Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Female board representation and corporate performance: A review and new estimates for Australia

TTPI - Working Paper 7/2023
June 2023

Nicholas Bayly
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

Robert Breunig
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

Chris Wokker
Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

Abstract

Recent years have seen increased interest in the effect of female board members on firm performance. Despite a conventional wisdom that female board members positively impact firm performance, a thorough examination of the research to date reveals no consensus that female board members have either a positive or negative effect on firm performance. We build the largest dataset of Australian board appointments assembled to date. We use our data to demonstrate how difficult it is to replicate existing research, with one example from Australia and one from the US. Using event studies and regression analyses we demonstrate that there is little evidence that female board representation affects firm performance.

Keywords: Firm performance, board of directors, gender representation, female directors
JEL codes: J16, N20, G32

**We are grateful to Renée Adams, Arnold Cowan, Ric Curnow, Nathan Deutscher, Shane Evans, Gigi Foster, Omar Majeed, Maria Racionero, Rhian Richardson, David Simmonds, Kristen Sobeck and Peter Varela for comments and suggestions. All errors remain our own. Corresponding author: robert.breunig@anu.edu.au*

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Asia and the Pacific

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

www.anu.edu.au

The Tax and Transfer Policy Institute (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

1. Introduction

In the United States women held 28.2 percent of board positions in 2020 (Milhomen, 2020) while in Australia almost 36 per cent of ASX200 company board positions are currently occupied by women (Australian Institute of Company Directors, 2022). Worldwide, the figure is considerably lower, at around 20 per cent (Milhomen, 2020). Policies to raise levels of female representation – from 'soft' encouragement to 'hard' quotas – have been implemented in many countries and justified in various ways. Although advocates often employ a 'social justice' argument when seeking to justify higher rates of female representation, there is also an 'economic argument' that greater representation of female directors improves company performance.

This article evaluates the 'economic argument', focusing on Australia, the UK, US and Norway. We begin by discussing how women on boards might affect company performance through impacts on board operations and outputs. We then discuss how female board representation can be measured and the accounting and market-based firm performance measures used in the literature.

We review the existing literature on the impact of female presence on boards on corporate performance. We first consider regression-based studies. We consider these studies through the lens of how convincingly they deal with the possibility of reverse causality and endogeneity. This is crucial when considering the relationship between female board members and firm performance. It may be, for example, that better performing firms are able to attract female board members more successfully. In this case, correlation between female board membership and firm performance does not reflect a causal impact of female board membership on performance. Our assessment is that most studies fail to convincingly identify a causal effect.

We then consider research which uses event studies to identify the effect of announcements of female board appointments on a firm's share price. If investors expect a female appointment to improve firm performance, the firm's share price should react positively to surprise announcements. We explain the event study approach and briefly discuss some important methodological issues around implementation and testing.

We demonstrate that there is no consensus in the existing literature about the existence of a statistically significant effect of female board representation on firm performance. This is the case for both the regression-based studies and the event studies. When statistically significant effects are found, there is similarly no consensus on whether the effects are positive or negative. Most regression studies in the literature do not look further than the correlation between board performance and female representation on boards. In the regression studies which convincingly deal with endogeneity, this lack of consensus is particularly pronounced with negative and insignificant results found more frequently than a positive impact of female board membership on firm performance.

We then turn to undertaking event study and regression analysis for Australia. To enable these analyses, we assemble the largest and most detailed dataset of Australian board appointments produced to date. Our motivation in assembling these data and undertaking our own analysis is three-fold. First, we want to shed light on the contradictory results that emerge from our review of the evidence. Are there particular outcome variables, particular sets of covariates or particular modelling approaches that lead to particular results? Second, we want to provide improved evidence for Australia.

Third, we want to provide a platform for further study of this issue. One problem encountered in the process of doing our research was the difficulty in replicating results from the papers we examined. Data and software code for replication were rarely available and we were often unable to follow what others had done. In the results sections below, we provide very detailed information on how to replicate our results. On our webpage, we provide computer code for all the results presented in the paper and, where possible, data. Some data are proprietary and in that case we provide careful instructions, alongside the computer code, for how to apply our code to that data to reproduce the results in the paper. [weblink to be added upon publication]

We first use our data to conduct a detailed event study analysis. We consider the impact of varying the period in which the stock market price reacts to the announcement of a female board member ("the event window") and conduct a variety of additional analysis. We find no strong evidence that a firm's share price reacts to the appointment of a female board member.

We then use our novel dataset to examine regression-based approaches to identifying the impact of female board members on firm performance. We begin by attempting to replicate two previous studies, one from Australia and one from the US. Both have garnered considerable attention. We are unable to replicate the results of the Australian study which had found a positive effect of female board membership on firm performance. The American study shows a negative relationship between female board membership and firm performance but we do not find this in Australia. Rather, we find no relationship when we use the approach of the American study.

Finally, we examine a wide range of different outcome measures, regression methodologies and control variables to discern the patterns, if any, which lead to contradictory conclusions about the impact of female board members on company performance.

Overall, our assessment of the evidence is that the conclusion most strongly supported in the Australian data is that female presence on boards has no effect on company performance, either positive or negative. While possible to find both negative and positive effects for certain specifications and performance measures, when a wide range of specifications and measures are considered, zero results are the dominant finding.

In what follows, we begin by considering how a female presence on boards might impact company performance. We then discuss how female representation on boards and board performance are measured in the literature. An extensive literature review follows, complemented by our own analysis. We finish with a discussion and conclusion.

2 Possible impacts of female representation on boards

Some advocates who push for greater representation of female directors argue that this will improve firm performance. For this 'economic argument' to be true, three propositions must hold. Firstly, the presence of women on boards must change board operations or be perceived to change operations. Secondly, these changes to board operations must change board outputs. Thirdly, these changes to board outputs must change company performance.

2.1 How women may change board operations

The literature suggests a range of mechanisms through which female representation could affect the performance of a board. For example, Dobbin and Jung (2011) suggest that there exists a 'female management style'. See the meta-analysis of this question by Eagly and Johnson (1990). Women's greater risk aversion and men's overconfidence are also regularly cited as reasons why female representation could affect the behaviour of corporate boards (Kim, Roden and Cox (2013); Larkin, Bernardi and Bosco (2012) and Huang and Kisgen (2013)). A range of other mechanisms have been put forward, with varying degrees of empirical support. Some of these mechanisms could be seen to echo common gender stereotypes, such as the claim by Zhang, Zhu and Bing (2013) that 'women possess more communal traits: they are affectionate, helpful, kind, sympathetic, interpersonally sensitive, nurturing, and concerned about others' welfare'.

However, even if we were to accept the proposition that there are inherent behavioural differences between the sexes *in general*, this does not necessarily mean that there will be inherent differences between male and female *board members*, who are unlikely to be representative of the broader population. Indeed, Chen, Crossland, and Huang (2016) find that survey evidence has provided 'little support for the claim that female directors are significantly more risk-averse than male directors'. Further, using a Swedish sample, Adams and Funk (2012) find that male directors are in fact more risk-averse and 'tradition and security oriented' than female ones.

Increased female representation on boards may change how boards operate because women have different 'skills, knowledge and perspectives' than men, as argued in Ali, Ng and Kulik (2014). Some have argued that such differences contribute positively to organisations. For example, Miller and Triana (2009) suggest that increased female representation brings a greater variety of ideas. However, Ahern and Dittmar (2012) note that new female directors hired in response to Norway's mandatory quota system were younger, less experienced and less likely to have CEO experience.[1] And Triana, Miller and Trzebiatowski (2013) posit that there will be lower levels of cohesion among more gender-diverse boards.

2.2 Changes to board outputs

Even if an increased female presence does change how boards operate, there may be no resultant change in outputs. For instance, additional female representation may change the deliberative process, but not the decisions arrived at by the board, hereafter referred to as *actual* outputs. But this definition of outputs is a narrow one. This is because outputs include not just the decisions made, but also how these decisions are *perceived* by different stakeholder groups.

The actual outputs of boards with greater female representation could differ for a range of reasons. For example, it may be the case that women directors are able to link to additional and different 'constituencies', such as female customers and female-owned suppliers (Hillman, Shropshire and Cannella 2007, Zhang 2012). Increased female board representation may also cause the board to make different decisions. For example, Triana, Millar and Trzebiatowski (2013) find that female directors are less likely to initiate strategic change than male directors when a company is

[1] Popular media also documented some of the odd results of the quota system: 'one Norwegian woman, lawyer Mimi Berdal, 55, at one point sat on a difficult-to-fathom 90 different boards' (Adams 2015).
https://www.forbes.com/sites/susanadams/2015/06/11/women-on-boards-slow-progress-and-marginalization-study-shows/?sh=5aadfc403745

performing poorly. And Chen, Crossland, and Huang (2016) find that firms with fewer female board members are less likely to acquire another firm. Dowling and Aribi (2013) and Huang and Kisgen (2013) report similar findings, and attribute them to a decreased tendency toward overconfidence among female executives.

There are a variety of reasons why a decision by a board with more women may be perceived differently. For example, in the past, the behaviour of firms with higher levels of female board representation may have been different from the behaviour of firms with lower levels of female board representation. Kathyayini, Tilt and Lester (2012) find a correlation between female board representation and the quality of environmental reporting. Similarly, Schnake, Williams and Fredenberger (2006) report that companies with greater female board representation are less likely to be investigated for suspected legal actions. And Hafsi and Turgut (2013) and Boulouta (2013) find that gender diverse boards perform more strongly on measures of corporate social performance. Board gender diversity may also send a signal to stakeholders that improves the board's reputation as a good corporate citizen among customers and as a meritocracy among current and potential employees (Wang and Clift, 2009), therefore increasing perceived outputs.

3 Measurement

Empirically addressing the question of whether female representation on boards affects firm performance requires measures which capture the degree of female presence on boards and firm performance measures. We consider each of these in turn.

3.1 Measuring female board representation

There are many measures of female board representation. The most straightforward is a simple percentage, running from zero to 100, or the corresponding proportion from zero to one. This measure is used in many papers, including Kim, Roden and Cox (2013) and Ali, Ng, and Kulik (2014). Other studies, such as Gul, Srinidhi and Ng (2011) as well as Wang and Clift (2009) use both a count of the number of women on a board, as well as the percentage. Dimovski, Lombardi and Cooper (2013) and Nguyen and Faff (2006) perform modelling with dummy variables for whether one or more board seats are occupied by women. Two studies (Bear, Rahman and Post (2010) and Dobbin and Jung (2011)) use a simple count of female directors instead of a proportion.

A more complicated measure of female board representation is the Blau Index, see Blau (1977). The Blau Index is equal to 1- $\Sigma_{i=1}^{2} p_i^2$ - where p_1 and p_2 represent the proportion of men and women, respectively on a firm's board. This measure is used in Ali, Ng and Kulik (2014) and Triana, Miller and Trzebiatowski (2013). With just two categories - male and female - the relationship between the proportion of women on a board and the Blau Index takes the form of an inverted parabola. Given that the regressions in these studies test for a linear relationship between their performance measures of choice and the Blau Index, a significant result indicates that the effect of adding female directors becomes weaker as parity is approached, and inverts if women have a greater share of directorships than men.

The diminishing returns implied by the Blau Index contrasts with the notion of 'critical mass' advanced by Erkut, Kramer and Konrad (2008). According to critical mass theory, one woman may suffer tokenism, and although the situation improves with two women, three or more women are considered a 'critical mass' and will have a noticeable impact on the functioning of the boardroom. Critical mass theory also admits the possibility that the link between company performance and female representation may in fact be 'U-shaped'. This contrasts with the relationship implied by the Blau Index, which is an inverted U-shape. In this vein, Joecks, Pull and Vetter (2013) find that female representation has a negative impact on firm performance up until a critical mass of 30 per cent, after which the effect becomes positive. And Brahma, Mwafor and Boateng (2021) find that having more than three female board members has a disproportionately positive effect on corporate performance compared to when there are fewer than three female board members.

3.2 Impact on corporate performance

A critical question in assessing the impact of female board membership on performance is deciding what constitutes "performance". Most measures of performance employed in the literature fall into two broad categories: accounting measures and market-based measures.

3.2.1 Accounting measures

As a measure of corporate performance, accounting measures are problematic both because they are backward looking and because of the potential scale of 'earnings management'. Dichev, Graham, Harvey and Rajgopal's (2016) survey of nearly 400 chief financial officers in the US found that CFOs believe that in any given year, one in five companies intentionally misrepresents its earnings using discretion within generally accepted accounting principles. Nevertheless, many studies consider how changes in the level of female board representation affect accounting returns, including return on assets (ROA), return on equity (ROE), and return on investment. Although there are a range of definitions employed in the literature, return on assets can be defined as net income divided by total assets; return on equity as the ratio of income to average shareholder equity; and return on investment as net income divided by net invested capital.

3.2.2 Market-based measures

Market-based measures of performance are more forward-looking, but are not without their own shortcomings. For instance, changes in a firm's market value could provide either evidence of *bias* in market valuation (for or against the appointment of female board members) or evidence of an *impact* of female appointees on expected performance, either in a positive or negative direction. The most common market-based measures are share price and Tobin's Q. Tobin's Q is commonly interpreted as a reflection of the market's expectations of future earnings.

4 Studies which use regression techniques to determine whether female presence on boards affects company performance

Most studies that consider the effects of female representation on corporate performance may, at least implicitly, be presenting correlation as causation. For example, Galbreath (2011) notes that better performing firms may be more likely to appoint women to boards, but oddly concludes that the low proportion of female directors in his sample somehow circumvents the issue of reverse causality.

Overall, we review 55 papers examining Australia, Norway, the UK and US, of which 24 found a positive relationship between female board representation and some measure of firm performance (including non-financial measures). Restricting this to papers that reviewed common financial measures of performance (ROA, ROE and Q), 14 of 34 analyses found a positive relationship. 17 papers made a serious attempt to address endogeneity, but only 9 of these examined financial performance. These are highlighted in section 4.2. A full list of the 55 papers are included in Appendix B. For each paper, we provide a brief summary of results along with information on the country studies, the sample of firms, the measure of female presence on boards

There is consistent evidence of a correlation between female representation and financial performance in many jurisdictions – one that has been noted at least as far back as Catalyst's (2004) US study. The event study and regression results which we present below reveal that this correlation exists in the Australian data. However, we find that endogeneity is a serious concern and that once this is addressed, there is little evidence of a causal relationship.

We first review studies that, in our view, do not convincingly address issues of reverse causality and endogeneity. We then summarise studies that do provide convincing causal identification of the impact of female board representation on firm performance.

4.1 Attempts to address endogeneity

Rather than provide a discussion of every paper we reviewed, our approach in this section is to discuss a broad set of papers to illustrate the various approaches these papers take to dealing with the possibility of reverse causation (better companies attract women to their boards) or endogeneity (that the presence of women on boards and better firm performance are driven by other, unobserved factors). Some papers ignore these problems, others deal with them in ways that are not particularly convincing. In section 4.2, we summarise the results from those papers that convincingly deal with endogeneity problems.

Horváth and Spirollari (2012) claim that lagging explanatory variables by one year is sufficient to address endogeneity, ignoring the likely autocorrelation of performance measures, especially over a short time period. Kathyayini, Tilt, and Lester (2012), having dismissed the availability of appropriate instrumental variables, rely on the approximately normal distribution of their ordinary least squares model's residuals as an 'informal sign that there are no obvious estimation issues'. This is clearly not a convincing test of exogeneity.

Terjesen, Couto and Francisco (2016) utilise a large dataset covering 47 countries, but their data are cross sectional, thus preventing the use of fixed effects techniques, which we demonstrate below to be consequential. Conyon and He (2017), find a positive and statistically significant relationship between the percentage of female board representation and firm performance using a quantile regression approach in the cross-section. They justify the omission of firm fixed effects by arguing that the presence of women on boards is correlated with the firm fixed effect. Logic would suggest that this correlation is a prime reason for the inclusion of firm-level fixed effects.

Cassells and Duncan (2020) report that an increase of "10 percentage points or more in female representation on the Boards of Australian ASX-listed companies leads to a 4.9% increase in company market value". They do not provide details of their regression specification. It is unclear whether the effect is an increase from female directors or just a `new blood' effect from any director change. They do not demonstrate that a female replacing a male director has a larger effect than a female replacing a female or a male replacing a male, which is what is required to make this claim.[2]

Among the studies reviewed, instrumental variables and two-stage least squares are the most common techniques used to control for the endogeneity that is regularly theorised and detected. But, there are significant difficulties in finding a valid instrument, given that the instrument must 1) be correlated with female board representation; and 2)

[2] Emails to the authors did not resolve these questions.

be uncorrelated with any unobservable factors affecting performance (Wooldridge 2009). Adams and Ferreira (2009) suggest 'the fraction of male directors on the board who sit on other boards on which there are female directors' meets these criteria. This instrument is shown to meet the first test, but the authors acknowledge that it may fail the second, as it may be correlated with performance, and may also be a proxy for the connectedness of the board, which could be correlated with performance. However, if board gender diversity impacts positively on performance, then a high level of connectedness between male board members and external female board members will indicate connections to the boards of high performing firms, and may just proxy for the quality of male directors. Causality is still at issue, as the instrument does not preclude the possibility that women self-select or are selected to better performing firms, which also have higher quality male directors. This issue is not substantively addressed by Adams and Ferreira (2009). Nor is it addressed by other studies that replicate this instrument or a variation thereof, such as Vafaei, Ahmed and Mather (2015) or Gregory-Smith, Main and O'Reilly (2014).

Anderson, Reeb, Upadhyay and Zhao (2011) use the heterogeneity of the county population of the firm's corporate headquarters as an instrument in their two stage least squares regression. Their reasons for suggesting that this will correlate with board diversity appear sound. But it could correlate with a range of other director, firm, industry and geographical characteristics that will impact performance and may be difficult to observe. Chen, Crossland and Huang (2016) use a similar instrumental variable – county-level female labour force participation – in their Heckman two-stage model. They argue that this instrument is unrelated to their outcome variable (acquisition intensity), but seem to ignore its potential correlation with unobservables. Their supplementary difference-in-differences analysis exploits the presumably exogenous change of board composition that follows the death of a director. However, the small number of deaths in their sample and the strong correlation between death and other director attributes (such as age and tenure) make it difficult to support the conclusion that any change in firms' performance can be attributed predominantly to exogenous changes in the board's gender ratio. Harjoto, Laksmana, and Lee (2015) and Srinidhi, Gul and Tsui (2011) also use a Heckman two-stage model. They both use the percentage of women employed in an industry to predict board diversity as it 'influences the likelihood of female participation in the boards of firms belonging to that industry'. This again seems problematic, as the instrument is just as likely to affect performance directly as is the original variable. That is, if one hypothesizes that board diversity directly impacts performance, it is strange to assume that workforce diversity does not.

Drawing on data before and after Norway's implementation of a quota on female directorships in 2003, Ahern and Dittmar (2012) use the proportion of female directors in 2002 interacted with year dummies as an instrument for female representation. They find that firms with a lower proportion of female directors before the announcement of the quota, and who therefore had to add more female directors by the compliance date in 2007, performed significantly worse than firms for which the quota was less binding. They argue that the only major pre-quota difference between firms with no female directors and those with at least one female director is firm size, and that this is captured by their use of firm fixed effects. Yang, Riepe, Moser, Pull and Terjesen (2019) also consider the effect of the introduction of quotas in Norway in 2003, and use a difference-in-difference approach with firms in Finland, Sweden and Denmark as the control group. They too find that the mandate had a negative effect on firm performance. Although these are perhaps the most convincing studies in the literature, they are only possible because of the sudden announcement of the quota. Further, the results of Ahern and Dittmar (2012) have been challenged by Eckbo, Nygaard and Thorburn (2021) who contend that after correcting for the cross-correlation in stock returns and the difficulty in attributing quota-related news to specific dates, there is no clear relationship between stock prices and the 2003 quota on female board representation. Maghin (2021) uses the introduction of gender quotas in France to consider the effect of greater female representation on firm profitability. Maghin finds that medium sized firms that increased female board representation became more profitable. However, the full implementation of these quotas took many years, and enforcement was haphazard, casting doubt on whether these results can be extrapolated to other contexts, such as larger firms.

