Master's Thesis Economics: International Business

The influence of gender diversity and earnings management on short- and long-term financial performance of European firms

Academic year 2018 - 2019

Nijmegen School of Management

Radboud University Nijmegen

Author:	Joey Janssen – s1014041
Supervisor:	dr. F. van Beest
Second reader:	dr. G. Braam
Date:	01-07-2019

Abstract

Gender diversity in boards of directors is an increasingly described topic in developed countries. This study contributes to this discussion by examining the effects of gender diversity and earnings management on financial firm performance in the short and long run. The data sample includes 951 listed companies in 2017. The companies are from 10 different countries of the European Economic Area without a gender quota. It is found that gender diversity has no influence on firm performance in the short run, and has a positive influence in the long run. In addition, has earnings management (ABEM) a positive influence on short-term firm performance, and no influence on long-term firm performance. Another part of earnings management, RAEM, has no influence in the short term, but it does have a negative influence on long-term firm performance. It further appears that earnings management mediates the relationship between gender diversity and firm performance. In addition, a moderation effect has also been found. Namely, that earnings management and gender diversity complement in explaining firm performance.

Keywords: accrual-based earnings management; discretionary accruals; earnings management; financial performance; gender diversity; mediation effect; moderation effect; real activities earnings management; quota

Table of Contents

1. Introduction

"Diversity in the 'man cave': Boardrooms gain women as minorities lag" (Reuters, 2019)

"Companies with female executives perform better, says research" (Financial Times, 2015)

It is generally known that men and women are physically different, but also in decision-making and risk aversion, men and women are different (Eckel & Grossman, 2008; Borghans et al., 2009). Moreover, are women more ethical (Dalton & Ortegren, 2011). A more diverse board will also make better managerial decisions and there will be better strategy formulation and more creativity (Dallas, 2001). This is partly due to the quality of the monitoring role, which increases when the diversity is higher (Campbell & Minguez-Vera, 2008). Joecks et al. (2013), find that the effect of gender diversity on firm performance follows a U-shape, and that a number of 3 women in a board is the "magic" number for a higher firm performance. This phenomenon is part of the critical mass theory. Many companies only appoint women as symbolic gesture. For example, by appointing one or two female board members. However, if the critical mass (≥ 30%) is attained, the chance for higher firm performance enhances (Torchia et al., 2011). According to Campbell & Minguez-Vera (2008), higher gender diversity in Spanish firms has a positive effect on firm value and "*the opposite causal relationship is not significant*"(p. 435). Moreover, they find that higher gender diversity can yield more economic advantages. Investors in Spain would also not disadvantage companies with high board diversity (Campbell & Minguez-Vera, 2008). In a related study is described that stock markets react positively in the short term, if they obtain information about female board hiring (Campbell & Minguez-Vera, 2010). Another study found no significant relationship between gender diversity and financial firm performance (Marinova et al., 2016). This time it concerned Dutch and Norwegian listed companies. It can therefore be concluded that a great deal of conflicting literature is available. This uncertainty about the effect on short and long-term financial performance is the reason for this study. If it is concluded that high board diversity has a positive impact on long-term financial performance, companies can make good use of this by hiring more women.

After a long period in which there were only men in a corporate board, more and more women are in boards in recent decades (European Commission, 2016). To continue this trend, the European Commission attempts to promote gender balance in all European states. In some countries such as Austria (35%), Belgium (33%), France (40%), and Germany (30%), gender quotas already apply to board compositions (European Commission, 2016). In the Netherlands, only a *target* of 30% female board members in large companies applies from 1 January 2020, which is defined as a soft quota within this study, since it is more a symbolic regulation (Senden, 2018). Hard quotas, on the other hand, are not targets, but actual quotas with sanctions if these quotas are not met. However, none of the EU-

members fully meet the set quotas yet. Would this situation change when it appears that a higher percentage of women in boards, increase firm performance?

There are of course many other factors that can influence firm performance. Earnings management, for example, can be an influential factor of financial performance (Nichols & Wahlen, 2004). This term is defined as the use of accounting techniques in ways that generate an overly positive view of the firm, whereby directives can obtain a private benefit (Kenton, 2018). There are two different forms of earnings management, namely; *real activities earnings manipulation (RAEM)* and *accrual-based earnings management (ABEM)* (Cohen & Zarowin, 2010; Zang, 2012). The former is defined as altering the execution of real transactions within a fiscal year, and the latter is defined as the change of accounting methods used for presenting transactions in the financial statements (Zang, 2012). Board members can use these sorts of earnings management to eventually obtain private benefit and make the financial results of a company appear more positive or negative, if desired. In this way, the profit or turnover of the company can be made worse or better. This can entail benefits regarding taxes, subsidies, personal bonuses, and attracting shareholders (Guidry et al., 1999; Dechow & Skinner, 2000; Phillips et al., 2003; Nichols & Wahlen, 2004).

Guidry et al. (1999) found that a significant number of managers manipulate short-term earnings, in order to increase their short-term bonuses. However, the expectation is that earnings management will be less if the percentage of women within the board of directors is higher, partly due to the more effective monitoring role (Campbell & Minguez-Vera, 2008). In addition, have Kaplan et al. (2009) found that female board members are more likely to report intentions about fraudulent financial reporting through an anonymous channel. This also includes cases such as earnings management. The reason women do this more could be related to their greater relationship with ethical behaviour. As a result, higher board diversity is expected to result in less earnings management, and less earnings management would in turn have a positive effect on the long-term financial performance of a company.

The research on the effects of gender diversity and earnings management on financial firm performance of European firms combined, is limited. Especially in combination with a study on the potential moderation effect between earnings management and gender diversity. However, there are many results of the effect of gender diversity on financial firm performance alone, but many results are divergent. Many studies examined American data (Campbell & Minguez-Vera, 2008). This study, on the other hand, examine listed European companies instead. In addition, gender diversity in boards of directors is increasingly in the news in recent years, and more countries are now taking sanctions against non-compliance with quotas (European Commission, 2016). From 2020 onwards, it is for example compulsory to have at least 1 female board member in listed companies in the state of California in the United States (Utrecht University, 2018). In some European countries as Austria, Belgium, France,

Germany, Greece, Italy and Spain, quotas between 30 and 40 percent are already in place (European Commission, 2016). After such legislative actions, there are increasing discussions as to whether such quotas are not just places to be filled by women, rather than looking at the individual's performance. Hence a study like this could show whether there is evidence of better business performance due to higher gender diversity. The results may be useful for the European Commission and national legislators. For example, if it turns out that gender diversity would have such a significant positive impact, a Europe-wide higher quota could be an option.

This study is aimed at filling these knowledge gaps. The research question this study answers is as follows:

"*What is the influence of gender diversity in corporate boards and earnings management on short- and long-term financial performance of European firms?"*

In section 2, literature regarding the beforementioned topics is described. Section 3 describes the methodology concerning this research. This includes the conceptual model, used variables, regressions, and data acquisition. Section 4 shows the results, and lastly, in section 5, the conclusions are formed.

2. Literature review and hypotheses

2.1 Financial performance

Financial performance is a commonly used measure of overall business performance (Delen et al., 2013). Earnings represent the "bottom-line" accounting measure of financial firm performance (Nichols & Wahlen, 2004). Earnings show the profit or loss of activities performed by the company, during a specified period. The capital market's measure for "bottom-line" financial performance is the firm's stock return (Δ market value + paid dividends). Reasons for the common use of earnings in combination with share prices are that current period earnings include useful information for forecasting future earnings, and that these future earnings provide information for dividend expectation development, which in turn determines share prices (Nichols & Wahlen, 2004). The figure below provides a detailed overview of these principles.

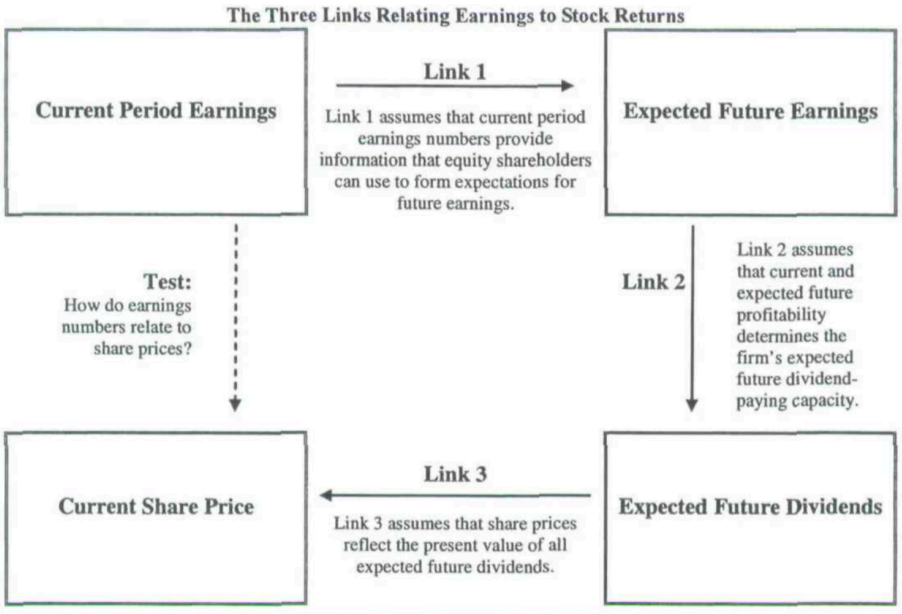

Figure 1: Three links relating earnings to stock returns (Nichols & Wahlen, 2004)

Since shareholders benefit most from good stock returns / share prices, it can be concluded that financial performance is a relevant variable for firm performance.

Financial performance can be divided into short and long-term performance. Both must be measured with different variables. An advantage of financial performance indicators is that after a number of years a comparison can be made with ratios of previous years. If this is done correctly, taking into account various aspects, such as the age of the company and investments, it is possible to measure whether a company performs better. However, a disadvantage of using financial ratios, is that they can be disrupted by various risks such as the interpretation of (tax) legislation (Wernerfelt & Montgomery, 1988) and earnings management (Healy & Wahlen, 1999; Dezsö & Ross, 2012). The latter is discussed in section 2.4.

2.1.1 Short-term financial performance

Short-term financial performance can be measured by financial statement ratios as ROA (Return over Assets) and ROE (Return on Equity). These accounting-based indicators for the greater part depend on the used asset-valuation method (Marinova et al., 2016). According to various studies, these appear to be frequently used instruments for measuring financial performance on short-term (Erhardt et al., 2003; Shrader et al., 1997). Opponents of these short-term methods criticize that these are limited by standards, established by the profession. In other words; impacted by the accounting practices (Al-Matari et al., 2014). Nevertheless, is ROA a broadly used indicator. This indicator measures the operating and financial performance of a firm. The higher ROA is, the more effective is the use of resources in favour of the shareholders (Al-Matari et al., 2014). A higher ROA also stands for the effective use of assets in favour of the economic interests of the shareholders (Ibrahim & AbdulSamad, 2011). Furthermore, when a company shows positive results through ROA, the previously planned performance has been achieved (Al-Matari et al., 2014; Nuryanah & Islam, 2011).

2.1.2 Long-term financial performance

Theory of finance shows that the current value on the stock market is nothing more than a proxy for all future cash flows. Either direct sales, via the future dividend flows, or simply the price increase of the shares. This is shown in figure 1 (Nichols & Wahlen, 2004). Long-term financial performance on the other hand, can be measured by multiple ratios, such as market value added (MVA), Market-to-book value (MTBV) and dividend yield (DY) (Al-Matari et al., 2014). However, the most frequently long-term value assessor is Tobin's Q. This is a commonly used unit of measurement for long-term financial performance and is defined as the ratio of the market value of a specific company to its assets' replacement costs (Wernerfelt & Montgomery, 1988; Hayes, 2019). This assessor is much more appealing than accounting returns (Wernerfelt & Montgomery, 1988). Mainly because it is less influenceable to accounting choices of board members and is the predominant measure used for long-term measurements (Marinova et al., 2016). Moreover, is Tobin's Q a predominant measure used in gender diversity studies, and relatively easy to interpret (Marinova et al., 2016). Wernerfelt & Montgomery (1988) further explain that Tobin's Q *"implicitly uses the correct risk-adjusted discount rate, imputes equilibrium returns, and minimizes distortions due to tax laws and accounting conventions"*, (p. 247). The three qualities quoted, can all be framed as long-term issues. On the other hand, a disadvantage of Tobin's Q is that intangible assets are omitted, which may lead to performance overstatement regarding firms with many of these intangible assets (Lindenberg & Ross, 1981).

