Master of Finance - Master thesis

Empirical analysis of hydrogen investments on market value.

AALBORG UNIVERSITET

Autor:
Bruno Magalhães
Petr Sigmund

Department of Business and Management
Faculty of Social Science
Aalborg University

May 28, 2023

Department of Business and Management
Aalborg University
http://www.aau.dk

Title:
Empirical analysis of hydrogen investments on market value.

Theme:
Valuation of companies based on Hydrogen investments

Project Period:
Spring Semester 2023

Participant(s):
Bruno Magalhães
Petr Sigmund

Supervisor(s):
Cesário Mateus

Page Numbers: 54

Date of Completion:
May 28, 2023

Abstract

In recent years modern society has been seen pushing further towards a better environment by, among others, developing green energy solutions. With the objective to bring down global warming and greenhouse gas emissions we have seen initiatives like the Paris Agreement being made. One promising solution could be Hydrogen, research and development has increased in recent years as stated by the IEA, and since this energy source can be both used and stored for energy generation investments related to hydrogen have increased around the world. This thesis aims to investigate if factual investments in this technology have real impacts on company value in the market. Utilizing panel data and applying random effects difference-in-difference estimation methods, this paper wants to verify the relationship between company market value and hydrogen investments considering the Paris agreement has the event that set in motion a bigger effort in the green transition.

The companies analyzed are all located in the USA and belong to the energy industry, treatment and control groups are divided in an even number and selection for treatment group was made according to the existence of hydrogen projects or investments made. Using financial data from 2010 to the end 2021 we apply statistical methods to evaluate the impact of any such events on company value after 2016.

After analysis, results suggest that hydrogen related investments in fact have a negative impact on company value for companies in the energy and fuel industry. Using a DiD model shows that firms that implemented or realized investments in green hydrogen technology experience an decrease in their respective market values when comparing to other peers that did not factually implement the same tier of projects and investments. This also pointed out that there is a negative effect on market value for companies that are more innovative and have a higher level of technological capabilities.

In summary, this thesis suggests that there could be important implications for firms in the energy industry. Since the results suggest that hydrogen related investments and projects have a negative impact on company value, it could be an important consideration for companies pondering about investing in more advanced and innovative technology.

Preface

Aalborg University,
May 28, 2023

Petr Sigmund & Bruno Magalhães
psigmu21@student.aau.dk
bteixe21@student.aau.dk

Contents

List of Figures

List of Tables

1 Introduction

As it is known global efforts to contradict climate change have grown in the past years, yet there are several challenges when considering the transition to cleaner and sustainable energy sources. There is a necessity to have a diversified energy mix everywhere, as this provides more stability and less risk but there are not a lot of choices for this because most technologies are immature. From this pool of new energy production methods hydrogen is considered to be a promising solution for the energy transition issue, it is a versatile energy source that can be used in several industries, like power generation, transportation and manufacturing industries. Given this potential many companies belonging to the energy and fuel industry have started looking in the direction of hydrogen as a form to accommodate the growing demand for clean energy

This thesis intends to investigate the impact of factual hydrogen investments on the company value of firms on the energy and fuel industry situated in the USA. More specifically, if companies that effectively directed funds and resources towards hydrogen investments have seen an increase in their market value compared to other firms that remained idle in this niche.

To arrive at our research question: "Do hydrogen investments cause company market value to increase in comparison with it's peers?". Financial data from 18 companies belonging to the USA energy sector will be used and will form a balanced panel data, data set, to which we will apply a random effects difference-in-difference (DiD) estimator and other factors. This will allow us to analyze the presence of hydrogen investments and relate them to the Paris Agreement.

Using the Paris Agreement as the main driver for the clean energy shift by corporations is particularly relevant as this document sets a clear target for countries to achieve regarding carbon emission, thus incentivizing companies to invest in clear energy sources like hydrogen. After this international legally binding treaty was signed, the exclusively economical decision to be part of the niche hydrogen market became also backed by political, environmental and legal reasons. By using the random effects DiD statistical model the unobserved heterogeneity can be controlled for all our different companies. It is important to mention that other major political and economical events that affected the industry sector that is discussed in this paper, events related to geopolitical tensions, presidential elections, OPEC regulations, volatility in oil prices and covid will be considered when analyzing and drawing conclusions from the data.

The research present in this study can be considered relevant as it will impact positively the literature available that is related to hydrogen investments in the energy industry. The key findings present in this thesis can provide valuable insights to firms considering investments in hydrogens projects and anyone interested in promoting the use of clean energy sources.

2 Motivation

In recent decades we have seen an increase in the need to take action to reduce global warming, greenhouse emissions while focusing on being more sustainable. In light of new policies and international efforts this necessity to take action is now more relevant in socio-economic terms as well as political and legal. There is a niche that is available to be taken and the energy and fuel industry as seen a growing interest to take part of this, one of the most probable paths to capitalize this will be hydrogen, as seen in IEA reports [1] . However, there is still some uncertainty if hydrogen related projects and investments are factually beneficial for a company and its growth from a market perspective.

A diverse energy mix and the green transition are fundamental topics to be considered when discussing the potential of hydrogen as the next viable step in becoming energy independent in a world that orbits around fossil fuels. The key takeaways from this thesis may provide the insights for firms in the industry looking to jump to or include hydrogen related investments in their portfolios and company mission.

3 Literature Review

3.1 Foundation of the Literature Review

Recently, hydrogen climbs the ranks as one of the most promising alternatives for clean energy, and this is found by looking at how much investment is being made in research and development of hydrogen related technologies when comparing it to other prototype technologies. Governments and companies seek to invest in hydrogen too, among other things, reduce their carbon footprint and transition to a green status quo. This can be seen by the creation of Paris agreement, which was supported by most of the governments in the world. This leads towards the question regarding performance of companies that have already taken the step further by investing their valuable resources into R&D. This is the main focus of this thesis, which aims at answering the research question, which is: "Do hydrogen investments cause company market value to increase in comparison with it's peers?"

Even though there has been a recent increase in the hydrogen investments, the number of relevant papers for this thesis is still limited. This can be seen when the hydrogen literature is compared with the other green investments possibilities, such as wind or solar solutions. The structure of this literature review is created while considering this limitation. There are 4 sections including this one. These sections focus on the current state of hydrogen from different perspectives. In addition to this based on the nature of this thesis there is a focus on financial statistical methods/models, which can be used to provide answer to previously mentioned research question.

The purpose of this literature review is to introduce to the reader the current state of hydrogen in energy sector and statistical models/methods that can be of use for analysis of the effect of hydrogen investments on the market valuation of a company.

3.2 Overview of the current state of Hydrogen

Hydrogen can be extracted from several substances. As an example of these substances can be hydrocarbons and water. One of the most promising chemical processes is water electrolysis. Another promising way to produce hydrogen is hydrocarbon steam reforming, method which is utilized globally in most of the hydrogen-production plants. The cost of the production of these two methods are discussed by Zhang [2]. The authors also focus on other methods which can be used to produce hydrogen and how cost efficient they are also. Zhang include methods that are based on fossil fuels, which provides great comparison from financial perspective with the methods that are based on renewable sources. Wang [3] also focuses on the efficiency of production based on the cost to produce 1 kg of hydrogen. However Zhang [2] works only with renewable methods. These papers introduce some of the most promising methods for the upcoming years, many scientists are trying to achieve the lowest cost of production. However due to the early stage most of these methods are in the phase of research and development without large industrial utilization.

On the other hand based on the IEA [4] report from 2019, which provides information about the development of demand for hydrogen in the last years. It is clear to see that larger number of companies and government institutions are getting involved with the methods that have been mentioned above and create their own research and development projects to get an edge over their peers. This happens in several areas, where one of these is hydrogen usage in energy sector. On the following figure it is possible to see that in the recent years there has been a great increase of these projects. Mostly aiming at vehicles, however there are also grid injections and electricity storage. [4]

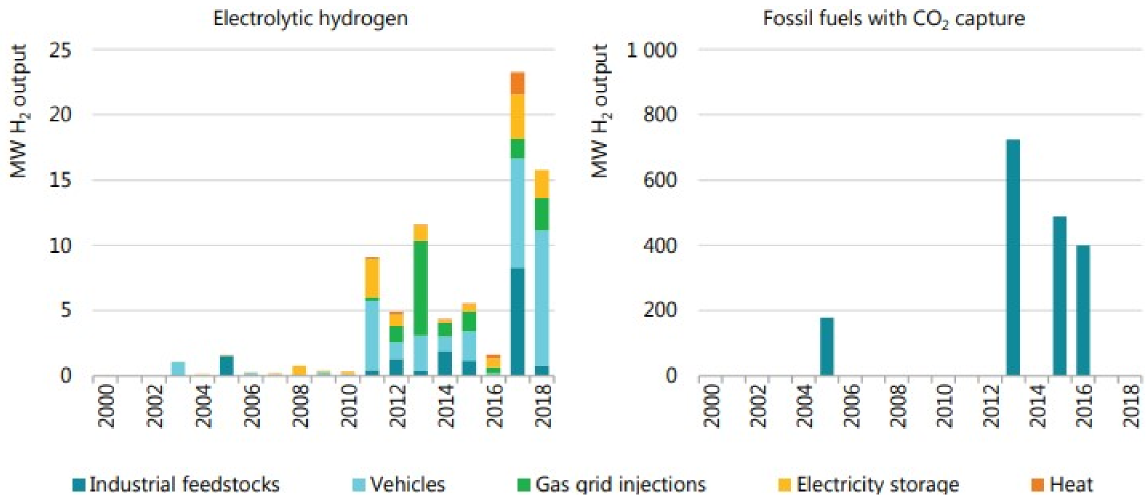


Figure 3.1: Capacity of new projects for hydrogen production for energy and climate purposes, by technology and start date

In the right part of the figure 3.1 can be seen the fossil fuel based hydrogen production for industrial application with CO2 capture. The values are higher in comparison with the output of electrolytic hydrogen. However this thesis focuses only on the companies that have been involved with the clean hydrogen, therefore there has not been any fossil fuel based projects. The picture is only for the reader to get an idea about the current state of the industry. Electrolytic hydrogen has been increasing since 2010 and this support the growing trend during the last years. However investing in these projects is far from easy based the unknown factors from the future.

Hydrogen and Investments

One of the crucial aspect that affect every investment is the timing, when the investment is made. Hydrogen investments do not differ, however there are also other factors, which are very important. Benthem [5] considers hydrogen infrastructure, which also affects the strategy of an investment. Infrastructure is one of the main concerns of other papers too. Zhu 2020 [6] and Zhao 2021 [7] are focusing on the importance of storage and logistics. All of the above mentioned are focusing on investments, which each company needs to carefully analyse before taking an action. Nevertheless, there are already companies, which took the extra step and these companies are focus of this paper, where they will be compared with other companies based on their market price movement. As previously mentioned the infrastructure is one of the main concerns for the development of the hydrogen investments. Based on this the U.S. Department of Energy have launched a H2USA, which is a collaboration with public and private federal agencies, automakers but also hydrogen providers and fuel cell developers. The focus is to advance the hydrogen infrastructure to support the possibilities of transportation of energy in the future. [8] However, this is not the only aspect that can motivate companies to invest into the hydrogen. The Paris Agreement can be seen as a milestone for the green transition that motivated several companies to invest into the future. This is directly tied to the research question of this paper, that focuses at the market valuation and if the companies that started investing are performing better than their peers. For the analysis of the impact/change after the Paris agreement this paper uses the DiD method, which is discussed in the following paragraph.

Difference in Differences Method

The model method that has been chosen for this paper was Difference in Differences. This decision was based on the fact, that the DiD model is very popular approach in economics and other social sciences, which aims at evaluating certain effects, which can differ in nature, regarding the the environment. In case of this study it is the policy change regarding the reduction of CO2 emissions and shift towards green energy, and how this policy change affected the market price of a companies, which have already invested into the hydrogen. The variety of usage of DiD can be seen from several scientific papers. Goodman-Bacon [9] focus on the treatment timing across groups of units or Athey [10], who aims at the alternative methods of the DiD model by introducing non-linear models. As an example from financial perspective can be Li [11], that applies the DiD model in strategic management research. Overall it can be said that the DiD model is still being improved and challenged by the researchers, however it is a strong tool for the evaluation of interventions, shocks and changes between treatment and control groups.
All the needed information was obtained from several books, which regards econometrics models/methods with the focus on panel data and basic of the DiD model as example of used authors can be Wooldridge, Damodaran, Baltagi, Badi H. As a searching engine for the scientific work was used google scholar.

