Breaking the Boardroom Ceiling: The Effect of Gender Diversity Mandates on Firm Performance

Moksh Garg

Sloan School of Management

May 21, 2024

Abstract

Firms have become increasingly receptive to marginalized groups including women in recent decades. For the most part, the significant progress made over the last few years can be attributed to state affirmative action – such as instituting quotas for women on corporate boards. These policies have not only reduced the scope of managerial discretion but also provided much-needed push for gender-inclusive corporate leadership. Intrigued by some of these timely changes in the corporate regulatory landscape, I analyze the causal effect of gender quotas on firm performance through this paper. While I acknowledge that there is plenty of research in this area, I also realize that a large chunk of it happens to be descriptive with mixed findings and therefore empirically and even theoretically inconclusive. As a result, through this paper, I try to empirically tease out the causal effect of the inclusion of women on corporate boards on firm performance by exploiting the amendment in the Indian Companies Act 2013, which came into force on April 1, 2014, and made it obligatory for all the listed concerns to appoint at least one woman director. While standard two-period DiD analysis yields negative but not statistically significant results, the staggered DiD analysis with no treatment reversals shows a significant and positive association at a 5% level of significance. The findings indicate that the inclusion of women on corporate boards improves firm performance by 2.30% over short to intermediate periods following treatment. However, these findings need to be interpreted with caution because staggered DiD analysis - 1) is ignorant of the underlying treatment assignment process, and 2) uses a disproportionately small number of never-treated firms to estimate ATT.

1 Introduction

The beginning of the 21st century has led to a growing emphasis on moral and social inclusion (Catalyst, 2022; UN Women, 2021). As a corollary, it is not surprising that even businesses have become increasingly receptive to historically marginalized groups including women. This is partly because of ethical considerations, i.e., it is the right thing to do, and partly because of underlying instrumental motives, i.e., embracing diversity and inclusion as a strategic lever to better performance outcomes (Nishii, 2013; Kossek et al., 2017). However, it is important to highlight that for many businesses, it is not the managerial willingness and volition that has propelled diversity but the obligation to remain legally compliant. Terjesen & Sealy (2016) emphasize how the promulgation of affirmative actions and policies over the recent years, such as the imposition of gender quotas on corporate boards, has been a major catalyst in improving the gender balance of top management teams. As a case in point, the amended Indian Companies Act, 2013[1] has led to a substantive increase in the percentage share of women directors on Indian corporate boards – almost quadrupling the percentage of women directors on corporate boards from 4% in 2014 to 16% in 2020 (Garg & Agarwal, 2023; Agarwal, 2023). To that extent, regulatory action has been successful in catalyzing positive social change through the use of sanction and legal-rational authority. Interestingly, while these laws have ushered in an outlook that promotes gender balance in top management teams, there is a lack of scholarly consensus about their impact on performance outcomes. As a matter of fact, the association between gender diversity and firm performance remains empirically as well as theoretically ambiguous in the extant literature. Keeping that in mind, I motivate this paper to explore and disentangle the effect of increased gender representation on firm performance with causal meaning and interpretation using a combination of tools and techniques learned this semester.

I specifically use difference-in-differences and a combination of some of its many variants – including both staggered and non-staggered – to systematically tease out the causal effect of increased gender diversity on firm performance. The use of DiD comes in handy in exploiting one of the landmark shifts in corporate law in recent decades: the amendment in the Indian Companies Act 2013 (henceforth, ICA 2013). The revised law came into force on April 1, 2014, and made it obligatory for all the listed concerns alongside other selected groups of companies to appoint at least one woman director. Interestingly, any failure to comply with the law would attract severe penalties, making it extremely costly for firms to digress. The revised ICA 2013, therefore, offers a suitable context to use the traditional two-period DiD estimator and ascertain the causal effect of gender diversity on firm performance. Furthermore, as outlined above, it was not the law alone that induced compliance. It can be observed from Figure 1 itself that a sizeable proportion of firms recruited women on corporate boards for reasons other than legal compliance, indicating a staggered adoption of treatment as opposed to one-shot treatment over an extended period of time.

Taken together, from a methodological standpoint, I believe that subject to underlying assumptions being satisfied, the use of DiD as a generalized framework will also help me deduce a more conclusive answer to the impending research question, which I believe is inherently endogenous and has been subject to empirical tautology in the prior work. The remainder of this paper is organized as follows. The subsequent section introduces research hypotheses and the major theoretical perspectives that guide the flow of my conceptual thinking. Then, section 3 explains the data and methodology involved in the empirical testing of hypotheses and also highlights the identification assumptions. Sections 4 and 5 provide a detailed assessment of results and supplementary analyses. Finally, section 6 concludes the paper and provides avenues for future research.

[1]As per the second Proviso to Section 149 (1) read with Rule 3 of the Companies (Appointment and Qualification of Directors) Rules, 2014, following set of companies has to appoint at least one women director:
(1) Every Listed Company
(2) Every Public Company having a paid-up share capital of INR 1 billion or more
(3) Every Public Company having a minimum turnover of INR 3 billion or more

——————————————
Insert Figure 1 here
——————————————

2 Research Hypotheses

As part of the bigger project that I am pursuing, I am interested in validating two specific hypotheses:

H1: All else equal, the inclusion of women on corporate boards improves firm performance.

I expect increased gender diversity at the board level to foster diversity of thought and ideas, which can translate to more creative and innovative decision-making improving firm performance and strategic prospects. However, a logical counterargument would be that increased diversity can turn dysfunctional as it may lead to greater conflicts, delay in decision-making, and disrupt organizational agility in a rapidly changing fast-moving business environment. Therefore, the relationship between gender diversity and firm performance remains puzzling as it is probable that increased diversity does greater harm than good. Yet, I surmise a positive association because it seems an inverse-U-shaped relationship exists and the increased levels of board diversity should heighten firm performance in the initial period before the positive impact tapers off. Given that boards are yet to become highly diverse, I do not see the inclusion of women to create negative spillovers for firm performance. Instead, women are likely to endow boards with unique skills and resources that can amplify firm performance (Hillman et al., 2007; Kim & Starks, 2016)

H2: All else equal, the improvement in firm performance due to increased board diversity is greater for family firms than non-family firms.

I am further interested in analyzing how family affiliation of businesses, either through ownership or management, moderate association between diversity and firm performance. This is because prior literature has established that family firms as structurally and behaviorally distinct from their non-family counterparts (Chrisman et al., 2004; Gomez-Mejia et al., 2011). This makes me wonder if the benefits from board diversity are going to be different, more specifically greater, for family firms than nonfamily firms. I hypothesize a positive interaction effect because – first, family firms are less professionalized and known to have sub-standard corporate governance practices and therefore stand a greater chance to gain through such initiative (Villalonga et al., 2015); and second, family firms have a long-standing reputation of being conservative and patriarchal and therefore the treatment effect is likely to be more pronounced (Amore et al., 2014; Mubaraka & Kammerlander, 2023). Overall, it looks plausible that the impact of gender diversity on corporate boards is going to be larger for family firms than non-family firms.

However, for this particular term paper, I shall exclusively focus on evaluating H1.

3 Data and Methods

I am interested in estimating the average treatment effect among treated (ATT) as my target estimand. To identify ATT using my baseline model, I make four important assumptions as described below.

Identification Strategy

I rely on two-period classic DiD using data for financial years ending on March 31st, 2014, and March 31st, 2015, i.e., for the years just before and after the law came into effect to estimate my baseline model. Specifying a fixed time window helps anchor the analysis for a statistical deep dive. While it may be argued that the two-way fixed effects (TWFE) model is rather too simplistic and not quite realistic given my sampling frame, I believe it offers a good starting point for me to estimate the average treatment effect ignoring the complications arising due to staggered adoption of treatment as well as treatment reversal observed in the underlying data.

$$\pi_{it}(D) = \alpha_i + \gamma_t + \tau_{2fe} D + \epsilon_{it}$$

where τ_{2fe} estimates the target estimand ATT. D is the treatment indicator and α_i and γ_t are unit and time fixed effects respectively.

I consider a slightly extended time window to identify treated and untreated units. The extended window includes three additional years before 2014. The underlying assumption is that the firm with no women on corporate boards from 2011 to 2014, i.e., four years before the legislation kicked in, is untreated, even though it might have been treated in the distant past. Following this logic, it is assumed that firms with no female directors during the pre-treatment window but at least one female director after the law came into effect are treated. Even though these assumptions ignore or gloss over the treatment history of firms until 2011, they are plausible to the extent we assume treatment affects short-term performance and the impact would have subsided fully and not carried over to subsequent years. Given this setup, ATT is identified if the following assumptions hold.