2.2 Gender diversity

There are many differences between men and women in the way they think. Women approach leadership, industrial design, conceptualization, management, delegation and investment strategy different as men do (Forbes, 2016). Moreover, are women more ethical, in most cases (Dalton &

Ortegren, 2011). Many of these differences are possibly visible in corporate boards. For example, decision-making and risk aversion between men and women are different (Grey, 2006; Eckel & Grossman, 2008; Borghans et al., 2009). A more diverse board will increase the quality of managerial decisions and there will be better strategy formulation and creativity (Dallas, 2001; Senden, 2018). This is partly due to the quality of the monitoring role, which increases when the diversity is higher (Campbell & Minguez-Vera, 2008). Yet it is strange that more than 50% of the world's population is female, and that only 16% are active in senior corporate or political positions (Grey, 2006).

Some scholars claimed that there is a certain threshold from when the female voice in boards starts to take effect. This principle is known as the critical mass theory. This concept only gained economic significance in the 1980s and became applied to situations in politics where women were not 30% part of political governing bodies (Dahlerup, 2006). Only after a critical mass around 30% of female board members is reached, gender diversity significantly influences firm performance in a positive way (Joecks et al., 2013). These results are based on a panel dataset of 151 listed firms in Germany for the years 2000 till 2005 (Joecks et al., 2013). Others argue that a critical mass is only useful if the belief is discarded that a single proportion of male/female covers all representation needs of women. Moreover, should views about that numbers alone, reach substantive changes in policy processes and outcomes, be rejected (Grey, 2006). For example, in New Zealand, woman with senior positions more readily acted as and for women in cases when they are in a team of sufficient size with feminist views. In addition, should these women find themselves in a pro-feminist environment (Grey, 2006). Another prominent scholar in the field of gender diversity, Dahlerup (2006), does not support the critical percentage of 30%. Dahlerup believes that it is necessary to do more specific empirical-based research on the importance of the size of female minorities under different conditions. Childs & Krook (2006), share this thought partly. The expectation of a sudden change as soon as women make up a certain percentage of an organization, is theoretically questionable and may incorrectly reflect the performance of existing female directors. On the other hand, they indicate that the critical mass concept proved to be very useful in increasing women's appointments, because this theory insists that only a few women are not enough for performance changes on large scale (Childs & Krook, 2006).

Some countries set quotas for the representation of women on corporate boards. Norway, for example, with a percentage of 40 percent (Wang & Kelan, 2013). This can have positive consequences for women, for example the acquisition of leadership roles. Correcting discrimination and offering equal opportunities to have a voice, just as men have, are valid reasons for quotas as well (Senden, 2018). Such countries indicate that women have difficulties, gaining access to dominant positions. And this while more and more women than men are graduating. However, these women do not hold influential positions with the ability of determining what society looks like (Senden, 2018). One wonders whether a quota with a preferential arrangement is a justified differentiation. Senden (2018) thinks so, provided

there is a clear under-representation of women. Especially for listed companies, i.e. the companies that are analysed in this study, it often applies that they affect a large and diverse group of people. If their boards of directors then reflects this group of people, this is a positive aspect (Senden, 2018).

2.3 Gender diversity & financial performance

Research by Joecks et al. (2013) shows that multiple years will pass until a certain number of women has a significant influence on financial firm performance. Furthermore, they have not been able to observe a significant short-term relationship in their study of 151 German companies over the years 2000-2005. It is found that at least three women are necessary for a significant change in performance (Joecks et al., 2013). Carter et al. (2010) have not been able to find a significant influence of diversity on financial performance in their sample of 500 American listed companies in the period 1998-2002 as well. They conclude that the relationship is endogenous. A comparable result can be found in a study by Chapple & Humphrey (2014). They find a weak insignificant negative correlation in some industries, and a weak insignificant positive correlation in others. A study of only Malaysian companies yielded inconsistent results to demonstrate a relationship between gender diversity and financial performance, measured by ROA (Marimuthu & Kolandaisamy, 2009). However, in some cases there appears to be a positive significant result. A study of 127 American companies (1993-1998) indicates that board diversity is positively correlated with financial firm performance on the short term, measured by ROA (Erhardt et al., 2003).

The majority of the literature discussed shows no significant influences on firm performance in the short run. It is therefore expected that gender diversity will not affect financial performance in the short term. Hypothesis 1a has been developed to test this expectation:

H1a: Gender diversity in European corporate boards has no influence on short-term financial performance of these firms.

In the long term, there appears to be a significant positive relationship between gender diversity and long-term financial performance in multiple cases. One study concerns Spanish companies, at a time when there was no quota (Campbell & Minguez-Vera, 2008). In addition, Erhardt et al. (2003) conclude that there is a significant positive influence of gender diversity on financial firm performance in the long term. This time they measure, based on return on investment (ROI). However, there are also cases where no significant relationship is measured in the long term. An empirical study on 186 listed Dutch and Danish companies in 2007 with an average of 5.4% female board members per company is such an example (Marinova et al., 2016). No less than 40 percent of the companies have at least 1 woman on their boards.

There are several reasons for the positive association between gender diversity and long-term financial firm performance. For example, studies by Eckel & Grossman (2008) and Borghans et al. (2009) show that women are more risk averse. This is reflected in less risky decisions by corporate boards. Although this ensures that the chance of extremely large incoming cash flows is reduced, it ensures that the certainty of smaller returns is guaranteed. This guarantees continuity and ensures healthy business operations in the long term (Eckel & Grossman, 2008). As described in section 2.2, women appear to be more ethical than men. This can be reflected in a better social reflection of the company through more ethical decisions (Dalton & Ortegren, 2011). A more diverse board will also make better managerial decisions, there will be better strategy formulation and more creativity (Dallas, 2001). This is partly due to the quality of the monitoring role, which increases when the diversity is higher (Campbell & Minguez-Vera, 2008). Since the majority of the discussed literature shows significant positive relationships on the long term, it is expected that the positive aspects that female board members carry, will be noticeable in the long term for European companies. This is translated into the hypothesis below:

H1b: Gender diversity in European corporate boards has a positive influence on the long-term financial performance of these firms.

2.4 Earnings management

Earnings management is defined as the use of accounting techniques in ways that generate an overly positive view of the firm, whereby directives can obtain a private benefit (Kenton, 2018). There are two different forms of earnings management, namely; real activities earnings manipulation (RAEM) and accrual-based earnings management (ABEM) (Cohen & Zarowin, 2010; Zang, 2012). Both are discussed below.

2.4.1 Accrual-based earnings management

ABEM is defined as altering the execution of real transactions within a fiscal year. This is done by changing accounting methods, used for presenting transactions in the financial statements (Zang, 2012). Board members could influence the used accounting methods to the most favourable one by manipulating accruals. Accruals are the difference between cash flows and net operating income in a specific period (Zang, 2012). These can be split into discretionary and non-discretionary accruals. Discretionary accruals can for example be influenced by adjusting inventory write-off (Healy & Wahlen, 1999). A disadvantage of ABEM, from the point-of-view of the users, is that this method is more visible for auditors and supervisors compared to RAEM (Zang, 2012). However, an advantage of this method is that it can be applied after the financial year is completed. Therefore , it can be decided at a later stage whether to proceed with a positive or negative correction through ABEM (Enomoto et al., 2015). The next years, this method could lose its effect. However, only when auditors or supervisors become aware of it (Zang, 2012). These harmful effects on the reputation can for example arise because investors realize that the result has been achieved by applying ABEM. If this is the case, a negative relationship

between ABEM and long-term financial performance is expected. On the other hand, if no one is aware of the use of ABEM, the relationship does not necessary change from the short-term expectation (Nichols & Wahlen, 2004).

2.4.2 Real activities earnings manipulation

Real activities earnings manipulation (RAEM) is defined as the change of accounting methods used for presenting transactions in the financial statements (Roychowdhury, 2006; Zang, 2012). RAEM is, for example, done by over-producing inventory with the aim of reducing costs of goods sold and by manipulating sales (Roychowdhury, 2006). These efforts are also associated with higher costs compared to ABEM. It is a very effective method that is likely to be used by directors which are mostly interested in the private benefits they can obtain (Kenton, 2018). The advantage of the manipulating board members, and the disadvantage for the firms, is the fact that RAEM is difficult to detect. This makes this form of earnings management a lot more harmful than ABEM in the long term (Achleitner et al., 2014). After all, transactions influenced by RAEM appear to be normal transactions in the short term when there is no suspicion. Main aspects of real activities earnings management are abnormal cash flows from operations, abnormal levels of discretionary expenses and abnormal levels of production costs (Roychowdhury, 2006; Achleitner et al., 2014). With regard to discretionary expenses, managers can for example increase earnings over a certain period through decreasing non-operating expenses such as advertising expenses, R&D and selling, general and administrative expenses (Achleitner et al., 2014). The above examples can also contribute to abnormal cash flows from operations, in cases where these expenses are paid directly.

2.5 Earnings management & financial performance

When considering the arguments of Nichols and Wahlen (2004), regarding the links between accounting earnings and stock returns, the management of these earnings is of relevance as well. Financial performance is logically explained by earnings; however, this does not apply to earnings *management*. Earnings management only generates a *temporary effect* and this will fade away, with as consequence that the expectations eventually need to be adjusted (Nichols & Wahlen, 2004). Since ABEM will not be noticed quickly in the short term by shareholders and auditors, the expectation is that the use will not have any negative reputation effects in the short term (Zang, 2012). In addition, is shown that ABEM has no negative effects on cash flows (Roychowdhury, 2006; Achleitner et al., 2014). Since board members have the option to decide later if they want to use ABEM, they probably attempt to get the best results out of it (Enomoto et al., 2015). The following hypothesis has been formed upon this basis:

H2a: Accrual-based earnings management has a positive influence on short-term financial performance of European companies.

The next years, this method could lose its effect since auditors or supervisors could scrutinize these practices (Zang, 2012). Manipulating earnings via ABEM is limited because the accounting flexibility of a company. Therefore, a different accounting method cannot be chosen every year (Becker et al., 1998). According to Zang (2012), this is a significant risk, since methods within ABEM will stand out if they are applied several times. However, this does not necessarily have to be the case. The literature on earnings management describes both options; awareness and non-awareness (Zang, 2012; Etomoto et al., 2015). It is therefore not possible to state with certainty whether the effect will be positive or negative. Both options remain possible in the long term (Nichols & Wahlen, 2004). Based on these contrasting arguments, it appears that the long-term effect depends on the awareness of earnings management by outsiders. It can be stated that a significant effect is expected in any case. If outsiders don't find out about the use of ABEM, it will continue, which has a positive effect on the results, and when they do find out, performance will be negatively influenced by the company's decreasing reputation. The following hypothesis is derived:

H2b: Accrual-based earnings management has a negative or positive influence (+-) on long-term financial performance of European companies.

For the other type of earnings management, RAEM, two hypotheses are formed too. Since REAM could be done by various, hardly detectable, practices, there is little risk for the concerned *board members* in the short term to be caught (Roychowdhury, 2006; Achleitner et al., 2014). As a consequence, it can have a big risk for the *firm* in the short term. RAEM is costly due to, for example, large discounts and increases in investments (Graham et al., 2005). The results may seem promising because certain targets are met, but actual cash flows are destroyed for the sake of reporting desired accounting numbers. These findings are translated in the following expectation:

H2c: Real activities earnings manipulation has a negative influence on short-term financial performance of European companies.