Dale-Olsen, Schøne and Verner (2013) also consider the effect of the 2003 quota in Norway, and like Yang, Riepe, Moser, Pull and Terjesen (2019) they use a differences-in-differences approach. Specifically, they compare the ROA of firms that were affected by the quota (public limited companies – PLCs) and those that were not (ordinary limited companies--LTDs). They find that the quota had a negligible effect on firm performance. But, despite the range of controls employed, it is difficult to imagine that all of the meaningful differences between listed and non-listed companies can be accounted for. Indeed, the authors note the very strongly negative performance of LTDs relative to PLCs during the Global Financial Crisis (GFC) in 2008 and 2009, and that they find it 'hard to argue ... this was related to women's board representation'. Also in the Norwegian context, Bøhren and Strøm (2010) suggest that a range of variables – including firm performance and female representation – are likely to be endogenous. While it seems reasonable to suspect that these variables are endogenous, it also seems likely that the instruments they select – firm ownership and directors' networks – are also endogenous. Because their dataset only covers the period prior to 2003, they are not able to take advantage of the introduction of the quota as others have.

Wintoki, Linck and Netter (2012) use the dynamic panel generalised method of moments estimator popularised by Arellano and Bond (1991) to investigate links between board structure and firm performance. We refer to this estimation technique in the rest of the paper as 'Arellano-Bond'; see Roodman (2009). They justify the use of Arellano-Bond by noting that board makeup and financial performance may be a function of past firm performance. Arellano-Bond nonetheless has some potential drawbacks including the potential for weak instruments and the assumption of no serial correlation. Carter, D'Souza, Simpkins and Simpson's (2010) three stage least squares estimation bears similarities to Arellano-Bond, in that they claim lagged values of the performance variables are used as instruments. However, they do not make explicit what is being instrumented, and the reporting of their results suggests that (one period) lagged performance is being used only as an explanatory variable.

Wintoki, Linck and Netter (2012) appear to have inspired the application of Arellano-Bond to the topic of gender and boards in Gregory-Smith, Main and O'Reilly (2014) and Pathan and Faff (2013). Both papers implement the Arellano and Bond (1991) test for serial correlation and the Sargan (1958) and Hansen (1982) tests of over-identifying restrictions – but only Pathan and Faff (2013) raise concerns about misspecification. They find a positive and statistically significant relationship between female representation and financial performance in the period 1997-2002, but not in the other periods they test. Conversely Gregory-Smith, Main and O'Reilly (2014) do not find female representation significant. They note that the insignificant link between performance and female representation they derive using the Arellano-Bond estimator is robust to a range of alternative models.

4.2 Studies that plausibly address endogeneity

Excluding event studies, we identify nine studies that rigorously attempt to address endogeneity. Of these, one finds an outright positive link between female board representation and financial performance, two find a mostly positive (but mixed) relationship, three find no significant relationship, and three find a negative relationship. Two of the negative results are from Norway, one of which considers the implementation of Norway's quota in 2003. Specifically, the ten studies are: Adams and Ferreira (2009); Ahern and Dittmar (2012); Anderson, Reeb and Upadhyay (2011); Bøhren and Strøm (2010); Carter, D'Souza, Simpkins and Simpson (2010); Dale-Olsen, Schøne and Verner (2013); Gregory-Smith, Main and O'Reilly (2014); Pathan and Faff (2013) and Vafaei, Ahmed and Mather (2015).

We summarise the key aspects of these studies in Table 1. Overall, there is no clear relationship between female board representation and firm performance.

TABLE 1
SUMMARY OF STUDIES THAT ATTEMPT TO ADDRESS ENDOGENEITY

Study	Sample	Methods	Findings
Adams and Ferreira (2009)	US. SP500, SP MidCap, SP SmallCap. 1,939 firms. N = 8,253	IV, firm fixed effects (and Arellano-Bond)	Negative
Ahern and Dittmar (2012)	Norway. 248 firms. N=1,230.	IV, firm fixed effects	Negative
Anderson, Reeb and Upadhyay (2011)	US. Russell 1000 industrial. 615 firms. N=1,230.	2SLS	Positive (but with some insignificant results)
Bøhren and Strøm (2010)	Norway. All OSE non-financial firms – 229 firms, N=1,290	2SLS, firm fixed effects	Negative
Carter, D'Souza, Simpkins and Simpson (2010)	US. SP500. 641 firms. N=2,563.	3SLS, firm fixed effects	Insignificant
Dale-Olsen, Schone, and Verner (2013)	Norway. PLCs (128 firms) and LTDs (36,924 firms)	Difference-in-Difference	Insignificant
Gregory-Smith, Main and O'Reilly (2014)	UK. FTSE350. 1,983 firms. N=11,515.	Arellano-Bond (primarily)	Insignificant
Pathan and Faff (2013)	US. Top 300 bank holding companies. 212 firms. N=2,640.	Arellano-Bond	Positive (but with some insignificant results)
Vafaei, Ahmed and Mather (2015)	Australia. ASX500. 224 firms. N=1,101.	IV and OLS	Positive

5. Event studies

This section introduces the event study methodology and key issues in the event study literature. We then review event studies pertaining to female board appointments. In the next section, we present the findings of our own event study using Australian data.

5.1 Event studies

Another approach to addressing endogeneity is to use an event study methodology. This is done by analysing the returns generated by a specific announcement (Hagendorff and Keasey 2012). Market returns across the days immediately before and after the announcement are compared with an estimated market model over a prior period, and a cumulative abnormal return (CAR) is calculated across the event window. Farrell and Hersch (2005) use this methodology to examine the impact of female directors' appointment on market-based performance. They find no significantly positive or negative CAR in the case of women's appointments to a board, even when there were previously no female board members.

It is important to note that event studies do not examine the effect that female directors have had on performance. Instead, event studies consider the effects that market participants believe female director appointment announcements will have in future. Most event studies draw on very small samples. We identified only two event studies — the paper on Singapore by Kang, Ding and Charoenwong (2010) and a study from Australia by Adams, Gray and Nowland (2011) — that have found a significant effect for female director appointment announcements. However both papers suffer from methodological issues which have been well-documented in the event study literature. See further discussion and our event study analysis below.

5.1.1 Event studies and the estimation of the market model

The basic 'market model' method employed by Fama, Fisher, Jensen and Roll (1969) still forms the basis of most event studies. This method regresses a security's returns against that of a market index in order to generate a prediction of the return on the date of a specific event. If the security's actual performance around the event date is 'abnormal', in that it is significantly higher or lower than the regression model predicts, it suggests the event has had an impact on the market's perception of the firm's underlying value.

The basic market model is expressed as (MacKinlay, 1997):

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \qquad (1)$$

$$E(\varepsilon_{it}) = 0$$

$$var(\varepsilon_{it}) = \sigma^2_{\varepsilon_i}$$

where:

- i indexes securities
- t is time measured in event time, and equals the number of trading days before $(-)$ or after $(+)$ the event $(t = 0)$
- R_{it} is the return of security i at time t
- R_{mt} is the overall market return at time t
- α_i and β_i are parameters to be estimated
- ε_{it} captures all other unobservable effects not associated with R_{mt}
- $\sigma^2_{\varepsilon_i}$ is also a parameter to be estimated, usually by assuming constant variance across firms.

Returns are generally calculated in log form. Following Fama, Fisher, Jensen and Roll (1969), dividends are included in return calculations on the ex-dividend date and closing prices are adjusted for factors such as share splits. Returns are therefore generally calculated as:

(1) $$R_{it} = ln\left[(adj.closing\ price_{it} + dividend_{it})/(adj.closing\ price_{i,t-1})\right]$$

The value of R_{it} approximates the percentage return, especially when the return is relatively small. For instance, a 1 per cent positive return generates R_{it}= 0.00995 (0.995 per cent).

5.1.2 Female board representation and financial performance – event study literature

As noted above, we only find one event study that finds a positive effect of the appointment of a female board member on a firm's share price in the countries that we cover in this review. Adams, Gray and Nowland (2011) examine cumulative abnormal returns (CAR) for a variety of event windows. They use a market model based on the top 500 companies in the Australian ASX. They compare male and female appointments and show that the market reacts positively to female appointments and does not react to male appointments. Their data cover the period 2004-2006 and their largest sample of female board appointments includes 67 events. In section 6 below, we conduct a similar event study with a substantially larger sample and show that while we can find a positive effect of female board appointments, this effect is not robust to non-parametric testing nor to the exclusion of a very small number of outliers. Adams, Gray and Nowland (2011) only use a t-test and the Wilcoxon test, both of which have been shown to be problematic in the literature. We discuss these problems further below.

A key limitation of previous event studies on this topic is small sample sizes. Farrell and Hersch (2005) implemented the first event study dealing explicitly with performance and female board representation. They fail to find a significant result, even when a female is appointed to a previously all-male board. However, their sample consists of only 111 announcements, and only 51 appointments to previously all-male boards. Kang, Ding and Charoenwong (2010) utilise a sample of 53 announcements. Although they find a statistically significant, positive impact on CAR across the day of and day after an announcement, their use of a value-weighted index along with the standard t-test without any robustness testing suggests the finding may actually be attributable to the non-normality of abnormal returns.

Although Brown and Warner (1985) suggest that a sample size of 50 can be sufficient for many parametric tests, Corrado and Truong (2008) demonstrate that this will be insufficient in many contexts. As such, Ding and Charoenwong's (2013) sample of 34 female board appointments in Singapore is likely to be too small. Lee and James (2007) restrict their analysis to the announcement of 'top executive' appointments including CEO, CFO, COO, President, and Executive Vice President. Their sample size is 17 female CEOs and 69 other female executives in the US. The final sample across both genders is 1,624. They find that both male and female CEO appointments lead to significantly negative share price effects, but that female appointments are significantly more negative. No significant effect is detected for male or female non-CEO appointments.

Nguyen, Hagendorff and Eshraghi (2015a) have a particularly small sample. Of the 252 appointments to US banks, 15 were female. They do not detect a significant effect of appointee's gender on share price. They do, however, detect a significantly positive cumulative abnormal return for new appointments regardless of gender across a five day event window. The authors' justification of the event window length is that:

> "investors require time to do their research on the appointee before they can accurately evaluate the appointee's impact on bank performance. Thus, this is likely to take a couple of days until a reliable and market price-moving assessment can be made."

Five days may be a long time in the share market and is certainly longer than a 'couple of days'. It is unclear whether a statistically significant result would occur with a shorter window. With a window length of five days, bias caused by serial dependence becomes a major issue. Cowan (1993), for instance, finds that the Patell test statistic is overstated by 0.7 per cent with a five-day event window and 240-day estimation window. Given that the Z statistic of 1.96 in

Nguyen, Hagendorff and Eshraghi (2015a) is only barely significant at the 5 per cent level, and the positive Wilcoxon test they report, it is unclear how solid their results are. Their relatively even split between positive and negative cumulative abnormal returns (53.2 per cent to 46.8 per cent) and the much lower median (0.24 per cent) than mean (0.99 per cent) cumulative abnormal returns indicate that outliers may be driving the results.

In addition to small sample sizes, there are four main shortcomings in the existing literature applying event studies to the question of the impact of female board membership on firm performance. First, the measurement of market returns is not always provided and some measures generate methodological problems. Some authors—e.g. Lee and James (2007) and Farrell and Hersch (2005)--do not detail their measure of market returns. Although Nguyen, Hagendorff and Eshraghi (2015a) use an equal-weighted Center for Research in Security Prices (CRSP) index in robustness testing, many authors use value-weighted indices, and Brown and Warner (1980) showed that value-weighted indices can result in severe misspecification. The recommendation to use equal-weighted indices, which are less affected by non-normal returns, has been consistent and strongly emphasised in the literature since Campbell and Wasley (1993). See also the discussion in Corrado and Truong (2008).

Second, there are a wide variety of estimation and event windows used, with the former ranging from 100 to 255 days, and the latter from 1 to 12 days. The choice of event window length is '[p]ossibly the most crucial research design issue ... in an event study' (McWilliams and Siegel, 1997). Although a value of n=250 days is usually chosen to correspond approximately to the number of trading days in a calendar year (Corrado 2011), a shorter estimation window may still be appropriate. The 12-day event window used by Hagendorff and Keasey (2012) is especially concerning. Other authors generally do not provide an explanation of their chosen window length.

Third, many studies only use parametric testing and two studies (Farrell and Hersch (2005) and Kang, Ding and Charoenwong (2010)) use the standard t-test without additional testing. Corrado and Truong (2008) show that the assumptions underlying parametric tests may not hold in some markets or under 'stressed' conditions, such as event clustering or event-induced variance. Parametric tests are not valid when a value-weighted index is used. Corrado (1989) and Corrado and Truong (2008) emphasise the importance of robustness testing with non-parametric tests; see also Corrado and Zivney (1992). It may also be problematic to employ the Wilcoxon test, as is done by Adams, Gray and Nowland (2011) and Nguyen, Hagendorff and Eshraghi (2015a), because it is known to over-reject the null.

Finally, Dimson and Marsh (1986) describe the so-called 'size effect', where an event study that focuses on smaller firms is likely to show positive abnormal returns relative to the market index. In the literature we have reviewed, accounting for the 'size effect' is generally performed through subsequent regression, rather than in assessing the validity of a finding of a significant cumulative abnormal return. Exceptions are Kang, Ding and Charoenwong (2010) who regress cumulative abnormal returns against assets, while Lee and James (2007) and Nguyen, Hagendorff and Eshraghi (2015a) regress cumulative abnormal returns against log assets. Below, we examine the effect of firm size through a regression approach.

One event study has looked at the effect of the announcement of a gender quota (rather than the actual appointment of a female board member). Greene, Intintoli and Kahle (2020) examine the effect of the 2018 California Senate Bill that required public companies headquartered in California to have at least one female director by the end of 2019 and at least two by the end of 2021. They find that the announcement of the quota caused returns to fall 1.2 per cent, and the effect was more negative when a greater number of females needed to be added to a firm's board. However, the authors find no effect on the day the Bill was introduced into the Senate, passed the Senate, or passed the Assembly. It is difficult to understand why the *signing* of the Bill had an effect on equity prices but that the Bill passing the earlier hurdles did not. This is especially puzzling given the Bill received considerable media attention prior to being signed (NBC 2018). This raises the question of whether there is some other factor driving the observed results.

6 *Event study using new Australian data*

In our literature review, we found that most event studies draw on very small samples, often with fewer than 50 appointments. Further, most use value-weighted indices when calculating market returns and many do not employ non-parametric testing. With these points in mind, we begin building a large dataset covering appointments to Australian boards. Our data cover all female appointments to boards between 2000 and 2016, irrespective of the prior board composition. We also create a matched sample of male appointments to test whether male and female appointments generate different impacts on company returns.

6.1 Data

Trading volumes and 'adjusted'[3] share prices are from Thomson Reuters Datastream. Market returns are calculated from the Morgan Stanley Capital International (MSCI) Australia Equal Weighted Price Index, sourced from Datastream. Time series firm financial data (such as assets by year) were obtained from Morningstar Datanalysis. Director appointment and termination data were obtained from Thomson Reuters Australia's Connect 4 Boardroom database. These data include each director's appointment date, name, gender, position title and age (although some data points are missing). Company announcement data are from the Thomson Reuters Australia's Connect 4 Company Announcement database but are also publicly available in searchable form on the ASX website.[4]

It is worth noting that accurate tracing of individual appointments must be done manually. This is because each announcement must be read in full to confirm it contains no confounding events, such as a director resignation. Each manually reviewed announcement has only a small chance of resulting in a useable event. Not only does the announcement need to be valid in itself (i.e. contain a single announcement of a female board appointment without any confounding news), it needs to survive subsequent testing for nearby confounding events, and needs to relate to a security that has sufficient returns data to allow estimation of the market model, including returns data for every day during the event window. As a result, many announcements cannot be used in an event study. These are the issues that often generate small samples in event studies.

The number of female board appointments in Connect 4 in our 2000-2016 data window is 6,152. Table 2 provides a summary of these appointments.

[3] 'Adjusted prices' account for dilution factors, such as 'bonus issues, rights issues and reconstructions', allowing 'comparability of current figures with those of previous years' (https://www.morningstar.com.au/s/documents/Glossary_morningstar.pdf).

[4] All data described in this section and in section 7, as well as our STATA code for reproducing the results, are available on our web page. Weblink to be provided upon publication.

TABLE 2

FEMALE AND MALE APPOINTMENT TYPE

	Female	Male
Non-Executive Director	1,721	19,508
Company Secretary	1,600	4,443
CFO	300	2,434
Company Secretary/CFO	294	2,042
Divisional Head	277	2,893
Company Secretary/Legal Counsel	276	401
Executive Director	241	4,118
Human Resources	174	131
MD/CEO	151	4,570
Chairman	131	5,626
Alternate Director	109	843
O/Seas Divisional/subsidiary head	69	936
COO	68	971
Chief Marketing Officer	54	209
CEO	51	1,252
Other	626	7,691
Total	**6,142**	**58,068**

Data source: Thomson Reuters Australia's Connect 4 Boardroom database

We focus on the non-executive director (NED) positions which has a large sample size for both men and women. We manually trace all 1,721 female NED appointments back to their respective announcements. Almost two-thirds of female NED appointments reviewed were unsuitable for use in the event study – with only 588 potentially valid, single announcements. Reasons for excluding the other observations are shown in Table 3.

TABLE 3

RESULTS OF MANUAL REVIEW OF FEMALE NED APPOINTMENTS

	Female
Single announcement (usable event)	588
Multiple announcements	710
Appointment prior to listing	289
Could not be located	119
Change of director's position only	13
Error in data	2
Total	**1,721**

These 588 announcements were then checked for the presence of confounding events. Defining events that are 'confounding' is not straightforward and a range of approaches have been taken in the past. Farrell and Hersch (2005) consider mergers, acquisitions, dividend changes, stock repurchases, stock splits, and earnings announcements during a 5-day window around the event day to be confounding. Nguyen, Hagendorff, and Eshraghi (2015a) refer only to simultaneous 'other corporate events'; Kang, Ding and Charoenwong (2010) exclude 'announcements with confounding events' but do not define them.

In Australia, the ASX reports some announcements as 'market sensitive'. This designation is based on the advice of the issuing firm. We note that director appointments are rarely marked as 'market sensitive'. We decided to eliminate events if the security went ex-dividend, if 'market sensitive' announcements were made or if any other announcements concerning board composition were made. For all of these, we consider the event to be confounding if it occurs within two days of the event window.

6.2 Initial analysis

We set the event window at a single trading day (*T*=0). In keeping with convention, the estimation window was set to 250 trading days, with securities to be excluded if fewer than 100 observations were available in the estimation window. Eliminating announcements due to confounding events, ex-dividend, or insufficient observations reduced the initial pool of 1,721 potential events to 302 usable events as documented in Table 4. We refer to this sample as our 'base sample'. Below, we consider other samples where we vary the event window.

