota (2019), I will focus my research on Canadian, US and EEA firms. Information from buy-side firms such as hedge funds, private equity firms and mutual funds are often limited due to protection of their clients. I will thus, unlike Lang et. al., include both buy-side and sell-side analyst coverage and how they are affected by factors such as financial measures and MiFID II. Firms from Great Britain will still be included in the EEA even though the EU separation agreement was signed January 28, 2020. Since data is only provided until February, 2020, this thesis will neither be able to study the impact of the 2019-20 Coronavirus Pandemic and Brexit on analyst coverage. Lastly, this paper will consider the ongoing financial crisis that has incurred due to SARS-CoV-2 and will thus include all firms from June, 1987 to February, 2020 even if some have become insolvent.

2 Literature Review

The literature review below intend to give the reader a broader insight in the research that has been conducted on MiFID II affect on analyst coverage. Few papers have been published on this field of study, which comes in no surprise since the regulations has only been in force since early 2018. Methodologies and data sources in this thesis are mainly inspired from three pioneering papers published independently by Fang et al. (2019), Lang et al. (2019) and Guo and Mota (2019). Although all the papers used similar approaches, they differ in research questions, variables and conclusions.

2.1 MiFID II Impact on Sell-Side and Buy-Side Analyst Coverage

Two papers published independently in 2019 by Fang et al. (2019) and Lang et al. (2019) provided evidence on how different financial measures and MiFID II affected the analyst coverage. However, contrary to first mentioned authors, Lang et. al. emphasizes a heavier focus on the sell-side coverage rather than buy-side.

In response to the new MiFID II regulations, Fang et. al. conducted a study on how investment firms were affected by it. Their study used a DiD approach where firms in the EEA countries would pose as the first treatment group and firms in North America would pose as the second. Furthermore, to proceed with the DiD method they introduced a post variable which was whether MiFID II had occurred. Lastly, to execute the and analyze the DiD, they had to compute a regression model including several financial measures from the firms such as earnings per share (EPS), return on asset (ROA) and firm size.

By applying DiD, Fang et al. (2019) concluded that MiFID II had a negative impact on the sell-side analyst coverage. Since the implementation of the regulation the buy-side analyst coverage had seen a sharp increase, as an effect of the decreasing researches done by the sell-side. Furthermore, according to Fang et. al. the phenomena within the investment firms that MiFID II had caused, did indirectly force sell-side analysts to purchase information and forecasts from buy-side analyst, implying that the new regulation had set a friction in the equilibrium between buy- and sell-side research (Fang et al. 2019).

Unlike Fang et. al. that focused on explaining how sell-side and buy-side had been affected by MiFID II, Lang et. al. conducted a study on the unbundling phenomena of sell-side analyst services in the aftermath of MiFID II. The latter paper had similar methodology as the first, which was a DiD approach with two treatment groups - EEA and North America - and post variable that was whether MiFID II had been implemented. However, the part that distinguishes these two authors research approach is that they both used different explanatory variables. Lang. et. al. included explanatory and macroeconomic variables such as firm age, logged GDP per capita and GDP growth for each country, thus counteracting potential omitted variable bias in their regression models.

2.2 Unbundling of Research in Investment Firms

In similarity to Fang et al. (2019) and Lang et al. (2019), Guo and Mota (2019) from Columbia Business School employed a DiD method with same treatment/control groups and post variable to determine whether MiFID II had an impact on analyst coverage. Guo & Mota's paper distinguishes from the remaining two papers by their usage of unemployment rate and unique analysts as explanatory variables. They identify each unique analyst coverage and define the variable as "the number of firms the analyst follows in a given fiscal year" (Guo and Mota 2019). Although choosing different variables, Guo & Mota came to comparable deductions as Fang et. al. and Lang et. al. Their research led to the conclusion that MiFID II and the unbundling of research costs had improved the information quality at the expense of reduced information quantity. Furthermore, Guo & Mota's results suggest that the decreasing profitability on the sell-side would increase the incentives for buy-side players to establish their own in-house research team. Consequently, enabling a larger proportion of sell-side information production to migrate to the buy-side (Guo and Mota 2019).

3 Empirical Framework

This section intend to present the empirical framework that will be used in this thesis. The foundation of the Difference-in-Differences Method is based on Multiple Regression Analysis. Thus, the reader will firstly be introduced to Multiple Regression Analysis in Section 3.1 followed by the Difference-in-Differences approach in Section 3.2.

3.1 Multiple Linear Regression Analysis

To model the relationship between a dependent variable and explanatory variables it can be useful to apply a Multiple Linear Regression Model which is an extension of the Ordinary Least Square (OLS) Model. The Multiple Linear Regression Model includes, unlike the OLS model, several explanatory variables and has the following mathematical notation,

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon_i, \tag{1}$$

where, Y_i is the dependent variable, β_0 is the intercept, β_k are the regression coefficients, β_0 is the intercept, X_k the explanatory variables and ϵ_i the random error terms.

The Multiple Linear Regression Model can be expressed in matrix form as,

$$Y = X\beta + \epsilon, \tag{2}$$

where the vectors Y, β, ϵ and matrix X are defined as,

$$Y = \begin{vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{vmatrix}, \quad X = \begin{vmatrix} 1 & X_{11} & X_{12} & ... & X_{13} \\ 1 & X_{21} & X_{22} & ... & X_{23} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & ... & X_{nk} \end{vmatrix} \quad \beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{vmatrix}, \quad \epsilon = \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{vmatrix} \tag{3}$$

Y is a $n \times 1$ vector of observations. X is a $n \times p$ matrix of the independent variables, β is $p \times 1$ vector of regression coefficients and ϵ the random errors in an $n \times 1$ vector.

Estimation of model parameters is done with the OLS method, where the least-square estimators are obtained by minimizing the sum of squares of the errors $\epsilon'\epsilon$. The underlying assumptions of the OLS is that the relationship between the dependent and explanatory variables should be linear, and that the error terms should be follow a normal distribution, i.e. homoscedasticity and multivariate normality. Furthermore, there should exists little or no correlation between the explanatory variables. By multiplying both sides in Equation (2) with the inverse of matrix X, following least-squares normal equations are obtained,

$$X'X\widehat{\beta} = X'Y \Rightarrow \quad \widehat{\beta}_{OLS} = (X'X)^{-1}X'Y, \tag{4}$$

There are two types of explanatory variables used in a regression model - quantitative and qualitative variables. The first mentioned variables are continuous, whereas the latter mentioned are typically considered as indicator or dummy variables and take the values 0 or 1 to indicate categorical effects that can shift an outcome (Montgomery et al. 2012).