RAEM, performed in the current period aimed on increasing earnings, could have negative effects on cash flows in the long run (Roychowdhury, 2006; Achleitner et al., 2014). For example, substantial discounts can contribute to increasing sales in the short term, but this can also create expectations for customers in the long term. This can have the long-term effect that customers buy less because they think that the products or services are "more expensive". This can negatively influence the financial performance in the long term (Roychowdhury, 2016). In addition, RAEM can for example mitigate shareholder power for higher dividends, since earnings could be reported downward. This problem mainly occurs at companies where the board wants to retain transgenerational control (Achleitner et al., 2014). It should also be borne in mind that when incorrect financial information is reported year after

year, this can have long-term effects on the interest for investors to invest in the specific company (Healy & Wahlen, 1999). In addition, the primary objective of RAEM is to achieve earnings thresholds (Roychowdhury, 2006). If continuity and operational management no longer have the highest priority, this can cause problems in the long term. These expectations are translated into the following hypothesis:

H2d: Real activities earnings manipulation has a negative influence on long-term financial performance of European companies.

2.6 Gender diversity, financial performance & earnings management

In section 2.2, is described that female board members have a more effective monitoring role than male members do (Campbell & Minguez-Vera, 2008). Because of their effective monitoring role on, for example, managers and fellow board members, it is expected that women will be able to recognize earnings management more quickly (Adams & Ferreira, 2009). And in particular accrual-based earnings management. Real-based earnings manipulation may also be recognized more quickly. Not all found cases of earnings management, can be traced to this monitoring role. The high ethical behaviour of women in combination with the lower tolerance of opportunism, also contribute to found cases (Srinidhi et al., 2011). Earnings management is more unethically experienced by women than by men, due to the misleading of shareholders. Barber & Odean (2001) stress that men trade 45% more than women, due to their overconfidence. As a result, men's net returns are reduced by 2.65 percentage points a year as opposed to 1.72 percentage points for women, due to failures (Barber & Odean, 2001). This once again reflects the risk aversion of women. This risk aversion, combined with all other traits of women, ensures that there is a positive correlation between higher board diversity and financial firm performance (Krishnan & Parsons, 2008). It is therefore expected that women not only recognize earnings management more quickly, but also will report it quicker than men do (Kaplan et al., 2009). According to Srinidhi et al (2011), the quality of earnings is increased, due to a better overview of managers' reports. This is the first study that found empirical evidence for a significant negative correlation between gender diversity and accrual-based earnings management. In addition, they also describe that this reduction in earnings management, due to higher gender diversity, has the effect of increasing corporate performance and governance.

2.6.1 Mediation effect

A mediator is characterized by absorbing an effect of the independent variable (Baron & Kenny, 1986). The mediation effect can be best explained with the help of figure 2.

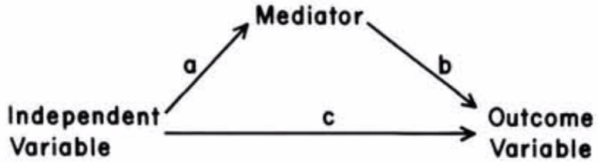

Figure 2: Mediation effect (Baron & Kenny, 1986, p. 1176)

In the case of this study, the independent variable is gender diversity, the mediator is earnings management (either ABEM or RAEM) and the outcome variable is financial performance (either short- or long-term). A commonly used way for testing mediating effects, is the "causal-step method" (Baron & Kenny, 1986). In the case of this study, four regression analyses could make clear whether mediation is involved. Firstly, it is tested whether there is a significant relationship between the independent variable and financial the outcome variable (arrow C in figure 2). Secondly, it is tested whether there is a significant link between the independent variable and the potential mediator (arrow A in figure 2). Thirdly, the significance of the relationship between the potential mediator and the outcome variable is tested (arrow B in figure 2). Fourthly, a multiple regression should be formed, which indicates whether there is really mediation. The significance of the relationship between the independent variable, the potential mediator and the outcome variable is tested simultaneously (Baron & Kenny, 1986).

A necessary condition for granting a mediation / absorption effect is testing of the relationship between the independent variable and the mediator. This relationship appears to be the case between gender diversity and earnings management according to prior studies (Campbell & Minguez-Vera, 2008; Kaplan et al., 2009). Both forms of earnings management will be less if the percentage of women within the board of directors is higher, partly due to the more effective monitoring role (Campbell & Minguez-Vera, 2008). In addition, women are more likely to report intentions about fraudulent financial reporting through an anonymous channel (Kaplan et al., 2009). The reason women do this more could be related to their greater relationship with ethical behaviour. As a result, higher board diversity leads to less earnings management, for both RAEM and ABEM. The effect of earnings management on financial firm performance must also remain significant after controlling for gender diversity. As soon as it appears that the relationship between gender diversity and financial firm performance is no longer significant, it can be stated that complete mediation is the case. When gender diversity still has a significant effect, it is referred to as a partial mediation effect (Baron & Kenny, 1986).
The following hypotheses are tested:

H3a: Earnings management mediates the relationship between gender diversity and short-term financial performance

And

H3b: Earnings management mediates the relationship between gender diversity and long-term financial performance

These hypotheses have been included purely for exploration. Little research has yet been done into the mediating role of specific forms of earnings management on the relationship between gender diversity

and firm performance. Hence, there is no expectation that a specific form of earnings management will play a significant mediating role. These hypotheses are therefore formed in a more general way and a potential mediation effect could take place.

2.6.2 Moderation effect

A moderator is a variable which influences the strength or direction between a dependent and independent variable which can either be complementary or substitutionary (Baron & Kenny, 1986). Moderation occurs when the moderator variable influences the direction or strength of the relationship between the independent and dependent variable, and the interaction term (predictor times moderator) is significant (Baron & Kenny, 1986). However, the effects of the moderator and the predictor on the dependent variable do not necessarily have to be significant to test the moderation hypothesis (Baron & Kenny, 1986).

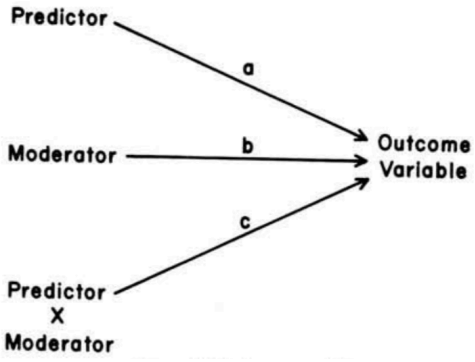

Figure 3: Moderation effect (Baron & Kenny, 1986, p. 1174)

It is expected that there is a negative moderation effect of gender diversity on the relationship between EM (either ABEM or RAEM) and financial performance (either short- or long-term). Since female board members are more ethical and creative, and possess more managerial advantages, than male members do (Campbell & Minguez-Vera, 2008; Dalton & Ortegren, 2011; Forbes, 2016). Because of their more effective monitoring role on, for example, managers and fellow board members, it is expected that women will be able to recognize earnings management more quickly (Adams & Ferreira, 2009). This is formulated in the following hypothesis:

H4: Gender diversity and earnings management complement in explaining financial performance.

The moderation effect within this research is only exploratory. In comparison with the other expected relationships included in the hypotheses, less existing research is available on moderation between these variables. This hypothesis is therefore formed in a more general way.

3. Methodology

3.1 Conceptual model

The research model, shown in figure 4, indicates the relationships that are expected to be found. A direct relationship is expected between gender diversity on long-term firm performance. On the short-term, no influence of gender diversity is expected. In addition, a direct relationship is expected between gender diversity and the extent to which there is RAEM and ABEM. In addition, it is expected that earnings management (ABEM & RAEM) mediates the effect of gender diversity on financial performance. Lastly, a moderation effect is expected, namely: the effect of gender diversity (moderator) on the relation between earnings management and short- and long-term firm performance.

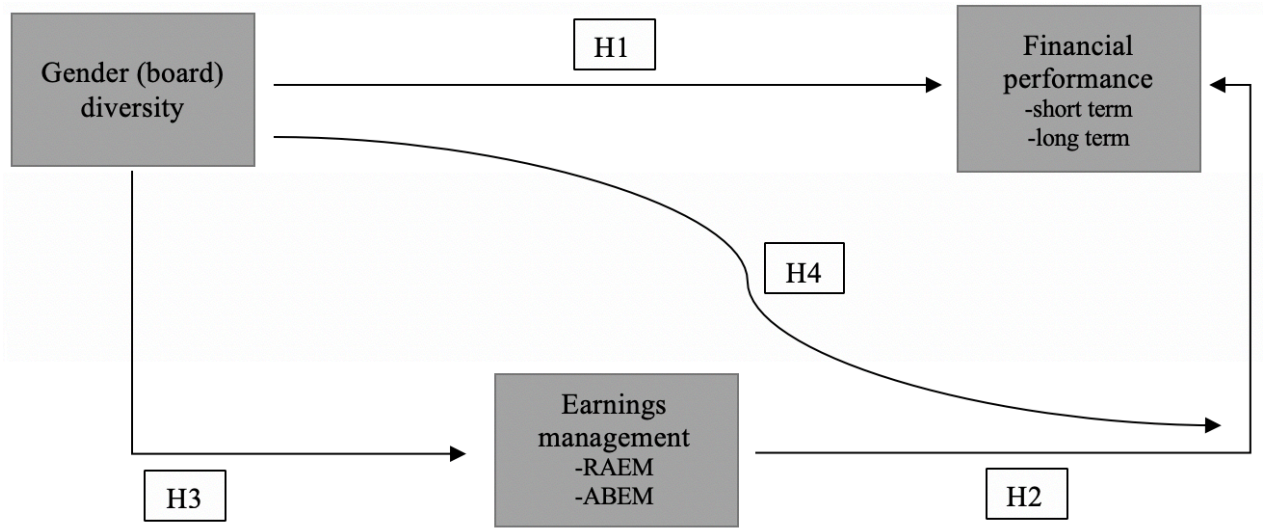

Figure 4: Research model (Glaum et al., 2013, adapted)

3.2 Research method

This empirical research explains the influence of gender diversity in corporate boards in combination with earnings management, on the financial performance of European companies in the short and long term, through a quantitative research method. Cross-sectional data analyses will be performed by ordinary least-squares (OLS) multivariate regressions. This method can estimate the effect of the predictor variables independently, by holding the other variables' effects constant. Data sets on firm level are used, obtained via Boardex and Eikon. In this way it is possible to gain insight into the composition of corporate boards. For this, data of the 31st of December 2017 is used, since this is the most recent date with complete data for all variables.

3.3 Variables

3.3.1 Dependent variable: financial performance

The dependent variable within this research, financial performance, is split into short-term financial performance and long-term financial performance, as the influence of gender diversity is expected to vary in the short and long term. *Short-term financial performance* is measured by ROA (Return over Assets). This financial ratio is measured by dividing the net income by total assets (Erhardt, Werbel, & Shrader, 2003; Carter et al., 2010);

$$ROA = \frac{Net\ income}{Total\ assets}$$

According to various studies, this appears to be a frequently used and appropriate instrument to measure financial firm performance on short-term by market and financial analysts and by other scholars (Erhardt et al., 2003; Shrader et al., 1997 (Carter et al., 2010).

Long-term financial performance, is measured by Tobin's Q. This variable is chosen because it is less influenceable to accounting choices of board members and is the predominant measure used for long-term measurements (Carter et al., 2010; Marinova et al., 2016). Tobin's Q is a commonly used unit of measurement for long-term financial performance and is defined as the ratio of the market value of a specific company to its assets' replacement costs (Carter et al., 2010; Hayes, 2019);

$$Tobin's\ Q = \frac{Market\ value}{Assets'\ replacement\ costs}$$

If Tobin's Q is higher than 1, which is the case when the market value is higher than its replacement costs, it is suspected that this company has future growth opportunities due to intangible assets (Sudarsanam, 2003; Marinova et al., 2016). The data of these variables is retrieved from Eikon.

3.3.2 Independent variables

3.3.2.1 Gender diversity

The necessary quantitative data for the variable *gender diversity* is retrieved from Boardex. Boardex gives insight in the board diversity of listed European companies. Board diversity can also indicate ethnic or age differences of directives (Carter et al., 2010). However, this is outside the scope of this study, so only gender diversity will be discussed. The terms board diversity and gender diversity are used interchangeably during this study. Board diversity will be measured by two measurements combined. Firstly, the percentage of female members in the board of directors, since many scholars find that this is the most effective measure (Erhardt, Werbel, & Shrader, 2003; Campbell & Minguez-Vera, 2008);

$$Gender\ diversity = \frac{number\ of\ female\ board\ members}{total\ of\ board\ members}$$

Secondly, the critical mass theory will be used as measure. This measure is included as a dummy variable. The value is 1 if the percentage of gender diversity is higher than 30% (Dahlerup, 2006; Joecks et al., 2013). If less, the value of the observation is 0.