Please note that for the assumptions stated below, G is a binary variable indicating whether a firm is affected by the law (1: if yes; 0: otherwise). T is also a binary indicator that takes the value of 1 for the period after treatment has occurred and or else 0. Finally, D, a binary variable again, indicates whether the firm is designated as treated or not.

1. *Assumption 1: Change in the expected value of firm performance π for control firms immediately before and after the year of treatment offers a credible counterfactual for the change in the expected value of π for treated firms had no treatment occurred. This assumption is commonly known as the parallel trends assumption.*

   In other words, had no treatment occurred:

$$\mathbb{E}[\pi_{i,1}(0) - \pi_{i,0}(0)|Z_i = 1] = \mathbb{E}[\pi_{i,1}(0) - \pi_{i,0}(0)|Z_i = 0]$$

2. *Assumption 2: There is perfect compliance, i.e., firms assigned to treatment take the treatment, and firms assigned to control do not take the treatment.*

   I assume,

$$D = ZT$$

3. *Assumption 3: Treatment effect is homogeneous and contemporaneous. In other words, there is a constant treatment effect is realized instantaneously with no spillovers,*

   By extrapolation, this also means that a firm with no women on corporate boards in any given year can be designated as untreated, even if it were treated in the immediate or distant past.

4. *Assumption 4: The unobserved confounding is additive and time-invariant.*

$$\pi_{1it}, \pi_{0it} \perp\!\!\!\perp Z_{it} | \alpha_i, \gamma_t$$

   Here α_i and γ_t indicate group-specific and time-specific fixed confounding respectively. After conditioning on unit-level fixed effects α_i and year-specific fixed effects γ_t, the potential outcomes are independent of treatment assignment.

Test for identification assumptions

Before estimating ATT, it is a must that the parallel trends assumption holds. The parallel trends assumption is the canonical assumption underlying the DiD framework. It posits that the expected value of the outcome variable for the treated units would have followed the same trajectory had it not been treated. Fundamentally, counterfactual for the change in outcomes for the treated group is imputed using the change in outcomes for the control group. Correspondingly, to evaluate the plausibility of this assumption, I test for the equality of means in ROA for treated and untreated firms in the pre-treatment time period, i.e., from 2010 to 2014, using the event study plot and placebo regressions. To undertake this analysis, I consider and analyze data for the extended time window – including four additional years before 2014. Based on the visualization of observed means of ROA for treated and untreated firms in the pre-treatment time period as well as the results derived using placebo regressions, the parallel trend assumption seems to hold.

———————————

Insert Figure 2 here

———————————

———————————

Insert Table 1 here

———————————

Assumptions 2 to 4 are also quite strong and restrictive. If we look at the data, it becomes obvious that assumption 4 does not hold as both treated and control firms, based on treatment assignment, take up treatment. Similarly, assumption 3 is also quite limiting as it assumes that increased diversity will not have carryover effects, i.e., treatment in a given year will not affect firm performance in future periods. Further as shown in Figure 3, even assumption 4 is less likely to hold considering treated and untreated firms are significantly different from each other based on firm size and age. Cumulatively speaking, these assumptions do not seem quite realistic. Therefore, I relax each of these assumptions, one at a time, by estimating a different model to test the robustness of my baseline estimates.

However, it is important to note that for each of these checks, I am trying to compute a different estimate that reflects a very different quantity than the quantity that my baseline model estimates. For example, in Model 4, I will be computing ATT for a very small subset of treated firms as matching would substantially drop treated firms with no comparison or untreated firms with similar observed covariates. Similarly, in Model 2, I will be calculating the ATT amongst treated compliers or LATE using the Wald DiD estimator[2]. For Model 3 with dynamic treatment effects, I compute the average treatment effect based on dynamic aggregation. However, following such a procedure, I also become oblivious to the treatment assignment process and rather assume that firms are getting treated at their own volition.

——————————

Insert Table 2 here

——————————

Sample Data

I test my research hypotheses using the data available for Indian publicly-listed firms. Consistent with the previous research in the Indian context, I use the proprietary database of CMIE Prowess to obtain financial information on Indian firms (Bertrand et al., 2002; Mani & Moody, 2014; Mishra & Suar, 2010). This yields an unbalanced panel dataset of 33,026 firm-year observations for ten years, i.e., from 2011 to 2020. To obtain board and director-level data for different firms, I use the National Stock Exchange's proprietary database called NSE Infobase. It is important to note that the database did not provide information about the gender of the director. Therefore, to construct my key independent variable of treatment onset, I inferred the gender of the director for each firm-year combination from their names using genderize.io API. The API uses more than a billion names as its training dataset and has been employed in prior gender-related research for its near-brilliant accuracy in predicting the gender of an individual based on their name (Koffi, 2024; Brück, 2023). This procedure helps in determining whether a firm i in a year t has at least one woman included on their board. This information is coded using the binary variable g_{it} that takes a value of 1 for firm i with women-included board in year t and 0 otherwise.

Dependent Variable

I use firm performance as a key outcome variable for my hypotheses. I acknowledge that the firm performance is subject to manifold expositions and could be measured differently depending on the precise meaning attributed to it. For my research, however, I define firm performance strictly in an economic sense. Consistent with previous literature on management and strategy, I choose Return on Assets (π) as the most suitable indicator of a firm's profitability or performance (Campbell & Mínguez-Vera, 2007; Kulich et al., 2011; Post & Byron, 2014b). This is to say that if a company witnesses an increase in its ROA on a year-on-year basis, then the firm performance has improved on a year-on-year basis or vice versa. To operationalize the construct, I use Profit after Tax (PAT) net of the prior period and extraordinary transactions expressed as the percentage of the total assets of the company at the end of the financial year to compute ROA (π) (Kowalewski, et. al., 2010). Correspondingly, π is computed as below:

[2]Given the complexity of my data, I have decided to refrain from considering treatment reversal, i.e., my analysis does not incorporate firms that switch from treated to untreated.

$$\pi_{it} = \frac{\text{Profit after Tax net of extraordinary items for firm i}}{\text{Total assets of firm i at the end of financial year t}}$$

where: π_{it} = ROA for firm *i* in year *t*

Independent Variables

For the standard two-period DiD analysis, it is important to code two binary indicator variables as key independent variables: one that categorizes units as treated and untreated and one that measures the onset of treatment[3]. ATT is then estimated using the interaction effect between these two binary indicators. The first variable is called *treat* and the second variable is called *post*[4]. They are coded as 1 or 0 based on the formulation below:

$$\text{treat} = \begin{cases} 1 & \text{if } g_{it} = 1 \\ 0 & \text{if otherwise} \end{cases}$$

$$\text{post} = \begin{cases} 1 & \text{if } t = 2015 \\ 0 & \text{if otherwise} \end{cases}$$

Covariates

I also benchmark treated and untreated firms on a bunch of covariates for some additional analyses to identify if these firms are systematically different from one another on any of these parameters:

- *Firm Age*: There is a possibility that differences in π accrue because of differences in cumulative experience (Bahk & Gort, 1993; Erhardt et al., 2003). I compute firm age using the natural log of the number of years elapsed since the date of incorporation, i.e., the current financial year minus the incorporation year.

- *Firm Size*: The differences in π could result from differences in economies of scale, access to resources, and size (Campbell & Minguez-Vera, 2008; Pitts & Hopkins, 1982). I ascertain firm size using the natural log value of the firm's total assets.

- *Industry Type*: It is also important to determine if we observe significant divergence in industries represented in treated and untreated sub-samples. The difference in industry mix could make treated and untreated fundamentally incomparable and nullify our results if any. I use two-digit NIC codes assigned to each firm to identify their industries.

- *Family*: The previous literature has established family firms as a distinctive form of organization compared to non-family firms. The extant literature also underscores how family firms could

[3] I consider a two-year window of 2014 and 2015 for my baseline model.