TABLE 4
DETERMINING THE BASE SAMPLE (EVENT WINDOW: T=O)

Description	Attrition	Usable sample
All female appointments		1,721
Potentially usable events (see Table 3)	1,133	588
Confounding events	165	423
Price data not available in Datastream	23	400
Missing overall market return in event window	1	399
Missing return in event window	67	332
Ex-dividend in event window	7	325
Insufficient observations in estimation window	23	302

Base sample defines the event window as a single trading day (*T*=0)

The requirements for an event to be usable for an event study frequently result in large reductions in sample size of the kind we show in Table 4. If we compare the 302 observations in the final estimation sample to the initial 1,721 female appointments to non-executive director board positions, they differ on two key characteristics. The first is that the 302 observations we keep are disproportionately from larger firms. About two-thirds of them are from the top two deciles of the firm asset distribution. Larger firms tend to appoint more female directors and more directors, generally, so this is not surprising. There is also a slight over-representation from the 'consumer discretionary' sector. We explore, in Table 8 below, whether there is any firm size effect in cumulative average returns for female board appointments. There is no statistically significant effect in our sample. This suggests that our results are unlikely to be affected by the 'size effect' documented by Dimson and Marsh (1986).

TABLE 5
EVENT STUDY RESULTS FOR BASE SAMPLE (EVENT WINDOW: T=0)

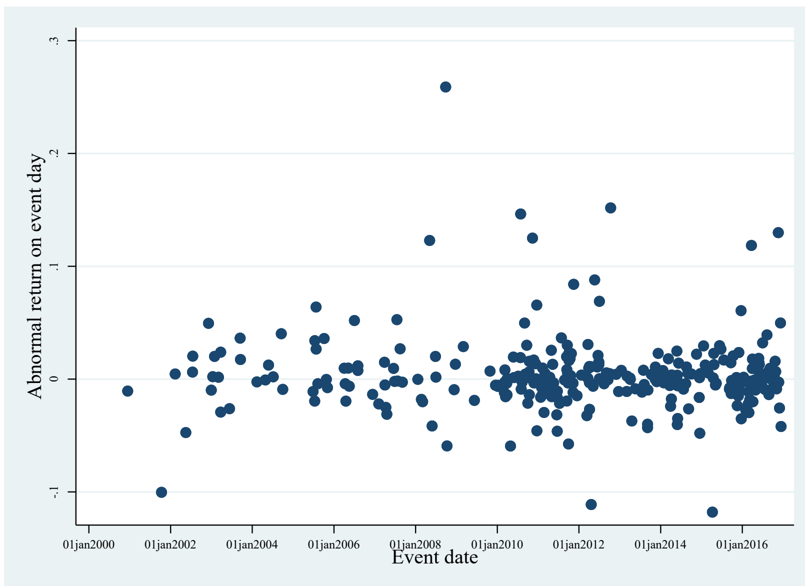
Description	Statistic
Sample size	302
Average abnormal return (event day *T*=0)	0.0033
Standard t-test	1.7597*
Portfolio return test	1.8300*
Patell test	2.1710**
BMP test	2.0567**
Generalised sign test	-0.2925
Rank test	1.3041

* Statistically significant at the 10 per cent level; ** Statistically significant at the 5 per cent level

To estimate equation (1) for the base sample, market returns data were converted to log form. A range of parametric and non-parametric test statistics were calculated using two-tailed tests. We use two-tailed tests as there are no compelling reasons to expect, *a priori*, that female appointments will have a positive or negative impact on prices. Table 5 reports these statistics for the event day (*T* = 0). The standard t-test and the portfolio return test both show a positive and significant impact of female board appointments on stock prices at the 10 per cent level. The Patell and BMP tests show a statistically positive impact at the 5 per cent level. The BMP test is another parametric test proposed by Boehmer, Musumeci and Pousen (1991) which is frequently used in event studies. Importantly, the two non-parametric tests show no statistically significant impact of female board appointments on firm share price.

It is particularly notable that the generalised sign test is negative when all other test statistics are positive, as it indicates that the positive parametric results are driven by outliers. A scatterplot (see Figure 1) of abnormal returns on the event day against time shows that the results are clustered around zero and most of the large outliers are positive. If we remove the two largest positive outliers and re-do the analysis of Table 5, all the parametric and non-parametric tests return statistically insignificant results.

FIGURE 1
ABNORMAL RETURNS BY FIRM ACROSS TIME FOR BASE SAMPLE



The literature highlights the sensitivity of the parametric tests to outliers and the risks of relying on those tests to draw conclusions. Our results here provide an illustration of this issue. The graphical evidence, the non-parametric tests and the parametric tests without the two largest outliers point to a strong conclusion of no effect of a female board appointment on a firm's share price.

If we broaden the definition of confounding events to include any company announcement apart from initial directors' interest notices, our usable sample shrinks to 152 events. We replicate the statistical tests of Table 5 on this reduced sample. The results are all statistically insignificant and we do not report them here.[5]

[5] These results and other results that are not reported are available from the authors upon request.

6.3 Varying the event window

Next, we examine whether varying the event window has any impact on our results. We first enlarge the event window beyond the day itself and remove events with confounding events within 2 days either side of the event window. As a result, the sample for each successive widening of the event window becomes smaller because we capture a larger number of confounding events. Our choice of windows is motivated by event windows which have been considered in the literature. Table 6, which shows a range of tests using different event windows, shows no consistent, statistically significant results. $\overline{CAR}$ is cumulative average return estimated over the event window. [6]

TABLE 6

RESULTS FOR DIFFERENT EVENT WINDOWS (CONFOUNDING EVENTS NARROWLY DEFINED)

Sample label	1B	1C	1D	1E	1F	1G	1H
Window (T=)	-1 to 0	0 to +1	-1 to +1	-2 to +2	0 to +2	0 to +3	0 to +4
Sample size	283	283	266	235	270	258	241
$\overline{CAR}$	0.0037	0.0046	0.0042	0.0012	0.0000	-0.0015	-0.0049
Standard t-test	1.3818	1.6802*	1.2586	0.2699	-0.0050	-0.3692	-1.0739
Portfolio return test	1.3595	1.7910*	1.2590	0.2642	-0.0053	-0.3882	-1.1188
Patell test	0.5293	1.3991	0.2728	-0.0068	0.1501	-0.2178	-1.0501
BMP test	0.4914	1.3578	0.2490	-0.0092	0.1534	-0.2394	-1.1447
Generalised sign test	-0.0478	1.0274	-0.7490	-0.5379	-0.1270	-1.2577	-1.0021
Rank test	-0.2485	0.8883	-0.3250	-0.4845	-0.3458	-0.7660	-1.5904

* Statistically significant at the 10 per cent level; ** Statistically significant at 5 per cent
$\overline{CAR}$ is cumulative average return estimated over the event window

The same process of varying event windows was also conducted using the broader definition of confounding events, that is, any company announcement with the exception of initial directors' interest notices. The results of this analysis are presented in Table 7. The results presented in Tables 6 and 7 show almost no statistically significant results. The only statistically significant result is for the event window defined between the day of the announcement and including the following three days. The Generalised Sign Test is statistically negative at the 10 per cent level.

From randomness, one would expect at least a few statistically significant tests even if the true value were zero. The lack of significance suggests a strong zero result—female non-executive director board appointments in Australia do not have an impact on a firm's share price.

[6] Restricting the analyses in Tables 6 and 7 to the smallest common sample of column 1H in Table 6 produces qualitatively similar results. The differences in results across columns are not being affected by changes in sample sizes per se, but rather by changes to the definition of the event window and the impact on confounding events.

TABLE 7
RESULTS FOR DIFFERENT EVENT WINDOWS (CONFOUNDING EVENTS BROADLY DEFINED)

Sample label	2B	2C	2D	2E	2F	2G	2H
Sample size	129	126	110	74	99	89	74
Window (T=)	-1 to 0	0 to +1	-1 to +1	-2 to +2	0 to +2	0 to +3	0 to +4
$\overline{CAR}$	0.0045	0.0051	0.0062	0.0055	0.0047	-0.0011	-0.0023
Standard t-test	1.1174	1.2146	1.1919	0.6680	0.7926	-0.1466	-0.2509
Portfolio return test	1.1171	1.3025	1.1960	0.6754	0.8766	-0.1554	-0.2617
Patell test	0.6082	0.9731	0.5364	0.4737	0.6312	-0.1888	-0.6861
BMP test	0.6657	1.0042	0.5616	0.5758	0.6975	-0.2110	-0.8698
Generalised sign test	0.2784	0.1855	-0.2331	0.7149	0.1922	-1.9302*	-0.4238
Rank test	0.4933	0.4664	0.1949	0.0140	0.2476	-0.9943	-1.2406

* Statistically significant at the 10 per cent level
$\overline{CAR}$ is cumulative average return estimated over the event window

6.4 Further analysis

Results from the previous section do not provide consistent evidence that the announcement of female board appointments affects security prices. Only one event window (T=0) provides any statistically significant result at the 5 per cent level, but it is heavily influence by outliers and not robust to non-parametric testing. Even for the tests that do provide a significant result, the significance evaporates once an alternative definition of confounding events is used.

Nonetheless, the literature also suggests that firms may differ systematically in respect to their propensity to appoint female directors or in respect of the response of their share price to such appointments. The previous literature has identified a range of relevant firm characteristics including: firm size; industry; board size; whether the board was previously all male; and year of event.

To investigate this hypothesis, we estimate a regression of the cumulative average return ($\overline{CAR}$) for each female board appointment on these factors. If any of these factors are statistically significant, this indicates that the return for a female board appointment is differing by firm characteristics. As the base sample from Table 5 is the only sample to return any statistically significant test results at the 5 per cent, this is the sample we use for this regression.[7] The results provide no indication that the market reaction to the appointment of a new female director differs according to any of the abovementioned firm characteristics (see Table 8). All of the coefficients are statistically indistinguishable from zero.

Overall, the results of sections 6.2 – 6.4 indicate that the announcement of the appointment of a female non-executive director to a firm's board does not have a statistically significant effect on a firm's security price.

[7] Note that the sample in Table 8 is slightly smaller than Table 5 because information on board size is missing prior to 2004 and some firms have missing asset information.

TABLE 8
REGRESSION OF CUMULATIVE AVERAGE RETURN ($\overline{CAR}$) ON VARIOUS FACTORS
BASE SAMPLE

ln(assets)	-0.0021
	(0.0013)
ln(boardsize)	0.0021
	(0.0130)
All male board	-0.0070
	(0.0048)
Year dummies	Included
Constant	0.0461**
	(0.0206)
Observations	283
Prob > F	0.4479
R^2	0.0563

"All male board" is an indicator equal to one if the board was previously entirely male.
Year dummy results not reported. All year dummies are insignificant at the 10 per cent level.
** Statistically significant at 5 per cent

6.5 Matched sample by male appointments

Even though we do not find that security prices respond to female board appointments, it is possible that announcements concerning *male* appointments could have a statistically significant effect. As the female sample is heavily skewed along various dimensions – most notably towards larger firms – we generate a matched sample and run a second event study with the same general settings as for the female base sample.

TABLE 9
IDENTIFYING USABLE EVENTS IN MATCHED MALE SAMPLE

Description	**Attrition**	**Running total**
Male appointments inspected		1,451
Multiple announcements	677	774
Appointment prior to listing	179	595
Could not be located	154	441
Confounding events	155	286
Price data not available in Datastream	16	240
Missing return in event window	6	280
Missing market return in event window	39	241
Ex-dividend in event window	3	238
Insufficient observations in estimation window	11	227

The criteria used for matching were: total assets by decile, calendar year of appointment and Global Industry Classification Standard (GICS) sector. The 302 female events contained 197 unique combinations of these 3 criteria, and the available male appointment data included a potential pool of 2,883 matching male events, with most female events generating multiple male matches. The male matches for each female event were then reviewed in random order until a valid match was found – that is, one which was not disqualified against other criteria such as multiple events, confounding events, or missing returns. We examined 1,451 male events that could potentially be matched to a female event.[8] As shown in Table 9, 227 of the 302 female events were able to be matched in this manner. For 75 of the female events, none of the matching male events provided a usable matched observation.

Event study results from this matched male sample are all statistically insignificant (Table 10).

TABLE 10
EVENT STUDY RESULTS FOR MATCHED MALE SAMPLE

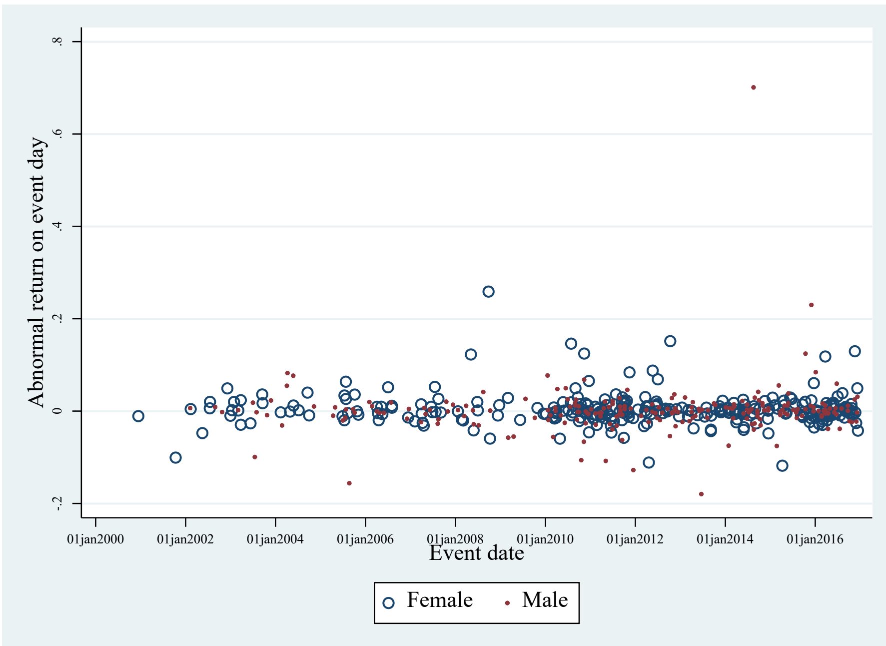
Description	Statistic
Final sample size	227
Average abnormal return (event day, *T*=0)	0.0018
Standard t-test	0.7682
Portfolio return test	0.7981
Patell test	0.2026
BMP test	0.1789
Generalised sign test	1.0347
Rank test	0.8028

We replicated this analysis after removing outliers. Their removal does not materially alter the results. Regressions of cumulative abnormal returns on ln(assets) provided no significant results, regardless of whether outliers were excluded or included. A direct t-test comparing abnormal returns between the matched male and female samples (whether using only the matched 227 events in each, or all 302 female and the matched 227 male events) did not reveal a statistically significant difference between the two.

A graphical comparison of the (full) female and the matched male sample reinforces the fact that abnormal returns on the day of both female and male appointment announcements tend to cluster around zero.

[8] The unused 1,432 matching male events could not be matched to any of the 85 female events for which we were unable to find a male match. Some of them were potentially matchable to female events but we only used one valid, matching male event per female event.

FIGURE 2
ABNORMAL RETURNS BY FIRM ACROSS TIME FOR FEMALE AND MALE SAMPLES



6.6 Discussion

One of the most striking aspects of the results presented above is the diversity of test outcomes within particular samples. It is common for test results to be highly significant using certain tests (particularly the t-test) but insignificant against others (particularly the non-parametric tests). Although Brown and Warner (1985) began to cast doubt on the reliability of the t-test under a range of circumstances in the 1980s (Binder 1998), it has still been commonly used in the recent literature on female board appointments, including in Lee and James (2007), Kang, Ding and Charoenwong (2010) and Adams, Gray and Nowland (2011).

Evidence of an impact of female board appointments on company share prices is not robust to non-parametric testing or removal of even a small number of outliers. None of the test results for the male sample were statistically significant. There is no evidence of a gender-differentiated impact of a new non-executive director (NED) appointment on a firm's security price. Overall, the results strongly suggest that, in Australia, there is no effect on share price of the appointment of a female NED to ASX-listed companies.

7. Regression Approaches: Replication Studies and New Australian Estimates

7.1 Introduction

Section 6 used an event study methodology to focus on the immediate share price effect of female board *appointments*. Although the sample size was much larger than most previous event studies, we were unable to detect a statistically significant effect of the appointment of a female board member on a firm's share price. Furthermore, there was no difference in the effect of appointing a new male or a new female member to a firm's board.

In this section, we return to whether the presence of women on boards or the number of women on boards has any effect on firm *outcomes*. We explore the regression-based approaches discussed in section 4, keeping in mind the difficulty in controlling for the potential endogeneity of female presence on boards and firm outcomes. As in section 6, we assemble a large dataset for Australia that allows us to replicate what others have done and to conduct our own analysis.

We begin by briefly describing our data and its sources. We then use this data to attempt to replicate two previous studies. We make modifications to our base data in line with the modifications made by the authors of those studies in order to match, as best we can, variable definitions, treatment of missing values and other sample exclusions. Finally, we use our full data to estimate a range of models to determine whether there is evidence that female board membership affects firm performance. We consider a large range of possible outcomes and an extensive set of control variables. We attempt to control for endogeneity.

7.2 Data

In what follows, we consider five different measures of firm performance: return on assets (ROA); return on equity (ROE); the ratio of cash flow from operations to total assets (CFO/TA or CFOTA); Tobin's Q (Q) and its natural logarithm (ln(Q)). ROA and ROE are taken from Morningstar's Datanalysis. CFO/TA and Q are calculated directly from Datanalysis fields consistent with the description provided by Vafaei, Ahmed and Mather (2015)—see section 7.3 below. Director appointment and departure data are from the Thomson Reuters Australia's Connect4 Boardroom database. These data include each director's appointment and departure date, name, gender, position title and age, and enable calculation of variables such as *Female%*, the percentage of female board members at the end of the relevant financial year. Firm financial data including assets and market capitalisation are taken from Datanalysis. Information on shareholders' significant holdings is obtained from Securities Industry Research Centre of Asia-Pacific's (SIRCA) data. *Industrysegments* is the number of different industries that the firm operates in and comes from Datanalysis. We use Connect4 for information on chief executive officers (CEO) and all board directors. Tenure is the average tenure of all directors. CEO Turnover is a dummy variable equal to one if the CEO has changed in the last 12 months. *Indep%* is the percentage of board members who are independent.

Table 11 provides descriptive statistics for this data for firms in the ASX 200, the ASX 500 and all firms in our database. [9] Table 11 excludes a small number of firms with values of variables which produce implausible values for

[9] All data described in this section and in section 6, as well as our STATA code for reproducing the results, are available on our web page. Weblink to be provided upon publication.

