Full Length Article

Women directors, firm performance, and firm risk: A causal perspective

Philip Yang[a], Jan Riepe[b], Katharina Moser[a], Kerstin Pull[a], Siri Terjesen[c,d,*]

[a] Department of Human Resource Management and Organization, School of Business and Economics, Tuebingen University, Germany
[b] Department of Banking, School of Business and Economics, Tuebingen University, Germany
[c] Department of Management, Kogod School of Business, American University, 4400 Massachusetts Ave., Washington, DC 20016, United States of America
[d] Department of Strategy and Management, Norwegian School of Economics, Helleveien 20, Bergen 5045, Norway

ABSTRACT

Norway was the first of ten countries to legislate gender quotas for boards of publicly traded firms. There is considerable debate and mixed evidence concerning the implications of female board representation. In this paper, we explain the main sources of biases in the existing literature on the effects of women directors on firm performance and review methods to account for these biases. We address the endogeneity problem by using a difference-in-differences approach to study the effects of women directors on firm performance with specific consideration of the common trend assumption, and we explicitly distinguish between accounting-based (i.e., operating income divided by assets, return on assets) and market-based (i.e., market-to-book ratio and Tobin's Q) performance measures in the Norwegian setting. The control group are firms from Finland, Sweden, and Denmark. We further extend the analysis of causal effects of women directors to firm risk. Our results imply a negative effect of mandated female representation on firm performance and on firm risk.

Introduction

Norway was the first country to legislate a gender-balancing quota for corporate boards of public limited firms on December 19, 2003; nine countries subsequently implemented quotas (Adams, 2016; Terjesen, Aguilera, & Lorenz, 2015). Advocates for gender-balancing quotas leverage equal opportunities perspectives to highlight a potential business case (Eagly, 2016) according to which firms with more (gender) diversity in the boardroom may perform better than their less-diverse counterparts (Kirsch, 2018; Terjesen & Sealy, 2016).

Extant *empirical* research is mixed and inconclusive with respect to the effects of gender-balancing quotas on firm performance, identifying positive, negative, or no performance effects (see for overviews: Kirsch, 2018; Pletzer, Nikolova, Kedzior, & Voelpel, 2015; Post & Byron, 2015; Joecks, Pull, & Vetter, 2013). This plurality of empirical results mirrors the *theoretical* literature in which upper echelons (Hambrick, 2007; Hambrick & Mason, 1984), resource dependence (Pfeffer & Salancik, 1978) and agency theories (Jensen & Meckling, 1976) suggest that firms with gender diverse boards might perform better, whereas, role incongruity theory (Eagly & Karau, 2002) and gender stereotyping on the part of investors (Haslam, Ryan, Kulich, Trojanowski, & Atkins, 2009) suggest a negative relationship between women directors and market-based measures of firm performance.

In light of the inconclusive empirical results and the absence of a comprehensive theory that explicitly conceptualizes a clear and precise relationship between women directors and firm performance within assumed boundaries (Bacharach, 1989; Durand & Vaara, 2009), several reviews and meta-analyses aim to provide clarity to the disparate findings. Post and Byron's (2015) meta-analysis reports that the link between women directors and firm performance depends on the choice of performance measures (accounting versus market-based) and country gender parities.

Existing reviews and meta-analyses fail to distinguish between empirical studies that are merely correlational in nature and those that seek to address the endogeneity problem inherent in the data. Because the presence of women directors does not result from exogenous variation, but rather from firm- and self-selection, it is essential to account for endogeneity when estimating the effects of women directors on firm performance (Adams, 2016; Brinkhuis & Scholtens, 2018).

Given the relevance of causality for addressing the effects of women directors, we offer a three-fold contribution to literature. *First*, we explain the main sources of biases in the existing literature on the effects of women directors on firm performance and review methods that account for these biases. *Second*, we add to the existing literature that addresses the endogeneity problem by using a difference-in-differences

approach to study the effects of women directors on firm performance with specific consideration of the common trends assumption and explicitly distinguishing between accounting-based and market-based performance measures. *Third*, we extend the analysis of causal effects of women directors to a further outcome variable highlighted in recent literature: firm risk (systematic and idiosyncratic). This extension is important for two reasons: First, firm risk is highly relevant for long-term firm success and survival (Graham & Harvey, 2001; Jeong & Harrison, 2017). Second, given the large literature on how risk preferences vary between men and women (Byrnes, Miller, & Schafer, 1999; Croson & Gneezy, 2009; Eckel & Grossman, 2006), it seems reasonable to expect that gender diversity in the boardroom affects firm risk.

By extending our analysis of the causal effects of women directors on firm risk, we respond to Adams' (2016) call for more research on how different preferences between male and female directors can affect firm strategy. Gender-balancing laws are particularly relevant for outcome variables that are arguably affected by gender differences in stable preferences such as risk aversion (Adams, 2016; Adams & Funk, 2012). Among others, Bernile, Bhagwat, and Yonker (2018) find board diversity decreases risk strategies when applying instrumental variable approaches to firm-level data. Studies also show this effect between female directorship and R&D risk when using fixed effects regressions (Chen, Ni, & Tong, 2016) or correlative meta-analyses (Jeong & Harrison, 2017). In contrast, more explicit studies on gender diversity and equity risk find no significant or direct relationship when reverse causality is accounted for (Sila, Gonzalez, & Hagendorff, 2016), but find strong indirect effects (Main, Gonzalez, & Sila, 2018). Further, recent literature emphasizes the role of selection that would limit the generalization of correlational studies if differences between the two genders in the boardroom did not mirror differences in the general population (Adams & Funk, 2012; Adams & Ragunathan, 2017; Sapienza, Zingales, & Maestripieri, 2009).

Endogeneity in the boardroom literature

Sources of endogeneity

Although the statement "correlation does not imply causation" has become a standard set phrase for researchers in the field of board diversity, few studies empirically address causality (Adams, 2016; Post & Byron, 2015). This absence is particularly troubling in light of the current popularity and effectiveness of quotas and the need to craft policy recommendations based on causal scientific findings (Adams, 2016; Antonakis, Bendahan, Jacquart, & Lalive, 2010; Eagly & Antonakis, 2015). Most causal studies measure the impact of increased female representation on firm performance through timing, true relationship, and non-spuriousness (Kenny, 1979). In contrast, correlational studies could be biased with respect to estimate direction, size, and significance because correlational studies are likely plagued by either one or multiple sources of endogeneity (Antonakis, 2017; Brinkhuis & Scholtens, 2018).

Endogeneity is present when the treatment variable correlates with the error term, because it is neither randomly assigned nor measured under conditional independence (Angrist & Pischke, 2010; Cameron & Trivedi, 2008). In these instances, the coefficients of the estimated regressions are not causal because whether the treatment variable or any other unobserved variable is responsible for the changes in the outcome variable is unclear (Antonakis et al., 2010). The most common sources of endogeneity in the current empirical research on gender diversity in boards are omitted variable and selection biases.

Omitted variable bias is present when regressions fail to control for variables that affect the treatment variable, the outcome variable, or both (Cameron & Trivedi, 2008; Wooldridge, 2010). Director ability and time-invariant firm characteristics are the most relevant variables that current literature on gender board representation tends to omit. In the presence of ability bias, regression estimates of female representations that do not account for the ability of female directors are likely upward biased for the positive association of the ability of (better-qualified) female directors with female representation and firm performance (Antonakis et al., 2010; Eagly & Antonakis, 2015). Omitted fixed effects occur if researchers choose not to use longitudinal data structure to distinguish within and between effects (Antonakis et al., 2010; Halaby, 2004). Even though the use of fixed effects comes at costs to multi-level analyses (Wooldridge, 2010), recent studies by Adams and Ferreira (2009) and Adams (2016) reveal that excluding fixed effects in gender board studies likely causes Simpson paradox related biases. Both studies demonstrate that regressions without these fixed effects show a positive relationship between female representation and firm performance, which turns negative as soon as fixed effects are included. Thus, clustering standard errors to account for repeated observations of firms does not solve this problem because standard error clustering only affects the significance, but not the size and direction of the estimated coefficients (Petersen, 2009).

Selection bias is one of the most discussed sources of endogeneity in the literature on gender diversity in boards (Adams, Hermalin, & Weisbach, 2010; Hermalin & Weisbach, 1998, 2003). In general, selection bias occurs when treatment and control groups differ in a systematic way and researchers fail to account for the selective process by which this treatment is assigned (Angrist & Pischke, 2010; Cameron & Trivedi, 2008). The endogenous nature of female representation is most striking with respect to selection in terms of firms' attributes (Ahern & Dittmar, 2012; Brinkhuis & Scholtens, 2018) and directors' characteristics (Adams & Funk, 2012; Adams & Ragunathan, 2017). Selection bias for firm's attributes is present when the share of female directorships varies as a function of differences in firms. A prominent illustration of this bias is firm size and the presence of female directors. If larger firms are systematically different from smaller firms with respect to their performance yet also differ in their likelihood of hiring female directors, any estimate that does not account for firm size is biased toward the difference between larger and smaller firms (Ahern & Dittmar, 2012; Brinkhuis & Scholtens, 2018).

Selection bias with respect to director characteristics refers to the systematic way that females with different core values and preferences select themselves in director positions. If gender differences in society do not mirror gender differences in the boardroom, it is possible that female directors are similar to their male counterparts or even score higher on supposedly "male" attributes for they would have not broken through the glass ceiling otherwise (Adams & Funk, 2012). One example is the "Lehman sisters" argument by policy makers to improve female board representation in the banking sector as firms with more women are expected to engage in less risky activities (Adams & Ragunathan, 2017). As reasonable as this idea sounds, it is only applicable if the female candidate pool for banking sector director positions is more risk averse compared with their male counterparts. In the worst case, such a gender policy could actually lead to the opposite outcome if only overly risky females choose the banking sector.

Unfortunately, endogeneity trickles down and further decreases the value of meta-analyses and systematic reviews (Antonakis, 2017; Ioannidis, 2016). Considering the large amount of endogeneity that feeds into these types of analyses, the credibility of meta-analyses decreases with the addition of any inconsistent finding. In the case of female directors, this bias can translate into positive mean effect sizes as in Post and Byron (2015) if the majority of studies exclude adjustments for selection and time invariant firm characteristics. In fact, Post and Byron (2015) conclude that their meta-analysis is unable to claim causality because only a minority of their meta-analytic sample studies address endogeneity. Although more recent meta-analyses such as Pletzer et al. (2015) find no significant relationship between women directorship and firm performance when only including published articles, these findings are not more informative as long as studies are not chosen and classified by the degree of endogeneity.