3.3.2.2 Earnings management: accrual-based earnings management

The other independent variable which is used, *earnings management,* is divided in accrual-based earnings management and real activities earnings manipulation (Cohen & Zarowin, 2010; Zang, 2012). In section 2.6 is described that both variants of earnings management can have a mediating effect on the relation between gender diversity and financial firm performance. Hence are both variants included as independent variables. For the measurement of ABEM, the following equation (adapted Jones model), utilized by Dechow et al. (1995) and Achleitner et al. (2014), is used:

$$\frac{ACC_t}{TA_{t-1}} = \alpha_0 + \alpha_1 \left(\frac{1}{TA_{t-1}}\right) + \alpha_2 \left(\left(\frac{\Delta REV_t}{TA_{t-1}}\right) - \left(\frac{\Delta REC_t}{TA_{t-1}}\right)\right) + \alpha_3 \frac{PPE_t}{TA_{t-1}} + \varepsilon_t$$

ACC_t, stands for total accruals at time t. This is measured by "net income before extraordinary items" less "cash flows regarding operations" at time t (Achleitner et al., 2014). $TA_{1\text{-}t}$ stands for total assets at the beginning of a year. This measure is added to scale ACC_t, ΔREV_t and ΔREC_t at t-1 (beginning of specific year). ΔREV_t stands for revenues in year t less revenues in t-1. ΔREC_t stands for net receivables in year t, less net receivables in year t-1 (Dechow et al., 1995). PPE_t stands for property plant and equipment. ε_t stands for the current discretionary accruals. The above model includes the assymmetric timelines of losses which provides non-linear abnormal accruals expectations (Dechow & Dichev, 2002). This aspect increases the explanatory power of this model.

3.3.2.3 Earnings management: real activities earnings manipulation

For defining the measurement of RAEM, three prior studies are used (Roychowdhury, 2006; Cohen & Zarowin, 2010; Achleitner et al., 2014). Following Roychowdhury (2006) and Cohen & Zarowin (2010), two proxies for RAEM are considered; abnormal cashflow from operations (AB_CFO) and abnormal production costs (AB_PROD). The way in which these proxies are operationalized is described below.

To find out the abnormal levels of cash flow from operations, the normal levels of cash flow from operations are measured with an OLS estimation model (Achleitner et al., 2014). The residue (ε_t) is the abnormal part of the cash flow.

$$\frac{CFO_t}{TA_{t-1}} = \alpha_0 + \alpha_1 \left(\frac{1}{TA_{t-1}}\right) + \alpha_2 \left(\frac{REV_t}{TA_{t-1}}\right) + \alpha_3 \left(\frac{\Delta REV_t}{TA_{t-1}}\right) + \varepsilon_t$$

where CFO_t, stands for cash flow from operations at time t; $TA_{t\text{-}1}$ stands for total assets at the beginning of a year (lagged total assets). REV_t stands for sales at time t. ΔREV_t stands for the change in sales

between t and t-1. All variables in the equation above are scaled by lagged total assets (TA_{t-1}). The proxy for AB_CFO is equal to actual CFO less the normal level of CFO. More negative values of AB_CFO indicate more RAEM (Achleitner et al., 2014).

In order to measure the second proxy, abnormal production costs, the measurement for normal product levels of production costs is operationalized below:

$$\frac{PROD_t}{TA_{t-1}} = \alpha_0 + \alpha_1 \left(\frac{1}{TA_{t-1}}\right) + \alpha_2 \left(\frac{REV_t}{TA_{t-1}}\right) + \alpha_3 \left(\frac{\Delta REV_t}{TA_{t-1}}\right) + \alpha_4 \left(\frac{\Delta REV_{t-1}}{TA_{t-1}}\right) + \varepsilon_t$$

The abnormal levels of production costs (AB_PROD) is measured by actual production costs minus the normal level of production costs at time t ($PROD_t$). The larger the value of AB_PROD, the more RAEM. $PROD_t$ is defined as the sum of cost of goods sold (COGS) and the change in inventory (ΔINV), all at time t. TA_{t-1} again stands for lagged total assets. REV_t stands for sales at time t; ΔREV_t stands for the change in sales (t – t-1) (Achleitner et al., 2014).

To construct these two proxies of RAEM in an unambiguous way, the negative values of AB_CFO were used, so that larger values of the two proxies represent more RAEM. (Achleitner et al., 2014).

Ultimately an overall summery measure is described (Cohen et al., 2008):

$RAEM\_AGG = (\text{AB\_PROD} - \text{AB\_CFO})$
The higher RAEM_AGG (Aggregate RAEM), the more RAEM is used.

3.3.3 Control variables

Not only gender diversity and earnings management exert influence on financial firm performance. Additional variables may add potential influence too. Prior literature mentioned firm size as influential factor of financial firm performance (Campbell & Minguez-Vera, 2008; Achleitner et al., 2014). The former scholars stress that firm size is the logarithm of the book value of the total assets of a firm. Larger firms would generally have a higher financial performance and could encourage the use of RAEM (Achleitner et al., 2014). Firm size (FSIZE) is measured by the logarithm of total assets of a company. The second added control variable is leverage (LEV). Leverage is measured by dividing total debts by total equity (Campbell & Minguez-Vera, 2008). The expectation is that a lower leverage, increases the attractiveness for potential shareholders. Another used control variable is potential firm loss. Firm loss is measured as a dummy (LOSS). If the company does not experience losses, the dummy value is "0". In the case it does, the value will be "1" (Arun et al., 2015). Leverage and firm loss are added to control for the risk of bankruptcy, which affects RAEM choices (Dyreng et al., 2012; Achleitner et al., 2014). The fourth control variable, firm growth (GROWTH), is measured by (total assets (t=0) – Total assets

(t=-1)) / Total assets (t=-1). If a company grows significantly for external reasons, this must of course be corrected. Moreover, could firm growth encourage board members to use RAEM (Achleitner et al., 2014). To provide clarity in the distinction between the results of the European countries, there is a dummy variable which checks for the country concerned. In this way it can be determined per country to what extent gender diversity and earnings management have an effect on performance. Lastly, there is controlled for the industry to which the companies belong. The Thomson Reuters Business Classification is used for this. It turned out that women in boards are not represented proportionally in every industry. For example, Carter et al. (2010) stressed that companies in the financial sector have the most female directors.

Control variable	**Abbreviation**	**Definition/measurement**
Firm size	FSIZE	Total assets
Leverage	LEV	Debts divided by total equity
Potential firm loss (dummy)	LOSS	Positive profit = 0 Negative profit = 1
Firm growth	GROWTH	(Total assets (t=0) – Total assets (t=-1)) / Total assets (t=-1)
Country (dummy)	Country	Concerned country: 1 Other country: 0
Industry (dummy)	Industry	TRBC (Thomson Reuters Business Classification) Industry name

Table 1: Control variables

3.4 Regression analyses based on hypotheses

To test the hypotheses discussed in section 2, multiple regressions are performed in this section. The definitions of the abbreviations of the used variables are described in section 3.3. The substantiation of the predicted signs within the regressions can be found in the relevant subjects in section 2.

The hypothesis regarding the direct effect of gender diversity on short-term financial performance, H1a is as follows:

H1a: Gender diversity in European corporate boards has no influence on short-term financial performance of these firms.

Regressions 1a-1 and 1a-2 are formed upon the basis of this hypothesis:

$$1a\text{-}1: \quad ROA = \alpha + \beta_1 Genderpercentage + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$$

$$1a\text{-}2: \quad ROA = \alpha + D_1 Criticalmass + \beta_1 FSIZE + \beta_2 LEV + \beta_3 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$$

With α as intercept, ε as error term, β_2 up to and including β_4 as the coefficients of the control variables, and D_2 up to and including D_4 as the coefficients of the control dummy variable (the interpretation of these symbols applies to all regressions in this study). There is no expected positive or negative association between gender diversity and short-term financial performance, so it is expected that the coefficients of β_1 and D_1 can be negative or positive. Furthermore, β_1 and D_1 do not have to be significant to support hypothesis 1a, since no significant relationship is expected between gender diversity and ROA.

Hypothesis 1b, which expresses the relationship between gender diversity and long-term financial performance, is as follows:
H1b: Gender diversity in European corporate boards has a positive influence on the long-term financial performance of these firms.
Regressions 1b-1 and 1b-2 are formed upon the basis of this hypothesis:

$$1b\text{-}1: \quad Tobin's\ Q = \alpha + \beta_1 Genderpercentage + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$$

$$1b\text{-}2: \quad Tobin's\ Q = \alpha + D_1 Criticalmass + \beta_1 FSIZE + \beta_2 LEV + \beta_3 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$$

In the long term, it is expected that the positive aspects (ethical and appropriate risk approach) that female board members carry, will be noticeable. The expectation is that this will result in a positive coefficient for β_1. The coefficient of the explanatory dummy variable (D_1), critical mass, is expected to be negative. This indicates that Tobin's Q will be lower if boards don't reach the critical mass of 30%. Moreover should β_1 and D_1 be significant for supporting H1b.

The first hypothesis about the relationship between earnings management and financial performance is H2a.
H2a: Accrual-based earnings management has a positive influence on short-term financial performance of European companies.
In order to measure the effect of accrual-based earnings management on the return over assets (short-term performance), the regression below will be performed.

$$2a: \quad ROA = \alpha + \beta_1 ABEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$$

There is a positive association between accrual-based earnings management and ROA, so a positive significant coefficient on β_1 is expected.

Based on H2b, *"Accrual-based earnings management has a negative or positive influence (+-) on long-term financial performance of European companies",* regression 2b is formed:

2b: $Tobin's\ Q = \alpha + \beta_1 ABEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

The long-term effect of ABEM depends on the awareness of earnings management by shareholders or auditors. So, it is expected that the coefficient of β_1 can be negative or positive. However, β_1 should be significant to support hypothesis 2b, since either a positive or negative relationship is plausible.

Regarding hypothesis 2c, *"real activities earnings manipulation has a negative influence on short-term financial performance of European companies",* regression 2c is formed:

2c: $ROA = \alpha + \beta_1 RAEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

The coefficient of β_1 should be negative and significant, since a negative association between real activities earnings management and ROA is expected. After all, these transactions appear to be ordinary transactions on the short term when there is no suspicion. Hence the expected negative coefficient.

H2d is formed as follows: *Real activities earnings manipulation has a negative influence on long-term financial performance of European companies.*

This hypothesis results in the forming of regression 2d:

2d: $Tobin's\ Q = \alpha + \beta_1 RAEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

Since RAEM has a direct influence on cash flows, there is a chance that this will cause disastrous problems in the future. This negative association between RAEM and Tobin's Q leads to the expectation of a negative coefficient on β_1.

Hypothesis 3a is formed to test the suspected mediation effect (short-term performance):

H3a: *Earnings management mediates the relationship between gender diversity and short- term financial performance.*

This hypothesis results in the forming of four regressions; 3a-1 up to and including 3a-4:

3a-1: $ROA = \alpha + \beta_1 Genderpercentage + \beta_2 ABEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

3a-2: $ROA = \alpha + D_1 Criticalmass + \beta_1 ABEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

3a-3: $ROA = \alpha + \beta_1 Genderpercentage + \beta_2 RAEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

3a-4: $ROA = \alpha + D_1 Criticalmass + \beta_1 RAEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

As described in H1a, gender diversity is expected to have no effect on ROA. So, it is expected that the coefficient of β_1 can be negative or positive. The same expectation is valid for the coefficient of the explanatory dummy variable (D_1), critical mass. Since it is expected that earnings management, so both ABEM and RAEM, may mediate the relationship between gender diversity and short-term financial performance in both a positive and negative direction, the coefficients of these variables may be negative or positive in both regressions either. A necessary condition for granting a mediation / absorption effect is testing of the relationship between gender diversity and both forms of earnings management. This relationship appears to be the case according to prior studies (see section 2.6). The effect of earnings management on financial firm performance must also remain significant after controlling for the effect of gender diversity. As soon as it appears that the relationship between gender diversity and financial firm performance is no longer significant, it can be stated that complete mediation is the case. When gender diversity still has a significant effect, it is referred to as a partial mediation effect (Baron & Kenny, 1986). To recognize the mediation effect, it is also required that significance of the associated variables has been demonstrated in previous regressions (steps 1, 2 and 3) (Baron & Kenny, 1986). Moreover, should all coefficients within at least one of the models be significant for supporting H3a.