[4] It is important to note that in my baseline analysis, a firm is designated as treated based on the actual receipt of treatment notwithstanding whether the firm was intended to be treated by the law or not. That is, a firm not impacted by the amendment could appoint a woman director and thus get treated even though the law did not apply to them. Therefore, I effectively undertake analysis of actually treated firms for my baseline model assuming treatment assignment corresponds to treatment receipt. I relax this assumption using instrumented or fuzzy DiD as an additional robustness check.

behave idiosyncratically because of controlling the family's involvement in business ownership and management. To that extent, I compare the balance of treated and untreated in their family firm composition and address issues that may arise due to any significant difference between them. I ascertain whether a firm is a family firm or not using Thomas Schmidhieny Centre for Family Enterprise's proprietary data[5].

- *Research Intensity*: Finally, I consider the research intensity of a firm as one more plausible driver confounding the association between gender diversity and firm performance. This is because research-centric firms are more likely to embrace innovation and diversity of thought as well as perform better than their peers. Thus, I account for a firm's long-term, entrepreneurial, and innovative nature by computing research intensity as the percentage of total expenditure on Research and Development (in both capital and current accounts) of total sales (Chen et al., 2016; Lopes et al., 1992).

Please refer to Figure 3 for the covariate balance plots for treated and untreated observations.

———————————
Insert Figure 3 here
———————————

4 Results

Table 3 reports descriptive statistics including means, standard deviations, and pairwise correlations among different variables considered for analysis. In line with my intuition, ROA is positively and significantly correlated with firm size ($\rho = 0.296$), firm age ($\rho = 0.031$), and research intensity ($\rho = 0.079$) at a 1% level of significance. This casts some doubt on assumed parallel trends as treatment assignment appears to be non-random and treated firms seem structurally different from control firms. This is because the ICA 2013 mandate fundamentally applies to bigger businesses in terms of overall shareholding and turnover. For the classical TWFE analysis, I still assume parallel trends. To mitigate against concerns arising from structural differences between treated and untreated firms, I shall invoke *conditional ignorability* assumption and combine it with DiD through the selection on observables to create a matched sample of firms for like-to-like comparison. I perform matched DiD in the supplementary analyses section.

———————————
Insert Tables 3 & 4 here
———————————

[5] According to TSCFE, a firm is identified as a family firm if the first condition of a minimum of 20 percent ownership is met and any one of the other two conditions are met: a) at least 20 percent of the equity shares were directly or indirectly- through holding companies, trusts, other corporations and/or people from the same family and b) either i) the management control was with the family or ii) ownership was passed on to the family members from generation to generation or iii) it was intended to pass on the ownership to the next generation.

Table 4 provides the ATT estimate derived using my baseline model with two-way fixed effects as specified under the identification section. We observe a negative coefficient of approximately 0.2% (β = -0.0018; SE = 0.0016; t = -1.15). However, the estimate is not statistically significant. This result indicates that treated firms' financial performance did not significantly differ from untreated firms' financial performance in the immediate year following law implementation.

——————————————
Insert Table 5 here
——————————————

I hold these results preliminary and not quite telling of the actual association between the treatment and outcome. This is precisely because some of my identification assumptions are not very realistic. In fact, some of them are directly contradicted by my data. Thus, I use the classic TWFE estimates as a starting point for my analysis. In the subsequent section, I will relax assumptions 2 to 4, one at a time, and re-estimate ATT or its variant, to see if the non-significant results hold throughout.

5 Supplementary Analyses

In this section, I relax some of my primary identification assumptions and observe how that changes my ATT estimate. As stated earlier, many of these methods will not estimate ATT as their estimand. Instead, they will estimate a slightly different quantity. I will specify the estimand that is being computed in each of the following sub-sections.

5.1 Fuzzy DiD and Imperfect Compliance

For my baseline model, I assume that both treatment assignment and uptake are random. However, they are not. The estimated ATT therefore is unbiased as units, treated or untreated, are self-selecting into treatment. The situation, therefore, could be re-conceptualized as an encouragement design with imperfect compliance. This implies that treatment assignment does not equal treatment delivered and that we observe two-sided non-compliance. Firms that are required by the law to the treatment group fail to take the treatment. In other words, not all firms that were required by law to include women directors on corporate boards complied with the statute. Conversely, some firms that are not legally bound by the amendment, and thus originally assigned to the control group, take the treatment by appointing women on their corporate boards. Therefore, the mismatch between treatment assignment and treatment take-up must be accounted for to obtain a more rigorous causal effect estimate.

I refer to the seminal work of de Chaisemartin and DHaultfoeuille (2018) and estimate a LATE ATT. The estimate effectively measures ATT among those who complied with the treatment assignment, i.e., firms that were assigned to treatment and were treated against firms that were assigned to control and were not treated. This approach is known as fuzzy DiD. The procedure recognizes that both the treated and control populations can receive treatment, albeit to a different extent. As a result, fuzzy DiD estimates ATT by explicitly accounting for one-sided non-compliance by combining the instrumental variable approach and DiD. Just like classical DiD using TWFE, fuzzy DiD can only be applied if certain assumptions are held. For the sake of brevity, I only specify the most critical assumptions and explain whether these assumptions hold in my context.

- $E(D_{11}) > E(D_{10})$

  This is the monotonicity assumption implying that every unit either retains its current status or gets treated, i.e., an untreated unit can become treated but no treated unit can become untreated. Interestingly, a very small percentage of firms in my dataset exhibit this trend (c. 1%). However, assuming it could be because of measurement error, I exclude those observations, which makes data adhere to this assumption.

- $E(D_{11}) - E(D_{10}) > E(D_{01}) - E(D_{00})$

  This assumption requires that the increase in the proportion of firms that receive treatment is greater for the treated group than for the control group. This assumption is satisfied but on the margins, i.e., the difference in probability of receiving treatment is only marginally greater for the treated sub-sample than the control sub-sample.

The Wald DiD statistic can be computed as below:

$$W_{DID} = \frac{E(Y_{11}) - E(Y_{10}) - \{E(Y_{01}) - E(Y_{00})\}}{E(D_{11}) - E(D_{10}) - \{E(D_{01}) - E(D_{00})\}}$$

It is important to note Wald DiD LATE statistic is a much narrower estimand as it estimates the treatment effect only on the sub-population of firms complying with the treatment[6]. Analogous to the instrumental variable framework, the DiD estimate derived using the baseline model can be interpreted as a reduced form DiD estimate. Because it is the reduced form estimate, the effect of increased diversity on firm performance is expected to be underestimated. The use of the formula above will yield a DiD estimate among compliers and can be intuitively interpreted as comparing two different difference-in-differences quantities – one that measures the impact on π divided by the one that measures the impact on the treatment uptake D. The W_{DID} can also be estimated using 2SLS regression. According to de Chaisemartin and DHaultfoeuille (2018), W_{DID} is the coefficient of D in a 2SLS regression of Y on D with Z and T as included instruments and Z x T as the excluded instrument[7].

Turns out, because $E(D_{11}) - E(D_{10}) - E(D_{01}) - E(D_{00}) \approx 9\%$, Z x T is a strong instrument with F-statistic substantially $\geq$ 10. In fact, the instrument is statistically significant in the first-order regression with an F-value of 288, satisfying the *relevance* criterion. This not only makes the LATE estimate reliable but also less susceptible to bias. Please note that I also include firm size as an additional covariate for 2SLS analysis to satisfy the *exclusion restriction* criterion. This is because if one reads into the provisions of ICA 2013, one would notice that the law specifically targets bigger firms. These firms are either listed or have threshold levels of shareholder capital and turnover. Therefore, the relationship between Z and Y is confounded because of firm size (S). See Figure 5 for the DAG. Below I present the 2-SLS regression results in the table below:

——————————

Insert Figure 5 here

——————————

[6]Using fuzzy-DiD has a distinct set of identifying assumptions over and above assumptions already made. It assumes that there is no treatment switch, i.e., once a firm becomes treated, it does not become untreated in the years that follow. Therefore, I exclude a small number of cases (1%) where treated firms become untreated before I estimate Wald DiD statistic.

[7]To clarify, here Z is a binary indicator for treatment assignment. It takes a value of 1 if a firm was required by law to appoint a woman director and 0 otherwise. On the contrary, D is a binary variable indicating actual treatment receipt.

—————————
Insert Tables 6 and 7 here
—————————

The 2-SLS estimates are similar to the estimates obtained using the baseline model except for the magnitude of the effect. Using 2-SLS yields a higher effect size (β_{2SLS} = -0.0341 vs β_{TWFE} = -0.0018), which is almost 20 times the effect observed using the traditional TWFE model. This is because 2-SLS estimates the effect amongst compliers in the treated group, i.e., for a small subset of firms included in the treated firms. Nonetheless, even the 2-SLS effect is not significant at the conventional levels of significance (z = -1.25; p = 0.211).