Methods to limit endogeneity

The best way to solve endogeneity is to randomize treatment (Angrist & Pischke, 2009) which is possible in experimental studies in which a researcher randomly assigns participants to a treatment and a comparable control group that hence eliminates selection and ability biases (Antonakis et al., 2010). However, these experiments are difficult to carry out in the context of female board representation, because firms will unlikely agree to assign leading roles at random (Adams, 2016).

The second-best solution is just as impractical and unrealistic. Referred to as the conditional independence assumption or selection-on-observables, any regression of female directorship on firms' performance approximates causality with the inclusion of control variables for any relevant confounding variable that affects either performance, female directorship, or both without being an outcome of the treatment itself (Angrist & Pischke, 2009; Cameron & Trivedi, 2008). However, this statistical adjustment can easily turn into a bottomless pit for it is almost impossible to identify and collect data on all of these variables.

In contrast to experiments and conditional-independence, empirical identification through instrumental variable approaches, regression discontinuity design (Cook, 2008), propensity scores (Rosenbaum & Rubin, 1983; Rubin, 2008), Heckman selection models (Heckman, 1979), and quasi-experimental methods are more realistic means to establish causality (Angrist & Pischke, 2009; Antonakis et al., 2010).

One prominent type of quasi-experimental methods that we use in our study is the difference-in-differences approach, which is applicable when an exogenous shock such as a law affects a treatment group, but not a comparable control group (Angrist & Krueger, 2007; Angrist & Pischke, 2009; de Cabo, Terjesen, Escot, & Gimeno, 2019). By comparing the pre- and post-reform differences between treatment and control groups of firms with respect to female directorship and firm performance, this method accounts for cross-sectional differences and time trends (Adams, 2016; Antonakis et al., 2010). As straightforward as the difference-in-differences approach is, its use is limited by the restrictive common trend assumption that requires graphical proofs of common trends with respect to the outcome variable between the treatment and control group before and after treatment (Angrist & Krueger, 2007; Imbens & Wooldridge, 2009; Lechner, 2010). Other notable adjustments for difference-in-differences estimations in the context of gender diversity are the inclusion of firm and year fixed effects to rule out time invariant firm characteristics and time trends (Eckbo, Nygaard, & Thorburn, 2018; Matsa & Miller, 2013) as well as standard errors clustered at the firm level to avoid serial correlation (Bertrand, Duflo, & Mullainathan, 2004). One seminal example is Matsa and Miller's (2013) study of the Norwegian quota to compare firm performance by using a triple difference between treated Norwegian firms, untreated Norwegians firms, and firms from neighboring countries.

Our empirical strategy is to use the Norwegian setting to analyze how the quota affects firm performance and risk through the increased share of female directorships by using a difference-in-differences approach. Our data restrictions and empirical identification are in line with Matsa and Miller (2013). We limit endogeneity by considering how the quota affects the relative performance (and risk) difference between the treatment group of Norwegian firms, and the control group of firms from Finland, Sweden, and Denmark. Various scholars use the Norwegian setting to analyze the effect of women directors on firm performance (Ahern & Dittmar, 2012; Eckbo et al., 2018; Matsa & Miller, 2013).

Method

Sample and data

We analyze the causal effect of the Norwegian quota on the various performance outcomes for firms by using data on their non-executive board members from BoardEx. BoardEx contains the share of female non-executive directors, as well as detailed information on their average tenure, experience, age, nationality, and educational degree. We combine the BoardEx dataset with financial data from the Thomson Reuter EIKON database that reports several firm level accounting- and market-based measures. We use yearly information from firm balance sheets for the whole sample because most firms only provide annual audited financial statements during our sample period. Both datasets are merged through the ISIN codes, year, and month of the report date.

Most restrictions are in line with Matsa and Miller (2013). Our data is limited to four countries for the difference-in-differences estimations with Norway as the treatment group, and Sweden, Denmark, and Finland as the control group. Like Matsa and Miller (2013), we exclude firms from financial and petroleum sectors and only consider firms with complete information on all board level and performance variables. We exclude the few firms that are subject to merger and acquisition (Martin & McConnell, 1991) or financial distress (Brown & Matsa, 2016; Opler & Titman, 1994). Finally, we apply a one-to-one matching method based on the performance variables of the respective firms in 2004 to increase the similarity within the sample (Leuven & Barbara, 2003; Rosenbaum & Rubin, 1983).

We then return to market data from the daily stock returns provided by Thomson Reuter EIKON for our risk variables. Data from firm return indices is already adjusted for dividends, stock splits, and equity issuing. Following Ince and Porter (2006), we exclude temporary and large price jumps from data errors as well as illiquid return series (Lesmond, Schill, & Zhou, 2004) and penny stocks. We aggregate the daily stock returns on a quarterly basis and merge market data to balance sheet information from the previous quarter, because balance sheet information is available to the public approximately three months after the end of the financial reporting period. The risk measures are then merged to the dataset for female board representation and performance and restricted to those firms with information on all board and risk variables. Notably, this procedure generates two subsamples that we use to analyze annual performance and quarterly risk. Because our sample restrictions for the risk sample are more restrictive than those in Matsa and Miller (2013), our risk sample includes fewer firms relative to the performance sample.

Firm performance

We assess firm performance through both accounting and market measures. Firms report accounting-based measures such as return on assets, return on equity, and return on invested capital according to legally enforceable and independently audited accounting principles (Combs, Crook, & Shook, 2005; Haslam et al., 2009; Post & Byron, 2015). In contrast, market-based measures are shaped by investor sentiments (Akerlof & Shiller, 2009; Barberis & Thaler, 2003; Haslam et al., 2009), behaviors, and beliefs (Haslam et al., 2009; Keynes, 1964 (originally 1936)) as well as analysts' views on future earnings potential (Dechow & Sloan, 1997; Haslam et al., 2009). Because market data take the investors' perspective (Brinkhuis & Scholtens, 2018), they are forward-looking whereas accounting variables only incorporate information from the reporting period. We use both accounting and market variables to distinguish these conceptually different outcomes and to enable comparison to previous studies.

As in Matsa and Miller (2013), our accounting measures are operating income divided by assets (OI/A) and return on assets (ROA). We calculate ROA with the ratio of earnings before interests and taxes (EBIT) to total assets, and winsorize the data at 1% and 99% levels to limit the influence of outliers on our regression coefficients. Our market-based performance measures follow Post and Byron (2015) and are Market-to-book ratio (MTBR) and Tobin's Q (the ratio of a firm's market valuation divided by its replacement value), which we also winsorize at 1% and 99% levels. The MTBR variable is divided by 100 to obtain more readable coefficients. MTBR exclusively takes the equity

investor's perspective and shows the relationship between the market value of a firm's equity divided by its book value.[1] MTBR reflects the expected value gains to the equity investors from the firm's past and present strategic decisions scaled by the time value of the equity amount injected into the firm. We also use Tobin's Q which reflects the market's assessment of a firm's total assets[2] by their replacement value (Tobin, 1969) and thereby yields a more comprehensive picture. Because Tobin's Q considers all the firm's assets, it can be easily compared across firms without adjusting for risk, leverage, or size (see, e.g., Stulz, 1994; Wernerfelt & Montgomery, 1988). Thus, Tobin's Q is preferable to other capital market measures such as the stock price. Although ratio variables bear the risk of spurious correlations (e.g., Kronmal, 1993), we use the ratio variable as a performance measure. We avoid biases from spurious correlations by carefully selecting our variables, avoiding scaling the dependent and independent variables by the same factor, and not using the scale of the dependent variable as a separate independent variable. Furthermore, we re-estimate the results without a control variable in the appendix and show that our results remain qualitatively unchanged.

Firm risk

Firm risk is highly relevant for long-term firm success and survival (Graham & Harvey, 2001; Jeong & Harrison, 2017). Various stakeholders benefit from lower firm risk (Cornell & Shapiro, 1987). Suppliers and employees are especially interested in idiosyncratic risk which is closely linked to a firm's default risk (Brown & Matsa, 2016; Hallikas, Karvonen, Pulkkinen, Virolainen, & Tuominen, 2004). Equity investors are concerned with lower systematic risk, which compensates the firm's equity investors for lower stock performance and also presents potential economic advantages to the firms' other stakeholders.

Our first risk measure is the volatility of a firm's equity returns: *Equivola* (Bernile et al., 2018; Jeong & Harrison, 2017; Perryman, Fernando, & Tripathy, 2016) which captures the overall risk with relevant consequences to the firms' equity investors. We use a rolling window of firm stock returns in the coming year to estimate their annual volatility. In a second step, we break up firm annual volatility into its systematic and idiosyncratic components (Bernile et al., 2018; Perryman et al., 2016). *Systematic risk* reflects the correlation between a firm's stock returns and the market return, which cannot be diversified by the equity investor, and consequently shapes the firm's cost of equity. The systematic risk is expressed by the correlation coefficient *Beta* and by the systematic volatility, and is highly relevant to the firm's equity investors. In contrast, *Idiosyncratic risk* of a firm's equity returns is less relevant to the equity investors, but of great concern to other stakeholders, such as loan investors, employees, and customers who would heavily suffer from the firm's default. We break-up the firm's volatility into systematic and idiosyncratic volatilities by applying a simple market model in which the EuroStoxx 50 serves as our market index:

$$Y_{i,t} = \beta_{i,t} * \text{EuroStoxx } 50_t + \epsilon_{i,t}$$

where $\beta_{i,t} * \text{EuroStoxx } 50_t$ captures the systematic risk of firm i at time t, and $\epsilon_{i,t}$ captures the idiosyncratic risk.

