More Women, Better Performance? – The Effect of Gender Quotas on Firm Performance

In the past decade, gender diversity has turned into a prominent area of debate in the world of business. Societal Pressure led the European Union to encourage the adoption of quotas to increase the share of women in the upper management. In this paper, I investigate the difference in the effect between a hard and soft quota on firm performance from 2011 until 2019. I use a Difference-in-Difference method where German listed firms represent the treatment and Dutch listed firms represent the control group. My results show that the hard quota significantly increases the share of women on the board and that it causes a performance-reducing effect. Nevertheless, other unobserved factors may be adding to the effect of the hard quota, which makes the interpretation ambiguous.

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Bachelor Thesis – International Bachelor in Economics and Business Economics

Name: Lucas Semprimoschnig
Student Number: 502634

Supervisor: Sam Hoey

Second Assessor: Max Coveney

Date Final Version: 09/07/2021

Table of Contents

Introduction

In recent years, gender diversity in board rooms entered the spotlight – changing norms and societal pressure motivated governments and firms to tackle the issue of female underrepresentation in the upper management. In fact, 62% of the European firms had at least one female executive in 2004, however, only 28% had more than one (Adams and Ferreira, 2009). Similarly, 65% of US firms had at least one female executive in 2003, whereas only 25% of them had more than one (Adams and Ferreira, 2009). This alarmed European states and incentivized the European Union to set concrete guidelines on the integration of more women into both the private and public sector (European Institute for Gender Equality, 2019). In the past 18 years, 10 countries introduced a statutory quota for women in board positions (Arndt and Wrohlich, 2019). Norway pioneered a quota of 40% of female board members for publicly traded and state-owned businesses since 2003 (Arndt and Wrohlich, 2019). Spain, Iceland Italy, Belgium and France quickly followed with a statutory quota as well. Since 2015, Germany, Portugal, Austria and the Netherlands (only in 2021) were added to the block of countries that have a statutory quota in place (AD, 2019; Arndt and Wrohlich, 2019). Despite all having a statutory quota, these countries differ in the percentage of female executives on the board and the corresponding sanctions for non-compliance. For instance, Germany and Austria have a quota of 30%, whereas France and Spain both instituted a quota of 40% (Arndt and Wrohlich, 2019).

Some countries use a 'soft quota' instead of a hard quota, which only serves as a strong recommendation for firms to reach a target value. Since is not obligatory for firms, noncompliance does not lead to sanctions. Scholars extensively researched the effect of the board composition on firm performance, but it remains unclear whether a hard quota proves to have a more negative or positive, let alone a different, effect than a soft quota on firm performance. A priori, a quota would influence the performance of a firm through its effect on the composition of the board. If the composition changes, tasks are carried out differently and would have different effects on firm performance. Thus, my paper examines the following question:

"Does the effect of a hard quota on firm performance differ from the effect of a soft quota on firm performance?"

This research is scientifically relevant because it takes into account the differences of a soft and hard quota and investigates the efficiency and effect of both quotas. Numerous parties claim a hard quota to be more efficient, but proof is scarce. It is socially relevant for the business environment and policymakers because of the implications I derive from my research. A difference in the effect may imply for policymakers to rethink their policies, as well as it implies for businesses to argue in favor or against a particular quota.

This paper gives an overview of the theoretical and empirical evidence of the effect of gender quotas on firm performance in the next section. Thereafter, I explain the data selection and methodology and analyze the dataset for the statistical method I intend to use. Subsequently, I present the results and discuss them thereafter.

Theoretical Framework

If one looks at gender quotas, three types of countries can be identified: Countries with a hard quota, countries with a soft quota and countries with no quota at all. In total, 10 countries belong to the first group, which can be split up further into three subcategories: First, four countries have a statutory quota in place alongside some hard sanctions. Examples of such sanctions include forced dissolution of the board or high monetary penalties. Second, three countries have a quota with moderate sanctions, which often refers to leaving the seat on the board empty until a female replacement is found. Third, three countries have a quota without any sanctions in place (Arndt and Wrohlich, 2019).

Roughly a dozen countries have no statutory quota in place, but instead use soft quotas to encourage businesses to reach a certain percentage of female executives on boards. Firms that ought to abide by soft quotas face no sanctions in case the targets are not met. Lastly, several Eastern and Southeastern European countries still have neither a quota nor recommendations in place (Arndt and Wrohlich, 2019).

The main tasks of corporate boards remain uniform across countries, though, and can be summarized in the following way: Supervising managers, providing information and counsel to managers, monitoring compliance with applicable laws and regulations, and connecting the firm to the external environment (Carter et al., 2010). As mentioned before, these tasks have direct consequences on firm performance and are influenced by the composition of the board. Thus, a more diverse board affects the decisions differently than a less diverse board, which influences the performance of the firm. Henceforth, the link between gender diversity and firm performance exists, however, no a priori knowledge predicts how the two are related to each other. Below, I will demonstrate the four prominent theories used in the literature to explain the relationship between gender diversity and firm performance (Carter et al., 2010). Every theory is based on different merits, thus, also predicts a different nature of the relationship.

Resource Dependence Theory

Developed by Pfeffer and Salancik (1978), the resource dependence theory asserts that boards bridge the gap between the firm and other external organizations. It benefits both organizations through the exchange of hard (material) and soft (knowledge) resources. The exchange of soft resources originates from labor mobility, for instance, when executives bring unique information with them that is beneficial for the firm. The degree to which such information is of added value depends on the circumstances and environment the firm operates in, but a broader range of resources will generally set up the firm to improve its financial performance in the long-term (Hillman et al., 2000). Diverse boards also enable access to a greater external talent pool, through which boards obtain different perspectives and uncommon approaches to enhance their decision-making process. Women may have a better understanding of certain concepts than men, which can ameliorate the problem-solving capacity of the firm. Therefore, this theory suggests a positive relationship between gender diversity and firm performance.

Human Capital Theory

Becker (1964) states in his Human Capital Theory that every individual has a unique stock of knowledge, skills and experience that can benefit an organization. Specifically, women may have different qualities than men that can be beneficial for firm performance. Often, the question of whether women have the necessary 'capital' (or qualities) to be in the upper management of firms is asked in connection to this theory. Terjesen et al. (2009) show, however, that women have the same qualities, including educational qualities, for executive positions. Women are only less likely to have the same experience as male business experts (Terjesen et al., 2009).

The contingency theory underlies this theory since certain qualities or skills may be more relevant for a firm at a particular moment, but not for other firms at that time or for the same firm at different points in time. Still, this theory suggests a positive relationship between diversity in boards and firm performance because of the unique capital one adds to the firm.

Agency Theory

The Agency Theory, conceived by Jensen and Meckling (1976), illustrates the relationship between a principal and his/her agent, who carries out actions for the principal. Carter et al. (2003) underline the monitoring and controlling role of the board on the executive team. To resolve the agency issues between the board (principal) and the executive team (agent), an independent board is needed because members will likely not shirk and monitor the executives thoroughly. In other words, a more diverse board will be relatively more reliable and monitor in an unbiased way. However, overmonitoring can also decrease shareholder value as it can disrupt the flow of the firm's operations. So, no clear direct link between the firm's financial performance and board diversity can be established.

Social Psychology Theory

The Social Psychology Theory questions whether demographic minorities can affect decisions in groups (Westphal and Milton, 2000). Majorities can crowd out the influence of minorities on group decisions because of their status, which infers that the board remains unaffected by a slight increase in group diversity. Diverse groups can increase social cohesion since differences in opinions and perspectives make decision-making more time-consuming and ineffective. On the other hand, minorities can foster divergent problem-solving through diverse perspectives and critical approaches, as mentioned before by the Resource Dependence Theory. Thus, this theory predicts a negative as well as a positive relationship between gender diversity and firm performance.

The section shows that a link between gender diversity and firm performance is plausible, but that the nature of the relationship remains ambiguous as different theories predict either positive or negative relationships. I will now turn to the empirical evidence, which is as diverse as the predictions of the theories.

Empirical Evidence

Despite the vast amount of debate this topic received recently, I will only focus on the most prominent papers that examine the exact relationship between board gender diversity and firm performance. Shrader et al. (1997) pioneered this field of research on a sample of US firms and find no significant relationship between the percentage of women on the board and a handful of accounting-based measures (representing firm performance). Similarly, Du Rietz and Henrekson (2000) find no significant relationship in their Swedish sample as well. Carter et al. (2003) show that a positive relationship between the share of women and Tobin's Q for their sample of US listed firms exists. Smith et al. (2006) investigate the relationship between gender diversity and numerous accounting-based measures on a sample of Danish firms and find no significant relationship. Rose (2007) affirms an insignificant relationship between gender diversity and Tobin's Q while using a different Danish sample. The same relationship is investigated by Campbell and Mínguez-Vera (2007) on a sample of Spanish firms, who find a significant, positive relationship. Adams and Ferreira (2009) use an IV panel regression method on a sample of US firms and find a significant, negative relationship. Haslam et al. (2010) analyze a sample of British firms and find a significant, negative relationship when using Tobin's Q as the dependent variable, but an insignificant relationship when using accounting-based measures as dependent variables. Based on a three-stage least squares approach, Carter et al. (2010) find no significant link between board gender diversity and firm performance. Liu et al. (2014) find a positive effect of gender diversity on firm performance based on a large sample of Chinese listed firms. Marinova et al (2016) merge a sample of Dutch and Danish firms and find no relationship between board diversity and firm performance.

All the papers above only investigate correlations between gender diversity and firm performance, however, the majority of them are prone to (1) omitted variable bias, (2) selection bias or (3) do not specifically address the effect of a quota on firm performance. Thus, the only two papers I could locate on the effect of a quota were Matsa and Miller (2013) and Yang et al. (2019). Yang et al. (2019) vary the approach by Matsa and Miller (2013) slightly using the Difference-in-Difference method for Norway in 2003. The Norwegian firms affected by the instituted quota are compared to firms from other Nordic countries, such that the quota represents the intervention that distinguishes the Norwegian treatment group from the Nordic control group. It is the first paper that establishes a causal perspective and shows that the hard quota negatively affects the performance of firms.

Based on the empirical evidence and the theoretical foundation, I hypothesize the following:

H1: The hard quota will significantly increase the share of women on the boards compared to the soft quota.

The EU claims statutory quotas to be more efficient than recommended target values in increasing the share of women on the boards (Arndt and Wrohlich, 2019). In fact, countries with a hard quota saw a fivefold increase in the share of women on the boards, whereas in countries without a legal quota the share of women on boards increased from 11% to 17%. Therefore, I believe in a greater increase of the number of women on the boards with a hard quota. Secondly, I hypothesize the following:

H2: The effect of the hard quota on firm performance will be significantly stronger than the effect of a soft quota.

I believe that the effect of a hard quota will be different from the effect of a soft quota because of the different rules and sanctions firms in both groups have to comply with. Firms can freely choose to reach the recommended percentage, hence, I would question whether one can attribute any effect on firm performance to the soft quota because several other economic and non-economic factors may play an unobserved role. Therefore, if any effect exists, I believe it to be stronger for the hard quota, which means either more negative or more positive. Lastly, I hypothesize the following:

H3: The effect of the hard quota on firm performance will be significantly more negative than the effect of a soft quota.