3.1.1 Multicollinearity

The presence of multicollinearity can impose uncertainty in the model by increasing the standard errors of estimated coefficients. A method to detect multicollinearity in Multiple Linear Regression models is by examining a correlation matrix of the explanatory variables. The elements (correlations) in the matrix are obtained by the unit length scaled values defined as,

$$w_{ij} = \frac{X_{ij} - \overline{X}_j}{s_{jj}^{1/2}},\ i = 1, 2, ..., n,\ j = 1, 2, ..., k \tag{5}$$

where k is the number of explanatory variables without the intercept, $\overline{X}_j$ is the mean of the explanatory variables in jth row and $s_{jj} = \sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2$. The correlation matrix is now obtained by multiplying two matrices W of the scaled values (Montgomery et al. 2012).

$$W'W = \begin{vmatrix} 1 & r_{12} & r_{13} & ... & r_{1k} \\ r_{12} & 1 & r_{23} & ... & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{1k} & r_{2k} & r_{3k} & ... & 1 \end{vmatrix} \tag{6}$$

3.1.2 Residual Diagnostics

As mentioned in section 3.1 there are a few assumptions that has to be met. Firstly, there has to exist a linear relationship between the dependent and explanatory variable. Secondly, the errors in the model should be normally distributed, $e \sim N(0, \sigma^2)$, and uncorrelated. The residuals are defined as,

$$e_i = Y_i - \hat{Y}_i,\ i = 1, 2, ..., n \tag{7}$$

where Y_i is the ith observation and $\hat{Y}_i$ is the fitted value. Plotting the residuals is a good way to examine the key assumptions underlined above. It is possible to assess whether the residuals are normally distributed and whether there are outliers or extreme values by either inspecting a QQ-plot or looking at the studentized residuals (Montgomery et al. 2012). Studentized residuals are defined as,

$$r_i = \frac{e_i}{\sqrt{MS_{res}(1 - h_{ii})}}, \ i = 1, 2, ...n \tag{8}$$

where h_{ii} is found in the ith diagonal element of the hat matrix

$$H = X(X'X)^{-1}X', \tag{9}$$

MS_{res} is the residual mean square defined as

$$MS_{res} = \frac{SS_{res}}{n - p}, \tag{10}$$

where SS_{res} is the residual sum of square defined as,

$$SS_{res} = \sum_i (y_i - f_i) = \sum_i e_i^2, \tag{11}$$

The Studentized residuals may be used in Quantile-Quantile (QQ) plots to examine if the errors are normally distributed. Furthermore, QQ-plots are sample order statistics plotted against theoretical quantiles from a standard normal distribution (Thode 2002).

3.2 The Difference-in-Differences Method

Difference-in-Differences (DiD) is a quasi-experimental method that estimates the effect of specific event, intervention or treatments such as introduction of new legislative frameworks. The technique is based on comparing the changes in outcomes over time between a population that is affected by the event (the treatment group) and a population that is unaffected by it (the control group). DiD estimates the outcome on event such as analyst coverage by comparing the average change over time in an outcome variable for a treatment group with the average change over time for a control group. Contrary to Time Series Analysis that estimates the effect on a specific population and over continuous time, DiD make use of panel data to estimate differences between treatment and control groups (Deschenes and Meng 2018).

The general definition of the Difference-in-Differences Model is a Multiple Linear Regression Model with a treatment/control group indicator variable $Treat_j$, a pre/post event indicator variable $Post_t$ and an interaction variable $Treat_j \times Post_t$. Corresponding estimate to the interaction variable with henceforth be defined as the DiD estimate. The DiD Model has the following mathematical notation,

$$Y_{i,t} = \beta_0 + \beta_1 Treat_j + \beta_2 Post_t + \beta_3 Treat_j \times Post_t + \beta_k X_{k,i,t} + \epsilon_{j,t} \tag{12}$$

Where $Y_{i,t}$ is the dependent variable of the model. β_0 is the intercept and β_3 is the DiD estimate for the interaction variable $Treat_j \times Post_t$. $Treat_j$ is the first indicator variable used in the DiD approach which represents the treatment/control group, where $j = 0$ is treatment population and $j = 1$ is control population. $Post_t$ is the second indicator variable used in the DiD interaction, where $t = 0$ is prior to an event and $t = 1$ is after an event. β_k is a vector consisting of k amount of beta coefficients for the corresponding explanatory variable, where $k = 4, 5, 6....$ $X_{k,i,t}$ is the single column vector that contains k amount of explanatory variables. $\epsilon_{j,t}$ is the error terms associated with the model. Similar assumptions of the OLS model apply to DiD. However, DiD requires a common trend (parallel trend) assumption that implies that the average change in the control group should represent the counterfactual change in the treatment group (Callaway and Li 2017; Fricke 2017; Rambachan and Roth 2019).

4 Data and Methodology

In this section I will present the data, describe the sample construction and provide the methodology of my research. The reader will be introduced to the two models that are going to be examined in 4.2.2, where the description and computation of the variables are thoroughly explained in Table 2. Data handling and statistical analysis was done with the programming software RStudio.

4.1 Sample Construction

The three data files used in this study was extracted from The Center for Research in Security Prices *CRSP/Compustat Merged*, Institutional Broker's Estimate System *(IBES) History Summary* and *IBES CRSP Link* through the Wharton Research Data Services (WRDS) over the time period June, 1987 to February, 2020.

CRSP/Compustat Merged consisted of data from countries around the world and had actual firm measures such as Earnings Before Interest and Taxes, Earnings per Share, Total Dividends, Liabilities and Employees. The IBES database, that was first systematized in 1976 and operated by Thomson Reuters, stores analyst estimates for more than 200 different types of firm measures. Forecasts include, but are not limited to Earnings Per Share, Liabilities, Net Sale and Total Assets. The *IBES History Summary* file had a total of 16 159 481 observations and 80 642 636 number of EPS estimates covering firms around the world for five fiscal years.

With the aid of the *IBES CRSP Link* file, matching was done by linking the firms' permanent identifier code PERMNO in the *CRSP/Compustat Merged* file with their respective Committee on Uniform Security Identification Procedures (CUSIP) number in the *IBES History Summary* file. After matching the IBES estimates with the firms' information in the CRSP/Compustat Merged file, there were a total of 513 EEA firms and 11 628 North American firms. Table 3 in Appendix illustrates the distribution of firms in both EEA and North American countries. The total Number of Estimates was reduced to 14 836 020 after matching the data, whereof 584 729 was for EEA firms and 14 251 291 was for NA firms. By inspecting the matched data, we can in Figure 1, observe that the number of estimates decreased drastically in the aftermath of the September 11 attacks and the introduction of both MiFID in 2007 and MiFID II in 2018. The sharp decrease of estimates after the first legislative framework MiFID could have also been caused by the financial crisis than incurred between 2007 and 2008.

Information on EPS forecast error was provided through the *IBES History Summary* file in the form of estimates that analysts had made and the actual value for a firm's fiscal year end. In this study the forecast error was computed as the quota of the absolute value of the difference of the actual value and the estimate that an analyst made on a firm i at time t, divided by the standard deviation. Figure 3 illustrates the distribution of the EPS forecast errors in a histogram, indicating that the EPS forecast error distribution resembles a standard normal distribution.

4.2 Methodology and Model Construction

4.2.1 Methodology

A similar approach will be used in this study as in the papers published by Fang et al. (2019), Lang et al. (2019) and Guo and Mota (2019). The impact of factors, and in particular MiFID II, on analyst coverage will be studied by applying a DiD approach with Multiple Linear Regression Analysis. The main difference with this study is that I intend to include more explanatory variables and have, unlike aforementioned authors, an extended data set to February, 2020. Computations and definitions of the dependent, explanatory and indicator variables are presented in Table 2. Ordinary OLS regressions will be run separately on the EEA and North American firms. In addition to how the authors in Section 2 conducted their study, I will analyze the diagnostics of the regressions models presented in the Equation below and potentially exclude or select variables that are more suitable.

4.2.2 Empirical Models

In the DiD approach firms in EEA will be used as the treatment group since MiFID II primarily affects the countries in the EU, while firms in Canada and The US will be used as the control group. Considering that the research is based on assessing the effects between two different timelines, pre-MiFID II will be coded as the pre-event, whereas post-MiFID II will be coded as post-event. The empirical models studied in this thesis will follow Equation (12) and are presented below.