Hypothesis 3b is formed to test the suspected mediation effect (long-term performance):
H3b: *Earnings management mediates the relationship between gender diversity and long-term financial performance.* This hypothesis results in the forming of four regressions; 3b-1 up to and including 3b-4:

3b-1: $Tobin's\ Q = \alpha + \beta_1 Genderpercentage + \beta_2 ABEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

3b-2: $Tobin's\ Q = \alpha + D_1 Criticalmass + \beta_1 ABEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

3b-3: $Tobin's\ Q = \alpha + \beta_1 Genderpercentage + \beta_2 RAEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

3b-4: $Tobin's\ Q = \alpha + D_1 Criticalmass + \beta_1 RAEM + \beta_2 FSIZE + \beta_3 LEV + \beta_4 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

As described in H1b, gender diversity is expected to have a positive influence on Tobin's Q. So, it is expected that the coefficient of gender percentage is positive. The coefficient of the explanatory dummy variable (D_1), critical mass, is expected to be positive either. This indicates that Tobin's Q will be lower if boards don't reach the critical mass of 30%. Since it is expected that earnings management, so both ABEM and RAEM, may mediate the relationship between gender diversity and long-term financial performance in both a positive and negative direction, the coefficients of ABEM and RAEM in all regressions can be negative or positive. To recognize the mediation effect, it is also required that significance of the associated variables has been demonstrated in previous regressions (steps 1, 2 and 3) (Baron & Kenny, 1986). Moreover, should all coefficients within at least one of the models of hypothesis 3b be significant for supporting H3b.

For measuring hypotheses 4, regarding a potential moderation effect, multiple regressions are formed. H4: *Gender diversity and earnings management complement in explaining financial performance.* This hypothesis results in the forming of regression 4-1 up to and including 4-8.

4-1: $ROA = \alpha + \beta_1 Genderpercentage + \beta_2 ABEM + \beta_3 Genderpercentage \cdot ABEM + \beta_4 FSIZE + \beta_5 LEV + \beta_6 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

4-2: $ROA = \alpha + D_1 Criticalmass + \beta_1 ABEM + \beta_2 Criticalmass \cdot ABEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

4-3: $ROA = \alpha + \beta_1 Genderpercentage + \beta_2 RAEM + \beta_3 Genderpercentage \cdot RAEM + \beta_4 FSIZE + \beta_5 LEV + \beta_6 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

4-4: $ROA = \alpha + D_1 Criticalmass + \beta_1 RAEM + \beta_2 Criticalmass \cdot RAEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

4-5: $Tobin's\ Q = \alpha + \beta_1 Genderpercentage + \beta_2 ABEM + \beta_3 Genderpercentage \cdot ABEM + \beta_4 FSIZE + \beta_5 LEV + \beta_6 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

4-6: $Tobin's\ Q = \alpha + D_1 Criticalmass + \beta_1 ABEM + \beta_2 Criticalmass \cdot ABEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

4-7: $Tobin's\ Q = \alpha + \beta_1 Genderpercentage + \beta_2 RAEM + \beta_3 Genderpercentage \cdot RAEM + \beta_4 FSIZE + \beta_5 LEV + \beta_6 GROWTH + D_1 LOSS + D_2 COUNTRY + D_3 INDUSTRY + \varepsilon$

4-8: $Tobin's\ Q = \alpha + D_1 Criticalmass + \beta_1 RAEM + \beta_2 Criticalmass{\cdot}RAEM + \beta_3 FSIZE + \beta_4 LEV + \beta_5 GROWTH + D_2 LOSS + D_3 COUNTRY + D_4 INDUSTRY + \varepsilon$

Gender diversity is expected to complement the effect of earnings management. Moderation occurs when the moderator variable influences the direction or strength of the relationship between the independent and dependent variable, and the interaction term (predictor times moderator) is significant (Baron & Kenny, 1986). However, the effects of the moderator and the predictor on the dependent variable do not necessarily have to be significant to test the moderation hypothesis (Baron & Kenny, 1986). So, there is no particular requirement for the coefficients in the models of hypothesis 4, except that the direction or strength changes compared to the regressions results of the previous hypotheses. Only the moderation term has to be significant. Since these regressions are based on an exploring hypothesis, there is no specific expectation which aspect of earnings management complements the most.

3.5 Dataset

The data concerning board composition and critical mass is obtained via Boardex, and the data of all other variables (dependent, independent and controls), are obtained via Eikon. Ten European countries of the European Economic Area (EEA) are selected, which are chosen at random. These countries have one thing in common; no hard gender quota. This means that there are no consequences for companies if they do not comply with a specific gender directive, which means that there is no real quota. The reason for analysing ten countries without mandatory quotas, (i.e. voluntary quotas) is that when there are female board members in these countries, they will on average have high qualities, on the basis of which the position was obtained. This could be a reason for the appointment as board member, and as a result they will have a greater influence (Campbell & Minguez-Vera, 2010). After all, gender diversity is much less embedded in these countries than in countries where there is an obligation. In addition, are three countries with a hard quota analysed. This is described in section 4.4.

Of each country, the stock market with the highest observations available within Eikon is chosen. Whether there is a quota, is based on data from the European Commission (2016). Countries where only a directive without sanction applies (soft quota), will be marked in the list below by "no", as companies do not have to expect consequences if they do not comply with the directive.

Country	**Quota**	**Name of stock market**	**Observations (N)**
Denmark	No	OMX Copenhagen (OMXC)	55
Finland	No	Nasdaq OMX / OMX Helsinki	80
Greece	No	Athens Stock Exchange / Athex Composite index	17
Ireland	No	Irish Stock Exchange	42
Netherlands	No	Euronext Amsterdam / AEX All share	88
Poland	No	Warsaw Stock Exchange / Warsaw General	45
Portugal	No	Euronext Lisbon / PSI All share	33
Spain	No	Bolsas y Mercados Espanoles, BME / Madrid SE General	114
Sweden	No	Nasdaq OMX	189
United Kingdom	No	FTSE	508
Total			**1171**

Table 2: Stock markets and observations per country (sample 1)

As described in table 2, there are 1171 observations in total for sample 1. Due to missing values of gender data, 220 observations are removed. Most of the deleted observations were companies from the United Kingdom. Of the remaining data, 4 observations had no value for total assets from the previous year (2016). As a result, the variable "growth" could not be calculated for these observations. This leads to a remainder of 947 observations. For the regressions which only tested the effects of gender diversity and control variables on financial performance, the total quantity of observations is therefore 947.

To calculate ABEM, 147 of the 951 observations were subsequently removed, due to influential cases and missing values of a part of the calculation for ABEM (PPE). Something similar applies to the observations of RAEM. Because there were many missing values for net revenues of 2015, the change in revenues for t-1 could not be calculated for multiple companies. As a result, 177 of the 951 observations were removed. In addition, observations were removed which were in an industry that had fewer than 5 observations. To calculate the discretionary accruals and the abnormal parts of the production costs and cash flows, it is necessary that every industry has at least 5 observations.

Sample 2 (table 3) includes three countries. The total of observations was 489 at the moment of retrieving of the raw data. Sample 2 is only used as a robustness check for testing hypothesis 1a and 1b. In the end, 149 observations were deleted because there was no data available on gender diversity for these firms. This left 340 observations to analyse.

Country	**Quota**	**Name of stock market**	**Observations (N)**
France	Yes, 40%	Euronext Paris / SBF	119
Germany	Yes, 30%	Deutsche Börse / Prime all share	292
Norway	Yes, 40%	Oslo Bors / Oslo Exchange All-share Index_GI	78
Total			**489**

Table 3: Stock markets and observations per country for the sensitivity analysis (sample 2)

4. Empirical results

The results of this study are described in this chapter. Firstly, the descriptive statistics of the dataset are shown. Secondly, the dataset is tested for multicollinearity with a correlation matrix. Thereafter, the results of the analysed regressions are described. In order to test if the performed core regressions behave different when only firms of countries with a gender quota are studied, sample 2 (table 3) is analysed in section 4.4.

4.1 Descriptive statistics

Table 4 shows the descriptive statistics of all used dependent, independent and control variables.

Descriptive statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent variables					
ROA	951	.0684053	.1170582	-1.0176	1.0956
Tobin's Q	951	.9949043	.0467448	.7621403	1.165163
Genderpercentage	951	.2290572	.1434451	0	.7142857
ABEM	804	-1.03e+07	3.85e+07	-3.56e+08	2.31e+08
RAEM	774	-1.24e+07	3.33e+07	-4.76e+08	4.96e+07
RAEM_CFO	930	1.31e+07	3.19e+07	-4.96e+07	4.76e+08
RAEM_PROD	774	-.0445039	.1060293	-.6471714	.381054
Genderperc*ABEM	804	-1433274	9897032	-6.67e+07	7.28e+07
Critmass*ABEM	804	-780985.8	2.11e+07	-1.33e+08	2.00e+08
Genderperc*RAEM	774	-2558291	8522049	-8.93e+07	1.24e+07
Critmass*RAEM	774	-3200003	1.69e+07	-2.28e+08	2.76e+07
FSIZE	951	20.81166	1.646818	15.66349	27.45172
LEV	951	.2522432	.3392214	-2.372	3.388
GROWTH	947	.0019469	.0102461	-.0689433	.0716773

Table 4: Descriptive statistics metric variables

All variables in table 4 are tested for normality, homoscedasticity, and influential cases. Appendix A describes the normality and homoscedasticity tests. Influential cases are analysed by scatterplots and these observations were deleted. However, this did not have a significant influence on the results due to the high amount of observations. The total of observations in the dataset for all non-quota countries is 951. In section 3.5 the reasons for deleting specific observations are described. The mean of ROA is 0.068 with an std. deviation of 0.117. These numbers are in line with comparable studies (Erhardt et al., 2003). In addition, the average of Tobin's Q approximately equals 1, which was expected. Namely, if this number would be significantly below 1, it is a reason for downgrading because the market value is lower than the book value. With regard to gender diversity, there are no strange values. The minimum is of course zero, and the maximum is 0.714. In addition, an average of 0.229 applies. For countries without a mandatory gender quota, this value is not particularly low. The values of RAEM and ABEM

on the other hand, are more difficult to interpret. These are in fact very small values and are calculated by saving the non-standardized errors of the formulas in section 3.3.2 as a new value. However, the calculations for these variables are exactly the same as in models of Dechow et al. (1995), Achleitner et al. (2014) and Cohen et al. (2008). Lastly, the three control variables, firm size, leverage, and growth, do not show strange values and are in line with comparable studies. However, it should be taken into account that firm size and growth are no longer naturally interpretable, due to the transformation to their logarithms. In addition to the above variables, two dummy variables have also been included in the study, critical mass and firm loss. Because these are dummies, it is not interesting to use the mean and std. deviation. It is, however, relevant to report that 269 of the 951 observations meet the critical mass of 30%. In addition, 90 of the 951 companies are making a loss (firmloss dummy).

4.2 Multicollinearity

To test the variables used in this study for multicollinearity, a correlation matrix is generated. Multicollinearity is the case if ≥ 2 independent variables correlate significantly with each other. This table also shows in which way interactions between given variables is important. If the correlation coefficient is lower than 0.3, the strength of the relationship is weak to nil. A coefficient between 0.3 and 0.7 indicates that there is a moderate relationship. If the coefficient is between 0.7 and 1, the relationship is strong. If the relationship is strong, without a valid reason, it should be further investigated. In case of this study, there is one strong relationship between variables, as seen in table 5. This is the relationship between critical mass and gender percentage. However, this relationship can be explained, since they both measure the quantity of female board members. This won't affect the outcome of regressions, since both proxies are not regressed simultaneously. So, there are no multicollinearity problems which should be corrected by excluding a specific variable. In addition to this correlation matrix, a VIF test has been carried out to detect multicollinearity (Appendix B). None of the variables exceed the upper limit of 5, so the results of the VIF test are in line with the correlation matrix.