Not only do I believe the effect of the hard quota to be stronger, but also more negative. Being obliged to add female executives to the board forces a change in the board that can disrupt the internal dynamics and qualities. As the Social Psychology Theory predicts, the addition of new (female) members can improve firm performance through including divergent problem-solving skills, but can also decrease firm performance through social cohesion within the board.

Data

I analyze the difference between the effect of a soft and a hard quota on firm performance based on a sample of German and Dutch listed firms. First, I retrieve financial and other firm-related information from Orbis, a database by Bureau van Dijk that gives insight into yearly data from financial statements. I select Dutch firms from the AEX, AMX and the AsCX for which the soft quota applies and German firms from the DAX, TecDax and MDAX for which the hard quota applies. Orbis does not contain information on the SDAX, hence, I exclude these firms from the analysis. Unlisted firms are also not considered in this paper. For instance, large, unlisted firms in Germany have to abide by objectives (but not a quota) to foster gender diversity on boards, which goes beyond the scope of this research (Deutscher Bundestag, 2015). The financial information I retrieve comprises Tobin's Q, the Return on Assets (ROA) using Net Income, the Return on Equity (ROE) using Net Income, the Total Assets, the Sales, the the Leverage Ratio, the Net Debt Level, the Firm Age and the Number of Employees (Adams & Ferreira, 2009).

I use BoardEx that provides information on the aggregate number of executives on all boards. It is the sum of directors in the executive and supervisory board. A company that has both boards has a so-called two-tier system, where the supervisory board represents the shareholder's interests and monitors the executive board. Besides the monitoring role, the latter is also appointed by the former. The Dutch government instituted a soft quota on both the "Raad van Bestuur" (Executive Board) and the "Raad van Commissarissen" (Supervisory Board), whereas the German hard quota applies to the "Aufsichtsrat" (Supervisory Board) exclusively. Therefore, I can only use the information on BoardEx for Dutch firms and collect information on German supervisory boards from annual reports to get the number of female executives and the respective total number of executives on the supervisory board. In the end, 10 Dutch firms lack gender information in BoardEx, for which I manually look up the values from annual reports.

I merge both datasets based on the company and the respective year from 2011 to 2019. The year 2020 is not considered because I fear that the Covid-19 Pandemic may add undesired, unobserved effects into the analysis which I cannot account for. Similar to Yang et al. (2019) and Matsa and Miller (2013), I exclude firms active in the financial sectors (banks) because I only examine firms with complete information on the selected variables. Furthermore, I remove roughly a dozen firms that went public after 2011 as these made information public after 2011 and not from 2011 onwards. I end up having a sample of 61 German listed and 33 Dutch listed firms with 846 observations in total.

In Table 2.1, I provide the descriptive statistics of my data and in Table 2.2 the correlation matrix of the variables. In the following subsections, I explain the variables used in my analysis and briefly discuss the corresponding statistics and correlations.

Table 2.1

Summary Statistics

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Ln (Tobin's Q)	846	-0.181	0.829	-2.465	2.456
ROA	846	4.789	7.900	-69.264	66.113
ROE	846	8.863	33.988	-656.566	104.895
Treatment	846	0.288	0.453	0.000	1.000
Treatment_Group	846	0.649	0.478	0.000	1.000
Share of Women	846	0.189	0.134	0.000	0.857
Total Assets	846	21300.000	44600.000	22223.970	340000.000
Sales	846	15200.000	29100.000	3.880	197000.000
Firm Age	846	77.830	51.302	15.000	175.000
Leverage Ratio	846	0.824	0.287	0.096	2.672
Number of Employees	846	43053.260	84039.860	17.000	547459.000
Net Debt	846	4954.336	16100.000	-6710.848	152000.000
Number of Directors	846	10.545	5.408	3.000	31.000

Note: This table contains the summary statistics of the performance measures, the variables of interest, the firm characteristics, and the control variable. In total, I use 846 observations. The ROA, the ROE, the Share of Women and the Leverage Ratio are reported in percentage points. Total Assets, Sales and Net Debt are reported in million dollars and Firm Age in years.

Table 2.2

Correlation Matrix

Variable	Ln (Tobin's Q)	ROA	ROE	Treatment	Treatment Group	Share of Women	Total Assets	Sales	Firm Age	Leverage Ratio	# Employees	Net Debt	# Directors
Ln (Tobin's Q)	1.000												
ROA	0.354	1.000											
ROE	0.181	0.563	1.000										
Treatment	0.077	0.027	0.007	1.000									
Treatment_Group	0.050	0.067	0.080	0.468	1.000								
Share of Women	-0.035	-0.042	0.004	0.327	0.082	1.000							
Total Assets	-0.327	-0.071	-0.017	0.144	0.239	0.144	1.000						
Sales	-0.318	-0.066	-0.009	0.107	0.241	0.144	0.945	1.000					
Firm Age	-0.119	0.037	0.057	-0.004	-0.009	0.112	0.056	0.092	1.000				
Leverage Ratio	-0.447	-0.267	-0.173	0.025	0.063	0.054	0.276	0.325	0.095	1.000			
Number of Employees	-0.205	-0.039	0.020	0.106	0.180	0.178	0.643	0.753	0.088	0.288	1.000		
Net Debt	-0.281	-0.066	0.006	0.129	0.179	0.106	0.909	0.799	-0.027	0.165	0.481	1.000	
Board Size	-0.397	-0.109	-0.059	0.193	0.379	0.229	0.607	0.629	0.199	0.405	0.567	0.432	1.000

Note: This table contains the correlations between the performance measures, the variables of interest, the firm characteristics and the control variable.

Firm Performance

In essence, two different categories exist for the performance measures: On the one hand, accounting-based measures, which are rather objective and backward-looking. On the other hand, market-based measures, that are rather subjective and forward-looking (Haslam et al., 2010). Accounting-based measures are heavily scrutinized by auditors to provide the market with complete, correct numbers on past performance. Market-based measures, however, are influenced by the beliefs and perspectives of investors and provide insight into the future potential of the firm (Campbell & Mínguez-Vera, 2007; Yang et al., 2019). Since these types of measures are fundamentally different, I decided to include both in the analysis.

As accounting-based measures, I use the Return on Assets (ROA) and the Return on Equity (ROE) based on net income before extraordinary items and discontinued operations. The denominator of the ROA is equal to the book value of the total assets and the denominator of the ROE to the book value of the total stockholder's equity. Essentially, both measures illustrate how well the company generates income from investments, only using a different reference base. In both cases, the higher the number, the more efficient it is using its resources to achieve higher earnings (Investopedia, 2021). A negative number, however, need not be worrisome because essentially the firm may be restructuring. In Table 2.1, the means of both the ROA and the ROE indicate an average efficient usage of the assets to generate income. The ROA reaches a minimum of -69% and a maximum of 66%, whereas the ROE reaches a minimum of -657% and a maximum of 105%. E.ON is the company that reports an ROE of -657% in 2016, which likely results from the transformation of their business into an increased sustainable corporate strategy (E.ON, 2017).

Besides, it is worth noting that the accounting-based measures are moderately, positively correlated with each other, but only weakly, positively correlated with Tobin's Q.

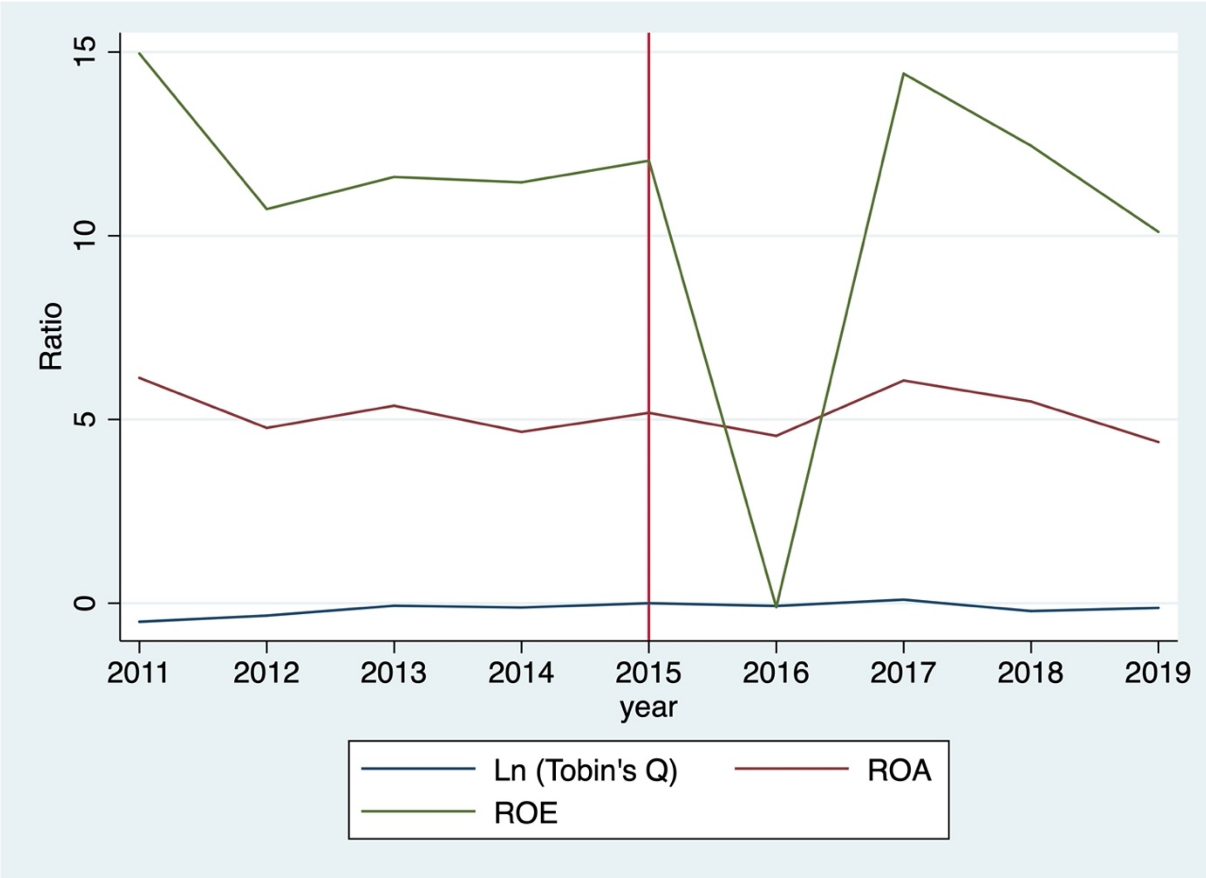
Tobin's Q represents the market-based measure and is calculated by dividing the sum of the market value of stock and the book value of debt by the number of total assets in that respective year. Taking total assets as the denominator makes it easy to compare the ratio across firms without adjusting for risk, leverage or size (Campbell & Mínguez-Vera, 2007; Yang et al., 2019). Subsequently, a ratio larger than 1 indicates that the firm adds value to its operations through efficient usage of the assets, whereas a ratio smaller than 1 signals inefficient usage of its assets. For the analysis, I use the natural logarithm since the variable is skewed to the right. The mean equals a value of 0.83 ($e^{-0.181}$), and Rational AG has the highest value for Tobin's Q of 11,66 in 2019.