TABLE 11 – SUMMARY STATISTICS (ASX200, ASX500, ALL FIRMS)

		ASX200						ASX500						All firms					
		Obs	Mean	Median	Std. Dev.	Min	Max	Obs	Mean	Median	Std. Dev.	Min	Max	Obs	Mean	Median	Std. Dev.	Min	Max
Performance	ROA	2,381	0.07	0.06	0.09	-0.55	1.80	5,920	0.04	0.06	0.19	-4.79	1.80	15,004	-0.04	0.02	0.26	-6.53	2.40
	ROE	2,379	0.14	0.11	0.22	-1.32	4.78	5,885	0.07	0.09	0.63	-32.99	9.53	14,881	-0.07	0.02	1.64	-33.57	164.59
	ln(Q)	2,380	0.48	0.32	0.58	-0.79	3.88	5,916	0.49	0.32	0.64	-1.36	3.88	14,980	0.18	0.10	0.78	-11.35	3.88
	Q	2,380	2.03	1.38	2.42	0.45	48.57	5,916	2.13	1.38	2.38	0.26	48.57	14,980	1.62	1.11	1.77	0.00	48.57
	CFOTA	2,377	0.09	0.07	0.11	-0.78	2.57	5,913	0.06	0.06	0.16	-1.83	2.57	14,759	0.00	0.02	0.19	-5.77	2.57
Female	%	2,294	13%	13%	12%	0%	63%	5,668	9%	0%	12%	0%	71%	14,098	6%	0%	11%	0%	80%
	Dummy	2,294	0.66	1	0.48	0	1	5,668	0.46	0	0.50	0	1	14,098	0.29	0	0.45	0	1
	Blau	2,294	0.20	0.22	0.16	0.00	0.50	5,668	0.14	0.00	0.16	0.00	0.50	14,098	0.09	0.00	0.15	0.00	0.50
Board size	Number	2,294	7.63	7	1.94	3	18	5,668	6.48	6	1.94	1	18	14,098	5.35	5	1.83	1	17
	ln(number)	2,294	2.00	1.95	0.25	1.10	2.89	5,668	1.82	1.79	0.30	0.00	2.89	14,098	1.62	1.61	0.33	0.00	2.89
Firm Size	ln(assets)	2,381	8.07	7.94	1.61	3.57	13.77	5,920	6.53	6.32	1.85	2.33	13.77	15,005	4.81	4.33	1.98	2.30	13.77
	ln(mktcap)	2,381	21.84	21.55	1.16	20.08	26.27	5,920	20.37	19.99	1.47	18.27	26.27	14,704	18.38	18.09	2.04	11.52	26.27
Leverage	Leverage	2,265	0.32	0.19	0.58	0	9.52	5,521	0.32	0.12	1.01	0	36.27	12,228	0.42	0.04	2.33	0	107.32
	Financial leverage	2,299	2.51	1.89	3.59	1.00	115.80	5,776	2.17	1.69	2.65	0	115.80	14,747	2.01	1.46	2.85	0	115.80
Other controls	Indep%	2,294	61%	67%	25%	0%	100%	5,668	53%	60%	26%	0%	100%	14,098	44%	50%	27%	0%	100%
	Tenure	2,244	4.98	4.50	2.67	0.03	19.11	5,556	4.85	4.22	3.09	0.02	24.82	13,857	4.46	3.60	3.47	0.00	29.55
	CEO tenure	2,285	5.15	3.68	5.08	0	34.62	5,631	5.24	3.41	5.67	0	44.98	13,949	4.89	2.95	5.75	0	44.98
	CEO gender	2,293	0.04	0	0.19	0	1	5,650	0.03	0	0.17	0	1	14,000	0.03	0	0.16	0	1
	CEO turnover	2,285	0.15	0	0.36	0	1	5,631	0.17	0.00	0.38	0	1	13,949	0.21	0	0.40	0	1
	Industry segments	2,377	2.86	3	1.52	1	8	5,878	2.36	2	1.39	1	8	14,905	1.95	1	1.26	1	10
	Firm age	2,381	20.21	14	18.97	1	136	5,920	16.17	12.00	15.16	1	136	15,012	13.83	10	12.64	1	136

ROA, ROE or Q. For instance, there is a firm with assets of $58 and net loss after tax of over $1 million, giving a negative ROA of 1.9 million per cent. These low levels of assets – including one firm reporting assets valued at $1, also affect measures of Tobin's Q. We drop firms that report equity below $2 million, as many of these firms have unreasonable ROE values. We drop firms from our sample if they have assets of less than $10 million because we found that this threshold produced ROA values that were within a reasonable range. Finally, we exclude firms with zero or negative liabilities.[10]

Our data cover 2,483 firms for the period 2005 – 2016 inclusive.

7.3 Analysis and replication of Vafaei, Ahmed and Mather (2015)

We attempted to replicate several papers as we conducted our literature review. We were almost never able to replicate the results that authors claimed and data and computer code were never available. The much publicised "replication crisis" (see for example the discussion in Camerer et al., 2018) in the social sciences seems particularly present in this area of study.

Here we present our attempts to replicate two papers which have been widely cited in the literature. The first is a paper by Vafaei et al. (2015) using Australian data and the second is a paper by Adams and Ferreira (2009) using data from the US. The dataset that we have compiled allow us to check if the results of Vafaei et al. (2015) are replicable and whether the conclusions of Adams and Ferreira (2009) apply in Australia.

We provide fairly extensive descriptions of our attempts to replicate these papers because we think that it is instructive for understanding some of the divergent results in the literature and in understanding how difficult replication can be. There are important decisions that need to be made in how to measure outcomes, how to measure female presence on boards, what variables to include as co-variates and how to measure them, how to deal with potential outliers and how to attempt to deal with endogeneity. All of this then needs to be communicated clearly by researchers in their publication.

Vafaei, Ahmed and Mather (2015), in their paper drawing on Australian data, first attempt to determine a relationship between performance and female board representation in Australia through a panel regression. They use four measures of performance (ROA, ROE, Q, CFO/TA) and two measures of female board representation.

[10] Vafaei, Ahmed and Mather (2015) include some of these observations but deal with their potential influence on regression results through Winsorisation. When we replicate their paper in the next section, we include these values and treat them as the authors did. See section 7.3.3 below.

Their model is defined as follows:

(2) $$Performance_{it} = \alpha_0 + \alpha_1 Female_{it} + \alpha_2 BHO_{it} + \alpha_3 MSO_{it} + \alpha_4 BoardSize_{it} + \alpha_5 FirmAge_{it} + \alpha_6 Leverage_{it} + \alpha_7 FirmSize_{it} + \alpha_8 NED\%_{it} + \alpha_9 AssetsInPlace_{it} + \alpha_{10} YearDummy_{it} + \alpha_{11} IndustryDummy_{it} + \varepsilon_{it}$$

They estimate this model using ordinary least squares (OLS) and two stage least squares (2SLS). In both cases, they pool their data across multiple years (2005-2011) and they do not exploit the panel nature of the data. This yields 16 sets of estimates (4 outcome measures X two measures of female board representation X two estimation techniques) all of which yield positive, significant coefficients for *Female*. Below we discuss variable definition and construction.

7.3.1 Variable measuring female board representation

The independent variable of interest in this model, female board representation (*Female*), is defined in two ways by Vafaei, Ahmed and Mather (2015): fist, as a dummy for whether any female board members are present, and second, as a continuous variable for the proportion of women on the board.

7.3.2 Other right-hand-side variables

Block-holder ownership (*BHO)* is the percentage of shares owned by shareholders who hold more than 5% of a company's total shares at reporting date. Managerial share ownership (*MSO)* is the percentage of ordinary shares owned by directors at reporting date. *BoardSize* is the natural logarithm of the number of directors at the reporting date. *FirmAge* is the natural logarithm of the firm age (the number of years the firm has been listed on the ASX). *Leverage* is the Ratio of long-term debt to market value of equity. *FirmSize* is the natural logarithm of total assets. NED% is the ratio of the number of non-executive directors (NED) to the number of directors. *AssetsInPlace* is the ratio of the sum of inventory, property, plant and equipment to total assets. Vafaei et al. (2015) write that their IndustryDummy variable is following the ASX industry classification, but it follows the Global Industry Classification Standard (GICS). This is obvious from the number of sectors – 10 rather than 24 – and the fact that the ASX classification was not used after 2002, well before the sample period in the paper.

7.3.3 Winsorisation

Vafaei et al. (2015) winsorise the upper and lower 2 per cent of: ROA; ROE; CFO/TA; *AssetsInPlace* and *Leverage*, as well as the upper and lower 10 per cent of Q and *FirmSize*. Winsorising Q and *Firmsize* in this manner may be drastic, as it changes the values of 20 per cent of the datapoints. Vafaei et al. (2015) do not report any results that are not winsorised nor do they detail the reasoning behind the chosen thresholds. As such, we are unable to determine whether the results in Vafaei et al. (2015) are sensitive to outliers.

7.3.4 Instrumental variables

Vafaei et al. (2015) use two instruments, both of which follow Adams and Ferreira (2009). Their main instrument, henceforth referred to as IV_{male}, is the fraction of male directors who sit on other boards on which there are other female directors. The other instrument used by Adams and Ferreira (2009) henceforth referred to as IV_{all}, is the fraction of all directors (male and female) on the board who sit on other boards on which there are other female directors. IV_{all} is the main instrument used by Vafaei, Ahmed and Mather (2015).

7.3.5 Replication dataset

We take ROA and ROE directly from Datanalysis as we are unable to follow the definitions and variable construction used in Vafaei, Ahmed and Mather (2015).[11] CFO/TA and Q are calculated directly from Datanalysis fields as per Vafaei, Ahmed and Mater's (2015) description. *Female* is defined as per Vafaei et al. (2015) as both a dummy and a percentage, with data taken from the Connect4 database. *BHO* was taken from the Securities Industry Research Centre of Asia-Pacific's (SIRCA) 'substantial shareholder' table, with values greater than 100 per cent amended or dropped depending on whether sufficient clarifying information was available from annual reports. *MSO* is calculated using the number of ordinary shares explicitly listed in SIRCA's 'personal shareholdings' table. *BoardSize* is defined as the number of board members measured at the firm's balance date (which usually aligns with the end of the Australian financial year). *FirmAge, Leverage* and *firmsize* (ln(total assets)) are taken from Datanalysis. *NED%* - the proportion of non-executive directors - is calculated from the Connect4 database. These all follow the source and calculation of Vafaei et al. (2015). *AssetsInPlace* is calculated from variables in Datanalysis as per their paper, though we consider two alternative treatments of missing (" '—' " in their paper) values. The first assigns a value of zero to '—' values, while the second treats '—' as missing values.

7.3.6 Replication

We restrict the data to the years considered by Vafaei et al. 2005-2011. The sample size they report is substantially smaller than what we find. Vafaei et al. (2015) consider the 500 firms in the ASX500 over seven years from 2005 up to and including 2011. If there were no missing values this would yield 3,500 observations. They initially report having dropped 586 observations from their original sample of 3,500 due to 'missing financial data' in Datanalysis. However, later in the paper, Vafaei et al. (2015) note 1,136 missing observations for *AssetsInPlace*. In either case, this level of attrition is difficult to explain, as there are very few gaps in Datanalysis' data for larger firms. The main potential exceptions are in those variables that are required for calculating *AssetsInPlace*, because inventory and PPE are relatively often recorded as '--' (missing). But even dropping all such values yields a substantial difference in sample size, as shown in Table 12 below. Vafaei et al. (2015) also report roughly half the

[11] While we had access to the same data, correspondence with the lead author did not clarify their approach to variable construction.

expected sample size for the variables *BoardSize*, *Female*, and *NED%*. This likely reflects an absence of manual matching between the Connect 4 dataset and the Datanalysis dataset.

TABLE 12

SAMPLE SIZE COMPARISON FOR FINANCIAL (DATANALYSIS) VARIABLES

Variable	Replication sample size	Vafaei, Ahmed and Mather 2015 sample size	Difference
ROA	3,500	2,915	585
ROE	3,472	2,884	588
Q	3,498	2,914	584
CFO/TA	3,498	2,914	584
Firm size	3,500	2,915	585
Leverage	3,248	2,648	600
Firm age	3,500	3,500	-
AssetsInPlace ('--' = missing)	2,805	2,364	441
AssetsInPlace ('--' = zero)	3,500	2,364	1,136

These issues notwithstanding, the summary statistics are generally fairly similar between our replication sample and the sample employed by Vafaei et al. (2015). These data are presented in Table 13, noting that ROA, ROE, CFO/TA, AssetsInPlace, Leverage, Q and FirmSize in the replication sample are winsorised to ease comparison between the samples.

Vafaei et al. (2015) find positive, statistically significant results in all 16 versions of their regression. This includes when *Female* is measured as a dummy and as a percentage, and whether using panel OLS or 2SLS. These results are significant at the 1 per cent level for ROA, ROE and CFO/TA. *Female* is also significant for *Q* at 1 per cent for both versions of *Female* using 2SLS, but at 5 and 10 per cent respectively for the dummy and percentage versions of *Female* using OLS. It is, however, not always clear what these levels of significance mean precisely. This is because Vafaei et al. (2015) write that 'one-tailed significance is applied where the directionality of the coefficient was predicted and two-tailed where directionality was not specified'. As their main hypothesis is that the presence of women on corporate boards is positively associated with firm financial performance, tests for ROA, ROE, Q and CFO/TA are presumably all one-tailed. This presumption of a positive relationship may be misplaced given the lack of consensus in the literature.

TABLE 13
COMPARISON OF REPLICATION SAMPLE WITH VAFAEI, AHMED AND MATHER (2015) SAMPLE

	Count		Min		Max		Mean		Median		Std Deviation	
	Rep. Sample	Vafaei et al.	Rep. Sample	Vafaei et al.	Rep. Sample	Vafaei et al.	Rep. Sample	Vafaei et al.	Rep. Sample	Vafaei et al.	Rep. Sample	Vafaei et al.
ROA	3,500	2,915	-0.411	-0.489	0.317	0.292	0.046	0.0376	0.059	0.0552	0.125	0.1306
ROE	3,472	2,884	-0.659	-0.693	0.674	0.7008	0.091	0.0832	0.096	0.0895	0.218	0.226
Q	3,498	2,914	0.876	0.829	4.272	3.865	1.921	1.793	1.485	1.404	1.100	0.984
CFO/TA	3,498	2,914	-0.339	-0.317	0.424	0.424	0.062	0.0581	0.054	0.0519	0.130	0.125
Female%	3,215	1,787	0	0	0.625	0.6	0.061	0.0678	0	0	0.093	0.0928
BHO1	2,828	2,078	0	0.05	1.000	1.000	0.343	0.3928	0.313	3.559	0.251	0.223
MSO	2,423	2,770	0	0	0.971	0.994	0.137	0.1125	0.043	0.0263	0.186	0.174
Board Size	3,215	1,787	1	3	17	20	6.438	7.9	6	8	1.971	2.751
Firm Age	3,500	3,500	1	1	136	140	15.454	17.024	11	12	14.963	15.871
Leverage	3,248	2,648	-	0	2.625	3.909	0.286	0.0372	0.108	0.133	0.505	0.719
Assets[a]	3,500	2,915	68.2m	72.1m	7,261.0m	8,470.0m	1,666.0m	2,120.0m	513.7m	783.0m	2,318.0m	2,740.0m
Ln(assets)	3,500	2,915	4.222	4.284	8.890	9.044	6.385	6.679	6.242	6.662	1.509	1.523
NED%	3,215	1,779	0	0.1	1.000	0.909	0.592	0.5867	0.600	0.6	0.155	0.1413
AssetsInPlace	2,805	2,364	0	0.00067	0.992	0.877	0.351	0.355	0.327	0.331	0.270	0.271

Note: figures from Vafaei, Ahmed and Mather (2015) are reported to the number of decimal places presented in their paper. Figures from the replication sample are reported to three decimal places unless this is inappropriate (e.g. median of integer values; $ millions). These statements apply to all further comparison tables.
[a]Vafaei, Ahmed and Mather (2015) describe the total asset figures in their table (Table 4 on p. 419) as being in '(Dollar, millions)'. This cannot be true, as the maximum value they record for firm size in the table is 8,470,000,000. We therefore assumed that their table reports raw dollar figures.

Using two-sided tests, we attempted to replicate the findings of Vafaei, Ahmed and Mather (2015). We estimate regressions using a wide range of variable definitions, including different definitions of ROA and ROE. Comparing the results shows that the coefficients derived from the replication sample are always - and often by orders of magnitude - lower than those reported by Vafaei, Ahmed and Mather (2015). Indeed, most of the models in the replication sample do not yield significant results even when using one-sided tests. Fixed effects and random effects specifications have very similar results.

Table 14 compares results obtained when using the replication sample compared to those reported by Vafaei, Ahmed and Mather (2015), where *Female* is measured as a dummy and employing the OLS specification. Table 15 shows results for: IV_{all} with *Female* measured as a dummy; OLS for *Female* measured as a percentage and IV_{all} for *Female* measured as a percentage. In Table 15, we do not present the results from Vafaei et al. (2015) to save space. For the same reason we do not report year or industry dummies. The reported results are derived using *AssetsInPlace* values of '--' treated as zero, although the significance of *Female* is unchanged regardless of how these variables are measured.

In Table 15, the only firm outcome measure for which we find any evidence of a significant impact of female presence on boards is CFO/TA. We only find an effect at the 10 per cent level and only for the two instrumental variable approaches. For ten of the other twelve estimates, we find no statistically significant impact of female presence on boards on firm outcomes. Vafaei et al. (2015) find statistically significant results for all of these.

Overall, we find that following the methodology described in Vafaei et al. (2015) as closely as possible does not provide compelling evidence that female representation significantly affects firms' financial performance. Correspondence with the lead author did not resolve the discrepancies between their results and our findings.

TABLE 14

COMPARISON OF RESULTS FOR PANEL OLS (FEMALE DUMMY) WITH VAFAEI ET AL.

	ROA		ROE		Q		CFO/TA	
Variable	Rep. sample	Vafaei et al. sample	Rep. sample	Vafaei et al. sample	Rep. sample	Vafaei et al. sample	Rep. sample	Vafaei et al. sample
Female dummy	0.001	0.0396***	0.005	0.0798***	0.055	0.119**	0.009	0.0355***
	(0.227)	(4.742)	(0.361)	(5.314)	(1.083)	(1.896)	(1.172)	(3.911)
BHO(1)	-0.005	0.0190	-0.023	0.0227	-0.036	-0.309**	0.003	-0.0064
	(-0.351)	(1.0305)	(-0.788)	(0.684)	(-0.327)	(-2.318)	(0.151)	(-0.3209)
MSO	0.062**	0.0214	0.102**	0.0337	0.209	0.343**	0.031	-0.0217
	(2.312)	(0.931)	(2.300)	(0.815)	(1.220)	(2.073)	(1.125)	(-0.8707)
BoardSize	-0.022	-0.0451***	-0.021	-0.0501**	-0.021	0.0683	-0.013	-0.0079
	(-1.413)	(-3.537)	(-0.783)	(-2.188)	(-0.192)	(0.744)	(-0.764)	(-0.5707)
FirmAge	0.004	-0.002	-0.007	-0.0111	-0.029	-0.0057	0.002	-0.00438
	(0.826)	(-0.408)	(-0.823)	(-1.205)	(-0.744)	(-0.155)	(0.403)	(-0.786)
Leverage	-0.035***	-0.0157**	-0.076***	-0.0207*	-0.281***	-0.44***	-0.036***	-0.0229***
	(-4.266)	(-2.394)	(-4.585)	(-1.755)	(-3.553)	(-9.291)	(-3.865)	(-3.201)
FirmSize	0.010**	0.0179***	0.017**	0.0188***	-0.374***	-0.253***	0.002	0.00324
	(2.203)	(5.347)	(2.200)	(3.117)	(-12.101)	(-10.460)	(0.517)	(0.889)
NED%	0.026	-0.0449	0.038	-0.0930*	-0.002	-0.103	0.024	-0.0623**
	(0.950)	(-1.592)	(0.789)	(-1.835)	(-0.014)	(-0.508)	(0.751)	(-2.033)
AssetsInPlace	0.039**	0.0185	0.066**	0.0106	-0.091	-0.296***	0.073***	0.00518***
	(2.306)	(1.369)	(2.165)	(0.437)	(-0.728)	(-3.090)	(3.708)	(3.519)
Intercept	0.058*	0.0345	0.135**	0.136***	4.406***	3.762***	0.096***	0.104***
	(1.815)	(1.975)	(2.231)	(2.634)	(16.721)	(18.118)	(2.728)	3.329
Observations	2,113	1,101	2,113	1,101	2,113	1,101	2,113	1,101
Groups	595	224	595	224	595	224	595	224

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively (Vafaei et al.'s levels of significance reported as per their paper). Unlike in other tables, Table 14 provides t-statistics in parentheses (not standard errors) to enable comparison with the statistics provided by Vafaei, Ahmed and Mather (2015).