5.2 Matched DiD and Time-Varying Unobserved Confounding

Next, I try to control for unobserved time-varying confounding by creating a matched sample sample of treated and untreated firms. The underlying intuition is that matching on pre-treatment outcomes partially balances unobserved confounders, which can subsequently mitigate some bias resulting from unobserved time-varying confounding. From Figure 3, it becomes obvious that control firms are younger in age and smaller in size compared to treated firms. There is a possibility of time-varying confoundedness which may differentially impact the bigger and older firms differently as opposed to smaller and younger firms. To minimize some of these concerns, I implemented a matched DID using coarsened exact matching (CEM). The idea is to compare treated and untreated forms that are very similar to each other, especially in dimensions such as firm age, size, and family ownership. I also make the exact match on two-digit industry codes. This substantially reduces the number of treated firms in search of the best fit. The CEM procedure yields a matched sample of 471 treated and control firms balanced on all the required dimensions. See Figure 4 for covariate balance plots for treated and untreated firms after matching firm size, firm age, research intensity, and family using CEM. I also do exact matching on industry codes.

—————————
Insert Figure 4 here
—————————

However, unfortunately, as shown in Figure 6, the parallel trend assumption does not seem any more plausible for the matched sample of firms. Furthermore, the instrument Z x T is no longer relevant. Therefore, both conceptually and empirically matched DiD does not seem to work in the context of my data. I still present the matched DiD results below. The results remain non-significant just as the estimates obtained using the baseline model.

As the non-parametric approach fails, I assume that adjusting for size in the 2SLS regression specification controls for all the time-varying unobserved confounding that might be affecting the potential outcomes across treated and untreated.

—————————
Insert Figure 6 here
—————————

—————————
Insert Tables 8 here
—————————

5.3 Staggered DiD and Carryover Effects

So far in my analysis, I have precluded the possibility that treatment uptake could be staggered across firms as opposed to one-shot. Conceptualizing treatment as staggered can be useful in two important ways: first, it can help ascertain dynamic treatment effects over an extended horizon; and second; it does not assume treatment to be constant across different time periods. Thus, employing staggered DiD helps relax assumption 3 in my baseline model that states treatment effect to be constant and contemporaneous. This could also be perhaps one of the reasons that I do not observe any significant association between gender diversity and firm performance when collapsing the entire data to just two time periods. In this section, I shall use a multi-period DiD estimator. As established by Imai & Kim (2021), TWFE produces biased and inconsistent estimates in a multi-period framework. Therefore, I use the estimator as proposed by Callaway and Sant'Anna (2019) to estimate and report dynamic treatment effects assuming conditional parallel trends on the industry code and state of firm headquarters between treated and untreated units for each year[8].

Here the target estimand is the group-time average treatment effect or $ATT(g, t)$. Please note that, in line with the procedure laid out by Callaway and Sant'Anna (2019), I exclude always-treated units and use never-treated units as the benchmark category to estimate $ATT(g, t)$ where g indicates the cohort when the firm was first treated and t represents event time ranging from -9 to 9.

$$ATT(g,t) = \mathbb{E}\left[Y_t - Y_{g-1} \mid G_g = 1\right] - \mathbb{E}\left[Y_t - Y_{g-1} \mid C = 1\right]$$

Refer to Figure 6 for the dynamic effect event study plot that visualizes the treatment effect by length of exposure. I also derive a combined treatment effect using the dynamic aggregation method of Callaway and Sant'Anna (2019), which essentially takes a weighted average of individual treatment effects by the number of units treated every year or cohort.

$$\frac{1}{\kappa}\sum_{g=2}^{\tau}\sum_{t=2}^{\tau}\mathbf{1}\{g \le t\}\text{ATT}(g,t)P(G=g)$$

$$\text{where } \kappa = \textstyle\sum_{g=2}^{\tau}\sum_{t=2}^{\tau}\mathbf{1}\{g \le t\}P(G=g).$$

This expression weights group-specific effects by the size of the groups. Compared to TWFE and its variants, the findings for multi-period DiD are relatively encouraging at both individual and combined levels. The estimates indicate that the inclusion of women on corporate boards creates a positive and significant impact on firm performance at a 5% level of significance. The ATT coefficient of 0.0231 with SE of 0.0115 derived using dynamic aggregation is not only statistically significant but also numerically substantive as the appointment of women directors on corporate boards is expected to increase ROA by 2.31%.

————————————

Insert Figures 7 and 8 here

————————————

[8]The parallel trend assumption for multi-period DID is satisfied when treated and untreated units are conditioned on the time-invariant covariates such as industry code and state where the firm is headquartered. The p-value for the pre-test of parallel trends assumption using the *att_gt* command in R comes out to be 0.1277.

———————————

Insert Tables 9 and 10 here

———————————

Reconciliation of Findings

The non-staggered DiD yields non-significant but negative estimates for the causal effect as opposed to staggered DiD which implies a positive and significant causal effect of increased gender diversity on firm performance. It is obvious to wonder why these findings are inconsistent in terms of their directionality if not statistical significance. Here I elaborate on why this might be the case:

- *Difference in Estimands*: While I am computing ATT using both non-staggered baseline DiD and staggered DiD, I am fundamentally estimating two very different quantities. The baseline model including Wald's DiD takes into account compliance and defines treated firms as ones targeted by the law. Even though the baseline model assumes perfect compliance, the fuzzy DiD specification relaxes the assumption and accounts for imperfect compliance. On the contrary, staggered DiD is oblivious to the problem of compliance. I could not find a model in the existing literature that extrapolates the staggered DiD framework to cases fraught with non-compliance. To that extent, staggered DiD undertakes the analysis of treated firms, regardless of whether they were intended to be treated or not, and therefore not immune to selection bias. This is because firms that have eventually included women on corporate boards could be systematically different from firms that did not, risking causal interpretation of findings derived using the staggered DiD specification.

- *Difference in Comparison Groups*: This relates to the point above. For non-staggered DiD, I am implicitly assuming firms not qualifying under the law but which may qualify in future years. i.e., not yet assigned to treatment firms, and firms that never qualify, i.e., never assigned to treatment firms, as my comparison group. On the contrary, staggered DiD uses firms that never take treatment as its comparison group. These two conceptualizations yield two very different control groups. For non-staggered DiD, the control group is defined based on whether a firm is targeted by the law whereas for staggered DiD the analysis is based on the actual receipt of treatment. For non-staggered DiD therefore it becomes important, even though the conditional parallel trends assumption holds, for me to benchmark firms that receive treatment vis-a-vis firms that never receive treatment across each cohot to diagnose if these two categories of firms are systematically different from one another. However, even this kind of analysis may not be very precise because of potential post-treatment bias, i.e., the key descriptives of the treated firms could be influenced because of the treatment. Hence I abstain from doing it here.

- *Small Number of Never Treated Firms*: It is important to note that the number of never-treated firms based on actual treatment receipt or uptake is disproportionately smaller than the number of firms at least treated once. On average, the number of never-treated firms is one-tenth of the firms that are ever treated. Therefore, due to a substantially smaller number of never-treated firms, the estimates might not be very robust.

Considering all of these factors, I am not very confident about the positive causal effect estimates derived using the staggered DiD analysis. I, therefore, request readers to be careful as they interpret this estimate knowing it ignores the treatment assignment process and relies on a very small set of never-treated firms for its imputation.

6 Conclusion and Discussion

The empirical results derived using different models and specifications have not been consistent. In fact, the magnitude and direction of causal association, even though not statistically significant, fluctuate depending on the tweaks made to the underlying DiD framework. However, there are a few important points to bear in mind as we interpret these results:

- The results are non-significant throughout for non-staggered DiD with a two-year window. It is positive and significant at a 5% level of significance if we consider staggered DiD, suggesting one-shot treatment effect may not be very telling of the positive impact that gender diversity creates. This is because the classic DiD model requires me to collapse the entire spectrum of data into a two-period window and does not allow observance of long-term effects after the onset of treatment. Even the positive and significant ATT estimate obtained using staggered DiD analysis is subject to skepticism on account of a disproportionately smaller number of never-treated firms and its obvious disregard for the treatment assignment process.