Number of Estimates

$$NumEst_{i,t} = \beta_0 + \beta_1 Treat_j + \beta_2 Post_t + \beta_3 Treat_j \times Post_t + \beta_k X_{k,i,t} + \epsilon_{j,t} \quad (13)$$

$NumEst_{i,t}$ is the analyst coverage (number of estimates) for a firm i at time t. $Treat_j$ is the first indicator variable used in the DiD approach which represents the treatment/control group, where $j = 0$ is EEA (the treatment group) and $j = 1$ is North America (the control group). $Post_t$ is the second indicator variable used in the DiD interaction, where $t = 0$ is pre-MiFID II and $t = 1$ is post-MiFID II. $X_{k,i,t}$ is the single column vector that contains k amount of explanatory variables such as a firm's Total Dividends over a Year, Earnings Before Interest and Taxes, Earnings Per Share, Loss, Liabilities, Return on Asset, Return on Equity and Firm Size. $\epsilon_{j,t}$ is the error terms associated with the model.

Forecast Error

$$ForErr_{i,t} = \beta_0 + \beta_1 Treat_j + \beta_2 Post_t + \beta_3 Treat_j \times Post_t + \beta_k X_{k,i,t} + \epsilon_{j,t} \quad (14)$$

$ForErr_{i,t}$ is the forecast error for a firm i at time t. Remaining variables are similar as in Equation 13.

5 Results

The results of my studies will be laid forth in two parts. Section 5.1 and 5.2 presents the results associated with the DiD approach on analyst coverage and forecast error respectively. Before proposing the DiD models *NumEst (Merged)* and *ForErr (Merged)*, the regressions models for EEA and NA firms will be introduced. The parallel trend assumption was considered to be met for both the DiD models as seen in Figure 2. Outliers and extremes was removed from all models prior to running the regressions. In particular, all forecast errors past 100 was removed from the model as can be illustrated by the cutoff on the residual vs. fitted and studentized residual vs. fitted plots in Figure 5 and 6. Furthermore, neither residuals for the *NumEst (Merged)* and *ForErr (Merged)* are normally distributed due to their heavy left and right tails as indicated in Figure 7. Albeit a violation to the normality assumption of the Multiple Linear Regression model, data was not chosen to be transformed nor further reduced of outliers since it would remove important information.

5.1 Number of Estimates

Studying the treatment and control group separately, we can in Table 4 detect that all variables except for Return on Equity are statistically significant for number of estimates on NA Firms. Analyst coverage in EEA are affected by coefficients such as whether a firm had made profits (Loss), the firm size (Size), Liabilities and Return on Assets. A noteworthy observation is the difference of coefficient signs and values on the Loss variable, a larger loss for EEA firms implied an increase of analyst coverage, whereas a larger loss for NA firms decreased the overall analyst coverage. Lastly, there is a large difference in the magnitude of the Size variable. Larger firms in the EEA tended to increase the analyst coverage at a bigger scale than in the NA (coefficient value 0.58 against 0.01).

The *NumEst (Merged)* DiD Model in Table 4 illustrates that the interaction coefficient is significant. A negative sign on the DiD coefficient indicates that the analyst coverage decreased sharply when MiFID II was implemented and specifically for the NA firms, which coincides well with the plot showing showing Number of Estimates over time in Figure 1. Furthermore, the magnitude of the $Post_t \times Treat_j$ variable implies a decrease of approximately 0.52 in number of estimates after the initiation of the legislative framework. All variables except for Loss, Dividends and Return on Equity were significant. Analyst coverage seemed to be unaffected by whether firm's reported profit or the amount of dividends announced to be distributed. Large and profitable firms tended to increase the Number of Estimates significantly as seen on the Firm Size, EBITDA and EPS coefficients. In particular, each thousand dollars of Earnings per Share increased the number of estimates for a firm by 2.38, whereas each thousand dollars of Earnings Before Interest and Taxes increased the Number of Estimates by 0.02. Liability on the other hand decreased a firm's Number of Estimates by 0.01. Lastly, The RoA coefficient were -0.32, implying a sharp decrease of firm estimates the higher Return on Asset are for a firm.

Neither Loss, Dividends and Return on Equity demonstrated any significance in the *NumEst (Merged)* DiD Model. As illustrated in Figure 4, the Dividends variable was highly correlated with both the EBITDA and Liabilities variable and was thus posing a large threat to the model. High collinearity implies that a small change in Dividends would change the EBITDA and Libabilities coefficient drastically, which could decrease the reliability of the model. As a consequence of reducing the model's predictive power, Loss, Dividends and Return on Equity was excluded from the final *NumEst (Merged)* DiD Model presented in Table 6. After the removal of redundant variables, the *NumEst (Final)* model demonstrated a smaller standard error in the intercept, Firm Size and EPS coefficients in contrast to the *NumEst (Merged)* model. As with the other models that are presented, the R^2 value remained low (22% for the final model).

5.2 Forecast Error

The regression models for EPS forecast error saw large increases in the Loss and Size variable. For the EEA firms we can in Table 5 observe that forecast error increased by 7.48 if it was a loss-making firm, whereas NA firms only saw an increase of 0.11. Firm Size shows an increase of 4.01 in EEA EPS forecast error in comparison to the NA firms coefficient 2.82.

Looking at the *ForErr (Merged)* model, the DiD coefficient in Table 5 demonstrates with the highest significance level that forecast errors increased after the implementation of MiFID II. In particular, the passing of MiFID II affected the forecast error by 2.49 in average for all firms. Furthermore, forecast error had large increases depending on whether a firm was small or large. Unlike the *NumEst (Merged)* model, the dividends coefficient was significant and increased the forecast error with 1.75 for each thousand dollars of dividends a company distributed to its shareholders. Albeit the low coefficient values, the remaining coefficients were significant to the *ForErr (Merged)* model.

In similarity to the *NumEst (Merged)* model, the dividend variable was highly correlated with EBITDA and Liabilities and was thus removed in the final forecast error model. Since the Return on Equity variable was significant to all *ForErr* models, and not correlated to any variables in the data set, it was chosen to be kept in the final model.

6 Analysis

In this section I will analyze the results and answer my researh questions in Section 1.2. Analogously to the order of how the results were presented, I will firstly discuss the impact of MiFID II on analyst coverage and secondly how other factors have affected analyst coverage.

6.1 Impact of MiFID II on Analyst Coverage

Results presented in Table 4 indicates that the quantity of Number of Estimates was negatively affected. Consistent with the previous studies conducted by Fang et al. (2019), Lang et al. (2019), and Guo and Mota (2019), the DiD estimate for the *NumEst (Merged)* model implies that the control group (NA firms) saw a significant decrease in Number of Estimates after the implementation of MiFID II. As research cost no longer can be bundled with other services that investment firms provide, profitability had shrunk considerably. Consequently, these analysts and brokerages can no longer provide as much coverage as they did prior to MiFID II, leading to a significant decrease of estimates since 2018. Surprisingly, the DiD estimate for the *ForErr (Merged)* model in Table 5 demonstrates a increase of forecast error in the aftermath of MiFID II, which contradicts the results of previous authors. The increase in forecast error could be due to a lag period after MiFID II. Previously published papers only had data to late 2018 or early 2019, which could have been the period where investment firms were as most cautious on their recommendations.