TABLE 15

RESULTS FOR IV WITH FEMALE MEASURED AS A DUMMY, OLS FOR FEMALE MEASURED AS A PERCENTAGE AND IV FOR FEMALE MEASURED AS A PERCENTAGE

	ROA			ROE			Q			CFO/TA		
	IV (dummy)	OLS (%)	IV (%)	IV (dummy)	OLS (%)	IV (%)	IV (dummy)	OLS (%)	IV (%)	IV (dummy)	OLS (%)	IV (%)
Female	0.075	0.003	0.232	0.156	0.042	0.495	0.238	0.391	0.088	0.231*	0.049	0.690*
	(0.101)	(0.030)	(0.292)	(0.185)	(0.063)	(0.546)	(0.713)	(0.279)	(2.197)	(0.135)	(0.036)	(0.365)
BHO(1)	-0.004	-0.005	-0.004	-0.021	-0.022	-0.020	-0.036	-0.034	-0.029	0.009	0.003	0.008
	(0.015)	(0.015)	(0.014)	(0.028)	(0.029)	(0.027)	(0.108)	(0.108)	(0.108)	(0.020)	(0.018)	(0.018)
MSO	0.075***	0.062**	0.069***	0.127***	0.102**	0.114***	0.244	0.210	0.196	0.070**	0.031	0.049*
	(0.027)	(0.027)	(0.021)	(0.049)	(0.044)	(0.038)	(0.186)	(0.171)	(0.156)	(0.036)	(0.028)	(0.026)
BoardSize	-0.033*	-0.021	-0.019	-0.047	-0.020	-0.018	-0.031	-0.010	-0.037	-0.049*	-0.011	-0.006
	(0.020)	(0.015)	(0.013)	(0.039)	(0.027)	(0.023)	(0.151)	(0.110)	(0.095)	(0.028)	(0.017)	(0.016)
FirmAge	0.006	0.004	0.005	-0.004	-0.007	-0.005	-0.027	-0.028	-0.026	0.008	0.002	0.005
	(0.006)	(0.005)	(0.005)	(0.010)	(0.009)	(0.009)	(0.040)	(0.039)	(0.040)	(0.008)	(0.006)	(0.006)
Leverage	-0.032***	-0.035***	-0.033***	-0.070***	-0.076***	-0.073***	-0.284***	-0.280***	-0.272***	-0.028**	-0.036***	-0.031***
	(0.008)	(0.008)	(0.007)	(0.015)	(0.016)	(0.014)	(0.060)	(0.079)	(0.055)	(0.011)	(0.009)	(0.009)
FirmSize	0.006	0.010**	0.008**	0.008	0.017**	0.013*	-0.376***	-0.374***	-0.381***	-0.010	0.003	-0.004
	(0.006)	(0.005)	(0.004)	(0.011)	(0.008)	(0.007)	(0.044)	(0.031)	(0.029)	(0.008)	(0.005)	(0.005)
NED%	0.011	0.026	0.015	0.010	0.037	0.018	-0.058	-0.010	0.023	-0.024	0.023	-0.007
	(0.031)	(0.027)	(0.025)	(0.058)	(0.048)	(0.047)	(0.223)	(0.173)	(0.192)	(0.042)	(0.032)	(0.032)
AssetsInPlace	0.045***	0.039**	0.042***	0.079***	0.066**	0.072***	-0.091	-0.090	-0.074	0.092***	0.072***	0.079***
	(0.017)	(0.017)	(0.014)	(0.031)	(0.031)	(0.026)	(0.117)	(0.125)	(0.105)	(0.022)	(0.020)	(0.017)
Constant	0.072**	0.058*	0.047	0.174***	0.133**	0.119**	4.387***	4.379***	4.461***	0.141***	0.093***	0.070*
	(0.032)	(0.032)	(0.031)	(0.062)	(0.060)	(0.056)	(0.244)	(0.261)	(0.244)	(0.044)	(0.035)	(0.039)
Observations	2,113	2,113	2,113	2,113	2,113	2,113	2,113	2,113	2,113	2,113	2,113	2,113
Groups	595	595	595	595	595	595	595	595	595	595	595	595

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively

7.4 Adams and Ferreira (2009)

Adams and Ferreira (2009) find that 'the average effect of [board] gender diversity on firm performance is negative'. In their conclusions, they state "We find that diversity has a positive impact on performance in firms that otherwise have weak governance, as measured by their abilities to resist takeovers. In firms with strong governance, however, enforcing gender quotas in the boardroom could ultimately decrease shareholder value." They draw on data from the S&P's 1,500 index for the 1996-2003 period. Their unbalanced panel includes 8,253 firm-year data-points for most variables across 1,939 firms. They include a control for 'firm complexity', proxied by the number of business segments in which a firm is active and they use log(sales) as a proxy for firm size. They also employ controls for: board size; the fraction of independent directors and the lagged performance variable. Two measures of performance are used, ROA and ln(Tobin's Q). Female board representation is measured as the proportion of female directors on the board. Year and industry dummies are included in the random effects version of their model. Regressions are performed using OLS, instrumental variables, random effects, fixed effects and the Arellano-Bond one-step estimator. The model they estimate is:

$$(2)\ Performance_{it} = \beta_0 + \beta_1 Female\%_{it} + \beta_2 BoardSize_{it} + \beta_3 Independent\%_{it} + \beta_4 FirmSize_{it} + \beta_5 \#BusinessSegments_{it} + \beta_6 Performance_{i,(t-1)} + \beta_7 IndustryDummy_{it} + \beta_8 YearDummy_{it} + \alpha_i + \varepsilon_{it}$$

Adams and Ferreira (2009) report a positive and statistically significant coefficient for *Female* in their OLS models using both performance measures. But this relationship is not robust to any method of addressing the endogeneity of gender diversity. If anything, the relation appears to be negative once steps are taken to address endogeneity.

Given that we do not use the Adams and Ferreira (2009) data in our replication, and we consider Australian rather than US firms, we do not describe their data. Further detail on their data can be found in Bayly (2020) and Adams and Ferreira (2009). The data we use are described in section 7.2 and the sample of firms that we use is described in the "all firms" columns of Table 11 above.

7.4.1 Do the Adams and Ferreira (2009) results hold in Australia?

Most of the variables used by Adams and Ferreira (2009) can be replicated using Australian data. The majority of the data used is as outlined in Section 7.2.1. However, Datanalysis does not provide the '*BusinessSegments*' data that Adams and Ferreira (2009) draw from Compustat, which is a proxy for firm complexity. Instead, we use Datanalysis's list of 'industry segmentals' (*#IndustrySegments)*. Also, Datanalysis does not contain the sales data that is used to proxy firm size in Adams and Ferreira (2009). Instead, we use total assets from Datanalysis to proxy firm size. In section 7.4.2 we explore the robustness of our results to the choice of firm size proxy. We also use a variable that captures the proportion of board members who are independent directors.

We implement the regressions described by Adams and Ferreira (2009) using our data from 2005-2016. However, the Arellano-Bond model requires some further explanation here. As per Adams and Ferreira (2009), the one step model is augmented by a one period lagged performance variable. The instruments used are two and all further period lagged left hand side variables and one period lags of all right hand side variables except year dummies. Standard errors are corrected for heteroskedasticity. Firm fixed effects are included. We replicate the Adams and Ferreira (2009) regressions on the "all firms" sample described in Table 11, but also estimate regressions with the ASX 200 and ASX500 samples.

Tables 16, 17 and 18 demonstrate that the random effects models all return a positive and statistically significant coefficient for *Female%* in the regressions with ln(Q) as the measure of performance. This is true whether we consider all firms (Table 16), ASX 500 firms (Table 17) or ASX 200 firms (Table 18). But the same models return consistently insignificant results when ROA is the measure of performance. All models that utilize fixed effects or instrumental variables yield insignificant results on both performance measures. All random effects models yield Hausman test p-values of 0.000, suggesting that fixed effects should be preferred, and this preference is supported by theoretical concerns regarding omitted firm characteristics that can be controlled for by the inclusion of fixed effects. For ASX500 firms, the Arellano-Bond model returns a negative and statistically significant coefficient for *Female%* against ROA. However, the Hansen (1982) test for this regression yields a p-value of 0.092, calling into question the validity of the instruments. Overall, the regressions performed provide strong evidence that there is no causal link between female board representation and firm performance.

TABLE 16

REGRESSION RESULTS (ALL FIRMS)

	OLS (RE)		OLS (FE)		IV (FE)		Arellano-Bond	
	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)
Female%	-0.001	0.176***	0.010	0.080	-0.248	-0.516	6.268	-0.156
	(0.026)	(0.067)	(0.039)	(0.079)	(0.243)	(0.450)	(7.375)	(0.271)
BoardSize	-0.016***	0.028***	-0.013***	0.019***	-0.013***	0.020***	0.447	0.128***
	(0.002)	(0.005)	(0.003)	(0.006)	(0.003)	(0.005)	(0.491)	(0.018)
Independent%	-0.019	0.025	-0.014	-0.005	-0.012	0.000	-0.567	0.044
	(0.013)	(0.021)	(0.015)	(0.024)	(0.014)	(0.022)	(0.513)	(0.039)
FirmSize [ln(assets)]	0.052***	-0.069***	0.087***	-0.153***	0.087***	-0.151***	-3.455	-0.329***
	(0.003)	(0.006)	(0.008)	(0.012)	(0.007)	(0.009)	(3.713)	(0.033)
# IndustrySegments	0.010***	0.031***						
	(0.003)	(0.009)						
Lag performance	-0.000	0.304***	0.000	0.264***	0.000	0.264***	-0.005	0.199***
	(0.000)	(0.019)	(0.000)	(0.020)	(0.000)	(0.017)	(0.006)	(0.022)
Constant	-0.102***	0.445***	-0.314***	0.885***				
	(0.016)	(0.042)	(0.033)	(0.054)				
Observations	13,788	13,757	13,788	13,757	13,462	13,429	12,026	11,988
No. groups	2,320	2,315	2,320	2,315	1,994	1,987	2,090	2,082
R2 (overall)	0.144	0.3792	0.097	0.195				
Industry dummies	Yes	Yes	No	No	No	No	No	No
Firm fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively

#IndustrySegments is time-invariant and drops out of the fixed effects models

TABLE 17
REGRESSION RESULTS (ASX500 FIRMS)

	OLS (RE)		OLS (FE)		IV (FE)		Arellano-Bond	
	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)
Female%	0.020	0.159*	0.006	-0.028	-0.199	0.061	-0.412**	0.222
	(0.023)	(0.081)	(0.025)	(0.095)	(0.138)	(0.438)	(0.174)	(0.356)
BoardSize	-0.008***	0.009*	-0.007***	-0.003	-0.007***	-0.003	-0.013	-0.019
	(0.002)	(0.005)	(0.002)	(0.006)	(0.002)	(0.005)	(0.010)	(0.018)
Independent%	-0.001	-0.006	-0.001	-0.024	0.000	-0.025	0.007	-0.057
	(0.016)	(0.023)	(0.017)	(0.024)	(0.015)	(0.023)	(0.026)	(0.046)
FirmSize [ln(assets)]	0.024***	-0.230***	0.024***	-0.297***	0.024***	-0.296***	0.139***	-0.127***
	(0.005)	(0.011)	(0.009)	(0.017)	(0.007)	(0.013)	(0.024)	(0.041)
# IndustrySegments	0.012***	0.049***						
	(0.003)	(0.010)						
Lag performance	0.125***	0.306***	0.050	0.292***	0.048	0.293***	0.000*	0.234***
	(0.033)	(0.020)	(0.033)	(0.021)	(0.032)	(0.018)	(0.000)	(0.033)
Constant	-0.007	1.648***	-0.051	2.317***				
	(0.024)	(0.076)	(0.049)	(0.110)				
Number of observations	5,561	5,553	5,561	5,553	5,325	5,318	4,817	4,807
No. groups	1,110	1,107	1,110	1,107	874	872	983	977
R2 (overall)	0.1552	0.5118	0.0705	0.4177				
Industry dummies	Yes	Yes	No	No	No	No	No	No
Firm fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively
#IndustrySegments is time-invariant and drops out of the fixed effects models

Table 18
Regression results (ASX200 firms)

	OLS (RE)		OLS (FE)		IV (FE)		Arellano-Bond	
	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)
Female%	0.019	0.209**	0.002	-0.017	-0.122	0.751	0.007	0.437
	(0.017)	(0.085)	(0.021)	(0.096)	(0.118)	(0.544)	(0.146)	(0.318)
BoardSize	0.000	0.008	0.001	0.003	0.000	0.005	-0.010	-0.010
	(0.001)	(0.006)	(0.002)	(0.007)	(0.001)	(0.007)	(0.010)	(0.016)
Independent%	-0.002	0.012	-0.004	0.001	-0.003	-0.006	0.047	-0.061
	(0.006)	(0.025)	(0.006)	(0.026)	(0.005)	(0.026)	(0.043)	(0.046)
FirmSize [ln(assets)]	-0.017***	-0.223***	-0.025***	-0.352***	-0.025***	-0.352***	0.198*	-0.251***
	(0.004)	(0.018)	(0.009)	(0.022)	(0.006)	(0.021)	(0.102)	(0.063)
# IndustrySegments	0.008***	0.023**						
	(0.002)	(0.010)						
Lag performance	0.174***	0.402***	0.119***	0.309***	0.118***	0.311***	-0.198	0.136***
	(0.048)	(0.030)	(0.046)	(0.032)	(0.036)	(0.027)	(0.136)	(0.047)
Constant	0.178***	1.843***	0.261***	3.109***				
	(0.026)	(0.133)	(0.067)	(0.175)				
Number of observations	2,267	2,266	2,267	2,266	2,194	2194	1,992	1,992
No. groups	411	410	411	410	338	338	376	376
R2 (overall)	0.3328	0.6576	0.1886	0.4943				
Industry dummies	Yes	Yes	No	No	No	No	No	No
Firm fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively
#IndustrySegments is time-invariant and drops out of the fixed effects models

7.4.2 Robustness to firm proxy size

The regressions estimated in the section above are now re-estimated using ln(revenue) and ln(market capitalization) as proxies for firm size instead of the natural log of firm assets. Each cell reports the coefficient on *Female%* from a separate estimation. ln(revenue) is missing for a substantial number of firms. To investigate whether the reduced sample which results when we use revenue as a proxy for firm size is affecting the results, we also estimate two more sets of regressions using assets and market capitalisation, but dropping firms for which revenue observations were unavailable. These regressions are marked with 'ltd'. These have the same, reduced sample size as the ln(revenue) estimates. The coefficients for *Female%* from each of these regressions are presented in Table 19.

TABLE 19
FEMALE% COEFFICIENTS WITH VARIOUS FIRM SIZE PROXIES

		Random effects		Fixed effects		Instrumental Variables		Arellano-Bond	
	Firm size proxy	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)	ROA	ln(Q)
ASX200	ln(assets)	0.019	0.209**	0.002	-0.017	-0.122	0.751	0.007	0.437
	ln(mktcap)	-0.004	-0.076	0.002	-0.018	-0.189	-0.146	-0.126	-0.067
	ln(revenue)	-0.001	0.068	0.005	-0.022	-0.208	0.248	0.071	0.437
	ln(assets) – ltd	0.021	0.159	0.001	-0.081	-0.132	0.816	-0.013	0.575
	ln (mktcap) – ltd	0.003	-0.082	0.009	0.029	-0.216	-0.227	-0.143	0.041
ASX500	ln(assets)	0.020	0.159*	0.006	-0.028	-0.199	0.061	-0.412**	0.222
	ln(mktcap)	0.016	-0.118	0.009	0.087	-0.159	-0.639	-0.498**	-0.330
	ln(revenue)	0.046*	0.006	0.033	0.078	-0.245	-0.819	-0.294*	-0.042
	ln(assets) – ltd	0.050*	0.144	0.027	-0.004	-0.233	-0.207	-0.401*	0.008
	ln (mktcap) – ltd	0.044*	-0.127	0.031	0.159	-0.229	-1.154*	-0.587**	-0.553
All firms	ln(assets)	-0.001	0.176***	0.010	0.080	-0.248	-0.516	6.268	-0.156
	ln(mktcap)	0.010	-0.213***	0.007	-0.083	-0.099	-1.786***	1.384	-1.381***
	ln(revenue)	0.036	0.085	0.025	0.037	-0.029	-1.230*	-0.092	-0.223
	ln(assets) – ltd	-0.008	0.143**	0.001	0.070	-0.262	-0.707	-1.116	-0.011
	ln (mktcap) – ltd	0.003	-0.209***	-0.008	-0.060	-0.006	-2.189***	-1.089	-1.426***

Note: *, ** and *** indicate significance at 10, 5 and 1 per cent respectively

It is clear from these results that the choice of firm size proxy can have a substantial effect. Although some coefficients are relatively insensitive to the firm size proxy chosen, many are not. For example, the *Female%* coefficient in regressions where ln(Q) is the measure of performance is generally much lower when the proxy for firm size is market capitalisation rather than assets. In one instance – the random effects regression for all firms using ln(Q) – *Female%* is found to be significant and negative when using market capitalisation and significant and positive when using assets.

The only consistently significant *Female%* coefficients of the same sign are those from the ASX500 Arellano-Bond model for ROA. These coefficients have a negative sign. However, the Hansen test results for all of these regressions - from p=0.002 to p=0.091 - suggest that the instruments are invalid. All of the random effects models return Hausman p-values of 0.000, suggesting fixed effects is necessary to control for unobserved firm heterogeneity. All of the fixed effects models produce statistically insignificant results. This indicates strongly that the significant random effects results are unreliable.

In contrast to the findings of Adams and Ferreira (2009), there is no strong suggestion from the results presented in Table 19 that fixed effects yields statistically negative results. There is also no evidence of the significant, negative effects from instrumental variable regressions found in Adams and Ferreira (2009). Broadly speaking, their insignificant results for the Arellano-Bond model are mostly replicated here, however, for the handful of models generating significant results, these appear far less convincing after further testing.

All of these models are expanded upon significantly in subsequent sections. The key finding at this stage is that applying the methods of Adams and Ferreira (2009) to Australian data does not yield similar results. Also notable is that the Australian results differ substantially depending on the firm size proxy employed, suggesting that the small number of significant results originally found using ln(assets) to proxy firm size are not robust. Importantly, many of the small number of significant models do not withstand basic post-estimation specification testing.

7.5 Exploring the regression approach

In this section, we estimate a wide variety of regressions using different measures of firm performance, different samples and different estimation techniques. Our purpose in doing this is to shed light on three questions: How frequently do we find a statistically significant effect of a female presence on boards on firm performance? When a statistically significant effect is found, is it more likely to be negative or positive? Are there any patterns in which measures of firm performance, which samples or which econometric techniques are most likely to produce statistically significant results?

7.5.1 Regression data

We begin by undertaking basic regressions which utilize five different measures of firm performance for firms in the ASX200, ASX500 and all firms listed on the ASX, respectively. The regressions presented in Table 20 use five different performance measures: return on assets (ROA); return on equity (ROE); the ratio of cash flow from operations to total assets (CFO/TA or CFOTA); Tobin's Q (Q) and its natural logarithm (ln(Q)). The construction of the data are as described in section 7.2 above and descriptive statistics are presented in Table 11.

7.5.2 Results of basic regressions

For each performance measure, the first column of Table 20 shows the results with no controls, the second column shows the results with a proxy for firm size (total assets of the firm in log form) and the third column shows the results with the proxy for firm size but also including firm fixed effects.