- That said, the non-significant results for non-staggered DiD could mean two things – *first*, maybe there is no significant causal linkage between board diversity and immediate or short-term firm performance; and *second*, the data for the two-year fixed window is scarce leading to noisy ATT estimates. Both seem to be partly true.

  - It is theoretically conceivable that the inclusion of women in corporate boards creates value but in the slightly longer term as opposed to concurrently. Had my dependent variable been a market-based performance measure such as stock price, then the impact would have been immediate assuming investors discounted the future benefits of increased board diversity to reprice the stock rather instantaneously.

  - Even the standard error estimates are disproportionately bigger for TWFE-based estimators compared to the staggered DID estimates – indicating greater imprecision in ATT estimation in a two-period setup as opposed to a fully flexible assessment. See Tables 5, 6, 7, and 8.

- It is also important for me to reiterate that the gender of board directors is inferred using an AI algorithm, more specifically *genderize.io* API[9]. While API does a good job at classifying English names, I suspect it not being as accurate at classifying Hindi names. This could be one major source of measurement error as we should not be oblivious to to AI-induced bias.

- The number of never-treated firms is exceptionally small. Moreover, as almost all of these firms are unlisted, they are relatively much smaller in size and age. Further, I suspect that unobserved time-varying confounding could differentially affect never-treated and treated firms. This is the reason we observe a statistically significant impact on ROA, if not for all the leads, for the event time equal to -5.

[9]Of the 14,824 unique names, the API classified 11,650 names (i.e., 80%) with an average probability of 0.94 and 0.89 for male and female names respectively.

- Theoretically speaking, firm performance is a multi-dimensional construct. My analysis uses ROA, an accounting-based measure to assess firm performance. However, the use of ROA could be limiting for being a backward-looking accounting-based computational artifact. While I could have used alternate market-based performance benchmarks other than accounting-based performance variables for a more comprehensive analysis, none of the never-treated firms are publicly listed rendering such analysis impossible.

- Another important concern that could be raised is about the generalizability of my findings. I acknowledge while this focus allows for a detailed analysis of a specific legislative change, it might not extend to other regulatory environments and cultural contexts.

Considering these limitations, I believe there are plenty of possibilities for scholars to further build on this stream of work through their future research. Admittedly, I fail to arrive at a theoretically and empirically consistent ATT estimate. Instead, my findings are inconclusive to the extent they report estimates of different signs and magnitudes. Methodologically, I believe there is an opportunity to build a model that takes into account not only staggered adoption of treatment but also whether units are complying with the treatment assignment or not.

References

[1] Agarwal, P. (2022). Research Report on Leadership Gender Balance in NSE 200 Companies.

[2] Amore, M. D., Garofalo, O., & Minichilli, A. (2014). Gender interactions within the family firm. Management Science, 60(5), 1083-1097.

[3] Bahk, B.-H., & Gort, M. (2015). Decomposing Learning by Doing in New Plants. Journal of Political Economy, 101(4), 561–583.

[4] Bertrand, M., Mehta, P., & Mullainathan, S. (2002). Ferreting out Tunneling: An Application to Indian Business Groups. The Quarterly Journal of Economics, 117(1), 121-148.

[5] Brück, O. (2023). A bibliometric analysis of the gender gap in the authorship of leading medical journals. Communications Medicine, 3(1), 179.

[6] Callaway, B., & SantAnna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

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

[8] Catalyst. (2022, March 1). Women Business Leaders: Global Statistics.

[9] Chen, S., Ni, X., & Tong, J. Y. (2016). Gender Diversity in the Boardroom and Risk Management: A Case of RD Investment. Journal of Business Ethics, 136(3), 599–621.

[10] Chrisman, J. J., Chua, J. H., & Litz, R. A. (2004). Comparing the agency costs of family and non–family firms: Conceptual issues and exploratory evidence. Entrepreneurship Theory and practice, 28(4), 335-354.

[11] De Chaisemartin, C., & dHaultfoeuille, X. (2018). Fuzzy differences-in-differences. The Review of Economic Studies, 85(2), 999-1028.

[12] Garg, M. & Agarwal, P.(2022). Numbers and Beyond: Gender Equity in Corporate India at Board Level. Financial Express.

[13] Gomez-Mejia, L. R., Cruz, C., Berrone, P., De Castro, J. (2011). The bind that ties: Socioemotional wealth preservation in family firms. The academy of management annals, 5(1), 653-707.

[14] Hillman, A. J., Shropshire, C., & Cannella, A. A. (2007). Organizational Predictors of Women on Corporate Boards. Academy of Management Journal , 50(4), 941–952.

[15] Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with time-series cross-sectional data. American Journal of Political Science, 67(3), 587-605.

[16] Imai, K., & Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. Political Analysis, 29(3), 405-415.

[17] Kim, D., & Starks, L. T. (2016). Gender Diversity on Corporate Boards: Do Women Contribute Unique Skills? American Economic Review, 106(5), 267–271.

[18] Koffi, M. (2021). Innovative ideas and gender inequality (No. 35). Working Paper Series.

[19] Kossek, E. E., Su, R., & Wu, L. (2017). Opting out or pushed out? Integrating perspectives on womens career equality for gender inclusion and interventions. Journal of Management, 43(1), 228-254.

[20] Kulich, C., Trojanowski, G., Ryan, M. K., Alexander Haslam, S., & Renneboog, L. D. R. (2011). Who gets the carrot and who gets the stick? Evidence of gender disparities in executive remuneration. Strategic Management Journal, 32(3), 301–321.

[21] Lopes, P. D. (1992). Innovation and Diversity in the Popular Music Industry, 1969 to 1990. American Sociological Review, 57(1), 56.

[22] Mani, D., & Moody, J. (2014). Moving beyond stylized economic network models: The hybrid world of the Indian firm ownership network. American Journal of Sociology, 119(6), 1629-1669.

[23] Mishra, S., & Suar, D. (2010). Does Corporate Social Responsibility Influence Firm Performance of Indian Companies? Journal of Business Ethics, 95(4), 571-601.

[24] Mubarka, K., & Kammerlander, N. H. (2023). A closer look at diversity and performance in family firms. Journal of Family Business Management, 13(4), 828-855.

[25] Nishii, L. H. (2013). The benefits of climate for inclusion for gender-diverse groups. Academy of Management Journal, 56(6), 1754-1774.

[26] Post, C., & Byron, K. (2014). Women on Boards and Firm Financial Performance: A Meta-Analysis. Academy of Management Journal, 58(5), 1546–1571.

[27] Pitts, R. A., & Hopkins, H. D. (1982). Firm Diversity: Conceptualization and Measurement. Academy of Management Review, 7(4), 620–629.

[28] Terjesen, S.,& Sealy, R. (2016). Board gender quotas: Exploring ethical tensions from a multi-theoretical perspective. Business Ethics Quarterly, 26(1), 23-65.

[29] UN Women. (n.d.). Facts and Figures: Women's Leadership and Political Participation. Retrieved March 29, 2022, from https://www.unwomen.org/en/what-we-do/facts-and-figures.

[30] Villalonga, B., Amit, R., Trujillo, M. A., & Guzmán, A. (2015). Governance of family firms. Annual Review of Financial Economics, 7, 635-654.

Table 1: Regression Results with using 1, 2, and 3-year lagged values of π as placebo outcomes

	Model 1†	Model 2†	Model 3†
	π_{t-3}	π_{t-2}	π_{t-1}
treat#post	-0.001		
	(-0.68)		
treat#post2		-0.001	
		(-0.60)	
treat#post3			-0.001
			(-0.21)
constant	0.018***	0.018***	0.018***
	(25.33)	(15.77)	(10.39)
N	8382	8382	8382

†Models use standard DID two-way fixed effects specification.
$^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$

Table 2: Assumptions and Supplementary Analyses[a]

Assumptions	**Robustness Checks**
2	Instrumented/Fuzzy DiD
3	Staggered DiD with dynamic treatment effects
4	Matched DiD

[a]Assumptions 2 to 4 are surely quite restrictive. If we look at the data, it will become clearer that Assumption 4 does not hold as both treated and control firms, based on treatment assignment, receive treatment. Similarly, assumption 3 is also quite limiting as it assumes that increased diversity will not have carryover effects, i.e., treatment in a given year will not affect firm performance in future periods. Further as shown in Figure 3, even assumption 4 is less likely to hold considering treated and untreated firms are significantly different from each other based on firm size and age. Cumulatively speaking, these assumptions do not seem quite realistic. Therefore, I relax each of these assumptions, one at a time, by estimating a different model to test the robustness of my baseline estimates.