6.2 Measures Affecting Analyst Coverage

According to all the models presented in Table 4, performance dictates how many estimates firms receive. The most notable difference for these models is that analyst coverage is highly dependent on whether a firm made profit (Loss) in EEA and NA firms separately, while it in the merged data set was insignificant. Furthermore, there was a considerable difference in the sign and magnitude of the loss variable for both EEA and NA firms in their respective regression models *NumEst (EEA)* and *NumEst (NA)*. The different significance level of the *NumEst (Merged)* could have been affected by the large deviation of observations between EEA and NA firms. NA firms made up for approximately 95% of the observations for both the $NumEst$ and $ForErr$ DiD models. Furthermore, The positive coefficient 0.31 for the Loss variable in the *NumEst (EEA)* indicates that analysts tended to cover more estimates over firms that goes into deficit rather than firms that profits. On the one hand, it may seem unlikely that analysts would research a firms with losses, while it on the other hand is plausible since there are other factors such as firms' plans for expansion and future dividends that can not be accounted for statistically. Considering that the Loss variable lacked statistical significance in the *NumEst (Merged)* DiD model, it was considered as redundant and discarded without any further remarks.

Inspecting the *NumEst (Merged)* model the size of a firm (Firm Size) had a significant and increasing impact on number of estimates. This could be explained by the EBITDA and EPS coefficients that are highly dependent on a firm's performance. The larger Earnings a firm accumulates, the higher Earnings per Share will become. Thus, the Firm Size coefficient is to an extent correlated to the Earnings of a firm. Albeit a low value, there is a certain correlation between the Firm Size and Earnings variables as indicated in Figure 4. The Liabilities and RoA coefficients had a significant and negative impact on the Number of Estimates. On the one hand, it is anticipated that firms with high liability should decrease their attractiveness for analysts and investors. On the other hand, RoA is unanticipated since a higher return on asset should entice more estimates for the firm. After all, RoA is an indicator on how efficient a firm's management are at using their assets to generate profit. There is no clear explanation to why RoA impact the number of estimates negatively, However, a postulation may be that analysts find loss-making firms more interesting than profit making. This sounds as a contradiction to all previous mentions that analysts aim is to increase the return of their investor, but there might be other incitement to hold positions in firms with low Return on Asset. A firm may have acquired a large amount of assets during a certain fiscal year when they perform weakly, only to invest on future projects.

Both final models demonstrated a lower standard error which could be a better fit due to the removal of variables with high collinearity. The fit of *NumEst (Final)* and *ForErr (Final)* had low variability at 22% respectively 8%. However, this does not pose a problem since the study conducted in this thesis is not on a natural phenomena, but rather the interaction in the market. Furthermore, the Forecast Error saw a decrease in the DiD coefficient after the removal of the Dividends variable implying that the model could have had a better predictive power with the introduction of other information.

7 Conclusions

The DiD regressions that was implemented in this study confirms the hypothesis, results and conclusions that Fang et al. (2019), Lang et al. (2019) and Guo and Mota (2019) laid forth in their study. With an extended data set up to February, 2020, I was able to confirm that MiFID II had a negative impact on the quantity of Number of Estimates. However, my results deviates from previous authors as EPS forecast errors increases rather than decreases after the implementation of MiFID II. In summary, the models demonstrates that the coverage of firms are highly dependent of a solid performance and well grounded balance sheet. Investment firms are dependent of their investors and give recommendations based on which companies that are deemed most profitable. A particular interesting observation is that analysts are indifferent between forecasting firms that reports negative Return on Asset and firms that reports solid Return on Asset. This phenomena could be explained by analysts' enticement of given recommendation on firms in the growth phase. Low Return of Asset can typically be a sign of a firm that are in it early stages of development.

8 Suggestions for Future Research

This study was conducted from late January to mid-May and aimed to further solidify the theories of Fang et al. (2019), Lang et al. (2019) and Guo and Mota (2019) that MiFID II decreased the number of analyst coverage while the quality of the estimates had increased. Since WRDS only update their database quarterly, there were only a handful of data points in 2020. Consequently, there was too few data points to conduct a study on how the 2019-20 Coronavirus Pandemic and Brexit had affected analyst coverage. A suggestion for future research could thus be to examine how events such as global pandemics and a specific country's withdrawal from a trading pact affects the research and profitability in investment firms. As an addition, future studies should include more or different variables such as firms' stock performances and macro variables to complement the empirical models presented in earlier papers and in this thesis.

Bibliography

Callaway, B. and Li, T. 2017. "Quantile Treatment Effects in Difference in Differences Models with Panel Data." *DETU Working Paper 12-01.*

Cheng, Y., Liu, M. H., and Qian, J. 2003. "Buy-Side Analysts, Sell-Side Analysts and Fund Performance: Theory and Evidence." *13th Annual Utah Winter Finance Conference.*

Deschenes, O. and Meng, K. C. 2018. "Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges." *Handbook of Environmental Economics 1st Edition* Volume 4.

Fang, B., Hope, O-K., Huang, Z., and Moldovan, R. 2019. "The Effects of MiFID II on Sell-Side Analysts, Buy-Side Analysts, and Firms." *Rotman School of Management Working Paper No. 3422155.*

Fricke, H. 2017. "Identification based on Difference-in-Differences Approaches with Multiple Treatments." *Oxford Bulletin of Economics and Statistics.*

Guo, Y. and Mota, L. 2019. "Should Information be Sold Separately? Evidence from MiFID II." *Columbia Business School.*

Harford, J., Jiang, F., Wang, R., and Xie, F. 2018. "Analyst career concerns, effort allocation, and firms' information environment." *Review of Financial Studies* 32(6):2179–2224.

Hobbs, J. and Singh, V. 2015. "A comparison of buy-side and sell-side analysts." *Review of Financial Economics* Volume 24:42–51.

Hong, H. and Kacperczyk, M. 2010. "Competition and bias." *The Quarterly Journal of Economics* 125(4):1683–1725.

Hong, H. and Kubik, J. D. 2003. "Analyzing the analysts: Career concerns and biased earnings forecasts." *The Journal of Finance* 58(1):313–351.

Karmaziene, E. 2019. "Incentives of financial analysts: trading turnover and compensation." *Swedish house of Finance Research Paper* 16-15.

Lang, M., Pinto, J., and Sul, E. 2019. "MiFID II unbundling and sell side analyst research." *Conference: 2019 Washington Area Research Symposium.*

Montgomery, D. C., Peack, E. A., and Vining, G. G. 2012. *Introduction to Linear Regression Analysis.* Wiley.

Rambachan, A. and Roth, J. 2019. "An Honest Approach to Parallel Trends." *Harvard University.*

The European Union. 2014. "Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU Text with EEA relevance." *Official Journal of the European Union.*

Thode, H. C. 2002. "Testing for normality." *Statistics: textbooks and monographs* Vol. 164.

Appendix

Appendix A. Acronym List

Table 1: Vocabulary and Acronyms associated with this Study.