3.3 Recent Themes and Trends

Most of the current papers that are focusing on the state of hydrogen acknowledge the positive hydrogen trend as an example of this can be Momirlan [12], focuses on the complex questions regarding the current status of technology, economics and environmental impact. Technology is one of the leading areas, that affects the hydrogen usage. Kalamaras [13] looks into the problematic regarding different ways of hydrogen production with a special focus on fossil and renewable biomass resources. These are only a few papers, which are supporting the green transition, especially by the usage of hydrogen, however there are some limitations of hydrogen that need to be addressed.

The current challenges that the hydrogen is facing now based on IEA [4] are:

1. The cost of production from low-carbon energy.

   Based on the IEA analysis, the cost of producing hydrogen from renewable energy sources could fall by 30 % in 2030. This is one of the main setbacks of green hydrogen, therefore one of the main concerns for the researchers. This is also mentioned in section 3.2 of this literature Review.

2. The speed of development of hydrogen infrastructure is slow, this hold back the widespread adoption.

   With better infrastructure it will be easier to achieve more efficient logistics, which would lead to decreasing the price for the consumers, however the initial investments for these projects are high. That is why there are projects like H2USA.

3. Regulations, which could potentially limit the development of clean hydrogen industry.

   It is about the collaboration between the government and the industry. The goal is to ensure that existing regulations will not be an unnecessary barrier for investors.

Based on this problems and challenges IEA [4] also analysed some of the near-term opportunities that would improve the path towards the widespread use of clean hydrogen. Some of these are:

1. Build on existing infrastructure.

   There is a possibility to introduce clean hydrogen to the natural gas suppliers. As much as 5% of the countries volume would significantly boost the demand for hydrogen. In addition to this it would severely drive down the costs.

2. Launch the Hydrogen trade's routes. The global LNG market can be a great example of how this utilization of shipping routes could improve the position of hydrogen in the world.

3.4 Research Gap

As previously mentioned in section 3.1 the biggest problem is the stage of the research. Many researchers are working on promising solutions for above mentioned challenges that comes with efficient implementation of hydrogen. However only time will tell if those solutions are cost efficient etc. The ongoing research of electrolysis has been improving in the recent years significantly but the projects, which aim at larger scale implementation have not results yet. These results could make it easier to analyse the potential investments.

4 Methodology

4.1 What types of data do we know?

In economics there are several ways to perceive data. It is usually tied to the specific methods, however in case of econometric methods, the data can be applied almost without or little modifications. Nevertheless, some type of datasets do have special features that need to be accounted for. The following part will introduce a few of the basic data structures that are often seen in the field of econometric. [14]

4.1.1 Cross Sectional Data

It can be a set that consists of a sample of households, firms, cities etc. or a variety of other units, which were taken at a certain point in time. It is possible that sometimes, the data on all units are not responding precisely to the same time period. One of the important features of cross sectional data is one of the core assumptions regarding the way of the collection of the data. The data is assumed to be obtained by random sampling. It is done because it simplifies statistical analysis. Cross section and is widely used especially in economics and other social sciences. Regarding economics it is closely tied to microeconomics fields such as labor economics, state and local finance etc. [14]

4.1.2 Time Series Data

The difference from cross sectional data is that during the time series analysis is the focus, which is put on several observations either from a single variable or several variables over time. Therefore, time series data is tied to time and the development of the observed variables during. One of the challenges that appears when working with time series data is that the economic observations are very rarely independent across time. Most of the economics data sets are in some ways related to their histories. One of the examples of this is a GDP from last quarter can provide information or likely the range of the following quarter. Due to this there have been developed several methods that provide ways, which exploit the dependent nature of economics time series and also address other issues. Data frequency is another feature that appears in time series data sets. Frequency specifies the rate at which the data is collected. The most common frequencies are weekly, monthly, annually, etc. In the case of stocks it can be even shorter such as days, hours, minutes. [14]

4.1.3 Panel Data

The panel data is a combination of a time series and cross section members in data sets. The key feature of panel data and the distinction from cross section data is the fact that the units are followed over a given time period. The panel data requires replication of the same units during the observing time. It can be more difficult to obtain than pooled cross section, however, observing same units leads towards several benefits over simple cross sectional data. The advantage leads towards having multiple observations on the same units, which allows control of certain unobserved characteristics. Another benefit of panel data is the possibility of studying the importance of lags in behavior or the result of decision making. This is important especially in an economic environment, where it is common that changes take a long time to be fully implemented. [14]

4.2 Why should Panel Data be used?

There are several benefits and assumptions that come, when working with panel data. The main benefits are: (1) The control of individual heterogeneity. Panel data propose that the data sets that consist of individuals, firms and countries are heterogeneous. This is a huge advantage in comparison with time series and cross section, which are not controlling the risk included in heterogeneity, which could potentially lead to biased results. [15] (2) Panel data provides more informative data with more variables, in addition to this there is less collinearity among the variables, more efficiency and more degrees of freedom. With more data it is usually easier to produce more reliable parameters. [15] (3) Dynamics of adjustment's study is supported more by panel data. The distributions of cross section, which look stable, hide possibly a multitude of changes. Panel data are also suitable for a study of the duration of certain economical effects just like unemployment and poverty, in addition to this if the duration is long enough it can provide information also about the adjustments and changes in the economics datasets. [15] (4) Panel data are also more precise in identifying and measuring effects that are simply not as detectable in pure time series or cross section. [15] (5) When looking at the complexity that can be achieved, the panel data allows us to test and construct more advanced behavioral models in comparison with purely cross section and time series data. [15] On the other hand there are some limitations that come with the usage of panel data. This can be an example of one of them: The problem can be tied to more difficult design and data collection. This problem can appear during surveys, that can provide observations, that are not as precise due to lack of cooperation, etc. [15] These problems can mostly appear during data collection for economics purposes. Asking about wage, jobs etc. When collecting data through servers that provide stock information it is possible to minimize these types of error, however due to the complexity of the models it is still more difficult to design them in comparison with time series and cross section.

4.3 What types of panel data there are?

Panel data can be presented in several ways. It is mostly regarding the number of observed entities and for how long. Since it is a connection of time series and cross section it can be created in many ways. These are some of the main groups, how can panel data be shown.

4.3.1 Balanced and Unbalanced Panel

Panel data can be seen as a balanced panel when all entities are present in all observed time periods. The Unbalanced panel is the opposite, where the entities that are observed in a data set are not present in all time periods, this can be more common than initially expected as incomplete data in data sets can be a common occurrence. It is quite clear that the balanced panel is more suitable, however sometimes, as mentioned in previous section section 4.1, there can be problems with data collection leading to missing values. Even though this can seem as a huge problem, there are several regression models that work for both balanced and unbalanced panel data.[16] Balanced panel can be seen in the following table (Same amount of entities and years). Another way of categorizing panel data is the Long and Short panel, it is closely tied to the amount of entities and observed periods. Where in the long panel data there are more periods than entities and in the short panel data it is the opposite. [17]

Entity ('E')	Year	Y	X1	X2
E1	2020	123	2	12
E1	2019	100	2	14
E1	2018	98	2	15
E2	2020	242	4	17
E2	2019	272	4	15
E2	2018	282	4	14
E3	2020	78	7	10
E3	2019	120	7	8
E3	2018	60	7	12

Table 1: Balanced Panel

4.3.2 Dynamics Panel

As previously mentioned section 4.1 panel data are able to define more sophisticated situations. This can be achieved especially while using a dynamic panel, which includes lagged dependent variables (Y).[16] It allows the model to provide more explanatory power. The dynamics panel can be seen below. Year 2018 is not part of the data, since it was included in the following year as lagged value.

Entity ('E')	Year	Y	Y-1	X1	X2
E1	2020	123	100	2	12
E1	2019	100	98	2	14
E1	2018	98		2	15
E2	2020	242	272	4	17
E2	2019	272	282	4	15
E2	2018	282		4	14
E3	2020	78	120	7	10
E3	2019	120	60	7	8
E3	2018	60		7	12

Table 2: Dynamic Panel

4.4 Regression Estimation

There are different types of regression depending on different types of data. This subchapter will focus on describing the basic regression estimators and why it is not possible to use certain regression on a previously mentioned table 1

4.4.1 Time Series regression

As previously mentioned time series are focusing on the development of the dependent variable throughout a certain time period, therefore if this regression would create 3 different regression models if applied on the table 1. The method that would be most likely used is OLS, that is one of the most common when working with Time series regression.[18] The following regression looks as follows [18]:

$$y_t = \beta_0 + \beta_1 z_t + u_t \; where \; t = 1, 2, ..., n \tag{1}$$

The problems that will emerge are connected to the disparate piece of information that would affect the way how it is possible to perceive a relationship among the explanatory variables and the dependent variable Y. In addition to this issue there might be a serial correlation that can emerge due to the time dependent nature of Y. [18]

4.4.2 Cross Section regression

The cross section regression can be described by a similar formula as in the case of time series, however due to the different nature of the data there are other problems included when applying on panel dataset. In the case of table insert number, there would be 3 different cross section regressions, which would be very limiting because there are only 3 firms that are being observed. In connection to this the parameters would have to be estimated from this data. This leads towards limitation of degrees of freedom and would potentially lead to not comprehensive analysis.

4.4.3 Panel regression

In comparison with the time series and cross section the panel regression has slightly different formula as follows [19]:

$$y_{it} = \alpha + X_{it}\beta + u_{it} \; where \; i = 1, ..., N \; t = 1, ..., T \tag{2}$$

The difference is the change of an error variable from u_t to $u_i t$, where $u_i t$ is a combination of more error terms.[19] The reason why this regression is more relevant is because of the ability to allow more observations, more degrees of freedom, in addition to this it incorporates changes not only within a firm but changes across firms. The last benefit is that this regression accounts for the impact of firm specific attributes.[19] Because of these possibilities the panel data regression is the most suitable regression for the model, which will be worked with in this master thesis.

4.5 How does Panel regression works?

The easiest way to model a panel regression is to use a simple linear model, these models contain parameters and variables from which is then created a regression. There is a minimum number of variables that are needed for the model to work. The simplest is one explanatory and one dependent variable. These two variables are shown in a panel regression formula from 4.4.3 , where Y is a representing dependent variable. The explanatory variable is represented by x. Nevertheless, it is not the only property that the simple model is made of. It would be impossible to assume that the explanatory variable covers the whole model perfectly because of this there is another variable that acknowledges the potential errors, that cover the information that was not explained. [20] In addition to the X variables there are two more coefficients one of them is α and the other is β. Alpha is a scalar and beta is a slope that affects the regression. Panel data differ from time series and cross section by the subscript, which in a case of panel data needs to account both i and t, where i denotes households, individuals, firms, countries etc. and t is denoting time. [19]

4.5.1 Error Terms

As previously mentioned error terms are something that regression models require and especially in panel data these errors are relevant because it is a combination of time series and cross section, therefore it needs to acknowledge both. This means that the error terms have two components attributed to them: individual specific effects and time specific effects. [21] While one of the components is individual specific effects that represent the variation or heterogeneity of the dependent variables across the different units, like starting conditions of units or their unobserved characteristics, the other component is the time specific effects which just observe the variation of the dependent variable over the specified time periods. [21] When considering the disturbance terms within panel data there are several assumptions that are significant for analysis, one of them being that disturbance terms are not correlated with the explanatory variable and have a zero mean, violation of this assumption can lead to biased estimates and inconsistent conclusions. [22] Another assumption states that error terms have to be identically and independently distributed across units and time periods and if this assumption is broken results might show incorrect standard errors and poor estimates. To summarize error terms in panel data are assumed to respect exogeneity and homoscedasticity and since there are several factors to take into consideration when doing a panel data analysis like serial correlation, cross section dependence or heteroscedasticity it is essential to have a careful consideration of the error terms as to guarantee the results obtained through analysis can be considerate reliable and valid given the specific case at hand. [21]

Now, it is worth mentioning that depending on the regression model applied, each error term present on these models might behave a little different from each other if compared, this is of course because each regression model has its own assumptions and specifications. Without going in depth to each specification the general look of the error terms for each regression model will be: Fixed effects - Error term consists of unit specific effect and idiosyncratic error that captures the within unit variation that is not explained by the model, this error term is assumed to be uncorrelated with the unit specific effect while being possibly correlated over time for the same unit. The resulting residual is the idiosyncratic disturbance term. [21]

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} \tag{3}$$

$$\alpha_i = \bar{y}_i - \beta \bar{x}_i \tag{4}$$

$$u_{it} = y_{it} - \bar{y}_i - \beta(x_{it} - \bar{x}_i) \tag{5}$$

Random effects - Components are the same as Fixed effects and while the error term is also assumed to be uncorrelated with the unit specific effect there is the added assumption of it being independently and identically distributed across units over a given time period within the panel data set. [22]