Table 20 shows that that when no controls are used there is a positive and statistically significant relationship between the percentage of female board members and four of the five measures of financial performance in the 'all firms' group. But, once controls for firm size and firm fixed effects are added, the relationship between the percentage of female board members and financial performance becomes negative and statistically significant. This is consistent with Adams (2016) who demonstrates that there is a positive correlation between ROE and the fraction of women on a firm's board in her S&P1500 sample from the U.S., but that its statistical significance disappears once firm size is controlled for using log sales. The magnitude of the relationship between ROE and the fraction of women on a firm's board drops further once industry dummy variables are added, and once firm fixed effects are included, this relationship becomes negative and significant at the 1 per cent level. This relationship also holds for ROA and ROE measures of performance in our sample of ASX500 firms but not for other measures of performance.

The results are more mixed when only ASX200 firms are considered. Once firm size and firm fixed effects are included, there is a negative relationship between the percentage of female board members and ROA, but a positive relationship between the percentage of female board members and Q. There is no relationship between the percentage of female board members and ROE, ln(Q) or CFOTA in the ASX200 sample once firm size and firm fixed effects are controlled for.

These basic regressions indicate that: firstly, the manner in which performance is measured alters the apparent relationship between performance and female representation. Secondly, for the 'all firms' group, the positive correlation between female representation and performance is both strong and consistent. Thirdly, and again at least for the 'all firms' group, this relationship consistently weakens and in fact becomes significantly negative once a firm size control and fixed effects are employed.

Table 20
Regression results for Female% (no controls, firm size only, and firm fixed effects)

	ROA			ROE			ln(Q)			Q			CFOTA		
	No controls	Firm size only	Firm fixed effects	No controls	Firm size only	Firm fixed effects	No controls	Firm size only	Firm fixed effects	No controls	Firm size only	Firm fixed effects	No controls	Firm size only	Firm fixed effects
ASX200															
Obs	2,294	2,294	2,294	2,292	2,292	2,292	2,293	2,293	2,293	2,293	2,293	2,293	2,293	2,293	2,293
Female% coefficient	-0.022	0.038	-0.035	0.093	0.151	0.020	-0.572	0.385	0.060	-1.568	1.390	0.678	-0.014	0.058	0.033
p-value	0.146	0.048	0.069	0.004	0.000	0.763	0.000	0.000	0.642	0.000	0.000	0.081	0.442	0.007	0.344
ASX500															
Obs	5,668	5,668	5,668	5,633	5,663	5,633	5,664	5,664	5,664	5,664	5,664	5,664	5,666	5,666	5,666
Female% coefficient	0.110	0.033	-0.081	0.409	0.288	-0.101	-0.422	0.546	-0.144	-1.524	1.456	0.073	0.071	0.034	-0.022
p-value	0.000	0.094	0.000	0.000	0.000	0.043	0.000	0.000	0.173	0.000	0.000	0.827	0.000	0.072	0.323
All firms															
Obs	14,092	14,092	14,092	13,980	13,980	13,980	14,087	14,087	14,087	14,087	14,087	14,087	14,086	14,086	14,086
Female% coefficient	0.206	-0.013	-0.111	0.411	0.068	-0.182	0.211	0.251	-0.337	0.093	0.628	-0.469	0.144	-0.009	-0.065
p-value	0.000	0.467	0.001	0.000	0.443	0.042	0.000	0.000	0.000	0.466	0.000	0.067	0.000	0.542	0.001

Note: results that are not statistically significant at the 10 per cent are shaded. Negative coefficients are shown in red.
Columns which say "Firm fixed effects" also include a time-varying control for firm size.

7.6 Regression analysis: robustness to performance measures and econometric techniques

We could not replicate the positive effect of female board representation on Australian company performance that was found by Vafaei, Ahmed and Mather (2015). Further, the negative effect of female board representation on US firms found by Adams and Ferreira (2009) does not appear to apply to Australian firms when we use their rich set of control variables. It is nonetheless worthwhile expanding and amending the models proposed by Vafaei, Ahmed and Mather (2015) and Adams and Ferreira, 2009. The aim of our next exercise is to cast a very wide net, testing a variety of models and statistical techniques to examine both the effect of female board representation on performance and the specification issues that appear to affect results that are found in different papers.

We expand upon the basic regressions of Table 20 by considering a range of additional control variables that have been suggested by the literature. We include the following additional variables in our analysis: membership of either the ASX200 index or the ASX500 index; the age of directors; the tenue of directors; CEO characteristics (gender, tenure and dummy variable for whether the CEO has been in place for more than 12 months) and lagged versions of the *Female%* variable. The augmented model is shown below.

$$(3)\ Performance_{it} = \gamma_0 + \gamma_1 Female_{it \text{ or } i(t-1) \text{ or } i(t-2)} + \gamma_2 BoardSize_{it} + \gamma_4 Independent\%_{it} + \gamma_5 BoardTenure_{it} + \gamma_6 CEOGender_{it} + \gamma_7 CEOTenure_{it} + \gamma_8 CEOTurnover_{it} + \gamma_9 \#BusinessSegments_{it} + \gamma_{10} FirmAge_{it} + \gamma_{11} FirmSize_{it} + \gamma_{12} Leverage_{it} + \gamma_{13} Performance_{i(t-1)} + \gamma_{14} YearDummy_{it} + \gamma_{15} IndustryDummy_{it} + \varepsilon_{it}$$

A wide range of the variations can be implemented with the Australian data available. *Performance* can be measured alternatively as: ROA; ROE; Q; ln(Q) and CFOTA. *Female* can be measured alternatively as: proportion of directors who are female (*Female%*); dummy for whether any of the directors are female and the Blau Index. *Female* can also be measured at different points in time: the present period; one-year lag and two-year lag. *BoardSize* can be measured alternatively as a raw number and in log form. *FirmSize* can be measured as ln(assets) and ln(market capitalisation). *Leverage* can be measured as 'financial leverage' as in Vafaei, Ahmed and Mather (2015), or using Datanalysis fields (dividing long term debt by market capitalization). *IndustrySegmentals* is used as a proxy for firm complexity. *Independent%* measures the proportion of board members who are independent directors.

All regressions are performed using the following samples: ASX200 firms only; ASX500 firms only; and all available firms. Regressions are performed using the following specifications: random effects; firm fixed effects; instrumental variables, with the instruments being alternatively IV_{male} (the fraction male directors on the board who sit on other boards where there are other female directors) and IV_{all} (the fraction of all directors who sit on other boards on which there are other female directors); and Arellano-Bond (AB) as specified by Adams and Ferreira (2009). This range of variables and regression specifications yields 5,400 regressions. This represents five different outcomes, three different samples, five different econometric specifications, three different ways of measuring female presence on boards, two different ways of measuring firm size, two different ways of measuring leverage, two different ways of measuring board size and three different time points at which *female* is measured.

Data construction and summary statistics for the three groups of firms are provided above in section 7.2 and Table 11. Firms in the ASX200 group are generally 'better' as defined by the performance measures, and their performance exhibits a much lower variance than the `all firms' group. ASX200 firms also tend to have larger boards, a greater proportion of both female and independent directors, and longer tenure for both directors and CEOs.

7.6.1 Overall regression results

Given the large number of regressions that we estimate, the first phase of analysis is primarily concerned with identifying models that yield statistically significant results for the coefficient on the *female* measure (i.e. percentage of female directors, the female dummy or Blau index). Across all 5,400 models, 982 (18 per cent) *female* coefficients were significant at 10 per cent, 628 (12 per cent) at 5 per cent and 301 (6 per cent) at 1 per cent. For ASX200 firms (11 per cent, 5 per cent and 1 per cent of models significant at the 10, 5 and 1 per cent levels respectively), the proportion of significant models for the ASX 200 is roughly what one would expect from random variation.

Insignificant results are most common, followed by statistically significant negative results. At the 10 per cent, 5 per cent and 1 per cent significance levels, negative results made up 67, 70 and 77 per cent of significant results respectively.[12]

Table 21 summarises the proportion of models yielding significant results at different levels by sample, lag and significance. Each bolded number represents the proportion of models for each performance measure that are statically significant at each level of statistical significance. These bolded numbers represent the average of the unbolded numbers in the three squares immediately below each bolded number. As each of the unbolded numbers represents 120 models, the bolded numbers represent 360 models.

This table illustrates that some previously hypothesised relationships are not supported by the regressions performed. Most clearly, for ASX200 firms, there is very rarely a statistically significant relationship between female board representation and ROA or ROE. Table 21 also shows that the proportion of significant models is highest for the Q-based measures. Thirdly, the findings are quite inconsistent, and relationships that may appear to be supported for one group of firms or for a particular performance measure may not be supported for others.

[12] For space reasons, details of all 5,400 regressions are not presented. Bayly (2020) provides preliminary estimates of these regressions. The percentages in this paragraph are based upon the final results presented in this paper.

TABLE 21
PROPORTION OF MODELS THAT ARE STATISTICALLY SIGNIFICANT BY LEVEL OF SIGNIFICANCE

	Level of significance								
	10%			5%			1%		
	ASX200	**ASX500**	**All firms**	**ASX200**	**ASX500**	**All firms**	**ASX200**	**ASX500**	**All firms**
ROA	**6.4%**	**19.7%**	**8.6%**	**0.8%**	**11.7%**	**5.3%**	**0.0%**	**6.9%**	**2.8%**
no lag	9.2%	21.7%	10.0%	0.0%	20.0%	5.8%	0.0%	10.8%	3.3%
lag 1	6.7%	23.3%	8.3%	2.5%	15.0%	6.7%	0.0%	10.0%	5.0%
lag 2	3.3%	14.2%	7.5%	0.0%	0.0%	3.3%	0.0%	0.0%	0.0%
ROE	**7.5%**	**32.2%**	**4.4%**	**2.2%**	**18.9%**	**1.9%**	**0.0%**	**7.2%**	**0.6%**
no lag	19.2%	53.3%	0.0%	6.7%	35.8%	0.0%	0.0%	18.3%	0.0%
lag 1	3.3%	36.7%	10.0%	0.0%	19.2%	4.2%	0.0%	3.3%	0.8%
lag 2	0.0%	6.7%	3.3%	0.0%	1.7%	1.7%	0.0%	0.0%	0.8%
ln(Q)	**20.3%**	**27.5%**	**51.9%**	**10.0%**	**20.3%**	**45.3%**	**1.1%**	**10.0%**	**36.1%**
no lag	20.0%	30.0%	59.2%	2.5%	16.7%	53.3%	0.0%	5.0%	40.0%
lag 1	20.8%	26.7%	52.5%	14.2%	26.7%	50.0%	3.3%	16.7%	41.7%
lag 2	20.0%	25.8%	44.2%	13.3%	17.5%	32.5%	0.0%	8.3%	26.7%
Q	**19.7%**	**14.7%**	**34.4%**	**10.0%**	**7.8%**	**28.9%**	**3.9%**	**0.3%**	**13.6%**
no lag	34.2%	8.3%	39.2%	20.0%	4.2%	33.3%	3.3%	0.0%	16.7%
lag 1	8.3%	19.2%	40.8%	5.0%	9.2%	35.0%	5.0%	0.0%	11.7%
lag 2	16.7%	16.7%	23.3%	5.0%	10.0%	18.3%	3.3%	0.8%	12.5%
CFOTA	**2.8%**	**10.3%**	**12.2%**	**0.6%**	**4.7%**	**6.1%**	**0.0%**	**0.0%**	**1.1%**
no lag	6.7%	15.8%	20.0%	1.7%	10.0%	15.0%	0.0%	0.0%	1.7%
lag 1	1.7%	15.0%	13.3%	0.0%	4.2%	3.3%	0.0%	0.0%	1.7%
lag 2	0.0%	0.0%	3.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 21 groups negative and positive results together.

Models were more likely overall to be significant for all firms (22.3 per cent were significant at 10 per cent) than for ASX200 firms (11.3 per cent). However, this may not be surprising given the large difference in sample size. As shown in Table 22, results differ considerably depending on the statistical technique employed. Random effects and Arellano-Bond models are generally much more likely to yield a significant result than fixed effects or instrumental variable models.

Table 22
Proportion of models that are statistically significant (at the 10% level) by model type

		Performance measure					
		ROA	**ROE**	**ln(Q)**	**Q**	**CFOTA**	**All measures**
All samples	Random effects	10.2%	30.1%	69.9%	49.5%	10.2%	**34.0%**
	Fixed effects	6.5%	14.8%	22.7%	12.0%	11.1%	**13.4%**
	IV_{all}	0.9%	8.3%	16.7%	13.0%	0.0%	**7.8%**
	IV_{male}	1.9%	5.6%	17.6%	22.2%	0.0%	**9.4%**
	Arellano-Bond	38.4%	14.8%	39.4%	18.1%	20.8%	**26.3%**
	All models	**11.6%**	**14.7%**	**33.2%**	**23.0%**	**8.4%**	**18.2%**
ASX200	Random effects	1.4%	12.5%	52.8%	40.3%	8.3%	**23.1%**
	Fixed effects	2.8%	0.0%	19.4%	11.1%	0.0%	**6.7%**
	IV_{all}	0.0%	0.0%	0.0%	5.6%	0.0%	**1.1%**
	IV_{male}	5.6%	0.0%	11.1%	33.3%	0.0%	**10.0%**
	Arellano-Bond	22.2%	25.0%	18.1%	8.3%	5.6%	**15.8%**
	All models	**6.4%**	**7.5%**	**20.3%**	**19.7%**	**2.8%**	**11.3%**
ASX500	Random effects	26.4%	77.8%	90.3%	54.2%	13.9%	**52.5%**
	Fixed effects	8.3%	44.4%	0.0%	0.0%	0.0%	**10.6%**
	IV_{all}	0.0%	16.7%	16.7%	0.0%	0.0%	**6.7%**
	IV_{male}	0.0%	16.7%	0.0%	0.0%	0.0%	**3.3%**
	Arellano-Bond	63.9%	5.6%	30.6%	19.4%	37.5%	**31.4%**
	All models	**19.7%**	**32.2%**	**27.5%**	**14.7%**	**10.3%**	**20.9%**
All firms	Random effects	2.8%	0.0%	66.7%	54.2%	8.3%	**26.4%**
	Fixed effects	8.3%	0.0%	48.6%	25.0%	33.3%	**23.1%**
	IV_{all}	2.8%	8.3%	33.3%	33.3%	0.0%	**15.6%**
	IV_{male}	0.0%	0.0%	41.7%	33.3%	0.0%	**15.0%**
	Arellano-Bond	29.2%	13.9%	69.4%	26.4%	19.4%	**31.7%**
	All models	**8.6%**	**4.4%**	**51.9%**	**34.4%**	**12.2%**	**22.3%**

Table 22 groups negative and positive results together.

7.6.2 Discussion of control variables and additional robustness checks

The most notable impact of controls comes from the choice of proxy for firm size (ln(assets) or ln(market capitalization)). 25 per cent of models give a different coefficient sign for *Female* depending on which proxy is used, and a further 10 per cent share the same sign but change from being statistically significant to insignificant (at 10 per cent significance). The effect of firm size proxy is even starker when Q-based models are used. For these models, 48 per cent of the models switch sign depending on the proxy and a further 14 per cent of those with the same sign lose or gain their significance (at 10 per cent). For Q-based measures, the coefficient is generally (69 per cent of the time) more positive when ln(assets) is used than it is when ln(market capitalisation) is used. But the effect is much more mixed for the other performance measures. These results are concerning given recent findings which indicate that empirical corporate finance findings may be dubious if they are not consistent across firm size measures (Dang, Li and Yang 2018).

Indeed, if we define a 'consistently significant result' as one in which a particular regression technique (such as fixed effects) returns a statistically significant coefficient for a particular *female* measure across all eight combinations of control variables, there are very few such results – just 24 (3.6 per cent) at 10 per cent and 13 (1.9 per cent) at both 5 and 1 per cent.

We conduct further analysis of these consistently significant results that hold across permutations of control variables and the results cast doubt on many of even this small number. Hausman Test results indicate that the random effects models that have statistically significant coefficients are biased. Many of the Arellano-Bond models that produce statistically significant results do not pass simple specification tests such as the Hansen (1982) test. Again, the strong conclusion that emerges is that there is no compelling evidence that female presence on boards affects firm performance in one direction or the other. The details of these robustness checks can be found at Appendix A.

8 Discussion and Conclusion

In the first part of the paper, we reviewed the literature which has tried to estimate the impact of female board membership on firm performance. Much of the literature is plagued by endogeneity problems and an inability to convincingly identify a causal effect of the gender composition of boards on firm performance. Looking at the small number of convincing studies, we found that three of them found no result, three found a negative impact of women on boards and three found a positive impact, although two of these presented a mix of positive and zero results. It seems clear that there is no strong consensus one way or the other.

We assembled a large dataset for Australia and using event study methodologies and regression approaches, attempted to answer the question for Australia and to analyse how the results vary by performance measure, regression technique, and model specification. Drawing on the previous literature, and after undertaking new event studies and regression analyses, we demonstrate that there is no robust relationship between female representation on corporate boards and firm performance. Simply put, we do not find support for the 'economic argument'.

Our event study showed no statistically significant effects of female board appointments on a firm's share price. With a matched male sample, we show that the appointment of a new male board member also has no statistically significant effect on a firm's share price. We do find that there is a large diversity of test outcomes across different sub-samples. Test results could be highly significant against certain tests (particularly the *t*-test) but insignificant against others (particularly the non-parametric tests), and results were highly sensitive to outliers. Overall, there is no clear evidence of a gender-differentiated impact of a new director appointment on a firm's share price given the statistically insignificant differences in abnormal returns between the matched male and female samples.

In our regression analysis, we are unable to replicate the results of Vafaei, Ahmed and Mather (2015), who claim a positive effect of female board appointments on firm performance in Australia. For Australian data, we fail to re-produce the results of Adams and Ferreira (2009) who showed a negative effect of female board appointments on firm performance in the U.S.

To explore how results might vary across different outcome measures and specifications, we estimated a wide range of regression models using the data that we assembled. We started with fifteen simple regressions (five performance measures and three different samples of firms) that used no control variables and then added firm size and fixed effects. Across the fifteen sets of estimates, when we use no controls, we find 8 cases where women have a positive and statistically significant effect on firm performance, 4 cases where women have a negative and statistically significant effect on firm performance and 3 cases where it is statistically insignificant. One we use fixed effects, the results change dramatically: we find only one case where women have a positive and statistically significant effect on firm performance, 8 cases where women have a negative and statistically significant effect on firm performance and 6 cases where the relationship is statistically insignificant. In many ways, the popular debate can be understood from these simple results. Media commentary, such as McCubbing (2023), take this positive correlation and then attach a causal interpretation which is not supported by further analysis.

We expand this simple analysis to consider a range of control variables, different measures of the presence of females on boards, different measures of firm size and a range of econometric techniques. Across all possible permutations, we estimate 5,400 regressions. The vast majority of results are statistically insignificant. When we do find statistically significant results, negative results dominate positive ones by a ratio of three to one. Overall, one can simply say that there is no strong evidence that female board appointments affect a firm's share price or a firm's financial performance in Australia.

What to make of these results?

One possibility is that boards have no effect on firm performance. In this case, we would expect to see zero effect of females on boards in most empirical work and a few cases where the effect is negative and a few cases where the effect is positive. This is fairly consistent with the evidence we present for Australia and a plausible interpretation of the evidence from the wider literature. As there is likely to be a bias against publishing zero results, the three out of nine studies which plausibly address endogeneity concerns which we present probably provide an underestimate of the true number of zero results that have been found by researchers.

