LUND UNIVERSITY
School of Economics and Management

The Business Case for Board Gender Diversity

Evidence from Swedish Firms

by

Eric Engerby Jönsson and Jesper Salomonson

June 2020

Master's Programme in International Strategic Management

Supervisor: Matts Kärreman

[This page is intentionally left blank]

Abstract

Gender diversity has gained a lot of attention in the corporate governance literature, reaching conflicting results whether board gender diversity has an impact on financial performance. As a consequence, we set out to investigate how board gender diversity affects financial performance through innovation using a mediation technique. We constructed three hypotheses grounded in the resource dependency theory and the agency theory arguing that board gender diversity will have an impact on a board's capability to execute the two main functions of a board; resource provision and monitoring.

Our study was applied in a Swedish context by studying 118 companies listed on OMXS over the years of 2008 to 2014. This implied a balanced micro panel data set, which enabled us to mitigate the risks of omitted variable bias and reversed causality. With the intention to investigate the mediation relationship, we used the Blau and the Shannon indices as our measures of board gender diversity, R&D expenses for innovation, and Tobin's Q and ROA for financial performance. Our findings indicated that there were no significant relationships between board gender diversity and innovation or financial performance, or innovation and financial performance, that were robust for reversed causality. Therefore, we had to reject all hypotheses and could not sustain a mediation relationship.

From a theoretical viewpoint, we concluded that RDT is too simplistic as it neglects how potential upsides of board gender diversity should be utilized while AT lacks sensitivity to account for contextual factors. For practitioners, practical tools are needed in order to benefit from board gender diversity. From a wider perspective, we concluded that the lack of significant results could be due to that gender diversity is irrelevant for a board's performance or that the board in itself lacks an influence on firm outcomes, making studies on demographic characteristics redundant since there is no relationship to discover. Additionally, while the business case of board gender diversity could not be supported, it may still exist positive effects of board gender diversity from other perspectives. This made us conclude that increased board gender diversity provides non-negative outcomes and thus, is desirable on a societal level.

Keywords: Board gender diversity, Innovation, Financial performance, Mediation, Panel data

Acknowledgement

We would like to show our deepest gratitude to our supervisor Matts Kärreman for showing dedication and interest in both the quality of our paper and the well-being of us. In addition to providing valuable insights on our work, we are grateful that he emphasized the importance of enjoying the writing process. Last, we hope that Matts has enjoyed our supervision meetings as much as we have done.

Eric Engerby Jönsson

Jesper Salomonson

Lund, 1st of June 2020

Table of contents

List of figures

List of tables

1 Introduction

1.1 Background

In 2014, there were more male directors named John, William, Robert or James than female directors occupying positions on the boards of the S&P 1500 companies (McGregor, 2015). In Sweden, a country which has come further in terms of gender equality, only 25.8 percent of the board seats were held by females at the same time (Larsson, Fristedt & Sundqvist, 2014). While the inclusion of females can be seen from a variety of viewpoints such as economic, societal, moral and cultural, there is an ongoing debate between those who think that it is the right thing to do and those who argue that it actually benefits shareholder value (Carter, D'Souza, Simkins & Simpson, 2010; Brunzell & Liljeblom, 2014). A former CEO stated the following: "When I'm sitting in that boardroom, my fiduciary responsibility is to the shareholders of that company – not social engineering. I can talk about diversity. But there ought to be a business case." (Dvorak, 2008, cited in Carter et al., 2010).

Furthermore, in the light of new regulatory frameworks addressing the underrepresentation of females on corporate boards, gender composition has gained increasing attention in the corporate governance literature (Ararat, Aksu & Tansel Cetin, 2015; Carter et al., 2010; Chen, Leung & Evans, 2018; Gordini & Rancati, 2017). More specifically, the link between board gender diversity and financial performance has been subject to investigation where researchers are disagreeing and reaching conflicting results (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2010; Eklund, Palmberg & Wiberg, 2009; Gordini & Rancati, 2017; Randøy, Thomsen & Oxelheim, 2006; Rose, 2007; Smith, Smith & Verner, 2006). As a result, it is debated whether the inclusion of female directors have any true benefits at all or if they are merely used as tokens signaling a firm's compliance with social expectations and governance codes (Bear, Rahman & Post, 2010; Carter, Simkins & Simpson, 2003; Konrad, Kramer & Erkut, 2008; Lafuente & Vaillant, 2019). Furthermore, as previous research remains contradicting, there are currently two views on the literature regarding the linkage between board gender diversity and firm performance (Torchia, Calabrò, Gabaldon & Kanadli, 2018); first, some authors argue that the research is reaching maturity and that other areas of diversity deserves more attention (Hillman, 2015); second, other authors argue that the research must become more fine-tuned, investigating moderating or mediating variables in order to understand how board gender diversity affects firm performance (Eagly, 2016; Gabrielsson & Huse, 2004; Milliken & Martins, 1996; Post & Byron, 2015).

The contribution to the second viewpoint has focused on investigating mediating and/or moderating variables such as the board's monitoring efforts (Ararat, et al., 2015), innovation productivity (Chen et al., 2018), innovation and reputation (Miller & Triana, 2009) a board's decision-making culture (Nielsen & Huse, 2010; Torchia et al., 2018) and a board's working structure (Nielsen & Huse, 2010). However, this new string of research remains scarce and needs to be contributed to. Interestingly, two of these articles focus on innovation as a mediating variable (Chen et al., 2018; Miller & Triana, 2009) and a third investigates a board's decision-making culture as a mediating variable between board gender diversity and innovation (Torchia et al., 2018). The focus on innovation is unsurprising as innovation is a survival strategy for firms to reach their financial targets (Caves & Ghemawat, 1992; Teece, Pisano & Shuen, 1997; Zahra & Garvis, 2000). Additionally, when investigating mediating variables between board gender diversity and financial performance, it is crucial that there is a relationship between the three variables. Subsequently, we find that innovation is intuitively connected to financial performance as well as the functions of the board as the board is responsible for formulating strategies (Tricker, 2019), provide valuable resources needed to execute the agreed strategy (Pfeffer & Salancik, 1978) and monitor managers' adherence to the strategy (Jensen & Meckling, 1976). Thus, we posit that innovation is important for further investigation as a potential mediating variable between board gender diversity and financial performance.

As we want to contribute to the research on innovation as a mediating variable of board gender diversity and firm performance, the research papers of Miller and Triana (2009) and Chen et al. (2018) will act as important references in this study. While the main rationale for investigating innovation as a mediating variable is due to previous research being scarce, we also identify potential areas of improvement when reading the papers of Miller and Triana (2009) and Chen et al. (2018). First, Miller and Triana (2009) use Fortune 500 firms as their sampling frame, which means that the results have low generalizability to other populations than the 500 largest firms in the US. It has been demonstrated in previous research that innovation is closely related to firm size where larger firms often obtain better access to resources and are better able to survive failing innovation projects (Adams & Ferreira, 2009; Gordini & Rancati, 2017; Schilling, 2017). Despite that Chen et al. (2018) use a wider sampling frame with more companies, they still investigate a US context. Hence, we see the need for using a different set of sample data. Second, Miller and Triana (2009) employ a cross-sectional data set, making the results sensitive to year-specific effects and creating an inability to control for unobserved firm-

specific effects, therefore the results become less reliable. By instead using panel data, we will be able to control for both firm- and time-specific effects and hence generate more robust results. As Chen et al. (2018) build on the work by Miller and Triana (2009), they have improved in terms of having a larger sample of a variety of firm sizes as well as adopting a panel data set. However, we find it peculiar that Chen et al. (2018) investigate a time period between 1998 and 2006 for a paper published in 2018. Consequently, we investigate a later time period of 2008 to 2014. Furthermore, common for previous research on board gender diversity, innovation and financial performance, is its disparity regarding how to operationalize variables. This further motivates us to perform this study, to see if our operationalizations yield unanimous results.

1.2 Purpose of our study

Based on the aforementioned discussion, the purpose of our panel data study is to investigate how board gender diversity affects financial performance mediated by innovation for publicly listed companies in Sweden. We incorporate the resource dependency theory (RDT) and the agency theory (AT), and based on these theories and previous research we form three hypotheses to investigate a potential causal relationship. The rationale for applying these theories is that they reflect the two main functions of the board, namely resource provision and monitoring. Additionally, RDT suggests that unique and valuable resources brought to the firm can result in improved capability to innovate while AT suggests that enhanced monitoring will increase managers' willingness to innovate and reduce the risk of managerial shirking. Following this, our research question is:

Does board gender diversity affect financial performance through innovation?

1.3 Contribution of our study

Our main contribution is a deeper understanding of how board gender diversity affects financial performance from a strategic management point of view. More specifically, our study provides three main contributions. First, our study responds to the extensive calls for more fine-tuned research by adopting a mediation technique and for country-specific studies on board gender diversity and financial performance by studying Swedish companies. Second, practitioners will gain a deeper understanding of the possible effects of board gender diversity on innovation and financial performance that can be utilized in strategy formulation. Last, our results become

relevant on a societal level as it contributes to the debate whether board gender diversity is economically relevant by having a causal effect on financial performance.

1.4 Outline of the paper

The remaining part of the paper proceeds as follows: Chapter 2 describes RDT and AT as our chosen theories and presents our hypotheses carefully based in previous research. Chapter 3 describes the methodology in terms of our chosen research design, data collection, variables, data analysis, and validity and reliability. Chapter 4 presents the empirical findings and how they should be interpreted. Chapter 5 discusses the results and chapter 6 concludes the paper with managerial and theoretical implications, accompanying limitations and suggestions for future research. However, before moving on to chapter 2, we find it important to emphasize that it may sound like the inclusion of female directors is the key for improving the main functions of the board which facilitates innovation. However, this is not the case. There is no clear evidence that female directors are superior to male directors (e.g. Adams & Ferreira, 2009; Nielsen & Huse, 2010; Randøy et al., 2006), or vice versa, rather it is the unique and added resources and abilities that females bring into a board that enhances board functions. This implies that there is no compelling evidence that a board consisting only of men is better or worse than a board consisting only of women. The benefit of including the minority gender into boards (which in almost all cases mean women) arises as a result of the heterogeneity and not as a result due to that one gender is a better board member than the other when taking account for different degrees of board experience. However, even if that would be the case, this study is not focused on one gender's potential superiority over the other.

2 Literature review and hypotheses

When conducting our literature search, we searched for board gender diversity and financial performance in order to get an overview of the topic. Early on, we discovered that previous research is urging for a more fine-tuned string of research and as a consequence, we searched for papers investigating moderating and/or mediating variables. For the entirety of the search process, we focused on articles published in well-known strategic management journals with more than 500 citations but later complemented these with newer articles, as the citation requirement implied a biased selection of older papers due to the time aspect. Simultaneously, we conducted citation chaining in order to identify papers with a large impact on this research area. Furthermore, all of our searches were carried out using Google Scholar or LUBSearch. During the entire process, we kept in mind to have geographically dispersed studies with an emphasis on articles from a Swedish and Scandinavian context. However, due to a limited amount of papers from Scandinavia, we complemented these with non-Scandinavian articles, mainly from other parts of Europe and the US.

2.1 Motivation for our theoretical viewpoints

Much of previous research investigating board gender diversity and firm performance has neglected the dimensions of board activities that can explain this relationship (Nielsen & Huse, 2010; Post & Byron, 2015; Zahra & Pearce, 1989). As monitoring and involvement in strategy formulation are two main responsibilities for a board (Canella, Finkelstein & Hambrick, 2009; Post & Byron, 2015; Zahra & Pearce, 1989), we applied AT as it gives the theoretical foundation for the monitoring function of the board (Fama & Jensen, 1983), and RDT as the resources directors bring influence a board's ability to be involved in strategy formulation (Pfeffer & Salancik, 1978). Hillman and Dalziel (2003) state that AT and RDT constrain focus to one board function while neglecting the effect of the other. This has resulted in an incomplete description of how boards affect firm performance, which we aim to overcome by combining the two theories. Due to the wide application of these theories in previous research, we deem them to be applicable in accordance with our hypotheses. In contrast, this paper could have chosen a different set of theories, such as signalling theory and behavioral theory of the firm as Miller and Triana (2009) did. However, as these theories neglect the two main functions of the board, we deem them inappropriate for this particular study.

Previously, efforts have been made in understanding how the overall composition of a board affects firm activities and firm outcomes (Baysinger & Butler, 1985; Dalton, Daily, Ellstrand & Johnson, 1998; Hillman, Keim & Luce, 2001). Hillman and Dalziel (2003) introduced the concept of board capital, defined as the combination of human and social capital, in order to explain the board's role as a resource provider and a monitor of the firm. Board capital represents the ability of the board to perform in these activities (Hillman & Dalziel, 2003). Under board capital, human capital consists of expertise, experience, knowledge, reputation, and skills which directors possess (Hillman & Dalziel, 2003) and these will affect how directors frame decisions and what they pay attention to (Johnson, Schnatterly & Hill, 2013). Social capital on the other hand, is the network of relationships directors possess to the external environment (Hillman & Dalziel, 2003). The following discussion is therefore based on how board capital affects board activities and firm outcomes.

2.1.1 A resource dependency theory perspective

Pfeffer and Salancik (1978) proposed the idea of directors bringing critical resources to the firm, seen as an open system relying on connections to the external environment. From an RDT perspective, the objective for a firm is to minimize its environmental uncertainty and dependence which is achieved by controlling resources such that the power of external actors over the firm is reduced. According to Pfeffer and Salancik (1978), a firm has five different tools to reduce the environmental uncertainty and dependence it faces, and the board is one of these tools. In accordance with this, a board provides four primary benefits; advice and counsel, legitimacy, communication channels for information between external organizations and the firm, and valuable access of support from external actors (Pfeffer & Salancik, 1978). These benefits, also denoted as provision of resources by Hillman and Dalziel (2003), are relevant to consider as they are positively associated with board capital. First, as boards mainly consist of members from previous decision-making roles higher up in organizations, they possess experience, expertise and skills that facilitate advice and counsel (Baysinger & Butler, 1985; Hillman, Cannella & Paetzold, 2000) just like directors with social ties to strategically related firms will provide better advice and counsel (Carpenter & Westphal, 2001). Second, by including renowned directors, the prestige of the board increases and subsequently the credibility and legitimacy of the firm (Certo, Daily & Dalton, 2001). At the same time, directors' social relationships also affect how the external perceives the firm and thus affects its legitimacy (Hillman & Dalziel, 2003; Hillman et al., 2000). This is closely related to the idea of tokenism, where a board member can provide legitimacy just by being present (Huse, 2007;

Konrad et al., 2008; Torchia, Calabrò & Huse, 2011). Third, empirical results indicate that boards with governmental connections lead to increased shareholder value as the connections ensured stable information flows and more open communication (Hillman & Dalziel, 2003; Hillman et al., 2000). Further, Hillman and Dalziel (2003) state that directors who sit on several boards are likely to reduce coordination costs and facilitate smooth information flows. Last, directorate ties help the firm to gain influence with critical stakeholder groups such that it can secure scarce resources on favorable terms (Hillman et al., 2000). These types of resource provisions that board capital enhances have been linked to firm performance in previous research (Hillman & Dalziel, 2003), indicating the significant value of board capital.

2.1.2 An agency theory perspective

From an AT-perspective, the ideas that human beings are subject to opportunistic behavior (self-interest seeking utility maximizers) and that information asymmetry characterizes the relationship between shareholders and managers make the separation of ownership and control in organizations problematic (Tricker, 2019). If managers prioritize their own interests at the expense of shareholder wealth, a conflict of interests and agency costs might occur (Jensen & Meckling, 1976). For example, unrelated diversification and R&D investment are normally two areas where managers and shareholders have conflicting interests that raise the need of governance mechanisms to ensure managers act on behalf of the shareholders (Shen & Gentry, 2014). This implies that the board function of monitoring managers is a fundamental activity for the protection of shareholders' interests and mitigation of the agency dilemma (Jensen & Meckling, 1976).

While it is evident that previous research has found clear linkages between board capital and the provision of resources (i.e. the RDT-perspective), there is a gap regarding the relationship between board capital and the monitoring function of the board (Hillman & Dalziel, 2003). This can be explained by how AT claims that incentives are directly affecting the monitoring function (Jensen & Meckling, 1976) which has led to the negligence of board capital's effect on monitoring. Even if incentives indeed are an important factor in a board's monitoring effectiveness, Jensen and Meckling (1976) stated that the monitoring ability depends on the skills of the board's members (i.e. the board capital). Hence, this implies that the monitoring ability is heterogenous between boards and that both board ability and incentives affect monitoring. More specifically, from a behavioral scientific perspective, Becker and Huselid (1992) mean that incentives will not alone determine a group's task performance, and Hunter

and Hunter (1984) found that ability is the most important factor in task performance. This means that the importance of incentives may need to be reconsidered for a board's monitoring role. Furthermore, regardless of how appealing an incentive may be, it will not lead to improved performance if the board member does not have the right skills or the desire to execute the activity to begin with (Hillman & Dalziel, 2003). Since incentives rather serve as a moderator between board capital and the board's two main functions (Hillman & Dalziel, 2003), we will focus on the effect of board capital. Based on the significant role of board capital, Carpenter and Westphal (2001) found that boards with experience in different situations and with specific expertise that improved its understanding of the inner workings of the firm, enhanced the monitoring effectiveness. Thus, board capital determines the ability of the board to be a successful monitor as well as securing provision of resources.

2.1.3 Board capital and demographic characteristics

While it is evident that the board capital has a significant effect on a board's two main functions, the factors affecting board capital need to be considered. Board size may be the most common proxy for board capital which depends on the statement by Pfeffer and Salancik (1978:172) saying that "the greater the need for effective external linkage, the larger the board should be". However, board size is a blunt instrument for board capital as it overlooks nuances in human and social capital (Haynes & Hillman, 2010). Since previous research denotes that board capital affects board functions, arguably, demographic characteristics are also affecting board capital as it explains variances in board members' human and social capital. The common rationale for applying demographic characteristics when analyzing the board is the underlying assumption that these characteristics affect directors' behaviors, cognition and decision making, and later, firm level outcomes (Forbes & Milliken, 1999). For example, Terjesen, Sealy and Singh (2009) state that gender differences will lead to differences in human capital of directors. Additionally, having a heterogeneous board implies a larger variety in the board's experience, knowledge, creativity and connections to the external environment which should lead to enhanced human and social capital (Haynes & Hillman, 2010). Consequently, the creation of a demographically heterogeneous board can be beneficial as directors with different age, gender and ethnicity arguably have different backgrounds, experiences and social connections (Golden & Zajac, 2001). This translates into a broader range of human and social capital as directors complement each other, leading to more strategic options being considered (Golden & Zajac, 2001). However, it is not self-evident that demographic heterogeneity will have a positive impact on board activities as it may be contingent on the context (Forbes & Milliken, 1999).

2.2 Board gender diversity's effect on board functions

Next to be examined is how board gender diversity affects the two board functions. Diversity can be defined as the variety of different statuses within a population (Blau, 1977). Following this, board diversity is defined as the variety inherent in a specific board's composition (Campbell & Mínguez-Vera, 2008; Gordini & Rancati, 2017). This variety can be measured in several dimensions where demographic characteristics are such a dimension. Gender is one demographic characteristic of directors affecting the strategic direction of the firm (Hillman, Cannella & Harris, 2002) and the monitoring ability of the board (Adams & Ferreira, 2009). We therefore define board gender diversity as the degree of heterogeneity among directors' biological gender, in line with previous research (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Gordini & Rancati, 2017; Miller & Triana, 2009; Post & Byron, 2015; Torchia et al., 2018).[1]

2.2.1 Board gender diversity's effect on resource provision

Rather than explaining directors as independent or dependent directors, as AT does, Hillman, et al. (2000) proposed a new taxonomy for directors based on the four primary benefits directors shall bring, making it applicable when discussing the role of directors as resource providers. The taxonomy is based on differences in occupational backgrounds and individual experiences as they indicate what type of resources each director possesses, such as information, skills, expertise and linkages to external actors. The four categories Hillman et al. (2000) identified were insiders, business experts, support specialists and community influentials. First, insiders are those who have served the company in a previous role such as a management position; second, business experts are those who have served in other firms and bring expertise from holding previous decision-making positions; third, support specialists are those who provide linkages to specifically important areas such as law or finance; and fourth, community influentials are those who possess important linkages to non-business organizations such as political parties, universities or other important institutions. In a subsequent study, it was shown that female directors are more likely than men to be support specialists and community influentials while men were more likely to be insiders and business experts (Hillman et al., 2002). The study also showed that female directors are more likely to hold advanced degrees and to have non-business backgrounds (Hillman et al., 2002). This implies that more unique

[1] While we are aware that gender in some contexts is not only seen as a biological question, but rather as a social construction with more aspects than purely a binary classification, this paper focuses on the biological separation of genders (i.e. the separation of males and females).

linkages can be exploited from a gender heterogenous board compared to a gender homogenous one, which is beneficial for two reasons. First, the inclusion of females can aid a firm's understanding of particular customer segments as they increase the board's collective knowledge about different customer segments (Carter et al., 2003; Gordini & Rancati, 2017; Smith et al., 2006). This follows the idea of Cox and Blake (1991), proposing that a heterogeneous team will have a superior understanding of diverse customer segments' needs and wants compared to a homogeneous team. As female directors have more experience from marketing and sales than their male counterparts (Groysberg & Bell, 2013), this can be combined with the improved understanding of particular customer segments to improve the customer fit of the firm's offerings. Similarly, Eklund et al. (2009) argue that heterogeneous boards can lead to better access to certain geographical markets, allowing firms to more easily expand their business. This can partly be explained by the fact that female directors, as more likely to be community influentials (Hillman et al., 2002), have more ties to regulation authorities which reduces transaction costs when expanding business to new markets and there is a need for negotiations with public institutions. Second, as a natural consequence of the heterogeneity in occupational backgrounds (Hillman et al., 2002), one of the most prominent arguments for having a gender diverse board is the additional information sources to tap (Peterson & Philpot, 2007; Randøy et al., 2006). This translates into a more exhaustive coverage of useful information (Granovetter, 1973) and hence the board being more proficient in problem-solving and decision-making (Gordini & Rancati, 2017). However, just as Forbes and Milliken (1999) stress, the effect of board gender diversity may be contingent on the context. This is especially relevant to consider with regard to the provision of resources; the increase in board capital arising from a gender diverse board is most applicable to countries with high gender parity since women's access to education, labor market and career development is more restricted in countries with large gender inequalities (Post & Byron, 2015; Tyrowicz, Terjesen & Mazurek, 2020). Consequently, the resource provision perspective of gender diversity may be most applicable for countries with high gender parity as women will be more included in society and the labor market.