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} \tag{6}$$

$$\alpha_i \sim N(0, \sigma_\alpha^2) \tag{7}$$

$$u_{it} \sim N(0, {\sigma_u}^2) \tag{8}$$

Pooled Model - In this regression model the error term only follows one assumption, which is its independence and identical distribution across all units over a given time period within the panel data. There is no consideration for unit specific effects or time invariant heterogeneity across units. [23]

$$y_{it} = \beta x_{it} + u_{it} \tag{9}$$

$$u_{it} \sim N(0, \sigma_u^2) \tag{10}$$

4.6 Difference in Differences Model

A Difference in differences (DiD) model is a method mostly used in econometric research and social studies. By studying the differential effect of a “treatment” on a “treatment group" versus a "control group" this statistical technique tries to identify positive or negative changes on a given dependent variable given a certain event and conditions. While it is a widely used model in empirical research the determination of this as an appropriate model depends on the data structure, research question and all the assumptions underlying the model. In this chapter, we discuss the selection of a DiD model for a balanced panel dataset with 2592 observations and go through all the considerations needed to have this method work with our research question. We argue that the random effects estimator is better suited for this dataset compared to fixed effects, pooling, and OLS. Difference in differences is a good decent non-experimental method for impact evaluation if when looking at the available data we can see that (i) randomization is difficult or expensive, (ii) the control group is well defined and (iii) assuming that the trends in treatment (event) and control in the absence of treatment would have follow the known pattern until the treatment even[24]

4.6.1 Data needed for DiD

As discussed previously we will be using panel data, this includes both a time series and cross sectional data. We will have a balanced panel dataset set which means that the number of observation for each time period and cross sectional unit are the same, in term of time series we look at data starting from January 2010 to the end of 2021 and for our cross section we will have 18 companies belonging to the energy industry in the USA. The 18 companies are divided into control and "treatment" groups in an even number and the consideration used is the factual existence of hydrogen investments. By using a DiD model on our data set we will estimate the average event effect as the difference that occurred between the variation of the outcome variable for both the control and "treatment" group relative to their pre-event levels. The event mentioned here is the Paris agreement.

4.6.2 Overview of DiD

In general when considering a Difference in difference approach we can assume a general model. Assume $i = 1, \cdots, n$ unit (companies in our case), and T as time periods, where $t = 1 \cdots T_0$ identify pre-treatment and $T_0 +_1, \cdots, T$ identify post-treatment. All the possible outcomes for i in period t (all potential outcomes for the companies during the time periods) with or without treatment are denoted by Y_{it}^1 and Y_{it}^0 respectively.[25] Consider D_{it} as the binary dummy variable that identifies if a company i is "treated" in period t or not. Given these consideration a general model for the possible outcome in the absence of treatment can be[25]:

$$Y_{it}^0 = X_{it}\beta + \lambda_t \mu_t + \delta_t + \epsilon_{it} \tag{11}$$

here X_{it} is a vector of observed time-varying covariates and μ_i the unobserved time-invariant characteristics whose effects, here represented as λ_t that may vary over time but are not expected to differ across units, δ_t is for the common time effects and the error term ϵ_{it} identifies exogenous unobserved idiosyncratic shocks.[25] Now, assuming a treatment effect as

τ_{it} we can propose an edit to the previous formula to represent the outcome of units under treatment[25]:

$$Y_{it}^1 = X_{it}\beta + \lambda_t \mu_t + \delta_t + \tau_{it}\epsilon_{it} \tag{12}$$

If we assume that treatment only affects the treated units in the periods that concern the post-treatment period, our observed outcome can be written as[25]:

$$Y_{it} = D_{it}Y^1 it + (1 - D_{it})Y_{it}^0 \tag{13}$$

4.6.3 Assumptions

For a Difference in Differences model to hold several assumptions need to be in place, essentially we will find the same assumptions as any OLS model but we also need to have a parallel trend assumption, which means that in the DiD model we expect that the “treatment” group and the control group will have followed the same trend in the outcome variable if no event had occurred.[25] In summary assumptions in this model will involve:

We assume “treatment” and control groups are comparable in terms of observable and unobservable characteristics. The event is independent of the outcome variable, this means that the “treatment” group was not selected based on their outcomes before the event.[25] There are no influences from one group to another.The parallel trend assumption holds and is not violated.[25]

4.6.4 Limitations

Like all statistical methods and research tools we are met with imperfections that need to be taken into consideration and weighed, we have some limitations to the Difference in differences model that need to be considered:

The event effect is assumed to be constant over time, while it is possible for the event effect to vary over time because of changes in the intervention and other external factors [25] It is assumed that other unobserved variables will not affect the outcome variable, while it is possible for one unobserved variable to be both correlated with the event and the outcome variable. This may lead to some bias on the estimated event effect. The DiD model assumes no measurement errors in the outcome variable. In reality measurement errors may occur due to external factors. [25]

4.6.5 DiD Summary

Overall the difference in difference model can be a powerful tool for estimation when used correctly. By considering all the limitations and assumptions carefully before applying it to the data we can reliably estimate the event effect on a particular outcome variable. In this study we will use a random effect estimator to address the unobserved heterogeneity across companies in the data set and draw reliable estimates of the effect of the Paris agreement and draw meaningful conclusions on its impact in our outcome variable.

4.7 Estimation Methods for Panel Data

To be able to apply a Difference in differences model with panel data we need to select the most appropriate estimator, this can be: fixed effects, random effects or pooled OLS. Each technique has its own limitations and assumptions, which can be a better fit given some specific conditions to each case, we will discuss this key points below.

4.7.1 Pooled OLS

OLS (Ordinary Least Squares) is a common regression that has multiple uses and can be used to estimate the differences in differences in the model. It is also likely that OLS will produce biased results if unobserved differences between the control and "treatment" group exist, which is the case. Pooled OLS, which is another method within the OLS regression can also be used to estimate the DiD by combining the control and "treatment" groups into one group and estimating the event effect as the difference in the outcome variable between the post and pre-event periods. This assumption among others makes pooling a less efficient method than DiD methods [22]

This model needs that all individuals have no correlation and it ignores time and other individual characteristics of the objects while focusing on the dependencies

$$Y_{it} = X_{it}\beta + \alpha_i + u_{it}; for\ t + 1, 2, ..., T\ and\ i = 1, 2, ..., N \tag{14}$$

Taking what was previously stated X_t and α_i can not be correlated, which translates to $Cov(X_{it}, \alpha_i) = 0$.These very unique details make Pooled OLS inappropriate for panel data, in most cases. [21]

4.7.2 Fixed Effects model

A Fixed Effects estimator will control for unobserved individual heterogeneity across the cross sectional units (Companies) by estimating a separate intercept for each unit. By assuming that the treatment effect is constant on all Companies this estimate eliminates time invariant issues but it can become biased if there is any time varying unobserved heterogeneity that can be correlated with the treatment status. [22]

Typically panel data is composed of a time period (time series) with several individual groups (cross-section) populating it with data and their own characteristic, observed or unobserved. These can stay constant over time like type of industry or geographical location, and the invariant characteristics are assumed to be correlated with the characteristics that in fact vary over time. This will influence the error terms which will cause an omitted variable bias.[22] By using a Fixed effects model where there is estimation of the common coefficients on the time varying regressors we avoid this bias. This model is particularly efficient when a research is pointed in only looking at variables that change over time. [22] A Fixed effects regression model will normally look like:

$$Y_{it} = \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \alpha_i + u_{it} \tag{15}$$

Where $i = 1, \ldots, n$ and $t = 1, \ldots, T$ and $Cov(X_{it}, \alpha_i) \neq 0$. We have entity-specific intercepts represented as the alphas and this allow to find heterogeneities across objects in the sample, and these individual intercepts are to be interpreted as the fixed effects of the object i. Having the assumption of exogeneity (independent variables X are not dependent on the dependent variable Y) on the explanatory variables, Fixed effect model become unbiased, so idiosyncratic error units should be uncorrelated with all explanatory variable that is used across all time periods. [22] Constant explanatory variables get swept away under the Fixed effect transformations.

4.7.3 Random Effects model

The Random Effects estimator assumes that the independent variable is random and that the differences between individuals are random and that the time invariant heterogeneity in the cross sectional units is not correlated with the event status (The Paris Agreement). This estimation method allows the event effect to vary across companies and estimates the average event effect as the difference in the outcome variable between the control and "treatment" group while being adjusted for time varying inaccuracies. Normally the Random Effects estimator has higher efficiency than the Fixed Effects estimator since it produces unbiased estimates if time varying unobserved heterogeneity exists that might be uncorrelated with the even status. [22]

By assuming that the entity's error term is not correlated with the predictors, Random effects will allow for time-invariant variables to play a role as explanatory variables. [22] Specifying those individual characteristics is important as they may not influence predictor variables. One of the issues with this estimator is the occurrence of variables that are not available which will lead to omitted variable bias in the model and by using the Random effects we will allow for generalization of inferences beyond the sample with this model. [22] The main assumptions present in this model involve the independent variable being random and the differences between individuals to also be random (assuming a distribution with random parameters). In the Random effects model there is a correlation in time between random elements relating to the same objects, but there should not be correlation of random elements of different objects in time periods that are different. [19] This model can outform others when we see individual effects not being correlated with explanatory variables. However, if the individual effects are in fact correlated with the explanatory variables then this model will be ineffective. While being a more effective model for a high number of factor levels it becomes less effective when there is a smaller number of factors, but this situation does not mean the model is not efficient or should be discarded. For example:

$$Y_{it} = \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \alpha_i + u_{it} \tag{16}$$

The equation here may become a random effect model if the assumption that unobserved effect α is uncorrelated with each explanatory variable occurs:

$$cov(x_{itj}, \alpha_i = 0);\ t = 1, 2, \ldots, T\ ;\ j = 1, 2, \ldots, k;\ i = 1, \ldots, n \tag{17}$$

In general when looking at the Random effects model we will find all the assumptions for the Fixed effect model already present, additionally there is the requirement for α_i to be independent for all explanatory variables in all occurring periods of time present in the data. The Random effects model will always assume that the unobserved effects are not correlated with all explanatory variables, regardless of whether these variables are fixed over time or not. [19]

4.8 DiD Valuation statistical tests

4.8.1 The Hausman test

The Hausman test is a method to decide between the Fixed effect model or the Random effect model where the null hypothesis assumes that Random effects is the one to use, so if rejected then Fixed effect is considered to be of best fit. [22] This test takes in the normal assumptions of both Fixed effects and Random effects, where Random effect's unit specific effects are uncorrelated with other regressors present in the model and where the Fixed effect has the assumption that the unit specific effects are in fact correlated with other regressors in the model and then tests the coefficients estimated from these two models and determines which one is more efficient, by looking at which one possesses the small variance. According to Woolridge [22] when the null is not rejected that in practical terms both models are sufficiently close so choosing one over the other might not really matter in the end. The null hypothesis also states that there is no correlation between unique errors and regressors found in the model. When we interpret the results we should look at the p-value, where the alpha significance level, usually 0,05, is below this level then we reject the null hypothesis. [26] A general take on this test can be given by the formula:

$$HausmanTest = (v_{FE} - v_{RE})'(VarCo_{FE} - VarCo_{RE})^{-1}(v_{FE} - v_{RE}) \tag{18}$$

where
v_{FE} is the vector of coefficients estimated from the fixed effects model.
v_{RE} is the vector of coefficients estimated from the random effects model.
$VarCo_{FE}$ is the variance-covariance matrix of the fixed effects estimator.
$VarCo_{RE}$ is the variance-covariance matrix of the random effects estimator.
$'$ denotes the transpose operator.
$^{-1}$ denotes the matrix inverse.

4.8.2 Levin-Lin-Chu (LLC) test

The Levin-Lin-Chu test is a version of the ADF (Augmented Dickey-Fuller) test which is used to test for unit roots in time series data. The LLC test is a robust test that can account for dependencies and heterogeneity across cross sections in the data while being a versatile test as it can be used in both stationary and non-stationary sets of panel data. This test has the assumptions that there is cross section independence, heterogeneous slopes and intercept and an homogeneous unit root. This test performs well with bigger data samples where a large T and N can be found, in other words a large number of time periods within the time series and unique entities within the cross section. [27]

4.8.3 Breusch-Pagan Lagrange Multiplier

The Breasch-Pagan test is a statistical test that researchers can apply to panel data in order to detect the presence of heteroscedasticity of the disturbance term within the data. Heteroscedasticity is seen as the variance of the error term (disturbance term) varying across different units or time periods in the data. This means that what can be expected as optimal is homoscedasticity, which is when variances of the error terms take the form of some constant number. The null hypothesis in this test assumes that variances of the disturbance terms are constant and independent of X, it states the following: H_0, the residuals are distributed with equal variance, $var[ux_1, x_2, \dots] : \sigma^2$ so whereas H_1, the variance is not a constant and is not independent of X in an unknown way. [28] The test can be performed in a few steps, first we estimate the regression model that is chosen, have the regression squared residuals calculated, then fitting a new model with the obtained squared residuals. After this we obtain R^2 and can calculate the Lagrange Multiplier (LM), afterwards using the chi-squared distribution we can ascertain the significance level and determine if the null is rejected or not. [22]

$$LM = T \cdot R^2 \tag{19}$$

where:
T is the number of time periods in the panel data.
R^2 is the coefficient of determination from a regression of the squared residuals from the original model on a set of explanatory variables that capture the panel-specific or time-specific effects.