Figures 2.2 and 2.3 provide an overview of the trend of the firm performance measures in Germany and the Netherlands, respectively. For Germany, the ROA and the natural logarithm of Tobin's Q remain relatively stable from 2011 until 2019. The ROA fluctuates at a value of 5 and the natural logarithm of Tobin's Q minimally fluctuates around a value of 0. The ROE remains above 10 in the first years and then experiences a large drop to a value of 0 in 2016, after which it recovers to a value of 10.

In terms of the natural logarithm of Tobin's Q, Dutch firms experience a similar, stagnant trend around 0. The ROA sees a slightly increasing trend to a value of roughly 6 in 2018. The ROE is lower than in Germany and illustrates two dips in 2013 and 2017 after which it reaches a value of approximately 13 in 2018. Eventually, it settles at a value just above 10, which may indicate a strong investment period.

Figure 2.2

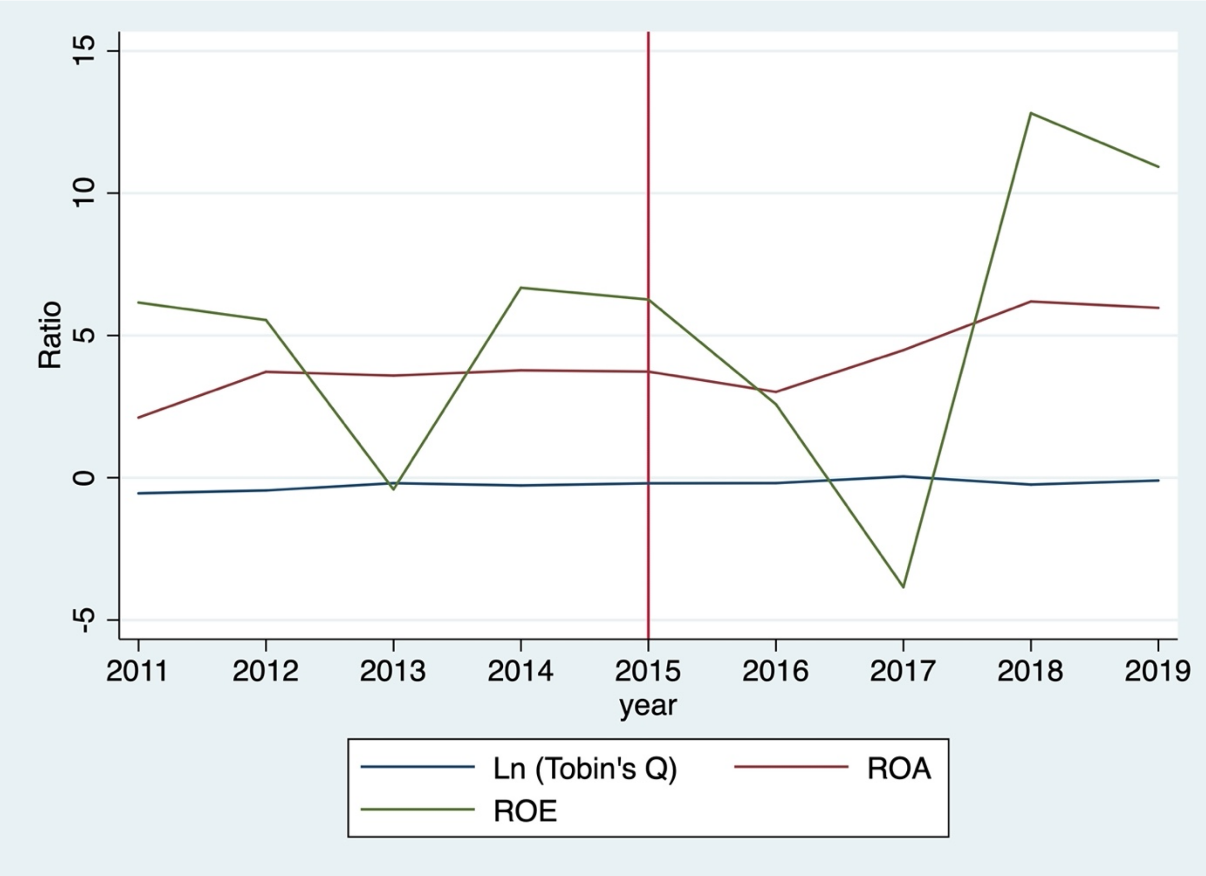
Descriptive Figure of Firm Performance Measures in Germany



Note: This figure illustrates the trend of the firm performance measures in Germany. The blue line indicates the trend for the natural logarithm of Tobin's Q, the red line for the ROA and the green line for the ROE. In 2015, Germany instituted a hard quota to increase the share of women on boards. This is illustrated with the red vertical line in the figure.

Figure 2.3

Descriptive Figure of Firm Performance Measures in the Netherlands



Note: This figure illustrates the trend of the firm performance measures in the Netherlands. The blue line indicates the trend for the natural logarithm of Tobin's Q, the red line for the ROA and the green line for the ROE. In 2015, Germany instituted a hard quota to increase the share of women on boards. This is illustrated with the red vertical line in the figure.

Gender Ratio

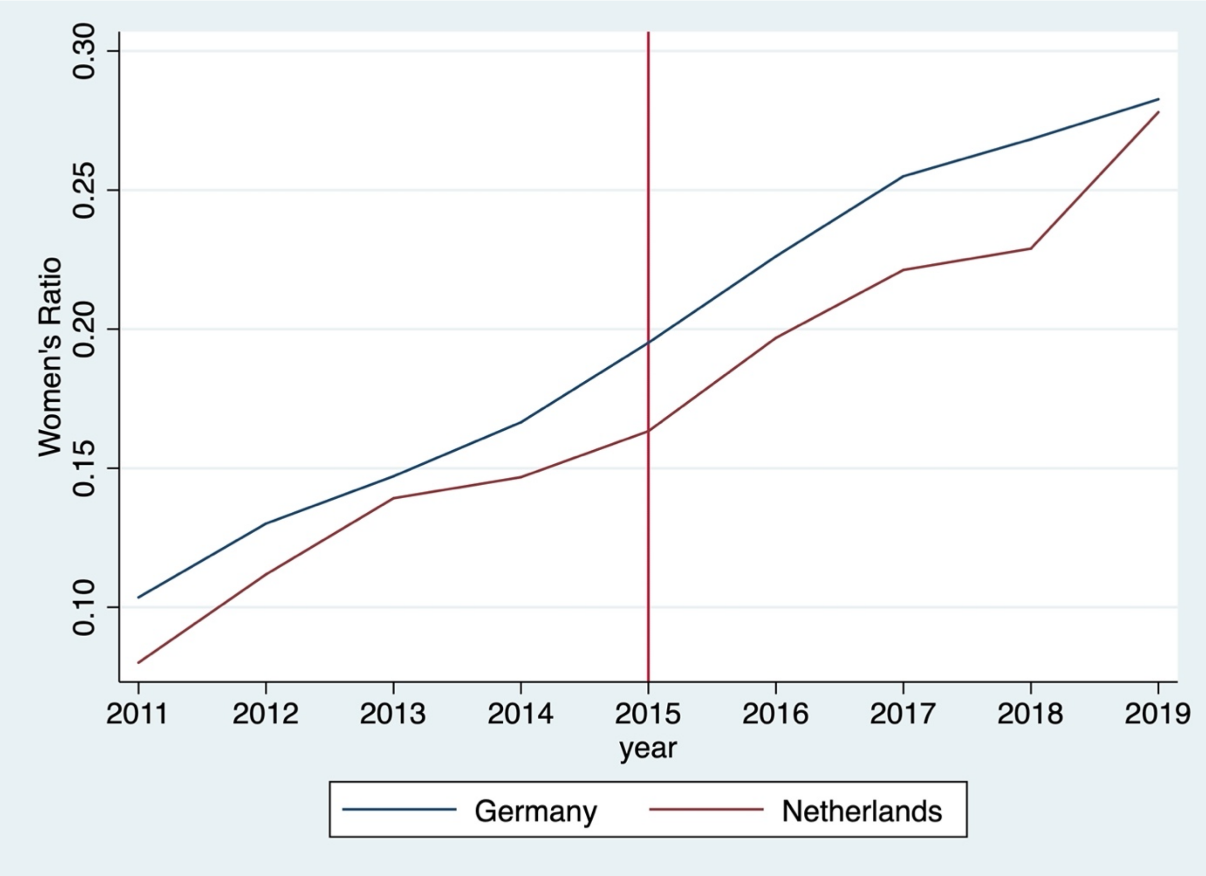
The gender ratio used for this analysis is a simple measure of the number of female executives on the board divided by the total number of executives on the board(s). Other scholars mostly use this as their base measure and extend their analyses with a dummy variable indicating whether at least one woman is present on the board (Adams & Ferreira, 2009), the absolute number of women on the board (Carter et al., 2010), or indexes, such as the Blau or Shannon index, that capture gender diversity (Campbell and Mínguez-Vera, 2007). Table 2.2 highlights that my gender ratio measure is weakly, negatively correlated with Tobin's Q and the ROA, but, weakly, positively correlated with the ROE.

The number that stands out is the maximum ratio of roughly 86%, which corresponds to the Koninklijke Boskalis Westminster N.V.. The board consisted of 6 members in 2014, 3 of which were women and 3 were men. After 2014, the board added one more female member and replaced two of these men with female executives, which constitutes a ratio of 6/7 or roughly 86%. Thereafter, the firm decreased the board by one female member, which reduces the measure to 71%. Frankly, such a gender ratio proves to be an exception rather than the norm as the mean indicates the share of women to be at roughly 19%.

Figure 2.4 illustrates the trend of the share of women on German and Dutch boards. This percentage is higher for German than for Dutch firms, but both experience a steady increase – both measures indicate that one out of ten board members are women in 2011, which develops to almost 3 out of 10 being female in 2019.

Figure 2.4

Trend on the Share of Women in Germany and the Netherlands



Note: This figure illustrates the trends of the share of women on the board for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. In 2015, Germany instituted a hard quota to increase the share of women on boards. This is illustrated with the red vertical line in the figure.

Control Variables and Firm Effects

According to Ahern and Dittmar (2012), Matsa and Miller (2013) and Yang et al. (2019), it is common to use characteristics of the boards as control variables. The board size, the education level, the average age and network sizes of the directors are prominent examples (Yang et al., 2019). I only include the board size because BoardEx does not provide yearly data on the other variables for this sample. The board size is measured by the total number of directors on the boards and averages 11 members per company. It ranges from having three to 31 directors on the board. For instance, 1 & 1 Drillisch AG and CTS Eventim AG & CO. KGAA are firms that have three directors, whereas RWE temporarily had 31 directors on the supervisory board in a year as several members joined and/or left mid-year and were still considered as a director of the board for that year.