Control variables

The use of control variables, even when adding potentially irrelevant ones (Cameron & Trivedi, 2008), competes with the risk of

including bad controls that may also be affected by the treatment (Angrist & Pischke, 2009). Hence, even well-identified studies such as Ahern and Dittmar (2012) and Matsa and Miller (2013) conduct most regressions with only firm fixed effects and industry-year effects to exclude time invariant differences. The robustness of those results is then tested with some control variables. Accordingly, we first estimate our regressions with only fixed effects. We then include control variables used in recent literature and compare how the estimates change with their inclusion to validate the robustness of our results. Because both results are very similar, we report the results with control variables in the main part of the paper and the estimates without controls in the appendix.

Our regressions control for *board size* in terms of the overall number of non-executive board directors (Ahern & Dittmar, 2012; Eckbo et al., 2018; Matsa & Miller, 2013). We also hold constant the effect of *average age* of non-executive directors (Ahern & Dittmar, 2012; Matsa & Miller, 2013). Further, we include several variables to account for non-executive board members' knowledge (Ahern & Dittmar, 2012; Matsa & Miller, 2013). As such, *average tenure on the specific board* measures the firm specific knowledge by averaging non-executive board members' years on the board. In a broader sense, *average experience in quoted and private boards* refers to non-executive board directors' overall experience in their functions as directors. Finally, we include *nationality mix*, an index variable that ranges from 0 to 1 to approximate the share of non-national directors, and *education* as a control variable that reports the average level of non-executive board directors' educational degree measured in terms of the number of educational degrees of non-executive directors above bachelor level (Ahern & Dittmar, 2012).

Analysis

Difference-in-differences estimates apply in natural experimental settings in which one policy reform, such as a board quota, affects a treated group but not a comparable control group (Angrist & Krueger, 2007; Angrist & Pischke, 2009; Bertrand et al., 2004). Therefore, we analyze the differences between the treatment and control groups before the reform and then how these differences change with the implementation of the reform.

We run several analyses to test for the common trend assumption in the outcome variables (Angrist & Krueger, 2007). Fig. 1 shows the final graphical results for our accounting and market variables. Our data violates this restrictive assumption for these variables when considering the same time periods as Matsa and Miller (2013) or Eckbo et al. (2018). In fact, the common trend in the treatment and control group with respect to performance is only met in our data for the reduced observation period between 2002 and 2008. The different trends in the pre-period might possibly come from the bust of the Dotcom-bubble that affected countries in different ways (Brunnermeier & Oehmke, 2013). Also, the post-2008 period is subject to the global financial crisis (Aiyar, 2012) that affected the sample countries in heterogeneous ways (Jensen & Johannesen, 2017). Following this sample restriction, our final sample for the difference-in-differences regressions with the performance outcomes includes 622 firm-year observations between 2002 and 2008.

We also test for common trends in the risk variables between firms in the treatment and control groups, and report corresponding graphs in Fig. 2. Whereas the common trend assumption in the volatility of a firm's equity returns, idiosyncratic risk, and systematic risk for the treatment and control groups is met for the same observation period, the common trend is less precise for the beta. The final sample for the difference-in-differences estimations with either risk variable as the outcome contains 2124 quarterly observations between 2002 and 2008.

Norway's gender quota was first proposed on a voluntary basis in 2003. Therefore, we use post 2003 years in Norway as our time treatment variable. Nevertheless, it was unclear in the beginning whether all firms would comply with the voluntary quota. For this reason, we

---

[1] The book value of equity consists of the accumulated amount of equity initially issued minus the share repurchased plus retained earnings minus the dividends paid to the equity investors.

[2] We measure Tobin's Q by the firm's market value of equity + book value of remaining assets all divided by the firm's book value of total assets.
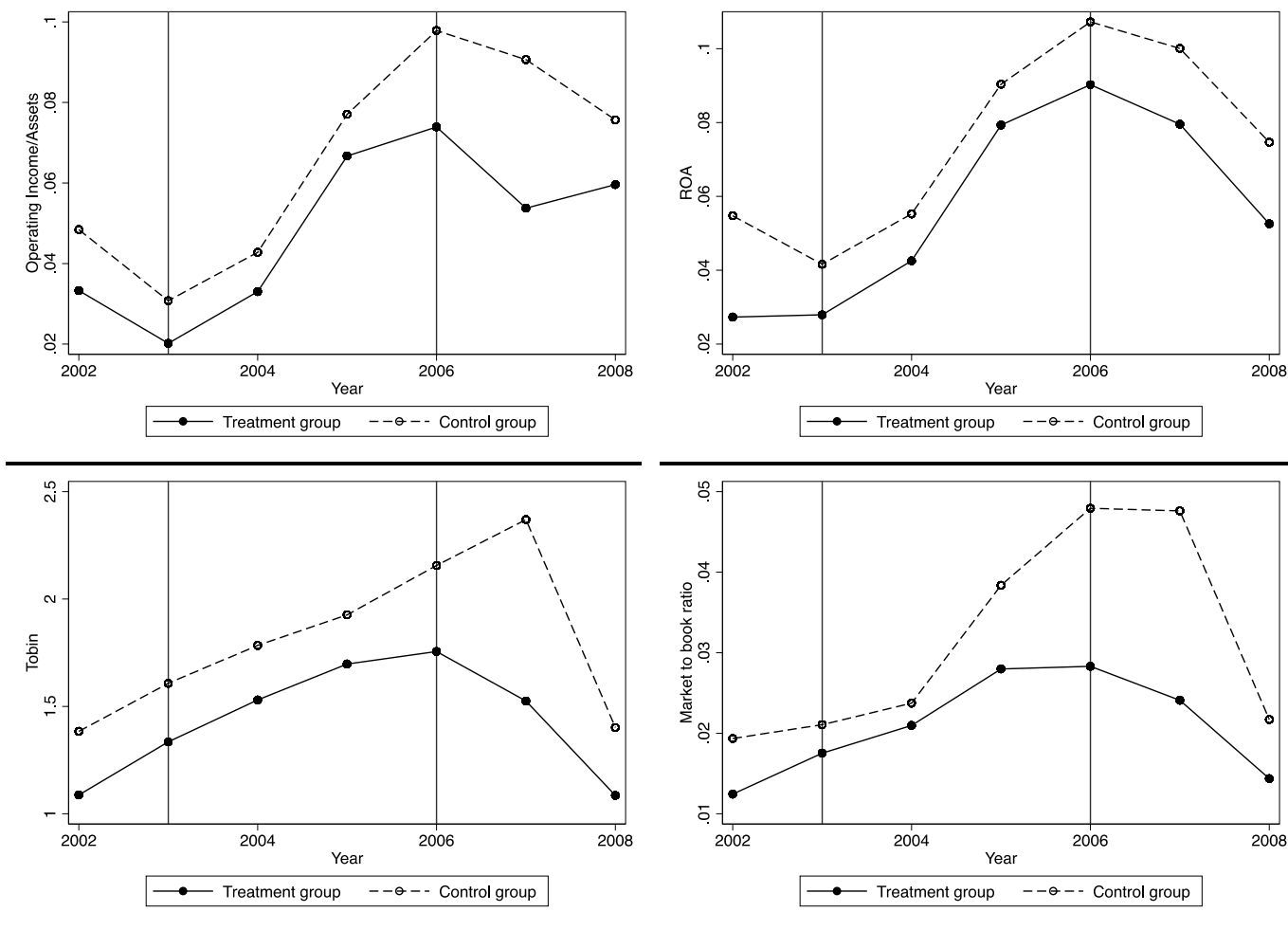
Fig. 1. Common trends in firm performance.

applied a graphical analysis to determine when female representation truly increased for the treatment group. After 2003, the government also converted the voluntary into a mandatory quota to underscore its intention to enforce the quota. Fig. 3 shows that the treatment group started to drastically increase the share of female non-executives directors only after 2004. For this reason, we also run a robustness analysis in the Appendices A3 to A6 in which we consider all years after 2004 as post-treatment years.

We report the summary statistics and the correlation matrix of the main variables used in Table 1. The correlations of the upper panel of variables are estimated using the yearly sample, whereas the correlations of the lower panel of variables are estimated with the quarterly sample. Overall, these correlates support the relevance of the control variables that we use for the subsequent analysis, because performance is positively associated with board size and director tenure and negatively associated with larger nationality mix. At the same time, we observe a negative association between female board representation and board directors' age and experience, but a positive association with directors' education. Whereas our accounting (and market) measures are strongly correlated among each other, the association *between* the accounting and market variables is considerably weaker.

These weaker associations illustrate the differences in the underlying economic concepts of accounting and market measures. Accounting measures reflect the amount of firm income generated within the previous reporting period, typically one year. Accounting measures are therefore backward looking (Altman, 1968; Beaver, 1966) and can be distorted by one-time effects such as the cost of restructuring. Long-term consequences, such as the profit from a long-

term investment project, only gradually enter the accounting measures. Because firms report accounting measures, they are also subject to managerial discretion as documented and discussed by Graham, Harvey, and Rajgopal (2005) and Dechow (1994). Market measures, in contrast, come from the capital market and reflect the equity investors' expectations on firm economic development. Market measures are thus entirely expectations based and forward looking. The future cash flows are discounted by the cost of capital and therefore consider a firm's systematic risk. Market performance might be biased by the general market conditions and investor sentiment (Baker & Wurgler, 2007). Because market measures directly incorporate the expected future economic consequences of a firm's strategic decisions, these measures might be better suited to study the consequences of female board members on a firm's economic development.