It is worth mentioning that this test can be applied to both Fixed effects and Random effects, the only difference that needs to be taken into account is that the formulation of the test statistic and the degrees of freedom will differ from one model to the other. Both follow the chi-squared distribution with the difference resting on the degrees of freedom, Fixed effects has its degree equal to the number of units in the data minus the amount of time invariant regressors, because this model absorbs the cross section variation it leave only the within unit variation to be tested for heteroscedasticity. While the Random effects has its degrees of freedom be the same as the number of panels minus the time invariant regressors and this is because this model allows for cross section heterogeneity.

4.8.4 Breusch-Godfrey

The Breash-Godfrey test allows to check for the existence of autocorrelation within a regression model. Given that this paper works with panel data the regression models with have their errors checked for the presence of autocorrelation. [29] While using this test first the R-squared is found using:

$$\epsilon_{t,1} = \alpha + \sum_{j=1}^{p} (\beta_j \cdot \epsilon_{t-j,i}) + \sum_{k=1}^{K} (\gamma_k \cdot X_{k,t,i}) + u_{t,i} \tag{20}$$

where:
$\epsilon_{t,i}$,i is the residual for individual i at time t from the original panel regression model
α is the intercept
β_j is the coefficient on the lagged residuals (j = 1, 2, ..., p)
$X_{k,t,i}$ are the k regressors in the panel regression model
γ_k is the coefficient on each regressor
p is the number of lags included in the auxiliary regression model
u_t,i is the error term

Afterwards the Breusch-Godfrey test can be calculated with the following formula:

$$BG = t \cdot R^2 \tag{21}$$

where:
BG is the Breusch-Godfrey test statistic
T is the number of time periods in the panel
R^2 is the coefficient of determination from the auxiliary regression model, which is estimated using the residuals from the original panel regression model.

The null hypothesis for this test states that there is no autocorrelation among the disturbance terms of the panel regression model. This means that the errors are uncorrelated across units and time in the panel data set. The alternative hypothesis states that there is factual autocorrelation among the errors in the regression model. [29]

4.8.5 Jarque Bera

The Jarque-Bera test is a test for normality that checks if the kurtosis and the skewness of the data sample match a normal distribution by testing the residuals, a normal distribution needs to have a skewness equal to zero and kurtosis equal to three. [30] Where residuals are to be seen as the differences between the observed points of data and the predicted points of data from a statistical model. The null hypothesis H_0 suggests that the kurtosis and skewness match a normal distribution, for panel data the statistical test can be seen as[30]:

$$JBtest = \frac{T}{6} \cdot [S^2 + 0.25 \cdot (K-3)^2 \cdot (N-1)] \quad (22)$$

where:
T he number of time periods
N is the number of cross-sectional unit
S is the sample skewness coefficient
K is the sample kurtosis coefficient

4.8.6 Robust T and Robust F

The Robust T and Robust F test are both important tests for panel data, they allow to check the significance of the coefficients present in regression models. To test the statistical significance of individual coefficients, which means we take into account how much does X (independent variable) impacts Y (dependent variable) in the regression model, we use the Robust T test. This test, which is also called other names like the panel-corrected standard errors (PCSE) T-test for example, considers the correlation between observations inside each panel and will provide more relevant results when compared to the standard versions of the t-test, the general formula can be viewed as[30]:

$$t = \frac{\hat{\beta} - \beta_0}{SE_{robust}} \quad (23)$$

where
$\hat{\beta}$ is the estimated coefficient for the independent variable of interest
β_0 is the null hypothesis value (usually zero)
SE_{robust} is the robust standard error of the estimated coefficient, calculated using the clustered standard errors approach.

The Robust F test, also known as the Wald test will test the overall relevancy of a group of coefficients, which will consider how much does a set of independent variables impact the one dependent variable. Similar to the Robust T test, this test also considers the correlation between observations inside each panel and provides better results than its standard counterparts by testing if at least one of the independent variables in the model has a statistically relevant impact on the dependent variable. The general formula for this test when considering panel data can be seen as[22]:

$$F = \frac{Rss_{res} - Rss_{Unres}}{q(N-K)} \tag{24}$$

where
Rss_{res} is the residual sum of squares from the restricted (null) model
Rss_{Unres} is the residual sum of squares from the unrestricted (alternative) model
q is the number of independent variables being jointly tested)
N is the total number of observations
K is the total number of regressors (including the intercept) in the unrestricted model.

4.9 Stock prices as model variable

There are a variety of financial and statistical models that are working with the variable made of stock prices. The stock price as itself can contribute a lot, especially by providing information regarding the changes in the asset price. These changes can show trends and public reactions against the changing environment in which the company that issued the stock operates. As the environment can be seen in certain markets, sectors, nations and so on. [31] The prices can be used in different perspectives depending on what is more suitable for the model in which they are the explanatory/explained variables. One of the most used ways is using them as absolute values or as a percentage when considering price variation. These two ways can be modified by the nature of the data depending on what the model is trying to achieve. As an example of these modifications the time frames can be different, which can be as small as hourly data up to yearly and so on. [31]

4.9.1 Asset Returns

Stock prices are mostly used to calculate returns of the underlying asset. Most of the financial studies are focusing on returns instead of absolute prices for several reasons.[32] One of these reasons can be the simplicity of interpretation for potential investors. Second reason is the nature of the data that is more attractive for the statistical properties, therefore it is easier to perform statistical testing and modelling. [33]

$$1 + R_t = \frac{P_t}{P_{t-1}} \tag{25}$$

This equation shows a one period simple return, which is a simple gross return of an asset in percentages. R_t stands for return of an asset and P stands for the price of an underlying asset [2], which can be and most of the time is a stock price. T that is specified by the nature of the data ranging usually from $t = 1, 2, 3, \ldots n$ As an example of a model that uses modified stock prices is the capital asset pricing model known as CAPM, which describes the relationship between expected return and systematic risk. CAPM can also be used to measure the portfolio performance.[31] In the case of the Capital asset pricing model the dependent variable can be the percentage change of adjusted stock prices.

4.9.2 Volatility

Price movements can be tied to more than just the absolute change between two periods. Another important statistical property is volatility, which shows the variance of changes of a price of a stock. In financial data it is mostly important because of forecasting models, which are easier to apply on volatility of a data instead of absolute values. It is because the absolute price movement of a stock, index etc. is in most of the cases unpredictable and follows statistical properties such as random walk. However, the variance is often used because it is not affected by these effects (RW).The returns and the standard deviations can show a lot about the price development and that is why it is so relevant in many models.[34]

4.9.3 Motivation to use 1

Since the discussed topic of this master thesis is closely tied up to financial data, the stock price data plays a huge role, which can provide a lot of information, however there are different ways on how to percieve this information. In connection to this master thesis's research question. The most suitable form of this variable will be as absolute values (market stock price). It is better to choose these values instead of variance of change because of the model (DiD), which is planned to be used.

4.10 Return on Invested Capital as model variable

Another interesting variable that can be applicable in the model is return on invested capital known as ROIC. Just like in the case of the stock prices, where it is more preferable to work with returns (preferably excess returns), the same trend can be seen in working with this variable. The ROIC ratio can be mostly used in company valuation. Due to the shift towards excess returns it is important to focus on measuring and forecasting returns obtained by business in two time frames. One of these frames is past investments and the other is expected future investments. [35]

4.10.1 Investment Returns and Accounting Returns

Returns on investments appear very frequently in the fields of finance and accounting, however there are different definitions.[35] In terms of accounting it is closely tied to the financial statements such as the balance sheet, income statements and cash flow statements. Balance sheet is the one that separates firms assets into two main categories Assets and Liabilities. The assets can be defined as a resource, which is controlled by an entity as

a result of past events, in addition to this it is also expected from these assets to bring future economic benefits.[36] As liabilities can be understood as a present obligation of the entity from past events, these are based on a settlement that results in an outflow from the entity's resources embodying economic benefits.[36] The sub-categories from these two main groups are estimated from the financial documents such as invoices, receipts, bank statements etc. In the end the values that can be seen as book values. In connection to this there are several ways how investments are measured from an accounting point of view. Due to the nature of the data, there are clear book values that are used for the calculation of financial ratios. In terms of investments ratios there are for example Return on Assets, Return on Equity and Return on Invested capital.[36] All of these ratios are based on the past data, which has been obtained during the fiscal year. That is one of the differences in comparison with the investment returns from the market perspective where the estimation is mostly based on expected profitability of an investment. There are often proxies and lagged values that are supposed to assume growth.[36] From this it is clear to assume that these are not as precise in comparison with the accounting book values, since these models are more based on forecasting. However both are relevant and lead towards answering the question: How good are the current and possible future investments in the company and if the returns that are produced by them exceed the cost of funding. [35]
The goal of return on invested capital is to measure the return, which has been earned on a capital invested in an investment. [35] The formula goes as follows. [37]

$$ReturnonInvestedCapital = \frac{NetIncome}{AverageTotalInvestedCapital} \tag{26}$$

where:

$$AVGTInvestedCapital = TShareholdersEquity + PreferredStocks + OperatingLeaseL \tag{27}$$

AVG Average
T Total
L Liabilities

Overall this financial ratio is supposed to provide information about all the investments that the company has in its books, therefore it can be used as a measurement that shows how well is the company managing new and old projects. Operating income of the most recent years can be a good proxy of the typical earnings on existing investments. The capital invested in these investments should be the book values. As operating income can be understood EBIT. [35]

4.10.2 Motivation to use 2

The motivation to use as variable the return on invested capital is mostly based on the information that can provide more insight into the potential development of a company. This development can be seen as potential new investments and how efficiently are these investments made. This does not only aim at forecasting towards the future but also the readjusted past projects. With more years of data it is possible to find some trends and see how different companies are performing, therefore providing details that can be used in the final conclusion, if proven significant.

4.11 Free Cash Flow margin as model variable

Cash flow is a widely used term and very important concept in the academic disciplines such as accounting, finance and economics and all business operations. It can be seen in a different perspective either as the level of absolute value or change in it. Cash flow itself can be a very valuable concept that provides information, which can be useful for insight into the performance of a firm's assets and the potential future direction. In comparison with the concept of net income, which is one of the most used ratios, that is closely watched by analysts, investors, media and all shareholders. The net income is reported on an income statement showing the result of an accrual basis in the accounting system, however it is not cash. The net income can be used in the analysis of potential future cash flows but still it cannot be used as cash, which is able to pay for wages, salaries, taxes, interests & dividends, services and debt. Inadequate cash can affect the overall liquidity of the company and its ability to pay for their payables, this increases the risk of default and could potentially lead to bankruptcy. Therefore, the cash flow models are widely used for valuation purposes regarding securities, mergers & acquisitions, capital assets and so on. [38]

4.11.1 Cash Flow statements

The cash flow statement is created out of four parts, that are describing different activities, which are tied to actions from the company during the fiscal year. These activities are:

Cash
Operating activities
Investing activities
Financing activities

Cash as previously mentioned can be seen as highly liquid short-term securities. That is why it is often referred to as cash & cash equivalents. Cash equivalents include treasury bills, certificates, bonds etc. Many companies are separating pure cash and these equivalents into two accounts. It is due to the nature of the activity, therefore, if these equivalents are separated from pure cash, they can be considered short-term investing activities. [39]
Operating activities are tied to the core business of the company. It includes producing

goods for sale, providing some sort of services and the cash effects of transactions that are tied to determination of income. [39] Investing activities can be separated into two groups. Firstly, it is acquiring and selling of securities as previously mentioned in the cash section and an asset group that produces/ is expected to produce benefits for the company in a long period. Secondly, it is tied to lending of money and collection of loans. [39] Financing activities are informing about the nature of the capital, therefore it shows borrowings from creditors and repayment of principal, in addition to this the financing activities include obtaining resources from owners and provision of investments that are owed to the owners. [39] Even Though, another very important step is to carefully select inflows and outflows in specific activities, as an example of such a selection of accounts can be seen in table 3. The inflows and outflows are changes that happened on specific accounts during a certain period of time. [39] All of these separations are needed to eventually reach the goal value of free cash flows, which is needed for additional calculations of the cash flow margin. [39]

Inflow	Outflow
-Asset Account	+Asset Account
+Liability Account	-Liability Account
+Equity Account	-Equity Account

Table 3: Inflows and Outflows

4.11.2 Free Cash Flow and FCF Margin

There are several ways to calculate free cash flow, the differences are mostly about what financial statements are used. There are varieties that are using balance sheets and income statements, on the other hand there are ones that are more connected with cash flow statements. [40]

One of keys for correct calculation of cash flow is the cash flow from operating activities, in future referred to as CFO [40], as previously defined what operating activities are. It is based on the core business of the company, that is why it is very relevant for these types of calculations.

$$FreeCashflow = CFO - CapitalExpenditures \tag{28}$$

The following formula is the free cash flow margin, it has been obtained from factset [37]. It modifies the value into percentages, which are more suitable for usage in statistical and financial models.

$$FreeCashFlowMargin = \frac{FreeCashFlow}{Sales} \tag{29}$$

4.11.3 Motivation to use 3

The motivation to use free cash flow margin as variable is due to the information that it provides about the current state of a company. It is closely tied to new and old investment, therefore it is providing relevant information in connection to our research question, which aims at the companies separated into two groups and their development regarding hydrogen future.