In addition to the board size, I include firm fixed effects and industry fixed effects to account for time-invariant differences (Yang et al., 2019). The information I use to distinguish between the effect of each firm (the firm fixed effect) is based on the Return on Assets (ROA) based on Net Income, the Return on Equity (ROE) based on Net Income, the Total Assets, the Sales, the Leverage ratio, the Net Debt and the Firm Age. For the industry fixed effects, I retrieve the corresponding Standard Industrial Classification (SIC) codes from Orbis, which classify the primary operations of firms into 10,000 categories (SIC Codes, 2021). Since I only have roughly 846 observations, I remove the last two digits of the 4-digit SIC Code to get the 2-digit primary classifications. This amounts to 10 broad categories, which can be seen in Table 2.3 in the Appendix.

Table 2.2 shows a negative correlation between the board size and the firm performance, but a positive correlation with the share of women on the board. Firm characteristics are predominantly negatively associated with firm performance, but mostly positively correlated with each other.

Methodology

Based on prior research, the relationship between gender diversity and firm performance is an area of research that comes short of causal inferences (Yang et al., 2019). These results rely on correlations rather than causations, which are one-dimensional association measures and suffer from endogeneity since the variable of interest is correlated with the error term. Selection bias and Omitted Variable Bias (OVB) are the most common sources of endogeneity – the former refers to the difference in outcomes in the absence of treatment, and the latter to omitted variables that are correlated with the treatment variable (Samad et al., 2009). Essentially, both issues make the two groups different before treatment, hence, they are not suitable control and treatment groups.

If one randomizes the treatment between the two groups, though, the two groups become similar in their baseline characteristics. They would only differ in the treatment, which removes the difference in outcomes in the absence of treatment. Adams (2016) recognizes, however, that this is not feasible for this research because board positions are not handed out randomly but rather earned through reputation and experience. Even if randomization was possible, selection bias may be a concern again since the participating women may be vastly different from the general population. Consequently, the results obtained from such a setting would not be externally valid.

For my analysis, I apply the Difference-in-Difference method where the treatment is not an individual's choice but made at a higher level of aggregation. This typically refers to a policy intervention that applies to one group, but not to another, similar group. In Card's and Krueger's (1994) famous paper, the intervention was the increase of the minimum wage in New Jersey, whereas it remained the same in the neighboring state of Pennsylvania. Yang et al. (2019) base their approach on Matsa and Miller (2013) and apply the DiD method to the introduction of a statutory quota on Norwegian firms in 2003, and use firms from Finland, Sweden and Denmark as the counterfactual. In my analysis, I chose Germany and the Netherlands as treatment and control groups, respectively, to examine the difference in the effect between a hard and a soft quota. The German government passed a law in 2015 that regulated a 30% minimum of women in supervisory boards for listed companies from 2016 onwards. It is important to mention that this 30% minimum did not have to be integrated immediately as some terms of board members still lasted for a few years. The firm was obliged, if necessary, to replace a male executive with a female executive only after the term of the respective male executive ended. It is similar to the Intention-To-Treat (ITT) principle because some firms, who ought to comply, do not abide by the gender ratio immediately. However, these firms must be integrated into the analysis as it provides an unbiased assessment of the efficacy of the quota (Montori and Guyatt, 2001).

In contrast to that, the Netherlands passed a law in 2013 which only recommends a target value of 30% women in both the executive and the supervisory board of listed firms.

The DiD method is suitable for causal inferences because it accounts for (1) unobserved, time-invariant differences between the control and the treatment group and (2) observed, time-variant factors that occur similarly to both groups (Samad et al., 2009). This method requires that unobserved characteristics between the two groups do not vary over time. In other words, in the absence of treatment, the difference between the trends would be constant over time, which is referred to as the parallel trends assumption (Samad et al., 2009). The effect I estimate can be captured in the following simple DiD regression:

$$(1)\ Y_{jit} = \beta * \text{Treatment} + \rho * \text{TGroup} + \lambda_t * \text{Year}_t + \theta_{jit} * \text{NDirectors} + \alpha_j + y_i + \varepsilon_{ijt}$$

Where Y_{jit} denotes the value for the firm performance measure of firm j in sector i at time t, driven by the treatment effect $\beta * \text{Treatment}$. The treatment effect is constructed through the interaction between belonging to the treatment group, i.e. when TGroup takes on value 1, and when Year takes on a value greater than 2015. Besides controlling for the size of the board $\theta_{ijt} * \text{NDirectors}$, I hold year fixed effects Year_t, firm fixed effects α_j and industry fixed effects γ_i constant.

All regressions, which include the treatment variable, use standard errors clustered on the firm level (Matsa and Miller, 2013; Yang et al., 2019). This approach underlies the criticism of Bertrand et al. (2004) who find autocorrelation in the variables and the error term when comparing outcomes over time. Clustering standard errors on firm-level allows for arbitrarily correlated data within the defined cluster, such that one does not underestimate the standard errors. It only affects the significance of the coefficients, whilst the magnitude and the direction remain unaffected (Petersen, 2009).

Analysis

Before examining the results, I analyze whether the parallel trends assumption holds for every outcome variable. Figure 4.1, Figure 4.2 and Figure 4.3 illustrate the trend of each firm performance variable. I report the graphical illustration for Tobin's Q in this section and enclose the corresponding graph of the natural logarithm of Tobin's Q in the Appendix. Clearly, the trends of the accounting-based measures do not run parallel, which is especially the case before the law passed in 2015. The strong drops of the ROE in 2013 and 2017 and the stark fluctuations before 2015 for the ROA are surprising, which motivates me to remove companies with large negative values. Companies with large negative values seem to be outliers as this may likely be an unique investment period or a random measurement error, hence, I remove E.ON and Pharming

Group N.V.. After excluding both, I get Figure 4.4 for the ROE. The trends become slightly more parallel, however, in the years of 2013 and 2014 the trends remain unparallel. A priori, I cannot identify any reason as to why the ROE is different between the two countries. Economic factors that affect the two countries differently must play a role. Interestingly, if I remove these two companies for the ROA, I get Figure 4.5. The trends seem to run parallel in the preintervention period as well, which would satisfy the parallel trends assumption. Besides, I removed these two companies to see whether the graphs for Tobin's Q and for the natural logarithm change, but they do not seem to differ. Figure 4.6 (Appendix) also confirms that the natural logarithm of Tobin's Q depicts parallel trends.

To investigate the parallel trends assumption further, I implement the leads of the treatment variable into the model. The idea is that if one shifts the intervention one period back, the parallel trends assumption would assume no differences between German and Dutch firms, except for the initial differences and the year fixed effects. In that case, the coefficient of the treatment (called lead) would be insignificant. It can be illustrated in the following way:

$$(2)\ Y_{jit} = \sum_{j=0}^{q} \beta_j * \text{Treatment}_{t+j} + \rho * \text{TGroup} + \lambda_t * \text{Year}_t + \theta_{ijt} * \text{NDirectors} + \alpha_j + \gamma_i + \varepsilon_{ijt}$$

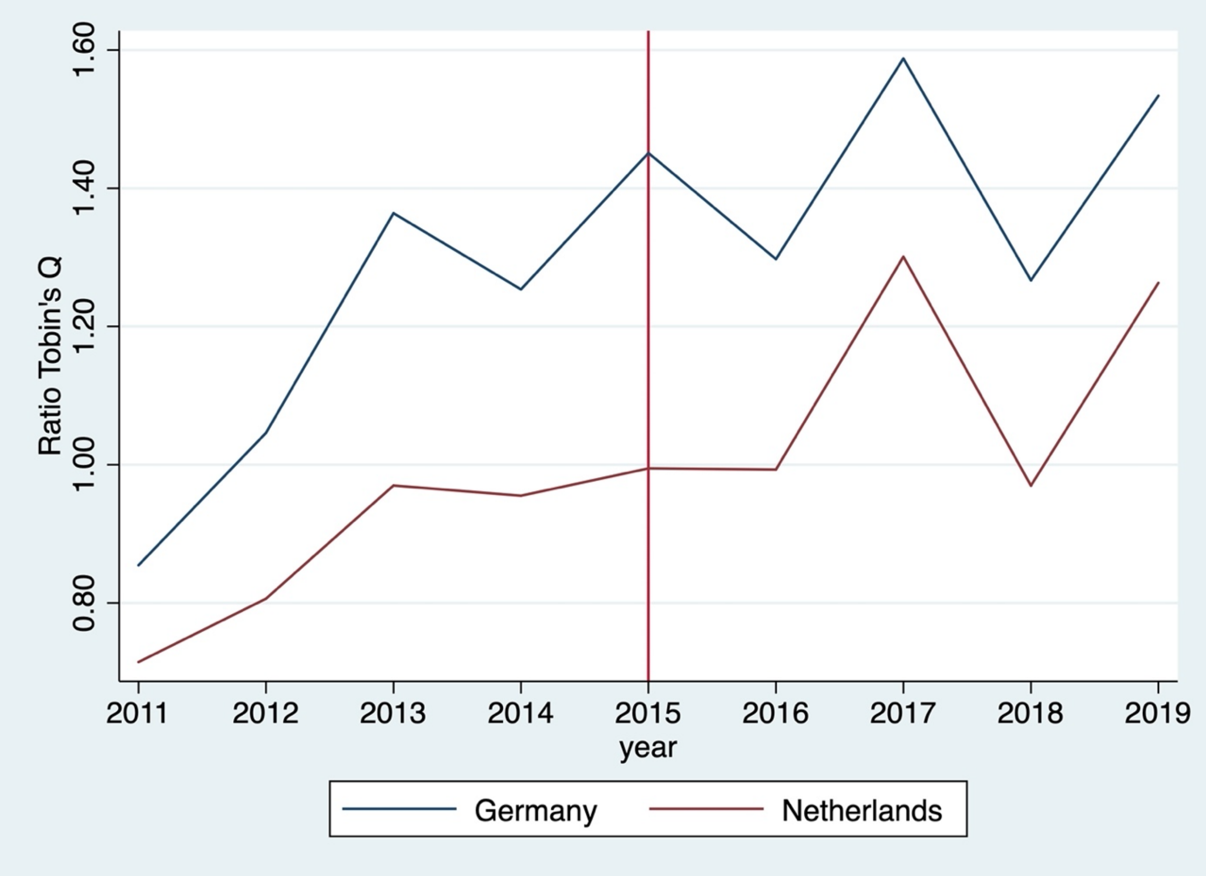
The corresponding regressions are included in the Appendix. For the accounting-based measures, I use the sample without the E.ON and Pharming Group N.V., but for Tobin's Q I use the entire sample. Table 4.1 illustrates the regression for Tobin's Q, Table 4.2 for the ROE and Table 4.3 for the ROA. None of the coefficients are significant, which means that the groups have no major differences on top of the year effects and initial differences. Thus, I can include all measures into the regression because the parallel trends assumption seems to be satisfied for all of them.

Additionally, I investigate whether the share of women on boards increased immediately after the implementation of the soft quota for Dutch firms and the hard quota for German firms (Yang et al., 2019). It could be the case that the change happened in later years, which means that the time of adaptation for firms would have induced a delayed effect. Figure 2.4 confirms that the female ratio increases in both countries after 2013, and more prominently after 2015. It is worth noting that from

2013 until 2015, the rate at which the female ratio increases is lower for Dutch firms. German firms have a relatively stable increase in the share of women over the entire period.