Vocabulary	**Acronym**
Difference-in-Differences	DiD
Earnings Per Share	EPS
Earnings Before Interest and Taxes	EBITDA
European Economic Area	EEA
European Union	EU
Forecast Error	ForErr
Markets in Financial Instruments Directive	MiFID
North America	NA
Number of Estimates	NumEst
Return on Asset	RoA
Return on Equity	RoE

Appendix B. Variable Description

Table 2: Description of Variables used in this Study

Code	Definition and source
Variables of Interest	
$Post_t$	The $Post_t$ variable is an indicator variable that is 0 for pre-MiFID II and 1 for post-MiFID II. Retrieved from IBES
$Treat_j$	The treatment/control group variable is an indicator variable used for the DiD approach, where $j = 0, 1$. $j = 0$ stands for EEA firm and $j = 1$ stands for North American firm. Retrieved from CRSP/Compustat Merged
Dependent Variables	
$NumEst_{i,t}$	Number of Estimates are the number of forecasts made on a particular firm i. Number of Estimates are often referred to as Analyst Coverage, Forecast or Estimate in the text. Analysts typically forecasts EPS. Retrieved from IBES
$ForErr_{i,t}$	Forecast Error are the quota of the absolute value of the difference of the actual value and the mean estimate that an analyst made on a firm i at time t, divided by the standard deviation. Computed as $ForErr = \frac{\lvert Actual-Estimate \rvert}{\lvert StandardDeviation \rvert}$. Analysts typically forecasts EPS. Retrieved from IBES
Explanatory Variables	
$Div_{i,t}$	Total Dividend Over a year t for firm i. Retrieved from CRSP/Compustat Merged
$EBITDA_{i,t}$	Earnings Before Interest, Taxes, Depreciation and Amortization year t for firm i. Retrieved from CRSP/Compustat Merged

Code	Definition and source
$EPS_{i,t}$	Earnings Per Share year t for firm i. Retrieved from CRSP/Compustat Merged
$LB_{i,t}$	Liabilities year t for firm i. Retrieved from CRSP/-Compustat Merged
$RoA_{i,t}$	Return on Assets year t for firm i. Computed as $RoA = \frac{NetIncome}{TotalAssets}$. Retrieved from CRSP/Compustat Merged
$RoE_{i,t}$	Return on Equity year t for firm i. Computed as $RoE = \frac{NetIncome}{Stockholder'sEquity}$. Retrieved from CRSP/-Compustat Merged
$Loss_{i,t}$	Indicator variable that indicates whether a firm i during time t made a loss. If $NetIncome_{i,t} < 0$, then $Loss_{i,t} = 1$ else, $Loss_{i,t} = 0$. Retrieved from CRSP/-Compustat Merged
$Size_{i,t}$	Indicator variable that indicates the size of firm i at time t. The two categories are small/mid enterprises (less than 1000 employees) and large enterprises (more than or equal to 1000 employees). The Size variable will also be called Firm Size in the text. If $Employees_{i,t} < 1000$, then $Size_{i,t} = 0$ else, $Size_{i,t} = 1$. Retrieved from CRSP/Compustat Merged

Appendix C. Descriptive

Table 3: Distribution of firms in the EEA and NA. The table only present firms that matched between the CRSP/Compustat Merged and IBES data files.

Country ISO Code	Country	Frequency	Percent
AUT	Austria	1	0.01%
BEL	Belgium	10	0.08%
CAN	Canada	168	1.38%
CHE	Switzerland	30	0.25%
CYP	Cyprus	2	0.02%
DEU	Germany	30	0.25%
DNK	Denmark	8	0.07%
ESP	Spain	10	0.08%
FIN	Finland	7	0.06%
FRA	France	38	0.31%
GBR	Great Britain	175	1.44%
GRC	Greece	31	0.26%
HUN	Hungary	2	0.02%
IRL	Ireland	50	0.41%
ISL	Iceland	1	0.01%
ITA	Italy	15	0.12%
LUX	Luxembourg	20	0.16%
NLD	Netherlands	58	0.48%
NOR	Norway	10	0.08%
PRT	Portugal	2	0.02%
SWE	Sweden	13	0.11%
USA	United States of America	11 460	94.39%
Total		**12 141**	**100.00%**

Figure 1: Illustrates EEA, NA and the Total Number of Estimates in thousands. The graph only includes estimates from IBES that are matched with CRSP/Compustat Merged file. Firms outside EEA, USA and Canada are excluded. The two vertical lines in 2007 and 2018 marks the introduction of MiFID and MiFID II respectively.

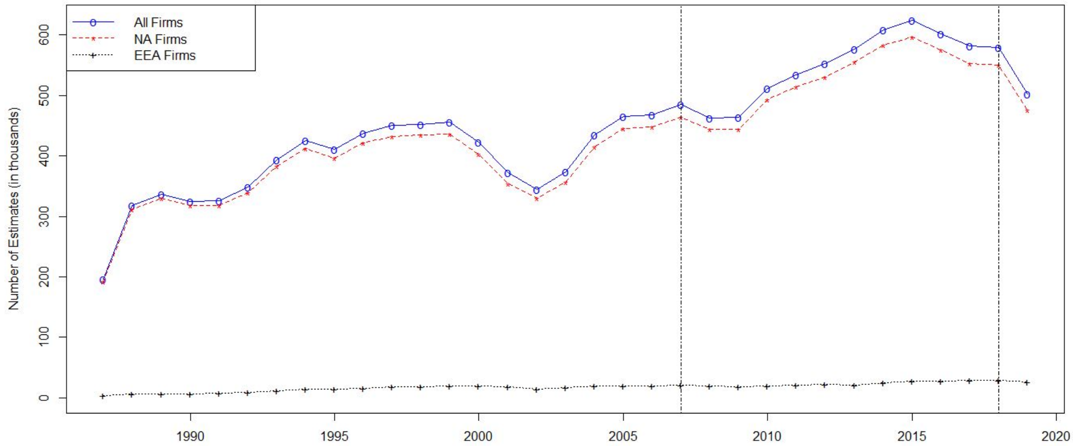
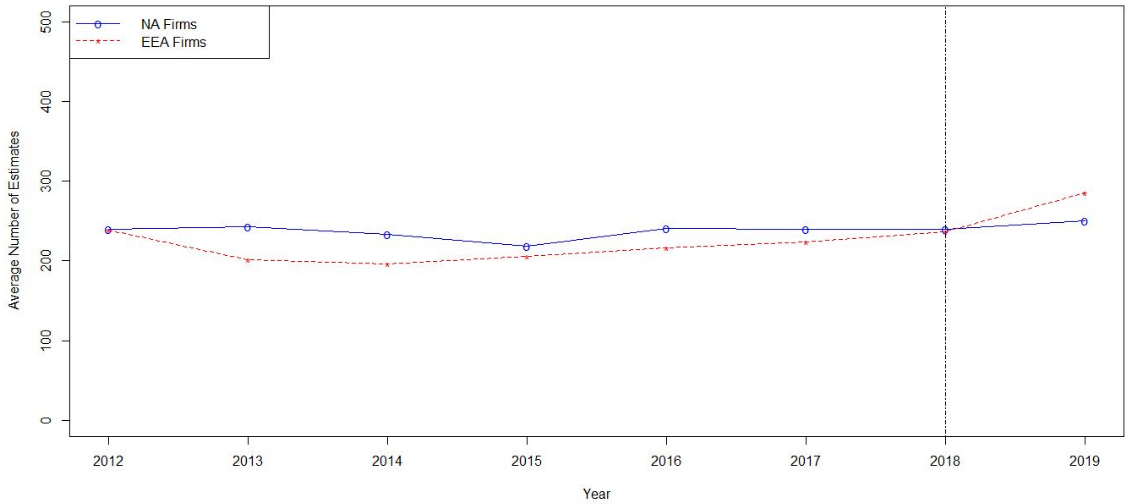


Figure 2: Illustrates the unweighted average yearly Number of Estimates and Forecast Error for EEA and NA firms from 2012 to 2019. In similarity to Figure 1, the graphs only includes observations from IBES that are matched with CRSP/Compustat Merged file.