2.2.2 Board gender diversity's effect on monitoring

The issue of how board gender diversity influences monitoring is then next to be examined. Eisenhardt (1989) means that a board acts as an information system for the shareholders to verify the behavior of managers, and the ability of a board to do this depends on its board capital (Hillman & Dalziel, 2003). As gender is a widely used proxy for differences in cognition

(Forbes & Milliken, 1999), it will have a profound impact on how well the board monitors and serves shareholders with an adequate information system. There are several indications that a more gender diverse board will affect the board's monitoring behavior. First, as Hillman et al. (2002) found an increase in the range of professional experience and advanced degrees by including more women onto the board, the effectiveness of monitoring is enhanced according to Carpenter and Westphal (2001). Second, Adams and Ferreira (2009) found that female directors are more likely to sit on audit and monitoring committees. In combination with women being more ethically oriented than men (Pan & Sparks, 2012) and more likely to ask critical questions or speak up when concerned about an issue (Huse & Solberg, 2006; Konrad et al., 2008), the inclusion of female directors will make the board a tougher monitor. Likewise, Post and Byron (2015) observed that female directors are positively related to increased focus on monitoring. Third, Adams and Ferreira (2009) further found that female directors have higher attendance than their male counterparts but also that the presence of women on the board increased the attendance by male directors. Moreover, female directors tend to be more prepared for board meetings (Huse & Solberg, 2006), perhaps partly due to higher pressure of proving their right to be on the board as chairmen are argued to be less satisfied with female directors than male ones according to Brunzell and Liljeblom (2014), which arguably can be seen as improved monitoring ability. However, previous research has stressed that differences in female and male director behaviors may be contingent on that female directors are a minority (Biernat & Kobrynowicz, 1997); for example, if male directors were a minority, it could be the case that they had a higher attendance behavior and generally were more prepared for board meetings as they felt a higher pressure to demonstrate their commitment and right to be on the board. Last, agency theorists mean that independent directors are more effective in monitoring managers as they face less conflicts of interests (Fama & Jensen, 1983) and since female directors are more likely to be independent directors (e.g. Adams & Ferreira, 2009; Carter et al., 2010; Simpson, Carter & D´Souza, 2010), this increased board independence should result in improved monitoring as well. Even though we have stated the increase in board capital from a gender diverse board as a significant determinant of monitoring, Carter et al. (2010) discuss whether ownership position may be a more powerful monitoring tool than board gender diversity. However, Adams and Ferreira (2009) concluded that board gender diversity indeed improved effectiveness of monitoring.

2.3 Resource provision and monitoring's effect on innovation

As gender diversity clearly has an effect on a board's ability to successfully provide resources and monitor the top management team, next to be examined is how these two functions of a board can affect innovation efforts. Defining innovation is a complex task with regard to the extensive definitions on innovation and their somewhat overlapping descriptions. We will therefore define innovation as a firm's commitment to create and commercialize new products or successfully implement new processes and organizational systems (Zahra & Garvis, 2000), characterized as an iterative process (Utterback, 1971). The degree of newness is key in defining innovation; Greenhalgh and Rogers (2010) argue that a true innovation must not only be new to the firm but also new to the relevant market, while others argue it to be sufficient if the innovation improves existing products, processes or systems (Nagji & Tuff, 2012; Utterback, 1971). Hence, we follow the latter standpoint in terms of level of newness.

2.3.1 Resource provision

As research posits that attitudes, cognitive functions and beliefs tend to vary with demographic variables (Robinson & Dechant, 1997), decisions on board level are not only based on past experiences but also demographic characteristics (Hambrick & Mason, 1984). Following this, researchers argue that demographic diversity (e.g. gender diversity) increases the number of ideas and promotes creativity which subsequently aids a firm's innovation process (Cox & Blake, 1991). In other words, the knowledge embodied in the human and social capital of a board (i.e. board capital), can become a competitive advantage for a firm in terms of significantly improving innovative efforts (Miller & Triana, 2009). Because of this, homogeneous boards may actually hamper innovation in terms of relying on few perspectives based on similar experience, while heterogeneous boards arguably boost innovation in terms of providing a broader range of ideas grounded in diverse experiences and perspectives. Similarly, Mintzberg, Raisinghani and Theoret (1976) propose that unique ideas and perspectives will have a positive impact on the identification, development and selection of firm decisions. This implies that the positive effect board gender diversity has on sources of information will benefit the firm from a larger range of ideas and perspectives which arguably have a positive effect on innovation. As the inclusion of females allows a board to understand particular customer segments (Carter et al., 2003; Cox & Blake, 1991; Gordini & Rancati, 2017; Smith et al., 2006), firms are better suited for innovating across and within different customer segments. Consequently, as Brown and Katz (2011) argue, insight and understanding of customer

behavior is an essential part in innovation work. Moreover, with detailed knowledge about customer behavior, boards can initiate more innovative ideas for the firm's strategy formulation.

While it is not only the human capital affecting innovation, social capital provides additional information that affects a firm's ability to innovate (Burt 1997; Granovetter, 1973). As females generally have to manage larger contact networks in order to obtain their career, their connections to external actors are much broader and diverse than those of males (Ibarra, 1993), and their possession of weak ties is more critical to prove their legitimacy compared to men (Ibarra, 1992). Hillman et al. (2002) found that female directors are more likely to be community influentials with better access to regulation authorities. This is likely to aid innovation efforts in terms of expanding the business to different geographical areas as Greenhalgh and Rogers (2010) argue that innovation only has to be new in the relevant context. Therefore, a female's possession of weak ties arguably aids the efforts and success of carrying out innovation activities. Furthermore, as board gender diversity can improve access to different geographic markets (Eklund et al., 2009), the firm will collect more ideas and inputs for creating new products (Schilling, 2017). Last, Granovetter (1973) means that weak ties facilitate a more comprehensive collection of information from the external environment. As gender heterogeneous boards provide additional information sources (Peterson & Philpot, 2007; Randøy et al., 2006) through more weak ties, leading to a more exhaustive coverage of useful information (Granovetter, 1973), boards become more proficient in solving problems and making decisions (Gordini & Rancati, 2017). In general, working with innovation implies shortage of information (Schilling, 2017), and shortage of information increases the environmental uncertainty the firm must deal with (Pfeffer & Salancik, 1978). Therefore, the more comprehensive information the board collects, the more it reduces environmental uncertainty which in turn will make innovative activities more predictable such that the firm becomes more inclined to innovate. On the other hand, with a variety of unique linkages, the risk of common information effect increases where board members tend to discard new and unique information to only discuss the information that is held by several members – regardless of the accuracy of the widely shared information (Stasser & Titus, 1985). This discussion bias implies a low utilization of unique and valuable information the minority group brings to the board and potentially leading to suboptimal decisions.

2.3.2 Monitoring

Despite that innovation is a key for competitive advantage and long-term survival (Brown & Katz, 2011; Katila & Ahuja, 2002), the associated risks from a manager perspective such as uncertainty regarding financial payback time and low success rate, indicate the conflict of interests (Shen & Gentry, 2012). Consequently, effective monitoring is a significant factor for enhanced governance of innovation (Chen et al., 2018). More specifically, there are two well-established theoretical arguments explaining why we believe there is a positive effect of monitoring on innovation. The first is based on the career concern model (Aghion, Van Reenen & Zingales, 2013; Holmström, 1999), meaning that managers are reluctant to risky innovation activities since failures are associated with inferior managerial skills (Chen et al., 2018). However, as monitoring increases the board's ability to separate managerial skills from external randomness, managers become more comfortable to innovate knowing that the board is able to distinguish between good managerial ability and stochastic reasons (Aghion et al., 2013). Hence, effective monitoring can increase managers willingness to innovate. The second argument stems from the quiet life model (Bertrand & Mullainathan, 2003), saying that managers are reluctant to costly innovation activities as it probably means deviation from status quo, implementation of new routines and more deliberate effort (Chen et al., 2018). However, enhanced monitoring can reduce this risk of managerial shirking and enforce innovation and value-creating activities. Consequently, female directors' contribution towards enhanced monitoring mitigates two agency problems related to managers' incentives to innovate. Enhanced monitoring will also have positive effects on minimizing innovation costs – one area to consider for successful innovation (Schilling, 2017). Thus, a more gender diverse board's effect on monitoring will both lead to a shift in managers' focus that facilitates long-term survival, dependent on innovation, and to higher pressure on efficient innovation work. As directors are more present, they will plausibly influence the direction of the firm to a higher degree that can facilitate a focus on innovation, and as more critical questions are asked, this will likely initiate a more creative process generating ideas for innovation. On the other hand, too much monitoring can also be detrimental for innovation since it normally requires some degree of flexibility (Adams & Ferreira, 2009; Schilling, 2017). Based on the above, our first hypothesis is as follows:

H1: Increased board gender diversity will have a positive causal effect on innovation.

2.4 Innovation's effect on financial performance

In order to define firm performance, previous literature highlights the use of accounting-based and market-based metrics (Adams & Ferreira, 2009; Post & Byron, 2015). Accounting metrics measure profitability in the past and present whereas market metrics measure expectations and speculations from external stakeholders (Post & Byron, 2015). Thus, the metrics cover different aspects of financial performance and should not be seen as interchangeable (Carter et al., 2010). We therefore argue that it is only when these two metrics are combined that one can fully understand a firm's financial performance. Innovation is widely seen as the foundation for firm survival and is a necessary activity that firms need to pursue in order to survive long-term as it enhances firm performance (Baumol, 2002; Brown & Katz, 2011; Cefis & Marsili, 2005; Katila & Ahuja, 2002; Schumpeter, 1942). Caves and Ghemawat (1992) stated that firms pursuing innovation efforts yield higher margins and increased market shares. Additionally, Teece et al. (1997) found that firms that are able to identify new opportunities and organize accordingly are performing better. These results are similar to those of Zahra and Garvis (2000) who found that firms pursuing innovation efforts are more proactive and accepting a higher risk-level yield better performance. The underlying reason for this comes from innovation as a way for firms to develop certain capabilities to ensure resilience when the external environment changes (Nelson & Winter, 1982; Teece et al., 1997).

The above elaboration stresses two important points; innovation has indeed an effect on firm performance (and possibly also a reversed; high performing firms have more resources to spend on innovation), and thus, investments in innovation activities are a strategic decision. Since the board is responsible for strategy formulation (Post & Byron, 2015; Zahra & Pearce, 1989), the board has an effect on firm performance through innovation as well. To support this connection, Zahra and Stanton (1988) argue that the board constitutes a substantial component in innovation activities; Zahra, Ireland and Hitt (2000) emphasize the significance of the board's influence on the level of innovation; and Miller and Triana (2009) argue that boards, consistent with RDT, shall allocate resources and provide ideas and relationships that in turn affects innovation. More specifically, the demographic diversity of the board is related to innovation as it brings strategic human and social capital resources to the board that positively affect its responsibilities, thereby enhancing innovation (Erhardt, Werbel & Shrader, 2003; Miller & Triana, 2009). Thus, the board plays an important role in initiating and organizing innovation.

Several studies have explicitly examined the effect of R&D investment on firm performance in different aspects. Parisi, Schiantarelli and Sembenelli (2006) found that R&D spending increased the probability of commercialization of new products. They also concluded that investments in R&D facilitated absorption of new technologies increasing productivity which follows the logic behind absorptive capacity; R&D spending positively affects a firm's ability to detect, assimilate and utilize new knowledge which is beneficial for subsequent innovation activities (Alexy, George, & Salter, 2013; Schilling, 2017). The benefits of absorptive capacity are also demonstrated by Cassiman and Golovko (2011) who found that exporting firms benefited more from investments in R&D since it enabled more efficient assimilation of the gained knowledge from the export markets. Finally, Lööf and Heshmati (2006) confirmed a causal relationship from R&D spending to innovation output and accompanying financial performance for Swedish firms. As a result, our second hypothesis is:

H2: Increased innovation will have a positive causal effect on financial performance.

2.5 Mediation of board gender diversity and financial performance

While the literature posits an array of positive attributes of board gender diversity, several researchers also highlight potential disadvantages. First, while heterogeneous boards are making more well-informed decisions, there is a lack of trust, co-operation and shared experiences which impede the speed of the decision-making process and thus impair strategy formulation (Adams & Ferreira, 2009; Nielsen & Huse, 2010). Second, for firms with weak shareholder protection, Adams and Ferreira (2009) found a small positive effect on shareholder value as increased monitoring effectiveness from gender diversity fulfilled a valuable role to shareholders, compared to firms with strong shareholder protection where the increased monitoring led to detrimental rigidity. This is in line with Jensen (1993) who argues that more diverse boards may suffer from less efficient and resolute monitoring capabilities. In contrast to this, Post and Byron (2015) found that the effect of more female directors on accounting-based performance is positive in countries with stronger shareholder protection. Additionally, they found that more female directors led to a positive effect on market-based performance in countries with greater gender parity and a negative effect in countries with lower gender parity. Last, Carter et al. (2003) provide another perspective of the topic when emphasizing that there is no clear-cut prediction of the link between board gender diversity and firm performance.

Thus, this indicates several sources of noise between a board's demographic characteristics and firm outcomes (Østergaard, Timmermans & Kristinsson, 2011).

While previous research is contradicting, we find adequate evidence supporting a positive relationship between board gender diversity and financial performance (Ararat et al., 2015; Campbell & Mínguez-Vera, 2008; Carter et al., 2003; Erhardt et al., 2003; Gordini & Rancati, 2017; Post & Byron, 2015). According to Cannella, Finkelstein and Hambrick (2009), contradicting results indicate that the relationship between board gender diversity and firm outcomes is not simple and direct but rather complex and indirect. The use of mediation partly counters this aspect as it identifies how one variable affects another and thus, creates a more robust explanation of possible causal relationship (Baron & Kenny, 1986). There are only two studies adopting a mediation approach in examining board gender diversity with the same scope as this paper (Chen et al., 2018; Miller & Triana, 2009). Miller and Triana (2009) found that board gender diversity and innovation was positively related, while they found no effect of board gender diversity on financial performance. However, as they performed a cross-sectional study, they had limited ability to account for possible time delays from gender diversity's effect on subsequent outcomes improving firm performance. In contrast, Chen et al. (2018) found that the fraction of female directors is related to innovation success and through this enhances financial performance in industries characterized as innovation intensive.

While Miller and Triana (2009) were unable to demonstrate innovation as a mediating variable, we recognize the research on gender diversity and innovation as well as innovation on financial performance in combination with the paper of Chen et al. (2018) to be sufficiently convincing that innovation indeed can mediate the relationship. First, research regarding board gender diversity and innovation is more univocal from an AT- and RDT-perspective; the inclusion of female directors facilitates monitoring and strategy formulation. In practical terms, a more gender diverse board induces a stronger focus on long-term innovation efforts by enhanced monitoring and possesses more perspectives, additional information and larger contact networks, resulting in improved decision-making. Second, research has demonstrated that innovation is clearly linked to financial performance (Caves & Ghemawat, 1992; Lööf & Heshmati, 2006; Teece et al., 1997; Zahra & Garvis, 2000). As we found adequate evidence that board gender diversity has a positive causal effect on financial performance, we argue that this effect is indirect as it is mediated by innovation. Therefore, our third hypothesis is:

H3: Innovation positively mediates the causal relationship between board gender diversity and financial performance.

2.6 Summary

Our suggested model involves four steps affecting each other (see Figure 2.1). First, gender diversity partly influences a board's ability to engage in monitoring and resource provision. The underlying reason for this is that females and males have different human and social capital. For example, resource provision is facilitated by female directors who tend to have different occupational backgrounds and individual experiences, and possess more advanced degrees and weak ties, compared to male directors. Monitoring is enhanced by female directors having higher attendance, being more inclined to ask critical questions and to sit on audit and monitoring committees, and overall be more likely to be independent directors. Second, the board's ability to engage in monitoring and resource provision will partly influence innovation. In terms of resource provision, board gender diversity brings a broader range of ideas, perspective and additional sources of information which undoubtedly aids the innovation process. With regard to monitoring, enhanced monitoring makes managers more inclined to innovate as the board's ability to distinguish skills from bad luck is improved (career concern model) and leads to less managerial shirking in favor of more effortful innovation activities (quiet life model). Last, innovation is necessary for long-term survival and can yield higher margins, increasing market shares and competitive advantage. Additionally, investments in R&D facilitate higher absorptive capacity and commercialization of new products. Therefore, innovation is likely to influence financial performance.

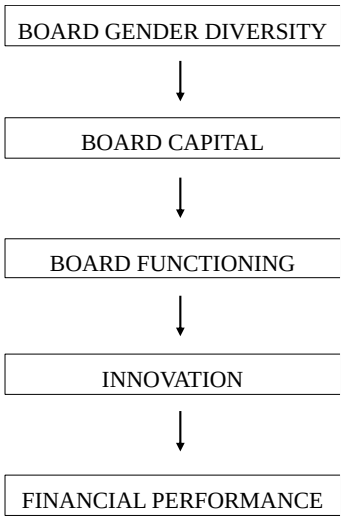

Figure 2.1 Illustration of the suggested relationship

3 Methodology

3.1 Research design

As the use of a quantitative design for studies concerning board gender diversity is a well-established approach in strategic management and in combination with the quantifiability of our variables, we found a quantitative design most suitable for our study. Moreover, our hypotheses are composed from established organizational theories and previous research, and are tested using quantifiable variables, which resulted in a deductive approach (Bell, Bryman & Harley, 2019; Creswell, 2014). We used multiple regression analysis since we had several independent variables; our variable of interest, either board gender diversity, innovation or both, and a couple of control variables with the intention to partial out their effect on the dependent variable (Wooldridge, 2016). The advantage of a multiple regression analysis is its ability to increase the likelihood of *ceteris paribus*, or all else being equal, a condition that only perfectly can be achieved in an experimental setting by keeping other variables fixed (Wooldridge, 2016).

With respect to previous literature, our research design's main contributions are the use of panel data and mediation. Since we identified that the vast majority of previous research emphasized that board gender diversity is likely to be endogenously influenced by economic and governance factors, leading to spurious results, (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2010; Chen et al., 2018; Randøy et al., 2006), we decided to test our hypotheses with a longitudinal approach by forming a micro panel data set. Likewise, there is a call for examining mediating variables to get a more fine-tuned understanding of the relationship between board gender diversity and firm performance (e.g. Bear et al., 2010; Miller & Triana, 2009; Post & Byron, 2015; Torchia et al., 2018; Tyrowicz et al., 2020). As a consequence, we decided to contribute to this area by investigating the mediating effect of innovation. By comparison, a case study would have faced limitations regarding its generalizability due to its narrow focus and an experimental study would have entailed practical difficulties in setting up an adequate experimental setting (Bell et al., 2019).

3.1.1 Panel data

Panel data, or equivalently longitudinal data, is defined as "a data set consisting of a time series for *each* cross-sectional member in the data set" (Wooldridge, 2016:9). As we decided to follow the same firms over several years, leading to multiple observations on the same unit, we worked

with a balanced data set.[2] The advantage with a balanced data set is increased efficiency in terms of lower variances and higher precision in estimations, resulting in smaller confidence intervals (Wooldridge, 2016).