Figure 4.1

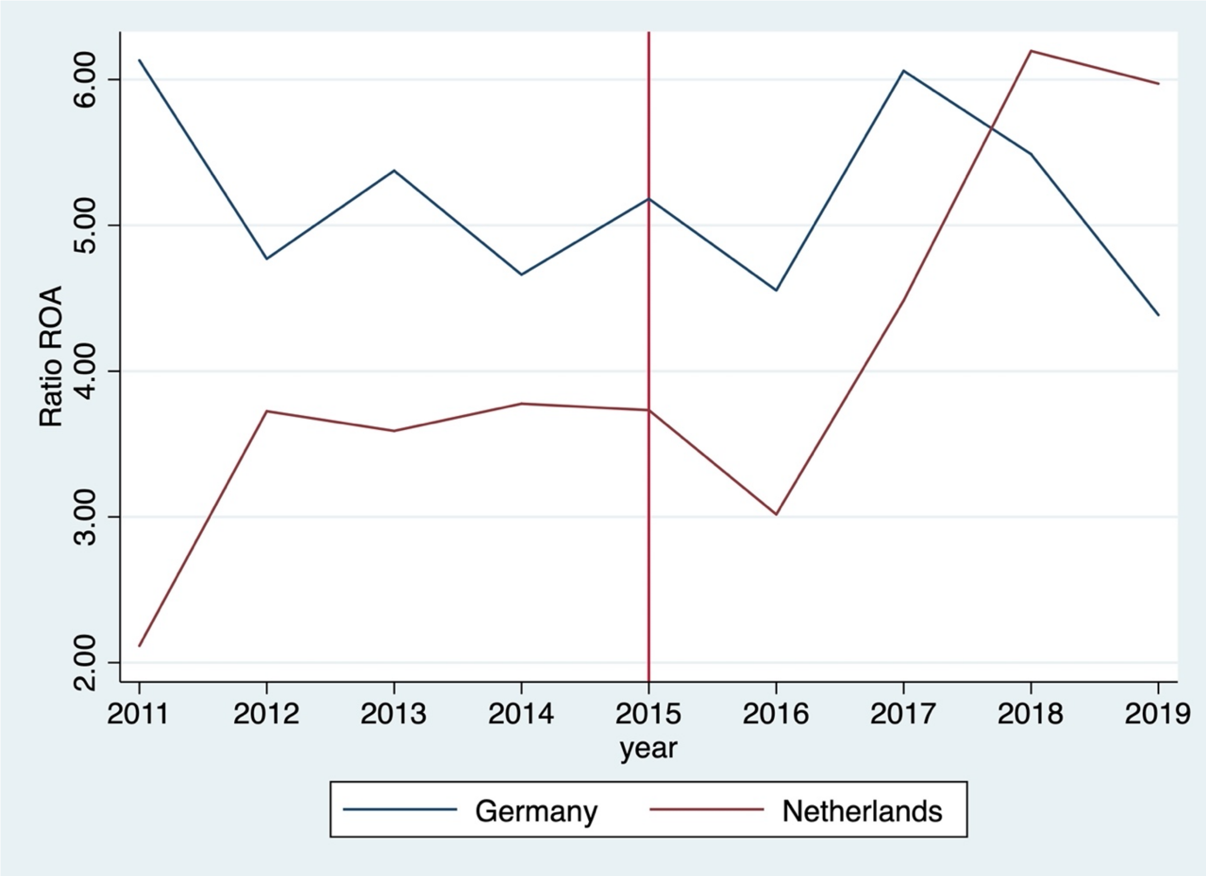
Graphical Illustration of the Parallel Trends Assumption for Tobin's Q



Note: This figure illustrates the trends of Tobin's Q for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The y-axis indicates the ratio for Tobin's Q and the x-axis the corresponding years. In total, 846 observations were used for this figure. The vertical red line illustrates the year of the intervention, that is to say when Germany instituted the hard quota.

Figure 4.2

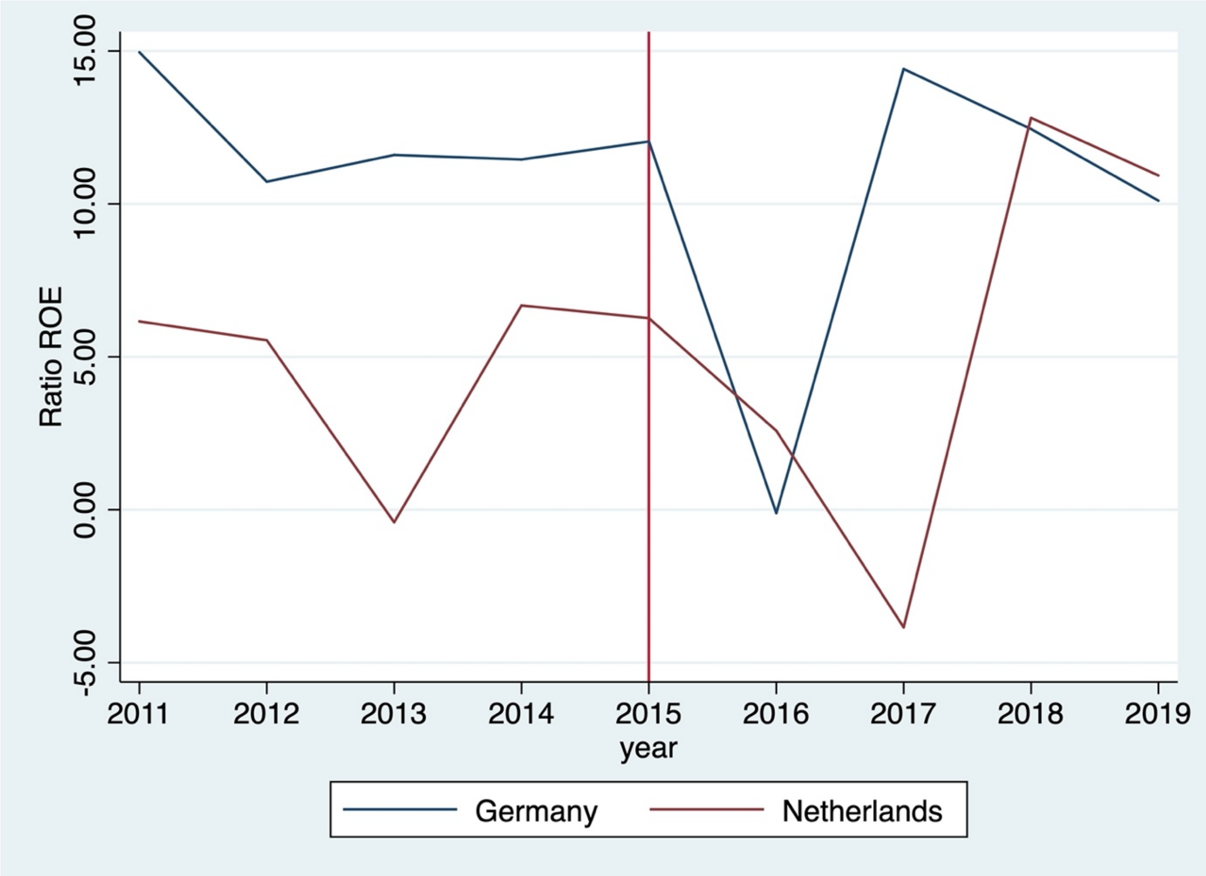
Graphical Illustration of the Parallel Trends Assumption for the ROA



Note: This figure illustrates the trends of the ROA for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The y-axis indicates the ratio for the ROA and the x-axis the corresponding years. In total, 846 observations were used for this figure. The vertical red line illustrates the year of the intervention, that is to say when Germany instituted the hard quota.

Figure 4.3

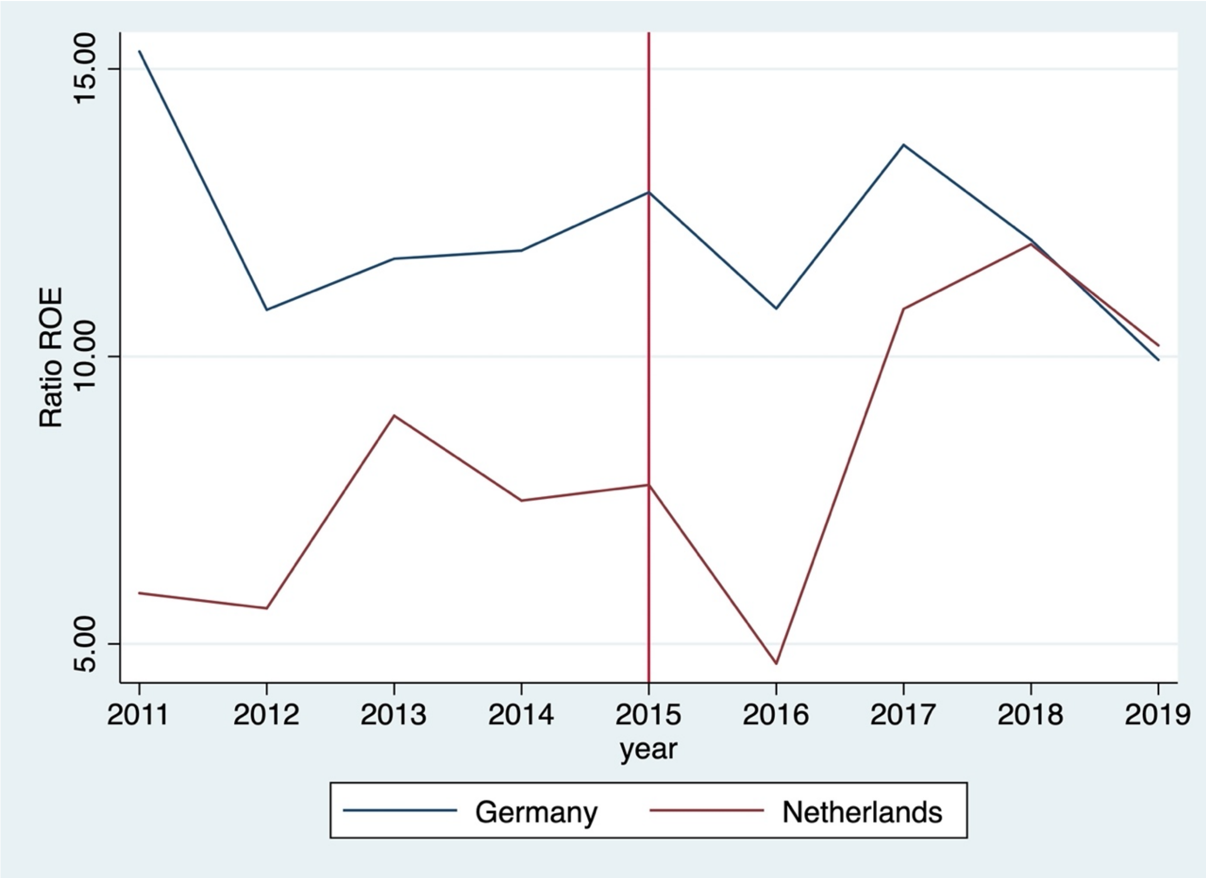
Graphical Illustration of the Parallel Trends Assumption for the ROE



Note: This figure illustrates the trends of the ROE for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The y-axis indicates the ratio for the ROE and the x-axis the corresponding years. In total, 846 observations were used for this figure. The vertical red line illustrates the year of the intervention, that is to say when Germany instituted the hard quota.