The selective response to the quota by firms is one main concern against the use of an ordinary difference-in-differences approach in the Norwegian context. Some Norwegian firms avoided the quota by delisting and becoming private (Ahern & Dittmar, 2012; Bøhren & Staubo, 2014); however, most of these firms already delisted before the quota came into effect (Eckbo et al., 2018; Nygaard, 2011). Matsa and Miller (2013) consider triple-difference estimations to control for this bias. Because Matsa and Miller (2013) find negligible differences between the simple difference-in-differences results and the triple-difference approach, we estimate the effect of mandated female representation with a simple difference-in-differences regression:

$$Y_{ijt} = \beta_1 Treat_j * Post2003_t + \lambda_i Year_t + X_{ijt} + \alpha_j + \tau_t + \epsilon_{ijt}$$

with Y_{ijt} denoting firm level outcome for the firm j in the sector i during
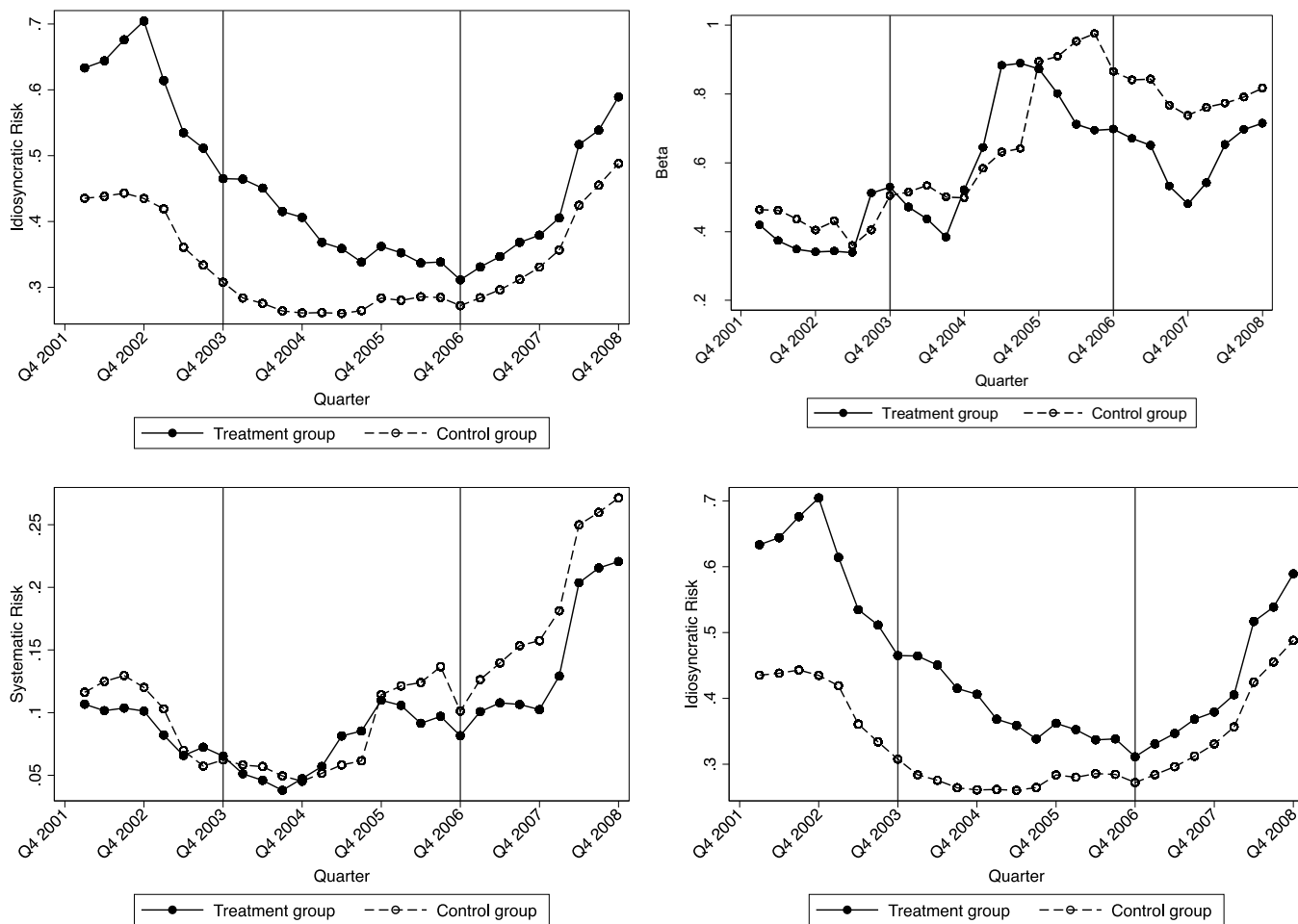
Fig. 2. Common trends in firm risk.

period t, explained by the mandated female board representation $Treat_j * Post2003_t$ while holding constant firm fixed effects α_j, industry specific time trends λ_i, and year effects τ_t. Further, we consider a set of control variables X_{ijt} that includes the overall board size as well as the

non-executive directors' average tenure, experience in quoted and private boards, age, nationality mix, and post-bachelor's educational degrees. All regressions are estimated with standard errors clustered by firms (Bertrand et al., 2004). To further validate our results, we run

Fig. 3. Female board representation.

Table 1
Summary statistics and correlation matrix.

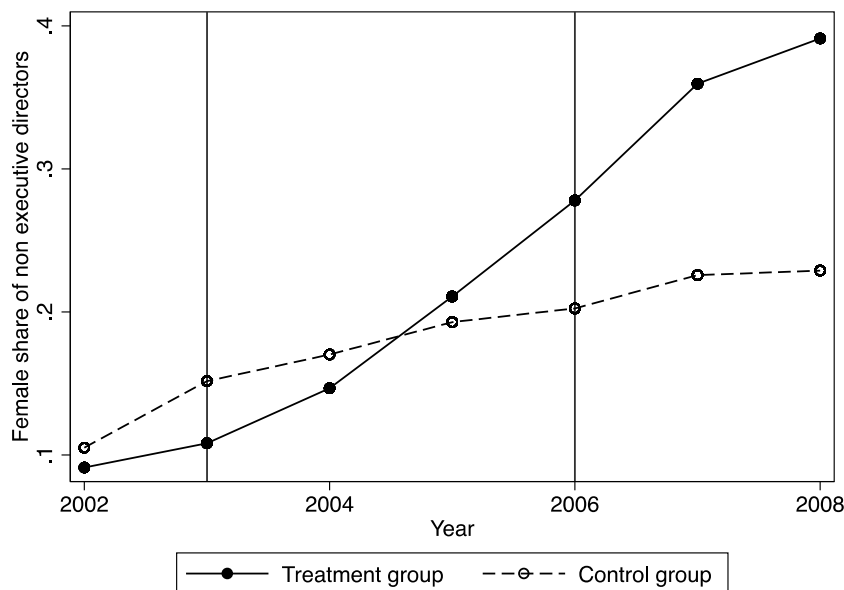|	Mean	(SD)	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) OI/A	0.061	(0.103)	−0.442	0.450	—																	
(2) ROA	0.071	(0.106)	−0.395	0.469	0.92	—																
(3) Tobin	1.700	(1.454)	0.349	27.950	0.09	0.11	—															
(4) MTBR/100	0.028	(0.059)	0.001	0.999	0.13	0.14	0.39	—														
(5) Treated	0.301	(0.459)	0	1	−0.07	−0.07	−0.13	−0.09	—													
(6) Post 2003	0.711	(0.453)	0	1	0.15	0.16	0.12	0.10	0.00	—												
(7) Female share	0.195	(0.154)	0	0.600	0.11	0.09	0.17	0.08	0.13	0.31	—											
(8) Board size	8.553	(3.156)	2	21	0.18	0.17	−0.09	−0.11	−0.32	−0.05	0.17	—										
(9) Tenure	4.297	(2.054)	0.300	12.500	0.17	0.19	0.02	0.02	−0.26	0.05	−0.03	0.17	—									
(10) Exp.-quoted	3.061	(1.594)	1	10.500	0.06	0.03	−0.01	0.14	−0.34	0.07	−0.01	−0.04	0.09	—								
(11) Exp.-private	7.082	(3.942)	1	32.100	0.01	−0.02	0.06	0.03	−0.29	0.07	−0.11	0.01	0.07	0.24	—							
(12) Age	53.883	(4.335)	39	67.200	0.07	0.08	−0.03	−0.07	−0.36	0.03	−0.21	0.03	0.25	0.28	0.08	—						
(13) Nationality mix	0.136	(0.205)	0	0.800	−0.08	−0.10	0.12	0.01	−0.06	0.07	0.06	0.08	−0.12	0.28	−0.06	0.19	—					
(14) Education	1.443	(0.511)	0.100	3.600	−0.05	−0.05	0.01	−0.01	−0.07	0.20	0.09	−0.20	−0.06	0.57	0.15	0.21	0.30	—				
(15) Equivola	0.390	(0.177)	0.116	1.228	−0.33	−0.39	−0.01	−0.04	−0.27	−0.12	−0.08	−0.35	−0.28	−0.10	−0.06	−0.16	−0.06	−0.06	—			
(16) Beta	0.626	(0.425)	−0.611	1.842	0.16	0.12	0.04	−0.01	0.31	0.38	0.14	0.10	0.06	0.24	0.08	0.27	0.16	0.25	0.07	—		
(17) Systematic risk	0.110	(0.090)	−0.110	0.431	0.12	0.06	−0.03	−0.05	0.11	0.31	0.13	0.07	0.04	0.23	0.06	0.23	0.15	0.18	0.36	0.77	—	
(18) Idiosyncratic risk	0.365	(0.172)	0.114	1.228	−0.38	−0.43	0.00	−0.04	−0.31	−0.19	−0.11	−0.39	−0.30	−0.15	−0.07	−0.21	−0.08	−0.10	0.98	−0.06	0.19	—

Notes: Correlations between variables 1 and 14 are estimated using the yearly sample (N = 662); variables 15–18 with the quarterly sample (N = 2124).

robustness checks using fractional logit models (Papke & Wooldridge, 1996) for those outcome variables bounded between 0 and 1.

Results

We report and discuss our results from the difference-in-differences estimations in four subsections. The first subsection reports our findings with respect to the changing board structure driven by mandated female representation. In the second subsection, we report results for the causal effect of the reform on firm performance. The third subsection considers the implications of mandated female representation to risk, and the fourth subsection contains our robustness checks.

Board structure

Table 2 reports the results of difference-in-differences regression with the various variables for board structure as outcomes. These regressions are conducted in the same manner as Matsa and Miller (2013) and include firm and industry-by-year fixed effects, but no further control variables. We find that the quota significantly increases female representation on the boards of the firms in the treatment group as compared to the control group ($\beta = 0.078$, $p = 0.005$). This result clearly emphasizes the effectiveness of quotas in promoting female representation, in particular a stringent "hard law" with sanctions, as is the case in Norway.