The possibility to observe the same firms over several years makes panel data superior for several reasons; first, it enables mitigation of omitted variable bias as we controlled for unobserved heterogeneity in terms of firm and year specific effects (Angrist & Pischke, 2009; Wooldridge, 2016); and second, it facilitates causal inferences where inferring causality would be difficult if each unit only was observed once as it enables control of a variable's lagged values (Bell et al., 2019; Wooldridge, 2016). These two issues, omitted variable bias and reversed causality, are the two most common endogeneity concerns influencing board gender diversity which can lead to spurious results (Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2010; Chen et al., 2018; Randøy et al., 2006). Omitted variable bias can occur as some firms have a more progressive firm culture, leading to both better governance and higher gender diversity that potentially influence innovation or financial performance which hinder us from trace out the true effect of board gender diversity (Adams & Ferreira, 2009). Reversed causality on the other hand, may arise due to self-selection of female and male directors to boards based on firms' earlier innovation efforts or financial performances. Farrell and Hersch (2005) argue that female directors can selectively choose to only join boards of successful firms due to a lower proportion of women than men with adequate experience. These aspects undermine the robustness of any conclusions that can be made, but panel data is an effective research design for dealing with these (Bell et al., 2019; Wooldridge, 2016). In comparison, a cross-sectional study cannot control for the unobserved specific effects (Angrist & Pischke, 2009; Wooldridge, 2016) and it lacks the ability to control for reversed causality since only one time period is investigated (Bell et al., 2019; Wooldridge, 2016). Therefore, the use of panel data allowed us to reach higher accuracy without having severe endogeneity issues significantly affecting our results (Angrist & Pischke, 2009).

Furthermore, as the impact of strategy formulation typically evolves over time and does not have an immediate effect on firm performance (Erhardt et al., 2003; Carter et al., 2010), this implies that the impact of board gender diversity arguably arises over time as well. Thus, a longitudinal approach accounts for the gender diverse board's potential to affect innovation and

[2] Balanced data set means that every year the same firms are studied (Wooldridge, 2016). Hence, the total number of observations will then be the product of number of firms and the number of years these firms are studied.

financial performance as it observes the companies over several years. Comparing with previous studies that are cross-sectional, they partly mitigate this by measuring performance in a later year than measuring board gender diversity (e.g. Bear et al., 2010; Carter et al., 2010; Chen et al., 2018; Miller & Triana, 2009) even if this method cannot be as robust as studying each firm over several years (Angrist & Pischke, 2009). Similarly, our research design is less vulnerable to temporal fluctuations in variables' values since several years are observed (Erhardt et al., 2003; Wooldridge, 2016). For example, it could be the case that one year is a particularly bad year for the economy but by using a longitudinal design, we could better control for factors such as market fluctuations and we reduced the effect of extreme exogenous effects as several years were included. Consequently, we argue that panel data provided us with more robust results. However, panel data is not without drawbacks. The overall disadvantage with panel data for a researcher is the increased complexity imposed on the data analysis (Wooldridge, 2016). This was realized to us as more data had to be collected, understood and adjusted in a consistent way. Additionally, we faced a risk of missing values since panel data requires observing the same companies over several time periods. As a result, we were forced to make adjustments to our sample in order to maintain a balanced data set – adjustments which are not necessary for a cross-sectional study and therefore, comes with less impact from the researchers.

3.1.2 Mediation

According to Baron and Kenny (1986), mediation occurs when a given variable fully or partly accounts for the relationship between two variables. Thus, a mediating variable explains how and why an independent variable causes an effect in a dependent variable. In comparison, a moderating variable affects the strength and/or direction of the effect between the independent and the dependent variable and therefore, they are not interchangeable (Baron & Kenny, 1986). We applied the mediation technique by Baron and Kenny (1986) as it has been adopted by previous mediation studies to examine board gender diversity (see Bear et al., 2010; Miller & Triana, 2009; Torchia et al., 2018). Applying this technique to our study means that the independent variable is board gender diversity, the mediating variable is innovation, and the dependent variable is financial performance. The three regressions we will run based on the mediation technique translates into the following:

1. The effect of board gender diversity on innovation (R1),
2. The effect of board gender diversity on financial performance (R2),
3. The effect of board gender diversity and innovation on financial performance (R3).

The requirements for proving a mediating relationship are that board gender diversity has a significant effect in R1 and R2; and the effect of innovation should be significant while the significant effect of board gender diversity should be partially or completely reduced in R3 (Baron & Kenny, 1986). The last requirement simply means that when we control for innovation, the effect of board gender diversity on financial performance is lower compared to the effect in R2 since the effect of board gender diversity in the last regression is mediated by innovation. If board gender diversity becomes non-significant while innovation is significant in R3, perfect mediation is achieved as it has no effect on financial performance when innovation is controlled for (Baron & Kenny, 1986).

However, we want to clarify that this mediation technique explicitly tested H1 and H3 but not H2 (i.e. innovation affects financial performance). In order to distinctly test H2, we added a fourth regression which is highly similar to R3 but it differs as we did not include the gender diversity variable (i.e. the independent variable in Baron and Kenny (1986)); it only tested whether innovation affects financial performance without including board gender diversity. This regression is consistently referred to as R4. Figure 3.1 illustrates the four regressions.

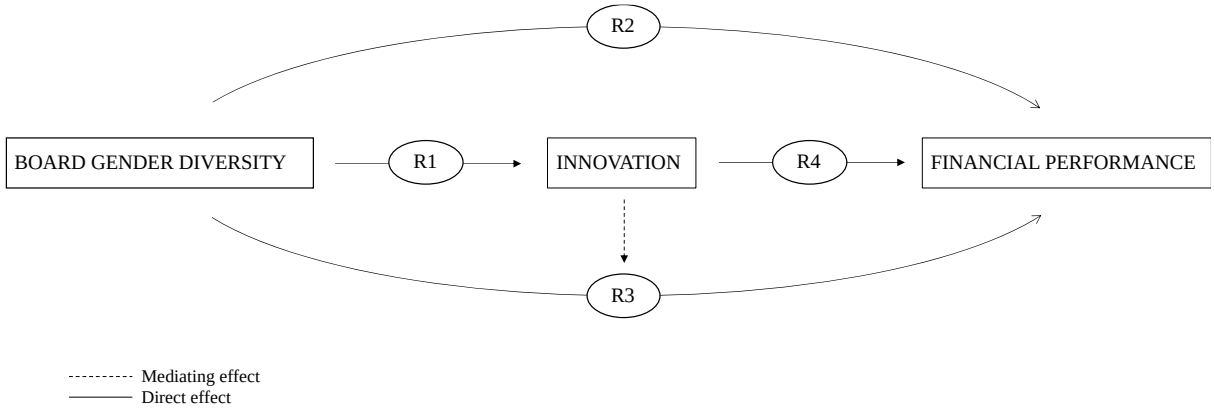

Figure 3.1 Illustration of regressions

3.2 Data collection and sample

3.2.1 Data collection method

We collected our data from two external data sources; *Standard and Poor's Compustat Global* (hereinafter Compustat) to collect financial data, and *Directors and Auditors in Sweden's Listed Companies* (hereinafter *Directors and auditors*) by Fristedt and Sundqvist (2008, 2009) and Fristedt et al. (2010–2014) to collect board composition data. Compustat is an online database,

containing secondary financial data of public firms worldwide provided by Standard & Poor's. As Compustat consists of more than 1 000 items, we allocated a lot of time in order to understand which items we should use for our variables by reading the manual, supervising with a teacher with experience from Compustat, and searching for how previous research had used the items of Compustat when calculating the same variables as we used. *Directors and auditors* is a collection of books, published on a yearly basis, containing data of Swedish public firms' board compositions. However, as the data from *Directors and auditors* was not digital, we manually transcribed the data into Excel. Both data sources collect data on a firm-level and yearly basis which facilitated our longitudinal design. Relying on secondary data enabled us to work with a large and high-quality data set in order to make conclusions with higher generalizability compared to a qualitative method (Bell et al., 2019). Finally, both sources are based on annual reports which guaranteed high quality regarding how data is reported.

3.2.2 Sample selection

The firms in our study were taken from the Nasdaq OMX Stockholm (OMXS) which is the largest stock exchange in Sweden in terms of market capitalization. Hence, the OMXS functioned as our sampling frame (Bell et al., 2019). We decided to constrain our sampling frame to firms from one country since we could not find any strong arguments to make a multi-country study based on the following advantages and disadvantages. First, with respect to our hypotheses, there was no explicit need to study board gender diversity in several countries in order to test them. Second, as with all quantitative studies using multivariate analysis, one should strive to achieve *ceteris paribus*, but if we were to include firms from different countries, we would automatically introduce more noise into our regression model. Third, calls have been made for longitudinal studies that only focus on one country to better trace the evolution of board gender diversity (Carter et al., 2010; Gordini & Rancati, 2017; Tyrowicz et al., 2020).

We decided to study firms in Sweden due to Sweden's high female labor participation rate as women are more likely to reach the corporate board in countries with high female labor participation rates (Tyrowicz et al., 2020). Subsequently, by using Swedish firms, we obtain a wide range of values on our independent variable. This is beneficial for two reasons; first, it allows for more efficient estimations as the independent variable has high individual specific variation (Wooldridge, 2016); second, we will be more able to detect a curvilinear relationship of board gender diversity's effect on innovation and performance. Erhardt et al. (2003) emphasize the benefit of having a wide range of board gender diversity values since it is likely

that the effect of increased board gender diversity has a diminishing marginal effect on performance. Other studies emphasize the characteristics of the corporate legal system, such as civil law versus common law (Campbell & Mínguez-Vera, 2008; Gordini & Rancati, 2017) which could affect differences in board composition regarding gender diversity. Indeed, we could have performed separate studies on different countries in order to increase the generalizability of our results, but the above arguments and the fact that we have only found two studies examining board gender diversity and financial performance in a Swedish context (Eklund et al., 2009; Randøy et al., 2006), are arguably adequate for us to focus on Sweden.

The studied time period spans over the years of 2008 to 2014. Deciding the length of the studied time period is one of the major issues with a longitudinal study (Bell et al., 2019). However, we based our choice of this time span that it is long enough to capture any potential causal effects of board heterogeneity on subsequent firm outcomes. Additionally, the rate of female director participation on Swedish boards increased from 19 to 26 percent (an increase of 36 percent) during these years (Fristedt & Sundqvist, 2008; Fristedt, et al., 2014) which makes it relevant to study whether the increased board gender diversity may impact firm outcomes. Lastly, in 2008, a revised code was implemented where firms listed on OMXS should strive for an equal distribution of gender on boards (Fristedt & Sundqvist, 2008). Therefore, we argue 2008 was a suitable starting year for our time period. A possible drawback of including 2008 and the following years is the financial crisis. This may of course impact the financial performance but as we work with a balanced data set, meaning that we followed all companies from 2008 to 2014, it implies that all of them survived the crisis. Hence, if a firm liquidates after this point, we find it unlikely to be a result of the financial crisis such that the crisis should not bias our data collection. Furthermore, since we lacked access to data on board composition from *Directors and auditors* later than 2014, this year became a natural ending point of the studied time period.

As a first step to determine our sample, we used a simple random sampling technique to select half of the companies that were present on OMXS in 2008, using the random function in Excel. This implied that the sampling frame consisted of 294 companies (Fristedt & Sundqvist, 2008) with the same likelihood to be included (Bell et al., 2019). We viewed 50 percent of the entire population as a sufficient size to draw conclusions for the entire population, even when taking into account that some of these selected companies would be excluded in the following steps. As a second step, we excluded 20 firms that were not listed on the OMXS over the entire time

period due to being delisted during the time period. Further, in accordance with Chen et al. (2018), we excluded nine companies denoted as "Financials" by Compustat which included banks and insurance and investment companies, as these are reported differently in Compustat leading to possible inconsistencies in how data is reported.[3] The final number of companies equaled 118 and they are listed by their Compustat ID (GVKEY) in Appendix A. As we had to do some adjustments to our sample, we acknowledge that we do not fully satisfy the requirements of a simple random sample (Bell et al., 2019). In retrospect, it would have been more appropriate to conduct the adjustments as a first step before using the random function in Excel in order to make a truer random sampling. However, as our adjustments to the sample have been minor, we deem our final sample to still obtain a sufficient degree of randomness.

Regarding sample size, Hair, Black, Babin and Anderson (2014) state that the number of observations should not fall short of 50 – and preferably not below 100 – for multiple regression analysis in order to have a high probability in detecting any significant relationships. In addition, one should take into account the number of explanatory variables used in the regression. As a minimum rule, Hair et al. (2014) point out five observations per explanatory variable but a ratio of 15:1 or as large as 20:1 is desirable in order to minimize the risk of overfitting the data (i.e. drawing conclusions from a small sample with low level of generalizability). As the size of our final sample was 118 companies studied over seven subsequent years, resulting in 826 observations and with five to six explanatory variables (see 3.4.2), we argue the sample size to be large enough to draw conclusions with statistical power.

3.2.3 Sample data

Once we had collected our data, we encountered some difficulties leading us to make some adjustments. First, we noticed that some firms reported their financial data from their annual reports in EUR instead of SEK, which we solved by using the average annual exchange rate by the Swedish Riksbank. Second, some firms had no data on the daily number of outstanding shares for some periods. Therefore, we complemented our data by manually reading the annual reports of those firms. However, in neither of the cases were there any announcements about changes in outstanding shares leading to the assumption that the outstanding shares were constant for the missing periods. Third, for some firms, the daily stock prices were reported in

[3] However, it should be noted that even after our exclusion of companies defined as financials by Compustat, our sample consisted of two companies defined as financials by OMXS.

several currencies. We decided to use the Swedish currency for all these companies in order to make as small changes to the original form of the data as possible. Fourth, we needed to ensure that the financial data was expressed in the same unit. For example, the daily stock prices were expressed in SEK while the financial data from annual reports was expressed in millions SEK. As Compustat consistently reports each item in the same unit, regardless of firm or time, we looked in the Compustat manual to see each item's reported unit. To ensure a correct adjustment, we manually compared five companies' values from Compustat with their annual reports. Last, since several companies have different classes of stocks, such as A, B, C and/or preferred stocks, we included all classes in the calculation of firm value as they all contribute to reflect the market valuation of a company. In contrast, we did not make any changes to the data from *Directors and auditors* given that no human errors occurred in the transcription process. We controlled for this by doing acceptance sampling on ten companies each year once the data was collected in Excel and no errors were found.

In our sample, a third of the observations lacked data on R&D expenses. Hence, lack of R&D values was set equal to zero in line with previous research (Chen et al., 2018; Miller & Triana, 2009). By doing so, we avoided biasing our results by excluding observations with low R&D spending (Miller & Triana, 2009; O'Brien, 2003). We argue that this is reasonable to do since companies not reporting R&D expenses are likely to spend a marginal amount of money on R&D. However, observations with missing values were still useful as they provided complementary data points to indicate that board gender diversity is not always equivalent to spending in R&D. If we would have excluded these observations, we would have manipulated our sample more than if we were to set missing values equal to zero. On the other hand, if a firm has no R&D expenses reported, it does not necessarily imply that this particular firm is not working with innovation. Instead, it is likely that the firm is reporting costs of innovation activities differently as there is no enforcement to report R&D expenses in Sweden. However, determining the value of observations with missing values is highly ambiguous and we could not find any theoretical or empirical support for how this should be done. Since it was still important to investigate the effect of equalizing missing values with zeros, we reran the regressions by excluding observations with missing values on R&D expenses in order to see if the results drastically changed.

Last, the sample included a couple of outliers for the financial performance variables Tobin's Q and ROA. Notably, all outliers came from companies listed as health care which required

some further investigation. The negative outliers were results of the companies being in a research phase with limited products to sell while the positive outliers were due to realization of successful findings. Hair et al. (2014) state that if an outlier can be explained by an extraordinary reason, such as the ones mentioned, the observations should be excluded unless variables reflecting this reason are included in the regression. Hence, as we explicitly investigate R&D expenses, we argue that this extraordinary reason is controlled for and we could not justify exclusion of these companies. However, we conducted our regressions without these companies as well (unreported), in line with suggested precautionary principle (Hair et al., 2014; Wooldridge, 2016), but could not prove any significant changes in our results. This made us secure that the existence of outliers in this case was explained by our included variables or otherwise do not have any severe impact on our estimations.

3.3 Variables

3.3.1 Independent variable – Board gender diversity

This paper is exclusively focusing on ordinary board members elected by the annual general meeting and employee directors as they occupy the same tasks and voting rights. The reason for excluding alternate board members is that they are only responsible to carry the voice of the ordinary board members in situations when they are unable to attend a board meeting. Moreover, since Sweden has a one-tier board system, we did not make any further limitations regarding which part of the corporate structure to focus on.

The operationalization of board gender diversity has mainly been done in three ways in previous research; the presence of female directors, the proportion of each gender on the board, and the proportion with respect to the evenness of the distribution of board members' gender (Post & Byron, 2015). In order to meet our definition of board gender diversity, we operationalized board gender diversity with the latest way. Hence, we applied the Blau and the Shannon indices; both argued to be appropriate measures of diversity (Harrison & Klein, 2007) as they reflect the inherent variations within a group of people (Blau, 1977; Shannon, 1948) – an important criteria when diversity is defined as degree of heterogeneity which the operationalization then needs to capture (Harrison & Klein, 2007). The Blau index is calculated as:

$$Blau_{i,t} = 1 - \sum_{i=1}^{n} P_i^2$$

Where P_i represents the share of board members in group i, i.e. female directors or male directors, and n is the total number of ordinary board members.

Since we define gender as a dichotomous variable (male or female), the Blau index ranges from 0 (when only one board member group is present) to 0.5 (when there is an equal distribution of men and women on the board). In contrast, the Shannon index is more sensitive to small differences in a board's gender composition due to its logarithmic function (Campbell & Mínguez-Vera, 2008) and therefore used to ensure robustness in our regressions. The Shannon index is calculated as:

$$Shannon_{i,t} = -\sum_{i=1}^{n} P_i * ln(1 + P_i)$$

Where P_i, i and n have the same meaning as in the Blau index.

The Shannon index ranges from 0 to approximately 0.69, from a perfect homogenous group to a perfect heterogenous group. Both the Blau (Ararat et al., 2015; Bear et al., 2010; Campbell & Mínguez-Vera, 2008; Miller & Triana, 2009; Gordini & Rancati, 2017) and the Shannon indices (Campbell & Mínguez-Vera, 2008; Gordini & Rancati, 2017) have been widely used in this research area. Since the Shannon index has a larger range for a small number of categories, it is advantageous to the Blau index (Harrison & Klein, 2007) even though it is harder to interpret due to its logarithmic feature. An increase in any of the indices should be interpreted as an increase in board gender diversity. It should be noted that the Blau and the Shannon indices are used as substitutes; we ran each regression with both the Blau and the Shannon indices. Hence, they are never used simultaneously in any regression as that would lead to high collinearity (Hair et al., 2014; Wooldridge, 2016) and it lacked purpose since they were aimed to operationalize the same phenomena.

3.3.2 Dependent variable – Financial performance

To be consistent with our definition of financial performance, we operationalized financial performance using one market-based and one accounting-based measure; Tobin's Q and ROA, respectively. In its most basic nature, Tobin's Q measures a firm's market value with respect to its assets while ROA measures how much income is generated from the firm's assets. Both Tobin's Q and ROA are ratios and therefore they are standardized performance measures, avoiding scale biases which simple difference measures suffer from (Chung & Pruitt, 1994).

According to Wernerfelt and Montgomery (1988), Tobin's Q is an appropriate proxy for a firm's competitive advantage as it is an indicator of the market's expectations of future earnings and long-term firm value. Following this, since innovation has been shown to be essential for long-term survival, we argue it is reasonable to apply in this context. Moreover, Tobin's Q is less subject to accounting manipulation, such as asset valuation techniques, and different report distortions (Campbell & Mínguez-Vera, 2008; Carter et al., 2010) and it has been widely used in the corporate governance literature (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2003; Carter et al., 2010; Ferrari, Ferraro, Profeta & Pronzato, 2016; Gordini & Rancati, 2017; Post & Byron, 2015; Rose, 2007). Despite that Tobin's Q may seem superior to accounting measures, we included ROA as it offers a complementary picture of firm performance with a focus on past performance (Post & Byron, 2015) and the firm's efficiency in generating profit in relation to its assets (Carter et al., 2010; Marton, Sandell & Stockenstrand, 2018). Thus, the inclusion of ROA provides a more holistic view of firm financial performance and has also been widely used in previous studies (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2010; Erhardt et al., 2003; Ferrari et al., 2016; Gordini & Rancati, 2017; Post & Byron, 2015; Randøy et al., 2006). ROA was calculated as the sum of earnings before interests and taxes (EBIT) and income interests divided by total assets (Marton et al., 2018).