Figure 4.4

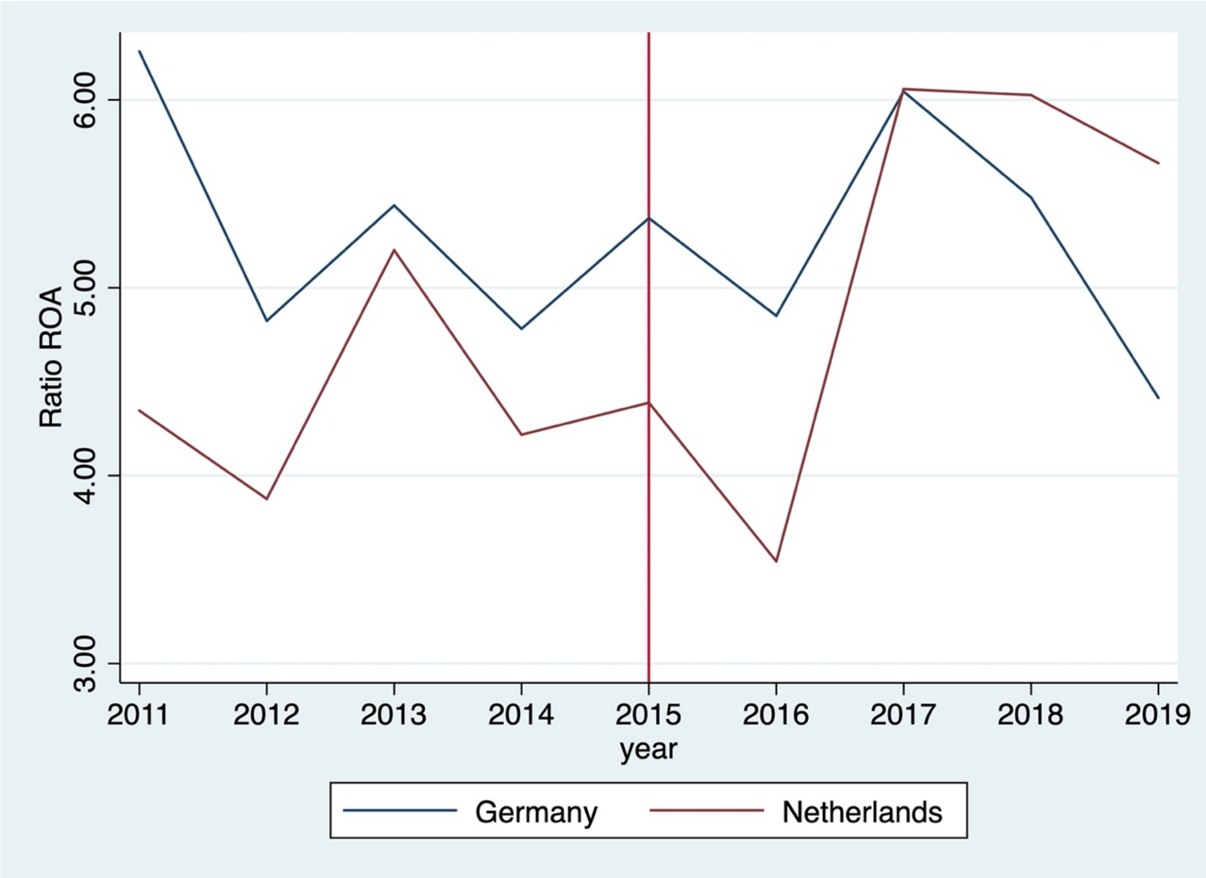
Graphical Illustration of the Parallel Trends Assumption for the ROE using an adjusted sample



Note: This figure illustrates the trends of the ROE for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The sample used for this illustration excludes the firms E.ON and Pharming Group N.V. to see whether the trends become similar at preintervention. The y-axis indicates the ratio for the ROE and the x-axis the corresponding years. In total, 828 observations were used for this figure. The vertical red line illustrates the year of the intervention, that is to say when Germany instituted the hard quota.

Figure 4.5

Graphical Illustration of the Parallel Trends Assumption for the ROA using an adjusted sample



Note: This figure illustrates the trends of the ROA for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The sample used for this illustration excludes the firms E.ON and Pharming Group N.V. to see whether the trends become similar at preintervention. The y-axis indicates the mean ratio for the ROE and the x-axis the corresponding years. In total, 828 observations were used for this figure. The vertical red line illustrates the year of the intervention, that is to say when Germany instituted the hard quota.

Results

In this section I present the results of my DiD analysis in three subsections: First, I illustrate the effects of the hard and soft quota on the composition of the boards. Second, I report the effect of both quotas on firm performance. Lastly, I perform a robustness analysis.

Board Composition

Following the approach of Yang et al. (2019), I investigate whether a hard quota significantly increases the share of women on boards, as opposed to a soft quota. Table 5.1 reports the results of the regression of the board characteristics on the treatment variable using the years after 2015 as post-treatment periods (Matsa & Miller, 2013; Yang et al., 2019). I find that the hard quota, as opposed to the soft quota, increases the share of women on the board significantly after 2015. It shows that the hard quota is effective, partly due to the sanctions and the normative pressure from society. Therefore, I accept my first hypothesis.

Furthermore, I find that the hard quota significantly increases the board size, which is in line with the findings of Yang et al. (2019). It implies that, on average, hard quotas make German firms increase the sizes of their supervisory boards to adjust to the minimum percentage. Combining both findings, the increase in the share of women does not result from the decrease of the board size, but rather from the increase of the number of female executives on the board. The correlation matrix also indicates a weak, but positive association between these two variables.

However, the positive effect of the hard quota on the board size is illogical. The gender quota does not regulate the size of the boards, but firms rather abide by other regulations that determine the number of directors on the boards based on the status of a company and the employee size. For instance, a German company with a maximum of 10,000 employees is required to have exactly 12 supervisory board members regardless of the gender quota (Bundesamt für Justiz, 2021). Increasing the board size just to meet the criteria is thus not possible. I can only think of an exception during the adaptation phase that temporarily allows the firm to do so, though, I cannot locate such regulation.

Table 5.1

Effect of the Quotas on the Board Composition

Variables	Share of Women	Board Size
Treatment (Post 2015)	0.110***	0.237***
	(0.000)	(0.000)
Treatment Group	-0.0618***	-1.772***
	(0.000)	(0.000)
Firm- & industry-fixed effects	Yes	Yes
Number of Firms	94	94
Number of Observations	846	846
R^2	0.6064	0.9468

Note: This table illustrates the regression of the board characteristics on the treatment variable using the years from 2016 until 2019 as post treatment period. The standard errors are clustered on firm level and are reported in parentheses below the coefficient. Firm and industry fixed effects are included in all regressions. The p-values are denoted in the following way: *** p < 0.01, ** p < 0.05, * p < 0.1

Firm Performance

In this section, I follow the approach of Matsa and Miller (2013) and Yang et al. (2019). Table 5.2 presents the results of the regression of the market-based and accounting-based measures on the treatment variables. Firm and industry fixed effects and the board size are included. For the accounting-based measures I use the subsample, whereas for Tobin's Q I use the entire sample. Regardless of the dependent variable, the DiD estimator is significantly negative ($\beta = -0.084, p = 0.004;\ \beta = -1.038, p = 0.007;\ \beta = -3.023, p = 0.000$). The results for the accounting-based measure are similar to Yang et al. (2019) but contrast their findings for the market-based measure since they find an insignificant, negative estimator. Thus, a hard quota reduces performance more than a soft quota which leads me to accept both the second and third hypotheses.

In Table 5.5 (Appendix), I report all three regressions to examine whether the exclusion of the company and industry fixed effects change the coefficients significantly. The treatment variables and their significance remain similar – only the size becomes slightly more negative when using the accounting-based measures and slightly less negative for Tobin's Q. Additionally, the control variable turns out to have a significant, negative effect on the performance measures. Thus, excluding the fixed effects does not lead to a significantly different outcome.

Table 5.2

Effect of the Quotas on Tobin's Q, the ROA and the ROE

Variables	Ln (Tobin's Q)	ROA	ROE
Treatment	-0.084***	-1.038***	-3.023***
	(0.001)	(0.011)	(0.001)
Treatment Group	1.379***	13.120***	22.628***
	(0.009)	(0.143)	(0.010)
Board Size	0.003	-0.120	-0.945***
	(0.005)	(0.083)	(0.004)
Firm- & industry-fixed effects	Yes	Yes	Yes
Number of Firms	94	92	92
Number of Observations	846	828	828
R^2	0.8616	0.4451	0.4075

Note: This table illustrates the regression of the board characteristics on the treatment variable using the years after 2015 as post treatment period. I use clustered standard errors on firm level, which are reported in parentheses beneath the coefficients. Firm and industry fixed effects are included in all regressions. The number of observations between the first and the second and third regression differ because a subsample with two firms less, which equals 18 observations, is used for the accounting-based measures. The p-values are denoted in the following way: *** p < 0.01, ** p < 0.05, * p < 0.1

Robustness Tests

Since the law for the soft quota passed for Dutch firms in 2013, it could be that before 2013 the two groups are already different. In that case, if I take Dutch firms as the treatment group and German firms as control group and create a new treatment (interaction) effect of belonging to the treatment group (a Dutch firm) and the year being larger than 2013, then the treatment effect needs to be insignificant for the two groups to be similar. This approach is related to the implementation of the leads – the difference is that I assign the German firms to the control and Dutch firms to the treatment, and that my treatment effect is the effect of a soft quota on firm performance as opposed to a hard quota. Table 5.3 summarizes the results, which indicate significant DiD estimators. The effect is positive for the accounting-based measures, but negative for Tobin's Q – it seems that the model I use before captures an additional effect of the soft quota in 2013. Thus, I restrict my sample from 2014 until 2019 and perform the same regression as in the subsection before. I include 2014 because otherwise I do not allow for anticipation before the intervention in 2015. The DiD method requires both a preintervention and postintervention period to estimate the difference in the treatment effect (Samad et al., 2009).

Table 5.4 reports the regression with the restricted sample and shows that the overall results remain unchanged. The magnitudes of the DiD estimators are less negative for the accounting-based measures, but still significant. The treatment coefficient when using Tobin's Q is more negative, but only significant at a 5% level. Thus, even with this subsample do I find a performance-reducing effect of the hard quota as opposed to the soft quota.