Table 3: Summary Statistics of Outcome Variables and Covariates

Variable	N	Mean	Std. Dev.	Min	Max
ROA	32983	0.006	0.134	-0.87	0.291
Firm Size*	33024	6.467	2.573	1.797	13.157
Firm Age*	33026	3.309	0.583	0	5.056
Research Intensity	33024	0.002	0.009	0	0.066
Family (1: Yes; 0: No)†	33026	0.865	-	0	1

* Variables measured using natural log
†SD omitted due to lack of its meaningfulness for a binary indicator such as family

Table 4: Pairwise correlations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ROA	1.000	0.296***	0.031***	0.079***	−0.005
(2) Firm Size		1.000	0.125***	0.238***	−0.139***
(3) Firm Age			1.000	0.049***	−0.070***
(4) Research Intensity				1.000	0.006
(5) Family (1: Yes; 0: No)					1.000

$^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$

Table 5: Difference-in-differences regression using TWFE model

	Coefficient	Std. Err.	t	P > |t|
ATET DiD (1 vs 0)	-0.0018	0.0016	-1.15	0.250

† DiD regression estimates using two-way fixed effects for FY 2014 and 2015

Note: Standard error adjusted for 4,208 clusters for firms.
Number of Observations = 8,382.

Table 6: 2SLS First-order Regression Output with Robust Standard Errors

Variable	Coefficient	Robust Std. Err.	t	P>|t|
Z x Post	0.0927	0.0244	3.79	0.000
Z	0.0542	0.0183	2.95	0.003
Post	0.1430	0.0226	6.32	0.000
Firm Size[a]	0.0297	0.0021	13.81	0.000
Constant	0.3392	0.0194	17.41	0.000

Note: Number of obs = 10,423, F(4, 10418) = 288.03
Prob > F = 0.0000, Root MSE = .4422
Included Instruments: Z, Post, Firm Size
Excluded Instrument: Z x Post

[a] Firm size is included as a covariate because it confounds the relationship between Z and ROA as presented in DAG in Figure 4.

Table 7: 2SLS Second-order Regression Output with Robust Standard Errors

Variable	Coefficient	Robust Std. Err.	t	P>|t|
$\hat{D}$	-0.0341	0.0273	-1.25	0.211
Z	-0.0118	0.0030	-3.87	0.000
Post	0.0070	0.0060	1.18	0.238
Firm Size[a]	0.0067	0.0008	7.87	0.000
Constant	-0.0019	0.0084	-0.23	0.821

Note: Number of obs = 10295, F(4, 10290) = 121.24
Prob > F = 0.0000, Root MSE = .0515
Included Instruments: Z, Post, Firm Size
Excluded Instrument: Z x Post

[a] Firm size is included as a covariate because it confounds the relationship between Z and ROA as presented in DAG in Figure 4.

Table 8: 2SLS Second-order Regression Output with Robust Standard Errors

Variable	Coefficient	Robust Std. Err.	t	P>|t|
$\hat{D}$	-0.0279	0.0577	-0.48	0.629
Z	0.1228	0.0401	3.06	0.002
Post	0.0205	0.0514	0.40	0.689
Constant	0.4031	0.0355	11.35	0.000

Note: Number of obs = 1808, F(3, 1804) = 4.85
Prob > F = 0.0023, Root MSE = .4985

Included Instruments: Z, Post
Excluded Instrument: Z x Post

Table 9: Overall Summary of ATT's Based on Event-Study/Dynamic Aggregation[†]

ATT	Std. Error	[95% Conf. Int. Low]	[95% Conf. Int. High]
0.023	0.0125	−0.0016	0.0475

[†] Unlike TWFE and its many variants estimated for my analysis, the findings for multi-period DiD are relatively encouraging at both individual and combined levels. The estimates indicate that the inclusion of women on corporate boards creates a positive and significant impact on firm performance at a 10% level of significance. The ATT coefficient of 0.0230 with SE of 0.0129 derived using dynamic aggregation is not only statistically significant but also numerically substantive as the appointment of women directors on corporate boards is expected to increase ROA by 2.30%.

Table 10: Dynamic Effects: Doubly Robust Estimation of Event Time

Event Time	Estimate	Std. Error	95% CI Lower	95% CI Upper
−9	−0.0482	0.0603	−0.2046	0.1083
−8	0.0476	0.0545	−0.0939	0.1891
−7	−0.0223	0.0457	−0.0745	0.2299
−6	−0.0116	0.0186	−0.0599	0.0366
−5	0.0291	0.0141	−0.0075	0.0658
−4	−0.0018	0.0103	−0.0286	0.0249
−3	−0.0014	0.0096	−0.0264	0.0237
−2	0.0034	0.0076	−0.0164	0.0232
−1	0.0004	0.0059	−0.0150	0.0159
0	0.0024	0.0157	−0.0190	0.0627
1	0.0263	0.0116	−0.0038	0.0564
2	0.0358	0.0161	−0.0060	0.0776
3	0.0127	0.0200	−0.0393	0.0646
4	0.0173	0.0164	−0.0254	0.0599
5	0.0100	0.0134	−0.0248	0.0449
6	0.0324	0.0172	−0.0123	0.0772
7	0.0259	0.0218	−0.0308	0.0826
8	0.0149	0.0226	−0.0437	0.0735
9	0.0324	0.0278	−0.0398	0.1047

Significance codes: '*' confidence band does not cover 0
Control Group: Never Treated, Anticipation Periods: 0

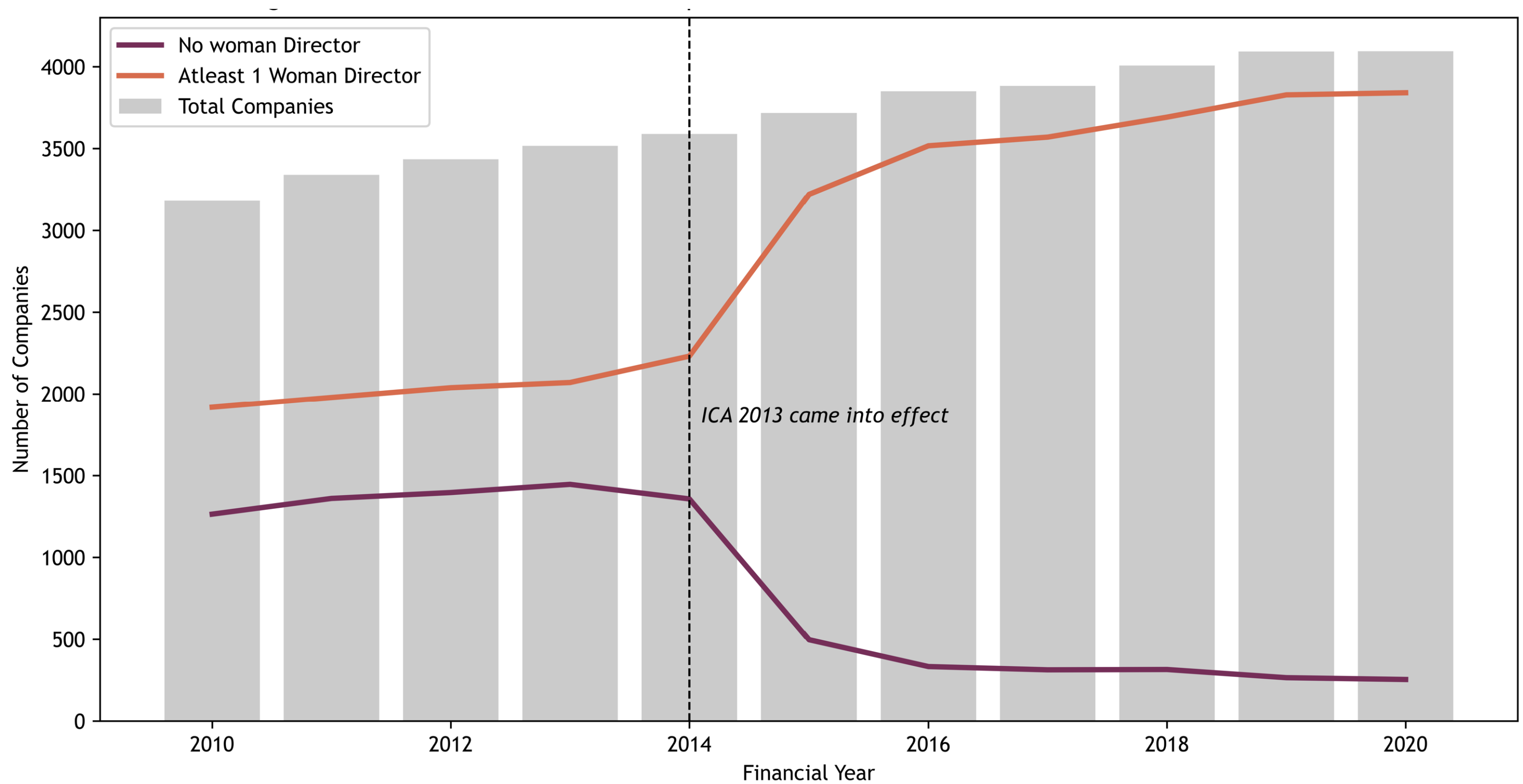

Figure 1: Number of Indian Companies with at least one woman director for FY 2010-2020