4.12 Market Capitalization as independent variable

Market capitalization was chosen as an independent variable to better explain our Y variable. This independent variable is considered to be relevant as it is the measure of total volume of outstanding shares of stock that a company has at any given period, this is obtained by multiplying the stock price at a given period by the total number of shares outstanding in the same given period. [40]

$$MarketCapitalization = S_o \cdot P_m \tag{30}$$

where:

S_o = Shares outstanding
P_m = market price of a share

By including the variable in our estimators for Difference in difference it is intended to help control for size bias of the firms being observed in our sample. Since it can be expected for larger companies to have greater impact on the overall market and sample, plus different characteristics can be present from smaller companies. Even though the sample that was gathered from the oil and energy industry for this study tries to get a balanced group of firms accounting for several aspects, one of them being size, by choosing this independent variable the differences between large and small companies in the analysis can be better accounted for. The inclusion of market capitalization in the estimators as a covariate will aid in controlling for potential confounding factors that can originate from company size which will factually help in better estimating the effect of clean hydrogen investments on the dependent variables surveid. Important to also note that other characteristics like financial resources, R&D capacity and market influence can be reflected in this variable, and since these characteristics can impact and influence how clean hydrogen business develops this variable is important to include in the analysis.

4.13 S&P 500 market index as independent value

Another choice of independent variable that was considered and inserted into the estimators in this paper is the S&P 500 market index, this financial market index can is normally considered by the research community and financial participants as a reference tracker for stock performance of the companies listed on the major stock exchanges in the United States, in other words it can be seen as the benchmark for companies in the stock market and their market value. By including this particular independent variable control for broader market trend is enabled, among other possible factors that can be present in the macroeconomic environment like inflation, economic growth and interest rates. Factors that impact performance of the firms analysed in this study will be captured by using this covariate in our Difference in difference estimator thus accounting for any potential bias that might originate from other market trends or influences. All in all both this independent variable and the former introduce market capitalization are intended to help our model with robustness and more accurate estimations.

4.14 Market Introduction

The energy and oil industry is a crucial component of any economy, as it contributes greatly to economic growth and contributes to the creation of jobs. Being such a driver of the economy of the USA this industry is also subject to several external factors that can impact the way it behaves and subsequently impact how firms within this branch behave and perform. By learning and understanding these key events we allow the reader to have a better understanding of possible data patterns that might come from the analysis in this paper.

2014

In 2014 the oil market experienced significant volatility regarding its prices, spiking up to 110 dollars a barrel and only to drop by 70 % of this value over the next few years. There are several reasons that could have caused it like the reduced demand for the commodity in the second half of 2014 from China and Europe could have driven it down while several Gulf states, Iraq and Saudi Arabia announced they were curtailing production. Nevertheless this impact can be seen in the industry and considered to have sufficient size to be relevant

2016-2018

In 2016 we see Trump [41] becoming the president and in general this had a positive impact for the market, specially regarding the oil and energy sector. The new office supported domestic energy production and eased regulations and supported domestic oil and gas production. In the same year we also saw the Organization of the Petroleum Exporting Countries (OPEC) openly deciding to cut back on production, intentionally, to raise prices. This cut figured a withhold of 1.2 million barrels per day in production, and curiously had support of other non-OPEC member countries.

Trade tension with China that started in 2018 also added up to the already existing geopolitical tensions with Iran, by imposing tariffs on Chinese goods, retaliatory tariffs were made against US energy exports which impacted the general demand for the US based commodities

2019

With the Covid-19 pandemic the global demand for energy commodities decreased and supply chain and production disruption. This had a strong impact in the US energy and oil industry and led to bankruptcies in the industry as oil prices reached an historical low in February 2020. In the same period US bond yields were at 1.66

4.15 Methodology plan for application

Research Question

The research, that is conducted in this master thesis, is based on what is the impact of hydrogen investments on the market value of companies and seek any other potential implications that can be extrapolated for the oil and energy industry and possibly other ventures.

Data collection

The collection started with selection of 18 companies from NYSE and NASDAQ. These companies then progressed to be evenly separated into two groups. For this selection it was mandatory to create criteria, which can be seen below.

1. Industry and Geolocation
2. Lifespan of Companies
3. Company Size
4. Factual Hydrogen Investments

Based on this information the companies were either put into the hydrogen group also known as the treatment group or non-hydrogen group which is considered the control group. The observed period ranges from 2010-1 to 2021-12. From these 12 years of monthly data the market price is obtained for each company. The monthly data for stock value was selected as it is believed that the more frequent observations, when possible, will contribute to a better Difference in difference model. The data regarding ROIC and FCFF were decided to feature yearly data instead, although not optimal it is the best available data form for these two factors, plus they were used for robustness testing, hence their criticality was evaluated and considered to be an ok measure to have this data in yearly steps. Therefore the total number of observations regarding the market value/ stock price is 2592, this means that the data set has a value for each company for each month during the full time period of twelve years, in addition to this and due to the nature of the data it can be said that the research will be mostly conducted from quantitative and empirical data while also referring to more qualitative aspects when necessary to make better sense of empirical analysis. Furthermore because of the characteristics of the data set, it is considered as panel data. This panel data, after the before mentioned approaches are done will be the final set on which the scientific methods will be performed. The main takeaway on why monthly data was preferred was because the monthly effect of the Paris agreement is to be analyzed on the data sample.

The main source for data collection is Factset, which provides relevant and trustworthy information that is necessary for precise analysis. Another used source is Yahoo finance that, given the problematic, helped with finding the companies which fit the criteria.

Methods

The methods that are used in this thesis derive from econometric research, precisely the

Difference in Differences model, which contributes with essential results for the analysed problematic of hydrogen impact on company's market value. As the treatment date is selected 12. December 2015 because of the Paris Agreement, which is expected to motivate a change in the market environment, especially in the industries on which this thesis focuses on.
To successfully perform the DiD it is important to implement the correct estimation method. These methods are based on the data set, which in this case is the panel data, therefore there are 3 tested estimation methods that are compared in this paper:

1. Pooled OLS
2. Fixed Effects - Within method
3. Random Effects

These are the tests that are used to analyse the performance of the model. The aim of the following test is to find the most suitable coefficient estimator, analyse the potential issues with heteroskedasticity, autocorrelation, normality and significance of the coefficients present in the regression model.

1. Hausman Test
2. Levin-Lin-Chu test
3. Breusch-Pagan Lagrange Multiplier
4. Breusch-Godfrey
5. Jarque Bera
6. Robust T and Robust F

To examine the significance of results, there are several robust tests to be performed. In the finance environment there are diverse economical variables, which can indicate changes in a company, therefore it is important to analyse more than just one variable. In this case there has been chosen 3 different economic variables as Y (explained variable) in the model to see how these variables perform and potentially if the measured results are comparable. Due to some differences in the selected properties it is necessary to add more explanatory variables, that improve the quality of the model:

1. Monthly returns of S&P500 during observed period
2. Market Capitalization

Last robust tests focus on the data set itself, where the tests will be aiming at the size bias. To test this, the model will be applied to the data set without 2 largest companies of each group. Another test will use the same course of action with only difference, which will be using the data set without 2 smallest companies of each group.

Tools

The tools that are used for the research are mostly R, which is a program that allows to perform statistical testing on which is this thesis based on. Another software that is used is

Microsoft excel, which was used for the data organization and data collection.

After analysing the results it is expected to come to conclusion, which helps to answer the initial research question on which this thesis focuses on.

5 Model Application

5.1 Company Selection

5.1.1 Criteria

When considering the study sample for our research there are several considerations to be taken. Since this pape figures the implementation of a Difference in differences model to analyse the outcome of a particular policy which in our case is the Paris agreement, it is crucial to establish clear criteria for sample group selection. By focusing on the quality of the sample this paper can contribute to a better and more precise modelling and outcomes. The main criteria used for selection are:

Industry and geolocation: The ensure the firms within our sample are comparable and to minimise the impact of external factors the sample selected accounted for the same location and sector. This compromise was realised by selecting only firms belonging to the American oil and energy industry and then was further improved by restricting the selection to companies that have been quoted in the New York Stock Exchange or the NASDAQ, since there are high standards for transparency and regulation in these platforms, additional security on our data sets is obtained like this.

Lifespan of companies: To guarantee sufficient observations and more viable modelling results, the selection of companies considered a long history in the oil and energy sector while being at the same time, publicly quoted. Selection was based on a minimum 11 years of existence in the market, 6 years before our policy event (The Paris Agreement) and 6 years after the policy event. This means that the data set has 6 years of pre-event data and 6 years of post-event data to consider in the analysis.

Afterwards more specific criteria is applied on a second phase to tawn out the original bigger data set obtained so far:

Company size: For comparability it is also important to group the companies by size, this was arguably the hardest step in the process, while there are absolute giants of firms entering the niche market and investing into green hydrogen to broaden their own portfolio, also it is found that smaller cap firms being successful in this field of green energy. Nevertheless, it was aimed to have both control and treatment group to have mainly medium cap firms.

Factual Hydrogen investments: The most important criteria in this study was dividing the firms into two groups. Since a control control group is needed, which figures companies that have not done anything for or engaged in clean hydrogen, and the treatment group needs to be only made up of companies that have factual investments in clean hydrogen. This will include green hydrogen production, production of any hydrogen related infrastructure, hydrogen fuel cell technologies, hydrogen refuelling, contribution to clean hydrogen production and R&D investments towards green hydrogen of at least $100 million.

5.1.2 Group Formation

After clearly defining our parameters for our selection process identification of the companies that fit our desired criteria to be implemented on the Difference in difference model analysis to see the impacts of clean hydrogen investments on market value and other financial ratios. Division of all the companies into two evenly distributed groups was performed. One is the Hydrogen Group which is to be seen as the treatment group regarding the DiD model and the other group is the non hydrogen or the Brown group and this can be considered as the control group for DiD model purposes.