	ROA	Tobin's Q	Genderperc.	Criticalmass	ABEM	RAEM	FSIZE	LEV	GROWTH	LOSS
ROA	1									
Tobin's Q	0.1787	1								
Genderpercentage	0.0335	0.0547	1							
Criticalmass	0.0184	0.0009	0.7720	1						
ABEM	0.1102	-0.3005	0.1690	0.1211	1					
RAEM	-0.0134	-0.2062	0.0600	0.0215	0.6166	1				
FSIZE	-0.0414	-0.3371	0.1733	0.0937	0.6408	0.1316	1			
LEV	-0.0247	-0.2002	-0.1284	-0.1065	0.1698	0.0774	0.1686	1		
GROWTH	0.3005	0.1071	0.0533	0.0267	-0.0198	-0.0542	0.0280	-0.1248	1	
LOSS	-0.2978	-0.1161	-0.1265	-0.0595	-0.1444	-0.0487	-0.2168	0.0705	-0.2215	1

Table 5: Correlation matrix

4.3 Regression analyses

This section contains the multiple regressions to test the formed hypotheses in this study. Table 6 shows the results of the regression analyses 1a-1, 1a-2, 1b-1, and 1b-2. If the direction of the coefficient is equal to the direction of the predicted sign in the regression analyses, a one-tailed test instead of two-tailed test is performed. This was for example the case with the effect of critical mass on ROA and Tobin's Q. The results of regression analyses 1a-1 and 1a-2 show that the quantity of women on boards does not significantly influence financial firm performance on the short-term, measured by ROA. Therefore, hypothesis 1a is supported. Moreover, explain the models 1a-1 and 1a-2 approximately 32% of short-term financial performance.

In addition, table 6 shows that the quantity of women in board of directors positively influences financial firm performance on the long-term ($p < 0.01$). A critical mass of women (>30%) has also a significant effect on Tobin's Q (* in two-tailed test, ** in one-tailed test). Based on model 1b-1 and 1b-2, a positive significant relationship between the two proxies and Tobin's Q can be confirmed. Hypothesis 1b is therefore supported.

Financial firm performance				
Model	1a-1	1a-2	1b-1	1b-2
	ROA	ROA	Tobin's Q	Tobin's Q
Genderpercentage	0.0401		0.0314***	
	(1.51)		(3.19)	
Criticalmass		0.0151		0.00574*
		(1.86)		(1.90)
Predicted sign	none	none	+	+
FSIZE	-0.00574**	-0.00543**	-0.00841***	-0.00800***
	(-2.39)	(-2.30)	(-9.45)	(-9.10)
LEV	0.00943	0.00895	-0.0212***	-0.0215***
	(0.83)	(0.79)	(-5.04)	(-5.09)
GROWTH	2.075***	2.078***	0.434***	0.433***
	(6.32)	(6.34)	(3.57)	(3.55)
LOSS	-0.0701***	-0.0708***	-0.0265***	-0.0273***
	(-5.45)	(-5.52)	(-5.56)	(-5.71)
Country	included	included	included	included
Industry	included	included	included	included
Constant	0.186** (2.55)	0.182** (2.50)	1.119*** (41.41)	1.114*** (41.18)
Observations	947	947	947	947
Adj. R-squared	0.323	0.324	0.418	0.414

Explanation: t-statistics are below the coefficients of each variable in parentheses.
Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)

Table 6: Regression analyses of hypotheses 1a & 1b

Table 7 shows the output of the regression analyses 2a t/m 2d. Model 2a explains 21.2% of ROA. The results show that the relationship between ABEM and ROA is slightly positive, and highly significant ($p<0.01$). Hypothesis 2a is therefore supported. Furthermore, are the relationships between all the control variables and ROA significant. On the other hand, the results of model 2b show that there is no significant relationship between ABEM and Tobin's Q. A significant positive or negative relationship was expected, and hypothesis 2b is therefore not supported. There is only an insignificant indication that a change in ABEM has a slightly negative effect on Tobin's Q. An alternative explanation than the expectation formed in section 2, could be the fact that ABEM does not affect cash flows, and therefore has no influence on long-term financial performance (Healy & Wahlen, 1999).

Models 2c and 2d analyse the effect of RAEM, instead of ABEM. Hypothesis 2c is rejected since RAEM has no significant influence on short-term financial performance. However, there is an indication that there is a negative effect. Next a one-tailed test was performed, but again no significant effect was found. An alternative explanation for these findings may be that one of the components of the aggregate effect of RAEM predominates and has no significant effect. Perhaps the costs for manipulating cash flows and investments were less than expected (Graham et al., 2005). In the long term, RAEM has a significant negative effect ($p<0.01$), which is in line with the expectation. Hypothesis 2d is therefore supported.

Financial firm performance				
Model	2a	2b	2c	2d
	ROA	Tobin's Q	ROA	Tobin's Q
ABEM	5.19e-10*** (4.08)	-4.54e-11 (-0.84)		
RAEM			-1.19e-10 (-1.14)	-1.52e-10*** (-3.27)
Predicted sign	+	+/-	-	-
FSIZE	-0.0140*** (-4.70)	-0.00772*** (-6.09)	-0.00441* (-1.94)	-0.00752*** (-7.46)
LEV	0.0289*** (2.62)	-0.0230*** (-4.89)	0.0348*** (3.24)	-0.0237*** (-4.99)
GROWTH	1.148*** (3.35)	0.550*** (3.77)	1.323*** (3.90)	0.603*** (4.01)
LOSS	-0.0784*** (-6.22)	-0.0274*** (-5.10)	-0.0743*** (-6.07)	-0.0301*** (-5.56)
Country	included	included	included	included
Industry	included	included	included	included
Constant	0.387*** (4.92)	1.153*** (34.40)	0.120 (1.20)	1.144*** (25.92)
Observations	804	804	774	774
Adj. R-squared	0.212	0.414	0.258	0.419
Explanation: t-statistics are below the coefficients of each variable in parentheses. Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)				

Table 7: Regression analyses of hypotheses 2a, 2b, 2c, and 2d

Table 8 shows the output of the regression analyses 3a-1 up to and including 3a-4. To test a potential mediation effect, the causal-step method, described in section 2.6.1 is used (Baron & Kenny, 1986). As the results in table 6 show, gender diversity has no effect on short-term financial performance (step 1). Furthermore, did earlier research show that gender diversity significantly influences earnings management (step 2) (Campbell & Minguez-Vera, 2008; Kaplan et al., 2009). From model 2a, it also appeared that a part of EM, namely ABEM, has a significant positive influence on ROA (step 3). Now that ABEM is being added to the models in table 8 (step 4), gender diversity appears to have a significant positive impact on ROA ($p < 0.10$). On the other hand, RAEM appears to have no effect on the relationship between gender diversity and short-term financial performance. Since the influence of gender percentage and critical mass on ROA does deviate significantly due to the presence of earnings management (ABEM), hypothesis 3a is supported. There appears to be a positive complete mediation effect on the short-term, based on the mediation of ABEM only. This does not make the support of the hypothesis any less strong, since this hypothesis was included for exploration, and there was no expectation which specific part of EM would have an influence. A complete mediation effect has been found because the relationship between gender diversity and financial performance is now significant in the short term.

Mediation effect				
Model	3a-1	3a-2	3a-3	3a-4
	ROA	ROA	ROA	ROA
Genderpercentage	0.0535* (1.84)		0.0390 (1.42)	
Criticalmass		0.0148* (1.72)		0.00972 (1.20)
ABEM	5.00e-10*** (3.92)	5.04e-10*** (3.96)		
RAEM			-1.27e-10 (-1.22)	-1.22e-10 (-1.16)
Predicted sign	+/-	+/-	+/-	+/-
FSIZE	-0.0147*** (-4.91)	-0.0142*** (-4.76)	-0.00513** (-2.20)	-0.00465** (-2.04)
LEV	0.0295*** (2.68)	0.0288** (2.61)	0.0349*** (3.25)	0.0346*** (3.22)
GROWTH	1.147*** (3.35)	1.152*** (3.37)	1.315*** (3.87)	1.325*** (3.90)
LOSS	-0.0769*** (-6.11)	-0.0782*** (-6.21)	-0.0735*** (-6.00)	-0.0742*** (-6.06)
Country	included	included	included	included
Industry	included	included	included	included
Constant	0.395*** (5.02)	0.388*** (4.93)	0.135 (1.35)	0.131 (1.31)
Observations	804	804	774	774
Adj. R-squared	0.214	0.214	0.259	0.259

Explanation: t-statistics are below the coefficients of each variable in parentheses.
Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)

Table 8: Regression analyses of hypothesis 3a

Table 9 shows that earnings management partly mediates the relationship between gender diversity and long-term financial performance. ABEM appears to have no significant influence on Tobin's Q, what was already apparent from the results of model 2b (Step 3). However, the coefficients of gender diversity and critical mass are both increasing due to the presence of ABEM. RAEM, on the other hand, does have a significant effect on Tobin's Q (model 2d). This confirms step 3 of the causal-step method (Baron & Kenny, 1986). Furthermore, it appears that RAEM increases the coefficient of gender percentage and critical mass (step 4). It can therefore be concluded that there is a partial mediation effect of earnings management on the relationship between gender diversity and long-term firm performance. For this reason, hypothesis 3b is supported. The reason that only a partial mediation effect is found, rather than a complete mediation effect, is because the effect of gender diversity on long-term financial performance was already significant and still is after including RAEM in this model.

Mediation effect				
Model	3b-1	3b-2	3b-3	3b-4
	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q
Genderpercentage	0.0442***		0.0402***	
	(3.59)		(3.34)	
Criticalmass		0.00728***		0.00614*
		(1.99)		(1.71)
Predicted sign	+	+	+	+
ABEM	-6.09e-11	-5.27e-11		
	(-1.13)	(-0.97)		
RAEM			-1.60e-10***	-1.53e-10***
			(-3.48)	(-3.31)
Predicted sign	+/-	+/-	+/-	+/-
FSIZE	-0.00834***	-0.00781***	-0.00826***	-0.00767***
	(-6.57)	(-6.17)	(-8.06)	(-7.59)
LEV	-0.0225***	-0.0230***	-0.0236***	-0.0238***
	(-4.83)	(-4.92)	(-5.01)	(-5.01)
GROWTH	0.549***	0.552***	0.594***	0.604***
	(3.79)	(3.79)	(3.98)	(4.02)
LOSS	-0.0262***	-0.0273***	-0.0292***	-0.0300***
	(-4.91)	(-5.09)	(-5.43)	(-5.55)
Country	included	included	included	included
Industry	included	included	included	included
Constant	1.160*** (34.83)	1.153*** (34.48)	1.160*** (26.32)	1.151*** (26.00)
Observations	804	804	774	774
Adj. R-squared	0.423	0.416	0.427	0.420
Explanation: t-statistics are below the coefficients of each variable in parentheses.				
Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)				

Table 9: Regression analyses of hypothesis 3b

Table 10 shows the output of the regression analyses 4-1 up to and including 4-4. These regressions test the potential moderation effect of gender diversity on the relationship between earnings management and firm performance on the short run. As expected, the effects of gender percentage and critical mass are both insignificant on the short run. Furthermore, it appears that ABEM has a highly significant effect on ROA ($p < 0.01$). RAEM on the other hand, has no significant effect. According to Baron & Kenny (1986), one of the conditions for recognizing a moderation effect is that the interaction term must be significant. In Table 10 this is only the case with Genderperc*ABEM and Genderperc*RAEM. The interaction terms that include critical mass are both not significant. The second condition is that gender diversity changes the direction or strength of the relationship between earnings management and financial performance. This also appears to be the case. The effect of ABEM on ROA changed from 5.19e-10 to 8.11e-10 and there is also an indication that the effect of RAEM changed from -1.19e-10 to 2.29e-10. However, the effects of the moderator and the predictor on the dependent variable do not necessarily have to be significant to test the moderation hypothesis (Baron & Kenny, 1986). There is a complementary moderation effect of gender diversity on the relation between earnings management and short-term financial performance.