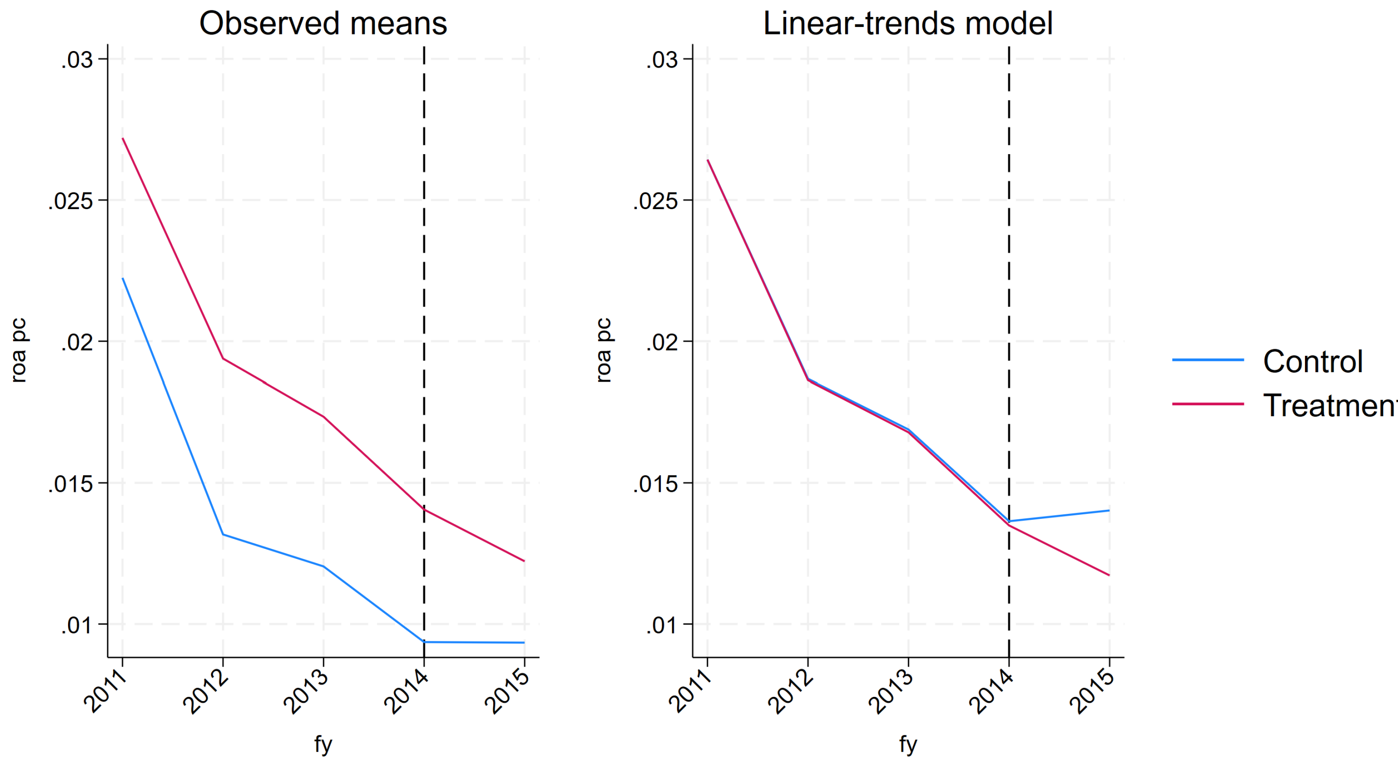

Figure 2: Visualization Check for Parallel Trends for baseline model using TWFE

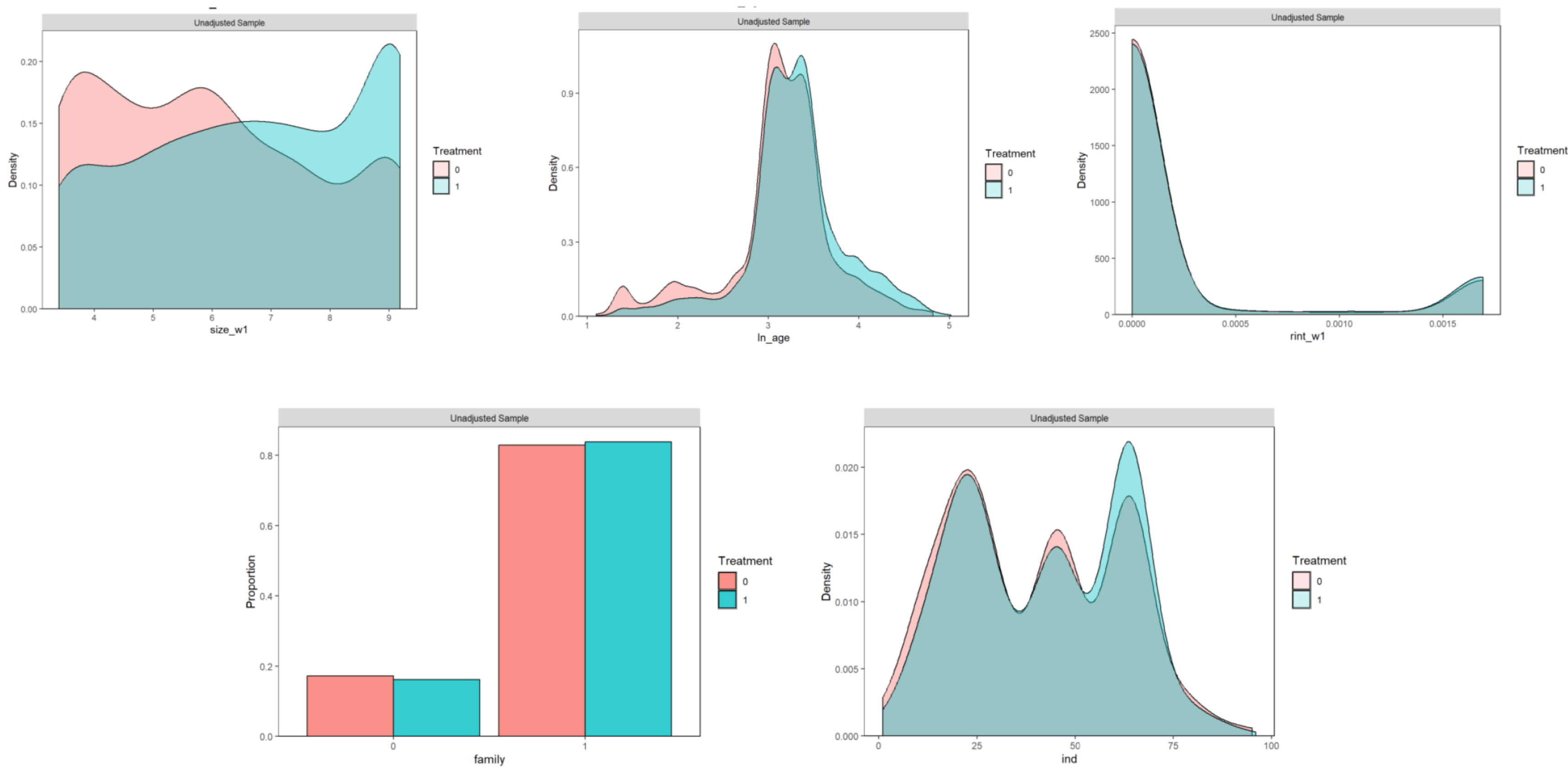

Figure 3: Covariate Balance Plots for treated and untreated observations before matching