Companies	Market Cap	Active in Green Hydrogen	Group
APD-US	63,562,346,808$	Yes	Hydrogen
CVX-US	328,786,864,560$	Yes	Hydrogen
ED-US	33,493,606,504$	Yes	Hydrogen
EIX-US	27,399,955,275$	No	Non Hydrogen
EQT-US	11,942,333,714$	No	Non Hydrogen
FCEL-US	900,725,262$	Yes	Hydrogen
GE-US	104,056,600,000$	Yes	Hydrogen
HAL-US	30,363,830,859$	No	Non Hydrogen
HP-US	3,891,135,000$	No	Non Hydrogen
MRO-US	16,201,002,566$	No	Non Hydrogen
MUR-US	6,103,431,800$	No	Non Hydrogen
NFG-US	5,150,621,938$	No	Non Hydrogen
OKE-US	29,991,890,000$	No	Non Hydrogen
OXY-US	57,910,461,648$	No	Non Hydrogen
PLUG-US	5,364,285,375$	Yes	Hydrogen
SLB-US	74,531,466,240$	Yes	Hydrogen
TS-US	17,494,000,000$	Yes	Hydrogen
XOM-US	472,437,800,000$	Yes	Hydrogen

Table 4: Table of companies

Both groups consist of 9 firms which follow the before mentioned criterias and are ready to have the Difference in differences model be applied to their data and findings analysed,this will generate 2592 observations across time for us. Initially the group was larger and contained 24 companies which were intended to be divided into 2 groups of 12, following the same initial criteria. On a second stage analysis where the hydrogen factors were double checked it was found that 3 of the initial companies present in the hydrogen group did not fill all expectations in a satisfactory manner, and of course were removed. This would also make it a necessity to remove the same number of members from the control group, for this the companies that were closest in terms of size were chosen. The initial list of 24 companies and subsequent triage looked like the following

Ticker	Company	Initial Triage
APD-US	AIR PRODUCERS AND CHEMICALS, INC.	Kept
CVX-US	CHEVRON CORPORATION	Kept
ED-US	CONSOLIDATED EDISON, INC.	Kept
EIX-US	EDISON INTERNATIONAL	Kept
EQT-US	EQT CORPORATION	Kept
FCEL-US	FUELCELL ENERGY, INC.	Kept
GE-US	GENERAL ELECTRIC COMPANY	Kept
HAL-US	HALLIBURTON COMPANY	Kept
HP-US	HELMERICH & PAYNE, INC.	Kept
MRO-US	MARATHON OIL CORPORATION	Kept
MUR-US	MURPHY OIL CORPORATION	Kept
NFG-US	NATIONAL FUEL GAS COMPANY	Kept
OKE-US	ONEOK, INC.	Kept
OXY-US	OCCIDENTAL PETROLEUM	Kept
PLUG-US	PLUG POWER INC.	Kept
SLB-US	SCHLUMBERGER N.V.	Kept
TS-US	TENARIS SA	Kept
XOM-US	EXXON MOBIL CORPORATION	Kept
ALB-US	ALBEMARLE CORPORATION	Rejected
CTRA-US	COTERRA ENERGY INC.	Rejected
CVI-US	CVR ENERGY, INC.	Rejected
GLNG-US	GOLAR LNG LIMITED	Rejected
RRC-US	RANGE RESOURCES CORPORATION	Rejected
RIG-US	TRANSOCEAN LTD	Rejected

Table 5: Table of all potential companies

5.2 Influence of the market

While clean hydrogen related investments could have the potential to influence the market positively by reducing the necessity of fossil fuels and reducing emissions, other external factors need to be taken into account to justify what the data is telling us. Factors such changes in regulations, cuts in productions or geopolitical tensions can have a bigger impact in the treatment group's performance.

5.2.1 Decision process of treatment

The event that delimits the Difference in difference "treatment" period in this thesis is the Paris Agreement. This international agreement was a landmark in securing a better carbon free future for the globe, and was initially signed by 196 countries on the twelfth of December of 2015. By requesting countries to determine their own objectives to reduce greenhouse gas emissions and reviewing them every 65 years, this treaty aims to limit the rise of global temperatures. By being an event that signals global efforts to address climate change it can be seen as the catalyst for green energy investments, including clean hydrogen technologies. Hydrogen related technologies play a significant role in the low-carbon transition because they offer a clean and efficient solution from fossil fuels for transportation and power grid requirements. Using this date as the event in our model it is possible able to evaluate the effectiveness of the clean hydrogen investments for companies in the oil and energy industry in the USA when it comes to company value, as well as other other factors, and in the end they will be tied to see if there is a direct influence of adherence to the objectives stated in this treaty and if they influence how the market sees firms that do so, mainly through their market value in the years following the treaty

5.3 Model Creation

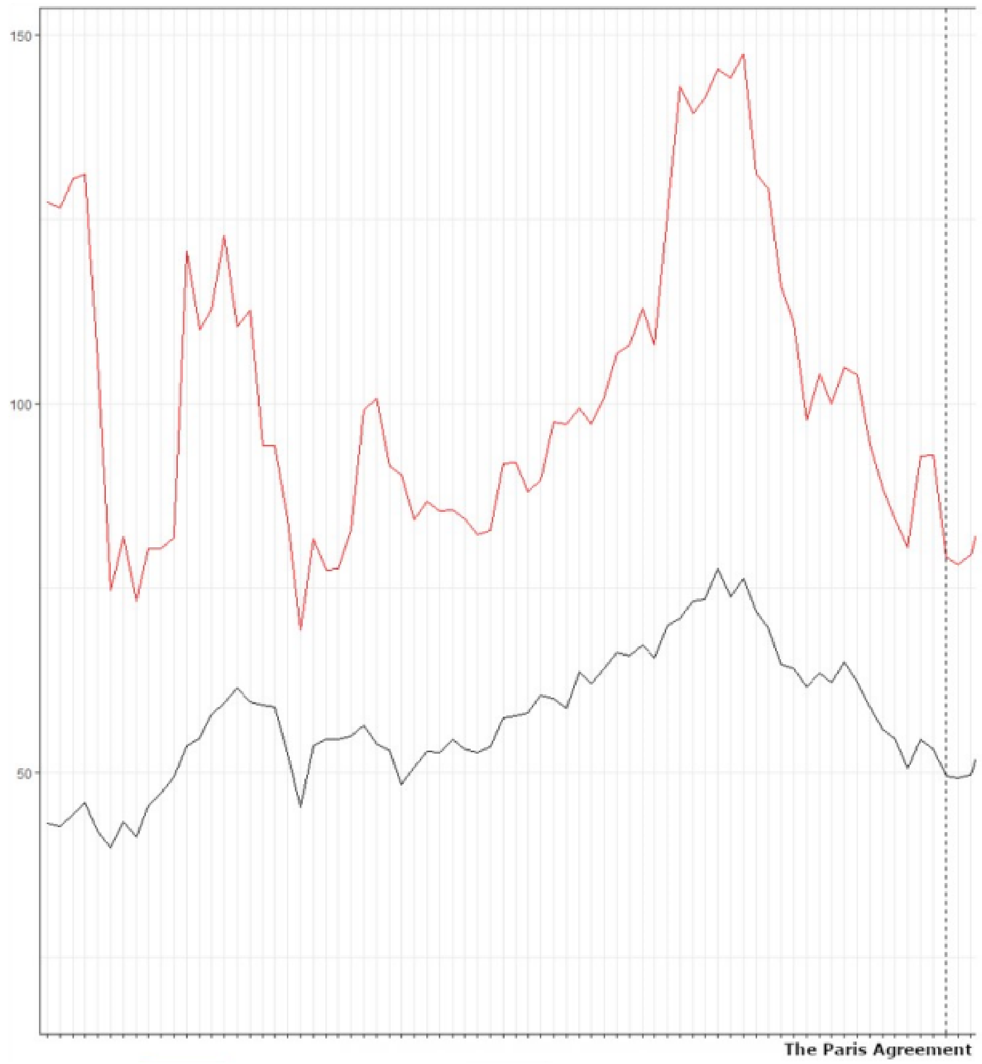
5.3.1 Optimal Model Selection

Pre-Model Selection

In the process of selection of the best model this research took into all possible and reasonable considerations given our specific case and data. As formerly introduced throughout this paper the reader will be looking at a sample group of companies within the oil and energy sector in the USA, from 2010 through 2021. Using a Difference in difference approach it is aimed to see if our observed event which is the Paris agreement had any significant impact on market value and other financial factors of these companies which are evenly divided into two groups as already explained. Before testing the models, a balanced panel data set is found, in which then the binary variables are added to make the data sample apt for DiD analysis. In this paper 3 dummy variables are created. A variable to identify the control and treatment groups, given our previous selection, another variable to flag data that is pre-event and post-event, in this case the event trigger is use, therefore is used 12/12/2015 as the date, and our final variable is the interaction between the two aforementioned variables.

Afterwards statistical tests test the data for stationarity as this characteristic is an important assumption in many econometric models that figure time series or panel data models. While conducting the Levin Lin Chu test our results shy away from desired 5% significance which might have some implications to our model, in this case considerations were made for altering the data so it would fit stationarity. After consideration it was concluded that this action would essentially change to data and how the results could be interpreted, which would mean the general results could not be good to be related to external events, hence it was taken into consideration two important assumptions of Difference in difference models. These assumptions are present in our model and this paper will later show proof and arguments for this, when the assumption of parallel trends and independence of treatment assignment are met then estimates of the treatment effect can be obtained and be unbiased even in the presence of non-stationarity. Of course this was decided as the p-value obtained was close enough to the significance level plus other accessory tests like the Augmented Dick Fuller test provided a significance of 1%.

Independence of treatment states that the treatment is to not be influenced by any other factor that may affect the outcome variable. When looking at the Paris agreement it can be ascertained with confidence that this event was not influenced by any factor that can be directly related to the market value or other financial ratios of the companies in the testing groups or their clean hydrogen investments. Given this it can be said that there is independence of treatment. Regarding the parallel trends assumption needed in Difference in difference models, this assumption states that the control and event groups should have followed a similar trend over time in the absence of the treatment, while this can not be known with certainty after the event takes place, since it never happened, it can be seen in the data before the event if both groups have followed a similar trend, which in our case the depends presents the following trends pre-event:



Red Line Market Value - Clean Hydrogen Group
Black Line Market Value - No Hydrogen Control Group

Figure 5.1: Paris Agreement

While we can see there are different proportions in the volumes in change from period to period it is quite clear that both groups followed the same trends, increasing and decreasing in the same periods and experiencing spikes and crashes in the same moments.

Decision of Model

Since our Difference in difference model is using panel data we needed to make use of panel data estimators. Our decision on what model to use was based on both statistical tests and qualitative decision making. Initially we generated our Difference in differences using the Fixed Effect estimator, a Random Effect estimator and a pooled OLS.

$$y_{it} = \beta_0 + \beta_1 Hydrogen_{it} + \beta_2 time_{it} + \beta_3 Hydrogen_{it} * time_{it} + \beta_4 X_{it1} + \beta_5 X_{it2} + u_{it} \tag{31}$$

where:
y= Market Price
$Hydrogen_{it}$ = Initial difference
$Time_{it}$ = Baseline change over time
$Hydrogen_{it} \cdot time_{it}$ = Treatment effect
X_{it1} = Market capitalization
X_{it2} = SP500 index values
u = Error term

It was conducted a Breush-Pagan Lagrange Multiplier test to verify if our models have any panel effect, result does not reject the null and this proves that since there is no evidence of significant unobserved effects in the individual level in the model, the OLS method is not a good choice. Afterwards it is necessary to have to test the Fixed effects model with the Random effects model, several statistical tests are conducted to verify what is the best model, where an F test for individual effects and the Hausman test are done. The F test for individual effects tries to see whether there are significant differences in the intercepts across the all individual units in our panel data and it. This test states that a Fixed effects model is to be taken over a Random effects or OLS models if the null hypothesis is not rejected, in our case, and with a p-value of 0.9 there is the indication that there are no significant difference in the intercepts across individual firms in the model and this mean that a pooled model or a Random effect model are more appropriate. Given that the findings on the Bresch-Pagan test above state that OLS is not a good fit, it is taken into consideration that the Random effects will be more appropriate than the Fixed effects. The Hausman test was also conducted to compare the efficiency of these two estimators. The null hypothesis states that both estimators are consistent and asymptotically efficient while the alternative hypothesis points that, at least on estimator is inconsistent and inefficient, by arriving at a p-value of 1 both Random effects and Fixed effects estimators are seen as good enough estimators for our Difference in difference model using panel data as they should produce results without any bias concerns.

At this point the statistical study sees a clear inclination towards Random effects, but this decision has to be backed by more data specific facts. Since the object of this thesis is subject to several unobserved confounding factors, like for example the changes in global fuel prices, political regulations, natural disaster or technological advancements that may affect the outcome variable and the event variable (treatment variable in the DiD context) that may lead to biassed estimates if they are not correctly accounted for in the chosen model. Using a Random Effects regression can help with this by assuming that there are individual specific effects that are unobserved but constant over time, in other words, this means that the unobserved confounding factors are captured by the individual specific effects, which it is, of course, consider to be important. Another important concept to take into consideration is unobserved heterogeneity in both time variant and time invariant forms, which in our case will be the variation of the event effect across all individual units in our study. By taking this into account the Random effects model allows for firm specific intercepts to vary randomly, thus capturing any unobserved factors that individually may affect each firm. Although, as initially intended a data set was created with units that already have a high standard of comparability there are other company specific factors that can influence the outcome variable in our data, time variant factors such as difference in investment/management style, access to resources or even technological expertise within the workforce and time invariant factors such as cultural factors or company background. By using the Random effects estimator it is more feasible to account for all these factor that may reduce efficient estimation of results and this is why, after taking all this into consideration, why the Random effects estimator for the Difference in difference model is considered the best fit for our particular data and case.