Our results further show that once the quota was implemented, directors' average educational degree level increases ($\beta = 0.139$, $p = 0.066$). Board size does not increase when we use "post 2003 years" as the time treatment variable, but it does increase when we use "post 2004 years" as a time treatment variable in the robustness check (see Appendix A3) which is in line with several other studies of the Norwegian setting (Ahern & Dittmar, 2012; Bertrand, Black, Jensen, & Lleras-Muney, 2018). We find no significant effect of the reform on board directors' average tenure, age, and nationality mix. However, the coefficient direction indicates a post-quota decrease in board directors' average age (Ahern & Dittmar, 2012).

Firm performance

Table 3 reports the findings with respect to firm performance. These regressions include firm and industry-by-year fixed effects, as well as board level controls.

We use a different data source and observation period but the same setting, similar restrictions, and same estimation method as in Matsa and Miller (2013). In comparison to Matsa and Miller (2013) and Eckbo et al.'s (2018) replication, our study carefully considers the common trends of the outcome variables and alternative post reform years to assess the robustness of those results. We find almost the same result with respect to the accounting measures. Although the coefficient size for the effect of the reform on the OI/A is almost identical ($\beta = -0.038$ in our study vs. $\beta = -0.034$ in Matsa and Miller (2013)), our result is slightly improved with respect to significance ($p = 0.002$). Our estimate on the effect of the mandated female representation on ROA is negative and significant ($\beta = -0.026$, $p = 0.036$). We find similar results for the reform when considering treated firms' relative market performance. The difference-in-differences estimates for Tobin's Q ($\beta = -0.210$, $p = 0.307$) and MTBR ($\beta = -0.010$, $p = 0.111$) are negative, albeit not statistically significant when considering post 2003 years as the treatment years. Because uncertainty in the market only resolves after firms engaged in quota-induced hiring of female board members, both coefficients turn marginally significant when we move the reform start date from 2004 to 2005, i.e. when we use "post 2004 years" instead of "post 2003 years" as the time treatment variable (see Appendix A4). If we use "post 2004 years" as the time treatment variable, the coefficients for Tobin's Q ($\beta = -0.346$, $p = 0.066$) and MTBR ($\beta = -0.013$, $p = 0.097$) are negative and larger in size. These effects are similar with respect to the significance and direction of the instrumental variable

Table 2
Changes in board structure.

Non-executive directors	Share of female directors	Board size	Average tenure	Average age	National. mix	Average educ.
Main explanatory variable						
DID (Norway × post 2003)	0.078***	0.305	0.355	−0.252	0.039	0.139*
	(0.027)	(0.336)	(0.333)	(0.635)	(0.031)	(0.075)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of firms	109	109	109	109	109	109
Observations	662	662	662	662	662	662
R-squared	0.464	0.105	0.110	0.138	0.128	0.273

Notes: This table summarizes the results from a firm- as well as industry-by-year fixed effects regression of firm's share of female non-executive directors, board size as well as average tenure, average age, nationality mix, and average amount of educational above-bachelor degrees for non-executive directors explained by the a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2003. Standard errors are clustered by firms and reported in parentheses.

*** p < 0.01.
* p < 0.1.

Table 3
Firm accounting and market performance.

Firm performance	OI/assets	ROA	Tobin	MTBR
Main explanatory variable				
DID (Norway × post 2003)	−0.038***	−0.026**	−0.210	−0.010
	(0.012)	(0.012)	(0.204)	(0.006)
Control variables				
Board size	−0.005	−0.006*	0.013	−0.001
	(0.003)	(0.003)	(0.025)	(0.001)
Tenure	0.003	0.002	−0.036	0.001
	(0.003)	(0.003)	(0.024)	(0.002)
Experience quoted	−0.009	−0.016*	−0.087	0.010
	(0.006)	(0.008)	(0.074)	(0.011)
Experience private	0.000	0.000	0.028	−0.004
	(0.002)	(0.003)	(0.036)	(0.004)
Age	0.000	0.002	0.024	−0.001
	(0.002)	(0.002)	(0.016)	(0.001)
Nationality	−0.002	−0.004	0.313	0.013
	(0.031)	(0.031)	(0.481)	(0.011)
Education	0.004	0.002	0.047	−0.005
	(0.016)	(0.018)	(0.183)	(0.007)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	109	109	109	109
Observations	662	662	662	662
R-squared	0.213	0.209	0.156	0.110

Notes: This table summarizes the results from a firm and industry-by-year fixed effects regressions of firm performance measured by operating profit/asset, return/asset, Tobin's Q, and MTBR, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2003. The estimate for MTBR is almost the same when applying a fractional logit model. Standard errors are clustered by firms and reported in parentheses.

*** p < 0.01.
** p < 0.05.
* p < 0.1.

estimates of Ahern and Dittmar (2012), although smaller in size. In fact, our coefficient size is close the estimates of the replication by Eckbo et al.'s (2018), which uses the same observation period as we do with an alternative instrument to find a non-significant and negative effect.

The coefficients are also significant in economic terms. An increase from one to two female board members on a board with four directors reduces that firm's operating income to assets by 12%.[3] This value equals twice the size of the standard deviation in the OI/A.

---

[3] In Table 2, we see the share of female board members increases by 7.8% faster for our treatment group and find an effect of −0.038 in the additional increase in OI/A. An increase from 25% to 50% would therefore lead to a decrease of −0.038 / 0.078 * 0.25 = −0.12. Thereby, the effect is twice as large as the OI/A's standard deviation (see Table 1).

Firm risk

Table 4 reports the difference-in-differences estimates of the regressions with firm risk as outcome variable. The regressions are conducted in the same manner as those for performance as the dependent variable. However, the sample size is larger as we use quarterly information on the risk variables.

Our results imply a negative effect of the reform on firm risk. As previously mentioned, the estimates for Beta (β = −0.199, p = 0.030) must be treated with caution as our graphical analysis suggested a potential violation of the common trend. Our results support Bernile et al.'s (2018) finding of decreased risk in board decisions (β = −0.114, p = 0.005) at the higher level of gender diversity. Further, our results show negative effects of the reform on systematic (β = −0.027, p = 0.170) and idiosyncratic risk (β = −0.100, p = 0.018) that are similar to findings by Perryman et al. (2016), Faccio, Marchica, and Mura (2016), and Sila et al. (2016). The effect on systematic risk turns marginally significant when we move the reform start date from 2004 to 2005, i.e., when we use "post 2004 years" instead of "post 2003 years" as the time treatment variable (see Appendix A5). Because systematic risk is relevant to equity investors (Graham & Harvey, 2001), lower systematic risk partially compensates these investors for the lower profitability and puts the lower performance caused by the gender-balancing quota into context. In addition, lower idiosyncratic risk provides economic benefits to other firm stakeholders such as suppliers (Hallikas et al., 2004) and employees (Brown & Matsa, 2016).

Robustness

Our difference-in-differences regressions determine the causal effect of mandated female board representation on firm performance using Norway as the treatment group and a control group that consists of three neighboring countries (Sweden, Denmark, and Finland). Because the control group is also treated at a constant rate by the reform, our difference-in-differences is fuzzy (Chaisemartin & D'HaultfŒuille, 2018), which allows the identification of the local average treatment effect of female representation on firm performance using the Wald difference-in-differences estimator. However, Fig. A1 shows that the female share for Finland does not follow a constant treatment rate over the observation years. For this reason, we employ a further robustness check and test the robustness of our results by estimating the Wald difference-in-differences estimate for all our firm performance variables with and without firms from Finland. Table 5 reports those results and shows no significant differences when excluding firms from Finland.

Table 4
Firm risk.

Firm risk	Equivola	Beta	Systematic risk	Idiosyncratic risk
Main explanatory variable				
DID (Norway × post 2003)	−0.114***	−0.199**	−0.027	−0.100**
	(0.040)	(0.090)	(0.020)	(0.042)
Control variables				
Board size	−0.003	0.001	−0.001	−0.002
	(0.005)	(0.012)	(0.002)	(0.004)
Tenure	−0.004	0.007	0.001	−0.004
	(0.006)	(0.014)	(0.003)	(0.005)
Experience quoted	0.034**	0.005	0.000	0.032**
	(0.015)	(0.037)	(0.007)	(0.014)
Experience private	−0.004	−0.002	−0.001	−0.003
	(0.007)	(0.013)	(0.003)	(0.006)
Age	−0.000	0.018***	0.004***	−0.002
	(0.004)	(0.007)	(0.001)	(0.004)
Nationality	0.026	0.031	0.006	0.030
	(0.058)	(0.168)	(0.038)	(0.055)
Education	−0.029	0.167**	0.031*	−0.038
	(0.034)	(0.079)	(0.016)	(0.032)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	87	87	87	87
Observations	2124	2124	2124	2124
R-squared	0.477	0.341	0.554	0.446

Notes: This table summarizes the results from firm and industry-by-year fixed effects regressions of firm market risk measured by equity volatility, market beta, systematic risk, and idiosyncratic risk, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2003. Standard errors are clustered by firms and reported in parentheses.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

Table 5
Local average treatment effects (LATE).

	OI/A	ROA	Tobin	MTBR	Equivola	Beta	Systematic risk	Idiosyncratic risk
LATE (with Finland)	−0.487	−0.333	−2.692	−0.128	−1.463	−2.557	−0.349	−1.284
LATE (without Finland)	−0.474	−0.333	−2.628	−0.128	−1.351	−2.168	−0.254	−1.173

Notes: This table summarizes the results of the local average treatment effects from a firm and industry-by-year fixed effects regressions of firm performance and risk measures by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2003. Because of potential concerns on simultaneous changes in one of the control countries, Finland, we compare results with and without the control firms from Finland.

Discussion

Key findings

Analyzing the causal effects of the Norwegian gender-balancing quota, we find the quota significantly increases the share of women directors on the boards of treated firms. Further, we find the quota significantly adversely affects the performance of treated firms and firm risk is significantly reduced.

Concerning the effects on *board structure*, the significant increase in the share of women directors is not associated with a change in average age or nationality mix, but is associated with a slight increase in average educational level. Interestingly, average tenure did not change for the treated firms even though bringing in new directors to replace existing ones will necessarily lead to a reduction in average tenure. Apparently, the untreated firms post-reform chose to exchange board members to a similar degree as the treated firms. Hence, besides affecting the share of women directors and board members' average educational level, the Norwegian gender-balancing quota did otherwise not affect board structure. *Board size* is only affected when we use post 2004 years as the treatment years.