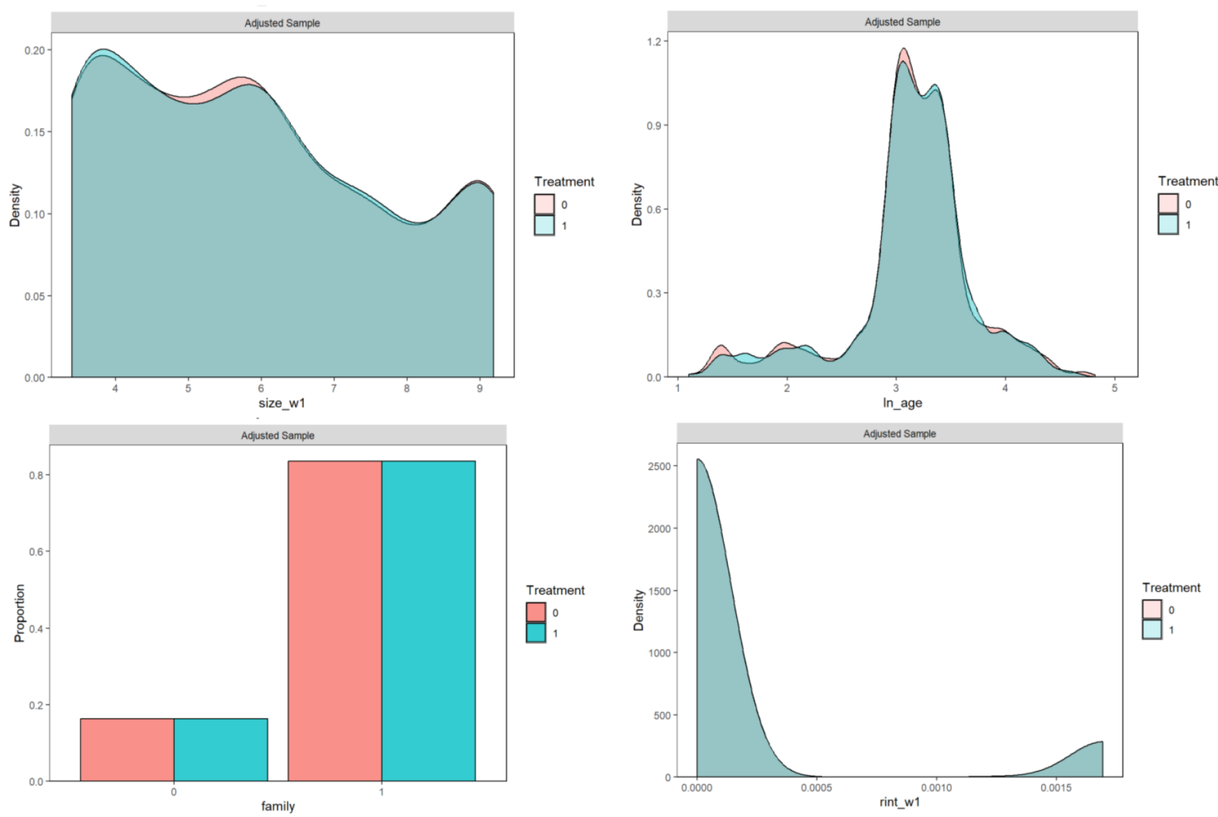

Figure 4: Covariate Balance Plots for treated and untreated observations after matching using CEM. Please note that I also do exact matching on two-digit industry codes. That is, treated and untreated observations are balanced across industry types.

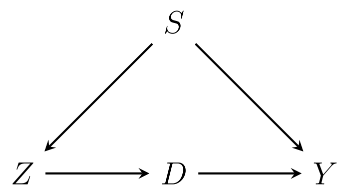

Figure 5: Directed Acyclic Graph (DAG) illustrating causal relationships among variables Treatment Assignment (Z), Treatment Receipt (D), ROA (Y), and Firm Size (S).

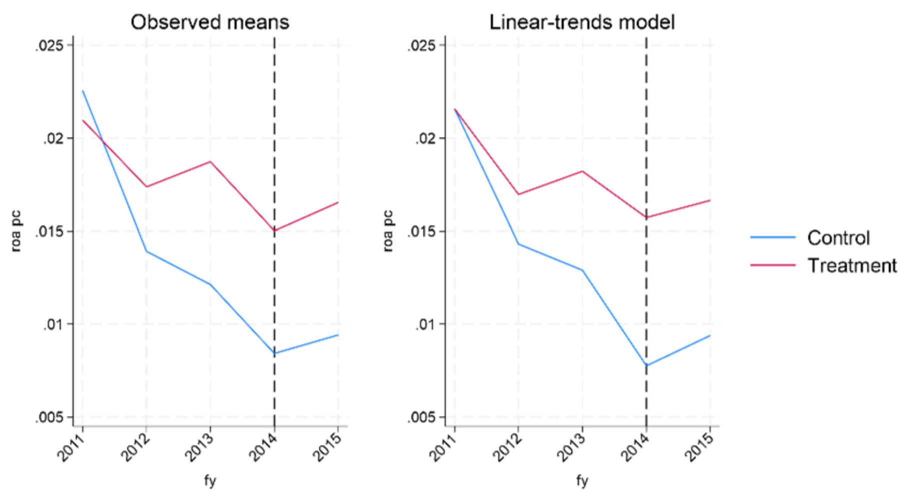

Figure 6: Visualization Check for Parallel Trends for TWFE model using CEM Matched Sample

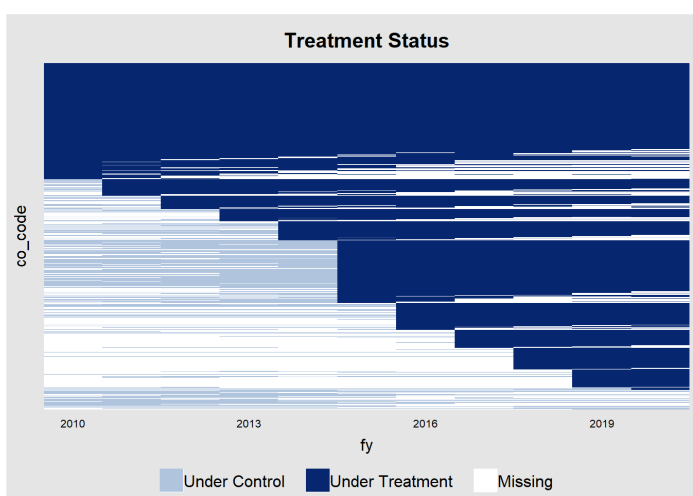

Figure 7: Treatment Status Plot for Staggered DiD

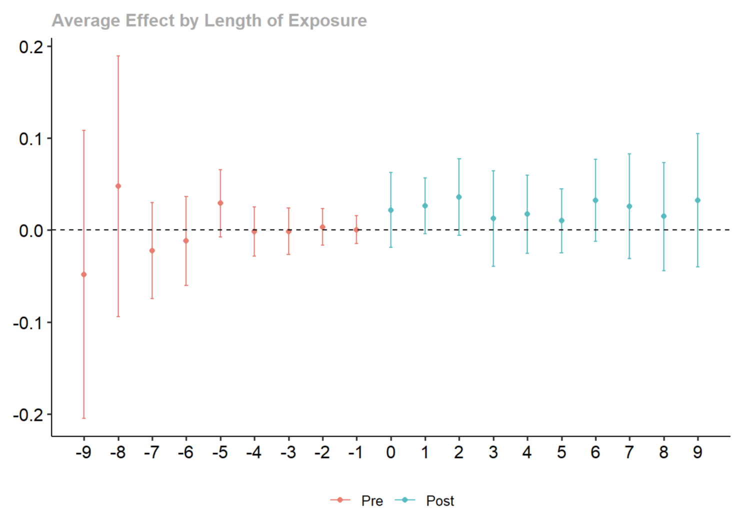

Figure 8: Event Study Plot with dynamic treatment effects for staggered DiD without treatment reversal