All previous studies addressed above have used an approximation of Tobin's Q since the original calculation of Tobin's Q is very complex and cumbersome (Chung & Pruitt, 1994). We followed a widely used approximation of Tobin's Q (e.g. Campbell & Mínguez-Vera, 2008; Chen et al., 2018; Gordini & Rancati, 2017; Rose, 2007) which was calculated as the market value of equity plus the book value of debt to the book value of assets. Thus, the equilibrium value is at one where the market value of the firm equals the replacement cost of its assets (Campbell & Mínguez-Vera, 2008). A ratio exceeding one is interpreted as the firm is expected to create more value from existing resources (Campbell & Mínguez-Vera, 2008) and as a possession of strong competitive advantage or growth opportunity (Gordini & Rancati, 2017). A ratio less than one indicates the opposite; poor utilization of available resources, a weak competitive position and expectations of degrowth. Hence, we expect that firms with high growth potential due to the possession of competitive advantage should also have a higher Tobin's Q.

3.3.3 Mediating variable – Innovation

Previous literature has stated that R&D is a solid proxy for firm innovation (Dalziel, Gentry & Bowerman, 2011; Hitt, Hoskisson & Kim, 1997; Hoskisson, Hitt, Johnson & Grossman, 2002). With regard to that, spending in R&D is necessary for entrepreneurial exploration (Zahra, 1996), ensuring a firm's long-term survival capability (Brown & Katz, 2011) and introduction of new products (Parisi et al., 2006). However, it is important to consider that R&D spending does not guarantee innovation since it is an inherently risky activity (Dalziel et al., 2011; Shen & Gentry, 2012) and since it is an input measure with no information regarding how successful the investments in R&D turn out to be (e.g. Chen et al., 2018; Richtnér, Brattström, Frishammar, Magnusson & Björk, 2017). Regardless of the success rate, early investments in R&D is beneficial for a firm's absorptive capacity (Cohen & Levinthal, 1990; Parisi et al., 2006) which indeed has a positive effect on a firm's future innovation capabilities (Schilling, 2017).

With regard to using an absolute measure (e.g. R&D expenses) or a ratio measure (e.g. R&D intensity) there are conflicting views in previous literature. R&D intensity has been accepted as a valid proxy for innovation (Hitt et al., 1997; Hoskisson et al., 2002) and has been used in similar studies as our (Miller & Triana, 2009). However, Rothaermel and Hess (2007) mean that an absolute measure is most suitable since a ratio measure of R&D contains high uncertainty whether the measure depends on the R&D expenses, as hoped for, or on the denominator (normally total assets). Consequently, we used the natural logarithm of R&D expenses as our operationalization of innovation which also was done by Chen et al. (2018). We used the natural logarithm as it is beneficial to ensure linearity and normal distribution of monetary variables (Wooldridge, 2016).

3.3.4 Control variables

The purpose of control variables is to partial out the effect they may have on the dependent variable in order to achieve a more precise and efficient estimation (Wooldridge, 2016).[4] As the mediation technique by Baron and Kenny (1986) implies that the dependent variable differs between the regression, the control variables could also slightly differ between the regressions. However, as we found strong evidence for adding the same control variables in all regressions

[4] Practically speaking, there is no difference between an independent and a control variable. In fact, all control variables are independent variables, but they can be separated based on their purpose of being included in the regression (Wooldridge, 2016).

from previous research, we felt secure to not differentiate the control variables with respect to the different regressions. When considering the inclusion of control variables, one needs to bear in mind that the more variables introduced to the regression, the more estimations need to be done given the same amount of data observations with higher risk of misleading results. Thus, this could have a detrimental effect on the efficiency of the estimations due to this trade-off (Hair et al., 2014; Wooldridge, 2016). Similarly, Bear et al. (2010) emphasize that a small sample size should imply a parsimonious approach to the inclusion of control variables. Followingly, we present the control variables used in our study.

First, we found that board size is an important control variable for both innovation and financial performance and was operationalized as the number of board members appointed by the annual general meeting, in accordance with previous research (Carter et al., 2003; Eklund et al., 2009; Erhardt et al., 2003; Randøy et al., 2006; Torchia et al., 2018). The number of directors on a board will arguably impact innovation since larger boards entail better access to more unique resources (Jackling & Johl, 2009) and influence the way directors perform their tasks (Fama & Jensen, 1983). Hence, board size will arguably affect a firm's strategic focus on innovation. Furthermore, Gordini and Rancati (2017) and a meta-analysis by Dalton et al. (1998) found that board size is positively and significantly related to financial performance. Second, firm size was operationalized as the log of total assets and used as a control variable for innovation as larger firms have more resources to spend on innovation (Schilling, 2017; Torchia et al., 2018). It was also controlled for in terms of financial performance as Gordini and Rancati (2017) and Randøy et al. (2006) established a positive and significant effect on Tobin's Q. Third, we controlled for independent directorship for both innovation and financial performance as Adams and Ferreira (2009) found a positive significant relationship between independent directorship and Tobin's Q as well as ROA. Baysinger and Butler (1985) further found that firms with independent directors outperform other firms, making it an important control variable both in terms of innovation and financial performance. Moreover, as females act as independent directors, we operationalized the control variable as the fraction of independent directors in order to avoid the possibility to have the relationships driven by the number of independent directors rather than the ratio of females on the board (Adams & Ferreira, 2009). Last, leverage was incorporated as a control variable and operationalized as the long-term debt and current liabilities to total assets. Leverage insinuates to use borrowed capital to pursue an investment opportunity, which arguably made it an important variable to control for in terms of

innovation as well as financial performance. A summary of all our variables are provided in Table 3.1.

Table 3.1 Summary of variables

Variable	Description	Data source
Independent variables		
Blau index	The inherent variation within a group of people.	Directors and auditors
Shannon index	The inherent variation within a group of people.	Directors and auditors
Mediating variable		
Innovation	The natural logarithm of R&D expenses.	Compustat Global
Dependent variables		
Tobin's Q	The relationship between the market value of a firm and its book value of assets.	Compustat Global
ROA	How efficient firm assets are used to generate profit.	Compustat Global
Control variables		
Board size	The number of ordinary board members.	Directors and auditors
Firm size	The natural logarithm of total assets.	Compustat Global
Independent directorship	The fraction of independent directors in relation to the company and major shareholders.	Directors and auditors
Leverage	The ratio of borrowed capital and total assets.	Compustat Global

3.4 Data analysis

3.4.1 Regression model

The use of more than one explanatory variable implies that we conducted a multiple regression analysis where the beta coefficients were estimated using Ordinary Least Squares (OLS) (Wooldridge, 2016). All statistical analyses were carried out using EViews version 11. Further, we applied a two-way error component model with fixed effects for our regression model as this allowed us to eliminate the effect of unobservable heterogeneity (Wooldridge, 2016). More specifically, unobservable heterogeneity means that we expected there to be firm specific and year specific characteristics which are unobserved, and thus are omitted variables (Wooldridge, 2016). For example, we expect heterogeneity exists among our firms as they have different cultures, different procedures to meet their objectives, and different backgrounds with subsequent path dependencies; all likely to influence board gender diversity. Likewise, as the focus on gender diversity has arguably increased in the first decades of the 21st century, we expect that heterogeneity over time will be present in our data as well.

There are several arguments to why we opted for fixed effects instead of random effects. First, previous literature is univocal in the superiority of fixed effects to deal with omitted variables and unobserved heterogeneity over time (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2010; Chen et al., 2018; Ferrari et al., 2016; Gordini & Rancati, 2017). Second, with respect to the different potential types of firm and year specific characteristics we exemplified above, it is intuitive that these are correlated with our independent variables and hence, fixed effects should be added (Angrist & Pischke, 2009; Wooldridge, 2016). Third, the interest of the study is not to estimate the effect the omitted variables have on our estimates but rather partial out the effect of unobservable heterogeneity, making fixed effects more suitable (Wooldridge, 2016). Note that the estimations of these effects are normally not reported (Wooldridge, 2016). Last, we also performed the Hausman test that tests whether the unobserved heterogeneity is correlated or not to our independent variables in cross-sectional level (Hausman, 1978). The test showed that the null hypothesis (random effects are most efficient) was rejected in each regression (except one, see table 4.7), implying that fixed effects indeed are suitable for our regression for firm specific effects. The test statistic for each regression is included in the tables presenting our results. The Hausman test is only applicable for testing cross-sectional hidden characteristics, therefore we used redundant fixed effects test which is pre-installed in EViews to test the year specific effects; when running our regressions with fixed effects, we rejected the null hypothesis which implied that fixed effects were not redundant on either time or cross-sectional level. In practical terms, the consequence of using fixed effects is that the adjusted R^2 (adjusted coefficient of determination) increases as 117 firm specific and six year specific variables are added to the model (Wooldridge, 2016).[5]

While fixed effects efficiently deal with the omitted variable concern, we still needed to consider the reversed causality concern. Even though there are heavy calls to control for reversed causality among previous researchers, Antonakis, Bendahan, Jacquart and Lalive (2010) found that 75 percent of quantitative studies published in the ten leading management journals did not control for reversed causality. Hence, we view our attempt to check for reversed causality as an advantage compared to other quantitative studies within management. Followingly, as our hypotheses aimed to examine causal relationships, we followed the design

[5] The reason behind why there is one less specific effect on both firm-level and year-level is that the fixed effects are dummy variables since they are unique for every company or year (Wooldridge, 2016). And as always when using dummy variable, the number of dummies equals one less the total number of units (Wooldridge, 2016).

by Campbell and Mínguez-Vera (2008) such that we conducted our regressions "in reverse", i.e. we regressed the variable of interest on the dependent variable. By doing this, we could easily conclude whether innovation and financial performance, respectively, had a significant effect on board gender diversity – lack of significance then indicated that reversed causality was not an issue for our data.

Our use of fixed effects affected how we controlled for industry. In order for fixed effects to work, there must be individual variation within each variable over time (Wooldridge, 2016). Hence, if any variable is possible to define as a constant over time for the same company, this results in perfect multicollinearity (Wooldridge, 2016). Because industry belonging was constant for all companies in our sample, we could not include industry as a dummy variable in the main regressions. Hence, we conducted subsample analysis for two industries (see 3.4.3).

3.4.2 Regression equations

We had three regressions to run in order to follow the mediation process by Baron and Kenny (1986) to test H1 and H3, and one regression (R4) to test H2. Before we carried out the correct specification of the model by using fixed effects (referred to as the correct model), we executed all regressions without any specific effects. This is called pooled OLS when no specific effects are used and implies that the time dimension of the data is ignored as each observation is seen as independent to all other observations (Wooldridge, 2016).[6] The reason for this is that we wanted to illustrate step-by-step the effect on the coefficient value, the level of significance and the adjusted R^2 of the inclusion of fixed effects. Additionally, the less impact we as researchers have on our data, the more robust our findings will be if they under different conditions which pooled OLS arguably is. For all regressions, we will use three levels of significance starting at a p-value of 0.10 indicated by one star *, 0.05 (**), and 0.01 (***). A p-value of 0.10 means that we have a ten percent risk of stating a significant relationship even if no exists.[7] The majority of previous research uses a significance level of ten percent (e.g. Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Chen et al., 2018; Gordini & Rancati, 2017; Torchia et al., 2018).

[6] To exemplify, pooled OLS implies that the data for Volvo in 2014 is seen as independent to the data for previous years of Volvo, even if this hardly is the case.
[7] This is called type 1 error or the rejection of a true null hypothesis.

As the range of Blau index goes from 0 to 0.5, and Shannon index from 0 to approximately 0.69, it becomes irrelevant to interpret the beta coefficient of these variables as a unit increase. For a more relevant interpretation of the estimation of the beta coefficient, we will constantly refer to it as a tenth unit increase's (0.1) effect on the dependent variable's expected value. All of our regressions included the four presented control variables in 3.3.4: board size, firm size, fraction of independent directorship, and leverage. Thus, the number of explanatory variables, in excess of the common intercept (β_1) and the specific effects, equals five in R1, R2 and R4, and six in R3. For a detailed explanation and mathematical outline of the regression equations for the correct model, we refer to Appendix B.

3.4.3 Subanalysis

As a complement to our main regressions, we executed two subanalyses as they cover aspects which are often considered in the light of board gender diversity, innovation and financial performance. Thus, one subanalysis was based on industry belonging and one based on the idea of critical mass. The results of these subanalyses are presented in table format in Appendix C but discussed in 5.4. We focused on the correct model (i.e. with fixed effects) and if significant relationships were found, we performed the regression in reverse to control for reversed causality. Hence, we did not carry out pooled OLS regressions or the regressions when excluding missing values of R&D expenses as the industries were selected due to their low ratio of missing values.

The first subanalysis based on industries was performed by dividing our sample into different industries, defined by OMXS. The rationale for investigating this paper's model for specific industries is a consequence of two factors; first, the level of innovation is widely different depending on which industry a specific firm adheres to (Chen et al., 2018; Miller & Triana, 2009; Torchia et al., 2018) and second, financial performance is related to industry (Chen et al., 2018; Randøy et al., 2006; Smith et al., 2006). Furthermore, we constrained this subanalysis to focus on the industries healthcare and technology due to the following reasons. First, previous research states that these industries are particularly innovation intensive (Chen et al., 2018; Dalziel et al., 2011; Schilling, 2017) which holds true for our sample as well (these industries

showed the highest average of R&D expenses). Second, the number of observations in these industries is sufficient for meeting the sample size ratio as suggested by Hair et al. (2014).[8]

With regard to the idea of critical mass, previous studies have stated that the real impact of female directors is not realized until the number of women exceeds three in order to break male stereotypes, change an all-male communication dynamic and overall to create change (Konrad et al., 2008; Torchia et al., 2011). Even if this is not the main focus of our study, we tested this idea by creating a dummy variable indicating the presence of three or more female directors on boards and used boards with less than three female directors as the base group. As mentioned in 3.3.1, measuring board gender diversity by the number of (female) directors is one of three main ways to operationalize board gender diversity. Several studies have used an absolute measure of female directors as an alternative measure of board gender diversity, although not with a focus on investigating critical mass (Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008; Carter et al., 2003; Gordini & Rancati, 2017). In our sample, a quarter of the observations had at least three female directors which implies that there is enough individual specific variation in order to use fixed effects and make robust estimations (Wooldridge, 2016).

3.4.4 Accuracy in data analysis

In order to ensure consistent estimations and accurate statistical inferences, Hair et al. (2014) state four assumptions that need to hold: linearity, normality, homoscedasticity, and independence. First, linearity is achieved as we did not find any curvilinear relationships among our variables and the use of logarithmic values on two variables further facilitates linearity (Hair et al., 2014; Wooldridge, 2016). Second, the assumption of normality is satisfied as most of our variables' mean and median values are close to each other, and we have a large number of observations leading to the central limit theorem (Wooldridge, 2016). On the other hand, financial data is typically not normally distributed (Wooldridge, 2016) – this is the case for our data as well – but when disregarding outliers, we found a clear indication of normal distribution. Third, we performed the Breusch-Pagan test manually by saving the residuals from each regression in the pooled model and regressed the squared residuals on the explanatory variables (i.e. an auxiliary regression, unreported) (Wooldridge, 2016). The Chi-Square statistic indicated that the null hypothesis of homoscedasticity was rejected and hence, the residuals do not have

[8] However, since the number of companies is not that large, we urge to keep in mind that the results derived from this subanalysis may lack high degree of generalizability.

constant variance conditional on the explanatory variables forcing us mitigate this as explained soon. Fourth, if the residuals are correlated over time, they are subject to autocorrelation and not independent. We found it logically intuitive that the residual of firm A in 2009 was likely dependent on the residual in 2008. As we both had heteroscedastic and autocorrelated residuals, we used White's robust diagonal standard errors for all regressions to ensure asymptotically consistent standard errors (Wooldridge, 2016).[9]

3.5 Validity and reliability

3.5.1 Validity

Validity is concerned with the appropriateness of the conclusions being drawn from the research (Bell et al., 2019), and there are different types of issues one must address with regard to this. First, construct validity refers to, for example, having hypotheses that are able to test what they claim to. As our hypotheses are based on two widely used theories in similar research papers in corporate governance, we arguably reached a high level of construct validity. Second, internal validity deals with the issue of causality. Since our research is interested in finding causal relationships between our variables, it was important to ensure that this causality holds. While there are many factors affecting our dependent variables, we controlled for several variables that are likely to affect and used fixed effects in order to reach *ceteris paribus* context. Third, external validity refers to the ability to generalize the findings beyond the specific context (Bell et al., 2019). However, since findings can only be generalized onto the population from which the sample was taken (Bell et al., 2019), the results of our study cannot be generalized beyond the listed firms on OMXS. Lastly, ecological validity is concerned with the ability to understand how things work in social settings. However, as our quantitative study examines the sample firms on an aggregate level, we are not able to apply the results on a particular company within our population. In order to understand the relationships on a deeper level, a more qualitative approach with in-depth interviews and observations are better suited (Bell et al, 2019).

Additionally, Bell et al. (2019) argue that face validity is particularly relevant when conducting quantitative studies. Face validity refers to the measures being used accurately reflects the core of the concept and thus mitigates the risk of spurious results (Bell et al., 2019). In terms of face

[9] The consequence of heteroscedasticity and autocorrelation is the same; the standard errors are incorrect which implies that all inferences being made are incorrect as well (Wooldridge, 2016). The use of White's robust standard errors is another way of calculating the standard errors which is made by the software and provides asymptotically consistent standard errors when the number of observations is large (Wooldridge, 2016).

validity, previous research has not agreed upon a standardized way to measure innovation, leading to the usage of a variety of proxies, including number of patents, percentage of sales coming from new products and R&D expenses (e.g. Chen et al., 2018; Miller & Triana, 2009; Rothaermel & Hess, 2007). While these are considered to be acceptable proxies for innovation, we used R&D expenses as a proxy for innovation as it has been the most extensively used proxy for innovation efforts (Chen et al., 2018; Dalziel et al., 2011; Rogers, 1998). A clear advantage of using R&D expenses over other proxies for innovation is its easiness to understand and the provision of a monetary figure to use in the following analysis (Rogers, 1998). However, a common critique towards using R&D expenses is its inability to capture all innovation efforts being made. According to the Swedish accounting and valuation principles, R&D spending is allowed to be accounted for as an intangible asset if the spending to a high probability will result in economic advantages (4 ch. 2 § ÅRL), which will hinder our ability to measure R&D spending because intangible assets normally are presented on an aggregate level. Thus, it may lie in the interest of managers to enter it as an intangible asset in order to level off R&D spending's impact on financial performance over several years rather than activate the entire cost immediately. However, since the requirements for entering R&D spending as an intangible asset are high, it is more likely that the majority of R&D spending will be shown in the income statement and therefore something we will capture with our measure of innovation. Another common critique towards R&D expenses as a proxy for innovation is that it only measures input while neglecting throughput and output (Richtnér et al., 2017; Rogers, 1998). In order to cope with this, several researchers have started to complement R&D expenses with an output measure such as the number of filed patents. However, the use of patent data does not necessarily improve the accuracy in measuring innovation as not all ideas are legally patentable and many firms choose to rely on secrecy rather than patents (Rogers, 1998). Additionally, heavy use of patents can signal that the firm has difficulties in appropriating the returns from its innovation efforts rather than being a more successful innovator (Toivanen, Stoneman & Bosworth, 2002).

Furthermore, as this paper is interested in understanding the relationship between board gender diversity and innovation, it was reasonable to use a proxy that is close to the board (in terms of possibility to affect the proxy). The relationship between boards and R&D expenses is promising as the board is able to exert influence (Dalziel et al., 2011; Zahra et al., 2000) and therefore the board's influence on key budgets such as R&D expenses requires attention. Arguably, a board is more able to exert influence on the input rather than the output of the

innovation process. Further, a firm's spending on R&D is an issue where managerial opportunism may come into play (Sanders & Carpenter, 2003; Shen & Gentry, 2012; Zahra, 1996) which implies that the board indeed has an important role to ensure that the management does not neglect it. Therefore, as R&D expenses are regarded as a suitable proxy for innovation, and being closely linked to the board, our operationalization is reaching a high degree of face validity.

3.5.2 Reliability

In order to improve the reliability of our study, we provide a clearly outlined and detailed description of the methodology used in order to make the study possible to replicate. More specifically, if researchers want to replicate our study, there may be a risk of defining and calculating key concepts differently. We mitigate this risk through explicitly providing the definitions being used in this paper as well as literature supporting each definition. With regard to the monetary variables (i.e. innovation, financial performance, and some of the control variables), we provide an exact description in Appendix A of how we calculated these using the Compustat items. Thereby, others can replicate our study using Compustat. In quantitative research, Bell et al. (2019) posit that inter-rater reliability refers to the problem of subjectivity when collecting data. While our study builds to a large extent on objectivity, on some occasions we had to make a judgement call when we were in doubt whether a particular board member was a male or a female. This occurred when the person in question had a gender-neutral or a foreign name and it was solved by searching for more information using mainly the company website. The use of secondary data from Compustat and *Directors and auditors* is strengthening the reliability of this study as they are considered to be reliable sources of information. However, when relying on secondary data, there is a risk of different interpretations of the data. We mitigate the risk of different interpretations by relying on manuals provided by Compustat and by clearly explaining the difficulties we interpreted when gathering the data in 3.2.3.