Moderation effect	**(short-term)**			
Model	4-1	4-2	4-3	4-4
	ROA	ROA	ROA	ROA
Genderpercentage	0.0324		0.0192	
	(1.05)		(0.65)	
Criticalmass		0.0134		0.00992
		(1.55)		(1.15)
Predicted sign	+/-	+/-	+/-	+/-
ABEM	8.11e-10***	5.66e-10***		
	(4.10)	(4.11)		
RAEM			2.29e-10	-1.25e-10
			(1.05)	(-1.07)
Predicted sign	+/-	+/-	+/-	+/-
Genderperc*ABEM	-1.43e-09**			
	(-2.06)			
Genderperc*RAEM			-1.58e-09*	
			(-1.85)	
Critmass*ABEM		-2.44e-10		
		(-1.19)		
Critmass*RAEM				1.63e-11
				(0.07)
Predicted sign	+/-	+/-	+/-	+/-
FSIZE	-0.0147***	-0.0141***	-0.00571**	-0.00465**
	(-4.93)	(-4.73)	(-2.43)	(-2.03)

LEV	0.0264**	0.0270*	0.0339***	0.0347***
	(2.38)	(2.43)	(3.15)	(3.22)
GROWTH	1.092***	1.145***	1.276***	1.326***
	(3.19)	(3.35)	(3.76)	(3.90)
LOSS	-0.0780***	-0.0798***	-0.0729***	-0.0742***
	(-6.20)	(-6.31)	(-5.96)	(-6.05)
Country	included	included	included	included
Industry	included	included	included	included
Constant	0.403*** (5.13)	0.387*** (4.93)	0.155 (1.54)	0.131 (1.31)
Observations	804	804	774	774
Adj. R-squared	0.218	0.214	0.262	0.258

Explanation: t-statistics are below the coefficients of each variable in parentheses.
Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)

Table 10: Regression analyses of hypothesis 4 (short-term)

Table 11 shows the output of the regression analyses 4-5 up to and including 4-8. These regressions test the potential moderation effect of gender diversity on the relationship between earnings management and firm performance on the long run. As expected, the effect of gender percentage is significant in both model 4-5 and 4-7. However, the effect of critical mass is only significant in model 4-6. Furthermore, it appears that RAEM has a significant effect on Tobin's Q ($p < 0.05$) in model 4-8. ABEM on the other hand, has no significant effect. To grant a moderation effect, the interaction term in a particular regression analysis should be significant (Baron & Kenny, 1986). This is again only the case with the interaction terms Genderperc * ABEM and Genderperc * RAEM. As a result, there can only be a moderation effect in models 4-5 and 4-7 in Table 11. The second condition is that gender diversity changes the direction or strength of the relationship between earnings management and financial performance. There is an indication that the effect of ABEM on Tobin's Q changed from -4.54e-11 to 4.91e-11 and an indication that the effect of RAEM changed from -1.52e-10 to 6.15e-11. The effects of the moderator and the predictor on the dependent variable do not necessarily have to be significant to test the moderation hypothesis (Baron & Kenny, 1986). There is a complementary moderation effect of gender diversity on the relation between earnings management and long-term financial performance. Since there is a complementary moderation effect in the short term and long term, hypothesis 4 is supported.

Moderation effect	**(long-term)**			
Model	4-5	4-6	4-7	4-8
	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q
Genderpercentage	0.0367***		0.0279**	
	(2.82)		(2.16)	
Criticalmass		0.00701*		0.00447
		(1.90)		(1.17)
Predicted sign	+/-	+/-	+/-	+/-
ABEM	4.91e-11	-4.03e-11		
	(0.59)	(-0.69)		
RAEM			6.15e-11	-1.25e-10**
			(0.64)	(-2.42)
Predicted sign	+/-	+/-	+/-	+/-
Genderperc*ABEM	-5.06e-10*			
	(-1.72)			
Genderperc*RAEM			-9.80e-10***	
			(-2.63)	
Critmass*ABEM		-4.90e-11		
		(-0.56)		
Critmass*RAEM				-1.32e-10
				(-1.26)
Predicted sign	+/-	+/-	+/-	+/-
FSIZE	-0.00835***	-0.00779***	-0.00862***	-0.00770***
	(-6.58)	(-6.14)	(-8.37)	(-7.62)
LEV	-0.0236***	-0.0234***	-0.0243***	-0.0242***
	(-5.02)	(-4.94)	(-5.15)	(-5.08)
GROWTH	0.530***	0.551***	0.570***	0.602***
	(3.65)	(3.78)	(3.83)	(4.01)
LOSS	-0.0266***	-0.0276***	-0.0289***	-0.0302***
	(-4.99)	(-5.12)	(-5.38)	(-5.58)
Country	included	included	included	included
Industry	included	included	included	included
Constant	1.163*** (34.92)	1.153*** (34.46)	1.172*** (26.56)	1.150*** (26.01)
Observations	804	804	774	774
Adj. R-squared	0.425	0.416	0.432	0.421

Explanation: t-statistics are below the coefficients of each variable in parentheses.

Significance of coefficients is indicated by no. of asterisks (* p<0.10, ** p<0.05, *** p<0.01)

Table 11: Regression analyses of hypothesis 4 (long-term)

4.4 Robustness checks

During this research, a robustness check was already built in, into all regression analyses where the effect of gender diversity was measured. There were two proxies used for measuring gender diversity included in this study: the percentage of women in relation to the total number of board members and the critical mass within a board. The model appears to be robust since critical mass indicates a comparable coefficient and the same direction as the gender percentage. However, in some cases the effect of critical mass is slightly less significant. As a second robustness check, a few regressions were performed where Tobin's Q was measured by market value divided by book value. This differs from the method used as standard (MV / Asset replacement costs). These results were largely the same in terms of quality and it therefore does not add any information to enclose these tables. The third robustness check is carried out to test whether the effect of gender diversity on financial performance remains the same in countries outside the sample, namely countries where a quota with sanctions applies. The purpose of this is to make a suggestion for future research. This split makes it possible to clarify the differences between social (sample 1) and legal pressure (sample 2). The countries concerned are; France, Germany and Norway. Table 3 includes the details about these three countries which have a gender quota set by the government. The data has also been tested for normality, multicollinearity and influential cases for this robustness check. Only hypotheses 1a and 1b will be tested again.

Financial firm performance				
Model	1a-1	1a-2	1b-1	1b-2
Sample 2	ROA	ROA	Tobin's Q	Tobin's Q
Genderpercentage	-0.140* (-2.53)		-0.0340 (-1.60)	
Criticalmass		-0.0438* (-2.37)		-0.0142* (-2.01)
Predicted sign	none	none	+	+
FSIZE	0.0160*** (4.54)	0.0142*** (4.20)	-0.00704*** (-5.20)	-0.00740*** (-5.72)
LEV	-0.0180*** (-3.45)	-0.0170** (-3.26)	-0.00919*** (-4.58)	-0.00893*** (-4.46)
GROWTH	0.267* (2.08)	0.262* (2.03)	0.00939 (0.19)	0.00677 (0.14)
LOSS	-0.0970*** (-5.97)	-0.0976*** (-6.01)	-0.0137* (-2.20)	-0.0136* (-2.19)
Country	included	included	included	included
Industry	included	included	included	included
Constant	-0.280** (-3.09)	-0.256** (-2.82)	1.114*** (32.02)	1.121*** (32.25)
Observations	340	340	340	340
Adj. R-squared	0.409	0.408	0.483	0.485
Explanation: t-statistics are below the coefficients of each variable in parentheses. Significance of coefficients is indicated by no. of asterisks (* p<0.05, ** p<0.01, *** p<0.001)				

Table 12: Regressions of sensitivity analysis

There appear to be a number of different results from the regressions between sample 1 and sample 2. In sample 1, gender percentage has no significant effect on short term financial performance (as expected), and in sample 2 (countries with a quota) a significant negative effect. This time the effect of gender percentage on Tobin's Q becomes insignificant. Furthermore, a critical mass now has a significant effect on both short-term and long-term performance ($p < 0.05$). This effect is negative, which was not expected. So, hypotheses 1a and 1b are both not supported in sample 2. This is an interesting finding that can be further investigated. It seems that the mandatory quota ensures that companies underperform in the short term. It may be possible to investigate whether a quota is really effective. Future research should review literature, mainly focused on countries with quotas. Aspects that have not been taken into account yet, seem to play a major role. With the help of additional research, there could be controlled for those relevant factors.

5. Conclusion

This study examines the ways in which board diversity affect financial performance of European companies in the short and long run. In addition, the mediating effect of earnings management on the relationship between gender diversity and financial performance, and the moderation effect of gender diversity on the relationship between earnings management and financial performance are examined. The dataset used for testing the hypotheses, includes 10 countries of the European Economic Area (EEA) without a gender quota. In addition are 3 countries of the EEA with a gender quota analysed in order to compare the results of the effect of gender diversity on financial performance alone. In this section an answer is given to the research question:

"*What is the influence of gender diversity in corporate boards and earnings management on short- and long-term financial performance of European firms?"*

This question can be answered based on existing literature and analyses results of this study. Nine hypotheses were formed to answer this research question. Based on prior research, no relationship between gender diversity and short-term financial performance is expected. The results of the analyses of this study are in line with this. In the long term, a positive relationship between gender diversity and financial performance is expected due to skills, ethical behaviour, risk-aversion, decision-making and creativity of female board members. The corresponding analysis in this study, confirms this expectation. In addition, the results show that the relationship between ABEM and short-term financial performance is slightly positive. However, a significant effect of ABEM on long-term financial performance is not found. This could be due to the fact that ABEM does not affect cash flows, and therefore has no influence on long-term financial performance (Healy & Wahlen, 1999). There is only an insignificant indication that a change in ABEM has a slightly negative effect on long-term financial performance.

Also, for the relationship between RAEM and short-term financial performance, no significant results are measured. There is only an indication that RAEM has a negative effect on short-term financial performance. An explanation for these unexpected findings may be that one of the components of the aggregate effect of RAEM predominates and has no significant relationship with ROA. However, in the long term, RAEM does appear to have a significant negative effect, which was expected. Another finding from this study is that earnings management mediates the relationship between gender diversity and financial performance in the short and long term. There appears to be a positive complete mediation effect in the short term, based on the mediation of ABEM only. In the long term, RAEM affects the coefficients of both proxies of gender diversity positively. Since the relationship was already significant, only a partial mediation effect on the long-term is measured.

The last main finding concerns the moderation effect of gender diversity. There is a complementary moderation effect of gender diversity on the relation between earnings management and short-term financial performance. The effect of ABEM on short term performance increased, when the moderator and the moderation term were included in the model. There is also an indication that the effect of RAEM increased. However, the effects of the moderator and the predictor on the dependent variable do not necessarily have to be significant to test the moderation hypothesis (Baron & Kenny, 1986). In addition, there also appears to be a complementary moderation effect in the long term. There are indications that the effects of ABEM and RAEM on long term performance increase, due to adding of the moderation term, which is significant too. Further investigation is required since the examination of details regarding a moderation effect is not within the scope of this study.

There are also some limitations to this research. First of all, the study is limited to one year, 2017. As a result, trends cannot be corrected and one-time events can have major consequences for the results. A panel data analysis could limit this problem. The second limitation concerns the number of control variables that are included. The analyses take into account the control variables which are used in similar studies. However, something could be overlooked and that may explain why none of the models is able to predict more than about 40% of the dependent variable. Of course, there is also the risk of errors in the data, which has been limited as far as possible by testing various assumptions. Furthermore, the robustness check shows that the model does not retrieve the expected results if only countries with a gender quota are measured. However, this is not due to the robustness of the model since the other two tests do show a robust model. Yet this remains an interesting finding since it is not in line with the theory from section 2. Further research into the performance of countries with a quota is required. However, these limitations do not constrain the filling of the research gap. Legislative powers in European countries benefit from this study, since it appears that setting a quota is not necessarily beneficial (section 4.4). The results are important for investors and firms themselves as well. Firms can benefit more when a gender quota is promoted, but when there are no sanctions if quotas are not met (according to the robustness check). Firms can already take action themselves by hiring more women with the aim of a higher financial performance. In addition, can firms appoint women to boards of directors in order to limit earnings management, since gender diversity has a negative impact on earnings management. If it is clear for investors that there are more women on the board of directors, and these investors are aware of the negative relation with earnings management, there may be higher chances that they will invest in a particular company. The latter is only an assumption, but perhaps an interesting future research topic. Moreover, could future research focus on an aggregate dataset for all economic developed countries in the world, instead of just Europe of the USA separately. In that way, the results will be more generalizable. Besides, should data from multiple years be used in order to control for trends. The research on mediating and moderating effects of gender diversity and earnings management should be examined in more detail as well, since these aspects within this study were only exploratory.