(a) Average Number of Estimates



(b) Average Forecast Error

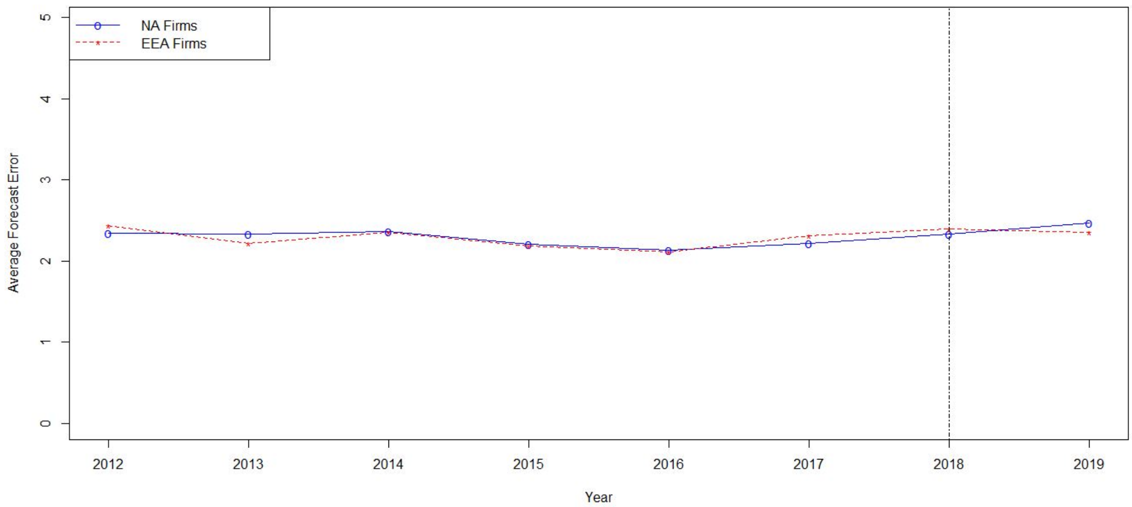
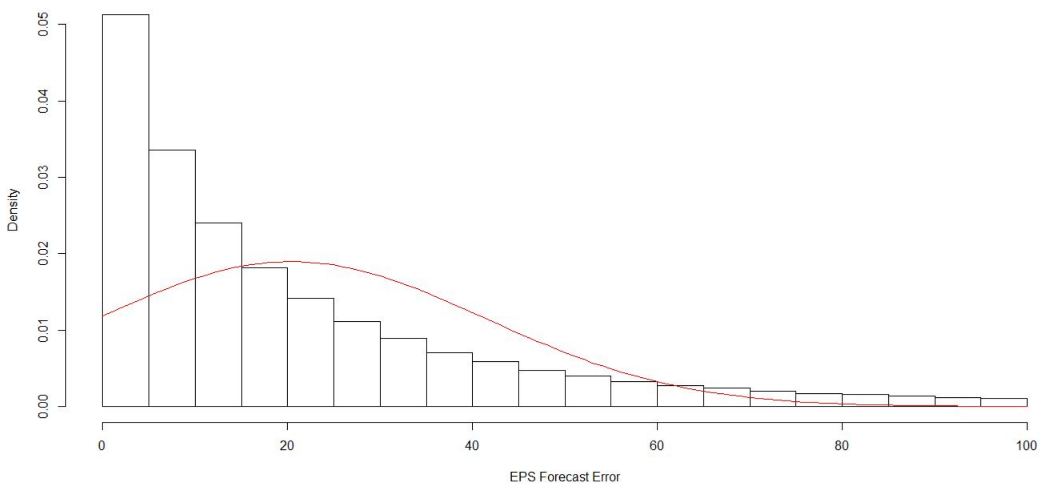


Figure 3: Histogram over EPS Forecast Error. Since Forecast Error is computed as an absolute value, there are no negative values in the illustration.



Appendix D. Summary Statistics

Table 4: Summary Statistics of the Initial Ordinary Least Squares Regression Models with Analyst Coverage (*NumEst*) as dependent variable. *NumEst (Merged)* is the DiD model with both EEA and NA firms, whereas *NumEst (EEA)* and *NumEst (NA)* are OLS models for EEA and NA firms respectively. Presented models are computed with the *lm()* function in Rstudio. Parenthesis is the Standard Error. '***' p-value < 0.001. '**' p-value < 0.01. '*' p-value < 0.05.

Coefficients	NumEst (Merged)	NumEst (EEA)	NumEst (NA)
(Intercept)	0.30	0.01	0.10
	(3.05)	(7.74)	(1.23)
$Treat_j$	0.81***		
	(2.87)		
$Post_t$	0.01***		
	(0.14)		
$Post_t \times Treat_j$	**-0.52*** **		
	(0.12)		
Loss	-0.11	0.31***	-4.22**
	(1.44)	(8.86)	(1.43)
Firm Size	0.01***	0.58***	0.01***
	(1.19)	(7.91)	(1.18)
Dividends	-0.01	-0.01	0.01***
	(0.01)	(0.01)	(0.01)
EBITDA	0.02***	0.01	0.03***
	(0.01)	(0.01)	(0.01)
EPS	2.38***	0.88	2.29***
	(0.15)	(0.49)	(0.15)
Liabilities	-0.01***	-0.01***	-0.01***
	(0.01)	(0.01)	(0.01)
RoA	-0.32***	-0.15**	-0.32***
	(0.65)	(5.62)	(0.64)
RoE	-0.01	-0.26	-0.01
	(0.01)	(0.16)	(0.01)
N	91 657	3664	87 987
R^2	0.22	0.04	0.26

Table 5: Summary Statistics of the Initial Ordinary Least Squares Regression Models with Forecast Error in percentage (*ForErr*) as dependent variable. *ForErr (Merged)* is the DiD model with both EEA and NA firms, whereas *ForErr (EEA)* and *ForErr (NA)* are OLS models for EEA and NA firms respectively. Presented models are computed with the *lm()* function in Rstudio. Parenthesis is the Standard Error. '***' p-value < 0.001. '**' p-value < 0.01. '*' p-value < 0.05

Coefficients	ForErr (Merged)	ForErr (EEA)	ForErr (NA)
(Intercept)	4.12	6.74	9.01
	(0.09)	(0.20)	(0.01)
$Treat_j$	4.80***		
	(0.08)		
$Post_t$	1.86***		
	(0.46)		
$Post_t \times Treat_j$	**2.49*** **		
	(0.44)		
Loss	0.11***	7.48***	0.11***
	(0.05)	(0.24)	(0.05)
Firm Size	2.94***	4.01***	2.82***
	(0.04)	(0.20)	(0.04)
Dividends	1.75***	0.01***	0.01***
	(0.01)	(0.01)	(0.01)
EBITDA	-0.01***	-0.01***	-0.01***
	(0.01)	(0.01)	(0.01)
EPS	0.51***	0.52***	0.52***
	(0.01)	(0.02)	(0.01)
Liabilities	-0.01**	0.01**	0.01***
	(0.01)	(0.01)	(0.01)
RoA	0.45***	1.02***	0.45***
	(0.02)	(0.15)	(0.02)
RoE	-0.01***	-0.01***	-0.01***
	(0.01)	(0.01)	(0.01)
N	1 584 868	66 445	1 518 417
R^2	0.08	0.07	0.08