With regard to *firm performance*, we find clear evidence that treated firms' performance is adversely affected. When we choose post 2003 years as the treatment years, we find a significant negative effect on accounting-based performance as measured by both return on assets and operating income divided by assets. The coefficients for market-based performance (Tobin's Q and MTBR) are negative as well. However, they only turn significant when we use post 2004 years as the treatment years.[4] Hence, we find evidence for a performance-reducing effect of the Norwegian gender-balancing quota, especially for accounting-based performance measures.

Lastly and with regard to *firm risk*, we find evidence that the Norwegian gender-balancing quota reduces firm risk. The risk-reducing effect refers to systematic as well as to idiosyncratic risk. However, with respect to systematic risk, the effect is less clear: Whereas the beta is significantly negatively affected irrespective of whether we use post 2003 or post 2004 as the treatment years, the coefficient for our additional proxy for systematic risk is only statistically significant in those regressions where we use post 2004 as the treatment years.

---

[4] Even after adjusting for potential value effects on corporate debt as in Riepe and Yang (2019), the coefficient remains statistically significant and negative, although it becomes smaller.

Implications for practice

With respect to practical implications, our results clearly show that the Norwegian gender-balancing quota was *extremely effective* in achieving its goal to increase the share of women directors and increase gender equality in the appointment of non-executive directors. Because our control group of countries (Sweden, Denmark and Finland) also envisaged (non-binding) initiatives that aimed at increasing female representation on corporate boards, our results thus convincingly highlight the effectiveness of hard versus soft law. That is, when politicians seek to increase gender equality in the boardroom, formulating non-binding targets for a more balanced gender representation as in Denmark and Sweden, will not be enough. Rather, to be effective, gender quotas need to be binding, and non-compliance has to be penalized.

Further, our results show that the *business case* for a more gender-balanced representation on corporate boards is not as easy to argue for as the proponents of more gender-balanced representation often assert. Accounting for endogeneity, we find firm performance decreases as a result of the Norwegian gender-balancing quota. That is, simply increasing the share of women directors on the board will not automatically lead to better firm performance. However, our results also do not undoubtedly hint that a more gender-balanced board performs worse. Rather, we find evidence for a more gender-balanced board to perform *differently* than a less gender-balanced board: Firms that are affected by the Norwegian gender-balancing quota score lower in terms of accounting-based performance (and, depending on the treatment year, also in terms of market-based performance), and they are characterized by less risk — which might positively affect firms' long-term success and survival.

Hence, our results suggest a *more differentiated* view on women's representation in the boardroom. While the Norwegian gender-balancing quota was extremely successful in increasing women's representation in the boardroom and thus fostering gender equality as an important societal goal, its economic effects are not so clear-cut and ambivalent: (accounting-based) performance went down following the quota, as did firm risk. Concerning the latter effect, it is unclear whether reduced firm risk is beneficial: whereas reduced risk will be positive from the perspective of various stakeholders (e.g., employees and debt holders), equity investors might consider rebalancing their investment portfolio to return to their (optimal) target risk level. Hence, from an economic perspective, it is not clear how the causal effects of the Norwegian gender-balancing quota should be evaluated. Further, it is not clear whether the effects are only visible in the short term or also in the long term.

Likewise, and concerning practical implications for *firms*, it is unclear whether firms that are not covered by any quota regulation should strive for more gender equality in the boardroom in their own vested interest. Rather, this will depend on a firm's comparative evaluation of performance and risk effects. Further, it is unclear whether the performance and risk effects that we measured in the context of the Norwegian gender-balancing quota are generalizable to a situation where firms, absent any quota regulation, *choose* to have a more gender-balanced board. Our specific study context has the advantage that endogeneity issues can be addressed, however, its specificity also restricts generalizability. Because the Norwegian gender-balancing quota simultaneously forced many firms to recruit a considerable number of women directors in a comparatively short time frame, there was an unprecedented boost in the demand for women who were considered qualified for a board directorship, and also ready and prepared to take such a position. This very specific situation may be responsible for the measured performance and risks effects, and the effects might not be measurable in a situation where – at the other extreme – only one single firm decides to appoint an additional woman to a board position that was formerly held by a man.

In any case, whether or not being affected by a quota (or the risk thereof), firms would seem to be well advised to invest into a sufficient pool of female talent and to actively search for qualified women who bring additional expertise to the boardroom. These activities will be of utmost importance in industries with a currently small pool of female candidates.

Implications for theory

With respect to theory, our results do not support the view that a more gender-balanced board will generally lead to a better corporate performance, as argued by the information and decision-making approach (e.g., Gruenfeld, Mannix, Williams, & Neale, 1996). Likewise, we do not find that a more gender-balanced board in general performs worse than an all-male board, which could have been rationalized by, for instance, the similarity attraction paradigm (Tajfel, 1974, 1981; Turner, 1975, 1987) or social categorization theory (Byrne, 1971).

Rather, our results hint that a more gender-balanced board performs *differently* due to distinct priorities and the pursuit of strategically different choices. Our finding that gender-diverse boards make different choices concerning a potential tradeoff between (short-term) performance and risk speaks to, among others, resource dependence theory (see e.g. Pfeffer, 1972; Pfeffer & Salancik, 1978) which suggests that more gender-diverse boards will benefit from broader perspectives, expertise, and networks. As a result of the different perspectives, expertise, and networks, a gender diverse board might well take *different* strategic decisions than an all-male board.

Most of all, our results highlight the need to develop a more differentiated and comprehensive theory that incorporates both the potential performance and the risk effects of a more gender-balanced boardroom representation. Moreover, future research should strive to provide a better understanding of the mechanisms behind those effects. We hope that by having cleanly identified the causal effects of a more gender-balanced board and by having assessed potential performance and potential risk effects, we inspire future theory development in this direction.

Limitations and suggestions for further research

Our study is limited in several respects, which we hope will be addressed in future research. First, we undertook our study in a very specific context: the Norwegian gender-balancing quota. This context helps us to address endogeneity problems (the main motivation of our study), however, it is unclear whether our results can be generalizable (a) to other quota regulations in other country contexts and (b) to a situation where a firm, absent any quota regulation, chooses to exchange a male director for a female director. As already stated above, the Norwegian gender-balancing quota is rather specific in that it simultaneously forced many firms to recruit a considerable number of women directors in a comparatively short time frame. We cannot exclude that this very specific situation is responsible for the measured performance and risk effects. Hence, future research should challenge our results and seek to replicate them in other contexts.

Second and as a result of explicitly testing for the common trends assumption (also post-reform), our analysis is rather short-term and does not enable us to make inferences concerning long-term effects. At present we do not know whether the measured effects on corporate performance and risk will hold true for the long run. Rather, once firms have adjusted to the new "regime" by systematically investing in a pool of female talent and ensuring that the supply of female talent does not fall short of the increased demand, women directors may not be distinguishable from male directors in terms of the expertise and networks they bring to the boardroom, and a board's gender composition might no longer affect firm performance or risk. Future studies should try to assess whether or not long-term effects are different from short-term effects.

In any case, we hope that our study inspires future research to focus

on endogeneity problems when studying the effects of gender diversity (at the board level and also in other types of teams). Likewise, future meta-analyses and reviews should consider the sample of studies in the context of empirical identification. In light of inconclusive empirical findings and contradictory theoretical predictions, empirical identification is particularly critical.

Finally, our results call for more research on the relevance of women directors for firm risk. Notwithstanding the relevance of the business case that focuses on (short-term) accounting and market-based performance, a much stronger case for female representation is rooted in (potentially stable) differences in preferences. Findings that mandatory female representation on boards decreases firm systematic and idiosyncratic risk hopefully inspires future work in that direction.

Conclusion

The present paper analyzes the causal effects of the Norwegian gender-balancing quota. We find the quota is extremely effective in increasing the share of women directors on the boards of treated firms. With respect to the effects on firm performance and risk, we find the quota adversely affects treated firms' performance and reduces treated firms' risk.
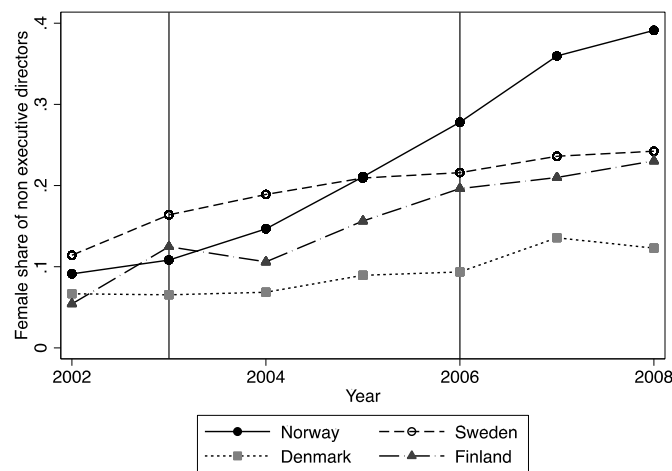
Appendix A

While fostering women's representation in the boardroom for equal opportunity reasons is beyond dispute, the evaluation of the quota's economic effects is rather ambiguous and less clear than the advocates and adversaries of gender-balancing quotas typically argue. The lack of rigor in previous research, especially concerning identification, and the resulting multitude of results allowed proponents as well as adversaries of a more gender-balanced board to push their ideological agendas. We hope to be able to contribute to a more objective and also more comprehensive discussion of the effects of a gender-balancing quota by carefully considering the causal identification of effects and by not only regarding potential performance, but also potential risk effects.

Acknowledgements

Fig. A1. Female share for treated and control groups by country.

Table A1

Firm accounting and market performance without controls.

Firm performance	OI/asset	ROA	Tobin	MTBR
Main explanatory variable				
DID (Norway × post 2003)	−0.037***	−0.027**	−0.209	−0.009
	(0.012)	(0.012)	(0.187)	(0.006)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	109	109	109	109
Observations	662	662	662	662
R-squared	0.200	0.187	0.151	0.089

Notes: This table summarizes the results from firm and industry-by-year fixed effects regressions of firm performance measured by operating profit/asset, return/asset, Tobin's Q, and market to book value, and explained by a difference-in-difference estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and post 2003 years. The estimate for MTBR very similar when applying a fractional logit model. Standard errors are clustered by firms and reported in parentheses.