Another possibility is that boards do have an effect on firm performance but there is no clear impact of female board members versus males because the females who have been appointed to boards are of a similar quality to male appointees. This seems likely if companies are exercising due diligence in board appointments. We spoke with CEOs of two large superannuation funds in Australia who felt that high-quality boards were key to firm performance but that there was no practical difference between having high-quality male board members and high-quality female board members. They were not surprised by our results and one pointed out to us that the fund would be pushing very hard to replace male board members with equally qualified female board members if this was such as easy way to increase firm performance and market value.

This does not mean that low representation of women on boards is of no concern. First of all, we only look at a small subset of possible firm performance measures. It may be that more gender diverse boards bring other benefits. Possibilities include: ability to withstand negative shocks or improved staff happiness and well-being. These may not be captured by the performance measures we have examined.

We are also not evaluating other social goals such as 'justice'. If board positions are simply paid sinecures, with no impact on firm performance, and these are controlled by old boy networks, then breaking them up with policies such as targets and quotas could be welfare enhancing.

The cautionary note that emerges from our study is that one should be wary of strong claims based on one study. One should also be cautious about interpreting correlations in data as if they are causation. Our event study and our extensive exploration of regression analyses show that one could cherry pick a strong negative result or a strong positive result quite easily. Such results, however, are uncommon. Such results are also not robust to minor specification changes or to post-estimation specification testing. They are not consistent across sub-samples, across different estimation techniques or across different outcome measures.

As a discipline, the best evidence that economics produces comes from strong theory backed by rigorous empirical work which documents the accuracy of the theoretical predictions in a wide variety of time periods, different circumstances, and different places. Consistent evidence of this type has yet to be generated for the proposition that increasing the number of women on boards improves firm performance.

References:

Adams, S. (2015). Women on boards: slow progress and marginalization, study shows, *Forbes* (online), available at https://www.forbes.com/sites/susanadams/2015/06/11/women-on-boards-slow-progress-and-marginalization-study-shows/?sh=24fe21053745. Viewed 2 June 2023.

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

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

Adams, R. and P. Funk (2012). Beyond the glass ceiling: does gender matter? *Management Science* **58**(2): 219-235.

Adams, R., S. Gray and J. Nowland (2011). Does gender matter in the boardroom? Evidence from the market reaction to mandatory new director announcements (November 2, 2011). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1953152.

Ahern, K. and A. Dittmar (2011). The changing of the boards: the impact on firm valuation of mandated female board representation, *Quarterly Journal of Economics* **127**(1): 137-197.

Ali, M., Y. Ng and C. Kulik (2014) Board age and gender diversity: a test of competing linear and curvilinear predictions, *Journal of Business Ethics* **125**(3): 497-512.

Anderson, R, Reeb, D, Upadhyay, A and Zhao, W 2011 'The economics of director heterogeneity', *Financial Management* **40**(1): 5-38.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**(2): 277–297.

Australian Institute of Company Directors (2022). Board diversity statistics, Available at https://aicd.companydirectors.com.au/advocacy/board-diversity/statistics. Retrieved 21 September 2022.

Bayly, N.L. (2020). Female Board Representation and Corporate Performance: an Australian Perspective, PhD thesis, Australian National University, Canberra.

Bear, S., N. Rahman, and C. Post (2010). The impact of board diversity and gender composition on corporate social responsibility and firm reputation, *Journal of Business Ethics* **97**(2): 207-221.

Blau, P.M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York, NY: The Free Press.

Boehmer, E., J. Musumeci and A. B. Poulsen (1991). Event Study Methodology Under Conditions of Event Induced Variance, *Journal of Financial Economics* **30**:253-72

Bøhren, Ø. and R.Ø. Strøm (2010). Governance and politics: Regulating independence and diversity in the board room, *Journal of Business Finance and Accounting* **37**(9-10): 1281-1308.

Brahma, S., C. Nwafor and A. Boateng (2021). Board gender diversity and firm performance: The UK evidence, *International Journal of Finance and Economics* **26**(4): 5704-5719.

Campbell, C. and C. Wasley (1993). Measuring security performance using daily NASDAQ returns, *Journal of Financial Economics* **33**(1): 73–92.

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

Cassells, R. and A. Duncan (2020). *Gender Equity Insights 2020: Delivering the Business Outcomes*, BCEC|WGEA Gender Equity Series, Issue #5, March 2020.

Catalyst (2004). *The Bottom Line: Connecting Corporate Performance and Gender Diversity*, Available at www.catalyst.org/knowledge/bottom-line-connecting-corporate-performance-and-gender-diversity.

Chen, G., C. Crossland and S. Huang (2016). Female board representation and corporate acquisition intensity, *Strategic Management Journal* **37**(2): 303-313.

Conyon, M.J. and L. He (2017). Firm performance and boardroom gender diversity: A quantile regression approach, *Journal of Business Research* **79**: 198-211.

Corrado, C. (1989). A nonparametric test for abnormal security-price performance in event studies, *Journal of Financial Economics* **23**(2): 385-395.

Corrado, C. (2011). Event studies: a methodology review, *Accounting and Finance* **51**(1): 207-234.

Corrado, C. and C. Truong (2008). Conducting event studies with Asia-Pacific security market data, *Pacific-Basin Finance Journal* **16**(5): 493-521.

Corrado, C. and T. Zivney (1992). The specification and power of the sign test in event study hypothesis tests using daily stock returns, *The Journal of Financial and Quantitative Analysis* **27**(3): 465-478.

Cowan, A. (1993). Test for cumulative abnormal returns over long periods: simulation evidence, *International Review of Financial Analysis* **2**(1): 51-68.

Dale-Olsen, H., P. Schøne and M. Verner (2013). Diversity among Norwegian boards of directors: does a quota for women improve firm performance? *Feminist Economics* **19**(4): 110-135.

Dichev, I., J. Graham, J, C. Harvey and S. Rajgopal (2016). The misrepresentation of earnings, *Financial Analyst's Journal* **72**(1): 22-35.

Dimovski, W., L. Lombardi and B. Cooper (2013). Women directors on boards of Australian real estate investment trusts, *Journal of Property and Investment Finance* **31**(2): 196-207.

Dimson, E. and P. Marsh (1986). Event study methodologies and the size effect: The case of UK press recommendations, *Journal of Financial Economics* **17**(1): 113-142.

Ding, D. and C. Charoenwong (2013). Stock market reaction when listed companies in Singapore appoint female directors, *International Journal of Management* **30**(1): 285-300.

Dobbin, F. and J. Jung (2011). Corporate board gender diversity and stock performance: the competence gap or institutional investor bias? *North Carolina Law Review* **89**(3): 809-838.

Dowling, M. and Z. Aribi (2013). Female directors and UK company acquisitiveness, *International Review of Financial Analysis* **29**(C): 79-86.

Eagly, A. and B. Johnson (1990). Gender and leadership style: a meta-analysis, *Psychological Bulletin* **108**(2): 233-256.

Eckbo, B.E., K. Nygaard and K.S. Thorburn (2021). Valuation effects of Norway's board gender-quota law revisited, *Management Science* **68**(1). Available at https://doi.org/10.1287/mnsc.2021.4031.

Erkut, S., V.W. Kramer and A.M. Konrad (2008). Critical mass: Does the number of women on a corporate board make a difference? *Women on corporate boards of directors: International research and practice. 350-366*.

Fama, E., L. Fisher, M. Jensen and R. Roll (1969). The adjustment of stock prices to new information, *International Economic Review* **10**(1): 1-21.

Farrell, K and P. Hersch (2005). Additions to corporate boards: the effect of gender, *Journal of Corporate Finance* **11**(1): 85-106.

Galbreath, J. (2011). Are there gender-related influences on corporate sustainability? A study of women on boards of directors, *Journal of Management and Organization* **17**(1); 7-38.

Gregory-Smith, I., B. Main, B and C. O'Reilly III (2014). Appointments, pay and performance in UK boardrooms by gender, *The Economic Journal* **124**(574): 109-128.

Greene, D., V.J. Intintoli and K.M. Kahle (2020). Do board gender quotas affect firm value? Evidence from California Senate Bill No. 826, *Journal of Corporate Finance* **60**: 101526.

Gul, F., B, Srinidhi and A. Ng (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics* **51(**3): 314-338.

Hafsi, T. and G. Turgut (2013). Boardroom diversity and its effect on social performance: conceptualization and empirical evidence, *Journal of Business Ethics* **112**(2): 463-479.

Hagendorff, J. and K. Keasey (2012). The value of board diversity in banking: evidence from the market for corporate control, *The European Journal of Finance* **18(**1): 41-58.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**(4): 1029–1055.

Harjoto, M., I. Laksmana and R. Lee (2015). Board diversity and corporate social responsibility, *Journal of Business Ethics* **132(**4): 641-660.

Hillman, A., C. Shropshire and A. Cannella Jr. (2007). Organizational predictors of women on corporate boards, *The Academy of Management Journal* **50(**4): 941-952.

Horváth, R. and P. Spirollari (2012). Do the board of directors' characteristics influence firm's performance? The US evidence, *Prague Economic Papers* **4**: 470-486.

Huang, J. and D. Kisgen (2013). Gender and corporate finance: are male executives overconfident relative to female executives? *Journal of Financial Economics* **108**(4): 822-839.

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

Kang, E., D. Ding and C. Charoenwong (2010). Investor reaction to women directors, *Journal of Business Research* **63**(8): 888-894.

Kathyayini, K., C. Tilt and L. Lester (2012). Corporate governance and environmental reporting: an Australian study, *Corporate Governance* **12**(2): 143-163.

Kim, J., D. Roden and S. Cox (2013). The composition and compensation of the board of directors as predictors of corporate fraud, *Accounting and Finance Research* **2(**3): 142-154.

Larkin, M., R. Bernardi and S. Bosco (2012). Board gender diversity, corporate reputation and market performance, *International Journal of Banking and Finance* **9**(1): 1-26.

Lee, P. and E. James (2007). She'-e-os: gender effects and investor reactions to the announcements of top executive appointments, *Strategic Management Journal* **28(**3): 227-241.

Maghin, H. (2022). Cracks in the Boards: The Opportunity Cost of Governance Homogeneity. CESifo Working Paper No. 9816. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4151172 retrieved 26 August 2022.

McCubbing, G. (2023). Appointing scientists to boards can be shortcut to productivity gains. *Australian Financial Review*, June 10 2023. Available at: https://www.afr.com/policy/economy/appointing-scientists-to-boards-can-be-shortcut-to-productivity-gains-20230607-p5deni

Milhomen, C. (2020). MSCI. Women on Boards 2020 Progress Report. https://www.msci.com/documents/10199/9ab8ea98-25fd-e843-c9e9-08f0d179bb85 retrieved 21 September 2021.

Miller, T. and M. Triana (2009). Demographic diversity in the boardroom: mediators of the board diversity-firm performance relationship, *Journal of Management Studies* **45(**5): 755-786.

NBC (2018). https://www.nbcnews.com/think/opinion/california-law-would-create-corporate-gender-quotas-textbook-case-government-ncna906601. Accessed 12 October 2021.

Nguyen, D., J. Hagendorff and A. Eshraghi (2015). Which executive characteristics create value in banking? Evidence from appointment announcements, *Corporate Governance: An International Review* **23**(2): 112-128.

Nguyen, H. and R. Faff (2006). Impact of board size and board diversity on firm value: Australian evidence, *Corporate Ownership and Control* **4**(2): 24-32.

Pathan, S. and R. Faff (2013). Does board structure in banks really affect their performance? *Journal of Banking and Finance* **37**(5): 1573-1589.

Roodman, D. (2008). A note on the theme of too many instruments, Centre for Global Development, Working Paper No. 125, August 2007, revised May 2008, viewed 18 April 2019,

Roodman, D. (2009) How to do xtabond2: an introduction to difference and system GMM in Stata, *The Stata Journal* **9**(1): 86-136.

Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables, *Econometrica* **26:** 393–415.

Schnake, M., R. Williams and W. Fredenberger (2006). Women on boards of directors: effects on firm social performance in the basic materials and financial services sectors, *Journal of Applied Business Research* **22**(1): 31-40.

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

Terjesen, S., E.B. Couto and P.M. Francisco (2016). Does the presence of independent and female directors impact firm performance? A multi-country study of board diversity, *Journal of Management and Governance* **20**(3): 447-483.

Triana, M., T. Miller and T. Trzebiatowski (2013). The double-edged nature of board gender diversity, *Organization Science* **25**(2): 609-632.

Vafaei, A., K. Ahmed and P. Mather (2015). Board diversity and financial performance in the top 500 Australian firms, *Australian Accounting Review* **25**(4): 413-427.

Wang, Y. and B. Clift (2009). Is there a "business case" for board diversity? *Pacific Accounting Review* **21**(2): 88-103.

Wintoki, M., J. Linck and J. Netter (2012). Endogeneity and the dynamics of internal corporate governance, *Journal of Financial Economics* **105**(3): 581-606.

World Bank (2019). Women, Business and the Law. Available at https://openknowledge.worldbank.org/bitstream/handle/10986/31327/WBL2019.pdf retrieved 21 September 2021.

Yang, P., J. Riepe, J., K. Moser, K., K. Pull and S. Terjesen (2019). Women directors, firm performance, and firm risk: A causal perspective, *The Leadership Quarterly* **30**(5): 101297.

Zhang, J., H. Zhu and H. Bing (2013). Board composition and corporate social responsibility: an empirical investigation in the post Sarbanes-Oxley era, *Journal of Business Ethics* **114**(3): 381-392.

Zhang, L. (2012). Board demographic diversity, independence, and corporate social performance, *Corporate Governance* **12**(5): 686-700.

Appendix A

Table A.1 summarises the small number of consistently significant results (at the 10 per cent level of significance) that hold across permutations of control variables. Positive results are marked '+', negative marked '-', and results that were not consistently significant for a particular measure of *Female*, despite being consistently significant for another measure of *Female* in the table, are marked 'I'. The ID numbers in the left-most column are used to identify these models in subsequent discussion.

TABLE A.1
SUMMARY OF CONSISTENTLY SIGNIFICANT REGRESSION RESULTS

					Measure of *Female*		
ID	Sample	Model	Lag	Performance measure	Female%	Dummy	Blau
1	ASX200	IV (all)	0	Q	+	+	+
2	ASX500	AB	0	ROA	-	-	-
3		AB	1	ROA	I	-	-
4		RE	0	ROE	+	+	+
5		FE	0	ROE	+	I	+
6		RE	1	ROE	+	+	+
7		FE	1	ROE	+	I	+
8	All firms	FE	0	CFOTA	I	-	-
9		AB	2	ln(Q)	-	-	-
10		FE	2	ln(Q)	I	-	I

Lag refers to the inclusion of a lagged firm performance measure in the regression.
"2" indicates that both the first and second lag of firm performance are included in the regression.

This is a strange collection of results that defies any simple generalisation. However, most of the consistently significant results do not withstand further scrutiny.

The major concern with random effects models is that they may omit time-invariant characteristics such as firm culture, and thus produce biased coefficients (Adams and Ferreira 2009). Stata's random effects post-estimation commands provide a version of the Hausman test that is robust to heteroscedasticity. Applying this test to the random effects models at hand yields fairly overwhelming rejections of the null. These results yield strong evidence that the random effects estimates are biased, that the significant results in respect of models ID4 and ID6 are unreliable, and that fixed effects models are to be preferred.

Stata also provides an instrumental variable post-estimation command that is a version of the Hausman test and is robust to heteroscedasticity. Although the results from this test are not completely consistent across the various instrumental variable models used with the sample here, it generally fails to reject the null (88 per cent of cases at 10 per cent significance). This suggests that despite the findings of Adams and Ferreira (2009) and Vafaei, Ahmed and Mather (2009), there may be no great benefit in treating female measures as endogenous in many of the models implemented here. This said, the significant endogeneity test results do tend to cluster. All bar one the 24 test results for the regressions using the ASX200 sample with no lag, Q as the measure of performance and instrumenting with 'IV all' are significant at 5 per cent, suggesting that in that specific case, both random effects and fixed effects models might be unreliable, and that the use of instrumental variables in model ID1 is justified. Half of the test results are significant (at 5 per cent) for the ASX500 sample with no lag and ROE as the measure of performance, regardless of which instrument is used. This suggests that the consistently significant results for models ID4 and ID5 may not actually be 'consistent'. But the results do not reject the ASX500 lag 1 version (models ID6 and ID7), nor the all firms no lag CFOTA model (ID8) and lag 2 ln(Q) model (ID10).

Arellano-Bond modelling

Results from Hansen tests for the Arellano-Bond models in Table A.1 are presented below in Table A.2.

TABLE A.2
HANSEN TEST P-VALUES FOR SIGNIFICANT ARELLANO-BOND MODELS

			Firm size	***ln(assets)***				***ln(market capitalisation)***				***Min. value***
			Leverage	***Financial leverage***		***Leverage***		***Financial leverage***		***Leverage***		
			Board size	***no.***	***ln***	***no.***	***ln***	***no.***	***ln***	***no.***	***ln***	
All firms	**Lag 2**	**ln(Q)**	**Female %**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			Dummy	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			Blau	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ASX500 firms	**No lag**	**ROA**	**Female %**	0.337	0.198	0.256	0.162	0.467	0.512	0.405	0.379	0.162
			Dummy	0.279	0.137	0.203	0.117	0.449	0.509	0.327	0.368	0.117
			Blau	0.274	0.154	0.247	0.147	0.522	0.553	0.418	0.408	0.147
	Lag 1	**ROA**	**Female %**	0.403	0.211	0.327	0.187	0.540	0.466	0.417	0.380	0.187
			Dummy	0.661	0.493	0.337	0.212	0.619	0.509	0.398	0.386	0.212
			Blau	0.453	0.274	0.304	0.172	0.584	0.487	0.390	0.351	0.172

For the models estimated across all firms with ln(Q) as the measure of performance, the Hansen test consistently yields highly significant results for all versions of the Arellano-Bond model. A significant value for this test indicates that the instruments formed by using two lags of the dependent variable are not valid. This casts doubt on the results from the models of Table A.1 which are summarised in line ID9.

Many of the higher p-values for the models estimated on ASX500 firms provide little comfort, given Roodman's (2008) caution that Hansen (1982) test results can be unreliable, and even p-values as high as 0.25 should be viewed with concern. Although the instrument count (around 200 in each of these models) is relatively low compared to the number of observations (around 4,000 for the ASX500 models and 8,000 for the `all firms' models), this does not safeguard the Hansen test from potential under-rejection of the null (Roodman, 2008). Overall, the Hansen test results undermine many of the individual results. And applying the criteria used for a 'consistently significant result' in section 7.5.1 suggests that none of the Arellano-Bond results hold across all permutations of controls.

Appendix B

Notes on Table B.1:

- Green rows: studies which implement econometric techniques to control for endogeneity and address financial performance
- Orange rows : studies which implement econometric techniques to control for endogeneity but do not address financial performance
- Rows highlighted blue: event studies (discussed in detail in section 5.1.2)

TABLE B.1
SUMMARY OF INDIVIDUAL FINDINGS FROM THE LITERATURE

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Adams & Ferreira 2009	US	S&P500, S&P MidCap, S&P SmallCap – 1,939 firms, N=8,253	1996-2003	Q, ROA	%	Negative	'[G]ender-diverse boards allocate more effort to monitoring … average effect of gender diversity on firm performance [ROA, Q] is negative' (p. 291).
Adams et al. 2011	Australia	ASX-listed firms – an initial total of 1,126 appointments (of both genders) resulting in samples of various sizes	2004-2006	Share price	Appointment (event study)	Positive	Significantly positive market reaction to female appointments (but not male appointments) in most samples, and evidence of a significantly positive effect relative to male appointments in some subsequent regression models.
Ahern & Dittmar 2012	Norway	All OSE PLCs – 248 firms, N=1,230	2003 event; 2001-2009 data	Q, share price	%, dummy (>0)	Negative	'[T]he quota caused a significant drop in the stock price at the announcement of the law and a large decline in Tobin's Q over the following years' (p. 137).
Ali et al. 2014	Australia	'Large' ASX firms – 288 firms	2012	Employee productivity, ROA	Blau Index	Positive	'The results indicate a positive linear relationship between gender diversity and employee productivity' (p. 497) but 'gender diversity did not have a significant effect on [ROA]' (p. 502). The linearity of the relationship weakens the critical mass argument (p. 507).
Anderson et al. 2011	US	Russell 1000 industrial – 615 firms, N=1,230	2003 & 2005	Q (EVA for robustness testing)	Index (quartiles for %, contributing to overall heterogeneity index)	Mixed	'[I]nvestors place valuation premiums on heterogeneous boards in complex firms but discount heterogeneity in less complex firms' (p. 5) but '[o]ccupational heterogeneity … appears to yield significantly larger, positive economic effect [*sic*] on firm performance than social [including gender] heterogeneity' (p. 27).