5.3.2 Stock Prices as dependent variable

The dependent variable of stock price is chosen as the variable to be explained by our Difference in difference random effect estimator with two independent variables. This is directly influenced by our research question "Do Hydrogen investments cause company market value to increase?", since the stock price of a certain company is in fact the market value of said company. The model will have market capitalization and the market index in which the firms are part of used as independent variables to control for size and market bias, as previously discussed. In terms of statistical relevance the following levels were obtained:

The p-value obtained in the T test of coefficient show if there is statistical relevance to each independent variable. Where moderate level of significance is seen as p-value ¡0.05, high level is p-value < 0.01 and high levels of significance are where the p-value < 0.001. Based on the results here shown, it can be seen that the Intercept, Hydrogen, Market Capitalization and the interaction between Hydrogen and Time are highly significant predictors of the dependent variable here considered, as we have very small p-values. Important to

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	51.0	1.47	34.7139	<0	***
Hydrogen	27.7	2.60	10.6496	<0	***
Time	1.19	1.78	0.6696	0.5032	
MarketCap	$1.71 \cdot 10^{-4}$	$1.18 \cdot 10^{-5}$	14.4789	<0	***
Index	$4.78 \cdot 10^{-4}$	$9.08 \cdot 10^{-4}$	0.526	0.5989	
Hydrogen:time	-47.9	2.72	-17.5667	<0	***

Table 6: Stock Prices Model - significance of variables

note that the coefficient in the interaction between Hydrogen and Time is negative, which means that with time hydrogen will have a negative development, this will be important when going forward with the analysis of results. Regarding the statistical tests necessary to verify several aspect of the model, the following results are found:

Tests	p-Value
Jarque Bera Test	$2.2 \cdot 10^{-16}$
Breusch Pagan Test	$2.2 \cdot 10^{-16}$
Wald Test	$2.2 \cdot 10^{-16}$
Breusch Godfrey or Wooldridge Test	$4.39 \cdot 10^{-14}$

Table 7: Stock Prices Model - Test Results

The Jarque Bera test, a test for goodness of fit that is used to determine if the data set follows a normal distribution, this is applied to the model residuals and it is found that it does not follow a normal distribution, this phenomena is in fact common when looking at financial data. The Breusch-Pagan test is testing for heteroscedasticity, the p-value suggest there is strong evidence of heteroscedasticity, which indicates the variance of the dependent variable is not constant across independent variable, this is as expected results, since stock price are being look at when there are events that vary across time, like market fluctuations, it is expected to see heteroscedasticity. The Wald test comes over to indicate if the parameters taken in the model are statistically significant, with the p-value obtained there is an added assurance on the parameters of the model being statistically significant. Finally the Breush-Godfrey/Wooldridge test will indicate if there is autocorrelation in the relation model, similar to the Breush-Pagan test, the p-value obtained suggests that there is evidence of autocorrelation in the model, this can be attributed to the fact that the data is grouped into 2 groups, since it's a Difference in difference model, then it can be expected that the errors in the regression model applied in the paper to be correlated within each group.

Overall, while initially it seems that the tests suggest some issues, they can be all seen as consequence of the particular assumptions and limitations of the statistical methods applied in this paper, furthermore since the parameters in this model can be considered statistically significant as seen in the before presented tests, the model can in fact efficiently explain the variation found in the dependent variable.

For the results of the Difference in difference model, the following values were obtained:

Variables	Market Value	Signif
Hydrogen	27.727	***
Time	1.192	
MarketCap	0.0002	***
Index	0.0005	
Hydrogen:Time	-47.860	***
Constant	50.974	***
Observations	2592	
R^2	0.171	
$Adjusted\ R^2$	0.169	
F Statistic	532.184	***

Table 8: DiD Model - Test Results

Now, from this table several observations can be made. There is a statistically relevant positive coefficient for Hydrogen, at 27.72, which suggests that the companies involved with clean hydrogen experience a higher market values than their non-clean hydrogen counterparts, this results can be explained by a growing interest in the market to have clean hydrogen used as a energy source where companies that do get involved in this industry benefit from government/fund related incentives. Looking at Market capitalization a positive and statistically relevant coefficient is found, this indicates that larger companies tend to have higher market values, in fact, this finding does not come as groundbreaking as it is quite expected to find these two topics closely related. When looking at the independent variable Index, which would consider the broader market using the S&P 500 index, it is seen a positive but non significant coefficient, this suggests that after controlling for the other variables in the model that the broader market trends do not influence the market value of the firms in the sample and that in fact that the market value is more drive by unit specific factors. This result also matches what was expected, since market value should be more influenced by unit specific factors, such as company size, investment in particular industries, etc… The R squared value of 0.17 could be considered small for a normal linear regression model, but accounting that panel data was used a value rounding 0.2 is considered acceptable, of course this value does not necessarily mean the model is of lower quality given other metrics already shown in this paper, this R2 is considered up to expectations. More relevant to the Difference in difference estimation is the interaction between Hydrogen and time, the results show a negative coefficient which is highly statistically significant and this suggests that the effect of clean hydrogen investments weaken or become negative over time. This can be due to several different factors, which will be discussed further. The visual representation of this Difference in difference model is as follows:

This representation has the years in the X axis and the mean of observations of the dependent variable for each period in the Y axis. The red line represents the clean hydrogen group and the black line the non clean hydrogen control group, a dashed line is seen in

Red Line Market Value - Clean Hydrogen Group
Black Line Market Value - No Hydrogen Control Group

Figure 5.2: Main DiD graph

the middle, and this represents the DiD event. This Difference in difference graph shows several indications that already were described previously, like the trend assumption before the event (treatment), which is one of the assumptions of DiD. Most importantly it is seen the relation between hydrogen and time, in fact after the event it is seen a reaction on the market price of the companies, although not an instant reaction, which makes sense since investments and corporate police changes take time to react to market environments. Afterwards from middle 2016 onwards a down trend can be seen, as mentioned before this suggests that clean hydrogen investments may decrease market value. This can be tied to other events that directly impact clean hydrogen investments and after 2016 an effect is verified that clearly has an impact. This could be because of a more crowded market in concern to hydrogen related industries or even government policies, of course this possibilities will have a more in depth observation in the upcoming conclusion section.

5.4 Robustness testing

In the section introduced here robustness tests will be conducted, to show how results and model efficiency would change if other variables and scenarios were considered. While robustness tests should not be seen as substitutes for careful consideration of the theoretical framework of the data used in any model and other tests like the ones explained in this paper, these tests do add value to the overall view on the methods applied.

5.4.1 Other dependent variables

Different dependent variables were considered, in attempts to better measure if the model's results can be considered consistent when using different outcome variables. If the coefficients of the independent variables remain stable and statistically significant that will be an indication of model robustness. The same model was used considering ROIC (Return on Invested Capital) as a dependent variable, yearly data, general tests reported similar performance and the significance tests reported the following:

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	1.3566	2.7782	0.4883	0.625819	
Hydrogen	-38.623	5.6364	-6.8523	$7.923 \cdot 10^{-11}$	***
Time	-7.3264	3.4638	-2.1151	0.035596	*
MarketCap	$1.3228 \cdot 10^{-4}$	$1.6715 \cdot 10^{-5}$	7.9138	$1.416 \cdot 10^{-13}$	***
Index	$8.526 \cdot 10^{-4}$	$1.6134 \cdot 10^{-3}$	0.5285	0.597722	
Hydrogen:time	15.056	5.5896	2.6937	0.007638	**

Table 9: ROIC - T test

While avoiding diving into deeper details here we can see there is still statistical significance in the model for this dependent variable. Another dependent variable considered was FCFM (Free Cash Flow Margin), yearly data, general diagnostic tests showed similar results also and the significance results are as follows:

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	-11.634	3.3985	-3.4232	0.0007439	***
Hydrogen	-14.323	5.7019	-2.5119	0.0127592	*
Time	1.3301	4.9154	0.2706	0.7869626	
MarketCap	$1.216 \cdot 10^{-4}$	$1.6253 \cdot 10^{-5}$	7.4817	$1.977 \cdot 10^{-12}$	***
Index	$2.250 \cdot 10^{-3}$	$1.5516 \cdot 10^{-3}$	1.4504	0.1484281	
Hydrogen:time	-4.5516	6.5214	-0.6979	0.4859843	

Table 10: FCFM - T test

Here it is seen that there is still statistical significance in independent variables, although different ones. Of course for different dependent variables other considerations will be taken into account, but the main takeaway point for this analysis is that the model indicates good robustness when tested with different dependent variables.

5.4.2 Accounting for size bias

Although size bias was accounted for in the original model with the inclusion of Market capitalization as an independent variable there was the decision to test the robustness of this initial consideration. For this, the original model was still conducted as initially presented but on a first run the 2 smallest companies from both groups are removed and on a second run the 2 biggest companies from both groups are removed. Then both are compared with the original results Regarding significance of Independent variables, for the first run where the 2 smallest companies from each group are removed t test provided the following:

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	48.580	1.4703	33.0414	<0	***
Hydrogen	4.3859	1.5091	2.9064	0.003696	**
Time	2.3381	1.4099	1.6583	0.097409	
MarketCap	$1.966 \cdot 10^{-4}$	$7.682 \cdot 10^{-6}$	25.5971	< 0	***
Index	$1.7389 \cdot 10^{-3}$	$8.0311 \cdot 10^{-4}$	2.1652	0.030493	*
Hydrogen:time	-17.790	2.3234	-7.6569	$2.939 \cdot 10^{-14}$	***

Table 11: Smallest companies removed - T test

While t tests for the second run where the biggest 2 companies from each group are removed shows:

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	40.873	1.1966	34.1565	<0	***
Hydrogen	25.115	4.1310	6.0798	$1.436 \cdot 10^{-9}$	***
Time	1.5206	1.5411	0.9867	0.323933	
MarketCap	$5.322 \cdot 10^{-4}$	$4.6025 \cdot 10^{-5}$	11.5632	< 0	***
Index	$2.4565 \cdot 10^{-3}$	$7.7601 \cdot 10^{-4}$	3.1656	0.001571	**
Hydrogen:time	-54.580	3.3623	-16.2330	< 0	***

Table 12: Largest companies removed - T test

Although both results prove to have smaller changes in level of significance it is still seen that the same independent variables remain significant in both scenarios when comparing with the original data sample. Since the results are still close and do not vary critically from the original model, we can conclude that there was very little to none size bias. One other interesting view to consider is how the visual representations of each model look.

Red Line Market Value - Clean Hydrogen Group
Black Line Market Value - No Hydrogen Control Group

Figure 5.3: DiD model size bias graph run 1

Visual representation of the Difference in difference graph of the first run.

Red Line Market Value - Clean Hydrogen Group
Black Line Market Value - No Hydrogen Control Group

Figure 5.4: DiD model size bias graph run 2

Visual representation of the Difference in difference graph of the second run.

It is seen higher level and faster responses to the market on the second run where the biggest companies are removed, while on the first run where the smallest companies are removed from both groups a more stable graph is seen, with less harsh increases and decreases. It's important to mention that all the graphs still follow the same trend, although on the second run scenario that does not have the biggest companies from both groups these fluctuations seem more accentuated, this can be attributed to having a sample with more companies that participate and engage more exclusively with clean hydrogen. By removing bigger and more established players that also develop clean hydrogen technologies it can be seen that the clean hydrogen group is more susceptible to market movements. Regarding the control group, no big change is noticed and this is what would be expected from a control group.

5.4.3 Non inclusion of independent variables

The final robustness test considered was the non-inclusion of independent variables, by doing this and analysing how the coefficients react to the remaining independent variables robustness of results can be taken into consideration. By removing Market capitalization and the Market Index as independent variables the sensitivity of the model to these independent variables can be ascertained, after removing the variables and running the tests, diagnostic tests proved similar to the original model and the following coefficient t tests were obtained:

Variables	Estimate	Std. Error	t value	Pr < *abs t*	Signif
Intercept	57.4883	1.0503	54.7356	<0	***
Hydrogen	44.24884	1.9066	23.2084	<0	***
Time	1.7189	1.2460	1.3795	0.1679	
Hydrogen:time	-53.5178	2.9020	-18.4419	<0	***

Table 13: Independent variables excluded - T Test

Here it can be seen that the model still proves statistical significance, but by doing this the bias accounted for previously with the inclusion of the two removed independent variables are not considered, which in turn indicate a worse model in terms of explanatory capability. This can be confirmed when looking at the Difference in difference results:

Variables	Market Value	Signif
Hydrogen	44.248	***
Time	1.719	
Hydrogen:Time	-53.518	***
Constant	57.488	***
Observations	2592	
R^2	0.096	
$Adjusted\ R^2$	0.095	
F Statistic	273.815	***

Table 14: DiD Model (independent variables excluded) - Test Results

With a lower R squared and a lower F statistic in comparison with the original model, the assumption that the two dependent variables considered initially to be a good decision is verified. Visually the model without the dependent variables looks almost the same, and it does not add value to this paper to be shared. In conclusion the robustness of the model when removing the dependent variables is verified by these findings.