Table 5.3

Effect of the Soft Quota on Tobin's Q, the ROA and the ROE

Variables	Ln (Tobin's Q)	ROA	ROE
Treatment	-0.086***	0.252*	1.127**
(Post 2013)	(0.001)	(0.037)	(0.026)
Treatment Group	1.494***	14.784**	27.104**
	(0.007)	(0.571)	(0.407)
Board Size	0.015	0.068	-0.651
	(0.004)	(0.366)	(0.261)
Firm- & industry-fixed effects	Yes	Yes	Yes
Number of Firms	94	92	92
Number of Observations	470	460	460
R^2	0.9070	0.5421	0.5598

Note: This table illustrates the regression of the board characteristics on the treatment variable. The period I consider ranges from 2011 until 2015. 2011 until 2013 is considered as the preintervention period, whereas 2014 until 2015 is seen as the postintervention period. I use clustered standard errors on firm level, which are reported in parentheses beneath the coefficients. Firm and industry fixed effects are included in all regressions. The number of observations between the first and the second and third regression differ because a subsample with two firms less, which equals 10 observations less, is used for the accounting-based measures. The p-values are denoted in the following way: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5.4

Effect of the Quotas on Tobin's Q, the ROA and the ROE based on the sample of 2014 until 2019

Variables	Ln (Tobin's Q)	ROA	ROE
Treatment (Post 2015)	-0.136**	-0.858***	-2.307***
	(0.003)	(0.009)	(0.003)
Treatment Group	1.448**	9.618***	14.467***
	(0.022)	(0.092)	(0.026)
Board Size	0.009	-0.214	-1.060***
	(0.011)	(0.047)	(0.008)
Firm- & industry-fixed effects	Yes	Yes	Yes
Number of Firms	94	92	92
Number of Observations	564	552	552
R^2	0.8922	0.5660	0.4841

Note: This table illustrates the regression of the board characteristics on the treatment variable using the restricted sample from 2014 until 2019. 2014 and 2015 are regarded as the preintervention period, and 2016 until 2019 as the postintervention period. I use clustered standard errors on firm level, which are reported in parentheses beneath the coefficients. Firm and industry fixed effects are included in all regressions. The number of observations between the first and the second and third regression differ because a subsample with two firms less, which equals 12 observations less, is used for the accounting-based measures. The p-values are denoted in the following way: *** p < 0.01, ** p < 0.05, * p < 0.1

Discussion

Key Findings

In the results section, I illustrate that the hard quota significantly increases the share of women on the boards, compared to the soft quota. A target value only recommends the integration of more women, whereas a legal quota makes firms increase their share of women. It does not necessarily mean that Dutch firms did not see an increase in the share of women (see Figure 2.4), but rather that the increase for German firms is significantly stronger compared to the Dutch firms after 2015. Since the hard quota also seems to enlarge the board size, the increase in the share of women results from the increase of the actual number of women on the board rather than from the decrease of the board size.

Regarding *firm performance*, I find that a hard quota is more performance-reducing than a soft quota. Specifically, the treatment effect is more negative for the accounting-based measures than for Tobin's Q. This finding is in line with the results of Matsa and Miller (2013) and Yang et al. (2019), though, one must be cautious with the corresponding interpretation. On the one hand, I expect this result because simply replacing male with female executives may be an impediment rather than a contribution in the short-term. Social cohesion may emerge from changing group dynamics, but this does not necessarily have to be related to gender. I argue that arbitrarily replacing key executives can disturb the composition of the board as groups need time to adjust to radical change. However, such a change cannot drive the negative effect on firm performance alone. Rather, it shows that increasing the share of women does not result in improved firm performance, but that boards with relatively more women score lower on the performance measures than boards with fewer women (Yang et al., 2019). Other factors must play an indirect or direct role such that the bad performance of the board continuously translates into a lower score in the performance measures. The necessary assumptions hold in my setting, but the interpretation of the results is not as straightforward as one would think.

Therefore, scholars and policymakers should take away that instituting a hard quota to simply increase the share of women can be troublesome for a firm. It is not clear how the causal effects of the hard quota should be interpreted, but it is likely that the negative effect cannot be solely attributed to the quota (Yang et al., 2019).

Implications for Practice and Theory

My results endorse the qualitative research by Arndt and Wrohlich (2019) that statutory quotas are more effective than soft quotas to integrate more women into the upper management. For a quota to work, compliance needs to be obligatory, and noncompliance must be sanctioned accordingly (Yang et al. 2019).

The performance-reducing effect I find does not support the Resource Dependence Theory or the Human Capital Theory – both theories mainly suggest that every outsider can contribute with his/her unique skills, but that is not always true. Some skills may not be relevant for a particular firm or at a particular point in time. Simply swapping a male with a female executive can cause disruptions in the entire firm, which translates into a lower score of the performance measures as well. As suggested by the Social Psychology Theory, such disruptions emerge from time-consuming and ineffective processes. Thus, despite being a more effective tool for policymaking, a hard quota can harm the daily operations of the firm in some cases.

Difference-in-Difference Method

The analysis of the parallel trends assumption and the lead implementation clarifies that the results can be correctly interpreted. In essence, a significant DiD estimate would mean that the difference between the treatment and control group can be attributed fully to the treatment effect. In other words, before the intervention both groups ought to have similar (parallel) trends, and after the intervention the groups experience a different effect only because of the treatment. As mentioned before, it is naïve to assume that this is the case. Other unobserved variables make the interpretation of the results unclear such that the negative effect can only be attributed partly to the quota. However, unobserved effects captured by the Dutch soft quota in 2013 are not a large issue. The robustness test shows that the effect remains similar and significantly negative when using the restricted sample from 2014 until 2019, hence, the performance-reducing effect is still justifiable. Therefore, other unobserved effects not captured by the soft quota would be a potential issue.

Therefore, I recommend extending this analysis by either (1) using multiple countries as a control group or (2) using a different setting. For (1), the issue with the

DiD method is that no single unit may be a suitable control group. Thus, it can be more efficient to create a 'synthetic' control group, i.e., including several countries and taking the weighted average of the predictors (estimators) of the outcome variables for the different units in the control group (Abadie et al., 2010). Consequently, the synthetic control group resembles the best possible control group for the treatment group. The difficulty in this approach is gathering the amount of data of comparable countries, which leads me to my second suggestion: Ideally, one would take both Germany and the Netherlands as treatment groups and evaluate the hard and the soft quota versus a country that has no quota at all. Currently, nine European countries still do not have a statutory quota, hence, one of them can serve as the counterfactual to estimate the effect of the respective quota individually (Arndt and Wrohlich, 2019). Frankly, I tried gathering data on such a country but financial and/or gender information is scarce for a comparable country.

Further Limitations

My analysis is limited in two additional points: First, this paper examines a period from 2011 until 2019 for the effect of a quota on firm performance. It makes the interpretation possible for the short-term only – whether these effects also hold in the long-term is unclear. Yang et al. (2019) suggest that long-term effects may vanish because the new norm may motivate firms to constantly hire women with expertise and networks similar to those of men. Thus, female directors may not be distinguishable from male directors anymore, which means that the effect on firm performance can become negligible. Therefore, I encourage further research to examine whether long-term effects are different from short-term effects.

Secondly, I question the generalizability of these results. Both countries instituted a hard and soft quota of 30%, but other European countries have different values for the quota, different sanctions, and a different set of firms that the rules apply to (Arndt and Wrohlich, 2019). In addition, German listed firms had one year to adjust to the quota, whereas other countries give firms more time. Norway, for instance, gave firms five years, France six years and Spain eight years of adaptation (Arndt and Wrohlich, 2019). Thus, my results are rather specific, so I encourage scholars to develop a setting that can be used across several countries.

Conclusion

While a substantial amount of prior research focused on examining the correlation between gender diversity on the boards and firm performance, I contribute to the literature by taking a causal approach. Specifically, I use a Difference-in-Difference method that examines the difference in the effects between a hard and soft quota of Germany and the Netherlands, respectively. The hard quota was instituted by Germany in 2015, whereas the soft quota came into effect in the Netherlands in 2013. The treatment group consists of German listed firms and the control group of Dutch listed firms. I find that the hard quota significantly increases the number of women in the upper management, but that it induces a performance-reducing effect. Social cohesion within the board can be a plausible driver. However, the interpretation remains ambiguous since I believe that other unobserved factors constitute this effect.

In addition to the suggestions made before, scholars can extend this research by integrating more board characteristics, such as average tenure, network capacity of executives, average age, etc. One can then conduct two regressions, namely one that uses only fixed effects and one that includes the board characteristics as well to further validate the robustness of the results (Yang et al., 2019).

Appendix

Table 2.3

Primary Classifications of the SIC Codes

Range of SIC Codes	Division
0100 – 0999	Agriculture, Forestry and Fishing
1000 – 1499	Mining
1500 – 1799	Construction
1800 – 1999	Not Defined
2000 – 3999	Manufacturing
4000 – 4999	Transportation, Communications, Electric, Gas and Sanitary Service
5000 – 5199	Wholesale Trade
5200 – 5999	Retail Trade
6000 – 6799	Finance, Insurance and Real Estate
7000 – 8999	Services
9100 – 9999	Public Administration

Note: This table provides an overview of the primary classifications of the SIC codes. The left column indicates the range of the 4-digit SIC Codes, and the right column the corresponding primary business activity of a company. The SIC codes range from 0100 to 9999.

Table 4.1

Regression including the leads of the treatment variable for Tobin's Q

Ln (Tobin's Q)	First Lead	Second Lead	Third Lead
Lead (f.Treatment)	0.086	0.081	0.079
	(0.067)	(0.054)	(0.057)
Year, Firm and industry fixed effects	Yes	Yes	Yes
Observations	752	658	564
R^2	0.8777	0.8893	0.8967

Note: This table summarizes the regression of firm performance using the natural logarithm of Tobin's Q as dependent variable on the leads of the treatment variable. Year, firm and industry fixed effects are implemented in the regression as well. These regressions use robust standard errors and are reported in the parentheses below the coefficient. The number of observations decline from the first column to the last column because the treatment variable is progressively shifted back from 2016 so that observations from 2019 cannot be used for the first, observations from 2019 and 2018 not for the second, and observations from 2019, 2018 and 2017 not for the third column. The p-values are denoted in the following way: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4.2

Regression including the leads of the treatment variable for the ROE

ROE	First Lead	Second Lead	Third Lead
Lead (f.Treatment)	-0.593	-1.205	-3.055
	(2.214)	(2.275)	(2.643)
Year, Firm and industry fixed effects	Yes	Yes	Yes
Observations	736	644	552
R^2	0.4641	0.5000	0.5162

Note: This table summarizes the regression of firm performance using the ROE as dependent variable on the leads of the treatment variable. Year, firm and industry fixed effects are implemented in the regression as well. These regressions use robust standard errors and are reported in the parentheses below the coefficient. The number of observations decline from the first column to the last column because the treatment variable is progressively shifted back from 2016 so that observations from 2019 cannot be used for the first, observations from 2019 and 2018 not for the second, and observations from 2019, 2018 and 2017 not for the third column. It is important to mention that the number of observations also differ from Table 4.1 because two firms were deleted for this regression. The p-values are denoted in the following way: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4.3