  *** $p < 0.01$.

  ** $p < 0.05$.

Table A2
Firm risk without controls.

Firm risk	Equivola	Beta	Systematic risk	Idiosyncratic risk
Main explanatory variable				
DID (Norway × post 2003)	−0.125***	−0.156*	−0.019	−0.113***
	(0.041)	(0.090)	(0.019)	(0.043)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	87	87	87	87
Observations	2124	2124	2124	2124
R-squared	0.457	0.318	0.536	0.426

Notes: This table summarizes the results from firm and industry-by-year fixed effects regressions of firm market risk measured by equity volatility, market beta, systematic risk, and idiosyncratic risk, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and post 2003 years. Standard errors are clustered by firms and reported in parentheses.
  *** $p < 0.01$.
  * $p < 0.1$.

Table A3
Changes in board structure.

Non-executive directors	Share of female directors	Board size	Average tenure	Average age	National. mix	Average educ.
Main explanatory variable						
DID (Norway × post 2004)	0.108***	0.535*	0.247	−0.865	0.031	0.158**
	(0.032)	(0.300)	(0.350)	(0.674)	(0.037)	(0.071)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of firms	109	109	109	109	109	109
Observations	662	662	662	662	662	662
R-squared	0.486	0.111	0.108	0.144	0.127	0.279

Notes: This table summarizes the results from firm and industry-by-year fixed effects regressions of firm share of female non-executive directors and board size, and non-executive directors' average tenure, average age, nationality mix, and average amount of educational above-bachelor degrees, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2004. Standard errors are clustered by firms and reported in parentheses.
  *** $p < 0.01$.
  ** $p < 0.05$.
  * $p < 0.1$.

Table A4
Firm accounting and market performance.

Firm performance	OI/assets	ROA	Tobin's Q	MTBR
Main explanatory variable				
DID (Norway × post 2004)	−0.038***	−0.018	−0.346*	−0.013*
	(0.013)	(0.012)	(0.186)	(0.008)
Control variables				
Board size	−0.004	−0.006*	0.018	−0.001
	(0.003)	(0.003)	(0.025)	(0.001)
Tenure	0.003	0.002	−0.033	0.002
	(0.003)	(0.003)	(0.024)	(0.002)
Experience quoted	−0.009	−0.016*	−0.090	0.010
	(0.006)	(0.008)	(0.076)	(0.011)
Experience private	0.001	0.000	0.029	−0.004
	(0.002)	(0.003)	(0.036)	(0.004)
Age	−0.000	0.002	0.021	−0.001
	(0.002)	(0.002)	(0.015)	(0.001)
Nationality	−0.002	−0.006	0.320	0.013
	(0.031)	(0.031)	(0.477)	(0.011)
Education	0.006	0.002	0.083	−0.004
	(0.016)	(0.019)	(0.193)	(0.007)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	109	109	109	109
Observations	662	662	662	662
R-squared	0.215	0.207	0.158	0.112

Notes: This table summarizes the results from a firm and industry-by-year fixed effects regressions of firm performance measured in terms of operating profit/asset, return/asset, Tobin's Q, and MTBR, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2004. The estimate for MTBR is almost the same when applying a fractional logit. Standard errors are clustered by firms and reported in parentheses.
  *** $p < 0.01$.
  * $p < 0.1$.

Table A5
Firm risk.

Firm risk	Equivola	Beta	Systematic risk	Idiosyncratic risk
Main explanatory variable				
DID (Norway × post 2004)	−0.133***	−0.198**	−0.036*	−0.114**
	(0.044)	(0.090)	(0.020)	(0.045)
Control variables				
Board size	−0.002	0.003	−0.001	−0.001
	(0.005)	(0.012)	(0.002)	(0.004)
Tenure	−0.003	0.008	0.001	−0.003
	(0.006)	(0.014)	(0.003)	(0.005)
Experience quoted	0.034**	0.005	0.000	0.032**
	(0.015)	(0.037)	(0.007)	(0.014)
Experience private	−0.004	−0.002	−0.001	−0.003
	(0.006)	(0.012)	(0.003)	(0.006)
Age	−0.001	0.017**	0.004**	−0.002
	(0.004)	(0.007)	(0.001)	(0.004)
Nationality	0.023	0.026	0.005	0.027
	(0.057)	(0.169)	(0.038)	(0.053)
Education	−0.026	0.169**	0.032**	−0.036
	(0.034)	(0.077)	(0.016)	(0.032)
Firm & industry-by-year FE	Yes	Yes	Yes	Yes
Number of firms	87	87	87	87
Observations	2124	2124	2124	2124
R-squared	0.477	0.341	0.554	0.446

Notes: This table summarizes the results from firm and industry-by-year fixed effects regressions of firm market risk measured by equity volatility, market beta, systematic risk, and idiosyncratic risk, and explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2004. Standard errors are clustered by firms and reported in parentheses.

  *** $p < 0.01$.
  ** $p < 0.05$.
  * $p < 0.1$.

Table A6
Local average treatment effects (LATE).

	OI/A	ROA	Tobin	MTBR	Equivola	Beta	Systematic risk	Idiosyncratic risk
LATE (with Finland)	−0.351	−0.166	−3.203	−0.120	−1.231	−1.833	−0.333	−1.056
LATE (without Finland)	−0.361	−0.175	−3.222	−0.130	−1.194	−1.537	−0.269	−1.028

Notes: This table summarizes the results of the local average treatment effects from firm and industry-by-year fixed effects regressions of firm performance and risk, and is explained by a difference-in-differences estimate that accounts for the interacted effect of treatment status for the gender quota law in Norway and years after 2004. Because of potential concerns on simultaneous changes in one of the control countries, Finland, we compare results with and without the control firms from Finland.

References

Adams, R. (2016). Women on boards: The superheroes of tomorrow? *The Leadership Quarterly, 27*(3), 371–386. https://doi.org/10.1016/j.leaqua.2015.11.001.

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. https://doi.org/10.1016/j.jfineco.2008.10.007.

Adams, R. B., & Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science, 58*(2), 219–235. https://doi.org/10.1287/mnsc.1110.1452.

Adams, R. B., Hermalin, B. E., & Weisbach, M. S. (2010). The role of boards of directors in corporate governance: A conceptual framework and survey. *Journal of Economic Literature, 48*(1), 58–107. https://doi.org/10.1257/jel.48.1.58.

Adams, R. B., & Ragunathan, V. (2017). Lehman sisters. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3046451 Advance online publication.

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. https://doi.org/10.1093/qje/qjr049.

Aiyar, S. (2012). From financial crisis to great recession: The role of globalized banks. *American Economic Review, 102*(3), 225–230. https://doi.org/10.1257/aer.102.3.225.

Akerlof, G. A., & Shiller, R. J. (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism.* Princeton: Princeton University Press.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.

Angrist, J. D., & Krueger, A. B. (2007). Empirical strategies in labor economics. In O. C. Ashenfelter, R. Layard, & D. E. Card (Vol. Eds.), *Handbook of labor economics. Vol. 3. Handbook of labor economics* (pp. 1277–1366). Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4463(99)03004-7.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton: Princeton Univ. Press. Retrieved from www.gbv.de/dms/zbw/57188833X.pdf.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3.

Antonakis, J. D. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly, 28*(1), 5–21. https://doi.org/10.1016/j.leaqua.2017.01.006.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086–1120. https://doi.org/10.1016/j.leaqua.2010.10.010.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496–515. https://doi.org/10.5465/amr.1989.4308374.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives, 21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.). *Handbook of the economics of finance* (pp. 1053–1128). Amsterdam: Elsevier. https://doi.org/10.1016/S1574-0102(03)01027-6.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71. https://doi.org/10.2307/2490171.

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics, 127*(3), 588–612. https://doi.org/10.1016/j.jfineco.2017.12.009.

Bertrand, M., Black, S. E., Jensen, S., & Lleras-Muney, A. (2018). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies, 48*, 58. https://doi.org/10.1093/restud/rdy032.

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.

https://doi.org/10.1162/003355304772839588.

Bøhren, Ø., & Staubo, S. (2014). Does mandatory gender balance work? Changing organizational form to avoid board upheaval. *Journal of Corporate Finance, 28*, 152–168. https://doi.org/10.1016/j.jcorpfin.2013.12.005.

Brinkhuis, E., & Scholtens, B. (2018). Investor response to appointment of female CEOs and CFOs. *The Leadership Quarterly, 29*(3), 423–441. https://doi.org/10.1016/j.leaqua.2017.08.002.

Brown, J., & Matsa, D. A. (2016). Boarding a sinking ship? An investigation of job applications to distressed firms. *The Journal of Finance, 71*(2), 507–550. https://doi.org/10.1111/jofi.12367.

Brunnermeier, M. K., & Oehmke, M. (2013). Bubbles, financial crises, and systemic risk. In G. Constantinides, M. Harris, & R. M. Stulz (Vol. Eds.), *Handbook of the economics of finance. Vol. 2. Handbook of the economics of finance* (pp. 1221–1288). Amsterdam, London: North Holland. https://doi.org/10.1016/B978-0-44-459406-8.00018-4.

Byrne, D. (1971). *The attraction paradigm.* New York: Academic Press.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367–383. https://doi.org/10.1037/0033-2909.125.3.367.

de Cabo, R. M., Terjesen, S., Escot, L., & Gimeno, R. (2019). Do 'soft law' board gender quotas work? Evidence from a natural experiment. *European Management Journal.* https://doi.org/10.1016/j.emj.2019.01.004.

Cameron, A. C., & Trivedi, P. K. (2008). *Microeconometrics: Methods and applications* (7. print)Cambridge: Cambridge University Press.

Chaisemartin, C.d., & D'HaultfŒuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies, 85*(2), 999–1028. https://doi.org/10.1093/restud/rdx049.

Chen, S., Ni, X., & Tong, J. Y. (2016). Gender diversity in the boardroom and risk management: A case of R&D investment. *Journal of Business Ethics, 136*(3), 599–621. https://doi.org/10.1007/s10551-014-2528-6.