Table 6: Summary Statistics of the Final Ordinary Least Squares Regression Models with Analyst Coverage (*NumEst*) and Forecast Error (*ForErr*) as dependent variables. No variables were discarded from the *ForErr* Model. Parenthesis is the Standard Error. '***' p-value < 0.001. '**' p-value < 0.01. '*' p-value < 0.05

Coefficients	NumEst (Final)	ForErr (Final)
(Intercept)	0.29	4.12
	(2.94)	(0.09)
$Treat_j$	0.81***	4.69***
	(2.87)	(0.08)
$Post_t$	0.01***	2.22***
	(0.10)	(0.46)
$Post_t \times Treat_j$	**-0.52*** **	**2.20*** **
	(0.12)	(0.44)
Loss		0.11***
		(0.05)
Firm Size	0.01***	2.92***
	(1.14)	(0.04)
Dividends		
EBITDA	0.02***	-0.01***
	(0.01)	(0.01)
EPS	2.83***	0.50***
	(0.14)	(0.01)
Liabilities	-0.01***	-0.01**
	(0.01)	(0.01)
RoA	-0.32***	0.44***
	(0.65)	(0.02)
RoE		-0.01***
		(0.01)
N	91 611	1 584 869
R^2	0.22	0.08

Table 7: Difference-in-Differences method. β_3 is the DiD estimator presented in Table 4 and 5. The two lower tables are presented with numeric.

	Treatment Group	**Control Group**	**Difference**
Pre-MiFID II	$\beta_0 + \beta_1$	β_0	β_1
Post-MiFID	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
Difference	$\beta_2 + \beta_3$	β_2	β_3

Numest	**EEA**	**NA**	**Difference**
Pre-MiFID II	1.11	0.30	0.81
Post-MiFID	0.60	0.31	0.29
Difference	-0.51	0.01	-0.52

ForErr	**EEA**	**NA**	**Difference**
Pre-MiFID II	8.92	4.12	4.80
Post-MiFID	13.27	5.98	7.29
Difference	4.35	1.86	2.49

Appendix E. Diagnostics

Figure 4: Correlation plot for the variables included in the Numbest of Estimates and Forecast Error data set. The variable Return on Assets (RoE) is removed since there were no correlation with the other variables.

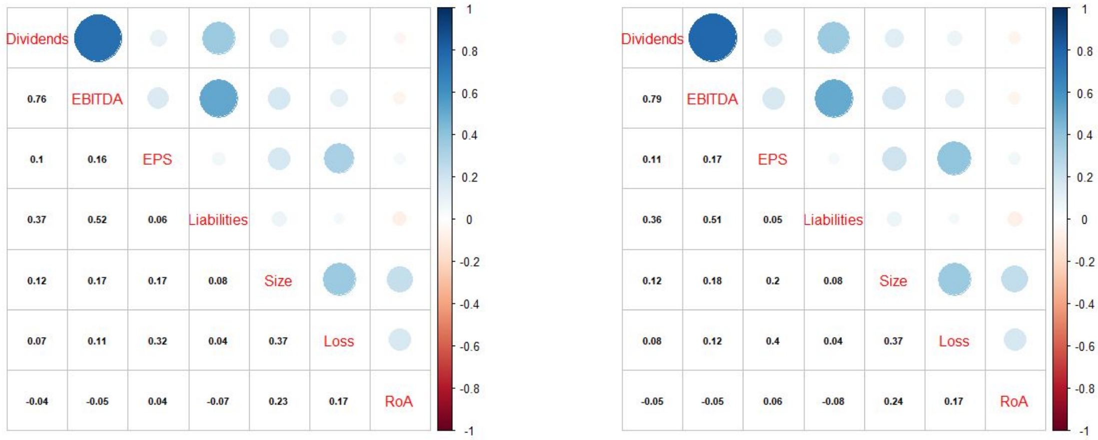
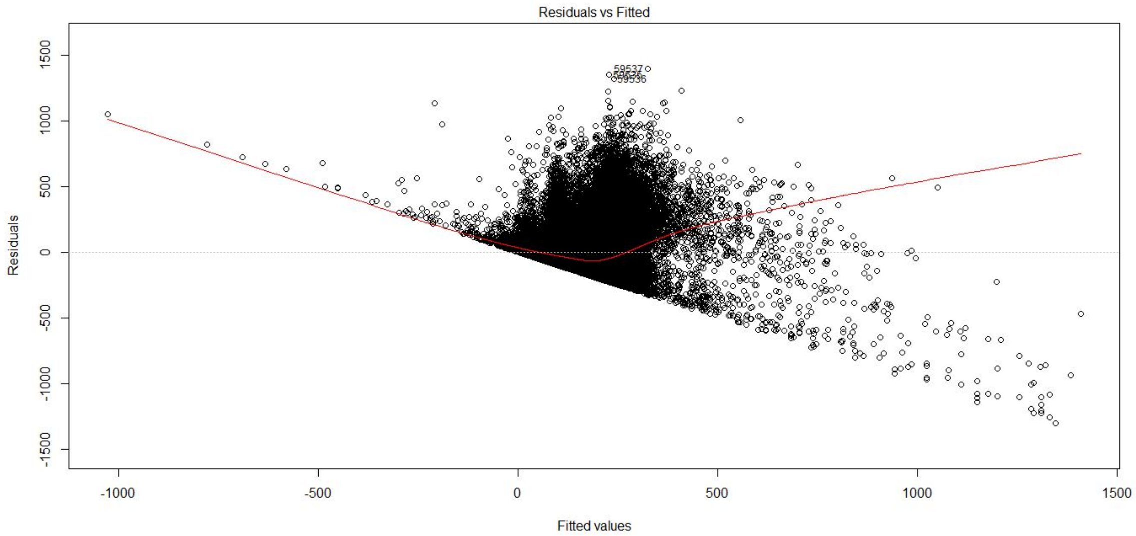


Figure 5: Illustrates Residuals vs Fitted Values for the *NumEst (Merged)* and *ForErr (Merged)* Model. A cutoff can be seen for Figure 6.b due to removal of all forecast errors above 100.