4 Results

4.1 Descriptive statistics and correlation matrix

4.1.1 Descriptive statistics

The descriptive statistics are presented in Table 4.1. Starting with the gender diversity variables, it can be stated that our sample consisted of both perfectly gender homogenous boards and perfectly gender heterogenous boards, as the values of the Blau and the Shannon indices cover the entire range of possible values they can attain. This implies that there are observations with only one gender present on the board and with an equal amount of both genders on the board. In detail, common to all observations that attain the minimum value (0) is that they consist of male directors. Further, while almost nine percent of the boards are perfectly gender homogenous, only three percent are perfectly gender heterogenous. Lastly, female directors are in majority in 2.4 percent of the observations, implying that the measure of these boards are not reaching its maximum value as perfect gender heterogeneity is not achieved.

The innovation variable indicates a large spread in its value range which we already highlighted for in the method section since a third of our observations had no R&D expenses reported. As a result, the standard deviation becomes relatively large and it is skewed towards zero as the mode of the variable is zero (unreported). It could be mentioned that the industry with the largest share of missing values on R&D expenses was the consumer services where 81 percent of the companies lacked data on R&D expenses (unreported).

The sample mean of Tobin's Q is 1.83, which is somewhat higher compared to the mean of Campbell and Mínguez-Vera (2008) who reported a mean of 1.6 but below the mean of Chen et al. (2018) at 2.03. The equilibrium value of Tobin's Q is at one, hence, as a couple of observations had a value above eight this indicates some outliers. As Wooldridge (2016) states, there is no formal definition of an outlier but rather outliers should be identified by looking for unusual observation values. However, the exclusion of these did not change the main results in any significant way. In comparison, the accounting-based performance measure, ROA, has a handful of negative outliers with a ROA below negative 100 percent. As explained in 3.2.3, all of these observations belong to the healthcare industry and the extreme values are due to being in a phase of research with limited revenue streams. The descriptive statistics of ROA are similar to the ones obtained by Adams and Ferreira (2009) who did not either adjust for outliers.

In addition to the presented descriptive statistics above, we collected data on the gender of the CEO as well as whether the CEO was present on the board. We find it noteworthy that among our sample, the CEO was a part of the board in 45 percent of the observations but in all of the cases the CEO was a man. Hence, if the CEO was a woman, she was never a director of the board, which contradicts the result of Matsa and Miller (2011) stating that female CEOs and their presence as directors are highly correlated.

Table 4.1 Descriptive statistics

Variables	Mean	Median	Maximum	Minimum	Standard deviation	Observations
Independent variables						
Blau index	0.33	0.355	0.5	0	0.135	826
Shannon index	0.495	0.54	0.693	0	0.184	826
Mediating variable						
Innovation	12.27	16.852	24.315	0	8.747	826
Dependent variables						
Tobin's Q	1.832	1.358	15.534	0.487	1.537	826
ROA[a]	0.049	0.076	0.849	–1.605	0.179	826
Control variables						
Board size	7.92	8	15	3	2.309	826
Firm size	21.704	21.312	26.671	17.342	2.089	826
Leverage[a]	0.191	0.176	0.892	0	0.157	826
Independent directorship[a]	0.526	0.5	1	0	0.216	826

[a] In decimal format

4.1.2 Correlation matrix

The correlations in Table 4.2 are based on Pearson correlations and we calculated the significance level of the correlations with EViews' covariance analysis tool. It should be noted that calculation of correlations disregards the time dimension as it assumes all observations are cross-sectional (Wooldridge, 2016). Hence, the panel data analysis which is the main focus of our results will provide a more accurate description. Additionally, correlation tells nothing about the direction of a relationship; it only indicates whether there is a relationship or not, making it to an insufficient technique to prove causality (Bell et al, 2019).

The results from the correlation analysis have several points to note. First, it is logical that the Blau and the Shannon indices indicate an almost perfect positive correlation; the two measures of board gender diversity are applied as they should reflect the inherent variation within a group

of people. In fact, the very strong correlation indeed indicates that this is the case. Second, insignificant and weak negative correlations between the gender diversity variables and innovation were found, telling us that the hypothesized positive effect of board gender diversity and innovation may not hold. Third, while both the Blau and the Shannon indices have highly significant correlations to ROA, the correlations to Tobin's Q are both less positive and only significant for the Blau index. This indicates that gender diverse boards are more likely to co-vary with higher ROA whereas we cannot be as confident with regard to Tobin's Q. Fourth, the negative correlation between Tobin's Q and ROA is remarkable as it is reasonable to expect that profitable companies should be higher valued by the market. Several studies have examined that these performance measures are statistically positively related (e.g. Campbell & Mínguez-Vera, 2008; Carter et al., 2003; Carter et al., 2010) which undoubtedly makes this finding noteworthy. We find it difficult to explain this and the idea that there may be some time lag between the two could not be supported; we ran a regression by lagging ROA since it could be the case that the effect of ROA will have a lagged effect on Tobin's Q (unreported). However, we applied Tobin's Q and ROA as two measures of financial performance due to their ability to cover different aspects of financial performance (i.e. growth opportunities and past performance, respectively) and this negative correlation indeed supports this attempt – even if we did not foresee this direct opposite relationship.[10] Fifth, the correlations between innovation and financial performance are equivocal; innovation seems to be positively correlated with market performance while negatively correlated to accounting-based performance. This could once again be a consequence of the time delay; spending on R&D affects current year's accounting measures but the potential positive effects may be delayed, and since the correlation calculation disregards the time dimension, this effect is overlooked. Sixth, both board size and independent directorship indicate relatively high positive correlations to board gender diversity, which supports previous research stating that larger boards tend to include more women (Bear et al., 2010; Carter et al., 2003) and that female directors are more likely to be independent directors (Adams & Ferreira, 2009; Carter et al., 2010; Simpson et al., 2010). Additionally, it is interesting to note that even firm size is positively correlated to board gender diversity. This confirms the overall pattern in our population, Fristedt and Sundqvist (2008, 2009) and Fristedt et al. (2010–2014) show that the largest companies on the OMXS have a higher degree of board gender diversity than the others. Last, the control variables' sign of correlation is laterally

[10] A mathematical oriented explanation of this is that the partial derivative with respect to total assets is positive for Tobin's Q while it is negative for ROA, meaning that Tobin's Q and ROA respond differently of an increase in total assets. Thus, this could partly explain the negative correlation we found.

reversed for Tobin's and ROA; board size, leverage and firm size are negatively correlated to Tobin's Q while positively correlated to ROA, and the reversed is true for independent directorship. This could be due to Tobin's Q and ROA being negatively correlated.

On an overall basis, what is striking about the absolute size of the correlation coefficients is that only two of them can be seen as strong, meaning a coefficient value above/below ± 0.70 (Bell et al., 2019). These two correlations are between the abovementioned Blau and Shannon indices, and board size and firm size. The latter correlation indicates that larger firms tend to have more members on the board. This could partly be explained by that the possibility to appoint employee directors of a board depends on the size of the firm. Consequently, as the majority of the correlations are weak, the issue of multicollinearity should not evoke any severe complications to achieve *ceteris paribus* since the independent variables do not seem to strongly co-vary (Hair et al., 2014; Wooldridge, 2016). Furthermore, a large share of the correlations is highly significant which partly depends on the autocorrelation between the same company's observation over time (Wooldridge, 2016).

Table 4.2 Correlation Matrix

Variables	*Blau*	*Shannon*	*Innovation*	*Tobin's Q*	*ROA*	*Board size*	*Firm size*	*ID*	*Leverage*
Blau	1								
Shannon	0.991***	1							
Innovation	–0.049	–0.031	1						
Tobin's Q	0.058*	0.049	0.158***	1					
ROA	0.091***	0.095***	–0.075**	–0.205***	1				
Board size	0.242***	0.273***	0.055	–0.169***	0.127***	1			
Firm size	0.178***	0.203***	0.101***	–0.289***	0.317***	0.74***	1		
ID	0.189***	0.18***	0.174***	0.171***	–0.084**	–0.261***	–0.184***	1	
Leverage	0.059***	0.075**	–0.045	–0.214***	0.037	0.221***	0.387***	–0.122***	1

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

ID = Independent directorship

4.2 Multiple regression analysis results

For each regression, we ran two complementary regressions with different designs; one where we applied pooled OLS as described in 3.4.2 in order to demonstrate the effect unobservable characteristics may have on our results (referred to as the pooled model), and one where we controlled for reversed causality by running the regression "in reverse", i.e. change place of the dependent and the independent variable (the reversed model). Additionally, for the regressions including the innovation variable, we ran the correct model in two different ways; one where we interpreted missing values as zero R&D expenses, and the other where we excluded

observations with missing values. Apart from presenting the coefficient values and standard errors of the variables, we include the Chi-Square statistic of the Hausman test, the adjusted R^2 and the F-statistic for each regression as well. The adjusted R^2 should be interpreted as the fraction of variation in the dependent variable that can be explained by the explanatory variables with respect to the number of explanatory variables included in the regression, and the F-statistic whether all explanatory variables are jointly insignificant (Wooldridge, 2016).[11] The purpose of the Hausman test is explained in 3.4.1.

As explained in 3.3.1, we have two overlapping measures of board gender diversity and they are never jointly included in the regressions. On an overall basis, the sign of the gender diversity coefficient and the significance level are similar regardless of whether the Blau or the Shannon indices are used. Therefore, we will refer to the gender diversity variable when there are no remarkable differences between the two measures, and in order to make legible tables, we present the regressions with the Blau index in the following sections while the corresponding regressions with the Shannon index instead are found in Appendix D (see Tables D.1–D3).

4.2.1 Board gender diversity and innovation

The results of R1 are presented in Table 4.3. In the first column, we can observe that the effect of board gender diversity is highly significant with a strong negative impact on innovation in the pooled model. The level of explained variation in R&D expenses is approximately six percent which should be compared to the adjusted R^2 in the correct model later on. Indeed, as the coefficient value is very high in absolute terms for both measures of gender diversity, it is reasonable to expect that this effect is driven by omitted variables which justifies the inclusion of fixed effects.

When controlling for unobserved firm and year specific effects, i.e. running the correct model, board gender diversity switches sign and becomes insignificant as seen in the second column. This indicates that the strong negative relationship in the pooled model is driven by omitted firm and year specific factors. Consequently, the lack of a causal effect of board gender diversity on innovation implies that H1 is rejected. Taking into account the mediation technique, this implies that we cannot prove that a mediating relationship of innovation between

[11] The difference between R^2 and adjusted R^2 is the consideration of added explanatory variables to the model which, regardless of its relevance, will never decrease R^2 (Wooldridge, 2016). Hence, adjusted R^2 is a more suitable measure of the explained variation for multiple regressions (Wooldridge, 2016).

board gender diversity and financial performance exists; if there is no causal relationship between the independent and the mediating variable, it is impossible for the mediating variable to mediate the independent variable by definition. Similarly, only one control variable remains significant in the correct model, which is firm size indicating that larger firms tend to spend more on R&D. As expected, the adjusted R^2 increased to 85 percent; a natural consequence of including one company unique and one year unique variable. Importantly, the F-statistic is highly significant meaning that the likelihood that all explanatory variables simultaneously have no effect on innovation is very low.

Even though we could not prove any causality from board gender diversity to innovation, we still performed the reversed regression. The result, as stated in the fourth column, means that we could neither find a statistically significant reversed relationship between the variables. It should be noted that independent directorship and leverage have strong significant effects on board gender diversity, but as the adjusted R^2 in a pooled reversed model is only 0.05 (unreported), there is low risk of high multicollinearity.

To compare the difference between equalizing missing values with zeros and excluding missing values (see 3.2.3), we found that this has limited impact on our results in the correct model as presented in the third column. However, it is worth noting that the p-value of board gender diversity slightly increases (i.e. becomes less close to being significant given a significance level of ten percent) while the possible effect on innovation decreases (i.e. the coefficient value is lower). We could therefore conclude that based on our sample, the method of setting missing values equal to zero, did not severely bias our results in the first hypothesis with respect to statistical conclusions – the economic relevance is indeed a bit different as it indicates that board gender diversity's effect in the later model is smaller and on a more reasonable level. The adjusted R^2 is extremely high in this regression (0.96), which should be questioned since it seems unlikely that board gender diversity and the four control variables explain almost all variation in companies' reported R&D expenses.

Table 4.3 Regression 1 (R1)

Explanatory variables	***Innovation***			***Board gender diversity***
	Pooled OLS	*Correct model*	*No missing values*	*Reversed model*
Board size	0.14645	0.012	0.063**	0.001
	(0.212)	(0.185)	(0.025)	(0.005)
Firm size	0.71***	1.786**	0.445***	0.02
	(0.238)	(0.503)	(0.077)	(0.013)
Independent directorship	9.299***	0.86	0.023	0.11***
	(1.372)	(0.96)	(0.115)	(0.027)
Leverage	–4.647**	0.282	–0.125	–0.105**
	(2.124)	(1.898)	(0.219)	(0.043)
Blau	–8.17***	1.979	0.195	
	(2.303)	(1.919)	(0.222)	
Innovation				0.0005
				(0.001)
Number of observations	826	826	555	826
Adjusted R^2	0.063	0.85	0.964	0.698
Hausman test		16.115***	20.245***	18.529**
F-test	12.186***	37.49***	151.428***	15.899***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

4.2.2 R2: Board gender diversity and financial performance

The results of R2 are presented in Table 4.4. The pooled model for testing the effect of board gender diversity on financial performance showed that with regard to market performance, board gender diversity had a positive significant effect on Tobin's Q. The adjusted R^2 indicates that approximately 11 percent of the variation in Tobin's Q is explained by the explanatory variables, where all of them are significant (except for board size when Tobin's Q is regressed on the Shannon index). Similarly, the effect of gender diversity on accounting-based performance is also positive significant and all control variables are significant as well. Interestingly, all control variables, except for leverage, have the opposite sign compared to when Tobin's Q is the dependent variable. Bearing in mind that Tobin's Q and ROA were negatively correlated, this may not be too surprising; with negative correlation it is reasonable to expect that some variables will change sign depending on which performance measure is used. The adjusted R^2 is slightly higher for ROA compared to Tobin's Q.

Conducting the correct model with fixed effects leads to somewhat different results, as presented in the second column named "Correct model". While we cannot prove that the Blau

index has any statistically significant effect on Tobin's Q, we found that it instead has a positive effect on ROA. At the same time, the Shannon index lacks significant effects on both Tobin's Q and ROA. However, in common for all of these cases is that the gender diversity variables possess a positive sign. The economic relevance of the (positive) significant relationship found between the Blau index and ROA should be highlighted; the estimated effect of a tenth unit increase in Blau translates into an expected increase in ROA with approximately 0.9 percent. This effect is arguably not that large and it should be kept in mind when interpreting this finding. With regard to the control variables, only firm size is significant in the model with Tobin's Q while firm size and leverage are positive and negative, respectively, significant on ROA. This shows that the correlations in Table 4.2 for firm size held even when taking account for the time dimension, as the sign of the correlations and the sign of the coefficients are the same, while the positive correlation between leverage and ROA turned out to a statistically significant negative relationship in a panel data analysis. Just as in the pooled model, most of the explanatory variables change sign when changing the dependent variable. The adjusted R^2 lies just over 70 percent compared to the adjusted R^2 in the pooled model of 11 to 14 percent. Regarding our hypotheses, since H1 was rejected in R1 it was already clear that H3 would be rejected. However, if H1 instead should not have been rejected, then the lack of positive significance in this regression would nevertheless imply that H3 would have been rejected.

Controlling for reversed causality seen in the third column, we notice that the detected positive relationship between the Blau index and ROA was not robust enough for a causal relationship: ROA has a positive significant effect on the Blau index as well. This could be interpreted as when companies are increasing their accounting-based performance through ROA, they are more likely to increase the gender heterogeneity of board members as well.

As innovation is not a part of this regression, we will not discuss the implications of excluding the observations with missing value on R&D expenses.

Table 4.4 Regression 2 (R2)

Explanatory variables	***Financial performance***				***Board gender diversity***	
	Pooled OLS		*Correct model*		*Reversed model*	
	Tobin's Q	ROA	Tobin's Q	ROA	Tobin's Q	ROA
Board size	0.063*	–0.023***	–0.04	–0.007	0.001	0.002
	(0.038)	(0.006)	(0.042)	(0.006)	(0.005)	(0.005)
Firm size	–0.229***	0.048***	–0.365*	0.01***	0.022*	0.016
	(0.056)	(0.007)	(0.221)	(0.026)	(0.013)	(0.013)
Independent directorship	0.793***	–0.078**	–0.05	–0.022	0.111***	0.112***
	(0.233)	(0.032)	(0.216)	(0.028)	(0.027)	(0.027)
Leverage	–1.027**	–0.146***	0.723	–0.272***	–0.105**	–0.089**
	(0.479)	(0.035)	(0.511)	(0.053)	(0.044)	(0.044)
Blau	0.859**	0.12**	0.047	0.088*		
	(0.39)	(0.049)	(0.491)	(0.053)		
Financial performance					–0.0004	0.055*
					(0.004)	(0.033)
Number of observations	826	826	826	826	825	826
Adjusted R^2	0.114	0.144	0.709	0.724	0.697	0.699
Hausman test			12.96**	24.054***	14.22**	12.941**
F-test	22.285***	28.84***	16.736***	17.924***	15.859***	15.962***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

4.2.3 R3: Board gender diversity, innovation and financial performance

With regard to that both R1 and R2 failed to prove significant relationships, we will go through the results of R3 more briefly as the main purpose of this regression was to examine the strength in the mediating relationship. As explained earlier, the purpose of including innovation in R3 was to function as a control variable and since we added a fourth regression, explicitly examining H2, we will present the results of the relationship between innovation and financial performance in 4.2.4. Nevertheless, the results of R3 are presented in table 4.5.

Just as in the pooled model in R^2, board gender diversity has a positive significant effect on both financial performance measures even when controlling for innovation as indicated in the first column. Interestingly, the coefficient value is somewhat higher which could be due to their negative correlation as the inclusion of innovation implies that we can keep innovation fixed when testing board gender diversity's effect on financial performance and not bias the effect downwards. The adjusted R^2 is also larger.

Performing the correct model in the second column, there are only minor changes in the effect and the p-values when controlling for innovation compared to R2. The most surprising aspect of the results may be that the adjusted R^2 marginally decreases. Thus, we therefore excluded each control variable one at a time to see whether this happened for the other variables as well. By doing so, we found that exclusion of firm size and leverage, respectively, led to a lower adjusted R^2 while exclusion of board size and fraction of independent directors, respectively, increased the adjusted R^2. Before making any conclusions from this, we want to point out that the difference is on the fourth decimal of the adjusted R^2-number. However, an increase in the adjusted R^2 when excluding a control variable can be interpreted as it does not contribute to explain the variation in the financial performance variable with respect to the effect of adding more explanatory variables. However, it is important to remember that the specific effects already explain a large share of the variation which could make it harder for more control variables to add new unique information to explain some further variation. The results of these exclusions of control variables are not presented in any table.

Table 4.5 Regression 3 (R3)

Explanatory variables	***Financial performance***			
	Pooled OLS		*Correct model*	
	Tobin's Q	ROA	Tobin's Q	ROA
Board size	0.059	–0.023***	–0.04	–0.007
	(0.037)	(0.005)	(0.042)	(0.006)
Firm size	–0.251***	0.049***	–0.378*	0.1***
	(0.058)	(0.007)	(0.224)	(0.026)
Independent directorship	0.505**	–0.057**	–0.056	–0.022
	(0.243)	(0.031)	(0.216)	(0.028)
Leverage	–0.883*	–0.157***	0.72	–0.272***
	(0.487)	(0.034)	(0.51)	(0.053)
Innovation	0.031***	–0.002***	0.007	0.0001
	(0.006)	(0.001)	(0.005)	(0.001)
Blau	1.112***	0.102**	0.033	0.088*
	(0.388)	(0.048)	(0.49)	(0.053)
Number of observations	826	826	826	826
Adjusted R^2	0.142	0.154	0.709	0.724
Hausman test			14.878**	25.799***
F-test	23.813***	26.063***	16.602***	17.76***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 6)

4.2.4 R4: Innovation and financial performance

The results of R4 are presented in Table 4.6 and starting with the pooled model (see the first column), we once again notice that the sign of the independent variable, in this case innovation, changes depending on which financial performance measure is used. More specifically, innovation has a positive effect on Tobin's Q but a negative effect on ROA, highly significant in both cases. Furthermore, the explained variations in the pooled regressions are slightly higher compared to the ones obtained in R1 when we regressed financial performance on gender diversity with the same control variables.