Combs, J. G., Crook, T. R., & Shook, C. L. (2005). The dimensionality of organizational performance and its implications for strategic management research. In D. J. Ketchen, & D. D. Bergh (Vol. Eds.), *Research methodology in strategy and management. Vol. 2. Research methodology in strategy and management* (pp. 259–286). Amsterdam, Oxford: Elsevier JAI. https://doi.org/10.1016/S1479-8387(05)02011-4.

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics, 142*(2), 636–654. https://doi.org/10.1016/j.jeconom.2007.05.002.

Cornell, B., & Shapiro, A. C. (1987). Corporate stakeholders and corporate finance. *Financial Management, 16*(1), 5. https://doi.org/10.2307/3665543.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature, 47*(2), 448–474. https://doi.org/10.1257/jel.47.2.448.

Dechow, P. M. (1994). Accounting earnings and cash flows as measures of firm performance. *Journal of Accounting and Economics, 18*(1), 3–42. https://doi.org/10.1016/0165-4101(94)90016-7.

Dechow, P. M., & Sloan, R. G. (1997). Returns to contrarian investment strategies: Tests of naive expectations hypotheses. *Journal of Financial Economics, 43*(1), 3–27. https://doi.org/10.1016/S0304-405X(96)00887-2.

Durand, R., & Vaara, E. (2009). Causation, counterfactuals, and competitive advantage. *Strategic Management Journal, 30*(12), 1245–1264. https://doi.org/10.1002/smj.793.

Eagly, A. H. (2016). When passionate advocates meet research on diversity, does the honest broker stand a chance? *Journal of Social Issues, 72*(1), 199–222. https://doi.org/10.1111/josi.12163.

Eagly, A. H., & Antonakis, J. (2015). Leadership. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology* (pp. 571–592). Washington: American Psychological Association. https://doi.org/10.1037/14341-018.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598. https://psycnet.apa.org/doi/10.1037/0033-295X.109.3.573.

Eckbo, B. E., Nygaard, K., & Thorburn, K. S. (2018). Board gender-balancing and firm value. *European Corporate Governance Institute (ECGI) — Finance working paper no. 463/2016*https://doi.org/10.2139/ssrn.2746786.

Eckel, C. C., & Grossman, P. J. (2006). Men, women and risk aversion: Experimental evidence. In C. Plott, & V. L. Smith (Vol. Eds.), *Handbook of experimental economics. Vol. 1. Handbook of experimental economics* (pp. 1061–1073). Elsevier: North Holland. https://doi.org/10.1016/S1574-0722(07)00113-8.

Faccio, M., Marchica, M.-T., & Mura, R. (2016). CEO gender, corporate risk-taking, and the efficiency of capital allocation. *Journal of Corporate Finance, 39*, 193–209. https://doi.org/10.1016/j.jcorpfin.2016.02.008.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics, 60*(2–3), 187–243. https://doi.org/10.1016/S0304-405X(01)00044-7.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics, 40*(1–3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes, 67*, 1–15. https://doi.org/10.1006/obhd.1996.0061.

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology, 30*(1), 507–544. https://doi.org/10.1146/annurev.soc.30.012703.110629.

Hallikas, J., Karvonen, I., Pulkkinen, U., Virolainen, V.-M., & Tuominen, M. (2004). Risk management processes in supplier networks. *International Journal of Production Economics, 90*(1), 47–58. https://doi.org/10.1016/j.ijpe.2004.02.007.

Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management*

Review, 32*(2), 334–343. https://doi.org/10.5465/amr.2007.24345254.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review, 9*(2), 193–206. https://doi.org/10.5465/amr.1984.4277628.

Haslam, S. A., Ryan, M. K., Kulich, C., Trojanowski, G., & Atkins, C. (2009). Investing with prejudice: The relationship between women's presence on company boards and objective and subjective measures of company performance. *British Journal of Management, 20*, 1. https://doi.org/10.1111/j.1467-8551.2009.00670.x.

Heckman, J. (1979). Sample bias as a specification error. *Econometrica, 47*(1), 153–162. https://doi.org/10.3386/w0172.

Hermalin, B., & Weisbach, M. (1998). Endogenously chosen boards of directors and their monitoring of the CEO. *The American Economic Review, 88*(1), 96–118.

Hermalin, B., & Weisbach, M. (2003). Boards of directors as an endogenously determined institution: A survey of the economic literature. *FRB New York — Economic policy review* (pp. 7–26). . https://doi.org/10.3386/w8161.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86. https://doi.org/10.1257/jel.47.1.5.

Ince, O. S., & Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care!. *Journal of Financial Research, 29*(4), 463–479. https://doi.org/10.1111/j.1475-6803.2006.00189.x.

Ioannidis, J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly, 94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210.

Jensen, T. L., & Johannesen, N. (2017). The consumption effects of the 2007–2008 financial crisis: Evidence from households in Denmark. *American Economic Review, 107*(11), 3386–3414. https://doi.org/10.1257/aer.20151497.

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. https://doi.org/10.1016/0304-405X(76)90026-X.

Jeong, S.-H., & Harrison, D. A. (2017). Glass breaking, strategy making, and value creating: Meta-analytic outcomes of women as CEOs and TMT members. *Academy of Management Journal, 60*(4), 1219–1252. https://doi.org/10.5465/amj.2014.0716.

Joecks, J., Pull, K., & Vetter, K. (2013). Gender diversity in the boardroom: What exactly constitutes a critical mass? *Journal of Business Ethics, 118*(1), 61–72.

Kenny, D. A. (1979). *Correlation and causality. A Wiley-Interscience Publication.* New York: Wiley.

Keynes, J. M. (1964). *The general theory of employment, interest, and money.* New York: Harcourt, Brace & World, Inc (originally 1936).

Kirsch, A. (2018). The gender composition of corporate boards: A review and research agenda. *The Leadership Quarterly, 29*(2), 346–364. https://doi.org/10.1016/j.leaqua.2017.06.001.

Kronmal, R. A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 156*(3), 379. https://doi.org/10.2307/2983064.

Lechner, M. (2010). The estimation of causal effects by difference-in-difference methods: Estimation of spatial panels. *Foundations and Trends in Econometrics, 4*(3), 165–224. https://doi.org/10.1561/0800000014.

Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics, 71*(2), 349–380. https://doi.org/10.1016/S0304-405X(03)00206-X.

Leuven, E., & Barbara, S. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching. *Statistical software components S432001.* Boston College Department of Economics.

Main, B., Gonzalez, A., & Sila, V. (2018). Gender equality and the spillover effects of women on boards. *Academy of Management Proceedings*10147. https://doi.org/10.5465/AMBPP.2018.10147abstract.

Martin, K. J., & McConnell, J. J. (1991). Corporate performance, corporate takeovers, and management turnover. *The Journal of Finance, 46*(2), 671. https://doi.org/10.2307/2328841.

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. https://doi.org/10.1257/app.5.3.136.

Nygaard, K. (2011). Forced board changes: Evidence from Norway. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.1793227 Advance online publication.

Opler, T. C. I. M., & Titman, S. (1994). Financial distress and corporate performance. *The Journal of Finance, 49*(3), 1015–1040. https://doi.org/10.1111/j.1540-6261.1994.tb00086.x.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*(6), 619–632. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1.

Perryman, A. A., Fernando, G. D., & Tripathy, A. (2016). Do gender differences persist? An examination of gender diversity on firm performance, risk, and executive compensation. *Journal of Business Research, 69*(2), 579–586. https://doi.org/10.1016/j.jbusres.2015.05.013.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22*(1), 435–480. https://doi.org/10.1093/rfs/hhn053.

Pfeffer, J. (1972). Merger as a response to organizational interdependence. *Administrative Science Quarterly, 17*, 382–394.

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

Pletzer, J. L., Nikolova, R., Kedzior, K. K., & Voelpel, S. C. (2015). Does gender matter? Female representation on corporate boards and firm financial performance: A meta-analysis. *PLoS One, 10*(6), e0130005. https://doi.org/10.1371/journal.pone.

0130005.

Post, C., & Byron, K. (2015). Women on boards and firm financial performance: A meta-analysis. *Academy of Management Journal, 58*(5), 1546–1571. https://doi.org/10.5465/amj.2013.0319.

Riepe, J., & Yang, P. (2019). *Empirical studies on gender diverse boards: Be aware of the value bias in corporate debt. Working paper.* Tuebingen University.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics, 2*(3), 808–840. https://doi.org/10.1214/08-AOAS187.

Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences, 106*(36), 15268–15273. https://doi.org/10.1073/pnas.0907352106.

Sila, V., Gonzalez, A., & Hagendorff, J. (2016). Women on board: Does boardroom gender diversity affect firm risk? *Journal of Corporate Finance, 36*, 26–53. https://doi.org/10.1016/j.jcorpfin.2015.10.003.

Stulz, R. M. (1994). Tobin's Q, corporate diversification, and firm performance. *Journal of Political Economy, 102*(6), 1248–1280. https://doi.org/10.1086/261970.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information, 13*, 65–93. https://doi.org/10.1177/2F053901847401300204.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology.* Cambridge: Cambridge University Press.

Terjesen, S., Aguilera, R. V., & Lorenz, R. (2015). Legislating a woman's seat on the board: Institutional factors driving gender quotas for boards of directors. *Journal of Business Ethics, 128*(2), 233–251. https://doi.org/10.1007/s10551-014-2083-1.

Terjesen, S., & Sealy, R. (2016). Board gender quotas: Exploring ethical tensions from a multi-theoretical perspective. *Business Ethics Quarterly, 26*(1), 23–65. https://doi.org/10.1017/beq.2016.7.

Tobin, J. (1969). A general equilibrium approach to monetary theory. *Journal of Money, Credit and Banking, 1*(1), 15–29. https://doi.org/10.2307/1991374.

Turner, J. C. (1975). Social comparison and social identity: Some comparisons for intergroup behavior. *European Journal of Social Psychology, 5*, 1–34. https://doi.org/10.1002/ejsp.2420050102.

Turner, J. C. (1987). *Rediscovering the social group: A social categorization theory.* Oxford: Blackwell.

Wernerfelt, B., & Montgomery, C. A. (1988). Tobin's Q and the importance of focus in firm performance. *The American Economic Review, 78*, 246–250. https://www.jstor.org/stable/1814713.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: MIT Press.