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Bear et al. 2010	US	FAMA healthcare only – 51 firms	2007, 2009 (lagged)	FAMA, mediators (KLD ratings)	No., Blau Index (for diversity generally)	Positive	'[T]he percentage of women on the board was positively associated with corporate reputation' and this effect was mediated by 'CSR institutional strength ratings' (p. 217).
Bilimoria 2006	US	Fortune 500 – 447 firms	1999-2000	NA	No. women officers, no. women holding line positions, dummy (women make up ≥25% of corporate officers)	NA	'[T]here is a positive relationship between the presence of women corporate directors and the representation of women executives in ... top management teams ... the number of women officers, women line officers, a critical mass of women in the executive team, women officers with clout titles, and top-earning women officers' (p. 57).
Bøhren & Strøm 2010	Norway	All OSE non-financial firms – N=1,290	1989-2002	Q, ROA, ROS (TRS)	%, % shareholder-elected	Negative	[T]he firm creates more value [Q, ROA, TRS] for its owners when ... gender diversity is low' (p. 1281).
Bonn 2004	Australia	ASX500 manufacturing only – 84 firms	1999, 2003	ROE, MBR	%	Positive	'[T]he ratio of female directors was positively related to [MBR]' (p. 20) but negative and insignificant for ROE (p. 19).
Boulouta 2013	US	S&P500 – 126 firms, N=594	1999-2003	KLD	%	Positive	'BGD [board gender diversity] has a significant impact on the negative social practices of CSP ... and the higher the BGD the lower these practices will be' (p. 193).
Brammer et al. 2009	UK	UK PLCs – 199 firms	2002	Britain's most admired companies survey	Dummy (>0)	Mixed	'[T]here is a positive reputational effect of a female presence at board level in only the consumer services sector ... a negative reputational effect ... in the producer services sector, and no significant reputational effect in all other industries' (pp. 25-26).
Cai et al. 2006	UK	Non-financial, non-FTSE350 firms – 114 firms	1999-2003	AS, PIN, MIN	Log %	Positive	'[T]he more diverse the board in terms of the presence of female directors, the less the information asymmetry in the market' (p. 782), but this result is only significant for one measure (AS) and insignificant for the other two.
Carter et al. 2010	US	S&P500 – 641 firms, N=2,563	1998-2002	Q, ROA	No., no. on various committees	Insignificant	'3SLS regressions provide no support for a link between ... gender ... diversity of the board and board committees and financial performance (p. 411).

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Catalyst 2004	US	Fortune 500 – 353 firms	1996-2000	ROE, TRS	% on 'top management teams'	Positive	'The group of companies with the highest [by quartile] representation of women on their top management teams experience better financial performance [ROE and TRS] than the group of companies with the lowest' (p. 2).
Chen et al. 2016	US	S&P1500 – N=13,248 (H1), N=2,825 (H2)	1998-2010	No. acquisitions and size thereof	%	NA	'[G]reater female board representation was negatively associated with both overall firm acquisitiveness and target acquisition size' (pp. 311-312).
Dale-Olsen et al. 2013	Norway	PLCs (128 firms) and LTDs (36,924 firms)	2003-2007	ROA	%	Insignificant	'The impact of the [quota] reform on firm performance is negligible [but] following the reform PLCs have to a larger extent accumulated capital financed by debt or debt and own capital' (p. 110).
Dickman et al. 2004	US	Wisconsin 50	1992-2002	ROI, ROE, EPS, TRS	Dummy (>0)	Positive	'Significant positive differences in 2002 between companies with no women directors and companies with women directors for each of the three accounting-based performance measures', but the results were not significant in 1992 (p. 15). TRS over the period 'for companies with women directors is 37 percent more than for those companies without' (p. 16).
Dimovski et al. 2013	Australia	37 Australian Real Estate Trusts, N=203	2006-2011	NA	No., dummy (>0)	NA	Real estate investment trusts based in Sydney and/or 'with larger boards are more likely to employ a woman director and indeed more women directors' (p. 196).
Dobbin & Jung 2011	US	Fortune 500 (supplemented) – 432 firms, N=2,883 to 3,016	1996-2007	Q, ROA	No.	Mixed	'The fact that board diversity has no effect on profits [ROA], but a negative effect on stock price [Q], lends support to our thesis that institutional investors may sell the stock of firms that appoint women to their boards ... because they are biased against women' (p. 828).
Dowling & Aribi 2013	UK	FTS100 (exc. financial firms) – 96 firms, N=681	2000-2011	No. acquisitions	%, no., dummy (≥2)	NA	'[T]he presence of female directors is related to reduced levels of large acquisitions' (p. 85), but there is not strong support for the 'critical mass' hypothesis.

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Farrell & Hersch 2005	US	Fortune 500 and Service 500 (excluding financial) firms – N=111 female director appointments	1990-1999	Share price	Appointment (event study)	Insignificant	'[W]omen tend to serve on better performing firms [and there are] insignificant abnormal returns on announcement of a women added to the board' (p. 85).
Galbreath 2011	Australia	ASX200 – 151 firms	2004, 2005-2007	ROE, ROA, MBR, environmental quality; social responsiveness	%, dummy (≥2)	Positive	The relationship between the proportion of female directors and ROE, MBR and social responsiveness is positive and significant, but is not significant for ROA or environmental quality. There is also 'some level of support for critical mass theory' (p. 26).
Greene et al. 2020	US (California)	Firms headquartered in California and listed on major exchanges – 602 firms	2018	Impact of California Senate Bill SB 826 on share price (event study)	Various measures in subsequent regressions, e.g.: no., 'gap', 'gap %', 'add female director' dummies	Negative (on Governor's signature of SB 826)	'[T]he negative reaction is likely driven by the mandated quota, which constrains board composition and imposes additional costs on the firm' (p. 11). '[F]irms in industries with a greater supply of female candidates and firms that can more easily replace existing directors or more easily attract female directors, are less negatively affected by the mandate' (p. 19).
Gregoric et al. 2009	Denmark Iceland Finland Norway Sweden	Listed Nordic firms – 431 to 757 firms, N=2,692	2001-2007	Q, ROA, annual asset growth	%, dummy (>0)	Insignificant	'[P]erformance differences are not statistically significant ... [F]emale board membership is associated with higher CEO pay levels ... whereas female remuneration committee membership is associated with lower CEO pay [in 2006]' (p. 4).
Gregory-Smith et al. 2014	UK	FTSE350 – 1,983 firms, N=11,515	1996-2011	Log PBR, ROA, ROE, TSR	%	Insignificant	'[N]o support for the argument that gender diverse boards enhance corporate performance' (p. 109).
Gul et al. 2011	US	Initial set not stated, exc. utilities and financials – N=4,084	2001-2006	Stock price informativeness	%, no., % NEDs, dummy (>0), dummy (≥5), dummies (female CEO, female chairman, female CEO duality)	Positive	There is 'a positive link between gender diversity and stock price informativeness' (p. 336) and 'the relationship is stronger for firms with weak corporate governance' (p. 314). The relationship is stronger when there are five or more female directors than merely a female director dummy (p. 324).
Hafsi & Turgut 2013	US	S&P500 – 95 firms	2005	KLD	Index (tercile split for diversity generally)	Positive	Gender diversity has significant positive impact on CSP, and this is mediated by 'structural diversity of boards' (p. 463).

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Hagendorff & Keasey 2012	US	Publicly listed commercial banks – N=148 bank merger announcements	1996-2004	Market reaction to bank merger announcements	% (at time of event)	Insignificant	The gender diversity of bidding banks' boards 'does not lead to measurable value effects' (announcement returns to mergers) (p. 41).
Harjoto et al. 2015	US	Set based on data availability – 1,489 firms, N=9,001	1999-2011	KLD	Blau Index	Positive	Board gender diversity 'is associated with the overall CSR score by increasing CSR strengths and reducing CSR concerns' and more strongly so in 'industries with greater need for stakeholder management' (p. 642).
Haslam et al. 2010	UK	FTSE100 – 126 firms	2001-2005	Q, ROA, ROE	%, dummy (>0)	Negative	'[T]here was no relationship between women's presence on boards and "objective" accountancy-based measures of performance … However, consistent with "glass cliff" research there was a negative relationship [with] "subjective" stock-based measures of performance' (p. 484).
Hillman et al. 2007	US	1000 largest US firms by sales – 950 firms N=9,722	1990-2003	NA	Dummy (>0)	NA	'[O]rganizational size, industry type, firm diversification strategy and network effects (links to other boards with women directors) significantly impact the likelihood of female representation on boards' (p. 941).
Horváth & Spirollari 2012	US	S&P500 – 136 firms, N=544	2005-2009	PBR	%, no., dummy (>0)	Insignificant	'[var.] measures for gender diversity are never found to be statistically significant' (p. 479).
Huse et al. 2009	Norway	Survey of board members – full data from 386 respondents	2006	Level of creative discussions	%, no.	Insignificant	'[T]he contribution of women to creative discussions only existed when the women had a different background from the men. However, this relationship was weak' (p. 592).
Kathyayini et al. 2012	Australia	100 largest ASX firms – 96 firms	2008	Environmental reporting	%	Positive	There exists 'a significant positive relationship between the extent of environmental reporting and the proportion of … female directors on a board' (p. 143).
Kim et al. 2013	US	128 firms with SEC violations matched to 128 controls	2003-2010	SEC violations (AAERs)	%, % on audit committee	Positive	'SEC violations are less likely when the board has more women' (p. 142).
Larkin et al. 2012	US	Fortune 500 – 449 firms	2010	Share price	Dummies (0, 1, >1)	Positive	Findings suggest an interactive effect between corporate reputation and the number of female directors – i.e. companies on 'most ethical' lists have more women and better financial performance.

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Lee & James 2007	US	Firms with available appointments in Factiva searches (p. 231) – 1,624 events, of which 17 were female CEO and 69 were female TMT appointments	1990-2000	Share price	Appointment (CEO or TMT)	Mixed	Market reaction is significantly more negative for female CEO appointments than for males; but there is no significant gendered difference for TMT appointments.
Mallin & Michelon 2011	US	Business Ethics 100 Best Corporate Citizens – 176 firms, N=278	2005-2007	KLD	%	Positive	'[T] proportion of … female directors [is] positively associated with corporate social performance' (p. 119) and with 'employee relations and human rights performance' (p. 120).
Mallin et al. 2013	US	Business Ethics 100 Best Corporate Citizens – 135 firms, N=221	2005-2007	KLD	%	Insignificant	'[N]o significant associations found between proportion of women and [social and environmental] disclosure' (p. 38).
Miller & Triana 2009	US	Fortune 500 – 326 firms	2002-2005	Fortune Corporate Reputation Survey, R&D expenditure, ROI, ROS	%, Blau Index (for diversity generally)	Insignificant	Board gender diversity is positively related to innovation, but not to reputation, and has no significant impact on firm performance.
Nguyen & Faff 2006	Australia	ASX500 – N=832	2000-2001	Q	%, dummy (>0)	Positive	'[T]he presence of women directors is associated with higher firm value' (p. 24).
Nguyen et al. 2015	US	145 banks – 52 external director appointments, of whom 15 were female	1999-2011	Share price (event study)	External director appointment	Insignificant	Although appointments are found to have a significantly positive effect on share price, the gender of the appointee is not significant in subsequent regressions.
Nielsen & Huse 2010	Norway	All OSE firms, publicly traded firms, and private companies with ≥50 employees – useful survey data from 272 respondents	2005	Board strategic involvement (6 factors)	NA	Mixed	'Women directors influence board strategic involvement through their contribution to board decision-making, which in turn depends on women directors' professional experiences and the different values they bring along … perception of women as unequal board members may limit their potential contribution to board decision-making' (p. 16).

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Pathan & Faff 2013	US	Top 300 bank holding companies – 212 firms, N=2,640	1997-2011	ROAA, ROAE, PTOI, NIM, Q, mean daily stock returns	%	Positive	The effect of female board representation on performance is different in different periods: positive pre-SOX (1997-2002), but insignificant post-SOX (2003-2006) and in crisis periods (2007-2011). It is positive overall. However, these findings do not hold for NIM and Q.
Schnake et al. 2006	US	SEC's Edgar database (financial services sector) – 192 firms	1998-2002	10K investigations	%, dummy (> mean %)	Positive	There are interaction effects 'such that a higher number of women coupled with longer average board tenure results in higher firm social performance (i.e., the fewer the number of 10K investigations brought against the firm). No link was found … for the basic materials firms' (p. 31)
Shrader et al. 1997	US	200 largest firms by market value	1992, 1993	ROA, ROE, ROI, ROS	%, no. (for board, management & top management)	Negative	Female board representation is negatively associated with all measures of performance in 1992, and all but ROE in 1993. However, a positive link across all measures exists for female managers in 1992, but only against ROS in 1993.
Singh et al. 2001	UK	FTSE100	1999-2000	Firm size, profit, turnover	Dummy (>0)	NA	'Female directors were to be found in the bigger, more profitable firms, and particularly those with the largest turnover' (p. 212).
Srinidhi et al. 2011	US	All firms with available data – N=2,480	2001-2007	Earnings quality	Dummies (female director, female NED, female director on the audit committee)	Positive	'[F]irms with greater female representation on their board exhibit higher earnings quality' and the 'results are consistent for each of the female board participation measures' (pp. 1638-9)
Torchia et al. 2011	Norway	OSE listed firms and others – 317 CEO survey responses	2005/2006	Innovation (board perception)	Dummy (0, 1, 2, ≥3)	Positive	At least three women board members, (but not one or two) 'makes it possible to enhance the level of firm innovation' but this is 'mediated by board strategic tasks' (p. 299).

Study	Country	Sample of firms	Year(s)	Performance measure(s)	Female directors measure(s)	Findings	
Triana et al. 2013	US	Fortune 500 – 462 firms	2002-2004	Strategic change	Blau Index, power	Mixed	'Results support a three-way interaction, indicating that when the board is not experiencing a threat as a result of low firm performance and women directors have greater power, the relationship between board gender diversity and amount of strategic change is the most positive. However, when the board is threatened by low firm performance and women directors have greater power, the relationship between board gender diversity and amount of strategic change is the most negative' (p. 609).
Vafaei et al. 2015	Australia	ASX500 – 224 firms, N=1,101	2005-2011	Q, ROA, ROE, CFO/TA	%, dummy (>0)	Positive	Board gender 'diversity is positively associated with financial performance' across all measures (p. 413).
Wang & Clift 2009	Australia	ASX500 – 243 firms	2000-2003;2003-2006	ROA, ROE, shareholder return	%, no.	Insignificant	[G]ender and racial diversity do not have significant influence on performance ... [but] larger firms tend to have relatively more female members' (p. 88).
Westphal & Bednar 2005	US	Mid-sized US companies (sales of $0.5m-$1m) – 228 board survey responses	Not stated	Strategy	Blau index	Negative	As outside director homogeneity decreases, the difference between the level of director's concern about strategy and their perception of others' concern increases for low performing firms. Pluralistic ignorance is reduced when boards are more homogeneous.
Williams 2003	US	Fortune 500 – 185 firms	1991-1994	Charitable contributions	%	NA	'[F]irms having a higher proportion of women serving on their boards do engage in charitable giving to a greater extent … in the areas of community service and the arts [but not in] education or public policy issues' (p. 1).
Zhang 2012	US	Fortune 500 – 475 firms	2007-2008	KLD	Blau index	Positive	'[B]oard gender diversity is positively related to institutional and technical strength [CSR] ratings' (p. 686).
Zhang et al. 2013	US	FAMA – 516 firms	2007	FAMA, KLD	%	Positive	'[T]he greater presence of … women directors is linked to better CSR performance within a firm's industry' (p. 381).

References: Appendix B

This reference list is for papers in Table B.1 that do not appear in the reference list of the main paper.

Bilimoria, D. (2006). The relationship between women corporate directors and women corporate officers, *Journal of Managerial Issues* **18**(1): 47-61.

Bonn, I. (2004). Board structure and firm performance: evidence from Australia, *Journal of Australian and New Zealand Academy of Management* **10**(1): 14-24.

Boulouta, I. (2013). Hidden connections: the link between board gender diversity and corporate social performance, *Journal of Business Ethics* **113**(2): 185-197.

Brammer, S., A. Millington and S. Pavelin (2009). Corporate reputation and women on the board, *British Journal of Management* **20**: 17-29.

Cai, C., K. Keasey and H. Short (2006). Corporate governance and information efficiency in security markets, *European Financial Management* **12**(5): 763-787.

Dickman, A., J. Schmidt, J. Fields and E. Van Dunk (2004). Diversity and Tomorrow's Profits: Women in Corporate Leadership. Available at http://publicpolicyforum.org/sites/default/files/WomenIncReport.pdf. Retrieved 16 October 2016.

Gregoric, A., L. Oxelheim, T. Randøy and S. Thomsen (2009). Corporate Governance as a Source of Competitiveness for Nordic Firms. Available at www.nordicinnovation.org/Global/_Publications/Reports/2009/Corporate%20governance%20as%20a%20source%20of%20competitiveness%20for%20Nordic%20firms.pdf. Retrieved 5 March 2017.

Haslam, S., M. Ryan, C. Kulich, G. Trojanowski and C. Atkins (2010). Investing with prejudice: the relationship between women's presence on company boards and objective and subjective measures of company performance, *British Journal of Management* **21**(2): 484-497.

Huse, M., S. Nielson and I. Hagen (2009). Women and employee-elected board members, and their contributions to board control tasks, *Journal of Business Ethics* **89**(4): 581-597.

Mallin, C. and G. Michelon (2011). Board reputation attributes and corporate social performance: an empirical investigation of the US best corporate citizens, *Accounting and Business Research*, **41**(2): 119-144.

Mallin, C., G. Michelon and D. Raggi (2013). Monitoring intensity and stakeholders' orientation: how does governance affect social and environmental disclosure? *Journal of Business Ethics* **114**(1): 29-43.

Nielsen, S. and M. Huse (2010). Women directors' contribution to board decision-making and strategic involvement: the role of equality perception, *European Management Review* **7**(1): 16-29.

Shrader, C., V. Blackburn and P. Iles (1997). Women in management and firm financial performance: an exploratory study, *Journal of Managerial Issues* **9**(3): 355-372.

Singh, V., S. Vinnicombe and P. Johnson (2001). Women directors on top UK boards, *Corporate Governance* **9**(3): 206-216.

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

Westphal, J. and M. Bednar (2005). Pluralistic ignorance in corporate boards and firms' strategic persistence in response to low firm performance, *Administrative Science Quarterly* **50**(2): 262-298.

Williams, R. (2003). Women on corporate boards of directors and their influence on corporate philanthropy, *Journal of Business Ethics* **42**(1): 1-10.