The results from running the correct model, see second column, indicate insignificant relationships on both financial performance measures. However, the p-values of innovation are remarkably different; when regressed with Tobin's Q, innovation is almost significant at the ten percent level (p-value = 0.11), but the p-value is close to one when ROA is regressed on innovation. As the coefficient value of innovation in the latter case is tremendously small, the following p-value will therefore be very high. Consequently, this made us consider whether there is some specification error in the model due to the very low estimated effect of innovation on ROA. We excluded one control variable at a time in order to investigate whether any of the control variables significantly affected the innovation variable. Interestingly, when excluding firm size, innovation becomes positive significant with a much larger effect on ROA compared to when firm size was included – a consequence we do not find when doing the same thing with Tobin's Q as dependent variable. Two possible explanations to this are the presence of collinearity or multicollinearity (Wooldridge, 2016). However, as the correlation between innovation and firm size was weak (0.1), albeit significant, this indicates low degree of collinearity. Additionally, the tendency for multicollinearity among all control variables on innovation was modest when we ran that as a regression. None of these extra regressions are reported in any table. In total, this implies that we cannot fully explain why this change in significance level occurs. Hence, based on the precautionary principle, we abstain from making any conclusions on the relationship between innovation and ROA from this model. However, as innovation was insignificant on Tobin's Q, although with a positive sign, we reject H2 with regard to market performance.

Conducting the regressions in reverse, displayed in the fourth column, does not contribute with any additional information compared to the correct model; Tobin's Q is close to be significant on innovation (p-value = 0.11) and ROA is just as insignificant as innovation was.

Lastly, the results when restricting our sample to observations with no missing values on R&D expenses are striking. These results, seen in the third column, show that innovation indeed has a significant effect on both Tobin's Q (positive effect) and ROA (negative effect) which could therefore imply that H2 cannot be rejected when excluding missing values. However, the reversed causality model for this setup (unreported) indicates that both Tobin's Q and ROA have significant effects on innovation as well. This means that rather than increased innovation positively (negatively) affects Tobin's Q (ROA), it could also be the case that companies with high Tobin's Q (ROA) are more likely to increase (decrease) innovation. Thus, no clear evidence of a causal relationship exists such that H2 needs to be rejected for this sample too. Lastly, compared to R1 when restricting the sample to observations with no missing values on R&D expenses, the adjusted R^2 is still high (around 0.7) but on a more reasonable level compared to the adjusted R^2 in R1 (0.96) with respect to the specific effects' explanation power of the variation.

Table 4.6 Regression 4 (R4)

Explanatory variables	***Financial performance***						***Innovation***	
	Pooled OLS		*Correct model*		*No missing values*		*Reversed model*	
	Tobin's Q	ROA	Tobin's Q	ROA	Tobin's Q	ROA	Tobin's Q	ROA
Board size	0.08**	–0.021***	–0.04	–0.007	–0.074	–0.004	0.020	0.015
	(0.038)	(0.005)	(0.041)	(0.006)	(0.059)	(0.007)	(0.185)	(0.185)
Firm size	–0.251***	0.049***	–0.377*	0.102***	–0.626**	0.153***	1.871***	1.835***
	(0.058)	(0.007)	(0.22)	(0.026)	(0.301)	(0.037)	(0.508)	(0.527)
Independent directorship	0.711***	–0.038	–0.053	–0.013	–0.194	–0.016	1.085	1.079
	(0.244)	(0.029)	(0.214)	(0.027)	(0.336)	(0.039)	(0.972)	(0.975)
Leverage	–0.864*	–0.155***	0.717	–0.281***	0.639	–0.216***	–0.013	0.061
	(0.496)	(0.034)	(0.498)	(0.052)	(0.742)	(0.061)	(1.897)	(1.9)
Innovation	0.029***	–0.002***	0.007	–0.00004	0.222**	–0.056***		
	(0.006)	(0.001)	(0.005)	(0.001)	(0.112)	(0.017)		
Financial performance							0.124	–0.051
							(0.078)	(0.805)
Number of observations	826	826	826	826	555	555	826	826
Adjusted R^2	0.135	0.15	0.71	0.723	0.671	0.73	0.85	0.85
Hausman test			13.402**	27.17***	4.105[a]	24.672***	14.689**	14.401**
F-test	26.764***	30.126***	16.756***	17.81***	12.552***	16.323***	37.448***	37.409***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

[a] In order to maintain consistency, we decided to use fixed effects for this regression despite that the Hausman test indicated that fixed effects were not the most efficient choice for this particular regression.

4.3 Hypotheses results

A summary of the outcome of our hypotheses is displayed in Table 4.7. The decisions whether the hypotheses were rejected or not hinge on the results of the correct model and eventually the reversed model if a positive significant relationship was found in first place. Additionally, for the regressions where innovation was studied, a complementary statement about the hypotheses from the model with no missing values is made. The first hypothesis was rejected as we could not find any positive significance of board gender diversity on innovation, which also held under the restricted sample with no missing values. The second hypothesis was also rejected since we could not find any positive statistical relationship between innovation and financial performance.[12] However, when we excluded observations with missing values, a positive significant relationship between innovation and market performance emerged, but as the reversed model indicated that market performance also had an effect on innovation, the causal relationship could not be statistically proved. Hence, H2 was rejected for the restricted sample as well. Taken together, the rejection of H1 and H2 therefore must imply that H3 is rejected as well; if board gender diversity lacks a causal effect on innovation (H1) and/or innovation lacks a causal effect on financial performance (H2), then it is impossible for innovation to mediate a causal relationship of board gender diversity and financial performance. This was further strengthened as we did not find any significant positive effect of board gender diversity on financial performance that was robust for reversed causality.

Table 4.7 Summary of hypotheses outcomes

Hypothesis	Regression	Outcome
1		
Increased board gender diversity will have a positive causal effect on innovation.	R1	Rejected
2		
Increased innovation will have a positive causal effect on financial performance in terms of accounting returns and market performance.	R4	Rejected
3		
Innovation positively mediates the causally indirect relationship between board gender diversity and financial performance.	R1, R2, R3	Rejected

[12] The most correctly would be to say that we reject H2 with respect to market performance but not to accounting-based performance since the regression with the latter measure indicated ambiguous results when firm size was excluded as a control variable.

5 Discussion

Relating back to the literature review, this paper is founded in the AT and RDT suggesting that board capital (i.e. human and social capital) will influence how board members make decisions and how they behave as resource providers and monitors. Further, it is posited that board capital varies within demographic characteristics such as gender. In reviewing the literature on RDT, we noted that gender heterogeneous boards benefit from additional resources as females tend to hold more advanced degrees, have better knowledge in marketing and sales, and come from different occupational backgrounds. As a result of these additional resources, the inclusion of females (i.e. becoming more gender heterogeneous) provides a more exhaustive coverage of available information which leads to a higher proficiency in decision-making and problem-solving. Additionally, the literature on AT posits that monitoring will improve in gender heterogeneous boards as female directors increase the range of professional experiences and education, are more likely to sit on auditing committees, ask more critical questions and have a higher rate of board meeting attendance which also increases the attendance of their male counterparts. In the following, we will discuss our empirical results in comparison to previous research.

5.1 H1: Board gender diversity and innovation

In reviewing the literature, it was stated that diversity tends to increase the number of ideas and promote creativity which arguably aids innovation. More specifically, diverse ideas will have a positive impact on the identification, development and selection of firm decisions. Similarly, having improved coverage of information will improve the inputs in the innovation process as innovation work implies a shortage of information. Additionally, the concept of the career concern model suggests that managers are concerned about their careers and therefore are reluctant to undertake risky innovation activities in fear of being seen as inferior managers. This fear is minimized if the manager in question is comfortable that the board will be able to separate managerial abilities from purely ambiguous reasons for a failed innovation project. Similarly, the quiet life model posits that managers are reluctant to participate in innovation activities as it is likely to result in new routines and discomfort which subsequently leads to managerial shirking. However, with a strong monitoring function, managerial shirking is reduced, and innovation can be benefitted.

Based on the abovementioned, we formulated our first hypothesis stating that board gender diversity will have a positive causal effect on innovation. As we had to reject this hypothesis based on our data analysis, we provide a contradicting finding to previous research on board gender diversity and innovation. The two previous research papers investigating innovation as a mediating variable between board gender diversity and financial performance found positive significant relationships in the first regression (i.e. between board gender diversity and innovation) (Chen et al., 2018; Miller & Triana, 2009). In terms of theoretical viewpoints, Chen et al. (2018) deduct their hypotheses from the AT, similar to this paper, but limit themselves to one theory. On the contrary, Miller and Triana (2009) adopt two distinctly different theories; the signalling theory and the behavioral theory of the firm. While the theoretical viewpoints slightly differ between Miller and Triana (2009), Chen et al. (2018) and this paper, the same arguments for constructing H1 can be found in all three papers. This makes us question why we find contradicting results and we offer three main explanations for this. First, we use a different set of sample data; Miller and Triana (2009) study the 500 largest firms in the US which arguably yield different results as firm size is closely related to innovation (something that was shown in our data analysis as well), and even though Chen et al. (2018) study a wider array of companies, they are also focused on the US. Second, while Miller and Triana (2009) adopt a cross-sectional approach, we use panel data similar to Chen et al. (2018). However, while Chen et al. (2018) investigated a time period of 1998–2006, we use a later period of 2008–2014 which also is likely to affect the results as the number of female directors increased by approximately five percent per year in our population (Fristedt & Sundqvist, 2008; Fristedt et al., 2014).[13] Third, and perhaps most important, the operationalization of variables differs significantly. While Miller and Triana (2009) operationalize innovation as R&D intensity in relation to industry, Chen et al. (2018) operationalize innovation as R&D expenses, number of citations and number of patents. Additionally, Chen et al. (2018) operationalize the independent variable as the fraction of females while Miller and Triana (2009) use the Blau index. The result of different operationalizations will undeniably provide inconsistent results; for example, the fraction of female directors and the Blau index investigate two distinctly different things; female board directors' effect and board gender heterogeneity's effect, respectively, on firm outcomes.

[13] The fraction of female directors in 2008 was 18.5 % and 25.8 % in 2014. This translates into an yearly average increase (x) of: $0.185 * x^7 = 0.258 \rightarrow x \approx 5\ (4.87)\ \%\ average\ increase\ per\ year$.

Regardless of the previous papers provided by Miller and Triana (2009) and Chen et al. (2018), we find several arguments why we did not find a significant relationship between board gender diversity and innovation. First, Adams and Ferreira (2009) posit that greater diversity is associated with a higher risk of conflicts and therefore provides slower decision-making. As innovation often implies being a first mover, slow decision-making will negatively impact innovation performance. This is strengthened by Nielsen and Huse (2010) who state that while gender balanced boards are better able to make high quality decisions, the cost of this often entails slower decision-making. Second, Brunzell and Liljeblom (2014) state that as females have less experience in board work, the inclusion of females can actually lead to lower board efficiency as the lack of experience impacts their skills as board members. However, this can also be due to male directors not knowing how to utilize the unique resources females bring, as they are marginalized. Last, Randøy et al. (2006) state that it is difficult to establish the effect board composition may have as there are several other factors that will arguably have an impact. While we have tried to cover these factors using control variables, one needs to understand that it is unlikely to capture all factors that might have an effect on the relationship between board gender diversity and innovation as there are several sources of noise that hinder the possibility to capture any potential effect. One source of noise that we must accept with this research design is the existence of many steps from board decision to innovation investment, even if we consistently argue that the causal link between board composition and R&D expenses is both theoretically and intuitively relevant. Similarly, Østergaard et al. (2011) state that innovation often requires interaction between several groups of people within an organization and therefore it can be untruthful to merely look at the composition of the board when one should look at the overall composition of the entire firm.

5.2 H2: Innovation and financial performance

There is a strong relationship between innovation and financial performance in previous studies where innovation is seen as the foundation for firm survival and is considered to be a crucial activity in order to enhance firm performance. The literature suggests that firms pursuing innovation activities yield higher margins and increasing market shares, are more proactive, and achieve overall better performance. Furthermore, it is stated that firms that are able to identify new opportunities and organize thereafter are performing better and possess resilience when the external environment changes.

It was hypothesized that increased innovation will have a positive causal effect on financial performance. We had to reject this hypothesis and contradict Miller and Triana (2009) and Chen et al. (2018) who found positive significant relationships. The contradicting results may be caused by different operationalizations of the variables. For example, Miller and Triana (2009) operationalize financial performance as return on investment and return on sales. Additionally, while Miller and Triana (2009) and Chen et al. (2018) set missing values equal to zero for the innovation variable, they do not provide information regarding how many observations they had with missing values. In our case, as a third of our observations had missing values on reported R&D expenses, we conducted two analyses; one where we set missing values equal to zero and one where the observations with missing values were excluded. This yielded two distinctly different results that we find important to emphasize; in the first case we found no significance while the second case indicated a positive significant relationship between innovation and Tobin's Q and a negative significant relationship between innovation and ROA. However, these relationships were subject to reversed causality, leading us to reject H2 for this case as well since we could not prove that the effect of innovation on financial performance was unidirectional.

While Tobin's Q and ROA reflect different aspects of financial performance, we find it worthy to discuss how innovation is positively related to Tobin's Q while negatively related to ROA when excluding observations with missing values. Shen and Gentry (2014) provide an explanation where they state that R&D expenses usually exhibit a low success rate in developing new technologies and even if it becomes successful, the financial benefits tend to be distant and uncertain. As ROA is a measure of past performance, it is likely that R&D expenses decrease the short-term financial performance. On the contrary, higher spending on R&D can send signals to the market that new innovations are forthcoming and therefore the future earnings of the firm are expected to increase. Similarly, Dalziel et al. (2011) state that R&D investments are risky and can dampen short-term performance. With the intention to account for this, we ran an additional regression where we considered a time lag of innovation on ROA (unreported), but we could not support the arguments provided by Shen and Gentry (2014) and Dalziel et al. (2011) as the lagged innovation variable was still insignificant (and negative). Additionally, it is important to highlight that one of the weaknesses with R&D expenses as a measure of innovation is that it does not reflect the actual outcome of innovation activities, but rather the focus on innovation. Therefore, it is reasonable to assume that H2 might

have yielded a different outcome if we would have access to data on an output measure on innovation, similar to Chen et al. (2018).

5.3 H3: Board gender diversity, innovation and financial performance

As mentioned in the literature review, there has been contradicting findings in previous research regarding board gender diversity and financial performance, indicating a complex and indirect relationship. The mediation technique provided by Baron and Kenny (1986) and used in this paper, partly counters this aspect and provides a more robust explanation of a possible relationship. The foundation of H3 was primarily based on two papers (Chen et al., 2018; Miller & Triana, 2009) due to the scarce number of previous studies examining innovation as a mediating variable between board gender diversity and financial performance. As H3 is dependent on the outcome in H1 and theoretically and intuitively connected to H2, we formulated H3 on the same arguments being used in H1 and H2. Additionally, as Chen et al. (2018) found that innovation indeed mediates the relationship between board gender diversity and financial performance, we felt confident to use this result as a basis for H3.

As a result, we hypothesized that innovation positively mediates the causal relationship between board gender diversity and financial performance. As we had to reject our hypothesis, we provide findings similar to Miller and Triana (2009) but contradicting to Chen et al. (2018). While we knew that rejecting H1 would lead to the rejection of H3, we continued with all regressions in order to provide transparency in the paper as well as to see if we could find any interesting results. The relationship of interest in this section is the one between board gender diversity and financial performance (R2). Similar to Miller and Triana (2009), we were unable to sustain a positive significant relationship between board gender diversity and financial performance. While we found that the Blau index had a positive significant effect on ROA, ROA also showed to have a positive significant effect on the Blau index. In addition, no significant effects were found when Tobin's Q was regressed on either board gender diversity indices. Followingly, we had to reject the hypothesis that increased board gender diversity has a causal effect on financial performance. We present three main reasons for this. First, Post and Byron (2015) argue that the relationship between board gender diversity and financial performance is not linear but rather it should be characterized as a reversed U-shaped relationship where too much monitoring and strategy involvement can become costly and unproductive at some point. Similarly, Adams and Ferreira (2009) posit that gender diverse

boards generally act as tougher monitors and that this can both create and destroy shareholder value. Firms with already strong monitoring functions are likely to destroy shareholder value whereas firms with weaker monitoring functions can instead increase shareholder value through the inclusion of females. Subsequently, it is plausible that the inclusion of females has less of an effect on monitoring functions in Sweden due to the relatively high shareholder protection rights in Sweden (the World Bank, 2014). This speculation is further strengthened as two other studies on the relationship between board gender diversity and financial performance in a Swedish context found no relationship (Randøy et al., 2006) or even a negative one (Eklund et al., 2009). Second, Rose (2007) posits that since males are considered to be the stereotypical board members, there might be a case of assimilation where female directors adopt the norms and behaviors of their male counterparts in order to be truly accepted and considered to be qualified as board members. However, Randøy et al. (2006) found that male directors in majority did not affect minorities' likelihood to assimilate to existing norms. Interestingly, these studies examine a Scandinavian context but still provide contradicting results. One possible explanation of this may be due to different board cultures, which affect the degree of assimilation of board minorities (Miller & Triana, 2009; Torchia et al., 2018). Third, while the above sections are derived from the belief that there are relevant differences between gender and its accompanying attributes for board performance, we find it reasonable to shortly discuss this from another starting point. As our results indicate, we are unable to impute gender any discrete attributes which implies that differences in gender may not be essential for board activities, making the degree of board gender diversity irrelevant. Rather it could be viewed from an individual perspective where individual attributes, randomly distributed across gender, have a larger impact on board activities than gender attributes and subsequently gender diversity. Followingly, it may be more fruitful to focus on the actual differences between female and male directors as gender is multi-faceted (Torchia et al., 2018). However, as presented in chapter 2, differences between gender have been confirmed by other studies in a board setting which indeed supports the fundamental idea that board gender diversity indeed is relevant to consider for board performance. Last, as suggested by Randøy et al. (2006), it can simply be the case that board composition does not matter much for firm performance but rather for individual board task performance and therefore it can be difficult to find relationships between the board and the outcome of a firm.

5.4 Subanalysis

As the main focus of the first subanalysis was to test the idea that board gender diversity has a particular strong effect in innovation intensive industries, we constrained this subanalysis to R1 and R4 which involved innovation to a high degree (Tables C.1 and C.2 in Appendix C). Overall, the results for both industries indicated that we could not prove the same effect as Chen et al. (2018); neither board gender diversity in R1 nor innovation in R4 had a significant positive effect on innovation and financial performance, respectively, in neither the health care nor the technology industry. For the technology industry, we even found a small negative relationship between innovation and ROA – just as we did in the main regressions with no missing values. In addition, even though our measures of board gender diversity had a significant negative effect on innovation in R1 for technology companies, we are cautious about making any inferences from this due to the very large coefficient values (in absolute terms).

Our second subanalysis focused on investigating the notion of critical mass (see Tables C.3).[14] Among our sample, 25 percent of the observations had at least three female directors (i.e. attained the critical mass) which means that the large majority of the sample is not able to achieve the potential positive effects of critical mass. Interestingly, the regression results are partially different compared to the ones obtained in the main regressions. Regarding that the critical mass variable is a dummy variable, the interpretation of the effects should be done in comparison with observations consisting of less than three female directors (i.e. not reaching the critical mass). While critical mass seems to lack any effect on innovation in R1, we found that the same pattern as detected in R2 with respect to the Blau index and ROA also holds for critical mass; critical mass has a positive significant effect on ROA, but it is not robust when we controlled for reversed causality. On the other hand, critical mass is insignificant on Tobin's Q – just as in the main model – but it has a negative sign, indicating a negative economic relevance on market performance. Overall, the lack of causal robust relationships was somewhat surprising as the main rationale to investigate critical mass is that the possession of diverse experiences, knowledge and values of gender heterogenous boards is only leveraged once the minority group reaches a number of at least three (Konrad et al., 2008). Thus, the existence of a critical mass number could have explained why we did not find robust positive

[14] The use of critical mass makes no difference on R4 compared to our main model (since gender diversity is not a part of this regression). Therefore, there was no reason to conduct R4 one more time as it would be identical to the original R4. Further, the lack of significant effects in R1 and R2 made R3 irrelevant to run and it is therefore not presented either.

significant effects in our main model; the true benefits are realized when the minority group reaches the critical mass, which is not reflected in the Blau or the Shannon indices. For example, these indices will yield a lower value on board gender diversity for a board with three women and seven men compared to a board with two women and four men, thus neglecting a potential critical mass effect. A perhaps more holistic measure would be to combine a measure like the Blau or the Shannon indices with the minimum number needed to realize the benefits of a gender diverse board. However, the results of this subanalysis did not support this idea. Despite that critical mass has been applied in legislative and political settings in recent years, there are few studies supporting its economic and statistical relevance (Torchia et al., 2011); Torchia et al. (2011) found a positive significant effect on firm innovation, while Post and Byron (2015) could only state that there are indications of critical mass' relevance for financial performance due to the limited amount of studies in their meta-study that examined critical mass. Moreover, the contemporary critical mass idea by Konrad et al. (2008) is based on interviews and discussions with 50 female US-directors. Bearing this in mind, we remain a bit critical whether three really is the magic number due to its small base of empirical support albeit the concept is intuitive and logical from our perspective. Consequently, we reran the subanalysis with two and four female directors as the magic number (unreported). These regressions showed that two or more female directors had no significant effects in either regression, however, four females or more had a small significant effect on ROA, but it was not robust for reversed causality either. This finding, while preliminary, is at least an indication whether three really is the correct magic number.