References

Achleitner, A., Günther, N., Kaserer, C., & Siciliano, G. (2014). Real earnings management and accrual-based earnings management in family firms. *European Accounting Review, 23*(3), 431-461.

Adams, R., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of financial economics, 94*(2), 291-309.

Al-Matari, E., Al-Swidi, A., & Fadzil, F. (2014). The Measurements of Firm Performance's Dimensions. *Asian Journal of Finance & Accounting, 6*(1), 1-26.

Arun, T., Almahrog, Y., & Aribi, Z. (2015). Female directors and earnings management: Evidence from UK companies. *International Review of Financial Analysis, 39*, 137-146.

Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics, 116*(1), 261-292.

Baron, R., & Kenny, D. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology, 51*(6), 1173-1182.

Becker, C., DeFond, M., Jiambalvo, J., & Subramanyam, K. (1998). The effect of audit quality on earnings management. *Contemporary accounting research, 15*(1), 1-24.

Borghans, L., Heckman, J., Golsteyn, B., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association, 7*(2-3), 649-658.

Campbell, K., & Minguez-Vera, A. (2008). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics, 83*, 435-451.

Campbell, K., & Minguez-Vera, A. (2010). Female board appointments and firm valuation: Short and long-term effects. *Journal of Management & Governance, 14*(1), 37-59.

Carter, D., D'Souza, F., Simkins, B., & Simpson, W. (2010). The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review, 18*(5), 396-414.

Chapple, L., & Humphrey, J. (2014). Does board gender diversity have a financial impact? Evidence using stock portfolio performance. *Journal of business ethics, 122*(4), 709-723.

Childs, S., & Krook, M. (2006). Should feminists give up on critical mass? A contingent yes. *Politics & Gender, 2*(4), 522-530.

Cohen, D., & Zarowin, P. (2010). Accrual-based and real earnings management activities around seasoned equity offerings. *Journal of accounting and Economics, 50*(1), 2-19.

Cohen, D., Dye, A., & Lys, T. (2008). Real and accrual-based earnings management in the pre and post-sarbanes- oxley periods. *The Accounting Review, 83*(3), 757–787.

Dahlerup, D. (2006). The story of the theory of critical mass. *Politics & Gender, 2*(4), 511-522.

Dallas, L. (2001). The new managerialism and diversity on corporate boards of directors. *Tulane Law Review, 76*, 1363-1405.

Dalton, D., & Ortegren, M. (2011). Gender differences in ethics research: The importance of controlling for the social desirability response bias. *Journal of Business Ethics, 103*(1), 73-93.

Dechow, P., & Dichev, I. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review, 77*, 35–59.

Dechow, P., & Skinner, D. (2000). Earnings Management: Reconciling the Views of Accounting Academics, Practitioners, and Regulators. *Accounting Horizons, 14*(2), 235-250.

Dechow, P., Sloan, R., & Sweeney, A. (1995). Detecting earnings management. *Accounting review, 70*(2), 193-225.

Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications, 40*(10), 3970-3983.

Dezsö, C., & Ross, D. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal, 33*(9), 1072-1089.

Dyreng, S., Hillegeist, S., & Penalva, F. (2012). Earnings management to avoid debt covenant violations and future performance (Working paper). *Duke University, Arizona State University and I.E.S.E.*

Eckel, C., & Grossman, P. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results, 1*, 1061-1073.

Enomoto, M., Kimura, F., & Yamaguchi, T. (2015). Accrual-based and real earnings management: An international comparison for investor protection. *Journal of Contemporary Accounting & Economics, 11*(3), 183-198.

Erhardt, N., Werbel, J., & Shrader, C. (2003). Board of director diversity and firm financial performance. *Corporate governance: An international review, 11*(2), 102-111.

European Commission. (2016). *Gender balance on corporate boards > Europe is cracking the glass ceiling.* Retrieved March 25, 2019, from European Commission: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwiQnfneuJ_hAhWnPOwKHb6oDswQFjAAegQIAhAC&url=https%3A%2F%2Fec.europa.eu%2Fnewsroom%2Fdocument.cfm%3Fdoc_id%3D46280&usg=AOvVaw3vS5Ma4NE4q43gpoR3zaLP

Financial Times. (2015). *Companies with female executives perform better, says research.* Retrieved April 9, 2019, from Financial Times: https://www.ft.com/content/96f65730-65da-11e5-a57f-21b88f7d973f

Forbes. (2016). *We Talk About Gender Diversity, But What We Really Mean Is This.* Retrieved April 10, 2019, from Forbes: https://www.forbes.com/sites/christopherskroupa/2016/12/02/we-talk-about-gender-diversity-but-what-we-really-mean-is/

Glaum, M., Schmidt, P., Street, D., & Vogel, S. (2013). Compliance with IFRS 3-and IAS 36-required disclosures across 17 European countries: company-and country-level determinants. *Accounting and business research, 43*(3), 163-204.

Graham, J., Harvey, C., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of accounting and economics, 40*(1-3), 3-73.

Grey, S. (2006). Numbers and beyond: The relevance of critical mass in gender research. *Politics & Gender, 2*(4), 492-502.

Guidry, F., Leone, A., & Rock, S. (1999). Earnings-based bonus plans and earnings management by business-unit managers. *Journal of accounting and economics, 26*(1-3), 113-142.

Hayes, A. (2019). *Q Ratio (Tobin's Q Ratio) Definition.* Retrieved March 26, 2019, from Investopedia: https://www.investopedia.com/terms/q/qratio.asp

Healy, P., & Wahlen, J. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting horizons, 13*(4), 365-383.

Ibrahim, H., & AbdulSamad, F. (2011). Corporate governance mechanisms and performance of public-listed family-ownership in Malaysia. *International Journal of Economics and Finance, 3*(1), 105–115.

Joecks, J., Pull, K., & Vetter, K. (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a "critical mass?". *Journal of business ethics, 118*(1), 61-72.

Kaplan, S., Pany, K., Samuels, J., & Zhang, J. (2009). An examination of the association between gender and reporting intentions for fraudulent financial reporting. *Journal of Business Ethics, 87*(1), 15-30.

Kenton, W. (2018). *Earnings Management*. Retrieved March 26, 2019, from Investopedia: https://www.investopedia.com/terms/e/earnings-management.asp

Krishnan, G., & Parsons, L. (2008). Getting to the Bottom Line: An Exploration of Gender and Earnings Quality. *Journal of Business Ethics, 78*, 65–76.

Lindenberg, E., & Ross, S. (1981). Tobin's q ratio and industrial organization. *Journal of business*, 1-32.

Marimuthu, M., & Kolandaisamy, I. (2009). Ethnic and gender diversity in boards of directors and their relevance to financial performance of Malaysian companies. *Journal of sustainable development, 2*(3), 139-148.

Marinova, J., Plantenga, J., & Remery, C. (2016). Gender diversity and firm performance: Evidence from Dutch and Danish boardrooms. *The International Journal of Human Resource Management, 27*(15), 1777-1790.

Nichols, D., & Wahlen, J. (2004). How do earnings numbers relate to stock returns? A review of classic accounting research with updated evidence. *Accounting Horizons, 18*(4), 263-286.

Nuryanah, S., & Islam, S. (2011). Corporate governance and performance: Evidence from an Emerging Market. *Malaysian Accounting Review, 10*(1), 17–42.

Phillips, J., Pincus, M., & Rego, S. (2003). Earnings management: New evidence based on deferred tax expense. *The Accounting Review, 78*(2), 491-521.

Reuters. (2019). *Diversity in the 'man cave': Boardrooms gain women as minorities lag.* Retrieved April 9, 2019, from Reuters: https://www.reuters.com/article/us-usa-directors-diversity/diversity-in-the-man-cave-boardrooms-gain-women-as-minorities-lag-idUSKCN1Q20E8

Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of accounting and economics, 42*(3), 335-370.

Senden, L. (2018). *"Een vrouwenquotum gaat om het rechtzetten van discriminatie".* Retrieved May 7, 2019, from Utrecht University: https://www.uu.nl/nieuws/een-vrouwenquotum-gaat-om-het-rechtzetten-van-discriminatie

Shrader, C., Blackburn, V., & Iles, P. (1997). Journal of Managerial Issues Women In Management And Firm Financial Performance: An Exploratory Study. *Journal of managerial issues, 9*(3), 355-372.

Srinidhi, B., Gul, F., & Tsui, J. (2011). Female directors and earnings quality. *Contemporary Accounting Research, 28*(5), 1610-1644.

Stevens. (2000). *Interaction Effects in Regression.* Retrieved May 17, 2019, from Radboud University: https://www.ru.nl/publish/pages/780171/interaction.pdf

Sudarsanam, S. (2003). Creating value from mergers and acquisitions: The challenges. London: Financial Times Prentice Hall.

Torchia, M., Calabrò, A., & Huse, M. (2011). Women directors on corporate boards: From tokenism to critical mass. *Journal of Business Ethics, 102*(2), 299-317.

Utrecht University. (2018). *"Een vrouwenquotum gaat om het rechtzetten van discriminatie".* Retrieved April 9, 2019, from Utrucht University: https://www.uu.nl/nieuws/een-vrouwenquotum-gaat-om-het-rechtzetten-van-discriminatie

Wang, M., & Kelan, E. (2013). The gender quota and female leadership: Effects of the Norwegian gender quota on board chairs and CEOs. *Journal of business ethics, 117*(3), 449-466.

Wernerfelt, B., & Montgomery, C. (1988). Tobin's q and the importance of focus in firm performance. *The American Economic Review*, 246-250.

Zang, A. (2012). Evidence on the trade-off between real activities manipulation and accrual-based earnings management. *The Accounting Review, 87*(2), 675-703.

Abbreviations

ABEM:	Accrual-based earnings management
CFO:	Cashflow from operations
EEA:	European Economic Area
EM:	Earnings Management
PPE:	Property, Plant and Equipment
PROD:	Production
REAM:	Real activities earnings manipulation
ROA:	Return on Assets

Appendices

Appendix A: Data-checks

Normality

All variables used in this study have been tested for normality to enable reliable regressions. This has been tested by visual analysis. Most used variables are normally distributed. However, were multiple variables found not to be normally distributed according to the plotted histograms. These variables have been transformed to their logarithm to make the distribution normal, since none of the unnormal distributed variables were scaled. Scaled numbers, such as Tobin's Q and ROA in this study, cannot be logged. This because numbers between 0 and 1 turn negative if logged. The variables which are transformed are: market value, net revenues, total assets, operational income, cash flow from operation activities, total debt, net receivables and PPE. After transforming, another check was done by generating the histograms again. This time, all variables were found to be normally distributed. After this, the required independent variables, whose distributions were not normally distributed in the first instance, were created again. The logarithmic transformation is the reason why the numbers in table 4 are less intuitive to interpret.

Homoskedasticity

In order to test for homoskedasticity, a Breusch-Pagan test was executed after every regression. At every test, the null-hypothesis which expects that there is no significant sign of heteroskedasticity, was rejected. Therefore, it does not add any relevant information to include all figures of these tests in this appendix. To solve the problem of heteroskedasticity, the regressions were executed using the robust function in STATA.

Appendix B: Variance inflation factor test (VIF)

VIF-test Variable	VIF	1/VIF
FSIZE	2.35	0.426168
ABEM	3.27	0.306049
Genderperc	2.62	0.381746
Critmass	2.47	0.405134
RAEM	2.04	0.489582
Tobin's Q	1.31	0.762901
LOSS	1.16	0.865004
LEV	1.09	0.914218
GROWTH	1.06	0.939999
Mean VIF	1.93	