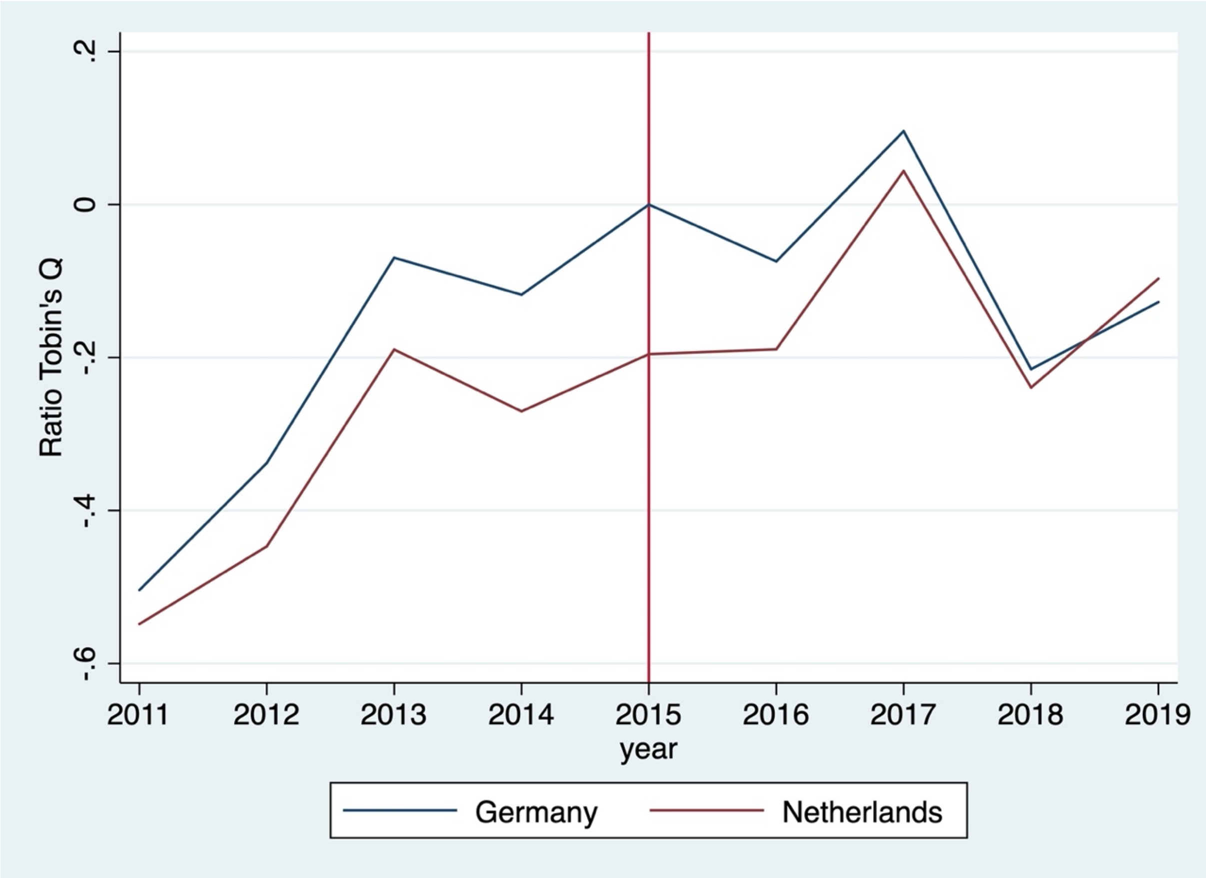
Regression including the leads of the treatment variable for the ROA

ROA	First Lead	Second Lead	Third Lead
Lead (f.Treatment)	0.039	-0.273	-0.822
	(1.115)	(1.168)	(1.297)
Year, Firm and industry fixed effects	Yes	Yes	Yes
Observations	736	644	552
R^2	0.4900	0.5210	0.5216

Note: This table summarizes the regression of firm performance using the ROA as dependent variable on the leads of the treatment variable. Year, firm and industry fixed effects are implemented in the regression as well. These regressions use robust standard errors and are reported in the parentheses below the coefficient. The number of observations decline from the first column to the last column because the treatment variable is progressively shifted back from 2016 so that observations from 2019 cannot be used for the first, observations from 2019 and 2018 not for the second, and observations from 2019, 2018 and 2017 not for the third column. It is important to mention that the number of observations also differ from Table 4.1 because two firms were deleted for this regression. The p-values are denoted in the following way: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 4.6

Graphical Illustration of the Parallel Trends Assumption for the natural logarithm of Tobin's Q



Note: This figure illustrates the trends of natural logarithm of Tobin's Q for both German and Dutch firms from 2011 until 2019. The blue line represents the German firms, whereas the red line represents the Dutch firms. The y-axis indicates the mean ratio for the natural logarithm of Tobin's Q and the x-axis the corresponding years. In total, 846 observations were used for this figure. The vertical red line illustrates the year when Germany instituted the hard quota.

Table 5.5

Effect of the Quotas on Tobin's Q, the ROA and the ROE without company and industry fixed effects

Variables	Ln (Tobin's Q)	ROA	ROE
Treatment	-0.077***	-1.025***	-3.091***
	(0.001)	(0.001)	(0.004)
Treatment Group	0.442*	1.841***	7.041**
	(0.040)	(0.023)	(0.127)
Board Size	-0.075*	-0.221**	-0.408**
	(0.009)	(0.005)	(0.031)
Firm- & industry-fixed effects	No	No	No
Number of Firms	94	92	92
Number of Observations	846	828	828
R^2	0.2459	0.0338	0.0443

Note: This table illustrates the regression of the board characteristics on the treatment variable using the years after 2015 as post treatment period. I use clustered standard errors on firm level, which are reported in parentheses beneath the coefficients. Firm and industry fixed effects are excluded in all regressions. The number of observations between the first and the second and third regression differ because a subsample with two firms less, which equals 18 observations, is used for the accounting-based measures. The p-values are denoted in the following way: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505.

AD. (2019). SER: Vrouwenquotum voor beursgenoteerde bedrijven. https://www.ad.nl/werk/ser-vrouwenquotum-voor-beursgenoteerde-bedrijven~af951b993/?referrer=https://www.google.com/

Adams, R. B. (2016). Women on boards: The superheroes of tomorrow? *The Leadership Quarterly*, *27*(3), 371–386.

Adams, R. B., & Ferreira, D. (2009). Women in the Boardroom and their Impact on Governance and Performance. *Journal of Financial Economics*, *94*(2), 291–309.

Ahern, K. R., & Dittmar, A. K. (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics*, *127*(1), 137–197.

Arndt, P., & Wrohlich, K. (2019). Gender Quotas In A European Comparison: Tough Sanctions Most Effective. *DIW Weekly Report*, *9*(38), 337–344.

Becker, G. (1964). *Human Capital*. Chicago, IL: University of Chicago Press.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Bundesamt für Justiz. (2021). *§ 7 MitbestG - Einzelnorm*. Gesetze im Internet. https://www.gesetze-im-internet.de/mitbestg/__7.html

Campbell, K., & Mínguez-Vera, A. (2007). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics*, *83*(3), 435–451.

Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, *84*(4), 772–793.

Carter, D. A., D'Souza, F., Simkins, B. J., & Simpson, W. G. (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.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate Governance, Board Diversity, and Firm Value. *The Financial Review*, *38*(1), 33–53.

Deutsche Börse AG. (2018). *Facts from 30 years of DAX*. Börse Frankfurt. https://www.boerse-frankfurt.de/en/know-how/about/geschichte-der-frankfurter-wertpapierboerse/facts-from-30-years-of-dax

Deutscher Bundestag. (2015). *Deutscher Bundestag - Frauenquote für die Top-Positionen beschlossen*. https://www.bundestag.de/dokumente/textarchiv/2015/kw10_de_frauenquote-363058

Du Rietz, A., & Henrekson, M. (2000). Testing the Female Underperformance Hypothesis. *Small Business Economics*, *14*(1), 1–10.

E.ON. (2017). *Annual Report 2016*. https://www.eon.com/content/dam/eon/eon-com/investors/annual-report/EON_Annual_Report_2016.pdf

European Institute for Gender Equality. (2019). *Browse Gender Statistics | Gender Statistics Database*. https://eige.europa.eu/gender-statistics/dgs/data-talks/legislative-quotas-can-be-strong-drivers-gender-balance-boardrooms

Haslam, S. A., Ryan, M. K., Kulich, C., Trojanowski, G., & Atkins, C. (2010). Investing with Prejudice: the Relationship Between Women's Presence on Company Boards and Objective and Subjective Measures of Company Performance. *British Journal of Management*. *21*(2), 484-497.

Hillman, A. J., Cannella, A. A., & Paetzold, R. L. (2000). The Resource Dependence Role of Corporate Directors: Strategic Adaptation of Board Composition in Response to Environmental Change. *Journal of Management Studies*, *37*(2), 235–256.

Investopedia. (2021). *How to Use Return on Assets When Analyzing a Company*. https://www.investopedia.com/terms/r/returnonassets.asp#:%7E:text=The%20Significance%20of%20Return%20on%20Assets%20(ROA),-ROA%2C%20in%20basic&text=The%20ROA%20figure%20gives%20investors,more%20money%20on%20less%20investment.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360.

Liu, Y., Wei, Z., & Xie, F. (2014). Do women directors improve firm performance in China? *Journal of Corporate Finance*, *28*, 169–184.

Marinova, J., Plantenga, J., & Remery, C. (2016). Gender Diversity and Firm Performance: Evidence from Dutch and Danish Boardrooms. *International Journal of Human Resource Management*, *27*(15), 1777–1790.

Matsa, D. A., & Miller, A. R. (2013). A Female Style in Corporate Leadership? Evidence from Quotas. *American Economic Journal: Applied Economics*, *5*(3), 136–169.

Montori, V. M., & Guyatt, G. H. (2001). Intention-To-Treat Principle. *CMAJ*, *165*(10), 1339–1341.

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies*, *22*(1), 435–480.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: a resource dependence perspective*. New York: Harper & Row.

Rose, C. (2007). Does female board representation influence firm performance? The Danish evidence. *Corporate Governance: An International Review*, *15*(2), 404–413.

Samad, H. A., Khandker, S. R., & Koolwal, G. B. (2009). *Handbook on Impact Evaluation: Quantitative Methods and Practices (World Bank Training Series)* (1st ed.). The World Bank Publications.

SIC Code. (2021). *SIC Code Lookup | SIC Code Search Tool*. SIC & NAICS Codes, Company Search, Business Lists - SICCODE.Com. https://siccode.com/sic-code-lookup-directory

Smith, N., Smith, V., & Verner, M. (2006). Do women in top management affect firm performance?A panel study of 2,500 Danish firms. *International Journal of Productivity and Performance Management*, *55*(7), 569–593.

Terjesen, S., Sealy, R., & Singh, V. (2009). Women Directors on Corporate Boards: A Review and Research Agenda. *Corporate Governance: An International Review*, *17*(3), 320–337.

Westphal, J. D., & Milton, L. P. (2000). Women In Management And Firm Financial Performance: An Exploratory Study. *Administrative Science Quarterly*, *9*(3), 366–398.

Yang, P., Riepe, J., Moser, K., Pull, K., & Terjesen, S. (2019). Women directors, firm performance, and firm risk: A causal perspective. *The Leadership Quarterly*, *30*(5).