As previous research has investigated the presence of at least one female director (Campbell & Mínguez-Vera, 2008; Gordini & Rancati, 2017), we carried out the same analysis and could not find any significant effects (unreported). Just as previous studies explain this, we believe that this is due that if the minority is alone (i.e. one female director) she is more likely to become marginalized or being selected due to tokenism rather than based on her competencies. Additionally, it could also be explained that when being alone in a minority group, it is more likely to assimilate to the existing norms and behaviors of the board, leading to poor utilization of the unique resources the minority may bring (Miller & Triana, 2009; Rose, 2007).

6 Conclusion

6.1 Implications and conclusions

This study has several implications for the application of AT and RDT in the corporate governance literature. First, RDT assumes that if unique resources are provided, these will be efficiently utilized and followingly benefit the firm without considering potential downsides. However, several researchers state that increasing board gender diversity often results in slower decision-making, and a risk of the common information effect. Consequently, we view RDT as somewhat simplistic as it neglects the degree to which additional unique resources can be utilized. Arguably, RDT would become more practically implementable if there was a clarification of how unique resources should be utilized. Second, AT neglects the contextual factors that impact the degree to which board gender diversity contributes to increased monitoring capabilities. Two contextual factors are the initial level of monitoring and the country-specific shareholder protection level. While the first factor indicates a reversed U-shaped relationship of monitoring and shareholder value, the second suggests that Swedish companies are not equally benefited in terms of monitoring due to relatively high shareholder protection. Thus, we posit that AT needs to become more sensitive to contextual differences in order to increase the understanding whether enhanced monitoring will benefit firms or rather have a detrimental or non-existing effect.

Our study also provides implications for practitioners. First, as our paper points out, one of the possible reasons why we did not find any positive relationship between board gender diversity and innovation or financial performance can be due to existing board members not knowing how to utilize the additional unique resources and capabilities that female directors bring. Consequently, practical tools and understanding of how different types of resources are best utilized are necessary for boards to develop. For example, this can be done by establishing an open board culture where the chairperson encourages the participation of all directors. Conditional on the advantages of board gender diversity are not efficiently utilized today and that they can lead to positive firm outcomes if correctly utilized, failure to realize this potential upside would lead to opportunity losses which no practitioner would deem desirable. Second, considering that board gender diversity does not negatively impact financial performance, firms should bear this in mind when appointing directors as there is no obvious economic downside by achieving gender heterogenous boards.

To conclude, we want to highlight two important conclusions for this paper. First, this study originates from board gender diversity which indicates two layers; gender diversity and the board. While the first has been widely discussed, it is relevant to present the idea that it may be possible that gender diversity indeed has an effect on firm outcomes but just not on the board. For example, gender diversity could be highly relevant for operative teams or the top management team where it has a profound impact on these teams' activities; or there may be individual attributes that have a bigger impact on boards' main responsibilities which are more relevant to investigate and not directly related to gender. Likewise, it could be that the board's possibility to influence firm outcomes is very weak such that there is no effect of whatever demographic characteristic that is under investigation, making studies as these redundant. While we find this last point important to consider, we deem it unlikely with respect to the literature on boards' connection to firm outcomes. Second, we consider it important to highlight what this paper was unable to find. As we did not find a negative or a positive relationship between board gender diversity and financial performance, we can neither reject nor support the business case for board gender diversity. However, given that board gender diversity can be seen from a variety of perspectives and provides value from an ethical, moral and societal perspective, and as there is no economic downside, one could argue that increased board gender diversity only provides non-negative outcomes.

6.2 Limitations

Our paper does not come without limitations. First, due to restricted access to secondary data on board composition, we limited our sample to the OMXS which impacts our ability to generalize our findings to a broader population than this. Further, as previous research together with our findings suggest that firm size is closely related to financial performance, smaller firms than those listed on the OMXS are most likely to display different relationships. Second, as a result of limited access to data, we restricted ourselves to operationalize innovation as R&D expenses. While R&D expenses are the most widely used measure for innovation, it still only represents the input. This is problematic as this paper indicates that innovation is highly sensitive to its operationalization and how a researcher deals with missing values. Along with the difficulties in capturing the core definition of innovation and how it should be measured, we find innovation to be a complex issue similarly to board gender diversity. Hence, robust findings regarding innovation need to be founded on several measures of innovation. Third, given that our hypotheses would not have been rejected, we would still have been unable to

conclude whether it was due to improved resource provision, monitoring or both. Thus, our research design does not possess the attributes to enable a detailed discussion whether RDT or AT best explained a potential positive relationship. Likewise, it may actually be the case that RDT and AT respond in the opposite direction of board gender diversity, implying that they outweigh each other's effects on firm outcomes, but which we are unable to detect. Overall, our possibility to support or reject any theory is therefore limited.

6.3 Suggestions for future research

Based on previous research, our results and limitations, the following section aims to present suggestions for future research. First, while we want to encourage future country-specific research in order to track historical development, we also want to highlight the importance of multi-country studies as all countries vary in terms of historical backgrounds, culture and regulations. There may also exist differences between countries in terms of the strength male directors have on minority groups' assimilation to current norms as mentioned in the discussion. Therefore, by investigating differences between countries, we would gain an increasing understanding of the effects of board gender diversity as the effect of board gender diversity may be more or less prominent for different countries. Second, the usefulness of case studies should be emphasized as they complement the lack of explanation power for individual observations which is a drawback with studies as ours. Thus, the rationale for carrying out case studies is to observe actual human behavior in order to get a more in-depth understanding of board gender diversity's effect on the inner workings of boards. We believe this to be beneficial as board gender diversity is argued to have a complex and indirect relationship with firm outcomes, such as innovation and financial performance. Hence, a case study has higher probability to explain this relationship on a deeper level whereas studies as this contribute with generalizability. Last, as there is no law enforcing firms to disclose R&D expenses, our measure of innovation was presented as a limitation of this paper. Followingly, we want to encourage future researchers to incorporate three measures of innovation; input, throughput and output measures, in order to better capture the entire innovation process. In addition, we found that ROA and Tobin's Q had a negative correlation, which is peculiar as it is plausible that better accounting-based performance leads to higher market value of the firm. Consequently, we propose future research to specifically investigate the relationship between these two measures.

References

Adams, R.B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance, *Journal of Financial Economics*, vol. 94, no. 2, pp. 291-309

Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership, *American Economic Review*, vol. 103, no. 1, pp. 277-304

Alexy, O., George, G., & Salter, A.J. (2013). Cui bono? The selective revealing of knowledge and its implications for innovative activity, *Academy of Management Review*, vol. 38, no. 2, pp. 270-291

Angrist, J.D., & Pischke, J.S. (2009). Mostly harmless econometrics: An empiricist's companion, New Jersey: Princeton University Press

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations, *The Leadership Quarterly*, vol. 21, no. 6, pp. 1086-1120

Ararat, M., Aksu, M., & Tansel Cetin, A. (2015). How board diversity affects firm performance in emerging markets: Evidence on channels in controlled firms, *Corporate Governance: An International Review*, vol. 23, no. 2, pp. 83-103

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations', *Journal of Personality and Social Psychology*, vol. 51, no. 6, pp. 1173–1182

Baumol, W.J. (2002). The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism, New Jersey: Princeton University Press

Baysinger, B.D., & Butler, H.N. (1985). Corporate governance and the board of directors: Performance effects of changes in board composition, *Journal of Law, Economics, & Organization*, vol. 1, no. 1, pp. 101-124

Bear, S., Rahman, N., & Post, C. (2010). The impact of board diversity and gender composition on corporate social responsibility and firm reputation, *Journal of Business Ethics*, vol. 97, no. 2, pp. 207-221

Becker, B.E., & Huselid, M.A. (1992). The incentive effects of tournament compensation systems, *Administrative Science Quarterly*, vol. 77, no. 3, pp. 336-350

Bell, E., Bryman, A., & Harley, B. (2019). Business Research Methods. Oxford: Oxford University Press

Bertrand, M., & Mullainathan, S. (2003). Enjoying the quiet life? Corporate governance and managerial preferences, *Journal of Political Economy*, vol. 111, no. 5, pp. 1043-1075

Biernat, M., & Kobrynowicz, D. (1997). Gender-and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups, *Journal of Personality and Social Psychology*, vol. 72, no. 3, p. 544-557

Blau, P.M. (1977). Inequality and Heterogeneity: A Primitive Theory of Social Structure. New York: The Free Press

Brown, T., & Katz, B. (2011). Change by design, *Journal of Product Innovation Management*, vol. 28, no. 3, pp. 381-383

Brunzell, T., & Liljeblom, E. (2014). Chairmen's perceptions of female board representation: A study on Nordic listed companies, *Equality, Diversity and Inclusion: An International Journal*, vol. 33, no. 6, pp. 523–534

Burt, R.S. (1997). The contingent value of social capital, *Administrative Science Quarterly*, vol. 42, no. 2, pp. 339-365

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance, *Journal of Business Ethics*, vol. 83, no. 3, pp. 435-451

Cannella, B., Finkelstein, F., & Hambrick, D.C. (2009). Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards, UK: Oxford University Press

Carpenter, M.A., & Westphal, J.D. (2001). The strategic context of external network ties: Examining the impact of director appointments on board involvement in strategic decision making, *Academy of Management journal*, vol. 44, no. 4, pp. 639-660

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*, vol. 18, no. 5, pp. 396-414

Carter, D.A., Simkins, B.J., & Simpson, W.G. (2003). Corporate governance, board diversity, and firm value, *Financial Review*, vol. 38, no. 1, pp. 33-53

Cassiman, B., & Golovko, E. (2011). Innovation and internationalization through exports, *Journal of International Business Studies*, vol. 42, no. 1, pp. 56-75

Caves, R.E., & Ghemawat, P. (1992). Identifying mobility barriers, *Strategic Management Journal*, vol. 13, no. 1, pp. 1-12

Cefis, E., & Marsili, O. (2005). A matter of life and death: Innovation and firm survival, *Industrial and Corporate Change*, vol. 14, no. 6, pp. 1167-1192

Certo, S.T., Daily, C.M., & Dalton, D.R. (2001). Signaling firm value through board structure: An investigation of initial public offerings, *Entrepreneurship Theory and Practice*, vol. 26, no. 2, pp. 33-50

Chen, J., Leung, W.S., & Evans, K.P. (2018). Female board representation, corporate innovation and firm performance, *Journal of Empirical Finance*, vol. 48, pp. 236-254

Chung, K.H., & Pruitt, S.W. (1994). A simple approximation of Tobin's q, *Financial Management*, vol. 23, no. 3, pp. 70-74

Cohen, W.M., & Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation, *Administrative Science Quarterly*, vol. 35, no. 1, pp.128-152

Cox, T.H., & Blake, S. (1991). Managing cultural diversity: Implications for organizational competitiveness, *Academy of Management Perspectives*, vol. 5, no. 3, pp. 45-56

Creswell, J.W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches, Los Angeles, California: Sage Publications Inc

Dalton, D.R., Daily, C.M., Ellstrand, A.E., & Johnson, J.L. (1998). Meta-analytic reviews of board composition, leadership structure, and financial performance, *Strategic Management Journal*, vol. 19, no. 3, pp. 269-290

Dalziel, T.D., Gentry, R.J., & Bowerman, M. (2011). An integrated agency–resource dependence view of the influence of directors' human and relational capital on firms' R&D spending, *Journal of Management Studies*, vol. 48, no. 6, pp. 1217-1242

Eagly, A.H. (2016). When passionate advocates meet research on diversity, does the honest broker stand a chance?, *Journal of Social Issues*, vol. 72, no. 1, pp.199-222

Eisenhardt, K.M. (1989). Making fast strategic decisions in high-velocity environments, *Academy of Management Journal*, vol. 32, no. 3, pp. 543-576

Eklund, J.E., Palmberg, J., & Wiberg, D. (2009). Ownership structure, board composition and investment performance, *Corporate Ownership and Control*, vol. 7, no. 1, p. 117

Erhardt, N.L., Werbel, J.D., & Shrader, C.B. (2003). Board of director diversity and firm financial performance, *Corporate governance: An international review*, vol. 11, no. 2, pp. 102-111

Fama, E.F., & Jensen, M.C. (1983). Agency problems and residual claims, *The Journal of Law and Economics*, vol. 26, no. 2, pp. 327-349

Farrell, K.A., & Hersch, P.L. (2005). Additions to corporate boards: The effect of gender, *Journal of Corporate Finance*, vol. 11, no. 1-2, pp. 85-106.

Ferrari, G., Ferraro, V., Profeta, P., and Pronzato, C. (2016). Gender quotas: Challenging the boards, performance, and the stock market, [Discussion paper no. 10239], The Institute for the Study of Labor

Forbes, D.P., & Milliken, F.J. (1999). Cognition and corporate governance: Understanding boards of directors as strategic decision-making groups, *Academy of Management Review*, vol. 24, no. 3, pp. 489-505

Fristedt, D., & Sundqvist, S. (2008). Styrelser och revisorer i Sveriges börsföretag 2008–2009. Stockholm: SIS Ägarservice AB

Fristedt, D., & Sundqvist, S. (2009). Styrelser och revisorer i Sveriges börsföretag 2009–2010. Stockholm: SIS Ägarservice AB

Fristedt, D., Larsson, Å., & Sundqvist, S. (2010). Styrelser och revisorer i Sveriges börsföretag. 2010–2011, Stockholm: SIS Ägarservice AB

Fristedt, D., Larsson, Å., & Sundqvist, S. (2011). Styrelser och revisorer i Sveriges börsföretag. 2011–2012, Stockholm: SIS Ägarservice AB

Fristedt, D., Larsson, Å., & Sundqvist, S. (2012). Styrelser och revisorer i Sveriges börsföretag. 2012–2013, Stockholm: SIS Ägarservice AB

Fristedt, D., Larsson, Å., & Sundqvist, S. (2013). Styrelser och revisorer i Sveriges börsföretag. 2013–2014, Stockholm: SIS Ägarservice AB

Fristedt, D., Larsson, Å., & Sundqvist, S. (2014). Styrelser och revisorer i Sveriges börsföretag. 2014–2015, Stockholm: SIS Ägarservice AB

Gabrielsson, J., & Huse, M. (2004). Context, behavior, and evolution: Challenges in research on boards and governance, *International Studies of Management & Organization*, vol. 34, no. 2, pp. 11-36

Golden, B.R., & Zajac, E.J. (2001). When will boards influence strategy? Inclination × power = strategic change, *Strategic Management Journal*, vol. 22, no. 12, pp. 1087-1111

Gordini, N., & Rancati, E. (2017). Gender diversity in the Italian boardroom and firm financial performance, *Management Research Review*, vol. 40, no. 1, pp. 75–94

Granovetter, M.S. (1973). The strength of weak ties, *American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380

Greenhalgh, C., & Rogers, M. (2010). Innovation, Intellectual Property, and Economic Growth. New Jersey: Princeton University Press

Groysberg, B., & Bell, D. (2013). Dysfunction in the boardroom, *Harvard Business Review*, vol. 91, no. 6, pp. 89-97

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). Multivariate Data Analysis, Essex: Pearson Education Limited.

Hambrick, D.C., & Mason, P.A. (1984). Upper echelons: The organization as a reflection of its top managers, *Academy of Management Review*, vol. 9, no. 2, pp. 193-206

Harrison, D.A., & Klein, K.J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations, *Academy of Management Review*, vol. 32, no. 4, pp. 1199-1228

Hausman, J.A. (1978). Specification tests in econometrics, *Econometrica*: *Journal of the Econometric Society*, vol. 46, no. 6, pp. 1251-1271

Haynes, K.T., & Hillman, A. (2010). The effect of board capital and CEO power on strategic change, *Strategic Management Journal*, vol. 31, no. 11, pp. 1145-1163

Hillman, A.J. (2015). Board diversity: Beginning to unpeel the onion, *Corporate Governance: An International Review*, vol. 23, no. 2, pp. 104-107

Hillman, A.J., & Dalziel, T. (2003). Boards of directors and firm performance: Integrating agency and resource dependence perspectives, *Academy of Management Review*, vol. 28, no. 3, pp. 383-396

Hillman, A.J., Cannella Jr, A.A., & Harris, I.C. (2002). Women and racial minorities in the boardroom: How do directors differ?, *Journal of Management*, vol. 28, no.6, pp. 747-763

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*, vol. 37, no. 2, pp. 235-256

Hillman, A.J., Keim, G.D., & Luce, R.A. (2001). Board composition and stakeholder performance: Do stakeholder directors make a difference?, *Business & Society*, vol. 40. No. 3, pp. 295-314

Hitt, M.A., Hoskisson, R.E., & Kim, H. (1997). International Diversification: Effects on Innovation and Firm Performance in Product-Diversified Firms, *The Academy of Management Journal*, vol. 40, no. 4, pp. 767

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective, *The Review of Economic Studies*, vol. 66, no. 1, pp. 169-182

Hoskisson, R., Hitt, M., Johnson, R., & Grossman, W. (2002). Conflicting voices: The effects of institutional ownership heterogeneity and internal governance on corporate innovation strategies, *The Academy of Management Journal*. Vol. 45, pp. 697-716

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors across studies, *Psychological Bulletin*, vol. 96, no. 1, pp. 72-98

Huse, M. (2007). Boards, Governance and Value Creation: The Human Side of Corporate Governance, Cambridge: Cambridge University Press

Huse, M., & Solberg, A.G. (2006). Gender-related boardroom dynamics: How Scandinavian women make and can make contributions on corporate boards, *Women in Management Review*, vol. 21, no. 2, pp. 113-130

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm, *Administrative Science Quarterly*, vol. 37, no. 3, pp. 422-447

Ibarra, H. (1993). Personal networks of women and minorities in management: A conceptual framework, *Academy of Management Review*, vol. 18, no. 1, pp. 56-87

Jackling, B., & Johl, S. (2009). Board structure and firm performance: Evidence from India's top companies, *Corporate Governance: An International Review*, vol. 17, no. 4, pp. 492-509

Jensen, M.C. (1993). The modern industrial revolution, exit, and the failure of internal control systems, *The Journal of Finance*, vol. 48, no. 3, pp. 831-880

Jensen, M.C., & Meckling, W.H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics*, vol. 3, no. 4, pp. 305-360

Johnson, S. G., Schnatterly, K., & Hill, A. D. (2013) Board composition beyond independence: Social capital, human capital, and demographics, *Journal of Management*, vol. 39, no. 1, pp. 232-262

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction, *Academy of Management Journal*, vol. 45, no. 6, pp. 1183-1194

Konrad, A.M., Kramer, V., & Erkut, S. (2008). The impact of three or more women on corporate boards, *Organizational Dynamics*, vol. 37, no. 2, pp. 145-164

Lafuente, E., & Vaillant, Y. (2019). Balance rather than critical mass or tokenism: Gender diversity, leadership and performance in financial firms, *International Journal of Manpower*, vol. 4, no. 5, pp. 894-916

Lööf, H., & Heshmati, A. (2006). On the relationship between innovation and performance: A sensitivity analysis, *Economics of Innovation and New Technology*, vol. 15, no. 4-5, pp. 317-344

Marton, J., Sandell, N., & Stockenstrand, A.K. (2018). Redovisning: Från Bokföring till Analys. Lund: Studentlitteratur

McGregor, J. (2015). There are more men on corporate boards named John, Robert, William or James than there are women on boards altogether, The Washington Post, 25 February, Available online: https://www.washingtonpost.com/news/on-leadership/wp/2015/02/25/there-are-more-men-on-corporate-boards-named-john-robert-william-or-james-than-there-are-women-altogether/ [Accessed 28 April 2020]

Miller, T., & del Carmen Triana, M. (2009). Demographic diversity in the boardroom: Mediators of the board diversity–firm performance relationship, *Journal of Management Studies*, vol. 46, no. 5, pp. 755-786

Milliken, F.J., & Martins, L.L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups, *Academy of Management Review*, vol. 21, no. 2, pp. 402-433