6 Discussion

Looking at the entire scientific research executed in this paper and conclusion, the main findings indicate that companies which invested in projects related to clean hydrogen did not see their market value increase due to these ventures. The Difference in difference model allows to, in a simplified way, to check if events have a positive or direct impact in certain factors present in individuals, this analysis does not support the theory that market value would increase by taking part into the clean hydrogen market, and in fact shows the opposite where market value decreased after the event in comparison to the control group. The answer to the thesis question in this paper: "Do hydrogen investments cause company market value to increase in comparison with it's peers? " is that market value in fact devalues when looking at clean hydrogen related projects. The findings that were reached are themselves the first in the field, since there is no previous paper that discusses similar implications. Other papers are found discussing related topics to this thesis like the Hydrogen report 2019 from the IEA but not in depth like found here. These implications imply that partaking into clean hydrogen ventures might not be the best decision in the short term of under 10 years. There is no data for the long term as not enough time has elapsed from the decision to take part of the green transition by several nations. This makes sense, as clean hydrogen technologies, although they see the biggest percentage of investment and promise are still a prototyping technology that is chasing more development everyday by several researchers which in turn means that from institutional investors point of view or long term investors companies associated with somewhat unproven technology become out of scope. Hence why support schemes and hydrogen specific investment funds are crucial for the success of this potential new renewable, akin to other technologies like wind or solar. These findings can imply that (i)Support schemes will be necessary to have corporate investors contribute to the clean hydrogen industry and (ii)Lower equity investments should be expected when entering these clean hydrogen ventures, at least for the long term. The data analyzed and the findings of course come with limitations, since the study group is exclusively based in the USA, these results might not be suitable for other economical areas around the world although could be used as a proxy with some special assumptions considered. Another limiting factor is the time elapsed from the Paris agreement, as of now, the treaty was signed 7 years ago, while its deadlines target 2050 as the objective and this means that data available regarding companies reacting and adjusting their business to it might be scarce or simply not available yet because it's too soon. Scientific investigations regarding renewable energies that will enable a better and faster green transition will be crucial in the coming years, papers that collectively point out the profitability, scalability and efficiency of new technologies, like clean hydrogen will help push the green transition forward. Specially, on the case of this paper, being able to execute a similar study in 5 or 10 years from now, might showcase very interesting findings that could point out to other necessary directions to better capitalize clean hydrogen investments

7 Conclusions

In summary it is concluded that the findings in this paper concerning the impact of clean hydrogen investments in firm market value, contribute to the ongoing debate regarding the efficiency of clean hydrogen investments in the energy sector when considering their added value to the market players that follow this type of investment path. What was observed, at least in the short term, is that the clean hydrogen investments do not mean an increase in market value. To verify this in the long term further research is needed and potentially other new findings will be present as further into the future these investments go. Several can be the causes that can suggest the way the market value of companies decreased, while initially it could be considered that a niche market which is involved in one of the biggest topics of this decade could be a good way to improve performance in the market, but it seems it's not the case. In current days clean hydrogen is a relatively new and still maturing technology and because of this investors will naturally perceive companies which participate in projects related to it to be more risky than others more in tune with the market counterparts. The high costs associated with the research and development as well as the initial investments of clean hydrogen project will naturally push market value down since it will translate into decreased short term profitability, and of course, there might have been external factors that might also impact but the findings in this thesis are clear and significant, in the short run the clean hydrogen investments are not contributing positively to the market value of companies. It is intended that this study will bring relevant and useful insights concerning the relationship between clean hydrogen and the market value of companies, while companies should consider this caliber of business expansion carefully other players such as policymakers need to consider this findings when developing laws and policies to help and support the green energy transition with the aid of green hydrogen.

References

[1] I. (2022). "Renewables 2022." (), [Online]. Available: https://www.iea.org/reports/renewables-2022. (accessed: 22.05.2023).

[2] B. Zhang, S.-X. Zhang, R. Yao, Y.-H. Wu, and J.-S. Qiu, "Progress and prospects of hydrogen production: Opportunities and challenges," *Journal of Electronic Science and Technology*, p. 100 080, Jan. 2021. doi: 10.1016/j.jnlest.2021.100080.

[3] M. Wang, G. Wang, Z. Sun, Y. Zhang, and D. Xu, "Review of renewable energy-based hydrogen production processes for sustainable energy innovation," *Global Energy Interconnection*, vol. 2, no. 5, pp. 436–443, Oct. 2019. doi: 10.1016/j.gloei.2019.11.019. [Online]. Available: https://doi.org/10.1016/j.gloei.2019.11.019.

[4] I. (2019). "The future of hydrogen." (), [Online]. Available: https://www.iea.org/reports/the-future-of-hydrogen. (accessed: 22.05.2023).

[5] A. van Benthem, G. Kramer, and R. Ramer, "An options approach to investment in a hydrogen infrastructure," *Energy Policy*, vol. 34, no. 17, pp. 2949–2963, Nov. 2006. doi: 10.1016/j.enpol.2005.05.006. [Online]. Available: https://doi.org/10.1016/j.enpol.2005.05.006.

[6] L. Zhu, L. Hu, S. Yüksel, H. Dinçer, H. Karakuş, and G. G. Ubay, "Analysis of strategic directions in sustainable hydrogen investment decisions," *Sustainability*, vol. 12, no. 11, p. 4581, Jun. 2020. doi: 10.3390/su12114581. [Online]. Available: https://doi.org/10.3390/su12114581.

[7] Y. Zhao, Y. Xu, S. Yüksel, H. Dinçer, and G. G. Ubay, "Hybrid IT2 fuzzy modelling with alpha cuts for hydrogen energy investments," *International Journal of Hydrogen Energy*, vol. 46, no. 13, pp. 8835–8851, Feb. 2021. doi: 10.1016/j.ijhydene.2020.12.211. [Online]. Available: https://doi.org/10.1016/j.ijhydene.2020.12.211.

[8] U. D. of Energy. "Hydrogen fueling infrastructure development." (), [Online]. Available: https://afdc.energy.gov/fuels/hydrogen_infrastructure.html. (accessed: 22.05.2023).

[9] A. Goodman-Bacon, "Difference-in-differences with variation in treatment timing," Tech. Rep., Sep. 2018. doi: 10.3386/w25018. [Online]. Available: https://doi.org/10.3386/w25018.

[10] S. Athey and G. W. Imbens, "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, vol. 74, no. 2, pp. 431–497, Mar. 2006. doi: 10.1111/j.1468-0262.2006.00668.x. [Online]. Available: https://doi.org/10.1111/j.1468-0262.2006.00668.x.

[11] C. H. Li and S. T. Certo, "The application of the difference-in-differences model in strategic management research," *Academy of Management Proceedings*, vol. 2021, no. 1, p. 13 503, Aug. 2021. doi: 10.5465/ambpp.2021.169. [Online]. Available: https://doi.org/10.5465/ambpp.2021.169.

[12] M. Momirlan and T. Veziroglu, "Current status of hydrogen energy," *Renewable and Sustainable Energy Reviews*, vol. 6, no. 1-2, pp. 141–179, Jan. 2002. DOI: 10.1016/s1364-0321(02)00004-7. [Online]. Available: https://doi.org/10.1016/s1364-0321(02)00004-7.

[13] C. M. Kalamaras and A. M. Efstathiou, "Hydrogen production technologies: Current state and future developments," *Conference Papers in Energy*, vol. 2013, pp. 1–9, Jun. 2013. DOI: 10.1155/2013/690627. [Online]. Available: https://doi.org/10.1155/2013/690627.

[14] J. Wooldridge, *Introductory Econometrics*, 7th ed. Florence, AL: South-Western College Publishing, Jan. 2019, ch. 1.

[15] B. H. Baltagi, "The one-way error component regression model," in *Econometric Analysis of Panel Data*. Cham: Springer International Publishing, 2021, ch. 1, ISBN: 978-3-030-53953-5. DOI: 10.1007/978-3-030-53953-5_2. [Online]. Available: https://doi.org/10.1007/978-3-030-53953-5_2.

[16] B. H. Baltagi, "The one-way error component regression model," in *Econometric Analysis of Panel Data*. Cham: Springer International Publishing, 2021, ch. 8-9, ISBN: 978-3-030-53953-5. DOI: 10.1007/978-3-030-53953-5_2. [Online]. Available: https://doi.org/10.1007/978-3-030-53953-5_2.

[17] A. C. Cameron and P. K. Trivedi, "Core methods," in *Microeconometrics*, Cambridge: Cambridge University Press, May 2005, ch. 21.

[18] J. Wooldridge, *Introductory Econometrics*, 7th ed. Florence, AL: South-Western College Publishing, Jan. 2019, ch. 10.

[19] B. H. Baltagi, "The one-way error component regression model," in *Econometric Analysis of Panel Data*. Cham: Springer International Publishing, 2021, ch. 2, ISBN: 978-3-030-53953-5. DOI: 10.1007/978-3-030-53953-5_2. [Online]. Available: https://doi.org/10.1007/978-3-030-53953-5_2.

[20] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[21] M. Arellano, *Panel Data Econometrics* (Advanced Texts in Econometrics), en. London, England: Oxford University Press, Jun. 2003.

[22] J. M. Wooldridge, *Econometric analysis of cross section and panel data*, en. London, England: MIT Press, Jan. 2002.

[23] W. H. Greene, *Econometric analysis*, 5th ed. Upper Saddle River, NJ: Pearson, Aug. 2002.

[24] P. Bharadwaj, *Quasi experimental method difference in differences*, Mar. 2010.

[25] S. O'Neill, N. Kreif, R. Grieve, M. Sutton, and J. S. Sekhon, "Estimating causal effects: Considering three alternatives to difference-in-differences estimation," *Health Services and Outcomes Research Methodology*, vol. 16, no. 1-2, pp. 1–21, May 2016. DOI: 10.1007/s10742-016-0146-8. [Online]. Available: https://doi.org/10.1007/s10742-016-0146-8.

[26] C. Kleiber and A. Zeileis, *Applied Econometrics with R* (Use R!), en, 2008th ed. New York, NY: Springer, Aug. 2008.

[27] A. Levin, C.-F. Lin, and C.-S. James Chu, "Unit root tests in panel data: Asymptotic and finite-sample properties," *Journal of Econometrics*, vol. 108, no. 1, pp. 1–24, 2002, ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(01)00098-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304407601000987.

[28] B. H. Baltagi, "The one-way error component regression model," in *Econometric Analysis of Panel Data*. Cham: Springer International Publishing, 2021, ch. 4, ISBN: 978-3-030-53953-5. DOI: 10.1007/978-3-030-53953-5_2. [Online]. Available: https://doi.org/10.1007/978-3-030-53953-5_2.

[29] M. Tsionas, *Panel data econometrics: Theory*, en. San Diego, CA: Academic Press, Jun. 2019.

[30] M. Hashem Pesaran, *Time Series and Panel Data Econometrics*, en. London, England: Oxford University Press, Aug. 2015.

[31] E. F. Fama and K. R. French, "The capital asset pricing model: Theory and evidence," *Journal of Economic Perspectives*, vol. 18, no. 3, pp. 25–46, Aug. 2004. DOI: 10.1257/0895330042162430. [Online]. Available: https://doi.org/10.1257/0895330042162430.

[32] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay, *The econometrics of financial markets*, en. Princeton, NJ: Princeton University Press, Dec. 1996.

[33] R. S. Tsay, *Analysis of financial time series* (Wiley Series in Probability and Statistics), en, 2nd ed. Chichester, England: Wiley-Blackwell, Sep. 2005.

[34] R. J. Shiller, *Market Volatility*, en. London, England: MIT Press, May 1990.

[35] A. Damodaran, "Return on capital (ROC), return on invested capital (ROIC) and return on equity (ROE): Measurement and implications," *SSRN Electronic Journal*, 2007. DOI: 10.2139/ssrn.1105499. [Online]. Available: https://doi.org/10.2139/ssrn.1105499.

[36] E. F. Fama and K. R. French, "Profitability, investment and average returns," *Journal of Financial Economics*, vol. 82, no. 3, pp. 491–518, 2006, ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2005.09.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304405X06001164.

[37] FactSet. "Financial database." (), [Online]. Available: https://www.factset.com/. (accessed: 23.04.2023).

[38] S. Bhandari and M. Adams, "On definition, measurement, and use of the free cash flow concept in financial reporting and analysis: A review and recommendations," *Journal of Accounting and Finance*, vol. 17, pp. 11–19, Jan. 2017.

[39] L. M. Fraser and A. Ormiston, *Understanding Financial Statements*, 10th ed. Upper Saddle River, NJ: Pearson, Jan. 2012.

[40] A. Damodaran, *Investment valuation* (Wiley finance), en, 3rd ed. Nashville, TN: John Wiley & Sons, Mar. 2012.

[41] E. (2016). "Impact of november 2016 opec agreement on oil market." (), [Online]. Available: https://www.ecb.europa.eu/pub/pdf/other/eb201608_focus01.en.pdf. (accessed: 23.04.2023).