(a) Residuals vs Fitted values for the *NumEst (Merged) Model*



(b) Residuals vs Fitted values for the *ForErr (Merged) Model*

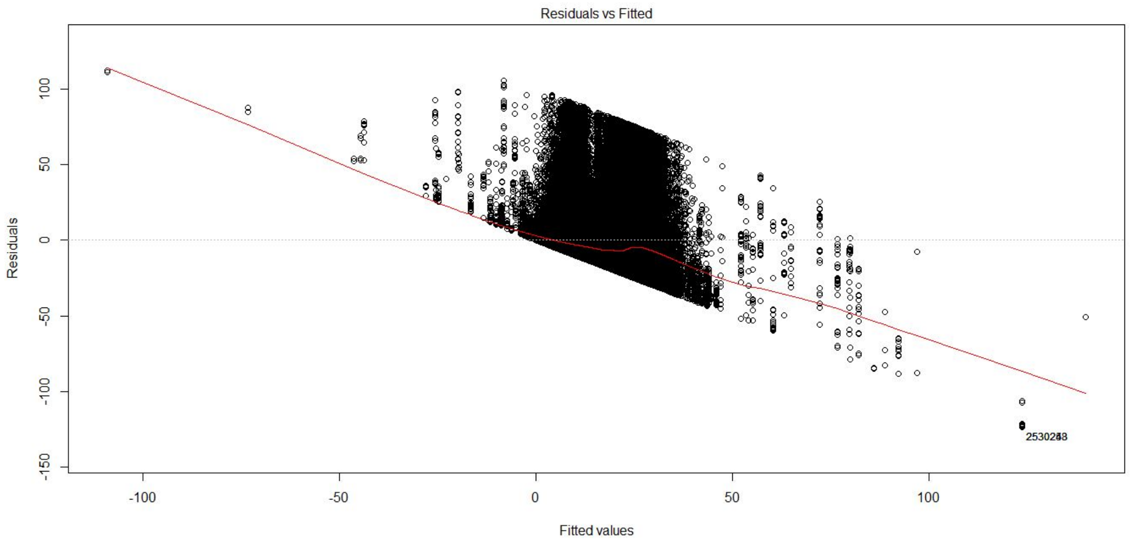
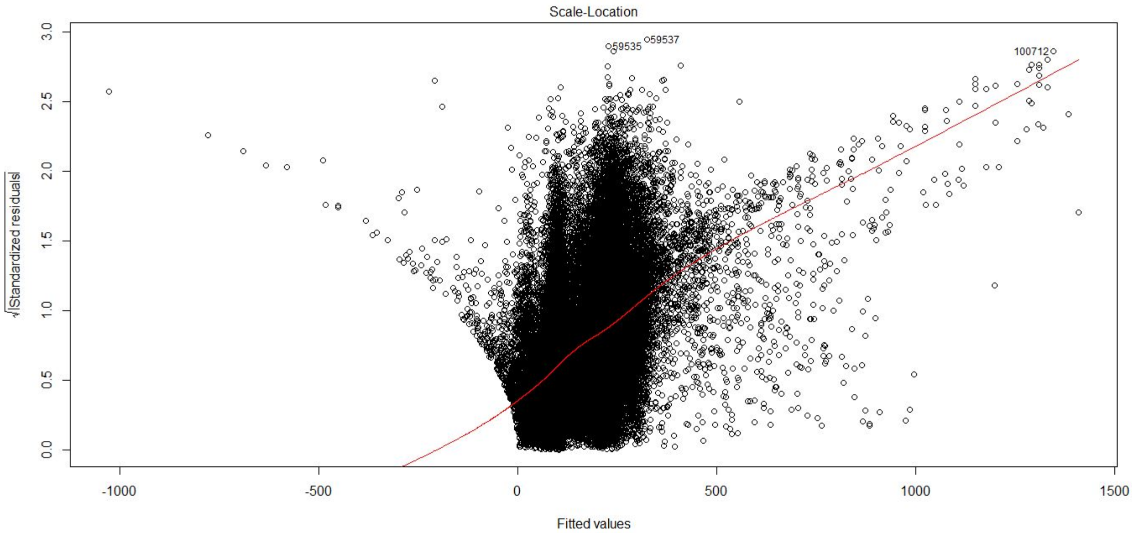


Figure 6: Illustrates Studentized Residuals vs Fitted Values for the *NumEst (Merged)* and *ForErr (Merged)* Model. A cutoff can be seen for Figure 7.b due to removal of all forecast errors above 100.

(a) Studentized Residuals vs Fitted values for the *NumEst (Merged) Model*



(b) Studentized Residuals vs Fitted values for the *ForErr (Merged) Model*

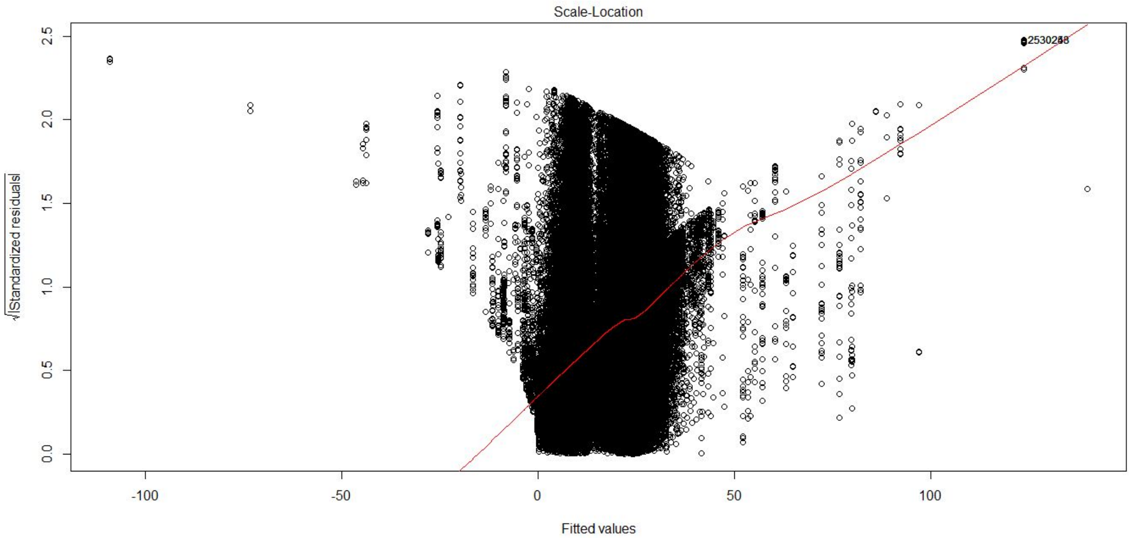
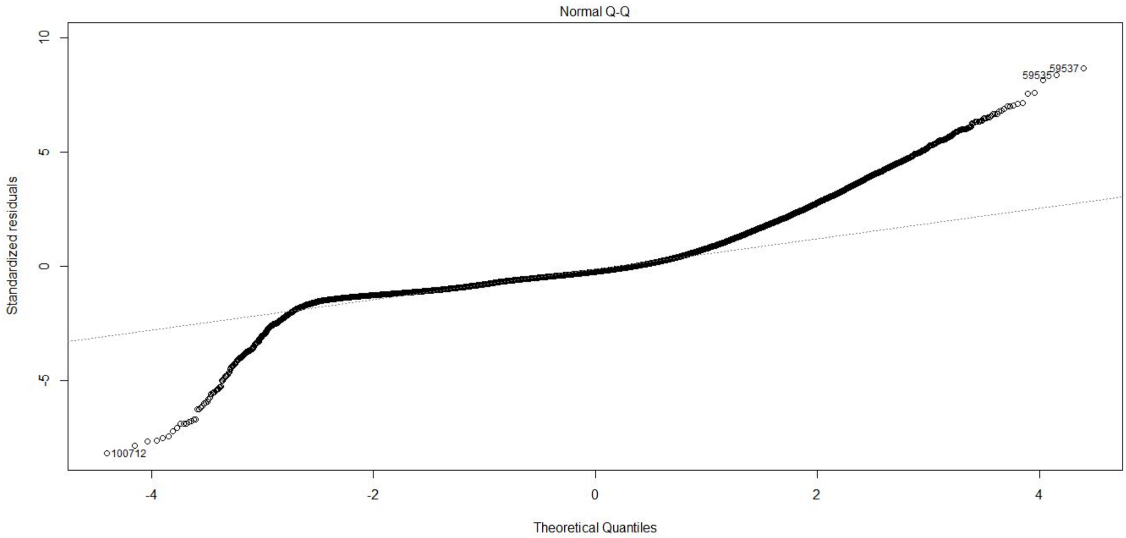


Figure 7: Illustrates QQ-plot for the *NumEst (Merged)* and *ForErr (Merged)* Model. None of the two dependent variables resembles a normal distribution. Neither models' residuals had resemblance of a normal distribution.

(a) QQ-plot for NumEst (Merged) Model



(b) QQ-plot for ForErr (Merged) Model