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of “unstructured" decision processes, *Administrative Science Quarterly*, vol. 21, no. 2, pp. 246-275

Nelson, R.R., & Winter, S.G. (1982). An Evolutionary Theory of Economic Change. Massachusetts: Harvard University Press

Nielsen, S., & Huse, M. (2010). The contribution of women on boards of directors: Going beyond the surface, *Corporate Governance: An International Review*, vol. 18, no. 2, pp. 136-148

O’Brien, J.P. (2003). The capital structure implications of pursuing a strategy of innovation, *Strategic Management Journal*, vol. 24, no. 5, p. 415-431

Pan, Y., & Sparks, J.R. (2012). Predictors, consequence, and measurement of ethical judgments: Review and meta-analysis, *Journal of Business Research*, vol. 65, no. 1, pp. 84-91

Parisi, M.L., Schiantarelli, F., & Sembenelli, A. (2006). Productivity, innovation and R&D: Micro evidence for Italy, *European Economic Review*, vol. 50, no. 8, pp. 2037-2061

Peterson, C.A., & Philpot, J. (2007). Women's roles on US Fortune 500 boards: Director expertise and committee memberships, *Journal of Business Ethics*, vol. 72, no. 2, pp. 177-196

Pfeffer, J., & Salancik, G. (1978). The External Control of Organizations: A Resource Dependence Perspective, New York: Harper & Row

Post, C., & Byron, K. (2015). Women on boards and firm financial performance: A meta-analysis, *Academy of Management Journal*, vol. 58, no. 5, pp. 1546-1571

Randøy, T., Thomsen, S., & Oxelheim, L. (2006). A Nordic perspective on corporate board diversity, [Working paper no. 05030], Nordic Innovation Centre, Oslo

Richtnér, A., Brattström, A., Frishammar, J., Magnusson, M., & Björk, J. (2017). Creating better innovation measurement practices, *MIT Sloan Management Review*, vol. 59, no. 1, pp. 45-53

Robinson, G., & Dechant, K. (1997). Building a business case for diversity, *Academy of Management Perspectives*, vol. 11, no. 3, pp. 21-31

Rogers, M. (1998). The definition and measurement of innovation, [Working paper no. 10/98], Melbourne Institute of Applied Economic and Social Research, Melbourne

Rose, C. (2007). Does female board representation influence firm performance? The Danish evidence, *Corporate Governance: An International Review*, vol. 15, no. 2, pp. 404-413

Rothaermel, F.T., & Hess, A.M. (2007). Building dynamic capabilities: Innovation driven by individual-, firm-, and network-level effects, *Organization Science*, vol. 18, no. 6, pp. 898-921

Sanders, W.G., & Carpenter, M.A. (2003). Strategic satisficing? A behavioral-agency theory perspective on stock repurchase program announcements. *Academy of Management Journal*, vol. 46, no. 2, pp. 160-178

Schilling, M. A. (2017). Strategic Management of Technological Innovation, New York: McGraw-Hill Education

Schumpeter, J. (1942). Capitalism, Socialism and Democracy. New York: Harper & Brothers

Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423

Shen, W., & Gentry, R.J. (2014). A cyclical view of the relationship between corporate governance and strategic management, *Journal of Management and Governance*, vol. 18, no. 4, pp. 959–973

Simpson, W.G., Carter, D., & D'Souza, F.P. (2010). What do we know about women on boards?, *Journal of Applied Finance*, vol. 20, no. 2, pp. 27-39

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*, vol. 55, no. 7, pp. 569–593

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion, *Journal of Personality and Social Psychology*, vol. 48, no. 6, pp. 1467-1478

Teece, D.J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management, *Strategic Management Journal*, vol. 18, no. 7, pp. 509-533

Terjesen, S., Sealy, R., & Singh, V. (2009). Women directors on corporate boards: A review and research agenda, *Corporate Governance: An International Review*, vol. 17, no. 3, pp. 320-337

The World Bank. (2014). Strength of investor protection, Available online: https://tcdata360.worldbank.org/indicators/h2e15b0d6?country=SWE&indicator=647&countries=BRA&viz=line_chart&years=2007,2017 [Accessed 15 May 2020]

Toivanen, O., Stoneman, P., & Bosworth, D. (2002). Innovation and the market value of UK Firms, 1989-1995, *Oxford Bulletin of Economics and Statistics*, vol. 64, no. 1, pp. 39–61

Torchia, M., Calabrò, A., & Huse, M. (2011). Women directors on corporate boards: From tokenism to critical mass. *Journal of Business Ethics*, vol. 102, no. 2, pp. 299-317

Torchia, M., Calabrò, A., Gabaldon, P., & Kanadli, S.B. (2018). Women directors contribution to organizational innovation: A behavioral approach, *Scandinavian Journal of Management*, vol. 34, no. 2, pp. 215-224

Tricker, B. (2019). Corporate Governance: Principles, Policies, and Practices, Oxford: Oxford University Press

Nagji, G., & Tuff, G. (2012). Managing your innovation portfolio, *Harvard Business Review*, vol. 90, no. 5, pp. 66-74

Tyrowicz, J., Terjesen, S., & Mazurek, J. (2020). All on board? New evidence on board gender diversity from a large panel of European firms, *European Management Journal*, [in press]

Utterback, J.M. (1971). The process of technological innovation within the firm, *Academy of Management Journal*, vol. 14, no. 1, pp. 75-88

Wernerfelt, B., & Montgomery, C.A. (1988). Tobin's Q and the importance of focus in firm performance, *The American Economic Review*, vol. 78, no. 1, pp. 246-250

Wooldridge, J.M. (2016). Introductory Econometrics: A Modern Approach, Boston: Cengage Learning

Zahra, S., & Stanton, W. (1988). The implications of board of directors' composition on corporate strategy and performance, *International Journal of Management*, vol. 5, no. 2, pp. 229-236

Zahra, S.A. (1996). Technology strategy and financial performance: Examining the moderating role of the firm's competitive environment, *Journal of Business Venturing*, vol. 11, no. 3, pp. 189-219

Zahra, S.A., & Garvis, D.M. (2000). International corporate entrepreneurship and firm performance: The moderating effect of international environmental hostility, *Journal of Business Venturing*, vol. 15, no. 5-6, pp. 469-492

Zahra, S.A., & Pearce, J.A. (1989). Boards of directors and corporate financial performance: A review and integrative model, *Journal of Management*, vol. 15, no. 2, pp. 291-334

Zahra, S.A., Ireland, R.D., & Hitt, M.A. (2000). International expansion by new venture firms: International diversity, mode of market entry, technological learning, and performance, *Academy of Management Journal*, vol. 43, pp. 925-950

Østergaard, C.R., Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation, *Research Policy*, vol. 40, no. 3, pp. 500-509

Appendix A

Calculation of variables using Compustat

Tobin's Q

Formula used: $\frac{Market\ value\ of\ equity + Book\ value\ of\ debt}{Book\ value\ of\ total\ assets}$

where $Book\ value\ of\ debt = Book\ value\ of\ total\ assets - Book\ value\ of\ equity.$

Compustat items (abbreviation within parentheses) used for market value of equity:

- Price – Close – Daily (PRCCD): The daily closing price of stock.
- Shares Outstanding (CSHOC): The number of daily outstanding shares.

Formula used for market value of equity per day: $PRCCD * CSHOC$

The market of firm i at year t was calculated as the average of the daily market value of equity.

Compustat items used for book value of debt:

- Assets – Total (AT): The book value of total assets. This item is used consistently whenever total assets are used in any calculation of the variables.
- Stockholders Equity – Total (TEQ): The book value of total equity. If TEQ was unavailable, we used Stockholders Equity – Parent (SEQ): The book value of total equity attributable to parent.

Formula used for book value of debt: $AT - TEQ$ or alternatively $AT - SEQ$

Return on assets (ROA)

Formula used: $\frac{Earnings\ before\ interest\ and\ taxes + Interest\ income}{Book\ value\ of\ total\ assets}$

Compustat items used:

- Earnings Before Interest and Taxes (EBIT)
- Interest and Related Income – Total (IDIT)
- Assets – Total (AT)

Firm size

Formula used: $Log(Book\ value\ of\ total\ assets)$ where $Book\ value\ of\ total\ assets > 0$

Leverage

Formula used: $\frac{Long-term\ debt + Current\ liabilities}{Book\ value\ of\ total\ assets}$

Compustat items used:

- Long-Term Debt – Total (DLTT)
- Debt in Current Liabilities – Total (DLC)
- Assets – Total (AT)

Innovation

Formula used: $Log(R\&D\ expenses)\ if\ R\&D\ expenses > 0, otherwise\ 0$

Compustat item used:

- Research and Development Expenses (XRD): All costs related to the development of new products or services.

About identifying companies in Compustat

The included companies in the study have the following Global Company Keys (GVKEY):
213323 230376 247318 103051 251921 239281 223492 17420 215406 239279 215361 244725 103219 101693 215363 245736 101585 249178 234163 249099 241543 273944 64894 285128 234178 290542 212830 234179 221471 287274 232089 221484 14620 221486 224231 215390 241562 215391 245320 242714 221806 103045 212844 102476 274480 102276 103065 288950 286889 100744 277941 232112 274196 234216 276568 234221 247323 228037 16299 104961 217164 213044 245222 221210 65423 239415 235543 102009 284609 234250 234251 234246 252285 213047 234243 286433 216617 274484 216621 216280 213075 234239 245001 234237 213048 237823 232138 224213 282053 250761 103314 222505 19591 220575 12368 234226 104981 216281 245746 221706 100980 11749 232119 234219 102915 245267 234211 100956 286911 222286 4439 220579 274451 101048 216867 234203 249124 11217

In order to identify the companies in the downloaded Excel sheet, we also included the company name using the Compustat item Company Name (CONM).

About currencies

Some firms reported their numbers in a different currency than SEK. Hence, we included the Compustat item ISO Currency Code (CURCD) to know which currency each number was reported in. (For the daily stock prices the code was CURCDD.)

About units

All the above mentioned financial Compustat items were expressed in MSEK, except for PRCCD and CSHOC. Hence, we multiplied the items in MSEK with 1 000 000.

Appendix B

Regression equations

R1: Board gender diversity and innovation

The first regression tested the relationship between board gender diversity and innovation. Therefore, this regression explicitly examined H1. The variable of interest was the gender diversity variable and its coefficient (β_2) should be interpreted as the expected effect of a tenth unit increase in board gender diversity on R&D expenses in percentage *ceteris paribus*. Thus, the following formula is used: $\beta_2 * 0.1 * 100$

$$Innovation_{i,t} = \beta_1 + \alpha_i + \mu_t + \beta_2 * Gender_{i,t} + \beta_j * CONTROL_{j,i,t} + \gamma_{i,t}$$

Where $Innovation_{i,t}$ represents the log R&D expenses of firm i in time t; β_1 is the common intercept for all firms before considering the impact of specific effects; α_i and μ_t is the firm and year specific effects of firm i in time t, respectively; $Gender_{i,t}$ is the Blau or Shannon index of the board gender diversity of firm i in time t; $CONTROL_{j,i,t}$ is the collection of control variables (j represents control variable j) presented in 3.3.4 which are board size in terms of number of ordinary board members, firm size in terms of log of total assets, independent directorship in terms of the fraction of independent directors, and leverage, of firm i in time t; $\gamma_{i,t}$ is the error term of firm i in time t.

R2: Board gender diversity and financial performance

To follow the mediation technique of Baron and Kenny (1986), the second regression should test the relationship between board gender diversity and financial performance. This regression did not on its own examine any of our hypotheses, however, it is a necessary component in order to test H2. The variable of interest was still the gender diversity variable and its accompanying coefficient (β_2) had a slightly different interpretation since the dependent variable (Tobin's Q and ROA, respectively) is stated as a ratio rather than in a logarithmic function. Hence, the beta coefficient of gender diversity should be interpreted as the expected effect of a tenth unit increase in board gender diversity on financial performance in percentage units *ceteris paribus*. The formula to calculate this is the same as in R1.

$$PERFORMANCE_{i,t} = \beta_1 + \alpha_i + \mu_t + \beta_2 * Gender_{i,t} + \beta_j * CONTROL_{j,i,t} + \gamma_{i,t}$$

Where $PERFORMANCE_{i,t}$ represent the financial performance in terms of Tobin's Q or ROA of firm i in time t; and the remaining variables are equivalent to the definition presented in board gender diversity and innovation.

R3: Board gender diversity, innovation and financial performance

The third regression aimed to estimate the mediating effect of innovation between board gender diversity and financial performance. As explained in 3.1.2, in order to prove a mediating relationship, the effect of board gender diversity needed to be reduced when including innovation. Consequently, by combining the results from R2 with R3, these regressions examined H3. Moreover, the variables of interest were now two; the gender diversity variable and the innovation variable. The coefficient of gender diversity (β_2) has the same interpretation as in regression two while the coefficient of innovation (β_3) measures the expected percentage unit change in financial performance for a percentage change in R&D expenses ceteris paribus and this effect is given by the coefficient itself (β_3).

$$PERFORMANCE_{i,t} = \beta_1 + \alpha_i + \mu_t + \beta_2 * Gender_{i,t} + \beta_3 * Innovation_{i,t} + \beta_j * CONTROL_{j,i,t} + \gamma_{i,t}$$

Where all variables have been defined in the previous two equations.

R4: Innovation and financial performance

The fourth and final regression investigated the relationship between innovation and financial performance in order to distinctly test hypothesis 2. The variable of interest was the innovation variable and its coefficient (β_2) should be interpreted in the same way as the innovation variable in R3.

$$PERFORMANCE_{i,t} = \beta_1 + \alpha_i + \mu_t + \beta_2 * Innovation_{i,t} + \beta_j * CONTROL_{j,i,t} + \gamma_{i,t}$$

Where all variables have been defined in the previous equations.

Appendix C

Subanalysis Tables

Industry – Health care

Tables C.1 Subanalysis Health Care R1 and R4

Explanatory variables	***Innovation***	
	Correct model	
Board size	0.369	0.378
	(0.284)	(0.293)
Firm size	0.324	0.302
	(0.226)	(0.236)
Independent directorship	1.369	1.409
	(1.438)	(1.45)
Leverage	3.002*	2.933*
	(1.778)	(1.744)
Blau	–3.409	
	(3.979)	
Shannon		–2.18
		(2.472)
Number of observations	133	133
Adjusted R^2	0.808	0.808
F-test	20.19***	20.147***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Explanatory variables	***Financial performance***	
	Correct model	
	Tobin's Q	ROA
Board size	–0.107	–0.026
	(0.149)	(0.019)
Firm size	–1.14**	0.21***
	(0.473)	(0.066)
Independent directorship	–1.227	0.013
	(0.844)	(0.133)
Leverage	5.837*	–0.238
	(3.052)	(0.198)
Innovation	–0.006	–0.004
	(0.036)	(0.005)
Number of observations	133	133
Adjusted R^2	0.616	0.696
F-test	8.303***	11.402***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Industry – Technology

Tables C.2 Subanalysis Technology R1 and R4

Explanatory variables	***Innovation***	
	Correct model	
Board size	–0.737	–0.788
	(0.496)	(0.491)
Firm size	5.502***	5.552***
	(1.428)	(1.435)
Independent directorship	1.314	1.468
	(2.938)	(2.938)
Leverage	–1.171	0.88
	(7.66)	(7.571)
Blau	–5.374**	
	(3.268)	
Shannon		3.3303
		(2.272)
Number of observations	112	112
Adjusted R^2	0.834	0.833
F-test	22.436***	22.313***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Explanatory variables	***Financial performance***	
	Correct model	
	Tobin's Q	ROA
Board size	0.042	–0.001
	(0.059)	(0.011)
Firm size	–0.574***	0.096**
	(0.189)	(0.04)
Independent directorship	0.406	0.033
	(0.355)	(0.053)
Leverage	–0.983*	–0.531***
	(0.715)	(0.154)
Innovation	0.018	–0.004*
	(0.012)	(0.002)
Number of observations	112	112
Adjusted R^2	0.87	0.674
F-test	29.527***	9.818***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Critical mass

Tables C.3 Subanalysis Critical Mass R1 and R2

Explanatory variables	***Innovation***
	Correct model
Board size	–0.003
	(0.181)
Firm size	1.849***
	(0.499)
Independent directorship	1.077
	(0.971)
Leverage	0.118
	(1.913)
Critical mass	0.404
	(0.536)
Number of observations	826
Adjusted R^2	0.85
Hausman test	13.893**
F-test	37.447***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

Explanatory variables	***Financial performance***	
	Correct model	
	Tobin's Q	ROA
Board size	–0.033	–0.008
	(0.043)	(0.005)
Firm size	–0.372*	0.103***
	(0.215)	(0.026)
Independent directorship	–0.044	–0.013
	(0.215)	(0.027)
Leverage	0.699	–0.278***
	(0.503)	(0.052)
Critical mass	–0.174	0.035**
	(0.129)	(0.018)
Number of observations	826	826
Adjusted R^2	0.71	0.725
Hausman test	15.075**	26.64***
F-test	16.798***	18.024***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

Appendix D

Multiple regression analysis results with the Shannon index

Table D.1 Regression 1 (R1) with the Shannon index

Explanatory variables	***Financial performance***				***Innovation***	
	Pooled model		*Correct model*		*Reversed model*	
	Tobin's Q	ROA	Tobin's Q	ROA	Tobin's Q	ROA
Board size	0.062	–0.024***	–0.041	–0.007	0.008	0.008
	(0.038)	(0.006)	(0.042)	(0.006)	(0.007)	(0.007)
Firm size	–0.229***	0.048***	–0.368*	0.01***	0.034*	0.026
	(0.056)	(0.007)	(0.221)	(0.026)	(0.019)	(0.018)
Independent directorship	0.795***	–0.078**	–0.065	–0.022	0.162***	0.163***
	(0.231)	(0.032)	(0.216)	(0.028)	(0.039)	(0.039)
Leverage	–1.033**	–0.147***	0.735	–0.274***	–0.139**	–0.118*
	(0.481)	(0.035)	(0.509)	(0.053)	(0.06)	(0.061)
Shannon	0.623**	0.089**	0.126	0.057		
	(0.281)	(0.037)	(0.342)	(0.038)		
Financial performance					0.002	0.07
					(0.005)	(0.046)
Number of observations	826	826	826	826	826	826
Adjusted R^2	0.114	0.144	0.709	0.724	0.683	0.684
Hausman test			12.889**	23.881***	13.498**	12.102**
F-test	22.236***	28.841***	16.741***	17.903***	14.861***	14.938***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

Table D.2 Regression 2 (R2) with the Shannon index

Explanatory variables	***Innovation***			***Board gender diversity***
	Pooled OLS	*Correct model*	*No missing values*	*Reversed model*
Board size	0.137	0.001	0.063**	0.008
	(0.213)	(0.185)	(0.025)	(0.007)
Firm size	0.712***	1.793***	0.446***	0.031*
	(0.238)	(0.502)	(0.077)	(0.018)
Independent directorship	9.113***	0.899	0.027	0.161***
	(1.376)	(0.955)	(0.114)	(0.039)
Leverage	–4.619**	0.228	–0.13	–0.138**
	(2.127)	(1.891)	(0.219)	(0.06)
Shannon	–5.211***	1.109	0.112	
	(1.688)	(1.329)	(0.148)	
Innovation				0.001
				(0.001)
Number of observations	826	826	555	826
Adjusted R^2	0.06	0.85	0.964	0.683
Hausman test		14.864***	19.852***	16.531***
F-test	11.496***	37.458***	151.347***	14.88***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 5)

Table D.3 Regression 3 (R3) with the Shannon index

Explanatory variables	***Financial performance***			
	Pooled OLS		*Correct model*	
	Tobin's Q	ROA	Tobin's Q	ROA
Board size	0.058	–0.023***	–0.041	–0.007
	(0.037)	(0.005)	(0.042)	(0.006)
Firm size	–0.251***	0.049***	–0.381*	0.1***
	(0.058)	(0.007)	(0.225)	(0.026)
Independent directorship	0.516**	–0.057**	–0.072	–0.022
	(0.241)	(0.031)	(0.216)	(0.028)
Leverage	–0.892**	–0.158***	0.734	–0.274***
	(0.489)	(0.034)	(0.508)	(0.053)
Innovation	0.031***	–0.002***	0.007	0.00001
	(0.006)	(0.001)	(0.005)	(0.001)
Shannon	0.782***	0.077**	0.118	0.057
	(0.277)	(0.037)	(0.341)	(0.038)
Number of observations	826	827	826	826
Adjusted R^2	0.141	0.154	0.709	0.724
Hausman test			14.612**	25.682***
F-test	23.661***	26.115***	16.607***	17.739***

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

Statistical significance is based on the White's robust standard errors reported in brackets.

Reported Hausman test is Chi-Square statistic (d.f. = 6)